

NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

Hybrid Genetic Algorithms
for
Reliability Assessment of Structural System

By
Jian Wang

A dissertation submitted to the Graduate Faculty in Engineering in Partial
fulfillment of the requirements for the degree of Doctor of Philosophy, the
City University of New York

2004

UMI Number: 3144148

Copyright 2004 by
Wang, Jian

All rights reserved.

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 3144148

Copyright 2004 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© 2004

Jian Wang

All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Engineering in Satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

May - 13 - 2004
Date

Michel Ghosn
Professor Michel Ghosn
Chair of examining Committee

5 - 13 - 2004
Date

Mumtaz K Kassir
Professor Mumtaz K. Kassir
Executive Officer

Michael Anshel
Professor Michael Anshel

Feng-Bao Lin
Professor Feng-Bao Lin

Anil Agrawal
Professor Anil Agrawal

Kolluru Subramaniam
Professor Kolluru Subramaniam

The City University of New York

Abstract

Hybrid Genetic Algorithms for Reliability Assessment of Structural Systems

By

Jian Wang

Advisor: Professor Michel Ghosn

Although the theory of structural system reliability has greatly matured over the last decades, widespread implementation of reliability methods in engineering practice has not yet taken place. One main reason for the lag between the theoretical developments and implementation is attributed to the limitations of most available reliability analytical techniques in their ability to account for one or more of these factors: 1) Accurately model the behavior of structural systems at high loads; 2) Consider different performance criteria; 3) Identify multiple equally important failure modes; 4) Account for load combinations; 5) Solve time dependent problems; and 6) Provide accurate solutions in a computationally efficient manner. To help resolve some of these perceived deficiencies, this Ph.D dissertation develops flexible yet efficient simulation-based methods that can be easily adapted for routine application when solving various types of structural reliability problems that are encountered in engineering practice.

Two hybrid Genetic Search Algorithms are developed to efficiently determine the probabilistically dominant failure modes of complex structural systems and determine their reliability index values. One of the proposed hybrid methods combines the benefits of the

Gene Expression Messy Genetic Algorithm (GEMGA) and the Shredding Genetic (SGA) operator to improve the efficiency of the search for failure modes through their linkage learning processes. The other proposed algorithm takes advantage of the pattern identification ability of Data Mining (DM) techniques to supplement the capacity of GA operators to explore new significant search domains. New data analysis schemes including an exploitation process based on the Tabu local search procedure are introduced in the algorithms to obtain accurate reliability index values and quantify the contributions of various random variables to the dominant failure modes.

The efficiency and accuracy of the proposed GA methods are verified by applying them to solve a range of benchmark reliability problems. By linking the proposed Genetic Algorithms to general-purpose finite element programs, the reliability of any structural system with any type of material behavior can be solved. This dissertation demonstrates the applicability of the proposed methods for solving realistic structural problems by performing the reliability analysis of cable stayed and suspension bridges subjected to combinations of loads and accounting for the geometric nonlinearity and the time-dependent deterioration of structural members.

Acknowledgments

I would like to express my deepest appreciation and special thanks to my advisor Professor Michel Ghosn for his continuous unselfish support, encouragement and patience throughout my studies. Thanks are due to Dr. LinZhong Deng for his help in reviewing and adapting the shredding operator that he has developed during his Ph.D. studies.

This study is partly supported by the National Science Foundation (NSF), this support is gratefully acknowledged.

I am also unconditionally grateful to my wife Lin Wang and my parents for their support and sincere encouragement during my study period in the USA.

TABLE OF CONTENTS

Abstract	
Acknowledgements	
List of Tables	
List of Figures	
Chapter 1 Introduction	1
1.1 Problem Statement	1
1.2 Review of Methods of Structural Reliability	4
1.3 Application of Genetic Algorithm in Structural Reliability Analysis	17
1.4 Research Objectives and Approach	21
Chapter 2 Linkage–Shredding Genetic Algorithm	27
2.1 Introduction	28
2.2 Structural Reliability Analysis Using Genetic Algorithms	31
2.3 Gene Expression Messy Genetic Algorithm (GEMGA)	33
2.4 Proposed Hybrid Genetic Algorithm.	39
2.5 Detailed Illustrative Example	43
2.6 Benchmark Tests for Evaluating the Performance of Proposed Hybrid GA	60
2.7 Further Test for Evaluating the Performance of Proposed Hybrid GA	65
2.8 Conclusions	67
Chapter 3 Time Dependent Structural Reliability	69
3.1 Introduction to Time-Dependent Reliability.	69
3.2 Time-Dependent Reliability Using Genetic Algorithms	73

3.3 Load Combinations	74
3.4 Illustrative Example	83
3.5 Engineering Example	88
3.6 Conclusions	90
Chapter 4 Hybrid Data Mining/Genetic Shredding Algorithm: Methodology	92
4.1 Introduction	93
4.2 Data Mining Background	96
4.3 Application of Data Mining to Reliability Analysis of Structural Systems	99
4.4 Proposed Hybrid Data Mining/Genetic Algorithm.	102
4.5 Illustrative Example	105
4.6 Tests for Evaluating the Performance of Proposed Algorithm	120
4.7 Conclusions	125
Chapter 5 Hybrid Data Mining/Genetic Shredding Algorithm: II: Applications & Sensitivity Analysis	127
5.1 Introduction	128
5.2 Analysis of A 10-bar Truss	129
5.3 Sensitivity Analysis	135
5.4 Analysis of a Simple Suspended Cable Structure	143
5.5 Analysis of Cable-Stayed Bridge Example	149
5.6 Conclusions	160
Chapter 6 Conclusions and Proposed Further Research	161
6.1 Conclusions	161
6.2 Further Research	163

Bibliography

List of Tables

Table 2.1 Input data for three-member truss example	44
Table 2.2 Reliability indices of failure modes	44
Table 2.3 Chromosomes of the first generation	45
Table 2.4 Initial linkage set of the first generation	48
Table 2.5 Clusters of linked genes	50
Table 2.6 Number of times that genes conditionally appear	51
Table 2.7 Conditional probability matrix	52
Table 2.8 Dominant failure modes	56
Table 2.9 Final search results	59
Table 2.10 Comparison of different reliability methods	64
Table 2.11. Input data for uniaxial bar example	66
Table 2.12 Comparison of different reliability methods for bar example	67
Table 3.1 Reliability indexes and direction cosines for time interval $\Delta t=15$ years	85
Table 3.2 Reliability indexes and direction cosines for time interval $\Delta t=5$ years	87
Table 3.3. Initial resistance, load and load effects	89
Table 4.1 Input data for three-member truss example	106
Table 4.2 Reliability indices of failure modes	107
Table 4.3 The first and second generations of chromosomes	110
Table 4.4 Linkage candidate c_1 items	112
Table 4.5 Linkage l_1 items	113
Table 4.6 Linkage candidate c_2 items	114

Table 4.7 Linkage l_2 items	115
Table 4.8 Final linkages after 5 generation of chromosomes	117
Table 4.9 Filtered failure modes	118
Table 4.10 Reliability indices of failure modes (after tabu local search)	119
Table 4.11 Accuracy of tabu search vs. structural evaluation times	120
Table 4.12 Comparison of different reliability methods	121
Table 5.1 Random variables of ten-bar truss structure	131
Table 5.2 Final failure equations for ten-bar truss	133
Table 5.3 Effect of linkage threshold on convergence	137
Table 5.4 Effect of linkage threshold cutoff criteria on convergence	138
Table 5.5 Sensitivity of results to pruning threshold of gene items	139
Table 5.6 Effect of population size on convergence	141
Table 5.7 Effect of reliability level on convergence	142
Table 5.8 Random variables of suspended structure	145
Table 5.9 Comparison of genetic algorithm results to available values	147
Table 5.10 Final search results for suspended structure	148
Table 5.11 Geometric properties of main girder of cable stayed bridge	151
Table 5.12 Random Variables of Cable Stayed Bridge	156

List of Figures

Figure 2.1 Search directions and their corresponding codes in 2-D	32
Figure 2.2 Flowchart of proposed algorithm	42
Figure 2.3 Three-member truss example	43
Figure 2.4. Transcription operation	46
Figure 2.5. Pre-Recombination-Expression operation	49
Figure 2.6 Recombination operation	54
Figure 2.7 Tabu local search	57
Figure 3.1 Schematic representation of load process and resistance degradation	70
Figure 3.2. Geometric representations of mean up-crossing rate	75
Figure 3.3 Illustration of combination of two load effects	81
Figure 3.4 Representation of strength degradation process	84
Figure 3.5 Simplified highway bridge model	88
Figure 4.1 Genetic coding system	100
Figure 4.2 Flowchart of proposed algorithm	103
Figure 4.3 Three-member truss example	105
Figure 4.4 Data mining process	108
Figure 5.1. Ten-bar truss example as adapted from Gorman (1979)	129
Figure 5.2. Suspended structure (pin-supported): (a) Geometry and loading as adapted from Frangopol & Imai (2000)	143
Figure 5.3 Suspended structure: (b) cross-sections: cable, hanger rope, and girder as adapted from Frangopol & Imai (2000)	144
Figure 5.4. Profile of cable stayed bridge	152

Figure 5.5. Typical cross section of main box girder	153
Figure 5.6 Support tower of cable stayed bridge	154

CHAPTER ONE

INTRODUCTION

1.1 PROBLEM STATEMENT

Current structural analysis and safety assessment methods focus on the behavior of new facilities using what are believed to be conservative estimates of the capacity of structural systems. These methods have evolved over decades of practice leading to the current generation of structural design specifications. The main concern of these specifications is to secure individual member safety although long-term member deterioration, the presence of system reserve strength, and the high uncertainties associated with predicting future loading conditions and other random factors are implicitly considered through the application of safety factors including load and resistance factors as well as system factors. However, the general dissatisfaction with the deteriorating state of the existing civil engineering infrastructure due to its exposure to harsh environmental conditions, unforeseen loads, and deferred maintenance, as well as recent interest in utilizing innovative materials with unknown long term performance, have led many researchers to explore the possibility of using new safety assessment tools based on performance-based engineering concepts. As stated by Ellingwood (2000), for performance based engineering to achieve its full potential, the following ingredients must be available:

- Computational tools to evaluate the performance of integrated systems;
- Relations that define performance in terms of structural response quantities that engineers can compute using available methods;
- Databases to measure uncertainties in the engineering parameters of interest;
- System reliability analysis procedures to provide the framework for managing these uncertainties and quantitative risk targets.

Because the theory of structural system reliability addresses many of the issues related to the ingredients listed above, the use of system reliability methods has recently gained considerable momentum. However, many of the available reliability analytical techniques have one or more limitations in their ability to:

- Accurately model the behavior of structural systems at high loads;
- Consider different performance criteria;
- Identify multiple equally important failure modes;
- Account for load combinations;
- Solve time dependent problems.

The application of simulation techniques in conjunction with general purpose finite element packages provides methods with a strong potential for resolving many of these outstanding issues [see for example the work of Maymon (1993), Borri & Speranzini (1997), Liu & Der Kiureghian (1991), Imai & Frangopol (2000), Guan & Melchers (1999), Grandhi & Lipping (1998), and many others]. In order to perform a simulation-based reliability

analysis, a systematic search algorithm that will identify the probabilistically dominant failure modes accounting for the time dependent randomness in applied loading and material properties is required. If all the failure modes (or at least the dominant ones) are known, the system reliability or at least bounds on the reliability can be obtained [Ditlevsen (1979), Thoft-Christensen, P., Murotsu, Y. (1986)]. Typical methods for identifying dominant failure modes can be broadly divided into three categories:

- Enumeration techniques;
- Plasticity-based procedures;
- Simulation-based methods.

[see for example Thoft-Christensen & Murotsu 1986; Moses 1982; Karamchandani 1987; Xiao and Mahadevan 1994; Ditlevsen (1982), Ditlevsen & Bjerager (1984, 1989), Karamchandani & Cornell (1992), Der Kiureghian (1996), Dey & Mahadevan (1998), Engelund & Rackwitz (1993), Frangopol (1985), Guan & Melchers (1999, 1993), Hohenbichler & Rackwitz (1983), Katsuki & Frangopol (1994), Rajashekhar & Ellingwood (1993), and many more].

But these techniques have various limitations. For example, it has been observed that:

- Searching for dominant failure modes of a large structure on a probabilistic basis is inefficient and requires too much computational time;
- Deterministic search strategies may miss important failure paths;

- Many methods rely on simplified representations of the structural response and may not be capable of modeling actual material behavior particularly at high loads or for defining different performance criteria.

Hence, more efficient and accurate methods need to be developed if reliability theory is to be applied on a regular basis for the practical safety evaluation of structural systems and as a tool for performance based design. The objective of this Ph.D thesis is then to develop an efficient simulation-based method for the reliability assessment of complex structural systems, which would realistically model their ultimate capacity, be implementable in practical situations, and provide accurate solutions using computationally efficient algorithms. The method should be sufficiently general to include ductile and brittle structural behavior and solve time-dependent problems. The next section describe current methods and discuss their advantages and drawbacks.

1.2 REVIEW OF METHODS OF STRUCTURAL RELIABILITY

Structural reliability deals with the uncertainties and variabilities associated with evaluating the safety of structural systems. These uncertainties that affect the loads and the structural resistances of a member may be modeled as random variables. The most important characteristic of a random variable is its probability distribution function which very often can be calculated based on its mean (or average) value and its standard deviation. The standard deviation of a random variable R with a mean or average value \bar{R} , is defined as σ_R .

A dimensionless measure of the uncertainty is the coefficient of variation (COV) that is defined as the ratio of the standard deviation divided by the mean value. For example the COV of the random variable R is defined as V_R such that:

$$V_R = \frac{\sigma_R}{R} \quad (\text{Eq.1.1})$$

In structural reliability, safety may be described as the situation where capacity (strength, resistance, fatigue life, etc.) exceeds demand (load, moment, stress ranges, etc.). Probability of failure, i.e., probability that capacity is less than applied load effects, may be formally calculated; however, its accuracy depends upon detailed data on the probability distributions of load and resistance variables. Since such data are often not available, approximate models are often used for calculation.

Let the margin of safety of a structure be defined as, Z , such that:

$$Z = R - S \quad (\text{Eq.1.2})$$

Where R is the resistance capacity, S is the total load effect. Probability of failure, P_f is the probability that the resistance R is less than or equal to the total applied load effect S or the probability that Z is less or equal to zero. This is symbolized by the equation:

$$P_f = \text{Pr} [R \leq S] \quad (\text{Eq.1.3})$$

Where P_f is the probability. If R and S follow independent normal distributions, then:

$$P_f = \Phi\left(\frac{0 - \bar{Z}}{\sigma_z}\right) = \Phi\left(-\frac{\bar{R} - \bar{S}}{\sqrt{\sigma_R^2 + \sigma_S^2}}\right) \quad (\text{Eq.1.4})$$

where Φ is the cumulative function of standard normal distribution. \bar{Z} is the mean of the safety margin and σ_z is the standard deviation of the safety margin. Thus, Equation 1.4 gives the probability that Z is less than 0. The reliability index, β , is defined such that:

$$P_f = \Phi(-\beta) \quad (\text{Eq.1.5})$$

which for the normal distribution case gives:

$$\beta = \frac{\bar{Z}}{\sigma_z} = \frac{\bar{R} - \bar{S}}{\sqrt{\sigma_R^2 + \sigma_S^2}} \quad (\text{Eq.1.6})$$

The reliability index β defined in Equations 1.5 and 1.6 provides an exact evaluation of risk (failure probability) if R and S follow normal (Gaussian) probability distributions. Although β was originally developed for normal distributions, similar calculations can be made if R and S follow other distributions.

When the failure function is nonlinear, the reliability assessment of a structural system can be evaluated through different available methods. Some of the most common methods include the First Order Reliability Method (FORM) which are also called First Order Second Moment methods known as FOSM and the Second Order Reliability Methods known as SORM. The traditional Monte Carlo Simulation (MCS) and its various variance reduction techniques are also often used in practice. A very popular simulation technique is the Adaptive Importance Sampling (AIS) method. Other methods are based on the Response Surface Method (RSM), neural network, and evolutionary algorithms. Hybrid methods that include various aspects of the abovementioned techniques are also used to improve the efficiency and accuracy. All these techniques are easily applied when the failure function of Equation 1.2 is known as a closed-form equation. For practical structural reliability applications, the failure function is not exactly known and one needs to rely on various analysis techniques such as the finite element method to evaluate the relationship between the loads and resistances and verify whether a particular structure is safe for various sample points of the pertinent random variables. Performing a nonlinear finite element analysis requires a large computational effort. Since all the reliability techniques mentioned above require a large number of structural analyses, the methods most likely to be utilized in practical applications are those that reduce the number of structural evaluations to a minimum. In the discussion below, we assume that the total computational cost is dominated by the cost of the structural analysis and all other costs such as generating the random variables and forming the necessary matrices are negligible. Below is a discussion of the most common methods used in practice along with their advantages and drawbacks.

FORM/ SORM

The objective of structural reliability methods is to calculate the probability of an occurrence of a structural failure (or safety) event. The event of interest is typically defined by an analytical function, $Z(x)$, where x is a vector of continuous random variables with continuous joint density function $f_x(x)$. The set of random variables may be correlated and/or conditionally dependent. The probability of failure is defined as $P_f(Z(x)<0)$ and can be represented by:

$$P_f = \int_{Z(x) \leq 0} f_x(x) dx \quad (\text{Eq.1.7})$$

The task of the structural reliability analysis is then to evaluate the multidimensional integral of the joint density function of x over the failure domain. We may evaluate the integral very efficiently by asymptotic first-order or second-order reliability methods (FORM/SORM). FORM and SORM map the joint distribution $f_x(x)$ into a standard normal space in which the failure function separates the overall space into a failure and safe domains. FORM and SORM would then find the location of the closest point on the failure surface to the mean value to define the “design point” or the most likely failure point. Since the reliability index is the minimum distance between the origin to the failure function, the reliability problem reduces to an optimization problem.

To represent the failure surface in the standardized normal space, we assume that a probability preserving transformation $x = T(\mathbf{u})$ exists where \mathbf{u} is an independent standard normal vector that transforms the probability integral into

$$P_f = \int_{Z(\mathbf{x}) \leq 0} f_X(\mathbf{x}) d\mathbf{x} = \int_{Z(T(\mathbf{u})) \leq 0} \varphi_U(\mathbf{u}) d\mathbf{u} \quad (\text{Eq.1.8})$$

where $\varphi_U(\mathbf{u})$ is the n -dimensional standard normal density with independent components.

An approximate transformations $\mathbf{x} = T(\mathbf{u})$ has been proposed by Rosenblatt. The probability of failure is obtained from:

$$P_f \approx \Phi(-\beta) \quad (\text{Eq.1.5'})$$

where $\beta = |\mathbf{u}^*|$ is the reliability index which gives the shortest distance between the failure point and the center of the normalized space which corresponds to the mean of the vector \mathbf{x} ., Φ is the cumulative probability function of the standard normal distribution and \mathbf{u}^* is found from

$$\mathbf{u}^* = \min |\mathbf{u}| \text{ for } \{\mathbf{u}: Z(\mathbf{u}) \leq 0\} \quad (\text{Eq.1.9})$$

The main computational task is to find the location of the \mathbf{u}^* -point (or β -point) by a suitable search algorithm. The failure function is replaced by its tangent hyperplane at \mathbf{u}^* to

facilitate the search process. This first order expansion led to designating this algorithm as the First Order Reliability Method (FORM).

A natural extension is to attempt second-order expansions of the limit state surfaces based on the presumption that a second-order expansion would lead to better estimates of the failure domain. This idea led to the development of Second Order Reliability Methods (SORM).

The main drawback of FORM/SORM is that the limit state function has to be expressed in closed form. But, for the reliability analysis of complex structures, this is not an easy thing. Additionally, there are some cases where FORM/SORM can't solve problems even when the limit state function is explicitly expressed. Rackwitz (2001) gives an example where FORM/SORM fail to converge when the limit state function is given by an equation of the form:

$$Z(\mathbf{x}) = \pm C m \sum_{i=1}^n \mathbf{x}_i \quad (\text{Eq.1.10})$$

where \mathbf{x}_i , $i=1,2,\dots,n$ are independent and exponentially distributed with the parameter λ .

The limit state function becomes highly non-linear in the U space.

$$Z(\mathbf{U}) = \pm C m \sum_{i=1}^n \ln(\Phi(-U_i)) / \lambda \quad (\text{Eq.1.11})$$

where U_i are independent and normally distributed. Although the exact probability of failure for the problem of Equation 1.11 is known to be $F_{Ga}(C; n; \lambda)$, where F_{Ga} denotes the Gamma distribution, FORM/SORM can only give a very crude approximation of β .

Importance Sampling

With few exceptions, FORM/SORM can perform well when the limit state function can be explicitly expressed. However, for many cases, the limit state function can only be implicitly expressed. In these cases, FORM/SORM cannot be directly applied and simulations techniques may be used.

A simulation calculates multiple scenarios of a model by repeatedly sampling values from the probability distributions for the random variables and using these as input. By counting the number of failures, unbiased estimates of the probability of failure are obtained. Simulation methods offer a feasible means to compute the probability of failure. In particular, Monte Carlo Simulation (MCS) is well known to be robust independent of the probability distribution types of the random variables and the dimension of the problem. Its main drawback, however, is that it is not suitable for finding small probabilities because the number of samples, and hence the number of system analyses, required to achieve a given accuracy is proportional to $1/P_f$. Essentially, finding small probabilities requires information from rare samples corresponding to failure, and on average it would require many samples before a failure occurs. Besides the computational effort that this would

require for very low probability of failures, common random number generators can also fail to produce truly independent samples.

Variance reduction techniques have been developed to improve the efficiency of the crude Monte Carlo method. In particular, the Importance Sampling technique has been used to shift the underlying distribution towards the failure region so as to gain information from rare failure events more efficiently. If an indicator function $I[Z_i(\mathbf{x}) < 0, i=1, \dots, m]$ is defined such that $I[]=1$ if \mathbf{x} is in failure domain D and zero otherwise, Equation 1.7 can be rewritten as:

$$P_f = \int \dots \int_D f_x(x) dx = \int \dots \int_D I[] f_x(x) dx \quad (\text{Eq.1.12})$$

Evidently, for low values of P_f few of the sample vectors \mathbf{x}_j will contribute to Equation 1.12 and as a result, very large sample sizes are required. If in addition the evaluation of $Z_i(\mathbf{x})$ is at all complicated the computational effort required to obtain a reasonable estimate of P_f becomes excessive. But, Equation 1.12 can be rewritten as:

$$P_f = \int \dots \int I[] \frac{f_x(v)}{h_v(v)} h_v(v) dv \quad (\text{Eq.1.13})$$

where $h_v(v)$ is the ‘‘Importance Sampling Density’’ function define as ISD as on next page, then P_f is estimated by the unbiased estimator of Equation 1.14:

$$P_f = \frac{1}{N} \sum_{j=1}^N \{I[Z_i(V_j) \leq 0] \frac{f_x(V_j)}{h_v(V_j)}\} \quad (\text{Eq.1.14})$$

where V_j is a vector of sample values taken from the importance sampling function $h_v()$.

$h_v()$ should be selected such that most information is extracted from the sample points V_j . This means that the samples should be taken in the vicinity of points of maximum likelihood of $f_x()$ and lying within $Z_i(x) < 0, i = 1, \dots, m$.

The efficiency of the method depends on the sampling density function. A “good” sampling density can be very efficient. On the other hand, a “poor” sampling density can be very inefficient and may even result in incorrect estimates of the failure probability with deceptively low estimated COV of the failure probability. Hence the success of the method relies on a prudent choice of the Importance Sampling Density (ISD), which undoubtedly requires knowledge of the system in the failure region.

Adaptive methods have been developed to focus the search on the failure points as more information is gathered. When the dimension n of the uncertain parameter space is not too large and the failure region D is relatively simple to describe, the method works well. When the dimension n is large and the complexity of the problem increases, it may be difficult to gain sufficient knowledge to improve the search focus and the adaptive method may fail to converge. Although, the Importance Sampling method can be used to identify the coordinates of the failure point, the failure modes can't be determined when the failure

functions are not known. Without a closed form expression, a finite element analysis is the only means available to determine whether failure has occurred.

Response Surface Methods

The limit state function of complex nonlinear structures can seldom be expressed in closed form and typically it needs to be evaluated implicitly through a numerical algorithm, such as a finite element method. Although the reliability analysis can be carried out using Monte Carlo Simulation (MCS), a large number of FE executions can make the computational cost prohibitively high. This is especially true for large and complex structures with high reliability (low probability of failure). In order to reduce the computational effort to an acceptable level, Response Surface Methods (RSM) have been developed. In RSM, the FE analysis is used to evaluate the actual limit at a grid of experimental points in the region where the design point is expected to be located on the actual (but implicit) limit state function. The actual limit state function is approximated by a polynomial function through a regression fit of the results at these pre-selected grid points. The resulting approximately fitted surface can then be used in a FORM or SORM algorithm to calculate the reliability index and estimate the corresponding probability of failure.

In practice, the actual limit state function $Z(\mathbf{x})$ is replaced by a polynomial. Typically a quadratic function is used without mixed terms of the form:

$$Z(\mathbf{x}) \approx a + \sum_{i=1}^n b_i x_i + \sum_{i=1}^n c_i x_i^2 \quad (\text{Eq.1.15})$$

$$(x_i = \mu_i + k_i \sigma_i)$$

Guan and Melchers (2001) showed that the estimated failure probability was very much influenced by the selected value of k_i , and the accuracy depended on the shape of the actual limit state function and the selection of the grid points that cannot be known a priori. Adaptive RSM are often used but convergence cannot be guaranteed.

Complex Structural Systems

For the reliability analysis of complex structures, the traditional approach includes three steps: Firstly, identify all possible failure modes for the structure. There are so many failure modes in complex structures with a high degree of redundancy that is impossible to identify all of them a priori. Hence, several methods, one of which is called the “truncated enumeration” method have been developed to identify the pertinent failure modes. The second step is to obtain the limit state equations for the structural failure condition. Node Replacement (Artificial Load) Technique and Incremental Load Technique are two major approaches that have been used to obtain the limit state equations. The third step is to estimate the probability of failure of the structural system. There are several approaches available for the estimation of probability of failure if the limit state equations have been obtained. Simulation methods (such as Monte Carlo Simulation Method) and upper and lower bound methods are often used. There are two more analysis approaches: Branch-and-

Bound method and β -unzipping method. Using the Branch-and-Bound method, the reliability of a structural system is evaluated by selecting the probabilistically dominant failure paths. The β -unzipping method is a method by which the reliability of structures can be estimated at a number of different levels. All the above methods require the engineer to identify the important failure modes or the dominant failure paths by structural elastic or plastic analysis, but they all have their limitations. Some of these methods may miss probabilistically important failure modes. Others are limited to plastic material and are unable to handle brittle behavior. For this reason general-purpose search methods that are applicable to any material type must be devised. Simulation techniques have been used for that purpose in conjunction with general-purpose F.E programs. However, these simulation methods such as Importance Sampling do not provide full information about the failure functions. Evolutionary type algorithms may on the other hand provide solutions to reliability problems while overcoming many of the deficiencies in current methods.

Evolutionary Algorithms

Evolutionary algorithms have been recently used for solving structural optimization problems. Nikolaos & Lagaros (2002) investigated the efficiency of various Evolutionary Algorithms (EA), such as Genetic Algorithms(GA) and Evolution Strategies(ES). Both types of algorithms imitate biological evolution in nature and combine the concept of artificial survival of the fittest with evolutionary operators to form a robust search mechanism. Both GA and ES have three characteristics that differ from other conventional

optimization algorithms: 1) In place of the usual deterministic operators, they use randomized operators: mutation, selection and recombination; 2) Instead of a single design point, they work simultaneously with a population of design points in the space of design variables; 3) They can handle, with minor modifications, continuous, discrete or mixed optimization problems (Papadrakakis 1998). The second characteristic allows for natural implementation of GA in the identification of multiple equally important failure modes in structural reliability. Modified versions of both Genetic Algorithms and Evolution Strategies have proved to be particularly promising to solve general optimization problems (Nikolaos & Lagaros. 2002). Since as mentioned earlier, reliability analysis reduces to an optimization problem consisting of a search for the minimum distance in the normalized U space, then GA seems to be a promising method for solving structural reliability problems as will be discussed in the next section.

1.3 APPLICATION OF GENETIC ALGORITHM IN STRUCTURAL RELIABILITY ANALYSIS

Recent developments in the field of artificial life provide several new computational techniques that are applicable for solving structural reliability problems involving large structural systems (Kaisheng et. al. 1996). For example, Genetic Algorithms, GA, have been developed following the principles of genetics and the natural selection process to guide a search into the regions of a domain that have the highest potential (or fitness) for improving the objective of the search (Goldberg. 1989). Shao & Murotsu (1999) and Deng. (2000) demonstrated the applicability of GA algorithms to the reliability analysis of structural systems. They showed that a genetic algorithm can be used to generate search directions in

the random variable space of the parameters related to structural loading and materials. A finite element program then performs a nonlinear structural analysis to follow the failure path in each search direction and find the corresponding system capacity. The safety level associated with the system's capacity is measured by the reliability index, β . Through evolution, GA finds the most critical combinations of loads and material properties that would produce the lowest value of the reliability index. Because of their random nature, GA can easily find alternative optima in repeated runs. Thus, the most critical structural failure paths (and modes) are identified. Although GA's ability to provide a versatile, robust, and simple technique for the reliability analysis of structural systems with multiple failure modes has been demonstrated, the shotgun search strategy that GA is based upon proved to be rather inefficient. In fact, GA requires that several searches be initiated simultaneously in several directions. Knowing that each step requires a nonlinear analysis of a complex structural system modeled by thousands of degrees of freedom, any improvements that can be made in narrowing down the search domain without missing out on important search directions are critical for future implementations of GA based reliability simulations in practical applications.

To ameliorate the GA search's efficiency, researchers have devised various forms of improved genetic algorithms depending on the particular application of interest. For example, improved GA algorithms have been used to solve various structural optimization problems including the topology of truss structures, transmission towers, and composite laminates (Goldberg & Samtani (1987) and Katsuki, et. al. (2001)). Many of these improved techniques center around reducing the dimensionality of the problem through the

introduction of linkages. The linkage process essentially weeds out less fit individuals by tightly binding together two or more closely correlated genes allowing them to travel as one unit under the action of GA operators (Harik, & Goldberg, 1996). Such linkages have been shown to overcome many of the limitations associated with GA such as building block disruption, inadequate exploration, spurious correlation, and many other perceived stumbling blocks to the widespread implementation of GA in structural applications. Similar improvements are sorely needed to help speed up the convergence of GA when solving reliability problems.

