U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

RELIABILITY-BASED ROBUST DESIGN OPTIMIZATION OF COMPOSITE STRUCTURES WITH APPLICATIONS

Gonçalo Maria Pereira das Neves Carneiro

Thesis submitted to the Faculty of Engineering of the University of Porto
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

FACULTY OF ENGINEERING
UNIVERSITY OF PORTO

2020

RELIABILITY-BASED ROBUST DESIGN OPTIMIZATION OF COMPOSITE STRUCTURES WITH APPLICATIONS

Gonçalo Maria Pereira das Neves Carneiro

Under the supervision of
Professor Doutor Carlos Alberto da Conceição António

RESUMO

A necessidade de incorporar robustez, fiabilidade e incerteza no projeto de sistemas estruturais de engenharia aeroespacial é reconhecida pela indústria. O projeto ótimo robusto baseado na fiabilidade (RBRDO, na sigla em inglês) tem o propósito de encontrar o melhor compromisso entre performance e robustez, ao aceitar a existência de incerteza quer nas variáveis de projeto, que na resposta do sistema estrutural. É usado para garantir a robustez estrutural, ao minimizar simultaneamente medidas de custo e de incerteza, ao mesmo tempo que satisfaz requisitos de fiabilidade. A imposição de restrições de fiabilidade no projeto de estruturas compósitas tem demonstrado ser bastante importante, devido à incerteza inerente ao processo de falha estrutural deste tipo de estruturas, frequentemente associada às propriedades materiais e geométricas de compósitos reforçados com fibras.

Uma vez que a resolução de problemas de projeto ótimo robusto e de projeto ótimo baseado na fiabilidade (em separado) envolve recorrentemente problemas de otimização de múltiplos ciclos, os custos computacionais a eles associados são muito elevados e, comummente, a precisão é preterida à eficiência. Por esse motivo, propomos nesta dissertação a resolução do problema de RBRDO de estruturas compósitas suportada exclusivamente por algoritmos de pesquisa evolucionária, em todos os seus ciclos. Para tal, são propostos novos desenvolvimentos algorítmicos, necessários para satisfazer o propósito de atingir um equilíbrio justo entre precisão e eficiência.

No estudo conduzido, são introduzidos dois modelos alternativos de RBRDO de estruturas laminadas compósitas, baseados em duas metodologias de avaliação da fiabilidade estrutural, nomeadamente, o *reliability índex approach* (RIA) e o *performance measure apporach* (PMA). O ciclo externo de projeto ótimo é formulado como um problema de otimização bi-objetivo de optimalidade e robustez e é resolvido por um algoritmo de pesquisa evolucionária de múltiplos-objetivos, de duas populações, baseado no conceito de dominância restringida. O ciclo interno de análise de fiabilidade é resolvido por algoritmos de pesquisa evolucionária, desenvolvidos especialmente para o efeito. A capacidade dos algoritmos convergirem, com probabilidade 1, para o ótimo global dos problemas de análise de fiabilidade, é demonstrada.

É apresentada uma análise comparativa entre os resultados dos dois modelos de RBRDO desenvolvidos. É também conduzido um estudo sobre os efeitos de diferentes fontes de incerteza sobre o RBRBDO de estruturas laminadas compósitas. Finalmente, a teoria da análise de importância de Sobol' é implementada com o propósito de reduzir a dimensionalidade do espaço de incerteza, melhorando assim a eficiência no sentido da convergência global da análise de fiabilidade baseada na pesquisa evolucionária, ao identificar as variáveis aleatórias mais importantes. Dois resultados teóricos fundamentais, que relacionam polinómios de Taylor de ordem k multilineares, de funções estocásticas multivariadas, com a sua própria decomposição ANOVA. Posteriormente, uma solução analítica aproximada dos índices de Sobol' é proposta, para funcionais de resposta implícitos. É apresentado um estudo de importância das variáveis aleatórias da análise de fiabilidade e os efeitos da redução dimensional comentados.

ABSTRACT

The need to incorporate robustness, reliability and uncertainty into the design of aerospace structural systems is recognized by the industry and commercial sectors. The reliability-based robust design optimization (RBRDO) aims to find the best design performance, by acknowledging the existence of uncertainty both in the design variables and structural response functionals. It is used to guarantee the structural robustness of optimal designs, simultaneously satisfying certain reliability requirements. The imposition of reliability constraints in the design of composite structures has proven to be quite important, due to the uncertainty inherent to the failure modes of this kind of structures, often associated with the material and geometric properties of fiber-reinforced composites.

Since the solution of robust design optimization (RDO) and reliability-based design optimization (RBDO) recurrently involves problems of multiple cycles, the computational costs associated with them are very high and, in general, efficiency is favored in detriment of accuracy. For that reason, in this thesis it is proposed the solution of the RBRDO problem of composite laminate structures supported exclusively by evolutionary algorithms (EA), in all cycles. To this end, new algorithmic developments, necessary to satisfy the purpose of achieving a fair balance between accuracy and efficiency, are proposed.

In the study conducted, two alternative formulations of the RBRDO problem of composite laminate structures are introduced, based on two reliability assessment methods, namely, the reliability index approach (RIA) and the performance measure approach (PMA). The outer cycle of design optimization is formulated as a bi-objective optimization of optimality and robustness and is solved by a two-population multi-objective EA, based on the concept of constrain-dominance. The inner cycle of reliability assessment is solved by two newly developed EAs. The convergence of the algorithms to the global optimal solution with probability 1 is demonstrated.

A comparative analysis between the results of the two design optimization models is presented. A study on the effects of different sources of uncertainty in the RBRDO of composite laminate structures is also conducted. Finally, the importance analysis theory of Sobol' is implemented as a means to reduce the dimensionality of the uncertainty space, improving the efficiency towards global convergence of evolutionary-based reliability assessment, by identifying the most important random variables in reliability assessment. Two fundamental theoretical results relating k-th order multilinear Taylor polynomials of multivariate stochastic functions of independent random variables with its own ANOVA functional decomposition are proposed. Then, an approximate analytical solution of the Sobol' indices is proposed, for implicit structural response functions. An importance study over the random variables in reliability assessment is executed and computing times of reliability assessment are compared.

ACKNOWLEDGEMENTS

I would like to take the opportunity to express my gratitude to all the people who contributed to the completion of my doctoral project. In particular, I would like to mention those people or entities, whose contribution was fundamental to the success of the project.

To my supervisor, Prof. Dr. Carlos Alberto da Conceição António, my sincere recognition for the guidance and for taking the time to share his expertise; for the help in the most difficult moments and, above all, for the trust and friendship.

To all the teachers of the doctoral program, with whom I had the pleasure of learning, my appreciation.

To all the teachers and colleagues of the Mathematics Section of the Mechanical Engineering Department, my special thanks for the friendship and valuable time.

To the Associated Laboratory for Energy, Transports and Aeronautics (LAETA) and specially to the Unit of Numerical Methods in Mechanics and Structural Engineering, my sincere appreciation for the funding and working conditions that allowed me to develop my project under an optimized environment.

To Miguel Barreira, my gratitude for the help and valuable discussions.

To all my friends, my special thanks for the good moments and friendship.

Most importantly, I owe my deepest gratitude to my family for the support, understanding and love and without whom the conclusion of this thesis would not have been possible. I am truly grateful.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Chapter II

$\mathbf{x}$	Vector of search variables
X	Search space
N	Dimensionality of the search space
$f(\mathbf{x})$	Objective function
$g_i(\mathbf{x})$	i-th inequality constraint function
$h_j(\mathbf{x})$	j-th equality constraint function
S	Feasible region of the search space
$\bar{\mathbf{x}}$	Candidate optimal solution
$\mathbf{d}$	Search direction
$\nabla f(\bar{\mathbf{x}})$	Gradient vector of f, at $\bar{\mathbf{x}}$
$\nabla g_i(\bar{\mathbf{x}})$	Gradient vector of g_i, at $\bar{\mathbf{x}}$
$\nabla h_j(\bar{\mathbf{x}})$	Gradient vector of h_j, at $\bar{\mathbf{x}}$
$\boldsymbol{\alpha}$	Feasible arc, leading from $\bar{\mathbf{x}}$
D	Cone of tangents to feasible arcs
$\boldsymbol{\alpha}'$	Tangent vector to an arc
D	Cone of tangents to feasible arcs
F	Cone of tangents to arcs improving f
F_0	Algebraic characterization of F
D_0	Algebraic characterization of D
$N(\nabla \mathbf{h}(\bar{\mathbf{x}}))$	Null space of the Jacobian matrix $\nabla \mathbf{h}$, at $\bar{\mathbf{x}}$
$\mathbf{u}$	Vector of multipliers associated with f and g_i
$\mathbf{v}$	Vectors o multipliers associated with h_j
C	Generic cone of search directions
G'	Closed cone of directions
$\bar{\mathbf{u}}$	Vector of Lagrange multipliers associated with g_i
$\bar{\mathbf{v}}$	Vector of Lagrange multipliers associated with h_j
$\phi(\mathbf{x}, \mathbf{u}, \mathbf{v})$	Lagrangian function
$\theta(\mathbf{u}, \mathbf{v})$	Objective function of the Lagrangian dual problem
$\nabla_{\mathbf{u}}\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$	Gradient vector of ϕ, with respect to $\mathbf{u}$, at $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$
$\nabla_{\mathbf{v}}\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$	Gradient vector of ϕ, with respect to $\mathbf{v}$, at $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$
$\nabla_{\mathbf{x}}\phi(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$	Gradient vector of ϕ, with respect to $\mathbf{x}$, at $(\bar{\mathbf{x}}, \bar{\mathbf{u}}, \bar{\mathbf{v}})$
$P(\mathbf{x})$	Penalty function
$\theta_P(\mu)$	Objective function of the penalty problem
μ	Penalty parameter
$G \rightarrow P$	Genotype-phenotype mapping function
p_m	Mutation probability

r_m	Mutation ratio
$p_{sel}(i)$	Selection probability of the i-th individual, in a population
$\mathbb{S}$	Finite genotype (search) space
$\mathbb{S}^*$	Set of all finite lists over $\mathbb{S}$
σ	n-tuple $(s_1, \dots, s_n) \in \mathbb{S}^*$
y_p	Parent list
$\mathbb{P}$	Set of all parent lists
$N(s)$	Neighborhood of an individual s
f_p	Parent-selection function
f_v	Variation function
f_s	Survivor-selection function
f_{tr}	Transition function
t	Generation number
σ_t	Population, at generation t
σ_0	Arbitrary initial population
$\mathbb{S}_{opt}$	Set of optimal solutions of $\mathbb{S}$
$\mathbf{f}$	Vector objective function, containing k objectives
Z	Objective space
S_X	Feasible (region of the) search space
S_Z	Feasible (region of the) objective space
$\mathcal{R}$	Binary relation on some set A
$(A, \mathcal{R})$	Ordered set
$\prec$	Pareto dominance relation
$\preccurlyeq$	Dominant or equal
$(Z, \preccurlyeq)$	Partially ordered objective space
Z_{PF}	Pareto front of Z, w.r.t. $\preccurlyeq$
$(X, \preccurlyeq_p)$	Preordered search space
X_{PO}	Pareto optimal set of X, w.r.t $\preccurlyeq_p$
$Z_{PF}(t)$	Current Pareto front
$X_{PO}(t)$	Current Pareto optimal set
$Z_{PF\,known}$	Known Pareto front, after convergence
$X_{PO\,known}$	Known Pareto optimal set, after convergence
$r(\mathbf{z})$	Ranking function of elements $\mathbf{z} \in Z$
$\preccurlyeq_r$	Pareto-ranking relation
F_i	Fronts on Z
$\prec_c$	Constrain-dominance relation
$\preccurlyeq_c$	Constrain-dominant or equal

Chapter III

$\mathbf{d}$	Vector of design variables
$\boldsymbol{\pi}$	Vector of system parameters
$\mathbf{u}$	Vector of state variables (displacements)
f_i	i-th objective function
φ_j	j-th system response functional
$\boldsymbol{\delta}$	Vector of random deviations
$\mathbf{x}$	Vector of random variables
$\boldsymbol{\mu}_{\mathbf{x}}$	Vector of mean-values of the random variables
ϕ_j	j-th stochastic system response functional
$\mathbf{C}_{\boldsymbol{\phi}}$	Variance-covariance matrix of the stochastic system response functionals
$\mathrm{Cov}(\phi_i, \phi_j)$	Components of $\mathbf{C}_{\boldsymbol{\phi}}$
$\det \mathbf{C}_{\boldsymbol{\phi}}$	Determinant of $\mathbf{C}_{\boldsymbol{\phi}}$
D_f	Failure space
D_s	Safety space
L	Failure surface
$(\Omega, \mathcal{F}, P)$	Uncertainty space
p_f	Probability of failure
$\bar{u}$	Critical displacement response of the structural system
R	Tsai number
X	Longitudinal tensile strength
X'	Longitudinal compressive strength
Y	Transverse tensile strength
Y'	Transverse compressive strength
S	Longitudinal shear strength
$\bar{R}$	Critical stress response functional (Tsai number)
g_1	Stochastic displacement limit-state function
g_2	Stochastic stress limit-state function
W	Total structural weight
u^a	Allowable critical displacement
R^a	Allowable critical Tsai number
p_f^a	Allowable probability of failure

Chapter IV

D_f	Failure space
p_f	Probability of failure
$g_2(\mathbf{x}, \boldsymbol{\pi})$	Stochastic stress limit-state function
$p_{\mathbf{x},\boldsymbol{\pi}}(\mathbf{x}, \boldsymbol{\pi})$	Joint probability density function of the random variables $(\mathbf{x}, \boldsymbol{\pi})$
$\hat{p}_f$	Numerical estimate of p_f
β_c	Cornell reliability index
$\mathbf{y}$	Independent standard normal random variables
$G_2(\mathbf{y})$	Standardized stress limit-state function
$\mathbf{y}_{\mathbf{MPP}}$	Most probable failure point
β_{HL}	Hasofer-Lind reliability index
β^a	Allowable reliability index
$\mathbb{S}$	Finite genotype space
$\mathbb{S}^*$	Set of all finite lists over $\mathbb{S}$
$\mathbf{P}^t$	Population, at generation t
$\mathbf{E}^t$	Elite group of the population, at generation t
$\mathbf{B}^t$	List of all offspring solutions, at generation t
$\mathbf{P}^t \cup \mathbf{B}^t$	Extended population, at generation t
$n_{\mathbf{P}}$	Number of individuals in $\mathbf{P}^t$
$n_{\mathbf{B}}$	Number of individuals in $\mathbf{B}^t$
β	Euclidean norm of a vector $\mathbf{y}$
$\mathbf{a}$	Vector of direction cosines
$\Gamma(\cdot)$	Penalty function
$\mathbb{B}^l$	Set of binary strings of length l
$s(\cdot)$	Real-to-binary mapping function
$(\beta, s(\mathbf{a}))$	Mixed genotype array
$\Delta\beta^k$	Nonlinear increment function
$\Delta\beta^k_{max}$	Amplitude of the increment function, at iteration k
$\mathcal{Z}$	Interest region of the uncertainty space
$\mathcal{Z}^0$	Initial search region
β_{min}, β_{max}	Euclidian distances measured from the origin of the uncertainty space
$\mathbf{E}^t_{\mathcal{Z}}$	List of solutions of high probabilistic failure content, ate generation t
$\mathcal{Z}^t$	Reduced search region
$\mathrm{Var}_{\mathbb{E}}(\beta)$	Sample variance of β
σ^2_{min}	Minimum level of diversity in the elite group
β^*	Approximation of the reliability index
$B(\beta^a)$	probability density hypersphere of radius β^a
$s(\mathbf{a})$	Binary genotype array

Chapter V

$(\mathbf{x}, \boldsymbol{\pi})$	Vector of random variables
$\mathbf{C}_{\boldsymbol{\phi}}$	Variance-covariance matrix of the stochastic system response functionals
$\mathrm{Cov}(\phi_i, \phi_j)$	Components of $\mathbf{C}_{\boldsymbol{\phi}}$
ϕ_i	i-th stochastic system response functional
$\boldsymbol{\delta}$	Vector of random deviations
$\varphi_i(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})$	Deterministic realization of ϕ_i, $(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})$
$s_{i,k}$	Sensitivity of φ_i with respect to k-th random variable, calculated at $(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})$
$\mathbf{S}$	Sensitivities matrix
$\mathbf{V}$	Diagonal variance matrix of the random variables
$\det \mathbf{C}_{\boldsymbol{\phi}}$	Determinant of $\mathbf{C}_{\boldsymbol{\phi}}$
$(\boldsymbol{\sigma}_{\mathbf{x}}, \boldsymbol{\sigma}_{\boldsymbol{\pi}})$	Standard deviations of the random variables
g_2	Stochastic limit-state function
$\mathbf{y}$	Vector of independent standard normal random variables
$G_2(\mathbf{y})$	Standardized stochastic limit-state function
p_f	Probability of failure
p_f^a	Allowable probability of failure
$\mathbf{y}_{\mathbf{MPP}}$	Most probable failure point
$G_{2_L}(\mathbf{y})$	First-order approximation of $G_2(\mathbf{y})$, at $\mathbf{y}_{\mathbf{MPP}}$
β_{HL}	Hasofer/Lind reliability index
β^a	Allowable reliability index
β^*	Approximation of the reliability index
$\not{g}_2$	Deterministic piecewise equivalent reliability limit-state function
$Q(\mathbf{y}, p)$	Quantile function
ξ	Constraint violation function
$\mathbf{SP}^t$	Short population, at generation t
$\mathbf{EP}^t$	Enlarged population, at generation t

Chapter VI

$(\Omega_2, \mathcal{F}, P_2)$	Uncertainty space
g	Stochastic limit-state function
$\mathbf{x}$	Vector of independent random variables
N	Number of random variables
P_2	Probability measure
g_0	Expectancy of $g(\mathbf{x})$ with respect to P_2
$g_{i_1,\ldots,i_s}$	Summands of the ANOVA decomposition
$S_{i_1,\ldots,i_s}$	Sobol' index associated with the random variables $x_{i_1}, \ldots, x_{i_s}$
α	Multi-index, such that $\lvert\alpha\rvert = \alpha_1 + \cdots + \alpha_N$ and $\alpha! = \alpha_1 \ldots \alpha_N$, $\forall \alpha_i \geq 0$
$\mu_{\alpha_i}(x_i)$	α_i-order central moment of x_i, with respect to P_2
ε	Nonnegative scalar
$(\Omega_r, \mathcal{F}_r, P_r)$	Reduced uncertainty space

Chapter VII

ϕ_i	i-th stochastic system response functional
$\mathbf{x}$	Random design variables
$N_{\mathbf{x}}$	Number of random design variables
$\boldsymbol{\pi}$	Vector of random parameters
$\boldsymbol{N_{\pi}}$	Number of random parameters
$\mathbf{m} \subseteq \boldsymbol{\pi}$	Vector of random mechanical properties
$\boldsymbol{\theta} \subseteq \mathbf{x}$	Vector of random ply angles
$\mathbf{h} \subseteq \mathbf{x}$	Vector of random thickness variables
$\mathbf{p} \subseteq \boldsymbol{\pi}$	Vector of random point loads
$\boldsymbol{\mu}_{\mathbf{x}}$	Vector of mean-values of the random design variables
$\boldsymbol{\mu}_{\boldsymbol{\pi}}$	Vector of mean-values of the random parameters
$\boldsymbol{\sigma}_{\mathbf{x}}$	Vector of standard deviations of the random design variables
$\boldsymbol{\mu}_{\boldsymbol{\pi}}$	Vector of standard deviations of the random parameters
$\mathbf{C}_{\boldsymbol{\phi}}$	Variance-covariance matrix
$\det \mathbf{C}_{\boldsymbol{\phi}}$	Determinant of $\mathbf{C}_{\boldsymbol{\phi}}$
g_1	Stochastic displacement limit-state function
g_2	Deterministic piecewise equivalent reliability limit-state function
E_1	Longitudinal elastic modulus
E_2	Transversal elastic modulus
G_{12}	In-plane shear modulus
ν_{12}	In-plane Poisson ration
X	Longitudinal tensile strength
X'	Longitudinal compression strength
Y	Transversal tensile strength
Y'	Transversal compression strength
S	Shear strength
β_{HL}	Hasofer-Lind reliability index
β^a	Allowable reliability index
R^a	Allowable Tsai number
u^a	Allowable critical displacement
$X_{PO_{known}}$	Known Pareto optimal set, after convergence
$Z_{PF_{known}}$	Known Pareto front, after convergence
$\bar{R}$	Critical Tsai number
$\mathbf{y_{MPP}}$	Most probable failure point
S_X	Feasible search space
S_Z	Feasible objective space
$CV(\cdot)$	Coefficient of variation
$\mathcal{Z}^{t_1}$	First reduced search region
Δt	Imposed number of generations after $\mathcal{Z}^{t_1}$
p_f	Probability of failure
γ	Confidence level

p_f^{low}	Lower bound of confidence interval
p_f^{up}	Upper bound of confidence interval
K_f	Number of failure events
N_s	Number of samples
ε	Threshold parameter
S_{m_i}	Sobol' index associated with the random mechanical property m_i
t_{total}	Time spent by the HmGA to achieve convergence
t_1	Time spent by the HmGA to find $\mathcal{Z}^1$
$t_{\leq}$	Time spent by the HmGA to achieve superior accuracy

I
INTRODUCTION

1.1 Motivation

Composite materials constitute a broad and important class of engineering materials. Effectively, many materials are composites, both of biological and artificial natures [1]. For that reason, there is not a widely accepted definition of *composite material*, since such a definition must be broad enough to include the vast number of materials that can be regarded as composites, in their area of application.

In the context of structural design, the following definition, due to Smith [2], provides an accurate description of the meaning of *composite material*:

> *"A composite material is a materials system composed of a mixture or combination of two or more micro- or macro-constituents that differ in form and chemical composition and are essentially insoluble in each other."*

The importance of composite materials, in engineering, is that two or more constituents of significantly different properties may combine to form a new material (composite) possessing properties that are superior to those of the constituents themselves [3]. Commonly, a constituent of high stiffness and strength (reinforcement) is embedded in a softer and ductile material (matrix) that binds and protects the reinforcement [4].

This is the case of *fiber-reinforced plastics* (FRP). These are composite materials made of a polymer matrix reinforced with fibers, usually of synthetic nature. The fibers of such composites are generally strong and stiff and therefore serve as the primary load-carrying constituent. The matrix holds the fibers together and serves as an agent to redistribute the loads from broken fibers to the adjacent ones in the material, when fibers start failing under excessive loads. FRPs represent a very important class of engineering materials, characterized by high stiffness and strength and low density. Their development dates to the beginnings of the twentieth century, but it's only after the Second World War, with the advent of the aerospace and aeronautic industries, that the application of these materials quickly expands into other engineering applications [5], as a result of a systematic scientific research and the increasing demand of industries to produce better products at reduced costs.

Currently, FRPs and composite materials in general are of widespread use in many industries, with applications in aerospace and aeronautics engineering, automotive engineering, civil engineering, electrical engineering, ballistics, medicine, among others [6]. Usually, in high-end applications and industries, preference is given to the so-called *advanced composite materials*, with reinforcements such as carbon fibers, aramid fibers, boron fibers [7]. In more commercial industrial segments, the predominant reinforcement constituent is fiberglass, because of its lower cost. Epoxy resins are commonly used as matrix materials for advanced composite materials, when the structural benefits surpass the cost. Alternatively, polyesters resins are usually not as strong as the epoxy ones, but are lower in cost and therefore widely used in FRPs. [2].

From the point-of-view of structural design, composite materials show very interesting characteristics and offer designers several technical advantages, in comparison with traditional advanced materials, such as titanium and aluminum. Composites allow the reduction of structural weight, due to high specific strength and stiffness; are ideal for precision engineering applications, due to good dimensional stability; allow to simplify production and reduce costs, mainly due to the reduced number of structural components and joints and the ease of producing molded polymers in comparison with cast aluminum or steel [8].

Commonly, composite materials show marked anisotropy or orthotropy. This usually arises because the fibrous reinforcement is preferentially aligned in particular directions [1]. Probably, the biggest advantage of artificial composites is their potential for controlled anisotropy that gives the ability to tailor their properties to the user's needs. Properties like stiffness and strength can be designed by selecting the constituents' material, their proportion and geometrical arrangement. The engineer is able to design the material concurrently with the structure [9].

A good example is that of *composite laminate materials*. A composite laminate is an assembly of two or more layers, each consisting of a softer matrix reinforced with a unidirectional, bidirectional or multidirectional constituent, designed to provide the required engineering properties. The layers may be composites themselves, such as FRPs layers placed so that different layers have different characteristics. This is the most commonly encountered laminated composite material used in the design of high-performance structures [3]. Layers of different materials may be used, resulting in hybrid laminates. Overall, the materials of the individual layers, the orientation of the fibers and the stacking sequence determine the properties of the laminate [7].

The directional nature of the fibers, in fiber-reinforced laminates, induces the directional dependence of the material properties. The matrix is generally homogeneous and isotropic. For all practical purposes, the fibers, which usually are much stiffer than the matrix, are homogenous and isotropic, as well. However, when combined, the properties of the composite are no longer isotropic, neither homogeneous [3]. In truth, composite laminates with oriented fibers are heterogeneous and orthotropic in nature, with properties defined in two directions in the plane of the layer – the direction along the fibers and the direction perpendicular to the fibers. Nevertheless, from a macroscopic perspective, the multiphasic composite laminate material may be approximated as being globally homogeneous, with equivalent averaged properties everywhere, and the theory of elastic orthotropy used to determine the stress and strain fields [7]. It is further assumed that the mechanical properties of each layer are well defined.

While composite laminate materials are attractive replacements for metallic materials, especially for lightweight structures that have rigorous stiffness and strength design requirements, the analysis and design of these materials and structures are considerably more complex, than those of the traditional isotropic materials. In terms of structural design, the orientation of the fibers, the thickness and material of the layer and the stacking sequence are variables commonly associated with the design of the composite laminate material itself. It means that finding an efficient structural design that meets all requirements of a specific application can be achieved not only by sizing the cross-sectional areas and member thicknesses, but also by tailoring of the material properties. The structural design of composite laminate structures shows therefore an additional degree-of-freedom, relative to the material characterization that is not seen with the traditional isotropic materials [3]. Furthermore, the most common design constraints are mostly written in terms of nonlinear functionals of the additional design variables of material characterization.

From an industrial perspective, it means the analysis of different design possibilities and of potential solutions is delayed. In practice, it ends up restricting the design process to certain fiber orientations, stacking sequences and material choices whose behaviors are well known by the designers.

The increased number of design variables is both a blessing and a curse for the designer. Composite laminates offer a superior ability to fine-tune the structure, but the selection of the important design variables and the choice of their values, to meet the design requirements, are a burden. The possibility of achieving an efficient design that is safe against multiple failure mechanisms, coupled with the difficulty in selecting the values of a large set of design variables makes *mathematical optimization* a natural tool for the design of composite laminate structures, with the purpose of turning the structural design a more systematic and well defined task, becoming less dependent on the designer sensitivity and achieving the maximum material performance [3]. The continuous research and development of new numerical tools, namely, finite element models suitable for composted structures, sensitivity analysis techniques for the differentiation of the objective and structural response functionals and efficient optimization algorithms, made computer aided design a powerful and seductive instrument for the design of composite laminate structures.

As any design optimization problem, the formulation and solution of *structural design optimization problems* is dependent on the technical and theoretical aspects of the problem itself. According to Eschenauer [10], in order to deal with optimization problems in the structural design process, a procedure should be established following the "*three columns concept"*. The three columns are:

1. Structural model;
2. Optimization algorithm;
3. Optimization model.

Accordingly, the *optimization model* is the search scheme for the optimal design that defines the linkage between the *optimization algorithm* and the *structural model*. It is a fundamental part of the structural design process. Therefore, the optimization model must include design variables and parameters, state variables, objective and system response functionals and design constraints.

Two fundamental concepts in structural design optimization are the notions of *structural integrity* and *structural stability*. The first refers to the ability of the structure to carry the applied loads without exceeding the materials' strength and collapsing. The latter, to the steadiness and predictability of the structural response. It is fundamental for the reproducibility of the designs, for example, in the industry. The increasing demand for composite laminates and the responsibility associated to it encouraged the analysis of the failure process of these materials, which is characterized by complex interactions between the matrices and fibers and the different layers [3]. On the other hand, the manufacturing process and any physical or human agent intervening on it, the quality control, the actual heterogeneity of composite materials and even the external loads applied to the structures will affect the quality of the final material and, therefore, the structural response, increasing the concerns regarding both structural integrity and stability [7].

Traditionally, deterministic philosophies of structural design rely on the over-sizing of the designs, to guarantee structural integrity. The explicit assessment of structural stability is often disregarded. However, currently, the necessity to take into consideration the uncertain nature of the most important design features, such as external loads, material strength and stiffness properties, the fibers orientation and geometrical parameters has been recognized by the scientific and the industry sectors, by the application of a stochastic design optimization philosophy.

In the literature, the optimization models developed for the design of composite laminate structures may be classified according to four alternative approaches: *deterministic models* [11], *probabilistic models* [12], *epistemic models* [13] and *fuzzy models* [14].

The design optimization models developed in this thesis belong to the category of probabilistic models. Specifically, this thesis deals with random uncertainty, in the sense that it is considered that the probability distributions governing uncertainty are known. In structural design, it is possible to

find essentially two approaches to the quantification of the effects of random uncertainty upon the design solutions, each having a distinct motivation.

Reliability assessment represents the probabilistic analysis of failure. The development of probabilistic methods of structural integrity, based on the theory of structural reliability, represents a turning point in the design of composite structures, in contrast to the traditional application of safety factors. Structural design optimization involving reliability constraints is a discipline called *reliability-based design optimization* (RBDO). The existence of reliability constraints in the design optimization aims to limit the frequency of failure events, due to random variations in the structural parameters. The traditional measure of structural reliability is a cumulative probability called the *probability of failure*. Often, it is represented by an approximate (and easier to calculate) measure known as the *reliability index*. Two alternative reliability assessment methods, known as the reliability index approach (RIA) and the performance measure approach (PMA), based on the concept of the reliability index, transform the integral form associated with cumulative probabilities into a minimization problem [15].

Robustness assessment represents the probabilistic analysis of variability (stability) of the structural responses. In concrete, it measures how much the behavior of a design is expected to deviate from its expected value. Structural design involving robustness measures is a discipline known as *robust design optimization* (RDO). The consideration of robustness in design optimization aims to limit the variability of the structural response of the design solutions, due to random variations in the structural parameters. There are several robustness measures proposed in the literature, the most common in design optimization being based on the statistical moments of the objective or structural response functionals, such as the variance. Robustness measures can be applied both at the objective and the constraints levels [16].

The application of both the RBDO and the RDO formulations to optimization models of large composite laminate structures does not come without computational difficulties. In terms of structural reliability, there is an increasing number of publications on its application to design optimization. However, the literature is saturated with fast gradient-based optimization algorithms to solve the RIA and the PMA [17]. It is believed that this fact is related with the high extra computational cost of reliability assessment, when introduced as an inner cycle of structural design optimization. Common alternatives found in the literature are the application of sampling methods, such as Monte Carlo simulation (MCS) and related methods, for the direct calculation of the probability of failure, which are known as being extremely expensive for large structures. Regarding RDO, approximate methods based on the theory of propagation of moments are usually preferred [16]. MCS and related methods are also applied for the direct calculation of the statistical moments of the objective and/or structural response functionals [16].

In recent years, a new research field in structural design optimization combining robustness and reliability assessment, known as *reliability-based robust design optimization* (RBRDO), has been explored. Few works have already been published, in the context of structural design, the main difficulty being the impractical computational cost of the design process. Application to composite laminate structures is still very reduced.

1.2 State-of-the-art review

In the following sections, a concise review of the recent and relevant publications on the topics of structural stochastic design optimization of composite laminate structures is presented.

1.2.1 RBDO of composite laminates

MCS is the foundation for sampling methods for reliability assessment. It consists in evaluating a traditional binary process, for which the frequency of failure events is an unbiased estimator of the probability of failure [18], with any desired accuracy. For most engineering problems, failure probabilities are too small and the number of required samples is huge. To cope with this, variance reduction techniques and improved surrogate models are proposed frequently. References [18-24] provide a collection of relevant and some recent advances in the application of MCS in structural reliability.

The inconvenience of MCS in reliability assessment is its global sense. Yet, random variations in structural variables are likely to happen near the respective mean-values. Local reliability methods focus their action in that region and are usually based on the concept of the reliability index. The first approaches were made by Cornell [25], Ditlevsen [26] and Lind [27]. These proposals were not invariant under different representations of the same failure criteria. Hasofer and Lind [28] proposed an invariant reliability index as the minimum distance, in standard deviation units, from the origin of the standard normal space of independent random variables to the failure surface. The corresponding point is named the most probable failure point (MPP). The authors presented a gradient-based iterative procedure for the solution of the minimization problem at hands. Rackwitz and Fissler [29] generalized the results of Hasofer and Lind to non-normal variables to what became known as the HL-RF algorithm.

The relation between the reliability index and the failure probability is documented. First- and second-order approximations of the failure surface at the MPP (FORM and SORM) give values of the probability of failure, assuming normally distributed uncertainty [15]. Recently a number of publications on methods to estimate the reliability index were proposed, based on gradient-based algorithms and response surface methods, to define an approximate explicit failure surface, in implicit problems [30-35]. Despite their efficiency, gradient methods often diverge for high dimensional problems and/or highly concave limit-state functions [36-39].

In the design of composite structures, reliability assessment is frequently an inner cycle of design optimization. The Reliability Index Approach (RIA) and the Performance Measure Approach (PMA) are the most common philosophies to impose and evaluate reliability constraints [17]. Reviews on techniques applied in the RBDO of composite structures are in references [40,41]. Despite the theoretical advances in the field of reliability analysis, serious computational obstacles arise when treating realistic problems. In particular, the RBDO of large-scale structural systems is an extremely computationally intensive task, as shown by Tsompanakis and Papadrakakis [42].

Zhou *et al.* [43] and Dehmous *et al.* [44] addressed the multi-scale design optimization of composite structures. The RIA was applied and solved by gradient-based algorithms. Chen *et al.* [45] applied the RBDO to optimize the stacking sequence of composite laminate structures, under reliability constraints. The RIA was solved by a gradient-based iterative procedure. Conceição António [7] and Conceição António et al. [46] developed an iterative scheme based on a gradient method to assess the reliability in RBDO of composite shells. Conceição António and Hoffbauer [47-48] preformed inverse reliability analysis, by minimizing the squared difference of the actual and target reliability index to achieve the desired reliability level of composite structures, with random material properties. The reliability index was estimated by a gradient-based iterative procedure.

Because the RBDO problem is by definition a double-cycle problem, several authors attempted to develop methodologies, usually based on the PMA, to avoid that issue. Chen *et al.* [49], considering the mean values of uncertainty parameters as normally distributed design variables, developed an iterative procedure, where the probabilistic constraints were replaced by an equivalent deterministic

one, reducing the RBDO to a single-loop problem. Another known method is the Sequential Optimization and Reliability Assessment (SORA), which aims to decouple both loops [50].

On structures with the possibility of discontinuous implicit response functionals, and/or their derivatives, gradient methods may also diverge. The use of meta-heuristics to find the global reliability index in structural problems is very scarce. Cheng and Li [51] proposed an artificial neural network (ANN)-based genetic algorithm (GA) to find the reliability index of structures. An ANN model trained from a data set obtained by the uniform design method extracted an explicit approximation of the limit-state function. Then a traditional GA was used to find MPP. The method was applied to a truss structure and proved to be efficient, for two random variables. Shao and Murotsu [51] proposed to find the reliability index with a GA, by searching only a finite number of prescribed directions. Deng *et al.* [53] then applied a shredding technique to the same algorithm to filter undesirable genetic characteristics, improving the efficiency of the GA. The method was applied to a truss problem and to a continuous nonlinear steel structure. A reasonable number of random variables was considered. Wang and Ghson [54] modified the previous shredding GA, to include linkage-learning operators to improve local and global convergence. In the context of composite structures, Conceição António [12] and Conceição António and Hoffbauer [55] developed a hierarchical GA for a unified design optimization and global MPP search, with critical displacement, critical stress and critical buckling constraints. The RBDO is performed in a single optimization cycle.

To the present date, it was not possible to find any publication on the solution of the PMA by EAs, neither combining the solution of reliability assessment by EAs with structural design optimization, in a RBDO model.

1.2.2 RDO of composite laminates

There are several formulations and concepts of robustness. Beyer *et al*. [16] present a complete discussion on the state-of-the art concepts in robustness assessment.

Conceição António and Hoffbauer [56] presented a new concept of feasibility robustness associated with the variability of design constraints. The RDO approach was defined as a bi-objective minimization problem of weight and the determinant of the variance-covariance matrix. It was shown that the Pareto-optimal front depends on the sources of uncertainty considered. Henrichsen *et al*. [57] studied the robust optimal design of laminated composite structures under buckling. Uncertainties were considered as geometric imperfections and modelled by specific shape functions. The chosen concept of robustness reflects the "worst-case scenario". The resultant optimization problems were solved using a nested approach, meaning that buckling analysis is not included in the optimization, but is solved explicitly when needed. Li *et al*. [58] developed a multi-objective RDO procedure for the crashworthiness design of foam filled thin-walled structures. Both random and interval types of uncertainty were considered for which two robustness measures were imposed as constraints. Specific Energy Absorption and Peak Crushing Force were considered as optimality objectives. Surrogate models were used to predict the both objectives and constraint functionals and MCS was used to estimate the robustness measures of problem. Kalantari *et al.* [59] considered a bi-objective optimization of cost and weight for the RDO of unidirectional carbon/glass hybrid composites, under flexural loading. Feasibility robustness was considered, which consisted of a "worst-case scenario". In this case, it means that a higher value of performance constraints is imposed, setting a safety margin for random uncertainty. The bi-objective optimization problem was then solved with a hybrid EA with local search, based on the NSGA-II. Uncertainty in ply-angles and thickness variables was considered. Xiao-Yi *et al.* [60] proposed a RDO model for variable curvilinear fibers composite structures in the presence of uncertainties in the constituent material properties and applied loads. The structural design problem was formulated as the bi-objective problem of performance and robustness,

where the latter is defined as the standard deviation of the objective function. The problem is then reduced to a single-objective RDO of the weighted sum of the two objectives and solved by a GA.

1.2.3 RBRDO of composite laminates

To the present date, publications on the RBRDO of composite structures are scarce. It is possible that it may be connected with the excessive computational cost required to be applied to large and complex structures. Only one publication, prior to this thesis, was found regarding the application of RBRDO to composite laminate structures. The remaining consider traditional isotropic materials.

Mourelatos and Liang [61] proposed a bi-objective model trading-off optimality and robustness subject to reliability constraints. Robustness was defined as the percentile difference of the objective function, as a measure of the spread of its probability density function, and was calculated by a variation of the advanced mean-value method. The two objectives were reduced into one, using a preference aggregation method as a way of reflecting the preference of designers and avoiding calculating the entire Pareto front. The original problem was then solved as a traditional RBDO problem, where reliability assessment was executed by the PMA. An efficient single-loop methodology based on gradient-based algorithms was applied to solve the resulting optimization problem. The method was then applied to an abstract mathematical example and an isotropic cantilever beam.

António and Hoffbauer [62] proposed an approach for RBRDO of angle-ply composites. It was considered as a bi-objective optimization problem, where a trade-off between performance target and robustness was expected to be achieved. The objective functions considered were the squared difference between the actual structural reliability index and its target value, as a performance measure, and the quotient of the coefficient of variation of the maximum applied load by its nominal value, as a robustness measure. Also, to make sure the optimal designs are reliable, the classical reliability constraint, based on RIA, was imposed. An optimization model was proposed to solve the problem based on a two-cycle optimization: one for reliability the assessment and another for the robustness assessment.

Motta and Afonso [63] proposed an optimization tool with the aim to obtain robust and reliable designs in short computational time. Robustness was defined by the standard deviation of the objective function. Reliability assessment was performed by the PMA and by an approximated approach, through the use of the statistical moments of the reliability constraints. Several other approximate methods were applied with the aim to reduce the computation cost. The authors concluded that the approach proved to be effective reducing the computational time by up to five orders of magnitude, in comparison with classical approaches of reliability and robustness assessment. The optimization model was applied to the design of 2D and 3D trusses.

Wang *et al.* [64] proposed a dynamic RBRDO model of a wind turbine's shrink disk. System reliability was defined as a combination of multiple limit-states. The Cornell [25] reliability index was applied as a simplified analytical reliability measure. The objective function was written in terms of a robustness measure defined as a function of the reliability sensitivities, with the intent to minimize the variability associated with the reliability of the structure. Both robustness and reliability were defined as a function of time.

1.3 Objectives and organization of the thesis

This thesis has as its primary goals the development of structural design optimization models combining structural reliability and robustness and their application to the design of composite laminate structures. The most important pre-requisite being that the developed models must be both *accurate* (in the sense that global optimality is attained) and *efficient* (achieving global optimality with economic costs). For that purpose, the execution of the developed optimization models is guaranteed by the exclusive use of evolutionary algorithms (EA).

The contribution of this research work to the design optimization of composite laminate structures is summarized in the following points:

1. Identification of *performance optimality* and *robustness* objectives and use of bi-objective optimization models;
2. Conjugation of robustness and reliability measures in the design optimization of composite laminate structures;
3. Development of EAs specifically for the efficient and global solution of the structural reliability assessment problems: the RIA and the PMA;
4. It constitutes a first approach to the solution of RBRDO problems with the exclusive use of EAs, in all cycles of the optimization model.
5. Study of the effects of different sources of uncertainty on the design optimization of composite laminate structures;
6. Identification of the most important random material properties in reliability assessment, by the analytical solution of Sobol' indices, and definition of an analytical dimensional reduction technique;
7. Application to large structures, with high number of variables.

The focus of this work is, therefore, on two of the three column concepts of structural design optimization, namely, the development of optimization algorithms and optimization models. This research work is, at the same time, restricted by the following practical implications:

1. Application to angle-ply composite laminates;
2. Definition of the structural limit-state by a first-ply failure philosophy;
3. Probabilistic reliability and robustness assessment;
4. Assumption of independent random variables.

The document is organized as follows.

Chapter 2 provides a concise introduction to mathematical optimization, in particular it focuses on the disciplines of *nonlinear programming* and *evolutionary computing*. Sect. 2.1 begins with a formal definition of the *nonlinear programming* (NLP) *problem*. Sect. 2.1.1 introduces the *necessary optimality conditions* for constrained optimization. Sect 2.1.2 introduces *Lagrangian duality* and the *exterior penalty formulation* of the NLP problem. Sect. 2.1.3 presents a historical overview of numerical optimization methods. The preference towards evolutionary algorithms (EAs), in the detriment of (faster) classic optimization methods, is explained. The evolutionary metaphor applied to metaheuristics in evolutionary computing is presented, followed by the conceptual structure that forms the common basis for the different variants of EAs. The main components of EAs are introduced, followed by notes on genetic diversity, hybridization and constraints handling are presented. At the end, the probabilistic convergence of EAs is studied.

Sect. 2.2 starts with a formal definition of the *multi-objective nonlinear programming problem*. Sect. 2.2.1 introduces *Pareto optimality* and the concept of *Pareto dominance* binary relation. Sect. 2.2.2 introduces multi-objective EAs (MOEAs), explains the preference for these algorithms and identify the three main goals MOEAs are expected to achieve. The *sorting mechanisms* of historical MOEAs are discussed. Brief notes on the hybridization of MOEAs and constraints handling are presented.

On Chapter 3 the concepts of the stochastic design optimization of composite laminate structures are introduced and a global RBRDO model is proposed. On Sect. 3.1 the concept of stochastic design optimization as an engineering design methodology is introduced, including the main components of the proposed RBRDO problem. The existence of uncertainty associated with structural design is discussed as well and specific definitions of random variables and stochastic response functionals are given. On Sect. 3.2, the concept of robustness is introduced and a global robustness measure based on the probability theory is defined. Similarly, on Sect. 3.3, the concept of probabilistic reliability, as a measure of structural integrity, is presented. The particular probability spaces associated with the probabilistic reliability assessment and the related probability measure of structural integrity are defined. Sect. 3.4 provides the definition of the response of the mechanical system in terms of displacements and stresses. The Tsai-Wu quadratic failure criterion for composite structures is presented. Sect. 3.5 provides a formal statement of the RBRDO problem, as a stochastic nonlinear programming problem.

Chapter 4 presents in detail the numerical computation of the structural reliability constraint and proposes new methodologies for reliability assessment, based on evolutionary computation. Sect. 4.1 presents a concise theoretical background on the numerical methods most commonly applied to solve the probability of failure integral: Monte Carlo Simulation, as a global reliability method, and the Reliability Index Approach (RIA) and the Performance Measure Approach (PMA), as local reliability methods. The preference for local reliability methods to assess the probabilistic reliability constraint in structural design optimization is explained. In Sect. 4.2, two novel evolutionary algorithms (EA) are specifically developed as numerical tools to solve both the RIA and the PMA problems, called the Hybrid micro Genetic Algorithm (HmGA) and the micro Genetic Algorithm (mGA), respectively. A new genetic repair operator, working on a continuous domain, is proposed for the local improvement of solutions and its convergence ability proven. A second operator for the progressive reduction and reallocation of the search space is also proposed, allowing for the iterative approximation of the unknown interest region of the uncertainty space associated with the reliability assessment problem. Such algorithms constitute a novelty in reliability assessment, since it is almost exclusively performed by gradient-based algorithms and sampling methods, in the literature. The preference for EAs over other optimization algorithms is discussed. The developed EAs are defined and discussed in depth and their ability to converge with probability 1 is demonstrated.

Chapter 5 defines approximate formulations of the proposed RBRDO problem, based on the local solution of the global robustness and reliability measures. Sect. 5.1 proposes the local approximation of the components of the variance-covariance matrix, by the theory of propagation of moments. On Sect. 5.2, the probabilistic reliability constraint is rewritten, in terms of the deterministic reliability measures calculated by the reliability index approach (RIA) and the performance measure approach (PMA). Sect. 5.3 introduces a MOEA for the solution of the proposed RBRDO problems of composite laminate structural systems. The search process is based on the coevolution of two populations that exchange information at two levels: first, through migration of solutions and, second, through genetic recombination. On each population, a particular ranking mechanism is applied, based on the concept of constrain-dominance.

In the reliability assessment of composite laminate structures with multiple components, the uncertainty space defined around design solutions easily becomes over-dimensioned and not all of the random variables are relevant. On Chapter 6, the importance analysis theory of Sobol' is

implemented to reduce the dimensionality of the uncertainty space, improving the efficiency towards global convergence of evolutionary-based reliability assessment. On Sect. 6.1., the importance analysis theory of Sobol' is introduced. Sect. 6.2 begins with the exposition of two fundamental results relating k-th order multilinear Taylor polynomials of multivariate stochastic functions of independent random variables with its own ANOVA functional decomposition. Then, Sobol' indices are formulated analytically for implicit structural response functions, following the theory of propagation of moments and without violating the fundamental principles presented by Sobol'. On its application to implicitly defined structural systems, the proposed formulation has the advantage of only requiring one adjoint system of equilibrium equations to be solved once. On Sect. 6.3, an analytical dimensional reduction criterion is presented. The theoretical developments assume independent random variables.

On Chapter 7, the RBRDO problem of composite laminate structures defined in Chapter 3 is solved by the algorithms proposed in Chapters 4 and 5. The developments assume the local approximation of the global robustness and reliability measures, based on gradient information of the stochastic response functionals of structural systems. However, the numerical estimate of derivatives of multivariate implicit response functionals is expensive, particularly in high dimensional problems. For that reason, Sect 7.1 introduces the adjoint variable method for the analytical estimate of gradient vectors of the stochastic response functionals. This method only requires an adjoint system of equilibrium equations to be solved, reducing the computational cost of differentiation.

Sect 7.2 begins by describing how uncertainty is propagated differently in robustness and reliability assessment and the local RBRDO problem is defined properly. Then, in Sect 7.2.1, a numerical physical model is introduced. In Sect 7.2.2, the local bi-objective RBRDO problem is solved. The outer cycle of design optimization is solved by the bi-level dominance multi-objective GA (MOGA-2D), introduced in Chapter 5. The inner cycle of reliability assessment conducted by the PMA and solved by the micro-GA (mGA), developed in Chapter 4. The validation of the mGA is integrated in the design optimization process. Sect. 7.2.3, presents the solution of the local RBRDO problem, with reliability assessment executed by the RIA. The outer cycle of design optimization is solved by the MOGA-2D and the inner-cycle of reliability assessment by the hybrid micro-GA (HmGA), developed in Chapter 4. The HmGA is first validated against two alternative methods. The RBRDO results are presented after.

Sect 7.3 presents a study on the effects of different sources of uncertainty in feasibility robustness. A simplified fuselage-like shell structure is introduced as numerical example. The outer cycle of design optimization is solved by the MOGA-2D. Reliability assessment is performed by the PMA and solved by the mGA. Lastly, in Sect. 7.4, the analytical dimensional reduction of the uncertainty space based on the analytical solution of the Sobol' indices, developed in Chapter 6, is applied to the HmGA as a powerful and simple tool to improve significantly the performance of the algorithm. Both in terms of accuracy and efficiency. The results of the importance analysis of the random variables that define the uncertainty space, in reliability assessment, are also discussed.

Lastly, Chapter 8 concludes the thesis. On Sect. 8.1, the main theoretical and algorithmic contributions achieved in this work are summarized, as well as the main results obtained from the solution of the developed RBRDO models. On sect. 8.2, future research perspectives are presented.

References

[1]. Hull D, Clyne TW. An introduction to composite materials. 2nd ed. Cambridge: Cambridge University Press; 1996.

[2]. Smith WF. Principles of materials science and engineering. 3rd ed. New York: McGraw-fill; 1996.

[3]. Gürdal Z, Haftka RT, Hajela P. Design and optimization of laminated composite materials. New York: Willey & Sons; 1999.

[4]. Callister WD, Rethwisch DG. Materials science and engineering. 10th ed. Willey & Sons; 2018.

[5]. Castro PT, Marques AT. Tendencies and needs for R&D in composites. Final report of European Atomic Energy Community, contract Nº 320105; 1991.

[6]. Visconti CI, Esposito R, Langella A, Baldi A. Manufacturing report. Composites Manufacturing 1990; 1(1).

[7]. Conceição António CC. Optimização de estruturas em materiais compósitos de matriz polimérica. PhD thesis. Porto: Universidade do Porto; 1995.

[8]. Watkins RI. Multilevel optimum design of large laminated composite structures. PhD thesis; 1986.

[9]. Elhajjar R, Saponara VL, Anastasia M. Smart composites: mechanics and design. 1st ed. Boca Raton: CRC Press; 2014.

[10]. Eschenauer HA. Aspects on finding optimal layouts of structures made of composite materials. Proceedings of the ninth international conference on composites materials 1993; 7: 651-60.

[11]. Narayana N, Gopalakrishnan S, Ganguli R. Design optimization of composites using genetic algorithms and failure mechanism based failure criterion. Comp Struct 2008; 83: 354-67.

[12]. Conceição António CC. A hierarchical genetic algorithm for reliability based design of geometrically nonlinear composite structures. Comp Struct 2001; 54: 37-47.

[13]. Huang ZL, Jiang C, Zhang Z, Fang T, Han X. A decoupling approach for evidence-theory-based reliability design optimization. Struct Multidisc Optim 2017.

[14]. Yin H, Yu D, Xia B. Reliability-based topology optimization for structures using fuzzy set model. Comput. Methods Appl. Mech. Engrg 2018; 333: 197-217.

[15]. Melchers RE. Structural Reliability Analysis and Prediction. 2nd ed. Willey & Sons; 1999.

[16]. Beyer H-G, Sendhoff B. Robust Optimization – A Comprehensive Survey. Comput Method Appl M 2007; 196(33-34): 3190-218.

[17]. Valdebenito MA, Schueller GI. A survey on approaches for reliability-based optimization. Struct Multidisc Optim 2010;42:645-63.

[18]. Huang C, El Hami A, Radi B. Overview of Structural Reliability Analysis Methods - Part II: Sampling Methods Incertitudes et fiabilité des systèmes multiphysiques 2017;1(Optimisation et Fiabilité).

[19]. Schueller G. Efficient Monte Carlo simulation procedures in structural uncertainty and reliability analysis-recent advances. Struct Eng Mech 2009;32(1):1–20.

[20]. Hurtado JE, Ramírez J. The estimation of failure probabilities as a false optimization problem. Struct Saf 2013;45:1–9.

[21]. Rashki M, Miri M, Moghaddam MA. A new efficient simulation method to approximate the probability of failure and most probable point. Struct Saf 2012;39:22–9.

[22]. Ditlevsen O, Melchers RE, Gluver H. General multi-dimensional probability integration by directional simulation. Comput Struct 1990;36(2):355–68.

[23]. Melchers R. Structural system reliability assessment using directional simulation. Struct Saf 1994;16(1):23–37

[24]. Papaioannou I, Papadimitriou C, Straub D. Sequential importance sampling for structural reliability analysis. Struct Saf 2016;62:66–75.

[25]. Cornell CA. A proposal for a reliability-based code suitable for immediate implementation. Memorandum to members of ASCE Task Committee on Structural Safety 1967.

[26]. Ditlevsen O. Structural reliability analysis and the invariance problem. Technical report No 22, Solid Mechanics Division, University of Waterloo, Ontario, Canada, 1973.

[27]. Lind NC. An invariant second-moment reliability format. Paper No 113 Solid Mechanics Division, University of Waterloo, Ontario, Canada, 1972.

[28]. Hasofer AM, Lind NC. An Exact and Invariant First Order Reliability. J Eng Mech-ASCE 1974;100(EM1):111-21.

[29]. Rackwitzt R, Fiessler B. Structural Reliability under combined random load sequences. Comput Struct 1978;9:489-94.

[30]. Periçaro GA, Santos SR, Ribeiro AA, Matioli LC. HLRF–BFGS optimization algorithm for structural reliability. Appl Math Model 2015;39(7):2025-35.

[31]. Keshtegar B A. hybrid conjugate finite-step length method for robust and efficient reliability analysis. Appl Math Model 2015;45:226-37.

[32]. Keshtegar B, Meng Z. A hybrid relaxed first-order reliability method for efficient structural reliability analysis. Struct Saf 2017;45:84-93.

[33]. Ezzati G, Mammadov M, Kulkarni S. A new reliability analysis method based on the conjugate gradient direction. Struct Multidisc Optim 2015;51(1):89-98

[34]. Acar E. A reliability index extrapolation method for separable limit states. Struct Multidisc Optim 2016;53:1099-1111.

[35]. Bai YC, Han X, Jiang C, Bi RG. A response-surface-based structural reliability analysis method by using non-probability convex model. Appl Math Model 2014;38(15-16):3834-47.

[36]. Meng Z, Li G, Yang D, Zhan D. A new directional stability transformation method of chaos control for first order reliability analysis. Struct Multidisc Optim 2017;55(2):601-12.

[37]. Keshtegar B. Limited conjugate gradient method for structural reliability analysis. Eng Comput 2017;33(3):621-29.

[38]. Jiang C, Han S, Ji M, Han X. A new method to solve the structural reliability index based on homotopy analysis. Acta Mech 2015;226(4):1067-83.

[39]. Keshtegar B. Stability iterative method for structural reliability analysis using a chaotic conjugate map. Nonlinear Dyn 2016;84(4):2161-74.

[40]. Díaz J, Montoya MC, Hernández S. Efficient methodologies for reliability-based design optimization of composite panels. Adv Eng Softw 2016;93:8-21.

[41]. Zhao Q, Chen X, Ma Z, Lin Y. A Comparison of Deterministic, Reliability-Based Topology Optimization under Uncertainties. Acta Mech Solida Sin 2016;29(1).

[42]. Tsompanakis Y, Papadrakakis M. Large-scale reliability-based structural optimization. Struct Multidisc Optim 2004;26(6):429–40.

[43]. Zhou X-Y, Gosling PD, Ullah Z, Kaczmarczyk L, Pearce CJ. Exploiting the benefits of multi-scale analysis in reliability analysis for composite structures. Compos Struct 2016;155:197-212.

[44]. Dehmous H, Duco F, Karama M, Welemane H. Multilevel assessment method for reliable design of composite structures. Mech Adv Mater Struc 2017;24(6):449-57.

[45]. Chen J, Ge R, Wei J. Probabilistic optimal design of laminates using improved particle swarm optimization. Eng Optimiz 2008;40(8):695-708.

[46]. Conceição António C, Torres Marques A, Gonçalves JF. Reliability based design with a degradation model of laminated composite structures. Struct Optim 1996;12:16-28.

[47]. Conceição António C, Hoffbauer LN. Uncertainty assessment approach for composite structures based on global sensitivity indices. Compos Struct 2013;99:202-12.

[48]. Conceição António C, Hoffbauer LN. Uncertainty propagation in inverse reliability-based design of composite structures. Int J Mech Mater Des 2010; 6: 89-102.

[49]. Chen X, Hasselman TK. Reliability based structural design optimization for practical applications. AIAA 1997; 1403: 2724-32.

[50]. Du X, Chen W. Sequential Optimization and Reliability Assessment method for efficient probabilistic design. Journal of Mechanical Design 2004; 126: 225-33.

[51]. Cheng J, Li QS. Reliability analysis of structures using artificial neural network based genetic algorithms. Comput Methods Appl Mech Engrg 2008;197:3742-50.

[52]. Shao S, Murotsu Y. Approach to failure mode analysis of large structures. Probabilist Eng Mech 1999;14(1-2):169-77.

[53]. Deng L, Ghosn M, Shao S. Development of a shredding genetic algorithm for structural reliability. Struct Saf 2005;27:113-31.

[54]. Wang J, Ghosn M. Linkage-shredding genetic algorithm for reliability assessment of structural systems. Struct Saf 2005;27:49-72.

[55]. Conceição António C, Hoffbauer LN. Reliability-based design optimization and uncertainty quantification for optimal conditions of composite structures with non-linear behavior. Eng Struct 2017;153:479-90.

[56]. Conceição António CC, Hoffbauer LN. Bi-level dominance GA for minimum weight and maximum feasibility robustness of composite structures. Compos. Struct 2016;135: 83-95.

[57]. Henrichsen SR, Lindgaard E, Lund E. Robust buckling optimization of laminated composite structures using discrete material optimization considering “worst” shape imperfections. Thin-Walled Structures 2015;94: 624-635.

[58]. Li F, Sun G, Huang X, Rong J, Li Q. Multiobjective robust optimization for crashworthiness design of foam filled thin-walled structures with random and interval uncertainties, Engineering Structures 2015;88: 111-124.

[59]. Kalantari M, Dong C, Davies IJ. Multi-objective robust optimization of unidirectional carbon/glass fiber reinforced fiber hybrid composites under flexural loading. Compos Struct 2016;138: 264-275.

[60]. Zhou X-Y, Ruan X, Goslin PD. Robust design optimization of variable angle tow composite plates for maximum buckling load in the presence of uncertainties. Compos. Struct 2019;223.

[61]. Mourelatos ZP, Liang J. A Methodology for Trading-Off Performance and Robustness Under Uncertainty. J. Mech. Des 2006; 128(4): 856-863.

[62]. Conceição António C, Hoffbauer LN: An approach for reliability-based robust design optimization of angle-ply composites. Compos. Struct 2009; 90: 53-59.

[63]. Motta RS, Afonso SMB. An efficient procedure for structural reliability-based robust design optimization. Struct Multidisc Optim 2016; 54: 511-530.

[64]. Wang J, Ning K, Xu J, Li Z. Reliability-based robust design of wind turbine’s shrink disk. Proc IMechE Part C: J Mechanical Engineering Science 2017.

II
PRINCIPLES OF OPTIMIZATION

Summary

*

The discipline of Optimization is one of extensive application in several fields of society. Today, optimization has become an important interdisciplinary area between *mathematics*, *computer science*, *engineering*, *operations research* and *economics*. The current standards in engineering design optimization often require the optimization of complex physical systems, in which the objective functions and/or the constraint functions are nonlinear. Such problems fall into the category of *nonlinear programming*. This chapter provides the reader with a concise introduction to nonlinear programming and evolutionary computing. It is divided into two parts.

Sect. 2.1 starts with a formal definition of the *nonlinear programming* (NLP) *problem*. Sect. 2.1.1 introduces the *necessary optimality conditions* for constrained optimization. Sect 2.1.2 introduces *Lagrangian duality* and the *exterior penalty formulation* of the NLP problem. Sect. 2.1.3 presents a historical overview of numerical optimization methods. The preference towards evolutionary algorithms (EAs), in the detriment of (faster) classic optimization methods, is explained. The evolutionary metaphor applied to metaheuristics in evolutionary computing is presented, followed by the conceptual structure that forms the common basis for the different variants of EAs. The main components of EAs are introduced, followed by notes on genetic diversity, hybridization and constraints handling. At the end, the probabilistic convergence of EAs is studied.

Sect. 2.2 starts with a formal definition of the *multi-objective nonlinear programming problem*. Sect. 2.2.1 introduces *Pareto optimality* and the concept of *Pareto dominance* binary relation. Sect. 2.2.2 introduces multi-objective EAs (MOEAs), explains the preference for these algorithms and identifies the three main goals MOEAs are expected to achieve. The *sorting mechanisms* of historical MOEAs are discussed. Brief notes on the hybridization of MOEAs and constraints handling are presented.

2.1 Single-objective Nonlinear Programming

Consider the following definition for the nonlinear programming (NLP) problem.

Definition 2.1 (*NLP problem*) [1]: Let f, g_i, for $i = 1, \dots, m$, and h_j, for $j = 1, \dots, l$, be functions defined in $\mathbb{R}^N$, X be an open subset of $\mathbb{R}^N$ and $\mathbf{x} \in X$ a vector of components $x_1, \dots, x_N$. Then, the general single-objective NLP problem is stated as follows:

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0 \quad \text{for } i = 1, \dots, m \\ & h_j(\mathbf{x}) = 0 \quad \text{for } j = 1, \dots, l \\ & \mathbf{x} \in X \end{aligned} \tag{2.1}$$

In the terminology of optimization, the function f is called the *objective function*, the constraints $g_i(\mathbf{x}) \leq 0$, for $i = 1, \dots, m$, are called *inequality constraints* and $h_j(\mathbf{x}) = 0$, for $j = 1, \dots, l$, are called *equality constraints*. The set X is one defining upper and lower bounds on the search variables $\mathbf{x}$, and is called the *search space*. Although the inequality and equality constraints may impose more restrictive bounds to the values of the variables than X itself, it plays a useful role in the application of optimization algorithms. A vector $\mathbf{x} \in X$ satisfying all the imposed constraints is called a *feasible solution* and the set S of all such solutions is called the *feasible region*. The feasible solution(s) which solve the NLP problem is (are) called *optimal solution(s)*. If $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$, for all $\mathbf{x} \in S$, then $\bar{\mathbf{x}}$ is called a global minimum. If such condition is only satisfied in a neighborhood $B(\bar{\mathbf{x}})$ of $\bar{\mathbf{x}}$, then it is called a local minimum. A point is *strict optimal* if the previous condition is satisfied for $<$ [2].

Throughout the next subsections, the solution of this class of problems is discussed. For that, we start by introducing the algebraic statements under which optimality is achieved, the *optimality conditions*. We then proceed to the algorithmic solution of optimization problems. First, a brief discussion on the need for numerical methods, followed by a more detailed discussion on evolutionary algorithms.

2.1.1 Optimality conditions

Optimality conditions can be divided into necessary or sufficient conditions. *Necessary optimality conditions* are those which any solution of the NLP problem has to satisfy. *Sufficient optimality conditions* are those that, when satisfied, guaranty that a feasible solution is also an optimal solution. Under certain conditions, some necessary conditions may be sufficient as well [2].

Optimality conditions of constrained problems are an extension of those for unconstrained problems. Hence, let f be differentiable at a point $\bar{\mathbf{x}}$. A vector $\mathbf{d}$ is said to be a *descent direction* if the dot product $\nabla f(\bar{\mathbf{x}}) \cdot \mathbf{d} < 0$, such that $f(\bar{\mathbf{x}} + \lambda \mathbf{d}) < f(\bar{\mathbf{x}})$, for $\lambda > 0$. Then, if $\bar{\mathbf{x}}$ is a local minimum, it is a *first-order necessary condition* that no descent direction from $\bar{\mathbf{x}}$ exists, that is, $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$. If f is twice differentiable at $\bar{\mathbf{x}}$, it is a *second-order sufficient condition* for $\bar{\mathbf{x}}$ to be a strict local minimum that $\nabla f(\bar{\mathbf{x}}) = \mathbf{0}$ and the Hessian matrix of f at $\bar{\mathbf{x}}$, $\mathbf{H}(\bar{\mathbf{x}})$, is positive definite [2].

In constrained problems, optimality conditions are expected to reflect information about the constraints as well. Regarding the NLP problem in (2.1), the feasible region S is delimited by a set of nonlinear (hyper)surfaces. Small feasible movements cannot then be fully described by a set of feasible directions alone, as in linear programming problems. This observation follows, because there is the need to describe feasible movements along curved surfaces as well. Associated with the concept of feasible direction from a point, should be the concept of feasible arc (curve) from a point [3,4].

Let S be a nonempty set in $\mathbb{R}^N$ and $\bar{\mathbf{x}} \in S$. A feasible arc $\boldsymbol{\alpha}(\lambda)$, leading from $\bar{\mathbf{x}}$, is a collection of points, parametrized by $\lambda \in [0, \delta]$, with $\delta > 0$, such that $\boldsymbol{\alpha}(0) = \bar{\mathbf{x}}$ and $\boldsymbol{\alpha}(\lambda) \in S$. An arc is said differentiable if $\boldsymbol{\alpha}'(\lambda) = (d\boldsymbol{\alpha}/d\lambda)$ exists. The cone of tangents to feasible arcs, at $\bar{\mathbf{x}}$, is given by [4]:

$$D = \{\boldsymbol{\alpha}'(0): \boldsymbol{\alpha}(\lambda) \in S, \forall \lambda \in [0, \delta], \delta > 0\} \tag{2.2}$$

Additionally, given a function f in $\mathbb{R}^N$, the set of all tangents to arcs improving f, from $\bar{\mathbf{x}}$**,** is given by [1]:

$$F = \{\boldsymbol{\alpha}'(0): f(\boldsymbol{\alpha}(\lambda)) < f(\boldsymbol{\alpha}(0)), \forall \lambda \in [0, \delta], \delta > 0\} \tag{2.3}$$

Analogously to unconstrained problems, it follows a *geometric optimality condition*.

Proposition 2.1 (*geometric optimality condition)* [1]: Let S be a nonempty set in $\mathbb{R}^N$ and $\bar{\mathbf{x}} \in S$. If $\bar{\mathbf{x}}$ is a local minimum, it is a necessary optimality condition that every improving arc is not a feasible arc, that is, $F \cap D = \emptyset$.

Although an important statement, the previous optimality condition is hard to verify, because in the majority of problems it is not possible to obtain an explicit representation of all feasible arcs, from $\bar{\mathbf{x}}$. We now seek a more explicit statement of the sets F and D that translates the geometric optimality condition into a more useful algebraic statement, using gradient information**.**

Let $S = \{\mathbf{x} \in X: g_i(\mathbf{x}) \leq 0, \text{for } i = 1, \dots, m \wedge h_j(\mathbf{x}) = 0, \text{for } j = 1, \dots, l\}$ and assume that f, g_i, h_j are continuously differentiable. Furthermore, let $I = \{i: g_i(\bar{\mathbf{x}}) = 0\}$ be the index set for the *active* inequality constraints. Given a feasible point $\bar{\mathbf{x}} \in S$, for any $\boldsymbol{\alpha}'(0) \in F$, f is a decreasing function of λ. Thus, by the chain rule of differentiation, it follows that $(df/d\lambda)_{\lambda=0} < 0 \Leftrightarrow \nabla f(\bar{\mathbf{x}}) \cdot \boldsymbol{\alpha}'(0) < 0$. Regarding the imposed constraint set, consider for now only feasible movements from points on the boundary of S towards its interior. Therefore, for $i \in I$, g_i are decreasing functions of λ and it follows $(dg_i/d\lambda)_{\lambda=0} < 0 \Leftrightarrow \nabla g_i(\bar{\mathbf{x}}) \cdot \boldsymbol{\alpha}'(0) < 0$, for any $\boldsymbol{\alpha}'(0) \in D$. Similarly, h_j are constant functions of λ and it follows $\left(dh_j/d\lambda\right)_{\lambda=0} = 0 \Leftrightarrow \nabla h_j(\bar{\mathbf{x}}) \cdot \boldsymbol{\alpha}'(0) = 0$, for any $\boldsymbol{\alpha}'(0) \in D$ [4].

An algebraic characterization of the sets F and D is given as follows [1]:

$$F_0 = \{\boldsymbol{\alpha}'(0): \nabla f(\bar{\mathbf{x}}) \cdot \boldsymbol{\alpha}'(0) < 0\}$$

$$G_0 = \{\boldsymbol{\alpha}'(0): \nabla g_i(\bar{\mathbf{x}}) \cdot \boldsymbol{\alpha}'(0) < 0, i \in I\}$$

$$H_0 = \{\boldsymbol{\alpha}'(0): \nabla h_j(\bar{\mathbf{x}}) \cdot \boldsymbol{\alpha}'(0) = 0, j = ,1 \dots, l\} \tag{2.4}$$

where $D_0 = G_0 \cap H_0$ may be interpreted as the cone of all tangents to feasible arcs, at $\bar{\mathbf{x}}$, pointing to the interior of S. Notice that the exclusion of feasible movements over the boundary of S allows us to refrain from making further considerations about the active inequality constraints.

The geometric optimality condition is still of difficult application because the set of all $\boldsymbol{\alpha}'(0)$ is still unknown in most NLP problems. However, by definition, it follows that $H_0 \subseteq N(\nabla \mathbf{h}(\bar{\mathbf{x}})) = \{\mathbf{d}: \nabla \mathbf{h}(\bar{\mathbf{x}})\mathbf{d} = \mathbf{0}\}$, where $N(\nabla \mathbf{h}(\bar{\mathbf{x}}))$ is the null space of the Jacobian matrix $\nabla \mathbf{h}(\bar{\mathbf{x}})$. Since it is easy to compute $\nabla \mathbf{h}(\bar{\mathbf{x}})$ and to generate its null space, the following condition is fundamental for the characterization of local minima in NLP problems [4].

Proposition 2.2 ***(regularity condition)* [4]:** Let $\nabla \mathbf{h}(\bar{\mathbf{x}})$ be the $(l \times N)$ Jacobian matrix, whose jth row is $\nabla h_j(\bar{\mathbf{x}})^t$. Given $\bar{\mathbf{x}} \in S$, if $\nabla h_j(\bar{\mathbf{x}})$, for $j = 1, \dots, l$ are linearly independent, then $H_0 = N(\nabla \mathbf{h}(\bar{\mathbf{x}}))$.

The regularity condition implies that no $\nabla h_j(\bar{\mathbf{x}})$ is colinear and that any tangent vector to a feasible arc, on the equality constraint surface, is simultaneously orthogonal to all $\nabla h_j(\bar{\mathbf{x}})$, for $j = 1, \dots, l$ [3,4]. Hence, in such case, it follows $D_0 = \{\mathbf{d}: \nabla g_i(\bar{\mathbf{x}}) \cdot \mathbf{d} < \mathbf{0}, i \in I \wedge \mathbf{d} \in N(\nabla \mathbf{h}(\bar{\mathbf{x}}))\}$.

It is easily demonstrated that, under the regularity condition, $F_0 \subseteq F$ and $D_0 \subseteq D$ and as a consequence $F \cap D = \emptyset \Rightarrow F_0 \cap D_0 = \emptyset$, for optimality [1]. We are now in a good position to reduce the geometric necessary optimality condition to a statement in terms of the gradients of the objective and constraint functions. Consider the following.

Proposition 2.3 [1]: Let $\bar{\mathbf{x}}$ be a regular point and suppose it is a local minimum. Furthermore, let **A** be the matrix whose rows are $\nabla f(\bar{\mathbf{x}})^t$ and $\nabla g_i(\bar{\mathbf{x}})^t$, for $i \in I$, and **C** the matrix whose rows are $\nabla h_j(\bar{\mathbf{x}})^t$, for $j = 1, .., l$. Then, $F_0 \cap D_0 = \emptyset$ if and only if that the system:

$$\mathbf{Ad} < \mathbf{0}, \quad \mathbf{Cd} = \mathbf{0} \tag{2.5}$$

has no solution.

A class of theorems, known as *Theorems of the Alternative*, establishes the relation between systems not having a simultaneous solution. One such theorem is *Motzkin's Transposition Theorem*.

Theorem 2.1 ***(Motzkin's Transposition Theorem)*** **[5]:** Let **A**, **B** and **C** be real constant matrices and **A** be nonempty. Then, either the system:

$$\mathbf{Ay} < \mathbf{0} \quad \mathbf{By} \leq \mathbf{0} \quad \mathbf{Cy} = \mathbf{0} \tag{2.6}$$

has a solution $\bar{\mathbf{y}}$, or the system:

$$\mathbf{A}^t\mathbf{z_1} + \mathbf{B}^t\mathbf{z_2} + \mathbf{C}^t\mathbf{z_3} = \mathbf{0}, \quad \mathbf{z_1} \geq \mathbf{0}, \quad \mathbf{z_1} \neq \mathbf{0}, \quad \mathbf{z_2} \geq \mathbf{0} \tag{2.7}$$

has a solution $\bar{\mathbf{z}}_\mathbf{1}$, $\bar{\mathbf{z}}_\mathbf{2}$, $\bar{\mathbf{z}}_\mathbf{3}$, but never both.

It follows an algebraic statement of the necessary optimality conditions of the NLP problem, due to Fritz John [6] and generalized by Mangasarian and Fromovitz [7] to include equality constraints.

