

PROBABILITY AND STATISTICS FOR ENGINEERS

Roberto Mínguez

December 17, 2012

Contents

1	Introduction to Probability Theory	1
1.1	Introduction	1
1.2	Probability	2
1.3	Axioms of Probability	4
1.4	Probabilistic Models	8
1.5	Univariate Random Variables	8
1.5.1	Probability Mass Function	8
1.5.2	Probability Density Function	9
1.5.3	Cumulative Distribution Function	10
1.5.4	Moments	12
1.5.5	Survival and Hazard Functions	16
1.5.6	Characteristic Function	18
1.6	Common Discrete Univariate Models	22
1.6.1	Bernoulli Random Variables	23
1.6.2	Binomial Random Variable	25
1.6.3	Geometric or Pascal Distribution	28
1.6.4	Negative Binomial Distribution	29
1.6.5	Hypergeometric Distribution	32
1.6.6	Poisson Distribution	34
1.6.7	Non-Zero Poisson Random Variable	37
1.7	Common Continuous Univariate Models	37
1.7.1	Uniform Distribution	38
1.7.2	Exponential Distribution	39
1.7.3	Gamma Random Variable	41
1.7.4	Beta Random Variable	44
1.7.5	Normal or Gaussian Distribution	46
1.7.6	Log-Normal Distribution	50
1.7.7	Logistic Distribution	52
1.7.8	Chi-Square and Chi Random Variables	54
1.7.9	Rayleigh Random Variable	55
1.7.10	Student's t Distribution	56
1.7.11	Extreme Value Distributions	57
1.7.12	Gumbel or Type I Extreme Value Distribution	60
1.7.13	Fréchet or Type II Extreme Value Distribution	62

1.7.14	Weibull of Type III Extreme Value Distribution	62
1.7.15	Generalized Extreme Value Distributions	63
1.7.16	Generalized Pareto Distributions	64
1.8	Truncated Distributions	64
1.9	Exercises**	67
2	Measures of Structural Reliability	71
2.1	Introduction	71
2.2	Deterministic Measures	73
2.2.1	Factor of safety	73
2.2.2	Load factor	75
2.2.3	Partial safety factor	76
2.2.4	Lack of invariance in some safety measures	77
2.2.5	Invariant safety measures	79
2.3	The Return Period: a Partial Probabilistic Safety Measure	80
2.4	Probabilistic Measures	82
2.4.1	Introduction	82
2.4.2	The basic reliability problem	84
2.4.3	Safety factors and characteristic values	88
2.4.4	Generalized reliability problem	91
3	Regression Analysis	97
3.1	Introduction	97
3.2	Simple Linear Regression Analysis	98
3.3	Simple Linear Regression Estimation	102
3.4	Residual Analysis	103
3.5	Multiple Linear Regression	105
3.5.1	Parameter confidence intervals	107
3.5.2	Residuals	110
3.6	Bad Data Detection	110
3.7	Maximum Likelihood Inference Method	111
3.8	Stepwise Regression	111
3.9	Linear Regression Using Transformed Models	112
3.10	Non-Linear Regression	113
3.10.1	A Model for Ultrasonic Data	114
3.10.2	A Model for Onion Yield Data	114
3.10.3	A Model for Calibration Data	115
3.11	Regression Sensitivity Analysis	118
3.11.1	Background	119
3.11.2	Least Squares Sensitivity Analysis	121
3.11.3	A Numerical Example	123
3.12	Minimax Regression	126
3.12.1	The Primal Minimax Regression Problem	126
3.12.2	A Numerical Example	130
3.12.3	The Dual Minimax Regression Problem	132

3.13	The Least-Absolute-Value (LAV) Regression	134
3.13.1	The Primal LAV Regression Problem	134
3.13.2	A Numerical Example	137
3.13.3	The Dual LAV Regression Problem	139
4	Structural Reliability Methods I	141
4.1	Introduction	141
4.2	The Tail Problem	143
4.3	Integration Methods	143
4.4	Simulation Methods	143
4.4.1	Importance sampling	144
4.4.2	Directional simulation	145
4.5	Second-Moment Methods	145
4.6	First-Order Second-Moment (FOSM) theory	146
4.6.1	Hasofer-Lind reliability index	146
4.6.2	Hasofer-Lind transformation	150
4.6.3	Limit state equation	151
4.6.4	Sensitivity factors	155
4.6.5	Numerical solution for non-linear limit state functions	155
4.6.6	Failure probability bounds	157
4.7	First-Order Reliability Methods (FORM)	158
4.7.1	Variable transformations	158
4.7.2	FORM algorithm	161
4.7.3	Rubblemound breakwater example	167
4.8	Second-Order Reliability-Methods (SORM)	174
4.8.1	Methods	174
4.9	Reliability Based Optimization	176
4.9.1	State of the art review	177
5	Structural Reliability Methods II. Structural Systems.	179
6	Duality	181
6.1	Introduction	181
6.2	Karush-Kuhn-Tucker First Order Optimality Conditions	181
6.3	Duality in Linear Programming	188
6.3.1	Obtaining the Dual from a Primal in Standard Form	189
6.3.2	Obtaining the Dual Problem	191
6.3.3	Duality Theorems	193
6.4	Duality in Nonlinear Programming	199
6.5	Illustration of Duality and Separability	213
6.6	Concluding Remarks	217
6.7	Exercises	217
7	Optimization Algorithms for Structural Reliability	223

8	Local Sensitivity Analysis	225
8.1	Introduction	225
8.2	A General Method for Sensitivity Analysis	225
8.2.1	Application Example: The Engineering Design.	228
8.3	Sensitivity in Constrained Optimization	231
8.4	Illustrative Examples	235
8.5	Exercises	252
9	Decomposition Techniques	255
9.1	Introduction	255
9.2	Benders Decomposition	256
9.2.1	Introduction	256
9.2.2	Benders Decomposition	256
9.3	Complicating Constraints Decomposition	261
9.3.1	Introduction	261
9.3.2	Complicating Constraints	261
9.3.3	Lagrangian relaxation	262
9.3.4	Augmented lagrangian relaxation	266
9.3.5	Optimality condition decomposition	266
9.3.6	Application to reliability problems	267
9.4	Complicating Variables and Constraints	267
9.5	Optimization Packages	268
9.5.1	The generalized reduced gradient method	269
9.6	Alternative Decomposition Methods	273
9.6.1	Relaxation Method	274
9.6.2	Hyperplane reconstruction approximation	280
9.7	Sensitivity Analysis	284
9.7.1	Introduction	284
9.7.2	Sensitivity in constraint optimization. General method.	284
10	Reliability Based Optimization Techniques (RBOT)	289
10.1	Introduction	289
10.2	Reliability Based Optimization Problems	289
10.2.1	Variable classification	290
10.2.2	Structural optimization problems subject to reliability constraints	291
10.3	RBO using Benders decomposition	298
10.3.1	Numerical example	306
10.4	The Probability-Safety Factor Method	306
10.4.1	Relaxation method	311
10.4.2	Hyperplane approximation method	316
10.5	Global and Partial Model Resolution	320
10.6	Sensitivity Analysis	322
10.6.1	Sensitivity analysis in the probability-safety factor method (PSFM)	323
10.6.2	Sensitivity analysis in the general model	326
10.7	Application: Bridge Crane Design	327

10.7.1	Obtaining some constraints	330
10.7.2	Numerical example	332
11	Applications	337
11.1	Retaining Wall	337
11.1.1	Introduction to the retaining wall design	337
11.1.2	Description of the retaining wall	338
11.1.3	Variable classification and failure modes considered	340
11.1.4	Forces acting on the wall	342
11.1.5	Design criteria and safety factor requirements	344
11.1.6	Objective function	348
11.1.7	Distributional assumptions of the model	348
11.1.8	Numerical solution and discussion	348
11.1.9	Sensitivity analysis	351
11.1.10	Monte Carlo simulation	354
11.2	Compound girder	354
11.2.1	The compound girder problem	354
11.2.2	Constraints	357
11.2.3	Objective function	358
11.2.4	Distributional assumptions of the model	359
11.2.5	Numerical example	360
11.2.6	Monte Carlo simulation.	362
11.2.7	Sensitivity analysis	363
11.2.8	Compound girder constraints	363
11.2.9	Notation	371
11.3	Rubble-Mound Breakwaters	375
11.3.1	Variable classification and failure modes considered	375
11.3.2	Design criteria	380
11.3.3	Objective function	381
11.3.4	Distributional assumptions for the model	381
11.3.5	Formulation and solution of the dual problem	384
11.3.6	Numerical example and analysis of the results	386
11.3.7	Monte Carlo simulation	391
11.3.8	Notation	392
11.4	Slope Stability	395
11.4.1	Calculus of variation approach	396
11.4.2	Discretization of the problem	398
11.4.3	Classical approach	400
11.4.4	Modern approach	403
11.4.5	The probability-safety factor method	407
11.4.6	Sensitivity analysis	409

A GAMS Codes	413
A.1 GAMS Code for the Relaxation Method	413
A.2 GAMS Code for the Hyperplane Approximation	417
A.3 GAMS Code for the Benders Decomposition	420
A.4 GAMS Code for the Retaining Wall Using the Relaxation Method	425
A.5 GAMS Code for the Hyperplane Approximation Method	429
A.6 GAMS Code for the Sensitivity Analysis	433
 Bibliography	 441

Chapter 1

Introduction to Probability Theory

1.1 Introduction

The concept of reliability, probability, or risk has been a thought-provoking subject for a long time. Pierre Simon, Marquis de Laplace [?] wrote a commentary on general intelligence, published as “A Philosophical Essay on Probabilities”, where he said: “*I present here without the aid of analysis the principles and general results of this theory, applying them to the most important questions of life, which are indeed for the most part only problems of probability. Strictly speaking it may even be said that nearly all our knowledge is problematical; and in the small number of things which we are able to know with certainty, even in the mathematical sciences themselves, the principal means of ascertaining truth—induction, and analogy—are based on probabilities; so that the entire system of human knowledge is connected with the theory set forth in this essay*”. He concluded, “*It is seen in this essay that the theory of probabilities is at bottom only common sense reduced to calculus; it make us appreciate with exactitude that which exact minds feel by a sort of instinct without being able oftentimes to give a reason for it. It leaves no arbitrariness in the choice of opinions and sides to be taken; and by its use can always be determined the most advantageous choice. Thereby it supplements most happily the ignorance and weakness of the human mind*”.

Obviously, Laplace’s concepts can be apply to civil engineering. Most observable phenomena in engineering systems contain a certain amount of uncertainty; that is, they can not be predicted with certainty. In fact, in engineering practice, variability, scatter, dispersion, or, uncertainty in the quantities of interest is the rule rather than the exception. Induced loadings to civil engineering systems are never completely known. Among there are uncertainties with respect to the frequency and intensity of earthquakes; the flow of surperficial water and groundwater and of toxic and hazardous materials; the action and variability of wind and waves, heat and cold, freezing and thawing, chemical and environmental factors; vibrations and shock; vehicular and pedestrian traffic; its distribution and weights; construction equipment, and processes. All engineering materials contain microcrystalline imperfections called ‘*dislocations*’, which initiate cracks or permit their propagation. No two manufactured objects can be exactly the same. Therefore, many physical quantities must be considered uncertain.

The trend in civil engineering today, more than ever before, is toward providing eco-

nomical designs at specified levels of safety, and in the development of engineering designs, decisions are often required irrespective of the state of completeness and quality of information, so that, the consequence of a given decision cannot be determined with complete confidence. For these reasons, decisions are made under conditions of uncertainty.

The effects of such uncertainty on design and planning are important and can induce to wrong decisions that can cause even the failure of the design with considerable economic and social consequences. To improve the confidence of the designs the proper quantification of the uncertainty, and evaluation of its effects should be assessed including the concepts and methods of probability.

In this light, we see that the role of probability is of primary importance in engineering, ranging from the description of information to the development of bases for design and decision making, supplementing the civil engineer's judgment.

1.2 Probability

In general, repeated measurements of physical phenomena generate multiple outcomes with different occurrence frequencies. The occurrence of multiple outcomes without any pattern is described by terms such as *uncertainty*, *randomness*, and *stochasticity*. Most of the parameters of interest in engineering analysis and design have some degree of uncertainty and thus may be considered to be random variables.

When we talk about a random variable, it is helpful to think of an associated *random experiment* or *trial*. A random experiment or trial can be thought of as any activity that will result in one and only one of several well-defined outcomes, but one does not know in advance which one will occur. The set of all possible outcomes of a random experiment E , denoted by $S(E)$ is called the *sample space* of the random experiment E and each individual outcome is a *sample point*. If the sample space contains a set of all possible real numbers in a certain interval, it is called *continuous* sample space, whereas if the sample points are discrete and countable entities it is called *discrete*.

A subset of a sample space is called an event. An event is therefore a set of sample points. If it contains no sample points it is called an *impossible* event and if it contains all the sample points in the sample space is called *certain* event.

Illustrative Example 1.1 (Simply supported beam). Consider the simply supported beam AB shown in Figure 1.1 with one concentrated force $P_1 = 10 \text{ kN}$ that can be placed everywhere along the beam. In this case, the reactions on the supports A and B , R_A and R_B , varies from 0 to 10. The sample space of both reactions are $S(R_A) = (0, 10)$ and $S(R_B) = (0, 10)$, respectively. Note that both are continuous sample spaces because reactions can take any value in the sample space.

An event of interest then may be the reactions in some specified interval; for example, $(R_A \geq 7), (1 \leq R_A \leq 5), (6 \leq R_B \leq 9)$ or $(R_B \leq 3)$. ■

Within the context of engineering usage there are two primary definitions of the concept of probability: *relative frequency* and *subject* interpretation.

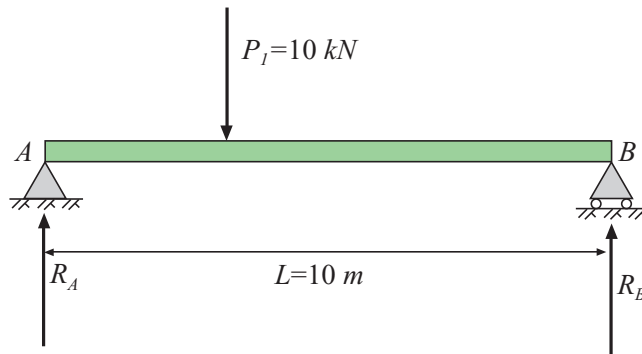


Figure 1.1: Simple supported beam.

Historically, the first measured of the probability of an outcome A was its relative frequency, thus if an outcome occurs T times in N trials, its probability is

$$P(A) = \frac{T}{N}, \quad (1.1)$$

where P is the probability operator, that implies that the probability of an outcome A equals the number of outcomes favorable to A divided by the total number of possible outcomes (size of the sample space). This is the frequency interpretation.

But in engineering applications structures will be built only once, and it will either fail or be successful during its design lifetime, it cannot be both. Thus the concept of repeated trials is meaningless, here it is applied the subjective interpretation of probability. That can be considered as a numerical measure of the likelihood that an event occurs relative to the set of all possible events.

The determination of the probability that an event occurs can be based on:

1. A priori assumptions about the underlying mechanisms governing events.
2. Relative frequency of empirical observations in the past.
3. Intuitive or subjective assumptions.
4. Combination of intuitive or subjective assumptions with experimental observations using the Bayes' theorem, whose result is called *Bayesian probability*.

Hence any probability depends on the state of knowledge or ignorance (or more generally 'the state of nature') at the time that the probability is calculated. Seen in this way, all probabilities are conditional (e.g. Tribus [?]).

Suppose that to finish a project's scheduled completion date (e.g., a bridge) at a certain stage of construction, the weather conditions can be classified into one of three categories: bad, normal, or good. Contractor examines the weather each day to assess its condition. This is a random experiment and its sample space, $S(E) = \{bad, normal, good\}$, has three elements.

Definition 1.1 (Random variable) A random variable can be defined as a real-valued function defined over a sample space of a random experiment. That is, the function assigns a real value to every element in the sample space of a random experiment. The set of all possible values of a random variable X , denoted by $S(X)$ is called the support or range of the random variable X .

Illustrative Example 1.2 (Weather condition). In the previous weather condition example, let X be $-1, 0$, or 1 , depending on whether is bad, normal, or good, respectively. Then X is a random variable with support $S(X) = \{-1, 0, 1\}$. The condition of the weather can also be assessed using a continuous scale, say, from 0 to 10 , to measure the weather quality, with 0 indicating the worst possible condition and 10 indicating the best. Let Y be the assessed condition of the structure. Then Y is a random variable with support $S(Y) = \{y : 0 \leq y \leq 10\}$. ■

We consistently use the customary notation of denoting random variables by uppercase letters such as X, Y , and Z or X_1, X_2, \dots, X_n , where n is the number of random variables under consideration. Realizations of random variables (that is, the actual values they may take) are denoted by the corresponding lowercase letters such as x, y , and z or x_1, x_2, \dots, x_n .

1.3 Axioms of Probability

Axiom I The probability of an event E is a real non-negative number:

$$0 \leq P(E) \leq 1. \quad (1.2)$$

Axiom II The probability of an inevitable event E is $P(E) = 1$. Whereas the probability of an impossible event is 0 .

Axiom III Addition rule. The probability of two events E_1 and E_2 occurs is

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2). \quad (1.3)$$

Hence, if the events are *mutually exclusive*, i. e., they can not occur simultaneously

$$P(E_1 \cup E_2) = P(E_1) + P(E_2). \quad (1.4)$$

Consider the design of an structure. After construction, only one of two outcomes can occur in the absolute structural sense: either it is successful or it fails. These are mutually exclusive outcomes, and *exhaustive* within the sense that no other outcomes are possible. Hence, the second axiom, requires

$$P(\text{success} \cup \text{failure}) = 1$$

And since they are mutually exclusive, the third axiom (1.4) specifies that

$$P(\text{success}) + P(\text{failure}) = 1.$$

Illustrative Example 1.3 (Bridge). Suppose that a bridge can just been subject to three types of surcharges A, B and C . A has half the chance that B has; B is two-thirds as likely as C to act under the structure. What is the probability for each to act if only one can be apply at the same time and there is always one surcharge acting on the structure?.

Since there are only three surcharges and one must be acting,

$$P(A \cup B \cup C) = 1.$$

As only one can act at the same time, they are mutually exclusive so

$$P(A) + P(B) + P(C) = 1.$$

Given $P(A) = P(B)/2$ and $P(B) = 2P(C)/3$, we have

$$P(B)/2 + P(B) + 3P(B)/2 = 1$$

whence, $P(A) = 1/6$, $P(B) = 1/3$ and $P(C) = 1/2$. ■

From the above axioms, the following derived results can be obtained:

1. Multiplication rule. The probability that events A and B both occur is

$$P(A \cap B) = P(A|B)P(B), \quad (1.5)$$

where $P(A|B)$ is the conditional probability of an outcome A , given that an outcome B has occurred.

If A and B are *independent*, i.e., if the occurrence or nonoccurrence of one has no effect on the probability of occurrence of the other, then

$$P(A \cap B) = P(A)P(B). \quad (1.6)$$

These equations can be generalized for a number n of events A_1, A_2, \dots, A_n

$$\begin{aligned} P(A_1 \cap A_2 \cap \dots \cap A_n) &= P(A_1)P(A_2 \cap \dots \cap A_n|A_1) = \\ &= P(A_1)P(A_2|A_1)P(A_3 \cap \dots \cap A_n|A_1 \cap A_2) = \\ &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1}) \end{aligned} \quad (1.7)$$

and if they are *independents*

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2)P(A_3) \dots P(A_n). \quad (1.8)$$

2. If \bar{E} is the event that E does not occur (its *complement*),

$$P(\bar{E} \cup E) = P(\bar{E}) + P(E) = 1$$

therefore,

$$P(\bar{E}) = 1 - P(E). \quad (1.9)$$

3. Conditional probability. It follows directly from (1.5) that

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (1.10)$$

4. Total probability theorem. Applying the multiplication rule it follows that for events A_i ; $i = 1, \dots, n$, mutually exclusive (i.e, covering all possibilities without overlapping),

$$P(E) = P(E|A_1)P(A_1) + P(E|A_2)P(A_2) + \dots + P(E|A_n)P(A_n). \quad (1.11)$$

5. Bayes' theorem. From the multiplication theorem, and considering a (set of) particular event(s) A_i and the event E ,

$$P(A_i \cap E) = P(A_i|E)P(E) = P(E|A_i)P(A_i), \quad (1.12)$$

but as $P(E)$ is (1.11)

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{P(E)} = \frac{P(E|A_i)P(A_i)}{\sum_{i=1}^n P(E|A_i)P(A_i)}. \quad (1.13)$$

6. Morgans' rule. The complement of the union of several events is equal to the intersection of the respective complements. Considering n events

$$\overline{E_1 \cup E_2 \cup \dots \cup E_{n-1} \cup E_n} = \overline{E_1} \cap \overline{E_2} \cap \dots \cap \overline{E_{n-1}} \cap \overline{E_n} \quad (1.14)$$

or, alternatively, the complement of the intersection is equal to union of the respective complements

$$\overline{E_1 \cap E_2 \cap \dots \cap E_{n-1} \cap E_n} = \overline{E_1} \cup \overline{E_2} \cup \dots \cup \overline{E_{n-1}} \cup \overline{E_n}. \quad (1.15)$$

Illustrative Example 1.4 (Quality control). A quality control program based on concrete cylinder tests is set up to inspect if 28-days unconfined compressive strength is not less than specified standards. Suppose that plant P_1 produces 60% and P_2 40% of the concrete used, and that 90% of the concrete produced by plant P_1 met the standard, whereas 95% did so from plant P_2 .

Thus, from these data, we know that the probability of concrete being produced by plant 1 is $P(P_1) = 0.6$, and by plant 2 is $P(P_2) = 0.4$. The probability to meet the standards if concrete is produced by plant 1 is $P(M|P_1) = 0.9$, whereas the same probability for plant 2 is $P(M|P_2) = 0.95$. Such probabilities are said to be conditionals.

Suppose that the concrete produce by both plants is used for constructing a bridge. If we take a sample cylinder, what is the probability of meet the standards and being from plant 1?

From equation (1.5) we have that

$$P(M \cap P_1) = P(M|P_1)P(P_1) = 0.9 \cdot 0.6 = 0.54.$$

What is the probability of meeting the standards?. Considering the total probability theorem (1.11),

$$P(M) = P(M|P_1)P(P_1) + P(M|P_2)P(P_2) = 0.9 \cdot 0.6 + 0.95 \cdot 0.4 = 0.92,$$

this indicates that, on average, 92 tests per 100 will prove adequate. Whereas from (1.9) we obtain that the probability of not meeting the standards (\bar{M}) is $P(\bar{M}) = 1 - P(M) = 1 - 0.92 = 0.08$.

Note that the probability of the concrete of being produced by plant 1 if the standards are met is, using Bayes' theorem (1.13),

$$P(P_1|M) = \frac{P(M|P_1)P(P_1)}{P(M)} = \frac{P(M|P_1)P(P_1)}{P(M|P_1)P(P_1) + P(M|P_2)P(P_2)} = \frac{0.54}{0.92} \approx 0.587.$$

Bridge specifications require, from a 100 test sample, the concrete be rejected if at least one cylinder among 5 selected at random do not meet the standards. What is the probability that the concrete produced by both plants will be considered acceptable?.

Let M_i be the standard fulfillment of test i -th, the probability of acceptance is the probability of the joint outcome $P(M_1 \cap M_2 \cap M_3 \cap M_4 \cap M_5)$. Applying (1.7) we have that this probability is,

$$P(M_1)P(M_2|M_1)P(M_3|M_1 \cap M_2)P(M_4|M_1 \cap M_2 \cap M_3)P(M_5|M_1 \cap M_2 \cap M_3 \cap M_4),$$

where the probability of meeting the standards for the first sample is $P(M_1) = 92/100$. After the event M_1 , there remain 99 samples of which 91 are acceptable, hence $P(M_2|M_1) = 91/99$. Continuing the reasoning, $P(M_3|M_1 \cap M_2) = 90/98$, $P(M_4|M_1 \cap M_2 \cap M_3) = 89/97$ and $P(M_5|M_1 \cap M_2 \cap M_3 \cap M_4) = 88/96$. Therefore,

$$P(\text{Acceptance}) = \frac{92}{100} \frac{91}{99} \frac{90}{98} \frac{89}{97} \frac{88}{96} \approx 0.6532. \quad \blacksquare$$

Illustrative Example 1.5 (Truss). Consider the statistically determinate structural system consisting of 5 members as it is shown in Figure 1.2. Let the event that element i fails be denoted by E_i ; obviously, its complement \bar{E}_i will be the survival of the i th member. Let the probability of failure of member i be equal to $P(E_i) = 0.01$ and considering that the individual members are statistically independent. What is the probability of failure of the system?.

As the structure is statically determinate, the failure of any member will result in the failure of the truss, that is $E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$. The complement of truss failure (survival of the truss) $\bar{E}_1 \cup \bar{E}_2 \cup \bar{E}_3 \cup \bar{E}_4 \cup \bar{E}_5$ means all five members must survive. Using Morgan's rule (1.14), the probability of survival is equal to

$$\begin{aligned} P(\text{Survival}) &= P(\overline{E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5}) = P(\bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_3 \cap \bar{E}_4 \cap \bar{E}_5) = \\ &= P(\bar{E}_1)P(\bar{E}_2)P(\bar{E}_3)P(\bar{E}_4)P(\bar{E}_5) = \\ &= (1 - P(E_1))(1 - P(E_2))(1 - P(E_3))(1 - P(E_4))(1 - P(E_5)) = \\ &= \prod_{i=1}^5 (1 - 0.01) \approx 0.951, \end{aligned}$$

and the probability of failure

$$P(\text{Failure}) = 1 - P(\text{Survival}) = 1 - 0.951 = 0.049. \quad \blacksquare$$

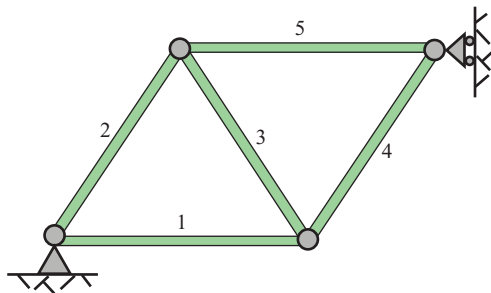


Figure 1.2: Statistically determinate truss.

1.4 Probabilistic Models

A random variable is said to be *discrete* if it can assume only a finite or countably infinite number of distinct values. Otherwise, it is said to be *continuous*. Thus, a continuous random variable can take an uncountable set of real values. The random variable X in Example 1.2 is discrete, whereas the random variable Y is continuous.

When we deal with a single random quantity, we have a *univariate* random variable. When we deal with two or more random quantities simultaneously, we have a *multivariate* random variable. Section 1.7 presents some probability functions of random variables, that are common to all discrete random variables with a discussion of some methods for defining the probability of univariate continuous random variables that include probability density, survival, hazard, and characteristic functions. Section 1.7 introduces examples of the commonly used discrete univariate random variables. Then commonly used continuous univariate random variables are presented in Section 1.7. Section ?? introduces discrete multivariate random variables. They are viewed with an eye on their application to reliability. Some multivariate continuous random variables are dealt with in Section ?. Some commonly used multivariate models are given in Section ?. Finally, Section ?? is devoted to truncated distributions, which have important applications.

1.5 Univariate Random Variables

To specify a random variable we need to know (a) its range or support, $S(X)$, which is the set of all possible values of the random variable, and (b) a tool by which we can obtain the probability associated with every subset in its support, $S(X)$. These tools are some functions such as the *probability mass function* (pmf), the *probability density function* (pdf), the *cumulative distribution function* (cdf), the *survival function*, the *hazard function* or the *characteristic function*. All these function and the so-called *moments* of random variables, are described in this section.

1.5.1 Probability Mass Function

Every discrete random variable has a *probability mass function* (pmf). The pmf of a discrete random variable X is a function that assigns to each real value x the probability of X having

Table 1.1: The Probability Mass Function (pmf) of a Random Variable X .

x	$P(x)$
-1	0.2
0	0.5
1	0.3
Total	1.0

the value x . That is $P_X(x) = \Pr(X = x)$. For notational simplicity we sometimes use $P(x)$ instead of $P_X(x)$. Every pmf $P(x)$ must satisfy the following conditions (1.2):

$$P(x) > 0, \quad \text{for all } x \in S(X), \quad \text{and} \quad \sum_{x \in S(X)} P(x) = 1. \quad (1.16)$$

Illustrative Example 1.6 (Weather condition). Suppose in Example 1.2 that 20% of the days have bad weather conditions, 50% have normal conditions, and the remaining 30% have good conditions, then at a random day, the probability that the weather condition in that day is bad is $P(-1) = 0.2$, the probability that it is normal is $P(0) = 0.5$, and the probability that it is good condition is $P(1) = 0.3$. ■

The pmf of a random variable X can be displayed in a table known as a *probability distribution table*. For example, Table 1.1 is the probability distribution table for the random variable X in Example 1.6. The first column in a probability distribution table is a list of the values of $x \in S(X)$, that is, only the values of x for which $P(x) > 0$. The second column displays $P(x)$. It is understood that $P(x) = 0$ for every $x \notin S(X)$.

1.5.2 Probability Density Function

As mentioned before, *continuous* random variables take an uncountable set of real values. Every continuous random variable has a *probability density function* (pdf). The pdf of a continuous random variable X is denoted by $f_X(x)$. For notational simplicity we sometimes use $f(x)$ instead of $f_X(x)$. Note that $f(x)$ is not $\Pr(X = x)$, as in the discrete case. But it is the height of the density curve at the point x (see 1.4 (a)).

Every pdf $f(x)$ must satisfy two conditions:

$$f(x) \geq 0, \forall x \quad (1.17)$$

and

$$\int_{x \in S(X)} f(x) dx = 1. \quad (1.18)$$

where, $S(X)$ is the support of the random variable X , the set of all values x for which $f(x) > 0$.

Table 1.2: The Probability Mass Function (pmf) and the Cumulative Distribution Function (cdf) of a Random Variable X .

x	$P(x)$	$F(x)$
-1	0.2	0.2
0	0.5	0.7
1	0.3	1.0

1.5.3 Cumulative Distribution Function

Every random variable also has a *cumulative distribution function* (cdf). The cdf of a random variable X , denoted by $F(x)$, is a function that assigns to each real value x the probability of X having values less than or equal to x , that is

$$F(x) = \Pr(X \leq x) = \sum_{a \leq x} P(a), \quad (1.19)$$

for the discrete case, and

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^x f(x)dx, \quad (1.20)$$

for the continuous case, which implies that

$$f(x) = \frac{dF(x)}{dx}. \quad (1.21)$$

Accordingly, the cdf can be obtained from the pmf and vice versa. For example, the cdf in the last column of Table 1.2 is computed from the pmf in Table 1.1 by accumulating $P(x)$ in the second column.

The pmf and cdf of any discrete random variable X can be displayed in probability distribution tables, such as Table 1.2, or they can be displayed graphically. For example, the graphs of the pmf and cdf in Table 1.2 are shown in Figure 1.3. In the graph of pmf, the height of a line on top of x is $P(x)$. The graph of the cdf for discrete random variable is a step function. The height of the step function is $F(x)$.

The cdf has the following properties as a direct consequence of the definitions of cdf and probability (see Figures 1.3 and 1.4 (b)):

1. $F(\infty) = 1$ and $F(-\infty) = 0$.
2. $F(x)$ is non-decreasing and right continuous.
3. For the discrete case, $P(x)$ is the jump of the cdf at x .
4. $\Pr(a < X \leq b) = F(b) - F(a)$.

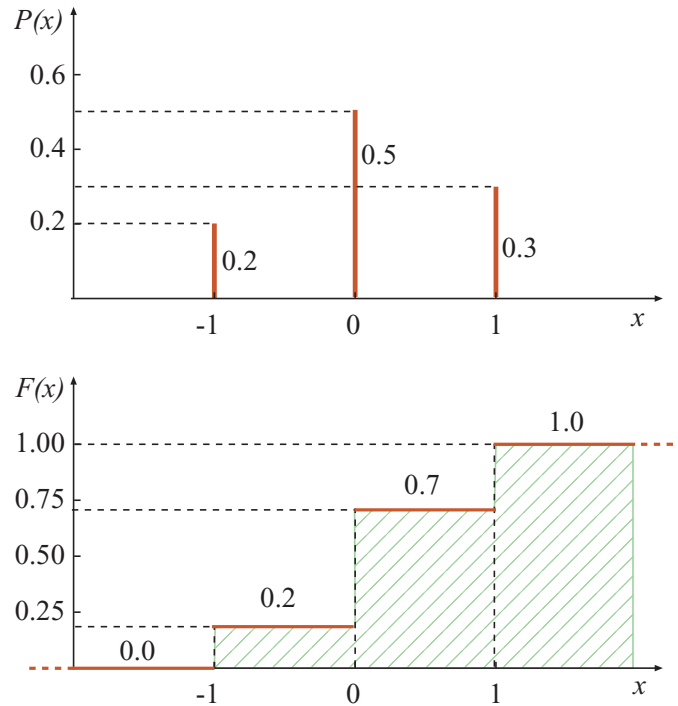


Figure 1.3: Graphs of the pmf and cdf of the random variable in Table 1.2.

Note that the fourth statement above is the probability that the random variable X takes values in the interval $(a, b]$, with $a \leq b$, that for the continuous case is given by

$$\Pr(a < X \leq b) = \int_a^b f(x)dx = F(b) - F(a). \quad (1.22)$$

Thus, $\Pr(a < X \leq b)$ is the area under the pdf on top of the interval $(a, b]$, as can be seen in Figure 1.4 (a), which shows the graphs of the pdf and cdf of a continuous random variable X . Note that, while $f(x)$ is the height of the density curve at x , $F(x)$ is the area under the curve to the left of x . From (1.18), the area under the pdf of any continuous random variable is 1.

Note also that

$$\Pr(X = x) = \Pr(x < X \leq x) = F(x) - F(x) = 0, \quad (1.23)$$

that is, while it is possible for a continuous random variable X to take a given value in its support, it is improbable that it will take this exact value. This is due to the fact that there are infinitely many possible values.

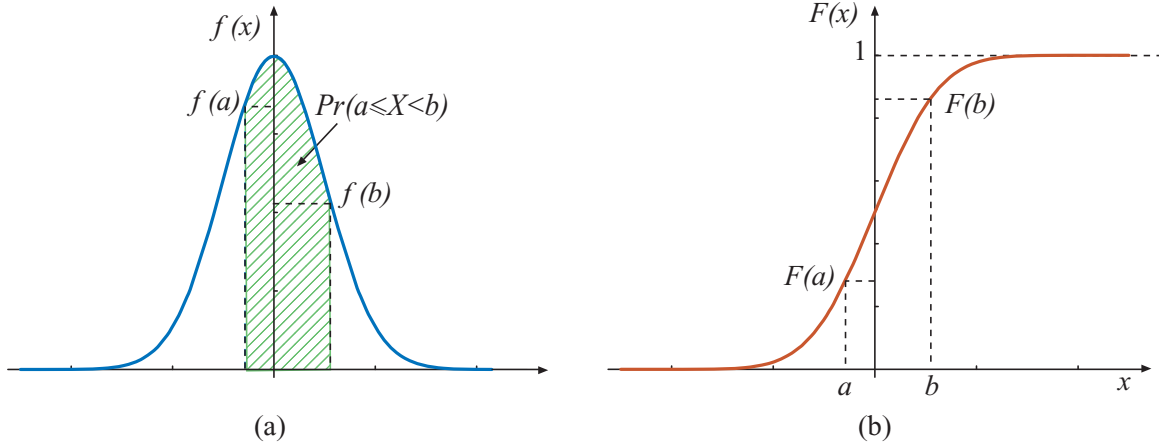


Figure 1.4: Graphs of the: (a) pdf and (b) cdf of a continuous random variable X .

1.5.4 Moments

Let $g(X)$ be a function of a discrete random variable X . The *expected value* of $g(X)$ is defined by

$$E[g(X)] = \sum_{x \in S(X)} g(x)P(x), \quad (1.24)$$

$$\text{or} = \int_{x \in S(X)} g(x)f(x)dx, \quad (1.25)$$

where (1.24) is the expression for the discrete random variable and the integral expression (1.25) is for the continuous case.

For example, letting $g(X) = X^r$, we obtain the so-called *r*th *moment* of the random variable X , with respect to the origin

$$E(X^r) = \sum_{x \in S(X)} x^r P(x), \quad (1.26)$$

$$\text{or} = \int_{x \in S(X)} x^r f(x)dx. \quad (1.27)$$

When $r = 1$, we obtain the *mean*, μ , of the discrete random variable X , that is a ‘weighted average’ of all the values the variable may take,

$$\mu = E(X) = \sum_{x \in S(X)} xP(x), \quad (1.28)$$

$$\text{or} = \int_{x \in S(X)} x f(x)dx. \quad (1.29)$$

Thus, the mean, μ , is the first moment of X with respect to the origin and it is analogous to the centroidal distance of the cross-section for a beam.

Letting $g(X) = (X - \mu)^r$, we obtain the r th *central moment*,

$$E[(X - \mu)^r] = \sum_{x \in S(X)} (x - \mu)^r P(x), \quad (1.30)$$

$$\text{or} = \int_{x \in S(X)} (x - \mu)^r f(x) dx. \quad (1.31)$$

When $r = 1$, it can be shown that the *first* central moment of any random variable X is zero, that is,

$$E(X - \mu) = 0. \quad (1.32)$$

When $r = 2$, we obtain the second central moment, better known as the *variance*, σ^2 , of the discrete random variable X , that is,

$$\sigma^2 = E[(X - \mu)^2] = \sum_{x \in S(X)} (x - \mu)^2 P(x), \quad (1.33)$$

$$\text{or} = \int_{x \in S(X)} (x - \mu)^2 f(x) dx, \quad (1.34)$$

and it is analogous to the moment of inertia of the cross-section for a beam.

The *standard deviation*, σ , of the random variable X is the positive square root of its variance and its equivalent concept in statics is the *radius of gyration*. The mean can be thought of as a measure of *center* and the standard deviation (or, equivalently, the variance) as a measure of *spread* or *variability*. It can be shown that the variance can also be expressed as

$$\sigma^2 = E(X^2) - \mu^2, \quad (1.35)$$

where

$$E(X^2) = \sum_{x \in S(X)} x^2 P(x),$$

$$\text{or} = \int_{x \in S(X)} x^2 f(x) dx,$$

is the second moment of X with respect to the origin. For example, the calculations of the mean and variance of the random variable X are shown in Table 1.3. Accordingly, the mean and variance of X are $\mu = 0.1$ and $\sigma^2 = 0.5 - 0.1^2 = 0.49$, respectively.

Other central tendency measures are the *mode* which is the most probable value, i.e., the value of x for which $P(x)$ or $f(x)$ is greatest, that for the continuous case can be estimated solving the following equation

$$\frac{df(x)}{dx} = 0, \quad (1.36)$$

and the *median* which is the value of x for which $F(x) = 0.5$. Accordingly, the mode and the median of the random variable X are 0 and 1, respectively.

Table 1.3: Calculations of the Mean and Variance of the Random Variable X .

x	$P(x)$	$xP(x)$	x^2	$x^2P(x)$
-1	0.2	-0.2	1	0.2
0	0.5	0.0	0	0.0
1	0.3	0.3	1	0.3
Total	1.0	0.1		0.5

The expected value, defined in (1.24) and (1.25), can be thought of as an *operator*, which has the following properties:

1. $E(c) = c$, for any constant c , that is, the expected values of a constant (a degenerate random variable) is the constant.
2. $E[cg(X)] = cE[g(X)]$.
3. $E[g(X) + h(X)] = E[g(X)] + E[h(X)]$, for any functions $g(X)$ and $h(X)$.

For example, $E(c + X) = E(c) + E(X) = c + \mu$. In words, the mean of a constant plus a random variable is the constant plus the mean of the random variable. As another example,

$$\sigma^2 = E[(X - \mu)^2] = E[(X^2 - 2\mu X + \mu^2)] = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2.$$

This is actually the proof of the identity in (1.35).

An extremely useful relative measure of the scatter of a random variable is the coefficient of variation defined as

$$v = \frac{\mu}{\sigma},$$

this value expresses a measure of the reliability of the central tendency and is very used for engineering applications. The higher the coefficient of variation, the greater will be the scatter. Table 1.4 gives some coefficients of variation for parameters common to civil engineering design.

For all random variables there are bounds on the relationship between the standard deviation and the mean. The most well-known is the *Chebyshev's inequality* (for a derivation see Lipschutz [?]),

$$P(|X - \mu| \geq h\sigma) \leq \frac{1}{h^2} \tag{1.37}$$

or

$$P(\mu - h\sigma \leq X \leq \mu + h\sigma) \geq 1 - \frac{1}{h^2} \tag{1.38}$$

where $h > 0$ is a real number. It asserts that for any probability distribution the probability that random values of the variate will lie within the range of h standard deviations on each side of the expected value is at least $1 - 1/h^2$. If the unknown probability distribution function is symmetrical with respect its expected value, and this is the maximum value (has a single peak, said to be *unimodal*), the expression (1.37) can be replaced by

$$P(\mu - h\sigma \leq X \leq \mu + h\sigma) \geq 1 - \frac{4}{9h^2}. \tag{1.39}$$

Table 1.4: Typical coefficients of variation for several engineering parameters.

Parameter	Coefficient of variation (v , %)	Parameter	Coefficient of variation (v , %)
Soil			
Porosity	10	Specific gravity	2
Water content		Degree of saturation	10
Silty clay	20	Unit weight	3
Clay	13	Compressibility factor	16
Pre. pressure	19	Compression index	
Friction angle (ϕ)		Sandy clay	26
Gravel	7	Clay	30
Sand	12	Cohesion (c)	40
SPT	26	Standard cone test	37
Structural loads, 50 year maximum			
Dead load	10	Live load	25
Snow load	26	Wind load	37
Steel structural resistance			
Yielding	11	Tensile strength	11
Uniform moment	13	Beam, column	15
Concrete structural resistance			
Flexure	11–14	Short columns	12–16

This is called the ‘Camp-Meidall’ inequality, sometimes called *Gauss’ inequality* (see Freeman [80]).

Illustrative Example 1.7 (Soil strength parameter). The expected value for the friction angle (ϕ) of a sand is $\mu_\phi = 25^\circ$. Estimate the minimum probability that a random sample of this sand will have a ϕ value between 19° and 31° .

Considering from Table 1.4 that $v_\phi = 0.12$, then the standard deviation is $\sigma_\phi = \mu_\phi v_\phi = 25 \cdot 0.12 = 3^\circ$ and $h = (\phi - \mu_\phi)/\sigma_\phi = 6/3 = 2$. Hence $P(19^\circ \leq \phi \leq 31^\circ) \geq 1 - 1/2^2 = 0.75$. Note that nothing else is known about the probability, if it is assumed that distribution of ϕ is symmetrical with its maximum at the mean value μ_ϕ , then using ‘Camp-Meidall’ inequality (1.39) we get that

$$P(19^\circ \leq \phi \leq 31^\circ) \geq 1 - \frac{4}{9 \cdot 2^2} \approx 0.89,$$

so recognizing symmetry, we can also claim that $P(\phi \leq 19^\circ) = P(\phi \geq 31^\circ) = (1 - 0.89/2) = 0.055$.

■

When $r = 3$, we obtain the third central moment, better known as the *skewness*, or lack

of symmetry, of the discrete random variable X , that is,

$$E[(X - \mu)^3] = \sum_{x \in S(X)} (x - \mu)^3 P(x), \quad (1.40)$$

$$\text{or} = \int_{x \in S(X)} (x - \mu)^3 f(x) dx, \quad (1.41)$$

$E[(X - \mu)^3]$ will be positive if there is greater dispersion of values $X > \mu$ than for values $X < \mu$; the sign and magnitude of $E[(X - \mu)^3]$ governs the sign and degree of ‘skewness’

$$\gamma_1 = \frac{E[(X - \mu)^3]}{\sigma^3}. \quad (1.42)$$

Positive skewness is indicated by the longer ‘tail’ of the distribution in the positive direction.

When $r = 4$, we obtain the fourth central moment, better known as the *kurtosis*, of the discrete random variable X , that is,

$$E[(X - \mu)^4] = \sum_{x \in S(X)} (x - \mu)^4 P(x), \quad (1.43)$$

$$\text{or} = \int_{x \in S(X)} (x - \mu)^4 f(x) dx, \quad (1.44)$$

that is a measure of the ‘flatness’ of a distribution. The greater the moment, the flatter (less peaked) is the distribution. The kurtosis is defined as

$$\gamma_2 = \frac{E[(X - \mu)^4]}{\sigma^4}, \quad (1.45)$$

with $\gamma_2 = 3$ for the Normal distribution. This measure is used only for large samples.

Higher order moments can be developed. In general the set of all moments of a probability mass function describes the function exactly. Thus, for some probability mass functions, a limited set of moments is sufficient to describe the function completely.

1.5.5 Survival and Hazard Functions

Let X be a non-negative random variable with pdf $f(x)$ and cdf $F(x)$. This happens, for example, when the random variable X is the lifetime of an object (e.g., a person or a machine). We have the following definitions:

Definition 1.2 (Survival function). *The function*

$$S(x) = \Pr(X > x) = 1 - F(x) \quad (1.46)$$

is called the survival function ■

The function $F(x)$ is called a survival function because it gives the probability that the object will survive beyond time x .

Definition 1.3 (Hazard function). *The hazard function (hazard rate or mortality function) is defined as:*

$$H(x) = \frac{f(x)}{\Pr(X > x)} = \frac{f(x)}{1 - F(x)}. \quad (1.47)$$

■

Assume that X is the lifetime of an element. Then, the hazard function can be interpreted as the probability, per unit time, that the item will fail just after time x given that the item has survived up to time x :

$$\begin{aligned} H(x) &= \lim_{\varepsilon \rightarrow 0} \frac{\Pr(x < X \leq x + \varepsilon | X > x)}{\varepsilon} \\ &= \frac{\lim_{\varepsilon \rightarrow 0} \frac{F(x + \varepsilon) - F(x)}{\varepsilon}}{1 - F(x)} \\ &= \frac{\frac{d}{dx} F(x)}{1 - F(x)} = \frac{f(x)}{1 - F(x)}. \end{aligned}$$

In other words, the hazard function can be interpreted as the probability of instantaneous failure given that the item has survived up to time x :

From (1.46) and (1.47), it can be seen that

$$f(x) = H(x)S(x), \quad (1.48)$$

that is, the pdf is the product of the hazard and survival functions. There is also a one-to-one correspondence between the cdf, $F(x)$, and the hazard function, $H(x)$. To see this, note that

$$\frac{d}{dx} \log(1 - F(x)) = -H(x),$$

and integrating from 0 to x we have,

$$\log(1 - F(x)) = - \int_0^x H(t) dt,$$

and

$$F(x) = 1 - \exp \left\{ - \int_0^x H(t) dt \right\}. \quad (1.49)$$

Consequently, there is a one-to-one correspondence between $F(x)$ and $H(x)$, hence one can be obtained from the other.

Note also that a comparison between (1.46) and (1.49) suggests the following relationship between the survival and hazard functions:

$$S(x) = \exp \left\{ - \int_0^x H(t) dt \right\}. \quad (1.50)$$

Illustrative Example 1.8 (Weibull). Consider a Weibull random variable with cdf

$$F(x) = 1 - \exp(-x^\beta); \quad \beta > 0, x > 0.$$

The corresponding pdf is given by

$$f(x) = \beta x^{\beta-1} \exp(-x^\beta).$$

Then, the hazard function is

$$H(x) = \frac{f(x)}{1 - F(x)} = \frac{\beta x^{\beta-1} \exp(-x^\beta)}{\exp(-x^\beta)} = \beta x^{\beta-1}.$$

Note that $H(x)$ is increasing, constant or decreasing according to $\beta > 1$, $\beta = 1$ or $\beta < 1$, respectively. ■

1.5.6 Characteristic Function

The approximative description of a random variable can be improved with the knowledge of its higher-order moments. Indeed, if all the moments of a random variable are known, its probability distribution would also be completely described. This means that an alternative way of describing the probability law of a random variable is using a function through which all the moments can be generated; such a function is the *moment-generating function*. Its complex form is a *characteristic function*, which is defined below.

Definition 1.4 (Characteristic function). *Let X be a unidimensional random variable with distribution function $F(x)$, its characteristic function, denoted $\psi_X(t)$ is defined, for the discrete case, as*

$$\psi(t) = \sum_{x \in S(X)} e^{itx} P(x), \quad (1.51)$$

where $P(x)$ is the pmf of X and $S(X)$ is the support of $P(x)$. For the continuous case, it is defined as

$$\psi(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx, \quad (1.52)$$

where $f(x)$ is the pdf of X . Note that, in both cases, $\psi(t)$ is a complex number. ■

Before we discuss the importance of the characteristic function, let us derive it for some random variables.

Illustrative Example 1.9 (Characteristic function of a discrete uniform random variable). The characteristic function of the discrete uniform random variable, which has the pmf

$$P(x) = \frac{1}{n}; \quad x = 1, 2, \dots, n,$$

is

$$\psi(t) = \sum_x e^{itx} P(x) = \sum_x e^{itx} \frac{1}{n} = \frac{1}{n} \sum_{i=1}^n e^{itx} = \frac{1}{n} \frac{e^{itn} - 1}{(e^{it} - 1)}.$$

■

Illustrative Example 1.10 (Characteristic function of a continuous uniform random variable). The characteristic function of the continuous uniform random variable, $U(0, a)$, with pdf function

$$f(x) = \frac{1}{a}; \quad 0 \leq x \leq a$$

is

$$\psi(t) = \int_0^a e^{itx} \frac{1}{a} dx = \frac{1}{a} \left. \frac{e^{itx}}{it} \right|_0^a = \frac{e^{ita} - 1}{ita}.$$

■

Illustrative Example 1.11 (Characteristic function of a binomial random variable). The characteristic function of the binomial random variable $B(n, p)$, with pmf function

$$P(x) = \binom{n}{x} p^x q^{n-x}; \quad x \in \{0, 1, \dots, n\}$$

is

$$\begin{aligned} \psi(t) &= \sum_{x=0}^n e^{itx} \binom{n}{x} p^x q^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (p e^{it})^x q^{n-x} = (p e^{it} + q)^n. \end{aligned} \quad (1.53)$$

■

Table 1.5 gives the characteristic functions of the most common distributions. Its main advantages are that it allows a simple calculation of moments and distributions of sums of independent random variables.

The most important properties (applications) of the characteristic function are:

1. The characteristic function always exists.
2. $\psi(0) = 1$.
3. $-1 \leq \psi(t) \leq 1$.
4. If $Z = aX + b$, where X is a random variable and a and b are real constants, we have

$$\psi_Z(t) = e^{itb} \psi_X(at), \quad (1.54)$$

where $\psi_Z(t)$ and $\psi_X(t)$ are the characteristic functions of Z and X , respectively.

5. The characteristic function of the sum of two random variables is the product of their characteristic functions:

$$\psi_{(X+Y)}(t) = \psi_X(t) \psi_Y(t). \quad (1.55)$$

Table 1.5: Characteristic Functions of Some Common Random Variables.

Discrete Distribution		Continuous Distribution	
Name	$\psi(t)$	Name	$\psi(t)$
Dirac	e^{its}	Uniform	$\frac{e^{itb} - e^{ita}}{it(b-a)}$
Bernouilli	$pe^{it} + q$	Exponential	$\left(1 - \frac{it}{\lambda}\right)^{-1}$
Binomial	$(pe^{it} + q)^n$	Gamma	$\left(1 - \frac{it}{\lambda}\right)^{-k}$
Geometric	$\frac{pe^{it}}{1 - (1-p)e^{it}}$	Normal	$\exp\left\{it\mu - \frac{\sigma^2 t^2}{2}\right\}$
Negative Binomial	$\left[\frac{pe^{it}}{1 - (1-p)e^{it}}\right]^r$	χ^2	$\frac{\exp\{it\lambda/(1-2it)\}}{(1-2it)^{n/2}}$
Poisson	$e^{\lambda(e^{it}-1)}$	Log-normal	**
Multinomial	$\left[\sum_{j=1}^k p_j e^{it_j}\right]^n$	Logistic	$e^{itx}\Gamma(1+i\beta t)\Gamma(1-i\beta t)$

6. Suppose that X_1, \dots, X_n is a set of n random variables with characteristic functions $\psi_{X_1}(t), \dots, \psi_{X_n}(t)$, respectively. Let $C = \sum_{i=1}^n a_i X_i$ be a linear convex combination of the random variables. Then, the characteristic function of C is the same linear convex combination of their characteristic functions, namely,

$$\psi_C(t) = \sum_{i=1}^n a_i \psi_{X_i}(t). \quad (1.56)$$

7. The characteristic function of the random variable R , which is the sum of a random number N of identically and independently distributed random variables X_1, X_2, \dots, X_N is given by

$$\psi_R(t) = \psi_N \left(\frac{\log \psi_X(t)}{i} \right), \quad (1.57)$$

where $\psi_X(t)$, $\psi_R(t)$ and $\psi_N(t)$ are the characteristic functions of X_i , R and N , respectively.

Illustrative Example 1.12 (Sum of normal random variables). Let Z_1, Z_2, \dots, Z_n be independent normal random variables $N(0, 1)$ with characteristic function

$$\psi_Z(t) = e^{-t^2/2}.$$

According to Property 5 the characteristic function of $S = Z_1 + Z_2 + \dots + Z_n$ is

$$\psi_S(t) = [\psi_Z(t)]^n,$$

and according to Property 4, since $M = S/\sqrt{n}$ we have

$$\begin{aligned} \psi_M(t) &= \psi_S(t/\sqrt{n}) = \psi_S(t\sqrt{n}) = [\psi_Z(t\sqrt{n})]^n \\ &= \exp \left(-\frac{1}{2} \left(\frac{t}{\sqrt{n}} \right)^2 n \right) = e^{-t^2/2}, \end{aligned} \quad (1.58)$$

which shows that $M = (Z_1 + Z_2 + \dots + Z_n)/\sqrt{n}$ has the same characteristic function as Z_i . ■

Illustrative Example 1.13 (Stability of the Gamma family with respect to scale changes). Let X be a Gamma $G(\theta, \lambda)$ random variable with characteristic function

$$\psi_X(t) = \left(1 - \frac{it}{\theta} \right)^{-\lambda}$$

and $Y = cX$. Then, by Property 4 we have

$$\begin{aligned} \psi_Y(t) &= \psi_X(ct) = \psi_{G(\theta, \lambda)}(ct) = \left(1 - \frac{ict}{\theta} \right)^{-\lambda} \\ &= \left(1 - \frac{it}{\theta/c} \right)^{-\lambda} = \psi_{G(\theta/c, \lambda)}(t), \end{aligned}$$

which shows that the random variable Y is $G(\theta/c, \lambda)$, that is, the Gamma family is stable with respect to scale changes. ■

Illustrative Example 1.14 (Stability of the uniform family). Let X be uniform $U(a, b)$, with characteristic function

$$\psi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

Then, by Property 4 the characteristic function of the random variable $Y = cX + d$ is

$$\begin{aligned} \psi_Y(t) &= e^{itd} \psi_X(ct) = e^{itd} \frac{e^{itbc} - e^{itac}}{ict(b-a)} \\ &= \frac{e^{it(cb+d)} - e^{it(ca+d)}}{it(cb-ca)} = \frac{e^{it(cb+d)} - e^{it(ca+d)}}{it[cb+d-(ca+d)]}, \end{aligned}$$

which shows that $Y \sim U(ca+d, cb+d)$. ■

As indicated, the k -th order moments μ_k with respect to the origin can be easily calculated from the characteristic function as

$$\mu_k = \frac{\psi^{(k)}(0)}{i^k}, \quad (1.59)$$

Illustrative Example 1.15 (Moments of the Bernoulli random variable). Since the characteristic function of the Bernoulli random variable is

$$\psi(t) = pe^{it} + q$$

using (1.59) we have

$$\mu_k = \frac{\psi^{(k)}(0)}{i^k} = \frac{pi^k}{i^k} = p.$$

that shows that all moments are equal to p . ■

Illustrative Example 1.16 (Moments of the Gamma random variable). Since the characteristic function of the Gamma $G(\theta, \lambda)$ random variable is

$$\psi(t) = \left(1 - \frac{it}{\theta}\right)^{-\lambda},$$

its moments with respect to the origin are:

$$\mu_k = \frac{\psi^{(k)}(0)}{i^k} = \frac{\lambda(\lambda-1)\dots(\lambda-k+1)}{\theta^k}.$$
■

1.6 Common Discrete Univariate Models

In this section we present several important discrete random variables that often arise in reliability applications. For a more detailed description and additional random variables, see for example, the books by Christensen (1984), [?], [?], Ross (1992), and Wackerly, Mendenhall, and Scheaffer (2002).

1.6.1 Bernoulli Random Variables

Problems of concern to engineers sometimes require the consideration of the potential occurrence or recurrence of an event in a sequence of repeated “trials” where only two possible mutually exclusive outcomes are possible, the *occurrence (success)* or *nonoccurrence (failure)* of the event.

Problems of this type may be modeled by a *Bernoulli* random variable. The probability of success in each trial is constant and equal to p and the probability of failure is $1 - p$. Note that the trials are statistically independent. This random experiment is called a *Bernoulli trial* or *experiment*. Define a random variable X by

$$X = \begin{cases} 0, & \text{if a failure is observed,} \\ 1, & \text{if a success is observed.} \end{cases}$$

This is called a *Bernoulli random variable*. The pmf of X is

$$P(x) = p^x(1 - p)^{1-x}; \quad x \in \{0, 1\}, \quad (1.60)$$

and its cdf is

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - p, & \text{if } 0 \leq x < 1, \\ 1, & \text{if } x \geq 1. \end{cases} \quad (1.61)$$

The mean and variance of a Bernoulli random variable are

$$\mu = p \quad \text{and} \quad \sigma^2 = p(1 - p). \quad (1.62)$$

Note that if $p = 1$, then X becomes a degenerate random variable (i.e., a constant) and the pmf of X is

$$P(x) = \begin{cases} 1, & \text{if } x = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (1.63)$$

This is known as the *Dirac* function.

Illustrative Example 1.17 (Concrete structure). Suppose we are interested in knowing whether or not a given concrete structure is in poor condition. Then, a random variable X can be defined as

$$X = \begin{cases} 1, & \text{if the condition is poor,} \\ 0, & \text{otherwise.} \end{cases}$$

This is a Bernoulli random variable. If 20% of structures are in poor condition. Then the pmf is

$$P(x) = 0.2^x(1 - 0.2)^{1-x}; \quad x \in \{0, 1\}.$$

The mean and variance of X are $\mu = p = 0.2$ and $\sigma^2 = p(1 - p) = 0.16$. ■

Bernoulli random variables arise frequently in the applications of extremes. In many occasions engineers are interested in events that cause failure such as exceedances of a random variable over a threshold value.

Definition 1.5 (Exceedances). Let X be a random variable and u a given threshold value. The event $\{X = x\}$ is said to be an exceedance at the level u if $x > u$. ■

For example, waves can destroy a breakwater when their heights exceed a given value, say 9 meters. Then it does not matter whether the height of a wave is 9.5, 10 or 12 meters because the consequences of these events are similar.

Let X be a random variable representing heights of waves and Y_u be defined as

$$Y_u = \begin{cases} 0, & \text{if no exceedance occurred,} \\ 1, & \text{if an exceedance occurred.} \end{cases}$$

Then Y_u is a Bernoulli random variable with success probability $p_u = \Pr(X > u)$.

Bernoulli random variables arise in many important practical engineering situations.

Illustrative Example 1.18 (Yearly maximum wave height). When designing a breakwater, civil engineers need to define the so called design wave height, which is the maximum wave height such that, when occurring, the breakwater will be able to withstand it without failure. Then, a natural design wave height would be the maximum wave height reaching the breakwater during its lifetime. However, this value is random and cannot be found. So, the only thing that an engineer can do is to choose this value with a small probability of being exceeded. In order to obtain this probability it is important to know the probability of exceedances of certain values during a year. Then, if we are concerned with whether the yearly maximum wave height exceeds a given threshold value h_0 , we have a Bernoulli experiment. ■

Illustrative Example 1.19 (Tensile strength). Suspension bridges are supported by long cables. However, long cables are much weaker than short cables, the only ones tested in the laboratory. This is so, because of the weakest link principle, which states that the strength of a long piece is the minimum strength of all its constituent pieces. Thus, the engineer has to extrapolate from lab results to real cables. The design of a suspension bridge requires the knowledge of the probability of the strength of the cable to fall below certain values. That is why values below a threshold are important. ■

Illustrative Example 1.20 (Nuclear power plant). When designing a nuclear power plant one has to consider the occurrence of earthquakes that can lead to disastrous consequences. Apart from the earthquake intensity, one of the main parameters to be considered is the distance from the earthquake epicenter to the location of the plant. Damage will be more severe for short distances than for long ones. Thus, engineers need to know whether this distance is below a given threshold value. ■

Illustrative Example 1.21 (Temperature). Temperatures have a great influence on engineering works and can cause problems either for large or small values. Then, values above or below given threshold values are important. ■

Illustrative Example 1.22 (Water flows). The water circulating through rivers also influence the life of humans. If the amount of water exceed a given level, large areas can be flooded. On the other hand, if the water levels are below given values the environment can be seriously damaged. ■

1.6.2 Binomial Random Variable

Suppose now that n Bernoulli experiments are run such that the following conditions hold:

1. The experiments are identical, i.e., the probability of success p is the same for all trials.
2. The experiments are independent, i.e., the outcome of an experiment has no influence on the outcomes of the others.

Let X be the number of successes in these n experiments. Then X is a random variable. To obtain the pmf of X , we first consider the event of obtaining x successes. If we obtained x successes, it also means that we obtained $n - x$ failures. Because the experiments are identical and independent, the probability of obtaining x successes and $n - x$ failures is:

$$p^x(1 - p)^{n-x}.$$

Note also that the number of possible ways of obtaining x successes (and $n - x$ failures) is obtained using the combinations formula:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Therefore, the pmf of X is:

$$P(x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x \in \{0, 1, \dots, n\}. \quad (1.64)$$

This random variable is known as the binomial random variable and is denoted by $X \sim B(n, p)$.

The mean and variance of a $B(n, p)$ random variable can be shown to be

$$\mu = np \quad \text{and} \quad \sigma^2 = np(1 - p). \quad (1.65)$$

Figure 1.5 shows the graphs of the pmf of three binomial random variables with $n = 50$ and three values of p . From these graphs it can be seen that when $p = 0.5$, the pmf is symmetric; otherwise it is skewed.

Since X is the number of successes in these n identical and independent Bernoulli experiments, one may think of X as the sum of n identical and independent Bernoulli random variables, that is $X = X_1 + X_2 + \dots + X_n$, where X_i is a Bernoulli random variable with probability of success equal to p . Note that when $n = 1$, then a $B(1, p)$ random variable is a Bernoulli random variable.

Another important property of binomial random variables is *reproductivity* with respect to the parameter n . This means that the sum of two independent binomial random variables with the same p , is also a binomial random variable. More precisely, if $X_1 \sim B(n_1, p)$ and $X_2 \sim B(n_2, p)$, then

$$X_1 + X_2 \sim B(n_1 + n_2, p).$$

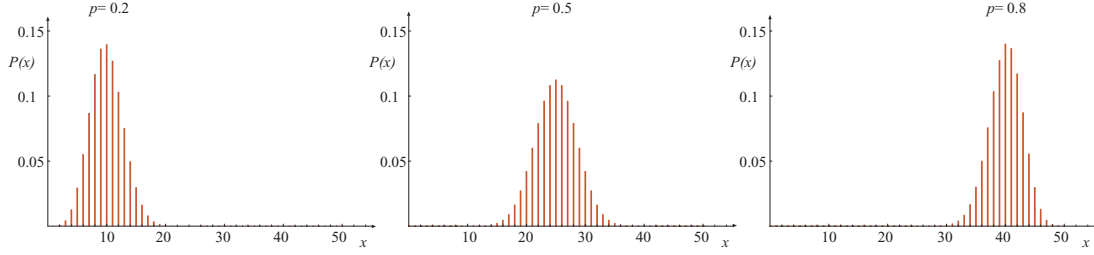


Figure 1.5: Examples of probability mass functions of binomial random variables with $n = 50$ and three values of p .

Illustrative Example 1.23 (Exceedances). An interesting practical problem consists of determining the probability of r exceedances over a value u in n identical and independent repetitions of the experiment. Since there are only two possible outcomes (exceedance or not exceedance), these are Bernoulli experiments. Consequently, the number of exceedances M_u over the value u of the associated random variable X is a $B(n, p_u)$ random variable with parameters n and p_u , where p_u is the probability of an exceedance over the level u of X . Therefore, the pmf of M_u is

$$\Pr(M_u = r) = \binom{n}{r} (p_u)^r (1 - p_u)^{n-r}, \quad r \in \{0, 1, \dots, n\}. \quad (1.66)$$

Moreover, since p_u can be written as

$$p_u = \Pr(X > u) = 1 - F(u), \quad (1.67)$$

where $F(\cdot)$ is the cdf of X , (1.66) becomes

$$\Pr(M_u = r) = \binom{n}{r} [1 - F(u)]^r [F(u)]^{n-r}. \quad (1.68)$$

■

Illustrative Example 1.24 (Concrete structures). Suppose that an engineer examined $n = 6$ concrete structures to determine which ones are in poor condition. As in Example 1.17, the probability that a given structure is in poor condition is $p = 0.2$. If X is the number of structures that are in poor condition, then X is a binomial random variable $B(6, 0.2)$. From (1.64) the pmf is

$$P(x) = \binom{6}{x} p^x (1 - p)^{6-x}, \quad x \in \{0, 1, \dots, 6\}. \quad (1.69)$$

For example, the probability that none of the six structures are found to be in poor condition is

$$P(0) = \binom{6}{0} 0.2^0 0.8^6 = 0.2621,$$

and the probability that only one of the six structures are found to be in poor condition is

$$P(1) = \binom{6}{1} 0.2^1 0.8^5 = 0.3932.$$

■

Illustrative Example 1.25 (Yearly maximum wave height). Consider a breakwater that is to be designed for a lifetime of 50 years. Assume also that the probability of yearly exceedance of a wave height of 9 meters is 0.05. Then, the probability of having 5 years with exceedances during its lifetime is given by

$$P(5) = \binom{50}{5} 0.05^5 (1 - 0.05)^{50-5} = 0.06584.$$

Note that we have admitted the two basic assumptions of the binomial model, that is, identical and independent Bernoulli experiments. In this case both assumptions are reasonable. Note however, that if the considered period were one day or one month instead of one year this would not be the case, because the wave heights in consecutive days are not independent events (Assume both days belong to the same storm, then the maximum wave heights would be both high. On the other hand, if the period were calm, both would be low).

It is well known that there is some periodical phenomena ruling the waves that can last for more than one year. For that reason it would be even better to consider periods of longer duration. ■

Illustrative Example 1.26 (Earthquake epicenter). From past experience, the epicenters of 10% of the earthquakes are within 50 km from a nuclear power plant. Now, consider a sequence of 10 such earthquakes and let X be the number of earthquakes whose epicenters are within 50 km from the nuclear power plant. Assume for the moment that the distances associated with different earthquakes are independent random variables and that all the earthquakes have the same probabilities of having their epicenters at distances below 50 km. Then, X is $B(10, 0.1)$ random variable. Accordingly, the probability that none of the 10 earthquakes will occur within 50 km is

$$P(0) = \binom{10}{0} 0.1^0 (1 - 0.1)^{10-0} = 0.348678.$$

Note that this probability is based on two assumptions:

1. The distances associated with any two earthquakes are independent random variables.
2. The occurrence of an earthquake does not change the possible locations for others.

Both assumptions are not very realistic, because once an earthquake has occurred some others usually occur in the same or nearby location until the accumulated energy is released. After then, the probability of occurrence at the same location becomes much smaller, because no energy has been built up yet. ■

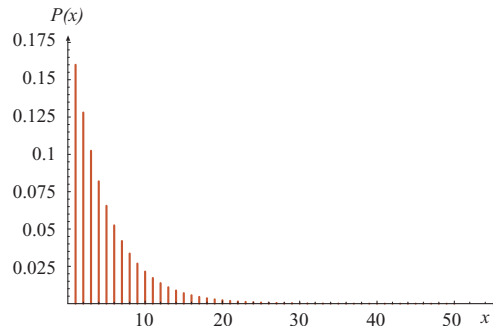


Figure 1.6: Graph of the probability mass function of the Geometric random variable with $p = 0.2$.

Illustrative Example 1.27 (Dam design). In planning the flood control system for a river, a dam has been design and construct with a projected life of 100 years. Supposing that the annual maximum flow of the river over a sequence of years is independent, what is the probability that a 500-year flood will occur once during the lifetime of the dam?.

Given only the information above the probability of a 500-year flood occurring in any one year is $p = 1/500 = 0.002$. If X is the number of 500-year floods during the lifetime (N) of the dam, then X is a binomial random variable $B(N, p) = B(100, 0.002)$, thus we have

$$P(1) = \binom{100}{1} 0.002^1 (1 - 0.002)^{100-1} = 0.164041.$$

The probability is about a 16.4%. ■

1.6.3 Geometric or Pascal Distribution

Consider again a series of identical and independent Bernoulli experiments, which are repeated until the first success is obtained. Let X be the number of trial on which the first success occurs. What is the pmf of the random variable X ? Note that if the first success has occurred at the trial number x , then the first $(x - 1)$ trials must have been failures. Since the probability of a success is p and the probability of the $(x - 1)$ failures is $(1 - p)^{(x-1)}$ (because the trials are identical and independent), the pmf of X is

$$P(x) = p(1 - p)^{x-1}, \quad x \in \{1, 2, \dots\}. \quad (1.70)$$

This random variable is called a *Geometric* or *Pascal* random variable and is denoted by $G(p)$. The pmf of $G(p)$ random variable is decreasing in x , which means that the largest value of $P(x)$ is at $x = 1$. A graph of the pmf of $G(0.2)$ is shown in Figure 1.6. The mean and variance of $G(p)$ are:

$$\mu = \frac{1}{p} \quad \text{and} \quad \sigma^2 = \frac{1 - p}{p^2}. \quad (1.71)$$

Illustrative Example 1.28 (Job interviews). A company has one vacant position to fill. It is known that 80% of job applicants for this position are actually qualified for the

job. The company interviews the applicants one at a time as they come in. The interview process stops as soon as one qualified applicant is found. How many interviews will have to be conducted until the first qualified applicant is found? This can be thought of having a series of identical and independent Bernoulli trials each with success probability $p = 0.8$. If X is the number of interviews required to find the first qualified applicant, then X is $G(0.8)$ random variable. For example, the probability that the first qualified applicant is found on the third interview is:

$$P(3) = 0.8(1 - 0.8)^{(3-1)} = 0.032.$$

Also, the probability that the company will conduct at least three interviews to find the first qualified applicant is

$$\Pr(X \geq 3) = 1 - \Pr(X \leq 2) = 1 - F(2) = 1 - [P(1) + P(2)] = 0.04.$$

If each interview costs the company \$500, then the expected costs of filling the vacant position is

$$500 E(X) = 500 (1/0.8) = \$625. \quad \blacksquare$$

Assume now that a given event (flood, dam failure, exceedance over a given temperature, etc.) is such that its probability of occurrence during a period of unit duration (normally one year) is a small value p . Assume also that the occurrences of such event in non-overlapping periods are independent. Then, as time passes, we have a sequence of identical Bernoulli experiments (occurrence or not of the given event). Thus, the time measured in the above units until the first occurrence of this event is the number of experiments until the first occurrence and then it can be considered as a geometric $G(p)$ random variable, whose mean is $1/p$. This means that on average the time between two consecutive occurrences of an event is equal to the reciprocal of the probability of the event during one time unit. This suggests the following definition.

Definition 1.6 (Return period). *Let A be an event, and T be the random time between successive occurrences of A . The mean value, μ , of the random variable T is called the return period of A (note that it is the mean time for the return of such an event).* \blacksquare

For the return period to be approximately $1/p$, the following conditions must hold:

1. The probability of one event occurring during a short period of time is small,
2. The probability of more than one event occurring during a short period of time is negligible.

1.6.4 Negative Binomial Distribution

The geometric distribution arises when we are interested in the number of Bernoulli trials that are required until we get the first success. Now suppose that we define the random variable X as the number of identical and independent Bernoulli trials that are required until we get the r th success. For the r th success to occur at the trial number x , we must

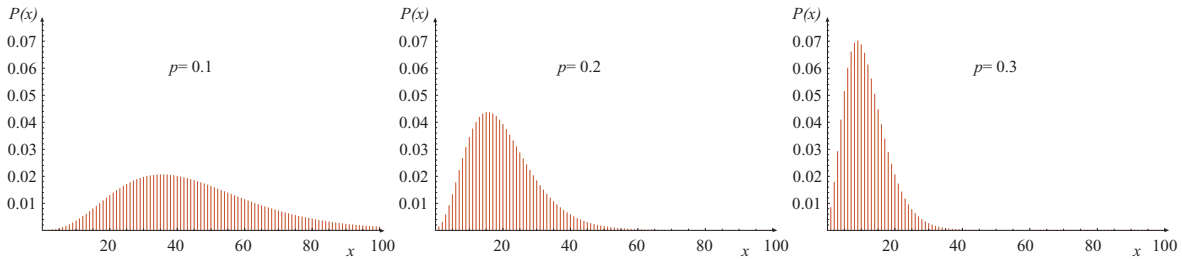


Figure 1.7: Examples of probability mass functions of negative binomial random variables with $r = 5$ and three values of p .

have $r - 1$ successes in the $x - 1$ previous trials and one success in the trial number x . The number of possible ways of obtaining $r - 1$ successes in $x - 1$ trials is obtained using the combinations formula:

$$\binom{x-1}{r-1} = \frac{(x-1)!}{(r-1)! (x-r)!}.$$

Therefore, the pmf of X is

$$P(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x \in \{r, r+1, \dots\}.$$

This random variable is called a *negative binomial* random variable and is denoted by $X \sim NB(r, p)$. It is also known as Pascal distribution. A graph of the pmf of $G(0.2)$ is shown in Figure 1.7. Note that the geometric distribution is a special case of the negative binomial distribution obtained by setting ($r = 1$), that is, $G(p) \sim NB(1, p)$. The mean and variance of an $NB(r, p)$ variable are

$$\mu = \frac{r(1-p)}{p} \quad \text{and} \quad \sigma^2 = \frac{r(1-p)}{p^2}. \quad (1.72)$$

Illustrative Example 1.29 (Radio transmission tower). A radio transmission tower has been designed for a 25 years wind. What is the probability that the third 25-year wind will occur in the ten year after construction?

If X is the number of years needed for three 25-year wind occurrences, then X is a negative binomial distribution $NB(3, 1/25)$. Thus the required probability is

$$P(10) = \binom{10-1}{3-1} 0.04^3 (1-0.04)^{10-3} = 0.00280825.$$

■

Illustrative Example 1.30 (Job interviews). Suppose that the company in Example 1.28 wishes to fill two vacant positions. Thus, the interview process stops as soon as two qualified applicants are found. If X is the number of interviews needed to fill the two vacant

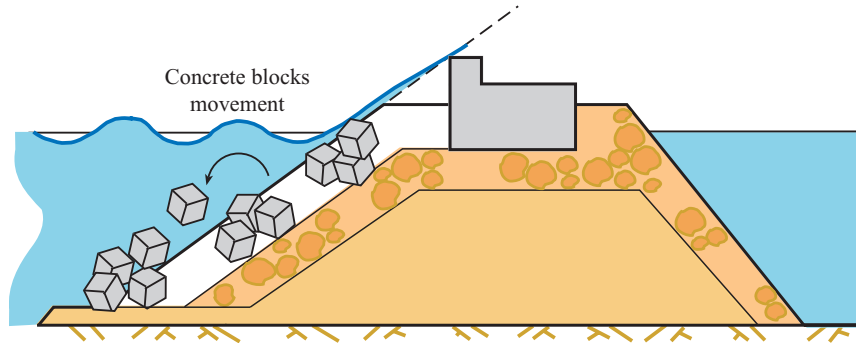


Figure 1.8: Illustration of rubble-mound breakwater.

positions, then X is an $NB(2, 0.8)$ random variable. For example, the probability that the second qualified applicant is found on the third interview is:

$$P(3) = \binom{3-1}{2-1} 0.8^2 0.2^{3-2} = 0.128.$$

■

Illustrative Example 1.31 (Job interviews). Suppose that the company in Example 1.28 wishes to fill two vacant positions. Thus, the interview process stops as soon as two qualified applicants are found. If X is the number of interviews needed to fill the two vacant positions, then X is an $NB(2, 0.8)$ random variable. For example, the probability that the second qualified applicant is found on the third interview is:

$$P(3) = \binom{3-1}{2-1} 0.8^2 0.2^{3-2} = 0.128.$$

■

Illustrative Example 1.32 (Rubble-mound breakwater). A rubble-mound breakwater is made of a supported crownwall on an earthfill that is protected by a mound and concrete block armor (large pieces of stone and cubic blocks to protect the earthfill from the waves) (see Figure 1.10). The geometrical connectivity and structural stress transmission in the armor occurs by friction and interlocking between units. While failure of rigid breakwaters occurs when a single wave exceeds a given threshold value, a rubble-mound breakwater fails after the occurrence of several waves above a given threshold value. This is because the failure is progressive, that is, the first wave produces some movement of the concrete pieces, the second increases the damage, etc.

Then, a failure occurs when the r -th Bernoulli event (wave height exceeding the threshold) occurs. Thus, the negative binomial random variable plays a key role. Assume that a wave produces some damage on the armor if its height exceeds 7 m., and that the probability of this event is 0.001. Then, if the rubble-mound breakwater fails after, say, 8 such waves,

then the number of waves occurring until failure is a negative binomial random variable. Consequently, the pmf of the number of waves until failure is

$$P(x) = \binom{x-1}{8-1} 0.001^8 (1-0.001)^{x-8}, \quad x \in \{8, 9, \dots\}.$$

■

Like the binomial random variable, the negative binomial random variable is also reproductive with respect to parameter r . This means that the sum of independent negative binomial random variables with the same probability of success p , is a negative binomial random variable. More precisely if $X_1 \sim NB(r_1, p)$ and $X_2 \sim NB(r_2, p)$, then

$$X_1 + X_2 \sim NB(r_1 + r_2, p).$$

1.6.5 Hypergeometric Distribution

Suppose we have a finite population consisting of N elements, where each element can be classified into one of two distinct groups. Say, for example, that we have N products of which D products are defective and the remaining N_D are acceptable (not defective). Suppose further that we wish to draw a random sample of size $n < N$ from this population without replacement. The random variable X , which is the number of defective items in the sample, is called a *hypergeometric* random variable and is denoted $HG(N, p, n)$, where $p = D/N$ is the proportion of defective items in the population.

It is clear that the number of defective elements, X , cannot exceed neither the total number of defective elements D , nor the sample size n . Also, it cannot be less than 0 or less than $n - (N - D)$. Therefore, the support of X is:

$$\max(0, n - qN) \leq X \leq \min(n, D),$$

where $q = 1 - p$ is the proportion of acceptable items in the population. The probability mass function of X is:

$$P(x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}}; \quad \max(0, n - qN) \leq x \leq \min(n, Np). \quad (1.73)$$

The numerator is the number of samples that can be obtained with x defective elements and $n - x$ acceptable elements. The denominator is the total number of possible samples of size n that can be drawn. The mean and variance of $HG(N, p, n)$ are:

$$\mu = np; \quad \sigma^2 = \frac{N-n}{N-1} np(1-p). \quad (1.74)$$

When N tends to infinity this distribution tends to the binomial distribution. Moreover, if n is small relative to N (about 5-10% of N in real problems), it can be shown that the hypergeometric distribution can be approximated by the binomial distribution. If the value

p is known precisely the use of the binomial distribution is appropriate but if p has to be estimated, then such a binomial distribution assumes more precision than actually exists, making the situation appear better than it really is. Since for using the binomial distribution a prior estimate of p is needed and this is rarely the case, the use of the binomial distribution is limited in many engineering applications using the hypergeometric distribution instead.

Note that the use of the hypergeometric distribution is the basic distribution underlying many sampling plans used in connection with acceptance sampling and quality control.

Illustrative Example 1.33 (Urn problem). A urn contains 20 balls, 5 white and 15 black. We draw a sample of size 10 without replacement. What is the probability that the drawn sample contains exactly 2 white balls? Here $N = 20$, $p = 5/20$, and $n = 10$. Letting X be the number of white balls, then X is $HG(20, 1/4, 10)$. From (1.73), we have,

$$P(2) = \frac{\binom{5}{2} \binom{15}{8}}{\binom{20}{10}} = 0.348.$$

■

Illustrative Example 1.34 (Quality control). To ensure quality at a large construction project where 200 concrete cylinders are to be taken daily, the acceptance criterion requires that from a 15 random sample testing of these cylinders, at least 14 must have a specified minimum crushing strength. The cylinders are provided by a plant where the 97% of the concrete meet the standards. What is the likelihood that the material from this plant will be rejected?.

The proportion of defective cylinders from plant is $p = 1 - 0.97 = 0.03$, thus if X is the number of cylinders from the sample not meeting the standards, then X is $HG(200, 0.03, 15)$. The material will be rejected if $X > 1$, thus

$$P(X > 1) = 1 - P(X \leq 1) = 1 - (P(X = 0) + P(X = 1))$$

where using (1.73), we have,

$$P(X > 1) = 1 - \left(\frac{\binom{15}{0} \binom{185}{15}}{\binom{200}{15}} + \frac{\binom{15}{1} \binom{185}{14}}{\binom{200}{15}} \right) = 0.06621.$$

■

Hence, the probability of rejecting the material is about 6.6%, that means that the acceptance criterion may be satisfactory for detecting poor-quality material.

Note that in this example n is 7.5% of N . Using the binomial approximation

$$P(X > 1) = 1 - \binom{15}{0} 0.03^0 (1 - 0.03)^{200-0} - \binom{15}{1} 0.03^1 (1 - 0.03)^{200-1} = 0.0729725,$$

where the error in the approximation is about 0.001%. A graph of the pmf of $HG(200, 0.03, 15)$ and the corresponding binomial approximation $B(15, 0.03)$ are shown in Figure 1.9.

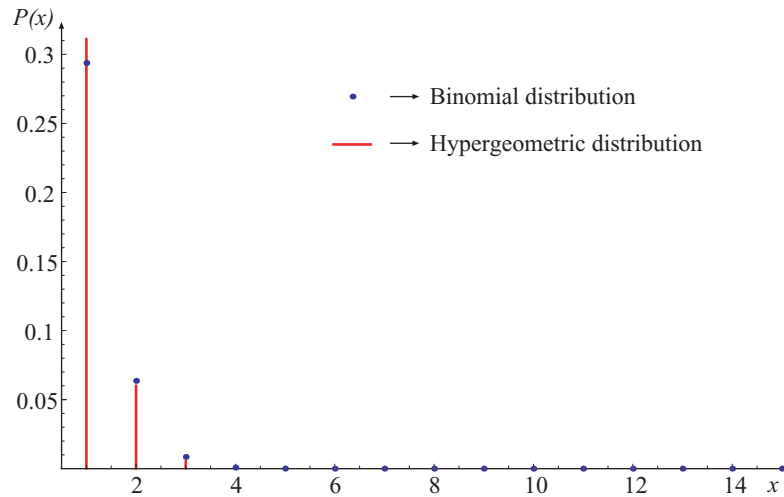


Figure 1.9: Example of the probability mass functions of the hypergeometric and the binomial approximation in the Illustrative Example 1.34.

1.6.6 Poisson Distribution

Another important distribution used frequently in engineering to evaluate the possible occurrence of events in any time and/or space is the *Poisson distribution*. Suppose we are interested in the number of occurrences of an event over a given interval of time or space. For example, let X be the number of tornados that strike a structure during a time interval t , or the number of fatigue cracks that may occur anywhere along a continuous weld. Then, X is a random variable and we are interested in finding its pmf. The experiment here consists of counting the number of times an event occurs during a given interval of duration t . Note that t does not have to be time; it could be location (fatigue cracks), area, volume, etc.

To derive the pmf of X we make the following *Poissonian* assumptions:

1. The probability p of the occurrence of a single event in a short interval d is proportional to its duration, that is, $p = \alpha d$, where α is a positive constant, known as the *arrival* or *intensity rate* that is the mean rate of occurrence of the event.
2. The probability of the occurrence of more than one event in the same interval is negligible.
3. The number of occurrences in one interval is independent of the number of occurrences in other non-overlapping intervals.
4. The number of events occurring in two intervals of the same duration have the same probability distribution.

Now, divide the interval t into n small and equal subintervals of duration $d = t/n$. Then, with the above assumptions, we may think of the n subintervals as n identical and

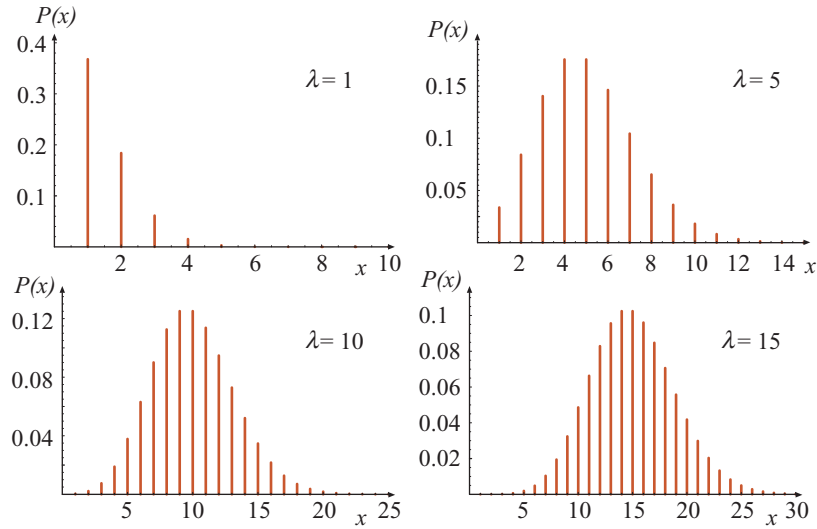


Figure 1.10: Some examples of probability mass functions of the Poisson random variable with four different values of λ .

independent Bernoulli trials X_1, X_2, \dots, X_n , with $\Pr(X_i = 1) = p$ for $i = 1, 2, \dots, n$. Letting $X = \sum_{i=1}^n X_i$, then X is a binomial random variable with parameters n and $p = \alpha t/n$. To guarantee that no more than a single event occurs in a subinterval d , the interval t may have to be divided into a very large number of intervals. So, we are interested in the pmf of X as $n \rightarrow \infty$. Under the above assumptions, one can show that

$$P(x) = \lim_{n \rightarrow \infty} \binom{n}{x} p^x (1-p)^{n-x} = \frac{e^{-\alpha t} (\alpha t)^x}{x!}; \quad x = 0, 1, 2, \dots, \lambda > 0. \quad (1.75)$$

Letting $\lambda = \alpha t$, we obtain

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}; \quad x = 0, 1, 2, \dots, \lambda > 0. \quad (1.76)$$

This random variable, which is the number of events occurring in period of a given duration t , is known as a Poisson random variable with parameter $\lambda = \alpha t$ and is denoted by $P_0(\lambda)$. Note that the parameter λ is equal to the intensity α times the duration t .

Figure 1.10 shows the graphs of the pmf of some Poisson random variables. It can be seen that as the parameter λ gets larger, the pmf becomes more symmetric.

The mean and variance of a $P_0(\lambda)$ variable are:

$$\mu = \lambda \quad \text{and} \quad \sigma^2 = \lambda. \quad (1.77)$$

Like the binomial random variable, the Poisson random variables are also reproductive, that is, if $X_1 \sim P_0(\lambda_1)$ and $X_2 \sim P_0(\lambda_2)$ are independent, then:

$$X_1 + X_2 \sim P_0(\lambda_1 + \lambda_2).$$

The Poisson random variable is particularly appropriate for modelling the number of occurrences of rare events such as storms, earthquakes, and floods.

Illustrative Example 1.35 (Storms). Suppose that storms of a certain level occur once every 50 years on the average. We wish to compute the probability of no such storm will occur during a single year. Assuming that X has a Poisson random variable with parameter λ , then $\lambda = 1/50 = 0.02$ and, using (1.76), we have

$$P(0) = \frac{e^{-1/50}(1/50)^0}{0!} = e^{-0.02} = 0.98.$$

That is, it is highly likely that no storms will occur in a single year.

For this model to be correct we need to check the assumptions of the Poisson model earlier. The first two assumptions are reasonable, because if several storms occur during a short interval, they could be considered as a single storm. The third and fourth assumptions are not true for close intervals, but they are true for far enough intervals. ■

Illustrative Example 1.36 (Parking garage). A parking garage has three car entrances. Assume that the number of cars coming into the garage using different entrances are independent Poisson random variables with parameters λ_1 , λ_2 and λ_3 . Then, using the reproductivity property of Poisson random variables, the total number of cars entering the garage is a Poisson random variable $P(\lambda_1 + \lambda_2 + \lambda_3)$.

The definition of reproductivity assumes that the random variables being considered in the sum are independent. Then, the number of cars entering at each entrance must be independent. This assumption must be checked before the above Poisson model is used. ■

Poisson Approximation of the Binomial Distribution

If X is a $B(n, p)$ random variable, but p is small, say $p \leq 0.01$, and $np \leq 5$, then the pmf of X can be approximated by the pmf of the Poisson random variable with $\lambda = np$, the mean of the binomial random variable, that is,

$$P(x) \approx \frac{e^{-\lambda} \lambda^x}{x!}; \quad x = 0, 1, \dots, n. \quad (1.78)$$

This is why the Poisson process is known as the *rare* events process.

Illustrative Example 1.37 (Storms). Consider the storms in Example 1.35, and suppose that we are interested in the number of years with storms over a 40-year period. Although X is $B(40, 1/50)$, it can be approximated by a $P(40/50)$ random variable. For example, $P(3)$ can be computed either exactly, using the binomial pmf

$$P(x) = \binom{40}{3} \left(\frac{1}{50}\right)^3 \left(1 - \frac{1}{50}\right)^{(40-3)} = 0.0374293,$$

or approximately, using the Poisson pmf,

$$P(x) \approx \frac{(40/50)^3 e^{-(40/50)}}{3!} = 0.0383427.$$

The error of approximation in this case is $0.0374293 - 0.0383427 = -0.0009134$. ■

Table 1.6: Some Discrete Random Variables that Arise in Engineering Applications, Together with Their Probability Mass Functions, Parameters and Supports.

Distribution	$P(x)$	Parameters and Support
Bernoulli	$P(x) = p^x(1-p)^{1-x}$	$0 < p < 1$ $x \in \{0, 1\}$
Binomial	$\binom{n}{x} p^x(1-p)^{n-x}$	$n \in \{1, 2, \dots\}$ $0 < p < 1$ $x \in \{0, 1, \dots, n\}$
Geometric	$p(1-p)^{x-1}$	$0 < p < 1$ $x \in \{1, 2, \dots\}$
Negative Binomial	$\binom{x-1}{r-1} p^r(1-p)^{x-r}$	$0 < p < 1$ $x \in \{r, r+1, \dots\}$
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$	$\lambda > 0$ $x \in \{0, 1, \dots\}$
Non-Zero Poisson	$\frac{e^{-\lambda}\lambda^x}{x!(1-e^{-\lambda})}$	$\lambda > 0$ $x \in \{1, 2, \dots\}$

1.6.7 Non-Zero Poisson Random Variable

In certain practical applications we are interested in the number of occurrences of an event over a period of duration t , but we also know that at least one event has to occur during the period. If the Poissonian assumptions hold, then it can be shown that the random variable X has the following pmf

$$P(x) = \frac{e^{-\lambda}\lambda^x}{x!(1-e^{-\lambda})}, \quad x \in \{1, 2, \dots\}, \lambda > 0. \quad (1.79)$$

This random variable is known as the *non-zero* or the *zero-truncated* Poisson distribution and is denoted by $P_0(\lambda)$. The mean and variance of $P_0(\lambda)$ are:

$$\mu = \frac{\lambda}{1-e^{-\lambda}}; \quad \sigma^2 = \frac{e^\lambda\lambda(-1+e^\lambda-\lambda)}{(1-e^\lambda)^2}.$$

A summary of the random variables we discussed in this section is given in Table 1.6.

1.7 Common Continuous Univariate Models

In this section we present several important continuous random variables that often arise in reliability engineering applications. For a more detailed presentations and additional random variables, see for example, the books by [?] and [?], [?], [?], and [?].

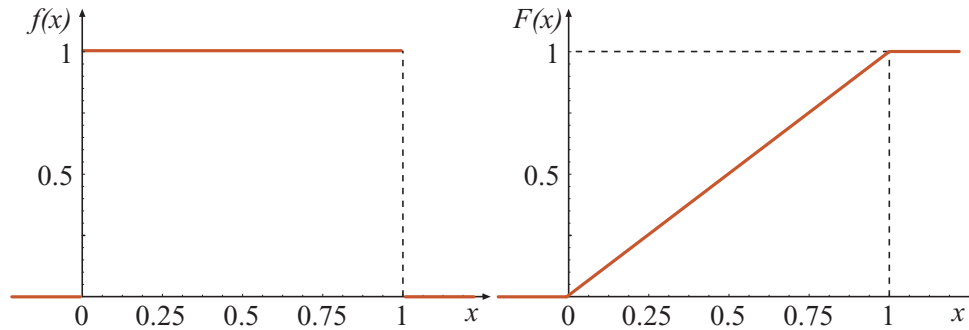


Figure 1.11: The pdf and cdf of the standard uniform random variable.

1.7.1 Uniform Distribution

The uniform random variable on the interval $[\alpha, \beta]$, denoted by $U(\alpha, \beta)$, has the following pdf

$$f(x) = \frac{1}{\beta - \alpha}; \quad \alpha \leq x \leq \beta, \quad (1.80)$$

from which it follows that the cdf can be written as

$$F(x) = \begin{cases} 0, & \text{if } x < \alpha \\ \frac{x - \alpha}{\beta - \alpha}, & \text{if } \alpha \leq x < \beta, \\ 1, & \text{if } x \geq \beta. \end{cases}$$

The mean and variance of X are

$$\mu = \frac{\alpha + \beta}{2} \quad \text{and} \quad \sigma^2 = \frac{(\beta - \alpha)^2}{12}. \quad (1.81)$$

A special case of $U(\alpha, \beta)$ is the *standard uniform* random variable, $U(0, 1)$ obtained by setting $\alpha = 0$ and $\beta = 1$. The pdf and cdf of $U(0, 1)$ are

$$f(x) = 1, \quad a \leq x \leq b, \quad (1.82)$$

and

$$F(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } 0 \leq x < 1, \\ 1, & \text{if } x \geq 1. \end{cases}$$

Figure 1.11 shows the pdf and cdf of the standard uniform random variable.

Illustrative Example 1.38 (Simply supported beam). Consider the same simply supported beam AB from Example 1.1 (see Figure 1.1) with one concentrated force P_1 that can be placed everywhere along the beam. This force is assumed to be uniform on the length L , because all the positions along the beam are equally likely to receive the force. Thus, the load position X is a uniform random variable, $U(0, L)$, with pdf

$$f(x) = x/L; \quad 0 \leq x \leq L.$$

■

The family of uniform random variables is stable with respect to changes of location and scale, that is, if X is $U(\alpha, \beta)$, then the variable $Y = cX + d$ is uniform $U(c\alpha + d, c\beta + d)$.

Illustrative Example 1.39 (Temperatures). Suppose that the temperature, in degrees Celsius, at a given time and location is $U(30, 40)$ random variable. Since $F = 1.8C + 32$, where F and C are the temperatures measured in degrees Fahrenheit and Celsius, respectively, then the temperature in degrees Fahrenheit is an $U(86, 104)$ random variable. ■

1.7.2 Exponential Distribution

Let X be the time between two consecutive Poisson events with intensity λ events per unit of time (see Section 1.6.6) such as the time between failures of machines or the time between arrivals at checkout counter. That is, we start at the time when the first event occurs and measure the time to the next event. In other words, X is the *inter-arrival* or *recurrence* time. Then X is a continuous random variable. What is the pdf and cdf of X ? Consider the event that X exceeds x , that is, the second event occurs after time x since the occurrence of the first event. The probability of this event is $\Pr(X > x) = 1 - \Pr(X \leq x) = 1 - F(x)$, where $F(x)$ is the cdf of the random variable X . This event, however, since the occurrences of an event in non-overlapping time intervals in a Poisson process are statistically independent, is equivalent to saying that no Poisson events have occurred before time x (implying no occurrence ($x = 0$) in time t). Replacing λ by λx in the Poisson pmf in (1.76), the probability of obtaining zero Poisson events is $P(0) = e^{-\lambda x}$. Therefore, we have

$$1 - F(x) = e^{-\lambda x},$$

from which it follows that the cdf of X is

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - e^{-\lambda x}, & \text{if } x \geq 0, \lambda > 0. \end{cases}$$

Taking the derivative of $F(x)$ with respect to x , we obtain the pdf

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} 0, & \text{if } x < 0, \\ \lambda e^{-\lambda x}, & \text{if } x \geq 0. \end{cases} \quad (1.83)$$

The random variable X whose pdf is given in (1.83) is called an *Exponential* random variable with parameter λ (also known as *negative exponential*). Note the exponential distribution is the continuous analogue of the Geometric distribution (see Section 1.6.3). When X is replaced by $-Y$ in (1.83), we obtain the pdf of the *reversed* Exponential random variable,

$$f(y) = \begin{cases} \lambda e^{\lambda y}, & \text{if } y < 0, \lambda > 0, \\ 0, & \text{if } y \geq 0. \end{cases} \quad (1.84)$$

The graphs of the pdf and cdf of two Exponential random variables are shown in Figure ???. It can be shown that the mean and variance of the Exponential random variable are

$$\mu = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2 = \frac{1}{\lambda^2}, \quad (1.85)$$

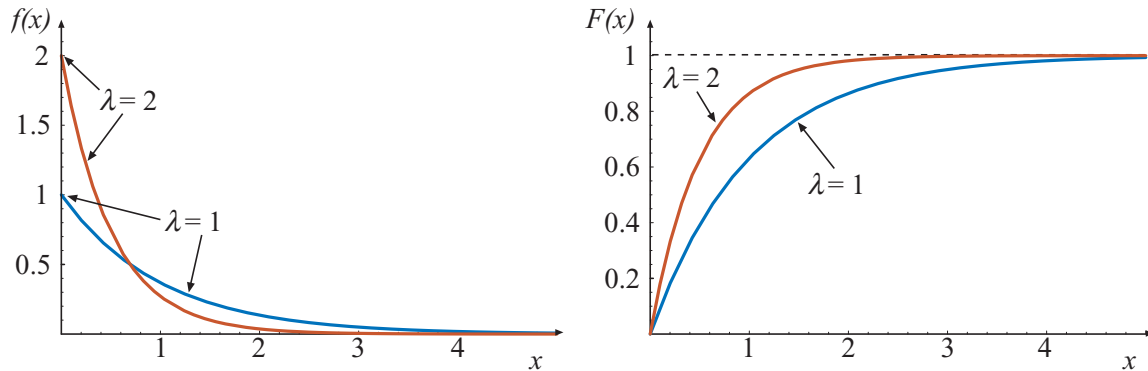


Figure 1.12: An example of the pdf and cdf of two Exponential random variables.

which means that the mean of the first occurrence time, or the mean recurrence time, or simply the return period for the Poisson model is $\frac{1}{\lambda}$.

The pdf of the Exponential distribution in (1.83) can also be expressed as

$$f(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{1}{\delta} e^{-x/\delta}, & \text{if } x \geq 0, \delta > 0. \end{cases} \quad (1.86)$$

This is simply a reparameterization of (1.83), where λ is replaced by $1/\delta$. In this form the cdf is

$$F(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - e^{-x/\delta}, & \text{if } x > 0, \delta > 0. \end{cases}$$

and the mean and variance are simply $\mu = \delta$ and $\sigma^2 = \delta^2$, respectively.

Exponential random variables have the so-called *memoryless* or *no-aging* property, that is,

$$\Pr(X > a + b | X > a) = \Pr(X > b).$$

In words, if X is associated with lifetime the probability of X exceeding a given time b is the same no matter which time origin a is considered, from which the terminology *no-aging* was derived.

Illustrative Example 1.40 (Waiting time at an intersection). When a car arrives at the intersection of two roads, it stops and then it needs a minimum time of t_0 seconds without passing cars to initiate the movement. If the arrival time, X , is assumed to be exponential with intensity λ cars/second, the probability of the waiting time to be zero is given by

$$\Pr(X > t_0) = 1 - \Pr(X \leq t_0) = 1 - F(t_0) = e^{-\lambda t_0}.$$

■

Illustrative Example 1.41 (Time between consecutive storms). Assume that the occurrence of storms is Poissonian with rate λ storms/year. Then, the time until the

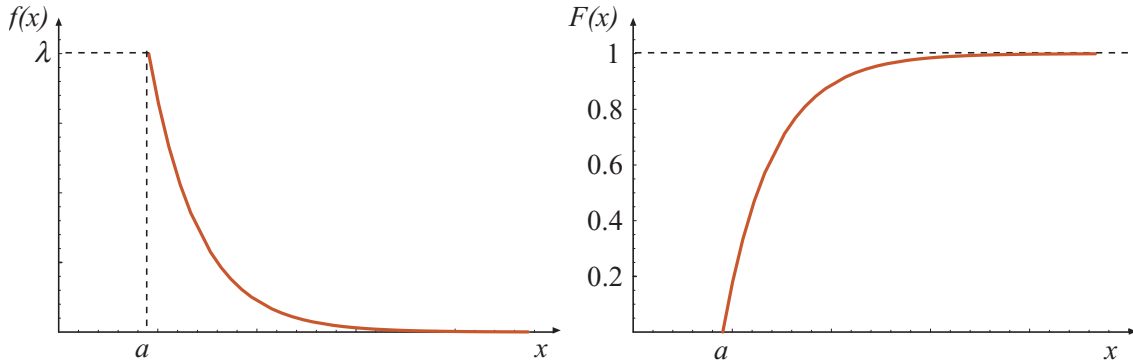


Figure 1.13: Pdf and cdf of the Shifted exponential distribution.

occurrence of the first storm and the time between consecutive storms are exponential random variables with parameter λ . For example, assume that $\lambda = 5$ storms/year. Then, the probability of the time until the occurrence of the first storm or the time between consecutive storms to be smaller than 1 month is

$$p = \Pr(X < 1/12) = F(1/12) = 1 - \lambda e^{-\lambda \times 1/12} = 1 - 5e^{-5/12} = 0.7253.$$

■

If in equation (1.83) the density function starts at any value a instead of $x = 0$, the resulting distribution is called *shifted exponential distribution*, with the corresponding pdf and cdf given as follows:

$$f(x) = \begin{cases} 0, & \text{if } x < a, \\ \lambda e^{-\lambda(x-a)}, & \text{if } x \geq a, \end{cases} \quad (1.87)$$

and

$$F(x) = \begin{cases} 0, & \text{if } x < a, \\ 1 - e^{-(x-a)/\lambda}, & \text{if } x \geq a, \lambda > 0. \end{cases} \quad (1.88)$$

Graphically these functions would appear as shown in Figure 1.13.

1.7.3 Gamma Random Variable

The Gamma random variable is a generalization of the exponential random variable. Consider a Poisson time process with intensity λ events per unit time. The time it takes for the first event to occur is an exponential random variable with parameter λ . Now, let X be the time up to the occurrence of θ Poisson events. If $\theta = 1$, then X is an exponential random variable, but if $\theta > 1$, then X is a Gamma random variable. What is then the pdf of a gamma random variable?

To derive the pdf of a Gamma random variable, we first introduce a useful function called the *Gamma function*, which is the generalization of the factorial for non-integer θ defined as

$$\Gamma(\theta) = \int_0^{\infty} y^{\theta-1} e^{-y} dy. \quad (1.89)$$

Some important properties of the Gamma function are:

$$\Gamma(0.5) = \sqrt{\pi}, \quad (1.90)$$

$$\Gamma(\theta) = (\theta - 1)\Gamma(\theta - 1), \quad \text{if } \theta \geq 1, \quad (1.91)$$

$$\Gamma(\theta) = (\theta - 1)!, \quad \text{if } \theta \text{ is a positive integer.} \quad (1.92)$$

Now, if X is the time it takes for the θ Poisson events to occur, then the probability that X is in the interval $(x, x + dx)$ is $\Pr(x \leq X \leq x + dx) = f(x)dx$. But this probability is equal to the probability of having $\theta - 1$ Poisson events occurred in a period of duration x times the probability of the occurrence of one event in a period of duration dx . Thus, we have

$$f(x)dx = \frac{e^{-\lambda x}(\lambda x)^{\theta-1}}{(\theta - 1)!} \lambda dx,$$

from which we obtain

$$f(x) = \frac{\lambda^\theta x^{\theta-1} e^{-\lambda x}}{(\theta - 1)!}; \quad 0 \leq x < \infty. \quad (1.93)$$

Using the property of the Gamma function in (1.92), Expression (1.93) can be written as

$$f(x) = \frac{\lambda^\theta x^{\theta-1} e^{-\lambda x}}{\Gamma(\theta)}; \quad 0 \leq x < \infty, \quad (1.94)$$

which is valid for any real positive θ . The pdf in (1.94) is known as the Gamma distribution with parameters θ and λ , and is denoted by $G(\theta, \lambda)$. Note that when $\theta = 1$, the pdf in (1.94) becomes (1.83), which is the pdf of the exponential random variable. For θ integer, the distribution is also known as *Erlang distribution*.

It may be recognized that the gamma distribution is the continuous analogue of the negative-binomial distribution, in the sense that it is related to the Poisson process in the same way than the negative binomial distribution is related to the Bernoulli sequence.

In general, the cdf of the Gamma random variable

$$F(x) = \int_0^x f(x)dx = \int_0^x \frac{\lambda^\theta x^{\theta-1} e^{-\lambda x}}{\Gamma(\theta)!} dx, \quad (1.95)$$

which is called the *incomplete Gamma function*, does not have a closed-form, but can be obtained by numerical integration. The pdf and cdf of the Gamma random variable is graphed in Figure 1.14.

The mean, variance and skewness coefficient of $G(\theta, \lambda)$ are:

$$\mu = \frac{\theta}{\lambda}, \quad \sigma^2 = \frac{\theta}{\lambda^2} \quad \text{and} \quad \gamma_1 = \frac{2}{\sqrt{\theta}}. \quad (1.96)$$

The family of gamma variable has the following important properties:

1. It is reproductive with respect to parameter θ , that is, if $X_i \sim G(\theta_i, \lambda)$; $i = 1, \dots, n$ are independent, then

$$\sum_{i=1}^n X_i \sim G\left(\sum_{i=1}^n \theta_i, \lambda\right).$$

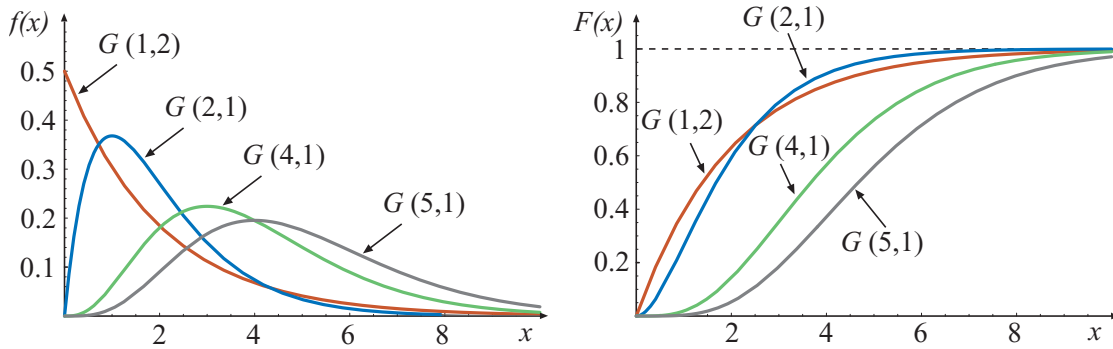


Figure 1.14: Examples of the pdf and cdf of some Gamma random variables.

2. It is stable with respect to scale changes, i.e., if X is $G(\theta, \lambda)$, then cX is $G(\theta, \lambda/c)$.
3. It is not stable with respect to changes of location. In other words, if X is a gamma random variable, then $X + a$ is not a Gamma random variable.

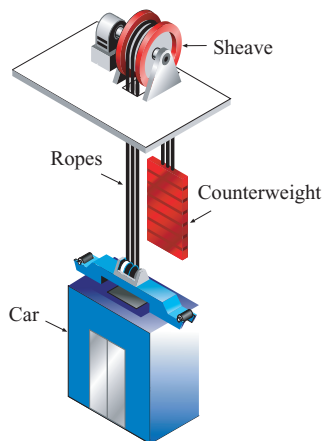


Figure 1.15: A rope elevator, consisting of a system of 3 parallel ropes.

Illustrative Example 1.42 (Elevator cable system). Consider the rope elevator shown in Figure 1.15, where the car is raised and lowered by traction steel ropes. The ropes that lift the car are also connected to a counterweight, which hangs on the other side of the sheave. The counterweight weighs about the same as the car filled to 40-percent capacity. In other words, when the car is 40 percent full (an average amount), the counterweight and the car are perfectly balanced. Elevators are built with several redundant safety systems that keep them in position. The first line of defense is the rope system itself. Each elevator rope is made from several lengths of steel material wound around one another. With this sturdy structure, one rope can support the weight of the elevator car and the counterweight

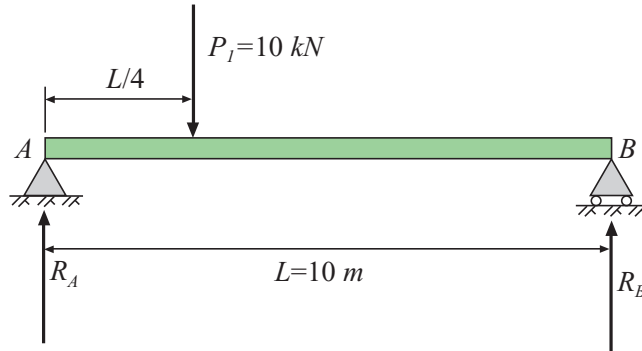


Figure 1.16: Simple supported beam.

on its own. But elevators are built with multiple ropes (between four and eight, typically). In the unlikely event that one of the ropes snaps, the rest will hold the elevator up. Assuming no replacement after failure, and that each rope is slightly longer than the other ones (only one rope is working at the same time) the useful life of the elevator is the sum of the lives of the individual ropes. If the life of each of the individual ropes is assumed to be $G(\theta, \lambda)$, then as the elevator has 3 ropes, its life X is $G(3\theta, \lambda)$. ■

Illustrative Example 1.43 (Simply supported beam). Consider the simply supported beam AB shown in Figure 1.16 with one concentrated force P_1 acting at a distance $L/4$ from the left support A . If the force P_1 is assumed to be $G(2, 1)$, it can be shown that the equilibrium conditions lead to supports reactions $R_A = 3P_1/4$ and $R_B = P_1/4$, respectively. Thus, both reactions will be Gamma distributed and since $G(\theta, \lambda)$ is stable with respect to scale changes, then, cP_1 is $G(2, 1/c)$. Therefore, $R_A \sim G(2, 4/3)$ and $R_B \sim G(2, 1/4)$. ■

Illustrative Example 1.44 (Farmers subsidy). Farmers receive a monetary subsidy after having three floods. Assume that the floods are Poisson events with rate 0.5 floods/year. Then, the time to the third flood is a $G(3, 0.5)$. Thus, the mean time until farmers receive subsidy is $\mu = \theta/\lambda = 3/0.5 = 6$ years. Also, the probability that a subsidy will be received after 10 years is

$$\Pr(X > 10) = 1 - F(10) = 1 - \int_0^{10} \frac{0.5(0.5x)^2 e^{-0.5x}}{2!} dx = 1 - 0.8753 = 0.1247.$$

1.7.4 Beta Random Variable

In 1895 Karl Pearson reasoned that if the binomial distribution would yield the normal distribution as a limit, the limit of the hypergeometric distribution might produce a more flexible continuous probability distribution capable of representing skewed variations. He

produced a parent distribution with a wide range of sibling probability distributions, exhibiting skewed distributions and symmetrical bell-shaped curves, the uniform distribution, and U-shaped, J-shaped and reverse J-shaped curves.

The chief advantage of the Beta random distribution is its great flexibility in being able to be fitted to experimental data. It is very useful for modelling experimental data with range limited to the interval $[a, b]$. For example, when X is the proportion of impurities in a chemical product or the proportion of time that a machine is under repair. Then X must satisfy $a \leq X \leq b$. The Beta random variable is used to model such experimental data. Its name is due to the presence of the *Beta function* in its pdf. The Beta function is defined as

$$\beta(\lambda, \theta) = \int_0^1 x^{\lambda-1}(1-x)^{\theta-1} dx; \quad \lambda > 0, \theta > 0. \quad (1.97)$$

Note that the Beta function is related to the Gamma function by

$$\beta(\lambda, \theta) = \frac{\Gamma(\lambda)\Gamma(\theta)}{\Gamma(\lambda + \theta)}, \quad (1.98)$$

and, for λ, θ integer

$$\beta(\lambda, \theta) = \frac{(\lambda - 1)!(\theta - 1)!}{(\lambda + \theta - 1)!}, \quad (1.99)$$

The pdf of a Beta random variable is given by

$$f(x) = \frac{(x-a)^{\lambda-1}(b-x)^{\theta-1}}{\beta(\lambda, \theta)} = \frac{\Gamma(\lambda + \theta)}{\Gamma(\lambda)\Gamma(\theta)}(x-a)^{\lambda-1}(b-x)^{\theta-1}; \quad a \leq x \leq b, \quad (1.100)$$

where $\lambda > 0$ and $\theta > 0$. The Beta random variable is denoted by $\beta(\lambda, \theta)$.

The cdf of the $\beta(\lambda, \theta)$ is

$$F(x) = \int_0^x f(x)dx = \int_0^x \frac{(x-a)^{\lambda-1}(b-x)^{\theta-1}}{\beta(\lambda, \theta)} dx, \quad (1.101)$$

using the new variable

$$Y = \frac{X-a}{b-a}$$

whose range is between 0 and 1. The preceding integral with this change of variable is equal to

$$F(x) = \int_0^{(x-a)/(b-a)} f(y)dy = \int_0^{(x-a)/(b-a)} \frac{y^{\lambda-1}(1-y)^{\theta-1}}{\beta(\lambda, \theta)} dy = \frac{I_x(\lambda, \theta)}{\beta(\lambda, \theta)}, \quad (1.102)$$

where $I_x(\lambda, \theta)$ is called the *incomplete Beta function*, which cannot be given in closed form, but can be obtained by numerical integration. Values of the *incomplete beta function ratio* have been tabulated; for example, by Pearson, and Pearson and Johnson.

If the values of the variate are limited between 0 and 1 ($a = 0$ and $b = 1$), equation (1.100) becomes

$$f(x) = \frac{x^{\lambda-1}(1-x)^{\theta-1}}{\beta(\lambda, \theta)} = \frac{\Gamma(\lambda + \theta)}{\Gamma(\lambda)\Gamma(\theta)}x^{\lambda-1}(1-x)^{\theta-1}; \quad 0 \leq x \leq 1, \quad (1.103)$$

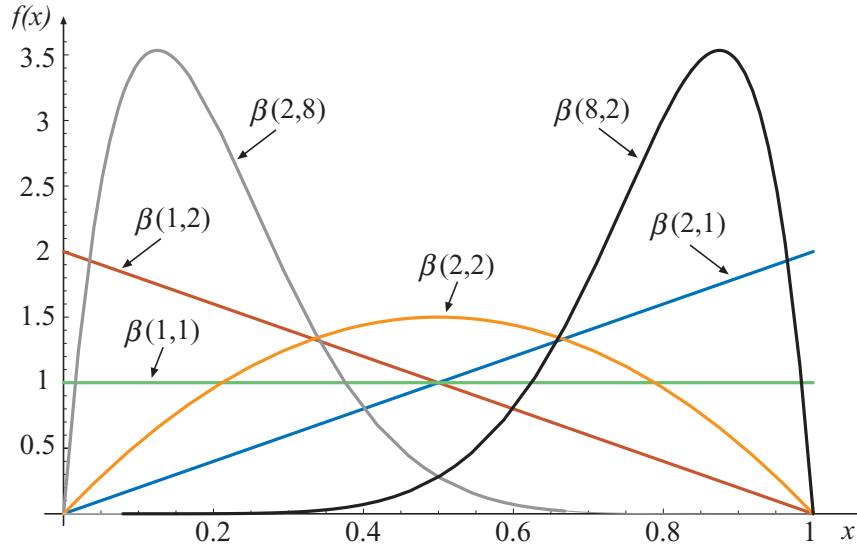


Figure 1.17: Examples showing that the probability density functions of Beta random variables take wide range of different shapes.

which can be called the standard beta distribution.

The mean, variance, mode and skewness coefficient of the Beta random variable are

$$\mu = a \frac{\lambda(b-a)}{\lambda+\theta}, \quad \sigma^2 = \frac{\lambda\theta(b-a)^2}{(\lambda+\theta+1)(\lambda+\theta)^2},$$

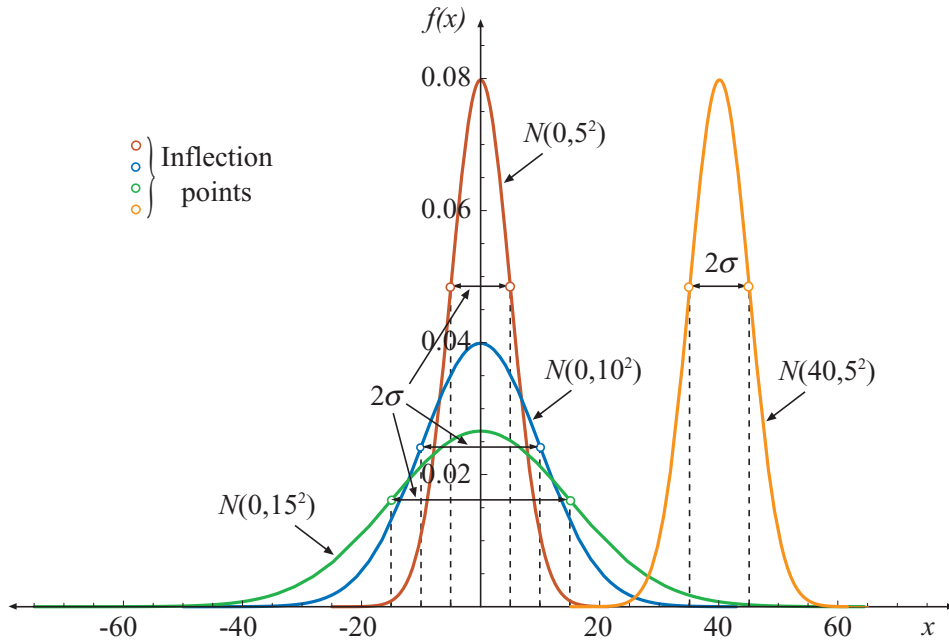
$$\tilde{x} = a + \frac{\lambda(b-a)(1-\lambda)}{2-\lambda-\theta} \quad \text{and} \quad \gamma_1 = \frac{2(\lambda-\theta)}{(\lambda+\theta)(\lambda+\theta+2)\sigma},$$

respectively.

In Figure 1.17 shows different examples of the pdf of the Beta random variable. Two particular cases of the Beta distribution are interesting. Setting $(\lambda = 1, \theta = 1)$, gives the *standard uniform* random variable, $U(0, 1)$, while setting $(\lambda = 1, \theta = 2)$ or $(\lambda = 2, \theta = 1)$ gives the *triangular* random variable whose cdf is given by $f(x) = 2x$ or $f(x) = 2(1-x)$, $0 \leq x \leq 1$.

1.7.5 Normal or Gaussian Distribution

One of the most important distributions in probability and statistics is the *normal* distribution (also known as the *Gaussian* distribution), which is a reasonable model for observations of many physical processes or properties. In practice it arises in various applications as a limiting case of other distributions. For example, consider the random variable, X , which is the sum of n *identically and independently distributed* (iid) random variables X_1, X_2, \dots, X_n . Then, by the *central limit theorem*, X is asymptotically (as $n \rightarrow \infty$) normal, regardless of the form of the distribution of the random variables X_1, X_2, \dots, X_n .

Figure 1.18: Normal probability density functions, $N(\mu, \sigma^2)$.

A precise mathematical statement and proof of the central limit theorem is given by Parzen [?]. In a qualitative way the *central limit theorem* (so named by G. Polya in 1920) states the phenomenon that the sum of many small chance effects is normally distributed. It does not require that the random variables have the same distribution or even that they be entirely independent.

The normal random variable with mean μ and variance σ^2 is denoted by $X \sim N(\mu, \sigma^2)$ and its pdf is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad -\infty < x < \infty, \quad (1.104)$$

where $-\infty < \mu < \infty$ and $\sigma > 0$. The mean and variance of a normal random variable are μ and σ^2 , respectively. Figure 1.18 is a graph of the pdfs of a $N(0, 5^2)$, $N(0, 10^2)$, $N(0, 15^2)$ and $N(40, 5^2)$. Note that the pdfs are symmetric about the means $\mu = 0$ and $\mu = 40$, respectively. Also, the pdfs have two inflection points; one on each side of the mean μ and equally distant from μ . The standard deviation σ are equal to the distances between the means and the inflection points. Like any other continuous random variables, the area under the curve is 1. The cdf of the normal random variable does not exist in closed form, but can be obtained by numerical integration and there are many approximations (see Abramowitz and Stegun [?] and Hastings [?]).

The effect of the parameters μ and σ on the pdf can be seen in Figure 1.18, which shows the pdf of four normal random variables, three of them with the same mean but different standard deviations and the other one with different mean. The higher the standard deviation the flatter the pdf.

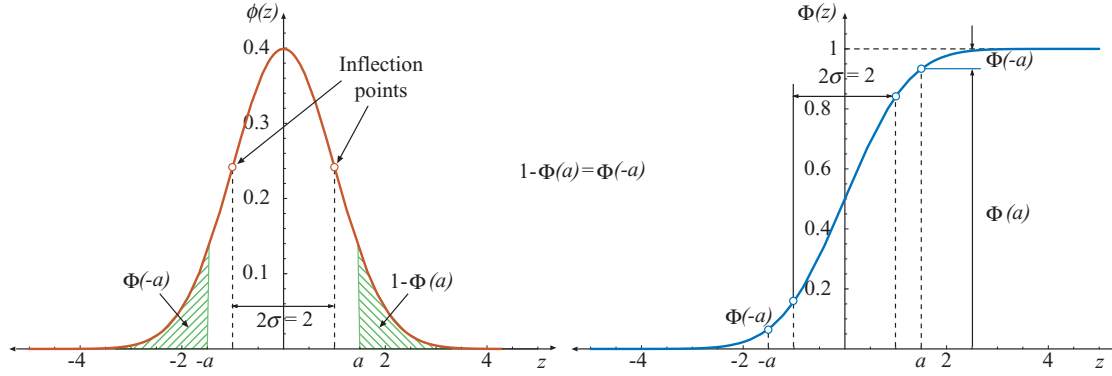


Figure 1.19: The pdf, $\phi(z)$, and the cdf, $\Phi(z)$ of the standard normal random variable, $N(0, 1^2)$.

If X is $N(\mu, \sigma^2)$ then the random variable

$$Z = \frac{X - \mu}{\sigma}, \quad (1.105)$$

is $N(0, 1^2)$. The normal random variable with mean 0 and standard deviation 1 is called the *standard normal* random variable. From (1.104), the pdf of Z is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}; \quad -\infty < z < \infty, \quad (1.106)$$

The corresponding cdf is

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx; \quad -\infty < z < \infty, \quad (1.107)$$

The pdf and cdf of the standard normal random variable are shown in Figure 1.19. This cdf also does not exist in closed form. However, it has been computed numerically and is given in the Appendix as Table ???. Note that because of the symmetry of the normal density $\Phi(-z) = 1 - \Phi(z)$.

The main interest of the change of variable in (1.105) is that we can use Table ??? to calculate probabilities for any other normal distribution. For example, if $X \sim N(\mu, \sigma^2)$, then

$$\Pr(X \leq x) = F(x) = \Pr\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = \Pr\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right),$$

and

$$\Pr(X = x) = f(x) = \frac{dF(x)}{dx} = d\Phi\left(\frac{x - \mu}{\sigma}\right) / dx = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right),$$

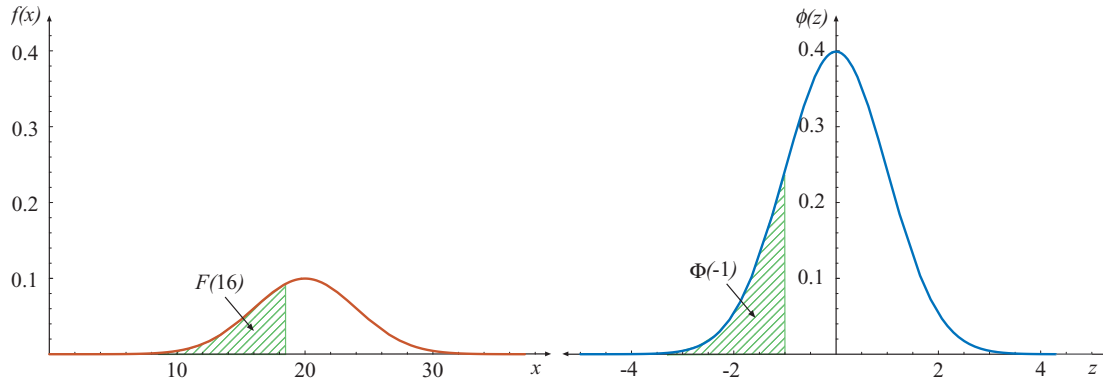


Figure 1.20: The pdf, $f(x)$, where $X \sim N(20, 4^2)$, and the pdf, $\phi(z)$, where $Z \sim N(0, 1^2)$. The shaded area under $f(x)$ to the left of $x = 16$ is equal to the shaded area under $\phi(z)$ to the left of $z = -1$, and the cdf, $\Phi(z)$ of the standard normal random variable, $N(0, 1)$.

where $\Phi(\bullet)$ and $\phi(\bullet)$ are the cdf and the pdf of the standard normal distribution in (1.107) and (1.106), respectively.

Note that the skewness and kurtosis coefficients are 0 (symmetrical distribution) and 3, respectively. The value of the kurtosis coefficient of the normal distribution is used as a criterion for distinguishing the relative flatness or peakedness of distributions.

Illustrative Example 1.45 (Normal variable). Assume that a simple compression strength is a normal random variable with mean $\mu = 20$ MPa and a standard deviation 4 MPa. Then, the probability that the compression strength is less than or equal to 16 MPa, is

$$\begin{aligned} \Pr(X \leq 16) &= F(16) = \Phi\left(\frac{16 - 20}{4}\right) \\ &= \Phi(-1) = 1 - \Phi(1) = 1 - 0.841345 = 0.158655. \end{aligned}$$

where $\Phi(1)$ is obtained from Table ???. Figure 1.20 shows that $\Pr(X \leq 16) = \Pr(Z \leq -1)$. This probability is equal to the shaded areas under the two curves. ■

The family of normal distributions is reproductive respect to the parameters μ and σ , that is, if $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, and X_1 and X_2 are independent, then

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

If the random variables X_j , $j = 1, 2, \dots, n$, are independent and normal $N(\mu_j, \sigma_j^2)$, then the random variable

$$Y = \sum_{j=1}^n c_j X_j; \quad c_j \in \mathbb{R}, \quad \forall j \in 1, 2, \dots, n \quad (1.108)$$

is normal with

$$\mu = \sum_{j=1}^n c_j \mu_j \quad \text{and} \quad \sigma^2 = \sum_{j=1}^n c_j^2 \sigma_j^2.$$

This shows that the normal family is stable with respect to linear combinations.

Normal Approximation to the Binomial Distribution

We know from Section 1.6.2 that the mean and variance of a Binomial random variable are $\mu = np$ and $\sigma^2 = np(1 - p)$. If the parameter n is large and neither p nor $(1 - p)$ are very close to zero, the variable

$$Z = \frac{X - np}{\sqrt{np(1 - p)}} \quad (1.109)$$

is approximately $N(0, 1^2)$. This allows approximating the binomial probabilities using the normal probabilities. In practice, good approximations are obtained if $np, n(1 - p) > 5$.

Illustrative Example 1.46 (Normal approximation). Knowing that 30% of patients entering a hospital with miocard infarct die in the hospital. If 2000 patients enter in one year and X is the number of these patients who will die in the hospital, then X is $B(2000, 0.3)$. Since n is large, $np = 600 > 5$, and $n(1 - p) = 1400 > 5$, we can use the normal approximation to the binomial. Since $\mu = 2000 \times 0.3 = 600$ patients and $\sigma^2 = 2000 \times 0.3 \times 0.7 = 420$ patients, then X can be approximated by $N(600, 420)$. Thus, for example, the probability that a maximum of 550 patients die in the hospital is:

$$\begin{aligned} \Pr(X \leq 550) &\approx \Pr(Z \leq (550 - 600)/\sqrt{420}) \\ &= \Phi(-2.44) = 1 - \Phi(2.44) \\ &= 1 - 0.9927 = 0.0073, \end{aligned}$$

where $\Phi(2.44)$ is obtained from Table ?? in the Appendix.

The exact result considering the Binomial distribution is the following

$$\Pr(X \leq 550) = \sum_{x=1}^{550} \binom{2000}{x} p^x (1 - p)^{2000-x} = 0.0075.$$

■

1.7.6 Log-Normal Distribution

In many engineering applications the variables involved can not have negative values due to the physical aspects of the problem. In this situations, modeling the variable as lognormal is more appropriate, eliminating the possibility of negative values.

A random variable X is log-normal $LN(\mu_{\ln X}, \sigma_{\ln X}^2)$ when its natural logarithm, $\ln(X)$, is normal $N(\mu_{\ln X}, \sigma_{\ln X}^2)$. The pdf of the log-normal random variable can be expressed as:

$$f(x) = \frac{1}{x\sigma_{\ln X}\sqrt{2\pi}} \exp\left[-\frac{(\ln(x) - \mu_{\ln X})^2}{2\sigma_{\ln X}^2}\right]; \quad x \geq 0, \quad (1.110)$$

where the parameters $\mu_{\ln X} = E[\ln X]$ and $\sigma_{\ln X} = \sqrt{E[(\ln X - \mu_{\ln X})^2]}$ are the mean and the standard deviation of the natural logarithm of the random variable X . The mean and variance (the first two moments) of the log-random variable X are

$$\mu_X = e^{\mu_{\ln X} + \sigma_{\ln X}^2/2} \quad \text{and} \quad (\sigma_X)^2 = e^{2\mu_{\ln X}}(e^{2\sigma_{\ln X}^2} - e^{\sigma_{\ln X}^2}). \quad (1.111)$$

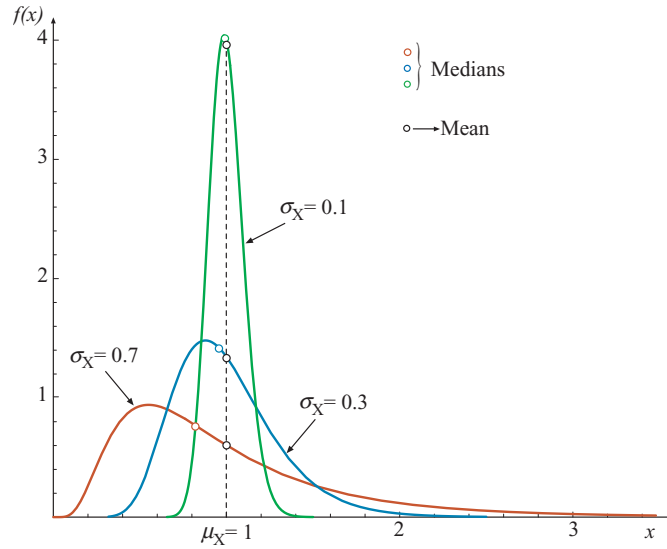


Figure 1.21: Log-normal probability density functions.

In Figure 1.21 are shown the probability density functions of three log-normal random variables with mean $\mu_X = 1$, and standard deviations $\sigma_X = 0.1$, $\sigma_X = 0.3$ and $\sigma_X = 0.7$, respectively.

Note that the parameters of the log-normal distribution can be calculated from the information on the two parameters from the normal distribution, the mean (μ_X) and the standard deviation (σ_X) of the sample population. It can be shown that

$$\mu_{\ln X} = \ln \frac{\mu_X}{\sqrt{(1 + v_X^2)}} \quad \text{and} \quad \sigma_{\ln X}^2 = \ln(1 + v_X^2), \quad (1.112)$$

where $v_X = \sigma_X/\mu_X$ is the coefficient of variation.

Note that if in expression (1.112), the coefficient of variation $v_X = \sigma_X/\mu_X$ is not very large (≤ 0.30), then

$$\sigma_{\ln X}^2 = \ln(1 + v_X^2) \approx v_X^2, \quad (1.113)$$

that is, the standard deviation of the natural logarithm of the random variable X is approximately equal to the coefficient of variation of the log-normal random variable X .

Denoting as x_m the median of a log-normal random variable X , it can be shown that,

$$\Pr(X \leq x_m) = 0.5 \rightarrow \Phi \left(\frac{\ln x_m - \mu_{\ln X}}{\sigma_{\ln X}} \right) = 0.5 \rightarrow \ln x_m = \mu.$$

Therefore, the median is

$$x_m = \exp \mu_{\ln X},$$

using expression (1.112) it can be concluded that

$$x_m = \frac{\mu_X}{\sqrt{(1 + v_X^2)}},$$

that is, the median value of a lognormal variable is always less than the mean value (see Figure 1.21). This statement implies that the skewness is always positive.

In some applications, some random variables are defined to be the products (instead of sums) of iid positive random variables. In these cases, taking the logarithm of the product yields the sum of the logarithms of its components. Thus, by the central limit theorem, the logarithm of the product of n iid random variables is asymptotically normal. In other words, the product of a large number of random variables approaches the log-normal distribution, irrespective of the individual distributions of the random variables.

The log-normal random variable is not reproductive with respect to its parameters $\mu_{\ln X}$ and $\sigma_{\ln X}^2$, but stable with respect to products of independent variables, that is, if $X_1 \sim LN(\mu_1, \sigma_1^2)$ and $X_2 \sim LN(\mu_2, \sigma_2^2)$, then

$$X_1 \times X_2 \sim LN(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Generally speaking, if $Y = \prod_{i=1}^n X_i$ where $X_i \sim LN(\mu_i, \sigma_i^2)$ with median x_i , then

$$Y \sim LN\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right) \text{ and } y_m = \prod_{i=1}^n x_i. \quad (1.114)$$

Illustrative Example 1.47 (Log-Normal variable). Consider the Example 1.45 but assuming that the simple compression strength is a log-normal random variable with mean $\mu_X = 20 \text{ MPa}$ and a standard deviation $\sigma_X = 4 \text{ MPa}$. What is the probability that the compression strength is less than or equal to 16 MPa ?

Using the expression (1.112) we first obtain the parameters $\mu_{\ln X}$ and $\sigma_{\ln X}$

$$\mu_{\ln X} = \ln \frac{20}{\sqrt{1 + 4/20^2}} = 2.97612 \quad \text{and} \quad \sigma_{\ln X} = \sqrt{\ln(1 + 4/20^2)} = 0.198042,$$

note that in this case $v_X = 4/20 = 0.2 \leq 0.3$, that is why $v_X \approx \sigma_X$. Then the probability that the compression strength is less or equal to 16 is

$$\begin{aligned} \Pr(X \leq 16) &= F(16) = \Phi\left(\frac{\ln 16 - \mu_{\ln X}}{\sigma_{\ln X}}\right) \\ &= \Phi(-1.02773) = 1 - \Phi(1.02773) = 1 - 0.847961 = 0.152039. \end{aligned}$$

■

1.7.7 Logistic Distribution

A random variable X is said to have a logistic distribution if its cdf is given by:

$$F(x) = \frac{1}{1 + \exp\{-(x - \alpha)/\beta\}}; \quad \beta > 0, \quad -\infty < x < \infty, \quad (1.115)$$

where α and β are location and scale parameters, respectively. Note that the logistic distribution (1.115) is symmetric about $x = \alpha$ and has a shape similar to that of the normal

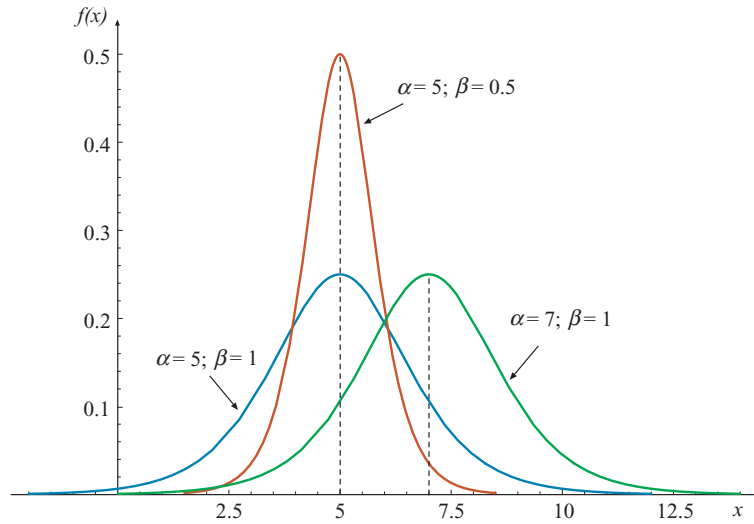


Figure 1.22: Logistic probability density functions.

distribution as it is shown in Figure 1.22. Note that the higher the scale parameter (β) the flatter the pdf.

The use of logistic function as a growth curve can be justified as follows. Consider the differential equation:

$$\frac{dF(x)}{dx} = k [F(x) - a] [b - F(x)], \quad (1.116)$$

where k , a and b are constants with $k > 0$ and $b > a$. In other words, the rate of growth is equal to the excess over the initial asymptotic value a times the deficiency compared with final asymptotic value b . The solution of the differential equation (1.116) with $a = 0$ and $b = 1$ (the asymptotic limits of the cdf) is

$$F(x) = \left(1 + ce^{-x/k}\right)^{-1},$$

where c is a constant. This is the same as the logistic distribution in (1.115) with $k = \beta$ and $c = e^{\alpha/\beta}$. Equation (1.116) is used as a model of autocatalysis (see [?]).

From (1.115), the pdf of the logistic random variable is

$$f(x) = \frac{\exp\{-(x-\alpha)/\beta\}}{\beta [1 + \exp\{-(x-\alpha)/\beta\}]^2}; \quad \beta > 0, \quad -\infty < x < \infty. \quad (1.117)$$

The corresponding characteristic function is:

$$\psi(t) = e^{it\alpha} \Gamma(1 + i\beta t) \Gamma(1 - i\beta t). \quad (1.118)$$

The mean and variance of the logistic random variable are:

$$\mu = \alpha, \quad \text{and} \quad \sigma^2 = \frac{\pi^2 \beta^2}{3}. \quad (1.119)$$

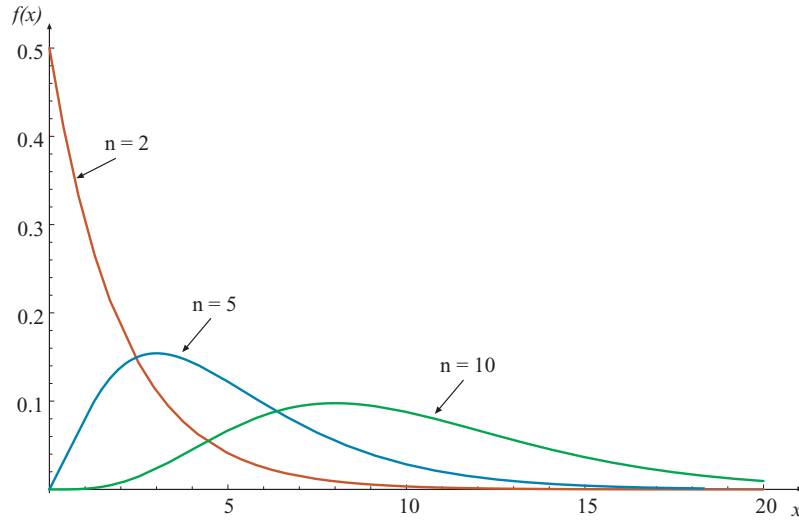


Figure 1.23: Central chi-square probability density functions.

A simple relationship between the cdf (1.115) and the pdf (1.117) is

$$f(x) = \frac{1}{\beta} F(x)[1 - F(x)].$$

This relation is useful to show several probabilistic properties.

1.7.8 Chi-Square and Chi Random Variables

Let Y_1, \dots, Y_n be independent random variables, where Y_i is distributed as $N(\mu_i, 1)$. Then, the variable

$$X = \sum_{i=1}^n Y_i^2$$

is called a *noncentral Chi-square* random variable with n degrees of freedom and *noncentrality parameter* $\lambda = \sum_{i=1}^n \mu_i^2$. It is denoted as $\chi_n^2(\lambda)$. When $\mu_i = 0$, for all i , then $\lambda = 0$ and we obtain the *central Chi-square* random variable, which is denoted by χ_n^2 . The pdf of the central Chi-squared random variable (see Figure 1.23) with n degrees of freedom is

$$f(x) = \frac{x^{(n/2)-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)}; \quad x \geq 0, \quad (1.120)$$

where $\Gamma(\cdot)$ is the Gamma function defined in (??). The cdf $F(x)$ cannot be given in closed form. However, it has been computed numerically and is given in the Appendix as Table ???. The mean and variance of a Chi-squared random variable with n degrees of freedom are:

$$\mu = n + \lambda \quad \text{and} \quad \sigma^2 = 2(n + 2\lambda). \quad (1.121)$$

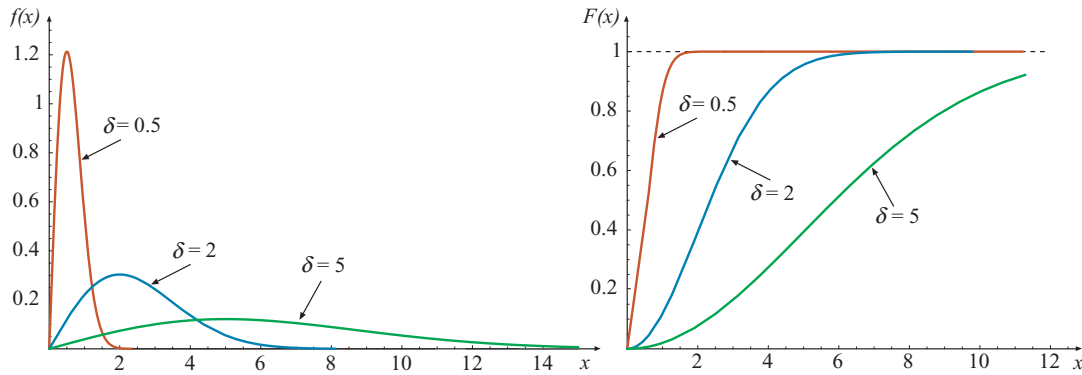


Figure 1.24: Rayleigh probability density functions.

The $\chi_n^2(\lambda)$ variable is reproductive with respect to n and λ , that is, if $X_1 \sim \chi_{n_1}^2(\lambda_1)$ and $X_2 \sim \chi_{n_2}^2(\lambda_2)$, then

$$X_1 + X_2 \sim \chi_{(n_1+n_2)}^2(\lambda_1 + \lambda_2).$$

The positive square root of a $\chi_n^2(\lambda)$ random variable is called a *Chi* random variable and is denoted by $\chi_n(\lambda)$.

Note that the Chi-squared distribution is very useful in the interval estimation of the population variance.

1.7.9 Rayleigh Random Variable

An interesting particular case of the χ_n random variable is the *Rayleigh* random variable, which is obtained when $n = 2$. The pdf of the Rayleigh random variable is given by

$$f(x) = \delta^{-2} x \exp \left[-\frac{1}{2} \left(\frac{x}{\delta} \right)^2 \right]; \quad \delta > 0, x \geq 0. \quad (1.122)$$

The corresponding cdf is

$$F(x) = 1 - \exp \left[-\frac{1}{2} \left(\frac{x}{\delta} \right)^2 \right]. \quad (1.123)$$

In Figure 1.24 are shown the pdfs and cdfs of several Rayleigh distributions.

The mean and variance are

$$\mu = \delta \sqrt{\pi/2} \quad \text{and} \quad \sigma^2 = \delta^2(4 - \pi)/2. \quad (1.124)$$

The Rayleigh distribution was originally derived by Lord Rayleigh in the late 19th century to describe the distribution of the intensity of sounds emitted from an infinite number of sources. Longuet-Higgins [?] demonstrated that this distribution is also applicable to the heights of waves in the sea, and since that time the Rayleigh distribution has been universally employed to describe wave heights.

Illustrative Example 1.48 (Overtopping by random sea waves). Consider a seawall built along an existing beach in order to protect the land area from high waves during storms. The main body of the seawall must be strong enough to withstand the wave attack of storm waves, and the crest must be high enough to prevent intrusion of sea water onto land by overtopping. Overtopping occurs when the incident waves are higher than $H = 7\text{ m}$. Taking into account that the relative wave height $h = H/H_s$, where H is the wave height and H_s is the significant wave height, that is the wave height of which is equal to the mean of heights of the highest one-third waves in a wave group, representative of a particular sea state, is a Rayleigh random variable with parameter $\delta = 1/2$. What is the probability that one wave overtops the seawall considering a sea state of $H_s = 5\text{ m}$ within a storm?

