Cost-Benefit Based Maintenance Optimization for Deteriorating Structures

Dissertation
zur Erlangung des akademischen Grades
Doktor–Ingenieur

an der Fakultät Bauingenieurwesen
der Bauhaus-Universität Weimar

vorgelegt von

Shoko Higuchi
aus Yokohama, Japan

Weimar

Gutachter:
1. Prof. Dr. techn. Christian Bucher
2. Prof. Dr. Tsuyoshi Takada
3. Prof. Dr.-Ing. Dipl.-Wirtsch.-Ing. Hans Wilhelm Alfen

Tag der Disputation: 11. April 2008

Acknowledgment

First and foremost, I would like to express my deepest sense of gratitude to my supervisor Prof. Christian Bucher for his guidance and excellent advice throughout this study. I would like to express my gratitude also to Prof. Tsuyoshi Takada from University of Tokyo for his advice and heart-full encouragement during his visits to ISM. Special thanks go to Prof. Hans Wilhelm Alfen for undertaking the third review.

My sincere thanks are extended to Dr. Michael Macke for his continuous guidance, outstanding advice and encouragement throughout this thesis. I also thank all my colleagues at ISM for sharing experiences and knowledge during the time of my study.

Finally, I take this opportunity to express my profound gratitude to my mother for her support and patience during my stay at Bauhaus-University Weimar.

Abstract

In recent years increasingly consideration has been given to the lifetime extension of existing structures. This is based on the fact that a growing percentage of civil infrastructure as well as buildings is threatened by obsolescence and that due to simple monetary reasons this can no longer be countered by simply re-building everything anew. Hence maintenance interventions are required which allow partial or complete structural rehabilitation. However, maintenance interventions have to be economically reasonable, that is, maintenance expenditures have to be outweighed by expected future benefits. Is this not the case, then indeed the structure is obsolete—at least in its current functional, economic, technical, or social configuration—and innovative alternatives have to be evaluated.

An optimization formulation for planning maintenance interventions based on cost-benefit criteria is proposed herein. The underlying formulation is as follows: (a) between maintenance interventions structural deterioration is described as a random process; (b) maintenance interventions can take place anytime throughout lifetime and comprise the rehabilitation of all deterioration states above a certain minimum level; and (c) maintenance interventions are optimized by taking into account all expected life-cycle costs (construction, failure, inspection and state-dependent repair costs) as well as state- or time-dependent benefit rates. The optimization is performed by an evolutionary algorithm. The proposed approach also allows to determine optimal lifetimes and acceptable failure rates.

Numerical examples demonstrate the importance of defining benefit rates explicitly. It is shown, that the optimal solution to maintenance interventions requires to take action before reaching the acceptable failure rate or the zero expected net benefit rate level. Deferring decisions with respect to maintenance not only results, in general, in higher losses, but also results in overly hazardous structures.

Kurzfassung

Die Verlängerung der Nutzungsdauer bestehender Tragwerke hat in den letzten Jahren zunehmend an Bedeutung gewonnen. Dies liegt in der Tatsache begründet, dass ein nicht unerheblicher Anteil der Infrastruktur wie auch an Gebäuden durch Überalterung bedroht ist, und dass es aus rein wirtschaftlichen Gründen nicht länger möglich ist diesen Zustand durch Neubau zu entgegnen. Es sind also Instandhaltungsstrategien notwendig, die eine teilweise oder vollständige Revitalisierung von Tragwerken erlauben. Allerdings müssen diese Instandhaltungsstrategien auch einen volkswirtschaftlichen Sinn haben, das heißt die entsprechenden Aufwendungen müssen durch einen zukünftig zu erwartenden Nutzen aufgewogen werden. Ist dies nicht der Fall, so sind die Tragwerke in der Tat veraltet—zumindest in ihrer momentanen funktionellen, wirtschaftlichen, technischen oder gesellschaftlichen Bedeutung—und Alternativvorschläge müssen untersucht werden.

In dieser Arbeit wird die Planung von Instandhaltungsmaßnahmen als Optimierungsaufgabe unter Verwendung von Kosten-Nutzen-Kriterien formuliert. Die zugrunde liegende Beschreibung ist wie folgt: (a) die Abnahme der Tragfähigkeit zwischen den Instandhaltungsmaßnahmen wird als Zufallsprozess beschrieben; (b) die Instandhaltungsmaßnahmen können jederzeit während der Nutzungsdauer stattfinden und bestehen in der Reparatur von Schadenszuständen eines gewissen Niveaus; (c) die Instandhaltungsmaßnahmen werden hinsichtlich aller Lebensdauerkosten (Errichtungs-, Versagens-, Inspektions- und schadensabhängiger Reparaturkosten) sowie des zustands- und zeitabhängigen Nutzens optimiert. Die Optimierung erfolgt mit Hilfe eines evolutionären Algorithmus. Die vorgeschlagene Formulierung erlaubt darüber hinaus auch die Bestimmung von optimalen Nutzungsdauern und zulässigen Versagensraten.

Die Rechenbeispiele weisen die Bedeutung einer expliziten Ausweisung des Nutzens aus. Es wird gezeigt, dass eine optimale Strategie für Instandhaltungsmaßnahmen ein aktiv werden vor Erreichen zulässiger Versagensraten oder dem Verschwinden des Nettonutzens je Zeiteinheit erfordert. Das Aufschieben von Entscheidungen bezüglich der Durchführung von Instandhaltungsmaßnahmen zieht in der Regel nicht nur höhere Folgekosten nach sich, sondern resultiert auch in Tragwerke mit unzulässig hohem Gefährdungspotential.

Contents

Basic Notation

a	transition rate, alternative (Ch. 3)
$\mathbf{A}$	transition rate matrix
$\mathcal{A}$	set of feasible acts
b, B	benefit
c	cost
C	cost, damage growth parameter (Ch. 5)
$\mathbf{C}$	rehabilitation matrix
d	damage levels (Ch. 5)
D	event of detection, damage (Ch. 5)
$\mathbf{D}$	inspection matrix
e	equality constraint
E	elastic modulus
$\mathrm{E}[\cdot]$	expected value
f	probability density function, fitness function (Ch. 4)
F	cumulative distribution function, failure event
g	objective function, net present benefit function
$\mathcal{G}$	set of design vectors
h	failure rate, inequality constraint (Ch. 2), rate function (Ch. 5)
$\mathbf{H}$	transformation matrix
i	index
I	inspection costs
$\mathbf{I}$	identity matrix
j	index
k	index
l	loss, length (Ch. 5), index
L	loss, failure cost
$\mathcal{L}$	set of design vectors
m	maintenance cost, number of states, renewal density (Ch. 2)
$\mathcal{M}$	set of design vectors
n	number of maintenance interventions, index
N	number of events, number of selections
$\mathbf{N}$	nilpotent matrix

p	probability, coefficient (Ch. 5)
$\mathbf{p}$	probability vector
$\mathbf{P}$	transition matrix
$\mathrm{Pr}(\cdot), \mathrm{Pr}(\cdot\|\cdot)$	probability, conditional probability
$\mathrm{PV}[\cdot]$	present value
q	coefficient, index
$\mathbf{Q}$	rehabilitation quality matrix
r	index
$R, \mathbf{R}$	rehabilitation cost, rehabilitation cost vector
$\mathbb{R}$	set of real numbers
s	string, length, index
$\mathcal{S}$	subset
$t, \mathbf{t}, T$	time, lifetime
u	utility function, sample of uniform random number (Ch. 4)
U	uniform random number
v	value function
V	volume, vertical loads (Ch. 5)
$\mathrm{Var}[\cdot]$	variance
W	strain energy
$x, \mathbf{x}$	design variables, consequences (Ch. 3)
$\alpha, \boldsymbol{\alpha}$	quality of inspection
β	reliability index
γ	discount rate
$\Delta, \boldsymbol{\Delta}$	rehabilitation level
ε	small parameter (Ch. 2), strain (Ch. 5)
$\boldsymbol{\zeta}$	'forcing' term
λ	occurrence rate, parameter (Ch. 4)
$\boldsymbol{\Lambda}$	Jordan block
μ	mean time (Ch. 2), parameter (Ch. 4)
ν	parameter
ξ	parameter
$\boldsymbol{\pi}$	probability
ϱ	mass density
$\sigma, \Delta\sigma$	stress, stress amplitude
Φ	normal cumulative distribution function
$\boldsymbol{\Psi}$	fundamental matrix
ω	impact (Ch. 3), eigenvalue (Ch. 5)
$\dot{(\cdot)}$	time derivative
$[\cdot]^*$	optimal, Laplace transform (Ch. 2)
$[\cdot]^{\mathrm{T}}$	transposed
$[\cdot]^{-1}$	inverse

1

Introduction

Structural and architectural engineers primarily design and analyze structures and structural components. They ensure that these structures satisfy a given design intent with respect to safety standards, that is, that the risk of structural collapse is below an acceptable level, and with respect to serviceability (for example, that the floor vibration and sway of buildings does not discomfort its occupants). However, structural engineers have not only to fulfill technical constraints, but also have to take into account on an ever increasing scale economical, environmental, esthetic and social aspects. That is, they are responsible for making efficient use of different types of resources to provide the best possible performance of structures throughout the (structural) lifetime. In fact, the life of a building is sometimes compared with the human body. As Levy and Salvadori (2002) write in the introduction to *Why Buildings Fall Down: How Structures Fail*:

> "A building is conceived when designed, born when built, alive while standing, dead from old age or unexpected accident. It breathes through the mouth of its windows and the lungs of its air-conditioning system. It circulates fluids through the veins and arteries of its pipes and sends messages to all parts of its body through the nervous system of its electric wires. A building reacts to changes in its outer or inner conditions through its brain of feedback systems, is protected by skin of its facade, supported by the skeleton of its columns, beams, and slabs, and rests on the feet of its foundations. Like most human bodies, most buildings have full lives, and then they die."

Clearly, the mission of an engineer is not only to design and construct (to conceive and give birth), but also to maintain the best performance of all structural and non-structural components throughout the entire life-cycle of a structure.

The decision to construct a building or not is, in general, based on the prospect that expected future benefits will outweigh expected costs. While building and civil engineering structures mostly show a high profitability after construction, this profitability quite often gradually diminishes with time, that is, the expected benefit per time decreases. The decrease may have its origin, for example, in the deteriorating performance of the structure, but maybe also originate from changing demands or changing public satisfaction. Hence, when designing a structure all costs and benefits as accumulated throughout lifetime have to be taken into account.

A similar situation is given in case of lifetime extensions of existing structures. This is a field of increasing activity, since a growing percentage of civil infrastructure as well as buildings is threatened by obsolescence, and due to simple monetary reasons this can no longer be countered by building everything anew. Hence maintenance interventions are required, which allow partial or complete structural rehabilitation and to restore profitability. However, the same economic reasoning which leads to the implementation of maintenance policies has also to be applied to the maintenance strategies themselves. In other words, maintenance interventions have to be economically reasonable, that is, maintenance expenditures have to be outweighed by expected future gains. If this is not the case, then indeed the structure is obsolete—at least in its current functional, economic, technical, or social configuration—and innovative alternatives have to be evaluated. Thus, the overall objective is to find an optimal balance between recovering the profitability of structural operation over a designated time horizon and the maintenance expenditures spent—without compromising safety issues.

A way to rationally balance advantages and disadvantages of proposed design solutions or maintenance strategies is cost-benefit analysis. Cost-benefit analysis—as well as its most general form, that is, decision analysis—is primarily normative in intent and, hence, serves as an aid to decision-making. It can be easily scrutinized and thus allows decision makers to evaluate each quantitative input or qualitative assumption with respect to possible alternatives and their consequences. Principally, cost-benefit analysis, as we apply it herein, consists of four steps:

i.) Modeling of all alternatives, that is, possible design solutions or sequences of maintenance actions and their consequences.

ii.) Assignment of probabilities to alternative-consequence relationships either based on past empirical data or derived from stochastic models.

iii.) Assignment of utilities to all consequences, that is, advantages as well as disadvantages, of proposed design solutions or maintenance strategies.

iv.) Optimization of structural design or maintenance strategies by maximizing expected utilities.

In cost-benefit analysis utility is measured in monetary units and is, but does not necessarily have to be, linear therein.

Although life-cycle considerations, lifetime performance, as well as designing for durability and sustainability gain more and more interest, the application of cost-benefit analysis still lies in its infancy. If cost considerations are performed at all, then in form of a so called life-cycle cost minimization, where—in the ideal case—all significant (adverse) costs (such as, for example, costs for construction, operation, maintenance, or failure) throughout lifetime are considered, but no benefit at all. Hence, the implicit assumptions hereby are that all alternatives have the same benefit, benefit neither depends on the condition of the structure nor on time, and benefit always outweighs all costs. The reason for not specifying benefit explicitly is the traditional focus of structural engineers on preventing structural collapse, though this is gradually changing through the utilization of performance-based criteria in design.

The aim of the present work is to apply consistently cost-benefit criteria in the planning of maintenance interventions and lifetime extensions of existing structures. We will show, that taking into account costs and benefits explicitly not only allows us to minimize expenditures, like can be done by life-cycle cost minimization, but enables us to determine whether expenditures spent for construction or maintenance are justified by future gains, allows us to assess the lifetimes of structures, and provides us with a rational method to define acceptable failure rates. In other words, the proposed analysis discloses the close relationship between monetary expenditures spent and the hazardousness of a structure through its entire lifetime, and shows the principal structure of how and when decisions for lifetime extensions and maintenance should be undertaken for maximum effect.

The remainder of this work is organized as follows. In Ch. 2 an overview on procedures in reliability-based structural optimization is given. The chapter describes the basic principles of reliability- and cost-based design, but also points out deficits of current formulations with respect to time-variant problems, that is, deteriorating structures, non-periodic maintenance policies and changing cost factors. Hence, in the following Ch. 3 a novel formulation for maintenance optimization is given, which is based on cost-benefit criteria. The time-variant formulation takes into account the different statistical characteristics of structural performance (state occupation probabilities and transition rates of performance states) as well as their corresponding cost factors. This not only allows a more refined optimization of maintenance interventions, but also allows to determine optimal lifetimes and acceptable failure rates. The proposed concept is likewise applicable for structures with and without systematic re-construction after failure.

When solving the in Ch. 3 formulated optimization problems, we have to maximize an objective function with respect to design variables which are not only continuous, but also discrete (for example, the number of maintenance times). These mixed-discrete problems are tackled herein by a hybrid of genetic algorithm and evolutionary strategy. The main operators of the proposed evolutionary algorithm and their working mechanisms are described in Ch. 4.

In the subsequent Ch. 5 the proposed cost-benefit criteria are prototypically applied to optimization problems in maintenance planning of deteriorating structures. These analyses take into account all life-cycle costs, such as construction, failure, quality-dependent inspection and state-dependent rehabilitation costs, as well as state- and time-dependent benefit rates. The numerical examples investigate the effects of different quantitative and qualitative assumptions on decision making. Ch. 6, finally, contains a summary of the important results of this work and discusses their implications for issues of lifetime extensions and sustainability.

2

Reliability-Based Optimization

2.1 INTRODUCTORY REMARKS

The ultimate goal of any engineer in designing or maintaining a structure or building can be principally described as achieving an ‘acceptable’ structural performance in the most efficient, that is, economic way. However, to specify what is ‘acceptable’ in terms of structural performance or safety raises important questions—not only in terms of structural mechanics or physical modeling, but also in terms of ethics, economics, and politics. Whereas the complex of issues dealing with questions like member sizing, structural redundancy allocation, or inspection and rehabilitation planning is of the more traditional engineering fare, the complex of issues dealing with, in principle, economical and political questions, such as the evaluation of consequences of component and structural failure, life-cycle costs, acceptable failure rates, etc. is mostly eschewed in the engineering community. Nevertheless, design or maintenance decisions are permanently made at the intersection of these mostly conflicting issues. This may not be as obvious as it should be, since questions of societal acceptance of technological risks are quite often treated in a form of ‘division of labor’. That is, problems of risk acceptance are delegated to commissions and committees, which in most cases provide standards and comprehensive guidelines based on operational experience and historically evolved rules. However, this approach tacitly assumes that past engineering practice is already nearly optimal. As has been vividly pointed out by Rackwitz (2002), since present rules were developed widely by trial and error, and are mainly formulated in terms of safety factors and cautiously selected nominal or representative design values, it is difficult to believe that design and maintenance rules based solely on past and present practice indeed provide structures which are at the same time economically optimal and ‘safe enough’. The dilemma of setting up rules based on historical knowledge gets even more evident when new conceptual problems arise, such as design for durability and the question of lifetime extensions, where only scarce previous operational experience is available.

When we want to tackle the problem of 'acceptability' of structural performance in a consistent way, then it is mandatory to take into account uncertainties. Clearly, compliance with structural performance criteria is almost never perfect. In structural operation or usage there is basically always some likeliness of non-performance or failure present—with all its associated adverse consequences. For example, the maximum load during a structure's lifetime or the evolution of the load carrying capacity of a deteriorating structure can not be predicted exactly. In fact, any prediction into the future is prone to randomness, and thus an absolute assurance of safety of a structure—in the sense of absence of failure—is realistically not feasible. The same holds for the assurance of serviceability. Hence, structural safety and serviceability may be warranted only in terms of certain probabilities. Such types of reliability and safety problems are typical not only in structural engineering, but are also quite common in industrial and mechanical engineering. An additional point to be mentioned is, that the determination and assurance of adequate system performance throughout its lifetime is quite often treated as a time-invariant problem. That is, it is assumed that the structural condition or the demand on the structure do not change with time. However, reliability and safety problems are usually time-dependent as, for example, in the case of structures whose capacities deteriorate with time and usage.

Until recently, the methods and strategies developed by engineers to manage the life cycle of a structure or building have been of purely technical nature. That is, design solutions or rehabilitation efforts have been proposed which allow to extend the structural lifetime by improving its reliability and durability, or which enable a more easily adaptation to other types of usage than initially anticipated. The development of these methods and strategies is certainly an important aspect of life-cycle engineering, but as in the case of safety considerations in initial structural design, also in life-cycle engineering questions of economic efficiency have to be asked. In fact, for most—at least still partially operable—structures there will be almost always a technical solution available to extend their lifetime. But are these solutions also justified in economic terms? Thus similar to the problem of 'how safe is safe enough?' in initial structural design we have to ask in life-cycle engineering the question of 'how much rehabilitation is too much rehabilitation?'. It should be emphasized, that this question is not meant in the sense that certain safety standards should be revised because of possible extensive costs of rehabilitation. What has to be answered, however, is, whether the expenditures spent on, say, maintenance and rehabilitation are justified by future gains or benefits from structural operation, or whether alternatives (for example, other types of usage, but also decommissioning) have to be addressed instead.

From the above follows, that to solve problems of life-cycle engineering principles of probability as well as economic theory have to be utilized. This gets evident from the fact that different costs and benefits are associated with different condition states of a structure or building. For example, due to wear and tear, fatigue, etc. bridges may have structural deficiencies which may affect structural safety. A bridge management option in such cases is to impose weight restrictions (Minchin, Jr. et al., 2006). However, this may require a re-routing of vehicles, resulting in extended travel or transportation times, additional accidents due to longer routes, increased environmental deterioration, and other economic or social losses (Minchin, Jr. et al., 2006; Sugimoto et al., 2002). In other words, the expected costs and benefits depend on the performance of the structure.

Expected costs accrue during the life cycle due to the occurrence of certain hazards (that is, they are related to the safety of the structure) and due to further, so to say, investment decisions in form of maintenance and rehabilitation efforts (which are related to the structural condition at the time of intervention). Similarly, expected benefits aggregate over time depending on the fulfillment of certain performance criteria. For a structure to be acceptable, the expected benefits have to outweigh the expected costs. That is, the structure has to give rise to welfare or human well-being during its lifetime. This is precisely what is meant by sustainability (Pearce, 2006; Rackwitz et al., 2005). And cost-benefit analysis—the principle of comparing benefits and costs in terms of social utility gains and losses—is nothing else than a quantitative approach for the evaluation of the possible decision alternatives (Boardman et al., 2006; Fuguitt and Wilcox, 1999).

However, before introducing cost-benefit criteria for life-cycle management, let us first give a critical overview of existing procedures or approaches in reliability-based optimization. We start with a short description of the historical development of reliability-based structural optimization, before we explain the components of reliability-based optimization formulations. This is followed by a more in-deep look at some state-of-the-art applications of reliability-based design and maintenance optimization. Finally, we draw some conclusions for our further work herein.

2.2 SHORT HISTORICAL OVERVIEW

According to Johnson (1953), the earliest formulation for a reliability-based economical design of civil engineering structures has been done, most likely, in the 1920s and 30s by Forssell and Gibrat. They proposed to minimize the expected total costs during the structural lifetime, that is, their objective function included not only the initial or construction costs, but also the expected future costs from failure or other damage. Also the work by Johnson (1953) is based on this approach. It should also be noted, that in Johnson's work all future costs are already properly discounted to their present value.

Whereas these early efforts remained basically unnoticed during the next decades, it was the influential paper by Freudenthal (1956) on structural safety in combination with the development of deterministic minimum weight design methods for aircrafts in the 1940s and 50s (see Kirsch, 1993), which led to the first widely recognized formulations of reliability-based optimum design in the 1960s. Due to their origin in aircraft and aerospace engineering, the design objectives have been formulated as minimum weight problems with constraints expressed in terms of probabilities of failure (Broding et al., 1964; Hilton and Feigen, 1960; Kalaba, 1962; Moses and Kinser, 1967; Switzky, 1965). Parallel to the maturing of mathematical programming techniques (Arora and Thanedar, 1986; Fletcher, 1987) and structural reliability theory (Madsen et al., 1986; Shinozuka, 1983) in the 1970s to mid 80s, these approaches become more widespread and were also increasingly applied in civil engineering (Frangopol and Moses, 1994). Although weight minimization has been also applied in civil engineering (for example, Frangopol, 1985), most applications utilize as objective function the expected total costs (Mau and Sexsmith, 1972; Moses, 1977; Moses, 1997; Parimi and Cohn, 1975; Pu et al., 1997; Rackwitz and Cuntze, 1987). This has also been reinforced in the last decade by a growing interest in performance-based design criteria (Ang and Lee, 2001; Ellingwood, 2001; Wen, 2001), as well as in maintenance or rehabilitation planning for deteriorating structures (Frangopol et al., 1997; Liu and Frangopol, 2004; Mori and Ellingwood, 1994a; Mori and Ellingwood, 1994b; Thoft-Christensen and Sørensen, 1987). In both cases life-cycle costs play a prominent role.

A quite singular position in reliability-based optimization is taken by the paper of Rosenblueth and Mendoza (1971). Contrary to all other papers up to this time and some time after, they propose not only to utilize a proper cost-benefit analysis, but they also take into account that any kind of safety or reliability problem is in principle time-variant, that is, they treat the load carrying capacities and loads consistently as stochastic processes. Interestingly, until the end of the last millennium this paper has been either ignored, or treated as a variant of the total expected cost approach (see, for example, Frangopol and Moses, 1994). Only recently the importance and the potential of this approach for risk acceptability, structural optimization and life-cycle engineering has been fully recognized (in a series of papers by Kuschel and Rackwitz, 2000; Rackwitz, 2000; Rackwitz et al., 2005; Streicher and Rackwitz, 2004).

2.3 FORMULATION OF OPTIMIZATION PROBLEMS

2.3.1 General Form

Let us look at reliability-based optimization problems in more technical detail. The formulation of any optimization problem requires the definition of an objective $g(\cdot)$, that is, a scalar quantity[1] which should be maximized or minimized, and the definition of a vector **x** of n design variables $x_1, x_2, \ldots, x_n$, which are allowed to be varied for maximizing or minimizing this objective. Since

$$\min_{\mathbf{x}\in\mathbb{R}^n}[g(\mathbf{x})] = -\max_{\mathbf{x}\in\mathbb{R}^n}[-g(\mathbf{x})] \tag{2.1}$$

the optimization problem can be formulated without loss of generality as

$$\mathbf{x}^* = \operatorname*{argmax}_{\mathbf{x}\in\mathbb{R}^n}[g(\mathbf{x})] \tag{2.2}$$

In addition, it may be necessary—either due to physical or due to functional limitations—to limit the range of the design variables **x** to a subset $\mathcal{S} \subseteq \mathbb{R}^n$. That is, the optimization problem of eq. (2.2) is modified to be

$$\mathbf{x}^* = \operatorname*{argmax}_{\mathbf{x}\in\mathcal{S}\subseteq\mathbb{R}^n}[g(\mathbf{x})] \tag{2.3}$$

Quite often, the subset $\mathcal{S}$ is formulated in terms of constraints, that is, equalities and inequalities involving **x**.

2.3.2 Design Variables

Any engineering structure or process is defined or controlled by a set of quantities. Some of them are fixed to a certain value at the outset of optimization, but to others we can freely assign different values from a given set of possible ones. These freely assignable quantities are called design variables. Design variables can be continuous or discrete. In addition, we can distinguish between structural and non-structural design variables.

Structural design variables include all variables directly related to the actual size, shape and topology of a structure. In fact, sizing, shape and topology optimization represent three principle strands in optimum structural design with their specific solution techniques (Kirsch, 1993). Hence, it is suitable to further classify structural design variables with respect to these categories:

i.) Sizing design variables: These are all quantities which are related to struc-

[1]This holds also for multi-objective optimization, where the ‘objective’ is simply a vector of scalar quantities (Rao, 1996).

tural component parameters such as material values, cross-sectional properties (area, moment of inertia, etc.) and thicknesses. Sizing only allows to find the best solution on component level, that is, for a given structural layout (geometry and topology).

ii.) Geometrical design variables: These design variables include all quantities which allow to modify the overall geometry of structures or structural components, that is, the shape of their boundaries. Examples are the coordinates of a shell structure, the location of supports in a bridge, the length of spans in a building, or the height of a truss.

iii.) Topological design variables: Whereas sizing and shape optimization results in geometrically different, but topological equivalent structures, topology optimization interferes with the connectivity pattern of components or boundaries within a structure. Topological design variables are, for example, the number of columns supporting a roof, the number of beams supporting a floor, etc. In principle, this includes also the optimization with respect to the structural system type, for example, whether a bridge is cable-stayed or uses suspension.

It goes without saying, that structural design problems can be, in general, a combination of the above three categories.

Beside the mentioned structural design variables, in problems involving performance criteria or deteriorating structures, also non-structural design variables are present. Non-structural design variables are related to the condition state of a structure, but they do not allow to change the overall structural design. Typical examples, as also used in this work, are maintenance times (that is, the time intervals between successive inspections and rehabilitations) or the quality of maintenance work (for example, threshold levels above which deteriorated condition states should be rehabilitated, or the efficiency of non-destructive inspection techniques), but also the number of maintenance interventions or the structural lifetime.

2.3.3 Constraints

Any realization of the design variables represents a design of the structure or an instance of the maintenance process. Clearly, some of these designs are useful solutions to the optimization problem, but others might be inadequate in terms of structural behavior or other considerations. If a design meets all the requirements placed on it, it is called a feasible design. The restrictions that must be satisfied in order to produce a feasible design are called constraints.

From a mathematical point of view, constraints are either expressed as a set of inequalities, written in vector form as

$$\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots]^{\mathrm{T}} \leq \mathbf{0} \tag{2.4}$$

or equalities

$$\mathbf{e}(\mathbf{x}) = [e_1(\mathbf{x}), e_2(\mathbf{x}), \ldots]^{\mathrm{T}} = \mathbf{0} \tag{2.5}$$

Hence, the feasible design space $\mathcal{S}$ for the design variables **x**, as used in eq. (2.3), is defined as

$$\mathcal{S} = \{\mathbf{x} \colon \mathbf{h}(\mathbf{x}) \leq \mathbf{0} \ \wedge \ \mathbf{e}(\mathbf{x}) = \mathbf{0}\} \tag{2.6}$$

The constraints may be linear or non-linear functions of the design variables.

From a physical point of view, we can distinguish the following two different types of constraints:

i.) Mechanical behavior constraints: These are constraints from behavior requirements, for example, acceptable stress, displacement, or buckling strength. Explicit behavior constraints are often given by formulas presented in design codes, standards or specifications. Implicit behavior constraints are, in general, derived from performance objectives, that is, failure or serviceability limit states, etc.

ii.) Side or technological constraints: These constraints restrict the range of design variables for reasons other than behavioral ones. They may derive from various considerations such as functionality, fabrication, or even esthetic. Examples are the minimum width of a beam or column, or the maximum height of a truss.

In deterministic structural optimization problems, the most typical constraints—beside side constraints—are certainly displacement and stress constraints. But also cost or weight constraints are used. In reliability-based structural optimization constraints are either deterministic as described above, or they are interpreted in probabilistic terms. That is, for example, stress and displacement constraints define performance or ultimate limit states, which are only allowed to be violated with a certain probability per unit time or whose expected number of violations per unit time should not exceed an acceptable level, respectively.

2.3.4 Objective Functions

Having given a set of feasible design vectors **x**, the best design solution for a specific criterion has to be found. Thereto, a function $g(\cdot)$ is introduced which allows to compare design alternatives consistently. This function is called objective and is defined in such a way, that a better design solution always corresponds to a larger (or smaller) value of the objective than a worse solution and attains its largest (or smallest) value for the optimal design solution $\mathbf{x}^*$. The objective function is usually non-linear in the design variables and it may represent the weight of a structure, its costs for construction and operation, or any other criterion by which a possible design is preferred to others. In general, the objective function represents the most important single property of a design, but it may also be given by a combination (for example, a weighted sum) of different properties.

The most widely used objective in structural design optimization is certainly the total structural weight (see also Sec. 2.2). The weight or mass of a structure with k components can be easily quantified as

$$g(\mathbf{x}) = \sum_{i=1}^{k} \varrho_i V_i \tag{2.7}$$

where ϱ_i is the mass density of the ith component, and V_i its volume, that is, the quantity of material used. Minimum weight design has been typically applied in aircraft and aerospace design, but also for building and civil engineering structures subjected to seismic or similar dynamic loads (Rao, 1996). However, minimum weight is only a convenient substitute criteria for cost. That is, the historical choice of, for example, minimizing gross take-off weight as the objective in commercial aircraft design is indeed intended to improve performance and subsequently lower operating costs, primarily through reduced fuel consumption (Peoples and Willcox, 2006).

A similar situation prevails in building and civil engineering design. If we take, for the moment, only the construction process, then the quantity of material used in construction constitute a significant part of its cost, but also other cost factors are important. For concrete structures, for example, not only the material costs for the concrete and reinforcement are significant, but also the costs for transportation, formwork or prestressing (Sarma and Adeli, 1998). Or in case of steel structures, for example, also costs related to the design of connections should be taken into account (Sarma and Adeli, 2002). Obviously, besides these costs, which are related to the structural capacity, there are also costs of similar magnitude for non-structural elements like technical facilities and electrical services (Kirk and Dell'Isola, 1995). Hence, strictly speaking, the weight-minimization problem should be formulated in terms of cost factors.

Let us assume that there are m different (random) cost factors C_j ($j = 1, 2, \ldots, m$), that is, the (random) initial cost C_0 is

$$C_0(\mathbf{x}) = \sum_{j=1}^{m} C_j(\mathbf{x}) \tag{2.8}$$

where **x** is the vector of design variables. There are now different approaches to utilize the initial cost C_0 as an objective function for optimization. These approaches differ in the way how we rank the possible outcomes, that is, how we express for different design vectors **x** our preferences for the respective probability distributions of $C_0(\mathbf{x})$ by a utility function (Luce and Raiffa, 1957). If there is a certain type of consistency in our preference description, then Von Neumann and Morgenstern (1947) showed that all our decisions should be solely based on the expected value of this utility function[2]. Thus, if we prefer to minimize the initial cost $C_0(\mathbf{x})$, but we do not care about possible large (and small) values of $C_0(\mathbf{x})$, that is, if we are so to say risk-neutral, then our utility is a linear function in the costs $C_j(\mathbf{x})$. Hence, the objective function $g(\mathbf{x})$ to be minimized is in this case

$$g(\mathbf{x}) = c_0(\mathbf{x}) = \mathrm{E}\big[C_0(\mathbf{x})\big] = \sum_{j=1}^{m} \mathrm{E}\big[C_j(\mathbf{x})\big] \tag{2.9}$$

with $c_0(\mathbf{x})$ denoting the expected initial cost. For obtaining an acceptable structural design, the objective function of eq. (2.9) has to be subjected, in general, to performance constraints. The minimum requirement is, at this, that the expected failure rate $h(\mathbf{x}, t)$ does not exceed a socially, economically or otherwise accepted limit $h_{\mathrm{a}}(t)$. Eq. (2.9) is the most common form of objective function when minimizing initial costs (Frangopol and Moses, 1994).

As getting evident from the linear form of eq. (2.9), we are indifferent with respect to design solutions **x** which have different probability distributions of the initial costs $C_0(\mathbf{x})$ and hence, for example, also different variances $\mathrm{Var}[C_0(\mathbf{x})]$, as long as the expected costs $c_0(\mathbf{x})$ are the same. This is not a weak point of utility theory, but only owed to our description of utility as a function linear in costs. If we want to be more risk-averse, we can do so, for example, by utilizing an exponential function for the utility (for other functions see also Levy and Markowitz, 1979). In this case, the objective function to be minimized is

$$g(\mathbf{x}) = \frac{1}{\varepsilon} \ln\{\mathrm{E}\left[\exp\left(\varepsilon C_0(\mathbf{x})\right)\right]\} \tag{2.10}$$

with $\varepsilon > 0$ as a parameter controlling our aversion. From the concept of certainty equivalence (Keeney and Raiffa, 1993), we see that

$$\exp\left(\varepsilon g(\mathbf{x})\right) = \mathrm{E}\left[\exp\left(\varepsilon C_0(\mathbf{x})\right)\right] \tag{2.11}$$

[2]More on this in Ch. 3.

Using Taylor's formula for the right-hand side of eq. (2.11), we can approximate the objective function $g(\mathbf{x})$ for small values of ε as

$$g(\mathbf{x}) \approx c_0(\mathbf{x}) + \frac{\varepsilon}{2}\,\mathrm{Var}\,[C_0(\mathbf{x})] \tag{2.12}$$

Thus, when using eq. (2.10) as an objective to minimize, we simultaneously minimize the expected initial cost and its variance—at least approximately.

A refinement of eq. (2.9) is to directly incorporate the failure consequences in the objective function. Let us denote with L the possible (future) loss, then the objective function for a structure without re-construction after failure is

$$g(\mathbf{x}, t, t_0) = c_0(\mathbf{x}) + \int_{t_0}^{t} L \exp(-\gamma\tau) h(\mathbf{x}, \tau)[1 - \Pr(F; \mathbf{x}, \tau, t_0)]\mathrm{d}\tau \tag{2.13}$$

where γ is the discount rate. For $\gamma = 0$, above formulation can be simplified to give

$$g(\mathbf{x}, t, t_0) = c_0(\mathbf{x}) + L \Pr(F; \mathbf{x}, t, t_0) \tag{2.14}$$

with $\Pr(F; \mathbf{x}, t, t_0)$ being the probability of first-passage to the failure domain F in the time interval $[t_0, t]$. The formulation of eq. (2.14) is the most common one in reliability-based structural design (Ang and Lee, 2001; Frangopol and Moses, 1994; Wen, 2001). It should be noted, however, that the respective structural optimization problem now refers to a specific period of time, that is, the time interval $[t_0, t]$.

A slight variation of the objective function of eq. (2.13) is given when the structure is systematically repaired (that is, re-constructed) after failure. In this case, we no longer operate with a single structure, but with an 'inventory' of structures which get activated one by one. For non-deteriorating structures the objective function is

$$g(\mathbf{x}, t, t_0) = c_0(\mathbf{x}) + \int_{t_0}^{t} (c_0(\mathbf{x}) + L) \exp(-\gamma\tau) h(\mathbf{x})\mathrm{d}\tau \tag{2.15}$$

For $\gamma = 0$, we can again simplify this to give

$$g(\mathbf{x}, t, t_0) = c_0(\mathbf{x}) + (c_0(\mathbf{x}) + L) \ln\left[1 + \frac{\Pr(F; \mathbf{x}, t, t_0)}{1 - \Pr(F; \mathbf{x}, t, t_0)}\right] \tag{2.16}$$

Although, in both eqs. (2.13) and (2.15) the failure consequences are incorporated directly, for an acceptable design we still require, in general, performance constraints. This is because these objective functions only contain 'negative' consequences to be minimized, but there is no term present which represents the 'positive' consequences of operating these structures and which is able to outweigh all 'negative' consequences. Or, to view it from a different perspective, when using an unconstrained objective of the form

$$g(\mathbf{x}, t, t_0) = c_0(\mathbf{x}) + l(\mathbf{x}, t, t_0) \tag{2.17}$$

to be minimized, where $l(\mathbf{x}, t, t_0)$ is the expected loss[3], we implicitly assume that all design solutions have the same (time-invariant) benefit rate, and that this benefit rate will always outweigh the cost rate $\dot{g}(\mathbf{x}, t, t_0)$.[4]

A formulation of the objective, which is sometimes proposed alternatively to the minimization of the weight according to eq. (2.7) or the expected costs according to eq. (2.14), is to minimize the probability of failure directly, that is,

$$g(\mathbf{x}, t, t_0) = \Pr(F; \mathbf{x}, t, t_0) \tag{2.18}$$

Obviously, for an economically reasonable design this objective has to be augmented with a cost constraint, or—as a substitute criteria—with a weight constraint (Frangopol and Moses, 1994; Rackwitz and Cuntze, 1987)[5].

As already mentioned in Sec. 2.2, a singular position in reliability-based optimization is taken by the approach in (Rosenblueth and Mendoza, 1971). Their objective to be maximized is a net present benefit function

$$g(\mathbf{x}, t, t_0) = b(\mathbf{x}, t, t_0) - c_0(\mathbf{x}) - l(\mathbf{x}, t, t_0) \tag{2.19}$$

where $b(\mathbf{x}, t, t_0)$ is the expected benefit, which depends on the vector of design variables **x** (Rosenblueth and Mendoza, 1971; Kuschel and Rackwitz, 2000). For a structure without re-construction after failure the expected benefit is

$$b(\mathbf{x}, t, t_0) = \int_{t_0}^{t} \dot{B}(\mathbf{x}, \tau) \exp(-\gamma\tau)[1 - \Pr(F; \mathbf{x}, \tau, t_0)]\mathrm{d}\tau \tag{2.20}$$

that is, the benefit aggregates only as long as the structure has not failed. For a time-constant benefit rate $\dot{B}(\mathbf{x}) = $ const. and a zero discount rate $\gamma = 0$, the expected benefit is determined as

$$b(\mathbf{x}, t, t_0) = \dot{B}(\mathbf{x}) \int_{t_0}^{t} [1 - \Pr(F; \mathbf{x}, \tau, t_0)]\mathrm{d}\tau \tag{2.21}$$

[3]The expected loss can be defined in a much wider sense than done herein so far. That is, it may not only include the costs from structural collapse, but also costs due to violating performance criteria, costs from operation and maintenance, etc.

[4]We investigate the subject of costs and benefits in structural design in more detail in Ch. 3.

[5]Strictly speaking, this type of optimization problem is not reliability-based, but indeed a cost- or weight-based reliability optimization problem.

and hence the net benefit function is

$$g(\mathbf{x}, t, t_0) = \dot{B}(\mathbf{x}) \int_{t_0}^{t} [1 - \Pr(F; \mathbf{x}, \tau, t_0)] \mathrm{d}\tau - c_0(\mathbf{x}) - L \Pr(F; \mathbf{x}, t, t_0) \quad (2.22)$$

In eqs. (2.20) to (2.22) we assumed that the benefit rate $\dot{B}(\cdot)$ does not depend on the condition state of the structure.

The most often employed variant of eq. (2.19) is with systematic re-construction after failure. In this case, the structure is always available to generate benefit. Hence, the expected benefit is, in general,

$$b(\mathbf{x}, t, t_0) = \int_{t_0}^{t} \dot{B}(\mathbf{x}, \tau) \exp(-\gamma\tau) \mathrm{d}\tau \quad (2.23)$$

and, with the same assumptions as above (that is, $\dot{B}(\mathbf{x}) =$ const. and $\gamma = 0$), the net benefit function is given as

$$g(\mathbf{x}, t, t_0) = \dot{B}(\mathbf{x})[t - t_0] - c_0(\mathbf{x}) - (c_0(\mathbf{x}) + L) \ln\left[1 + \frac{\Pr(F; \mathbf{x}, t, t_0)}{1 - \Pr(F; \mathbf{x}, t, t_0)}\right] \quad (2.24)$$

See also eq. (2.16) for the failure cost term. The objective function of eq. (2.19) has been originally used without any constraints (Rosenblueth and Mendoza, 1971), but has been recently augmented by reliability or cost constraints (Rackwitz, 2001; Rackwitz, 2004).