Theorem 2.2 ***(Fritz John necessary optimality conditions)*** **[7]:** Let S be a nonempty set in $\mathbb{R}^N$, $\bar{\mathbf{x}} \in S$ and $I = \{i: g_i(\bar{\mathbf{x}}) = 0\}$. Furthermore, assume that f, g_i, for $i = 1, \dots, m$, and h_j, for $j = 1, \dots, l$ are continuously differentiable, at $\bar{\mathbf{x}}$. If $\bar{\mathbf{x}}$ is a local minimum of the NLP problem, then there exist vectors $\mathbf{u} = (u_0, u_1, \dots, u_m)$ and $\mathbf{v} = (v_1, \dots, v_l)$, such that:

$$\begin{aligned} u_0\nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} u_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{j=1}^{l} v_j \nabla h_j(\bar{\mathbf{x}}) &= \mathbf{0} \\ u_i g_i(\bar{\mathbf{x}}) &= 0 \quad \text{for } i = 1, .., m \\ \mathbf{u} &\geq \mathbf{0} \\ (\mathbf{u}, \mathbf{v}) &\neq \mathbf{0} \end{aligned} \tag{2.8}$$

Proof. If $\nabla h_j(\bar{\mathbf{x}})$ are linearly dependent, there is a vector $\mathbf{v} \neq \mathbf{0}$ such that $\sum_{j=1}^{l} v_j \nabla h_j(\bar{\mathbf{x}}) = \mathbf{0}$. Letting u_0 and u_i be equal to zero, the optimality conditions hold trivially. Now, if $\nabla h_j(\bar{\mathbf{x}})$ are linearly independent, $\bar{\mathbf{x}}$ is a regular point and Proposition 2.3 holds. Thus, by Theorem 2.1, with **B** empty, there exist vectors $\bar{\mathbf{z}}_\mathbf{1} = \mathbf{u}$, with $u_i \geq 0$, for $i \in I$, and $\bar{\mathbf{z}}_\mathbf{3} = \mathbf{v}$, such that (2.7) holds. The condition $u_i g_i(\bar{\mathbf{x}}) = 0$, for $i = 1, \dots, m$, follows immediately by letting $u_i = 0$, for $i \notin I$. ■

Any point satisfying the Fritz John (FJ) conditions is called a FJ point. Regarding the constraint set, apart from the linear dependence of the gradients $\nabla h_j(\bar{\mathbf{x}})$, there are other circumstances leading to the triviality of the FJ conditions. Generally, the FJ conditions will hold trivially for any feasible point satisfying $F_0 \cap D_0 = \emptyset$. For any regular $\hat{\mathbf{x}} \in S$, if the gradients $\nabla g_i(\hat{\mathbf{x}})$, for $i \in I$, are linearly dependent, then, by definition, the system $\sum_{i \in I} u_i \nabla g_i(\hat{\mathbf{x}}) = \mathbf{0}$ has a solution with $\mathbf{u} \geq \mathbf{0}$ and $\mathbf{u} \neq \mathbf{0}$. Thus, by Theorem 2.1, the system $\nabla g_i(\bar{\mathbf{x}}) \cdot \mathbf{d} < 0$, for $i \in I$, has no solution, which implies $G_0 = \emptyset \Rightarrow D_0 = \emptyset \Rightarrow F_0 \cap D_0 = \emptyset$. In such cases, u_i, for $i \in I$, can be any positive real and $(u_0, \mathbf{v}) = \mathbf{0}$ and the FJ conditions will hold trivially. It means that the optimality condition will not be written in terms of the gradient of f, since $u_0 = 0$ [1].

Motivated by the previous observations, we are led to state sufficient conditions for $u_0 > 0$. These are called *Constraint Qualifications* (CQs). There are essentially two types of CQs. The first type makes stronger assumptions (easier to verify) but are weaker (less applicable), assuming special characteristics of the constraint set [8]. A classic example is the *nondegeneracy condition* (or, Linear Independence CQ (LICQ)).

Proposition 2.4 ***(nondegeneracy condition)*** **[1,8]:** Let $\bar{\mathbf{x}} \in S$ be a regular point. If the gradient vectors $\nabla g_i(\bar{\mathbf{x}})$, for $i \in I$, are linearly independent, then $G_0 \neq \emptyset \Rightarrow D_0 \neq \emptyset$.

This result follows from Theorem 2.1. The second type of CQs characterizes the cone of feasible directions, at $\bar{\mathbf{x}}$. These make weaker assumptions and are therefore stronger.

By Theorem 2.1, if $\mathbf{A} = \nabla f(\bar{\mathbf{x}})$, $\mathbf{B}$ is the matrix whose rows are $\nabla g_i(\bar{\mathbf{x}})^t$, for $i \in I$, and $\mathbf{C}$ is the matrix whose rows are $\nabla h_j(\bar{\mathbf{x}})^t$, for $j = 1, \dots, l$, then, by (2.7), $u_0 > 0$. Hence, let C be a cone and $G' = \{\mathbf{d}: \nabla g_i(\bar{\mathbf{x}}) \cdot \mathbf{d} \leq 0, i \in I\}$ a closed cone, such that $\text{int}(G') \cap H_0 \subseteq C \subseteq G' \cap H_0$. Then, it is a sufficient condition for $u_0 > 0$ that $C = G' \cap H_0$ [1,8]. It means that certain elements of the boundary of G' are added to $\text{int}(G')$ to define C and that the CQ is satisfied when C contains all the elements of the interior and boundary of $G' \cap H_0$. Different CQs arise from different definitions of C. We consider two of them.

Proposition 2.5 ***(Mangasarian-Fromovitz CQ)*** **[1,7]:** Let $\bar{\mathbf{x}} \in S$. It is sufficient for $u_0 > 0$ that there exists a $\mathbf{d} \in \mathbb{R}^N$ such that $D_0 \neq \emptyset$ and $\bar{\mathbf{x}}$ is a regular point.

This CQ follows from the fact that, since $D_0 = \text{int}(G') \cap H_0$, then $\text{cl}(D_0) = G' \cap H_0 \Leftrightarrow D_0 \neq \emptyset$ [1]. Hence, in this case, $C = \text{cl}(D_0)$. A stronger CQ with weaker assumptions follows.

Proposition 2.6 ***(Kuhn-Tucker CQ)*** **[8,9]:** Let $\bar{\mathbf{x}} \in S$. It is sufficient for $u_0 > 0$ that there exists a $\mathbf{d} \in \mathbb{R}^N$ such that $D \neq \emptyset$ and $D = G' \cap H_0$.

In this case, $C = D$. As a final comment, notice that by definition $D_0 \subseteq D \subseteq G' \cap H_0$. Also, notice that the LICQ is a particular case of the MFCQ. Therefore, it is concluded that LICQ ⇒ MFCQ ⇒ KTCQ.

The Karush-Kuhn-Tucker (KKT) necessary optimality conditions then follow.

Theorem 2.3 ***(KKT necessary optimality conditions)*** **[9]:** Let S be a nonempty set in $\mathbb{R}^N$ and $\bar{\mathbf{x}} \in S$. Furthermore, assume that f, g_i, for $i = 1, \dots, m$, and h_j, for $j = 1, \dots, l$ are continuously differentiable, at $\bar{\mathbf{x}}$. Further assume that the KTCQ holds true. If $\bar{\mathbf{x}}$ is a local minimum of the NLP problem, then there exist vectors $\bar{\mathbf{u}} = (\bar{u}_1, \dots, \bar{u}_m)$ and $\bar{\mathbf{v}} = (\bar{v}_1, \dots, \bar{v}_l)$, such that:

$$\begin{aligned} \nabla f(\bar{\mathbf{x}}) + \sum_{i=1}^{m} \bar{u}_i \nabla g_i(\bar{\mathbf{x}}) + \sum_{j=1}^{l} \bar{v}_j \nabla h_j(\bar{\mathbf{x}}) &= \mathbf{0} \\ \bar{u}_i g_i(\bar{\mathbf{x}}) &= 0 \quad \text{for } i = 1, .., m \\ \bar{\mathbf{u}} &\geq \mathbf{0} \end{aligned} \tag{2.9}$$

Proof. By the imposition of the KTCQ, we have $F_0 \cap G' \cap H_0 = \emptyset$; that is, the system $\nabla f(\bar{\mathbf{x}}) \cdot \mathbf{d} < \mathbf{0}$, $\mathbf{Bd} \leq \mathbf{0}$ and $\mathbf{Cd} = \mathbf{0}$ has no solution, where $\mathbf{B}$ is the matrix whose rows are $\nabla g_i(\bar{\mathbf{x}})^t$, for $i \in I = \{i: g_i(\bar{\mathbf{x}}) = 0\}$, and $\mathbf{C}$ the matrix whose rows are $\nabla h_j(\bar{\mathbf{x}})^t$, for $j = 1, \dots, l$. Thus, by Theorem 2.1, there exist $\bar{z}_1 = u_0$, with $u_0 > 0$, $\bar{\mathbf{z}}_\mathbf{2} = \mathbf{u}$, with $u_i \geq 0$, for $i \in I$, and $\bar{\mathbf{z}}_\mathbf{3} = \mathbf{v}$, such that (2.7) holds, and $\bar{u}_0 = 1$, $\bar{\mathbf{u}} = \mathbf{u}/u_0$ and $\bar{\mathbf{v}} = \mathbf{v}/u_0$. The condition $u_i g_i(\bar{\mathbf{x}}) = 0$, for $i = 1, \dots, m$, follows by letting $u_i = 0$, for $i \notin I$. ■

Any point $\bar{\mathbf{x}}$, satisfying the KKT conditions is said a KKT point and the vectors $\bar{\mathbf{u}}$ and $\bar{\mathbf{v}}$ are called Lagrange multipliers. This section ends with two thoughts about the presented optimality conditions. The first thought is that not every point satisfying the FJ or the KKT conditions is optimal, but rather that optimal points should at least satisfy one of them. Both optimality conditions are limited by the constraint set and the kind of constraints. The KKT require a certain smoothness of the constraint set, while the FJ conditions don't. Hence, while the FJ end up being trivially satisfied, the KKT may be only partially satisfied, depending on the imposed CQ, or not satisfied at all. Not all optimal points are KKT points, but usually such points are FJ points. The second thought is that under suitable convexity assumptions both FJ and KKT conditions are also sufficient.

2.1.2 Notes on Lagrangian duality and Penalty

In the previous section, the necessary optimality conditions of the NLP problem were introduced from a geometric point-of-view. Focusing now on the KKT conditions, we briefly present an alternative perspective (the Lagrangian duality) and explore why these conditions can fail to be satisfied. To overcome the limitations of Lagrangian duality, the Penalty formulation of the NLP problem is introduced and an important result is stated, which will prove to be very useful in the application of Evolutionary Algorithms to constrained optimization problems.

Associated with the NLP problem (2.1), the *primal problem*, there is another NLP problem, the *dual problem.* In particular, we are interested in the so called *Lagrangian dual problem*, defined after $\phi(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \sum_{i=1}^{m} u_i g_i(\mathbf{x}) + \sum_{j=1}^{l} v_j h_j(\mathbf{x})$, known as the *Lagrangian function* [1].

Definition 2.2 ***(Lagrangian dual problem)*** **[1]:** Let f, g_i, for $i = 1, \dots, m$, and h_j, for $j = 1, \dots, l$, be functions in $\mathbb{R}^N$ and X be an open subset of $\mathbb{R}^N$. The Lagrangian dual problem is stated as follows:

$$\begin{aligned} \sup \quad & \theta(\mathbf{u}, \mathbf{v}) = \inf\{\phi(\mathbf{x}, \mathbf{u}, \mathbf{v}): \mathbf{x} \in X\} \\ \text{subject to} \quad & \mathbf{u} \geq \mathbf{0} \end{aligned} \tag{2.10}$$

where $\mathbf{u} \geq \mathbf{0}$ means the components of $\mathbf{u}$ are greater than or equal to 0. The process of integrating the constraints into the objective functions using Lagrange multipliers is referred to as dualization. Duality theory provides the relationships between the primal and dual problems. A fundamental result is the *weak duality theorem.*

Theorem 2.4 ***(weak duality theorem)*** **[1]:** Let $\mathbf{x}$ be a feasible solution to the primal NLP problem and $(\mathbf{u}, \mathbf{v})$ be a feasible solution of the Lagrangian dual problem. Then, $f(\mathbf{x}) \geq \theta(\mathbf{u}, \mathbf{v})$.

As a consequence, it is straightforward to conclude that $\inf\{f(\mathbf{x}): \mathbf{x} \in S\} \geq \sup\{\theta(\mathbf{u}, \mathbf{v}): \mathbf{u} \geq \mathbf{0}\}$. That is, the optimal objective value of the primal problem is always greater than or equal to the optimal objective value of the dual problem. In the case a strict inequality holds true, it is said to exist a *duality gap*, between the optimal objectives of the primal and dual problems. Moreover, given $\bar{\mathbf{x}} \in S$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, such that $\bar{\mathbf{u}} \geq 0$, if $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}}, \bar{\mathbf{v}})$, then $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}}, \bar{\mathbf{v}})$ are solutions of the primal and dual problems respectively and there is no duality gap [1].

An interesting feature of $\theta(\mathbf{u},\mathbf{v})$ is that it is always concave [1]. So, the maximization of θ becomes an attractive way to solve the primal problem, if there exists a null duality gap. We ask: under which conditions is the duality gap null? To answer such question, consider the following problem.

Definition 2.3 ***(saddle value problem)*** **[9]:** To find vectors $\bar{\mathbf{x}}$ and $(\bar{\mathbf{u}},\bar{\mathbf{v}})$, such that:

$$\phi(\bar{\mathbf{x}},\mathbf{u},\mathbf{v}) \le \phi(\bar{\mathbf{x}},\bar{\mathbf{u}},\bar{\mathbf{v}}) \le \phi(\mathbf{x},\bar{\mathbf{u}},\bar{\mathbf{v}}) \quad ,\forall\, \mathbf{x}\in X, \mathbf{u}\ge 0 \tag{2.11}$$

A solution $(\bar{\mathbf{x}},\bar{\mathbf{u}},\bar{\mathbf{v}})$ of the saddle value problem is said a saddle point of the Lagrangian function. Also, it follows that a saddle point is such that $\phi(\bar{\mathbf{x}},\bar{\mathbf{u}},\bar{\mathbf{v}}) = \min\{\phi(\mathbf{x},\bar{\mathbf{u}},\bar{\mathbf{v}}): \mathbf{x}\in X\} = f(\bar{\mathbf{x}})$, given that $\bar{\mathbf{x}}\in S$ and $\bar{u}_i g_i(\bar{\mathbf{x}}) = 0$, for $i = 1,\dots,m$, and thus $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}},\bar{\mathbf{v}})$. Kuhn and Tucker [9] gave the necessary conditions for a saddle point.

Proposition 2.7 [9]: Let $\phi(\mathbf{x},\mathbf{u},\mathbf{v})$ be continuously differentiable, at $(\bar{\mathbf{x}},\bar{\mathbf{u}},\bar{\mathbf{v}})$. The following conditions are necessary for a saddle point:

$$\bar{\mathbf{x}}\in X \quad ,\nabla_{\mathbf{u}}\phi(\bar{\mathbf{x}},\bar{\mathbf{u}},\bar{\mathbf{v}}) \le \mathbf{0} \quad ,\nabla_{\mathbf{v}}\phi(\bar{\mathbf{x}},\bar{\mathbf{u}},\bar{\mathbf{v}}) = \mathbf{0} \tag{2.12}$$

$$\nabla_{\mathbf{x}}\phi(\bar{\mathbf{x}},\bar{\mathbf{u}},\bar{\mathbf{v}}) = \mathbf{0} \quad ,\nabla_{\mathbf{u}}\phi(\bar{\mathbf{x}},\bar{\mathbf{u}},\bar{\mathbf{v}})\cdot\bar{\mathbf{u}} = \mathbf{0} \quad ,\mathbf{u}\ge\mathbf{0} \tag{2.13}$$

where $\nabla_{\mathbf{x}}\phi$, $\nabla_{\mathbf{u}}\phi$ and $\nabla_{\mathbf{v}}\phi$ are the gradient vectors of ϕ, with respect to $\mathbf{x}$, $\mathbf{u}$ and $\mathbf{v}$, respectively.

From the previous statement, it is immediate to show that the KKT conditions and the necessary saddle point conditions, for the Lagrangian function, are equivalent. Notice that, while conditions (2.12) guarantee that $\bar{\mathbf{x}}\in S$, conditions (2.13) are equivalent to (2.9). That is, the necessary saddle point conditions are necessary optimality conditions for the primal NLP problem and the KKT conditions are also necessary saddle point conditions, for the Lagrangian function. Hence, the KKT conditions are necessary conditions for $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}},\bar{\mathbf{v}})$. However, they are not sufficient for null duality gap, just like they are not sufficient for optimality.

It is then concluded that the KKT conditions guarantee the optimality of $\bar{\mathbf{x}}\in S$ if and only if $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}},\bar{\mathbf{v}})$. Or, equivalently, if and only if $(\bar{\mathbf{x}},\bar{\mathbf{u}},\bar{\mathbf{v}})$ is a saddle point for the Lagrangian function. A sufficient condition for null duality gap is given by the *strong duality theorem*, also known as *KKT sufficient optimality conditions.*

Theorem 2.5 ***(strong duality theorem)*** **[1]:** Let S be a nonempty set in $\mathbb{R}^N$. Suppose $\bar{\mathbf{x}}\in S$ satisfies the KKT conditions, that f, g_i, for $i\in I$, are convex at $\bar{\mathbf{x}}$ and that h_j, for $j = 1,\dots,l$, are affine (i.e. of the form $h_j(\mathbf{x}) = \mathbf{a}\mathbf{x} - b$), with $\bar{\mathbf{v}}\ne\mathbf{0}$. Then, $f(\bar{\mathbf{x}}) = \theta(\bar{\mathbf{u}},\bar{\mathbf{v}})$ and $(\bar{\mathbf{x}},\bar{\mathbf{u}},\bar{\mathbf{v}})$ is a saddle point for the Lagrangian function.

In particular, if the convexity assumptions are restricted to a neighborhood of $\bar{\mathbf{x}}$, then $\bar{\mathbf{x}}$ is a local minimum of the primal problem. The previous theorem shows that the KKT conditions are only sufficient for optimality under strong convexity assumptions on the objective and constraint functions. However, most NLP problems are not convex problems. Indeed, Lagrangian duality is limited by the duality gap phenomenon. To see how, please consider the following brief geometric interpretation of the Lagrangian dual problem.

Without loss of generality, consider only one equality constraint, such that the primal problem is $\min\{f(\mathbf{x}): h(\mathbf{x}) = 0, \mathbf{x}\in X\}$. Now, let $H = \{(y,z): y = h(\mathbf{x}), z = f(\mathbf{x}), \text{for } \mathbf{x}\in X\}$ be the image of X in the space of coordinates (y,z). If h and f are both convex functions, then it is easily shown that H is a convex set. Otherwise, it is not convex. Figure 2.1 shows the representation of a convex problem, (a), and a nonconvex one, (b). The primal problem asks to find the point with the minimum ordinate z, for which $y = 0$. Such point is $(\bar{y},\bar{z})$. Now, let $v\in\mathbb{R}$ and consider the straight line of equation $z + vy = \alpha$. For its part, the Lagrangian dual problem is $\max\{\min\{z + vy, \mathbf{x}\in X\}, v\in\mathbb{R}\}$.

In other words, it asks to find the slope $(-v)$ of the straight line supporting the set H from bellow, for which the intersection with the z-axis, α, is maximum. As shown in (a), of Figure 2.1, for convex problems, the optimal solution of the primal and dual problems is the same. That is, there is no duality gap. However, for nonconvex problems, in (b), it is not possible to have the same solution for both problems, because no straight line can support H, at point $(\bar{y}, \bar{z})$. In agreement with the preceding discussion, if the problem is nonconvex, there will not be a saddle point for the Lagrangian function and the KKT conditions are not sufficient for optimality. It is demonstrated graphically that the reason for this limitation is the fact that the Lagrangian dual approach can only offer linear support to the set H. As observed in (b), a nonlinear support of H would possibly overcome such limitation [1].

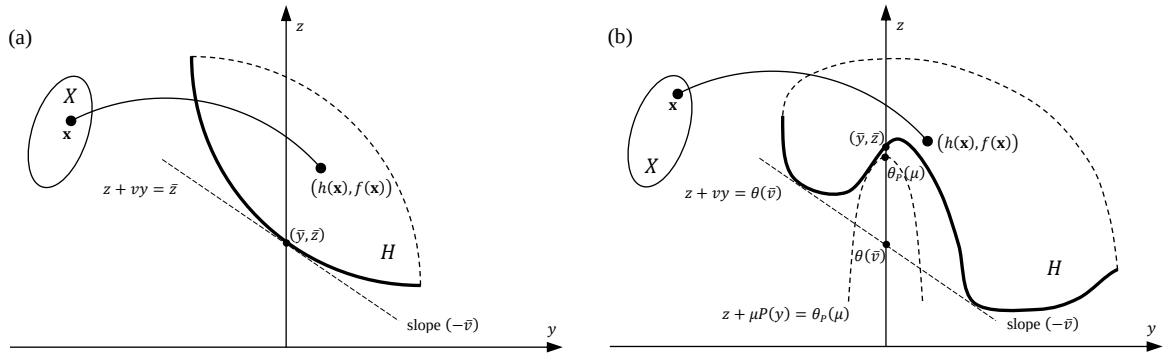

Figure 2.1: Geometric interpretation of Lagrangian dual and exterior penalty, for (a) convex problems and (b) nonconvex problems.

A generalization of Lagrangian duality, allowing for the nonlinear support of nonconvex sets H, is known as *penalty formulation* [1]. Associated with the *primal problem*, there is another NLP problem, the *penalty problem.* In particular, we are interested in the so called *exterior penalty problem*, defined after $\phi_P(\mathbf{x}, \mu) = f(\mathbf{x}) + \mu P(\mathbf{x})$, where $\mu \geq 0$ and $P(\mathbf{x}) = \sum_{i=1}^{m} \left[\max\left(0, g_i(\mathbf{x})\right)\right]^p + \sum_{j=1}^{l} \left|h_j(\mathbf{x})\right|^p$ is called *penalty function*, satisfying $P(\mathbf{x}) > 0$, if $\mathbf{x} \notin S$, and $P(\mathbf{x}) = 0$, if $\mathbf{x} \in S$.

Definition 2.4 ***(exterior penalty problem)*** **[1]:** Let $P(\mathbf{x})$, f, g_i, for $i = 1, \ldots, m$, and h_j, for $j = 1, \ldots, l$, be continuous functions in $\mathbb{R}^N$ and X be an open subset of $\mathbb{R}^N$. The exterior penalty problem is as follows:

$$\begin{aligned} \sup \quad & \theta_P(\mu) = \inf\{\phi_P(\mathbf{x}, \mu) : \mathbf{x} \in X\} \\ \text{subject to} \quad & \mu \geq 0 \end{aligned} \tag{2.14}$$

Notice that $\phi_P(\mathbf{x}, \mu)$ and $\theta_P(\mu)$ play an analogous role to that of $\phi(\mathbf{x}, \mathbf{u}, \mathbf{v})$ and $\theta(\mathbf{u}, \mathbf{v})$, allowing however for the nonlinear dualization of the primal problem. Similarly, it is also demonstrated that $\inf\{f(\mathbf{x}) : \mathbf{x} \in S\} \geq \sup\{\theta_P(\mu) : \mu \geq 0\}$. On the other hand, suppose that for each μ there exists an $\mathbf{x}_\mu \in X$ such that $\theta_P(\mu) = f(\mathbf{x}_\mu) + \mu P(\mathbf{x}_\mu)$. Then, $\theta_P(\mu)$ and $f(\mathbf{x}_\mu)$ are nondecreasing functions of μ and $P(\mathbf{x}_\mu)$ is a nonincreasing function of μ. Then, it follows the result.

Theorem 2.6 ***(Zangwill convergence theorem)*** **[1]:** Let $P(\mathbf{x})$, f, g_i and h_j be continuous functions and X be a compact subset of $\mathbb{R}^N$. Suppose the primal problem has a solution $\bar{\mathbf{x}} \in S$ and that for each μ there exists a $\mathbf{x}_\mu \in X$ such that $\theta_P(\mu) = f(\mathbf{x}_\mu) + \mu P(\mathbf{x}_\mu)$. Then,

$$\inf\{f(\mathbf{x}) : \mathbf{x} \in S\} = \sup\{\theta_P(\mu) : \mu \geq 0\} = \lim_{\mu \to \infty} \theta_P(\mu) \tag{2.15}$$

and $\mathbf{x}_\mu \to \bar{\mathbf{x}}$ as $\mu \to \infty$.

The theorem states that there exist convergent sequences $\{\mathbf{x}_\mu\}$ whose limit is $\bar{\mathbf{x}}$. Hence, the optimal solutions of the penalty and the primal problems can be made arbitrarily close. It requires X to be compact, as well, in opposition to the preceding considerations. Although initially written to prove the convergence properties of an algorithm solving a sequence of exterior penalty problems with sequentially increasing values of μ, this theorem is important in the definition of fitness (or merit) functions. Some classes of optimization algorithms use fitness functions to compare different solutions, in terms of their degree of optimality. As discussed in the next section, Evolutionary Algorithms are one of the classes that benefit from penalty formulations of the primal problem.

2.1.3 Evolutionary Algorithms

The optimality conditions presented so far can be used either to assess the optimality of particular sets of solutions or to determine candidate optimal solutions, by solving the systems in (2.12) and (2.13). In order to guarantee global (or local) optimality of the solutions, Theorem 2.5 states that strong convexity assumptions and a constraint qualification are required. However, engineering problems usually are nonlinear, nonconvex, high dimensional and may present several constraint functions. The resulting systems are difficult to solve. Sometimes, only a numerical solution exists. In addition, it is frequent that the functions and the variables do not have an explicit relationship, making such systems hard to define.

The applicability of the optimality conditions ends up being restricted to simpler, mainly academic, problems. Instead, efficient and elegant algorithms are available to solve the NLP problem directly. In times of lower computational power, *iterative optimization algorithms* were, and still are, an efficient alternative to find optimal solutions to some problems. These methods can be classified into two different categories: *direct search methods* and *indirect search methods*. The first handle the constraints explicitly, while the latter transform the original constrained problem into a sequence of penalty problems. Generally, these algorithms are based on an iterative process, as follows [2]:

$$\mathbf{x}^{t+1} = \mathbf{x}^t + \lambda^t \mathbf{s}^t \tag{2.16}$$

where $\mathbf{x}^t$ is the starting solution for the t-th iteration, $\mathbf{s}^t$ is the search direction vector, λ^t is the step length and $\mathbf{x}^{t+1}$ is the new solution from which the next iteration starts. Different methods are distinguished by the way both $\mathbf{s}^t$ and λ^t are calculated, but share a common goal: to sequentially improve $\mathbf{x}^t$, at each iteration, until convergence to a feasible solution. To this purpose, there are essentially two classes of methods: zeroth order methods and gradient-based methods (first or second orders). Ultimately, upon convergence, iterative optimization algorithms attempt to satisfy some optimality condition, fully or partially. Notable examples are Sequential Linear Programming, Sequential Quadratic Programming , Zoutendijk's method and Gradient projection method (*direct search methods*); Interior penalty function method, exterior penalty function method and Augmented Lagrangian method (*indirect search methods*); Powell's method , Cauchy's method, conjugate gradient method, Newton's method and quasi-Newton methods (*unconstrained search methods*). However, despite their efficiency, the ability of such methods to achieve global convergence is only guaranteed under narrow conditions. Indeed, only local convergence is expected. Moreover, numerical instability and divergence are known issues, particularly in high dimensional and nonlinear problems, with multiple optima and discontinuous domains. Complete surveys about iterative optimization algorithms can be found in [1,2,10].

At the present time, there is an increasing demand for metaheuristics to solve complex engineering NLP problems. Historically, metaheuristics have a parallel evolution to iterative optimization algorithms. Sorensen *et al*. [11] present a survey on the history of metaheuristics. According to the authors, in the field of optimization, the term "metaheuristic" may have two definitions:

- a set of concepts and strategies combined to form a high-level problem-independent algorithmic framework;
- a problem-specific implementation of an algorithm based on such a framework.

The advent of metaheuristics occurred in the 1960s, with the development of the first algorithms in the field of *artificial intelligence*, which mimicked the human ability to solve problems. Soon, in the field of optimization, new metaheuristics were developed emulating physical and natural phenomena. The paradigm was the *optimization* of some kind of concept that would serve as a metaphor to the actual optimization problem.

Evolutionary Algorithms (EAs) initially appeared with the intent of studying the phenomenon of natural evolution [12]. Researchers then understood that evolution itself is a powerful tool of nature to solve the environmental adaptation problems of species, over time. Evolution by natural selection tends to *minimize* the discrepancy between a specie's characteristics and the necessary attributes to survive on its environment, developing fitter individuals. This metaphor provides a simple, and yet very effective, framework to solve many kinds of problems, particularly combinatorial optimization problems. Several EAs appeared since then, each with its own interpretation of the evolutionary metaphor. Some of them are still relevant today and are the foundation of many recent EAs. Evolution Strategy and Evolutionary Programming [13] were developed as single-solution algorithms, relaying on a basic *mutation* operator applied to the solution, at each generation. Holland [12] later recognized the importance of a population of solutions and the positive effects of the *recombination* of solutions in the efficiency of the optimization process. His *schemata* theory was popularized by Goldberg [14] and originated the term "*Genetic Algorithm"* (GA), as an algorithm encompassing several concepts of natural evolution. Other relevant EA's are Estimation of a Distribution Algorithm [15], Differential Evolution [16] and Genetic Programming [17]. Other historical population-based metaheuristics, constructed after some metaphor, are Simulated Annealing, Ant Colony Optimization, Particle Swarm Optimization, Taboo Search and Scatter Search [11].

Currently, the development of metaheuristics in the field of optimization is focused on the combination of concepts from different algorithms and frameworks. In evolutionary computation, special attention has been given to the hybridization of EAs to create frameworks of higher level, capable of converging faster. Local search methods and heuristics are used to improve the population, exploring the knowledge that in the neighborhood of each solution there might be a superior one (local optimum), or to transform infeasible solutions into feasible ones. Niching techniques [18], coevolution of species or populations [13] are also amongst the most important frameworks applied to EAs. Such metaheuristics are known as *hybrid EAs*. A distinctive class of hybrid EAs are the *Memetic Algorithms* [19].

The development of new metaheuristics in the field of optimization is still problem-dependent and relies on the experience of the practitioner. Historically, metaheuristics were more of industrial and practical application, while optimality conditions and iterative optimization algorithms were developed and studied in academic environments. For that reason, scholars were skeptical about their value. Time eventually proved that metaheuristics are in fact valuable, and frequently more appropriate, to solve real-world optimization problems, but, unlike their counterparts, their theoretical study is much more recent. As a result, many new algorithms are developed based on new metaphors without knowing their actual ability to converge and, more importantly, under what conditions it may happen [11].

For that reason, this thesis is focused solely on EAs, in particular on the hybridization of GAs. These algorithms are between the few whose global convergence is actually studied and proven in several scientific publications and by different authors. Throughout the next subsections *the evolutionary metaphor* is introduced and its translation to the field of evolutionary computation is explained. A conceptual EA is introduced and the main evolutionary operators are described. Then, probabilistic

convergence (i.e. the limit behavior) is proven by a Markov Chain analysis of EAs and the conditions for global convergence are discussed. Finally, a brief discussion on the genetic diversity maintenance, exploitation of the search space and constraint handling is presented.

2.1.3.1 The evolutionary metaphor

Evolutionary Algorithms a class of population-based metaheuristics, whose design is inspired by observations about the natural evolution of species. Although there exist may kinds of EAs proposed in the literature, motivated by different interpretations of the evolutionary metaphor, all of them respect fully, or in part, the following set of frameworks [13,20]:

- species evolve in populations of individuals, which will compete to survive in an environment;
- each individual is defined by a composite of hereditary information (genotype) and by a composite of observable features (phenotype), which are determined by the genotype;
- the adaptation of individuals to the environment depends largely of their phenotypical characteristics;
- Darwin's *theory of natural selection* establishes that the better adapted individuals have greater probability of remaining in the population (physical survival) and reproduce (genetic survival);
- less adapted individuals, on average, perish faster and eventually are eliminated from the population;
- new individuals (offspring) are generated by genetic recombination of previous individuals (parents);
- individuals may suffer mutations on their genes, which will influence their phenotypes.

From the previous ideas, the following set of principles is organized, which will govern the action of EAs in general:

- evolution is defined as a *sequence of populations* generated iteratively. Each iteration is called a *generation*;
- the environment is defined by the problem to be solved, i.e. by the *search space*, including feasible and unfeasible search regions, and by the *objective function*;
- individuals in the environment represent *solutions in the search space*, i.e. realizations of the vector of search variables, **x**;
- individuals are coded by an ordered set of genes, written in a *specific alphabet*;
- the phenotype of individuals is obtained by a *genotype-phenotype mapping function*, assuming that the phenotype is influenced by both the genes and the environment;
- to each individual is associated a *fitness value*, as a measure of its adaptation to the environment. Fitness depends on the objective function and on the phenotype, among others;
- new individuals are generated by *variation operators*, which alter the population, corresponding to movements in the search space. Evolution is materialized by the action of *selection operators*, which choose groups of individuals for each stage of the evolutionary process;
- the inherent randomness of natural evolution is simulated by *random sampling processes*, making EAs stochastic optimization methods.

The traditional evolutionary model applied by EAs is called *artificial evolution* [20]: evaluate, variate and select populations of individuals (solutions), at each generation (iteration), with the purpose of finding the genetic configuration which maximizes a fitness function. The application of the evolutionary metaphor as framework to solve the NLP problem is what distinguishes EAs from other optimization metaheuristics. In this section, we explore the following conceptual structure of an EA.

```
Make initial population
Evaluate population
UNTIL stop DO
    Select parents from the population
    Genetic variation of parents and/or individuals
    Evaluate population
    Select survivors for the next generation
END
Output optimum solution
```

Figure 2.2: Conceptual structure of an EA.

In general, EAs share the same concepts, totally or partially. If one interprets the conceptual EA as the set of all concepts and heuristics materializing the evolutionary metaphor, then every EA can be defined as a subset of the conceptual EA. Each EA is thus characterized by its own set of *evolutionary operators*. In the field of evolutionary computation, an operator is a heuristic comprehending a series of stochastic and /or deterministic operations applied iteratively to the elements of the population. Being very versatile, EAs are allowed to be applied to many different classes of problems. However, it means that each operator has to be designed to work on a specific kind of *data structure*, which suites best the problem to be solved. In the next paragraphs, we present and discuss the most applied types of data structure and evolutionary operators. In general, the operators can be divided into *variation*, *selection* and *evaluation* operators.

Data Structure

To specify a representation for the search variables is a necessary step to allow EAs to generate and evaluate different configurations on the search space. For practical reasons, it is common to distinguish between the representation used to generate and manipulate search points and the representation used to evaluate such solutions. The first is the genotype and the latter the phenotype. The translation from one to another is made by a genotype-phenotype ($G \rightarrow P$) mapping function. However, the need for a map, in the mathematical sense, depends on the problem itself [20]. Generically, it is possible to divide the optimization problems into two categories: planning and numerical optimization problems, which differ in the nature of the respective variables. In planning optimization problems, the vector of search variables usually represents a sequence of integers (task identifiers) and the objective function represents a program simulating the planning, from the sequence, and returning a measure of quality (of time, cost, risk, etc.) [13]. In numerical optimization problems, the vector of search variables is comprised of real-valued scalars and the objective function a scalar function (implicit or explicit) to be optimized.

Problems with integer variables (Integer Programming) are discrete by nature and represent combinatorial optimization problems suitable to be solved by metaheuristics, like EAs. On the other hand, problems of real-valued variables (Linear or Nonlinear Programing) are associated with continuous search spaces. To be solved by EAs, traditionally it is applied a discretization of the search space, transforming the continuous problem into a combinatorial optimization problem. This is achieved by the application of a genotype-phenotype distinction, defined explicitly by a $G \rightarrow P$

mapping function. Today, there are also algorithmic solutions that allow EAs to work directly on a continuous search space. In this case, the genotype-phenotype distinction is avoided. The following examples introduce different types of data structure, based on some classic problems.

In *resource allocation problems*, it is given a set of N resources, each with a profit $p_i > 0$ and cost $w_i > 0$. The goal is to determine the configuration of resources maximizing the sum of the profits of the allocated resources, without exceeding an upper limit of the total cost value. The Knapsack problem is a classic example. In such problems, the search variables must indicate clearly which resources are allocated and which are not, having two possible values only. Boolean variables are ideal to deal with such problems. The solutions are represented by an *array of Nbinary variables*, where the value 1 at the i-th position indicates that the i-th resource is allocated and 0 means the opposite. For example, the binary string

$$\{0\ 1\ 1\ 0\ 0\ 1\}$$

represents a configuration where only the resources 2, 3 and 6 are allocated. Here, the genotype-phenotype distinction is purely conceptual. The binary string being the genotype, while its interpretation (the actual resource allocation) is the phenotype.

In *vehicle routing problems*, it is given a set of locations to be visited and the distances/costs between each pair of them. The goal is to determine the itinerary which minimizes the travel distance/cost, knowing that all the locations must be visited. The Travelling Salesman is a classic example. In this case, the search variables should reflect the order in which each location is visited, from a common starting and arrival point. A natural way of representing these solutions is by means of *permutations* (i.e. ordered arrays of integers), such as the following example, for 6 locations:

$$\{5\ 1\ 2\ 6\ 4\ 3\}$$

Again, the genotype-phenotype distinction is conceptual. In this case, the permutations are the genotype of the solutions, from which different interpretations, phenotypes, can be developed. One possibility is permutations establishing traveling sequences between consecutive values of the array. Another one is the value in each position of the array indicating the position of the respective location in the travelling plan. Still regarding the same problem, other coding schemes may be applied to describe the genotype, as well. For instance, permutations may be represented as *arrays of binary strings*, encoding integers, such that the previous permutation is represented by:

$$\{101, 001, 010, 110, 100, 011\}$$

This genotype-phenotype distinction requires an explicit $G \rightarrow P$ mapping function, such as [2]:

$$x_i = \sum_{k=0}^{l-1} 2^k b_k \quad , \forall x_i \in \mathbb{N} \tag{2.17}$$

where l is the length of the binary string and b_k its the k-th bit. The phenotype is defined by the calculated permutation of integers, accompanied by an appropriate interpretation. A similar genotype-phenotype distinction may be applied also to resource allocation problems. In such case, the algorithm must accept arrays of variable length, for both genotype and phenotype.

Regarding *real-valued problems*, these may be generalized by the NLP problem. Usually, to solve NLP problems, EAs consider *arrays of binary strings* to represent the genotype of solutions [14]. A genotype-phenotype distinction, defined explicitly by the following $G \rightarrow P$ mapping function [2]:

$$x_i = x_i^l + \frac{x_i^u - x_i^l}{2^l - 1} \sum_{k=0}^{l-1} 2^k b_k \quad , \forall x_i \in \mathbb{R} \tag{2.18}$$

allows to discretize the search space, transforming the NLP problem into a combinatorial optimization problem. In (2.18), l is the length of the binary string, b_k its k-th bit and x_i^u and x_i^l represent upper and lower bounds on the search variables x_i and may be understood as the numerical realization of the set X (see (2.1)), in the discretized problem. These are geometric constraints, usually referred to as *size constraints*. Notice that both the mapping functions in (2.17) and (2.18) contain genetic information of the solutions, as well as information of the environment (problem). In the first case, the number of bits is determined by the number of locations to visit. In the latter, x_i^u and x_i^l are parameters of the problem and l determines the resolution of the discretization of the search space. For the same size constraints, the larger the number of bits the smaller the incremental step between two consecutive values of the variables. Such behavior is translated by the following inequality [2]:

$$2^l \geq \frac{x_i^u - x_i^l}{\Delta x_i} \tag{2.19}$$

where Δx_i is the incremental step, associated with the search variable x_i. As referred previously, there are approaches, proposed in the literature, to apply EAs directly into continuous search spaces [21]. The advantage of these algorithms is not requiring a genotype-phenotype distinction, as traditional EAs do, although they need a very specific set of operators. EAs are also suitable to work with alternative data structures, beyond vectors and arrays. Depending on the problem, it might be better to work with matrices [22], trees [17], graphs [23], etc.

Independently of the kind of data structure, each evolutionary operator shall apply upon the population a set of actions reflecting the purpose for which they were conceived. Otherwise, in practice, the structure of the conceptual EA may not be satisfied. Acknowledging that this thesis concerns about NLP problems, throughout the following subsections, we will focus on the review of the main evolutionary operators applied to the resolution of real-valued optimization problems. In particular, those exploring a binary-to-real genotype-phenotype distinction, as described in (2.18). For completeness of this survey, references on the application of evolutionary operators to other kinds of data structure will be included, as well.

Fitness function

Let $\mathbb{S}$ be the genotype space of individuals $s \in \mathbb{S}$, with any kind of data structure and $X \subseteq \mathbb{R}^N$ be the phenotype space of variables $\mathbf{x} \in X$. Also, let $(G \to P): \mathbb{S} \to X$ be a genotype-phenotype mapping function and $f: X \to \mathbb{R}$ the objective-function. Then, a *fitness function* $F \equiv f \circ (G \to P): \mathbb{S} \to \mathbb{R}$ is a function mapping any configuration of individuals in a population into a positive quality measure.

Individuals are evaluated, at any generation, to decide whether or not they take part in the evolutionary process. Many different definitions for the fitness function can be found in the literature. The most common procedure is to assign a fitness that is computed exactly as, or directly from, the objective function [2]. It may also include other relevant informations, like the degree of violation of constraints or some measure of genetic diversity [24,25].

Genetic Variation Operators

Genetic variation operators are the heuristics responsible for producing changes in the population, generating new individuals. In practice, they are responsible for producing movements in the search space, searching for potentially more promising regions. Variation operators can be categorized as *recombination operators* and *mutation operators*.

Recombination operators

Genetic recombination, or crossover, is the operation that produces new solutions (*offspring*), from the crossing of genetic information of two or more solutions (*parents*), allowing the algorithm to exploit the search region formed between the parent solutions. The application of recombination operators is a distinctive trait of EAs, in relation to other metaheuristics. In particular, of GAs [14]. Differential Evolution [16] and modern formulations of Evolution Strategy [13] also apply their own recombination operators. On the other hand, Evolutionary Programming and early formulations of Evolution Strategy [13] give preference to the careful design of single-solution variation operators (mutation).

Regarding binary-coded GAs, the traditional recombination operator is the *one-point crossover* [14,20]. Given two parent strings, the operator randomly selects one point inside the strings (crossover point), with equal probability ($p_{sel} = 1/l$), and combines the "head" of a parent with the "tail" of the other, and *vice-versa*, generating two offspring solutions. Figure 2.3 depicts the process.

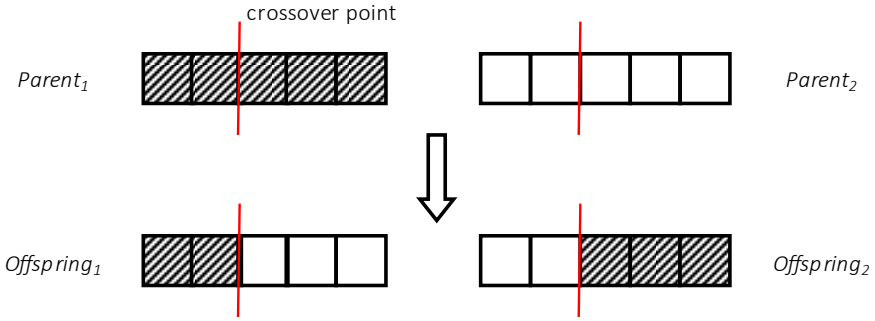

Figure 2.3: One-point Crossover.

The *two-point crossover* is also common [13]. Similarly, two different crossover points are randomly selected and the resulting (inner and outer) segments of genes combined to generate two offspring solutions. The process is shown in Figure 2.4. It is generalized by the *k-point crossover*, for any positive $k \in \mathbb{N}$, choosing k crossover points.

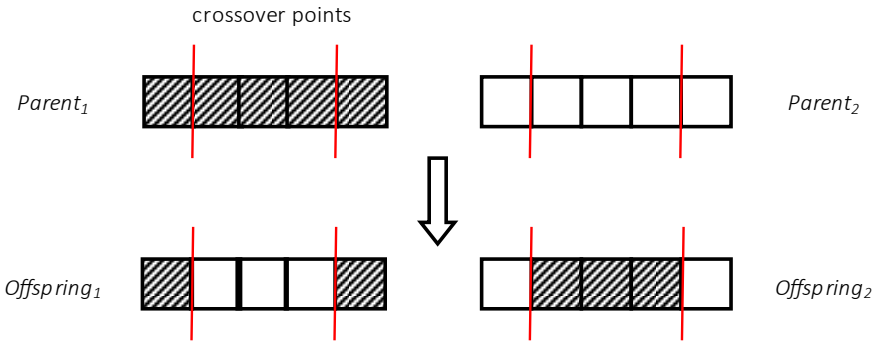

Figure 2.4: Two-point Crossover.

With the *uniform crossover* operator [26], the genes of each parent are compared individually and, if different, a random choice decides which parent a gene should be inherited from, with a given ratio r_{uc} (usually 0.5). After the first offspring is generated, the second is created using the inverse mapping, as shown in Figure 2.5. By applying other mixing ratios, it is then possible to bias the genetic information of one offspring towards one of the parents. Other recombination operators are the *half-uniform crossover* [27], the *parametrized uniform crossover* [28], the *line crossover* [13] and the *multi-parent crossover* [13].

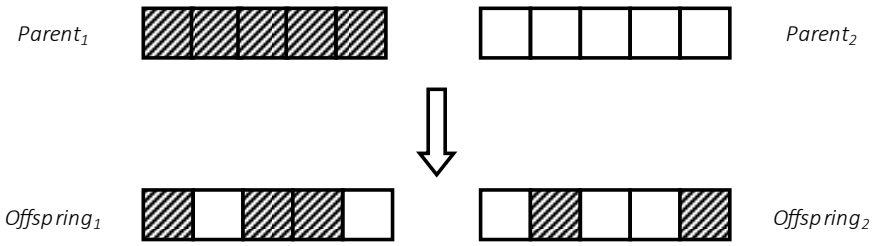

Figure 2.5: Uniform crossover. The array $\{\mathbf{0.2}, \mathbf{0.7}, \mathbf{0.1}, \mathbf{0.4}, \mathbf{0.9}\}$ of random numbers and $\boldsymbol{p} = \mathbf{0.5}$ were used to decide inheritance.

Regarding permutation-coded GAs, it is necessary that crossover operators accept permutations, as input, and return valid permutations, as output (e.g., without repetition of integers). The previous operators, developed to be applied to binary strings, do not guarantee such requirements The *partially mapped crossover* [29] and the *order crossover* [30] are two operators found in the literature. As for real-coded GAs, real-coded versions of the one-point, two-point and uniform crossovers are easily adapted from their binary-coded counterparts. Other recombination operators are designed to take advantage of the continuous search space, exploring the ability of the operator to converge or expand the evolution, or even to correlate changes on the search variables. Herrera *et al.* [32] and Eiben and Smith [33] present complete surveys on real-coded crossover operators.

Regarding other EAs, the *exponential crossover* and the *binomial crossover* [34] are of standard application in Differential Evolution, resembling the two-point crossover and the uniform crossover, respectively. Modern formulations of Evolution Strategy use real encoding and apply different operators for the recombination of parents. A common heuristic is to generate offspring solutions by linear combination of parents [13]. The coefficients of the combination may be equal to all the parents, obtaining an "average" offspring, or weighted by the parents' fitness.

Mutation operators

Mutation is the operation that produces new solutions by local and random changes of the genetic code of individuals. Its principle is analogous to that of biological mutation and is used to maintain, or introduce, genetic diversity in the population. Traditionally, the recombination of solutions has greater influence upon the evolution, while mutation acts more discreetly in the background, being applied less frequently [20]. It doesn't mean, however, that the mutation operator is less important. Indeed, as we shall conclude posteriorly, the mutation operator is necessary to achieve global convergence, in the long run. On the other hand, the recombination operator is responsible for accelerating the convergence rate of the algorithm.

The action of mutation operators is similar to movements in search space, starting from an initial solution. While recombination operators exploit regions of the search space limited by the genetic information of a set of parents, mutation operators explore the neighborhood around one solution. The extension of such neighborhood depends on the magnitude of the mutation and on the heuristic itself. It can be applied both at gene-level or at individual-level and is usually regulated by a parameter

$p_m \in [0,1]$, called *mutation probability* [33]. Just like the recombination operators, there are several mutation operators and their choice is also dependent on the adopted type of data structure.

For binary-coded GAs, mutation consists in *flipping* the bits in the genetic sequence of an individual. The operator acts independently over each individual and the event of flipping a specific bit on a string is stochastically independent. The process is divided into two parts: selecting genes to mutate, with probability p_{sel}, and flipping the respective bits, with probability r_m (mutation ratio), such that the probability of mutating a specific gene is equal to the probability of selecting and flipping that same gene, $p_m = p_{sel} r_m$. Hence, the probability of transitioning from a string b, of length l, to another b', by flipping k specific bits is given by [35]:

$$\Pr(b \to b') = {p_m}^k (1 - p_m)^{l-k} > 0 \tag{2.20}$$

There are essentially two heuristics of mutation. *Gene-by-gene mutation* [33] visits each gene separately and allows each bit to flip with probability r_m. In practice, for a binary string of length l, it means each gene is selected with probability $p_{sel} = 1$, in a sequence of l independent selection events, and mutated with probability r_m. The mutation probability of a specific gene is $p_m = r_m$. With this operator, the actual number of mutated genes is not fixed, but depends on a drawn sequence of l uniformly distributed random numbers and the mutation ratio, as shown in Figure 2.6.

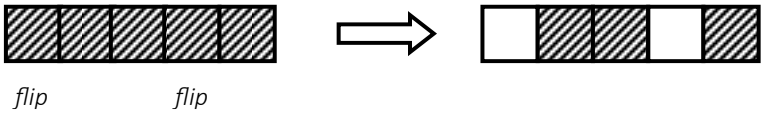

Figure 2.6: Gene-by-gene mutation. The values $\boldsymbol{k = 1}$, $\boldsymbol{l = 5}$, $\boldsymbol{p_m = k/l = 0.2}$ and the array of random numbers $\{\boldsymbol{0.15, 0.6, 0.3, 0.1, 0.8}\}$ are used to decide mutation. The first and fourth genes are mutated.

On the other hand, *k-point mutation* allows greater control over the number of mutated genes. For a binary string of length l, k genes are selected for mutation, without replacement, in a series of k sampling events. The probability of one gene being selected, anywhere among the k trials, is described by a hypergeometric distribution:

$$p_{sel} = \frac{\binom{l-1}{k-1}}{\binom{l}{k}} = \frac{k}{l} \tag{2.21}$$

and the respective mutation probability is $p_m = \frac{k}{l} r_m$.

It is then concluded that, for equal values of r_m, the mutation probability is always inferior for k-point mutation than for gene-by-gene mutation, for a given k and l. It means that k-point mutation offers smaller probabilities of transitioning from one string to another with significant changes in the genetic code, as follows from (2.20). For that reason, in gene-by-gene mutation, r_m is usually set to smaller values, to avoid the GA of becoming a sophisticated random walk heuristic [20]. Corne and Lones [13] indicate typical values of 4-8%. From the perspective of the evolution, gene-by-gene mutation is more explorative, while k-point mutation tends to be more exploitative. Moreover, the latter allows to determine an exact number of mutated genes, by simply making $r_m = 1$, as in Figure 2.7.

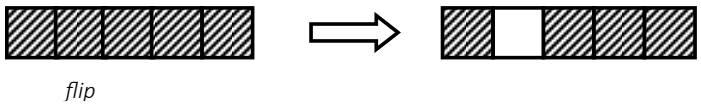

Figure 2.7: $\boldsymbol{k}$-point mutation. The values $\boldsymbol{r_m = 1}$, $\boldsymbol{k = 1}$, $\boldsymbol{l = 5}$, $\boldsymbol{p_m = k/l = 0.2}$ and the random number $\boldsymbol{0.35 \in [0.2, 0.4]}$ are used to decide mutation. Only the second gene is mutated.

For permutation-coded GAs, the mutation operator must ensure that the individual after mutation is a permutation of the individual prior to mutation, preserving order and meaning [36]. The action of assigning a random integer value to a gene can no longer be independent from other genes. Rather, the application of mutation to permutations is a matter of moving genes around in an individual's array [33]. Therefore, gene-by-gene heuristics are not allowed and the mutation probability is interpreted as the probability that some heuristic will permutate an individual's genetic sequence. The most common heuristics applied to permutations are *swap mutation*, *insert mutation*, *scramble mutation*, *inversion mutation* and *assorted mutation* [33,36]. Real-coded GAs require mutation heuristics that are able to locally perturb the solutions, in a continuous domain, and hence the forms of mutation described above are not applicable. The common procedure is to generate a vector of random deviations, with a symmetric distribution around zero, which will be added to a real-coded solution. Gene-by-gene and k-point mutation are applicable, either generating a full, or an incomplete, vector of random deviations. *Uniform mutation* and *Gaussian perturbation* [33] are found in the literature. *Nonuniform mutation*, controlling the magnitude of mutation with the number of generations, is also common [37].

Regarding other EAs, Evolutionary Programming and Evolution Strategy algorithms usually rely on real encoding. Gaussian perturbation is commonly applied together with self-adaptation strategies to each search variable, according to a set of parameters that determine the magnitude of mutation [13,33]. Differential Evolution uses a distinctive stochastic geometric approach, based on the Nelder-Mead simplex search method, known as *differential mutation* [38].

Selection mechanisms

Selection mechanisms are the computational counterpart of *natural selection*. They are responsible for setting evolution in motion and improving the overall fitness of the population, in time. As with natural life, the act of selection happens at several stages in the evolution cycle of an EA. Recovering the conceptual EA depicted in Figure 2.2, one finds two explicit moments of selection, *parent selection* and *survivor selection*, defined by independent operators, which may or may not complement each other. Selection is also found at a smaller scale, as a part of larger heuristics, like recombination and mutation.

The act of selection may be either stochastic or deterministic. EAs mostly rely on stochastic processes as a means to simulate the inherent randomness of natural evolution. However, it is acknowledged that a sensitive balance between stochastic and deterministic mechanisms is fundamental for performance efficiency, because while stochasticity is often associated with diversity maintenance (exploration), determinism is associated with selective pressure (exploitation).

Implementing stochastic selection mechanisms in EAs is straightforward and involves random sampling. The simplest way of randomly select k samples from a population of size n is by the *roulette wheel algorithm* [33]. For its simplicity, the roulette wheel algorithm is traditionally implemented to perform random sampling processes on EAs.

It assumes a certain order within the population, so that to each element corresponds a selection probability $p_{sel}(i)$ and a cumulative probability $P_i = \sum_1^i p_{sel}(i)$. The procedure consists on dividing the interval [0,1] into n subintervals $[P_{i-1}, P_i\ ]$, drawing a uniformly distributed random number r and identifying on which interval it falls. It is repeated k times. The structure of the algorithm is given in Figure 2.8.

```
! Given n subintervals [P_{i-1}, P_i ]
! select k samples from the population
set j = 1
WHILE (j ≤ k) DO
    r = U(0,1)
    set i = 1
    WHILE (P_i < r) DO
        set i = i + 1
    END
    set sample(j) = element(i)
    set j = j + 1
END
Output sample
```

Figure 2.8: Pseudocode for the roulette wheel algorithm [33].

Some authors prefer *remainder sampling methods* [33,39], designed to have lower values of bias and bounded spread, when drawing samples with replacement, as in the parent and survivor selection mechanisms. The goal is to obtain samples that better represent the selection probability distribution, in an attempt to improve the performance of EAs. Baker [40] proposes the s*tochastic universal sampling method*, with zero bias and minimum spread.

Several selection mechanisms can be found in the literature. Some are completely stochastic, while others are deterministic and some a combination of the two. Parent and survivor selection can be totally independent and applied at different stages, or they can work as a sequence, simply complementing each other. Just like any other evolutionary operator, selection mechanisms are plastic and can be designed in different ways to better suite a specific problem. The principles behind them, however, never change.

Parent selection mechanism

The role of the *parent selection mechanism* is to classify the individuals of the population, based on their fitness level, and to assign greater probability of reproduction to the best individuals, allowing their genetic survival. Weaker individuals are not without a chance. In nature, such individuals may have genetic information that may prove useful to future generations. An individual becomes a parent if selected to undergo recombination and to generate offspring. This mechanism is typically applied as a two-fold stochastic process, in the sense that sequences of random events will determine which parents are selected and whether or not they reproduce. First, a set of individuals, known as the *mating pool*, is drawn from the population. Then, from those, a percentage undergoes recombination.

Fitness proportional selection [33] is the most straightforward selection mechanism in GAs. It determines that the probability of an individual being selected to be in the mating pool depends on its fitness ratio, i.e., the quotient between an individual's fitness and the total fitness of the population, as follows:

$$p_{sel}(i) = \frac{f(i)}{\sum_{j=1}^{n} f(j)} > 0 \tag{2.22}$$

The mating pool is then generated applying any sampling method, as discussed before. This mechanism presents some issues, that must be considered when designing a GA. It applies a strong selective pressure in outstanding individuals, but almost no selective pressure in balanced populations. GAs may suffer from excessive exploitation early on in the evolution, when fitness variance is high, following a quick stagnation, when individuals become too similar too fast.

Ranking proportional selection is an alternative mechanism to preserve limited selective pressure. It assigns selection probabilities to individuals based on a performance measure (ranking), rather than the magnitude of their fitness, and requires the population to be sorted by fitness. The mapping from ranking to selection probability is not unique [33]. Baker [41] proposes the following linear map:

$$p_{sel}(i) = \frac{2(s-1)}{n(n-1)} i + \frac{(2-s)}{n} \geq 0 \tag{2.23}$$

where n is the population size, $i \in [0, n-1]$ is the individual ranking, representing the number of worse individuals in the sorted population, and $s \in [1,2]$ is a real parameter to control the boundaries of the probability mass function, assigning minimum and maximum values to the selection probabilities. Hence, it follows that $p_{sel}(0) \in [0, (1/n)]$ and $p_{sel}(n-1) \in [(1/n), (2/n)]$ and the selection probability of each individual increases linearly, at constant rate, on the interval $[0, n-1]$.

Tournament selection [33] is a popular method among GA developers, for its simplicity and because it does not require statistical nor quality knowledge of the entire population. It is particularly helpful, whenever the population is very large or obtaining such knowledge is highly time consuming [33]. Instead, it only relies on some qualitative relationship that can compare any two individuals and on a parameter t_{size}, the tournament size, that allows to control selective pressure. The higher t_{size} is the more above-average individuals are expected to enter the tournament [20]. Figure 2.9 shows the algorithm.

```
! given an ordering relationship between individuals
! select k samples from the population
set t_size
set i = 1
WHILE (i ≤ k) DO
    draw t_size individuals universally at random, with or without replacement
    rank the drawn individuals by quality
    denote the best individual as individual(1)
    set sample(i) = individual(1)
    set i = i + 1
END
Output sample
```

Figure 2.9: Pseudocode for the tournament selection algorithm [33].

Traditionally, in GAs, after generating the mating pool, it is introduced the concept of *crossover probability* $p_c \in [0,1]$, which determines how many individuals will actually undergo recombination. Parents are selected from the mating pool to create offspring via recombination with probability p_c, otherwise remain in the population with probability $(1 - p_c)$ [29]. Such step is taken as a way of preserving some genetic material from the current generation and control selective pressure. In that

sense, the size of the mating pool itself is debatable and may range from k $(< n)$ individuals to n or more individuals, just like the number of offspring individuals.

In relation to other EAs, in Evolutionary Programming parent selection is deterministic [42], each parent generates one child by mutation. In Evolution Strategies parent selection tends to be stochastic [33], it may be deterministic as well [13]. Differential Evolution does not construct any probability distribution. Individuals taking part in the variation operation are uniformly sampled from the population [13].

Survivor selection mechanism

The *survivor selection mechanism* represents the second moment of selection, in EAs. It is applied after variation operators and ultimately aims to select those individuals that will make up the population on the next generation, according to some selection criteria. As opposed to the previous operators, survivor selection mechanisms depend on neither data structure nor algorithmic structure. Hence, they are often shared between EAs. In contrast to the parent selection mechanism, which is typically characterized as a stochastic process, survivor selection is often applied for the most part as a deterministic process, thus being more connoted with the exploitation of the search space. In the literature, we find two traditional evolution models: the *generational model* and the *steady-state model* [20]. In the former, an entirely new population of n individuals is generated from the offspring set $(k \geq n)$, at each generation. The idea is that the genetic material of the parents may largely survive and remain in the population, although the parents themselves do not. In the latter, only a part of the population is replaced $(k < n)$, at each generation, in such a way that parents and offspring can and typically do co-exist. Whatever the evolution model, the survival of individuals may be determined on the basis of age (time) or fitness (quality).

The essence of *age-based replacement* [33] relies on the lifetime of individuals, rather than their fitness, sustained by the hypothesis that the average fitness of the population increases gradually over time. It is a *deterministic mechanism* that allows individuals to stay on the population only for a limited number of generations, after which they are replaced. This way, it is possible to exert some control on the selective pressure, by reducing the number of *copies* of highly fit individuals, at each generation. The canonic GA [14] applies age-based replacement to the generational model, trading an entire population of size n by the same number of offspring $(k = n)$, each individual having the lifetime of exactly one generation. Age-based replacement can also be applied to steady-state models $(k < n)$, where only the k oldest individuals of the current population are replaced by the k offspring solutions.

Alternatively, f*itness-based replacement* decides survival based on the fitness of individuals. Stochastic mechanisms like fitness proportional selection, ranking proportional selection and tournament selection can be used as stochastic survivor selection mechanisms. *Round-robin tournament selection* is a variant of tournament selection, introduced as survival mechanism within Evolutionary Programming [33]. It was applied to GAs by Harik *et al.* [43]. Stochastic replacement schemes can be applied both to generational and steady-state models. In the latter, the remaining population is uniformly randomly selected from the parent solutions [44]. As mentioned, it is often preferable to balance the stochasticity applied early in the evolution cycle with a deterministic mechanism of survival selection. (n, k) *selection* [33] is a deterministic fitness-based generational model combining both fitness and age-based replacement, commonly used in Evolutionary Strategies, where $k > n$. It consists of discarding all the parents, at the end of each generation, and rank the k offspring solutions according to fitness. Of those, the best n survive.

Elitist strategy

An undesirable effect of the traditional evolution models is the possibility of a new population being worse, in terms of fitness than the previous one, as a side effect of stochastic selection and variation operators. A straightforward solution to this issue is the application of what is known as an *elitist strategy*, or simply *elitism* [44]. Broadly speaking, elitism is the act of saving the best individual(s) of the current population, for the next generation, without any variation. That way, it allows to the population to preserve the genetic features that give it higher levels of fitness and force the evolution towards even better solutions. There are innumerous strategies under the umbrella of elitism, proposed in the literature, both for generational and steady-state models, some of which we present next. Rudolph [35] studied the effect, in terms of the limit-behavior of the canonic GA, of keeping record of the best solution found over time, as extra member of the population, which didn't take part in the evolution cycle. He deemed this form of elitism as "an algorithmic trick", alluding to the fact it is not part of the heuristic of any operator, but rather an extra algorithmic step. In e*litist selection* [45], the fittest individual survives with probability 1. Not only is it maintained in the population, it is also allowed to generate new individuals. If no solution has equal or better fitness, then the fittest individual, it is artificially inserted into the new population and one of the offspring is discarded. This scheme is usually applied in conjunction with any of the previous replacement schemes. Elitist selection is known to create a large selective pressure, which may lead to premature convergence [44].

Other replacement schemes do not employ an explicit elitist strategy. Instead, they automatically guarantee that the best individual remains in the population. Whitley's GENITOR algorithm [46] is a fitness-based steady-steady GA, where the worst k individuals of the population are replaced by the new k offspring solutions. This scheme allows to invariably keep the fittest solutions in the population but has the tendency to apply excessive selective pressure towards those solutions, since they are kept in the population for longer periods of time. It is usually applied with large populations and a "no duplicates" policy. The $(n + k)$ *selection* [33] is a fitness-based replacement mechanism, common in Evolution Strategies. This scheme considers the union of both the parent and the offspring populations $(n + k)$ and selects the top-n individuals to survive, for the next generation.

2.1.3.2 Controlling genetic diversity

In EAs, a good performance results from a proper balance between selective pressure (exploration) and genetic diversity (exploration). Exerting selective pressure is a natural feature of EAs. It prevents the search of becoming totally random, by favoring the most promising combinations of genes, to survive and reproduce. On its side, genetic diversity concerns about the degree of similarity between the solutions in a population. In population-based EAs, selective pressure is known to cause the loss of genetic information available in the population. A number of measures have been proposed for quantifying selective pressure, often in terms of the loss genetic diversity [33]. Goldberg [14] concludes that the sudden decrease in diversity may lead to what is known as *premature convergence*. The genetic operators fail to produce enough diversity to tackle new search regions and the algorithm reiterates over the known regions. Finding the global optimum becomes a rare random event and any possible improvement in the solution quality justifies the resources used.

Controlling genetic diversity is therefore fundamental for the performance of EAs. Over the years, innumerous approaches have been proposed to achieve greater genetic diversity. There are essentially two kinds of methods: methods to generate raw diversity and methods to maintain diversity. The first explicitly introduce raw diversity, whenever the search stagnates or some minimum level of diversity is exceeded, by restarting the population, totally or partially. These also include several initialization techniques to generate diverse initial populations. The latter, on the other hand, seek to introduce and

maintain diversity at every generation, such that the populations are capable of investigating many optima in parallel. These methods can be applied either at selection level or at spatial distribution level. Diversity control methods may also be classified as *fitness-based*, if they manipulate explicitly the fitness of solutions to account for some diversity measure; or as *non-fitness-based*, if the control of genetic diversity is imposed by an independent heuristic. Table 2.1 shows a collection of current and relevant diversity control methods, organized in accordance with their taxonomic classification.

Table 2.1: Taxonomic classification of diversity management techniques.

		Fitness-based	Non-fitness-based
Raw diversity generation		-	restart operators [27,43], initialization techniques [50].
Diversity maintenance	Parent selection and recombination	fitness sharing [33], ranking proportional selection [41], FUSS [47], disruptive selection [48].	negative assortative mating: [51] outbreeding mating: [51] speciation [52]
	Survivor selection and replacement	fitness sharing [33], repelling algorithm [49], FUDS [47].	standard crowding [44], deterministic crowding [53], restrictive tournament selection [54], similarity control [55], implicit mutation [56].
	Spatial separation [57]	-	island EAs, cellular EAs, clustering EAs.

Of particular interest are the so called *niching methods*, based on the mechanics of natural ecosystems, where nature organizes stable subpopulations (niches) of related individuals by forcing them to share finite resources. Within the new metaphor, a niche is local region of the search space containing solutions of interest that will compete within that niche but not between niches, preserving global diversity without lowering local selective pressure. Niching has been applied to EAs for a long time and techniques include *crowding methods* [44,53,54], *fitness sharing* [33] and *spatial separation methods* [57]. For comparative studies, see [58]. Originally developed for GAs, niching methods have been applied as well in Differential Evolution and, less commonly, in Evolution Strategies [13].

Alternatively, *non-random mating schemes* are biologically inspired selection mechanisms, based on the degrees of parenthood or similarity of the agents involved in the reproduction game. The *negative assortative mating* (NAM) is a parent selection mechanism, that induces dissimilar individuals to mate more often [51]. *Outbreeding mating* [51] determines that individuals mate preferentially outside their family tree. Contrary to NAM schemes, outbreeding does not exclude recombination between similar individuals explicitly, but surely decreases its frequency, since relatives tend to share large amounts of genetic information. In opposition to the two previous schemes are *speciation schemes* [51]. Often, the diversity introduced by niching methods starts to slow down the convergence rate, as soon as good solutions associated with different niches start to produce inferior offspring. In such cases, recombination between niches should be reduced.

With the exception of fitness sharing, all the methods discussed so far do not rely on fitness manipulation. Fitness-based methods are usually in the form of disruptive selection mechanisms that favor the spread of solutions along the search space, by manipulating their fitness. Other techniques include ranking proportional selection [41], disruptive selection, fitness uniform selection scheme (FUSS) and fitness uniform deletion scheme (FUDS) [47] and the repelling algorithm [49].

2.1.3.3 Hybrid Evolutionary Algorithms

A hybrid metaheuristic is an algorithm combining several components of different metaheuristics [20]. The hybridization of metaheuristics is attractive, because it can take advantage of the complementary features of different methods to produce an algorithm that is both more robust and efficient. In the field of evolutionary computation, hybrid EAs has been successfully applied to solve real-world problems [59,60]. The plasticity of EAs allows the development of hybrid algorithms that either offer a superior balance between exploration and exploitation or intensify one of them. A concise analysis and a taxonomic classification of hybrid EAs can be found in [61]. There are many different instantiations of hybrid EAs, each one with a distinct framework. Yet, these can be classified under two categories: collaborative and integrative hybrid EAs.

Collaborative hybrid EAs are a set of self-contained metaheuristics, where at least one EA is included, that interchange information between them, while running in parallel sequentially. In Island EAs (see Table 2.1), a population is partitioned into several subpopulations (species), having different exploration and exploitation rules. The migration of solutions between subpopulations unifies the process. In Coevolutionary EAs [13], populations evolve independently, but their fitness is interdependent. Other collaborative schemes rely on a single population, such as in [62], where a GA, tabu search and a local optimizer run in parallel. In [64], a GA developed for exploration is run and, as a final search stage, the best solutions found with the GA are improved with a local search method. On the other hand, in [64], the local refinement of solutions is executed by a second highly exploitative GA.

Integrative hybrid EAs are a category of metaheuristics where one master algorithm is in charge of the search process, while a subordinate algorithm is embedded in the master's procedure. EAs may work either as the master or the subordinate algorithm. Perhaps, the best-known instantiation of integrative hybrid EAs are the *Memetic Algorithms* (MAs), named after the concept of *meme* introduced by Richard Dawkins [19] as a unit of cultural evolution that self-propagates within populations and through generations. Initially, MAs relied on the integration of local optimizers (subordinate) to improve solutions obtained by an EA (master), achieving a trade-off between local exploitation and global exploration, respectively [65]. Currently, MAs have evolved into more sophisticated metaheuristics and can categorized under the concepts of *Lamarckism* or *Baldwinism* [66]. In evolutionary computation, Lamarckism refers to the explicit manipulation of the solutions' strings, seen as acquired knowledge that will be inherited by future offspring, while Baldwinism refers to the idea that an individual's potential to improve will affect its reproductive success and, consequently, the genetic composition of future populations through natural selection, without the explicit genetic manipulation. A recent review on MAs can be found in [67]. Lastly, EAs can be integrated in the procedures of other master algorithms, such as in [68], where a GA plays the role of the neighborhood operator in Simulated Annealing, and in [69], where a version of the CHC algorithm plays the role of a perturbation operator, in iterated local search.

3.1.3.4 Constraints handling

The vast majority the NLP problems found in engineering are constrained. Performance-wise, optimization algorithms are largely affected by the existence of constraints, both in terms of speed and quality of the solutions. In general, the difficulties in solving constrained optimization problems are associated with the fact that the feasible region is often much smaller than the whole search space and that the unconstrained optimum is often located outside the feasible region and the constrained optimum near the constraints' boundaries.

EAs were originally conceived as global optimizers for unconstrained problems. The incorporation of constraints in the evolutionary process is somewhat artificial and takes its toll in the performance of these algorithms, because a one-to-one correspondence (bijection) between the set of feasible parents and the set of possible offspring is not always available, but rather a many-to-many correspondence. It means that, from a set of feasible phenotypes, the variation operators can produce new genotypes with infeasible phenotypes and that such genotypes can be produced by different sets of parents. Thus, generating feasible solutions becomes a hard task for EAs [70]. Over the years, several constraint handling methods have been proposed in the literature. Eiben and Smith [33] identify two approaches: indirect and direct constraint handling, which are not mutually exclusive.

Indirect constraint handling methods represent the classical way of incorporating constraints into EAs. In this case, constraints are transformed into optimization objectives and the fitness function is modified by means of a penalty function, transforming the primal problem into a penalty problem. Consistent with section 2.1.2, the penalized objective function to be minimized is defined as follows:

$$\phi_P(\mathbf{x}) = f(\mathbf{x}) + \mu P(\mathbf{x}) \tag{2.24}$$

where $\mu \geq 0$ and $P(\mathbf{x})$ a penalty function, satisfying $P(\mathbf{x}) > 0$, $\forall \mathbf{x} \notin S$, and $P(\mathbf{x}) = 0$, $\forall \mathbf{x} \in S$. Such as with iterative optimization algorithms (IOAs), penalty functions in evolutionary computation can be classified either as interior or exterior [70]. In EAs, exterior penalty functions are largely preferred, simply because the problem of finding feasible solutions in the search space is itself NP-hard and exterior penalty methods do not require feasible initial solutions, as opposed to interior penalty methods [71]. However, IOAs and EAs have different philosophies. Applying Theorem 2.6, IOAs approximate successive infeasible optima of a sequence of penalty problems to the feasible optimum of the primal problem, starting from an infeasible solution and by increasing the value of μ successively at each step [1]. On the other hand, EAs penalize the fitness function of infeasible solutions, in order to increase the (stochastic) selective pressure towards the feasible solutions. Hence, Theorem 2.6 is not applicable to EAs, in the sense that no convergent sequence of infeasible optima is constructed by the consecutive populations. With time, one expects the best solutions to guide the search from the interior of the feasible region S, where the penalty function is null.

The application of exterior penalty functions to EAs has its own specificities. Penalty functions are indeed not only a function of the search variables, but also of a set of parameters known as *penalty parameters* that must be accurately determined for a proper balance between exploration and exploitation. In general, penalty functions are defined in terms of metrics $d_i(\mathbf{x})$ of constraint violation, as follows:

$$P(\mathbf{x}) = \sum_i K_i \big(d_i(\mathbf{x})\big)^{q_i} \tag{2.25}$$

where K_i are penalty parameters and $q_i \in \{1,2\}$ are penalty exponents. As for the metrics $d_i(\mathbf{x})$, there are essentially two definitions: either they represent the absolute degree of constraint violation, or they are a function of the number of violated constraints [70]. The effectiveness of exterior penalty methods is however largely determined by the penalty parameters, μ and K_i [33]. If excessive, the fitness values of solutions become too similar and EAs tend to converge to suboptimal feasible solutions. If too low, EAs tend to converge to infeasible solutions, due to an excessive exploration of the infeasible region. Researchers have proposed different ways to automate the definition of good penalty parameters.

Penalty functions are said *static* if the penalty parameters are kept constant, during the entire evolutionary process. The simplest approach is death penalty. Infeasible solutions are assigned very high penalties, or simply excluded from the population. Although very simple, this approach does not

distinguish between infeasible solutions. Kuri and Quezada [72] proposes a piecewise fitness function, where the fitness of infeasible solutions is defined solely as a function of the number of constraints violated. Homaifar *et al.* [73] defined several levels of constraint violation, each with increasing penalty parameters, for each constraint. Conceição António [56] defined two degrees of constraint violation (weak and strong), and the associated values of $P(\mathbf{x})$, to solve systems of two exponential equations in order to K_i and q_i. The main problem of using static penalty functions remains setting of the penalty parameters, that requires some knowledge of the problem.

Penalty functions are said *dynamic* if the penalty parameters are a function of time, i.e., generations. A typical approach is to replace the static penalty parameters by a function $K_i(n) = (K_i n)^{\alpha}$, where $\alpha \in \{1,2\}$ and n is the current generation [74] Other dynamin penalty functions include those of Kazarlis and Petridis [75]. Alternative dynamic methods use the analogy of simulated annealing [70], where the penalty parameter increases over time, while the cooling temperature decreases following a cooling law. In general, the issues associated with static penalties are also present with dynamic penalties [70], although now it is possible to vary the intensity of selective pressure over time.