The probability that a single wave overtops the structure is equal to the probability of h being greater than $7/5$, that is,

$$\begin{aligned}\Pr(h > 7/5) &= 1 - \Pr(h \leq 7/5) \\ &= 1 - \left(1 - \exp\left[-\frac{1}{2}\left(\frac{7/5}{1/2}\right)^2\right]\right) = 0.0198411.\end{aligned}$$

■

1.7.10 Student's t Distribution

Let Z be the normal $N(\lambda, 1)$ and Y be a χ_n^2 . If Z and Y are independent, then, the random variable

$$X = \frac{Z}{\sqrt{Y/n}}$$

is called the *noncentral Student's t* random variable with n degrees of freedom and non-centrality parameter λ and is denoted by $t_n(\lambda)$. When $\lambda = 0$, we obtain the *central Student's t* random variable, which is denoted by t_n and its pdf is

$$f(x) = \frac{\Gamma(n+1)/2}{\Gamma(n/2)\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{(n+1)/2}; \quad -\infty < x < \infty, \quad (1.125)$$

where $\Gamma(\cdot)$ is the Gamma function defined in (1.89). A family of t -distributions with various values of n degrees of freedom is shown in Figure 1.25. It may be observed that it has a bell-shape density function similar to the normal curve and it is symmetrical with respect to the origin. The t -distribution is flatter than the standard normal distribution but it tends toward the standard normal distribution as n increases. The cdf $F(x)$ is not simple. However, it has been computed numerically and is given in the Appendix as Table ???. The mean and variance of the central t random variable are $\mu = 0$, for $n > 1$, and $\sigma^2 = n/(n-2)$, for $n > 2$, respectively. Whereas the skewness and kurtosis coefficients are $\gamma_1 = 0$ (symmetrical distribution) and $\gamma_2 = 3 + 6/(n-4)$, for $n > 4$, respectively.

Note that the t -distribution is useful for determining the confidence interval of the population mean with unknown variance.

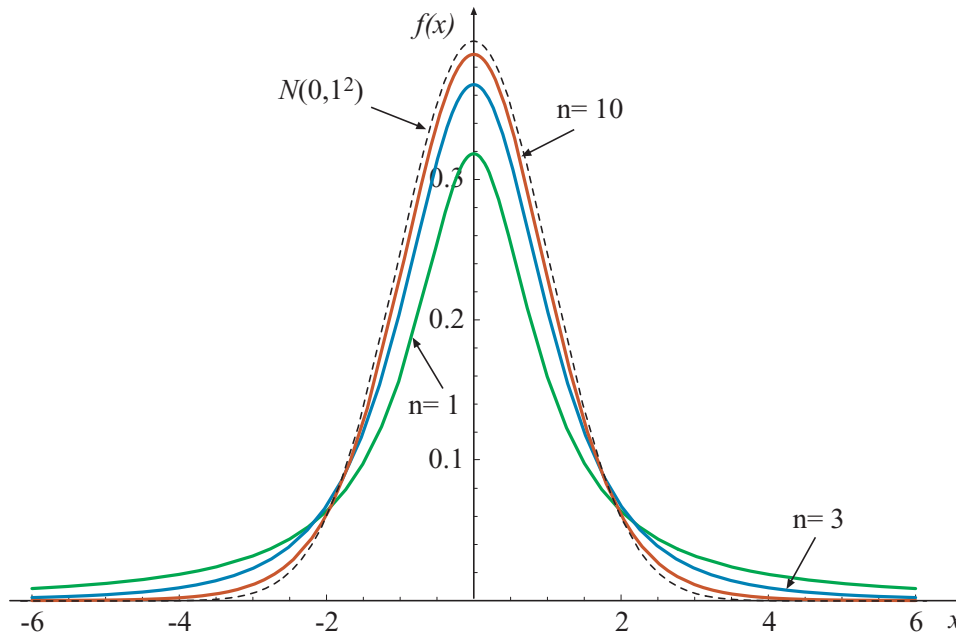


Figure 1.25: Noncentral Student's t probability density functions.

1.7.11 Extreme Value Distributions

Often, when natural calamities of great magnitude happen, we are left wondering about their occurrence and frequency, and whether anything could have been done at all either to avert them or at least to have been better prepared for them. These could include devastating earthquakes, massive snowfalls, etc. It is evident that the characteristics of interest in all these cases are *extremes* in that they correspond to either minimum (e.g. minimum amount of precipitation) or maximum (e.g. maximum amount of water flows) values.

In many engineering applications, we are primarily concerned with extremes. These include: maximum wind speed during storms (used to categorize the storm), minimum stress at which components break, minimum weight at which a structure develops a crack, minimum strength of materials, maximum wave height on the shoreline, etc.

Since the primary issues of interest in all the above examples are concerning the occurrence of such events and their frequency, a careful statistical analysis would require the availability of data on such extremes (preferably of a large size, for making predictions accurately) and an appropriate statistical model for those extremes (which would lead to correct predictions). The detailed mathematical aspects of extreme distributions can be found elsewhere (Gumbel [?], Castillo [28] and Castillo et al. [?]).

Order Statistics and Extremes

Let X_1, X_2, \dots, X_n be a sample of size n drawn from a common pdf $f(x)$ and cdf $F(x)$. Arrange (X_1, X_2, \dots, X_n) in an increasing order of magnitude and let $X_{1:n} \leq \dots \leq X_{n:n}$ be the ordered values. The r -th element of this sequence, $X_{r:n}$, is called the r -th *order statistic*

in the sample. The first and the last order statistics, $X_{1:n}$ and $X_{n:n}$ are the *minimum* and the *maximum* of (X_1, X_2, \dots, X_n) , respectively. The minimum and the maximum order statistics are called the *extremes*.

Order statistics are very important in practice and specially so for the minimum, $X_{1:n}$, and the maximum, $X_{n:n}$, because they are the critical values used in engineering, physics, medicine, etc (see for example, Castillo and Hadi [?, ?, ?]).

Distribution of Extremes

Let X be a random variable with pdf $f(x)$ and cdf $F(x)$. Let (x_1, x_2, \dots, x_n) be a iid sample drawn from $F(x)$. Then considering independence, their joint pdf is

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i), \quad (1.126)$$

and their cdf is

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i). \quad (1.127)$$

The cdf of the maximum order statistic, $X_{n:n}$, is

$$\begin{aligned} F_{max}(x) &= \Pr(X_{n:n} \leq x) \\ &= \Pr(\max(X_1, X_2, \dots, X_n) \leq x) \\ &= \Pr(\text{all } X_i \leq x) \end{aligned} \quad (1.128)$$

$$\begin{aligned} &= \prod_{i=1}^n \Pr(X_i \leq x) \\ &= \prod_{i=1}^n F(x) = [F(x)]^n. \end{aligned} \quad (1.129)$$

Thus, the cdf of the maximum can be obtained by raising the original cdf to the n -th power. The pdf of $X_{n:n}$ is obtained by differentiating (1.129) with respect to x , that is,

$$f_{max}(x) = n f(x) [F(x)]^{n-1}. \quad (1.130)$$

Figure 1.26 shows the pdf and cdf of the maximum order statistics, $X_{n:n}$, in samples of sizes 1, 2, 10, 100, 1000 and 10000 drawn from the Standard Normal distribution $N(0, 1^2)$. Note that the curve for $n = 1$ is of the parents pdf $f(x)$ and cdf $F(x)$. Observe that the pdf and cdf of the maximum order statistics change location and scale with increasing n . In particular, the pdf moves to the right with decreasing scale as n increases, and the cdf moves to the right and with increasing slope as n increases. As can be inferred from the graphs, at $n = \infty$, $X_{n:n}$ becomes a degenerate random variable. This fact can be obtained directly from (1.128).

The cdf of the minimum order statistic, $X_{1:n}$, can be obtained in a similar way, as follows:

$$F_{min}(x) = \Pr(X_{1:n} \leq x)$$

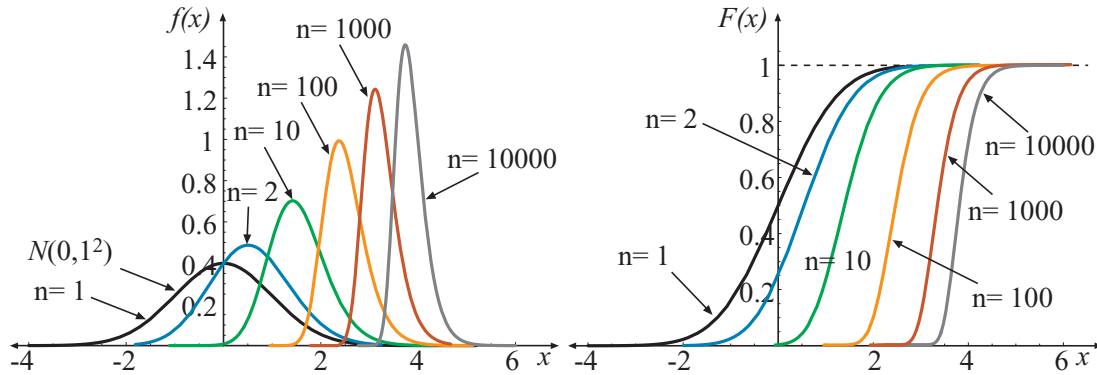


Figure 1.26: The pdf and cdf of the maxima of a sample of sizes 1, 2, 10, 100, 1000 and 10000 drawn from the Standard Normal $N(0, 1^2)$.

$$\begin{aligned}
 &= \Pr(\min(X_1, X_2, \dots, X_n) \leq x) \\
 &= 1 - \Pr(\min(X_1, X_2, \dots, X_n) > x) \\
 &= 1 - \Pr(\text{all } X_i > x) = 1 - \prod_{i=1}^n \Pr(X_i > x) \quad (1.131)
 \end{aligned}$$

$$= 1 - \prod_{i=1}^n (1 - F(x)) = 1 - [1 - F(x)]^n \quad (1.132)$$

$$= 1 - [S(x)]^n, \quad (1.133)$$

where $S(x)$ is the survival function defined in (1.46). Thus, the cdf of the minimum order statistics is one minus the survival function raised to the power n .

This implies that

$$S_{min} = 1 - F_{min}(x) = S^n(x)$$

or, in other words, that the survival function of the minimum is the survival function of the parent raised to the n -th power.

The pdf is obtained by differentiating (1.132) with respect to x as

$$f_{min}(x) = nf(x)[1 - F(x)]^{n-1}. \quad (1.134)$$

Figure 1.27 shows the pdf and cdf of the minimum order statistics, $X_{1:n}$, in sample sizes 1, 2, 10, 100, 1000 and 10000 drawn from the Standard Normal distribution $N(0, 1^2)$. Note that the curve for $n = 1$ is of the parent's pdf $f(x)$ and cdf $F(x)$. Compare the pdf and cdf in Figure 1.27 to those in Figure 1.26 and note the similarities and differences. Note also that all the graphs in Figure 1.27 can be obtained from those in Figure 1.26 after one or two symmetric transformations. The pdf and cdf of the minimum order statistics also change location and scale with increasing n , but they move in the opposite direction. That is, the pdf moves to the left with decreasing scale as n increases, and the cdf moves to the left and with increasing slope as n increases. It can also be inferred from the graphs, at $n = \infty$, $X_{1:n}$ becomes a degenerate random variable. This can also be obtained directly from (1.131).

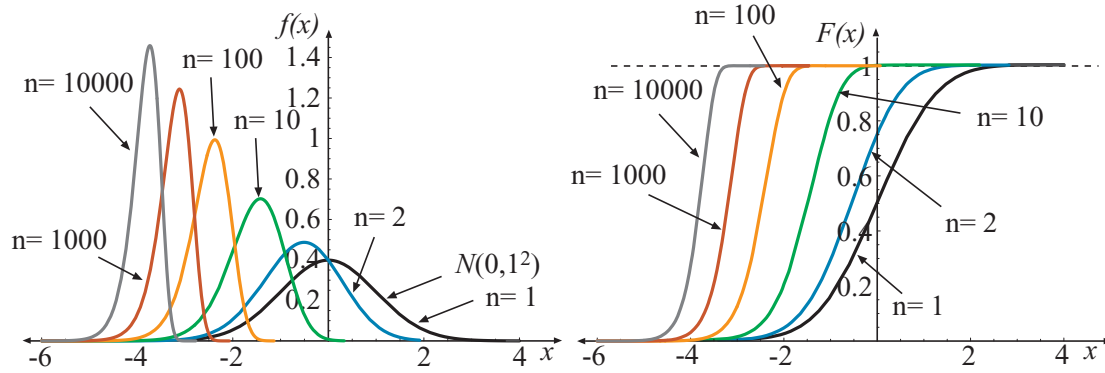


Figure 1.27: The pdf and cdf of the minimum of a sample of sizes 1, 2, 10, 100, 1000 and 10000 drawn from the Standard Normal $N(0, 1^2)$.

Asymptotic Distributions

In many engineering applications the expressions derived above are not very useful. This happens, for example, in the following cases:

- When the sample size is very large or goes to infinity.
- When the cdf, $F(x)$, of the parent population is unknown.
- When the sample size is unknown.

So the study of limiting distributions of order statistics, that is, when $n \rightarrow \infty$ is very useful and have a wide range of applications in engineering problems.

Among other statistics, the minimum and the maximum are the most relevant to engineering applications.

Extreme value distributions are treated no differently than any other distributions discussed earlier. And can be uniquely defined in terms of its pdf or cdf and the parameters of the distribution. Once an extreme value distribution is uniquely defined, probabilistic information can be extracted from it.

1.7.12 Gumbel or Type I Extreme Value Distribution

The Type I asymptotic distribution, also referred as the Gumbel, ‘double exponential’ and ‘Fisher-Tippett Type I’ appears very frequently in practical problems when we observe data that represent maxima values. It can be shown that for many parent populations with limited or unlimited left tail the limit of independent samples converges to a Gumbel distribution. The pdf of the Gumbel random variable is given by

$$f(x) = \frac{1}{\delta} \exp \left[\frac{\lambda - x}{\delta} - \exp \left(\frac{\lambda - x}{\delta} \right) \right]; \quad -\infty < x < \infty \quad (1.135)$$

and the cdf by

$$F(x) = \exp \left[-\exp \left(\frac{\lambda - x}{\delta} \right) \right]; \quad -\infty < x < \infty \quad (1.136)$$

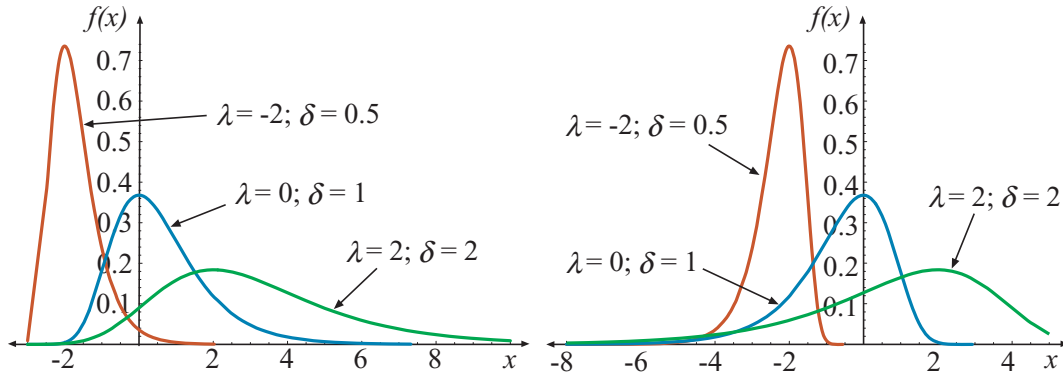


Figure 1.28: Pdfs of the Gumbel distribution for different values of the parameters (left). Reverse Gumbel distribution for the same parameters (right).

where the parameters λ and δ are the mode or characteristic value of the initial distribution, and the shape parameter that is a measure of the dispersion of the distribution, respectively. δ is sometimes known as the ‘slope’ of the distribution (obtained when plotting the distribution on the so-called ‘Gumbel’ paper). The first three moments, mean, variance and skewness are:

$$\mu = \lambda + \gamma\delta, \quad \sigma^2 = \frac{\pi^2\delta^2}{6} \quad \text{and} \quad \gamma_1 = \frac{12\sqrt{6}\zeta(3)}{\pi^3} \approx 1.13955, \quad (1.137)$$

where $\gamma = 0.577216$ is Euler’s constant and ζ is the Riemann zeta function. Note that the skewness is seen to be independent of λ and δ .

Also of interest is the *reversed* Gumbel distribution, used for the smallest value with pdf

$$f(x) = \frac{1}{\delta} \exp \left[\frac{x - \lambda}{\delta} - \exp \left(\frac{x - \lambda}{\delta} \right) \right]; \quad -\infty < x < \infty \quad (1.138)$$

and cdf

$$F(x) = 1 - \exp \left[-\exp \left(\frac{x - \lambda}{\delta} \right) \right]; \quad -\infty < x < \infty \quad (1.139)$$

with moments

$$\mu = \lambda - \gamma\delta, \quad \sigma^2 = \frac{\pi^2\delta^2}{6} \quad \text{and} \quad \gamma_1 = -\frac{12\sqrt{6}\zeta(3)}{\pi^3} \approx -1.13955. \quad (1.140)$$

In Figure 1.28 the pdfs of the three different Gumbel (left hand) and *reversed* Gumbel (right hand) distributions are shown. The extreme value distribution for the minimum value has less practical applications than that for the maximum value.

In general, the Type I asymptotic form is obtained by the convergence of distributions with an exponential tail. The normal, gamma and exponential distributions are of this type.

1.7.13 Fréchet or Type II Extreme Value Distribution

The Type II asymptotic distribution, also referred as the Fréchet distribution appears very frequently in practical problems when we observe data that represent maxima values. Many parent populations with unlimited left tail the limit of independent samples converges to a Fréchet distribution. The Type II asymptotic form is obtained as n goes to infinity from an initial distribution that has a polynomial tail in the direction of the extreme value.

The pdf of the Fréchet random variable is given by

$$f(x) = \frac{\beta\delta}{x^2} \exp \left[- \left(\frac{\delta}{x-\lambda} \right)^\beta \right] \left(\frac{\delta}{x-\lambda} \right)^{\beta-1}; \quad x > \lambda \quad (1.141)$$

and the cdf by

$$F(x) = \exp \left[- \left(\frac{\delta}{x-\lambda} \right)^\beta \right]; \quad x > \lambda \quad (1.142)$$

with mean and variance

$$\mu = \lambda + \delta\Gamma \left(1 - \frac{1}{\beta} \right); \quad \text{for } \beta > 1 \quad (1.143)$$

and

$$\sigma^2 = \delta^2 \left[\Gamma \left(1 - \frac{2}{\beta} \right) - \Gamma^2 \left(1 - \frac{1}{\beta} \right) \right]; \quad \text{for } \beta > 2. \quad (1.144)$$

Also of interest is the it reversed Fréchet distribution with pdf

$$f(x) = \frac{\beta\delta}{x^2} \exp \left[- \left(\frac{\delta}{\lambda-x} \right)^\beta \right] \left(\frac{\delta}{\lambda-x} \right)^{\beta-1}; \quad x < \lambda \quad (1.145)$$

and cdf

$$F(x) = 1 - \exp \left[- \left(\frac{\delta}{\lambda-x} \right)^\beta \right]; \quad x < \lambda \quad (1.146)$$

with mean and variance

$$\mu = \lambda + \delta\Gamma \left(1 - \frac{1}{\beta} \right); \quad \text{for } \beta > 1 \quad (1.147)$$

and

$$\sigma^2 = \delta^2 \left[\Gamma \left(1 - \frac{2}{\beta} \right) - \Gamma^2 \left(1 - \frac{1}{\beta} \right) \right]; \quad \text{for } \beta > 2. \quad (1.148)$$

1.7.14 Weibull or Type III Extreme Value Distribution

The Type III asymptotic distribution, also referred as Weibull distribution appears very frequently in practical problems when we observe data that represent minima values. The reason why this occurs will be justified in Chapter ??, where it is shown that for many parent populations with limited left tail the limit of independent samples converges to a Weibull distribution. The pdf of the Weibull random variable is given by

$$f(x) = \frac{\beta}{\delta} \exp \left[- \left(\frac{x-\lambda}{\delta} \right)^\beta \right] \left(\frac{x-\lambda}{\delta} \right)^{\beta-1}; \quad x > \lambda, \quad (1.149)$$

and the cdf by

$$F(x) = 1 - \exp \left[- \left(\frac{x - \lambda}{\delta} \right)^\beta \right]; \quad x \geq \lambda, \quad (1.150)$$

with mean and variance

$$\mu = \lambda + \delta \Gamma \left(1 + \frac{1}{\beta} \right) \quad \text{and} \quad \sigma^2 = \delta^2 \left[\Gamma \left(1 + \frac{2}{\beta} \right) - \Gamma^2 \left(1 + \frac{1}{\beta} \right) \right]. \quad (1.151)$$

Also of interest is the *reversed* Weibull distribution with pdf

$$f(x) = \frac{\beta}{\delta} \exp \left[- \left(\frac{\lambda - x}{\delta} \right)^\beta \right] \left(\frac{\lambda - x}{\delta} \right)^{\beta-1}; \quad x < \lambda, \quad (1.152)$$

and cdf

$$F(x) = \exp \left[- \left(\frac{\lambda - x}{\delta} \right)^\beta \right]; \quad x \leq \lambda, \quad (1.153)$$

with mean and variance

$$\mu = \lambda - \delta \Gamma \left(1 + \frac{1}{\beta} \right) \quad \text{and} \quad \sigma^2 = \delta^2 \left[\Gamma \left(1 + \frac{2}{\beta} \right) - \Gamma^2 \left(1 + \frac{1}{\beta} \right) \right]. \quad (1.154)$$

1.7.15 Generalized Extreme Value Distributions

The Generalized Extreme Value Distributions include all distributions that can be obtained as the limit of sequences of maxima and minima values (see Chapter ??). The cdf of the maximal generalized extreme value distribution (GEV) is given by

$$H(x; \lambda, \delta, \kappa) = \begin{cases} \exp \left\{ - \left[1 - \kappa \left(\frac{x - \lambda}{\delta} \right) \right]^{\frac{1}{\kappa}} \right\}; & 1 - \kappa \left(\frac{x - \lambda}{\delta} \right) \geq 0, \kappa \neq 0, \\ \exp \left\{ - \exp \left(\frac{\lambda - x}{\delta} \right) \right\}; & -\infty < x < \infty, \kappa = 0, \end{cases} \quad (1.155)$$

where the support is $x \leq \lambda + \delta/\kappa$, if $\kappa > 0$, or $x \geq \lambda + \delta/\kappa$, if $\kappa < 0$. The corresponding p -quantiles are

$$x_p = \begin{cases} \lambda + \delta [1 - (-\log p)^\kappa] / \kappa, & \text{if } \kappa \neq 0, \\ \lambda - \delta \log(-\log p), & \text{if } \kappa = 0. \end{cases} \quad (1.156)$$

The Gumbel, reversed Weibull and Fréchet distributions are particular cases of the maximal GEV.

Also of interest is the Minimal GEV distribution with cdf, $H(x; \lambda, \delta, \kappa)$, which is given by

$$\begin{cases} 1 - \exp \left\{ - \left[1 + \kappa \left(\frac{x - \lambda}{\delta} \right) \right]^{1/\kappa} \right\}; & 1 + \kappa \left(\frac{x - \lambda}{\delta} \right) \geq 0, \kappa \neq 0, \\ 1 - \exp \left[- \exp \left(\frac{x - \lambda}{\delta} \right) \right]; & -\infty < x < \infty. \end{cases} \quad (1.157)$$

where the support is $x \geq \lambda - \delta/\kappa$, if $\kappa > 0$, or $x \leq \lambda - \delta/\kappa$, if $\kappa < 0$.

The corresponding p -quantiles are

$$x_p = \begin{cases} \lambda - \delta [1 - (-\log(1-p))^\kappa] / \kappa, & \text{if } \kappa \neq 0, \\ \lambda + \delta \log(-\log(1-p)), & \text{if } \kappa = 0. \end{cases} \quad (1.158)$$

The reversed Gumbel, Weibull and reversed Fréchet distributions are particular cases of the minimal GEV.

We conclude the discussion of univariate continuous random variable with a summary of them, which is given in Tables 1.7 and 1.8.

1.7.16 Generalized Pareto Distributions

The generalized Pareto distribution arises when you consider excesses of a random variable above or below given thresholds (see Chapter ??). The cdf of the generalized Pareto distribution is given by

$$F(x; \lambda, \kappa) = \begin{cases} 1 - \left(1 - \frac{\kappa x}{\lambda}\right)^{1/\kappa}; & \left(1 - \frac{\kappa x}{\lambda}\right) \geq 0, \\ & \kappa \neq 0, \lambda > 0, \\ 1 - e^{-x/\lambda}; & x \geq 0, \kappa = 0, \lambda > 0, \end{cases} \quad (1.159)$$

where λ and κ are scale and shape parameters, respectively. For $\kappa \neq 0$, the range of x is $0 \leq x \leq \lambda/\kappa$, if $\kappa > 0$, and $x \geq 0$, if $\kappa \leq 0$.

Also of interest is the reversed generalized Pareto distribution with cdf

$$F(x; \lambda, \kappa) = \begin{cases} \left(1 + \frac{\kappa x}{\lambda}\right)^{1/\kappa}; & \left(1 + \frac{\kappa x}{\lambda}\right) \geq 0, x \leq 0, \\ & \kappa \neq 0, \lambda > 0, \\ e^{x/\lambda}; & x \leq 0, \kappa = 0, \lambda > 0, \end{cases} \quad (1.160)$$

where λ and κ are scale and shape parameters, respectively. For $\kappa \neq 0$, the range of x is $-\lambda/\kappa \leq x \leq 0$, if $\kappa > 0$, and $x < 0$, if $\kappa \leq 0$.

1.8 Truncated Distributions

In this section we introduce truncated distributions that have a lot of interest when dealing with extremes because then, only values above or below certain threshold values are of interest.

Definition 1.7 (Truncated Distributions). *Let X be a random variable. We call the random variables $X|X \leq x_0$, $X|X > x_0$ and $X|x_0 \leq X < x_1$ truncated at x_0 from the right, from the left or truncated at x_0 from the right and at x_1 from the left, respectively. ■*

The following theorem gives the corresponding cdfs as a function of the cdf $F_X(x)$ of X .

Theorem 1.1 ((Cdf of a truncated distribution)). *The cdf of the truncated random variable $X|X \leq x_0$ is:*

$$F_{X|X \leq x_0}(x) = \begin{cases} F_X(x)/F_X(x_0), & \text{if } x < x_0, \\ 1, & \text{if } x \geq x_0. \end{cases} \quad (1.161)$$

The cdf of the truncated random variable $X|X > x_0$ is:

$$F_{X|X > x_0}(x) = \begin{cases} 0, & \text{if } x < x_0, \\ \frac{F_X(x) - F_X(x_0)}{1 - F_X(x_0)}, & \text{if } x \geq x_0. \end{cases} \quad (1.162)$$

and finally, the cdf of the truncated random variable $X|x_0 \leq X < x_1$ is:

$$F_{X|x_0 < X \leq x_1}(x) = \begin{cases} 0, & \text{if } x < x_0, \\ \frac{F_X(x) - F_X(x_0)}{F_X(x_1) - F_X(x_0)}, & \text{if } x_0 \leq x < x_1, \\ 1, & \text{if } x \geq x_1. \end{cases}$$

Proof. Because of the definition of conditional probability we have:

$$\begin{aligned} F_{X|X \leq x_0}(x) &= \Pr(X \leq x | X \leq x_0) \\ &= \frac{\Pr((X \leq x) \cap (X \leq x_0))}{\Pr(X \leq x_0)} \\ &= \frac{\Pr((X \leq x) \cap (X \leq x_0))}{F_X(x_0)} \\ &= \begin{cases} \frac{\Pr(X \leq x)}{F_X(x_0)} = \frac{F_X(x)}{F_X(x_0)}, & \text{if } x < x_0, \\ \frac{\Pr(X \leq x_0)}{F_X(x_0)} = 1, & \text{if } x \geq x_0. \end{cases} \end{aligned}$$

For the left truncation we have:

$$\begin{aligned} F_{X|X > x_0}(x) &= \Pr(X \leq x | X > x_0) \\ &= \frac{\Pr((X \leq x) \cap (X > x_0))}{\Pr(X > x_0)} \\ &= \frac{\Pr((X \leq x) \cap (X > x_0))}{1 - F_X(x_0)} \\ &= \begin{cases} 0, & \text{if } x < x_0, \\ \frac{F_X(x) - F_X(x_0)}{1 - F_X(x_0)}, & \text{if } x \geq x_0. \end{cases} \end{aligned}$$

and finally, for both sides truncation:

$$\begin{aligned}
 F_{X|x_0 < X \leq x_1}(x) &= \Pr(X \leq x | x_0 < X \leq x_1) \\
 &= \frac{\Pr((X \leq x) \cap (x_0 < X \leq x_1))}{\Pr(x_0 < X \leq x_1)} \\
 &= \frac{\Pr((X \leq x) \cap (x_0 < X \leq x_1))}{F_X(x_1) - F_X(x_0)} \\
 &= \begin{cases} 0, & \text{if } x < x_0, \\ \frac{F_X(x) - F_X(x_0)}{F_X(x_1) - F_X(x_0)}, & \text{if } x_0 \leq x < x_1, \\ 1, & \text{if } x \geq x_1. \end{cases}
 \end{aligned}$$

■

Illustrative Example 1.49 (Lifetime). The remaining lifetime, X , in years of a patient after suffering a heart attack has a cdf

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{1 - e^{-\lambda x}}{1 - e^{-50\lambda}}, & \text{if } 0 \leq x < 50, \\ 1, & \text{if } x \geq 50. \end{cases}$$

If a given patient suffered a heart attack 30 years ago, determine the cdf of the remaining lifetime.

Before solving this problem, it is worthwhile mentioning that the given variable is a right truncated exponential distribution, which implies that no patient can survive above 50 years after suffering a heart attack.

Since the patient suffered a heart attack 30 years ago, we must determine the cdf of the random variable $(X - 30)$ conditioned by $X \geq 30$. Then, we ask for the truncated distribution on the left at 30 and translated 30 units, that is,

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ \frac{e^{-30\lambda} - e^{-\lambda(x+30)}}{e^{-30\lambda} - e^{-50\lambda}}, & \text{if } 0 \leq x < 20, \\ 1, & \text{if } x \geq 20. \end{cases}$$

■

Illustrative Example 1.50 (Hospital). The age, X , in years, of the persons entering a hospital have pdf

$$f_X(x) = \begin{cases} \frac{\pi}{200} \sin\left(\frac{\pi x}{100}\right), & \text{if } 0 \leq x < 100, \\ 0, & \text{otherwise.} \end{cases}$$

Obtain:

1. Pdf of the age, X , in years of the children younger than 5 years old that enter the hospital

2. Do the same as above for the persons above 60 years of age.
1. It is the same density but truncated at $X = 5$ on the right and then we have

$$f_X(x) = \begin{cases} \frac{\pi \sin(\pi x/100)}{100 [1 - \cos(5\pi/100)]}, & \text{if } x < 5, \\ 0, & \text{otherwise.} \end{cases}$$

2. Similarly, it is the same density but truncated at $X = 60$ on the left and then we have

$$f_Z(z) = \begin{cases} \frac{\pi \sin(\pi z/100)}{100 [1 + \cos(60\pi/100)]}, & \text{if } 60 \leq z < 100 \\ 0, & \text{otherwise.} \end{cases}$$

■

Illustrative Example 1.51 (Screw strength). A factory producing screws states that its strength, R^* , in Kg/cm^2 has an exponential distribution $E(\lambda)$. If all the screws are subject to a quality test consisting of applying a test stress of $10 Kg/cm^2$ and those failing are discarded, determine the pdf of the strength, R , of the accepted screws.

Since after the test the screws with a strength less than $10 Kg/cm^2$, are discarded, the resulting strength is truncated on the left at $10 Kg/cm^2$, so that we have:

$$f_R(x) = \begin{cases} \frac{\lambda e^{-\lambda x}}{e^{-10\lambda}} = \lambda e^{-\lambda(x-10)}, & \text{if } y > 10 Kg/cm^2, \\ 0, & \text{otherwise.} \end{cases}$$

■

1.9 Exercises**

- 1.1 Show that the first central moment of any discrete random variable is zero.
- 1.2 Show that the mean and variance of a Bernoulli random variable, with success probability p , are $\mu = p$ and $\sigma^2 = p(1 - p)$.
- 1.3 Show that the mean and variance of a $B(n, p)$ random variable are $\mu = np$ and $\sigma^2 = np(1 - p)$.
- 1.4 For the $B(6, 0.2)$ random variable in Exercise 1.10, whose pmf is given in (??), compute the probability of each of the following events:
 - (a) At least one structure is in poor condition.
 - (b) At most four structures are in poor conditions.
 - (c) Between 2 and 4 structures are in poor conditions.
- 1.5 Let $X = X_1, X_2, \dots, X_n$, where $X_i; i = 1, 2, \dots, n$, are identical and independent Bernoulli random variables with probability of success equal to p . Use the reproductive property of the binomial random variable to show that $X \sim B(n, p)$

- 1.6 Show that the pmf of any $G(p)$ random variable is decreasing in x .
- 1.7 Suppose that the company in Example ?? wishes to fill two vacant positions.
- What is the probability that at least four interviews will be required to fill the two vacant positions?
 - If an interview costs the company \$500 on the average, what is the expected costs of filling the two vacant positions?
- 1.8 Prove the result in Equation (1.75).
- 1.9 Use (??) to show that the pmf of the non-zero Poisson random variable is given by (1.79).
- 1.10 The probability of a river to be flooded at a certain location during a period of one year is 0.02.
- Find the pmf of the number of floods to be registered during a period of 20 years.
 - Find the pmf of the number of years required until the first with a flood.
 - Find the pmf of the number of years required until the fifth with a flood.
- 1.11 Assume that the occurrence of earthquakes with intensity above a given threshold is Poissonian with rate 2 earthquakes/year. Compute:
- The probability of having no earthquakes in a period of 6 months (the time a dam is being repaired).
 - The pmf of the number of earthquakes occurring in a period of 5 years.
 - Discuss the validity of the assumptions for using the Poisson model.
- 1.12 It is said that a system (for example a dam or a dike) has been designed for the N -year if it withstand floods that occur once in N years, i.e., its probability of occurrence is $1/N$ in any year. Assuming that floods in different years are independent, calculate.
- The probability of having a flood larger or equal to the N -year flood during a period of 50 years.
 - The probability of having one or more such a floods in 50 years.
 - If a company designs 20 independent systems (located far enough) for the 500-year flood, what is the cdf of the number of systems that will fail in 50 years?
- 1.13 Compute the mean vector, covariance matrix, and correlation matrix for each of the following multinomial random variables:
- The three variables in Example ??.
 - The four variables in Example ??.
- 1.14 In a $M(n; p_1, p_2, \dots, p_k)$ random variables, show that the correlation coefficient between X_i and X_j is

$$\rho_{ij} = -\sqrt{p_i p_j / [(1 - p_i)(1 - p_j)]}.$$

Table 1.7: The Probability Density Functions of Some Continuous Random Variables that Frequently Arise in Engineering Applications.

Distribution	$f(x)$
Uniform	$(\beta - \lambda)^{-1}$
Exponential	$\lambda e^{-\lambda x}$
Gamma	$\frac{\lambda^\theta x^{\theta-1} e^{-\lambda x}}{\Gamma(\theta)!}$
Beta	$\frac{x^{\lambda-1} (1-x)^{\theta-1}}{\beta(\lambda, \theta)}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right]$
Log-normal	$\frac{1}{x\sigma\sqrt{2\pi}} \exp \left[-\frac{(\log(x)-\mu)^2}{2\sigma^2} \right]$
Logistic	$\frac{\exp \{-(x-\alpha)/\beta\}}{\beta [1 + \exp \{-(x-\alpha)/\beta\}]^2}$
Central Chi-Square	$\frac{x^{(n/2)-1} e^{-x/2}}{2^{n/2} \Gamma(n/2)}$
Rayleigh	$\delta^{-2} x \exp \left[-\frac{1}{2} \left(\frac{x}{\delta} \right)^2 \right]$
Student's t	$\frac{\Gamma(n+1)/2}{\Gamma(n/2)\sqrt{n\pi}} \left(1 + \frac{x^2}{n} \right)^{-(n+1)/2}$
Weibull	$\frac{\beta}{\delta} \exp - \left(\frac{x-\lambda}{\delta} \right)^\beta \left(\frac{x-\lambda}{\delta} \right)^{\beta-1}$
Reversed Weibull	$\frac{\beta}{\delta} \exp - \left(\frac{\lambda-x}{\delta} \right)^\beta \left(\frac{\lambda-x}{\delta} \right)^{\beta-1}$
Gumbel	$\frac{1}{\delta} \exp \left[\frac{\lambda-x}{\delta} - \exp \left(\frac{\lambda-x}{\delta} \right) \right]$
Reversed Gumbel	$\frac{1}{\delta} \exp \left[\frac{x-\lambda}{\delta} - \exp \left(\frac{x-\lambda}{\delta} \right) \right]$
Fréchet	$\frac{\beta\delta}{x^2} \exp - \left(\frac{\delta}{x-\lambda} \right)^\beta \left(\frac{\delta}{x-\lambda} \right)^{\beta-1}$
Reversed Fréchet	$\frac{\beta\delta}{x^2} \exp - \left(\frac{\delta}{\lambda-x} \right)^\beta \left(\frac{\delta}{\lambda-x} \right)^{\beta-1}$
Maximal GPD	$\frac{1}{\lambda} \left(1 - \frac{\kappa x}{\lambda} \right)^{\frac{1}{\kappa} - 1}$
Minimal GPD	**

Table 1.8: The Means and Variances of Some Continuous Random Variables that Frequently Arise in Engineering Applications.

Distribution	Mean	Variance
Uniform	$(\lambda + \beta)/2$	$(\beta - \lambda)^2/12$
Exponential	$1/\lambda$	$1/\lambda^2$
Gamma	θ/λ	θ/λ^2
Beta	$\lambda/(\lambda + \theta)$	$\lambda\theta/[(\lambda + \theta + 1)(\lambda + \theta)^2]$
Normal	μ	σ^2
Log-normal	$e^{\mu + \sigma^2/2}$	$e^{2\mu}(e^{2\sigma^2} - e^{\sigma^2})$
Logistic	**	**
Central Chi-Square	n	$2n$
Rayleigh	$\delta\sqrt{\pi/2}$	$\delta^2(4 - \pi)/2$
Student's t	0	$n/(n - 2)$
Weibull	$\lambda + \delta\Gamma\left(1 + \frac{1}{\beta}\right)$	$\delta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right]$
Reversed Weibull	$\lambda - \delta\Gamma\left(1 + \frac{1}{\beta}\right)$	$\delta^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right]$
Gumbel	$\lambda + 0.57772\delta$	$\pi^2\delta^2/6$
Reversed Gumbel	$\lambda - 0.57772\delta$	$\pi^2\delta^2/6$
Fréchet	$\lambda + \delta\Gamma\left(1 - \frac{1}{\beta}\right)$	$\delta^2 \left[\Gamma\left(1 - \frac{2}{\beta}\right) - \Gamma^2\left(1 - \frac{1}{\beta}\right) \right]$
Reversed Fréchet	$\lambda + \delta\Gamma\left(1 - \frac{1}{\beta}\right)$	$\delta^2 \left[\Gamma\left(1 - \frac{2}{\beta}\right) - \Gamma^2\left(1 - \frac{1}{\beta}\right) \right]$
Maximal GPD	**	**
Minimal GPD	**	**

Chapter 2

Measures of Structural Reliability

2.1 Introduction

In general engineering design consists of proportioning the elements of a system so that it satisfies various criteria of performance, safety, serviceability, and durability under various demands. During its lifetime the structure is subject to changes of conditions or *state* of the structure, going from the undamage state to a state of deterioration, or wear, damage in various degrees, excessive deflections or deformations, failure, or collapse. It is assumed that malfunction in the design stage can occur in only a finite number of *modes*, each mode of malfunction may be termed a *limit state*, that gives rise to a design inequality. The ‘violation’ of a limit state, its failure, can be defined as the inability of the structure to perform adequately under the imposed requirements, reaching an undesirable condition for the structure.

Definition 2.1 (Limit state). *Limit state of a structure is the situation once it has been overcome is considered that the structure does not fulfill one of the requirements it was designed for.* ■

The failure of a structure is not easily defined. For example, some structures fail immediately and it is not reversible, such as the buckling of a structural column. In other structures, failure occurs gradually, and the determination of the stage at which the system ceases to perform adequately is highly subjective. Moreover, as the consequences of failure can be very different depending on the type of the structure and the environment, the limit states can be classified in:

1. **Operating suspension.** It consists in the disruption of normal use of the structure, usually produced by atmospheric agents, such as storms, hurricanes, etc. In these situations the safety of using the structure decreases considerably but the structure recovers the normal use once the external agent disappears. This limit state is widely used in harbors, where the ships can not be used during sea storms.
2. **Serviceability.** It consists in the disruption of normal use of the structure, owing to excessive deformations, vibrations, local damage, etc.

3. **Damage.** Situations where important repairs are required to avoid the failure or collapse of the structure, these include, for example, excessive or premature cracking, deformation or permanent inelastic deformation, etc.
4. **Ultimate.** These limit states refer to the collapse of all or part of the structure, such as, tipping or sliding, rupture, progressive collapse, plastic mechanism, instability, corrosion, fatigue, deterioration, fire, etc. Damage limit states often are included in the category.

From observation, it is known that very few structures collapse, or require major repairs, so that the occurrence of violation of the most serious limit states is relatively rare (see Vrouwenvelder [172], CIRIA [?]). But when it occurs the consequences may be extreme, causing *losses* that can be economical, cultural, injuries and even deaths. The relative importance of these components varies greatly with the type of structure and its exposure.

As a whole, the minimum requirements that structures must fulfill during their lifetimes are: mechanical resistance and stability, safety in case of accident and use safety. The acceptable level of reliability must be viewed within the context of possible costs, risks, and associated social benefits. The reliability of a system or structure is the probabilistic assessment of the *likelihood* of its adequate performance for a specified period of time (*lifetime*, or *useful life*) under proposed *operating conditions*. In particular, the study of structural safety and its calculation and prediction is concerned with the violation of the ultimate or safety limit states for the structure.

Definition 2.2 (Lifetime or useful life). *Lifetime or useful life of an structure is the period of time from its start-up in which the structure must to maintain acceptable safety, functionality and aspect conditions. During this time it will require normal maintenance without important rehabilitation or reparation works.* ■

There are numerous sources of uncertainty in the functioning of structural systems, incorporating the information of uncertainty into actual design problems is the subject of this book.

There are several ways to measure the reliability of structural systems, since the classical methods based on safety factors where the random variables involved are considered deterministic an equal to some conservative values about which it is assumed that there is no uncertainty, to the probabilistic approach, where the uncertainty related to the involved variables is taken into account by means of their probability density functions.

A number of terms has been proposed in the literature to designate the ability of the system to perform under the imposed requirements. Among these are resistance, strength, capacity, load, input and demand. In this text we will use resistance (R) and load (S) but the concept could be extended to other disciplines in engineering replacing resistance and load by supply and demand for example.

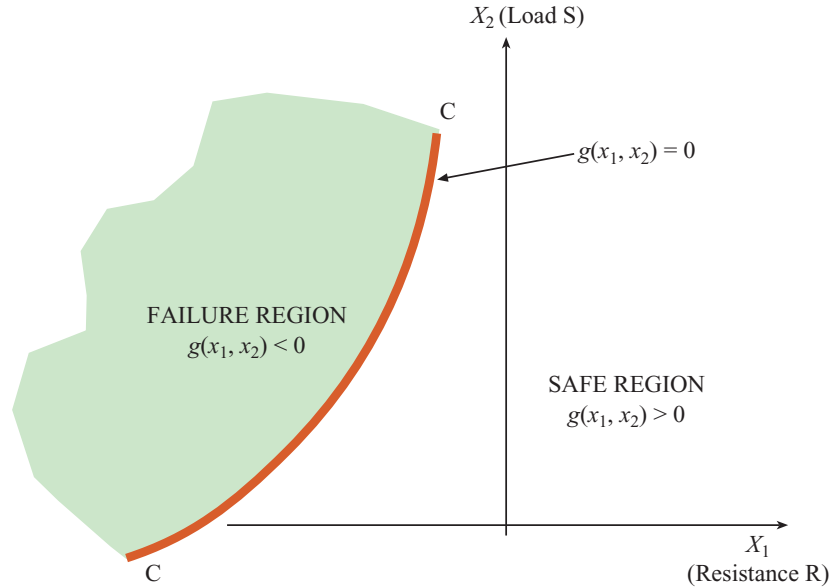


Figure 2.1: Graphical illustration of the safety and failure regions and the limit state equation for the bidimensional case.

2.2 Deterministic Measures

2.2.1 Factor of safety

Traditionally in civil engineering, analysis of the risk of failure are made on the basis of allowable *factors of safety*. Safety factors were establish only by means of engineering judgement as a matter of expediency, not necessity. This may be selected on the basis of experimental observations, previous practical experience, economic and, perhaps, political considerations. Usually, its selection is the responsibility of a code committee.

Consider the set of basic variables (X_1, X_2, \dots, X_n) that represent the design variables or project factors including parameters and uncertain variables (strengths, loads, dimensions, ...), that belong to a n -dimensional space and that are considered random. Deterministic variables are considered particular cases of random variables. The n -dimensional space can be divided in two regions with respect a failure mode (see Figure 2.1): safety region where the project requirements with respect the corresponding failure mode are satisfied, and the failure region, where the structure fails.

$$\begin{aligned} \text{Safe region:} \quad \mathcal{S} &\equiv \{(x_1, x_2, \dots, x_n) | g(x_1, x_2, \dots, x_n) > 0\}, \\ \text{Failure region:} \quad \mathcal{F} &\equiv \{(x_1, x_2, \dots, x_n) | g(x_1, x_2, \dots, x_n) \leq 0\}, \end{aligned} \quad (2.1)$$

where $g(x_1, x_2, \dots, x_n)$ is called *failure function* or *limit state function* that usually can be expressed as a function of the quotient between two non-dimensional quantities,

$$g(x_1, x_2, \dots, x_n) = \frac{h_R(x_1, x_2, \dots, x_n)}{h_S(x_1, x_2, \dots, x_n)} - 1 \quad (2.2)$$

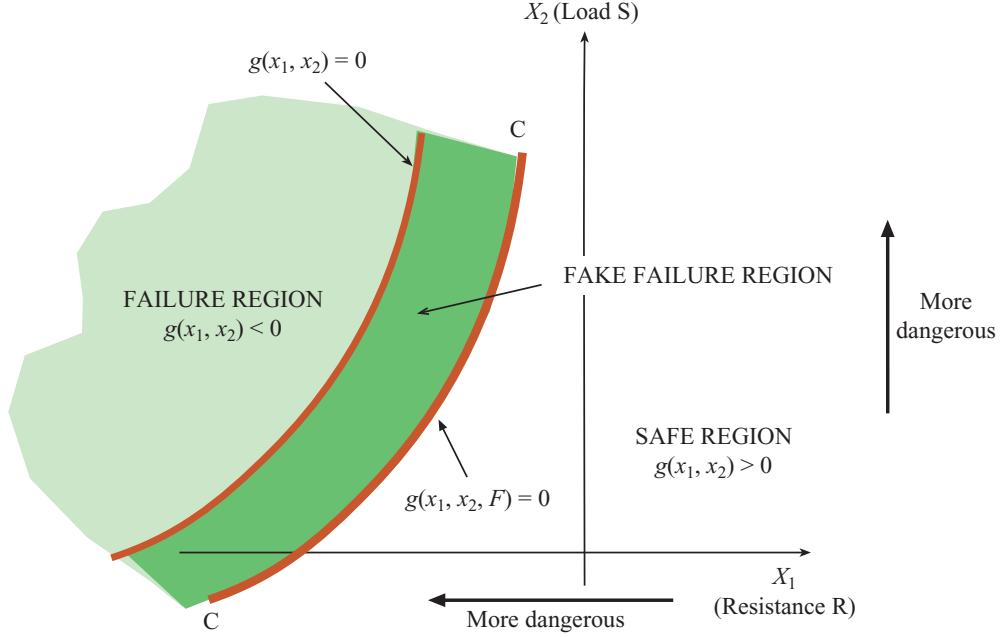


Figure 2.2: Graphical illustration of the failure region increase using the global safety factor.

where $h_R(x_1, x_2, \dots, x_n)$ and $h_S(x_1, x_2, \dots, x_n)$ represent the magnitudes related to resistance or capacity and the loads or stresses, respectively.

The boundary between the safe and the failure region is the limit state. It is often convenient to include the limit states in the failure set.

As equation (2.1) is a limit constraint, using the global safety factor, the variables (X_1, X_2, \dots, X_n) are considered deterministic and equal to their nominal values. If we talk about resistances is usually a conservative value, one, two or three standard deviation below the mean, whereas if we talk about loads it is also a conservative value; however it is several times above the mean value. The extra safety is introduced by means of the safety factor as follows:

$$g(x_1, x_2, \dots, x_n, F) = \frac{h_R(x_1, x_2, \dots, x_n)}{h_S(x_1, x_2, \dots, x_n)} - F \geq 0 \quad (2.3)$$

where F is the safety factor ($F > 1$).

In Figure 2.2 is shown how the failure region is increased artificially using the safety factor F . This means that there is a safety region (dark green in Figure 2.2) where failure is considered so that the design will be farther to the failure region. Note that the new limit state equation $g(x_1, x_2, F) = 0$ move right with respect the real limit state equation $g(x_1, x_2) = 0$.

Combining expressions (2.1) and (2.3), the condition of limit state violation can be rewritten as

$$\frac{h_R(x_1, x_2, \dots, x_n)}{F} \leq h_S(x_1, x_2, \dots, x_n). \quad (2.4)$$

Expression (2.4) is a limit state equation when the inequality sign is replaced by an equality. Therefore, the design will be acceptable from the safety point of view as long as condition (2.4) holds.

Heretofore, the term ‘safety factor’ has been used in relation to overturning, sliding, etc., of structures as a whole, or as in geomechanics. But traditionally the safety factor has been associated with resistance-load (R,S) , or elastic stress. To do this, functions $h_R(x_1, x_2, \dots, x_n)$ and $h_S(x_1, x_2, \dots, x_n)$ have to be interpreted appropriately.

For example, considering elastic stress analysis, expression (2.1) is:

$$\begin{aligned} \text{Safe region:} & \quad \sigma_i(\epsilon) \leq \sigma_{pi}, \\ \text{Failure region:} & \quad \sigma_i(\epsilon) > \sigma_{pi}, \end{aligned} \quad (2.5)$$

that implies no failure occurs as long as the i th applied stress component at the generic point ϵ , $\sigma_i(\epsilon)$, is lower than the permissible stress σ_{pi} . This is usually defined in structural design codes, derived from material strengths σ_{ui} but reduced by the safety factor:

$$\sigma_{pi} = \frac{\sigma_{ui}}{F}. \quad (2.6)$$

Conceptually, in a deterministic design the safety factor can be applied to the resistance, as in (2.6), to the load, or both. The allowable stress design methods use a safety factor to compute the permissible stresses σ_{pi} from the ultimate stresses σ_{ui} , and a successful design ensures that the stresses caused by the nominal values of the loads $\sigma_i(\epsilon)$ do not exceed the permissible stresses:

$$\sigma_{pi} = \frac{\sigma_{ui}}{F} > \sigma_i(\epsilon) \quad \text{or} \quad \frac{\sigma_{pi}}{\sigma_i(\epsilon)} > 1. \quad (2.7)$$

2.2.2 Load factor

In the ultimate strength design method, the loads are multiplied by certain load factors to determine the ultimate loads, and the members are required to resist various design combinations of the ultimate loads.

The ‘load factor’ λ is a special kind of safety factor developed for use originally in the plastic theory of structures. It is defined as the theoretical value by which a set of loads acting on the structure must be multiplied just enough to cause the structure to collapse. Strength is determined from idealized plastic material strength properties (see Heyman [94]) whereas the loads are those acting on the structure during service load conditions.

Thus, for a given failure mode i , the structure is considered to have ‘fail’ if the plastic resistances (R_{pi}) and the factor loads (λQ_j) are related as follows

$$W_R(\mathbf{R}_P) \leq W_Q(\lambda \mathbf{Q}) \quad (2.8)$$

where \mathbf{R}_P is the vector of all plastic resistances (e.g. plastic moments) and \mathbf{Q} is the vector of applied loads. $W_R(\cdot)$ and $W_Q(\cdot)$ are the internal and external works, respectively, associated with the failure mode considered.

If proportional loading is assumed, as is usual, the load factor can be taken out of parentheses. Also the loads usually consist of several components, such as dead, live, wind,

snow, etc. Thus (2.8) can be written in the form of a limit state equation:

$$\frac{W_R(\mathbf{R}_P)}{\lambda W_Q(\mathbf{Q}_D + \mathbf{Q}_L + \dots)} = 1 \quad (2.9)$$

where ‘failure’ occurs if the left-hand side is less or equal than unity.

2.2.3 Partial safety factor

A development of the above two measures of safety derived in the use of multiple safety factors. Instead of using the safety factor for the resistance alone, as in the working stress method, or for the loads alone, as in the ultimate strength method, it is more rational to apply safety factors to both resistance and loads.

Such formats were introduced in Denmark in the 1940s in reinforced concrete and geotechnical standards under the name of *partial safety factors formats*; other institutions as the American Concrete Institute in the 1960s, the National Building Code of Canada since 1977, under different names, such as *ultimate strength design*, *limit state design* (LSD) (NBCC [?]), *load and resistance factor formats*. It has been difficult to find a terminology that is universally suitable. But from the reliability point of view, the most powerful characteristic of this method is the flexibility that the presence of many adjustable factors gives. All these methods are jointly called multiple-factor formats.

The verification of a structure with respect to a limit state uses a mathematical model of the limit state in the form of a failure or limit state function as in (2.2). The relevant parameters X_1, X_2, \dots, X_n are usually divided in three categories: load parameters Q_i , strength parameters R_j , and nominal geometric parameters L_k defining the structure. The limit state equation is as follows:

$$g(Q_i, R_j, L_k) = 0 \quad (2.10)$$

where, conventionally, $g \leq 0$ signifies exceeding the limit state, i. e., failure.

The design values of the basic random variables are denoted by q_{id} , r_{jd} and l_{kd} , respectively. The design values may be written as

$$q_{id} = \gamma_i q_{ic}, \quad r_{jd} = \phi_j r_{jc} \quad \text{and} \quad l_{kd} = \zeta_k l_{kc}, \quad (2.11)$$

where q_{ic} , r_{jc} and l_{kc} are the *characteristic values*, γ_i are *load factors*, ϕ_j *resistance factors*, and ζ_k are *geometrical factors*.

Verification of the structure is done inserting the design and the dimension parameters into the limit state function to ascertain that

$$g(q_{id}, r_{jd}, l_{kd}) = g(\gamma_i q_{ic}, \phi_j r_{jc}, \zeta_k l_{kc}) \geq 0. \quad (2.12)$$

Load and resistance factors can be decomposed further according to basic sources of uncertainty, thus, for example, the resistance factor ϕ_j can take into account in NKB [?] dispersion in material strength, model uncertainty, uncertainty due to use of a vicarious strength parameter, type of consequences of failure and degree of inspection and control other than the statistical quality control of the material. On the other hand, load factors

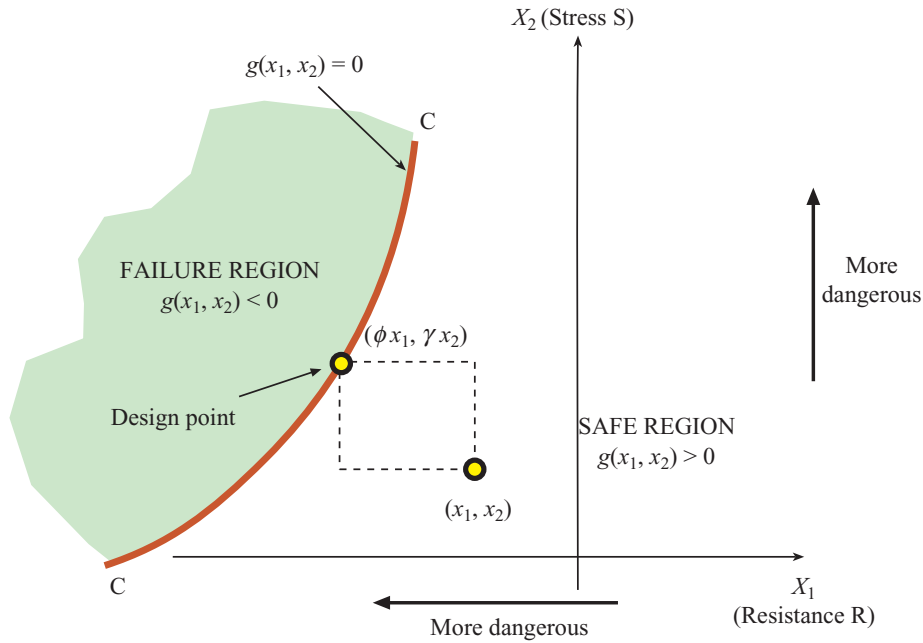


Figure 2.3: Design using partial safety factors.

can consider separately dead loads (D), life loads (L), snow, wind, and earthquake loads (Q). Limit state equation 2.12 can be rewritten as:

$$\phi_j R_j \leq \gamma_{Di} Q_{Di} + \gamma_{Li} Q_{Li} + \dots \quad (2.13)$$

The characteristic values used are often the mean value for dead load, the 98% fractile in the distribution of the annual maxima for movable loads, the 2% or 5% fractile for strength parameters, and the mean value for geometrical parameters. The load factors are usually greater than unity except when the load has an stabilizing effect. The resistance factors are usually less than unity. The geometrical factors are generally equal to unity.

To illustrate we represent in Figure 2.3 the case of two variables X_1 and X_2 , the first one associated with a resistance (R), and the other one related to a stress (S). Obviously, the design becomes more dangerous as they decrease and increase their values, respectively. Then, a code based on partial safety coefficients fixes two values x_1 and x_2 (the characteristic values), and states that such a set of values (x_1, x_2) is safe if and only if the point $(\phi x_1, \gamma x_2)$ is in the safe region, where ϕ and γ are the partial safety coefficients for both variables. The partial safety factors have to fulfill the condition

$$\phi < 1, \quad \gamma > 1,$$

that is, they are decreasing (resistances) and increasing (loads) factors, respectively.

2.2.4 Lack of invariance in some safety measures

One of the problems of the preceding safety measures is that they suffer to a degree from arbitrariness, lack of invariance, or both, under a change of formulation of the failure criteria

consistent with the laws of algebra and mechanics. This lack of uniqueness is due to the ambiguity in the choice of the failure function. This ambiguity was first explained by Ditlevsen [60]. It arises because there are different ways in which the relationships between resistances and loads may be defined. Ideally, the safety measure should not depend on the way in which the loads and resistances are defined. Some examples of this are given below.

Illustrative Example 2.1 (Lack of invariance, adapted from Melchers [123]). Figure 2.10 shows a simply supported beam loaded at the midpoint by a concentrate force W and in the left support by a moment M . Using the ‘factor of safety’, the safety of the structure can be measured in several ways:

1. **Overturning resistance about A**

$$F_1 = \frac{\text{resisting moment about A}}{\text{overturning moment about A}} = \frac{d \times R}{M + Wd/2} = \frac{10 \times 24}{100 + 4 \times 5} = 2.0 \quad (2.14)$$

2. **Capacity of support B**

$$F_2 = \frac{\text{compression resistance of support B}}{\text{compressive load on support B}} = \frac{R}{M/d + W/2} = \frac{24}{10 + 2} = 2.0 \quad (2.15)$$

3. **Net capacity of column B**

$$F_3 = \frac{\text{compression resistance of support B}}{\text{net compressive load on support B}} = \frac{R - W/2}{M/d} = \frac{24 - 2}{10} = 2.2 \quad (2.16)$$

All these factors of safety for support B are applied to the same structure with the same loading conditions, so that the differences in the values are due entirely to the different definitions of resistance and applied load (or stress). However, it is easily verified that the same result is obtained $F_1 = F_2 = F_3 = 1$ if a partial safety factor $\phi = 0.5$ is applied to the resistance R :

$$F_1 = \frac{d\phi R}{M + Wd/2} = \frac{10 \times 0.5 \times 24}{100 + 4 \times 10/2} = 1, \quad (2.17)$$

$$F_2 = \frac{\phi R}{M/d + W/2} = \frac{0.5 \times 24}{100/10 + 4/2} = 1, \quad (2.18)$$

$$F_3 = \frac{\phi R - W/2}{M/d} = \frac{0.5 \times 24 - 4/2}{100/10} = 1. \quad (2.19)$$

The same result would be achieved also using the partial safety factor $\gamma = 2$ applied to the loads:

$$F_1 = \frac{dR}{\gamma(M + Wd/2)} = \frac{10 \times 24}{2(100 + 4 \times /2)} = 1, \quad (2.20)$$

$$F_2 = \frac{R}{\gamma(M/d + W/2)} = \frac{24}{2(100/10 + 4/2)} = 1, \quad (2.21)$$

$$F_3 = \frac{R - \gamma W/2}{\gamma M/d} = \frac{24 - 24/2}{2100/10} = 1. \quad (2.22)$$

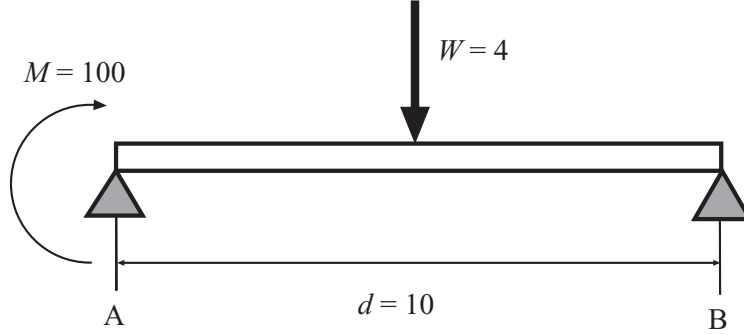


Figure 2.4: Simply supported beam from Illustrative Example 2.1.

In fact, any ϕ and γ combination such that $F_i = 1$ could be chosen:

$$F_1 = \frac{d\phi R}{\gamma(M + Wd/2)}, \quad F_2 = \frac{\phi R}{\gamma(M/d + W/2)}, \quad F_3 = \frac{\phi R - \gamma W/2}{\gamma M/d}. \quad (2.23)$$

This problem is due to the fact that the safety measures as in (2.14)-(2.16) depends on the specific choice of the failure function, not just on the failure surface. To avoid arbitrariness due to the choice of the failure function only information about the failure surface must be used, i. e., if the safety factor is calculated as a point in the failure surface, that is, when $F_i = 1$.

An alternative way to measure safety is the margin of safety, which measures the excess resistance compared with the stress resultant. Generally speaking, it is obtained reordering terms in equation (2.3), thus:

$$z = h_R(x_1, x_2, \dots, x_n) - h_S(x_1, x_2, \dots, x_n). \quad (2.24)$$

Some authors consider the safety margin to the random variable $Z = G(X_1, X_2, \dots, X_n)$.

Analogously as in (2.14)-(2.16), if information just about the failure surface is used, that in this case implies $z_i = 0$, then the three safety margins are equivalents. For example, considering $R = 24 \times 0.8 = 19.2$, $M = 1.6 \times 100 = 160$, $W = 1.6 \times 4 = 6.4$ and $d = 10$ the safety margins are:

$$\begin{aligned} z_1 &= dR - (M + Wd/2) = 10 \times 19.2 - (160 + 6.4 \times 10/2) = 0, \\ z_2 &= R - (M/d + W/2) = 19.2 - (160/10 + 6.4/2) = 0, \\ z_3 &= R - W/2 - M/d = 19.2 - 6.4/2 - 160/10 = 0. \end{aligned} \quad (2.25)$$

■

2.2.5 Invariant safety measures

As it has been shown in the Illustrative Example 2.1, one form of invariant safety measure is obtained if resistances R_i and loads Q_j acting on the structure are so factored that the ratio between any relevant pair $\phi_i R_i$ and $\gamma_j Q_j$ is unity at the point of limit state violation. This means that the design values used are in the failure surface. In simple terms, this means that all variables be reduced to a common base before being compared.

2.3 The Return Period: a Partial Probabilistic Safety Measure

In the historical development of engineering design, loads due to natural phenomena such as winds, waves, storms, floods, earthquakes, etc. were recognized quite early as having a randomness in time as well as in space. Since the primary issues of interest in all the above examples concern the occurrence of such events and their frequency, a careful statistical analysis would require the availability of data on such extremes (preferably of a large size, for making predictions accurately) and an appropriate statistical model for those extremes (which would lead to correct predictions).

The randomness in time was considered in terms of the ‘return period’. If $F(x)$ is the cdf of the yearly maxima of a random variable X , the return period, τ_x , of the event $\{X > x\}$ is $1/[1 - F(x)]$ years. Similarly, if $F(x)$ is the cdf of the yearly minima of a random variable X , the return period τ_x of the event $\{X < x\}$ is $1/F(x)$ years. Then, we have

$$\tau_x = [1 - F(x)]^{-1} \quad (2.26)$$

for exceedances, and

$$\tau_x = [F(x)]^{-1} \quad (2.27)$$

for shortfalls.

Consider a time interval of small duration τ (unit time) and the Bernoulli experiment consisting of determining whether or not the event $\{X > x\}$ occurs in such interval. Consider now a sequence of time intervals of the same duration and the corresponding Bernoulli experiments. The number of required Bernoulli experiments for the event to occur for the first time is a Pascal or geometric random variable $Ge(p)$, which means $1/p$ is the return period measured in τ units. For this to be valid, p must be small, to guarantee that no more than one event occurs in each time interval (a small probability of having more than one occurrence of the event).

Note that if an engineering work fails if and only if the event A occurs, then the mean life of the engineering work is the return period of A .

Definition 2.3 (Return period) *Return period is defined as the average (or expected) time between two successive statistically independent events.*

The importance of return period in engineering is due to the fact that many design criteria use return periods, that is, an engineering work is designed to withstand, on the average, return periods of 50, 100, or 500 years.

In addition, the probability of occurrence of the event A before the return period is (see the geometric distribution):

$$F(\tau) = 1 - (1 - p)^\tau = 1 - (1 - p)^{1/p}, \quad (2.28)$$

which for $\tau \rightarrow \infty$ ($p \rightarrow 0$) tends to 0.63212.

Remark 2.1 *If instead of recording yearly maxima of X , the maxima for a period of k years were recorded, then τ_x in (2.26) and (2.27) would be measured in periods of k years.* ■

Illustrative Example 2.2 (Return period of a flood). The cdf of the yearly maximum flood (in m^3/sec) in a given cross-section of a river is given by

$$F(x) = \exp \left[- \exp \left(- \frac{x - 38.5}{7.8} \right) \right].$$

Then, the return periods of floods of 60 and 70 m^3/sec are

$$\tau_{60} = \frac{1}{1 - F(60)} = 16.25 \text{ years}$$

and

$$\tau_{70} = \frac{1}{1 - F(70)} = 57.24 \text{ years}$$

This means that yearly maximum floods of 60 and 70 m^3/sec occur, on average, once every 16.25 and 57.24 years, respectively. ■

Illustrative Example 2.3 (Design wave height for a breakwater). If a breakwater is to be designed to withstand a mean useful life of 50 years, and the yearly maximum wave height h (in feet) is known, from previous experience, to be:

$$F(h) = \exp \left[- \exp \left(- \frac{h - 15}{4} \right) \right]. \quad (2.29)$$

Then, the design wave height must satisfy the equation

$$\frac{1}{1 - F(h)} = 50,$$

which leads to a design wave height of $h = 30.61$ feet. ■

Let $F_k(x)$ and $F_r(x)$ be the cdf of the maximum of a random variable X which is observed during periods of k and r years, respectively, where $r = sk$, and assume that the maximum in different years are independent random variables.

Then, we have

$$F_r(x) = F_k^s(x). \quad (2.30)$$

The return periods $\tau_x^{(k)}$ and $\tau_x^{(r)}$ of a value x of X , measured in periods of k and r years, respectively, are

$$\tau_x^{(k)} = \frac{1}{1 - F_k(x)} \quad \text{and} \quad \tau_x^{(r)} = \frac{1}{1 - F_r(x)}. \quad (2.31)$$

Then, using (2.31) we get

$$\tau_x^{(r)} = \frac{1}{1 - F_r(x)} = \frac{1}{1 - F_k^s(x)} = \frac{1}{1 - (1 - S_k(x))^s} \approx \frac{1}{sS_k(x)} \quad (2.32)$$

and, since

$$\tau_x^{(k)} = \frac{1}{1 - F_k(x)} = \frac{1}{1 - (1 - S_k(x))} = \frac{1}{S_k(x)}, \quad (2.33)$$

we get

$$\tau_x^{(k)} = s\tau_x^{(r)}. \quad (2.34)$$

For example, if r and k refer to one year and one month periods, respectively, then $s = 12$, and the return period measured in months is 12 times the return period measured in years.

Note that this approximation, based on the geometric distribution, is valid only if the function $F(x)$ at x is very small, so that the approximation

$$[1 - F(x)]^s \approx 1 - sF(x)$$

is valid. For that to hold, the length of the unit period considered must be small enough for the probability of one event to be small, and the probability for two or more events to be negligible. Otherwise, the model is invalid.

Illustrative Example 2.4 (Return periods based on different data durations).

Assume that the yearly maximum wave height at a given location has the Gumbel distribution in (2.29), where the return period of a wave of height $h = 39.855$ feet is 500 years. If the maximum yearly waves are assumed to be independent, the maximum wave height in a decade has the following cdf

$$F_{10}(h) = F^{10}(h) = \exp \left[-\exp \left(-\frac{h - 15 - 4 \log 10}{4} \right) \right].$$

The return period of $h = 39.855$ feet is

$$\begin{aligned} \tau_{10} &= \frac{1}{1 - F(30.61)} \\ &= \frac{1}{1 - \exp \left[-\exp \left(-(30.61 - 15 - \log 10)/4 \right) \right]} \\ &= 50.4588 \text{ decades,} \end{aligned}$$

which is approximately 50 decades. ■

2.4 Probabilistic Measures

2.4.1 Introduction

In Section 2.2 it has been shown a deterministic measure of the safety by means of safety factors. In Section 2.3 temporal uncertainty is considered using the return period concept, transforming ‘level crossing’ problems, randomly distributed in time, into deterministically design problems that can be solved using conventional design procedures. This is a useful improvement over deterministic descriptions of loading but ignores the fact that, even at a given point in time, the actual loading is uncertain.

In the presence of uncertainty, the information about the variable randomness is in the corresponding probability density functions. Resistance, geometric, workmanship and many other variables may be described in probabilistic terms.

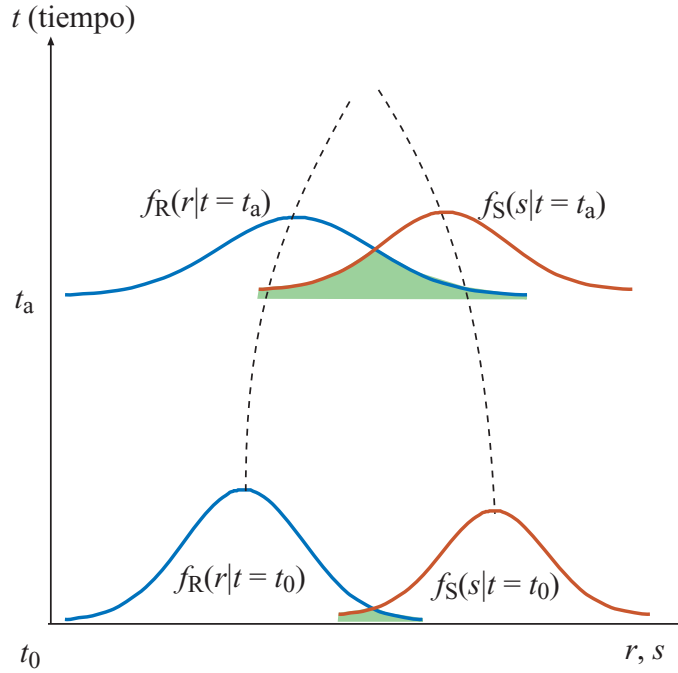


Figure 2.5: Time-dependent reliability problem schema.

Consider that the effects of all the variables with a certain level of uncertainty can be reduced to two random variables, associated with resistance R and loads or stresses S . Whose probability density functions f_R and f_S are obtained from the probability density functions of the design variables, respectively. In general, both loads and resistance are functions of time, though the loads depends in a fluctuating way, in such a way that loads have a tendency to increase and resistances to decrease with time. It is also usual for the uncertainty in both these quantities to increase with time. Therefore, the probability density functions become flatter and wider with time and the mean values also change with time approaching one to each other as it is shown in Figure 2.5.

The safety limit state will be violated whenever, at any time t ,

$$R(t) - S(t) \leq 0 \quad \text{or} \quad \frac{R(t)}{S(t)} \leq 1.$$

The probability that this occur at any time is the probability of limit state violation, or probability of failure p_f . A qualitative measure of the probability of failure may be obtained by considering the area of overlap between the two curves. This are depends on three factors:

1. The relative position of the two curves. As the relative position between the two curves increases, reducing the overlapped area, the probability of failure decreases. The position of the curves may be represented by the mean values μ_R and μ_S . Since relative positions between curves use to decrease with time, probability of failure increases with time.

2. The dispersion of the two curves. If the curves are narrow, the overlapping is lower, and the probability of failure is lower as well. The dispersion may be characterized by the standard deviation. Analogously as in the previous case, probability of failure increases with time because uncertainty increases.
3. The shapes of the two curves. These are represented by the probability density functions $f_R(r)$ and $f_S(s)$.

In many situations, however, it is convenient to assume that the probability of failure does not depend on time, this will be the case if the load is applied only once on the structure and the probability of limit state violation is sought for that load application only. If the load is applied many times and the resistance R is considered constant. Then it is assumed that the structure can fail under the application of the (only once) maximum load during its lifetime. In this case, loads are more properly defined by extreme value distributions. If this is done, the effect of time may be ignored in the reliability calculations.

2.4.2 The basic reliability problem

The basic structural reliability problem considers only load effect S resisted by the resistance R . The uncertainties in both load and resistance variables are expressed in the form of the probability density functions, $f_S(s)$ and $f_R(r)$, respectively.

The probability of the failure event p_f , that in this case is equal to the probability of the resistance R to be lower than the stress resultant S can be stated in one of the following ways:

$$p_f = P(R \leq S) \quad (2.35)$$

$$= P(R - S \leq 0) \quad (2.36)$$

$$= P\left(\frac{R}{S} \leq 1\right) = \quad (2.37)$$

$$= P(\log R - \log S \leq 0) \quad (2.38)$$

$$= P(g(R, S) \leq 0) \quad (2.39)$$

where equation (2.39) is the limit state equation or failure function (2.2) particularized for the case where just two random variables R and S are considered. The probability of failure is equal to the probability of limit state violation.

In Figure 2.6 the joint probability density function $f_{RS}(r, s)$ of the variables R and S over the safe region, and their corresponding marginal probability density functions are shown. Equations (2.35)-(2.39) are represented in Figure 2.6 by the domain D , so that the probability of failure becomes:

$$p_f = P(g(R, S) \leq 0) = \int \int_D f_{RS}(r, s) dr ds = \int_{-\infty}^{\infty} \int_{-\infty}^s f_{RS}(r, s) dr ds \quad (2.40)$$

that is the volume of the joint density function over the failure domain D . This probability of failure is the probability that a sample point falls in the failure region. Conversely, the

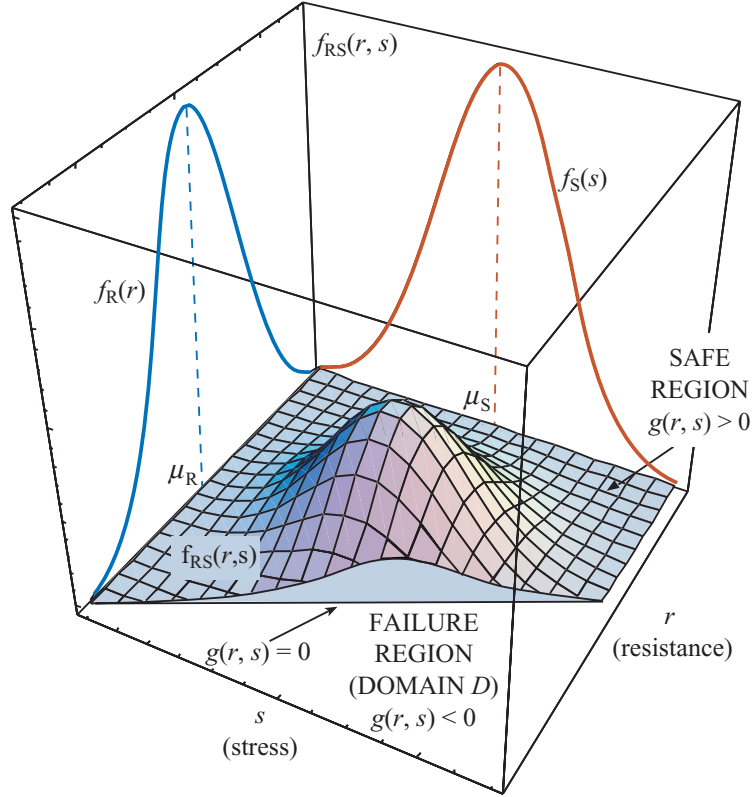


Figure 2.6: Graphical illustration of the joint density function $f_{RS}(r, s)$, marginal density functions $f_R(r)$ and $f_S(s)$ and safe and failure (D) domains for the basic reliability problem.

reliability p_r of the structure is the probability that the sample point falls in the safe region. Both, reliability and probability of failure are related as follows:

$$p_r = 1 - p_f.$$

If the random variables R and S are independent, the joint density function is obtained as the product of the marginal density functions ($f_{RS}(r, s) = f_R(r)f_S(s)$) and expression (2.40) becomes:

$$p_f = P(g(R, S) \leq 0) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^s f_R(r) dr \right) f_S(s) ds. \quad (2.41)$$

Note that the integral over R in equation (2.41) is the cumulative distribution function of R particularized for $r = s$, thus, equation (2.41) can be rewritten as follows:

$$p_f = \int_{-\infty}^{\infty} F_R(s) f_S(s) ds. \quad (2.42)$$

This way, the probability of failure is just expressed as a function of the load effect S . This function is known as ‘convolution integral’ and its meaning is easily explained by

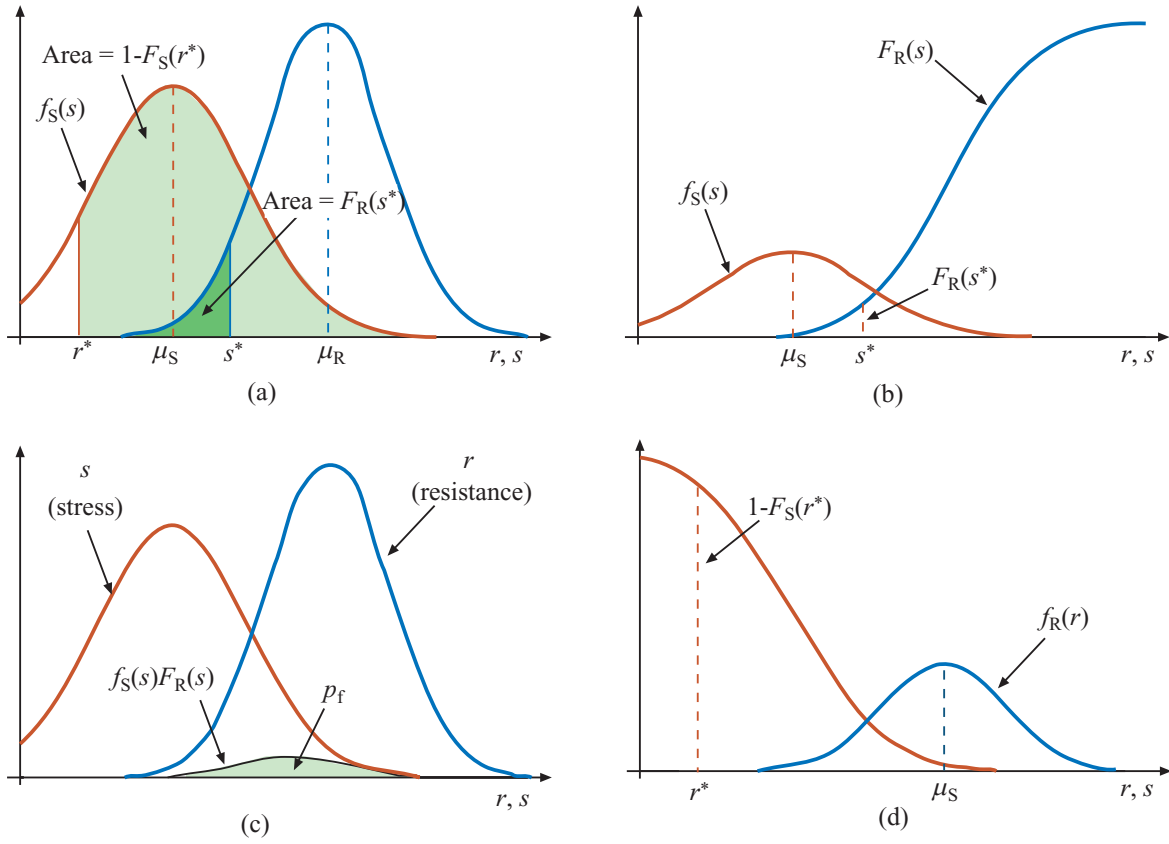


Figure 2.7: Graphical illustration of the marginal density and distribution functions, and failure probability in the basic reliability problem.

reference to Figure 2.7(b). The term $F_R(s)$ is the probability that the actual resistance is less than the load effect s , that is, the probability of failure if the actual load effect or stress is equal to s . $f_S(s)$ represent the probability of S to be equal to the value s . Function $F_R(s)f_S(s)$ is shown in Figure 2.7(c). By considering all possible values of s , i. e. integrating over all s , the total probability of failure is obtained.

Alternatively, the probability of failure can be expressed as a function just of the resistance r , equation (2.41) becomes (see Figure 2.7(d)):

$$p_f = \int_{-\infty}^{\infty} [1 - F_S(r)] f_R(r) dr, \quad (2.43)$$

where the term $1 - F_S(r)$ is the probability that load effect or stress s is bigger than resistance r , that is, failure when the actual resistance is equal to r .

It should be noted that the probability of failure p_f is not given by the area of overlap between the two density functions $f_R(r)$ and $f_S(s)$. This is a common misconception. However overlapping is a qualitative measure of the probability of failure.

Illustrative Example 2.5 (Normal random variables special case). For the general

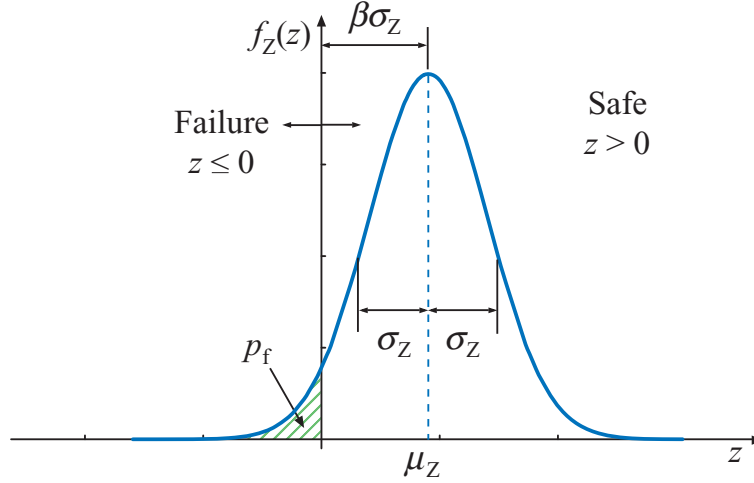


Figure 2.8: Illustration of the reliability index and the safety margin Z density function.

case, closed-form solutions do not exist for the integrals in equation (2.40). There are, however, a number of special cases. If both R and S are considered independent normal random variables with means μ_R and μ_S and standard deviations σ_R and σ_S , respectively. The safety margin $Z = R - S$ (shown in Figure 2.8) is another normal random variable $Z \sim N(\mu_Z, \sigma_Z^2)$ where:

$$\begin{aligned}\mu_Z &= \mu_R - \mu_S \\ \sigma_Z^2 &= \sigma_R^2 + \sigma_S^2\end{aligned}\quad (2.44)$$

Equation (2.35) becomes:

$$p_f = P(R - S \leq 0) = P(Z \leq 0) = \Phi\left(\frac{0 - \mu_Z}{\sigma_Z}\right) \quad (2.45)$$

where $\Phi(\cdot)$ is the standard distribution function. Using (2.44) and (2.45) it follows that (Cornell [48]):

$$p_f = \Phi\left(\frac{0 - (\mu_R - \mu_S)}{\sqrt{\sigma_R^2 + \sigma_S^2}}\right) = \Phi(-\beta) \quad (2.46)$$

where $\beta = \mu_Z/\sigma_Z$ is the ‘reliability’ or ‘safety’ index.

Note that the reliability index β may now be defined as the number of standard deviations by which the mean value of the safety margin μ_Z exceeds zero. If the standard deviations increase, the reliability index decrease and hence p_f will increase. Similarly, if the difference between the the mean of the load effect or stress and the mean of the resistance is reduced, p_f increases.

For the more general case where R and S are jointly normally distributed with correlation coefficient ρ , equation (2.45) still holds but the standard deviation of the safety margin in given by:

$$\sigma_Z = \sqrt{\sigma_R^2 + \sigma_S^2 + 2\rho\sigma_R\sigma_S}.$$

■

Illustrative Example 2.6 (Log-Normal random variables special case). If the variables can not take negative values considering the physical aspects of the problem, both R and S can be considered independent lognormal random variables with means $\mu_{\ln R}$ and $\mu_{\ln S}$ and standard deviations $\sigma_{\ln R}$ and $\sigma_{\ln S}$, respectively. The variable $Z = R/S$ is another log-normal random variable $Z \sim LN(\mu_{\ln Z}, \sigma_{\ln Z}^2)$, taking natural logarithms we get:

$$\ln Z = \ln R - \ln S \quad (2.47)$$

and equation (2.37) becomes:

$$p_f = P\left(Z = \frac{R}{S} \leq 1\right) = P(\ln R - \ln S \leq 0) = \Phi\left(\frac{0 - \mu_{\ln Z}}{\sigma_{\ln Z}}\right) = \Phi\left(\frac{\mu_{\ln S} - \mu_{\ln R}}{\sqrt{\sigma_{\ln R}^2 + \sigma_{\ln S}^2}}\right) \quad (2.48)$$

where $\Phi(\cdot)$ is the standard distribution function. Using the relationship between the mean, standard deviation, and coefficient of variation and the parameters of the lognormal distribution (1.112) it follows that :

$$p_f = \Phi\left(\frac{\ln\left(\frac{\mu_S}{\mu_R} \sqrt{\frac{1+v_S^2}{1+v_R^2}}\right)}{\sqrt{\ln((1+v_S^2)(1+v_R^2))}}\right) = \Phi(-\beta) \quad (2.49)$$

If the coefficients of variation are small (< 0.3) the approximate expression (1.113) can be used, equation (2.49) simplifies to:

$$p_f \approx \Phi\left(\frac{\ln\left(\frac{\mu_S}{\mu_R}\right)}{\sqrt{v_S^2 + v_R^2}}\right) = \Phi(-\beta) \quad (2.50)$$

■

There are some other analytical expressions for the probability of failure p_f for some other combinations of distributions of R and S . But in structural analysis, this simple representation is rarely of much practical use, so we shall not consider them here. In general, equation (2.40) needs to be evaluated by numerical methods.

2.4.3 Safety factors and characteristic values

The traditional safety measures reviewed in the previous sections can be related directly to the probability of failure p_f of limit state violation.

If we consider the central safety factor λ_0 as the ratio of the mean values of the load and resistance

$$\lambda_0 = \frac{\mu_R}{\mu_S}, \quad (2.51)$$

a relationship can be establish with the probability of failure p_f or the reliability index β . Obviously, this relationship will depend on the probability distribution for R and S , so that not general result can be given. If both are normal distributions as in the Illustrative Example 2.5, then taking expression (2.46) and dividing by μ_S we get:

$$p_f = \Phi \left(\frac{0 - (\mu_R - \mu_S)}{(v_R^2 \mu_R^2 + v_S^2 \mu_S^2)^{1/2}} \right) = \Phi \left(\frac{0 - (\lambda_0 - 1)}{(v_R^2 \lambda_0^2 + v_S^2)^{1/2}} \right) = \Phi(-\beta) \quad (2.52)$$

it follows that:

$$\lambda_0 = \frac{1 + \beta (v_R^2 + v_S^2 - \beta^2 v_R^2 v_S^2)^{1/2}}{1 - \beta^2 v_R^2}. \quad (2.53)$$

Alternatively, if we consider Illustrative Example 2.6 where the variables involved are log-normally distributed. From expression (2.49) we get that:

$$\lambda_0 = \frac{\mu_R}{\mu_S} = \frac{\exp \left(\beta \sqrt{\ln \left((1 + v_S^2)(1 + v_R^2) \right)} \right)}{\sqrt{\frac{1 + v_R^2}{1 + v_S^2}}}. \quad (2.54)$$

Using the approximative expression (2.50) it follows that

$$\lambda_0 \approx \exp \left(\beta \sqrt{v_S^2 + v_R^2} \right). \quad (2.55)$$

Note that this definition of safety factor does not accord with conventional usage, where generally some upper range value of applied load or stress is compared with some range lower value of strength. This values, usually called ‘nominal’ or ‘characteristic’ are measurable fractiles.

For resistances, the characteristic values are usually the 2nd, 5th, or 10th percentiles ($\alpha = 0.02, \alpha = 0.05, \alpha = 0.1$), and are the strengths that most (98%, 95% and 90%, respectively) samples will exceed. In other words is the resistance value r_k such that the probabilities of obtaining lower resistances are 2%, 5% and 10%, respectively. It is defined in the low side of the mean resistance μ_R :

$$r_k = \mu_R(1 - k_R v_R) \quad (2.56)$$

where $v_R = \sigma_R/\mu_R$ is the coefficient of variation, and k_R is a constant obtained as:

$$\alpha = \Phi \left(\frac{r_k - \mu_R}{\sigma_R} \right) = \Phi(-k_R); \quad k_R = -\Phi^{-1}(\alpha) = -\Phi^{-1}(0.05) \quad (2.57)$$

Note that this way the sign of k_R is positive. From equation (2.57) operating with the term

$$\frac{r_k - \mu_R}{\sigma_R} = -k_R \quad (2.58)$$

equation (2.56) is obtained (see Figure 2.9). Note that the characteristic value is located at a distance of k_R standard deviations (σ_R) below the mean

$$r_k = \mu_R - k_R \sigma_R$$

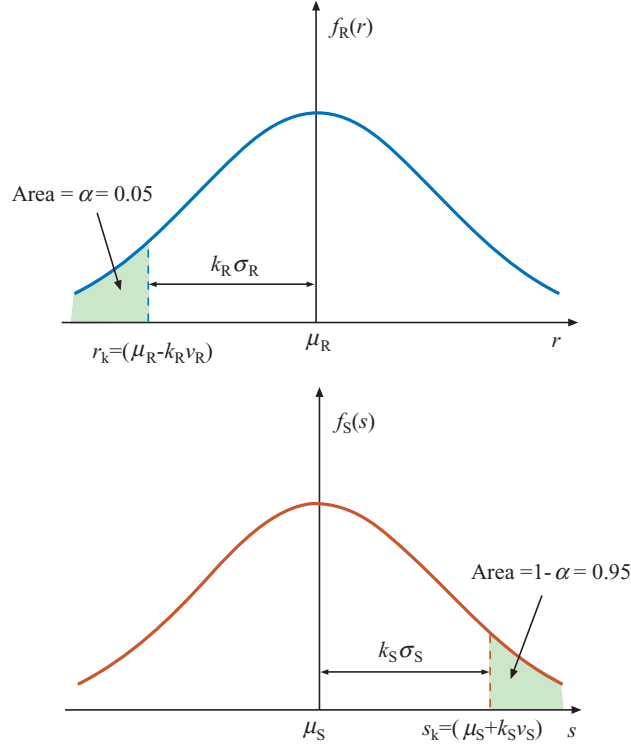


Figure 2.9: Definition of the characteristic values for resistances R and loads or stresses S .

Similarly, for the load effect or stress S the characteristic value is estimated on the high hand side of the mean (see Figure 2.9). The characteristic load or stress value s_k is the value such that the probabilities of obtaining bigger loads or stresses are 2%, 5% and 10%, respectively. It is defined in the high side of the mean load or stress μ_S :

$$s_k = \mu_S(1 + k_S v_S) \quad (2.59)$$

where $v_S = \sigma_S/\mu_S$ is the coefficient of variation, and k_S is a constant obtained as:

$$\alpha = \Phi\left(\frac{s_k - \mu_S}{\sigma_S}\right) = \Phi(k_S); \quad k_S = \Phi^{-1}(\alpha) = \Phi^{-1}(0.95) \quad (2.60)$$

From equation (2.60) operating with the term

$$\frac{s_k - \mu_S}{\sigma_S} = k_S \quad (2.61)$$

equation (2.59) is obtained. The characteristic value is located at a distance of k_S standard deviations (σ_S) above the mean

$$s_k = \mu_S + k_S \sigma_S$$

Using the nominal or characteristic values r_k and s_k the nominal or characteristic safety factor λ_k can be defined as:

$$\lambda_k = \frac{r_k}{s_k} \quad (2.62)$$

A relationship can be established between the characteristic safety factor λ_k and the central λ_0 , using (2.51), (2.59), (2.56), and (2.62):

$$\lambda_k = \frac{r_k}{s_k} = \frac{\mu_R(1 - k_R v_R)}{\mu_S(1 + k_S v_S)} = \frac{1 - k_R v_R}{1 + k_S v_S} \lambda_0, \quad (2.63)$$

so that a relationship between p_f or β , and λ_k for given v_R , v_S , k_R and k_S follows immediately:

$$\lambda_k = \frac{1 - k_R v_R}{1 + k_S v_S} \frac{1 + \beta (v_R^2 + v_S^2 - \beta^2 v_R^2 v_S^2)^{1/2}}{1 - \beta^2 v_R^2} \quad (2.64)$$

if the random variables R and S are considered normal, and

$$\lambda_k = \frac{1 - k_R v_R}{1 + k_S v_S} \frac{\exp\left(\beta \sqrt{\ln\left((1 + v_S^2)(1 + v_R^2)\right)}\right)}{\sqrt{\frac{1 + v_R^2}{1 + v_S^2}}} \quad (2.65)$$

if the random variables R and S are considered log-normally distributed.

Expressions (2.53), (2.54), (2.55), (2.64) and (2.65) indicate that the factors are strongly dependent on the uncertainty associated with R and S ; if the probability of failure p_f or β is wanted to be constant when v_R and v_S are greater, the factors k_R and k_S need to be increased. This demonstrates again the deficiencies of the deterministic measures of limit state violation. They ignore the available information about uncertainties in the random variables involved.

This relationship is very important from the calibration point of view because the ‘code parameters’ (characteristic values, safety factors, confidence levels, etc.) has to be selected for considering the uncertainty in the random variables involved. This procedure is called ‘calibration’ of the code format.

2.4.4 Generalized reliability problem

Only in the simplest structures can the reliability be expressed as a function of the two random variables R and S . The best option is to replace the simple R - S form of limit state function with a generalized version expressed in terms of the basic variables (X_1, X_2, \dots, X_n) as in (2.2). Where the terms R and S are substituted by the functions $h_R(x_1, x_2, \dots, x_n)$ and $h_S(x_1, x_2, \dots, x_n)$ represent the magnitudes related to resistance or capacity and the loads or stresses, respectively.

Mathematically, the functional relationships among the basic variables corresponding to each performance criterion or limit state, can be described as

$$Z = g(x_1, x_2, \dots, x_n). \quad (2.66)$$

The failure surface for the limit state under consideration can be defined as $Z = 0$. This $(n - 1)$ dimensional hyper-surface in the n - dimensional basic variable space is the boundary between the safe ($Z > 0$) and unsafe regions ($Z < 0$) representing the state beyond which a structure can no longer fulfill the function for it was designated.

Considering the probability density function of all the variables involved:

$$f(\mathbf{x}) = f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \Theta) \quad (2.67)$$

where Θ is a parametric vector containing the statistical parameters, such us, mean values, standard deviations, correlation coefficients, etc., data perfectly defined in the design, etc. The probability of failure is given by the integral:

$$p_f(\Theta) = \int_{g(x_1, x_2, \dots, x_n) \leq 0} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \Theta) dx_1 dx_2 \dots dx_n. \quad (2.68)$$

Note that the integral is over the failure region. The reliability of the structure may then be expressed as

$$\mathcal{R}(\Theta) = 1 - p_f(\Theta). \quad (2.69)$$

If, and only if, the basic random variables are statistically independent, then the joint probability density function may be replaced by the product of the individual probability density functions, and expression (2.68) is as follows:

$$p_f(\Theta) = \int_{g(x_1, x_2, \dots, x_n) \leq 0} f_{X_1}(x_1; \Theta_1) f_{X_2}(x_2; \Theta_2) \dots f_{X_n}(x_n; \Theta_n) dx_1 dx_2 \dots dx_n, \quad (2.70)$$

where the parametric vector $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_n)$ is decomposed by components.

Both probabilities of failure in expressions (2.68) and (2.70) are conditional because usually the statistical information of the random variables (Θ) is estimated or not known with precision. Thus, the probability of failure obtained is a ‘point estimate’, given a particular set of assumptions about the probability distributions of \mathbf{X} .

If an unconditional probability of failure (as it is usually desirable in decision making) is required the total probability theorem (1.11) can be invoked. For the present this can be done by taking the expected value of the conditional probability estimate (Der Kiureghian, [54]):

$$p_f = E[p_f(\Theta)] = \int_{\Theta} p_f(\Theta) f_{\Xi}(\Theta) d\Theta \quad (2.71)$$

where $E[\cdot]$ is the expected value, and $f_{\Xi}(\Theta)$ is the joint probability density function of the parametric vector Θ .

Illustrative Example 2.7 (Simple supported beam with lognormal random variables). Consider the simply supported beam in Figure 2.10, where P is a point load applied in the middle of the beam, L is the beam length, and b and c are the dimensions of its rectangular cross section.

To analyze the beam problem we consider the following initial set of variables:

- **Random variables:**

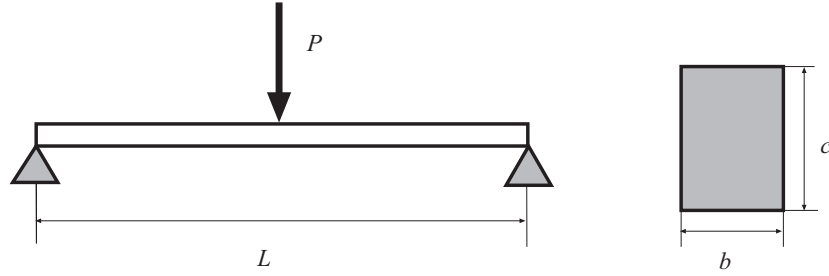


Figure 2.10: Simple supported beam and cross sectional area.

- P : Maximum value of the point load that occurs during the service life of the structure.
- s : Actual beam strength.
- L : The actual beam length.
- b : The actual width depth.
- c : The actual beam depth.

- **Design variables:**

- P_0 : Point load fixed by the code for this type of structure.
- s_0 : Beam strength fixed by the code.
- L_0 : The design beam length.
- b_0 : The design beam width.
- c_0 : The design beam depth.
- γ : Safety factor used to majorize the load P .
- ϕ : Safety factor used to minorize the strength s .

We note that we have distinguished the design variables L_0, b_0 and c_0 , which are fixed values, from the actual values L, b and c , which are random variables (normally assumed to be with mean values L_0, b_0 and c_0 , respectively, and small variance).

The aim of the example is obtained the dimensions of the cross section for given values of the beam strength, length of the beam L_0 to support the load with reasonable margin of safety.

In a classical design, i.e., based on safety coefficients. Then, the geometric design of the beam, i.e., the determination of the values of b_0 and c_0 , can be done using the well known strength of materials expression:

$$\phi s_0 = \frac{M}{W_0} = \frac{\gamma P_0 L_0 / 4}{\frac{I_0}{c_0 / 2}} = \frac{3\gamma P_0 L_0}{2b_0 c_0^2}, \quad (2.72)$$

where W_0 and I_0 are the resistant moment and the moment of inertia, respectively, associated with b_0 and c_0 .

Table 2.1: Mean and standard deviation of the natural logarithms of the random variables involved in the Illustrative Example 2.7.

Variable	Mean	Standard deviation
P	$\mu_{\ln P} = \ln(400000) = 12.8992$	$\sigma_{\ln P} = 0.257984$
s	$\mu_{\ln s} = \ln(25000000) = 17.0344$	$\sigma_{\ln s} = 0.085172$
L	$\mu_{\ln L} = \ln(10) = 2.30259$	$\sigma_{\ln L} = 0.023026$
b	$\mu_{\ln b} = \ln(0.576) = -0.551648$	$\sigma_{\ln b} = 0.011033$
c	$\mu_{\ln c} = \ln(1) = 0$	$\sigma_{\ln c} = 0.011033$

The failure region and the limit state equation of this example is obtained considering that the random variable maximum stress s_{max} is related to the random variables by the expression

$$s_{max} = \frac{3PL}{2bc^2}, \quad (2.73)$$

the system will fail if $s_{max} \geq s$. Then, the failure region will be:

$$\frac{3PL}{2bc^2} \geq s, \quad (2.74)$$

i.e.,

$$\begin{aligned} \text{Safe Region:} & \quad s - (3PL)/(2bc^2) > 0, \\ \text{Failure Region:} & \quad s - (3PL)/(2bc^2) \leq 0. \end{aligned} \quad (2.75)$$

For the sake of simplicity and with the aim of illustrating the design process and being able to calculate the exact probability of failure, we assume in this case that P, s, L, b and c are independent lognormal distributions:

$$\begin{aligned} \ln P & \sim N(\mu_{\ln P}, \sigma_{\ln P}^2), \\ \ln s & \sim N(\mu_{\ln s}, \sigma_{\ln s}^2), \\ \ln L & \sim N(\mu_{\ln L}, \sigma_{\ln L}^2), \\ \ln b & \sim N(\mu_{\ln b}, \sigma_{\ln b}^2), \\ \ln c & \sim N(\mu_{\ln c}, \sigma_{\ln c}^2) \end{aligned}$$

such that the means and standard deviations for their logarithms are shown in Table 2.1.

The exact probability of failure associated with this classical design can be obtained using the properties of the lognormal distributions and the equation (2.74), as follows:

$$\begin{aligned} p_f & = Prob[s_{max} \geq s] = Prob\left[s - \frac{3PL}{2bc^2} \leq 0\right] = Prob\left[-\frac{3PL}{2bc^2s} \leq 1\right] = \\ & = Prob\left[-\ln P + \ln s - \ln L + \ln b + 2\ln c + \ln \frac{2}{3} \leq 0\right] = \\ & = F N(-\mu_P + \mu_s - \mu_L + \mu_b + 2\mu_c + \log \frac{2}{3}, \sigma_P^2 + \sigma_s^2 + \sigma_L^2 + \sigma_b^2 + 4\sigma_c^2) \quad (0) = \end{aligned}$$

$$= \Phi \left[\frac{-\ln \frac{2}{3} + \mu_P - \mu_s + \mu_L - \mu_b - 2\mu_c}{\sqrt{\sigma_P^2 + \sigma_s^2 + \sigma_L^2 + \sigma_b^2 + 4\sigma_c^2}} \right], \quad (2.76)$$

where Φ is the cdf of the standard normal $N(0, 1)$, random variable.

Assume that the values given by the code are:

$$s_0 = 25Mp; \quad P_0 = 400000N; \quad L_0 = 10m; \quad \gamma = 1.6; \quad \phi = 1/1.5.$$

Then, according to Expression (2.72) we get

$$b_0 c_0^2 = 0.576. \quad (2.77)$$

So, we are free to choose any pair of values for variables b and c such that (2.77) holds. For example, we can design using $b_0 = 0.576m$ and $c_0 = 1.0m$.

According to Expression (2.76) the probability of failure associated with the classical design is

$$P_f = \Phi(-\beta) = \Phi(-3.19785) = 0.000692289.$$

■

Chapter 3

Regression Analysis

3.1 Introduction

Regression analysis is a statistical technique which allows modeling the functional dependency between two variables and make predictions of one variable when other variables are known. It has been widely used in many different engineering and science applications.

Linear regression is widely used in biological, behavioral and social sciences to describe possible relationships between variables. It ranks as one of the most important tools used in these disciplines. Another important application is the definition of trends in technical analysis, i.e, the long-term movement in time series data after other components have been accounted for. It tells whether a particular data set (say GDP, oil prices or stock prices) have increased or decreased over the period of time. A trend line could simply be drawn by eye through a set of data points, but more properly their position and slope is calculated using statistical techniques like linear regression. Trend lines typically are straight lines, although some variations use higher degree polynomials depending on the degree of curvature desired in the line. Trend lines are sometimes used in business analysis to show changes in data over time. This has the advantage of being simple. Trend lines are often used to argue that a particular action or event (such as training, or an advertising campaign) caused observed changes at a point in time. This is a simple technique, and does not require a control group, experimental design, or a sophisticated analysis technique. However, it suffers from a lack of scientific validity in cases where other potential changes can affect the data.

Regression analysis is also important in epidemiology, for instance, early evidence relating tobacco smoking to mortality and morbidity came from studies employing regression. Researchers usually include several variables in their regression analysis in an effort to remove factors that might produce spurious correlations. For the cigarette smoking example, researchers might include socio-economic status in addition to smoking to ensure that any observed effect of smoking on mortality is not due to some effect of education or income. However, it is never possible to include all possible confounding variables in a study employing regression. For the smoking example, a hypothetical gene might increase mortality and also cause people to smoke more. For this reason, randomized controlled trials are often able to generate more compelling evidence of causal relationships than correlational analysis using linear regression. When controlled experiments are not feasible, variants

of regression analysis such as instrumental variables and other methods may be used to attempt to estimate causal relationships from observational data.

Financial applications also include regression analysis. The capital asset pricing model uses linear regression as well as the concept of Beta for analyzing and quantifying the systematic risk of an investment. This comes directly from the Beta coefficient of the linear regression model that relates the return on the investment to the return on all risky assets. Regression may not be the appropriate way to estimate beta in finance given that it is supposed to provide the volatility of an investment relative to the volatility of the market as a whole. This would require that both these variables be treated in the same way when estimating the slope. Whereas regression treats all variability as being in the investment returns variable, i.e. it only considers residuals in the dependent variable.

In environmental sciences, linear regression finds application in a wide range of environmental science applications. For example, recent work published in the Journal of Geophysical Research used regression models to identify data contamination, which led to an overstatement of global warming trends over land. Using the regression model to filter extraneous, non-climatic effects reduced the estimated 1980-2002 global average temperature trends over land by about half, and so on.

Note that determining the linear relationship between two variables is called *simple linear regression* analysis. Between more than two variables is called *multiple linear regression* analysis, and determining the nonlinear relationship between random variables is called *nonlinear regression* analysis. All these concepts are discussed in the following sections.

3.2 Simple Linear Regression Analysis

In this section we introduce the simple linear regression model using an illustrative example. The standard simple linear regression model is

$$Y = \beta_0 + \beta_1 X + \varepsilon, \quad (3.1)$$

where Y is the response variable (random), X is the predictor variable, $\beta = (\beta_0 \beta_1)^T$ is the vector of regression parameters, and ε is an independent random error.

We illustrate the method using the star cluster data set, which is a well-known data set in the area of sensitivity analysis and outliers detection (influential data) and it has been analyzed by many authors. The data set can be found in [147], p. 57. Two variables are measured for each of 47 stars: the effective temperature at the surface of a star (x) and the light intensity of the star (y). The scatter plot of Y versus X in Figure 3.1 shows there is a direct relationship between the two variables for all the stars except for four unusual stars at the upper-left corner of the graph. These stars, which are known as the giant stars, have low temperature with high light intensity. They can be expected to exert undue influence on the estimated regression parameters. We have selected a two-dimensional example with the purpose of illustrating concepts using some graphical displays that are only possible for two dimensions.

In the context of linear regression analysis (3.1) is expressed as:

$$E(Y|X = x) = E(Y|x) = \beta_0 + \beta_1 x, \quad (3.2)$$

i	x	y	i	x	y
1	4.37	5.23	24	4.38	5.02
2	4.56	5.74	25	4.42	4.66
3	4.26	4.93	26	4.29	4.66
4	4.30	5.19	27	4.38	4.90
5	4.46	5.46	28	4.22	4.39
6	3.84	4.65	29	3.48	6.05
7	4.57	5.27	30	4.38	4.42
8	4.26	5.57	31	4.56	5.10
9	4.37	5.12	32	4.45	5.22
10	3.49	5.73	33	3.49	6.49
11	4.43	5.45	34	4.23	4.34
12	4.48	5.42	35	4.62	5.62
13	4.01	4.05	36	4.53	5.10
14	4.29	4.26	37	4.45	5.22
15	4.42	4.58	38	4.53	5.18
16	4.23	3.94	39	4.43	5.57
17	4.42	4.18	40	4.38	4.62
18	4.23	4.18	41	4.45	5.06
19	3.49	5.89	42	4.50	5.34
20	4.29	4.38	43	4.45	5.34
21	4.29	4.22	44	4.55	5.54
22	4.42	4.42	45	4.45	4.98
23	4.49	4.85	46	4.42	4.50

Table 3.1: Star cluster data set. Giant stars data are boldfaced.

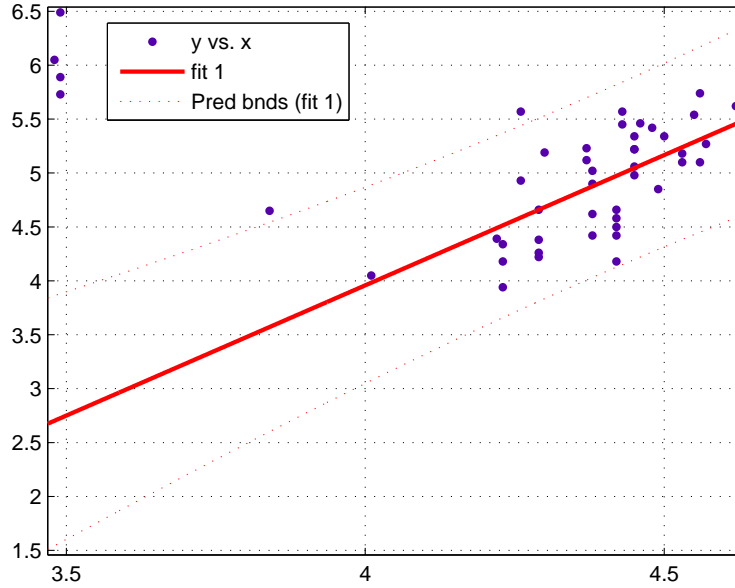


Figure 3.1: Scatter plot of Y versus X for the star cluster data. Note that the regression line excludes the giant stars.

which provides the expected value of the *dependent* or *response* variable Y for a given value x of the *predictor*, *regressor* or *independent* variable.

Regression analysis consists on, based on data, calculating the regression coefficients based on the following assumptions:

1. There is a linear relationship between dependent and independent variables, as the scatter diagram indicates.
2. X observations are error free, but predictor variable Y has an error term ε which has zero mean and an unknown but constant variance σ^2 implying that the spread in data about the regression line is constant (see Figure 3.2).
3. Errors are uncorrelated and normally distributed.

Remark 3.1 Note that the term *linear regression* accounts only for the nonlinearity for the regression parameters, so that nonlinearities on the dependent variables do not affect the model, like in the following examples:

Polynomial linear regression

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_p x^p + \varepsilon$$

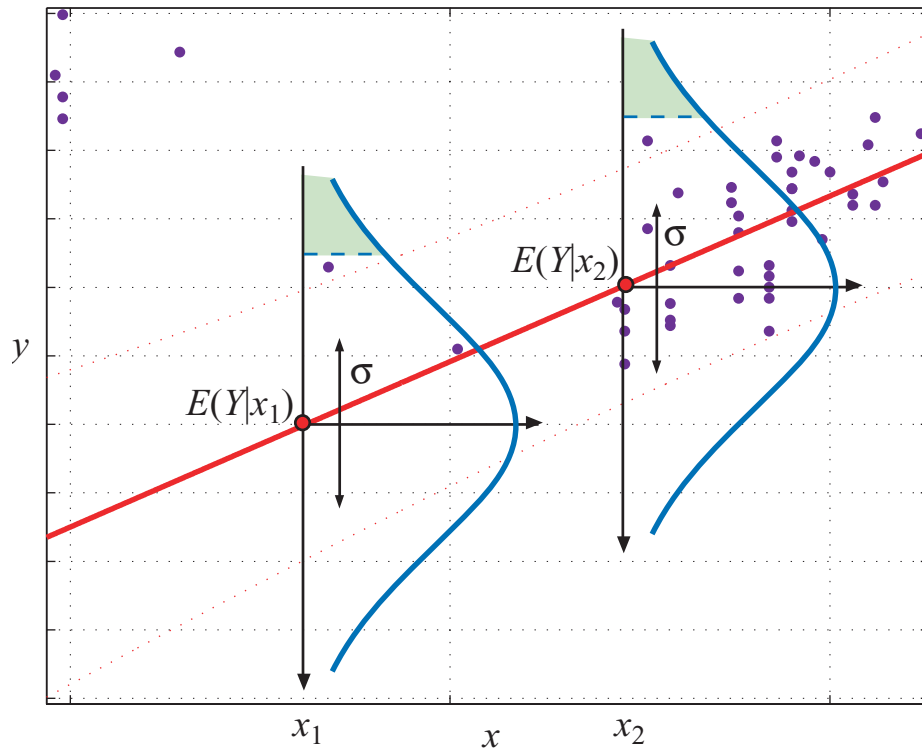


Figure 3.2: Regression line and probability density functions of the response variable Y for given values of the regressor variable X .

Trigonometric linear regression

$$y = \beta_0 + \beta_1 \sin(x) + \beta_2 \cos(x) + \dots + \beta_{2p-1} \sin(px) + \beta_{2p} \cos(px) + \varepsilon$$

Logarithm linear regression

$$y = \beta_0 + \beta_1 \log(x) + \varepsilon$$

■

3.3 Simple Linear Regression Estimation

The main objective of the regression analysis is to estimate the parameters β_0, β_1 which best fit the data. The most widely used method is the least-squares analysis that minimizes the sum of squares of the differences (errors) between the observed and predicted values of the dependent variable as follows:

$$\text{Minimize}_{\beta_0, \beta_1} \text{SS}_E = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2, \quad (3.3)$$

where SS_E is the *error sum of squares*.

Using the Lagrange multipliers method, the equations for the two unknown regression coefficients are:

$$\frac{\partial \text{SS}_E}{\partial \beta_0} = \sum_{i=1}^n -2(y_i - \beta_0 - \beta_1 x_i) = 0, \quad (3.4)$$

$$\frac{\partial \text{SS}_E}{\partial \beta_1} = \sum_{i=1}^n -2x_i(y_i - \beta_0 - \beta_1 x_i) = 0. \quad (3.5)$$

The first estimate parameter $\hat{\beta}_0$ is obtained from the first equation:

$$\hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^n y_i - \frac{\hat{\beta}_1}{n} \sum_{i=1}^n x_i = \bar{y} - \hat{\beta}_1 \bar{x}, \quad (3.6)$$

where \bar{x} and \bar{y} are the estimated mean values, whereas from (3.5):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}. \quad (3.7)$$

Note that (3.7) can be conveniently expressed as:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \quad (3.8)$$

where S_{xy} and S_{xx} are the covariance and variance with respect to x , respectively.

Computational Example 3.1 (Star data set regression coefficients). Based on star data given in Table 3.1, without considering giant stars (10,19,29,33), the regression coefficient estimates are:

$$\begin{aligned}\bar{x} &= 4.3824, \\ \bar{y} &= 4.9017, \\ \hat{\beta}_1 &= \frac{S_{xy}}{S_{xx}} = \frac{0.0430}{0.0220} = 1.9553, \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} = 4.9017 - 1.9553 \times 4.3824 = -3.6673.\end{aligned}$$

■

Once the parameters of the regression line are estimated it is also of interest the *error mean square* or *residual variance*, whose unbiased estimator is as follows:

$$MS_E = S_{Y|X}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}. \quad (3.9)$$

Note that in the denominator the number of degrees of freedom equal to $n - 2$ are subtracted from the total number of data, since there are two parameters to be estimated.

The adequacy of the regression model in representing the relationship between regressor and response variables is of great interest and one of the parameters used for this task is known as *coefficient of determination* or *correlation coefficient* R^2 . In order to derive its expression let consider the total variability of the variable Y given by S_{yy} , which has two components:

$$\sum_{i=1}^n (y_i - \bar{y}_i)^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (3.10)$$

or

$$S_{yy} = SS_R + SS_E, \quad (3.11)$$

First component accounts for the variability in the observations due to the regression model (SS_R), and the second component accounts for the residual variation unexplained by the regression model (SS_E). The coefficient of determination is a non-dimensional quantity defined as:

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}}. \quad (3.12)$$

R^2 has always a positive value and ranges from 0 to 1. Note that if it is equal to 1, the error sum of squares is zero, which means that data points fits perfectly the regression line.

3.4 Residual Analysis

Before a regression model is accepted it is important to evaluate the underlying assumptions about residuals. The residual is defined as the observed value of the dependent variable minus the predicted value and is expressed as:

$$\varepsilon_i = y_i - \hat{y}_i; \quad i = 1, \dots, n. \quad (3.13)$$

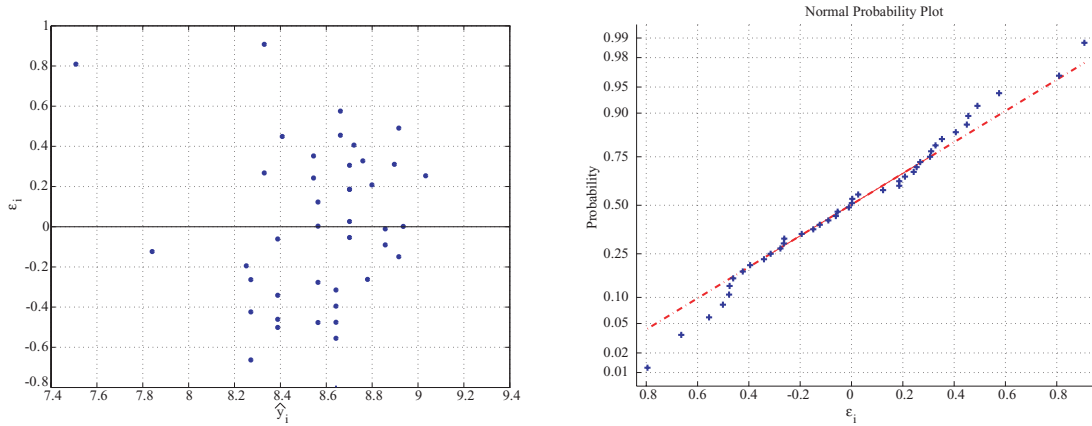


Figure 3.3: Residual analysis of Star data set: a) Residuals versus predicted values, and b) normal probability plot.

Note that the residual is a measure of the variability in the dependent variable Y not considered by the regression model. Since the residuals must be independent zero mean normal random variables. There are several statistical analysis techniques for residual analysis, but the most informative are:

1. The scatter plot residuals versus predicted values. If residuals have an horizontal band on both sides of the zero axis, the mean value is approximately zero, and the variance is constant about the regression line, as shown in Figure 3.3 a). Note that plot like those shown in Figure the regression analysis is unsatisfactory and higher order terms, or transformations on the dependent variables are required.
2. Residual normal probability plot, so that if plot appears to be linear, it satisfies the normality assumption of the residuals. See Figure 3.3 b).

In Figure 3.4 the scatter plot residual versus predicted variables for two different cases where the regression analysis hypothesis are not fulfilled are shown. Note that in the first case (left graph) the variance is not constant, and for the second case (right graph) there is a nonlinear relationship which makes the model invalid.

For those cases which do not hold the homoscedasticity assumption (constant variance) data is transformed to stabilize variance as follows:

$$Z = g(Y) = K \int \frac{1}{\sigma(Y)} dY, \quad (3.14)$$

where K is a constant and $\sigma(Y)$ is the evolution function of σ . The most widely used functions corresponds to the model $\sigma(Y) = Y^\gamma$ where:

$$Z = g(Y) = \begin{cases} KY^{1-\gamma} & \text{si } \gamma \neq 1 \\ K \log(Y) & \text{si } \gamma = 1. \end{cases} \quad (3.15)$$

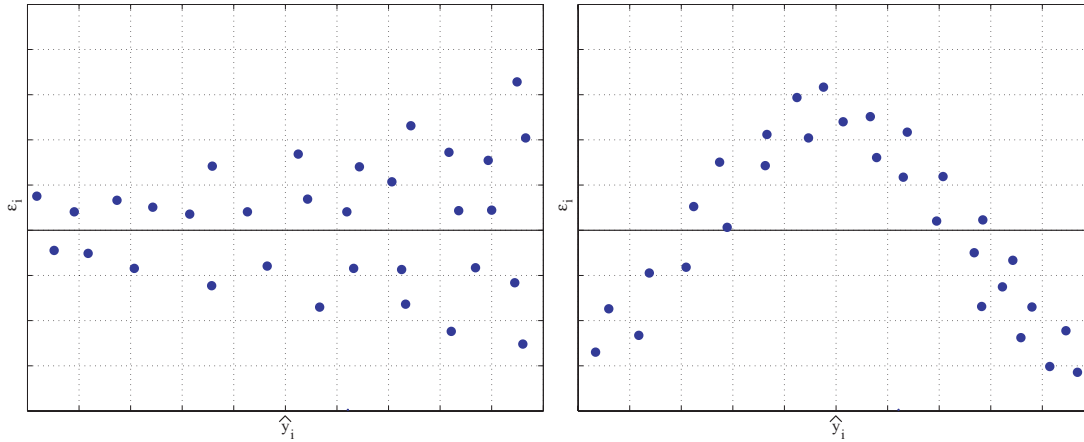


Figure 3.4: Scatter plot residual analysis.

3.5 Multiple Linear Regression

Linear regression analysis can be generalized to consider multiple independent variables X_1, X_2, \dots, X_p . The regression model (3.1) is rewritten as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (3.16)$$

where

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \dots & x_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{pn} \end{pmatrix}; \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}; \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}; \quad (3.17)$$

where \mathbf{Y} contains n observations of the dependent variable, \mathbf{X} is the design matrix of size $n \times (p+1)$ where x_{ij} is the observation of the i th regressor variable in the j th data set, $\boldsymbol{\beta}$ is the regression coefficients vector to be determined with p components, and $\boldsymbol{\varepsilon}$ are uncorrelated and identically distributed normal random variables with zero mean and unknown constant variance. This notation allows writing several complex expressions easily:

$$\begin{aligned} \varepsilon_1^2 + \dots + \varepsilon_n^2 &= \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \\ y_1^2 + \dots + y_n^2 &= \mathbf{Y}^T \mathbf{Y} \\ y_1 + \dots + y_n &= \mathbf{1}^T \mathbf{Y} \\ (y_1^2 + \dots + y_n^2)/n &= \mathbf{Y}^T \mathbf{Y}/n \end{aligned} \quad (3.18)$$

where $\mathbf{1}$ is a column vector with ones.

If the least-squares method is used to estimate the parameters the following problem is solved:

$$\text{Minimize}_{\boldsymbol{\beta}} \quad SS_E = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) . \quad (3.19)$$

The objective function can be expanded as follows:

$$\begin{aligned}
 \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \sum_{i=1}^n \left(y_i - \sum_{j=0}^p \beta_j x_{ij} \right)^2 \\
 &= \mathbf{Y}^T \mathbf{Y} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \\
 &= \mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}.
 \end{aligned} \tag{3.20}$$

Deriving expression (3.20) with respect to $\boldsymbol{\beta}$ and considering that at the minimum the derivative must be zero:

$$\frac{\partial \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = 0, \tag{3.21}$$

the system of *normal equations* [42] is obtained. Assuming that matrix $\mathbf{X}^T \mathbf{X}$ is nonsingular, the parameter estimator is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \tag{3.22}$$

which demonstrates that the regression parameter estimator is a linear function of the predictor variables \mathbf{Y} .

The parameter estimator variance-covariance matrix can be obtained from (3.22) considering that the variance-covariance matrix of \mathbf{Y} is known. Note that for linear combinations of normal random variables, the variance-covariance matrix is obtained as follows:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \mathbf{C} \boldsymbol{\Sigma}_{\mathbf{Y}} \mathbf{C}^T. \tag{3.23}$$

From (3.22) and since $\hat{\boldsymbol{\beta}}$ is a linear combination of \mathbf{Y} , given by $\mathbf{C} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$, it results in the following variance-covariance matrix:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{Y}} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}, \tag{3.24}$$

and since $\boldsymbol{\Sigma}_{\mathbf{Y}} = \sigma^2 \mathbf{I}$:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \sigma^2 \left[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{I} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \right] = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}, \tag{3.25}$$

where σ^2 is the constant residual variance.

Analogously to (3.9), an unbiased estimate for the conditional variance can be calculated as follows:

$$\text{MS}_E = S_{\mathbf{Y}|\mathbf{X}}^2 = \frac{\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}{n - p - 1} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p - 1}. \tag{3.26}$$

Note that the number of degrees of freedom associated with the $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$ is equal to $n - p - 1$, since the mean value of the residuals must be zero and there are $p + 1$ estimated parameters.

It is also possible to calculate the coefficient of determination given by (3.10), starting from (3.20) and using the estimated parameters (3.22):

$$\begin{aligned}
\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} &= \mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \\
&= \mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \\
&= \mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} \\
&= \mathbf{Y}^T \mathbf{Y} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y},
\end{aligned} \tag{3.27}$$

and adding the term $-n\bar{y}^2$ in both sides of the last equation and rearranging:

$$\mathbf{Y}^T \mathbf{Y} - n\bar{y}^2 = (\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} - n\bar{y}^2) + \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}, \tag{3.28}$$

where the correspondence with the terms in (3.20) is as follows:

$$\begin{aligned}
S_{yy} &= \sum_{i=1}^n (y_i - \bar{y}_i)^2 = \mathbf{Y}^T \mathbf{Y} - n\bar{y}^2, \\
SS_R &= \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2 = \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} - n\bar{y}^2, \\
SS_E &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon},
\end{aligned} \tag{3.29}$$

and the correlation coefficient for the multivariate case is equal to:

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}} = \frac{\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} - n\bar{y}^2}{\mathbf{Y}^T \mathbf{Y} - n\bar{y}^2}. \tag{3.30}$$

Computational Example 3.2 (Star data set regression coefficients). Based on star data given in Table 3.1, without considering giant stars (10,19,29,33), and the regression coefficient estimates given in Computational Example 3.1, the coefficient of determination is:

$$R^2 = 0.3458,$$

and the parameter variance-covariance matrix is:

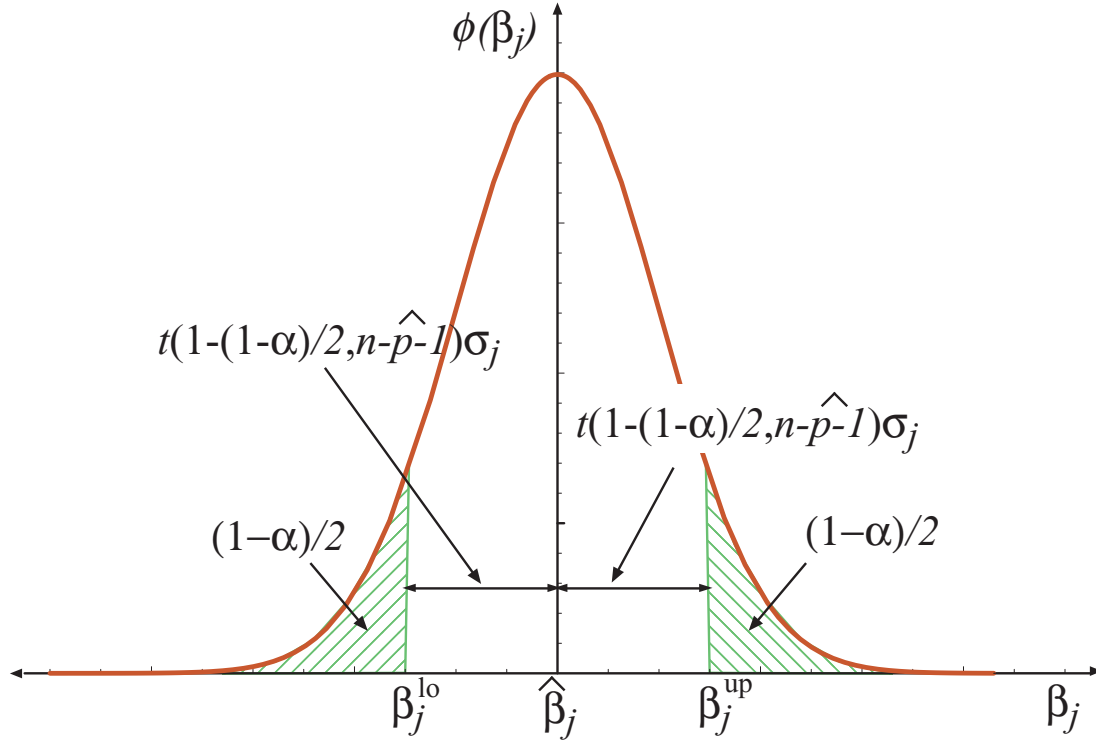
$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \begin{pmatrix} 3.4768 & -0.7925 \\ -0.7925 & 0.1808 \end{pmatrix}.$$

■

3.5.1 Parameter confidence intervals

The solution to the normal equations provides the mean values of the estimated parameters, and considering i) the variance-covariance matrix given by (3.25) and ii) that observational errors are normally distributed, the estimated parameter vector is exactly distributed as follows:

$$\hat{\boldsymbol{\beta}} \sim N\left(\hat{\boldsymbol{\beta}}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}\right), \tag{3.31}$$

Figure 3.5: α confidence interval for each parameter.

where N denotes the multivariate normal distribution.

The α (0.9, 0.95 or 0.99) confidence interval for each parameter is equal to (see Figure 3.5):

$$\begin{aligned}\beta_j^{\text{up}} &= \hat{\beta}_j + t_{(1-(1-\alpha)/2, n-p-1)} \hat{\sigma}_j, \quad j = 0, 1, \dots, p \\ \beta_j^{\text{lo}} &= \hat{\beta}_j - t_{(1-(1-\alpha)/2, n-p-1)} \hat{\sigma}_j, \quad j = 0, 1, \dots, p,\end{aligned}\quad (3.32)$$

where $t_{(1-(1-\alpha)/2, n-p-1)}$ is the Student's t -distribution $(1 - (1 - \alpha)/2)$ quantile with $n - p - 1$ degrees of freedom and $\hat{\sigma}_j$ is the estimated standard deviation for parameter j :

$$\hat{\sigma}_j = \sqrt{\frac{\hat{\boldsymbol{\epsilon}}^T \hat{\boldsymbol{\epsilon}}}{n - p - 1} \left[(\mathbf{X}^T \mathbf{X})^{-1} \right]_{jj}}, \quad (3.33)$$

where $\left[(\mathbf{X}^T \mathbf{X})^{-1} \right]_{jj}$ is the j -th diagonal element of the inverse matrix. This formula corresponds to the square root of the diagonal matrix obtained from expression (3.25).

Computational Example 3.3 (Star data set parameter confidence intervals). Based on the regression coefficient estimates given in Computational Example 3.2, the parameter confidence intervals (95%) are:

$$\begin{aligned}\beta_0^{\text{up}} &= 0.1012; & \beta_1^{\text{up}} &= 2.8148; \\ \beta_0^{\text{lo}} &= -7.4359; & \beta_1^{\text{lo}} &= 1.0959.\end{aligned}$$

■

It is also possible to calculate mean response and predicted response confidence intervals using the delta method, which given the parameter variance-covariance matrix $\Sigma_{\hat{\beta}}$ allows obtaining the covariance matrix of $\psi = h(\beta)$ as given in the following theorem.

Theorem 3.1 (The Delta Method). *Let $\hat{\beta}$ be the maximum likelihood estimate of β with variance and covariance matrix $\Sigma_{\hat{\beta}}$. Let $\psi = h(\beta)$ be the new parameter vector defined as a function of β . Then, for large sample size n , $\hat{\psi} = h(\hat{\beta})$ is asymptotically normal, that is,*

$$\hat{\psi} \sim N(h(\hat{\beta}), \nabla_{\beta}^T \psi \Sigma_{\hat{\beta}} \nabla_{\beta} \psi) \quad (3.34)$$

where $\nabla_{\beta} \psi$ is the $(p+1)$ vector of partial derivatives of ψ with respect to β , which is given by

$$\nabla_{\beta} \psi = \left[\frac{\partial \psi}{\partial \beta_0} \quad \frac{\partial \psi}{\partial \beta_1} \quad \cdots \quad \frac{\partial \psi}{\partial \beta_p} \right]^T. \quad (3.35)$$

Thus, the covariance matrix of ψ is $\nabla_{\beta}^T \psi \Sigma_{\hat{\beta}} \nabla_{\beta} \psi$.

Given a value of the independent variable, \mathbf{x}_d , the predicted response is:

$$y_d = \mathbf{x}_d^T \hat{\beta} = \sum_{i=0}^p x_{d_i} \hat{\beta}_i, \quad (3.36)$$

where $x_{d_0} = 1$, which corresponds to the expected value or mean response. Since $\psi = y_d$ and thus $\nabla_{\beta} \psi = \mathbf{x}_d$, the mean response variance is equal to:

$$\sigma_{y_d}^2 = \mathbf{x}_d^T \Sigma_{\hat{\beta}} \mathbf{x}_d = \hat{\sigma}^2 \mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_d, \quad (3.37)$$

so that the corresponding confidence intervals are:

$$\mathbf{x}_d^T \hat{\beta} \pm t_{(1-(1-\alpha)/2, n-p-1)} \hat{\sigma} \sqrt{\mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_d}. \quad (3.38)$$

If on the other hand, the predicted response confidence intervals are looked for, the uncertainty not explained by the regression model should be included in the expected value variance, so that, the predicted response confidence intervals are:

$$\mathbf{x}_d^T \hat{\beta} \pm t_{(1-(1-\alpha)/2, n-p-1)} \hat{\sigma} \sqrt{1 + \mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_d}. \quad (3.39)$$

Note that 100 α % of data should be between these bands.

3.5.2 Residuals

The residual is defined as

$$\hat{\mathbf{r}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}, \quad (3.40)$$

$$\hat{\mathbf{r}} = \mathbf{Y} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (3.41)$$

thus

$$\hat{\mathbf{r}} = \left[\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \right] \mathbf{Y} = \mathbf{S}\mathbf{Y}, \quad (3.42)$$

where $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is known as *hat* or *projection* matrix, which is an *idempotent* matrix [89], i.e.,

$$\left[\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \right]^m = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T,$$

and \mathbf{S} is known as the residual sensitivity matrix, which is also an idempotent matrix.

Since residuals are a linear combination of \mathbf{Y} , its variance-covariance matrix is:

$$\begin{aligned} \boldsymbol{\Omega} &= \text{E}[\hat{\mathbf{r}}\hat{\mathbf{r}}^T] = \text{E}[\mathbf{S}\mathbf{Y}(\mathbf{S}\mathbf{Y})^T] \\ &= \text{E}[\mathbf{S}\mathbf{Y}\mathbf{Y}^T\mathbf{S}^T] \\ &= \mathbf{S}\text{E}[\mathbf{Y}\mathbf{Y}^T]\mathbf{S}^T \\ &= \mathbf{S}\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{S}^T = \sigma^2\mathbf{S}\mathbf{S}^T \end{aligned} \quad (3.43)$$

The normalized residual ($N(0, 1^2)$) of data i is:

$$r_i^N = \frac{\hat{r}_i}{\sqrt{\Omega_{ii}}}. \quad (3.44)$$

The largest residual identifies a bad measurement with a α confidence level (e.g., 0.99) if $|r_i^N| > \Phi^{-1}(1 - (1 - \alpha)/2)$. Further details can be found in [3].

3.6 Bad Data Detection

If errors are Gaussian-distributed unbiased and independent, square sum of residuals $\hat{\mathbf{r}}^T\hat{\mathbf{r}}$ is a χ^2 with $n - p - 1$ degrees of freedom [42]. Thus, we can write

$$\text{Prob} \left(\hat{\mathbf{r}}^T\hat{\mathbf{r}} \leq \chi^2(\alpha, n - p - 1) \right) = \alpha, \quad (3.45)$$

where α is the confidence level.

Therefore, for a given α (e.g., 0.99), the value $\chi^2(\alpha, m + r - n)$ can be computed and the χ^2 test applied, i.e., if $\hat{\mathbf{r}}^T\hat{\mathbf{r}} < \chi^2(\alpha, n - p - 1)$ there is no bad measurement at the α confidence level, otherwise there is. Further details can be found in [3].

3.7 Maximum Likelihood Inference Method

One advantage of using least squares method for parameter estimation is that the solution corresponds to the maximum likelihood estimate. Note that the log-likelihood function for the ϵ independent and normally distributed errors is:

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \quad (3.46)$$

Differentiating with respect to the parameter $\boldsymbol{\beta}$, we get:

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} (\mathbf{X}^T \mathbf{Y} - \mathbf{X}^T \mathbf{X} \boldsymbol{\beta}) = 0, \quad (3.47)$$

and setting this $p + 1$ equations to zero and solving for $\boldsymbol{\beta}$ gives:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}, \quad (3.48)$$

which is equal to expression (3.22).

It can be checked that it is a maximum by looking at the Hessian matrix of the log-likelihood function:

$$\frac{\partial^2 \ell(\boldsymbol{\beta}, \sigma^2)}{\partial^2 \boldsymbol{\beta}} = \frac{-\mathbf{X}^T \mathbf{X}}{\sigma^2}. \quad (3.49)$$

Note also that using the method of maximum likelihood the parameters covariance matrix is equal to the inverse of the *Fisher information matrix*, $\mathbf{I}_{\boldsymbol{\beta}}$, which under certain regularity conditions is equal to the Hessian matrix with the sign changed:

$$\mathbf{I}_{\boldsymbol{\beta}} = -\frac{\partial^2 \ell(\boldsymbol{\beta}, \sigma^2)}{\partial^2 \boldsymbol{\beta}} = \frac{\mathbf{X}^T \mathbf{X}}{\sigma^2}, \quad (3.50)$$

whose inverse is the variance-covariance matrix:

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}, \quad (3.51)$$

which is equal to expression (3.25).

3.8 Stepwise Regression

There are cases in multivariate regression analysis when it is not clear which are the dependent variables truly significant in the models. To overcome this difficulty stepwise regression methods are used, and the choice of predictive variables is carried out by an automatic procedure. Usually, this takes the form of a sequence of F -tests, but other techniques are possible, such as t -tests, adjusted R -square, Akaike information criterion, Bayesian information criterion, Mallows statistic C_p , or false discovery rate.

At each stage in the process, after a new variable is added, a test is made to check if some variables can be deleted without appreciably increasing the residual sum of squares (SS_E). The procedure terminates when the measure is (locally) maximized, or when the available improvement falls below some critical value.

Let consider two models:

$$\text{Model 1 : } Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (3.52)$$

$$\text{Model 2 : } Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \beta_{p+1} x_{p+1} + \dots + \beta_q x_q + \varepsilon. \quad (3.53)$$

$$(3.54)$$

The minimum residual sum of squares for each model are SS_1 and SS_2 . Note that model 2 is always able to fit data better than model 1, since it includes parameters from model 1. For this reason the following statistic can be calculated:

$$F = \frac{\frac{SS_1 - SS_2}{q - p}}{\frac{SS_2}{n - p - 1}}, \quad (3.55)$$

where SS_i is the residual sum of squares of model i . Under the null hypothesis that model 2 does not provide a significantly better fit than model 1, F will have an F -Snedecor distribution with $(q - p, n - p - 1)$ degrees of freedom. The null hypothesis is rejected if the F calculated from the data is greater than the critical value of the F -Snedecor distribution for some desired false-rejection probability.

Alternative statistics to evaluate the best model are:

1. *Adjustment coefficient of determination or correlation*, which is defined as:

$$R_a^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}, \quad (3.56)$$

which penalize models with a high number of parameters ($p + 1$). The best model is the one whose R_a is minimum.

2. Mallows C_p statistic, which was proposed only as a way of facilitating comparisons among many alternative subset regressions, but it can be used as a subsetting criterion in selecting a reduced model without such problems. C_p is defined as:

$$C_p = \frac{SS_{p+1}}{s^2} - (n - 2(p + 1)), \quad (3.57)$$

where $p + 1$ is the number of parameters and s^2 is a estimate of σ^2 . Note that $E[C_p] = p + 1$.

3.9 Linear Regression Using Transformed Models

The proposed regression models shown in previous sections were linear with respect parameters, which allows using all derivations with respect estimation parameters and confidence intervals and so on. However there are non-linear regression models which can be transformed into linear through simple transformations. Some examples are the following:

- **Exponential:** The model

$$y = \beta_0 e^{\beta_1 x}, \quad (3.58)$$

which considering the transformation $y' = \log(y)$ becomes:

$$y' = \log(\beta_0) + \beta_1 x = \beta'_0 + \beta_1 x. \quad (3.59)$$

- **Power:** The model

$$y = \beta_0 x^{\beta_1}, \quad (3.60)$$

which considering the transformation $y' = \log(y)$ and $x' = \log(x)$ becomes:

$$y' = \log(\beta_0) + \beta_1 \log(x) = \beta'_0 + \beta_1 x'. \quad (3.61)$$

3.10 Non-Linear Regression

Alternatively to linear and linearizable models there are intrinsically (nonlinearizable) non-linear regression models, which can be written as

$$y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (3.62)$$

where y_i is the i th value of the response variable, \mathbf{x}_i is a $k \times 1$ vector of predictor variables corresponding to the i th observation, and ε_i is a random error. The function f is known and nonlinear in the parameter vector $\boldsymbol{\beta}$. Suppose we wish to estimate the parameters of the nonlinear model in (3.62) using the LS, LAV, and MM methods. These can be formulated as optimization problems as follows.

The most popular method for estimating the regression parameters $\boldsymbol{\beta}$ is the least squares (LS) method, where we minimize the sum of squared distances between observed and predicted values, that is,

$$\text{Minimize}_{\boldsymbol{\beta}} Z_{LS} = \sum_{i=1}^n (y_i - f(\mathbf{x}_i; \boldsymbol{\beta}))^2. \quad (3.63)$$

Two less common alternatives to the least squares method are the minimax method (MM) and the least absolute value (LAV) method. In the MM regression problem, we minimize the maximum of the distances between observed and predicted values, i.e.:

$$\text{Minimize}_{\boldsymbol{\beta}, \varepsilon} Z_{MM} = \varepsilon \quad (3.64)$$

subject to

$$\begin{aligned} y_i - f(\mathbf{x}_i; \boldsymbol{\beta}) &\leq \varepsilon, & i = 1, \dots, n, \\ f(\mathbf{x}_i; \boldsymbol{\beta}) - y_i &\leq \varepsilon, & i = 1, \dots, n. \end{aligned} \quad (3.65)$$

The LAV method minimizes the sum of the distances between observed and predicted values, i.e.:

$$\text{Minimize}_{\boldsymbol{\beta}, \varepsilon_i} Z_{LAV} = \sum_{i=1}^n \varepsilon_i \quad (3.66)$$

subject to

$$\begin{aligned} y_i - f(\mathbf{x}_i; \boldsymbol{\beta}) &\leq \varepsilon_i, & i = 1, \dots, n, \\ f(\mathbf{x}_i; \boldsymbol{\beta}) - y_i &\leq \varepsilon_i, & i = 1, \dots, n. \end{aligned}$$

In the following sections several non-linear regression models are shown.

3.10.1 A Model for Ultrasonic Data

The data we use here are the result of a NIST study involving ultrasonic calibration. The data consists on 54 observations on two variables. The response variable (y) is ultrasonic response and the predictor variable (x) is metal distance. The data in Table 3.2 were taken from <http://www.itl.nist.gov/div898/strd/nls/data/LINKS/DATA/Chwirut2.dat>. In this Web site, a nonlinear regression model of the form

$$y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) + \varepsilon_i = \frac{e^{-\beta_1 x_i}}{\beta_2 + \beta_3 x_i} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (3.67)$$

was fitted to the data.

The Least Square objective function is:

$$Z_{LS} = \sum_{i=1}^n \left(y_i - \frac{e^{-\beta_1 x_i}}{\beta_2 + \beta_3 x_i} \right)^2$$

which can be solve using any non-linear mathematical programming solver existing in the literature.

The following LS estimates for the ultrasonic data are obtained:

$$Z_{LS} = 513.05, \quad \hat{\beta}_1 = 0.16658, \quad \hat{\beta}_2 = 0.00516, \quad \hat{\beta}_3 = 0.01215.$$

Figure 3.6 shows the least squares model for the Ultrasonic Calibration Data

3.10.2 A Model for Onion Yield Data

The onion yield data for is taken from Ratkowsky (1983), p. 58 (the Uraidla variety). Here y is the onion yield (g/plant) and x is the plant density (plans/m²). Ratkowsky (1983) fit the model

$$\log(y) = -\log(\alpha + \beta x) \quad (3.68)$$

to the data, by least squares.