In addition to the theoretical advances made on GA, the last few years have witnessed the development of new techniques to speed up the process of decoding the human gene through the support of the National Institute of Health (NIH) and through private funding. In fact, researchers involved in the human gene decoding project realized that traditional methods for identifying genes in a genome constitute agonizingly slow processes forcing scientists to typically spend years locating and decoding a single one (Lemonick & Thompson 1999; Thompson 1999). By using a technique called shot-gunning, Dr. Venter and his team were able to find genes in a much easier and more efficient manner than traditional techniques (Lemonick & Thompson 1999). The process involves shredding a chromosome into pieces; identifying the properties of the genes in each segment; then replacing the segments back into their original location in the string. Because of the success of this gene shredding process, and the successful implementation of the linkaged GA search methods, it may seem reasonable to adopt into the GA based reliability algorithm some of the ideas developed during the human genome project and the recent theoretical developments in gene grouping

methods.

1.4 RESEARCH OBJECTIVES AND APPROACH

Based on the observations made in the previous paragraph, it seems reasonable to expect that an efficient reliability method can be developed using a modified GA that would be adapted to search for the minimum reliability index in a domain bounded by a set of failure functions associated with each of the modes that may cause the failure of the structure. To improve the efficiency of GA, the objective of this study is to introduce into GA new operators that simulate the process followed by Venter (1999) and his colleagues and take advantage of some aspects of the linkage algorithm developed by Harik & Goldberg (1996). The shredding algorithm first proposed by Deng (2000) has been found to have features closely related to those of the linkage operator as proposed by Harik and Goldberg (1996). Hence, it is natural to develop a new GA method that will combine the advantages of both the shredding and linkage methods. This dissertation will show that indeed such a combined algorithm will provide dramatic improvements in the search efficiency. Furthermore, the application of GA to solving structural reliability problems was based on executing the search in a discrete set of pre-determined directions. Such an approach, however, would only produce rough estimates of the reliability index and the corresponding probability of failure. To improve the estimates of the reliability, this thesis proposes an exploitation of the search so that the original GA-produced solution is further refined to provide accurate estimates of the global and local minima. The GA search as proposed by Shao (1999) and Deng (2000) was shown to be compatible with any F.E

analysis package so that the GA determines a direction and the F.E analysis is subsequently used to find the failure point in this particular search direction. In this manner, the GA will be able to solve any problem that is solvable by F.E for a variety of element types including ductile and brittle materials. This research will further improve the GA to solve time-dependent problems that might arise from member degradation and/or load combination problems. It is expected that additional improvements to the search method can also be made by using recently developed Data Mining (DM) techniques as part of a hybrid Genetic Algorithm.

Data Mining is a fast growing field that deals with the research and practice of detecting patterns from data. Using machine learning, statistics, and other techniques, Data Mining represents the process of detecting relations among the data. The relations are typically captured in terms of rules, similarity based subsets, and associations among the search space dimensions. Therefore, a Data Mining algorithm can also be viewed as a search for appropriate association rules. The core idea behind the introduction of a linkage learning process in GA algorithms is to detect the relations of the variables among the chromosomes. Hence, from a structural reliability point of view, it should be possible to use Data Mining techniques to detect linkages between the random variables. These linkages will contribute to efficiently identifying the variables that are most likely to control the dominant failure modes.

Since GA works with a coding of variables, not the variables themselves, the search is restricted to some discrete directions. When dealing with continuous variables, the

traditional GA can only find the relative minimum not the exact global or local minimum. Adding the number of coding of variables can only significantly increase the computation effort not the accuracy of the search. So, a new process called exploitation is introduced to resolve the problem between the accuracy and efficiency of GA algorithms and solve optimization problems involving continuous variables.

In summary, this research study will develop hybrid GA techniques for solving structural reliability problems. The proposed GA's will have two important benefits in efficiency and design, namely:

Efficiency: The linkage-learning process will play a very important role in improving the efficiency of GA. Linkage in the context of GAs represents the ability of building blocks to bind tightly together and thus travel as one under the action of the crossover operator. The application of linkage-learning has been intricately tied with defeating many of the limitations of GAs. These limitations encountered when using traditional GA techniques include: building block disruption, inadequate exploration, and spurious correlation. By reducing the dimensionality of the problem through linkage learning, the number of generations needed to reach convergence are greatly reduced. The linkage-based GAs generally include two operators known as transcription and recombination. The application of the transcription and recombination operators is to evaluate whether a new chromosome will have improved fitness through small changes in its gene structure. The shredding operator previously introduced by Deng (2000) has demonstrated that an assessment of the fitness of a chromosome can be achieved from the information contained in the fitness

matrix that evaluate the fitness of strings of genes that describe the chromosome. Thus, by combining the linkage-learning process with the shredding algorithm, large improvements in computational time efficiency can be introduced. The shredding operator is essentially an expert system that helps estimate the fitness without requiring the time consuming full-fledged structural analysis. By shredding a chromosome, the fitness of each segment can be estimated from the genes encountered in this segment and when replaced in the original chromosome these segment fitnesses are combined to provide reasonable estimates of the reliability of the failure mode represented by the chromosome. This shredding process has been found to improve the efficiency of GA significantly.

Design: In engineering practice, engineers can use the information about linkages to design systems with adequate reliability levels and minimum cost, or to choose the optimum method to modify the reliability of existing structures. Most importantly, armed with the linkage information, engineers will be able to identify which variables have most control on a system's reliability and which of these variables are strongly linked to each other. Hence, the linkage information is extremely useful for both improving the efficiency and for the practical application of the results.

Time-Dependent Reliability using Genetic Algorithms

When solving time-dependent reliability problems, the traditional approach used in the past consisted of discretizing the service life of structures into several time intervals and

taking the sum of the failure probabilities of the different time intervals as an estimate of the total probability of failure. This method simplified the time-dependent problem by assuming that the resistance and loads are all independent at different time intervals. This assumption obviously neglects the fact that the member and system resistances of a structure cannot be independent within the structure's service life. The proposed GA methods can be used to solve time dependent problems without the need to assume independence between the elements affecting the reliability of a structure over the discretized time intervals. Thus, better estimates of the probability of failure of time-dependent problems can be obtained with high efficiency.

1.5 SUMMARY

In summary, the purpose of this study is to develop an efficient simulation-based method for the reliability assessment of complex structural systems based on artificial intelligent techniques. The method should be sufficiently general to include ductile and brittle structural behavior and solve time-dependent problems. An improved genetic algorithm, GA, will be developed to include a filtration operator based on the principle of genetic elitism. Data Mining techniques will also be included to further improve the efficiency of the search. The proposed improved GA will be applied in time increments so that the time dependent random nature of the problem can be solved. The improved efficiency, robustness, and stability of the proposed simulation-based structural reliability method will be demonstrated through benchmark tests and demonstrations of its

applicability for the solution of large scale structural systems. The main contributions of this study will be:

- The application of the linkage-learning process in the solution of structural reliability problems.
- The application of a newly proposed shredding operator for estimating the reliability index to improve the efficiency of GA algorithm during the linkage learning and recombination processes.
- The GA algorithm will be extended so that it will help identify the most likely failure modes and identify the variables that contribute the most to the failure of structural systems.
- It will be shown that the proposed GA will converge faster than the standard GA without losing its effectiveness and will have comparable efficiency to that observed when using the Importance Sampling technique.
- Data Mining will be applied in the linkage learning process to improve the efficiency of this process that accounts for more than half the computational effort of GA.
- The GA procedure will be generalized to solve time dependent reliability problems including deteriorating structural systems as well as load combinations.
- The algorithm will be applicable for the reliability analysis of any structural linear or nonlinear system that can be analyzed using any available finite element analysis package.

1.6 DISSERTATION OUTLINE

This document provides a dissertation for a Ph.D. study that was conducted in the Department of Civil Engineering of the City College and the Graduate Center of CUNY. The aim of the study is to develop improved simulation routines for the reliability assessment of structural systems. The proposed routines are based on the latest advances in artificial intelligence methods including Genetic Algorithms and Data mining Techniques. To achieve the goals of the study, this dissertation is divided into the following Chapters: Chapter 1 gave the introduction and the problem statement for the dissertation. Chapter 2 develops an improved hybrid linkage-shredding algorithm for the reliability analysis of structural systems. Chapter 3 describes the application of the proposed GA to solve time-dependent problems. Chapter 4 develops an improved hybrid Data Mining-Shredding Algorithm. Chapter 5 provides sensitivity analyses and describes the application of the proposed Data Mining-Shredding Algorithm for the reliability assessment of nonlinear structural systems. Chapter 6 gives the conclusions and suggested future research.

CHAPTER TWO

LINKAGE-SHREDDING GENETIC ALGORITHM

ABSTRACT

The identification of the probabilistically dominant failure modes of complex structures and the estimation of their reliability can be reduced to the solution of an optimization problem in the standardized normal space of the variables that control the random loads applied on the structure and the random strengths of its members. Hence, the ability of Genetic Algorithms to identify local and global optima makes them especially suitable for solving structural reliability problems. To reduce the known inefficiencies of traditional Genetic algorithms, this chapter proposes the use of a hybrid genetic search algorithm that is capable of efficiently identifying the dominant failure modes of a structural system and estimating their reliability index values.

The proposed algorithm combines the benefits and the efficiency of the linkage learning process of the Gene Expression Messy Genetic Algorithm (GEMGA) to the ability of the Shredding Genetic operator to explore new significant search domains. By reducing the dimensionality of the problem through linkage learning, the number of generations needed to reach convergence is greatly reduced. To further ameliorate the GA search's efficiency, the Shredding operator is used to estimate the value of the reliability index in a given search direction by building and updating a fitness value matrix based on an evolutionary learning process. This will drastically reduce the number of structural analyses that are usually required during the reliability assessment of structural systems. Furthermore, an exploitation process is implemented during the search for the local optima

to obtain accurate reliability indexes and quantify the contributions of various variables to the structural failure modes identified during the search process. Thus, the proposed algorithm provides detailed information about a structure's failure modes that would be helpful for optimizing the design and improving the structure's safety against local and global failures. Examples are provided to demonstrate the high efficiency and accuracy of the proposed hybrid Genetic Algorithm.

Keywords: Gene Expression Messy Genetic Algorithm, Shredding operator, Tabu local search, Linkage learning, Structural failure modes, Structural Reliability, Time-dependent reliability.

2.1 INTRODUCTION

Recent developments in the field of artificial life provide several new computational techniques that are applicable for solving structural reliability problems involving large structural systems (Kaisheng et. al. 1996). For example, Genetic Algorithms, GA, have been developed following the principles of genetics and the natural selection process to guide a search into the regions of a domain that have the highest "fitness" or potential for improving the objective of the search (Goldberg, 1989). Recent studies by Shao & Murotsu (1999) and Deng et. al. (2000,2003), demonstrated the applicability of GA algorithms to the reliability analysis of structural systems. They showed that a genetic algorithm can be used to generate search directions in the random variable space of the parameters related to structural loading and member strengths. A finite element program then performs a nonlinear structural analysis to follow the failure path in each search direction and find the corresponding system

capacity. The safety level associated with the system's capacity is measured by the reliability index, β . Through evolution, GA finds the most critical combinations of loads and material properties that would produce the lowest value of the reliability index. Because of their random nature, GA can easily produce alternative optima in repeated runs. Thus, the most critical structural failure paths (and modes) are identified.

Although GA's ability to provide a versatile, robust, and simple technique for the reliability analysis of structural systems with multiple failure modes has been demonstrated, the shotgun search strategy that GA is based upon proved to be rather inefficient. In fact, GA requires that several searches be initiated simultaneously in several directions. Knowing that each step requires a nonlinear analysis of a complex structural system modeled by a large number of degrees of freedom, any improvements that can be made in narrowing down the search domain without missing out on important search directions are critical for future implementations of GA based reliability simulations in practical applications.

To improve a GA search's efficiency, researchers have devised various forms of improved genetic algorithms depending on the particular application of interest. For example, improved GA algorithms have been used to solve various structural optimization problems including the topology of truss structures, transmission towers, and composite laminates (e.g. Goldberg & Samtani (1986) and Katsuki, et. al. 2001). Many of these improved techniques center around reducing the dimensionality of the problem through the introduction of linkages.

The linkage process essentially weeds out less fit individuals by tightly binding together two or more closely correlated genes allowing them to travel as one unit under the action of GA operators (Harik, & Goldberg, 2000). Such linkages have been shown to

overcome many of the limitations associated with GA such as building block disruption, inadequate exploration, spurious correlation, and many other perceived stumbling blocks to the widespread implementation of GA in structural applications. Similar improvements are sorely needed to help speed up the convergence of GA when solving reliability problems. The shredding Genetic algorithm developed by Deng et. al. (2000,2003) goes one step in this direction. The process is based on shredding the genetic chromosomes that model a given search direction in the standardized random variable space. The application of the shredding operator was found to improve the convergence of Genetic algorithms to the global optimum reliability. In addition, the process helps further improve the efficiency by reducing the number of structural analyses required during the search process using a self-learning algorithm that would eliminate the need to perform an incremental structural analysis for each generated search direction.

This chapter builds on the operators proposed by Deng et. al. (2000,2003) and creates a hybrid method that combines the benefits of the shredding and learning operators of Deng et. al. (2000,2003) to the linkage process developed by Harik & Goldberg (1996). The object is to develop a structural reliability algorithm that is capable of efficiently determining the probabilistically dominant failure modes of a structural system. The proposed algorithm will also include an exploitation process to refine the search for the local optima and thus obtain an accurate estimate of the reliability index for each of the identified failure modes.

2.2 STRUCTURAL RELIABILITY ANALYSIS USING GENETIC ALGORITHMS

The identification of the probabilistically dominant failure modes of complex structures and the estimation of their reliability can be reduced to the solution of an optimization problem in the standardized normal space of the variables that control the random loads applied on the structure and the random strengths of its members (see for example Thoft-Christensen & Morutsu (1986)). The ability of Genetic algorithms to identify local and global optima makes them suitable for solving structural reliability problems as demonstrated by Shao & Morutsu (1999).

The approach proposed by Shao & Morutsu (1999) is based on the observation that the failure modes $g_1, g_2 \dots g_n$ of a structural system in the independent standard normal U-space can be reached by following the failure paths in directions I_i as illustrated in Figure 2.1. If the search is restricted to the paths at 45° angle intervals, the unit vectors may then be expressed by a three-digit system where each of the digits belongs to the set $(-1, 0, 1)$. Thus, each search direction may be described by a set of variables each of which can have a value equal to $-1, 0$ or 1 . By digitizing a search vector using the three-digit system proposed by Shao & Morutsu (1999), the rectangular coordinates of a direction vector would constitute a string of digits known as a chromosome in GA parlance where each digit (having the values $-1, 0$ or 1) may be labeled as a gene. Other coding schemes can be easily developed to produce narrower search sweeps since any vector component can be represented in a binary system. In structural reliability applications, the fitness of a chromosome is defined in terms of how fast it would lead to reaching a failure point. Hence, the most fit chromosomes are those associated with a low reliability index.

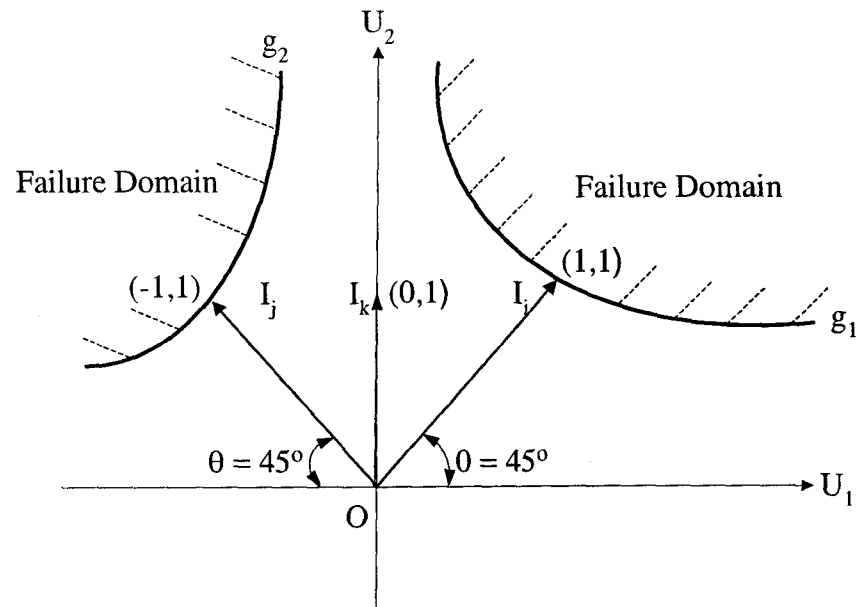


Figure 2.1 Search directions and their corresponding codes in 2-D

The classical Genetic search algorithm proposed by Shao and Murotsu (1999) has been successfully applied to solve structural reliability problems but was found to be rather inefficient. To improve the efficiency of the GA search, Deng (2000) proposed a modified genetic algorithm, referred to as the Shredding Genetic Algorithm (SGA). In summary, the process proposed by Deng (2000) consists of shredding the string of digits (the chromosome) and identifying the strong pairs of genes that have been observed to occur simultaneously in the strongest (most fit) chromosomes. In addition, Deng (2000) observed that a reasonably good approximation to the overall fitness (reliability index) of a chromosome can be obtained based on information assembled from the fitness of previously studied chromosomes having similar genes or pairs of genes. This observation was found to lead to large improvements in the efficiency of the algorithm as the actual calculation of the fitness (reliability index) would require extensive computational effort when the structural system is nonlinear and involves a large number of variables. The shredding and pairing of

the genes as developed by Deng (2000) could be extended for groups involving more than two genes. Other researchers who had previously made similar observations to those of Deng (2000) have developed a linkage learning operator encompassing many of the features included in Deng's algorithm. The linkage learning process was subsequently included in an algorithm known as Gene Expression Messy Genetic Algorithm (GEMGA), which will be discussed in the next section.

2.3 GENE EXPRESSION MESSY GENETIC ALGORITHM (GEMGA)

The Gene Expression Messy Genetic Algorithm (GEMGA) is a method developed to improve the GA search process based on a linkage identifying process. The linkage process improves the efficiency of a search by tightly binding together two or more closely correlated genes allowing them to travel as one unit. Kargupta (1998) demonstrated that linkage learning brings back the GA to the domain of linear sample complexity without sacrificing its performance. This is achieved by having the algorithm look for relations that capture similarities in the fitness landscape. In GEMGA, the problem of associating similarities based on the chromosomes' fitness values is posed as a problem of detecting approximate symmetry. Symmetry can be defined as invariance in the fitness when some variation is applied to the chromosome. The GEMGA exploits the ability of detecting relations among the search space members and uses these to guide the future search directions (Kargupta et. al. 1998,1995, 1999)

To execute the principles stated above, GEMGA relies on two primary operators (Bandyopadhyay, 1998), namely: (1) transcription, and (2) Recombination-Expression, as described below.

Transcription: The purpose of the transcription operator is to initialize the linkage list. Given a chromosome formed by a number of genes, the transcription operator applies a random change in the value of every gene one at a time. The value of the gene is flipped and the corresponding change in the fitness value is noted. Since each gene can have three possible values, the flipping is done from one value to one of the other two values randomly with a 50% chance. The objective is to spot "local" symmetry of the fitness landscape in a statistical sense. If a change in a gene value causes an improvement in the fitness of the chromosome, then the original instance of the gene certainly does not belong to the best schema. Transcription sets the corresponding "Capacity" of the gene to one indicating that the gene has the capacity to improve. On the other hand if the fitness worsens, then the original gene may belong to a good class (or at least that observation does not say it is otherwise). The corresponding "Capacity" of the gene is set to zero, indicating that the value at that gene position cannot be improved in a statistical sense. After executing the changes and determining the capacity of a gene to improve, the value of that gene is set back to its original value and the fitness of the chromosome is set back to the original fitness.

All the genes whose capacities are reduced to zeroes are collected in one list, called "the initial linkage list", which is stored as the first element of the linkage set associated with the chromosome. Thus, the transcription operator does not change anything in a chromosome. But, it evaluates the capacities of the genes to improve and initiates the formation of the linkage lists. The same process is continued deterministically for all the genes in every chromosome of the population (Kargupta, 1998).

Recombination-Expression: This operator primarily consists of two phases: the Pre-Recombination-Expression phase and the Recombination phase. The Pre-Recombination-Expression phase determines the clusters of genes encountered in the linkages of any two chromosomes of a generation. During the first generation, a pair of chromosomes is selected and one of the two is marked. Of the genes present in the initial linkage lists of both the marked chromosome and its pair, only those that have the same gene value in both the marked and its pair are extracted and grouped as a separate cluster of linked genes. The process is repeated to check all possible pairs of chromosomes existing in the first generation.

Based on the set of all linkage clusters assembled above, an $n \times n$ Conditional Probability Matrix (CPM) is formed such as n is the number of genes in one chromosome. The entry $CPM_{i,j}$ indicates the probability of the occurrence of gene i , when gene j is present in a linkage set. This matrix will be used to determine the final linkage sets using the Get-Final-Linkage operator.

For each row i of CPM, the maximum value in the row (other than i) is identified. If the maximum probability is above a threshold value, then the genes that have their probability values within ϵ of the maximum are grouped in one linkage list connected to gene i . This list will be known as L_i , with i varying between 1 to n . The list will be associated with a weight which is calculated as the average of the probabilities $CPM_{i,j}$ for all the j 's included in the list L_i .

Comparing the initial linkage list of each chromosome with L_i , the genes that appear in the initial linkage and also in L_i are extracted and grouped in a new list called the final linkage list of gene i , known as FL_{ik} where k is the number of the chromosome studied with

k varying between 1 and the number of chromosomes in one generation. Because the order in which the genes are listed in a linkage is not important, the final set of lists may appear to contain several $FL_{i,k}$ lists associated with the same gene i for the same chromosome k .

Each final linkage list will be associated with two characteristics. The first is the weight of the linkage as calculated above. The other property of the final linkage is its goodness. The goodness value indicates how good the linkage of the genes is in terms of its contribution to the fitness. For the first generation, the goodness values of all linkages are set to zero. The goodness value is changed during the recombination process if it improves the fitness of a chromosome as will be explained further below.

After selecting the set of final linkage lists for each chromosome in the first generation, the initial linkage list is deleted and never used again.

Recombination: Following the Pre-Recombination-Expression phase, the GEMGA Recombination operator is applied iteratively on randomly selected pairs of chromosomes. First, copies of each pair are made. An element (one list) of the linkage set of the first chromosome is selected based on a linear combination of its weight and goodness value (for the first generation, the goodness of all linkages are 0). The process is repeated for the second chromosome. The best of the two selected linkage lists is identified as the linkage list that will be used to execute the cross over operation. The chromosome that was associated with the identified linkage list is marked. All the genes of the marked chromosome that appear in the identified linkage list are swapped in-group with those of its pair.

After the cross-over, the fitnesses of the new chromosomes are calculated. The fitness of the new chromosome containing the identified linkage list is compared to that of the original unmarked chromosome's fitness. If the new chromosome shows an improvement in the fitness, then the goodness of the identified linkage list is improved by increasing its value by 0.5. Otherwise, its goodness is decreased by 0.5.

Finally, the fitnesses of all four chromosomes (the original two and the new chromosomes) are compared and only the two with the best fitnesses are retained in the new generation.

SUMMARY: The performance of GEMGA was compared by Kargupta and Bandyopadhyay (1998), to a large number of different optimization algorithms. The comparison executed for large problems with millions of local optima and bounded variable interactions, demonstrated the superior performance of the Algorithm.

EXPLOITATION OPERATOR

The GEMGA operator has been successfully applied to solving local and global optimization problems, and Deng (2000) has shown that the shredding algorithm reduces the required number of operations required to reach a global optimum for structural reliability problems. However, as shown in Figure 2.1, when using GA algorithms to solve reliability problems, the search is executed along previously defined angles. For example, Deng (2000) following the suggestion of Shao (1999) performed the searches along 45° angles in the normalized random space. Restricting the search to specific angles gives very rough results. One way to improve the accuracy of the reliability analysis is to decrease the angle between

two search directions. But, the total possible search directions will increase sharply rendering the search extremely inefficient. To solve the efficiency it is proposed to use a local search algorithm based on the concept of a bi-population scheme developed by Tsutsui (1997). The concept is based on a two-phased search involving an explorer sub-GA and an exploiter sub-GA to solve highly multimodal problems involving real-valued genes. The explorer sub-GA mainly performs global exploration of the search space. The exploiter sub-GA exploits the fit local areas of the search space around the neighborhood of the best-so-far solution (Tsutsui, 1997).

Unlike the approach proposed by Tsutsui (1997) that executes the exploration and the exploitation operators at every generation, which would drastically increase the computational effort, it is herein proposed to use the exploiter operator only after the search has converged to the optimum explorers. In this proposed scheme, the exploration process uses a combination of the GEMGA and Shredding operators and the exploiter process is used to give an accurate solution of the search outside the 45° angle restrictions that have been utilized during the exploration process.

In this chapter, it is proposed to use the Tabu Search algorithm to exploit the search around the previously determined approximate local optima. The Tabu Search (TS) algorithm has been extensively used and described in the literature and has been found to be one of the most appropriate tool for finding local optima (Battiti, 1994). The main concept behind TS is to avoid entrainment in cycles by forbidding or penalizing moves that take the next generation search to points in the solution space previously visited (hence "tabu"), (Glover 1986). To avoid retracing the steps used during the search for a local minimum, TS records recent moves in one or more Tabu lists. The application of TS for solving structural

reliability problems is illustrated in the example provided in this chapter. Further information about TS can be found in the literature. For example one can refer to the text by Glover and Laguna (1997).

2.4 PROPOSED HYBRID GENETIC ALGORITHM.

The high efficiency of GEMGA in detecting multiple local optima and important linkages has been proven (Kargupta and Bandyopadhyay, 1998). This is especially important when solving complex structural reliability problems because of the large number of possible failure modes that may significantly contribute to the probability of system failure. However, when large numbers of random variables dominate the problem, the number of operations in the GEMGA transcription and recombination operators would still require intensive computational effort. The Shredding Genetic Algorithm (SGA) has been found to improve the efficiency in detecting the global optimum by focusing the search around the most important genes. Another advantage of SGA is its reasonable accuracy in predicting the fitness value of a new chromosome without necessarily executing a complete structural analysis. This is done through the learning process associated with SGA. By focusing on improving the convergence toward the global optimum, SGA has been found to miss out on local optima (Deng et. al. 2000) which, as mentioned above, may be very important for the reliability analysis of structural systems.

Observing the strengths and weakness in both GEMGA and SGA, it is quickly noticed that these two methods would greatly complement each other. Hence, it is herein proposed to use a hybrid algorithm that would combine the benefits of GEMGA in detecting multiple failure modes and the ability of SGA to minimize the structural analysis effort by

predicting the fitness of new chromosomes. This chapter will demonstrate that the new hybrid algorithm will show high efficiency in detecting multiple important failure modes in a computationally efficient manner. The proposed hybrid GA can be summarized using the following steps:

1. Explorer Process

- a) Create randomly the first generation of chromosomes in the standard normal space. The chromosomes define the direction of a search for the modes of failure.
- b) Establish the fitness index matrix where fitness is the inverse of the reliability index ($1/\beta$) or the inverse of the distance between the origin and the point on the failure surface along the pre-set search direction.
- c) Execute the transcription operator to detect schemata that capture local fitness symmetry, and predict the fitness value of each new chromosome.
- d) Execute Pre-Recombination-Expression operator to determine the linkages.
- e) GEMGA recombination: use a recombination operator that combines the effects of selection and crossover. Reconstruct modified schema linkage sets and their parameters. At the same time, use the fitness index matrix or structural analysis to estimate the fitness value. If a structural analysis is performed, update the fitness matrix to improve the accuracy of future predictions of fitness values.

2. Exploiter Process

- a) Determine the important failure modes.
- b) Use Tabu Search to exploit fit local areas of the search space to improve estimates of the reliability indices and obtain approximate expressions of the failure mode equations.
- c) Determine the approximate expression of the failure mode equations and the probability of system failure.

The proposed search algorithm, as shown in the flowchart of Figure 2.2, is illustrated in the next section through a simple detailed example.

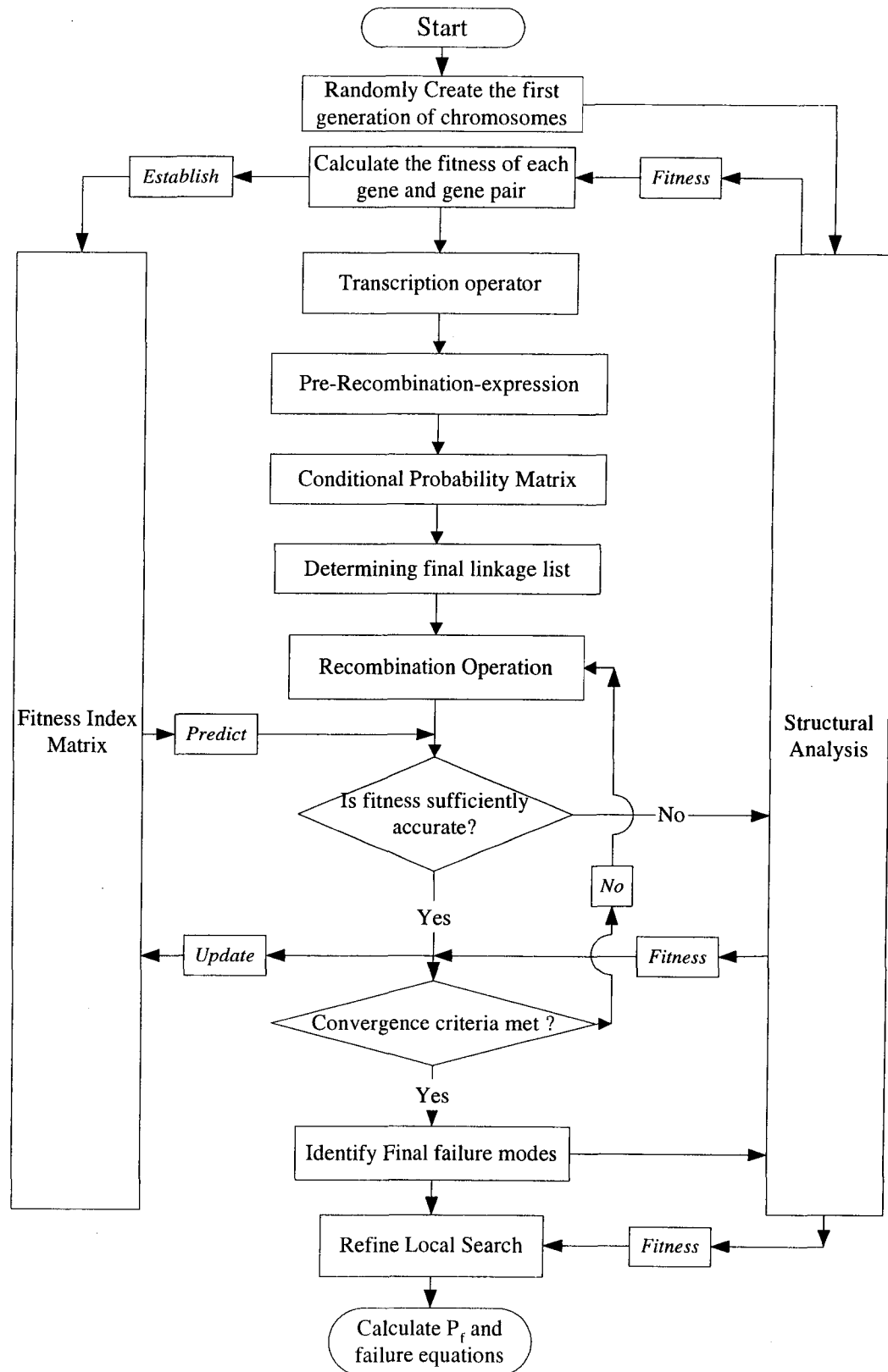


Figure 2.2 Flowchart of proposed algorithm

2.5 DETAILED ILLUSTRATIVE EXAMPLE

A simple example is presented to illustrate the proposed method. The example shown in Figure 2.3, consists of the same three-bar truss analyzed by Deng (2000).