The typical optimization problem in structural design involves a single objective function like weight or expected cost and multiple constraints formulated in terms of displacements, stresses, frequencies etc. (Kirsch, 1993). As we have seen above, under conditions of uncertainty, probabilistic measures such as probabilities of failure or failure rates enter the objective function or the constraints. This becomes, in general, an additional burden in the optimization process, since the calculation of these probabilistic measures poses quite often a formidable task in itself. For example, when utilizing the first order reliability method, the determination of the reliability index is again an optimization problem to be solved. Different methods have been developed to solve such types of problems, like the bi-level approach (Enevoldsen and Sørensen, 1994), the mono-level approach (Kuschel and Rackwitz, 2000) or the decoupling approach (Royset et al., 2001). Nevertheless, reliability-based design certainly remains also in the foreseeable future a challenging task. In case of reliability-based maintenance optimization the formulations are computationally more amenable, since the problem of determining the probabilistic measures can be solved, in general, independently from the optimization task.

2.4 STATE OF THE ART IN RELIABILITY-BASED OPTIMIZATION

2.4.1 Structural Design Optimization

Now, let us take a more in-deep look at some state-of-the-art applications in reliability-based optimization. We start with the papers by Rosenblueth and Mendoza (1971) and Rackwitz (2000), since these will be most instrumental for our formulation of cost-benefit criteria in maintenance optimization herein. In fact, these papers are so far the only ones which formulate an optimal structural design problem as a cost-benefit problem and, at the same time, take into account the time-variant nature of reliability problems. However, they are only interested in the time interval $[0, \infty)$, that is, in the value of $g(\mathbf{x}, \infty, 0)$, but not in the evolution of the net present benefit throughout time t. Whereas the formulation of the present benefit in case of no re-construction after failure by Rosenblueth and Mendoza (1971) is identical to eq. (2.20), Rackwitz (2000) gives a somehow different description:

$$b(\mathbf{x}, \infty, 0) = \int_0^\infty \int_0^\tau \dot{B}(\mathbf{x}, \xi) \exp(-\gamma\xi) \mathrm{d}\xi f(\mathbf{x}, \tau) \mathrm{d}\tau \tag{2.25}$$

where $f(\cdot)$ is the probability density function of the time to first failure. The 'shortcoming' of eq. (2.25) is, that the benefit only accrues when the structure fails. Thus, eq. (2.25) is only correct for the written case $t \to \infty$, that is, when the structure has failed with probability one, and gives in fact in this case the same value like eq. (2.20). The identity of both equations for $t \to \infty$ can also be shown by integrating eq. (2.25) py parts. However, eq. (2.25) can not be used for calculating the present benefit for $t < \infty$ by just modifying the upper integration limit in the first integral.

The expected present loss is in (Rosenblueth and Mendoza, 1971; Rackwitz, 2000) also written in terms of the probability density function of the time to first failure as

$$l(\mathbf{x}, \infty, 0) = \int_0^\infty L \exp(-\gamma\tau) f(\mathbf{x}, \tau) \mathrm{d}\tau \tag{2.26}$$

Eq. (2.26) is identical to the loss calculation in eq. (2.13), since the failure rate $h(\cdot)$ is defined as

$$h(\mathbf{x}, t) = \frac{f(\mathbf{x}, t)}{1 - F(\mathbf{x}, t)} \tag{2.27}$$

and the cumulative distribution is

$$F(\mathbf{x}, t) = \Pr(F; \mathbf{x}, t, 0) \tag{2.28}$$

Hence,

$$f(\mathbf{x},t) = h(\mathbf{x},t)[1 - \mathrm{Pr}(F;\mathbf{x},t,0)] \tag{2.29}$$

The integrals on the right hand sides of eqs. (2.25) and (2.26) are nothing else than Laplace transforms, where the discount rate γ is the Laplace variable. Thus, the net benefit function is written in (Rosenblueth and Mendoza, 1971; Rackwitz, 2000) in terms of the Laplace transform $f^*(\cdot)$ of $f(\cdot)$ as

$$g(\mathbf{x},\infty,0) = \frac{\dot{B}(\mathbf{x})}{\gamma}(1 - f^*(\mathbf{x},\gamma)) - c_0(\mathbf{x}) - Lf^*(\mathbf{x},\gamma) \tag{2.30}$$

with $\dot{B}(\cdot)$ is assumed to be constant in time. In the important case that the failure events occur according to a Poisson process with intensity $\lambda(\cdot)$ we simply get

$$g(\mathbf{x},\infty,0) = \frac{\dot{B}(\mathbf{x})}{\gamma + \lambda(\mathbf{x})} - c_0(\mathbf{x}) - \frac{L\lambda(\mathbf{x})}{\gamma + \lambda(\mathbf{x})} \tag{2.31}$$

The main focus in (Rosenblueth and Mendoza, 1971; Rackwitz, 2000), however, is on systematic re-construction after structural failure. For this purpose they utilize a renewal process, that is, the j-th event ($j = 1, 2, \ldots$) in the process occurs at time $T_1 + T_2 + \cdots + T_j$, where $\{T_1, T_2, \ldots\}$ are identically distributed with probability density function $f(\cdot)$, but where T_1 may have a different density function $f_1(\cdot)$. In fact, we can distinguish three types of renewal processes (Cox and Miller, 2001):

i.) *Ordinary renewal process:* All random variables T_j are identically distributed, hence, $f_1(t) = f(t)$.

ii.) *Modified renewal process:* The probability density functions $f_1(t)$ and $f(t)$ are not necessarily the same.

iii.) *Equilibrium (or stationary) renewal process:* The probability density function $f_1(t)$ has the special form:

$$f_1(t) = \frac{1 - F(t)}{\mu} \tag{2.32}$$

where $F(t)$ is the distribution function corresponding to $f(t)$ and $\mu = \mathrm{E}[T_j]$ (with $j = 2, 3, \ldots$).

The third type of renewal process is certainly the form which is entirely in agreement with the basic argument of 'stationarity' made in (Rosenblueth and Mendoza, 1971; Rackwitz, 2000).

In case of systematic re-construction, the cost and benefit functions are formulated in terms of the renewal density $m(t)$, which is defined as

$$m(t) = \sum_{r=1}^{\infty} f^{(j)}(t) \tag{2.33}$$

where $f^{(j)}(\cdot)$ is the j-fold convolution of the density with itself, that is,

$$f^{(j)}(t) = \int_{0}^{t} f^{(j-1)}(t-\tau) f(\tau) \mathrm{d}\tau \tag{2.34}$$

whereby $f_1(\cdot)$ has to be taken into account in the first convolution. Now taking the Laplace transform of the renewal density $m(t)$ gives for the ordinary renewal process

$$m^*(\gamma) = \frac{f^*(\gamma)}{1 - f^*(\gamma)} \tag{2.35}$$

for the modified renewal process

$$m^*(\gamma) = \frac{f_1^*(\gamma)}{1 - f^*(\gamma)} \tag{2.36}$$

and for the equilibrium renewal process

$$m^*(\gamma) = \frac{1}{\gamma\mu} \tag{2.37}$$

Just mentioning the equilibrium renewal process, we get for the net present benefit the following relation:

$$g(\mathbf{x}, \infty, 0) = \frac{\dot{B}(\mathbf{x})}{\gamma} - c_0(\mathbf{x}) - \frac{c_0(\mathbf{x}) + L}{\gamma\mu(\mathbf{x})} \tag{2.38}$$

Thus, the only probabilistic measure entering the objective function is the mean time to failure $\mu(\mathbf{x})$. In case of the Poisson process with intensity $\lambda(\cdot)$ above equation reads

$$g(\mathbf{x}, \infty, 0) = \frac{\dot{B}(\mathbf{x})}{\gamma} - c_0(\mathbf{x}) - \frac{(c_0(\mathbf{x}) + L)\lambda(\mathbf{x})}{\gamma} \tag{2.39}$$

In (Rosenblueth and Mendoza, 1971) the objective functions of eqs. (2.30) and (2.38) have been proposed without any constraints. In (Rackwitz, 2000), however, the objective functions are initially subjected to constraints formulated in terms of failure rates, and later, for example in (Rackwitz, 2004), to a constraint formulated in terms of a social indicator called life-quality index (Lind, 1993; Pandey and Nathwani, 1997). If these constraints are active, then, in both cases, a more expensive design solution is selected than would be necessary according to the net-benefit criterion (see Fig. 2.1). The failure-rate constrains, however, have been added

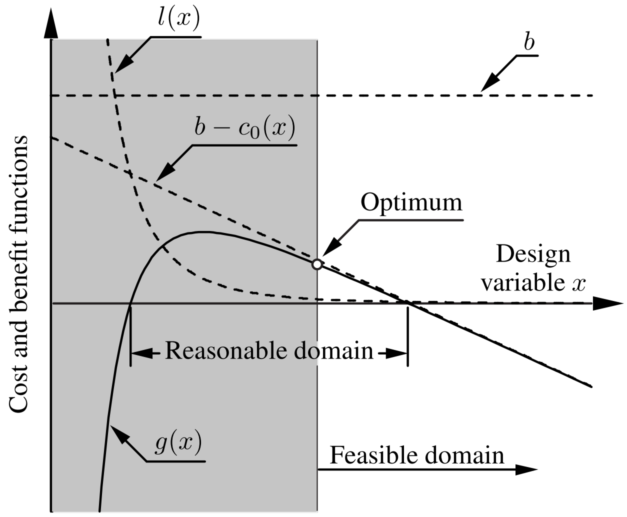

Figure 2.1: Schematic sketch of cost and benefit functions with optimal solution for constrained problem according to (Rackwitz, 2000; Rackwitz, 2004).

only for reliability verification purposes or compliance with codified admissible failure rates (Rackwitz, 2000) and can indeed also "[. . .] be absent" (Rackwitz, 2001) from the optimization problem, since in case of an already optimal design practice they would be always fulfilled anyway. A similar argument holds for the life-quality-index constraint—at least as long as the life-saving costs are included in L (Rackwitz, 2004). It should be mentioned, however, that in (Rackwitz, 2004) it is the life-quality index which determines what is acceptable for a technical facility, and not the net present benefit function of, say, eq. (2.38).

In (Streicher and Rackwitz, 2004) there has been made an attempt to apply eq. (2.38) to maintenance problems. However, the only maintenance policy which fits to the formulation based on renewal processes is obviously age replacement (Barlow and Proschan, 1965; Streicher and Rackwitz, 2004), that is, the structure will be newly re-constructed at failure or some fixed replacement time t_{R}, whichever comes first. This restrictive capability of the renewal formulation with respect to maintenance planning has a simple reason. The structures under investigation can be essentially only in one of two possible states: the safe state S or the failure state F. In other words, there are no deterioration or condition states present which could be directly addressed as would be necessary for any kind of condition-based approach. Consequently, only the probability density function of the time to first failure can be manipulated. In case of age replacement this simply means that the probability density function $f(\cdot)$ is no longer integrated until $t \to \infty$, but only until the replacement time t_{R} (Streicher and Rackwitz, 2004). A further restriction in the maintenance policy arises due to the usage of systematic renewals and infinite time horizons. This leads necessarily to a stationary policy, that is, constant replacement times t_{R}. Thus, the important problem of remaining lifetimes of existing structures can not be addressed satisfactorily.

Further below herein, we will model the probability evolution of deteriorating structures by continuous-time Markov chains (Ross, 1970; Ross, 2003). At the moment it may suffice to show exemplarily, that such approach leads to the same results of the net present benefit as when using the probability density functions of the time to first failure. Let us denote the probability of being at time t either still in the safe state S or already in the failure state F as $\pi_{\mathrm{S}}(t) = \Pr(S;t)$ and $\pi_{\mathrm{F}}(t) = \Pr(F;t)$, respectively. In case of no re-construction and failure events occurring according to a Poisson process with intensity λ, we get the following differential equation

$$\dot{\boldsymbol{\pi}}(t) = \begin{bmatrix} \dot{\pi}_{\mathrm{S}} \\ \dot{\pi}_{\mathrm{F}} \end{bmatrix} = \begin{bmatrix} -\lambda & 0 \\ \lambda & 0 \end{bmatrix} \begin{bmatrix} \pi_{\mathrm{S}} \\ \pi_{\mathrm{F}} \end{bmatrix} = \mathbf{A}\boldsymbol{\pi}(t), \quad \boldsymbol{\pi}(t=0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{2.40}$$

Inserting the solution of eq. (2.40) in the net present benefit function gives

$$\begin{aligned} g(t,0) &= \int_0^t \dot{B} \exp(-\gamma\tau)\pi_{\mathrm{S}}(\tau)\mathrm{d}\tau - c_0 - \int_0^t L \exp(-\gamma\tau)\dot{\pi}_{\mathrm{F}}(\tau)\mathrm{d}\tau \\ &= \frac{\dot{B}}{\gamma+\lambda}\{1 - \exp[-(\gamma+\lambda)t]\} - c_0 \\ &\quad - \frac{L\lambda}{\gamma+\lambda}\{1 - \exp[-(\gamma+\lambda)t]\} \end{aligned} \tag{2.41}$$

Performing the passage to the limit $t \to \infty$ finally results in

$$g(\infty,0) = \frac{\dot{B}}{\gamma+\lambda} - c_0 - \frac{L\lambda}{\gamma+\lambda} \tag{2.42}$$

which is the solution reported in eq. (2.31). Clearly, in case of deteriorating structures the intensity λ is a function of time and, hence, the two-state transition matrix **A** of eq. (2.40) is also time-dependent.

In case of systematic re-construction 'additional' structures become available at a rate proportional to the failure rate (see Fig. 2.2). This can be interpreted as a

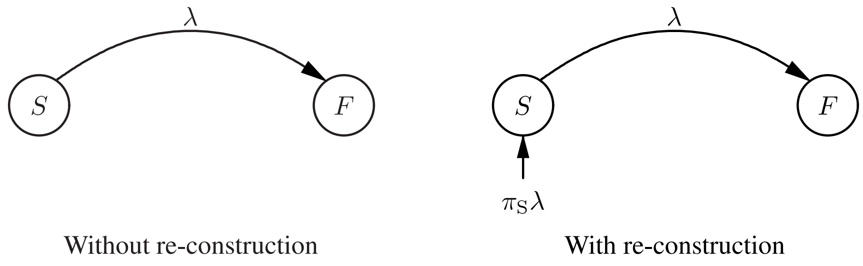

Figure 2.2: Transition rate diagrams for a two-state structure (S: safe state; F: failure state).

'forcing' term $\boldsymbol{\zeta}(t)$ which is added to the autonomous eq. (2.40) to give

$$\dot{\boldsymbol{\pi}}(t) = \mathbf{A}\boldsymbol{\pi}(t) + \underbrace{\begin{bmatrix} \pi_S\lambda \\ 0 \end{bmatrix}}_{=\boldsymbol{\zeta}(t)} = \begin{bmatrix} 0 & 0 \\ \lambda & 0 \end{bmatrix} \boldsymbol{\pi}(t), \quad \boldsymbol{\pi}(t=0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{2.43}$$

Analogously, the net present benefit is

$$\begin{aligned} g(t,0) &= \int_0^t \dot{B} \exp(-\gamma\tau)\pi_S(\tau)\mathrm{d}\tau - c_0 - \int_0^t (c_0 + L) \exp(-\gamma\tau)\dot{\pi}_F(\tau)\mathrm{d}\tau \\ &= \frac{\dot{B}}{\gamma}[1 - \exp(-\gamma t)] - c_0 - \frac{(c_0 + L)\lambda}{\gamma}[1 - \exp(-\gamma t)] \end{aligned} \tag{2.44}$$

which results for $t \to \infty$ in the solution of eq. (2.39):

$$g(\infty, 0) = \frac{\dot{B}}{\gamma} - c_0 - \frac{(c_0 + L)\lambda}{\gamma} \tag{2.45}$$

As already mentioned above in a different context, for designing condition-based maintenance policies the Markov chain has to be augmented by deterioration or condition states which are (directly or indirectly) observable.

With respect to this identified necessity of including deterioration states in the analysis, the paper of Ang and Lee (2001) is of interest. They formulate an earthquake-resistant design of reinforced concrete structures as a minimization problem of the expected total life-cycle costs subject to failure rate constraints. The first interesting aspect of their formulation is, that the life-cycle costs include also the costs for rehabilitating damage due to (minor or moderate) earthquakes. The condition state of the reinforced concrete structures are determined with the help of the damage index of Park et al. (1985). The rehabilitation costs are modeled as an increasing function of this damage index, that is, the more overall damage present in the structure, the higher the rehabilitation costs for restoring the structure to an 'as new' state. By this approach the authors take into account the requirement from performance-based design to consider different levels of performance objectives for reducing the high costs associated with the loss of use and rehabilitation of heavily damaged structures (Ghobarah, 2001).

The second interesting aspect of the paper by Ang and Lee (2001) is, that they give an unusually detailed breakdown of the losses or failure costs. They differentiate between direct and indirect losses. The direct losses include the above mentioned damage repair costs, the costs for replacement of property and non-structural components, costs for avoiding injuries, and life-saving costs. The costs for life saving and injury avoidance are determined from the relations between structural collapse rates and fatality or injury rates, the expected number of persons affected, and appropriate economic estimates of the value of statistical injury and statistical life as taken from (Viscusi, 1993). The indirect losses are attributed to the economic effects caused by structural failure, the so called ripple effects. In fact, such effects can have quite substantial and sustained impact, whereby variations of the total impact are explained by the seriousness of the respective failure and the a priori riskiness assigned to it (see thereto also Broder, 1990).

Since the paper by Ang and Lee (2001) has been written for the purpose of earthquake-resistant design, the rehabilitation actions are assumed to be event-based. That is, the earthquake is the only event which causes damage or failure, and this event is always followed by rehabilitation or re-construction of the structure to an 'as new' state. Hence, despite the sophisticated cost modeling, in the actual calculation the effect of taking into account structural damage is just a simple modification of the cost function—an additional term like the indirect losses. In fact, Ang and Lee (2001) do not take into account structural deterioration during the life cycle, nor does their approach allow a more refined maintenance or rehabilitation policy.

2.4.2 Optimization of Maintenance Strategies

Although there are many publications in reliability engineering dealing with the problem of optimal maintenance planning for deteriorating single- or multi-unit systems (for a comprehensive overview see Barlow and Proschan, 1965; Pierskalla and

Voelker, 1976; Valdez-Flores and Feldman, 1989; Wang, 2002), reliability-based maintenance optimization for structures or infrastructure did not receive any considerable interest until the late 1980s (Thoft-Christensen and Sørensen, 1987; Madsen et al., 1990). Madsen et al. (1990) propose to minimize the expected present life-cycle costs, which are composed of initial costs c_0, inspection costs $I(\cdot)$, rehabilitation costs R and failure costs L, throughout a given lifetime $[0, T]$. The optimization variables are the number n of maintenance interventions, the maintenance times $\mathbf{t} = \{t_1, t_2, \ldots, t_n\}$ and the quality $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ of the inspection procedure. The optimization problem in (Madsen et al., 1990) is formulated as[6]

$$\begin{aligned}\min_{n,\mathbf{t},\boldsymbol{\alpha}} \; & c_0 + \sum_{j=1}^{n} \{I(\alpha_j)[1 - \Pr(F; t_j)] + R\mathrm{E}[M_j]\} \frac{1}{(1+\gamma)^{t_j}} \\ & + \sum_{s=1}^{T} L[\Pr(F; s) - \Pr(F; s-1)] \frac{1}{(1+\gamma)^{s}}\end{aligned} \tag{2.46}$$

subject to

$$\beta(T) = -\Phi^{-1}(\Pr(F; T)) \geq \beta_{\min} \tag{2.47}$$

In eq. (2.46) the term $\mathrm{E}[M_j]$ denotes the expected number of rehabilitations at the j-th maintenance intervention. Also, the inspection costs $I(\cdot)$ are a function of the inspection quality α_j, whereas the rehabilitation costs R are independent of the amount of damage present. Thus, the cost modeling indicates more likely a pure replacement strategy.

The objective function of eq. (2.46) is subjected to the constraint of eq. (2.47) formulated in terms of the probability of failure (or reliability index β) for the entire lifetime $[0, T]$. This or similar formulations are quite typical for most applications in structural engineering. However, the probability of failure is an 'integral measure', that is, it allows to trade-off an overly hazardous structure at, say, the last period of the life cycle against an overly safe structure from earlier on. Hence, instead of a time-dependent probability of failure specified for an arbitrary horizon, it is the failure rate which should be utilized for setting safety targets (Rackwitz, 2000). This also allows to avoid principal inconsistencies like in (Thoft-Christensen and Sørensen, 1987), where the objective function to be minimized is additionally constrained in terms of the reliability indices $\beta(t_j)$ ($j = 1, 2, \ldots, n$) to always the same minimum reliability index $\beta_{\min}$ between maintenance interventions, although

[6] It should be noted that we modified the failure cost term given in (Madsen et al., 1990) to have proper yearly discounting. Moreover, in (Madsen et al., 1990) it is assumed that the failure cost L is a function of time, although no description of this function is given.

the time intervals $(t_{j-1}, t_j]$ between the interventions are to be optimized and differ considerable.

In (Mori and Ellingwood, 1994b) the formulation developed in (Thoft-Christensen and Sørensen, 1987; Madsen et al., 1990) is applied to the optimal maintenance planning for concrete beams. In their paper not only the inspection costs are dependent on its quality, but also the repair costs depend on the amount of damage present in the structure. The optimization is again performed with respect to inspection qualities (in their case a detection threshold) and maintenance times. They do not discount any costs, because they claim that "[...] the future discount rate is unknown" (Mori and Ellingwood, 1994b). More interesting, however, are their optimization results when the cost factors for maintenance and failure are varied. In case of high failure-cost/maintenance-cost ratios the structure is subjected to extensive maintenance interventions to keep the structure in a 'as new' state throughout the given lifetime $[0, T]$, whereby the probability-of-failure constraint of the type like eq. (2.47) does not become active. In case of low failure-cost/maintenance-cost ratios, however, only a minimum of number maintenance interventions are performed at the beginning of the life cycle just to ensure compliance with the probability-of-failure constraint at time T.

The latter is in contrast to the findings in (Frangopol et al., 1997). Although Frangopol et al. (1997) refer explicitly to the work in (Thoft-Christensen and Sørensen, 1987) and (Mori and Ellingwood, 1994b), and utilize similar cost models, an expected present life-cycle cost as objective to be minimized and probability-of-failure constraints as in (Thoft-Christensen and Sørensen, 1987), their optimal solutions indicate that maintenance interventions should not be performed in the beginning of the lifetime, but at the end of the lifetime when deterioration becomes more manifest. Unfortunately, the descriptions in (Frangopol et al., 1997) are somehow incongruous (for example, failure costs are not discounted, although all other costs are discounted; or the probability of failure in the lifetime is determined as maximum of the probabilities of failure between maintenance interventions) for being able to perform a more detailed comparison. But what is most notable at the analyses of Mori and Ellingwood (1994b), Frangopol et al. (1997) and similar ones is, that the probability-of-failure constraints quite often strongly affect the solutions. Nevertheless, so far there has basically no attempt been made to utilize cost-benefit analysis for maintenance optimization—with the exception of Streicher and Rackwitz (2004) mentioned above—which would, among others, allow to avoid inconsistencies between failure costs and reliability constraints. Nor are there any formulations available which take into account other performance criteria than failure—as done recently in structural design optimization.

2.5 CONCLUSIONS

During the past decades there has been tremendous progress towards solving life-cycle-based and cost-optimal design or maintenance problems. Nevertheless, from the short review of state-of-the-art applications given above it becomes evident that the following four topics require further development:

i.) Whereas the minimization of the expected life-cycle costs is an important contribution to determine cost-effective design or maintenance solutions, it has some shortcomings in so far as it always implicitly assumes that all solutions have the same benefit and that the benefit always outweighs the costs. In general, however, we have to justify why we want to allocate resources for a new or existing structure. And this can only be done by showing that the 'advantages' (benefits) outweigh the 'disadvantages' (costs). For example, in case of maintenance planning we have to determine the conditions under which is it worth to extend the lifetime of an existing structure by rehabilitation efforts and, equally important, when not, that is, when is the structure indeed obsolete and warrants no further investments. Similarly, "[...] the concept of an 'acceptable risk' is without proper meaning unless benefits are known" (Pearce, 1981). Hence, in both the decision on investments and in the setting of acceptable limits a cost-benefit analysis is required.

ii.) In reliability-based maintenance optimization, but also in reliability-based structural design, it is still common practice to utilize the probability of failure specified for an arbitrary time horizon as a criterion of 'risk acceptability'—for example, in the form of constraints. Such a practice can be certainly justified for non-deteriorating structures subjected to stationary loads or demands, since in this case the general time-variant reliability problem can be transformed in a time-invariant one in terms of the just mentioned probabilities of failure. However, in general cases, like when deterioration is present, the problem has to be treated as an intrinsic time-variant one. That is, the 'acceptability' of the structure has to be verified for each point in the entire lifetime. Hence, the appropriate choice for setting safety targets is the failure rate. Only thereby we can avoid the implicit trade-offs between overly safe and overly hazardous periods in the life cycle of a structure.

iii.) In basically all applications of reliability-based design or maintenance optimization it is assumed that the optimal lifetime of the structure is known. But what about an existing structure which has been used already for some time? Is its lifetime the remaining time of use? Or does it have the same lifetime like a new structure? And what is the lifetime of an existing structure which will be used in a different way than originally designed for? Clearly, these kind of questions are closely related with the acceptable failure rate of a structure. For example, if we have a non-deteriorating structure which is designed in such a way that its failure rate is smaller than the acceptable failure rate, then, under the condition that the loads or demands acting on the structure will not increase in time, the failure rate of the structure will be for any future point in time smaller than the acceptable failure rate. In other words, the lifetime of a non-deteriorating structure is infinite[7]. This should not be misunderstood in such a way, that the structure will not fail at all. Rather, the probability of failure per, say, year, under the condition, that the structure has not failed so far, is the same in each year, and is always less than what is deemed acceptable. In case of deterioration, or more precisely, increasing failure rates, the lifetime is certainly finite. But in both cases the lifetime is related to the acceptable failure rate and the current condition of the structure.

iv.) Beside preventing structural collapse, an important aspect of maintenance planning and structural design is to assure a certain performance of the structure throughout its lifetime. That is, the benefit from using or operating a structure may be reduced due to structural deficiencies. The same holds for functional aging, where the structure is still technical sound, but does no longer meet its current demands. Such effects can only be consistently described when the condition- and time-dependency of benefits is explicitly taken into account.

In the following we will develop along the line of these four topics a novel formulation for maintenance optimization based on cost-benefit criteria. This formulation will not only allow a more refined optimization of maintenance interventions, but will simultaneously allow to determine optimal lifetimes and acceptable failure rates.

[7] The use of a finite time interval in traditional design is only due to the use of the probability of failure as an acceptance criteria, which is always related to a time interval.

3

Fundamentals of Cost-Benefit Analysis

3.1 VALUE FUNCTIONS AND UTILITY FUNCTIONS

Before we introduce cost-benefit analysis, let us formally investigate the process of decision making. Given is a set $\mathcal{A}$ of all feasible acts or alternatives a. To each act $a \in \mathcal{A}$ we associate n indices of value: $X_1(a)$, $X_2(a)$, $\ldots$, $X_n(a)$. In principle, we can think of the n evaluators $X_j(a)$ as mapping each a from the act space into a point

$$\mathbf{x} = (x_1 = X_1(a), x_2 = X_2(a), \ldots, x_n = X_n(a)) \tag{3.1}$$

in the n-dimensional consequence space (Fig. 3.1). Now, the decision maker's problem is how to choose a in such a way, that he will be 'happiest' with the resulting consequences $\mathbf{x}$. In some cases there may exist a dominant solution a_{dom} which is

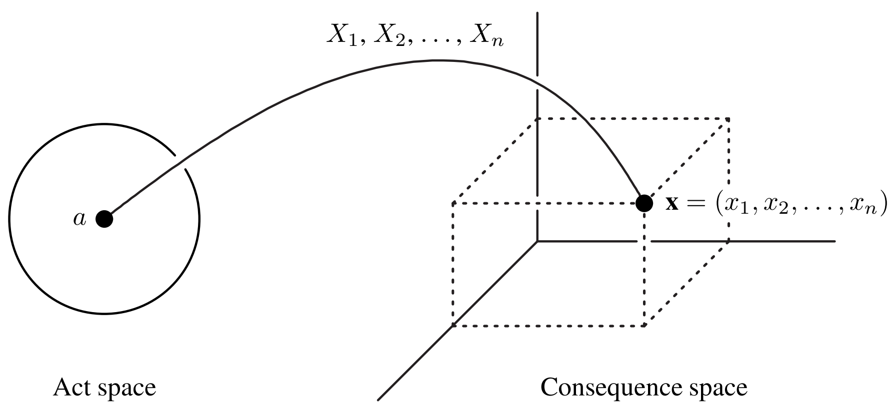

Figure 3.1: Mapping of acts into consequences, as adopted from (Keeney and Raiffa, 1993).

better than all alternatives $\tilde{a} \in \mathcal{A} \backslash \{a_{\text{dom}}\}$ in terms of the evaluators X_j, that is,

$$X_j(a_{\text{dom}}) \geq X_j(\tilde{a}) \ \text{ for all } j \tag{3.2}$$

and

$$X_j(a_{\text{dom}}) > X_j(\tilde{a}) \ \text{ for some } j \tag{3.3}$$

(In eqs. (3.2) and (3.3) we assumed, without any loss of generality, that the evaluators are defined in such a way, that higher values are preferred.)

In general, however, it is not possible to maximize all evaluators (or objectives) simultaneously. In such cases, a scalar-valued function $v(\cdot)$ has to be constructed, representing the decision maker's preference structure. That is, if the decision maker is indifferent between two consequences $\mathbf{x}'$ and $\mathbf{x}''$ (written $\mathbf{x}' \sim \mathbf{x}''$), then the value functions $v(\mathbf{x}')$ and $v(\mathbf{x}'')$ are equal, and vice versa:

$$\mathbf{x}' \sim \mathbf{x}'' \ \Leftrightarrow \ v(\mathbf{x}') = v(\mathbf{x}'') \tag{3.4}$$

And if the decision maker prefers outcome $\mathbf{x}'$ over outcome $\mathbf{x}''$ (written $\mathbf{x}' \succ \mathbf{x}''$), then $v(\mathbf{x}')$ is greater than $v(\mathbf{x}'')$, and vice versa:

$$\mathbf{x}' \succ \mathbf{x}'' \ \Leftrightarrow \ v(\mathbf{x}') > v(\mathbf{x}'') \tag{3.5}$$

With such a value function, the problem of choosing the best alternative a^* is put into the format of the standard optimization problem

$$a^* = \underset{a \in \mathcal{A}}{\operatorname{argmax}}\, v(\mathbf{x}) \tag{3.6}$$

Clearly, the difficulty of the original multi-objective decision problem is, thereby, shifted to the construction of the value function $v(\cdot)$. Some techniques for assessing multi-attribute value functions are given in (Keeney and Raiffa, 1993). Our focus later on, however, will not be on decisions in general, but on problems concerning the welfare of a community, a region or society at large. Thereby, the value function to be maximized, that is, welfare, is given by the net benefit function (Adler and Posner, 2006; Boardman et al., 2006).

Thus far, we have only treated the problem of decisions under certainty. In case of decisions with respect to optimal maintenance planning we have to make predictions about the future—predictions about future loads, structural deterioration, etc. These predictions are, in general, uncertain. Consequently, when we choose an act a, the resulting consequences of this act are also uncertain. This is shown in an illustrative way in Fig. 3.2 for the two acts a' and a''. Their respective uncertain (uni-variate) consequences x' and x'' are described by the probability density

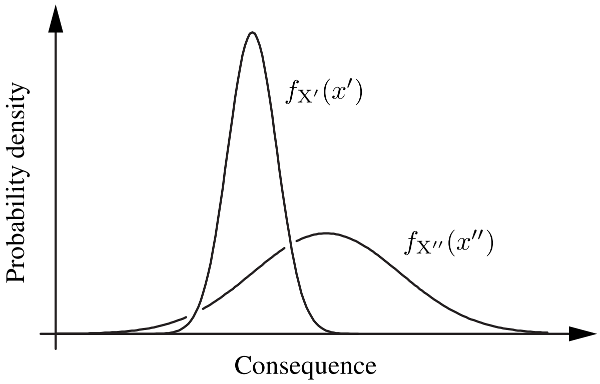

Figure 3.2: Illustration of uncertain consequences x' and x''.

functions $f_{X'}(x')$ and $f_{X''}(x'')$, respectively. Similar to the case of decisions under certainty, the decision maker's problem is again to choose between different acts $a \in \mathcal{A}$, however, now in such a way, that he will be 'happiest' with the resulting probability distribution function $F_{\mathbf{X}}(\mathbf{x})$ of the consequences $\mathbf{x}$. And as before, there may be cases when a solution a_{dom}, now, probabilistically dominates all alternatives (Keeney and Raiffa, 1993), but, in general, we need again a function $u(\cdot)$ representing the decision maker's preference structure. In case of decisions under uncertainty such a function is called utility function. The salient characterizing property of the utility function is, that the decision maker's preference structure with respect to different probability distribution functions is replaced by the comparision of expected utilities. Thus, the case of indifference is

$$F_{\mathbf{X}'}(\mathbf{x}') \sim F_{\mathbf{X}''}(\mathbf{x}'') \;\Leftrightarrow\; \mathrm{E}[u(\mathbf{x}')] = \mathrm{E}[u(\mathbf{x}'')] \tag{3.7}$$

and the case of preference is

$$F_{\mathbf{X}'}(\mathbf{x}') \succ F_{\mathbf{X}''}(\mathbf{x}'') \;\Leftrightarrow\; \mathrm{E}[u(\mathbf{x}')] > \mathrm{E}[u(\mathbf{x}'')] \tag{3.8}$$

where the expected utility is given as

$$\mathrm{E}[u(\mathbf{x})] = \int\limits_{\mathbb{R}^n} u(\mathbf{x})\,\mathrm{d}F_{\mathbf{X}} \tag{3.9}$$

Consequently, the best alternative a^* has the highest expected utility, that is,

$$a^* = \underset{a \in \mathcal{A}}{\operatorname{argmax}}\, \mathrm{E}[u(\mathbf{x})] \tag{3.10}$$

It should be noted, that the ordering of preferences as determined by the expected utility function $\mathrm{E}[u(\mathbf{x})]$ generally differs from the one by taking the expected value $\mathrm{E}[v(\mathbf{x})]$ of the value function—unless $v(\mathbf{x})$ is a linear transformation of $u(\mathbf{x})$ (Keeney and Raiffa, 1993; Schoemaker, 1982).

The idea of utility goes back to Bernoulli (1738), but it was Von Neumann and Morgenstern (1947) who showed that the utility function can be derived from a set of preference axioms: completeness, transitivity, continuity, monotonicity, and substitution (Von Neumann and Morgenstern, 1947; Schoemaker, 1982). That is, as long as all these preference axioms hold, then there exists an utility function such that an ordering of possible alternatives with respect to the expected utility fully coincides with the decision maker's actual preferences. It should be noted, that there exist also alternative formulations of preference axioms (see, for example, Fishburn, 1987; Luce and Raiffa, 1957; Pratt et al., 1964; Savage, 1954) which nevertheless result in the same expected utility theorem.

The utility functions $u(\cdot)$ differ in the way they represent the decision maker's preferences towards risky situations. For example, a decision maker may reject an uncertain situation (often called a lottery or a gamble) which would give him a certain expected value of return in favor of a smaller value he will receive for sure. In this case we say that the decision maker or his preference structure is risk-averse. In cost-benefit analysis it is assumed that the utility function is linear in wealth $v(\cdot)$. Hence, we can write without any loss of generality:

$$u(\mathbf{x}) = v(\mathbf{x}) \tag{3.11}$$

That is, society at large is risk-neutral. Eq. (3.11) also implies that decisions are not affected by changes in wealth position. Applying now the expected utility theorem for societal decision making means, that that alternative should be chosen which maximizes the expected wealth of society. For other utiliy functions for wealth and their implied preference structures see (Bell and Fishburn, 2000).

3.2 COST-BENEFIT ANALYSIS AS DECISION-AIDING RATIONALE

We now want to focus on those decisions which are related to the well-being or welfare of society. To ensure efficiency in resource allocation and to achieve a maximum rise in social welfare, it may be necessary to use decision-aiding rationales that are based on systematic and careful assessment of all alternatives or options under consideration. Decision-making theory distinguishes thereby between descriptive theories, which are concerned with understanding and predicting how decisions are actually made, and normative theories dealing with the problem of how people ideally should act (Fischhoff et al., 1981). Nevertheless, normative theories may also contain descriptive elements in the form of empirically observed or expressed values of beliefs, which, however, have to fulfill certain consistency requirements—the above mentioned preference axioms.

The most common types of decision-aiding approaches are displayed in Fig. 3.3,

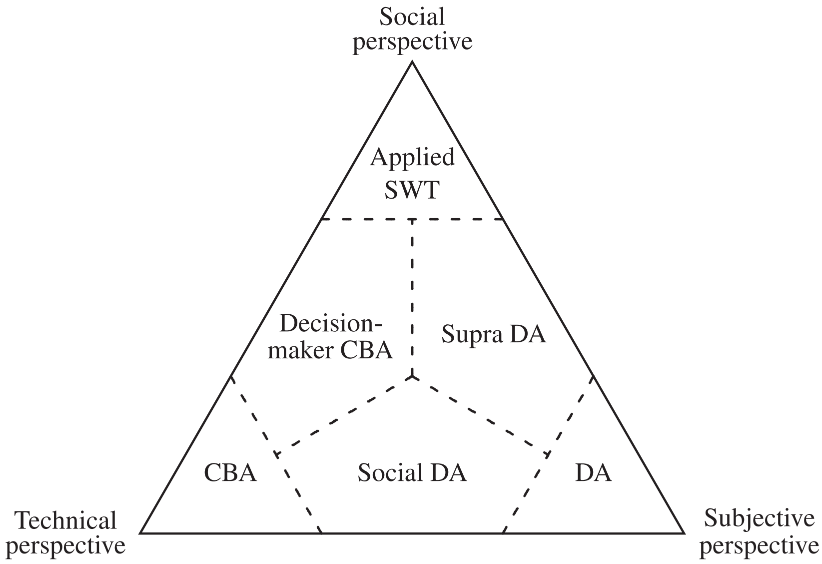

Figure 3.3: Partition of the space of decision-aiding rationales, as adopted from (Merkhofer, 1987). [CBA: cost-benefit analysis; DA: decision analysis; SWT: social welfare theory.]

schematically separating the space of rationales. The approaches differ in terms of the decision criterion adopted for identifying the 'best' alternative and in terms of procedures utilized to implement the respective criterion (Merkhofer, 1987). Thus, cost-benefit analysis utilizes the expected net present value as decision criterion and deduces preference structures from choices and prices observed in the marketplace—either directly or indirectly. Decision analysis, on the other hand, employs subjective value judgements, that is, utilities, and the preferred choice is indicated by the maximum expected utility, instead of an expected monetary value. Applied social welfare theory, finally, aggregates individual preferences to form a social welfare function to be maximized. According to the classification given in (Merkhofer, 1987), cost-benefit analysis has a more technical perspective, whereas decision analysis shows a more subjective one, and applied welfare theory a more social one.

Clearly, there is also a transition between or a blending of approaches. For example, a supra decision maker (Keeney and Raiffa, 1993) aggregates the individual utility functions to a social utility function, social decision analysis may use willingness-to-pay and efficiency arguments to construct an utility (or, more precisely, value) function, or a decision-maker cost-benefit analysis adopts procedures from social choice or decision theory for representing preferences. Similarly, the decision criterion associated with one approach can, in principle, be derived from each of the other theories by making appropriate assumptions. For example, cost-benefit analysis may be interpreted as a special kind of decision analysis whose utility function can be decomposed into costs and benefits, is measured in monetary terms and is linear therein (see previous Sec. 3.1).