In this case the objective function is:

$$Z_{LS} = \sum_{i=1}^n (\log y_i + \log(\alpha + \beta x_i))^2,$$

The LS estimates found by Ratkowsky (1988), p. 59, are

$$\hat{\alpha} = 0.003462, \quad \text{and} \quad \hat{\beta} = 0.000129.$$

Table 3.2: Ultrasonic Calibration Data.

Index	y	x	Index	y	x	Index	y	x
1	92.9000	0.500	19	13.12	3.00	37	3.75	5.75
2	57.1000	1.000	20	59.90	0.75	38	11.81	3.00
3	31.0500	1.750	21	14.60	3.00	39	54.70	0.75
4	11.5875	3.750	22	32.90	1.50	40	23.70	2.50
5	8.0250	5.750	23	5.44	6.00	41	11.55	4.00
6	63.6000	0.875	24	12.56	3.00	42	61.30	0.75
7	21.4000	2.250	25	5.44	6.00	43	17.70	2.50
8	14.2500	3.250	26	32.00	1.50	44	8.74	4.00
9	8.4750	5.250	27	13.95	3.00	45	59.20	0.75
10	63.8000	0.750	28	75.80	0.50	46	16.30	2.50
11	26.8000	1.750	29	20.00	2.00	47	8.62	4.00
12	16.4625	2.750	30	10.42	4.00	48	81.00	0.50
13	7.1250	4.750	31	59.50	0.75	49	4.87	6.00
14	67.3000	0.625	32	21.67	2.00	50	14.62	3.00
15	41.0000	1.250	33	8.55	5.00	51	81.70	0.50
16	21.1500	2.250	34	62.00	0.75	52	17.17	2.75
17	8.1750	4.250	35	20.20	2.25	53	81.30	0.50
18	81.5000	0.500	36	7.76	3.75	54	28.90	1.75

The scatter plot of y versus x , with the estimated least square curve, is shown in Figure 3.7. An outlier (observation 38) clearly appears from the graph as an outlier. Because we have only two variables in this case, outliers can be easily spotted upon the inspection of the scatter plot. But in higher dimensions, the detection of outliers becomes very difficult.

Ratkowsky (1988) remarks that observation 38 is suspected outlier and is omitted from further analysis. For the purpose of sensitivity analysis, we include the outlier to see if any of the sensitivity measures will be able to detect it. With the outlier included, we found the LS estimates to be

$$Z_{LS} = 0.912, \quad \hat{\alpha} = 0.003462, \quad \beta = 0.000129,$$

which agree with those found by Ratkowsky (1983).

3.10.3 A Model for Calibration Data

According to Seber and Wild (1989), Tiede and Pagano (1979) fit the calibration model

$$y_i = \alpha + \beta(1 + \gamma x_i^\delta)^{-1} + \varepsilon_i, \quad i = 1, 2, \dots, 14, \quad (3.69)$$

to the set of data in Table 3.4. As can be seen from the scatter plot of y_i versus x_i in Figure 3.8, observation 9 is a clear outlier.

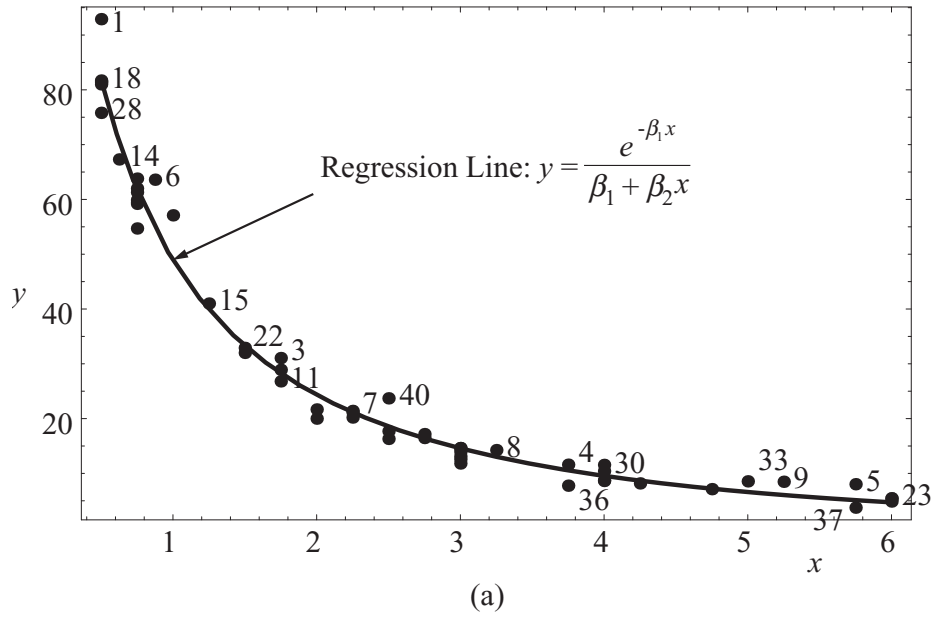


Figure 3.6: Scatter plot of y_i versus x_i , for the Ultrasonic Calibration data set in Table 3.2 and the least squares fitted model.

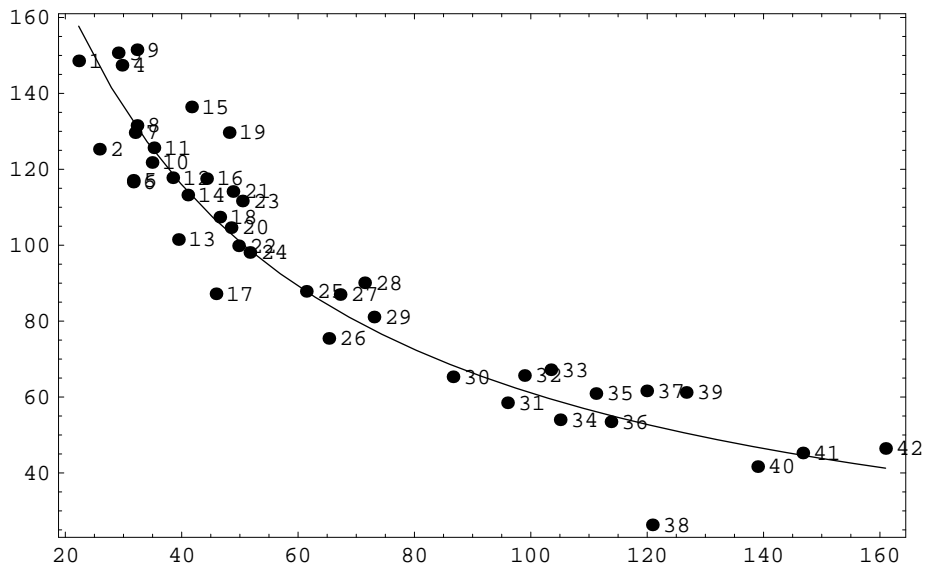


Figure 3.7: Scatter plot of y_i versus x_i , for the Onion Yield Data set in Table 3.2 and the least squares fitted model.

Table 3.3: Onion yield data.

i	x_i	y_i	i	x_i	y_i
1	22.30	148.57	22	49.78	99.85
2	25.86	125.30	23	50.43	111.65
3	29.09	150.69	24	51.72	98.09
4	29.74	147.42	25	61.42	87.85
5	31.68	117.10	26	65.29	75.45
6	31.68	116.64	27	67.23	87.01
7	32.00	129.66	28	71.44	90.10
8	32.32	131.54	29	73.05	81.08
9	32.32	151.50	30	86.63	65.33
10	34.91	121.80	31	96.00	58.49
11	35.23	125.67	32	98.91	65.67
12	38.47	117.78	33	103.44	67.19
13	39.44	101.50	34	105.05	54.01
14	41.05	113.22	35	111.19	60.92
15	41.70	136.43	36	113.78	53.48
16	44.28	117.54	37	119.92	61.62
17	45.90	87.20	38	120.89	26.32
18	46.55	107.41	39	126.71	61.21
19	48.16	129.68	40	138.99	41.67
20	48.49	104.63	41	146.75	45.26
21	48.81	114.15	42	160.97	46.45

Table 3.4: Calibration data.

i	x_i	y_i	i	x_i	y_i
1	0	7.720	8	10	3.208
2	0	8.113	9	20	4.478
3	2	6.664	10	20	2.396
4	2	6.801	11	50	1.302
5	5	4.994	12	50	1.377
6	5	4.948	13	100	1.025
7	10	3.410	14	100	1.096

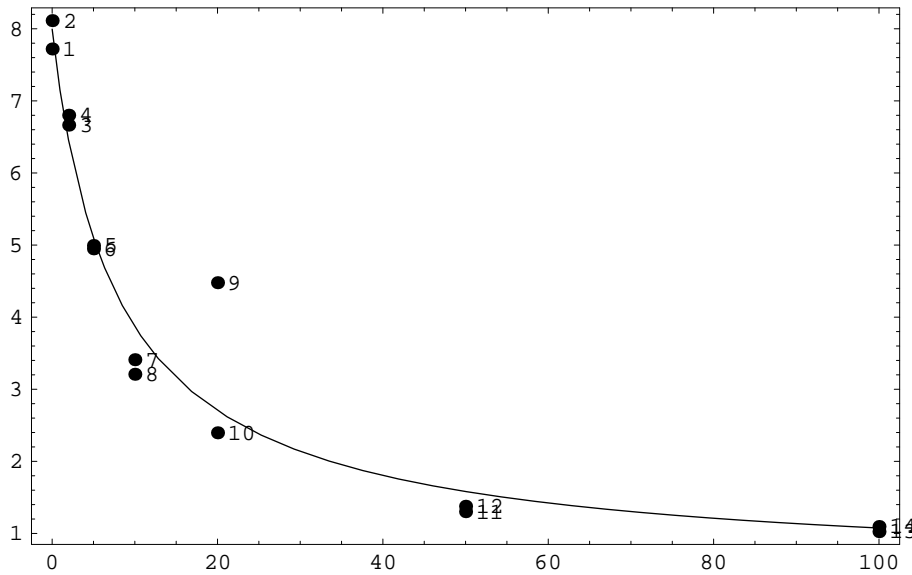


Figure 3.8: Least squares model for the Calibration Data showing data and least squares regression line.

In this case the objective function is:

$$Z_{LS} = \sum_{i=1}^n \left(y_i - \alpha - \beta(1 + \gamma x_i^\delta)^{-1} \right)^2,$$

which derives in the following LS estimates of the calibration data are obtained:

$$Z_{LS} = 4.31, \hat{\alpha} = 0.442, \hat{\beta} = 7.554, \hat{\gamma} = 0.133, \hat{\delta} = 0.957.$$

Figure 3.8 shows the least squares model for the Calibration Data.

3.11 Regression Sensitivity Analysis

It has long been recognized that statistical conclusions drawn from an analysis can be sensitive to changes in a model, deviations from assumptions, or other perturbations in the inputs of a statistical analysis (e.g., observations in a data set). Actually, more often than not, conclusions drawn from an analysis can be completely turned around if one or two data points are changed.

In other words, one needs to know the influence of each data item (explanatory or response variables), weight, model, etc. on the final results so as to make the adequate corrections when necessary. It is therefore essential for data analysts to be able to assess the sensitivity of their conclusions to various perturbations in the inputs. Sensitivity analysis provides confidence in the model and the conclusions from an analysis because it allows the analyst to assess the effects of departures from the assumptions made and the data values,

detect outliers or wrong data values, define testing strategies, optimize resources, reduce costs, and avoid unexpected problems.

Sensitivity analysis has seen a great surge of research activities over the last three decades. Most of the work, however, focused almost exclusively on least squares linear regression. An evidence supporting this observation is the number of books and articles that have appeared in the literature. See, for example, the books by Hawkins (1980), Belsley, Kuh, and Welsch (1980), Cook and Weisberg, (1982), Atkinson, (1985), Chatterjee and Hadi (1988), and Barnett and Lewis (1994), and the articles by Pregibon (1981), Gray and Ling, (1984), Gray (1986), Cook(1986), Jones and Ling (1988), Weissfeld and Schneider (1990a, 1990b), Schwarzmann (1991), Paul and Fung (1991), Simonoff (1991), Nyquist (1992), Hadi and Simonoff (1993), Atkinson (1994), Peña and Yohai (1995), Barrett and Gray (1997), and Mayo and Gray (1997), Billor, Chatterjee, and Hadi (2001), Winsnowski, Montgomery, and James (2001), Castillo and Kjaerulff (2003) and Castillo et al. (2003,2004a,2004b). By contrast to the area of least squares linear regression, sensitivity analysis in nonlinear regression has received very little attention.

3.11.1 Background

Consider the following general nonlinear *primal problem* (P):

$$\underset{\boldsymbol{\beta}}{\text{Minimize}} \quad Z_P = f(\boldsymbol{\beta}; \mathbf{z}) \quad (3.70)$$

subject to

$$\mathbf{h}(\boldsymbol{\beta}; \mathbf{z}) = \mathbf{0} : \boldsymbol{\lambda} \quad (3.71)$$

$$\mathbf{g}(\boldsymbol{\beta}; \mathbf{z}) \leq \mathbf{0} : \boldsymbol{\mu}, \quad (3.72)$$

where boldfaced letters refer to vectors, $\boldsymbol{\beta} \in \mathbb{R}^n$, $\mathbf{z} \in \mathbb{R}^p$, $\mathbf{h}(\boldsymbol{\beta}; \mathbf{z}) \in \mathbb{R}^\ell$ and $\mathbf{g}(\boldsymbol{\beta}; \mathbf{z}) \in \mathbb{R}^m$, and $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are the dual variables to be introduced below.

Every primal nonlinear programming problem P of the form (3.70)–(3.72), has an associated dual problem D , which is defined as:

$$\underset{\boldsymbol{\lambda}, \boldsymbol{\mu}}{\text{Maximize}} \quad Z_D = \text{Inf}_{\boldsymbol{\beta}} \{ \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{z}) \} \quad (3.73)$$

subject to

$$\boldsymbol{\mu} \geq \mathbf{0}, \quad (3.74)$$

where

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{z}) = f(\boldsymbol{\beta}; \mathbf{z}) + \boldsymbol{\lambda}^T \mathbf{h}(\boldsymbol{\beta}; \mathbf{z}) + \boldsymbol{\mu}^T \mathbf{g}(\boldsymbol{\beta}; \mathbf{z}), \quad (3.75)$$

is the Lagrangian function associated with the primal problem (3.70)–(3.72), and $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are called dual variables and they are vectors of dimensions ℓ and m , the number of equalities and inequalities in the primal problem, respectively.

We assume here that all the inequalities are active and the corresponding μ multipliers are non-null. Note that after solving problem (3.70)–(3.72), it is very easy to check which of the inequalities in (3.72) are active, and then we can ignore the inactive constraints because

they do not alter the optimal solution. Now, let $\mathbf{q}(\boldsymbol{\beta}; \mathbf{z}) = \mathbf{g}(\boldsymbol{\beta}; \mathbf{z}) \cup \mathbf{h}(\boldsymbol{\beta}; \mathbf{z})$, that is, the set of all equality constraints in (3.71) and the inequality constraints in (3.72). Also, let $\boldsymbol{\eta}$ be the corresponding $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. Castillo et al. (2005) show that, if the problem is regular, i.e., the gradients of the active constraints are linearly independent at the optimal point, and the Lagrange multipliers of the active inequality constraints are non-null, the sensitivity of the optimal solution $(\boldsymbol{\beta}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, Z^*)$ of (3.70)–(3.72) to changes in the parameters can be determined by the system of equations:

$$\begin{bmatrix} \mathbf{F}_{\boldsymbol{\beta}} & | & \mathbf{F}_{\mathbf{z}} & | & \mathbf{0} & | & -1 \\ \hline \mathbf{F}_{\boldsymbol{\beta}\boldsymbol{\beta}} & | & \mathbf{F}_{\boldsymbol{\beta}\mathbf{z}} & | & \mathbf{Q}_{\boldsymbol{\beta}}^T & | & 0 \\ \hline \mathbf{Q}_{\boldsymbol{\beta}} & | & \mathbf{Q}_{\mathbf{z}} & | & \mathbf{0} & | & 0 \end{bmatrix} \begin{bmatrix} d\boldsymbol{\beta} \\ dz \\ d\boldsymbol{\eta} \\ dZ \end{bmatrix} = \mathbf{0}, \quad (3.76)$$

where $d\boldsymbol{\beta}$, dz , $d\boldsymbol{\eta}$ and dZ are the differential perturbations, and all the matrices in (8.66) are evaluated at the optimal solution and are defined below (with the corresponding dimensions in parenthesis):

$$\mathbf{F}_{\boldsymbol{\beta}}_{(1 \times n)} = \left(\nabla_{\boldsymbol{\beta}} f(\boldsymbol{\beta}^*, \mathbf{z}) \right)^T, \quad (3.77)$$

$$\mathbf{F}_{\mathbf{z}}_{(1 \times p)} = \left(\nabla_{\mathbf{z}} f(\boldsymbol{\beta}^*, \mathbf{z}) \right)^T, \quad (3.78)$$

$$\mathbf{F}_{\boldsymbol{\beta}\boldsymbol{\beta}}_{(n \times n)} = \nabla_{\boldsymbol{\beta}\boldsymbol{\beta}} f(\boldsymbol{\beta}^*, \mathbf{z}) + \sum_{k=1}^{\ell+m} \eta_k^* \nabla_{\boldsymbol{\beta}\boldsymbol{\beta}} q_k(\boldsymbol{\beta}^*, \mathbf{z}), \quad (3.79)$$

$$\mathbf{F}_{\boldsymbol{\beta}\mathbf{z}}_{(n \times p)} = \nabla_{\boldsymbol{\beta}\mathbf{z}} f(\boldsymbol{\beta}^*, \mathbf{z}) + \sum_{k=1}^{\ell+m} \eta_k^* \nabla_{\boldsymbol{\beta}\mathbf{z}} q_k(\boldsymbol{\beta}^*, \mathbf{z}), \quad (3.80)$$

$$\mathbf{Q}_{\boldsymbol{\beta}}_{((\ell+m) \times n)} = \left(\nabla_{\boldsymbol{\beta}} \mathbf{q}(\boldsymbol{\beta}^*, \mathbf{z}) \right)^T, \quad (3.81)$$

$$\mathbf{Q}_{\mathbf{z}}_{((\ell+m) \times p)} = \left(\nabla_{\mathbf{z}} \mathbf{q}(\boldsymbol{\beta}^*, \mathbf{z}) \right)^T. \quad (3.82)$$

Condition (8.66) can be written as

$$\mathbf{U} \begin{bmatrix} d\boldsymbol{\beta} \\ d\boldsymbol{\eta} \\ dZ \end{bmatrix} = \mathbf{S} dz, \quad (3.83)$$

where

$$\mathbf{U} = \begin{bmatrix} \mathbf{F}_{\boldsymbol{\beta}} & | & \mathbf{0} & | & -1 \\ \hline \mathbf{F}_{\boldsymbol{\beta}\boldsymbol{\beta}} & | & \mathbf{Q}_{\boldsymbol{\beta}}^T & | & 0 \\ \hline \mathbf{Q}_{\boldsymbol{\beta}} & | & \mathbf{0} & | & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{S} = - \begin{bmatrix} \mathbf{F}_{\mathbf{z}} \\ \mathbf{F}_{\boldsymbol{\beta}\mathbf{z}} \\ \mathbf{Q}_{\mathbf{z}} \end{bmatrix}. \quad (3.84)$$

Thus, given a unit direction vector dz , we can solve (3.83) to obtain the corresponding directional or partial derivatives. If the solution exists and is unique, we obtain the corresponding directional derivatives. For a partial derivative to exist the corresponding

directional derivatives must exist. If the system (3.83) has no solution, the corresponding directional derivative does not exist.

Three cases are of special interest:

1. If the optimization problem has no constraints, the system (3.83) leads to

$$\begin{aligned}\mathbf{F}_\beta d\beta - dZ &= -\mathbf{F}_z dz \\ \mathbf{F}_{\beta\beta} d\beta &= -\mathbf{F}_{\beta z} dz\end{aligned}$$

which if $\mathbf{F}_{\beta\beta}$ is invertible gives

$$\frac{\partial\beta}{\partial\mathbf{z}}_{(n \times p)} = -\mathbf{F}_{\beta\beta}^{-1} \mathbf{F}_{\beta z}, \quad (3.85)$$

$$\frac{\partial Z}{\partial\mathbf{z}}_{(1 \times p)} = -\mathbf{F}_\beta \mathbf{F}_{\beta\beta}^{-1} \mathbf{F}_{\beta z} + \mathbf{F}_z = \mathbf{F}_z, \quad (3.86)$$

because $\mathbf{F}_\beta = 0$.

2. If the matrix \mathbf{Q}_β is invertible, the solution of the system (3.83) becomes

$$\frac{\partial\beta}{\partial\mathbf{z}}_{(n \times p)} = -\mathbf{Q}_\beta^{-1} \mathbf{Q}_z, \quad (3.87)$$

$$\frac{\partial\eta}{\partial\mathbf{z}}_{(n \times p)} = \left(\mathbf{Q}_\beta^{-1}\right)^T \left(\mathbf{F}_{\beta\beta} \mathbf{Q}_\beta^{-1} \mathbf{Q}_z - \mathbf{F}_{\beta z}\right), \quad (3.88)$$

$$\frac{\partial Z}{\partial\mathbf{z}}_{(1 \times p)} = -\mathbf{F}_\beta \mathbf{Q}_\beta^{-1} \mathbf{Q}_z + \mathbf{F}_z = \mathbf{F}_z + \eta^T \mathbf{Q}_z. \quad (3.89)$$

3. If the matrix \mathbf{U} is invertible, we have

$$\begin{pmatrix} \frac{\partial\beta}{\partial\mathbf{z}} \\ \frac{\partial\eta}{\partial\mathbf{z}} \\ \frac{\partial Z}{\partial\mathbf{z}} \end{pmatrix} = \mathbf{U}^{-1} \mathbf{S}. \quad (3.90)$$

The formulas developed above are applicable to any optimization problem. Castillo et al. (2004) use these results to obtain the sensitivity of the objective function with respect to changes in the data. In this paper, we show that the results can also be used to derive the sensitivity of the estimated parameters with respect to changes in the data. We apply them in Sections 3.10.2–3.13 to three regression estimation problems: the least squares (LS), minimax (MM) and least absolute value (LAV).

3.11.2 Least Squares Sensitivity Analysis

The least squares method leads to the following optimization problem

$$\text{Minimize}_{\beta} Z_{LS} = \sum_{i=1}^n (y_i - \mathbf{x}_i^T \beta)^2. \quad (3.91)$$

Since the optimization problem in (3.91) does not have any constraints, the system in (8.66) can be solved by replacing

$$\mathbf{F}_{\beta} \beta_{(k \times k)} = \nabla_{\beta} f(\bar{\beta}, z) = 2\mathbf{X}^T \mathbf{X}, \quad (3.92)$$

$$\mathbf{F}_{\beta z} \beta z_{(k \times n(k+1))} = \nabla_{\beta z} f(\bar{\beta}, z) = 2(-\mathbf{X}^T | \beta^T \otimes \mathbf{X}^T - \mathbf{I}_k \otimes \mathbf{e}^T), \quad (3.93)$$

$$\mathbf{F}_{\beta} \beta_{(1 \times k)} = \left(\nabla_{\beta} f(\bar{\beta}, z) \right)^T = -2\mathbf{e}^T \mathbf{X} = \mathbf{0}, \quad (3.94)$$

$$\mathbf{F}_{z} z_{(1 \times (k+1))} = \left(\nabla_z f(\bar{\beta}, z) \right)^T = 2(\mathbf{e}^T | -\mathbf{e}^T \otimes \beta^T), \quad (3.95)$$

where \mathbf{e} is the vector of errors, and it has been taken into account that for the LS method $\mathbf{e}^T \mathbf{X} = \mathbf{0}$, in (3.85) and (3.86), leading to

$$\frac{\partial \beta}{\partial \mathbf{y}_{k \times n}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T, \quad (3.96)$$

$$\frac{\partial \beta}{\partial \mathbf{x}_{k \times nk}} = -(\mathbf{X}^T \mathbf{X})^{-1} (\beta^T \otimes \mathbf{X}^T - \mathbf{I}_k \otimes \mathbf{e}^T), \quad (3.97)$$

$$\frac{\partial Z_{LS}^*}{\partial \mathbf{y}_{1 \times n}} = 2\mathbf{e}^T, \quad (3.98)$$

$$\frac{\partial Z_{LS}^*}{\partial \mathbf{x}_{1 \times nk}} = -2(\mathbf{e}^T \otimes \beta^T). \quad (3.99)$$

Note that the matrix with the sensitivities of the β with respect to \mathbf{y} is known in the outlier detection literature as the *catcher* matrix.

From (3.96) to (3.99) one immediately obtains the following formulas for the sensitivities:

$$\frac{\partial \beta_j}{\partial y_i} = \sum_{r=1}^k c_{jr} x_{ir}, \quad (3.100)$$

$$\frac{\partial \beta_j}{\partial x_{st}} = -\sum_{r=1}^k c_{jr} [\beta_t x_{sr} - \delta_{tr} e_s], \quad (3.101)$$

$$\frac{\partial Z_{LS}^*}{\partial y_i} = 2e_i, \quad (3.102)$$

$$\frac{\partial Z_{LS}^*}{\partial x_{st}} = -2e_s \beta_t, \quad (3.103)$$

where c_{ij} are the elements of matrix $(\mathbf{X}^T \mathbf{X})^{-1}$, δ_{tr} is the Kronecker delta function, and e_i is the i th residual. To be meaningfully interpreted and comparable, we use the standardized versions of the above sensitivities by subtracting their means and dividing by their standard deviations. For example, the standardized version of (3.102) becomes

$$S_{LS}(y_i) = \frac{e_i}{\hat{\sigma} \sqrt{1 - p_{ii}}}, \quad (3.104)$$

where p_{ii} is the i th leverage value (the i th diagonal element of $\mathbf{P} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$) and

$$\hat{\sigma}^2 = \frac{\mathbf{e}^T \mathbf{e}}{n - k}. \quad (3.105)$$

Similarly, the standardized version of (3.103) becomes

$$S_{LS}(x_{ij}) = \frac{e_i \hat{\beta}_j}{\hat{\sigma} \sqrt{(1 - p_{ii}) [\hat{\sigma}^2 c_{jj} + \hat{\beta}_j^2]}}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, k, \quad (3.106)$$

where c_{jj} is the j th diagonal element of $(\mathbf{X}^T \mathbf{X})^{-1}$.

It is interesting to note that these standardized sensitivities possess the following properties:

1. Apart from degenerate cases, all the standardized sensitivities with respect to all data points are different from zero.
2. The standardized sensitivities $\frac{\partial Z_{LS}^*}{\partial y_i}$ and $\frac{\partial Z_{LS}^*}{\partial x_{ij}}$ coincide in absolute value, but the sign of the second depends on the sign of the corresponding β_j .
3. The standardized sensitivities $\frac{\partial \beta_s}{\partial y_i}$ and $\frac{\partial \beta_r}{\partial y_i}$ coincide in absolute value, but their signs depend on the point positions.
4. The standardized sensitivities $\frac{\partial \beta_s}{\partial x_{ij}}$ and $\frac{\partial \beta_r}{\partial x_{ij}}$ coincide in absolute value, but their signs depend on the point positions.

3.11.3 A Numerical Example

Now, fitting a regression line by least squares to the star cluster data including giant stars, one gets the model (see Figure 3.1):

$$y = 6.979 - 0.455x + \varepsilon.$$

Note here the effects of the four giant stars on the least squares regression line. The estimated line has a negative slope, which is contrary to the expectation of the relationship between light and temperature.

Using (3.100)–(3.103) the objective function and the parameters sensitivities with respect to the data $\mathbf{z} = (\mathbf{y}, \mathbf{x})$ have been obtained. Table 3.5 shows the standardized sensitivities of Z_{LS} , β_0 , β_1 with respect to data. It can be observed that the properties of these sensitivities mentioned above hold in this table.

Figure 3.9 shows the scatter plot of the star cluster data, where the points are sorted by their objective function sensitivities (upper graph) and by the slope or intercept sensitivities (lower graph). Thus, the higher the number next to a point the more sensitive the results with respect to changes in the data point.

The sensitivity contours in these plots have been obtained as follows. A new data point (x_{n+1}, y_{n+1}) has been assumed to enter the sample and then the sensitivities associated with this point have been re-calculated as a function of its coordinates. In this way, these contours permit determining the sensitivity of a new point entering the sample, or, approximately, determining the sensitivity of any point already existing in the sample. Notice that the closer the point to the regression line the lower the objective function sensitivity. Note also

Table 3.5: Standardized Sensitivities of Z , β_0, β_1 With Respect to Data.

Index	Z_{LS}			β_0			β_1		
	Y	X	(X,Y)	Y	X	(X,Y)	Y	X	(X,Y)
1	0.43	0.43	0.608	-0.209	-0.466	0.511	0.209	0.466	0.511
2	1.495	1.495	2.114	-0.869	-1.653	1.868	0.869	1.653	1.868
3	-0.195	-0.195	0.276	0.174	0.229	0.288	-0.174	-0.229	0.288
4	1.495	1.495	2.114	-0.869	-1.653	1.868	0.869	1.653	1.868
5	0.302	0.302	0.427	0.035	-0.286	0.288	-0.035	0.286	0.288
6	0.914	0.914	1.292	-0.521	-1.008	1.135	0.521	1.008	1.135
7	-1.036	-1.036	1.465	1.634	1.381	2.139	-1.634	-1.381	2.139
8	0.664	0.664	0.939	-0.904	-0.852	1.242	0.904	0.852	1.242
9	0.947	0.947	1.34	0.174	-0.883	0.9	-0.174	0.883	0.9
10	0.234	0.234	0.331	-0.209	-0.275	0.345	0.209	0.275	0.345
11	0.607	0.607	0.859	2.85	0.058	2.851	-2.85	-0.058	2.851
12	0.871	0.871	1.232	-0.417	-0.944	1.032	0.417	0.944	1.032
13	0.859	0.859	1.214	-0.591	-0.971	1.136	0.591	0.971	1.136
14	-1.969	-1.969	2.784	1.043	2.154	2.393	-1.043	-2.154	2.393
15	-1.366	-1.366	1.932	0.07	1.346	1.348	-0.07	-1.346	1.348
16	-0.689	-0.689	0.975	-0.382	0.584	0.698	0.382	-0.584	0.698
17	-1.986	-1.986	2.809	0.278	1.997	2.017	-0.278	-1.997	2.017
18	-1.403	-1.403	1.985	-0.382	1.279	1.335	0.382	-1.279	1.335
19	-1.558	-1.558	2.203	0.278	1.58	1.604	-0.278	-1.58	1.604
20	0.893	0.893	1.262	2.85	-0.22	2.859	-2.85	0.22	2.859
21	-1.152	-1.152	1.629	0.07	1.138	1.14	-0.07	-1.138	1.14
22	-1.438	-1.438	2.033	0.07	1.416	1.417	-0.07	-1.416	1.417
23	-0.975	-0.975	1.379	-0.382	0.862	0.943	0.382	-0.862	0.943
24	-0.151	-0.151	0.213	-0.626	0.004	0.626	0.626	-0.004	0.626
25	0.063	0.063	0.09	-0.243	-0.117	0.27	0.243	0.117	0.27
26	-0.547	-0.547	0.773	-0.382	0.445	0.587	0.382	-0.445	0.587
27	-0.652	-0.652	0.923	0.07	0.651	0.655	-0.07	-0.651	0.655
28	-0.151	-0.151	0.213	-0.243	0.091	0.26	0.243	-0.091	0.26
29	-1.191	-1.191	1.685	0.313	1.231	1.27	-0.313	-1.231	1.27
30	1.17	1.17	1.655	2.885	-0.482	2.925	-2.885	0.482	2.925
31	-1.008	-1.008	1.425	-0.243	0.926	0.957	0.243	-0.926	0.957
32	0.352	0.352	0.498	-0.869	-0.541	1.024	0.869	0.541	1.024
33	0.477	0.477	0.675	-0.487	-0.575	0.754	0.487	0.575	0.754
34	1.964	1.964	2.777	2.85	-1.263	3.117	-2.85	1.263	3.117
35	-1.272	-1.272	1.799	0.278	1.302	1.331	-0.278	-1.302	1.331
36	1.329	1.329	1.88	-1.077	-1.54	1.879	1.077	1.54	1.879
37	0.328	0.328	0.464	-0.765	-0.494	0.91	0.765	0.494	0.91
38	0.477	0.477	0.675	-0.487	-0.575	0.754	0.487	0.575	0.754
39	0.471	0.471	0.666	-0.765	-0.633	0.992	0.765	0.633	0.992
40	1.086	1.086	1.535	-0.417	-1.152	1.225	0.417	1.152	1.225
41	-0.651	-0.651	0.92	-0.243	0.578	0.627	0.243	-0.578	0.627
42	0.192	0.192	0.271	-0.487	-0.297	0.57	0.487	0.297	0.57
43	0.732	0.732	1.035	-0.66	-0.863	1.087	0.66	0.863	1.087
44	0.691	0.691	0.978	-0.487	-0.784	0.923	0.487	0.784	0.923
45	1.13	1.13	1.598	-0.834	-1.29	1.536	0.834	1.29	1.536
46	0.049	0.049	0.069	-0.487	-0.158	0.512	0.487	0.158	0.512
47	-0.832	-0.832	1.177	-0.382	0.723	0.818	0.382	-0.723	0.818

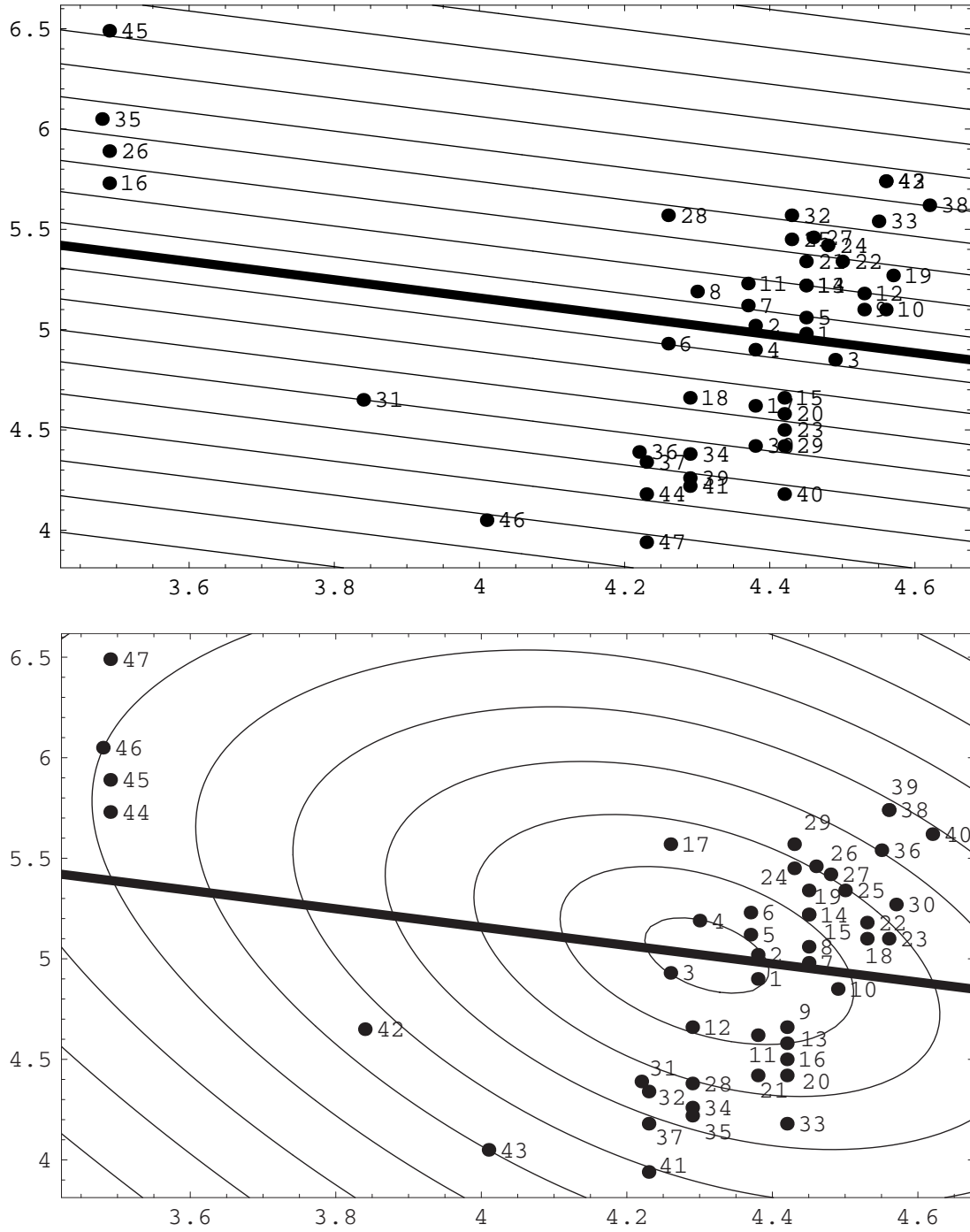


Figure 3.9: Scatter plot of the star cluster data with the sensitivity contours. The number next to a point refers to the rank of the point according its sensitivity with respect the objective function (upper graph) and the slope parameter (lower graph).

that the closer the points to the center of gravity the smaller the sensitivity with respect to the beta parameters.

The most interesting revelation of sensitivity analysis can be seen in Figure 3.9. The upper graph shows that only one of the four giant stars exert undue sensitivity on the objective function estimates. However, the lower graph shows that the four giant stars are the ones with the greatest sensitivities on the parameters.

For a comparison purpose with existing influence measures, Figure 3.10 shows the index plot of the internally studentized residual, Cook's distance (cook, 1977), and the sensitivities in the last three columns in Table 3.5. It can be seen from these graphs that the studentized residuals fail to detect any of the four giant starts. Cook's distance nominates at most three of the four stars. Existing sensitivity methods based on least squares fails to detect the influence of the four stars on the least squares result. This is due to the well-known problem of masking (the least squares estimates are not-robust). On the contrary, the four stars are clearly separated from the rest of the points in the index plot of the Y - and XY -sensitivities. Thus, revealing the superiority of the proposed method with respect to the existing ones. We should note here that the proposed sensitivity measures have been tested with other data sets and have shown similar performance, but the results are not reported here because of lack of space.

3.12 Minimax Regression

3.12.1 The Primal Minimax Regression Problem

The minimax method estimates the regression coefficient by minimizing the maximum error, that is,

$$\underset{\boldsymbol{\beta}}{\text{Minimize}} \quad Z_{MM} = \max_i |y_i - \mathbf{x}_i^T \boldsymbol{\beta}|, \quad (3.107)$$

which is equivalent to the linear programming problem:

$$\underset{\boldsymbol{\beta}, \varepsilon}{\text{Minimize}} \quad Z_{MM} = \varepsilon \quad (3.108)$$

subject to

$$y_i - \mathbf{x}_i^T \boldsymbol{\beta} \leq \varepsilon : \mu_i^{(1)}, \quad i = 1, \dots, n, \quad (3.109)$$

$$\mathbf{x}_i^T \boldsymbol{\beta} - y_i \leq \varepsilon : \mu_i^{(2)}, \quad i = 1, \dots, n, \quad (3.110)$$

where $\mu_i^{(1)}$ and $\mu_i^{(2)}$ are the dual variables.

To obtain the sensitivities of the β estimates with respect to data it is convenient to assume that we are not in a degenerate case, i.e., we assume that a total of exactly k constraints in (3.109) and (3.110) are active. It is also convenient to reduce the analysis to the sensitivities that are known to be non-null. Then, following the steps in Section 3.11.1, we have

$$\mathbf{q}_i(\bar{\mathbf{x}}, \mathbf{z}) = y_i - \mathbf{x}_i^T \boldsymbol{\beta} - \varepsilon = 0, \quad : \eta_i^+ \quad i \in I^+, \quad (3.111)$$

$$\mathbf{q}_i(\bar{\mathbf{x}}, \mathbf{z}) = \mathbf{x}_i^T \boldsymbol{\beta} - y_i - \varepsilon = 0, \quad : \eta_i^-, \quad i \in I^-, \quad (3.112)$$

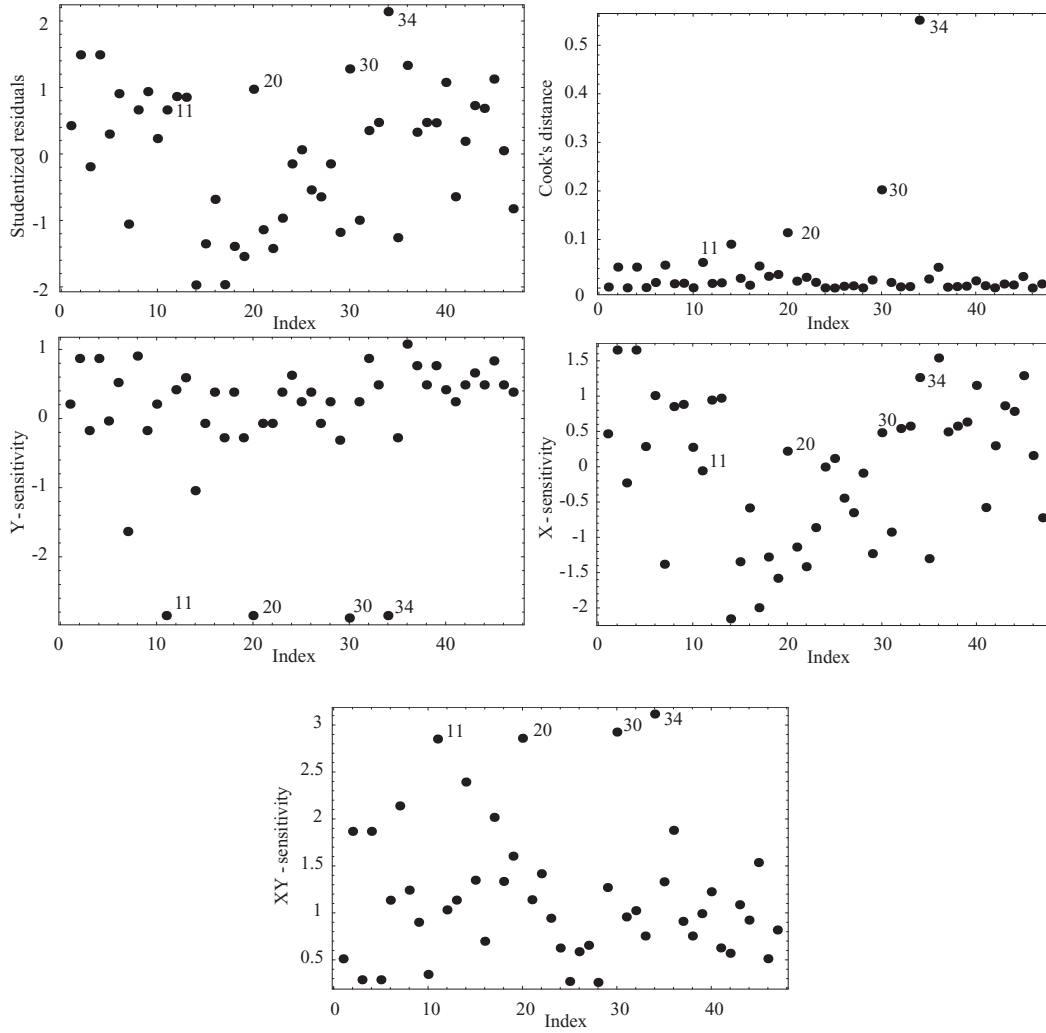


Figure 3.10: Index plot of the studentized residuals, Cook's distance, and the sensitivities in the last three columns in Table 3.5.

i.e., the sets I^+ and I^- , with cardinals p^+ and p^- , respectively, give the data points that correspond to the active constraints. Note that $\boldsymbol{\eta}^+$ and $\boldsymbol{\eta}^-$ are the column vectors of the dual variables associated with the sets I^+ and I^- , respectively. We define the ordered set L as $L = I^+ \cap I^-$. Apart from degenerate cases, we have $p^+ + p^- = k + 1$.

To obtain the matrices $\mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta}$ to $\mathbf{F}\mathbf{z}$ in Section 3.11.1, we need the following derivatives

$$\frac{\partial q_i(\boldsymbol{\beta}_r; \mathbf{y}, \mathbf{x})}{\partial \beta_r} = \begin{cases} -x_{ir}, & i \in I^+, \\ x_{ir}, & i \in I^-, \end{cases} \quad (3.113)$$

$$\frac{\partial q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial y_s} = \begin{cases} 1, & i \in I^+, \quad s = i, \\ -1, & i \in I^-, \quad s = i, \\ 0, & \text{otherwise,} \end{cases} \quad (3.114)$$

$$\frac{\partial q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial x_{rs}} = \begin{cases} -\beta_s, & i \in I^+, \quad r = i, \\ \beta_s, & i \in I^-, \quad r = i, \\ 0, & \text{otherwise,} \end{cases} \quad (3.115)$$

$$\frac{\partial q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \varepsilon} = \begin{cases} -1, & i \in I^+ \cup I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.116)$$

$$\frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \beta_r \partial x_{pq}} = \begin{cases} -1, & i \in I^+, \quad p = i, \quad q = r, \\ 1, & i \in I^-, \quad p = i, \quad q = r, \\ 0, & \text{otherwise,} \end{cases} \quad (3.117)$$

$$\frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \beta_r \partial \beta_s} = 0, \quad (3.118)$$

$$\frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \beta_r \partial y_s} = \frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial y_i \partial x_{rs}} = \frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \varepsilon \partial x_{rs}} = \frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \varepsilon \partial y_r} = 0, \quad (3.119)$$

$$\frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \varepsilon \partial \beta_r} = 0, \quad (3.120)$$

and all derivatives of $f(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})$ with respect to $\boldsymbol{\beta}$, \mathbf{y} and \mathbf{x} are null.

Note that it is not necessary to calculate the derivatives with respect to ε because the objective function is precisely ε .

Let \mathbf{X}^+ and \mathbf{X}^- denote the submatrices of \mathbf{X} which contain the rows whose numbers are in I^+ and I^- , respectively. Then, we have

$$\mathbf{F}\boldsymbol{\beta}\mathbf{z} = \begin{pmatrix} \mathbf{0}_{k \times (k+1)} & | & -\boldsymbol{\eta}^{+T} \otimes \mathbf{I}_k & | & \boldsymbol{\eta}^{-T} \otimes \mathbf{I}_k \\ \text{-----} & + & \text{-----} & + & \text{-----} \\ \mathbf{0}_{1 \times (k+1)} & | & \mathbf{0}_{1 \times kp^+} & | & \mathbf{0}_{1 \times kp^-} \end{pmatrix}_{(k+1) \times (k+1)(k+1)}, \quad (3.121)$$

$$\mathbf{Q}\boldsymbol{\beta} = \begin{pmatrix} -\mathbf{X}^+ & | & -\mathbf{1}_{p^+ \times 1} \\ \text{-----} & + & \text{-----} \\ \mathbf{X}^- & | & -\mathbf{1}_{p^- \times 1} \end{pmatrix}_{(k+1) \times (k+1)}, \quad (3.122)$$

$$\mathbf{Q}\mathbf{z} = \begin{pmatrix} \mathbf{I}_{p^+} & | & \mathbf{0}_{p^+ \times p^-} & | & -\mathbf{I}_{p^+} \otimes \boldsymbol{\beta}^T & | & \mathbf{0}_{p^+ \times kp^-} \\ \text{-----} & + & \text{-----} & + & \text{-----} & + & \text{-----} \\ \mathbf{0}_{p^- \times p^+} & | & -\mathbf{I}_{p^-} & | & \mathbf{0}_{p^- \times kp^+} & | & \mathbf{I}_{p^-} \otimes \boldsymbol{\beta}^T \end{pmatrix}_{(k+1) \times (k+1)(k+1)} \quad (3.123)$$

$$\mathbf{F}\boldsymbol{\beta} = (\mathbf{0}_{1 \times k} \quad | \quad 1)_{1 \times (k+1)}, \quad (3.124)$$

$$\mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta} = \mathbf{0}_{(k+1) \times (k+1)}, \quad (3.125)$$

$$\mathbf{F}\mathbf{z} = \mathbf{0}_{1 \times (k+1)(k+1)}. \quad (3.126)$$

Once these matrices are known, one can obtain all the sensitivities using (3.87) to (3.89), that is:

$$\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}_{(n \times p)}} = -\mathbf{Q}_{\boldsymbol{\beta}}^{-1} \mathbf{Q} \mathbf{z}, \quad (3.127)$$

$$\frac{\partial \boldsymbol{\eta}}{\partial \mathbf{z}_{(n \times p)}} = -(\mathbf{Q}_{\boldsymbol{\beta}}^{-1})^T \mathbf{F} \boldsymbol{\beta} \mathbf{z}, \quad (3.128)$$

$$\frac{\partial Z}{\partial \mathbf{z}_{(1 \times p)}} = -\mathbf{F} \boldsymbol{\beta} \mathbf{Q}_{\boldsymbol{\beta}}^{-1} \mathbf{Q} \mathbf{z}. \quad (3.129)$$

Since the first $k + 1$ columns of matrix $\mathbf{F} \boldsymbol{\beta} \mathbf{z}$ vanish, Expression (3.128) proves that the sensitivities of the dual variables with respect to the \mathbf{y} are null.

From (3.121)–(3.126) one immediately obtains the following formulas for the sensitivities:

$$\frac{\partial \beta_j}{\partial y_i} = \begin{cases} -\rho_{j\ell(i)}, & i \in I^+, \\ \rho_{j\ell(i)}, & i \in I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.130)$$

$$\frac{\partial \beta_j}{\partial x_{st}} = \begin{cases} \rho_{j\ell(s)} \beta_t, & s \in I^+, \\ -\rho_{j\ell(s)} \beta_t, & s \in I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.131)$$

$$\frac{\partial \eta_j}{\partial y_i} = 0, \quad (3.132)$$

$$\frac{\partial \eta_j}{\partial x_{st}} = \begin{cases} \rho_{tj} \eta_{\ell(s)}, & s \in I^+, \\ -\rho_{tj} \eta_{\ell(s)}, & s \in I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.133)$$

$$\frac{\partial Z_{MM}^*}{\partial y_i} = \begin{cases} -\rho_{k+1,\ell(i)}, & i \in I^+, \\ \rho_{k+1,\ell(i)}, & i \in I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.134)$$

$$\frac{\partial Z_{MM}^*}{\partial x_{st}} = \begin{cases} \rho_{k+1,\ell(s)} \beta_t, & s \in I^+, \\ -\rho_{k+1,\ell(s)} \beta_t, & s \in I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.135)$$

where ρ_{ij} are the elements of $\mathbf{Q}_{\boldsymbol{\beta}}^{-1}$, and the $\ell(i)$ refers to the order of i in the set L .

Since at most one of $\mu_i^{(1)}$ and $\mu_i^{(2)}$ are non-null for each value of i , these sensitivities are null, for those data points such that the constraints (3.109) and (3.110) are inactive, and different from zero otherwise. Note also that the sensitivities are proportional to the corresponding regression coefficient β_j and to the dual variable $\mu_i^{(s)}$ value.

The standardized sensitivities of the MM objective function with respect to the response variable in (3.134) values are,

$$S_{MM}(y_i) = \frac{(\partial Z_{MM}^* / \partial y_i) - m}{s}, \quad i = 1, 2, \dots, n, \quad (3.136)$$

where m and s are the mean and standard deviation of $\partial Z_{MM}^* / \partial y_i, i = 1, 2, \dots, n$.

Similarly, the standardized sensitivities with respect to the predictor variables in (3.135) are

$$S_{MM}(x_{ij}) = \frac{(\partial Z_{MM}^* / \partial x_{ij}) - m_j}{s_j}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, k, \quad (3.137)$$

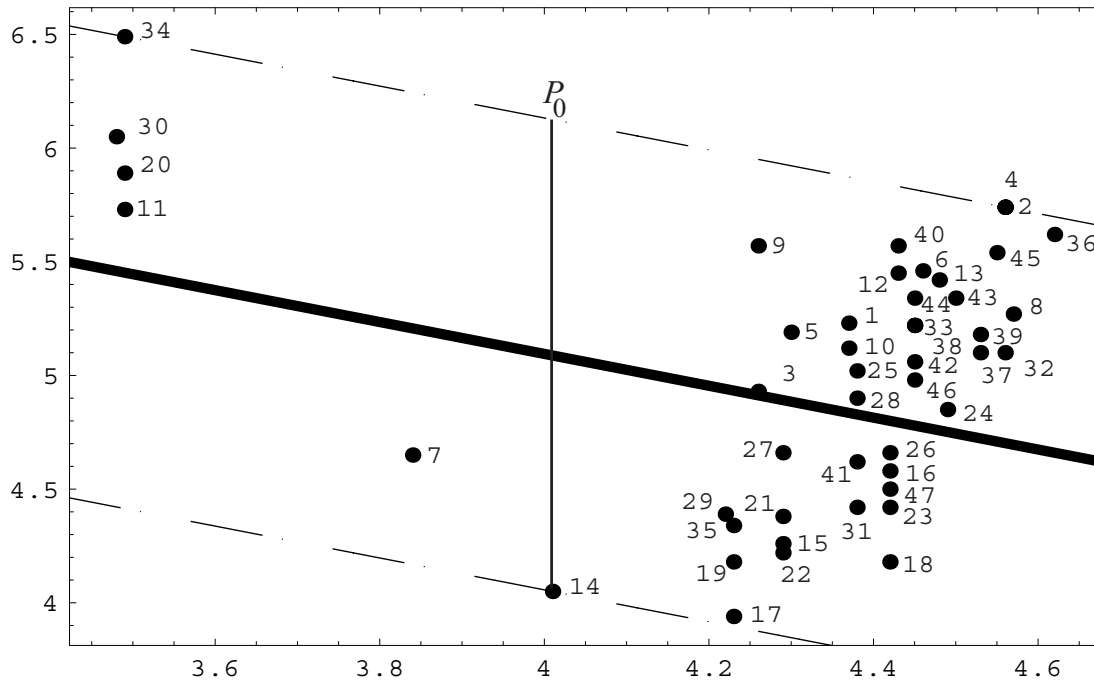


Figure 3.11: Minimax regression. The number next to a point refers to the rank of the point according its sensitivity with respect the objective function.

where m_j and s_j are the mean and standard deviation of the sensitivities in (3.135), after replacing β_j by its MM estimate.

3.12.2 A Numerical Example

Now fitting a regression line to the star cluster data using the Minimax method and solving the optimization problem (3.108)–(3.110) one gets the model

$$y = 7.89850 - 0.70093x + \varepsilon$$

with an optimal value $\varepsilon^* = 1.03776$. These estimates are associated with the points in the sets $I^+ = \{2, 34\}$ and $I^- = \{14\}$, i.e., only data points 2, 14 and 34 have active constraints. The optimal values of the dual variables are:

$$\boldsymbol{\eta}^+ = (-0.243, -0.257)^T; \quad \boldsymbol{\eta}^- = \{-0.50\}.$$

Figure 3.11 shows the data points, the minimax regression line, and the corresponding parallel bands at a vertical distance $\varepsilon = 1.03776$ up and down from the regression line. Note that the data points 2, 14 and 34 are at the bands.

Since point 2 and 4 are coincident, we have a degenerate case. However, we can eliminate the degeneration problem by removing point 4, because it has no influence on the final solution. In addition, the left directional derivatives of the optimal solution with respect

to these two points are null, because the other points leads to the optimal solution. Then, there are no partial derivatives with respect to these two points.

In this case, using the formulas in (3.121) to (3.123) we have the following matrices

$$\mathbf{F}_{\beta z} = \begin{pmatrix} 0 & 0 & 0 & 0.243 & 0 & 0.257 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.243 & 0 & 0.257 & 0 & -0.5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad (3.138)$$

$$\mathbf{Q}_{\beta} = \begin{pmatrix} -1. & -4.56 & -1. \\ -1. & -3.49 & -1. \\ 1. & 4.01 & -1. \end{pmatrix}, \quad (3.139)$$

$$\mathbf{Q}_z = \begin{pmatrix} 1. & 0 & 0 & -7.898 & 0.701 & 0 & 0 & 0 & 0 \\ 0 & 1. & 0 & 0 & 0 & -7.898 & 0.701 & 0 & 0 \\ 0 & 0 & -1. & 0 & 0 & 0 & 0 & 7.899 & -0.701 \end{pmatrix}, \quad (3.140)$$

where here we have

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \varepsilon), \quad (3.141)$$

$$\mathbf{z} = (y_2, y_{34}, y_{14}, x_{1,2}, x_{2,2}, x_{1,34}, x_{2,34}, x_{1,14}, x_{2,14}), \quad (3.142)$$

and from (3.127)–(3.129) the following sensitivities can be obtained:

$$\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}} = \begin{pmatrix} -3.505 & 4.005 & 0.5 & 27.682 & -2.457 & -31.631 & 2.807 & -3.949 & 0.35 \\ 0.935 & -0.935 & 0 & -7.382 & 0.655 & 7.382 & -0.655 & 0 & 0 \\ 0.243 & 0.257 & -0.5 & -1.919 & 0.17 & -2.03 & 0.18 & 3.949 & -0.35 \end{pmatrix},$$

$$\frac{\partial \boldsymbol{\eta}}{\partial \mathbf{z}} = \begin{pmatrix} 0 & 0 & 0 & -0.852 & 0.227 & -0.901 & 0.24 & 1.752 & -0.467 \\ 0 & 0 & 0 & 0.973 & -0.227 & 1.029 & -0.24 & -2.002 & 0.467 \\ 0 & 0 & 0 & -0.121 & 0 & -0.128 & 0 & 0.25 & 0 \end{pmatrix},$$

$$\frac{\partial Z}{\partial \mathbf{z}} = (0.243 \ 0.257 \ -0.500 \ -1.919 \ 0.170 \ -2.030 \ 0.180 \ 3.949 \ -0.350).$$

Table 3.6 shows the sensitivities of Z^* , β_0 , and β_1 with respect to data for the Minimax regression method. They have been extracted from the above matrices.

The results in Table 3.6 lead to the following conclusions that are not particular to the data in this example, but of general validity:

1. One of the hyperplane bands (upper or lower depending on the case) passes through k data points ($k = 2$, here, and the data points are 2 and 34 as can be seen in Figure 3.11). They are all points associated with one of the active constraints (3.109) or (3.110). We call these points the band points, because they define the corresponding hyperplane band.
2. There exist one point (point 14 in Figure 3.11) associated with the only active constraint in the other set of the pair (3.109) or (3.110). We call this point the ε point, because it gives the optimal value of ε , that is, the vertical distance from it to the hyperplane defining the band above (see Figure 3.11).

Table 3.6: Sensitivities of Z^* , β_0 , and β_1 With Respect to Data for the Minimax Regression Method.

Index	Z^*		β_0		β_1	
	Y	X	Y	X	Y	X
2	0.243	0.170	-3.505	-2.457	0.935	0.655
14	-0.500	-0.350	0.500	0.350	0	0
34	0.257	0.180	4.005	2.807	-0.935	-0.655

3. With the exception of degenerate cases, no more active constraints exist. This implies a total of exactly $k + 1$ active constraints. This is due to the fact that in linear programming optimal solutions coincide with basic solutions, that are defined by $k + 1$ constraints if the space of the unknowns (the regression coefficients β and ε) is of dimension $k + 1$.
4. The sensitivities of the estimated regression parameters with respect to the ε -data point (in our example $\frac{\partial \beta_1}{\partial y_{14}}$ and $\frac{\partial \beta_1}{\partial x_{14}}$) are null because a small change in the ε -data point does not alter the estimated regression hyperplane or the bands, which is defined only by the band points.
5. The sensitivities of the objective function ε with respect to the y coordinate of the ε point has always absolute value $1/2$, because the regression line does not change when changing only the ε point ordinate and is half way from it to the band. The sign of this sensitivity is positive or negative depending on whether the ε point is below or above the regression hyperplane, respectively.
6. The absolute values of the estimated regression parameters sensitivities with respect to the band points ($\frac{\partial \beta_1}{\partial y_2}$, $\frac{\partial \beta_1}{\partial y_{34}}$, $\frac{\partial \beta_1}{\partial x_2}$, and $\frac{\partial \beta_1}{\partial x_{34}}$ in our example) are identical but with opposite signs, because the changes in the slope depend on those points.

3.12.3 The Dual Minimax Regression Problem

The problem (3.108)–(3.110) can be written in matrix form as

$$\begin{array}{l} \text{Minimize} \\ \beta, \varepsilon \end{array} \quad Z_{MM} = \varepsilon \quad (3.143)$$

subject to

$$\left(\begin{array}{c|c} -\mathbf{X}_{n \times k} & -\mathbf{1}_{n \times 1} \\ \hline - & - \\ \mathbf{X}_{n \times k} & -\mathbf{1}_{n \times 1} \end{array} \right) \begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} \leq \begin{pmatrix} -\mathbf{y}_{n \times 1} \\ - \\ \mathbf{y}_{n \times 1} \end{pmatrix} : \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ - \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}. \quad (3.144)$$

The corresponding dual problem is

$$\begin{array}{l} \text{Maximize} \\ \boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)} \end{array} \quad \sum_{i=1}^n y_i \left(\mu_i^{(2)} - \mu_i^{(1)} \right) \quad (3.145)$$

subject to

$$\left(\begin{array}{c|c} -\mathbf{X}_{k \times n}^T & \mathbf{X}_{k \times n}^T \\ \hline -\mathbf{1}_{1 \times n} & -\mathbf{1}_{1 \times n} \end{array} \right) \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{k \times 1} \\ 1 \end{pmatrix}, \quad (3.146)$$

$$\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)} \leq \mathbf{0}, \quad (3.147)$$

where $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ are the dual variables.

Letting $\lambda_i^{(j)} = -2\mu_i^{(j)}$; $j = 1, 2$, this dual problem can be written as

$$\text{Minimize } \sum_{i=1}^n y_i \lambda_i^{(2)} - \sum_{i=1}^n y_i \lambda_i^{(1)} \quad (3.148)$$

$$\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)}$$

subject to

$$-\sum_{i=1}^n \lambda_i^{(1)} x_{ij} + \sum_{i=1}^n \lambda_i^{(2)} x_{ij} = 0, \quad j = 2, 3, \dots, k, \quad (3.149)$$

$$\sum_{i=1}^n \lambda_i^{(1)} = \sum_{i=1}^n \lambda_i^{(2)}, \quad (3.150)$$

$$\sum_{i=1}^n \lambda_i^{(1)} + \sum_{i=1}^n \lambda_i^{(2)} = 2, \quad (3.151)$$

$$\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)} \geq \mathbf{0}. \quad (3.152)$$

Because of (3.150) and (3.152) we can divide (3.148), (3.149) and (3.151) by $\sum_{i=1}^n \lambda_i^{(1)}$ and letting $\rho_s^{(r)} = \lambda_s^{(r)} / \sum_{i=1}^n \lambda_i^{(r)}$; $r = 1, 2$, then (3.148)–(3.152) become

$$\text{Minimize } \sum_{i=1}^n y_i \rho_i^{(2)} - \sum_{i=1}^n y_i \rho_i^{(1)} \quad (3.153)$$

$$\boldsymbol{\rho}^{(1)}, \boldsymbol{\rho}^{(2)}$$

subject to

$$\sum_{i=1}^n \rho_i^{(1)} x_{ij} = \sum_{i=1}^n \rho_i^{(2)} x_{ij}, \quad j = 2, 3, \dots, k, \quad (3.154)$$

$$\sum_{i=1}^n \rho_i^{(1)} = \sum_{i=1}^n \rho_i^{(2)} = 1, \quad (3.155)$$

$$\boldsymbol{\rho}^{(1)}, \boldsymbol{\rho}^{(2)} \geq \mathbf{0}, \quad (3.156)$$

showing that the dual variables $\boldsymbol{\rho}^{(1)}$ and $\boldsymbol{\rho}^{(2)}$ can be interpreted as two probability mass functions on the set

$$\mathcal{S} = \{(y_i, \mathbf{x}_i) | i = 1, 2, \dots, n\}.$$

Hence, the objective function in (3.153) can be interpreted as the difference of marginal means $E^{(2)}[Y] - E^{(1)}[Y]$. Similarly, the constraints in (3.154) can be interpreted as the

equality of marginal means, $E^{(1)}[X_j] = E^{(2)}[X_j]$ $j = 2, 3, \dots, k$. Accordingly, one can think of the dual as a problem of finding two probability mass functions on the set S such that they minimize the difference of marginal means $E^{(2)}[Y] - E^{(1)}[Y]$ subject to the equality of expectations $E^{(1)}[X_j] = E^{(2)}[X_j]$ $j = 2, 3, \dots, k$.

The above interesting interpretations of the dual problem can be illustrated using Figure 3.11, as an example. In this case, the probability mass function $\rho^{(1)}$ assigns probability 0.486 to point 2, and probability 0.514 to point 34, and the probability mass functions $\rho^{(2)}$ assigns probability 1 to point 14. Other points are assigned probability zero by both probability measures.

These two probabilities assigned to points 2 and 14 are inversely proportional to the distances of these points to the point P_0 in the same band whose abscissa coincides with x_{14} . Minimizing the difference of marginal expectations $E^{(2)}[Y] - E^{(1)}[Y]$ means minimizing the vertical distance between point P_0 and point 14. Note that in fact the set of sample points is partitioned into two sets: those above and those below the regression line. The supports of these two probabilities are inside these two sets.

3.13 The Least-Absolute-Value (LAV) Regression

3.13.1 The Primal LAV Regression Problem

In the LAV regression problem (see, for example, Arthanari and Dodge (1993)), we minimize the sum of the distances between observed and predicted values instead of their squares, i.e.:

$$\underset{\boldsymbol{\beta}}{\text{Minimize}} \quad Z_{LAV} = \sum_{i=1}^n |y_i - \mathbf{x}_i^T \boldsymbol{\beta}|. \quad (3.157)$$

This method treats all errors equally. Thus, this method must be used when the user is concerned about any level of error. In fact, what is important is the sum of all absolute errors, not a single error.

Due to the presence of the absolute-value function, it is difficult to solve (3.157) using standard regression techniques. The LAV problem in (3.157) is equivalent to the following problem

$$\underset{\boldsymbol{\beta}, \varepsilon_i}{\text{Minimize}} \quad Z_{LAV} = \sum_{i=1}^n \varepsilon_i \quad (3.158)$$

subject to

$$y_i - \mathbf{x}_i^T \boldsymbol{\beta} \leq \varepsilon_i, \quad i = 1, \dots, n, \quad (3.159)$$

$$\mathbf{x}_i^T \boldsymbol{\beta} - y_i \leq \varepsilon_i, \quad i = 1, \dots, n. \quad (3.160)$$

To obtain the sensitivities of the β estimates with respect to data it is convenient to assume that we are not in a degenerate case, i.e., a total of exactly n constraints in (3.159) and (3.160) are active, and for exactly k points both are active. Let I^+ and I^- the sets of indices associated with the active constraints in (3.159), and (3.160), respectively, and $L = I^+ \cup I^-$ and $I^0 = I^+ \cap I^-$, where we keep the order of the elements in I^+ and I^- .

It is also convenient to reduce the analysis to the sensitivities that are known to be non-null. Then, following the steps in Section 3.11.1, we have

$$\mathbf{q}_i^+(\bar{\mathbf{x}}, \mathbf{z}) = y_i - \mathbf{x}_i^T \boldsymbol{\beta} - \varepsilon_i = 0, \quad : \quad \eta_i^+ \quad i \in I^+ \quad (3.161)$$

$$\mathbf{q}_i^-(\bar{\mathbf{x}}, \mathbf{z}) = \mathbf{x}_i^T \boldsymbol{\beta} - y_i - \varepsilon_i = 0, \quad : \quad \eta_i^- \quad i \in I^-, \quad (3.162)$$

where the sets I^+ and I^- , have cardinals p^+ and p^- , respectively, where for non-degenerate cases $p^+ + p^- = n + k$. Note that $\boldsymbol{\eta}^+$ and $\boldsymbol{\eta}^-$ are the column vectors of dimension n with the dual variables associated with the sets I^+ and I^- , respectively and null values, otherwise.

To obtain the matrices $\mathbf{F}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ to $\mathbf{F}_{\mathbf{z}}$ in Section 3.11.1, we need the following derivatives:

$$\frac{\partial q_i(\boldsymbol{\beta}_r; \mathbf{y}, \mathbf{x})}{\partial \beta_r} = \begin{cases} -x_{ir}, & i \in I^+, \\ x_{ir}, & i \in I^-, \end{cases} \quad (3.163)$$

$$\frac{\partial q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial y_s} = \begin{cases} 1, & i \in I^+, \quad s = i, \\ -1, & i \in I^-, \quad s = i, \\ 0, & \text{otherwise,} \end{cases} \quad (3.164)$$

$$\frac{\partial q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial x_{rs}} = \begin{cases} -\beta_s, & i \in I^+, \quad r = i, \\ \beta_s, & i \in I^-, \quad r = i, \\ 0, & \text{otherwise,} \end{cases} \quad (3.165)$$

$$\frac{\partial q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \varepsilon_r} = \begin{cases} -1, & i = r, \\ 0, & \text{otherwise,} \end{cases} \quad (3.166)$$

$$\frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \beta_r \partial x_{pq}} = \begin{cases} -1, & i \in I^+, \quad p = i, \quad q = r, \\ 1, & i \in I^-, \quad p = i, \quad q = r, \\ 0, & \text{otherwise,} \end{cases} \quad (3.167)$$

$$\frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \beta_r \partial \beta_s} = 0, \quad (3.168)$$

$$\frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \beta_r \partial y_s} = \frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial y_i \partial x_{rs}} = \frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \varepsilon_r \partial x_{rs}} = \frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \varepsilon_s \partial y_r} = 0, \quad (3.169)$$

$$\frac{\partial^2 q_i(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})}{\partial \varepsilon_s \partial \beta_r} = 0, \quad (3.170)$$

$$\frac{\partial f(\boldsymbol{\beta}_r; \mathbf{y}, \mathbf{x})}{\partial \varepsilon_r} = 1, \quad (3.171)$$

and all derivatives of $f(\boldsymbol{\beta}; \mathbf{y}, \mathbf{x})$ with respect to $\boldsymbol{\beta}$, \mathbf{y} and \mathbf{x} are null.

Let \mathbf{X}^+ and \mathbf{X}^- denote the the submatrices of \mathbf{X} such that they contain the rows whose numbers are in I^+ and I^- , respectively. Then, we have:

$$\mathbf{F}_{\boldsymbol{\beta}\mathbf{z}} = \left(\begin{array}{c|c} \mathbf{0}_{k \times n} & (\boldsymbol{\eta}^- - \boldsymbol{\eta}^+)^T \otimes \mathbf{I}_k \\ \hline \mathbf{0}_{n \times n} & \mathbf{0}_{n \times kn} \end{array} \right)_{(k+n) \times (k+1)n}, \quad (3.172)$$

$$\mathbf{Q}_{\boldsymbol{\beta}} = \left(\begin{array}{c|c} -\mathbf{X}_{p^+ \times k}^+ & -\mathbf{H}_{p^+ \times n}^+ \\ \hline \mathbf{X}_{p^- \times k}^- & -\mathbf{H}_{p^- \times n}^- \end{array} \right)_{(k+n) \times (k+n)}, \quad (3.173)$$

$$\mathbf{Q}_{\mathbf{z}} = \left(\begin{array}{c|c} \mathbf{H}_{p^+ \times n}^+ & -\mathbf{H}_{p^+ \times n}^+ \otimes \boldsymbol{\beta}^T \\ \hline \mathbf{H}_{p^- \times n}^- & -\mathbf{H}_{p^- \times n}^- \otimes \boldsymbol{\beta}^T \end{array} \right)_{(k+n) \times (k+1)n}, \quad (3.174)$$

$$\mathbf{F}\boldsymbol{\beta} = (\mathbf{0}_{1 \times k} \mid \mathbf{1}_{1 \times n})_{1 \times (k+n)}, \quad (3.175)$$

$$\mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta} = \mathbf{0}_{(k+n) \times (k+n)}, \quad (3.176)$$

$$\mathbf{F}\mathbf{z} = \mathbf{0}_{1 \times (k+1)n}, \quad (3.177)$$

and $\mathbf{H}_{p^+ \times n}^+$ is such that

$$h_{ij} = \begin{cases} 1, & \text{if the } i\text{th element of } I^+ \text{ occupies the } j\text{th position in set } L, \\ 0, & \text{otherwise.} \end{cases}$$

and $\mathbf{H}_{p^- \times n}^-$ is such that

$$h_{ij} = \begin{cases} 1, & \text{if the } i\text{th element of } I^- \text{ occupies the } j\text{th position in set } L, \\ 0, & \text{otherwise.} \end{cases}$$

Once these matrices are known, one can obtain all the sensitivities using (3.87)–(3.89), that is:

$$\frac{\partial \boldsymbol{\beta}}{\partial \mathbf{z}_{(k+n) \times (k+1)n}} = -\mathbf{Q}\boldsymbol{\beta}^{-1}\mathbf{Q}\mathbf{z}, \quad (3.178)$$

$$\frac{\partial \boldsymbol{\eta}}{\partial \mathbf{z}_{(k+n) \times (k+1)n}} = -(\mathbf{Q}\boldsymbol{\beta}^{-1})^T \mathbf{F}\boldsymbol{\beta}\mathbf{z}, \quad (3.179)$$

$$\frac{\partial Z}{\partial \mathbf{z}_{1 \times (k+1)n}} = -\mathbf{F}\boldsymbol{\beta}\mathbf{Q}\boldsymbol{\beta}^{-1}\mathbf{Q}\mathbf{z}. \quad (3.180)$$

Since the first n columns of matrix $\mathbf{F}\boldsymbol{\beta}\mathbf{z}$ vanish, Expression (3.172) proves that the sensitivities of the dual variables with respect to the \mathbf{y} are null. From (3.172)–(3.177) one immediately obtains the following formulas for the sensitivities:

$$\frac{\partial \beta_j}{\partial y_i} = \begin{cases} -\rho_{j,\ell(i)}, & i \in I^+, \\ \rho_{j,\ell(i)}, & i \in I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.181)$$

$$\frac{\partial \beta_j}{\partial x_{st}} = \begin{cases} \rho_{j,\ell(s)}\beta_t, & i \in I^+, \\ -\rho_{j,\ell(s)}\beta_t, & i \in I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.182)$$

$$\frac{\partial \eta_j}{\partial y_i} = 0, \quad (3.183)$$

$$\frac{\partial \eta_j}{\partial x_{st}} = \begin{cases} \rho_{tj}\eta_{\ell(s)}, & s \in I^+, \\ -\rho_{tj}\eta_{\ell(s)}, & s \in I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.184)$$

$$\frac{\partial Z_{LAV}^*}{\partial y_i} = \begin{cases} -\sum_{j=k+1}^n \rho_{j,\ell(i)}, & i \in I^+, \\ \sum_{j=k+1}^n \rho_{j,\ell(i)}, & i \in I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.185)$$

$$\frac{\partial Z_{LAV}^*}{\partial x_{st}} = \begin{cases} \sum_{j=k+1}^n \rho_{j,\ell(s)}\beta_t, & i \in I^+, \\ -\sum_{j=k+1}^n \rho_{j,\ell(s)}\beta_t, & i \in I^-, \\ 0, & \text{otherwise,} \end{cases} \quad (3.186)$$

Table 3.7: Sensitivities of $Z_{LAV}^*, \beta_0, \beta_1$ With Respect to Data for the LAV Regression Method.

<i>Index</i>	Z_{LAV}		β_0		β_1	
	<i>Y</i>	<i>X</i>	<i>Y</i>	<i>X</i>	<i>Y</i>	<i>X</i>
1	1.	0.693	0	0	0	0
2	1.	0.693	0	0	0	0
3	-1.	-0.693	0	0	0	0
4	1.	0.693	0	0	0	0
5	1.	0.693	0	0	0	0
6	1.	0.693	0	0	0	0
7	-1.	-0.693	0	0	0	0
8	1.	0.693	0	0	0	0
9	1.	0.693	0	0	0	0
10	0.205	0.142	-3.966	-2.749	1.136	0.788
11	0.795	0.551	4.966	3.442	-1.136	-0.788
12	1.	0.693	0	0	0	0
13	1.	0.693	0	0	0	0
14	-1.	-0.693	0	0	0	0
15	-1.	-0.693	0	0	0	0
16	-1.	-0.693	0	0	0	0
17	-1.	-0.693	0	0	0	0
18	-1.	-0.693	0	0	0	0
19	-1.	-0.693	0	0	0	0
20	1.	0.693	0	0	0	0
21	-1.	-0.693	0	0	0	0
22	-1.	-0.693	0	0	0	0
23	-1.	-0.693	0	0	0	0
24	-1.	-0.693	0	0	0	0
25	-1.	-0.693	0	0	0	0
26	-1.	-0.693	0	0	0	0
27	-1.	-0.693	0	0	0	0
28	-1.	-0.693	0	0	0	0
29	-1.	-0.693	0	0	0	0
30	1.	0.693	0	0	0	0
31	-1.	-0.693	0	0	0	0
32	1.	0.693	0	0	0	0
33	1.	0.693	0	0	0	0
34	1.	0.693	0	0	0	0
35	-1.	-0.693	0	0	0	0
36	1.	0.693	0	0	0	0
37	1.	0.693	0	0	0	0
38	1.	0.693	0	0	0	0
39	1.	0.693	0	0	0	0
40	1.	0.693	0	0	0	0
41	-1.	-0.693	0	0	0	0
42	-1.	-0.693	0	0	0	0
43	1.	0.693	0	0	0	0
44	1.	0.693	0	0	0	0
45	1.	0.693	0	0	0	0
46	-1.	-0.693	0	0	0	0
47	-1.	-0.693	0	0	0	0

1. The LAV regression hyperplane passes through k data points (in our example $k = 2$ and the points are 10 and 11). We call these points regression hyperplane-points.
2. Apart from degenerate cases, infinitesimal changes in the remaining points produce no change in the regression hyperplane, and so, the corresponding sensitivities are null.
3. With the exception of degenerate cases, a total of $k+n$ active constraints exists, which correspond to constraints (3.159) and (3.160). This is due to the fact that optimal solutions in linear programming coincide with basic solutions, that are defined by $k+n$ constraints if the space of the unknowns (the k regression coefficients β and the n ε -variables) is of dimension $k+n$. This is the reason why in Table 3.7 the sensitivities $\frac{\partial\beta_0}{\partial y_i}$, $\frac{\partial\beta_1}{\partial y_i}$, $\frac{\partial\beta_0}{\partial x_i}$, and $\frac{\partial\beta_1}{\partial x_i}$ for the data points not in the LAV regression hyperplane vanish.
4. Infinitesimal changes in the regression hyperplane points lead to changes in the regression hyperplane and thus in their slopes and intercept parameters. So, their corresponding sensitivities are not null.
5. The sensitivities of the objective function with respect to the regression hyperplane points are different from zero. The sensitivities with respect to y_i take value 1 or -1 , depending on whether they are above or below the LAV regression hyperplane.
6. The sensitivities of the objective function with respect to x_i are different from zero and have positive or negative sign depending on whether they are on the left or the right hand side of the LAV regression hyperplane. All of them have the same absolute value but the sign can be different, as indicated.

The above conclusions are valid in general, i.e., they are not particular to this example.

3.13.3 The Dual LAV Regression Problem

The primal problem in (3.158)–(3.160) can be written in matrix form as:

$$\begin{array}{l} \text{Minimize } Z_{LAV} = \sum_{i=1}^n \varepsilon_i \\ \beta, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \end{array} \quad (3.191)$$

subject to

$$\left(\begin{array}{c|c} -\mathbf{X}_{n \times k} & -\mathbf{I}_n \\ \hline & \\ \mathbf{X}_{n \times k} & -\mathbf{I}_n \end{array} \right) \begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} \leq \begin{pmatrix} -\mathbf{y}_{n \times 1} \\ \mathbf{y}_{n \times 1} \end{pmatrix} : \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}. \quad (3.192)$$

The corresponding dual problem is

$$\begin{array}{l} \text{Maximize } \sum_{i=1}^n y_i (\mu_i^{(2)} - \mu_i^{(1)}) \\ \boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)} \end{array} \quad (3.193)$$

subject to

$$\left(\begin{array}{c|c} -\mathbf{X}_{k \times n}^T & \mathbf{X}_{k \times n}^T \\ \hline - & - \\ -\mathbf{I}_n & -\mathbf{I}_n \end{array} \right) \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ - \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{k \times 1} \\ - \\ \mathbf{1}_{n \times 1} \end{pmatrix}, \quad (3.194)$$

$$\boldsymbol{\mu}^{(1)}, \boldsymbol{\mu}^{(2)} \leq \mathbf{0}, \quad (3.195)$$

where $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2)}$ are the dual variables.

Letting $\lambda_i^{(j)} = -\mu_i^{(j)}$; $j = 1, 2$, this dual problem can be written as

$$\begin{aligned} & \text{Minimize} \quad \sum_{i=1}^n y_i \lambda_i^{(2)} - \sum_{i=1}^n y_i \lambda_i^{(1)} \\ & \boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)} \end{aligned} \quad (3.196)$$

subject to

$$\sum_{i=1}^n \lambda_i^{(1)} = \sum_{i=1}^n \lambda_i^{(2)}, \quad (3.197)$$

$$-\sum_{i=1}^n \lambda_i^{(1)} x_{ij} + \sum_{i=1}^n \lambda_i^{(2)} x_{ij} = 0, \quad j = 2, 3, \dots, k, \quad (3.198)$$

$$\lambda_i^{(1)} + \lambda_i^{(2)} = 1, \quad i = 1, 2, \dots, n, \quad (3.199)$$

$$\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)} \geq \mathbf{0}, \quad (3.200)$$

which shows that the sets of dual variables $\boldsymbol{\lambda}^{(1)}$ and $\boldsymbol{\lambda}^{(2)}$ can be interpreted as a set of n probability mass functions defined on the set

$$\mathcal{S} = \{(-y_i, -\mathbf{x}_i) | i = 1, 2, \dots, n\} \cup \{(y_i, \mathbf{x}_i) | i = 1, 2, \dots, n\}$$

and such that

$$\Pr((-y_i, -\mathbf{x}_i)) = \lambda_i^{(1)}, \quad \Pr((y_i, \mathbf{x}_i)) = \lambda_i^{(2)}.$$

Then, the constraints (3.198) can be interpreted as zero means, $E[X_{ij}]$, and the objective function as the sum of n means $\sum_{i=1}^n E[Y_i]$. So the dual problem consists of assigning n different probability mass functions to the set \mathcal{S} such that they minimize the sum of expectations $\sum_{i=1}^n E[Y_i]$ subject to the equality $\sum_{i=1}^n \lambda_i^{(1)} = \sum_{i=1}^n \lambda_i^{(2)}$.

It is interesting to interpret the dual with the help of Figure 3.12. The $\boldsymbol{\lambda}^{(1)}$ and $\boldsymbol{\lambda}^{(2)}$ assign non-zero probability to points above, below and on the regression line, respectively (see the correspondence with sets I^+ and I^-).

Chapter 4

Structural Reliability Methods I

4.1 Introduction

En el capítulo ?? se han revisado los principios básicos de la fiabilidad estructural. Desde los trabajos pioneros de Freudenthal [81] en la década de los cincuenta, en los que comenzó a introducir los conceptos estadísticos para calcular la probabilidad de fallo, se han desarrollado metodologías que nos permiten dividir los métodos para tratar los problemas relativos a la fiabilidad en varios niveles:

1. NIVEL 1: Se seleccionan coeficientes de seguridad parciales para cada una de las variables (cargas, resistencias, etc.). Es el método clásico y el más utilizado en los códigos actuales.
2. NIVEL 2: Alternativamente se puede obtener la probabilidad de fallo, p_f , que puede calcularse usando la función de densidad conjunta $f(\mathbf{x}) = f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$ de todas las variables que intervienen mediante la expresión:

$$p_f = \int_{g(x_1, x_2, \dots, x_n) \leq 0} f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n. \quad (4.1)$$

El problema es que la integral normalmente es difícil de calcular, debido a dos razones principales: (a) lo complicado de la función de densidad $f(\mathbf{x})$, y (b) la complejidad de la región de fallo $g(\mathbf{x}) \leq 0$. Por tanto, se han de utilizar métodos aproximados, que se basan en aproximaciones de la función de densidad $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$, de la región de fallo o ecuación de estado límite $g(x_1, x_2, \dots, x_n) \leq 0$, o de ambas. A este nivel, se utiliza una aproximación de la distribución de la función de probabilidad usando los dos primeros momentos de la función de distribución conjunta.

3. NIVEL 3: Para el cálculo de la probabilidad de fallo se utiliza la función de densidad conjunta global, y se trata de calcular la probabilidad de fallo exacta. Estos métodos requieren fórmulas especiales de integración y metodologías específicas.

En la tabla 4.1 se presenta un esquema detallado de las metodologías existentes.

Table 4.1: Jerarquía de las medidas de fiabilidad estructural.

Nivel	Métodos de cálculo	Distribuciones estadísticas	Ecuaciones de estado límite	Incertidumbre asociada	Resultados
1:	Calibración de códigos usando métodos de nivel 2 o 3	No se usan	Ecuaciones lineales usualmente	Factores arbitrarios	Cofec. parciales
2: Momentos de segundo orden	Álgebra de segundo orden	Sólo distribuciones normales	Lineales o aprox. lineales	Puede incluirse como distribuciones normales	Probab. de fallo
3: Métodos exactos	Transformaciones	Distribuciones normales equivalentes	Lineales o aproximadamente lineales	Puede incluirse	Probab. de fallo
	Integración numérica y simulación	Cualesquiera	Cualesquiera	variables aleatorias	
4: Métodos de decisión					Mínimo coste o máximo beneficio (RBO)

Los métodos de nivel 2 para aproximar la integral (4.1) utilizan los dos primeros momentos de las distribuciones de las variables que intervienen en el caso de que estén incorrelacionadas, en caso de que haya correlación entre las variables, primeramente se transforman en variables independientes y posteriormente se trabaja con los dos primeros momentos de las variables resultantes, con lo cual se trabaja con variables normales independientes. Se emplea una aproximación lineal de la región de fallo (FOSM ‘First Order Second Moment’).

Los trabajos de Mayer [120], Freudenthal [81], Rzhantitsyn [151] y Basler [14] contenían conceptos asociados con los dos primeros momentos de las distribuciones. Pero no fue hasta los trabajos de Cornell [48] cuando la metodología se asentó y cobró la relevancia que merecía.

Los ‘Métodos de Fiabilidad de Primer Orden’ (FORM ‘First Order Reliability Methods’), usan también aproximaciones lineales de las ecuaciones de estado límite $g(x_1, x_2, \dots, x_n) \leq 0$ pero trabajan con las funciones de densidad exactas de las variables que intervienen, surgieron en el campo de la fiabilidad estructural con Freudenthal [81] en 1956, y han sido ampliados por Hasofer y Lind [92], Rackwitz y Flessler [140], Hohenbichler y Rackwitz [95], Ditlevsen ([61], etc.

Los métodos de segundo orden (SORM ‘Second Order Reliability Methods’) utilizan un desarrollo en serie de Taylor para aproximar las regiones de fallo (véase, por ejemplo, Davies [52], Field [75], Breitung [25], Tvedt [167, 168], Der Kiureghian et al. [57], Katsuki y Frangopol [101], Koyluoglu y Nielsen [105], Cai y Elishakoff (1994), Papadimitriou [134] o Zhao y Ono [181]). Para una completa descripción de estos métodos y algunos ejemplos ilustrativos véase Ditlevsen y Madsen [64] y Madsen, Krenk y Lind [116]. Estos métodos han demostrado dar resultados muy precisos y son mucho más eficientes que las técnicas de simulación de Monte Carlo para estimar percentiles extremos (véase, por ejemplo, Wirsching y Wu [175], o Haskin, Staple y Ding [91]).

Una alternativa muy extendida es tratar el problema mediante simulación ponderada (véase por ejemplo, Siegmund [156], Rubinstein [150], Zhang [180], o Givens y Raftery [88]). Estas técnicas pueden usarse para disminuir el tiempo y número de simulaciones con

respecto al método básico de Monte Carlo (véase Hesterberg [93]).

4.2 The Tail Problem

El lector debe tener presente la problemática que surge en todas estas metodologías cuando las probabilidades de fallo son muy pequeñas, que es lo normal en el análisis de la fiabilidad de estructuras. En estos casos se dice que se está tratando con las colas de las distribuciones que es bien sabido son muy sensibles al modelo supuesto (véase Galambos [83] y Castillo [28]). En otras palabras, la probabilidad de fallo es extremadamente dependiente de la cola de las distribuciones. Dado que siempre se trata con problemas con muy pequeñas probabilidades de fallo, se necesitan métodos adecuados de simulación o cálculo (la simulación directa con Monte Carlo es muy ineficiente). Algunos de ellos se pueden estudiar en Castillo, Solares y Gómez [40, 39, 41], Wirsching y Wu [175] y Wu, Burnside y Cruse [179]).

4.3 Integration Methods

La solución analítica de la integral de convolución (??) o de la integral (4.1) es sólo posible en un número muy reducido de situaciones. Por tanto, en la mayoría de los casos se precisa la integración numérica. La aproximación más simple utiliza la regla trapezoidal, que ha dado resultados bastante aceptables, principalmente porque la infravaloración de la integral en torno a los valores medios de las variables se compensa por la sobrestimación en el dominio de integración restante (Dalhquist y Björck, [51]). Metodologías mucho más refinadas, tales como la regla de Simpson, o fórmulas de cuadratura como Gauss-Laguerre, Gauss-Hermite, etc. pueden ser más apropiados (véase Davis y Rabinowitz, [53]).

El problema de la integración numérica es que el coste computacional y la imprecisión aumentan considerablemente con la dimensión del problema, de tal forma que no han tenido un papel predominante en la determinación de la fiabilidad y se han desarrollado métodos alternativos.

4.4 Simulation Methods

Una alternativa para el cálculo de la probabilidad de fallo definida por la integral (4.1) son los métodos de simulación. Consisten, como su propio nombre indica, en simular artificialmente un gran número de realizaciones de las variables aleatorias $\hat{\mathbf{x}} = \hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$ (experimentos), y comprobar si se produce superación del estado límite ‘fallo’ ($g(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) \leq 0$) o si no se produce ($g(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) > 0$). La probabilidad de fallo se puede aproximar mediante la expresión:

$$p_f \approx \frac{n(g(\hat{\mathbf{x}}) \leq 0)}{N} \quad (4.2)$$

donde $n(g(\hat{\mathbf{x}}) \leq 0)$ es el número de experimentos n en los que $g(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) \leq 0$ (fallos) y N es el número total de realizaciones.

El procedimiento descrito anteriormente es el más sencillo de realizar con esta metodología, y se conoce como método de Monte Carlo ‘normal’, sin embargo para mejorar los resultados han surgido procedimientos alternativos que han dado como resultado metodologías muy eficientes. Para poder aplicar estas técnicas a problemas de fiabilidad estructural es necesario:

1. Desarrollar metodologías sistemáticas para generar vectores aleatorios para las variables básicas o de proyecto \mathbf{x} .
2. Seleccionar una estrategia de simulación adecuada.
3. Desarrollar técnicas para evaluar la función $g(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$.
4. Seleccionar el número de experimentos a realizar para obtener una estimación aceptable de la probabilidad de fallo p_f .

No se va a entrar en detalle sobre los puntos 1, 3 y 4, información amplia y detallada sobre esos temas aplicados a la fiabilidad estructural se puede encontrar en Melchers [123] y en Box y Muller [24]. Pero sí se va a comentar la selección de la estrategia de simulación, que es el punto en el que los métodos de simulación más han mejorado, y que los ha convertido en una herramienta muy útil en el campo de la fiabilidad.

4.4.1 Importance sampling

Partiendo de la metodología básica de la simulación de Monte Carlo, vista en la sección anterior, en el año 1983 se publicaron dos artículos (Harbitz [90] y Shinozuka [155]) en los que se desarrollaban sendas técnicas basadas en conceptos de simulación, que hicieron que la comunidad científica que trataba los temas de fiabilidad, aceptara la aplicabilidad práctica de las mismas. Su principal ventaja con respecto a otras metodologías es que proporcionaban de forma eficiente y simultánea una estimación de la probabilidad de fallo y del error. Además de permitir trabajar con funciones de estado límite no diferenciables.

Se denominaron ‘exactas’ en el sentido de que tienden al resultado exacto cuando N (el número de realizaciones) tiende a infinito. El procedimiento más sencillo de implementación es el siguiente:

$$\begin{aligned}
 p_f &= \int_{g(\mathbf{x}) \leq 0} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_{g(\mathbf{x}) \leq 0} f_{\mathbf{X}}(\mathbf{x}) \frac{\psi(\mathbf{x})}{\psi(\mathbf{x})} d\mathbf{x} \\
 &\approx \frac{1}{N} \sum_{i=1}^N \ell(g(\mathbf{x}_i)) \frac{f_{\mathbf{X}}(\mathbf{x}_i)}{\psi(\mathbf{x}_i)}
 \end{aligned} \tag{4.3}$$

donde $\ell(g(\mathbf{x}_i))$ es la función indicadora para el dominio de fallo, y $\psi(\mathbf{x}_i)$ es la función de ponderación seleccionada. Se elige de forma que la región de interés, en el entorno de la zona de rotura, sea la zona más importante y predominante.

Actualmente se han desarrollado sofisticadas estrategias de simulación que pueden sustituir a los métodos más utilizados en la actualidad (FORM y SORM, que se tratarán en las secciones siguientes) con un pequeño aumento del tiempo de computación.

Para obtener más detalles sobre este método y los tipos de funciones de ponderación véase Melchers [122, 123], Engelund y Rackwitz [72] y Melchers y Li [124]. Para lectores

interesados en la búsqueda secuencial o en técnicas adaptativas de mejora del método consultar Melchers [121], Harbitz [90], Bucher [26], Karamchandani et al. [99], Der Kiureghian y Dakessian [55].

Para el estudio de sensibilidad mediante técnicas de simulación, es decir, como afectan los parámetros o variables en la probabilidad de fallo véase Rubinstein [150], Melchers [122] y Karamchandani y Cornell [100].

4.4.2 Directional simulation

Los métodos de simulación vistos hasta ahora han sido formulados en el sistema cartesiano exclusivamente. La idea de trabajar en coordenadas polares surgió de los trabajos de Deák (1980) para tratar de evaluar la función de distribución normal multivariada. Esto llevo a la simulación en el espacio multinormal estándar \mathbf{y} en coordenadas polares, a lo que se llamó ‘simulación direccional’ (Ditlevsen et al [65]).

Dado un vector \mathbf{Z} en el espacio Gaussiano n -dimensional (\mathbf{z}), de tal forma que $\mathbf{Z} = R\mathbf{A}$ ($R \geq 0$), donde \mathbf{A} es un vector compuesto por cosenos directores aleatorios independientes, que representa una dirección en el espacio \mathbf{Z} . El vector \mathbf{A} está uniformemente distribuido sobre la hiperesfera n -dimensional, y R es tal que R^2 sigue una distribución chi-cuadrado con n grados de libertad. Si R es independiente de \mathbf{A} , la integral (4.1) queda (Bjerager [20, 21]):

$$p_f = \int_{\mathbf{a}} P[g(R\mathbf{A}) \leq 0 | \mathbf{A} = \mathbf{a}] f_{\mathbf{A}}(\mathbf{a}) d\mathbf{a} \quad (4.4)$$

donde $f_{\mathbf{A}}(\cdot)$ es función de densidad de \mathbf{A} sobre la hiperesfera unitaria y $g(\cdot) = 0$ es la función de estado límite en el espacio \mathbf{Z} .

El procedimiento de simulación procede generando un vector $\hat{\mathbf{a}}_i$ y calculando el valor de r tal que $g(r\hat{\mathbf{a}}) = 0$. La probabilidad de fallo de la estimación será:

$$\hat{p}_{f_i} = P[g(r\hat{\mathbf{a}})_i \leq 0] = 1 - \chi_n^2(r) \quad (4.5)$$

donde $\chi_n^2(\cdot)$ es la función de distribución chi-cuadrado con n grados de libertad.

Repetiendo el procedimiento con las N realizaciones permite estimar la probabilidad aproximada como:

$$p_f \approx E(\hat{p}_{f_i}) = \frac{1}{N} \sum_{i=1}^N \hat{p}_{f_i} \quad (4.6)$$

Esta técnica es especialmente útil con regiones de fallo con mucha curvatura, y se comporta mal con funciones lineales (Engelund y Rackwitz [72]).

Otra alternativa de la simulación direccional es su combinación con el método de ponderación en una determinada dirección (véase Ditlevsen y Bjerager [63]).

4.5 Second-Moment Methods

Una alternativa a los métodos de integración y simulación vistos en las secciones 4.3 y 4.4, respectivamente, son los métodos que tratan de estimar la fiabilidad de un sistema basándose

en los dos primeros momentos de la función de densidad de cada una de las variables que intervienen, esto es, la media y la desviación típica. Por tanto cada variable está representada por variables normales cuyas distribuciones quedan perfectamente definidas con los dos primeros momentos.

En el ejemplo ilustrativo ?? se mostró que para el caso de dos variables normales de medias μ_R y μ_S y varianzas σ_R^2 y σ_S^2 , respectivamente, su margen de seguridad $G = R - S$ era una variable aleatoria con una distribución normal $G \sim N(\mu_G, \sigma_G^2)$ con la que se puede calcular la probabilidad de fallo p_f como:

$$p_f = P(R - S \leq 0) = P(G \leq 0) = \Phi\left(\frac{0 - \mu_G}{\sigma_G}\right) \quad (4.7)$$

donde $\beta = \mu_G/\sigma_G$ es el índice de fiabilidad.

Evidentemente, la ecuación (4.7) proporciona el valor exacto, ya que tanto R como S están normalmente distribuidas. En el caso de que tengan otras distribuciones se obtiene una probabilidad de fallo ‘nominal’. De hecho, conceptualmente es mejor hablar en términos de ‘índice de fiabilidad’ o ‘seguridad’ β que de probabilidad de fallo.

Las ideas descritas anteriormente se pueden extender al caso de ecuaciones de estado límite con combinaciones lineales de un conjunto de variables aleatorias normales de la forma:

$$G_{\mathbf{X}}(\mathbf{X}) = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n \quad (4.8)$$

donde $G_{\mathbf{X}}(\mathbf{X})$ está normalmente distribuida, por lo que se puede calcular su índice de fiabilidad β y su probabilidad de fallo p_f .

Dado que $\sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_{X_i}, \sum_{i=1}^n a_i^2 \sigma_{X_i}^2\right)$,

$$p_f \leq P\left(\sum_{i=1}^n a_i X_i \leq -a_0\right) = F_{N\left(\sum_{i=1}^n a_i \mu_{X_i}, \sum_{i=1}^n a_i^2 \sigma_{X_i}^2\right)}(-a_0) = \Phi\left(\frac{-a_0 - \sum_{i=1}^n a_i \mu_{X_i}}{\sqrt{\sum_{i=1}^n a_i^2 \sigma_{X_i}^2}}\right) = \Phi(-\beta), \quad (4.9)$$

donde $F_{N(\cdot)}(x)$ es la función de distribución de la $N\left(\sum_{i=1}^n a_i \mu_{X_i}, \sum_{i=1}^n a_i^2 \sigma_{X_i}^2\right)$, y $\Phi(\cdot)$ es la función de distribución de la variable normal estándar $N(0, 1)$.

En el caso de que la ecuación de estado límite $G(\mathbf{X}) = 0$ fuera no lineal, no se podrían calcular sus dos primeros momentos, por lo que se lineariza $G(\mathbf{X}) = 0$ mediante desarrollo en serie de Taylor de primer orden en el entorno del punto \mathbf{x}^* , con lo que se obtiene $G_L(\mathbf{X}) = 0$. Esta aproximación se conoce como método de primer orden (véase la figura 4.1).

4.6 First-Order Second-Moment (FOSM) theory

4.6.1 Hasofer-Lind reliability index

En el capítulo ?? en el que se estudiaron las medidas de la fiabilidad estructural, se vió que uno de los problemas que existían era la falta de invarianza de algunas de las metodologías

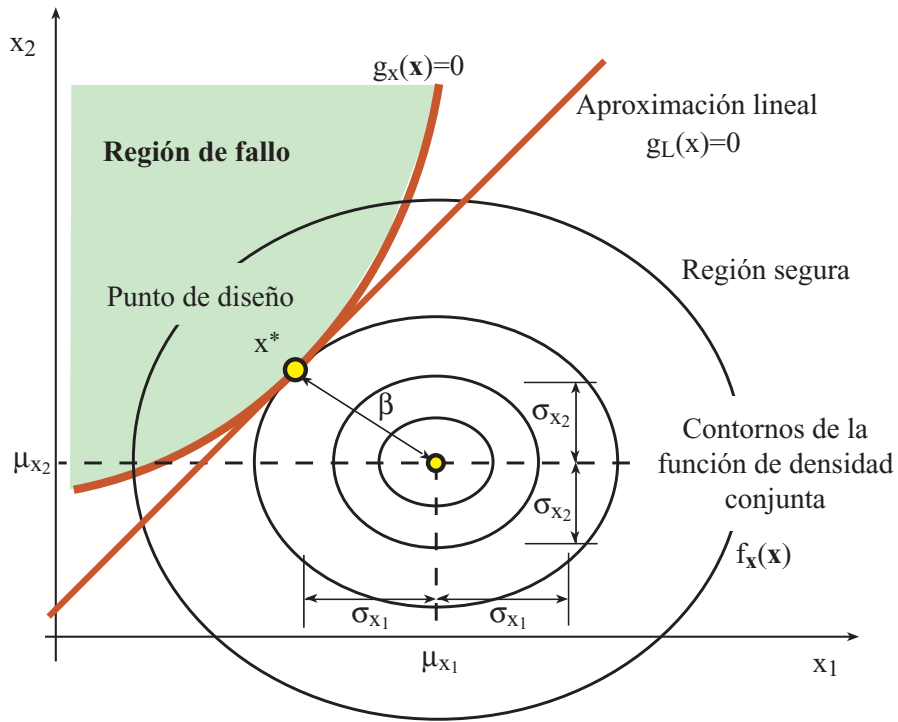


Figure 4.1: Ilustración gráfica de la superficie de estado límite $G_{\mathbf{X}}(\mathbf{X}) = 0$ y su aproximación lineal en el entorno del punto \mathbf{x}^* para el caso bidimensional. Interpretación gráfica del índice de fiabilidad β de Hasofer-Lind.

empleadas. Lógicamente, es deseable, que la medida que se utilice para determinar si un sistema es lo suficientemente seguro o no, sea independiente de la definición de la ecuación de estado límite, es decir, que sea invariante. Para solucionar este problema, Hasofer y Lind [92] propusieron una definición invariante del índice de fiabilidad β .

Sea \mathbf{X} un vector que contiene las variables aleatorias que intervienen, $\boldsymbol{\mu}_{\mathbf{X}}$ su vector de medias y $\boldsymbol{\sigma}_{\mathbf{X}}$ su matriz de varianzas covarianzas. La formulación matricial del índice de fiabilidad propuesto por Hasofer y Lind es (véase Ditlevsen [62], Veneciano [170] y Low y Tang [113]):

$$\beta = \underset{\mathbf{x}}{\text{Mínimo}} \sqrt{(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^T \boldsymbol{\sigma}_{\mathbf{X}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{X}})^T} \quad (4.10)$$

sometido a

$$g_{\mathbf{X}}(\mathbf{x}) = 0. \quad (4.11)$$

El valor óptimo de $\mathbf{x} = \mathbf{x}^*$ (véase la figura 4.1) se conoce como ‘punto de diseño’ o ‘punto de máxima verosimilitud’, y es el punto dentro de la región de fallo en el que el valor de la función de densidad conjunta de las variables que intervienen $f_{\mathbf{X}}(\mathbf{x}, \boldsymbol{\sigma}_{\mathbf{X}})$, que es una función normal multivariada, es máximo. Es decir, es el valor más probable de las variables de proyecto con el que se produce fallo o superación del estado límite.

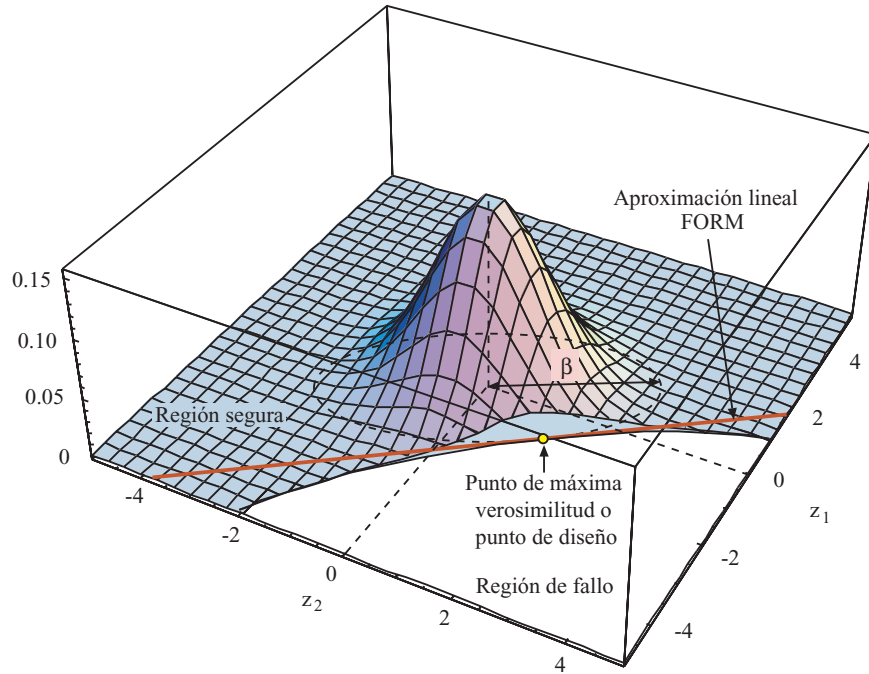


Figure 4.2: Ilustración gráfica de la función de densidad normal estándar bivariada, del índice de fiabilidad de Hasofer-Lind y del ‘punto de diseño’ o de ‘máxima verosimilitud’ en el espacio \mathbf{z} .

Se puede obtener una interpretación más intuitiva del índice de fiabilidad. Si se transforman las variables aleatorias \mathbf{X} al espacio normal estándar multivariado \mathbf{Z} de forma que:

$$\boldsymbol{\mu}_Z = \mathbf{0}, \quad \boldsymbol{\sigma}_Z = \mathbf{I},$$

donde $\boldsymbol{\mu}_Z$ es el vector de medias de las nuevas variables, $\boldsymbol{\sigma}_Z$ es la nueva matriz de varianzas-covarianzas, e \mathbf{I} es la matriz identidad. El problema (4.10)-(4.11) se transforma en:

$$\beta = \underset{\mathbf{z}}{\text{Mínimo}} \sqrt{\mathbf{z}^T \mathbf{z}} \quad (4.12)$$

sometido a

$$g_Z(\mathbf{z}) = 0. \quad (4.13)$$

En la figura 4.2 se muestra la interpretación gráfica del índice de fiabilidad en el espacio \mathbf{z} , en la que se comprueba que el índice de fiabilidad es la distancia mínima del origen de coordenadas al punto más próximo de la ecuación de estado límite. El punto de diseño \mathbf{z}^* es el transformado del punto \mathbf{x}^* .

Illustrative Example 4.1 (Índice de fiabilidad de Hasofer-Lind para ecuaciones de estado límite lineales). Se ha visto que en caso de ecuaciones de estado límite lineales

$G_{\mathbf{X}}(\mathbf{X}) = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n$ como en (4.8), el índice de fiabilidad se podía calcular como (4.9):

$$\beta = \left(a_0 + \sum_{i=1}^n a_i \mu_{X_i} \right) / \sqrt{\sum_{i=1}^n a_i^2 \sigma_{X_i}^2}. \quad (4.14)$$

Se va a tratar de obtener este resultado resolviendo el problema (4.10)-(4.11) suponiendo que las variables \mathbf{X} son linealmente independientes pero considerando como función objetivo β^2 , ya que la solución será la misma:

$$\beta^2 = \underset{x_i; i=1, \dots, n}{\text{Mínimo}} \frac{\sum_{i=1}^n (x_i - \mu_{X_i})^2}{\sigma_{X_i}^2} = \sum_{i=1}^n z_i^2 \quad (4.15)$$

sometido a

$$\begin{aligned} g_{\mathbf{X}}(\mathbf{x}) &= a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n = \\ &= a_0 + a_1(\mu_{X_1} + \sigma_{X_1}z_1) + \dots + a_n(\mu_{X_n} + \sigma_{X_n}z_n) = 0 \end{aligned} \quad (4.16)$$

La función Lagrangiana del problema anterior es:

$$\mathcal{L}(\mathbf{z}, \lambda) = \sum_{i=1}^n z_i^2 + \lambda(a_0 + a_1(\mu_{X_1} + \sigma_{X_1}z_1) + \dots + a_n(\mu_{X_n} + \sigma_{X_n}z_n)) \quad (4.17)$$

donde λ es el multiplicador de Lagrange asociado a la restricción.

Derivando con respecto a las variables \mathbf{z} y λ

$$\frac{\partial \mathcal{L}(\mathbf{z}, \lambda)}{\partial z_i} = 2z_i + \lambda a_i \sigma_{X_i} = 0; \quad i = 1, \dots, n \quad (4.18)$$

$$\frac{\partial \mathcal{L}(\mathbf{z}, \lambda)}{\partial \lambda} = a_0 + \sum_{i=1}^n a_i(\mu_{X_i} + \sigma_{X_i}z_i) = 0 \quad (4.19)$$

De la ecuación (4.18) se obtiene

$$z_i = \frac{-\lambda a_i \sigma_{X_i}}{2},$$

sustituyendo estos valores en la ecuación (4.19)

$$a_0 + \sum_{i=1}^n a_i \left(\mu_{X_i} + \sigma_{X_i} \frac{-\lambda a_i \sigma_{X_i}}{2} \right) = 0,$$

y despejando el valor de λ ,

$$\lambda = 2 \frac{\left(a_0 + \sum_{i=1}^n a_i \mu_{X_i} \right)}{\sum_{i=1}^n a_i^2 \sigma_{X_i}^2}.$$

Sustituyendo los valores de z_i y λ en la función objetivo

$$\beta^2 = \sum_{i=1}^n z_i^2 = \sum_{i=1}^n \left(\frac{-\lambda a_i \sigma_{X_i}}{2} \right)^2 = \frac{4\lambda^2}{4} \sum_{i=1}^n a_i^2 \sigma_{X_i}^2 = \frac{\left(a_0 + \sum_{i=1}^n a_i \mu_{X_i} \right)^2}{\sum_{i=1}^n a_i^4 \sigma_{X_i}^4} \sum_{i=1}^n a_i^2 \sigma_{X_i}^2 \quad (4.20)$$

simplificando se llega a la conclusión de que el índice de fiabilidad es:

$$\beta = \frac{a_0 + \sum_{i=1}^n a_i \mu_{X_i}}{\sqrt{\sum_{i=1}^n a_i^2 \sigma_{X_i}^2}} \quad (4.21)$$

expresión que es idéntica a (4.14). ■

4.6.2 Hasofer-Lind transformation

El cambio de variable propuesto por Hasofer y Lind, no es esencial para que el método funcione, ahora bien es muy útil porque en el espacio transformado es más fácil calcular el punto de diseño y el índice de fiabilidad. Por lo tanto, un aspecto fundamental en los métodos FOSM es la transformación de las variables normales multivariadas, en variables normales estándar independientes $N(0, 1)$. En el caso de que las variables que intervienen sean independientes se puede utilizar la bien conocida transformación:

$$Z_i = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}}; \forall i = 1, \dots, n. \quad (4.22)$$

En el caso de que las variables estén correlacionadas se puede emplear la transformación ortogonal de variables aleatorias normales.

Orthogonal transformation of normal random variables

Sea $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vector que contiene las variables aleatorias correlacionadas que intervienen en un sistema. Sus valores medios, las varianzas y covarianzas son, respectivamente

$$\boldsymbol{\mu}_X = E(\mathbf{X}) = (E(X_1), E(X_2), \dots, E(X_n)) = (\mu_{X_1}, \mu_{X_2}, \dots, \mu_{X_n}),$$

y

$$\boldsymbol{\sigma}_X = \text{cov}(X_i, X_j) = \sigma_X^2(i, j). \quad (4.23)$$

La matriz de varianzas-covarianzas (4.23) será estrictamente diagonal, si las variables (X_1, X_2, \dots, X_n) están incorrelacionadas, es decir, si son independientes.

El objetivo de la transformación es obtener un vector linealmente independiente \mathbf{U} , y una matriz de transformación \mathbf{B} de forma que

$$\mathbf{U} = \mathbf{B}\mathbf{X} \quad (4.24)$$

Es necesario que la transformación sea ortogonal, de forma que las distancias en ambos espacios \mathbf{X} y \mathbf{U} sean constantes para que el índice de fiabilidad sea el mismo.

Observese, que el vector \mathbf{U} sólo será incorrelacionado si la matriz de varianzas-covarianzas σ_U transformada es diagonal. Ésta, bajo la transformación lineal (4.24) queda como:

$$\sigma_U = \text{cov}(\mathbf{U}, \mathbf{U}^T) = \text{cov}(\mathbf{B}\mathbf{X}, \mathbf{X}^T \mathbf{B}^T) = \quad (4.25)$$

$$= \mathbf{B} \text{cov}(\mathbf{X}, \mathbf{X}^T) \mathbf{B}^T = \mathbf{B} \sigma_X \mathbf{B}^T \quad (4.26)$$

Del análisis matricial, y teniendo en cuenta que en los problemas ingenieriles, la matriz de varianzas-covarianzas es simétrica y definida positiva, se puede descomponer usando la descomposición de Cholesky, con lo que no tienen que calcularse los autovalores y autovectores, así

$$\sigma_X = \mathbf{L}\mathbf{L}^T \quad (4.27)$$

donde \mathbf{L} es una matriz triangular inferior, tal que su inversa $\mathbf{B} = \mathbf{L}^{-1}$ será también triangular inferior y podrá obtenerse fácilmente. Entonces

$$\mathbf{B} \sigma_X \mathbf{B}^T = (\mathbf{B} \mathbf{L}) (\mathbf{L}^T \mathbf{B}^T) = \mathbf{I} \quad (4.28)$$

donde \mathbf{I} es la matriz identidad.

Sustituyendo (4.28) en (4.26) se obtiene

$$\sigma_U = \mathbf{I}. \quad (4.29)$$

El último paso es transformar las variables $\mathbf{U} \sim N(\boldsymbol{\mu}_U, \mathbf{I})$ en el conjunto $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_n)$, para lo cual:

$$\mathbf{Z} = \frac{\mathbf{Z} - \boldsymbol{\mu}_Z}{\mathbf{I}} = \mathbf{B}(\mathbf{X} - \boldsymbol{\mu}_X) \quad (4.30)$$

Expresión esta última que permite pasar del espacio \mathbf{X} al \mathbf{Z} .

4.6.3 Limit state equation

El método FOSM por tanto, lineariza $G_{\mathbf{X}}(\mathbf{X}) = 0$ mediante desarrollo en serie de Taylor de primer orden en el entorno del punto \mathbf{x}^* obtenido, que es el ‘punto de diseño’ en el espacio \mathbf{X} y que se corresponde con el punto \mathbf{z}^* . Además, calcula el índice de fiabilidad β (véase la figura 4.3).

Hay una relación directa entre el punto de diseño \mathbf{z}^* y el índice de fiabilidad β . El vector unitario normal a la ecuación de verificación $g_{\mathbf{Z}}(\mathbf{z}) = 0$ en el punto de diseño viene dado por la expresión:

$$\boldsymbol{\alpha} = \frac{\frac{\partial g_{\mathbf{Z}}}{\partial \mathbf{z}}}{\sqrt{\frac{\partial g_{\mathbf{Z}}}{\partial \mathbf{z}}^T \frac{\partial g_{\mathbf{Z}}}{\partial \mathbf{z}}}} \quad (4.31)$$

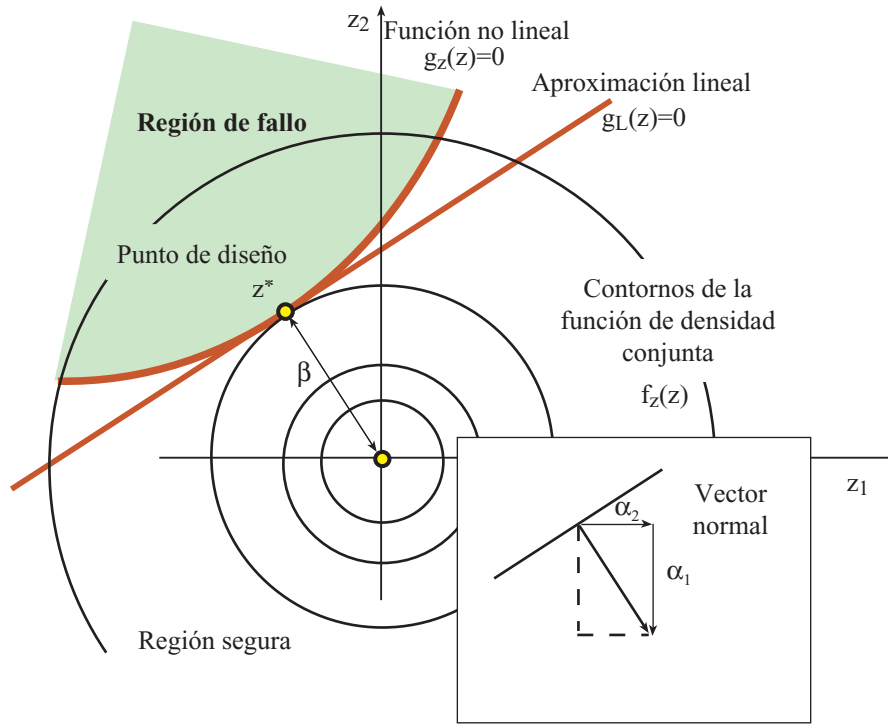


Figure 4.3: Ilustración gráfica del índice de fiabilidad β , de las ecuaciones de estado límite, tanto lineales como no lineales $g_{\mathbf{Z}}(\mathbf{z})$ y de los contornos de la función de densidad normal estándar bivariada.

Con α (véase la figura 4.3) conocido, las coordenadas del punto de diseño son:

$$\mathbf{z}^* = -\alpha\beta \quad (4.32)$$

donde el signo negativo proviene del hecho de que el vector α es positivo en el sentido creciente de $g_{\mathbf{Z}}(\mathbf{z})$.

En el caso de tener una ecuación de estado límite lineal de la forma $G_{\mathbf{X}}(\mathbf{X}) = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n$ es relativamente fácil encontrar el punto de fallo en el espacio \mathbf{Z} que verifique (4.32) y (4.12)-(4.13) ya que se conoce de antemano la dirección del vector α . La ecuación de verificación en el espacio \mathbf{Z} queda utilizando (4.22) de la forma:

$$g_{\mathbf{Z}}(\mathbf{z}) = a_0 + a_1(\mu_{X_1} + \sigma_{X_1}z_1) + a_2(\mu_{X_2} + \sigma_{X_2}z_2) + \dots + a_n(\mu_{X_n} + \sigma_{X_n}z_n) \quad (4.33)$$

$$= a_0 + \sum_{i=1}^n a_i\mu_{X_i} + \sum_{i=1}^n a_i\sigma_{X_i}z_i \quad (4.34)$$

$$= b_0 + \sum_{i=1}^n b_i z_i \quad (4.35)$$

donde b_i es el coeficiente i -ésimo del hiperplano transformado de $G_{\mathbf{X}}(\mathbf{X})$ al espacio $G_{\mathbf{Z}}(\mathbf{Z})$.

Calculando el vector normal al nuevo hiperplano se tiene que:

$$\boldsymbol{\alpha}^T = \frac{1}{\sqrt{\sum_{i=1}^n b_i^2}} (b_1, b_2, \dots, b_n) \quad (4.36)$$

Sustituyendo esta expresión en (4.35), resulta:

$$g_{\mathbf{z}}(\mathbf{z}) = b_0 + \sqrt{\sum_{i=1}^n b_i^2} \boldsymbol{\alpha}^T \mathbf{z} \quad (4.37)$$

Dado que en el punto óptimo se ha de cumplir (4.13) se llega a la siguiente expresión:

$$g_{\mathbf{z}}(\mathbf{z}^*) = \frac{b_0}{\sqrt{\sum_{i=1}^n b_i^2}} + \boldsymbol{\alpha}^T \mathbf{z}^* = 0, \quad (4.38)$$

en la que se ha dividido la expresión (4.37) por $\sqrt{\sum_{i=1}^n b_i^2}$, que al estar igualada a cero no varía el resultado.

Despejando \mathbf{z}^* , resulta

$$\mathbf{z}^* = -\boldsymbol{\alpha} \frac{b_0}{\sqrt{\sum_{i=1}^n b_i^2}},$$

identificando términos entre esta última expresión y (4.32) llegamos a conclusión de que β es igual a

$$\beta = \frac{b_0}{\sqrt{\sum_{i=1}^n b_i^2}} \quad (4.39)$$

y teniendo en cuenta (4.35)

$$\beta = \frac{a_0 + \sum_{i=1}^n a_i \mu_{X_i}}{\sqrt{\sum_{i=1}^n a_i^2 \sigma_{X_i}^2}} \quad (4.40)$$

que como no podía ser de otra manera es idéntica a las expresiones (4.21) y (4.14).

Cuando la ecuación de estado límite es no lineal no se pueden obtener los dos primeros momentos de $G_{\mathbf{X}}(\mathbf{X})$ en el espacio \mathbf{X} , ni en el espacio \mathbf{Z} . Esto se debe a que una combinación no lineal de variables normales no genera otra variable normal. Por tanto se procederá a realizar un desarrollo en serie de Taylor en el entorno del punto de diseño, para lo cual se ha de resolver el problema (4.12)-(4.13).

Introduciendo la función Lagrangiana el problema se transforma en:

$$\mathcal{L}(\mathbf{z}, \lambda) = \underset{\mathbf{z}, \lambda}{\text{Mínimo}} \sqrt{\mathbf{z}^T \mathbf{z}} + \lambda g_{\mathbf{Z}}(\mathbf{z}) \quad (4.41)$$

donde λ es el multiplicador de Lagrange asociado a la restricción (4.13).

Derivando con respecto a las variables \mathbf{z} y λ e igualando a cero:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial z_i} &= z_i (\mathbf{z}^T \mathbf{z})^{(-1/2)} + \lambda \frac{\partial g_{\mathbf{Z}}}{\partial z_i} = 0; \quad i = 1, \dots, n \\ \frac{\partial \mathcal{L}}{\partial \lambda} &= g_{\mathbf{Z}}(\mathbf{z}) = 0 \end{aligned} \quad (4.42)$$

que de forma compacta queda:

$$0 = \mathbf{z} \delta^{-1} + \lambda \nabla g_{\mathbf{Z}}(\mathbf{z}) \quad (4.43)$$

$$0 = g_{\mathbf{Z}}(\mathbf{z}) \quad (4.44)$$

donde $\nabla g_{\mathbf{Z}}(\mathbf{z}) = (\partial g_{\mathbf{Z}}/\partial z_1, \partial g_{\mathbf{Z}}/\partial z_2, \dots, \partial g_{\mathbf{Z}}/\partial z_n)^T$ y $\delta = (\mathbf{z}^T \mathbf{z})^{1/2}$. La expresión (4.44) se cumple por definición, por lo tanto despejando \mathbf{z} de la expresión (4.43) (véase Horne y Price [97]):

$$\mathbf{z}^* = -\lambda \nabla g_{\mathbf{Z}}(\mathbf{z}) \delta \quad (4.45)$$

Si el punto obtenido es un máximo, mínimo o un punto de silla depende de la función $\nabla g_{\mathbf{Z}}(\mathbf{z})$. Suponiendo que es un mínimo e identificando términos de la fórmula (4.45) con los de la expresión (4.32) se tiene que $\lambda = \pm (\nabla g_{\mathbf{Z}}(\mathbf{z})^T \nabla g_{\mathbf{Z}}(\mathbf{z}))^{-1/2}$ y $\delta = \beta$, despejando β , se llega a la siguiente expresión:

$$\beta = \frac{-\mathbf{z}^{*T} \nabla g_{\mathbf{Z}}(\mathbf{z})}{(\nabla g_{\mathbf{Z}}(\mathbf{z})^T \nabla g_{\mathbf{Z}}(\mathbf{z}))^{1/2}} = -\mathbf{z}^{*T} \boldsymbol{\alpha} \quad (4.46)$$

Proof. Se demostrará mediante el desarrollo en serie de Taylor que efectivamente el β calculado en (4.46) es la mínima distancia del origen a la región de fallo. Linearizando la ecuación de estado límite $g(\cdot)$ en el entorno del supuesto punto de diseño \mathbf{z}^* :

$$g_{\mathbf{Z}}(\mathbf{z}) \approx g_{\mathbf{Z}}(\mathbf{z}^*) + (\mathbf{z} - \mathbf{z}^*)^T \nabla g_{\mathbf{Z}}(\mathbf{z}) \quad (4.47)$$

como $g_{\mathbf{Z}}(\mathbf{z}^*)$ está sobre la ecuación de estado límite $g_{\mathbf{Z}}(\mathbf{z}^*) = 0$, con lo cual

$$g_{\mathbf{Z}}(\mathbf{z}) \approx -\mathbf{z}^{*T} \nabla g_{\mathbf{Z}}(\mathbf{z}) + \mathbf{z}^T \nabla g_{\mathbf{Z}}(\mathbf{z}) \quad (4.48)$$

que es una combinación lineal de variables normales, por lo que pueden calcularse los dos primeros momentos de la aproximación lineal de $g_{\mathbf{Z}}(\mathbf{z})$:

$$\mu_{g_{\mathbf{Z}}(\mathbf{z})} = -\mathbf{z}^{*T} \nabla g_{\mathbf{Z}}(\mathbf{z}) \quad (4.49)$$

$$\sigma_{g_{\mathbf{Z}}(\mathbf{z})}^2 = \nabla g_{\mathbf{Z}}(\mathbf{z})^T \nabla g_{\mathbf{Z}}(\mathbf{z}) \quad (4.50)$$

Como $\beta = \mu_{g_{\mathbf{Z}}(\mathbf{z})} / \sigma_{g_{\mathbf{Z}}(\mathbf{z})}$ se llega a la conclusión de que:

$$\beta = \frac{\mu_{g_{\mathbf{Z}}(\mathbf{z})}}{\sigma_{g_{\mathbf{Z}}(\mathbf{z})}} = \frac{-\mathbf{z}^{*T} \nabla g_{\mathbf{Z}}(\mathbf{z})}{(\nabla g_{\mathbf{Z}}(\mathbf{z})^T \nabla g_{\mathbf{Z}}(\mathbf{z}))^{1/2}} = -\mathbf{z}^{*T} \boldsymbol{\alpha} \quad (4.51)$$

que es idéntica a la expresión (4.46), con lo que queda demostrado. ■

4.6.4 Sensitivity factors

Los cosenos directores α_i calculados en la sección anterior representan la sensibilidad de la función de estado límite $g_{\mathbf{z}}(\mathbf{z})$ en \mathbf{z}^* a cambios en la variable z_i (Hohenbichler y Rackwitz [96] y Bjerager y Krenk [22]). Es muy importante para reducir la dimensionalidad de los problemas, y para estudiar que variables tienen más influencia en la fiabilidad de un sistema. De tal forma que si el valor de la sensibilidad es pequeño se puede tratar la variable asociada como determinista.

4.6.5 Numerical solution for non-linear limit state functions

En muchas situaciones, principalmente cuando el número de variables es muy grande o se tienen ecuaciones de estado límite muy complejas, es preciso recurrir a soluciones numéricas. Si las restricciones son lineales el problema se reduce a un problema de programación cuadrática, para los que existen algoritmos muy eficientes. Si la restricción $g_{\mathbf{z}}(\mathbf{z}) = 0$ es no lineal hay varios algoritmos que funcionan bien con ese tipo de metodologías (Liu y Der Kiureghian [112]). Uno de los que mejor funciona es el método del gradiente proyectado, originalmente desarrollado por Rosen [145], éste es una modificación del método del descenso para optimización no restringida. Los métodos de penalización no son adecuados para problemas de fiabilidad. El método del lagrangiano aumentado se comporta mejor que los métodos de penalización, pero son complicados de implementar, sobre todo a la hora de dar valores iniciales a los multiplicadores de Lagrange de las restricciones. Los métodos de programación cuadrática secuencial fueron introducidos en el campo de la fiabilidad por Shinozuka [155]. Schittkowski [153] basándose en el método de Han, Powel y Wilson (véase Gill, Murray y Wright [87]) propuso una rutina basada en programación secuencial cuadrática que se ha hecho muy popular en el campo de la fiabilidad. La idea esencial del método es reemplazar el Hessiano del Lagrangiano por una matriz aproximada, y usar la función Lagrangiana aumentada para determinar el tamaño de paso. Otro algoritmo de programación secuencial cuadrática ampliamente extendido en el campo de la fiabilidad estructural es el debido a Powell [138].

A continuación se va a desarrollar el método propuesto por Hasofer y Lind [92], que se puede considerar más un procedimiento iterativo que una clase de algoritmo.

Algorithm 4.1 (Método de Hasofer-Lind (FOSM)). Partiendo de la aproximación lineal de la función $g_{\mathbf{z}}(\mathbf{z})$ en el punto $\mathbf{z}^{(i)}$:

$$g_{\mathbf{z}}(\mathbf{z}^{(i+1)}) \approx g_{\mathbf{z}}(\mathbf{z}^{(i)}) + (\mathbf{z}^{(i+1)} - \mathbf{z}^{(i)})^T \nabla g_{\mathbf{z}}(\mathbf{z}^{(i)}) = 0, \quad (4.52)$$

donde $\nabla g_{\mathbf{z}}(\mathbf{z}^{(i)}) = \partial g_{\mathbf{z}} / \partial \mathbf{z}^{(i)}$. Y teniendo en cuenta que:

$$\mathbf{z}^{(i)} = -\boldsymbol{\alpha}^{(i)} \beta^{(i)}; \quad \boldsymbol{\alpha}^{(i)} = \frac{\nabla g_{\mathbf{z}}(\mathbf{z}^{(i)})}{\left(\nabla g_{\mathbf{z}}(\mathbf{z}^{(i)})^T \nabla g_{\mathbf{z}}(\mathbf{z}^{(i)})\right)^{1/2}} \quad (4.53)$$

Sustituyendo (4.53) en (4.52) y reorganizando la expresión se obtiene la siguiente fórmula

recurrente:

$$\mathbf{z}^{(i+1)} = -\boldsymbol{\alpha}^{(i)} \left(\beta^{(i)} + \frac{g_{\mathbf{Z}}(\mathbf{z}^{(i)})}{\left(\nabla g_{\mathbf{Z}}(\mathbf{z}^{(i)})^T \nabla g_{\mathbf{Z}}(\mathbf{z}^{(i)}) \right)^{1/2}} \right) \quad (4.54)$$

El algoritmo queda por tanto:

- **Entrada:** Conjunto inicial de variables $\{X_1, X_2, \dots, X_n\}$, su función de densidad conjunta $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$, y la región de fallo $g_{\mathbf{X}}(x_1, x_2, \dots, x_n) \leq 0$.
- **Salida:** La probabilidad de fallo asociada a la región de fallo dada.

Paso 1: Transformación de las variables $\{X_1, X_2, \dots, X_n\}$ al espacio multinormal estándar $\{Z_1, Z_2, \dots, Z_n\}$.

Paso 2: Transformación de la región de fallo $g_{\mathbf{X}}(\mathbf{x}) = 0$ a $g_{\mathbf{Z}}(\mathbf{z}) = 0$.

Paso 3: Selección del punto de inicio $(\mathbf{x}^{(1)}, \mathbf{z}^{(1)})$.

Paso 4: Obtención de la distancia del punto $\mathbf{z}^{(i)}$ al origen, éste será la aproximación i -ésima del índice de fiabilidad, $\beta^{(i)} = \sqrt{\mathbf{z}^{(i)T} \mathbf{z}^{(i)}}$.

Paso 5: Obtención de $\nabla g_{\mathbf{Z}}(\mathbf{z}^{(i)})$ y los cosenos directores $\boldsymbol{\alpha}^{(i)}$ del vector normal a la región de fallo $g_{\mathbf{Z}}(\mathbf{z}^{(i)}) = 0$ mediante la fórmula (4.53).

Paso 6: Cálculo de $g_{\mathbf{Z}}(\mathbf{z}^{(i)})$.

Paso 7: Obtención de la nueva aproximación del punto de diseño $\mathbf{z}^{(i+1)}$ mediante la fórmula (4.54) y el índice de fiabilidad $\beta^{(i+1)} = \sqrt{\mathbf{z}^{(i+1)T} \mathbf{z}^{(i+1)}}$.

Paso 8: Si el valor de β o de $g_{\mathbf{Z}}(\mathbf{z})$ se estabiliza, se para el procedimiento. Si no, incrementar i en una unidad y continuar con el algoritmo en el **paso 5**.

En la figura 4.4 se muestra la interpretación geométrica de cada una de las etapas del procedimiento iterativo. ■

También es posible transformar la relación (4.54) y expresarla en el espacio original \mathbf{X} , con lo que no se requiere la transformación al espacio \mathbf{Z} (véase Parkinson [135] y Li y Lumb [108]):

$$\mathbf{x}^{(i+1)} = \boldsymbol{\mu}_X - \boldsymbol{\sigma}_X \nabla g_{\mathbf{X}}(\mathbf{x})^{(i)} \frac{(\mathbf{x}^{(i)} - \boldsymbol{\mu}_X)^T \nabla g_{\mathbf{X}}(\mathbf{x})^{(i)}}{\nabla g_{\mathbf{X}}(\mathbf{x})^{(i)T} \boldsymbol{\sigma}_X \nabla g_{\mathbf{X}}(\mathbf{x})^{(i)}} \quad (4.55)$$

donde $\nabla g_{\mathbf{X}}(\mathbf{x}) = (\partial g_{\mathbf{X}}/\partial x_1, \partial g_{\mathbf{X}}/\partial x_2, \dots, \partial g_{\mathbf{X}}/\partial x_n)^T$, $\boldsymbol{\mu}_X$ es el vector de medias de las variables aleatorias y $\boldsymbol{\sigma}_X$ es la matriz de varianzas-covarianzas.

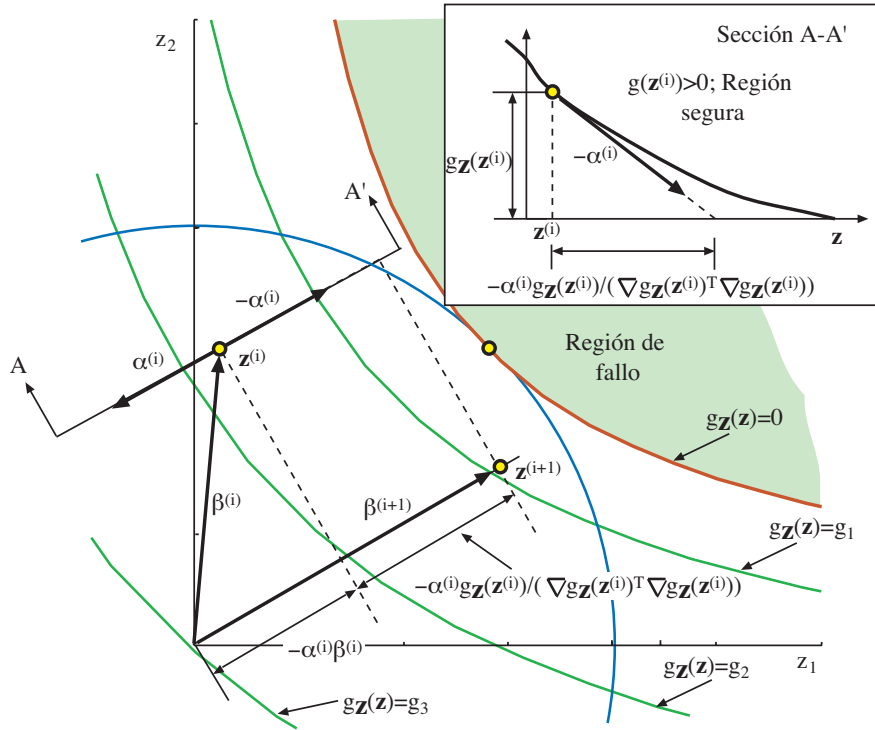


Figure 4.4: Ilustración gráfica del método de Hasofer-Lind para el caso bidimensional, en la que se aprecia la interpretación geométrica de todas las variables que intervienen.

4.6.6 Failure probability bounds

La teoría FOSM se centraba en obtener un valor indicativo de la fiabilidad del sistema, que además fuera invariante. Pero, ¿cómo se puede relacionar el índice de fiabilidad β obtenido con la probabilidad de superación o fallo del estado límite?

Todo depende de la forma de la función de estado límite, ya que si es lineal, se obtiene la probabilidad de fallo exacta sin más que aplicar $p_f = \Phi(-\beta)$. En cambio si se aplica esta formulación a una ecuación de estado límite no lineal, pueden pasar dos cosas: si la región de fallo es convexa con respecto a origen (véase la figura 4.6) la aproximación es un límite inferior de la probabilidad de fallo, y si es cóncava como en la figura 4.5 es una cota superior.

En el caso de que la función sea cóncava y con una curvatura muy grande (véase la figura 4.5) se puede calcular una cota superior de la probabilidad de fallo. En este caso se aproxima la región de fallo por una hiperesfera con centro en el origen y radio el índice de fiabilidad β . Como $\sum_{i=1}^n Z_i^2 \sim \chi_n^2$

$$p_f \leq P\left(\sum_{i=1}^n Z_i^2 > \beta^2\right) = 1 - F_{\chi_n^2}(\beta^2), \quad (4.56)$$

donde $F_{\chi_n^2}$ es la función de distribución de la variable aleatoria χ^2 con n grados de libertad.

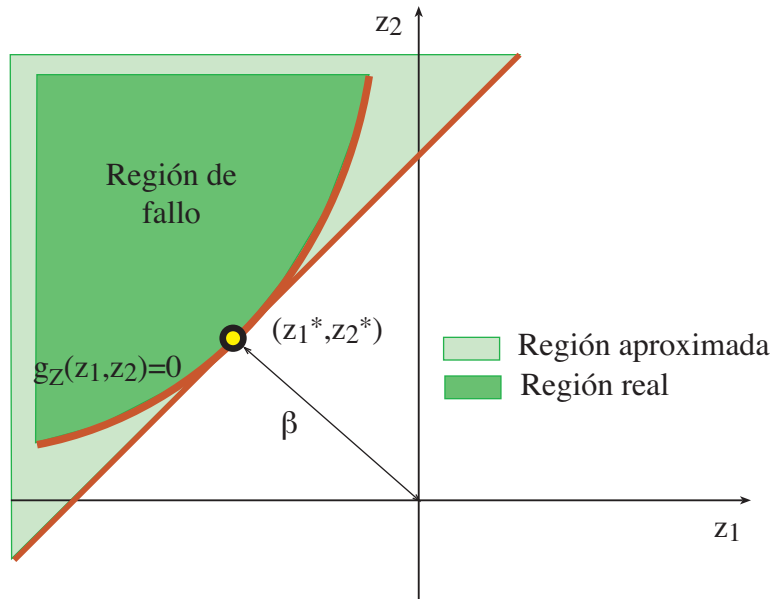


Figure 4.5: Cota superior de la probabilidad de fallo cuando la región de fallo es convexa con respecto al origen.

Nótese que esta cota está muy del lado de la seguridad, por eso apenas se utiliza.

Sin embargo, se han propuesto otros muchos métodos, como los de Freudenthal [81], Hasofer y Lind [92], Breitung [25], Castillo, Solares, y Gómez [40, 39, 41], y Haskin, Staple y Ding [91], por ejemplo.

4.7 First-Order Reliability Methods (FORM)

En la sección 4.6 hemos visto la metodología empleada para calcular la probabilidad de fallo de un sistema considerando sólo los dos primeros momentos de las variables aleatorias. En esta sección se estudiará la aproximación lineal del problema pero trabajando con las distribuciones reales de las variables.

La metodología es exactamente la misma que en el método FOSM lo único que cambia es la transformación de las variables aleatorias $\mathbf{X} = (X_1, X_2, \dots, X_n)$ al espacio normal estándar multivariado $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)$, que al trabajar con las funciones de densidad exactas requieran métodos de transformación especiales.

4.7.1 Variable transformations

La transformación de una variable aleatoria X independiente con una distribución cualesquiera en una variable Z normalmente distribuida se conoce como transformación normal y se puede expresar matemáticamente como:

$$F_X(x) = \Phi(z) \text{ o bien } z = \Phi^{-1}(F_X(x)) \quad (4.57)$$

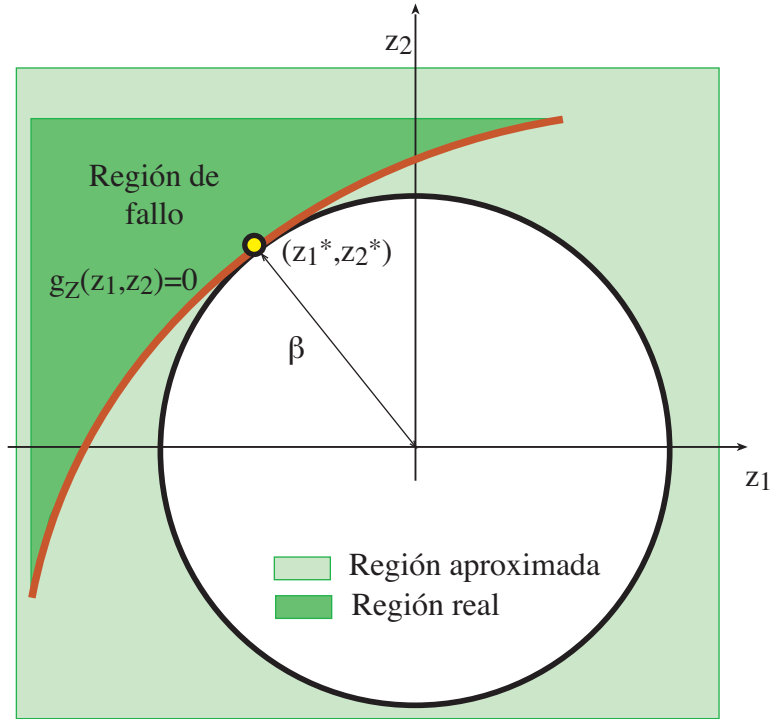


Figure 4.6: Cota superior de la probabilidad de fallo cuando la región de fallo es cóncava con respecto al origen.

donde $F_X(x)$ es la función de distribución de la variable X y Φ es la función de distribución de la variable normal estándar $Z \sim N(0, 1)$.

Obsérvese que la transformación normal requiere una transformación intermedia en la variable uniforme $U = F_X(X)$ (véase la figura ??).

En el caso de que las variables no sean independientes se puede aplicar la transformación de Rosenblatt [146]:

$$\begin{aligned}
 u_1 &= F_1(x_1) \\
 u_2 &= F_2(x_2|x_1) \\
 &\vdots \\
 u_n &= F_n(x_n|x_1, x_2, \dots, x_{n-1}),
 \end{aligned}
 \tag{4.58}$$

con lo que se obtienen las variables uniformes independientes $U(0, 1)$, primer paso de la figura ?. Posteriormente se aplica la transformación normal a las variables uniformes y se obtienen las Z , segundo paso de la figura 4.7:

$$\begin{aligned}
 z_1 &= \Phi^{-1}(F_1(x_1)) \\
 z_2 &= \Phi^{-1}(F_2(x_2|x_1)) \\
 &\vdots \\
 z_n &= \Phi^{-1}(F_n(x_n|x_1, x_2, \dots, x_{n-1})).
 \end{aligned}
 \tag{4.59}$$

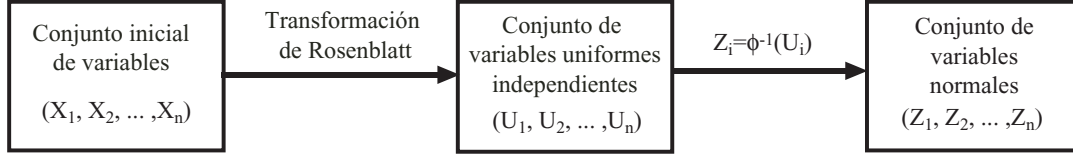


Figure 4.7: Ilustración de cómo se transforma el conjunto inicial de variables en un conjunto de variables uniformes independientes primero, y en un conjunto de variables normales independientes estándar después.

donde $F_1(x_1), F_2(x_2|x_1), \dots, F_n(x_n|x_1, x_2, \dots, x_{n-1})$ son las funciones de distribución marginales de X_1 y de las variables condicionales indicadas, respectivamente.

Antes de incorporar la transformación de las variables al algoritmo, es necesario transformar la ecuación de estado límite de $g_{\mathbf{X}}(\mathbf{x})$ a $g_{\mathbf{Z}}(\mathbf{z})$ y calcular el jacobiano \mathbf{J} de la transformación (4.59) cuyos elementos son de la forma:

$$j_{ij} = \frac{\partial z_j}{\partial x_i} = \frac{1}{\phi(z_j)} \frac{\partial F_j(x_j|x_1, x_2, \dots, x_{j-1})}{\partial x_i}. \quad (4.60)$$

Evidentemente, si $i > j$, $\partial F_j/\partial x_i = 0$, por lo tanto la matriz jacobiana es triangular superior, y puede calcularse su inversa por simple sustitución hacia atrás.