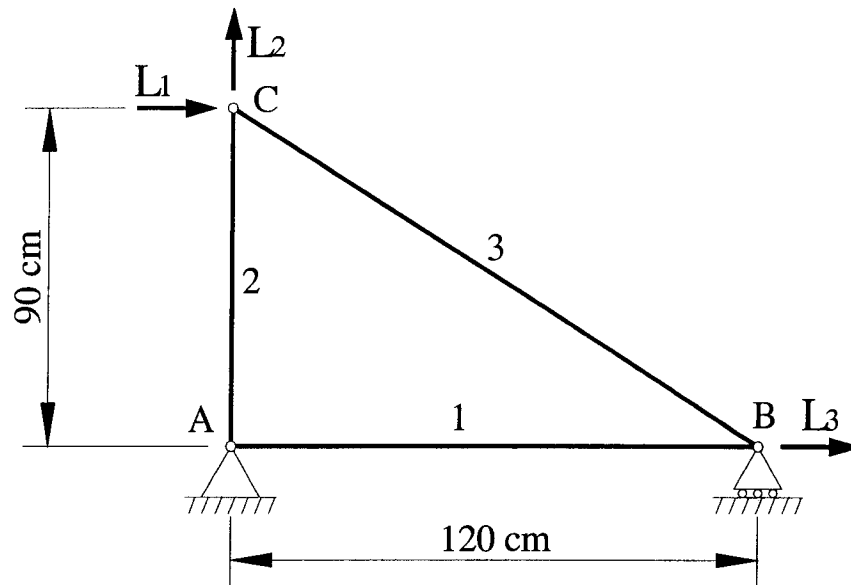


Figure 2.3 Three-member truss example.

The input data for the loads and member strengths are listed in Table 2.1. The mean and the COV of the random variables identified as the loads L_1 , L_2 and L_3 and the yielding strengths for members 1, 2 and 3 respectively identified as R_1 , R_2 and R_3 . The potential member failure modes are labeled Z_1^+ , Z_1^- , Z_2^+ , Z_2^- , Z_3^+ and Z_3^- , where the subscript indicates which member fails. The “+” sign denotes failure in tension, while the “-” sign denotes failure in compression. It is assumed that the yielding strengths for tension and compression are independent random variables with the same mean and standard deviations. The reliability indexes for all the failure modes can be calculated using FOSM and the

results are listed in Table 2.2. From the table, we observe that the first two failure modes are the dominant failure modes. Table 2.1 Input data for three-member truss example.

Table 2.1 Input data for three-member truss example

Random Variables	Mean	COV	Distribution type
L ₁	35 KN	10%	Normal
L ₂	24 KN	10%	Normal
L ₃	15 KN	10%	Normal
R ₁	63.48 KN	5%	Normal
R ₂	63.48 KN	5%	Normal
R ₃	63.48 KN	5%	Normal

Table 2.2 Reliability indices of failure modes

Mode	β
Z ₁₊	2.719
Z ₂₊	2.775
Z ₃₋	3.670
Z ₃₊	19.94
Z ₁₋	22.93
Z ₂₋	23.86

The objective of this section is to verify the ability of the proposed hybrid Genetic Algorithm to identify the most dominant modes of the above problem. In the proposed hybrid GA, the search for the failure modes is executed by following the failure paths that represent the directions $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6$, which correspond to the sequence of random variables L₁, L₂, L₃ and R₁, R₂, R₃ in that order.

A rule of thumb that was established based on experience consists of using a population size that is double to triple the number of random variables. For this example,

the population size is chosen to be 20. The first generation of chromosomes (representing possible search directions) is created randomly. An example of a set of first generation chromosomes and their reliability indices (distance from the origin to the failure point along a search direction) are listed in Table 2.3.

Table 2.3 Chromosomes of the first generation

Index	Chromosome						β
1	1	0	0	-1	0	-1	3.498
2	1	1	-1	1	0	0	5.266
3	0	-1	1	-1	1	1	6.449
4	1	1	0	-1	0	-1	4.040
5	0	1	1	1	0	1	11.025
6	0	-1	-1	1	1	-1	13.900
7	1	1	1	-1	1	-1	4.040
8	1	0	-1	1	-1	0	4.563
9	0	0	-1	1	-1	0	7.220
10	0	0	1	0	-1	0	5.895
11	1	0	1	0	0	1	4.670
12	1	1	-1	1	-1	0	3.608
13	0	1	-1	1	-1	0	4.747
14	0	0	-1	1	-1	1	8.336
15	1	1	1	0	-1	0	3.227
16	1	0	0	1	-1	1	4.563
17	0	1	1	-1	0	0	4.995
18	0	-1	-1	-1	0	0	13.947
19	0	-1	-1	-1	1	-1	13.900
20	0	0	0	1	-1	1	7.220

Transcription: The binary value of each gene of a chromosome is randomly changed and the new reliability index of the whole chromosome is evaluated.

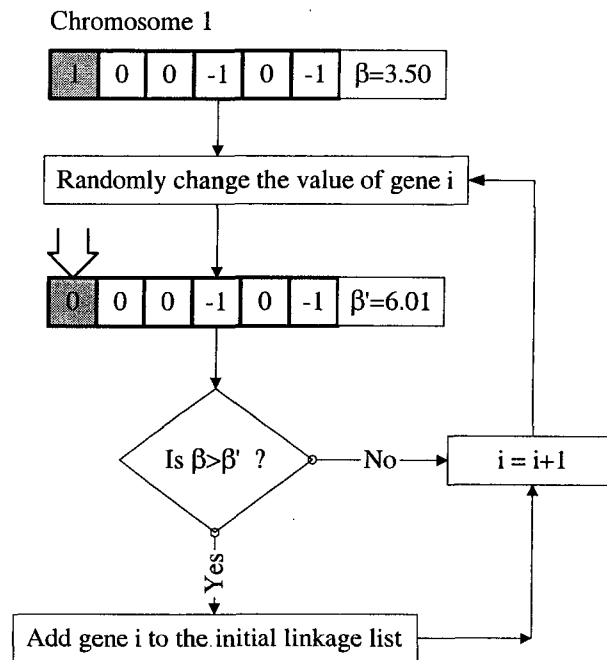


Figure 2.4 Transcription operation

For example in Figure 2.4, for the chromosome (1 0 0 -1 0 -1) with a $\beta=3.50$, the value of the first gene is 1. This number is randomly changed to 0 or -1. If the value 0 is selected, then we evaluate the reliability index of the transformed chromosome (0 0 0 -1 0 -1) to be $\beta=6.01$. Since the change caused no improvement in the fitness (new $\beta=6.01 > 3.50$), then the original gene may belong to a good class. The corresponding capacity of that first gene is set to zero because this gene is not a good candidate for change. The value of that gene is set back to its original value (i.e. 1) and the fitness of the chromosome is set

to the original fitness ($\beta=3.50$). All the genes whose capacities are reduced to zero are collected in one list called the initial linkage list. This list is stored as the first element of the linkage set associated with the chromosome. The transcription operator does not change anything in a chromosome except the capacities and initiates the formation of the linkage set. The initial linkage lists for all the chromosomes of Table 2.3 are given in Table 2.4.

Table 2.4 Initial linkage set of the first generation

Chromosome	Initial Linkage List
<i>1</i>	1,2,3,4,5
<i>2</i>	1,2,5,6
<i>3</i>	1,3,4
<i>4</i>	1,3,4
<i>5</i>	2
<i>6</i>	6
<i>7</i>	1,3,4
<i>8</i>	1,5,6
<i>9</i>	5,6
<i>10</i>	1,4,5,6
<i>11</i>	1,2,3,4,5
<i>12</i>	1,2,5,6
<i>13</i>	2,5,6
<i>14</i>	1,2,5
<i>15</i>	1,2,4,5,6
<i>16</i>	1,3,5
<i>17</i>	1,3,4,6
<i>18</i>	4,5
<i>19</i>	6
<i>20</i>	1,2,3,5

Pre-Recombination-Expression: This operator determines the clusters of genes defining the linkages within the chromosomes of the first generation. First, a pair of chromosomes is randomly selected and one of them is marked. Of the genes present in the initial linkage list

of the marked chromosome, only the genes that have the same values and capacities (0) in both the marked and unmarked chromosomes are extracted and grouped as a separate set.

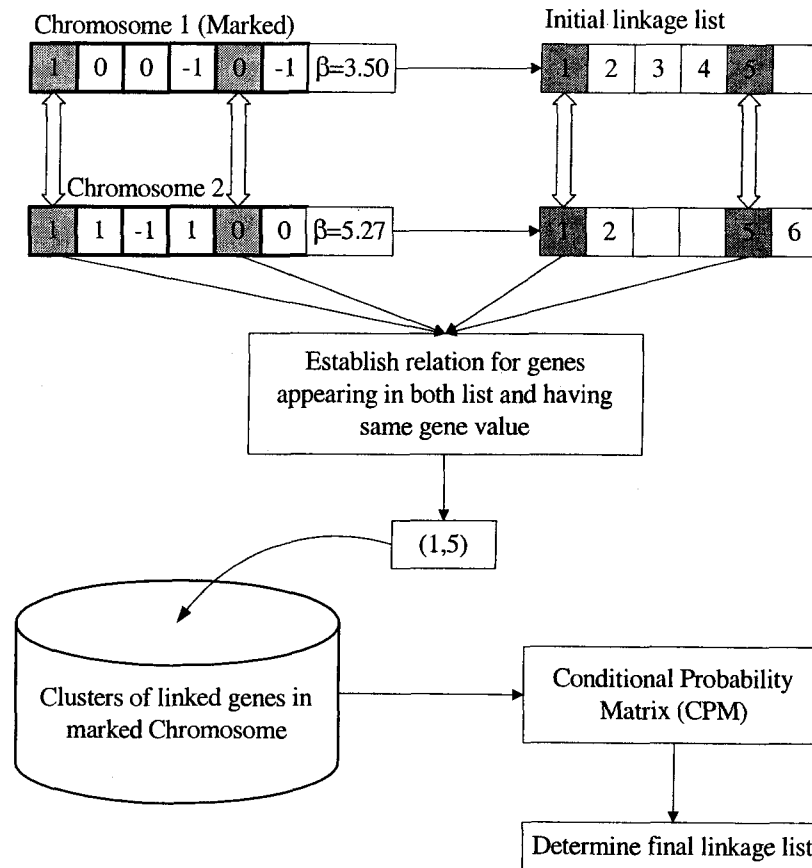


Figure 2.5. Pre-recombination-expression operation

For example, in Figure 2.5, chromosome 1 is marked. From Table 2.3 we observe that the genes have the values (1 0 0 -1 0 -1). Genes 1, 2, 3, 4 and 5 form the initial linkage list as given in Table 2.4. Chromosome 2 has the gene values (1 1 -1 1 0 0) and genes 1, 2, 5, and 6 form its initial linkage list. Comparing chromosomes 1 and 2, we find that genes 1 and 5 have the same values and capacities. Consequently we designate this group of genes (1,5) as a new cluster of linked genes. Other linked genes are obtained by comparing

chromosome 1 with all the other chromosomes of the generation. The process is repeated with all the chromosomes checked in pairs. This produces a cluster of linked genes such as the one shown in Table 2.5.

Table 2.5 Clusters of linked genes

Index	Linkage Set					
	No.1	No.2	No.3	No.4	No.5	No.6
1	1,5	1,4	1,2,5	1,3	4,5	2,3
2	1,5	1,6	1,2,6			
3	3,4	1,3,4				
4	1,4					
5						
6						
7	1,4	3,4	1,3			
8	1,6	5,6	1,5,6			
9	5,6					
10	5,6	1,5	4,5,6	1,6		
11	1,2,5	1,3	1,4			
12	1,2,6	1,3	1,4			
13	1,2,6	1,5,6	2,5,6	1,2,5,6		
14	1,5	1,2,5				
15	1,2,6	1,5,6	4,5,6	1,4	1,2,5,6	
16	1,3	1,5	3,5			
17	1,3,4	1,6				
18	4,5					
19						
20	2,3	1,5	1,2,5	3,5		

The number of gene occurrences in Tables 2.4 and 2.5 are assembled in a matrix as shown in Table 2.6. The entries in (i,j) provide the number of times that genes i and j appear together in any of the linkage clusters provided in Tables 2.5 and 2.6. At this point, the matrix is symmetric.

Table 2.6 Number of times that genes conditionally appear

Gene	1	2	3	4	5	6
1	47	16	13	15	25	18
2	16	23	5	3	16	12
3	13	5	19	9	6	1
4	15	3	9	22	9	5
5	25	16	6	9	40	20
6	18	12	1	5	20	31

By dividing each term of every row of Table 2.6 by the diagonal term of that row, a $n \times n$ conditional probability matrix is formed (Table 2.7). The entries i,j of Table 2.7 indicate the probability of the occurrence of gene i , when gene j is present in a linkage list. Notice that the matrix shown is no longer symmetric.

Table 2.7 Conditional probability matrix.

Gene	1	2	3	4	5	6
1	1.000	0.340	0.277	0.319	0.532	0.383
2	0.696	1.000	0.217	0.130	0.696	0.522
3	0.684	0.263	1.000	0.474	0.316	0.053
4	0.682	0.136	0.409	1.000	0.409	0.227
5	0.625	0.400	0.150	0.225	1.000	0.500
6	0.581	0.387	0.032	0.161	0.645	1.000

The final linkage lists of each chromosome are computed using the Get-Final-Linkage operator. The average value of all the non-diagonal terms of the conditional probability matrix is obtained as 0.381 and 80% of this average value (0.305 in the case of Table 2.7) is used as a threshold to determine which of the linkages should be kept.

The maximum value from each row of the conditional probability matrix is selected, and the genes that have their probability values within 0.30 of the maximum are included in the linkage list for gene *i*. The weight factor for this particular linkage list is set to the average value of the conditional probabilities of every gene in the list.

To get the final linkage list we will use the list of Table 2.4 along with the CPM matrix of Table 2.7. Thus, for chromosome 1, the initial linkage list will consist of genes No. 1,2,3,4, and 5. For the row $i = 1$, the maximum value in the conditional probability matrix is = 0.532. The average probability for this row for the appropriate genes is $0.367 = (0.340+0.277+0.319+0.532)/4$. The average row value 0.367 is above the first threshold 0.305 (80% of average value from the whole matrix). Hence the linkage set should be kept for gene 1 of chromosome 1 of the first generation. Also, the individual probabilities of the

entries (position indexes No.2, No.3, No.4 and No. 5) are all greater than $0.532 - 0.300 = 0.232$. Hence, the final linkage list for gene 1 of chromosome No.1 is (1,2,3,4,5) indicating that genes 1, 2, 3, 4 and 5 are strongly linked. For the case when the first test is passed but not the second, the list is kept but the genes that fail the second are dropped from the list.

Recombination: The recombination operator is the combination of crossover and selection operators. Two chromosomes are randomly selected and their genes are exchanged without disrupting, in a probability sense, the strongest linkages previously identified. Continuing with our example, assume that chromosomes 1 and 2 are randomly selected. Referring to Figure 2.6, we find that the linkage set (2,1,5) in chromosome No.1 has the maximum value of combination of goodness and weight for all the lists of chromosomes 1 and 2 (for the first generation, the initial goodness of each final linkage is always 0.0). Hence, chromosome No.1 is marked. The cross over is executed to keep the integrity of the elements found in the linkage list consisting of genes 2,1,5. The chromosomes with their initial fitness are listed as:

Chromosome No.1: gene sequence:(**1** **0** **0** -1 **0** -1) $\beta_{old}^m = 3.50$

Chromosome No.2: gene sequence:(**1** **1** -1 **1** **0** **0**) $\beta_{old}^{um} = 5.27$

Where the superscript m indicates marked chromosome, um indicate unmarked chromosome. The bolded numbers indicate that these genes belong to the “best” linkage set of these 2 chromosomes.

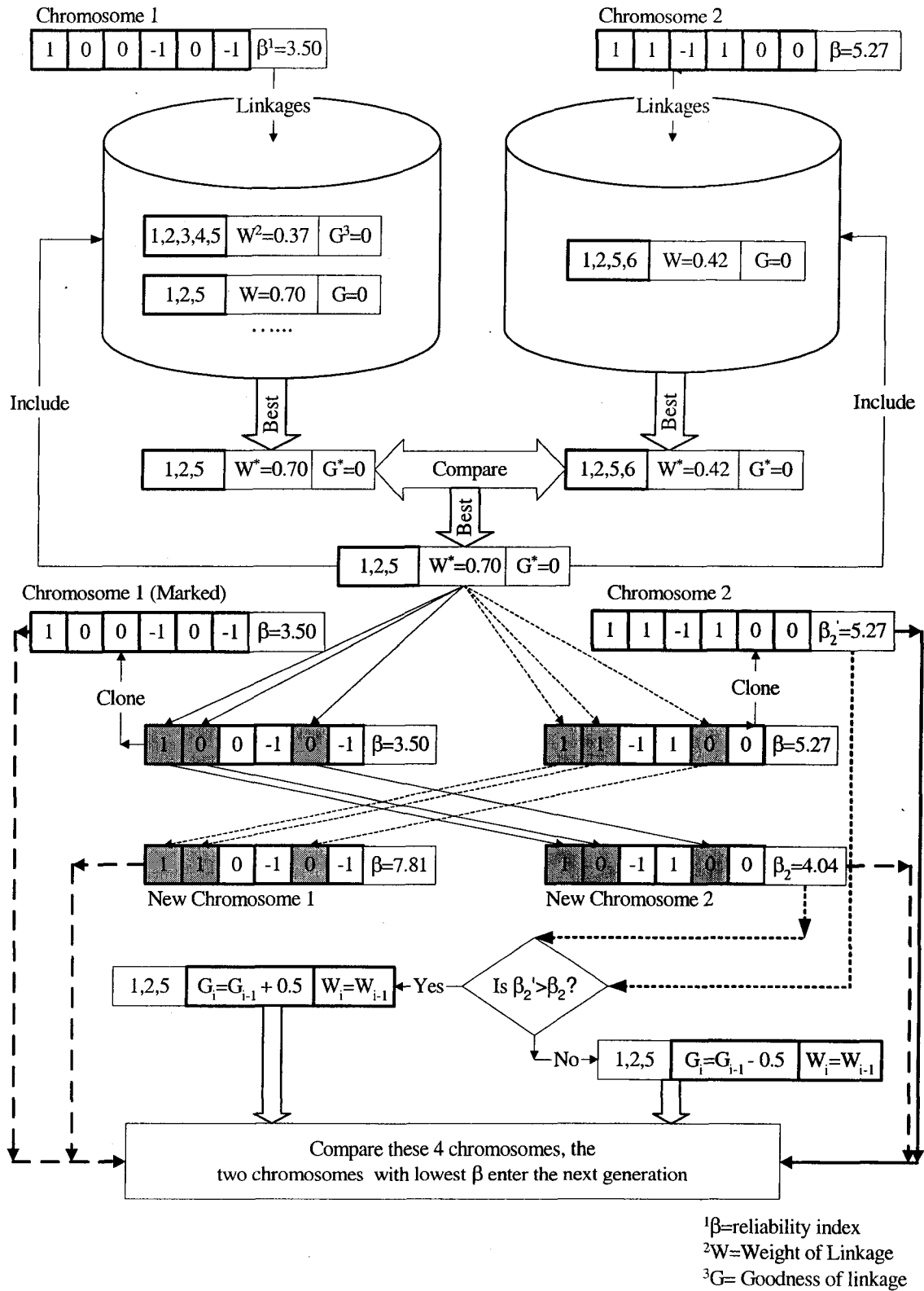


Figure 2.6 Recombination operation

The recombination process proceeds as follows:

1. Clone two chromosomes C_1 and C_2 that originally have the same features and fitness values of chromosomes 1 and 2.
2. Replace the linkage list of the unmarked chromosome with the strongest list of the marked chromosome.
3. Crossover by exchanging the values of the genes that are not linked and estimate the new fitness.

$$\text{New chromosome No.1 (1 1 0 -1 0 -1)} \Rightarrow \beta_{new}^m = 7.81,$$

$$\text{New chromosome No.2: (1 0 -1 1 0 0)} \Rightarrow \beta_{new}^{um} = 4.04$$

4. If $\beta_{new}^{um} < \beta_{old}^{um}$, that means the fitness of the unmarked chromosome has improved indicating that the selected linkage list (2,1,5) is good. The goodness value of the linkage list (2,1,5) in Chromosome C_1 and the new chromosome No.2 is increased by 0.5. Then, we include the improved cloned chromosome C_1 and new chromosome No.2 in the next generation. If $\beta_{new}^{um} > \beta_{old}^{um}$, that means that the marked linkage list is not good, so we decrease its goodness value and we keep C_2 and the marked chromosome that has the lower reliability index for the next generation. In summary, we always select for the next generation the best 2 chromosomes of the four original or transformed four chromosomes.

The explorer search process is continued until:

1. The average reliability index of the current generation does not show significant improvement over the former generation: $\beta_{average}^{(k+1)generation} > \gamma \beta_{average}^{k generation}$, γ can be set to 0.95; and
2. The first three different minimum reliability indexes of the current generation remain the same as those of previous generation.

Two dominant failure modes are identified by the proposed algorithm as shown in Table 2.8.

Table 2.8 Dominant failure modes

Failure Mode	Gene value						β
<i>1</i>	1	1	0	0	-1	0	2.795
<i>2</i>	1	0	1	-1	0	0	2.856

Exploiter Process

The purpose of the exploiter process is to refine the estimate of the reliability index after it converges within the directions set by the original GA coding. See for example Figure 2.7. This process is executed for the dominant modes of failure. The process that uses Tabu Search algorithm is executed for the truss example as described in this section.

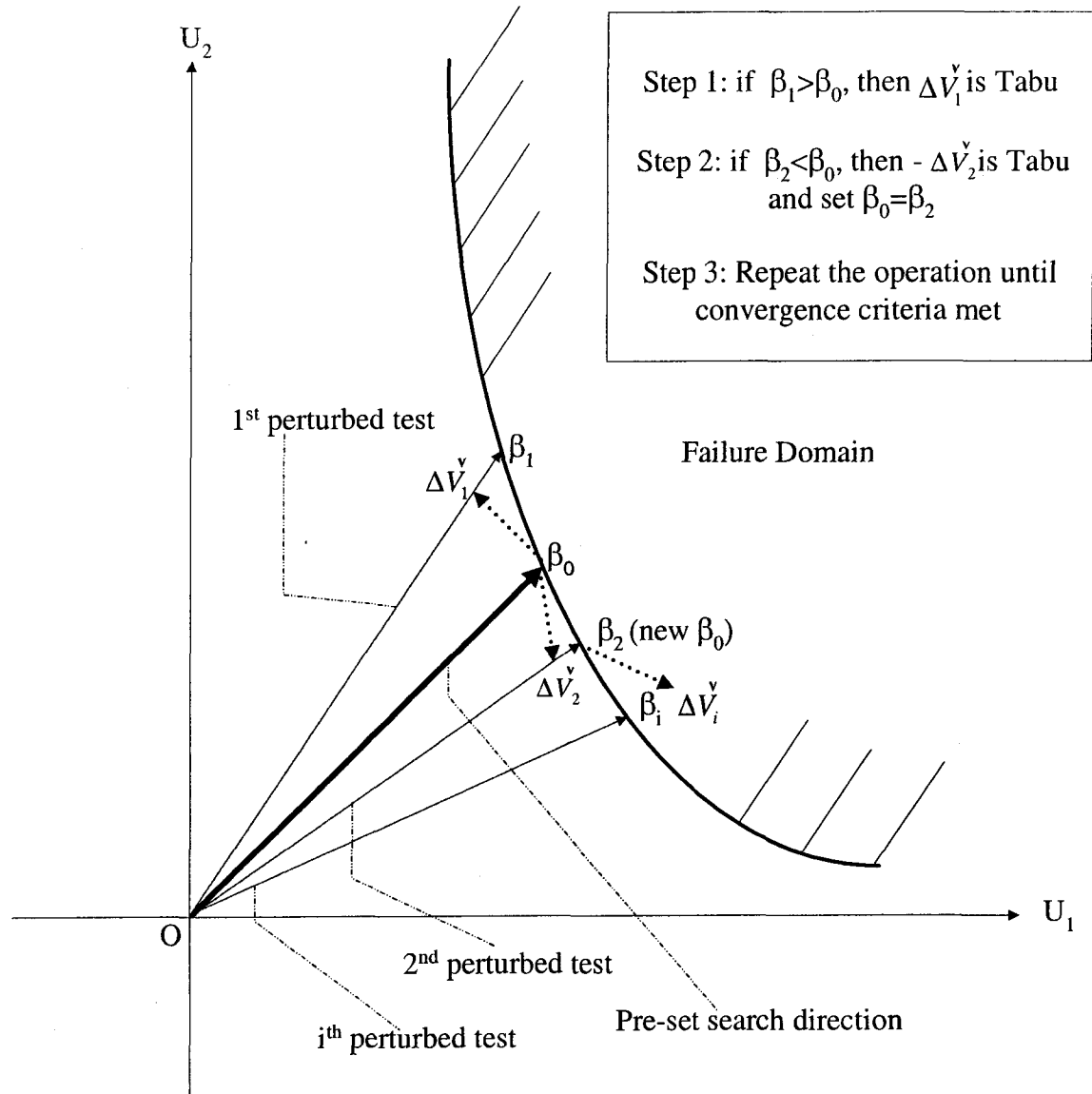


Figure 2.7 Tabu local search

The chromosomes of mode 2 are (1,0,1,-1,0,0) these indicate that the unit vector for the search direction for failure mode No. 2 is $\vec{V}_0 = (0.577, 0, 0.577, -0.577, 0, 0)$ with $\beta=2.856$. The first step consists of randomly perturbing the search direction to get a new search direction \vec{V}_1 .

$$\check{V}_1 = \check{V}_0 + \Delta\check{V}_1 \quad (\text{Eq. 2.1})$$

$$\check{V}_1 = (0.577+0.035, 0.000, 0.577+0.076, -0.577+0.056, 0.000, 0.000) \quad (\text{Eq. 2.2})$$

Normalizing, \check{V}_1 becomes $\check{V}_1 = (0.591, 0, 0.630, -0.503, 0, 0)$ with $\beta=2.892$. Since the new search doesn't result in a lower reliability index, the incremental search direction $\Delta\check{V}_1$ is not a good candidate for further search. Therefore, the normalized search direction $\Delta\check{V}_1$ is added to the list of Tabu directions. Any subsequent searches along $\Delta\check{V}_1$ should be prohibited. The same logic is repeated until the normalized local search direction $\Delta\check{V}_k$ results in an improvement in the reliability index. In this case the negative of the corresponding $-\Delta\check{V}_k$ is added to the Tabu list to prevent the search from going back.

The Tabu local search process is continued until no improvement in the reliability index is observed after a certain number of structural analysis iterations. In this case, the threshold is set to be 10 iterations. The reliability indices after the Tabu local search for the two most important failure modes are listed in Table 2.9. The difference in β between the exact values and those obtained from the proposed algorithm is less than 0.001.

Table 2.9 Final search results

Failure mode	Unit vector of design point						β
	1	0.549 (0.551)	0.509 (0.504)	0 (0)	0 (0)	-0.662 (-0.669)	
2	0.701 (0.706)	0 (0)	0.318 (0.303)	-0.638 (-0.640)	0 (0)	0 (0)	2.720 (2.719)

(Note: the value inside parenthesis is the exact solution)

Identification of failure equations: Once the Tabu search converges to the final failure modes, the approximate failure equations can be obtained based on the FORM or SORM, the two failure equations are obtained as follows,

$$Z_1 = R_1 - 0.996*L_1 - 1.055*L_3 + 0.740 \quad (\text{Eq. 2.3})$$

$$Z_2 = R_2 - 0.765*L_1 - 1.034*L_2 + 1.425 \quad (\text{Eq. 2.4})$$

Which are compared to the exact failure equations of failure modes No.1 and No.2 given as:

$$Z_1 = R_1 - L_1 - L_3 \quad (\text{Eq.2.5})$$

$$Z_2 = R_2 - 0.75*L_1 - L_2 \quad (\text{Eq. 2.6})$$

Comparing Equations 2.3, 2.4 with Equations 2.5, 2.6, it can be seen that Equation 2.3 and Equation 2.4 give very good approximations to the exact failure Equations. A more refined local search criterion associated with a large number of structural analysis iterations would lead to ever more accurate results.