Employing the (more technical) perspective of cost-benefit analysis in determining optimal solutions to maintenance implies by no means, that differing (non-technical) perceptions of risk (Starr, 1969; Slovic, 1987) should be ignored in the decision-making process. Quite to the contrary, taking into account public risk perception is a key component in successful risk communication and risk management (Fischhoff et al., 1981; Slovic, 2000). However, cost-benefit analysis is in our view normative in intent, not predictive or explanatory. Like any normative analysis, cost-benefit analysis articulates norms and principles that its agents 'ought' to follow. But 'ought' means 'can', not 'will' or 'have to'. As is convincingly argued in (Adler and Posner, 2006), cost-benefit analysis is not a super-procedure. It tracks overall well-being, not rights, distributive considerations, or putative moral factors. But this means also, that cost-benefit analysis has always to be imbedded in a political and institutional context, which allows to monitor those external components throughout the decision-making process and to determine whether alternatives are morally advisable in its entirety (see also Hubin, 1994; Sunstein, 2002).

The origin of applying cost-benefit to assess public expenditure decisions can be traced back to the U.S. Flood Control Act of 1936. After a series of disastrous floods in the 1920s and 30s, the U.S. Congress declared in Section 1 of the Flood Control Act (as cited in Arnold, 1988),

> "[...] that the Federal Government should improve or participate in the improvement of navigable waters or their tributaries, including watersheds thereof, for flood-control purposes *if the benefits to whomsoever they may accrue are in excess of the estimated costs*[a], and if the lives and social security of people are otherwise adversely affected."
>
> [a]Emphasis added.

This authorization of cost-benefit analysis became even more widespread by the U.S. Reclamation Project Act of 1939. In both cases, cost-benefit analysis has been seen as a useful source of information in supporting the decision process, not as its sole determinant (Fuguitt and Wilcox, 1999).

In the 1950s and 60s application of cost-benefit analysis in the U.S. not only widened to military expenditures and health care, although in the derivate form of cost-effectiveness analysis, but spread also to other countries like the U.K., where it was utilized for investment planning in transportation projects. In the 1980s and 90s, finally, cost-benefit analysis has also been incorporated into environmental evaluations as exemplified by the U.S. Clean Air Act Amendments of 1990, which allows marketable emission permits and auctions of emission rights. In Japan cost-benefit analyses are performed for diverse public projects as dams, roads, railways, or coastal fishing grounds. Also cost-benefit criteria are employed in the re-evaluation of government policies (Yamada, 2006). In summary, cost-benefit analysis became over the decades a standard economic tool in policy making, and its use in the future will rather increase than decline (Fuguitt and Wilcox, 1999).

3.3 BASIC ANALYSIS STEPS

To summarize, cost-benefit analysis is a quantitative approach that discloses any input to the decision-making process as well as any qualitative assumption made and, at the same time, provides formal criteria for identifying the 'best' alternative under situations of risk. For the cost-benefit analysis of maintenance interventions we distinguish four basic analysis steps:

i.) Modeling of all alternatives, that is, possible types or sequences of maintenance actions and their consequences.

Since in most cost-benefit analyses no explicit optimization is performed, usually only a small number of in advance already 'fixed' alternatives are compared among each other and the status quo. But in most technical problems, like maintenance planning for deteriorating structures herein, the alternatives can vary with respect to many design variables, for example, the time and number of maintenance interventions, the quality of inspection, the rehabilitation strategy, etc. Hence, we actually have to develop a 'parameterized' model of the maintenance process and its consequences.

ii.) Assignment of probabilities to alternative-consequence relationships, either based on past empirical data, or derived from stochastic models.

When planning future maintenance interventions we have to take into account all uncertainties with respect to the condition states of the structures and the imperfections of maintenance actions. Although there exist some empirical data on the deterioration of structures (as in the case of bridge management systems), future structural condition states depend on a such a multitude of factors like, for example, structural design, loads, environmental conditions, etc., but also on the maintenance interventions themselves, that it is not overly realistic to assume that there will be ever enough data available to describe stochastic deterioration without recursion to a physical or chemical model. Hence, structural reliability methods are mandatory to determine how the structure will evolve in time.

iii.) Monetization of all impacts, that is, all alternatives with their favorable as well as detrimental consequences (benefits and costs).

Benefits of infrastructure projects, for example, are reduction of travel or transportation times and, therewith, a reduction in transportation costs, number of accidents as well as environmental deterioration. But benefits can also be more indirect like in the case of economic strengthening of a certain region. Costs may include construction costs, costs for operation and maintenance, costs due to failure, etc. For being able to compare the benefits and costs, they have to be expressed in a common unit. In cost-benefit analysis this is some monetary unit like dollars, yens or euros. The monetary valuing is done via direct or indirect market methods (Boardman et al., 2006). It should be also noted, that when specifying benefits and costs this has to be done with due diligence to avoid ‘double counting’.

iv.) Optimization of maintenance interventions by maximizing the expected net present benefit.

The decision criterion—the objective function—in cost-benefit analysis is the expected net present benefit, that is, benefits minus costs—with proper discounting. This net present benefit should be positive for a project to get accepted. If there are several alternatives, then the project with the largest net present benefit is selected. Herein we will not just select the ‘best’ maintenance policy from a set of possible alternatives, but we will indeed optimize the maintenance interventions in such a way that the net present benefit is getting maximized throughout the structural lifetime.

The term ‘net present benefit’ in above analysis step *(iv.)* needs some further explanation. As already mentioned, the net benefit $g(t, t_0)$ throughout the time period $[t_0, t]$ is the difference between the benefits $b(t, t_0)$ and the losses or costs $l(t, t_0)$:

$$g(t, t_0) = b(t, t_0) - l(t, t_0) \tag{3.12}$$

In general, decisions have important consequences that extend over time, that is, we expect to derive benefits and incur costs over a number of years. However, for different projects these benefits and costs may arise in different future time periods, so that we are required to make intertemporal comparisons. Thereto, we discount future benefits and costs such that all monetary amounts are in a common metric, the so called present value (Boardman et al., 2006). If we take the beginning t_0 of the lifetime as the 'present', then the present value $\mathrm{PV}_{t_0}[\cdot]$ of the net benefit is

$$\mathrm{PV}_{t_0}\left[g(t,t_0)\right] = \mathrm{PV}_{t_0}\left[b(t,t_0) - l(t,t_0)\right] = \mathrm{PV}_{t_0}\left[b(t,t_0)\right] - \mathrm{PV}_{t_0}\left[l(t,t_0)\right] \quad (3.13)$$

In other words, the present value of the net benefit—the net present benefit—equals the difference between the present value of the benefits and the present value of the costs. How to calculate present values, that is, how to discount, is discussed next.

3.4 DISCOUNTING

Most discounting is done annually at an annual discount rate $\gamma \geq 0$. Hence, the present value $\mathrm{PV}_{t=0}$ of an impact $w(t)$—either a benefit $b(t)$ derived, or a cost $l(t)$ incured—after one year ($t = 1$) is determined as

$$\mathrm{PV}_{t=0}\left[w(t=1)\right] = \frac{w(t=1)}{1+\gamma} \quad (3.14)$$

Consequently, the present value of an impact $w(t = n)$ occuring after $t = n$ years is given as

$$\mathrm{PV}_{t=0}\left[w(t=n)\right] = w(t=n)\left[1+\gamma\right]^{-n} \quad (3.15)$$

The term $[1+\gamma]^{-n}$ is also called present value factor, or discount factor.

If we want to discount for a different resolution of time than years, we can do so by adjusting eq. (3.15). Let us assume we have a given time period $[t_0, t]$, and we want to discount ξ times per year. Then the present value is

$$\mathrm{PV}_{t_0}\left[w(t)\right] = w(t)\left[1+\frac{\gamma}{\xi}\right]^{-\xi(t-t_0)} \quad (3.16)$$

or after expansion

$$\mathrm{PV}_{t_0}\left[w(t)\right] = w(t)\left\{\left[1+\frac{\gamma}{\xi}\right]^{\xi/\gamma}\right\}^{-\gamma(t-t_0)} \quad (3.17)$$

In fact, in our following analysis we want to discount all monetary amounts continously. That is, we perform the following passage to the limit for $\xi/\gamma \rightarrow \infty$:

$$\lim_{\xi/\gamma \rightarrow \infty} \left[1 + \frac{\gamma}{\xi} \right]^{\xi/\gamma} = e \tag{3.18}$$

where e is the base of the natural logarithm, also sometimes called Euler's number, which equals 2.71828 truncated to five decimal places. Hence when continuously discounting, the present value of an impact $w(t)$ is given as

$$\mathrm{PV}_{t_0} \left[w(t) \right] = w(t) \exp[-\gamma(t - t_0)] \tag{3.19}$$

Thus far we tacitly assumed that the (annual) discount rate is undisputedly known. However, there exists quite a controversy in the literature about discounting and the appropriate values of discount rates to be utilized.[1] Proponents of discounting, that is, non-zero discount rates, argue mostly along the lines of opportunity cost or time preference (Boardman et al., 2006; Morrison, 1998). The basic opportunity-cost argument for discounting is, that decision-makers have more options to produce future benefits than just prolonging the status quo or implementing the project under investigation. Thus, discounting alerts the decision-maker to these additional alternatives. In other words, if discounted net benefits are less than zero, then there is at least a third option, for example, some different project, available that is better for overall welfare than the project currently under investigation.

In case of the social rate of time preference, economic theory hypothesizes, albeit backed by empirical evidence, that members of society prefer present welfare impacts to future ones. Thereby, the discount rate depends on the welfare function utilized, even though most popular welfare models result in a discount rate that contains two additive terms: one related to pure time preference and one related to economic growth (Boardman et al., 2006; Rackwitz, 2006). As pointed out in (Arrow et al., 1996; Morrison, 1998), the choice between the two competing approaches of opportunity cost and time preference is primarily a matter of regulatory policy.

As already mentioned above, there exists also a controversy whether social discounting should be performed at all. Opponents of discounting argue mainly along the line of intertemporal distributive fairness, that is, that future generations should not be worse off than present one, and vice versa (Cowen and Parfit, 1992; Heinzerling, 1999). As mentioned by Adler and Posner (2006), however, any discount rate,

[1]For a more than typical example of this controversy see the correspondence between Donohue III (1999) and Heinzerling (1999).

for example, also a zero value, may produce conflicts between overall welfare and distributive norms. Since cost-benefit analysis is only designed for tracking overall welfare, Adler and Posner (2006) suggest to accompany cost-benefit analysis by a separative distributive analysis to determine whether certain projects or policies are on balance morally advisable.[2]

3.5 COST-BENEFIT ANALYSIS OF MAINTENANCE INTERVENTIONS

3.5.1 Expected Net Present Benefit

With these prerequisites we can now formulate our maintenance optimization problem in the framework of cost-benefit analysis. For assessing the acceptability of a certain maintenance policy we have to take into account not only the expected maintenance costs m_j, but all expected future losses $l(t, t_0)$ and benefits $b(t, t_0)$ throughout the time interval $[t_0, t]$. Moreover, we include also the expected initial cost c_0 in our calculations to ensure that the structure has been originally designed in an economically reasonable way. All these cost factors are aggregated in the net benefit function $g(t, t_0)$ as

$$g(t, t_0) = b(t, t_0) - l(t, t_0) - \sum_{j=1}^{n} m_j - c_0 \tag{3.20}$$

Whereas the expected initial cost c_0 and the expected maintenance costs m_j accrue at discrete times t_0 and t_j, respectively, benefits and losses aggregate continuously throughout the entire lifetime. Utilizing the principle of cost-benefit analysis, the maintenance interventions will be justified when the total expected costs are outweighed by the total expected benefits.

For being able to calculate all benefits and costs, we need to have a description of structural deterioration in terms of directly or indirectly observable damage or deterioration states, which can be related to specific structural performance conditions—including their effect on the load carrying capacity and the remaining lifetime—as well as possible rehabilitation actions (Hearn, 1998a; Das, 1998). Moreover, since deterioration is uncertain over time due to the variability inherent in load effects and operating environments, it should be ideally modeled as a (monotonically increasing) stochastic process $\{X(t),\ t \geq t_0\}$, where t denotes time.

Let us describe the condition of the deteriorating structure by a finite number of, say, m discrete states: $1,\ 2, \ldots,\ m$. With such a description we can achieve

[2] In our following numerical analyses we will utilize a constant, non-zero discount rate. However, it should be mentioned, that our formulation is general. Thus, any type of discount rate, that is, also a zero discount rate, can be utilized.

a certain compatibility with the condition rating of existing infrastructure management systems (Hawk and Small, 1998; Thompson et al., 1998; De Brito et al., 1997; Söderqvist and Veijola, 1998; Miyamoto et al., 2001; Roelfstra et al., 2004). The states are numbered in our case in such a way, that state 1 corresponds to no deterioration, state 2 corresponds to minor deterioration, and so on, until state m, which denotes structural failure, that is, structural collapse. Thus, the probability distribution $\boldsymbol{\pi}(t)$ of being in one of the m possible states at time t is

$$\boldsymbol{\pi}(t) = \begin{bmatrix} \pi_1(t) \\ \pi_2(t) \\ \vdots \\ \pi_m(t) \end{bmatrix} = \begin{bmatrix} \Pr(X(t) = 1) \\ \Pr(X(t) = 2) \\ \vdots \\ \Pr(X(t) = m) \end{bmatrix} \tag{3.21}$$

The explicit inclusion of the failure state m as a condition state allows us to take into account the effect of structural deterioration on structural safety. This inclusion clearly differs from existing infrastructure management system, where failure is addressed only indirectly via the verbal description of the most severe condition state as 'requiring immediate action' (see, for example, Roelfstra et al., 2004; Scherer and Glagola, 1994).

For calculating the benefits and losses, two basic quantities of the deterioration process $X(t)$ are required:

i.) *Pointwise availability* (Barlow and Proschan, 1965): The probability that the structure is at a given instant of time t in one of the $(m-1)$ deterioration states, that is, that the structure is still operable. These probabilities are given as $\pi_k(t) = \Pr(X(t) = k)$ with $k = 1, 2, \ldots, m-1$.

ii.) *Number of failures per time interval*: The probability that the structure, which is still operable at time t, will fail in the following time interval $(t, t + \mathrm{d}t]$, where $\mathrm{d}t$ is a small time increment. This probability is given as $h(t)\mathrm{d}t = \dot{\pi}_m(t)\mathrm{d}t/(1 - \pi_m(t)) = \mathrm{d}\Pr(X(t) = m)/(1 - \Pr(X(t) = m))$.

With these two quantities we can calculate the costs and benefits which accrue during the existence of the structure, as well as the ones which occur in case of failure.

3.5.2 Expected Benefits

The reason why we use or operate a structure is, that we expect a certain benefit from doing so. For getting this benefit, the structure has to be usable or operable, that is, the structure has to be available. As mentioned abobe, the availability of the structure is determined by the probabilities $\pi_k(t)$ with $k = 1, 2, \ldots, m-1$. But this is only part of the picture in the benefit calculation, since the benefits generated may also depend on the condition of the structure. For example, due to wear and tear, fatigue, etc. bridges may have structural deficiencies which may affect structural safety. Bridge management systems deal usually with such deficiencies by imposing weight restrictions (Minchin, Jr. et al., 2006). Such weight restrictions, however, decrease the benefit we can derive from the structure, because this may require a re-routing of vehicles, resulting in extended travel or transportation times, additional accidents due to longer routes, increased environmental deterioration, and other economic or social losses (Minchin, Jr. et al., 2006; Sugimoto et al., 2002). Hence, the received benefit $\mathrm{d}b(t)$ per time interval $(t, t + \mathrm{d}t]$ is determined as the sum of the products of the state-dependent benefit rates $\dot{B}_k$ and the probabilities $\pi_k(t)$ of the structure being at time t in one of the deterioration states $k = 1, 2, \ldots, m-1$:

$$\mathrm{d}b(t) = \sum_{k=1}^{m-1} \dot{B}_k \pi_k(t) \mathrm{d}t \tag{3.22}$$

The problem of structural deficiency is aggravated by the additional problem of functional aging. By functional aging we mean the case that structures may still be technically sound, but may simply no longer meet current demands. For example, also comparatively recently constructed bridges and highways serve traffic loads and volumes which exceed by far the anticipated values (Minchin, Jr. et al., 2006), and thus require likewise management interventions as, for example, weight or speed limit restrictions—with all its economic consequences. Thus, the benefit rates in eq. (3.22) are, in general, not only state-, but also age- or time-dependent, such that the received benefit $\mathrm{d}b(t)$ per time interval $(t, t + \mathrm{d}t]$ is given as

$$\mathrm{d}b(t) = \sum_{k=1}^{m-1} \dot{B}_k(t) \pi_k(t) \mathrm{d}t \tag{3.23}$$

Since the received benefits arise at different future time intervals, we continuously discount all future benefits with the discount rate γ. Hence, the expected present benefit $b(t, t_0)$ for the time interval $[t_0, t]$, as used in eq. (3.20), is

$$b(t, t_0) = \int_{t_0}^{t} \exp(-\gamma\tau) \mathrm{d}b(\tau) = \sum_{k=1}^{m-1} \int_{t_0}^{t} \exp(-\gamma\tau) \dot{B}_k(\tau) \pi_k(\tau) \mathrm{d}\tau \tag{3.24}$$

3.5.3 Expected Losses

When using or operating a structure, we are well aware, that there is, in general, no risk-free structure. In other words, there is always a small probability that a structure may collapse. The probability that a structure collapses in the time interval $(t, t + \mathrm{d}t]$ is given as $\mathrm{d}\pi_m(t)$. This probability of failure can be expressed with the help of two quantities. First, the structure may only fail, if it has not failed so far. The probability of no failure until time t is $(1 - \pi_m(t))$, which is nothing else than the probability of the structure being still in one of its operable states, that is,

$$1 - \pi_m(t) = \sum_{k=1}^{m-1} \pi_k(t) \tag{3.25}$$

Second, the probability of failure, conditional that the structure is still operable at time t, is given by the failure rate $h(t)$ as $h(t)\mathrm{d}t$. Combining both quantities gives the probability of failure $\mathrm{d}\pi_m(t)$ in the time interval $(t, t + \mathrm{d}t]$:

$$\mathrm{d}\pi_m(t) = h(t)(1 - \pi_m(t))\mathrm{d}t \tag{3.26}$$

The failure consequences are composed of indirect failure costs, such as economic losses, and direct failure costs, such as life saving costs, environmental protection costs, costs for de-commissioning, etc. (Rackwitz, 2000; Rackwitz, 2004). All these cost factors are aggregated in the failure cost L. Thus, the loss $\mathrm{d}l(t)$ in the time interval $(t, t + \mathrm{d}t]$ is given as

$$\mathrm{d}l(t) = L\mathrm{d}\pi_m(t) = Lh(t)(1 - \pi_m(t))\mathrm{d}t \tag{3.27}$$

that is, the failure consequences are multiplied with the probability of failure. Like in the case of benefits, the losses may occur at different time intervals. Hence, also the losses have to be discounted to give present values. Thus, the expected present loss $l(t, t_0)$ for the time interval $[t_0, t]$, as used in eq. (3.20), is

$$l(t, t_0) = \int_{t_0}^{t} \exp(-\gamma\tau)\mathrm{d}l(\tau) = \int_{t_0}^{t} \exp(-\gamma\tau)Lh(\tau)(1 - \pi_m(\tau))\mathrm{d}\tau \tag{3.28}$$

3.5.4 Expected Maintenance Costs

Maintenance actions can be performed anytime before failure. The expected discounted maintenance costs m_j at time t_j are composed of inspection and rehabilitation costs. Thereby, rehabilitation work is always preceded by inspection. That is, we first have to assess the deterioration state of an existing structure by (mostly) non-destructive inspection techniques (Aktan et al., 1996; Rens et al., 1997; Hearn, 1998b; Pandey, 1998). However, such procedures are, in general, of imperfect nature due to limited resolutions, partial observability of the structure, measurement errors, humen errors, imperfect interrelations between measured and sought-for quantities, etc. That is, there is always the possibility given, that the inspection techniques do not disclose the actual condition state of the structure.

The standard method for characterizing and validating the quantified detection/discrimination capability of a non-destructive inspection technique is the probability of detection (Rummel, 1998; Achenbach, 2000). The probability of detecting a given deterioration state (event D) is a conditional probability, that is, $\Pr(D|X(t) = k, \alpha)$, in so far as the capability of the inspection technique depends, in general, on the degree of deterioration present, say, $X(t) = k$—in general, a higher degree of deterioration is more likely to be detected than only minor deterioration—and the amount of effort spent in assessing the structural condition. For example, utilizing different, partial complementary inspection methods (Horn and Mayo, 2000) may lead to an overall improved inspection capability, described herein by the parameter of inspection quality α, but may also result in higher inspection costs $I(\alpha)$. It should go without saying, that the inspection efforts can vary from maintenance to maintenance, that is, for each maintenance at time t_j we can have a different parameter α_j of inspection quality.

Since inspection or condition assessment makes only sense as long as the structure has not failed, the cost for inspection $I(\alpha_j)$ has to be multiplied by the probability of no failure $(1-\pi_m(t_j^-))$, where t_j^- is the time just prior to the j-th rehabilitation. Thus, the expected cost of inspection is $I(\alpha_j)(1-\pi_m(t_j^-))$. The rehabilitation costs depend on whether a damaged structure has been identified during inspection as being actually damaged. The probability that a structure has been identified as being in a certain damage state k is the product of the probability $\pi_k(t_j^-)$ of being in this state and the probability $\Pr(D|k, a_j)$ of just detecting such damage, that is,

$$\mathbf{D}_j \boldsymbol{\pi}(t_j^-) = \begin{bmatrix} 0 \\ \Pr(D|2, a_j)\pi_2(t_j^-) \\ \Pr(D|3, a_j)\pi_3(t_j^-) \\ \vdots \\ \Pr(D|m-1, a_j)\pi_{m-1}(t_j^-) \\ \pi_m(t_j^-) \end{bmatrix} \tag{3.29}$$

The $(m{\times}m)$-matrix $\mathbf{D}_j = \mathbf{D}(\alpha_j)$ in eq. (3.29) is defined as

$$\mathbf{D}_j = \operatorname{diag}\{0, \Pr(D|2, \alpha_j), \Pr(D|3, \alpha_j), \ldots, \Pr(D|m-1, \alpha_j), 1\} \tag{3.30}$$

where we assume, without any restrictions, that a failed structure (state m) is detected with probability one.

Having performed a condition rating of the structure, we have to decide to which extent rehabilitation work should be performed. Since we are interested in a more or less simple rehabilitation policy, which can be easily implemented in practice, we base our policy on a structural damage or deterioration threshold Δ. That is, if the condition assessment reveals, that the overall structural deterioration has reached the value, say, $X(t) = k$ and if this value equals or exceeds the threshold Δ, then the structure will be rehabilitated. If, however, $X(t) = k$ is less than the threshold Δ, then no action is performed. We call this deterioration threshold herein rehabilitation level. The above described maintenance policy as such can be interpreted as a variant of the classical failure limit policy (Barlow and Proschan, 1965; Pierskalla and Voelker, 1976; Valdez-Flores and Feldman, 1989; Wang, 2002).

From the above follows, that for determining the probability vector of the structural condition states which will be de-facto rehabilitated, we have to delete those entries of the probability vector of eq. (3.29) which are not related to any possible rehabilitation work. Hence, we pre-multiply the vector $\mathbf{D}_j\boldsymbol{\pi}(t_j^-)$ with a matrix $\mathbf{C}_j$ to get

$$\mathbf{C}_j\mathbf{D}_j\boldsymbol{\pi}(t_j^-) = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \Pr(D|\Delta_j, a_j)\pi_k(t_j^-) \\ \Pr(D|\Delta_j+1, a_j)\pi_{k+1}(t_j^-) \\ \vdots \\ \Pr(D|m-1, a_j)\pi_{m-1}(t_j^-) \\ 0 \end{bmatrix} \tag{3.31}$$

That is, the $(m{\times}m)$-matrix $\mathbf{C}_j = \mathbf{C}(\Delta_j)$, which can also vary from maintenance intervention to maintenance intervention, just filters out the deterioration states to be rehabilitated. It has the following diagonal form:

$$\mathbf{C}_j = \operatorname{diag}\{0, \ldots, 0, c_{\Delta_j\Delta_j} = 1, \ldots, 1, 0\} \tag{3.32}$$

where Δ_j denotes the rehabilitation level at the j-th maintenance intervention. That is, all deterioration states $k \geq \Delta_j$ are rehabilitated. The (m,m)-th entry in matrix $\mathbf{C}_j$ is equal to zero, because we do not re-build any failed structure.

The expected rehabilitation cost is then the scalar product of the probability vector $\mathbf{C}_j\mathbf{D}_j\boldsymbol{\pi}(t_j^-)$ and the rehabilitation cost vector $\mathbf{R} = [0, R_2, R_3, \ldots, R_{m-1}, 0]$, that is, $\mathbf{RC}_j\mathbf{D}_j\boldsymbol{\pi}(t_j^-)$. With proper discounting, the expected present maintenance costs at time t_j are finally

$$m_j = \exp(-\gamma t_j^-)\mathbf{RC}_j\mathbf{D}_j\boldsymbol{\pi}(t_j^-) + \exp(-\gamma t_j^-)I(\alpha_j)(1 - \pi_m(t_j^-)) \tag{3.33}$$

3.6 ACCEPTABLE FAILURE RATE AND OPTIMAL LIFETIME

Optimal solutions in structural design and maintenance planning are obtained by maximizing the expected net present benefit

$$g(T, t_0) = b(T, t_0) - l(T, t_0) - \sum_{j=1}^{n} m_j - c_0 \tag{3.34}$$

throughout a designated time horizon $[t_0, T]$, where T denotes the lifetime of the structure. As can be seen from eq. (3.34), the quantity $g(T, t_0)$ depends on the difference between expected benefits and expected costs. In fact, these factors are a description of our preferences towards the structure. Optimally these preferences have to be fulfilled for any time interval. In other words, we are indeed interested in maximizing the expected net present benefit rate

$$\dot{g}(t) = \sum_{k=1}^{m-1} \exp(-\gamma t)\dot{B}_k(t)\pi_k(t) - \exp(-\gamma t)Lh(t)(1 - \pi_m(t)) \tag{3.35}$$

for all t in $[t_0, T]$. Thus, eq. (3.35) evaluates the present trade-off between expected benefit and expected costs per unit time, whereas eq. (3.34) determines whether the investments in terms of expected initial costs c_0 and expected maintenance costs m_j are economically reasonable for the time horizon $[t_0, T]$.

For a structure to be acceptable with respect to our preferences, the expected benefit has to outweigh the expected costs for all t, that is, the expected net present benefit rate has to be positive:

$$\dot{g}(t) > 0 \tag{3.36}$$

Since $\exp(-\gamma t) > 0$, this condition can also be written in terms of the non-discounted, expected net benefit rate

$$\dot{g}_{\mathrm{n}}(t) = \sum_{k=1}^{m-1} \dot{B}_k(t)\pi_k(t) - Lh(t)(1 - \pi_m(t)) \tag{3.37}$$

as

$$\dot{g}_{\mathrm{n}}(t) > 0 \tag{3.38}$$

The requirement of a positive net benefit rate for all t is especially relevant since most of the quantities that influence the decision-making process change during the lifetime of a structure, such as failure rates, structural performance, operation costs or public satisfaction. It should also be noted that for assessing a structure as being acceptable according to eq. (3.38) only the present preferences towards the structure and the present state of the structure have to be known.

From eq. (3.38) follows that the lifetime T is reached as soon as the expected loss starts to be prevalent:

$$\dot{g}_{\mathrm{n}}(T) = 0 \tag{3.39}$$

The zero net benefit rate criterion of eq. (3.39) implicitly defines the acceptable failure $h_{\mathrm{a}}(t)$ rate of the structure. Rearranging eq. (3.37), that is, $\dot{g}_{\mathrm{n}}(t) = 0$, for the failure rate gives

$$h_{\mathrm{a}}(t) = \sum_{k=1}^{m-1} \frac{\dot{B}_k(t)}{L} \frac{\pi_k(t)}{1 - \pi_m(t)} \tag{3.40}$$

Thus, as long as the expected net benefit rate is positive, the failure rate $h(t)$ of the structure is less than the acceptable failure rate $h_{\mathrm{a}}(t)$:

$$h(t) < h_{\mathrm{a}}(t) \tag{3.41}$$

This shows the close relation between cost-benefit criteria and structural safety, since a violation of eq. (3.38) leads inevitable to a violation of eq. (3.41), and vice versa. If the benefit is state-independent, that is, $\dot{B}_k(t) = \dot{B}(t)$ for all k, the acceptable failure rate is simply given as

$$h_{\mathrm{a}}(t) = \frac{\dot{B}(t)}{L} \tag{3.42}$$

Eqs. (3.40) and (3.42) also show that a non-deteriorating structure will never overstay its lifetime as long as the preference towards the structure, in terms of benefit rates, and the demand on the structure, in terms of the failure rate $h(t)$, is not changing with time t. As already mentioned above, that this does not imply, that the structure will not fail at all, but the probability of failure per, say, year, under the condition, that the structure has not failed so far, is the same in each year, and is always less than what is deemed acceptable.

3.7 OPTIMIZATION PROBLEM

3.7.1 Maximization of Net Present Benefit Rate

As already mentioned, for a structure to be built, operated and maintained, all investments also have to be economically reasonable in total. Hence, the expected net present benefit has to be positive at the end of lifetime T:

$$g(T) > 0 \tag{3.43}$$

Moreover, maintenance interventions are only economically reasonable when the maintenance costs m_j are outweighed by the net benefit accumulated until the next maintenance intervention, that is,

$$b(t_{j+1}, t_j) - l(t_{j+1}, t_j) - m_j \geq 0 \tag{3.44}$$

or the end of lifetime T, whichever comes first. Thus the optimal maintenance plan for a given number n of maintenance interventions is found by solving the following maximization problem:

$$g^*(T^*) = \max_{\mathbf{t}^*, \boldsymbol{\Delta}^*, \boldsymbol{\alpha}^*, T^*} \int_{t_0}^{T} \dot{g}(t)\mathrm{d}t - \sum_{j=1}^{n} m_j - c_0 \tag{3.45}$$

subject to

$$g^*(T^*) \geq 0 \tag{3.46}$$

and

$$b^*(t^*_{j+1}, t^*_j) - l^*(t^*_{j+1}, t^*_j) - m^*_j \geq 0 \tag{3.47}$$

with $(\cdot)^*$ denoting optimal values. In eq. (3.45), $\mathbf{t}^* = \{t^*_1, t^*_2, \ldots, t^*_n\}$, $\boldsymbol{\Delta}^* = \{\Delta^*_1, \Delta^*_2, \ldots, \Delta^*_n\}$ and $\boldsymbol{\alpha}^* = \{\alpha^*_1, \alpha^*_2, \ldots, \alpha^*_n\}$ are the optimal sequences of maintenance times, rehabilitation levels and inspection qualities, respectively. It should be also noted, that the formulation of eq. (3.45) with respect to the lifetime T to be optimized imposes the end constraint

$$\dot{g}^*(T^*) = 0 \tag{3.48}$$

on the maximization problem. That is, the optimal lifetime T^* is reached as soon as the expected net present benefit rate becomes zero, or, equivalently, as soon as the failure rate $h(t)$ reaches the acceptable failure rate $h_\mathrm{a}(t)$.

3.7.2 Budget Constraints

In real applications there are quite often constraints with respect to the available budget on maintenance efforts (Miyamoto et al., 2003). In other words, for an optimal solution to be actually realizable the additional constraint

$$\tilde{m}^*_j \leq m_\mathrm{b} \tag{3.49}$$

has to be fulfilled, where m_b is the maximum available maintenance budget and $\tilde{(\cdot)}$ denotes the solution of the optimization problem of eq. (3.45) taking into account the constraints of eqs. (3.46), (3.47) and (3.49). However, whereas the constraints of eqs. (3.46) and (3.47) just assure that the expected future net benefits will outweigh the initial cost c_0 and the maintenance costs m_j, respectively, the constraint of eq. (3.49) indeed restricts the scope of maintenance actions and thus the long-term availability of the structure. Hence, a budget constraint, if active, leads, in general, to a sub-optimal solution, that is, smaller net benefits. The following scenarios are examples of possible consequences of budget constraints:

i.) For a given number n of maintenance interventions the lifetime $\tilde{T}^*$ is shortened, that is, $\tilde{T}^* \leq T^*$, since not all rehabilitation work necessary can be performed.

ii.) For a fixed time period $[t_0, T]$ the structure is, because of deterioration, from time to time in such a bad condition (that is, structural deficient), that—until budgetary funds become again available—it can be used only in a restricted way or even not at all (Minchin, Jr. et al., 2006; Miyamoto et al., 2003), at least not without violating safety standards.

iii.) Maintenance has to be performed more often ($\tilde{n} \geq n$), resulting in the long run in higher than necessary overall maintenance costs, that is, $\sum_{j=1}^{\tilde{n}} \tilde{m}_j^* \geq \sum_{j=1}^{n} m_j^*$.

The usefulness of the optimization formulation of eqs. (3.45) to (3.47), be it with or without eq. (3.49), is indeed, that it discloses this relation between benefits, costs and failure rates. Thus, if budget constraints exist, it also shows the resulting consequences: smaller net benefits, shorter lifetimes and reduced periods of usability, or, if no management intervention is performed, for example, in terms of weight restrictions or complete closure of bridges, the prolonged usage of non-acceptable, that is, potentially hazardous structures. In the following we will utilize the formulation without any budgetary constraint, that is, our optimization problem is given by eqs. (3.45) to (3.47) only.

4

Evolutionary Algorithms

4.1 MIXED-DISCRETE OPTIMIZATION PROBLEMS

There are many applications in structural engineering where we have to maximize (or minimize) an objective function $g(\mathbf{x})$ with respect to a vector **x** of design variables x_i ($i = 1, 2, \ldots, m$) which are not continuous, but which have to be selected from a given set of possible values. Typical examples of such discrete design variables are: member cross-sections to be taken from commercially available standard sizes, material properties which have to correspond to available materials, connectivity patterns of structural members, or in our case herein, number of maintenance actions to be performed, different rehabilitation strategies to be utilized, and so on (see Sec. 3.7.1). At first glance, it seems that, since there exist only a finite number of feasible points, optimization problems with discrete variables may be easier to solve than those with continuous variables. However, even in moderately sized problems the number of possible solutions is too large to make an exhaustive search for the best solution $\mathbf{x}^*$. Moreover, discrete optimization problems do not have the smoothness properties of continuous problems that allows us to use objective and constraint information at a point **x** to deduce information about the function's behavior at all points close to **x**. In other words, in the discrete case, points that are 'close' in some sense to **x** may have markedly different function values. Hence, discrete or mixed-discrete optimization problems are indeed far more difficult to solve than those involving only continuous variables (Arora et al., 1994; Arora, 2002).

The type of mixed-discrete optimization problem to be solved herein is characterized by the fact, that the discrete variables can not be simply 'simulated' as being continuous, that is, they can have only discrete values during optimization. Moreover, the objective function is not continuously differentiable. Hence, without performing the quite difficile task to approximate the optimization problem as one with a continuously differentiable objective function involving only continuous design variables, solution techniques like, for example, branch and bound, rounding-off, penalty approach, Lagrangian relaxation, or sequential linearization are not directly applicable (Arora, 2002; Thanedar and Vanderplaats, 1995). Thus, currently, only random search methods, such as simulated annealing, genetic algorithms, or evolutionary strategies, qualify as practicable, albeit computationally expensive techniques.

The most effective and customizable random search methods are presumably evolutionary algorithms. Historically, four different types of evolutionary algorithms originated: genetic algorithms (Holland, 1992), genetic programming (Koza, 1992), evolutionary strategies (Schwefel, 1995), and evolutionary programming (Fogel et al., 1966). Although developed separately, they are viewed nowadays just as algorithmic variants in evolutionary computing (Eiben and Smith, 2003). A historical review of the different evolutionary algorithms is given in (De Jong et al., 2000). Evolutionary algorithms are successfully applied for optimal sizing, shape and topology optimization in structural design (Cai and Thierauf, 1996; Grierson and Pak, 1993; Pezeshk and Kamp, 2002; Rafiq et al., 2003; Soh and Yang, 2000; Yang and Soh, 1997—among others), but also for scheduling problems (for example, Jong and Schonfeld, 2001; Vergara et al., 2002) and maintenance planning (for example, Levitin and Lisnianski, 2000).

The main difference between local random search methods and evolutionary algorithms is, that the latter operate simultaneously on a set of design vectors. This set, thereby, defines implicitly a non-uniform probability distribution for generating new design vectors. Possible interactions are taken into account by combining partial solutions from two or more members of the set. This clearly contrasts with the globally uniform distribution functions of simple random search methods, or the locally uniform distribution functions utilized by many other stochastic algorithms such as, for example, simulated annealing.

4.2 PRINCIPAL STRUCTURE OF EVOLUTIONARY ALGORITHMS

Although genetic algorithms, genetic programming, evolutionary strategies and evolutionary programming have been designed for different purposes with, in part, highly specialized operators, these algorithms, nevertheless, are based on the same two underlying ideas:

i.) Instead of a single design vector $\mathbf{x}$, evolutionary algorithms utilize a set $\mathcal{M} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(\mu)}\}$ of μ different design vectors $\mathbf{x}^{(j)}$ $(j = 1, 2, \ldots, \mu)$, which are processed simultaneously.

ii.) In place of the usual deterministic operators for finding a search direction and step size to improve the design vector, evolutionary algorithms utilize randomized ones like selection, recombination and mutation, which generate new and, hopefully, but not necessarily, improved design vectors.

The difference between variants of evolutionary algorithms are mainly in the emphasis they put on different operators and how they represent the optimization problem. For example, prototypical genetic algorithms utilize binary representations, whereas evolutionary strategies work directly with real-valued design vectors. Or, genetic algorithms emphasize recombination as the most important search operator and apply mutation only with a very small probability as a so called 'background operator', whereas evolutionary strategies qualify both mutation and recombination as essential operators and apply them not only to design vectors, but also to strategy parameters, such as, for example, mutation rates and distribution parameters. Nevertheless, these differences are mainly of historical origin, rather than an unalterable technical necessity (Eiben and Smith, 2003). Thus, in the following, we will not differentiate too strongly between the different variants, but we will give a description of the principle structure and main components of evolutionary algorithms.

Given an objective function $g(\mathbf{x})$ to be maximized with respect to the m design variables $\mathbf{x} = [x_1, x_2, \ldots, x_m]^{\mathrm{T}}$, a set $\mathcal{M}$ of μ different design vectors $\mathcal{M} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(\mu)}\}$ is created randomly—all of them possible solutions to the optimization problem. This set $\mathcal{M}$ then starts the evolutionary cycle, as sketched in Fig. 4.1 ('Initialization'). To decide how to proceed, the quality of all design vec-

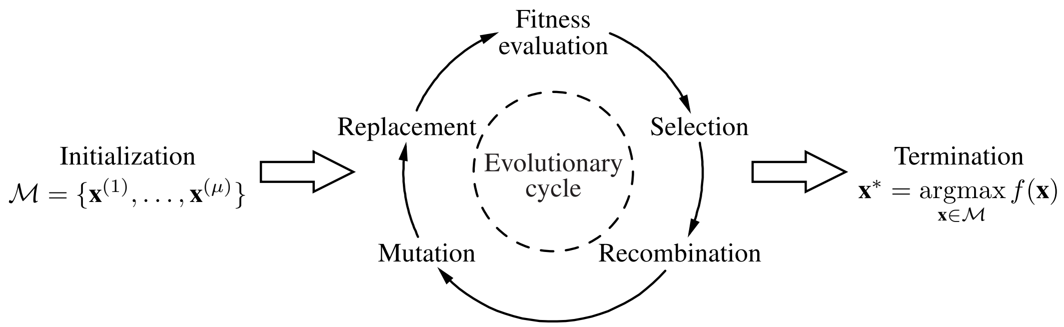

Figure 4.1: Principal structure of an evolutionary algorithm.

tors in $\mathcal{M}$ is determined (Fig. 4.1, 'Fitness evaluation'). Therefore, the objective function $g(\mathbf{x})$ is mapped to a fitness function $f(\mathbf{x})$, such that a 'better' design vector corresponds to a larger value of the fitness function than an 'inferior' design vector and, hence, has a higher probability of being selected for recombination and mutation. Although, in principle, the objective function $g(\mathbf{x})$, or a linear mapping of it, can be used directly as a fitness function, special types of fitness functions $f(\mathbf{x})$ are available, which allow to control the distribution of the selection probabilities and, as a consequence, the operation of the algorithm.