El cálculo de los gradientes $\nabla g_{\mathbf{Z}}(\mathbf{z})$ puede ser difícil porque en general no se dispone de una expresión explícita de $g_{\mathbf{Z}}(\mathbf{z})$. En esos casos se procede de la siguiente manera, teniendo en cuenta que:

$$\frac{\partial g_{\mathbf{X}}(\mathbf{x})}{\partial x_j} = \sum_{i=1}^n \frac{\partial g_{\mathbf{Z}}(\mathbf{z})}{\partial z_i} \frac{\partial z_i}{\partial x_j} = [\mathbf{J}] \frac{\partial g_{\mathbf{Z}}(\mathbf{z})}{\partial \mathbf{z}} \quad (4.61)$$

de tal forma que

$$\begin{pmatrix} \frac{\partial g_{\mathbf{X}}(\mathbf{x})}{\partial x_1} \\ \frac{\partial g_{\mathbf{X}}(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial g_{\mathbf{X}}(\mathbf{x})}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_2}{\partial x_1} & \cdots & \frac{\partial z_n}{\partial x_1} \\ \frac{\partial z_1}{\partial x_2} & \frac{\partial z_2}{\partial x_2} & \cdots & \frac{\partial z_n}{\partial x_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial z_1}{\partial x_n} & \frac{\partial z_2}{\partial x_n} & \cdots & \frac{\partial z_n}{\partial x_n} \end{pmatrix} \begin{pmatrix} \frac{\partial g_{\mathbf{Z}}(\mathbf{z})}{\partial z_1} \\ \frac{\partial g_{\mathbf{Z}}(\mathbf{z})}{\partial z_2} \\ \vdots \\ \frac{\partial g_{\mathbf{Z}}(\mathbf{z})}{\partial z_n} \end{pmatrix} \quad (4.62)$$

calculando la inversa del jacobiano

$$\begin{pmatrix} \frac{\partial g_{\mathbf{z}}(\mathbf{z})}{\partial z_1} \\ \frac{\partial g_{\mathbf{z}}(\mathbf{z})}{\partial z_2} \\ \vdots \\ \frac{\partial g_{\mathbf{z}}(\mathbf{z})}{\partial z_n} \end{pmatrix} = \mathbf{J}^{-1} \begin{pmatrix} \frac{\partial g_{\mathbf{x}}(\mathbf{x})}{\partial x_1} \\ \frac{\partial g_{\mathbf{x}}(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial g_{\mathbf{x}}(\mathbf{x})}{\partial x_n} \end{pmatrix} \quad (4.63)$$

En los casos en los que no se dispone de la función de densidad conjunta de las variables que intervienen, sino que se tienen las funciones de distribución marginales de cada una de las variables $F_{\mathbf{x}}(\mathbf{x})$ y la matriz de correlación ρ no se puede aplicar Rosenblatt, pero se dispone de la transformación de Nataf [128].

4.7.2 FORM algorithm

Se puede extender el algoritmo 4.1 para determinar el punto de diseño y el índice de fiabilidad β de la teoría FOSM al caso FORM (véase Hohenbichler y Rackwitz [95]).

Algorithm 4.2 (Método FORM).

- **Entrada:** Conjunto inicial de variables $\{X_1, X_2, \dots, X_n\}$, su función de densidad conjunta $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n)$, y la región de fallo $g_{\mathbf{x}}(x_1, x_2, \dots, x_n) \leq 0$.
- **Salida:** La probabilidad de fallo asociada a la región de fallo dada y el punto de diseño.

Paso 1: Iniciación $i = 1$ y selección del punto de inicio $\mathbf{x}^{(i)}$.

Paso 2: Transformación de las variables $\{X_1, X_2, \dots, X_n\}$ al espacio multinormal estándar $\{Z_1, Z_2, \dots, Z_n\}$ mediante (4.57) o (4.59) y obtención de $\mathbf{z}^{(i)}$.

Paso 3: Obtención del jacobiano \mathbf{J} mediante la expresión (4.60), y su inversa \mathbf{J}^{-1} por sustitución hacia atrás.

Paso 4: Cálculo de los cosenos directores $\boldsymbol{\alpha}^{(i)}$ de acuerdo a (4.62) y (4.31), evaluación de $g_{\mathbf{x}}(\mathbf{x}^{(i)}) = g_{\mathbf{z}}(\mathbf{z}^{(i)})$ y estimación de $\beta^{(i)}$ conforme a (4.32).

Paso 5: Obtención de la nueva aproximación del punto de diseño $\mathbf{z}^{(i+1)}$ mediante la fórmula (4.54) y de su índice de fiabilidad $\beta^{(i+1)} = \sqrt{\mathbf{z}^{(i+1)T} \mathbf{z}^{(i+1)}}$. Se puede calcular el punto $\mathbf{x}^{(i+1)}$ usando la transformación inversa de (4.59).

Paso 6: Si el valor de β o de $g_{\mathbf{z}}(\mathbf{z})$ se estabiliza, se para el procedimiento. Si no, incrementar i en una unidad y continuar con el algoritmo en el **paso 3**.

■

Computational Example 4.1 (Método FORM). El siguiente ejemplo se encuentra en Melchers [123] y está adaptado de Dolinsky [66], y Hohenbichler y Rackwitz [95]. Es uno de los pocos casos en los que se puede tratar analíticamente todo el problema. Considérese la ecuación de estado límite

$$g_{\mathbf{x}}(\mathbf{x}) = 6 - 2x_1 - x_2$$

donde \mathbf{x} son variables aleatorias con la siguiente función de densidad conjunta:

$$f_{\mathbf{x}}(\mathbf{x}) = \begin{cases} (ab - 1 + ax_1 + bx_2 + x_1x_2) \exp(-ax_1 - bx_2 - x_1x_2) & \text{si } x_1, x_2 \geq 0 \\ 0 & \text{en otro caso.} \end{cases} \quad (4.64)$$

Las funciones de densidad marginales se obtienen integrando de la forma siguiente:

$$f_{X_1}(x_1) = \int_{x_2=0}^{\infty} f_{\mathbf{x}}(x_1, x_2) dx_2 = a \exp(-ax_1) \quad (4.65)$$

$$f_{X_2}(x_2) = \int_{x_1=0}^{\infty} f_{\mathbf{x}}(x_1, x_2) dx_1 = b \exp(-bx_2), \quad (4.66)$$

la transformación de Rosenblatt particularizada para este ejemplo será:

$$\Phi(z_1) = F_1(x_1) = \int_{w=0}^{x_1} f_{X_1}(w) dw = 1 - \exp(-ax_1); \quad \forall x_1 \geq 0 \quad (4.67)$$

$$\begin{aligned} \Phi(z_2) = F_2(x_2|x_1) &= \frac{\int_0^{x_2} f_{\mathbf{x}}(x_1, w) dw}{f_{X_1}(x_1)} \\ &= 1 - \left(1 + \frac{x_2}{a}\right) \exp(-bx_2 - x_1x_2). \end{aligned} \quad (4.68)$$

Por lo que la ecuación de estado límite en función de \mathbf{z} (véase la figura 4.8) queda como:

$$g_{\mathbf{z}}(\mathbf{z}) = 6 - 2F_1^{-1}(\Phi(z_1)) - F_2^{-1}(\Phi(z_2)) \quad (4.69)$$

El jacobiano de la expresión (4.62) queda de la siguiente manera:

$$\mathbf{J} = \begin{pmatrix} \frac{\partial z_1}{\partial x_1} & \frac{\partial z_2}{\partial x_1} \\ \frac{\partial z_1}{\partial x_2} & \frac{\partial z_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\phi(z_1)} \frac{\partial F_1(x_1)}{\partial x_1} & \frac{1}{\phi(z_2)} \frac{\partial F_2(x_2|x_1)}{\partial x_1} \\ \frac{1}{\phi(z_1)} \frac{\partial F_1(x_1)}{\partial x_2} & \frac{1}{\phi(z_2)} \frac{\partial F_2(x_2|x_1)}{\partial x_2} \end{pmatrix} \quad (4.70)$$

Calculando las expresiones analíticas correspondientes se tiene:

$$\mathbf{J} = \begin{pmatrix} \frac{1}{\phi(z_1)} \frac{a}{\exp(ax_1)} & \frac{1}{\phi(z_2)} \frac{x_2(a+x_2)}{a \exp(x_2(b+x_1))} \\ 0 & \frac{1}{\phi(z_2)} \frac{(b+x_1)(a+x_2)-1}{a \exp(x_2(b+x_1))} \end{pmatrix} \quad (4.71)$$

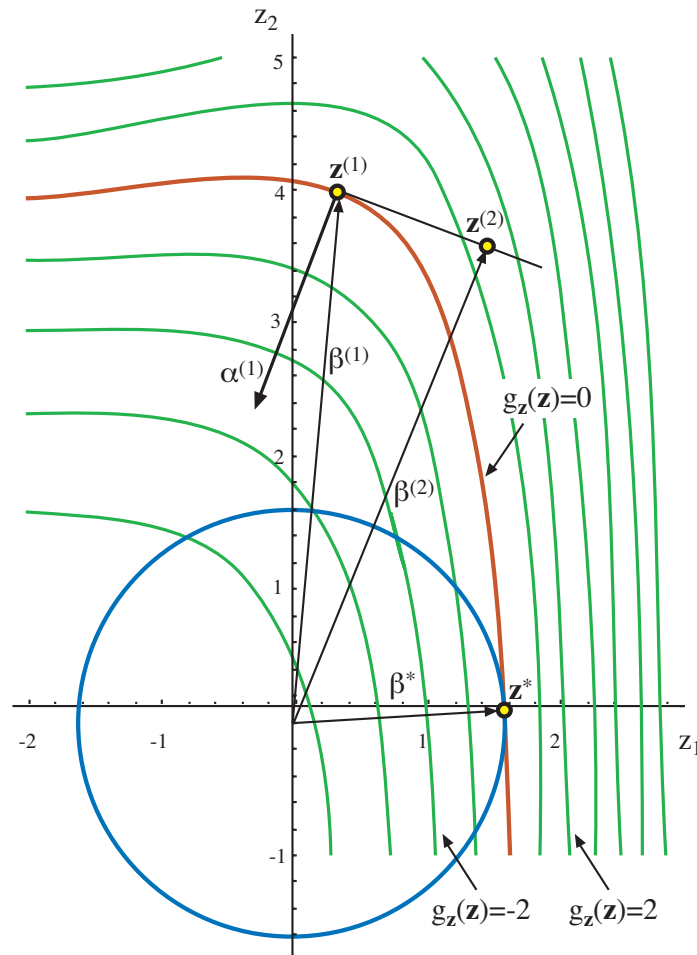


Figure 4.8: Ilustración gráfica del ejemplo computacional 4.1 en el que muestran diferentes contornos de la ecuación de estado límite $g_{\mathbf{z}}(\mathbf{z})$, el punto de inicio $\mathbf{z}^{(1)}$, el correspondiente a la segunda iteración $\mathbf{z}^{(2)}$ y el punto solución \mathbf{z}^* .

mientras que la inversa, calculada por sustitución hacia atrás es

$$\mathbf{J}^{-1} = \begin{pmatrix} \phi(z_1) \frac{\exp(ax_1)}{a} & -\phi(z_2) \frac{\exp(ax_1)x_2(a+x_2)}{a((a+x_2)(b+x_1)-1)} \\ 0 & \phi(z_2) \frac{a \exp(b+x_1)x_2}{(a+x_2)(b+x_1)-1} \end{pmatrix} \quad (4.72)$$

Supóngase que $a = 1$ y $b = 2$, el algoritmo 4.2 procedería de la siguiente manera:

1. Iniciación $i = 1$ y selección del punto de inicio $\mathbf{x}^{(1)} = (1, 4)^T$.
2. Transformación del vector $\mathbf{x}^{(1)}$ al $\mathbf{z}^{(1)} = (0.337475, 4.0072)^T$ (véase la figura 4.8) mediante (4.67) y (4.68).
3. Obtención del jacobiano \mathbf{J} y su inversa \mathbf{J}^{-1} mediante las expresiones (4.71) y (4.72),

$$\mathbf{J} = \begin{pmatrix} 0.976172 & 0.945074 \\ 0 & 0.661552 \end{pmatrix}; \quad \mathbf{J}^{-1} = \begin{pmatrix} 1.02441 & -1.46344 \\ 0 & 1.5116 \end{pmatrix}.$$

4. Cálculo de los cosenos directores $\boldsymbol{\alpha}^{(1)}$ (véase la figura 4.8) de acuerdo a (4.62) y (4.31), evaluación de $g_{\mathbf{x}}(\mathbf{x}^{(1)}) = g_{\mathbf{z}}(\mathbf{z}^{(1)})$ y estimación de $\beta^{(1)}$ conforme a (4.32).

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial g_{\mathbf{z}}(\mathbf{z})}{\partial z_1} \\ \frac{\partial g_{\mathbf{z}}(\mathbf{z})}{\partial z_2} \end{pmatrix} = \mathbf{J}^{-1} \begin{pmatrix} \frac{\partial g_{\mathbf{x}}(\mathbf{x})}{\partial x_1} \\ \frac{\partial g_{\mathbf{x}}(\mathbf{x})}{\partial x_2} \end{pmatrix} = \mathbf{J}^{-1} \begin{pmatrix} -2 \\ -1 \end{pmatrix} = \begin{pmatrix} -0.585377 \\ -1.5116 \end{pmatrix}$$

de tal forma que como $\ell = 1.62099$,

$$\boldsymbol{\alpha}^{(1)} = \begin{pmatrix} -0.361124 \\ -0.932518 \end{pmatrix}; \quad g_{\mathbf{x}}(\mathbf{x}^{(1)}) = 0; \quad \beta^{(1)} = (0.337475^2 + 4.0072^2)^{(1/2)} = 4.02139.$$

5. Obtención de la nueva aproximación del punto de diseño $\mathbf{z}^{(2)}$ mediante la fórmula (4.54). Se puede calcular el punto $\mathbf{x}^{(2)}$ usando la transformación inversa de (4.59),

$$\mathbf{z}^{(2)} = -\boldsymbol{\alpha}^{(1)} * \left(\beta^{(1)} + g_{\mathbf{x}}(\mathbf{x}^{(1)})/\ell \right) = (1.45222, 3.75002)^T$$

y

$$\mathbf{x}^{(2)} = (2.61428, 2.28018)^T.$$

El algoritmo se repite hasta que los valores de \mathbf{x} o de \mathbf{z} se estabilizan, en la tabla 4.2 se muestra la evolución de las variables durante la ejecución del algoritmo. Los valores solución correspondientes a la última fila de la tabla son:

$$\mathbf{x}^* = (2.90434, 0.191314)^T, \quad \mathbf{z}^* = (1.60013, 0.084909)^T \text{ y } \beta^* = 1.60238.$$

■

Table 4.2: Evolución del algoritmo 4.2 para el ejemplo computacional 4.1.

i	$x_1^{(i)}$	$x_2^{(i)}$	$z_1^{(i)}$	$z_2^{(i)}$	$\beta^{(i)}$	$gz(\mathbf{z}^{(i)})$
1	4.00000	1.000000	0.33747	4.007200	4.02139	0.00000
2	2.61428	2.280180	1.45222	3.750020	3.50844	-1.50875
3	7.74292	0.226621	3.33027	1.103810	2.13942	-9.71246
4	4.12096	0.142202	2.13873	0.054346	1.65564	-2.38413
5	3.01625	0.179809	1.65478	0.053338	1.60332	-0.21232
6	2.90687	0.189931	1.60138	0.078868	1.60239	-3.68037×10^{-3}
7	2.90442	0.191209	1.60017	0.084425	1.60238	-4.32695×10^{-5}
8	2.90435	0.191307	1.60013	0.084875	1.60238	-2.64198×10^{-7}
9	2.90434	0.191314	1.60013	0.084909	1.60238	-1.32413×10^{-9}

Este mismo problema se podría haber resuelto mediante algún método de optimización, a continuación se presenta el código perteneciente al programa de optimización GAMS (General Algebraic Modelling System) para resolver el problema con el método del gradiente reducido generalizado.

```

$title Ejemplo computacional

file out /libro.out/;
put out;

SETS
  I numero de variables aleatorias /1*2/;
SCALARS
  pi /3.1415926535898/;
PARAMETERS
  a /1.0/
  b /2.0/;
VARIABLES
  beta indice de fiabilidad
  pf probabilidad de fallo (FORM)
  z(I) variables aleatorias normales;
POSITIVE VARIABLES
  x(I) variables aleatorias;
EQUATIONS
  betadef definicion del indice de fiabilidad
  Zdef1 transformacion de Rosenblatt para la variable 1
  Zdef2 transformacion de Rosenblatt para la variable 2
  verdef ecuacion de estado limite;

betadef..beta=e=sqrt(sum(I,sqr(z(I))));

```

```

Zdef1..errorf(z('1'))=e=1-exp(-a*x('1'));
Zdef2..errorf(z('2'))=e=1-(1+x('2')/a)*exp(-b*x('2')-x('1')*x('2'));
verdef..0=e=6-2*x('1')-x('2');

MODEL ejemplo/betadef,Zdef1,Zdef2,verdef/;

* Valores iniciales de las variables
x.l('1')=1.0;
x.l('2')=4.0;

put "Valores iniciales"//;
loop(I,
  put "x(",I.tl:1,")= ",x.l(I):12:8/;
);

*Resolucion del modelo
SOLVE ejemplo USING nlp MINIMIZING beta;

put "pf= ",(errorf(-beta.l)):12:8,", modelstat= ",ejemplo.modelstat,
  ", solvestat= ",ejemplo.solvestat/;
put "beta= ",(beta.l):12:8/;
loop(I,
  put "z(",I.tl:1,")= ",z.l(I):12:8/;
);
loop(I,
  put "x(",I.tl:1,")= ",x.l(I):12:8/;
);

```

Tras ejecutar se obtendrá la solución en el fichero 'libro.out', que tendrá la forma siguiente:

Valores iniciales

```

x(1)= 1.00000000
x(2)= 4.00000000
pf= 0.05453556, modelstat= 2.00, solvestat= 1.00
beta= 1.60238010
z(1)= 1.60012890
z(2)= 0.08490869
x(1)= 2.90434319
x(2)= 0.19131363

```

Como se puede comprobar la solución es la misma que la del ejemplo computacional 4.1 pero se observa la simplicidad de utilizar un paquete de optimización estándar.

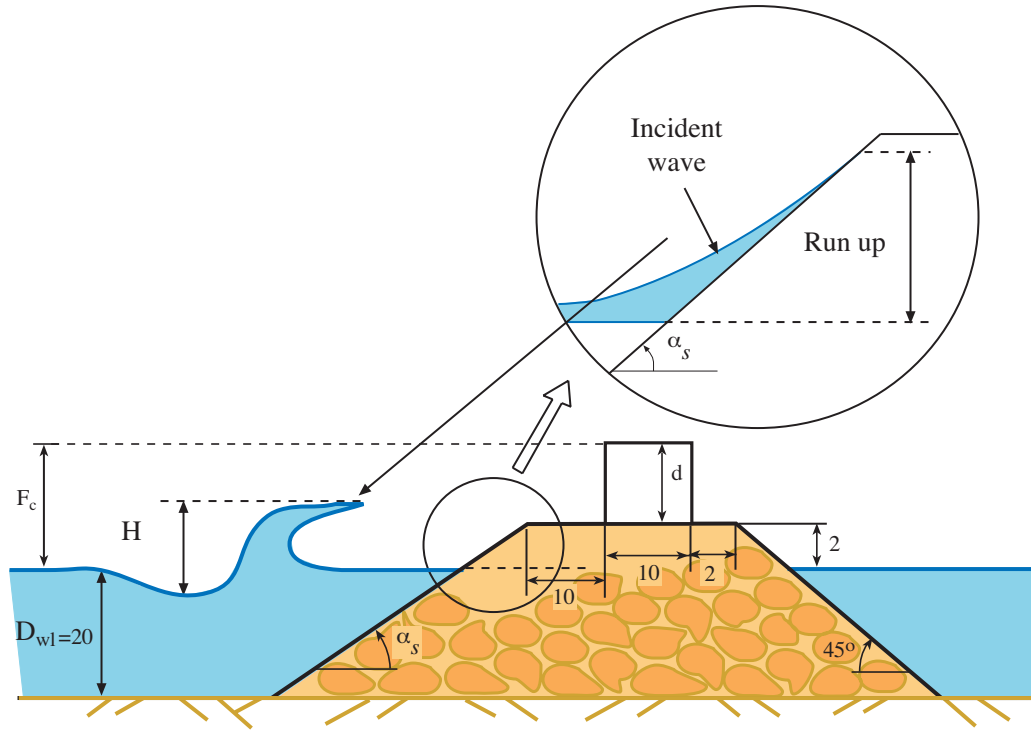


Figure 4.9: Ilustración del dique en talud y sus principales elementos.

4.7.3 Rubblemound breakwater example

Rubblemound breakwaters (see Figure 4.9) are constructed along the seacoast to protect the coastal or harbor areas from the wave attack during storms. The breakwater must be strong enough to survive the attack of storm waves, and the crest must be high enough to prevent the intrusion of sea water onto the harbor by overtopping. The volume of water that overtops the structure can be estimated from the volume of water that would rise over the extension of the slope exceeding the freeboard level. With this approximation, overtopping (failure) occurs whenever the difference between the maximum excursion of water over the slope, R_u , called wave 'run-up', exceeds the freeboard F_c .

Based on experiments, Losada y Giménez Curto (1979) proposed the following equation to evaluate the dimensionless quantity R_u/H :

$$R_u = A_u H (1 - \exp(-B_u I_r)), \quad (4.73)$$

where

I_r : Iribarren number $I_r \approx 1.25T \tan \beta / \sqrt{H}$, for shallow waters.

A_u y B_u : are given experimental parameters that depend on the armor units,

H y T : incident wave height and period, respectively,

α_s : is the seaside slope angle.

Thus, the limit state equation that verify if overtopping failure occurs is

$$g(A_u, B_u, H, T) = F_c - R_u < 0, \quad (4.74)$$

where F_c is the freeboard. Note that if F_c is lower than R_u then $g(A_u, B_u, H, T) < 0$, and failure occurs.

Statistical hypothesis

To get the reliability against overtopping, the following random variables are considered:

1. A_u and B_u are normal independent random variables. Independent between them, even though they have been determined by a fitting procedure, because the resulting correlation was negligible. And independent of (H, T) , because they just depend on the armor units (nature and positioning).

2. The relative wave height is a Raileigh random variable with cumulative distribution function

$$F_H(h) = 1 - \exp(-1.416^2 h^2), \quad (4.75)$$

where $h = H/H_s$ is the relative wave height and H_s is the significant wave height of the sea state.

3. The joint probability density function $(H/H_s, T/T_z)$ is a Longuet-Higgins distribution $(LH(\nu))$:

$$f(t, h; \nu) = 2 \left[\frac{1}{2} \left(1 + \frac{1}{\sqrt{1 + \nu^2}} \right) \right]^{-1} \exp \left\{ -h^2 \left[1 + \left(\left(1 - \frac{1}{t} \right) \frac{1}{\nu} \right)^2 \right] \right\} \frac{h^2}{\nu t^2 \sqrt{\pi}}; \quad h, t \geq 0, \quad (4.76)$$

where $t = T/T_z$ is the relative wave period and T_z is the mean zero-crossing period.

Finally, the period distribution for a range of wave height can be obtained against the conditional density function as

$$f(t|h) = \frac{f(t, h; \nu)}{f(h)} = \frac{1.416h}{\sqrt{\pi}\nu} \exp \left(-\frac{1.416^2 h^2}{\nu^2} (t-1)^2 \right). \quad (4.77)$$

This density function is a normal with mean unity. The standard deviation is obtained in a straightforward manner as:

$$\sigma(t) = \frac{\nu}{1.416\sqrt{2}h}.$$

This equation shows that the range of period distributions for a given value of wave height is inversely proportional to the wave height. Thus, lower wave heights have wider period distributions.

These assumptions are valid just working with narrow spectrum swell, i.e., low values of parameter ν ($\nu^2 < 0.36$).

In Table 4.3 are shown the statistical hypothesis for the random variables considered A_u, B_u, H y T .

Table 4.3: Statistical distributions for the random variables considered in the overtopping example.

Variable	Distribution
A_u	$N(1.05, 0.210^2)$
B_u	$N(-0.67, 0.134^2)$
H/H_s	Raileigh($H_s = 5$)
$(H/H_s, T/T_z)$	$LH(\nu = 0.25), T_z = 10$

FORM standard application

To get a different interpretation of the reliability problem, algorithm 4.2 is modified. The method proceeds as follows:

1. Obtaining independent uniform variables using Rosenblatt transformation

$$\begin{aligned}
 u_1 &= \Phi((A_u - m_a)/s_a) \\
 u_2 &= \Phi((B_u - m_b)/s_b) \\
 u_3 &= F_H(H/H_s) \\
 u_4 &= \Phi\left(\frac{T/T_z - 1}{\nu/(1.416\sqrt{2}h)}\right)
 \end{aligned} \tag{4.78}$$

2. Transforming variables into the multinormal standard space \mathbf{Z}

$$\begin{aligned}
 z_1 &= (A_u - m_a)/s_a \\
 z_2 &= (B_u - m_b)/s_b \\
 z_3 &= \Phi^{-1}(F_H(H/H_s)) \\
 z_4 &= \frac{T/T_z - 1}{\nu/(1.416\sqrt{2}h)}
 \end{aligned} \tag{4.79}$$

3. Starting the iterative process:

- (a) An initial point in terms of random variables is selected

$$\mathbf{x}^{(1)} = (A_u^{(1)}, B_u^{(1)}, H^{(1)}, T^{(1)})^T,$$

thus using (4.79) it is obtained the point

$$\mathbf{z}^{(1)} = (z_1^{(1)}, z_2^{(1)}, z_3^{(1)}, z_4^{(1)})^T.$$

- (b) Linearization of the limit state equation $g_{\mathbf{z}}(\mathbf{z})$, transformed from (4.74), using Taylor series expansion

$$g_{\mathbf{z}}(\mathbf{z}) = g_{\mathbf{z}}(\mathbf{z}^{(1)}) + \frac{\partial g_{\mathbf{z}}(\mathbf{z}^{(1)})}{\partial \mathbf{z}} (\mathbf{z} - \mathbf{z}^{(1)}).$$

The limit state equation value $g_{\mathbf{z}}(\mathbf{z}^{(1)}) = g_{\mathbf{x}}(\mathbf{x}^{(1)})$ is obtained in a straightforward manner from (4.74), but the partial derivative $\partial g_{\mathbf{z}}(\mathbf{z}^{(1)})/\partial \mathbf{z}$ is more difficult to obtain,

$$\begin{aligned} \frac{\partial g_{\mathbf{z}}(\mathbf{z})}{\partial \mathbf{z}} &= \begin{pmatrix} \frac{\partial g_{\mathbf{x}}}{\partial A_u} \frac{\partial A_u}{\partial z_1} + \frac{\partial g_{\mathbf{x}}}{\partial B_u} \frac{\partial B_u}{\partial z_1} + \frac{\partial g_{\mathbf{x}}}{\partial H} \frac{\partial H}{\partial z_1} + \frac{\partial g_{\mathbf{x}}}{\partial T} \frac{\partial T}{\partial z_1} \\ \frac{\partial g_{\mathbf{x}}}{\partial A_u} \frac{\partial A_u}{\partial z_2} + \frac{\partial g_{\mathbf{x}}}{\partial B_u} \frac{\partial B_u}{\partial z_2} + \frac{\partial g_{\mathbf{x}}}{\partial H} \frac{\partial H}{\partial z_2} + \frac{\partial g_{\mathbf{x}}}{\partial T} \frac{\partial T}{\partial z_2} \\ \frac{\partial g_{\mathbf{x}}}{\partial A_u} \frac{\partial A_u}{\partial z_3} + \frac{\partial g_{\mathbf{x}}}{\partial B_u} \frac{\partial B_u}{\partial z_3} + \frac{\partial g_{\mathbf{x}}}{\partial H} \frac{\partial H}{\partial z_3} + \frac{\partial g_{\mathbf{x}}}{\partial T} \frac{\partial T}{\partial z_3} \\ \frac{\partial g_{\mathbf{x}}}{\partial A_u} \frac{\partial A_u}{\partial z_4} + \frac{\partial g_{\mathbf{x}}}{\partial B_u} \frac{\partial B_u}{\partial z_4} + \frac{\partial g_{\mathbf{x}}}{\partial H} \frac{\partial H}{\partial z_4} + \frac{\partial g_{\mathbf{x}}}{\partial T} \frac{\partial T}{\partial z_4} \end{pmatrix} = \\ &= \mathbf{J}^{-1} \frac{\partial \mathbf{g}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial A_u}{\partial z_1} & \frac{\partial B_u}{\partial z_1} & \frac{\partial H}{\partial z_1} & \frac{\partial T}{\partial z_1} \\ \frac{\partial A_u}{\partial z_2} & \frac{\partial B_u}{\partial z_2} & \frac{\partial H}{\partial z_2} & \frac{\partial T}{\partial z_2} \\ \frac{\partial A_u}{\partial z_3} & \frac{\partial B_u}{\partial z_3} & \frac{\partial H}{\partial z_3} & \frac{\partial T}{\partial z_3} \\ \frac{\partial A_u}{\partial z_4} & \frac{\partial B_u}{\partial z_4} & \frac{\partial H}{\partial z_4} & \frac{\partial T}{\partial z_4} \end{pmatrix} \begin{pmatrix} \frac{\partial g_{\mathbf{x}}}{\partial A_u} \\ \frac{\partial g_{\mathbf{x}}}{\partial B_u} \\ \frac{\partial g_{\mathbf{x}}}{\partial H} \\ \frac{\partial g_{\mathbf{x}}}{\partial T} \end{pmatrix} \end{aligned}$$

where \mathbf{J}^{-1} is the inverse of the Jacobian \mathbf{J} of the Rosenblatt transformation,

$$\begin{aligned} \mathbf{J} &= \begin{pmatrix} \frac{\partial z_1}{\partial A_u} & \frac{\partial z_2}{\partial A_u} & \frac{\partial z_3}{\partial A_u} & \frac{\partial z_4}{\partial A_u} \\ \frac{\partial z_1}{\partial B_u} & \frac{\partial z_2}{\partial B_u} & \frac{\partial z_3}{\partial B_u} & \frac{\partial z_4}{\partial B_u} \\ \frac{\partial z_1}{\partial H} & \frac{\partial z_2}{\partial H} & \frac{\partial z_3}{\partial H} & \frac{\partial z_4}{\partial H} \\ \frac{\partial z_1}{\partial T} & \frac{\partial z_2}{\partial T} & \frac{\partial z_3}{\partial T} & \frac{\partial z_4}{\partial T} \end{pmatrix} = \\ &= \begin{pmatrix} \frac{1}{\sigma_a} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_b} & 0 & 0 \\ 0 & 0 & \frac{\partial \Phi^{-1}(F_H(H/H_s))}{\partial H} & -\frac{T/T_z - 1}{(\nu/\sqrt{2}(H^2/H_s)1.416)} \\ 0 & 0 & 0 & \frac{1/T_z}{(\nu/\sqrt{2}(H/H_s)1.416)} \end{pmatrix} \quad (4.80) \end{aligned}$$

whereas $\frac{\partial \mathbf{g}}{\partial \mathbf{x}}$ is equal to,

$$\begin{pmatrix} \frac{\partial g_{\mathbf{x}}}{\partial A_u} \\ \frac{\partial g_{\mathbf{x}}}{\partial B_u} \\ \frac{\partial g_{\mathbf{x}}}{\partial H} \\ \frac{\partial g_{\mathbf{x}}}{\partial T} \end{pmatrix} = \begin{pmatrix} -H(1 - \exp(B_u I_r)) \\ H A_u I_r \exp(B_u I_r) \\ -A_u(1 - \exp(B_u I_r)) + H A_u B_u \exp(B_u I_r) \frac{\partial I_r}{\partial H} \\ H A_u B_u \exp(B_u I_r) \frac{\partial I_r}{\partial T} \end{pmatrix} \quad (4.81)$$

- (c) As at the solution point the condition $g_{\mathbf{z}}(\mathbf{z}) = 0$ has to be fulfilled, the following estimation of the final design point can be obtained as

$$\mathbf{z}^{(2)} = \mathbf{z}^{(1)} - \frac{g_{\mathbf{z}}(\mathbf{z}^{(1)})}{\partial g_{\mathbf{z}}(\mathbf{z}^{(1)}) / \partial \mathbf{z}}. \quad (4.82)$$

Note that the equation is the same as (4.52) and, therefore, equal to (4.54).

- (d) The iterative process continue until changes in \mathbf{z} or β are smaller than the tolerance.

FORM application using optimization techniques

Using standard optimization packages for solving the same problem, the global problem approach is:

$$\begin{array}{l} \text{Minimize} \\ z_1, z_2, z_3, z_4 \end{array} \left(\sum_{i=1}^n z_i^2 \right)^{1/2} \quad (4.83)$$

subject to

$$\begin{aligned} z_1 &= (A_u - m_a) / s_a \\ z_2 &= (B_u - m_b) / s_b \\ \Phi(z_3) &= F_H(H/H_s) \\ z_4 &= \frac{T/T_z - 1}{(\nu/\sqrt{2}h1.416)} \\ 0 &= F_c - A_u H \left(1 - \exp(1.25 B_u T \tan \beta / \sqrt{H}) \right). \end{aligned} \quad (4.84)$$

The GAMS code for solving this problem is the following:

```
$title Rebase
```

```
file out /rebase.out/;
put out;
```

```
SETS
```

```
I number of random variables /1*4/;
```

SCALARS

```
pi /3.1415926535898/
a Raileigh distribution parameter /1.416/;
```

PARAMETERS

```
ma mean value of Au /1.05/
sa standard deviation of Au /0.210/
mb mean value of Bu /-0.67/
sb standard deviation of Bu /0.134/
Hs significant wave height of the sea state (m) /5.0/
Tz mean period of the sea state (s) /10.0/
nu spectral width coefficient /0.25/
tanb seaside slope
Fc freeboard /10/;
tanb=1/1.5;
```

VARIABLES

```
beta reliability index
z(I) standard normal random variables
Au experimental parameter that depend on the armor units
Bu experimental parameter that depend on the armor units
H wave height
T wave period;
```

EQUATIONS

```
betadef reliability index definition
Zdef1 Rosenblatt transformation related to Au
Zdef2 Rosenblatt transformation associated with Bu
Zdef3 Rosenblatt transformation related to H
Zdef4 Rosenblatt transformation associated with T
verdef overtopping limit state equation definition;
```

```
betadef..beta=e=sqrt(sum(I,sqr(Z(I))));
Zdef1..Z('1')=e=(Au-ma)/sa; Zdef2..Z('2')=e=(Bu-mb)/sb;
Zdef3..errorf(Z('3'))=e=1-exp(-sqr(a*H/Hs));
Zdef4..Z('4')=e=(T/Tz-1)*(sqrt(2)*a*H/Hs)/nu;
verdef..0=e=Fc-Au*H*(1-exp(Bu*1.25*T*tanb/sqrt(H)));
```

```
MODEL overtopping/betadef,Zdef1,Zdef2,Zdef3,Zdef4,verdef/;
```

```
* Initial values for the variables involved
Z.l(I)=0; Au.l=ma; Bu.l=mb; H.l=2.93981; T.l=Tz;
```

```
* Writing in external file
```

```

put "Initial values"/;
put "Au= ",Au.1:12:8/;
put "Bu=",Bu.1:12:8/;
put "H= ",H.1:12:8/;
put "T= ",T.1:12:8/;
loop(I,
    put "Z(",I.t1:1,")= ",Z.1(I):12:8/;
);
put/;

*      Solving the optimization problem
SOLVE overtopping USING nlp MINIMIZING beta;

*      Writing the results in the external files
put " modelstat= ",overtopping.modelstat,",solvestat=
",overtopping.solvestat/;

put "pf= ",(errorf(-beta.1)):12:8/; put "beta= ",beta.1:12:8//;
loop(I,
    put "Z(",I.t1:1,")= ",Z.1(I):12:8/;
) put ""/;

put "Design point or point of maximum likelihood"/;

put "Au= ",Au.1:12:8/;
put "Bu= ",Bu.1:12:8/;
put "H= ",H.1:12:8/;
put "T= ",T.1:12:8/;

```

Ejemplo numérico

Applying both methods and assuming the following values for the variables involved:

$$\begin{aligned}
 m_a = 1.05; \quad s_a = 0.210; \quad m_b = -0.67; \quad s_b = 0.134; \quad H_s = 5; \\
 T_z = 10; \quad \nu = 0.25; \quad \tan \beta = 1/1.5; \quad F_c = 10,
 \end{aligned}
 \tag{4.85}$$

the same result is obtained. The reliability index is $\beta^* = 3.07867$, and the probability of failure is $p_f^* = 0.0010396$. This implies that the probability of overtopping of one wave in the sea state defined by H_s y T_z is approximately 0.1%. The design or maximum likelihood point in both spaces are:

$$\begin{aligned}
 z_1^* = 1.6287; \quad z_2^* = -0.5308; \quad z_3^* = 2.5483; \quad z_4^* = 0.2233; \\
 A_u^* = 1.3920; \quad B_u^* = -0.7411; \quad H^* = 8.0669; \quad T^* = 10.1728.
 \end{aligned}$$

The out file 'rebase.out', is the following:

Initial values

```

Au= 1.05000000
Bu= -0.67000000
H= 2.93981000
T= 10.00000000
Z(1)= 0.00000000
Z(2)= 0.00000000
Z(3)= 0.00000000
Z(4)= 0.00000000

modelstat= 2.00, solvestat= 1.00 pf= 0.00103957
beta= 3.07866894

Z(1)= 1.62869708
Z(2)= -0.53081131
Z(3)= 2.54831685
Z(4)=
0.22331340

Design point or point of maximum likelihood
Au= 1.39202639
Bu=-0.74112872
H= 8.06687811
T= 10.17279893

```

4.8 Second-Order Reliability-Methods (SORM)

Los métodos de primer orden estudiados en las secciones precedentes, pueden no ser satisfactorios si la ecuación de estado límite tiene una curvatura significativa. Incluso, ecuaciones de estado límite aparentemente lineales o con poca curvatura en el espacio original \mathbf{X} pueden tener una gran curvatura en el espacio transformado \mathbf{Z} .

Para estos casos hay otras aproximaciones posibles, como muestran Breitung [25], Castillo, Solares, y Gómez [39, 40, 41], Haskin, Staple y Ding [91]. Muchas de ellas están basadas en la utilización de aproximaciones de segundo orden (SORM, ‘second order reliability methods’) como en Breitung [25], Tvedt [167], Der Kiureghian et al. [57], Koyluoglu y Nielsen [105], etc.)

4.8.1 Methods

Los métodos de segundo orden tratan de aproximar las ecuaciones de estado límite último por aproximaciones parabólicas o esféricas en el entorno del punto de diseño o de máxima verosimilitud.

Tras una aproximación parabólica de la región de fallo, la integral (4.1) puede expresarse

de la siguiente manera (véase Madsen et al. [116]):

$$R = 1 - p_f = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \phi(x_1) \dots \phi(x_{n-1}) \int_{-\infty}^{\beta + \sum_{i=1}^{n-1} k_i x_i^2 / 2} \phi(x_n) dx_n dx_{n-1} \dots dx_1 \quad (4.86)$$

donde R es la fiabilidad del sistema, $\phi(\cdot)$ es la función de densidad de la variable normal estándar y $k_i = \partial^2 z_n / \partial^2 z_i$ son las curvaturas principales de la ecuación de estado límite $g\mathbf{z}(\mathbf{z}^*)$ en el punto de diseño \mathbf{z}^* , que pueden calcularse como los autovalores del Hessiano.

Breitung [25] obtuvo una fórmula asintótica que aproxima la probabilidad exacta de fallo cuando $\beta \rightarrow \infty$:

$$p_f = \Phi(-\beta) \prod_{i=1}^{n-1} (1 + \beta k_i)^{-1/2}, \quad (4.87)$$

donde $\Phi(\cdot)$ es la función de distribución de la variable normal estándar. Esta fórmula es válida sólo si $\beta k_i > -1$ y da buenos resultados con valores altos de β .

Para valores moderados del índice de fiabilidad β , Tvedt [167] introdujo una aproximación de tres términos en la que los dos últimos pueden interpretarse como correctores de la fórmula de Breitung:

$$p_f = A_1 + A_2 + A_3 \quad (4.88)$$

$$A_1 = \Phi(-\beta) \prod_{i=1}^{n-1} (1 + \beta k_i)^{-1/2} \quad (4.89)$$

$$A_2 = [\beta \Phi(-\beta) - \phi(\beta)] \left\{ \prod_{i=1}^{n-1} (1 + \beta k_i)^{-1/2} - \prod_{i=1}^{n-1} (1 + (\beta - 1)k_i)^{-1/2} \right\} \quad (4.90)$$

$$A_3 = (\beta + 1) [\beta \Phi(-\beta) - \phi(\beta)] \left\{ \prod_{i=1}^{n-1} (1 + \beta k_i)^{-1/2} - \operatorname{Re} \left[\prod_{i=1}^{n-1} (1 + (\beta + j)k_i)^{-1/2} \right] \right\} \quad (4.91)$$

donde j es la variable imaginaria. El segundo y tercer término no proporcionan una corrección suficientemente precisa para valores bajos de β en todos los casos. Y al igual que la fórmula de Breitung no funciona bien con curvaturas negativas, e incluso hay singularidades si la curvatura se aproxima a la de la esfera con centro el origen y radio β .

Posteriormente Tvedt [168] obtuvo una nueva formulación en el dominio complejo en la que se calcula la fiabilidad invirtiendo la función característica para la aproximación cuadrática. Tras introducir una nueva variable aleatoria $B = X_n - \sum_{i=1}^{n-1} k_i X_i^2 / 2$ cuya función característica es

$$M_B(\theta) = E[\exp(j\theta B)] = \exp\left(-\frac{1}{2}\theta^2\right) \prod_{i=1}^{n-1} (1 + j k_i \theta)^{-1/2} \quad (4.92)$$

la fórmula propuesta simplificada es

$$\begin{aligned}
 R = 1 - p_f &= P(B \leq \beta) = \int_{-\infty}^{\beta} \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(ju\theta) M_B(\theta) d\theta du \\
 &= 0.5 + \frac{1}{\pi} \int_0^{\infty} \sin \left(\beta\theta + \frac{1}{2} \sum_{i=1}^{n-1} \arctan(k_i\theta) \right) \frac{\exp(-\theta^2/2)}{\theta \prod_{i=1}^{n-1} (1 + k_i^2\theta^2)^{1/4}} d\theta
 \end{aligned} \tag{4.93}$$

Tvedt derivó la expresión en el dominio complejo pero se ha transformado al dominio real para simplificar. Esta fórmula es exacta para integrales SORM, y aplicable para todo tipo de curvaturas y valores de β sin embargo requiere la integración numérica de (4.93), aunque es una integral unidimensional.

Una versión más sencilla y que parece prometedora es la aproximación propuesta por Koyluoglu y Nielsen [105], que utiliza una aproximación polinomial del gradiente de la función de densidad conjunta en el punto de diseño.

4.9 Reliability Based Optimization

La era moderna de la optimización estructural comenzó hace 35 años con el trabajo pionero de Smicht [157], que mostró claramente que los problemas de diseño estructural pueden plantearse como problemas de programación matemática. Y dado que las técnicas de optimización se han desarrollado enormemente, parece lógico que se traten de utilizar estas metodologías para el diseño.

Hay dos filosofías de diseño, íntimamente ligadas a las distintas medidas de la fiabilidad estructural (véase el capítulo ??): optimización estructural determinista (DSO, ‘Deterministic Structural Optimization’), y la optimización estructural basada en fiabilidad (RBSO, ‘Reliability-Based Structural Optimization’). Dado que los problemas ingenieriles son claramente no deterministas nos vamos a centrar en esta última.

En las últimas tres décadas se ha publicado una gran cantidad de literatura relativa a estos asuntos Murotsu, Kishi, Okada, Yonezawa y Taguchi [127], Sorensen [158, 159], para una revisión exhaustiva de la misma véase Frangopol [78]. Algunas de ellas incluyen las conferencias ICOSSAR-IASSAR: Kobe 1985 (Konishi et al. [104]), San Francisco 1989 (Ang et al. [7]), Insbruck 1993 (Schÿeller et al. [154]). Las últimas conferencias ICASP-CERRA: Florencia 1983 (Augusti et al. [9]), Vancouver 1987 (Lind [111]), México 1991 (Esteva y Ruiz [73]) y Paris 1995 (Lemaire et al. [107]). Las Conferencias Especiales sobre Mecánica Probabilista y Fiabilidad Estructural: Berkeley 1984 (Wen [174]), BlacksBurg 1988 (Spanos [161]) y Denver 1992 (Lin [109]). Las seis conferencias sobre Fiabilidad y Optimización de Sistemas Structurales del Grupo de Trabajo IFIP 7.5: Aalborg 1987 (Thoft-Christensen [164]), Londres 1988 (Thoft-Christensen [165]), Berkeley 1990 (Der Kiureghian y Thoft Christensen [59]) Munich 1991 (Rackwitz y Thoft-Christensen [141]), Takamatsu-shi 1993 (Thoft Christensen y Ishikawa [166]) y Assisi 1994 (Rackwitz et al. [139]).

El principal objetivo en el diseño estructural es conseguir estructuras con niveles de fiabilidad satisfactorios y lo más baratas posibles. Este requerimiento contradictorio usualmente se consigue mediante métodos deterministas basados en códigos, existen una gran cantidad

de métodos numéricos que resuelven este tipo de problemas (véase Arora [8], Vanderplaats [169], Adeli [5], Bazaraa et al. [16], Castillo et al. [30], Luenberger [?]), pero actualmente, estamos en disposición de incluir la parte estadística del problema (véase Blockley [23], Ditlevsen y Madsen [64], Eurocódigo [74], ROM [143], Freudenthal [81], Madsen et al. [116], Melchers [123], Steward y Melchers [162], Wirsching y Wu [175], Wu, Burnside y Cruse [179]). La minimización del coste no es la única función objetivo propuesta, de hecho en la optimización determinista usualmente se minimiza el volumen o peso sujeto a las restricciones impuestas por los códigos para los esfuerzos, desplazamientos, etc. Y puede ser de interés un diseño que maximice la utilidad de la estructura, u otra función objetivo que se nos ocurra. Uno de los aspectos más complicados a la hora de definir la función objetivo de un problema basado en la fiabilidad es dar un coste monetario a los heridos o muertos que se pudieran producir, sin embargo, cuando se prescinde de ese término, se pueden diseñar estructuras óptimas durante la vida útil de la obra teniendo en cuenta costes directos, costes de reparación, mantenimiento, etc.

4.9.1 State of the art review

En esta subsección se recoge una recopilación de trabajos relacionados con la optimización estructural (véase Enevoldsen [68] y Enevoldsen y Sorensen [71]). En el capítulo 10 se presenta una recopilación de los problemas tratados en esta tesis.

Los problemas básicos que se presentan en la optimización estructural tratan de minimizar una función objetivo teniendo en consideración la fiabilidad del sistema, es decir que se añaden restricciones de fiabilidad, pero no sólo asociadas a modos de fallo independientes sino también relativas a sistemas tanto en serie como en paralelo (véase Enevoldsen y Sorensen [70], Royset, Der Kiureghian y Polak [149] o Der Kiureghian y Polak [58]). Las primeras aproximaciones al problema se concentraron en una optimización con dos niveles usando los métodos FORM, y también en una aproximación en un nivel donde el problema asociado a la fiabilidad se reemplazaba por las condiciones de Karush-Kuhn-Tucker del problema FORM (Madsen y Friis Hansen [117] y Kuschel y Rackwitz [106]). La primera aproximación tiene dos problemas (Kirjner-Neto et al. [103]): el primero que no es posible probar la convergencia de algoritmos estándar. El segundo es que requiere la evaluación de muchos índices de fiabilidad, que puede ser computacionalmente caro, además está limitado al uso de FORM, su extensión a SORM u otros métodos de fiabilidad es problemática. La segunda aproximación tiene varias desventajas, el método requiere las derivadas segundas de la ecuación de estado límite, y está limitada a aproximaciones FORM y SORM y a problemas con fiabilidades asociadas a componentes. Para sistemas, sólo se puede aplicar a sistemas en serie separables, es decir, a aquellos en los que la correlación entre modos de fallo es despreciable. El método de la superficie de respuesta ('Response Surface Method') se ha utilizado para obtener una solución aproximada usando dos niveles (Gasser y Schueller [85]). También ha habido esfuerzos para usar optimización estocástica, Marti [119] y Der Kiureghian y Polak [58] han reformulado los problemas de diseño relativos a fiabilidad a problemas de optimización semiinfinitos (Polak [136]). Kirjener-Neto et al. [103] desarrollaron una aproximación externa para resolver problemas con restricciones en componentes, y Der Kiureghian y Polak lo extendieron a sistemas estructurales introduciendo los cálculos relativos a fiabilidad independientemente del procedimiento de optimización.

Chapter 5

Structural Reliability Methods II. Structural Systems.

Chapter 6

Duality

6.1 Introduction

This chapter deals with duality in mathematical programming. Given a mathematical programming problem, called the primal problem, there exists another associated mathematical programming problem, called the dual problem, closely related with it. Since duality is symmetric, i.e., the dual of the dual problem is the primal, both problems are said to be dual to each other. In the following, we provide simple examples that lead to a better understanding of this important concept.

Under certain assumptions, the primal and dual problems have the same optimal objective function value; hence it is possible to solve the primal problem indirectly by solving its corresponding dual problem. This can lead to important computational advantages.

As it will be shown, the values of the dual variables give the sensitivities of the objective function with respect to changes in the constraints.

Furthermore, dual variables constitute the information interchanged between master problems and subproblems in most decomposition schemes; therefore, duality is important in this book.

In Section 6.2 the important Karush-Kuhn-Tucker optimality conditions are described. Since the mathematical statement of dual problems have substantial differences in linear and nonlinear programming, we deal with them in Sections 6.3 and 6.4 respectively. However, later, their equivalence is shown. Finally, Section 6.5 illustrates duality and separability.

6.2 Karush-Kuhn-Tucker First Order Optimality Conditions

The general problem of mathematical programming, also referred to as the nonlinear programming problem (NLPP), can be stated as follows.

$$\begin{array}{ll} \text{Minimize} & z = f(x_1, \dots, x_n) \\ & x_1, \dots, x_n \end{array} \quad (6.1)$$

subject to

$$\begin{aligned} h_1(x_1, \dots, x_n) &= 0 \\ &\vdots \\ h_\ell(x_1, \dots, x_n) &= 0 \\ g_1(x_1, \dots, x_n) &\leq 0 \\ &\vdots \\ g_m(x_1, \dots, x_n) &\leq 0 \end{aligned}$$

In compact form the previous model can be stated as follows.

$$\begin{aligned} \text{Minimize } z &= f(\mathbf{x}) \\ \mathbf{x} \end{aligned} \tag{6.2}$$

subject to

$$\mathbf{h}(\mathbf{x}) = \mathbf{0} \tag{6.3}$$

$$\mathbf{g}(\mathbf{x}) \leq \mathbf{0} \tag{6.4}$$

where $\mathbf{x} = (x_1, \dots, x_n)^T$ is the vector of the decision variables, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is the objective function, and $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^\ell$ and $\mathbf{r} : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Note that $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), \dots, h_\ell(\mathbf{x}))^T$ and $\mathbf{r}(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_m(\mathbf{x}))^T$ are the equality and the inequality constraints, respectively. For this problem to be nonlinear, at least one of the functions involved in its formulation must be nonlinear. Any vector $\mathbf{x} \in \mathbb{R}^n$ that satisfies the constraints is said to be a feasible solution, and the set of all feasible solutions is referred to as the feasible region.

One of the most important theoretical results in the field of nonlinear programming are the conditions of Karush, Kuhn and Tucker. They must be satisfied at any constrained optimum, local or global, of any linear and most nonlinear programming problems. Moreover, they form the basis for the development of many computational algorithms. In addition, the criteria for stopping many algorithms, specifically, for recognizing when a local constrained optimum has been achieved, are derived directly from them.

In unconstrained differentiable problems the gradient is equal to zero at the minimum. In constrained differentiable problems the gradient is not necessarily equal to zero. This is due to the constraints of the problem. Karush–Kuhn–Tucker conditions generalize the necessary conditions for unconstrained problems to constrained problems.

Definition 6.1 (Karush–Kuhn–Tucker Conditions (KKTC)). *The vector $\mathbf{x} \in \mathbb{R}^n$ satisfies the KKTCs for the NLPP (6.2)–(6.4) if there exists vectors $\boldsymbol{\mu} \in \mathbb{R}^m$ and $\boldsymbol{\lambda} \in \mathbb{R}^\ell$ such that*

$$\nabla f(\mathbf{x}) + \sum_{k=1}^{\ell} \lambda_k \nabla h_k(\mathbf{x}) + \sum_{j=1}^m \mu_j \nabla g_j(\mathbf{x}) = \mathbf{0} \tag{6.5}$$

$$h_k(\mathbf{x}) = 0, k = 1, \dots, \ell \tag{6.6}$$

$$g_j(\mathbf{x}) \leq 0, j = 1, \dots, m \tag{6.7}$$

$$\mu_j g_j(\mathbf{x}) = 0, j = 1, \dots, m \tag{6.8}$$

$$\mu_j \geq 0, j = 1, \dots, m \tag{6.9}$$

■

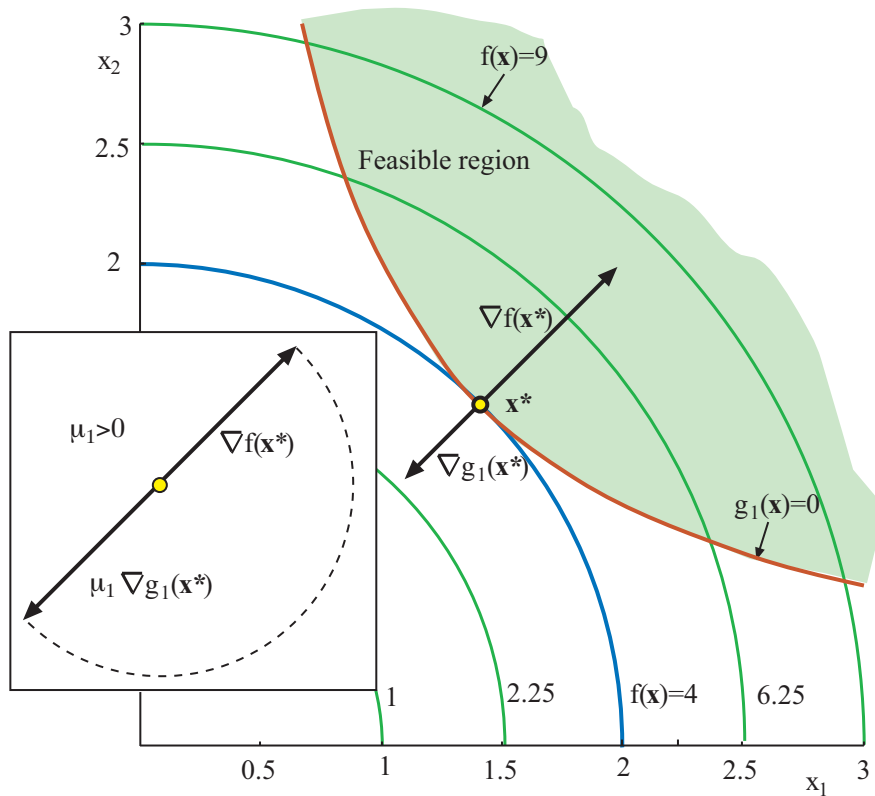


Figure 6.1: Illustration of the Karush-Kuhn-Tucker conditions for the case of one inequality constraint in the bidimensional case.

The vectors μ and λ are called the Kuhn–Tucker multipliers . Condition (6.8) is known as the complementary slackness condition and requires the nonnegativity of the multipliers of the inequality constraints (6.9), and is referred to as the dual feasibility condition, and (6.6)–(6.7) are called the primal feasibility conditions.

The genesis of these first-order optimality conditions (KKTC) can be motivated in the case of two independent variables for the case of (i) one equality or inequality constraint, (ii) two inequality constraints and (iii) one equality and one inequality constraint, as shown in Figures 6.1, 6.2, and 6.3, respectively.

Consider the case of an inequality constraint (see Figure 6.1), which separates the plane \mathbb{R}^2 into two regions. In one of them the constraint is satisfied, and in the other one it is not. Feasible points are those in the feasible region including its border curve. If the minimum of the objective function is attained at the interior of the feasible region, the constraint is not binding and the associated multiplier vanishes. If, on the contrary, the minimum is attained at the boundary, the constraint is binding. The problem is then equivalent to one with the corresponding equality constraint, and at the minimum the gradients of the objective function and the constraint must be parallel (see Figure 6.1). They must be pointing in opposite directions because the objective function increases as we move toward the interior of the feasible region, while the constraint function becomes negative and thus it

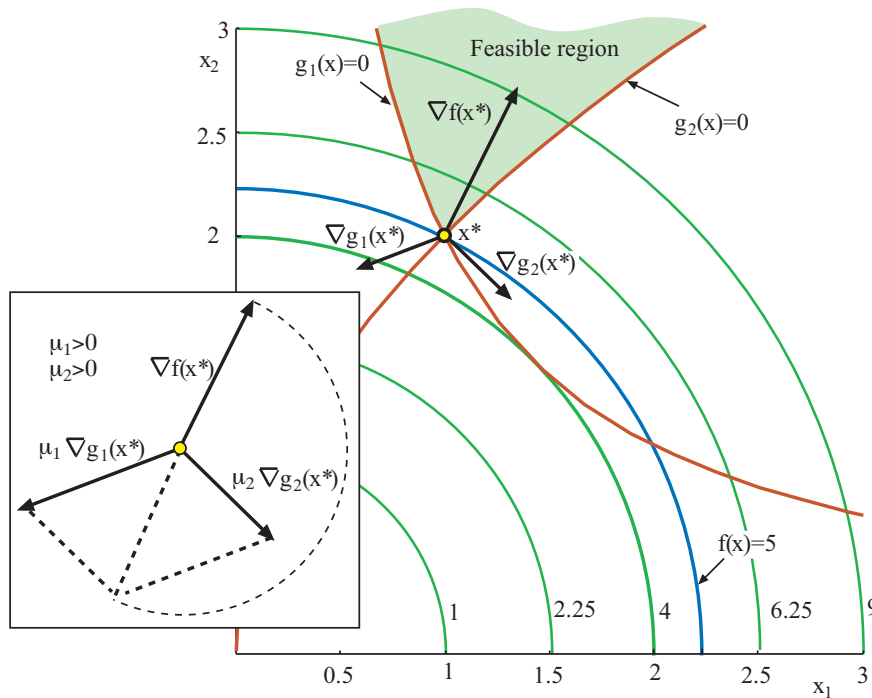


Figure 6.2: Illustration of the Karush-Kuhn-Tucker conditions for the case of two inequality constraints in the bidimensional case.

decreases. The multiplier is then either positive or zero. This is what optimality conditions say in this situation.

If we add a new inequality constraint (see Figure 6.2), equation (6.5) requires that, if we multiply the gradient of each binding constraint ($g_1(\mathbf{x}^*)$ and $g_2(\mathbf{x}^*)$) by its corresponding Lagrange multiplier, the vector sum of these two vectors must be equal the negative of the gradient of the objective function, as it is shown in Figure 6.2.

Now consider the case of one equality constraint and one inequality constraint (see Figure 6.3). Satisfying the constraints is equivalent to moving along the curve that represents the equality constraint inside the feasible region defined by the inequality constraint. While moving along this curve, the objective function contour curves must intersect in such a way that the objective function value decreases. When the inequality constraint is strictly satisfied the gradients of the objective function and the constraints are linearly dependent. This is what the first order optimality conditions state (see Figure 6.3).

Remark 6.1 (Special cases.) *If a type of constraint is lacking in a NLPP, then the multiplier associated with the “absent” constraint is equal to zero, and the constraint is dropped from the formulation of the KKTs. The form of the KKTs for these cases are:*

1. **(Unconstrained problems.)** *In this case we have only the condition*

$$\nabla f(\mathbf{x}) = \mathbf{0}$$

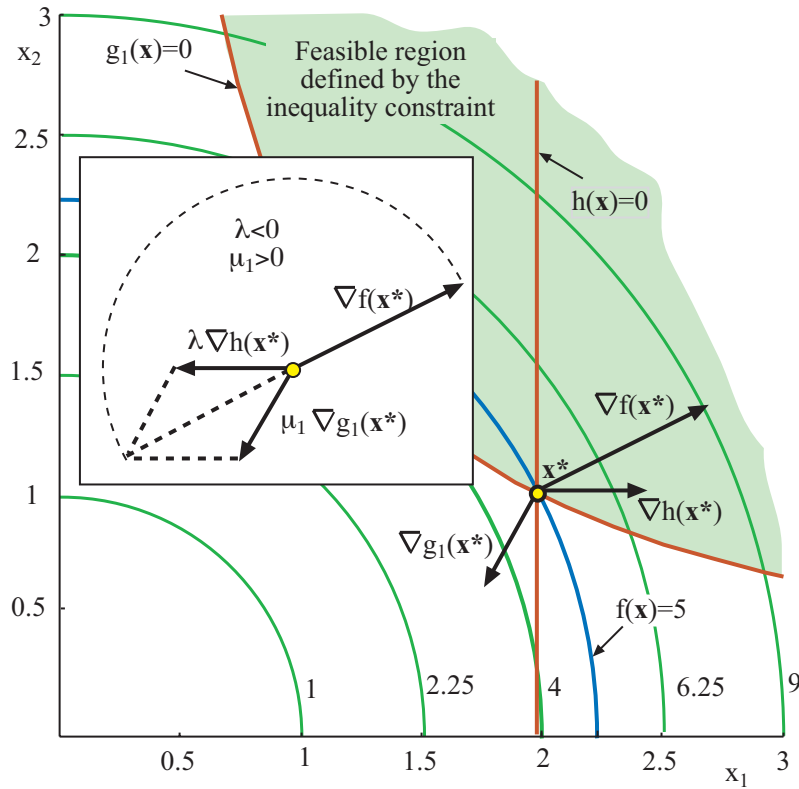


Figure 6.3: Illustration of the Karush-Kuhn-Tucker conditions for the case of one inequality constraint and one equality constraint in the bidimensional case.

2. **(Problems with only equality constraints).** The KKTs are an extension of the classical principle of the multiplier method. This method appears when NLPP have only equality constraints, and the KKTs has the form

$$\begin{aligned} \nabla f(\mathbf{x}) + \sum_{k=1}^{\ell} \lambda_k \nabla h_k(\mathbf{x}) &= \mathbf{0} \\ h_k(\mathbf{x}) &= 0; \quad k = 1, \dots, \ell \end{aligned} \tag{6.10}$$

3. **(Problems with only inequality constraints).** KKT conditions read

$$\begin{aligned} \nabla f(\mathbf{x}) + \sum_{j=1}^m \mu_j \nabla g_j(\mathbf{x}) &= \mathbf{0} \\ g_j(\mathbf{x}) &\leq 0; \quad j = 1, \dots, m \\ \mu_j g_j(\mathbf{x}) &= 0; \quad j = 1, \dots, m \\ \mu_j &\geq 0; \quad j = 1, \dots, m \end{aligned} \tag{6.11}$$



If we define the Lagrangian function by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x})$$

we can write the KKT conditions in compact form as

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= \mathbf{0} \\ \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= \mathbf{0} \\ \nabla_{\boldsymbol{\mu}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) &\leq \mathbf{0} \\ \boldsymbol{\mu}^T \nabla_{\boldsymbol{\mu}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) &= 0 \\ \boldsymbol{\mu} &\geq \mathbf{0} \end{aligned}$$

Note that $\boldsymbol{\mu}^T \nabla_{\boldsymbol{\mu}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0$ is equivalent to $\mu_j g_j(\mathbf{x}) = 0, j = 1, \dots, m$ only because $\nabla_{\boldsymbol{\mu}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \leq \mathbf{0}$ and $\boldsymbol{\mu} \geq \mathbf{0}$.

Illustrative Example 6.1 (KKT conditions). Consider the following optimization problem:

$$\begin{aligned} \text{Minimize} \quad & -x_1^2 - x_2^2 \\ \text{subject to} \quad & x_1, x_2 \end{aligned} \tag{6.12}$$

subject to

$$\begin{aligned} x_1 + x_2 &\leq 1 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned} \tag{6.13}$$

The Lagrangian function is:

$$\mathcal{L}(x_1, x_2; \mu_1, \mu_2, \mu_3) = -x_1^2 - x_2^2 + \mu_1(x_1 + x_2 - 1) + \mu_2(-x_1) + \mu_3(-x_2)$$

and the Karush-Kuhn-Tucker conditions become:

$$\frac{\partial \mathcal{L}(x_1, x_2; \mu_1, \mu_2, \mu_3)}{\partial x_1} = -2x_1 + \mu_1 - \mu_2 = 0 \Rightarrow x_1 = \frac{\mu_1 - \mu_2}{2} \tag{6.14}$$

$$\frac{\partial \mathcal{L}(x_1, x_2; \mu_1, \mu_2, \mu_3)}{\partial x_2} = -2x_2 + \mu_1 - \mu_3 = 0 \Rightarrow x_2 = \frac{\mu_1 - \mu_3}{2} \tag{6.15}$$

$$x_1 + x_2 \leq 1 \tag{6.16}$$

$$-x_1 \leq 0 \tag{6.17}$$

$$-x_2 \leq 0 \tag{6.18}$$

$$\mu_1(x_1 + x_2 - 1) = 0 \tag{6.19}$$

$$\mu_2(-x_1) = 0 \tag{6.20}$$

$$\mu_3(-x_2) = 0 \tag{6.21}$$

$$\mu_1, \mu_2, \mu_3 \geq 0 \tag{6.22}$$

which has two solutions:

$$x_1 = 1; \quad x_2 = 0; \quad \mu_1 = 2; \quad \mu_2 = 0; \quad \mu_3 = 2; \quad z = -1$$

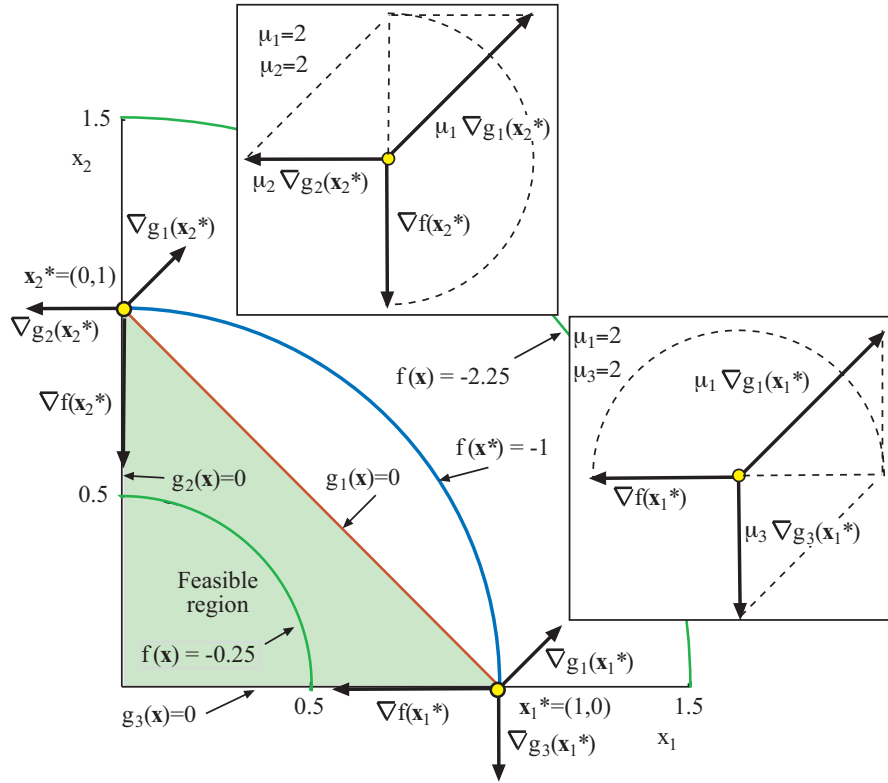


Figure 6.4: Illustration of the minimization problem in Example 6.1.

and

$$x_1 = 0; \quad x_2 = 1; \quad \mu_1 = 2; \quad \mu_2 = 2; \quad \mu_3 = 0; \quad z = -1$$

These solutions are illustrated in Figure 6.4, where the two optimal points are shown together with the gradients of the objective function and the active constraints as the KKT conditions state. Note that they are linearly dependent. ■

Equality Constraints and Newton Algorithm

The KKTs for problem.

$$\text{Minimize } z = f(\mathbf{x}) \tag{6.23}$$

$$\mathbf{x}$$

subject to

$$\mathbf{h}(\mathbf{x}) = \mathbf{0} \tag{6.24}$$

constitute the system of nonlinear equations

$$\begin{aligned} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) &= \mathbf{0} \\ \mathbf{h}(\mathbf{x}) &= \mathbf{0} \end{aligned} \tag{6.25}$$

where $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x})$.

The system of $n + \ell$ nonlinear equations presented above can be solved by the Newton algorithm. If \mathbf{z} denotes $(\mathbf{x}, \boldsymbol{\lambda})$ and $\mathbf{F}(\mathbf{z})$ denotes system (6.25), the Taylor expansion of this system is

$$\mathbf{F}(\mathbf{z} + \Delta \mathbf{z}) \approx \mathbf{F}(\mathbf{z}) + \nabla_{\mathbf{z}} \mathbf{F}(\mathbf{z}) \Delta \mathbf{z}$$

for $\|\Delta \mathbf{z}\|$ sufficiently small.

To achieve $\mathbf{F}(\mathbf{z}) = \mathbf{0}$, it is convenient to find a direction $\Delta \mathbf{z}$ so that $\mathbf{F}(\mathbf{z} + \Delta \mathbf{z}) = \mathbf{0}$. This direction can be computed from

$$\nabla_{\mathbf{z}} \mathbf{F}(\mathbf{z}) \Delta \mathbf{z} = -\mathbf{F}(\mathbf{z})$$

where $\nabla_{\mathbf{z}} \mathbf{F}(\mathbf{z})$ can be expressed as

$$\nabla_{\mathbf{z}} \mathbf{F}(\mathbf{z}) = \nabla_{(\mathbf{x}, \boldsymbol{\lambda})} \mathbf{F}(\mathbf{x}, \boldsymbol{\lambda}) = \begin{pmatrix} \nabla_{\mathbf{x}\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) & \nabla_{\mathbf{x}}^T \mathbf{h}(\mathbf{x}) \\ \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}) & \mathbf{0} \end{pmatrix}$$

and where $\nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x})$ is the Jacobian of $\mathbf{h}(\mathbf{x})$.

The matrix above is denominated the *KKT* matrix of problem (6.23)–(6.24), and the system

$$\begin{pmatrix} \nabla_{\mathbf{x}\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) & \nabla_{\mathbf{x}}^T \mathbf{h}(\mathbf{x}) \\ \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}) & \mathbf{0} \end{pmatrix} \begin{pmatrix} \Delta \mathbf{x} \\ \Delta \boldsymbol{\lambda} \end{pmatrix} = - \begin{pmatrix} \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \\ \mathbf{h}(\mathbf{x}) \end{pmatrix}$$

is denominated the *KKT* system of problem (6.23)–(6.24). It constitutes a Newton iteration to solve system (6.25).

6.3 Duality in Linear Programming

In this section we deal with duality in linear programming. We start by giving the definition of a dual problem.

Definition 6.2 (Dual problem). *Given the linear programming problem*

$$\begin{aligned} \text{Minimize } z &= \mathbf{c}^T \mathbf{x} \\ \mathbf{x} & \end{aligned} \tag{6.26}$$

subject to

$$\begin{aligned} \mathbf{A}\mathbf{x} &\geq \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned} \tag{6.27}$$

its dual problem is

$$\begin{aligned} \text{Maximize } z &= \mathbf{b}^T \mathbf{y} \\ \mathbf{y} & \end{aligned} \tag{6.28}$$

subject to

$$\begin{aligned} \mathbf{A}^T \mathbf{y} &\leq \mathbf{c} \\ \mathbf{y} &\geq \mathbf{0} \end{aligned} \tag{6.29}$$

where $\mathbf{y} = (y_1, \dots, y_m)^T$ are called dual variables. ■

We identify the first problem as the primal problem, and the second one as its dual counterpart. Notice how the same elements (the matrix \mathbf{A} , and the vectors \mathbf{b} and \mathbf{c}) determine both problems. Although the primal problem is not in its standard form, this format enables us to see more clearly the symmetry between the primal and the corresponding dual problem. The following theorem shows that the dual of the dual is the primal.

Theorem 6.1 (Symmetry of the dual relationship). *Duality is a symmetric relationship, that is, if problem D is the dual of problem P , then P is also the dual of D .*

Proof. To see this, we rewrite the dual problem (6.28)-(6.29) as a minimization problem with constraints of the form \geq , as in (6.26)-(6.27).

$$\begin{array}{ll} \text{Minimize} & z = -\mathbf{b}^T \mathbf{y} \\ & \mathbf{y} \end{array}$$

subject to

$$\begin{array}{ll} -\mathbf{A}^T \mathbf{y} & \geq -\mathbf{c} \\ \mathbf{y} & \geq \mathbf{0} \end{array} \quad (6.30)$$

Then, according to Definition 6.2, its dual is

$$\begin{array}{ll} \text{Maximize} & z = -\mathbf{c}^T \mathbf{x} \\ & \mathbf{x} \end{array}$$

subject to

$$\begin{array}{ll} -\mathbf{A}\mathbf{x} & \leq -\mathbf{b} \\ \mathbf{x} & \geq \mathbf{0} \end{array} \quad (6.31)$$

which is equivalent to the primal problem (6.26)-(6.27). ■

Remark 6.2 *As it can be observed, every constraint of the primal problem has a dual variable associated with it, the coefficients of the objective function of the primal problem are the RHS terms of the constraints of the dual problem and vice versa, and the coefficient matrix of the constraints of the dual problem is the transpose of the constraint matrix of the primal one. The primal is a minimization problem but its dual is a maximization problem.* ■

6.3.1 Obtaining the Dual from a Primal in Standard Form

In this section we address the problem of finding the dual problem when the primal is given in its standard form. To answer this question all we have to do is using Definition 6.2 of dual problem.

Theorem 6.2 (Dual in standard form). *The dual of the problem in standard form*

$$\begin{array}{ll} \text{Minimize} & z = \mathbf{c}^T \mathbf{x} \\ & \mathbf{x} \end{array} \quad (6.32)$$

subject to

$$\begin{aligned} \mathbf{Ax} &= \mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned} \tag{6.33}$$

is

$$\begin{aligned} \text{Maximize } z &= \mathbf{b}^T \mathbf{y} \\ \mathbf{y} \end{aligned} \tag{6.34}$$

subject to

$$\mathbf{A}^T \mathbf{y} \leq \mathbf{c} \tag{6.35}$$

Proof. The equality $\mathbf{Ax} = \mathbf{b}$ in (6.33) can be replaced by the two equivalent inequalities $\mathbf{Ax} \geq \mathbf{b}$ and $-\mathbf{Ax} \geq -\mathbf{b}$. Then, we can write the problem (6.32)-(6.33) as

$$\begin{aligned} \text{Minimize } z &= \mathbf{c}^T \mathbf{x} \\ \mathbf{x} \end{aligned}$$

subject to

$$\begin{aligned} \mathbf{Ax} &\geq \mathbf{b} \\ -\mathbf{Ax} &\geq -\mathbf{b} \\ \mathbf{x} &\geq \mathbf{0} \end{aligned}$$

Using Definition 6.2, the dual of this problem is

$$\begin{aligned} \text{Maximize } z &= \mathbf{b}^T \mathbf{y}^{(1)} - \mathbf{b}^T \mathbf{y}^{(2)} \\ \mathbf{y}^{(1)}, \mathbf{y}^{(2)} \end{aligned}$$

that is equivalent to

$$\begin{aligned} \text{Maximize } z &= \mathbf{b}^T \mathbf{y} \\ \mathbf{y} \end{aligned}$$

where $\mathbf{y} = \mathbf{y}^{(1)} - \mathbf{y}^{(2)}$ is not restricted in sign, subject to

$$\mathbf{A}^T \mathbf{y}^{(1)} - \mathbf{A}^T \mathbf{y}^{(2)} \leq \mathbf{c}$$

that is equivalent to

$$\mathbf{A}^T \mathbf{y} \leq \mathbf{c}$$

which ends the proof. ■

Remark 6.3 *The relationship between primal and dual problems (Theorem 6.2) could be taken as the definition of dual problem, and then, the initially given definition would have resulted as a theorem.* ■

6.3.2 Obtaining the Dual Problem

A linear programming problem whose formulation is not in the form (6.26)-(6.27) also has a dual problem. To facilitate the obtention of the dual of a given problem in any form or the primal from the dual, one can use the following sets of rules:

Rules to be used when the primal is a minimization problem:

Rule 1. A minimization primal problem leads to a maximization dual problem, and a maximization dual problem leads to a minimization primal problem.

Rule 2. An equality constraint in the primal (dual) implies that the corresponding dual (primal) variable is unrestricted.

Rule 3. A \geq (\leq) inequality constraint in the primal (dual) implies that its dual (primal) variable is nonnegative.

Rule 4. A \leq (\geq) inequality constraint in the primal (dual) implies that its dual (primal) variable is nonpositive.

Rule 5. A primal (dual) nonnegative variable leads to a \leq (\geq) inequality constraint in the dual (primal) problem.

Rule 6. A primal (dual) nonpositive variable leads to a \geq (\leq) inequality constraint in the dual (primal) problem.

Rule 7. A primal (dual) unrestricted variable leads to an equality constraint in the dual (primal) problem.

Rules to be used when the primal is a maximization problem:

Rule 1. A maximization primal problem leads to a minimization dual problem, and a minimization dual problem leads to a maximization primal problem.

Rule 2. An equality constraint in the primal (dual) implies that the corresponding dual (primal) variable is unrestricted.

Rule 3. A \geq (\leq) inequality constraint in the primal (dual) implies that its dual (primal) variable is nonpositive.

Rule 4. A \leq (\geq) inequality constraint in the primal (dual) implies that its dual (primal) variable is nonnegative.

Rule 5. A primal (dual) nonnegative variable leads to a \geq (\leq) inequality constraint in the dual (primal) problem.

Rule 6. A primal (dual) nonpositive variable leads to a \leq (\geq) inequality constraint in the dual (primal) problem.

Figure 6.5: Rules for obtaining dual problems from primal ones and vice versa.

Rule 7. A primal (dual) unrestricted variable leads to an equality constraint in the dual (primal) problem.

The above rules are summarized in Figure 6.5.

Illustrative Example 6.2 (Dual problem). The dual of the linear programming problem

$$\begin{array}{ll} \text{Minimize} & z = x_1 + x_2 - x_3 \\ & x_1, x_2, x_3 \end{array}$$

subject to

$$\begin{array}{rcll} 2x_1 & +x_2 & & \geq 3 \\ x_1 & & -x_3 & = 0 \\ & x_2 & +x_3 & \leq 2 \\ & & x_3 & \geq 0 \\ & & x_2 & \leq 0 \end{array} \quad (6.36)$$

is

$$\begin{array}{ll} \text{Maximize} & z = 3y_1 + 2y_3 \\ & y_1, y_2 \end{array}$$

subject to

$$\begin{array}{rcll} 2y_1 & +y_2 & & = 1 \\ y_1 & & +y_3 & \geq 1 \\ & -y_2 & +y_3 & \leq -1 \\ & & y_1 & \geq 0 \\ & & y_3 & \leq 0 \end{array} \quad (6.37)$$

To see this, we apply the rules, as follows:

Rule 1. Since the primal is a minimization problem, the dual is a maximization problem.

Rule 2. Since the second constraint in the primal is an equality constraint, the second dual variable y_2 is unrestricted.

Rule 3. Since the first constraint in the primal is \geq , the first dual variable y_1 is nonnegative.

Rule 4. Since the third constraint in the primal is \leq , the third dual variable y_3 is nonpositive.

Rule 5. Since the third primal variable x_3 is nonnegative, the third dual constraint is \leq .

Rule 6. Since the second primal variable x_2 is nonpositive, the second dual constraint is \geq .

Rule 7. Since the first primal variable x_1 is unrestricted, the first dual constraint is an equality.

Using the same rules to the dual problem, we can recover the primal one as follows:

Rule 1. Since the dual is a maximization problem, the primal is a minimization problem.

Rule 2. Since the first constraint in the dual is an equality constraint, the first primal variable x_1 is unrestricted.

Rule 3. Since the third constraint in the dual is \leq , the third primal variable x_3 is nonnegative.

Rule 4. Since the second constraint in the dual is \geq , the second primal variable x_2 is nonpositive.

Rule 5. Since the first dual variable y_1 is nonnegative, the first primal constraint is \geq .

Rule 6. Since the third dual variable y_3 is nonpositive, the third primal constraint is \leq .

Rule 7. Since the second dual variable y_2 is unrestricted, the second primal constraint is an equality.

■

6.3.3 Duality Theorems

The importance of the primal-dual relationship is established in the following theorems.

Lemma 6.1 (Weak duality lemma) *Let P be the LPP (6.26)-(6.27), and D its dual (6.28)-(6.29). Let \mathbf{x} be a feasible solution of P and \mathbf{y} a feasible solution of D . Then*

$$\mathbf{b}^T \mathbf{y} \leq \mathbf{c}^T \mathbf{x} \quad (6.38)$$

Proof. The proof is simple. If \mathbf{x} and \mathbf{y} are feasible for P and D , respectively, then

$$\mathbf{A}\mathbf{x} \geq \mathbf{b}, \quad \mathbf{y} \geq \mathbf{0}, \quad \mathbf{A}^T \mathbf{y} \leq \mathbf{c}, \quad \mathbf{x} \geq \mathbf{0}$$

Note, by the nonnegativity of \mathbf{x} and \mathbf{y} , that

$$\mathbf{b}^T \mathbf{y} \leq \mathbf{x}^T \mathbf{A}^T \mathbf{y} \leq \mathbf{x}^T \mathbf{c} = \mathbf{c}^T \mathbf{x}$$

■

Corollary 6.1 *If $\mathbf{b}^T \bar{\mathbf{y}} = \mathbf{c}^T \bar{\mathbf{x}}$ for some particular vectors $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, feasible for P and D , respectively, then the optimal solutions of the dual and primal problems are $\bar{\mathbf{y}}$ and $\bar{\mathbf{x}}$, respectively. Furthermore, if $\bar{\mathbf{x}}$ is an optimal solution for the primal problem, then, there is an optimal solution $\bar{\mathbf{y}}$ for the dual problem, and the optimal objective function value of both problems coincide with the common value $\mathbf{b}^T \bar{\mathbf{y}} = \mathbf{c}^T \bar{\mathbf{x}}$. Otherwise, one or both of the two sets of feasible solutions is empty.*

Table 6.1: Data of the mill planning problem

	Table type		Machine-hours available per day
	1	2	
Cutting time [hours]	4	3	40
Finishing time [hours]	4	7	56
Profit [\$]	70	90	

Proof. Proving this corollary is straightforward if we realize that, using (6.38), we obtain:

$$\begin{aligned}
 \mathbf{c}^T \bar{\mathbf{x}} = \mathbf{b}^T \bar{\mathbf{y}} &\leq \max_{\mathbf{y}} \{ \mathbf{b}^T \mathbf{y} \mid \mathbf{A}^T \mathbf{y} \leq \mathbf{c} \} \leq \\
 &\leq \min_{\mathbf{x}} \{ \mathbf{c}^T \mathbf{x} \mid \mathbf{A} \mathbf{x} \geq \mathbf{b}, \mathbf{x} \geq \mathbf{0} \} \leq \\
 &\leq \mathbf{c}^T \bar{\mathbf{x}} = \mathbf{b}^T \bar{\mathbf{y}}
 \end{aligned}$$

Hence, all inequalities are in reality equalities and $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ must be optimal solutions of P and D respectively, as claimed. ■

The Corollary 6.1 asserts that both problems, the primal P and the dual D , admit optimal solutions simultaneously.

In summary, one of the following two statements holds:

1. Both problems have optimal solution and their optimal values coincide.
2. One problem has unbounded optimum and the other one has an empty feasible region.

Illustrative Example 6.3 (Primal and dual of a mill problem). A small mill manufactures two types of wood tables. Each table of type 1 requires 4 machine-hours for cutting from the wood stock to size and 4 machine-hours finishing time (assembly, painting, etc.). Similarly, each table of type 2 requires 3 machine-hours of cutting time and 7 machine-hours of finishing time. The cutting and finishing equipment capacities available per day are 40 and 56 machine-hours, respectively. Finally, each type 1 table provides \$70 profit, while each type 2 table provides \$90 profit. These data are summarized in Table 6.1.

Primal and dual problems are analyzed below.

Primal:

In order to maximize the daily total profit, we wish to know the quantity of each type of tables produced per day. The linear programming model of the problem is

$$\begin{aligned}
 \text{Maximize } z &= 70x_1 + 90x_2 \\
 x_1, x_2 &
 \end{aligned}$$

subject to

$$\begin{aligned}
 4x_1 + 3x_2 &\leq 40 \\
 4x_1 + 7x_2 &\leq 56
 \end{aligned} \tag{6.39}$$

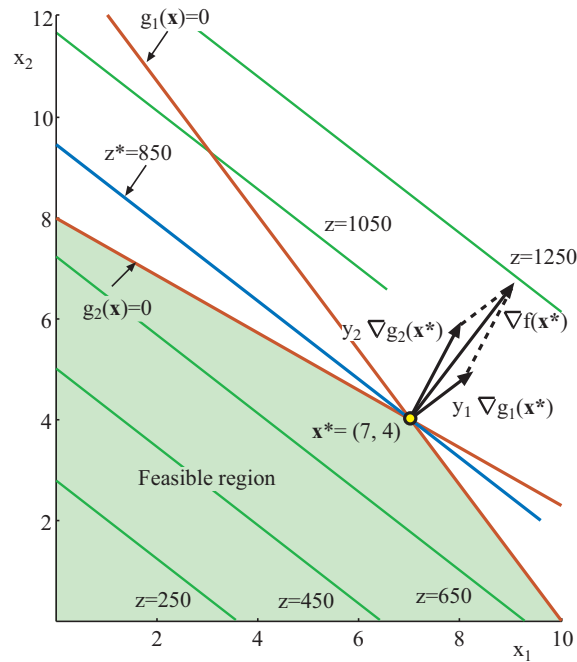


Figure 6.6: Graphical analysis of the mill planning problem.

$$x_1, x_2 \geq 0$$

where x_1 and x_2 are the quantities of table types 1 and 2, respectively, to be produced per day.

The optimal solution of this model, as can be seen in Figure 6.6, indicates a daily production of 7 type 1 tables and 4 type 2 tables, and a daily maximum profit of \$850. This result indicates that both cutting and finishing resources are fully utilized, because both constraints are binding.

Now, suppose that we wish to obtain higher daily profits. Then, an expansion of the available capacities is necessary. Suppose that the available finishing capacity can be expanded from 56 machine-hours to 72 machine-hours per day. How does this expansion affect the total daily profit? The answer is given by the solution of the following linear programming problem, where we

$$\begin{array}{ll} \text{Maximize} & z = 70x_1 + 90x_2 \\ & x_1, x_2 \end{array}$$

subject to

$$\begin{array}{rcl} 4x_1 + 3x_2 & \leq & 40 \\ 4x_1 + 7x_2 & \leq & 72 \\ x_1, x_2 & \geq & 0 \end{array} \quad (6.40)$$

In this case the optimal solution is $x_1 = 4$ and $x_2 = 8$ with a maximum daily profit of \$1000 (see Figure 6.7). This solution indicates that we can increase the daily profit by

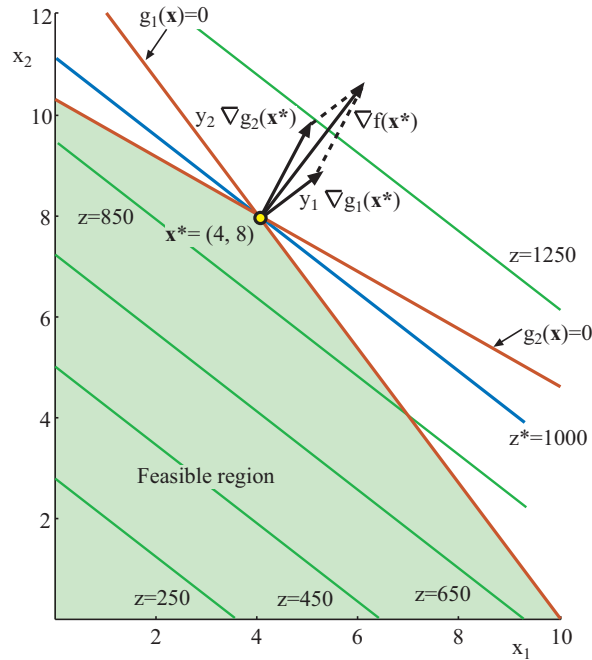


Figure 6.7: Graphical analysis of the expanded mill planning problem.

\$150 if we increase the daily finishing capacity by $72 - 56 = 16$ machine-hours. The rate $\$(1000-850)/16 = \$150/16 = \$75/8$, at which the objective function value will increase as the finishing capacity is increased by 1 hour, is called the sensitivity or shadow price (also dual price) of the finishing resource constraint.

A formal treatment of sensitivities is provided in Theorem 6.5 in page 211.

In general, the shadow price of a constraint is the change in the optimal value of the objective function per unit change in the RHS value of the constraint, assuming that all other parameter values of the problem remain unchanged. For many linear programming applications the shadow prices are at least as important as the solution of the problem itself; they indicate whether certain changes in the use of productive capacities affect the objective function value. Shadow prices can be obtained directly from the solution of the primal problem or determined by formulating and solving the dual problem.

Note that for both problems (6.39) and (6.40), as the same constraints remain active, the dual variable values are identical.

Dual:

The dual of the planning mill problem (6.39) is stated as follows.

$$\begin{aligned} \text{Minimize } & z = 40y_1 + 56y_2 \\ & y_1, y_2 \end{aligned}$$

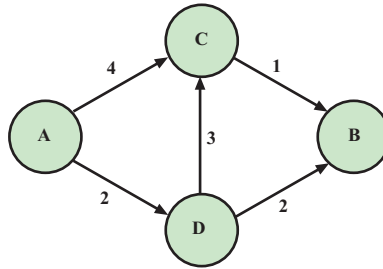


Figure 6.8: Communication net in Example 6.4.

subject to

$$\begin{aligned}
 4y_1 + 4y_2 &\geq 70 \\
 3y_1 + 7y_2 &\geq 90 \\
 y_1, y_2 &\geq 0
 \end{aligned} \tag{6.41}$$

The optimal solution of this problem is $y_1 = 65/8$, $y_2 = 75/8$ and the optimal objective function value is 850. Note that y_1 and y_2 are the shadow prices of cutting and finishing resources, respectively, and optimal objective function values of problems (6.39) and (6.41) coincide.

The dual problem (6.41) can be interpreted as follows. Assume that we are willing to sell machine time (rather than produce and sell tables), provided that we can obtain at least the same level of total profit. In this case, producing tables and selling machine time would be two equivalent alternatives; the y_1 and y_2 variables would represent the sale prices of one hour of work on the cutting and finishing machines, respectively. In order to maintain a competitive business we should set the total daily profit as low as possible, that is, we should minimize the function $40y_1 + 56y_2$, where the 40 and 56 coefficients represent the machine-hours available per day for sale of cutting and finishing machines. The constraints in (6.41) state that the total prices of the cutting and finishing hours required for producing one unit of each type of table must not be less than the profit expected from the sale of one unit of each type of table, and that prices must be positive quantities. ■

The primal dual relationship is further illustrated through the example below.

Illustrative Example 6.4 (Communication net). We are interested in sending a message through a communication net from one node A to some other node B with minimal cost. The network consists in four nodes, A , B , C , and D , and there are channels connecting A and C , A and D , D and C , C and B , and D and B , with associated costs per unit of message 4, 2, 3, 1, and 2, respectively (see Figure 6.8).

Primal

If variable x_{PQ} denotes the fraction of the message going through the channel connecting any node P with any other node Q , the following constraints must be satisfied:

1. No fraction of the total message is lost in node C :

$$x_{AC} + x_{DC} = x_{CB} \quad (6.42)$$

2. No part of the message is lost in node D :

$$x_{AD} = x_{DC} + x_{DB} \quad (6.43)$$

3. The complete message arrives to B :

$$1 = x_{CB} + x_{DB} \quad (6.44)$$

If we write the above constraints in matrix form, we get

$$\begin{array}{rccccrcr} x_{AC} & & +x_{DC} & -x_{CB} & & = & 0 \\ & x_{AD} & -x_{DC} & & -x_{DB} & = & 0 \\ & & & x_{CB} & +x_{DB} & = & 1 \\ x_{AC}, x_{AD}, x_{DC}, x_{CB}, x_{DB} & & & & & \geq & 0 \end{array} \quad (6.45)$$

Note that if we add the last two constraints and subtract the first one, we get $x_{AC} + x_{AD} = 1$. This means that the whole message departs from A . Since it can be obtained as a linear combination of the other three, it must not be included to avoid redundancy.

Consequently, the primal approach to this problem consists of minimizing the total cost of sending the messages

$$\begin{array}{ll} \text{Minimize} & z = 4x_{AC} + 2x_{AD} + 3x_{DC} + x_{CB} + 2x_{DB} \\ & x_{AC}, x_{AD}, x_{DC}, x_{CB}, x_{DB} \end{array}$$

subject to

$$\begin{array}{rccccrcr} x_{AC} & & +x_{DC} & -x_{CB} & & = & 0 \\ & x_{AD} & -x_{DC} & & -x_{DB} & = & 0 \\ & & & x_{CB} & +x_{DB} & = & 1 \\ x_{AC}, x_{AD}, x_{DC}, x_{CB}, x_{DB} & & & & & \geq & 0 \end{array} \quad (6.46)$$

Dual

Alternatively, we can assume that we can buy the information of the messages at each node and sell them at a different node. Thus, the dual approach can be interpreted as finding the prices y_A, y_B, y_C, y_D of the fractions of the total message in each node, assuming that fractions of messages are bought in one node and sold in another. The net benefit obtained in such a transaction would be the difference of prices between the arriving and departing nodes. If we normalize these conditions by setting $y_A = 0$, the benefits are

$$\begin{array}{l} y_C - y_A = y_C \\ y_D - y_A = y_D \\ y_C - y_D \\ y_B - y_C \\ y_B - y_D \end{array}$$

We must ensure that these benefits are not greater than the prices we already know:

$$\begin{aligned} y_C &\leq 4 \\ y_D &\leq 2 \\ y_C - y_D &\leq 3 \\ y_B - y_C &\leq 1 \\ y_B - y_D &\leq 2 \end{aligned}$$

The objective is now to maximize the revenue when taking the total message from node A to node B , that is, the dual problem becomes:

$$\begin{aligned} &\text{Maximize } y_B \\ & y_C, y_D, y_B \end{aligned}$$

subject to

$$\begin{aligned} y_C &\leq 4 \\ y_D &\leq 2 \\ y_C - y_D &\leq 3 \\ -y_C + y_B &\leq 1 \\ -y_D + y_B &\leq 2 \end{aligned} \tag{6.47}$$

■

6.4 Duality in Nonlinear Programming

In this section the results obtained for linear programming problems are generalized to nonlinear programming problems. Consider the following general nonlinear primal problem

$$\begin{aligned} &\text{Minimize } z_P = f(\mathbf{x}) \\ & \mathbf{x} \end{aligned} \tag{6.48}$$

subject to

$$\begin{aligned} \mathbf{h}(\mathbf{x}) &= \mathbf{0} \\ \mathbf{g}(\mathbf{x}) &\leq \mathbf{0} \end{aligned} \tag{6.49}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{h} : \mathbb{R}^n \rightarrow \mathbb{R}^\ell$, $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

The dual problem requires the introduction of the dual function defined near $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$ as

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \text{Infimum}_{\mathbf{x}} \{f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x})\} \tag{6.50}$$

where $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$ are the multipliers associated with the constraints (6.49) for the optimal solution \mathbf{x}^* of problem (6.48)-(6.49).

The dual problem of the primal problem (6.48)-(6.49) is then defined as follows:

$$\begin{aligned} &\text{Maximize } z_D = \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ & \boldsymbol{\lambda}, \boldsymbol{\mu} \end{aligned} \tag{6.51}$$

subject to

$$\boldsymbol{\mu} \geq \mathbf{0} \tag{6.52}$$

Using the Lagrangian

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x})$$

we can rewrite the dual problem as

$$\text{Maximize}_{\boldsymbol{\lambda}, \boldsymbol{\mu}; \boldsymbol{\mu} \geq 0} \left[\text{Infimum}_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \right] \quad (6.53)$$

Illustrative Example 6.5 (Dual function). Consider the same problem as the Illustrative Example 6.1. To obtain the dual function $\phi(\mu_1, \mu_2, \mu_3)$ one needs to obtain the Infimum:

$$\text{Infimum}_{x_1, x_2} \mathcal{L}(x_1, x_2; \mu_1, \mu_2, \mu_3)$$

However, a global infimum does not exist for any possible set of (μ_1, μ_2, μ_3) , because the only candidate for relative minimum, from (6.14) and (6.15), is:

$$x_1 = \frac{\mu_1 - \mu_2}{2}; \quad x_2 = \frac{\mu_1 - \mu_3}{2} \quad (6.54)$$

but the Hessian at this point becomes

$$\begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix}$$

which is negative definite.

Consequently, the infimum must be found at the boundary. Since (6.13) is a triangle we must try first to find local minimum at each side and then at the vertices. The Lagrangian function on the side $x_1 + x_2 = 1$ becomes:

$$\begin{aligned} \mathcal{L}(x_1, x_2; \mu_1, \mu_2, \mu_3) &= -x_1^2 - (1 - x_1)^2 + \mu_2(-x_1) + \mu_3(x_1 - 1) \\ &= -2x_1^2 + x_1(2 - \mu_2 + \mu_3) - 1 - \mu_3 \end{aligned}$$

which has no local minimum for $0 < x_1 < 1$.

The Lagrangian function on the side $x_1 = 0$ becomes:

$$\mathcal{L}(x_1, x_2; \mu_1, \mu_2, \mu_3) = -x_2^2 + \mu_1(x_2 - 1) - \mu_3x_2$$

which has no local minimum for $0 < x_1 < 1$.

The Lagrangian function on the side $x_2 = 0$ becomes:

$$\mathcal{L}(x_1, x_2; \mu_1, \mu_2, \mu_3) = -x_1^2 + \mu_1(x_1 - 1) - \mu_2x_1$$

which has no local minimum for $0 < x_1 < 1$.

Then, we need to calculate the values at the vertices and obtain the minimum.

The Lagrangian function on the vertex $(x_1, x_2) = (0, 0)$ is $-\mu_1$.

The Lagrangian function on the vertex $(x_1, x_2) = (0, 1)$ is $-1 - \mu_3$.

The Lagrangian function on the vertex $(x_1, x_2) = (1, 0)$ is $-1 - \mu_2$.

Since the minimum sought is the minimum of the values at the three vertices, the dual function is:

$$\phi(\mu_1, \mu_2, \mu_3) = \begin{cases} -\mu_1 & \text{if } \mu_1 \geq \mu_3 + 1; \mu_1 \geq \mu_2 + 1 \\ -1 - \mu_2 & \text{if } \mu_2 \geq \mu_1 - 1; \mu_2 \geq \mu_3 \\ -1 - \mu_3 & \text{if } \mu_3 \geq \mu_1 - 1; \mu_3 \geq \mu_2 \end{cases}$$

which clearly is not differentiable.

Then, the dual problem becomes:

$$\begin{array}{ll} \text{Maximize} & \phi(\mu_1, \mu_2, \mu_3) \\ & \mu_1, \mu_2, \mu_3 \end{array}$$

subject to

$$\mu_1, \mu_2, \mu_3 \geq 0$$

which solution is

$$\mu_1 = 1; \mu_2 = 0; \mu_3 = 0; z = -1$$

■

Remark 6.4 We assume that f, \mathbf{h} , and \mathbf{g} are such that the infimum of the Lagrangian function is always attained at some \mathbf{x} , so that the “infimum” operation in (6.50) and (6.53) can be replaced by the “minimum” operation. Then, problem (6.53) is referred to as the *max-min dual problem*.

In this situation, if we denote by $\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ a point where the minimum of the Lagrangian is attained (considered as a function of the multipliers), we can write

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}))$$

■

On the other hand, under the assumption of local convexity of f and \mathbf{g} , the Lagrangian is locally convex in \mathbf{x} , and therefore at the minimum its gradient should vanish

$$\nabla \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \nabla f(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) + \boldsymbol{\mu}^T \nabla \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) = \mathbf{0} \quad (6.55)$$

This last identity can be utilized to derive an expression for the gradient of the dual function as well as its Hessian. Indeed, according to the chain rule

$$\begin{aligned} \nabla_{\boldsymbol{\mu}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= [\nabla f(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) + \boldsymbol{\lambda}^T \nabla \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) + \boldsymbol{\mu}^T \nabla \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}))]^T \times \\ &\quad \nabla_{\boldsymbol{\mu}} \mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) \end{aligned} \quad (6.56)$$

and by (6.55)

$$\boldsymbol{\mu} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) \quad (6.57)$$

One must be careful in using expression (6.57) because of the discontinuity of (6.56) with respect to $\boldsymbol{\mu}$. Nevertheless, this is not a problem in practice.

Similarly

$$\nabla_{\boldsymbol{\lambda}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) \quad (6.58)$$

Note that the gradient of the dual function is the vector of mismatches of the corresponding constraints.

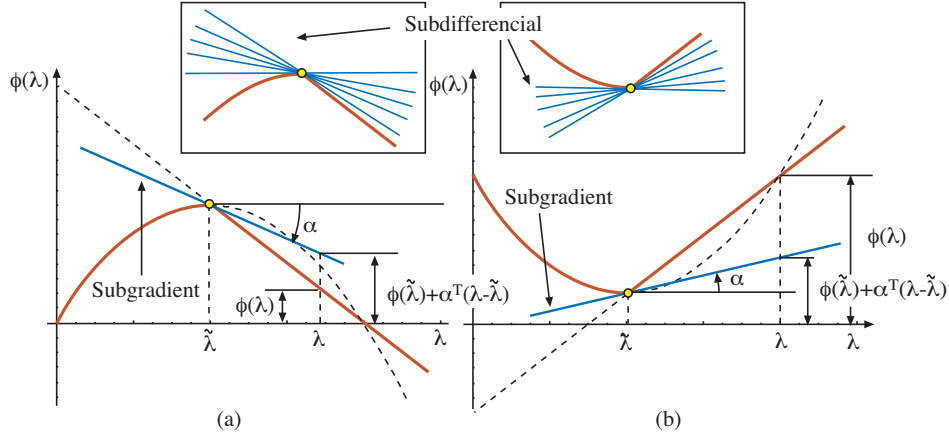


Figure 6.9: Subgradient and subdifferential graphical illustration: (a) concave function and (b) convex function.

Remark 6.5 For non differentiable functions, the role that gradients play for differentiable functions is done by the subgradient. Therefore, the important concept of subgradient is defined below. ■

Definition 6.3 (Subgradient and subdifferential). Let \mathcal{C} be a convex nonempty set in \Re^n and let $\phi : \mathcal{C} \rightarrow \Re$ be convex. Then, α is called a subgradient of $\phi(\lambda)$ at $\tilde{\lambda} \in \mathcal{C}$ if

$$\phi(\lambda) \geq \phi(\tilde{\lambda}) + \alpha^T(\lambda - \tilde{\lambda}); \quad \forall \lambda \in \mathcal{C} \quad (6.59)$$

Analogously, let $\phi : \mathcal{C} \rightarrow \Re$ be concave. Then, α is called a subgradient of $\phi(\lambda)$ at $\tilde{\lambda} \in \mathcal{C}$ if

$$\phi(\lambda) \leq \phi(\tilde{\lambda}) + \alpha^T(\lambda - \tilde{\lambda}); \quad \forall \lambda \in \mathcal{C} \quad (6.60)$$

The set of all subgradients of $\phi(\lambda)$ in $\tilde{\lambda}$ is a convex set known as the subdifferential of $\phi(\lambda)$ at $\tilde{\lambda}$.

Figure 6.9 illustrates the concepts of subgradient and subdifferential. ■

Illustrative Example 6.6 (Subgradient). Consider the following problem:

$$\begin{aligned} &\text{Minimize} && -2x_1 + x_2 \\ &&& x_1, x_2 \end{aligned}$$

subject to

$$\begin{aligned} x_1 + x_2 &= \frac{5}{2} \\ (x_1, x_2) &\in \mathbf{X} \end{aligned}$$

where $\mathbf{X} = \{(0, 0), (0, 2), (2, 0), (2, 2), (5/4, 5/4)\}$.

The dual function $\phi(\lambda)$ is obtained solving the following problem:

$$\phi(\lambda) = \underset{x_1, x_2}{\text{minimum}} \quad -2x_1 + x_2 + \lambda \left(x_1 + x_2 - \frac{5}{2} \right)$$

subject to

$$(x_1, x_2) \in \mathbf{X}$$

The explicit expression of the dual function for this example is given by

$$\phi(\lambda) = \begin{cases} -2 + \frac{3\lambda}{2} & \text{if } \lambda \leq -1 \\ -4 - \frac{\lambda}{2} & \text{if } -1 \leq \lambda \leq 2 \\ -\frac{5\lambda}{2} & \text{if } \lambda \geq 2 \end{cases}$$

Differentiating with respect to λ we obtain

$$\alpha(\lambda) = \frac{d\phi(\lambda)}{d\lambda} = \begin{cases} \frac{3}{2} & \text{if } \lambda \leq -1 \\ -\frac{1}{2} & \text{if } -1 \leq \lambda \leq 2 \\ -\frac{5}{2} & \text{if } \lambda \geq 2 \end{cases}$$

that is the subgradient of the dual function. Note that at the values $\lambda = -1$ and $\lambda = 2$ there are many possible subgradients, thus their corresponding subdifferentials are

$$\begin{aligned} \tilde{\lambda} = -1; \quad \alpha &\in [-1/2, 3/2] \\ \tilde{\lambda} = 2; \quad \alpha &\in [-5/2, -1/2] \end{aligned}$$

as it is shown in Figure 9.4. ■

Concerning the Hessian, notice that by differentiating (6.57) and (6.58) with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, respectively, we obtain

$$\nabla_{\boldsymbol{\mu}}^2 \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) \quad \nabla_{\boldsymbol{\mu}} \mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \quad (6.61)$$

and

$$\nabla_{\boldsymbol{\lambda}}^2 \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) \quad \nabla_{\boldsymbol{\lambda}} \mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}) \quad (6.62)$$

To obtain an expression for $\nabla_{\boldsymbol{\mu}} \mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})$ and $\nabla_{\boldsymbol{\lambda}} \mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})$, we differentiate (6.55) with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$, respectively. The results are

$$\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}), \boldsymbol{\lambda}, \boldsymbol{\mu}) \quad \nabla_{\boldsymbol{\mu}} \mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) = \mathbf{0} \quad (6.63)$$

and a similar equation exchanging $\boldsymbol{\lambda}$ by $\boldsymbol{\mu}$ and \mathbf{h} by \mathbf{g} , i.e.,

$$\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}), \boldsymbol{\lambda}, \boldsymbol{\mu}) \quad \nabla_{\boldsymbol{\lambda}} \mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}) + \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) = \mathbf{0} \quad (6.64)$$

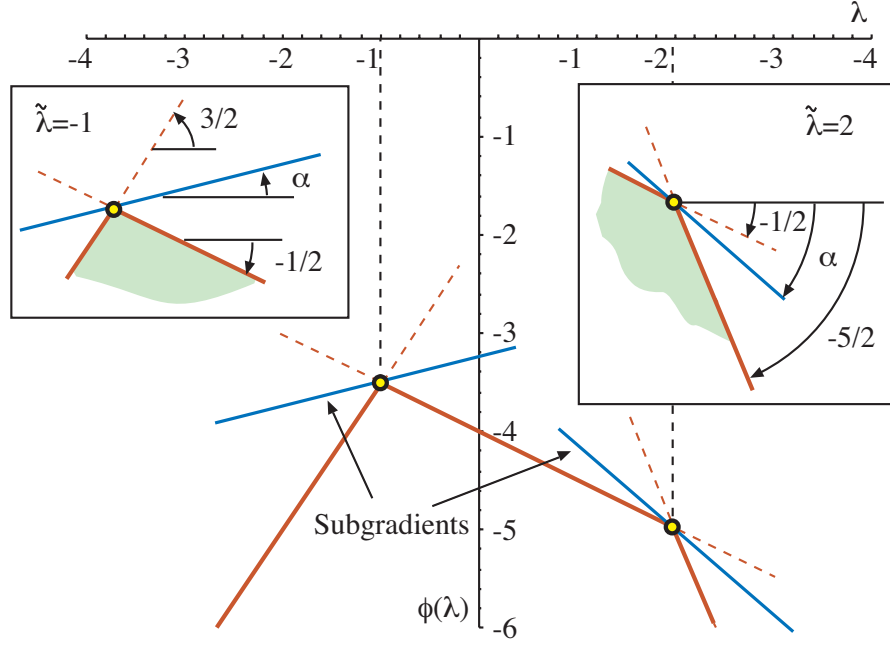


Figure 6.10: Graphical illustration of the dual function, subgradient and subdifferential for the Illustrative Example 6.6.

Taking these equations into (6.61) and (6.62) respectively, we obtain

$$\nabla_{\boldsymbol{\mu}}^2 \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) [\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}), \boldsymbol{\lambda}, \boldsymbol{\mu})]^{-1} \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) \quad (6.65)$$

and the parallel formula for $\nabla_{\boldsymbol{\lambda}}^2 \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$:

$$\nabla_{\boldsymbol{\lambda}}^2 \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) [\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}), \boldsymbol{\lambda}, \boldsymbol{\mu})]^{-1} \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) \quad (6.66)$$

Finally, if we are interested in the mixed second partial derivatives

$$\nabla_{\boldsymbol{\lambda} \boldsymbol{\mu}}^2 \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

similar computations lead to

$$\nabla_{\boldsymbol{\lambda} \boldsymbol{\mu}}^2 \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) [\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}), \boldsymbol{\lambda}, \boldsymbol{\mu})]^{-1} \nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) \quad (6.67)$$

and similarly

$$\nabla_{\boldsymbol{\lambda} \boldsymbol{\mu}}^2 \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\nabla_{\mathbf{x}} \mathbf{h}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) [\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu}), \boldsymbol{\lambda}, \boldsymbol{\mu})]^{-1} \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}(\boldsymbol{\lambda}, \boldsymbol{\mu})) \quad (6.68)$$

To show that the above definition of a dual problem is equivalent to Definition 6.2, valid for linear programming problems, we provide the following example.

Illustrative Example 6.7 (Linear programming). Consider the dual problem of the linear programming primal problem:

$$\text{Minimize } \mathbf{c}^T \mathbf{x} \quad (6.69)$$

subject to

$$\mathbf{Ax} = \mathbf{a} \quad (6.70)$$

$$\mathbf{Bx} \leq \mathbf{b}. \quad (6.71)$$

Using the rules in Section 6.3.2, page 191 the dual problem becomes:

$$\text{Maximize } \lambda^T \mathbf{a} + \mu^T \mathbf{b} \quad (6.72)$$

$$\lambda, \mu$$

subject to

$$\mathbf{A}^T \lambda + \mathbf{B}^T \mu = \mathbf{c} \quad (6.73)$$

and

$$\mu \leq \mathbf{0} \quad (6.74)$$

On the other hand, the Lagrangian function of (6.69)-(6.71) becomes

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \lambda, \mu) &= \mathbf{c}^T \mathbf{x} + \lambda^T (\mathbf{Ax} - \mathbf{a}) + \mu^T (\mathbf{Bx} - \mathbf{b}) \\ &= (\mathbf{c}^T + \lambda^T \mathbf{A} + \mu^T \mathbf{B}) \mathbf{x} - \lambda^T \mathbf{a} - \mu^T \mathbf{b} \end{aligned} \quad (6.75)$$

For this function to have a minimum we must have

$$(\mathbf{c}^T + \lambda^T \mathbf{A} + \mu^T \mathbf{B}) = \mathbf{0} \quad (6.76)$$

that leads to

$$\lambda^T \mathbf{A} + \mu^T \mathbf{B} = -\mathbf{c}^T \quad (6.77)$$

and the minimum is

$$\phi(\lambda, \mu) = -\lambda^T \mathbf{a} - \mu^T \mathbf{b} \quad (6.78)$$

Thus, the dual problem becomes

$$\text{Maximize } -\lambda^T \mathbf{a} - \mu^T \mathbf{b} \quad (6.79)$$

$$\lambda, \mu$$

subject to

$$\lambda^T \mathbf{A} + \mu^T \mathbf{B} = -\mathbf{c}^T \quad (6.80)$$

and

$$\mu \geq \mathbf{0} \quad (6.81)$$

that, changing signs to λ and μ , is equivalent to (6.72)-(6.74).

Note that the dual variables are λ and μ .

■

Illustrative Example 6.8 (Numerical example). Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & z_P = 40x_1 + 56x_2 \\ & x_1, x_2 \end{array}$$

subject to

$$\begin{array}{rcl} 4x_1 & +4x_2 & \geq 70 \\ 3x_1 & +7x_2 & \geq 90 \\ x_1, x_2 & & \geq 0 \end{array}$$

The optimal solution of this problem is $x_1^* = 65/8$, $x_2^* = 75/8$, and the optimal objective function value is 850.

The above problem can be written as

$$\begin{array}{ll} \text{Minimize} & z = 40x_1 + 56x_2 \\ & x_1, x_2 \end{array}$$

subject to

$$\begin{array}{rcl} -4x_1 & -4x_2 & \leq -70 \\ -3x_1 & -7x_2 & \leq -90 \\ -x_1 & & \leq 0 \\ & -x_2 & \leq 0 \end{array}$$

The Lagrangian function for this problem is:

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}) &= 40x_1 + 56x_2 + \mu_1(-4x_1 - 4x_2 + 70) + \mu_2(-3x_1 - 7x_2 + 90) \\ &\quad - \mu_3x_1 - \mu_4x_2 \\ &= (40 - 4\mu_1 - 3\mu_2 - \mu_3)x_1 + (56 - 4\mu_1 - 7\mu_2 - \mu_4)x_2 \\ &\quad + 70\mu_1 + 90\mu_2 \end{aligned}$$

that has a minimum only if (note the non-negativity of x_1, x_2):

$$\begin{array}{rcl} 4\mu_1 & +3\mu_2 & +\mu_3 & = & 40 \\ 4\mu_1 & +7\mu_2 & & +\mu_4 & = & 56 \end{array}$$

The minimum is

$$\phi(\boldsymbol{\mu}) = 70\mu_1 + 90\mu_2$$

Then, we conclude that the dual problem is:

$$\begin{array}{ll} \text{Maximize} & \phi(\boldsymbol{\mu}) = 70\mu_1 + 90\mu_2 \\ & \mu_1, \mu_2 \end{array}$$

subject to

$$\begin{array}{rcl} 4\mu_1 & +3\mu_2 & +\mu_3 & = & 40 \\ 4\mu_1 & +7\mu_2 & & +\mu_4 & = & 56 \\ \mu_1 & & & & \geq & 0 \\ & \mu_2 & & & \geq & 0 \\ & & \mu_3 & & \geq & 0 \\ & & & \mu_4 & \geq & 0 \end{array}$$

Table 6.2: Numerical example: Primal and dual problems with their solutions.

Primal			Dual		
Minimize	$z_P = 40x_1 + 56x_2$		Maximize	$z_D = 70\mu_1 + 90\mu_2$	
	x_1, x_2			μ_1, μ_2	
subject to			subject to		
	$4x_1$	$+4x_2 \geq 70$		$4\mu_1$	$+3\mu_2 \leq 40$
	$3x_1$	$+7x_2 \geq 90$		$4\mu_1$	$+7\mu_2 \leq 56$
	x_1	≥ 0		μ_1	≥ 0
		$x_2 \geq 0$			$\mu_2 \geq 0$
Optimal solution					
	$x_1^* = 65/8; x_2^* = 75/8$			$\mu_1^* = 7; \mu_2^* = 4$	

that, considering μ_3 and μ_4 as slack variables is equivalent to:

$$\begin{aligned} \text{Maximize } z_D &= 70\mu_1 + 90\mu_2 \\ \mu_1, \mu_2 & \\ & \\ & 4\mu_1 + 3\mu_2 \leq 40 \\ & 4\mu_1 + 7\mu_2 \leq 56 \\ & \mu_1 \geq 0 \\ & \mu_2 \geq 0 \end{aligned}$$

The optimal solution of this problem is $\mu_1^* = 7$, $\mu_2^* = 4$, and the optimal objective function value is 850.

This means (see Table 6.2) that the primal objective function z_P will change in 7 units per each unit of change of the right hand side which value is 70, and 4 units per each unit of change of the right hand side which value is 90, and that dual objective function z_D will change in $65/8$ units per each unit of change of the right hand side 40, and $75/8$ units per each unit of change of the right hand side 56. ■

The following theorem shows that any value of the objective function of the dual problem is a lower bound of the optimal value of the objective function of the primal problem. This may be used to terminate computations in an iterative algorithm where such values are available.

Theorem 6.3 (Weak duality). *For any feasible solution \mathbf{x} of the primal problem (6.48)-(6.49) and for any feasible solution, $\boldsymbol{\lambda}, \boldsymbol{\mu}$, of the dual problem (6.51)-(6.52), the following holds*

$$f(\mathbf{x}) \geq \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \tag{6.82}$$

Proof. From the definition of λ, μ , for every feasible \mathbf{x} and for every $\mu \geq 0, \mu \in \mathbb{R}^m$ and $\lambda \in \mathbb{R}^\ell$, we have

$$\phi(\lambda, \mu) = \underset{\mathbf{y}}{\text{Infimum}} [f(\mathbf{y}) + \lambda^T \mathbf{h}(\mathbf{y}) + \mu^T \mathbf{g}(\mathbf{y})] \leq f(\mathbf{x}) + \lambda^T \mathbf{h}(\mathbf{x}) + \mu^T \mathbf{g}(\mathbf{x}) \quad (6.83)$$

and

$$f(\mathbf{x}) + \lambda^T \mathbf{h}(\mathbf{x}) + \mu^T \mathbf{g}(\mathbf{x}) \leq f(\mathbf{x}) \quad (6.84)$$

because $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ and $\mu^T \mathbf{g}(\mathbf{x}) < 0$ for being feasible solutions, therefore

$$\phi(\lambda, \mu) \leq f(\mathbf{x}) \quad (6.85)$$

■

If the feasible regions of both primal and dual problems are nonempty, then (6.82) implies that $\sup(\phi)$ and $\inf(f)$, both taken over their respective constraint sets, are finite. Certainly, neither the supremum nor the infimum need to be attained at any feasible point. If they are, then (6.82) yields the result that if both primal and dual are feasible, both have optimal solutions. A corollary of high practical interest is stated below.

Corollary 6.2 (Primal and dual optimality)

1. The following relation always holds

$$\sup\{\phi(\lambda, \mu) | \mu \geq \mathbf{0}\} \leq \inf\{f(\mathbf{x}) | \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\} \quad (6.86)$$

2. If $f(\mathbf{x}^*) = \phi(\lambda^*, \mu^*)$ for some feasible solution \mathbf{x}^* of the primal problem (6.48)-(6.49) and for some feasible solution (λ^*, μ^*) of the dual problem (6.51)-(6.52), then \mathbf{x}^* and (λ^*, μ^*) are optimal solutions of the primal and dual problems, respectively. This solution defines a saddle point of the Lagrangian that corresponds to a maximum with respect to λ and μ and a minimum with respect to \mathbf{x} (see Figure 6.11).

3. If $\sup\{\phi(\lambda, \mu) : \mu \geq \mathbf{0}\} = +\infty$, then the primal problem has no feasible solution.

4. If $\inf\{f(\mathbf{x}) : \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\} = -\infty$, then $\phi(\lambda, \mu) = -\infty$ for every $\lambda \in \mathbb{R}^\ell$ and $\mu \geq \mathbf{0}$.

The proof of this corollary can be found for instance in [16].

If equation (6.86) does not hold with equality, the definition of duality gap, stated below, is of interest.

Definition 6.4 (Duality gap). The difference

$$\sup\{\phi(\lambda, \mu) | \mu \geq \mathbf{0}\} - \inf\{f(\mathbf{x}) | \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$$

is called the duality gap of the dual problems (6.48)-(6.49) and (6.51)-(6.52).

■

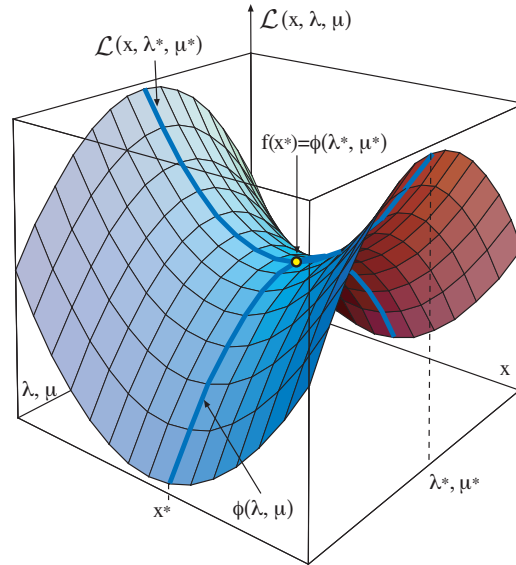


Figure 6.11: Graphical illustration of saddle point $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ of the Lagrangian.

Illustrative Example 6.9 (Duality gap). Consider the same problem as the one considered in the Illustrative Example 6.6:

$$\begin{aligned} &\text{minimize} && -2x_1 + x_2 \\ &&& x_1, x_2 \end{aligned}$$

subject to

$$\begin{aligned} x_1 + x_2 &= \frac{5}{2} \\ (x_1, x_2) &\in \mathbf{X} \end{aligned}$$

where $\mathbf{X} = \{(0, 0), (0, 2), (2, 0), (2, 2), (5/4, 5/4)\}$. Its dual function is given by the explicit expression

$$\phi(\lambda) = \begin{cases} -2 + \frac{3\lambda}{2} & \text{if } \lambda \leq -1 \\ -4 - \frac{\lambda}{2} & \text{if } -1 \leq \lambda \leq 2 \\ -\frac{5\lambda}{2} & \text{if } \lambda \geq 2 \end{cases}$$

shown in Figure 9.5. The optimal solution of the dual problem is $\lambda^* = -1$ with objective function value $\phi(\lambda^*) = -7/2$.

It is easy to verify that $x_1^* = 5/4$ and $x_2^* = 5/4$ is the optimal solution of the primal problem with an objective function value equal to $f(\mathbf{x}^*) = -5/4$.

Therefore, the difference $f(\mathbf{x}^*) - \phi(\lambda^*) = -5/4 + 7/2 = 9/4$ is the duality gap shown in Figure 9.5.

■

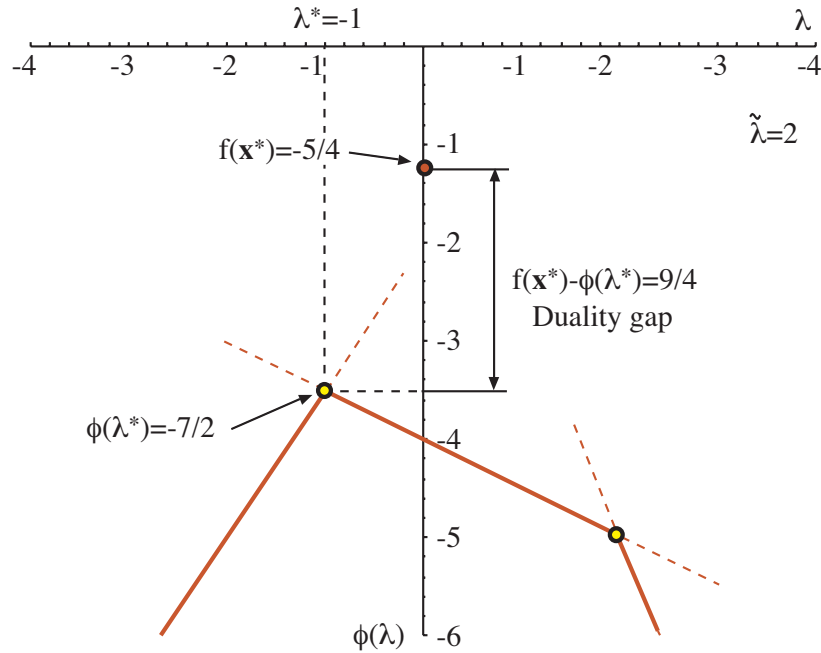


Figure 6.12: Graphical illustration of the duality gap for the Illustrative Example 6.9.

If a multiplier vector solves the dual problem and there is no duality gap, that is, if the relationship (6.86) is strictly satisfied, then, the solutions of the Lagrangian associated with this multiplier vector are optimal solutions of the primal problem. This result introduces a way to solve the primal problem by means of solving the dual problem. The main result consists of obtaining conditions that guarantee the nonexistence of the duality gap. This is the situation of problems in which some convexity conditions hold.

Theorem 6.4 (Duality theorem for convex problems). *Consider a locally convex mathematical programming problem as (6.48)-(6.49). If \mathbf{x}^* solves the primal problem, then its associated vector of multipliers $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ solves the dual problem, and $(\boldsymbol{\mu}^*)^T \mathbf{g}(\mathbf{x}^*) = 0$. Conversely, if $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ solves the dual problem, and there exists a solution, \mathbf{x}^* , of the Lagrangian associated to this vector of multipliers such that $(\boldsymbol{\mu}^*)^T \mathbf{g}(\mathbf{x}^*) = 0$, then \mathbf{x}^* is an optimal solution of the primal problem.*

Proof. If \mathbf{x}^* solves the primal problem and $(\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*)$ is its associated vector of multipliers, we know that they should be a solution of the KKT conditions

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^* \nabla \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^* \nabla \mathbf{g}(\mathbf{x}^*) &= \mathbf{0} \\ \boldsymbol{\mu}^* \mathbf{g}(\mathbf{x}^*) &= \mathbf{0} \\ \boldsymbol{\mu}^* &\geq \mathbf{0} \\ \mathbf{g}(\mathbf{x}^*) &\leq \mathbf{0} \\ \mathbf{h}(\mathbf{x}^*) &= \mathbf{0} \end{aligned}$$

The minimization problem defining $\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is

$$\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \text{Inf}_{\mathbf{x}} \{f(\mathbf{x}) + \boldsymbol{\lambda}^{*T} \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^{*T} \mathbf{g}(\mathbf{x})\}$$

The condition coming from the above KKT system

$$\nabla f(\mathbf{x}^*) + \boldsymbol{\lambda}^* \nabla \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^* \nabla \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$$

together with the local convexity of the objective function (note that $\boldsymbol{\mu}^* \geq \mathbf{0}$)

$$f(\mathbf{x}) + \boldsymbol{\lambda}^{*T} \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^{*T} \mathbf{g}(\mathbf{x})$$

imply that \mathbf{x}^* is a local minimizer for the problem to determine $\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ and therefore

$$\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\mathbf{x}^*)$$

and by Corollary 6.2 the pair $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is an optimal solution of the dual problem. Note that the condition

$$\boldsymbol{\mu}^* \mathbf{g}(\mathbf{x}^*) = \mathbf{0}$$

also holds.

Conversely, assume that the pair $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is an optimal solution for the dual problem, and that \mathbf{x}^* is a corresponding solution of the associated Lagrangian such that $(\boldsymbol{\mu}^*)^T \mathbf{g}(\mathbf{x}^*) = 0$. Applying the KKT conditions to the dual, it can be concluded that

$$\nabla_{\boldsymbol{\mu}} \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq \mathbf{0}, \quad \nabla_{\boldsymbol{\lambda}} \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$$

But taking into account (6.57) and (6.58), we obtain

$$\mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}; \quad \mathbf{h}(\mathbf{x}^*) = \mathbf{0}$$

This implies that \mathbf{x}^* is feasible for the primal. Then

$$\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\mathbf{x}^*) + \boldsymbol{\lambda}^{*T} \mathbf{h}(\mathbf{x}^*) + \boldsymbol{\mu}^{*T} \mathbf{g}(\mathbf{x}^*) = f(\mathbf{x}^*)$$

and by Corollary 6.2 we conclude the proof. ■

In the following, relevant sensitivity results are derived.

Theorem 6.5 (Sensitivities are given by dual variables). *Consider the following general nonlinear primal problem*

$$\text{Minimize}_{\mathbf{x}} \quad z_P = f(\mathbf{x}) \tag{6.87}$$

subject to

$$\begin{aligned} \mathbf{h}(\mathbf{x}) &= \mathbf{a} \\ \mathbf{g}(\mathbf{x}) &\leq \mathbf{b} \end{aligned} \tag{6.88}$$

and its dual problem with $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ as dual variables. Then, the values of the dual variables, also called shadow prices, give the sensitivity of the objective function optimal value to changes in the primal constraints, i.e.,

$$\left. \frac{\partial f(\mathbf{x})}{\partial a_k} \right|_{\mathbf{x}^*} = -\lambda_k; \quad \left. \frac{\partial f(\mathbf{x})}{\partial b_j} \right|_{\mathbf{x}^*} = -\mu_j, \tag{6.89}$$

Proof.

If the optimal values for the primal and dual exist, they satisfy:

$$\begin{aligned} f(\mathbf{x}) &= \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}) - \boldsymbol{\lambda}^T \mathbf{a} - \boldsymbol{\mu}^T \mathbf{b} \\ &= f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{h}(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{g}(\mathbf{x}) - \sum_{k=1}^{\ell} \lambda_k a_k - \sum_{j=1}^m \mu_j b_j \end{aligned}$$

then, we have (6.89), which proves that the values of the dual variables are the sensitivities of the objective function to changes in the right hand sides of (6.88). ■

Illustrative Example 6.10 (Dual problem). Consider the following (primal) problem.

$$\begin{aligned} \text{Minimize} \quad & z_P = x_1^2 + x_2^2 \\ & x_1, x_2 \end{aligned}$$

subject to

$$\begin{aligned} x_1 + x_2 &\leq 4 \\ -x_1 &\leq 0 \\ -x_2 &\leq 0 \end{aligned}$$

The unique optimal solution to this problem is $(x_1^*, x_2^*) = (0, 0)$ with the optimal objective function value equal to 0. Let us formulate the dual problem. We need first to find the dual function:

$$\begin{aligned} \phi(\mu_1, \mu_2, \mu_3) &= \min_{\mathbf{x} \in \mathbb{R}^2} \{x_1^2 + x_2^2 + \mu_1(x_1 + x_2 - 4) + \mu_2(-x_1) + \mu_3(-x_2)\} \\ &= \min_{x_1} \{x_1^2 + (\mu_1 - \mu_2)x_1\} + \min_{x_2} \{x_2^2 + (\mu_1 - \mu_3)x_2\} - 4\mu_1 \end{aligned}$$

After algebra manipulation we explicitly find the dual function

$$\phi(\boldsymbol{\mu}) = -\frac{1}{4}[(\mu_1 - \mu_2)^2 + (\mu_1 - \mu_3)^2] - 4\mu_1$$

an the dual problem becomes

$$\begin{aligned} \text{Supremum} \quad & \phi(\boldsymbol{\mu}) \\ & \boldsymbol{\mu} \geq 0 \end{aligned}$$

Note that the dual function is concave. The dual problem optimum is attained when $\mu_1^* = \mu_2^* = \mu_3^* = 0$ and $z_D^* = \phi(\boldsymbol{\mu}^*) = 0$. ■

It is relevant at this point to stress that in some instances it may be important to look at duality partially, in the sense that some but not all constraints are dualized. Those restrictions not dualized are incorporated as such in the definition of the dual function. This is in fact the underlying driving motivation for the Lagrangian decomposition techniques.

6.5 Illustration of Duality and Separability

To illustrate the relationship between primal and dual problems and separability, consider a single commodity that is produced and supplied under the following conditions:

1. Different producers compete in a single market.
2. Demand in every time period of a study horizon has to be strictly satisfied.
3. No storage capacity is available.
4. Producers have different production cost functions and production constraints.

We consider two different approaches to this problem (i) the centralized or primal approach and (ii) the competitive or dual approach.

A centralized (primal) approach to this problem is as follows. Producers elect a system operator which is in charge of guaranteeing the supply of demand every time period. To this end, the system operator:

1. Is entitled to know the actual production cost function of every producer and its production constraints.
2. Knows the actual commodity demand in every time period.
3. Has the power to dictate the actual production of every producer.

The main elements of this problem are

1. Data.

$c_{ti}(p_{ti})$: production cost of producer i in period t to produce p_{ti} ,

Π_i : producer i feasible operating region

d_t : demand in period t

T : number of time periods

I : number of producers

2. Variables.

p_{ti} : amount of commodity produced by producer i during period t

3. Constraints.

$$\sum_{i=1}^I p_{ti} = d_t; \quad t = 1, \dots, T \quad (6.90)$$

$$p_{ti} \in \Pi_i; \quad t = 1, \dots, T; i = 1, \dots, I \quad (6.91)$$

This enforces for every time period the strict supply of demand, and the operational limits of every producer.

4. **Function to be optimized.** Since the operator has the objective of minimizing total production cost, he has to solve the following problem

$$\text{Minimize}_{p_{ti}} \sum_{t=1}^T \sum_{i=1}^I c_{ti}(p_{ti}) \quad (6.92)$$

which is the sum of the production costs of all producers over all time periods.

Thus, the primal problem is as follows.

$$\text{Minimize}_{p_{ti}} \sum_{t=1}^T \sum_{i=1}^I c_{ti}(p_{ti}) \quad (6.93)$$

subject to

$$\sum_{i=1}^I p_{ti} = d_t; \quad t = 1, \dots, T \quad (6.94)$$

$$p_{ti} \in \Pi_i; \quad t = 1, \dots, T; i = 1, \dots, I \quad (6.95)$$

The solution of this problem provides the optimal level of production of every producer. Then, the system operator communicates its optimal production to every producer, which actually produces this optimal value.

Note that the approach explained is a centralized one. The system operator has knowledge of the production cost functions and the production constraints of every producer, can determine the actual (optimal) production of every producer, and has the power to force every producer to produce its centrally determined optimal value.

We analyze then the structure of the dual problem of the cost minimization primal problem (6.93)–(6.95).

If partial duality over constraints (6.94) is considered, the Lagrangian function has the form

$$\mathcal{L}(p_{ti}, \lambda_t) = \sum_{t=1}^T \sum_{i=1}^I c_{ti}(p_{ti}) + \sum_{t=1}^T \lambda_t \left(d_t - \sum_{i=1}^I p_{ti} \right) \quad (6.96)$$

Arranging terms, this Lagrangian function can be expressed as

$$\mathcal{L}(p_{ti}, \lambda_t) = \sum_{i=1}^I \left[\sum_{t=1}^T (c_{ti}(p_{ti}) - \lambda_t p_{ti}) \right] + \sum_{t=1}^T \lambda_t d_t \quad (6.97)$$

The evaluation of the dual function for a given value of the Lagrange multipliers $\lambda_t = \widehat{\lambda}_t$; $t = 1, \dots, T$, is achieved by solving the problem.

$$\text{Minimize}_{p_{ti}} \mathcal{L}(p_{ti}, \widehat{\lambda}_t) \quad (6.98)$$

subject to

$$p_{ti} \in \Pi_i; \quad t = 1, \dots, T; i = 1, \dots, I \quad (6.99)$$

Using the explicit expression (6.97) for the Lagrangian function, the problem presented above now entails

$$\text{Minimize}_{p_{ti}} \sum_{i=1}^I \left[\sum_{t=1}^T (c_{ti}(p_{ti}) - \hat{\lambda}_t p_{ti}) \right] + \sum_{t=1}^T \hat{\lambda}_t d_t \quad (6.100)$$

subject to

$$p_{ti} \in \Pi_i; \quad t = 1, \dots, T; \quad i = 1, \dots, I \quad (6.101)$$

Taking into account that the expression $\sum_{t=1}^T \hat{\lambda}_t d_t$ is constant, the previous problem reduces to

$$\text{Minimize}_{p_{ti}} \sum_{i=1}^I \left[\sum_{t=1}^T (c_{ti}(p_{ti}) - \hat{\lambda}_t p_{ti}) \right] \quad (6.102)$$

subject to

$$p_{ti} \in \Pi_i; \quad t = 1, \dots, T; \quad i = 1, \dots, I \quad (6.103)$$

This problem has the very remarkable property of being separable; that is, it decomposes by producer. The individual problem of producer j is therefore (note the sign change of the objective function):

$$\text{Maximize}_{p_{tj}} \sum_{t=1}^T [\hat{\lambda}_t p_{tj} - c_{tj}(p_{tj})] \quad (6.104)$$

subject to

$$p_{tj} \in \Pi_j; \quad t = 1, \dots, T \quad (6.105)$$

Therefore, the evaluation of the dual function can be achieved by solving problem (6.104)-(6.105) for every producer, namely, I times.

What is again particularly remarkable is the interpretation of problem (6.104)–(6.105); it represents the maximization of the profit of producer j subject to its production constraints. Lagrange multipliers λ_t ; $t = 1, \dots, T$ are naturally interpreted as the selling prices of the commodity in periods $t = 1, \dots, T$.

It can be concluded that to evaluate the dual function, every producer has to solve its own profit maximization problem, and then, the operator adds the maximum profit of every producer and the constant term $\sum_{t=1}^T \hat{\lambda}_t d_t$.

The duality theorem states that, under certain convexity assumptions, the solutions of the primal and the dual problems are the same in terms of the objective function value. It also states that the evaluation of the dual function for the optimal values of Lagrangian multipliers λ_t^* ; $t = 1, \dots, T$, results in optimal values for productions p_{ti}^* ; $t = 1, \dots, T$; $i = 1, \dots, I$. As a consequence, the solution of the primal problem can be obtained solving the dual one.

Duality theory also states that under very general conditions in the structure of the primal problem, the dual function is concave, and therefore, it can be maximized using a simple steepest-ascent procedure.

To use a steepest-ascent procedure, it is necessary to know the gradient of the dual function. Duality theory provides us with a simple expression for the gradient of the dual function (see (6.57) and (6.58)). For productions p_{ti} ; $t = 1, \dots, T$; $i = 1, \dots, I$, the components of the gradient of the dual function are

$$d_t - \sum_{i=1}^I p_{ti}; \quad t = 1, \dots, T \quad (6.106)$$

That is, the component t of the gradient of the dual function is the actual demand imbalance in period t .

A steepest-ascent procedure to solve the dual problem (and therefore the primal one) includes the following steps:

1. The system operator sets up initial values of Lagrange multipliers (selling prices).
2. Using the Lagrange multipliers presented above, every producer maximizes its own profit meeting its own production constraints.
3. The operator calculates the demand imbalance in every time period (components of the gradient of the dual function).
4. If the norm of the gradient of the dual function is not sufficiently small, a steepest-ascent iteration is performed and the procedure continues in 2. Otherwise the procedure concludes, the dual solution has been attained.

The preceding four-step algorithm can be interpreted in the framework of a competitive market. This is done below.

1. The market operator broadcasts to producers the initial ($k = 1$) value of selling price for every period, $\lambda_t^{(k)}$; $t = 1, \dots, T$.
2. Every producer j solves its profit maximization problem:

$$\text{Maximize}_{p_{tj}} \sum_{t=1}^T \left(\lambda_t^{(k)} p_{tj} - c_{tj}(p_{tj}) \right) \quad (6.107)$$

subject to

$$p_{tj} \in \Pi_j; \quad \forall t \quad (6.108)$$

and sends to the market operator its optimal production schedule, $p_{tj}^{(k)}$; $t = 1, \dots, T$; $j = 1, \dots, I$.

3. The market operator computes the demand imbalance in every time period

$$d_t - \sum_{i=1}^I p_{ti}^{(k)}; \quad t = 1, \dots, T \quad (6.109)$$

and update prices proportionally to imbalances

$$\lambda_t^{(k+1)} = \lambda_t^{(k)} + K \left(d_t - \sum_{i=1}^I p_{ti}^{(k)} \right); \quad t = 1, \dots, T \quad (6.110)$$

where K is a proportionality constant.

4. If prices remain constant in two consecutive rounds, stop; the optimal solution (market equilibrium) has been attained. Otherwise, the algorithm continues in Step 2.

This dual solution algorithm can be interpreted as the functioning of a multiround auction. In fact, economists call this iterative procedure a Walrasian auction.

A Walrasian auction preserves the privacy of the corporate information (cost function and production constraints) of every producer and allows them to maximize their respective own profits.

Some comments are in order. The primal approach is centralized, and the system operator has complete information of the production cost function and operation constraints of every producer. Moreover, it is given the power of deciding how much should produce every producer.

As opposed to the primal approach, the dual approach is decentralized in nature and every producer maintains the privacy of its production cost function and operation constraints. Furthermore, every producer decides how much to produce to maximize its own benefits.

Under some assumptions, duality theory guarantees that both approaches yield identical results.

6.6 Concluding Remarks

This chapter addresses the classical topic of duality in both linear and nonlinear programming. After stating the Karush-Kuhn-Tucker optimality conditions the dual problem is derived for both linear and nonlinear problems. The chapter concludes illustrating how duality can be used to attain a separable dual problem.

Dual variables are the crucial communication instruments used by the subproblems that result from applying any decomposition procedure to a decomposable problem. Moreover, dual variables are sensitivity parameters that inform on the objective function changes due to changes on the right hand sides of the constraints.

This chapter provides background material of high interest to address the next chapters.

Additional material on duality can be found in the classical manuals by Bazaraa et al. [16] and by Luenberger [?].

6.7 Exercises

Exercise 6.1. Solve the following problems by applying KKTs.

1.

$$\begin{array}{ll} \text{Maximize} & z = x_1 - \exp(-x_2) \\ & x_1, x_2 \end{array}$$

subject to

$$\begin{array}{rcl} -\sin x_1 + x_2 & \leq & 0 \\ x_1 & \leq & 3 \end{array}$$

2.

$$\begin{array}{ll} \text{Minimize} & z = (x_1 - 4)^2 + (x_2 - 3)^2 \\ & x_1, x_2 \end{array}$$

subject to

$$\begin{array}{rcl} x_1^2 - x_2 & \leq & 0 \\ x_2 & \leq & 4 \end{array}$$

3.

$$\begin{array}{ll} \text{Maximize} & z = x_2 \\ & x_1, x_2 \end{array}$$

subject to

$$\begin{array}{rcl} x_1^2 - x_2^2 & \leq & 4 \\ -x_1^2 + x_2 & \leq & 0 \\ x_1, x_2 & \geq & 0 \end{array}$$

Exercise 6.2. Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & z = 3x_1 - x_2 + x_4 \\ & x_1, x_2, x_3, x_4 \end{array}$$

subject to

$$\begin{array}{rcl} x_1 + x_2 - x_3 - x_4 & = & 4 \\ 2x_1 - x_2 & + & x_4 \leq 0 \\ 3x_2 + x_3 - 2x_4 & \geq & 1 \\ x_1, x_2 & \geq & 0 \\ x_4 & \leq & 0 \end{array}$$

Build the dual problem, solve both problems, and check that the optimal values coincide.

Exercise 6.3. Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & z = x_1 \\ & x_1, x_2 \end{array}$$

subject to

$$x_1^2 + x_2^2 = 1$$

Compute the dual function and show that it is a concave function. Draw it. Find the optimal solutions for the dual and primal problems by means of comparing their objective functions.

Exercise 6.4. Consider the following problem:

$$\text{Minimize } z = e^{-x}$$

subject to

$$x \geq 0$$

Solve the primal problem. Compute explicitly the dual function and solve the dual problem.

Exercise 6.5. Consider the following problem:

$$\text{Minimize } z = (x_1 - 4)^2 + (x_2 - 1)^2$$

subject to

$$\begin{aligned} 2x_1 + x_2 &\geq 6 \\ -x_1^2 + x_2 &\geq -2 \end{aligned}$$

1. Find the optimal solution for this primal problem geometrically.
2. Obtain graphically the optimal multipliers values associated with both constraints, i. e., the optimal solution of the dual problem.
3. Verify that KKT conditions hold.

Exercise 6.6. Consider the following problem:

$$\text{Minimize } z = 2x_1^2 + x_2^2 - 2x_1x_2 - 6x_2 - 4x_1$$

subject to

$$\begin{aligned} x_1^2 + x_2^2 &= 1 \\ -x_1 + 2x_2 &\leq 0 \\ x_1 + x_2 &\leq 8 \\ x_1, x_2 &\geq 0 \end{aligned}$$

1. Solve the problem using the KKTs
2. Compute explicitly the dual function.

3. Does a duality gap exist?

Exercise 6.7. Consider the same problem as the one in Illustrative Example 6.6, page 202. Obtain the explicit expression of the dual function.

Exercise 6.8. Consider the following problem:

$$\begin{array}{ll} \text{Minimize} & z = x_1 + x_2 \\ & x_1, x_2 \end{array}$$

subject to

$$\begin{array}{ll} x_1 - x_2 & = 0 \\ (x_1, x_2) & \in \mathbf{X} \end{array}$$

where $\mathbf{X} = \{(-4, 0), (4, 0), (0, -4), (0, 4), (-1, -1), (1, 1)\}$.