Identification of probability of system failure: Once the final failure modes are determined by the Tabu local search, the correlation coefficient between each pair of modes is calculated. In this case, only two modes are obtained resulting in a correlation coefficient $\rho_{12} = 0.385$. The system's probability of failure is then obtained by:

$$\begin{aligned} P_f(Z_1 \cup Z_2) &= P_f(Z_1) + P_f(Z_2) - P_f(Z_1 \cap Z_2) \\ &= 3.26 \cdot 10^{-3} + 2.75 \cdot 10^{-3} - 1.35 \cdot 10^{-4} \\ &= 5.88 \cdot 10^{-3} \end{aligned} \quad (\text{Eq.2.7})$$

Resulting in a system reliability index, $\beta_{\text{sys}} = 2.519$

Using a crude Monte Carlo Simulation (MCS) with one million cycles, the value $\beta_{\text{sys}} = 2.515$ is obtained along with a probability of failure $P_f = 5.95 \cdot 10^{-3}$ proving the accuracy of the proposed algorithm. Other tests of accuracy are provided in the next Section.

2.6 BENCHMARK TESTS FOR EVALUATING THE PERFORMANCE OF PROPOSED HYBRID GA

Engelund and Rackwitz (1993) gave several examples that can be used as standard tests for evaluating the performance of reliability analysis methods. Originally, these benchmark tests were used to verify the applicability of alternate Importance Sampling algorithms. In this chapter, these same examples are used to verify the adequacy of the genetic algorithm to converge in an efficient manner to the exact solution.

Benchmark test No. 1: Number of variables and probability level

The failure function, Z_1 , is given as a function of the number of random variables, n , and the reliability index β by:

$$Z = \beta\sqrt{n} - \sum_{i=1}^n U_i, \quad U_i \sim N(0,1) \quad (\text{Eq.2.8})$$

where U_i designates the variable i in the standard normal space with mean 0 and standard deviation 1.

For $n=10$, using the hybrid algorithm proposed herein, convergence was reached after 255 function evaluation times. The exact $\beta = 5.0$ was obtained on the 9th generation.

The second problem uses the same equation with 50 random variables $n=50$. The hybrid GA required a total of 754 function evaluations. The reliability index obtained was $\beta = 10.646$. When the number of evaluations is increased to 858, $\beta = 10.475$ is obtained. These are deemed as acceptable solutions given the very high value of the exact reliability index $\beta=10.0$ corresponding to a probability of failure $P_f = 8.15 \times 10^{-24}$.

Benchmark test No. 2: multiple β -points

In this case the limit state function is a hyperbola having the following properties:

$$Z = X_1 X_2 - PL \quad (\text{Eq. 2.9})$$

$$P = 14.614, L=10.0 \quad (\text{Eq. 2.10})$$

$$X_1 \sim N(78064.4, 11709.7) \quad (\text{Eq. 2.11})$$

$$X_2 \sim N(0.0104, 1.56 \times 10^{-3}) \quad (\text{Eq. 2.12})$$

Using a MCS, the probability of failure is obtained as $P_f = 1.84 \times 10^{-7}$ with a reliability index $\beta = 5.18$. The Proposed Algorithm produces a reliability index $\beta = 5.21$, which for this highly nonlinear problem is only slightly different than the value obtained from the MCS.

Benchmark test No. 3: Parallel System

Four parallel failure modes are used in this example:

$$\begin{aligned}
 Z_1 &= 2.677 - U_1 - U_2 \\
 Z_2 &= 2.500 - U_2 - U_3 \\
 Z_3 &= 2.323 - U_3 - U_4 \\
 Z_4 &= 2.250 - U_4 - U_5 \\
 U_i &\sim N(0,1). \quad i=1,\dots,5
 \end{aligned}
 \tag{Eq. 2.13}$$

The correct reliability indices are

$$\begin{aligned}
 \beta_1 &= 1.893 \\
 \beta_2 &= 1.768 \\
 \beta_3 &= 1.643 \\
 \beta_4 &= 1.591
 \end{aligned}
 \tag{Eq. 2.14}$$

resulting in a system reliability index $\beta_{\text{sys}}=3.529$. The hybrid algorithm produced a $\beta_{\text{sys}}=3.527$. Convergence was achieved after 213 function evaluations.

Benchmark test No. 4: noisy limit state function

The noisy function with six random variables is given as:

$$\begin{aligned}
 Z_4 &= X_1 + 2X_2 + 2X_3 + X_4 - 5X_5 - 5X_6 + 0.001 \sum_{i=1}^6 \sin(100X_i) \\
 X_1 &= X_2 = X_3 = X_4 \sim N(120,12) \\
 X_5 &\sim N(50,15). \quad X_6 \sim N(40,12)
 \end{aligned}
 \tag{Eq. 2.15}$$

The Monte Carlo simulation produced a reliability index $\beta=2.614$. The hybrid algorithm converged to $\beta=2.615$ after 97 iterations.

Benchmark test No. 5: series system

Two failure modes in series with 6 random variables are given as:

$$\begin{aligned}
 Z_1 &= X_1 + 2X_2 + X_4 + X_5 - 5X_6 \\
 Z_2 &= X_2 + 2X_3 + X_4 - 5X_6 \\
 X_1 &= X_2 = X_3 = X_4 \sim LN(60,6) \\
 X_5 &\sim Gumbel(20,6) \quad X_6 \sim Gumbel(25,7.5)
 \end{aligned}
 \tag{Eq. 2.16}$$

250 function evaluation times were needed to converge to the following reliability index values and direction cosines values for each mode:

$$\begin{aligned}
 \beta_1 &= 2.476 \quad \alpha = (-0.050, -0.122, 0.009, -0.060, -0.013, 0.989) \\
 \beta_2 &= 2.240 \quad \alpha = (-0.008, -0.076, -0.138, -0.070, -0.040, 0.984)
 \end{aligned}
 \tag{Eq. 2.17}$$

For the series system the hybrid algorithm resulted in a final reliability index is $\beta=2.238$ whereas the “exact” value is $\beta=2.235$. Comparison of the results from different reliability methods for the five benchmark problems is provided in Table 2.10.

Table 2.10 Comparison of different reliability methods

No.	Benchmark Problem	P_f			β			Evaluation times	
		Exact	Importance Sampling	GEMGA	Exact	Importance Sampling	GEMGA	Importance Sampling	GEMGA
1	# of variables (10)	$2.92 \cdot 10^{-7}$	$2.89 \cdot 10^{-7}$	$2.92 \cdot 10^{-7}$	5.0	5.0	5.0	500	255
1	# of variables (50)		N/A		10	NA	10.65	NA	754
2	multiple β points	$18.4 \cdot 10^{-8}$	Good estimate	$9.65 \cdot 10^{-8}$	5.085	NA	5.206	500	40
3	Parallel System	$2.087 \cdot 10^{-4}$	$2.024 \cdot 10^{-4}$	$2.103 \cdot 10^{-4}$	3.529	3.537	3.527	500	213
4	Noisy limit state function	$4.47 \cdot 10^{-3}$	Good estimate	$4.46 \cdot 10^{-3}$	2.614	NA	2.615	500	97
5	Series system	$1.27 \cdot 10^{-2}$	$1.28 \cdot 10^{-2}$	$1.26 \cdot 10^{-2}$	2.235	2.232	2.238	500	250

2.7 FURTHER TEST FOR EVALUATING THE PERFORMANCE OF PROPOSED HYBRID GA

In order to compare the performance of the proposed hybrid genetic algorithm with that of gradient-based and other commonly used structural reliability methods, such as FORM, Response Surface Method (RSM), Importance Sampling Method (ISM), and different Neural Network techniques, one more example is analyzed in this paper. The selected simple example is that of a uniaxial bar with plastic behavior that is subjected to an axial load. The problem was solved by Gomes & Awruch (2004) for different techniques and a comparison between the efficiency of their solutions and the solution obtained using the hybrid genetic algorithm is compared. The failure function is given by the equation:

$$Z = Af_y - P \quad (\text{Eq.2.18})$$

where A is the cross sectional area of the bar, f_y is the yielding stress, and P is the applied load. The input data for the random variables are listed in Table 2.11.

Table 2.11. Input data for uniaxial bar example.

Random Variables	Mean	COV	Distribution type
A	0.004 m ²	12.5%	Normal
f _y	5.0 Mpa	5%	Normal
P	10,000 N	20%	Normal

Using the hybrid algorithm proposed herein, convergence was reached after a total 30 function evaluation times. The “exact” $\beta = 3.049$ was obtained. Comparison of the results from different reliability methods for this problem is provided in Table 2.12. When compared to the results of Gomes & Awruch (2004), it is observed that the hybrid genetic algorithm is as efficient as the Neural Network technique that they used and is more efficient than the Response Surface Method or the Importance Sampling technique. The FORM algorithm obviously converged after only 7 iterations in this problem where the limit state function is known. However, as explained earlier, the advantage of the FORM algorithm quickly dissipates for problems involving multiple failure modes and for problems where the limit state function is not explicitly known and the search gradients have to be numerically evaluated using finite element analysis programs. On the other hand, the ability of the proposed algorithm to solve multiple failure mode problems has been clearly demonstrated in this paper. This example helps to further demonstrate the high accuracy and efficiency of the proposed algorithm.

Table 2.12 Comparison of different reliability methods for bar example.

Technique	Reliability Index β	Probability of Failure P_f	Average Evaluation Times	Reference
Analytical solution	3.049	0.001148	--	(Ang & Tang, 1984)
Direct MCS	3.075	0.001050	100,000	Gomes & Awruch (2004)
MCS with Importance Sampling	3.050	0.001131	14,000	Gomes & Awruch (2004)
FORM (explicit limit function)	3.049	0.001147	7	Gomes & Awruch (2004)
Response Surface Method	3.049	0.001147	50	Gomes & Awruch (2004)
Neural Network	3.046	0.001141	30	Gomes & Awruch (2004)
Proposed Hybrid Algorithm	3.049	0.001148	30	This dissertation

2.8 CONCLUSIONS

This chapter utilizes a hybrid genetic algorithm to solve structural reliability problems. The algorithm, that is a modification of the genetic shredding algorithm (Deng et. al. 2003), includes linkage-learning operators to improve the convergence to the local optima as well as the global optimum. An exploiter process is added to refine the local search process and accurately determine solutions that may not necessarily lie in the directions that were pre-set during the original genetic coding.

The accuracy and efficiency of the algorithm was tested using a three-bar truss system that was described in detail. In addition, five benchmark tests previously proposed by Engelund & Rackwitz (1993) for testing the efficiency of importance sampling

methods have been used to verify the accuracy and efficiency of the proposed algorithm.

The algorithm has been found to have the following benefits:

- It is able to find the reliability indexes and identify multiple equally important failure modes as well as the variables and sub-mechanisms that contribute the most to the failure of structural systems
- It is able to refine the local search to obtain accurate estimates of the reliability indexes in directions other than those of the originally discretized search directions.
- Its efficiency is shown to be of the same order and often better than that of Importance Sampling Techniques.

CHAPTER THREE

TIME DEPENDENT STRUCTURAL RELIABILITY

ABSTRACT

This Chapter develops a hybrid genetic search algorithm for evaluating the time-dependent reliability of structural systems subjected to stochastic loads. Structural deterioration is also taken into account. This chapter offers a new outlook on time-dependent problems and an approximate method for their solution. Examples show that the proposed hybrid Genetic Algorithm can solve time-dependent and load combination problems very efficiently with high accuracy.

3.1 INTRODUCTION TO TIME-DEPENDENT RELIABILITY.

Structural loads and strengths vary in time due to the nature of the loading process and the possible degradation and environmental changes that affect structural member capacities. A general model to describe the effect of time on structural reliability can be represented as shown in Figure 3.1. In the figure, a structural system is subjected to a sequence of discrete stochastic load events. Three different load types are represented. The first load type is the permanent load represented by the dashed rectangle in the lower part of the figure. The second load type is represented by the pointed rectangles with varying intensities over different periods of time that cover the whole service life of the structure.

The third type of loads is represented by spikes that may occur randomly with random intensities during the service life.

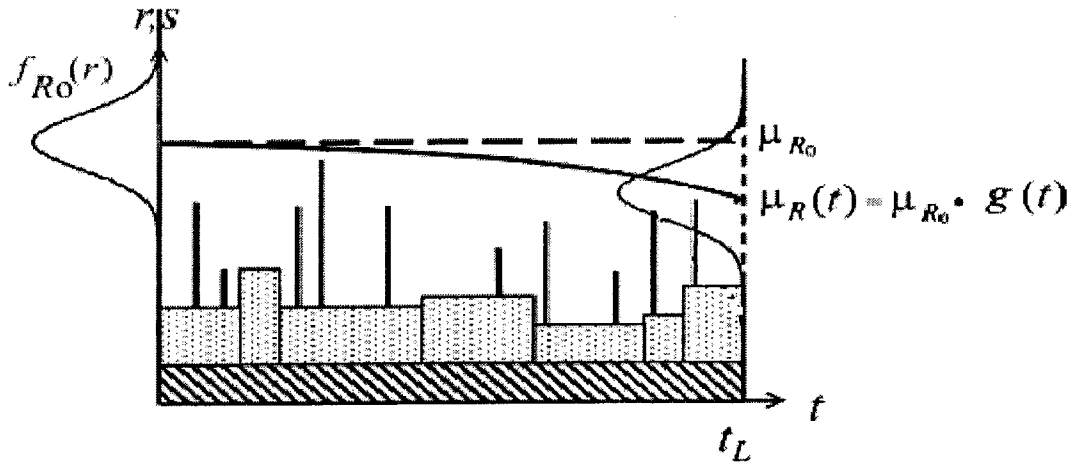


Figure.3.1 Schematic representation of load process and resistance degradation.

The strength of a structure may deteriorate with time due to environmental factors such as corrosion, sulfate, or other chemical attacks. The resistance can then be modeled as:

$$R(t) = R_0 G(t) \quad (\text{Eq.3.1})$$

in which R_0 is the capacity in the undegraded (original) state, and $G(t)$ is the time-dependent degradation function defining the fraction of the initial strength remaining at time t . Both R_0 and $G(t)$ are random. It is also often assumed that $G(t)$ is independent of the load history (i.e. fatigue problems cannot be represented using this model).

The safety margin is also a function of time that can be represented by:

$$Z(t) = R(t) - S(t) \quad (\text{Eq.3.2})$$

where $S(t)$ is the total load effect on the structural component under investigation.

To establish the probability that $Z(t)$ is less than or equal to zero within the lifetime t_L of the structure we need to solve the so-called “crossing” problem. The time, t_f , at which $Z(t)$ becomes less than zero for the first time is called the ‘time to failure’ and is a random variable. The probability that $Z(t) < 0$ within t_L is called the “first-passage” probability. In time-variant reliability one is rarely interested in the instantaneous failure probability. It is rather the distribution of the time of first passage of the time-dependent failure function $Z(t)$ into the failure domain that is of interest. But such distributions are known only for a few very special problems. For this reason, the out-crossing approach is usually used. The conditional out-crossing rate is defined by

$$v^+(\tau) = \lim_{\Delta t \rightarrow 0} \frac{P(\{Z(\tau) > 0\} \cap \{Z(\tau + \Delta t) \leq 0\})}{\Delta t} \quad (\text{Eq.3.3})$$

where $v^+(\tau)$ is the instantaneous rate at which the load exceeds the resistance.

The out-crossing rate can be defined only if the point process of crossings is a regular process, i.e. if the probability of having more than one out-crossing in a small time interval is negligibly small.

Let $N^+(t_1, t_2)$ be the number of crossings in the time interval $[t_1, t_2]$, the mean number of regular crossings $N^+(t_1, t_2)$ in the time interval $[t_1, t_2]$ can be determined from

$$E[N^+(t_1, t_2)] = \int_{t_1}^{t_2} v^+(\tau) d\tau \quad (\text{Eq.3.4})$$

It has been shown that the failure probability can be approximated as:(Cramer, Leadbetter, 1967)

$$P_f(t_1, t_2) \approx 1 - \exp[-E[N^+(t_1, t_2)]] \quad (\text{Eq.3.5})$$

Several researchers (Rackwitz, 2001, Veneziano 1977, Breitung and Rackwitz 1992) have found it computationally very difficult to determine the out-crossing rate from equation 3.3. In fact the traditional out-crossing calculation approach can only provide upper and lower reliability bounds. Furthermore, the class of random processes that can be handled is very limited. Alternatively, some researchers have chosen to estimate the outcrossing rate directly from the geometry of the failure surface.

The method proposed in this Chapter to estimate the outcrossing rate is based on the work by Kiureghian (2000) who has used the geometry of the failure function to solve random vibration problems. In a series of papers by Kiureghian and Li (1996, 2000) on the subject, problem was modeled as a discrete representation of input process in terms of a vector of standard normal random variables rather than random processes. In this case, many statistical quantities of interest in classical random vibration theory can be given geometric interpretations in the standard normal space. These include the distribution of the process and its envelope, correlation functions and statistics of crossing rates. Exact solutions can be found when the response is linear and the excitation is Gaussian. For nonlinear and non-Gaussian problems only approximate solutions can be obtained. Nevertheless, these geometric interpretations provide a new outlook to random vibration problems and potentially offer new alternatives for their approximate solution by the use of FORM and SORM. In fact, Kiureghian (2000) showed that these FORM and SORM methods that are well developed for time-invariant reliability problems could also be efficiently used to solve classical problems of random vibrations.

3.2 TIME-DEPENDENT RELIABILITY USING GENETIC ALGORITHMS

Consider an m -vector of time dependent random variables $\mathbf{x} = \mathbf{x}(t)$. These $\mathbf{x}(t)$ may represent the time dependent load or resistance variables of a structural component or system. The failure domain D is represented by $D \equiv \{X : Z(\mathbf{x}) \leq 0\}$, where $Z(\mathbf{x})$ is a limit-state function. The failure domain is crossed at time $t+\Delta t$ if the failure function is greater

than zero at time t but then becomes less than zero at time $t+\Delta t$. Hence, the crossing rate as a function of time t is can be given as:

$$v(t) = \lim_{\Delta t \rightarrow 0} \frac{P(\{Z(t) > 0\} \cap P(\{Z(X(t+\Delta t)) \leq 0\})}{\Delta t} \quad (\text{Eq.3.6})$$

In the standard normal space U , the numerator in the above equation can be seen as a parallel-system reliability problem involving the two limit-state functions:

$$\begin{aligned} Z_1 &= -Z(\mathbf{x}(t)) \leq 0 \\ Z_2 &= Z(\mathbf{x}(t+\Delta t)) \leq 0 \end{aligned} \quad (\text{Eq.3.7})$$

Approximations for the intersection of Z_1 and Z_2 can be developed by fitting the limit-state surfaces $Z_1(\mathbf{x}) = 0$ and $Z_2(\mathbf{x}) = 0$ using FORM or SORM. The event of interest then has the form of a wedge (the shaded part in Figure 3.2).

For the general limit-state function $Z = R(t) - S(t)$, we consider two realizations of Z at two different points in time t_1 and t_2 . Figure 3.2 shows the corresponding limit-state surface (hyper-planes), unit normal, reliability indices and design points. The failure probability of the structure in both time t_1 and t_2

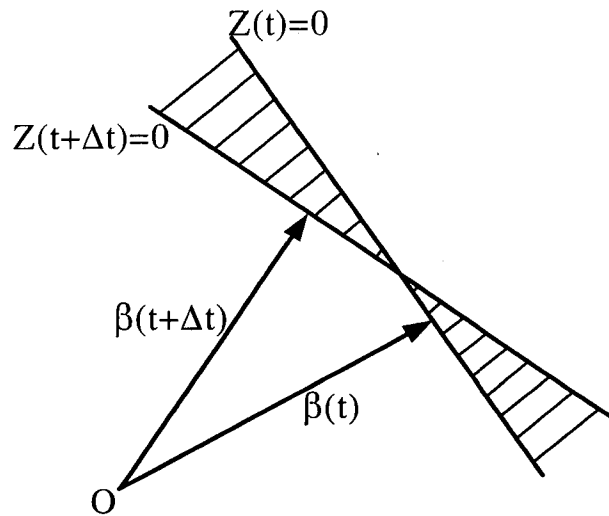


Figure 3.2. Geometric representations of mean up-crossing rate

$$\begin{aligned}
 P_f(Z(t_1) \cup Z(t_2)) &= P_f(Z(t_1)) + P_f(Z(t_2)) - P_f(Z(t_1) \cap Z(t_2)) \\
 &= \Phi(-\beta_1) + \Phi(-\beta_2) - \Phi(-\beta_1, -\beta_2, \rho)
 \end{aligned}
 \tag{Eq.3.8}$$

where $\Phi(-\beta_1, -\beta_2, \rho)$ denotes the bivariate standard normal cumulative distribution function.

β_i is the reliability index at time t_i , $\beta_i = \beta(t_i)$. The correlation coefficient

$\rho = \rho_{xx}(t_1, t_2)$ gives the correlation between the reliability at different times.

3.3 LOAD COMBINATIONS

Structural loads vary in space and in time. When second loads act on a structure, the expected maximum value of the combined effects needs to be considered to verify whether

the structure will be safe or unsafe within its intended service life. The problem can be modeled as shown in Figure 3.3. Several methods of varying degrees of accuracy and simplicity have been developed in the past to address this load combination problem. Among these methods are Turkstra's rule, Ferry-Borges model and Wen's Load Coincidence method. Below is a brief discussion on each of these methods.

Turkstra's rule

Turkstra's rule for the combination of two loads can be expressed as (Turkstra & Madsen, 1980):

$$X_{\max,T} = \max \left\{ \begin{array}{l} [\max(x_1) + \bar{x}_2] \\ [\bar{x}_1 + \max(x_2)] \end{array} \right\} \quad (\text{Eq.3.9})$$

Where $X_{\max,T}$ is the maximum value for the combined load effects in a period of time T, $\max(x_1)$ is the maximum of all possible x_1 values, $\max(x_2)$ is the maximum of all possible x_2 values, \bar{x}_1 is the mean value of x_1 and \bar{x}_2 is the mean value of x_2 .

Turkstra's rule is often used because of its simplicity. However, Turkstra's rule is generally found to provide inconsistent results and is often unconservative (Melchers, 1999). The approximation is unconservative because it ignores the possibility that the maximum combined effect may occur when neither of the load processes is at its maximum value but rather when one of the processes is at "near-maximum" value.

Wen's Load Coincidence Method

The Wen Load Coincidence model can be represented by the equation (Wen, 1981):

$$P(E,T) \approx 1 - \left\{ \exp \left\{ - \left[\sum_{i=1}^n \nu_i p_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \nu_{ij} p_{ij} + \dots \right] T \right\} \right\} \quad (\text{Eq.3.10})$$

Where $P(E,T)$ is the probability of reaching limit state E (e.g. probability of failure or the probability of exceeding a response level denoted by E) in a time period T, n is the total number of load types each designated by the subscripts i and j, ν_i is the rate of occurrence of load type i, p_i is the probability of failure given the occurrence of load type i only, λ_{ij} is the rate of occurrence of load types i and j simultaneously, p_{ij} is the probability of failure given the occurrence of load types i and j simultaneously. The process can be extended for three or more loads.

In Wen's load coincidence method, the exceedence probability of the combination is estimated based on the empirically determined incidence ratio of load combination. Wen's model is valid when the load intensities are pulse like functions of time that last for very short durations.

Ferry Borges-Castanheta Model

The Ferry-Borges model assumes that each load effect is formed by a sequence of independent load events each with an equal duration. The service life of the structure is then divided into equal intervals of time, each interval being equal to the time duration of load 1, t_1 . The probability of load 1 occurring in an arbitrary time interval can be calculated from the occurrence rate of the load. Simultaneously, the probability distribution of the intensity of load 1 given that the load has occurred can be calculated from statistical information on load intensities. The probability of load 2 occurring in the same time interval as load 1 is calculated from the rate of occurrence of load 2 and the time duration of loads 1 and 2. After calculating the probability density for load 2 given that it has occurred, the probability of the intensity of the combined loads can be easily calculated (Turkstra & Madsen, 1980; Thoft-Christensen & Baker, 1982).

The load combination problem consists of predicting the maximum value of the combined load effect X , namely $X_{\max,T}$, that is likely to occur in the lifetime of the structure, T . In the lifetime of the structure there will be n_1 independent occurrences of the combined load, X . The maximum value of the n_1 possible outcomes is represent by:

$$X_{\max,T} = \max_{n_1}[x] \quad (\text{Eq. 3.11})$$

The maximum value of x_2 that is likely to occur within a time period t_1 (i.e. when load 1 is on) is defined as $x_{2 \max, t_1}$. Since load 2 occurs a total of n_2 times within the time period t_1 , $x_{2 \max, t_1}$ is represented by:

$$x_{2 \max, t_1} = \max_{n_2}[x_2] \quad (\text{Eq.3.12})$$

$X_{\max, T}$ can then be expressed as;

$$X_{\max, T} = \max_{n_1}[x_1 + x_{2 \max, t_1}] \quad (\text{Eq.3.13})$$

Or:

$$X_{\max, T} = \max_{n_1}[x_1 + \max_{n_2}[x_2]] \quad (\text{Eq.3.14})$$

The problem reduces then to finding the maximum of n_2 occurrences of load 2, adding the effect of this maximum to the effect of load 1, then taking the maximum of n_1 occurrences of the combined effect of x_1 and the n_2 maximum of load 2. This approach assumes that x_1 and x_2 have constant intensities during the duration of one of their occurrences. Notice that x_1 or x_2 could possibly have magnitudes equal to zero. If the intensities of x_1 and x_2 are random variables with known probability distribution functions, then the probability distribution functions of the maximum of several events can be calculated using Equation 3.14.

The cumulative distribution of a single load event, Y , can be represented as $F_Y(Y^*)$. $F_Y(Y^*)$ gives the probability that the variable Y takes a value less than or equal to Y^* . Most load combination studies assume that the load intensities are independent from one occurrence to the other. In this case, the cumulative distribution of the maximum of m events that occur in a time period T can be calculated from the probability distribution of one event by:

$$F_{Y_{\max},m}(Y^*) = F_Y(Y^*)^m \quad (\text{Eq.3.15})$$

Where m is the number of times the load Y occurs in the time period T . The Ferry-Borges model is illustrated as shown in Figure 3.3 (Ghosn, Moses and Wang. 2002).

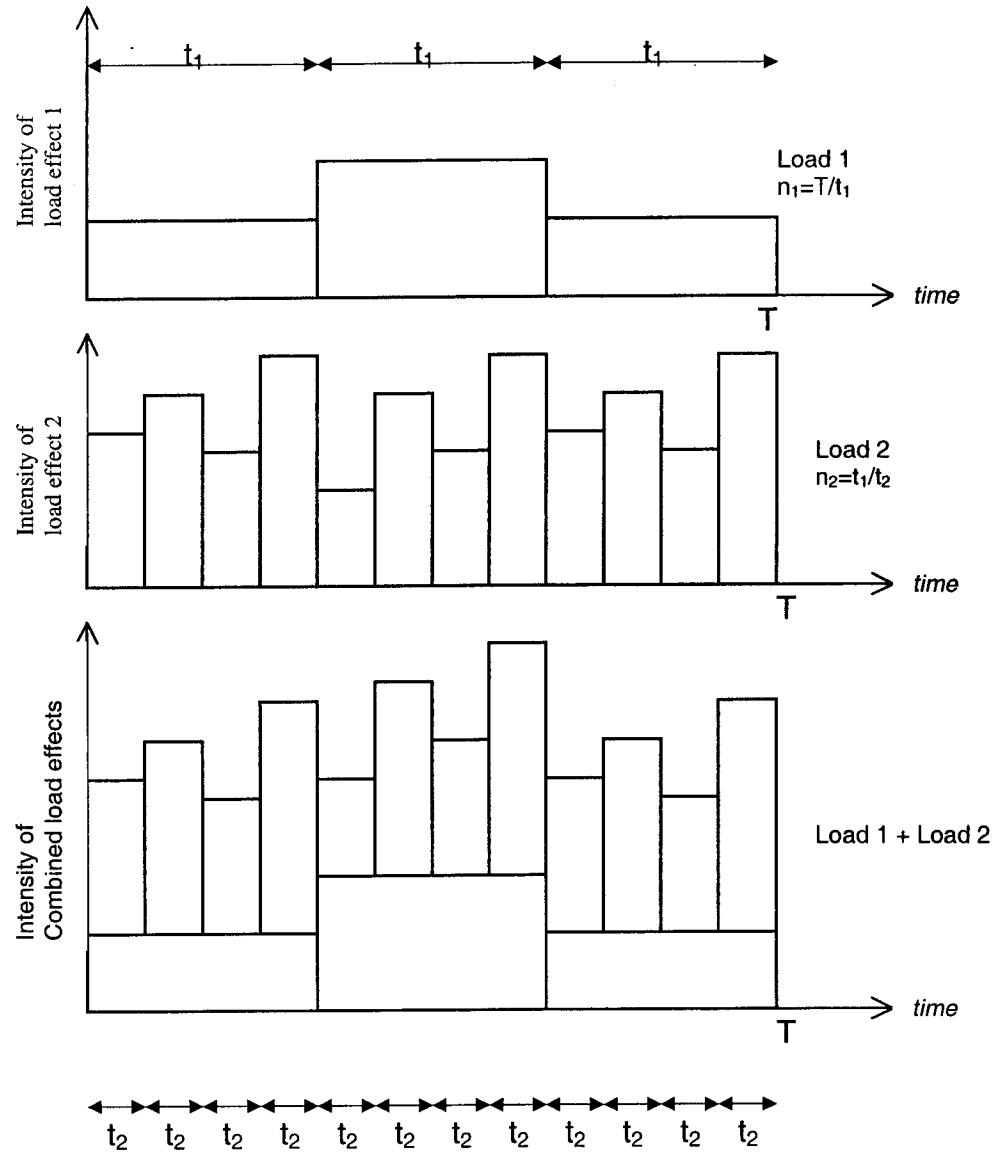


Figure 3.3 Illustration of combination of two load effects

This study uses the Ferry-Borges model as it provides a more intuitive approach to the load combination problem than the mathematical formulation of Wen's Load coincidence method. The Ferry-Borges method is directly implementable into FORM as demonstrated by Turkstra & Madsen (1980) and can be modified to account for the correlation from the modeling uncertainties using conditional probability distribution functions. Similarly, simulation-based GA can be easily applied to use the Ferry-Borges model including the consideration of correlation of load effects from different time intervals.

As shown in the Equation 3.14, the combination of two loads can be implemented in the proposed GA through the following steps:

- Let $x_3 = \max_{n_2}[x_2]$.
- Find the equivalent normal distribution function of $x_3 \sim N(\mu_{x_2}, \sigma_{x_2})$
- Let $x_4 = x_1 + x_3$.
- Find the equivalent normal distribution function of x_4
 $\sim N(\mu_{x_1} + \mu_{x_2}, \sqrt{\sigma_{x_1}^2 + (\sigma_{x_2})^2})$
- Let $x_5 = \max_{n_1}[x_4]$
- Find the equivalent normal distribution function of $x_5 \sim N(\mu_{x_4}, \sigma_{x_4})$

All approximations are made at the design point in the normalized coordinate system. Once the approximations are made, the search for the reliability index can be

executed using GA following the same steps outlined in Chapter 2. The method can be easily extended to the combination of more than two loads.