Having assigned to each design vector in $\mathcal{M}$ its selection probability, or at least its fitness value, λ design vectors are chosen from $\mathcal{M}$ to form a new set $\mathcal{L}$ (Fig. 4.1, 'Selection'). The selection is done most of the time by random sampling from $\mathcal{M}$ according to the selection probabilities, but also by fitness-related deterministic procedures operating on a randomly drawn subset of $\mathcal{M}$. The number λ of design vectors in $\mathcal{L}$ can be smaller, equal or larger than μ, that is, also multiple selections of the same design vector are, in principle, allowed to take place.

Thus far, the set $\mathcal{L}$ contains only design vectors which are also in $\mathcal{M}$, also on the average the design vectors in $\mathcal{L}$ are more 'fit'. To find new and 'better' solutions, evolutionary algorithms utilize two stochastic search operators: first recombination, and then mutation. Recombination generates new design vectors $\tilde{\mathbf{x}} \in \mathcal{L}'$ by exchanging information between two (or more) design vectors from the set $\mathcal{L}$ (Fig. 4.1, 'Recombination'). Obviously, the exchange of information is only effective, if at least parts of the design vectors chosen for recombination differ. Now, since we have only selected 'good' design vectors for being in $\mathcal{L}$, non-differing parts of the design vectors indicate subspaces of the design variable space, where on the average higher fitness values can be expected. Thus, recombination can be interpreted as a directed stochastic search operator.

After recombination, mutation is applied to the new design vectors $\tilde{\mathbf{x}}$ to get $\tilde{\tilde{\mathbf{x}}} \in \mathcal{L}''$ (Fig. 4.1, 'Mutation'). Whereas recombination is a directed search operator, mutation can be interpreted as a random walk. That is, the design vectors $\tilde{\mathbf{x}}$ are subjected with a certain probability to random perturbations, such that each feasible design vector can be reached—at least, in the long run. From the description of the working mechanisms of recombination and mutation, it follows that, for being effective, recombination has to be performed before mutation. However, there is no restriction about the number of different recombination or mutation operators that can be applied back-to-back.

To finish the evolutionary cycle, from the λ newly generated design vectors $\tilde{\tilde{\mathbf{x}}}$ in $\mathcal{L}''$ and the μ old design vectors $\mathbf{x}$ in $\mathcal{M}$, a new set $\mathcal{M}'$ has to be formed (Fig. 4.1, 'Replacement'). This set $\mathcal{M}'$, then, replaces $\mathcal{M}$ in the next cycle. There are two principle ways for replacement: age-based or fitness-based. In age-based replacement, the new design vectors (with $\lambda \leq \mu$) simply replace the old ones—independent whether the new design vectors have a higher fitness value or not. In fitness-based replacement, new design vectors are generated (with $\lambda \geq \mu$), and either the best μ new design vectors from $\mathcal{L}''$ are utilized to form the new set $\mathcal{M}'$, or the best μ design vectors from the unified set $\mathcal{L}'' \cup \mathcal{M}$.

The combined application of selection and variation (that is, recombination and mutation) generally leads to improving fitness values from evolutionary cycle to evolutionary cycle, until a design vector $\mathbf{x}^*$ is found that fulfills the termination criteria (Fig. 4.1, ‘Termination’). As already mentioned, the different variants of evolutionary algorithms do not differ with respect to the above described principal structure, but only with respect to which actual type of operators are used for its components. In Table 4.1 the components of prototypical genetic algorithms and

Table 4.1: Components of prototypical genetic algorithms and evolutionary strategies.

Component	Genetic algorithm	Evolutionary strategy
Representation	Binary	Real-valued
Fitness evaluation	Fitness-proportional scaling	Uniform selection probability
Selection	Random sampling	Random sampling
Recombination	Single-point crossover	Discrete or intermediary
Mutation	Bit flip	Perturbation with normal distribution
Replacement	Age-based	Fitness-based

evolutionary strategies are listed. In the following Secs. 4.3 to 4.8, we explain the most important operators and their working mechanism in more detail, before specifying in Sec. 4.9 the evolutionary algorithm we utilize herein.

4.3 FITNESS EVALUATION

4.3.1 Fitness-Proportional Scaling

In cost-benefit analysis our aim is to maximize the utility or net present benefit function $g(\mathbf{x})$. However, this function will not be non-negative for all $\mathbf{x}$ as required for the fitness function $f(\mathbf{x})$ in evolutionary algorithms. Thus it is necessary to map the function $g(\mathbf{x})$ to a fitness function $f(\mathbf{x})$ in one way or the other. The most simple form is certainly given by just adding all possible cost factors, such as, for example, the initial cost c_0 and the loss L, to the net present benefit function:

$$f(\mathbf{x}) = g(\mathbf{x}) + c_0 + L \geq 0 \tag{4.1}$$

Utilizing the above mapping, the probability for the design vector $\mathbf{x}^{(j)}$ (with $j = 1, 2, \ldots, \mu$) to be selected from population $\mathcal{M} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(\mu)}\}$ is

$$\Pr(j) = \frac{f(\mathbf{x}^{(j)})}{\sum\limits_{k=1}^{\mu} f(\mathbf{x}^{(k)})} \tag{4.2}$$

The probability of the design vector $\mathbf{x}^{(j)}$ to be selected for reproduction or reinsertion is proportional to its fitness value as compared to the fitness values of all other elements in $\mathcal{M}$. However, this type of fitness-proportional scaling has some known problems (Eiben and Smith, 2003):

i.) Design vectors with large fitness values take over the entire population very quickly leading to premature convergence of the algorithm.

ii.) On the other hand, when the fitness values of the design vectors are all very similar, then there is almost no selection pressure and the improvement of design vectors progresses quite slowly.

iii.) Transposed versions of the same objective or fitness function lead to different behaviors of the algorithm, that is, its performance is scaling dependent.

To level down above effects on evolutionary algorithms, different modifications of the fitness scaling procedure have been suggested. These range from logarithmic, power law, or exponential mappings of the objective function, depending on whether the fitness values should be more centered or more spread out, to dynamic scaling procedures utilizing statistical estimators of the gain distribution of the current or last few populations (Grefenstette, 2000a). A typical example of the latter is sigma scaling (Goldberg, 1989), defined as

$$f(\mathbf{x}) = \max\left\{ g(\mathbf{x}) - \mathrm{E}[g(\mathbf{x})] + \xi\,\sqrt{\mathrm{Var}[g(\mathbf{x})]}\,, 0 \right\} \tag{4.3}$$

with ξ being a scaling parameter usually chosen between 1 and 3. Nevertheless, none of the above mentioned modifications allows a robust control of the fitness values and, therewith, the selection pressure in general settings, that is, without prior knowledge of the population dynamics. Therefore, fitness-proportional scaling has been abandoned more and more in recent history and has been nowadays replaced almost exclusively by rank-based schemes.

4.3.2 Rank-Based Scheme

For any rank-based scheme the μ design vectors $\mathbf{x}$ in $\mathcal{M}$ are sorted increasingly with respect to the values of their objective functions (or fitness values). Then the rank k is assigned to the design vector $\mathbf{x}^{(j)}$ according to its (sorted) position k, with rank 1 being assigned to the worst design vector and rank μ to the best one. Note, that all design vectors get a different rank, even if they have the same objective value. Ranking eliminates the problems of fitness-proportional scaling (premature convergence, loss of selection pressure, dependency on formulation of fitness function) by preserving a constant selection pressure, since the selection probabilities are assigned solely with respect to the rank of the design vectors.

The mapping from the rank to the selection probability can be done again in different ways. The most common mappings are linear and exponential ranking (Grefenstette, 2000b). In the former, the selection probability for the design vector $\mathbf{x}^{(j)}$ is

$$\Pr(j) = \frac{1}{\mu}\left\{(2-\xi) + 2(\xi - 1)\,\frac{\mathrm{rank}(\mathbf{x}^{(j)}) - 1}{\mu - 1}\right\}, \quad 1 \le \xi \le 2 \tag{4.4}$$

Hence, the probability of an average good design vector to be selected is μ^{-1}, whereas the probabilities of the best and worst design vector being selected are, respectively, $\xi\mu^{-1}$ and $(2-\xi)\mu^{-1}$. If a higher selection pressure is required, an exponential ranking scheme is often used. In this case the selection probability is defined as

$$\Pr(k) = \frac{(1-\xi)\,\xi^{\mu - \mathrm{rank}(\mathbf{x}^{(j)})}}{1-\xi^{\mu}}\,, \quad 0 < \xi < 1 \tag{4.5}$$

The closer ξ is to 1, the lower the 'exponentiality' of the ranking procedure.

4.4 SELECTION

4.4.1 Roulette Wheel Selection

Selection is one of the main operators used in evolutionary algorithms. There are, in general, two instances in the algorithm where a selection mechanism is utilized. The first instance is, when we derive for recombination a set $\mathcal{L}$ of λ 'good' vectors from a given set $\mathcal{M}$ of μ design vectors. This is most typical for genetic algorithms. The second instance, being more typical for evolutionary strategies, is in the replacement phase, that is when we join the sets $\mathcal{L}$ and $\mathcal{M}$ to form a new, 'better' set $\mathcal{M}'$.

One way to perform the selection, is simply by sampling randomly from the set $\mathcal{M} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(\mu)}\}$. Since the probability of selecting the design vector

$\mathbf{x}^{(j)}$ should be proportional to $\Pr(j)$, we generate a random variate U, which is uniformly distributed over $(0, 1)$, and set

$$\mathbf{X} = \begin{cases} \mathbf{x}^{(1)} & \text{if } U < \Pr(1) \\ \mathbf{x}^{(2)} & \text{if } \Pr(1) \le U < \Pr(1) + \Pr(2) \\ \vdots & \\ \mathbf{x}^{(j)} & \text{if } \sum_{k=1}^{j-1} \Pr(k) \le U < \sum_{k=1}^{j} \Pr(k) \\ \vdots & \\ \mathbf{x}^{(\mu)} & \text{if } \sum_{k=1}^{\mu-1} \Pr(k) \le U \end{cases} \tag{4.6}$$

where $\mathbf{X}$ is our (random) selection. This procedure is repeated, until the required number λ of selections is made. The expected number $\mathrm{E}[N^{(j)}]$ of times the design vector $\mathbf{x}^{(j)}$ is thereby chosen is $\mathrm{E}[N^{(j)}] = \lambda \Pr(j)$.

In evolutionary computing, such sampling scheme is called roulette wheel selection, since we can think of the probability distribution as defining a roulette wheel. That is, each design vector is represented by a pocket of the wheel, where the size of the j-th pocket corresponds to the selection probability $\Pr(j)$. The sampling can then be visualized as spinning the one-armed roulette wheel by a random amount $U \in (0, 1)$ and selecting the design vector whose pocket ends up at the top. As an example, let us select $(\lambda = 2)$ times a design vector from the set $\mathcal{M} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(9)}\}$ (see Fig. 4.2). Thus, we spin the wheel by the random

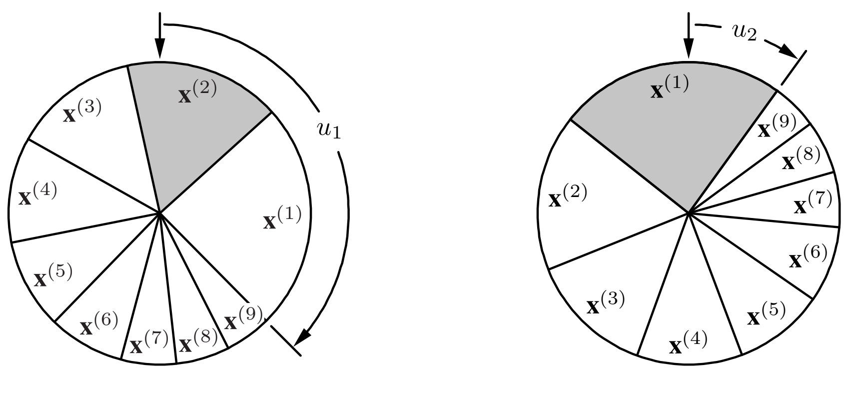

Figure 4.2: Example of roulette wheel selection ($\lambda = 2$).

amount $U_1 \in (0, 1)$ and select the design vector in the upper position, that is in our example $\mathbf{x}^{(2)}$. Then we turn back the wheel to its initial position and spin it again by a random amount $U_2 \in (0, 1)$. This results, finally, in the set of selected design vectors $\mathcal{L} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\}$.

4.4.2 Stochastic Universal Sampling

When selecting λ design vectors from the set $\mathcal{M}$ by making λ calls to the roulette wheel procedure, then the number $N^{(j)}$ of times an individual vector $\mathbf{x}^{(j)}$ is chosen shows a large variance. To reduce this variance, Baker (1987) developed an algorithm called stochastic universal sampling. Therefore, only a single random variate U_1 is generated from the uniform distribution in $(0, \lambda^{-1})$. The remaining variates $U_2, U_3, \ldots, U_\lambda$ are then calculated systematically by

$$U_i = U_1 + \frac{i-1}{\lambda}\,, \quad i = 2, 3, \ldots, \lambda \tag{4.7}$$

Conceptually, this is equivalent to making a single spin of the roullette wheel with λ equally spaced arms, rather than λ spins with a one-armed wheel. Obviously, for $\lambda = 1$, roulette wheel selection and stochastic universal sampling are identical.

Fig. 4.3 shows an exemplary realization of stochastic universal sampling. The

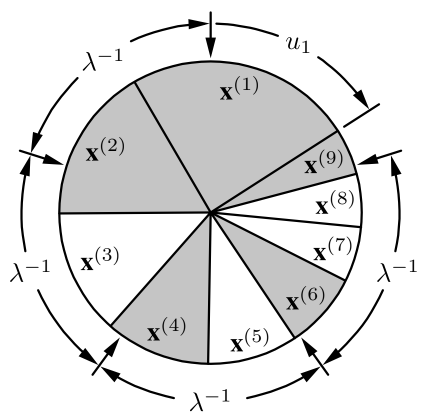

Figure 4.3: Example of stochastic universal sampling ($\lambda = 5$).

set $\mathcal{M}$ consists of $\mu = 9$ design vectors $\mathcal{M} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(9)}\}$ with different selection probabilities $\Pr(j)$. For selecting $(\lambda = 5)$ times a design vector, the roulette wheel is spun once by a random amount $U_1 \in (0, \lambda^{-1})$. The design vector in the upper position is selected first. The remaining design vectors are then found by rotating the wheel $(\lambda - 1)$ times by the (deterministic) amount λ^{-1}, finally resulting in the set $\mathcal{L} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(4)}, \mathbf{x}^{(6)}, \mathbf{x}^{(9)}\}$.

For stochastic universal sampling, the expected number of times the design vector $\mathbf{x}^{(j)}$ is chosen is $\mathrm{E}[N^{(j)}] = \lambda \Pr(j)$, as before for the roulette wheel selection. Due to the systematic nature of the scheme, however, the actual number of times $\mathbf{x}^{(j)}$ is selected in a call of the procedure is either $\lfloor \lambda \Pr(j) \rfloor$ or $\lceil \lambda \Pr(j) \rceil$, thus reducing the variance drastically. Moreover, if stochastic universal sampling is utilized in combination with a ranking scheme, also multiple selection of design vectors can be easily controlled.

4.4.3 Tournament Selection

The previous two selection methods are based on sampling λ times from the entire set $\mathcal{M}$ according to the explicitly given selection probabilities $\Pr(j)$. A selection mechanism, which only utilizes an implicit definition of selection probabilities, is tournament selection. Thereto, a set $\mathcal{G} \subset \mathcal{M}$ of γ design vectors are randomly drawn from $\mathcal{M}$, that is, the design vectors are chosen with equal probability μ^{-1}. Then, the design vectors in $\mathcal{G}$ take part in a so called tournament. That is, they are compared pairwise according to their fitness or rank, and the lower ranking design vector is removed from the set $\mathcal{G}$ (deterministic tournament), or, a game of chance is played to select the design vector to be deleted, with the higher ranking design vector having a higher probability of winning the game (stochastic tournament). The winner of the tournament is then inserted in the set $\mathcal{L}$. The procedure is repeated λ times.

A description of the main properties of tournament selection is given in (Blickle, 2000). Tournament selection is translation and scaling invariant, that is, scaling techniques as used for fitness-proportional selection are not required. The selection probability is implicitly controlled via the tournament size γ. If the tournament size is larger, 'bad' design vectors have a smaller chance of being selected. In most applications γ is chosen between 6 and 10 (Blickle, 2000). Since tournament selection only needs knowledge about the relative ranking of the design vectors in set $\mathcal{G}$, but not of the absolute ranking with respect to set $\mathcal{M}$, it is widely used in parallel processing applications. However, the algorithm leads, like roulette wheel selection, to a high variance in the expected number of times a design vector is chosen, because λ independent trials are carried out from the implicitly given probability distribution.

4.4.4 Truncation Selection

A selection method which is, in general, associated with evolutionary strategies, and mostly applied in the replacement phase, is truncation selection. Truncation selection is a deterministic rank-based scheme. The μ design vectors $\mathbf{x}^{(j)}$ in the set

$\mathcal{M}$ are given by their rank $\mathrm{rank}(\mathbf{x}^{(j)})$ for $j = 1, 2, \ldots, \mu$. Then a fraction $\lambda\mu^{-1}$ is specified, such that only the 'best' λ design vectors are selected for the set $\mathcal{L}$, that is,

$$\mathcal{L} = \left\{\mathbf{x}^{(j)}:\ \mathrm{rank}(\mathbf{x}^{(j)}) > \mu - \lambda\right\} \tag{4.8}$$

There exists also a stochastic version of the method (Grefenstette, 2000b). In this case the following selection probabilities

$$\Pr(j) = \begin{cases} \lambda^{-1} & \text{if } \mathrm{rank}(\mathbf{x}^{(j)}) > \mu - \lambda \\ 0 & \text{otherwise} \end{cases} \tag{4.9}$$

are assigned to the design vectors $\mathbf{x}^{(j)}$. The assignment of the λ design vectors to the set $\mathcal{L}$ is done in this case by roulette wheel selection. Since in both versions of truncation selection the 'worst' design vectors are excluded from being selected, truncation selection is clearly biased.

4.5 RECOMBINATION

4.5.1 Binary Crossover

What distinguishes evolutionary algorithms fundamentally from gradient-based optimization methods are their search operators. Instead of utilizing gradients of the objective and constraint functions to arrive at an improved design vector, as is the case in gradient-based optimization methods (see, for example, Fletcher, 1987; Nocedal and Wright, 1999), evolutionary algorithms generate new and, hopefully, but not necessarily, 'better' solutions directly from the set $\mathcal{L}$ of different design vectors. This creation of new design vectors is done, thereby, in two different ways: either by recombination, or by mutation.

Recombination or crossover combines information of different design vectors by swapping information between them. The basic idea behind recombination is like this: having given two design vectors with good performance but different features, we would like to generate a new design vector which combines the best features from each. Since we do not know which features are responsible for the good performance, the best thing we can do is to perform the recombination at random. Clearly, some of these new design vectors will have undesirable combinations of traits, most of them may be no better or worse than their origins, but, nevertheless, there is a certain chance that some may have an improved performance.

Traditionally, recombination is done for binary representations of design parameters (Goldberg, 1989). Having given a pair $(s^{(j)}, s^{(k)})$ of strings, each having l bits, a common random position $U \in [1, 2, \ldots, l-1]$ along the two strings is chosen and the bits from one side of this position are swapped between the two strings to create a new pair $(\tilde{s}^{(j)}, \tilde{s}^{(k)})$ of strings (see Fig. 4.4). Obviously, this operator—called

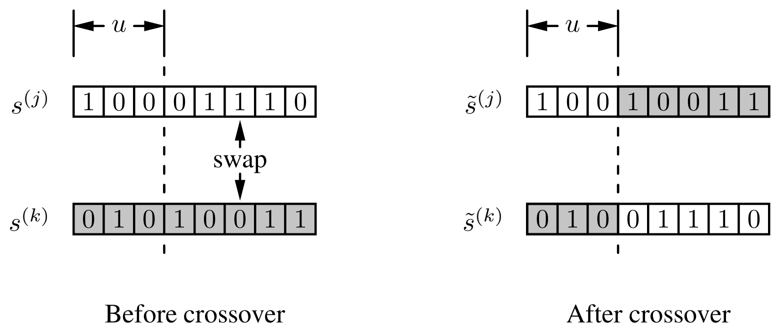

Figure 4.4: Principle of single-point crossover (8-bit string).

single-point crossover—can be generalized to multi- or n-point crossover. In this case the strings are cut into $n+1$ segments, which are then swapped alternately. There exist even operators which utilize more than two strings for recombination (Eiben and Smith, 2003).

The overall effect of recombination is that of focusing the search on hyperplanes whose points (that is, parts of the bit string) show on average the best performance. Over time this search becomes more and more focused as more and more elements of the strings converge and the creation of new strings is restricted to the remaining variation in these strings. To avoid premature convergence, recombination is, in general, combined with mutation, since the latter produces variations independent of the convergence status of the string elements.

In our problem of maintenance optimization, the design variables have integer- or real-valued representations. Hence, to apply binary crossover, we would have to binary encode these variables before performing crossover, and then decode them again after crossover. As an example, let us encode an integer-valued variable $x \in [0, 1, \ldots, 255]$ in an 8-bit string s. Two possible 8-bit strings $s^{(j)}$ and $s^{(k)}$ are given in Fig. 4.4. There respective integer values $x^{(j)}$ and $x^{(k)}$ are depicted in Fig. 4.5 as solid circles. Fig. 4.5 displays also all seven possible outcomes $(\tilde{x}^{(j)}, \tilde{x}^{(k)})$ of single-point crossover as open circles. It should be noted that two of these pairs are doubly, since the two strings $s^{(j)}$ and $s^{(k)}$ have at two positions the same bit (that is, the third and the seventh bit are identical).

As can be shown, the arithmetic mean of the decoded values of the original integer pair $(x^{(j)}, x^{(k)})$ equals the arithmetic mean of each integer pair $(\tilde{x}^{(j)}, \tilde{x}^{(k)})$

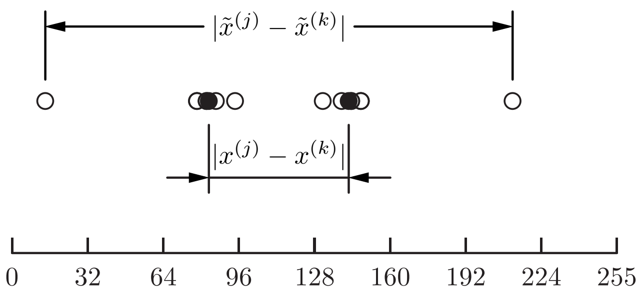

Figure 4.5: Spread of single-point crossover (using the two 8-bit strings of Fig. 4.4).

after crossover (Deb and Agrawal, 1995), that is

$$\frac{x^{(j)} + x^{(k)}}{2} = \frac{\tilde{x}^{(j)} + \tilde{x}^{(k)}}{2} \tag{4.10}$$

Moreover, the pair $(\tilde{x}^{(j)}, \tilde{x}^{(k)})$ lies either inside or outside the region bounded by the original pair $(x^{(j)}, x^{(k)})$. As can also be seen from Fig. 4.5, most of crossed-over pairs $(\tilde{x}^{(j)}, \tilde{x}^{(k)})$ lie in the neighborhood of the original pair $(x^{(j)}, x^{(k)})$. That is, if the original pair has spread $|x^{(j)} - x^{(k)}|$, then it is more likely that the crossed-over pairs have spread $|\tilde{x}^{(j)} - \tilde{x}^{(k)}|$ of similar magnitude. The only exception in the present example is when the crossover is performed after the first bit. In this case the spread is much larger than the original spread. This is a known problem from (natural) binary coding, that is, different bits have so called different significance (Goldberg, 1989; Eiben and Smith, 2003).

A variation of the natural binary coding from above is Gray coding, which is a binary numeral system where two successive values differ in only one bit. To see the effect of different binary encoding schemes on the spread after crossover, Figs. 4.6 and 4.7 display all possible combinations for an integer-valued variable $x \in [0, 1, \ldots, 63]$ encoded for a 6-bit natural binary code and 6-bit Gray code, respectively. As can be seen, for Gray coding the possible spreads after crossover are more confined to the region around the spread of the original pair as is the case for natural binary coding.

4.5.2 Simulated Binary Crossover

Instead of utilizing binary encdoing and decoding to be able to perform crossover of integer- and real-valued representations of design variables, specially constructed recombination operators can also be applied directly to the original variables. One of these crossover operators is called simulated binary crossover (Deb and Agrawal,

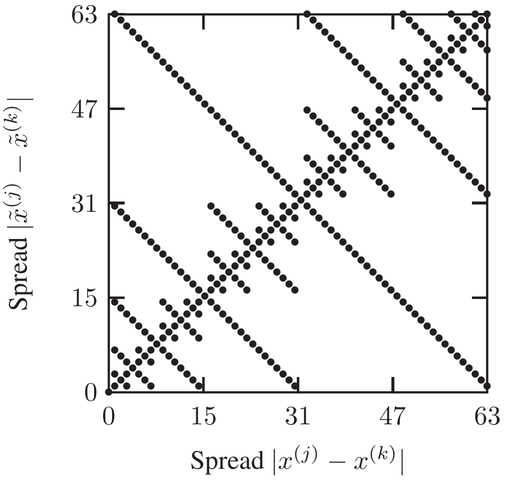

Figure 4.6: Possible combinations of spreads $|x^{(j)} - x^{(k)}|$ and $|\tilde{x}^{(j)} - \tilde{x}^{(k)}|$ for 6-bit natural binary code.

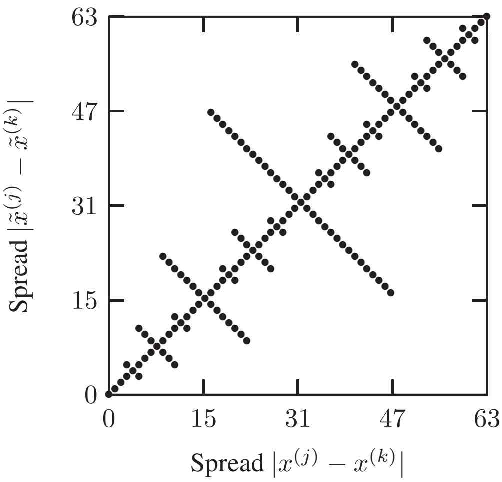

Figure 4.7: Possible combinations of spreads $|x^{(j)} - x^{(k)}|$ and $|\tilde{x}^{(j)} - \tilde{x}^{(k)}|$ for a 6-bit Gray code.

1995), which can be applied to real-valued pairs $(x^{(j)}, x^{(k)})$. Simulated binary crossover is constructed to mimic binary crossover, that is:

i.) The arithmetic mean of the original pair equals the arithmetic mean of each crossed-over pair.

ii.) Each crossed-over pair lies either inside or outside the region bounded by the original pair.

iii.) There is a higher likelihood that the crossed-over pair lies in the vicinity of the original pair, than more far away.

Whereas the first two conditions can be easily fulfilled by using a linear combination of the arithmetic mean $(x^{(j)} + x^{(k)})/2$ and the spread $|x^{(j)} - x^{(k)}|$, the third condition is taken into account by a probability distribution function. Hence, for a given real-valued pair $(x^{(j)}, x^{(k)})$, the crossed-over pair $(\tilde{x}^{(j)}, \tilde{x}^{(k)})$ is generated by

$$
\begin{aligned}
\tilde{x}^{(j)} &= \frac{x^{(j)} + x^{(k)}}{2} + \xi \, \frac{x^{(j)} - x^{(k)}}{2} \\
\tilde{x}^{(k)} &= \frac{x^{(j)} + x^{(k)}}{2} - \xi \, \frac{x^{(j)} - x^{(k)}}{2}
\end{aligned}
\tag{4.11}
$$

where

$$
\xi = \begin{cases} (2u)^{1/(1+\nu)} & \text{if } u \leq 1/2 \\ [2(1-u)]^{-1/(1+\nu)} & \text{otherwise} \end{cases}
\tag{4.12}
$$

In eq. (4.12) u denotes a random sample drawn from U uniformly distributed over $(0, 1)$, and $\nu > 0$ is a parameter controlling the variation of the spread. Fig. 4.8

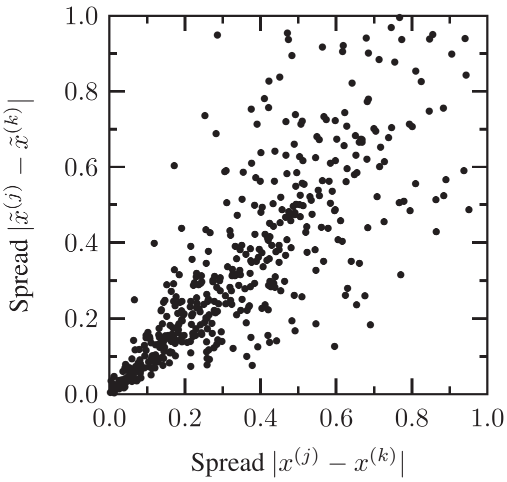

Figure 4.8: Scatter plot of spreads $|x^{(j)} - x^{(k)}|$ and $|\tilde{x}^{(j)} - \tilde{x}^{(k)}|$ for simulated binary crossover of two independent and uniformly distributed random variates $X^{(j)}$ and $X^{(k)}$ ($\nu = 2$, $\lambda = 1000$).

shows a scatter plot for a real-valued random vector $\mathbf{X} = [X^{(j)}, X^{(k)}]^{\mathrm{T}}$, whose components are independent and uniformly distributed over $(0, 1)$. The preference of the crossed-over pairs to lie in the vicinity of the original pairs can be clearly seen. It should also be noted, that the crossed-over variates $\tilde{X}^{(j)}$ and $\tilde{X}^{(k)}$ can have values—and, therewith, spreads—outside the range $(0, 1)$.

4.6 MUTATION

Mutation is a stochastic operator which causes a random, unbiased change to a variable. As in the case of recombination, mutation operates, traditionally, on binary codes. Different to recombination, however, mutation utilizes only a single bit-string, that is, the mutation is peformed without taking into account the properties of any other bit-string. Mutation allows, thereby, each bit to flip, that is, to change its value from 1 to 0, or from 0 to 1, with a small probability, the so called bitwise mutation rate. An illustration of this procedure is shown in Fig. 4.9 for an 8-bit string $\tilde{s}^{(j)}$, where the third and the sixth bit are flipped.

To apply mutation to integer- or real-valued design vector $\tilde{\mathbf{x}}^{(j)}$, we have, again, to investigate the properties of binary mutation. Obviously, mutation is a random walk through the string space. For example, in Fig. 4.10 fifty mutations of an integer-valued pair $(\tilde{x}^{(j)}, \tilde{x}^{(k)})$ are shown. The initial pair $(\tilde{x}^{(j)} = 31$, $\tilde{x}^{(k)} = 31)$ and its mu-

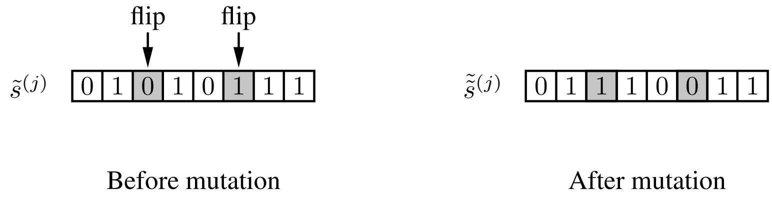

Figure 4.9: Principle of mutation (8-bit string).

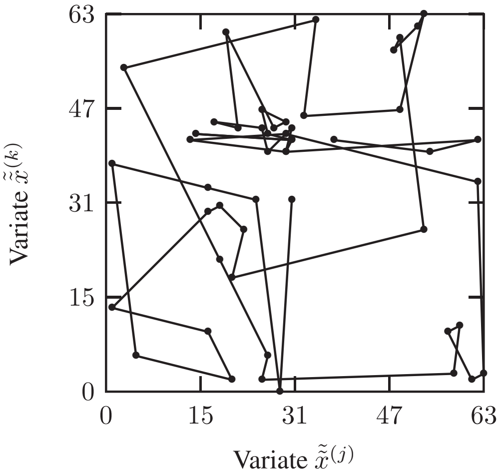

Figure 4.10: Mutation walk (50 mutations, 6-bit natural binary code).

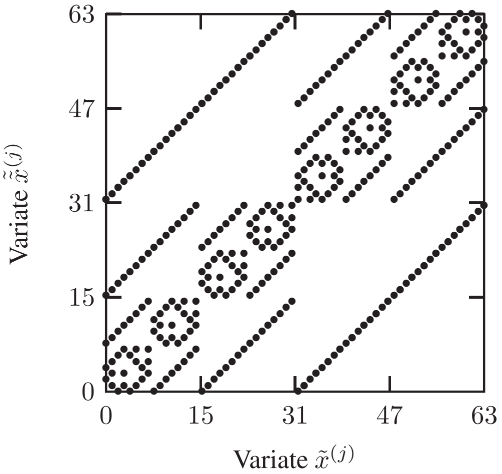

Figure 4.11: Possible mutations for a 6-bit natural binary code.

tations $\tilde{\tilde{x}}^{(j)}, \tilde{\tilde{x}}^{(k)} \in [0, 1, \ldots, 63]$ are encoded as a 6-bit natural binary code. Only one bit of each component is flipped in each mutation step. As can be seen, the mutation walk tends to fill in the entire integer space $[0, 63] \times [0, 63]$. In fact, mutation assures, in principle at least, that each feasible design vector can be reached. Hence, mutation is combined, in general, with recombination to avoid premature convergence (Eiben and Smith, 2003).

When displaying all possible mutations of an integer-valued design variable, as done, for example, in Fig. 4.11 for a 6-bit natural binary code, we see that mutation results in new design values $\tilde{\tilde{x}}^{(j)}$ which have a larger likelihood of being in the vicinity of the original design value $\tilde{x}^{(j)}$, than being more far away. Taking this into account, we can construct mutation operators for integer- or real-valued representations of design variables. One of these operators is non-uniform mutation with fixed distribution (Eiben and Smith, 2003). Thereby, the current design vector $\tilde{\mathbf{x}}^{(j)}$ is perturbed by a small amount. This can be achieved by drawing for each component $\tilde{x}_i^{(j)}$ $(i = 1, 2, \ldots, n)$ of $\tilde{\mathbf{x}}^{(j)}$ a value $\xi_i^{(j)}$ randomly from a probability distribution function which is symmetric about zero, and more likely to generate small changes than large ones. This random value $\xi_i^{(j)}$ is then added to the component $\tilde{x}_i^{(j)}$ to give

$$\tilde{\tilde{x}}_i^{(j)} = \tilde{x}_i^{(j)} + \xi_i^{(j)} \tag{4.13}$$

For real-valued representation, most often, a normal distribution function with zero mean and user-specified standard deviation is utilized. For integer-valued representations a corresponding probability mass function has to be constructed. A short overview of other types of mutation operators can be found in (Eiben and Smith, 2003).

4.7 REPLACEMENT

4.7.1 Age-Based Replacement

After recombination and mutation we have a set $\mathcal{L}''$ of λ design vectors, which are different—in terms of their components, but also their fitness values—from the μ design vectors in $\mathcal{M}$. To start the next evolutionary cycle, we have to decide which design vectors should be taken from $\mathcal{L}''$ and $\mathcal{M}$ to form a new, 'better' set $\mathcal{M}'$—to replace the old set $\mathcal{M}$. Although replacement is conceptually similar to selection, there are two noteworthy differences. First, the decision which design vectors to choose is not always based on their fitness values, favoring those with 'better' performance, but in certain variants of evolutionary algorithms (like, for example, simple genetic algorithms) only on their age. And second, whereas selection is typically stochastic, replacement is quite often deterministic. For example, the design vectors in the unified set $\mathcal{L}'' \cup \mathcal{M}$ are ranked according to their fitness values and only the 'best' design vectors are taken (fitness-based). Or, the design vectors for $\mathcal{M}'$ are only selected from the set $\mathcal{L}''$ (age-based).

The idea of age-based replacement is that each design vector exists in the set $\mathcal{M}$ for the same number of evolutionary cycles. This does not necessarily preclude the continuing presence of highly preferable solutions in $\mathcal{M}$. But it is dependent on their being chosen at least once in the selection phase and then surviving intact recombination and mutation. Age-based replacement is typically utilized in simple genetic algorithms (Goldberg, 1989). Since in this variant of evolutionary algorithms the number λ of design vectors in $\mathcal{L}''$ is the same as the number μ of design vectors in $\mathcal{M}$, that is, $\mu = \lambda$, each design vector exists just for one cycle. In other words, the set $\mathcal{M}$ is simply replaced by the set $\mathcal{L}''$. We say that the sets $\mathcal{M}$ and $\mathcal{M}'$ are non-overlapping, since, in general, $\mathcal{M} \cap \mathcal{M}' = \emptyset$. A variant of this replacement strategy are generation-gap methods (Sarma and De Jong, 2000). In this case the sets $\mathcal{M}$ and $\mathcal{M}'$ are overlapping. That is, the set $\mathcal{L}''$ contains $\lambda < \mu$ design vectors. Hence, only a fraction λ/μ of $\mathcal{M}$ is replaced by the new design vectors in $\mathcal{L}''$ and, thus, $\mathcal{M} \cap \mathcal{M}' = \mathcal{L}''$. For age-based strategies the replacement is, typically, done on a first-in-first-out basis.

4.7.2 Fitness-Based Replacement

Whereas genetic algorithms are based, mostly, on age-based replacement, evolutionary strategies utilize fitness-based replacement. Likewise, fitness-based replacement can be grouped in overlapping and non-overlapping strategies. A non-overlapping strategy is the (μ, λ) or comma strategy. Thereby $\lambda > \mu$ (in general, λ is a multiple of μ) new design vectors are generated via recombination and mutation, and the 'best' μ vectors are selected deterministically—by truncation selection of the set $\mathcal{L}''$—to represent the new set $\mathcal{M}'$. In evolutionary programming, stochastic tournament is used to select the design vectors for $\mathcal{M}'$, and, hence, the elitist policy is not quite as strong as in the case of evolutionary strategies.

An overlapping strategy is the $(\mu + \lambda)$ or plus strategy. Thereby, $\lambda \geq \mu$ new design vectors are generated for the set $\mathcal{L}''$. Different to the comma strategy, however, all $(\mu+\lambda)$ design vectors compete directly. That is, the 'best' μ design vectors are chosen from the unified set $\mathcal{L}'' \cup \mathcal{M}$ via truncation selection. Thus, the $(\mu + \lambda)$ strategy is a strong elitism method, since it always retains the 'best' design vectors unless they are replaced by 'superior' ones. Although this method can lead to very rapid improvements in the mean fitness, it can also lead to premature convergence as the design vectors tend to rapidly focus on the 'fittest' member currently present. Thus, it should be utilized in combination with sufficiently large sets and rank-based schemes for selection.

4.8 INITIALIZATION, OPERATION AND TERMINATION

Initialization of the set $\mathcal{M}$ is kept simple in most evolutionary algorithm applications, that is, the first instance of the set $\mathcal{M}$ is generated by randomly generated design vectors—mostly uniform distributed over a certain range of values. In principle, however, also problem-specific heuristics can be used in the initialization phase to generate an initial set $\mathcal{M}$ with higher mean fitness. Whether this is worth the extra computational effort or not, very much depends on the application at hand. It should also be noted, that recombination is most effective when the population is diverse. Thus, an initial set $\mathcal{M}$, which is too focused around a possible solution, makes recombination, potentially, ineffective, and leaves mutation as the only operator to avoid premature convergence. Nevertheless, it is certainly profitable to start with a set of feasible design vectors, than to squander the initial evolutionary cycles on finding such feasible vectors.

A weak point in the theoretical formulation of evolutionary algorithms are their termination criteria. Whereas in gradient-based optimization procedures the necessary (and sufficient) optimality conditions are checked, evolutionary algorithms can only utilize information on the evolution of the design vectors $\mathbf{x}^{(j)}$ and their corresponding fitness values $f(\mathbf{x}^{(j)})$. Consequently, the most often used termination criteria in evolutionary algorithms are:

i.) The total number of fitness evaluations or evolutionary cycles reaches a given limit.

ii.) For a given period of time (that is, for a given number of evolutionary cycles or fitness evaluations), the fitness improvement remains under a threshold value.

iii.) The diversity, that is, the number of different design vectors in the set $\mathcal{M}$, drops under a given threshold value.

That is, the termination criteria in evolutionary algorithms are either related to the elapsed computational time, or how the design vectors and their fitness values change from one evolutionary cycle to the next.