1. Compute explicitly the dual function.
2. Obtain graphically the solution of the dual problem.
3. Solve the primal problem. Does a duality gap exist?

Exercise 6.9. Write and solve the dual of the problem:

$$\begin{array}{ll} \text{Maximize} & z = 3x_1 - 4x_2 + 9x_3 + x_4 \\ & x_1, x_2, x_3, x_4 \end{array}$$

subject to

$$\begin{array}{llll} x_1 & -5x_2 & +x_3 & \geq 0 \\ 3x_1 & -5x_2 & +x_3 & \geq 10 \\ & x_2, x_3 & & \geq 0 \\ & & x_4 & \leq 0 \end{array}$$

Exercise 6.10. Find the minimum length of a ladder that must lean against a wall if a box of dimensions a and b is placed right at the corner of that same wall. Formulate and solve the primal and dual problems.

Exercise 6.11. Find the closest point of the surface $xyz = 1$ to the origin. Solve the primal and dual problems.

Exercise 6.12. Consider the case of 3 producers of a good competing in a single market during a time horizon of 4 months with demands 5, 6, 7 and 4, respectively, and assume that no storage capacity is available. Suppose that the producers have the following production cost functions

$$c_{t1}(p_{t1}) = 1 + 2p_{t1}^2, \quad c_{t2}(p_{t2}) = 2 + 3p_{t2}^2, \quad c_{t3}(p_{t3}) = 3 + p_{t3}^2$$

where the subindex ti refers to producer i and period t , and it has been assumed that the production cost are independent of the period t . Assume also that the maximum productions of the 3 producers are respectively:

$$C_1 = 2, \quad C_2 = 3, \quad C_3 = 3$$

1. Solve the problem using the centralized approach described in Section 6.5 and show that the optimal solution gives a minimum cost of \$95.1636 for the following productions:

Producer	Time period			
	1	2	3	4
1	1.36	1.80	2.00	1.09
2	0.91	1.20	2.00	0.73
3	2.73	3.00	3.00	2.18

2. Solve the problem using the decentralized approach and two different values of the K constant to update the prices. Compare the number of iterations required for a given error in the prices of the good.
3. Use a large value of K and check that the process blows up.

Chapter 7

Optimization Algorithms for Structural Reliability

Chapter 8

Local Sensitivity Analysis

8.1 Introduction

Sensitivity analysis is the study of how the variation in the output of a model can be apportioned, qualitatively or quantitatively, to different sources of variation. Sensitivity analysis aims to determine how the model depends upon the data or information fed into it, upon its structure and upon the framing assumptions made to build it. As a whole, sensitivity analysis is used to increase the confidence in the model and its predictions, by providing an understanding of how the model response variables respond to changes in the inputs. Adding a sensitivity analysis to an study means adding quality to it.

In this chapter, some methods for sensitivity analysis are discussed.

8.2 A General Method for Sensitivity Analysis

In this section we show how the duality methods can be applied to sensitivity analysis in a straightforward manner. We would like to emphasize that the method to be presented in this section is of general validity. The basic idea is simple. Assume that we desire to know the sensitivity of the objective function to changes in some data values. If we convert the data in artificial variables and lock them, by means of constraints, to their actual values, we obtain a problem that is equivalent to the initial optimization problem but has a constraint such that the values of the dual variables associated with them give the desired sensitivities.

To be more precise, consider the general optimization problem

$$\text{Minimize } f(\mathbf{x}, \boldsymbol{\eta}_0) \quad (8.1)$$

\mathbf{x}

subject to

$$\mathbf{h}(\mathbf{x}, \boldsymbol{\eta}_0) = \mathbf{0} \quad (8.2)$$

$$\mathbf{g}(\mathbf{x}, \boldsymbol{\eta}_0) \leq \mathbf{0} \quad (8.3)$$

where $\boldsymbol{\eta}_0$ is the vector of data values.

Then, this problem is equivalent to the following one:

$$\begin{aligned} \text{Minimize } & f(\mathbf{x}, \boldsymbol{\eta}) \\ \mathbf{x}, \boldsymbol{\eta} & \end{aligned} \quad (8.4)$$

subject to

$$\mathbf{h}(\mathbf{x}, \boldsymbol{\eta}) = \mathbf{0} \quad (8.5)$$

$$\mathbf{g}(\mathbf{x}, \boldsymbol{\eta}) \leq \mathbf{0} \quad (8.6)$$

$$\boldsymbol{\eta} = \boldsymbol{\eta}_0 \quad (8.7)$$

but in this last formulation the values of the dual variables give the partial derivatives of the objective function with respect to the components of $\boldsymbol{\eta}_0$.

Illustrative Example 8.1 (Sensitivities in nonlinear programming). Consider the following optimization problem:

$$\begin{aligned} \text{Minimize } & z = a_1x^2 + a_2y^2 \\ \mathbf{x}, \mathbf{y} & \end{aligned} \quad (8.8)$$

subject to

$$\begin{aligned} xy & \geq a_3 \\ a_4x + y & = a_5 \\ (x - 4)^2 + a_6(y - 2)^2 & \leq a_7 \end{aligned} \quad (8.9)$$

where $a_1 = 1$, $a_2 = 1$, $a_3 = 4$, $a_4 = 1$, $a_5 = 5$, $a_6 = 1$ and $a_7 = 1$.

The Lagrangian function is

$$\mathcal{L}(x, y; \mu, \lambda) = x^2 + y^2 + \lambda(x + y - 5) + \mu_1(4 - xy) + \mu_2((x - 4)^2 + (y - 2)^2 - 1)$$

and the KKT conditions are

$$\frac{\partial \mathcal{L}}{\partial x} = 2x + \lambda - \mu_1y + 2\mu_2(x - 4) = 0 \quad (8.10)$$

$$\frac{\partial \mathcal{L}}{\partial y} = 2y + \lambda - \mu_1x + 2\mu_2(y - 2) = 0 \quad (8.11)$$

$$x + y = 5 \quad (8.12)$$

$$xy \geq 4 \quad (8.13)$$

$$(x - 4)^2 + (y - 2)^2 \leq 1 \quad (8.14)$$

$$\mu_1(4 - xy) = 0 \quad (8.15)$$

$$\mu_2((x - 4)^2 + (y - 2)^2 - 1) = 0 \quad (8.16)$$

$$\mu_1 \geq 0 \quad (8.17)$$

$$\mu_2 \geq 0 \quad (8.18)$$

The only point that satisfies the KKT conditions is $(3, 2)$ that is the optimal point, as it is illustrated in Figure 8.1. The corresponding values of the multipliers are: $\lambda = -4$, $\mu_1 = 0$ and $\mu_2 = 1$.

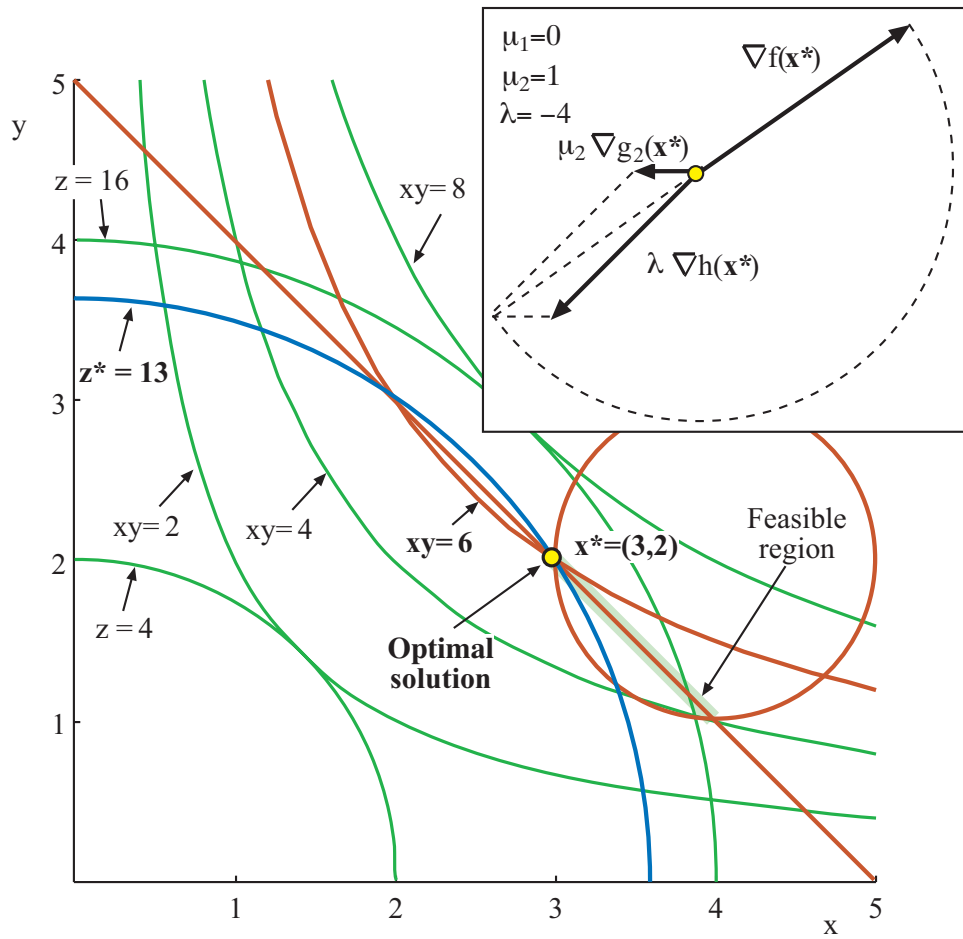


Figure 8.1: Illustration of the minimization problem in Example 8.1.

To study the sensitivity of the objective function optimal value to the \mathbf{a} -parameters, the problem (8.8)-(8.9) can be written as:

$$\begin{aligned} \text{Minimize } z &= a_1x^2 + a_2y^2 \\ x, y, \mathbf{a} \end{aligned}$$

subject to

$$\begin{aligned} xy &\geq a_3 \\ a_4x + y &= a_5 \\ (x - 4)^2 + a_6(y - 2)^2 &\leq a_7 \\ \mathbf{a} &= \mathbf{a}_0 \end{aligned} \tag{8.19}$$

where, since the parameters a_3, a_5 and a_7 already appear on the right hand sides of some constraints, $\mathbf{a} = (a_1, a_2, a_4, a_6)$, and $\mathbf{a}_0 = (1, 1, 1, 1)$.

The resulting values of the sensitivities, i.e., the dual variable values are:

$$\frac{\partial z^*}{\partial a_1} = 9, \quad \frac{\partial z^*}{\partial a_2} = 4, \quad \frac{\partial z^*}{\partial a_3} = 0, \quad \frac{\partial z^*}{\partial a_4} = -12, \quad \frac{\partial z^*}{\partial a_5} = 4, \quad \frac{\partial z^*}{\partial a_6} = 0, \quad \frac{\partial z^*}{\partial a_7} = -1$$

where the asterisks refer to optimal values. ■

8.2.1 Application Example: The Engineering Design.

In this section it is explained how a sensitivity analysis can be done for the engineering design problem described in pages ?? and ?? with practically no additional cost, using standard optimization procedures. The method, as described in the previous section, consists of writing the initial problem in an equivalent form such that the values of some dual variables are the sensitivities looked for. Two different alternatives are considered.

Alternative 1. Updating safety factor bounds

This method is based on observing that the classical problem (??)-(??):

$$\text{Minimizing } c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}) \tag{8.20}$$

$$\bar{\mathbf{d}}$$

subject to

$$g_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}) \geq F_i^0; \forall i \in I \tag{8.21}$$

$$r_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}) \leq 0; \forall j \in J \tag{8.22}$$

is equivalent to the problem:

$$\text{Minimizing } c(\bar{\mathbf{d}}, \boldsymbol{\eta}) \tag{8.23}$$

$$\bar{\mathbf{d}}, \boldsymbol{\eta}$$

subject to

$$g_i(\bar{\mathbf{d}}, \boldsymbol{\eta}) \geq F_i^0; \forall i \in I \tag{8.24}$$

$$r_j(\bar{\mathbf{d}}, \boldsymbol{\eta}) \leq 0; \forall j \in J \tag{8.25}$$

and

$$\boldsymbol{\eta} = \tilde{\boldsymbol{\eta}} \quad (8.26)$$

Since constraints (8.26) involve the data in their right hand sides, and the dual variables are the sensitivities of the objective function value to changes in the constraints right hand side terms, the desired sensitivities are the values of the corresponding dual variables. In other words, the values of the dual variables associated with the constraints (8.26) give how much the objective function $c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}})$ changes with a small unit increment change of the corresponding data parameter.

One of the interesting features of the proposed procedure is that it provides the engineer with information of how the solution varies as the values of the project variables are changed.

Remark 8.1 *Note that problem (8.23)-(8.26) need to be solved only once, i.e., one can use problem (8.20)-(8.22) for the iterative process and, once convergence is obtained, switch to this modified problem for obtaining the sensitivities, after one more iteration.* ■

Similarly, the problem that allows obtaining the reliability index associated with failure mode i is:

$$\beta_i = \underset{\mathbf{d}_i, \boldsymbol{\eta}_i}{\text{Minimize}} \sum_{k=1}^n z_k^2(\mathbf{d}_i, \boldsymbol{\eta}_i, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \boldsymbol{\kappa}) \quad (8.27)$$

subject to

$$\begin{aligned} z_k &= z_k(\mathbf{d}_i, \boldsymbol{\eta}_i, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \boldsymbol{\kappa}); \quad k = 1, 2, \dots, n \\ g_i(\mathbf{d}_i, \boldsymbol{\eta}_i) &= 1 \end{aligned} \quad (8.28)$$

where $\bar{\mathbf{d}}$ and $\tilde{\boldsymbol{\eta}}$ refer to the mean and characteristic values of \mathbf{d} and $\boldsymbol{\eta}$, respectively. This problem is equivalent to:

$$\underset{\mathbf{d}_i, \boldsymbol{\eta}_i, \mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\kappa}_{\text{aux}}}{\text{Minimize}} \quad \beta_i = \sum_{k=1}^n z_k^2(\mathbf{d}_i, \boldsymbol{\eta}_i, \mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\kappa}_{\text{aux}}) \quad (8.29)$$

subject to

$$z_k = z_k(\mathbf{d}_i, \boldsymbol{\eta}_i, \mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\kappa}_{\text{aux}}); \quad k = 1, 2, \dots, n \quad (8.30)$$

$$g_i(\mathbf{d}_i, \boldsymbol{\eta}_i) = 1 \quad (8.31)$$

$$(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\kappa}_{\text{aux}}) = (\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \boldsymbol{\kappa}) \quad (8.32)$$

where $\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\kappa}_{\text{aux}}$ are artificial variables.

Then, the dual variables associated with the constraints (8.32) give the sensitivities of the reliabilities β_i to the $\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \boldsymbol{\kappa}$ parameters.

Numerical example

Using this method for the wall problem in page ??, the results in Table 8.1 are obtained. This table shows the cost and the β sensitivities. Note that the reliabilities decrease as the dispersions (σ_h and σ_t) increase. Note also that an increase of μ_t and μ_h increases the cost.

Table 8.1: Sensitivities.

x	$\frac{\partial c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}})}{\partial x}$	$\frac{\partial \beta}{\partial x}$
∂a	--	4.893
∂b	--	1.223
$\partial \mu_h$	0.080	-0.062
$\partial \mu_t$	2.227	-1.784
$\partial \sigma_h$	--	-0.175
$\partial \sigma_t$	--	-1.909
$\partial \gamma$	0.000	0.000
∂F_o^0	0.000	--
$\partial \beta^0$	--	--

This is a valuable information for code makers and construction engineers, because they can know the cost or reliability increase due to a change in safety factors, failure probability bounds, or uncertainty parameters.

For illustration purposes, the implementation of this algorithm for the wall example is shown in Appendix ??, page ??.

Alternative 2. Using cutting hyperplanes

For the iterative method presented in Section 8.2.1, the cost sensitivities with respect to F_i^0 and $\boldsymbol{\eta}$, and the β -sensitivities with respect to $\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \boldsymbol{\kappa}$ can be determined, but the cost sensitivities with respect to β_i^0 and $\boldsymbol{\kappa}$ cannot be determined. The alternative method given below provides such sensitivities and in addition exhibit a better convergence. The method consists of transforming the following problem:

$$\text{Minimize}_{\bar{\mathbf{d}}} c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}) \quad (8.33)$$

subject to

$$g_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}) \geq F_i^0; \quad \forall i \in I \quad (8.34)$$

$$\beta_i^{(s)} + \boldsymbol{\lambda}_i^{(s)T} (\bar{\mathbf{d}} - \bar{\mathbf{d}}^{(s)}) \geq \beta_i^0; \quad \forall i \in I; \quad s = 1, 2, \dots, \nu - 1 \quad (8.35)$$

$$r_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}) \leq 0; \quad \forall j \in J \quad (8.36)$$

where $\forall i$:

$$\beta_i^{(\nu)} = \text{Minimum}_{\mathbf{d}_i, \boldsymbol{\eta}_i, \bar{\mathbf{d}}} \sum_{k=1}^n z_k^2(\mathbf{d}_i, \boldsymbol{\eta}_i, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \boldsymbol{\kappa}) \quad (8.37)$$

subject to

$$z_k = z_k(\mathbf{d}_i, \boldsymbol{\eta}_i, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \boldsymbol{\kappa}); \quad k = 1, 2, \dots, n \quad (8.38)$$

$$g_i(\mathbf{d}_i, \boldsymbol{\eta}_i) = 1 \quad (8.39)$$

$$\bar{\mathbf{d}} = \bar{\mathbf{d}}^{(\nu)} : \boldsymbol{\lambda}_i^{(\nu)} \quad (8.40)$$

is equivalent to the problem:

$$\begin{aligned} \text{Minimize } & c(\bar{\mathbf{d}}, \boldsymbol{\eta}) \\ & \bar{\mathbf{d}}, \boldsymbol{\eta}, \boldsymbol{\kappa}_{\text{aux}} \end{aligned} \quad (8.41)$$

subject to

$$g_i(\bar{\mathbf{d}}, \boldsymbol{\eta}) \geq F_i^0; \quad \forall i \in I \quad (8.42)$$

$$\beta_i^{(s)} + \lambda_i^{(s)T}(\bar{\mathbf{d}} - \bar{\mathbf{d}}^{(s)}) + \mu_i^{(s)T}(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}) + \mathbf{X}_i^{(s)T}(\boldsymbol{\kappa}_{\text{aux}} - \boldsymbol{\kappa}) \geq \beta_i^0; \quad \forall i \in I; s \leq \nu - 1 \quad (8.43)$$

$$r_j(\bar{\mathbf{d}}, \boldsymbol{\eta}) \leq 0; \quad \forall j \in J \quad (8.44)$$

$$\boldsymbol{\eta} = \tilde{\boldsymbol{\eta}} \quad (8.45)$$

$$\boldsymbol{\kappa}_{\text{aux}} = \boldsymbol{\kappa} \quad (8.46)$$

where $\forall i$:

$$\beta_i^{(\nu)} = \text{Minimum}_{\mathbf{d}_i, \boldsymbol{\eta}_i, \bar{\mathbf{d}}, \boldsymbol{\eta}, \boldsymbol{\kappa}_{\text{aux}}} \sum_{k=1}^n z_k^2(\mathbf{d}_i, \boldsymbol{\eta}_i, \bar{\mathbf{d}}, \boldsymbol{\eta}, \boldsymbol{\kappa}_{\text{aux}}) \quad (8.47)$$

subject to

$$z_k = z_k(\bar{\mathbf{d}}, \mathbf{d}_i, \boldsymbol{\eta}_i, \boldsymbol{\eta}, \boldsymbol{\kappa}_{\text{aux}}); \quad k = 1, 2, \dots, n \quad (8.48)$$

$$g_i(\mathbf{d}_i, \boldsymbol{\eta}_i) = 1 \quad (8.49)$$

$$\bar{\mathbf{d}} = \bar{\mathbf{d}}^{(\nu)} : \lambda_i^{(\nu)} \quad (8.50)$$

$$\boldsymbol{\eta} = \tilde{\boldsymbol{\eta}} : \mu_i^{(\nu)} \quad (8.51)$$

$$\boldsymbol{\kappa}_{\text{aux}} = \boldsymbol{\kappa} : \mathbf{X}_i^{(\nu)} \quad (8.52)$$

It should be noted that this statement of the problem allows determining the sensitivities of the objective function to parameters $F_i^0, \beta_i^0, \tilde{\boldsymbol{\eta}}$ and $\boldsymbol{\kappa}$ in a straightforward manner. Note also that this new formulation of the problem can be implemented easily in most conventional optimization software packages.

Using this method for the wall problem the results in Table 8.1 are obtained. This table shows the cost and the β sensitivities. Note that the cost increases and the reliabilities decrease as the dispersions (σ_h and σ_t) increase. Note also that an increase of μ_t and μ_h increases the cost. This is a valuable information for code makers and construction engineers, because they can know the cost or reliability increase due to a change in safety factors, failure probability bounds, or uncertainty parameters.

Application of this method to the wall example leads to the results in Table 8.2. The GAMS implementation of the wall problem is given in Appendix ??, page ??.

8.3 Sensitivity in Constrained Optimization

In this section we deal with local sensitivity from a general and formal perspective.

Consider the following NLPP:

$$\begin{aligned} \text{Minimize } & z = f(\mathbf{x}, \mathbf{a}) \\ & \mathbf{x}, \mathbf{a} \end{aligned} \quad (8.53)$$

Table 8.2: Sensitivities.

x	$\frac{\partial c(\mathbf{d}, \tilde{\boldsymbol{\eta}})}{\partial x}$	$\frac{\partial \beta}{\partial x}$
∂a	--	4.893
∂b	--	1.223
$\partial \mu_h$	0.080	-0.062
$\partial \mu_t$	2.277	-1.784
$\partial \sigma_h$	0.223	-0.175
$\partial \sigma_t$	2.437	-1.909
$\partial \gamma$	0.000	0.000
∂F_0^0	0.000	--
$\partial \beta^0$	1.277	--

subject to

$$\begin{aligned} \mathbf{h}(\mathbf{x}, \mathbf{a}) &= \mathbf{0} \\ \mathbf{g}(\mathbf{x}, \mathbf{a}) &\leq \mathbf{0} \end{aligned} \quad (8.54)$$

where $f : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$, $\mathbf{h} : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^\ell$, $\mathbf{g} : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^m$ with $\mathbf{h}(\mathbf{x}, \mathbf{a}) = (h_1(\mathbf{x}, \mathbf{a}), \dots, h_\ell(\mathbf{x}, \mathbf{a}))^T$ and $\mathbf{g}(\mathbf{x}, \mathbf{a}) = (g_1(\mathbf{x}, \mathbf{a}), \dots, g_m(\mathbf{x}, \mathbf{a}))^T$ are continuously twice differentiable functions over the feasible region $S(\mathbf{a}) = \{\mathbf{x} | \mathbf{h}(\mathbf{x}, \mathbf{a}) = \mathbf{0}, \mathbf{g}(\mathbf{x}, \mathbf{a}) \leq \mathbf{0}\}$.

The vector $\bar{\mathbf{x}} \in \mathbb{R}^n$ satisfies the Karush–Kuhn–Tucker conditions (KKTCS) for the NLPP (8.53)–(8.54) if there exists a pair of vectors $\boldsymbol{\mu} \in \mathbb{R}^m$ and $\boldsymbol{\lambda} \in \mathbb{R}^\ell$ such that (see Bazaraa, Sherali and Shetty [16] or Luenberger [115]):

$$\begin{aligned} \nabla_{\mathbf{x}} f(\bar{\mathbf{x}}, \mathbf{a}) + \sum_{k=1}^{\ell} \lambda_k \nabla_{\mathbf{x}} h_k(\bar{\mathbf{x}}, \mathbf{a}) \\ + \sum_{j=1}^m \mu_j \nabla_{\mathbf{x}} g_j(\bar{\mathbf{x}}, \mathbf{a}) &= \mathbf{0} \end{aligned} \quad (8.55)$$

$$h_k(\bar{\mathbf{x}}, \mathbf{a}) = 0; \quad k = 1, 2, \dots, \ell \quad (8.56)$$

$$g_j(\bar{\mathbf{x}}, \mathbf{a}) \leq 0; \quad j = 1, 2, \dots, m \quad (8.57)$$

$$\mu_j g_j(\bar{\mathbf{x}}, \mathbf{a}) = 0; \quad j = 1, \dots, m \quad (8.58)$$

$$\boldsymbol{\mu} \geq \mathbf{0} \quad (8.59)$$

As detailed in Chapter 6, page 183, the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are called the Karush–Kuhn–Tucker multipliers. Conditions (8.56)–(8.57) are called the primal feasibility conditions. Condition (8.58) is known as the complementary slackness condition, condition (8.59) requires the nonnegativity of the multipliers of the inequality constraints, and is referred to as the dual feasibility conditions.

In this section we aim at determining the sensitivity of the optimal solution $(\mathbf{x}^*, \mathbf{a}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*, z^*)$ of (8.55) - (8.59) to changes in the parameters, i.e., we perturb or modify $\mathbf{a}, \mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}$ and f in such a way that the KKTCS still hold. Thus, to obtain the sensitivity equations we

differentiate (8.53) and (8.55)-(8.59), as follows:

$$(\nabla_{\mathbf{x}}f(\mathbf{x}, \mathbf{a}))^T d\mathbf{x} + (\nabla_{\mathbf{a}}f(\mathbf{x}, \mathbf{a}))^T d\mathbf{a} - dz = 0 \quad (8.60)$$

$$\begin{aligned} & (\nabla_{\mathbf{x}}f(\bar{\mathbf{x}}, \mathbf{a}) + \sum_{k=1}^{\ell} \lambda_k \nabla_{\mathbf{x}\mathbf{x}}h_k(\bar{\mathbf{x}}, \mathbf{a}))d\mathbf{x} \\ & \quad + \sum_{j=1}^m \mu_j \nabla_{\mathbf{x}\mathbf{x}}g_j(\bar{\mathbf{x}}, \mathbf{a})d\mathbf{x} \\ & + (\nabla_{\mathbf{x}\mathbf{a}}f(\bar{\mathbf{x}}, \mathbf{a}) + \sum_{k=1}^{\ell} \lambda_k \nabla_{\mathbf{x}\mathbf{a}}h_k(\bar{\mathbf{x}}, \mathbf{a}))d\mathbf{a} \\ & \quad + \sum_{j=1}^m \mu_j \nabla_{\mathbf{x}\mathbf{a}}g_j(\bar{\mathbf{x}}, \mathbf{a}))d\mathbf{a} \\ & + \nabla_{\mathbf{x}}\mathbf{h}(\bar{\mathbf{x}}, \mathbf{a})d\boldsymbol{\lambda} + \nabla_{\mathbf{x}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a})d\boldsymbol{\mu} = \mathbf{0}_n \end{aligned} \quad (8.61)$$

$$(\nabla_{\mathbf{x}}\mathbf{h}(\bar{\mathbf{x}}, \mathbf{a}))^T d\mathbf{x} + (\nabla_{\mathbf{a}}\mathbf{h}(\bar{\mathbf{x}}, \mathbf{a}))^T d\mathbf{a} = \mathbf{0}_{\ell}; \quad (8.62)$$

$$(\nabla_{\mathbf{x}}g_j(\bar{\mathbf{x}}, \mathbf{a}))^T d\mathbf{x} + (\nabla_{\mathbf{a}}g_j(\bar{\mathbf{x}}, \mathbf{a}))^T d\mathbf{a} \leq 0; \text{ if } g_j(\bar{\mathbf{x}}, \mathbf{a}) = 0 \quad (8.63)$$

$$\begin{aligned} & \text{diag}[\boldsymbol{\mu}] (\nabla_{\mathbf{x}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a}))^T d\mathbf{x} \\ & + \text{diag}[\boldsymbol{\mu}] (\nabla_{\mathbf{a}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a}))^T d\mathbf{a} + \text{diag}[\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a})]d\boldsymbol{\mu} = \mathbf{0}_m; \end{aligned} \quad (8.64)$$

$$-d\mu_j \leq 0, \text{ if } \mu_j = 0 \quad (8.65)$$

where *diag* refers to a diagonal matrix, and all the matrices are evaluated at the optimal solution. For convenient reference, the dimensions of all these matrices are given in Table 8.3. In matrix form, the system (8.60)-(8.65) can be written as:

$$\mathbf{A} \begin{bmatrix} \frac{d\mathbf{x}}{d\mathbf{a}} \\ \frac{d\boldsymbol{\lambda}}{d\boldsymbol{\mu}} \\ \frac{dz}{dz} \end{bmatrix} = \mathbf{0} \quad (8.66)$$

$$\left[\begin{array}{c|c|c|c|c} \mathbf{D}(\nabla_{\mathbf{x}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a}))^T & \mathbf{D}(\nabla_{\mathbf{a}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a}))^T & \mathbf{0} & \mathbf{0} & 0 \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{I}_m^* & 0 \end{array} \right] \begin{bmatrix} \frac{d\mathbf{x}}{d\mathbf{a}} \\ \frac{d\boldsymbol{\lambda}}{d\boldsymbol{\mu}} \\ \frac{dz}{dz} \end{bmatrix} \leq \mathbf{0} \quad (8.67)$$

Table 8.3: Main matrices for sensitivity analysis and their respective dimensions.

$\nabla_{\mathbf{x}}f(\mathbf{x}, \mathbf{a})$	$\nabla_{\mathbf{a}}f(\mathbf{x}, \mathbf{a})$	$\nabla_{\mathbf{x}\mathbf{x}}f(\bar{\mathbf{x}}, \mathbf{a})$	$\nabla_{\mathbf{x}\mathbf{x}}h_k(\bar{\mathbf{x}}, \mathbf{a})$	$\nabla_{\mathbf{x}\mathbf{x}}g_j(\bar{\mathbf{x}}, \mathbf{a})$
$n \times 1$	$p \times 1$	$n \times n$	$n \times n$	$n \times n$
$\nabla_{\mathbf{x}\mathbf{a}}f(\bar{\mathbf{x}}, \mathbf{a})$	$\nabla_{\mathbf{x}\mathbf{a}}h_k(\bar{\mathbf{x}}, \mathbf{a})$	$\nabla_{\mathbf{x}\mathbf{a}}g_j(\bar{\mathbf{x}}, \mathbf{a})$	$\nabla_{\mathbf{x}}\mathbf{h}(\bar{\mathbf{x}}, \mathbf{a})$	$\nabla_{\mathbf{x}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a})$
$n \times p$	$n \times p$	$n \times p$	$n \times \ell$	$n \times m$
$\nabla_{\mathbf{a}}\mathbf{h}(\bar{\mathbf{x}}, \mathbf{a})$	$\nabla_{\mathbf{x}}g_j(\bar{\mathbf{x}}, \mathbf{a})$	$\nabla_{\mathbf{a}}g_j(\bar{\mathbf{x}}, \mathbf{a})$	$diag[\boldsymbol{\mu}]$	$\nabla_{\mathbf{x}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a})$
$p \times \ell$	$n \times 1$	$p \times 1$	$m \times m$	$n \times m$
$\nabla_{\mathbf{a}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a})$	$diag[\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a})]$	$d\mathbf{x}$	$d\mathbf{a}$	$d\boldsymbol{\lambda}$
$p \times m$	$m \times m$	$n \times 1$	$p \times 1$	$\ell \times 1$
$d\boldsymbol{\mu}$				
$m \times 1$				

where \mathbf{A} is equal to

$$\begin{array}{|c|c|c|c|c|}
 \hline
 \nabla_{\mathbf{x}}f(\mathbf{x}, \mathbf{a})^T & \nabla_{\mathbf{a}}f(\mathbf{x}, \mathbf{a})^T & \mathbf{0} & \mathbf{0} & -1 \\
 \hline
 (\nabla_{\mathbf{x}\mathbf{x}}f(\bar{\mathbf{x}}, \mathbf{a})) & (\nabla_{\mathbf{x}\mathbf{a}}f(\bar{\mathbf{x}}, \mathbf{a})) & & & \\
 + \sum_{k=1}^{\ell} \lambda_k \nabla_{\mathbf{x}\mathbf{x}}h_k(\bar{\mathbf{x}}, \mathbf{a}) & + \sum_{k=1}^{\ell} \lambda_k \nabla_{\mathbf{x}\mathbf{a}}h_k(\bar{\mathbf{x}}, \mathbf{a}) & \nabla_{\mathbf{x}}\mathbf{h}(\bar{\mathbf{x}}, \mathbf{a}) & \nabla_{\mathbf{x}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a}) & 0 \\
 + \sum_{j=1}^m \mu_j \nabla_{\mathbf{x}\mathbf{x}}g_j(\bar{\mathbf{x}}, \mathbf{a}) & + \sum_{j=1}^m \mu_j \nabla_{\mathbf{x}\mathbf{a}}g_j(\bar{\mathbf{x}}, \mathbf{a}) & & & \\
 \hline
 (\nabla_{\mathbf{x}}\mathbf{h}(\bar{\mathbf{x}}, \mathbf{a}))^T & (\nabla_{\mathbf{a}}\mathbf{h}(\bar{\mathbf{x}}, \mathbf{a}))^T & \mathbf{0} & \mathbf{0} & 0 \\
 \hline
 diag[\boldsymbol{\mu}](\nabla_{\mathbf{x}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a}))^T & diag[\boldsymbol{\mu}](\nabla_{\mathbf{a}}\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a}))^T & \mathbf{0} & diag[\mathbf{g}(\bar{\mathbf{x}}, \mathbf{a})] & 0 \\
 \hline
 \end{array}$$

and \mathbf{D} is a unit matrix after removing rows j such that $g_j(\bar{\mathbf{x}}, \mathbf{a}) < 0$, and $-I_m^*$ is a unit matrix after removing rows j such that $\mu_j \neq 0$.

Conditions (8.66)-(8.67) define the feasible set for the changes $d\mathbf{x}, d\mathbf{a}, d\boldsymbol{\lambda}, d\boldsymbol{\mu}$ and dz . Since these conditions constitute an homogeneous linear system of equalities and inequalities in $d\mathbf{x}, d\mathbf{a}, d\boldsymbol{\lambda}, d\boldsymbol{\mu}$ and dz , its general solution is a cone (see Castillo et al. [29, 30]):

$$\begin{pmatrix} d\mathbf{x} \\ d\mathbf{a} \\ d\boldsymbol{\lambda} \\ d\boldsymbol{\mu} \\ dz \end{pmatrix} = \sum_{i=1}^p \rho_i \mathbf{v}_i + \sum_{j=1}^q \pi_j \mathbf{w}_j \quad (8.68)$$

where $\rho_i \in \mathbb{R}; i = 1, 2, \dots, p$; and $\pi_j \in \mathbb{R}^+; j = 1, 2, \dots, q$, and \mathbf{v}_i and \mathbf{w}_j are vectors that generate (a) the linear space and (b) the acute cone parts of the cone, respectively.

Expression (8.68) allows determining:

1. The set of feasible perturbations $\{d\mathbf{x}, d\mathbf{a}, d\boldsymbol{\lambda}, d\boldsymbol{\mu}, dz\}$ for moving from a KKT solution to another KKT solution.

2. Whether or not certain partial derivatives exist. Partial derivatives require modifying one variable while the remaining variables are kept constant. This requirement needs to be possible and compatible with optimality. Otherwise, the partial derivative does not exist. Note that Expression (8.68) allows determining whether or not this is possible. This is illustrated in Example 8.2 below.
3. Directions where partial derivatives do not exist. Similarly, Expression (8.68) defines all possible feasible directions. So, if a direction is not feasible, the corresponding directional derivative does not exist.

Remark 8.2 *To perform a local sensitivity analysis, the objective function (8.53) and the nonlinear constraint in (8.54) can be replaced by the corresponding quadratic approximations that share tangent hyperplanes and Hessian matrices at the solution point.* ■

8.4 Illustrative Examples

The remaining of the chapter provides relevant examples that illustrate results stated in this section.

Illustrative Example 8.2 (KKT conditions). Consider the following optimization problem:

$$\text{Minimize}_{x_1, x_2} (x_1 - 1)^2 + (x_2 - 1)^2 \quad (8.69)$$

subject to

$$\begin{aligned} x_1 + x_2 &\leq 1 \\ x_1 &\geq 0 \\ x_2 &\geq 0 \end{aligned} \quad (8.70)$$

The Lagrangian function is:

$$\mathcal{L}(\mathbf{x}; \boldsymbol{\mu}) = (x_1 - 1)^2 + (x_2 - 1)^2 + \mu_1(x_1 + x_2 - 1) + \mu_2(-x_1) + \mu_3(-x_2)$$

and the Karush-Kuhn-Tucker conditions become:

$$\frac{\partial \mathcal{L}(\mathbf{x}; \boldsymbol{\mu})}{\partial x_1} = 2(x_1 - 1) + \mu_1 - \mu_2 = 0 \Rightarrow x_1 = \frac{\mu_2 - \mu_1}{2} + 1 \quad (8.71)$$

$$\frac{\partial \mathcal{L}(\mathbf{x}; \boldsymbol{\mu})}{\partial x_2} = 2(x_2 - 1) + \mu_1 - \mu_3 = 0 \Rightarrow x_2 = \frac{\mu_3 - \mu_1}{2} + 1 \quad (8.72)$$

$$x_1 + x_2 \leq 1 \quad (8.73)$$

$$-x_1 \leq 0 \quad (8.74)$$

$$-x_2 \leq 0 \quad (8.75)$$

$$\mu_1(x_1 + x_2 - 1) = 0 \quad (8.76)$$

$$\mu_2(-x_1) = 0 \quad (8.77)$$

$$\mu_3(-x_2) = 0 \quad (8.78)$$

$$\mu_1, \mu_2, \mu_3 \geq 0 \quad (8.79)$$

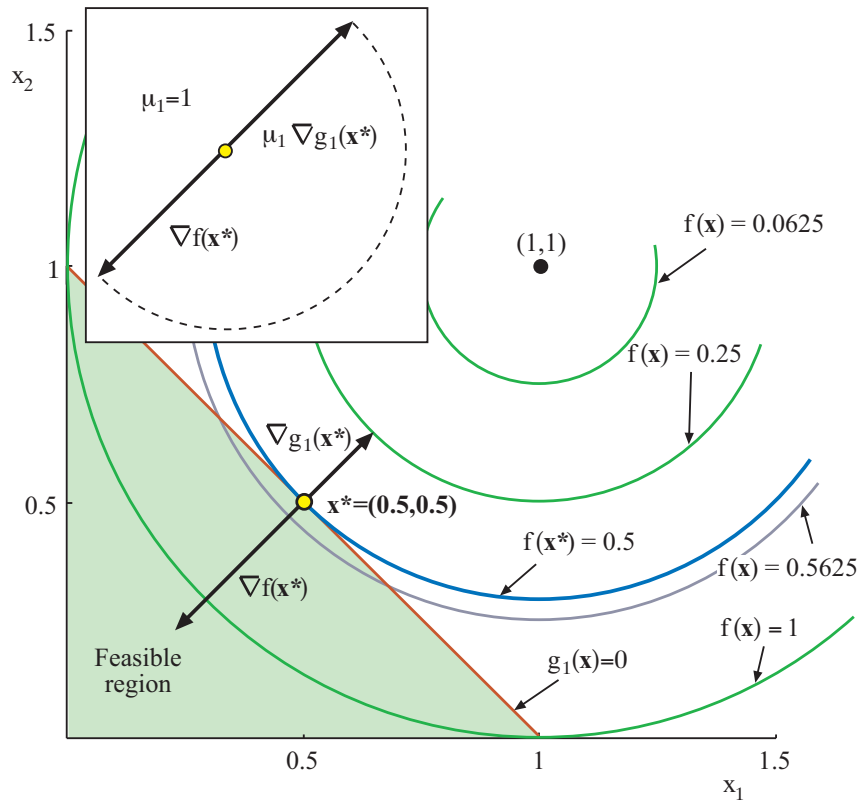


Figure 8.2: Illustration of the minimization problem in Example 8.2.

which have one solution:

$$x_1 = 1/2, \quad x_2 = 1/2, \quad \mu_1 = 1, \quad \mu_2 = 0, \quad \mu_3 = 0$$

This is illustrated in Figure 8.2, where the optimal point is shown together with the gradient of the objective function and the active constraint. Note that they have opposite directions.

To obtain the dual function $\theta(\mu_1, \mu_2, \mu_3)$ one needs to obtain:

$$\text{Infimum}_{x_1, x_2} \mathcal{L}(x_1, x_2; \mu_1, \mu_2, \mu_3)$$

which is attained at the point (see (8.71) and (8.72)):

$$x_1 = \frac{\mu_2 - \mu_1}{2} + 1; \quad x_2 = \frac{\mu_3 - \mu_1}{2} + 1 \quad (8.80)$$

This solution is a minimum because the Hessian of $\mathcal{L}(\mathbf{x}; \boldsymbol{\mu})$ at this point becomes

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

which is positive definite.

Then, the dual function is:

$$\phi(\boldsymbol{\mu}) = (-2\mu_1^2 - 4\mu_2 - \mu_2^2 - \mu_3(4 + \mu_3) + 2\mu_1(2 + \mu_2 + \mu_3)) / 4$$

and the dual problem becomes:

$$\begin{aligned} &\text{Maximize} && (-2\mu_1^2 - 4\mu_2 - \mu_2^2 - \mu_3(4 + \mu_3) + 2\mu_1(2 + \mu_2 + \mu_3)) / 4 \\ &&& \mu_1, \mu_2, \mu_3 \end{aligned}$$

subject to

$$\mu_1, \mu_2, \mu_3 \geq 0$$

which solution is:

$$\mu_1 = 1, \mu_2 = 0, \mu_3 = 0, z = 1/2$$

Consider the following parametric optimization problem which is a generalization of problem (8.69)-(8.70):

$$\begin{aligned} &\text{Minimize} && (x_1 - 1)^2 + (x_2 - 1)^2 \\ &&& x_1, x_2 \end{aligned} \tag{8.81}$$

subject to

$$\begin{aligned} x_1 + x_2 &\leq a_1 \\ x_1 &\geq a_2 \\ x_2 &\geq a_3 \end{aligned} \tag{8.82}$$

which for $a_1 = 1, a_2 = 0, a_3 = 0$ leads to the following optimal solution:

$$x_1 = 1/2, x_2 = 1/2, \mu_1 = 1, \mu_2 = 0, \mu_3 = 0, z = 1/2$$

To analyze the existence of partial derivatives we solve the system of inequalities (8.66)-(8.67), which in this case becomes

$$\left(\begin{array}{ccc|ccc|ccc} -1 & -1 & | & 0 & 0 & 0 & | & 0 & 0 & 0 & | & -1 \\ \hline 2 & 0 & | & 0 & 0 & 0 & | & 1 & -1 & 0 & | & 0 \\ 0 & 2 & | & 0 & 0 & 0 & | & 1 & 0 & -1 & | & 0 \\ \hline 1 & 1 & | & -1 & 0 & 0 & | & 0 & 0 & 0 & | & 0 \\ \hline 0 & 0 & | & 0 & 0 & 0 & | & 0 & -\frac{1}{2} & 0 & | & 0 \\ \hline 0 & 0 & | & 0 & 0 & 0 & | & 0 & 0 & -\frac{1}{2} & | & 0 \end{array} \right) \begin{pmatrix} dx_1 \\ dx_2 \\ \hline da_1 \\ da_2 \\ da_3 \\ \hline d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ \hline dz \end{pmatrix} = \mathbf{0} \tag{8.83}$$

and

$$\left(\begin{array}{ccc|ccc|ccc} 1 & 1 & | & -1 & 0 & 0 & | & 0 & 0 & 0 & | & 0 \\ \hline 0 & 0 & | & 0 & 0 & 0 & | & 0 & -1 & 0 & | & 0 \\ 0 & 0 & | & 0 & 0 & 0 & | & 0 & 0 & -1 & | & 0 \end{array} \right) \begin{pmatrix} dx_1 \\ dx_2 \\ \hline da_1 \\ da_2 \\ da_3 \\ \hline d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ \hline dz \end{pmatrix} \leq \mathbf{0} \tag{8.84}$$

leading to the cone

$$\begin{pmatrix} dx_1 \\ dx_2 \\ \text{---} \\ da_1 \\ da_2 \\ da_3 \\ \text{---} \\ d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ \text{---} \\ dz \end{pmatrix} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ \text{---} & \text{---} & \text{---} \\ 0 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ \text{---} & \text{---} & \text{---} \\ 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \text{---} & \text{---} & \text{---} \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} \quad (8.85)$$

which in fact is a linear space. How to derive this cone is explained in detail in [29].

For discussing the existence of directional and partial derivatives with respect to a_1 to a_3 one needs to know which directional vectors belong to the above linear space. This requires obtaining the orthogonal set (dual cone):

$$\begin{pmatrix} da_1 \\ da_2 \\ da_3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} \quad (8.86)$$

which is the vector $\mathbf{0}$. This proves that all derivatives with respect to a_1, a_2 and a_3 exist.

To obtain them we write

$$\begin{pmatrix} dx_1 \\ dx_2 \\ \text{---} \\ da_1 \\ da_2 \\ da_3 \\ \text{---} \\ d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ \text{---} \\ dz \end{pmatrix} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ \text{---} & \text{---} & \text{---} \\ 0 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ \text{---} & \text{---} & \text{---} \\ 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \text{---} & \text{---} & \text{---} \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ \text{---} & \text{---} & \text{---} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \text{---} & \text{---} & \text{---} \\ -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \text{---} & \text{---} & \text{---} \\ -1 & 0 & 0 \end{pmatrix}$$

which proves that

$$\frac{\partial z^*}{\partial a_1} = -1, \quad \frac{\partial z^*}{\partial a_2} = 0, \quad \frac{\partial z^*}{\partial a_3} = 0$$

■

Illustrative Example 8.3 (Sensitivities). Consider the following parametric optimization problem:

$$\begin{aligned} \text{Minimize} \quad & -x_1^2 - x_2^2 \\ & x_1, x_2 \end{aligned} \quad (8.87)$$

subject to

$$\begin{aligned} x_1 + x_2 &\leq a_1 \\ x_1 &\geq a_2 \\ x_2 &\geq a_3 \end{aligned} \tag{8.88}$$

which for $a_1 = 1, a_2 = 0, a_3 = 0$ leads to the following optimal solution (see Example 6.1 in page 186):

$$x_1 = 0, \quad x_2 = 1, \quad \mu_1 = 2, \quad \mu_2 = 2, \quad \mu_3 = 0, \quad z = -1$$

To analyze the existence of partial derivatives we solve the system of inequalities (8.66)-(8.67), which in this case becomes

$$\left(\begin{array}{ccc|ccc|c} 0 & -2 & & 0 & 0 & 0 & -1 \\ - & - & + & - & - & - & - \\ -2 & 0 & & 0 & 0 & 0 & 1 \\ 0 & -2 & & 0 & 0 & 0 & 1 \\ - & - & + & - & - & - & - \\ 2 & 2 & & -2 & 0 & 0 & 0 \\ - & - & + & - & - & - & - \\ -2 & 0 & & 0 & 2 & 0 & 0 \\ - & - & + & - & - & - & - \\ 0 & 0 & & 0 & 0 & 0 & -1 \end{array} \right) \begin{pmatrix} dx_1 \\ dx_2 \\ da_1 \\ da_2 \\ da_3 \\ d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ dz \end{pmatrix} = \mathbf{0} \tag{8.89}$$

and

$$\left(\begin{array}{ccc|ccc|c} 1 & 1 & & -1 & 0 & 0 & 0 \\ -1 & 0 & & 0 & 1 & 0 & 0 \\ - & - & + & - & - & - & - \\ 0 & 0 & & 0 & 0 & 0 & -1 \end{array} \right) \begin{pmatrix} dx_1 \\ dx_2 \\ da_1 \\ da_2 \\ da_3 \\ d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ dz \end{pmatrix} \leq \mathbf{0} \tag{8.90}$$

The general solution of (8.89) and (8.90) is the linear space:

$$\begin{pmatrix} dx_1 \\ dx_2 \\ da_1 \\ da_2 \\ da_3 \\ d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ dz \end{pmatrix} = \begin{pmatrix} -1 & 0 & -1 \\ 0 & 0 & -1 \\ -1 & 0 & -2 \\ -1 & 0 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} \tag{8.91}$$

For discussing the existence of directional and partial derivatives with respect to a_1 to a_3 we must obtain the orthogonal set to the linear space:

$$\begin{pmatrix} da_1 \\ da_2 \\ da_3 \end{pmatrix} = \begin{pmatrix} -1 & 0 & -2 \\ -1 & 0 & -1 \\ 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} \quad (8.92)$$

which is the vector $\mathbf{0}$. This proves that all derivatives with respect to a_1, a_2 and a_3 exist.

To obtain them we write

$$\begin{pmatrix} dx_1 \\ dx_2 \\ \text{---} \\ da_1 \\ da_2 \\ da_3 \\ \text{---} \\ d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ \text{---} \\ dz \end{pmatrix} = \begin{pmatrix} -1 & 0 & -1 \\ 0 & 0 & -1 \\ \text{---} & \text{---} & \text{---} \\ -1 & 0 & -2 \\ -1 & 0 & -1 \\ 0 & 2 & 0 \\ \text{---} & \text{---} & \text{---} \\ 0 & 0 & -2 \\ 2 & 0 & 0 \\ 0 & 0 & 0 \\ \text{---} & \text{---} & \text{---} \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} -1 & 0 & -2 \\ -1 & 0 & -1 \\ 0 & 2 & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ \text{---} & \text{---} & \text{---} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \text{---} & \text{---} & \text{---} \\ 2 & -2 & 0 \\ 2 & -4 & 0 \\ 0 & 0 & 0 \\ \text{---} & \text{---} & \text{---} \\ -2 & 2 & 0 \end{pmatrix}$$

which proves that

$$\frac{\partial z^*}{\partial a_1} = -2, \quad \frac{\partial z^*}{\partial a_2} = 2, \quad \frac{\partial z^*}{\partial a_3} = 0$$

■

Illustrative Example 8.4 (Sensitivities). Consider the following parametric optimization problem:

$$\begin{array}{ll} \text{Minimize} & -2a_1x - y \\ & x, y \end{array} \quad (8.93)$$

subject to

$$\begin{array}{rcl} x + y & = & a_2 \\ x & \leq & a_3 \\ y & \leq & a_4 \\ x + 4y/3 & \leq & 4 \\ x & \geq & 0 \\ y & \geq & 0 \end{array} \quad (8.94)$$

which for $a_1 = 1, a_2 = 2, a_3 = 1, a_4 = 1$ leads to the following optimal solution:

$$x = 1, \quad y = 1, \quad \lambda = 1, \quad \mu_1 = 1, \quad \mu_2 = 0, \quad \mu_3 = 0, \quad \mu_4 = 0, \quad \mu_5 = 0, \quad z = -3$$

shown in Figure 8.3.

To analyze the existence of partial derivatives we solve the system of inequalities (8.66)-(8.67), which in this case becomes

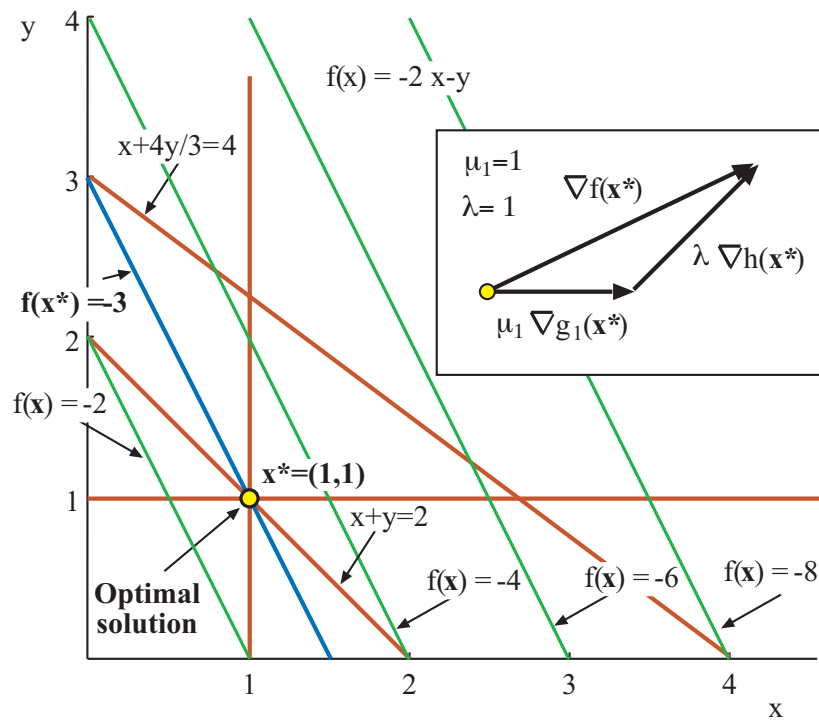


Figure 8.3: Illustration of the minimization problem in Example 8.4.

$$\left(\begin{array}{cccc|cccc|cccc|c}
 2 & 1 & & & 2 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 & 0 & -1 \\
 \hline
 0 & 0 & & & 2 & 0 & 0 & 0 & 1 & & 1 & 0 & 1 & -1 & 0 & 0 \\
 0 & 0 & & & 0 & 0 & 0 & 0 & 1 & & 0 & 1 & \frac{4}{3} & 0 & -1 & 0 \\
 \hline
 1 & 1 & & & 0 & -1 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 & 0 & 0 \\
 \hline
 1 & 0 & & & 0 & 0 & -1 & 0 & 0 & & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & & & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & & & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & -\frac{5}{3} & 0 & 0 & 0 \\
 0 & 0 & & & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & -1 & 0 & 0 \\
 0 & 0 & & & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 & -1 & 0
 \end{array} \right) \begin{pmatrix} dx \\ dy \\ \hline da_1 \\ da_2 \\ da_3 \\ da_4 \\ \hline d\lambda \\ \hline d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ d\mu_4 \\ d\mu_5 \\ \hline dz \end{pmatrix} = \mathbf{0} \quad (8.95)$$

and

$$\left(\begin{array}{cccc|cccc|cccc|c}
 1 & 0 & & & 0 & 0 & -1 & 0 & 0 & & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & & & 0 & 0 & 0 & -1 & 0 & & 0 & 0 & 0 & 0 & 0 & 0 \\
 \hline
 0 & 0 & & & 0 & 0 & 0 & 0 & 0 & & 0 & -1 & 0 & 0 & 0 & 0 \\
 0 & 0 & & & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & -1 & 0 & 0 & 0 \\
 0 & 0 & & & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & -1 & 0 & 0 \\
 0 & 0 & & & 0 & 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 & -1 & 0
 \end{array} \right) \begin{pmatrix} dx \\ dy \\ \hline da_1 \\ da_2 \\ da_3 \\ da_4 \\ \hline d\lambda \\ \hline d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ d\mu_4 \\ d\mu_5 \\ \hline dz \end{pmatrix} \leq \mathbf{0} \quad (8.96)$$

leading to the cone

$$\begin{pmatrix} dx \\ dy \\ \text{---} \\ da_1 \\ da_2 \\ da_3 \\ da_4 \\ \text{---} \\ d\lambda \\ \text{---} \\ d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ d\mu_4 \\ d\mu_5 \\ \text{---} \\ dz \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ \text{---} & \text{---} & \text{---} \\ -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \\ \text{---} & \text{---} & \text{---} \\ 0 & 0 & 0 \\ \text{---} & \text{---} & \text{---} \\ -1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \text{---} & \text{---} & \text{---} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \text{---} & \text{---} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ \text{---} & \text{---} \\ -1 & 0 \\ \text{---} & \text{---} \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \text{---} & \text{---} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} \quad (8.97)$$

For discussing the existence of directional and partial derivatives with respect to a_1 to a_4 we must obtain the dual of the cone:

$$\begin{pmatrix} da_1 \\ da_2 \\ da_3 \\ da_4 \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} \quad (8.98)$$

which is the cone:

$$\begin{pmatrix} 0 \\ 1 \\ -1 \\ -1 \end{pmatrix}_\pi \quad (8.99)$$

For a directional derivative to exist, the $(da_1, da_2, da_3, da_4)^T$ corresponding variation must belong to the cone (8.98), i.e., its dot product by the cone generator in (8.99) must be nonpositive. Similarly, for a partial derivative to exist the corresponding left and right directional derivatives must exist. This implies that the only existing partial derivative is the one with respect to a_1 (the one associated with the zero row in (8.99)). In addition, there exist the right derivatives of a_3 and a_4 (the negative rows in (8.99)), and the left derivative with respect to a_2 (the positive row in (8.99)).

The geometric interpretation of the non existence of partial derivatives with respect to a_2, a_3 and a_4 is clear, since an increment of a_2 or a decrement of a_3 or a_4 leads to infeasibility (see Figure 8.4).

To obtain the numerical values of the partial derivatives we need to find the values of ρ and π in (8.98) that lead to values of $(da_1, da_2, da_3, da_4)^T$ equal to the column vectors in

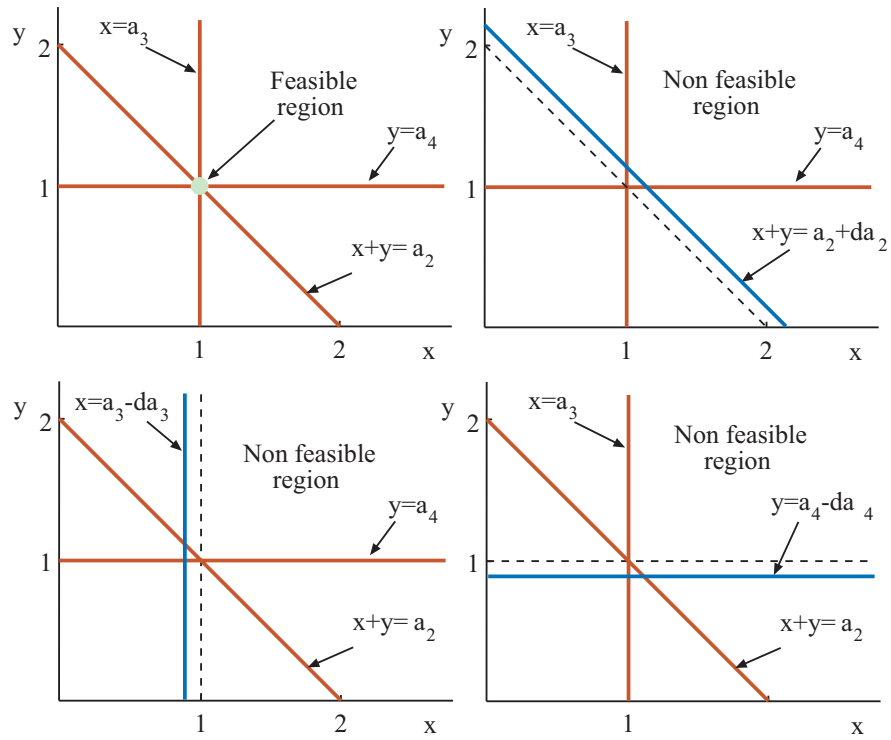


Figure 8.4: Geometric interpretation of the non existence of the partial derivatives with respect to a_2 , a_3 and a_4 in Example 8.4.

the matrix that were already known to exist:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{8.100}$$

i.e., the following matrix equation must be solved:

$$\begin{pmatrix} -1/2 & -1/2 & -1/2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix} X_\rho + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} Y_\pi = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad Y_\pi \geq \mathbf{0}$$

where X_ρ and Y_π are matrices of dimensions 3×4 and 2×4 , respectively. The solution is:

$$X_\rho = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -2 & 1 & -1 & 0 \end{pmatrix}, \quad Y_\pi = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

and replacing X_ρ and Y_π into Equation (8.97) one gets:

$$\begin{pmatrix} dx \\ dy \\ \text{---} \\ da_1 \\ da_2 \\ da_3 \\ da_4 \\ \text{---} \\ d\lambda \\ \text{---} \\ d\mu_1 \\ d\mu_2 \\ d\mu_3 \\ d\mu_4 \\ d\mu_5 \\ \text{---} \\ dz \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 \\ \text{---} & \text{---} & \text{---} & \text{---} \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \text{---} & \text{---} & \text{---} & \text{---} \\ 0 & 0 & 0 & 0 \\ \text{---} & \text{---} & \text{---} & \text{---} \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \text{---} & \text{---} & \text{---} & \text{---} \\ -2 & 1 & -1 & 0 \end{pmatrix} \tag{8.101}$$

which allows obtaining the partial and directional derivatives sought after, i.e.,:

$$\frac{\partial z^*}{\partial a_1} = -2, \quad \frac{\partial z^{*-}}{\partial a_2} = -1, \quad \frac{\partial z^{*+}}{\partial a_3} = -1, \quad \frac{\partial z^{*+}}{\partial a_4} = 0$$

where the asterisks refer to optimal values. In addition, the same directional and partial derivatives of $x, y, \lambda, \mu_1, \mu_2, \mu_3, \mu_4$ and μ_5 , can be obtained from the corresponding rows in (8.101).



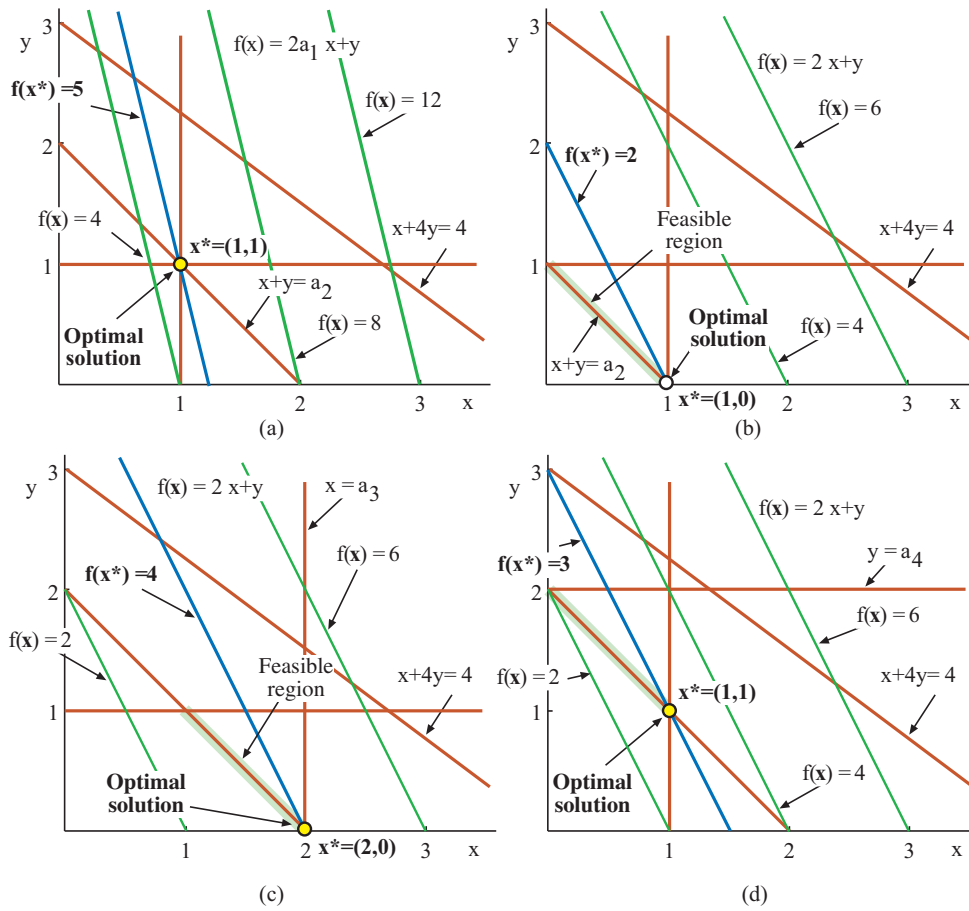


Figure 8.5: .

Illustrative Example 8.5 (Example 8.1 revisited). In this example the methodology explained in Section 8.3 is illustrated by its application to the non linear problem of Example 8.1. In this case, Equations (8.66)-(8.67) become:

$$\left(\begin{array}{cccc|cccc|cc|c} 6 & 4 & 9 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ - & - & + & - & - & - & - & - & + & - & + & - \\ 4 & 0 & 6 & 0 & 0 & -4 & 0 & 0 & 0 & 1 & -2 & -2 & 0 \\ 0 & 4 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 1 & -3 & 0 & 0 \\ - & - & + & - & - & - & - & - & + & - & + & - & - \\ 1 & 1 & 0 & 0 & 0 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 \\ - & - & + & - & - & - & - & - & + & - & + & - & - \\ -2 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{array} \right) \begin{pmatrix} dx \\ dy \\ - \\ da_1 \\ da_2 \\ da_3 \\ da_4 \\ da_5 \\ da_6 \\ da_7 \\ - \\ d\lambda \\ - \\ d\mu_2 \\ d\mu_2 \\ - \\ dz \end{pmatrix} = \mathbf{0}$$

and

$$\left(\begin{array}{cccc|cccc|cc|c} -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ - & - & + & - & - & - & - & - & + & - & + & - & - \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{array} \right) \begin{pmatrix} dx \\ dy \\ - \\ da_1 \\ da_2 \\ da_3 \\ da_4 \\ da_5 \\ da_6 \\ da_7 \\ - \\ d\lambda \\ - \\ d\mu_2 \\ d\mu_2 \\ - \\ dz \end{pmatrix} \leq \mathbf{0}$$

The solution of this system of inequalities is the cone:

$$\begin{pmatrix} dx \\ dy \\ \hline da_1 \\ da_2 \\ da_3 \\ da_4 \\ da_5 \\ da_6 \\ da_7 \\ \hline d\lambda \\ \hline d\mu_2 \\ d\mu_2 \\ \hline dz \end{pmatrix} = \begin{pmatrix} 0 & 0 & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ - & - & - & - & - & - & - \\ 0 & 0 & \frac{1}{3} & \frac{1}{9} & 0 & 0 & \frac{1}{9} \\ 0 & 0 & 0 & -\frac{1}{4} & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{5}{12} & 0 & -\frac{1}{2} & \frac{1}{6} \\ 0 & 0 & -\frac{1}{2} & \frac{5}{4} & 1 & -\frac{3}{2} & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ - & - & - & - & - & - & - \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ - & - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ - & - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \rho_5 \\ \rho_6 \\ \rho_7 \end{pmatrix} \quad (8.102)$$

Choosing a collection of ρ -values in such a way that the \mathbf{a} values become the unit matrix, i.e., multiplying by the inverse matrix of the submatrix in (8.102) associated with the \mathbf{a} -values, one gets:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & -3 & 1 & 0 & \frac{1}{2} \\ - & - & - & - & - & - & - \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ - & - & - & - & - & - & - \\ 0 & -4 & 0 & 12 & -4 & 0 & -2 \\ - & - & - & - & - & - & - \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & -2 & 0 & 4 & -2 & 0 & -2 \\ - & - & - & - & - & - & - \\ 9 & 4 & 0 & -12 & 4 & 0 & -1 \end{pmatrix}$$

Note that the last row gives the sensitivities of z with respect to the \mathbf{a} -values, that are

coincident with those obtained in Example 8.1. Additionally, rows 1, 2, 10, 11 and 12 give the sensitivities of x, y, λ, μ_1 and μ_2 with respect to the same \mathbf{a} -values. ■

Illustrative Example 8.6 (A degenerated case). Consider again the optimization problem (8.8)-(8.9) but with $a_3 = 6$ instead of $a_3 = 4$. Then, the KKT conditions lead to the optimal solution:

$$x = 3, y = 2, \lambda = -4, \mu_1 = 0, \mu_2 = 1, z = 13$$

and the resulting sensitivities using the method described in Section 8.1, page 225 are:

$$\frac{\partial z^*}{\partial a_1} = 9, \frac{\partial z^*}{\partial a_2} = 4, \frac{\partial z^*}{\partial a_3} = 0, \frac{\partial z^*}{\partial a_4} = -12, \frac{\partial z^*}{\partial a_5} = 4, \frac{\partial z^*}{\partial a_6} = 0, \frac{\partial z^*}{\partial a_7} = -1$$

If the method presented in Section 8.3, page 231 is used, Equations (8.66)-(8.67) become:

$$\left(\begin{array}{cccccccc|cccc|c} 6 & 4 & | & 9 & 4 & 0 & 0 & 0 & 0 & | & 0 & 0 & | & -1 \\ - & + & - & - & - & - & - & - & + & - & + & - & - & + & - \\ 4 & 0 & | & 6 & 0 & 0 & -4 & 0 & 0 & | & 0 & 1 & | & -2 & -2 & | & 0 \\ 0 & 4 & | & 0 & 4 & 0 & 0 & 0 & 0 & | & 0 & 1 & | & -3 & 0 & | & 0 \\ - & - & + & - & - & - & - & - & - & + & - & + & - & - & + & - \\ 1 & 1 & | & 0 & 0 & 0 & 3 & -1 & 0 & | & 0 & 0 & | & 0 & 0 & | & 0 \\ 0 & 0 & | & 0 & 0 & 0 & 0 & 0 & 0 & | & 0 & 0 & | & 0 & 0 & | & 0 \\ - & - & + & - & - & - & - & - & - & + & - & + & - & - & + & - \\ -2 & 0 & | & 0 & 0 & 0 & 0 & 0 & 0 & -1 & | & 0 & | & 0 & 0 & | & 0 \end{array} \right) \begin{pmatrix} dx \\ dy \\ - \\ da_1 \\ da_2 \\ da_3 \\ da_4 \\ da_5 \\ da_6 \\ da_7 \\ - \\ d\lambda \\ - \\ d\mu_2 \\ d\mu_2 \\ - \\ dz \end{pmatrix} = \mathbf{0}$$

and

$$\left(\begin{array}{cccccccc|cccc|c} -2 & -3 & | & 0 & 0 & 1 & 0 & 0 & 0 & | & 0 & | & 0 & 0 & | & 0 \\ -2 & 0 & | & 0 & 0 & 0 & 0 & 0 & 0 & -1 & | & 0 & | & 0 & 0 & | & 0 \\ - & - & + & - & - & - & - & - & - & + & - & + & - & - & + & - \\ 0 & 0 & | & 0 & 0 & 0 & 0 & 0 & 0 & 0 & | & 0 & | & -1 & 0 & | & 0 \end{array} \right) \begin{pmatrix} dx \\ dy \\ - \\ da_1 \\ da_2 \\ da_3 \\ da_4 \\ da_5 \\ da_6 \\ da_7 \\ - \\ d\lambda \\ - \\ d\mu_2 \\ d\mu_2 \\ - \\ dz \end{pmatrix} \leq \mathbf{0}$$

The solution of this system of inequalities is the cone:

$$\begin{pmatrix} dx \\ dy \\ \hline da_1 \\ da_2 \\ da_3 \\ da_4 \\ da_5 \\ da_6 \\ da_7 \\ \hline d\lambda \\ \hline d\mu_2 \\ d\mu_2 \\ \hline dz \end{pmatrix} = \begin{pmatrix} 0 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & -5/4 & 1 & 3/2 & -1/2 \\ \hline 0 & 1/3 & 1/9 & 0 & 0 & 1/9 \\ 0 & -1/2 & 1 & -1 & -3/2 & 1/2 \\ 0 & 1/2 & -15/4 & 3 & 9/2 & -3/2 \\ 0 & 0 & 5/12 & 0 & -1/2 & 1/6 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \rho_5 \\ \rho_6 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 3 \\ \hline 0 & -1/3 \\ 0 & -9/4 \\ -1 & 9 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \hline 0 & 0 \\ \hline 0 & 1 \\ 0 & 0 \\ \hline 0 & 0 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} \quad (8.103)$$

To check the existence of directional derivatives, we need the dual cone of the cone:

$$\begin{pmatrix} da_1 \\ da_2 \\ da_3 \\ da_4 \\ da_5 \\ da_6 \\ da_7 \end{pmatrix} = \begin{pmatrix} 0 & 1/3 & 1/9 & 0 & 0 & 1/9 \\ 0 & -1/2 & 1 & -1 & -3/2 & 1/2 \\ 0 & 1/2 & -15/4 & 3 & 9/2 & -3/2 \\ 0 & 0 & 5/12 & 0 & -1/2 & 1/6 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \rho_5 \\ \rho_6 \end{pmatrix} + \begin{pmatrix} 0 & -1/3 \\ 0 & -9/4 \\ -1 & 9 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} \quad (8.104)$$

that is the cone

$$\begin{pmatrix} 0 \\ 0 \\ 2 \\ 18 \\ -6 \\ 0 \\ -1 \end{pmatrix} \pi \quad (8.105)$$

For a directional derivative to exist, the corresponding variation must belong to the cone (8.104), i.e., its dot product with the generator cone in (8.105) must be nonpositive. Similarly, for a partial derivative to exist the corresponding left and right directional derivatives must exist. This implies that the only existing partial derivatives are those with respect to a_1, a_2 and a_6 (those associated with the zero rows in (8.105)). In addition, there exist the right derivatives of a_5 and a_7 (the negative rows in (8.105)), and the left derivatives of a_3 and a_4 (the positive rows in (8.105)).

The geometric interpretation of the non existence of partial derivatives with respect to a_3, a_4, a_5 and a_7 is clear, since an increment of a_3 or a_4 or a decrement of a_5 or a_7 leads to infeasibility.

To obtain the partial derivatives we need to find the values of ρ and π in (8.103) that lead to values of $(da_1, da_2, \dots, da_7)^T$ equal to the column vectors in the matrix, that were

already known to exist:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (8.106)$$

i.e., to solve the matrix equation:

$$\begin{pmatrix} 0 & 1/3 & 1/9 & 0 & 0 & 1/9 \\ 0 & -1/2 & 1 & -3/2 & 1/2 \\ 0 & 1/2 & -15/4 & 3 & 9/2 & -3/2 \\ 0 & 0 & 5/12 & 0 & -1/2 & 1/6 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} X_\rho + \begin{pmatrix} 0 & -1/3 \\ 0 & -9/4 \\ -1 & 9 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} Y_\pi = \begin{pmatrix} 10 & 0 & 0000 \\ 01 & 0 & 0000 \\ 00 & -1 & 0000 \\ 00 & 0 & -1000 \\ 00 & 0 & 0100 \\ 00 & 0 & 0010 \\ 00 & 0 & 0001 \end{pmatrix}$$

where X_ρ and $Y_\pi \geq \mathbf{0}$ are matrices of dimensions 6×7 and 2×7 , respectively. The solution is

$$X_\rho = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & -4 & 0 & -12 & -4 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 3 & -2 & 0 & -4 & -2 & 0 & -2 \\ 9 & 4 & 0 & 12 & 4 & 0 & -1 \end{pmatrix}, \quad Y_\pi = \begin{pmatrix} 0 & 0 & 1 & 9 & 3 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Replacing X_ρ and Y_π into Equation (8.104) one gets:

$$\begin{pmatrix} dx \\ dy \\ \dots \\ da_1 \\ da_2 \\ da_3 \\ da_4 \\ da_5 \\ da_6 \\ da_7 \\ \dots \\ d\lambda \\ \dots \\ d\mu_2 \\ d\mu_2 \\ \dots \\ dz \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -1/2 \\ 0 & 0 & 0 & 3 & 1 & 0 & 1/2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & -4 & 0 & -12 & -4 & 0 & -2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & -2 & 0 & -4 & -2 & 0 & -2 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 9 & 4 & 0 & 12 & 4 & 0 & -1 \end{pmatrix} \quad (8.107)$$

which allows obtaining the partial and directional derivatives sought after, i.e.,

$$\frac{\partial z^*}{\partial a_1} = 9, \quad \frac{\partial z^*}{\partial a_2} = 4, \quad \frac{\partial z^{*-}}{\partial a_3} = 0, \quad \frac{\partial z^{-*}}{\partial a_4} = -12, \quad \frac{\partial z^{*+}}{\partial a_5} = 4, \quad \frac{\partial z^*}{\partial a_6} = 0, \quad \frac{\partial z^{*+}}{\partial a_7} = -1$$

where the asterisks refer to optimal values. In addition, the same directional and partial derivatives of x, y, λ, μ_1 and μ_2 , can be obtained from the corresponding rows in (8.107). ■

8.5 Exercises

Exercise 8.1. Consider the following optimization problem:

$$\begin{array}{ll} \text{Minimize} & -(x_1 + 1)^2 - (x_2 + 1)^2 \\ & x_1, x_2 \end{array}$$

subject to

$$x_1^2 + x_2^2 \leq 1$$

1. Solve the problem graphically.
2. Obtain the Lagrangian function.
3. Solve the problem using the KKT conditions and show that the optimal value of the objective function, attained at the point $(1/\sqrt{2}, 1/\sqrt{2})$, is $-(1 + \sqrt{2})^2$.
4. Show that the dual function is $\theta(\mu) = \frac{\mu(1 + \mu)}{1 - \mu}$ and discuss its range.
5. Formulate the dual problem.
6. Solve the dual problem and show that the optimal value of the objective function, attained at the point $\mu = 1 + \sqrt{2}$, is $-(1 + \sqrt{2})^2$.
7. Check that there is no duality gap.
8. Replace the 1 in the right hand side of the constraint by parameter a and discuss the existence of derivatives with respect to a .
9. Finally, obtain the derivatives of the optimal objective function value and the coordinates of the solution point with respect to a .

Exercise 8.2. Consider the water supply problem in Section ?? with the data in Figure ?. Modify the optimization problem for the sensitivities of the objective function with respect to the capacities r_i and flows q_j to be directly obtained as the values of dual variables.

Exercise 8.3. Consider the problem

$$\begin{array}{ll} \text{Minimize} & Q = (x - x_0)^2 + (y - y_0)^2 \\ & x, y \end{array}$$

subject to

$$(x - a)^2 + (y - b)^2 = r^2$$

where x_0, y_0, a, b and r are given constants.

1. Give a geometrical interpretation to this problem and plot a graph to explain it.

2. Solve the problem analytically.
3. Obtain the sensitivity of Q with respect to x_0, y_0, a, b and r based on the analytical solution.
4. Modify the initial problem to obtain the above sensitivities using the techniques developed in this chapter.

Exercise 8.4. Consider the problem

$$\begin{array}{ll} \text{Maximize} & xy \\ & x, y \end{array}$$

subject to

$$\begin{array}{ll} xy & \leq a^2 \\ y & = bx \\ x & \leq c \\ y & \leq d \\ x & \geq 0 \\ y & \geq 0 \end{array}$$

where a, b, c and d are given constants.

1. Give a geometrical interpretation to this problem and plot a graph to explain it.
2. Solve the problem analytically.
3. Obtain the sensitivity of Q with respect to a, b, c and d based on the analytical solution.
4. Modify the initial problem to obtain the above sensitivities using the techniques developed in this chapter.
5. Discuss the existence of partial derivatives of Q with respect to a, b, c and d , for the case $a = 4; b = 1; c = 2; d = 2$.

Exercise 8.5. Consider the operation of a multi-area electricity network example in Section ??, page ??, and the numerical data in Tables ??, ?? and ??. Modify the problem to obtain the sensitivities of the objective function with respect to all data values.

Exercise 8.6. A circle of minimum radius must be found such that it contains a given set of points. To this end:

1. Simulate 30 points with random coordinates $N(0, 1)$.
2. State the problem as one optimization problem. Note that the variables are the coordinates of its center (x_0, y_0) and its radius r .
3. Solve the problem using GAMS.

4. Use the technique of converting the data to artificial variables and incorporating constraints stating that these variables are equal to their corresponding data values, to obtain the sensitivities of the minimum radius with respect to the point coordinates.
5. Identify the three points that define the circle based on the above sensitivities.
6. Generalize the problem to the case of n dimensions.

Chapter 9

Decomposition Techniques

9.1 Introduction

The size of a linear programming problem can be very large. One can encounter in practice problems with several hundred thousands of equations and/or unknowns. To solve these problems the use of some special techniques is either convenient or required. Alternatively, a distributed solution of large problems may be desirable for technical or practical reasons. Decomposition techniques allow certain type of problems to be solved in a decentralized or distributed fashion. Alternatively, they lead to a drastic simplification in the solution procedure of the problem under study.

Uno de los casos en los que estas técnicas pueden tener gran interés es en los problemas de optimización combinados con estudios de fiabilidad, que se caracterizan por tener dos niveles: el nivel 1, es la optimización global de todas las variables de diseño \mathbf{d} , para determinar los valores deterministas, esperados o característicos de las mismas, y el nivel 2 en el que se estima la fiabilidad mediante un problema de optimización que se estudiará más adelante. Este problema, claramente está descompuesto en dos etapas.

Otro campo en el que pueden tener gran interés es en el tratamiento desacoplado de problemas, como por ejemplo, problemas asociados a la optimización y fiabilidad de estructuras en combinación con elementos finitos (véase Enevoldsen y Sorensen ([70]) y Frangopol y Hendawi ([79])). Problema en el que se han logrado resolver con éxito ejemplos de dimensiones pequeñas y moderadas, pero de los que se precisan técnicas para resolución de problemas mayores.

To apply advantageously a decomposition technique, the considered problem should have the appropriate structure. Two such structures arise in practice. The first is characterized by complicating constraints, and the second by complicating variables. The complicating constraints and variables are those that complicate the solution of the problem, or prevent a straightforward solution of the problem or a solution by blocks, i.e., they make the problem more difficult to solve.

9.2 Benders Decomposition for Complicating Variables

9.2.1 Introduction

Benders decomposition is analyzed below to address nonlinear problems with decomposable structure and complicating variables.

9.2.2 Benders Decomposition

The problem structure required to apply advantageously Benders decomposition is

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}, \mathbf{y}) \\ & \mathbf{x}, \mathbf{y} \end{array} \quad (9.1)$$

sujeito a

$$\mathbf{c}(\mathbf{x}) \leq \mathbf{0} \quad (9.2)$$

$$\mathbf{d}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \quad (9.3)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \quad (9.4)$$

$$\mathbf{y}^{\text{lo}} \leq \mathbf{y} \leq \mathbf{y}^{\text{up}} \quad (9.5)$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$, $f(\mathbf{x}, \mathbf{y}) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$, $\mathbf{c}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^p$ y $\mathbf{d}(\mathbf{x}, \mathbf{y}) : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^q$. The problem includes both equality and inequality constraints. Variables \mathbf{x} are complicating variables, i.e., variables that if fixed to given values render a simple or decomposable problem. Auxiliary function $\alpha(\mathbf{x})$ is defined below. Function $\alpha(\mathbf{x})$ expresses the objective function of the original problem (total costs) as a function solely of the complicating variables.

$$\alpha(\mathbf{x}) = \underset{\mathbf{y}}{\text{minimum}} f(\mathbf{x}, \mathbf{y}) \quad (9.6)$$

subject to

$$\mathbf{d}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \quad (9.7)$$

$$\mathbf{y}^{\text{lo}} \leq \mathbf{y} \leq \mathbf{y}^{\text{up}}. \quad (9.8)$$

Using function $\alpha(\mathbf{x})$, the original problem (9.1)-(9.5) can be expressed as

$$\begin{array}{ll} \text{Minimize} & \alpha(\mathbf{x}) \\ & \mathbf{x} \end{array} \quad (9.9)$$

subject to

$$\mathbf{c}(\mathbf{x}) \leq \mathbf{0} \quad (9.10)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}}. \quad (9.11)$$

The procedure explained below produces iteratively better and better approximations to function $\alpha(\mathbf{x})$. If complicating variables are fixed to specific values using constraints of

the form, $\mathbf{x} \leftarrow \mathbf{x}^{(i)}$, and so that $\mathbf{c}(\mathbf{x}^{(i)}) \leq 0$, the resulting problem is easy to solve. This problem has the form:

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}, \mathbf{y}) \\ & \mathbf{y} \end{array} \quad (9.12)$$

subject to

$$\mathbf{d}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \quad (9.13)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \quad (9.14)$$

$$\mathbf{x} = \mathbf{x}^{(i)} : \boldsymbol{\lambda}^{(i)}. \quad (9.15)$$

The problem above is denominated subproblem. Typically, it decomposes in many subproblems. The solution of the problem above provides values for the non complicating variables, $\mathbf{y}^{(i)}$, and the dual variable vector associated to those constraints that fix the complicating variables to given values. This sensitivity vector is denoted by $\boldsymbol{\lambda}^{(i)}$. An upper bound of the optimal objective function value is readily available because problem (9.12)-(9.15) is more constrained than the original one. This upper bound is $z_{\text{up}}^{(i)} = f(\mathbf{x}^{(i)}, \mathbf{y}^{(i)})$. The information obtained solving the subproblem allows reproducing more and more accurately the original problem. Moreover, if function $\alpha(\mathbf{x})$ is convex, the following problem approximates from below the original one.

$$\begin{array}{ll} \text{Minimize} & \alpha \\ & \alpha, \mathbf{x} \end{array} \quad (9.16)$$

sujeito a

$$\mathbf{c}(\mathbf{x}) \leq \mathbf{0} \quad (9.17)$$

$$\alpha \geq f(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) + \sum_{k=1}^n \lambda_k^{(i)} (x_k - x_k^{(i)}) \quad (9.18)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \quad (9.19)$$

The last constraint of the problem above is called Benders cut. The problem itself is denominated master problem. Note that the optimal objective function value of this problem is a lower bound of the optimal objective function value of the original problem. This is so because master problem is a relaxation of the original problem. The solution of this master problem provides new values for the complicating variables that are used for solving a new subproblem. In turn, this subproblem provides information to formulate a more accurate master problem that provides new values of complicating variables. The procedure continues until upper and lower bounds of the objective function optimal value are close enough.

Benders decomposition works as follows:

Algorithm 9.1 (Benders Decomposition).

Step 0: Initialization. Find feasible values for the complicating variables \mathbf{x}_0 , so that $\mathbf{c}(\mathbf{x}_0) \leq \mathbf{0}$.

Set $\nu = 1$, $\mathbf{x}^{(\nu)} = \mathbf{x}_0$, $z_{\text{up}}^{(\nu-1)} = +\infty$, $z_{\text{down}}^{(\nu-1)} = -\infty$.

Step 1: Subproblem solution. Solve subproblem or subproblems:

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}, \mathbf{y}) \\ & \mathbf{y} \end{array}$$

subject to

$$\begin{array}{ll} \mathbf{d}(\mathbf{x}, \mathbf{y}) & \leq \mathbf{0} \\ \mathbf{y}^{\text{lo}} & \leq \mathbf{y} \leq \mathbf{y}^{\text{up}} \\ \mathbf{x} & = \mathbf{x}^{(\nu)} : \boldsymbol{\lambda}^{(\nu)} \end{array}$$

The solution of this subproblem provides $\mathbf{y}^{(\nu)}$, $f(\mathbf{x}^{(\nu)}, \mathbf{y}^{(\nu)})$, $\boldsymbol{\lambda}^{(\nu)}$.

Update the objective function upper bound, $z_{\text{up}}^{(\nu)} \leftarrow f(\mathbf{x}^{(\nu)}, \mathbf{y}^{(\nu)})$.

Step 2: Convergence check. If $|z_{\text{up}}^{(\nu)} - z_{\text{lo}}^{(\nu)}| / |z_{\text{lo}}^{(\nu)}| \leq \varepsilon$, the solution with a level of accuracy ε of the objective function is

$$\begin{array}{l} \mathbf{x}^* = \mathbf{x}^{(\nu)} \\ \mathbf{y}^* = \mathbf{y}^{(\nu)}. \end{array}$$

Otherwise, set $\nu \leftarrow \nu + 1$ and go to Step 3.

Step 3: Master problem solution. Solve the problem.

$$\begin{array}{ll} \text{Minimize} & \alpha \\ & \alpha, \mathbf{x} \end{array}$$

subject to

$$\begin{array}{ll} \alpha & \geq f(\mathbf{x}^{(i)}, \mathbf{y}^{(i)}) + \sum_{k=1}^n \lambda_k^{(i)} (x_k - x_k^{(i)}); \quad \forall i = 1, \dots, \nu \\ \mathbf{c}(\mathbf{x}) & \leq \mathbf{0} \\ \mathbf{x}^{\text{lo}} & \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \end{array}$$

Note that at every iteration a new constraint is added. The solution of the master problem provides $\mathbf{x}^{(\nu)}$, and $\alpha^{(\nu)}$.

Update objective function lower bound, $z_{\text{lo}}^{(\nu)} \leftarrow \alpha^{(\nu)}$ and go to Step 1.

■

Convergence is guaranteed if $\alpha(\mathbf{x})$ is convex.

Illustrative Example 9.1 (Benders decomposition for a nonlinear problem). Consider the problem

$$\begin{array}{ll} \text{Minimize} & z = xy_2 + y_1y_2 \\ & x, y_1, y_2 \end{array}$$

subject to

$$\begin{array}{rcl} 4y_1 - x^2 & \leq & 5 \\ 2y_2^2 - \frac{xy_2}{3} & \leq & 2 \\ y_1 & \geq & 0 \\ x & \geq & 2 \\ x & \leq & 12 \end{array}$$

The optimal values of this example are $x^* = 12$, $y_1^* = -37.25$ and $y_2^* = -0.41$ with objective function value equal to $z^* = -30.34$. If variable x is considered to be a complicating variable, the above problem is solved using the Benders decomposition algorithm.

Step 0: Initialization. The iteration counter is initialized, $\nu = 1$. The initial value for the complicating variable x is $x^{(1)} = 2$. The lower bound of the objective function is $z_{\text{lo}}^{(1)} = -\infty$.

Step 1: Subproblem solution. The subproblem below is solved.

$$\begin{array}{ll} \text{Minimizes} & z = xy_2 + y_1y_2 \\ & x, y_1, y_2 \end{array}$$

subject to

$$\begin{array}{rcl} 4y_1 - x^2 & \leq & 5 \\ 2y_2^2 - \frac{xy_2}{3} & \leq & 2 \\ y_1 & \geq & 0 \\ x & = & 2 : \lambda \end{array}$$

which solution is $y_1^{(1)} = 2.25$, $y_2^{(1)} = -0.85$ and $\lambda^{(1)} = -2.81$ with an objective function value $z = -6.99$. The upper bound of the objective function optimal value is $z_{\text{up}}^{(1)} = -6.99$.

Step 2: Convergence checking. The expression $|z_{\text{up}}^{(1)} - z_{\text{lo}}^{(1)}|/|z_{\text{lo}}^{(1)}| = 1.0$ is not small enough, therefore the iteration counter is updated $\nu \leftarrow \nu + 1 = 2$ and the procedure continues in Step 3.

Step 3: Master problem solution. The master problem below is solved.

$$\begin{array}{ll} \text{Minimize} & \alpha \\ & \alpha, x \end{array}$$

subject to

$$\begin{array}{rcl} -6.99 - 2.81 \times (x - 2) & \leq & \alpha \\ \alpha & \geq & -50 \\ x & \geq & 2 \\ x & \leq & 12 \end{array}$$

The solution of this problem is $x^{(1)} = 12.00$ and $\alpha^{(1)} = -35.13$. The lower bound of the objective function optimal value is $z_{1o}^{(2)} = -35.13$. Procedure continues in Step 1.

Step 1: Subproblem solution. The subproblem below is solved.

$$\begin{array}{ll} \text{Minimize} & z = xy_2 + y_1y_2 \\ & x, y_1, y_2 \end{array}$$

subject to

$$\begin{array}{rcl} 4y_1 - x^2 & \leq & 5 \\ 2y_2^2 - \frac{xy_2}{3} & \leq & 2 \\ y_1 & \geq & 0 \\ x & = & 12 : \lambda \end{array}$$

which solution is $y_1^{(2)} = 37.25$, $y_2^{(2)} = -0.41$ and $\lambda^{(2)} = -1.94$ with an objective function value $z = -30.34$. The upper bound of the objective function optimal value is $z_{up}^{(2)} = -30.34$.

Step 2: Convergence checking. The expression $|z_{up}^{(2)} - z_{1o}^{(2)}|/|z_{1o}^{(2)}| = 0.136$ is not small enough, therefore the iteration counter is updated $\nu \leftarrow \nu + 1 = 3$.

Step 3: Master problem solution. The master problem below is solved.

$$\begin{array}{ll} \text{Minimize} & \alpha \\ & \alpha, x \end{array}$$

subject to

$$\begin{array}{rcl} -6.99 - 2.81 \times (x - 2) & \leq & \alpha \\ -30.34 - 1.94 \times (x - 12) & \leq & \alpha \\ \alpha & \geq & -50 \\ x & \geq & 2 \\ x & \leq & 12 \end{array}$$

The solution of this problem is $x^{(3)} = 12$ and $\alpha^{(3)} = -30.34$. The lower bound of the objective function optimal value is $z_{1o}^{(3)} = -30.34$.

Step 1: Subproblem solution. The subproblem below is solved.

$$\begin{array}{ll} \text{Minimize} & z = xy_2 + y_1y_2 \\ & x, y_1, y_2 \end{array}$$

subject to

$$\begin{array}{rcl} 4y_1 - x^2 & \leq & 5 \\ 2y_2^2 - \frac{xy_2}{3} & \leq & 2 \\ y_1 & \geq & 0 \\ x & = & 12 : \lambda \end{array}$$

which solution is $y_1^{(3)} = 37.25$, $y_2^{(3)} = -0.41$ and $\lambda^{(3)} = -1.94$ with an objective function value $z = -30.34$. The upper bound of the objective function optimal value is $z_{up}^{(3)} = -30.34$.

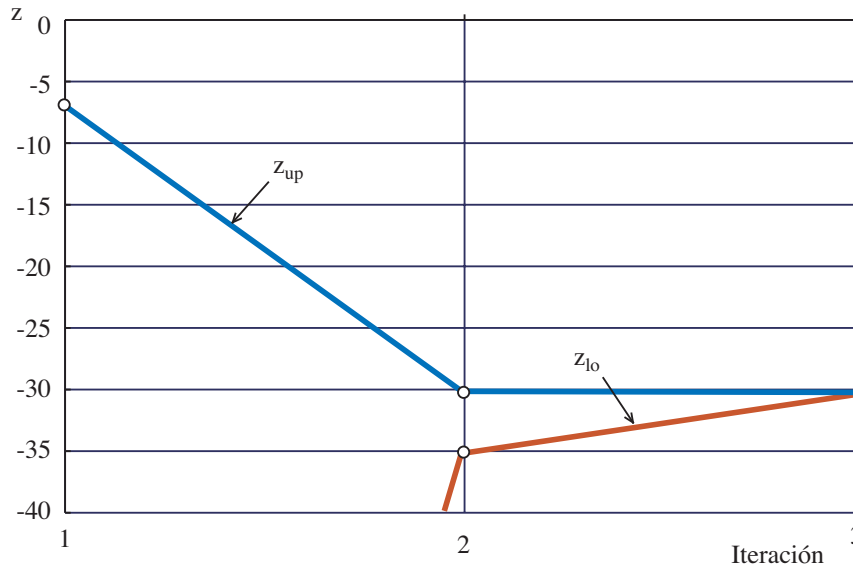


Figure 9.1: Evolution of the upper and lower limits of the objective function value in the Illustrative Example 9.1.

Step 2: Convergence checking. The expression $|z_{up}^{(3)} - z_{lo}^{(3)}| / |z_{lo}^{(3)}| = 0.0$ is small enough, therefore, the optimal solution $x^* = 12$, $y_1^* = -37.25$ y $y_2^* = -0.41$ has been found.

In Figure 9.1 it is shown the convergence of the iterative process. In this case the solution of the problem is exact because the optimum value of the complicating variable x coincides with its upper bound. ■

9.3 Decomposition Techniques for Complicating Constraints

9.3.1 Introduction

This section analyzes nonlinear problems with decomposable structure. The complicating constraint and the complicating variable cases are both considered. For dealing with the complicating constraint case, three procedures are analyzed: Lagrangian relaxation (LR), augmented Lagrangian decomposition (ALD) and optimality condition decomposition (OCD). In the last section a new relaxation procedure for dealing with reliability based optimization problems.

9.3.2 Complicating Constraints

Complicating constraints are constraints that if relaxed, the resulting problem decomposes in several simpler problems. Alternatively, the resulting problem attains such a structure

that its solution is simple. The decomposable case is the most interesting one in practice and the one mostly considered in this section.

For the Lagrangian relaxation technique to be applied advantageously to a mathematical programming problem, the problem should have the following structure

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}) \\ \mathbf{x} & \end{array} \quad (9.20)$$

sujeito a

$$\mathbf{a}(\mathbf{x}) = \mathbf{0} \quad (9.21)$$

$$\mathbf{b}(\mathbf{x}) \leq \mathbf{0} \quad (9.22)$$

$$\mathbf{c}(\mathbf{x}) = \mathbf{0} \quad (9.23)$$

$$\mathbf{d}(\mathbf{x}) \leq \mathbf{0} \quad (9.24)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \quad (9.25)$$

where $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{a}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^{n_a}$, $\mathbf{b}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^{n_b}$, $\mathbf{c}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^{n_c}$, $\mathbf{d}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^{n_d}$, n_a, n_b, n_c and n_d are scalars. Constraints $\mathbf{c}(\mathbf{x}) = \mathbf{0}$, and $\mathbf{d}(\mathbf{x}) \leq \mathbf{0}$ are complicating constraints, i.e., constraints that if relaxed, problem (9.20)-(9.25) becomes drastically simplify.

In Sections 9.3.3, ?? and ??, three different decomposition algorithms are studied. Sections 9.3.3 and ?? presents the basic theory of the Lagrangian relaxation and the augmented Lagrangian decomposition procedures, respectively. Section ?? presents a partitioning technique based on decomposing the optimality conditions of the original problem; it is denominated optimality condition decomposition (OCD).

9.3.3 Lagrangian relaxation

In this subsection Lagrangian relaxation method for solving problem (9.20)-(9.24) is briefly introduced. Constraints (9.23) and (9.24) are remove from the initial problems by means of the Lagrangian function (LF), that is defined as (see Bazaraa [16] and Luenberger [?]):

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{c}(\mathbf{x}) + \boldsymbol{\mu}^T \mathbf{d}(\mathbf{x}), \quad (9.26)$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are Lagrange multiplier vectors. Under local convexity assumptions ($\nabla_{\mathbf{x}}^2 \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) > \mathbf{0}$, where \mathbf{x}^* , $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$ are the optimal primal and dual values) the dual function (DF) is defined as

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \underset{\mathbf{x}}{\text{minimum}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \quad (9.27)$$

subject to

$$\mathbf{a}(\mathbf{x}) = \mathbf{0} \quad (9.28)$$

$$\mathbf{b}(\mathbf{x}) \leq \mathbf{0} \quad (9.29)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \quad (9.30)$$

The dual function is concave and in general non-differentiable [115]. This is a fundamental fact that is exploited in the algorithms described below. The dual problem (DP) is then defined as

$$\begin{aligned} & \text{Maximize } \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ & \boldsymbol{\lambda}, \boldsymbol{\mu} \end{aligned} \quad (9.31)$$

subject to

$$\boldsymbol{\mu} \geq \mathbf{0} \quad (9.32)$$

The LR decomposition procedure is attractive if the dual function is easily evaluated for given values $\bar{\boldsymbol{\lambda}}$ and $\bar{\boldsymbol{\mu}}$ of the multiplier vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ respectively. In other words, if it is easy to solve the so-called relaxed primal problem (RPP) for given $\bar{\boldsymbol{\lambda}}$ and $\bar{\boldsymbol{\mu}}$, i.e., the problem

$$\begin{aligned} & \text{Minimize } \mathcal{L}(\mathbf{x}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}) \\ & \mathbf{x} \end{aligned} \quad (9.33)$$

The above problem typically decomposes into subproblems. That is

$$\mathbf{a}(\mathbf{x}) = \mathbf{0} \quad (9.34)$$

$$\mathbf{b}(\mathbf{x}) \leq \mathbf{0} \quad (9.35)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \quad (9.36)$$

The above problem typically decomposes into subproblems. That is

$$\begin{aligned} & \text{Minimize } \sum_{i=1}^n \mathcal{L}_i(\mathbf{x}_i, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}) \\ & \mathbf{x}_i; i = 1, \dots, n \end{aligned} \quad (9.37)$$

subject to

$$\mathbf{a}_i(\mathbf{x}_i) = \mathbf{0}; \quad i = 1, \dots, n \quad (9.38)$$

$$\mathbf{b}_i(\mathbf{x}_i) \leq \mathbf{0}; \quad i = 1, \dots, n \quad (9.39)$$

$$\mathbf{x}_i^{\text{lo}} \leq \mathbf{x}_i \leq \mathbf{x}_i^{\text{up}}; i = 1, \dots, n \quad (9.40)$$

This decomposition facilitates its solution, and normally allows physical and economical interpretations. Problem (9.37)-(9.40) is called the decomposed primal problem (DPP). Then, the resulting subproblems can be solved in parallel.

Under local convexity assumptions, the local duality theorem says that

$$f(\mathbf{x}^*) = \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \quad (9.41)$$

where \mathbf{x}^* is the minimizer for the primal problem and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ the maximizer for the dual problem.

In the non-convex case, given a feasible solution for the primal problem, \mathbf{x} , and a feasible solution for the dual problem, $(\boldsymbol{\lambda}, \boldsymbol{\mu})$, the weak duality theorem says that

$$f(\mathbf{x}^*) \geq \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \quad (9.42)$$

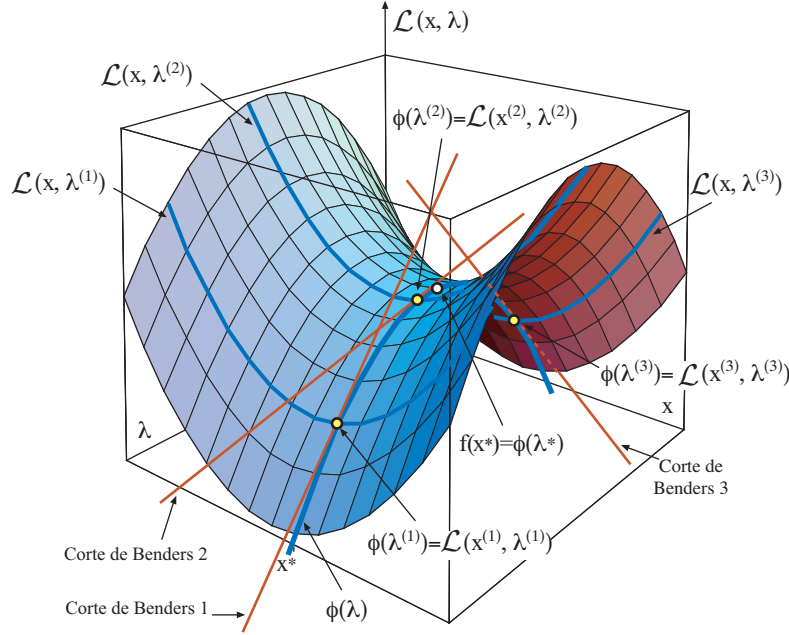


Figure 9.2: Ilustración gráfica de la reconstrucción por hiperplanos de la función dual $\phi(\lambda)$.

Therefore, in the convex case, the solution of the dual problem provides the solution of the primal problem. In the non-convex case the objective function value at the optimal solution of the dual problem provides a lower bound to the objective function value at the optimal solution of the primal problem. Most engineering and science mathematical programming problems are non-convex. The solution of the dual problem is called Phase 1 of the LR procedure.

Several multiplier updating procedures exist. In the following, for the sake of clarity, multiplier vectors λ and μ are renamed as $\theta = \text{column}(\lambda, \mu)$:

Subgradient Method. The column vector of constraint mismatches at iteration ν constitutes a subgradient of the dual function [16], i.e.,

$$\mathbf{s}^{(\nu)} = \text{column} \left[\mathbf{c}(\mathbf{x}^{(\nu)}), \mathbf{d}(\mathbf{x}^{(\nu)}) \right] \quad (9.43)$$

The multiplier vector is updated as (see Polyak [137]):

$$\boldsymbol{\theta}^{(\nu+1)} = \boldsymbol{\theta}^{(\nu)} + k^{(\nu)} \frac{\mathbf{s}^{(\nu)}}{|\mathbf{s}^{(\nu)}|} \quad (9.44)$$

Cutting Plane Method (CP). The updated multiplier vector is obtained by solving the linear programming problem below

$$\begin{aligned} &\text{Maximize } z \\ &z, \boldsymbol{\theta} \in C \end{aligned} \quad (9.45)$$

subject to

$$z \leq \phi^{(k)} + \mathbf{s}^{(k)T} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}); k = 1, \dots, \nu \quad (9.46)$$

where C is a convex and compact set. It is made up of the ranges of variations of the multipliers, i.e., $C = \{\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{down}} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_{\text{up}}\}$. It should be noted that the above constraints represent half-spaces (hyperplanes) on the multiplier space (see Figure 9.2). Values $\boldsymbol{\theta}_{\text{down}}$ and $\boldsymbol{\theta}_{\text{up}}$ are, in general, easily obtained from the physical or economical properties of the system which is modelled. Note that this updating procedure is exactly the same as the master problem of the Benders decomposition technique treated in Section ??, but with a concave function and maximizing the objective function instead of minimizing it. It should be noted that the number of constraints of the above problem grows with the number of iterations. The above problem is a relaxed dual problem (RDP) which gets closer to the actual dual problem as the number of iterations grows. The CP method achieves a dual optimum by reconstructing the dual function. It reconstructs the region of interest and regions of no interest. This reconstruction is computationally expensive and therefore the CP method computational burden is high. This algorithm is typically stopped when the multiplier vector difference between two consecutive iterations is below a pre-specified threshold.

Bundle Method (BD). The updated multiplier vector is obtained solving the relaxed dual quadratic programming problem below

$$\begin{aligned} & \text{Maximize} && z - \alpha^{(\nu)} |\boldsymbol{\theta} - \boldsymbol{\Theta}^{(\nu)}|^2 \\ & z, \boldsymbol{\theta} \in C \end{aligned} \quad (9.47)$$

subject to

$$z \leq \phi^{(k)} + \mathbf{s}^{(k)T} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}); k = 1, \dots, \nu \quad (9.48)$$

where α is a penalty parameter, $|\cdot|$ is the Euclidean norm and $\boldsymbol{\Theta}$, “the center of gravity”, is a vector of multipliers centered in the feasibility region so that oscillations are avoided. It should be noted that the number of constraints of the problem above grows with the number of iterations.

Trust Region Method (TR). The updated multiplier vector is obtained solving the relaxed dual linear programming problem below

$$\begin{aligned} & \text{Maximize} && z \\ & z, \boldsymbol{\theta} \in C^{(\nu)} \end{aligned} \quad (9.49)$$

subject to

$$z \leq \phi^{(k)} + \mathbf{s}^{(k)T} (\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}); k = 1, \dots, n; n \leq \bar{n} \quad (9.50)$$

where \bar{n} is the maximum number of constraints considered when solving the problem above, and $C^{(\nu)}$ is the dynamically updated set defining the feasibility region for the multipliers. If the number of iterations is larger than the specified maximum number of constraints, the excess constraints are eliminated as stated below.

9.3.4 Augmented lagrangian relaxation

La relajación lagrangiana aumentada es una variante de la relajación lagrangiana en la que se trabaja con una función lagrangiana aumentada de la forma (véase Luenberger [?], Cohen[46] o Bertsekas [18]):

$$\mathcal{A}(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \alpha, \beta) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{c}(\mathbf{x}) + \boldsymbol{\mu}^T \hat{\mathbf{d}}(\mathbf{x}, \mathbf{z}) + \frac{1}{2}\alpha \|\mathbf{c}(\mathbf{x})\|^2 + \frac{1}{2}\beta \|\hat{\mathbf{d}}(\mathbf{x}, \mathbf{z})\|^2 \quad (9.51)$$

where penalty parameters α and β are large enough scalars to ensure local convexity, the component i of function $\hat{\mathbf{d}}(\mathbf{x}, \mathbf{z})$ is defined as $\tilde{d}_i(\mathbf{x}, \mathbf{z}) = d_i(\mathbf{x}) + z_i^2$, and the z_i 's ($i = 1, \dots, n_d$) are additional variables to transform inequality constraints into equality constraints. It should be noted that the quadratic terms confer the Augmented Lagrangian function good local convexity properties.

9.3.5 Optimality condition decomposition

The decomposition technique analyzed in this Section can be interpreted as a particular implementation of the Lagrangian relaxation procedure. It is motivated by a natural decomposition of the optimality conditions of the original problem. An in-depth analysis of this technique can be found in Conejo [47].

For convenience, problem (9.20)-(9.24) can be written as

$$\underset{\mathbf{x}_a}{\text{minimize}} \quad \sum_{a=1}^A f_a(\mathbf{x}_a) \quad (9.52)$$

subject to

$$\mathbf{h}(\mathbf{x}_1, \dots, \mathbf{x}_A) \leq \mathbf{0} \quad (9.53)$$

$$\mathbf{g}_a(\mathbf{x}_a) \leq \mathbf{0} \quad a = 1, \dots, A \quad (9.54)$$

$$\mathbf{x}_a^{\text{lo}} \leq \mathbf{x}_a \leq \mathbf{x}_a^{\text{up}} \quad a = 1, \dots, A \quad (9.55)$$

where \mathbf{x}_a are the variables for each block a in which the original problem decomposes. It should be noted that the sets of equations (9.53) and (9.54) represent both equality and inequality constraints. Equations (9.53) are complicating constraints. These equations contain variables from different blocks and prevent each subproblem from being solved independently. If these equations are removed from problem (9.52)-(9.55), the resulting problem can be trivially decomposed into one subproblem for each block. Constraints (9.54) contain only variables belonging to block a for $a = 1, \dots, A$. Problem (9.52)-(9.54) is equivalent to the problem below

$$\text{Minimize} \quad \sum_{a=1}^A f_a(\mathbf{x}_a) + \sum_{a=1}^A \boldsymbol{\lambda}_a^T \mathbf{h}_a(\mathbf{x}_1, \dots, \mathbf{x}_A) \quad (9.56)$$

subject to

$$\mathbf{h}(\mathbf{x}_1, \dots, \mathbf{x}_A) \leq \mathbf{0} \quad a = 1, \dots, A \quad (9.57)$$

$$\mathbf{g}_a(\mathbf{x}_a) \leq \mathbf{0} \quad a = 1, \dots, A \quad (9.58)$$

$$\mathbf{x}_a^{\text{lo}} \leq \mathbf{x}_a \leq \mathbf{x}_a^{\text{up}} \quad a = 1, \dots, A \quad (9.59)$$

where constraints (9.53) have been separated in different blocks. Note that the way in which these constraints are distributed does not affect the solution of the problem, i.e., they can be distributed based on engineering insight.

Given trial values to all variables and multipliers (indicated by overlining) different than those in block a , problem (9.56)-(9.58) reduces to

$$\text{Minimize}_{\mathbf{x}_a} \quad k + f_a(\mathbf{x}_a) + \sum_{b=1, b \neq a}^A \bar{\lambda}_b^T \mathbf{h}_b(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_{a-1}, \mathbf{x}_a, \bar{\mathbf{x}}_{a+1}, \dots, \bar{\mathbf{x}}_A) \quad (9.60)$$

subject to

$$\mathbf{h}_a(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_{a-1}, \mathbf{x}_a, \bar{\mathbf{x}}_{a+1}, \dots, \bar{\mathbf{x}}_A) \leq \mathbf{0}, \quad (9.61)$$

$$\mathbf{g}_a(\mathbf{x}_a) \leq \mathbf{0}, \quad (9.62)$$

$$\mathbf{x}_a^{\text{lo}} \leq \mathbf{x}_a \leq \mathbf{x}_a^{\text{up}}, \quad (9.63)$$

where $k = \sum_{b=1, b \neq a}^A f_b(\bar{\mathbf{x}}_b)$ is a constant. The dual variable vector corresponding to constraint

((9.61) is denoted by λ_a . The reduced problem (9.60)-(9.62) can be obtained for every block of the original problem. The proposed decomposition technique is actually based on the solutions of these reduced block-related problems.

If the first-order optimality conditions of every block reduced subproblem (9.60)-(9.62) ($a = 1, \dots, A$) are stucked together, it can be observed that they are identical to the first-order optimality conditions of the global problem (9.52)-(9.54). It should be emphasized that this is a relevant result that is exploited in the algorithm below. The main difference between the Lagrangian Relaxation algorithm and the proposed decomposition one is that Lagrangian Relaxation adds all the complicating constraints into the objective function. Therefore it needs auxiliary procedures to update the Lagrange multipliers. On the contrary, the analyzed technique does not need any procedure to update the multipliers because this updating is automatic and results from keeping the complicating constraints (9.61) in every block subproblem. The proposed approach has the advantage that convergence properties do not require to attain an optimal solution of the subproblems at each iteration of the algorithm. It is enough to perform a single iteration for each subproblem, and then to update variable values. As a consequence, computation times can be significantly reduced with respect to other methods that require the computation of the optimum for the subproblems in order to attain convergence.

9.3.6 Application to reliability problems

The method explained previously can be applied to the reliability based optimization problems (RBSO) shown in Chapter 10.

9.4 Complicating Variables and Constraints

When the problem considered has both complicating variables and constraints. In such situation a nested decomposition can be used. For instance, an outer Benders Decomposi-

tion algorithm may deal with complicating variables while an inner Lagrangian Relaxation procedure may deal with complicating constraints.

9.5 Optimization Packages

Dado que las técnicas explicadas en las secciones anteriores se pueden aplicar mediante cualquier paquete o rutina de optimización, se presentarán a continuación una serie de alternativas existentes en el mercado que permitirían resolver todos los problemas relativos a fiabilidad.

Los paquetes de optimización son programas con un lenguaje de programación propio que permiten el modelado, análisis y resolución de múltiples problemas de optimización. La ventaja que tienen es que trabajan con algoritmos de optimización estándar contrastados y no es preciso que el usuario tenga un conocimiento exhaustivo del funcionamiento de los mismos aunque, lógicamente, cuanto más familiarizado se este con los métodos que utilice, mayor será el rendimiento.

Algunos de los existentes son:

AIMMS. El paquete AIMMS [129, 19] del Paragon Decision Technology B.V. en Haarlem.

AMPL. El paquete AMPL [76, 130] del Dash Optimization de Nueva York.

GAMS. El paquete GAMS [131] pertenece a la GAMS Development Corp. de Washington. GAMS también está disponible para clientes europeos en el ARKI Consulting & Development A/S en Bagsvaerd, Dinamarca, o en GAMS Software GmbH en Colonia, Alemania.

LINDO y What's Best. El paquete [132] pertenece a LINDO Systems, Inc. en Chicago.

MPL. Este paquete [133] pertenece al Maximal Software Inc, en Arlington.

En todos los trabajos realizados para esta tesis se ha optado por el GAMS (General Algebraic Modeling System) dada la familiaridad con su lenguaje de programación. De todos los algoritmos que puede utilizar el programa se ha optado por utilizar siempre el optimizador CONOPT, especialmente diseñado para trabajar con problemas no lineales de larga escala, y que ha demostrado tener un comportamiento muy bueno con problemas relativos a fiabilidad. Está basado en el método del gradiente reducido generalizado (GRG, 'Generalized Reduced Gradient Method') y ha sido elegido por su eficiencia y fiabilidad para resolver un amplio abanico de modelos. El método GRG original mejora la convergencia para modelos con un gran número de no linealidades, es decir, con modelos complicados. De ahí que sea preferible la utilización de este optimizador en modelos muy no lineales que en aquellos en los que es complicado alcanzar la factibilidad. Además se han introducido subrutinas de todo tipo, como las de programación lineal secuencial, que hacen que el método sea eficiente en problemas poco o medianamente no lineales.

Este optimizador está especialmente diseñado para la resolución de problemas grandes de tipo 'sparse', y puede trabajar con problemas que tienen del orden de 10000 restricciones. De hecho, el factor limitante del método es el número de variables superbásicas, más que

el número total de variables o restricciones, es decir, el número de grados de libertad en el entorno del punto óptimo. Con más de 150 variables superbásicas puede dar problemas de convergencia, especialmente si está mal escalado.

Como vemos, por las características del optimizador, parece claro que es un método adecuado para la resolución de problemas relativos a la fiabilidad, en los que intervienen un gran número de variables en su mayoría auxiliares y sometidos a grandes no linealidades. Este comportamiento se debe o bien a la transformación de Rosenblatt, o a la obtención de las ecuaciones de estado límite. Además, en problemas asociados a la optimización en ingeniería independientemente de la fiabilidad, la mayoría de las variables son dependientes y con una gran cantidad de ecuaciones de igualdad, es decir, que el número de variables superbásicas es reducido.

9.5.1 The generalized reduced gradient method

En esta subsección se presenta el método del gradiente reducido generalizado (para más detalles véase VanderPlaats [169] o Bazaraa, Jarvis y Sherali [15]). Este método procede del método del gradiente reducido desarrollado por Wolfe [176] para la resolución de problemas no lineales con restricciones de igualdad lineales, que posteriormente fue extendido por Abadie y Carpentier [1] para implementar restricciones no lineales y que se consolidó finalmente tras el trabajo de Gabriele y Ragsdell [82].

Es un método similar al método del gradiente proyectado de Rosen [145], método recomendado por Liu y Der Kiureghian [112] para su empleo en problemas de fiabilidad. Trata de buscar una dirección de búsqueda tal, que las restricciones activas permanezcan así para un movimiento pequeño en esa dirección. Si debido a la no linealidad, alguna de las restricciones pasa a ser inactiva, se utiliza un método de Newton para regresar al contorno restringido.

Considérese el siguiente problema de programación matemática no lineal:

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) \\ & \mathbf{x} \end{array} \quad (9.64)$$

sometido a

$$g_j(\mathbf{x}) + x_{j+n} = 0; \quad j = 1, \dots, m \quad (9.65)$$

$$h_k(\mathbf{x}) = 0; \quad k = 1, \dots, \ell \quad (9.66)$$

$$x_i^{\text{lo}} \leq x_i \leq x_i^{\text{up}}; \quad i = 1, \dots, n \quad (9.67)$$

$$x_{j+n} \geq 0; \quad j = 1, \dots, m \quad (9.68)$$

en el que las restricciones (9.65) representan las desigualdades no lineales, transformadas a igualdades mediante variables de holgura positivas (restricción (9.68)). Las restricciones (9.66) son las de igualdad, y (9.67) son los límites de las variables. El número total de variables de diseño será de $n + m$.

La idea básica del método es tratar de identificar las variables dependientes asociadas a las restricciones de igualdad, con lo cual se puede reducir el número de variables independientes. De esta manera se puede generar un problema de optimización no restringida sujeto sólo a los límites de las variables dependientes e independientes. A las variables

independientes se las conoce como variables de decisión o superbásicas, mientras que a las dependientes se las conoce como variables de estado o no básicas.

Considerando de nuevo el problema (9.64)-(9.68), sólo con restricciones de igualdad. El vector \mathbf{x} , que contiene a las variables iniciales (n) más las de holgura (m) se divide en dos:

$$\mathbf{x} = \begin{pmatrix} \mathbf{z} \\ \mathbf{y} \end{pmatrix}; \quad \begin{array}{l} n - \ell \text{ variables independientes} \\ m + \ell \text{ variables dependientes} \end{array} \quad (9.69)$$

De momento no se discute el criterio para dividir las variables, y puesto que ahora todas las restricciones son de igualdad se va a agrupar (9.65) y (9.66) en una única restricción,

$$h_k(\mathbf{x}) = 0; \quad k = 1, \dots, m + \ell \quad (9.70)$$

y lo mismo con las variables (9.67) y (9.68) para obtener

$$x_i^{\text{lo}} \leq x_i \leq x_i^{\text{up}}; \quad i = 1, \dots, n + m \quad (9.71)$$

donde los límites superiores de las variables de holgura serán valores muy grandes (tienden a infinito).

El problema de optimización (9.64) y (9.68) queda:

$$\begin{array}{ll} \text{Minimizar} & f(\mathbf{x}) = f(\mathbf{z}, \mathbf{y}) \\ & \mathbf{z}, \mathbf{y} \end{array} \quad (9.72)$$

sometido a

$$h_k(\mathbf{x}) = 0; \quad k = 1, \dots, m + \ell \quad (9.73)$$

$$x_i^{\text{lo}} \leq x_i \leq x_i^{\text{up}}; \quad i = 1, \dots, n + m \quad (9.74)$$

Diferenciando la función objetivo (9.72) y las restricciones (9.73) resulta

$$df(\mathbf{x}) = \nabla_{\mathbf{z}}^T f(\mathbf{x}) d\mathbf{z} + \nabla_{\mathbf{y}}^T f(\mathbf{x}) d\mathbf{y} \quad (9.75)$$

$$dh_k(\mathbf{x}) = \nabla_{\mathbf{z}}^T h_k(\mathbf{x}) d\mathbf{z} + \nabla_{\mathbf{y}}^T h_k(\mathbf{x}) d\mathbf{y}; \quad k = 1, \dots, m + \ell \quad (9.76)$$

donde $\nabla_{\mathbf{z}}$ y $\nabla_{\mathbf{y}}$ son los gradientes con respecto a las variables independientes y dependientes, respectivamente.

Suponiendo que inicialmente se cumplen las restricciones de igualdad, para cualquier cambio en las variables superbásicas, éstas han de seguir activas para que se cumpla la factibilidad. Esto implica que $dh_k(\mathbf{x}) = 0; \quad k = 1, \dots, m + \ell$. La ecuación (9.76) puede escribirse en forma matricial como

$$d\mathbf{h}(\mathbf{x}) = \begin{pmatrix} \nabla_{\mathbf{z}}^T h_1(\mathbf{x}) \\ \nabla_{\mathbf{z}}^T h_2(\mathbf{x}) \\ \vdots \\ \nabla_{\mathbf{z}}^T h_{m+\ell}(\mathbf{x}) \end{pmatrix} d\mathbf{z} + \begin{pmatrix} \nabla_{\mathbf{y}}^T h_1(\mathbf{x}) \\ \nabla_{\mathbf{y}}^T h_2(\mathbf{x}) \\ \vdots \\ \nabla_{\mathbf{y}}^T h_{m+\ell}(\mathbf{x}) \end{pmatrix} d\mathbf{y} = \mathbf{A}d\mathbf{z} + \mathbf{B}d\mathbf{y} = 0 \quad (9.77)$$

Ahora para cualquier cambio en las variables independientes \mathbf{z} , como ha de cumplirse la condición (9.77), se pueden calcular los valores de las variables \mathbf{y} para mantener la factibilidad de las restricciones,

$$d\mathbf{y} = -\mathbf{B}^{-1}\mathbf{A}d\mathbf{z} \quad (9.78)$$

Sustituyendo la ecuación (9.78) en la ecuación (9.75), se tiene:

$$\begin{aligned} df(\mathbf{x}) &= \nabla_{\mathbf{z}}^T f(\mathbf{x})d\mathbf{z} - \nabla_{\mathbf{y}}^T f(\mathbf{x}) [\mathbf{B}^{-1}\mathbf{A}] d\mathbf{z} \\ &= \{ \nabla_{\mathbf{z}}^T f(\mathbf{x}) - \nabla_{\mathbf{y}}^T f(\mathbf{x}) [\mathbf{B}^{-1}\mathbf{A}] \} d\mathbf{z} \end{aligned} \quad (9.79)$$

de donde

$$\mathbf{G}_R = \frac{df(\mathbf{x})}{d\mathbf{z}} = \nabla_{\mathbf{z}} f(\mathbf{x}) - [\mathbf{B}^{-1}\mathbf{A}]^T \nabla_{\mathbf{y}} f(\mathbf{x}) \quad (9.80)$$

donde \mathbf{G}_R es el ‘gradiente reducido generalizado’. Éste puede ser utilizado para obtener una dirección nueva de búsqueda para las variables superbásicas:

$$\mathbf{z}^q = \mathbf{z}^{q-1} + \alpha \mathbf{G}_R; \quad d\mathbf{z} = \alpha \mathbf{G}_R \quad (9.81)$$

El problema es que para cada valor α las variables dependientes \mathbf{y} han de ser actualizadas mediante la ecuación (9.78), que es una aproximación lineal, con lo que es posible que con el nuevo valor de \mathbf{y} no se cumpla (9.77). Por tanto se debe encontrar un nuevo $d\mathbf{y}$ que cumpla la condición

$$h_k(\mathbf{x}) + dh_k(\mathbf{x}) = 0; \quad k = 1, \dots, m + \ell \quad (9.82)$$

donde $\mathbf{x} = (\mathbf{z} + d\mathbf{z}, \mathbf{y} + d\mathbf{y})^T$.

Es decir, se ha calcular el $d\mathbf{y}$ tal que

$$dh_k(\mathbf{x}) = -h_k(\mathbf{x}); \quad k = 1, \dots, m + \ell \quad (9.83)$$

Sustituyendo la ecuación (9.83) en la expresión (9.77) y despejando el $d\mathbf{y}$ se obtiene

$$d\mathbf{y} = \mathbf{B}^{-1} \{ \mathbf{h}(\mathbf{x}) - \mathbf{A}d\mathbf{z} \} \quad (9.84)$$

Por tanto las variables dependientes son actualizadas añadiendo $d\mathbf{y}$ a los valores previos, y es preciso evaluar de nuevo las restricciones. Hasta que todas las restricciones cumplan $h_k(\mathbf{x}) \leq \epsilon$; $k = 1, \dots, m + \ell$, donde ϵ es una tolerancia especificada. Nótese que este procedimiento no es más que el método de Newton para resolver sistemas no lineales, sin embargo se supone que la información del gradiente en \mathbf{A} y \mathbf{B} es constante, es decir que no se actualiza, con lo cual el método es similar a una iteración simple de punto fijo. Así, sin considerar como se separan las variables, y suponiendo calculadas las variables de holgura iniciales, el algoritmo queda de la siguiente manera:

Algorithm 9.2 (Método del gradiente reducido generalizado).

- **Entrada:** Función objetivo (9.64), restricciones (9.65)-(9.68), valor inicial de las variables $\mathbf{x}^{(0)}$ en la región factible y del parámetro α . Iniciar $i = 0$.
- **Salida:** Valor de las variables \mathbf{x}^* óptimo que satisface las condiciones de Karush-Kuhn-Tucker (definición ??).

Paso 1: Hacer $i = i + 1$ y calcular las matrices $\mathbf{A}^{(i-1)}$ y $\mathbf{B}^{(i-1)}$ mediante la fórmula (9.77) en el punto $\mathbf{x}^{(i-1)} = (\mathbf{z}^{(i-1)}, \mathbf{y}^{(i-1)})^T$.

Paso 2: Obtención de la dirección de búsqueda $\mathbf{G}_R^{(i-1)}$ mediante (9.80), si $\mathbf{G}_R^{(i-1)} = 0$ se ha localizado un punto de KKT, el procedimiento termina con la solución $\mathbf{x}^* = \mathbf{x}^{(i-1)}$ en caso contrario ir al paso siguiente.

Paso 3: Calcular el nuevo valor de las variables superbásicas $\mathbf{z}^{(i)} = \mathbf{z}^{(i-1)} + \alpha \mathbf{G}_R^{(i-1)}$

Paso 4: Actualización de las variables dependientes $\mathbf{y}^{(i)}$, con el procedimiento iterativo siguiente:

1. Iniciar $j = 0$.
2. Hacer $j = j + 1$, $\mathbf{x}^{(j-1)} = (\mathbf{z}^{(i)}, \mathbf{y}^{(i-1)})^T$.
3. Evaluación del nuevo incremento de las variables dependientes mediante la fórmula (9.84),

$$d\mathbf{y}^{(j)} = \mathbf{B}^{(i-1)^{-1}} \left\{ \mathbf{h}(\mathbf{x}^{(j-1)}) - \mathbf{A}^{(i-1)} d\mathbf{z} \right\}.$$
4. Establecer $\mathbf{y}^{(j)} = \mathbf{y}^{(j-1)} + d\mathbf{y}^{(j)}$ y $\mathbf{x}^{(j)} = (\mathbf{z}^{(i)}, \mathbf{y}^{(j)})^T$.
5. Evaluar $\mathbf{h}(\mathbf{x}^{(j)})$, si $h_k(\mathbf{x}^{(j)}) \leq \epsilon$; $\forall k = 1, \dots, m + \ell$ ir al **paso 1**, en caso contrario $j = j + 1$ e ir a 2.

■

Uno de los condicionantes importantes para el correcto funcionamiento del método es la división de las variables en dependientes e independientes, dado que la matriz asociada a las variables dependientes \mathbf{y} (\mathbf{B}), ha de ser no singular para poder calcular su inversa. Y además hay que procurar que las variables dependientes seleccionadas estén lejos de sus límites. Por tanto, se tiene que efectuar un preproceso para cumplir estos requerimientos. Para ello se va a partir de la matriz de gradientes,

$$\mathbf{Q} = \begin{pmatrix} \nabla_{\mathbf{x}}^T h_1(\mathbf{x}) \\ \nabla_{\mathbf{x}}^T h_2(\mathbf{x}) \\ \vdots \\ \nabla_{\mathbf{x}}^T h_{m+\ell}(\mathbf{x}) \end{pmatrix}_{(m+\ell) \times (n+m)} \quad (9.85)$$

como hay más columnas que filas, se quiere extraer una submatriz no singular que corresponderá a la matriz \mathbf{B} , para ello se usa una eliminación de Gauss con pivoteo partiendo de la ecuación

$$\mathbf{Q}\mathbf{X} = \mathbf{I} \quad (9.86)$$

donde \mathbf{I} es la matriz identidad. Si se efectúa el pivoteo, y se busca el máximo pivote parcial en cada fila, la columna seleccionada será la correspondiente a la variable dependiente seleccionada. Normalizando y pivotando con las filas inferiores tanto en la matriz \mathbf{Q} como en la \mathbf{I} , al final del proceso la transformada de la matriz identidad será \mathbf{B}^{-1} , y las columnas restantes de \mathbf{Q} contendrán $\mathbf{B}^{-1}\mathbf{A}$. Nótese que desde el punto de vista computacional, se puede almacenar la transformada de la identidad en las columnas que contienen al pivote.

Para evitar el tomar como variables dependientes las que están cerca de sus límites inferior o superior, se prescinde de utilizar éstas como columnas pivote. Aún así, esto no nos libra de que en función del valor de α seleccionado, las variables dependientes actualizadas sean infactibles. Habrá que controlar ese parámetro. Las últimas versiones de este método, utilizan una secuencia discreta de tamaños de paso, calculan en nuevo punto \mathbf{x} y el valor de la función objetivo $f(\mathbf{x})$, y mediante interpolación cuadrática calculan el tamaño de paso que conduce al valor mínimo de la función objetivo. Esta técnica hace que el algoritmo sea más fiable y robusto.

9.6 Alternative Decomposition Methods

In Chapter 9 several standard decomposition techniques have been described for nonlinear problems, for the cases of complicating constraints and variables. In this chapter, two other techniques are described and illustrative examples are given, including the very important engineering design method, which has an important practical interest.

Consider the problem:

$$\text{Minimize } c(\mathbf{x}, \mathbf{y}) \quad (9.87)$$

$$\mathbf{x}$$

subject to

$$\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{h}_0 \quad (9.88)$$

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \quad (9.89)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \quad (9.90)$$

where the function $\mathbf{h}(\mathbf{x}, \mathbf{y})$ cannot be easily evaluated. For example,

$$h_i(\mathbf{x}, \mathbf{y}) = \text{Minimum}_{\mathbf{u}, \mathbf{v}} \ell_i(\mathbf{x}, \mathbf{y}; \mathbf{u}, \mathbf{v}) \quad (9.91)$$

subject to

$$r_i(\mathbf{u}, \mathbf{v}) = k. \quad (9.92)$$

or perhaps $\mathbf{h}(\mathbf{x})$ is the output of a complicated finite element program. Note that constraint (9.95) can be both equality and inequality constraints.

The main difficulty of this problem is that the constraint (9.88) cannot be incorporated into standard optimization frameworks, so that decomposition techniques are required. Since two related optimization problems are solved, it is called a bi-level decomposition. In this section two different alternative methods for solving this type of problems are discussed.

9.6.1 Relaxation Method

Under some circumstances, the solution of the problem (9.87)-(9.90) can be obtained as the limit of the solutions of a sequence $\{P^{(n)} : n = 1, 2, \dots\}$ of problems, where $P^{(n)}$ consists of:

$$\begin{aligned} \text{Minimize } & c(\mathbf{x}, \mathbf{y}) \\ & \mathbf{x} \end{aligned} \quad (9.93)$$

subject to

$$\mathbf{r}(\mathbf{x}, \mathbf{y}) = \mathbf{r}_0^{(n)} \quad (9.94)$$

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \quad (9.95)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \quad (9.96)$$

where $\mathbf{h}(\mathbf{x}, \mathbf{y})$ has been replaced by $\mathbf{r}(\mathbf{x}, \mathbf{y})$, which is easily implementable in standard optimization algorithms, and a rule to obtain $\mathbf{r}_0^{(n)}$ is given:

$$\mathbf{r}_0^{(n+1)} = \mathbf{r}_0^{(n)} + \rho(\mathbf{h}_0 - \mathbf{h}^{(n)})$$

where $\mathbf{h}^{(n)}$ is the vector of solutions of the problems (9.91)-(9.92) for the optimal solution (\mathbf{x}, \mathbf{y}) of the problem $P^{(n)}$.

Alternatively, it could be used the dual variable $\boldsymbol{\lambda}$ associated with constraint (9.92) herewith Taylor series expansion to update the value $\mathbf{r}_0^{(n)}$:

$$\mathbf{h}_0 = \mathbf{h}^{(n)} + \frac{\partial \mathbf{h}^{(n)}}{\partial \mathbf{r}_0^{(n)}} \left(\mathbf{r}_0^{(n+1)} - \mathbf{r}_0^{(n)} \right) \quad (9.97)$$

$$\mathbf{h}_0 = \mathbf{h}^{(n)} + \boldsymbol{\lambda} \left(\mathbf{r}_0^{(n+1)} - \mathbf{r}_0^{(n)} \right) \quad (9.98)$$

thus

$$\mathbf{r}_0^{(n+1)} = \mathbf{r}_0^{(n)} + \frac{(\mathbf{h}_0 - \mathbf{h}^{(n)})}{\boldsymbol{\lambda}}. \quad (9.99)$$

Esto requiere la existencia de una correspondencia regular incremental $\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{q}(\mathbf{r}(\mathbf{x}, \mathbf{y}))$, entre $\mathbf{h}(\mathbf{x}, \mathbf{y})$ y $\mathbf{r}(\mathbf{x}, \mathbf{y})$, que se llamará $\mathbf{h} = \mathbf{q}(\mathbf{r})$.

This requires the existence of a regular one-to-one (increasing) unknown correspondence, $\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{q}(\mathbf{r}(\mathbf{x}, \mathbf{y}))$, between $\mathbf{h}(\mathbf{x}, \mathbf{y})$ and $\mathbf{r}(\mathbf{x}, \mathbf{y})$, which will be denoted $\mathbf{h} = \mathbf{q}(\mathbf{r})$.

More precisely, to solve the initial problem (9.87)-(9.90) we propose the algorithm below:

Algorithm 9.3 (Relaxation algorithm).

Input. The problem (9.87)-(9.90), a relaxation factor ρ , initial values \mathbf{r}_0 , and the error ϵ to check convergence.

Output. The solution of the problem (9.87)-(9.90) with the given error.

Step 0: Initialization. Initialize the iteration counter, $n = 1$, and fix the values of $\mathbf{r}_0^{(n)}$ for the first iteration to the initial values \mathbf{r}_0 .

Step 1: Master problem solution. Solve the master problem:

$$\text{Minimize}_{\mathbf{x}} c(\mathbf{x}, \mathbf{y}) \quad (9.100)$$

subject to

$$\mathbf{r}(\mathbf{x}, \mathbf{y}) = \mathbf{r}_0^{(n)} \quad (9.101)$$

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \quad (9.102)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \quad (9.103)$$

and obtain its optimal solution $(\mathbf{x}^*, \mathbf{y}^*)$.

Step 2: Subproblems solution. Obtain, for all i , the values

$$h_i^{(n)}(\mathbf{x}^*, \mathbf{y}^*) = \text{Minimum}_{\mathbf{u}, \mathbf{v}} \ell_i(\mathbf{x}^*, \mathbf{y}^*; \mathbf{u}, \mathbf{v}) \quad (9.104)$$

subject to

$$r_i(\mathbf{u}, \mathbf{v}) = k; \lambda_i \quad (9.105)$$

where λ_i is the dual variable related to constraint (9.105).

Step 3: Check convergence. If $\|\mathbf{h}^{(n)} - \mathbf{h}_0\| < \epsilon$ stop the procedure and output the solution. Otherwise, continue with Step 4.

Step 4: Update the \mathbf{r}_0 bounds. Use the formula:

$$\mathbf{r}_0^{(n+1)} = \mathbf{r}_0^{(n)} + \rho(\mathbf{h}_0 - \mathbf{h}^{(n)}),$$

or

$$\mathbf{r}_0^{(n+1)} = \mathbf{r}_0^{(n)} + \frac{(\mathbf{h}_0 - \mathbf{h}^{(n)})}{\lambda},$$

increase by one the iteration counter $n = n + 1$, and go to Step 1. ■

Theorem 9.1 (Convergence of Algorithm 9.3). *Under some regularity conditions, that guarantee the Taylor series expansion of $\mathbf{q}(\mathbf{r})$ and $\|\mathbf{I}_m - \rho \nabla \mathbf{q}(\mathbf{r}_+^{(n)})\| < 1$, the above algorithm leads to the solution of the problem (9.87)-(9.92).*

Proof. Assuming some regularity conditions, the convergence of the algorithm can be justified as follows.

$$\begin{aligned}
\mathbf{h}^{(n+1)} &= \mathbf{q}(\mathbf{r}_0^{(n+1)}) = \mathbf{q}(\mathbf{r}_0^{(n)} + \rho(\mathbf{h}_0 - \mathbf{h}^{(n)})) \\
&= \mathbf{q}(\mathbf{r}_0^{(n)}) + \rho \nabla \mathbf{q}(\mathbf{r}_+^{(n)})(\mathbf{h}_0 - \mathbf{h}^{(n)}) \\
&= \mathbf{h}^{(n)} + \rho \nabla \mathbf{q}(\mathbf{r}_+^{(n)})(\mathbf{h}_0 - \mathbf{h}^{(n)}) \\
&= (\mathbf{I}_m - \rho \nabla \mathbf{q}(\mathbf{r}_+^{(n)})) \mathbf{h}^{(n)} + \rho \nabla \mathbf{q}(\mathbf{r}_+^{(n)}) \mathbf{h}_0 \\
&= (\mathbf{I}_m - \rho \mathbf{Q}_n) \mathbf{h}^{(n)} + \rho \mathbf{Q}_n \mathbf{h}_0
\end{aligned}$$

where $\mathbf{r}_+^{(n)}$ is the usual intermediate point of the Taylor series, \mathbf{I}_m is the identity matrix whose dimension m is that of \mathbf{h} , and $\mathbf{Q}_n = \nabla \mathbf{q}(\mathbf{r}_+^{(n)})$. Then, we have:

$$\mathbf{h}^{(n)} = \left(\prod_{i=1}^{n-1} (\mathbf{I}_m - \rho \mathbf{Q}_i) \right) \mathbf{h}^{(1)} + \rho \left(\sum_{j=1}^{n-1} \mathbf{Q}_j \prod_{s=j+1}^{n-1} (\mathbf{I}_m - \rho \mathbf{Q}_s) \right) \mathbf{h}_0$$

which for adequate values of ρ and regularity conditions, including $\|\mathbf{I}_m - \rho \mathbf{Q}_i\| < 1$, converges to \mathbf{h}_0 . ■

Illustrative Example 9.2 (Relaxation method). Consider the problem:

$$\begin{aligned}
&\text{Minimize} && (x_1 + x_2 - 2)^2 + (x_3 + x_4 - 2)^2 && (9.106) \\
&x_1, x_2, x_3, x_4
\end{aligned}$$

subject to

$$h_1(\mathbf{x}, \mathbf{y}) = 1.2 \quad (9.107)$$

$$h_2(\mathbf{x}, \mathbf{y}) = 6 \quad (9.108)$$

where function $\mathbf{h}(\mathbf{x}, \mathbf{y})$ is defined as:

$$h_1(\mathbf{x}, \mathbf{y}) = \text{Minimum}_{u_1, u_2, u_3, u_4} \sum_{i=1}^4 (u_i - x_i)^2 \quad (9.109)$$

subject to

$$3u_1 + u_2 + 2u_3 + u_4 = 6 \quad (9.110)$$

and

$$h_2(\mathbf{x}, \mathbf{y}) = \text{Minimum}_{u_1, u_2, u_3, u_4} \sum_{i=1}^4 (u_i - x_i)^2 \quad (9.111)$$

subject to

$$u_1 + u_2 + u_3 + 2u_4 = 7 \quad (9.112)$$

Table 9.1: Procedimiento iterativo hasta que se obtiene la solución en el ejemplo ilustrativo 9.2.

n	x_1	x_2	x_3	x_4	r_1	r_2	h_1	h_2	c	$error$
1	0.50	1.50	2.00	0.00	7.56	6.36	0.07	1.29	0.00	23.49816
2	1.68	0.40	0.00	2.14	8.05	9.33	0.22	0.06	0.03	36.25063
3	1.03	1.64	0.00	3.33	8.46	11.88	0.38	0.90	2.21	26.64688
4	0.47	2.71	0.00	4.35	8.67	12.82	0.78	4.11	6.91	3.76708
5	0.29	3.08	0.00	4.73	8.42	12.72	1.71	6.21	9.31	0.30644
6	0.19	3.15	0.00	4.69	8.15	12.65	1.73	6.15	9.03	0.29960
7	0.08	3.25	0.00	4.66	7.84	12.55	1.82	6.20	8.83	0.42641
8	0.00	3.14	0.00	4.71	7.59	12.47	1.69	6.15	8.61	0.26849
9	0.00	2.72	0.00	4.88	7.56	12.48	1.27	5.98	8.79	0.00513
10	0.00	2.64	0.00	4.92	7.55	12.48	1.21	6.00	8.93	0.00022
11	0.00	2.63	0.00	4.93	7.55	12.48	1.20	6.00	8.95	0.00001
12	0.00	2.62	0.00	4.93	7.55	12.48	1.20	6.00	8.96	0.00000
13	0.00	2.62	0.00	4.93	7.55	12.48	1.20	6.00	8.96	0.00000

The solution of the problem (9.106)-(9.112) can be obtained as the limit of the solutions of the sequence $\{P^{(n)} : n = 1, 2, \dots\}$ of problems:

$$\text{Minimize}_{x_1, x_2, x_3, x_4} (x_1 + x_2 - 2)^2 + (x_3 + x_4 - 2)^2 \tag{9.113}$$

subject to

$$3x_1 + x_2 + 2x_3 + x_4 = r_1^{(n)} \tag{9.114}$$

$$x_1 + x_2 + x_3 + 2x_4 = r_2^{(n)} \tag{9.115}$$

The updating parameter $\mathbf{r}_0^{(n)}$ formula is:

$$\begin{pmatrix} r_1^{(n+1)} \\ r_2^{(n+1)} \end{pmatrix} = \begin{pmatrix} r_1^{(n)} \\ r_2^{(n)} \end{pmatrix} + 0.5 \begin{pmatrix} 1.2 - h_1^{(n)} \\ 6 - h_2^{(n)} \end{pmatrix}$$

La tabla 9.1 muestra la evolución del procedimiento iterativo en la que se pueden ver los valores de las variables y del error $\|\mathbf{h}_0 - \mathbf{h}^{(n)}\|$, para los datos siguientes:

$$\rho = 0.5; r_{01}^{(1)} = 7; r_{02}^{(1)} = 4; \epsilon = 0.00001.$$

■

Computational Example 9.1 (Aplicación de la técnica de relajación). Dado el problema

$$\text{Minimizar}_{\mathbf{x}} c(\mathbf{x}, \mathbf{y}) = \text{Minimizar}_{x_1, x_2} \left(\frac{x_1}{y_1} \right)^2 + \left(\frac{x_2}{y_2} \right)^2 \tag{9.116}$$

sujeto a

$$g_1(\mathbf{x}, \mathbf{y}) = \frac{x_1 x_2}{y_1 y_2} \geq g_1^0 \quad (9.117)$$

$$g_2(\mathbf{x}, \mathbf{y}) = \frac{x_2}{y_2} \sqrt{\frac{y_1}{x_1}} \geq g_2^0 \quad (9.118)$$

$$\mathbf{h}(\mathbf{x}, \mathbf{y}) \geq \mathbf{h}_0 \quad (9.119)$$

$$(9.120)$$

donde la función $\mathbf{h}(\mathbf{x}, \mathbf{y})$ es la solución del siguiente problema para $i = 1, 2$

$$h_i(\mathbf{x}, \mathbf{y}) = \underset{u_1, u_2, v_1, v_2}{\text{Mínimo}} \sqrt{\left(\frac{u_1 - x_1}{x_1 v_{x_1}}\right)^2 + \left(\frac{u_2 - x_2}{x_2 v_{x_2}}\right)^2 + \left(\frac{v_1 - y_1}{y_1 v_{y_1}}\right)^2 + \left(\frac{v_2 - y_2}{y_2 v_{y_2}}\right)^2} \quad (9.121)$$

sujeto a

$$g_i(\mathbf{u}, \mathbf{v}) = 1, \quad (9.122)$$

donde

$$g_1(\mathbf{u}, \mathbf{v}) = \frac{u_1 u_2}{v_1 v_2} \quad \text{y} \quad g_2(\mathbf{u}, \mathbf{v}) = \frac{u_2}{v_2} \sqrt{\frac{v_1}{u_1}}.$$

Suponiendo que los datos del problema son:

$$y_1 = 1.0; \quad y_2 = 1.0; \quad v_{x_1} = 0.02; \quad v_{x_2} = 0.02; \quad v_{y_1} = 0.1; \quad v_{y_2} = 0.1;$$

$$g_1^0 = 1.2; \quad g_2^0 = 1.6; \quad h_1^0 = 3.0; \quad h_2^0 = 4.0,$$

y tomando como valores iniciales de la variable $\mathbf{x} = (5.0, 5.0)^T$ se resuelve el problema mediante la actualización automática del parámetro $\mathbf{g}^{(n)}$ con la fórmula (9.99). De esta forma no hay que definir el coeficiente de relajación ρ . En la figura 9.3 se muestra la interpretación gráfica del problema y se muestra que la solución óptima se alcanza para los valores $\mathbf{x}^* = (0.951, 1.560)^T$, con un valor de la función objetivo de $f(\mathbf{x}^*, \mathbf{y}^*) = 3.340$. En la tabla 9.2 se muestra la evolución de método hasta alcanzar la solución del problema, se puede comprobar que converge en sólo 6 iteraciones.