Adaptive penalty functions are dynamic methods that take feedback from the search process, to better adjust the penalty parameters and reduce the danger of a poor choice of the penalty parameters [33]. Hadj-Alouane and Bean [76] proposed a method that increases the penalty parameter if the elite group of the last n_k generations was feasible and vice-versa. If the elite group is partially feasible only, then the penalty parameter remains constant. In the Stepwise Adaptation of Weights algorithm [33] a fixed penalty increment ΔK is added to the parameters K_i, for each constraint violated by the best individuals in the population. Other methods include those of Eiben *et al.* [77], Rasheed [78] and Crossley and Williams [79]. Although, these methods represent an improvement in the automatic assignment of penalty parameters, their increased complexity also requires extra parameters to be determined, which once again requires knowledge of the problem.

As an alternative, *direct constraint handling methods* englobe heuristics that explicitly enforce the constraints of the NLP problem, such that every solution is feasible without the need for any penalization. Under limited conditions of linearity of the constraint functions and convexity of the feasible region, it is possible to construct evolutionary operators that are *closed under the feasible region* [33]. That is, operators that transform feasible solutions into other feasible solutions. Alternative methods that guarantee feasibility are the *decoder functions*. Decoders are a class of $G \rightarrow P$ mappings from the genotype space directly to the feasible region. Theoretically, these methods offer great advantages, such as: no parameters, no need to evaluate infeasible solutions, easy evaluation of solution near the constraint boundaries, no need for special operators. However, decoders are requested to follow some strict rules [33,70]. The main drawback of decoders is the fact that the resulting mapping function is often many-to-one, meaning a number of potentially radically different genotypes may be mapped to the same genotype, introducing redundancy in the search process [33]. Another fact is that decoders are almost exclusively applied to binary problems, and their application to NLP problems is scarce.

Repair functions remain as the most applied direct constraint handling methods. The use of repair operators in EAs can be seen as a special case of adding local search to the algorithm. In this case, local improvement of means reducing (or removing) the constraint violation, rather than simply improve the value of the fitness function [33]. Liepins *et al.* [80] found an increased performance of EAs equipped with a repair operator, on a diverse set of constrained combinatorial optimization problems. Like MAs, repair functions can be understood as being Lamarckian or Baldwinian. Indeed, the GENOCOP III algorithm for continuous domains [81] combines both concepts, by changing the genotype only a fraction of the repaired solutions, while the remaining are only affected in terms of fitness. In terms of replacement rates, Orvosh and Davis [82] recommend a 5% rate for combinatorial optimization problems. Michalewicz [83] reported, instead, that a 15% rate suites numerical

optimization problems best. On the other hand, Liepiens *et al.* [80] never returned the repaired solution to the population, while Nakano [84] always did so. Other authors that have used repair algorithms are Le Riche and Haftka [85], Tate and Smith [86] and Mühlenbein [87]. There are no standards for the design of repair operators. As useful as it may seem, repairing infeasible solutions increases the number of function evaluations and the success of this approach relies on a carefully designed heuristic that best suites the problem. Also, for some problems, it is reported that repair operators may introduce some undesired bias in the search, harming the evolutionary process [77].

2.1.3.5 Limit-behavior of EAs: a convergence analysis

This section provides a short analysis of the limit-behavior, when $t \to \infty$, of EAs with finite search spaces and discrete time. By characterizing EAs as stochastic processes, it becomes possible to evaluate their ability to converge to the global optimal solution with probability 1, when $t \to \infty$, and hence when algorithms are no longer problem-dependent. Every algorithm converging in finite time will also do it in infinite time. The opposite, however, is not necessarily true. Therefore, the limit-behavior analysis of EAs will allow to determine under which conditions convergence is guaranteed.

Consider the following fundamental definitions. Let $\mathbb{S}$ be a finite search space and $\mathbb{A}$ a potentially empty set containing additional search information. An individual s is an element of $\mathbb{S} \times \mathbb{A}$. Now, let $\mathbb{S}^*$ be the set of all finite lists over $\mathbb{S} \times \mathbb{A}$. A population σ is a n-tuple $(s_1, \dots, s_n) \in \mathbb{S}^*$ keeping some order and allowing repetition of elements. A parent-list y_p is a list of individuals selected for genetic variation. $\mathbb{P} \subseteq \mathbb{S}^*$ denotes the set of all parent lists. A neighborhood of s, $N(s) \in \mathbb{S}^*$, is a list of individuals with a certain degree of similarity with s, including s itself. It is introduced to determine the extension of the action of variation operators. Without loss of generality, it is now assumed that $\mathbb{A} = \emptyset$.

The action of evolutionary operators can be described, in an abstract manner, by a set of stochastic functions, whose outputs represent changes in the population. Consider the following definition.

Definition 2.5 ***(stochastic function)*** **[45]:** Let (Ω, A, P) be a probability space and let $\mathbb{f}$ be the set of all functions $f: \mathcal{X} \to \mathcal{Y}$. Then, $g: \Omega \to \mathbb{f}$ is a random variable, whose realization $g(\omega): \mathcal{X} \to \mathcal{Y}$ is a stochastic function $f(\omega, x)$, $\forall \omega \in \Omega$, and ω is a random parameter obtained by a random sampling process.

Given the structure of the conceptual EA, represented in Figure 2.2, consider the following sequence of stochastic functions.

Definition 2.6 ***(parent-selection function)*** **[45]:** Let $\alpha \in \Omega_{\mathrm{A}}$ be a random parameter. A parent-selection function $f_p: \Omega_{\mathrm{A}} \times \mathbb{S}^* \to \mathcal{P}(\mathbb{P})$ is a stochastic function that returns a mating pool, comprised of several parent-lists, $y_p \in \mathbb{P}$, from a population $\sigma \in \mathbb{S}^*$, such that:

$$y_p \in f_p(\alpha, \sigma) \Rightarrow y_p \subseteq \sigma \Rightarrow y_p \neq \emptyset \tag{2.26}$$

where $\mathcal{P}(\mathbb{P})$ is the power set of $\mathbb{P}$ containing all possible mating pools $f_p(\alpha, \sigma)$.

Definition 2.7 ***(variation function)*** **[45]:** Let $\beta \in \Omega_{\mathrm{B}}$ be a random parameter. A variation function $f_v: \Omega_{\mathrm{B}} \times \mathbb{P} \to \mathcal{P}(\mathbb{S})$ is a stochastic function that returns an offspring set, belonging to the joint neighborhood of the elements in a parent-list $y_p \in f_p(\alpha, \sigma)$, such that:

$$f_v(\beta, y_p) \subseteq \bigcup_{s \in y_p} N(s) \tag{2.27}$$

where $\mathcal{P}(\mathbb{S})$ is the power set of $\mathbb{S}$.

The variation function represents an abstraction of the recombination and mutation operators, presented in the previous section. To each operator there corresponds a specific variation function, as follows:

$$f_v(\beta, y_p) = \begin{cases} cross(\beta, y_p) & , \text{if } \#y_p > 1 \\ mut(\beta, y_p) & , \text{if } \#y_p = 1 \end{cases} \tag{2.28}$$

where $\#y_p$ is the cardinality of the parent-list. From the previous definition it is easily concluded that the structure of the neighborhoods defined in the recombination and mutation operators is different. A neighborhood represents the list of the possible offspring to be generated by the individuals in a parent-list, by a particular variation function. The degree of similarity between the possible offspring and the respective parents is determined by the heuristic subjacent to each operator. Hence, different notions of locality are attained, with each operator. Regarding recombination, the neighborhood contains the individuals possibly reached by the stochastic crossing of the genetic information of a set of parents. On the other hand, with mutation, the neighborhood contains the solutions possibly reached by the stochastic alteration of genes in the genotype of an individual. Notice that $N(s) = \mathbb{S}$ is a correct definition, which frees us from being restricted to local search [45].

Definition 2.8 ***(survivor selection function)*** **[45]:** Let $\gamma \in \Omega_\Gamma$ be a random parameter. A survivor selection function $f_s: \Omega_\Gamma \times \mathbb{S}^* \to \mathbb{S}^*$ is a stochastic function that returns a set of n individuals from the extended population σ', of parents and offspring solutions, to survive for the next generation such that:

$$f_s(\gamma, \sigma') \subseteq \sigma' \quad , \forall \sigma' = \sigma \cup \left(\bigcup_{y_p \in f_p(\alpha, \sigma)} f_v(\beta, y_p) \right) \tag{2.29}$$

Given a formal definition of the evolutionary operators, it is now possible to present a proper definition of evolution, as follows.

Definition 2.9 ***(evolution)*** **[45]:** Let $\alpha_t \in \Omega_\mathrm{A}$, $\beta_t \in \Omega_\mathrm{B}$ and $\gamma_t \in \Omega_\Gamma$ be random parameters generated at an instant $t \in \mathbb{N}$, in generation units. A transition function $f_{tr}: (\Omega_\mathrm{A}, \Omega_\mathrm{B}, \Omega_\Gamma) \times \mathbb{S}^* \to \mathbb{S}^*$ is a stochastic function that returns a new population in a single operation as follows:

$$f_{tr}((\alpha_t, \beta_t, \gamma_t), \sigma_t) = f_s\left(\gamma_t, \sigma_t \cup \left(\bigcup_{y_{p_t} \in f_p(\alpha_t, \sigma_t)} f_v\left(\beta_t, y_{p_t}\right) \right) \right) \tag{2.30}$$

then, the evolution of an EA is defined by the following sequence of states, in time:

$$\{\sigma_t \mid t \geq 0\} : \begin{cases} \sigma_0 \in \mathbb{S}^* & , t = 0 \\ \sigma_{t+1} = f_{tr}((\alpha_t, \beta_t, \gamma_t), \sigma_t) & , t > 0 \end{cases} \tag{2.31}$$

where σ_0 is an arbitrary initial population.

By the previous definition, it is therefore concluded that the evolution between two consecutive populations (or states) is a stochastic process, in discrete time, where each state only depends on the immediately preceding state, in a probabilistic manner [88]. That is, a stochastic process that possesses the *Markov property*. Hence, to each realization of the transition function corresponds a

transition probability, reflecting the probability of transitioning between two states, at a given $t \in \mathbb{N}$. Assuming that the transition function is independent from t, then the evolution is classified as a *time-homogeneous Markov chain with a finite search space* [88], with constant transition probabilities between states, over time [35].

The theory of Markov chains allows to evaluate the limit-behavior of EAs and under what conditions global convergence happens. Specific Markov models were developed for some variants of EAs [88]. However, in order to investigate the limit-behavior of EAs, as an entire class of algorithms, it is necessary to abstract the evolutionary operators and determine a set of qualitative conditions that the evolution must satisfy, in order to achieve global convergence. Notice that the deterministic concept of "converging to the optimum" is not appropriate, because the state transitions of an EA are of stochastic nature. That said, please consider the following definition of stochastic convergence.

Definition 2.10: Let $\mathbb{S}_{opt} = \{s \in \mathbb{S} \mid s \text{ is an optimum of } f\}$. An EA is said to converge to the global optimal solution with probability 1, if and only if $\Pr\left(\lim_{t\to\infty} \sigma_t \cap \mathbb{S}_{opt} \neq \emptyset \mid \sigma_0\right) = 1$.

The previous definition simply states that, for an EA to converge to the optimal solution, the probability of transitioning from an arbitrary initial population to a state containing at least one optimal solution must be equal to 1, when $t \to \infty$. Finally, consider the following definitions [45].

Definition 2.11: The evolution is monotone if $\max\{f(s)|\forall s \in \sigma_{t+1}\} \geq \max\{f(s)|\forall s \in \sigma_t\}\ \forall t \in \mathbb{N}$.

Definition 2.12: An individual $s \in \mathbb{S}$ is accessible by the evolution if $\Pr(\exists t \in \mathbb{N}: s \in \sigma_t \mid \sigma_0) > 0$. A state $y \in \mathbb{S}^*$ is accessible by the evolution if $\Pr(\exists t \in \mathbb{N}: \sigma_t = y \mid \sigma_0) > 0$. The set of all states that can occur in the evolution, or possible successors of σ_0, is $succ(\sigma_0) = \{y \in \mathbb{S}^* \mid y \text{ is accessible}\}$.

Given these definitions, Eiben *et al.* [45] presented the following sufficient conditions for the stochastic convergence of EAs, when $t \to \infty$. The idea behind such conditions is to determine upper bounds on the probability of transitions between states not reaching, at least, one global optimal solution.

Theorem 2.7 ***(conditions for stochastic convergence)*****:** Let $\sigma_0 \in \mathbb{S}^*$ be an arbitrary initial population and assume the following conditions are satisfied:

a) $\{\sigma_t \mid t \geq 0\}$ is monotone;

b) for $t_k \in \mathbb{N}$ and $\varepsilon_k \in [0,1]$, such that $k \to \infty \Rightarrow t_k \to \infty$ and $\prod_{k=0}^{\infty} \varepsilon_k = 0$, the probability $\Pr\left(\sigma_{t_{k+1}} \cap \mathbb{S}_{opt} = \emptyset \mid \sigma_{t_k} = y\right) \leq \varepsilon_k, \forall y \in succ(\sigma_0)$, holds for every $k \in \mathbb{N}$.

Then, the evolution reaches an optimum with probability 1.

The previous theorem simply states that if the evolution is monotone and if it is possible to construct sequences $\{t_1, t_2, \dots\}$ and $\{\varepsilon_1, \varepsilon_2, \dots\}$ such that the probability of transitioning to a suboptimal state, from any state $y \in succ(\sigma_0)$, is bounded above by an ever decreasing ε_k, then the EA achieves stochastic convergence. Such conditions are sufficient for the stochastic convergence of EAs, when $t \to \infty$, but are only necessary for convergence is finite time.

Under the assumption of homogeneity of the evolution, Eiben *et al.* [45] and Rudolph [88] determine a set of assumptions, about the properties of the neighborhood and the transition functions, that imply stochastic convergence of EAs, which we adapt to the present formulation:

(A1) Every individual of a population $\sigma \in \mathbb{S}^*$ has positive probability of being selected as a parent, both for recombination and mutation, that is, $\Pr\left(s \in f_p(\alpha, \sigma)\right) \geq \delta_p > 0, \forall s \in \sigma$.

(A2) Every individual in a parent-list $y_p \in f_p(\alpha, \sigma)$ has positive probability of not being altered by the recombination process, that is, $\Pr\left(s \in cross\left(\beta, y_p\right)\right) \geq \delta_{cross} > 0, \forall s \in y_p$.

(A3) Every individual $t \in N(s)$ has positive probability of being generated after mutation, that is, $\Pr\left(t \in mut\left(\beta, y_p\right)\right) \geq \delta_{mut} > 0, \forall t \in N(s)$ and $s \in y_p$.

(A4) The neighborhood structure is connective, i.e., $\forall s_1 \in \mathbb{S}$ and $\forall s_2 \in N(s_1)$, $N(s_1) \cap N(s_2) \neq \emptyset$ holds, such that stepping from neighbor to neighbor, it is possible to get from one individual to any other individual in $\mathbb{S}$, by mutation.

(A5) Every individual of the extended population $\sigma' \in \mathbb{S}^*$, has positive probability of surviving for the next generation, that is, $\Pr\left(s \in f_s(\gamma, \sigma')\right) \geq \delta_s > 0, \forall s \in \sigma'$.

(A6) The survivor selection function is conservative, that is, $\Pr\left(M_{\sigma'} \cap f_s(\gamma, \sigma')\right) = 1$, where $M_{\sigma'}$ represents the set containing the top solution of any extended population $\sigma' \in \mathbb{S}^*$.

Theorem 2.8: Let $N(s) \subseteq \mathbb{S}$, $\forall s \in \mathbb{S}$. An EA satisfying assumptions (A1)-(A6) converges to the global optimal solution with probability 1.

Proof. By (A6), $\{\sigma_t \mid t \geq 0\}$ is monotone and assuming that $\forall t \in \mathbb{N}$ the probability distribution of the random parameters α_t, β_t and γ_t is the same, then the Markov chain is homogeneous.

By (A4), $\forall y \in succ(\sigma_0)$, there are finite paths that allow the transition between an arbitrary $s_0 \in y$ and some $s_{opt} \in \mathbb{S}_{opt}$. Thus, $\forall y \in succ(\sigma_0)$, there is at least one accessible optimum, such that $\Pr\left(\exists t \in \mathbb{N}: s_{opt} \in \sigma_t \mid \sigma_0\right) > 0$ holds for some s_{opt}. Since $\mathbb{S}^*$ is finite, for any $y \in succ(\sigma_0)$, there is a finite minimum number of mutation steps required to find an optimum with positive probability, $m_y = \min\left\{t \in \mathbb{N} \mid \Pr\left(\sigma_t \cap \mathbb{S}_{opt} \neq \emptyset | \sigma_0 = y\right) > 0\right\} < \infty$. Because there are as many shortest paths as many $y \in succ(\sigma_0)$, there is a finite maximum shortest path $m = \max\left\{m_y | y \in succ(\sigma_0)\right\} < \infty$.

Consider now an arbitrary solution s_0, in some population $y \in succ(\sigma_0)$, and s_1 as the next step in its shortest path to s_{opt}. By assumptions (A1)-(A3) and (A5), the probability of such transition after one generation is, at least, $\delta_c \delta_m \delta_s > 0$, where $\delta_c = \delta_p \delta_{cross}$ and $\delta_m = \delta_p \delta_{mut}$. Hence, by monotony, from any $y \in succ(\sigma_0)$, it is possible to transition into $\mathbb{S}_{opt}$ in m mutation steps, with a positive probability, such that:

$$\delta_y = \Pr\left(\sigma_m \cap \mathbb{S}_{opt} \neq \emptyset | \sigma_0 = y\right) \geq (\delta_c \delta_m \delta_s)^{m-1} \delta_c \delta_m > 0 \tag{2.32}$$

$\forall y \in succ(\sigma_0)$.Then, the probability of not visiting the global optimal solution after m generations is bounded by:

$$\Pr\left(\sigma_m \cap \mathbb{S}_{opt} = \emptyset | \sigma_0 = y\right) \leq p \tag{2.33}$$

$\forall y \in succ(\sigma_0)$, where $p = \max\left\{\left(1 - \delta_y\right) \mid y \in succ(\sigma_0)\right\}$. By the homogeneity, the transition between states is always equal, meaning the transition probabilities are constant $\forall t \in \mathbb{N}$. Thus, between two consecutive generations, the probability of not visiting an optimum is, at most, $p^{(1/m)}$. Then, by making $\varepsilon_k = p^{(1/m)}$, it follows that:

$$\Pr\left(\sigma_{t+1} \cap \mathbb{S}_{opt} = \emptyset | \sigma_t = y\right) = \Pr\left(\sigma_1 \cap \mathbb{S}_{opt} = \emptyset | \sigma_0 = y\right) \leq \varepsilon_k \tag{2.34}$$

$\forall y \in succ(\sigma_0)$. Then, there are sequences $\{n_0, n_1, \dots\}$ and $\{\varepsilon_1, \varepsilon_2, \dots\}$, such that $k \to \infty \Rightarrow t_k \to \infty$ and

$$\lim_{k\to\infty} \Pr\left(\sigma_{t_k} \cap \mathbb{S}_{opt} = \emptyset | \sigma_0 = y\right) = \prod_{k=0}^{\infty} \varepsilon_k = p^{(k/m)} = 0 \tag{2.35}$$

$\forall y \in succ(\sigma_0)$. The proof is complete. ■

Corollary 2.1: Let $N(s) = \mathbb{S}$, $\forall s \in \mathbb{S}$, be the neighborhood associated with the mutation process. An EA satisfying assumptions (A3) and (A6) converges to the global optimum with probability 1.

Proof. By (A6), $\{\sigma_t \mid t \geq 0\}$ is monotone and assuming that $\forall t \in \mathbb{N}$ the probability distribution of the random parameters α_t, β_t and γ_t is the same, then the Markov chain is homogeneous. $N(s) = \mathbb{S}$, $\forall s \in \mathbb{S}$, implies that assumption (A4) is satisfied trivially. Together with assumption (A3), it means that the probability of transitioning from any $s \in \mathbb{S}$ to any other $t \in \mathbb{S}$, in a single mutation step, is positive and it follows that $m = m_y = 1$, $\forall y \in succ(\sigma_0)$. Therefore, the considerations between (2.32) and (2.35) remain valid, even if assumptions (A1), (A2) and (A5) are neglected, because if an optimum is attained after mutation, it survives according to (A6). ■

This last result allows to conclude that genetic variation of individuals by mutation alone, has the ability to reach the optimum, given an appropriate connectivity of the neighborhood sets. The potential positive effects of recombination are completely neglected in this analysis. At most, it is required that any solution should survive recombination intact. Actually, notice that an EA without recombination ($\delta_c = 1$) and with any connective mutation operator will always visit the optimum, whereas and EA without mutation but with a typical recombination operator does not have this guarantee [89].

2.2 Multi-objective Nonlinear Programming

Consider the following definition.

Definition 2.13 (*MO-NLP problem*) [90]: Let $\mathbf{f}: \mathbb{R}^N \to \mathbb{R}^k$, $Z \subset \mathbb{R}^k$ and $\mathbf{z} \in Z$ a vector of components $f_1(\mathbf{x}), \dots, f_k(\mathbf{x})$. Also, let g_i, for $i = 1, \dots, m$, and h_j, for $j = 1, \dots, l$, be scalar functions defined in $\mathbb{R}^N$, $X \subset \mathbb{R}^N$ and $\mathbf{x} \in X$ a vector of components $x_1, \dots, x_N$. Then, the general multi-objective NLP problem is stated as follows:

$$\begin{aligned} \min \quad & \mathbf{f}(\mathbf{x}) = \left(f_1(\mathbf{x}), \dots, f_k(\mathbf{x})\right)^T \\ \text{subject to} \quad & g_i(\mathbf{x}) \leq 0 \quad \text{for } i = 1, \dots, m \\ & h_j(\mathbf{x}) = 0 \quad \text{for } j = 1, \dots, l \\ & \mathbf{x} \in X \end{aligned} \tag{2.36}$$

In the terminology of multi-objective optimization, $\mathbf{f}(\mathbf{x})$ is a *vector objective function* containing k objective functions to be optimized simultaneously. The set X, known as the *search space*, defines upper and lower bounds on the search variables $\mathbf{x}$, while $Z = \{\mathbf{z} = \mathbf{f}(\mathbf{x}): \mathbf{x} \in X\}$ is the corresponding *objective space*. Additional conditions are imposed to the search variables in the shape of *inequality constraints* $g_i(\mathbf{x}) \leq 0$, $\forall i$, and *equality constraints* $h_j(\mathbf{x}) = 0$, $\forall j$. The *feasible search region* is defined by the set $S_X = \left\{\mathbf{x} \in X: g_i(\mathbf{x}) \leq 0, \forall i \wedge h_j(\mathbf{x}) = 0, \forall j\right\}$ and $S_Z = \{\mathbf{z} \in Z: \mathbf{x} \in S_X\}$ is the corresponding *feasible objective space*.

In contrast to single-objective optimization, solving a multi-objective problem is not a straightforward exercise. The multiple objective functions form a mathematical description of performance that must be shared by distinct criteria, usually in conflict with each other. Typically, there is no such thing as a global optimum, but rather a set of solutions sharing equivalent levels of "*quality*", all satisfying a predetermined concept of *optimality* [66]. It is then required a decision maker to assign preference (a biased notion of optimality) towards a certain subset of the solutions found, which may include one or more, based on knowledge of the problem or personal experience [90].

Finding optimal solutions, for any number of objectives, is a task generalized as *"finding the minimal elements of a partially ordered set"* [91]. Throughout the next section, we assign meaning to this past sentence, in the light of single and multi-objective optimization. We discuss the differences between the two cases and the need for a suitable optimality concept, under the existence of multiple objectives. Then, a final section on the application of Evolutionary Algorithms (EAs) to MO-NLP problems is presented.

2.2.1 Pareto Optimality

Start by considering the notion of *binary relation*, as an association between elements of two nonempty sets. A binary relation $\mathcal{R}$ on some set A is defined as a subset of $A \times A$. We write $\mathbf{a_1}\mathcal{R}\mathbf{a_2} \Leftrightarrow (\mathbf{a_1}, \mathbf{a_2}) \in \mathcal{R}$, by which we mean $\mathbf{a_1}$ is related to $\mathbf{a_2}$ [91].

Definition 2.14 [92]: A binary relation $\mathcal{R}$ on A is called:

- Reflexive if $\mathbf{a_1}\mathcal{R}\mathbf{a_1}$ for all $\mathbf{a_1} \in A$
- Anti-reflexive if $\mathbf{a_1}\mathcal{R}\mathbf{a_2} \Rightarrow \mathbf{a_1} \neq \mathbf{a_2}$ for all $\mathbf{a_1}, \mathbf{a_2} \in A$
- Symmetric if $\mathbf{a_1}\mathcal{R}\mathbf{a_2} \Rightarrow \mathbf{a_2}\mathcal{R}\mathbf{a_1}$ for all $\mathbf{a_1}, \mathbf{a_2} \in A$
- Antisymmetric if $\mathbf{a_1}\mathcal{R}\mathbf{a_2} \wedge \mathbf{a_2}\mathcal{R}\mathbf{a_1} \Rightarrow \mathbf{a_1} = \mathbf{a_2}$ for all $\mathbf{a_1}, \mathbf{a_2} \in A$
- Asymmetric if $\mathbf{a_1}\mathcal{R}\mathbf{a_2} \Rightarrow \neg(\mathbf{a_2}\mathcal{R}\mathbf{a_1})$ for all $\mathbf{a_1}, \mathbf{a_2} \in A$
- Transitive if $\mathbf{a_1}\mathcal{R}\mathbf{a_2} \wedge \mathbf{a_2}\mathcal{R}\mathbf{a_3} \Rightarrow \mathbf{a_1}\mathcal{R}\mathbf{a_3}$ for all $\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3} \in A$
- Connected if $\mathbf{a_1}\mathcal{R}\mathbf{a_2} \vee \mathbf{a_2}\mathcal{R}\mathbf{a_1}$ for all $\mathbf{a_1}, \mathbf{a_2} \in A$.

The last property is also known as the *totality axiom* and simply states that, if satisfied, every pair of elements $\mathbf{a_1}, \mathbf{a_2} \in A$ is always related. Such elements are *comparable*. In opposition, elements $\mathbf{a_1}, \mathbf{a_2} \in A$ are *incomparable* if $\neg(\mathbf{a_1}\mathcal{R}\mathbf{a_2}) \vee \neg(\mathbf{a_2}\mathcal{R}\mathbf{a_1})$, which is denoted by $\mathbf{a_1} \parallel \mathbf{a_2}$.

There exist several types of binary relations which are important in optimization theory, namely *order relations* and *equivalence relations*. These generalize the intuitive notions of an ordering and equality between elements on a set and are characterized as follows.

Definition 2.15 [93]: A binary relation $\mathcal{R}$ on a set A is:

- an *order relation* if it is antisymmetric and transitive
- an *equivalence relation* if it is reflexive, symmetric and transitive.

For now, let us focus on order relations.

Definition 2.16 [91]: A binary relation $\mathcal{R}$ on a set A is called a *partial order* if it is reflexive, antisymmetric and transitive. The pair $(A, \mathcal{R})$ is called a *partially ordered set* (poset).

An example of a partial order is the inclusion relation $\subseteq$ on the power set of a set [91]. Based on the definition of partial order relation, stronger statements defining stronger order relations can be written as follows.

Definition 2.17 [94]: Let $(A, \mathcal{R})$ be a poset. If, additionally, the totality axiom is satisfied, then $\mathcal{R}$ is a *total order* and $(A, \mathcal{R})$ is a *totally ordered set*, or a *chain*. On the contrary, if every pair on A is incomparable, $(A, \mathcal{R})$ is called an *anti-chain*.

A classic example of a total order is the usual order of the real numbers, $\leq$ [94]. Notice that it is only the totality axiom that distinguishes partial orders from total orders. In fact, it also explains the name *partial*. The partiality essentially refers to the fact that not all pairs in a set can be compared and, thus, as opposed to total orders, there exist incomparable pairs. A natural consequence of the previous definition is that, if $\mathcal{R}$ is a total order on a set, it implies a partial order on the same set.

Another stronger statement based on partial orders is written as follows.

Definition 2.18 [94]: A binary relation $\mathcal{R}$ on a set A is called a *strict partial order* if it is anti-reflexive and transitive (and therefore asymmetric).

An example of a strict partial order is the strict inclusion between two sets $\subset$ [93]. A strict partial order is not, in general, a partial order. It is irreflexive, while a partial order is reflexive [94]. It implies, however, a partial order relation. The link between the two relations is given next.

Proposition 2.8 [91]: Let A be some set and let $\mathcal{R}$ denote a partial order relation on A. Then, the relation $\mathcal{R}_{strict}$ defined by:

$$\mathbf{a_1}\mathcal{R}_{strict}\mathbf{a_2} \Leftrightarrow (\mathbf{a_1}\mathcal{R}\mathbf{a_2}) \wedge \neg(\mathbf{a_2}\mathcal{R}\mathbf{a_1}) \quad , \forall \mathbf{a_1}, \mathbf{a_2} \in A \tag{2.37}$$

is a strict partial order.

A weaker statement, defining a weaker binary relation is written as follows.

Definition 2.19 [92]: A binary relation $\mathcal{R}$ in a set A is called a *preorder* if it is reflexive and transitive. The pair $(A, \mathcal{R})$ is called a *preordered set.*

The term *preorder* reflects the idea that such a binary relation is nearly a (partial) order, but not quite; it is neither necessarily antisymmetric nor asymmetric. Partial orders may be constructed from preorders simply by enforcing antisymmetry. That is, any subset of a preordered set that satisfies antisymmetry is a poset [95]. Hence, a partial order always implies a preorder on a set. Similarly, a strict partial order may be constructed from a preorder, imposing asymmetry, by considering $\mathcal{R}$ as a preorder relation, in Proposition 2.8 [92]. In such case, the binary relation is anti-reflexive as well. Equivalence relations may be constructed as well from both partial orders and preorders, by imposing symmetry, as follows.

Proposition 2.9 [92]: Let A be some set and let $\mathcal{R}$ denote a partial order, or a preorder, on A. Then, the binary relation $\sim$ defined by:

$$\mathbf{a_1} \sim \mathbf{a_2} \Leftrightarrow (\mathbf{a_1}\mathcal{R}\mathbf{a_2}) \wedge (\mathbf{a_2}\mathcal{R}\mathbf{a_1}) \quad , \forall \mathbf{a_1}, \mathbf{a_2} \in A \tag{2.38}$$

is an equivalence relation. Moreover, if $\mathcal{R}$ is a partial order, then $\mathbf{a_1} \sim \mathbf{a_2} \Rightarrow \mathbf{a_1} = \mathbf{a_2}$, due to the antisymmetry property.

We are now in condition to discuss the concept of optimality. In *single-objective optimization*,

$$Z \subset \mathbb{R} \tag{2.39}$$

Then, there is a total order relation on Z, induced by the usual *lesser-than-or-equal* order of the real numbers, $\leq$. Besides being total, the usual order on $\mathbb{R}$ has some valuable properties, as well. It is compatible with the algebraic operations of addition and scalar multiplication [96].

Proposition 2.10: Let $(A, \mathcal{R})$ be a poset. Then, for any $\mathbf{a_1}, \mathbf{a_2}, \mathbf{a_3} \in A$ and $c \in \mathbb{R}$,

$$\mathbf{a_1} \leq \mathbf{a_2} \Rightarrow \mathbf{a_1} + \mathbf{a_3} \leq \mathbf{a_2} + \mathbf{a_3} \text{ and } c\mathbf{a_1} \leq c\mathbf{a_2} \text{ if } c \geq 0 \text{ and } c\mathbf{a_1} \geq c\mathbf{a_2} \text{ if } c \leq 0 \qquad (2.40)$$

In addition, it has the *Archimedean property* [96].

Proposition 2.11: Let $(A, \mathcal{R})$ be a poset compatible with addition and scalar multiplication. For any $\mathbf{a_1}, \mathbf{a_2} \in A$, with $\mathbf{a_1}$ strictly positive w.r.t. $\mathcal{R}$, there is a $k \in \mathbb{N}$ such that $k\mathbf{a_2}\mathcal{R}\mathbf{a_1}$.

Which means no element in $\mathbb{R}$ is infinitesimal with respect to any other. In terms of the elements in Z, it is possible to establish a metric that compares any two objective values, w.r.t. $\leq$. Lastly, it also has the *greatest lower bound property*[1] [96].

Proposition 2.12: Every $B \subseteq A \neq \emptyset$, that is bounded bellow, w.r.t. $\mathcal{R}$, has an infimum in the poset A.

Satisfying this property sets the conditions for the existence of $\inf(f)$, in the objective space. Therefore, it is intuitive to define an optimality concept, in the following manner:

$$\begin{aligned} \mathbf{x}^* \text{ is optimum} \quad &\Leftrightarrow f(\mathbf{x}^*) = \inf\{f(\mathbf{x}) : \mathbf{x} \in X\} \\ &\Leftrightarrow f(\mathbf{x}^*) \leq f(\mathbf{x}), \forall \mathbf{x} \in X \end{aligned} \qquad (2.41)$$

The existence of $f(\mathbf{x}^*)$, and of the corresponding $\mathbf{x}^*$, is guaranteed by the *extreme-value theorem*, for continuous scalar fields [97], because of the greatest lower bound property. Such theorem is naturally satisfied if X is a compact subset of $\mathbb{R}^N$, which is common in optimization.

In *multi-objective optimization*, however,

$$Z \subset \mathbb{R}^k \qquad (2.42)$$

with $k > 1$. One difference between $\mathbb{R}$ and $\mathbb{R}^k$ is the apparent absence of a natural order on $\mathbb{R}^k$ [96]. This fact implies that the definition of an appropriate concept of optimality, for multiple objectives, is not as intuitive as in the case of single-objective optimization.

It is possible to construct total orders on $\mathbb{R}^k$, compatible with the algebraic operations of addition and scalar multiplication, but these do not satisfy neither the Archimedean nor the greatest lower bound properties [96]. An example is the *lexicographic order* on $\mathbb{R}^k$, $k > 1$. Often used in the context of multi-objective optimization [98], this order relation admits infinitesimal objective vectors and does not guarantee the existence of an infimum, on every bounded $Z \subset \mathbb{R}^k$ [99]. The application of the lexicographic order to multi-objective optimization problems requires the decision maker to rank the different objectives in terms of their importance, such that a sequence of single-objective problems is solved, for each objective function, according to the established ranking [98]. That is, the lexicographic ordering of the objectives is used to define the sequence of single-objective problems, rather than to compare solutions in the objective space. This way, to each problem with objective function $z_i = f_i(\mathbf{x})$, the objective space is reduced to $Z_i \subset \mathbb{R}$, where the usual order $\leq$ on $\mathbb{R}$ exists. The possibility of any tradeoff between conflicting objectives is eliminated. This approach, however, only returns a single optimum vector, whose components are the optimum of each single-objective problem [99]. Frequently, it is difficult for the decision maker to rank the objectives and, instead, it is more convenient to have a range of quality solutions before deciding. In this sense, it is more adequate to construct partial order relations on Z that satisfy both the Archimedean and the greatest lower bound properties and assign equal importance to the objective functions. Indifference towards

[1] The dual statement is the *least upper bound property*, regarding the existence of a supremum.

the objectives proves to be important, because in multi-objective optimization the objectives may be in conflict or not and that notion is not always known *a priori* [98].

In multi-objective optimization, the leading concept of optimality is called *Pareto optimality*, whose definition is based on an order relation known as *(Pareto) dominance*. For now, we introduce the dominance relation generically as a strict partial order in Z (see Definition 2.18), represented by $\prec$, and with $\mathbf{z_1} \prec \mathbf{z_2}$ we mean $\mathbf{z_1}$ *dominates/is partially less than* $\mathbf{z_2}$.

From the dominance relation, we introduce the following partial order, $\preccurlyeq$ (*dominant or equal)*:

$$\mathbf{z_1} \preccurlyeq \mathbf{z_2} \Leftrightarrow \mathbf{z_1} \prec \mathbf{z_2} \vee \mathbf{z_1} = \mathbf{z_2} \quad , \forall \mathbf{z_1}, \mathbf{z_2} \in Z \tag{2.43}$$

which is equivalent to the product order on $\mathbb{R}^k$, $\leq$, that is [70]:

$$\mathbf{z_1} \preccurlyeq \mathbf{z_2} \Leftrightarrow z_{1_i} \leq z_{2_i}, \forall i \in \{1, \dots, k\} \quad , \forall \mathbf{z_1}, \mathbf{z_2} \in Z \tag{2.44}$$

such that $(Z, \preccurlyeq)$ is a *partially ordered objective space*. As a convention, every appearance of $\preccurlyeq$ tacitly assumes its definition via (2.43). Notice that this equation is the converse of (2.37) and that partial and strict partial order relations can be defined from each other [100].

Proposition 2.13[96]: The product order on $\mathbb{R}^k$, $k > 1$: (i) is compatible with addition and scalar multiplication; (ii) has the Archimedean property; and (iii) has the greatest lower bound property.

In partially ordered sets, the notion of minimum (maximum) is generalized by the concepts of least element and minimal element. Please, consider the following definitions, w.r.t. $\preccurlyeq$. An element $\mathbf{z_m}$ is the *least element* of Z, if $\mathbf{z_m}$ is the infimum of Z and $\mathbf{z_m} \in Z$. That is, $\mathbf{z_m} \in Z : \mathbf{z_m} \preccurlyeq \mathbf{z}$, for every $\mathbf{z} \in Z$. On the other hand, an element $\mathbf{z_m}$ is a *minimal element* of Z, if no element in Z is strictly smaller than $\mathbf{z_m}$. The *greatest element* and the *maximal elements* of a poset are defined dually [100]. Although related concepts, a minimal element of a poset must not be misinterpreted as the least element of the same set. The notion of least element implies comparability with every other element in Z. However, due to the partiality of the imposed order relation, not all pairs in Z are comparable. In this sense, for $\mathbf{z_m}, \mathbf{z_1} \in Z$, if $\mathbf{z_m}$ is a minimal element of Z and $\mathbf{z_m} \parallel \mathbf{z_1}$, then $\mathbf{z_1}$ is also a minimal element of Z [92], which leaves the possibility for the existence of many minimal elements. Therefore, a poset can have many minimal elements, without having a least element. However, if a poset has a least element, then it is the unique minimal element [92]. Furthermore, a poset can have one minimal element, without having a least element, if such element is not comparable with every other element in the poset. Consequently, in totally ordered sets, it is easily concluded that the concepts of least and minimal elements are coincident [92], just like in single-objective optimization.

Regarding the existence of minimal elements, there are many conditions formulated in the literature, often labelled as *existence theorems*. Given that the partial order $\preccurlyeq$ constructed in Z has the greatest lower bound property, then consider the following condition.

Proposition 2.14 [92]: Let $(Z, \preccurlyeq)$ be a poset. For any compact and nonempty $Z \subset \mathbb{R}^k$, $k \geq 1$, there exists a nonempty set of minimal elements, w.r.t. $\preccurlyeq$.

We can now relate the concepts presented so far with the purpose of multi-objective optimization: *to find the set of all minimal elements of the partially ordered objective space* $(Z, \preccurlyeq)$. In the vocabulary of multi-objective optimization, minimal elements of Z, w.r.t. the partial order $\preccurlyeq$, are called *nondominated points* and the set of all nondominated points is called the *nondominated set*, or the *Pareto front* [70].

Definition 2.20 [92]: Let $(Z, \preccurlyeq)$ be a partially ordered objective space. The set

$$Z_{PF} = \{\mathbf{z} \in Z : \nexists \mathbf{z_1} \in Z \text{ such that } \mathbf{z_1} \prec \mathbf{z}\} \tag{2.45}$$

is the *Pareto front* of Z, w.r.t. $\preccurlyeq$.

About the location of the Pareto front, Ehrgott [92] proves the following.

Proposition 2.15: $Z_{PF} \subset \partial Z \subset Z$.

That is, the Pareto front is a proper subset of the boundary of Z. In fact, for minimization problems and assuming all the objectives are nonnegative, the nondominated points are in the "bottom left part" of Z [92]. Deb [98] offers a comprehensive discussion on the location and shape of the Pareto front in the objective space, depending on the orientation of objectives (if minimization and/or maximation).

The previous results define optimality on Z. To have a complete definition of optimality, it is necessary to extend these concepts to the search space X and identify the corresponding $\mathbf{x} \in X : \mathbf{z} = \mathbf{f}(\mathbf{x}) \in Z_{PF}$. By the definitions of Z and X, both sets are linked as follows.

Proposition 2.16 [101]: Let $(Z, \preccurlyeq)$ be the partially ordered objective space, with $Z \subseteq \mathbb{R}^k$, and $X \subseteq \mathbb{R}^N$ the search space. Then, it follows that the map $\mathbf{f}: X \to Z$ induces a preorder on X, $\forall \mathbf{x_1}, \mathbf{x_2} \in X$, such that:

$$\begin{aligned} \mathbf{x_1} \prec_p \mathbf{x_2} \quad &\Leftrightarrow \mathbf{f}(\mathbf{x_1}) \prec \mathbf{f}(\mathbf{x_2}) \\ \mathbf{x_1} \sim_p \mathbf{x_2} \quad &\Leftrightarrow \mathbf{f}(\mathbf{x_1}) = \mathbf{f}(\mathbf{x_2}) \\ \mathbf{x_1} \preccurlyeq_p \mathbf{x_2} \quad &\Leftrightarrow \mathbf{x_1} \prec_p \mathbf{x_2} \vee \mathbf{x_1} \sim_p \mathbf{x_2} \end{aligned} \tag{2.46}$$

and $\left(X, \preccurlyeq_p\right)$ is the preordered search space.

The minimal elements of X, w.r.t. the preorder $\preccurlyeq_p$, are called *Pareto optimal solutions*, which are a generalization of the concept of optimum solution, in single-objective optimization, and the set of all such solutions is called the *Pareto optimal set* [70].

Definition 2.21: Let $\left(X, \preccurlyeq_p\right)$ be the preordered search space. The set

$$X_{PO} = \{\mathbf{x} \in X : \mathbf{z} = \mathbf{f}(\mathbf{x}) \in Z_{PF}\} \tag{2.47}$$

is the *Pareto optimal set* of X, w.r.t. $\preccurlyeq_p$.

An equivalent definition of (2.47) is $X_{PO} = \left\{\mathbf{x} \in X : \nexists \mathbf{x_1} \in S_X \text{ such that } \mathbf{x_1} \prec_p \mathbf{x}\right\}$ [92]. Note that the cardinality of X_{PO} is at least as big as the cardinality of Z_{PF}, but not *vice-versa*, because $\preccurlyeq_p$ is not antisymmetric, and so $\mathbf{x_1} \sim_p \mathbf{x_2}$ does not imply equality between two vectors; only indifference, w.r.t. $\preccurlyeq_p$. Thus, there can be more than one element in X_{PO} with the same image in Z_{PF} [92]. Regarding the existence of Pareto optimal solutions, consider the following statement.

Proposition 2.17 [102]: Let $X \subset \mathbb{R}^N$ be compact and $\mathbf{f}: X \to \mathbb{R}^k$ a continuous map. Then, $\mathbf{f}(X)$ is compact.

By definition $X \neq \emptyset \Rightarrow Z \neq \emptyset$, so X is compact $\Rightarrow Z$ is compact $\Rightarrow Z_{PF} \neq \emptyset \Rightarrow X_{PO} \neq \emptyset$. Although $X_{PO} \neq \emptyset$ seems trivial from (2.47), in truth it requires Z to be compact beforehand. However, such condition alone is hardly guaranteed, while the compactness of X is very common.

A final comment on the cardinality of the sets of optimal points: the existence of conflicting objectives implies the existence of incomparable points in Z, because at a certain point it is not possible to improve one objective without deteriorating the others. Hence, Z_{PF} contains all the incomparable elements in Z, i.e., the minimal elements. On the other hand, if the objectives are not in conflict, then every element in Z is comparable with any other and Z_{PF} contains the least element of Z [98].

This presentation is concluded with an exact definition of the dominance relation.

Definition 2.22 (*Pareto Dominance*) [70]: Let $\mathbf{z_1}, \mathbf{z_2} \in Z$. Then,

$$\mathbf{z_1} \prec \mathbf{z_2} \Leftrightarrow \forall i \in \{1, \dots, k\}, z_{1_i} \leq z_{2_i} \wedge \exists i \in \{1, \dots, k\} : z_{1_i} < z_{2_i} \tag{2.48}$$

In the literature, it is customary to find alternative definitions of the dominance relation. As referred in [70,90], such definitions appeared out of necessity. Some related with the difficulty in finding or converging to the Pareto front and the Pareto optimal set, in numerical optimization [70]. Others related with the handling of constraints [98]. In the next section, regarding the application of EAs to the solution of the MO-NLP problem, we will refer to the latter case.

2.2.2 Multi-objective Evolutionary Algorithms

Multi-objective optimization methods can be divided into three major categories: methods with *a priori*, *a posterior* or no articulation of preferences, referring to the fact that the decision maker may be urged to assign more importance to certain objectives, before or after solving the problem, or to treat them equally. Many of such methods convert the original problem into a single-objective problem and return only one optimal solution, at a time. Hence, requiring multiple runs to yield (a part of) the Pareto front. A concise review on the subsect can be found in [90].

When the entire Pareto front is required (or a representative part of it), more direct approaches are preferable. Multi-objective EAs (MOEAs) are, by far, the most popular search heuristics, for several reasons. MOEAs can converge to the global Pareto front, in a single run [70]; do not require the user to prioritize, scale or weight objectives [103]; do not require gradient information and are effective regardless of the nature of the objectives and constraints functions [90]. Furthermore, MOEAs can be constructed directly from EAs. In fact, both differ essentially in the application of distinct concepts of optimality [98,103].

The ultimate goal in multi-objective optimization is to identify the entire Pareto front. However, Veldhuizen [104] demonstrates that the Pareto front is composed of at most an infinite number of nondominated points, if Z is a continuous set. A practical approach is to determine a finite set of solutions that approximates the Pareto front as well as possible. In fact, the accuracy of a decision maker's choice depends on how well the Pareto front is approximated. With these concerns in mind, MOEAs are expected to accomplish the following three main goals [70]:

(i) Identify and preserve nondominated points and the corresponding Pareto optimal points
(ii) Generate and maintain diversity on the Pareto front and/or on the Pareto optimal set
(iii) Gradually progress towards the global Pareto front

In order to discuss the algorithmic progress of MOEAs, we introduce the following Pareto notation [104]. During the execution of a MOEA, a *current Pareto optimal set*, $X_{PO}(t)$, and a *current Pareto front*, $Z_{PF}(t)$, are determined at each generation t, with respect to the *current population*. Acknowledging that any population is a finite subset of X, consider the following.

Proposition 2.18 [94]: Any nonempty finite subset of a poset $(Z, \preccurlyeq)$ has at least one minimal element, w.r.t. $\preccurlyeq$.

Hence, $Z_{PF}(t) \neq \emptyset \Rightarrow X_{PO}(t) \neq \emptyset$, for all $t \in \mathbb{N}$. At the end of the evolutionary process, we define $Z_{PF_{known}} = Z_{PF}(t)$ and $X_{PO_{known}} = X_{PO}(t)$ as the final sets returned by the MOEA. Ideally, after convergence, it is expected that $Z_{PF_{known}} \subseteq Z_{PF}$ and $X_{PO_{known}} \subseteq X_{PO}$.

2.2.2.1 Fitness assignment and diversity maintenance

Under the evolutionary metaphor, Pareto optimal solutions are those genotypes, whose corresponding phenotypical features cannot be improved, w.r.t. the preorder $\preccurlyeq_p$ [70].

While genetic variation operators generate new solutions, selection operators are responsible for the progression of the evolutionary process and establish the algorithmic structure of EAs and MOEAs. A fundamental process of the selection operators is the *sorting mechanism*. The term "sorting" is clear and means *to arrange things according to their type*. Although related, it shall not be confused with "ordering", which in turn implies a pre-established order between elements. In EAs, sorting populations is trivial, since the objective space is totally ordered, and is associated with a well-defined notion of fitness. In such case, "sorting" and "ordering" are interchangeable. On the other hand, in MOEAs, fitness is somehow an abstract concept, due to a partially ordered objective space. As a result, one cannot order all the elements in a population, but can sort them into equivalence classes. Several sorting mechanisms have been proposed in the literature, usually in association with the processes of *fitness assignment* and/or *diversity maintenance*, in order to comply with the three main goals above. Also, unlike EAs, the introduction of elitism can deeply change the algorithmic structure of MOEAs. Hence, the following discussion is split into non-elitist and elitist MOEAs.

Non-elitist Multi-objective EAs

Early MOEAs are extensions of EAs. Initially, the paradigm was to maintain the structure of EAs and to replicate the natural order of the objective space, found in single-objective optimization. The *Vector Evaluated GA* (VEGA) [105] is as a modification of the canonic GA. At every generation, the population is randomly dived into *k* subpopulations, referring to each of the *k* objectives. Individuals are assigned a fitness value based on the respective objective function. Fitness proportional selection is applied independently to each subpopulation and crossover and mutation are applied as usual on the whole population, to yield the next generation. Schaffer [105] proposed two selection heuristics to maintain diversity in the Pareto front. A generalization of VEGA, the *Weight-based GA* (WBGA) [106] assigns fitness values equal to the weighted sum of the normalized *k* objective functions. Weight vectors are embedded within the solutions genotypes. Populations maintain multiple weight vectors and are evolved by a canonical GA, thereby finding multiple Pareto-optimal solutions in a single run. Fitness sharing is applied explicitly to the weight vectors of each individual, for diversity maintenance. In practice, both algorithms have convergence limitations in nonconvex problems, because the sorting mechanisms, and therefore the selection operator, rely on a set of weights rather than the dominance relation [98].

The application of the dominance partial order to sort populations is limited by the existence of incomparable elements in the objective space. To cope with this limitation, authors constructed an alternative binary relation in Z, called the *Pareto-ranking relation* [107].

Proposition 2.19: Let $r: Z \to \mathbb{N}$ be a ranking function and $\preccurlyeq_r$ represent the Pareto-ranking relation, such that, for all $\mathbf{z_1}, \mathbf{z_2} \in Z$:

$$\begin{aligned} \mathbf{z_1} \prec_r \mathbf{z_2} &\Leftrightarrow r(\mathbf{z_1}) < r(\mathbf{z_2}) \\ \mathbf{z_1} \sim_r \mathbf{z_2} &\Leftrightarrow r(\mathbf{z_1}) = r(\mathbf{z_2}) \\ \mathbf{z_1} \preccurlyeq_r \mathbf{z_2} &\Leftrightarrow \mathbf{z_1} \prec_r \mathbf{z_2} \vee \mathbf{z_1} \sim_r \mathbf{z_2} \end{aligned} \tag{2.49}$$

then, $\preccurlyeq_r$ is a *total preorder* on Z.

Proof: It is immediate to conclude that $\preccurlyeq_r$ is both reflexive and transitive. Since $\mathbf{z_1} \sim_r \mathbf{z_2}$ is true for all $\mathbf{z_1}, \mathbf{z_2} \in Z$, including $\mathbf{z_1} \neq \mathbf{z_2}$, $\preccurlyeq_r$ is not antisymmetric. Then, $\preccurlyeq_r$ is a preorder. It also satisfies the totality axiom, because $\mathbb{N}$ is totally ordered, therefore being a total preorder. ■

Proposition 2.20 [107]: The Pareto-ranking relation induces a total preorder in X.

Being a total preorder, the Pareto-ranking relation allows to compare and sort all the elements in Z into disjoint equivalence classes

$$F_i = \{\mathbf{z} \in Z : r(\mathbf{z}) = i\} \tag{2.50}$$

called *fronts*. We say elements of F_i are solutions of *rank* i. Also, for notational purposes, we assume (in a minimization sense) that $F_1 = Z_{PF}$. Then, $r(\mathbf{z}) = 1 \Leftrightarrow \mathbf{z} \in Z_{PF}$

The *Multi-objective GA* (MOGA) [108] was the first to combine Pareto-ranking with diversity maintenance among the nondominated points. At every generation, individuals are sorted by the *dominance rank* $r(\mathbf{z}) = 1 + n_{\prec}(\mathbf{z})$, where $n_{\prec}$ is the number of other individuals dominating $\mathbf{z}$. An average fitness, inversely proportional to the rank, is assigned and fitness sharing is then applied front-wise, in the objective space. Stochastic universal selection, crossover and mutation are applied as usual. The *Nondominated Sorting GA* (NSGA) [109] ranks solutions in a *nondominated sorting* procedure. Nondominated individuals of the whole population are allocated in F_1 and assigned rank 1. This process is repeated for the remaining population, until every individual is ranked. Fitness is assigned front-wise using fitness sharing, in the search space, such that the worst fitness assigned to F_i is better than the best fitness assigned to F_{i+1}. Stochastic remainder selection, crossover and mutation are operators applied as usual. The *Niched-Pareto GA* (NPGA) [110] neither assigns fitness nor sorts the population. It applies a ranking-based binary tournament, for parent selection, with two possible outcomes. Two individuals of the population are compared against a subset of the population. Of the two, if one has $r = 1$ and the other has $r > 1$, the former is chosen. However, if both have either $r = 1$ or $r > 1$, both are checked against the current offspring population. The one with the smaller niche count is chosen as parent. Crossover and mutation operators are applied in the usual manner.

Elitist Multi-objective EAs

Elitism in the context of EAs means that the fittest individual(s) found so far always survives to the next generation. In this respect, all nondominated individuals discovered by a MOEA are considered elite solutions. However, in this case, applying an elitist strategy is not as straightforward, due to the large number of possible elite solutions [103]. The *Vector-optimized Evolution Strategy* (VOES) [98] is an ES with a fitness assignment similar to the VEGA. The *Elitist NSGA* (NSGA-II) [111] combines a $(n + n)$ elitist nondominated sorting strategy, with a dynamic niching strategy, based on the crowding distance [98]. No fitness is assigned. The new population is filled by the best fronts of the parent-plus-offspring population, one at a time, until n individuals are allocated. In the last allowed

front, only the individuals in the least crowded region of that front survive. Crowded tournament selection [98], crossover and mutation generate n offspring solutions, at every generation.

Many MOEAs implementations use an *archive population* that stores all, or a part of, the Pareto optimal solutions generated in the current population, up until generation t. As with single-population MOEAs, we define a *known Pareto optimal set*, $X_{PO_{known}}(t)$, and a *known Pareto front* $Z_{PF_{known}}(t)$ [104], with respect to the archive, such that $Z_{PF_{known}} = Z_{PF_{known}}(t)$ and $X_{PO_{known}} = X_{PO_{known}}(t)$ are now the final sets returned by the MOEA. Therefore, after convergence, it is expected that $Z_{PF_{known}} \subseteq Z_{PF}$ and $X_{PO_{known}} \subseteq X_{PO}$.

The updating scheme of the archive population is rather universal among MOEAs. The basic idea is to sort the current population and to add the rank 1 individuals to the archive. At every generation the archive must be sorted, in order to rank the new solutions. For practicality, some authors limit the archive population to just a few solutions. Diversity measures are then applied for that purpose. The *Strength Pareto EA* (SPEA) [112] applies a clustering algorithm [98] to keep just a predetermined number of the most spread rank 1 solutions of the archive, with an *elitist selection* strategy. Fitness is assigned to the members of both populations by means of the *strength parameter*, which favors solutions dominating, and dominated by, the least number of individuals. Binary tournament selection is applied to the combined current and archive populations. Crossover and mutation are applied as usual. Similarly, in the *Pareto Envelope-based Algorithm* (PESA) [113], a cell-based density measure [98] is used to maintain an archive of fixed size. No fitness is assigned. Offspring solutions are generated by crossover and mutation of the archive members, based on their density information, as well. The PESA-II [114] follows a more direct selection mechanism. Instead, it selects cells in the objective space, giving preference the sparsely occupied ones. Solutions within a cell are then randomly selected as parents. Also, in the $(1+1)$-*Pareto-archived Evolution Strategy* (PAES) [115], no fitness is assigned. The acceptance of an offspring solution in the archive is based on its dominance status and its cell-based density, with respect to the archived solutions. To determine the parent for the next generation, the current parent and offspring are compared for dominance. If none dominates the other, the least dense of the two becomes parent. At each generation, one offspring solution is generated by Gaussian mutation of the parent solution.

2.2.2.2 Hybrid Multi-objective EAs

Hybridization of MOEAs with alternative metaheuristics has been suggested to improve the approximation of the set $Z_{PF_{knwon}}$ to the actual Pareto front and to generate a more distributed set of nondominated points in $Z_{PF_{knwon}}$ [70]. In principle, hybrid MOEAs offer the ability to better balance the exploration and the exploitation or to intensify one of the features. However, managing the binomial exploration/exploitation, in MOEAs, may lead to a deteriorated computational efficiency, if the hybrid framework is not carefully planned [70]. In particular, because in multi-objective optimization the convergence goals must be attained in two spaces, *X* and *Z* [70]. Following the taxonomy adopted to classify hybrid EAs, in Sect 2.1, we consider collaborative and integrative hybrid MOEAs.

Collaborative hybrid MOEAs are a set of self-contained metaheuristics, where at least one MOEA is included, that interchange information between them, while running in parallel or sequentially. In Coevolutionary MOEAs [70] various populations evolve independently, through a MOEA. Examples can be found in [116,117]. In [70], it is suggested the idea of coevolving a MOEA with alternative metaheuristics. In [118,119] authors propose to run a MOEA and apply a local search procedure afterwards to improve the nondominated set. The Multi-objective Hybrid Optimization Algorithm [120] is an elitist framework that adaptively alternates between three MOEAs to evolve a single

population, based on their success. *Integrative hybrid MOEAs* are a category of metaheuristics where one master algorithm commands the search process, while a subordinate algorithm is embedded in the master's procedure. MOEAs may work either as the master or the subordinate algorithm. *Multi-objective Memetic Algorithms* (MOMAs) are the most popular instantiation. *Lamarckian MOMAs* improve solutions by the explicit manipulation of genetic strings with local optimizers, during evolution. Example are the *Multi-objective Genetic Local Search* [121] *Memetic PAES* [122], among others. *Baldwinian MOEAs* improve individual selection probabilities, based on individuals' potential to improve, without explicit genetic manipulation. Examples include the Multi-objective Hybrid Optimization Algorithm [120], Ishibuchi *et al.* hybrid MOEA [12], among others.

2.2.2.3 Constraint handling

In multi-objective optimization, constraints divide both the objective and the search spaces into two regions – the feasible and infeasible regions. Like in single-objective optimization, all nondominated points, and the respective Pareto optimal solutions, must be feasible. So far, from (2.39) on, the theorical developments assumed an unconstrained formulation of the MO-NLP problem (we considered only the spaces Z and X), but these remain true if, in addition to $Z \subseteq \mathbb{R}^k$ and $X \subseteq \mathbb{R}^N$, we consider $S_Z \subseteq Z$ and $S_X \subseteq X$.

However, in most constrained engineering problems, both S_Z and S_X are unknown and the same technical difficulties faced by EAs, when dealing with constrained problems, are also found in MOEAs, yet aggravated by the existence of a partially ordered objective space and multiple minimal elements. Like EAs, MOEAs were conceived as global optimizers for unconstrained problems and work over the entire search space X. An additional procedure is needed to distinguish between feasible and infeasible solutions and to steer the evolution towards the feasible regions. The use of penalty functions is popular among practitioners, defining penalized objective functions, from which the dominance relation between solutions can be established and any MOEA constructed for unconstrained optimization is applicable [98]. Alternative and more systematic sorting mechanisms, based on dominance and niching, exist in the literature, such as the ones by Jimenez *et al*. [124] and Ray *et al*. [125]. In particular, Deb [98] proposed an alternative to the Pareto dominance strict partial order on Z, called *constrain-dominance*, as follows.

Definition 2.23 (*constrain-dominance*) [98]: Let $\mathbf{z_1}, \mathbf{z_2} \in Z$ and $\xi(\mathbf{z}): Z \to \mathbb{R}$ be a constraint-violation function. A solution $\mathbf{z_1}$ is said to '*constrain-dominate*' a solution $\mathbf{z_2}$ (or $\mathbf{z_1} \prec_c \mathbf{z_2}$), if any of the following conditions is true:

1) $\mathbf{z_1}$ is feasible and $\mathbf{z_2}$ is not.
2) $\mathbf{z_1}$ and $\mathbf{z_2}$ are both infeasible and $\xi(\mathbf{z_1}) < \xi(\mathbf{z_2})$.
3) $\mathbf{z_1}$ and $\mathbf{z_2}$ are both feasible and $\mathbf{z_1} \prec \mathbf{z_2}$.

Similar to $\preccurlyeq$, the partial order $\preccurlyeq_c$ (*constrain-dominant or equal*) is constructed from (2.43) (not (2.44)) and $(Z, \preccurlyeq_c)$ is a partially ordered objective space. Also $(X, \preccurlyeq_c)$ is a preordered search space. Notice that $\preccurlyeq_c$ and $\preccurlyeq$ are the same in the feasible region but that $\preccurlyeq_c$ will order the infeasible region according to the scalar function $\xi(\mathbf{z})$ [98]. Indeed, only the feasible region is partially ordered, w.r.t. $\preccurlyeq_c$, and therefore the set of minimal elements of $(S_Z, \preccurlyeq)$ and $(Z, \preccurlyeq_c)$ coincide. In the presence of constraints, we can also apply the same Pareto-ranking relation and the associated sorting mechanisms described above to classify the population of a MOEA into different classes of domination (fronts), noting that infeasible solutions belong to the same front only if their constraint-violation is equal. The constrain-dominance principle is generic and can be used with any MOEA [98].

References

[1]. Bazaraa M., Sherali H.D., Shetty C.M. Nonlinear programming: theory and algorithms. 3rd ed. John Willey & Sons; 2006.

[2]. Rao SS. Engineering Optimization: Theory and Practice. 4th ed. John Willey & Sons; 2009.

[3]. Luenberger DG. Linear and nonlinear programming. 2nd ed. Springer-Verlag US; 2003.

[4]. Nash SG, Sofer A. Linear and nonlinear programming. 1st ed. McGraw-Hill; 1996

[5]. Motzkin TS. Two consequences of the transposition theorem of linear Inequalities. Econometrica 1951; 19: 184-185.

[6]. John F. Extremum problems with inequalities as side conditions. In: Friedrichs KO, Neugebauer OE, Stoker JJ, editors. Studies and Essays, Courant Anniversary Volume, New York: Wiley (Interscience); 1948, p. 187-204.

[7]. Mangasarian OL, Fromovitz S. The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints. J. Math. Anal. Appl 1967; 17: 37-47

[8]. Bazaraa MS, Goode JJ, Shetty CM. Constraint qualifications revisited. Management Science 1972; 18(9): 567-73.

[9]. Kuhn HW, Tucker AW. Nonlinear programming. In: J. Neyman, editor. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley: University of California Press; 1951, p. 481-92.

[10]. Arora JS. Introduction to optimum design. 3rd ed. Elsevier; 2012.

[11]. Sörensen K, Sevaux M, Glover F. A history of metaheuristics. In: Martí R, Pardalos PM, Resende GC, editors. Handbook of Heuristics, Springer; 2018, 791-806.

[12]. Holland J. Adaptation in natural and artificial systems. The University of Michigan Press; 1975.

[13]. Corne D, Lones MA. Evolutionary algorithms. In: Martí R, Pardalos PM, Resende GC, editors. Handbook of Heuristics, Springer; 2018, 409-30.

[14]. Goldberg DE. Genetic algorithms in search, optimization and machine learning. Addison-Wesley; 1989.

[15]. Hauschild M, Pelikan M. An introduction and survey of estimation of distribution algorithms. Swarm Evol Comput 2011; 1(3):111–128.

[16]. Das S, Suganthan PN. Differential evolution: a survey of the state-of-the-art. IEEE Trans Evol Comput 2011; 15(1):4–31.

[17]. Koza J. Genetic programming: on the programming of computers by means of natural selection. MIT Press; 1992.

[18]. Singh G, Deb K. Comparison of multi-modal optimization algorithms based on evolutionary algorithms. ACM; 2006.

[19]. Moscato P. On evolution, search, optimization, genetic algorithms and martial arts: towards memetic algorithms. Caltech concurrent computation program, C3P report 826; 1989.

[20]. García-Martínez C, Rodriguez F, Lozano M. Genetic algorithms. In: Martí R, Pardalos PM, Resende GC, editors. Handbook of Heuristics, Springer; 2018, 431-64.

[21]. Herrera F, Lozano M, Verdegay J. Tackling real-coded genetic algorithms: operators and tools for behavioural analysis. Artif Intell Rev 1998; 12:265–319.

[22]. Lones MA, Tyrrell AM. Regulatory motif discovery using a population clustering evolutionary algorithm. IEEE/ACM Trans Comput Biol Bioinform 2007; 4(3):403–414.

[23]. Miller J.F. Cartesian Genetic Programming. In: Miller J, editor. Cartesian Genetic Programming. Natural Computing Series. Springer; 2011.

[24]. Smith A, Coit D, Baeck T, Fogel D, Michalewicz Z. Penalty functions. In: Bäck T, Fogel DB, Michalewics Z, editors. Handbook on evolutionary computation. Oxford University Press; 1997, p. C5.2:1–C5.2:6.

[25]. Yeniay Ö. Penalty function methods for constrained optimization with genetic algorithms. Math Comput Appl 2005; 10(1): 45–56.

[26]. Syswerda G. Uniform crossover in genetic algorithms. In: Schaffer D J, editor. Proceedings of the Third International Conference on Genetic Algorithms, Morgan Kaufmann; 1989, p. 2-9.

[27]. Eshelman L. The CHC adaptive search algorithm: how to have safe search when engaging in nontraditional genetic recombination. Foundations of genetic algorithms 1991; 1: 265-83.

[28]. Spears WM, DeJong SK. On the virtues of parametrized uniform crossover. In: Belew RK, Booker LB, editors. Proceedings of Fourth International Conference on Genetic Algorithms. Morgan Kaufmann; 1991. p. 230–6.

[29]. Goldberg D, Lingle R. Alleles, Loci and the traveling salesman problem. In: Proceedings of the international conference on genetic algorithms; 1985, p. 154–9.

[30]. Davis L. Adaptive algorithms to epistactic domains. In: Proceedings of the international conference on artificial intelligence; 1985, p. 162–4.

[31]. Herrera F, Lozano M. Fuzzy adaptive genetic algorithms: design, taxonomy, and future directions. Soft Comput 2003; 7(8): 545–62.

[32]. Herrera F, Lozano M, Sánchez A. A taxonomy for the crossover operator for real-coded genetic algorithms: an experimental study. Int J Intell Syst 2003; 18(3):309–338.

[33]. Eiben A, Smith J. Introduction to evolutionary computation. Natural computing series. Springer; 2003.

[34]. Araujo L, Merelo J. Diversity through multiculturality: assessing migrant choice policies in an island model. IEEE Trans Evol Comput 2011; 15(4):456–469.

[35]. Rudolph G. Convergence analysis of canonical genetic algorithms. IEEE Transactions on Neural Networks 1994; 5(1): 96-101.

[36]. Serpell M, Smith J. Self-adaption of mutation operator and probability for permutation representations in genetic algorithms. Evol Comput 2010; 18(3):491–514.

[37]. Janikow C, Michalewicz Z. An experimental comparison of binary and floating-point representation in genetic algorithms. In: Proceedings of the fourth international conference on genetic algorithms; 1991, p. 31–36.

[38]. Storn R, Price K. Differential evolution - a simple and efficient adaptive scheme for global optimization over continuous spaces. Technical Report TR-95-012, ICSI, Berkeley; 1995.

[39]. Goldberg DE, Deb K. A Comparative Analysis of Selection Schemes Used in Genetic Algorithms. Foundations of Genetic Algorithms; 1991, p. 69–93.

[40]. Baker JE. Reducing bias and inefficiency in the selection algorithm. In: Proceedings of the Second International Conference on Genetic Algorithms; 1987, p. 14-21.

[41]. Baker JE. Adaptive selection methods for genetic algorithms. In: Proceedings of an International Conference on Genetic Algorithms and Their Applications; 1985, p. 100-111.

[42]. Fogel DB. Evolutionary computation: the fossil record. Wiley-IEEE Press; 1998.

[43]. Harik, GR, Lobo FG, Goldberg E. The Compact Genetic Algorithm. IEEE Transactions On Evolutionary Computation 1999; 3(4): 287-97.

[44]. De Jong K. Analysis of the behavior of a class of genetic adaptive systems. PhD thesis. University of Michigan; 1975.

[45]. Eiben AE, Aarts EHL, Van Hee KM. Global convergence of genetic algorithms: A Markov chain analysis. In: Schwefel H-P, editor. Parallel Problem Solving from Nature.

[46]. Whitley LD, Kauth J. Genitor: A different genetic algorithm. In: Proceedings of the Rocky Mountain Conference on Artificial Intelligence; 1988, p. 118–130.

[47]. Hutter M, Legg S. Fitness uniform optimization. IEEE Trans Evol Comput 2006; 10(5): 568–589.

[48]. Kuo T, Hwang S. A genetic algorithm with disruptive selection. IEEE Trans Syst Man Cybern 1996; 26(2):299–307.

[49]. Wong YY, Lee KH, Leung KS, Ho CW. A novel approach in parameter adaptation and diversity maintenance for genetic algorithms. Soft Comput 2003; 7(8):506–515.

[50]. Kazimipour B, Li X, Qin A. A review of population initialization techniques for evolutionary algorithms. In: Proceedings of the IEEE congress on evolutionary computation; 2014, p. 2585–2592.

[51]. Fernandes C, Rosa AC. Self-adjusting the intensity of assortative mating in genetic algorithms. Soft Comput 2008; 12: 955-979.

[52]. Petrowski A. A new selection operator dedicated to speciation. In: Proceedings of the 7th international conference on genetic algorithms; 1997, p. 144–151.

[53]. Mahfoud S. Crowding and preselection revised. In: Männer R, Manderick B, editors. Parallel problem solving from nature, volume 2. Elsevier Science; 1992, p. 27–36.

[54]. Harik G. Finding multimodal solutions using restricted tournament selection. In: Proceedings of the international conference on genetic algorithms; 1995, p. 24–31.

[55]. Conceição António C, Castro CF, Sousa LC. Optimziation of metal forming processes. Comput Struct 2004; 82: 1425-1433.

[56]. Conceição António CC. A hierarchical genetic algorithm for reliability based design of geometrically nonlinear composite structures. Comp Struct 2001; 54: 37-47.

[57]. Lim TY. Structured population genetic algorithms: a literature survey. Artif Intell Rev 2014; 41: 385-399.

[58]. Zaharie D. Influence of crossover on the behavior of differential evolution algorithms. Appl Soft Comput 2009; 9(3):1126–1138.

[59]. Bhandarkar S, Zhang H. Image segmentation using evolutionary computation. IEEE Trans Evol Comput 1999; 3(1):1–21.

[60]. Gao J, Sun L, Gen M. A hybrid genetic and variable neighborhood descent algorithm for flexible job shop scheduling problems. Comput Oper Res 2008; 35(9): 2892–2907.

[61]. Talbi E. A taxonomy of hybrid metaheuristics. J Heuristics 2002; 8(5):541–564.

[62]. Talbi EG, Bachelet V. Cosearch: a parallel cooperative metaheuristic. J Math Model Algorithms 2006; 5(1):5–22.

[63]. Chelouah R, Siarry P. Genetic and Nelder-Mead algorithms hybridized for a more accurate global optimization of continuous multiminima functions. Eur J Oper Res 2003; 148(2): 335–348.

[64]. García-Martínez C, Lozano M, Herrera F, Molina D, Sánchez A. Global and local real-coded genetic algorithms based on parent-centric crossover operators. Eur J Oper Res 2008; 185(3):1088–1113.

[65]. Molina D, Lozano M, García-Martínez C, Herrera F. Memetic algorithms for continuous optimization based on local search chains. Evol Comput 2010; 18(1):27–63.

[66]. Coello Coello CA, Lamont GB, Van Veldhuizen DA. Evolutionary Algorithms for Solving Multi-Objective Problems. 2nd ed. Springer; 2007.

[67]. Neri F, Cotta C. Memetic algorithms and memetic computing optimization: a literature review. Swarm Evol Comput 2012; 2: 1–14.

[68]. García-Martínez C, Lozano M, Rodríguez-Díaz F. A simulated annealing method based on a specialised evolutionary algorithm. Appl Soft Comput 2012; 12(2):573–588;

[69]. Lozano M, García-Martínez C. Hybrid metaheuristics with evolutionary algorithms specializing in intensification and diversification: overview and progress report. Comput Oper Res 2010; 37:481–497.

[70]. Coello Coello CA. Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art. Comput. Methods Appl. Mech. Engrg 2002; 191: 1245-1287.

[71]. Smith AE, Coit DW. Constraint handling techniques – penalty functions, in: Back T, Fogel DB, Michalewic, editors. Handbook of Evolutionary Computation, Oxford University Press and Institute of Physics Publishing; 1997.

[72]. Morales AK, Quezada CV. A universal eclectic genetic algorithm for constrained optimization. In: Proceedings of the 6th European Congress on Intelligent Techniques and Soft Computing, EUFIT'98; 1998, p. 518–522.

[73]. Homaifar A, Lai SHY, Qi X, Constrained optimization via genetic algorithms. Simulation 1994; 62 (4) 242–254.

[74]. Joines JA, Houck CR. On the use of non-stationary penalty functions to solve nonlinear constrained optimisation problems with GA's. In: ICEC-94 [229], p. 579–584.

[75]. Kazarlis S, Petridis V, Varying fitness functions in genetic algorithms: Studying the rate of increase of the dynamic penalty terms, in: Eiben AE, Back T, Schoenauer M, Schwefel H-P, editors. Parallel Problem Solving from Nature V – PPSN V, Springer; 1998.

[76]. Hadj-Alouane AB, Bean JC. A genetic algorithm for the multiple-choice integer program, Operations Research 1997; 45: 92–101.

[77]. Eiben AE, Raué PE, Ruttkay Zs. GA-easy and GA-hard constraint satisfaction problems. In: M. Meyer, editor. Proceedings of the ECAI'94 Workshop on Constraint Processing, Springer;1995, p. 267–284.

[78]. Rasheed K. An adaptive penalty approach for constrained genetic-algorithm optimization. In: Koza JR, Banzhaf W, Chellapilla W, Deb K, Dorigo M, Fogel DB, Garzon MH, Goldberg DE, Iba H, Riolo RL, editors. Proceedings of the Third Annual Genetic Programming Conference, Morgan Kaufmann; 1998, p. 584–590.

[79]. Crossley WA, Williams EA. A study of adaptive penalty functions for constrained genetic algorithm based optimization. In: AIAA 35th Aerospace Sciences Meeting and Exhibit, AIAA; 1997.

[80]. Liepins GE, Vose MD. Representational issues in genetic optimization. J. Exp. Theoret. Comput. Sci. 1990; 2 (2): 4–30.