3.4 ILLUSTRATIVE EXAMPLE

To illustrate how the GEMGA-shredding algorithm can be used to solve the time-dependent reliability problem including strength degradation and load combination, a simple example is presented herein. In this example, the time-dependent failure $Z(t)$ is given in terms of the resistance $R(t)$ and two loads $L_1(t)$ and $L_2(t)$

$$Z(t) = R(t) - L_1(t) - L_2(t) = R_0 G(t) - L_1(t) - L_2(t) \quad (\text{Eq.3.16})$$

$$\begin{aligned} L_1 &: N(25, 2.5), & L_2 &: N(10, 1) \\ R_0 &= A\sigma, & A &= 2.3, & \sigma &: N(27.6, 1.38) \\ G(t) &= 1 - v\sqrt{t}, & v &: N(0.03, 0.006) \end{aligned} \quad (\text{Eq.3.17})$$

The number of repetitions for load 1 L_1 is assumed to be $1/15 \text{ yr}$. Similarly, the number of repetitions of L_2 is $1/\text{yr}$. The service life of the structure is 75 years. Hence, the limit-state function is:

$$Z(t) = A\sigma(1 - v\sqrt{t}) - \max_{x=1 \dots 5}(L_1 + \max_{x=1 \dots 15}(L_2)) \quad (\text{Eq.3.18})$$

Using a Crude Monte Carlo Simulation with a 1,000,000 simulations, the following results are obtained:

$$P_f = 5.05 \cdot 10^{-2}$$

$$\beta = -\Phi^{-1}(P_f) = 1.640 \quad (\text{Eq.3.19})$$

When using the proposed algorithm, the 75-year service life of the structure is discretized into five intervals $\Delta t=15$ years. The reliability and its direction cosine are showed as Table 3.1. This discretization process is illustrated as shown in Figure 3.4. The results for each time increment are provided in Table 3.1 including the direction cosines α_i for the minimum β in each time interval.

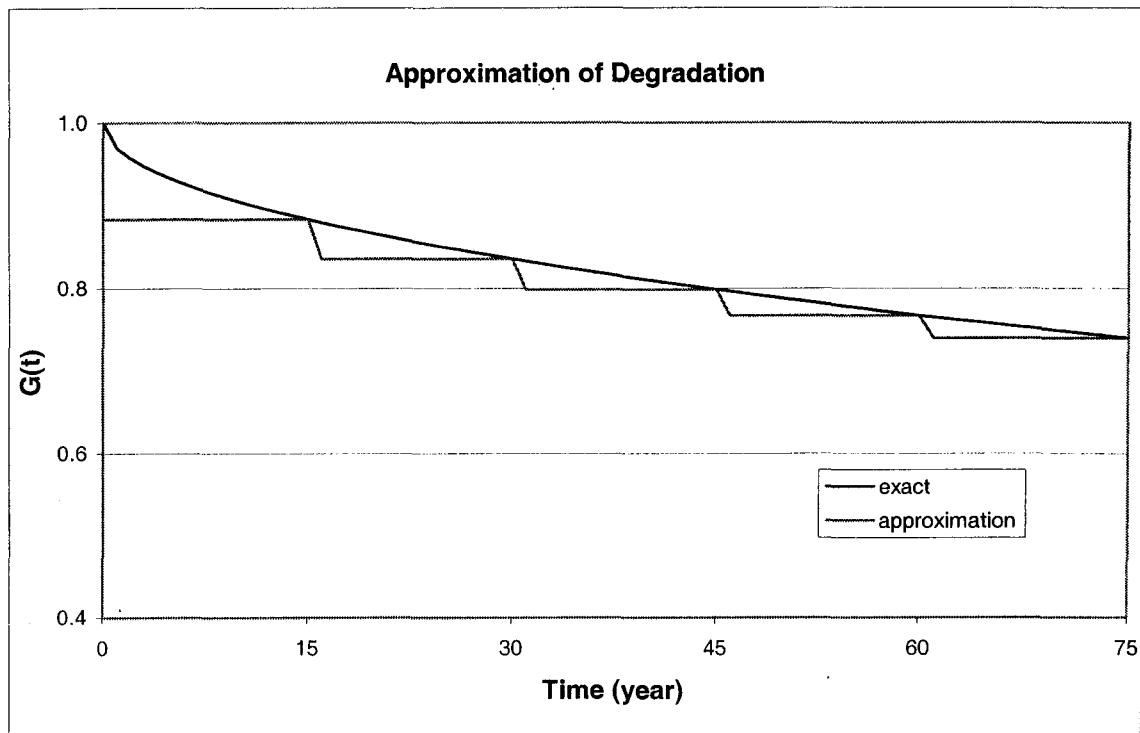


Figure 3.4. Representation of strength degradation process

Table 3.1 Reliability indexes and direction cosines for time interval $\Delta t=15$ years

	α_σ	α_v	α_{L_1}	α_{L_2}	β
1 st 15Yr	-0.730	0.310	0.576	0.196	4.523
2 nd 15Yr	-0.676	0.483	0.532	0.164	3.555
3 rd 15Yr	-0.613	0.599	0.498	0.127	2.793
4 th 15Yr	-0.583	0.648	0.472	0.131	2.168
5 th 15Yr	-0.526	0.735	0.396	0.162	1.646

We assume the event of structural failure in the first 15 years is F_1 , the second 15 years' is F_2 , etc. The probability of failure in 75 years is the union of the failure events of all five intervals and can be calculated from:

$$\begin{aligned}
 P(F_1 \cup F_2 \cup F_3 \cup F_4 \cup F_5) = & P(F_1) + P(F_2) + P(F_3) + P(F_4) + P(F_5) \\
 & - P(F_1 \cap F_2) - P(F_1 \cap F_3) - P(F_1 \cap F_4) - P(F_1 \cap F_5) \\
 & - P(F_2 \cap F_3) - P(F_2 \cap F_4) - P(F_2 \cap F_5) - P(F_3 \cap F_4) \\
 & - P(F_3 \cap F_5) - P(F_4 \cap F_5) \\
 & + P(F_1 \cap F_2 \cap F_3) + \dots
 \end{aligned} \tag{Eq.3.20}$$

Since the intersection of three events is very small, it can be ignored when we calculate the probability of failure. By plugging in the values from Table 3.1 into Equation 3.21, the final probability of failure and reliability index is obtained as:

$$\begin{aligned}
 P_f &= 5.0 \cdot 10^{-2} \\
 \beta &= -\Phi^{-1}(P_f) = 1.645
 \end{aligned}
 \tag{Eq.3.21}$$

The results when the time increment is reduced to $\Delta t = 5$ years are given in Table 3.2.

By plugging in the values from Table 3.2 into Equation 3.21, the final probability of failure and reliability index is obtained as:

$$\begin{aligned}
 P_f &= 4.895 \cdot 10^{-2} \\
 \beta &= -\Phi^{-1}(P_f) = 1.655
 \end{aligned}
 \tag{Eq.3.22}$$

Comparing the results from different time intervals, we can find that the both of two different time intervals can result in good estimation of probability of failure. Considering the efficiency of the algorithm, the bigger time interval ($\Delta t=15$ years) is more efficient. In this case, the chromosome size of each generation is 10, the generation size is 5, then the evaluation times of rough search step is 93. In the refine search step, for each time interval, evaluation times are limited to 10, so the total evaluation times is $93+10*5=143$.

Table 3.2 Reliability indexes and direction cosines for time interval $\Delta t=5$ years

	α_σ	α_v	α_{L_1}	α_{L_2}	β
1 st 5Yr	-0.781	0.150	0.580	0.178	5.543
2 nd 5Yr	-0.758	0.239	0.579	0.183	5.058
3 rd 5Yr	-0.729	0.309	0.582	0.185	4.666
4 th 5Yr	-0.706	0.377	0.567	0.191	4.321
5 th 5Yr	-0.673	0.441	0.570	0.166	4.008
6 th 5Yr	-0.653	0.461	0.560	0.163	3.721
7 th 5Yr	-0.638	0.515	0.547	0.159	3.451
8 th 5Yr	-0.620	0.555	0.536	0.140	3.200
9 th 5Yr	-0.606	0.600	0.502	0.146	2.966
10 th 5Yr	-0.586	0.617	0.502	0.150	2.747
11 th 5Yr	-0.550	0.651	0.502	0.148	2.539
12 th 5Yr	-0.540	0.661	0.501	0.143	2.346
13 th 5Yr	-0.529	0.668	0.504	0.138	2.165
14 th 5Yr	-0.525	0.707	0.459	0.117	1.992
15 th 5Yr	-0.511	0.713	0.469	0.103	1.831

3.5 ENGINEERING EXAMPLE

Consider the reinforced concrete T-beam highway bridge adopted by Enright and Frangopol (1998). The bridge consists of three 30ft simply supported spans . each span has five girders equally spaced 8.25 ft. more details about the bridge can be found in the reference of Enright and Frangopol (1998). Simplified mode is applied as shown in Figure 3.5.

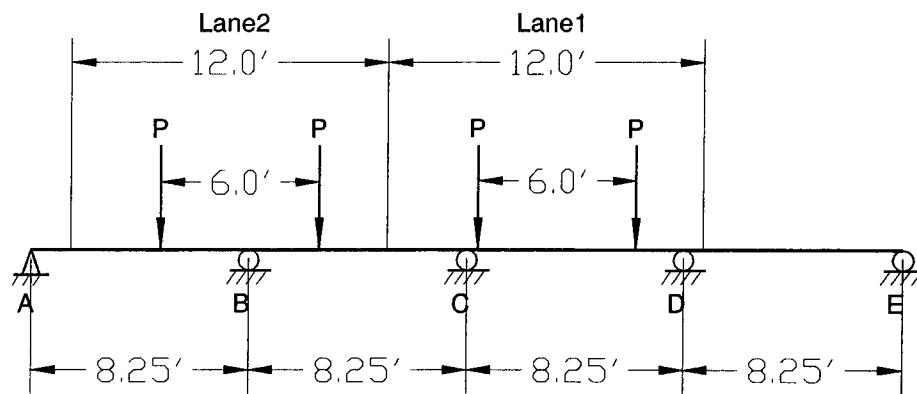


Figure 3.5. Simplified highway bridge model

The nominal value, bias, COV and distribution type of the initial resistance and load effects are indicated in Table 3.3.

Table 3.3. Initial Resistance, load and load effects

Variable	Nominal value	Bias	COV	Distribution Type
Initial Bending Resistance (kip-ft)	1086.75	1.12	0.12	Lognormal
Initial Shear Resistance (kip)	136.82	1.28	0.10	Lognormal
Dead Moment (kip-ft)	237.81	1.05	0.10	Normal
Dead Shear (kip)	63.40	1.05	0.10	Normal
Live load (one Truck) for Moment (kip)	72	0.90	0.19	Normal
Live load (one Truck) for Shear (kip)	72	0.90	0.23	Normal

The failure of bridge is defined as any one of girder failure due to moment or shear. Following the advise of Enright and Frangopol (1998). Two modified resistance degradation cases are considered as follows

$$\text{For moment: } g(t) = 1 - \frac{0.005}{3}t, \quad E(T_1 = 5.0 \text{ yrs}) \quad (\text{Eq.3.23})$$

$$\text{For Shear: } g(t) = 1 - \frac{0.0075}{3}t, \quad E(T_1 = 4.0 \text{ yrs}) \quad (\text{Eq.3.24})$$

The “exact” value of system probability of failure and reliability index based on one million crude Monte Carlo Simulation are:

$$P_f = 4.611 \times 10^{-2} \quad (\text{Eq.3.25})$$

$$\beta = 1.684 \quad (\text{Eq.3.26})$$

The 75-year service life of the highway bridge is discretized into five intervals $\Delta t=15$ years. From the reliability analysis, the shear failure mode is the critical failure mode which

dominate the probability of system failure. The system probability of failure and reliability index are:

$$P_f = 4.515 \times 10^{-2} \quad (\text{Eq.3.27})$$

$$\beta = 1.694 \quad (\text{Eq.3.28})$$

If the 75-year service life is discretized into fifteen intervals $\Delta t = 5$ years. More accurate result obtains as follows:

$$P_f = 4.622 \times 10^{-2} \quad (\text{Eq.3.29})$$

$$\beta = 1.683 \quad (\text{Eq.3.30})$$

Comparing Equation 3.27, 3.28 with Equation 3.29, 3.30, the proposed GA algorithm provided accurate result.

3.6 CONCLUSIONS

In this chapter a method to analyze the reliability of time dependent problem has been presented by application of the algorithm proposed in Chapter 2 to discretize time intervals. The method is applicable to solving deteriorating structural problem under the effects of combinations of loads. Major features of the method include:

1. Through discretizing the service life of the structure into several time intervals, the failure probability of time-dependent problem can be estimated with high accuracy. While traditional approximation methods depend on the assumption that the failure events at different time intervals are independent.
2. Using the Ferry-Borges model, the load combination problem can be easily solved by the proposed GA.
3. The illustrative example shows that the proposed GA is very efficient in dealing with the time-dependent problem.

CHAPTER FOUR

HYBRID DATA MINING/GENETIC SHREDDING ALGORITHM

PART I: METHODOLOGY

ABSTRACT

Recent studies have successfully introduced Genetic Algorithms (GA) to identify the important failure modes of complex structures and quantify their contributions to the reduction of the reliability of structural systems. In these studies, the efficiency of traditional GA techniques was substantially improved by incorporating linkage-learning operators that were used to explore relations among the random variables controlling the safety of a structural system. However, the currently used linkage learning methods, such as the Shredding Genetic Algorithm (SGA) and Gene Expression Messy Genetic Algorithm (GEMGA) were found to be either too complex for easy implementation in routine reliability analyses or too narrowly focused on finding the global optimal solution and thus reduced the capacity of GA to identify local optima. Other techniques especially suitable for exploring relations and linkages among data are known as Data Mining (DM) algorithms. Several types of DM algorithms are available in the Computer Science literature. For example, one of the most popular and successful DM tools is the Apriori algorithm. However, when used on their own DM algorithms may fail in some instances when the original database is skewed, biased, or too narrow. Hence, the purpose of this chapter is to propose a hybrid algorithm that would combine the benefits of the pattern identification ability of Data Mining techniques to the capacity of GA operators to explore new significant

search domains. The implementation of the shredding operator and its ability to reduce the computational effort through its self-learning process will further lead to the development of an efficient and robust reliability algorithm. This chapter will demonstrate that the proposed algorithm will significantly reduce the computational effort associated with determining the probabilistically dominant failure modes of structural systems without missing important failure modes. Furthermore, a new failure mode pruning criterion and a refined Tabu local search algorithm are also implemented to improve the efficiency and accuracy of the reliability analysis. Examples are provided in this chapter and the accompanying chapter to demonstrate the high efficiency and accuracy of the proposed hybrid Data Mining-Genetic Shredding Algorithm for failure mode exploration, identification, and exploitation.

4.1 INTRODUCTION

The identification of the probabilistically dominant failure modes of a complex structure and the estimation of its system reliability can be reduced to the solution of an optimization problem in the standardized normal space of the variables that control the random loads applied on the structure and the random strengths of its members (see for example Thoft-Christensen & Murotsu (1986)). Hence, many optimization algorithms can potentially be applied to solve structural reliability problems. For example, various types of modified Genetic Algorithms (GA) have been recently applied by Shao & Murotsu (1999), Deng et. al. (2000), and Wang & Ghosn (2003) for the reliability analysis of structural systems. These studies showed that a genetic algorithm can be used to generate search directions in the random variable space composed of the parameters that control structural loading and member strengths. A finite element program is used to follow the failure path in

each search direction and find the corresponding structural capacity. The safety level associated with the structural capacity is measured by the reliability index, β . Through evolution, GA finds the most critical combinations of loads and material properties that would produce the lowest value of the reliability index. Because of their random nature, GA can easily produce alternative optima in repeated runs. Thus, the most critical structural failure paths and modes are identified.

Through initiating multiple searches simultaneously in different directions, GA provides a robust technique for the reliability analysis of structural systems with multiple failure modes. However, the shotgun search strategy that GA is based upon proved to be rather inefficient for solving large-scale problems because each search direction requires a nonlinear analysis of a complex structural system modeled by a large number of degrees of freedom. To improve a GA search's efficiency, many techniques were proposed. These are centered on reducing the dimensionality of the problem through the introduction of linkages. The GA linkage learning process essentially weeds out less fit individuals by tightly binding together two or more closely correlated genes allowing them to travel as one unit under the action of GA operators (Harik & Goldberg; 2000).

Recent research by Deng et. al. (2003) proposed the use of a GA shredding operator which has been shown to help further improve the search's efficiency by reducing the number of structural analyses required during the search process. The most important feature of the shredding operator is its use of a self-learning algorithm that would eliminate the need to perform a structural analysis for each generated search direction. Based on these

observations, Wang and Ghosn (2003) used a hybrid Gene Expression Messy Genetic Algorithm (GEMGA) and demonstrated that combining the linkage learning process with the shredding operator will greatly improve the performance of Genetic Algorithms and their ability to solve structural reliability problems. However, when large numbers of random variables dominate the problem, the number of operations in GEMGA would still require intensive computational effort. Furthermore, investigations of the modified GEMGA (Wang & Ghosn; 2003) have shown that the maximum number of failure modes that can be identified cannot exceed the number of random variables. For example, for a ten-bar truss with one degree of redundancy, having 45 possible rigid-plastic collapse modes, the hybrid GEMGA can only identify at most 10 of the total failure modes.

Since the main idea behind the improved GA algorithm is to introduce the process of linkage exploration and identification, then any improvement in the ability of the algorithm to explore and identify linkages among the variables will be helpful to further speedup the search process. Data Mining is a tool that has been developed in the field of computer science to detect relations among data sets. Since, the problem of determining the failure modes of a structural system also involves the detection of patterns from large sets of data, the principles of Data Mining should also be applicable for determining the important failure modes of a structure accounting for the uncertainties in load intensities and structural response. This chapter explores the possibility of utilizing Data Mining techniques during the linkage learning process when a GEMGA algorithm is used to solve structural reliability problems. The chapter will demonstrate that the application of Data Mining will help improve the efficacy of the search for multiple failure modes of structural systems and help

overcome many of the deficiencies of GEMGA including the limits on the number of failure modes that can be identified.

4.2 DATA MINING BACKGROUND

Data Mining is a fast growing field in Computer Science that deals with the research and practice of detecting patterns from data. In practice, the field has been regularly applied to the decision support problem faced by most large retail organizations like supermarkets. Supermarkets collect and store massive amounts of sales data referred to as baskets. These are carefully analyzed to produce important pieces of the marketing infrastructure. Frequent item-sets play an essential role in many Data Mining routines that try to find interesting patterns from databases. These important patterns may include association rules, correlations, sequences, classifiers and clusters. The original motivation for searching for association rules came from the need to analyze so called supermarket transaction data. In the supermarket problem, association rules describe how often items are purchased together. For example, an association rule “diaper -> beer (80%)” states that four out of five customers that bought diapers also bought beer. Such rules can be useful for decisions concerning product pricing, promotions, store layout and many others. In a similar manner, the association rules for structural systems would involve the identification of the elements that fail in a given mode under the pertinent applied loading conditions. Thus, the Data Mining tools that have been found to be so valuable in detecting association rules in all sorts of applications promise to be as useful in structural reliability applications.

The main problem with association rule inductions is that there are immense

numbers of possible rules, which would require an efficient algorithm to prune them and identify the most valuable ones (Borgelt and Kruse 2002). For example, for the product range of a supermarket, which may consist of several thousand different products, there are billions of possible association rules. In a similar manner, a complex structural system is composed by thousands of elements that may interact in billions of ways to form possible failure mechanisms and collapse modes. It is obvious that such a vast amount of rules cannot be processed by inspecting each one in turn. Therefore, efficient algorithms are needed that restrict the search space and check only a subset of all the rules, if possible, without missing the important ones. One such algorithm that has been successfully applied is the well-known Apriori algorithm originally developed by Agrawal and Srikant (1994).

The Apriori algorithm is founded on the observation that if any given set of attributes, S , is not adequately supported, any superset containing S will not be adequately supported either. Consequently any effort to calculate the support for such a superset is wasted. The principle of Apriori Algorithm is that “Every subset of a frequent item set has to be at least as frequent as the set itself” (Agrawal and Srikant 1994). For example, if we know that $\{A, B\}$ is not adequately supported, it follows that $\{A, B, C\}$ and $\{A, B, D, \text{etc.}\}$ will not be adequately supported either. The Apriori algorithm is a powerful candidate-set-pruning technique, and it reduces the list of important candidates dramatically in a few operations. For the supermarket problem, the Apriori algorithm works as follows:

1. Assume that there are n supermarket baskets. Each of the baskets is filled by a different number of items. However, the number of times an item occurs in a

single basket is not important. Hence, the database consists of the number of baskets, n , and the list of the items that could appear.

2. Define a linkage set to be the set of items that have close bonding. i.e. that are very likely to appear together in a basket.
3. Define k as the number of items in a linkage set.
4. Set $k=1$. In other words, start by identifying the individual (single) items that appear in the database.
5. Define a set C_k that will contain the list of all the possible groups of k items that appear together in any basket of the database along with the number of baskets in which each group is encountered.
6. Define L_k to be a pruned subset of C_k . L_k will include a list of the group of k items whose count number in step 5 exceeds a pre-set threshold value.
7. Set $k = k+1$;
8. Form C_k only from the elements that are encountered in L_{k-1} ;
9. Repeat steps 5-8 until C_k is empty;
10. The complete sets L_1 through L_k provide the list of frequent sets of elements.

As discussed below this algorithm can be easily adapted to finding linkages between random variables during a structural reliability analysis.

4.3 APPLICATION OF DATA MINING TO RELIABILITY ANALYSIS OF STRUCTURAL SYSTEMS

Following the general principle of the Apriori algorithm, an appropriate equivalent principle for the structural reliability problem can be stated as follows: “if a partial failure

mechanism has a high probability of occurrence, then any subset of elements (sub-mechanism) from that partial mechanism must also have a high probability of occurrence". In this context, a partial mechanism is formed by a combination of several elements that contribute to the partial or full failure of a structure.

To solve the structural reliability problem, we use the genetic algorithm formulation described by Shao & Morutsu (1999) and Deng et. al. (2000 & 2003) where each gene represents a random variable and each gene can take the values 1, 0 or -1 indicating that the direction cosine of the search vector under investigation lies along the positive direction, is invariant, or lies along the negative direction of the random variable, respectively. In this context, the Linkage Candidate set, C_1 , will consist of the total number of genes in a chromosome times 2; 2 being the number of possible nonzero values that a single gene can take (i.e. 1 or -1). Remember that each gene represents a random variable and the number of genes is equal to the number of independent random variables. In our nomenclature, the combination of gene value and position will be referred to as a gene item. Figure 4.1 gives a representation of the search process along with gene coding.

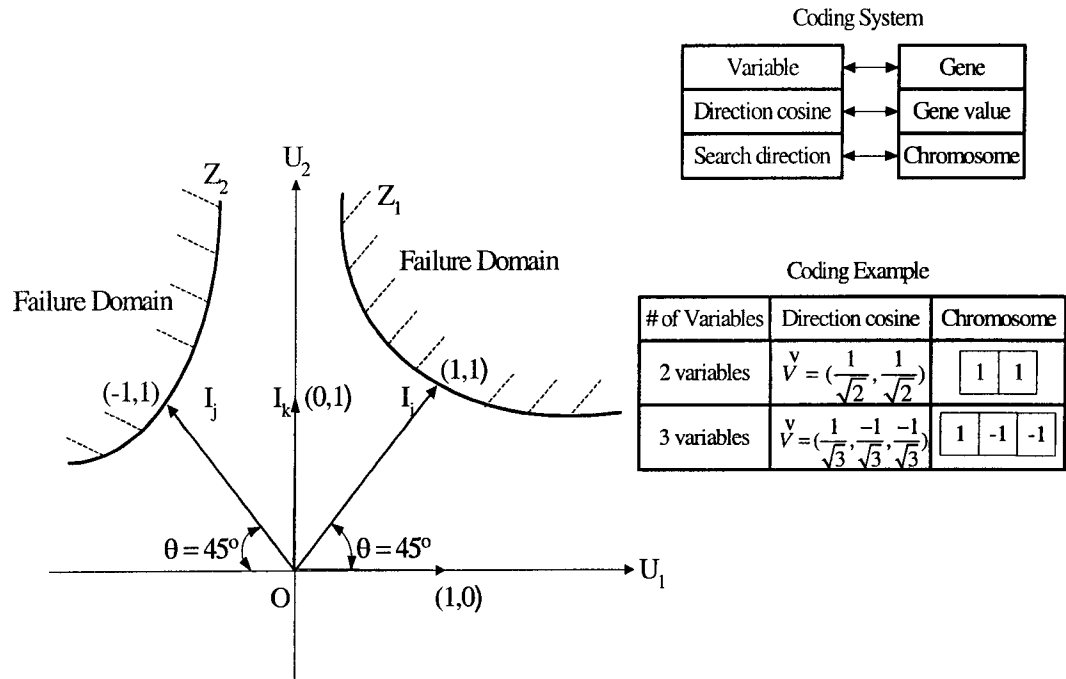


Figure 4.1 Genetic coding system

Unlike the supermarket problem where every item has equal weight during the counting process, it is herein proposed to put weighting factors for each parameter as a function of its contribution to the failure probability. Specifically, the fitness function defined by Shao & Morutsu (1999) and Deng et. al (2000 & 2003) as $1/\beta^2$ is used as the weighting function. Hence, when an item, j , (formed by k genes) is encountered in n_j chromosomes, its count in L_k is obtained by:

$$F_{kj} = \frac{\frac{1}{\beta_1^2} + \frac{1}{\beta_2^2} + \dots + \frac{1}{\beta_i^2} + \dots + \frac{1}{\beta_{n_j}^2}}{n_j} \quad (\text{Eq. 4.1})$$

Where, F_{kj} , is the weighted count number of item j in L_k , β_i is the reliability index of chromosome i and i varies between 1 and n_j . n_j being the number of times an item j is encountered within a group of chromosomes.

In the proposed approach, the exploitation part of the search can be further refined using a Tabu Search (TS) algorithm to focus the search on the exact failure point rather than simply follow the pre-determined (-1, 0, 1) directions. Additionally, a failure mode enumeration filtering procedure is used to remove repeated modes from the count as described below.

The Genetic Algorithm often results in a listing of chromosomes (search directions) having similar genes except for minor differences in one or two genes. A thorough investigation of these cases shows that the similarities between these search directions are indicating that the true search direction is somewhere between the directions indicated by the by the pre-set directions. To filter out such cases and avoid repeated local searches, a failure-mode filtering algorithm is used following the structural reliability principle mentioned earlier as follows: If adding one or more variables into a search direction results in a significant decrease in the reliability index, then, the old direction is not close to the direction of the local minimum. Alternatively, if adding one or more variables into a search direction results in a significantly higher reliability index, then, the old search direction is close to the local minimum while the new one is not. The directions that are identified as being close to local minima are subsequently used in an exploitation algorithm to determine the actual value of the minimum.

4.4 PROPOSED HYBRID DATA MINING/GENETIC ALGORITHM.

The Apriori algorithm is the core idea of frequent item-set mining, and is the subject of much current research in the field of Data Mining. However, the Apriori algorithm may fail in some instances when the original database is skewed, biased, or too narrow. In such cases, the Apriori algorithm may fail because it can only “detect and prune” linkages in a given database but is not able to “generate” new search directions. The ability of exploring new important search directions of the Shredding Genetic Algorithm (SGA) has been proven (Deng et. al. 2000 & 2003)). Observing the strengths and weakness in both Data Mining and SGA it is quickly noticed that these two methods would greatly complement each other. By combining the shredding algorithm with the Apriori search it will be demonstrated that it is possible to overcome the deficiencies of the Apriori algorithm by enriching the database with new important information. The examples given in this chapter and the accompanying chapter will demonstrate that the new hybrid algorithm will show high efficiency in detecting multiple important failure modes in a computationally efficient manner. Furthermore, the exploitation part of the search can be further refined using a Tabu algorithm to focus the search on the exact failure point rather than simply follow the pre-determined directions. Additionally, a failure mode enumeration filtering procedure is used to remove repeated modes. The proposed hybrid data-mining/genetic algorithm can be summarized using the following steps, which are also shown in the flow chart of Figure 4.2.