Posteriormente se han realizado varias ejecuciones del mismo problema usando distintos valores del coeficiente de relajación $\rho = (0.05, 0.1, 0.2, 0.3, 0.4)$. Todos ellos convergen a la misma solución, excepto con $\rho = 0.4$. En la figura 9.4 se muestra la evolución del error con los distintos valores de ρ , de la cual se pueden extraer las siguientes conclusiones:

1. El procedimiento siempre converge a la solución si el coeficiente ρ es muy pequeño, aun así, lo hace de manera lenta para valores excesivamente pequeños.
2. Conforme aumentamos el valor de ρ , el método se comporta mejor convergiendo cada vez más rápido, hasta que se alcanza un valor óptimo ρ^* en el que el número de iteraciones es mínimo. En este ejemplo está en torno al $\rho^* \approx 0.2$.
3. Si se aumenta el valor de ρ por encima del valor óptimo, la convergencia empeora hasta que llega un momento ($\rho \approx 0.4$) en el que la solución comienza a oscilar y no converge.

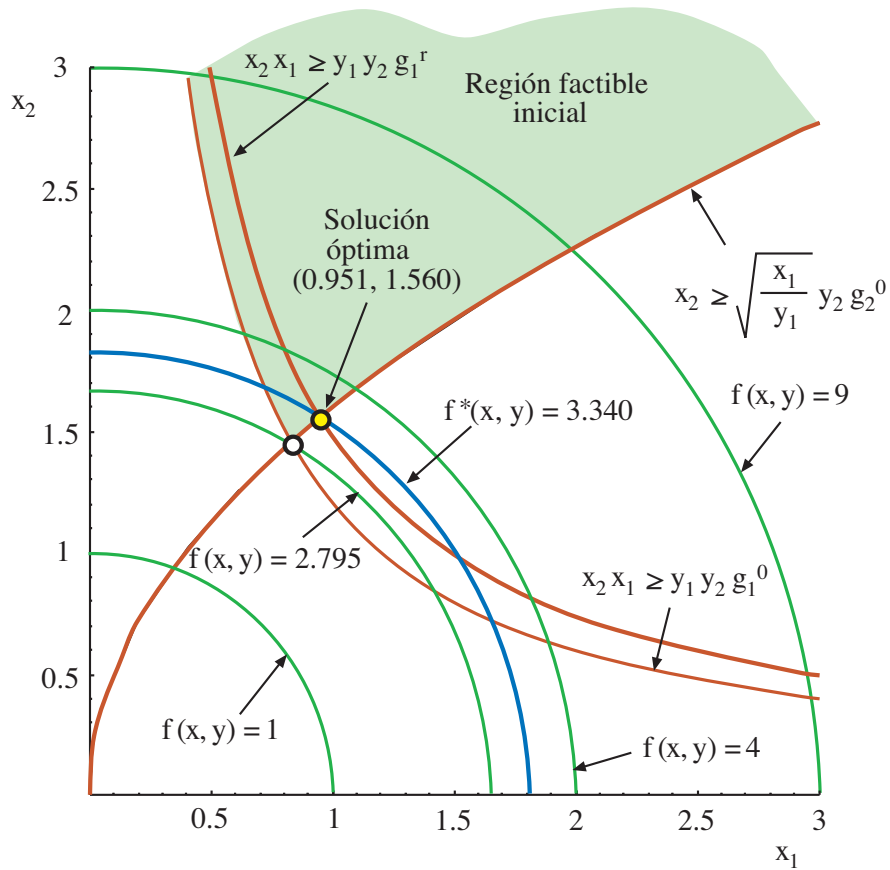


Figure 9.3: Ilustración gráfica del ejemplo computacional 9.1.

4. La resolución del problema mediante la fórmula de actualización (9.99) es aproximadamente equivalente (siempre mejor) a la actualización con el coeficiente de relajación óptimo ρ^* en términos de velocidad de convergencia. Pero sólo cuando se está cerca de la solución óptima, en ciertos casos, si se está lejos de la solución final la utilización de la fórmula puede provocar grandes variaciones en los valores actualizados que hacen que los modelos sean infactibles. Hay que ser muy cuidadoso utilizando ambas metodologías.
5. Los problemas que presentan ambos métodos de actualización pueden utilizarse para generar modelos mixtos que combinen las dos técnicas para que el parámetro inicial ρ actúe de factor de relajación de la variable dual λ actualizadora, de forma que se tome un parámetro actualizador combinado diferente para cada parámetro.

■

En el apéndice A.1 se presenta la implementación en GAMS del método de relajación.

Table 9.2: Ilustración del procedimiento iterativo para el ejemplo computacional 9.1.

n	$x_1^{(n)}$	$x_2^{(n)}$	$g_1^{(n)}$	$g_1^{(n)}$	$h_1^{(n)}$	$h_1^{(n)}$	$error^{(\nu)}$
1	0.825	1.454	1.200	1.600	1.319	4.419	1.2752713
2	0.943	1.554	1.466	1.600	2.895	4.419	0.5445521
3	0.951	1.560	1.484	1.600	2.998	4.419	0.0342347
4	0.951	1.560	1.484	1.600	3.000	4.419	0.0007747
5	0.951	1.560	1.484	1.600	3.000	4.419	0.0000155
6	0.951	1.560	1.484	1.600	3.000	4.419	0.0000003

9.6.2 Hyperplane reconstruction approximation

Considérese el mismo problema (9.87)-(9.90), o cualquier otro problema con restricciones de complicación. El principal inconveniente de esos problemas es que las restricciones de complicación no pueden incorporarse en los paquetes de optimización estándar, por lo que se necesita un procedimiento de descomposición.

Una posible solución alternativa al método anterior es la aproximación de la restricción en el problema global mediante hiperplanos aproximantes (Castillo et al. [?]) de la siguiente manera. Sea ν la iteración actual:

$$\underset{\mathbf{x}}{\text{Minimizar}} \quad c(\mathbf{x}, \mathbf{y}) \quad (9.123)$$

sujeto a

$$\mathbf{h}^{(s)} + \sum_{\forall i} \frac{\partial h_i^{(s)}}{\partial \mathbf{x}^{(s)}} (\mathbf{x} - \mathbf{x}^{(s)}) = \mathbf{h}_0; \quad s = 1, 2, \dots, \nu - 1 \quad (9.124)$$

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \quad (9.125)$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \quad (9.126)$$

de donde se obtiene el valor óptimo de $\mathbf{x}^{(\nu)}$, que se emplea para calcular el nuevo valor de la función $\mathbf{h}(\mathbf{x}, \mathbf{y})$

$$h_i^{(\nu)}(\mathbf{x}^{(\nu)}, \mathbf{y}) = \underset{\mathbf{u}, \mathbf{v}}{\text{Mínimo}} \quad \ell_i(\mathbf{x}, \mathbf{y}; \mathbf{u}, \mathbf{v}) \quad (9.127)$$

sujeto a

$$r_i(\mathbf{u}, \mathbf{v}) = k \quad (9.128)$$

$$\mathbf{x} = \mathbf{x}^{(\nu)} : \boldsymbol{\lambda}_i^{(\nu)} \quad (9.129)$$

donde $\boldsymbol{\lambda}_i^{(\nu)} = \partial h_i^{(\nu)} / \partial \mathbf{x}^{(\nu)}$ es un vector con las variables duales asociadas a la restricción 9.129, que fija los valores de las variables de diseño \mathbf{x} a los valores obtenidos del problema maestro (9.99)-(9.126), $\mathbf{x}^{(\nu)}$.

Algorithm 9.4 (Algoritmo de los hiperplanos aproximantes).

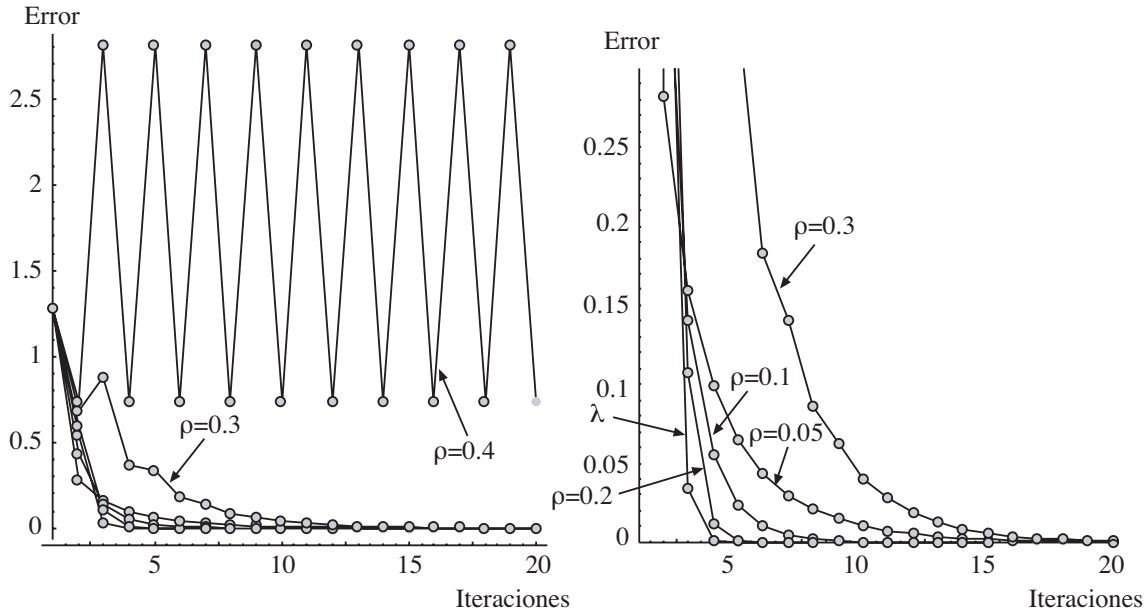


Figure 9.4: Evolución del error para distintos valores de coeficiente de relajación ρ y con el método de actualización dual λ .

Entrada. El problema (9.87)-(9.90), y la tolerancia ϵ para controlar la convergencia.

Salida. La solución del problema (9.87)-(9.90) con su error.

Paso 0: Inicialización. Iniciar el contador de iteraciones, $\nu = 1$.

Paso 1: Solución del problema maestro. Resolver el problema:

$$\text{Minimizar } c(\mathbf{x}, \mathbf{y}) \tag{9.130}$$

$$\mathbf{x}$$

sujeto a

$$\mathbf{h}^{(s)} + \sum_{\forall i} \frac{\partial h_i^{(s)}}{\partial \mathbf{x}^{(s)}} (\mathbf{x} - \mathbf{x}^{(s)}) = \mathbf{h}_0; \quad s = 1, 2, \dots, \nu - 1 \tag{9.131}$$

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \tag{9.132}$$

$$\mathbf{x}^{\text{lo}} \leq \mathbf{x} \leq \mathbf{x}^{\text{up}} \tag{9.133}$$

y obtener su solución $(\mathbf{x}^{(\nu)}, \mathbf{y}^{(\nu)})$. Inicialmente, cuando $\nu = 1$ la restricción (9.131) no contiene ningún hiperplano. Posteriormente se van añadiendo, así se va reconstruyendo la restricción de complicación mediante hiperplanos.

Paso 2: Solución de los subproblemas. Obtener, para todos los valores de i , los valores

$$h_i^{(\nu)}(\mathbf{x}^{(\nu)}, \mathbf{y}) = \underset{\mathbf{u}, \mathbf{v}}{\text{Mínimo}} \ell_i(\mathbf{x}, \mathbf{y}; \mathbf{u}, \mathbf{v}) \tag{9.134}$$

sujeto a

$$r_i(\mathbf{u}, \mathbf{v}) = k \quad (9.135)$$

$$\mathbf{x} = \mathbf{x}^{(\nu)} : \boldsymbol{\lambda}_i^{(\nu)} \quad (9.136)$$

donde $\boldsymbol{\lambda}_i^{(\nu)}$ es el vector dual asociado a la restricción (9.136).

Paso 3: Comprobación de la convergencia. Si $\|\mathbf{h}^{(\nu)} - \mathbf{h}_0\| < \epsilon$ parar el proceso, se ha alcanzado la solución. Si no, incrementar el contador de iteraciones en una unidad $\nu = \nu + 1$, e ir al paso 1. ■

Nótese, que la principal ventaja de este método con respecto al de relajación de la subsección 9.6.1 es que no necesita actualización de ningún parámetro, conforme se resuelven los subproblemas con distintos valores de las variables de diseño, se va reconstruyendo la restricción de complicación mediante hiperplanos que se incorporan en el problema maestro o global.

Computational Example 9.2 (Aplicación de la aproximación por hiperplanos).

A continuación vamos a resolver el mismo ejemplo del ejemplo computacional 9.1 con la aproximación por hiperplanos:

$$\underset{\mathbf{x}}{\text{Minimizar}} \quad c(\mathbf{x}, \mathbf{y}) = \underset{x_1, x_2}{\text{Minimizar}} \quad \left(\frac{x_1}{y_1}\right)^2 + \left(\frac{x_2}{y_2}\right)^2 \quad (9.137)$$

sujeto a

$$g_1(\mathbf{x}, \mathbf{y}) = \frac{x_1 x_2}{y_1 y_2} \geq g_1^0 \quad (9.138)$$

$$g_2(\mathbf{x}, \mathbf{y}) = \frac{x_2}{y_2} \sqrt{\frac{y_1}{x_1}} \geq g_2^0 \quad (9.139)$$

$$\mathbf{h}(\mathbf{x}, \mathbf{y}) \geq \mathbf{h}_0 \quad (9.140)$$

$$(9.141)$$

donde la función $\mathbf{h}(\mathbf{x}, \mathbf{y})$ es la solución del siguiente problema para $i = 1, 2$

$$h_i(\mathbf{x}, \mathbf{y}) = \underset{u_1, u_2, v_1, v_2}{\text{Mínimo}} \quad \sqrt{\left(\frac{u_1 - x_1}{x_1 v_{x_1}}\right)^2 + \left(\frac{u_2 - x_2}{x_2 v_{x_2}}\right)^2 + \left(\frac{v_1 - y_1}{y_1 v_{y_1}}\right)^2 + \left(\frac{v_2 - y_2}{y_2 v_{y_2}}\right)^2} \quad (9.142)$$

sujeto a

$$g_i(\mathbf{u}, \mathbf{v}) = 1, \quad (9.143)$$

donde

$$g_1(\mathbf{u}, \mathbf{v}) = \frac{u_1 u_2}{v_1 v_2} \quad y \quad g_2(\mathbf{u}, \mathbf{v}) = \frac{u_2}{v_2} \sqrt{\frac{v_1}{u_1}}.$$

Suponiendo que los datos del problema son los mismos:

$$y_1 = 1.0; \quad y_2 = 1.0; \quad v_{x_1} = 0.02; \quad v_{x_2} = 0.02; \quad v_{y_1} = 0.1; \quad v_{y_2} = 0.1;$$

Table 9.3: Ilustración del procedimiento iterativo para el ejemplo computacional 9.2.

ν	$x_1^{(\nu)}$	$x_2^{(\nu)}$	$g_1^{(\nu)}$	$g_2^{(\nu)}$	$h_1^{(\nu)}$	$h_2^{(\nu)}$	$error^{(\nu)}$
1	0.825	1.454	1.200	1.600	1.319	4.419	1.2752713
2	0.950	1.559	1.481	1.600	2.980	4.419	0.5576132
3	0.951	1.560	1.484	1.600	3.000	4.419	0.0065065
4	0.951	1.560	1.484	1.600	3.000	4.419	0.0000009

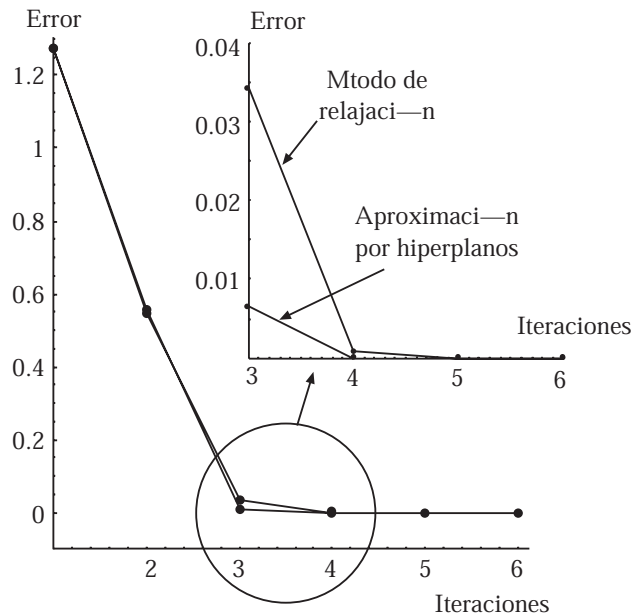


Figure 9.5: Evolución del error con el método de aproximación por hiperplanos y con el de relajación con el método de actualización dual λ .

$$g_1^0 = 1.2; \quad g_2^0 = 1.6; \quad h_1^0 = 3.0; \quad h_2^0 = 4.0,$$

y tomando como valores iniciales de la variable $\mathbf{x} = (5.0, 5.0)^T$. En la tabla 9.3 se muestra la evolución de método hasta alcanzar la solución del problema, que como era de esperar es la misma que la obtenida anteriormente con el método de relajación, $\mathbf{x}^* = (0.951, 1.560)^T$, con un valor de la función objetivo de $f(\mathbf{x}^*, \mathbf{y}^*) = 3.340$.

En este caso el método se comporta incluso mejor, convergiendo en cuatro iteraciones. Este comportamiento no es casual, y se puede concluir que este es el método que mejor comportamiento tiene. Al igual que el método de Newton para resolver raíces de ecuaciones tiene una convergencia cuadrática. En la figura 9.5 se ilustra una comparación de la evolución del error entre este método y el de relajación con actualización automática mediante variables duales, si bien el comportamiento es comparable, el primero funciona mejor y será preferible aplicarlo en todos los casos.



En el apéndice A.2 se presenta la implementación en GAMS del método de la aproximación por hiperplanos.

9.7 Sensitivity Analysis

9.7.1 Introduction

El análisis de sensibilidad es el estudio de como la variación de los parámetros que intervienen en un determinado modelo, afectan, cuantitativa o cualitativamente, a la solución del problema. El objetivo de los estudios de sensibilidad es determinar la dependencia del modelo a los datos que intervienen, con lo cual se aumenta la confianza del mismo y de sus predicciones, proporcionando el conocimiento de cómo afectan los cambios en los datos en la respuesta del modelo.

9.7.2 Sensitivity in constraint optimization. General method.

Todo problema de optimización genérico como el (??)-(??) se dice que ha alcanzado las condiciones de optimalidad cuando se cumplen las condiciones de Karush-Kuhn-Tucker (véase la definición ??) en el punto solución.

La sensibilidad de un parámetro se determina teniendo en cuenta que las condiciones de Karush-Kuhn-Tucker han de satisfacerse si se introduce una distorsión pequeña en el parámetro (véase Castillo et al. [?], Vanderplaats [169], Bazaraa, Sherali y Shetty [16] o Luenberger [?]).

Prescindiendo de un estudio más exhaustivo y riguroso sobre métodos estándar de sensibilidad se va a introducir un método general muy útil desde el punto de vista práctico. La idea es muy simple, asumiendo que se desea obtener la sensibilidad de la función objetivo a cambios en los datos, se convierten éstos en variables artificiales y se añaden restricciones que fijen los valores de estas variables a los valores de los parámetros, con lo cual las variables duales asociadas a esas restricciones nos darán la sensibilidad buscada.

De forma más precisa, considérese el siguiente problema de optimización

$$\text{Minimizar } f(\mathbf{x}, \boldsymbol{\eta}_0)$$

\mathbf{x}

sujeto a

$$\mathbf{h}(\mathbf{x}, \boldsymbol{\eta}_0) = \mathbf{0}$$

$$\mathbf{g}(\mathbf{x}, \boldsymbol{\eta}_0) \leq \mathbf{0}$$

donde $\boldsymbol{\eta}_0$ es el vector de datos o parámetros con respecto a los cuales se desea obtener la sensibilidad.

Para ello, se plantea el problema anterior de la siguiente manera:

$$\text{Minimizar } f(\mathbf{x}, \boldsymbol{\eta})$$

$\mathbf{x}, \boldsymbol{\eta}$

sujeto a

$$\mathbf{h}(\mathbf{x}, \boldsymbol{\eta}) = \mathbf{0}$$

$$\mathbf{g}(\mathbf{x}, \boldsymbol{\eta}) \leq \mathbf{0}$$

$$\boldsymbol{\eta} = \boldsymbol{\eta}_0 : \boldsymbol{\lambda}$$

donde los valores de las variables duales $\boldsymbol{\lambda}$ dan la sensibilidad o derivadas parciales de la función objetivo con respecto a los parámetros $\boldsymbol{\eta}_0$.

Un aspecto muy importante desde el punto de vista computacional es la utilización adecuada del método de sensibilidad. Dado que se van a introducir más variables, que complican la resolución de los modelos con las técnicas de optimización estándar, sólo se transformará el problema original una vez que se haya alcanzado la convergencia, de tal forma que el problema auxiliar con las variables ficticias se resuelve partiendo de la solución óptima, con lo cual la obtención de las sensibilidades está garantizada.

Illustrative Example 9.3 (Aplicación del método de sensibilidad al ejemplo computacional 9.1). Considérese el mismo problema que en el ejemplo computacional 9.1

$$\underset{\mathbf{x}}{\text{Minimizar}} \quad c(\mathbf{x}, \mathbf{y}) = \underset{x_1, x_2}{\text{Minimizar}} \quad \left(\frac{x_1}{y_1}\right)^2 + \left(\frac{x_2}{y_2}\right)^2 \quad (9.144)$$

sujeto a

$$g_1(\mathbf{x}, \mathbf{y}) = \frac{x_1 x_2}{y_1 y_2} \geq g_1^0 \quad (9.145)$$

$$g_2(\mathbf{x}, \mathbf{y}) = \frac{x_2}{y_2} \sqrt{\frac{y_1}{x_1}} \geq g_2^0 \quad (9.146)$$

$$\mathbf{h}(\mathbf{x}, \mathbf{y}) \geq \mathbf{h}_0 \quad (9.147)$$

$$(9.148)$$

donde la función $\mathbf{h}(\mathbf{x}, \mathbf{y})$ es la solución del siguiente problema para $i = 1, 2$

$$h_i(\mathbf{x}, \mathbf{y}) = \underset{u_1, u_2, v_1, v_2}{\text{Mínimo}} \quad \sqrt{\left(\frac{u_1 - x_1}{x_1 v_{x_1}}\right)^2 + \left(\frac{u_2 - x_2}{x_2 v_{x_2}}\right)^2 + \left(\frac{v_1 - y_1}{y_1 v_{y_1}}\right)^2 + \left(\frac{v_2 - y_2}{y_2 v_{y_2}}\right)^2} \quad (9.149)$$

sujeto a

$$g_i(\mathbf{u}, \mathbf{v}) = 1, \quad (9.150)$$

donde

$$g_1(\mathbf{u}, \mathbf{v}) = \frac{u_1 u_2}{v_1 v_2} \quad \text{y} \quad g_2(\mathbf{u}, \mathbf{v}) = \frac{u_2}{v_2} \sqrt{\frac{v_1}{u_1}}.$$

Supóngase que se quiere calcular el mayor número de sensibilidades de los parámetros del modelo, con lo cual se llega a un mayor entendimiento del mismo. Considerando los mismos valores numéricos que en el ejemplo 9.1 y resolviendo el problema mediante la aproximación por hiperplanos plantearíamos los siguientes problemas de sensibilidad en el óptimo \mathbf{x}^* :

Para el clásico:

$$\text{Minimizar}_{\mathbf{x}, \mathbf{y}^{\text{aux}}, \mathbf{cv}^{\text{aux}}} c(\mathbf{x}, \mathbf{y}^{\text{aux}}) = \text{Minimizar}_{x_1, x_2} \left(\frac{x_1}{y_1^{\text{aux}}} \right)^2 + \left(\frac{x_2}{y_2^{\text{aux}}} \right)^2 \quad (9.151)$$

sujeto a

$$g_1(\mathbf{x}, \mathbf{y}^{\text{aux}}) = \frac{x_1 x_2}{y_1^{\text{aux}} y_2^{\text{aux}}} \geq g_1^0 : \beta_1 \quad (9.152)$$

$$g_2(\mathbf{x}, \mathbf{y}^{\text{aux}}) = \frac{x_2}{y_2^{\text{aux}}} \sqrt{\frac{y_1^{\text{aux}}}{x_1}} \geq g_2^0 : \beta_2 \quad (9.153)$$

$$\mathbf{h}^{(s)} + \sum_{\forall i} \frac{\partial h_i}{\partial \mathbf{x}^*} (\mathbf{x} - \mathbf{x}^*) + \quad (9.154)$$

$$+ \sum_{\forall i} \frac{\partial h_i}{\partial \mathbf{y}} (\mathbf{y}^{\text{aux}} - \mathbf{y}) + \quad (9.155)$$

$$+ \sum_{\forall i} \frac{\partial h_i}{\partial \mathbf{cv}} (\mathbf{cv}^{\text{aux}} - \mathbf{cv}) \geq \mathbf{h}_0 : \boldsymbol{\theta}; s = 1, 2, \dots, \nu - 1 \quad (9.156)$$

$$\mathbf{cv}^{\text{aux}} = \mathbf{cv} : \boldsymbol{\phi} \quad (9.157)$$

$$\mathbf{y}^{\text{aux}} = \mathbf{y} : \boldsymbol{\psi} \quad (9.158)$$

con lo cual se obtienen las derivadas de la función objetivo del problema original β con respecto a los límites inferiores \mathbf{g}^0 , las derivadas parciales $\boldsymbol{\theta}$ respecto a los límites inferiores \mathbf{h}_0 , las derivadas parciales $\boldsymbol{\psi}$ con respecto a los datos \mathbf{y} y las sensibilidades $\boldsymbol{\phi}$ con respecto a los parámetros $\mathbf{cv} = (v_{x_1}, v_{x_2}, v_{y_1}, v_{y_2})^T$.

Para los subproblemas: $\forall i$:

$$\text{Mínimo}_{\mathbf{u}, \mathbf{v}, \mathbf{x}^{\text{aux}}, \mathbf{y}^{\text{aux}}, \mathbf{cv}^{\text{aux}}} \sqrt{\sum_{j=1}^2 \left(\frac{u_j - x_j^{\text{aux}}}{x_j^{\text{aux}} v_{x_j}^{\text{aux}}} \right)^2 + \sum_{j=1}^2 \left(\frac{v_j - y_j^{\text{aux}}}{y_j^{\text{aux}} v_{y_j}^{\text{aux}}} \right)^2} \quad (9.159)$$

sujeto a

$$g_i(\mathbf{u}, \mathbf{v}) = 1 \quad (9.160)$$

$$\mathbf{x}^{\text{aux}} = \mathbf{x}^* : \boldsymbol{\lambda}_i \quad (9.161)$$

$$\mathbf{y}^{\text{aux}} = \mathbf{y} : \boldsymbol{\psi}_i \quad (9.162)$$

$$\mathbf{cv}^{\text{aux}} = \mathbf{cv} : \boldsymbol{\chi}_i \quad (9.163)$$

donde $\boldsymbol{\lambda}_i$, $\boldsymbol{\psi}_i$ y $\boldsymbol{\chi}_i$ son las variables duales asociadas a las restricciones que fijan los valores auxiliares de las variables a los valores fijos de las variables objetivo y de los parámetros, respectivamente. Y que representan las sensibilidades de las funciones \mathbf{h} con respecto a los mismos

$$\boldsymbol{\lambda}_i = \frac{\partial h_i}{\partial \mathbf{x}^*}; \quad \boldsymbol{\psi}_i = \frac{\partial h_i}{\partial \mathbf{y}}; \quad \boldsymbol{\chi}_i = \frac{\partial h_i}{\partial \mathbf{cv}},$$

que se emplean en el clásico en la reproducción del hiperplano tangente a la función \mathbf{h} en el punto óptimo \mathbf{x}^* .

Table 9.4: Sensibilidades del ejemplo ilustrativo 9.3.

Parámetro	$\partial f(\mathbf{x}, \mathbf{y})$	$h_1(\mathbf{x}, \mathbf{y})$	$h_2(\mathbf{x}, \mathbf{y})$
∂x_1	--	8.711	-4.943
∂x_2	--	5.310	6.026
∂y_1	0.000	-8.286	4.701
∂y_2	0.000	-8.286	-9.403
∂v_{x_1}	1.436	-4.203	-1.922
∂v_{x_2}	1.436	-4.203	-8.085
∂v_{y_1}	4.836	-14.159	-19.623
∂v_{y_2}	4.836	-14.159	-22.567
∂g_1^0	0.000	--	--
∂q_2^0	1.275	--	--
∂h_1^0	0.342	--	--
∂h_2^0	0.000	--	--

En la tabla 9.4 se muestran los valores numéricos de las sensibilidades buscadas. Nótese la cantidad de información que proporciona la tabla, así por ejemplo, se puede comprobar que los únicos límites inferiores con sensibilidad distinta de cero son q_2^0 y h_1^0 , lo cual no es de extrañar porque de la solución del ejemplo 9.1 se puede comprobar que son las únicas restricciones activas, y por lo tanto las únicas que pueden influir en la solución con una pequeña perturbación de esos límites. ■

Pese haberse aplicado el estudio de sensibilidad usando la aproximación por hiperplanos, es posible obtener los mismos resultados mediante la resolución por el método de relajación.

Chapter 10

Reliability Based Optimization Techniques (RBOT)

10.1 Introduction

En el capítulo 4 se ha revisado el estado del arte de los métodos empleados en fiabilidad, y se ha llegado a la conclusión de que las metodologías más interesantes desde el punto de vista práctico eran las basadas en métodos FORM y SORM. Ambas aproximaciones requieren el cálculo previo del índice de fiabilidad y del punto de máxima verosimilitud, que se reducía a un problema de optimización.

Abdo y Rackwitz [2] y Liu y Der Kiureghian [112] presentaron sendos estudios de análisis del comportamiento de algoritmos de optimización aplicados a la fiabilidad estructural. Der Kiureghian y De Stafeno [56] presentaron un algoritmo basado en el métodos de gradiente que permite calcular el punto de diseño \mathbf{z}^* , el índice de fiabilidad β y además las curvaturas principales, con lo cual la aproximación FORM y SORM es inmediata usando los métodos de la sección 4.8.1. El algoritmo iterativo desarrollado por Rackwitz y Fiessler, véase Madsen, Krend y Lind [116], ha demostrado ser muy rápido y efectivo en el análisis FORM.

En este capítulo, no se van a discutir los diferentes métodos aplicables a problemas de fiabilidad, sino que se propondrá y analizará cómo se puede utilizar cualquiera de los paquetes de optimización estándar existentes en el mercado, para la resolución de problemas de optimización basados en técnicas de optimización, en particular en aquellos en los que el cálculo de la solución del problema (4.12)-(4.13) sea sólo una parte del objetivo del problema.

10.2 Reliability Based Optimization Problems

En la sección 4.9 se introdujo el estado del arte de los métodos de optimización aplicados a la optimización estructural. En esta sección se presentarán de forma más concreta los diferentes problemas tratados en esta tesis.

10.2.1 Variable classification

In the design and reliability analysis of a structure, there are some random variables (X_1, \dots, X_n) involved. They include geometric variables, material properties, loads, etc. Without loss of generality, we make no distinction between random and deterministic variables. So, it is assumed that all variables involved are random, and deterministic variables are only particular cases of them.

It is important to distinguish between design values of the random variables X_i , and actual values x_i ($i = 1, 2, \dots, n$). The design values are those values selected by the engineer at the design stage for the geometric variables (dimensions), the material properties (strengths, stiffness, etc.), that do not necessarily correspond with those in the real work. Thus, in this paper the design values are assumed to be the means or the characteristic values (extreme percentiles) of the corresponding random variables, and are denoted \bar{x}_i (mean) and \tilde{x}_i (characteristic), respectively. Some of these design values are chosen by the engineer or given by the design codes, and some (associated with the design variables) are selected by the optimization procedure to be presented. The set of variables (X_1, \dots, X_n) will be partitioned in four sets:

1. **Optimized design variables \mathbf{d} :** Their mean or characteristic values are to be chosen by the optimization procedure. Normally, they describe the dimensions of the work being designed, such as width, thickness, height, cross sections, etc., but can include material properties, etc.
2. **Non-optimized design variables $\boldsymbol{\eta}$ (parameters):** Their mean or characteristic values are fixed by the engineer or the code guidelines as input data to the optimization program. Some examples are costs, material properties (unit weights, strength, Young modulus, etc.), and other geometric dimensions of the work being designed (bridge length, platform width, etc.) that are fixed.
3. **Agents $\boldsymbol{\phi}$:** This set includes all the variables not controlled by the engineer, mainly actions acting on the structure (wind, surcharges, earthquake intensity, etc.) and experimental parameters. Agent is any entity, as wind, waves, earthquakes, etc., that can act on or significantly affect the safety, serviceability, and use of a structure and its environment. Their mean and/or characteristic values will be denoted $\tilde{\boldsymbol{\phi}}$.
4. **Random model parameters $\boldsymbol{\kappa}$:** Set of parameters used in the probabilistic design, defining the random spatial and temporal variability and dependence structure of the random variables involved (standard deviations, variation coefficients, correlations, covariance matrices, etc.).
5. **Dependent or non-basic variables $\boldsymbol{\psi}$:** Dependent variables which can be written in terms of the basic variables \mathbf{d} , $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ to facilitate the calculations and the statement of the problem constraints.

The corresponding means of \mathbf{d} will be denoted $\bar{\mathbf{d}}$, and the mean or the characteristic values of $\boldsymbol{\eta}$ is denoted $\tilde{\boldsymbol{\eta}}$.

10.2.2 Structural optimization problems subject to reliability constraints

The phases that the structure undergoes are: construction, useful life, maintenance and repair, dismantling, etc. Each phase has an associated duration. During each of these phases, the structure and the environment experiment a continuous sequence of outcomes, that have to be analyzed in the project. To simplify the analysis of the consequences of these outcomes, discrete time intervals, during which this behavior is described, characterized and evaluated by means of the project factors, are used. A project state associated with such behavior and period is a realization of the project factors, i.e., a fixed value for each of the project factors. This discretization of the problem in project states is a simplification of the actual stochastic process that allow describing, in a simple form, the structural and environmental behavior by deterministic values and statistical descriptors (project factors). During each of these periods, the shape, the structural response and the exploitation of the structure are assumed to be stationary processes.

The set of project states can be partitioned in three different subsets, that are associated with normal, extreme and accidental work and operating conditions (see Figure 10.1). The former conditions include the project states that habitually occur and for which the maritime structure offers the service it was designed for. The extreme conditions include the project states associated with the most severe actions due to project factors.

Limit states. The objective of the project design is to verify that the structure fulfills the project requirements in each and every one of the project states. For the sake of simplifying the verification process, only some of all the possible project states are verified, namely those representing limit situations of the structure with respect to stability, shape, use and exploitation. For this reason, these states are known as limit states.

As it was shown in Chapter 2, there are several types of limit states which can be classified in two different groups: (1) those affecting the safety and serviceability of the structure, that define the failure modes affecting the previously mentioned aspects of the structure; and (2) those associated with the use and exploitation of the structure, in which there is no structural failure, i.e., once the cause of the stoppage has disappeared, the structure totally recovers its use and exploitation requirements.

Failure modes. A mode describes the form or mechanism in which the failure or the operational stoppage of a part of the structure or one of its elements is produced. The modes that occur in a similar way or by the same mechanism will be assigned to the same structural or operational limit state (see Figure 10.2).

Verification equations. A verification equation is an equation defining a failure mode. It is necessary to establish a verification equation for each failure mode assigned to an ultimate or serviceability limit state and for each operational mode belonging to an operational limit state. Generally speaking, this equation is a state equation, and therefore, it is applied assuming that the outcomes of the set of project factors are stationary and uniform, from a statistical point of view.

Usually, there are two ways for establishing a verification equation: (a) the global or overall safety coefficient, and (b) the safety margin. Traditionally in engineering, the veri-

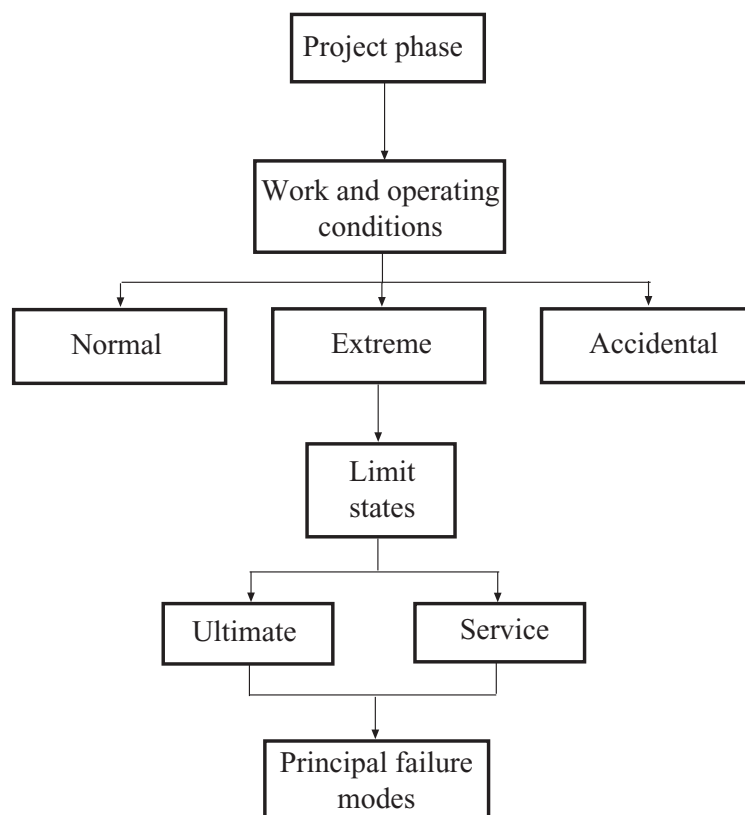


Figure 10.1: General methodology for computing the safety factor and the failure probability.

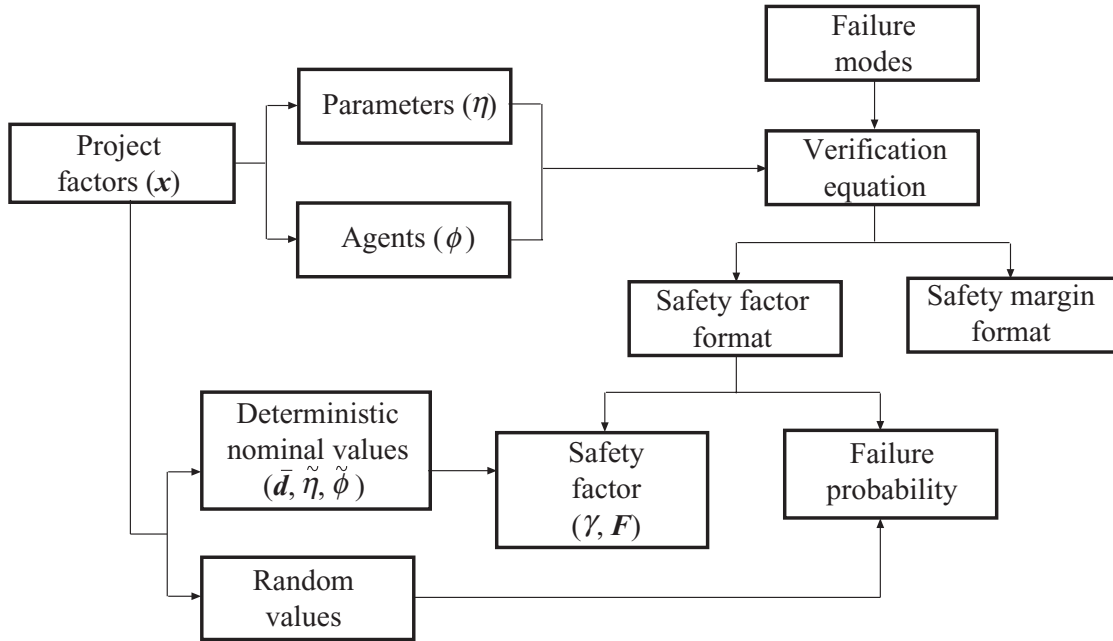


Figure 10.2: Variables, factors, project stages and reliability assessment methods.

Verification of the fulfillment of safety requirements of a structure in relation to a failure mode is carried out through the global safety coefficient, F , that is defined as the quotient of favorable terms and unfavorable terms. To consider that the structure or element has been favorably verified against a given failure mode, the corresponding safety coefficient should be greater than a given minimum value, i.e., $F > F_0$ (see Figure 10.2).

Deterministic and random design factors. The magnitude (and direction) of the project factors and consequently, the response and shape of the structure and its level of use and exploitation varies over time. A project factor can be assigned a nominal (deterministic) value, or a value based on a probability model or some other procedure, such as previous experience, predesign laboratory testing, etc. Thus, during the verification process the project factors can be treated as random or deterministic variables depending, among other things, on the verification method.

Optimal design. Engineering design of structural elements is a complicated and highly iterative process that usually requires an extensive experience. Iterations consist of a trial-and-error selection of the design variables or parameters, together with a check of the safety and functionality constraints, until reasonable structures, in terms of cost and safety, are obtained.

Over the last few years design methods have been improved by applying optimization techniques. The main advantage is that these techniques lead to optimal design and automation, i.e., the values of the design variables are provided by the optimization procedure (the

optimal values) and not fixed by the engineer. Designer's concerns are only the constraints to be imposed on the problem and the objective function to be optimized.

Engineering design can be done from three different points of view denoted here as (1) classic or deterministic, based on safety factors (DSO, 'Deterministic Structural Optimization'), (2) probabilistic, based on failure probabilities considering uncertainty models and random variables (RBSO, 'Reliability-Based Structural Optimization'), and (3) dual (classic-probabilistic), a method that combines the advantages of both design, the classic and the probabilistic (DRBSO, 'Dual Reliability-Based Structural Optimization').

Un aspecto muy importante antes de describir los distintos problemas tratados en esta tesis es que en la aproximación clásica se utilizan restricciones con coeficientes de seguridad donde las variables se asumen deterministas (valores nominales) e iguales a sus valores medios o característicos (percentiles extremos).

In this book problems dealing with reliability constraints associated with individual and time independent failure modes are deeply analyzed (see Frangopol [77], Murotsu et al. [127] and Sorensen [158]). Thus, the following cases are considered:

1. **Classical design using global safety factors.** The engineer minimizes a given objective function subject to some reliability constraints related to global safety factors associated with different failure modes (see Section 2.2.1), that is

$$\text{Minimize}_{\bar{\mathbf{d}}} c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \quad (10.1)$$

subject to

$$g_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_i^0) \geq 0; \forall i \in I \quad (10.2)$$

$$r_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \leq 0; \forall j \in J \quad (10.3)$$

$$\mathbf{h}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) = \boldsymbol{\psi} \quad (10.4)$$

$$\bar{\mathbf{d}}^{\text{lo}} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.5)$$

where $c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi})$ is the objective function (cost function), I is the set of failure modes, $g_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_i^0)$; $i \in I$ are the actual values of the global safety factors associated with the different failure modes I , J is the set of the geometric and code constraints, of the form $r_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \leq 0$, $\mathbf{h}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}})$ are the equations that allow obtaining the values of the intermediate variables $\boldsymbol{\psi}$ (they simplify the statement of the problem) and (10.5) are design variable bounds. Note that only the mean or characteristic values $\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}$ of the random variables $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ are used in this problem. The optimal solution $\bar{\mathbf{d}}$ obtained from the optimization problem is assumed to be the mean of the random variables \mathbf{d} . Note that these are the design values but not the real ones, that are assumed random with mean $\bar{\mathbf{d}}$, i.e., centered at this point.

Note that in the safety factor paradigm one minimizes the cost or expected cost subject to the safety factor and other geometric or code constraints.

2. **Classical design using partial safety factors.** Nowadays, most of the existing engineer codes are based on partial safety factors associated with the random variables

involved instead of failure modes (see Section 2.2.3). The new problem is as follows:

$$\text{Minimize}_{\bar{\mathbf{d}}} c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \quad (10.6)$$

subject to

$$g_i(\boldsymbol{\gamma}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \geq 0; \forall i \in I \quad (10.7)$$

$$r_j(\boldsymbol{\gamma}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \leq 0; \forall j \in J \quad (10.8)$$

$$\mathbf{h}(\boldsymbol{\gamma}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) = \boldsymbol{\psi} \quad (10.9)$$

$$\bar{\mathbf{d}}^{\text{lo}} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.10)$$

$$\boldsymbol{\gamma}^{\text{lo}} \leq \boldsymbol{\gamma} \leq \boldsymbol{\gamma}^{\text{up}} \quad (10.11)$$

where $\boldsymbol{\gamma}$ is the partial safety factor vector whose components are associated with the different random variables. Their values will be greater or lower than 1 depending on their status from the safety point of view, safe for the structure or unsafe, respectively.

3. **Probabilistic design.** Alternatively, probabilistic design minimize an objective function (cost, utility, etc.) subject to some reliability constraints based on probabilities of failure or reliability indexes (FORM), that is

$$\text{Minimize}_{\bar{\mathbf{d}}} c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \quad (10.12)$$

subject to

$$\beta_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) \geq \beta_i^0; \forall i \in I \quad (10.13)$$

$$r_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \leq 0; \forall j \in J \quad (10.14)$$

$$\mathbf{h}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) = \boldsymbol{\psi} \quad (10.15)$$

$$\bar{\mathbf{d}}^{\text{lo}} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.16)$$

where β_i is the reliability index related to failure mode i , and $\beta_i^0; i \in I$ are the allowed lower bounds for the corresponding β -values. Note that in this model classical constraints (10.2) and (10.7) are replaced by probabilistic constraints (10.13).

4. **Dual global design.** To incorporate the advantages of both the optimal classic design and the optimal probability design a new model that is the base of the probability of failure-safety factor (PFSF) method is presented (see [?, ?, 31, ?]). It includes global safety factors and probabilities of failure:

$$\text{Minimize}_{\bar{\mathbf{d}}} c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \quad (10.17)$$

subject to

$$g_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_i^0) \geq 0; \forall i \in I \quad (10.18)$$

$$\beta_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) \geq \beta_i^0; \forall i \in I \quad (10.19)$$

$$r_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \leq 0; \forall j \in J \quad (10.20)$$

$$\mathbf{h}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) = \boldsymbol{\psi} \quad (10.21)$$

$$\bar{\mathbf{d}}^{\text{lo}} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.22)$$

that is exactly the same problem as (10.1)-(10.5) but including the probability constraints (10.19).

5. **Dual partial design.** If partial safety factors are used instead (see Casas [27]) the problem can be stated as:

$$\text{Minimize}_{\bar{\mathbf{d}}} c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \quad (10.23)$$

subject to

$$g_i(\boldsymbol{\gamma}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \geq 0; \forall i \in I \quad (10.24)$$

$$\beta_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) \geq \beta_i^0; \forall i \in I \quad (10.25)$$

$$r_j(\boldsymbol{\gamma}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \leq 0; \forall j \in J \quad (10.26)$$

$$\mathbf{h}(\boldsymbol{\gamma}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) = \boldsymbol{\psi} \quad (10.27)$$

$$\bar{\mathbf{d}}^{\text{lo}} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.28)$$

$$\boldsymbol{\gamma}^{\text{lo}} \leq \boldsymbol{\gamma} \leq \boldsymbol{\gamma}^{\text{up}} \quad (10.29)$$

that is exactly the same problem as (10.6)-(10.11) but including the probability constraints (10.25).

6. **Global reliability based design.** The most general problem, including global safety factors, that is going to be deal with in this book is minimizing the expected total cost of a structure, including maintenance and construction, as follows:

$$\text{Minimize}_{\bar{\mathbf{d}}} C_{to}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) = C_{co}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) + C_{\beta}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) \quad (10.30)$$

subject to

$$g_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_i^0) \geq 0; \forall i \in I \quad (10.31)$$

$$\beta_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) \geq \beta_i^0; \forall i \in I \quad (10.32)$$

$$r_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \leq 0; \forall j \in J \quad (10.33)$$

$$\mathbf{h}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) = \boldsymbol{\psi} \quad (10.34)$$

$$\bar{\mathbf{d}}^{\text{lo}} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.35)$$

where C_{to} is the total expected cost during the lifetime of the structure, C_{co} is the initial/construction cost and C_{β} is the cost associated to the probability of failure (maintenance, repair, insurance, etc.).

7. **Partial reliability based design.** If in the previous model global safety factors are replaced by partial safety coefficients:

$$\text{Minimize}_{\bar{\mathbf{d}}} C_{to}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) = C_{co}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) + C_{\beta}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) \quad (10.36)$$

subject to

$$g_i(\gamma, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \psi) \geq 0; \forall i \in I \quad (10.37)$$

$$\beta_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \psi, \boldsymbol{\kappa}) \geq \beta_i^0; \forall i \in I \quad (10.38)$$

$$r_j(\gamma, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \psi) \leq 0; \forall j \in J \quad (10.39)$$

$$\mathbf{h}(\gamma, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) = \boldsymbol{\psi} \quad (10.40)$$

$$\bar{\mathbf{d}}^{\text{lo}} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.41)$$

$$\gamma^{\text{lo}} \leq \gamma \leq \gamma^{\text{up}} \quad (10.42)$$

Note that the last two models are the most general ones, so that it is possible to get the other ones as particular cases. For example, Model 1 (classical design with global safety factors) can be obtained from Model 6 (global reliability based design) eliminating constraint (10.32) and the cost associated with the failure probability $C_{\boldsymbol{\beta}}$.

Unfortunately, the model with probability constraints or cost functions dependent on reliability indexes $\boldsymbol{\beta}$ cannot be solved directly because each of the constraints (10.13), (10.19), (10.25), (10.32) and (10.38) involve a complicated integral or another optimization problem associated with each failure mode, i.e.:

$$\beta_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \psi, \boldsymbol{\kappa}) = \underset{\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}}{\text{Minimum}} \beta_i(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \psi, \boldsymbol{\kappa}) \quad (10.43)$$

subject to

$$g_i(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \psi; 1) = 0 \quad (10.44)$$

$$\mathbf{h}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}) = \boldsymbol{\psi} \quad (10.45)$$

Replacing $\beta_i(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \psi, \boldsymbol{\kappa})$ by its corresponding value given in equation(4.12), problem (10.43)-(10.45) becomes:

$$\beta_i = \underset{\mathbf{z}}{\text{Minimum}} \sqrt{\mathbf{z}^T \mathbf{z}} \quad (10.46)$$

subject to

$$g_i(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \psi; 1) = 0 \quad (10.47)$$

$$\mathbf{h}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}) = \boldsymbol{\psi} \quad (10.48)$$

$$\mathbf{T}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\kappa}) = \mathbf{z} \quad (10.49)$$

where $g_i(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \psi) = 1$ is the limit state equation (strict failure), that it is the same independently if global or partial safety factors are used (reliability index invariance safety measure property), and equation (10.49) is the transformation (Rosenblatt (4.59) or Nataf [128]) leading to the standard unit normal \mathbf{Z} variables used in FORM. Note that variables $\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}$ are realizations of the corresponding random variables, and that the optimal point obtained from problem (10.46)-(10.49), $\mathbf{d}^*, \boldsymbol{\eta}^*, \boldsymbol{\phi}^*$ is the point of maximum likelihood or design point for the corresponding failure mode.

As it has been shown, the reliability-based optimization problems can be characterized as bi-level optimization problems (see Mínguez [125] and Mínguez et al. [126]):

Level 1: The upper level is the overall optimization in the design variables \mathbf{d} , in this case the minimization of the objective function (master problem).

Level 2: The lower level is the reliability estimation in the \mathbf{z} variables necessary for evaluating the probability of failure or the reliability index solving problem (10.46)-(10.49) (subproblems).

Decomposition procedures for solving this bi-level structure are shown in Chapters 9 and ??.

10.3 Reliability based optimization of a rubblemound breakwater using Benders decomposition

In this Chapter several models related to Reliability-Based Structural Optimization (RBSO) has been introduced. The problems where the cost function include cost associated with the probability of failure can be solved using decomposition techniques (see Benders [17] and Geoffrion [86]). This methods were applied to reliability optimization problems by Mínguez [125], and Mínguez et al. [126]. The price that has to be paid for such a simplification is iteration. That is, instead of solving the original problem with complicating constraints at once, two simpler problems are solved iteratively: a simple problem called master problem (approximation of the original one) and the subproblem or subproblems (inner optimization), so that complicating constraints are being progressively taken into account into the master problem. In fact, the expected cost function is approximated by an increasing number of hyperplanes. For a detailed analysis of decomposition techniques see Conejo et al. [?].

To show the applicability of the method an example based on the ‘Global reliability based design’ model (10.30)-(10.35) but eliminating the constraints related to reliability indexes (10.32) and safety factors (10.32). Subsequently, additional methods for dealing with constraints (10.32) (called complicating constraints) are introduced.

Consider the construction of a rubblemound breakwater (see Figure 10.3) to protect a harbor area from high waves during a storm. Our goal is an optimal design of the breakwater based on minimizing the construction and the insurance costs against overtopping damage of the internal structures and ships.

The construction cost is

$$C_{co} = c_c v_c + c_a c_a$$

where v_c and c_a are the concrete and armor volumes, respectively, and c_c and c_a are the respective construction costs per unit volume.

For the sake of simplicity the insurance cost is evaluated considering the probability of overtopping failure, P_f^D , during the breakwater lifetime, D . To carry out a rigorous analysis, not only the probability of failure but how much water exceeds the freeboard level should be considered as well. Thus, the insurance cost is

$$C_{in} = 5000 + 1.25 \times 10^6 P_f^D{}^2$$

where the numerical constants are typical values.

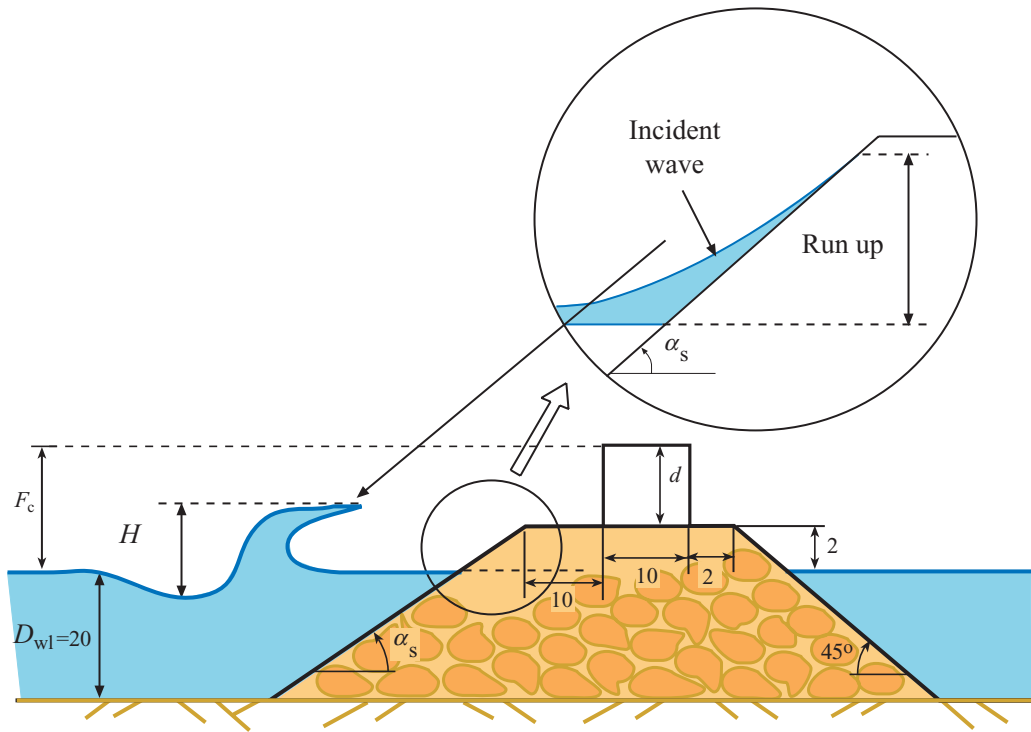


Figure 10.3: Parameterized rubblemound breakwater used in the example.

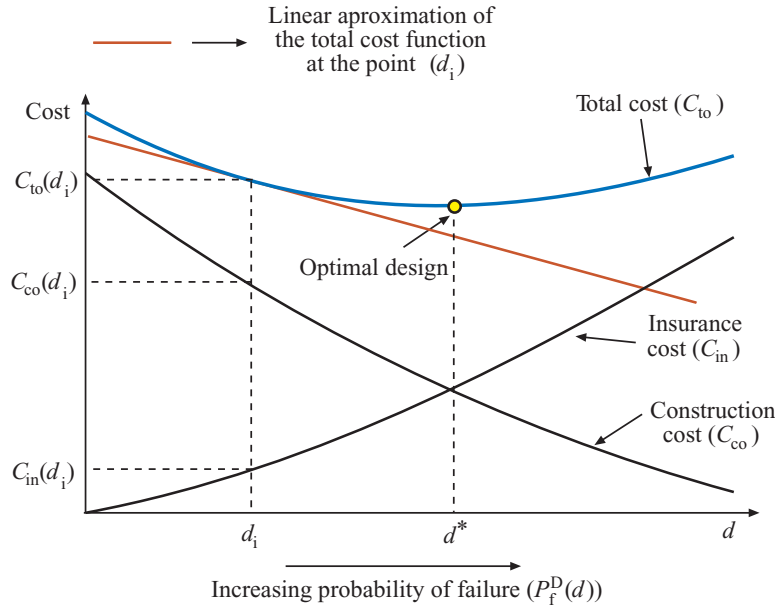


Figure 10.4: Graphical illustration of the cost functions for the breakwater design.

The construction, insurance and total cost versus the probability of failure are shown in Figure 10.4. Note the decreasing and increasing character of the construction and insurance costs, respectively, when the failure probability increases, and the convex character of the total cost.

For a rubblemound breakwater of slope $\tan \alpha_s$ and freeboard F_c , (see Figure 10.3), and a given wave of height H and period T , the volume of water that overtops the structure can be estimated from the volume of water that would rise over the extension of the slope exceeding the freeboard level. With this approximation, overtopping (failure) occurs whenever the difference between the maximum excursion of water over the slope, R_u , called wave run-up, exceeds the freeboard F_c . Thus, overtopping failure occurs if

$$F_c - R_u < 0. \tag{10.50}$$

Based on experiments, the following equation has been proposed to evaluate the dimensionless quantity R_u/H :

$$\frac{R_u}{H} = A_u (1 - e^{B_u I_r})$$

where A_u and B_u are given coefficients that depend on the armor units and I_r is the Iribarren number

$$I_r = \frac{\tan \alpha_s}{\sqrt{H/L}}$$

where α_s is the seaside slope angle and L is the wave length, obtained from the dispersion equation

$$\left(\frac{2\pi}{T}\right)^2 = g \frac{2\pi}{L} \tanh \frac{2\pi D_{wl}}{L}$$

where D_{wl} is the design water level and g is the gravity constant.

In addition, due to construction reasons the slope α_s is limited by:

$$1/5 \leq \tan \alpha_s \leq 1/2$$

The set of variables and parameters involved in this problem, as in the wall design problem, can be partitioned into the following subsets (see Section 10.2.1):

d: *Optimization design variables.* The design random variables which values are to be chosen by the optimization program to optimize the objective function (minimize the cost):

$$\mathbf{d} = \{F_c, \tan \alpha_s\}$$

η : *Non-optimization design variables (Parameters).* The set of variables and parameters which mean or characteristic values are fixed by the engineer or the code and must be given as data to the optimization program:

$$\boldsymbol{\eta} = \{A_u, B_u, D_{wl}, g, c_c, c_a\}$$

ϕ : *Agents.* Agent is any entity, as wind, waves, earthquakes, etc., that can act on or significantly affect the safety, serviceability, and use of a structure and its environment. In this case the ones associated with swell

$$\boldsymbol{\phi} = \{H, T\}$$

κ : *Random model parameters.* The set of parameters used in the probabilistic design, defining the random variability and dependence structure of the variables involved, in this case sea state descriptors:

$$\boldsymbol{\kappa} = \{H_s, \bar{T}, d_{st}\}$$

where H_s is the significant wave height, \bar{T} is the average value of the period of the sea waves and d_{st} is the sea state duration.

ψ : *Auxiliary or non-basic variables.* The auxiliary variables which values can be obtained from the basic variables \mathbf{d} , $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$, using some formulas:

$$\boldsymbol{\psi} = \{I_r, c_a, v_c, C_{co}, C_{in}, R_u, L, d\}$$

The basic random variables in this problem are H and T , that are assumed to be independent and with cumulative distribution functions:

$$F_H(H) = 1 - e^{-2(H/H_s)^2}; \quad H \geq 0 \quad (10.51)$$

and

$$F_T(T) = 1 - e^{-0.675(T/\bar{T})^4}; \quad T \geq 0 \quad (10.52)$$

If P_f is the probability of overtopping failure due to a single wave, the lifetime breakwater failure probability becomes

$$P_f^D(\mathbf{d}) = 1 - (1 - P_f(\mathbf{d}))^N \quad (10.53)$$

where $N = d_{st}/\bar{T}$ is the mean number of waves during the design sea state for period D , and d_{st} is its duration. Note that the waves in the sea state are considered independent.

Then, the design problem consists of:

$$\begin{aligned} \text{Minimize } & C_{to} = c_c v_c + c_a v_a + 5000 + 1.25 \times 10^6 P_f^{D^2} \\ & F_c, \tan \alpha_s \end{aligned}$$

subject to

$$\begin{aligned} 1/5 & \leq \tan \alpha_s \leq 1/2 \\ F_c & = 2 + d \\ v_c & = 10d \\ v_a & = \frac{1}{2}(D_{wl} + 2) \left(46 + D_{wl} + \frac{D_{wl} + 2}{\tan \alpha_s} \right) \\ P_f(\mathbf{d}) & = \Phi(-\beta) \\ P_f^D(\mathbf{d}) & = 1 - (1 - P_f(\mathbf{d}))^{(d_{st}/\bar{T})} \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable and β is the reliability index that cannot be obtained directly and involve another optimization problem

$$\text{Minimize } \beta = \sqrt{z_1^2 + z_2^2} \\ H, T$$

subject to

$$\begin{aligned} \frac{R_u}{H} & = A_u (1 - e^{B_u I_r}) \\ I_r & = \frac{\tan \alpha_s}{\sqrt{H/L}} \\ \left(\frac{2\pi}{T} \right)^2 & = g \frac{2\pi}{L} \tanh \frac{2\pi D_{wl}}{L} \\ \Phi(z_1) & = 1 - e^{-2(H/H_s)^2} \\ \Phi(z_2) & = 1 - e^{-0.675(T/\bar{T})^4} \\ F_c & = R_u \end{aligned}$$

where z_1 and z_2 are independent standard normal random variables.

As it has been shown, the reliability-based optimization problems can be characterized as bi-level optimization problems (see Mínguez [125] and Mínguez et al. [126]). The upper level is the overall optimization in the design variables \mathbf{d} , in this case the minimization of the total cost function. The lower level is the reliability estimation in the \mathbf{z} variables necessary for evaluating the insurance cost. This bi-level structure is the reason why the decomposition procedure is needed.

In this Section we give a possible way of solving the corresponding optimization problem using Benders decomposition. We consider as complicating variables the design variables \mathbf{d} , our goal is the hyperplane reconstruction of the total cost function $C_{\text{to}}(\mathbf{d})$. More precisely, the method proceeds as follows:

Step 0: Initialization. Select initial values for the design variables (complicating variables) $\mathbf{d}^{(0)} = \{F_c^{(0)}, \tan \alpha_s^{(0)}\}$.

Set the iteration counter to $\nu = 1$, $\mathbf{d}^{(\nu)} = \mathbf{d}^{(0)}$, and the total cost lower bound to $C_{\text{down}}^{(\nu)} = -\infty$.

Step 1: Evaluating the reliability index β (subproblem solution). Based on the actual design values (complicating variables), the reliability index is calculated solving the problem

$$\text{Minimize}_{H, T} \quad \beta^{(\nu)} = \sqrt{z_1^2 + z_2^2}$$

subject to

$$\begin{aligned} \frac{R_u}{H} &= A_u (1 - e^{B_u I_r}) \\ I_r &= \frac{\tan \alpha_s}{\sqrt{H/L}} \\ \left(\frac{2\pi}{T}\right)^2 &= g \frac{2\pi}{L} \tanh \frac{2\pi D_{wl}}{L} \\ \Phi(z_1) &= 1 - \exp(-2(H/H_s)^2) \\ \Phi(z_2) &= 1 - \exp(-0.675(T/\bar{T})^4) \\ R_u &= F_c \\ \mathbf{d} &= \mathbf{d}^{(\nu)} : \boldsymbol{\lambda}^{(\nu)} \end{aligned}$$

The solution of this subproblem provides $\beta^{(\nu)}$, and the partial derivatives of the reliability index with respect the design variables $\boldsymbol{\lambda}^{(\nu)}$.

Then, it is possible to evaluate the probability of failure of one wave (P_f), the probability of failure in the design sea state (P_f^D), the volumes of material (v_c , v_a), the construction cost C_{co} , the insurance cost C_{in} and the total cost $C_{\text{to}}^{(\nu)}$ (see Figure 10.5) for the actual values of the complicating (design) variables $\mathbf{d}^{(\nu)}$ and the reliability index $\beta^{(\nu)}$ as

$$\begin{aligned} v_c &= 10h \\ v_a &= \frac{1}{2}(D_{wl} + 2)(46 + D_{wl} + \frac{(D_{wl} + 2)}{\tan \alpha_s^{(\nu)}}) \\ F_c^{(\nu)} &= 2 + h \\ P_f &= \Phi(-\beta^{(\nu)}) \\ P_f^D &= 1 - (1 - P_f)^{d_{\text{st}}/\bar{T}} \\ C_{\text{co}} &= c_c v_c + c_a v_a \end{aligned}$$

$$\begin{aligned} C_{\text{in}} &= 5000 + 1.25 \times 10^6 P_f^D \\ C_{\text{to}}^{(\nu)} &= C_{\text{co}} + C_{\text{in}} \end{aligned}$$

Thus, we have just calculated one point of the total cost function, for reconstructing the function using hyperplanes, we need the partial derivatives of function C_{to} with respect the design variables $\mathbf{d}^{(\nu)}$, $\boldsymbol{\Omega}^{(\nu)}$ shown in Figure 10.5. These partial derivatives can be calculated using the following expression:

$$\boldsymbol{\Omega}^{(\nu)} = \frac{\partial C_{\text{to}}^{(\nu)}}{\partial \mathbf{d}^{(\nu)}} = \frac{\partial C_{\text{co}}}{\partial \mathbf{d}^{(\nu)}} + \frac{\partial C_{\text{in}}}{\partial \mathbf{d}^{(\nu)}}$$

where $\frac{\partial C_{\text{co}}}{\partial \mathbf{d}^{(\nu)}}$ is the partial derivative of the construction cost, given in analytical form or calculated using the auxiliary problem

$$\begin{aligned} \text{Minimize } & C_{\text{co}} = c_c v_c + c_a v_a \\ & F_c, \tan \alpha \end{aligned}$$

subject to

$$\begin{aligned} v_c &= 10h \\ v_a &= \frac{1}{2}(D_{\text{wl}} + 2)(46 + D_{\text{wl}} + \frac{D_{\text{wl}} + 2}{\tan \alpha_s}) \\ F_c &= 2 + h \\ \mathbf{d} &= \mathbf{d}^{(\nu)} : \boldsymbol{\theta}^{(\nu)} \end{aligned}$$

where $\boldsymbol{\theta}^{(\nu)}$ are the required derivatives.

The partial derivative of the insurance cost is obtained as follows

$$\begin{aligned} \frac{\partial C_{\text{in}}}{\partial \mathbf{d}^{(\nu)}} &= \frac{\partial C_{\text{in}}}{\partial P_f^D} \frac{\partial P_f^D}{\partial \mathbf{d}^{(\nu)}} = \\ &= -\frac{\partial C_{\text{in}}}{\partial P_f^D} \frac{d_{\text{st}}}{\bar{T}} (1 - P_f)^{(d_{\text{st}}/\bar{T}-1)} \frac{\partial P_f}{\partial \mathbf{d}^{(\nu)}} = \\ &= \frac{\partial C_{\text{in}}}{\partial P_f^D} \frac{d_{\text{st}}}{\bar{T}} (1 - P_f)^{(d_{\text{st}}/\bar{T}-1)} \frac{\exp(-\beta^2/2)}{\sqrt{2\pi}} \frac{\partial \beta}{\partial \mathbf{d}^{(\nu)}} \end{aligned}$$

where for this example

$$\frac{\partial C_{\text{in}}}{\partial P_f^D} = 2.5 \times 10^6 P_f^D, \text{ and } \frac{\partial \beta}{\partial \mathbf{d}^{(\nu)}} = \boldsymbol{\lambda}^{(\nu)}$$

Set the total cost upper bound to $C_{\text{up}}^{(\nu)} = C_{\text{to}}^{(\nu)}$.

Step 2: Convergence check. If $|C_{\text{up}}^{(\nu)} - C_{\text{down}}^{(\nu)}|/|C_{\text{up}}^{(\nu)}| \leq \varepsilon$, the solution with a level of accuracy ε of the objective function and design variables is

$$C_{\text{to}}^* = C_{\text{to}}^{(\nu)}; \quad \mathbf{d}^* = \mathbf{d}^{(\nu)}$$

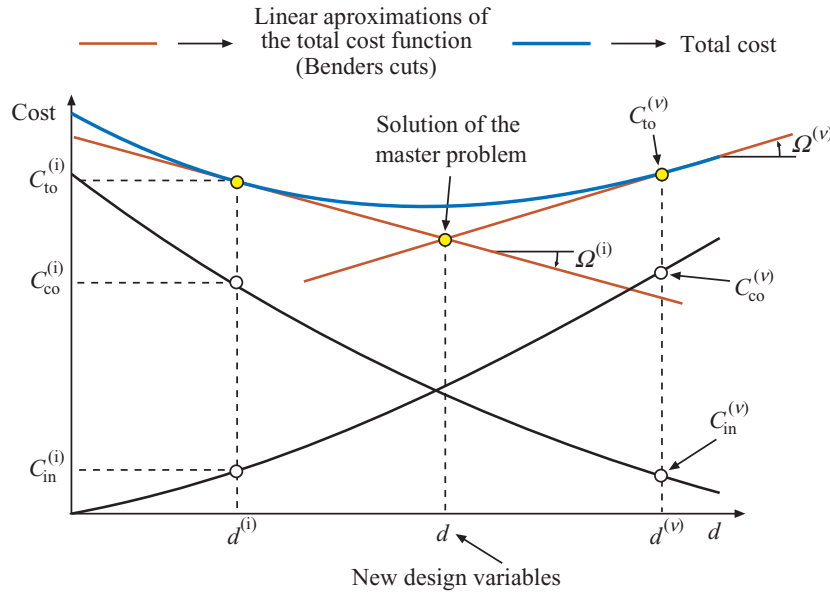


Figure 10.5: Graphical illustration of the reconstruction of the total cost function using Benders cuts.

Otherwise, set $\nu \leftarrow \nu + 1$ and go to Step 3.

Step 3: Master problem. The hyperplane reconstruction of the total cost function is used for calculating the new values of the complicating variables \mathbf{d} .

$$\begin{aligned} & \text{Minimize} && \alpha_{\text{cost}} \\ & \alpha_{\text{cost}}, \mathbf{d} \end{aligned}$$

subject to

$$\begin{aligned} \alpha_{\text{cost}} & \geq C_{\text{to}}^{(\nu)} + \sum_{i=1}^n \Omega_i^{(\nu)} (d_i - d_i^{(\nu)}); \quad \forall i = 1, \dots, \nu - 1 \\ 2 & \leq \frac{1}{\tan \alpha} \leq 5 \\ \alpha_{\text{cost}} & \geq 5000 \end{aligned}$$

The solution of this master problem provides the new values of the design variables $\mathbf{d}^{(\nu)}$ for iteration ν , and the corresponding lower bound of the total cost function $C_{\text{down}}^{(\nu)} = \alpha_{\text{cost}}$. The algorithm continues in Step 1.

As it has been shown, the reliability-based optimization problems characterized as bi level optimization problems, can be solved easily using the Benders decomposition procedure.

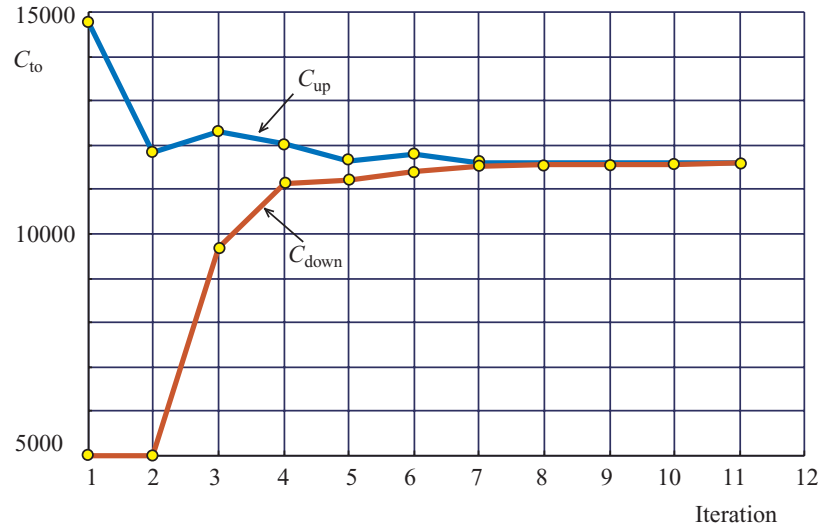


Figure 10.6: Evolution of the objective function for the rubblemound breakwater example.

10.3.1 Numerical example

To perform a reliability-based design of an individual rubblemound breakwater, assume the following values for the variables involved:

$$D_{wl} = 20 \text{ m}; A_u = 1.05; B_u = -0.67; c_c = 60 \text{ \$/m}^3;$$

$$g = 9.81 \text{ m/s}^2; c_a = 2.4 \text{ \$/m}^3; H_s = 5 \text{ m}; \bar{T} = 10 \text{ s}; d_{st} = 1 \text{ h}$$

The solution of this problem, using the method above, it is shown in Table 10.1 . It turns out that convergence of the process requires only 11 iterations. The evolution of the total cost function bounds during the process is illustrated in Figure 10.6. The optimal reliability index and probability of run-up for a single wave and during the design sea state are, respectively:

$$\beta^* = 4.738; P_f^* = 0.00000111 \text{ and } P_f^{D*} = 0.00039845.$$

The GAMS code for solving this rubblemound breakwater example is given in the Appendix A.3.

10.4 The Probability-Safety Factor Method. Reliability Based Optimization of a Vertical Wall

In Chapter 9, several methods for solving special problems with complicating constraints were introduced. Considering the generic problem (9.87)-(9.90), it is clear that its structure is analogous to the reliability-based models stated in Section 10.2.2, and therefore those methods will be suitable for solving reliability base problems including complicating constraints.

Table 10.1: Illustration of the iterative procedure.

ν	F_c	$\tan \alpha_s$	C_{co}	C_{in}	C_{to}	C_{down}	C_{up}	error
1	7.00	0.33	6484.8	8297.8	14782.6	5000.0	14782.6	1.9565
2	5.65	0.20	6836.4	5000.0	11836.4	5000.0	11836.4	1.3673
3	9.32	0.50	7296.0	5000.0	12296.0	9682.5	12296.0	0.2699
4	6.52	0.29	6489.7	5542.8	12032.5	11116.5	12032.5	0.0824
5	6.66	0.29	6571.1	5077.2	11648.3	11197.9	11648.3	0.0402
6	7.02	0.29	6786.8	5000.0	11786.8	11413.5	11786.8	0.0327
7	5.98	0.24	6598.6	5007.5	11606.1	11521.9	11606.1	0.0073
8	6.40	0.26	6583.4	5021.2	11604.5	11570.4	11604.5	0.0030
9	6.00	0.24	6553.6	5033.9	11587.5	11571.8	11587.5	0.0014
10	5.67	0.22	6578.7	5020.4	11599.1	11584.6	11599.1	0.0013
11	5.88	0.23	6571.3	5019.8	11591.1	11585.8	11591.1	0.0005

Considering the ‘Dual partial design’ the following correspondences with problem (9.87)-(9.90) can be appointed:

$$c(\mathbf{x}, \mathbf{y}) \iff c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \quad (10.54)$$

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \iff g_i(\boldsymbol{\gamma}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \geq 0; \forall i \in I \quad (10.55)$$

$$\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{h}_0 \iff \beta_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) \geq \beta_i^0; \forall i \in I \quad (10.56)$$

$$\mathbf{x}^{lo} \leq \mathbf{x} \leq \mathbf{x}^{up} \iff \bar{\mathbf{d}}^{lo} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{up} \quad (10.57)$$

Note that the dual models proposed (10.17)-(10.22) and (10.23)-(10.29), have been never used before for designing specific engineering works, but they are very similar to code calibration problems. This models are the base of the ‘Probability-Safety Factor Method’ (PSFM) for designing engineering works.

There are two principal ways of dealing with the design of engineering works:

1. *The classical approach.* This is based on safety factors, which are used to guarantee the required safety of the structures to be designed. The engineer, when faced with the problem of designing an engineering work, identifies all possible failure modes and chooses the design variable values for the corresponding engineering work to make it reasonably safe with respect to these modes. The greater the damage associated with the failure mode, the greater the level of safety required for this mode (see EUROCODE [74], ROM [143]).

A classical design fixes the values of the safety factors and chooses the values of the design variables to satisfy these safety conditions. All the variables involved are assumed to be deterministic.

2. *The probability-based approach.* This works with probabilities of failure. Normally, a global probability of failure is used as the basic design criteria. However, working with failure probabilities is difficult because (a) it requires the definition of the joint

probability of all variables involved, and (b) the evaluation of the failure probability is not an easy task. The problem becomes even more difficult if several failure modes are analyzed, because the failure region is the union of the different failure mode regions, and regions defined as unions are difficult to deal with because of their irregular and non-differentiable boundaries (see Melchers [123]). As an alternative design criteria, the probabilities of failure for the different modes can be used. Nevertheless, one may easily obtain an upper bound for the global failure probability by summing all the failure mode probabilities.

Nowadays, both approaches are questioned: the classical approach because it does not give a clear idea of how far we are from failure, and the probability-based approach because it is very sensitive to the assumed joint distribution and tail assumptions.

Defenders of the classical and probabilistic approaches have serious difficulties in working together because they speak different languages. In this paper, we defend the coexistence of safety factors and failure probabilities and present a method that solves this problem, limiting both safety factors and probabilities of failure for the proposed designs. The method consists of a sequence of optimal (in the sense of minimizing the cost or an alternative objective function) classical designs, based on sets of safety factors bounds, that are adequately modified in each iteration to satisfy both the safety factors and the failure mode probability bound requirements.

The method provides an automatic design tool by optimizing an objective function subject to the standard geometric and code constraints, and two more sets of constraints, that guarantee some given safety factors and failure probability bounds, associated with a given set of failure modes. This implies a double safety check (see Castillo et al. [?, 31, ?, ?]) that leads to safer structures and designs less prone to wrong or unrealistic probability assumptions, and to excessively small (unsafe) or large (costly) safety factors.

To illustrate these concepts, consider the wall problem given in Figure 10.7, where a and b are the width and the height of the wall, γ is the unit weight of the wall, w is its weight per unit length, t is the horizontal force acting on its right hand side, h is the corresponding offset with respect to the soil level, σ_{mean} is the medium stress at the foundation level (it is used instead of the maximum stress in order to simplify the problem), s is the soil strength, and k is the friction coefficient between soil and wall. These are the actual values of the corresponding random variables that are denoted using the corresponding capital letters.

In this section we assume that a, b, γ, t, h, s and k are independent normal random variables. Note that from the random character point of view there are two sets of variables $\{t, h, s, k\}$, that have a large importance because of its large dispersion, and $\{a, b, \gamma\}$ that are random but with small variability. In fact they could be considered as deterministic. Note also that a and b are random because in the real wall they will not take the exact values desired by the engineer due to constructions errors and imprecisions.

According to the classification stated in Section 10.2.1, the set of variables involved in the problem can be partitioned into five subsets:

- d:** *Optimization design variables.* They are the design variables which values are to be chosen by the optimization program to optimize the objective function (minimize the cost). Normally, they define the dimensions of the work being design, as width,

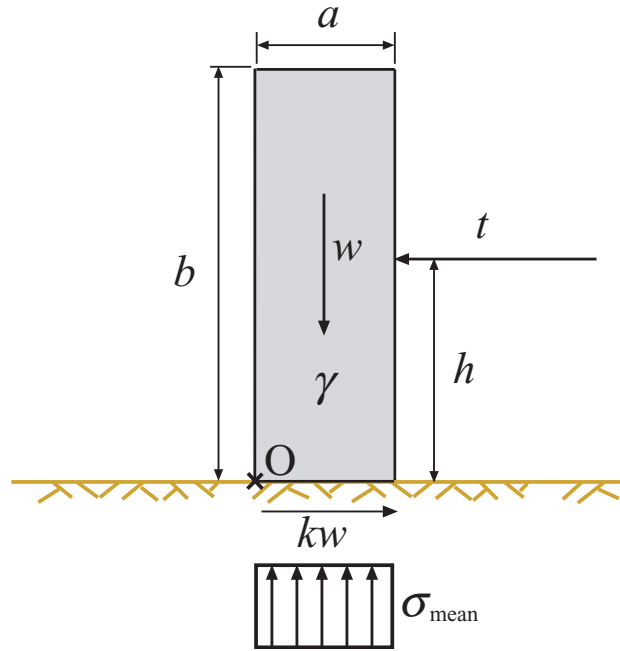


Figure 10.7: Wall and acting forces.

thickness, height, cross sections, etc. In our wall example these variables are:

$$\mathbf{d} = \{a, b\}$$

$\boldsymbol{\eta}$: *Non-optimization design variables.* They are the set of variables (deterministic or random) which mean or characteristic values are fixed by the engineer or the code and must be given as data to the optimization program. Some examples are costs, material properties (unit weights, strength, Young modula, etc.), and other geometric dimensions of the work being designed. In our wall example:

$$\boldsymbol{\eta} = \{\gamma\}$$

$\boldsymbol{\phi}$: *Agents.* This set includes all the variables not controlled by the engineer, such as actions acting on the structure

$$\boldsymbol{\phi} = \{t, h, s, k\}$$

$\boldsymbol{\kappa}$: *Random model parameters.* They are the set of parameters used in the probabilistic design, defining the random variability and dependence structure of the variables involved. In our wall example:

$$\boldsymbol{\kappa} = \{\sigma_a, \sigma_b, \sigma_\gamma, \sigma_t, \sigma_h, \sigma_s, \sigma_k\}$$

where σ refers to the standard deviation of the corresponding variable.

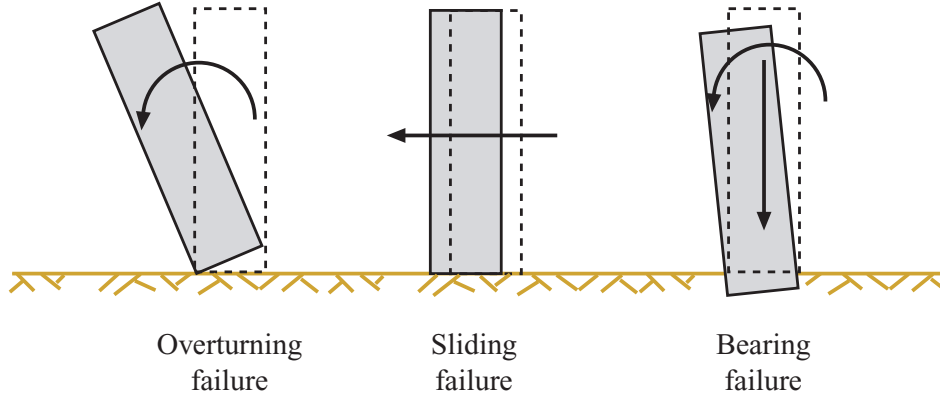


Figure 10.8: Illustration of the wall three modes of failure.

ψ : *Auxiliary or non-basic variables.* They are auxiliary variables which values can be obtained from the basic variables (\mathbf{d} , $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$), using some formulas. They are used to facilitate the calculations and the statement of the problem constraints. In the wall example:

$$\boldsymbol{\psi} = \{w, \sigma_{\text{mean}}\}$$

The corresponding mean of \mathbf{d} , and the mean or characteristic values of $\boldsymbol{\eta}$ and $\boldsymbol{\phi}$ will be denoted $\bar{\mathbf{d}}$, $\tilde{\boldsymbol{\eta}}$ and $\tilde{\boldsymbol{\phi}}$, respectively.

In the classical approach the safety factors are used as constraints and the variables are assumed to be deterministic, i.e., the mean or characteristic (extreme percentiles) values of the variables are considered.

Assume that the following three failure modes are considered (see Figure 10.8):

1. *Overturning failure mode.* The overturning safety factor F_o is defined as the ratio of the stabilizing to the overturning moments with respect to some point (O in Figure 10.7), as

$$F_o = g_o(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) = \frac{\text{stabilizing moment}}{\text{overturning moment}} = \frac{w\bar{a}/2}{\tilde{h}\tilde{t}} = \frac{\bar{a}^2\tilde{b}\tilde{\gamma}}{2\tilde{h}\tilde{t}} \geq F_o^0 \quad (10.58)$$

where F_o^0 is the corresponding lower bound associated with the overturning failure, and the bars and tildes refer to the means and the characteristic values of the corresponding variables, respectively.

2. *Sliding failure mode.* The sliding safety factor F_s is the ratio of the stabilizing to the sliding forces as

$$F_s = g_s(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) = \frac{\text{stabilizing force}}{\text{sliding force}} = \frac{\tilde{k}w}{\tilde{t}} = \frac{\bar{a}\tilde{b}\tilde{k}\tilde{\gamma}}{\tilde{t}} \geq F_s^0 \quad (10.59)$$

where F_s^0 is the corresponding lower bound associated with the sliding failure.

3. *Bearing capacity failure mode.* The bearing capacity safety factor F_b is the ratio of the bearing capacity to the maximum stress at the bottom of the wall (considered constant for illustration purposes)

$$F_b = g_b(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) = \frac{\text{bearing capacity}}{\text{maximum stress}} = \frac{\tilde{S}}{\sigma_{\text{mean}}} \geq F_b^0 \quad (10.60)$$

where F_b^0 is the corresponding lower bound associated with a foundation failure.

The wall is safe if and only if $F_o, F_s, F_b \geq 1$.

10.4.1 Relaxation method

In this Section the method presented in Section 9.6.1 is particularized for solving the dual models. The method uses an iterative procedure that consists of repeating a sequence of three steps (see Castillo et al. [31, ?, ?]): (1) an optimal (in the sense of optimizing an objective function) classic design, based on given safety factors, is done, (2) reliability indices or bounds for all failures modes are determined, and (3) all mode safety factor bounds are adjusted. The three steps are repeated until convergence, i.e., until the safety factors lower bounds and the failure mode probability upper bounds are satisfied. More precisely, considering the ‘Dual global design’ (10.17)-(10.22), the method proceeds as follows:

Algorithm 10.1 (Updating safety factor bounds).

Input. The nonlinear programming problem (10.17)-(10.22).

Output. The solution of the problem obtained by iteratively updating the safety factor bounds.

Step 0: Initialization. Set the iteration counter to $\nu = 1$, and the global safety factors $F_i^{(\nu)}$ to their lower bound F_i^0 .

Step 1: Solving the optimal classic design. An optimal classic design based on the actual safety factors, that are fixed initially to their corresponding lower bounds, is done. In other words, the following problem is solved :

$$\underset{\bar{\mathbf{d}}}{\text{Minimizar}} \quad c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \quad (10.61)$$

sujeito a

$$g_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_i^{(\nu)}) \geq 0 : \forall i \in I \quad (10.62)$$

$$r_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \leq 0 : \forall j \in J \quad (10.63)$$

$$\mathbf{h}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) = \boldsymbol{\psi} \quad (10.64)$$

$$\bar{\mathbf{d}}^{\text{lo}} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.65)$$

The result of this process is a set of values of the design variables (their means) that satisfy the safety factor constraints (10.62) and the geometric and code ones (10.63).

Step 2: Evaluating the reliability indices β_i .

The actual reliability indices β_i associated with all modes of failure are evaluated, based on the design values of Step 1, solving the problem

$$\beta_i^{(\nu)} = \underset{\mathbf{z}}{\text{Mínimo}} \sqrt{\mathbf{z}^T \mathbf{z}} \quad (10.66)$$

subject

$$g_i(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi}; 1) \geq 0 \quad (10.67)$$

$$\mathbf{h}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}) = \boldsymbol{\psi} \quad (10.68)$$

$$\mathbf{T}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\kappa}) = \mathbf{z} \quad (10.69)$$

At this step as many optimization problems as the number of modes of failure are solved. Note that the design variables $\vec{\mathbf{d}}^{(\nu)}$ are included in the Rosenblatt transformation (10.69).

Step 3: Check convergence. If $\|(\boldsymbol{\beta}^{(\nu)} - \boldsymbol{\beta}^{(\nu-1)})/\boldsymbol{\beta}^{(\nu)}\| < \epsilon$ then, stop. Otherwise, go to Step 4.

Step 4: Updating safety factor values.

The safety factors bounds are adequately updated. To this end, the safety factors are modified using the increments

$$F_i^{(\nu+1)} = \max \left(F_i^{(\nu)} + \rho(\beta_i^0 - \beta^{(\nu)}), F_i^0 \right) \quad (10.70)$$

where ρ is a small positive constant (relaxation factor). Next, the iteration counter is updated $\nu \leftarrow \nu + 1$ and the process continues with Step 1.

Note that values of the actual reliability indices $\beta_i^{(\nu+1)}$ below the desired bound levels β_i^0 , lead to an increase of the associated safety factor bound.

In addition, if, using this formula, any safety factor $F_i^{(\nu+1)}$ becomes smaller than the associated lower bound, it is kept equal to F_i^0 . ■

In Figure 10.9 it is shown the diagram of the method.

Remarks:

1. The result of the iterative scheme is an optimal design that satisfies the required safety factors and β -values bounds. The optimal classic design and the optimal probability based design are particular cases of the present dual design procedure. If a classic design is looked for, large bounds for the probabilities of failure ($p_i^0; \forall i \in I$) or small bounds for the reliability indices ($\beta_i^0; \forall i \in I$) must be selected. If, on the contrary, a probability based design is looked for, low safety factors lower bounds ($F_i^0 = 1; \forall i \in I$) must be chosen.
2. One of the problems of the classical designs based on safety factors alone consists of its lack of invariance of the corresponding safety measures, that should not depend on the way in which the loads and resistances are defined. In simple terms, this

requires that all variables be reduced to a common base before being compared. Note that the optimal classic design (Step 1 of the PFSF method) does not satisfy this condition. However, in the proposed method, this is solved within the probabilistic step, in which the “design points” or “points of maximum likelihood” are calculated. Thus, the partial safety factors (γ_{d_i} and γ_{η_i}) that must be used in a classical design (based on existing codes) can be calculated for each mode of failure to obtain the same reliability (equivalent designs)

$$\gamma_{d_i} = \frac{d_i^*}{d} \quad \text{and} \quad \gamma_{\eta_i} = \frac{\eta_i^*}{\tilde{\eta}}. \tag{10.71}$$

3. The use of this method is similar to establishing a code calibration for each specific design, instead of using the existing codes that define the partial safety factors for many design situations within a selected scope first step in code calibrations (Lind [110], Baker [10], Ellinwood et al. [67]).

Illustrative Example 10.1 (Wall design using the relaxation method). To perform a probabilistic design in the wall example presented in Section 10.4, the joint probability density of all variables is required. Assume for example that all the variables involved are independent normal random variables, i.e.,

$$a \sim N(\bar{a}, \sigma_a); \quad b \sim N(\bar{b}, \sigma_b); \quad k \sim N(\tilde{k}, \sigma_k);$$

$$t \sim N(\tilde{t}, \sigma_t); \quad \gamma \sim N(\tilde{\gamma}, \sigma_\gamma); \quad h \sim N(\tilde{h}, \sigma_h); \quad s \sim N(\tilde{s}, \sigma_s)$$

where $\bar{a}, \bar{b}, \tilde{k}, \tilde{t}, \tilde{\gamma}, \tilde{h}$ and \tilde{s} are the mean values and $\sigma_a, \sigma_b, \sigma_k, \sigma_t, \sigma_\gamma, \sigma_h$ and σ_s are the standard deviations of a, b, k, t, γ, h and s , respectively. The numerical values are given in Table 10.2.

Table 10.2: Data for the wall example.

Variable x							
	a [m]	b [m]	k	t [kN]	γ [kN/m ³]	h [m]	s [kN/m ²]
\bar{x} o \tilde{x}	\bar{a}	\bar{b}	0.3	50	23	3	220
σ_x	0.01	0.01	0.05	15	0.46	0.2	16

Using the Rosenblatt [146] transformation, this set is transformed into a set of standard normal random variables z_1, z_2, \dots, z_7 by

$$\begin{aligned} z_1 &= \frac{a - \bar{a}}{\sigma_a}; & z_2 &= \frac{b - \bar{b}}{\sigma_b}; & z_3 &= \frac{k - \tilde{k}}{\sigma_k}; & z_4 &= \frac{t - \tilde{t}}{\sigma_t}; \\ z_5 &= \frac{\gamma - \tilde{\gamma}}{\sigma_\gamma}; & z_6 &= \frac{h - \tilde{h}}{\sigma_h}; & z_7 &= \frac{s - \tilde{s}}{\sigma_s} \end{aligned} \tag{10.72}$$

Assume that the required safety factors and reliability bounds are:

$$F_o^0 = 1.5; \quad F_s^0 = 1.6; \quad F_b^0 = 1.5; \quad \beta_o^0 = 3; \quad \beta_s^0 = 3; \quad \beta_b^0 = 3$$

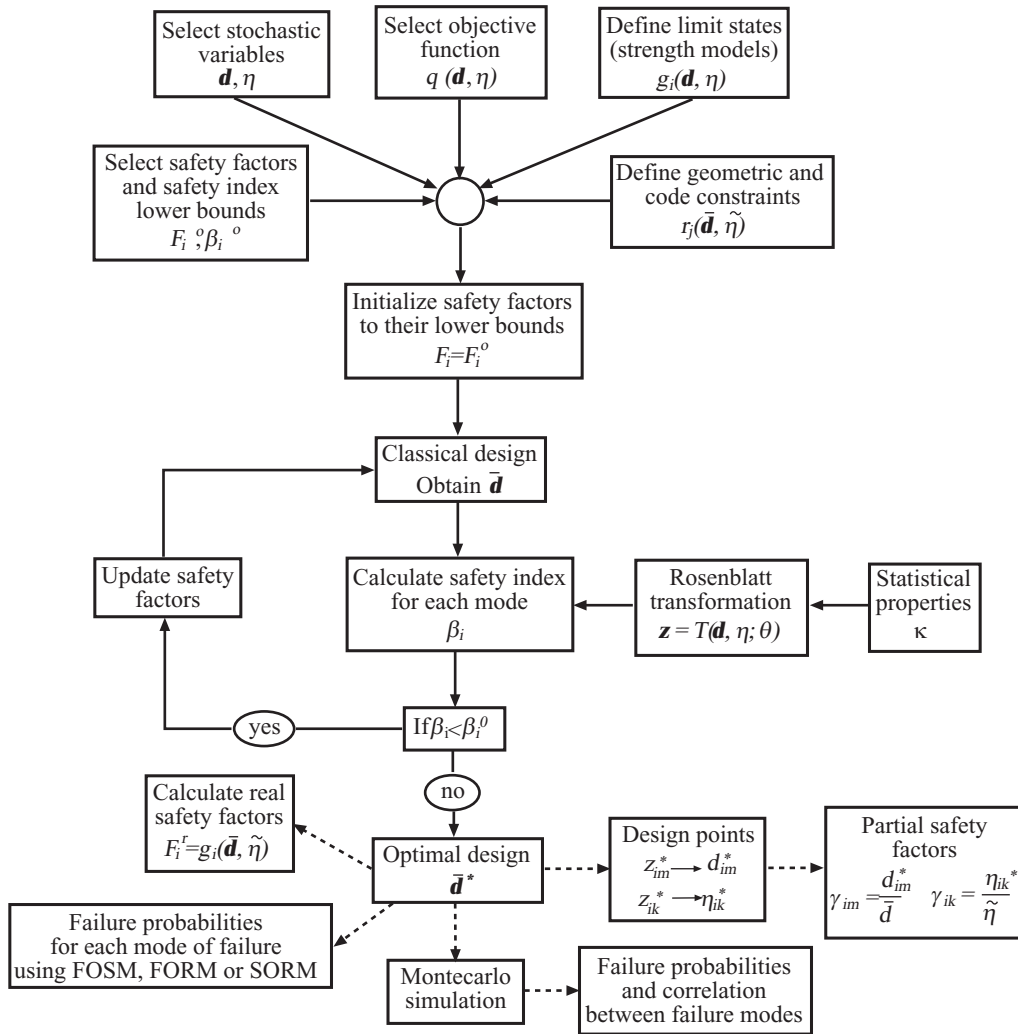


Figure 10.9: Flow diagram of the proposed ‘Probability-Safety Factor Method’ (PSFM).

For these numerical data, Algorithm 9.3 in this case consists of the following steps:

Step 0: Initialization. Let

$$\nu = 1; F_o^{(1)} = 1.5; F_s^{(1)} = 1.6; F_b^{(1)} = 1.5$$

Step 1: Solve the classical problem.

$$\text{Minimize } \bar{a}\bar{b} \quad (10.73)$$

$$\bar{a}, \bar{b}$$

subject to

$$\frac{\bar{a}^2\bar{b}\tilde{\gamma}}{2\tilde{h}\tilde{t}} \geq F_o^{(\nu)} \quad (10.74)$$

$$\frac{\bar{a}\bar{b}\tilde{k}\tilde{\gamma}}{\tilde{t}} \geq F_s^{(\nu)} \quad (10.75)$$

$$\frac{\tilde{s}}{\sigma_{\text{mean}}} \geq F_b^{(\nu)} \quad (10.76)$$

$$\bar{b} = 2\bar{a} \quad (10.77)$$

Step 2: Solve the subproblems. For $i = 1, 2, 3$:

$$\beta_i^{(\nu)} = \text{Minimum}_{a, b, k, t, \gamma, h, s} \sum_{j=1}^7 z_j^2 \quad (10.78)$$

subject to (10.72) and

$$\frac{a^2b\gamma}{2ht} = 1 \quad (10.79)$$

or

$$\frac{abk\gamma}{t} = 1 \quad (10.80)$$

or

$$\frac{s}{\sigma_{\text{mean}}} = 1 \quad (10.81)$$

depending on $i = 1, 2$ or 3 , i.e., overturning, sliding or foundation failures, respectively, are considered.

Step 3: Check convergence. If $\|(\beta^{(\nu)} - \beta^{(\nu-1)})/\beta^{(\nu)}\| < \epsilon$ then, stop. Otherwise, go to Step 4.

Step 4: Update safety factors. Using

$$F_i^{(\nu+1)} = \max \left(F_i^{(\nu)} + \rho(\beta_i^0 - \beta^{(\nu)}), F_i^0 \right)$$

the safety factors are updated. Next, the iteration counter is updated $\nu \leftarrow \nu + 1$ and the process continues with Step 1.

Table 10.3: Illustration of the iterative process of the original algorithm.

ν	cost	a	b	Actual bounds			Actual values			Actual values		
				F_o^0	F_s^0	F_b^0	F_o	F_s	F_b	β_o	β_s	β_b
1	11.594	2.408	4.816	1.500	1.600	1.500	2.140	1.600	1.986	3.456	1.491	6.763
2	20.332	3.189	6.377	1.500	1.600	1.500	4.970	2.806	1.500	10.132	3.245	4.508
3	18.918	3.076	6.152	1.500	1.600	1.500	4.461	2.611	1.555	9.084	3.042	4.832
4	18.672	3.056	6.112	1.500	1.600	1.500	4.374	2.577	1.565	8.900	3.005	4.890
5	18.645	3.054	6.107	1.500	1.600	1.500	4.365	2.573	1.566	8.879	3.000	4.897
6	18.642	3.053	6.107	1.500	1.600	1.500	4.364	2.573	1.566	8.877	3.000	4.897

Using this algorithm assuming that the required safety factors and reliability bounds are:

$$F_o^0 = 1.5; F_s^0 = 1.6; F_b^0 = 1.5; \beta_o^0 = 3; \beta_s^0 = 3; \beta_b^0 = 3$$

the results shown in Table 10.3 are obtained. The solution is: $\bar{a} = 3.053 m$ and $\bar{b} = 6.107 m$. This table shows the progress and convergence of the process, that requires only 6 iterations when the tolerance is equal to $\epsilon = 0.001$, and that only β_s is active, and F_o, F_s, F_b, β_o and β_b are inactive. This is illustrated in the last row of the table, where the active values have been boldfaced. This means that the reliability index for sliding is more restrictive than the corresponding safety factor.

Since the final overturning actual safety factor (4.364) and beta value (8.877) are very high compared with the design bounds (1.5 and 3, respectively), this implies that overturning is almost impossible. ■

The GAMS implementation of the relaxation method for solving the Probability-Safety Factor Method is shown in Appendix A.4.

10.4.2 Hyperplane approximation method

The iterative method presented in the previous Section requires a relaxation factor ρ that needs to be fixed by trial and error. An adequate selection leads to a fast convergence of the process, but an inadequate selection can lead to lack of convergence. In this section an alternative method (see Castillo et al. [?]) is given that solves this shortcoming, and in addition exhibit a better convergence. The method is as follows.

The problem (10.17)-(10.22) can be organized in the following steps:

Step 1: Solve the classical problem.

$$\text{Minimize } c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \quad (10.82)$$

$$\bar{\mathbf{d}}$$

subject to

$$g_i(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_i^0) \geq 0 : \forall i \in I \quad (10.83)$$

$$\beta_i^{(s)} + \lambda_i^{(s)T} (\bar{\mathbf{d}} - \bar{\mathbf{d}}^{(s)}) \geq \beta_i^0 : \forall i \in I; s = 1, 2, \dots, \nu - 1 \quad (10.84)$$

$$r_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \leq 0 : \forall j \in J \quad (10.85)$$

$$\mathbf{h}(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) = \boldsymbol{\psi} \quad (10.86)$$

$$\bar{\mathbf{d}}^{\text{lo}} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.87)$$

The result of this process is a set of values of the design variables $\bar{\mathbf{d}}^{(\nu)}$ (their means) that satisfy the safety factor constraints (10.83) and the geometric and code ones (10.85). It also includes a linear approximation of constraints (10.84) associated with the reliability index bounds.

Step 2: Solve the subproblems. Reliability indexes for the actual design variable values are evaluated solving the following problem:

$$\beta_i^{(\nu)} = \text{Minimum}_{\mathbf{z}, \bar{\mathbf{d}}} \sqrt{\mathbf{z}^T \mathbf{z}} \quad (10.88)$$

subject to

$$g_i(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi}; 1) = 0 \quad (10.89)$$

$$\mathbf{h}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}) = \boldsymbol{\psi} \quad (10.90)$$

$$\mathbf{T}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\kappa}) = \mathbf{z} \quad (10.91)$$

$$\bar{\mathbf{d}} = \bar{\mathbf{d}}^{(\nu)} : \boldsymbol{\lambda}_i^{(\nu)} \quad (10.92)$$

where $\boldsymbol{\lambda}_i^{(\nu)}$ is the dual variable vector associated with constraint (10.92) which fixes the mean values of the design variables for iteration ν to the values obtained in the classical design $\bar{\mathbf{d}}^{(\nu)}$.

Step 3: Check convergence. If $\|(\boldsymbol{\beta}^{(\nu)} - \boldsymbol{\beta}^{(\nu-1)})/\boldsymbol{\beta}^{(\nu)}\| < \epsilon$ then, stop. Otherwise, the iteration counter is updated $\nu \leftarrow \nu + 1$ and the process continues with Step 1.

It should be noted that problem (10.82)-(10.87) is a relaxation of problem (10.17)-(10.22) in the sense that functions $\beta_i^{(s)}(\cdot)$ are approximated using cutting hyperplanes. Functions $\beta_i^{(s)}(\cdot)$ become more precisely approximated as the iterative procedure progresses, which implies that problem (10.82)-(10.87) reproduces more exactly problem (10.17)-(10.22) (see Kelly [102]). Observe, additionally, that cutting hyperplanes are constructed using the dual variable vector associated with constraint (10.92) in problems (10.88)-(10.92) (the subproblems). Constraints (10.92) in problems (10.88)-(10.92) fix to given values the optimization variable vector of problem (10.82)-(10.87) (the master problem).

Illustrative Example 10.2 (Wall design using the hyperplane approximation).

For the wall example, this process consists of the following steps:

Step 0: Initialization. Let

$$\nu = 1; F_o^{(1)} = F_o^0 = 1.5; F_s^{(1)} = F_s^0 = 1.6; F_b^{(1)} = F_b^0 = 1.5$$

Step 1: Solve the classical problem.

$$\text{Minimize } \bar{a}\bar{b} \quad (10.93)$$

$$\bar{a}, \bar{b}$$

subject to

$$\frac{\bar{a}^2\bar{b}\tilde{\gamma}}{2\tilde{h}\tilde{t}} \geq F_o^0 \quad (10.94)$$

$$\frac{\bar{a}\bar{b}\tilde{k}\tilde{\gamma}}{\tilde{t}} \geq F_s^0 \quad (10.95)$$

$$\frac{\tilde{s}}{\sigma_{\text{mean}}} \geq F_b^0 \quad (10.96)$$

$$\beta_i^{(s)} + \lambda_i^{(s)T} \left(\begin{pmatrix} a \\ b \end{pmatrix} - \begin{pmatrix} a^{(s)} \\ b^{(s)} \end{pmatrix} \right) \geq \beta_i^0; \forall i \in I; s = 1, \dots, \nu - 1 \quad (10.97)$$

$$\bar{b} = 2\bar{a} \quad (10.98)$$

obtaining $\bar{a}^{(\nu)}$ and $\bar{b}^{(\nu)}$.

Step 2: Solve the subproblems. For $i = 1, 2, 3$:

$$\beta_i^{(\nu)} = \text{Minimum}_{a, b, k, t, \gamma, h, s} \sum_{j=1}^7 z_j^2 \quad (10.99)$$

subject to (10.72),

$$a = \bar{a}^{(\nu)} : \lambda_{1i}^{(\nu)}$$

$$b = \bar{b}^{(\nu)} : \lambda_{2i}^{(\nu)}$$

and

$$\frac{a^2b\gamma}{2ht} = 1 \quad (10.100)$$

or

$$\frac{abk\gamma}{t} = 1 \quad (10.101)$$

or

$$\frac{s}{\sigma_{\text{mean}}} = 1 \quad (10.102)$$

depending on $i = 1, 2$ or 3 , i.e., overturning, sliding or foundation failures, respectively, are considered.

Step 3: Check convergence. If $\|(\beta^{(\nu)} - \beta^{(\nu-1)})/\beta^{(\nu)}\| < \epsilon$ then, stop. Otherwise, the iteration counter is updated $\nu \leftarrow \nu + 1$ and the process continues with Step 1.

The iterative procedure leads to the results shown in Table 10.4, that provides the same information as Table 10.3 using the alternative procedure. In this case the process converges after 5 iterations for the same value of the tolerance $\epsilon = 0.001$.

■

Table 10.4: Illustration of the iterative process for the alternative algorithm.

ν	cost	a	b	Actual bounds			Actual values			Actual values		
				F_o^0	F_s^0	F_b^0	F_o	F_s	F_b	β_o	β_s	β_b
1	11.594	2.408	4.816	1.500	1.600	1.500	2.140	1.600	1.986	3.456	1.491	6.763
2	17.463	2.955	5.910	1.500	1.600	1.500	3.956	2.410	1.619	7.989	2.807	5.180
3	18.603	3.050	6.100	1.500	1.600	1.500	4.350	2.567	1.568	8.848	2.994	4.906
4	18.642	3.053	6.107	1.500	1.600	1.500	4.363	2.573	1.567	8.877	3.000	4.897
5	18.642	3.053	6.107	1.500	1.600	1.500	4.363	2.573	1.567	8.877	3.000	4.897

An additional advantage of the proposed methodology is that partial safety factors associated with random variables can be used, instead of global safety factors. Thus, the method can be used as a calibration tool of the standard codes based on partial safety factors. Particularizing the method to the ‘dual partial design’ (10.23)-(10.29) the three basic steps for iteration ν are as follows:

Step 1: Solve the classical problem.

$$\text{Minimize}_{\bar{\mathbf{d}}} c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \quad (10.103)$$

subject to

$$g_i(\boldsymbol{\gamma}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; 1) \geq 0; \forall i \in I \quad (10.104)$$

$$\beta_i^{(s)} + \boldsymbol{\lambda}_i^{(s)T} (\bar{\mathbf{d}} - \bar{\mathbf{d}}^{(s)}) \geq \beta_i^0; \forall i \in I; s = 1, 2, \dots, \nu - 1 \quad (10.105)$$

$$r_j(\boldsymbol{\gamma}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \leq 0; \forall j \in J \quad (10.106)$$

$$\mathbf{h}(\boldsymbol{\gamma}, \bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) = \boldsymbol{\psi} \quad (10.107)$$

$$\bar{\mathbf{d}}^{\text{lo}} \leq \bar{\mathbf{d}} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.108)$$

$$\boldsymbol{\gamma}^{\text{lo}} \leq \boldsymbol{\gamma} \leq \boldsymbol{\gamma}^{\text{up}} \quad (10.109)$$

The result of this process is a set of values of the design variables $\bar{\mathbf{d}}^{(\nu)}$ (their means) that satisfy the safety factor constraints (10.104) and the geometric and code ones (10.106). It also includes a linear approximation (10.105) of constraints associated with the reliability index bounds.

Step 2: Solve the subproblems. Reliability indexes for the actual design variable values are evaluated solving the following problem:

$$\beta_i^{(\nu)} = \text{Minimum}_{\mathbf{z}, \bar{\mathbf{d}}} \sqrt{\mathbf{z}^T \mathbf{z}} \quad (10.110)$$

subject to

$$g_i(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi}; 1) = 0 \quad (10.111)$$

$$\mathbf{h}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}) = \boldsymbol{\psi} \quad (10.112)$$

$$\mathbf{T}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\kappa}) = \mathbf{z} \quad (10.113)$$

$$\bar{\mathbf{d}} = \bar{\mathbf{d}}^{(\nu)} : \boldsymbol{\lambda}_i^{(\nu)} \quad (10.114)$$

where $\boldsymbol{\lambda}_i^{(\nu)}$ is the dual variable vector associated with constraint (10.114) which fixes the mean values of the design variables for iteration ν to the values obtained in the classical design $\bar{\mathbf{d}}^{(\nu)}$.

Step 3: Check convergence. If $\|(\boldsymbol{\beta}^{(\nu)} - \boldsymbol{\beta}^{(\nu-1)})/\boldsymbol{\beta}^{(\nu)}\| < \epsilon$ then, stop. Otherwise, the iteration counter is updated $\nu \leftarrow \nu + 1$ and the process continues with Step 1.

The GAMS implementation of the hyperplane approximation for solving the Probability-Safety Factor Method is shown in Appendix A.5.

10.5 Global and Partial Model Resolution

In the problems where there are both complicating variables and constraints, that is, global (10.30)-(10.35) and partial (10.36)-(10.42) reliability based designs, respectively, special procedures are needed. The best solution from the computational point of view is an external Benders decomposition for dealing with complicating variables, and the hyperplane approximation for the complicating constraints.

Particularizing for the global reliability based design (10.30)-(10.35), the method proceeds as follows:

Step 0: Initialization. Select initial values for the design variables (complicating variables) $\mathbf{d}^{(0)}$.

Set the iteration counter to $\nu = 1$, $\mathbf{d}^{(\nu)} = \mathbf{d}^{(0)}$, and the total cost lower bound to $C_{\text{down}}^{(\nu)} = -\infty$.

Step 1: Evaluating the reliability indexes $\beta_i^{(\nu)}$ (subproblem solution). Based on the actual design values (complicating variables), the reliability index is calculated solving the problem:

$$\beta_i^{(\nu)} = \underset{\mathbf{z}, \bar{\mathbf{d}}}{\text{Minimum}} \sqrt{\mathbf{z}^T \mathbf{z}} \quad (10.115)$$

subject to

$$g_i(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi}; 1) = 0 \quad (10.116)$$

$$\mathbf{h}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}) = \boldsymbol{\psi} \quad (10.117)$$

$$\mathbf{T}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\kappa}) = \mathbf{z} \quad (10.118)$$

$$\bar{\mathbf{d}} = \bar{\mathbf{d}}^{(\nu)} : \boldsymbol{\lambda}_i^{(\nu)} \quad (10.119)$$

where $\boldsymbol{\lambda}_i^{(\nu)}$ is the vector containing the partial derivatives of the reliability index with respect the design variables associated to constraint (10.119) that fixes the mean values of

the design variables to the ones obtained from the master problem $\bar{\mathbf{d}}^{(\nu)}$.

Step 2: Total cost evaluation. The total cost $C_{\text{to}}^{(\nu)}$ for the actual values of the complicating (design) variables $\mathbf{d}^{(\nu)}$ is evaluated

$$C_{\text{to}}^{(\nu)}(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) = C_{\text{co}}(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) + C_{\boldsymbol{\beta}^{(\nu)}}(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa}) \quad (10.120)$$

and the partial derivatives of function C_{to} with respect the complicating variables

$$\begin{aligned} \boldsymbol{\Omega}^{(\nu)} &= \frac{\partial C_{\text{to}}^{(\nu)}(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa})}{\partial \bar{\mathbf{d}}^{(\nu)}} = \frac{\partial C_{\text{co}}(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi})}{\partial \bar{\mathbf{d}}^{(\nu)}} + \frac{\partial C_{\boldsymbol{\beta}^{(\nu)}}(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa})}{\partial \bar{\mathbf{d}}^{(\nu)}} = \\ &= \frac{\partial C_{\text{co}}(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi})}{\partial \bar{\mathbf{d}}^{(\nu)}} + \sum_{\forall i} \lambda_i^{(\nu)} \frac{\partial C_{\boldsymbol{\beta}^{(\nu)}}(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}, \boldsymbol{\kappa})}{\partial \beta_i^{(\nu)}} \end{aligned} \quad (10.121)$$

Set the total cost upper bound to $C_{\text{up}}^{(\nu)} = C_{\text{to}}^{(\nu)}$.

Step 2: Convergence check. If $|C_{\text{up}}^{(\nu)} - C_{\text{down}}^{(\nu)}|/|C_{\text{up}}^{(\nu)}| \leq \varepsilon$, the solution with a level of accuracy ε of the objective function and design variables is

$$C_{\text{to}}^* = C_{\text{to}}^{(\nu)}; \quad \mathbf{d}^* = \mathbf{d}^{(\nu)}$$

Otherwise, set $\nu \leftarrow \nu + 1$ and go to Step 4.

Step 4: Master problem. The hyperplane reconstruction of the total cost function is used for calculating the new values of the complicating variables \mathbf{d} .

$$\begin{aligned} &\text{Minimize} \quad \alpha_{\text{cost}} \\ &\alpha_{\text{cost}}, \bar{\mathbf{d}}^{(\nu)} \end{aligned} \quad (10.122)$$

subject to

$$\alpha_{\text{cost}} \geq C_{\text{to}}^{(j)} + \boldsymbol{\Omega}^{(j)T}(\bar{\mathbf{d}}^{(\nu)} - \bar{\mathbf{d}}^{(j)}); \quad \forall j = 1, \dots, \nu - 1 \quad (10.123)$$

$$0 \leq g_i(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_i^0); \quad \forall i \in I \quad (10.124)$$

$$\beta_i^0 \leq \beta_i^{(j)} + \boldsymbol{\lambda}_i^{(j)T}(\bar{\mathbf{d}}^{(\nu)} - \bar{\mathbf{d}}^{(j)}); \quad \forall i \in I; \quad j = 1, 2, \dots, \nu - 1 \quad (10.125)$$

$$0 \leq r_j(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}); \quad \forall j \in J \quad (10.126)$$

$$\boldsymbol{\psi} = \mathbf{h}(\bar{\mathbf{d}}^{(\nu)}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}) \quad (10.127)$$

$$\bar{\mathbf{d}}^{\text{down}} \leq \bar{\mathbf{d}}^{(\nu)} \leq \bar{\mathbf{d}}^{\text{up}} \quad (10.128)$$

The solution of this master problem provides the new values of the design variables $\mathbf{d}^{(\nu)}$ for iteration ν , and the corresponding lower bound of the total cost function $C_{\text{down}}^{(\nu)} = \alpha_{\text{cost}}$. The algorithm continues in Step 1.

If constraint (10.125) changes its curvature inside the feasibility region lack of convergence can occur. In this situation is useful to introduce just the last approximative hyperplane

$$\beta_i^0 \geq \beta_i^{(j)} + \lambda_i^{(j)T} (\bar{\mathbf{d}}^{(\nu)} - \bar{\mathbf{d}}^{(j)}); \quad \forall i \in I; \quad j = \nu - 1.$$

Analogously, the partial reliability based model (10.36)-(10.42) can be solved using the same method but replacing constraints (10.124), (10.126) and (10.127) in the master problem by (10.37), (10.39), (10.40) and adding the corresponding partial safety factors bounds (10.42).

10.6 Sensitivity Analysis in Structural Optimization Problems with Reliability Constraints

It has been shown that multiple engineering problems can be stated and solved using first order reliability techniques, and optimization methods. The solution of these problems provides the optimal design values \mathbf{x}^* , the optimal objective function, and the corresponding constraint mismatches. As every nonlinear programming problem the conditions of Karush, Kuhn and Tucker must be satisfied at any constrained optimum, local or global (Frangopol [77] and Enevoldsen [69]).

Sensitivity analysis consists of determining “how” and “how much” specific changes in the parameters of an optimization problem influence the optimal objective function value and the point (or points) where the optimum is attained.

The problem of sensitivity analysis in nonlinear programming has been discussed by several authors, as, for example, Vanderplaats [169], Sobiesky et al. [?], Enevoldsen, [69], Roos, Terlaky and Vial [144], Bjerager and Krand [22], etc. There are at least three ways of deriving equations for the unknown sensitivities: (a) the Lagrange multiplier equations of the constrained optimum (see Sobiesky et al. [?]), (b) differentiation of the Karush–Kuhn–Tucker conditions to obtain the sensitivities of the objective function with respect to changes in the parameters (see Sorensen and Enevoldsen [160] or Enevoldsen, [69]), and (c) the extreme conditions of a penalty function (see Sobiesky et al. [?]). Subsequently, Castillo et al. [?, ?] present a method with several advantages with respect the previous ones: (a) to perform a general analysis, without assuming neither the existence of partial derivatives, nor that the active inequality constraints remain active, and (b) to give an integrated approach that gives all sensitivities (objective function values, primal and dual variables values) at once. As a consequence, a distinction between equality and inequality constraints is necessary, because the active inequality constraints may become inactive, but equality constraints are always active. This consideration leads to a set of feasible changes that has the structure of a cone. Furthermore, this analysis allows determining whether or not partial or directional derivatives with respect to parameters exist.

Without considering theoretical sensitivity aspects, in this Section the application of the general method presented in Section 9.7.2 is shown.

10.6.1 Sensitivity analysis in the probability-safety factor method (PSFM)

Se va a aplicar el análisis de sensibilidad con el método general para la resolución de problemas mediante el método combinado de probabilidades de fallo coeficientes de seguridad (PSFM). La aplicación en los demás modelos es totalmente extrapolable de forma inmediata.

Dado que el estudio de sensibilidad se realizará una vez se tenga la solución óptima del problema, se va a plantear los modelos auxiliares que me permiten obtener todas las sensibilidades partiendo de la solución óptima.

Particularizando al planteamiento general del método de relajación para resolver el diseño mixto global (10.17)-(10.22) y teniendo en cuenta que la solución óptima del problema es $\bar{\mathbf{d}}^*$, \mathbf{d}_i^* , $\boldsymbol{\eta}_i^*$ y $\boldsymbol{\phi}_i^*$; $\forall i \in I$:

Sensibilidad en el modelo mixto global. Se plantea el siguiente problema auxiliar

$$\begin{array}{l} \text{Minimizar} \\ \bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\kappa}^{\text{aux}} \end{array} \quad c(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\psi}) \quad (10.129)$$

sujeto a

$$g_i(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\psi}) \geq F_i^0 : \Gamma_i; \forall i \in I \quad (10.130)$$

$$\sqrt{\mathbf{z}_i^T \mathbf{z}_i} \geq \beta_i^0 : \Lambda_i; \forall i \in I \quad (10.131)$$

$$\mathbf{T}(\mathbf{d}_i^*, \boldsymbol{\eta}_i^*, \boldsymbol{\phi}_i^*, \boldsymbol{\kappa}^{\text{aux}}) = \mathbf{z}_i; \forall i \in I \quad (10.132)$$

$$r_j(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\psi}) \leq 0; \forall j \in J \quad (10.133)$$

$$\mathbf{h}(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}) = \boldsymbol{\psi} \quad (10.134)$$

$$\tilde{\boldsymbol{\eta}}^{\text{aux}} = \tilde{\boldsymbol{\eta}} : \boldsymbol{\varsigma} \quad (10.135)$$

$$\tilde{\boldsymbol{\phi}}^{\text{aux}} = \tilde{\boldsymbol{\phi}} : \boldsymbol{\varrho} \quad (10.136)$$

$$\boldsymbol{\kappa}^{\text{aux}} = \boldsymbol{\kappa} : \boldsymbol{\nu} \quad (10.137)$$

de donde se obtiene la sensibilidad $\boldsymbol{\Gamma}$ del coste con respecto a los límites inferiores de los coeficientes de seguridad \mathbf{F}^0 , las derivadas parciales $\boldsymbol{\Lambda}$ con respecto a los límites inferiores de los índices de fiabilidad β^0 , las sensibilidades $\boldsymbol{\varsigma}$, $\boldsymbol{\varrho}$ y $\boldsymbol{\nu}$ del coste con respecto $\tilde{\boldsymbol{\eta}}$, $\tilde{\boldsymbol{\phi}}$ y $\boldsymbol{\kappa}$, respectivamente. Con lo cual el ingeniero proyectista conoce la influencia que tienen todas las variables de proyecto y las restricciones de fiabilidad, tanto clásicas como modernas, en el coste. Con las ventajas que eso conlleva.

Sensibilidad con respecto a los índices de fiabilidad β_i . Se plantea el siguiente problema

$$\beta_i^* = \begin{array}{l} \text{Mínimo} \\ \mathbf{z}, \bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\kappa}^{\text{aux}} \end{array} \quad \sqrt{\mathbf{z}^T \mathbf{z}} \quad (10.138)$$

sujeto a

$$g_i(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\psi}) = 1 \quad (10.139)$$

$$\mathbf{h}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}) = \boldsymbol{\psi} \quad (10.140)$$

$$\mathbf{T}(\mathbf{d}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\kappa}^{\text{aux}}) = \mathbf{z} \quad (10.141)$$

$$\bar{\mathbf{d}}^{\text{aux}} = \bar{\mathbf{d}}^* : \boldsymbol{\lambda}_i \quad (10.142)$$

$$\tilde{\boldsymbol{\eta}}^{\text{aux}} = \tilde{\boldsymbol{\eta}} : \boldsymbol{\vartheta}_i \quad (10.143)$$

$$\tilde{\boldsymbol{\phi}}^{\text{aux}} = \tilde{\boldsymbol{\phi}} : \boldsymbol{\varphi}_i \quad (10.144)$$

$$\boldsymbol{\kappa}^{\text{aux}} = \boldsymbol{\kappa} : \boldsymbol{\Upsilon}_i \quad (10.145)$$

de tal forma que se obtienen las derivadas $\boldsymbol{\lambda}_i$ del índice de fiabilidad β_i con respecto a los valores medios o característicos óptimos de las variables de diseño $\bar{\mathbf{d}}^*$, las sensibilidades $\boldsymbol{\vartheta}_i$ con respecto a los valores característicos de los parámetros $\tilde{\boldsymbol{\eta}}$, las derivadas parciales $\boldsymbol{\varphi}_i$ con respecto a los valores característicos de los agentes $\tilde{\boldsymbol{\phi}}$ y por último las sensibilidades $\boldsymbol{\Upsilon}_i$ de los parámetros estadísticos $\boldsymbol{\kappa}$. Estos valores son extremadamente útiles para el proyectista porque sabe en todo momento la influencia que tienen todas las variables de proyecto en cada uno de los modos de fallo, y puede saber cuáles requieren mayor control, o que variables ha de modificar para cambiar la fiabilidad con respecto al modo de fallo considerado. Estas sensibilidades nos dan la influencia de la parámetros en la probabilidad de fallo, y son las sensibilidades del vector Θ en el cálculo de la probabilidad de fallo condicionada dada por la ecuación (??).

Nótese que los valores de $\bar{\mathbf{d}}^{\text{aux}}$, $\tilde{\boldsymbol{\eta}}^{\text{aux}}$, $\tilde{\boldsymbol{\phi}}^{\text{aux}}$ están implícitos en la transformación de Rosenblatt (10.141), y que la solución del modelo dará los puntos de máxima verosimilitud ya conocidos $\bar{\mathbf{d}}^*$, $\boldsymbol{\eta}_i^*$ y $\boldsymbol{\phi}^*$.

Si se quisieran calcular las mismas sensibilidades con el método de aproximación por hiperplanos, sólo se tendría que modificar el problema clásico, que quedaría de la siguiente manera:

Sensibilidad en el modelo mixto global. Se plantea el siguiente problema auxiliar

$$\text{Minimizar}_{\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\kappa}^{\text{aux}}} c(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\psi}) \quad (10.146)$$

sujeto a

$$g_i(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\psi}) \geq F_i^0 : \Gamma_i; \forall i \in I \quad (10.147)$$

$$\beta_i^* + \boldsymbol{\lambda}_i^T (\bar{\mathbf{d}}^{\text{aux}} - \bar{\mathbf{d}}^*) + \boldsymbol{\vartheta}_i^T (\tilde{\boldsymbol{\eta}}^{\text{aux}} - \tilde{\boldsymbol{\eta}}) + \boldsymbol{\varphi}_i^T (\tilde{\boldsymbol{\phi}}^{\text{aux}} - \tilde{\boldsymbol{\phi}}) + \boldsymbol{\Upsilon}_i^T (\boldsymbol{\kappa}^{\text{aux}} - \boldsymbol{\kappa}) \geq \beta_i^0; \Lambda_i; \forall i \in I \quad (10.148)$$

$$r_j(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\psi}) \leq 0; \forall j \in J \quad (10.149)$$

$$\mathbf{h}(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}) = \boldsymbol{\psi} \quad (10.150)$$

$$\tilde{\boldsymbol{\eta}}^{\text{aux}} = \tilde{\boldsymbol{\eta}} : \boldsymbol{\varsigma} \quad (10.151)$$

$$\tilde{\boldsymbol{\phi}}^{\text{aux}} = \tilde{\boldsymbol{\phi}} : \boldsymbol{\rho} \quad (10.152)$$

$$\boldsymbol{\kappa}^{\text{aux}} = \boldsymbol{\kappa} : \boldsymbol{\nu} \quad (10.153)$$

en el que las restricciones (10.131) y (10.132) se han sustituido por la aproximación del hiperplano (10.148). Nótese que las pendientes del hiperplano son precisamente las variables duales de las restricciones (10.142), (10.143), (10.144) y (10.145) ($\boldsymbol{\lambda}_i$, $\boldsymbol{\vartheta}_i$, $\boldsymbol{\varphi}_i$ y $\boldsymbol{\Upsilon}_i$). Lógicamente los valores de las sensibilidades por ambos métodos son los mismos.

Illustrative Example 10.3 (Sensibilidad en el ejemplo del muro mediante el método de relajación). Aplicando el estudio de sensibilidad al muro del ejemplo ilustrativo 10.1, se obtienen los siguientes resultados mostrados en la tabla 10.5.

Table 10.5: Sensibilidades en el ejemplo ilustrativo del muro.

x	$\frac{\partial c}{\partial x}$	$\frac{\partial \beta_t}{\partial x}$	$\frac{\partial \beta_s}{\partial x}$	$\frac{\partial \beta_b}{\partial x}$
\bar{a}	--	6.095	0.950	0.000
\bar{b}	--	1.521	0.475	-1.440
$\bar{\nu}$	-105.047	0.000	16.339	0.000
\bar{T}	0.246	-0.057	-0.038	0.000
$\bar{\gamma}$	-0.813	0.418	0.127	-0.376
\bar{H}	0.000	-2.409	0.000	0.000
\bar{S}	0.000	0.000	0.000	0.062
σ_a	0.174	-3.294	-0.027	0.000
σ_b	0.043	-0.205	-0.007	-0.101
σ_ν	257.455	0.000	-40.045	0.000
σ_T	0.423	-0.430	-0.066	0.000
σ_γ	0.142	-0.713	-0.022	-0.318
σ_H	0.000	-10.295	0.000	0.000
σ_S	0.000	0.000	0.000	-0.297
F_o	0.000	--	--	--
F_s	0.000	--	--	--
F_b	0.000	--	--	--
β_0	0.000	--	--	--
β_s	6.429	--	--	--
β_b	0.000	--	--	--

De la que se obtiene una gran cantidad de información, por ejemplo, se puede comprobar cómo efectivamente la única restricción de seguridad activa es la asociada al índice de fiabilidad por deslizamiento β_s , y por tanto los parámetros estadísticos que más influencia (mayor sensibilidad) tienen en el coste son los asociados al deslizamiento, el valor medio del coeficiente de fricción $\bar{\nu}$ (-105.047) y su desviación típica σ_ν (257.455). En los modelos de fiabilidad también se comprueba cómo el aumento de la incertidumbre en las variables disminuye el índice de fiabilidad con el consiguiente aumento de la probabilidad de fallo (valores negativos en todas las sensibilidades), y cómo el aumento de los valores medios de las acciones estabilizadoras aumenta la fiabilidad mientras que el aumento de las desestabilizadoras la disminuye, así por ejemplo, el aumento del peso específico γ aumentan la fiabilidad con respecto a vuelco (0.418) y deslizamiento (0.127), mientras en el caso de hundimiento la disminuye (-0.376).

■

Para el caso del diseño mixto parcial (10.23)-(10.29), es decir, empleando coeficientes de seguridad parciales, la solución devolvería los coeficientes parciales óptimos γ^* y el problema

auxiliar global sería:

Sensibilidad en el modelo mixto parcial. Se plantea el siguiente problema auxiliar

$$\text{Minimizar}_{\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\kappa}^{\text{aux}}, \boldsymbol{\gamma}^{\text{aux}}} c(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\psi}) \quad (10.154)$$

sujeto a

$$g_i(\boldsymbol{\gamma}^{\text{aux}T}(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}), \boldsymbol{\psi}) \geq 1; \forall i \in I \quad (10.155)$$

$$\beta_i^* + \boldsymbol{\lambda}_i^T(\bar{\mathbf{d}}^{\text{aux}} - \bar{\mathbf{d}}^*) + \boldsymbol{\vartheta}_i^T(\tilde{\boldsymbol{\eta}}^{\text{aux}} - \tilde{\boldsymbol{\eta}}) + \boldsymbol{\varphi}_i^T(\tilde{\boldsymbol{\phi}}^{\text{aux}} - \tilde{\boldsymbol{\phi}}) + \boldsymbol{\Upsilon}_i^T(\boldsymbol{\kappa}^{\text{aux}} - \boldsymbol{\kappa}) \geq \beta_i^0; \Lambda_i; \forall i \in I \quad (10.156)$$

$$r_j(\boldsymbol{\gamma}^{\text{aux}T}(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}), \boldsymbol{\psi}) \leq 0; \forall j \in J \quad (10.157)$$

$$\mathbf{h}(\boldsymbol{\gamma}^{\text{aux}T}(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}})) = \boldsymbol{\psi} \quad (10.158)$$

$$\tilde{\boldsymbol{\eta}}^{\text{aux}} = \tilde{\boldsymbol{\eta}} : \boldsymbol{\varsigma} \quad (10.159)$$

$$\tilde{\boldsymbol{\phi}}^{\text{aux}} = \tilde{\boldsymbol{\phi}} : \boldsymbol{\rho} \quad (10.160)$$

$$\boldsymbol{\kappa}^{\text{aux}} = \boldsymbol{\kappa} : \boldsymbol{\nu} \quad (10.161)$$

$$\boldsymbol{\gamma}^{\text{aux}} = \boldsymbol{\gamma}^* : \boldsymbol{\tau} \quad (10.162)$$

con el que se obtienen las mismas sensibilidades que en el modelo mixto global pero sustituyendo las derivadas con respecto a los límites inferiores de los coeficientes de seguridad ($\boldsymbol{\Gamma}$) por las sensibilidades $\boldsymbol{\tau}$ con respecto a los valores óptimos de los coeficientes de seguridad parciales, que podrán ser iguales o no a los límites impuestos por la restricción (10.29).

Igualmente utiliza la aproximación por el hiperplano (10.156) para la restricción de complicación asociada a los índices de fiabilidad que emplea las variables duales $\boldsymbol{\lambda}_i, \boldsymbol{\vartheta}_i, \boldsymbol{\varphi}_i$ y $\boldsymbol{\Upsilon}_i$.

En el apéndice A.6 se presentan los códigos GAMS para el estudio de las sensibilidades aplicados al ejemplo del muro.

10.6.2 Sensitivity analysis in the general model

En el caso de querer obtener la sensibilidad de los modelos generales se han de resolver dos modelos auxiliares, uno asociado a los problemas de fiabilidad y que coincide con el (10.138)-(10.145) del apartado anterior, y otro para el modelo global de coste:

$$\text{Minimizar}_{\alpha_{cost}, \bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\kappa}^{\text{aux}}} \alpha_{cost} \quad (10.163)$$

sujeto a

$$\alpha_{cost} \geq C_{to}^* + \boldsymbol{\Omega}^T(\bar{\mathbf{d}}^{\text{aux}} - \bar{\mathbf{d}}^*) + \boldsymbol{\xi}^T(\tilde{\boldsymbol{\eta}}^{\text{aux}} - \tilde{\boldsymbol{\eta}}) + \boldsymbol{\delta}^T(\tilde{\boldsymbol{\phi}}^{\text{aux}} - \tilde{\boldsymbol{\phi}}) + \boldsymbol{\chi}^T(\boldsymbol{\kappa}^{\text{aux}} - \boldsymbol{\kappa}) \quad (10.164)$$

$$F_i^0 \leq g_i(\bar{\mathbf{d}}^{\text{aux}}, \tilde{\boldsymbol{\eta}}^{\text{aux}}, \tilde{\boldsymbol{\phi}}^{\text{aux}}, \boldsymbol{\psi}^{\text{aux}}) : \Gamma_i; \forall i \in I \quad (10.165)$$

$$\beta_i^0 \leq \beta_i^* + \boldsymbol{\lambda}_i^T(\bar{\mathbf{d}}^{\text{aux}} - \bar{\mathbf{d}}^*) + \boldsymbol{\vartheta}_i^T(\tilde{\boldsymbol{\eta}}^{\text{aux}} - \tilde{\boldsymbol{\eta}})$$

$$+ \varphi_i^T (\tilde{\phi}^{\text{aux}} - \tilde{\phi}) + \Upsilon_i^T (\kappa^{\text{aux}} - \kappa) : \zeta_i; \forall i \in I \quad (10.166)$$

$$0 \geq r_j(\bar{d}^{\text{aux}}, \tilde{\eta}^{\text{aux}}, \tilde{\phi}^{\text{aux}}, \psi); \forall j \in J \quad (10.167)$$

$$\psi = \mathbf{h}(\bar{d}^{\text{aux}}, \tilde{\eta}^{\text{aux}}, \tilde{\phi}^{\text{aux}}) \quad (10.168)$$

$$\tilde{\eta}^{\text{aux}} = \tilde{\eta} \quad (10.169)$$

$$\tilde{\phi}^{\text{aux}} = \tilde{\phi} \quad (10.170)$$

$$\kappa^{\text{aux}} = \kappa \quad (10.171)$$

de donde se obtienen las derivadas parciales del coste total Γ y ζ con respecto a los límites inferiores F^0 y β^0 , respectivamente. Las derivadas del coste total con respecto a los parámetros son:

$$\Omega = \frac{\partial C_{to}^* (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi, \kappa)}{\partial \bar{d}^*} = \quad (10.172)$$

$$= \frac{\partial C_{co} (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi)}{\partial \bar{d}^*} + \sum_{\forall i} \lambda_i \frac{\partial C_{\beta^*} (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi, \kappa)}{\partial \beta_i^*} \quad (10.173)$$

$$\xi = \frac{\partial C_{to}^* (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi, \kappa)}{\partial \tilde{\eta}} = \quad (10.174)$$

$$= \frac{\partial C_{co} (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi)}{\partial \tilde{\eta}} + \sum_{\forall i} \vartheta_i \frac{\partial C_{\beta^*} (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi, \kappa)}{\partial \beta_i^*} \quad (10.175)$$

$$\delta = \frac{\partial C_{to}^* (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi, \kappa)}{\partial \tilde{\phi}} = \quad (10.176)$$

$$= \frac{\partial C_{co} (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi)}{\partial \tilde{\phi}} + \sum_{\forall i} \varphi_i \frac{\partial C_{\beta^*} (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi, \kappa)}{\partial \beta_i^*} \quad (10.177)$$

$$\chi = \frac{\partial C_{to}^* (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi, \kappa)}{\partial \kappa} = \quad (10.178)$$

$$= \frac{\partial C_{co} (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi)}{\partial \kappa} + \sum_{\forall i} \Upsilon_i \frac{\partial C_{\beta^*} (\bar{d}^*, \tilde{\eta}, \tilde{\phi}, \psi, \kappa)}{\partial \beta_i^*}. \quad (10.179)$$

Se pueden obtener de esta forma las derivadas del coste global (en la vida útil) con respecto a los factores de diseño, variables de diseño (ϖ), parámetros (ς), agentes (ϱ) y parámetros estadísticos (ν).

10.7 Bridge Crane Design Using the Probability-Safety Factor Method. Sensitivity Analysis

Modern industrial requirements needs the application of equipment for handling large, heavy, or bulky objects. That is why there are engineers specializing in overhead mate-

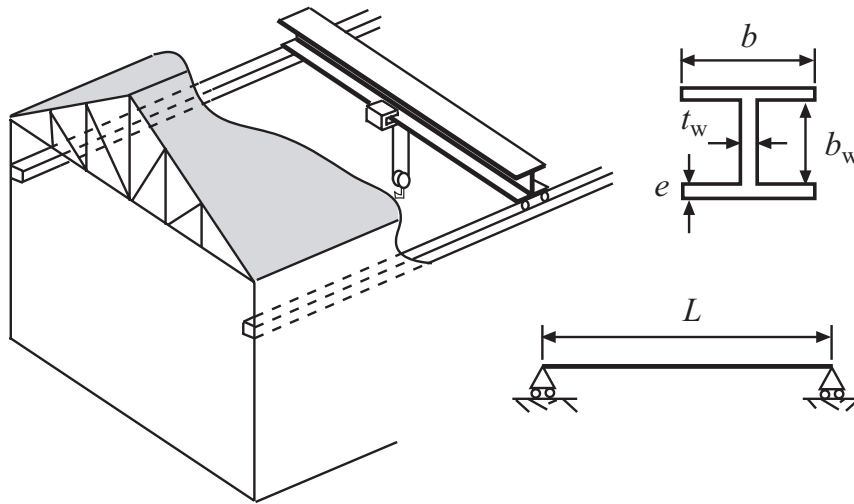


Figure 10.10: Graphical illustration of the bridge crane design and its cross section.

rial handling: bridge cranes, hoists, and monorails.

Focussing on crane bridges, an under running overhead crane with single girder is shown in Figure 10.10. All its structural elements then must be manufactured in accordance with current mandatory requirements of the National Safety and Health Act, OSHA Section 1910.179 and 1910.309 as applicable. Additionally, all ACECO cranes are manufactured in accordance with the appropriate standard of ANSI specifications, the National Electric Code, and the Crane Manufacturers Association of America (CMAA) specifications. Crane girders are designed using, structural steel beams (reinforced as necessary) or fabricated plate box sections. Bridge girder to end truck connections are designed for loadings, stresses and stability in accordance with current CMAA design specifications.

In this Section we will apply the Dual global design (10.17)-(10.22) (PSFM), for designing an overhead crane for an industrial nave (see Figure 10.10). In particular, we calculate the bridge girder dimensions that allow trolley travelling horizontally. It should be a box section fabricated from plate of structural steel, for the web, top plate and bottom plate, so as to provide for maximum strength at minimum dead weight. Maximum allowable vertical girder deflection should be a function of span.

Consider the girder and the cross section shown in Figure 10.10, where L is the span or distance from centerline to centerline of runway rails, b and e are the flange width and thickness, respectively and h_w and t_w are the web height and thickness, respectively.

The set of variables involved in the problem can be partitioned into four subsets:

- d:** *Optimization design variables.* They are the design variables which values are to be chosen by the optimization program to optimize the objective function (minimize the cost). In this crane example these variables are (see Figure 10.10):

$$\mathbf{d} = \{b, e, t_w, h_w\}$$

- η :** *Non-optimization design variables.* They are the set of variables which mean or char-

acteristic values are fixed by the engineer or the code and must be given as data to the optimization program. In this bridge girder example:

$$\boldsymbol{\eta} = \{f_y, L, \gamma_y, E, \nu, c_y\}$$

where f_y is the elastic limit of structural steel, L is the length of the span, γ_y is the steel unit weight, E is the Young modulus of the steel, ν is the Poisson modulus and c_y is the steel cost.

ϕ : *Agents*. Agent is any entity, as wind, waves, earthquakes, etc., that can act on or significantly affect the safety, serviceability, and use of a structure and its environment. In this case the ones associated with the maximum load supported by the girder

$$\phi = \{P\}$$

κ : *Random model parameters*. They are the set of parameters defining the random variability and dependence structure of the variables involved. In our example:

$$\boldsymbol{\kappa} = \{\sigma_{f_y}, \sigma_P, \sigma_L, \sigma_{\gamma_y}\}$$

where σ refers to the standard deviation of the corresponding variable.

ψ : *Auxiliary or non-basic variables*. They are auxiliary variables which values can be obtained from the basic variables \boldsymbol{d} and $\boldsymbol{\eta}$, using some formulas. In the example:

$$\boldsymbol{\psi} = \{W, I_{xx}, I_{yy}, I_t, G, \sigma, \tau, M_{cr}, \delta\}$$

where its meaning is described later.

In the classical approach the safety factors are used as constraints and the variables are assumed to be deterministic, i.e., the mean or characteristic (extreme percentiles) values of the variables are considered.

Assume that the following four failure modes are considered (see Figure 10.11):

1. *Maximum allowed deflection*. The maximum deflection safety factor F_d is defined (see Figure 10.11(a)) as

$$F_d = \frac{\delta_{\max}}{\delta} \quad (10.180)$$

where δ is the maximum deflection on the center of the girder, and δ_{\max} is the maximum deflection allowed by codes.

2. *Damage limit state of the steel upper and lower flanges*. We consider the ratio of the actual strength to actual stresses as

$$F_u = \frac{f_y}{\sqrt{\sigma^2 + 3\tau^2}} \quad (10.181)$$

where F_u is the corresponding safety factor, and σ and τ are the normal and tangential stresses at the center of the beam, respectively.

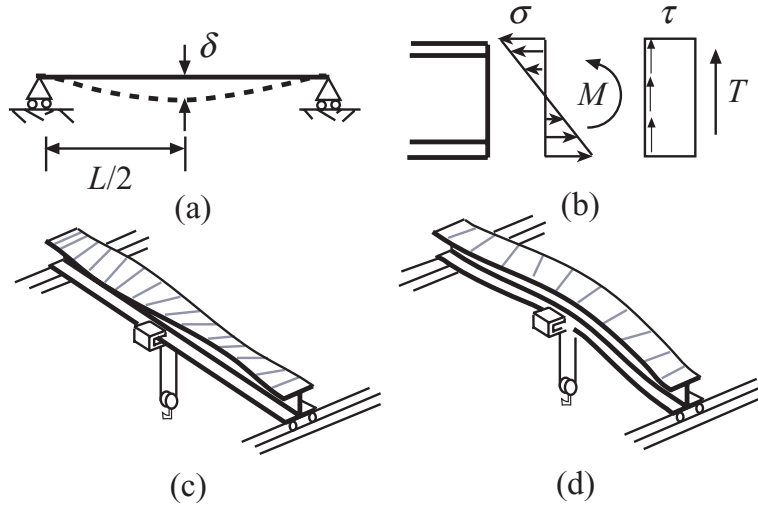


Figure 10.11: Illustration of the bridge girder modes of failure.