[81]. Michalewicz Z, Nazhiyath G. Genocop III: A coevolutionary algorithm for numerical optimisation problems with nonlinear constraintrs. In: Proceedings of the 1995 IEEE Conference on Evolutionary Computation. IEEE Press; 1995, p. 647–651

[82]. Orvosh D, Davis L. Using a genetic algorithm to optimize problems with feasibility constraints, In: Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE Press; 1994, p. 548–553.

[83]. Michalewicz Z, Genetic Algorithms + Data Structure = Evolution Programs, 3rd ed. Springer; 1996.

[84]. Nakano R. Conventional genetic algorithm for job shop problems. In: Belew RK, Booker LB, wditors. Proceedings of the Fourth International Conference on Genetic Algorithms, Morgan Kaufmann; 1991, p. 474–479.

[85]. Le Riche RG, Haftka RT. Improved genetic algorithm for minimum thickness composite laminate design, Compos Engrg 1994; 3 (1): 121–139.

[86]. Tate DM, Smith AE. A genetic approach to the quadratic assignment problem. Computers and Operations Research (1995; 22 (1): 73–78.

[87]. Muhlenbein, Parallel genetic algorithms in combinatorial optimization, in: Balci, Sharda, Zenios, editors. Computer Science and Operations Research. Pergamon Press; 1992, p. 441–456.

[88]. Rudolph G. Finite Markov chain results in evolutionary computation: a tour d'horizon. Fundamenta Informaticae 1998; 35: 67-89.

[89]. Fogel DB. Asymptotic convergence properties of genetic algorithms and evolutionary programming: analysis and experiments. Cybernetics and Systems; 1994; 25(3): 389-407.

[90]. Marler RT, Arora JS. Survey of multi-objective optimization methods for engineering. Struct Multidisc Optim 2004; 26: 369-395.

[91]. Rudolph G. Evolutionary search for minimal elements in partially ordered finite sets. In: Porto VW, Saravanan N, Waagen D, Eiben AE, editors. Evolutionary Programming VII. EP 1998. Lecture Notes in Computer Science, vol 1447. Springer; 1998.

[92]. Ehrgott M. Multicriteria optimization. 2nd ed. Springer; 2003.

[93]. Fernandes RL, Ricou M. Introdução à álgebra. 2ª ed. IST Press; 2014.

[94]. Simovici DA, Djeraba C. Mathematical Tools for Data Mining. Springer; 2008.

[95]. Schöder BS. Ordered sets: an introduction. Springer Science+Business Media; 2003.

[96]. Ghorpade SR, Limaye BV. A course in multivariable calculus and analysis. Springer; 2010.

[97]. Apostol TM. Calculus: volume II. 2nd ed. John Willey & Sons; 1969.

[98]. Deb K. Multi-objective optimization using evolutionary algorithms. John Willey & Sons; 2001.

[99]. Chankong V, Haimes YY. Multiobjective decision making: theory and methodology. Elsevier Science publishing; 1983.

[100]. Halmos PR. Naïve set theory. D. Van Nostrand Company; 1960.

[101]. Kumar R, Banerjee N. Running time analysis of a multiobjective evolutionary algorithm. In: Wright AH, De Jong K, Vose MD, Scmitt LM. Foundations of Genetic Algorithms: 8th International Workshop; 2005.

[102]. Rudin W. Principles of mathematical analysis. 3rd ed. McGraw-Hill; 1964.

[103]. Konak A, Coit DW, Smith AE. Multi-objective optimization using genetic algorithms: A tutorial. Reliability Engineering and System Safety 2006; 91: 992-1007.

[104]. Van Veldhuizen DA. Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations. PhD thesis. Air Force Institute of Technology; 1999.

[105]. Schaffer JD. Some experiments in machine learning suing vector evaluated genetic algorithms. PhD thesis. TN: Vanderbilt University; 1984.

[106]. Hajela P, Lin C-Y. Genetic algorithms in multi-criterion optimal design. Structural Optimization 1993; 4(2): 99-107.

[107]. Loshchilov I, Schoenauer M, Sebag M. Not all parents are equal for the MO-CMA-ES. In: Takahashi RHC, Deb K, Wanner EF, Greco S, editors. Proceedings of the Evolutionary Multi-criterion Optimization: 6th International Conference, EMO 2011. Springer; 2011.

[108]. Fonseca CM, Fleming PJ. Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization. In: Forrest S, editor. Genetic Algorithms: Proceedings of the Fifth International Conference. Kaufmann; 1993.

[109]. Srinivas N, Deb K. Multi-objective function optimziation using nondominated sorting genetic algorithms. Evolutionary Computation Journal 1994; 2(3): 221-248.

[110]. Horn J, Nafploitis N, Goldberg D. A niched Pareto genetic algorithm for multi-objective optimization. In: Proceedings of the first IEEE Conference on Evolutionary Computation; 1994, p. 82-87.

[111]. Deb K, Pratap S, Meyarivan T. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In: Proceedings of the Parallel Problem Solving from Nature VI; 2000, p.849-858.

[112]. Zitzler E, Thiele L. An evolutionary algorithm for multiobjective optimization: the strength Pareto approach. Technical report 43, Computer Engineering and Network Laboratory, Swiss Federal Institute of Technology; 1998.

[113]. Corne DW, Knowels JD, Oates MJ. The Pareto Envelope-Based Selection Algorithm for Multiobjective Optimization. In: Parallel Problem Solving from Nature PPSN VI. PPSN 2000. Lecture Notes in Computer Science, vol 1917. Springer; 2000.

[114]. Corne DW, Jerram NR, Knowels JD. PESA-II: region-based selection in evolutionary multiobjective optimization. In: Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation; 2001, p. 283-290.

[115]. Knowles J, Corne D. The Pareto archived evolution strategy: a new baseline algorithm for Pareto multiobjective optimization. In: Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406); 1999, p. 98-105.

[116]. Neef M, Thierens D, Arciszewski H. A Case Study of a Multiobjective Recombinative Genetic Algorithm with Coevolutionary Sharing. In: 1999 Congress on Evolutionary Computation; 1999, p. 796–803.

[117]. Coello Coello CA and Reyes Sierra. A Coevolutionary Multi-Objective Evolutionary Algorithm. In: Proceedings of the 2003 Congress on Evolutionary Computation (CEC'2003); 2003, p. 482–489.

[118]. Deb K, Goel T. A Hybrid Multi-Objective Evolutionary Approach to Engineering Shape Design. In: First International Conference on Evolutionary Multi-Criterion Optimization; 2001, p. 385–399.

[119]. Leiva HA, Esquivel SC, Gallard RH. Multiplicity and Local Search in Evolutionary Algorithms to Build the Pareto Front. In: Proceedings of the XX International Conference of the Chilean Computer Science Society; 2000, p. 7-13.

[120]. Moral RJ, Dulikravich GS, Multi-Objective Hybrid Evolutionary Optimization with Automatic Switching Among Constituent Algorithms. AIAA Journal 2008; 46(3): 673- 681.

[121]. Ishibuchi H, Murata T. Multi-Objective Genetic Local Search Algorithm. In: Proceedings of the 1996 International Conference on Evolutionary Computation; 1996.

[122]. Keerativuttiumrong N, Chaiyaratana N, Varavithya V. Multi-objective Co-operative Co-evolutionary Genetic Algorithm. In: Parallel Problem Solving from Nature—PPSN VII; 2002.

[123]. Ishibuchi H, Doi T, Nojima Y. Incorporation of Scalarizing Fitness Functions into Evolutionary Multiobjective Optimization Algorithms. In: Lecture Notes in Computer Science; 2006, p. 493–502.

[124]. Jimenez F, VErdegay JL, Skarmeta AFG. Evolutionary Techniques for Constrained Multiobjective Optimization Problems. In: Proceedings of Workshop on Multi-criterion Optimization Using Evolutionary Methods; 1999.

[125]. Ray T, Tai K, Seow KC. An evolutionary algorithm for multiobjective optimization. Engineering Optimization 2001; 33(3): 399-424.

III
RELIABILITY-BASED ROBUST DESIGN OPTIMIZATION

Summary

*

This chapter presents the *Reliability-based Robust Design Optimization* (RBRDO) problem of composite laminate mechanical systems, to which we propose new algorithmic solutions, in the next chapters. In Sect. 3.1 the concept of stochastic design optimization as an engineering design methodology is introduced, including the main components of the proposed RBRDO problem. The existence of uncertainty associated with structural design is discussed as well and specific definitions of random variables and stochastic response functionals are given. In Sect. 3.2, the concept of robustness is introduced and a global robustness measure based on the probability theory is defined. Similarly, in Sect. 3.3, the concept of probabilistic reliability, as a measure of structural integrity, is presented. The particular probability spaces associated with the probabilistic reliability assessment and the related probability measure of structural integrity are defined. Sect. 3.4 provides the definition of the response of the mechanical system in terms of displacements and stresses. The Tsai-Wu quadratic failure criterion for composite structures is presented. Sect. 3.5 provides a formal statement of the RBRDO problem, as a stochastic nonlinear programming problem.

3.1 Stochastic design optimization

The process of *engineering design* consists of four stages [1]:

1) Identification of the design targets and needs;
2) Formal statement of the problem;
3) Creation of one or more design alternatives and the analysis of their performance, using engineering sciences;
4) Application of a methodology for the selection of the design that satisfies the targets and needs the best.

Such sequential description of the engineering design is in truth incomplete, because the meaning of "best" is vague and the choice of a methodology for the selection of design alternatives influences each of the stages and the decisions made there. Different methodologies require different statements of the design problem and often of proper design targets and requirements. The number of design alternatives to be analyzed and how each alternative is constructed also depend on the chosen methodology.

Design optimization is an engineering design methodology that formulates design problems as *mathematical optimization* problems, with the purpose of identifying the *optimal design* among many design alternatives. Given the complexity of many of the current engineering problems, design optimization is, in general, formulated as a *nonlinear programming problem*. Therefore, it is usually associated with a set of optimization heuristics, which automate the design methodology, once the first two design stages are established. In that sense, a design optimization problem is not complete (well-posed) without the previous definition of the following design elements:

- *Design variables*, $\mathbf{d} \in D$, are the search variables of a design optimization problem and define the search space, called the *design space* $D \subset \mathbb{R}^{N_\mathbf{d}}$;
- *System parameters*, $\boldsymbol{\pi} \in \Pi$, are variables that condition the search process, in the sense that their values influence the state of the system. These are fixed by the designer(s) at the beginning of the search process. $\Pi \subset \mathbb{R}^{N_\boldsymbol{\pi}}$ is called the *parameter space*.
- *State variables*, $\mathbf{u} \in U$, are quantities that completely determine the state of a system. Often, a state is determined by $\mathbf{u} = \mathbf{u}(\mathbf{d}, \boldsymbol{\pi})$. $U \subset \mathbb{R}^{N_\mathbf{u}}$ is called the *state space.*
- *Objective functions* $f_i : D \times \Pi \times U \to \mathbb{R}$ are the functional representation of the i-th design target, usually translated as *performance* or *cost* measures.
- *System response functionals* $\varphi_j : D \times \Pi \times U \to \mathbb{R}$ are the functional representation of the state of the system.
- *Design Constraints* are mathematical relations, in the shape of *inequalities* and *equalities*, that relate the design variables, system parameters and state variables, dividing the design space into two disjoint partitions: the *feasible space* and the *infeasible space*. These usually result from multidisciplinary topics imposed by design codes and/or technical limitations and are often written is terms of the system response functionals.

In general, each objective function, system response functional and design constraint can be:

- an *explicit expression* of the variables. These are usually derived directly from de laws of engineering sciences or empirical models.
- an *implicit expression* of the variables. These usually represent a formal statement of a complex process of numerical simulation, involving an inner cycle of calculations. In engineering, typical examples are the numerical solutions of systems of differential equations, describing physical phenomena.

Regarding the nature of the design elements, design optimization methodologies can be either *deterministic* or *stochastic*. Determinism involves the assumption that a system is completely determined by a finite set of known mathematical relations and variables. However, the determinism lies in a system's mathematical description, often based on a limited understanding of reality. Indeed, different mathematical models of the same physical system may predict different responses, which in turn may vary from the observed reality, depending on their accuracy. Deterministic design methodologies have proven very useful, in design optimization, but often rely on a simplistic representation of reality. In truth, there is uncertainty associated with the majority of engineering design problems. Uncertainty can be found at every level of the design process and is usually modeled as noise in the variables of the problem and/or directly in the functions of the optimization model. There are also several kinds of uncertainty, each representing a way of modeling our partial knowledge of reality [2,3]. In particular, in this thesis, we are concerned with *random uncertainty* [4].

Stochastic design optimization is the methodology that combines design optimization with *stochastic programming*. In the field of mathematical optimization, stochastic programming is an optimization framework involving random uncertainty, in the sense that it takes advantage of the fact that the probability distributions governing uncertainty are *known* [5]. Unlike other formulations, such as linear, nonlinear or convex programming, there is not a general statement for the stochastic programming problem, because there are several concepts of uncertainty propagation and quantification [5]. A well-posed stochastic design optimization problem requires a complete definition of both.

Regarding *uncertainty propagation*, we consider uncertainty to exist at the most elementary level in engineering design, by defining an additional set of *random variables*, as follows.

Definition 3.1 (*random variable*): Let $\Delta \subseteq \mathbb{R}^{N_{\mathbf{x}}}$, where $N_{\mathbf{x}}$ is the number of random variables, and (Δ, A, P) be a probability space. For a vector of random deviations $\boldsymbol{\delta} \in \Delta$, $\mathbf{x} : \Delta \to \Omega$ is a vector of random variables of the kind:

$$\mathbf{x} = \boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\delta} \tag{3.1}$$

where $\Omega \subseteq \mathbb{R}^{N_{\mathbf{x}}}$ and $\boldsymbol{\mu}_{\mathbf{x}} \in \Omega$ is the vector of mean-values, around which random variables oscillate.

This way, there is a consistent propagation of random uncertainty to the functions in a design optimization problem, which become random as well, whenever dependent on the set of random variables **x**. As an adaptation of Definition 2.5, consider the following definition.

Definition 3.2 (*stochastic response functional*): Let (Ω, A, P) be a probability space and let $\mathbb{f}$ be the set of all response functionals $\varphi: D \times \Pi \times U \to \mathbb{R}$. Then, $\phi: \Omega \to \mathbb{f}$ is a random variable, whose realization $\phi(\mathbf{x}) : D \times \Pi \times U \to \mathbb{R}$ is a *stochastic response functional* $\varphi(\mathbf{x}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u})$, for all $\mathbf{x} \in \Omega$. To $\varphi(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u})$ we call the *deterministic realization of* ϕ, for $\boldsymbol{\mu}_{\mathbf{x}} \in \Omega$.

In the two previous definitions, random variables were assumed as additional variables of the design problem. Yet, both definitions remain valid if, **d**, **π** and **u** are themselves random variables. In such case, the respective mean-values, describe the deterministic configuration of the system.

Regarding *uncertainty quantification*, we consider the concepts of *robustness* and *reliability*. The first is related with the notion of variability of a design. The latter is related with the notion of integrity. Both concepts have several interpretations, in different branches of science and engineering. In the realm of structural design, we restrain ourselves to a probabilistic one. The subject is introduced in the following sections.

3.2 Robustness assessment

In design optimization, *robustness* is the concept that measures the variability associated with the performance and/or the response of a system. Variability appears in real physical systems due to several factors, some of which we discuss in the following topics:

- the objective functions and design constraints often represent approximations of reality, by virtue of simplified mathematical models. Thus, not only the *true* optimal design of the system is unknown, but the estimated performance and response of the *found* optimal design may be inaccurate as well;
- the uncertainty associated with the manufacturing processes, and the prohibitive costs to eliminate them, make the exact reproduction of the *found* optimal solution not viable;
- optimization problems are static, but reality is dynamic. With time, the conditions under which a system operates may vary from those for which it was optimized.

We acknowledge that changes in the operating conditions of a system are likely to happen, during its lifecycle. Therefore, regarding the purpose of engineering design, we would like to pose the following question: *is it always desirable to find optimal solutions, in the design space, that solve a design optimization problem in a traditional sense, i.e., deterministically?*

It is common understanding among engineers that engineering design is not necessarily about creating the best design, but rather the one that suites the problem best. Design solutions that maintain constant levels of performance and response, under uncertainty, are called *robust*. Frequently, in engineering design, robustness is given preference in detriment of pure optimality. This is so, because, in a traditional sense, optimal designs are *minima* (or *minimal elements*) in the design space and are expected to be sensitive to small variations in the variables of a system. Although, in a sense, robustness may be perceived as contrary to optimization [6], both concepts can be combined in a design methodology called *robust design optimization* (RDO), whose optimal solutions satisfy both *optimality* and *robustness*. The concept of robustness is generally introduced into design optimization by manipulating the original problem at the level of the objective function(s) or design constraints. From this manipulation results a functional of the objective function(s) or design constraints, called the *robust counterpart function*, whose purpose is to steer the optimization process towards robust optimal solutions. There exist several formulations of robust design optimization. Beyer and Sendhoff [8] present a comprehensive survey.

To cope with the existence of random variables in the design optimization process, we define robustness as a measure based on the probability theory. More specifically, defined as a function of the second-order moments of the system response functionals. Please, consider the following definitions.

Definition 3.3 (*variance-covariance matrix*) [8]: Let $\mathbf{x} \in \Omega$ be a vector of random variables defined in a probability space (Ω, A, P) and $\phi_i : \Omega \to \mathbb{R}$ the i-th stochastic system response functional, integrable on Ω with respect to the probability measure $P(\mathbf{x})$. The *variance-covariance matrix*, $\mathbf{C_\phi}$, is the matrix whose components $\mathrm{Cov}(\phi_i, \phi_j)$ are given by:

$$\mathrm{Cov}(\phi_i, \phi_j) = \int_\Omega \big(\phi_i - E(\phi_i)\big)\Big(\phi_j - E\big(\phi_j\big)\Big)\, dP(\mathbf{x}) \tag{3.2}$$

where $E(\phi_i)$ is the expected value of ϕ_i.

Definition 3.4 (*global robustness measure*): On a probability space (Ω, A, P), the global robustness measure is defined as the *determinant of the variance-covariance matrix*, $\det \mathbf{C_\phi} : \Omega \to \mathbb{R}$.

With *"global robustness measure"* we mean one that quantifies the robustness of a design solution over the entire sample space Ω. Such condition is clearly satisfied in (3.2), where the integration is performed over Ω.

3.3 Reliability assessment

In engineering design, *reliability* is commonly interpreted as the concept that measures the *proximity to failure* of physical systems. Conceptually, reliability is closely related with the notion of *feasibility* in the design space, as *failure* is understood as the *state* of physical systems violating the *integrity constraints* imposed in the design problem. The notion of design integrity is not unique and has several interpretations, in engineering design. We restrain our future analysis to the topic of *structural integrity*. Reliability assessment introduces a specific nomenclature [9]:

- *Limit-state functions* are system response functionals associated with the integrity of physical systems;
- *Integrity constraints* are specific inequality constraints written in terms of the limit-state functions;
- Integrity constraints divide $D \times \Pi \times U$ in two disjoint regions, called the *safety space* (feasibility) and the *failure space* (infeasibility). Contrary to the usual convention in nonlinear programming, in reliability assessment, feasibility is associated with nonnegative values of the limit-state functions;
- The constraint surface separating the failure and the safety spaces is called the *failure surface* (or *limit-state surface*).

Commonly, optimal design solutions are expected to be close to, or over, the failure surface, since the imposed integrity constraints tend to be contrary to the design targets. For that reason, the existence of random uncertainty is identified as one of the main reasons for the unexpected failure of real physical systems that are theoretically feasible in terms of design integrity. Notice that, although design solutions may be in the safe space by a deterministic definition of integrity, random uncertainty may induce significant deviations in the predicted system response that violate the integrity constraints.

To cope with the existence of random uncertainty, we consider a probabilistic definition of reliability, providing a probabilistic measure of integrity associated with specific failure events. The introduction of reliability assessment in design optimization allows a systematic analysis of design integrity under uncertainty, identifying those designs that are probabilistically *infeasible*, but deterministically *feasible*. Such a design optimization methodology is called *reliability-based design optimization* (RBDO), whose optimal solutions satisfy both *optimality* and (*probabilistic*) *reliability*. The concept of reliability is generally introduced into design optimization by manipulating the original deterministic integrity constraints and transforming them into probabilistic constraints. Please, consider the following formal definitions.

Definition 3.5: Let $\mathbf{x} \in \Omega$ be a vector of random variables defined in a probability space $(\Omega, \mathcal{F}, P)$ and $\phi_i : \Omega \to \mathbb{R}$ a stochastic limit-state function. The sets $D_f = \{\mathbf{x} \in \Omega : \phi_i(\mathbf{x}) < 0\}$ and $D_s = \overline{D_f}$ are disjoint subsets of Ω, called the *failure space* and the *safety space*, respectively. The set $L = \{\mathbf{x} \in \Omega : \phi_i(\mathbf{x}) = 0\}$ is called the *failure surface*.

Definition 3.6: The probability space $(\Omega, \mathcal{F}, P)$, for which $\mathcal{F} = \{\emptyset, D_f, D_s, \Omega\}$ and $P : \mathcal{F} \to [0,1]$ is called the *uncertainty space*. Moreover, $P(\emptyset) = 0$, $P(\Omega) = 1$,

$$P(D_f) = \int_{D_f} dP(\mathbf{x}) \tag{3.3}$$

and $P(D_s) = 1 - P(D_f)$.

Definition 3.7 (*global reliability measure*): On the uncertainty space $(\Omega, \mathcal{F}, P)$, the global reliability measure is defined as the system's *probability of failure* $p_f = P(D_f)$.

With *"global reliability measure"* we mean one that quantifies the reliability of a design solution over the entire failure space. Such condition is clearly satisfied in (3.3).

3.4 Structural response

In this section, we define the system response functionals in terms of the *displacement response* and the *stress response* of composite laminate structural systems, as well as the respective deterministic response functionals.

3.4.1 Displacement response

The structural analysis of composite laminate structures is based on a displacement formulation of the Finite Element Method (FEM), in particular the shell finite element model developed by Ahmad *et al.* [10]. This shell element is obtained from a 3-dimensionmal finite element using a degenerative procedure. It is an isoparametric element with eight nodes and five degrees-of-freedom (DOF) per node, based on a Mindlin-type shell theory. The shell consists of a number of perfectly bonded plies and each individual ply is assumed homogeneous and anisotropic. Let $\mathbf{u}(\boldsymbol{\mu}_\mathbf{x}, \mathbf{d}, \boldsymbol{\pi}) \in U$ be the vector of displacements (state variables) and $\Psi: D \times \Pi \times U \to U$ a mapping representing the state of physical systems [11]. In this work, it is considered the linear equilibrium of structures under static loading conditions, such that:

$$\Psi(\boldsymbol{\mu}_\mathbf{x}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u}) = 0 \Leftrightarrow \mathbf{K}\mathbf{u} - \mathbf{f} = \mathbf{0} \quad , \forall \boldsymbol{\mu}_\mathbf{x} \in \Omega \tag{3.4}$$

where $\mathbf{K} \equiv \mathbf{K}(\boldsymbol{\mu}_\mathbf{x}, \mathbf{d}, \boldsymbol{\pi})$ is the stiffness matrix and $\mathbf{f}$ is the vector of external loads applied to the system. The vectors of design variables $\mathbf{d} \in D$ and system parameters $\boldsymbol{\pi} \in \Pi$ contain variables related with the material and geometric properties of the structural system and are specified later. The solution $\mathbf{u}^*$ of the equilibrium equation (3.4) is unique for each realization $(\boldsymbol{\mu}_\mathbf{x}^*, \mathbf{d}^*, \boldsymbol{\pi}^*)$ [11].

The applied geometric discretization of the physical system [10] implies the calculation of the displacements field in specific nodal points identified by the coordinates (e, k), where e is the element number and k is the nodal point number of the element e. Hereafter, we identify the components of the displacements vector as $u_{i_{(e,k)}}$, for $i = 1, \ldots, 5$, and define the associated response functional, as follows.

Definition 3.8 (*displacement response functional*): The deterministic system response functional $\varphi_1(\boldsymbol{\mu}_\mathbf{x}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u})$, associated with the *critical displacement* of the structural system, is defined by:

$$\varphi_1(\boldsymbol{\mu}_\mathbf{x}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u}) \equiv \bar{u}(\boldsymbol{\mu}_\mathbf{x}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u}) = \max u_{i_{(e,k)}} \tag{3.5}$$

3.4.2 Stress response

From the previous structural analysis, the prediction of the associated stress field (state) is trivial. The structural integrity analysis of composite laminate structures now introduced is based on the stress response of the structural system under consideration. The whole stress state is characterized by a scalar measure of structural integrity, associated with specific failure criteria. Contrary to isotropic materials, the failure modes of a composite laminate are not independent, for many possible interactions between them may occur. For that reason, we opt for an *interactive failure criterion*. Just like with the stress components, failure criteria for anisotropic composite materials are defined in the *fiber axis system*, being necessary to project them into the *laminate axis system*. These criteria are essentially empirical, yet consistent with the laws of mechanics, and rely on the *strength/stress ratio* $R \in \mathbb{R}^+$, playing an important role in measuring structural integrity. For linear-elastic materials, we write:

$$\boldsymbol{\sigma}^{\mathbf{max}} = R\boldsymbol{\sigma}^{\mathbf{apl}} \tag{3.6}$$

where $\boldsymbol{\sigma}^{\mathbf{max}}$ is the vector of the maximum allowable stresses and $\boldsymbol{\sigma}^{\mathbf{apl}}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u})$ is the vector of the applied stresses, such that $R \geq 1$ means structural safety and, for linear elastic structures, it is a multiplicative factor by which the applied stress components, $\sigma_i^{apl}(\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u})$, can be multiplied before failure occurs. On the other hand, $R < 1$ means failure but it lacks any physical meaning, because the maximum level of structural strength was exceeded and the system entered in a new regime other than the elastic one. Still, it provides the designer valuable information about how much the applied stress state should be reduced. Equation (3.6) reflects a failure criterion of independent failure modes called the *maximum stress criterion*. To account for the coupling between failure modes, the *Tsai-Wu quadratic failure criterion* is applied. On the stress space, the failure surface is defined by the following functional relation [12]:

$$\sum\nolimits_{i,j} \left[F_{ij}\sigma_i\sigma_j\right] + \sum\nolimits_i \left[F_i\sigma_i\right] = 1 \tag{3.7}$$

where σ_i represents the i-th component of the stress tensor and F_{ij} and F_i are the corresponding quadratic and linear strength parameters, defined on the *laminate axis system*. For thin orthotropic plies, under plane stress state, it is plausible to neglect stresses along the thickness coordinate. Thus, a reduced number of strength parameters is defined, in the *fiber axis system*, as [12]:

$$F_{XX} = \frac{1}{XX'} \quad F_{YY} = \frac{1}{YY'} \qquad F_{XY} = F_{XY}^{*}\sqrt{F_{XX}F_{YY}} \quad , -0.5 \leq F_{XY}^{*} \leq 0$$

$$F_{SS} = \frac{1}{S^2} \qquad F_X = \frac{1}{X} - \frac{1}{X'} \quad F_Y = \frac{1}{Y} - \frac{1}{Y'} \qquad F_S = 0 \tag{3.8}$$

where X and X' are the longitudinal tensile and compressive strengths, Y and Y' are the transverse tensile and compressive strengths and S is the longitudinal shear strength. To be consistent with (3.7), the strength parameters and the stress components need to be projected onto the *laminate axis system*, such that, under the assumption of plane stress state on the plies, only the principle stress and strength components $i = 1,2,6$ are considered [12]. The equality in (3.7) reflects the functional relation between the applied and the maximum allowable stress components, describing the boundary between the safety and the failure spaces. Hence, introducing (3.6) into (3.7) leads to:

$$\left[F_{ij}\sigma_i^{apl}\sigma_j^{apl}\right]R^2 + \left[F_i\sigma_i^{apl}\right]R = 1 \quad i,j = 1,2,6 \tag{3.9}$$

which solves to:

$$R = \left(-\frac{b}{2a} + \sqrt{\left(\frac{b}{2a}\right)^2 + \frac{1}{a}} \right) \tag{3.10}$$

where $a = F_{ij}\sigma_i^{apl}\sigma_j^{apl}$ and $b = F_i\sigma_i^{apl}$, for $i, j = 1,2,6$. The parameter R, as in (3.10), is known as the *Tsai number* [12]. On the stress space, for constant values of longitudinal shear stress, the failure surface associated with the Tsai-Wu criterion assumes the closed shape of an ellipse [12], along which R is equal to one.

The applied geometric discretization of the physical system [10] implies calculating the stress field and the Tsai number in specific integration points identified by the coordinates (e, p, k), as defined previously. For a first-ply failure philosophy, the first ply to fail is the one possessing the point with the critical Tsai number, representing the point of the structure with the most aggravated stress state (i.e. the one closer to failure). Thus, to each vector of applied stresses $\boldsymbol{\sigma}^{\mathbf{apl}}_{(e,p,k)}$ there is a $R_{(e,p,k)}$ and we define the associated response functional as follows.

Definition 3.9 (*stress response functional*): The deterministic system response functional $\varphi_2(\boldsymbol{\mu}_\mathbf{x}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u})$, associated with the *critical Tsai number* of the structural system, is defined by:

$$\varphi_2(\boldsymbol{\mu}_\mathbf{x}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u}) \equiv \bar{R}(\boldsymbol{\mu}_\mathbf{x}, \mathbf{d}, \boldsymbol{\pi}, \mathbf{u}) = \min R_{(e,p,k)} \tag{3.11}$$

3.5 The reliability-based robust design optimization problem

Taking into consideration the concepts presented in the previous sections, we now introduce a stochastic design optimization methodology combining the RDO and the RBDO frameworks, called *reliability-based robust design optimization* (RBRDO).

In the following definition, our concern is to show only the variables that intervene in the RBRDO methodology. Therefore, the dependency on the state variables $\mathbf{u} \in U$ is omitted, for the sake of simplicity, although it always exists. Additionally, we intend to highlight the explicit and implicit dependencies of each functional in the problem. For that, we introduce the following notation: *a functional $f(a, b|c)$ is said explicitly dependent on (a, b) and implicitly dependent on c. Moreover, $f(a, b, c)$ implies explicit dependence on (a, b, c), while f implies implicit dependence.*

The proposed RBRDO problem is defined as the bi-objective minimization of the structural weight W (*performance optimality*) and the determinant of the variance-covariance matrix $\det \mathbf{C}_\boldsymbol{\Phi}$ (*robustness optimality*). The design process is executed over the mean-values $\boldsymbol{\mu}_\mathbf{x} \in D$ of the *random design variables* $\mathbf{x} \in \Omega_D$. Additionally, a set of *random system parameters* $\boldsymbol{\pi} \in \Omega_\Pi$ is considered, with constant mean-values $\boldsymbol{\mu}_\boldsymbol{\pi} \in \Pi$ as an input of the problem. The inclusion of deterministic variables in the design methodology is also immediate. In agreement with Definition 3.1, a deterministic variable is one with null random deviation. The structural response is measured by two system response functionals associated with the critical displacement $\bar{u}$ and the critical Tsai number $\bar{R}$. Here, robustness is defined as a measure of the variability associated with the system response functionals, due to uncertainty, and includes the correlation effects between $\bar{u}$ and $\bar{R}$. The design space is also limited by the imposition of upper bounds to both the displacement and stress responses of the structural system. To accommodate for the different nature of displacements and stresses, the design constraints are written in the shape of normalized stochastic response functionals. The displacement constraint is evaluated deterministically. Probabilistic reliability assessment is

introduced in the design optimization upon the stress limit-state function. The introduction of a stress reliability constraint attempts to guide the optimization process in way that a minimum probabilistic level of structural integrity is imposed to the Pareto optimal design solutions. In general, it is also considered that, in robustness and reliability assessment, uncertainty is propagated into different groups of variables, such that the robustness and the reliability measures are defined with respect to different probability spaces.

Definition 3.10 (global *RBRDO problem*): Let (Ω_1, A, P_1) be a probability space with $\Omega_1 \subseteq \Omega_D \times \Omega_\Pi$, $\det \mathbf{C}_{\boldsymbol{\phi}} : \Omega_1 \to \mathbb{R}$ a global robustness measure and $\boldsymbol{\phi}(\mathbf{x}, \boldsymbol{\pi}) = (\bar{u}, \bar{R})$ the associated vector of stochastic response functionals. Also, let $(\Omega_2, \mathcal{F}, P_2)$ be an uncertainty space with $\Omega_2 \subseteq \Omega_D \times \Omega_\Pi$, $p_f \in [0,1]$ a global reliability measure and $\mathbf{g}(\mathbf{x}, \boldsymbol{\pi})$ a vector of stochastic limit-state functions, with components:

$$g_1(\mathbf{x}, \boldsymbol{\pi}) = \frac{\bar{u}(\mathbf{x}, \boldsymbol{\pi})}{u^a} - 1 \tag{3.12}$$

and

$$g_2(\mathbf{x}, \boldsymbol{\pi}) = \frac{\bar{R}(\mathbf{x}, \boldsymbol{\pi})}{R^a} - 1 \tag{3.13}$$

where u^a and R^a are the allowable values of $\bar{u}$ and $\bar{R}$. The RBRDO problem is then stated as follows:

$$\begin{aligned} \min_{\boldsymbol{\mu}_\mathbf{x}} \quad & \mathbf{f}(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_{\boldsymbol{\pi}}) = \left(W(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_{\boldsymbol{\pi}}),\ \det \mathbf{C}_{\boldsymbol{\phi}}\right) \\ \text{subject to:} \quad & g_1(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_{\boldsymbol{\pi}}) \leq 0 \\ & p_f \equiv P_2(g_2(\mathbf{x}, \boldsymbol{\pi}) < 0) \leq p_f^a \\ & \boldsymbol{\mu}_\mathbf{x} \in D \end{aligned} \tag{3.14}$$

where p_f^a is the allowable value of p_f.

This definition of robustness, imposed in the design problem as an objective measuring the variability of the structural responses associated with the design constraints, is hereby called *feasibility robustness*. The objectives W and $\det \mathbf{C}_{\boldsymbol{\phi}}$ are mostly related through the stiffness of the laminates, which is dependent on the material properties, geometric properties, among others. In this thesis, the conjugation of feasibility robustness and probabilistic reliability assessment in the design optimization of composite laminate structures is studied.

The evaluation of both the robustness and the reliability measures associated with complex structural system, like composite laminate structures, is performed by numerical methods yielding approximate values of the components of the variance-covariance matrix and of the probability of failure of each design solution. Hence, design optimization problem becomes one with three computing cycles: the main cycle regarding the optimization process itself and two others regarding the numerical evaluation of the robustness and the reliability measures. The numerical solution of the integrals in (3.2) and (3.3) requires a number of extra numerical simulations of the structural model and can be a burden in terms of computational cost. Overall, a balance between numerical accuracy and the computing time needed to perform both the robustness and the reliability assessment is desired.

Throughout the following chapters, we approach these difficulties and propose new algorithmic developments to solve the RBRDO problem that are solely based on evolutionary algorithms (EA). Our choice for EAs is justified by their ability to achieve global convergence, as discussed in Chapter 2, while other optimization algorithms only guarantee local convergence at best.

References

[1]. Papalambros PY, Wilde DJ. Principles of optimal design: modelling and computing. 2nd ed. Cambridge University Press; 2000.

[2]. Aven T. On Different Types of Uncertainties in the Context of the Precautionary Principle. Risk Analysis 2011; 31(10); 1515-25.

[3]. Nikolaidis E. Types of Uncertainty in Design Decision Making. In: Nikolaidis E, Ghiocel DM, Singhal S, editors. Engineering Design Reliability Handbook. Taylor & Francis; 2005.

[4]. Cacuci DG. Sensitivity and uncertainty Analysis: theory. 1st ed. Chapman and Hall/CRC; 2003.

[5]. Shapiro A, Dentcheva D, Ruszczynki A. Lectures on stochastic programming: modelling and theory. Society for Industrial and Applied Mathematics, Mathematical Programming Society.; 2009.

[6]. Marczyk J. Stochastic multidisciplinary improvement: beyond optimization. American Institute of Aeronautics and Astronautics. AIAA.

[7]. Beyer H-G, Sendhoff B. Robust Optimization – A Comprehensive Survey. Comput Method Appl M 2007; 196(33-34): 3190-218.

[8]. Huang B and Du X. Analytical robustnss assessment for robust design. Struct Multidisc Optim 2007; 34: 123-37.

[9]. Melchers RE. Structural Reliability Analysis and Prediction. 2nd ed. Willey & Sons; 1999.

[10]. Ahmad S, Irons S, Zienkiewicz OC. Analysis of thick and thin shell structures by curved finite elements. Int J Numer Methods Eng 1970;2(3):419-51.

[11]. Plessix, RE. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophys J Int 2006; 167: 495-503.

[12]. Tsai SW, Wu EM. A General Theory of Strength for Anisotropic Materials. J Compos Mater 1971;5(1):58-80.

IV
EVOLUTIONARY RELIABILITY ASSESSMENT

Summary

*

The previous chapter gave an overview of the basic principles of stochastic design optimization and formulated a particular instantiation of the reliability-based robust design optimization problem. This chapter presents in detail the numerical computation of the reliability constraint and propose novel methodologies for reliability assessment, based on evolutionary computation. Sect. 4.1 presents a concise theoretical background on the numerical methods most commonly applied to solve the probability of failure integral: Monte Carlo Simulation, as a global reliability method, and the Reliability Index Approach (RIA) and the Performance Measure Approach (PMA), as local reliability methods. The preference for local reliability methods to assess the probabilistic reliability constraint in structural design optimization is explained. In Sect. 4.2, two novel evolutionary algorithms (EA) are specifically developed as numerical tools to solve both the RIA and the PMA problems. Such algorithms constitute a novelty in reliability assessment, since it is almost exclusively performed by gradient-based algorithms and sampling methods, in the literature. The preference for EAs over other optimization algorithms is discussed. The developed EAs are defined and discussed in depth and their ability to converge with probability 1 is demonstrated.

4.1 Approximation models for structural reliability

We begin our discussion by defining explicitly the probability of failure integral, in (3.3). Let $(\Omega_2, \mathcal{F}, P_2)$ be an uncertainty space and $D_f = \{\mathbf{x}, \boldsymbol{\pi} \in \Omega_2 : g_2(\mathbf{x}, \boldsymbol{\pi}) < 0\} \in \mathcal{F}$. Then, it follows [1]:

$$p_f = \Pr(g_2(\mathbf{x}, \boldsymbol{\pi}) < 0) = \int_{g_2(\mathbf{x},\boldsymbol{\pi})<0} p_{\mathbf{x},\boldsymbol{\pi}}(\mathbf{x}, \boldsymbol{\pi})\, d\mathbf{x} d\boldsymbol{\pi} \tag{4.1}$$

where $p_{\mathbf{x},\boldsymbol{\pi}}(\mathbf{x}, \boldsymbol{\pi})$ is the joint probability density function (PDF) of the random variables $(\mathbf{x}, \boldsymbol{\pi})$.

The analytical solution of the integral in (4.1) can only be attempted when both the PDF and the domain of integration are known and explicitly defined. For most structural design problems, physical systems are usually numerically simulated. Therefore, limit-state functions are only known pointwise and become implicitly defined. In addition, knowledge about the distributions and correlations of the random variables may be incomplete. Numerical methods are the most common techniques to perform structural reliability assessment [2]. These are usually divided into global and local reliability methods. *Global reliability methods* analyze the whole sample space Ω_2, or a relevant part of it, in order to provide an unbiased estimator of the probability of failure. Monte Carlo Simulation is the most straightforward approach and sets the basis for other sampling methods in reliability assessment. On the other hand, *local reliability methods* focus their action on the region of the sample space in the vicinity of the mean-values of the random variables. These are based on the concept of the *reliability index* and obviate the integration process by transforming the integrand into a multi-normal joint PDF, for which special results are available.

4.1.2 Monte Carlo Simulation

We call Monte Carlo Simulation (MCS) to a broad class of stochastic methods for the numerical computation of definite integrals, based on random sampling. In reliability assessment, it is applied for the direct computation of the probability of failure, as defined in (4.1) [1]. The basic principle of MCS methods is the redefinition of the probability of failure integral, as follows [1,2]:

$$p_f = \int_{\Omega_2} \mathbb{I}\big(g_2(\mathbf{x}, \boldsymbol{\pi})\big) p_{\mathbf{x},\boldsymbol{\pi}}(\mathbf{x}, \boldsymbol{\pi})\, d\mathbf{x} d\boldsymbol{\pi} \tag{4.2}$$

where $\mathbb{I}(\cdot)$ is an indicator function defined as:

$$\mathbb{I}\big(g_2(\mathbf{x}, \boldsymbol{\pi})\big) = \begin{cases} 1 & , \text{if } g_2(\mathbf{x}, \boldsymbol{\pi}) < 0 \\ 0 & , \text{if } g_2(\mathbf{x}, \boldsymbol{\pi}) \geq 0 \end{cases} \tag{4.3}$$

By using the above indicator function, the integration domain is changed from the failure domain D_f to the entire sample space Ω_2 and so, from (4.2), we see that the probability of failure is thus defined as the expectation of $\mathbb{I}\big(g_2(\mathbf{x}, \boldsymbol{\pi})\big)$, that is:

$$p_f = \mathrm{E}\left(\mathbb{I}\big(g_2(\mathbf{x}, \boldsymbol{\pi})\big)\right) \tag{4.4}$$

Consequently, from sample statistics, and by interpreting the calculation of cumulative probabilities as a traditional binomial experiment on a finite subset $\widehat{\Omega}_2 \subset \Omega_2$ of K random realizations of $(\mathbf{x}, \boldsymbol{\pi})$, following $p_{\mathbf{x},\boldsymbol{\pi}}(\mathbf{x}, \boldsymbol{\pi})$ and originating K_f failure events, the measure:

$$\hat{p}_f = \frac{1}{K}\sum_{i=1}^{K} \mathbb{I}\left(g_2(\mathbf{x_i}, \boldsymbol{\pi_i})\right) = \frac{K_f}{K} \tag{4.5}$$

is an unbiased estimator of the *unknown* true value of p_f with any desired accuracy [3], such that $\lim_{K\to\infty} \hat{p}_f = p_f$.

The formulation in (4.5) represents the simplest instantiation of MCS, known as *crude* MCS. The method is simple to implement and can be applied to all class of reliability problems [3]. However, its performance is dependent on the actual true value of p_f, which is unknown, and on the complexity of the limit-state function to be evaluated. Commonly, in structural design problems, the imposed probabilities of failures are extremely small, requiring huge samples for the estimate $\hat{p}_f$ to converge, and the limit-state functions result from expensive implicit models of structural systems, such as in finite element analysis, leading to considerable computational problems [2]. To cope with these limitations, variance reduction techniques, such as *importance sampling*, *directional simulation* and *subset sampling*, have been developed [1,2]. References [4-9] provide a collection of relevant and some recent advances in the application of MCS in structural reliability.

4.1.3 Local reliability methods

The hindrance of global reliability methods is their global sense (i.e. the domain of integration is the entire Ω_2). In the context of structural design optimization, the repeated application of global reliability methods to calculate extremely low probabilities of failure of innumerous design solutions often makes the design optimization process not feasible. On the other hand, local reliability methods focus their action in the region around the mean-values of the random variables. These are based on the concept of the *reliability index* and simplify the integration process by transforming the integrand into a multi-normal joint PDF.

On this section, we introduce two alternative local reliability methods, called the *reliability index approach* and the *performance measure approach*. Although lacking the accuracy of MCS methods, these do not require the exhaustive evaluation of the uncertainty space and are not as expensive, allowing the recurrent evaluation of the probabilistic reliability constraint in structural design optimization problems.

4.1.3.1 The reliability index approach

The *reliability index approach* (RIA) is a local reliability assessment methodology that returns a quantitative measure of reliability. The notion of a reliability index was first proposed by Cornell [10], as follows. Let μ_{g_2} and σ_{g_2} be the first two moments of $g_2(\mathbf{x}, \boldsymbol{\pi})$ and assume $g_2 \sim N\left(\mu_{g_2}, \sigma_{g_2}\right)$. Accordingly, the probability of failure is given by:

$$p_f = \Pr(g_2 < 0) = \Phi\left(\frac{0 - \mu_{g_2}}{\sigma_{g_2}}\right) = \Phi(-\beta_c) \tag{4.6}$$

here $\Phi(\cdot)$ is the standard normal distribution function. In this case, the analytical integration of (4.1) is possible and one-dimensional, since both the integration limits and the distribution of the random variable g_2 are known. The measure

$$\beta_c = \mu_{g_2}/\sigma_{g_2} \tag{4.7}$$

is named the *Cornell reliability index* and represents the distance, in standard-deviation units, from μ_{g_2} to $g_2 = 0$, as show in Figure 4.1. It is also *invariant*, by definition, only requiring the knowledge of the moments of the random normal variable g_2.

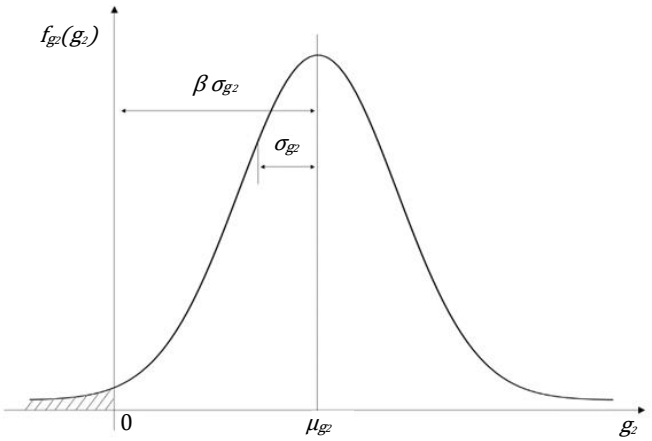

Figure 4.1: One-dimensional representation of the Cornell reliability index, associated to a normally distributed limit-state function.

The previous idea is readily extended to the case where g_2 is a *multilinear* function of $(\mathbf{x}, \boldsymbol{\pi})$, for which the first two moments can be exactly calculated. Thus, (4.6) is exact if $g_2(\mathbf{x}, \boldsymbol{\pi})$ is the result of the addition, or product, of independent normal $(\mathbf{x}, \boldsymbol{\pi})$ [1] and preserves invariance [11]. In general, $g_2(\mathbf{x}, \boldsymbol{\pi})$ is nonlinear and the failure surface is a level set given by a nonlinear equation $g_2(\mathbf{x}, \boldsymbol{\pi}) = 0$ in the uncertainty space. Hence, the first two moments of g_2 are not readily calculated. A suitable approach is the so called *first-order second moment* (FOSM) theory that estimates the moments of the linearized limit-state function, approximated by a first-order Taylor polynomial about a point $(\hat{\mathbf{x}}, \hat{\boldsymbol{\pi}})$ [1]. Invariance however is lost, since β_c becomes explicitly dependent on g_2 and on the expansion point $(\hat{\mathbf{x}}, \hat{\boldsymbol{\pi}})$ [11]. Historically, Cornell [10], Lind [12] and Ditlevsen [13] proposed preliminary approaches to determine a reliability measure that are not invariant, under nonlinear limit-state functions.

Hasofer and Lind [14] proposed a generalization of the FOSM theory that preserves invariance under nonlinear limit-state functions, known as the *first-order reliability method* (FORM). The authors propose to calculate the Cornell reliability index at the point of the failure surface with the greatest probability density, around which failure is expected to occur. In the space of *independent standard normal random variables*, $\mathbf{y} \sim N(\mathbf{0}, \mathbf{1})$, the identification of such point is strictly a minimization problem, given by:

$$\begin{aligned} \min_{\mathbf{y}} \quad & \|\mathbf{y}\| \\ \text{subject to:} \quad & G_2(\mathbf{y}) = 0 \end{aligned} \tag{4.8}$$

where $\mathbf{y} = T(\mathbf{x}, \boldsymbol{\pi})$ and $G_2(\mathbf{y}) = g_2(\mathbf{x}, \boldsymbol{\pi})$ are the transformed vector of random variables and limit-state function, respectively.

The standardization of the uncertainty space is a fundamental step of the underlying theory, allowing to characterize uncertainty by a rotationally symmetric multivariate standard normal distribution, for which useful properties are known. In case $(\mathbf{x}, \boldsymbol{\pi})$ are independent normal random variables, the transformation $\mathbf{y} = T(\mathbf{x}, \boldsymbol{\pi})$ is linear and exact [14]. For non-normal random variables, additional distributional information is required. A detailed discussion on the subject can be found in [1,15].

The solution to (4.8) is called the *most probable failure point*, $\mathbf{y_{MPP}}$, and $\|\mathbf{y_{MPP}}\|$ represents the shortest distance, in standard-deviation units, from the origin of the standardized uncertainty space to the failure surface. A fundamental result follows.

Proposition 4.1 [1]: $\beta_c(\mathbf{y_{MPP}}) = \|\mathbf{y_{MPP}}\|$.

Proof: First, from the application of the KKT conditions to (4.8), it results [16]:

$$\|\mathbf{y_{MPP}}\| = -\frac{\mathbf{y_{MPP}} \cdot \nabla G_2}{\|\nabla G_2\|} \tag{4.9}$$

where ∇G_2 is the gradient-vector of $G_2(\mathbf{y})$, at $\mathbf{y_{MPP}}$. Now, let $G_2(\mathbf{y})$ be approximated, at $\mathbf{y_{MPP}}$, by a first-order Taylor polynomial:

$$G_{2_L}(\mathbf{y}) = G_2(\mathbf{y_{MPP}}) + \sum_{i=1}^{N^y} (y_i - y_{MPP_i}) \left(\frac{\partial G_2}{\partial y_i}\right)_{\mathbf{y_{MPP}}} \tag{4.10}$$

Since by definition $G_2(\mathbf{y_{MPP}}) = 0$, $\mu_{y_i} = 0$ and $\sigma_{y_i} = 1$, it follows:

$$\mathrm{E}(G_{2_L}) = \int_{\Omega_\mathbf{y}} G_{2_L}(\mathbf{y}) \phi_\mathbf{y}(\mathbf{y}) d\mathbf{y} = -\mathbf{y_{MPP}} \cdot \nabla G_2 \tag{4.11}$$

and

$$\mathrm{Var}(G_{2_L}) = \int_{\Omega_\mathbf{y}} \left[G_{2_L}(\mathbf{y}) - \mathrm{E}(G_{2_L})\right]^2 \phi_\mathbf{y}(\mathbf{y}) d\mathbf{y} = \nabla G_2 \cdot \nabla G_2 \tag{4.12}$$

where $\phi_\mathbf{y}(\mathbf{y})$ is the multivariate standard normal PDF. Thus, by (4.7), we obtain:

$$\beta_c(\mathbf{y_{MPP}}) = -\frac{\mathbf{y_{MPP}}^T \nabla G_2}{\|\nabla G_2\|} \tag{4.13}$$

■

From (4.8) and (4.13), we conclude that $\|\mathbf{y_{MPP}}\|$ is an invariant reliability measure, only requiring knowledge about the failure surface, since ∇G_2, calculated at $\mathbf{y_{MPP}}$, is completely defined by the hyperplane tangent to the failure surface. Furthermore, $\|\mathbf{y_{MPP}}\|$ is unique in the standardized uncertainty space, given the optimality of $\mathbf{y_{MPP}}$ [11]. Hence, it is a design solution's property and we write:

$$\beta_{HL}(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi}) = \|\mathbf{y_{MPP}}\| \tag{4.14}$$

called the *Hasofer-Lind reliability index.* In practice, reliability assessment is reduced to find the point $\mathbf{y_{MPP}}$ and its norm $\|\mathbf{y_{MPP}}\|$, in the standardized uncertainty space. It then results that, by approximating the failure surface as a hyperplane, at $\mathbf{y_{MPP}}$, the probability of failure integral has the following approximate solution:

$$\hat{p}_f = \int_{G_{2_L}(\mathbf{y})<0} \phi_\mathbf{y}(\mathbf{y}) d\mathbf{y} = \Phi(-\beta_{HL}) \tag{4.15}$$

Notice that the multivariate integration over the failure domain is reduced to the one-dimensional integration along the direction of the vector $\mathbf{y_{MPP}}$, which is only possible due to the linearization of the failure surface and to the rotational symmetry of $\phi_\mathbf{y}(\mathbf{y})$, as shown in Figure 4.2.

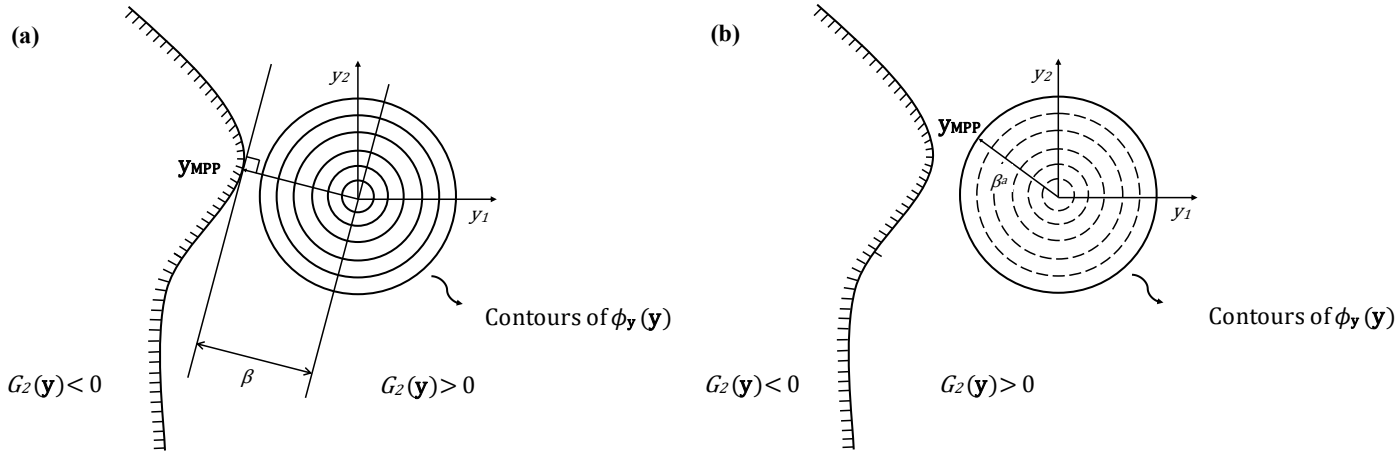

Figure 4.2: Two-dimensional representation of the standardized uncertainty space, with a representation of the MPP, for (a) the RIA and (b) the PMA.

4.1.3.2 The performance measure approach

The previous approach estimates the probability of failure directly from the reliability index of design solutions. Assuming the linearization of the failure surface, at $\mathbf{y_{MPP}}$, equation (4.15) approximates (4.1) as a function of the minimum Euclidean distance from the origin of the standardized uncertainty space to the failure surface. From the invariant RIA formulation in (4.8), one can define an alternative local reliability assessment method, known as the *performance measure approach* (PMA). Contrary to the RIA, the PMA does not return a quantitative reliability measure, but rather a qualitative one, determining if design solutions effectively satisfy the imposed level of reliability.

The reliability assessment problem is now formulated as the inverse problem of the RIA, in (4.8). Specifically, it is the problem of finding the point with the smallest limit-state value, on a predetermined probability density level set. In the standardized uncertainty space, the identification of such point is given by the following inverse minimization problem [17]:

$$\begin{aligned} \min_{\mathbf{y}} \quad & G_2(\mathbf{y}) \\ \text{subject to:} \quad & \|\mathbf{y}\| = \beta^a \end{aligned} \tag{4.16}$$

where β^a is the radius, in standard deviation units, of the imposed probability density hypersphere. The standardization of the uncertainty space allows therefore to explore the known properties of the rotationally symmetric multivariate standard normal distribution, as before. Hence, by the linearization of $G_2(\mathbf{y})$, at each $\mathbf{y}$, it follows from (4.15) that $\beta^a = \Phi^{-1}(p_f^a)$. In such case, the problem is known as the *inverse FORM* (iFORM) and its solution $\mathbf{y_{MPP}}$ represents the point on the probability density hypersphere, centered at the origin of the uncertainty space, that is closer to failure in terms of the limit-state value.

4.2 Evolutionary models for structural reliability

The application of the previous reliability assessment methods to complex structural systems, defined by expensive implicit numerical models, requires the evaluation of finite sets of solutions in the uncertainty space, following well-structured algorithms. MCS and related methods rely on sampling heuristics and can achieve great accuracy. However, due to the low probabilities of failure imposed in structural design optimization, sampling methods become impractical to be applied sequentially, mainly in high dimensional problems [2]. The RIA and the PMA are usually preferred as means to assess reliability in structural design optimization and are historically solved by gradient-based algorithms, which are characterized by algorithmic simplicity and convergence speed. However, these cannot guarantee global convergence, often anchoring in local optima and underestimating structural reliability, mainly in high dimensional problems. Numerical divergence is also possible in the presence of discontinuous implicit response functionals, and/or the respective derivatives, and highly concave failure surfaces [18,19] Contrary to gradient-based algorithms, evolutionary algorithms (EA) do not require gradient information, convexity and continuity of the search space. Moreover, global convergence is proven to be achieved, provided the algorithms satisfy certain conditions (see Chapter 2). Yet, convergence is slower, because a series of stochastic operators is applied to a set of solutions, at every iteration, rather than just one.

A common struggle to all these numerical methods is the difficulty to deal with highly dimensional search spaces. In this matter, EAs appear to offer the best compromise between numerical accuracy and efficiency. With that intent, we begin by presenting the *basic evolutionary operators* (BEOs), which constitute the core of the EAs developed to solve the local reliability assessment problems.

4.2.1 Basic evolutionary operators

The demanding nature of the local reliability methods asks for a robust EA, capable of solving the reliability assessment problem of innumerous design solutions. On its application to design optimization, reliability assessment must be both accurate and efficient, in a way that the design optimization problem not only is correctly solved but is viable, as well.

4.2.1.1 Algorithmic description

A set of diverse operators is now introduced, combining high selective pressure (supported by a strong elitist strategy) with a strong exploratory capability (imposed by highly disruptive operators). The benefits of such a polarized approach are discussed by Eshelman [20]. The goal is to build EAs capable of achieving good accuracy, while working with a very reduced-size populations and genotypes, on a large range of problems with the same parameter settings. We start by defining the basic list structures of the BEOs.

Definition 4.1: Let $\mathbb{S}$ be a finite genotype space and $\mathbb{S}^*$ be the set of all finite lists over $\mathbb{S}$. A list $\mathbf{P}^t \in \mathbb{S}^*$, containing $n_{\mathbf{P}}$ genotypes, defines the *population*, at generation t. Similarly, the list $\mathbf{E}^t \subset \mathbf{P}^t$, containing the n_{top} genotypes of $\mathbf{P}^t$ with the highest fitness, defines the *elite group* of the population and $\mathbf{B}^t$ defines the list of all offspring solutions.

Elitist parent selection

The parent selection mechanism is responsible for the stochastic selection of $n_{\mathbf{B}}$ parent-lists that compose the mating pool, at each generation. Each parent-list contains two individuals chosen at random by two independent fitness-proportional selection processes: one parent selected from the elite group $\mathbf{E}^t$ and another from the remaining population $\mathbf{P}^t \backslash \{\mathbf{E}^t\}$.

Such a bipartite selection procedure implies that, for each pair of parents, one (and only one) parent belongs necessarily to the elite group, hence being deemed *elitist*. With it, it is expected to achieve two simultaneous goals: first, an increased selective pressure (bias) towards the best fitted individuals of the population; second, an increased reproduction probability of the weakest individuals of the population.

Uniform crossover

For each parent-list, one (and only one) offspring solution is constructed and stored in the offspring population $\mathbf{B}^t$, hence generating a total of $n_{\mathbf{B}}$ solutions. In the presence of (strong) elitism, we look for an operator that provides productive genetic recombinations and not simply preserves the genetic features of the parent solutions [20], which is already guaranteed by elitism. Therefore, the offspring genetic material is obtained by the application of the *uniform crossover* operator: each gene being determined in a biased way, according to the crossover ratio r_{uc} that represents the probability of selecting genes from the elite parent (see Chapter 2).

The interesting property of uniform crossover is that a genetic feature is not heritable unless it is common to both parents [21]. As such, uniform crossover is a highly disruptive recombination operator that only preserves genetic similarity and has the ability to create new linkages among the genes of the offspring solutions, induced by the conflicting genes of the respective parents, acting as a macro-mutation operator [22]. Indeed, the greater the genetic difference between two parents the more explorative uniform crossover becomes.

Elitist survivor selection

At an intermediate state, after recombination, the entire population is extended to the list $\mathbf{P}^t \cup \mathbf{B}^t$. The survivor selection mechanism is responsible for the composition of the population, in the next generation, reducing the number of individuals to the initial population size $n_{\mathbf{P}}$. Survivor selection is performed in *three* stages by *three* elitist replacement mechanisms, each with a specific task.

Similarity control

The growth of instances of good genetic features among the individuals of the population is of little value if it leads to premature convergence. Since superior individuals are expected to produce more offspring, it is likely that, eventually, an individual will be recombined with one of its near relatives. As this leads to crossing over individuals sharing a considerable number of genetic features, exploration via genetic recombination may quickly degenerate [20]. Although uniform crossover is able to slow down such phenomenon, sometimes individuals that have few differences among them are paired for recombination, particularly in advanced stages of the evolutionary process. If the offspring survives together with one or both parents, it will be likely that such event will occur more often in future generations.

We introduce a mechanism of genetic similarity control, to promote diversity among the members of $\mathbf{P}^t \cup \mathbf{B}^t$. The *similarity control operator* is presented as an *elitist* and *highly disruptive* replacement mechanism. Similarity is evaluated variable-by-variable. After fitness-based ranking of $\mathbf{P}^t \cup \mathbf{B}^t$, each individual is compared against all the weaker individuals, one at a time. For each pair of similar individuals, the weakest one is removed from the population. At the end of the comparison process, the size of $\mathbf{P}^t \cup \mathbf{B}^t$ is recovered with randomly generated individuals. Hence, the process is elitist (always keeps the best individuals) and highly disruptive (it does not preserve the genetic features of individuals with high levels of similarity).

Elitist $(n_{\mathbf{P}} + n_{\mathbf{B}})$ replacement mechanism

In order to achieve a monotone evolution, it is adopted an elitist $(n_{\mathbf{P}} + n_{\mathbf{B}})$ replacement mechanism, where the newly created offspring and the existent population members undergo a cross-generational competition: the extended population $\mathbf{P}^t \cup \mathbf{B}^t$ is ranked by fitness, after which only the top $n_{\mathbf{P}}$ individuals are selected to survive.

Unlike the elitist parent selection, characterized by two independent moments of stochastic selection, the adopted elitist replacement scheme imposes a deterministic selection rule, steered by fitness. From the joint application of elitism at both the parent and survivor selection, not often seen in the literature [23,24], results a strong selective pressure biased, at first (parent selection), in favor of the best performing individuals in $\mathbf{P}^t$ and, later on (survivor selection), against the worst performing individuals in $\mathbf{P}^t \cup \mathbf{B}^t$.

The interesting property of the $(n_{\mathbf{P}} + n_{\mathbf{B}})$ replacement scheme is its robustness against the quality loss of the solutions generated by the recombination operator, in the sense that both top and average individuals are allowed to survive longer, if not outperformed by the offspring. Therefore, it does not require a crossover operator with a low disruption probability. Indeed, it allows genetic recombination to be used as a controlled exploration mechanism that balances the strong selective pressure induced by elitism [20].

Implicit mutation

The combination of uniform crossover and similarity control may have a positive effect delaying premature convergence. But these mechanisms cannot guarantee desirable levels of genetic diversity throughout the entire evolution. This is because the main task of similarity control is to avoid the existence of (quasi-) clone solutions in the population and not to generate raw diversity. The latter is instead a consequence. Eventually, it is expected that the evolutionary process manages to produce solutions that are as similar as possible. In such case, the similarity control operator loses importance as it is no longer called to inject raw diversity in the population enough times. A productive mutation operator is needed.

In the presence of highly disruptive operators, such as uniform crossover and the similarity control operator, the traditional mutation (see Chapter 2), where a reduced number of genes are changed, is not expected to produce significant effects, because the amount of genetic diversity it can generate is potentially inferior to that of the former two operators [20].

In order to prevent the search from stagnating, we introduce an operator called *implicit mutation* [23], in the shape of an *elitist* and *highly disruptive* replacement mechanism. At each generation, after the $(n_{\mathbf{P}} + n_{\mathbf{B}})$ replacement mechanism, the resulting population is ranked according to individual fitness. Then, the worst n_{bot} individuals are eliminated and the same number of solutions is randomly generated. The operator is elitist (always keeps the best individuals) and highly disruptive (it does not preserve the genetic features of the worst individuals).

Contrary to the similarity control, the main goal of implicit mutation is to inject raw diversity into the population, affecting the evolution in a medium to long-term perspective, regarding the productive exploration of the search space via genetic recombination [24,25]. Indeed, its combination with the elitist parent selection of the next generation promotes the recombination between the elite and the newly generated individuals. Also, the joint application of similarity control, the $(n_{\mathbf{P}} + n_{\mathbf{B}})$ replacement mechanism and implicit mutation only allows average to top individuals, with enough genetic diversity, to survive at each generation.

4.2.1.2 Limit-behavior of the BEOs

The algorithms proposed in the next sections, to solve the inner cycle of local reliability assessment (RIA and PMA), are based on the set of BEOs proposed above. Therefore, we start this short convergence analysis by formalizing the algorithmic sequence constructed from the BEOs, as follows in Algorithm 4.1.

Algorithm 4.1: (*basic evolutionary operators*):

repeat
for $\mathbf{P}^t$ **do**
  Fitness-based ranking
  Elitist parent selection of $n_{\mathbf{B}}$ parent-lists
    one parent from $\mathbf{E}^t$ *and another from* $\mathbf{P}^t \backslash \{\mathbf{E}^t\}$
  Uniform crossover
    generate $n_{\boldsymbol{B}}$ *offspring solutions, one per parent-list*
  Allocate offspring solutions into $\mathbf{B}^t$
end do
for $\mathbf{P}^t \cup \mathbf{B}^t$ **do**
  Fitness-based ranking
  Similarity control
    for each pair of similar individuals, replace the worst by a random one
  Fitness-based ranking
  $(n_{\boldsymbol{P}} + n_{\boldsymbol{B}})$ replacement mechanism
    eliminate the worst $n_{\boldsymbol{B}}$ *individuals*
end do
for $\mathbf{P}^t$ **do**
Implicit mutation
    replace the worst n_{bot}*individuals by* n_{bot} *random individuals*
end do
Set $t \coloneqq t + 1$
until *stopping criterion*

The limit-behavior (i.e. when $t \to \infty$) of the BEOs is demonstrated in the following theorem.

Theorem 4.1: Let $\mathbb{S}$ be a finite genotype space of individuals, $\mathbb{S}^*$ the set of all finite lists over $\mathbb{S}$ and $\bar{f}: \mathbb{S} \to \mathbb{R}$ a fitness function to be maximized. For an arbitrary initial population $\mathbf{P}^0 \in \mathbb{S}^*$ and a homogeneous evolution, the BEOs converge to the global optimal of $\bar{f}$ with probability 1.

Proof: We begin by demonstrating that the $(n_{\boldsymbol{P}} + n_{\boldsymbol{B}})$ replacement mechanism implies assumption (A6) (see Sect. 2.1.3.5). Let $M_{\mathbf{P}^t} \in \mathbb{S}^*$ be the list containing the optimal solutions of $\mathbf{P}^t \in \mathbb{S}^*$, such that $M_{\mathbf{P}^t} \subseteq \mathbf{E}^t$. The elitist parent selection mechanism implies $\#\mathbf{E}^t < n_{\mathbf{P}}$. Hence:

$$M_{\mathbf{P}^t} \subseteq \mathbf{E}^t \subset \mathbf{P}^t \tag{4.17}$$

The deterministic replacement law says that, after fitness-based ranking of $\mathbf{P}^t \cup \mathbf{B}^t$, the best $n_{\mathbf{P}}$ individuals survive. Thus, from (4.17), it follows:

$$\Pr\left(M_{\mathbf{P}^t} \cap f_s(\gamma, \mathbf{P}^t \cup \mathbf{B}^t)\right) = \Pr\left(\mathbf{P}^t \cap f_s(\gamma, \mathbf{P}^t \cup \mathbf{B}^t)\right) = 1 \tag{4.18}$$

and (A6) is satisfied, since naturally $M_{\mathbf{P}^t \cup \mathbf{B}^t} = M_{\mathbf{P}^t}$.

We are now left to demonstrate that assumption (A3) is satisfied by the algorithm and that the neighborhood associated with the similarity control and implicit mutation operators is $\mathbb{S}$ itself. It must be noted that replacing an individual by a randomly generated genotype is equivalent to perform gene-by-gene mutation with a mutation ratio $r_m = 0.5$. Thus, by (2.20), for a genotype of length l, the probability of transitioning from an $s \in \mathbf{P}^t$, selected for replacement, to any $t \in N(s)$ is positive and (A3) holds. Furthermore, $N(s) = \mathbb{S}$ and (A4) is trivially satisfied. The conditions of Corollary 2.1 are satisfied. The BEOs converge to the global optimal of $\bar{f}$ with probability 1. ■

4.2.2 Evolutionary-based reliability index approach

The RIA, as stated in (4.8), is quite challenging to solve for real-scenario structural reliability problems and asks for a compromise between accuracy and efficiency. One major difficulty is related with the lack knowledge about the failure surface. While in the stress space (see (3.9)) the quadratic failure surface is defined as an ellipsoid, meaning it exists in all directions, it may have a different shape in the uncertainty space, depending on the functional relationship between the random variables and the Tsai number, R. Thus, we identify an additional *implicit objective* of the RIA problem: *to identify the directions for which the failure surface exists.*

Through the next sections, we present a new EA specifically developed to solve the RIA. A set of new evolutionary operators and algorithmic strategies was developed to overcome the natural inability of EAs to handle equality constraints, while actively pursuing the solution to both objectives of the RIA: to find the minimum distance to and the direction of the MPP.

Novelties include the decomposition of random variables, the redefinition of the RIA as an equivalent penalty problem, a mixed real-binary coding of the search variables and two new evolutionary operators: one for the genetic repair of the solutions (to impose the equality constraint of the reliability assessment problem) and another for the progressive reduction and reallocation of the search domain (implicitly steering the evolution and focusing the search in regions of the failure surface with higher probability density, where the MPP is expected to be). The application of these operators is more efficient if some knowledge of the problem is known. For that purpose, an initial estimate of the reliability index is calculated. The next sections will detail these procedures.

The proposed method is suitable to be applied to any EA, with elitist strategy, and is expected to set the basis for further developments on the design optimization of more complex structures with multiple failure criteria.

4.2.2.1 Decomposition of random variables and redefinition of the RIA

The first step in evolutionary search is to define the fitness function, which is related to the objective functions and the constraints of the problem. To improve the efficiency of the search process, the random variables are decomposed into their magnitude (norm) and direction (direction cosines) components, in the standardized uncertainty space. Any coordinate in the standardized uncertainty space is defined by:

$$\mathbf{y} = \beta\mathbf{a} \quad ,\forall\, \mathbf{a}: \begin{cases} -1 \le a_i \le 1 \\ \sum_{i=1}^{N_\mathbf{y}} (a_i)^2 = 1 \end{cases} \tag{4.19}$$

where $\mathbf{a}$ is the vector of direction cosines and $N_\mathbf{y} = N_\mathbf{x} + N_{\boldsymbol{\pi}}$ is the total number of random variables. Assuming the hypothesis that $\mathbf{y} = T(\mathbf{x}, \boldsymbol{\pi})$ is both known and invertible, then we write:

$$G_2(\mathbf{y}) = g_2\big(T^{-1}(\mathbf{y})\big) \tag{4.20}$$

After (4.19) and (4.20), the primal RIA problem, in (4.8), is converted into a penalty formulation, in the shape of the unconstrained maximization of a fitness functional $\bar{f}: \Omega_2 \rightarrow \mathbb{R}_0^+$, suitable to be solved by EAs, as follows:

$$\max_{\beta, \mathbf{a}} \quad \bar{f} = C - \beta - \lambda \Gamma\left(g_2\left(T^{-1}(\mathbf{y})\right)\right) \quad , \forall\, \mathbf{y} = \beta \mathbf{a} \tag{4.21}$$

where C is a high valued positive constant, λ is a scaling factor and $\Gamma(\cdot)$ is a penalty function, according to the degree of violation of the equality constraint $g_2\left(T^{-1}(\mathbf{y})\right) = 0$, given by [24,25]:

$$\Gamma(g_2) = \begin{cases} 0 & , \text{if } |g_2| \leq \eta \\ K|g_2|^q & , \text{if } |g_2| > \eta \end{cases} \tag{4.22}$$

where η is a very small positive real, representing the allowable constraint violation. The exponential form of the penalty function allows to achieve a gradual penalization of the solutions, organizing them according to their level of violation in a continuous manner. For a greater control over the penalty applied, the constants K and q are defined considering two degrees of violation and penalty: weak penalty p_1, corresponding to a small constraint violation $g_{2,1}$, and strong penalty p_2, corresponding to a significant constraint violation $g_{2,2}$. Introducing the two degrees of violation, in (4.22), and equating to K and q, one obtains [24,25]:

$$q = \frac{\ln(p_1/p_2)}{\ln\left(g_{2,1}/g_{2,2}\right)}$$

$$K = \frac{p_1}{\left|g_{2,1}\right|^q} \tag{4.23}$$

One of the characteristics of the proposed penalty problem is the independence between β and $\mathbf{a}$ as search variables, allowing to easily manipulate the genotypes of the solutions, in order to achieve the two objectives of the RIA. This independence is fundamental to the application of the new evolutionary operators presented throughout the next sections.

Furthermore, in (4.21), notice that g_2 is defined in the (non-standard) uncertainty space, while β and $\mathbf{a}$ are defined in the standardized uncertainty space. However, this is only possible if the transformation between both spaces is known, avoiding the transformation of the implicit limit-state function to the standardized uncertainty space and ultimately saving computing time.

4.2.2.2 Mixed genotype structure

Another fundamental step in evolutionary search is the definition of an appropriate data structure, to represent the search variables and further manipulate the solutions, during the search process. The kind of data structure limits the kind of evolutionary operators to be applied and, for that reason, must respect the characteristics of the problem at hands. A suitable data structure is one that allows algorithms to perform a more efficient search.

Regarding the RIA problem, we introduce the concept of *mixed genotype*.

Definition 4.2: Let $\mathbb{S} = \mathbb{B}^l$ be the set of binary strings of length l, whose elements $s(\mathbf{a}) \in \mathbb{B}^l$ are the binary representation of the direction cosines $\mathbf{a}$, where $l = \sum_{i=1}^{N_y} l_i$, l_i is the number of bits allocated to each component a_i of $\mathbf{a}$ and $s(\cdot)$ is the real-to-binary mapping function. An array $\left(\beta, s(\mathbf{a})\right) \in \mathbb{R} \times \mathbb{B}^l$ is called a *mixed genotype* and represents a point in the standardized uncertainty space, as defined in (4.19).