To get a better understanding of above termination criteria, let us have an exemplary look at the operation of evolutionary algorithms. Assume, we want to maximize the objective function

$$g(\mathbf{x}) = 99 - 10\,x_1^4 + 20\,x_1^2 x_2 - 10\,x_2^2 - x_1^2 + 2\,x_1 \tag{4.14}$$

with respect to the design variables x_1 and x_2. In Fig. 4.12 the isolines of the objective function $g(\mathbf{x})$ are displayed. The function has a maximum at ($x_1^* = 1$, $x_2^* = 1$) with $g^*(\mathbf{x}^*) = 100$. We start with a set $\mathcal{M}$ of $\mu = 20$ real-valued design vectors $\mathbf{x}$, which are randomly spread over the search space $[-2, 2] \times [-1, 3]$, as can be seen in Fig. 4.13. We utilize an exponential ranking scheme with $\xi = 0.9$ and stochastic universal sampling to select $\lambda = 20$ design vectors for the set $\mathcal{L}$. Each randomly picked pair is subjected to simulated binary crossover with $\nu = 2$. Moreover, each design vector is subjected at a mean rate of 0.2 to non-uniform mutation with a normal distribution, having zero mean and standard deviation $\sigma = 0.1$. A plus strategy is used for recombination.

The progress of the maximization problem with the number of evolutionary cycles is shown in Fig. 4.14. The curves in Fig. 4.14 are quite characteristic for

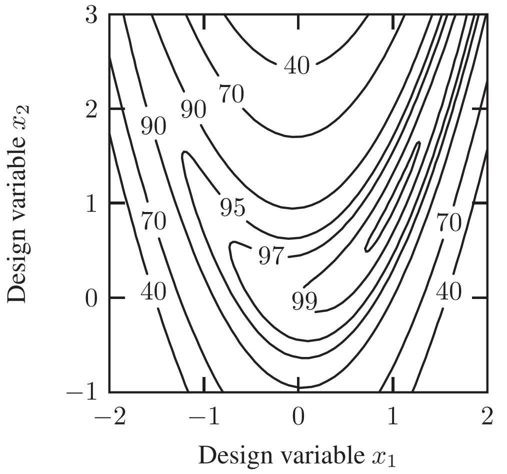

Figure 4.12: Isolines of objective function $f(\mathbf{x})$ of eq. (4.14).

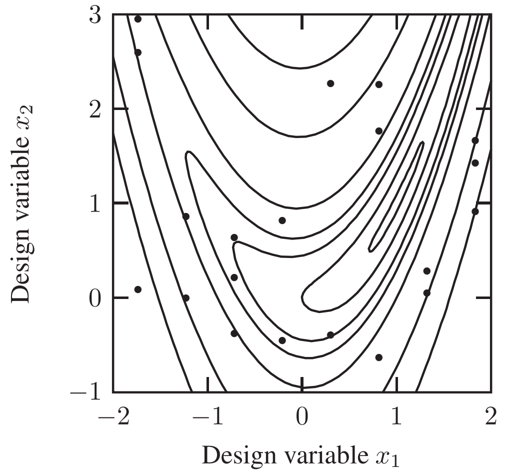

Figure 4.13: Initial design vectors (•), randomly scattered in $[-2, 2] \times [-1, 3]$.

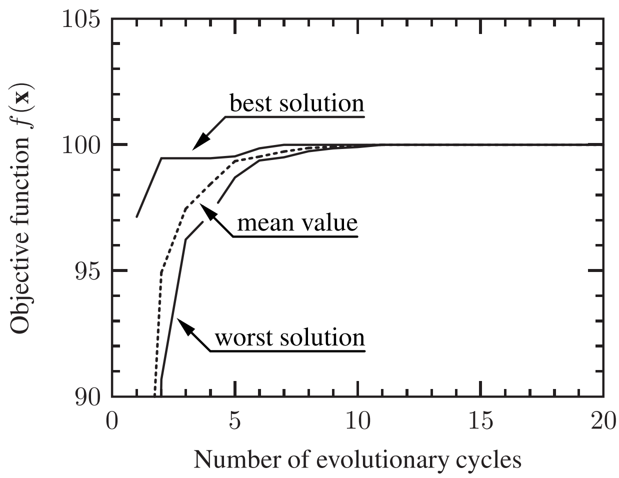

Figure 4.14: Progress of the maximization problem with number of evolutionary cycles.

evolutionary algorithms, that is, there is a rapid progress in the beginning, where recombination is most effective, followed by a dying-out of the progress, where the search is pressed ahead mainly by mutation. This can also seen by inspecting Figs. 4.15 and 4.16. After ten evolutionary cycles, the design vectors have abandoned the low-fitness regions and are scattered in the region above a certain minimum objective function value of approximately 99.9 (see Fig. 4.15). At this time the difference—in terms of objective function values—between the worst design vector and the best design vector in the set $\mathcal{M}$, or the mean value of all design vectors and the best design vector, almost completely vanished (Fig. 4.14). Nevertheless, the

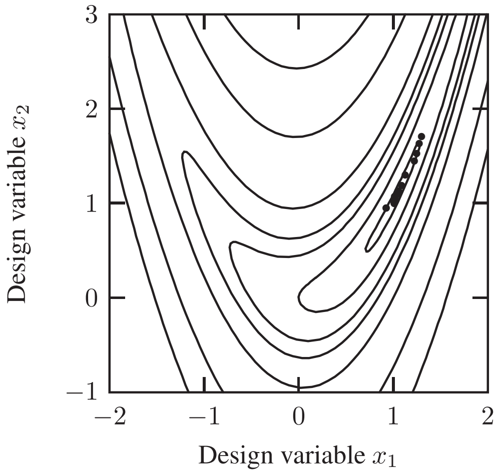

Figure 4.15: Design vectors (•) after ten evolutionary cycles ($\mu = 20$).

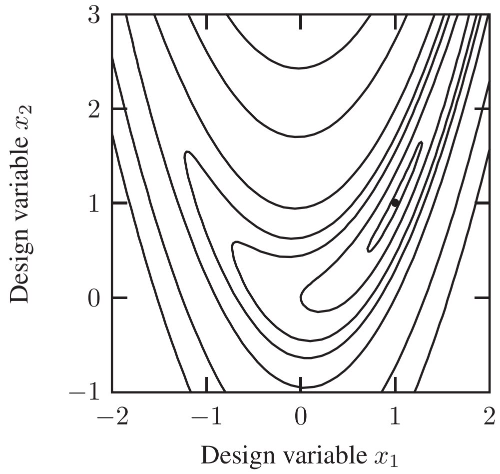

Figure 4.16: Design vectors (•) after twenty evolutionary cycles ($\mu = 20$).

design vectors in $\mathcal{M}$ still show some diversity (Fig. 4.15). Only at a later time, after twenty cycles, say, the design vectors are concentrated clearly around the optimal solution $\mathbf{x}^*$, as can be seen in Fig. 4.16.

As the above maximization example shows, an effective termination criteria should be a combination between the stalling of the fitness improvement and the loss of diversity of the design vectors in $\mathcal{M}$. It should also be noted, that although carefully constructed evolutionary algorithms have the feature to converge almost surely to the global optimum for $t \to \infty$ (Rudolph, 1994a; Rudolph, 1994b), in practical (finite-time) applications it is, nevertheless, possible that either only a local optimum is reached, or that the search is terminated too early, due to the slow improvement rates in the final stage of evolutionary algorithms.

4.9 SUMMARY

An overview of the main components and the principle working mechanisms of evolutionary algorithms has been given. There is no definite, universally applicable evolutionary algorithm, but rather a wide variety of algorithms, that can be created from above described components to fit different problem definitions. Whereas, historically four separate types of evolutionary algorithms developed (that is, genetic algorithms, genetic programming, evolutionary strategies, and evolutionary programming), nowadays a convergence of these different algorithms and an interchange of their components can be observed. A typical example of such convergence are self-adaptive genetic algorithms with real-valued representations (Deb and Beyer, 2001). Similarly, in this thesis we do not use one of the prototypical algorithms, but a hybrid of genetic algorithm and evolutionary strategy. Its main components are listed in Table 4.2.

Table 4.2: Main components of evolutionary algorithm utilized herein.

Component	Utilized herein
Representation	Integer and real-valued design variables
Fitness evaluation	Exponential ranking
Selection	Stochastic universal sampling
Recombination	One-point crossover (between vector components) and simulated binary crossover
Mutation	Non-uniform mutation with fixed distribution
Replacement	Plus strategy

5

Optimizing Maintenance Interventions

5.1 CLASSIFICATION OF MAINTENANCE STRATEGIES

Maintenance interventions allow to ensure sustained integrity and serviceability of structures. However, maintenance interventions can only be justified when the monetary expenditures spent on, say, inspection and rehabilitation are outweighed by expected future benefits from structural operation. If this is not the case, then the structures under investigation are indeed obsolete—at least in their current functional, economic, technical or social configuration—and innovative alternatives have to be evaluated. Thus, the key element in any maintenance effort is the definition of rational criteria which enable us to decide whether these maintenance interventions are economically reasonable or not. For this purpose, we developed in Ch. 3 an optimization formulation based on cost-benefit criteria, which takes into account all significant life-cycle costs, such as construction, failure, inspection and rehabilitation costs, as well as state- or time-dependent benefit rates. Utilizing the principle of cost-benefit analysis, the maintenance interventions will be justified when the total expected (discounted) costs are outweighed by the total expected (discounted) benefits. The proposed formulation also allows to determine optimal lifetimes and acceptable failure rates, that is, it enables us to disclose the close relation between monetary expenditures spent and the hazardousness of structures throughout their entire lifetime.

Before applying our proposed formulation in optimizing maintenance interventions, let us shortly summarize existing maintenance models. Maintenance strategies can, in principle, be categorized into two major classes: corrective maintenance and preventive maintenance (Wang, 2002). Corrective maintenance takes place when a structure or structural component fails. That is, corrective maintenance—

or, as it is most often simply called, repair—means all actions performed as a result of failure, to restore a structure to a specified condition. By contrast, preventive maintenance denotes all actions performed in an attempt to preserve a still operating structure in a specified condition. This includes timely inspection, detection and prevention of incipient failures, that is, decisions to rehabilitate deteriorated components. It goes without saying, that due to the generally disastrous consequences of structural failure, only preventive maintenance strategies are generally acceptable policies in structure and infrastructure engineering.

Preventive maintenance policies can be further classified in age-dependent preventive maintenance, periodic preventive maintenance, sequential preventive maintenance and failure limit preventive maintenance—just to name the most prominent ones (Pierskalla and Voelker, 1976; Valdez-Flores and Feldman, 1989; Wang, 2002). These models assume, in general, that statistical descriptors of failure—as, for example, the mean time to failure, the expected number of failures or the mean residual life—are readily available. The models are then optimized with respect to a cost criterion, like, for example, total cost (for finite planning horizons) or long-run expected cost per unit time (for infinite planning horizons). A coarse description of the above mentioned preventive maintenance policies is as follows:

i.) Age-dependent preventive maintenance has its origin in age replacement policies, where a structure is always replaced when it reaches a certain age (that is, deterioration level) or repaired at failure, whichever occurs first. Extensions thereof led to age-dependent preventive maintenance policies, where the structure is preventively maintained at some predetermined age, or repaired at failure, until perfect maintenance is undertaken. Maintenance, thereby, can be either minimal, imperfect, or perfect.

ii.) In case of periodic preventive maintenance, structures are preventively maintained at fixed time intervals. Most common in this class of maintenance policies are periodic replacement with minimal repair and imperfect maintenance with minimal repair. Also combinations thereof are utilized, for example, an imperfect repair until a specified number of repairs are reached and the structure is replaced.

iii.) Sequential preventive maintenance is based on the state of the structure at maintenance. The next time for preventive maintenance is selected to minimize the expected expenditure during the remaining lifetime. This policy allows a more flexible scheduling then periodic preventive maintenance and, hence, also more cost-effective maintenance solutions.

iv.) Finally, failure limit preventive maintenance utilizes predetermined levels of failure rates or other reliability measures. That is maintenance is undertaken only when the structure reaches an acceptable failure rate or an equivalent reliability measure. This policy can be combined with inspections, when maintenance interventions are based on measurements of some increasing state variable like, for example, wear or accumulated damage.

Above described preventive maintenance models cost-optimize, in general, always only one specific aspect of maintenance: maintenance times, deterioration levels, rehabilitation levels, acceptable failure rates, etc. Since however all these different aspects are closely related, we will optimize herein the above mentioned aspects of maintenance intervention simultaneously. (For our objective function see Sec. 3.7.1.)

5.2 MODELING DETERIORATION WITH CONTINUOUS-TIME MARKOV CHAINS

When formulating our cost-benefit analysis for maintenance interventions in Sec. 3.5, we described the condition state of a deteriorating structure by the time-dependent discrete probability distribution $\boldsymbol{\pi}(t)$, without specifying further how to derive such a distribution. We want to catch up on this now. Another derivation utilizing time distributions will be given in Sec. 5.7.2.

A large variety of structures deteriorate as a result of cumulative effects induced by a sequence of random shocks occurring over time. Each shock causes a random amount of damage which accumulates additively until rehabilitation or failure. As before, and in agreement with the method of condition rating of structural elements (Cesare et al., 1992; Madanat et al., 1995), let us consider a structure which at any given time can be identified as being in one of m possible discrete states: $1, 2, \ldots,$ m. For a single random shock, the probability $p_{i,j}$ that the structure is in state i after the shock, provided the structure was in state j before the shock, is defined by the $(m \times m)$-transition matrix

$$\mathbf{P} = \begin{bmatrix} p_{1,1} & 0 & \cdots & 0 \\ p_{2,1} & p_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ p_{m,1} & p_{m,2} & \cdots & 1 \end{bmatrix} \tag{5.1}$$

Hence, specifying the initial probability distribution $\boldsymbol{\pi}_0$ of the structure as being in one of the possible states, the probability distribution $\boldsymbol{\pi}_q$ after q shocks is given by

Markov chain theory as

$$\boldsymbol{\pi}_q = \mathbf{P}_q \cdots \mathbf{P}_2\mathbf{P}_1\boldsymbol{\pi}_0 \tag{5.2}$$

with $\mathbf{P}_k$ ($k = 1,2,\ldots,q$) being the transition matrix for the k-th shock. For a discussion of the appropriateness of the Markovian hypothesis in deterioration modeling and some relaxations thereof see (Sobczyk and Spencer, Jr., 1992). The entries in the transition matrix **P** can either be determined from experimental data (Bogdanoff and Krieger, 1978; Cesare et al., 1992; Madanat et al., 1995) or by utilizing analytical models (Gansted et al., 1991; Lassen, 1991).

In the following we assume that the number $N(t)$ of random shocks in the time interval $[t_0, t]$ follows an inhomogeneous Poisson counting process with occurrence rate $\lambda(t)$. Thus $N(t)$ is a Poisson variable with mean

$$\tau(t) = \int_{t_0}^{t} \lambda(s)\,\mathrm{d}s \tag{5.3}$$

In general, matrix **P** has eigenvalues $\omega_1, \omega_2, \ldots, \omega_k$ with multiplicities $m_1, m_2, \ldots,$ m_k, respectively, that is,

$$\det(\mathbf{P} - \omega\mathbf{I}) = \prod_{i=1}^{k}(\omega_i - \omega)^{m_i} \tag{5.4}$$

where **I** is the identity matrix. Hence matrix **P** is similar to a block-diagonal matrix of the form (Cox and Miller, 2001; Gusella, 2000)

$$\mathbf{H}^{-1}\mathbf{P}\mathbf{H} = \mathbf{J} = \begin{bmatrix} \boldsymbol{\Lambda}_1 & 0 & \cdots & 0 \\ 0 & \boldsymbol{\Lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\Lambda}_k \end{bmatrix} = \boldsymbol{\Lambda}_1 \oplus \boldsymbol{\Lambda}_2 \oplus \cdots \oplus \boldsymbol{\Lambda}_k = \bigoplus_{i=1}^{k} \boldsymbol{\Lambda}_i \tag{5.5}$$

In eq. (5.5) the so called Jordan blocks $\boldsymbol{\Lambda}_i$ are $(m_i \times m_i)$-matrices of type

$$\boldsymbol{\Lambda}_i = \omega_i\,\mathbf{I}_{m_i \times m_i} + \mathbf{N}_{m_i \times m_i} \tag{5.6}$$

with $\mathbf{N}_{m_i \times m_i}$ being a nilpotent matrix (Horn and Johnson, 1990), that is,

$$\mathbf{N}^{j}_{m_i \times m_i} = \mathbf{0} \quad \text{for} \quad j \geq m_i \tag{5.7}$$

The usefulness of above representation is that powers of **P** can be simply expressed as powers of Jordan blocks:

$$\mathbf{P}^q = \mathbf{H}\mathbf{J}^q\mathbf{H}^{-1} = \mathbf{H}\bigoplus_{i=1}^{k}\boldsymbol{\Lambda}_i^q\mathbf{H}^{-1} \tag{5.8}$$

with

$$\mathbf{\Lambda}_i^q = \omega_i^q \, \mathbf{I}_{m_i \times m_i} + \sum_{j=1}^{m_i-1} \binom{q}{j} \omega_i^{q-j} \, \mathbf{N}_{m_i \times m_i}^j \tag{5.9}$$

and the binomial coefficients defined as

$$\binom{q}{j} = \begin{cases} \dfrac{q!}{j!\,(q-j)!} & \text{for} \quad 0 \le j \le q \\ 0 & \text{for} \quad 0 \le q < j \end{cases} \tag{5.10}$$

Since the number $N(t)$ of random shocks follows an inhomogeneous Poisson counting process, the powers of the Jordan blocks $\mathbf{\Lambda}_i$ have to be weighted by its probability of occurrence:

$$\begin{aligned} \sum_{n=0}^{\infty} \exp(-\tau(t)) \frac{\tau^n(t)}{n!} \mathbf{\Lambda}_i^n &= \Bigg\{ \exp(-\tau(t)) \sum_{n=0}^{\infty} \frac{(\omega_i \tau(t))^n}{n!} \mathbf{I}_{m_i \times m_i} \\ &\quad + \sum_{j=1}^{m_i-1} \frac{\tau^j(t)}{j!} \sum_{n=j}^{\infty} \frac{(\omega_i \tau(t))^{n-j}}{(n-j)!} \mathbf{N}_{m_i \times m_i}^j \Bigg\} \\ &= \exp[(\omega_i - 1)\tau(t)] \\ &\quad \times \left\{ \mathbf{I}_{m_i \times m_i} + \sum_{j=1}^{m_i-1} \frac{\tau^j(t)}{j!} \mathbf{N}_{m_i \times m_i}^j \right\} \end{aligned} \tag{5.11}$$

Hence, given the initial propbability distribution $\boldsymbol{\pi}_0 = \boldsymbol{\pi}(t_0)$, the probability distribution $\boldsymbol{\pi}(t)$ at time t is

$$\boldsymbol{\pi}(t) = \mathbf{H} \bigoplus_{i=1}^{k} \exp[(\omega_i - 1)\tau(t)] \left\{ \mathbf{I}_{m_i \times m_i} + \sum_{j=1}^{m_i-1} \frac{[\tau(t)]^j}{j!} \mathbf{N}_{m_i \times m_i}^j \right\} \mathbf{H}^{-1} \boldsymbol{\pi}_0 \tag{5.12}$$

Eq. (5.12) can be compactly written as

$$\boldsymbol{\pi}(t) = \mathbf{\Psi}(t, t_0) \boldsymbol{\pi}_0 \tag{5.13}$$

with $\mathbf{\Psi}(t, t_0)$ being the fundamental matrix. Moreover, eq. (5.12) is nothing else than the solution of the differential equation

$$\dot{\boldsymbol{\pi}}(t) = \lambda(t) \, \mathbf{H} \bigoplus_{i=1}^{k} [(\omega_i - 1) \, \mathbf{I}_{m_i \times m_i} + \mathbf{N}_{m_i \times m_i}] \, \mathbf{H}^{-1} \boldsymbol{\pi}(t) \tag{5.14}$$

or simply

$$\dot{\boldsymbol{\pi}}(t) = \lambda(t)(\mathbf{P} - \mathbf{I})\boldsymbol{\pi}(t) = \mathbf{A}(t)\boldsymbol{\pi}(t) \quad \text{with} \quad \boldsymbol{\pi}(t_0) = \boldsymbol{\pi}_0 \tag{5.15}$$

where $\mathbf{A}(t)$ denotes the (time-dependent) system matrix:

$$\mathbf{A}(t) = \begin{bmatrix} a_{1,1}(t) & 0 & \cdots & 0 \\ a_{2,1}(t) & a_{2,2}(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}(t) & a_{m,2}(t) & \cdots & 0 \end{bmatrix} \tag{5.16}$$

with $a_{mm}(t) = 0$, since state m is an absorbing state. The matrix elements $a_{lk}(t)$ have to fulfill the conditions

$$a_{k,k}(t) = -\sum_{l=k+1}^{m} a_{l,k}(t) \quad \text{and} \quad a_{l,k}(t) \geq 0 \quad (l > k = 1, 2, \ldots, m-1) \tag{5.17}$$

Thus, when optimizing maintenance interventions we have to solve the differential eq. (5.15) to get the probability evolution in time—at least between interventions. The only thing still to be done is to describe the effect of inspection and rehabilitation on the vector $\boldsymbol{\pi}(t)$, which we will do next.

5.3 INSPECTION AND REHABILITATION

The lifetime of a deteriorating structure is reached as soon as its failure rate $h(t)$, given in the present case by

$$h(t) = \frac{\dot{\pi}_m(t)}{1 - \pi_m(t)} \tag{5.18}$$

exceeds an acceptable limit $h_{\mathrm{a}}(t)$, or when the profitability of structural operation is exhausted. To recover profitability, or to reduce the failure rate, maintenance interventions, that is, inspections and rehabilitations, have to be performed. Using eq. (5.15), the probability evolution of a deteriorating structure subject to a sequence of n maintenance interventions at times $t_1 < t_2 < \ldots < t_j < \ldots < t_n$ can be described as

$$\dot{\boldsymbol{\pi}}(t) = \mathbf{A}(t)\boldsymbol{\pi}(t), \quad t \neq t_j \tag{5.19}$$

$$\delta\boldsymbol{\pi}(t) = \mathbf{Q}\mathbf{C}_j\mathbf{D}_j\boldsymbol{\pi}(t), \quad t = t_j \tag{5.20}$$

The probability distribution $\boldsymbol{\pi}(t_j^+)$ after inspection and rehabilitation is determined from the probability distribution $\boldsymbol{\pi}(t_j^-)$ before the jth maintenance intervention as

$$\boldsymbol{\pi}(t_j^+) = \boldsymbol{\pi}(t_j^-) + \delta\boldsymbol{\pi}(t_j) = [\mathbf{I} + \mathbf{Q}\mathbf{C}_j\mathbf{D}_j]\,\boldsymbol{\pi}(t_j^-) \tag{5.21}$$

with $\mathbf{I}$ denoting the identity matrix. The matrices $\mathbf{Q}$, $\mathbf{C}_j$ and $\mathbf{D}_j$ describe the rehabilitation quality, the extent of rehabilitation to be done and the inspection quality, respectively.

The matrices $\mathbf{C}_j$ and $\mathbf{D}_j$ have already been introduced in Sec. 3.5.4 when determining the expected maintenance costs. Here we give only a short description of the imperfect nature of non-destructive inspection methods. There is quite a considerable body of analytical as well as experimental work available on how to model non-destructive inspections (Mori and Ellingwood, 1994a; Rummel, 1998; Achenbach, 2000; Thoft-Christensen and Sørensen, 1987; Frangopol et al., 1997; Simola and Pulkkinen, 1998). In general, the probability of detection function is a non-decreasing function of damage (Mori and Ellingwood, 1994a). In fact, quite often experimental data shows a sigmoidal trend, which can be modeled, for example, by the following function:

$$\Pr(D|k, \alpha_j) = 1 - \exp\left[-\frac{\alpha_j^2(k-1)^2}{(m-1)^2}\right] \tag{5.22}$$

In Eq. (5.22) the probability of detection $\Pr(D|k, \alpha_j)$ is a function of the possible deterioration states $k = 2, 3, \ldots, m-1$ and the inspection quality $\alpha_j \geq 0$, which can vary from inspection to inspection. Other, but similar models can be found in (Mori and Ellingwood, 1994a; Thoft-Christensen and Sørensen, 1987; Frangopol et al., 1997).

The probability of detect function $\Pr(D|k, \alpha_j)$ of eq. (5.22) is displayed in Fig. 5.1 for different values of α_j, which cover the range of possible values in the

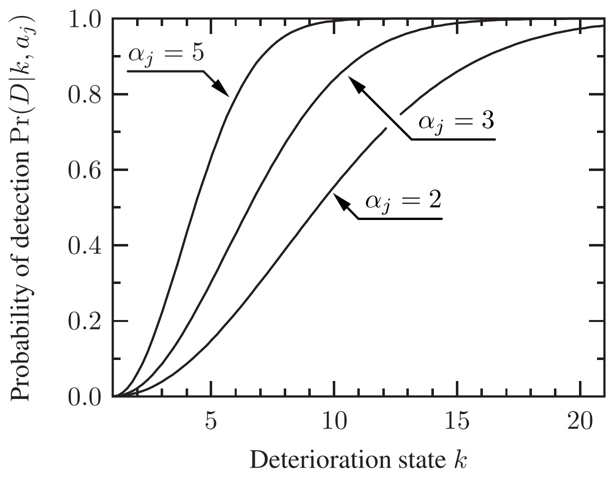

Figure 5.1: Probability of detection $\Pr(D|k, \alpha_j)$ as a function of the condition state k for different inspection qualities α_j (number of states $m = 21$).

following examples. Typically, we will utilize a value of $\alpha_j = 5$. As can be seen from Fig. 5.1, this allows to detect damage states $X(t) \geq 7$ with a probability of 0.9

or higher. It should be noted, however, that the actual shape of the probability of detection function is quite often of secondary importance and can be replaced by a unit step function centered around the median value of the probability of detection curve (Mori and Ellingwood, 1994a). Thus, in our case this would mean, that all damage states $X(t) \geq 5$ would be detected (see Fig. 5.1). In case of perfect inspection, that is, for $\alpha_j \to \infty$, the probability of detection is simply given as:

$$\lim_{\alpha_j \to \infty} \Pr(D|k, \alpha_j) = 1 \quad (k = 2, 3, \ldots, m-1) \tag{5.23}$$

Because of technological limitations, inadequate rehabilitation solutions or faulty execution of construction work, among others, any kind of rehabilitation work is, in general, also imperfect. Hence, we model the rehabilitation of deteriorated states by the $(m \times m)$-matrix **Q**, which is defined as

$$\mathbf{Q} = [q_{l,k}]_{m \times m} \tag{5.24}$$

where the matrix elements $q_{l,k}$ have to fulfill the conditions ($l < k = 2, 3, \ldots, m-1$)

$$-1 \leq q_{k,k} = -\sum_{l=1}^{k-1} q_{l,k} \leq 0 \tag{5.25}$$

and

$$0 \leq q_{l,k} \leq 1 \tag{5.26}$$

All other matrix elements are equal to zero. The description of eqs. (5.24) to (5.26) can be interpreted as a generalization of the classical imperfect repair model (Brown and Proschan, 1983; Pham and Wang, 1996). In case of perfect maintenance eqs. (5.25) and (5.26) reduce to

$$-1 = q_{k,k} = -q_{1,k} \quad (k = 2, 3, \ldots, m-1) \tag{5.27}$$

that is, when applying matrix **Q** the state k is transformed to the initial or non-deteriorated state 1. In case of imperfect maintenance only a fraction $|q_{k,k}|$ of the deteriorated state k is rehabilitated and, due to the imperfect nature of rehabilitation, this fraction is distributed to the states $l < k$ 'proportionally' to $q_{l,k}$.

The joint effect of imperfect inspection and rehabilitation work is exemplary shown in Fig. 5.2. (The shown example corresponds to the optimal solution for $n = 6$ in Table 5.7.) Just before the first maintenance at time $t_1^* = 38.5$ years, we have a distribution of damage states with its median value around $X(t_1^{*-}) = 8$. After inspection and rehabilitation (with a rehabilitation level of $\Delta_1^* = 5$), the probability

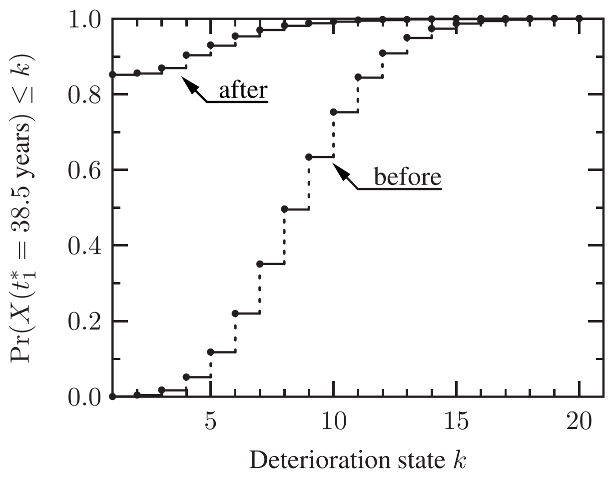

Figure 5.2: Probability distribution $\Pr(X(t) \leq k)$ before and after maintenance at $t_1^* = 38.5$ years.

that the structure is 'as new' is $\Pr(X(t_1^{*+}) = 1) \approx 0.85$ (see Fig. 5.2). However, due to the imperfections, there is also a probability of $\Pr(X(t_1^{*+}) \geq 5) = 1 - \Pr(X(t_1^{*+}) \leq 4) \approx 0.1$, that the structure is still in a damage state larger than $X(t_1^{*+}) = 4$. Nevertheless, this probability is considerable smaller than before maintenance, where this probability had a value of $\Pr(X(t_1^{*-}) \geq 5) \approx 0.95$.

5.4 OBJECTIVE FUNCTIONS

When optimizing the maintenance interventions we will mainly utilize the formulation given in Sec. 3.7.1, that is, we solve the following maximization problem throughout the lifetime $[0, T]$:

$$g^*(T^*) = \max_{\mathbf{t}^*, \boldsymbol{\Delta}^*, \boldsymbol{\alpha}^*, T^*} \left\{ \int_0^T \underbrace{\exp(-\gamma\tau) \left[\sum_{k=1}^{m-1} \dot{B}_k(\tau)\pi_k(\tau) - L\dot{\pi}_m(\tau) \right]}_{= \dot{g}(\tau)} \mathrm{d}\tau \right.$$
$$\left. - \sum_{j=1}^{n} \underbrace{\exp(-\gamma t_j^-) \left[\mathbf{R}\mathbf{C}_j \mathbf{D}_j \boldsymbol{\pi}(t_j^-) + I(\alpha_j)(1 - \pi_m(t_j^-)) \right]}_{= m(t_j^-)} - c_0 \right\} \tag{5.28}$$

subject to

$$\int_{t_j^*}^{t_{j+1}^*} \dot{g}(\tau) \mathrm{d}\tau - m(t_j^{*-}) \geq 0 \tag{5.29}$$

and

$$g^*(T^*) \geq 0 \tag{5.30}$$

The vectors $\mathbf{t}^* = \{t_1^*, t_2^*, \ldots, t_n^*\}$, $\boldsymbol{\Delta}^* = \{\Delta_1^*, \Delta_2^*, \ldots, \Delta_n^*\}$ and $\boldsymbol{\alpha}^* = \{\alpha_1^*, \alpha_2^*, \ldots, \alpha_n^*\}$ are the optimal sequence of maintenance times, rehabilitation levels and inspection qualities, respectively, and T^* is the optimal lifetime.

A slightly different optimization problem can be formulated when the length of the designated time horizon, that is, $[0, T]$ is known (fixed terminal time). In this case we have the following maximization problem with respect to the optimal number of inspections n^*:

$$g^*(T) = \max_{\mathbf{t}^*, \boldsymbol{\alpha}^*, \boldsymbol{\Delta}^*, n^*} \left\{ \int_0^T \dot{g}(\tau)\mathrm{d}\tau - \sum_{j=1}^{n} m(t_j^-) - c_0 \right\} \tag{5.31}$$

subject to the constraints of eqs. (5.29) and (5.30) and the additional constraint

$$\dot{g}^*(t) > 0 \tag{5.32}$$

for all $t \in [0, T]$. It should be recalled, that the objective function of eq. (5.28) fulfills the type of constraint of eq. (5.32) implicitly. In the following examples we will mainly utilize the formulation of eqs. (5.28) to (5.30) with the exception of the first example in Sec. 5.6.1.

5.5 SETTING OF COST FACTORS

Before we can finally start with our numerical examples, we still have to discuss how to set the basic cost factors in the cost-benefit analysis. Thus, let us start with the social discount rate γ. As already mentioned in Sec. 3.4, the choice between the two competing approaches of opportunity cost and time preference for setting the discount rate is primarily a matter of regulatory policy (Morrison, 1998; Arrow et al., 1996). What is more interesting, however, is which value is normally utilized by agencies. In the past, most discount rates for long-term regulations ranged between values of $\gamma = 0.02$ to 0.10 per year, however, these rates are trending to get revised downward (Boardman et al., 2006; Morrison, 1998). In fact, recalculations of these discount rates—based on the respective theories employed by these agencies—give, in general, smaller values between $\gamma = 0.01$ to 0.05 per year (Boardman et al., 2006; U. S. Federal Emergency Management Agency, 1994b). This is also reflected in some more recent recommendations. For example, the British Treasury Board recommends a discount rate of $\gamma = 0.035$ per year for a time horizon of 30 years, time-declining discount rates beyond 30 years, and a minimum discount rate of $\gamma = 0.01$ per year for more than 300 years (Great Britain H. M. Treasury, 2003). Or, the Federal Ministry of Transport, Building and Housing of the Federal Republic of Germany (2005) utilizes a constant discount rate of $\gamma = 0.03$ per year for transport infrastructure planning. Such values also correspond to a recent study on societal discount rates of most more developed countries, suggesting an intermediate compromise value of $\gamma = 0.03$ per year (Rackwitz, 2006). Thus, in the following we adopt for all our calculations a constant discount rate of $\gamma = 0.03$ per year.

Now, let us discuss the setting of the initial cost c_0, the failure cost L, and the benefit rate $\dot{B}$, respectively. Although these cost factors will not be taken from a specific case, they are, nevertheless, consistent with realistic values. We set all cost factors in relation to the initial construction cost c_0 (Kanda, 1996; Kanda and Shah, 1997), whose value is chosen, just for convenience, as $c_0 = 10$ monetary units (m.u.). Assuming, for the moment, that there are no structural failures, losses or maintenance costs, then, based on purely economical grounds, for a structure to be build the minimum requirement is, that for a given service life the expected benefits outweigh the initial cost c_0. Taking, for example, the typically anticipated average lifetime of bridges of $t_\mathrm{a} = 50$ years (Federal Ministry of Transport, Building and Housing of the Federal Republic of Germany, 2005), the normalized benefit rate has to be at least $\dot{B}/c_0 \geq \gamma/[1 - \exp(-\gamma t_\mathrm{a})] \approx 0.039$ per year. The equality sign corresponds thereby to a benefit-(initial) cost ratio of $b(t_\mathrm{a}, 0)/c_0 = 1$.

Benefits are defined as all of the effects of a project on its users or society at large. For example, in transport infrastructure planning the benefits commonly considered are from reduction of travel time or transportation costs, increased traffic safety, environmental relief, impacts from induced traffic, and, sometimes also, economic effects (Sugimoto et al., 2002; Federal Ministry of Transport, Building and Housing of the Federal Republic of Germany, 2005). Different analyses show, that in actual project evaluations these benefits are of similar magnitude (that is, equal or a little bit higher) than the corresponding costs (Sugimoto et al., 2002; Rackwitz, 2002). Hence, we set the normalized benefit rate to a value of $\dot{B}/c_0 = 0.07$ per year. With this value, the benefit-(initial) cost ratio has a value of $b(t_\mathrm{a}, 0)/c_0 \approx 1.81$ for a lifetime of $t_\mathrm{a} = 50$. Taking into account the losses from structural failure, as is done in the following examples, the net benefit-(initial) cost ratio is indeed $[b(T^*, 0) - l(T^*, 0)]/c_0 \approx 1.67$ for the optimal lifetime of $T^* = 46.1$ years (see Sec. 5.6.1). This corresponds to an internal rate of return of 0.067 per year. This is in conformance with the funding of public projects, which should not be expected to have excessive returns on investments, otherwise the private sector would be always able to finance these projects. It should also be mentioned, that in public projects benefits get quite regularly overestimated (Flyvbjerg et al., 2005)—and costs get quite regularly underestimated (Flyvbjerg et al., 2002).

With respect to the failure cost or loss L we normally differentiate between direct and indirect losses (Ang and Lee, 2001). The direct loss includes, for example, costs for replacement of property and non-structural components, costs for avoiding injuries, or life saving costs (Ang and Lee, 2001; Kanda, 1996; Kanda and Shah, 1997). The costs for life saving and injury avoidance can be determined from the relations between structural collapse rates and fatality or injury rates (Ang and Lee, 2001), the expected number of persons affected (Ang and Lee, 2001; Rackwitz, 2004), and appropriate economic estimates of the value of statistical injury and statistical life (Kniesner and Leeth, 1991; Viscusi, 1993; Viscusi and Aldy, 2003; Adler and Posner, 2000; Adler, 2006). The indirect loss is described by the economic effects caused by structural failure, so-called ripple effects (Ang and Lee, 2001; Broder, 1990). As already mentioned in Sec. 2.4.1, such effects can have quite substantial and sustained impact. Variations in the indirect losses are explained by the seriousness of the respective failure and the a priori riskiness assigned to it (Broder, 1990). A typical example along this line, that is, tremendous indirect costs, is certainly the Silver Bridge collapse (see Lichtenstein, 1993).

Although, there have been some attempts to quantify such direct and indirect losses (Ang and Lee, 2001; Rackwitz, 2004), we choose herein a different approach. The advantage of a cost-benefit analysis actually is, that it discloses the relation between cost factors and safety issues. Thus, setting the normalized failure cost herein to a value of $L/c_0 = 200$, gives an acceptable failure rate of $h_{\mathrm{a}}(t) = 3.5 \cdot 10^{-4}$, which compares quite favorably to a proposed medium consequence failure rate (Rackwitz, 2000) of $h(t) = 3.7 \cdot 10^{-4}$, or a failure rate averaged over some reference period as set in International Standard ISO 2394 (1998).

Beside the above discussed basic cost factors, there are also costs related to maintenance interventions. Whereas in practical applications there may exist a quite refined breakdown of maintenance costs, we identify herein only two main components (Thoft-Christensen and Sørensen, 1987; Mori and Ellingwood, 1994b; Frangopol et al., 1997): inspection costs I and rehabilitation costs R_k. Data from damaged buildings suggests that the rehabilitation costs are dependent on the overall damage state (Ang and Lee, 2001; U. S. Federal Emergency Management Agency, 1994a; Kang, 2001), that is, the more overall damage is present in a structure, the higher are the rehabilitation costs for restoring the structure to the 'as new' state $X(t) = 1$. The dependency on the damage state $X(t) = k$ is modeled, in most cases, either as a linear function with a limit of repairable damage (Ang and Lee, 2001; U. S. Federal Emergency Management Agency, 1994a), or as a non-linearly increasing function of damage (Mori and Ellingwood, 1994b). We follow herein the latter approach by setting the rehabilitation costs R_k as

$$R_k = \frac{c_0}{m+1-k} \tag{5.33}$$

with $k = 2, 3, \ldots, m-1$, that is, the rehabilitation costs R_k are non-linear dependent on the amount of damage $X(t) = k$ to be removed. But it should go without saying, that any type of cost model could be utilized in the following examples. The same holds for the inspection and assessment cost I, which we set to a value of $I = 0.2c_0$. This may be a quite high value, but we are interested in showing, that even when maintenance interventions require substantial monetary investments, such investments may nevertheless pay off. A summary of the common cost factors for the following examples is given in Table 5.1.

Table 5.1: Common cost factors for optimization.