3. *Damage limit state of the steel web.* The bearing capacity safety factor F_w is the ratio of the shear strength capacity to actual shear stress at the center of the beam

$$F_w = \frac{f_y}{\sqrt{3}\tau} \quad (10.182)$$

4. *Global Buckling.* The global buckling safety factor F_b is the ratio of the critical moment against buckling of the cross section to the maximum moment applied at the center of the beam

$$F_b = \frac{M_{cr}}{M} \quad (10.183)$$

The girder bridge is safe if and only if F_d, F_u, F_w and $F_b \geq 1$.

10.7.1 Obtaining some constraints

In the following subsection we give full detail of how to obtain the constraints for the girder bridge.

Geometrical and mechanical properties of the girder

The moments of inertia I_{xx} and I_{yy} are obtained as

$$I_{xx} = \frac{1}{12} (b(h_w + 2e)^3 - (b - t_w)h_w^3) \quad (10.184)$$

$$I_{yy} = \frac{1}{12} (2eb^3 + h_w t_w^3) \quad (10.185)$$

whereas the torsional moment of inertia is obtained using

$$I_t = \frac{1}{3} (2be^3 - h_w t_w^3) \quad (10.186)$$

The deflection at the center of the beam is calculated using:

$$\delta = \frac{PL^3}{48EI_{xx}} + \frac{5WL^4}{384EI_{xx}} \quad (10.187)$$

where W is the girder bridge weight per unit length

$$W = \gamma_s(2eb + t_w h_w) \quad (10.188)$$

The stresses at the center of the beam are calculated considering:

$$T = P/2 \quad (10.189)$$

$$M = PL/4 \quad (10.190)$$

where T and M are the shear force and moment, respectively. Thus,

$$\sigma = \frac{M(h_w + e)}{2I_{xx}} \quad (10.191)$$

$$\tau = \frac{T}{h_w t_w} \quad (10.192)$$

The critical moment for global buckling is

$$M_{cr} = \frac{\pi}{L} \sqrt{EG I_{yy} I_t} \quad (10.193)$$

with the auxiliary parameter

$$G = \frac{E}{2(1 + \nu)}$$

Así,

$$\sigma = \frac{M(h_w + e)}{2I_{xx}} \quad (10.194)$$

$$\tau = \frac{T}{h_w t_w} \quad (10.195)$$

El momento crítico frente al pandeo lateral es

$$M_{cr} = \frac{\pi}{L} \sqrt{EG I_{yy} I_t} \quad (10.196)$$

con el parámetro auxiliar

$$G = \frac{E}{2(1 + \nu)}.$$

Code and other requirements

The following constraints are fixed by the codes.

The steel thickness must satisfy

$$0.008 \leq e \leq 0.15 \quad (10.197)$$

$$0.008 \leq t_w \leq 0.15 \quad (10.198)$$

and the maximum deflection allowed is

$$\delta_{\max} = L/888$$

To avoid local buckling (see Figure 10.11(c)) the design satisfy the following restriction

$$\frac{b}{2e} \leq 15 \sqrt{\frac{276}{f_y}} \quad (10.199)$$

where f_y is the steel strength in *MPa*.

To support the trolley, unit which travels on the bottom flange of the bridge girder and carries the hoist, the minimum flange width is 0.30 *m*.

10.7.2 Numerical example

To perform a probabilistic design in the bridge girder example, the joint probability density of all variables is required. Assume for example that all the variables involved are independent normal random variables, i.e.,

$$f_y \sim N(\mu_{f_y}, \sigma_{f_y}); \quad P \sim N(\mu_P, \sigma_P); \quad L \sim N(\mu_L, \sigma_L); \quad \gamma_y \sim N(\mu_{\gamma_y}, \sigma_{\gamma_y})$$

The means are:

$$\mu_{f_y} = 400 \text{ MPa}; \quad \mu_P = 600 \text{ kN}; \quad \mu_L = 6 \text{ m}; \quad \mu_{\gamma_y} = 78.5 \text{ kN/m}^3$$

and the standard deviations:

$$\sigma_{f_y} = 251 \text{ MPa}; \quad \sigma_P = 90 \text{ kN}; \quad \sigma_L = 0.05; \quad \sigma_{\gamma_y} = 0.785 \text{ kN/m}^3$$

The constant parameters are:

$$E = 210000 \text{ MPa}; \quad \nu = 0.3; \quad c_y = 0.24 \text{ \$/kN}$$

Assume also that the required safety factors and reliability bounds are:

$$F_d^0 = 1.15; \quad F_u^0 = 1.5; \quad F_w^0 = 1.5; \quad F_b^0 = 1.3$$

and

$$\beta_d^0 = 1.5; \quad \beta_u^0 = 3.7; \quad \beta_w^0 = 3.7; \quad \beta_b^0 = 3.2$$

Note that “violation” of limit states with more serious consequences are associated with higher reliability indices.

Using the Rosenblatt [146] transformation, this set is transformed into a set of standard normal random variables z_1, z_2, \dots, z_4 by

$$z_1 = \frac{f_y - \mu_{f_y}}{\sigma_{f_y}}; \quad z_2 = \frac{P - \mu_P}{\sigma_P}; \quad z_3 = \frac{L - \mu_L}{\sigma_L}; \quad z_4 = \frac{\gamma_y - \mu_{\gamma_y}}{\sigma_{\gamma_y}} \quad (10.200)$$

Then using the proposed method, the results shown in Table 10.6 are obtained. It shows the progress and convergence of the process, that is attained after 7 iterations. The first iteration column shows the values of the design variables, and the actual safety factors and failure β -values, associated with the optimal classical design for the safety factors. Note that no lower safety factor bound is active. Note also that the β_b and β_d constraints (boldfaced in Table 10.6) are active. Then, the associated safety margins are increased until all constraints hold.

The last column of the table shows the design values of the design variables

$$b, e, t_w, h_w$$

together with the safety factors and associated β -values.

The active values appear boldfaced in Table 10.6, from which the following conclusions can be drawn.

1. The process converges in only 7 iterations.
2. The list of actual safety factors and β -reliability indices.
3. No safety factor is active.
4. Due to the terribly strict constraints imposed by the serviceability limit state constraint (maximum deflection), and global buckling the probability bounds β_b and β_d are active.
5. The final design (iteration 7) is more expensive than the initial design (iteration 1), because this does not satisfy the β_b and β_d constraints.

The sensitivities for the composite beam example are given in Table 10.7 that gives the cost sensitivities associated with the optimal classical design. It allows to know how much a small change in a single data value changes the total cost of the composite beam. This information is extremely useful during the construction process to control the cost, and for analyzing how the changes in the safety factors required by the codes influence the total cost of engineering works. For example, a change of one dollar in the unit cost c_y of the steel leads to a cost increase of 9842.876 euros (see the corresponding entry in Table 10.7). Similarly, an increase in the safety factor lower bound F_d does not change the cost, and an increase of one unit in the bridge span leads to an increase of the cost of 746.662 euros.

It gives the sensitivities associated with the β -values too. It is useful to know how much a small change in a single data value changes the corresponding β -value, for example, the means, standard deviations, etc. In this table the designer can easily analyze how the quality of the material (reduced standard deviations in f_y) or precision in the applied loads (reduced standard deviations in P) influence the safety of the beam. Note that an increase in the dispersion (standard deviations or coefficients of variation) leads to a decrease of the β indices.

Table 10.6: Illustration of the iterative process.

ν	Units	1	2	3	4	5	6	7
$\text{cost}^{(\nu)}$	[\$]	2245.0	2243.2	2325.6	2354.8	2361.2	2362.2	2362.3
$b^{(\nu)}$	[cm]	41.81	30.00	36.14	39.25	40.43	40.67	40.68
$e^{(\nu)}$	[mm]	16.78	23.65	20.44	19.13	18.64	18.54	18.53
$t_w^{(\nu)}$	[mm]	8.00	8.00	8.00	8.00	8.00	8.00	8.00
$h_w^{(\nu)}$	[cm]	72.91	70.70	72.49	72.64	72.72	72.74	72.74
$F_u^{(\nu)}$	--	2.23	2.17	2.30	2.33	2.33	2.34	2.34
$F_t^{(\nu)}$	--	4.49	4.35	4.46	4.47	4.48	4.48	4.48
$F_b^{(\nu)}$	--	1.30	1.30	1.42	1.47	1.48	1.49	1.49
$F_d^{(\nu)}$	--	1.15	1.11	1.20	1.22	1.22	1.22	1.22
$\beta_u^{(\nu)}$	--	6.014	5.795	6.235	6.342	6.370	6.375	6.375
$\beta_t^{(\nu)}$	--	10.968	10.794	10.935	10.948	10.954	10.955	10.955
$\beta_b^{(\nu)}$	--	1.980	1.980	2.761	3.106	3.186	3.199	3.200
$\beta_d^{(\nu)}$	--	1.001	0.725	1.333	1.461	1.494	1.500	1.500
$\text{error}^{(\nu)}$	--	0.6627	0.3815	0.4563	0.1111	0.0253	0.0041	0.0002

Table 10.7: Sensitivities.

	∂cost	$\partial \beta_u$	$\partial \beta_t$	$\partial \beta_b$	$\partial \beta_d$
∂b	--	12.851	0.000	46.864	17.717
∂e	--	280.902	0.000	993.835	408.587
∂t_w	--	352.458	698.939	122.267	108.587
∂h_w	--	11.974	7.687	0.448	23.088
$\partial \mu_{f_y}$	0.000	0.000	0.000	0.000	0.000
$\partial \mu_P$	1.268	-0.008	-0.005	-0.011	-0.011
$\partial \mu_L$	746.662	-0.975	0.000	-3.218	-2.722
$\partial \mu_{\gamma_y}$	30.125	0.000	0.000	0.000	-0.001
$\partial \sigma_{f_y}$	0.000	0.000	0.000	0.000	0.000
$\partial \sigma_P$	3.312	-0.036	-0.027	-0.035	-0.016
$\partial \sigma_L$	149.935	-0.303	0.000	-1.657	-0.556
$\partial \sigma_{\gamma_y}$	0.000	0.000	0.000	0.000	0.000
∂E	0.000	0.000	0.000	0.000	0.000
$\partial \nu$	290.378	0.000	0.000	-3.730	0.000
∂c_y	9842.876	0.000	0.000	0.000	0.000
∂F_u	0.000	--	--	--	--
∂F_t	0.000	--	--	--	--
∂F_b	0.000	--	--	--	--
∂F_d	0.000	--	--	--	--
$\partial \beta_u$	0.000	--	--	--	--
$\partial \beta_t$	0.000	--	--	--	--
$\partial \beta_b$	77.858	--	--	--	--
$\partial \beta_d$	37.611	--	--	--	--

Chapter 11

Applications

11.1 Retaining Wall

In this Section the Reliability-Based Design of a retaining wall is performed.

11.1.1 Introduction to the retaining wall design

Retaining structures are designed to hold back soil where an abrupt change in ground elevation occurs. The retained material or backfill exerts a push on the structure and thus tends to overturn or slide it, or both. The stem, heel and toe of such a wall act as cantilever beams, which need to be designed to withstand the soil pressures. Thus, a wall design implies:

1. Performing the stability check for the structure.
2. Computing the maximum and minimum soil pressures present under the toe and heel, comparing them with the allowable soil pressure provided as data.
3. Designing the reinforcing steel for the toe, heel and stem considering the corresponding bending and shear.

There are several well-known commercial computer programs, such as Correct Surcharge, DDRW-1, EPRES, FREW, GRETA, GWALL, HEAVE, Kzero2, LPRES, RETAIN, Retaining Wall Design, RetWall, ReWaRD, SHEET, Sheetpile-2, SHORING, SPUNT-A2, SPW 911, STAWALL, UNIBEAR, WALLAP, etc., that compute the soil bearing pressures under the base of a wall supporting any kind of backfill material with additional surcharge and concentrated external loads acting on the wall. In addition, they analyze the stability of the whole structure and perform the concrete design based on several design methods, in turn based on the working or the ultimate strength.

These programs normally offer the option of considering other external loads applied to the wall in addition to the backfill pressure. Though a surcharge can be defined as an equivalent height of backfill, they calculate earth pressures at key locations down the retaining wall, bending moments, shear forces, wall displacements, ground settlements, etc., and implement safety factors.

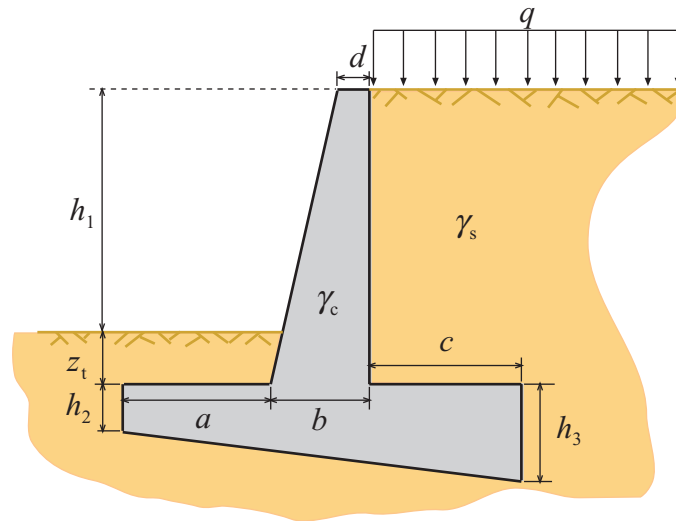


Figure 11.1: Geometrical description of the wall.

Some computer programs additionally design retaining walls according to certain international standards, such as BS 8002, Eurocode 7, CIRIA 104, Hong Kong Geoguide, CP2, British Steel's Piling Handbook and Highway Agency's BD42/94.

Most existing software for wall design uses the classical design based on safety factors, i.e. the design corresponds to level I. In other words, no random variables are allowed and, consequently, no probability of failure is obtained. Nowadays however the level II and III procedures have been substantially developed, and it is highly recommendable to use them.

An additional problem of these programs is that the design parameters, such as the toe or heel lengths, the stem widths, etc. have to be provided by the engineer, and an acceptable design of the wall is obtained only after a trial and error procedure. Once a wall design satisfying all the constraints is obtained, it is not easy to know whether we are far from or close to the optimal design. In this section we deal with an optimal design that allows an automatic design of the wall.

11.1.2 Description of the retaining wall

A retaining wall problem involves many variables, such as all wall dimensions, backfill slope, concrete and steel strength, allowable soil bearing pressure, backfill properties, etc. For example, consider the wall in Figure 11.1 that defines the geometry of the wall (note that the wall is defined in parametric form).

1. Geometric data.

h_1 : height of the wall.

2. Acting agents data.

q : surcharge.

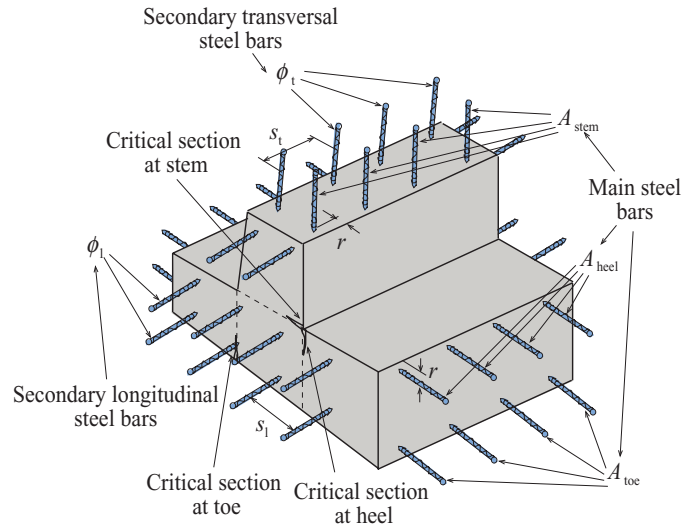


Figure 11.2: Illustration of the reinforced bar positions.

3. Material definition data.

γ_c : unit weight of the concrete.

γ_s : unit weight of the soil.

γ_{st} : unit weight of steel.

f_c : concrete design strength.

f_y : steel design strength.

k_a : active pressure coefficient.

k_p : passive pressure coefficient.

σ_{soil} : soil strength.

μ_{crit} : critical friction factor.

τ_{max} : concrete shear strength.

4. Steel bar data. See Figure 11.2, which gives the details of the bar reinforcement.

r : bar cover depth.

ϕ_ℓ : secondary longitudinal steel bar diameter (minimum).

ϕ_t : secondary transversal steel bar diameter (minimum).

s_ℓ : secondary longitudinal bar spacing.

s_t : secondary transversal bar spacing.

5. Mode safety factors lower bounds.

F_t^0 : safety factor against overturning.

- F_s^0 : safety factor against sliding.
 F_b^0 : safety factor for the bearing capacity constraint.
 F_{stem}^0 : safety factor for the bending moment at stem.
 F_{toe}^0 : safety factor for the bending moment at toe.
 F_{heel}^0 : safety factor for the bending moment at heel.
 F_{sstem}^0 : safety factor for the shear at stem.
 F_{stoe}^0 : safety factor for the shear at toe.
 F_{sheel}^0 : safety factor for the shear at heel.

6. Cost data.

- c_c : cost per cubic meter of concrete.
 c_{st} : cost per Newton of steel.
 c_t : cost per square meter of timber.
 c_{ex} : cost per cubic meter of excavation.

7. Design variables.

These are the variables whose mean values are automatically selected by the optimization procedure.

- a : toe length.
 c : heel length.
 b : stem width, bottom.
 d : stem width, top.
 z_t : soil cover.
 h_2 : toe height
 h_3 : heel height.
 A_{stem} : cross section of the reinforcing bars at stem.
 A_{toe} : cross section of the reinforcing bars at toe.
 A_{heel} : cross section of the reinforcing bars at heel.

11.1.3 Variable classification and failure modes considered

As the PFSM is going to be apply to the retaining wall example, the first step is the variable classification. It is important to distinguish between design values (those desired by the engineer), which in this paper are assumed to be the expectations, $E(X_i)$ or \bar{x}_i , or characteristic values, of the random variables $X_i : i = 1, 2, \dots, n$, and actual values x_i (those existing in reality). Some of these expectations are chosen by the engineer or the design codes, and some are selected by the optimization procedure to be presented. The set of actual values associated with the expectations chosen by the engineer will be denoted by the vector

$$\boldsymbol{\eta} = (h_1, q, \gamma_c, \gamma_s, \gamma_{st}, f_c, f_y, k_a, k_p, \sigma_{soil}, \mu_{crit}, \tau_{max}, r, \phi_\ell, \phi_t, s_\ell, s_t, c_c, c_{st}, c_t, c_{ex}),$$

and the set of actual values associated with the design variables, chosen by the optimization procedure, will be denoted by the vector

$$\mathbf{d} = (a, b, c, d, z_t, h_2, h_3, A_{stem}, A_{toe}, A_{heel}).$$

Note that vector $\boldsymbol{\eta}$ can be decomposed as in Section 10.2.1. The corresponding mean or characteristic vectors will be denoted $\tilde{\boldsymbol{\eta}}$ and $\bar{\mathbf{d}}$, respectively.

In this wall example, the following failure models are considered:

1. *Sliding failure mode.* The analysis of this failure can be done by considering the ratio of the stabilizing to the sliding forces as

$$g_s(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_s) = \frac{\text{stabilizing forces}}{\text{sliding forces}} - F_s \quad (11.1)$$

where F_s is the global safety coefficient associated with the sliding failure.

2. *Overturning failure mode.* The analysis of this failure can be done by considering the ratio of the stabilizing to the overturning moments with respect to some point, as

$$g_t(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_t) = \frac{\text{stabilizing moments}}{\text{overturning moments}} - F_t \quad (11.2)$$

where F_t is the safety coefficient associated with the overturning failure.

3. *Bearing capacity failure mode.* The analysis of this failure can be done by considering the ratio of the bearing capacity to the actual stress as

$$g_b(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_b) = \frac{\text{bearing capacity}}{\text{actual stress}} - F_b \quad (11.3)$$

where F_b is the safety coefficient associated with a foundation failure.

4. *Flexural failure at stem, toe and heel.* The analysis of this failure can be done by considering the ratio of the steel strength capacity to the actual steel stress as

$$g_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_j) = \frac{\text{steel strength capacity}}{\text{actual stress}} - F_j; \quad j = \text{stem, toe, heel} \quad (11.4)$$

where F_{stem} , F_{toe} and F_{heel} are the safety coefficients associated with the flexural failures at the corresponding sections.

5. *Shear strength failure at stem, toe and heel.* The analysis of this failure can be done by considering the ratio of the shear concrete strength capacity to the actual shear force as

$$g_j(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}; F_j) = \frac{\text{shear strength capacity}}{\text{actual shear force}} - F_j; \quad j = \text{sstem, stoe, sheel} \quad (11.5)$$

where F_{sstem} , F_{stoe} and F_{sheel} are the safety coefficients associated with the shear strength failures at the corresponding sections.

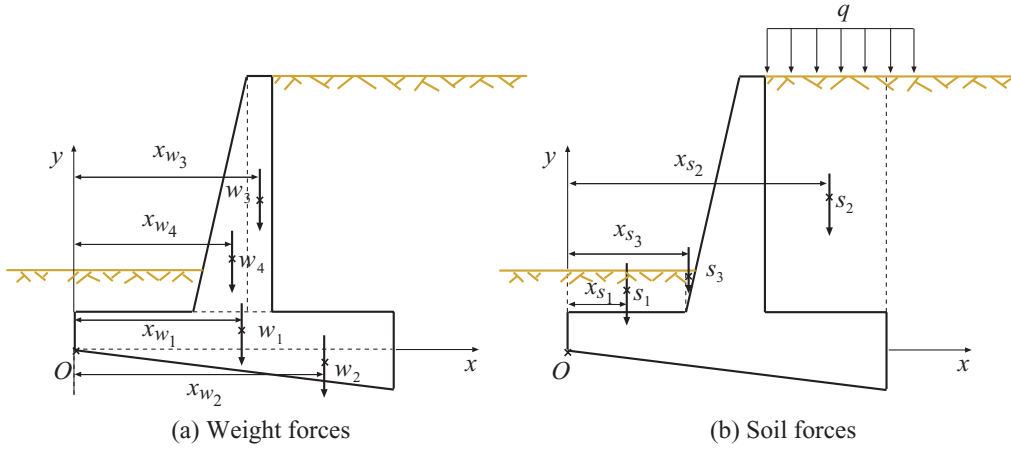


Figure 11.3: Decomposition of the cross section of the cantilever wall into simpler geometrical pieces (triangles and rectangles) to calculate: (a) the weights of the concrete, and (b) the soil weights.

Since, including the three flexural and three shear, we have 9 different failure modes, from now on, we define the set I_f of failure modes as

$$I_f = \{s, t, b, stem, toe, heel, sstem, stoe, sheel\}$$

The wall will be safe if and only if $F_s, F_t, F_b, F_{stem}, F_{toe}, F_{heel}, F_{sstem}, F_{stoe}, F_{sheel} \geq 1$.

Note that only the usually relevant failure modes have been considered, but other failures are also possible, as shallow or deep soil failures; however, for the sake of clarity, they have been ignored in this paper.

It is important to mention that we consider not the serviceability states, but the ultimate limit states.

11.1.4 Forces acting on the wall

Before deriving the set of constraints, we determine the total weight of the wall and soil, and the earth pressures acting on the wall, together with the corresponding points of application. Then, we state the constraints. To this end, we use intermediate or auxiliary variables that facilitate the work.

Note that optimization programs allow the use of these auxiliary variables, and that we need not to obtain an explicit expression for the constraints in terms of the design variables. This is an important advantage, because such a process is complicated.

Total weight of the wall. From Figure 11.3(a) we obtain the following wall weights and their corresponding locations:

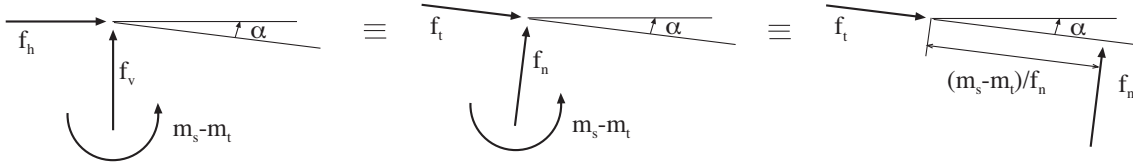


Figure 11.5: Ilustración de la localización de la reacción en la base del muro.

$$\begin{aligned}
 s_p &= k_p \gamma_s (h_2 + z_t)^2 / 2 \\
 y_{s_p} &= (h_2 + z_t) / 3 \\
 s_a &= k_a \gamma_s (h_1 + z_t + h_3)^2 / 2 + k_a q (h_1 + z_t + h_3) \\
 y_{s_a} &= k_a (h_1 + z_t + h_3)^2 (q + \gamma_s (h_1 + z_t + h_3) / 3) / (2s_a) - h_3 + h_2
 \end{aligned} \tag{11.8}$$

where s_p and s_a are the horizontal soil pressures, and y_{s_p} and y_{s_a} are the associated locations.

Resultant forces with respect to the toe. From Figure 11.4(b) we obtain:

$$f_h = s_a - s_p; \quad f_v = s + w; \quad m_t = s_a y_{s_a}; \quad m_s = s_p y_{s_p} + w x_w + s x_s, \tag{11.9}$$

where f_h and f_v are the horizontal and vertical components and m_t and m_s the overturning and stabilizing moments, respectively.

11.1.5 Design criteria and safety factor requirements

Basándose en los cálculos previos es fácil escribir las restricciones de estado límite.

Restricción de vuelco. Esta restricción establece que el ratio momento estabilizador, momento volcador ha de ser mayor que el coeficiente de seguridad correspondiente:

$$m_s / m_t \geq F_t \tag{11.10}$$

Restricción de deslizamiento. Esta restricción establece que el ratio coeficiente de rozamiento crítico, coeficiente de rozamiento existente ha de ser mayor que el coeficiente de seguridad correspondiente:

$$\mu_{\text{crit}} / \mu \geq F_s \tag{11.11}$$

donde, de la figura 11.5 se obtiene:

$$\begin{aligned}
 \alpha &= \arctan [(h_3 - h_2) / (a + b + c)] \\
 f_n &= f_v \cos \alpha + f_h \sin \alpha \\
 f_t &= f_h \cos \alpha - f_v \sin \alpha \\
 \mu &= f_t / f_n
 \end{aligned} \tag{11.12}$$

Restricción de la capacidad portante. Esta restricción establece que la máxima tensión (σ_i) en el extremo izquierdo del pie del muro ha de ser menor que la capacidad portante del terreno dividida por el coeficiente de seguridad correspondiente:

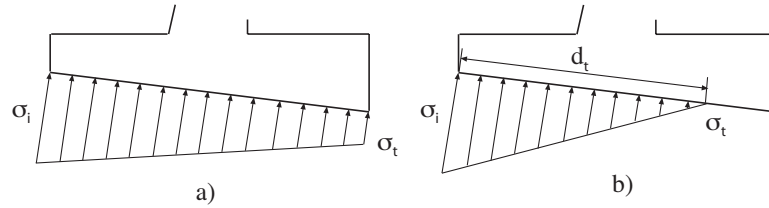


Figure 11.6: Presiones actuando en la base del muro. Caso a) ambas presiones son positivas. Caso b) σ_t es nula, y no existen tracciones.

$$\frac{\sigma_{\text{soil}}}{\sigma_i} \geq F_b \quad (11.13)$$

Donde, si $d_t = 3(m_s - m_t)/f_n \geq (a + b + c)/\cos \alpha$, de la figura 11.6(a) se obtiene

$$\begin{aligned} m_f &= (m_s - m_t) - f_n(a + b + c)/(2 \cos \alpha) \\ \sigma_i &= f_n \cos \alpha / (a + b + c) - 6m_f / ((a + b + c)/\cos \alpha)^2 \\ \sigma_t &= f_n \cos \alpha / (a + b + c) + 6m_f / ((a + b + c)/\cos \alpha)^2 \end{aligned} \quad (11.14)$$

y, de la figura 11.6(b) se obtiene

$$\begin{aligned} d_t &= 3(m_s - m_t)/f_n \\ \sigma_i &= 2f_n^2/(3(m_s - m_t)) \\ \sigma_t &= \sigma_i(1 - (a + b + c)/(d_t \cos(\alpha))). \end{aligned} \quad (11.15)$$

Como en el segundo caso σ_t es negativo, y este es un valor ficticio para calcular otras variables, ya que se considera que el terreno no resiste tracciones, se tomará como $\sigma_t = 0$.

Restricciones de flexión y esfuerzo cortante en el alzado, la puntera y el talón.

Estas restricciones establecen que no se producirá fallo ni por flexión, ni por esfuerzo cortante en las localizaciones correspondientes.

Lo primero que se hace es calcular la armadura que necesita una sección general sometida a una fuerza normal (n), un esfuerzo cortante (v) y un momento flector (m) para resistir (véase la figura 11.7):

$$\begin{aligned} \sum M_A = 0: & \quad f_c 0.8x(h - r - 0.4x) = m + n(h/2 - r) \\ \sum F_H = 0: & \quad t = f_c 0.8x - n \\ & \quad \frac{A_s f_y}{t} \geq F_k; \quad k = \text{stem, toe, heel} \\ & \quad \frac{\tau_{\text{max}}}{v/(h - r)} \geq F_j; \quad j = \text{sstem, stoe, sheel} \end{aligned} \quad (11.16)$$

Lo siguiente es particularizar para el alzado, la puntera y el talón, con lo que se obtiene:

Sección del alzado. De la figura 11.8(a) se obtiene

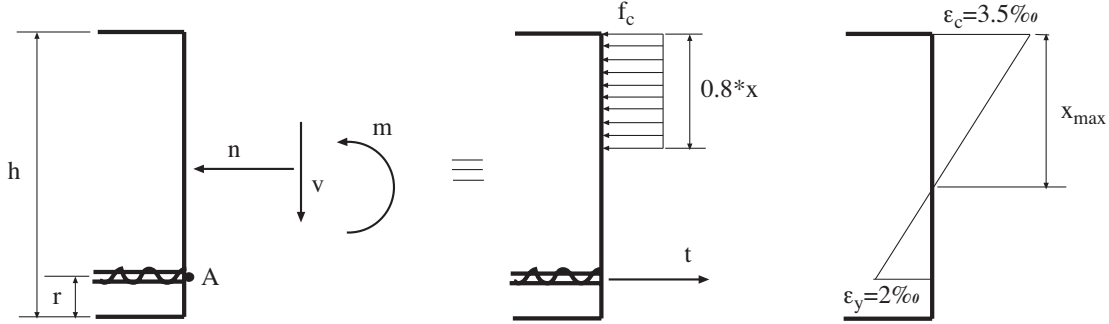


Figure 11.7: Ilustración del cálculo de las armaduras.

$$\begin{aligned}
 n_{stem} &= w_3 + w_4 + s_3 \\
 v_{stem} &= k_a q (h_1 + z_t) + k_a \gamma_s * (h_1 + z_t)^2 / 2 - k_p \gamma_s z t^2 / 2 \\
 m_{stem} &= k_a \gamma_s (h_1 + z t)^3 / 6 + k_a * q * (h_1 + z t)^2 / 2 - k_p \gamma_s z t^3 / 6 \\
 &\quad + n_{stem} (a + b / 2) - (\sum_{i=3}^4 w_i x_{w_i} + s_3 x_{s_3}) \\
 h &= b
 \end{aligned} \tag{11.17}$$

Sección de la puntera. De la figura 11.8(b) se obtienen

$$\begin{aligned}
 c_{aux} &= h_2 + a \tan \alpha \\
 d_{aux} &= a / \cos \alpha \\
 p_{aux} &= \gamma_c a (h_2 + c_{aux}) / 2 \\
 \sigma_x &= \sigma_i + a (\sigma_t - \sigma_i) / (a + b + c) \\
 n_{toe} &= k_p \gamma_s h_2 (z_t + h_2 / 2) + (\sin \alpha + \mu \cos \alpha) (\sigma_i + \sigma_x) d_{aux} / 2 \\
 v_{toe} &= (\cos \alpha - \mu \sin \alpha) (\sigma_i + \sigma_x) d_{aux} / 2 - p_{aux} - s_1 \\
 m_{toe} &= -a s_1 / 2 + k_p \gamma_s z_t h_2 (c_{aux} - 0.5 h_2) \\
 &\quad + k_p \gamma_s h_2^2 (c_{aux} - 2 h_2 / 3) / 2 + \sigma_i d_{aux}^2 / 2 + (\sigma_x - \sigma_i) d_{aux}^2 / 3 \\
 &\quad - n_{puntera} c_{aux} / 2 - \gamma_c a^2 h_2 / 2 - \gamma_c (c_{aux} - h_2) a^2 / 6 \\
 h &= c_{aux}
 \end{aligned} \tag{11.18}$$

Sección del talón. La figura 11.8(c) lleva a

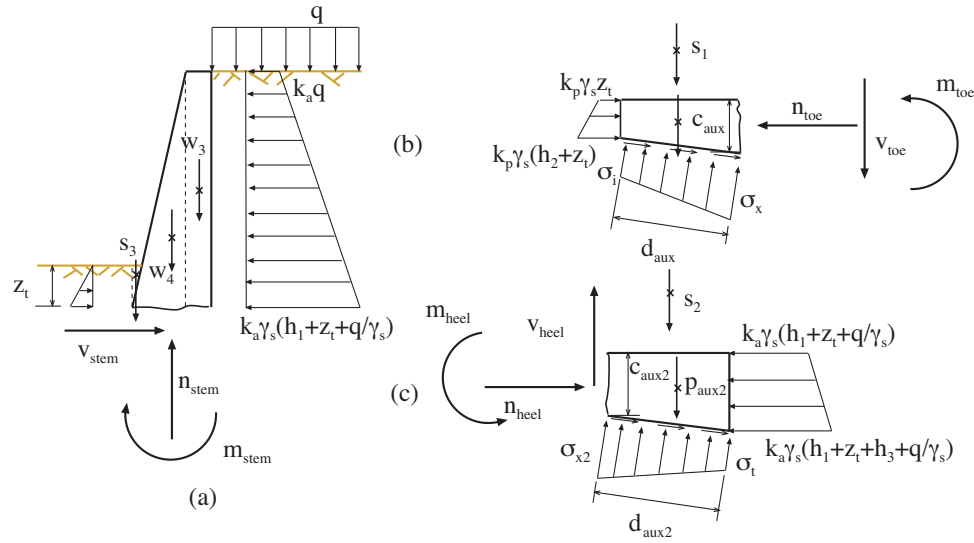


Figure 11.8: Ilustración de las fuerzas que intervienen en el cálculo de las armaduras en: (a) la sección crítica del alzado, (b) sección crítica de la puntera y (c) en la sección crítica del talón.

$$\begin{aligned}
 c_{aux2} &= h_2 + (a + b) \tan \alpha \\
 d_{aux2} &= c / \cos \alpha \\
 p_{aux2} &= \gamma_c c (h_3 + c_{aux2}) / 2 \\
 \sigma_{x2} &= \sigma_i + (a + b) (\sigma_t - \sigma_i) / (a + b + c) \\
 n_{heel} &= k_a \gamma_s h_3 (h_1 + z_t + q / \gamma_s + h_3 / 2) - (\sin \alpha + \mu \cos \alpha) (\sigma_i + \sigma_{x2}) d_{aux2} / 2 \\
 v_{heel} &= -(\cos \alpha - \mu \sin \alpha) (\sigma_i + \sigma_{x2}) d_{aux2} / 2 + p_{aux2} - s_2 \\
 m_{heel} &= s_2 c / 2 - k_a \gamma_s (h_1 + z_t + q / \gamma_s) h_3 (c_{aux2} - h_3 / 2) \\
 &\quad - 0.5 k_a \gamma_s h_3^2 (c_{aux2} - 2h_3 / 3) - \sigma_t d_{aux2}^2 / 2 \\
 &\quad - 0.5 (\sigma_{x2} - \sigma_t) d_{aux2}^2 / 3 + n_{heel} c_{aux2} / 2 \\
 &\quad + \gamma_c c^2 a_{ux2} / 2 + \gamma_c (h_3 - c_{aux2}) c^2 / 2 \\
 h &= c_{aux2}
 \end{aligned} \tag{11.19}$$

Reemplazando estos valores en (11.16) se obtienen las secciones transversales de armadura necesarias para el alzado, A_{stem} , la puntera, A_{toe} , y el talón, A_{heel} .

Razones constructivas. Estas restricciones se deben a motivos constructivos, ya que el proceso de optimización tiende a que los espesores en la puntera y en lo alto del alzado sean nulos.

$$d \geq 0.3 \tag{11.20}$$

$$z_t \geq 0.3 \tag{11.21}$$

11.1.6 Objective function

En esta sección se calcula el coste total del muro por metro lineal, incluyendo coste de hormigón, acero, excavación y encofrados:

$$Cost = h(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}) = v_c c_c + s_t c_t + w_{st} c_{st} + v_{ex} c_{ex}$$

Como éste es un ejemplo ilustrativo, y para mayor simplicidad, no se considera la vida útil del muro, ni los costes de reparación y mantenimiento.

Para tal fin, se deben calcular los volúmenes de hormigón, v_c , excavación, v_{ex} , la superficie total de encofrado, s_t , y el peso total de acero para las armaduras, w_{st} .

$$\begin{aligned} v_c &= w/\gamma_c \\ s_t &= h_2 + h_3 + h_1 + z_t + \sqrt{(b-d)^2 + h_1^2} \\ w_{st} &= \gamma_{st} (A_1(h_1 + z_t + c_{aux2} + 0.5) + A_2(d_{aux} + d_{aux2} + d_{aux3}) + A_3(a + b + c)) \\ &\quad + (h_2 + h_3 + d + \sqrt{(b-d)^2 + (h_1 + z_t)^2} + 4)\pi\phi_t^2\gamma_{st}/(4s_t) + \phi_l^2\gamma_{st}/(4s_l)(h_2 + \\ &\quad + \sqrt{(b-d)^2 + (h_1 + z_t)^2} + d + h_1 + z_t + c + h_3 + d_{aux} + d_{aux2} + d_{aux3} + a) \\ v_{ex} &= v_c + (h_1 + z_t)c + z_t a + (h_1 + z_t + h_3)^2/2 + (h_2 + z_t)^2/2 \end{aligned} \quad (11.22)$$

donde $\gamma_{st} = 78.5kN/m^3$.

11.1.7 Distributional assumptions of the model

El diseño de un muro mediante métodos de Nivel II y III requiere la definición de las propiedades estadísticas de las variables que intervienen. En este ejemplo todas las variables, geométricas, cargas externas, y definiciones de material se toman como normales e independientes, cuyos valores de diseño se corresponden con sus medias, a las que se hará referencia con una barra sobre la variable correspondiente, y con las desviaciones típicas mostradas en la tabla 11.1. Se estará empleando, por tanto el método FOSM descrito en la sección 4.6.

Para mayor simplicidad, en este ejemplo, el siguiente conjunto de variables se asumirá determinista:

$$\{r, \phi_l, \phi_t, s_l, s_t, c_c, c_{st}, c_t, c_{ex}\}$$

11.1.8 Numerical solution and discussion

Supóngase el siguiente muro vertical:

$$\begin{array}{lllll} \bar{h}_1 = 5m; & \bar{\gamma}_c = 25kN/m^3; & \bar{\gamma}_s = 20kN/m^3; & \bar{f}_c = 30MPa; & \bar{f}_y = 450MPa; \\ \bar{k}_a = 0.5; & \bar{k}_p = 3; & \bar{\nu}_{crit} = 0.5; & \bar{\sigma}_{soil} = 0.25MPa; & \bar{q} = 40kN/m; \\ r = 0.05m; & F_t^0 = 1.2; & F_s^0 = 1.15; & F_b^0 = 1.2; & F_{stem}^0 = 1.4; \\ F_{toe}^0 = 1.4; & F_{heel}^0 = 1.4; & F_{stem}^0 = 1.2; & F_{stoe}^0 = 1.2; & F_{sheel}^0 = 1.2; \\ c_c = 66euro/m^3; & c_{st} = 0.06euro/N; & c_t = 12euro/m^3; & c_{ex} = 3.6euro/m^3; & \tau_{max} = 0.4MPa, \end{array}$$

donde éstos son los datos elegidos por el diseñador.

El ejemplo propuesto ha sido implementado en GAMS (General Algebraic Modeling System) (véase Castillo, Conejo, Pedregal, García and Alguacil [30]).

Table 11.1: Distribución estadística de las variables del ejemplo del muro.

Tipo	Variable	Desviación típica	Tipo	Variable	Desviación típica
Variables geométricas	a	$0.025m$	Resistencias	f_c	$3MPa$
	b	$0.025m$		f_y	$22MPa$
	c	$0.025m$		σ_{soil}	$0.25MPa$
	d	$0.025m$		τ_{max}	$0.04MPa$
	z_t	$0.025m$		Pesos específicos	γ_c
	h_1	$0.025m$	γ_s		$2kN/m^3$
	h_2	$0.025m$	γ_{st}		$0.2kN/m^3$
	h_3	$0.025m$	Otras variables	k_a	0.05
	A_{stem}	$0.00001m^2$		k_p	0.1
	A_{toe}	$0.00001m^2$		ν_{crit}	0.05
	A_{heel}	$0.00001m^2$		q	$8kN/m$

La tabla 11.2 muestra la convergencia del método, que es alcanzada tras 11 iteraciones. La última columna de la tabla muestra los valores de diseño óptimos de las variables de diseño $a, b, c, d, h_2, h_3, z_t, A_{stem}, A_{toe}$ y A_{heel} , junto con los coeficientes de seguridad y los correspondientes valores del índice de fiabilidad β . El diseño se realizó garantizando los siguientes valores de los coeficientes de seguridad clásicos $F_t \geq 2.0; F_s \geq 2.0; F_b \geq 2.0; F_{stem} \geq 1.3; F_{toe} \geq 1.3; F_{heel} \geq 1.3; F_{sstem} \geq 1.4; F_{stoe} \geq 1.4; F_{sheel} \geq 1.4$ y $\beta_t, \beta_s, \beta_b, \beta_{stem}, \beta_{toe}, \beta_{heel}, \beta_{sstem}, \beta_{stoe}$ y β_{sheel} mayores o iguales a $\beta_i^0 = 3.71$. Los valores activos aparecen subrayados en la tabla 11.2.

De los resultados se pueden obtener las siguientes conclusiones:

1. El proceso converge en 11 iteraciones, pero prácticamente se obtienen los mismos resultados a partir de la iteración 5.
2. El coeficiente de seguridad F_s^0 es activo, es decir, es más restrictivo que su correspondiente probabilidad de fallo. Nótese que la probabilidad de fallo para este modo es mayor que $\beta_s^0 = 3.719$.
3. Las cotas probabilistas $\beta_b^0, \beta_{stem}^0, \beta_{toe}^0, \beta_{heel}^0, \beta_{sstem}^0, \beta_{stoe}^0, \beta_{sheel}^0$ son activas, es decir, son más restrictivas que las asociadas a los coeficientes de seguridad.
4. El coeficiente de seguridad mínimo F_t^0 y su correspondiente cota probabilista β_t^0 son ambas inactivas. Las restricciones relativas a los demás modos de fallo hacen que se cumplan éstas. Nótese que su coeficiente de seguridad y su índice de fiabilidad son mayores que las cotas inferiores.
5. Los valores óptimos para d y z_t son $d = 0.3$ y $z_t = 0.3$. Esto implica que las restricciones (11.20) y (11.21) son activas. En otras palabras, se puede obtener un menor coste eliminando estas condiciones, entonces la solución óptima sería $d = z_t = 0$.

Table 11.2: Ilustración del proceso iterativo. Los valores finales están en negrita.

		ITERACIONES				
Variable	Unidades	1	2	5	10	11 (fin)
Cost	<i>euro</i>	1105.2	1246.4	1280.5	1282.8	1282.8
h_1	<i>m</i>	5.00	5.00	5.00	5.00	5.00
z_t	<i>m</i>	0.30	0.30	0.30	0.30	0.30
a	<i>m</i>	1.55	1.72	1.80	1.81	1.81
b	<i>m</i>	0.90	1.15	1.23	1.23	1.23
c	<i>m</i>	4.83	4.23	3.98	3.97	3.97
d	<i>m</i>	0.30	0.30	0.30	0.30	0.30
h_2	<i>m</i>	0.50	0.60	0.62	0.62	0.62
h_3	<i>m</i>	1.04	1.38	1.51	1.52	1.52
A_{alzado}	m^2	0.00166	0.00167	0.00159	0.00158	0.00158
$A_{puntera}$	m^2	0.00044	0.00042	0.00037	0.00037	0.00037
A_{heel}	m^2	0.00180	0.00161	0.00143	0.00142	0.00142
F_t	–	6.45	6.05	5.95	5.96	5.96
F_s	–	2.00	2.00	2.00	2.00	2.00
F_b	–	2.00	2.00	2.01	2.02	2.02
F_{stem}	–	1.30	1.79	1.86	1.86	1.86
$F_{puntera}$	–	1.30	1.63	1.49	1.50	1.50
F_{heel}	–	1.30	1.81	1.89	1.89	1.89
F_{sstem}	–	1.40	1.80	1.93	1.94	1.94
F_{stoe}	–	1.40	1.74	1.82	1.82	1.82
F_{sheel}	–	1.40	1.79	1.93	1.94	1.94
β_t	–	–	24.63	24.12	24.13	24.13
β_s	–	–	3.87	3.80	3.79	3.79
β_b	–	–	3.71	3.71	3.72	3.72
β_{stem}	–	–	3.50	3.71	3.72	3.72
β_{toe}	–	–	4.44	3.68	3.72	3.72
β_{heel}	–	–	3.45	3.71	3.72	3.72
β_{sstem}	–	–	3.33	3.70	3.72	3.72
β_{stoe}	–	–	3.48	3.71	3.72	3.72
β_{sheel}	–	–	3.31	3.70	3.72	3.72

Table 11.3: Sensibilidades con respecto a los datos del problema.

c_c	c_t	c_{st}	c_{ex}	r	ϕ_l
11.531	12.815	2726.431	56.756	738.240	3.668
ϕ_t	s_l	s_t	F_t	F_s	F_b
5.245	-110.042	-157.347	0.000	193.519	55.568
F_{alzado}	F_{toe}	F_{heel}	F_{sstem}	F_{stoe}	F_{sheel}
27.290	8.298	24.804	1.432	160.687	129.427
$\bar{\gamma}_c$	$\bar{\gamma}_s$	$\bar{\gamma}_{st}$	\bar{f}_c	\bar{f}_y	$\bar{\sigma}_{soil}$
-5.593	22.063	2.084	-0.024	-0.245	-448.275
$\bar{\tau}_{max}$	$\bar{\mu}_{crit}$	\bar{q}	\bar{k}_a	\bar{k}_p	\bar{h}_1
-1364.251	-774.074	11.679	2495.998	-18.428	490.769

11.1.9 Sensitivity analysis

Las sensibilidades para el ejemplo numérico se dan en las Tablas 11.3, 11.4 and 11.5, respectivamente.

La tabla 11.3 da las sensibilidades de la solución óptima con respecto a los datos, con lo cual se sabe cómo cambia la función objetivo (coste de construcción óptimo) cuando el valor del dato únicamente aumenta en una pequeña cantidad. Esta información es extremadamente útil durante el proceso de construcción para controlar el coste, y para analizar cómo los cambios en los datos afectan al mismo. Entre ellos se tiene la influencia del coste de los materiales, la influencia de los coeficientes de seguridad, etc. Así por ejemplo, un cambio de un euro en el coste por metro cúbico de hormigón (c_c) conlleva un incremento de coste de construcción de 11.531 euro (véase la casilla correspondiente en la tabla 11.3). De forma similar, un aumento en el límite inferior del coeficiente de seguridad F_t^0 no afecta al coste, ya que en el óptimo no es activa la restricción correspondiente (11.10), pero un aumento del límite inferior del coeficiente de seguridad F_s^0 supone un aumento de coste de 193.519 euro por unidad de aumento.

Es importante resaltar que el estudio de sensibilidad tiene carácter local, es decir, que sólo vale en el entorno de la solución, ya que grandes aumentos en alguno de los parámetros, por ejemplo en F_t^0 puede suponer que en el nuevo valor óptimo la restricción (11.10) se active, con lo que el anterior análisis de sensibilidad no valdría.

La tabla 11.4 da las sensibilidades de los índices de fiabilidad (β) con respecto a los datos, y a los parámetros que definen las distribuciones estadísticas de las variables aleatorias (medias y desviaciones típicas). Se puede por tanto analizar cómo influyen todos esos parámetros en la fiabilidad y por tanto en la seguridad de cada uno de los modos de fallo. Con lo cual se sabe que datos o parámetros influyen más en cada uno de los modos de fallo y se podrá actuar en consecuencia. Con esta tabla, el ingeniero puede analizar cómo influye, por ejemplo, la calidad del material (desviación típica en f_c o f_y) o la precisión en la construcción (desviación típica en las dimensiones de diseño) en la seguridad del muro.

Si se prefieren valores no dimensionales de las sensibilidades se puede reemplazar el valor

Table 11.4: Sensibilidades $\frac{\partial \beta_i}{\partial x}$, con respecto a los datos x , en el ejemplo del muro.

Dato x	$\frac{\partial \beta_t}{\partial x}$	$\frac{\partial \beta_s}{\partial x}$	$\frac{\partial \beta_b}{\partial x}$	$\frac{\partial \beta_{stem}}{\partial x}$	$\frac{\partial \beta_{toe}}{\partial x}$	$\frac{\partial \beta_{heel}}{\partial x}$	$\frac{\partial \beta_{sstem}}{\partial x}$	$\frac{\partial \beta_{stoe}}{\partial x}$	$\frac{\partial \beta_{sheel}}{\partial x}$
r	0.00	0.00	0.00	-5.93	-12.34	-7.10	-4.59	-7.13	-5.67
ϕ_l	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ϕ_t	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
s_l	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
s_t	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
μ_{h_1}	-6.13	-0.95	-1.55	-3.26	-3.27	-3.12	-1.56	-1.95	-1.83
μ_a	4.65	-0.06	1.89	0.00	-7.99	2.17	0.00	0.04	0.75
μ_b	5.33	0.38	1.36	6.98	2.33	2.39	4.59	1.34	1.10
μ_c	6.49	1.07	0.68	0.00	-0.54	-1.21	0.00	0.35	0.12
μ_d	0.79	0.49	-0.52	0.34	-2.02	0.37	0.00	-0.72	0.46
μ_{h_2}	-6.49	-0.91	-1.27	0.00	18.44	2.38	0.00	4.48	2.38
μ_{h_3}	2.75	1.34	-0.75	0.00	4.17	2.95	0.00	1.15	1.75
μ_{z_t}	-6.09	0.17	-1.84	-3.24	-1.61	-3.04	-1.25	-1.42	-1.57
$\mu_{A_{stem}}$	0.00	0.00	0.00	4261.54	0.00	0.00	0.00	0.00	0.00
$\mu_{A_{toe}}$	0.00	0.00	0.00	0.00	25649.34	0.00	0.00	0.00	0.00
$\mu_{A_{heel}}$	0.00	0.00	0.00	0.00	0.00	4671.60	0.00	0.00	0.00
μ_{γ_c}	0.25	0.09	-0.05	0.03	-0.14	0.03	0.00	-0.04	0.03
μ_{γ_s}	0.19	-0.03	-0.07	-0.14	-0.11	-0.13	-0.13	-0.10	-0.14
$\mu_{\gamma_{st}}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
μ_{f_c}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
μ_{f_y}	0.00	0.00	0.00	0.02	0.02	0.02	0.00	0.00	0.00
$\mu_{\sigma_{soil}}$	0.00	0.00	29.03	0.00	0.00	0.00	0.00	0.00	0.00
$\mu_{\tau_{max}}$	0.00	0.00	0.00	0.00	0.00	0.00	18.53	20.16	18.70
$\mu_{\mu_{crit}}$	0.00	15.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00
μ_q	-0.05	-0.03	-0.04	-0.08	-0.07	-0.08	-0.05	-0.05	-0.06
μ_{k_a}	-14.44	-11.84	-7.02	-12.24	-4.75	-12.71	-9.24	-7.61	-7.76
μ_{k_p}	0.00	0.13	0.03	0.00	0.12	0.03	0.02	0.03	0.03
σ_{h_1}	-22.70	-0.08	-0.22	-0.99	-0.99	-0.91	-0.23	-0.35	-0.31
σ_{z_t}	-22.36	0.00	-0.32	-0.97	-0.24	-0.86	-0.14	-0.19	-0.23
σ_a	-13.05	0.00	-0.33	0.00	-5.94	-0.44	0.00	0.00	-0.05
σ_b	-17.16	-0.01	-0.17	-4.53	-0.51	-0.53	-1.96	-0.17	-0.11
σ_c	-25.40	-0.11	-0.04	0.00	-0.03	-0.14	0.00	-0.01	0.00
σ_d	-0.37	-0.02	-0.03	-0.01	-0.38	-0.01	0.00	-0.05	-0.02
σ_{h_2}	-25.43	-0.08	-0.15	0.00	-31.62	-0.53	0.00	-1.86	-0.53
σ_{h_3}	-4.58	-0.17	-0.05	0.00	-1.62	-0.81	0.00	-0.12	-0.29
$\sigma_{A_{stem}}$	0.00	0.00	0.00	-675.39	0.00	0.00	0.00	0.00	0.00
$\sigma_{A_{toe}}$	0.00	0.00	0.00	0.00	-24465.36	0.00	0.00	0.00	0.00
$\sigma_{A_{heel}}$	0.00	0.00	0.00	0.00	0.00	-811.61	0.00	0.00	0.00
σ_{γ_c}	-1.46	-0.03	-0.01	0.00	-0.07	0.00	0.00	-0.01	0.00
σ_{γ_s}	-1.68	-0.01	-0.03	-0.14	-0.09	-0.13	-0.12	-0.07	-0.16
$\sigma_{\gamma_{st}}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
σ_{f_c}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
σ_{f_y}	0.00	0.00	0.00	-0.02	-0.04	-0.02	0.00	0.00	0.00
σ_{σ_s}	0.00	0.00	-94.03	0.00	0.00	0.00	0.00	0.00	0.00
σ_{τ_m}	0.00	0.00	0.00	0.00	0.00	0.00	-51.05	-60.43	-52.02
$\sigma_{\mu_{cr}}$	0.00	-43.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00
σ_q	-0.43	-0.03	-0.04	-0.18	-0.13	-0.17	-0.07	-0.07	-0.09
σ_{k_a}	-251.42	-26.61	-9.16	-27.85	-4.20	-30.02	-15.89	-10.77	-11.21
σ_{k_p}	0.00	-0.01	0.00	0.00	-0.01	0.00	0.00	0.00	0.00

Table 11.6: Ilustración de la simulación de Monte Carlo.

<i>Modo</i>	$P_f(\times 10^{-4})$	β_{FOSM}	β_{MC}
β_t	0.0000	24.127	$-\infty$
β_s	0.6893	3.794	-3.812
β_b	0.9934	3.719	-3.721
β_{stem}	0.9428	3.719	3.734
β_{toe}	0.8399	3.719	3.763
β_{heel}	0.9397	3.719	3.735
β_{sstem}	1.0117	3.719	3.716
β_{stoe}	1.0135	3.719	3.716
β_{sheel}	1.0269	3.719	3.712
β_{tot}	4.9208	-	3.295

$\frac{\partial \beta_t}{\partial x}$ por $\frac{\bar{x} \partial \beta_t}{\partial x}$ en la tabla 11.4 y obtener la tabla 11.5.

11.1.10 Monte Carlo simulation

Para determinar la probabilidad de fallo de cada uno de los modos de rotura, así como la probabilidad de fallo global, se ha realizado una simulación de Monte Carlo con 10^8 réplicas. Habida cuenta del gran número de modos de fallo, y pese a ser un sistema en serie, no se han calculado las probabilidades de las intersecciones ya que ascenderían a $2^8 - 1 = 511$.

Se puede concluir que la probabilidad de fallo global del muro es de 0.00049208, mientras que las probabilidades de fallo de cada uno de los modos de rotura, junto con los índices de fiabilidad aproximados mediante FOSM (β_{FOSM}) y los equivalentes de la simulación de Monte Carlo ($\beta_{MC} = \Phi^{-1}(1 - P_f)$) se encuentran en la tabla 11.6.

11.2 Compound girder

En esta sección se describe con detalle el cálculo de una viga mixta perteneciente al tablero de un puente biapoyado basado en técnicas probabilistas.

11.2.1 The compound girder problem

Para ilustrar los conceptos que se presentarán en las secciones siguientes, se usará el ejemplo de una viga mixta simplemente apoyada (véanse las Figuras 11.9 y 11.10), con una geometría dada (nótese que está paramétricamente definida). No obstante, se clarifica al lector que el objetivo del presente ejemplo es ilustrar conceptos y no realizar un análisis exhaustivo de todos los posibles modos de fallo de la viga (que los árboles no escondan al bosque).

Las vigas mixtas tienen un uso muy extendido en la construcción de puentes, plantas comerciales, etc. Con su utilización, se pretende reemplazar en las zonas de compresión el acero, que es un material caro, por hormigón armado, que es más barato. Para ello la losa de hormigón se une al ala superior de la viga de acero mediante conectores.

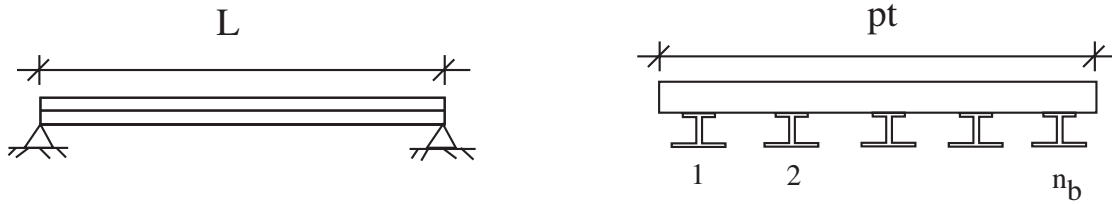


Figure 11.9: Puente mixto biapoyado utilizado en el Ejemplo.

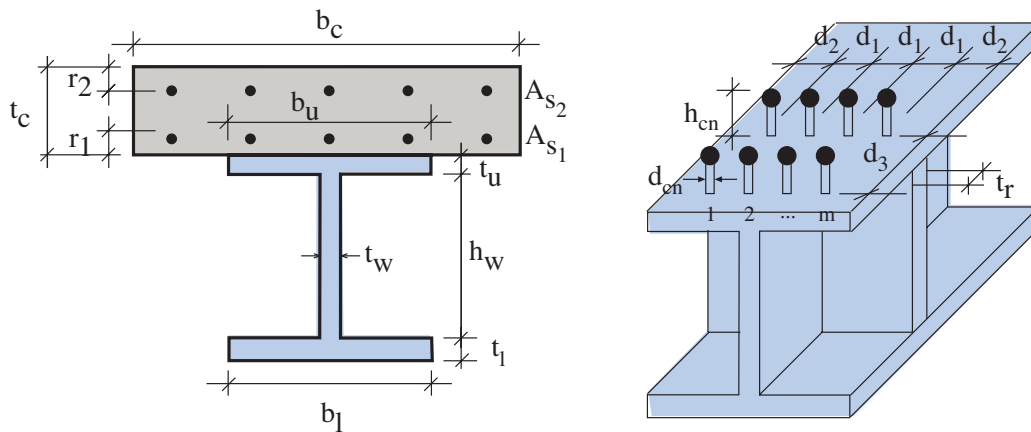


Figure 11.10: Definición geométrica de la viga mixta.

El diseño de una viga mixta requiere seleccionar los siguientes parámetros: espesores de chapa de la viga mixta, espaciamiento entre vigas, espesor y posición de los rigidizadores, número y dimensiones de los conectores, etc.

Para facilitar la descripción del problema, se considera el siguiente conjunto de índices:

e : Localizaciones: $\{c = \text{losa de hormigón}, u = \text{ala superior}, l = \text{ala inferior}\}$

t : Periodos de tiempo considerados: $\{0, \infty\}$.

p : Posiciones longitudinales en el puente: $\{c = \text{centro de vano}, s = \text{apoyo}\}$

En este ejemplo, no se discute el método usado para la selección de los índices de fiabilidad (β), existen en la literatura amplias discusiones acerca de este tema (Royal Society Study Group [152], Blockley [23], Rowe [148] y Stewart y Melchers [162]).

Independientemente del método elegido para diseñar, el primer paso es seleccionar los modos de fallo o estados límites. En este ejemplo se considerarán los siguientes:

1. Estado límite de servicio.

- Flecha máxima en centro luz.

2. Estado límite de daño.

- Deformaciones plásticas locales en el hormigón a tiempo 0 y ∞ .
- Deformaciones plásticas locales en las alas superiores e inferiores de la viga de acero a tiempo 0 y ∞ .
- Deformaciones plásticas locales en el alma de acero.

3. Estado límite último.

- Momento último resistido por la sección.
- Esfuerzo cortante último resistido por la sección.

Las ecuaciones de estado límite se obtendrán en la subsección 11.2.2.

Al igual que en la sección anterior (11.1), y dado que se va a emplear el método PSFM, se va a clasificar las variables que intervienen. El conjunto de variables de diseño (X_1, \dots, X_n) se dividirá de forma más detallada que en el ejemplo anterior en cinco subconjuntos (véase la subsección 11.2.9 para la notación):

d: Variables de diseño optimizadas.

$$\mathbf{d} = (b_e, t_e, h_w, t_w, d_1, d_2, d_3, A_{s1}, A_{s2}, h_{st}, t_r)$$

η : Conjunto de parámetros.

$$\boldsymbol{\eta} = (L, pl, d_{st}, r_1, r_2, n_b, m, E_t, E_s, \phi, \nu, e_p, e_s, c_c, c_s, c_y, c_{st})$$

ϕ : Variables aleatorias básicas usadas en los modelos probabilistas.

$$\boldsymbol{\phi} = (f_c, f_s, f_y, \gamma_c, \gamma_s, w_d, w_p, p_s)$$

κ : Parámetros estadísticos.

$$\boldsymbol{\kappa} = (\mu_{w_d}, \mu_{w_p}, \mu_{p_s}, \mu_{f_c}, \mu_{f_s}, \mu_{f_y}, \mu_{\gamma_c}, \mu_{\gamma_s}, v_{w_d}, v_{w_p}, v_{p_s}, v_{f_c}, v_{f_s}, v_{f_y}, v_{\gamma_c}, v_{\gamma_s})$$

ψ : Variables auxiliares o no básicas.

$$\boldsymbol{\psi} = (V_p, \sigma_{e,t}, \tau_p, M_p, M_{un}, V_{max}, w_1, l_d, C_p, s_u, V_t, A_{e,t}, A_w, I_t, I_s, I_r, Y_e, Y_w, EI_t, \delta_t, \delta, \delta_{max}, A_{min}, e_{max}, C_{un}, T_{un}, X_{un}, \tau_{cr}, \lambda_w, m_t, \tau_{max}, P_{stc}, P_{st}, N_{cs})$$

donde

$$e \in \{c, u, l\}; \quad t \in \{0, \infty\}; \quad p \in \{c, s\}.$$

11.2.2 Constraints

Para el diseño de la viga mixta se han de considerar, entre otras, las siguientes restricciones:

1. Restricciones geométricas e impuestas por el código.
2. Restricciones asociadas a la probabilidad de fallo (una por cada modo de fallo).
3. Restricciones asociadas a los coeficientes de seguridad. Están asociadas a cada uno de los modos de fallo que se consideran (estados límites de servicio, daño y último).

- (a) **Estado límite de daño del hormigón a tiempo 0 y ∞ .** El análisis de este modo de fallo se puede hacer considerando el ratio resistencia, esfuerzos actuantes

$$\frac{f_c}{\sigma_{c,t}} = \frac{f_c}{\frac{M_c V_t E_t}{EI_t}} \geq F_{ct}; t \in \{0, \infty\} \quad (11.23)$$

donde F_{c0} y $F_{c\infty}$ son los coeficientes de seguridad asociados a cada una de las condiciones.

- (b) **Estado límite de daño de las alas superior e inferior de la viga de acero a tiempo 0 y ∞ .** Se considera el ratio resistencia, esfuerzos actuantes

$$\frac{f_y}{\sqrt{\sigma_{u,t}^2 + 3\tau_c^2}} \geq F_{ut}; t \in \{0, \infty\} \quad (11.24)$$

y

$$\frac{f_y}{\sqrt{\sigma_{l,t}^2 + 3\tau_c^2}} \geq F_{lt}; t \in \{0, \infty\} \quad (11.25)$$

donde F_{u0} , $F_{u\infty}$, F_{l0} y $F_{l\infty}$ son los coeficientes de seguridad asociados con estas condiciones.

- (c) **Estado límite de daño del alma de acero.** En este caso se considera el ratio resistencia cortante de acero, esfuerzo cortante actuante

$$\frac{f_y}{\sqrt{3\tau_s^2}} = \frac{f_y}{\frac{V_s \sqrt{3}}{h_w t_w}} \geq F_w \quad (11.26)$$

donde F_w es el coeficiente de seguridad correspondiente.

- (d) **Estado límite último asociado al máximo momento flector resistido por la sección.** El análisis de este modo de fallo se puede hacer considerando el ratio máximo momento que resiste la sección, momento actuante

$$\frac{M_{un}}{M_c} \geq F_m \quad (11.27)$$

donde F_m es el coeficiente de seguridad asociado a este modo de fallo.

- (e) **Estado límite último asociado al máximo esfuerzo cortante resistido por la sección.** El análisis de este modo de fallo se puede hacer considerando el ratio máximo esfuerzo cortante que resiste la sección, esfuerzo cortante actuante

$$\frac{V_{max}}{V_s} \geq F_v \quad (11.28)$$

donde F_v es el coeficiente de seguridad asociado a este modo de fallo.

- (f) **Estado límite de servicio de la máxima flecha en centro vano.** El análisis se puede hacer considerando el ratio máxima flecha permitida, flecha existente

$$\frac{\delta_{max}}{\delta} \geq F_\delta \quad (11.29)$$

donde F_δ es el coeficiente de seguridad asociado a esta condición.

Dado que se tienen 10 modos de fallo diferentes, se define el conjunto I_f de modos de fallo como

$$I_f = \{\delta, c_0, c_\infty, u_0, u_\infty, \ell_0, \ell_\infty, w, m, v\}$$

La viga se considerará válida si los coeficientes de seguridad $F_\delta, F_{c_0}, F_{c_\infty}, F_{u_0}, F_{u_\infty}, F_{\ell_0}, F_{\ell_\infty}, F_w, F_m$ y F_v son mayores que los valores mínimos requeridos.

Todos estos modos de fallo están correlacionados porque tienen agentes comunes. Estas correlaciones se observarán en la simulación de Monte Carlo en la subsección 11.2.7.

Obsérvese que sólo se han considerado los modos de fallo más relevantes, pero hay otros posibles; sin embargo, han sido ignorados por motivos de claridad.

11.2.3 Objective function

En esta sección se calcula el coste total de construcción de la viga. Dado que se trata de un ejemplo ilustrativo y con propósito de no complicar el problema en exceso, no se considera el coste a largo plazo, ni mantenimiento, ni reparaciones.

Por tanto, se necesita calcular el peso total de hormigón y acero, así como el número de conectores por unidad de longitud, y los rigidizadores del alma.

$$\begin{aligned} \text{Coste} = & 1000\gamma_s \left[b_u t_u \left(c_y + \frac{2.404 \times 10^{-9}}{(0.04 - t_u)^4} \right) + b_\ell t_\ell \left(c_y + \frac{2.404 \times 10^{-9}}{(0.04 - t_\ell)^4} \right) \right] \\ & + 1000\gamma_s \left[h_w t_w \left(c_y + \frac{2.404 \times 10^{-9}}{(0.04 - t_w)^4} \right) + t_r h_w (b_\ell - t_w) \left(c_y + \frac{2.404 \times 10^{-9}}{(0.04 - t_r)^4} \right) \right] \\ & + 1000c_s \gamma_s (A_{s_1} + A_{s_2}) + c_{st} \frac{m}{d_3} + c_c b_c t_c (b_c - b_\ell)^2 \end{aligned} \quad (11.30)$$

Como puede observarse en la ecuación (11.30) el precio del acero tiene una penalización en función del espesor, de tal forma que a mayores espesores, mayor coste. De esta manera se tiene en cuenta el aumento de coste en manipular y cortar las chapas.

11.2.4 Distributional assumptions of the model

Diseñar con métodos de nivel II y III implica la definición de las propiedades estadísticas de las variables aleatorias que intervienen.

Todas las variables estocásticas son independientes, así la transformación de Rosenblatt es muy simple. La independencia entre los parámetros geométricos, las propiedades mecánicas y las físicas es natural porque proceden de distintas fuentes. La hipótesis de independencia entre las variables relativas a diferentes materiales parámetros geométricos es también bastante realista, aunque no supondría ningún problema para el modelo considerar la dependencia entre determinadas variables.

Es importante recalcar que la correlación entre los diferentes modos de fallo proviene del hecho de que tienen variables comunes. Así, aunque las variables se consideren independientes, los modos de fallo podrán estar correlacionados. En otras palabras, la principal fuente de correlación entre modos de fallo procede de la dependencia en variables comunes y no de la dependencia de las variables en sí mismas.

Las hipótesis de las distribuciones son:

1. **Parámetros geométricos.** Las variables de diseño $b_u, b_\ell, t_u, t_\ell, h_w, t_w, d_1, d_2, d_3, A_{s1}, A_{s2}, h_{st}$ y t_r) se consideran deterministas e iguales a los valores de diseño clásicos (con probabilidad uno). Esto se debe al alto control al que se someten estas variables durante la construcción.

Las variables de diseño b_c y t_c se consideran normalmente distribuidas con media igual al valor clásico y desviaciones típicas de 0.01 m y 0.005 m, respectivamente. Se asume que el control sobre los elementos de hormigón es menor que el de las piezas de acero.

2. **Propiedades mecánicas de los materiales.** La resistencia del hormigón, f_c , y las resistencias del acero, f_s y f_y , se toman como normalmente distribuidas. Sus valores medios se calculan a partir de los valores característicos f_{ck}, f_{sk} y f_{yk} , y de sus coeficientes de variación v_{f_c}, v_{f_s} y v_{f_y} (hay una probabilidad 0.95 de que el valor real sea mayor que el característico), respectivamente:

$$\begin{aligned}\mu_{f_c} &= \frac{f_{ck}}{1 - 1.64v_{f_c}} = f_{ck} + 8, \\ \mu_{f_s} &= \frac{f_{sk}}{1 - 1.64v_{f_s}}, \\ \mu_{f_y} &= \frac{f_{yk}}{1 - 1.64v_{f_y}}.\end{aligned}$$

Las variables w_d, w_p y p_s se suponen normales independientes con valores medios dados μ_{w_d}, μ_{w_p} y μ_{p_s} , y coeficientes de variación v_{w_d}, v_{w_p} y v_{p_s} , respectivamente. Todas estas variables toman sus valores característicos para el diseño clásico.

3. **Propiedades físicas del material.** Los pesos específicos del hormigón y del acero, γ_c y γ_s se consideran variables normales también. Sus medias y coeficientes de variación son $\mu_{\gamma_c}, \mu_{\gamma_s}, v_{\gamma_c}$ y v_{γ_s} , respectivamente. Estas variables toman sus valores medios para el clásico.

11.2.5 Numerical example

El método propuesto en la sección 10.4 ha sido implementado en GAMS (General Algebraic Modeling System) (véase Castillo, Conejo, Pedregal, García y Alguacil [30]) para un ejemplo concreto. Considérese la siguiente viga mixta:

$$\begin{array}{llll}
 L = 30 \text{ m} & pl = 10 \text{ m} & e_p = 2 & e_s = 1.7 \\
 E_s = 210000 \text{ MPa} & \phi = 2 & r_1 = 0.05 & r_2 = 0.05 \text{ m} \\
 m = 4 & d_{st} = 0.02 \text{ m} & n_b = 4 & p_{sk} = 4 \text{ KN/m}^2 \\
 w_{dk} = 40 \text{ KN/m} & w_{pk} = 600 \text{ KN} & \mu_{\gamma_c} = 25 \text{ KN/m}^3; & \mu_{\gamma_s} = 78.5 \text{ KN/m}^3; \\
 \mu_{f_c} = 38 \text{ MPa} & f_{ck} = 30 \text{ MPa} & f_{sk} = 400 \text{ MPa} & f_{yk} = 235 \text{ MPa} \\
 v_{ps} = 0.3 & v_{wd} = 0.2 & v_{wp} = 0.3 & v_{\gamma_c} = 0.1 \\
 v_{\gamma_s} = 0.01 & v_{f_y} = 0.03 & v_{f_s} = 0.04 & c_y = 0.24 \text{ euros/N} \\
 c_s = 0.063 \text{ euros/N} & c_{st} = 1 \text{ euros} & c_c = 60.1 \text{ euros/m}^3 &
 \end{array}$$

donde todos estos valores son seleccionados por el diseñador.

Se han utilizado los siguientes límites inferiores para los coeficientes de seguridad y para los valores del índice de fiabilidad

$$F_\delta = 1.0; F_{c0} = F_{c\infty} = 1.15; F_{u0} = F_{u\infty} = 1.15; F_{\ell0} = F_{\ell\infty} = 1.15; F_w = 1.15; F_m = F_v = 1.3 \quad (11.31)$$

y

$$\beta_\delta = 2.58; \beta_{c0} = \beta_{c\infty} = \beta_{u0} = \beta_{u\infty} = \beta_{\ell0} = \beta_{\ell\infty} = \beta_w = 3.29; \beta_m = \beta_v = 3.72 \quad (11.32)$$

correspondientes a probabilidades de fallo de 0.5×10^{-2} , 0.5×10^{-4} y 10^{-4} , respectivamente. Obsérvese que la ‘violación’ de los estados límites con mayores consecuencias negativas en caso de fallo están asociados a menores probabilidades de fallo.

La tabla 11.7 muestra la convergencia del método, que con la tolerancia exigida se alcanza tras 8 iteraciones. La columna de la primera iteración muestra los valores iniciales de las variables de diseño, los valores actuales de los coeficientes de seguridad y los valores de los índices de fiabilidad (β) asociados con el diseño óptimo para los coeficientes de seguridad en (11.31). Obsérvese que sólo las restricciones $F_\delta = 1.0$ y $F_{\ell\infty} = 1.15$ (en negrita en la tabla 11.7) son activas. Y que la restricción asociada a β_δ (en negrita en la tabla 11.7) no se satisface. Por tanto los límites inferiores de los coeficientes de seguridad son actualizados para la segunda iteración. Se pueden hacer comentarios similares para la segunda columna. En cambio, en la iteración 3 la restricción que no se satisface es la β_δ . Por lo tanto, se requieren más iteraciones hasta que la solución converge de forma que todas las restricciones se cumplen.

La última columna de la tabla muestra los valores óptimos de las variables

$$b_c, b_u, b_\ell, t_c, t_u, t_\ell, t_r, h_w, t_w, h_{st}, A_{s1}, A_{s2}, d_1, d_2, d_3,$$

junto con los coeficientes de seguridad y valores del índice de fiabilidad (β) asociados.

Los valores activos en la solución final aparecen subrayados en la tabla 11.7, de la cual se pueden extraer las siguientes conclusiones:

Table 11.7: Ilustración del proceso iterativo. Los valores de diseño finales aparecen en negrita.

		ITERACIONES			
Variable	Unidades	1	2	3	8 (fin)
Coste	<i>euros</i>	1212.4	1218.2	1220.1	1221.1
b_c	<i>m</i>	2.500	2.500	2.500	2.500
b_u	<i>m</i>	0.441	0.442	0.442	0.442
b_l	<i>m</i>	0.441	0.442	0.442	0.442
t_c	<i>cm</i>	27.95	28.57	28.78	28.88
t_u	<i>cm</i>	0.88	0.88	0.88	0.88
t_l	<i>cm</i>	2.11	2.10	2.10	2.10
h_w	<i>m</i>	1.972	1.974	1.974	1.975
t_w	<i>cm</i>	1.60	1.61	1.61	1.61
h_{st}	<i>cm</i>	8.00	8.00	8.00	8.00
t_r	<i>cm</i>	0.80	0.80	0.80	0.80
As_1	<i>cm²</i>	6.29	6.43	6.47	6.50
As_2	<i>cm²</i>	6.29	6.43	6.47	6.50
d_1	<i>cm</i>	11.37	11.40	11.41	11.41
d_2	<i>cm</i>	5.00	5.00	5.00	5.00
d_3	<i>cm</i>	18.31	17.92	17.79	17.72
F_δ	–	1.00	1.01	1.01	1.02
F_{c0}	–	3.60	3.62	3.63	3.63
$F_{c\infty}$	–	4.86	4.90	4.91	4.92
F_{u0}	–	9.66	9.99	10.11	10.17
$F_{u\infty}$	–	3.05	3.12	3.14	3.15
F_{l0}	–	1.26	1.26	1.26	1.26
$F_{l\infty}$	–	1.15	1.15	1.15	1.15
F_w	–	4.18	4.16	4.16	4.15
F_m	–	1.80	1.80	1.80	1.80
F_v	–	2.63	2.62	2.62	2.62
β_δ	–	2.44	2.53	2.56	2.58
β_{c0}	–	6.40	6.41	6.41	6.41
$\beta_{c\infty}$	–	6.78	6.78	6.79	6.79
β_{u0}	–	29.18	29.36	29.42	29.44
$\beta_{u\infty}$	–	18.37	18.68	18.79	18.84
β_{l0}	–	5.82	5.82	5.82	5.82
$\beta_{l\infty}$	–	4.63	4.63	4.64	4.64
β_w	–	23.07	23.04	23.03	23.02
β_m	–	7.12	7.12	7.12	7.12
β_v	–	19.32	19.28	19.26	19.26

Table 11.8: Resultados de la simulación de Monte Carlo para 10^7 realizaciones.

Tipo de fallo	Frecuencia	Prob. of failure ($\times 10^{-4}$)	
		Montecarlo	FORM
δ	50564	0.50564×10^{-3}	0.50011×10^{-3}
c_0	0	0	7.152×10^{-11}
c_∞	0	0	5.66658×10^{-12}
u_0	0	0	0
u_∞	0	0	0
ℓ_0	0	0	3.0018×10^{-9}
ℓ_∞	17	1.7×10^{-6}	1.78201×10^{-6}
w	0	0	0
m	0	0	5.5117×10^{-13}
v	0	0	0

1. El proceso converge tras 8 iteraciones.
2. Se obtiene un listado con los coeficientes de seguridad reales y los índices de fiabilidad asociados a cada modo de fallo.
3. Únicamente el coeficiente de seguridad F_{ℓ_∞} , es activo.
4. Debido a la restricción impuesta por el estado límite de servicio, el único límite probabilista activo es β_δ .
5. El diseño final (iteración 8) es más caro que el inicial, ya que el primero no satisfacía la restricción asociada al β_δ .

En la subsección 11.2.7 se puede obtener más información, ya que las restricciones activas están marcadas con el signo “•”.

11.2.6 Monte Carlo simulation.

Para estimar la probabilidad de fallo global de la estructura, comprender mejor la interacción y correlación entre los distintos modos de fallo, y estimar las probabilidades de fallo de cada combinación de modos de fallo, se ha realizado una simulación de Monte Carlo con 10^7 réplicas, determinándose las probabilidades de fallo de cada uno de los modos de fallo. Los resultados se presentan en la tabla 11.8.

La probabilidad de exceder la máxima flecha permitida en centro luz es 0.50564×10^{-3} , muy similar a la obtenida mediante la aproximación FORM. Las probabilidades asociadas a estado límite de daño y último son 1.7×10^{-6} y 0, respectivamente. Cabe señalar que las situaciones en las que se producen daños, se excede siempre la máxima flecha permitida.

Table 11.9: Sensibilidades al coste con respecto a los datos en el ejemplo ilustrativo del muro.

d_{st}	L	r_1	r_2	pl	n_b	e_p	e_u	ϕ
-2257.21	25.73	0.00	0.00	79.28	-276.31	128.38	154.81	98.21
E_s	m	f_{ck}	f_{sk}	f_{yk}	c_y	c_s	$c_s t$	c_c
98.21	0.00	0.36	0.00	0.45	4054.13	102.03	22.57	3.06
μ_{w_d}	μ_{p_s}	μ_{w_p}	μ_{f_c}	μ_{f_s}	μ_{f_y}	μ_{γ_c}	μ_{γ_s}	F_δ
1.85	98.26	0.64	-2.12	0.00	0.00	4.02	13.21	537.34
F_{c_0}	F_{c_∞}	F_{u_0}	F_{u_∞}	F_{l_0}	F_{l_∞}	F_w	F_m	F_v
0.00	0.00	0.00	-0.09	0.00	132.79	0.00	0.00	0.00

11.2.7 Sensitivity analysis

Las sensibilidades para el ejemplo numérico de la viga se dan en las tablas 11.9 y 11.10. La tabla 11.9 muestra las sensibilidades al coste asociadas al diseño clásico. Y nos permite conocer cómo influyen los cambios en los parámetros en el coste total óptimo de la viga mixta. Esta información es extremadamente útil durante el proceso de construcción para controlar el coste, y cómo influyen en el coste los coeficientes de seguridad globales. Por ejemplo, un cambio de un euro en el coste unitario c_s del acero, produce un incremento de coste de 102.03 euros por metro lineal de viga (véase la entrada correspondiente en la tabla 11.9). De forma similar, un incremento en el límite inferior del coeficiente de seguridad F_{c_0} no cambia el coste, lo cual es lógico ya que F_{c_0} no está activo en la solución final, y un incremento de una unidad en la longitud del puente supone un aumento de coste de 25.73 euros por metro lineal.

La tabla 11.10 muestra la sensibilidades asociadas a los valores de β . Es muy útil conocer la influencia de los parámetros en la seguridad asociada a cada uno de los modos de fallo, y saber cómo varía el índice de fiabilidad cuando alguno de los parámetros cambia en una unidad. Con esta tabla, el diseñador puede analizar la influencia de la calidad de los materiales (reducidas desviaciones típicas en f_c o f_y) o la precisión en la determinación de las acciones de proyecto (reducidas desviaciones típicas en v_{p_s} o v_{w_p}) en la seguridad de la viga. Observese que un aumento en la dispersión de las variables (desviaciones típicas o coeficientes de variación) supone una disminución de los índices de fiabilidad β y por consiguiente un aumento de la probabilidad de fallo.

11.2.8 Compound girder constraints

En este apartado se da una descripción detallada de cómo se han obtenido las restricciones asociadas a la viga. Las restricciones de desigualdad activas para el ejemplo numérico de la subsección 11.2.6 están señaladas con el signo “●”; si no, son inactivas. La lista de variables auxiliares utilizadas se da en la subsección 11.2.9.

Table 11.10: Sensibilidades $\frac{\partial \beta_i}{\partial x}$, con respecto a los datos x , en el ejemplo ilustrativo de la viga.

x	$\frac{\partial \beta_6}{\partial x}$	$\frac{\partial \beta_{c0}}{\partial x}$	$\frac{\partial \beta_{c\infty}}{\partial x}$	$\frac{\partial \beta_{u0}}{\partial x}$	$\frac{\partial \beta_{u\infty}}{\partial x}$	$\frac{\partial \beta_{t0}}{\partial x}$	$\frac{\partial \beta_{t\infty}}{\partial x}$	$\frac{\partial \beta_w}{\partial x}$	$\frac{\partial \beta_m}{\partial x}$	$\frac{\partial \beta_v}{\partial x}$
l	-0.29	-0.08	-0.06	-0.23	-0.74	-0.73	-0.70	-0.24	-0.03	-0.39
r_1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.02	0.00
r_2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
pl	0.39	0.06	0.04	0.32	0.72	0.59	0.55	0.56	0.02	0.93
n_b	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
e_p	-1.72	-0.19	-0.14	-1.39	-2.89	-2.21	-2.02	-2.23	-0.06	-3.76
e_u	-2.20	-0.19	-0.14	-0.62	-2.38	-2.01	-1.86	-1.99	-0.06	-3.33
ϕ	-1.46	0.00	0.10	0.00	-4.04	0.00	-0.45	0.00	0.00	0.00
μ_{w_d}	-0.01	-0.01	0.00	-0.01	-0.05	-0.05	-0.05	-0.04	0.00	-0.06
μ_{p_s}	-1.40	-0.12	-0.09	-0.39	-1.51	-1.28	-1.18	-1.26	-0.04	-2.11
μ_{w_p}	-0.01	0.00	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	-0.02
μ_{f_c}	0.03	0.03	0.03	0.05	0.11	0.01	0.01	0.00	0.02	0.00
μ_{f_s}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
μ_{f_y}	0.00	0.00	0.00	0.02	0.06	0.05	0.05	0.05	0.00	0.05
μ_{γ_c}	-0.04	-0.02	-0.01	-0.03	-0.10	-0.13	-0.13	-0.10	-0.01	-0.15
μ_{γ_s}	0.00	0.00	0.00	0.00	-0.01	-0.01	-0.01	0.00	0.00	-0.01
μ_{bc}	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.10	0.00
μ_{tc}	12.24	1.05	1.14	27.04	48.14	-3.02	-2.22	-8.10	0.32	-12.54
v_{w_d}	-0.08	-0.04	-0.03	-0.57	-2.69	-1.60	-1.32	-2.25	-0.01	-4.00
v_{p_s}	-4.46	-0.15	-0.09	-1.95	-9.23	-5.49	-4.52	-7.74	-0.02	-13.74
v_{w_p}	-3.95	-0.20	-0.12	-6.44	-14.02	-7.71	-6.34	-10.71	-0.03	-19.02
v_{f_c}	0.00	-49.11	-52.46	0.00	0.00	0.00	0.00	0.00	-55.47	0.00
v_{f_s}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
v_{f_y}	0.00	0.00	0.00	-871.1	-319.3	-35.75	-24.48	-544	-0.01	-245.1
v_{γ_s}	0.00	0.00	0.00	-0.01	-0.04	-0.02	-0.02	-0.03	0.00	-0.06
v_{γ_c}	-0.24	-0.12	-0.07	-1.42	-6.57	-4.62	-3.80	-6.70	-0.02	-12.00
σ_{bc}	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00	0.00
σ_{tc}	-1.93	-0.04	-0.04	-107.6	-218.3	-0.27	-0.11	-7.55	0.00	-15.14

Propiedades mecánicas y geométricas de los elementos

En esta apartado se dan los detalles relativos a características mecánicas y geométricas.

Módulo de elasticidad del hormigón a tiempos 0 y ∞ :

$$E_0 = 10000\sqrt[3]{\mu f_c} \quad (11.33)$$

$$E_\infty = \frac{E_0}{1 + \phi} \quad (11.34)$$

Secciones transversales (medidas como área de acero equivalente):

$$A_{c,t} = \frac{b_c t_c E_t}{E_s} \quad (11.35)$$

$$A_{u,t} = b_u t_u \quad (11.36)$$

$$A_{\ell,t} = b_\ell t_\ell \quad (11.37)$$

$$A_w = h_w t_w \quad (11.38)$$

Coordenadas de los centros de gravedad de cada uno de los elementos que conforman la viga:

$$Y_c = t_c/2 \quad (11.39)$$

$$Y_u = t_c + t_u/2 \quad (11.40)$$

$$Y_\ell = t_c + t_u + h_w + t_\ell/2 \quad (11.41)$$

$$Y_w = t_c + t_u + h_w/2 \quad (11.42)$$

Momentos de inercia (medidos como área de acero equivalente):

$$I_t = \frac{b_c t_c^3 E_t}{12 E_s} \quad (11.43)$$

$$I_s = \frac{b_u t_u^3}{12} + \frac{b_\ell t_\ell^3}{12} + \frac{t_w h_w^3}{12} \quad (11.44)$$

Centro de gravedad de la sección transversal incluyendo todos los materiales:

$$V_t = \frac{\sum_e A_{e,t} Y_e + A_w Y_w}{\sum_e A_{e,t} + A_w} \quad (11.45)$$

Rigideces:

$$EI_t = E_s \left(I_t + I_s + \sum_e A_{e,t} (Y_e - V_t)^2 + A_w (Y_w - V_t)^2 \right) \quad (11.46)$$

Flechas en centro luz (asumiendo la presencia de apeos durante la construcción):

$$\delta = \delta_\infty - \delta_0 \quad (11.47)$$

éstas se calculan mediante la fórmula:

$$\delta_\infty = 2 \int_0^{L/2} \frac{(w_1 + l_d + s_u)(L - x)x^2 + C_p x^2}{4000 EI_\infty} dx \quad (11.48)$$

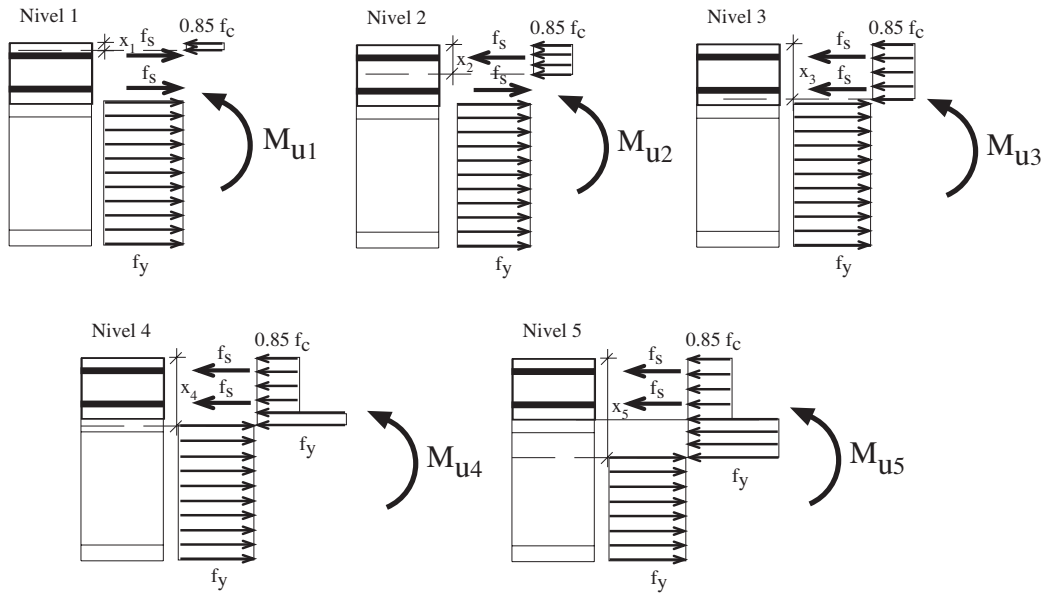


Figure 11.11: Posiciones posibles del eje neutro de plastificación de la sección.

y

$$\delta_0 = 2 \int_0^{L/2} \frac{(w_1 + l_d)(L - x)x^2}{4000EI_0} dx \quad (11.49)$$

La figura 11.11 muestra los momentos últimos resistidos por la sección (M_{un}) para diferentes posiciones de la fibra neutra. Este momento se calcula usando las ecuaciones de equilibrio de fuerzas y de momentos donde se utilizan las siguientes variables intermedias C_{un} , T_{un} y X_{un} (véase la figura 11.12). El conjunto n se utiliza para definir los intervalos en los que puede situarse la fibra neutra: $n \equiv \{1, 2, 3, 4, 5\}$.

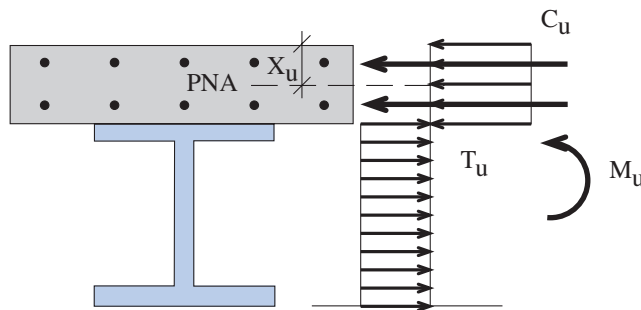


Figure 11.12: Plastificación por flexión.

Las tensiones, en las posiciones u y ℓ en el tiempo t , en el centro de la viga son:

$$\sigma_{u,t} = \frac{M_c(V_t - t_c)E_s}{EI_t} \quad (11.50)$$

$$\sigma_{\ell,t} = \frac{M_c(t_c + t_w + h_w + t_i - V_t)E_s}{EI_t} \quad (11.51)$$

$$\tau_c = \frac{V_c}{h_w t_w} \quad (11.52)$$

Acciones sobre la viga

En este caso se considerará el peso muerto de la viga, una sobrecarga uniformemente repartida sobre el tablero (p_s), una carga en faja repartida en todo el ancho del tablero (w_d) y una sobrecarga puntual (w_p).

El peso por unidad de longitud correspondiente a la viga es

$$w_1 = \gamma_c b_c t_c + \gamma_s [b_u t_u + b_\ell t_\ell + h_w t_w + (b_\ell - t_w) h_w t_r] \quad (11.53)$$

y la parte proporcional de la carga en faja es (véase la figura 11.13)

$$l_d = \frac{w_d b_c}{p_\ell} \quad (11.54)$$

la parte proporcional de la carga puntual considerando el factor de excentricidad (e_p) para tener en cuenta la existencia de n_b vigas es

$$C_p = \frac{w_p b_c e_p}{p_\ell} \quad (11.55)$$

y la parte proporcional de la sobrecarga uniformemente repartida considerando el factor de excentricidad (e_s) es

$$s_u = p_s b_c e_s \quad (11.56)$$

Por tanto, los momentos flectores y esfuerzos cortantes actuantes tanto en centro luz como en los apoyos son

$$M_c = \frac{(w_1 + l_d + s_u)L^2}{8000} + \frac{C_p L}{4000} \quad (11.57)$$

$$M_s = 0 \quad (11.58)$$

$$V_c = \frac{C_p}{2000} \quad (11.59)$$

$$V_s = \frac{(w_1 + l_d + s_u)L}{2000} + \frac{C_p}{1000} \quad (11.60)$$

Restricciones adicionales

Para simplificar se asume un puente esbelto, es decir,

$$\frac{pl}{n_b} \leq 0.1L \quad (11.61)$$

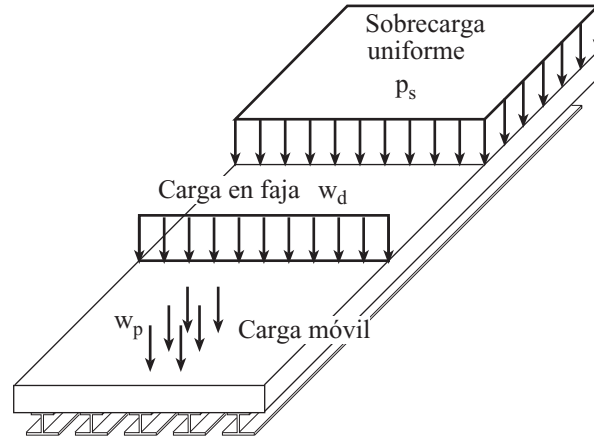


Figure 11.13: Cargas consideradas que actúan en el puente mixto.

y que el hormigón trabaja a compresión

$$V_t \geq t_c; t \in \{0, \infty\}. \quad (11.62)$$

Adicionalmente, para facilitar el refuerzo del alma, se impone

$$\bullet \quad b_u \leq b_\ell \quad (11.63)$$

y

$$b_c = \frac{pl}{n_b} \quad (11.64)$$

Requerimientos del código

Las siguientes restricciones están fijadas por el código español.

La anchura de acero debe satisfacer

$$e_{max} = 0.150\text{m} \quad \text{if} \quad f_{yk} = 235\text{MPa} \quad (11.65)$$

$$e_{max} = 0.084\text{m} \quad \text{if} \quad f_{yk} = 275\text{MPa} \quad (11.66)$$

$$e_{max} = 0.038\text{m} \quad \text{if} \quad f_{yk} = 355\text{MPa} \quad (11.67)$$

$$0.008 \leq t_u \leq e_{max} \quad (11.68)$$

$$0.008 \leq t_\ell \leq e_{max} \quad (11.69)$$

$$0.008 \leq t_w \leq e_{max} \quad (11.70)$$

$$\bullet \quad 0.008 \leq t_r \leq e_{max} \quad (11.71)$$

$$0.180 \leq t_c \quad (11.72)$$

$$b_\ell \leq 30t_\ell \quad (11.73)$$

La cantidad mínima de armadura es

$$\bullet \quad A_{s_1} \geq 0.0009b_c t_c \quad (11.74)$$

$$\bullet \quad A_{s_2} \geq 0.0009b_c t_c \quad (11.75)$$

Estado límite de deformaciones del alma: para evitar la abolladura del alma se tienen que cumplir las siguientes restricciones:

$$\frac{h_w}{t_w} \leq 0.55 \frac{E_s}{f_y} \sqrt{K} \quad (11.76)$$

$$\bullet \frac{h_w}{t_w} \leq 100 \sqrt{\frac{355}{f_y}} \quad (11.77)$$

donde $K = \frac{h_w t_w}{b_u t_u}$.

Restricciones asociadas a los conectores

Un correcto diseño de los conectores requiere que se cumplan las siguientes restricciones geométricas:

$$\bullet 4d_{st} \leq h_{st} \leq 0.75t_c \quad (11.78)$$

$$d_3 \geq 5d_{st} \quad (11.79)$$

$$2.5d_{st} \leq d_1 \leq 0.8 \text{ m} \quad (11.80)$$

$$d_2 = 0.05 \text{ m} \quad (11.81)$$

$$d_3 \leq 0.8 \text{ m} \quad (11.82)$$

$$d_{st} \leq 2.5t_u \quad (11.83)$$

$$d_1(m-1) = b_u - 2d_2 \quad (11.84)$$

$$22t_u \geq d_3 \sqrt{\frac{f_y}{235}} \quad (11.85)$$

$$35t_u \geq d_1 \sqrt{\frac{f_y}{235}} \quad (11.86)$$

$$9t_u \geq d_2 \sqrt{\frac{f_y}{235}} \quad (11.87)$$

Para evitar el fallo debido al hormigón que rodea a los conectores se ha de cumplir (véase la figura 11.14).

$$\bullet 0.6 \times 0.29 \frac{d_{st}^2}{1.25} \sqrt{f_c E_c} \geq N_{cs} d_3 / (mL/2) \quad (11.88)$$

y para evitar el fallo directo de los conectores

$$0.8 \times 450\pi \frac{d_{st}^2}{4 \times 1.25} \geq N_{cs} d_3 / (mL/2) \quad (11.89)$$

donde

$$N_{cs} = 0.85 f_c b_c t_c + (A_{s1} + A_{s2}) f_s \quad (11.90)$$

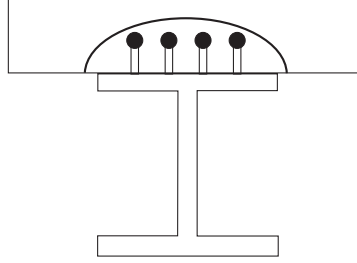


Figure 11.14: Ilustración del fallo de los conectores debido al hormigón circundante.

Restricciones asociadas al refuerzo del alma

Para el diseño del refuerzo del alma se han utilizado las siguientes restricciones:

$$I_r = \frac{1}{12}t_r b_u^3 + \frac{30t_w^4}{12} \sqrt{\frac{235}{f_y}} \quad (11.91)$$

$$\bullet I_r \geq \left[1 + 2 \left(\frac{h_w/t_w - 75}{75} \right) \right] \left(21h_w - \frac{15}{h_w} \right) \frac{h_w t_w^3}{12(1 - \nu^2)}; \nu = 0.3 \quad (11.92)$$

$$I_r \geq 1.5 \left(\frac{h_w}{50} \right)^4 \quad (11.93)$$

$$60t_r \geq b_\ell - t_w \quad (11.94)$$

Restricciones asociadas a las soldaduras de las chapas

Para una correcta soldadura entre chapas se ha de cumplir que:

$$\bullet t_u \geq 0.008 + \frac{t_w - 0.01}{7.5} \quad (11.95)$$

$$t_\ell \leq 0.01 + \frac{t_w - 0.008}{7.5} \quad (11.96)$$

Máximos esfuerzos cortantes

Finalmente, las restricciones por esfuerzo cortante son:

$$\tau_{cr} = 5.34E_s 0.9(t_w/h_w)^2 \quad (11.97)$$

$$\lambda_w = \sqrt{\frac{f_y}{\sqrt{3}\tau_{cr}}} \quad (11.98)$$

$$m_t = \exp(-0.227086\lambda_w^2) \quad (11.99)$$

$$\tau_{\max} = \frac{m_t f_y}{\sqrt{3}} \quad (11.100)$$

$$V_{\max} = \tau_{\max} h_w t_w \quad (11.101)$$

11.2.9 Notation

1. Datos

(a) Constantes de diseño

L : longitud de vano (m).

pl : ancho de plataforma (m).

d_{st} : diámetro del conector (m).

r_1 : recubrimiento inferior de la armadura (m).

r_2 : recubrimiento superior de la armadura (m).

n_b : número de vigas de acero.

m : número de conectores por fila (m).

(b) Propiedades de los materiales

f_{ck} : resistencia característica del hormigón (MPa).

f_{sk} : resistencia característica del acero para armaduras (MPa).

f_{yk} : resistencia característica del acero estructural (MPa).

E_t : módulo de elasticidad del hormigón a tiempo $t \in \{0, \infty\}$ (MPa).

E_s : módulo de elasticidad del acero (MPa)..

ϕ : coeficiente de fluencia del hormigón.

ν : módulo de Poisson.

(c) Propiedades estadísticas

μ_{w_d} : valor medio de w_d .

μ_{w_p} : valor medio de w_p .

μ_{p_s} : valor medio de p_s .

μ_{f_c} : valor medio de f_c .

μ_{f_s} : valor medio de f_s .

μ_{f_y} : valor medio de f_y .

μ_{γ_c} : valor medio de γ_c .

μ_{γ_s} : valor medio de γ_s .

v_{w_d} : coeficiente de variación de w_d .

v_{w_p} : coeficiente de variación de w_p .

v_{p_s} : coeficiente de variación de p_s .

v_{f_c} : coeficiente de variación de f_c .

v_{f_s} : coeficiente de variación de f_s .

v_{f_y} : coeficiente de variación de f_y .

v_{γ_c} : coeficiente de variación de γ_c .

v_{γ_s} : coeficiente de variación de γ_s .

(d) Datos relativos a las sobrecargas

e_p : coeficiente de excentricidad de la carga puntual (adimensional).

e_s : coeficiente de excentricidad de la sobrecarga uniformemente repartida (adimensional).

(e) **Límites inferiores de los coeficientes de seguridad.**

F_{c0}^0 : coeficiente de seguridad asociado al hormigón en el centro de la viga a tiempo 0.

$F_{c\infty}^0$: coeficiente de seguridad asociado al hormigón en el centro de la viga a tiempo ∞ .

F_{u0}^0 : coeficiente de seguridad asociado al ala superior de acero a tiempo 0.

$F_{u\infty}^0$: coeficiente de seguridad asociado al ala superior de acero a tiempo ∞ .

$F_{\ell 0}^0$: coeficiente de seguridad asociado al ala inferior de acero a tiempo 0.

$F_{\ell \infty}^0$: coeficiente de seguridad asociado al ala inferior de acero a tiempo ∞ .

F_w^0 : coeficiente de seguridad asociado al alma.

F_m^0 : coeficiente de seguridad asociado al momento último resistido por la sección.

F_v^0 : coeficiente de seguridad asociado al cortante último resistido por la sección.

(f) **Datos relativos al coste**

c_c : coste por metro cúbico de hormigón (euros/m³).

c_s : coste por Newton de acero para armaduras (euros/N).

c_y : coste por Newton de acero estructural (euros/N).

c_{st} : coste por conector (euros).

2. Variables

(a) **VARIABLES DE DISEÑO**

b_e : anchura del elemento $e \in \{c, u, \ell\}$ (m).

t_e : espesores o altura del elemento $e \in \{c, u, \ell\}$ (m).

h_w : altura del alma de acero (m).

t_w : espesor del alma de acero (m).

d_1, d_2, d_3 : separaciones entre conectores 1, 2, 3 (m).

A_{s_1} : armadura en la parte inferior de la losa de hormigón (m).

A_{s_2} : armadura en la parte superior de la losa de hormigón (m).

h_{st} : altura del conector (m).

t_r : espesor o altura del refuerzo del alma (m).

(b) **VARIABLES ALEATORIAS**

w_d : carga muerta (en faja) actuando sobre el ancho de plataforma (KN/m).

w_p : sobrecarga (carga puntual) (KN).

p_s : sobrecarga por unidad de superficie (KN/m²).

f_c : resistencia del hormigón usada en los cálculos (MPa).

f_s : resistencia del acero para armaduras usada en los cálculos (MPa).

f_y : resistencia del acero estructural usada en los cálculos (MPa).

γ_c : peso específico del hormigón usado en los cálculos (KN/m³).

γ_s : peso específico del acero (kN/m^3).

(c) **Variables auxiliares o intermedias**

V_p : esfuerzo cortante en la posición $p \in \{c, s\}$ (KN).

$\sigma_{e,t}$: tensiones normales en la posición $e \in \{c, u, \ell\}$ a tiempo $t \in \{0, \infty\}$ (MPa).

τ_p : tensión tangencial del alma en la posición $p \in \{c, s\}$ (MPa).

M_p : momento en la posición $p \in \{c, s\}$ (MN.m).

M_{un} : momento último resistido por la sección cuando se asume que la fibra neutra está en la posición $n \in \{1, 2, 3, 4, 5\}$ (MN).

V_{max} : máximo esfuerzo cortante soportado por el acero estructural (MN/m^2).

w_1 : peso total de la viga por unidad de longitud (KN/m).

l_d : carga muerta actuando sobre todo el ancho de la viga (KN/m).

C_p : carga puntual actuando sobre la viga (KN).

s_u : sobrecarga actuando sobre la viga por unidad de longitud (KN/m).

V_t : situación del centro de gravedad de la viga a tiempo $t \in \{0, \infty\}$ (m).

$A_{e,t}$: área de hormigón en la posición $e \in \{c, u, \ell\}$ a tiempo $t \in \{0, \infty\}$ (m^2).

A_w : área del alma (m^2).

I_t : momento de inercia de la losa de hormigón a tiempo $t \in \{0, \infty\}$ (m^4).

I_s : momento de inercia del alma (m^4).

I_r : momento de inercia del refuerzo del alma (m^4).

Y_e : coordenada del centro de gravedad de la losa de hormigón en la posición $e \in \{c, u, \ell\}$ (m).

Y_w : localización del centro de gravedad del alma de acero (m).

EI_t : rigidez del hormigón a tiempo $t \in \{0, \infty\}$ (MNm^2).

δ_t : flecha de la viga en centro luz a tiempo $t \in \{0, \infty\}$ (m).

δ : flecha de la viga en centro luz a tiempo ∞ tras sustraerla la flecha a tiempo 0 (m).

δ_{max} : flecha máxima permitida en centro luz (m).

A_{min} : cuantía mínima de armadura (m^2).

e_{max} : espesor máximo de las chapas de acero (m).

C_{un} : esfuerzo de compresión cuando se asume que la fibra neutra está en la posición $n \in \{1, 2, 3, 4, 5\}$ (MN).

T_{un} : esfuerzo de tracción cuando se asume que la fibra neutra está en la posición $n \in \{1, 2, 3, 4, 5\}$ (MN).

X_{un} : posición actual del eje neutro cuando se asume que la fibra neutra está en la posición $n \in \{1, 2, 3, 4, 5\}$ (MN).

τ_{cr} : máxima tensión tangencial soportada por el alma de la viga (MPa).

λ_w : coeficiente auxiliar para calcular la máxima tensión tangencial soportada por el alma.

m_t : coeficiente auxiliar para calcular la máxima tensión tangencial soportada por el alma.

τ_{\max} : máxima tensión tangencial soportada por el acero estructural (MPa).

P_{st} : máxima fuerza soportada por los conectores (MN).

N_{cs} : máxima fuerza soportada por la losa de hormigón (MN).

Figure 11.15: Parametrización del dique en talud usada en el ejemplo ilustrativo.

11.3 Rubble-Mound Breakwaters

En esta sección se describe con detalle el cálculo de un dique en talud basado en técnicas probabilistas. Existen en la literatura trabajos basados en técnicas de optimización y en coeficientes de seguridad, o en probabilidades de fallo (véase por ejemplo, USACE, PIANC, IWR, etc.). Incluso Vrouwenvelder [173] propuso una combinada aplicada a la calibración de códigos.

En el Manual de Ingeniería Costera de la USACE (U.S.A Corps of Engineers) y en el resultado final del proyecto MAST III (Marine Science and Technology Programme), ‘Probabilistic Design Tools for Vertical Breakwaters’ de la Unión Europea (1996-1999), se incluyen aspectos sobre el modelado físico de los procesos relevantes que afectan a este tipo de estructuras, considerando los aspectos estadísticos de las variables que intervienen. Se proponen dos métodos alternativos: el primero en el que se propone un procedimiento de optimización para el cálculo del nivel de seguridad óptimo, y el segundo basado en una aproximación clásica mediante coeficientes de seguridad. La metodología presentada en este ejemplo puede considerarse una continuación de las investigaciones llevadas a cabo por el USACE y el MAST III, en la que se combinan las dos metodologías prevaleciendo la más restrictiva.

El diseño paramétrico del dique se muestra en la figura 11.15, mientras que el listado de todas las variables y parámetros que intervienen se da en la subsección 11.3.8. Obsérvese que todos estos parámetros definen geoméricamente los diferentes elementos de la sección transversal del dique y deben definirse en los planos constructivos del proyecto. El objetivo del presente ejemplo es lograr un diseño óptimo desde el punto de vista económico, sujeto a un conjunto de restricciones que garantizan la seguridad de la estructura con respecto a los modos de fallo considerados, con una perspectiva clásica, es decir, mediante la utilización de coeficientes de seguridad y desde la óptica del diseño probabilista. Para lograr un diseño óptimo desde el punto de vista económico deberían considerarse los costes de construcción, mantenimiento, reparación, paradas operativas, desmantelamiento al final de su vida útil, etc.; sin embargo, para simplificar el problema y su exposición, se ha limitado el presente estudio al coste constructivo.

11.3.1 Variable classification and failure modes considered

El conjunto de variables de diseño (X_1, \dots, X_n) se dividirá de forma detallada en los cinco subconjuntos siguientes (véase la Subsección 11.3.8 para la notación):

d: Variables de diseño a optimizar.

$$\mathbf{d} = (a, b, c, d, e, F_c, f, n, p, q, r, s, t, \alpha_\ell, \alpha_s)$$

η : Conjunto de parámetros dato utilizados en el diseño clásico.

$$\boldsymbol{\eta} = (C_{al}, C_c, C_{co}, C_{ul}, D_{WL}, g, \gamma_c, \gamma_s, \gamma_w)$$

ϕ : Variables aleatorias básicas usadas en los modelos probabilistas.

$$\phi = (H_s, A_r, A_u, B_r, B_u, C_f, H, T, \mu_c, h_1, h_2)$$

κ : Parámetros estadísticos que definen la variabilidad y dependencia de las variables aleatorias.

$$\kappa = (D, v_{A_r}, v_{B_r}, v_{C_f}, \mu_{A_r}, \mu_{B_r}, \mu_{A_u}, \mu_{B_u}, \mu_{C_f}, \mu_{\mu_c}, \mu_{h_1}, \mu_{h_2}, \sigma_{A_u}, \sigma_{B_u}, \sigma_{\mu_c}, \sigma_{h_1}, \sigma_{h_2})$$

ψ : Variables auxiliares o no básicas cuyos valores se obtienen a partir de las variables pertenecientes a los conjuntos $(\mathbf{d}, \boldsymbol{\eta}, \phi)$ mediante alguna fórmula.

$$\psi = (A_c, F_h, F_v, I_r, I_{r_0}, L, \ell, \ell_e, P_{S_0}, R, R_u, S_0, V_c, V_1, V_2, V_3, W, W_1, W_f, \alpha, \lambda, \phi_e)$$

Para el presente estudio sólo se han considerado como variables aleatorias las básicas ϕ y las correspondientes no básicas, es decir que todas las variables ψ que se obtienen mediante alguna fórmula, la cual contiene una variable aleatoria, será considerada aleatoria.

Aunque para el estudio de un dique en talud se pueden y deben considerar modos de fallo tales como, por ejemplo, estabilidad de las piezas de los mantos, estabilidad del núcleo, fallos geotécnicos, etc. Sólo se consideran en este ejemplo tres modos de fallo, que se incluirán en el procedimiento de optimización: (1) rebase, (2) estabilidad del manto principal y (3) deslizamiento del espaldón.

Algunos de los modos considerados están correlacionados, debido a que poseen agentes comunes, o porque la ocurrencia de alguno de ellos puede inducir a otros. Así por ejemplo el hecho de que se produzca una extracción de las piezas del manto principal, favorece la extracción de las piezas del manto secundario y posteriormente el núcleo, y puede dar lugar a una socavación progresiva que lleve a la ruina total del dique. En este ejemplo sólo se considerará la correlación debida a agentes comunes, ya que la interacción es complicada y aún no se comprende de forma adecuada, seguramente ésta será un área de investigación muy activa en un futuro próximo. Por tanto la fiabilidad calculada será una estimación incompleta, no porque la metodología empleada no pueda tenerla en consideración, sino por la falta de información adecuada. Las necesidades proyectistas del futuro requerirán un gran trabajo experimental en este campo.

Estrictamente hablando, algunos de los modos de fallo considerados, tales como la extracción de piezas del manto, o el rebase, no se corresponden con la definición clásica de fallo estructural (colapso de la estructura), ya que por un lado, el manto es una estructura granular donde la conectividad y transmisión de esfuerzos se produce por la fricción y el grado de acoplamiento intergranular, y por otro, en el caso del rebase, no sólo importa si se produce o no, sino el volumen de agua que sobrepasa la coronación. Así pues, estos modos de fallo están relacionados con unos criterios preestablecidos o umbrales, que nos dan una indicación sobre la resistencia residual de la estructura antes del colapso. Por tanto como otras muchas disciplinas de la ingeniería civil, la ocurrencia de fallo no necesariamente significa que la estructura haya colapsado, sino que responde a otros criterios como que su resistencia se ha visto considerablemente mermada, o que se ha reducido considerablemente la funcionalidad para la que fue diseñada.

Para el caso de diques en talud, hay al menos tres criterios de fallo posible: inicio de avería, avería de Iribarren y destrucción, (véase Losada, 1990 para una discusión completa). Un desplazamiento de las piezas del orden de un 2-5% es un criterio típico para considerar inicio de avería. La elección de uno u otro criterio para el diseño, depende de varios factores, entre ellos, la capacidad o incapacidad de poder efectuar reparaciones cuando sea preciso, las consecuencias ambientales y económicas del fallo, etc.

Rebase. Para un dique en talud de pendiente $\tan \alpha_s$ y francobordo F_c , (véase la figura 11.15), y para una altura de ola H y un periodo T dados, el volumen de agua que sobrepasa la estructura puede estimarse del volumen de agua que sobrepasaría el nivel del francobordo, si el talud fuera indefinido. Con esta aproximación, el rebase se produce siempre que la diferencia entre la máxima cota alcanzada por el agua en el talud imaginario, llamada run-up (R_u), exceda el nivel del francobordo F_c .

Losada (1990) propuso la siguiente ecuación de verificación basada en la experimentación para evaluar el término adimensional R_u/H :

$$\frac{R_u}{H} = A_u (1 - e^{B_u I_r}) \quad (11.102)$$

donde A_u y B_u son coeficientes que dependen del tipo de piezas del manto e I_r es el número de Iribarren

$$I_r = \frac{\tan \alpha_s}{\sqrt{H/L}} \quad (11.103)$$

donde L es la longitud de onda de la ola y α_s es el ángulo del talud de barlovento.

Bajo tales condiciones, la ocurrencia de fallo se puede comprobar con la siguiente ecuación de verificación:

$$\frac{F_c}{R_u} \geq F_o \quad (11.104)$$

donde F_o es el coeficiente de seguridad al rebase.

Obsérvese, que para clarificar y simplificar el problema, la ecuación de verificación (11.104) sólo indica si se produce o no rebase, para un estudio adecuado debería calcularse el volumen de agua que sobrepasa la estructura, ya que el nivel de daño en el intradós del espaldón depende de esa cuantía.

Estabilidad del manto principal. Este fallo hace referencia a la extracción de las piezas (cubos, escollera, dolos, ...) del manto principal. Basándose en resultados experimentales, Losada (1990) propuso la siguiente ecuación de verificación para evaluar el término adimensional $\frac{W}{\gamma_w H^3}$:

$$\frac{W}{\gamma_w H^3} = R \Phi_e \quad (11.105)$$

donde γ_w es el peso específico del agua, Φ_e es la función de estabilidad, R es una constante adimensional, que depende de γ_c y γ_w , y W es el peso individual de los bloques, que están dados por

$$W = \gamma_c \ell^3 \quad (11.106)$$

$$R = \frac{\gamma_c/\gamma_w}{\left(\frac{\gamma_c}{\gamma_w} - 1\right)^3} \quad (11.107)$$

$$\Phi_e = A_r(I_r - I_{r_0}) \exp[B_r(I_r - I_{r_0})] \quad (11.108)$$

$$I_r \geq I_{r_0} \quad (11.109)$$

donde γ_c es el peso específico del hormigón (cubos), ℓ es la longitud característica del lado del cubo (se supone que el manto principal está constituido por dos capas de bloques cúbicos), y la expresión (11.109) define el rango de validez del modelo,

$$I_{r_0} = 2.656 \tan \alpha_s, \quad (11.110)$$

y A_r y B_r dependen de $\cot \alpha_s$ por medio de las expresiones experimentales

$$A_r = 0.2566 - 0.177047 \cot \alpha_s + 0.0342933 \cot^2 \alpha_s \quad (11.111)$$

$$B_r = -0.0201 - 0.4123 \cot \alpha_s + 0.055 \cot^2 \alpha_s \quad (11.112)$$

que son válidas siempre y cuando se cumpla que $1.5 \leq \cot \alpha_s \leq 3$.

Bajo tales condiciones, la ocurrencia de fallo por extracción de piezas del manto puede determinarse mediante la ecuación de verificación:

$$\frac{W}{\gamma_w R \Phi_e H^3} \geq F_a \quad (11.113)$$

donde F_a es el coeficiente de seguridad a fallo del manto principal.

Deslizamiento del espaldón. Este fallo ocurre cuando el espaldón desliza sobre su base por efecto de la presión del agua. Se produce cuando se cumple la siguiente ecuación de verificación (véase la figura 11.16(b))

$$\mu_c(W_1 - F_v) = F_h \quad (11.114)$$

donde μ_c es el coeficiente de rozamiento al deslizamiento, y (véase Martin et al. (1999)):

$$F_h = (S_0 - A_c)P_{S_0} + (W_f + A_c)\lambda P_{S_0} \quad (11.115)$$

$$F_v = \frac{1}{2}\lambda P_{S_0} e \quad (11.116)$$

$$W_1 = V_c \gamma_c - W_f e \gamma_w \quad (11.117)$$

$$V_c = pq + se \quad (11.118)$$

donde F_h y F_v son la fuerza total horizontal y vertical debido a la presión del agua, S_0 es el nivel de agua que alcanzaría la ola debido al run-up si el talud fuera indefinido, A_c es la cota a la que está situada la berma, P_{S_0} es la presión que ejerce la ola sobre el espaldón, W_f es la profundidad de agua a la que está sumergido el espaldón, e es el ancho del espaldón, $V_c \gamma_c$ es el peso total de hormigón del espaldón, y W_1 es su peso actual (parte seca y parte sumergida).

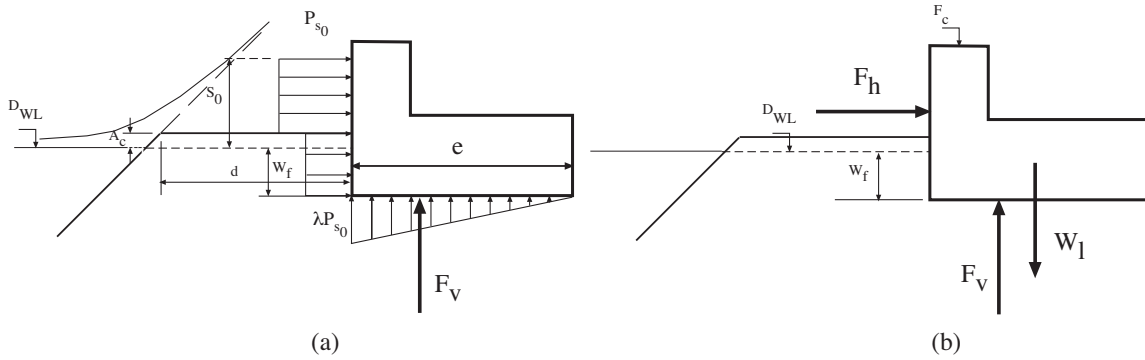


Figure 11.16: Fuerzas que actúan sobre el espaldón y que pueden producir el deslizamiento con respecto a su base.

Para el caso en el que $I_r \geq 2$, las presiones que actúan en el espaldón al nivel z son (véase la figura 11.16):

$$P_d(z) = \begin{cases} \lambda P_{S_0} & \text{if } z < A_c \\ P_{S_0} & \text{if } z > A_c \end{cases} \quad (11.119)$$

donde

$$I_r \geq 2 \quad (11.120)$$

$$P_{S_0} = \alpha \gamma_w S_0 \quad (11.121)$$

$$S_0 = H \left(1 - \frac{A_c}{R_u} \right) \quad (11.122)$$

$$\alpha = 2C_f \left(\frac{R_u}{H} \cos \alpha_s \right)^2 \quad (11.123)$$

$$\lambda = 0.8 \exp \left(-10.9 \frac{d}{L} \right) \quad (11.124)$$

$$\left(\frac{2\pi}{T} \right)^2 = g \frac{2\pi}{L} \tanh \left(\frac{2\pi D_{WL}}{L} \right) \quad (11.125)$$

donde g es la aceleración de la gravedad, d es la anchura de la berma, α es un coeficiente adimensional, D_{WL} es el nivel del mar de diseño, y C_f es un coeficiente experimental.

Obsérvese que sólo se verifica el primer pico de presión; pero la presión reflejada puede actuar a posteriori (véase Martin et al. (1999) para más detalles). El fallo por deslizamiento puede verificarse mediante la siguiente expresión

$$\frac{\mu_c (W_1 - F_v)}{F_h} \geq F_s \quad (11.126)$$

donde F_s es el coeficiente de seguridad frente al deslizamiento.

11.3.2 Design criteria

En ingeniería oceanográfica hay ciertas normas de buena práctica que deben tenerse en consideración. Algunas son propias de cada país y otras tienen raíces históricas. Las usadas en este ejemplo son¹ (véase la figura 11.15):

1. Espesores de capa y ancho de la berma:

$$a = 2\ell \quad (11.127)$$

$$b = 2\ell_e \quad (11.128)$$

$$\bullet d \geq 2\ell \quad (11.129)$$

$$\bullet f \geq 2\ell_e \quad (11.130)$$

2. Condiciones de filtro:

$$\bullet \frac{W}{20} \leq \ell_e^3 \gamma_s \quad (11.131)$$

$$\ell_e^3 \gamma_s \leq \frac{W}{10} \quad (11.132)$$

3. Condicionamientos constructivos:

$$b + c \leq D_{WL} \quad (11.133)$$

$$D_{WL} \leq a + b + c \quad (11.134)$$

$$t = 1m \quad (11.135)$$

$$r = 2t \quad (11.136)$$

$$p = 2 \quad (11.137)$$

4. Razones económicas:

$$1.5 \leq \cot \alpha_s \leq 3 \quad (11.138)$$

$$\bullet 1.5 \leq \cot \alpha_\ell \leq 3 \quad (11.139)$$

5. Rotura del oleaje y control del rebase:

$$\bullet A_c \geq 3H_s/4 \quad (11.140)$$

$$F_c \geq S_0 \quad (11.141)$$

6. Identidades geométricas:

$$D_{WL} + F_c = b + c + s + q \quad (11.142)$$

$$W_f = s + q - F_c \quad (11.143)$$

$$A_c = a + b + c - D_{WL} \quad (11.144)$$

¹ Círculos negros frente a las inecuaciones indican que en la solución numérica del ejemplo éstas son activas, es decir, degeneran en igualdades

7. Requerimientos del modelo:

$$s \leq 12(\bullet); \quad 2 \leq q \leq 8(\bullet); \quad e \geq 10 \quad (11.145)$$

donde ℓ y ℓ_e son la longitud de cubo equivalente para los mantos principal y secundario, respectivamente, y H_s es la altura de ola significante, parámetro descriptor del estado de mar de cálculo.

11.3.3 Objective function

Para el ejemplo que nos ocupa, la función a optimizar es el coste de construcción:

$$C = C_c V_c + C_{al} V_1 + C_{ul} V_2 + C_{co} V_3 \quad (11.146)$$

donde V_c, V_1, V_2 y V_3 son los volúmenes de hormigón, manto principal, manto secundario y núcleo, respectivamente, y C_c, C_{al}, C_{ul} y C_{co} son los respectivos costes por unidad de volumen.

Considerando el dique en talud de la figura 11.15, los volúmenes se calculan con las siguientes expresiones:

Hormigón del espaldón:

$$V_c = pq + se \quad (11.147)$$

Manto principal:

$$V_1 = ad + a \frac{b+c-t}{\sin \alpha_s} + \frac{a^2}{2 \tan \alpha_s} \quad (11.148)$$

Manto secundario:

$$V_2 = b \left(e + d - a \tan \frac{\alpha_s}{2} \right) + b \frac{c-t}{\sin \alpha_s} + \frac{b^2}{2 \tan \alpha_s} + \left(2f + \frac{n+b}{\tan \alpha_\ell} \right) \frac{n+b}{2} + \frac{cb}{\sin \alpha_\ell} \quad (11.149)$$

Núcleo:

$$V_3 = c \left(f + e + d - (a+b) \tan \frac{\alpha_s}{2} - \frac{b}{\sin \alpha_\ell} + \frac{n+b}{\tan \alpha_\ell} \right) + \frac{c^2}{2} \left(\frac{1}{\tan \alpha_s} + \frac{1}{\tan \alpha_\ell} \right) + t \left(r + \frac{a+b}{\sin \alpha_s} \right) \quad (11.150)$$

11.3.4 Distributional assumptions for the model

Asumiendo una vida útil del dique en talud de 50 años, y con el fin de poder aplicar el método propuesto, es necesario definir el conjunto de variables deterministas y aleatorias, y sus dependencias.

Parámetros:

1. Geométricos:

- (a) El francobordo, F_c , y las pendientes de los taludes α_s y α_ℓ vienen dados por sus correspondientes valores nominales, y se consideran deterministas.

2. Propiedades de los materiales:

- (a) Los pesos específicos γ_c , γ_w y γ_s de hormigón, agua y escollera vienen dados por sus valores nominales correspondientes, y se consideran deterministas.

3. Propiedades mecánicas:

- (a) El coeficiente de rozamiento μ_c se considera una variable normal con media $\mu_{\mu_c} = 0.6$ y valor característico $\mu_{c0.05} = 0.55$.

4. Experimentales:

- (a) Modelo de run-up

$$\frac{R_u}{H} = A_u (1 - e^{B_u I_r}) \quad (11.151)$$

- A_u es una variables normal $N(\mu_{A_u}, \sigma_{A_u}^2)$, donde $\mu_{A_u} = 1.05$ y $\sigma_{A_u} = 0.21$.
- B_u es una variables normal $N(\mu_{B_u}, \sigma_{B_u}^2)$, donde $\mu_{B_u} = -0.67$ y $\sigma_{B_u} = 0.134$. Estas variables se han obtenido por un procedimiento de ajuste, y dado que su correlación es prácticamente nula se considerarán independientes.

- (b) Modelo de estabilidad del manto

$$\Phi_e = A_r (I_r - I_{r0}) \exp[B_r (I_r - I_{r0})] \quad (11.152)$$

- A_r es una variable normal $N(\mu_{A_r}, \sigma_{A_r}^2)$, donde

$$\mu_{A_r} = 0.2566 - 0.177047 \cot \alpha_s + 0.0342933 \cot^2 \alpha_s \quad (11.153)$$

- B_r es una variable normal $N(\mu_{B_r}, \sigma_{B_r}^2)$, donde

$$\mu_{B_r} = -0.0201 - 0.4123 \cot \alpha_s + 0.055 \cot^2 \alpha_s \quad (11.154)$$

5. Modelo de deslizamiento

- (a) C_f es una variable normal con media μ_{C_f} y desviación típica σ_{C_f} .
- (b) α es una variable normal cuya distribución puede obtenerse a partir de la ecuación (11.123).
- (c) λ es determinista y viene dada por su valor nominal.
- (d) S_0 es una variable aleatoria cuya distribución puede obtenerse de la ecuación (11.122).
- (e) P_d y P_{S_0} son variables aleatorias cuyas distribuciones pueden obtenerse a partir de las ecuaciones (11.119) y (11.121), respectivamente.

Agentes:

1. Climáticos:

- (a) La marea astronómica h_1 se considera una variable uniforme $U(zp, zp+tr)$, donde zp es el cero del puerto y tr es la carrera de marea.
- (b) La marea meteorológica h_2 se considera una variable normal $N(\mu_{h_2}, \sigma_{h_2}^2)$. Así, el nivel del mar será $D_{WL} = h_1 + h_2$.
- (c) La altura máxima de ola H_D y el periodo de pico T_p en un estado de mar son variables aleatorias con funciones de distribución:

$$F_{H_D}(x; H_s) = 1 - \exp \left[- \left(\frac{x - 1.263 - 0.326H_s - 0.172H_s^2}{1.465} \right)^{2.507} \right];$$

$$x \geq 1.263 + 0.326H_s + 0.172H_s^2 \quad (11.155)$$

y

$$F_{T_p}(x; H_s) = \exp \left[- \left(\frac{11.176 + 3.756H_s - 0.415H_s^2 - x}{7.597} \right)^{2.992} \right]; \quad (11.156)$$

$$x \leq 11.176 + 3.756H_s - 0.415H_s^2$$

donde H_s es una variable aleatoria con función de densidad exponencial:

$$f_{H_s}(x) = 1.58247 \exp(-1.58247(x - 3)); \quad x \geq 3;$$

donde sólo se consideran estados de mar superiores a tres metros ($H_s > 3$).

Estas distribuciones de H_D y T_p como función de H_s , y la distribución de H_s se han derivado de los datos de la boya de Gijón. Los datos recogen la información de 5.69 años, 260 tormentas ($H_s > 3$), y 1857 estados de mar (de una hora de duración).

Dependencia entre las variables:

Como se explicó en la subsección 11.3.1, todas las variables aleatorias mencionadas se asumen independientes, y las demás se considerarán deterministas. Es importante recalcar que el motivo de esta hipótesis no se debe a la imposibilidad del método para tratar correlación entre variables, sino a la falta de entendimiento y de datos sobre la interacción entre ellas. Además, los modos de fallo podrán estar correlacionados, debido a la presencia de variables comunes. En otras palabras, la principal fuente de correlación entre modos de fallo se debe a su dependencia en variables comunes y no a la dependencia de las variables en sí mismas.

La elección de las funciones estadísticas y de sus parámetros se ha elegido con propósitos ilustrativos. Para aplicar el método a casos reales se ha de realizar una selección mucho más cuidadosa, usando registros de datos de periodos largos. Sólo unos pocos países tienen información suficiente para obtener estas funciones de forma adecuada.

11.3.5 Formulation and solution of the dual problem

Siguiendo el procedimiento descrito en el capítulo 9, el primer paso es seleccionar los límites inferiores de los coeficientes de seguridad F_o^0, F_s^0 y F_a^0 , y los límites inferiores de los índices de fiabilidad β_o^0, β_s^0 y β_a^0 ; el siguiente es comenzar con el procedimiento iterativo hasta que se cumplan todas las restricciones.

Diseño clásico óptimo. Para obtener el diseño clásico en la iteración k , se minimiza la función de coste

$$c(\bar{\mathbf{d}}, \tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\phi}}, \boldsymbol{\psi}) \quad (11.157)$$

sujeta a las restricciones asociadas a los coeficientes de seguridad:

Rebase:

$$\frac{F_c}{R_u} \geq F_o^k \quad (11.158)$$

Estabilidad del manto principal:

$$\frac{W}{\gamma_w R \Phi_e H^3} \geq F_a^k \quad (11.159)$$

Deslizamiento del espaldón:

$$\frac{\mu_c(W_1 - F_v)}{F_h} \geq F_s^k \quad (11.160)$$

y a las restricciones de diseño.

Diseño óptimo probabilista. Para evaluar los límites inferiores de los índices de fiabilidad β , que es equivalente a calcular los límites superiores de las probabilidades de fallo de cada uno de los modos de fallo considerados, se aplica el procedimiento descrito en la sección 4.7. Nótese que se va a calcular la probabilidad de fallo durante la vida útil de la obra. Consecuentemente, la probabilidad de fallo para el modo i es:

$$P_{f_i} = 1 - \left(1 - \int_3^\infty f_{H_s}(H_s) P_{f_{ss_i}}(H_s) dH_s \right)^{n_{ss}D} \quad (11.161)$$

donde $P_{f_{ss_i}}(H_s)$ es la probabilidad de fallo para el modo i en un estado de mar definido por H_s , y n_{ss} es el número equivalente de estados de mar al año ($n_{ss} = 326$ para los datos de Gijón) y D es la vida útil. La integral (11.161) puede calcularse usando la fórmula de cuadratura de Gauss-Legendre como:

$$P_{f_i} \approx 1 - \left(1 - \frac{H_s^{max} - H_s^{min}}{2} \sum_{j=1}^n w_j f_{H_s}(H_{s_j}) P_{f_{ss_i}}(H_{s_j}) \right)^{n_{ss}D} \quad (11.162)$$

donde w_j y H_{s_j} son respectivamente los puntos y los pesos de Gauss.

Como se necesita $P_{f_{ss_i}}(H_{s_j})$ se aplica el siguiente procedimiento para cada H_{s_j} :

1. Transformación de Rosenblatt: dado que en el presente ejemplo todas las variables aleatorias se han considerado independientes la transformación se simplifica mucho

$$\begin{aligned}
u_1 &= \Phi((A_u - \mu_{A_u})/\sigma_{A_u}) = \Phi(z_1) \\
u_2 &= \Phi((B_u - \mu_{B_u})/\sigma_{B_u}) = \Phi(z_2) \\
u_3 &= \Phi((A_r - \mu_{A_r})/\sigma_{A_r}) = \Phi(z_3) \\
u_4 &= \Phi((B_r - \mu_{B_r})/\sigma_{B_r}) = \Phi(z_4) \\
u_5 &= F_{T_p}(T; H_{s_j}) = \Phi(z_5) \\
u_6 &= F_{H_D}(H; H_{s_j}) = \Phi(z_6) \\
u_7 &= \Phi((\mu_c - \mu_{\mu_c})/\sigma_{\mu_c}) = \Phi(z_7) \\
u_8 &= \Phi((C_f - \mu_{C_f})/\sigma_{C_f}) = \Phi(z_8) \\
u_9 &= (h_1 - zp)/tr = \Phi(z_9) \\
u_{10} &= \Phi((h_2 - \mu_{h_2})/\sigma_{h_2}) = \Phi(z_{10})
\end{aligned} \tag{11.163}$$

2. Transformación a variables normales tipificadas (normalizadas $N(0, 1)$):

$$\begin{aligned}
z_1 &= (A_u - \mu_{A_u})/\sigma_{A_u} \\
z_2 &= (B_u - \mu_{B_u})/\sigma_{B_u} \\
z_3 &= (A_r - \mu_{A_r})/\sigma_{A_r} \\
z_4 &= (B_r - \mu_{B_r})/\sigma_{B_r} \\
z_5 &= \Phi^{-1}(F_{T_p}(T; H_{s_j})) \\
z_6 &= \Phi^{-1}(F_{H_D}(H; H_{s_j})) \\
z_7 &= (\mu_c - \mu_{\mu_c})/\sigma_{\mu_c} \\
z_8 &= (C_f - \mu_{C_f})/\sigma_{C_f} \\
z_9 &= \Phi^{-1}((h_1 - zp)/tr) \\
z_{10} &= (h_2 - \mu_{h_2})/\sigma_{h_2}
\end{aligned} \tag{11.164}$$

3. Resolución de los siguientes problemas de optimización para cada H_{s_j} :

$$\begin{aligned}
&\text{Minimizar} && \beta_{i,j}^2 = \sum_{k=1}^{10} z_k^2 \\
&A_u, B_u, H, T, A_r, B_r, C_f, \mu_c, h_1, h_2
\end{aligned} \tag{11.165}$$

sujeto a (11.163) y a alguna de las siguientes restricciones.

Rebase:

$$1 = \frac{F_c}{R_u} \tag{11.166}$$

Estabilidad del manto principal:

$$1 = \frac{W}{\gamma_w R \Phi_e H^3} \tag{11.167}$$

Deslizamiento del espaldón:

$$1 = \frac{\mu_c(W_1 - F_v)}{F_h} \tag{11.168}$$

4. La probabilidad de fallo para el modo i en un estado de mar definido por H_{s_j} se calcula como:

$$P_{f_{ss_i}}(H_{s_j}) = \Phi(-\beta_{i,j}) \quad (11.169)$$

Una vez que las probabilidades de fallo para las distintas alturas de ola significativa están calculadas, la probabilidad de fallo durante la vida útil de la obra se calcula usando (11.162). Y su índice de fiabilidad β asociado es:

$$\beta_i^k = -\Phi^{-1}(P_{f_i}) \quad (11.170)$$

Comprobación de la convergencia y actualización de los coeficientes de seguridad. En esta etapa se comprueba si los índices de fiabilidad obtenidos (β^k) satisfacen los límites deseados. Si la respuesta es afirmativa, el procedimiento se detiene, y el diseño final es el obtenido por el diseño clásico. Si es negativa, se actualizan los coeficientes de seguridad mediante las ecuaciones (10.70) o (??) y se vuelve a la etapa 1.

11.3.6 Numerical example and analysis of the results

El método propuesto ha sido implementado en GAMS (General Algebraic Modeling System)(véase Castillo, Conejo, Pedregal, García and Alguacil (2001)), con un dique en talud con las siguientes características:

1. Los límites inferiores de los coeficientes de seguridad y de los índices de fiabilidad han de depender de las consecuencias asociadas al fallo de cada uno de los modos considerados durante la vida útil de la obra (se ha considerado un periodo de 50 años) (véase Vrouwenvelder (2002)). Así, los límites inferiores de los coeficientes de seguridad iniciales ($k = 0$) son:

$$F_o^0 = 1.05; \quad F_s^0 = 1.5; \quad F_a^0 = 1.05, \quad (11.171)$$

y los límites inferiores de los índices de fiabilidad (β) son:

$$\beta_o^0 = 2.32; \quad \beta_s^0 = 3.08; \quad \beta_a^0 = 3.08, \quad (11.172)$$

que se corresponden con probabilidades de fallo de 0.01 y 0.001, respectivamente.

2. Para el diseño clásico, se toma un estado de mar de diseño de $H_s = 6.5$. La altura de ola máxima y el periodo de pico de la ola se toman como los cuantiles 0.99 de ambas distribuciones, las demás variables aleatorias se igualan a sus valores medios, aunque podrían igualarse a sus valores característicos. El estado de mar de cálculo se toma como $D_{WL} = zp + tr$.
3. Los valores nominales de las variables, los parámetros estadísticos y los costes del ejemplo numérico se dan en la tabla 11.11.

Table 11.11: Valores nominales, estadísticos y costes del ejemplo numérico.

VALORES DE DISEÑO
<i>Valores nominales</i>
$p = 2m; q = 4m; e = 8m; r = 2m; t = 1m;$ $zp = 20m; tr = 5m; H_s = 5m; D_{WL} = 25m;$
$\gamma_c = 23.5KN/m^3; \gamma_s = 26KN/m^3; \gamma_w = 10.25KN/m^3$
<i>Propiedades estadísticas</i>
$\mu_{A_u} = 1.05; \sigma_{A_u} = 0.21; \mu_{B_u} = -0.67; \sigma_{B_u} = 0.134; \mu_{\mu_c} = 0.6; \sigma_{\mu_c} = 0.01941;$ $\mu_{C_f} =; \sigma_{C_f} = 1.45; \mu_{h_2} = 0.02414; \sigma_{h_2} = 0.11597;$ $\mu_{A_r} = 0.2566 - 0.177047 \cot \alpha_s + 0.0342933 \cot^2 \alpha_s; \sigma_{A_r} = 0.21;$ $\mu_{B_r} = -0.0201 - 0.4123 \cot \alpha_s + 0.055 \cot^2 \alpha_s; \sigma_{B_r} = 0.134$
<i>Coste</i>
$C_{al} = 818.4euros/m^3; C_{ul} = 18.72euros/m^3; C_{co} = 2.4euros/m^3; C_c = 60.1euros/m^3$

Análisis de resultados La tabla 11.12 muestra la convergencia del método, que para este caso se alcanza tras 14 iteraciones. La primera columna muestra los valores de las variables geométricas, los coeficientes de seguridad y los correspondientes índices de fiabilidad, es decir, aquellos que resultan de utilizar los coeficientes de seguridad en (11.171) sin consideración alguna sobre las probabilidades de fallo (índices de fiabilidad en (11.172)). Sin embargo, como los índices de fiabilidad obtenidos con este diseño son menores que los límites inferiores elegidos (11.172), los coeficientes de seguridad deben incrementarse usando (10.70) o (??), y repetirse el diseño clásico con los nuevos coeficientes de seguridad, en la iteración 2. Como ocurre lo mismo con los índices de fiabilidad de la segunda iteración, el procedimiento continúa hasta que se llega a la iteración 14, en la que se satisfacen las restricciones (11.172).

La última columna de la tabla muestra los valores de las variables de diseño, junto con los factores de seguridad y sus índices de fiabilidad asociados. Los valores activos aparecen subrayados en la tabla.

A la vista de la tabla se puede concluir lo siguiente. Los límites inferiores de los coeficientes de seguridad F_o^0 , F_s^0 y F_a^0 son inactivos, lo mismo que β_o^0 . Mientras que los límites β_s^0 y β_a^0 son activos, lo que implica que los límites probabilistas asociados a sus modos de fallo son más restrictivos que los asociados a los coeficientes de seguridad, y que la seguridad frente a rebase queda garantizada con la seguridad frente a estos modos de fallo.

Nótese que el valor óptimo del recubrimiento del espaldón a sotavento (n) es 0. Esto conlleva a un ahorro en la construcción. Pero desde un punto de vista práctico, no es una solución razonable, y no se daría si se consideraran otros modos de fallo como el provocado por motivos geotécnicos. Pero nos ilustra que el método nos avisa de la presencia de variables sin restringir.

Las sensibilidades para el ejemplo ilustrativo se muestran en las Tablas 11.13 y 11.14. La tabla 11.13 da las sensibilidades del coste asociadas al diseño clásico. Permite saber cuánto cambia el coste de construcción del dique en talud cuando se produce un pequeño cambio en alguno de los parámetros. Esta información es extremadamente útil durante el proceso constructivo para controlar el coste, y para analizar la influencia de los factores de seguridad requeridos por los códigos en el mismo. Por ejemplo, un cambio de una unidad en D_{WL} conlleva un aumento aproximado en el coste total de 14174.2 euros (véase la correspondiente entrada en la tabla 11.13). Similarmente, un aumento de una unidad en el peso específico del hormigón γ_c supondría una disminución aproximada de 22562.5 euros.

La tabla 11.14 muestra las sensibilidades de los índices de fiabilidad. Es muy útil conocer la influencia de los parámetros, como desviaciones típicas, medias, coeficientes de variación, etc, en la probabilidad de fallo de cada modo. Con esta tabla el diseñador puede analizar cómo la calidad de los materiales, o la precisión en la construcción, influyen en la fiabilidad de la estructura. Por ejemplo, un incremento de 1 unidad en el coeficiente de fricción μ_{cr} , sin cambiar los demás parámetros, supone un aumento aproximado del índice de fiabilidad β_s de 0.385, con la disminución de probabilidad de fallo que esto supone. Y un incremento del parámetro μ_{Au} de 1 unidad, disminuye el índice de fiabilidad β_o en 4.145.

Similarmente, también se podrían dar las sensibilidades de los índices de fiabilidad con respecto a las variables de diseño.

Table 11.12: Ilustración del proceso iterativo. Los valores de diseño finales están en negrita, mientras que los coeficientes de seguridad o índices de fiabilidad activos se encuentran subrayados.

Variable	Unidades	ITERACIONES			
		1	2	3	14 (fin)
Coste	<i>euros</i> × 10 ³	272.2	281.9	291.1	299.7
<i>a</i>	<i>m</i>	6.11	6.18	6.16	6.16
<i>b</i>	<i>m</i>	2.18	2.20	2.19	2.19
<i>c</i>	<i>m</i>	16.59	16.50	17.27	17.47
<i>r</i>	<i>m</i>	2.00	2.00	2.00	2.00
<i>t</i>	<i>m</i>	1.00	1.00	1.00	1.00
<i>e</i>	<i>m</i>	10.00	14.51	16.13	23.54
<i>p</i>	<i>m</i>	2.00	2.00	2.00	2.00
<i>s</i>	<i>m</i>	6.61	12.00	12.00	12.00
<i>q</i>	<i>m</i>	7.13	8.00	8.00	8.00
<i>d</i>	<i>m</i>	6.11	6.18	6.16	6.16
<i>f</i>	<i>m</i>	2.18	2.20	2.19	2.19
<i>n</i>	<i>m</i>	0.00	0.00	0.00	0.00
β	<i>m</i>	0.46	0.46	0.46	0.46
α	<i>m</i>	0.59	0.59	0.59	0.59
F_c	<i>m</i>	12.50	18.70	19.46	19.67
F_o	–	1.21	1.81	1.88	1.90
F_s	–	<u>1.50</u>	3.92	7.41	12.58
F_a	–	<u>1.50</u>	1.55	1.53	1.54
β_o	–	2.70	2.88	3.23	3.32
β_s	–	1.87	1.33	2.27	3.08
β_a	–	2.92	3.13	3.07	3.08

Table 11.13: Sensibilidades del coste $\frac{\partial C}{\partial x}$ con respecto a los datos ($\boldsymbol{\eta}$) del ejemplo del dique en talud.