A schematic representation of a mixed genotype is shown in Figure 4.3. From the previous definition another one follows.

Definition 4.3: Let $\mathbb{S}^*$ be the set of all finite lists over $\mathbb{R} \times \mathbb{B}$. A list $\mathbf{P}^t \in \mathbb{S}^*$, containing $n_{\mathbf{P}}$ mixed genotypes, defines the *population*, at generation t. Similarly, the list $\mathbf{E}^t \subset \mathbf{P}^t$, containing the n_{top} mixed genotypes of $\mathbf{P}^t$ with the highest fitness, defines the *elite group* of the population.

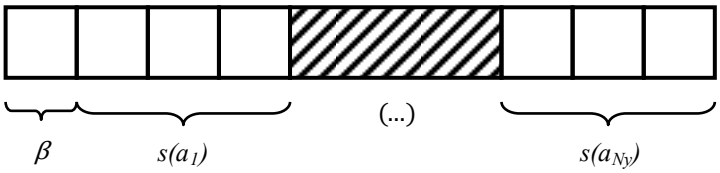

Figure 4.3: Schematic representation of a mixed genotype.

4.2.2.3 Individual genetic repair operator

The need to satisfy an implicit equality constraint makes the RIA problem hard to solve. Moreover, EAs are not capable *per se* of automatically satisfying equality constraints and often rely on the successive penalization of infeasible solutions, which largely deteriorates the performance of the algorithms. Dealing with equality constraints in continuous NLP problems implies the reduction of the feasible space to a hyper-surface (the failure surface), defined in a hyper-volumetric search space (the uncertainty space).

To address the infeasibility of the solutions generated by the BEOs, at any generaion, before every fitness-based ranking step in Algorithm 4.1, we introduce a novel repair mechanism, called the *individual genetic repair operator* (IGRO). To repair a genotype with an infeasible phenotype means to transform that same genotype into another with a feasible phenotype. The repair may be *total*, if the constraint violation is eliminated, or *partial*, if the constraint violation is reduced. In specific, the action of the IGRO, upon each solution, consists on the manipulation of the genetic information referent to β, while the genes relative to vector **a** are kept intact. Hence, the IGRO can be seen as a special case of adding local search to the BEOs. More precisely, as a *Lamarckian*-kind repair mechanism with a twofold purpose:

1. *local improvement:* eliminates the individuals' constraint violation, finding the value of β, along direction **a**, for which the failure surface is defined;
2. *identification of promising search directions*: bias the search towards the failure surface, penalizing individuals whose constraint violation cannot be eliminated.

For convenience, consider the following notation:

$$g_2(\beta|\mathbf{a}) = g_2\left(T^{-1}(\mathbf{y})\right) \quad ,\forall\, \mathbf{y} = \beta\mathbf{a} \tag{4.24}$$

as a simplified representation of g_2 as a function of β, given a fixed search direction **a** in the uncertainty space.

Given a point in the uncertainty space (i.e. an individual), the IGRO consists on the following *constrained one-dimensional inverse minimization problem* of the limit-state function:

$$\min_{\beta} \ |g_2(\beta|\mathbf{a})| \tag{4.25}$$

becoming an inner cycle of the RIA itself. The previous problem is said *constrained*, because the search direction **a** of the solution is kept constant, and *inverse*, since the sought optimal value ($|g_2(\beta|\mathbf{a})| = 0$) is known and the objective is to determine the corresponding *unknown* value of β. The numerical solution of the problem (4.25) is achieved by a deterministic iterative process, represented by the following asymptotic sequence:

$$|g_2(\beta^0|\mathbf{a})|, |g_2(\beta^1|\mathbf{a})|, \dots, \left|g_2\left(\beta^k|\mathbf{a}\right)\right|, \left|g_2\left(\beta^{k+1}|\mathbf{a}\right)\right|, \dots, \rightarrow 0 \tag{4.26}$$

with:

$$\beta^{k+1} = \beta^k \pm \Delta\beta^k \tag{4.27}$$

where $k = 0, \dots, k_{max}$ is the number of the current iteration, β^0 is the initial value of β and k_{max} is the maximum number of iterations. The repair mechanism becomes completely characterized by a nonlinear increment function $\Delta\beta^k$, given by the following proposed exponential law:

$$\Delta\beta^k = \Delta\beta^k_{max}\left(2\, e^{\frac{|g_2(\beta^k|\mathbf{a})|-g_{2,0}}{g_{2,0}}\ln(2)} - 1\right) \tag{4.28}$$

where $g_{2,0} = g_2(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi})$ is the deterministic realization of the stochastic stress limit-state function and $\Delta\beta^k_{max}$ is a user-defined parameter relative to the amplitude of the increment. The increment function is strictly decreasing on $[0, g_{2,0}]$, with maximum value $\Delta\beta^k_{max}$, at $|g_2(\beta^k|\mathbf{a})| = g_{2,0}$, and minimum value 0, at $|g_2(\beta^k|\mathbf{a})| = 0$. Furthermore, $\Delta\beta^k > 0$, if $g_2(\beta^k|\mathbf{a}) > 0$ and $\Delta\beta^k < 0$ otherwise. Overall, the exponential law allows the iterative process to run more efficiently towards convergence, varying from larger to more refined increments, as the sequence approaches zero.

In what follows, we make some considerations about the behavior of the IGRO. In particular, we establish the conditions under which sequence (4.26) converges.

A sequence of real numbers is said to converge if it approaches some limit. A sufficient condition is given by the Monotone Convergence Theorem: "*every bounded monotone sequence of real numbers has a limit"* [26].

The IGRO is said to converge if $\lim_{k\to\infty}|g_2(\beta^k|\mathbf{a})| = 0$. Thus, it suffices to satisfy:

$$|g_2(\beta^k|\mathbf{a})| \in [0, M] \quad , \forall k = 0,1,2, \dots \tag{4.29}$$

$$|g_2(\beta^{k+1}|\mathbf{a})| \leq |g_2(\beta^k|\mathbf{a})| \quad , \forall k = 0,1,2, \dots \tag{4.30}$$

in order to achieve convergence.

Equations (4.26)-(4.28) show that the iterative process of the IGRO depends on two factors: the relation between the values of $\Delta\beta^k_{max}$ and $|g_2(\beta^{k+1}|\mathbf{a})|$ and the functional relation between g_2 and β, along a given search direction $\mathbf{a}$. Thus, in order to satisfy conditions (4.29) and (4.30) it is necessary to characterize and establish conditions on $\Delta\beta^k_{max}$ and $g_2(\beta|\mathbf{a})$.

Determining appropriate values of $\Delta\beta^k_{max}$ is crucial, in terms of the efficiency and numerical stability of the IGRO. If too small, the iterative process becomes cumbersome. If too large, stability may be lost, and the method diverge. The Monotone Convergence Theorem above is a rather strong statement. In truth, every convergent sequence is bounded, but not necessarily monotone [26]. Yet, in terms of practical efficiency of the IGRO, it makes little sense for sequence (4.26) not to be monotone, in addition to ensuring numerical stability and convergence. For these reasons, we impose the monotony condition (4.30) to the IGRO and use it to derivate a relation between $\Delta\beta^k_{max}$ and $|g_2(\beta^{k+1}|\mathbf{a})|$.

At any iteration k, given $|g_2(\beta^k|\mathbf{a})|$, the calculation of the next term in the sequence requires the calculation of $\Delta\beta^k$, as an exponential function of the current term. Solving (4.28) for $|g_2(\beta^k|\mathbf{a})|$ and replacing on the right-hand side of (4.30), after some algebra, the following inequality results:

$$\left(2e^{\frac{|g_2(\beta^{k+1}|\mathbf{a})|-g_{2,0}}{g_{2,0}}\ln(2)} - 1\right) \leq \frac{\Delta\beta^k}{\Delta\beta^k_{max}} \tag{4.31}$$

The previous inequality is necessary and sufficient for the monotony of sequence (4.26). A parameter $\Delta\beta_{max}^{k}$ satisfying such condition is said to be *sufficiently small*. If not, it must be reduced.

We now establish a sufficient condition for the convergence of the IGRO, as follows.

Theorem 4.2: Let $g_2(\beta|\mathbf{a})$: $\mathbb{R} \to \mathbb{R}$ be a continuous function for which there is an $M > 0$, such that $g_2(\beta|\mathbf{a})$ is decreasing on $[0, M]$. If $g_2(\beta^0|\mathbf{a}) = M$ and $\Delta\beta_{max}^{k}$ is sufficiently small, then the IGRO converges asymptotically to zero.

Proof: The continuity of $g_2(\beta|\mathbf{a})$ implies that the increment function $\Delta\beta^k$, in (4.28), is defined for all β, at any k. Furthermore, being a decreasing function on $[0, M]$ implies that there exists a $\bar{\beta}$ such that $g_2(\bar{\beta}|\mathbf{a}) = 0$, for a given search direction $\mathbf{a}$, and at least one interval containing zero, where it is possible to generate a decreasing sequence of real numbers.

We now demonstrate that sequence (4.26) is bounded. To impose $g_2(\beta^0|\mathbf{a}) = M > 0$ implies the sequence starts on the positive side of $g_2(\beta|\mathbf{a})$. If $\Delta\beta_{max}^{k}$ is sufficiently small, the sequence is decreasing and $|g_2(\beta^k|\mathbf{a})| \leq M$, for all k. Then, the sequence is bounded above by M. Similarly, if $\Delta\beta_{max}^{k}$ is sufficiently small, the increment function $\Delta\beta^k$, in (4.28), generates sequences $\{\Delta\beta^k\} \to 0$, as $|g_2(\beta^k|\mathbf{a})| \to 0$, for $k \to \infty$. Then, $|g_2(\beta^k|\mathbf{a})| \geq 0$, for all k, and sequence (4.26) is bounded below by zero. The IGRO produces a bounded decreasing sequence and therefore converges asymptotically to zero.■

A result of the past theorem is $\lim_{k\to\infty}|g_2(\beta^k|\mathbf{a})| = 0 \Rightarrow \lim_{k\to\infty}\Delta\beta^k = 0 \Rightarrow \lim_{k\to\infty}\beta^k = \bar{\beta}$.

Regarding the behavior of the IGRO towards convergence, we identify two main modes of convergence: strong and weak convergence. Both satisfy the conditions of Theorem 4.2. *Strong convergence* represents the ideal behavior designed for the IGRO. The algorithm describes a descent and asymptotic search path, on the positive side of $g_2(\beta|\mathbf{a})$, starting from a positive initial estimate $g_2(\beta^0|\mathbf{a}) = M$ until $g_2(\bar{\beta}|\mathbf{a}) = 0$. *Weak convergence* consists in an oscillatory and asymptotic path, around $g_2(\bar{\beta}|\mathbf{a}) = 0$, starting from a positive initial estimate $g_2(\beta^0|\mathbf{a}) = M$ and characterized by sequentially smaller amplitudes $|g_2(\beta^k|\mathbf{a})|$, at every k. This oscillatory behavior originates in increments $\Delta\beta^k$ large enough to alternate between positive and negative values of $g_2(\beta|\mathbf{a})$, in consecutive iterations, while $\Delta\beta_{max}^{k}$ is still sufficiently small. In agreement with Theorem 4.2, due to the nonlinearity of the IGRO, *combined modes of convergence* are also allowed, where oscillatory and descent search paths may interchange.

About the previous theorem, notice that reliability assessment is only applied to design solutions that are deterministically safe, i.e. for which $g_{2,0} > 0$. Therefore, it is reasonable to impose a positive initial term of sequence (4.26). However, Theorem 4.2 assumes the idealized situation where the limit-state function is decreasing on $[0, M]$. In most cases, it is not possible to assert whether $g_2(\beta|\mathbf{a})$ is actually decreasing on $[0, M]$, but this should be an acceptable hypothesis given the measurable proximity between the origin of the uncertainty space and the failure surface, particularly for optimized structural systems.

In case $g_2(\beta|\mathbf{a})$ is non-decreasing on $[0, M]$, the IGRO is capable of producing convergent sequences, provided they are bounded by some interval $[0, M]$, with $M > 0$, and there is a positive $M_1 \leq M$, such that $g_2(\beta|\mathbf{a})$ is decreasing on $[0, M_1]$. That is to say the algorithm accepts the initial violation of inequality (4.31), for a finite number of iterations, producing a positive non-decreasing transient sequence that must be followed by a decreasing subsequence converging to zero.

In fact, the parameter $\Delta\beta_{max}^{k}$ is only updated if sequence (4.26) alternates between positive and negative values of the limit-state function and monotony is lost. That is, if:

$$g_2(\beta^{k+1}|\mathbf{a})g_2(\beta^k|\mathbf{a}) < 0 \quad \wedge \quad |g_2(\beta^{k+1}|\mathbf{a})| > |g_2(\beta^k|\mathbf{a})| \tag{4.32}$$

Otherwise, it means the sequence is decreasing (displaying strong, weak, or combined convergence) or it is producing non-decreasing sequences, on the positive side of $g_2(\beta|\mathbf{a})$. Given the dependence of both $\Delta\beta^k$ and $\left|g_2\left(\beta^{k+1}|\mathbf{a}\right)\right|$ on $\Delta\beta^k_{max}$, the updating process becomes a inner cycle of the current iteration, increasing the number of finite element model analysis. Therefore, instead of using (4.31) to iteratively update $\Delta\beta^k_{max}$, we propose a shortcut solution. Assuming $g_2(\beta|\mathbf{a})$ is decreasing on $[0, g_{2,0}]$, then $\Delta\beta^k_{max}$ defines the largest possible increment value. The principle behind the proposed updating scheme is that $\Delta\beta^k_{max}$ should never be larger than $\lim_{k\to\infty}\beta^k = \bar{\beta}$. Hence, we propose the following recurrence formula:

$$\Delta\beta^k_{max} = \alpha\,\beta^k \quad ,\alpha \in [0,1] \tag{4.33}$$

to be applied iteratively, until $\Delta\beta^k_{max}$ is sufficiently small. We observed $\alpha = 3/4$ to be the most effective value, achieving sufficiently small values within one or few cycles, without decreasing the parameter too much.

Additionally, in order to further increase the efficiency of the IGRO, whenever sequence (4.26) displays weak convergence, the new value β^{k+1} is obtained by the bisection of the range between β^{k-1} and β^k, replacing equations (4.27) and (4.28), as follows:

$$\beta^{k+1} = \frac{\beta^k + \beta^{k-1}}{2} \tag{4.34}$$

Overall, the proposed iterative process runs efficiently towards convergence, varying from larger to more refined increments, as the limit-state function approaches zero. In practice, convergence is achieved, in a finite number of iterations, after the following inequality is verified:

$$\left|g_2\left(\beta^k|\mathbf{a}\right) - 0\right| \le \eta \quad k = 1, \dots, k_{max} \tag{4.35}$$

where η is a very small valued non-negative real, as defined in (4.22).

Stopping conditions are also applied, at various levels and with different purposes. These are activated independently of the convergence state of the IGRO. The iterative process is aborted if (1) $\beta^k \ge \beta_{max}$ and $g_2\left(\beta^k|\mathbf{a}\right) > 0$, for any k, where β_{max} is an imposed maximum allowed value for β^k. It may happen either because the solution is not pointing to the failure surface (if it is not a closed surface) or if it is too far, along the search direction $\mathbf{a}$; (2) if $\Delta\beta^k_{max}$ is not sufficiently small after k_{count} consecutive attempts, for the same reasons of the previous condition; (3) if the $k = k_{max}$, for same reasons of the two previous conditions and to avoid the iterative process to become too long; (4) if $g_2\left(\beta^k|\mathbf{a}\right) > g_{2,0}$, for any k. This condition prevents the occurrence of numerical errors in the finite element code due to theoretical limitations of the physical model, such as the locking effect. Whenever the IGRO stops, before convergence is achieved, the respective design solutions are penalized as defined in (4.22). Figure 4.4 represents the flowchart of the IGRO.

We close this section with a comment on the relation between the IGRO and the need for a mixed genotype. If the genetic information about β was written as a binary-coded variable, the resolution of the repair mechanism would be limited by both the allocated number of genes (bits) and the respective size-constraints, ultimately making the procedure unviable, because $g(\beta|\mathbf{a}) = 0$ may not be defined for every combination of bits. On the contrary, a real-coded gene allows a theoretically infinite resolution of the repair mechanism, only limited by the floating-point format.

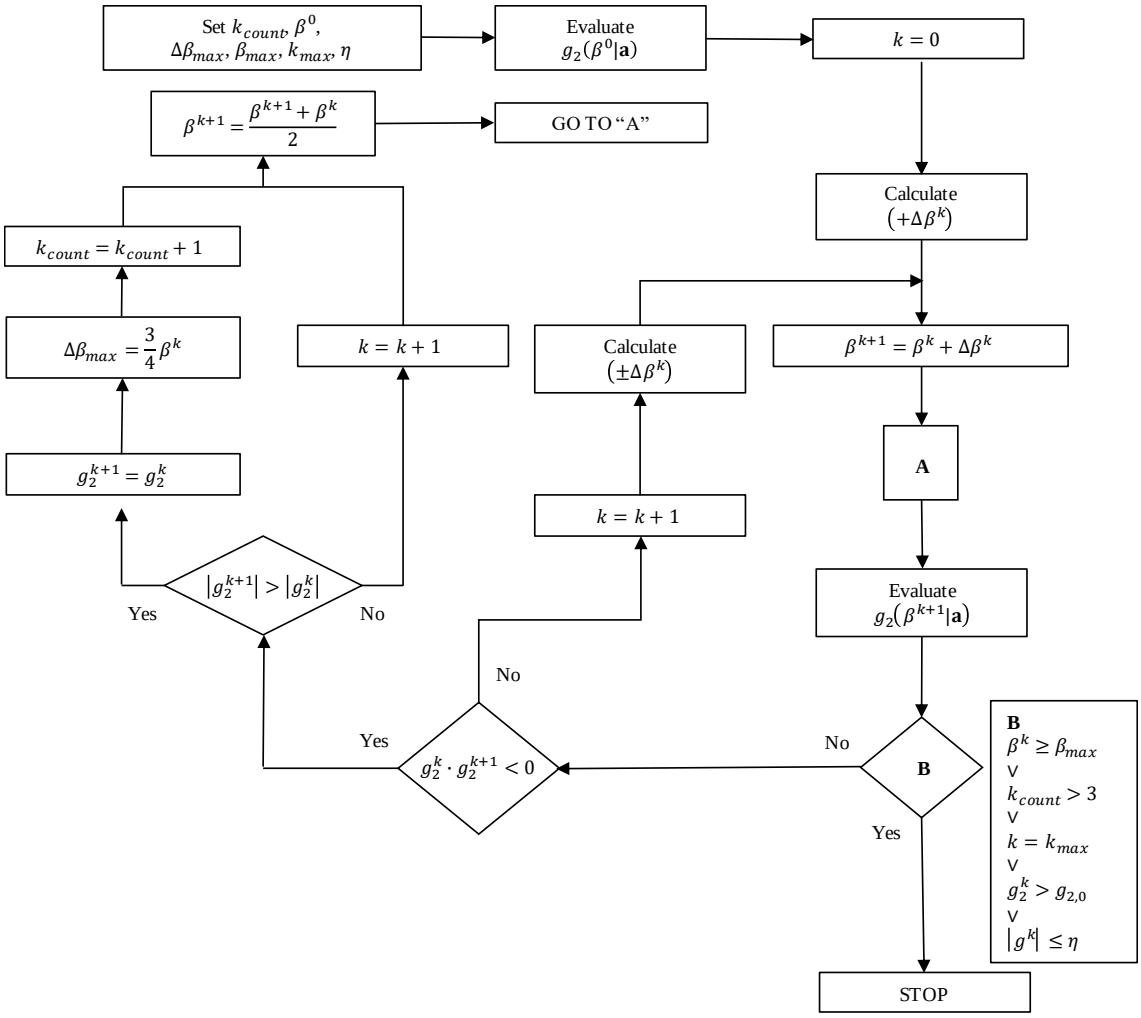

Figure 4.4: Flowchart of the individual genetic repair operator.

4.2.2.4 Reduction and reallocation of the search space operator

To satisfy the equality constraint of the RIA is mandatory to any method designed to solve the problem. At the same time, it is the pitfall of any method performance-wise. So far, the IGRO acts *explicitly* on the genotypes, by correcting the real-coded gene related to β, while keeping the binary-coded genes relative to vector **a** fixed. The operator improves the solutions by translating them to the failure surface and increasing their potential within the population. However, not all solutions will satisfy the imposed constraint, after being repaired. Those must be penalized. But, given the size of the uncertainty space, it is not of practical interest to keep searching along directions not pointing towards the MPP.

A second novel evolutionary operator is now introduced, with an *implicit* action on the binary-coded genes related to the direction cosines **a** of solutions, called the *reduction and reallocation of the search space operator* (RSO). The operator is complementary to the IGRO and has a twofold purpose:

1. to identify the target region of the uncertainty space, where the MPP is expected to lie;
2. to focus the search on the target region, increasing the resolution of the numerical discretization imposed by the binary data structure. From an evolutionary point-of-view, it implies that certain phenotypical characteristics will be given preference over others.

In the scope of the current problem, the target region of the uncertainty space, is defined as follows.

Definition 4.4: Let $B(\mathbf{y_{MPP}};\epsilon)$ be a ϵ-neighborhood centered at the MPP and L the limit-state surface (see general Definition 3.5). Then, the set:

$$\mathcal{Z} = \{\mathbf{y} \in \mathbb{R}^{N_\mathbf{y}} : \mathbf{y} \in B(\mathbf{y}_{\mathrm{MPP}};\epsilon) \ \wedge \ T^{-1}(\mathbf{y}) \in L\} \tag{4.36}$$

is known as the *interest region*.

In practice, the limit-state function is implicit and the location of the interest region is unknown. The RSO follows a search strategy divided into *three* stages, in order to achieve a satisfactory approximation of the *true* interest region. The operator uses *knowledge of the problem* acquired from the deterministic design solution, to determine initial bounds to the search space, and updates its knowledge based on the data generated during the evolutionary process, sequentially updating the current search region. The initial bounded search region is defined as follows.

Definition 4.5: Let β_{min} and β_{max} be two Euclidian distances measured from the origin of the uncertainty space, such that $\beta(\mathbf{y_{MPP}}) \in [\beta_{min}, \beta_{max}]$. Then, the set:

$$\mathcal{Z}^0 = \{\mathbf{y} \in \mathbb{R}^{N_\mathbf{y}} : \beta \in [\beta_{min}, \beta_{max}], \mathbf{a} \in [-1,1]\} \tag{4.37}$$

is called the *initial search region* and $\mathcal{Z} \subset \mathcal{Z}^0$.

The first stage of evolution is run on $\mathcal{Z}^0$, seeking solutions that are potentially on $\mathcal{Z}$.Thus, the search space is initially reduced to the space between two hyperspheres of radius β_{min} and β_{max} respectively (see next section), while the BEOs are allowed to freely explore any search direction. The goal is to generate a comprehensive reference list of solutions of *high probabilistic failure content*.

Definition 4.6: A solution in the uncertainty space is said to be of *high probabilistic failure content* if and only if it satisfies $\beta \in [\beta_{min}, \beta_{max}] \wedge T^{-1}(\mathbf{y}) \in L$, for all $\mathbf{y} = \beta\mathbf{a}$. Furthermore, let $\mathbf{E}^t \subseteq \mathbf{P}^t$ be the elite group of the population, at generation t. Then, $\mathbf{E}_\mathcal{Z}^t \subseteq \mathbf{E}^t$ is the list containing the mixed genotypes in $\mathbf{E}^t$, whose phenotypes are solutions of high probabilistic failure content.

Appropriately choosing $\mathbf{E}_\mathcal{Z}^t$, at generation t, as the reference list within the population, once the equality $\mathbf{E}_\mathcal{Z}^t = \mathbf{E}^t$ is verified (i.e., elite group only contains solutions of high probabilistic failure content) a new search region is defined.

Definition 4.7: Let $\mathbf{E}_\mathcal{Z}^t = \mathbf{E}^t$and n_{top} be the number of elite individuals. Then, the set:

$$\mathcal{Z}^t = \left\{ \mathbf{y} \in \mathbb{R}^{N_\mathbf{y}} : \mathbf{y} \in \mathbf{E}_\mathcal{Z}^t \wedge a_i \in \left[a_{i_{min}}, a_{i_{max}}\right], i = 1, \dots, N_\mathbf{y}\right\} \tag{4.38}$$

where $a_{i_{min}} = \min\left(a_i^j, j = 1,..,n_{top}\right)$ and $a_{i_{max}} = \max\left(a_i^j, j = 1,..,n_{top}\right)$, is called a *reduced search region.*

The reduction of the search region, from $\mathcal{Z}^0$ to $\mathcal{Z}^t$, sets the beginning of the second stage of evolution. For the new search region, the elite mixed genotypes are recoded (considering the new size-constraints) and the remaining population is randomly generated, introducing the needed genetic diversity to explore the new search region. The advantage of restarting the procedure with an already defined elite group is the possibility to bias the future evolution of the population. In practice, it instigates the algorithm to choose the most promising evolutionary path available and steer the search to the most promising areas of the reduced search region.

After the reduction of the search space, the population evolves near the failure surface, requiring only small corrections of β. The numerical burden of the IGRO is alleviated, on the one side, by reducing the number of genetic repairs. However, on the other side, every individual in the population needs to be repaired, since all the individuals are of high probabilistic failure content. Because solutions are no longer penalized, the single differentiating fitness factor is their β-value.

The reduction of the search space around the elite group, at generation t, poses two issues:

(1) the fact that the search process restarts with an elite group, may cause the evolution of the population to nest quickly and to favor certain genetic characteristics that could otherwise be improved.
(2) even though, ideally, $\mathcal{Z} \subset \mathcal{Z}^0$ and $\mathcal{Z}^t$ aims to be a good approximation of $\mathcal{Z}$, it is not possible to guarantee that $\mathbf{y}_{\mathrm{MPP}} \in \mathcal{Z}^t$, because the MPP may be outside of the reduced size constraints.

Issue (1) is a consequence of the loss of diversity in the population, resulting from the forced manipulation of the genetic characteristics of the population. Issue (2) results from the implicit definition of the limit-state function and the fact that the location of the true interest region $\mathcal{Z}$ and of the MPP itself are unknown. To cope with these issues:

(1) *the elite's global diversity is measured, at every generation*. It is defined at phenotypic level as the sample variance of the β-values within the elite, $\mathrm{Var}_{\mathbf{E}}(\beta)$, measuring the potential of the elite group to create improved search regions, since the phenotypical similarity between the search directions of different solutions is predictable, after the reduction of the search space.
(2) *the search region, $\mathcal{Z}^t$, is upgraded sequentially, at every pre-determined number of generations, $t_{\mathcal{Z}}$, for an optimal maturation of the elite*. The upgrade process is repeated, while a minimum level of diversity in the elite, σ_a^2, is preserved.

If $\mathrm{Var}_{\mathbb{E}}(\beta) < \sigma_{min}^2$, then there are no advantages in proceeding with the iterative reduction of the search space. On the third stage of evolution, the population evolves on the last reduced search region, until final convergence.

Overall, to prevent the hypothesis of dimensions being reduced to a single point, or being too small, a minimum diameter is imposed to each size constraint, such that:

$$\left|a_{i_{max}} - a_{i_{min}}\right| \geq \Delta a \tag{4.39}$$

for $i = 1, .., N_{\mathbf{y}}$. Thus, the search space is not only progressively reduced but also reallocated. Furthermore, the reduction operation allows to increase the resolution of the search process, making possible to have a precision similar to that of gradient-based methods, with only few bits per variable. The maximum possible quality of the resolution is thus defined by the diameter Δa and the number of bits per variable.

There are two critical steps to be taken before final convergence. The first is to define the first reduced search region, $\mathcal{Z}^{t_1}$, at generation t_1. The second is the loss of diversity in the elite group. A lower level of diversity means the potential to find better individuals is smaller and may be a sign that the population is getting closer to the MPP or converging to local minima. Then, the loss of diversity should be progressive, allowing to define search regions with a higher likelihood of containing the MPP. For that, a good combination of σ_a^2, $t_{\mathcal{Z}}$ and Δa is vital.

The goal is to find a finite sequence $\{\mathcal{Z}^0, \mathcal{Z}^{t_1}, \mathcal{Z}^{t_1+t_z}, \ldots, \mathcal{Z}^{t_1+nt_z}\}$, whose final term, $\mathcal{Z}^{t_1+nt_z}$, is a good approximation of $\mathcal{Z}$ and for which $\mathbf{y}_{\mathrm{MPP}} \in \mathcal{Z}^{t_1+nt_z}$. Convergence of the method is considered to happen only in the third stage of evolution. To be achieved, $\mathcal{Z}^{t_1+nt_z}$ must be defined and a minimum number of generations Δt has to be completed, in the first place. Then, a measure of similarity between elite groups, at consecutive generations, must be satisfied. The convergence criterion is written as follows:

$$\mathbf{E}^{t_1+\Delta t} \sim \mathbf{E}^{t_1+\Delta t+1} \quad , \text{if } \mathrm{Var}_{\mathbf{E}}(\beta) < \sigma_a^2 \tag{4.40}$$

where $\sim$ is a symbol representing similarity between two lists.

4.2.2.5 Knowledge of the problem

The proposed evolutionary operators, as presented in the previous sections, were developed to help the BEOs to explore the vast uncertainty space in a more efficient manner, avoiding uninteresting regions, such as those far from the failure surface. Yet, the heuristics supporting the proposed operators require some parameters to be correctly pre-determined for each design solution, to make their application more efficient, namely, $\Delta\beta_{max}^{0}$, β_{min} and β_{max}. Being related with the notion of the reliability index, the definition of these parameters would largely benefit if an estimate of the true value of $\beta(\mathbf{y_{MPP}})$ is available.

Let $g_2(\mathbf{x}, \boldsymbol{\pi})$: $\Omega_2 \rightarrow \mathbb{R}$ be differentiable at the origin of the uncertainty space. Hasofer and Lind [14] give the following approximation to the reliability index, based on a first-order Taylor expansion of the limit-state function, at $(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})$, as follows:

$$\beta^* = \left(\frac{g_2(\mathbf{x}, \boldsymbol{\pi})}{\sqrt{\sum_{i=1}^{N_x} \left(\frac{\partial g_2}{\partial x_i} \, \sigma_{x_i} \right)^2 + \sum_{i=1}^{N_{\boldsymbol{\pi}}} \left(\frac{\partial g_2}{\partial \pi_i} \, \sigma_{\pi_i} \right)^2}} \right)_{(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})} \tag{4.41}$$

In the previous equation, it is important to notice that β^* differs from β^0 found in (4.26), where the latter is a reference to the β-value of a solution before being repaired. Both parameters, however, can match. This estimate is then used as "knowledge of the problem" to define the three necessary parameters. For the current problem, it suffices that:

$$\beta_{min} = \beta_{ref} - 1 \, ; \quad \beta_{max} = \beta_{ref} + 1 \, ; \quad \Delta\beta_{max} = \beta_{max} \tag{4.42}$$

with:

$$\beta_{ref} = \begin{cases} \mathrm{Int}(\beta^*) & , \text{if } \beta^* - \mathrm{Int}(\beta^*) < 0.5 \\ \mathrm{Int}(\beta^*) + 0.5 & , \text{otherwise} \end{cases} \tag{4.43}$$

where $\mathrm{Int}(\cdot)$ represents the function returning the integer part of a real number.

To determine this estimate, the gradient vector of the limit-state function is only calculated once, at the mean values of the variables, before the evolutionary process starts. In fact, as will be shown in Chapter 5, the gradient information about each design solution will be already available, before reliability assessment starts. Yet, to avoid excessive model evaluations, the gradient vector with respect to the random variables is calculated analytically by the Adjoint Variable Method (see Chapter 5).

4.2.2.6 Genetic similarity

We now provide an explicit definition of genetic similarity between two mixed genotypes, based on a *variable-by-variable similarity* concept. We start by defining similarity between variables.

Definition 4.8: Let a_i be the ith component of vector $\mathbf{a}$, for $i = 1, .., N_{\mathbf{y}}$. Two individuals are said to have an equal value on a variable if $s(a_i)^{j_1} = s(a_i)^{j_2}$, for $j_1, j_2 = 1, \ldots, n_{\mathbf{P}}$, with $j_1 \neq j_2$.

Hence, we define similarity between individuals as follows.

Definition 4.9: Two solutions in the population are said to be similar if the number of different variables is less than or equal to a prescribed $\varepsilon_{\mathbf{P}} \in \mathbb{N}$.

The similarity control operator has a fundamental task in preventing individuals from becoming too similar. However, similarity control may have the perverse effect of eliminating solutions of probabilistic failure content from the population, inhibiting the RSO of concluding the first stage of evolution, on the initial search region $\mathcal{Z}^0$, with efficiency. Such behavior is enhanced if the number of bits allocated to each binary search variable a_i is low, because the number existing genetic combinations that produce high probabilistic failure content is low and naturally the genetic similarity between these solutions is expectedly higher. Such phenomenon is observed and discussed with more detail in Chapter 7.

In order to balance the disruptiveness of the similarity control operator and the necessary selective pressure to find solutions on the failure surface, during the first stage of evolution, the accepted degree of similarity between solutions in $\mathbf{E}_Z^t$ is higher than for those in the remaining population, after genetic repair. With such measure we attempt to increase the selective pressure towards solutions of high probabilistic failure content.

Definition 4.10: Two solutions of high probabilistic failure content, in $\mathbf{E}_Z^t$, are said to be similar if the number of different variables is less than or equal to a prescribed $\varepsilon_{\mathbf{E}_Z^t} < \varepsilon_{\mathbf{P}}$.

4.2.2.7 Algorithmic structure

The combination of the BEOs with the novel developments proposed above results in a powerful hybridized genetic algorithm capable of solving the RIA in a practical time. The ability of the BEOs in working with reduced-size genotypes and populations is only potentiated by the new evolutionary operators. Hence, the hybridized algorithm achieves both numerical accuracy and efficiency for a large number of different design solutions, with a single set of user-defined parameters, working with very low numbers of genes per variable and individuals in the population. For that reason, it is named the *Hybrid micro Genetic Algorithm* (HmGA). Algorithm 4.2 details the structure of the method.

The algorithmic structure of the HmGA is divided in three stages of evolution, each designed to achieve a certain goal, as imposed by the RSO. On the first stage, the successive populations search for solutions lying on the failure surface, at an acceptable distance from the origin of the uncertainty space, using the IGRO as the main tool for the task. On the second stage the RSO becomes the most important operator, as successive reduced search regions are defined, promoting a guided "zooming effect" that successively increases the resolution of the discrete search process on regions of the uncertainty space with higher probability densities. Lastly, on the third stage, the algorithm refines the search until final convergence. The flowchart of the HmGA in shown in Figure 4.5.

4.2.2.8 Limit-behavior of the HmGA

Regarding the ability of the HmGA to find the global optimal solution of the RIA with probability 1, when $t \to \infty$, it should be noted that the evolutionary process is carried out by the BEOs, as presented in Algorithm 4.1, which possess such property. However, the RSO imposes a different algorithmic structure to the HmGA and convergence to the global optimal MPP may not be guaranteed. For that reason, we state a (rather trivial) necessary sufficient and condition for the convergence of the HmGA.

Theorem 4.3: Let $\mathbb{S} = \mathbb{B}^l$ be the set of binary strings of length l, $\mathbb{S}^*$ the set of all finite lists over $\mathbb{R} \times \mathbb{B}^l$ and $\bar{f}: \mathbb{R} \times \mathbb{B}^l \to \mathbb{R}$ a fitness function to be maximized. For an arbitrary initial population $\mathbf{P}^0 \in \mathbb{S}^*$, if $\mathbf{y}_{\mathbf{MPP}} \in \mathcal{Z}^{t_1+nt_z}$ and the evolutionary process is executed by the BEOs, then the HmGA converges to the global optimal of $\bar{f}$ with probability 1.

The proof of the theorem is immediate, since the IGRO is a deterministic operator and does not interfere with the stochasticity of the evolution. Moreover, the RSO preserves the conditions

required in Corollary 2.1, as the similarity control and the implicit mutation operators do. Hence, it is necessary and sufficient that $\mathbf{y_{MMP}}$ belongs to the final search region $\mathcal{Z}^{t_1+nt_z}$ for the HmGA to be able to converge to the global optimal MPP, with probability 1. For that, a suitable combination of the user-defined parameters σ_a^2, $t_\mathcal{Z}$ and mainly Δa is vital.

Algorithm 4.2: (*Hybrid micro Genetic Algorithm*):

Initialization of 1st stage of evolution
Set parameters β_{min}, β_{max}, $n_\mathbf{P}$, n_{top}, n_{bot}, Δa, σ_a^2, t_z
Set $t \coloneqq 1$
Initialization of population $\mathbf{P}^t$
Set $\beta^j = \beta_{min}$ and randomly generate $s(\mathbf{a})^j \sim U(0,1)$, for $j = 1, \dots, n_\mathbf{P}$
Set size-constraints $a_i \in [-1,1]$, for $i = 1,..,N_\mathbf{y}$
Set initial search region $\mathcal{Z}^0$
repeat
Apply BEOs + IGRO (prior to fitness-based ranking)
Set $t \coloneqq t + 1$.
until $(\mathbf{E}_\mathcal{Z}^t = \mathbf{E}^t)$

Initialization of 2nd stage of evolution
repeat
Set new size-constraints $a_i \in \left[a_{i_{min}}, a_{i_{max}}\right]$, for $i = 1,..,N_\mathbf{y}$
Migrate solutions of $\mathbf{E^t}$ to new population $\mathbf{P^t}$
Randomly generate $\left[n_\boldsymbol{P} - \left(n_{top} + 1\right)\right]$ solutions
Set $\beta^j = \beta_{min}$ and randomly generate $s(\mathbf{a})^j \sim U(0,1)$, for $j = 1, \dots, n_\mathbf{P}$
Set search region $\mathcal{Z}^t$
Set $t_{ref} \coloneqq t$
repeat
Apply BEOs + IGRO (prior to fitness-based ranking)
Measure the elite's sample variance $\mathrm{Var}_\mathbf{E}(\beta)$
Set $t \coloneqq t + 1$
until $\left(t - t_{ref}\right) = t_\mathcal{Z}$
Set $t \coloneqq t + 1$.
until $(\mathrm{Var}_\mathbf{E}(\beta) < \sigma_a^2)$

Initialization of 3rd stage of evolution
repeat
Apply BEOs + IGRO (prior to fitness-based ranking)
Set $t \coloneqq t + 1$.
until $\left(\mathbf{E}^{t_1+\Delta t} \sim \mathbf{E}^{t_1+\Delta t+1}\right)$

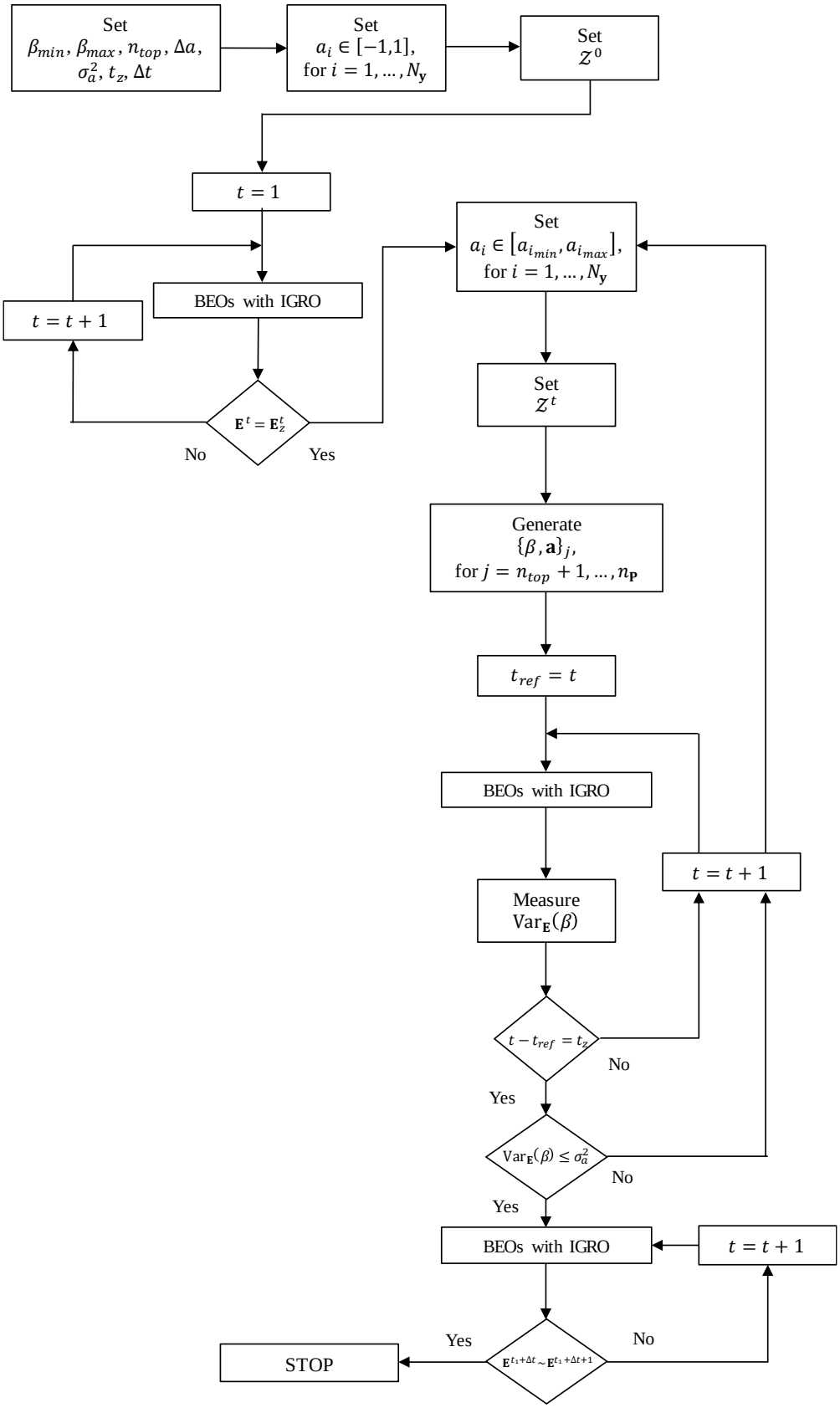

Figure 4.5: Flowchart of the HmGA.

4.2.3 Evolutionary-based performance measure approach

The PMA as stated in (4.16) is a challenging NLP problem that requires a compromise between accuracy and efficiency. Although sharing common features with the RIA, namely the notion of reliability index defined on a standardized uncertainty space, both methods differ from each other on the meaning of their objectives and optimal solutions. While the RIA searches for the point with maximum probability density, on a limit-state level set (the failure surface), the PMA searches for the point with the minimum limit-state value, on a probability density level set.

For that reason, when solved by gradient-based algorithms, as reported in several publications on the subject [27], the PMA shows higher numerical stability and efficiency, in comparison with the RIA, for design solutions for which the probabilistic reliability constraint is inactive. That is, when the reliability index of design solutions is higher than the allowable value, β^a. The PMA does not require a blind search for an implicit surface. Indeed, any probability density hypersphere of radius β^a, in the standardized uncertainty space, is explicitly defined as:

$$B(\beta^a) = \{\mathbf{y} \in \mathbb{R}^{N_\mathbf{y}} : \|\mathbf{y}\| - \beta^a = 0\} \tag{4.44}$$

and does not require any knowledge of the failure surface. Hence, whenever design solutions respect the allowable reliability level, the effective search space of the PMA is smaller than that of the RIA and the problem is in theory easier to solve by gradient-based algorithms. The opposite is also true whenever design solutions violate the allowable reliability level.

Based on the previous discussion, we present an EA developed to solve the PMA. Contrary to the RIA, in order to solve the PMA, only minor additions to the BEOs are needed, mainly related with the problem and the search variables' definition and how the equality constraint is handled.

4.2.3.1 Decomposition of random variables and redefinition of the PMA

We start by the definition of the fitness-function of the problem. In order to improve the efficiency of the search process, the standard normal random variables are reduced to their direction components, with an imposed norm. From (4.44), let $B(\beta^a)$ be the probability density hypersphere of radius β^a, centered at the origin of the standardized uncertainty space. Any coordinate $\mathbf{y} \in B(\beta^a)$ is given by:

$$\mathbf{y} = \beta^a \mathbf{a} \quad , \forall\, \mathbf{a}: \begin{cases} -1 \le a_i \le 1 \\ \sum_{i=1}^{N_\mathbf{y}} (a_i)^2 = 1 \end{cases} \tag{4.45}$$

where $\mathbf{a}$ is the vector of direction cosines and $N_\mathbf{y} = N_\mathbf{x} + N_\boldsymbol{\pi}$ is the total number of random variables. Because $\mathbf{y} = \mathbf{y}(\mathbf{a})$ and assuming that $\mathbf{y} = T(\mathbf{x}, \boldsymbol{\pi})$ is both known and invertible, then we write:

$$G_2(\mathbf{y}) = g_2\big(T^{-1}(\mathbf{a})\big) \quad , \forall\, \mathbf{y} = \beta^a \mathbf{a} \tag{4.46}$$

Thus, the primal PMA problem, in (4.16), is converted into an unconstrained maximization of the positive fitness function $\bar{f}: \Omega_2 \to \mathbb{R}_0^+$, suitable for EAs, as follows:

$$\max_{\mathbf{a}} \quad \bar{f} = C - \lambda\, g_2\big(T^{-1}(\mathbf{a})\big) \tag{4.47}$$

where C is a high valued positive constant and λ is a scaling factor. Notice that there is not the need to use any penalty function in the current formulation, because the search space is reduced to the hypersphere $B(\beta^a)$. Hence, the PMA becomes simply a directional search problem, having the direction cosines of any $\mathbf{y} \in B(\beta^a)$ as search variables. As before, notice that g_2 is defined in

the (non-standard) uncertainty space, while β^a and $\mathbf{a}$ are defined in the standardized uncertainty space, saving computer resources.

4.2.3.2 Binary genotype structure

A suitable data structure is one that allows algorithms to perform a more efficient search. Because the set of search variables is reduced to the direction cosines in the standardized uncertainty space, then solutions in the uncertainty space are fully represented with binary genotypes.

Definition 4.11: Let $\mathbb{S} = \mathbb{B}^l$ be the set of binary strings of length l. An array $s(\mathbf{a}) \in \mathbb{B}^l$ is called a *genotype* and represents a point in the standardized uncertainty space, as defined in (4.45), where $l = \sum_{i=1}^{N_y} l_i$, l_i is the number of bits allocated to each component a_i of $\mathbf{a}$ and $s(\cdot)$ is the real-to-binary mapping function.

From the previous definition, it follows.

Definition 4.12: Let $\mathbb{S}^*$ be the set of all finite lists over $\mathbb{B}^l$. A list $\mathbf{P}^t \in \mathbb{S}^*$, containing $n_{\mathbf{P}}$ genotypes, defines the *population*, at generation t. Similarly, the list $\mathbf{E}^t \subset \mathbf{P}^t$, containing the n_{top} genotypes of $\mathbf{P}^t$ with the highest fitness, defines the *elite group* of the population.

4.2.3.3 Genetic similarity

The concept of genetic similarity between individuals in the HmGA only considered the binary segment of the mixed genotypes. Therefore, the new algorithm applies the same *variable-by-variable* similarity control, as given by Definition 4.8 and Definition 4.9.

4.2.3.4 Algorithmic structure and stopping criterion

The algorithmic developments applied to the BEOs, in order to solve the PMA are notoriously simpler than those developed to solve the RIA. This is due to the inherent explicit definition of the equality constraint of the PMA, given by (4.44). The algorithm is able to work with reduced-size genotypes and populations and achieves a satisfactory convergence, in terms of accuracy and efficiency, for a large number of design solutions, with a single set of user-defined parameters. For that reason, it is called the *micro Genetic Algorithm* (mGA). The flowchart of the mGA is shown in Figure 4.6.

Contrary to the HmGA, the structure of the mGA comprises a single stage of evolution. Hence, the algorithm is said to converge, in a finite number of generations, as soon as the BEOs satisfy the following criterion. Let $s(\mathbf{a})^{n_{top}} \in \mathbf{E}^t$ represent the last member of the elite group, for a given generation t. The mGA is said to converge, in finite time, if it satisfies the following inequality:

$$|s(\mathbf{a})^{n_{top}} \in \mathbf{E}^t - s(\mathbf{a})^{n_{top}} \in \mathbf{E}^{t+\alpha}| \leq \varepsilon \quad \text{, for } \alpha = 1, \dots, \alpha_{max} \tag{4.48}$$

that is, if the absolute difference between the last member of the elite, at t, and the last member of the elite group, at $t + \alpha$, approximate zero with a desired accuracy, for α_{max} consecutive generations.

4.2.3.5 Limit-behavior of the mGA

Regarding the ability of the mGA to find the global optimal solution of the PMA with probability 1, when $t \to \infty$, it should be noted that the evolutionary process is carried out by the BEOs as presented in Algorithm 4.1, which possess such property, as demonstrated in Theorem 4.1.

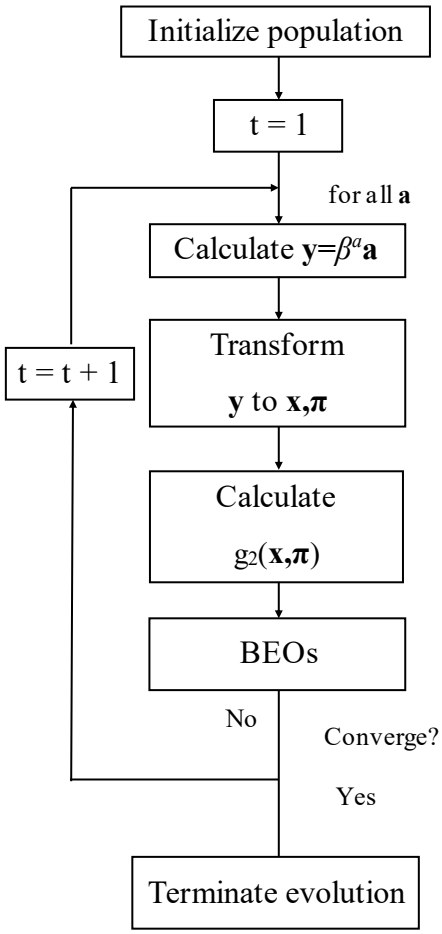

Figure 4.6: Flowchart of the mGA.

4.3 Concluding notes

This chapter begins with concise theoretical discussion on the numerical methods most commonly applied to solve the probability of failure integral: Monte Carlo Simulation, as a global reliability method, and the Reliability Index Approach (RIA) and the Performance Measure Approach (PMA), as local reliability methods. The preference for local reliability methods to assess the probabilistic reliability constraint in structural design optimization is explained.

Two novel evolutionary algorithms (EA) are specifically developed as numerical tools to solve both the RIA and the PMA problems. Such algorithms constitute a novelty in reliability assessment, since it is almost exclusively performed by gradient-based algorithms and sampling methods, in the literature. The preference for EAs over other optimization algorithms is justified by the fact that EAs do not require gradient information, convexity and continuity of the search space. Moreover, global convergence is proven to be achieved, provided the algorithms satisfy certain conditions. Yet, convergence is slower particularly in the presence of equality constraints.

In particular, the demanding nature of the local reliability methods asks for robust EAs, capable of solving the reliability assessment problem of innumerous design solutions. On its application to design optimization, reliability assessment must be both accurate and efficient, in a way that the optimal design problem not only is correctly solved but is viable, as well. A set of diversified operators called the basic evolutionary operators (BEOs) is introduced, combining high selective pressure (supported by a strong elitist strategy) with a strong exploratory capability (imposed by highly disruptive operators). These operators are the core of the two EAs developed to solve the

RIA and the PMA. A detailed description of the BEOs is provided and their convergence to the global optimal solution with probability 1 is demonstrated.

Regarding the RIA, the biggest challenge is to satisfy an implicitly defined equality constraint. Alone, the BEOs cannot provide a satisfying solution to the problem. A set of new evolutionary operators and algorithmic strategies was developed to overcome the natural inability of EAs to handle equality constraints. Novelties include the decomposition of random variables, the redefinition of the RIA as an equivalent penalty problem, with a mixed real-binary coding of the search variables and two new evolutionary operators: one for the genetic repair of the solutions and another for the progressive reduction and reallocation of the search space. Knowledge of the problem, in the shape of an initial estimate of the reliability index of design solutions is used to improve the efficiency of the algorithm.

On its turn, the PMA does not require a blind search for an implicit surface, through the search space. The method enjoys a higher numerical stability, due to an explicit equality constraint, not written in terms of an implicit structural response functional, but rather in terms of a symmetric probability density hypersurface in the space of standard normal random variables. The resulting EA, developed to solve the PMA, is therefore much simpler and bases its evolutionary process mostly on the BEOs. Yet, it largely benefits from the redefinition of the PMA as an unconstrained maximization problem, without the need for any sort of penalty function, reducing the search space to the surface of the hypersphere of radius equal to the allowable reliability index.

References

[1]. Melchers RE. Structural Reliability Analysis and Prediction. 2nd ed. Willey & Sons; 1999.

[2]. Aldosary M, Wang J, Chenfeng L. Structural reliability and stochastic finite element methods: State-of-the-art review and evidence-based comparison. Eng Computations 2018; 35(6); 2165-214.

[3]. Nikolaidis E, Ghiocel DM, & Singhal S. Engineering design reliability applications: for the aerospace, automotive and ship industries. CRC Press, 2007.

[4]. Schueller G. Efficient Monte Carlo simulation procedures in structural uncertainty and reliability analysis-recent advances. Struct Eng Mech 2009;32(1):1–20.

[5]. Hurtado JE, Ramírez J. The estimation of failure probabilities as a false optimization problem. Struct Saf 2013;45:1–9.

[6]. Rashki M, Miri M, Moghaddam MA. A new efficient simulation method to approximate the probability of failure and most probable point. Struct Saf 2012;39:22–9.

[7]. Ditlevsen O, Melchers RE, Gluver H. General multi-dimensional probability integration by directional simulation. Comput Struct 1990;36(2):355–68.

[8]. Melchers R. Structural system reliability assessment using directional simulation. Struct Saf 1994;16(1):23–37

[9]. Papaioannou I, Papadimitriou C, Straub D. Sequential importance sampling for structural reliability analysis. Struct Saf 2016;62:66–75.

[10]. Cornell CA. A proposal for a reliability-based code suitable for immediate implementation. Memorandum to members of ASCE Task Committee on Structural Safety 1967.

[11]. Ditlevsen O, Madsen HO. Structural reliability analysis. 1st ed. John Wiley & Sons; 1996.

[12]. Lind NC. An invariant second-moment reliability format. Paper No 113 Solid Mechanics Division, University of Waterloo, Ontario, Canada, 1972.

[13]. Ditlevsen O. Structural reliability analysis and the invariance problem. Technical report No 22, Solid Mechanics Division, University of Waterloo, Ontario, Canada, 1973.

[14]. 17. Hasofer AM, Lind NC. An Exact and Invariant First Order Reliability. J Eng Mech-ASCE 1974;100(EM1):111-21

[15]. Rackwitzt R, Fiessler B. Structural Reliability under combined random load sequences. Comput Struct 1978;9:489-94.

[16]. Horne MR, Price PH. Commentary on the level 2 procedure. Rationalization of Safety and Serviceability Factors in Structural Codes, report No. 63. Construction Industry Research and Information Association; 1977, p. 209-226.

[17]. Youn BD, Choi KK, Du L. Enriched Performance Measure Approach for Reliability-Based Design Optimization. AIAA Journal 2005; 43(4): 874-884.

[18]. Jiang, C., Han, S., Ji, M. and Han, X. (2015) “A new method to solve the structural reliability index based on homotopy analysis”, Acta Mech, Vol. 226 No. 4, pp.1067-83.

[19]. Keshtegar, B. (2017), “Limited conjugate gradient method for structural reliability analysis”, Eng Computations, Vol. 33 No. 3, pp. 621-29.

[20]. Eshelman L. The CHC adaptive search algorithm: how to safe search when engaging in nontraditional genetic recombination. In: Foundations of Genetic Algorithms 1991 (FOGA1); 1991, p. 265-283.

[21]. Falkenauer E. The worth of the uniform. In: Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406); 1999, p. 776-782.

[22]. Watson RA, Pollack JB. Recombination without respect: schema combination and disruption in genetic algorithm crossover. In: Proceedings of the 2nd Annual Conference on Genetic and Evolutionary Computation; 2000, p. 112-119.

[23]. Conceição António C, Castro CF, Sousa LC. Optimziation of metal forming processes. Comput Struct 2004; 82: 1425-1433.

[24]. Conceição António C. A hierarchical genetic algorithm for reliability based design of geometrically nonlinear composite structures. Comp Struct 2001; 54: 37-47.

[25]. Conceição António C. A multilevel genetic algorithm for optimization of geometrically nonlinear stiffened composite structures. Struct Multidisc Optim 2002; 24: 372-386.

[26]. D'Angelo JP, West DB. Mathematical thinking: problem solving and proofs. 2nd ed. Prentice Hall; 1997.

[27]. Chen J, Ge R, Wei J. Probabilistic optimal design of laminates using improved particle swarm optimization. Eng Optimiz 2008;40(8):695-708.

V
EVOLUTIONARY RELIABILITY-BASED ROBUST DESIGN OPTIMIZATION

Summary

*

Chapter 3 introduced a particular instantiation of the reliability-based robust design optimization problem (RBRDO) of composite laminate structural systems, combining performance optimality, feasibility robustness and probabilistic reliability assessment. Chapter 4 discussed in detail the numerical computation of the reliability constraint and proposed new algorithms for the global optimal solution of local reliability methods, based on evolutionary computation. The current chapter, defines approximate formulations of the proposed RBRDO problem, based on the local solution of the global robustness and reliability measures. Sect. 5.1 proposes the local approximation of the components of the variance-covariance matrix, by the theory of propagation of moments. On Sect. 5.2, the probabilistic reliability constraint is rewritten, in terms of the deterministic reliability measures calculated by the reliability index approach (RIA) and the performance measure approach (PMA). Sect. 5.3 introduces a multi-objective evolutionary algorithm (MOEA) for the solution of the proposed RBRDO problems of composite laminate structural systems.

5.1 Sensitivities-based robustness assessment

We begin our discussion by defining explicitly the integral in (3.2), referring to the component $\mathrm{Cov}(\phi_i, \phi_j)$ of the variance-covariance matrix. Let $\mathbf{x}, \boldsymbol{\pi} \in \Omega_1$ be vectors of random variables defined in a probability space (Ω_1, A, P_1) and $\phi_i : \Omega_1 \to \mathbb{R}$ the i-th stochastic system response functional. The component $\mathrm{Cov}(\phi_i, \phi_j)$ of the *variance-covariance matrix* is given by:

$$\mathrm{Cov}(\phi_i, \phi_j) = \int_{\Omega_1} \left(\phi_i - E(\phi_i)\right)\left(\phi_j - E(\phi_j)\right) p_{\mathbf{x},\boldsymbol{\pi}}(\mathbf{x}, \boldsymbol{\pi}) d\mathbf{x} d\boldsymbol{\pi} \tag{5.1}$$

where $p_{\mathbf{x},\boldsymbol{\pi}}(\mathbf{x}, \boldsymbol{\pi})$ is the joint probability density function (PDF) of the random variables $(\mathbf{x}, \boldsymbol{\pi})$. As with the probability of failure integral, the analytical solution of the previous integral is most likely impossible. For most structural design problems, physical systems are usually numerically simulated and the respective system response functionals implicit. Regarding the numerical solution of (5.1), Monte Carlo simulation (MCS) is a very straightforward method. A comprehensive review on sampling methods for the numerical estimate of variances can be found in the following reference [1]. Although very accurate, sampling methods are computationally expensive and become impractical to be applied recursively in real-scenario structural design optimization problems.

Instead, we opt for the approximate solution of the integrals in (5.1), via the *theory of propagation of moments*. The aforementioned theory provides a systematic methodology for the analytical calculation of the components of the variance-covariance matrix, via the sensitivities (i.e. derivatives) of the system response functionals. It allows to decompose the measures of variance into summands referent to each variable (main effects) or groups of variables (interaction effects), by means of first and higher order sensitives, respectively [2].

The theory of propagation of moments bases its approach on the construction of k-th order Taylor polynomials of the response functionals. Taylor polynomials are written as truncated power series and therefore differentiation and integration can be easily performed summand by summand, allowing for an analytical, yet approximate, solution of the moments of the system response functionals. For a detailed presentation on the subject, including higher-order developments, please refer to [2].

Of particular interest, for the current problem, is the approximation of the system response functionals by first-order Taylor polynomials, for which only the main effect terms are non-null. The following presentation starts by considering only one system response functional and is later generalized for the case of multiple system response functionals.

From (3.1), let the vector of random deviations be written as $\boldsymbol{\delta} = (\Delta\mathbf{x}, \Delta\boldsymbol{\pi})$, where $\Delta\mathbf{x} = (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})$ and $\Delta\boldsymbol{\pi} = (\boldsymbol{\pi} - \boldsymbol{\mu}_\boldsymbol{\pi})$. The first-order Taylor polynomial of the i-th stochastic system response functional ϕ_i, at point $(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi})$, is given by:

$$\begin{aligned} \phi_i(\mathbf{x}, \boldsymbol{\pi}) \quad &\cong \varphi_i(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi}) + \sum_{k_1=1}^{N_\mathbf{x}} \left(\frac{\partial \varphi_i}{\partial x_{k_1}}\right)_{(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi})} \Delta x_{k_1} + \sum_{k_2=1}^{N_\boldsymbol{\pi}} \left(\frac{\partial \varphi_i}{\partial \pi_{k_2}}\right)_{(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi})} \Delta \pi_{k_2} \\ &\cong \varphi_i(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi}) + \sum_{k=1}^{N_\mathbf{x}+N_\boldsymbol{\pi}} s_{i,k}\delta_k \end{aligned} \tag{5.2}$$

where $s_{i,k}$ is the sensitivity of φ_i with respect to k-th random variable, calculated at $(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi})$.

In order to proceed to the calculation of the integral in (5.1), we require a previous consideration. For a vector $\boldsymbol{\delta}$ of random deviations about the mean values, the following result is known:

$$\begin{aligned}\int_{\Omega_1}\left(\sum_k \delta_k\right)^2 dP_1 &= \int_{\Omega_1}\left(\sum_k \delta_k\right)\left(\sum_l \delta_l\right) dP_1 \\ &= \int_{\Omega_1}\left[\sum_i \delta_k\left(\delta_k + \sum_{l\neq k}\delta_l\right)\right] dP_1 \\ &= \int_{\Omega_1}\sum_i {\delta_k}^2\, dP_1 + \int_{\Omega_1}\sum_k\sum_{l\neq k}\delta_k\delta_l\, dP_1 \end{aligned} \tag{5.3}$$

Here, we make the assumption of independence of random variables. Then, the second term of the sum, on the right-hand side of the last equation, is zero and the next equality is written:

$$\int_{\Omega_1}\left(\sum_k \delta_k\right)^2 dP_1 = \sum_k \int_{\Omega_1} {\delta_k}^2 dP_1 \tag{5.4}$$

Another fundamental result in the first-order propagation of moments is that the expected value of ϕ_i is equal to is deterministic realization, that is $E(\phi_i) = \varphi_i(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi})$ [2]. Now, from (5.2) and (5.4) and after some algebra, the integral in (5.1) has the following explicit solution:

$$\mathrm{Cov}(\phi_i, \phi_i) = \mathrm{Var}(\phi_i) = \sum_{k_1=1}^{N_\mathbf{x}} {s_{i,k_1}}^2\, \mathrm{Var}\left(x_{k_1}\right) + \sum_{k_2=1}^{N_\boldsymbol{\pi}} {s_{i,k_2}}^2\, \mathrm{Var}\left(\pi_{k_2}\right) \tag{5.5}$$

representing the first-order approximation of the variance of the i-th system response functional. The previous concepts are readily extended to the case of m system response functionals, dependent on $(\mathbf{x}, \boldsymbol{\pi})$. Hence, the first-order approximation of the component $\mathrm{Cov}(\phi_i, \phi_j)$ of the *variance-covariance matrix* is given by:

$$\mathrm{Cov}(\phi_i, \phi_j) = \sum_{k_1=1}^{N_\mathbf{x}} s_{i,k_1} s_{j,k_1}\, \mathrm{Var}\left(x_{k_1}\right) + \sum_{k_2=1}^{N_\boldsymbol{\pi}} s_{i,k_2} s_{j,k_2}\, \mathrm{Var}\left(\pi_{k_2}\right) \tag{5.6}$$

Therefore, the first-order approximation of the variance-covariance matrix is explicitly defined by:

$$\mathbf{C}_\boldsymbol{\phi} = \mathbf{S}\mathbf{V}\mathbf{S}^T \tag{5.7}$$

where $\mathbf{S}$ is the $(m \times N)$ sensitivities matrix, whose components $s_{i,k}$ are the sensitivities of the i-th system response functional, with respect to the k-th random variable, calculated at $(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi})$, and $\mathbf{V}$ is the diagonal variance matrix of the random variables, with components:

$$v_{k,k} = \begin{cases} {\sigma_{x_k}}^2 & \text{, if } 1 \le k \le N_\mathbf{x} \\ {\sigma_{\pi_k}}^2 & \text{, if } (N_\mathbf{x} + 1) \le k \le N \end{cases} \tag{5.8}$$

where $N = N_\mathbf{x} + N_\boldsymbol{\pi}$.

Higher-order moments of the response functions are usually avoided in practice, as the equations become very complex [2]. As seen in (5.5) and (5.6), the second moment equations depend on the first-order sensitivities of the system response functionals, with respect to the random variables. Hence, these results constitute a first-order local approximation of the global robustness measure with dependence on the expansion point $(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})$ and on the standard deviations of the random variables $(\boldsymbol{\sigma}_{\mathbf{x}}, \boldsymbol{\sigma}_{\boldsymbol{\pi}})$. Hence, we write:

$$\det \mathbf{C}_{\boldsymbol{\phi}} \cong f(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}}, \boldsymbol{\sigma}_{\mathbf{x}}, \boldsymbol{\sigma}_{\boldsymbol{\pi}}) \tag{5.9}$$

5.2 Equivalent deterministic reliability constraints

We began the previous chapter by introducing global and local reliability assessment methods; namely, Monte Carlo simulation (MCS), the reliability index approach (RIA) and the performance measure approach (PMA). If on the one hand, MCS (and related methods) return a direct estimate of the probability of failure, given by (4.5), on the other hand, both the RIA and the PMA only return indirect measures of reliability. In such case, the probabilistic reliability constraint, must be replaced by equivalent inequalities, written in terms of the respective reliability measures. We now recover some useful concepts.

Let $\mathbf{x}, \boldsymbol{\pi} \in \Omega_2$ be random variables defined in an uncertainty space $(\Omega_2, \mathcal{F}, P_2)$ and $g_2 \colon \Omega_2 \to \mathbb{R}$ a stochastic limit-state function, associated with the critical Tsai number of the composite laminate structural system. Furthermore, let $\mathbf{y} = T(\mathbf{x}, \boldsymbol{\pi})$ be the transformed vector of independent random variables, such that $\mathbf{y} \sim N(\mathbf{0}, \mathbf{1})$, and $G_2(\mathbf{y}) = g_2(\mathbf{x}, \boldsymbol{\pi})$ the transformed stochastic limit-state function. The probabilistic reliability constraint, in (3.14), may be written as:

$$\Pr(G_2(\mathbf{y}) < 0) \leq p_f^a \tag{5.10}$$

Although having different meanings, to the optimal solutions of both the RIA and the PMA problems is given the name of *most probable failure point* (MPP), represented by $\mathbf{y}_{\mathbf{MPP}}$.

5.2.1 The reliability index approach equivalent constraint

Both in the RIA and the PMA it is assumed the approximation of $G_2(\mathbf{y})$ by a first-order Taylor polynomial, at $\mathbf{y}_{\mathbf{MPP}}$, which we now recover from (4.10):

$$G_{2_L}(\mathbf{y}) = G_2(\mathbf{y}_{\mathbf{MPP}}) + \sum_{i=1}^{N^{\mathbf{y}}} \left(y_i - y_{MPP_i}\right) \left(\frac{\partial G_2}{\partial y_i}\right)_{\mathbf{y}_{\mathbf{MPP}}} \tag{5.11}$$

It follows that any linear combination of independent standard normal random variables is itself a normal random variable [3], i.e., $\mathbf{y} \sim N(\mathbf{0}, \mathbf{1}) \Rightarrow G_{2_L}(\mathbf{y}) \sim N\left(\mu_{G_{2_L}}, \sigma_{G_{2_L}}\right)$. Then, directly from (4.15), it follows:

$$\begin{aligned} & \Pr(G_2(\mathbf{y}) < 0) \leq p_f^a \\ \Leftrightarrow \quad & \Phi^{-1}(\Pr(G_2(\mathbf{y}) < 0)) \geq \Phi^{-1}\left(p_f^a\right) \\ \Leftrightarrow \quad & \beta_{HL} \geq \beta^a \end{aligned} \tag{5.12}$$

where β^a is the allowable reliability index, referent to the allowable probability of failure p_f^a.

The advantage of such an equivalent reliability constraint is twofold: it is written in terms of a deterministic reliability measure, defined in the standardized uncertainty space, and it only requires the knowledge of the failure surface, hence, being invariant. In practice, design solutions with large β values do not require an exhaustive reliability assessment, because the respective probabilities of failure are so small that the chances of being infeasible approximate zero. Similarly, design solutions that are predictably infeasible, in terms of reliability, do not need to be assessed. With the sole purpose of reducing the number of implicit model evaluations run during the whole RBRDO process, we define a piecewise equivalent reliability constraint, as follows.

Let $g_2(\mathbf{x}, \boldsymbol{\pi}): \Omega_2 \rightarrow \mathbb{R}$ be differentiable at the origin of the uncertainty space. A prior estimate of the actual reliability index is given by:

$$\beta^*(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_{\boldsymbol{\pi}}) = \left(\frac{g_2(\mathbf{x}, \boldsymbol{\pi})}{\sqrt{\sum_{i=1}^{N_x} \left(\frac{\partial g_2}{\partial x_i} \, \sigma_{x_i} \right)^2 + \sum_{i=1}^{N_\pi} \left(\frac{\partial g_2}{\partial \pi_i} \, \sigma_{\pi_i} \right)^2}} \right)_{(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_{\boldsymbol{\pi}})} \tag{5.13}$$

and let β^{sup} be a user-defined parameter. Then, the *piecewise RIA equivalent limit-state function* is defined as follows:

$$\mathcal{g}_2(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_{\boldsymbol{\pi}}) = \begin{cases} \dfrac{\beta_{HL}(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_{\boldsymbol{\pi}})}{\beta^a} - 1 & \text{, if } \beta^a \leq \beta^* \leq \beta^{sup} \\ \dfrac{\beta^*(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_{\boldsymbol{\pi}})}{\beta^a} - 1 & \text{, otherwise} \end{cases} \tag{5.14}$$

where β_{HL} is the solution of problem (4.21).

5.2.2 The performance measure approach equivalent constraint

Contrary to the RIA, we intend to define an equivalent reliability constraint explicitly in terms of $G_2(\mathbf{y_{MPP}})$. With this purpose, we introduce the concept of *quantile function* and derive an alternative and exact formulation of the probabilistic reliability constraint.

Definition 5.1 [4]: Let G_2 be a stochastic limit-state function, whose distribution function $F_{G_2}: \mathbb{R} \rightarrow [0,1]$ is continuous and strictly monotone. The quantile function associated with F_{G_2} is given by:

$$Q(\mathbf{y}, p) = \min\{a \in \mathbb{R}: F_{G_2}(a) \geq p\} \quad = F_{G_2}^{-1}(p) \tag{5.15}$$

From the notion of structural integrity, it follows that [5]:

$$Q(\mathbf{y}, p_f) = F_{G_2}^{-1}(\Pr(G_2(\mathbf{y}) < 0)) = 0 \tag{5.16}$$

Then, from (5.10), it results:

$$Q(\mathbf{y}, p_f^a) \geq 0 \tag{5.17}$$

which is the *exact quantile formulation* of the probabilistic reliability constraint.

To derive an equivalent reliability constraint, based on the solution of the PMA, we consider again the first-order Taylor polynomial of G_2, at $\mathbf{y_{MPP}}$. Thus, combining (5.15) and (5.11), we obtain [5]:

$$Q(\mathbf{y}, p_f^a) \cong \min\{a \in \mathbb{R}: \Pr(G_2(\mathbf{y_{MPP}}) + \nabla^T G_2(\mathbf{y_{MPP}})\Delta\mathbf{y} < a) \geq p_f^a\} \tag{5.18}$$

where $\Delta\mathbf{y} = (\mathbf{y} - \mathbf{y_{MPP}})$, which simplifies to:

$$Q(\mathbf{y}, p_f^a) \cong \min\left\{a \in \mathbb{R}: \Phi\left(\frac{a - \mu_{G_{2_L}}}{\sigma_{G_{2_L}}}\right) \geq p_f^a\right\} \tag{5.19}$$

The last result requires the first two moments of the linearized limit-state function, calculated at $\mathbf{y_{MPP}}$, which are given by the integrals in (4.11) and (4.12). Inserting in (5.19), it results:

$$Q(\mathbf{y}, p_f^a) \cong \min\left\{a \in \mathbb{R}: \Phi\left(\frac{a - G_2(\mathbf{y_{MPP}}) + \nabla^T G_2(\mathbf{y_{MPP}})\mathbf{y_{MPP}}}{\|\nabla G_2\|}\right) \geq p_f^a\right\} \tag{5.20}$$

noting that for the PMA $G_2(\mathbf{y_{MPP}}) \neq 0$. Now rearranging the previous expression, on obtains:

$$Q(\mathbf{y}, p_f^a) \cong \min\{a \in \mathbb{R}: a \geq G_2(\mathbf{y_{MPP}}) - \nabla^T G_2(\mathbf{y_{MPP}})\mathbf{y_{MPP}} + \Phi^{-1}(p_f^a)\|\nabla G_2\|\} \tag{5.21}$$

Taking into consideration the meaning of $\min\{a \in \mathbb{R}\}$, a first-order approximation of the quantile function is obtained, as follows:

$$Q(\mathbf{y}, p_f^a) \cong G_2(\mathbf{y_{MPP}}) - \nabla^T G_2(\mathbf{y_{MPP}})\mathbf{y_{MPP}} + \Phi^{-1}(p_f^a)\|\nabla G_2\| \tag{5.22}$$

Given that $\mathbf{y_{MPP}}$ is the optimal solution of the PMA problem, in (4.16), from the application of the KKT necessary optimality conditions (see Chapter 2), it results the following equality:

$$\nabla^T G_2(\mathbf{y_{MPP}})\mathbf{y_{MPP}} - \Phi^{-1}(p_f^a)\|\nabla G_2\| = 0 \tag{5.23}$$

Thus, from (5.22), it follows that:

$$Q(\mathbf{y}, p_f^a) \cong G_2(\mathbf{y_{MPP}}) \tag{5.24}$$

That is, the first-order approximation of the quantile function, obtained via the first-order Taylor polynomial of $G_2(\mathbf{y})$, is equal to the optimal value of PMA problem, in (4.16). Finally, from (5.17), the PMA equivalent reliability constraint is given the following inequality:

$$G_2(\mathbf{y_{MPP}}) \geq 0 \tag{5.25}$$

Similar to the RIA, design solutions that are predictably infeasible, in terms of reliability, do not need to be assessed. In order to reduce the number of implicit model evaluations run during the whole RBRDO process, we define the *piecewise PMA equivalent limit-state function*, as follows:

$$g_2(\boldsymbol{\mu_x}, \boldsymbol{\mu_\pi}) = \begin{cases} \dfrac{\beta^*(\boldsymbol{\mu_x}, \boldsymbol{\mu_\pi})}{\beta^a} - 1 & \text{, if } \beta^a \leq \beta^* \\ g_2\left(T^{-1}(\mathbf{a_{MPP}})\right) & \text{, otherwise} \end{cases} \tag{5.26}$$

where $\mathbf{a_{MPP}}(\boldsymbol{\mu_x}, \boldsymbol{\mu_\pi})$ is the solution of problem (4.47).