Description	Variable	Numerical value
Initial cost	c_0	10 [m.u.]
Loss	L	2000 [m.u.]
Inspection cost	I	2 [m.u.]
Discount rate	γ	0.03 [per year]

5.6 NUMERICAL EXAMPLES

5.6.1 Constant Benefit Rate

The first three examples investigate the effect of different benefit rates (constant, state-dependent or time-dependent) on optimum maintenance planning. The fourth example studies the effect of a time-variant system matrix $\mathbf{A}(t)$. All four examples assume perfect inspection and rehabilitation. The last two examples investigate the effect of imperfect inspection and rehabilitation. The deteriorating structure investigated is described by $m = 21$ states, that is, it has one intact or non-deteriorated state, nineteen deterioration states, and one failure state. This is quite a large number of states as compared to bridge management systems which utilize normally only five to seven condition states (Hawk and Small, 1998; Roelfstra et al., 2004; Scherer and Glagola, 1994; Söderqvist and Veijola, 1998; Thompson et al., 1998). However, we want to achieve in our examples a sufficient 'fine-graininess' with respect to possible maintenance policies. Nevertheless, we can always lump our

$m = 21$ states to a smaller number of, say, seven states. Thus, utilizing $m = 21$ states is in no ways restrictive. In fact, such lumping is also done in some bridge management systems with even more refined scales (Miyamoto et al., 2001).

As a first example, let us assume that the benefit rates are state- and time-independent with $\dot{B}_k = 0.7$ m.u. per year ($k = 1, 2, \ldots, m-1$). This is the general assumption made in life-cycle cost-based maintenance planning (Frangopol et al., 1997; Mori and Ellingwood, 1994b). The structure is subjected to random shocks following a homogeneous Poisson counting process such the system matrix is

$$\mathbf{A} = \lambda \begin{bmatrix} -p & 0 & \cdots & 0 & 0 \\ p & -p & \cdots & 0 & 0 \\ 0 & p & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & -p & 0 \\ 0 & 0 & \cdots & p & 0 \end{bmatrix} \tag{5.34}$$

with a state transition rate of $\lambda p = 0.2$ per year. We get such a description of the matrix **A**, for example, when utilizing a probabilistic linear damage accumulation law equivalent to Miner's rule (Bogdanoff, 1978). Initially the structure is entirely intact, that is, $\pi_1(t_0 = 0) = 1.0$. Without any further maintenance interventions the structure reaches its optimal lifetime at $T^* = 46.1$ years. This is the time when the expected net present benefit reaches its maximum of $g^*(T^*) = 6.698$ m.u. as can be seen in Fig. 5.3. The probability of failure is $\pi_m(T^*) = 1.4 \cdot 10^{-3}$ (see Fig. 5.4).

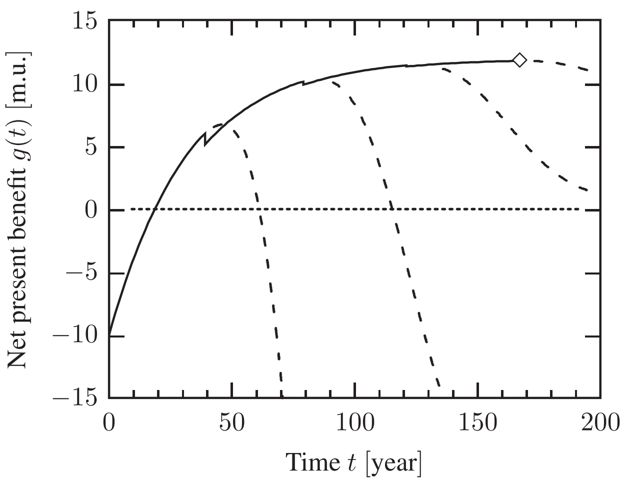

Figure 5.3: Expected net present benefit $g(t)$ for constant benefit rate under optimal maintenance interventions ($n = 3$, solid line). Dashed lines indicate net present benefit evolution without further maintenance interventions.

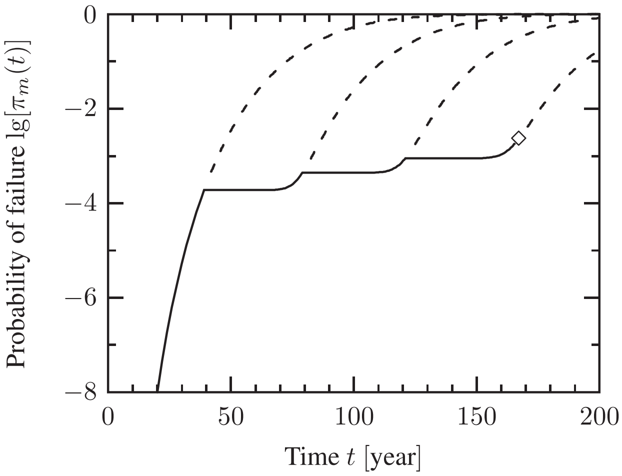

Figure 5.4: Probability of failure for constant benefit rate under optimal maintenance interventions ($n = 3$, solid line). Dashed lines indicate probability of failure evolution without further maintenance interventions.

In the following we want to extend the lifetime through maintenance interventions. Hence, for a given number n of maintenance interventions we determine the optimal lifetime T^* as well as the optimal sequence of maintenance times $\mathbf{t}^*$ and minimum repair levels $\boldsymbol{\Delta}^*$ by maximizing the expected net present benefit rate $\dot{g}(t)$ according to eqs. (5.28) to (5.30). The results up to $n = 4$ are displayed in Table 5.2. As can be seen, the expected net present benefit $g^*(T^*)$ increases with time.

Table 5.2: Optimal maintenance interventions for constant benefit rate.

n	0	1	2	3	4
T^* [year]	46.1	88.1	128.0	167.1	206.0
$g^*(T^*)$ [m.u.]	6.698	10.351	11.419	11.744	11.846
t_1^* [year]	–	41.9	39.9	39.1	38.9
t_2^* [year]	–	–	81.9	79.1	78.0
t_3^* [year]	–	–	–	121.0	117.9
t_4^* [year]	–	–	–	–	159.8
Δ_1^*	–	2	3	3	3
Δ_2^*	–	–	2	3	3
Δ_3^*	–	–	–	2	3
Δ_4^*	–	–	–	–	2

Indeed, in the present case the expected net present benefit increases monotonically, although the absolute increase diminishes due to discounting (see also Fig. 5.3). In other words, proper maintenance allows to extend the economically reasonable lifetime of the structure 'infinitely'. Such a result is not surprising since neither our preference with respect to the structure (in terms of cost factors) nor the system description (in form of matrix **A**) is changing with time. Hence, under the condition that the structure has survived until a certain time, maintenance expenditures spent are always justified by the profitability of further structural operation.

That such lifetime extension by no means compromises safety issues can be seen from Figs. 5.5 and 5.6. In Fig. 5.5 the expected net benefit rate $\dot{g}_{\mathrm{n}}(t)$ for $n = 3$

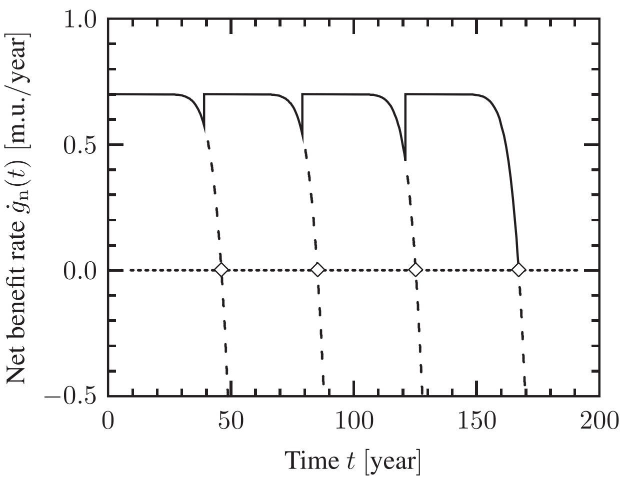

Figure 5.5: Expected net benefit rate $\dot{g}_{\mathrm{n}}(t)$ for constant benefit rate under optimal maintenance interventions ($n = 3$, solid line).

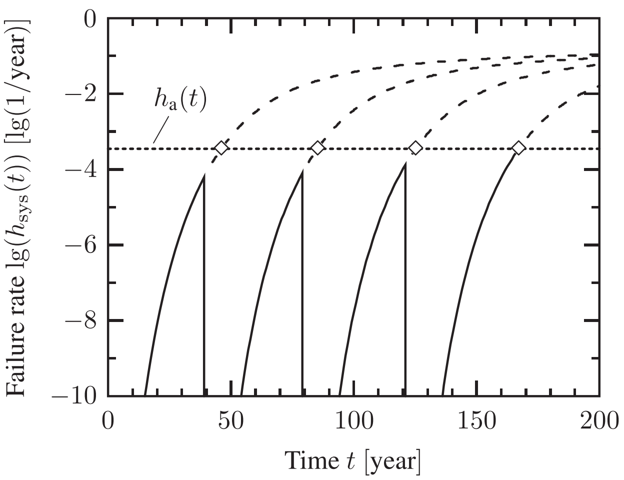

Figure 5.6: Comparison of failure rate $h(t)$ for constant benefit rate under optimal maintenance interventions ($n = 3$, solid line) with acceptable failure rate $h_{\mathrm{a}}(t)$ (dotted line).

optimal maintenance interventions is displayed. Due to structural deterioration also the profitability of structural operation deteriorates. During maintenance interventions all defects about a certain minimum level are rehabilitated, thereby recovering the initial expected benefit rate almost completely. Moreover, the maintenance interventions ensure that the expected net benefit rate remains positive. The dashed lines in Fig. 5.5 correspond to the case that no further maintenance interventions are performed. The points of zero expected net benefit rate are also shown in Fig. 5.6 as dotted line. As can be seen, structural deterioration results in an increase in the failure rate $h(t)$ which would exceed the acceptable failure rate $h_{\mathrm{a}}(t) = 3.5 \cdot 10^{-4}$ per year without maintenance interventions. Thus maintenance interventions not only assure profitability, but also guarantee a certain safety level.

As mentioned above, the problem formulation of eq. (5.28) constraints the optimum solution to a zero expected net benefit rate at the end of the lifetime T^*. This can be clearly observed, for example, from the solution for $n = 3$ in Fig. 5.5. That such a constraint does not necessarily maximize the expected net present benefit for time T^* can be seen when we solve the maximization problem of eqs. (5.31) to (5.32). The results for the fixed lifetime $T = 167.1$ years are given in Table 5.3.

Table 5.3: Optimal maintenance interventions for constant benefit rate (fixed lifetime).

n	3	4	5
T [year]	167.1	167.1	167.1
$g^*(T)$ [m.u.]	11.744	11.756	11.743
t_1^* [year]	39.1	38.7	38.7
t_2^* [year]	79.1	77.4	77.4
t_3^* [year]	121.0	115.9	115.9
t_4^* [year]	–	153.7	153.7
t_5^* [year]	–	–	167.1
Δ_1^*	3	3	3
Δ_2^*	3	3	3
Δ_3^*	2	3	3
Δ_4^*	–	12	12
Δ_5^*	–	–	(no repair)

For $n^* = 4$ maintenance interventions the expected net present benefit at time T becomes a maximum. The result differs mainly from the previous solution for a free terminal time by an additional maintenance intervention at $t_4 = 153.7$ years just before reaching the designated lifetime. Consequently, only deterioration levels greater or equal than $\Delta_4 = 12$ are rehabilitated. Further maintenance interventions, however, are not effective. For $n \geq 5$ all further maintenance interventions would consist of no rehabilitation and would be executed at the end of lifetime T at cost $I\exp(-\gamma T)$.

In Figs. 5.7 and 5.8 the expected net benefit rate and the failure rate, respectively, are shown for $n^* = 4$ (fixed terminal time). The expected net benefit rate and the failure rate at time of inspection or the end of lifetime are almost constant, whereas for the formulation with a free terminal time (Figs. 5.5 and 5.6) the values approach evermore closely the zero expected net benefit rate or the acceptable failure rate

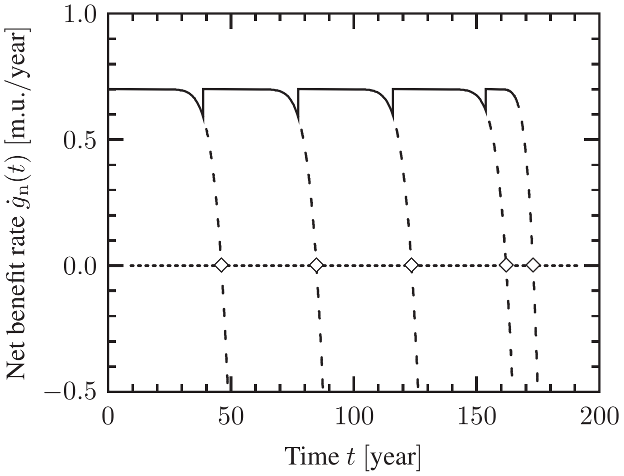

Figure 5.7: Expected net benefit rate $\dot{g}_{\rm n}(t)$ for constant benefit rate under optimal maintenance interventions ($n^* = 4$, solid line) for fixed lifetime $T = 167.1$ years.

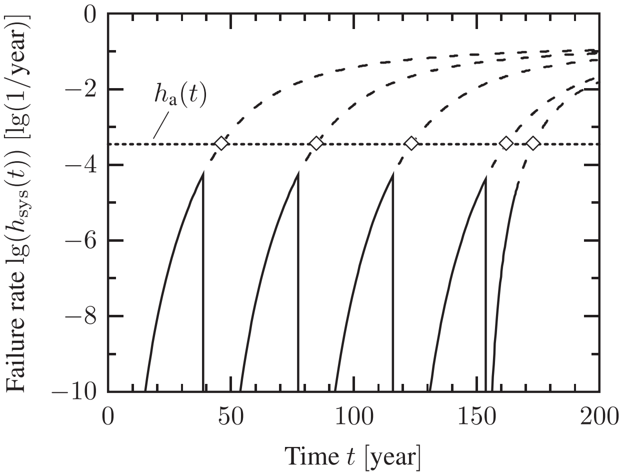

Figure 5.8: Comparison of failure rate $h(t)$ for constant benefit rate under optimal maintenance interventions ($n^* = 4$, solid line) with acceptable failure rate $h_{\rm a}(t)$ (dotted line) for fixed lifetime T = 167.1 years.

from maintenance time to maintenance time. Nevertheless, both cases clearly show that an optimal maintenance policy requires to take actions before reaching the acceptable failure rate or the zero expected net benefit rate level, respectively.

5.6.2 State-Dependent Benefit Rate

For the second example we assume that the benefit rates decrease linearly with the deterioration state k, that is,

$$\dot{B}_k = 0.8 \times \frac{m-k}{m-1} \text{ m.u. per year} \quad (k = 1, 2, \ldots, m-1) \tag{5.35}$$

All other quantities remain the same, that is, the cost factors are given again by Table 5.1 and the system matrix **A** by eq. (5.34). Without maintenance interventions, the optimal life time is reached at $T^* = 44.1$ years. The corresponding failure rate is $h(T^*) = 2.2 \cdot 10^{-4}$ per year.

The optimal maintenance interventions for state-dependent benefit rates and different number n of inspections are given in Table 5.4. As in the previous example,

Table 5.4: Optimal maintenance interventions for state-dependent benefit rates.

n	0	1	2	3	4	5
T^* [year]	44.1	82.2	115.8	146.4	175.6	204.0
$g^*(T^*)$ [m.u.]	5.683	9.366	10.611	11.086	11.280	11.361
t_1^* [year]	–	38.1	33.5	30.6	29.2	28.6
t_2^* [year]	–	–	71.7	64.1	59.8	57.8
t_3^* [year]	–	–	–	102.3	93.3	88.3
t_4^* [year]	–	–	–	–	131.5	121.8
t_5^* [year]	–	–	–	–	–	159.9
Δ_1^*	–	2	2	2	2	2
Δ_2^*	–	–	2	2	2	2
Δ_3^*	–	–	–	2	2	2
Δ_4^*	–	–	–	–	2	2
Δ_5^*	–	–	–	–	–	2

the expected net present benefit $g^*(T^*)$ increases monotonically with time. As can be seen from Figs. 5.9 and 5.10 the qualitative behavior of the optimal solution is like the one for constant benefit rates (Figs. 5.5 and 5.6). The main difference, however, is that more economic pressure is present to keep the structure in a non-deteriorating state since the benefit rates decrease linearly with the states. This results in more often maintenance interventions than in the first example. Note that only the benefit rates are different in the present example, whereas the state description is identical. Thus performing maintenance interventions more regularly is not

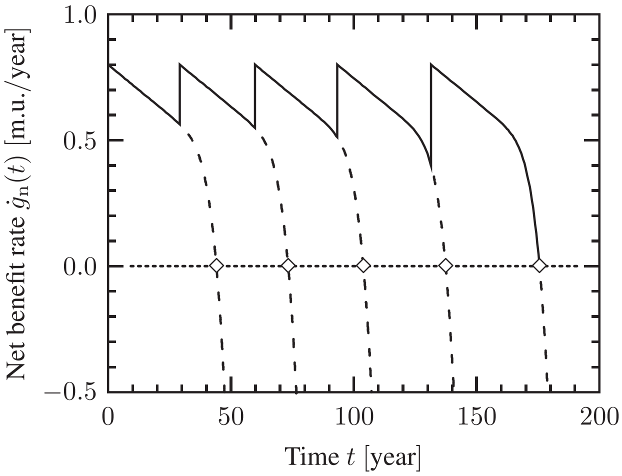

Figure 5.9: Expected net benefit rate $\dot{g}_{\mathrm{n}}(t)$ for state-dependent benefit rate under optimal maintenance interventions ($n = 4$, solid line).

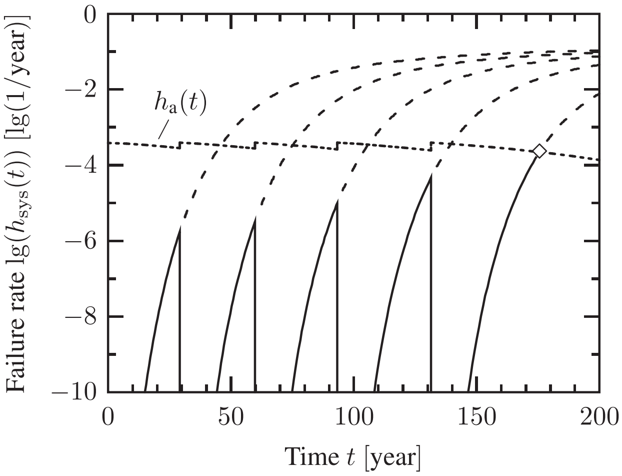

Figure 5.10: Comparison of failure rate $h(t)$ for state-dependent benefit rate under optimal maintenance interventions ($n = 4$, solid line) with acceptable failure rate $h_{\mathrm{a}}(t)$ (dotted line).

due to safety issues as can be seen by inspecting Fig. 5.10, which shows that the structure is rehabilitated way before the acceptable failure rate of $h_{\mathrm{a}}(t) \approx 2.2 \cdot 10^{-4}$ per year is reached, but due to economic reasons expressed by the different benefit rates.

5.6.3 Time-Dependent Benefit Rate

In our third example we assume that the benefit rates are constant in the states, but decrease with time, that is,

$$\dot{B}_k(t) = 1.0 \times \exp(-0.015\,t) \text{ m.u. per year} \quad (k = 1, 2, \ldots, m-1) \qquad (5.36)$$

This can be interpreted as a changing attitude towards the structure. All other quantities are the same as in the previous two examples. The optimal solutions for different numbers n of inspections are presented in Table 5.5. In contrast to the two

Table 5.5: Optimal maintenance interventions for time-dependent benefit rates.

n	0	1	2	3	4
T^* [year]	44.7	82.3	118.2	154.1	190.2
$g^*(T^*)$ [m.u.]	8.682	10.565	10.794	10.801	10.791
t_1^* [year]	–	40.0	38.9	38.7	38.7
t_2^* [year]	–	–	78.0	77.4	77.4
t_3^* [year]	–	–	–	115.8	115.9
t_4^* [year]	–	–	–	–	153.9
Δ_1^*	–	3	3	3	3
Δ_2^*	–	–	3	3	3
Δ_3^*	–	–	–	3	3
Δ_4^*	–	–	–	–	4

previous examples, there is an optimal number $n^* = 3$ of maintenance interventions giving a maximum value of the expected net present benefit of $g^*(T^*) = 10.801$ m.u. for the lifetime $T^* = 154.1$ years. The corresponding expected net benefit rate is shown in Fig. 5.11 and the corresponding failure rate in Fig. 5.12.

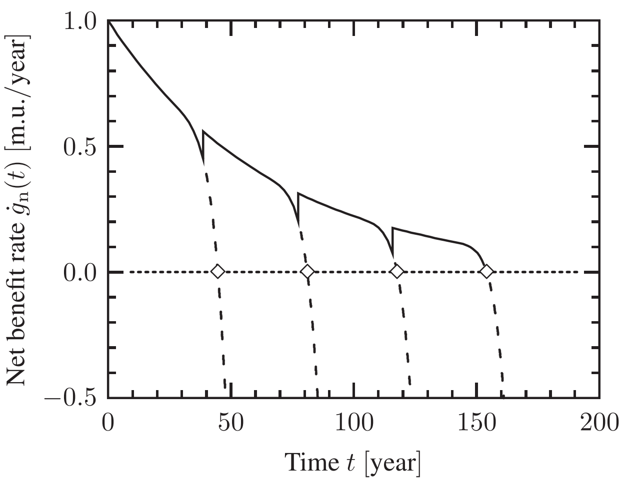

Figure 5.11: Expected net benefit rate $\dot{g}_n(t)$ for time-dependent benefit rate under optimal maintenance interventions ($n^* = 3$, solid line).

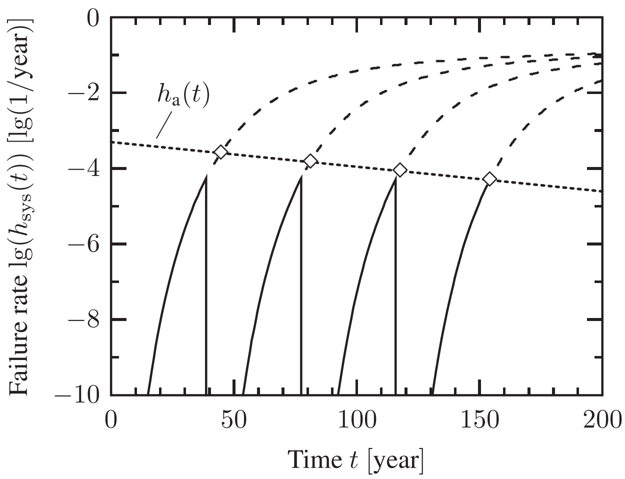

Figure 5.12: Comparison of failure rate $h(t)$ for time-dependent benefit rate under optimal maintenance interventions ($n^* = 3$, solid line) with acceptable failure rate $h_a(t)$ (dotted line).

As getting evident, due to the benefit rates decreasing with time, the repair of states no longer results in a complete recovery of the initial expected net benefit rate. In fact, Fig. 5.11 shows a qualitative similar behavior of the expected net benefit rate as in Fig. 5.5, only modulated by the exponential function given in eq. (5.36). Comparing Tables 5.2, 5.3 and 5.5, there are only slight differences in the optimal maintenance times although in the latter the benefit rates decrease by one order of magnitude throughout lifetime. Thus, obviously, the optimal maintenance actions are quite insensitive to the absolute values of the benefit rates, however quite sensitive to its relative (state-dependent) values as can be seen by comparison with Table 5.4.

Nevertheless, the above mentioned (modulated) decrease in the expected net benefit rate has two consequences. First, since the expected net present benefit between maintenance times is determined by an integration of the net present benefit rate over the respective time interval, this net benefit decreases with time such that for $n \geq 4$ the maintenance expenditures exceed the future benefit, that is, the structure becomes obsolete at $T^* = 154.1$ years. And second, since the benefit rates are part of the description of preferences towards the structure, also the acceptable failure rate $h_a(t)$ decreases with time.

5.6.4 Inhomogeneous Poisson Process

Next, let us assume that the benefit rate is constant with $\dot{B}_k = 0.7$ m.u. per year, but the system is subjected to an inhomogeneous Poisson process with time-dependent intensity $\lambda(t) = 0.01t^2$ per year. The system matrix is defined as

$$\mathbf{A}(t) = \lambda(t) \begin{bmatrix} -p & 0 & \cdots & 0 & 0 \\ p & -2p & \cdots & 0 & 0 \\ 0 & 2p & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & -(m-1)p & 0 \\ 0 & 0 & \cdots & (m-1)p & 0 \end{bmatrix} \tag{5.37}$$

with $p = 0.01$. The optimal solutions for different numbers n of maintenance interventions are given in Table 5.6.

Table 5.6: Optimal maintenance interventions for inhomogeneous Poisson process.

n	0	1	2	3	4
T^* [year]	32.2	39.9	45.2	49.5	53.1
$g^*(T^*)$ [m.u.]	4.181	5.088	5.405	5.504	5.498
t_1^* [year]	–	31.6	31.3	31.2	31.1
t_2^* [year]	–	–	39.4	39.3	39.2
t_3^* [year]	–	–	–	44.9	44.8
t_4^* [year]	–	–	–	–	49.3
Δ_1^*	–	2	2	2	2
Δ_2^*	–	–	2	2	2
Δ_3^*	–	–	–	2	2
Δ_4^*	–	–	–	–	2

The optimal number of maintenance interventions is $n^* = 3$, the maximum value of the expected net present benefit is $g^*(T^*) = 5.504$ m.u. and the optimal lifetime is $T^* = 49.5$ years. As can be seen from the corresponding expected net benefit rate in Fig. 5.13 and the failure rate in Fig. 5.14, the inhomogeneous Poisson process results in an accelerated deterioration of the structure leading to a considerable shortening of the time intervals between maintenance interventions. Consequentially, the expected net present benefit accumulated between the maintenance interventions decreases until for $n \geq 4$ it no longer outweighs the maintenance expenditures, that is, the structure becomes again obsolete.

5.6.5 Imperfect Inspections and Rehabilitations

In the next two examples, we investigate the effect of imperfect inspection methods and imperfect rehabilitation actions on optimal maintenance planning. Thus, the quality of the inspection method is described for all n interventions by the parameter

$$\alpha_j = 5.0 \quad (j = 1, 2, \ldots, n) \tag{5.38}$$

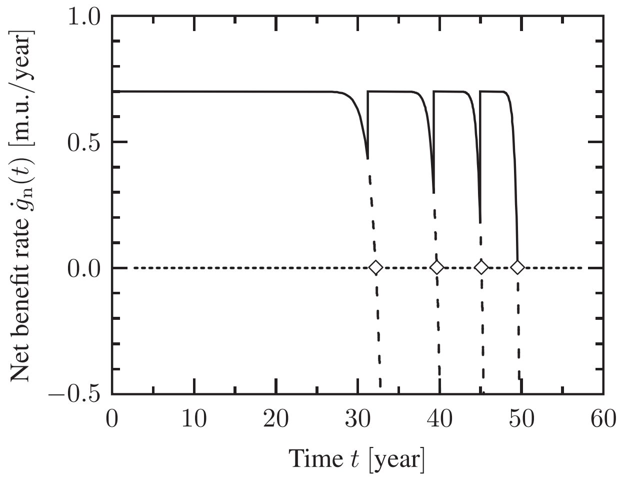

Figure 5.13: Expected net benefit rate $\dot{g}_{\mathrm{n}}(t)$ for inhomogeneous Poisson process under optimal maintenance interventions ($n^* = 3$, solid line).

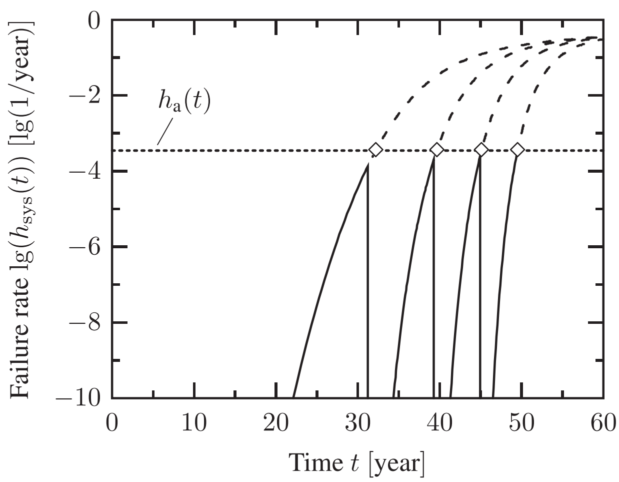

Figure 5.14: Comparison of failure rate $h(t)$ for inhomogeneous Poisson process under optimal maintenance interventions ($n^* = 3$, solid line) with acceptable failure rate $h_{\mathrm{a}}(t)$ (dotted line).

in $\mathbf{D}_j$, whereas the (imperfect) rehabilitation matrix $\mathbf{Q}$ is defined by

$$q_{k,k} = -q_{1,k} = -0.97 \quad (k = 2, 3, \ldots, m-1) \tag{5.39}$$

All other elements equal zero. The system matrix $\mathbf{A}$, the initial probability distribution $\boldsymbol{\pi}(t_0)$ and all cost factors are the same like in the first example (Sec. 5.6.1). Again, we maximize the net benefit rate throughout lifetime for a given number n of maintenance interventions. The optimal solutions are listed in Table 5.7.

Table 5.7: Optimal solution for maintenance interventions in case of imperfect inspection and repair.

n	1	2	3	4	5	6	7
T^* [year]	71.0	96.6	120.5	143.7	166.1	188.9	210.9
$g^*(T^*)$ [m.u.]	8.705	10.084	10.607	10.907	11.043	11.117	11.153
t_1^* [year]	39.7	39.8	38.8	38.7	38.5	38.5	38.4
t_2^* [year]	–	60.7	59.0	56.5	55.7	55.1	54.8
t_3^* [year]	–	–	87.3	85.1	82.6	81.9	81.4
t_4^* [year]	–	–	–	108.6	105.4	102.4	101.4
t_5^* [year]	–	–	–	–	132.3	129.6	126.7
t_6^* [year]	–	–	–	–	–	154.0	150.6
t_7^* [year]	–	–	–	–	–	–	176.9
Δ_1^*	5	3	5	5	5	5	5
Δ_2^*	–	4	3	4	5	5	5
Δ_3^*	–	–	5	3	5	5	5
Δ_4^*	–	–	–	4	3	4	5
Δ_5^*	–	–	–	–	5	3	5
Δ_6^*	–	–	–	–	–	4	3
Δ_7^*	–	–	–	–	–	–	5

As before, the expected net present benefit increases with the number of n inspections, that is, also in case of imperfect rehabilitation, the lifetime of the structure can be extended, in principle, infinitely. However, maintenance interventions have to be performed more often, resulting in a smaller total value of the expected net present benefit as compared with the case of perfect rehabilitation. A consequence of performing maintenance interventions more often is, that the optimal rehabilitation levels Δ_j^* increase, because there are now more possibilities to detect and rehabilitate deteriorated states.

In Figs. 5.15 and 5.16 the evolution of the expected net benefit rate and the failure rate, respectively, are displayed exemplary for $n = 6$ maintenance interventions. As can be seen, the first maintenance intervention is closely followed by a second maintenance intervention. Due to imperfect rehabilitation, the structure can only be partially rehabilitated. Hence, not only the expected net benefit rate is not completely recovered after each rehabilitation, but also the failure rate decreases always only by approximately two orders of magnitude. The remaining severe structural defects are then the cause of untimely additional inspection and rehabilitation efforts. Nevertheless, with an increasing number of maintenance interventions a more and more periodic maintenance pattern evolves, which is only perturbed in Fig. 5.15, Fig. 5.16 and Table 5.7 by the effects of reaching the terminal time T^*. When comparing Fig. 5.6 and Fig. 5.16, we can recognize, that in

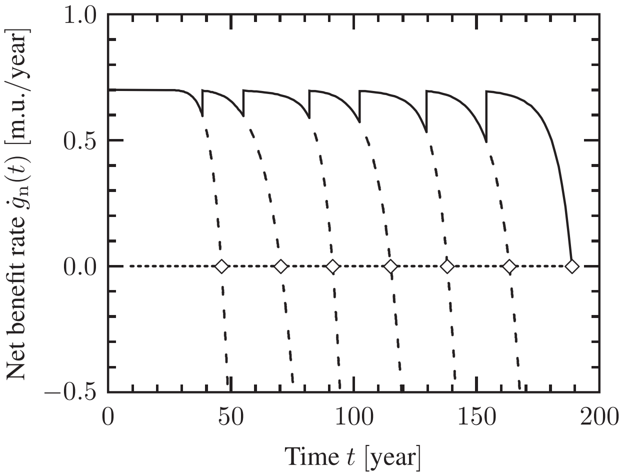

Figure 5.15: Expected net benefit rate $\dot{g}_\mathrm{n}(t)$ in case of imperfect inspection and rehabilitation ($n = 6$, solid line).

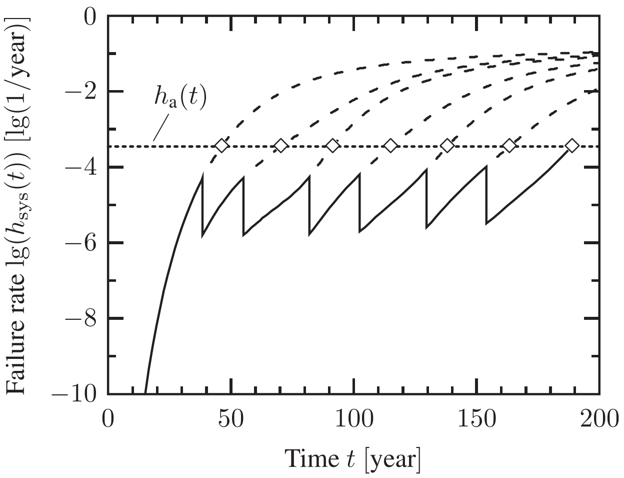

Figure 5.16: Failure rate $h(t)$ in case of imperfect inspection and rehabilitation ($n = 6$, solid line).

both cases the optimal maintenance interventions are performed at times t_j when reaching a failure rate $h(t_j^-) < h_\mathrm{a}(t_j^-)$ of similar magnitude.

5.6.6 Inspection Quality

Finally, let us optimize the maintenance interventions not only with respect to the sequences of maintenance times $\mathbf{t}^* = \{t_1^*, t_2^*, \ldots, t_n^*\}$ and rehabilitation levels $\mathbf{\Delta}^* = \{\Delta_1^*, \Delta_2^*, \ldots, \Delta_n^*\}$, but also with respect to the inspection qualities $\boldsymbol{\alpha}^* = \{\alpha_1^*, \alpha_2^*,$

$\ldots, \alpha_n^*\}$. The cost for inspection is now given as

$$I(\alpha_j) = \frac{1+\alpha_j}{3} \text{ m.u.} \quad (\alpha_j \geq 0) \tag{5.40}$$

Moreover, we assume again that the benefit rates are state- and time-dependent:

$$\dot{B}_k(t) = \frac{m-k}{m-1} \exp(-0.015t) \text{ m.u. per year} \tag{5.41}$$

with $k = 1, 2, \ldots, m-1$. The remaining cost factors are the same as before, as are the system matrix **A** and the initial probability distribution $\boldsymbol{\pi}(t_0)$. The (imperfect) repair matrix **Q** is described by eq. (5.39).

In Table 5.8 the optimal solutions are given for $1 \leq n \leq 6$ number of maintenance interventions. The expected net present benefit reaches its maximum at $n^* = 5$ maintenance interventions. Hence, the structure becomes obsolete at time $T^* = 116.8$ years (Fig. 5.17). It should be noted from Table 5.8, that by optimizing not only the maintenance times, but also the rehabilitation levels and inspection qualities, the time of obsolescence is put further back by increasing the minimum

Table 5.8: Optimal solution for maintenance interventions in case of state- and time-dependent benefit rates.

n	1	2	3	4	5	6
T^* [year]	60.3	78.8	92.2	105.1	116.8	127.7
g^* [m.u.]	6.76	7.18	7.29	7.32	7.33	7.32
t_1^* [year]	34.2	34.0	33.2	32.8	32.8	32.7
t_2^* [year]	–	52.6	51.0	50.1	49.8	49.7
t_3^* [year]	–	–	69.2	67.5	66.5	66.2
t_4^* [year]	–	–	–	84.8	83.3	82.7
t_5^* [year]	–	–	–	–	98.6	97.8
t_6^* [year]	–	–	–	–	–	112.2
Δ_1^*	3	3	3	3	3	3
Δ_2^*	–	3	3	3	3	3
Δ_3^*	–	–	4	3	3	3
Δ_4^*	–	–	–	4	4	4
Δ_5^*	–	–	–	–	6	4
Δ_6^*	–	–	–	–	–	8
α_1^*	4.7	4.4	4.1	4.0	3.9	3.9
α_2^*	–	4.0	3.6	3.4	3.3	3.3
α_3^*	–	–	3.7	3.5	3.3	3.2
α_4^*	–	–	–	3.3	3.2	3.1
α_5^*	–	–	–	–	2.9	2.9
α_6^*	–	–	–	–	–	2.6

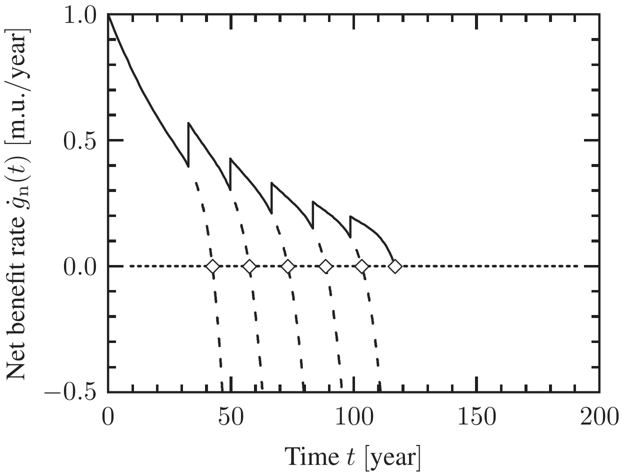

Figure 5.17: Expected net benefit rate $\dot{g}_{\rm n}(t)$ for optimal inspection qualities ($n^* = 5$, solid line).

rehabilitation levels and lowering inspection qualities, that is, by decreasing maintenance costs. That such procedure is not compromising safety issues at all can be seen from the failure rate in Fig. 5.18.

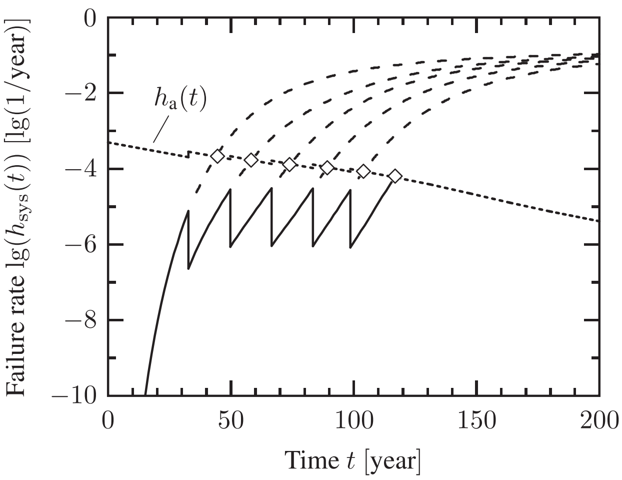

Figure 5.18: Failure rate $h(t)$ for optimal inspection qualities ($n^* = 5$, solid line).

Like in the previous examples, the optimal maintenance interventions are performed at times t_j when the failure rate reaches a value of almost similar magnitude. Since due to the decreasing benefit rate, however, also the acceptable failure rate $h_{\rm a}(t)$ decreases according to eq. (3.40) with time, the lifetime of the structure can not be extended infinitely, but reaches its optimal value at $T^* = 116.8$ years.

5.7 APPLICATION TO BRIDGE MANAGEMENT

5.7.1 Bridge Management Systems

In order to rationalize decisions with respect to maintenance or rehabilitation, bridge management systems have been developed and implemented in North America, Europe and Japan (Hawk and Small, 1998; Thompson et al., 1998; De Brito et al., 1997; Söderqvist and Veijola, 1998; Roelfstra et al., 2004; Miyamoto et al., 2001). The generic components of these management systems can be coarsely summarized as: (a) assessment of bridge conditions, (b) forecasting of further bridge deterioration, and (c) identification and prioritization of maintenance needs and their corresponding financial requirements. However, these systems have been repeatedly criticized for mainly two reasons. The first point of criticism is, that the assessment of bridge conditions is done commonly by so called condition ratings (verbal descriptors) on structural element level made during routine visual inspections. Therefore, these condition ratings mostly indicate the relative health of structural elements only, but they do not identify the physical or chemical processes that cause the deterioration, nor are they directly related to structural behavior, that is, structural safety and serviceability (Roelfstra et al., 2004; Hearn, 1998a). Structural safety is only indirectly mentioned in the description of the most severe condition state as ‘need for immediate intervention’.