$\partial C / \partial x$						
x	D_{WL}	γ_c	γ_s	γ_w	F_o	F_s
	14174.2	-22562.5	-47.3	51848.6	0.000	1599.5
x	F_a	c_{al}	c_{ul}	c_{co}	c_c	H_s
	62258.2	335.9	226.0	1070.2	298.5	40391.0

Table 11.14: Sensibilidades de los índices de fiabilidad $\frac{\partial \beta_i}{\partial x}$, con respecto a los datos $x = (\boldsymbol{\eta}, \boldsymbol{\kappa})$ del ejemplo del dique en talud. Valores positivos indican el aumento del índice cuando el parámetro correspondiente aumenta en una unidad.

Dato x	$\frac{\partial \beta_o}{\partial x}$	$\frac{\partial \beta_s}{\partial x}$	$\frac{\partial \beta_a}{\partial x}$
γ_c	0.000	0.021	2.198
γ_s	0.000	0.000	0.000
γ_w	0.000	-0.048	-6.204
μ_{A_u}	-4.145	-0.708	0.000
μ_{B_u}	2.239	0.431	0.000
$\mu_{\mu_{cr}}$	0.000	0.385	0.000
μ_{A_r}	0.000	0.000	-308.402
μ_{B_r}	0.000	0.000	-0.158
μ_{c_f}	0.000	-0.336	0.000
μ_{h_1}	-0.028	-0.167	-0.258
μ_{h_2}	-0.028	-0.167	-0.258
σ_{A_u}	-32.994	-3.343	0.000
σ_{B_u}	-6.032	-0.783	0.000
$\sigma_{\mu_{cr}}$	0.000	-0.091	0.000
v_{A_r}	0.000	0.000	-47.323
v_{B_r}	0.000	0.000	-23.132
v_{C_f}	0.000	-0.587	0.000
σ_{h_1}	-0.019	-0.162	-0.245
σ_{h_2}	-0.001	-0.104	-0.114

Table 11.15: Simulación de Monte Carlo para estimar las probabilidades de fallo para cada combinación posible de modos de fallo.

Tipo de fallo	Probabilidad
\emptyset	0.99890
$\{o\}$	0.00010
$\{s\}$	0.00039
$\{a\}$	0.00044
$\{o\} \cup \{s\}$	0.00015
$\{o\} \cup \{a\}$	0.00000
$\{s\} \cup \{a\}$	0.00002
$\{o\} \cup \{s\} \cup \{a\}$	0.00000

11.3.7 Monte Carlo simulation

Para determinar la probabilidad de fallo global, entender mejor la interacción y correlación entre los modos de fallo, y estimar las probabilidades de fallo de todas las combinaciones de fallo posibles, se realiza una simulación de Monte Carlo de 10^9 réplicas. La tabla 11.3.7 muestra los fallos de todas las combinaciones posibles, y la probabilidad de fallo global que es de 0.0011. Las probabilidades asociadas a al rebase, deslizamiento y estabilidad de las piezas del manto son 0.00025, 0.00056 y 0.00046, respectivamente.

Obsérvese en primer lugar, que las probabilidades de deslizamiento y extracción de piezas del manto son menores que las determinadas mediante el método FORM que coinciden con las máximas permitidas de 0.001, pero son del mismo orden de magnitud, con lo que puede considerarse un diseño adecuado a este respecto. Sin embargo, la probabilidad de rebase es mucho menor que la máxima exigida de 0.01. Por qué entonces, no trata de reducir la altura de ataque del muro del espaldón para que sea más barato. La respuesta es sencilla, y se debe a las simplificaciones del modelo. Dado que se ha de aumentar el peso del espaldón para reducir la probabilidad de deslizamiento, y como no hemos impuesto restricciones relativas al hundimiento de la cimentación, al modelo le resulta más barato aumentar el peso del espaldón aumentando el muro de ataque en vez de aumentar gradualmente todo el espaldón.

11.3.8 Notation

A_c : cota de la berma.	F_c : francobordo.
A_r : constante adimensional para la definición de la función de estabilidad.	F_h : fuerza horizontal actuante sobre el espaldón debido a la presión del oleaje.
A_u : coeficiente adimensional que depende del tipo de piezas del manto principal.	F_o : coeficiente de seguridad al rebase.
a : espesor del manto principal.	F_o^0 : límite inferior del coeficiente de seguridad al rebase.
B_r : constante adimensional para la definición de la función de estabilidad.	F_s : coeficiente de seguridad frente al deslizamiento del espaldón.
B_u : coeficiente adimensional que depende del tipo de piezas del manto principal.	F_s^0 : límite inferior del coeficiente de seguridad frente al deslizamiento del espaldón.
b : espesor del manto secundario.	F_v : subpresión bajo el espaldón.
C_{al} : coste del manto principal por unidad de volumen.	f : anchura de la berma en el lado de sotavento.
C_c : coste del hormigón por unidad de volumen.	g : aceleración de la gravedad.
C_{co} : coste del núcleo por unidad de volumen.	H : altura de ola.
C_f : coeficiente experimental aleatorio.	H_D : altura de ola máxima.
C_{ul} : coste del manto secundario por unidad de volumen.	H_s : altura de ola significativa.
c : altura del núcleo.	h_1 : marea astronómica.
D : vida útil.	h_2 : marea meteorológica.
D_{WL} : nivel del mar de diseño.	I_r : número de Iribarren.
\mathbf{d} : variables de diseño.	I_{r_0} : constante adimensional.
d : anchura de la berma en el lado de barlovento.	L : longitud de onda.
e : anchura del espaldón.	ℓ : longitud de lado del cubo del manto principal.
F_a : coeficiente de seguridad a la estabilidad del manto principal.	α_e : longitud de lado equivalente del manto secundario.
F_a^0 : límite inferior del coeficiente de seguridad a la estabilidad del manto principal.	n : recubrimiento del espaldón a sotavento.
	p_i^0 : límite superior de la probabilidad de fallo del modo i .
	P_{S_0} : presión del agua sobre el espaldón inducida por el oleaje.

p : espesor en la coronación del espaldón.	β_a : índice de fiabilidad respecto a la estabilidad del manto principal.
q : altura de la coronación del espaldón.	β_o : índice de fiabilidad respecto al rebase.
R : constante adimensional que depende de γ_c y γ_w .	β_s : índice de fiabilidad respecto al deslizamiento del espaldón.
R_u : run-up.	β_i^k : límite inferior del índice de fiabilidad del modo i para la iteración k .
r : longitud de la lengua del dique.	β_a^0 : límite inferior del índice de fiabilidad respecto a la estabilidad del manto principal.
S_0 : altura de ola equivalente debido al run-up.	β_o^0 : límite inferior del índice de fiabilidad respecto al rebase.
s : altura del bloque principal del espaldón.	β_s^0 : límite inferior del índice de fiabilidad respecto al deslizamiento del espaldón.
T : periodo de la ola.	γ_c : peso específico del hormigón.
T_p : periodo de pico de la ola.	γ_s : peso específico de la escollera.
t : altura de la lengua del dique.	γ_w : peso específico del agua.
tr : carrera de marea.	η : conjunto de parámetros del diseño clásico.
U : variable aleatoria uniforme.	ϕ : conjunto de variables básicas usadas en los modelos probabilistas.
V_c : volumen total de hormigón.	ψ : variables auxiliares (no básicas) cuyos valores se obtienen de las básicas, de los parámetros y de las de diseño (\mathbf{d}, η, ϕ).
V_1 : volumen total del manto principal.	κ : conjunto de parámetros que definen la variabilidad y dependencia de las variables aleatorias.
V_2 : volumen total del manto secundario.	$\tilde{\eta}$: valores medios de η .
V_3 : volumen total del núcleo.	$\tilde{\phi}$: valores medios de ϕ .
v_{A_r} : coeficiente de variación de A_r .	$\tilde{\psi}$: valores medios de ψ .
v_{B_r} : coeficiente de variación de B_r .	λ : constante adimensional.
v_{C_f} : coeficiente de variación de C_f .	μ_{A_r} : valor medio de A_r .
W : peso de las piezas del manto principal.	μ_{B_r} : valor medio de B_r .
W_f : peso sumergido del espaldón.	
W_1 : peso del espaldón.	
z_p : cero del puerto.	
α : constante adimensional.	
α_ℓ : ángulo del talud a sotavento.	
α_s : ángulo del talud a sotavento.	

μ_{A_u} : valor medio de A_u .

μ_{B_u} : valor medio de B_u .

μ_{C_f} : valor medio de C_f .

μ_c : coeficiente de rozamiento.

μ_{μ_c} : valor medio de μ_c .

σ_{A_u} : desviación típica de A_u .

σ_{B_u} : desviación típica de B_u .

σ_{μ_c} : desviación típica de μ_c .

Φ_e : función de estabilidad.

11.4 Slope Stability

En esta sección se describe con detalle la aplicación de las técnicas de optimización al cálculo de la estabilidad de taludes. El estudio de la estabilidad de taludes ha sido un tema muy tratado por investigadores geotécnicos en sus dos vertientes: (a) aproximación clásica, basada en coeficientes de seguridad, y (b) la moderna, que utiliza probabilidades de fallo (véase Wu y Kraft [178], Cornell [49], Li y Lumb [108], Alonso [6], Tang, Yüemen y Ang [163], Venmarcke [171], Wolff [177], Barabosa, Morris y Sarma [13], Dai, Fredlund y Stolte [50] y Christian, Ladd y Baecher [45]). La aproximación clásica se cuestiona porque no da una idea de lo lejos que se está del fallo, y se demanda un diseño basado en técnicas probabilistas (véase Ditlevsen y Madsen [64], Wirsching y Wu [175], y Wu, Burnside y Cruse [179]). Sin embargo, la aproximación moderna también se critica por la sensibilidad a las hipótesis estadísticas, especialmente al comportamiento en las colas (véase Galambos [83], Castillo [28], y Castillo et al. [40, 39]).

Un análisis clásico del problema consiste en determinar los coeficientes de seguridad asociados a una colección de líneas seleccionadas por el ingeniero, y elegir aquella que tiene el mínimo coeficiente que será tomado como el coeficiente de seguridad del talud. El conjunto de líneas de deslizamiento es normalmente una familia paramétrica. Sin embargo, dado que este conjunto no cubre todas las líneas posibles, este procedimiento no garantiza obtener el mínimo. Esto en un principio no es problema si el coeficiente está cerca del mínimo, pero puede serlo si está lejos.

Alternativamente, puede usarse una aproximación no paramétrica. La más importante es la basada en el cálculo de variaciones, que se ha aplicado a la mecánica de suelos en el pasado (véase, por ejemplo, Chen y Giger [43], Chen y Snitbhan [44], Baker y Garber [11, 12, 84], Revilla y Castillo [142], y Castillo y Luceño [35, 36]). Esta técnica es una generalización del problema de máximos y mínimos, en el que se busca el valor máximo o mínimo de funcionales, en vez de funciones. Un funcional es una aplicación del conjunto de funciones $y(x)$ que representa la línea de deslizamiento, en un número real que representa el coeficiente de seguridad del talud. Por lo tanto, el funcional asocia a cada línea de deslizamiento su correspondiente coeficiente de seguridad. En la mayoría de las aproximaciones, el funcional se define como el cociente entre dos integrales, que representan el ratio entre fuerzas y momentos estabilizadores y desestabilizadores.

Pero esto no es todo, como puntualizó Cornell [49], la fiabilidad de un sistema es la de todas sus superficies potencialmente deslizantes, y por tanto la probabilidad de fallo de un sistema será mayor que la probabilidad para una simple línea. La diferencia entre ellas dependerá de la correlación entre las probabilidades de fallo de las diferentes líneas.

La aproximación mediante el cálculo de variaciones tiene claras ventajas sobre otras metodologías, porque: (a) en la aproximación clásica, se garantiza la obtención de la línea pésima, es decir, no hay ninguna línea con menor coeficiente de seguridad, y (b) en la aproximación moderna, se garantiza la obtención de la línea pésima con menor índice de fiabilidad, es decir, con mayor probabilidad de fallo.

Finalmente, una vez que se ha seleccionado la línea pésima, es interesante efectuar un estudio de sensibilidad para saber cómo y cuánto influyen los parámetros en la seguridad.

En esta sección se presentan tres contribuciones: (a) una aproximación variacional al problema de la estabilidad de taludes, que evita la necesidad de utilizar familias paramétricas

particulares, (b) un nuevo método que combina coeficientes de seguridad y probabilidades de fallo, y (c) herramientas potentes para el estudio de sensibilidad.

El capítulo se estructura de la siguiente manera. En la subsección 11.4.1 se introduce el método del cálculo variacional. En las subsecciones 11.4.2 y 11.4.3 se describe como pueden resolver diferentes problemas asociados a la estabilidad de taludes ingenieros clásicos y probabilistas mediante el cálculo variacional. En la subsección 11.4.4 se presenta el método combinado. La subsección 11.4.5 tratan problemas de sensibilidad. Y finalmente, en la subsección 11.4.6 se recogen varias conclusiones relativas a la metodología.

11.4.1 Calculus of variation approach

En una publicación reciente de la revista *Structural Safety*, Malkawi et al. [118] realizaron un análisis muy interesante de la fiabilidad de la estabilidad de taludes mediante técnicas de simulación de Monte Carlo, y métodos de primer (FORM) y segundo orden (SORM) con cuatro métodos diferentes (ordinario y simplificado de Bishop, Janbu y Spencer) para el análisis de la estabilidad de taludes, y llegaron a la conclusión que con las hipótesis estadísticas con las que trabajaban no había diferencias significativas en los resultados obtenidos con cada uno de los métodos mencionados. Este hecho, justifica la selección del método de Janbu en nuestro análisis, aunque no sería complicado implementar otro método de estabilidad.

Revilla y Castillo [142], Castillo y Revilla [37], o Castillo y Luceño [32, 33, 34], basándose en el método de Janbu (véase Janbu [98]), propusieron el siguiente funcional:

$$F = \frac{\int_{x_1}^{x_{n+1}} \left[\frac{c}{\gamma} + (\bar{y}(x) - y(x)) \tan \phi \right] (1 + y'^2(x))}{1 + \frac{y'(x) \tan \phi}{F}} dx}{\int_{x_1}^{x_{n+1}} (\bar{y}(x) - y(x)) y'(x) dx} \quad (11.173)$$

donde F es el coeficiente de seguridad, $\bar{y}(x)$ es el perfil del talud (ordenada en el punto x), c es la cohesión del terreno, ϕ es el ángulo de fricción interna, γ es el peso específico del terreno y x_1 y x_{n+1} son las abscisas del principio y el final de la línea de rotura o deslizamiento.

Un aspecto muy importante en el que el ingeniero ha de ser muy cuidadoso es la selección del funcional asociado a la estabilidad. De hecho, Castillo y A. Luceño [33] demostraron que ciertos funcionales propuestos en el pasado no están acotados, y no son válidos para el diseño. En concreto, probaron que otros funcionales, como el (11.173), están acotados, y consecuentemente, son válidos para el diseño.

Nótese que (11.173), para un talud predefinido $\bar{y}(x)$, involucra la utilización de cinco variables importantes ϕ, c, γ, H y F , es decir, se está tratando con un espacio de 5 dimensiones. Por supuesto se puede trabajar con estas variables, pero se complica las cosas innecesariamente y no permite comprender en su totalidad la estructura de los problemas de estabilidad.

De hecho, utilizando el análisis dimensional, mediante el teorema de II, se puede comprobar que la expresión (11.173) puede reescribirse en función de tres variables adimensionales

$$\{F, N = \frac{c}{\gamma H}, \psi = \tan \phi\} \quad (11.174)$$

como

$$F = \frac{\int_{u_1}^{u_{n+1}} \frac{[N + (\bar{z}(u) - z(u))\psi] (1 + z'^2(u))}{1 + \frac{\psi z'(u)}{F}} du}{\int_{u_1}^{u_{n+1}} (\bar{z}(u) - z(u)) z'(u) du} \quad (11.175)$$

donde

$$u = \frac{x}{H} \quad (11.176)$$

es la variable adimensional asociada a la coordenada x , y

$$z(u) = \frac{y(uH)}{H} \quad (11.177)$$

$$\bar{z}(u) = \frac{\bar{y}(uH)}{H} \quad (11.178)$$

son el perfil del talud y la línea de deslizamiento adimensionales, respectivamente.

Esto nos permite estudiar el problema en un espacio tridimensional, revelándonos que, mientras los valores adimensionales de N y ψ en la ecuación (11.174) permanezcan constantes, tanto el factor de seguridad, como la línea de deslizamiento y el perfil adimensionales permanecen también constantes.

Por ejemplo, los coeficientes de seguridad asociados a dos taludes con $\psi = 0.6, c = 20kN/m^2, \gamma = 21kN/m^3$ y $H = 10m$, y $\psi = 0.6, c = 2kN/m^2, \gamma = 21kN/m^3$ y $H = 1m$ son idénticos.

La expresión (11.175) muestra que dados $\bar{z}(u)$ y $z(u)$, hay una relación entre F, N y ψ :

$$F = H(N, \psi) \quad (11.179)$$

Sin embargo, se puede ir más lejos, si se da uno cuenta que en la expresión (11.175), N, ψ y F pueden multiplicarse por una constante arbitraria k , y se sigue manteniendo la misma relación, es decir,

$$kF = kH(N, \psi) = H(kN, k\psi) \quad (11.180)$$

en otras palabras, satisface una ecuación funcional, conocida como ecuación funcional de las funciones homogéneas, cuya solución general puede escribirse como ² (véase Aczél [4], Castillo y Ruiz [38]):

$$\frac{F}{\psi} = h\left(\frac{N}{\psi}\right) \Leftrightarrow F^* = h(N^*) \quad (11.181)$$

²Se supone que la función $H(\cdot, \cdot)$ realmente depende de las variables N y ψ .

Esto permite expresar nuestro problema en términos de dos variables adimensionales $F^* = F/\psi$ y $N^* = N/\psi$. De hecho, (11.175) puede escribirse como

$$1 = \frac{\int_{u_1}^{u_{n+1}} \frac{[N^* + (\bar{z}(u) - z(u))](1 + z'^2(u))}{F^* + z'(u)} du}{\int_{u_1}^{u_{n+1}} (\bar{z}(u) - z(u)) z'(u) du} \quad (11.182)$$

Por supuesto, se podría haber comenzado la exposición del capítulo con la fórmula (11.182) en vez de (11.173), pero esta decisión haría pasar por alto la belleza y potencia del análisis dimensional, por un lado, y de las ecuaciones funcionales por otro.

Una conclusión importante es que nuestro problema ha quedado reducido a dos variables que definen la naturaleza y fiabilidad del talud. Por tanto, nuestro problema se reduce a

$$\begin{array}{l} \text{Minimizar} \quad F^* \\ u_1, u_{n+1}, z(u) \end{array}$$

sometido a (11.182) y $N^* = N_0$. Si se resuelve el problema con diferentes valores de N_0 se puede elaborar una gráfica relacionando F^* y N^* con la que se puede obtener el coeficiente de seguridad F asociado a cualquier combinación de variables dato ϕ, c, γ y H .

La implementación del cálculo de variaciones a la estabilidad de taludes consiste en minimizar, como ya se ha dicho antes, F^* con respecto a u_1, u_{n+1} y $z(u)$. Este problema puede resolverse analíticamente, sin embargo, conduce a sistemas de ecuaciones muy complicados y condiciones de transversalidad (véase Luceño [114]). La alternativa obvia consiste en una aproximación numérica, que es la opción seleccionada en esta sección.

11.4.2 Discretization of the problem

Usando las siguientes fórmulas (véase la figura 11.17):

$$\Delta u = \frac{u_{n+1} - u_1}{n} \quad (11.183)$$

$$u_i = u_1 + (i - 1)\Delta u \quad (11.184)$$

$$v_i = \frac{z_{i+1} + z_i}{2} \quad (11.185)$$

$$v'_i = \frac{z_{i+1} - z_i}{\Delta u} \quad (11.186)$$

$$\bar{v}_i = \bar{z}(u_i + \Delta u/2) \quad (11.187)$$

la ecuación (11.182) puede ser fácilmente discretizada como

$$1 = \frac{\sum_{i=1}^n \frac{[N^* + (\bar{v}_i - v_i)](1 + v_i'^2)\Delta u}{F^* + v'_i}}{\sum_{i=1}^n (\bar{v}_i - v_i)v'_i\Delta u} \quad (11.188)$$

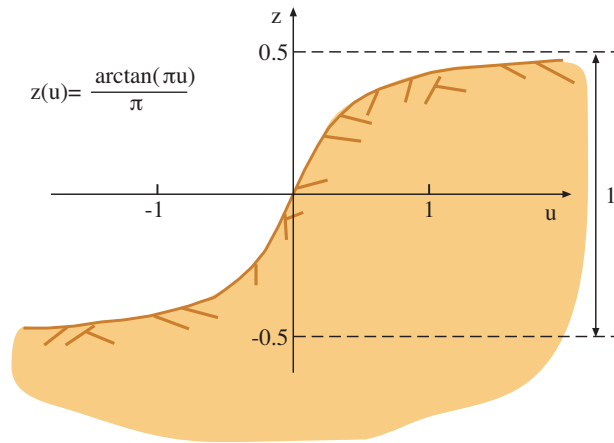


Figure 11.18: Perfil del talud seleccionado para el ejemplo ilustrativo.

11.4.3 Classical approach

En esta sección se resuelve el problema de la estabilidad mediante el método clásico basado en coeficientes de seguridad, es decir, se supone que todos los parámetros son deterministas.

Derivation of the stability relation between N^* and F^*

Como ejemplo, en esta subsección se supone un perfil de talud arcotangente (véase la figura 11.18)

$$y(x) = H \arctan(\pi x/H)/\pi; \quad -\infty < x < \infty$$

que implica

$$z(u) = \arctan(\pi u)/\pi; \quad -\infty < u < \infty.$$

Resolviendo el problema (11.189)-(11.192), para todos los valores posibles del factor adimensional N^* , se puede obtener una gráfica que determine el coeficiente de seguridad para cualquier N^* . Este gráfico se muestra en la figura 11.19, donde se muestran los valores óptimos de F^* en función de N^* ($F^* = h(N^*)$). Las líneas de deslizamiento correspondientes se muestran en la figura 11.20. Estos gráficos permiten resolver gráficamente cualquier problema por el método clásico.

Classical solution of the stability problems

Los problemas asociados a la estabilidad pueden presentarse en formas diferentes, tales como:

Problema tipo 1: Obtención del máximo valor permitido de N^* para un valor fijo de ψ y un coeficiente de seguridad F . En este caso se calcula $F^* = F/\psi$, se entra directamente en el gráfico de la figura 11.19 y se determina el valor de N^* buscado.

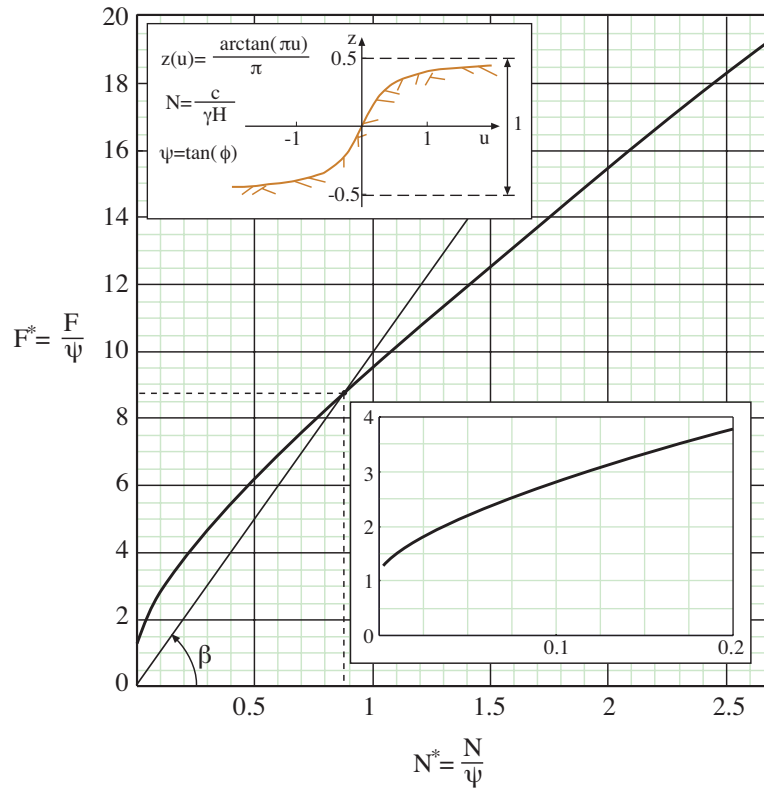


Figure 11.19: Valores óptimos de F^* en función de N^* y zona ampliada en el entorno de origen.

Una vez que se tiene N^* , se puede calcular el valor de H en función de los parámetros del terreno c, γ, ψ , como $H = \frac{c}{N^* \gamma \psi}$, o el valor de c , en función de H, γ and ψ , como $c = N^* \gamma H \psi$.

Computational Example 11.1 (Problema tipo 1). Supóngase que se desea diseñar un talud con un coeficiente de seguridad $F = 2$ y $\phi = 10^\circ \equiv 1/\psi = 1/\tan(\phi) = 5.67$. En este caso $F^* = 2 \times 5.67 = 11.34$, entrando en el gráfico de la figura 11.19 se obtiene $N^* = 1.27$.

Si $c = 200kN/m^2$ y $\gamma = 23kN/m^3$, entonces la altura del talud debe ser de $H = \frac{200 \times 5.67}{1.27 \times 23} = 38.83m$, o, si $H = 30m, \gamma = 23kN/m^3$ y $\psi = 1/5.67 = 0.176$, entonces, el valor de diseño es $c = 1.27 \times 23 \times 30 \times 0.176 = 154.5kN/m^2$. ■

Problema tipo 2: Obtención del valor ψ requerido para unos valores dados de H, γ, c y del coeficiente de seguridad F . En este caso puede parecer en un principio que se necesite un método iterativo, porque ψ aparece tanto en N^* como en F^* . Para evitar este procedimiento se puede descomponer el gráfico de la figura 11.19 en uno

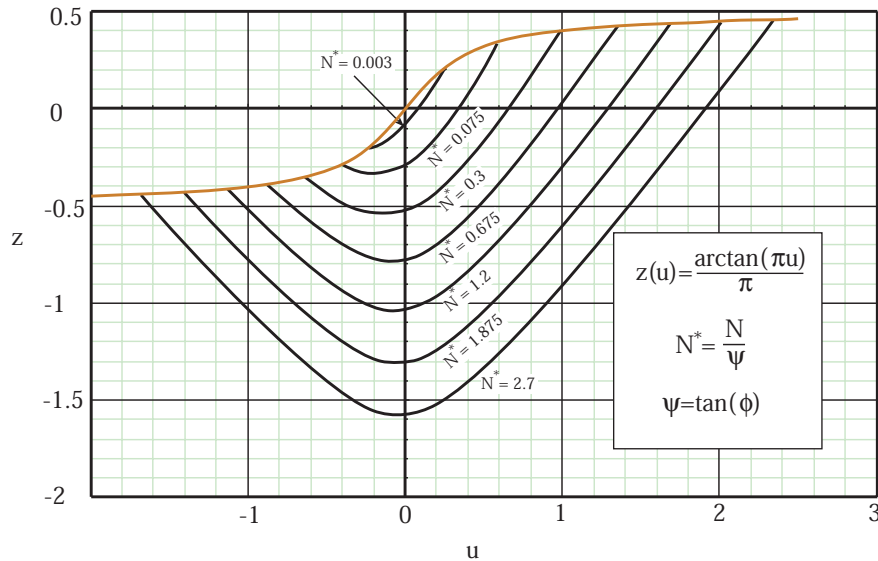


Figure 11.20: Líneas de deslizamiento críticas para diferentes valores de N^* .

equivalente (véase la figura 11.21), donde la relación F^* y N^* se transforma en muchas relaciones entre ψ , N y F , rompiendo la relación (11.181), con lo cual sólo se ha de entrar en la gráfica con N y F y obtener el valor de ψ . Aunque se puede utilizar también el gráfico inicial (figura 11.19) con la recta que parte del origen con pendiente $\beta = F^*/N^* = F/N$ (véase la figura 11.19), con lo cual se elimina el parámetro ψ . La intersección de esa recta con la curva crítica será el punto buscado, no hay más que despejar ψ de cualquiera de las variables N^* o F^* .

Computational Example 11.2 (Problema tipo 2). Supóngase que $H = 30\text{m}$, $\gamma = 23\text{kN/m}^3$, $c = 200\text{kN/m}^2$ y $F = 2$. Entrando en el gráfico de la figura 11.21, con $N = \frac{200}{23 \times 30} = 0.2899$ y $F = 2$, se obtiene $\psi = 0.08$. Se conseguiría el mismo resultado trazando la recta de pendiente $\beta = 2/0.2899$ y tomando el punto intersección de la misma con la curva crítica. ■

Problema tipo 3: Obtención del valor del coeficiente de seguridad F para unos valores de H, γ, c y ψ conocidos. En este caso se calcula $N^* = \frac{c}{H\gamma\psi}$, se entra en el gráfico de la figura 11.19, se determina F^* y se calcula el valor de $F = F^*\psi$.

Computational Example 11.3 (Problema tipo 3:). Si $H = 30\text{m}$, $\gamma = 23\text{kN/m}^3$, $c = 200\text{kN/m}^2$ y $\psi = 0.176$, se calcula $N^* = \frac{200}{30 \times 23 \times 0.176} = 1.647$, entrando en el gráfico correspondiente, se obtiene $F^* = 13.35$ y finalmente se obtiene el valor de F , $F = 13.35 \times 0.176 = 2.355$. ■

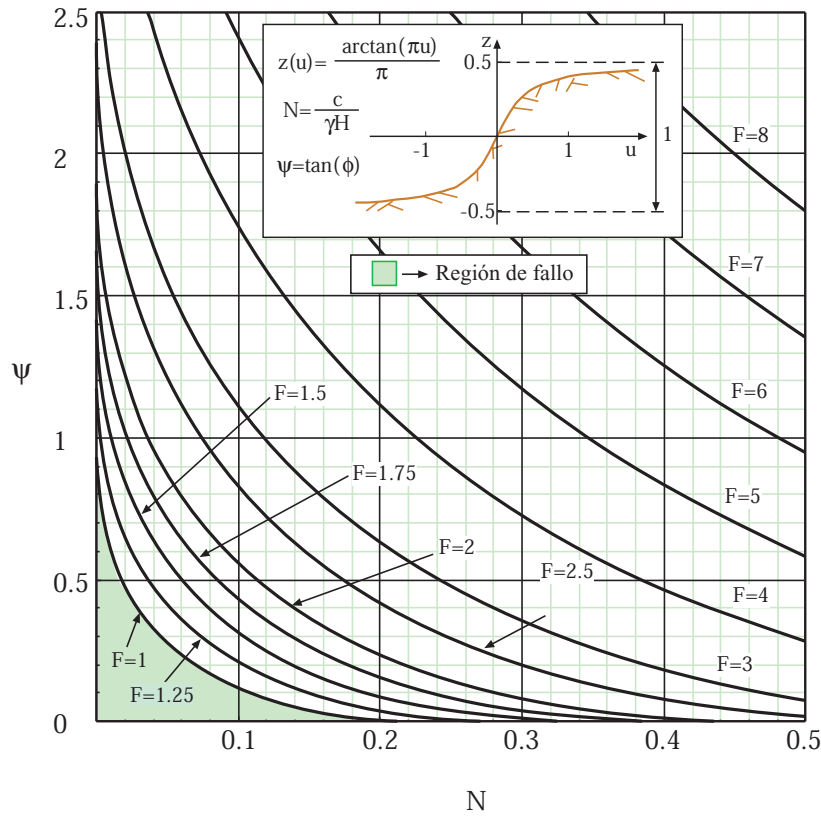


Figure 11.21: Valores óptimos de F en función N .

11.4.4 Modern approach

Contrariamente a las hipótesis de la subsección anterior, se supone que todas las variables son aleatorias. Dado que el problema de la probabilidad de fallo depende sólo de dos variables $F^+ = 1/\psi$ y N^* , ó de ψ y N se puede trabajar sólo con ellas. Nótese que dado que en punto de fallo $F = 1$, F^+ es distinto de F^* . Posiblemente uno esté tentado en trabajar con c, γ, ϕ y H , y hacer ciertas hipótesis sobre la función de densidad conjunta. Sin embargo, dado que el problema depende sólo de dos variables, el resultado final dependerá sólo de la función de densidad conjunta de las variables (F^+, N^*) ó (ψ, N) , y sin pérdida de generalidad, se puede trabajar sólo con ellas.

Todos los trabajos existentes en la actualidad no tienen en cuenta este hecho, y tratan de obtener distribuciones que se adecúen a los valores de c, γ, ϕ cuando su influencia en la estabilidad del sistema depende de N y ψ .

Calculating the reliability index of a given design

Para simplificar nuestro tratamiento se supone que F^+ y N^* son variables normales, pero dependientes, porque el parámetro ψ aparece en F^+ y N^* . De forma más precisa se con-

sidera:

$$F^+ \sim N(\mu_{F^+}, \sigma_{F^+}^2); \quad N^*|F^+ \sim N(\mu_{N^*} + \beta_0(F^+ - \mu_{F^+}), \sigma_{N^*|F^+}^2)$$

donde μ_{F^+} y μ_{N^*} son las medias de F^+ y N^* , respectivamente, $\sigma_{F^+}^2$ es la varianza de F^+ , $N^*|F^+$ es N^* condicionada sobre F^+ , $\sigma_{N^*|F^+}^2$ es la varianza de N^* dada F^+ , y β_0 es el coeficiente de regresión de N^* sobre F^+ . En lo que sigue, se supone que $\mu_{F^+} = F_0^+$ y $\mu_{N^*} = N_0^*$, es decir, que son iguales a los valores de diseño clásicos F_0^+ y N_0^* , respectivamente.

Una vez más, para calcular la probabilidad de fallo se necesita transformar el conjunto inicial de variables en un conjunto de variables normales estándar independientes usando la transformación de Rosenblatt (véase Rosenblatt [146], Ditlevsen y Madsen [64] y Madsen, Krenk y Lind [116]):

$$\begin{aligned} Z_{F^+} &= \frac{F^+ - \mu_{F^+}}{\sigma_{F^+}} = \frac{F^+ - \mu_{F^+}}{v_{F^+}\mu_{F^+}} \\ Z_{N^*} &= \frac{N^* - \mu_{N^*} - \beta_0(F^+ - \mu_{F^+})}{\sigma_{N^*|F^+}} = \frac{N^* - \mu_{N^*} - \beta_0(F^+ - \mu_{F^+})}{v_{N^*|F^+}\mu_{N^*}} \end{aligned} \quad (11.193)$$

donde v_{F^+} es el coeficiente de variación de F^+ , y $v_{N^*|F^+}$ es el coeficiente de variación de N^* , dado F^+ .

Así, el cálculo de índice de fiabilidad β se reduce a:

$$\text{Minimizar } \beta = \sqrt{Z_{F^+}^2 + Z_{N^*}^2} \quad (11.194)$$

sujeto a

$$\begin{aligned} Z_{F^+} &= \frac{F^+ - \mu_{F^+}}{\sigma_{F^+}} = \frac{F^+ - \mu_{F^+}}{v_{F^+}\mu_{F^+}} \\ Z_{N^*} &= \frac{N^* - \mu_{N^*} - \beta_0(F^+ - \mu_{F^+})}{\sigma_{N^*|F^+}} = \frac{N^* - \mu_{N^*} - \beta_0(F^+ - \mu_{F^+})}{v_{N^*|F^+}\mu_{N^*}} \\ F^+ &= h(N^*) \end{aligned} \quad (11.195)$$

Dado que se resuelve el problema de forma adimensional, se están resolviendo infinitos problemas a la vez. Para aplicar el método a un caso concreto basta transformar los valores óptimos adimensionales a los correspondientes dimensionales.

A continuación se van a presentar dos métodos diferentes: aproximaciones 1 y 2.

Approach 1

Desafortunadamente, no existe una expresión explícita que relacione F^+ en términos de N^* ($F^+ = h(N^*)$). Así, una forma de abordar el problema es obtener (usando mínimos cuadrados, etc.) una expresión analítica aproximada $F^* = \tilde{h}(N^*)$, por ejemplo un polinomio de grado m . El siguiente polinomio de grado 8 proporciona una buena aproximación

$$\begin{aligned} F^+ &= 1.35 + 18.19N^* - 46.6N^{*2} + 111.26N^{*3} - 155.56N^{*4} \\ &\quad + 127.52N^{*5} - 60.23N^{*6} + 15.14N^{*7} - 1.565N^{*8} \end{aligned}$$

Approach 2

Dado que la región de fallo se determina mediante un procedimiento de optimización (11.189)-(11.192), el problema ha de resolverse mediante un procedimiento iterativo, como el dado en el siguiente algoritmo basado en el método de la bisección:

Algorithm 11.1 (Cálculo del índice de fiabilidad β).

Entrada: Parámetros estadísticos del problema $\mu_{F^+}, \mu_{N^*}, v_{F^+}, v_{N^*|F^+}$ y β_0 , y una tolerancia ϵ para controlar la convergencia del proceso.

Salida: Valores óptimos de N^* y F^+ y el índice de fiabilidad β .

Etapas 1. Selección de dos valores iniciales, uno muy pequeño $N_1^* = 0.001$ y otro muy alto $N_2^* = 5$, y calcular sus respectivos valores de fallo F_1^+ y F_2^+ , resolviendo el problema (11.189)-(11.192) para cada valor. Esto es equivalente a entrar en el gráfico de la figura 11.19.

Etapas 2. Cálculo de los valores β_1 y β_2 asociados con los pares (N_1^*, F_1^+) y (N_2^*, F_2^+) , respectivamente, utilizando las expresiones (11.193) y (11.194).

Etapas 3. Si $|\beta_1 - \beta_2| < \epsilon$, ir a la etapa 7. Si no, hacer $N_3 = 0.51N_1 + 0.49N_2$ y $N_4 = 0.49N_1 + 0.51N_2$.

Etapas 4. Cálculo de sus respectivos valores de fallo F_3^+ and F_4^+ , resolviendo el problema de optimización (11.189)-(11.192), o entrando en el gráfico de la figura 11.19.

Etapas 5. Cálculo de los índices de fiabilidad β_3 y β_4 asociados a los puntos (N_3^*, F_3^+) y (N_4^*, F_4^+) , respectivamente.

Etapas 6. Si $\beta_4 < \beta_3$ hacer $N_1^* = N_3$; $F_1^+ = F_3^+$ y $\beta_1 = \beta_3$. Si no, $N_2^* = N_4$; $F_2^+ = F_4^+$ y $\beta_2 = \beta_4$, e ir a la etapa 3.

Etapas 7. Devolver $N^* = (N_1^* + N_2^*)/2$, $F^+ = (F_1^+ + F_2^+)/2$ y $\beta = (\beta_1 + \beta_2)/2$. ■

Modern solution of the stability problems

Dado que en esta aproximación se requiere que el talud diseñado tenga un índice de fiabilidad mayor que un valor mínimo dado β^0 , como es posible que el diseño seleccionado no cumpla esta restricción, el proceso debe repetirse hasta que se cumple la restricción.

El problema puede plantearse de dos formas distintas como:

Problema tipo 1: Obtención del valor μ_{N^*} para unos valores dados de $\mu_{F^+}, v_{F^+}, v_{N^*|F^+}$, y límite inferior del índice de fiabilidad β^0 . El problema puede plantearse como:

$$\text{Minimizar } \beta = \sqrt{Z_{F^+}^2 + Z_{N^*}^2} \quad (11.196)$$

$Z_{F^+}, Z_{N^*}, \mu_{N^*}$

Table 11.16: Ilustración del procedimiento iterativo para el ejemplo computacional 11.4.

Iteración k	$\mu_{N^*}^{(k)}$	$\beta^{(k)}$
1	2.000	2.931
2	1.534	2.126
3	1.471	2.012
4	1.465	2.001
5	1.465	2.000
6	1.465	2.000

sujeto a

$$\begin{aligned}
 Z_{F^+} &= \frac{F^+ - \mu_{F^+}}{\sigma_{F^+}} = \frac{F^+ - \mu_{F^+}}{v_{F^+} \mu_{F^+}} \\
 Z_{N^*} &= \frac{N^* - \mu_{N^*} - \beta_0(F^+ - \mu_{F^+})}{\sigma_{N^*|F^+}} = \frac{N^* - \mu_{N^*} - \beta_0(F^+ - \mu_{F^+})}{v_{N^*|F^+} \mu_{N^*}} \quad (11.197) \\
 F^+ &= h(N^*) \\
 \beta &> \beta^0
 \end{aligned}$$

y resuelto mediante la expresión aproximada $\tilde{h}(N^*)$ for $h(N^*)$. Si no se dispone de la aproximación, se procede de la siguiente manera: se comienza con un valor inicial μ_{N^*} , se resuelve el problema (11.194)-(11.195), y se obtiene β . Si el valor resultante de β es menor que el valor β^0 requerido, se aumenta el valor de μ_{N^*} y se repite el proceso hasta que se consigue que $\beta = \beta^0$. Como aplicaciones de este caso se puede determinar H en términos de μ_{N^*}, γ, ψ y c , o c en términos de μ_{N^*}, γ, ψ y H .

Computational Example 11.4 (Problema tipo 1). Asumiendo que se tiene $\mu_{F^+} = 5.67$, $v_{F^+} = 0.1$, $v_{N^*|F^+} = 0.1$, $\beta_0 = 1$, y un límite inferior del índice de fiabilidad $\beta^0 = 2$. Entonces, la solución de (11.196)-(11.197) es

$$F^+ = 4.584; \quad \mu_{N^*} = 1.464; \quad N^* = 0.294; \quad \beta = 2.$$

Si, alternativamente, se usa el procedimiento descrito anteriormente, con un valor inicial de $\mu_{N^*} = 2$, se resuelve el problema (11.194)-(11.195), obteniendo $\beta = 2.931$. Como el resultado obtenido para β es mayor que el requerido $\beta^0 = 2$, se disminuye el valor de μ_{N^*} mediante la fórmula:

$$\mu_{N^*}^{(k+1)} = \mu_{N^*}^{(k)} + \rho(\beta^0 - \beta^{(k)}) = 2 + 0.5(2 - 2.931) = 1.534.$$

Después se repite el proceso hasta que $\beta = \beta^0$, como se ilustra en la tabla 11.16. ■

Problema tipo 2: Obtención del μ_{F^+} requerido para un valor dado de β^0 y parámetros $\mu_N, v_{F^+}, v_{N^*|F^+}, \beta_0$.

Este problema puede plantearse como:

$$\begin{array}{l} \text{Minimizar} \\ Z_{F^+}, Z_{N^*}, \mu_{F^+} \end{array} \quad \beta = \sqrt{Z_{F^+}^2 + Z_{N^*}^2} \quad (11.198)$$

sometido a

$$\begin{aligned} Z_{F^+} &= \frac{F^+ - \mu_{F^+}}{\sigma_{F^+}} = \frac{F^+ - \mu_{F^+}}{v_{F^+} \mu_{F^+}} \\ Z_{N^*} &= \frac{N^* - \mu_{N^*} - \beta_0(F^+ - \mu_{F^+})}{\sigma_{N^*|F^+}} = \frac{N^* - \mu_{N^*} - \beta_0(F^+ - \mu_{F^+})}{v_{N^*|F^+} \mu_{N^*}} \\ F^+ &= h(N^*) \\ \mu_{N^*} &= \mu_N \mu_{F^+} \\ \beta &> \beta^0 \end{aligned} \quad (11.199)$$

y resuelto usando la aproximación $\tilde{h}(N^*)$ for $h(N^*)$. Si no se dispone de la expresión aproximada se procede de la forma siguiente: se selecciona un valor inicial de μ_{F^+} , se calcula $\mu_{N^*} = \mu_N \mu_{F^+}$, se resuelve el problema (11.194)-(11.195), y se obtienen β . Si el valor obtenido de β es menor que el valor mínimo requerido β^0 , se disminuye el valor de μ_{F^+} y se repite el proceso hasta que se consigue que $\beta = \beta^0$. El valor final de μ_{F^+} es el valor buscado.

Computational Example 11.5 (Problema tipo 1). Asumiendo que $\mu_N = 0.28$, $v_{F^+} = 0.1$, $v_{N^*|F^+} = 0.1$, $\beta_0 = 1$, y un límite inferior del índice de fiabilidad $\beta^0 = 2$. Entonces, la solución de (11.198)-(11.199) es

$$F^+ = 7.141; \quad N^* = 0.643; \quad \mu_{N^*} = 2.459; \quad \beta = 2.$$

■

Problema tipo 3: Obtención del índice de fiabilidad β para los parámetros estadísticos $\mu_{N^*}, \mu_{F^+}, v_{F^+}, v_{N^*|F^+}, \beta_0, \mu_N$. En este caso, basta con resolver el problema (11.194)-(11.195), y obtener β .

Computational Example 11.6 (Problema tipo 3). Asumiendo que $\mu_{N^*} = 2$, $\mu_{F^+} = 5.67$, $v_{F^+} = 0.1$, $v_{N^*|F^+} = 0.1$ y $\beta_0 = 1$, entonces, resolviendo el problema (11.194)-(11.195), se obtienen $\beta = 2.931$. ■

11.4.5 The probability-safety factor method

En esta subsección se presenta el método que combina la aproximación clásica y la moderna, particularizado para el caso de estabilidad de taludes. La idea es diseñar un talud (selección

del valor de μ_{N^*}) tal que su coeficiente de seguridad F y su índice de fiabilidad β satisfagan $F > F^0$ y $\beta > \beta^0$, es decir:

$$\begin{aligned} & \text{Minimizar} & \beta &= \sqrt{Z_{F^+}^2 + Z_{N^*}^2} \\ & Z_{F^+}, Z_{N^*}, \mu_{N^*}, F \end{aligned} \quad (11.200)$$

sujeto a

$$\begin{aligned} Z_{F^+} &= \frac{F^+ - \mu_{F^+}}{\sigma_{F^+}} = \frac{F^+ - \mu_{F^+}}{v_{F^+} \mu_{F^+}} \\ Z_{N^*} &= \frac{N^* - \mu_{N^*} - \beta_0(F^+ - \mu_{F^+})}{\sigma_{N^*|F^+}} = \frac{N^* - \mu_{N^*} - \beta_0(F^+ - \mu_{F^+})}{v_{N^*|F^+} \mu_{N^*}} \\ F^+ &= h(N^*) \\ F \mu_{F^+} &= h(\mu_{N^*}) \\ F &> F^0 \\ \beta &> \beta^0 \end{aligned} \quad (11.201)$$

Para su solución puede utilizarse el siguiente algoritmo:

Algorithm 11.2 (Método combinado con probabilidades de fallo y coeficientes de seguridad).

Entrada: Parámetros estadísticos $\mu_{F^+} = \mu_{1/\psi}$, v_{F^+} , $v_{N^*|F^+}$ y β_0 , límites inferiores del coeficiente de seguridad y del índice de fiabilidad F^0 y β^0 , un coeficiente actualizador ρ y una tolerancia ϵ para controlar la convergencia del proceso.

Salida: Coeficientes de seguridad real F^{actual} e índice de fiabilidad β , y los valores de diseño para los valores clásicos F^* y μ_{N^*} , y los modernos F^+ y N^* .

Etapas 1. Hacer $F^{actual} = F^0$ y $F^{previous} = F^0 + 2\epsilon$.

Etapas 2. Si $|F^{actual} - F^{previous}| < \epsilon$ ir a la etapa 6. Si no, hacer $F^{previous} = F^{actual}$, $F^* = F^{actual} \times \mu_{F^+}$.

Etapas 3. Resolver el problema (11.189)-(11.190), o entrar en el gráfico de la figura 11.19 con F^* , para obtener μ_{N^*} .

Etapas 4. Resolver el problema de optimización (11.194)-(11.195), y obtener los valores de diseño F^+ y N^* , y β .

Etapas 5. Si $\beta > \beta^0$, entonces ir a la etapa 2. Si no, hacer $F^{actual} = F^{actual} + \rho(\beta^0 - \beta)$ e ir a la etapa 2.

Etapas 6. Devolver F^{actual} , β , F^* , μ_{N^*} , F^+ y N^* .



11.4.6 Sensitivity analysis

Para obtener las sensibilidades de β con respecto a μ_{F^+} , v_{F^+} , $v_{N^*|F^+}$ y β_0 se consideran como variables artificiales y se incluyen las restricciones que fijan esas variables a sus valores dato. En otras palabras, se transforma el problema de optimización de la etapa 2 en:

$$\text{Minimizar } \beta = \sqrt{Z_{F^+}^2 + Z_{N^*}^2} \quad (11.202)$$

$$Z_{F^+}, Z_{N^*}, \mu_{F^+}^+, \mu_{N^*}^+, v_{F^+}^+, v_{N^*|F^+}^+, \beta_0^+$$

sujeto a

$$\begin{aligned} Z_{F^+} &= \frac{F^+ - \mu_{F^+}^+}{v_{F^+} \mu_{F^+}^+} \\ Z_{N^*} &= \frac{N^* - \mu_{N^*}^+ - \beta_0^+ (F^+ - \mu_{F^+}^+)}{v_{N^*|F^+} \mu_{N^*}^+} \\ F^+ &= h(N^*) \\ F \mu_{F^+} &= h(\mu_{N^*}^+) \\ F &> F^0 \\ \beta &> \beta^0 \\ \mu_{F^+}^+ &= \mu_{F^+} \\ v_{F^+}^+ &= v_{F^+} \\ v_{N^*|F^+}^+ &= v_{N^*|F^+} \\ \beta_0^+ &= \beta_0 \end{aligned} \quad (11.203)$$

Entonces, los valores de las variables duales correspondientes son las sensibilidades deseadas.

Las sensibilidades de F^* con respecto a N_0^* , pueden obtenerse directamente del problema dual (11.189)-(11.190), porque N_0^* aparece en la última restricción.

Los algoritmos precedentes se han implementado en el GAMS (véase Castillo et al. [30]), y usados para resolver el siguiente ejemplo.

Computational Example 11.7 (Método combinado). En este ejemplo se ilustra el método exacto y el aproximado. Para el exacto se usan los datos:

$$\begin{aligned} F^0 &= 2.2; \beta^0 = 2.0; \mu_{F^+} = \mu_{1/\psi} = 5.67; \\ v_{F^+} &= 0.1; v_{N^*|F^+} = 0.1; \beta_0 = 1. \end{aligned}$$

Resolviendo el problema (11.200)-(11.201) se obtiene:

$$F^+ = 4.562; N^* = 0.292; \beta = 2.04; F = 2.2; \mu_{N^*} = 1.487.$$

y las sensibilidades de la tabla 11.17.

Y para el método aproximado se asume

$$\begin{aligned} F^0 &= 1.2; \beta^0 = 2.0; \rho = 0.5; \mu_{F^+} = \mu_{1/\psi} = 5.67; \\ v_{F^+} &= 0.1; v_{N^*|F^+} = 0.1; \beta_0 = 1. \end{aligned}$$

Entonces, aplicando el algoritmo 11.2, se obtiene:

$$F^* = 12.34; \mu_{N^*} = 1.465; F^+ = 4.58; N^* = 0.294; \beta_{actual} = 2.0; F_{actual} = 2.18.$$

Table 11.17: Sensibilidades del índice de fiabilidad con respecto a los parámetros del modelo.

$\frac{\partial \beta}{\partial F^0}$	$\frac{\partial \beta}{\partial \mu_{F^+}}$	$\frac{\partial \beta}{\partial v_{N^* F^+}}$	$\frac{\partial \beta}{\partial v_{F^+}}$	$\frac{\partial \beta}{\partial \beta_0}$
1.734	0.104	-1.675	-18.72	-2.136

Table 11.18: Evolución del procedimiento iterativo.

Iteración	C. seguridad F	Clásico		Punto de diseño		β
		μ_{N^*}	μ_{F^*}	N^*	F^+	
1	1.200	0.594	6.806	0.411	5.491	0.320
2	2.040	1.335	11.568	0.310	4.707	1.761
3	2.159	1.448	12.246	0.296	4.599	1.969
4	2.175	1.463	12.334	0.295	4.585	1.996
5	2.177	1.465	12.345	0.294	4.584	1.999
6	2.177	1.465	12.347	0.294	4.583	2.000
7	2.177	1.465	12.347	0.294	4.583	2.000

Table 11.19: Sensibilidades del índice de fiabilidad β con respecto a los parámetros estadísticos.

$\frac{\partial\beta}{\partial\mu_{N^*}}$	$\frac{\partial\beta}{\partial\mu_{F^+}}$	$\frac{\partial\beta}{\partial v_{N^* F^+}}$	$\frac{\partial\beta}{\partial v_{F^+}}$	$\frac{\partial\beta}{\partial\beta_0}$
1.82	-0.562	-1.6	-18.4	-2.1

El proceso se ilustra en la tabla 11.18, que muestra la rápida convergencia del método.

Nótese que la restricción asociada al coeficiente de seguridad $F > F^0 = 1.2$ no es activa, mientras que la restricción asociada al índice de fiabilidad $\beta > \beta^0 = 2$ es activa.

Finalmente, las sensibilidades del índice de fiabilidad β con respecto a los parámetros estadísticos $\mu_{N^*}, \mu_{F^+}, v_{N^*|F^+}, v_{F^+}$ y β_0 se han calculado utilizando el método descrito en 11.4.4, y los resultados se muestran en la tabla 11.19, donde se observa que el índice de fiabilidad aumenta en 1.82 unidades por incremento unitario de μ_{N^*} , y disminuye $-0.562, -1.6, -18.4$ y -2.1 , por incremento unitario de $\mu_{F^+}, v_{N^*|F^+}, v_{F^+}$ y β_0 , respectivamente.

■

Appendix A

GAMS Codes

En el presente apéndice se presentan y explican las implementaciones en GAMS de las aplicaciones tratadas en los capítulos anteriores. El capítulo está estructurado de la siguiente manera: en la sección A.1 se trata el método de relajación, en la sección A.2 se presenta la implementación de la aproximación por hiperplanos, en la sección A.3 se muestra la descomposición de Benders, en la A.4 se muestra de nuevo el método de relajación pero aplicado al ejemplo del muro, en la sección A.5 se trata la implementación del muro resuelta mediante la aproximación por hiperplanos, y por último en A.6 se trata el análisis de sensibilidad.

Las explicaciones del código irán con un asterisco (*) al comienzo de cada línea, es la forma de poner comentarios en el GAMS. O bien, cuando los comentarios ocupan varias líneas de la forma:

```
$Ontext
Comentarios ...
...
...
...
$Offtext
```

A.1 GAMS Code for the Relaxation Method

En esta sección se presenta la implementación del método de relajación para resolver problemas de fiabilidad basándose en el método combinado (PSFM). Está aplicado para la resolución del ejemplo computacional 9.1:

```
$title Metodo de relajacion aplicada al PSFM
* Inicializo el fichero de escritura de resultados
file out1 /relaja.out/;
```

```
SETS
  X variables x /mu_x1,mu_x2/
  A parametros y /mu_y1,mu_y2/
  B parametros v /v_x_1,v_x_2,v_y_1,v_y_2/
```

```

M restricciones h /first,second/
IT numero maximo de iteraciones /1*20/;

ALIAS(M,Maux);

SCALARS
  epsilon maxima tolerancia permitida /1e-5/
  error parametro de almacenamiento del error /1/
  iteration contados de iteraciones /0/
  rho coeficiente de relajacion /0.2/;

PARAMETERS
  Srho(M) multiplicador de lagrange asociado a gi para la actualizacion automatica
  Flo(M) limites inferiores de las restricciones g1 y g2/
    first      1.2
    second     1.6
  /
  Fr(M) valores reales de las restricciones g1 y g2
  Fpar(M) valores actuales de los limites inferiores de las restricciones g1 y g2
  betalo(M) limites inferiores de las restricciones h1 y h2/
  first       3
  second      4
  /
  betaa(M) valores reales de las restricciones h1 y h2
  betaux(M) valores auxiliares para el calculo del error
  DataX(X) valores iniciales de las variables x/
    mu_x1     5
    mu_x2     5
  /
  DataA(A) valores de los parametros y/
    mu_y1     1
    mu_y2     1
  /
  DataB(B) valores de los parametros v/
    v_x_1    0.02
    v_x_2    0.02
    v_y_1    0.1
    v_y_2    0.1
  /
  errores(IT) parametro del almacenamiento del error en cada iteracion;

VARIABLES
  zc variable objetivo de cada uno de los modelos
  VarX(X) variables de diseno x
  VarU(X) variables u asociadas a las x
  VarV(A) variables v asociadas a las y
  Z(B) variables auxiliares z;

EQUATIONS
* Ecuaciones del modelo para resolucion del problema maestro (clasico)
  ZClassic, firste, seconde

```



```

* Ecuaciones para la resolucio de los subproblemas
  Zbeta, ZUdef, ZVdef, Ffirst, Fsecond;

* Ecuaciones del modelo para resolucio del problema maestro (clasico)
  ZClassic..zc=e=sqr(VarX('mu_x1')/DataA('mu_y1'))+sqr(VarX('mu_x2')/DataA('mu_y2'));
  firste..VarX('mu_x1')*VarX('mu_x2')=g=DataA('mu_y1')*DataA('mu_y2')*Fpar('first');
  seconde..sqrt(DataA('mu_y1')*VarX('mu_x2'))=g=DataA('mu_y2')*sqrt(VarX('mu_x1'))...
  ...*Fpar('second');

* Ecuaciones para la resolucio de los subproblemas
  Zbeta..zc=e=sqrt(sum(B,sqr(Z(B)))));
  ZUdef(X,B)$(ORD(B)=ORD(X))..Z(B)=e=(VarU(X)-DataX(X))/(DataB(B)*DataX(X));
  ZVdef(A,B)$(ORD(A)+CARD(X)=ORD(B))..Z(B)=e=(VarV(A)-DataA(A))/(DataB(B)*DataA(A));
  Ffirst..VarU('mu_x1')*VarU('mu_x2')=e=VarV('mu_y1')*VarV('mu_y2');
  Fsecond..sqrt(VarV('mu_y1'))*VarU('mu_x2')=e=VarV('mu_y2')*sqrt(VarU('mu_x1'));

* Planteamiento de los modelos clasico y subproblemas
  MODEL classic /ZClassic,firste,seconde/;
  MODEL mfirst/Zbeta,ZUdef,ZVdef,Ffirst/;
  MODEL msecond/Zbeta,ZUdef,ZVdef,Fsecond/;

* Inicializacion de los limites inferiores de las restricciones g1 y g2
* a los valores iniciales
  Fpar(M)=Flo(M);
  betaa(M)=betalo(M);

* Comienzo del procedimiento iterativo
loop(IT$(error>epsilon),
  iteration=iteration+1;
* Escritura del numero de iteraciones
  put out1;
  put "  ITERATION ",iteration:5//;
  put "----- CLASSICAL -----"/;
* Almacenamiento de los valores de h1 y h2 para controlar el error
  betaux(M)=betaa(M);
* Valores iniciales de las variables x
  VarX.l(X)=DataX(X);
* Resolucio del model clasico
  SOLVE classic USING nlp MINIMIZING zc;
* Escritura de resultados
  put "cost=",zc.l:8:3, " modelstat= ",classic.modelstat,"...
  ... solvestat= ",classic.solvestat/;
  put " cpu=",classic.resusd:9:5//;
  put "Los valores medios son A=",VarX.l('mu_x1'):8:3," y B=",VarX.l('mu_x2'):8:3/;
  put "Actual safety factors"/;
* Calculo del valor actual de las restricciones g1 y g2
  Fr('first')=(VarX.l('mu_x1')*VarX.l('mu_x2'))/(DataA('mu_y1')*DataA('mu_y2'));
  Fr('second')=sqrt(DataA('mu_y1')*VarX.l('mu_x2'))/...
  ... (DataA('mu_y2')*sqrt(VarX.l('mu_x1')));
* Escritura de los valores actuales
  put "Fr('first')=", Fr('first'):8:3, " Fr('second')=", Fr('second'):8:3/;

```

```

* Almacenamiento en el parametro asociado de los valores optimos de x
  DataX(X)=VarX.l(X);
* Resolucion de los subproblemas
  loop(M,
* Valores iniciales de las variables u y v
  VarU.l(X)=DataX(X);
  VarV.l(A)=DataA(A);
  if(ORD(M) eq 1,
* Resolucion del subproblema asociado a h1
  SOLVE mfirst USING nlp MINIMIZING zc;
* Escritura de resultados
  put "beta=",zc.l:8:3, " modelstat= ",mfirst.modelstat,",...
  ... solvestat= ",mfirst.solvestat/;
  put " cpu=",mfirst.resusd:9:5//;
* Almacenamiento de la variable dual asociada a la restriccion g1
  Srho(M)=Ffirst.m;
  else
* Resolucion del subproblema asociado a h2
  SOLVE msecond USING nlp MINIMIZING zc;
* Escritura de resultados
  put "betas=",zc.l:8:3, " modelstat= ",msecond.modelstat,",...
  ... solvestat= ",msecond.solvestat/;
  put " cpu=",msecond.resusd:8:3//;
* Almacenamiento de la variable dual asociada a la restriccion g2
  Srho(M)=Fsecond.m;
  );
* Almacenamiento de los valores optimos de h1 y h2
  betaa(M)=zc.l;
  );
* Escritura de valores intermedios
  loop(M,
  put "beta('",M.tl,"')=",betaa(M):6:3," Fpar('",M.tl,"')=",Fpar(M):6:3,...
  ..." Fr('",M.tl:5,"')=",Fr(M):8:3/;
  );
* Actualizacion de los limites inferiores de g1 y g2
* con el coeficiente de relajacion
* Fpar(M)=Fpar(M)+rho*(betalo(M)-betaa(M));
* Actualizacion de los limites inferiores de g1 y g2 con el metodo dual
  Fpar(M)=Fpar(M)-(betalo(M)-betaa(M))/Srho(M);
* Comprobacion de que los nuevos limites no son inferiores a los requeridos
  loop(M,
  if(Fpar(M)<Flo(M),Fpar(M)=Flo(M));
  );
* Calculo del error
  error=0.0;
  loop(M,
  if((abs(betaa(M)-betaux(M))/betaa(M))>error and betaa(M)>0,
  error=(abs(betaa(M)-betaux(M))/betaa(M));
  );
  );
* Almacenamiento del error

```

```

    errores(IT)=error;
* Escritura
    put "error(",iteration:5,")= ",error:14:12/;

    put "-----"//;
);
* Fin del procedimiento iterativo

```

A.2 GAMS Code for the Hyperplane Aproximation

En esta subsección se presenta la implementación del método de aproximación por hiperplanos para resolver problemas de fiabilidad basándose en el método combinado (PSFM). Está aplicado para la resolución del ejemplo computacional 9.1:

```

$title Metodo de relajacion aplicada al PFSM
* Inicializo el fichero de escritura de resultados
file out1 /hiper.out/;

SETS
    X variables x /mu_x1,mu_x2/
    A parametros y /mu_y1,mu_y2/
    B parametros v /v_x_1,v_x_2,v_y_1,v_y_2/
    M restricciones h /first,second/
    IT numero maximo de iteraciones /1*20/
    ITER(IT) conjunto dinamico para controlar los hiperplanos
    FIRST(M) conjunto dinamico para la primera iteracion;

* Desactivacion de los hiperplanos
ITER(IT)=no;
FIRST(M)=no;
ALIAS(M,Maux);

SCALARS
    epsilon maxima tolerancia permitida /1e-5/
    error parametro de almacenamiento del error /1/
    iteration contados de iteraciones /0/;

PARAMETERS
    Flo(M) limites inferiores de las restricciones g1 y g2/
        first          1.2
        second         1.6
    /
    Fr(M) valores reales de las restricciones g1 y g2
    betalo(M) limites inferiores de las restricciones h1 y h2/
    first            3
    second           4
    /
    betaa(M) valores reales de las restricciones h1 y h2
    betaux(M) valores auxiliares para el calculo del error
    DataX(X) valores iniciales de las variables x/

```

```

mu_x1  5
mu_x2  5
/
DataA(A) valores de los parametros y/
mu_y1  1
mu_y2  1
/
DataB(B) valores de los parametros v/
v_x_1  0.02
v_x_2  0.02
v_y_1  0.1
v_y_2  0.1
/
errores(IT) parametro del almacenamiento del error en cada iteracion
lambdaK(IT,M,X) pendientes de los hiperplanos aproximantes
XesK(IT,X) valores optimos de las variables x
betaK(IT,M) valores de las restricciones h para cada valor de x;

VARIABLES
zc variable objetivo de cada uno de los modelos
VarX(X) variables de diseno x
VarU(X) variables u asociadas a las x
VarV(A) variables v asociadas a las y
Z(B) variables auxiliares z
beta(M) valores actuales de las restricciones h1 y h2;

EQUATIONS
* Ecuaciones del modelo para resolucion del problema maestro (clasico)
  ZClassic, firste, seconde, betadef, auxbeta
* Ecuaciones para la resolucion de los subproblemas
  Zbeta, ZUdef, ZVdef, Ffirst, Fsecond, fixedX;

* Ecuaciones del modelo para resolucion del problema maestro (clasico)
  ZClassic..zc=e=sqr(VarX('mu_x1')/DataA('mu_y1'))+sqr(VarX('mu_x2')/DataA('mu_y2'));
  firste..VarX('mu_x1')*VarX('mu_x2')=g=DataA('mu_y1')*DataA('mu_y2')*Flo('first');
  seconde..sqrt(DataA('mu_y1'))*VarX('mu_x2')=g=DataA('mu_y2')*sqrt(VarX('mu_x1'))*...
    ...Flo('second');
  betadef(ITER,M)$(FIRST(M))..beta(M)=l=betaK(ITER,M)+...
    ...sum(X,lambdaK(ITER,M,X)*(VarX(X)-XesK(ITER,X)));
  auxbeta(M)$(FIRST(M))..beta(M)=g=betaLo(M);

* Ecuaciones para la resolucion de los subproblemas
  Zbeta..zc=e=sqrt(sum(B,sqr(Z(B)))));
  ZUdef(X,B)$(ORD(X)=ORD(B))..Z(B)=e=(VarU(X)-VarX(X))/(DataB(B)*VarX(X));
  ZVdef(A,B)$(ORD(A)+CARD(X)=ORD(B))..Z(B)=e=(VarV(A)-DataA(A))/(DataB(B)*DataA(A));
  Ffirst..VarU('mu_x1')*VarU('mu_x2')=e=VarV('mu_y1')*VarV('mu_y2');
  Fsecond..sqrt(VarV('mu_y1'))*VarU('mu_x2')=e=VarV('mu_y2')*sqrt(VarU('mu_x1'));
  fixedX(X)..VarX(X)=e=DataX(X);

* Planteamiento de los modelos clasico y subproblemas
  MODEL classic /ZClassic,firste,seconde,betadef,auxbeta/;

```

```

MODEL mfirst /Zbeta,ZUdef,ZVdef,Ffirst,fixedX/;
MODEL msecond /Zbeta,ZUdef,ZVdef,Fsecond,fixedX/;

* Inicializacion de los limites inferiores de las restricciones g1 y g2
* a los valores iniciales
  betaa(M)=betalo(M);
  lambdaK(IT,M,X)=0.0;
  XesK(IT,X)=0.0;
  betaK(IT,M)=0.0;

* Comienzo del procedimiento iterativo
loop(IT$(error>epsilon),
  iteration=iteration+1;
* Escritura del numero de iteraciones
  put out1;
  put "  ITERATION ",iteration:5//;
  put "----- CLASSICAL -----"/;
* Almacenamiento de los valores de h1 y h2 para controlar el error
  betaux(M)=betaa(M);
* Valores iniciales de las variables x
  VarX.l(X)=DataX(X);
* Resolucion del model clasico
  SOLVE classic USING nlp MINIMIZING zc;
* Escritura de resultados
  put "cost=",zc.l:8:3, " modelstat= ",classic.modelstat,"...
  ... solvestat= ",classic.solvestat/;
  put " cpu=",classic.resusd:9:5//;
  put "Los valores medios son A=",VarX.l('mu_x1'):8:3," y B=",VarX.l('mu_x2'):8:3/;
  put "Actual safety factors"/;
* Calculo del valor actual de las restricciones g1 y g2
  Fr('first')=(VarX.l('mu_x1')*VarX.l('mu_x2'))/(DataA('mu_y1')*DataA('mu_y2'));
  Fr('second')=sqrt(DataA('mu_y1')*VarX.l('mu_x2')/...
  ... (DataA('mu_y2')*sqrt(VarX.l('mu_x1')));
* Escritura de los valores actuales
  put "Fr('first')=", Fr('first'):8:3," Fr('second')=", Fr('second'):8:3/;
* Almacenamiento en el parametro asociado de los valores optimos de x
  DataX(X)=VarX.l(X);
  XesK(IT,X)=DataX(X);
* Resolucion de los subproblemas
  loop(M,
* Valores iniciales de las variables u y v
  VarU.l(X)=DataX(X);
  VarV.l(A)=DataA(A);
  if(ORD(M) eq 1,
* Resolucion del subproblema asociado a h1
  SOLVE mfirst USING nlp MINIMIZING zc;
* Escritura de resultados
  put "beta=",zc.l:8:3, " modelstat= ",mfirst.modelstat,"...
  ... solvestat= ",mfirst.solvestat/;
  put " cpu=",mfirst.resusd:9:5//;
  else

```

```

* Resolucion del subproblema asociado a h2
    SOLVE msecond USING nlp MINIMIZING zc;
* Escritura de resultados
    put "betas=",zc.1:8:3, " modelstat= ",msecond.modelstat,",...
    ... solvestat= ",msecond.solvestat/;
    put " cpu=",msecond.resusd:8:3//;
);
* Almacenamiento de los valores optimos de h1 y h2
    betaa(M)=zc.1;
    lambdaK(IT,M,X)=fixedX.m(X);
    betaK(IT,M)=zc.1;
);
* Escritura de valores intermedios
    loop(M,
        put "beta('",M.tl,"')=",betaa(M):6:3," Fr('",M.tl:5,"')=",Fr(M):8:3/;
    );
* Calculo del error
    error=0.0;
    loop(M,
        if((abs(betaa(M)-betaux(M))/betaa(M))>error and betaa(M)>0,
            error=(abs(betaa(M)-betaux(M))/betaa(M));
        );
    );
    FIRST(M)=yes;
    ITER(IT)=yes;
* Almacenamiento del error
    errores(IT)=error;
* Escritura
    put "error(",iteration:5,")= ",error:14:12/;

    put "-----"//;
);
* Fin del procedimiento iterativo

```

A.3 GAMS Code for the Benders Decomposition

A continuación se describe el código GAMS en el que se implementa la resolución del problema del dique de escollera mediante la descomposición de Benders.

```
$title Ejemplo del dique de escollera
```

```

* Inicialización del fichero de salida de resultados
file out /rebasebenders.out/;
put out;

```

```
SETS
```

```

I variables aleatorias/1*2/
IT numero maximo de iteraciones del metodo de Benders /1*20/
ITER(IT) conjunto dinamico para controlar los cortes de Benders;

```

```

* Desactivacion de todos los hiperplanos aproximantes
ITER(IT)=no;

SCALARS
pi /3.1415926535898/
gra gravedad terrestre/9.81/
error valor inicial del error/1/
Csup cota superior del coste total optimo
Cinf cota inferior del coste total optimo /5000/
Toler tolerancia admisible /1e-3/

SCALARS
* Parametros
Dwl nivel del mar de diseno (m) /20/
Au parametro experimental /1.05/
Bu parametro experimental /-0.67/
cc coste del hormigon (euro) /60/
ca coste de la escollera (euro) /2.4/

* Random model parameters, sea state descriptors
Hs altura de ola significativa (m) /5/
Tz periodo medio (s) /10/
dst duracion del estado de mar (s) /3600/

* Auxiliary scalars
pf probabilidad de fallo de una ola
pfD probabilidad de fallo en la vida util;

PARAMETERS
Ctotal(IT) coste total para cada iteracion
ValorFc(IT) francobordos de cada iteracion
LambdaFc(IT) derivadas parciales del coste total...
...con respecto al francobordo en cada iteracion
ValorTan(IT) pendientes del talud de cada iteracion
LambdaTan(IT) derivadas parciales del coste total...
...con respecto a la pendiente del talud en cada iteracion;

VARIABLES
cd coste de construccion
ci coste de seguro
beta indice de fiabilidad
alfacost variable auxiliar para el problema maestro
z(I) variables normales estandar;

POSITIVE VARIABLES
* Variables aleatorias
H altura de ola, T periodo de la ola

* Variables de diseño, variables de complicacion
tan pendiente del talud, Fc francobordo
auxtan variables auxiliar asociada a la pendiente...

```

```

...del talud para el problema maestro
auxFc variables auxiliar asociada al francobordo...
...para el problema maestro

* Variables auxiliares
Ir numero de Iribarren, L longitud de onda, Ru run up, d altura del espaldon
vc volumen de hormigon, va volumen de escollera;

* Limites de las variables,
* muy importante para conseguir la convergencia
auxtan.lo=1/5; auxtan.up=1/2; L.lo=10; H.up=2.2*Hs; T.up=2.2*Tz;

EQUATIONS
* Optimizacion global
cddf definicion del coste de construccion
ddf definicion de la altura del espaldon
vcdf definicion del volumen de hormigon
vadf definicion del volumen de escollera

* Optimizacion asociada a la fiabilidad
betadef definicion del indice de fiabilidad
Zdef1 transformacion de Rosenblatt para la variable H
Zdef2 transformacion de Rosenblatt para la variable T
Iridf definicion del numero de Iribarren
Ldf ecuacion de la dispersion
rudf definicion del run up
verdf ecuacion de estado limite asociado al rebase

* Problema maestro
Restric(IT) cortes de Benders
auxmaster cota inferior del coste optimo;

* Optimizacion global
cddf..cd=e=cc*vc+ca*va;
ddf..Fc=e=2+d;
vcdf..vc=e=10*d;
vadf..va=e=0.5*(Dwl+2)*(Dwl+46+(Dwl+2)/tan);

* Optimizacion asociada a la fiabilidad
betadef..beta=e=sqrt(sum(I,sqr(z(I))));
Zdef1..errorf(z('1'))=e=1-exp(-2*sqr(H/Hs));
Zdef2..errorf(z('2'))=e=1-exp(-0.675*power((T/Tz),4));
Iridf..Ir*sqr(H/L)=e=tan;
Ldf..2*pi*L*(exp(2*pi*Dwl/L)+exp(-2*pi*Dwl/L))=e...
...T*T*gra*(exp(2*pi*Dwl/L)-exp(-2*pi*Dwl/L));
rudf..Ru=e=H*Au*(1-exp(Bu*Ir));
verdf..Fc=e=Ru;

* Problema maestro
Restric(ITER)..alfacost=g=Ctotal(ITER)+...
...LambdaFc(ITER)*(auxFc-ValorFc(ITER))+...

```



```

...LambdaTan(ITER)*(auxtan-ValorTan(ITER));
auxmaster..alfacost=g=5000;

MODEL sublevel/betadef,Zdef1,Zdef2,Iridf,Ldf,rudf,verdf/;
MODEL cdirect/cddf,ddf,vcdf,vadf/;
MODEL Master/Restric,auxmaster/;

* Valores iniciales de los parametros
  Cttotal(IT)=0.0; LambdaFc(IT)=0.0; ValorFc(IT)=0.0;
  LambdaTan(IT)=0.0; ValorTan(IT)=0.0;

* Valores iniciales de las variables de complicacion
Fc.fx=7; tan.fx=1/3;

* Comienzo del procedimiento iterativo
loop(IT$(error gt TOLER),

* Escritura del numero de iteracion
put "      Iteration= ",ord(IT):12:8//;

  if(ORD(IT)>1,

* Resolucion del problema maestro para obtener los nuevos valores
* de las variables de complicacion
  SOLVE Master USING lp MINIMIZING alfacost;

* Escritura de resultados
  put "alfacost= ",alfacost.l:12:4,", modelstat= ",...
  ...Master.modelstat,", solvestat= ",Master.solvestat/;

* Actualización de los nuevos valores de las variables de complicacion
  Fc.fx=auxFc.l; tan.fx=auxtan.l;

* Actualizacion del limite inferior del coste para la iteracion IT
  Cinf=alfacost.l;
);

* Almacenamiento de las variables de complicacion para los cortes...
...de Benders
  ValorFc(IT)=Fc.l; ValorTan(IT)=tan.l;

* Activacion del corte de Benders para el siguiente problema maestro
  ITER(IT)=yes;

* Escritura
  put "Complicating variables: Fc=",Fc.l:6:3,", tan=",tan.l:6:3/;

* Valores iniciales para la resolucion del problema de fiabilidad...
*muy importante para la convergencia del modelo
  H.l=1.5*Hs; T.l=1.1*Tz; L.l=136.931; Ir.l=tan.l/sqrt(H.l/L.l);
  Ru.l=H.l*Au*(1-exp(Bu*Ir.l)); z.l('1')=2.28; z.l('2')=0.32;

```

```

beta.l=sqrt(sum(I,sqr(z.l(I))));

* Resolucion subproblema asociado a la fiabilidad para los nuevos valores fijos...
* de las variables de complicacion
SOLVE sublevel USING nlp MINIMIZING beta;

* Escritura
put "pf= ",(errorf(-beta.l)):12:8," , modelstat= ",...
...sublevel.modelstat," , solvestat= ",sublevel.solvestat/;

* Calculo de las probabilidades de fallo
pf=errorf(-beta.l); pfD=1-(1-pf)**(dst/Tz);

* Derivadas parciales de pfD con respecto a las variables de complicacion
LambdaFc(IT)=-dst*(1-pf)**(dst/Tz-1)*exp(-beta.l*beta.l/2)*...
...(Fc.m)/(Tz*sqrt(2*pi));
LambdaTan(IT)=-dst*(1-pf)**(dst/Tz-1)*exp(-beta.l*beta.l/2)*...
...(tan.m)/(Tz*sqrt(2*pi));

* Coste del seguro en funcion de pfD
ci.l=5000+125000000*pfD**2;

* Escritura
put "pfD= ",pfD:12:8/;
put "insurance cost= ",(ci.l):12:4/;
put "LambdaFc1(",ord(IT):2," )=",LambdaFc(IT):12:4,...
..."LambdaTn1(",ord(IT):2," )=",LambdaTan(IT):12:4/;

* Sumo al coste total el coste del seguro
Ctotal(IT)=ci.l;

* Derivadas parciales del coste del seguro con respecto a las
* variables de complicacion
LambdaFc(IT)=LambdaFc(IT)*(2*125000000*pfD);
LambdaTan(IT)=LambdaTan(IT)*(2*125000000*pfD);

* Escritura
put "LambdaFc2(",ord(IT):2," )=",LambdaFc(IT):12:4,...
..."LambdaTn2(",ord(IT):2," )=",LambdaTan(IT):12:4/;

* Resolucion del modelo auxiliar para calcular las derivadas
* del coste directo con respecto a
* las variables de complicacion
SOLVE cdirect USING nlp MINIMIZING cd;

* Escritura
put "direct cost= ",cd.l:12:4," , modelstat= ",...
...cdirect.modelstat," , solvestat= ",cdirect.solvestat/;

* Calculo del coste total
Ctotal(IT)=Ctotal(IT)+cd.l;

```

A.4. GAMS CODE FOR THE RETAINING WALL USING THE RELAXATION METHOD 425

```
* Derivadas parciales del coste total con respecto a las
* variables de complicacion
LambdaFc(IT)=LambdaFc(IT)+Fc.m;
LambdaTan(IT)=LambdaTan(IT)+tan.m;

* Escritura
put "Ctotal(",ord(IT):2," )=",Ctotal(IT):12:4/;
put "LambdaFc(",ord(IT):2," )=",LambdaFc(IT):12:4,...
..."LambdaTn(",ord(IT):2," )=",LambdaTan(IT):12:4/;

* Limite superior ddel optimo del coste total
Csup=Ctotal(IT);

* Calculo del error
error=(abs(Csup-Cinf)/Cinf);

* Escritura
put "Upper bound= ",Csup:12:4/;
put "Lower bound= ",Cinf:12:4/;
put "error= ",error:15:10//;
);
```

A.4 GAMS Code for the Retaining Wall Using the Relaxation Method

En esta sección se muestra el código GAMS del ejemplo del muro resuelto mediante el método de relajación:

```
$title Muro

file out1 /murorel.out/;

SETS
  X variables de diseno /mua,mub/
  A parametros y agentes /munu,muT,mugamma,muH,muS/
  B parametros estadisticos/sigmaa,sigmab,sigmanu,sigmaT,sigmagamma,sigmaH,sigmaS/
  M modos de fallo /turning,sliding,bearing/
  IT numero maximo de iteraciones /1*20/;

ALIAS(M,Maux);

SCALARS
  epsilon maxima tolerancia permitida /1e-8/
  error parametro de almacenamiento del error /1/
  iteration contador de iteraciones /0/
  rho coeficiente de relajacion /0.8/;

PARAMETERS
```

```

Flo(M) limites inferiores de los coeficientes de seguridad/
turning      1.5
sliding      1.6
bearing      1.5
/
Fr(M) valores reales de los coeficientes de seguridad
Fpar(M) valores actuales de los limites inferiores de los coeficientes de seguridad
betalo(M) limites inferiores de los indices de fiabilidad/
turning      3.0
sliding      3.0
bearing      3.0
/
betaa(M) valores reales de los indices de fiabilidad
betaux(M) valores auxiliares para el calculo del error
DataX(X) valores iniciales de las variables de diseno/
  mua 3
  mub 6
/
DataA(A) valores de los parametros y agentes/
  munu 0.3
  muT 50.0
  mugamma 23.0
  muH 3.0
  muS 220.0
/
DataB(B) valores de los parametros estadisticos/
  sigmaa 0.01
  sigmab 0.01
  sigmanu 0.05
  sigmaT 15.0
  sigmagamma 0.46
  sigmaH 0.2
  sigmaS 16.0
/;

VARIABLES
  zc variable objetivo decada uno de los modelos
  VarX(X) variables de diseno x
  VarU(X) variables u asociadas a las x
  VarV(A) variables v asociadas a las y
  Z(B) variables auxiliares z;

EQUATIONS
* Ecuaciones del modelo para resolucion del problema maestro (clasico)
ZClassic, turning, sliding, bearing, geometric
* Ecuaciones para la resolucion de los subproblemas
Zbeta, ZUdef, ZVdef, Fturning, Fsliding, Fbearing;

* Ecuaciones del modelo para resolucion del problema maestro (clasico)
ZClassic..zc=e=VarX('mua')*VarX('mub');
turning..VarX('mua')*VarX('mua')*VarX('mub')*DataA('mugamma')=g=...

```

A.4. GAMS CODE FOR THE RETAINING WALL USING THE RELAXATION METHOD427

```

...2*DataA('muH')*DataA('muT')*Fpar('turning');
sliding..VarX('mua')*VarX('mub')*DataA('munu')*DataA('mugamma')=g...
...DataA('muT')*Fpar('sliding');
bearing..DataA('muS')=g=VarX('mub')*DataA('mugamma')*Fpar('bearing');
geometric..VarX('mua')*2=e=VarX('mub');

* Ecuaciones para la resolucio de los subproblemas
Zbeta..ZC=e=sqrt(sum(B,sqr(Z(B))));
ZUdef(X,B$(ORD(X)=ORD(B))..Z(B)=e=(VarU(X)-DataX(X))/DataB(B);
ZVdef(A,B$(ORD(A)+CARD(X)=ORD(B))..Z(B)=e=(VarV(A)-DataA(A))/(DataB(B));
Fturning..VarU('mua')*VarU('mua')*VarU('mub')*VarV('mugamma')=e...
...2*VarV('muH')*VarV('muT');
Fsliding..VarU('mua')*VarU('mub')*VarV('munu')*VarV('mugamma')=e=VarV('muT');
Fbearing..VarU('mub')*VarV('mugamma')=e=VarV('muS');

* Planteamiento de los modelos clasico y subproblemas
MODEL classic /ZClassic,turning,sliding,bearing,geometric/;
MODEL mturning /Zbeta,ZUdef,ZVdef,Fturning/;
MODEL msliding /Zbeta,ZUdef,ZVdef,Fsliding/;
MODEL mbearing /Zbeta,ZUdef,ZVdef,Fbearing/;

* Inicializacion de los limites inferiores de las restricciones
* asociadas a los coeficientes de seguridad
Fpar(M)=Flo(M);
betaa(M)=betalo(M);

* Comienzo del procedimiento iterativo
loop(IT$(error>epsilon),
iteration=iteration+1;
* Escritura del numero de iteraciones
put out1;
put " ITERATION ",iteration:5//;

put "----- CLASICO -----"/;
* Almacenamiento de los indices de fiabilidad para controlar el error
betaux(M)=betaa(M);
* Valores iniciales de las variables x
VarX.l(X)=DataX(X);
* Resolucion del model clasico
SOLVE classic USING nlp MINIMIZING zc;
* Escritura de resultados
put "coste=",zc.l:12:8, " modelstat= ",classic.modelstat,...
... solvestat= ",classic.solvestat/;
put " cpu=",classic.resusd:9:5//;
put "Los valores medios son A=",VarX.l('mua'):8:3," y B=",VarX.l('mub'):8:3/;
* Calculo del valor actual de los coeficientes de seguridad
put "Valores actuales de los coeficientes"/;
Fr('turning')=(VarX.l('mua')*VarX.l('mua')*VarX.l('mub')*DataA('mugamma')...
.../(2*DataA('muH')*DataA('muT')));
Fr('sliding')=(VarX.l('mua')*VarX.l('mub')*DataA('munu')*DataA('mugamma')...
.../DataA('muT'));

```

```

    Fr('bearing')=DataA('muS')/(VarX.l('mub')*DataA('mugamma'));
* Escritura de los valores actuales
  put "Fr('turning')=", Fr('turning')," Fr('sliding')=", Fr('sliding'),...
  ..." Fr('bearing')=", Fr('bearing')/;
* Almacenamiento en el parametro asociado de los valores optimos de x
  DataX(X)=VarX.l(X);
* Resolucion de los subproblemas
  loop(M,
* Valores iniciales de las variables u y v
  VarU.l(X)=DataX(X);
  VarV.l(A)=DataA(A);
  if(ORD(M) eq 1,
    put "----- VUELCO -----"/;
* Obtencion del indice de fiabilidad para rebase
  SOLVE mturning USING nlp MINIMIZING zc;
* Escritura de resultados
  put "beta=",zc.l:8:3, " modelstat= ",mturning.modelstat,", solvestat= ",...
  ...mturning.solvestat/;
  put " cpu=",mturning.resusd:9:5//;
  else
  if(ORD(M) eq 2,
    put "----- DESLIZAMIENTO -----"/;
* Obtencion del indice de fiabilidad para deslizamiento
  SOLVE msliding USING nlp MINIMIZING zc;
* Escritura de resultados
  put "betas=",zc.l:8:3, " modelstat= ",msliding.modelstat,", solvestat= ",...
  ...msliding.solvestat/;
  put " cpu=",msliding.resusd:8:3//;
  else
  put "----- HUNDIMIENTO -----"/;
* Obtencion del indice de fiabilidad para hundimiento
  SOLVE mbearing USING nlp MINIMIZING zc;
* Escritura de resultados
  put "betas=",zc.l:8:3, " modelstat= ",mbearing.modelstat,", solvestat= ",...
  ...mbearing.solvestat/;
  put " cpu=",mbearing.resusd:8:3//;
  );
);
* Almacenamiento de los valores optimos de los indices de fiabilidad
  betaa(M)=zc.l;
);
* Escritura de valores intermedios
  loop(M,
  put "beta('",M.tl,"')=",betaa(M):6:3," Fpar('",M.tl,"')=",Fpar(M):6:3,...
  ..." Fr('",M.tl:5,"')=",Fr(M):8:3/;
  );
* Actualizacion de los limites inferiores de los coeficientes de seguridad
  Fpar(M)=Fpar(M)+rho*(betalo(M)-betaa(M));
* Comprobacion de que los nuevos limites no son inferiores a los requeridos
  loop(M,
  if(Fpar(M)<Flo(M),Fpar(M)=Flo(M)););

```

```

);
* Calculo del error
error=0.0;
loop(M,
    if(abs((betaa(M)-betaux(M))/betaa(M))>error,
        error=abs((betaa(M)-betaux(M))/betaa(M));
    );
);
* Escritura
put "error(",iteration:5,")= ",error:14:12/;

put "-----"//;
);
* Fin del procedimiento iterativo

```

A.5 GAMS Code for the Hyperplane Approximation Method

A continuación se muestra el código GAMS del ejemplo del muro resuelto mediante el método de aproximación por hiperplanos:

```

$title Muro por hiperplanos
* Inicializo el fichero de escritura de resultados

file out1 /murohiper.out/;

SETS
    X variables de diseno /mua,mub/
    A parametros /munu,muT,mugamma,muH,muS/
    B parametros y agentes /sigmaa,sigmas,sigmanu,sigmaT,sigmagamma,sigmaH,sigmaS/
    M modos de fallo /turning,sliding,bearing/
    IT numero maximo de iteraciones /1*20/
    ITER(IT) conjunto dinamico para controlar los hiperplanos
    FIRST(M) conjunto dinamico para la primera iteracion;

ALIAS(M,Maux);
* Desactivacion de los hiperplanos
ITER(IT)=no;
FIRST(M)=no;

SCALARS
    epsilon maxima tolerancia permitida /1e-8/
    error parametro de almacenamiento del error /1/
    iteration contados de iteraciones /0/;

PARAMETERS
    Flo(M) limites inferiores de los coeficientes de seguridad/
        turning          1.5
        sliding          1.6
        bearing          1.5
    /

```

```

Fr(M) valores reales de los coeficientes de seguridad
betalo(M) limites inferiores de los indices de fiabilidad/
  turning      3.0
  sliding      3.0
  bearing      3.0
/
betaa(M) valores reales de los indices de fiabilidad
betaux(M) valores auxiliares para el calculo del error
DataX(X) valores iniciales de las variables de diseno /
  mua 3
  mub 6
/
DataA(A) valores de los parametros y agentes /
  munu 0.3
  muT 50.0
  mugamma 23.0
  muH 3.0
  muS 220.0
/
DataB(B) valores de los parametros estadisticos/
  sigmaa 0.01
  sigmab 0.01
  sigmanu 0.05
  sigmaT 15.0
  sigmagamma 0.46
  sigmaH 0.2
  sigmaS 16.0
/
lambdaK(IT,M,X) pendientes de los hiperplanos aproximantes
XesK(IT,X) valores optimos de las variables x
betaK(IT,M) valores de los indices de fiabilidad para cada valor de x;

VARIABLES
  zc variable objetivo decada uno de los modelos
  VarX(X) variables de diseno x
  VarU(X) variables u asociadas a las x
  VarV(A) variables v asociadas a los parametros y agentes
  Z(B) variables auxiliares z
  beta(M) valores actuales de los indices de fiabilidad;

EQUATIONS
* Ecuaciones del modelo para resolucion del problema maestro (clasico)
ZClassic,turning,sliding,bearing,geometric,betadef,auxbeta
* Ecuaciones para la resolucion de los subproblemas
Zbeta,ZUdef,ZVdef,Fturning,Fsliding,Fbearing,fixedX;

* Ecuaciones del modelo para resolucion del problema maestro (clasico)
ZClassic..zc=e=VarX('mua')*VarX('mub');
turning..VarX('mua')*VarX('mua')*VarX('mub')*DataA('mugamma')=g...
...2*DataA('muH')*DataA('muT')*Flo('turning');
sliding..VarX('mua')*VarX('mub')*DataA('munu')*DataA('mugamma')=g...

```



```

...DataA('muT')*Flo('sliding');
bearing..DataA('muS')=g=VarX('mub')*DataA('mugamma')*Flo('bearing');
geometric..VarX('mua')*2=e=VarX('mub');
betadef(ITER,M)$(FIRST(M))..beta(M)=1=betaK(ITER,M)+...
...sum(X,lambdaK(ITER,M,X)*(VarX(X)-XesK(ITER,X)));
auxbeta(M)$(FIRST(M))..beta(M)=g=betao(M);

* Ecuaciones para la resolucio de los subproblemas
Zbeta..zc=e=sqrt(sum(B,sqr(Z(B))));
ZUdef(X,B)$(ORD(X)=ORD(B))..Z(B)=e=(VarU(X)-VarX(X))/(DataB(B));
ZVdef(A,B)$(ORD(A)+CARD(X)=ORD(B))..Z(B)=e=(VarV(A)-DataA(A))/(DataB(B));
Fturning..VarU('mua')*VarU('mua')*VarU('mub')*VarV('mugamma')=e=...
...2*VarV('muH')*VarV('muT');
Fsliding..VarU('mua')*VarU('mub')*VarV('munu')*VarV('mugamma')=e=VarV('muT');
Fbearing..VarU('mub')*VarV('mugamma')=e=VarV('muS');
fixedX(X)..VarX(X)=e=DataX(X);

* Planteamiento de los modelos clasico y subproblemas
MODEL classic /ZClassic,turning,sliding,bearing,geometric,betadef,auxbeta/;
MODEL mturning/Zbeta,ZUdef,ZVdef,Fturning,fixedX/;
MODEL msliding/Zbeta,ZUdef,ZVdef,Fsliding,fixedX/;
MODEL mbearing/Zbeta,ZUdef,ZVdef,Fbearing,fixedX/;

* Inicializacion de los limites inferiores de los indices de fiabilidad
betaa(M)=betao(M);
lambdaK(IT,M,X)=0.0;
XesK(IT,X)=0.0;
betaK(IT,M)=0.0;

* Comienzo del procedimiento iterativo
loop(IT$(error>epsilon),
  iteration=iteration+1;
* Escritura del numero de iteraciones
  put out1;
  put "  ITERATION ",iteration:5//;
  put "----- CLASICO -----"/;
* Almacenamiento de los indices de fiabilidad para controlar el error
  betaux(M)=betaa(M);
* Valores iniciales de las variables x
  VarX.l(X)=DataX(X);
* Resolucion del model clasico
  SOLVE classic USING nlp MINIMIZING zc;
* Escritura de resultados
  put "coste=",zc.l:8:3, " modelstat= ",classic.modelstat,",...
  ... solvestat= ",classic.solvestat/;
  put " cpu=",classic.resusd:8:3//;
  put "Los valores medios son A=",VarX.l('mua'):8:3,",y B=",VarX.l('mub'):8:3/;
  put "Valores actuales de los coeficientes de seguridad"/;
* Calculo del valor actual de los coeficientes de seguridad
  Fr('turning')=(VarX.l('mua')*VarX.l('mua')*VarX.l('mub')*DataA('mugamma')...
  .../(2*DataA('muH')*DataA('muT')));

```

```

    Fr('sliding')=(VarX.l('mua')*VarX.l('mub')*DataA('munu')*DataA('mugamma')...
    .../DataA('muT'));
    Fr('bearing')=DataA('muS')/(VarX.l('mub')*DataA('mugamma'));
* Escritura de los valores actuales
    put "Fr('turning')=", Fr('turning')," Fr('sliding')=", Fr('sliding'),...
    ..." Fr('bearing')=", Fr('bearing')/;
* Almacenamiento en el parametro asociado de los valores optimos de x
    DataX(X)=VarX.l(X);
    XesK(IT,X)=DataX(X);
* Resolucion de los subproblemas
    loop(M,
* Valores iniciales de las variables u y v
        VarU.l(X)=DataX(X);
        VarV.l(A)=DataA(A);
        if(ORD(M) eq 1,
            put "----- VUELCO -----"/;
* Resolucion del subproblema asociado a rebase
            SOLVE mturning USING nlp MINIMIZING zc;
* Escritura de resultados
            put "beta=",zc.l:8:3, " modelstat= ",mturning.modelstat,",...
            ... solvestat= ",mturning.solvestat/;
            put " cpu=",mturning.resusd:9:5//;
            else
                if(ORD(M) eq 2,
                    put "----- DESLIZAMIENTO -----"/;
* Resolucion del subproblema asociado a deslizamiento
                    SOLVE msliding USING nlp MINIMIZING zc;
* Escritura de resultados
                    put "betas=",zc.l:8:3, " modelstat= ",msliding.modelstat,",...
                    ... solvestat= ",msliding.solvestat/;
                    put " cpu=",msliding.resusd:8:3//;
                    else
                        put "----- HUNDIMIENTO -----"/;
* Resolucion del subproblema asociado a hundimiento
                        SOLVE mbearing USING nlp MINIMIZING zc;
* Escritura de resultados
                        put "betas=",zc.l:8:3, " modelstat= ",mbearing.modelstat,",...
                        ... solvestat= ",mbearing.solvestat/;
                        put " cpu=",mbearing.resusd:8:3//;
                    );
                );
* Almacenamiento de los valores optimos de los indices de fiabilidad
* y las derivadas parciales con respecto x
                betaa(M)=zc.l;
                lambdaK(IT,M,X)=fixedX.m(X);
                betaK(IT,M)=zc.l;
            );
* Escritura de valores intermedios
        loop(M,
            put "beta('",M.tl:8,"')=",betaa(M):6:3," Fr('",M.tl:8,"')=",Fr(M):6:3/;
        );

```

```

* Calculo del error
  error=0.0;
  loop(M,
    if((abs(betaa(M)-betaux(M))/betaa(M))>error and betaa(M)>0,
      error=(abs(betaa(M)-betaux(M))/betaa(M));
    );
  );
* Escritura
  put "error(",iteration:5,")= ",error:10:7/;
* Acitvacion del hiperplano para las siguientes iteraciones
  FIRST(M)=yes;
  ITER(IT)=yes;
);
* Fin del procedimiento iterativo

```

A.6 GAMS Code for the Sensitivity Analysis

En esta sección se presentan los códigos GAMS para el estudio de las sensibilidades aplicados al ejemplo del muro. Dado que el estudio de sensibilidad sólo se aplica cuando el procedimiento ha convergido sólo se incluirá la parte de código que se necesita añadir a los programas presentados en las secciones A.4 y A.5.

Método de relajación. En el caso del método de relajación habría que añadir los siguientes elementos

```

PARAMETERS
* Puntos de maxima verosimilitud
  PMLU(M,X)
  PMLV(M,A)
* Parametros para almacenar las sensibilidades
  sensFclas(M)
  sensAclas(A)
  sensBclas(B)
  sensMclas(M)
  sensF(M,M)
  sensX(M,X)
  sensA(M,A)
  sensB(M,B);

VARIABLES
  VarA(A) variable auxiliar para los parametros y agentes
  VarB(B) variables auxiliar para los parametros estadisticos
  Zeta(M,B) variable auxiliar para las variables normales estandar
  Fa(M) variables auxiliares para los limites inferiores de los coeficientes...
  ... de seguridad;

EQUATIONS
* Ecuaciones del modelo para obtener las sensibilidades del problema maestro
* (clasico)

```

```

turningS, slidingS, bearingS, ZUsens, ZVsens, betasens
* Ecuaciones para las sensibilidades de los subproblemas
ZUdefS, ZVdefS;

* Ecuaciones del modelo para obtener las sensibilidades del problema maestro
* (clasico)
turningS..VarX('mua')*VarX('mua')*VarX('mub')*VarA('mugamma')=g=...
...2*VarA('muH')*VarA('muT')*Fa('turning');
slidingS..VarX('mua')*VarX('mub')*VarA('munu')*VarA('mugamma')=g=
...VarA('muT')*Fa('sliding');
bearingS..VarA('muS')=g=VarX('mub')*VarA('mugamma')*Fa('bearing');
ZUsens(M,X,B)$ (ORD(X)=ORD(B))..Zeta(M,B)=e=(PMLU(M,X)-VarX(X))/(VarB(B));
ZVsens(M,A,B)$ (ORD(A)+CARD(X)=ORD(B))..Zeta(M,B)=e=...
...(PMLV(M,A)-VarA(A))/(VarB(B));
betasens(M)..sqrt(sum(B,Zeta(M,B)*Zeta(M,B)))=g=betao(M);

* Ecuaciones para las sensibilidades de los subproblemas
ZUdefS(X,B)$ (ORD(X)=ORD(B))..Z(B)=e=(VarU(X)-VarX(X))/VarB(B);
ZVdefS(A,B)$ (ORD(A)+CARD(X)=ORD(B))..Z(B)=e=(VarV(A)-VarA(A))/(VarB(B));

* Planteamiento de los modelos de sensibilidad
MODEL classicS /ZClassic,turningS,slidingS,bearingS,geometric,ZUsens,ZVsens,...
...betasens/;
MODEL mturningS /Zbeta,ZUdefS,ZVdefS,Fturning/;
MODEL mslidingS /Zbeta,ZUdefS,ZVdefS,Fsliding/;
MODEL mbearingS /Zbeta,ZUdefS,ZVdefS,Fbearing/;

* Inicialización de los parametros de almacenamiento de los puntos de
* maxima verosimilitud (antes del comienzo del procedimiento iterativo)
PMLU(M,X)=0;
PMLV(M,A)=0;

* Almacenamiento de los puntos de maxima verosimilitud
* (en el loop M dentro del procedimiento iterativo, despues del calculo
* de los indices de fiabilidad)
PMLU(M,X)=VarU.l(X);
PMLV(M,A)=VarV.l(A);

* (Tras el procedimiento iterativo)
* Comienzo del estudio de sensibilidad
* Fijo las variables auxiliares a los valores de los paramentros
VarA.fx(A)=DataA(A);
VarB.fx(B)=DataB(B);
Fa.fx(M)=Flo(M);

* Inicializo los valores de las variables normales estandar
loop((M,X,B)$ (ORD(X)=ORD(B)),
Zeta.l(M,B)=(PMLU(M,X)-VarX.l(X))/(VarB.l(B));
);
loop((M,A,B)$ (ORD(A)+CARD(X)=ORD(B)),
Zeta.l(M,B)=(PMLV(M,A)-VarA.l(A))/(VarB.l(B));
);

```

```

);

* Resolucion del modelo clasico
SOLVE classicS USING nlp MINIMIZING cost;
* Escritura de resultados
put "cost=",cost.l:12:5, " modelstat= ",classicS.modelstat,",...
... solvestat= ",classicS.solvestat/;
put " cpu=",classicS.resusd:8:3//;
* Almacenamiento de las sensibilidades
sensFclas(M)=Fa.m(M);
sensAclas(A)=VarA.m(A);
sensBclas(B)=VarB.m(B);
sensMclas(M)=betasens.m(M);
* Escritura de las sensibilidades
put "CLASSIC SENSITIVITIES"/;
put "-----"/;
loop(M,
  put " Fa.m('",M.tl:8,"')",Fa.m(M):8:3/;
);
put /;
loop(A,
  put " VarA.m('",A.tl:8,"')",VarA.m(A):8:3/;
);
put /;
loop(B,
  put " VarB.m('",B.tl:8,"')",VarB.m(B):8:3/;
);
put /;
loop(M,
  put " betasens.m('",M.tl:8,"')",betasens.m(M):8:3/;
);
put /;
* Puesta a cero de las sensibilidades
Fa.m(M)=0;
VarA.m(A)=0;
VarB.m(B)=0;
betasens.m(M)=0;
* Fijo las variables de diseño al valor optimo
VarX.fx(X)=DataX(X);
* Ciclo para las sensibilidades de los modelos de fiabilidad
loop(M,
  VarU.l(X)=DataX(X);
  VarV.l(A)=DataA(A);
  if(ORD(M) eq 1,
* Obtencion del indice de fiabilidad para vuelco
  SOLVE mturningS USING nlp MINIMIZING zc;
* Escritura de resultados
  put "beta=",zc.l:12:5, " modelstat= ",mturningS.modelstat,",...
... solvestat= ",mturningS.solvestat/;
  put " cpu=",mturningS.resusd:8:3//;
  put "TURNING SENSITIVITIES"/;

```

```

        put "-----"/;
    else
        if(ORD(M) eq 2,
* Obtencion del indice de fiabilidad para deslizamiento
        SOLVE mslidingS USING nlp MINIMIZING zc;
* Escritura de resultados
        put "beta=",zc.l:12:5, " modelstat= ",mslidingS.modelstat,",...
... solvestat= ",mslidingS.solvestat/;
        put " cpu=",mslidingS.resusd:8:3//;
        put "SLIDING SENSITIVITIES"/;
        put "-----"/;
        else
* Obtencion del indice de fiabilidad para hundimiento
        SOLVE mbearingS USING nlp MINIMIZING zc;
* Escritura de resultados
        put "beta=",zc.l:12:5, " modelstat= ",mbearingS.modelstat,",...
... solvestat= ",mbearingS.solvestat/;
        put " cpu=",mbearingS.resusd:8:3//;
        put "BEARING SENSITIVITIES"/;
        put "-----"/;
    );
);
* Almacenamiento de las sensibilidades
sensX(M,X)=VarX.m(X);
sensA(M,A)=VarA.m(A);
sensB(M,B)=VarB.m(B);
* Escritura de las sensibilidades
loop(X,
    put " VarX.m('",X.tl:8,"')",VarX.m(X):8:3/;
);
put /;
loop(A,
    put " VarA.m('",A.tl:8,"')",VarA.m(A):8:3/;
);
put /;
loop(B,
    put " VarB.m('",B.tl:8,"')",VarB.m(B):8:3/;
);
put /;
* Puesta a cero de las sensibilidades
VarX.m(X)=0;
VarA.m(A)=0;
VarB.m(B)=0;
);

```

Método de aproximación por hiperplanos. En el caso del método de aproximación por hiperplanos habría que añadir los siguientes elementos:

PARAMETERS

* Pendientes de los hiperplanos aproximantes con respecto a parametros y agentes...

```

... y a paramentros estadisticos
    muK(M,A)
    deltaK(M,B)
* Parametros para almacenar las sensibilidades
    sensFclas(M)
    sensAclas(A)
    sensBclas(B)
    sensMclas(M)
    sensF(M,M)
    sensX(M,X)
    sensA(M,A)
    sensB(M,B);

VARIABLES
    VarA(A) variable auxiliar para los parametros y agentes
    VarB(B) variables auxiliar para los parametros estadisticos
    Zeta(M,B) variable auxiliar para las variables normales estandar
    Fa(M) variables auxiliares para los limites inferiores de los...
... coeficientes de seguridad;

EQUATIONS
* Ecuaciones del modelo para obtener las sensibilidades del problema maestro
* (clasico)
turningS,slidingS,bearingS,betadefS,fixedA(A),fixedB(B)
* Ecuaciones para las sensibilidades de los subproblemas
ZUdefS,ZVdefS;

* Ecuaciones del modelo para obtener las sensibilidades del problema maestro
* (clasico)
turningS..VarX('mua')*VarX('mua')*VarX('mub')*VarA('mugamma')=g=...
...2*VarA('muH')*VarA('muT')*Fa('turning');
slidingS..VarX('mua')*VarX('mub')*VarA('munu')*VarA('mugamma')=g=...
...VarA('muT')*Fa('sliding');
bearingS..VarA('muS')=g=VarX('mub')*VarA('mugamma')*Fa('bearing');
betadefS(ITER,M)$(FIRST(M))..beta(M)=1=betaK(ITER,M)+...
...sum(X,lambdaK(ITER,M,X)*(VarX(X)-XesK(ITER,X)))+...
...sum(A,muK(M,A)*(VarA(A)-DataA(A)))+sum(B,deltaK(M,B)*(VarB(B)-dataB(B)));

* Ecuaciones para las sensibilidades de los subproblemas
ZUdefS(X,B)$(ORD(X)=ORD(B))..Z(B)=e=(VarU(X)-VarX(X))/(VarB(B));
ZVdefS(A,B)$(ORD(A)+CARD(X)=ORD(B))..Z(B)=e=(VarV(A)-VarA(A))/(VarB(B));

* Planteamiento de los modelos de sensibilidad
MODEL mturningS /Zbeta,ZUdefS,ZVdefS,Fturning,fixedX/;
MODEL mslidingS /Zbeta,ZUdefS,ZVdefS,Fsliding,fixedX/;
MODEL mbearingS /Zbeta,ZUdefS,ZVdefS,Fbearing,fixedX/;
MODEL classicS /ZClassic,turningS,slidingS,bearingS,geometric,betadefS,auxbeta/;

* Comienzo del estudio de sensibilidad
* Fijo las variables auxiliares a los valores de los paramentros
VarA.fx(A)=DataA(A);

```

```

VarB.fx(B)=DataB(B);
* Inicializacion de parametros
muK(M,A)=0;
deltaK(M,B)=0;
* Ciclo para las sensibilidades de los modelos de fiabilidad
loop(M,
  VarU.l(X)=DataX(X);
  VarV.l(A)=DataA(A);
  if(ORD(M) eq 1,
* Obtencion del indice de fiabilidad para vuelco
  SOLVE mturningS USING nlp MINIMIZING zc;
* Escritura de resultados
  put "beta=",zc.l:12:5, " modelstat= ",mturningS.modelstat,",...
... solvestat= ",mturningS.solvestat/;
  put " cpu=",mturningS.resusd:8:3//;
  put "TURNING SENSITIVITIES"/;
  put "-----"/;
  else
  if(ORD(M) eq 2,
* Obtencion del indice de fiabilidad para deslizamiento
  SOLVE mslidingS USING nlp MINIMIZING zc;
* Escritura de resultados
  put "beta=",zc.l:12:5, " modelstat= ",mslidingS.modelstat,",...
... solvestat= ",mslidingS.solvestat/;
  put " cpu=",mslidingS.resusd:8:3//;
  put "SLIDING SENSITIVITIES"/;
  put "-----"/;
  else
* Obtencion del indice de fiabilidad para hundimiento
  SOLVE mbearingS USING nlp MINIMIZING zc;
* Escritura de resultados
  put "beta=",zc.l:12:5, " modelstat= ",mbearingS.modelstat,",...
... solvestat= ",mbearingS.solvestat/;
  put " cpu=",mbearingS.resusd:8:3//;
  put "BEARING SENSITIVITIES"/;
  put "-----"/;
  );
);
* Almacenamiento de las sensibilidades
muK(M,A)=VarA.m(A);
deltaK(M,B)=VarB.m(B);
sensX(M,X)=fixedX.m(X);
sensA(M,A)=VarA.m(A);
sensB(M,B)=VarB.m(B);
* Escritura de las sensibilidades
loop(X,
  put " VarX.m('",X.tl:8,"')",fixedX.m(X):8:3//;
);
put /;
loop(A,
  put " VarA.m('",A.tl:8,"')",VarA.m(A):8:3//;

```



```

);
put /;
loop(B,
    put " VarB.m('",B.tl:8,"')",VarB.m(B):8:3/;
);
put /;
* Puesta a cero de las sensibilidades
VarX.m(X)=0;
VarA.m(A)=0;
VarB.m(B)=0;
);
* Fijo las variables auxiliares a los valores de los limites
* inferiores de los coeficientes de seguridad
Fa.fx(M)=Flo(M);
* Resolucion del modelo clasico
SOLVE classicS USING nlp MINIMIZING zc;
* Escritura de resultados
put "z=",zc.l:8:3, " modelstat= ",classicS.modelstat,",...
... solvestat= ",classicS.solvestat/;
put " cpu=",classicS.resusd:8:3//;
* Almacenamiento de las sensibilidades
sensFclas(M)=Fa.m(M);
sensAclas(A)=VarA.m(A);
sensBclas(B)=VarB.m(B);
sensMclas(M)=auxbeta.m(M);
* Escritura de las sensibilidades
put "CLASSIC SENSITIVITIES"/;
put "-----"/;

loop(M,
    put " Fa.m('",M.tl:8,"')",Fa.m(M):8:3/;
);
put /;
loop(A,
    put " VarA.m('",A.tl:8,"')",VarA.m(A):8:3/;
);
put /;
loop(B,
    put " VarB.m('",B.tl:8,"')",VarB.m(B):8:3/;
);
put /;
loop(M,
    put " Auxbeta.m('",M.tl:8,"')",auxbeta.m(M):8:3/;
);
put /;
put ///;

```


Bibliography

- [1] ABADIE, J., AND CARPENTIER, J. Generalization of the wolfe reduced gradient method to the case of nonlinear constraints. In *Optimization*, R. Fletcher, Ed. Academic Press, New York, 1969, pp. 37–47.
- [2] ABDO, R., AND RACKWITZ, R. A new beta-point algorithm for large variable problems in time-invariant and time-variant reliability problems. In *Proceedings 3rd IFIP WG7.5 Working Conf. on Reliability and Optimization of Structural Systems (1990)*, A. Der Kiureghian and P. Thoft-Christensen, Eds., 112, Springer.
- [3] ABUR, A., AND EXPÓSITO, A. G. *Electric Power System State Estimation. Theory and Implementations*. Marcel Dekker, New York, 2004.
- [4] ACZÉL, J. Lectures on functional equations and their applications. In *Mathematics in Science and Engineering*, vol. 19. Academic Press, 1966.
- [5] ADELI, H. *Advances in design Optimization*. Chapman & Hall, New York, 1994.
- [6] ALONSO, E. E. Risk analysis of slopes and its application to slopes in canadian sensitive clays. *Geotechnique* 26, 3 (1976), 453–472.
- [7] ANG, A. H. S., SHINOZUKA, M., AND SCHUÉLLER, G. I. Structural safety and reliability. In *Proceedings ICOSSAR '89, 5th International Conference on Structural Safety and Reliability* (San Francisco, California, August 7–11 1989), vol. 1–3, ASCE. New York, 1990.
- [8] ARORA, S. Computational design optimization: A review and future directions. *Structural Safety* 7, 2–4 (1990), 131–148.
- [9] AUGUSTI, G., BORRI, A., AND VANNUCCHI, G. Applications of statistics and probability in soil and structural engineering. In *Proceedings ICASP* (Florence, Italy, June 13–17 1983), vol. 1–3, Pitagora Editrice. Bologna.
- [10] BAKER, M. J. Evaluation of partial safety factors for level i codes. Example of application of methods to reinforced concrete beams. In *Bulletin d'Information No. 112, Comité Européen du Béton* (Paris, 1976), vol. 112, pp. 190–211.
- [11] BAKER, R., AND GARBER, M. Variational approach to slope stability. In *Proceedings of the 9th International Conference on Soil Mechanics and Foundations Engineering* (Tokyo, 1977), vol. 2, pp. 9–12.

- [12] BAKER, R., AND GARBER, M. Theoretical analysis of the stability of slopes. *Geotechnique* 28, 4 (1978), 395–411.
- [13] BARABOSA, M. R., MORRIS, D. V., AND SARMA, S. K. Factor of safety and probability of failure of rockfill embankments. *Geotechnique* 39, 3 (1989), 471–483.
- [14] BASLER, B. Untersuchungen über den sicherheitsbegriff von bauwerken. *Schweiz. Arch.* 27, 4 (1961), 133–160.
- [15] BAZARAA, M. S., JARVIS, J. J., AND SHERALI, H. D. *Linear Programming and Network Flows*, second ed. John Wiley & Sons, New York, 1990.
- [16] BAZARAA, M. S., SHERALI, H. D., AND SHETTY, C. M. *Nonlinear Programming. Theory and Algorithms*, second ed. John Wiley & Sons, New York, 1993.
- [17] BENDERS, J. F. Partitioning procedures for solving mixed-variable programming problems. *Numerische Mathematik* 4 (1962), 238–252.
- [18] BERTSEKAS, D. P. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, 1982.
- [19] BISSCHOP, J., AND ROELOFS, M. *AIMMS, The User's guide*. Paragon Decision Technology, 1999.
- [20] BJERAGER, P. Probability integration by directional simulation. *J. Engineering Mechanics, ASCE* 114, 8 (1988), 1285–1302.
- [21] BJERAGER, P. On computational methods for structural reliability analysis. *Structural Safety* 9, 2 (1990), 79–96.
- [22] BJERAGER, P., AND KRENK, S. Parametric sensitivity in first order reliability theory. *J. Engineering Mechanics, ASCE* 115, 7 (1989), 1577–1582.
- [23] BLOCKLEY, D. I. *Engineering Safety*. McGraw-Hill, London, 1992.
- [24] BOX, G. E. P., AND MULLER, M. E. A note on the generation of normal deviates. *Ann. Math. Stat.* 29 (1958), 610–611.
- [25] BREITUNG, K. Asymptotic approximations for multinormal integrals. *J. Engrg. Mech.* 100, 3 (1984), 357–366.
- [26] BUCHER, C. G. Adaptive sampling, an iterative fast monte carlo procedure. *Structural Safety* 5, 2 (1988), 119–126.
- [27] CASAS, J. E. Reliability-based partial safety factors in cantilever construction of concrete bridges. *Journal of Structural Engineering* 123, 3 (1997), 305–312.
- [28] CASTILLO, E. *Extreme Value Theory in Engineering*. Academic Press, New York, 1988.

- [29] CASTILLO, E., COBO, A., JUBETE, F., AND PRUNEDA, R. E. *Orthogonal Sets and Polar Methods in Linear Algebra: Applications to Matrix Calculations, Systems of Equations and Inequalities, and Linear Programming*. John Wiley & Sons, New York, 1999.
- [30] CASTILLO, E., CONEJO, A. J., PEDREGAL, P., GARCÍA, R., AND ALGUACIL, N. *Building and Solving Mathematical Programming Models in Engineering and Science*. John Wiley & Sons Inc., New York, 2001. Pure and Applied Mathematics: A Wiley-Interscience Series of Texts, Monographs and Tracts.
- [31] CASTILLO, E., LOSADA, M., MÍNGUEZ, R., CASTILLO, C., AND BAQUERIZO, A. An optimal engineering design method that combines safety factors and failure probabilities: Application to rubble-mound breakwaters. *Journal of Waterways, Ports, Coastal and Ocean Engineering, ASCE* 130, 2 (2004), 77–88.
- [32] CASTILLO, E., AND LUCEÑO, A. One application of the calculus of variations to bearing capacity of foundations. In *Second International Conference on Applied Numerical Modeling* (Madrid, 1978), pp. 1–11.
- [33] CASTILLO, E., AND LUCEÑO, A. Application of the calculus of variations to the vertical cut off in cohesive frictionless soil (discussion). *Geotechnique* 30, 1 (1980), 1–16.
- [34] CASTILLO, E., AND LUCEÑO, A. Application of the calculus of variations to the vertical cut-off in cohesive frictionless soil (discussion by j. de jong). *Geotechnique* 30 (1981), 295–296.
- [35] CASTILLO, E., AND LUCEÑO, A. A critical analysis of some variational methods in slope stability analysis. *International Journal for Numerical and Analytical Methods in Geomechanics* 6 (1982), 195–209.
- [36] CASTILLO, E., AND LUCEÑO, A. Variational methods and the upper bound theorem. *Journal of Engineering Mechanics, ASCE* 109, 5 (1983), 1157–1174.
- [37] CASTILLO, E., AND REVILLA, J. The calculus of variations and the stability of slopes. In *Proceedings of the 9th International Conference on Soil Mechanics and Foundations Engineering* (Tokyo, 1977), vol. 2, pp. 25–30.
- [38] CASTILLO, E., AND RUÍZ, R. *Functional Equations in Science and Engineering*. Marcel Dekker, 1992.
- [39] CASTILLO, E., SOLARES, C., AND GÓMEZ, P. Estimating extreme probabilities using tail simulated data. *International Journal of Approximate Reasoning: official publication of the North American Fuzzy Information Processing Society* 17, 2-3 (1996), 163–190.
- [40] CASTILLO, E., SOLARES, C., AND GÓMEZ, P. Tail sensitivity analysis in bayesian networks. In *Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (UAI'96)* (Portland (Oregon), 1996), Morgan Kaufmann Publishers, San Francisco, California, pp. 133–140.

- [41] CASTILLO, E., SOLARES, C., AND GÓMEZ, P. High probability one-sided confidence intervals in reliability models. *Nuclear Science and Engineering* 126 (1997), 158–167.
- [42] CHATTERJEE, S., HADI, A. S., AND PRICE, B. *Regression Analysis by Example*. John Wiley & Sons, New York, 1990.
- [43] CHEN, W. F., AND GIGER, M. W. Limit analysis of stability of slopes. *Journal of the Soil Mechanics and Foundations Division, ASCE* (1971).
- [44] CHEN, W. F., AND SNITBHAN, N. On slip surface and slope stability analysis. *Soils and Foundations* 15 (1975), 41–49.
- [45] CHRISTIAN, J. T., LADD, C. C., AND BAECHER, G. B. Reliability applied to slope stability analysis. *Journal of Geotechnical Engineering Division, ASCE* 120, 12 (1994), 2180–2207.
- [46] COHEN, G. Auxiliary problem principle and decomposition of optimization problems. *Journal of Optimization Theory and Applications* 32, 3 (November 1980).
- [47] CONEJO, A. J., NOGALES, F. J., AND PRIETO, F. J. A decomposition procedure based on approximate Newton directions. *Math. Program. A* 93 (2002), 495–515.
- [48] CORNELL, C. A. A probability based structural code. *J. Amer. Concrete Inst.* 66, 12 (1969), 974–985.
- [49] CORNELL, C. A. First-order uncertainty analysis of soil deformation and stability. In *Proceedings of the 1st International Conference on Application of Statistics and Probability to Soil and Structural Engineering* (Hong Kong, 1971), pp. 129–144.
- [50] DAI, Y. F., G., D., AND STOLTE, W. J. A probabilistic slope stability analysis using deterministic computer software. In *Proceedings of the Conference on Probabilistic Methods in Geotechnical Engineering* (Canberra, Australia, February 1993), pp. 10–12.
- [51] DALHQUIST, G., AND BJÖRCK, A. *Numerical Methods*. Prentice-Hall, Englewoods, New York, 1974.
- [52] DAVIES, R. B. Numerical inversion of a characteristic function. *Biometrika* 60 (1973), 415–417.
- [53] DAVIS, P. J., AND RABINOWITZ, P. *Methods of Numerical Integration*. Academic Press, New York, 1975.
- [54] DER KIUREGHIAN, A. Bayesian analysis of model uncertainty in structural reliability. In *Proc. 3rd IFIP WG7.5 Conf. Reliability and Optimization of Structural Systems* (Berlin, 1990), Springer, pp. 211–221.
- [55] DER KIUREGHIAN, A., AND DAKESSIAN, T. Multiple design points in first and second-order reliability. *Structural Safety* 20, 1 (1998), 37–50.

- [56] DER KIUREGHIAN, A., AND DE STAFENO, M. Efficient algorithm for second-order reliability analysis. *J. Engineering Mechanics, ASCE* 117, 12 (1991), 2904–2923.
- [57] DER KIUREGHIAN, A., LIN, H. Z., AND HWANG, S. J. Second order reliability approximations. *Journal of Engineering Mechanics, ASCE* 113, 8 (1987), 1208–1225.
- [58] DER KIUREGHIAN, A., AND POLAK, E. Reliability-based optimal design: A decouple approach. In *Reliability and Optimization of Structural Systems*, A. S. Nowak, Ed. Book Crafters, Chelsea, Mich., 1998, pp. 197–205.
- [59] DER KIUREGHIAN, A., AND THOFT-CHRISTENSEN, P. Reliability and optimization of structural systems '90. In *Proceedings of the third IFIP WG 7.5 Conference on Reliability and Optimization of Structural Systems* (Berkeley, California, USA, March, 26–28 1990), vol. 61 of *Lecture Notes in Engineering*, Springer-Verlag, Berlin.
- [60] DITLEVSEN, O. Structural reliability and the invariance problem. Solid Mechanics Report 22, University of Waterloo, 1973.
- [61] DITLEVSEN, O. Principle of normal tail approximation. *J. Engineering Mechanics Div., ASCE* 107, 6 (1981), 1191–1208.
- [62] DITLEVSEN, O. *Uncertainty modeling with applications to multidimensional engineering systems*. McGraw-Hill Book Co., New York, 1981.
- [63] DITLEVSEN, O., AND BJERAGER, P. Plastic reliability analysis by directional simulation. *J. Engineering Mechanics Div., ASCE* 115, 6 (1989), 1347–1362.
- [64] DITLEVSEN, O., AND MADSEN, H. O. *Structural reliability methods*. Wiley, Chichester, New York, 1996.
- [65] DITLEVSEN, O., OLESEN, R., AND MOHR, G. Solution of a class of load combination problems by directional simulation. *Structural Safety* 4 (1987), 95–109.
- [66] DOLINSKY, K. First order second-moment approximation in reliability of structural systems: critical review and alternative approach. *Structural Safety* 1, 3 (1983), 211–231.
- [67] ELLINWOOD, B., GALAMBOS, T. V., MACGREGOR, J. C., AND CORNELL, C. A. Development of a probability based load criteria for american national standard A58: Building code requirements for minimum design loads in buildings and other structures. In *NBS Special Publication* (Washington, DC., 1980), no. 577, National Bureau of Standards, US Department of Commerce.
- [68] ENEVOLDSEN, I. *Reliability-based structural optimization*. PhD thesis, University of Aalborg, Aalborg, Denmark, 1991.
- [69] ENEVOLDSEN, I. Sensitivity analysis of reliability-based optimal solutions. *Journal of Engineering Mechanics* 120, 1 (1994), 198–205.

- [70] ENEVOLDSEN, I., AND SORENSEN, J. D. Reliability-based optimization of series systems of parallel systems. *Journal of Structural Engineering* 119, 4 (1993), 1069–1084.
- [71] ENEVOLDSEN, I., AND SORENSEN, J. D. Reliability-based optimization in structural engineering. *Structural Safety* 15 (1994), 169–196.
- [72] ENGELUND, S., AND RACKWITZ, R. A benchmark study on importance sampling techniques in structural reliability. *Structural Safety* 12, 4 (1993), 255–276.
- [73] ESTEVA, L. Y RUÍZ, S. E. Proceedings icasp 6. In *Sixth International Conference on Applications of Statistics and Probability in Civil Engineering* (Mexico City, June 17–21 1991), vol. 1–3.
- [74] Eurocode, 1991. Part 1.
- [75] FIELD, C. Tail areas of linear combinations of chi-squares and non-central chi-squares. *Journal of Statistical Computation and Simulation* 15 (1996), 243–248.
- [76] FOURER, R., GAY, D. M., AND KERNIGHAN, B. W. *AMPL: A Modelling Language for Mathematical Programming*. Boyd & Fraser publishing company, Massachusetts, 1993.
- [77] FRANGOPOL, D. M. Sensitivity of reliability-based optimum design. *J. Struct. Engrg., ASCE* 111, 8 (1985), 1703–1721.
- [78] FRANGOPOL, D. M. Reliability-based optimum structural design. In *Probabilistic Structural Mechanics Handbook*, C. Sundarajan, Ed. Chapman & Hall, New York, 1995, ch. 16, pp. 352–387.
- [79] FRANGOPOL, D. M., AND HENDAWI, S. Incorporation of corrosion effects in reliability-based optimization of composite hybrid plate girders. *Structural Safety* 16, 1+2 (1994), 145–169.
- [80] FREEMAN, H. *An Introduction to Statistical Inference*. Addison-Wesley, Reading, MA, 1963.
- [81] FREUDENTHAL, A. M. Safety and the probability of structural failure. *Transactions, ASCE* 121 (1956), 1337–1397.
- [82] GABRIELE, G. A., AND RAGSDALL, K. The generalized reduced gradient method: A reliable tool for optimal design. *ASME J. Engin. Ind.* 99, 2 (May 1977), 394–400.
- [83] GALAMBOS, J. *The Asymptotic Theory of Extreme Order Statistics*. Robert E. Krieger Publishing Company, Malabar, Florida, 1987.
- [84] GARBER, M., AND BAKER, R. Bearing capacity by variational method. *Journal of the Geotechnical Engineering Division, ASCE* 103 (1977), 1209–1225.

- [85] GASSER, M., AND SCHUËLLER, G. I. Some basic principles in reliability-based optimization (rbo) of structures and mechanical components. In *Stochastic programming methods and technical applications*, K. Marti and P. Kall, Eds. Springer-Verlag, Berlin, Germany, 1998, pp. 11–103. Lecture Notes in Economics and Mathematical Systems 458.
- [86] GEOFFRION, A. M. Generalized benders decomposition. *JOTA* 10, 4 (1972), 237–260.
- [87] GILL, P. E., MURRAY, W., AND WRIGHT, M. H. *Practical Optimization*. Academic Press, New York, 91.
- [88] GIVENS, G. H., AND RAFTERY, A. E. Local adaptive importance sampling for multivariate densities with strong nonlinear relationships. *Journal of the American Statistical Association* 91, 433 (1996), 132–141.
- [89] GOLUB, G. H. AND VAN LOAN, C. F. *Matrix Computations*, third ed. The Johns Hopkins University Press, Baltimore and London, 1996.
- [90] HARBITZ, A. An efficient sampling method for probability of failure calculation. *Structural Safety* 3, 2 (1986), 109–115.
- [91] HASKIN, F. E., STAPLE, B. D., AND DING, C. Efficient uncertainty analyses using fast probability integration. *Nuclear Engineering and Design* 166, 2 (1996), 225–248.
- [92] HASOFER, A. M., AND LIND, N. C. Exact and invariant second moment code format. *J. Engrg. Mech.* 100, EM1 (1974), 111–121.
- [93] HESTERBERG, T. Weighted average importance sampling and defensive mixture distributions. *Technometrics* 37, 2 (1995), 185–194.
- [94] HEYMAN, J. *Plastic design of frames*, vol. 2. Cambridge University Press, 1971.
- [95] HOHENBICHLER, M., AND RACKWITZ, R. Non-normal dependent vectors in structural safety. *J. Engineering Mechanics Div., ASCE* 107, 6 (1981), 1227–1238.
- [96] HOHENBICHLER, M., AND RACKWITZ, R. Sensitivity and importance measures in structural reliability. *Civil Engineering Systems* 3, 4 (1986), 203–209.
- [97] HORNE, M. R., AND H., P. P. Commentary on the level 2 procedure, rationalization of safety and serviceability factors in structural codes. Report 63, Construction Industry Research and Information Association, London, 1977.
- [98] JANBU, N. Application of composite slip surfaces for stability analysis. In *Proc. European Conf. Stability of Earth Slopes* (Stockholm, 1954), vol. 3, pp. 43–49.
- [99] KARAMCHANDANI, A., BJERAGER, P., AND CORNELL, C. A. Adaptive importance sampling. In *Proceedings 5th International Conference on Structural Safety and Reliability* (San Francisco, 1989), ASCE, pp. 855–862.
- [100] KARAMCHANDANI, A., AND CORNELL, C. Sensitivity estimation within first and second order reliability methods. *Structural Safety* 11, 1 (1994), 59–74.

- [101] KATSUKI, S., AND FRANGOPOL, D. M. Hyperspace division method for structural reliability. *Journal of Engineering Mechanics, ASCE* 120, 11 (1994), 2405–2427.
- [102] KELLEY, J. E. The cutting plane method for solving convex programs. *J. Soc. Ind. and Appl. Math.* 8, 4 (1960), 703–712.
- [103] KIRJNER-NETO, C., POLAK, E., AND DER KIUREGHIAN, A. An outer approximations approach to reliability-based optimal design of structures. *J. Optimization Theory and Application* 98, 1 (1998), 1–17.
- [104] KONISHI, I., ANG, A. H.-S., AND SHINOZUKA, M. Structural safety and reliability. In *Proceedings 4th International Conference on Structural Safety and Reliability, IASSAR* (Kobe, Japan, May 10–13 1985), vol. I–III.
- [105] KOYLUOGLU, H. U., AND NIELSEN, R. K. New approximations for sorm integrals. *Structural Safety* 13 (1994), 235–246.
- [106] KUSCHEL, N., AND RACKWITZ, R. Optimal design under time-variant reliability constraints. *Structural Safety* 22, 2 (2000), 113–128.
- [107] LEMAIRE. Applications of statistics and probability. In *Proceedings ICASP 7* (Paris, France, JULY 10–13 1995), vol. 1–2, Balkema, Rotterdam.
- [108] LI, K. S., AND LUMB, P. Probabilistic design of slopes. *Can. Geotech. J.* 24 (1987), 520–535.
- [109] LIN, Y. K. Reliability and risk analysis in civil engineering. In *Proceedings 6th Specialty Conference on Probabilistic Mechanics and Structural and Geotechnical Reliability* (Denver, Colorado, July, 8–10 1992), Y. K. Lin, Ed., ASCE, New York.
- [110] LIND, N. C. Application to design of level i codes. In *Bulletin d'Information No. 112, Comite Européen due Béton* (Paris, 1976), vol. 112, pp. 73–89.
- [111] LIND, N. C. Reliability and risk analysis in civil engineering. In *Proceedings ICASP 5, Fifth International Conference on Applications of Statistics and Probability in Civil Engineering* (University of British, Columbia, Vancouver, May 25–29 1987), University of Waterloo, Waterloo, Canada.
- [112] LIU, P.-L., AND DER KIUREGHIAN, A. Optimization algorithms for structural reliability. *Structural Safety* 9 (1991), 161–177.
- [113] LOW, B. K., AND TANG, W. H. Efficient reliability evaluation using spreadsheet. *Journal of Engineering Mechanics, ASCE* 123, 7 (1994), 749–752.
- [114] LUCEÑO, A. *Análisis de los métodos variacionales aplicados a los problemas de estabilidad en mecánica del suelo. Utilización del teorema de la cota superior.* PhD thesis, University of Cantabria, Santander, Spain, 1979.
- [115] LUENBERGER, D. G. *Linear and Nonlinear Programming*, second ed. Addison-Wesley, Reading, Massachusetts, 1984.

- [116] MADSEN, H. D., KRENK, S., AND LIND, N. C. *Methods of structural safety*, second ed. Prentice Hall, Inc., Englewood Cliffs, New York, 1986.
- [117] MADSEN, H. O., AND FRIIS HANSEN, P. A comparison of some algorithms for reliability-based structural optimization and sensitivity analysis. In *Reliability and Optimization of Structural Systems*, R. Rackwitz and P. Thoft-Christensen, Eds. Springer, Berlin, 1992, pp. 443–451.
- [118] MALKAWI, A. I. H., HASSAN, W. F., AND ABDULLA, F. A. Uncertainty and reliability analysis applied to slope stability. *Structural Safety* 22 (2000), 161–187.
- [119] MARTI, K. Solving stochastic structural optimization problems by rsm-based stochastic approximation methods: Gradient estimation in case of intermediate variables. *Mathematical Methods of Operations Res.* 46 (1997), 409–434.
- [120] MAYER, H. *Die Sicherheit der Bauwerke*. Springer, Berlin, 1926.
- [121] MELCHERS, R. E. Improved importance sampling for structural reliability calculation. In *Proceedings 5th International Conference on Structural Safety and Reliability* (New York, 1989), ASCE, pp. 1185–1192.
- [122] MELCHERS, R. E. Simulation in time-invariant and time-variant reliability problems. In *Proceedings 4th IFIP Conference on Reliability and Optimization of Structural Systems* (Berlin, 1991), Springer, pp. 39–82.
- [123] MELCHERS, R. E. *Structural reliability analysis and prediction*, second ed. John Wiley & Sons, New York, 1999.
- [124] MELCHERS, R. E., AND LI, C. Q. Discussion of engelund, s. and rackwitz, r. *Structural Safety* 14, 4 (1993), 299–302.
- [125] MÍNGUEZ, R. *Seguridad, Fiabilidad y Análisis de Sensibilidad en Obras de Ingeniería Civil Mediante Técnicas de Optimización por Descomposición. Aplicaciones*. PhD thesis, Universidad de Cantabria, Santander, España, 2003.
- [126] MÍNGUEZ, R., CASTILLO, E., AND HADI, A. S. Solving the inverse reliability problem using decomposition techniques. *Structural Safety* 27 (2005), 1–23.
- [127] MUROTSU, Y., KISHI, M., OKADA, H., YONEZAWA, M., AND TAGUCHI, K. Probabilistic optimum design of frame structure. In *11th IFIP Conference on System Modeling and Optimization* (Berlin, 1984), Springer-Verlag, pp. 545–554.
- [128] NATAF, A. Détermination des distribution de probabilités dont les marges sont données. *Comptes Rendus de l'Academie des Sciences* 225 (1962), 42–43.
- [129] <http://www.aimms.com/>.
- [130] <http://www.dashopt.com>.
- [131] <http://www.gams.com>.

- [132] <http://www.lindo.com>.
- [133] <http://www.maximal-usa.com>.
- [134] PAPADIMITRIOU, C. BECK, J. L., AND KATAFYGIOTIS, L. S. Asymptotic expansions for reliability and moments of uncertain systems. *Journal of Engineering Mechanics, ASCE* 123, 12 (1997), 1219–1229.
- [135] PARKINSON, D. B. Computer solution for the reliability index. *Engineering Structures* 2 (1980), 57–62.
- [136] POLAK, E. *Optimization: Algorithms and Consistent Approximations*. Springer, New York, 1997.
- [137] POLYAK, B. T. *Introduction to Optimization*. Optimization Software, New York, 1987.
- [138] POWELL, M. J. D. Vmcwd: A fortran subroutine for constrained optimization. Report DAMPT 1982/NA4, Cambridge University, United Kingdom, 1982.
- [139] RACKWITZ, R., AUGUSTI, G., AND BORRI, A. Reliability and optimization of structural systems. In *Proceedings of the Sixth IFIP WG 7.5 Conference on Reliability and Optimization of Structural Systems* (Assisi, Perugia, Italy, September, 7–9 1994), Chapman & Hall, London.
- [140] RACKWITZ, R., AND FIESSLER, B. Structural reliability under combined load sequences. *Comput. Struct.* 9 (1978), 489–494.
- [141] RACKWITZ, R., AND THOFT-CHRISTENSEN, P. Reliability and optimization of structural systems ‘91. In *Proceedings of the fourth IFIP WG 7.5 Conference on Reliability and Optimization of Structural Systems* (Munich, Germany, September, 11–13 1991), vol. 76 of *Lecture Notes in Engineering*, Springer-Verlag, Berlin.
- [142] REVILLA, J., AND CASTILLO, E. The calculus of variations applied to stability of slopes. *Geotechnique* 27, 1 (1977), 1–1.
- [143] ROM 0.0. Procedimiento general y bases de cálculo en el proyecto de obras marítimas y portuarias. Tech. rep., Puertos del Estado, Madrid, España, Noviembre 2001. pp 245.
- [144] ROOS, C., TERLAKY, T., AND VIAL, J. P. *Theory and Algorithms for Linear Optimization: An Interior Point Approach*. John Wiley & Sons, Chichester, UK, 1997. Wiley-Interscience Series in Discrete Mathematics and Optimization.
- [145] ROSEN, J. B. The gradient projection method for nonlinear programming, part i, linear constraints. *SIAM J. Appl. Math.* 8 (1960), 181–217.
- [146] ROSENBLATT, M. Remarks on a multivariate transformation. *Ann. Math. Stat.* 23, 3 (1952), 470–472.

- [147] ROUSSEEUW, P. J., AND LEROY, A. M. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, 1987.
- [148] ROWE, W. D. *An Anatomy of Risk*. John Wiley & Sons, New York, 1977.
- [149] ROYSET, J. O., DER KIUREGHIAN, A., AND POLAK, E. Reliability-based optimal design of series structural systems. *Journal of Engineering Mechanics, ASCE* 127, 6 (2001), 607–614.
- [150] RUBINSTEIN, B. V. *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York, 1981.
- [151] RZHANITZYN, R. It is necessary to improve the standards of design of building structures. a statistical method of design of building structures. Technical Translation, by Allan D. E. 1368, National Research Council of Canada, Ottawa, 1957.
- [152] RZHANITZYN, R. Risk: analysis, perception and management. Tech. rep., Royal Society Study Group, London, 1991.
- [153] SCHITKOWSKI, K. Nlpql: A fortran subroutine solving constrained non-linear programming problems. *Ann. Oper. Res.* 5 (1985), 485–500.
- [154] SCHUËLLER, G. I., SHINOZUKA, M., AND YAO, J. T. P. Structural safety and reliability. In *Proceedings ICOSSAR '93, 6th International Conference on Structural Safety and Reliability* (Innsbruck, Austria, August 9–13 1993), vol. 1–3, Balkema. Rotterdam, 1994.
- [155] SHINOZUKA, M. Stochastic methods in structural dynamics. Stochastic fields and their digital simulations, Martinus Nijhoff, The Hague, 1987.
- [156] SIEGMUND, D. Importance sampling in the monte carlo study of sequential tests. *The Annals of Statistics* 4 (1976), 673–684.
- [157] SMICHT, L. A. Structural design by systematic synthesis. In *Proceedings of the Second ASCE Conference on Electronic Computation* (Pittsburgh, Pennsylvania, 1960), pp. 105–122.
- [158] SORENSEN, J. D. Reliability-based optimization of structural elements. Structural reliability theory, The Uni. of Aalborg, Aalborg, Denmark, 1986. Paper No. 18.
- [159] SORENSEN, J. D. Reliability-based optimization of structural systems. Structural reliability theory, The Uni. of Aalborg, Aalborg, Denmark, 1987. Paper No. 32.
- [160] SORENSEN, J. D., AND ENEVOLDSEN, I. Sensitivity analysis in reliability-based shape optimization. In *NATO ASI* (Dordrecht, The Netherlands, 1992), B. H. V. Topping, Ed., vol. 221 of *E*, Kluwer Academic Publishers, pp. 617–638.
- [161] SPANOS, P. D. Probabilistic methods in civil engineering. In *Proceedings 5th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability* (Blacksburg, Virginia, May, 25–27 1988), P. D. Spanos, Ed., ASCE, New York.

- [162] STEWART, M. G., AND MELCHERS, R. E. *Probabilistic Risk Assessment of Engineering Systems*. Chapman & Hall, London, 1997.
- [163] TANG, W. H., YÜCEMEN, M. S., AND ANG, A. H. S. Probability-based short term design of soil slopes. *Canadian Geotechnical Journal* 13, 3 (1976), 201–215.
- [164] THOFT-CHRISTENSEN, P. Reliability and optimization of structural systems. In *Proceedings of the first IFIP WG 7.5 Conference on Reliability and Optimization of Structural Systems* (Aalborg, Denmark, May 1987), vol. 33 of *Lecture Notes in Engineering*, Springer-Verlag, Berlin.
- [165] THOFT-CHRISTENSEN, P. Reliability and optimization of structural systems '88. In *Proceedings of the second IFIP WG 7.5 Conference on Reliability and Optimization of Structural Systems* (London, UK, September, 26–28 1988), vol. 48 of *Lecture Notes in Engineering*, Springer-Verlag, Berlin.
- [166] THOFT-CHRISTENSEN, P., AND ISHIKAWA, H. Reliability and optimization of structural systems '91. In *Proceedings of the fifth IFIP WG 7.5 Conference on Reliability and Optimization of Structural Systems* (Takamatsu-shi, Kagawa, Japan, March, 24–26 1993), *Lecture Notes in Engineering*, North-Holland, Amsterdam.
- [167] TVEDT, L. Two second-order approximations to the failure probability. Veritas rep. rdiv/20-004083, Der norske Veritas, Oslo, 1983.
- [168] TVEDT, L. Distribution of quadratic forms in normal space. *Journal of Engineering Mechanics* 116, 6 (1990), 1183–1197.
- [169] VANDERPLAATS, G. N. *Numerical Optimization Techniques for Engineering Design*. McGraw-Hill, New York, 1984.
- [170] VENECIANO, D. Contributions to second moment reliability. Res. Rep. R74-33, Dep. of Civ. Engrg., Mass. Inst. of Technol., Mass, 1974.
- [171] VENMARCKE, E. H. Reliability of earth slopes. *Journal of Geotechnical Engineering Division, ASCE* 103, 11 (1997), 1227–1246.
- [172] VROUWENVELDER, T. The jcss probabilistic model code. *Structural Safety* 19, 3 (1997), 245–251.
- [173] VROUWENVELDER, T. *Reliability Based Code calibration. The use of the JCSS Probabilistic Model Code.*, march 21/22 ed. Joint Committee of Structural Safety Workshop on Code Calibration, Zurich, 2002.
- [174] WEN, Y.-K. Probabilistic mechanics and structural reliability. In *Proceedings 4th ASCE Specialty Conference on Probabilistic Mechanics and Structural Reliability* (Berkeley, California, January, 11–13 1984), Y.-K. Wen, Ed., ASCE, New York.
- [175] WIRSCHING, P. H., AND WU, Y. T. Advanced reliability methods for structural evaluation. *J. Engineering Mechanics Division, ASCE* 109, 2 (1987), 19–23.

- [176] WOLFE, P. Methods of nonlinear programming. In *Recent Advances in Mathematical Programming*, R. L. Graves and P. Wolfe, Eds. McGraw-Hill, New York, 1963, pp. 76–77.
- [177] WOLFF, T. F. *Analysis and design of embankment dam slopes: a probabilistic approach*. PhD thesis, Purdue University, West Lafayette, 1985.
- [178] WU, T. H., AND KRAFT, L. M. Safety analysis of slopes. *Journal of Soil Mechanics and Foundation Division, ASCE* 96, 2 (1970), 609–630.
- [179] WU, Y. T., BURNSIDE, O. H., AND CRUSE, T. A. Probabilistic methods for structural response analysis. In *Computational Mechanics of Probabilistic and Reliability Analysis* (Annandale, Va, 1989), W. K. Liu and T. Belytschko, Eds., Elmepress International, pp. 181–196.
- [180] ZHANG, P. Nonparametric importance sampling. *Journal of the American Statistical Association* 91, 435 (1996), 1245–1253.
- [181] ZHAO, Y. G., AND ONO, T. A general procedure for first/order reliability method (form/sorm). *Structural Safety* 21 (1999), 95–112.