5.3 The bi-level dominance multi-objective genetic algorithm

The proposed algorithm to solve the bi-objective RBRDO problem is the *bi-level dominance multi-objective genetic algorithm* (MOGA-2D), developed by Conceição António and Hoffbauer [6]. The algorithm searches the design space to find multiple Pareto-optimal solutions in parallel, using two simultaneous populations, namely, the short population ($\mathbf{SP}^t$) and the enlarged population ($\mathbf{EP}^t$). It performs using the concept of local dominance at the $\mathbf{SP}^t$ and storing the newly generated nondominated solutions (rank 1) into the $\mathbf{EP}^t$. This enlarged population is continuously updated based on global dominance concepts and has two main functionalities: to build the global Pareto front and to transmit its best solutions' genetic properties to the next small generations of the evolutionary process. The existence of two parallel populations is fundamental for a successful convergence of the method, since it allows to reduce the number of nondominated solutions stored in the $\mathbf{SP}^t$, while keeping that information in the $\mathbf{EP}^t$, and proceeding with the evolutionary search granting diversity and always keeping the best Pareto-optimal front.

The algorithm applies the concept of constrain-dominance, based on the total constrain violation of individuals (see Chapter 2). The total constraint violation of an individual is defined as the sum of the absolute values of the violated constraint functions of the RBRDO problem and is defined by:

$$\xi(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}}) = \Gamma_1 |g_1(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})| + \Gamma_2 |g_2(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})| \tag{5.27}$$

with

$$\Gamma_{\mathrm{i}} = \begin{cases} 0 & \text{, if feasible} \\ 1 & \text{, if infeasible} \end{cases} \tag{5.28}$$

The concept of constrain-dominance is introduced in both the $\mathbf{SP}^t$ and the $\mathbf{EP}^t$ populations, however, with different strategies. In the $\mathbf{EP}^t$, it is applied directly to compare the solutions and to rank them according to their level of dominance (global dominance). Given the size and history of this population, the dominance is applied in a global sense, allowing the progressive construction of the global Pareto front. As the process is continuously applied at every generation, it is possible that an individual with non-dominated status will then be dominated and removed from the evolutionary process. This procedure is independent from the ranking process done within the $\mathbf{SP}^t$. In the $\mathbf{SP}^t$, solutions are ranked according to the concept of shared fitness, based on local constrain-dominance.

Fitness assignment based on local dominance: In the $\mathbf{SP}^t$, solutions are ranked according to their fitness, which no longer depends on an absolute value related to a certain fitness function, but rather on the concept of dominance. The individual fitness is calculated according to the niche occupied by the solution and also depending on the number of individuals with the same level of dominance in its neighborhood. It is called shared fitness. A sharing function is used to improve the distribution of rank 1 solutions along the local Pareto front, at the $\mathbf{SP}^t$ level, during the evolutionary process. Though the elitist strategy adopted in the $\mathbf{SP}^t$ is based on fitness, it is also based on constrain-dominance, implicitly. The procedure to assign a fitness value to each solution is described in a previous work by Conceição António [7].

The evolutionary process of the MOGA-2D comprises a stochastic elitist parent selection mechanism, the uniform crossover operator and a classic elitist strategy coupled with implicit mutation and similarity control. The innovation of the algorithm is the exchange of information between the two populations happening at two levels: first through migration of nondominated solutions from the $\mathbf{SP}^t$ to the $\mathbf{EP}^t$; and second through the recombination of genes between individuals of both populations. A detailed description of these mechanisms is found in Conceição António and Hoffbauer [6] and Conceição António [8]. The algorithmic structure of the MOGA-2D is presented next.

Algorithm 5.1: (*bi-level dominance multi-objective genetic algorithm*):

Initialization of Local Search:
Set $t \coloneqq 1$;
Set size constraints for ply-angle and thickness variables;
Set $\beta_a, \bar{R}_a$ and $\bar{u}_a$.
repeat
for SP^t do
Ranking of solutions based on constrain-dominance;
Fitness assignment based on local dominance;
Pseudo-fitness-based ranking of solutions (implicitly based on dominance);
Definition of the elite group, of the $\boldsymbol{N_{top}}$ fittest individuals.
end do
for EP^t do
Migrate rank 1 solutions from **SP**^{t};
Ranking of solutions based on constrain-dominance.
end do
for SP^{t+1} do
Migrate the elite group from **SP**^{t};
Select pairs of parents from **SP**^{t}
one from the elite and another from the remaining population;
Offspring generation and selection mechanisms ;
Implicit mutation
Complete **SP**$^{t+1}$ *with random generation of new* $\boldsymbol{N_{bot}}$ *individuals;*
Ranking of solutions based on constrain-dominance;
Fitness assignment based on local dominance;
Pseudo-fitness-based ranking of solutions (implicitly based on dominance);
Similarity control, between individuals;
For each pair of similar solutions, eliminate the worst by a random one.
end do
for EP^{t+1} do
Migrate rank 1 solutions from **SP**^{t};
Ranking of solutions based on constrain-dominance.
end do
for SP^{t+2} do
Migrate the elite group from **SP**$^{t+1}$;
Select pairs of parents:
one from the elite of **SP**$^{t+1}$ *and another from* **EP**$^{t+1}$*, with rank* ≤ 3*;*
Offspring generation and selection mechanisms ;
Implicit mutation
Complete **SP**$^{t+2}$ with random generation of new $\boldsymbol{N_{bot}}$ individuals;
Ranking of solutions based on constrain-dominance;
Fitness assignment based on local dominance;
Pseudo-fitness-based ranking of solutions (implicitly based on dominance);
end do
for EP^{t+2} do
Migrate rank 1 solutions from **SP**$^{t+1}$;
Ranking of solutions based on constrain-dominance.
end do
until $t = N_{gen}$

5.4 Concluding notes

On this chapter, approximate formulations of the RBRDO problem of composite laminate structural systems, defined in Chapter 3, are proposed, based on the local approximation of the determinant of the variance-covariance matrix and on the equivalent deterministic formulation of the probabilistic reliability constraint, based on local reliability measures.

The components of the variance-covariance matrix are locally approximated via the theory of propagation of moments. It consists on the construction of Taylor polynomials of the system response functionals, allowing integration to be easily performed summand by summand.

Regarding reliability assessment, both the reliability index approach (RIA) and the performance measure approach (PMA) only return deterministic indirect measures of reliability. In such case, the probabilistic reliability constraint, must be replaced by equivalent inequalities, written in terms of the respective reliability measures. For both methods, the equivalent inequalities are derived after a first-order approximation of the limit-state function, at the respective most probable failure point (MPP), hence, constituting a local approximation of the original probabilistic reliability constraint. Furthermore, with the sole purpose of reducing the number of implicit model evaluations, run during the whole design optimization process, piecewise equivalent reliability constraints are defined for both the RIA and the PMA, based on an analytic first-order estimate of the reliability index of the design solutions.

The proposed algorithm to solve the bi-objective RBRDO problem is the bi-level dominance multi-objective genetic algorithm (MOGA-2D), developed by António and Hoffbauer [6]. The algorithm searches the design space to find multiple Pareto-optimal solutions in parallel, using two simultaneous populations, a local population and an enlarged (archive) population. Contrary to the classic population-based multi-objective evolutionary algorithms, the MOGA-2D promotes the interaction of both populations, during the evolutionary process, in order to increase its efficiency towards the Pareto front. Furthermore, each population has its own sorting mechanism, based on the concept of constrain-dominance. In the local population, solutions are sorted according to a fitness assignment process, based on local dominance. On the other hand, in the enlarged population individuals are sorted according to the concept of global constrain-dominance.

The evolutionary process of the MOGA-2D comprises a stochastic elitist parent selection mechanism, the uniform crossover operator and a classic elitist strategy coupled with implicit mutation and similarity control. The innovation of the algorithm is the exchange of information between the two populations. The complete algorithmic structure of the MOGA-2D is presented.

References

[1]. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M and Tarantola S. Variance based sensitivity analisys of a model output. Design and estimator for the total sensitivity index. Computer Physiscs Communications. 2010; 181: 259-70.

[2]. Cacuci DG. Sensitivity and uncertainty analysis: theory. 1st ed. Chapman and Hall/CRC; 2003

[3]. Melchers RE. Structural Reliability Analysis and Prediction. 2nd ed. Willey & Sons; 1999.

[4]. Gilchrist W. Statistical modelling with quantile functions. 1st ed. Chapman and Hall/CRC; 2000.

[5]. Chiralaksanakul A, Mahadevan S. Decoupled approach to multidisciplinary design optimization under uncertainty. Optim Eng 2007; 8: 21-42.

[6]. António CC, Hoffbauer LN. Bi-level dominance GA for minimum weight and maximum feasibility robustness of composite structures. Compos Struct 2016; 135: 83-95.

[7]. Conceição António C. Local and global Pareto dominance applied to optimal design and material selection of composite structures. Struct Multidisc Optim 2013; 48: 73-94.

[8]. Conceição António C. A study on synergy of multiple crossover operators in a hierarchical genetic algorithm applied to structural optimization. Struct Multidisc Optim 2009; 38: 117-135.

VI
SOBOL' INDICES AS ANALYTICAL DIMENSIONAL REDUCTION TECHNIQUE

Summary

*

In the reliability assessment of composite laminate structures with multiple components, the uncertainty space defined around design solutions easily becomes over-dimensioned and not all of the random variables are relevant. In this chapter, the importance analysis theory of Sobol' is implemented to reduce the dimension of the uncertainty space, improving the efficiency towards global convergence of evolutionary-based reliability assessment. On Sect. 6.1., the importance analysis theory of Sobol' is introduced. Sect. 6.2 begins with the exposition of two fundamental results relating k-th order multilinear Taylor polynomials of multivariate stochastic functions of independent random variables with its own ANOVA functional decomposition. Then, Sobol' indices are formulated analytically for implicit structural response functions, following the theory of propagation of moments and without violating the fundamental principles presented by Sobol'. On its application to implicitly defined structural systems, the proposed formulation has the advantage of only requiring one adjoint system of equilibrium equations to be solved once. On Sect. 6.3, an analytical dimensional reduction criterion is presented. The theoretical developments assume independent random variables.

6.1 The importance analysis theory of Sobol'

Importance analysis concerns about the quantification of the effects of variation of the input variables upon the response functionals of a system. From a set of data, it allows to identify those parameters which are relevant to the construction, or the characterization, of a mathematical model and those which are not. On its turn, uncertainty analysis allows to infer how well a model predicts the observed data. In practice, the exact forms of mathematical models are unknown, as well as the complete characterization of the uncertainty existent on the observed/estimated data [1].

In the reliability assessment of structural systems, a proper characterization of the uncertainty space is vital for the accuracy and efficiency of the methods. In constructing the uncertainty space, defined around a particular deterministic design solution, it is understood that if it is too large the computing costs will be high and the action of the applied numerical methods may be compromised.

By combining importance analysis with uncertainty analysis, it is possible to achieve a simplification of the uncertainty space, by identifying which variables contribute the most to the variation of the system response functionals and by excluding or paralyzing the less important variables [2]. Thus, in the context of importance analysis, it is fundamental to give a proper definition of *importance*.

Definition 6.1: Let $(\Omega_2, \mathcal{F}, P_2)$ be an uncertainty space and $g_2: \Omega_2 \to \mathbb{R}$ a stochastic limit-state function. *Importance* is the concept measuring the reduction in the variance of g_2, if one or more variables are fixed on a given value.

From this definition, a variable is deemed as important only if it is responsible for a considerable fraction of the total variance of the system. Otherwise, it is not. The meaning of "a considerable fraction" is not rigid and will be addressed later, in Sect. 6.3. Thus, it seems a natural choice that variance-based importance measures are considered to execute importance analysis. These measures are global, in the sense that they assess importance as a characteristic of an entire domain, rather than pointwise.

Remark: For the sake of simplicity, we represent $g_2(\mathbf{x}, \boldsymbol{\pi})$ as $g(\mathbf{x})$, collapsing the $(\mathbf{x}, \boldsymbol{\pi})$ into $\mathbf{x}$, such that $N = N_{\mathbf{x}} + N_{\boldsymbol{\pi}}$ is the total number of random variables, and cancelling the subscript numeral.

Let $\mathbf{x} \in \Omega_2$ be a vector of random variables defined in an uncertainty space $(\Omega_2, \mathcal{F}, P_2)$ and consider a stochastic limit-state function $g: \Omega_2 \to \mathbb{R}$. Furthermore, assume g is integrable over Ω_2 with respect to the probability measure P_2 and that $dP_2(\mathbf{x}) = \prod_{k=1}^{N} p_{x_k}(x_k) dx_k$ (meaning independence between the random variables), where $p_{x_k}(x_k)$ is the marginal PDF of x_k. Then, the limit-state function can be expanded as [3]:

$$g(\mathbf{x}) = g_0 + \sum_{i=1}^{N} g_i + \sum_{i<j}^{N} g_{i,j} + \cdots + g_{1,2,\ldots,N} \tag{6.1}$$

such that the integral of each summand $g_{i,j,\ldots,l} = g(x_i, x_j, \ldots, x_l)$, for $1 \leq i < j < \cdots < l \leq N$, with respect to any of their own variables is null; that is:

$$\int_{\Omega_2} g_{i,j,\ldots,l} \, dP_2(x_k) = 0 \quad , i \leq k \leq l \tag{6.2}$$

From the previous definition, it results that g_0 is constant and equal to the expectancy of $g(\mathbf{x})$, defined by:

$$g_0 = \int_{\Omega_2} g(\mathbf{x})\, dP_2(\mathbf{x}) \tag{6.3}$$

and all the summands of the expansion are orthogonal, with respect to the following inner product:

$$\int_{\Omega_2} g_{i_1,j_1,\dots,l_1}\, g_{i_2,j_2,\dots,l_2}\, dP_2(\mathbf{x}) = 0 \tag{6.4}$$

if $\{i_1, j_1, \dots, l_1\} \neq \{i_2, j_2, \dots, l_2\}$. Thus, for a generic set of indices $\{i_1, \dots, i_s\}$ associated with the random variables, the expansion is unique and each summand in (6.1) is given by:

$$g_{i_1,\dots,i_s} = \int_{\Omega_2} g(\mathbf{x}) \prod_{k \neq i_1,\dots,i_s} dP_2(x_k) - \sum_{i=i_1}^{i_s} g_i - \sum_{i<j=i_1}^{i_s} g_{i,j} - \cdots - g_0 \tag{6.5}$$

noting that, on the right-hand side, only summands of dimension lower than that of $g_{i_1,\dots,i_s}$ exist.

This functional decomposition is known as the ANOVA decomposition. Now, adding the assumption that $g(\mathbf{x})$ is square-integrable, the following ratios can be defined:

$$S_{i_1,\dots,i_s} = \frac{\mathrm{Var}\left(g_{i_1,\dots,i_s}\right)}{\mathrm{Var}\left(g(\mathbf{x})\right)} = \frac{\int_{\Omega_2} \left(g_{i_1,\dots,i_s}\right)^2 dP_2(\mathbf{x})}{\int_{\Omega_2} g(\mathbf{x})^2\, dP_2(\mathbf{x}) - {g_0}^2} \quad \in [0,1] \tag{6.6}$$

as global importance measures. These ratios explain the total variance $\mathrm{Var}\left(g(\mathbf{x})\right)$ as a contribution of the partial variances $\mathrm{Var}\left(g_{i_1,\dots,i_s}\right)$, associated with each summand, in (6.1). That is, they measure the reduction in the total variance of the system, induced by freezing the random variables, of indices $i_1, \dots, i_s$ [4].

The indices (6.6) are named after Sobol' [4] who first derived them. The indices, as derived by Sobol', are not necessarily a measure of uncertainty *per se*, but rather a deterministic measure of importance, given that they were originally formulated in the N-dimensional unit cube. In practice, however, this is equivalent to the case of independent uniformly distributed variables on the interval $[0,1]$, and their interpretation as the quotient between variances still holds. This way, the definitions presented in (6.1)-(6.6) are a generalization of the Sobol' indices to any probability space of independent random variables.

As with the components of the variance-covariance matrix (see Sect. 5.1), Sobol' indices are extremely difficult to calculate, particularly for large-scale physical problems, with implicit system response functionals. In the literature, there are several proposals for the numerical estimate of the indices, usually based on sampling methods. Please, see [4-6]. To avoid exhaustive sampling methods, the theory of propagation of moments is applied to the Sobol' indices, which not only proves to be useful, but also adequate. An issue discussed next.

6.2 Propagation of moments applied to Sobol' indices

The theory of propagation of moments provides a systematic methodology combining UA with IA, via the sensitivities of the system responses in relation to their (random) variables. It allows to decompose the measures of uncertainty into summands referent to each variable (main effects) or groups of variables (interaction effects), by means of first or higher order sensitivities (derivatives) [1].

While the measures defined in such a way are of local character and depend on the point where sensitivities are calculated, they have the ability to offer a pointwise quantification of the uncertainty measures. On the other hand, global measures will homogenize those quantities throughout the domain (see the integral definition of statistical measures), i.e., the scalar quantities provided by global measures are constants describing the overall/average phenomena on the sample space. They may be understood as the expectancy of certain functional measures. For example, given a square-integrable function f and its expected value $\mathrm{E}(f)$, the variance of f may be seen as the expectancy of the dispersion functional $\left(f - \mathrm{E}(f)\right)^2$. However, locally, these measures are allowed to deviate from their global expectations and, in some cases, the local assessment of the uncertainty measures may be appropriate.

Considering the reliability assessment problem, as defined by the reliability index approach (RIA), in (4.8), it may be interpreted as a quasi-local problem, focusing its action in the area of the uncertainty space around the origin. Thus, it seems coherent to consider local information, at least, as useful as global information, because while it is only valid on the point at which it is calculated, it may still be relevant in a neighbourhood of such point, indicating a trend.

The theory of propagation of moments bases its approach on the construction of k-th order Taylor polynomials of the response functionals. Taylor polynomials are written as truncated power series and therefore differentiation and integration can be easily performed summand by summand, allowing for an analytical, yet approximate, solution of the statistical moments of functions of random response functionals.

In the context of the application of the theory of propagation of moments to Sobol' indices, we propose two fundamental results relating k-th order multilinear Taylor polynomials with the ANOVA functional decomposition of multivariate stochastic functions of independent random variables.

Theorem 6.1: Let $\mathbf{x} \in \Omega_2$ be a vector of N independent random variables defined in an uncertainty space $(\Omega_2, \mathcal{F}, P_2)$ and assume $g: \Omega_2 \to \mathbb{R}$ is integrable over Ω_2 with respect to the probability measure P_2 and is differentiable at $\boldsymbol{\mu}_\mathbf{x}$. Then, the N-th order Taylor polynomial of $g(\mathbf{x})$ is coincident with <u>its own</u> ANOVA decomposition if and only if it is multilinear.

Proof: The proof consists in showing that non-linear Taylor polynomials are not coincident with their own ANOVA decomposition. Start by considering a complete N-th order Taylor polynomial of g at point $\boldsymbol{\mu}_\mathbf{x}$, written in compact notation as follows:

$$g(\mathbf{x}) \cong\ g(\boldsymbol{\mu}_\mathbf{x}) + \sum_{|\alpha|=1}^{N} \frac{1}{\alpha!} \left(\frac{\partial^{|\alpha|} g}{\prod_{i=1}^{N} \partial {x_i}^{\alpha_i}} \right)_{\boldsymbol{\mu}_\mathbf{x}} \prod_{i=1}^{N} \Delta {x_i}^{\alpha_i} \tag{6.7}$$

where α is a multi-index, such that $|\alpha| = \alpha_1 + \cdots + \alpha_N$ and $\alpha! = \alpha_1 \ldots \alpha_N$, $\forall \alpha_i \geq 0$, and $\sum_{|\alpha|=1}^{N}(\cdot)$ represents all the summations existent for each value of $|\alpha|$. From (6.3), it results:

$$g_0 = g(\boldsymbol{\mu}_\mathbf{x}) + \sum_{|\alpha|=2}^{N} \frac{1}{\alpha!} \left(\frac{\partial^{|\alpha|} g}{\prod_{i=1}^{N} \partial {x_i}^{\alpha_i}} \right)_{\boldsymbol{\mu}_\mathbf{x}} \prod_{i=1}^{N} \mu_{\alpha_i}(x_i) \tag{6.8}$$

where $\mu_{\alpha_i}(x_i)$ is the α_i-order central moment of x_i, with respect to the probability measure. This result represents an approximation to the expected value of g, truncated to the N-th order, and the summands are corrective terms considering local gradient information and the distributional parameters of the random variables.

By inserting (6.8) with (6.5) and (6.1), the summands on the right-hand side of (6.8) are eventually cancelled out, but, by definition, these are non-null and are necessary to satisfy the orthogonality condition stated in (6.4). Thus, the summands of the N-th order Taylor polynomial of g, in (6.7), will be equal to the summands of its own ANOVA decomposition, if the mixed partial derivatives with $\alpha_i > 1$, for all $i = 1, \dots N$, are null. That is, the summands on (6.1) and (6.7) are equal if the Taylor polynomial is multilinear. ■

This subject is also mentioned, by other means, in [7]. The multilinearity assumption implies that the remaining higher-order derivatives are null, such that, the main effects of the ANOVA decomposition, g_i, are associated only with the first-order partial derivatives and the interaction effects, $g_{i,j,k,\dots}$, with the mixed partial derivatives for which α_i is either 1 or 0, for all $i = 1, \dots, N$. Thus, a multilinear Taylor polynomial of g is exact only if g is itself multilinear. Otherwise, it constitutes an approximation, which is only valid in a neighbourhood of $\boldsymbol{\mu}_\mathbf{x}$. Such local approximation may still be reasonable, if the function g is locally smooth, as the first-order terms will dominate over higher-order terms [7].

From Theorem 6.1, we propose a broader statement, as follows.

Corollary 6.1: In an uncertainty space of N independent random variables, any multilinear k-th order Taylor polynomial is coincident with its own ANOVA decomposition, for all $k \leq N$.

Proof: Consider the following *multilinear* k-th order Taylor polynomial of g, over N variables, written in compact notation:

$$g(\mathbf{x}) \cong\ g(\boldsymbol{\mu}_\mathbf{x}) + \sum_{|\alpha|=1}^{k} \left(\frac{\partial^{|\alpha|} g}{\prod_{i=1}^{|\alpha|} \partial x_i} \right)_{\boldsymbol{\mu}_\mathbf{x}} \prod_{i=1}^{|\alpha|} \Delta x_i \tag{6.9}$$

where $|\alpha| = \sum_{i=1}^{N} \alpha_i$, and each α_i is either 1 or 0. Let $I = \{i_1, \dots, i_N\}$ be a set of indices, in ascending order (by which we mean $i_s < i_p \Leftrightarrow s < p$). Then, consider subsets $I' \subseteq I$, with cardinality $\#I' \coloneqq \alpha_{I'} \leq k \leq N$, containing combinations of $\alpha_{I'}$ ordered indices and, similarly, subsets $I^c \subseteq I$, such that $\#I' <\ \#I^c \leq N$. Then, from Theorem 6.1, by combining (6.9) with (6.5) it results a constant term:

$$g_0 =\ g(\boldsymbol{\mu}_\mathbf{x}) \tag{6.10}$$

plus a set of non-null summands:

$$g_{I'} = \left(\frac{\partial^{\alpha_{I'}} g}{\prod_{i \in I'} \partial x_i} \right)_{\boldsymbol{\mu}_\mathbf{x}} \prod_{i \in I'} \Delta x_i \tag{6.11}$$

which are orthogonal with respect to the probability measure, and also a set of null summands:

$$g_{I^c} = 0 \tag{6.12}$$

hence, orthogonal with all other summands with respect to the probability measure. Since any k-th Taylor polynomial, over N variables, is one whose partial derivatives of order higher than k are null, then the ANOVA summands in (6.10)-(6.12) are coincident with the summands in (6.9). ■

The last theorem and corollary just demonstrate that the local approximation of a function by multilinear Taylor polynomials allows to transform the integral ANOVA decomposition into an equivalent differential decomposition, valid within a small neighbourhood. These results demonstrate that the complex integration process of the ANOVA decomposition is not required, since multilinear Taylor polynomials already satisfy the necessary condition in (6.2).

The ANOVA decomposition is now written in polynomial form, allowing for an analytical solution of the Sobol' indices in (6.6). First, the variance of each ANOVA summand is immediate, assuming independent random variables, and is given by:

$$\begin{aligned}
\mathrm{Var}(g_{I'}) &= \int_{\Omega_2} \left[\left(\frac{\partial^{\alpha_{I'}} g}{\prod_{i \in I'} \partial x_i} \right)_{\boldsymbol{\mu}_\mathbf{x}} \prod\nolimits_{i \in I'} \Delta x_i \right]^2 dP_2 \\
&= \left(\frac{\partial^{\alpha_{I'}} g}{\prod_{i \in I'} \partial x_i} \right)^2_{\boldsymbol{\mu}_\mathbf{x}} \int_{\Omega_2} \prod\nolimits_{i \in I'} {\Delta x_i}^2 \, dP_2 \\
&= \left(\frac{\partial^{\alpha_{I'}} g}{\prod_{i \in I'} \partial x_i} \right)^2_{\boldsymbol{\mu}_\mathbf{x}} \prod\nolimits_{i \in I'} \mathrm{Var}(x_i)
\end{aligned} \tag{6.13}$$

On its turn, the calculation of the total variance of g, $\mathrm{Var}\big(g(\mathbf{x})\big)$, requires a previous consideration. For a set of random increments $\Delta\mathbf{x} = (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x})$, of independent random variables, about the mean value, the following result is known:

$$\begin{aligned}
\int_{\Omega_2} \left(\sum_i \Delta x_i \right)^2 dP_2 &= \int_{\Omega_2} \left(\sum_i \Delta x_i \right) \left(\sum_j \Delta x_j \right) dP_2 \\
&= \int_{\Omega_2} \left[\sum_i \Delta x_i \left(\Delta x_i + \sum_{j \neq i} \Delta x_j \right) \right] dP_2 \\
&= \int_{\Omega_2} \sum_i {\Delta x_i}^2 \, dP_2 + \int_{\Omega_2} \sum_i \sum_{j \neq i} \Delta x_i \Delta x_j \, dP_2
\end{aligned} \tag{6.14}$$

Since the assumption of independence between the random variables is made, then the second term of the sum, on the right-hand side of the last equation, is zero and the next equality is written:

$$\int_{\Omega_2} \left(\sum_i \Delta x_i \right)^2 dP_2 = \sum_i \int_{\Omega_2} {\Delta x_i}^2 dP_2 \tag{6.15}$$

From this result, and after some algebra, it follows that, if g is approximated by a multilinear k-th order Taylor polynomial, over N independent random variables, its variance is given by:

$$\begin{aligned}\mathrm{Var}\big(g(\mathbf{x})\big) &= \mathrm{E}(g(\mathbf{x})^2) - \mathrm{E}\big(g(\mathbf{x})\big)^2 \\ &= \sum_{|\alpha|=1}^{k} \left(\frac{\partial^{|\alpha|} g}{\prod_{i=1}^{|\alpha|} \partial x_i} \right)_{\boldsymbol{\mu}_\mathbf{x}}^2 \prod_{i=1}^{|\alpha|} \mathrm{Var}(x_i)\end{aligned} \tag{6.16}$$

remembering that $\sum_{|\alpha|=1}^{k}(\cdot)$ is the compact notation for all the summations existent on each value of $|\alpha|$ and that, since the polynomial is multilinear, α_i is either 1 or 0, for $i = 1, \ldots, N$. Comparing (6.16) and (6.13), it is seen that $\mathrm{Var}\big(g(\mathbf{x})\big)$ is equal to the summation of the variances of all the decomposition terms, i.e.:

$$\mathrm{Var}\big(g(\mathbf{x})\big) = \sum\nolimits_{I'} \mathrm{Var}(g_{I'}) \tag{6.17}$$

Finally, after (6.6), the Sobol' indices associated with the set of indices I' are defined as follows:

$$S_{I'} = \frac{\left(\frac{\partial^{\alpha_{I'}} g}{\prod_{i \in I'} \partial x_i} \right)_{\boldsymbol{\mu}_\mathbf{x}}^2 \prod_{i \in I'} \mathrm{Var}(x_i)}{\sum_{|\alpha|=1}^{k} \left(\frac{\partial^{|\alpha|} g}{\prod_{i=1}^{|\alpha|} \partial x_i} \right)_{\boldsymbol{\mu}_\mathbf{x}}^2 \prod_{i=1}^{|\alpha|} \mathrm{Var}(x_i)} \quad \in [0,1] \tag{6.18}$$

or

$$S_{I'} = \frac{\mathrm{Var}(g_{I'})}{\sum_{I'} \mathrm{Var}(g_{I'})} \quad \in [0,1] \tag{6.19}$$

The proposed analytical solution of the Sobol' indices allows for different degrees of polynomial approximation. In practice, in implicit structural design problems, lower-order approximations are often preferred, because differential analysis is demanding for implicit response functions. Of particular interest, for the current problem, is the approximation by a first-order Taylor polynomial, for which only the main effect indices are non-null, as follows:

$$S_{x_i} = \left(\frac{\partial g}{\partial x_i} \right)_{\boldsymbol{\mu}_\mathbf{x}}^2 \frac{\mathrm{Var}(x_i)}{\mathrm{Var}(g)} \tag{6.20}$$

which turns out to be coincident with the importance index proposed by Helton [8].

6.3 Analytical dimensional reduction

Many of the proposed methods in the literature for space dimensional reduction are of numerical nature and may be associated with relevant computational costs, in spite of their accuracy (Carreira-Perpiñán, 1997). In the context of structural design optimization, a field where reliability assessment is commonly applied, such costs associated with the dimensional reduction of the uncertainty space of each design solution may become excessive in the overall optimization process. The purpose of the theoretical results presented above is to conjugate the analytical solution of Sobol' indices with structural reliability assessment, as a straightforward dimensional reduction technique of the uncertainty space, requiring only one extra adjoint system of equations to be solved (see Chapter 7).

Now, we formalize a reliability assessment dimension reductional criterion, satisfying Definition 6.1.

Definition 6.2: Let $I = \{i_1, \dots, i_N\}$ be a set of indices. A random variable x_{i_k}, with $i_k \in I$, defined in the uncertainty space $(\Omega_2, \mathcal{F}, P_2)$, is said to be important if the sum of all Sobol' indices with respect to x_{i_k} is greater than or equal to a nonnegative scalar ε, that is:

$$S_{i_k} + \sum_{j \in I} S_{i_k,j} + \sum_{j<l \in I} S_{i_k,j,l} + \cdots + S_{i_1,\dots,i_N} \geq \varepsilon \tag{6.21}$$

otherwise it can be fixed.

Assuming that some random variables are *not important*, then we define the reduced uncertainty space for reliability assessment, as follows:

Definition 6.3: The set $\Omega_r \subseteq \Omega_2$, of all the $n_r \leq N$ important random variables, is called *reduced sample space* and defines the *reduced uncertainty space* $(\Omega_r, \mathcal{F}_r, P_r)$, where $\mathcal{F}_r$ is a σ-algebra on Ω_r and P_r the respective probability measure.

It naturally comes that if ε is correctly chosen, then it is expected that $P_r \cong P_2 \Rightarrow p_{f_r} \cong p_f$ and also $\beta_{HL_r} \cong \beta_{HL}$, where p_{f_r} is the probability of failure of a design solution, calculated in the reduced uncertainty space, and β_{HL_r} the respective Hasofer-Lind reliability index.

6.4 Concluding notes

In the reliability assessment of composite laminate structures with multiple components, the uncertainty space defined around design solutions easily becomes over-dimensioned and not all the random variables are relevant. In this chapter, the importance analysis theory of Sobol' is implemented to reduce the dimension of the uncertainty space, improving the efficiency towards global convergence of evolutionary-based reliability assessment.

An initial conceptual definition of *importance* was given, where a variable is deemed as important only if it is responsible for a considerable fraction of the total variance of the structural system. To comply with the above conceptual definition of importance, the importance analysis theory of Sobol' indices is introduced, providing appropriate variance-based importance measures, called Sobol' indices.

The difficulty to calculate exact values of the Sobol' indices is on par with their own complex integral definition. Many authors have proposed several sampling methods to produce accurate estimates of the indices. However, as with the probability of failure integral and the components of the variance-covariance matrix, the numerical estimate of Sobol' indices by sampling methods is not suitable to be integrated as an inner cycle in structural design optimization problems, of complex structural systems, like the one defined in Chapters 3-5.

In this chapter, it is therefore proposed the application of the theory of propagation of moments, to perform the analytical solution of Sobol' indices, based on the local approximation of the limit-state function. The theoretical developments are built over two fundamental results relating the k-th order multilinear Taylor polynomial of multivariate stochastic functions of independent random variables with its own ANOVA functional decomposition. These results demonstrate that the complex integration process of the ANOVA decomposition is not required.

Finally, a straightforward dimensional reduction criterion of the uncertainty space, for reliability assessment is proposed. Such criterion may as well be applied to other areas with different probability spaces, although it was not explored.

References

[1]. Cacuci DG. Sensitivity and uncertainty Analysis: theory. 1st ed. Chapman and Hall/CRC; 2003.

[2]. Saltelli A, Ratto M, Tarantola S, Campolongo F. Sensitivity analysis practices: Strategies for model-based inference. Reliab Eng Syst Safe 2006; 91 (10-11): 1109-25.

[3]. Efron B, Stein C. The Jackknife Estimate of Variance. Ann Stat 1981; 9(3): 586-96.

[4]. Sobol IM. Sensitivity Estimates for Nonlinear Mathematical Models. MMCE 1993; 1(4): 407-14.

[5]. Conceição António C, Hoffbauer LN. From local to global importance measures of uncertainty propagation in composite structures. Compos Struct 2008; 85: 213-25.

[6]. Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S. Variance based sensitivity analisys of a model output. Design and estimator for the total sensitivity index. Comput Phys Commun 2010; 181: 259-70.

[7]. Borgonovo E. Sensitivity Analysis: An Introduction for the Management Scientist, Springer International Publishing; 2017.

[8]. Helton JC. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliab Eng Syst Safe 1993; 42: 327-67.

VII
APPLICATION TO COMPOSITE LAMINATE STRUCTURES

Summary

*

On this chapter, the RBRDO problem of composite laminate structures defined in Chapter 3 is solved by the algorithms proposed in Chapters 4 and 5. The developments assume the local approximation of the global robustness and reliability measures, based on gradient information of the stochastic response functionals of structural systems. However, the numerical estimate of derivatives of multivariate implicit response functionals is expensive, particularly in high dimensional problems. For that reason, Sect 7.1 introduces the adjoint variable method for the analytical estimate of gradient vectors of the stochastic response functionals. This method only requires an adjoint system of equilibrium equations to be solved, reducing the computational cost of differentiation. Sect 7.2 begins by describing how uncertainty is propagated differently in robustness and reliability assessment and the local RBRDO problem is defined properly. Then, in Sect 7.2.1, a numerical physical model is introduced. In Sect 7.2.2, the local RBRDO problem is solved. The outer cycle of bi-objective design optimization is solved by the MOGA-2D. The inner cycle of reliability assessment (PMA) is solved by the mGA. The validation of the mGA is integrated in the design optimization process. Sect. 7.2.3, presents the solution of the local RBRDO problem, with reliability assessment executed by the RIA. The outer cycle of design optimization is solved by the MOGA-2D and the inner-cycle of reliability assessment by the HmGA. The HmGA is first validated against two alternative methods. The RBRDO results are presented after. Sect 7.3 presents a study on the effects of different sources of uncertainty in feasibility robustness. A simplified fuselage-like shell structure is introduced as numerical example. The outer cycle of design optimization is solved by the MOGA-2D. Reliability assessment is performed by the PMA and solved by the mGA. Lastly, in Sect. 7.4, the analytical dimensional reduction of the uncertainty space based on the analytical solution of the Sobol' indices, developed in Chapter 6, is applied to the HmGA as a powerful and simple tool to improve significantly the performance of the algorithm. Both in terms of accuracy and efficiency. The results of the importance analysis of the uncertainty space, in reliability assessment, are also discussed.

7.1 The adjoint variable method

On Chapter 5, the global robustness measure is solved by the theory of propagation of moments, based on first-order Taylor polynomials. Such approximated formulation requires the calculation of the sensitivities (derivatives) of the displacement and stress system response functionals, at the mean-values of the random variables. On Chapter 4, local formulations of the global probabilistic reliability measure are presented. Although neither the reliability index approach (RIA) nor the performance measure approach (PMA) require the use of sensitivities of the limit-state function, the developed evolutionary algorithms make good use of the available sensitivities to calculate an initial estimate of the β-value of design solutions and improve the efficiency of the reliability assessment process: it is used by the MOGA-2D as a differentiating factor to choose which design solutions need a refined reliability assessment and which don't, saving computational resources, and by the HmGA, during the RIA inner cycle, at several levels as knowledge of the problem. Overall, the use of sensitivities does not harm the efficiency of the reliability assessment inner cycle, as this is information already available from robustness assessment, calculated only once for each design solution. On Chapter 6, Sobol' indices are solved by theory of propagation of moments, based on k-th order multilinear Taylor polynomials. A first-order approximation is chosen, requiring the sensitivities of the limit-state function, calculated at the mean values of the design solutions. A criterion for the dimensional reduction of the uncertainty space, in reliability assessment, is then presented. The integration of this criterion on the reliability assessment cycle is immediate, because, again, the sensitives are already available from robustness assessment. Nevertheless, sensitivities must be calculated once for each design solution generated by the MOGA-2D, during the RBRDO evolutionary process.

Any instantiation of the finite differences method provides a straightforward calculation of the sensitives of the system response functionals, with both explicit and implicit dependence on its variables. However, the number of model evaluations required is proportional to the number of variables of the structural system, losing its efficiency for high-dimensional problems, such as the one proposed. As an alternative, the *adjoint variable method* is a general semi-analytical method to compute the gradient of system response functionals with explicit dependence on a set of state variables and with explicit and implicit dependence on a set of design variables. Numerically, this approach only requires an extra system of equations to be solved, defining an adjoint state-of-equilibrium. Arora and Cardoso [1] proposed a formulation the method in connection with structural analysis, based on the principles presented in [2,3]. The version proposed by Conceição António [4], adapted to the Finite Elements Analysis of composite laminate structures, is used here.

Remark: For the sake of simplicity, we represent a generic system response functional $\phi(\mathbf{x}, \boldsymbol{\pi})$ as $\phi(\mathbf{x})$, collapsing the $(\mathbf{x}, \boldsymbol{\pi})$ into $\mathbf{x}$, such that $N = N_{\mathbf{x}} + N_{\boldsymbol{\pi}}$ is the total number of random variables.

Let U represent the state space, of state variables $\mathbf{u}(\mathbf{x})$, and Ω a sample space of variables $\mathbf{x}$. Let $H: U \times \Omega \to \mathbb{R}$ be a functional with both implicit and explicit dependence on $\mathbf{x}$, such that, $\phi(\mathbf{x}) = H(\mathbf{u}(\mathbf{x}), \mathbf{x})$, and $\Psi: U \times \Omega \to U$ be a mapping representing the state of physical systems. In this work, it is considered the linear equilibrium state of structures under static loading conditions (in agreement with (3.4)), such that:

$$\Psi(\mathbf{u}(\mathbf{x}), \mathbf{x}) = 0 \Leftrightarrow \mathbf{K}\mathbf{u} - \mathbf{f} = \mathbf{0} \tag{7.1}$$

where $\mathbf{K} \equiv \mathbf{K}(\mathbf{x})$ is the stiffness matrix, $\mathbf{f}$ is the vector of external loads. The solution of the equilibrium equation, $\mathbf{u}^*$, is unique for any realization in the sample space, $\mathbf{x}^*$, and so, since H originates from any variational principle, for a given design $\mathbf{x}^*$, the state $\mathbf{u}^*$ is also the solution of the following inverse problem [5]:

$$\min_{\mathbf{u}} \quad H(\mathbf{u}(\mathbf{x}), \mathbf{x})$$

$$\text{subject to} \quad \mathbf{Ku} - \mathbf{f} = \mathbf{0} \tag{7.2}$$

The Lagrangian function $\mathcal{L}: U \times \Lambda \times \Omega \to \mathbb{R}$ associated with the minimization problem is defined as follows:

$$\mathcal{L}(\mathbf{u}(\mathbf{x}), \mathbf{x}, \boldsymbol{\lambda}) = H(\mathbf{u}(\mathbf{x}), \mathbf{x}) - \boldsymbol{\lambda}^{\boldsymbol{T}}[\mathbf{Ku} - \mathbf{f}] \tag{7.3}$$

where $\boldsymbol{\lambda} \in \Lambda$ is vector of Lagrange multipliers, also known as the adjoint state variables, and Λ is the dual state space. Then, $(\mathbf{u}^*, \boldsymbol{\lambda}^*)$ is a stationary point of $\nabla\mathcal{L}$, thus verifying the following optimality conditions:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{u}} = \frac{\partial H}{\partial \mathbf{u}} - \boldsymbol{\lambda}^{\boldsymbol{T}} \frac{\partial}{\partial \mathbf{u}}[\mathbf{Ku} - \mathbf{f}] = \mathbf{0} \tag{7.4}$$

and

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}} = \mathbf{0} \tag{7.5}$$

Notice that $\partial\mathcal{L}/\partial\mathbf{x}$ is not necessarily null, because $\mathbf{x}^*$ is not necessarily the vector that minimizes $\mathcal{L}(\mathbf{u}(\mathbf{x}), \mathbf{x}, \boldsymbol{\lambda})$, or $H(\mathbf{u}(\mathbf{x}), \mathbf{x})$, in the sample space, Ω. Given that the state equation, in (7.1) is null everywhere, in the sample space, then its derivative with respect to $\mathbf{x}$ is also null everywhere [6]. Thus, the following equality is written [1]:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \frac{\mathrm{d}H}{\mathrm{d}\mathbf{x}} \tag{7.6}$$

where $\mathrm{d}H/\mathrm{d}\mathbf{x}$ refers to the total derivative of H with respect to $\mathbf{x}$. The Lagrangian function, defined in terms of the adjoint states, $\boldsymbol{\lambda}$, allows to calculate the implicit components of the total derivatives of H, with respect to $\mathbf{x}$. Considering the independence of the load vector $\mathbf{f}$ to the state variables $\mathbf{u}$, the adjoint-state equations are obtained from (7.4), as follows:

$$\mathbf{K}(\mathbf{x})\boldsymbol{\lambda} - \frac{\partial H}{\partial \mathbf{u}} = \mathbf{0} \tag{7.7}$$

where $\mathbf{K}(\mathbf{x})$ is given by (7.5),which is equivalent to the equilibrium (7.1). Differentiating (7.3) with respect to $\mathbf{x}$, it follows that:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \frac{\partial H}{\partial \mathbf{x}} + \frac{\partial H}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}} - \boldsymbol{\lambda}^{\boldsymbol{T}}\left[\frac{\partial}{\partial \mathbf{x}}[\mathbf{Ku} - \mathbf{f}] + \frac{\partial}{\partial \mathbf{u}}[\mathbf{Ku} - \mathbf{f}]\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\right] \tag{7.8}$$

which can be simplified using equality of (7.4), yielding:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \frac{\partial H}{\partial \mathbf{x}} - \boldsymbol{\lambda}^{\boldsymbol{T}} \frac{\partial}{\partial \mathbf{x}}[\mathbf{Ku} - \mathbf{f}] \tag{7.9}$$

Considering the independence of $\mathbf{f}$ to te design variables $\mathbf{x}$, from the equality in (7.6), it results:

$$\frac{\mathrm{d}H}{\mathrm{d}\mathbf{x}} = \frac{\partial H}{\partial \mathbf{x}} - \boldsymbol{\lambda}^{\boldsymbol{T}} \frac{\partial \mathbf{K}}{\partial \mathbf{x}}\mathbf{u} \tag{7.10}$$

From the last equation, it is concluded that the implicit component of the total derivative of H, with respect to $\mathbf{x}$, is given by $\left(-\boldsymbol{\lambda}^{\boldsymbol{T}} \frac{\partial \mathbf{K}}{\partial \mathbf{x}} \mathbf{u}\right)$. This methodology is twofold [1,4]:

1. Solve the adjoint system of equations, in (7.7);
2. Obtain the sensitivities from (7.10).

Finally, in the sample space, it results that [5][3]:

$$\frac{\mathrm{d}\phi}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{d}H}{\mathrm{d}\mathbf{x}} \tag{7.11}$$

7.2 RBRDO of composite laminate structures

In the following sections, we study the numerical solution of the RBRDO problem of composite laminate structures, presented in Definition 3.10. Over the Chapters 4 and 5, we introduced a number of algorithmic and analytical developments to allow the efficient solution of the problem with the exclusive use of evolutionary algorithms (EAs). As a result, the original RBRDO problem must be rewritten, relying on local approximations of the global robustness and reliability measures. The process is detailed as follows.

In the present study, the uncertainty in the system response functionals is associated with four possible sources of uncertainty, divided into two groups of variables: random design variables $\mathbf{x} \in \Omega_D$ and random parameters $\boldsymbol{\pi} \in \Omega_{\Pi}$. They are organized as follows:

Source 1: mechanical properties $\mathbf{m} \subseteq \boldsymbol{\pi}$, defined as random parameters;
Source 2: ply angles of the laminates $\boldsymbol{\theta} \subseteq \mathbf{x}$, defined as random design variables;
Source 3: laminate thicknesses $\mathbf{h} \subseteq \mathbf{x}$, defined as random design variables;
Source 4: point loads $\mathbf{p} \subseteq \boldsymbol{\pi}$, defined as random parameters.

Depending on the uncertainty quantification measure (robustness or reliability), the sources of uncertainty differ. In *robustness assessment*, the uncertainty is propagated through the random design variables $\mathbf{x}$ and/or the random parameters $\boldsymbol{\pi}$. In agreement with Definition 3.10, the mean values of the random design variables, $\boldsymbol{\mu}_{\mathbf{x}} \in D$, are the design variables of the RBRDO problem, thus varying between design solutions. The mean values of the random parameters, $\boldsymbol{\mu}_{\boldsymbol{\pi}} \in \Pi$, and the standard deviations $\boldsymbol{\sigma}_{\mathbf{x}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\pi}}$ are constant. In *reliability assessment* the uncertainty of the system is propagated only through the mechanical properties, $\mathbf{m}$, for which the mean values, $\boldsymbol{\mu}_{\mathbf{m}} \in \Pi$, and the standard deviations, $\boldsymbol{\sigma}_{\mathbf{m}}$, are constant. Furthermore, the inner evaluation of the limit-state function is performed for fixed values of $\boldsymbol{\mu}_{\mathbf{x}}$ and $\boldsymbol{\mu}_{\boldsymbol{\pi}}$. Reliability assessment is only applied to the design solutions that satisfy either (5.14) or (5.26).

Such distribution of uncertainty assumes that the reliability of composite laminate structural systems is mainly affected by the mechanical properties, which is explained as follows. On the one hand, the point loads are static and the variability in their amplitudes controlled, in the design process, by the robustness measure, by designing structural systems less sensitive to random oscillations in loads amplitudes. Similarly, the ply angles (fiber misalignment) and the laminate thickness (dimensional instability from polymeric cure) characterize the structural system at a macro-scale laminate-level and are easier to control during the production stage. Hence, controlled by the robustness measure. On the other hand, mechanical properties define the strength properties of the material and characterize the structural system at a micro-scale level (by micromechanics). The associated uncertainty relates with the manufacturing process of the composite material itself, rather than the

assembly of the laminate layers, thus having superior importance on the structural reliability than the other variables. Yet, they are included in the robustness measure as they are expected to affect the variability of the design solutions as well.

In reliability assessment, the standard deviations $\boldsymbol{\sigma}_{\mathbf{m}}$ of the mechanical properties are necessary to apply the inverse transformation $\mathbf{m} = T^{-1}(\mathbf{y}_{\mathbf{m}})$, in order to comply with (4.20) and (4.46). Assuming the original random variables follow $\mathbf{m} \sim N(\boldsymbol{\mu}_{\mathbf{m}}, \boldsymbol{\sigma}_{\mathbf{m}})$, then the following known transformation is applied:

$$\mathbf{m} = \boldsymbol{\mu}_{\mathbf{m}} + \mathbf{y}_{\mathbf{m}}\boldsymbol{\sigma}_{\mathbf{m}} \tag{7.12}$$

with $\mathbf{y}_{\mathbf{m}} \sim N(\mathbf{0}, \mathbf{1})$.

Now, we rewrite the global RBRDO problem, in Definition 3.1, as follows.

Definition 7.1 (*local RBRDO problem*): Let $\det \mathbf{C}_{\boldsymbol{\phi}} : D \times \Pi \to \mathbb{R}$ be a local robustness measure, given by (5.7), where $\boldsymbol{\phi}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}}) = (\bar{u}, \bar{R})$ is the vector of deterministic system response functionals, given by (3.5) and (3.11). Furthermore, let $\mathbf{g}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})$ be a vector of deterministic limit-state functions, with components:

$$g_1(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}}) = \frac{\bar{u}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})}{u^a} - 1 \tag{7.13}$$

as the deterministic displacement limit-state function, and $g_2(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}})$ as the deterministic piecewise equivalent reliability limit-state function, given by (5.14) or (5.26). The local formulation of the RBRDO problem is stated as follows:

$$\begin{aligned} \min_{\boldsymbol{\mu}_{\mathbf{x}}} \quad & \mathbf{f}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}}, \boldsymbol{\sigma}_{\mathbf{x}}, \boldsymbol{\sigma}_{\boldsymbol{\pi}}) = \left(W(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}}),\ \det \mathbf{C}_{\boldsymbol{\phi}}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}}, \boldsymbol{\sigma}_{\mathbf{x}}, \boldsymbol{\sigma}_{\boldsymbol{\pi}})\right) \\ \text{subject to:} \quad & g_1(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}}) \leq 0 \\ & g_2(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\mu}_{\boldsymbol{\pi}}) \geq 0 \\ & \boldsymbol{\mu}_{\mathbf{x}} \in D \end{aligned} \tag{7.14}$$

In the following sections, we present the solution of the previous problem using solely the EAs developed in Chapters 4 and 5. Two instantiations of the problem are considered: one for the reliability assessment performed by the PMA and another by the RIA. We begin with the description of a physical model, common to both instantiations. Then, we discuss the results obtained by the two RBRDO methodologies and compare them.

7.2.1 Physical model

The structural analysis of laminate composite structures is based on a displacement formulation of the FEM, in particular, using the shell finite element model developed by Ahmad et al. [7]. It is an isoparametric element with eight nodes, four integration points per ply, and five freedom degrees per node, based on the Mindlin shell theory. The shell consists of a number of perfectly bonded plies. Each individual ply is assumed homogeneous and anisotropic.

A clamped cylindrical shell laminated structure is considered as show in Figure 7.1. Nine vertical loads of mean value $P_k = 11.5\ KN$ are applied along the free linear side (AB) of the structure. This side is constrained in the y-axis direction (see appendix B). The structure is divided into four laminates, grouping all elements, and there are two elements per laminate. The laminate/element

distribution is as shown also in Figure 7.1. Balanced angle-ply laminates with five layers on a symmetric stacking sequence $[+\theta/-\theta/0/-\theta/+\theta]$ are considered. The ply angle θ is referenced to the x-axis of the Cartesian referential. Variables h_i, $i = 1,2,3,4$, denote the thicknesses of the laminates. A smoothing procedure is performed at the boundaries of the laminates.

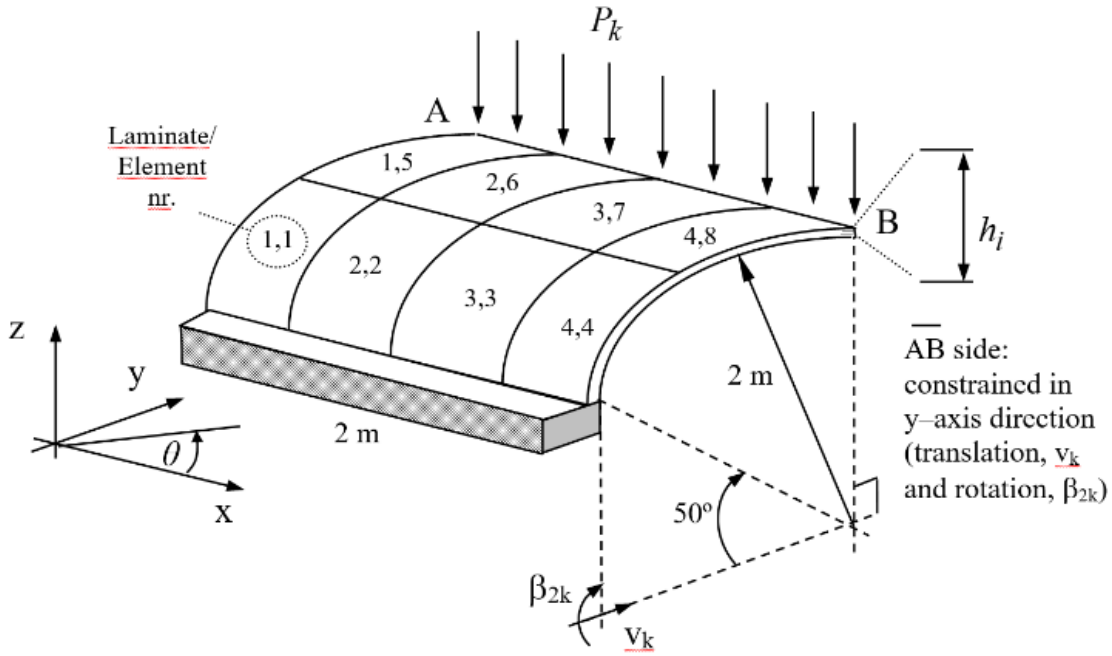

Figure 7.1: Geometric definition of the cylindrical shell structure and composite laminates distribution.

A composite material built with the carbon/epoxy system denoted T300/N5208 [8] is used in the current analysis. This is a unidirectional composite of long carbon fibers aggregated in an epoxy matrix. The macro mechanics' mean values of the elastic and strength properties of the ply material used in the laminates of the structure are presented in Table 7.1. The elastic constants of the orthotropic ply are the longitudinal elastic modulus E_1, the transversal elastic modulus E_2, the in-plane shear modulus G_{12} and the in-plain Poisson's ration ν_{12}. The ply strength properties are the longitudinal tensile strength X, the longitudinal compression strength X', the transversal tensile strength Y, the transversal compression strength Y' and the shear strength S.

Table 7.1: Mean values of the mechanical properties of the T300/N5208 composite layers [8].

$E_1\ [GPa]$	$E_2\ [GPa]$	$G_{12}\ [GPa]$	ν_{12}
181.00	10.30	7.17	0.28
$X; X'[MPa]$	$Y; Y'[MPa]$	$S\ [MPa]$	$\rho\ [kg/m^3]$
1500; 1500	40; 246	68	1600

Also, according to an importance study of the mechanical properties themselves, in António *et al.* [9], the most important parameters were shown to be E_{1j}, E_{2j}, Y_j and S_j. Therefore, the vector of random mechanical properties, $\mathbf{m}$, only includes the random parameters E_{1j}, E_{2j}, Y_j and S_j, where the subscript $j = 1,2,3,4$ denotes the laminate number. Being four the number of laminates, there are sixteen random mechanical properties aggregated in vector $\boldsymbol{\pi}$, plus nine random vertical loads. Five random design variables are considered in vector $\mathbf{x}$: one ply-angle θ, equal for all symmetric laminates, and four laminates thickness variables h_i, for $i = 1,2,3,4$. The constant values of the standard deviations of the random variables are given by Table 7.2.

Table 7.2: Standard deviations of the random variables.

Group 1	$\sigma_{\mathrm{m}_i} = 6\%\ \mu_{\mathrm{m}_i}, i = 1, \ldots, 16$
Group 2	$\sigma_\theta = 5°$
Group 3	$\sigma_{h_j} = 5 \times 10^{-4}, j = 1, \ldots, 4$
Group 4	$\sigma_{\mathrm{P}_k} = 6\%\ P_k, k = 1, \ldots, 9$

7.2.2 RBRDO with imposed probability density value

In this section, we present the numerical solution of the local RBRDO problem, where the reliability assessment problem is solved by the PMA and the reliability constraint is written in terms of the piecewise equivalent PMA limit-state function, in (5.26). The proposed problem is solved exclusively by two EAs: the MOGA-2D and the mGA. The parameters of both algorithms are defined in Table 7.3 (a) and (b). Moreover, the allowable critical displacement is $u^a = 8 \times 10^{-2}$ [m] and the allowable reliability index is $\beta^a = 3.0$.

In order to observe the effects of the reliability constraint in the design optimization of composite laminate structures, the same design problem, is solved without reliability assessment (robust design optimization (RDO)), where $g_2(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi})$ is replaced by the deterministic realization of the stress limit-state function, in (3.13), as follows:

$$g_2(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi}) = \frac{\bar{R}(\boldsymbol{\mu}_\mathbf{x}, \boldsymbol{\mu}_\boldsymbol{\pi})}{R^a} - 1 \tag{7.15}$$

where R^a is the allowable critical Tsai number and is equal to 1.

Moreover, regarding the reliability assessment, the PMA only provides a qualitative reliability measure, since from its optimal solution it is not possible to calculate any estimate of the actual probability of failure. For that reason, the validation of the mGA is integrated in the RBRDO process itself, in the sense that if the developed algorithms are well constructed, the design solutions in the resulting Pareto optimal set $X_{PO_{known}}$ do not violate the imposed reliability constraint and the resulting Pareto front $Z_{PF_{known}}$ satisfies the three main goals identified in Sect. 2.2.2.

Table 7.3: Parameters and size constraints of the MOGA-2D (a) and of the mGA (b).

(a) MOGA-2D		(b) mGA	
Short Population size, $n_{\mathbf{SP}}$	30	Population size, $n_{\mathbf{P}}$	12
Elite group size, n_{top}	0.33 $n_{\mathbf{SP}}$	Elite group size, n_{top}	0.33 $n_{\mathbf{P}}$
Mutation group size, n_{bot}	0.20 $n_{\mathbf{SP}}$	Mutation group size, n_{bot}	0.33 $n_{\mathbf{P}}$
Number of generations	300	Stop if elite intact after/ Maximum number of generations	6/ 30
Code format (bits per variable)/ /size constraint for ply angle	4/ /[0°, 90°]	Similarity control: allowed nº of different variables	1
Code format (bits per variable)/ /size constraint for laminate thicknesses	5/ /[0.005m, 0.040 m]	Code format (bits per variable)/ /size constraint for **m**	3/ /see (4.45)

The proposed methodology proved to be efficient, as it took about 17 hours (in debug mode) to complete the design optimization procedure, using an Intel(R) Core(TM) i7-6700 CPU @ 3.40Ghz processor. The most significant part of the RBRDO problem, in terms of computational cost, is clearly the reliability assessment. Without the reliability constraint, the RDO problem, is solved in approximately 35 minutes. The main contributions to the efficiency of the method were the redefinition of the original PMA problem, in (4.16), to the equivalent unconstrained maximization of the positive fitness functional $\bar{f}$, in (4.47), suitable to be solved by EAs, and the decomposition of the random variables that allowed to eliminate the equality constraint of the PMA, reducing the search space to the surface of the hypersphere of radius β^a. Without these developments, the reliability assessment would be cumbersome and the design optimization impracticable.

The numerical results show an interesting behavior of the studied structural system, in terms of its structural integrity, which we display on the following set of figures. Figure 7.2 shows the Pareto fronts (PF) obtained with and without the reliability constraint (RBRDO and RDO, respectively). It is observed that both fronts are close to each other, meaning the optimal region of the objective space is of similar shape, in both the RBRDO and RDO problems. However, solutions tend to become slightly heavier with increasing values of system variability, in the RBRDO formulation, suggesting the reliability constraint is eliminating design solutions that are probabilistically infeasible.

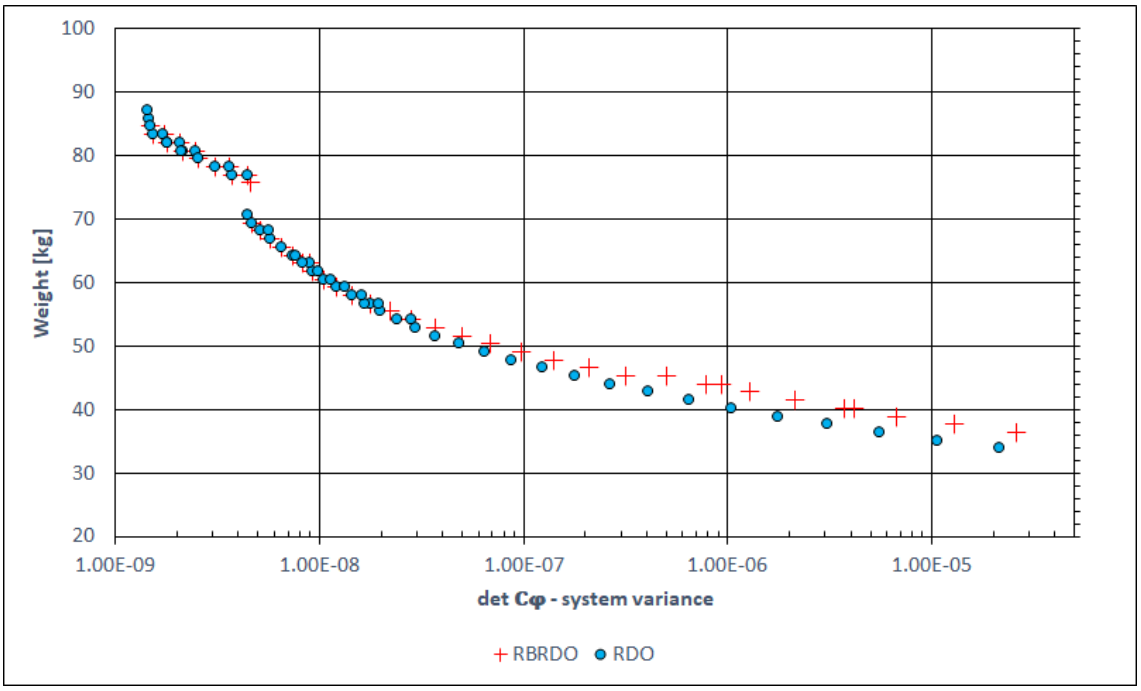

Figure 7.2: Pareto fronts obtained with and without reliability assessment (PMA).

Figure 7.3 shows the distribution of the critical Tsai number, along the PFs obtained with and without reliability assessment. In the figure, the effect of the reliability constraint is certainly seen. First, it shows that, between 70-90 kg of weight, the critical Tsai numbers are mostly coincident, meaning that robustness leads the design optimization, in this weight range. On the other hand, for the remaining design solutions, there are significant differences in the values of $\bar{R}$. In the RBRDO, the reliability constraint does not allow the critical Tsai number of the Pareto optimal solutions to fall under the value of 1.19 and imposes a jump on its value, in the lower portion of the PF. In the RDO, the critical Tsai number can be as lower as 1, satisfying the deterministic stress constraint.

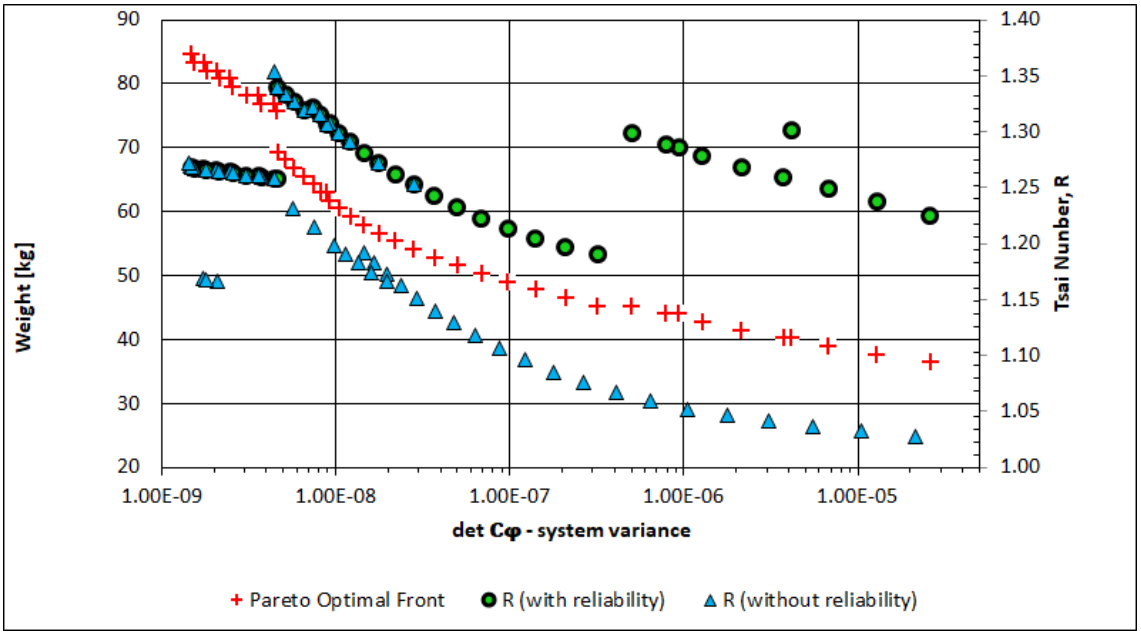

Figure 7.3: Distribution of the critical Tsai number, along the Pareto fronts, obtained with and without reliability assessment (PMA).

This same behavior is seen on the distribution of the reliability index, along the PF of the RBRDO problem, displayed in Figure 7.4. The actual values of β_{HL} were calculated *a posteriori*, for each Pareto optimal solution, by a gradient-based algorithm of the RIA [4,10], and were used to validate the proposed methodology. As seen, the reliability index is greater than $\beta^a = 3.0$, for every Pareto optimal solution, meaning the proposed evolutionary-based PMA was successfully applied. Similarly, the distribution of the optimal solution of the transformed PMA problem, in (4.47), is shown in Figure 7.5.

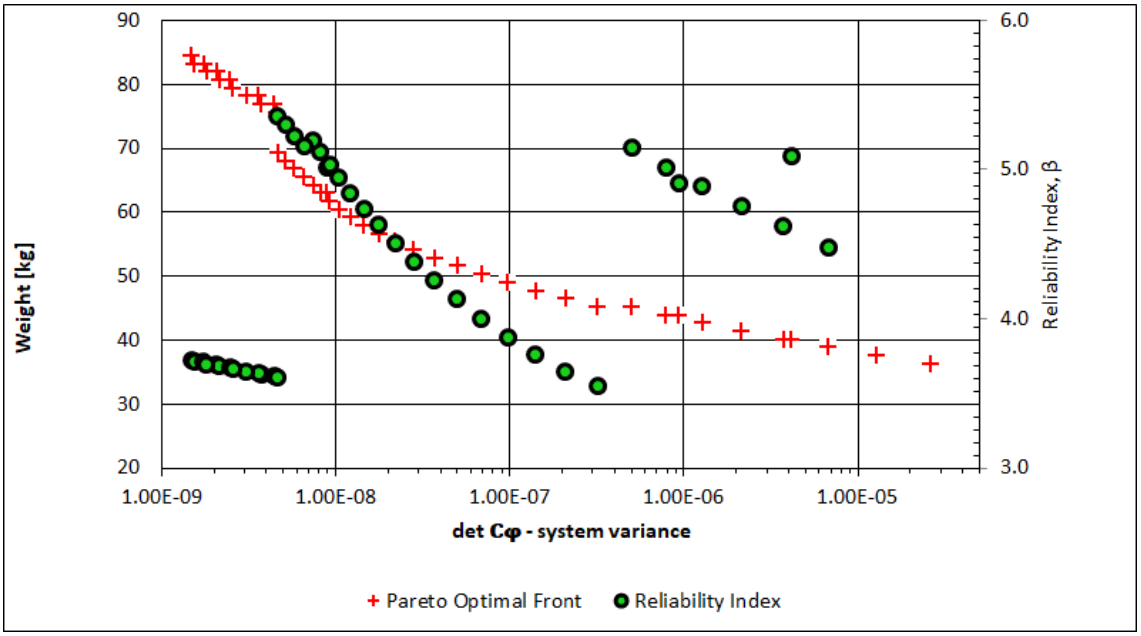

Figure 7.4: Distribution of the reliability index of the Pareto optimal solutions, along the Pareto front.

Comparing Figure 7.3, Figure 7.4 and Figure 7.5, it is possible to observe the parallelism between the distributions of $\bar{R}$, β_{HL} and $g_2\left(T^{-1}(\mathbf{a_{MPP}})\right)$, along the PF of the RBRDO problem, indicating that a functional relationship might exist between the three integrity measures. An interesting fact can be observed on each one of these figures. The heaviest Pareto optimal solutions (70-90 kg) have lower levels of structural integrity, than several of the lighter solutions, with higher levels of variability. Furthermore, there are regions of the PF where sudden jumps in the structural integrity measures happen, meaning that Pareto optimal solutions with similar weight and variability values can have very different integrity levels. Complementary studies are needed, in order to provide a proper explanation for this behavior. We suspect it relates to the anisotropy of the composite materials and the fact that the structural system comprises four independent laminates, while the PF considers the total structural weight. That is, there might exist a many-to-one relationship between the feasible search space S_X and the feasible objective space S_Z.

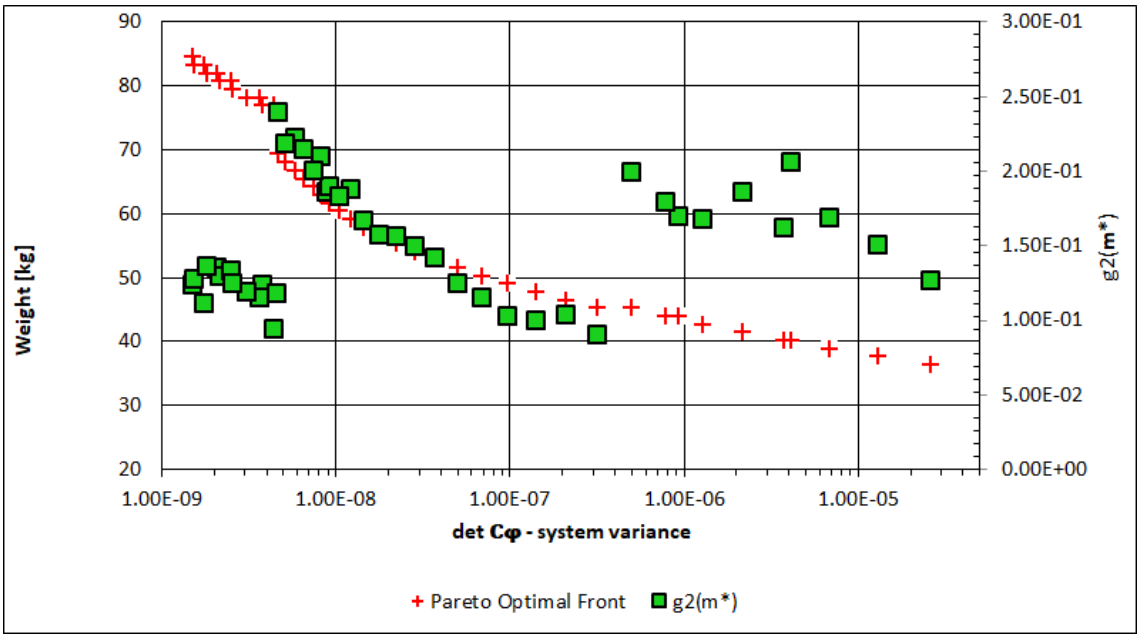

Figure 7.5: Distribution of the solution of the transformed PMA, in (4.47), along the Pareto front.

Figure 7.6 shows the distribution of the individual coefficients of variation of displacement and stress, respectively, given by:

$$CV(\bar{u}) = \sqrt{\mathrm{Var}(\bar{u})}/u^a$$

$$CV(\bar{R}) = \sqrt{\mathrm{Var}(\bar{R})}/R^a \tag{7.16}$$

These coefficients are related with the diagonal elements of the variance-covariance matrix, $\mathrm{Var}(\bar{u})$ and $\mathrm{Var}(\bar{R})$ (given in (5.5)), and are used to analyze the individual contribution of the displacement and stress limit-state functions, in the total variability of the Pareto optimal solutions. As it can be seen, $CV(\bar{R})$ is larger but constant along the PF, varying about 25%. On the contrary, $CV(\bar{u})$ increases exponentially with the decrease in the total structural weight. It is concluded that the variability associated with the displacement response has a larger responsibility on the increase of the system variability (i.e., decrease of robustness) of the Pareto optimal solutions.