Whereas more objective and accurate structural condition assessments can be performed by utilizing concepts of structural identification (Aktan et al., 1996), the complex of problems related to structural safety can only be addressed by structural reliability theory. Thus, what is required is a consistent description of the time-variant condition of a structure in terms of both deterioration and ultimate failure states, as done, for example, in (Mori and Ellingwood, 1994a; Mori and Ellingwood, 1994b; Pandey, 1998; Stewart, 2001; Thoft-Christensen and Sørensen, 1987; Val, 2005). A structural state description in terms of only ‘failure’ or ‘no failure’, as proposed, among others, in (Frangopol et al., 2001), via so called reliability profiles, is not sufficient for optimal maintenance planning, since it does not allow to relate (directly or indirectly) observable deterioration states to specific performance conditions—including its effect on the load carrying capacity and the remaining lifetime—as well as rehabilitation actions to be performed, as is mandatory for effective bridge maintenance systems (Hearn, 1998a; Das, 1998). The addition of a separate condition profile in (Neves and Frangopol, 2005) tries to remedy this shortcoming, although it remains unclear why condition and safety should be treated as separate entities. Moreover, the reliability and condition profiles in (Frangopol et al., 2001; Neves and Frangopol, 2005) are not calculated with the help of structural reliability theory, but are either directly determined by so called experts, or estimated from statistical data. However, as shown in (Czarnecki and Nowak, 2008), time-variant reliability profiles depend on a multitude of factors (structural design, loads, environmental conditions, deterioration mechanisms, etc.) which will be hardly reflected in its entirety in expert knowledge, nor does it seem to be overly realistic to assume, that there will be ever enough data available to directly estimate reliability profiles, that is, without recursion to a physical or chemical model.

Thus, to summarize the first point of criticism, when addressing the problem of optimal maintenance planning for bridges, a consistent probabilistic description of the condition of a structure—including not only deterioration states, but also structural collapse—is indispensable. This requires the explicit modeling of structures, deterioration processes, condition assessments and maintenance interventions. For practical applicability, condition states in the probabilistic analysis should be selected compliant with experimental condition assessment techniques. This allows not only to utilize inspection data for modeling purposes, but also to define optimal maintenance actions in terms of experimentally observable—whether directly or indirectly—indicators of structural deterioration. We will show this in the following exemplarely for a truss-type bridge under fatigue loading, where the overall structural damage state is determined with the help of static load tests.

The second set of criticism of existing bridge management systems is centered around the models utilized for deterioration forecasting. Commonly, discrete-time Markov chains, with time-homogeneous transition probabilities, are employed as a statistical model, based on the above mentioned visual inspection data (Hawk and Small, 1998; Thompson et al., 1998; De Brito et al., 1997; Söderqvist and Veijola, 1998; Roelfstra et al., 2004; Miyamoto et al., 2001). Due to their sole reliance on inspection data, these Markov chains, evidently, inherit also the above mentioned shortcomings of the subjective nature of condition ratings and their lack of information on structural behavior. But also the data itself is problematic, since most often it does not make reference to differences in the structural characteristics of bridges, environmental conditions, past rehabilitation efforts or even time intervals between inspections, thereby compromising the accuracy of its estimates. Also there is theoretical as well as experimental evidence available that the transition probabilities are, in general, time-inhomogeneous, that is, that age—the time since construction or rehabilitation—has a significant impact on the deterioration rate (Jiang et al., 1989; Madanat et al., 1995; Kallen and van Noortwijk, 2006). However, it should be also mentioned, that we do not follow the general rejection of Markov processes as being not able to model such behavior at all, as has been done in (Frangopol et al., 2001; Neves and Frangopol, 2005). Utilizing continuous-time Markov chains, with time-inhomogeneous transition probability rates, indeed allows to model age dependency (Kallen and van Noortwijk, 2006), as well as maintenance effects like delays in deterioration or changes in deterioration rates. For this purpose we propose in the following to utilize the time distributions of reaching defined damage or deterioration states—as determined from a probabilistic analysis—to build up the transition matrix. It should go again without saying, that this requires an explicit modeling of the deterioration process and respective maintenance interventions, that is, it can not be done by solely utilizing inspection data or expert knowledge.

5.7.2 Determining Transition Rates from Time Distributions

We want to describe the probability evolution of the discrete deterioration states by a continuous-time Markov chain with time-inhomogeneous (age-dependent) transition rates. Hence, the probability distribution vector $\boldsymbol{\pi}(t, t_0)$ is governed by the differential equation

$$\dot{\boldsymbol{\pi}}(t, t_0) = \mathbf{A}(t, t_0)\boldsymbol{\pi}(t, t_0)\,, \quad \boldsymbol{\pi}(t_0, t_0) = \mathbf{p}(t_0) \tag{5.42}$$

where $\mathbf{A}(t, t_0)$ is the matrix of time-inhomogeneous transition probability rates and $\mathbf{p}(t_0)$ is the initial probability vector of the structure being in one of the m states at time t_0. A simplified version of eq. (5.42) with only two states, that is, 'no failure' (state 1) and 'failure' (state 2) underlies the deterioration modeling in (Frangopol et al., 2001; Neves and Frangopol, 2005). Therein, the probability evolution for a structure, without re-building after failure, is assumed to be

$$\dot{\boldsymbol{\pi}}(t, t_0) = \begin{bmatrix} -\Phi(-\beta(t, t_0)) & 0 \\ \Phi(-\beta(t, t_0)) & 0 \end{bmatrix} \boldsymbol{\pi}(t, t_0)\,, \quad \mathbf{p}(t_0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{5.43}$$

where $\Phi(-\beta(t, t_0))$ denotes the failure rate, $\Phi(\cdot)$ is the standard normal distribution function, and $\beta(t, t_0)$ is a piecewise linear, monotonically decreasing function in time t.

As repeatedly mentioned, since it is quite unlikely, that there will be enough data available to directly estimate all transition rates $a_{l,k}(t, t_0)$ of $\mathbf{A}(t, t_0)$, a probabilistic analysis has to be performed. Thus, let us assume that the deterioration process starts in state 1 at time t_0 with probability one, that is, $\Pr(X(t_0) = 1) = 1$. The time distribution functions until certain deterioration states are reached or failure, that is, structural collapse, occurs are

$$F_k(t, t_0) = \Pr(X(t) > k | X(t_0) = 1)\,, \quad k = 1, 2, \ldots, m - 1 \tag{5.44}$$

Hence, the probability of the deterioration process $\{X(t), t \geq t_0\}$ being in state k is

$$\pi_k(t,t_0) = \Pr(X(t) = k|X(t_0) = 1)$$
$$= \begin{cases} 1 - F_1(t,t_0)\,, & k = 1 \\ F_{k-1}(t,t_0) - F_k(t,t_0)\,, & 1 < k < m \\ F_{m-1}(t,t_0)\,, & k = m \end{cases} \tag{5.45}$$

Utilizing the (non-negative) rate function

$$h_k(t,t_0) = \frac{f_k(t,t_0)}{1 - F_k(t,t_0)} \tag{5.46}$$

with $f_k(t,t_0)$ being the probability density function of the time until $X(t) \geq k$, the distribution functions can be re-written as

$$F_k(t,t_0) = 1 - \exp\left[-\int_{t_0}^{t} h_k(\tau,t_0)\,\mathrm{d}\tau\right] \tag{5.47}$$

Differentiating eq. (5.45) with respect to time t, we get the elements $a_{l,k}(t,t_0)$ of matrix $\mathbf{A}(t,t_0)$ in terms of the rate functions $h_k(t,t_0)$ after some re-arrangement as

$$a_{l,k}(t,t_0) = \begin{cases} -h_k(t,t_0)\,, & l = k \neq m \\ 0\,, & l = k = m \\ h_{l-1}(t,t_0) - h_l(t,t_0)\,, & l \neq m,\ k < l \\ h_{m-1}(t,t_0)\,, & l = m,\ k < m \end{cases} \tag{5.48}$$

5.7.3 Truss-Type Bridge Structure under Fatigue Loading

Let us exemplify the above procedure for the truss-type bridge structure shown in Fig. 5.19. The pin-jointed truss has dimensions $l = 8\,\text{m}$, $s_1 = 5\,\text{m}$ and $s_2 = 2\,\text{m}$. All structural elements have cross sectional areas of $0.04\,\text{m}^2$—with the exception of elements (a) and (b) having $0.06\,\text{m}^2$. The elastic modulus is $E = 2.1 \cdot 10^{11}\,\text{N/m}^2$. The structure is subjected to the external loads $V_1(t)$ and $V_2(t)$. The loads $V_1(t)$ and $V_2(t)$ are assumed to form an independent and identically distributed random load sequence with a mean occurrence rate of 20,000 per year. Both load amplitudes are Rayleigh distributed with mean $500\,\text{kN}$, standard deviation $100\,\text{kN}$ and a mutual correlation of 0.3.

Fatigue damage accumulation in structural elements is determined in the following by a continuous damage mechanics approach (Simo and Ju, 1987; Chaboche,

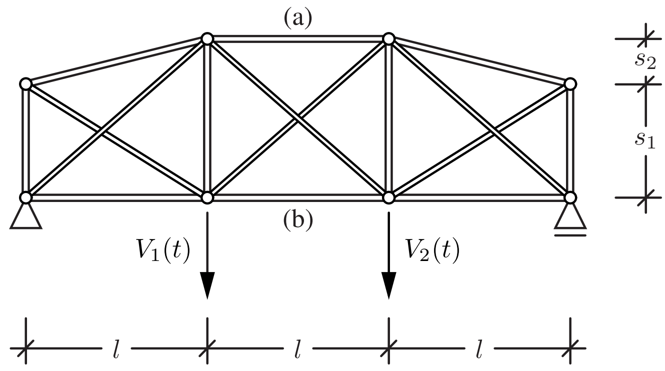

Figure 5.19: Truss-type bridge under random load processes $V_1(t)$ and $V_2(t)$.

1988; Lemaitre and Chaboche, 1990). The damage growth δD per load cycle δn is modeled as

$$\delta D = C(1 - D)^{-p}(\Delta\sigma)^q \, \delta n \tag{5.49}$$

where $0 \leq D \leq 1$ denotes damage, $\Delta\sigma$ is the stress amplitude, C is the damage growth rate, and p and q are coefficients, chosen herein as $p = 0.5$ and $q = 2$, respectively. The damage growth rate is assumed to be lognormally distributed with mean $5 \cdot 10^{-9}\,\text{MPa}^{-2}$ and standard deviation $1 \cdot 10^{-9}\,\text{MPa}^{-2}$.

Eq. (5.49) describes the damage evolution on a local level, that is, in our case on the level of structural elements. When we assume that the deformation behavior of the material is only affected by the damage in the form of effective stresses, then the uniaxial elastic damage law is

$$\sigma = E(1 - D)\varepsilon_{\text{e}} \tag{5.50}$$

where E is the elastic modulus of the undamaged material and ε_{e} is the elastic strain. To combine the damages on structural member level to form an indicator of overall structural damage D_{sys} we utilize the elastic strain energy. Hence, overall structural damage is defined as (Lemaitre and Chaboche, 1990; Hanganu et al., 2002)

$$D_{\text{sys}} = 1 - \frac{W}{W_0} \tag{5.51}$$

where W is the actual elastic strain energy and W_0 is the fictitious elastic strain energy of the undamaged structure due to the actual strains. For determining the strain energies we subject the structure to a static load test. As reference load case the most likely load configuration (V_1^*, V_2^*) is utilized. A detailed discussion on the effect of reference loads for determining global damage indices is given in (Hanganu et al., 2002).

In analogy to current condition rating procedures, we select five damage levels: $d_1 = 0.2$, $d_2 = 0.4$, $d_3 = 0.6$, $d_4 = 0.8$ and $d_5 = 1.0$, the latter denoting structural collapse. Fig. 5.20 shows the determined time distributions of the truss-type bridge

$$F_k(t, 0) = \Pr(D_{\text{sys}}(t) \geq d_k)\,, \quad k = 1, 2, \ldots, 5 \tag{5.52}$$

for exceeding the different overall damage levels d_k. The corresponding non-diagonal terms of $\mathbf{A}(t, 0)$ are displayed in Fig. 5.21. It should be noted, that the quantity $a_{6,k}(t, 0)$ is nothing else than the hazard or failure rate of the structure without any maintenance.

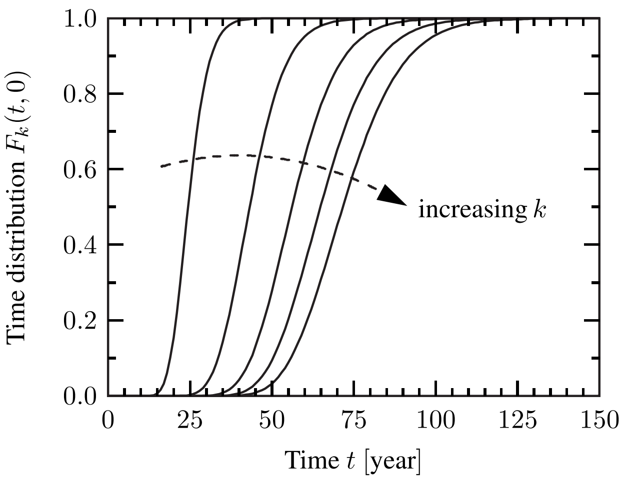

Figure 5.20: Time distributions $F_k(t, 0)$ for $k = 1, 2, \ldots, 5$.

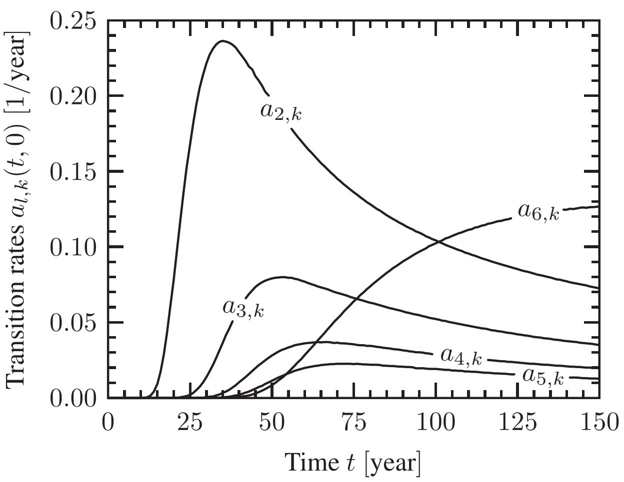

Figure 5.21: Transition rates $a_{l,k}(t, 0)$ for $l = 2, 3, \ldots, 6$.

5.7.4 Modeling of Maintenance Interventions

Since the matrix **A** is now time-inhomogeneous, the maintenance modeling given in Sec. 5.3 needs some minor modification. As before, the probability evolution until the first maintenance intervention at time t_1 is

$$\dot{\boldsymbol{\pi}}(t,t_0) = \mathbf{A}(t,t_0)\boldsymbol{\pi}(t,t_0)\,, \quad t_0 \le t < t_1 \tag{5.53}$$

with initial condition $\boldsymbol{\pi}(t_0,t_0) = \mathbf{p}(t_0)$. At time t_1 a decision is made with respect to the amount of rehabilitation work to be done. In general, the structure is inspected, its condition is assessed and a minimum level of rehabilitation Δ_1 is selected such that all deterioration states $k \ge \Delta_1$ are restored. Since matrix $\mathbf{A}(t,\cdot)$ is time-dependent, we have to split the probability vector $\boldsymbol{\pi}(t_1^-,t_0)$ in two vectors:

$$\boldsymbol{\pi}(t_1^+,t_0) = \boldsymbol{\pi}(t_1^-,t_0) - \mathbf{C}_1\boldsymbol{\pi}(t_1^-,t_0) = \boldsymbol{\pi}(t_1^-,t_0) - \delta\boldsymbol{\pi}(t_1,t_0) \tag{5.54}$$

containing the probabilities of the non-rehabilitated states, whose evolution for $t_1 < t < t_2$ is continued to be governed by the matrix $\mathbf{A}(t,t_0)$, and

$$\boldsymbol{\pi}(t_1,t_1) = \mathbf{Q}\mathbf{C}_1\boldsymbol{\pi}(t_1^-,t_0) \tag{5.55}$$

the probabilities of the rehabilitated states, which can be interpreted as the initial conditions $\boldsymbol{\pi}(t_1,t_1) = \mathbf{p}(t_1)$ of a new structure, whose evolution is governed by a, so to say, restarted matrix $\mathbf{A}(t,t_1)$, as will be shown below.

The above scheme can be continued, such that for n maintenance interventions at times $t_1 < t_2 < \ldots < t_n$ the governing equations are given as ($s = 0,1,\ldots,n$; $r = 1,2,\ldots,n$; $r > s$)

$$\dot{\boldsymbol{\pi}}(t,t_s) = \mathbf{A}(t,t_s)\boldsymbol{\pi}(t,t_s)\,, \quad t \ge t_s,\ \ t \ne t_r \tag{5.56}$$

$$\delta\boldsymbol{\pi}(t,t_s) = -\mathbf{C}_r\boldsymbol{\pi}(t,t_s)\,, \qquad t = t_r \tag{5.57}$$

with initial conditions

$$\mathbf{p}(t_s) = \begin{cases} \mathbf{p}(t_0)\,, & s = 0 \\ \mathbf{Q}\mathbf{C}_s \sum\limits_{j=0}^{s-1} \boldsymbol{\pi}(t_s^-,t_j)\,, & s > 0 \end{cases} \tag{5.58}$$

The probabilities of the structure being in one of the m states at time t, and the corresponding derivatives with respect to time are

$$[\Pr(X(t)=1), \Pr(X(t)=2), \ldots, \Pr(X(t)=m)]^{\mathrm{T}} = \sum_{s:\ t>t_s} \boldsymbol{\pi}(t,t_s) \tag{5.59}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\Pr(X(t)=1),\Pr(X(t)=2),\ldots,\Pr(X(t)=m)\right]^{\mathrm{T}} = \sum_{s:\ t>t_s} \dot{\boldsymbol{\pi}}(t,t_s) \quad (5.60)$$

respectively. The failure rate $h_{\text{sys}}(t)$ of the structure is given as

$$h_{\text{sys}}(t) = \frac{\mathrm{d}\Pr(X(t)=m)/\mathrm{d}t}{1-\Pr(X(t)=m)} \quad (5.61)$$

It should be mentioned, that the above modeling implicitly assumes that if rehabilitation is performed it is perfect, that is, the rehabilitated structure is 'as new'.

5.7.5 Optimization of Bridge Management Strategies

Analogous to Sec. 5.6, we will optimize the maintenance strategies for the above described truss-type bridge for different benefit rates. The basic cost factors are set to the same values as in Sec. 5.5, that is, initial cost, loss, inspection cost and discount rate are given, respectively, as $c_0 = 10$ m.u., $L = 2000$ m.u., $I = 2$ m.u. and $\gamma = 0.03$ per year. The rehabilitation costs are assumed to be $R_k = c_0/(7-k)$ with $k = 1, 2, \ldots, 5$. At time $t_0 = 0$ the structure is free of deterioration, that is, $D_{\text{sys}}(t_0) = 0$.

As a first example, let us assume a constant benefit rate of $\dot{B}_k = 0.7$ m.u. per year. Without inspection and rehabilitation, the bridge reaches its optimal lifetime at $T^* = 38.1$ years. At this time the expected net present benefit attains its maximum value of $g^*(T^*) = 5.370$ m.u. (see Table 5.9). The corresponding probability of failure up to this time is $\Pr(X(T^*) = m) = 7.7 \cdot 10^{-4}$.

Table 5.9: Optimal solutions for bridge maintenance in case of state-independent benefit rate.

n	0	1	2	3	4
T^* [year]	38.1	74.4	109.8	144.9	179.8
$g^*(T^*)$ [m.u.]	5.370	9.038	10.287	10.721	10.873
t_1^* [year]	–	36.3	35.4	35.1	35.0
t_2^* [year]	–	–	71.7	70.5	70.1
t_3^* [year]	–	–	–	106.8	105.5
t_4^* [year]	–	–	–	–	141.8
Δ_1^*	–	1	1	1	1
Δ_2^*	–	–	1	1	1
Δ_3^*	–	–	–	1	1
Δ_4^*	–	–	–	–	1

To extend the lifetime of the bridge beyond its 38.1 years, we maximize the expected net present benefit rate by optimizing the maintenance times $\mathbf{t}^*$ and the rehabilitation levels $\boldsymbol{\Delta}^*$ for a given number n of maintenance interventions. The results are listed in Table 5.9 and shown exemplarily for $n = 3$ in Figs. 5.22 and 5.23. As in previous examples with a time-homogeneous matrix $\mathbf{A}$, the optimal strategy requires to perform rehabilitation efforts at almost periodic intervals, and before reaching the acceptable failure rate of $h_{\mathrm{a}}(t) = 3.5 \cdot 10^{-4}$. Due to the small

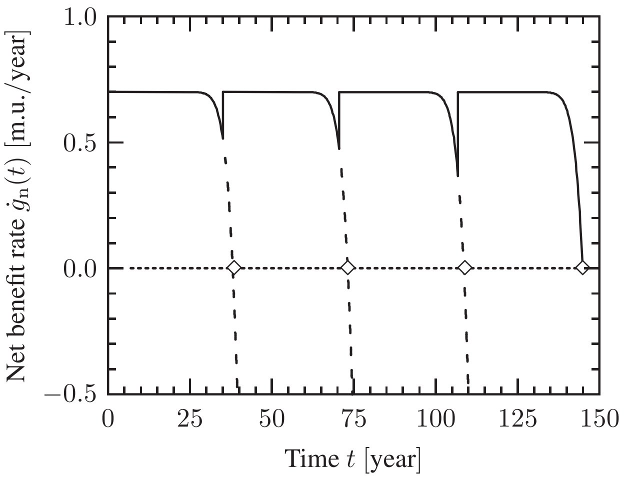

Figure 5.22: Expected net benefit rate $\dot{g}_{\mathrm{n}}(t)$ for bridge management problem in case of state-independent benefit rate ($n = 3$, solid line).

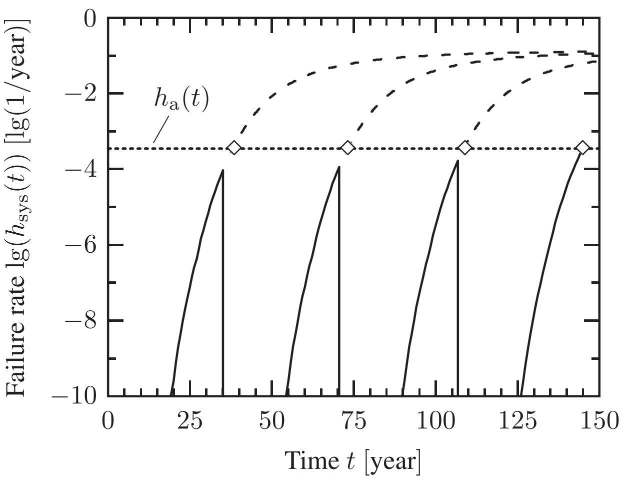

Figure 5.23: Failure rate $h_{\mathrm{sys}}(t)$ for bridge management problem in case of state-independent benefit rate ($n = 3$, solid line).

number of condition states ($m = 6$) there is no variation in the optimal rehabilitation level Δ_j^* present, that is, at maintenance time t_j all structures with a condition state 2 or higher get rehabilitated to an 'as new' state.

Next we assume that the benefit rates decrease linearly with the deterioration state k as

$$\dot{B}_k = 1.0 \times \frac{m-k}{m-1} \text{ m.u. per year} \quad (k = 1, \ldots, 5) \tag{5.62}$$

All other quantities remain the same as in the previous example. The optimal maintenance interventions for state-dependent benefit rates and different number n of interventions are given in Table 5.10. As before, the expected net present benefit

Table 5.10: Optimal solutions for bridge management problem in case of state-dependent benefit rates.

n	1	2	3	4
T^* [year]	73.7	107.4	139.7	171.5
$g^*(T^*)$ [m.u.]	16.546	18.504	19.230	19.508
t_1^* [year]	35.5	33.6	32.4	31.7
t_2^* [year]	–	69.2	66.0	64.1
t_3^* [year]	–	–	101.5	97.7
t_4^* [year]	–	–	–	133.3
Δ_1^*	1	1	1	1
Δ_2^*	–	1	1	1
Δ_3^*	–	–	1	1
Δ_4^*	–	–	–	1

$g^*(T^*)$ increases monotonically with time. Also, the qualitative behavior of the optimal solution is like the one for constant benefit rates, although, the times between maintenance interventions got shortened somehow due to the state-dependency of the benefit rates. In fact, the discussion from Sec. 5.6.2 could be reiterated here in its entirety.

The same also holds when the benefit rate is modeled as being state- and time-dependent:

$$\dot{B}_k(t) = 1.0 \times \frac{m-k}{m-1} \exp(-0.015t) \text{ m.u. per year} \quad (k = 1, \ldots, 5) \tag{5.63}$$

In Table 5.11 the optimal solutions are given for $1 \leq n \leq 4$ number of maintenance interventions. As can be seen, the expected net present benefit reaches its maximum at $n^* = 2$ maintenance interventions. Hence, the optimal lifetime is $T^* = 102.0$ years. From the display of the expected net benefit rate $\dot{g}_{\mathrm{n}}(t)$ in Fig. 5.24, small dents at around 25 years after, respectively, construction or rehabil-

Table 5.11: Optimal solutions for bridge management problem in case of state- and time-dependent benefit rates.

n	1	2	3	4
T^* [year]	69.9	102.0	134.7	168.0
$g^*(T^*)$ [m.u.]	8.961	9.201	9.179	9.146
t_1^* [year]	34.1	33.1	33.0	33.0
t_2^* [year]	–	67.1	66.8	66.9
t_3^* [year]	–	–	100.8	101.0
t_4^* [year]	–	–	–	134.9
Δ_1^*	1	1	1	1
Δ_2^*	–	1	1	1
Δ_3^*	–	–	1	1
Δ_4^*	–	–	–	1

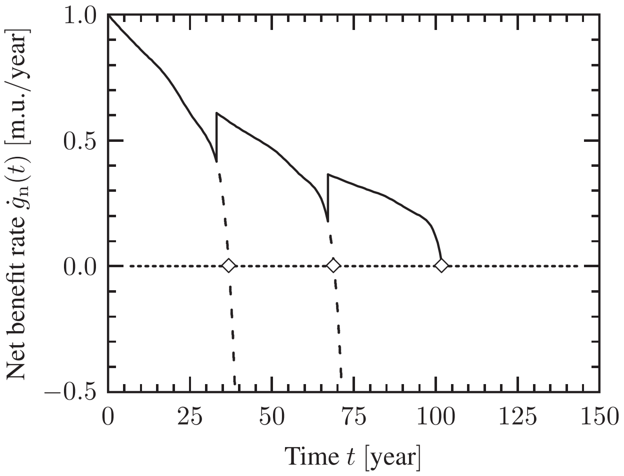

Figure 5.24: Expected net benefit rate $\dot{g}_n(t)$ for bridge management problem in case of state- and time-dependent benefit rates ($n^* = 2$, solid line).

itation can be noted. At this time the most likely damage state is state 2. And due to the 'coarseness' of the damage states and the corresponding benefit rates this transition from state 1 to state 2 can be seen in Fig. 5.24. Or, in the similar example of Sec. 5.6.3 such dents were not visible doe to the higher number of condition states and the more smooth transitions in the corresponding benefit rates.

Finally, let us investigate the influence of different discount rates γ on the optimal solution of $n = 2$ maintenance interventions from the previous example. In Table 5.12 the solutions for $\gamma = 0.00, 0.01, \ldots, 0.06$ per year are listed. As can be seen, the effect of different discount rates have a tremendous effect on the value of

Table 5.12: Optimal solutions for bridge management problem in case of state- and time-dependent benefit rates (with $n = 2$) for different discount rates γ.

γ [per year]	0.00	0.01	0.02	0.03	0.04	0.05	0.06
T^* [year]	102.6	101.7	101.5	102.0	102.7	103.5	104.2
$g^*(T^*)$ [m.u.]	31.07	20.10	13.49	9.20	6.24	4.08	2.43
t_1^* [year]	33.5	32.8	32.7	33.1	33.6	34.1	34.5
t_2^* [year]	67.7	66.8	66.6	67.1	67.8	68.6	69.4
Δ_1^*	1	1	1	1	1	1	1
Δ_2^*	1	1	1	1	1	1	1

the net present benefit $g^*(T^*)$, as should be expected when discounting on a time horizon of more than one hundred years. However, the maintenance times and the lifetimes differ only by an almost negligible amount. Also, there is no general—even minor—trend visible like, for example, that higher discount rates would shift all maintenance interventions to later points in time.

6
Conclusions

Although there has been in recent years an increasing interest in the lifetime extension of existing structures, the offered solutions so far have been of purely technical nature. That is, rehabilitation methods have been developed which allow to extend the structural lifetime by improving, say, its reliability and durability, but the question whether these efforts are also justified in economic terms has been basically not answered. This is all the more astonishing, since it have been originally monetary reasons—the realization that a growing percentage of civil infrastructure and buildings is threatened by obsolescence, and that it is no longer economically feasible to counter this by re-building everything anew—that have been the cause for the interest in lifetime extensions.

The problem gets further aggravated by the fact that maintenance or rehabilitation planning requires forecasts into the future with respect to, for example, maximum loads or the evolution of the load carrying capacity in case of deterioration, which are all inherently random processes. Hence, for a consistent treatment of these uncertainties structural reliability methods are indispensable. However, structural reliability methods traditionally focus on the probability of failure as a measure of safety, which is specified for its lifetime of, say, fifty years. This criterion is also quite unanimously utilized for maintenance planning, although, strictly speaking, it is the failure rate which characterizes at each point in time the hazardousness of a (deteriorating) structure.

As has been shown in this work, both problems—the economic justification of rehabilitation efforts and the consistent selection of acceptable safety criteria—are closely related. For this purpose the maintenance optimization problem has been formulated in the framework of cost-benefit analysis. The utilization of cost-benefit analysis as a decision-aiding rationale allows us not only to disclose all cost and benefit streams, but also provides us with rational criteria for, on the one hand, decisions on investments and, on the other hand, determining optimal structural lifetimes and acceptable failure rates—the latter when used in combination with structural reliability theory. For achieving this, all significant life-cycle costs—such as construction, failure, inspection and rehabilitation costs—as well as time- and state-dependent benefit rates have to be taken into account.

The optimization of the maintenance interventions is performed herein by maximizing the net benefit rate throughout the lifetime of the structure. Design variables are maintenance times, rehabilitation levels, inspection qualities, the structural lifetime and the number of maintenance interventions. In this way a maintenance policy which is quite straight forward to implement can be determined in terms of directly or indirectly observable indicators of structural deterioration. Since the resulting optimization problem for maintenance planning is a mixed-discrete one, we utilize an evolutionary algorithm for its solution. Such algorithms can be interpreted—without any recursion to biological terminology—as adaptively-directed random search methods. It is also shown, that there is no need to utilize binary representations, since operators for recombination and mutation can be analogously constructed for integer or real-valued representations of design variables.

For being able to describe consistently the time-variant condition of a structure in terms of both performance and ultimate failure states, we explicitly model the deterioration process and the respective maintenance interventions, since it can not be done by solely utilizing inspection data or so called expert knowledge. To achieve a certain compatibility with existing condition rating procedures of existing infrastructure or building management systems, we describe the structural condition by a finite number of states. The probability evolution between maintenance interventions is governed by a continuous-time Markov chain. The time-homogeneous or time-inhomogeneous transition matrices are derived, respectively, from Markovian deterioration models or, more generally, from time distributions of reaching defined deterioration states. The maintenance interventions at certain points in the lifetime then act like pulses 'pushing back' the structural condition to lower deterioration or 'as new' states.

The numerical examples demonstrate the importance of defining benefit rates explicitly. As has been shown, state-dependent benefit rates result in markedly different optimal maintenance times as compared to benefit rates which are state-independent, whereas only slight differences in optimal maintenance times can be observed between time-dependent and time-independent benefit rates, respectively. Also, imperfect inspection or rehabilitation results in performing maintenance interventions more often than in the case of perfect inspection and rehabilitation. If the preference towards the structure as expressed in terms of benefit rates and costs is not changing with time, maintenance interventions allow, in principle, to extend the lifetime of a structure infinitely. This does not mean that there is no failure at all, but that the failure rate—the probability of failure per, say, year, under the condition, that the structure has not failed so far—is always less than the acceptable failure rate.

If, however, benefit rates, costs or the system matrix are time-variant, then it may happen that at a certain time the expenditures spent on maintenance no longer outweigh the expected future net benefit, that is, the structure becomes obsolete. Nevertheless, in all these cases the optimal solution to maintenance interventions requires to take action before reaching the acceptable failure rate or the zero expected net benefit rate level. This is also an interesting finding, since so far reliability-based optimization procedures quite often implicitly assume that an optimal maintenance solution is one where the maintenance actions are executed when the failure rate reaches the acceptable level. In general we have to say, that deferring decisions with respect to maintenance not only results in higher economic losses, but also compromises safety issues, that is, non-action leads to potentially hazardous structures. Thus, to summarize, the approach for cost-benefit based maintenance optimization presented herein allows to determine consistently maintenance policies which are at the same time economically optimal and 'safe enough'. And, as is equally important for decision-making, it discloses the interaction between economic decisions concerning further investments (in the form of inspection and rehabilitation expenditures), the benefits that can be gained, and the implied risks to society.

Nevertheless, there still remain some topics which warrant further investigation. Most closely related to the work herein is the need for efficient methods for calculating first-passage densities or failure rates. A first attempt in this direction has been made in (Joanni and Rackwitz, 2005), where it was shown that not only different approximation methods, but also Monte Carlo sampling can in principle be utilized to solve the required multi-variate normal integrals. But this needs certainly further investigation, especially with respect to somehow more complex problems as given in (Joanni and Rackwitz, 2005). Also, to optimize jointly initial structural design and maintenance interventions remains an interesting issue to explore.

With respect to the cost parameters, more clarification is needed with respect to failure costs and how to determine benefits. Whereas, for example, benefit forecasts in public road transportation projects are on the average quite reasonable, in rail transportation projects benefits get systematically and strongly overestimated (Flyvbjerg et al., 2005). A solution to such problems is certainly not to dismiss benefit forecasts at all, since in road planning they are more or less working, but to use mechanisms of transparency and accountability to mitigate such systematic misrepresentations. In case of failure costs the question remains whether indirect costs should be included or not. This is closely related to the determination of acceptable failure rates. In this work we have chosen the failure costs such that they correspond approximately to a failure rate as specified in International Standard ISO 2394 (1998). But this results in comparatively high failure costs, so seemingly indicating that indirect costs should indeed be included—at least as long as the current acceptable failure rates are not getting revised.

References

Achenbach, J. D. (2000), “Quantitative Nondestructive Evaluation,” *International Journal of Solids and Structures,* **37**, 13–27.

Adler, M. D. (2006), “QALYs and Policy Evaluation: A New Perspective,” *Yale Journal of Health Policy, Law, and Ethics,* **6**, 1–92.

Adler, M. D. and E. A. Posner (2000), “Implementing Cost-Benefit Analysis When Preferences Are Distorted,” *The Journal of Legal Studies,* **29**, 1105–1147.

Adler, M. D. and E. A. Posner (2006), *New Foundations of Cost-Benefit Analysis*, Harvard University Press, Cambridge.

Aktan, A. E., D. N. Farhey, D. L. Brown, V. Dalal, A. J. Helmicki, V. J. Hunt and S. J. Shelley (1996), “Condition Assessment for Bridge Management,” *Journal of Infrastructure Systems,* ASCE*,* **2**, 108–117.

Ang, A. H.-S. and J.-C. Lee (2001), “Cost Optimal Design of R/C Buildings,” *Reliability Engineering and System Safety,* **73**, 233–238.

Arnold, J. L. (1988), *The Evolution of the 1936 Flood Control Act*, Office of History, United States Army Corps of Engineers, Fort Belvoir.

Arora, J. S. (2002), “Methods for Discrete Variable Structural Optimization,” in: S. A. Burns (ed.), *Recent Advances in Optimal Structural Design*. ASCE, Reston*,* 1–40.

Arora, J. S., M. W. Huang and C. C. Hsieh (1994), “Methods for Optimization of Nonlinear Problems with Discrete Variables: A Review,” *Structural Optimization,* **8**, 69–85.

Arora, J. S. and P. B. Thanedar (1986), “Computational Methods for Optimum Design of Large Complex Systems,” *Computational Mechanics,* **1**, 221–242.

Arrow, K. J., W. R. Cline, K.-G. Maler, M. Munasinghe, R. Squitieri and J. E. Stiglitz (1996), “Intertemporal Equity, Discounting, and Economic Efficiency,” in: J. P. Bruce, H. Lee, and E. F. Haites (eds.), *Climat Change 1995: Economic and Social Dimensions of Climate Change*. Cambridge University Press, Cambridge*,* 125–144.

Baker, J. E. (1987), “Reducing Bias and Inefficiency in the Selection Algorithm,” in: J. J. Grefenstette (ed.), *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms*. Erlbaum, Hillsdale, 14–21.

Barlow, R. E. and F. Proschan (1965), *Mathematical Theory of Reliability*, Wiley, New York.

Bell, D. E. and P. C. Fishburn (2000), "Utility Functions for Wealth," *Journal of Risk and Uncertainty,* **20**, 5–44.

Bernoulli, D. (1738), "Specimen Theoriae Novae de Mensura Sortis," *Commentarii Academiae Scientiarum Imperialis Petropolitanae,* **5**, 175–192. Translated from Latin into English by L. Sommer as "Exposition of a New Theory on the Measurement of Risk," *Econometrica,* **22**, 23–36 (1954).

Blickle, T. (2000), "Tournament Selection," in: T. Bäck, D. B. Fogel, and Z. Michalewicz (eds.), *Evolutionary Computation 1: Basic Algorithms and Operators*. Institute of Physics Publishing, Bristol, 181–186.

Boardman, A. E., D. H. Greenberg, A. R. Vining and D. L. Weimer (2006), *Cost-Benefit Analysis: Concepts and Practice*, Pearson Prentice Hall, Upper Saddle River.

Bogdanoff, J. L. (1978), "A New Cumulative Damage Model—Part 1," *Journal of Applied Mechanics,* ASME, **45**, 246–250.

Bogdanoff, J. L. and W. Krieger (1978), "A New Cumulative Damage Model—Part 2," *Journal of Applied Mechanics,* ASME, **45**, 251–257.

Broder, I. E. (1990), "The Cost of Accidental Death: A Capital Market Approach," *Journal of Risk and Uncertainty,* **3**, 51–63.

Broding, W. C., F. W. Diederich and P. S. Parker (1964), "Structural Optimization and Design Based on a Reliability Design Criterion," *Journal of Spacecraft and Rockets,* **1**, 56–61.

Brown, M. and F. Proschan (1983), "Imperfect Repair," *Journal of Applied Probability,* **20**, 851–859.

Cai, J. and G. Thierauf (1996), "Evolution Strategies for Solving Discrete Optimization Problems," *Advances in Engineering Software,* **25**, 177–183.

Cesare, M. A., C. Santamarina, C. Turkstra and E. H. Vanmarcke (1992), "Modeling Bridge Deterioration with Markov Chains," *Journal of Transportation Engineering,* ASCE, **118**, 820–833.

Chaboche, J. L. (1988), "Continuum Damage Mechanics: Part I—General Concepts," *Journal of Applied Mechanics,* ASME, **55**, 59–64.

Cowen, T. and D. Parfit (1992), "Against the Social Discount Rate," in: P. Laslett and J. S. Fishkin (eds.), *Justice Between Age Groups and Generations*. Yale University Press, New Haven, 144–161.

Cox, D. R. and H. D. Miller (2001), *The Theory of Stochastic Processes*, Chapman & Hall/CRC, Boca Raton.

Czarnecki, A. A. and A. S. Nowak (2008), “Time-Variant Reliability Profiles for Steel Girder Bridges,” *Structural Safety,* **30**, 49–64.

Das, P. C. (1998), “New Developments in Bridge Management Methodology,” *Structural Engineering International,* **8**, 299–302.

De Brito, J., F. A. Branco, P. Thoft-Christensen and J. D. Sørensen (1997), “An Expert System for Concrete Bridge Management,” *Engineering Structures,* **19**, 519–526.