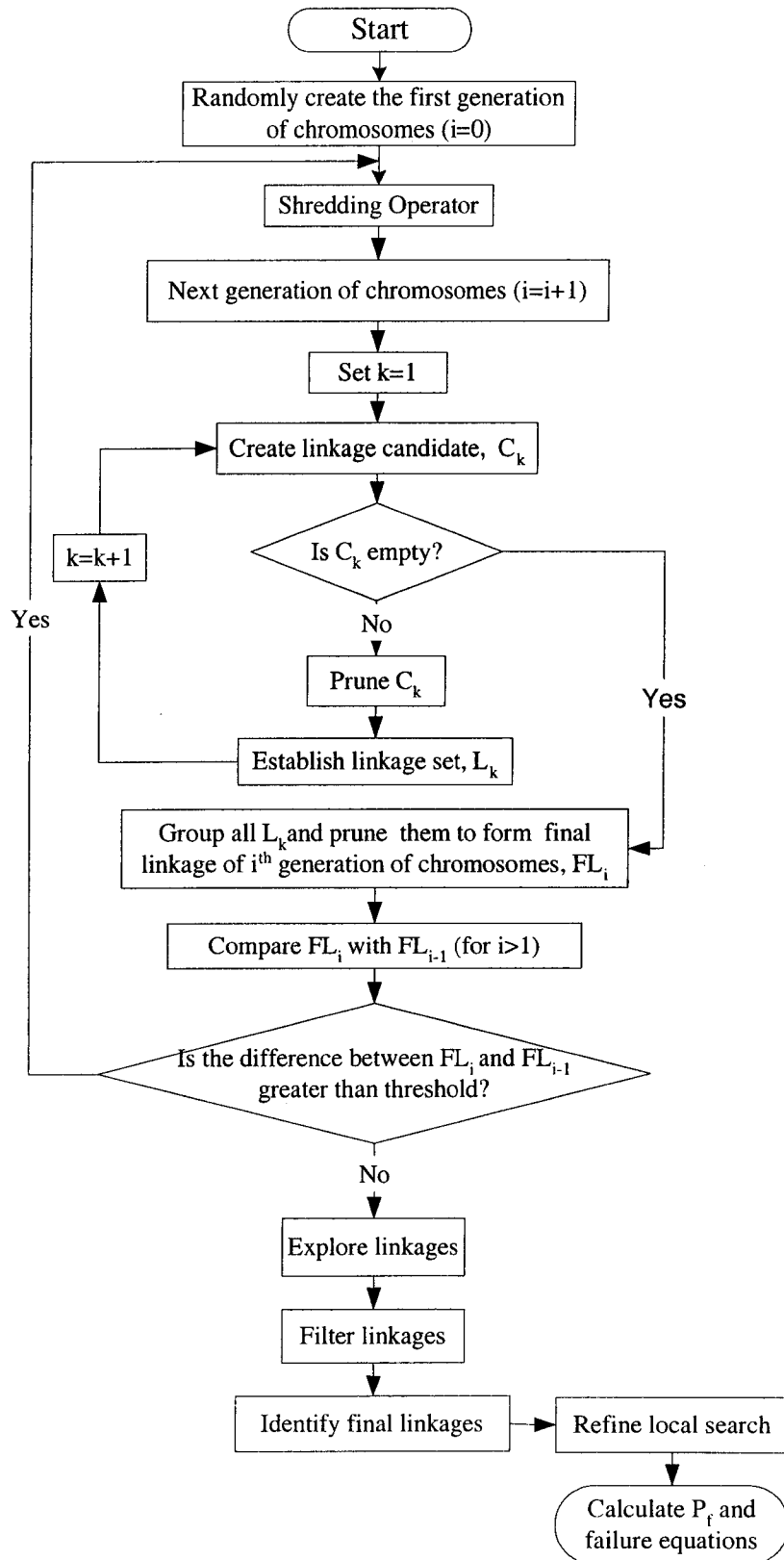


Figure 4.2 Flowchart of proposed algorithm

1. Randomly generate the first population of chromosomes representing search directions in the standardized random variable space.
2. Generate the next generation of chromosomes using the shredding operator. The purpose is to improve the robustness of the population.
3. Set $k=1$; and determine the linkage candidate set (C_k) of all genes from the database encountered in all the chromosomes of the current population.
4. Based on fitness, prune the “good” linkages and establish L_k as a subset of C_k .
5. Set $k = k+1$;
6. Form new C_k from L_{k-1} ;
7. Repeat steps 4 through 6 until C_k is empty;
8. Repeat steps 2 through 7 until the convergence of linkage sets;
9. Use the mode enumeration filter to identify the important failure modes and reduce the number of repeated modes;
10. Use the Tabu algorithm to refine the local search.
11. Obtain the failure mode expression for all the pertinent modes and calculate each mode’s probability of failure and the system reliability.

Despite its simplicity, the proposed algorithm provides a powerful and robust tool for solving structural reliability problems as demonstrated in the next two sections of this chapter and the accompanying chapter that illustrates the application of the algorithm for solving structural reliability problems and provides a thorough sensitivity analysis.

4.5 ILLUSTRATIVE EXAMPLE

Problem statement

A simple example is presented to illustrate the proposed method. The example shown in Figure 4.3 consists of the same three-bar truss analyzed by Deng (2000). The input data for the loads and member strengths are listed in Table 4.1. The mean and the COV of the random variables are identified as the loads L_1 , L_2 and L_3 and the yielding strengths for members 1, 2 and 3 respectively are identified as R_1 , R_2 and R_3 .

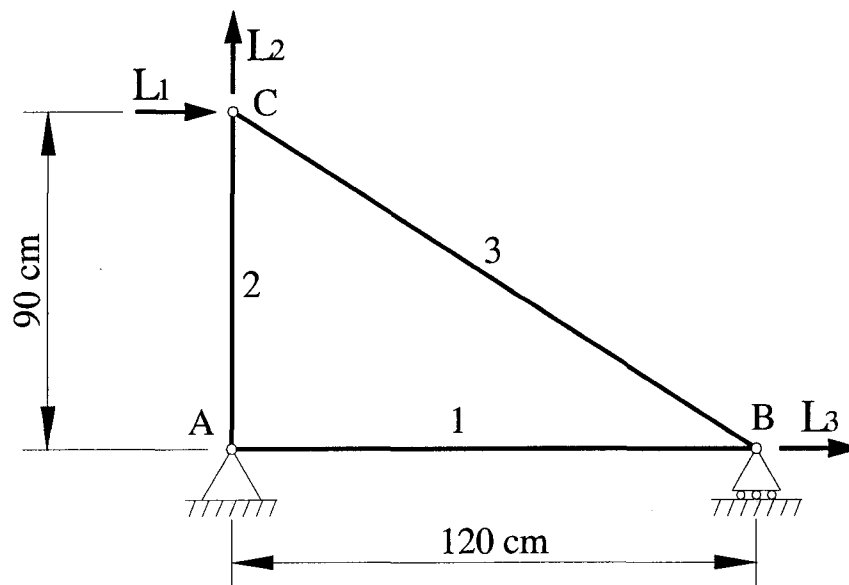


Figure 4.3 Three-member truss example.

Table 4.1 Input data for three-member truss example

Random variables	Mean (KN)	COV	Distribution type
L ₁	35	10%	Normal
L ₂	24	10%	Normal
L ₃	15	10%	Normal
R ₁	63.48	5%	Normal
R ₂	63.48	5%	Normal
R ₃	63.48	5%	Normal

The potential member failure modes are labeled Z_1^+ , Z_1^- , Z_2^+ , Z_2^- , Z_3^+ and Z_3^- , where the subscript indicates which member fails. The “+” sign denotes failure in tension, while the “-” sign denotes failure in compression. It is assumed that the yielding strengths for tension and compression are independent random variables with the same mean and standard deviations. The reliability indexes for all the failure modes are listed in Table 4.2. From the table, we observe that the first two failure modes are the dominant failure modes.

Table 4.2 Reliability Indices of failure modes

Failure Mode	β
Z ₁₊	2.719
Z ₂₊	2.775
Z ₃₋	3.670
Z ₃₊	19.94
Z ₁₋	22.93
Z ₂₋	23.86

The objective of this section is to verify the ability and the efficiency of the proposed hybrid Data Mining/Shredding Genetic Algorithm to identify the most dominant modes of the above problem whose exact solution is known. The analysis process follows the steps outlined in the previous section as well as Figure 4.4.

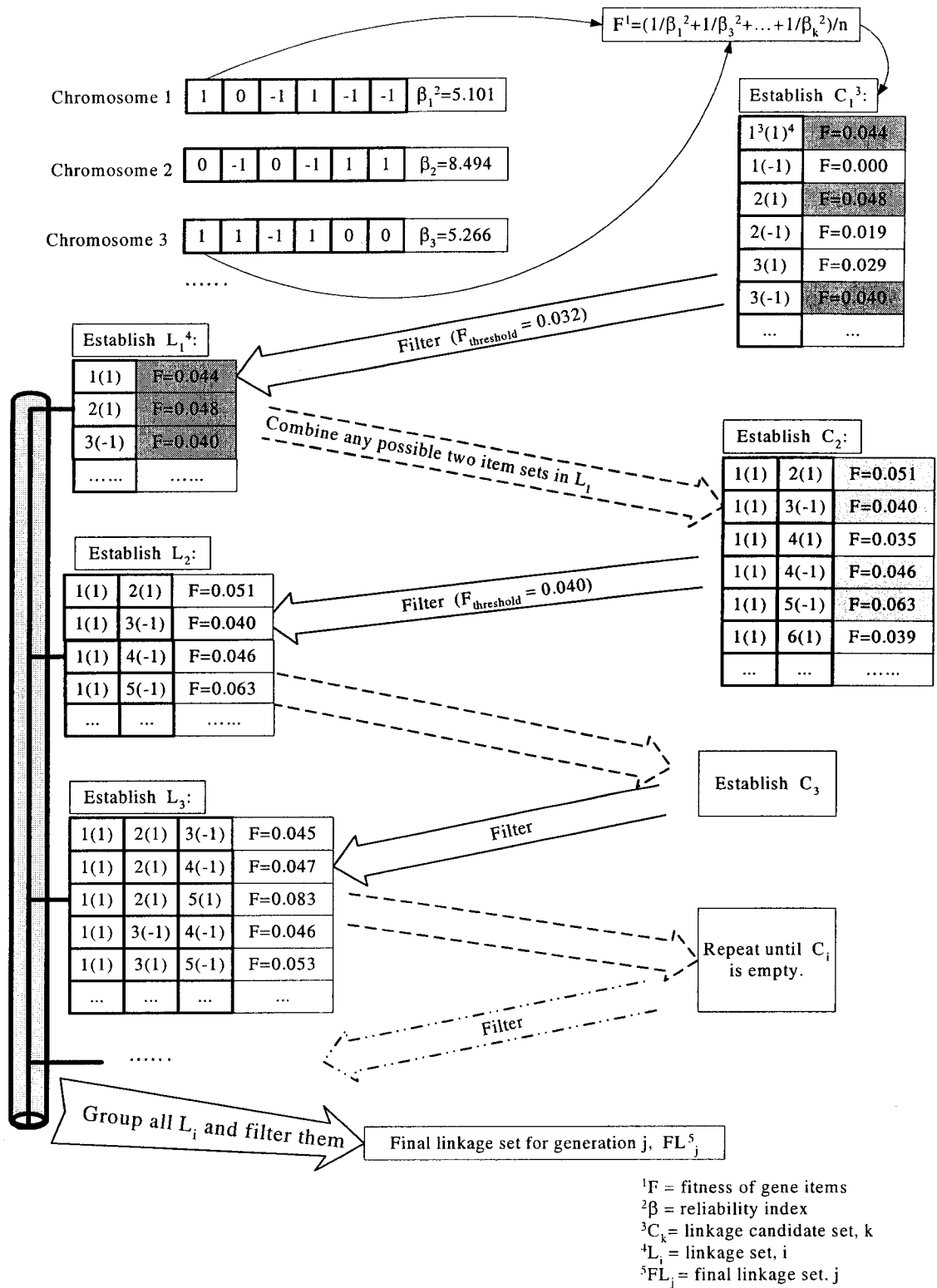


Figure 4.4 Data mining process

Exploration process

The first step in the method is to set the problem using the genetic coding of Shao and Morutsu (1999). The first randomly generated population is listed in Table 4.3. Before the data-mining search is initiated, a second population is created using the genetic shredding algorithm in order to refine the population toward more focused search directions than those that were originally randomly created. The chromosomes of the second generation are also listed in Table 4.3. Each population is formed by 10 chromosomes. A sensitivity analysis is performed in the accompanying chapter to illustrate the effect of the population size on the convergence of the algorithm.

Table 4.3 The first and second generations of chromosomes

Generation no.	Chromosome no.	Gene position and value						β
		1	2	3	4	5	6	
1	1	1	0	-1	1	-1	-1	5.101
	2	0	-1	0	-1	1	1	8.494
	3	1	1	-1	1	0	0	5.266
	4	1	-1	-1	0	1	0	9.019
	5	1	-1	1	1	0	1	16.507
	6	1	1	1	1	1	-1	6.402
	7	1	1	-1	-1	-1	1	3.953
	8	0	-1	0	0	-1	-1	10.767
	9	1	1	-1	-1	1	-1	6.382
	10	0	-1	1	-1	1	1	6.449
2	11	1	1	1	1	-1	1	3.953
	12	1	1	1	0	1	1	6.028
	13	1	0	-1	-1	-1	1	5.101
	14	1	-1	0	-1	-1	1	4.516
	15	0	1	-1	-1	-1	-1	5.307
	16	0	1	0	-1	1	0	7.356
	17	1	1	-1	-1	1	1	6.382
	18	1	1	-1	-1	-1	0	3.608
	19	1	1	0	0	-1	0	2.795
	20	1	0	-1	1	-1	0	4.563

Using the shredding operator of Deng et. al. (2003), the average fitnesses ($1/\beta^2$) for all the gene items of populations 1 and 2 are obtained as listed in Table 4.4. The mean fitness of all the gene items of Table 4 is 0.035. Hence, the threshold is set to be equal to $0.9 \times \text{mean} = 0.032$. The next step is to prune the gene items whose average fitness is less than the 0.032 threshold. The set of surviving genes L_1 is listed in Table 4.5. A sensitivity

analysis is performed in the accompanying chapter to illustrate the effect of the threshold on the speed of convergence.

Table 4.4 Linkage candidate C_1 items

Item no.	Gene position and value						Fitness
	1	2	3	4	5	6	
1	1						0.044
2	-1						0.000
3		1					0.048
4		-1					0.019
5			1				0.029
6			-1				0.040
7				1			0.036
8				-1			0.037
9					1		0.021
10					-1		0.055
11						1	0.034
12						-1	0.026

Table 4.5 Linkage L_1 items

Item no.	Gene position and value						Fitness
	1	2	3	4	5	6	
1	1						0.044
2		1					0.048
3			-1				0.040
4				1			0.036
5				-1			0.037
6					-1		0.055
7						1	0.034

The list of items containing 2 gene pairs, C_2 , is obtained as shown in Table 4.6. The item sets no. 4 and 5 in Table 4.5 cannot be combined because they have the same gene, the difference being only in the value of the gene. Using equation (1) and the method described by Deng et. al. (2003), the fitness of each possible item is calculated as given in the last column of Table 4.6..

Table 4.6 Linkage candidate C_2 items

Item no.	Gene position and value						Fitness
	1	2	3	4	5	6	
1	1	1					0.051
2	1		-1				0.040
3	1			1			0.035
4	1			-1			0.046
5	1				-1		0.063
6	1					1	0.039
7		1	-1				0.044
8		1		1			0.041
9		1		-1			0.041
10		1			-1		0.074
11		1				1	0.045
12			-1	1			0.041
13			-1	-1			0.044
14			-1		-1		0.050
15			-1			1	0.042
16				1	-1		0.050
17				1		1	0.034
18				-1	-1		0.053
19				-1		1	0.031
20					-1	1	0.054

The mean fitness of all gene items of Table 6 is 0.045. Hence, the threshold is set to be equal to $0.9 \times \text{mean} = 0.040$. The gene items whose average fitness is less than the 0.040 threshold are pruned and the set of surviving pairs of genes, L_2 , is listed in Table 4.7.

Table 4.7 Linkage L_2 items

Item no.	Gene position and value						Fitness
	1	2	3	4	5	6	
1	1	1					0.051
2	1		-1				0.040
3	1			-1			0.046
4	1				-1		0.063
5		1	-1				0.044
6		1		1			0.041
7		1		-1			0.041
8		1			-1		0.074
9		1				1	0.045
10			-1	1			0.041
11			-1	-1			0.044
12			-1		-1		0.050
13			-1			1	0.042
14				1	-1		0.050
15				-1	-1		0.053
16					-1	1	0.054

Using the same logic, $C_4, L_4, C_5, L_5, \dots$ are obtained and the process continues until C_i is empty. At that point, L_1, L_2, \dots, L_i are combined in one global list that is pruned again to obtain the final linkage set of the first and the second generations. These are assembled in a final linkage list for the first population, FL_1 .

Subsequently, a new generation of chromosomes is assembled using the Genetic Shredding operator. The Apriori algorithm is applied again to obtain the second final linkage list, FL_2 . FL_2 and FL_1 are compared to find how many of the listed items are repeated. The comparison involves the “important” linked items that are defined to be the items with fitnesses exceeding 0.9 times the average fitness. If the number of new important linkages is

less than 5% of those detected in the previous set, then the process of detecting important linkages is stopped. A sensitivity analysis is performed in the accompanying chapter to illustrate the effect of the convergence criterion on the accuracy of the algorithm.

For the truss example of this section, the linkage detection process is stopped after 5 generations (where each generation is composed of 10 chromosomes). The final failure modes identified are listed in Table 4.8. The process required a total of 67 structural analysis iterations.

Table 4.8 Final linkages after 5 generation of chromosomes

Linkage no.	Gene position and value						β
	1	2	3	4	5	6	
1	1	1			-1		2.794
2	1		1	-1			2.856
3	1			-1			2.856
4	1				-1		3.226
5	1	1	-1		-1		3.227
6	1		1	-1	-1		3.298
7		1			-1		3.357
8	1			-1	-1		3.498
9	1	1	-1	-1	-1		3.608
10	1		1	-1	-1	-1	3.688

Mode Filtering

As previously mentioned, several of the final chromosomes are not independent but are different representations of the same failure mode. To refine the failure mode list, the enumeration filtering criterion previously described is applied. As an example, by comparing the linkage set No.4, to that of No.1, it is observed that when the gene in position 2 with a value equal to 1 (the same value observed in chromosome 1 for gene no. 2) is added into chromosome No.4, the reliability index is reduced. This strongly indicates that failure mode No.4 is not independent from failure mode No.1. Using the same logic, failure modes No. 9 and No.5 are not independent either. After the filtering operation is completed, the remaining seemingly independent failure modes are listed in Table 4.9.

Table 4.9 Filtered failure modes

Failure mode no.	Gene position and value						β
	1	2	3	4	5	6	
1	1	1			-1		2.794
2	1		1	-1			2.856

Tabu local search

The chromosomes in Table 4.9 are along pre-determined search directions. To find the actual direction for the minimum reliability indexes, the Tabu local search algorithm previously described by Wang & Ghosn (2003) is used. The Tabu local search process is continued until no improvement in the reliability index is obtained. This occurred of eleven iterations. The reliability indices after the Tabu local search for the two most important failure modes are listed in Table 4.10. The difference between the exact values of the reliability index shown in () of Table 4.10, and those obtained from the proposed Data Mining/Genetic Shredding algorithm is less than 0.002. Table 10 also shows the direction cosine for these two modes of failure and compares them to the exact values.

Table 4.10 Reliability indices of failure modes (after tabu local search)

Failure mode no.	Unit vector of design point						β
	1	2	3	4	5	6	
1	0.535 (0.551)	0.484 (0.504)	0.000 (0.000)	0.000 (0.000)	-0.693 (-0.669)	0.000 (0.000)	2.777 (2.775)
2	0.695 (0.706)	0.000 (0.000)	0.304 (0.303)	-0.651 (-0.640)	0.000 (0.000)	0.000 (0.000)	2.720 (2.719)

Note: the value inside parenthesis is the exact solution

Determination of Failure Equations

From the failure modes listed in Table 4.10, the corresponding failure equations can be obtained based on FORM. For example, for the failure mode No.2, The failure equation is

$$Z_2 = R_1 - 0.968L_1 - 0.988L_3 - 1.524 \quad (\text{Eq. 4.2})$$

Which is compared to the exact failure equation of failure mode No.2 given as:

$$Z_2 = R_1 - L_1 - L_3 \quad (\text{Eq. 4.3})$$

Comparing Equation 4.2 and Equation 4.3, it can be seen that Equation 4.2 gives a very good approximation to the exact failure Equation 4.3. A more refined local search criterion associated with a large number of structural analysis iterations would lead to more accurate results as listed in Table 4.11.

Table 4.11 Accuracy of tabu search vs. structural evaluation times

No.	Failure equation	β	Evaluation times (Tabu Search)
Exact	$Z_2=R_1 - L_1 - L_3$	2.7192	--
1	$Z_2 = R_1 - 0.968L_1 - 0.988L_3 - 1.524$	2.7200	11
2	$Z_2 = R_1 - 0.970L_1 - 0.988L_3 - 1.475$	2.7198	20
3	$Z_2 = R_1 - 0.986L_1 - 1.078L_3 - 0.682$	2.7197	35
4	$Z_2 = R_1 - 0.990L_1 - 1.050L_3 - 0.487$	2.7195	52
5	$Z_2 = R_1 - 0.998L_1 - 1.001L_3 - 0.060$	2.7193	158
6	$Z_2 = R_1 - 1.003L_1 - 0.999L_3 - 0.010$	2.7193	220
7	$Z_2 = R_1 - L_1 - 1.001L_3 - 0.010$	2.7193	228

In summary, it is observed that the proposed algorithm provides accurate results in a reasonably efficient manner. Further comparisons with structural analysis problems and other benchmark problems are provided in the following section and the accompanying chapter.

4.6 TESTS FOR EVALUATING THE PERFORMANCE OF PROPOSED ALGORITHM

To test the efficiency and accuracy of the proposed Data Mining/Shredding Genetic Algorithm, the benchmark problems presented by Engelund & Rackwitz (1993) are solved in this section and compared to the solutions obtained using Importance Sampling Algorithms as provided by Engelund & Rackwitz (1993) and the GEMGA method described by Wang and Ghosn (2003). The comparison is also summarized in Table 4.12.

Table 4.12 Comparison of different reliability methods

No.	Benchmark problem	P_f				β				Evaluation times		
		Exact	Importance Sampling	GEMGA	Data Mining	Exact	Importance Sampling	GEMGA	Data Mining	Importance Sampling	GEMGA	Data Mining
1	Number of variables (10)	$2.92 \cdot 10^{-7}$	$2.89 \cdot 10^{-7}$	$2.92 \cdot 10^{-7}$	$2.92 \cdot 10^{-7}$	5.000	5.000	5.000	5.000	500	255	120
1	Number of variables (50)	--	--	--	--	10.000	--	10.650	10.890	--	754	350
2	multiple β points	$18.4 \cdot 10^{-8}$	Good estimate	$9.65 \cdot 10^{-8}$	$9.55 \cdot 10^{-8}$	5.085	--	5.206	5.200	500	40	30
3	Parallel system	$2.09 \cdot 10^{-4}$	$2.02 \cdot 10^{-4}$	$2.10 \cdot 10^{-4}$	$2.08 \cdot 10^{-4}$	3.529	3.537	3.527	3.530	500	213	96
4	Noisy limit state function	$4.47 \cdot 10^{-3}$	Good estimate	$4.46 \cdot 10^{-3}$	$4.45 \cdot 10^{-3}$	2.614	--	2.615	2.616	500	97	63
5	Series system	$1.27 \cdot 10^{-2}$	$1.28 \cdot 10^{-2}$	$1.26 \cdot 10^{-2}$	$1.26 \cdot 10^{-2}$	2.235	2.232	2.238	2.238	500	250	80

Benchmark test No. 1: Number of variables and probability level

The failure equation, Z , is given as a function of the number of random variables, n , and the reliability index β by:

$$Z = \beta\sqrt{n} - \sum_{i=1}^n U_i, \quad U_i \sim N(0,1) \quad (\text{Eq. 4.4})$$

$$n = 10, \quad \beta = 5.0, \quad P_f = 2.92 * 10^{-7} \quad (\text{Eq. 4.5})$$

Where U_i designates the variable i in the standard normal space with mean 0 and standard deviation 1.0.

For $n=10$, the hybrid algorithm converged after 120 function evaluation times.

The exact $\beta=5.00$ was obtained. The direction cosines obtained were:

$$\alpha = (\alpha_1 \dots \alpha_i \dots \alpha_{10}), \quad \alpha_i = 0.316 \quad (\text{Eq. 4.7})$$

which are the exact values for the problem described in Equations 4.4 and 4.5. Convergence was reached with about half the iterations needed for GEMGA and about one quarter those used by Engelund & Rackwitz (1993) to solve the problem using Importance Sampling.

For $n=50$ convergence was reached after 350 function evaluation times. $\beta=10.89$ was obtained rather than the exact value of 10.00, which is acceptable given the high value of $\beta=10.0$.

Benchmark test No. 2: multiple β points

The limit state equation is a hyperbola:

$$\begin{aligned} Z &= X_1 X_2 - PL \\ P &= 14.614, L = 10.0 \\ X_1 &\sim N(78064.4, 11709.7) \\ X_2 &\sim N(0.0104, 1.56 * 10^{-3}) \end{aligned} \quad (\text{Eq. 4.8})$$

In this case, the hybrid algorithm converged after 30 function evaluation times resulting in a reliability index $\beta=5.20$ as compared to the exact $\beta=5.085$ while GEMGA required about 80 iterations.

Benchmark test No. 3: Parallel System

The following four parallel failure modes are analyzed in this example:

$$\begin{aligned} Z_1 &= 2.677 - U_1 - U_2 \\ Z_2 &= 2.500 - U_2 - U_3 \\ Z_3 &= 2.323 - U_3 - U_4 \\ Z_4 &= 2.250 - U_4 - U_5 \\ U_i &\sim N(0,1), \quad i=1,\dots,5 \end{aligned} \quad (\text{Eq. 4.9})$$

The correct reliability indices for each mode are:

$$\begin{aligned}
\beta_1 &= 1.893 \\
\beta_2 &= 1.768 \\
\beta_3 &= 1.643 \\
\beta_4 &= 1.591
\end{aligned}
\tag{Eq. 4.10}$$

In this case the Data Mining/Shredding Algorithm converged to accurate answers in a total of 96 iterations which is again less than $\frac{1}{2}$ the number of iterations required by GEMGA while the accuracy is practically the same.

Benchmark test No. 4: noisy limit state function

The noisy function with six random variables is given as:

$$\begin{aligned}
Z &= X_1 + 2X_2 + 2X_3 + X_4 - 5X_5 - 5X_6 + 0.001 \sum_{i=1}^6 \sin(100X_i) \\
X_1 = X_2 = X_3 = X_4 &\sim N(120,12) \\
X_5 &\sim N(50,15). \quad X_6 \sim N(40,12)
\end{aligned}
\tag{Eq. 4.11}$$

In this case the Data Mining/Shredding algorithm converged to a beta of 2.616 in a total of 63 iterations as compared to the 123 iterations required by GEMGA.

Benchmark test No. 5: series system

Two failure modes in series with 6 random variables are given as:

$$\begin{aligned}
Z_1 &= X_1 + 2X_2 + X_4 + X_5 - 5X_6 \\
Z_2 &= X_2 + 2X_3 + X_4 - 5X_6 \\
X_1 = X_2 = X_3 = X_4 &\sim LN(60,6) \\
X_5 &\sim \text{Gumbel}(20,6) \quad X_6 \sim \text{Gumbel}(25,7.5)
\end{aligned}
\tag{Eq. 4.12}$$

In this case the Data Mining/Shredding algorithm converged to a reliability index in a total of 80 iterations as compared to the 250 iterations that GEMGA required for the same level of accuracy.

4.7 CONCLUSIONS

This chapter proposed a new hybrid Data Mining/Genetic Shredding Algorithm to solve structural reliability problems. The method is based on the Apriori algorithm originally proposed by Agrawal (1994) and modified for the identification of different structural failure modes in complex structural systems. In addition, the algorithm utilizes the genetic shredding operator first proposed by Deng (2000). A new linkage learning process is developed in this chapter to explore the standard normalized random variable space for “good” search directions and to improve the convergence to the local optima as well as the global optimum. A new filtering criterion is proposed to identify the independent failure modes. A Tabu local search process is added to the algorithm to refine the search and find accurate solutions that may not necessarily lie in the directions that were pre-set during the original genetic coding.

The accuracy and efficiency of the algorithm was tested using a simple three-bar truss problem. In addition, five benchmark tests previously proposed by Engelund & Rackwitz (1993) for testing the efficiency and accuracy of importance sampling methods have been used in this chapter to verify the robustness of the proposed algorithm. Additional sensitivity analyses and further demonstrations of the applicability of the

method for solving structural reliability problems are provided in the accompanying chapter.

CHAPTER FIVE

HYBRID DATA MINING/GENETIC SHREDDING ALGORITHM: APPLICATIONS & SENSITIVITY ANALYSIS

ABSTRACT

This is the second part of effort describe the methodology and application of a hybrid Data Mining/Genetic Shredding Algorithm for the reliability analysis of structural systems. This chapter illustrates the application of the method described in the accompanying chapter and concentrates on performing sensitivity analyses to study the effect of various parameters on the robustness of the proposed algorithm and its speed of convergence. A ten-bar truss example with material nonlinearity is solved to illustrate the high efficiency and accuracy of the proposed algorithm in identifying multiple failure modes. In addition, the proposed algorithm is combined with a general-purpose finite element program to find the important failure modes of two-geometrically nonlinear cable systems. These examples include a simplified model of a suspension bridge and a realistic model of a cable stayed bridge. The system reliability is computed in each case to demonstrate the applicability of the proposed algorithm for finding the solutions of complex structural systems. The results of the algorithm are compared to those of Monte Carlo Simulations or those of available published solutions to verify the accuracy of the proposed method.

5.1 INTRODUCTION

In the preceding chapter, a hybrid Data Mining/ Genetic Shredding Algorithm is proposed to identify multiple important failure modes and calculate the reliability of structural systems. In this chapter, sensitivity analyses are performed to study the impact of various factors such as convergence criteria, pruning fitness threshold, and population size on the proposed algorithm. The sensitivity analyses are performed on a ten-bar truss example and an appropriate set of parameters and convergence criteria are suggested.

Furthermore, this chapter demonstrates the application of the algorithm to realistic models of long-span cable stayed bridges. The cabled stayed bridge example along with a simplified suspension bridge problem adapted from the work of Frangopol and Iami (2000) are solved to verify the capability of the proposed algorithm to solve geometrically nonlinear problems and to demonstrate the ability of linking the algorithm to general-purpose finite element packages. Such capacity increases the applicability of the proposed algorithm for finding the reliability of realistic structural problems.

5.2 ANALYSIS OF A 10-BAR TRUSS

PROBLEM STATEMENT

This example consists of a truss problem solved herein to illustrate the efficiency and accuracy of the proposed algorithm in identifying multiple failure modes. The problem is adapted from Gorman (1979) who has identified all its failure modes using the incremental loading technique. Comparing the results obtained using the proposed failure mode search

methodology to the solution provided by Gorman (1979) will confirm that the proposed Data Mining/Genetic Shredding Algorithm will converge to the correct solution given a sufficient number of iterations.