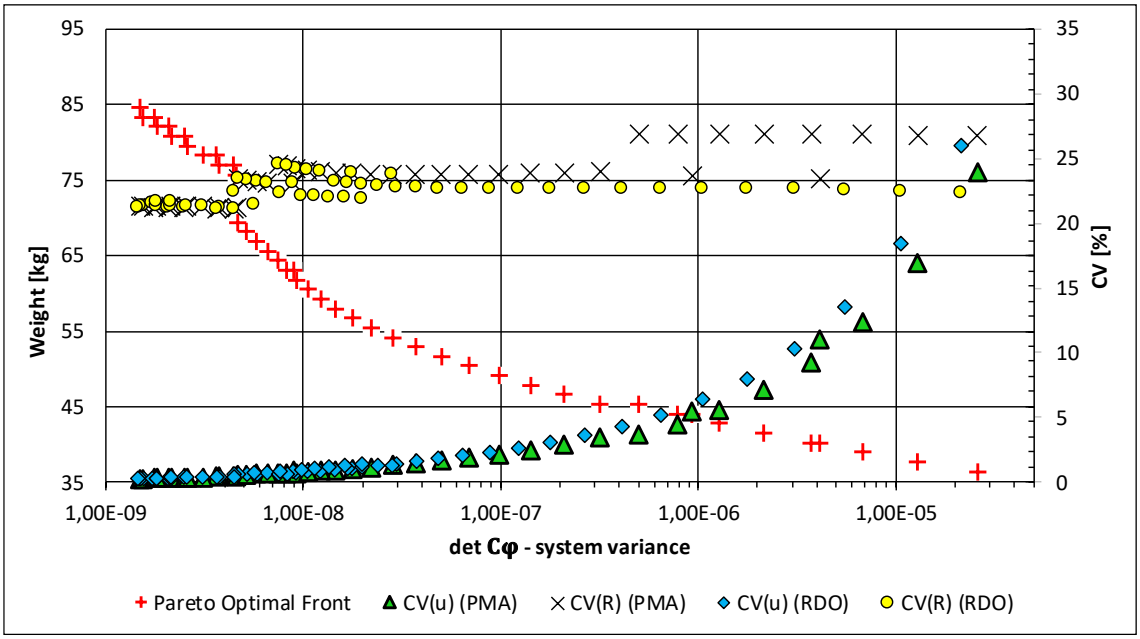

Figure 7.6: Distribution of the coefficients of variation, $\boldsymbol{CV}(\bar{\boldsymbol{u}})$ and $\boldsymbol{CV}(\bar{\boldsymbol{R}})$, along the Pareto front (PMA).

By comparing the coefficients obtained for the RBRDO and the RDO problems, it is possible to see how the reliability constraint affects the robustness of the system. One observed fact is that the existence of a reliability constraint, associated with the stress state of the system, results in a small increase in the variability of the stochastic stress response. Thus, we conclude that there is a positive trade-off between reliability and robustness, in the sense that, in the RBRDO, Pareto optimal solutions have significantly improved levels of structural integrity, at the expense of a small reduction in robustness.

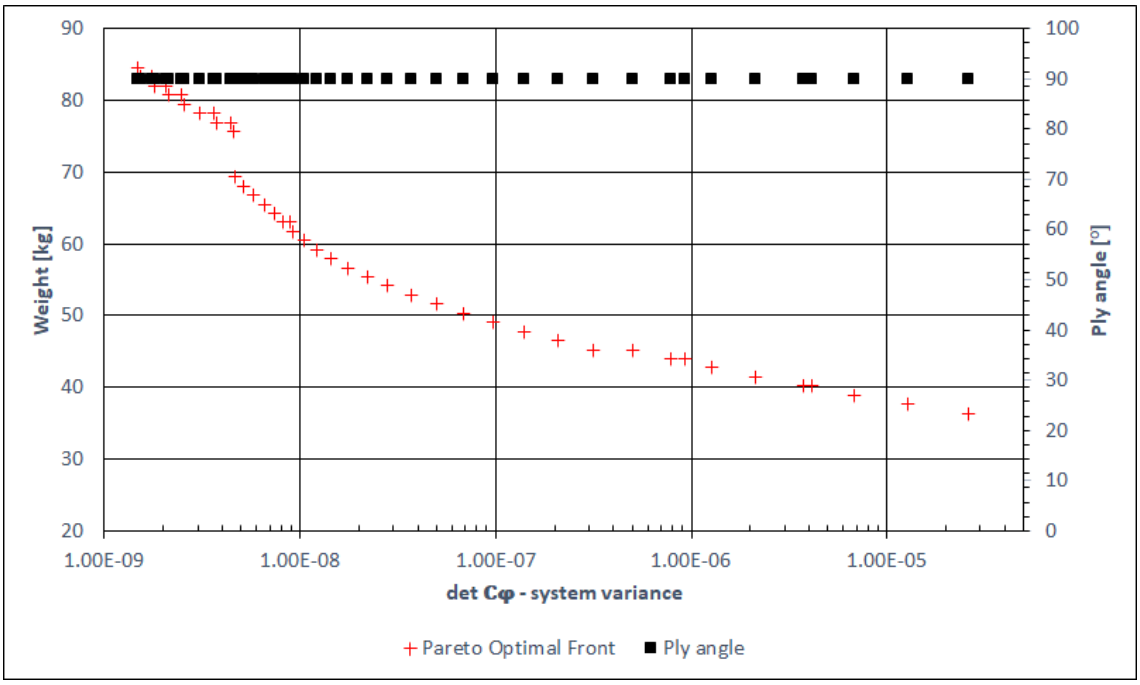

Figure 7.7: Distribution of the mean-values of the random ply-angles, along the Pareto front (PMA).

Regarding the Pareto optimal solutions, Figure 7.7 shows the distribution of the mean-values of the random play-angles, along the PF. It is observed that the optimal orientation of the fibers is at 90°, for all Pareto optimal solutions, meaning that the fibers are optimally aligned along the y-axis. Figure 7.8 shows the distribution of the thickness variables, along the PF. As seen, the design optimization process evolved towards increasing mainly the thickness of the fourth laminate, h_4, as a means to increase the total structural weight. The reason for this behavior would need further investigation, but one may conjecture the possibility that the chosen size-constraints may be too wide and that, in the absence of any geometrical constraint, the selective pressure during the evolution determined an evolutionary path that favored this configuration.

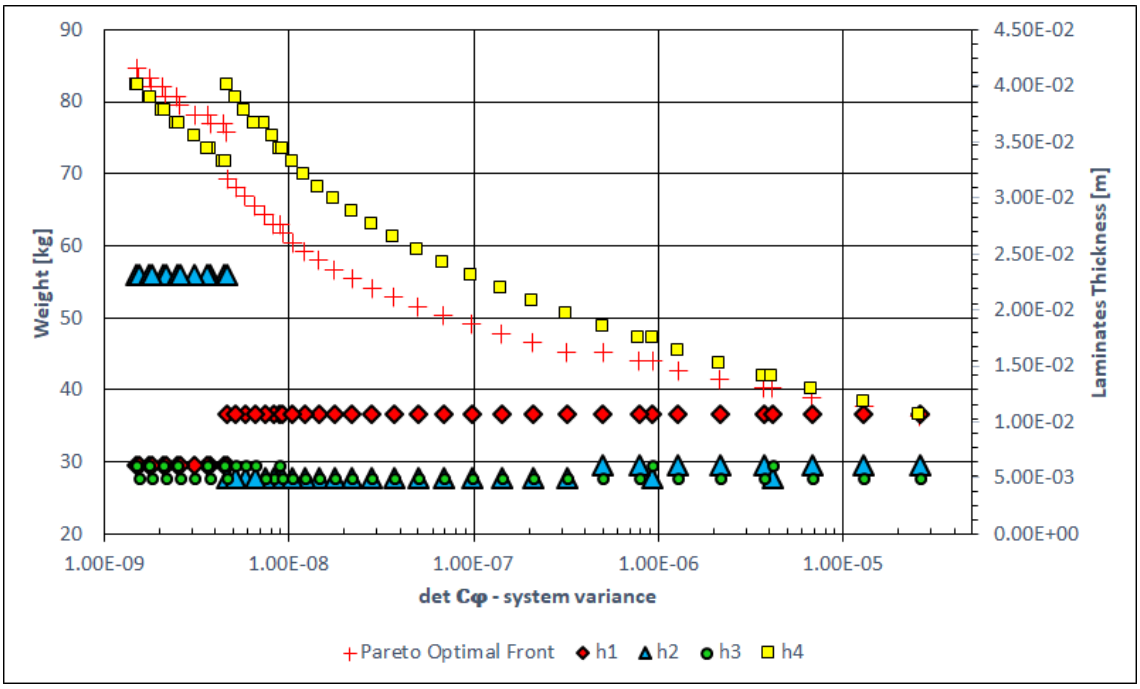

Figure 7.8: Distribution of the mean-values of the random thickness variables, of the four laminates, along the Pareto front (PMA).

The obtained results show that the proposed method is an efficient tool of stochastic design optimization. As a final comment to the presented results, it is important to notice that the mentioned computational effort corresponds to an evolutionary process run for 300 generations. Regarding the three main goals of MOEA (see Sect. 2.2.2), this number proves to be a good compromise between a gradual progression towards the true Pareto front and the ability to generate and maintain diversity on the Pareto front. As observed in Figure 7.9, from early on in the evolution, the set of nondominated points tends to the known Pareto front, at generation 300, demonstrating the convergence ability of the MOGA-2D. It is seen that, at generations 75 and 100, convergence is almost achieved. In truth, 300 generations may exceed the necessary number of generations for a satisfactory convergence.

Yet, we would like to emphasize that the results presented are obtained for the particular physical model presented in Sect. 7.2.1 and for the algorithm parameters set in Table 7.3. As with most EAs, both the MOGA-2D and the mGA are problem-dependent and require the fine-tuning of the respective parameters and appropriate stopping criteria. As future work, it would be interesting to test different size constraints, or to impose a geometric constraint, that allowed an even distribution of the laminate thicknesses. It would be important, as well, to study the many-to-one relationship between the feasible search space and the feasible objective space, in the case of multi-laminate composite structures, to understand how different Pareto optimal solutions, with very different levels of structural integrity, can have very similar Pareto fronts.

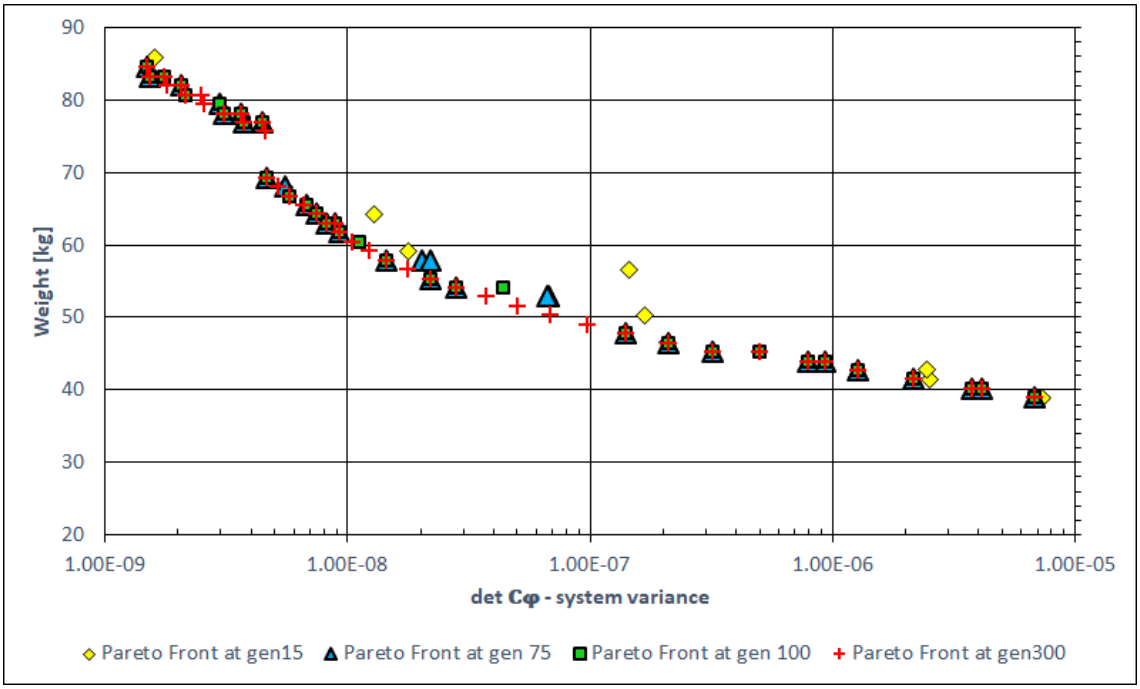

Figure 7.9: Convergence of the nondominated set, at different stages of the evolution: 15, 75, 100 and 300 generations.

7.2.3 RBRDO with imposed limit-state value

In this section, we present the numerical solution of the local RBRDO problem, where the reliability assessment problem is solved by the RIA and the reliability constraint is written in terms of the piecewise equivalent RIA limit-state function, in (5.14). The proposed problem is solved exclusively by two EAs: the MOGA-2D and the HmGA, to solve the RIA. The parameters of both algorithms are defined in Table 7.4 (a) and (b). Moreover, the allowable critical displacement is $u^a = 8 \times 10^{-2}$ [m] and the allowable reliability index is $\beta^a = 3.0$.

Contrary to the PMA, the RIA returns a quantitative measure of the structural reliability. For that reason, in Sect. 7.2.3.1, we begin by validating the HmGA, by comparison of the algorithm against alternative methods of reliability assessment. As a reference set of solutions, we use Pareto optimal solutions, obtained in the previous section, for which the reliability index was calculated as a post-operation procedure, by a gradient-based algorithm developed in [4,10]. Such results are shown in Figure 7.4 Furthermore, for global validation of the HmGA, we select three of the referred Pareto optimal solutions and estimate two-sided confidence intervals of the unknown parameter p_f. We also discuss the performance of the HmGA against the alternative methods. Then, in Sect. 7.2.3.2, we present the results of the local RBRDO problem, with reliability assessment performed by the RIA.

7.2.3.1 Validation of the HmGA

To execute the HmGA an Intel(R) Core(TM) i7-6700 CPU @ 3.40Ghz processor was used. It was imposed that convergence is achieved after at least $\Delta t = 50$ generations were run after the formation of the first reduced search region $\mathcal{Z}^{t_1}$. When evaluating the accuracy of EAs, it is important to take into account that these methods act on a discretized search space and that their resolution is limited, mainly, by the number of genes per variable and the size-constraints, while for gradient-based methods it is only limited by floating-point format of the mathematical operations.

Table 7.4: Parameters and size constraints of the MOGA-2D (a) and of the HmGA (b).

(a) MOGA-2D		(b) HmGA	
Short Population size, $n_{\mathbf{SP}}$	30	Population size, $n_{\mathbf{P}}$	15
Elite group size, n_{top}	0.33 $n_{\mathbf{SP}}$	Elite group size, n_{top}	0.33 $n_{\mathbf{P}}$
Mutation group size, n_{bot}	0.20 $n_{\mathbf{SP}}$	Mutation group size, n_{bot}	0.33 $n_{\mathbf{P}}$
Number of generations	300	$\Delta\beta^0_{max}/k_{max}/\Delta a/$ $t_z/\sigma_a^2/\Delta t$	β_{max}/50/0.15/ 4/0.01/50
Code format (bits per variable)/ /size constraint for ply angle	4/ /$[0°, 90°]$	Similarity control: nº of different variables if $g_2 = 0$/ otherwise	1 var/ 3 var
Code format (bits per variable)/ /size constraint for laminate thicknesses	5/ /[0.005m, 0.040 m]	Code format (bits per variable)/ /size constraint for **m**	3/ /see (4.45)

Figure 7.10 compares the distribution of the reliability index calculated by the HmGA against the results already shown in Figure 7.4. Results show a very good proximity between the estimated values of β_{HL}. Performance-wise, the HmGA was able to achieve convergence at generation $t_1 + 50$, for each of the tested Pareto optimal solutions.

Calculating the relative difference, of the values of β, between the two methods as:

$$\varepsilon = \frac{|\beta_{\mathrm{HmGA}} - \beta_{\nabla}|}{\beta_{\nabla}} \tag{7.17}$$

where β_{HmGA} is the reliability index estimated by the HmGA and β_{∇} the reliability index estimated by the gradient-based method, it is seen in Figure 7.11 that the maximum relative difference between the two methods is inferior to 5% and the mean relative difference is about 0.6%. It is also worth to notice that, the majority of the estimated values have a relative difference inferior to the mean.

Figure 7.12 displays the distribution of the computing times of the HmGA, as a function of β_{HL}. We identify three intervals worth the attention. For $\beta_{HL} \in [3.552, 3.758]$ there is a cluster of solutions whose computing times vary between 5 and 8 minutes. For $\beta_{HL} \in [3.887, 4.483]$ computing times decrease and are relatively stable with a maximum difference of 2 minutes between the faster and the slower runs. Finally, for $\beta_{HL} \in [4.508, 5.363]$ there is a significant increase in the computing time of the solutions, with values varying between 4 and 12 minutes. A polynomial regression of the computing times shows a quadratic behavior of the algorithm. On the contrary, gradient-based are known to increase the computing times in proportion with the reliability index. On the following paragraphs we attempt to explain this particular behavior of the HmGA.

The computing time needed by the HmGA to evaluate each Pareto optimal solution depends essentially on two-time factors: the time spent on each generation and the time spent until convergence. The *time spent on each generation* is the time required by the HmGA to evaluate the entire population. Since finite elements analysis is the most significative operation executed during the evolutionary process, it primarily depends on the number of iterations run by the individual genetic repair operator (IGRO) and also on the execution of the basic evolutionary operators (BEOs). On the other hand, *the time spent until convergence* depends on the number of generations t_1 necessary to find the first reduced search region $\mathcal{Z}^{t_1}$ and on the imposed minimum number of generations Δt, until final convergence.

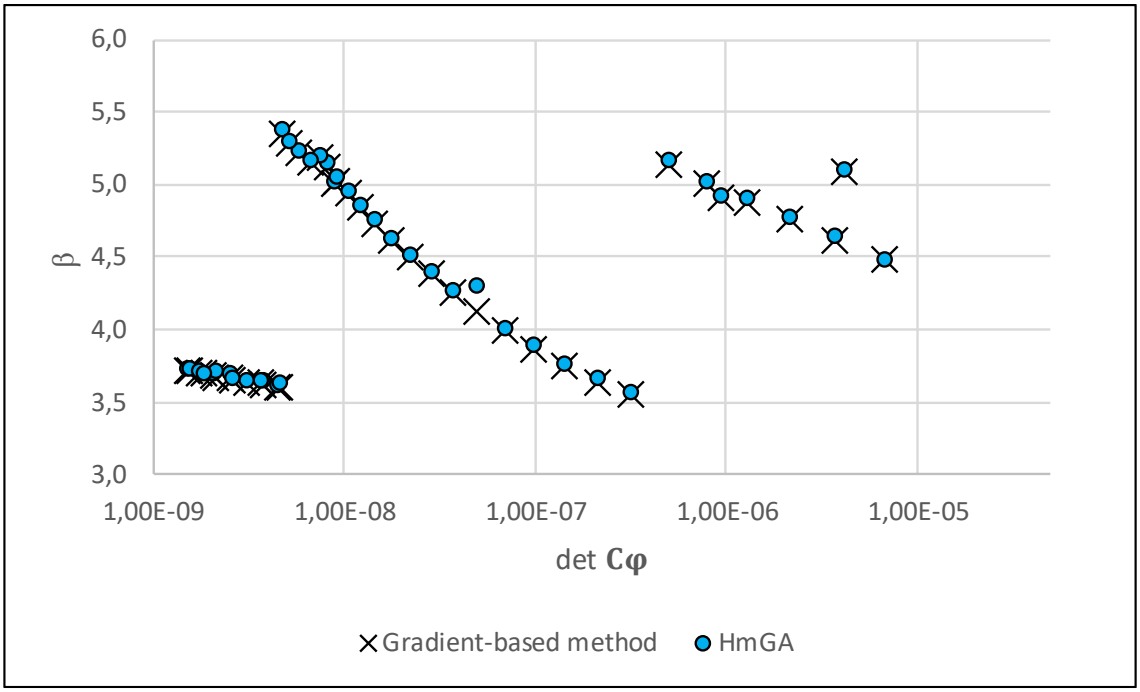

Figure 7.10: Comparison of the reliability index values calculated by the HmGA and a gradient-based algorithm.

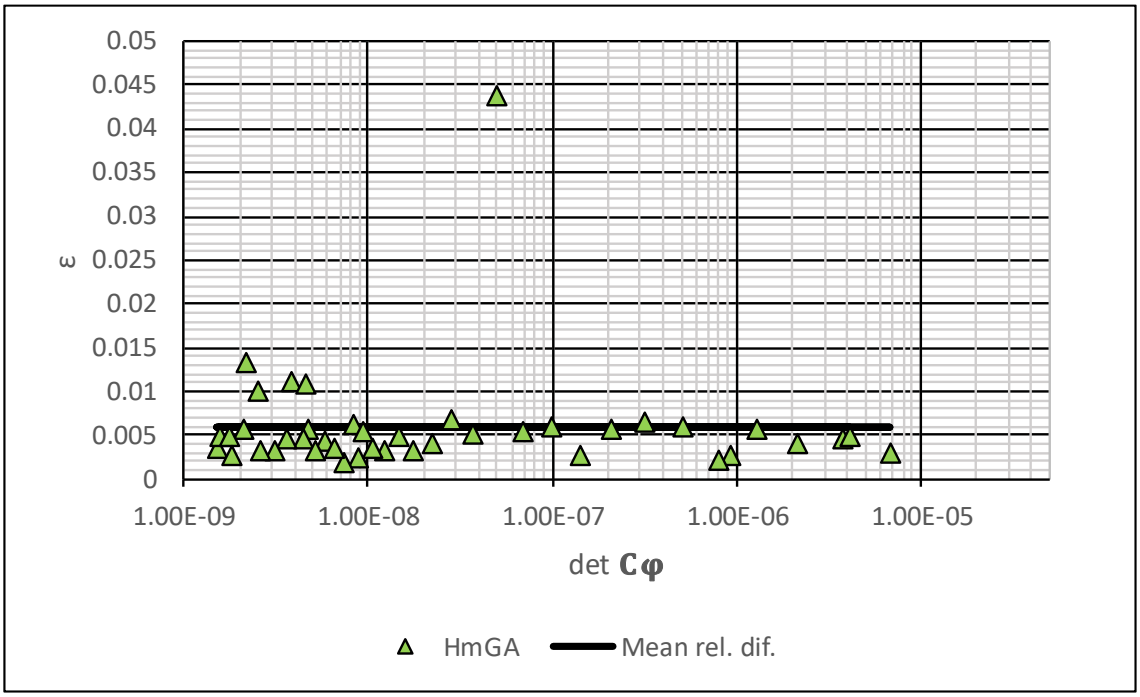

Figure 7.11: Relative difference between the values of $\boldsymbol{\beta_{HL}}$ calculated by the HmGA and a gradient-based algorithm.

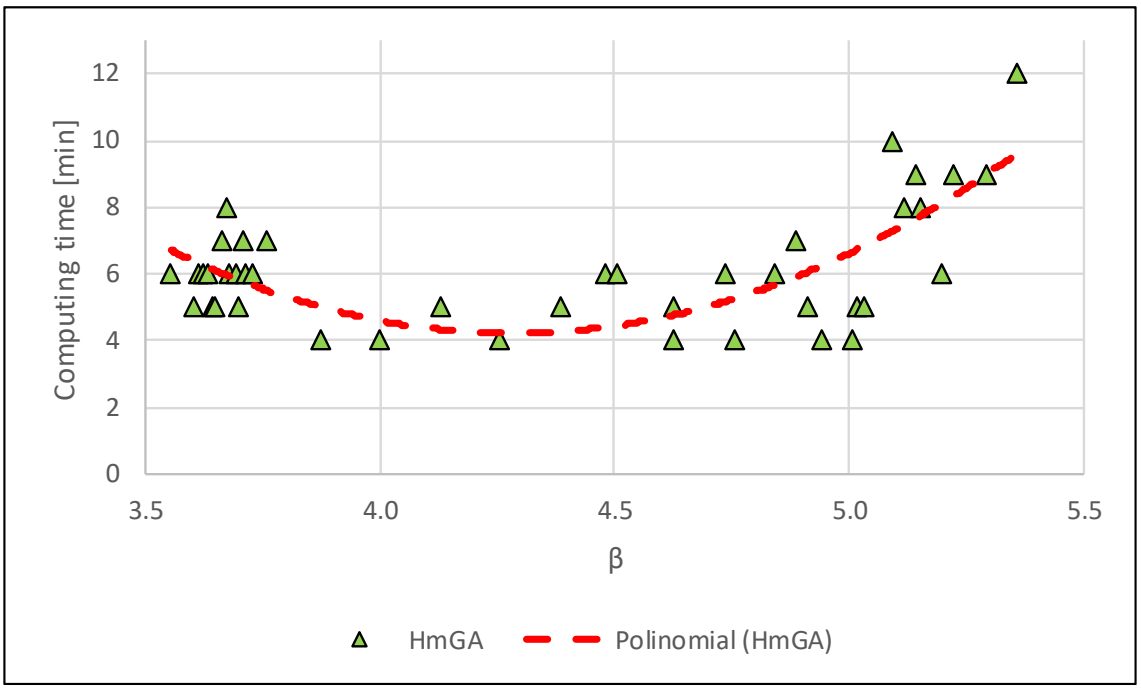

Figure 7.12: Distribution and quadratic polynomial regression of the computing times of the HmGA as a function of $\boldsymbol{\beta_{HL}}$.

Both factors are related to each other. Prior to generation t_1, the IGRO is often terminated before convergence (see end of Sect. 4.2.2.3). In opposition, after generation t_1, every solution in the population will be of high probabilistic failure content, after being repaired. While the number of genetic corrections is expectedly lower for these solutions, for being naturally closer to the failure surface, the IGRO must run until convergence. There is an increase in the number of model evaluations and therefore most of the computing time is spent on the Δt generations after t_1. However, it also means that, on this generational period, the HmGA has a more constant and predictable behavior, for all Pareto optimal solutions it evaluates.

In agreement with the following analysis, it can be concluded that the factor that contributes the most to the variations in the computing times is the number of generations needed to find $\mathcal{Z}^{t_1}$. Figure 7.13 plots t_1 as a function of the β_{HL} of each Pareto optimal solution. As is seen, there is an evident parallelism between the distribution of t_1 (in generations) and the distribution of the total computing times (in minutes). While the time spent on each generation is expected to be short, particularly before t_1, because the BEOs only evaluate the newly generated solutions, at each generation, there is in an interesting phenomenon happening before t_1, due to the similarity control operator, that largely increases the number of generations needed to find $\mathcal{Z}^{t_1}$.

During the search process of the HmGA, it is possible to observe what we call a phenomenon of *triangular similarity*, where a superior elite solution has a considerable degree of genetic similarity with more than one elite solutions of inferior fitness, which are not similar between them. As a result, the population fails to preserve those inferior elite solutions and the number of solutions of high probabilistic failure content declines considerably, between consecutive generations. Even so, with the adopted strategy for similarity control, the HmGA is able to converge in practical computing times, demonstrating the robustness of the newly developed evolutionary operators. With further developments it will be possible to reduce even more the computing times of the algorithm. Such issue is discussed later in Sect. 7.4.

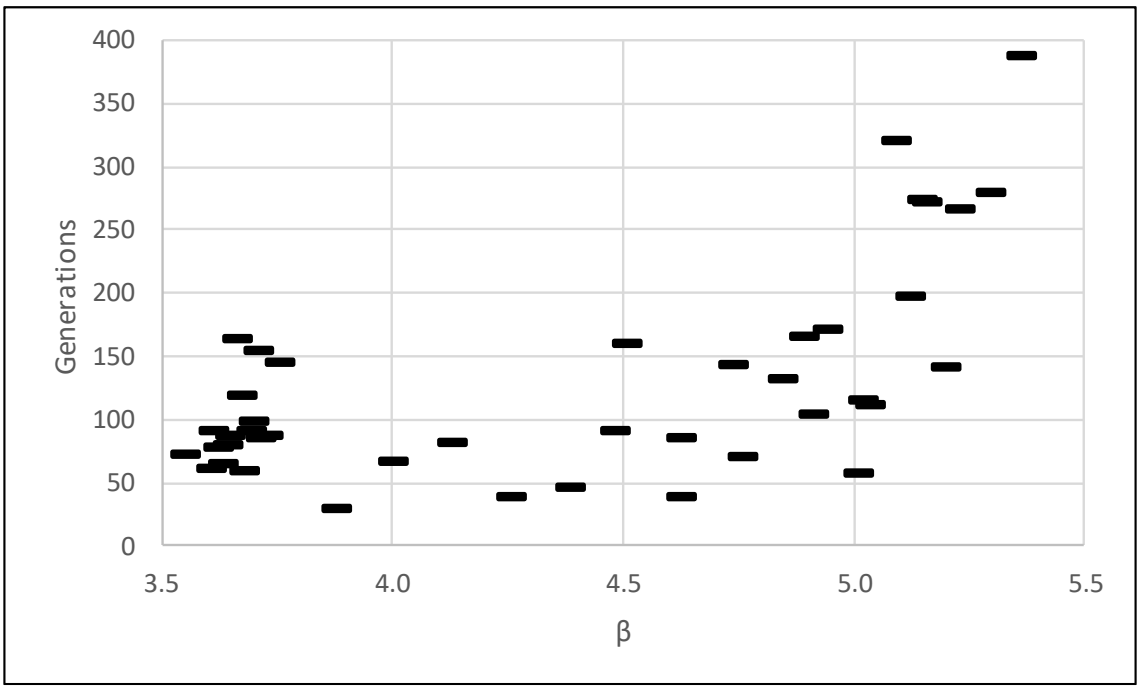

Figure 7.13: Number of generations needed to find $\boldsymbol{z^{t_1}}$, as a function of $\boldsymbol{\beta_{HL}}$.

Regarding the optimal solution of the RIA (the MPP), it is of our interest to analyze how close the points found by the HmGA and the gradient-based algorithm are. Given the number of design solutions evaluated, it is not possible to plot all of them. Only the most significant results are represented graphically. The relative difference of the coordinates of the MPP obtained by the two methods is given by:

$$\varepsilon = \frac{\left| m_{i_{\text{HmGA}}} - m_{i_\nabla} \right|}{m_{i_\nabla}} \tag{7.18}$$

where $m_{i_{\text{HmGA}}}$ is the ith coordinate of the MPP estimated by the HmGA and m_{i_∇} is the ith coordinate of the MPP by the gradient-based algorithm, for $i = 1, \ldots, 16$. It was found that, unsurprisingly, the Pareto optimal solution with the highest relative difference in the value of β_{HL} is the also the one with the highest relative difference in the coordinates of the MPP. Figure 7.14 plots the relative differences between the coordinates of the MPP estimated by the two methods, for such solution. It is seen that the highest relative difference happens for the coordinate referent to the longitudinal Young modulus of the first laminate, E_{11}. The value of 4% is neglectable and it is also the highest among all the Pareto optimal solutions.

Figure 7.15 shows a uniform radial distribution of the Pareto optimal solutions, allowing to plot the values of E_{11} of all the MPPs and to compare them against the respective mean values. Each point on these plots, here referred to as α-plots, is defined by $\left(E_{11_j} \cos\alpha \, , E_{11_j} \sin\alpha\right) \quad , \forall\, \alpha = 2\pi\,(j-1)/j$, for $j = 1, \ldots, 42$. Starting in $0\,rad$, the Pareto optimal solutions are ordered counterclockwise. Overall, the E_{11}-th coordinate of the MPPs, obtained by the two methods, present similar deviations from the mean values, with neglectable differences in magnitude.

The final step in validating the HmGA is to perform a crude MCS analysis. Since huge number of samples are needed to predict accurately very low probabilities of failure, only three representative Pareto optimal solutions are evaluated by MCS. Then, based on the performed MCS analysis, two-

sided confidence intervals of the unknown parameter p_f are calculated with confidence γ, for which the lower and upper bounds are given by [11]:

$$p_f^{low,up} = \frac{K_f}{N_s} \mp \Phi^{-1}(\gamma) \sqrt{\frac{\frac{K_f}{N_s} - \left(1 - \frac{K_f}{N_s}\right)}{N_s}} \tag{7.19}$$

where K_f is the number of failure events and N_s is the number of samples.

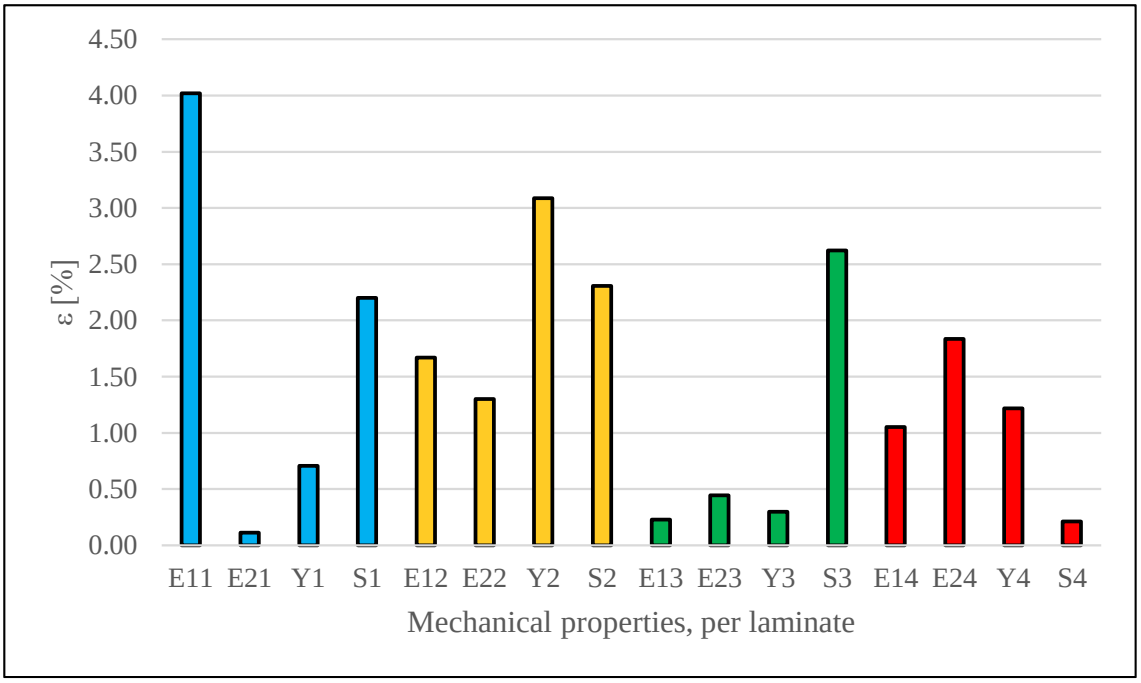

Figure 7.14: Relative difference, in percentage, of the coordinates of the MPP estimated by the HmGA and the gradient-based algorithm.

A total of N_s normally distributed random samples, in the uncertainty space, were generated using the Box-Muller transform [12]. Table 7.5 presents the values of β_{HL} of the representative solutions, together with the corresponding estimate of the probability of failure $p_{f_{RIA}} = \Phi(-\beta_{HL})$, the number of samples used for each solution N_s, the number of failure events K_f, the lower and upper bounds of the intervals with $\gamma = 95\%$ and the estimate of the failure probability calculated via the MCS, $p_{f_{MCS}}$.

Figure 7.16 plots the calculated confidence intervals and the position of the estimate $p_{f_{RIA}}$ within the interval, for the three representative Pareto optimal solutions. It is seen that the estimate $p_{f_{RIA}}$ of the three representative Pareto optimal solutions falls inside the calculated confidence intervals. From a frequentist perspective, 95% of the random confidence intervals, constructed from random samples, are expected to contain the unknown true value of p_f. The fact that the estimate $p_{f_{RIA}}$ lies inside a confidence interval is a good measure of the convergence of the HmGA.

Table 7.5: Input and output parameters of the MCS of three representative Pareto optimal solutions.

	β	$\Phi(-\beta)$	N_s	K_f	p_f^{low}	p_f^{up}	$p_{f_{MCS}}$
(1)	3.552	1.913E-04	$3E^6$	587	0.000182384	0.000208949	1.957E-04
(2)	4.002	3.144E-05	$3E^6$	90	1.67174E-05	3.52014E-05	3.000E-05
(3)	4.508	3.266E-06	$3E^6$	6	0	3.34302E-06	2.000E-06

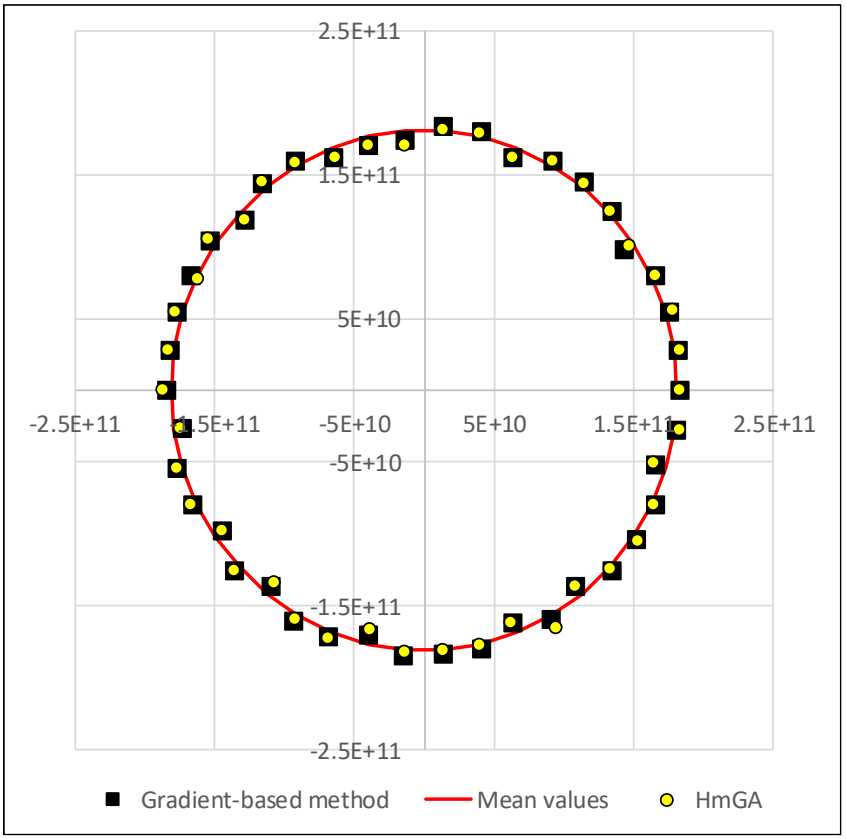

Figure 7.15: α-plots of the longitudinal Young modulus of the first laminate, E_{11}.

The pitfall regarding the crude MCS is the huge number of samples required to estimate very low probabilities and the computing times necessary to evaluate all of them. For the three solutions in Table 7.5, three million simulations were run. It took 35 hours to assess the probability of failure of each design solution. As seen in Table 7.5 and in Figure 7.16, solution (1) clearly does not require so many samples for the MCS analysis to converge. Yet, even a few hundred thousand samples would take several hours to be concluded, because of the need to consecutively run a finite elements model. On the other hand, solution (3) requires a lot more than three million simulations, since a very low number of failure events was observed. This happened, because the probability of failure is already extremely low for a reliability index around 4.5. However, the reliability index of most of the forty-two Pareto optimal solutions found by the RBRDO procedure, in the previous section, is higher than 4.5.

Computing times of this magnitude are incomprehensible, if reliability assessment is an inner cycle of design optimization. By comparison, it shows the efficiency and accuracy of the HmGA in the solution of the RIA, with computing times peaking at 12 minutes, for reliability indexes higher than 4.5. Even though the proposed method is slower than gradient-based algorithms, the advantages of EAs have to be taken into account. In particular, the possibility to achieve global convergence. The proposed evolutionary-based methodology is expected to set the basis for further developments on the design optimization of more complex systems, with multiple failure criteria, where gradient-based algorithms are expected to fail, or converge to local optima.

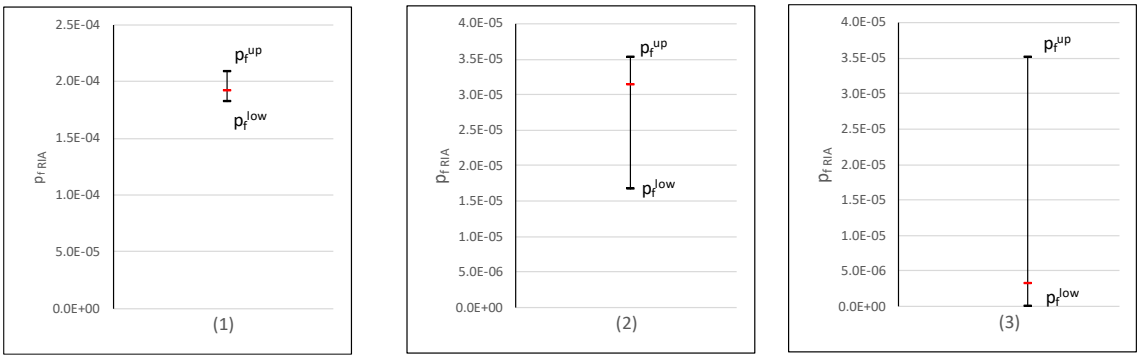

Figure 7.16: Confidence intervals (black) of $\boldsymbol{p_f}$, with $\boldsymbol{\gamma = 95}$%, and estimate $\boldsymbol{p_{f_{RIA}}}$ (red) of the representative Pareto optimal solutions.

7.2.3.2 RBRDO results

In order to observe the effects of the reliability constraint, the design optimization problem, is solved without reliability assessment (RDO), where $\boldsymbol{g_2(\mu_x, \mu_\pi)}$ is replaced by (7.15).

The proposed method proved to be efficient. It took 27 hours (in debug mode) to complete the optimization procedure, for the considered numerical example, using an Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz processor, for an elevated number of generations run (300 generations). The most expensive part of the RBRDO problem is again the reliability assessment, since the related RDO problem is solved in approximately 35 minutes. The main contribution to the efficiency of the method is the IGRO and the RSO operators, that allowed the HmGA to quickly identify the most promising regions of the uncertainty space. Without these operators, the solution of the RIA, via evolutionary search, would be a tedious task, most likely suffering from pre-convergence issues, and the design optimization impracticable.

Figure 7.17 plots the Pareto fronts (PF) obtained with and without the reliability assessment (RBRDO and RDO, respectively), after 300 generations of the MOGA-2D. As it is seen, the PFs are close to each other, meaning the optimal region of the objective space is of similar shape, in both the RBRDO and RDO problems. It is observed that the heaviest Pareto optimal solutions (70-90 kg) are slightly lighter in the RBRDO, while the lightest Pareto optimal solutions tend to be heavier, in the RBRDO.

Figure 7.18 shows the distribution of the critical Tsai number, along the PFs obtained with the RBRDO and the RDO formulations. It is seen that there are significant differences in the values of $\bar{R}$, for most Pareto optimal solutions. In the RBRDO, solutions have considerably higher values of $\bar{R}$ and the distribution of the along the PF is not as smooth as in the RDO. Furthermore, the reliability constraint does not allow the critical Tsai number to fall under the value of 1.1. In the RDO, the critical Tsai number can be as low as 1.0, satisfying the deterministic stress constraint.

Figure 7.19 shows the distribution of the β_{HL} calculated by the HmGA, along the PF. As it is seen, the reliability index is greater than $\beta^a = 3.0$, for every Pareto optimal solution. Again, comparing Figure 7.18 and Figure 7.19, it is possible to observe the parallelism between the distributions of $\bar{R}$ and β_{HL}, indicating that a functional relationship might exist between the two integrity measures. It is also seen that, for heavier solutions, β_{HL} varies between 3.0 and 10.5, while for lighter solutions it has varies between 3.0 and 5.0. Such triangular distribution of β_{HL} (and $\bar{R}$, as well) indicates that solutions with similar weight and variability can have very different integrity levels. As with the results in Sect. 7.2.2, we believe such distribution of the structural integrity measures is due to a many-to-one relationship between the feasible search space S_X and the feasible objective space S_Z. Further investigation is required.

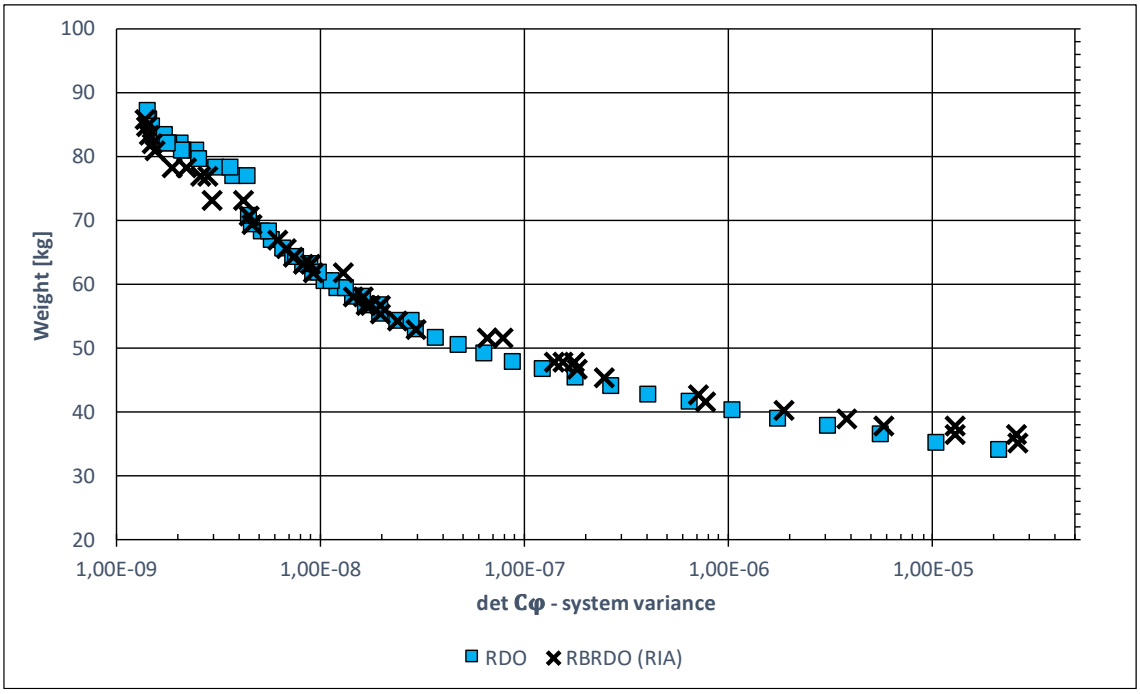

Figure 7.17: Pareto fronts obtained with and without reliability assessment (RIA).

In relation to the robustness of the system, it is interesting to see how it is affected by the reliability constraint. Figure 7.20 shows the distribution of the individual coefficients of variation of displacement and stress, given by (7.16), for the RBRDO and the RDO problems. These coefficients are used to analyze the individual contribution of the displacement and stress limit-state functions, in the total variability of the Pareto optimal solutions.

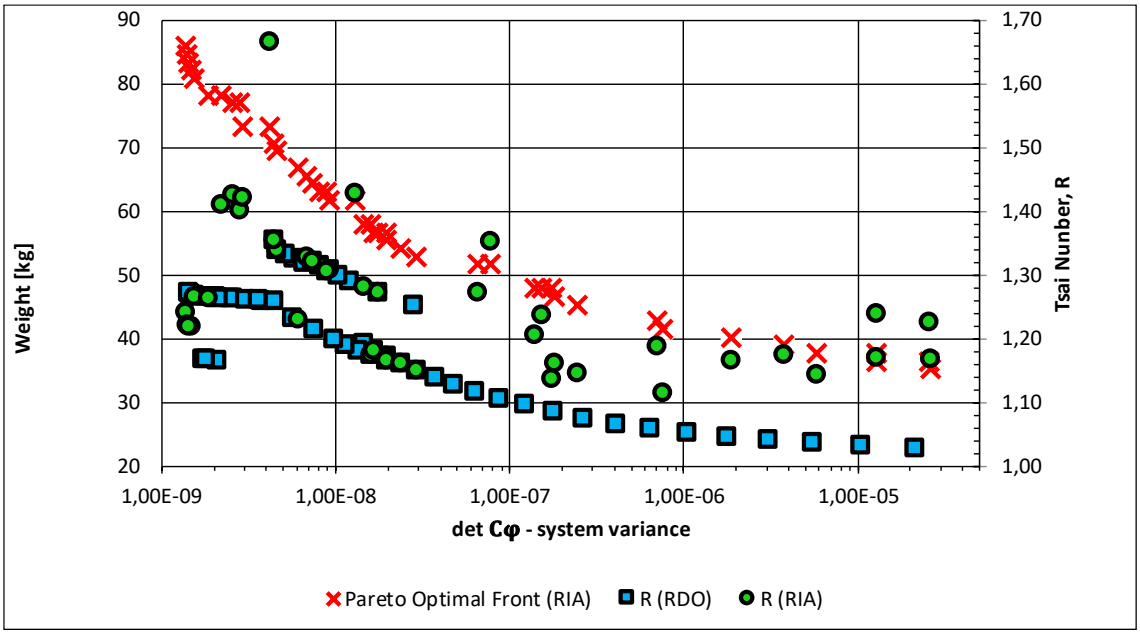

Figure 7.18: Distribution of the critical Tsai number, along the Pareto fronts, obtained with and without reliability assessment (RIA).

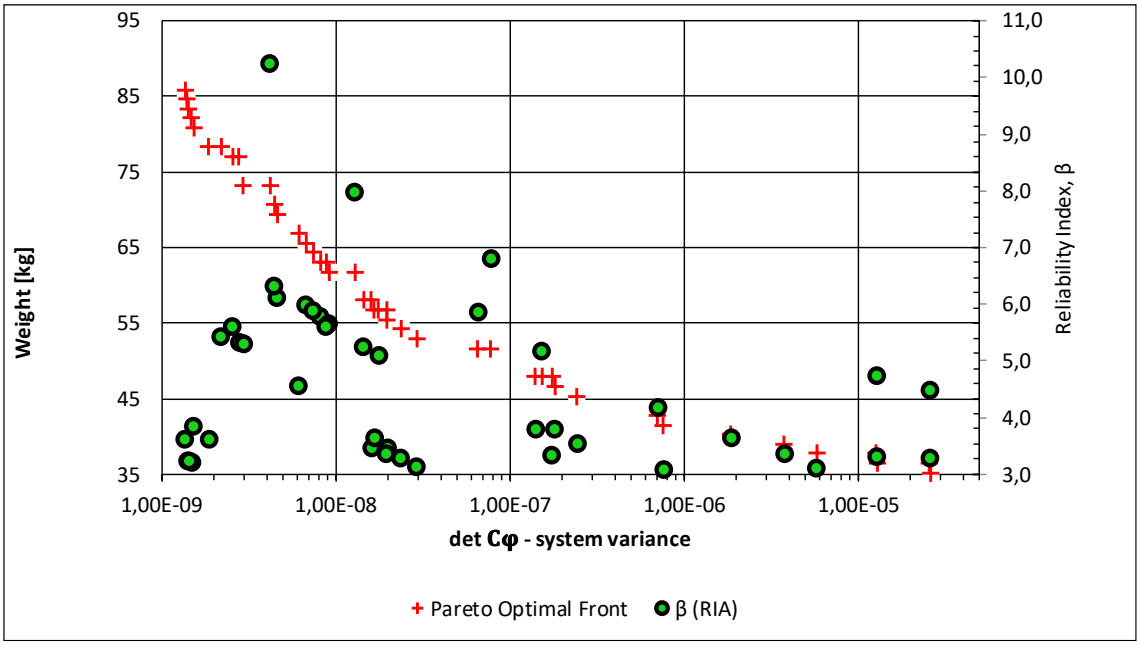

Figure 7.19: Distribution of the reliability index of the Pareto optimal solutions, calculated by the HmGA, along the Pareto front.

As it is seen, $CV(\bar{R})$ is larger but constant along the PF, with values about 25%. On the contrary, $CV(\bar{u})$ increases exponentially with the decrease in the total structural weight. It is concluded that the variability associated with the displacement response has a larger responsibility for the increase in the system variability (i.e., decrease of robustness) of the Pareto optimal solutions. Furthermore, as with the PMA, there is a positive trade-off between reliability and robustness, in the sense that Pareto optimal solutions have significantly improved levels of structural integrity, at the expense of a small reduction in the robustness associated with the stress response.

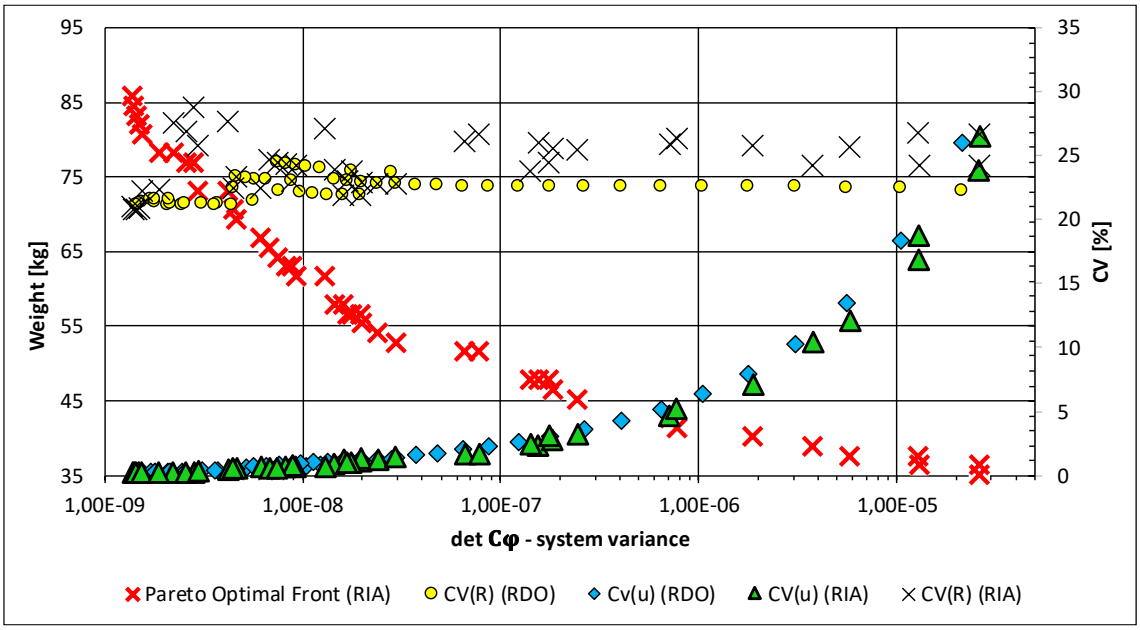

Figure 7.20: Distribution of the coefficients of variation, $\boldsymbol{CV(\bar{u})}$ and $\boldsymbol{CV(\bar{R})}$, along the Pareto front (RIA and RDO).

Regarding the Pareto optimal solutions, Figure 7.21 shows the distribution of the mean-values of the random ply-angles, along the PF. It is observed that the optimal orientation of the fibers is at 90°, for most Pareto optimal solutions, with the exception of six solutions, for which the optimal mean ply angle is of 84°.

Figure 7.22 shows the distribution of the thickness variables, along the PF. Again, the design optimization process evolved towards increasing mainly the thickness of the fourth laminate, h_4, to increase the total structural weight. The reason for this behavior would need further investigation. It is plausible to infer that that the chosen size-constraints may be too wide and that, in the absence of any geometrical constraint, the selective pressure, during the evolution determined an evolutionary path that favored this configuration.

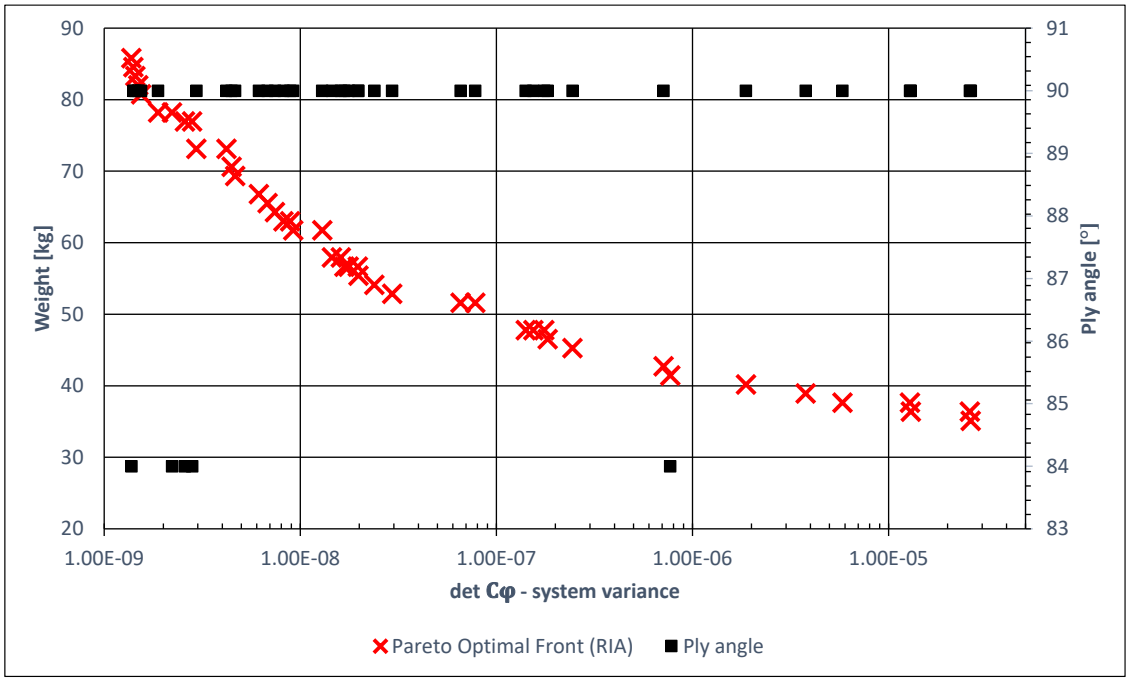

Figure 7.21: Distribution of the mean-values of the random ply-angles, along the Pareto front (RIA).

7.2.3.3 PMA vs. RIA in robust design optimization

As closing comments, we present a short discussion on the differences between solving the reliability assessment problem with the PMA or the RIA, for the current design optimization problem.

In terms of computing times, it is concluded that the same RBRDO problem is solved 1.5 times faster with the PMA, than with the RIA. As reported in the literature (for example, see [13]), the PMA is more efficient when probabilistic constraints are inactive, while the RIA performs better when the probabilistic constraint is violated. However, the PMA is unable to calculate the reliability index of the designs, during the optimization procedure (it needs to be calculated as a post operation), and the evolutionary process is never guided by the actual reliability level of the solutions.

It is interesting to observe that, also between both RBRDO problems, the distribution of the structural integrity measures $\bar{R}$ and β_{HL} are significantly difference, while the respective PFs are very close. Comparing Figure 7.3 and Figure 7.18, it is seen that the RBRDO problem with the PMA achieves a much smoother distribution of $\bar{R}$, with values confined to a much smaller interval. At the same time, the PMA demonstrates to be more conservative than the RIA, with lower values of $\bar{R}$.

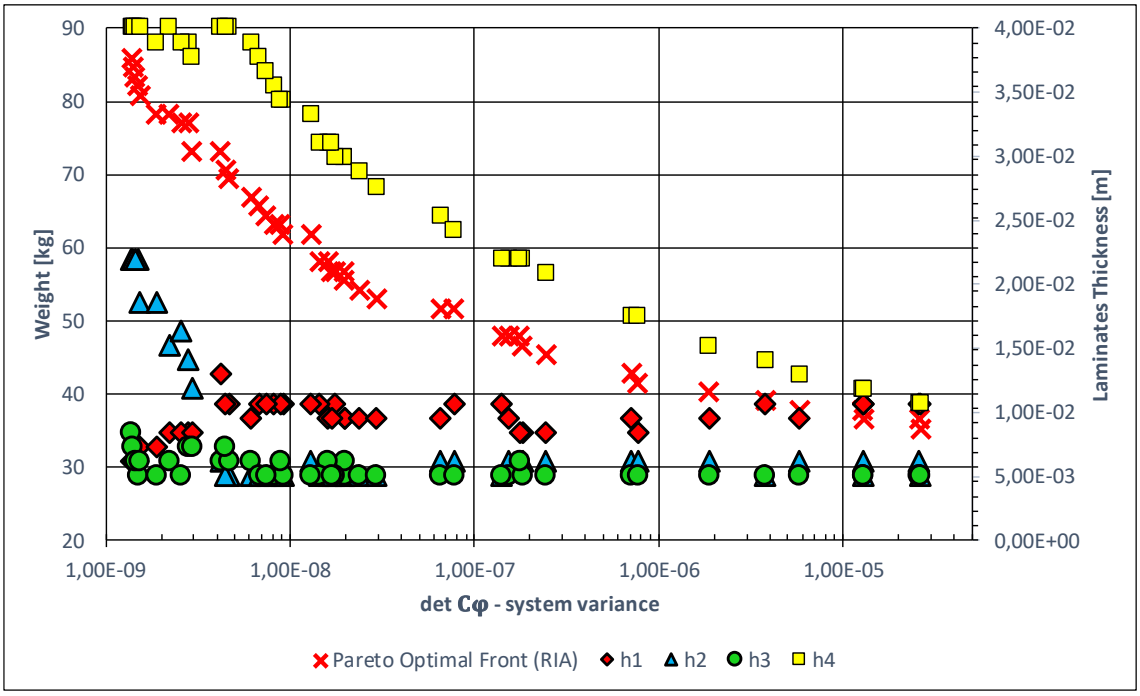

Figure 7.22: Distribution of the mean-values of the random thickness variables, of the four laminates, along the Pareto front (RIA).

Similar results were found for β_{HL}. With the PMA, the obtained distribution of the reliability index of the Pareto optimal solutions is smoother, than that obtained with the RIA. The values of the reliability index vary between smaller intervals, as well. It is also observed that there is a parallelism between the distribution of $\bar{R}$ and β_{HL}, as already mentioned above, meaning that Pareto optimal solutions with similar values of structural weight and variability can have significantly different levels of structural integrity.

Comparing Figure 7.8 and Figure 7.22, it is seen that the distribution of the mean-values of the random thickness design variables, along the PF, is also smoother with the PMA, than with the RIA, although the values are of the same magnitude. Yet, it is on the distribution of the thickness variables that the biggest differences can be found, particularly for the Pareto optimal solutions between 70 and 90 kg. For this weight range, the RBRDO with the RIA succeeded to find slightly lighter solutions, for which the decrease in the total weight results mainly from lighter second and fourth laminates. For the remaining Pareto optimal solutions, it is seen that the thickness values oscillate a little, between the two methods.

The similarity between the PF, obtained with the PMA and the RIA, compared with the significant differences in the respective distributions of $\bar{R}$ and β_{HL}, reveals the sensitivity of the measures of structural integrity to the random design variables, which is corroborated with the calculated values of $CV(\bar{R})$. Given the similarity between the ply angles and since the material properties are constant, the random thickness variables appear to be the most influent ones, on the values of $\bar{R}$ and β_{HL}. This is justified by the fact that the critical Tsai number is an inverse function of the cubic power of the laminate thickness. Regarding the configuration of the uncertainty space, it means that the failure surface and the origin of the uncertainty space are significantly altered with small variations in the thickness values, resulting in different reliability levels.

This behavior (Pareto optimal solutions of similar weight and variability displaying such distinct performance in the stress space) reveals perhaps two interesting phenomena: 1) the existence of a loose degree-of-freedom in the present design optimization problem; 2) an already reported [13]

accelerated convergence of design optimization with the PMA in relation to the RIA, in design optimization problems.

Issue 1) refers to the possible many-to-one relationship between the feasible search space S_X and the feasible objective space S_Z, that we mentioned before. From the observed data, it can be concluded that, for one structural weight value, there are many feasible combinations of the thickness variables, which will originate different critical points on the structure, possibly leading to different values of $\bar{R}$ and β_{HL}. How this translates to similar values in the system variability needs further investigation. Thus, in the true sense of *optimization*, for a given optimal realization in the objective space, the combination of the thicknesses of the four laminates of the structure should perhaps be optimized, as well, to match a specific common criterion (or at least constrained). For example, to maximize structural integrity, within a specific weight value. Regarding issue 2), it is possible that, albeit the large number of generations run, the RBRDO with the RIA would still need more generations to achieve a better degree of convergence that would allow to have smoother distributions along the PF, like those obtained with the PMA. Yet, Issue 2) does not alter the fact that issue 1) should be also analyzed.

7.3 Effects of different sources of uncertainty in RBRDO

In this section, we conduct a study on the effects of different sources of uncertainty upon the solutions of the RBRDO problem stated in Definition 7.1. More precisely, we aim to understand how different sources of uncertainty affect the *feasibility robustness* of composite laminate structures. The solution of the bi-objective outer cycle of stochastic design optimization is solved by the MOGA-2D. Structural reliability is evaluated by the PMA, here solved by the mGA, since the results in the previous section demonstrated that the PMA allowed smoother distributions of the measured quantities along the PF and such quality favors the comparison of different results. The parameters of both algorithms are defined in Table 7.3 (a) and (b). The reliability constraint is written in terms of the piecewise equivalent PMA limit-state function, in (5.26). The allowable reliability index is $\beta^a = 3.0$ and the allowable critical displacement is $u^a = 8 \times 10^{-2}$ [m].

We start by introducing a new physical model and follow with a discussion of the results obtained.

7.3.1 Physical model

As a numerical example consider an aircraft fuselage-like composite shell structure, with a window and five degrees of freedom (three displacements and two rotations), as shown in Figure 7.23. Assuming it is integrated in a complete structure, and not isolated, diversified boundary conditions are considered, to allow for some flexibility and emulate the connection between consecutive components.

The shell structure is fully supported on points A and B and supported along the x and z-axis between them. On points C and D, it is supported along the x and y-axis. Between points A and C and between B and D, the shell is simply supported along the y-axis. To induce vertical bending, the structure is only simply supported along the x-axis, between points C and D (see appendix B). To allow for only a certain level of bending deformation on the window and on the side CD, displacement constraints are considered on pre-determined points, marked with red dots in Figure 7.23. From this set of points, the point with maximum displacement is the one defining the displacement constraint.

In order to simulate the exterior pressure acting of the fuselage, nodal (point) loads of equal magnitude $P_k = 11{,}5\ KN$, always normal to the surface, are applied on every node. Point loads acting as

distributed loads are applied for numerical convenience. At every node, the load vector is projected onto the x and z-axis.

The structure consists of four laminates, corresponding to four macro-elements with total thickness variables, h_i, for $i = 1,2,3,4$, such that each layer has equal thickness. A balanced angle-ply laminate with five layers distributed in the symmetric stacking sequence $[+\theta/-\theta/0°/-\theta/+\theta]$ is considered for each macro-element. The ply-angle θ is referenced to the x-axis and is a design variable of the structure, meaning it assumes the same value for all laminates. Material properties are those of a carbon/epoxy composite system denoted T300/N5208 [8]. The elastic constants of each orthotropic ply are the longitudinal elastic modulus E_1, the transversal elastic modulus E_2 , the in-plane shear modulus G_{12} and the in-plain Poisson's ratio ν_{12}. The ply strength properties are the longitudinal tensile strength X, the longitudinal compression strength X', the transversal tensile strength Y and the transversal compression strength Y' and the shear strength S. The mean-values of the mechanical properties are defined in Table 7.1.

According to António *et al.* [9], the most important parameters were shown to be E_{1j}, E_{2j}, Y_j and S_j. Therefore, the vector of random mechanical properties, $\mathbf{m}$, only includes the random parameters E_{1j}, E_{2j}, Y_j and S_j, where the subscript $j = 1,2,3,4$ denotes the laminate number. Hence, there are sixteen random mechanical properties aggregated in vector $\boldsymbol{\pi}$, plus nine random vertical loads. Five random design variables are considered in vector $\mathbf{x}$: one ply-angle θ, equal for all symmetric laminates, and four laminates thickness variables h_i, for $i = 1,2,3,4$. The constant values of the standard deviations of the random variables are given by Table 7.2.

Several sets of sources of uncertainty are tested. We call an *uncertainty set* to any set containing from one to several combinations of different sources. The following sets are considered in this study:

$$\{\mathbf{m}\}, \{\mathbf{m}, \mathbf{h}\}, \{\mathbf{m}, \theta\}, \{\mathbf{m}, \mathbf{h}, \theta\}, \{\mathbf{m}, \mathbf{h}, \theta, \mathbf{p}\} \tag{7.20}$$

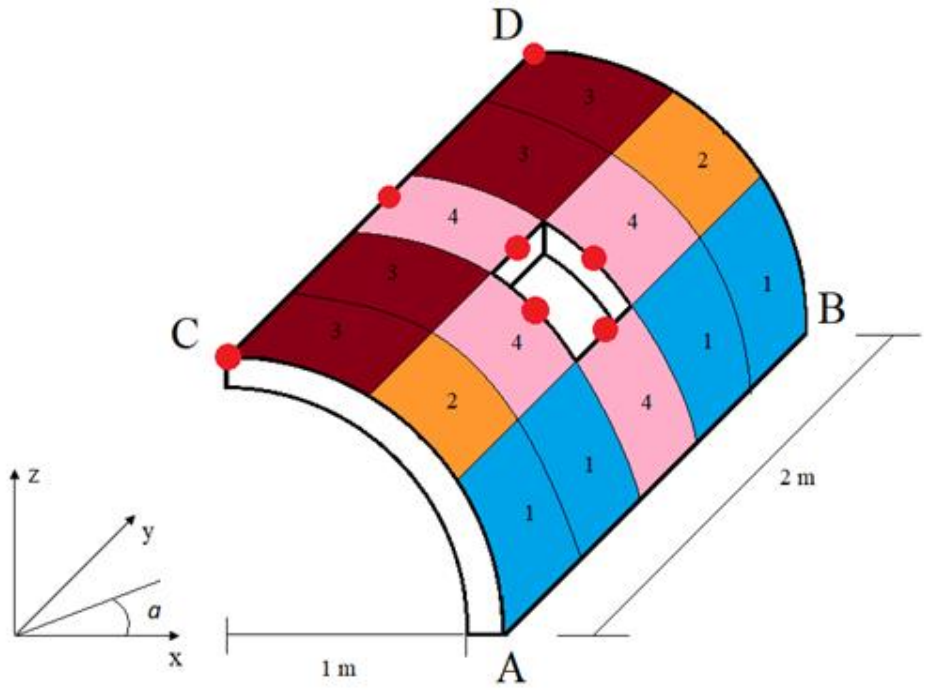

Figure 7.23: Geometric definition of a fuselage-like shell structure, distribution of the laminates and representation of the points of evaluation of the displacement constraint (red dots).

The various uncertainty sets are considered to only have explicit influence on the feasibility robustness, while reliability assessment is only affected by the set of random mechanical properties, by the same motives explained in Sect 7.2. Further studies on the explicit influence of different sources of uncertainty on the reliability assessment and on the combined structural feasibility and reliability assessment of composite laminate structures need to be taken in the future.

7.3.2 Results

The proposed method proved to be efficient, as it took around 12 hours to complete the design optimization procedure, for the presented numerical example, using an Intel(R) Core(TM) i7-6700 CPU @ 3.40Ghz processor. The results obtained show interesting behaviours that appear to be specific to the current RBRDO problem and physical model. Therefore, these should be considered valid under their own circumstances. Nevertheless, some general, but important, conclusions about the stochastic design optimization of composite laminate composites can be made, from the next discussion.

Figure 7.24 displays the different PFs obtained with the tested uncertainty sets. For the current numerical example, the random mechanical properties appear to have a very reduced influence on feasibility robustness, when compared to the other random variables. On a linear scale, as in the figure, the PF obtained with the uncertainty set {**m**} is almost vertical and the Pareto optimal solutions display almost null system variability. It means that the components of the variance-covariance matrix $\mathbf{C}_{\boldsymbol{\Phi}}$, related with the mechanical properties, have very low values, because the related derivatives (sensitivities) are close to zero. A *possible* justification arises: the reliability constraint, being defined solely for {**m**}, possibly acts as an implicit constraint over the importance of the random mechanical properties on the system variability. In other words, by restraining structural integrity to a minimum level of probabilistic reliability, in terms of the mechanical properties, it may be that the variability associated with these random variables becomes almost null, since failure becomes associated with events on the tail of the associated probability distribution.

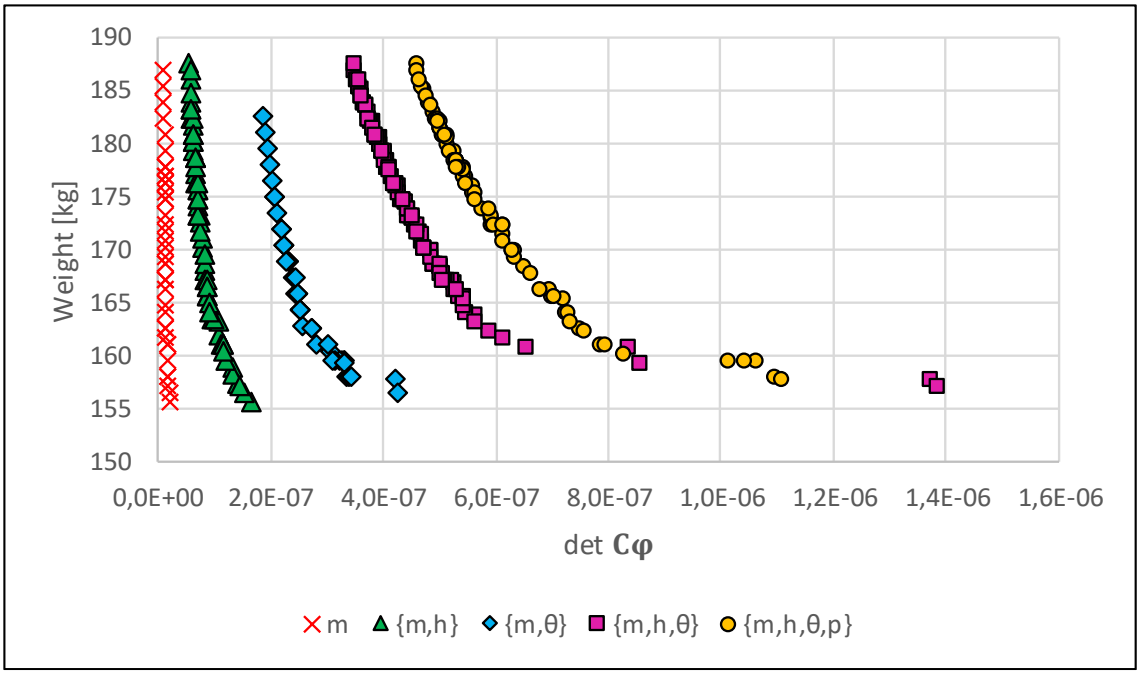

Figure 7.24: Pareto fronts of the tested uncertainty sets (linear scale).

The other PFs, in Figure 7.24, show increasing values of the system variability with the cardinality of the uncertainty sets, as would be expected. From the uncertainty sets $\{\mathbf{m}, \mathbf{h}\}$ and $\{\mathbf{m}, \theta\}$ result small to moderate increases in the value of $\det \mathbf{C}_{\boldsymbol{\phi}}$, respectively. However, the uncertainty set $\{\mathbf{m}, \mathbf{h}, \theta\}$ results a significant increase of the system variability, higher than the sum of the parts. This phenomenon is possibly caused by the existence of synergetic effects between the displacement and stress responses, that amplify the determinant of the variance-covariance matrix. That is, although the components of $\mathbf{C}_{\boldsymbol{\phi}}$ are linearized (see (5.5) and (5.6)), it follows that:

$$\det \mathbf{C}_{\boldsymbol{\phi}} = \mathrm{Var}(\bar{u})\mathrm{Var}(\bar{R}) - \mathrm{Cov}(\bar{u}, \bar{R})^2 \tag{7.21}$$

is indeed a quadratic measure of robustness and, thus, the difference between the product $\mathrm{Var}(\bar{u})\mathrm{Var}(\bar{R})$ and $\mathrm{Cov}(\bar{u}, \bar{R})^2$ increases nonlinearly, when more random variables are added to the equation. Furthermore, by increasing the number of uncertainty variables in feasibility robustness, there is the possibility that Pareto optimality is satisfied for different critical points of displacement and stress with different values of the respective sensitives. Further studies are needed.