De Jong, K., D. B. Fogel and H.-P. Schwefel (2000), “A History of Evolutionary Computation,” in: T. Bäck, D. B. Fogel, and Z. Michalewicz (eds.), *Evolutionary Computation 1: Basic Algorithms and Operators*. Institute of Physics Publishing, Bristol, 40–58.

Deb, K. and R. B. Agrawal (1995), “Simulated Binary Crossover for Continuous Search Space,” *Complex Systems,* **9**, 115–148.

Deb, K. and H.-G. Beyer (2001), “Self-Adaptive Genetic Algorithms with Simulated Binary Crossover,” *Evolutionary Computation,* **9**, 197–221.

Donohue III, J. J. (1999), “Why We Should Discount the Views of Those Who Discount Discounting,” *Yale Law Journal,* **108**, 1901–1910.

Eiben, A. E. and J. E. Smith (2003), *Introduction to Evolutionary Computing*, Springer, Berlin.

Ellingwood, B. R. (2001), “Acceptable Risk Bases for Design of Structures,” *Progress in Structural Engineering and Materials,* **3**, 170–179.

Enevoldsen, I. and J. D. Sørensen (1994), “Reliability-Based Optimization in Structural Engineering,” *Structural Safety,* **15**, 169–196.

Federal Ministry of Transport, Building and Housing of the Federal Republic of Germany (2005), *Federal Transport Infrastructure Plan 2003: Macroeconomic Evaluation Methodology*, Federal Ministry of Transport, Building and Housing, Berlin.

Fischhoff, B., S. Lichtenstein, P. Slovic, S. L. Derby and R. L. Keeney (1981), *Acceptable Risk*, Cambridge University Press, Cambridge.

Fishburn, P. C. (1987), “Reconsiderations in the Foundations of Decision under Uncertainty,” *The Economic Journal,* **97**, 825–841.

Fletcher, R. (1987), *Practical Methods of Optimization*, Wiley, Chichester.

Flyvbjerg, B., M. K. S. Holm and S. L. Buhl (2005), “How (In)accurate Are Demand Forecasts in Public Work Projects?,” *Journal of the American Planning Association,* **71**, 131–146.

Flyvbjerg, B., M. S. Holm and S. Buhl (2002), “Underestimating Costs in Public Work Projects: Error or Lie?,” *Journal of the American Planning Association,* **68**, 279–295.

Fogel, L. J., A. J. Owens and M. J. Walsh (1966), *Artificial Intelligence Through Simulated Evolution*, Wiley, New York.

Frangopol, D. M. (1985), "Structural Optimization Using Reliability Concepts," *Journal of Structural Engineering,* ASCE, **111**, 2288–2301.

Frangopol, D. M., J. S. Kong and E. S. Gharaibeh (2001), "Reliability-Based Life-Cycle Management of Highway Bridges," *Journal of Computing in Civil Engineering,* ASCE, **15**, 27–34.

Frangopol, D. M., K.-Y. Lin and A. C. Estes (1997), "Life-Cycle Cost Design of Deteriorating Structures," *Journal of Structural Engineering,* ASCE, **123**, 1390–1401.

Frangopol, D. M. and F. Moses (1994), "Reliability-Based Structural Optimization," in: H. Adeli (ed.), *Advances in Design Optimization*. Chapmann and Hall, London, 492–570.

Freudenthal, A. M. (1956), "Safety and the Probability of Structural Failure," *Transactions of the ASCE,* **121**, 1337–1397.

Fuguitt, D. and S. J. Wilcox (1999), *Cost-Benefit Analysis for Public Sector Decision Makers*, Quorum, Westport.

Gansted, L., R. Brincker and L. P. Hansen (1991), "Fracture Mechanical Markov Chain Crack Growth Model," *Engineering Fracture Mechanics,* **38**, 475–489.

Ghobarah, A. (2001), "Performance-Based Design in Earthquake Engineering: State of Development," *Engineering Structures,* **23**, 878–884.

Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Boston.

Great Britain H. M. Treasury (2003), *The Green Book: Appraisal and Evaluation in Central Government*, TSO, London.

Grefenstette, J. (2000a), "Proportional Selection and Sampling Algorithms," in: T. Bäck, D. B. Fogel, and Z. Michalewicz (eds.), *Evolutionary Computation 1: Basic Algorithms and Operators*. Institute of Physics Publishing, Bristol, 172–180.

Grefenstette, J. (2000b), "Rank-Based Selection," in: T. Bäck, D. B. Fogel, and Z. Michalewicz (eds.), *Evolutionary Computation 1: Basic Algorithms and Operators*. Institute of Physics Publishing, Bristol, 187–194.

Grierson, D. E. and W. H. Pak (1993), "Optimal Sizing, Geometrical and Topological Design Using a Genetic Algorithm," *Structural Optimization,* **6**, 151–159.

Gusella, V. (2000), "Eigen-Analysis of Cumulative Damage," *Structural Safety,* **22**, 189–202.

Hanganu, A. D., E. Oñate and A. H. Barbat (2002), "A Finite Element Methodology for Local/Global Damage Evaluation in Civil Engineering Structures," *Comput-*

ers and Structures, **80**, 1667–1687.

Hawk, H. and E. P. Small (1998), "The BRIDGIT Bridge Management System," *Structural Engineering International,* **8**, 309–314.

Hearn, G. (1998a), "Condition Data and Bridge Management Systems," *Structural Engineering International,* **8**, 221–225.

Hearn, M. D. (1998b), "Integration of Bridge Management Systems and Nondestructive Evaluations," *Journal of Infrastructure Systems,* ASCE*,* **4**, 49–55.

Heinzerling, L. (1999), "Discounting Life," *Yale Law Journal,* **108**, 1911–1915.

Hilton, H. H. and M. Feigen (1960), "Minimum Weight Analysis Based on Structural Reliability," *Journal of the Aerospace Sciences,* **27**, 641–652.

Holland, J. H. (1992), *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge.

Horn, D. and W. R. Mayo (2000), "NDE Reliability Gains from Combining Eddy-Current and Ultrasonic Testing," *NDT & E International,* **33**, 351–362.

Horn, R. A. and C. R. Johnson (1990), *Matrix Analysis*, Cambridge University Press, Cambridge.

Hubin, D. C. (1994), "The Moral Justification of Benefit/Cost Analysis," *Economics and Philosophy,* **10**, 169–194.

International Standard ISO 2394 (1998), *General Principles on Reliability of Structures*, International Organization for Standardization, Geneva.

Jiang, Y., M. Saito and K. C. Sinha (1989), "Bridge Performane Prediction Model using the Markov Chain," *Transportation Research Record,* **1180**, 25–32.

Joanni, A. E. and R. Rackwitz (2005), "High-Dimensional Multivariate Normal Integrals for a Discrete Approximation of the First-Passage Density," in: J. D. Sørensen and D. M. Frangopol (eds.), *Proceedings of the Twelfth WG 7.5 Working Conference on Reliability and Optimization of Structural Systems*. Taylor and Francis, London.

Johnson, A. I. (1953), *Strength, Safety and Economical Dimensions of Structures*, Bulletins of the Division of Building Statics and Structural Engineering at the Royal Institute of Technology, Stockholm.

Jong, J.-C. and P. Schonfeld (2001), "Genetic Algorithm for Selecting and Scheduling Interdependent Projects," *Journal of Waterway, Port, Coastal, and Ocean Engineering,* ASCE*,* **127**, 45–52.

Kalaba, R. (1962), "Design of Minimal-Weight Structures for Given Reliability and Cost," *Journal of the Aerospace Sciences,* **29**, 355–356.

Kallen, M. J. and J. M. van Noortwijk (2006), "Statistical Inference for Markov Deterioration Models of Bridge Conditions in the Netherlands," in: P. J. S. Cruz, D. M. Frangopol, and L. C. Neves (eds.), *Proceedings of the Third International*

Conference on Bridge Maintenance, Safety and Management. Taylor and Francis, London.

Kanda, J. (1996), “Normalized Failure Cost as a Measure of a Structure Importance,” *Nuclear Engineering and Design,* **160**, 299–305.

Kanda, J. and H. Shah (1997), “Engineering Role in Failure Cost Evaluation for Buildings,” *Structural Safety,* **19**, 79–90.

Kang, Y. K. W. Y. J. (2001), “Minimum Building Life-Cycle Cost Design Criteria. II: Applications,” *Journal of Structural Engineering,* ASCE*,* **127**, 338–346.

Keeney, R. L. and H. Raiffa (1993), *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, Canbridge University Press, Cambridge.

Kirk, S. J. and A. J. Dell’Isola (1995), *Life Cycle Costing for Design Professionals*, McGraw-Hill, New York.

Kirsch, U. (1993), *Structural Optimization: Fundamentals and Applications*, Springer, Berlin.

Kniesner, T. J. and J. D. Leeth (1991), “Compensating Wage Differentials for Fatal Injury Risk in Australia, Japan, and the United States,” *Journal of Risk and Uncertainty,* **4**, 75–90.

Koza, J. R. (1992), *Genetic Programming*, MIT Press, Cambridge.

Kuschel, N. and R. Rackwitz (2000), “Optimal Design under Time-Variant Reliability Constraints,” *Structural Safety,* **22**, 113–127.

Lassen, T. (1991), “Markov Modelling of the Fatigue Damage in Welded Structures under In-Service Inspection,” *International Journal of Fatigue,* **13**, 417–422.

Lemaitre, J. and J.-L. Chaboche (1990), *Mechanics of Solid Materials*, Cambridge University Press, Cambridge.

Levitin, G. and A. Lisnianski (2000), “Optimization of Imperfect Preventive Maintenance for Multi-State Systems,” *Reliability Engineering and System Safety,* **67**, 193–203.

Levy, H. and H. M. Markowitz (1979), “Approximating Expected Utility by a Function of Mean and Variance,” *American Economic Review,* **69**, 308–317.

Levy, M. and M. Salvadori (2002), *Why Buildings Fall Down: How Structures Fail*, Norton, New York.

Lichtenstein, A. G. (1993), “The Silver Bridge Collapse Recounted,” *Journal of Performance of Constructed Facilities,* ASCE*,* **7**, 249–261.

Lind, N. C. (1993), “A Compound Index of National Development,” *Social Indicators Research,* **28**, 267–284.

Liu, M. and D. M. Frangopol (2004), “Optimal Bridge Maintenance Planning Based on Probabilistic Performance Prediction,” *Engineering Structures,* **26**, 991–1002.

Luce, R. D. and H. Raiffa (1957), *Games and Decisions*, Wiley, New York.

Madanat, S., R. Mishalani and W. H. Wan Ibrahim (1995), "Estimation of Infrastructure Transition Probabilities from Condition Rating Data," *Journal of Infrastructure Systems,* ASCE, **1**, 120–125.

Madsen, H. O., S. Krenk and N. C. Lind (1986), *Methods of Structural Safety*, Prentice-Hall, Englewood Cliffs.

Madsen, H. O., J. D. Sørensen and R. Olesen (1990), "Optimal Inspection Planning for Fatigue Damage of Offshore Structures," in: A. H.-S. Ang, M. Shinozuka, and G. I. Schuëller (eds.), *Proceedings of the Fifth International Conference on Structural Safety and Reliability*. ASCE, New York, 2099–2106.

Mau, S.-T. and R. G. Sexsmith (1972), "Minimum Expected Cost Optimization," *Journal of the Structural Division,* ASCE, **98**, 2043–2058.

Merkhofer, M. W. (1987), *Decision Science and Social Risk Management: A Comparative Evaluation of Cost-Benefit Analysis, Decision Analysis, and Other Formal Decision-Aiding Approaches*, Reidel, Dordrecht.

Minchin, Jr., R. E., T. Zayed, A. J. Boyd and M. Mendoza (2006), "Best Practices of Bridge System Management—A Synthesis," *Journal of Management in Engineering,* **22**, 186–195.

Miyamoto, A., K. Kawamura and H. Nakamura (2001), "Development of a Bridge Management System for Existing Bridges," *Advances in Engineering Software,* **32**, 821–833.

Miyamoto, A., K. Kawamura and K. C. G. Ong (2003), "Bridge Management System for Existing Bridge Groups," in: H. Furuta, M. Dogaki, and M. Sakano (eds.), *Proceedings of the Tenth IFIP WG7.5 Working Conference on Reliability and Optimization of Structural Systems*. Balkema, Lisse, 69–78.

Mori, Y. and B. R. Ellingwood (1994a), "Maintaining Reliability of Concrete Structures. I: Role of Inspection/Repair," *Journal of Structural Engineering,* ASCE, **120**, 824–845.

Mori, Y. and B. R. Ellingwood (1994b), "Maintaining Reliability of Concrete Structures. II: Optimum Inspection/Repair," *Journal of Structural Engineering,* ASCE, **120**, 846–862.

Morrison, E. R. (1998), "Judicial Review of Discount Rates Used in Regulatory Cost-Benefit Analysis," *University of Chicago Law Review,* **65**, 1333–1369.

Moses, F. (1977), "Structural System Reliability and Optimization," *Computers and Structures,* **7**, 283–290.

Moses, F. (1997), "Problems and Prospects of Reliability-Based Optimization," *Engineering Structures,* **19**, 293–301.

Moses, F. and D. E. Kinser (1967), "Optimum Structural Design with Failure Probability Constraints," *AIAA Journal,* **5**, 1152–1158.

Neves, L. C. and D. M. Frangopol (2005), "Condition, Safety and Cost Profiles for Deteriorating Structures with Emphasis on Bridges," *Reliability Engineering and System Safety,* **89**, 185–198.

Nocedal, J. and S. J. Wright (1999), *Numerical Optimization*, Springer, New York.

Pandey, M. D. (1998), "Probabilistic Models for Condition Assessment of Oil and Gas Pipelines," *NDT & E International,* **31**, 349–358.

Pandey, M. D. and J. S. Nathwani (1997), "Measurement of Socio-Economic Inequality Using the Life-Quality Index," *Social Indicators Research,* **39**, 187–202.

Parimi, S. R. and M. Z. Cohn (1975), "Optimal Criteria in Probabilistic Structural Design," in: A. Sawczuk and Z. Mróz (eds.), *Optimization in Structural Design*. Springer, Berlin, 278–293.

Park, Y.-J., A. H.-S. Ang and Y. K. Wen (1985), "Seismic Damage Analysis of Reinforced Concrete Buildings," *Journal of Structural Engineering,* ASCE, **111**, 740–757.

Pearce, D. (2006), "Is the Construction Sector Sustainable?: Definitions and Reflections," *Building Research and Information,* **34**, 201–207.

Pearce, D. W. (1981), "Risk Assessment: Use and Misuse," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences,* **376**, 181–192.

Peoples, R. and K. Willcox (2006), "Value-Based Multidisciplinary Optimization of Commercial Aircraft Design and Business Risk Assessment," *Journal of Aircraft,* **43**, 913–921.

Pezeshk, S. and C. V. Kamp (2002), "Sate of the Art in the Use of Genetic Algorithms in Design of Steel Structures," in: S. A. Burns (ed.), *Recent Advances in Optimal Structural Design*. ASCE, Reston, 55–79.

Pham, H. and H. Wang (1996), "Imperfect Maintenance," *European Journal of Operational Research,* **94**, 425–438.

Pierskalla, W. P. and J. A. Voelker (1976), "A Survey of Maintenance Models: The Control and Surveillance of Deteriorating Systems," *Naval Research Logistics Quarterly,* **23**, 353–388.

Pratt, J. W., H. Raiffa and R. Schlaifer (1964), "The Foundations of Descision Under Uncertainty: An Elementary Exposition," *Journal of the American Statistical Association,* **59**, 353–375.

Pu, Y., P. K. Das and D. Faulkner (1997), "A Strategy for Reliability-Based Optimization," *Engineering Structures,* **19**, 276–282.

Rackwitz, R. (2000), "Optimization—The Basis of Code-Making and Reliability

Verification," *Structural Safety,* **22**, 27–60.

Rackwitz, R. (2001), "Optimizing Systematically Renewed Structures," *Reliability Engineering and System Safety,* **73**, 269–279.

Rackwitz, R. (2002), "Optimization and Risk Acceptability Based on the Life Quality Index," *Structural Safety,* **24**, 297–331.

Rackwitz, R. (2004), "Optimal and Acceptable Technical Facilities Involving Risks," *Risk Analysis,* **24**, 675–3695.

Rackwitz, R. (2006), "The Effect of Discounting, Different Mortality Reduction Schemes and Predictive Cohort Life Tables on Risk Acceptability Criteria," *Reliability Engineering and System Safety,* **91**, 469–484.

Rackwitz, R. and R. Cuntze (1987), "Formulations of Reliability-Oriented Optimization," *Engineering Optimization,* **11**, 69–76.

Rackwitz, R., A. Lentz and M. Faber (2005), "Socio-Economically Sustainable Civil Engineering Infrastructures by Optimization," *Structural Safety,* **27**, 187–229.

Rafiq, M. Y., J. D. Mathews and G. N. Bullock (2003), "Conceptual Building Design—Evolutionary Approach," *Journal of Computing in Civil Engineering,* ASCE*,* **17**, 150–158.

Rao, S. S. (1996), *Engineering Optimization: Theory and Practice*, Wiley, New York.

Rens, K. L., T. J. Wipf and F. W. Klaiber (1997), "Review of Nondestructive Evaluation Techniques of Civil Infrastructure," *Journal of Performance of Constructed Facilities,* ASCE*,* **11**, 152–160.

Roelfstra, G., R. Hajdin, B. Adey and E. Brühwiler (2004), "Condition Evolution in Bridge Management Systems and Corrosion-Induced Deterioration," *Journal of Bridge Engineering,* ASCE*,* **9**, 268–277.

Rosenblueth, E. and E. Mendoza (1971), "Reliability Optimization in Isostatic Structures," *Journal of the Engineering Mechanics Division,* ASCE*,* **97**, 1625–1641.

Ross, S. M. (1970), *Applied Probability Models with Optimization Applications*, Holden-Day, San Francisco.

Ross, S. M. (2003), *Introduction to Probability Models*, Academic Press, Amsterdam.

Royset, J. O., A. Der Kiureghian and E. Polak (2001), "Reliability-Based Optimal Structural Design by the Decoupling Approach," *Reliability Engineering and System Safety,* **73**, 213–221.

Rudolph, G. (1994a), "Convergence Analysis of Canonical Genetic Algorithms," *IEEE Transactions on Neural Networks,* **5**, 96–101.

Rudolph, G. (1994b), “Convergence of Non-Elitist Strategies,” in: Z. Michalewicz, J. D. Schaffer, H.-P. Schwefel, D. B. Fogel, and H. Kitano (eds.), *Proceedings of the First IEEE Conference on Evolutionary Computation*. IEEE Press, Piscataway, 63–66.

Rummel, W. D. (1998), “Back to Basics: Probability of Detection as a Quantitative Measure of Nondestructive Testing End-to-End Process Capabilities,” *Materials Evaluation,* **56**, 29.

Sarma, J. and K. De Jong (2000), “Generation Gap Methods,” in: T. Bäck, D. B. Fogel, and Z. Michalewicz (eds.), *Evolutionary Computation 1: Basic Algorithms and Operators*. Institute of Physics Publishing, Bristol, 205–211.

Sarma, K. C. and H. Adeli (1998), “Cost Optimization of Concrete Structures,” *Journal of Structural Engineering,* ASCE, **124**, 570–578.

Sarma, K. C. and H. Adeli (2002), “Life-Cycle Cost Optimization of Steel Structures,” *International Journal for Numerical Methods in Engineering,* **55**, 1451–1462.

Savage, L. (1954), *The Foundations of Statistics*, Wiley, New York.

Scherer, W. T. and D. M. Glagola (1994), “Markovian Models for Bridge Maintenance Management,” *Journal of Transportation Engineering,* ASCE, **120**, 37–51.

Schoemaker, P. J. H. (1982), “The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations,” *Journal of Economic Literature,* **20**, 529–563.

Schwefel, H.-P. (1995), *Evolution and Optimum Seeking*, Wiley, New York.

Shinozuka, M. (1983), “Basic Analysis of Structural Safety,” *Journal of Structural Engineering,* ASCE, **109**, 721–740.

Simo, J. C. and J. W. Ju (1987), “Strain- and Stress-Based Continuum Damage Models—I. Formulation,” *International Journal of Solids and Structures,* **23**, 821–840.

Simola, K. and U. Pulkkinen (1998), “Models for Non-Destructive Inspection Data,” *Reliability Engineering and System Safety,* **60**, 1–12.

Slovic, P. (1987), “Perception of Risk,” *Science,* **236**, 280–285.

Slovic, P. (2000), *The Perception of Risk*, Earthscan, London.

Sobczyk, K. and B. F. Spencer, Jr. (1992), *Random Fatigue: From Data to Theory*, Academic Press, Boston.

Söderqvist, M.-K. and M. Veijola (1998), “The Finnish Bridge Management System,” *Structural Engineering International,* **8**, 315–319.

Soh, C. K. and Y. Yang (2000), “Genetic Programming-Based Approach for Structural Optimization,” *Journal of Computing in Civil Engineering,* ASCE, **14**, 31–37.

Starr, C. (1969), "Social Benefit versus Technological Risk: What is Our Society Willing to Pay for Safety?," *Science,* **165**, 1232–1238.

Stewart, M. G. (2001), "Reliability-Based Assessment of Ageing Bridges Using Risk Ranking and Life Cycle Cost Decision Analyses," *Reliability Engineering and System Safety,* **74**, 263–273.

Streicher, H. and R. Rackwitz (2004), "Time-Variant Reliability-Oriented Structural Optimization and a Renewal Model for Life-Cycle Costing," *Probabilistic Engineering Mechanics,* **19**, 171–183.

Sugimoto, H., S. Sudo, A. Goto, T. Watanabe and T. Tamura (2002), "On Quantification of User Costs of Bridges in Hokkaido and Their Applications," *Structural Engineering/Earthquake Engineering,* JSCE*,* **19**, 33s–44s.

Sunstein, C. R. (2002), *Risk and Reason: Safety, Law, and the Environment*, Cambridge University Press, Cambridge.

Switzky, H. (1965), "Minimum Weight Design with Structural Reliability," *Journal of Aircraft,* **2**, 228–232.

Thanedar, P. B. and G. N. Vanderplaats (1995), "Survey of Discrete Variable Optimization for Structural Design," *Journal of Structural Engineering,* ASCE*,* **121**, 301–306.

Thoft-Christensen, P. and J. D. Sørensen (1987), "Optimal Strategy for Inspection and Repair of Structural Systems," *Civil Engineering Systems,* **4**, 94–100.

Thompson, P. D., E. P. Small, M. Johnson and A. R. Marshall (1998), "The Pontis Bridge Management System," *Structural Engineering International,* **8**, 303–308.

U. S. Federal Emergency Management Agency (1994a), *Seismic Rehabilitation of Federal Buildings: A Benefit/Cost Model. Volume 1: A User's Manual*, Federal Emergency Management Agency, Washington.

U. S. Federal Emergency Management Agency (1994b), *Seismic Rehabilitation of Federal Buildings: A Benefit/Cost Model. Volume 2: Supporting Documentation*, Federal Emergency Management Agency, Washington.

Val, D. V. (2005), "Effect of Different Limit States on Life-Cycle Cost of RC Structures in Corrosive Environment," *Journal of Infrastructure Systems,* ASCE*,* **11**, 231–240.

Valdez-Flores, C. and R. M. Feldman (1989), "A Survey of Preventive Maintenance Models for Stochastically Deteriorating Single-Unit Systems," *Naval Research Logistics,* **36**, 419–446.

Vergara, F. E., M. Khouja and Z. Michalewicz (2002), "An Evolutionary Algorithm for Optimizing Material Flow in Supply Chains," *Computers and Industrial Engineering,* **43**, 407–421.

Viscusi, W. K. (1993), "The Value of Risks to Life and Health," *Journal of Economic Literature,* **31**, 1912–1946.

Viscusi, W. K. and J. E. Aldy (2003), "The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World," *Journal of Risk and Uncertainty,* **27**, 5–76.

Von Neumann, J. and O. Morgenstern (1947), *Theory of Games and Economic Behavior*, Princeton University Press, Princeton.

Wang, H. (2002), "A Survey of Maintenance Policies of Deteriorating Systems," *European Journal of Operational Research,* **139**, 469–489.

Wen, Y. K. (2001), "Reliability and Performance-Based Design," *Structural Safety,* **23**, 407–428.

Yamada, H. (2006), "Role of Cost-Benefit Analysis in Public Projects," *Rippō to Chōsa,* **256**, in Japanese.

Yang, J. and C. K. Soh (1997), "Structural Optimization by Genetic Algorithms with Tournament Selection," *Journal of Computing in Civil Engineering,* ASCE*,* **11**, 195–200.

Zusammenfassung in Deutsch

EINFÜHRUNG

Die Entscheidung ein Tragwerk beziehungsweise eine bauliche Anlage zu errichten, instand zu halten oder umzunutzen basiert im Wesentlichen darauf, dass man sich davon einen zukünftig Nutzen erwartet, der die möglicherweise anfallenden Kosten aufwiegt oder übersteigt. Ein rationales Verfahren zur Unterstützung derartiger Entscheidungsfindungen ist die Kosten-Nutzen-Analyse, die, wie auch andere normative Entscheidungstheorien, Kriterien bestimmt, nach denen mögliche Alternativen zu vergleichen und hinsichtlich ihrer Konsequenzen zu bewerten sind, um die optimale Lösung zu einer Problemstellung zu finden. Obwohl Kostenbetrachtungen im Lebenszyklus-Management von Bauwerken zunehmend an Interesse gewinnen, wurden bisher Kosten-Nutzen-Betrachtungen zur optimalen Planung von Instandhaltungsmaßnahmen praktisch nicht eingesetzt.

Das Ziel der vorliegenden Arbeit ist es daher konsistent Kosten-Nutzen-Kriterien bei der Planung von Instandhaltungsmaßnahmen beziehungsweise der Verlängerung der Nutzungsdauer anzuwenden. Dabei werden sowohl die Unsicherheiten bezüglich zukünftiger Lasten und sich verändernder Tragfähigkeiten berücksichtigt, als auch der im Allgemeinen von der Gebrauchsfähigkeit abhängige Nutzen, die vom Tragwerkszustand abhängigen Kosten der Instandhaltung und die Kosten möglichen Tragwerkversagens. Dieses Vorgehen erlaubt auch die Festlegung von optimalen Nutzungsdauern und zulässigen Versagens- oder Ausfallraten, und macht somit die miteinander in Beziehung stehenden Anforderungen an die Wirtschaftlichkeit, Gebrauchsfähigkeit und Sicherheit von Tragwerken transparent.

ZUVERLÄSSIGKEITSORIENTIERTE OPTIMIERUNG

Optimierungsprobleme in der Strukturmechanik bestehen im Allgemeinen aus der Minimierung oder Maximierung einer Zielfunktion wie Gewicht oder Kosten, wobei die möglichen Lösungen in der Form von Nebenbedingungen bezüglich der auftretenden Spannungen, Verformungen, Schwingungszahlen, etc. restringiert sein können. Sind Unsicherheiten vorhanden, so treten in der Zielfunktion oder den Nebenbedingungen zusätzlich probabilistische Maße wie Erstversagenswahrscheinlichkeiten oder Versagensraten auf. Stand der Technik in der zuverlässigkeitsorientierten Strukturoptimierung ist es, für eine vorgegebene Nutzungsdauer den Erwartungswert aus Errichtungs- und Versagenskosten zu minimieren—meist unter der Nebenbedingung, das eine zulässiges Sicherheitsniveau eingehalten wird. Dabei wird bei der Formulierung der Zielfunktion zwischen den beiden Fällen der Aufgabe eines Tragwerks nach dem Versagen und des systematischen Wiederaufbaus unterschieden. Für den letzteren Fall kommen neuerdings auch Kosten-Nutzen-Kriterien zum Einsatz. Allerdings sind diese eben nur für unbeschränkte Zeithorizonte formuliert, erlauben also nicht die Bestimmung von Restnutzungsdauern und die Planung von Instandhaltungsmaßnahmen einzelner bestehender Bauwerke, sondern liefern nur Aussagen bezüglich der mittleren Nutzungsdauer und der mittleren Zeitdauer zwischen zwei Instandhaltungen. Auch werden nicht verschiedene Bauwerkszustände unterschieden, so dass nur die prophylaktische Erneuerung als Instandhaltungsmaßnahme möglich ist.

Auf dem Gebiet der zuverlässigkeitsorientierten Optimierung von Instandhaltungsplänen von Tragwerken werden zwar meist verschiedene Tragwerkszustände und die davon abhängigen Kosten der Inspektion und Instandsetzung unterschieden, aber es wird nicht der ebenfalls davon abhängige Nutzen berücksichtigt, weshalb in dieser Formulierung eine Nebenbedingung in Form eines zulässigen Sicherheitsniveaus erforderlich ist. Dies hat zur Folge, dass zwar die Kosten der Instandhaltungsmaßnahme minimiert werden, aber ob die Maßnahme selbst wirtschaftlich sinnvoll ist kann nicht festgestellt werden. Zudem wird das zulässige Sicherheitsniveau in den Nebenbedingungen einmütig mittels der Erstversagenswahrscheinlichkeit angegeben, was bei alternden Bauwerken und veränderlichen Zeitintervallen zu Inkonsistenzen führt. Es besteht also ein erheblicher Bedarf bei der Optimierung von Instandhaltungsplänen den jeweiligen Nutzen—gerade auch in Abhängigkeit von der Gebrauchsfähigkeit—zu berücksichtigen, sowie die Ausfallrate als zulässiges Sicherheitsmaß zu verwenden. Auch Aussagen zur (wirtschaftlich) optimalen Nutzungsdauer sind zu treffen.

GRUNDLAGEN DER KOSTEN-NUTZEN-ANALYSE

Bei der Entscheidungsfindung ist aus einer Menge möglicher Alternativen diejenige auszuwählen, deren Auswirkungen am ‘besten’ den jeweiligen Präferenzen entsprechen. Bei Entscheidungen unter Unsicherheiten sind auch die Auswirkungen unsicher, das heißt zufällig. Erfüllen jedoch die erwähnten Präferenzen die Axiome der Rationalität, so kann das Entscheidungsproblem auf die Maximierung des Erwartungswerts der Nutzenfunktion reduziert werden. In der Kosten-Nutzen-Analyse wird diese Nutzenfunktion in einen Kosten- und einen Nutzenanteil aufgespaltet, ist eine lineare Funktion dieser Anteile und wird in Geldeinheiten gemessen. Die Kosten-Nutzen-Analyse ist ein Instrument der Entscheidungsunterstützung, das heißt überwiegt für ein Vorhaben der zu erwartende Nutzen die zu erwartenden Kosten, ist also ein Nutzenüberschuss vorhanden, so sollte aus wohlfahrtstheoretischen Gründen das Vorhaben realisiert werden. Da dieser Ansatz normativ ist, kann der Entscheidungsträger jedoch von dieser Vorgabe abweichen. Kosten-Nutzen-Analysen werden in den Industriestaaten in zunehmenden Maße bei öffentlichen Vorhaben einer gewissen wirtschaftlichen Größenordnung eingesetzt und erlauben es den Entscheidungsprozess nachprüfbar zu gestalten.

Der Erwartungswert des Nutzens eines Bauwerks je Zeiteinheit bestimmt sich aus der Verfügbarkeit, das heißt der Wahrscheinlichkeit, dass das Bauwerk zum herausgegriffenen Zeitpunkt im gebrauchsfähigen Zustand ist, und dem Nutzen je Zeiteinheit, der aus der Existenz des Bauwerks gezogen wird. Die Höhe des Nutzens ist im Allgemeinen nicht nur vom technischen Zustand, sondern auch von den an ein Bauwerk gestellten Anforderungen abhängig, die sich im Laufe der Zeit ändern können. Der Erwartungswert der Versagenskosten je Zeiteinheit bestimmt sich aus der Versagensrate, der Wahrscheinlichkeit des Nichtversagens bis zum herausgegriffenen Zeitpunkt und den direkten und indirekten Kosten im Versagensfall. Die Kosten für Instandhaltungsmaßnahmen setzen sich zusammen aus den Kosten zur Schadensbestimmung und den Kosten zur Schadensbeseitigung. Die Höhe der Kosten sind dabei von den eingesetzten Verfahren und der Größe des Schadens abhängig. Dabei ist allerdings zu beachten, dass aufgrund der Unsicherheiten der Verfahren zur Zustandserfassung von Bauwerken nicht notwendigerweise alle Schäden erfasst und somit möglicherweise beseitigt werden, und dass es ein optimaler Instandhaltungsplan unter Umständen gar nicht zwingend erfordert jedweden Schaden zu beseitigen. In Übereinstimmung mit existierenden Managementsystemen für Bauwerke wird der Bauwerkszustand in diskreten Stufen dargestellt.

Der während des gesamten Lebenszyklus zu erwartende Nutzen als auch alle zu erwartenden Kosten werden kontinuierlich abgezinst und zum Nettonutzen zusammengefasst. Das Prinzip der Kosten-Nutzen-Analyse fordert, dass der Nettonutzen immer positiv ist. Eine optimale Lösung ist gegeben, wenn der Nettonutzen maximal wird. Dies ist idealerweise für jedes beliebige Zeitintervall zu gewährleisten. Daraus folgt, dass in der Tat der Nettonutzen je Zeiteinheit zu maximieren ist. Das Ende der optimalen Nutzungsdauer eines Bauwerks ist erreicht, wenn der Nettonutzen je Zeiteinheit nicht länger positiv ist. Dieses Kriterium des Verschwindens des Nettonutzens je Zeiteinheit definiert auch die zu jedem Zeitpunkt zulässige Versagensrate. Zur Kosten-Nutzen orientierten Optimierung von Instandhaltungsstrategien ist somit der Nettonutzen für die optimalen Nutzungsdauer zu maximieren, unter den Nebenbedingungen eines positiven Nettonutzens für die gesamte optimale Nutzungsdauer und eines die Instandhaltungskosten jeweils aufwiegenden Nettonutzens zwischen zwei Instandhaltungsmaßnahmen. Die Variablen der Zielfunktion sind die Anzahl der Instandhaltungsmaßnahmen und die Zeitintervalle zwischen diesen, die Nutzungsdauer, das jeweilige Mindestschadensniveau, das einer Instandsetzung bedarf, und der jeweilige Umfang beziehungsweise die jeweilige Qualität der Bauwerkszustandserfassung.

EVOLUTIONÄRE ALGORITHMEN

Da das beschriebene gemischt-ganzzahlige Optimierungsproblem nicht ohne weiteres durch kontinuierliche Hilfsprobleme angenähert werden kann, wird zur Lösung ein evolutionärer Algorithmus eingesetzt. Evolutionäre Algorithmen sind stochastische Suchverfahren, deren Suchrichtungen jedoch adaptiv angepasst werden. Der hier verwendete Algorithmus basiert auf einer natürlichen Repräsentation der Variablen. Entsprechend wird der stochastischen Operator zur Mutation als Irrfahrt modelliert und der zur Rekombination als Resampling-Verfahren mit einer an Hand eines gegebenen Stichprobenpaars parametrisierten Wahrscheinlichkeitsverteilung. Die Bewertung der einzelnen Lösungsvektoren erfolgt rangbasiert, deren Auswahl zur Rekombination und Mutation aus der Menge der möglichen Lösungsvektoren geschieht zufällig mittels einer varianzmindernden Stichprobenerhebung. Die Menge der Lösungsvektoren für den nächsten Zyklus wird mittels Plusselektion bestimmt, das heißt die Auswahl ist streng elitär.

OPTIMIERUNG VON INSTANDHALTUNGSMASSNAHMEN

Die Beschreibung der Entwicklung der Zustandswahrscheinlichkeiten des Bauwerks erfolgt mittels eines Markovschen Prozesses mit endlich vielen Zuständen, das heißt einer zeit-kontinuierlichen Markovschen Kette. Die Übergangsintensitäten für die verschiedenen Bauwerkszustände werden entweder mit Hilfe Markovscher Alterungsmodelle bestimmt oder aus den Verteilungsfunktionen der Zeit bis zum erreichen definierter Schadenszustände. Die Übergangsintensitäten sind im Allgemeinen zeitvariant. Innerhalb dieser Beschreibung können die Instandhaltungsstrategien als einfache Matrix-Vektor-Operationen dargestellt werden.

An Hand von Rechenbeispielen wird der Einfluss des zustands- oder zeitabhängigen Nutzens eines Bauwerks auf die Bestimmung optimaler Instandhaltungspläne untersucht. Es zeigt sich, dass im Falle des zustandsabhängigen Nutzens bei sonst gleichbleibenden Bedingungen die Zeitpunkte der Instandhaltung wesentlich kürzer sind als im Falle eines zustandsunabhängigen Nutzens. Hingegen hat die Zeitabhängigkeit des Nutzens auf die Instandhaltungsintervalle keinen deutlichen Einfluss. Allerdings erfolgt in letzterem Fall auch eine zeitliche Abnahme der zulässigen Versagensrate, so dass die Instandhaltungsstrategien es nicht erlauben das Bauwerk endlos zu nutzen. Ab einem gewissen Zeitpunkt ist das Bauwerk veraltet. Eine Veraltung erfolgt auch, wenn das Bauwerk zeitlich wachsenden Anforderungen ausgesetzt wird. In allen anderen Fällen hingegen erlaubt es der wiederholte Einsatz von Instandhaltungsmaßnahmen mit wirtschaftlichen Gewinn die Nutzungsdauer der Bauwerke beliebig zu verlängern.

Obige Aussagen gelten auch, wenn die Unsicherheiten in der Zustandserfassung und in der Schadensbeseitigung berücksichtigt werden. In beiden Fällen führen die Unsicherheiten jedoch zu frühzeitigen Wiederholungen der Instandhaltungsmaßnahmen. Auch zeigt sich, dass in diesen Fällen die Anforderung an die zerstörungsfreien Prüfverfahren bezüglich der Erkennung kleinster Schäden geringer sind, da wesentlich öfter Instandhaltungsmaßnahmen durchgeführt werden müssen. Dessen ungeachtet ist auch hier ein fortwährender Gebrauch des Bauwerks wirtschaftlich sinnvoll. Die Wahl des Zinsfußes hat keinen merklichen Einfluß auf die Ausgestaltung der optimalen Instandhaltungsstrategie.

SCHLUSSFOLGERUNGEN

Kosten-Nutzen-Kriterien erlauben gleichzeitig die Optimierung von Instandhaltungsstrategien und die Bestimmung zulässiger Versagensraten und optimaler Nutzungsdauern. Wie gezeigt wurde, stehen dabei Wirtschaftlichkeit, Gebrauchsfähigkeit und Sicherheit in enger Beziehung zueinander. Das entscheidende probabilistische Maß zur Beurteilung der Sicherheit von Bauwerken ist die Versagensrate. Die Nutzungsdauer der Bauwerke kann mit dem Einsatz von Instandhaltungsmaßnahmen im Prinzip beliebig verlängert werden, so lange sich nicht die Anforderungen an das Bauwerk mit der Zeit erhöhen. Optimale Instandhaltungsstrategien erfordern es dabei vor Erreichen der zulässigen Versagensrate aktiv zu werden. Ein Aufschieben von Instandhaltungsmaßnahmen führt nicht nur zu unzulässig gefährlichen Bauwerken, sondern ist auch wirtschaftlich nachteilig.

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegeben Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise entgeltlich/unentgeltlich geholfen:

1. ...
2. ...
3. ...

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich versichere ehrenwörtlich, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Weimar, 25. Oktober 2007

樋口聖子

(Shoko Higuchi)

Curriculum Vitae

Personal Data	Name:	Shoko Higuchi
	Date of Birth:	5 August 1971
	Place of Birth:	Yokohama, Japan
	Citizenship:	Japan
Education	04.78–03.84:	Public Elementary School Tokyo, Japan
	04.84–03.85:	Umegaoka Junior High School Tokyo, Japan
	04.85–03.87:	Hachioji Dairoku Junior High School Tokyo, Japan
	04.87–03.90:	Toin Gakuen High School Kanagawa, Japan
	04.90–03.94:	Tokyo University of Science Faculty of Engineering Tokyo, Japan
	03.94:	Bachelor of Engineering (Architecture)
	04.94–03.96:	University of Tokyo Graduate School of Engineering Tokyo, Japan
	03.96:	Master of Engineering (Architecture)
	Since 01.03:	Ph.D. Student Bauhaus-University Weimar Faculty of Civil Engineering Weimar, Germany
Professional Experience	04.96–03.99:	Structural Design Engineer NTT Facilities Tokyo, Japan
	04.99–12.02:	Research Engineer NTT Facilities Tokyo, Japan