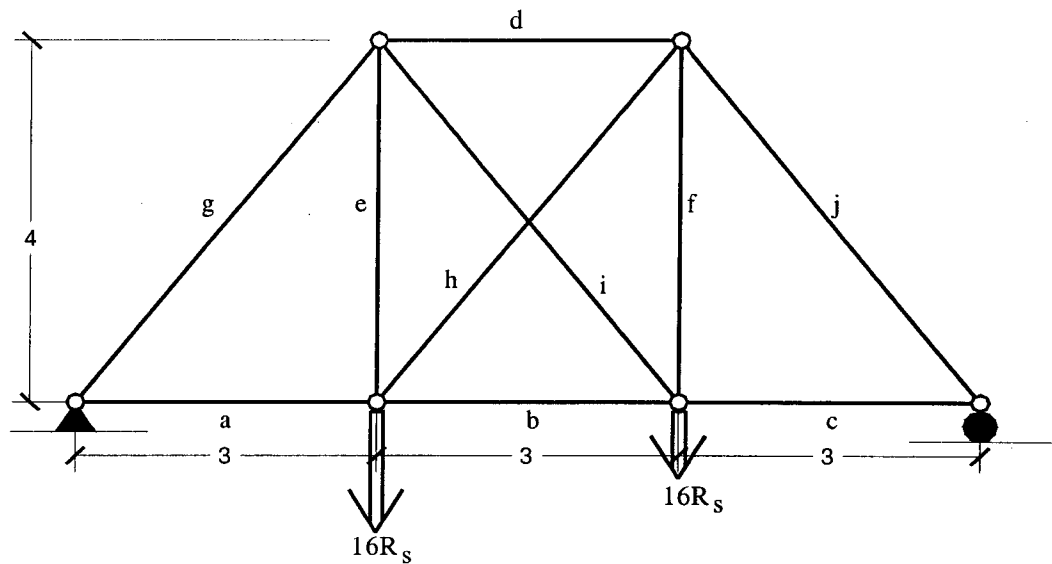


Figure 5.1. Ten-bar truss example as adapted from Gorman (1979)

The 10 bar truss is shown in Figure 5.1. The structure has a single degree of redundancy. Hence, the maximum number of rigid-plastic collapse modes is:

$$n = \binom{10}{2} = \frac{10!}{8!2!} = 45 \quad (\text{Eq. 5.1})$$

However, several collapse modes consist of the failure of one element, and other possible collapse modes are inconsistent with the loading condition. By eliminating these modes, 19 different significant collapse mechanisms are counted (Gorman;1979).

In this example, all element strengths are independent random variables and follow normal distributions. The mean value of each element's strength is shown in Table 5.1. The COV of each element's strength is assumed to be 10%. For more details about the problem, the reader is advised to consult the reference by Gorman (1979).

Table 5.1 Random variables of ten-bar truss structure

Random Variables	Mean	COV	Distribution Type
Strength of member A	15	1.5	Normal
Strength of member B	15	1.5	Normal
Strength of member C	15	1.5	Normal
Strength of member D	15	1.5	Normal
Strength of member E	20	2.0	Normal
Strength of member F	20	2.0	Normal
Strength of member G	25	2.5	Normal
Strength of member H	10	1.0	Normal
Strength of member I	10	1.0	Normal
Strength of member J	25	2.5	Normal

FAILURE MODES IDENTIFICATION

The application of the proposed algorithm to identify the failure modes of the 10-bar truss system begins by randomly creating the first generation of search directions represented by a set of chromosomes. In this case, the number of chromosomes in each generation is selected to be 20 based on the rule of thumb of using twice the number of random variables. The second generation is obtained using the classical Genetic Algorithm applying the shredding operator. The next step consists of identifying the linkage sets and pruning these to get the final linkage set for the current generation of chromosomes. A new generation of chromosomes is then created and the corresponding final linkage set is identified. The process is assumed to have reached convergence when the number of new linkages identified from a new generation is less than 10% of the number of linkages already

identified. For the 10-bar truss being analyzed, the proposed algorithm was stopped after six generations. In this case, the total number of linkages detected was 439.

The final linkage set identified after six generations was subsequently filtered to eliminate the cases where two linkages that have completely the same genes but one or more of those genes have different gene values. In these cases, the linkage that has the higher reliability index is filtered out. The next step consisted of pruning the failure modes to keep only the independent modes. This is done based on the failure search principle provided in the accompanying chapter which states that “if a partial failure mechanism has a high probability of occurrence, then any subset of elements forming a sub-mechanism of that partial mechanism must also have a high probability of occurrence”. The filtration process resulted in the identification of the 19 failure modes listed in Table 5.2.

Table 5.2 Final failure equations for ten-bar truss

Failure mode No. (ordered as identified by the algorithm)	Failure mode No. (as listed by Gorman; 1979)	Failure Equation	Calculated β	Exact β
1	1	$Z = R_a - 9.00$	4.000	4.000
2	2	$Z = R_g - 15.00$	4.000	4.000
3	9	$Z = R_c - 6.00$	6.000	6.000
4	10	$Z = R_j - 10.00$	6.000	6.000
5	5	$Z = R_b + R_d - 15.00$	7.080	7.070
6	18	$Z = R_b + 0.45R_e - 14.50$	5.432	5.400
7	3	$Z = R_d + 0.45R_e - 14.50$	5.432	5.400
8	19	$Z = R_b + 1.05R_f - 13.80$	8.602	8.600
9	11	$Z = R_d + 1.05R_f - 13.80$	8.603	8.600
10	16	$Z = R_e + 2.33R_f - 30.20$	7.180	7.07
11	6	$Z = R_b + 2.25R_h - 14.95$	8.341	8.321
12	15	$Z = R_h + R_j - 5.01$	10.60	10.607
13	14	$Z = R_b + 2.25 R_i - 14.93$	8.352	8.321
14	8	$Z = R_d + 2.25 R_i - 14.96$	8.338	8.321
15	7	$Z = R_e + 3.00 R_h - 28.60$	6.085	6.037
16	13	$Z = R_d + 2.25 R_h - 16.82$	8.396	8.321
17	12	$Z = R_f + 2.14R_h - 13.22$	9.624	9.615
18	17	$Z = R_f + 2.14 R_i - 13.12$	9.654	9.615
19	4	$Z = R_e + 5.00 R_i - 26.48$	6.080	6.037
β_{system}			3.829	3.829

As seen in Table 5.2, the algorithm has clearly identified all the 19 failure modes listed by Gorman (1979). At the same time, the first two failure modes are recognized as the critical failure modes for the structural system. Finally, the Tabu algorithm is used to refine the local search and obtain better estimates of the failure mode directions and the reliability

indexes for the two critical failure modes and the failure mode equations are obtained as listed in Table 5.2.

For the system reliability of the structure, if only two critical failure modes are considered, the result from the proposed algorithm gave a $\beta_{\text{sys}} = 3.833$, with a probability of failure $P_f = 6.33 \times 10^{-5}$. These are compared to $\beta_{\text{sys}} = 3.829$ and $P_f = 6.43 \times 10^{-5}$ obtained from 10 million cycles of a Monte Carlo Simulation. If local searches are extended to all 19-failure modes, better estimation of the system's probability of failure and the system's reliability index are obtained. The proposed algorithm provided a $\beta_{\text{sys}} = 3.829$, with a probability of failure $P_f = 6.43 \times 10^{-5}$.

SPEED OF CONVERGENCE

The total number of structural analyses required for obtaining the system failure probability and system reliability index was 314. The Genetic search algorithm using the benefits of the shredding operator required a total of 120 structural evaluations. The process of obtaining accurate reliability index values for the identified important linkages and additional 174 structural evaluations are needed. Finally, during the Tabu local search 10 iterations are used for each critical failure mode resulting in 20 structural analyses. This large number of iterations was needed in this example in order to identify all the failure modes. In practical situations, only the first two failure modes have significant contributions to the total probability of failure and system reliability. Hence, a total of about 150 structural analyses would have been adequate to obtain a good approximation of this truss system's

reliability. The next section will provide more information about the convergence criteria and the sensitivity of the results to the threshold parameters used during the search process.

5.3 SENSITIVITY ANALYSIS

POPULATION CONVERGENCE CRITERIA

The convergence criterion used above stopped the population generation process when the number of newly observed linkages is less than 10% of the total previously identified linkages. Table 5.3 shows how the number of identified linkages changes when the initial random generator seed used to create the initial population of chromosomes is changed. For example, for the first case studied (seed 1), we observe that the total number of linkages identified after the second generation is 266 while the total number of linkages becomes 326 with the third generation. The final number of modes identified after pruning is 18 for both cases. The number of failure modes increases to 19 starting with the fourth generation when the number of new linkages increases by an additional 12.9%. Similar observations are made for other runs executed using different initial seeds. Table 5.4 shows how the convergence criterion affects on the average the number of final mode shapes. In four of the five cases studied, the number of modes identified was 19 when the change in the number of linkages was less than 10%. The other case led to 18 modes when the number of new linkages was less than 10%, which in most situations would still lead to an acceptable assessment of the overall system reliability. Table 5.4 summarizes how, on the average, the total number of identified failure modes changes with the cut-off value. From Table 5.4, it

is observed that choosing 10% of the total number of linkages as the threshold is reasonable since for the cases where the 10% criterion is used, an average of 18.8 failure modes are identified compared to the actual total of 19 failure modes. It is noted that in all the cases investigated, the total number of modes would have been identified in 7 or fewer generations.

Table 5.3 Effect of linkage threshold on convergence

Generation Number		2	3	4	5	6	7	8
Seed 1	No. of Linkages before filtering	266	326	368	415	439	456	488
	No. of Linkages after filtering	18	18	19	19	19	19	19
	Ratio= $(N_i - N_{i-1})/N_{i-1}$ (%)	--	22.6%	12.9%	12.8%	5.8%	3.9%	7.0%
Seed 2	No. of Linkages before filtering	255	364	405	446	493	513	527
	No. of Linkages after filtering	14	14	14	15	17	19	19
	Ratio (%)	--	42.7%	11.3%	10.1%	10.5%	4.1%	2.7%
Seed 3	No. of Linkages before filtering	264	322	367	404	435	452	474
	No. of Linkages after filtering	16	16	18	18	18	19	19
	Ratio (%)	--	22.0%	14.0%	10.1%	7.7%	3.9%	4.9%
Seed 4	No. of Linkages before filtering	277	341	467	503	537	620	643
	No. of Linkages after filtering	18	18	19	19	19	19	19
	Ratio (%)	--	23.1%	37.0%	7.7%	6.8%	15.5%	3.7%
Seed 5	No. of Linkages before filtering	240	300	334	351	369	377	471
	No. of Linkages after filtering	16	17	18	19	19	19	19
	Ratio (%)	--	25.0%	11.3%	5.1%	5.1%	2.2%	24.9%

Table 5.4 Effect of linkage threshold cutoff criteria on convergence

Convergence Criteria	15%	10%	5%
Average number of failure modes	17.6	18.8	19.0
Number of Generation to Convergence	4.2	5.8	7.2
β_{system}	3.829	3.829	3.829

LINKAGE PRUNING FITNESS THRESHOLD

The next sensitivity analysis tested the effectiveness of the pruning fitness threshold value. As mentioned in the accompanying chapter, the criterion used to prune the weak genes based on their fitness, before assembling the remaining ones into the linkage set C_i and the final linkage sets L_i , consisted of keeping the genes whose fitness index value was above 0.9 times the average fitness. In this section, a sensitivity analysis is performed to verify that the 0.90-criterion is a reasonable cut-off value. Hence, different cut-off values ranging from 0.80 to 1.00 are compared. The results are provided in Table 5.5 that shows how the number of identified failure modes decreases as the cut-off value is increased from 0.80 to 1.0. It is noted that when the threshold is set at 1.0, only two final linkages are identified which lead to two modes of failure. However, these two modes are the two most critical failure modes having the lowest reliability index values for the truss system analyzed in this section.

Table 5.5 Sensitivity of results to pruning threshold of gene items

Factor (*mean of fitness value)	0.8	0.85	0.9	0.94	0.95	1.00
Number of Linkages identified	19	19	19	17	16	2
Number of structural analyses needed (excluding Tabu local Search)	322	315	294	279	274	232
β_{system}	3.829	3.829	3.829	3.829	3.829	3.833

POPULATION SIZE

For the GA algorithm to produce an efficient and robust search strategy, an appropriate number of chromosomes must be chosen to form each generation of search directions. If there are too few chromosomes in one generation, the proposed algorithm has only few possibilities to perform the crossover operation that is necessary for the exploration of new search directions, and only a small part of the search space is explored. This may lead to either a large number of generations to reach convergence or even worse, the proposed algorithm may fail altogether. On the other hand, if there are too many chromosomes in one generation, the proposed algorithm will be very slow since it would be difficult for the proposed algorithm to focus on the important search directions.

Several studies have investigated this problem. For example, Deng (2000) gave a crude population scaling law, which states that the optimum population size should be related to the number of genes in a chromosomes and the number of possible values that a gene can take. This law may be represented by an equation of the form

$$N_{\text{population size}} = \text{order}(\lambda * b^k / k) \quad (\text{Eq. 5.2})$$

Where $N_{\text{population size}}$ is the optimum population size, λ =length of chromosomes, $b=2$ for binary coding and k is the average size of the schema of interest (effectively the average number of bits per parameter). Deng (2000) found that Equation 5.2 is conservative and one can most likely get by with population at least twice as small.

In this 10-bar truss example, the number of variables is $\lambda=10$, each of which can be either 0, 1 or -1 . Thus, b is equal to 3. Hence, following Deng's (2000) suggestion, the optimum population size should be:

$$\text{Population size} = \text{Order} (10 * 3^1 / 1) / 2 = \text{Order} (30) / 2 = \text{Order} (15) \quad (\text{Eq. 5.3})$$

From Equation 5.3 it is observed that the optimum population size for example 1 reduces to about 15. To remain on the conservative side, a population size of 20 was selected.

To further check the appropriateness of the proposed approach to determining the ideal population size, the algorithm is repeated for different population sizes. The results are shown in Table 5.6, which shows that in this example, a population size of 20 is actually the optimum value resulting in the algorithm converging after 116 structural analyses. It should be noted that the number of structural analyses in Table 5.6 does not include the number of structural analyses needed during the filtration of the local search directions or during the local Tabu search.

Table 5.6 Effect of population size on convergence

Population size	Structural analyses to convergence
10	320
16	220
20	116
24	144
30	240
40	242

EFFECT OF RELIABILITY LEVEL

Unlike Monte Carlo simulation methods, the reliability level of the problem does not affect the speed of convergence of the proposed algorithm. To verify this observation, the mean values of the truss member resistances is increased or decreased by multiplying them by a factor ranging from 0.80 to 1.40. The average number of genetic iterations needed for convergence for each of the cases studied is listed in Table 5.7, which shows that the average number of generations needed to converge is almost the same regardless of the actual value of the mean even if the system reliability level increases from $\beta_{\text{sys}}=2.243$ to $\beta_{\text{sys}}=5.595$.

Table 5.7 Effect of reliability level on convergence

Ratio of new mean to actual mean γ	0.8	1.0	1.2	1.4
Number of generations needed for convergence	6.0	5.8	6.2	6.6
β_{system}	2.243	3.829	4.917	5.595

Further verifications of the efficiency and accuracy of the proposed algorithm are provided in the next sections, which present the solutions of realistic large scale complex structural problems.

5.4 ANALYSIS OF A SIMPLE SUSPENDED CABLE STRUCTURE

The hybrid algorithm proposed in this and the accompanying chapter is used to find the reliability of a geometrically nonlinear elastic structure. The problem solved in this section is adopted from the work of Frangopol and Imai (2000). The problem consists of the symmetric suspended cable structure shown in Figure 5.2. Frangopol and Imai (2000) calculated the reliability index using a mode enumeration scheme by modeling the structure as a series of parallel systems having failure surfaces approximated by hyperplanes (Frangopol and Imai; 2000). In their example, they assumed that the structure is loaded by two random loads P_1 and P_2 applied at the girder. The other random variables that control the reliability of the structure are the yielding stress of the cable, the yielding stress of the hangers, and the yielding stress of the girder. The statistical input data is provided in Table

5.8. The problem first solved by Frangopol & Imai (2000) is repeated in this chapter using the proposed hybrid genetic algorithm linked to the commercial general purpose finite element program ANSYS. The purpose is to verify the capability of the method to solve geometrically nonlinear problems and to demonstrate the ability of linking the algorithm to a general-purpose finite element package. This would enhance the potential use of such a reliability method for solving realistic structural analysis problems. Such capabilities would prove the flexibility of the proposed algorithm and increase the likelihood of future implementation of system reliability techniques in practical structural design and analysis applications.

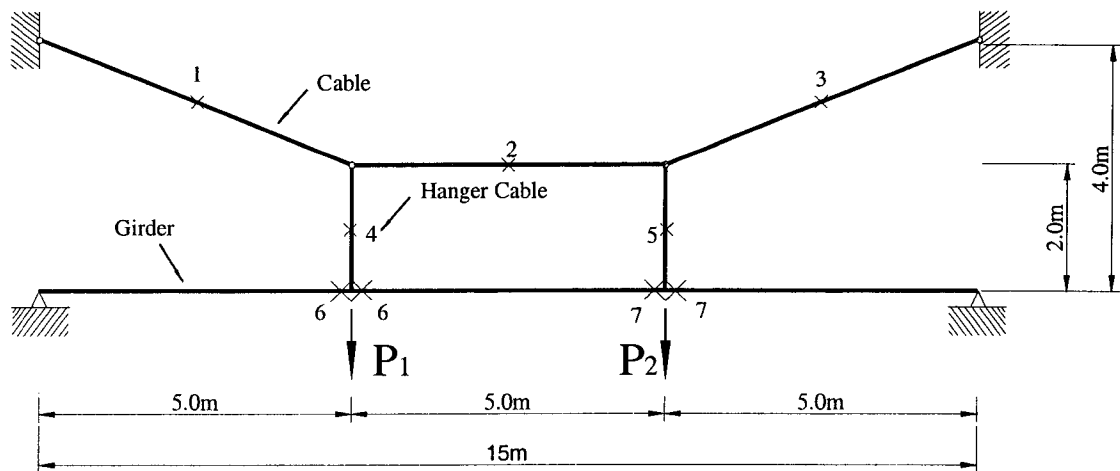


Figure 5.2. Suspended structure (pin-supported): (a) geometry and loading as adapted from Frangopol & Imai (2000)

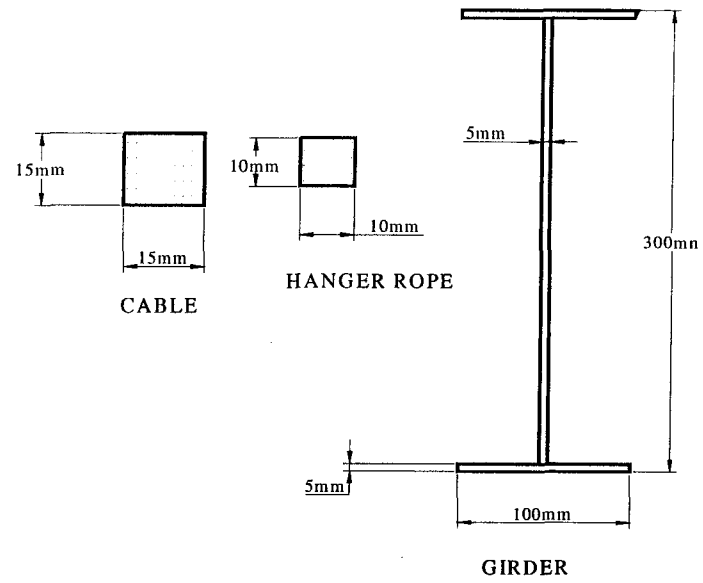


Figure 5.3 Suspended structure: (b) cross-sections: cable, hanger rope, and girder as adapted from Frangopol & Imai (2000)

Table 5.8 Random variables of suspended structure

Random Variables	Mean	C.O.V	Distribution Type
Load P_1 (KN)	50	0.3	Normal
Load P_2 (KN)	50	0.3	Normal
Yield Stress (cable) σ_y (Mpa)	1500	0.05	Normal
Yield Stress (hanger) σ_y (Mpa)	1000	0.05	Normal
Yield Stress (girder) σ_y (Mpa)	250	0.05	Normal

The structural system consists of one main suspension cable divided into three elements. The cable carries a single girder supported by two hanger ropes. The cross-sectional area of the cable is deterministic ($A_c=225\text{mm}^2$). The same cross sectional area $A_h = 100\text{mm}^2$ is assumed for the two hanger ropes. The girder's area, A_g is $2,450\text{mm}^2$ with the cross section shown in Figure 5.3. The deterministic modulus of elasticity of the cable and hanger ropes is assumed to be the same, $E=150\text{GPa}$. The cross-section of the girder has the moment of inertia $I_z = 31,920,920 \text{ mm}^4$ and the plastic section modulus $Z_p = fI_z/(h/2)$, where f , the ratio of the fully plastic moment to the yield moment, is equal to 1.167. The girder is pin-supported at both ends and its modulus of elasticity is $E = 200 \text{ GPa}$. The statistics of the random variables considered in the reliability analysis are listed in Table 5.8. All the variables are assumed to follow normal distributions. The resistances of the girder cross-sections at points 6 and 6', and 7 and 7' are assumed to be fully correlated. These are the same values assigned in the example solved by Frangopol & Imai (2000).

During their enumeration, Frangopol & Imai (2000) defined the system failure of the structure using three different criteria as follows:

- Criterion A: failure of any of the seven cross-sections;
- Criterion B: failure of any two of cross-sections 4–7;
- Criterion C: failure of any three of cross-sections 4–7.

For the geometrically nonlinear elastic behavior, based on First-Order Reliability Methods (FORM), Frangopol & Imai (2000) obtained a reliability index value $\beta=2.054$ for Criterion A. The reliability index $\beta = 5.188$ and $\beta = 5.438$ are calculated for Criteria B, C respectively.

To study the possibility of solving the same problem using the proposed hybrid algorithm, the general structural Analysis software ANSYS 7.0 is applied to model the failure of the same suspension bridge. The same three failure criteria used by Frangopol & Imai (2000) are adopted to verify the accuracy of the proposed algorithm. The results for criterion A, B and C obtained from the Genetic Algorithm are $\beta=2.06$, 5.214 & 5.527 respectively which are reasonably close to the values represented by Frangopol & Imai (2000). Furthermore, an additional criterion defined as the collapse of the structure is also introduced. Theoretically, collapse is observed when as a consequence of an incremental increase in the applied load, the displacement of the structure goes to infinity. This occurs when the stiffness matrix becomes singular. In the cases where a numerical value is obtained from the program, infinity is assumed to coincide to be 100-fold (or larger) increase in the displacement when the load is increased by ten percent.

The final results of the reliability analysis after completing the refined Tabu local search are compared with those of Frangopol and Imai (2000). The results are listed in Table 5.9, which once again prove the accuracy of the proposed algorithm. Convergence was reached after about 70 structural evaluations at which time the algorithm identified three important failure modes. For the criterion “collapse of the structure”, the direction cosines and the reliability index values for this case are listed in Table 5.10.

Table 5.9 Comparison of genetic algorithm results to available values

System Reliability Index	Frangopol & Iami (2000)	Proposed Algorithm
Criterion A	2.054	2.060
Criterion B	5.188	5.214
Criterion C	5.438	5.517
Collapse of Structure	--	5.232

Note: pin-supported geometrically nonlinear ductile system

Table 5.10 Final search results for suspended structure

Failure Mode	Direction Cosines					β
	1	2	3	4	5	
1	0.572	0.668	-0.075	-0.471	0	5.266
2	0.529	0.609	0.139	-0.53	-0.22	5.418
3	0.751	0.66	0	0	0	5.848

If all three failure modes are considered, the probability of failure of the system is obtained as:

$$\begin{aligned}
 P_f(Z_1 \cup Z_2 \cup Z_3) &= P_f(Z_1) + P_f(Z_2) + P_f(Z_3) \\
 &\quad - P_f(Z_1 \cap Z_2) - P_f(Z_1 \cap Z_3) - P_f(Z_2 \cap Z_3) + P_f(Z_1 \cap Z_2 \cap Z_3) \\
 &= \Phi(-\beta_1) + \Phi(-\beta_2) + \Phi(-\beta_3) \\
 &\quad - \Phi(-\beta_1, -\beta_2, \rho_{1,2}) - \Phi(-\beta_1, -\beta_3, \rho_{1,3}) - \Phi(-\beta_2, -\beta_3, \rho_{2,3}) + \Phi(-\beta_1, -\beta_2, -\beta_3, \rho_{1,2,3}) \\
 &= 8.53 \cdot 10^{-8}
 \end{aligned} \tag{Eq. 5.4}$$

Resulting in a system reliability index β_{sys} equal to:

$$\beta_{\text{sys}} = -\Phi^{-1}(P_f) = 5.226 \tag{Eq. 5.5}$$

If only the first and second failure modes are considered, the probability of failure and the system reliability index become:

$$\begin{aligned}
 P_f(Z_1 \cup Z_2) &= P_f(Z_1) + P_f(Z_2) - P_f(Z_1 \cap Z_2) \\
 &= \Phi(-\beta_1) + \Phi(-\beta_2) - \Phi(-\beta_1, -\beta_2, \rho_{1,2}) \\
 &= 8.39 \cdot 10^{-8}
 \end{aligned} \tag{Eq. 5.6}$$

$$\beta_{\text{sys}} = -\Phi^{-1}(P_f) = 5.232 \tag{Eq. 5.7}$$

If only the first failure mode is considered, the results become:

$$P_f = 6.98 \cdot 10^{-8}, \beta_{\text{sys}} = 5.266 \quad (\text{Eq. 5.8})$$

This shows that the contribution of the third failure mode to the total probability of failure is less than 5%.

The example solved in this section demonstrates that the proposed algorithm provides an accurate and efficient method for solving practical structural reliability problems. A very important feature is the ability of linking the algorithm with available finite element packages to facilitate its use by practicing engineers and use realistic models of the structures being analyzed.

5.5 ANALYSIS OF CABLE-STAYED BRIDGE EXAMPLE

The hybrid algorithm proposed in this and the preceding chapter is finally applied to demonstrate the capability of the proposed algorithm to identify the important failure mode and the system reliability of a large geometrically nonlinear structure. The realistic model of a cable-stayed bridge consists of one support tower and the symmetric cable stayed system shown in Figure 5.4. The steel box girder has the cross section shape shown in Figure 5.5. The dimensions of the girder, however, vary so that the girder sectional properties can be divided into three categories as listed in Table 5.11. The deterministic modulus of elasticity of the three girders is assumed to be the same, $E = 2.1 \cdot 10^5$ MPa, and the yielding stress is $\sigma_y = 345$ MPa. The total length of the bridge is 1024 meters. The height of the main tower is 269.95 meters as shown in Figure 5.6. In this analysis, since the tower is very stiff

compared to the other members, the components of the tower are modeled as rigid elements. There are total 35 stay cables whose design tension stress is 1860MPa and modulus of elasticity $E=1.95 \times 10^5$ MPa. The typical span between the cables is 15 meters except for the leftmost part of the bridge, where the cables are spaced 9 meters apart.

Table 5.11 Geometric properties of main girder of cable stayed bridge

	Self Weight (KN/M)	Area (m ²)	I _x (m ⁴)	I _y (m ⁴)
Section 1	215.6	2.544	10.164	261.748
Section 2	223.4	2.278	9.078	226.203
Section 3	245.0	1.743	6.755	168.786

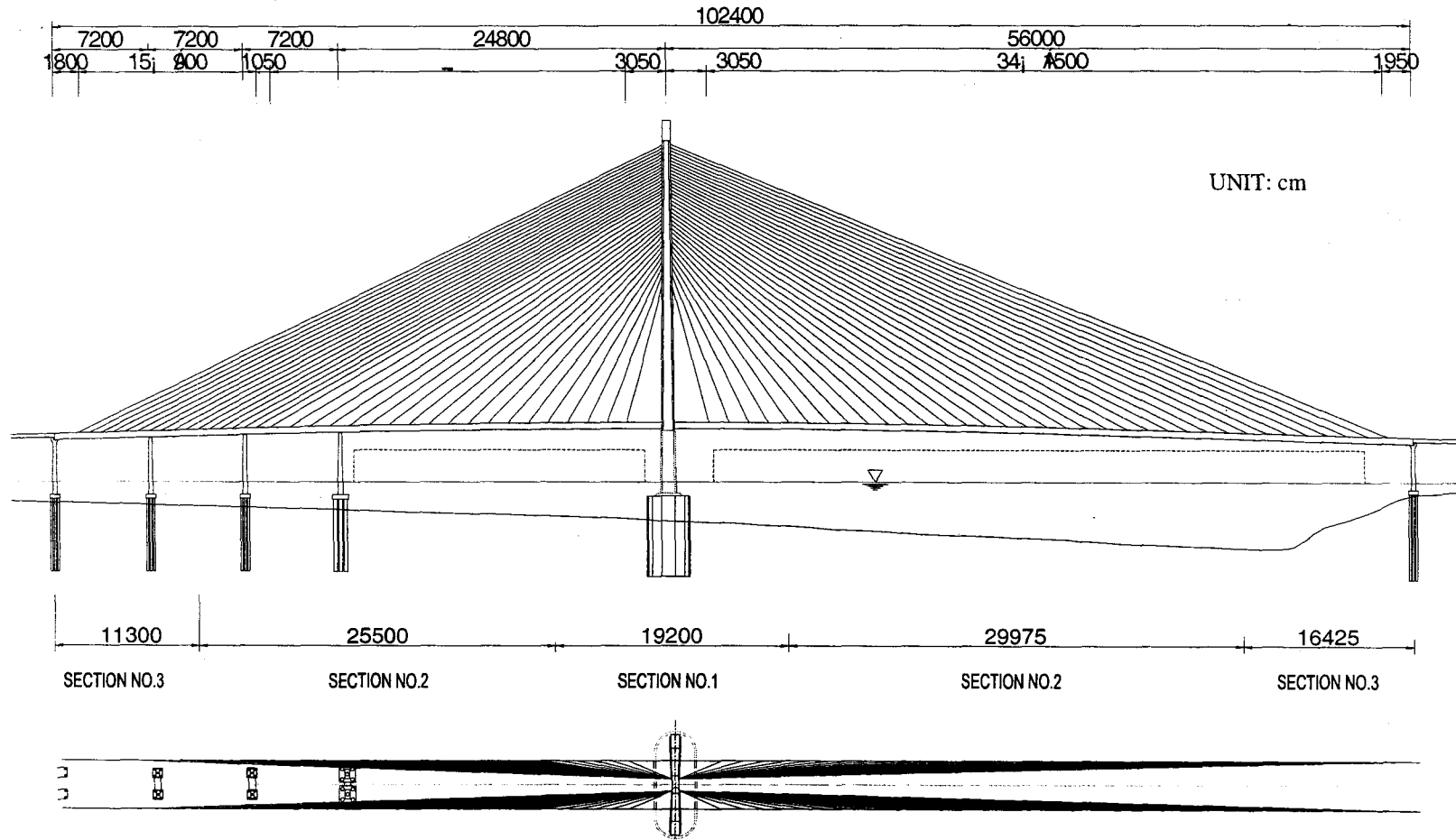


Figure 5.4. Profile of cable stayed bridge

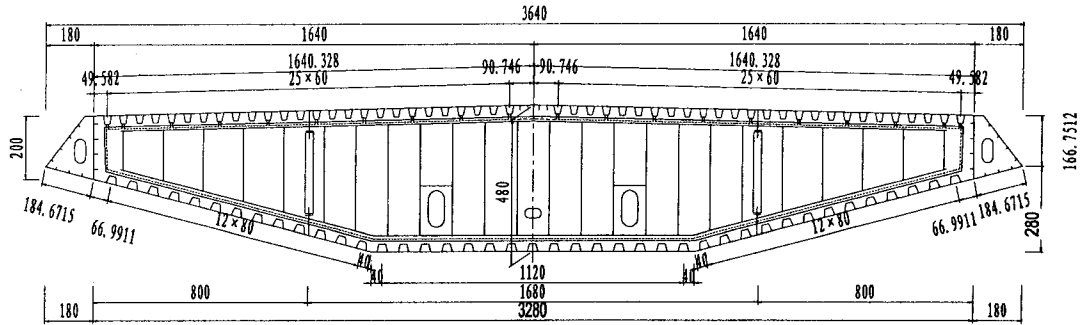


Figure 5.5. Typical cross section of main box girder

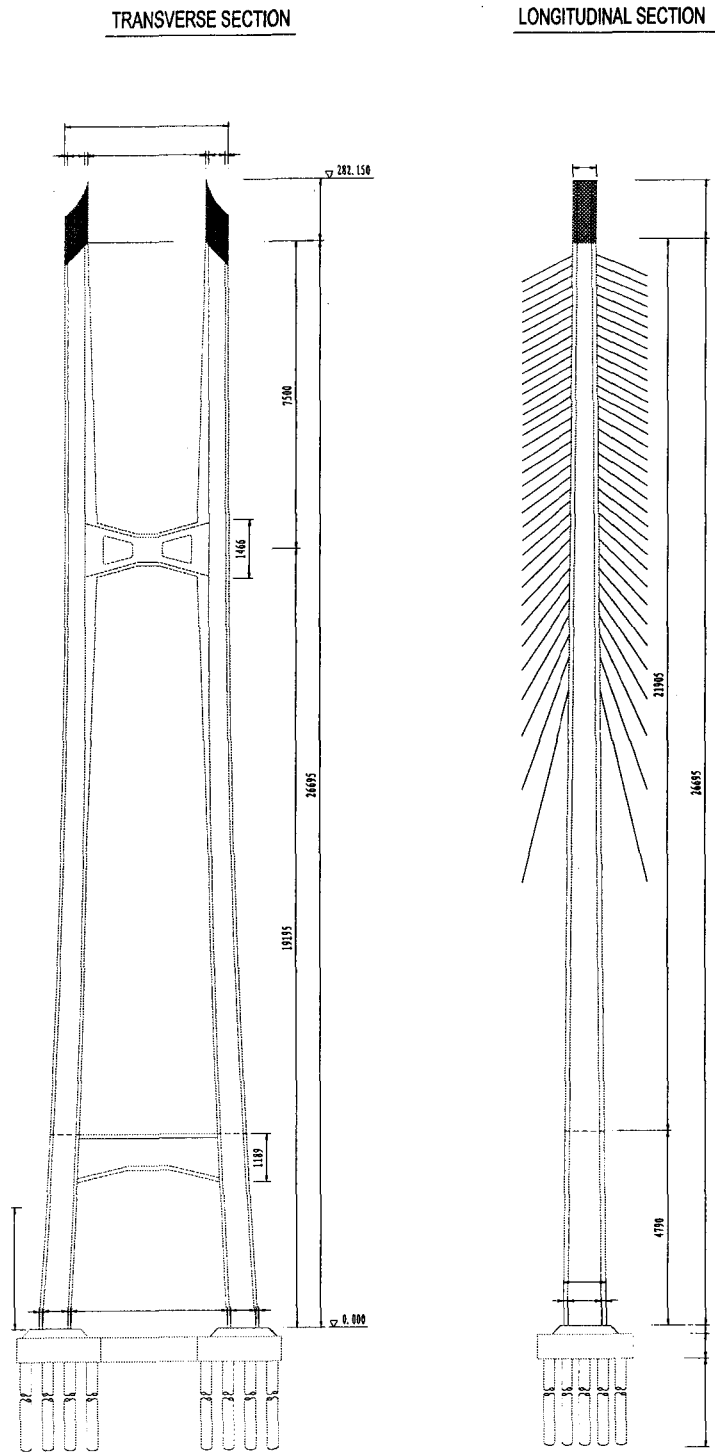


Figure 5.6 Support tower of cable stayed bridge

The random variables considered in this example include the live load, dead load, moment capacities of the three different cross sections of the main box girder, as well as the strengths of the 35 stay cables. The nominal values, biases, COV's and distribution types of all the variables are listed in Table 5.12. The biases and COV's of the live and dead loads are the same as those used by Imai & Frangopol (2002) who also assumed that these variables follow extreme type I, and normal distributions respectively. In Table 12, assuming structural degradation of the stay cables, the biases of all cables' strengths are assumed to be 0.74 for the 75-year service life. The structural reliability of service life is a time dependent problem. However, to simplify the problem, it is assumed that the biases of cable strengths are constant over the whole service life. All 40 random variables are assumed to be statistically independent.

Table 5.12 Random variables of cable stayed bridge

Variable Number	Variables	Nominal Value	Bias	COV	Distribution Type
1	Live Load	129.78KN/M	1.05	0.2	Extreme Type I
2	Dead Load	See Table 11	1.03	0.08	Normal
3	M _{cap} of Main Girder Section 1	1,353,456 KN-M	1.05	0.11	Normal
4	M _{cap} of Main Girder Section 2	1,205,740 KN-M	1.05	0.11	Normal
5	M _{cap} of Main Girder Section 3	889,777 KN-M	1.05	0.11	Normal
6	Yield Strength Cable # 1 (KN)	16,665.6	0.74	0.2	Normal
7	Yield Strength Cable # 2 (KN)	13,020.0	0.74	0.2	Normal
8	Yield Strength Cable # 3 (KN)	13,020.0	0.74	0.2	Normal
9	Yield Strength Cable # 4 (KN)	18,228.0	0.74	0.2	Normal
10	Yield Strength Cable # 5 (KN)	18,228.0	0.74	0.2	Normal
11	Yield Strength Cable # 6 (KN)	18,228.0	0.74	0.2	Normal
12	Yield Strength Cable # 7 (KN)	18,228.0	0.74	0.2	Normal
13	Yield Strength Cable # 8 (KN)	18,228.0	0.74	0.2	Normal
14	Yield Strength Cable # 9 (KN)	19,269.6	0.74	0.2	Normal
15	Yield Strength Cable # 10 (KN)	20,311.2	0.74	0.2	Normal
16	Yield Strength Cable # 11 (KN)	21,352.8	0.74	0.2	Normal
17	Yield Strength Cable # 12 (KN)	22,394.4	0.74	0.2	Normal
18	Yield Strength Cable # 13 (KN)	23,956.8	0.74	0.2	Normal
19	Yield Strength Cable # 14 (KN)	26,560.8	0.74	0.2	Normal
20	Yield Strength Cable # 15 (KN)	26,560.8	0.74	0.2	Normal
21	Yield Strength Cable # 16 (KN)	26,560.8	0.74	0.2	Normal
22	Yield Strength Cable # 17 (KN)	26,560.8	0.74	0.2	Normal
23	Yield Strength Cable # 18 (KN)	27,081.6	0.74	0.2	Normal
24	Yield Strength Cable # 19 (KN)	27,081.6	0.74	0.2	Normal
25	Yield Strength Cable # 20 (KN)	31,248.0	0.74	0.2	Normal
26	Yield Strength Cable # 21 (KN)	31,248.0	0.74	0.2	Normal
27	Yield Strength Cable # 22 (KN)	31,248.0	0.74	0.2	Normal
28	Yield Strength Cable # 23 (KN)	28,644.0	0.74	0.2	Normal
29	Yield Strength Cable # 24 (KN)	31,768.8	0.74	0.2	Normal
30	Yield Strength Cable # 25 (KN)	31,768.8	0.74	0.2	Normal
31	Yield Strength Cable # 26 (KN)	31,768.8	0.74	0.2	Normal
32	Yield Strength Cable # 27 (KN)	31,768.8	0.74	0.2	Normal
33	Yield Strength Cable # 28 (KN)	31,768.8	0.74	0.2	Normal
34	Yield Strength Cable # 29 (KN)	31,768.8	0.74	0.2	Normal
35	Yield Strength Cable # 30 (KN)	31,248.0	0.74	0.2	Normal
36	Yield Strength Cable # 31 (KN)	31,248.0	0.74	0.2	Normal
37	Yield Strength Cable # 32 (KN)	31,248.0	0.74	0.2	Normal
38	Yield Strength Cable # 33 (KN)	30,727.2	0.74	0.2	Normal

39	Yield Strength Cable # 34 (KN)	30,727.2	0.74	0.2	Normal
40	Yield Strength Cable # 35 (KN)	30,727.2	0.74	0.2	Normal

In a first step of the analysis, all the resistance variables are multiplied by a reduction factor equal to 0.3. This is done to reduce the probability of failure, P_f , so that the analysis can be performed using classical Monte Carlo Simulation. The failure equation can then be roughly represented by the equation:

$$Z = 0.3R - S \quad (\text{Eq. 5.9})$$

Applying the Monte Carlo Simulation direct sampling option in ANSYS 7.0 with 300 cycles, the probability of failure and the system's reliability index were found to be:

$$P_f = 1.93 \cdot 10^{-1} \quad (\text{Eq. 5.10})$$

$$\beta_{\text{sys}} = 0.866 \quad (\text{Eq. 5.11})$$

Using the proposed Data Mining/Genetic Algorithm, the failure modes are identified, and the system reliability index as well as the probability of structural failure are obtained as:

$$P_f = 1.87 \cdot 10^{-1} \quad (\text{Eq. 5.12})$$

$$\beta_{\text{sys}} = 0.889 \quad (\text{Eq. 5.13})$$

In this case, the total number of structural analyses required for convergence happened to be equal to 300. Comparing the results of the proposed algorithm with those of the Monte Carlo Simulation, it is noticed that the proposed algorithm provides reasonably accurate results. The analysis was stopped after identifying the first three failure modes.

Improved accuracy can still be obtained if the simulation is continued so that a larger number of modes are identified. Another reason for the difference is the assumption that all failure modes can be approximated by first order equations. For this particular problem, this is not necessarily the case.

Ignoring the 0.3 reduction factor used in the previous test case, the reliability analysis of the cable-stayed bridge is performed using the proposed algorithm with the original data of Table 5.12. The analysis was stopped after the three most important failure modes are identified. If L represents the live load, D represents the dead load, G_i represents the resistance of girder section i and C_j represents the resistance of cable j , the most critical failure equation, Z_1 , and the reliability index, β_{z1} , are given as follows:

$$Z_1 = 3.374G_1 + 2.4G_2 + 1.7G_3 + 0.79C_1 + 0.451C_3 + 0.72C_7 + 1.026C_{10} + 1.227C_{12} + 1.082C_{14} + 1.381C_{17} + 0.473C_{21} + 0.711C_{26} + 1.284C_{30} + 1.085C_{33} + 0.85C_{34} + 0.741C_{35} - 1.421L - 10.89 \quad (\text{Eq. 5.14})$$

$$\beta_{z1} = 4.791 \quad (\text{Eq. 5.15})$$

Some valuable information can be obtained from this dominant failure mode. For example, the dead load is found not to be a critical parameter that would affect the system reliability. This is most probably due to its low COV. From a design optimization point of view, the cables that don't show in this failure mode can be considered to have been overdesigned. More sensitivity analyses can be performed to analyze different girder box sections, different load conditions, and cable strength to obtain a more efficient design.

For the second failure mode, the failure equation and reliability index are obtained as:

$$Z_2 = 2.646G_1 + 2.292G_2 + 1.392C_9 + 1.986C_{12} + 1.859C_{15} + 1.219C_{17} + 1.356C_{21} + 1.158C_{31} + 0.83C_{32} + 0.81C_{35} - 1.336L - 1.427D - 4.592 \quad (\text{Eq. 5.16})$$

$$\beta_{Z2} = 5.379 \quad (\text{Eq. 5.17})$$

For the third failure mode, the failure equation and reliability are:

$$Z_3 = 1.223G_1 + 2.694G_2 + 0.45C_4 + 0.535C_7 + 0.615C_{12} + 0.445C_{13} + 1.836C_{15} + 0.743C_{16} + 1.866C_{17} + 1.347C_{20} + 1.193C_{21} + 1.018C_{22} + 0.572C_{26} + 0.648C_{28} + 1.417C_{32} - 1.089L - 6.822 \quad (\text{Eq. 5.18})$$

$$\beta_{Z3} = 5.931 \quad (\text{Eq. 5.19})$$

The system reliability when all three failure modes are considered is:

$$\beta_{\text{sys}} = 4.7819 \quad (\text{Eq. 5.20})$$

$$P_f = 8.68 \times 10^{-7} \quad (\text{Eq. 5.21})$$

When only the two failure modes are considered, the system reliability becomes:

$$\beta_{\text{sys}} = 4.7820 \quad (\text{Eq. 5.22})$$

$$P_f = 8.68 \times 10^{-7} \quad (\text{Eq. 5.23})$$

This shows that in this example, the first failure mode is dominant and the system's reliability can be well estimated when the first two modes are considered.

The long span cable-stayed bridge example solved in this section demonstrates the accuracy, efficiency and high flexibility of the proposed algorithm and its ability to be combined with general-purpose finite element programs to perform the reliability analysis of

complex structural system. As briefly mentioned above, the results of such analyses can be used to optimize the design of such structures and identify the most critical failure modes.

5.6 CONCLUSIONS

This chapter solved several structural reliability problems to illustrate the applicability of the hybrid Data Mining/Genetic Shredding Algorithm proposed in the preceding chapter. The algorithm is tested to verify its robustness. Sensitivity analyses are performed to investigate the parameters that affect the convergence speed and accuracy of the proposed procedure. Combined with a general-purposed finite element program, the proposed algorithm is successfully applied to solve reliability problems of various structural systems considering typical geometrically nonlinear elastic behavior or plastic behavior. Furthermore, the solutions of a complex cable stayed bridge example is provided to demonstrate the applicability of the proposed algorithm for estimating the system reliability of complex structural systems. Due to its flexibility, it is expected that the implementation of algorithms similar to the one proposed in this chapter will be helpful in increasing the application of reliability methods in the process of structural design, rehabilitation, and strengthening.

CHAPTER SIX

CONCLUSIONS AND PROPOSED FURTHER RESEARCH

6.1 CONCLUSIONS

The goal of this Ph.D. Dissertation was to develop efficient and accurate simulation methods to evaluate the reliability of complex structural systems. To achieve the goals of this research, two hybrid genetic search algorithms have been proposed. The first method is based on the linkage learning process described by Kargupta (1995), while the second is based on the principle of data mining. The main achievements of the study include:

- A modified hybrid linkage-learning/shredding operator algorithm is developed to join the benefits of the linkage operators first proposed by Kargupta (1995) to those of the shredding operator proposed by Deng (2000). The algorithm is able to identify multiple equally important failure modes and find their reliability indexes.
- New efficient and robust convergence criteria applicable to reliability problems have been proposed to improve the convergence to local optima as well as the global optimum.
- The efficiency of genetic search algorithms has been further improved by proposing a new hybrid Data Mining/Genetic Shredding Algorithm. The method is based on the Apriori algorithm originally proposed by Agrawal (1994) and modified for the identification of different structural failure modes in complex structural systems. A new linkage learning process is developed based on the Apriori algorithm to explore the

standard normalized random variable space for “good” search directions. The proposed Data Mining/GA algorithm is shown to be efficient reducing the number of operations needed to reach convergence.

- A new filtering criterion based on the general principle of structural failure mechanisms is proposed to identify the independent failure modes.
- A Tabu local search process is added to the hybrid algorithms to refine the search and find accurate solutions that may not necessarily lie in the directions that were pre-set during the original genetic coding.
- The hybrid GA algorithm was extended to help identify the most likely failure modes and identify the variables and sub-mechanisms that contribute the most to the failure of structural systems.
- Sensitivity analyses were performed to study the effect of various parameters on the robustness of the proposed algorithms and their speed of convergence.
- The GA procedure was generalized to solve time dependent reliability problems including deteriorating structural systems by discretizing the life of the structure into several time intervals.
- Illustrative examples and benchmark test problems have been solved to verify the accuracy and efficiency of the proposed method.
- The proposed algorithm was combined with a general-purpose finite element program to find the important failure modes of two-geometrically nonlinear structural cable systems. The results demonstrated the suitability of the proposed methods for practical implementation in engineering practice

6.2 FURTHER RESEARCH

This study developed hybrid genetic search algorithms for identifying the multiple failure modes of complex structural systems. The new methods are developed to combine the benefits of the pattern identification ability of Data Mining techniques to the capacity of the Shredding GA operators to explore new significant search domains. The results obtained confirmed the efficiency and accuracy of the proposed methods for application during the reliability analysis of complex structural systems. The long-term benefits of the proposed techniques will help engineers design new structural systems to meet optimum levels of safety and reliability. The methods can also help in the reliability assessment of existing structures to guide the inspection and rehabilitation process. More work is needed to extend the range of the application of the proposed methods. Specifically, the following additional tasks could be performed in the future:

- In the current study, the limit state functions were assumed to be linear in order to simplify the process of calculating the probability of failure P_f . However, generally speaking, the limit state functions are not planar in the normalized random variable, U , space nor are they independent. The nonlinearity of the limit state functions can be accounted for by using higher order methods based on SORM. Simultaneously, the work of Melchers (2001) can be extended to further improve the estimation of the joint probability of failure between piecewise linear segments of the limit states by exploring the intersection between the discretized failure mode segments.

- The efficiency of the structural analysis during the search process can be drastically improved by developing efficient re-analysis techniques that can use the information obtained from previous steps of the search process. For example, at each step of Tabu local search, only a limited number of structural members are changed. The solver developed by Deng and Ghosn (2001) proved to provide a dramatic improvement of computational efficiency for structural redesign and optimization problems by reducing the computational effort by a factor 4. The application of such algorithms in the reliability analysis would drastically improve the efficiency of the reliability analysis to manageable levels so that real world problems with large numbers of random variables can be solved on desktop computers.
- Future research should also demonstrate how the proposed methods can be Applied to improve the structural design and assessment of complex structural systems. The objective would be to optimize the design process using reliability constraints and provide guidelines for advanced performance-based design techniques based on reliability principles.
- In previous research, it has been found that the solution of random vibration problems using FORM methods are difficult to implement due to the inefficiencies of existing algorithms. For example, Der Kiureghian (1996) investigated the geometry of random vibration problems in the space of standard normal random variables by discretizing the input process. His research showed that for linear systems subjected to Gaussian or non-Gaussian excitations, the problems of interest can be characterized by linear or non-linear geometric forms, such as vectors, planes, half spaces, wedges and ellipsoids. Approximate solutions for such problems could be obtained by first- and second-order

reliability methods (FORM and SORM). Through discretizing the service life of the structure into several time intervals, the failure probability of a time-dependant problem can be estimated with high accuracy. However, the process is found to be extremely time-consuming. Future research relying on the methods developed in this dissertation may lead to improving the efficiency of time dependent analyses by using hybrid genetic algorithms.

BIBLIOGRAPHY

- Agrawal, R. and Srikant, R. (1994). "Fast Algorithms for Mining Association Rules in Large Databases." VLDB 1994, 487-499.
- Bandyopadhyay, S. & Kargupta, H. & Wang, G. (1998) Revisiting the GEMGA: Scalable Evolutionary Optimization Through Linkage Learning. Proceedings of the IEEE International Conference on Evolutionary Computation. IEEE Press. p.603-608.
- Battiti, R. and Tecchiolli, G. (1994) Simulated annealing and Tabu search in the long run: A comparison on QAP tasks. Computer & Mathematics with Applications. Volume 28. Issue 6.
- Becquet, C. and Balachon, S. (2002) Strong-association-rule mining for large-scale gene-expression data analysis: a case study on human SAGE data. Genome Biology 3(12).
- Bennett, R.M. and Ang. A. H.-S. (1983) Investigation of methods for structural system reliability, structural research series, No. 510, University of Illinois, Urbana, IL.
- Bjerager, P. (1988) Probability integration by directional simulation, J. Engineering Mechanics, ASCE, Vol. 114, No. 8, 1285-1302.
- Borgelt, C. and Kruse, R. (2002). "Induction of Association Rules: Apriori Implementation." The 14th Conference on Computational Statistics. Compstat 2002, Berlin, Germany
- Borri, A. and Speranzini, E. (1997) Structural reliability analysis using a standard deterministic finite element code, Structural Safety, Volume 19, Issue 4, Pages 361-382
- Breitung, K. and Rackwitz, R. (1982) Nonlinear combination of load processes. J Struct Mech 10 2, pp. 145-166
- Bruneau, M., (1992) Evaluation of System-Reliability Methods for Cable-Stayed Bridge Design, Journal of Structural Engineering, Vol. 118, No. 4, pp. 1106-1120
- Deng, L., Ghosn, M., and Shao, S. (2004). "Shredding Genetic Algorithm for Reliability Analysis of Structural Systems." accepted by Journal of Structural Safety
- Deng, L. and Ghosn, M., (2001). "Pseudoforce Method for Nonlinear Analysis and Reanalysis of Structural Systems" Journal of Structural Engineering, ASCE, Vol. 127 (5), pp. 570-578
- Deng, L.Z. (2000) Artificial Intelligence Techniques for Reliability Assessment of Nonlinear Bridge Systems. Ph.D dissertation, the City University of New York, New York, NY.

- Der Kiureghian, A., (1996) Structural reliability methods for seismic safety assessment: a review, *Engineering Structures*, Volume 18, Issue 6, Pages 412-424
- Dey, A. and Mahadevan, S. (1998) Ductile structural system reliability analysis using adaptive importance sampling *Structural Safety*, Volume 20, Issue 2, Pages 137-154
- Ditlevsen, O. (1979) Narrow reliability bounds for structural systems. *J. Struct. Mech.*, Vol. 7. pp.435-451.
- Ditlevsen, O. (1982) System Reliability Bounding by Conditioning, *Journal of the Engineering Mechanics Division*, Vol. 108, No. 5, September/October 1982, pp. 708-718
- Ditlevsen, O. and Bjerager, P. (1984) Reliability of highly redundant plastic structures, *J. Engineering Mechanics*, ASCE, vol. 110, no.5, 671-693
- Ditlevsen, O. and Bjerager, P. (1989) Plastic reliability analysis by directional simulation, *J. Engineering Mechanics*, ASCE, vol. 115, no.6, 1347-1362.
- Ellingwood, B. (2000) LRFD: implementing structural reliability in professional practice. *Engr. Struct.* 22(2) pp.106-115.
- Engelund, S. and Rackwitz, R. (1993) A benchmark study on importance sampling techniques in structural reliability, *Structural Safety*, Vol. 12, No. 4, 255-276.
- Enright, M. and Frangopol, D. (1998) Probabilistic analysis of resistance degradation of reinforced concrete bridge beams under corrosion. *Engineering structures*, Vol. 20(11). pp. 960-971
- Frangopol, D.M. and Imai, K. (2000) Geometrically nonlinear finite element reliability analysis of structural systems. II: applications. *Computers and Structures* 77 6, pp. 693-709.
- Frangopol, D.M., (1985) Sensitivity studies in reliability based analysis of redundant structures, *Structural safety*, Vol. 3, No. 1, 13-22.
- Ghosn, M., Moses, F. and Wang, J. (2002) Design of highway bridge for extreme events. Report to the NCHRP.
- Glover and Laguna. *Tabu Search*, Kluwer Academic Publishers, Boston, ISBN 0-7923-9965-X, 1997.
- Goldburg, D.E., (1989) *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company, Inc., Massachusetts.
- Gomes, H.M. and Awruch, A. M. (2004) Comparison of response surface and neural network with other methods for structural reliability analysis. *Journal of Structural Safety*. Vol.26. pp. 49-67

- Gorman, Mark. R. (1979) Reliability of structural systems. Report No. 79-2. Department of Civil Engineering, Case Western Reserve University, Cleveland, Ohio.
- Grandhi, R. and Liping, W. (1998) Reliability-based structural optimization using improved two-point adaptive nonlinear approximations, *Finite Elements in Analysis and Design*, Volume 29, Issue 1, Pages 35-48.
- Guan X. L. and Melchers, R. E. (1999) A load space formulation for probabilistic finite element analysis of structural reliability, *Probabilistic Engineering Mechanics*, Volume 14, Issues 1-24, Pages 73-81
- Guan, X. L. and Melchers, R. E. (2001) Effect of response surface parameter variation on structural reliability estimates. *Structural Safety*. Vol.23. Issue 4. pp.429-444.
- Harik, G.R., Goldberg, D.E.(1996) Learning linkage. *Foundations of Genetic Algorithms Workshop 4* pp.247-262.
- Hohenbichler, M. and Rackwitz, R. (1983) First-order concepts in systems reliability, *Structural Safety*, Vol. 1, No. 3 177-188.
- Imai K. and Frangopol, D.M. (2000) Geometrically nonlinear finite element reliability analysis of structural systems. I: theory, *Computers & Structures*, Volume 77, Issue 6, Pages 677-691
- Imai, K. and Frangopol, D.M. (2002) System Reliability of Suspension Bridges. *Structural Safety*. Volume 24, Issues 2-4. pp. 219-259
- Kaisheng, C., Shengkun, Z., and Wei, H. (1996) Artificial intelligence \square -unzipping method in structural system reliability analysis, *Computers & Structures*, Volume 60, Issue 4, Pages 619-626
- Karamchandani, A. (1987), "Structural System Reliability Analysis Methods." Report No. 83, Department of Civil Engineering, Stanford University.
- Karamchandani, A., and Cornell, C.A. (1992) Reliability Analysis of Truss Structures with Multistate Elements. II, *Journal of Structural Engineering*, Vol. 118, No. 4, pp. 910-925
- Kargupta, H. and Bandyopadhyay, S. A Perspective On the Foundation and Evolution of the Linkage Learning Genetic Algorithm. Special Issue on Genetic Algorithms, the *Journal of Computer Methods in Applied Mechanics and Engineering (CMAME)*. Guest Eds: David E. Goldberg and Kalyanmoy Deb. 186, 1999. p. 266-294.
- Kargupta, H., Sanseverino, E.R., Johnson, E., and Agrawal, S. (1998) the genetic algorithm, linkage learning, and scalable data mining.

- Kargupta, K. H. and Bandyopadhyay, S. Further Experimentations on the scalability of the GEMGA. Lecture Notes in Computer Science. Springer Verlag. Vol. 1498. 1998. p.315-324.
- Katsuki S. and Frangopol, D.M., (1994) Hyperspace Division Method for Structural Reliability, Journal of Engineering Mechanics, Vol. 120, No. 11, pp. 2405-2427
- Katsuki, S., Fukawa, G. and Nagadori, N., (2001) An Application of Genetic Algorithm for Interactive Optimal Design Method of Truss Structures, ICOSSAR'01, Newport Beach CA
- Kiureghian, A. D. (2000) The geometry of random vibrations and solutions by FORM and SORM. Prob. Eng. Mech. 15 , pp. 81–90
- Kiureghian, A.D. and Li,C-C. (1996) Nonlinear random vibration analysis through optimization. Proc. Of the 7th IFIP WG 7.5 Conference Mechanics Divisions,ASCE 96(4), pp.483-493
- Lemonick, M.D. and Thompson, D, (1999), “Racing to Map Our DNA”, Time Magazine, January 11, 44-50.
- Lin, T.S. and Corotis, R.B. (1985) Reliability of Ductile Systems with Random Strengths, Journal of Structural Engineering, Vol. 111, No. 6, June 1985, pp. 1306-1325
- Liu, P.L. and Der Kiureghian, A. (1991) Finite element reliability of geometrically nonlinear uncertain structures, J. Engineering Mechanics, ASCE, Vol. 117 No. 8 1806-1825
- Maymon, G. (1993), Probability of failure of structures without a closed-form failure function. Computers and Structures 49 2, pp. 301-313.
- Melchers, R. E. and Ahammed, M.(2001), Estimation of failure probabilities for intersections of non-linear limit states. Structural Safety Volume 23, Issue 2, pp.123-135
- Melchers, R.E., (1999) "Structural Reliability: Analysis and Prediction", John Wiley & Sons, New York, NY.
- Moses, F. (1982), “System Reliability Developments in Structural Engineering,” Structural Safety, 1(1), 3-13.
- Nikolaos, D. and Lagaros (2002) Structural optimization using evolutionary algorithms, Computers & Structures. Vol. 80. Issue 7-8.
- Papadrakakis, M., Lagaros, N.D., Thierauf, G. and Cai, J.(1998) Advanced solution methods in structural optimization based on evolution strategies. J Eng Comput 15 1, pp. 12–34.
- Rackwitz, R. (2001) Reliability analysis—a review and some perspectives, Structural safety. Vol. 23. Issue 4.

Robinson, D.G., and Toussaint, G., (1993), "Reliability of Unidirectional Composites Using Genetic Algorithms", Probabilistic Structural Mechanics: Advances in Structural Reliability, IUTAM Symposium San Antonio/USA, Springer-Verlag, 455-465.

Shao, S. and Murotsu, Y. (1999) "Approach to Failure Mode Analysis of Large Structures." Probabilistic Engineering Mechanics. Vol 14. 169-177.

Thoft-Christensen, P. and Baker, M. (1982) Structural Reliability Theory and its Applications, Springer Verlag, Berlin, Germany.

Thoft-Christensen, P., Murotsu, Y. (1986). "Application of Structural System Reliability Theory." Spring-Verlag Berlin Heidelberg New York.

Thompson, D. (1999) Gene Maverick, Time Magazine, January 11, pp.54-55. US Dept. of Energy, Human Genome Program Report.

Tsutsui, Shigeyoshi, Ghosh, Ashish, Corne, David and Fujimoto, Yoshiji. A Real Coded Genetic Algorithm with an Explorer and an Exploiter Populations. Proceedings of the 7th International Conference on Genetic Algorithms (ICGA-97). 1997. p.238-245.

Turkstra, C.J. and Madsen, H. (1980) Load Combinations in Codified Structural Design, ASCE, J. of Structural Engineering, Vol. 106 No. 12, pp. 2527-2543.

Veneziano, D., Grigoriu, M. and Cornell, C.A. (1977) Vector-process models for system reliability. J Eng Mech Div ASCE 103 EM 3, pp. 441-460

Xiao, Q. and Mahadevan, S., (1994), Fast failure mode identification for ductile structural system reliability, Structural Safety, Vol. 13, No. 4, 207-226.