Finally, as expected, the uncertainty set $\{\mathbf{m}, \mathbf{h}, \theta, \mathbf{p}\}$ produces the highest values of $\det \mathbf{C}_{\boldsymbol{\phi}}$. The increase in the system variability, from $\{\mathbf{m}, \mathbf{h}, \theta\}$ to $\{\mathbf{m}, \mathbf{h}, \theta, \mathbf{p}\}$, is just moderate and it is not possible to assess if it is due to the variance of the random point-loads or to any synergetic effect. As a last comment, it is seen that despite the increase of the system variability, the weight range of the Pareto optimal solutions is constant, for all the uncertainty sets, varying between 155 from 190 kg.

Regarding the distribution of the critical Tsai number and the Hasofer-Lind reliability index, along the PF, for the different uncertainty sets, it is observed that, first, the distributions of both $\bar{R}$ and β_{HL} are not significantly changed by the uncertainty sets, and, second, that there is a similarity between the shape of the distributions of both integrity measures (as found in Sect. 7.2). For this reason, in the following discussion, we support the description of the results with the most representative graphics. The remaining can be found in Appendix A.

In order to demonstrate the effects of the reliability constraint, we compare the distributions of the critical Tsai number, along the PF, obtained with RDO and RBRDO. Table 7.6 presents the lower and upper bounds of the values of $\bar{R}$, obtained with the tested uncertainty sets. It is seen that the reliability constraint imposes higher lower bounds of $\bar{R}$, than those obtained with RDO, Furthermore, it is seen that in the RBRDO the lower bound of $\bar{R}$ increases with the cardinality of the uncertainty sets, while in the RDO the highest lower bound happens for $\{\mathbf{m}, \theta\}$. On the other hand, the upper bound of $\bar{R}$ is not significantly altered. In terms of shape, the distributions of $\bar{R}$ assumes a V-shape, for all uncertainty sets. As a representative example, shown in Figure 7.25, consider the case of $\{\mathbf{m}, \mathbf{h}\}$.

Table 7.6: Lower and upper bounds of $\bar{R}$, obtained with RDO and RBRDO, for the tested uncertainty sets.

	RDO		RBRDO	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
$\{\mathbf{m}\}$	1.05	1.47	1.23	1.55
$\{\mathbf{m}, \mathbf{h}\}$	1.11	1.51	1.22	1.55
$\{\mathbf{m}, \theta\}$	1.19	1.49	1.24	1.51
$\{\mathbf{m}, \mathbf{h}, \theta\}$	1.18	1.51	1.26	1.57
$\{\mathbf{m}, \mathbf{h}, \theta, \mathbf{p}\}$	1.14	1.51	1.29	1.51

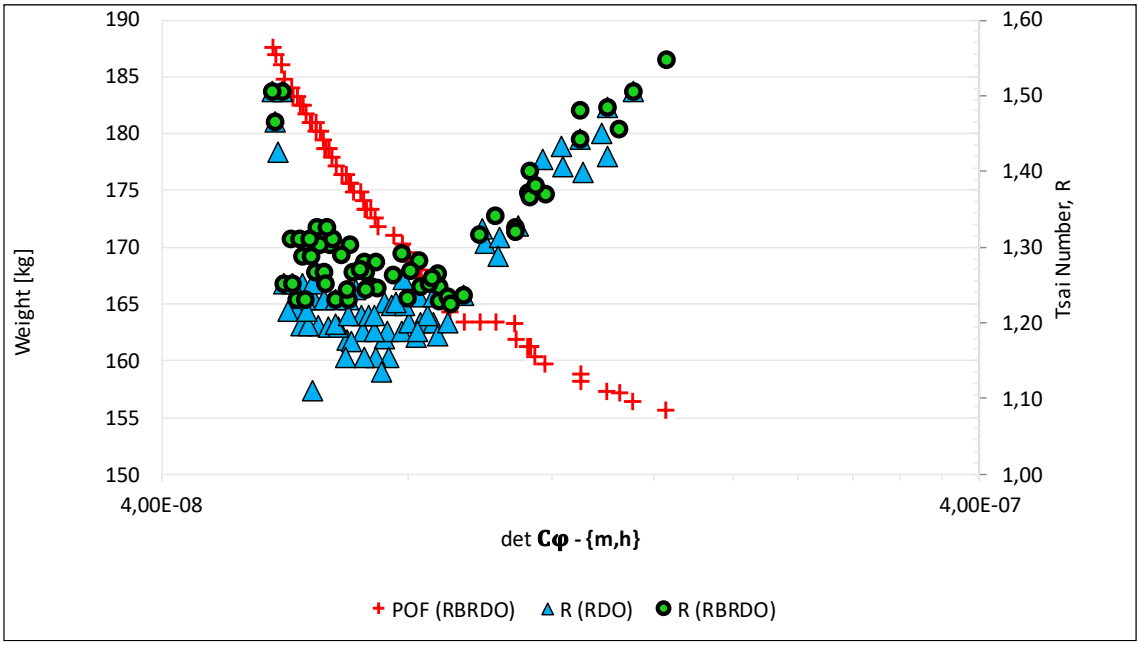

Figure 7.25: Distribution of the critical Tsai number, along the Pareto front, obtained with the uncertainty set $\{\mathbf{m}, \mathbf{h}\}$.

Although not represented in the figure, there is a good match between the PFs obtained with RBRDO and RDO. Like the results obtained in Sect. 7.2, it means that there is a many-to-one relationship between the feasible search space S_X and the feasible objective space S_Z.

Regarding the distribution of β_{HL}, there is a clear parallelism with the distribution of $\bar{R}$, as shown in Figure 7.26. This feature is observed for the tested uncertainty sets. As seen in Table 7.7, the lower bound of β_{HL} increases with the cardinality of the uncertainty sets. The values of β_{HL} were calculated by a gradient-based algorithm [4,10]. Due to technical difficulties, the reliability index of the Pareto optimal solutions was not calculated, for $\{\mathbf{m}, \mathbf{h}, \theta\}$.

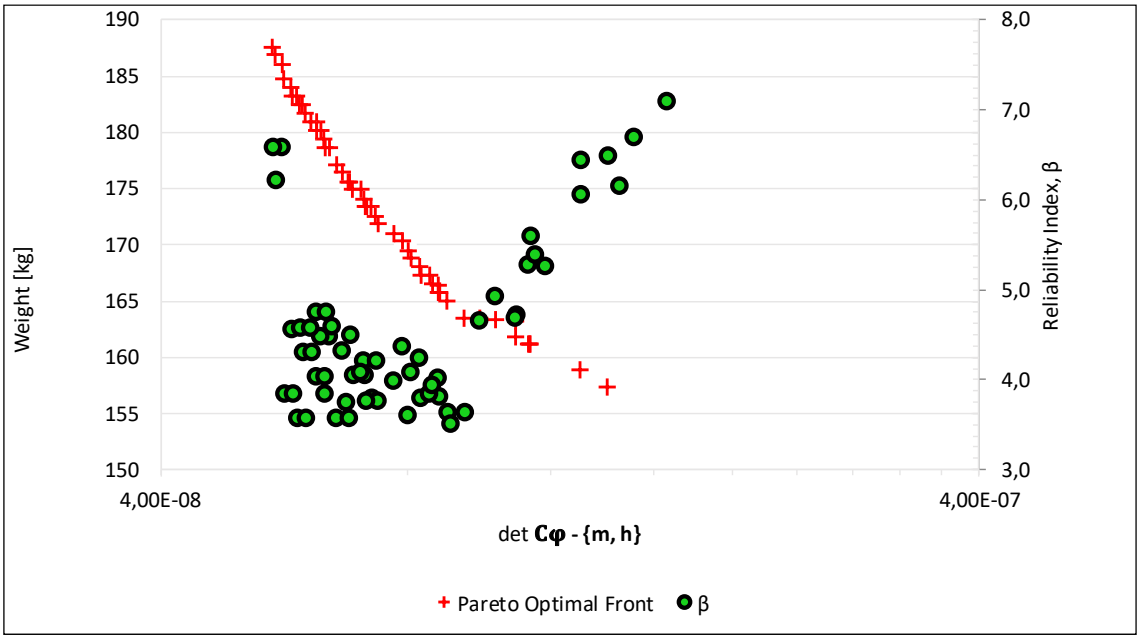

Figure 7.26: Distribution of the Hasofer-Lind reliability index, along the Pareto front, obtained with the uncertainty set $\{\mathbf{m}, \mathbf{h}\}$.

Table 7.7: Lower and upper bounds of $\boldsymbol{\beta_{HL}}$, obtained with RBRDO, for the tested uncertainty sets.

	RBRDO	
	Lower Bound	Upper Bound
$\{\mathbf{m}\}$	3.57	7.09
$\{\mathbf{m}, \mathbf{h}\}$	3.51	7.09
$\{\mathbf{m}, \theta\}$	3.71	6.63
$\{\mathbf{m}, \mathbf{h}, \theta\}$	-	-
$\{\mathbf{m}, \mathbf{h}, \theta, \mathbf{p}\}$	4.31	6.59

The V-shape seen in the distributions of both $\bar{R}$ and β_{HL}, for all uncertainty sets, indicates that the solutions in the extremes of the PF have higher values of structural integrity than the intermediate ones. While on Figure 7.3 and Figure 7.4 we see sudden jumps on the values of $\bar{R}$ and β_{HL}, these are followed by a decrease of the respective values. Instead, for the current physical mode, we see an increasing tendency in the lower values of system variability. Although further tests are needed, to make proper conclusions on why this behaviour happens, we try to explain this phenomenon with the available results.

Figure 7.27 shows the distribution of the random ply-angle θ, along the PFs, for the tested uncertainty sets. From this figure, we can relate the increase in the values of $\bar{R}$ and β_{HL} with the decrease in the value of θ. In general, there is a tendency to decrease the values of θ until 0º, except for the case $\{\mathbf{m}, \theta\}$, to which we will refer later. For now, consider the distribution relative to the uncertainty set $\{\mathbf{m}, \mathbf{h}, \theta, \mathbf{p}\}$. In this case, we find that all Pareto optimal solutions have an optimal ply angle of 12º, except for the five solutions with the highest system variability (lowest weight), for which $\theta = 0$º. As it can be seen in Figures A.12 and A.13, these are precisely the five clustered solutions having abnormally high values of $\bar{R}$, from 1.44 to 1.47, and β_{HL}, from 6.04 to 6.38. Similarly, for $\{\mathbf{m}, \mathbf{h}, \theta\}$, we have four Pareto optimal solutions for which $\theta = 0$º. As seen in Figure A.10, these have the highest values of $\bar{R}$, from 1.53 to 1.56. Furthermore, there is one solution with $\theta = 6$º, and $\bar{R} = 1.42$.

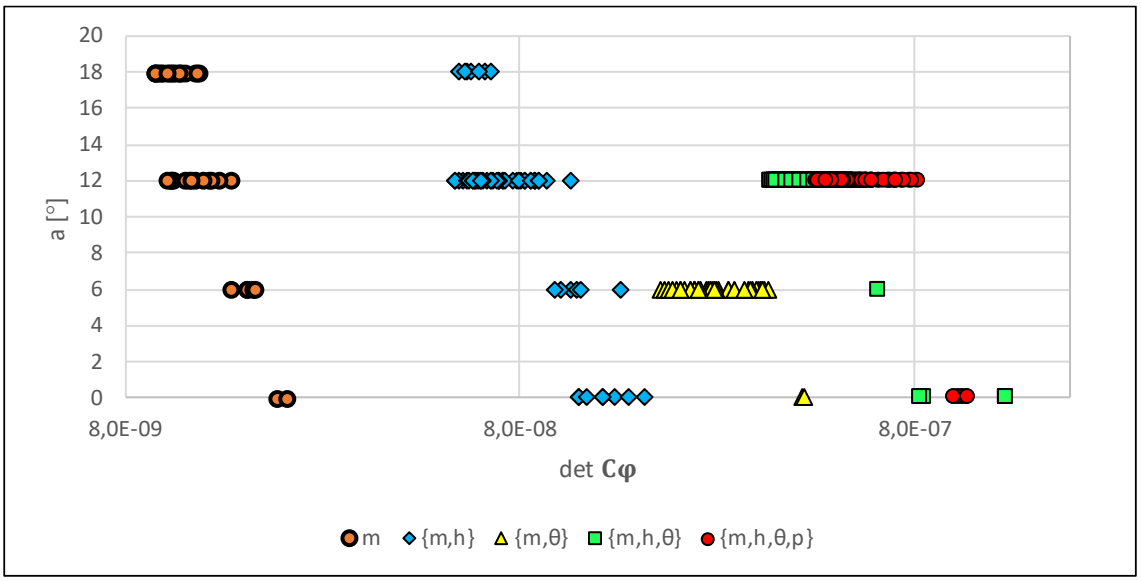

Figure 7.27: Distribution of the random ply-angle, along the Pareto front, for the tested uncertainty sets.

In our explanation, the decrease of the value of the ply-angle until 0º justifies the increase of the integrity parameters, in the right side of the PF. On the left side, where Pareto optimal solutions are heavier and the ply angle has higher values, we find that structural strength is mostly determined by the thickness of the laminates. Due to its clearance, consider the distribution of the thickness variables along the PF, for the case $\{\mathbf{m}\}$, shown in Figure 7.28. It is seen that most of the structural strength is assured by laminates 2 and 4 (see Figure 7.23), which preserve the maximum value of thickness (0,040 m), forming a kind of central frame. On the other hand, laminate 3 decreases its thickness until moderate values (0.032 m) and laminate 1 until lower values (0.027 m). Thus, it suggests that until a certain point, the structural strength is influenced by the laminate thickness, hence decreasing in parallel with the decrease of the thickness variables $h_1$and h_3, after which the ply-angle θ approaches 0º and the structural strength increases.

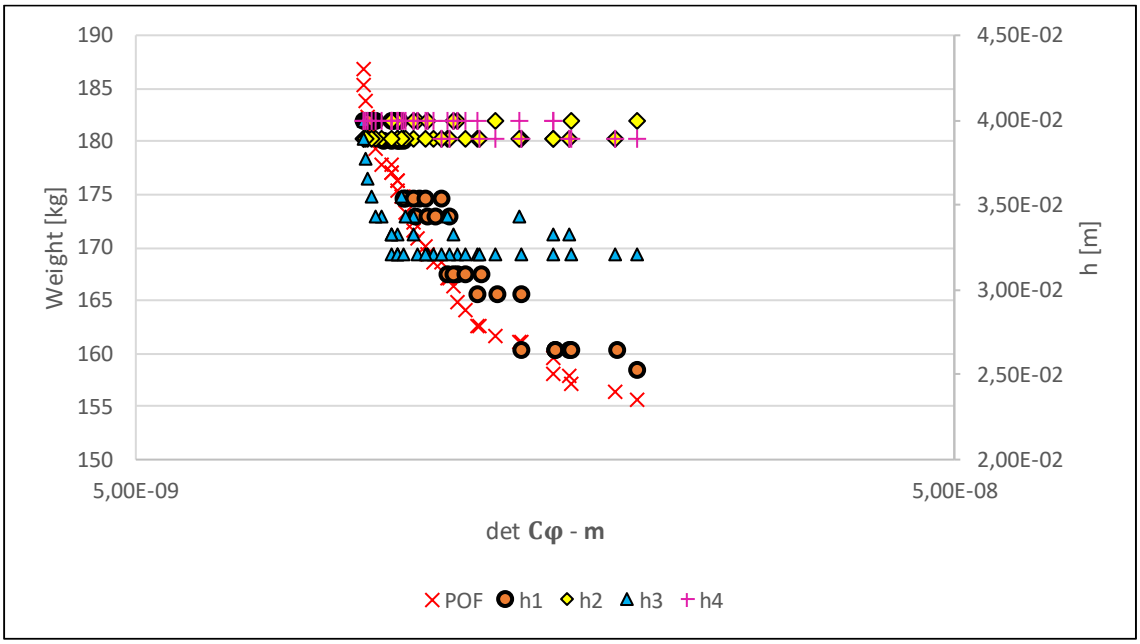

Figure 7.28: Distribution of the random thickness variables, along the Pareto front, obtained with the uncertainty set $\{\mathbf{m}\}$.

Similar behaviors can be found, for the remaining uncertainty sets, the exception being $\{\mathbf{m}, \theta\}$. In this case, only one Pareto optimal solution has $\theta = 0º$, while all the others have $\theta = 6º$. However, the distributions of $\bar{R}$ and β_{HL} still follow the same pattern (see Figures A.7 and A.8). As seen in Figure A.9, the optimal thickness values of laminate 2 have lower values (0.031 m) and increase until large values (0.39 m). For the case $\{\mathbf{m}, \mathbf{h}, \theta\}$, structural strength is guaranteed by laminates 2 and 3, instead, while the thickness of laminates 1 and 4 decreases. The increase of structural strength, in the lighter designs, is still related with lower values of θ.

The V-shape of the distribution of the Tsai number and the reliability index allows to conclude that it is possible to obtain lighter composite laminate structures, if the fibers are aligned with the optimal ply-angle. However, in the particular case of this numerical example, it shows as well that the design optimization process has a loose degree-of-freedom. Clearly, for all the tested uncertainty sets, the heaviest Pareto optimal solutions, although optimal, could have higher values of structural integrity if the respective optimal ply-angles were set at 0º.

7.4 Application of Sobol' indices to reliability assessment

On Chapter 6, we developed an analytical solution of the Sobol' indices, based on two theoretical results relating k-th order multilinear Taylor polynomials of multivariate stochastic functions of independent random variables with their own ANOVA functional decomposition. These results allow to replace the complex integration process of the ANOVA decomposition by point-wise estimates of the derivatives of the stochastic response functionals, offering a local estimate of the Sobol' indices. We then proposed a dimensional reduction criterion, for reliability assessment (see Definition 6.2), that only requires an additional adjoint system of equations to be solved, by the Adjoint Variable Method (see Sect 7.1).

In this section, the analytical solution of Sobol' indices and the related dimensional reduction criterion are applied to the HmGA. The goal is to improve the convergence rate of the algorithm, by reducing the expected computing times in finding the global optimal solution of the RIA. This study is conducted over twenty (20) of the forty-two (42) Pareto optimal solutions that were used to validate the HmGA, in Sect. 7.2.3.1. These are called *reference solutions*. The structural reliability of the said solutions was guaranteed by the PMA, as an inner cycle of the RBRDO problem solved in Sect. 7.2.2.

Naturally, the physical model considered in this study is the one introduced in Sect. 7.2.1. As before, in reliability assessment, the uncertainty of the system is propagated only through the mechanical properties, $\mathbf{m}$, for which the mean values, $\boldsymbol{\mu}_{\mathbf{m}} \in \Pi$, and the standard deviations, $\boldsymbol{\sigma}_{\mathbf{m}}$, are constant. Furthermore, the inner evaluation of the limit-state function is performed for fixed values of the mean-values of the random design variables, $\boldsymbol{\mu}_{\mathbf{x}}$, and the mean values of the random design parameters, $\boldsymbol{\mu}_{\boldsymbol{\pi}}$, where the random mechanical properties are included as well. The standard deviations $\boldsymbol{\sigma}_{\mathbf{x}}$ and $\boldsymbol{\sigma}_{\boldsymbol{\pi}}$ are constant. The applied geometric discretization of the physical system implies the calculation of the stress field in specific integration points identified by the coordinates (e, p, k), where $e = 1, \dots, 8$ is the element number (see Figure 7.1), $p = 1, \dots, 5$ is the ply number (bottom to top) of the element e and $k = 1, \dots, 4$ is the integration point number of the ply p. The Tsai number is calculated on each of these points. As in Sect. 3.4.2, the stress response functional is evaluated in the critical point of the structure, where the stress state is the most aggravated.

The HmGA is now instrumented with the proposed analytical dimensional reduction criterion (see Definition 6.2). When the threshold parameter ε is null, i.e. $\varepsilon = 0$, all the random mechanical properties are classified as important. Then, $n_r = N_{\mathbf{m}} \Rightarrow \Omega_r = \Omega_2$ and the HmGA performs equally to its original version. On the other hand, to perform the proposed analytical dimensional reduction, it is considered that $\varepsilon = 10^{-3}$. This small value is explained with the need to include the majority of the uncertainty space, guarantying a good accuracy of the reliability assessment. At the same time, it allows a considerable dimensional reduction of the problem. This subject is discussed in the next section. Table 7.8 presents the mean-values of the random design variables, $\boldsymbol{\mu}_{\mathbf{x}}$, of the reference solutions, and the respective reliability index, as calculated by the HmGA, for $\varepsilon = 0$, in Sect. 7.2.3.1. The solutions are labelled and sorted in ascending order of their reliability index value. The parameters of the HmGA are found in Table 7.4.

On the next section, we present the results obtained by the improved HmGA. We begin by presenting the calculated Sobol' indices and discuss how uncertainty is explained by the random mechanical properties. We also identify the three factors that seem to influence the most the observed explanation of uncertainty. Finally, we analyse the performance of the HmGA equipped with and without the proposed dimensional reduction criterion. Both the accuracy and the efficiency of the algorithm are demonstrated.

Table 7.8: Mean-values of the random design variables and $\boldsymbol{\beta_{HL}}$ of the reference Pareto optima solutions, for $\boldsymbol{\varepsilon = 0}$, after convergence of the HmGA. For comments on the values of $\boldsymbol{\beta_{HL}}$, for $\boldsymbol{\varepsilon = 10^{-3}}$, see the next section.

Design Solution	θ	h_1 $[m]$	h_2 $[m]$	h_3 $[m]$	h_4 $[m]$	β_{HL} $(\varepsilon = 0)$	β_{HL} $(\varepsilon = 10^{-3})$
1	90º	6.13E-03	2.31E-02	5.00E-03	3.44E-02	3.665	3.626
2	90º	6.13E-03	2.31E-02	5.00E-03	3.66E-02	3.675	3.664
3	90º	6.13E-03	2.31E-02	6.13E-03	3.66E-02	3.710	3.674
4	90º	6.13E-03	2.31E-02	5.00E-03	3.77E-02	3.730	3.682
5	90º	1.06E-02	5.00E-03	5.00E-03	2.31E-02	3.900	3.898
6	90º	1.06E-02	5.00E-03	5.00E-03	2.42E-02	4.023	4.012
7	90º	1.06E-02	5.00E-03	5.00E-03	2.65E-02	4.280	4.268
8	90º	1.06E-02	5.00E-03	5.00E-03	2.53E-02	4.310	4.140
9	90º	1.06E-02	5.00E-03	5.00E-03	2.76E-02	4.415	4.399
10	90º	1.06E-02	6.13E-03	5.00E-03	1.29E-02	4.496	4.487
11	90º	1.06E-02	5.00E-03	5.00E-03	2.98E-02	4.642	4.636
12	90º	1.06E-02	6.13E-03	5.00E-03	1.40E-02	4.647	4.630
13	90º	1.06E-02	6.13E-03	5.00E-03	1.52E-02	4.780	4.766
14	90º	1.06E-02	5.00E-03	5.00E-03	3.21E-02	4.861	4.855
15	90º	1.06E-02	5.00E-03	5.00E-03	3.32E-02	4.963	4.954
16	90º	1.06E-02	6.13E-03	5.00E-03	1.74E-02	5.029	5.026
17	90º	1.06E-02	5.00E-03	5.00E-03	3.55E-02	5.154	5.131
18	90º	1.06E-02	5.00E-03	5.00E-03	3.66E-02	5.209	5.218
19	90º	1.06E-02	5.00E-03	6.13E-03	3.89E-02	5.313	5.308
20	90º	1.06E-02	5.00E-03	6.13E-03	4.00E-02	5.394	5.374

7.4.1 Results

An Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz processor was used. First, it is important to clarify some concepts:

- *Reference solutions* are the optimal designs, used as reference in this study;
- *Solutions* are those found by the HmGA, in the uncertainty space, during the reliability assessment search process;
- *(Global) convergence* is the status achieved by the HmGA, when the criterion in equation (4.40) is met, independently of the value of ε; *Superior accuracy* is the status achieved by the HmGA, for $\varepsilon = 10^{-3}$, when the reliability index of at least one solution is less than or equal to the one found for $\varepsilon = 0$, after convergence, for the same reference solution;
- t_{total} is the time, measured in minutes or generations, spent by the HmGA to achieve convergence, for any ε; t_1 is the time, measured in minutes or generations, spent by the HmGA to find the first reduced search region; $t_{\leq}$ is the time, measured in minutes or generations, after t_1, spent by the HmGA to achieve superior accuracy, for $\varepsilon = 10^{-3}$;
- *Dimensional reduction* of the uncertainty space refers to the decrease in the number of random variables, in the reliability assessment problem, after the Sobol' importance analysis;
- *Reduction of the search space* refers to the decrease of the amplitude of the size-constraints, caused by the RSO, during the search process.

The application of the proposed analytic dimensional reduction technique proved to be very successful. The efficiency and accuracy of the evolutionary-based reliability assessment were greatly improved, meaning that estimates of β and $\mathbf{y_{MPP}}$ were faster to obtain and better calculated with $\varepsilon = 10^{-3}$ than with $\varepsilon = 0$. This outcome is a consequence of a significant reduction of the uncertainty space, by elimination of several random variables that have null, or almost null, interference in the probabilistic failure of the structural system.

Note that the calculated Sobol' indices are here estimated after a first-order approximation of the limit-state function and are therefore local estimates of a global measure. Higher-order estimates could have been considered, but were unnecessary. The local quality of the indices suites the current problem, since only one adjoint system of state equations had to be solved. Such computational effort is negligible. The required gradient information is already calculated by the HmGA, on its original instantiation, to calculate the estimate β^*, in (4.41). Furthermore, if integrated in the proposed RBRDO, gradient information is already available from the calculation of the determinant of the variance-covariance matrix, in (5.7). Table 7.9 presents the calculated Sobol' indices, with respect to each random variable, for every reference solution. The percentage of uncertainty explained by the important random variables (highlighted) and the location of the critical point of the structure is also shown.

It is seen that, for all the reference solutions, the value of n_r is significantly smaller than $N_{\mathbf{m}}$ (=16), varying between 5 and 8 important variables. In terms of the performance of the HmGA, it means that the number of possible genetic combinations is exponentially reduced, increasing the possibility of the evolutionary process to find its path towards global convergence faster. However, as it is seen in the table, the important random mechanical properties are responsible for explaining more than 99.8% of the uncertainty propagated into the uncertainty space. Therefore, the proposed dimensional reduction proves to be well controlled, since the reduced uncertainty space still contains the MPP with 99.8% of likelihood.

The importance of the random variables is affected mainly by 3 factors: the location of the critical point; the existence of implicit and/or explicit components of the total derivative of the limit-state function; the values of the design variables.

The integration point where the Tsai number takes its most critical value is that where the limit-state function is defined, in the reliability assessment, and where its derivatives and the Sobol' indices are calculated. The differentiation in order to m_i involves explicit and implicit components, given by $\partial g_2/\partial m_i$ and $\left(-\boldsymbol{\lambda}^{\boldsymbol{T}}\frac{\partial \mathbf{K}}{\partial m_i}\mathbf{u}\right)$, respectively (see (7.10)). The explicit component represents a direct measure of the sensitivity of g_2 in order to one of the random variables. In this problem, there exists only explicit component for the derivatives with respect to the mechanical properties associated with the laminate of the critical point. On the other hand, the implicit component is an indirect measure of sensitivity, quantifying the effects that variations on the random mechanical properties have upon the displacements field. Here, these may be related either with material properties of the laminate of the critical point or of the other laminates and are propagated through the vector of structural displacements, $\mathbf{u}$, and the vector of adjoint displacements, $\boldsymbol{\lambda}$. Although not evident in (7.10), only the derivatives in order to the elastic constants E_{1i} and E_{2i}, for $i = 1,2,34$, have implicit component, because the stiffness matrix $\mathbf{K}$ does not depend on the strength properties Y_i and S_i. Thus, the total derivatives $(\mathrm{d}g/\mathrm{d}Y_i)_j$ and $(\mathrm{d}g/\mathrm{d}S_i)_j$, calculated in the laminate j, are null if $i \neq j$. Therefore, the respective Sobol' indices are null as well, as shown by the results in Table 7.9. Only two different critical points were found. For reference solutions 1 to 4, in Table 7.9 (above the red line), the critical point is located on the upper layer of the sixth element, corresponding to the second laminate of the structure. For reference solutions 5 to 20, it is located on the upper layer of fifth element, corresponding to the first laminate. The location of the critical point appears to be the most influent factor in the magnitude of the Sobol' indices. As seen in Table 7.9, the indices referring to the laminate of the critical point have higher values.

Table 7.9: Sobol' indices with respect to each random variable (important variables are highlighted), of the tested reference solutions. Plus, percentage of uncertainty explained by the important variables and location of the critical point of the structural system.

	$S_{E_{11}}$	$S_{E_{21}}$	S_{Y_1}	S_{S_1}	$S_{E_{12}}$	$S_{E_{22}}$	S_{Y_2}	S_{S_2}	$S_{E_{13}}$	$S_{E_{23}}$	S_{Y_3}	S_{S_3}	$S_{E_{14}}$	$S_{E_{24}}$	S_{Y_4}	S_{S_4}	$\sum_{i:S_{x_i}\geq\varepsilon} S_{x_i}$	(e,p,k)
1	0.045087	0.000075	0.000000	0.000000	0.002576	0.079531	0.869864	0.002465	0.000310	0.000023	0.000000	0.000000	0.000007	0.000062	0.000000	0.000000	99.952%	(6,5,2)
2	0.045120	0.000081	0.000000	0.000000	0.002498	0.079459	0.869876	0.002505	0.000370	0.000028	0.000000	0.000000	0.000004	0.000059	0.000000	0.000000	99.946%	(6,5,2)
3	0.044835	0.000119	0.000000	0.000000	0.002716	0.079639	0.869888	0.002326	0.000383	0.000027	0.000000	0.000000	0.000004	0.000063	0.000000	0.000000	99.940%	(6,5,2)
4	0.045127	0.000083	0.000000	0.000000	0.002459	0.079428	0.869886	0.002523	0.000401	0.000031	0.000000	0.000000	0.000003	0.000058	0.000000	0.000000	99.942%	(6,5,2)
5	0.011232	0.052879	0.901506	0.000649	0.016280	0.000000	0.000000	0.000000	0.009348	0.001781	0.000000	0.000000	0.000105	0.006222	0.000000	0.000000	99.925%	(5,5,2)
6	0.011936	0.054055	0.899907	0.000521	0.015555	0.000015	0.000000	0.000000	0.010294	0.001937	0.000000	0.000000	0.000116	0.005663	0.000000	0.000000	99.935%	(5,5,2)
7	0.013354	0.056028	0.896805	0.000340	0.014194	0.000100	0.000000	0.000000	0.012147	0.002202	0.000000	0.000000	0.000130	0.004701	0.000000	0.000000	99.943%	(5,5,2)
8	0.012648	0.055101	0.898340	0.000420	0.014857	0.000051	0.000000	0.000000	0.011225	0.002078	0.000000	0.000000	0.000125	0.005157	0.000000	0.000000	99.940%	(5,5,2)
9	0.014045	0.056847	0.895304	0.000277	0.013572	0.000156	0.000000	0.000000	0.013065	0.002311	0.000000	0.000000	0.000132	0.004291	0.000000	0.000000	99.944%	(5,5,2)
10	0.012190	0.032951	0.917755	0.001564	0.018329	0.000484	0.000000	0.000000	0.000949	0.000488	0.000000	0.000000	0.000125	0.015165	0.000000	0.000000	99.795%	(5,5,2)
11	0.015346	0.058201	0.892416	0.000187	0.012453	0.000277	0.000000	0.000000	0.014916	0.002481	0.000000	0.000000	0.000130	0.003593	0.000000	0.000000	99.941%	(5,5,2)
12	0.013134	0.034693	0.915930	0.001115	0.018715	0.000347	0.000000	0.000000	0.001553	0.000614	0.000000	0.000000	0.000113	0.013786	0.000000	0.000000	99.893%	(5,5,2)
13	0.014059	0.036481	0.913975	0.000813	0.018862	0.000219	0.000000	0.000000	0.002273	0.000743	0.000000	0.000000	0.000089	0.012484	0.000000	0.000000	99.813%	(5,5,2)
14	0.016496	0.059229	0.889710	0.000129	0.011487	0.000393	0.000000	0.000000	0.016808	0.002593	0.000000	0.000000	0.000121	0.003034	0.000000	0.000000	99.936%	(5,5,2)
15	0.017005	0.059638	0.888434	0.000108	0.011051	0.000445	0.000000	0.000000	0.017774	0.002631	0.000000	0.000000	0.000115	0.002798	0.000000	0.000000	99.933%	(5,5,2)
16	0.015954	0.039954	0.909865	0.000451	0.018549	0.000046	0.000000	0.000000	0.003949	0.001011	0.000000	0.000000	0.000038	0.010183	0.000000	0.000000	99.946%	(5,5,2)
17	0.017880	0.060268	0.886056	0.000078	0.010258	0.000536	0.000000	0.000000	0.019750	0.002673	0.000000	0.000000	0.000102	0.002399	0.000000	0.000000	99.928%	(5,5,2)
18	0.018247	0.060496	0.884959	0.000066	0.009893	0.000575	0.000000	0.000000	0.020758	0.002680	0.000000	0.000000	0.000095	0.002231	0.000000	0.000000	99.926%	(5,5,2)
19	0.017179	0.065700	0.874294	0.000142	0.014626	0.000466	0.000000	0.000000	0.020603	0.003886	0.000000	0.000000	0.000339	0.002764	0.000000	0.000000	99.905%	(5,5,2)
20	0.017568	0.066029	0.873407	0.000125	0.014191	0.000519	0.000000	0.000000	0.021337	0.003894	0.000000	0.000000	0.000315	0.002615	0.000000	0.000000	99.904%	(5,5,2)

Table 7.10: Estimates of $\mathbf{y_{MPP}}$ relative to the 20th reference solution, for $\boldsymbol{\varepsilon = 0}$ and $\boldsymbol{\varepsilon = 10^{-3}}$, at $\boldsymbol{t_{total}}$, and the respective relative differences, in percentage.

ε	E_{11} $[GPa]$	E_{21} $[GPa]$	Y_1 $[MPa]$	S_1 $[MPa]$	E_{12} $[GPa]$	E_{22} $[GPa]$	Y_2 $[MPa]$	S_2 $[MPa]$	E_{13} $[GPa]$	E_{23} $[GPa]$	Y_3 $[MPa]$	S_3 $[MPa]$	E_{14} $[GPa]$	E_{24} $[GPa]$	Y_4 $[MPa]$	S_4 $[MPa]$
0	186.830	10.9722	27.5449	68.5938	178.391	10.5155	39.4212	67.8068	174.709	10.3546	40.0433	67.6513	180.817	10.1870	39.6159	68.6470
10^{-3}	185.251	10.8797	27.4618	68.0000	176.988	10.3000	40.0000	68.0000	175.627	10.4146	40.0000	68.0000	181.000	10.1624	40.0000	68.0000
Relative difference %	0.8524	0.8502	0.3026	0.8732	0.7927	2.0922	1.4470	0.2841	0.5227	0.5761	0.1083	0.5128	0.1011	0.2421	0.9602	0.9515

In the case of reference solutions 1 to 4, whose critical points lie on the 2nd laminate, it is seen that all the corresponding random mechanical properties are important. The strength property Y_2 is the most important random variable, explaining 87% of the uncertainty, followed by E_{22} with 8%. Although these mechanical properties have lower values of variance, the high values of their Sobol' indices, S_{Y_2} and $S_{E_{22}}$, are justified with higher values of the total derivatives. On the other hand, the elastic property E_{11}, of the 1st laminate, is more important than E_{12}, of the 2nd laminate. Having the same variance, this result is interpreted as a consequence of a strong implicit component of the total derivative $\mathrm{d}g_2/\mathrm{d}E_{11}$, which superimposes both the explicit and implicit components of $\mathrm{d}g_2/\mathrm{d}E_{12}$. The explanation lies on the fact that h_1 is one order of magnitude inferior to h_2, meaning that the second moment of area in matrix **K** is amplified.

In the case of reference solutions 5 to 20, whose critical points lie on the 1st laminate (below the red line in Table 7.9), there is a similar (but not equal) explanation of the uncertainty. It is seen that Y_1 is the most important random variable, explaining between 87% and 90% of the uncertainty, followed by E_{21}, explaining between 4% and 6% of the uncertainty. Both variables have lower values of variance, meaning their importance results from high valued total derivatives. Regarding the longitudinal Young moduli, there is an equilibrium between the importance levels of E_{11}, E_{12} and E_{13}, again justified by the lower thickness values of the 2nd and 3rd laminates, amplifying the implicit component of the total derivatives. In this case, the importance of E_{13} results from both the inner laminates being thinner and, so, subject to larger deflections. Thus, both the thickness and the displacements amplify the importance of this variable. Also, the transversal elastic moduli, E_{24} and E_{23}, have significant levels of importance resulting from strong implicit components, amplified by the displacements and lower thickness of the laminates.

The application of the proposed dimensional reduction has a very positive impact on the computational efficiency of the reliability assessment. Figure 7.29 and Figure 7.30 show the comparison of computing times, calculated for $\varepsilon = 0$ and $\varepsilon = 10^{-3}$. Specific evolutionary instants (t_1, t_{total} and $t_{\leq}$) are evaluated and compared in terms of generations and minutes.

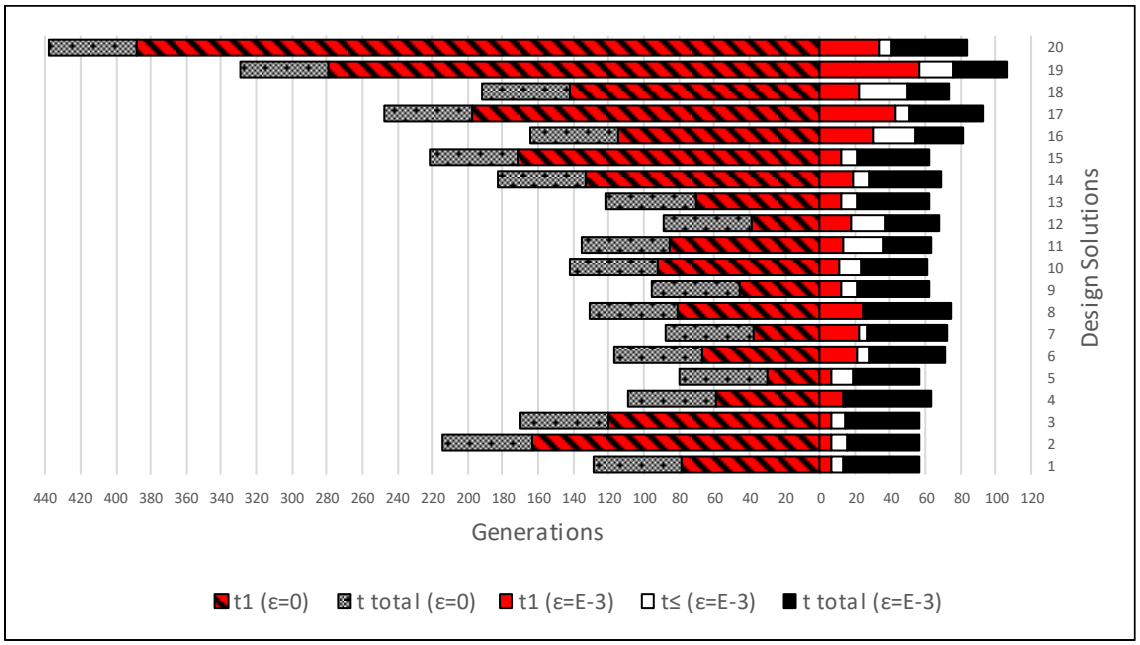

Figure 7.29: Comparison of the computing times, in generations, of reliability assessment, for $\boldsymbol{\varepsilon = 0}$ and $\boldsymbol{\varepsilon = 10^{-3}}$.

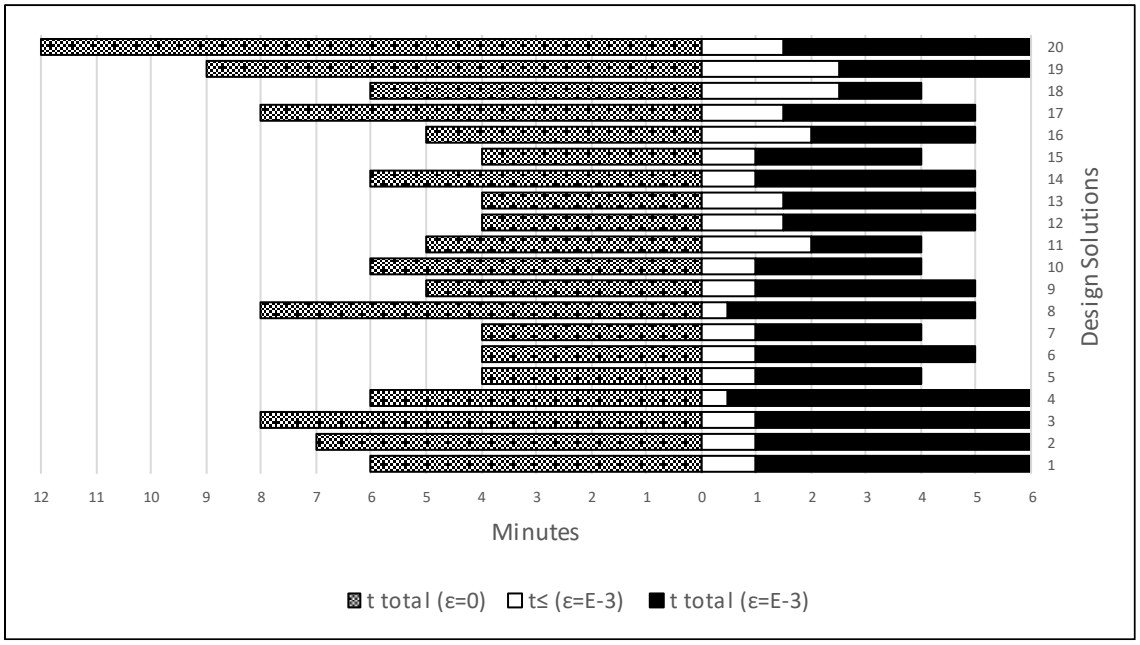

Figure 7.30: Comparison of the computing times, in minutes, of reliability assessment, for $\boldsymbol{\varepsilon = 0}$ and $\boldsymbol{\varepsilon = 10^{-3}}$.

Regarding the performance of the HmGA, to achieve convergence it is imposed that the population evolves during a predetermined number of generations, $t_{total} = t_1 + \Delta t$. Since Δt is constant, t_1 becomes the differentiating factor between a large or a short value of t_{total}. As it is shown in Figure 7.29, for $\varepsilon = 0$, large values of t_1 were required. A comparison with the respective total times, in Figure 7.30, shows that the long first stage of the evolution, in generations, was responsible for the large total computing times, for $\varepsilon = 0$.

For $\varepsilon = 10^{-3}$, it is seen that the proposed dimension reduction of the uncertainty space causes a significant decrease and homogenization of t_{total}, both in number of generations and in minutes. The maximum decrease happens for the 20th design, from 440 generations ($\varepsilon = 0$) to 84 generations ($\varepsilon = 10^{-3}$). Or, from 12 minutes to 6 minutes. As it is seen in Figure 7.29, the values of t_1 are significantly lower, meaning the improved HmGA was efficient in finding the first reduced search region $\mathcal{Z}^{t_1}$, i.e., it succeeded in finding and keeping solutions of high probabilistic failure content.

However, it also means that the number of genetic repairs performed by the IGRO, during this stage of the evolution, is larger. In practice, it implies that more model evaluations are performed and the number of the ratio minutes/generation increases. The computational cost becomes more influenced by the number of genetic repairs, than by t_1. Thus, there is not a visible parallelism between the distributions of t_{total}, in generations and in minutes, for $\varepsilon = 10^{-3}$. It is seen that, although in general t_{total} in minutes is lower, it remained the same or increased slightly for some reference solutions.

The fact that, for $\varepsilon = 10^{-3}$, the HmGA performs faster and steadier until convergence is a consequence of a reduced uncertainty space, in the reliability assessment problem. As discussed, less random variables imply exponentially less allowable genetic combinations. If $\mathbf{y_{MPP}} \in \Omega_r$, than the evolutionary path until the MPP is expected to be shorter. Table 7.8 shows the estimates of the reliability index of each reference solution, at t_{total}, for $\varepsilon = 0$ and $\varepsilon = 10^{-3}$. The results are very satisfactory: the values of β_{HL} are slightly lower for the latter case, meaning the HmGA was able to find a point on the failure surface that is closer to the actual MPP. That is, the algorithm achieved *superior accuracy* upon convergence.

Regarding the estimate of the MPP, Table 7.10 shows the calculated values of the components of the vector $\mathbf{y_{MPP}}$, for $\varepsilon = 0$ and $\varepsilon = 10^{-3}$, at t_{total}. The results also match, with relative differences no larger than 2% (generally less than 1%). It is seen however that the largest relative differences occur for the mechanical variables that were not considered important.

Considering that the results obtained for $\varepsilon = 0$ were validated in Sect. 7.2.3.1, then, for $\varepsilon = 10^{-3}$, there exists a $t_{\lesssim}$ such that $(t_1 + t_{\lesssim}) \leq t_{total}$. That is, the number of generations and minutes needed to achieve similar degrees of accuracy to those obtained with $\varepsilon = 0$ is lesser than t_{total}. As shown in Figure 7.29 and Figure 7.30, if $t_{\lesssim}$ is considered in the global convergence criterion defined in (4.40), instead of Δt, the computational efficiency of the HmGA improves significantly. Table 7.11 summarizes the performance of the HmGA, for $\varepsilon = 10^{-3}$.

Table 7.11: Average performance of the HmGA at the evolutionary instant ($\boldsymbol{t_1 + t_{\lesssim}}$), in generations and in minutes, for $\boldsymbol{\varepsilon = 10^{-3}}$.

$\mathbf{mean}\, t_{\lesssim}$	$\mathbf{min}\, t_{\lesssim}$	$\mathbf{max}\, t_{\lesssim}$	$\mathbf{mean}\, t_1 + t_{\lesssim}$		$\mathbf{min}\, t_1 + t_{\lesssim}$		$\mathbf{max}\, t_1 + t_{\lesssim}$	
gen.	gen.	gen.	gen.	min.	gen.	min.	gen.	min.
11.25	1	27	30.9	1.3	13	0.5	76	2.5

7.5 Concluding notes

On this chapter, the RBRDO problem of composite laminate structures proposed in Chapter 3 is solved. In accordance with Chapters 4 and 5, the global robustness and reliability measures are replaced by local approximations, with the purpose of avoiding sampling methods, like Monte Carlo simulation and related methods. Such local approximations, however, are based on pointwise gradient information of the stochastic system response functionals that must be calculated recurrently, during the design optimization process.

The numerical estimate of the gradient vectors of the implicit displacement and stress response functionals of composite laminate structural systems are in general hard to calculate by the traditional finite differences method, mainly due to the high dimensionality of the problem that imposes an elevated number of model evaluations only for the sensitivities analysis of a single design solution. Instead, we opt to apply an alternative semi-analytical method, called the adjoint variable method, that only requires an adjoint system of equilibrium equations to be solved for each design solution.

Then, a local approximation of the RBRDO problem is defined properly and the propagation of uncertainty in robustness and reliability assessment is explained. One of the purposes of this chapter is to compare, albeit for a particular numerical problem, how different the stochastic design optimization process behaves if reliability assessment is executed by the performance measure approach (PMA) or the reliability index approach (RIA). For that purpose, we introduce a composite laminate shell, as a numerical example. The outer bi-objective design optimization cycle is solved by the MOGA-2D (see Chapter 5). The inner cycle of reliability assessment is solved by the mGA (applied to the PMA) or by the HmGA (applied to the RIA).

Overall, the proposed methodology to solve the RBRDO problem of composite laminate structures proved to be very efficient, given the complexity of the problem. Direct comparison allows to conclude at first that, for this particular physical model, the MOGA-2D together with the PMA is faster to conclude the same predetermined number of generations and returns smoother distributions of the structural integrity parameters, along the Pareto front (PF) of the problem.

It is seen that, between both RBRDO problems, the distribution of the structural integrity measures $\bar{R}$ and β_{HL} are significantly difference, while the respective PFs are very close. The PMA allows much smoother distribution of $\bar{R}$, with values confined to a much smaller interval. At the same time, the PMA demonstrates to be more conservative than the RIA, with higher values of $\bar{R}$. Similar results were found for β_{HL}. With the PMA, the obtained distribution of the reliability index of the Pareto optimal solutions is smoother, than that obtained with the RIA. The values of the reliability index vary between smaller intervals, as well. It is also observed that there is a parallelism between the distribution of $\bar{R}$ and β_{HL}, meaning that Pareto optimal solutions with similar values of structural weight and system variability can have significantly different levels of structural integrity. Regarding the mean-values of the random design variables, it is seen that their distribution along the PF is also smoother with the PMA, than with the RIA, although the values are of the same magnitude. Yet, it is on the distribution of the thickness variables that the biggest differences can be found.

We suggest, therefore, the existence of a loose degree-of-freedom in the proposed RBRDO problem, in the sense that there exists a many-to-one relationship between the feasible search space and the feasible objective space of the problem and that, due to the high sensitivity of the structural integrity measures to the mean-values of the random design variables, the integrity measures are free to take any value that does not violate the reliability constraint without any kind of positive differentiation between them. It seems clear that, given two optimal designs of equivalent structural weight and system variability, the one with higher structural integrity is better. A second level of optimization or a third objective should be considered, in order to optimize the structural integrity values as well. Or, at least, structural integrity should be constrained from bellow.

In this chapter, a study on the effects of different sources of uncertainty in the RBRDO of composite laminate structures is also presented. We defined several combinations of sources of uncertainty, to which we call of uncertainty sets. Here the various uncertainty sets are considered to only have explicit influence on the feasibility robustness, while in reliability assessment (PMA) the uncertainty is only propagated through the random mechanical properties. Further studies on the explicit influence of different sources of uncertainty on the reliability assessment and on the combined structural feasibility and reliability assessment of composite laminate structures need to be taken in the future.

A numerical example of an aircraft fuselage-like composite shell structure is introduced. The results obtained show to be specific to this numerical example, but reveal interesting phenomena that possibly affect the design optimization of all composite laminate structures. For the current numerical example, the random mechanical properties appear to have a very reduced influence on feasibility robustness, when compared to the other random variables. A possible explanation is that it may be possible that the reliability constraint is acting as an implicit constraint over the importance of the random mechanical properties on the system variability. In other words, by restraining structural integrity to a minimum level of probabilistic reliability, in terms of the mechanical properties, it may be that the variability associated with these random variables becomes almost null, since failure becomes associated with events on the tail of the associated probability distribution.

As expected, it is seen that the combined effect of all random variable produces the largest values of system variability or, in other words, produces the small values of robustness. However, it is the combined effect of the random mechanical properties, the random ply angle and the random thickness variables that produce the greatest increment in the system variability. Greater than the sum of the parts, which suggests the existence of synergetic effects, between the displacement and stress responses of the structure, when these variables are considered.

Regarding the distributions of $\bar{R}$ and β_{HL}, along the PF, it is observed once more the existence of a parallelism between the two measures. Overall, it is found that the reliability constraint increases the

values of the lower bounds of both $\bar{R}$ and β_{HL} along the PF. Furthermore, it is seen that in the RBRDO the lower bound of $\bar{R}$ increases with the cardinality of the uncertainty sets.

For all the tested uncertainty sets, the distributions of $\bar{R}$ and β_{HL} present a V-shape, clearly indicating the lightest Pareto optimal solutions have superior structural integrity than the intermediate ones and often than the heaviest ones. A closer analysis allows to observe that such phenomenon is accompanied by a decrease of the optimal ply angles along the PF, for all uncertainty sets. It is seen that the lightest Pareto optimal solutions have lower ply angles and that the closer to 0º the higher the structural strength of the designs. Hence, explaining the V-shape of the distributions.

The V-shape of the distribution of the Tsai number and the reliability index allows to conclude that it is possible to obtain lighter composite laminate structures, if the fibers are aligned with the optimal ply-angle. However, in the particular case of this numerical example, it shows as well that the design optimization process has a loose degree-of-freedom. Clearly, for all the tested uncertainty sets, the heaviest Pareto optimal solutions, although optimal, could have higher values of structural integrity if the respective optimal ply-angles were set at 0º.

Finally, we study the effects of the developed analytical dimensional reduction technique, based on the analytical solution of the Sobol' indices. The goal is to improve the convergence rate of the HmGA, by reducing the expected computing times in finding the global optimal solution of the RIA. The study is made by comparing the performance of the algorithm with and without the said technique. A subset of the Pareto optimal solutions obtained with the RBRDO together with the PMA are used as reference solutions of the study. As before, the uncertainty is propagated in reliability assessment through the random mechanical properties.

The analytical solution of the Sobol' indices is based on the first-order approximation of the stress response functional, hence, requiring gradient information. As before, differentiation is performed by the adjoint variable method. The costs of such operation are negligible, because the gradient vector of the stress response is already available, since the HmGA already makes use of to calculate an initial estimate of β_{HL}.

The goodness of the proposed methodology is twofold. First, the dimensional reduction of the uncertainty space only requires one adjoint set of equilibrium equations to be solve once, at the beginning of the reliability assessment procedure. Second, it allowed to significantly increase the efficiency of the HmGA. The results of the importance analysis show that about 99.8% of the uncertainty propagation is explained by a small number of random mechanical properties. From a total of 16 variables, only 5 to 8 were considered important. It was also found that, for the tested numerical example, the location of the critical point of the structure is the most important factor to determine the importance of the random mechanical properties. The values of the design variables also play an important role, affecting the implicit component of the total derivatives.

Regarding the ability to achieve global convergence, it is shown that the HmGA performs faster and steadier if only the important random variables rare considered, in reliability assessment. It is also capable of achieving a superior accuracy than without any dimensional reduction, long before the global convergence criterion is satisfied.

Therefore, it is expected that the application of evolutionary-based reliability assessment, as an inner cycle of structural design optimization, becomes significantly more affordable. The reliability assessment of complex structural problems with multiple failure modes, where gradient-based are expected to fail and where Monte Carlo Simulation becomes impractical, may have an efficient solution with EAs. Further work needs to be done in this respect.

References

[1]. Arora JS, Cardoso JB. Variational principle for shape design sensitivity analysis. AIAA Journal 1992; 30: 538-47.

[2]. Pontryagin LS, Boltyanskii VG, Gamkrelidze RV. The mathematical theory of optimal processes. Pergamon Press; 1964.

[3]. Lions JL. Optimal control of systems governed by partial differential equations. Springer-Verlag; 1971.

[4]. Conceição António CC. Optimização de estruturas em materiais compósitos de matriz polimérica. PhD thesis. Porto: Universidade do Porto; 1995.

[5]. Plessix, RE. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophys J Int 2006; 167: 495-503.

[6]. Bradley AM. PDR-constrained optimization and the adjoint method; 2013. http://cs.stanford.edu/~ambrad/adjoint_tutorial.pdf; Accessed 20 April 2017.

[7]. Ahmad S, Irons S, Zienkiewicz OC. Analysis of thick and thin shell structures by curved finite elements. Int J Numer Methods Eng 1970;2(3):419-51.

[8]. Tsai SW. Composite Design. Think Composites, USA 1987.

[9]. António C, Soeiro A, Marques AT. Influence of physical properties randomness in laminated composites strength. ICCM-9. In: Proceedings of the Ninth International Conference on Composite Materials; 1995.

[10]. Conceição António C, Torres Marques A, Gonçalves JF. Reliability based design with a degradation model of laminated composite structures. Struct Optim 1996;12:16-28.

[11]. Casella G, Berger LB. Statistical Inference. Thomson Learning; 2000.

[12]. Box GEP, Muller ME. A note on the generation of random normal deviates. The Annals of Mathematical Statistics 1958: 610-11.

[13]. Tu J, Choi KK, Park YH. Design Potential Method for Robust System Parameter Design. AIAA Journal 2001; 39(4): 667-677.

VIII
CONCLUSIONS

8.1 General considerations and conclusions

The main conclusions of the work presented in this thesis and potential research topics directly related with the developed work are presented in this chapter.

8.1.1 RBRDO with imposed probability density value

A new approach to the RBRDO of ply-angle composite laminate shell structures is proposed. Design optimization is considered as the bi-objective minimization problem of the structural weight (optimality) and the determinant of the variance-covariance matrix (robustness). The outer cycle of design optimization is solved by the *bi-level dominance multi-objective genetic algorithm* (MOGA-2D). Regarding reliability assessment, the *performance measure approach* (PMA) is rewritten as an unconstrained maximization problem, reducing the search space to the surface of the hypersphere of radius β^a, and solved by the *micro-genetic algorithm* (mGA). Furthermore, to skip unnecessary reliability assessment inner cycles, the reliability constraint is defined in terms of the piecewise equivalent PMA limit-state function.

Throughout the entire RBRDO process, the derivatives are analytically estimated by the Adjoint Variable Method, only requiring one adjoint system of equilibrium equations to be solved, for each design solution. The method proves to be very efficient and a powerful tool for designers to make decisions, establishing the priorities between optimality, robustness and reliability. A physical model of a balanced angle-ply laminate composite shell is presented.

The results show that the reliability constraint has a visible influence on the evolutionary process, generating designs with higher values of structural weight at the lowest portion of the Pareto front, in comparison with the corresponding RDO problem. This phenomenon happens because the reliability constraint prevents the excessive decrease of the reliability index, along to Pareto front, to reach levels too close of the failure domain. In fact, it is seen a parallelism between the distribution of the structural integrity measures, along the Pareto front, namely, the critical Tsai number, the reliability index and the optimal solution of the PMA.

Regarding the robustness of the Pareto optimal solutions, the coefficient of variation of the critical Tsai number has moderate but constant values, along the Pareto front. On the other hand, the coefficient of variation of the critical displacements shows an exponential growth along the Pareto front and appears to have a greater influence on the increase of the system variability.

In relation to the optimal value of the design variables, the optimal mean-values of the ply angles are constant and equal to 90º. The mean-values of the random thickness variables of the four laminates show different behaviors. It is seen that the design optimization process evolves towards increasing mainly the thickness of the fourth laminate, to increase the total structural weight.

The proposed procedure proves to be successful in increasing the efficiency of the RBRDO of composite structures. If the PMA was to be applied in its standard form, and its optimization cycle solved directly by an evolutionary algorithm (EA), without any adaptation, the computing times would have been exponentially higher. The piecewise equivalent PMA limit-state function also plays an important role acting as a surrogate-like reliability assessment model, for the safest design solutions.

8.1.2 RBRDO with imposed limit-state value

Contrary to the PMA, the *reliability index approach* (RIA) returns a quantitative measure of structural reliability. For that reason, the *hybrid micro-genetic algorithm* (HmGA) is first validated. The proposed method is expected to set the basis for further developments on the design optimization of more complex systems, with multiple failure criteria, where gradient-methods are expected to fail. Random variables are decomposed into magnitude and direction components, the RIA redefined as a penalty problem, with a mixed real-binary genotype, and two new evolutionary operators are introduced: one for the genetic repair of the solutions (to impose the equality constraint of the reliability assessment problem) and another for the progressive reduction and reallocation of the search domain (implicitly guiding the evolution to a region of the failure surface with higher probability content)

To study the ability of the proposed method, a large set of Pareto-optimal solutions is evaluated, resulting from the RBRDO problem discussed in the previous section, on which reliability assessment was solved by the PMA. The reliability index, β_{HL}, of each design solution is estimated by the proposed evolutionary method and by a gradient-based method. Then, confidence intervals for the probability of failure are estimated, with *Monte Carlo simulation* (MCS).

The results show very good accuracy of the proposed method in predicting the value of β_{HL} and the location of the *most probable failure point* (MPP) on the failure surface. Computing times of the HmGA are considered to be practical and it is concluded that the most consuming part of the search process happens during the first stage of evolution, because of the similarity control operator. Yet, it demonstrates the efficiency and the capability of the new evolutionary operators to solve the RIA problem.

Finally, confidence intervals of the probability of failure are calculated by the crude MCS method, for three representative optimal solutions. The probabilities of failure estimated by means of β_{HL} fall within the confidence intervals. MCS requires a huge number of samples to achieve convergence, taking about 35 hours to complete the reliability assessment cycle. By comparison, it shows the efficiency of the proposed evolutionary methodology, which is able to solve the problem in few minutes with similar levels of accuracy. Even though the HmGA performs slower than gradient-based methods, the advantages of EAs must be considered, mainly the ability to achieve global convergence.

Considering the RBRDO of composite laminate structures, design optimization is considered as the bi-objective minimization problem of the weight (optimality) and the determinant of the variance-covariance matrix of the response functionals of the system (robustness). Reliability assessment is executed by the RIA, as an inner cycle of design optimization. The key concept of this methodology is the exclusive use of EAs with elitist strategy, allowing for the global convergence of each of the two cycles of the optimization process. Furthermore, to skip unnecessary reliability assessment inner cycles, the reliability constraint is defined in terms of the piecewise equivalent RIA limit-state function.

Throughout the entire RBRDO process, the derivatives are analytically estimated by the Adjoint Variable Method, only requiring one adjoint system of equilibrium equations to be solved, for each

design solution. The method proved to be very efficient and a powerful tool for designers to make decisions, establishing the priorities between optimality, robustness and reliability. A physical model of a balanced angle-ply laminate composite shell is presented.

The results show the efficiency of the method. The Pareto front converges in practical computing times. The Pareto fronts obtained with and without the reliability assessment (RBRDO and RDO, respectively) are close to each other, meaning the optimal region of the objective space is of similar shape, in both problems. It is observed that the heaviest Pareto optimal solutions are slightly lighter in the RBRDO, while the lightest Pareto optimal solutions tend to be heavier, in the RBRDO.

The distribution of the reliability index, along the Pareto front, shows the reliability constraint is satisfied for every Pareto optimal solution, meaning that the reliability assessment inner cycle is applied successfully. However, the distribution is very irregular, contrary to that obtained with the PMA. Again, it is seen a parallelism between the distribution of the structural integrity measures, along the Pareto front, namely, the critical Tsai number and the reliability index.

In relation to the robustness of the Pareto optimal solutions, the coefficient of variation of the critical Tsai number has moderate but constant values, along the Pareto front. On the other hand, the coefficient of variation of the critical displacements shows an exponential growth along the Pareto front and appears to have a greater influence on the increase of the system variability.

Regarding the Pareto optimal solutions, it is observed that the optimal orientation of the fibers is at 90°, for most Pareto optimal solutions, with few exceptions at 84º. The distribution of the thickness variables along the Pareto front shows that the optimization process evolves towards increasing mainly the thickness of the fourth laminate, to increase the total structural weight.

8.1.3 PMA vs. RIA in robust design optimization

Although the Pareto fronts obtained with both RBRDO models, with the PMA and the RIA, are very close to each other we found significant differences that must be reported and discussed.

In terms of computing times, it is concluded that the same RBRDO problem is solved 1.5 times faster with the PMA, than with the RIA. It is interesting to observe that, between both RBRDO problems, the distributions of the critical Tsai number and the reliability index are significantly difference, while the respective Pareto fronts are very close. Overall, it is seen that the PMA allows for much smoother distributions of the structural integrity measures, with values confined to much smaller intervals than the RIA. However, at the same time, the PMA shows to be a more conservative approach, since the RIA allows Pareto optimal solutions to have reliability values much closer to the allowable value β^a.

The similarity between the Pareto fronts, obtained with the PMA and the RIA, compared with the significant differences in the respective distributions of the critical Tsai number and of the reliability index, reveals the sensitivity of the measures of structural integrity to the random design variables, in particular to the random thickness variables, which is corroborated with the calculated values of the stress coefficient of variation. This is justified by the fact that the critical Tsai number is an inverse function of the cubic power of the laminate thickness. Regarding the configuration of the uncertainty space, it means that the failure surface and the origin of the uncertainty space are significantly altered with small variations in the thickness values, resulting in different reliability levels.

This behavior reveals the possibility of the existence of a loose design degree-of-freedom, in the formulated RBRDO problem. It refers to a possible many-to-one relationship between the feasible search space and the feasible objective space. That is, Pareto optimal solutions of equivalent weight and system variability can have significantly different levels of structural integrity, due to different realizations of the design variables. Thus, in the true sense of *optimization*, for a given optimal

realization in the objective space, the distribution of the thicknesses among the four laminates, should perhaps be optimized, as well, for a fixed value of structural weight.

8.1.4 Effects of different sources of uncertainty in RBRDO

A study on the effects of different sources of uncertainty in the RBRDO of composite laminate structures is conducted. Several combinations of sources of uncertainty are defined, to which we call uncertainty sets. The various uncertainty sets are considered to only have explicitly influence on feasibility robustness, while in reliability assessment (PMA) the uncertainty is only propagated through the random mechanical properties. A numerical example of an aircraft fuselage-like composite shell structure is presented.

For the current numerical example, the random mechanical properties appear to have a very reduced influence on feasibility robustness, when compared to the other random variables. A possible explanation is that it may be possible that the reliability constraint is acting as an implicit constraint over the importance of the random mechanical properties on the system variability. In other words, by restraining structural integrity to a minimum level of probabilistic reliability, in terms of the mechanical properties, it may be that the variability associated with these random variables becomes almost null, since failure becomes associated with events on the tail of the associated probability distribution.

As expected, it is seen that the combined effect of all random variable produces the largest values of system variability or, in other words, produces the smallest values of robustness. However, it is the combined effect of the random mechanical properties, the random ply angle and the random thickness variables that produces the greatest increment in the system variability. Greater than the sum of the parts, which suggests the existence of synergetic effects, between the displacement and stress responses of the structure, when these variables are considered.

Regarding the distributions of the critical Tsai number and the reliability index, it is observed once more the existence of a parallelism between the two measures. Overall, it is found that the reliability constraint increases the values of the lower bounds of both measures. For all the tested uncertainty sets, it is found that the distributions of the structural integrity measures display a V-like shape, clearly indicating the lightest Pareto optimal solutions have superior structural integrity than the intermediate ones and often than the heaviest ones. A closer analysis allows to observe that such phenomenon is accompanied by a decrease of the optimal ply angles along the PF, for all uncertainty sets. It is seen that the lightest Pareto optimal solutions have lower ply angles and that the closer to 0^{o} the higher the structural strength of the designs.

The V-shape of the distribution of the Tsai number and the reliability index allows to conclude that it is possible to obtain lighter composite laminate structures, if the fibers are aligned with the optimal ply-angle. However, in the particular case of this numerical example, it shows as well that the design optimization process has a loose degree-of-freedom. Clearly, for all the tested uncertainty sets, the heaviest Pareto optimal solutions, although optimal, could have higher values of structural integrity if the respective optimal ply-angles were set at 0^{o}.

8.1.5 Application of Sobol' indices to reliability assessment

The importance analysis theory of Sobol' is implemented to reduce the dimensionality of the uncertainty space, improving the efficiency towards global convergence of evolutionary-based reliability assessment. Sobol' indices are formulated analytically following the theory of propagation of moments. The HmGA is instrumented with the Sobol' indices. A threshold parameter is introduced

to identify the important random variables. A set of Pareto-optimal solutions of a multi-laminate composite structure is evaluated.

It is demonstrated that, in a probability space of N independent random variables, any k-th order Taylor polynomial is coincident with its own ANOVA decomposition if and only if it is multi-linear, for $k \leq N$. This result allows to consider multi-linear Taylor approximations of the limit-state function to perform local importance analysis without the need to calculate the integral terms of the ANOVA decomposition. It also allows to use only low order gradient information, which is useful if the structural response functionals are implicit defined. Of particular interest is the first-order Taylor polynomial. The first-order total derivatives are analytically calculated by the Adjoint Variable Method.

The goodness of the proposed methodology is twofold. First, the dimensional reduction of the uncertainty space only requires one adjoint set of equilibrium equations to be solved once, at the beginning of the reliability assessment procedure. Second, it allows to significantly increase the efficiency of the HmGA. The results of the Importance Analysis show that about 99.8% of the uncertainty propagation is explained by a small number of random variables. From a total of 16 random variables, only 5 to 8 random variables are considered as important. It is also found that, for the tested numerical example, the location of the critical point of the structure is the most important factor to determine the importance of the random variables. The values of the design variables also play an important role, affecting the implicit component of the total derivatives.

Regarding the ability to achieve global convergence, it is shown that the HmGA performs faster and steadier if only the important random variables are considered in the reliability assessment. It is also capable of achieving superior accuracy, long before the global convergence criterion is satisfied.

As a concluding remark, with the application of the proposed dimensional reduction to evolutionary-based reliability assessment, structural design optimization is expected to become more affordable. Furthermore, the reliability assessment of complex structural problems with multiple failure modes, where gradient-based are expected to fail and where MCS becomes impractical, may have an efficient solution with EAs. Further work needs to be done in this respect.

8.2 Perspectives of future research

Future work related with the design optimization of composite structures shall be directed to the following research topics:

1) Formulation of new structural design optimization models;
2) Inclusion of geometric nonlinearity and fracture analysis, in the RBRDO of composite laminate structures;
3) Application of stochastic structural design optimization to real-scenario structural systems;
4) Study the functional relation between the Tsai number and the Hasofer-Lind reliability index;
5) Application and formulation of new reliability and robustness measures.

Topic 1) is vast and constitutes a mandatory step to be taken on every design optimization problem. It means that the designer must evolve the existent models, or seek new ones, to provide the best answers to the constantly new demands and challenges posed by the industry and the commercial sectors.

Regarding the results obtained in this thesis, it was reported the existence of an additional design degree-of-freedom, in the global RBRDO problem introduced in Chapter 3. Although the numerical models introduced are simplified representations of real-scenario structures, they are sufficiently complex in terms of the number of components and of the materials to reveal the possibility of a

many-to-one relation between the feasible search space of the problem and the respective objective space.

The possibility that points of the same neighborhood in the objective space can have significantly different levels of structural integrity indicates that the nondominated points (the ones on the Pareto front) should be further optimized in terms of the structural integrity measures. That is, to an optimal realization in the objective space (in terms of the objective functions) should correspond the optimal realization in the search space (in terms of structural integrity).Such problem is called an *inverse optimization problem* and is (most likely) extremely hard to be solved directly. The direct approach to it would be the implementation of a second level of optimization, after convergence of the Pareto front, where the design variables would be the thickness variables only, subject to a weight-related constraint. In this case, the stricter the constraints the harder the problem. An indirect approach would be to consider a third objective, maximizing the critical Tsai number, for example. The complexity of such problem would be exponentially greater than the one formulated in Chapter 3, since instead of a two-dimensional curve, the Pareto front would be a three-dimensional surface. In both cases, new optimization algorithms would have to be developed.

Topic 2) refers to the necessity of accounting for a more realistic description of reality in the structural models used in the stochastic design optimization of structural systems and, particularly, of composite laminate structures. In truth, the application of the RBRDO to more complex structures is limited by the ability to produce accurate simulations of the underlying physical phenomena. This leads us to Topic 3). The more complex the structural simulation the less efficient, and probably the less accurate, the same optimization model will be. For the past 20 years, the use of surrogate models has been studied and very interesting developments have been published in the literature. However, one should notice that the application of surrogate models in the context of structural design optimization must be carefully thought and adapted to each problem.

Topic 4) refers to the observable parallelism between the distributions of the critical Tsai number and the reliability index. We question if this phenomenon of parallelism is happening only under the circumstances of the current RBRDO problem, or not.

Topic 5) relates with Topic 1), in the sense, that from the alternative formulations of robustness and reliability result new optimization models necessarily. The inclusion of time in structural robustness and reliability assessment may be very important, in the sense that it would allow to link the definition of robustness and reliability with the concept of fatigue resistance, which is an important topic in composite materials. It would allow as well to correct the probability distributions of the random variables that may change with time, reflecting changes in the operation conditions of the structure, thus allowing for more accurate predictions of robustness and reliability over the expected life-time of the structure.

APPENDIX A

A.1 Uncertainty set {m}

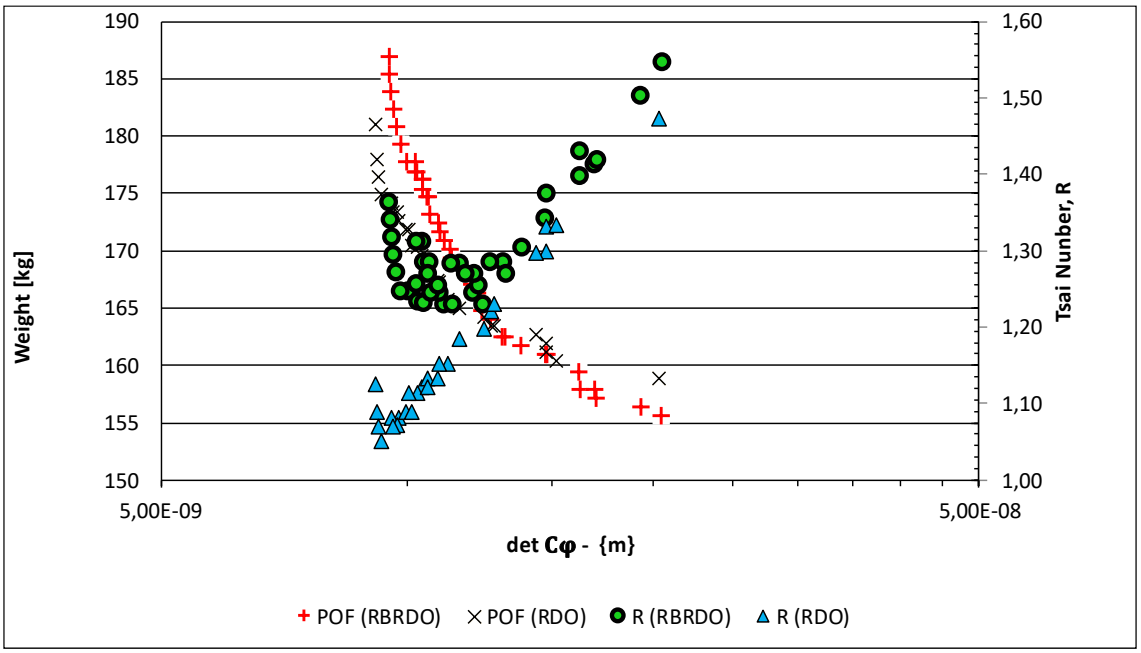

Figure A.1: Distribution of the critical Tsai number, along the Pareto front, obtained with the uncertainty set {**m**}.

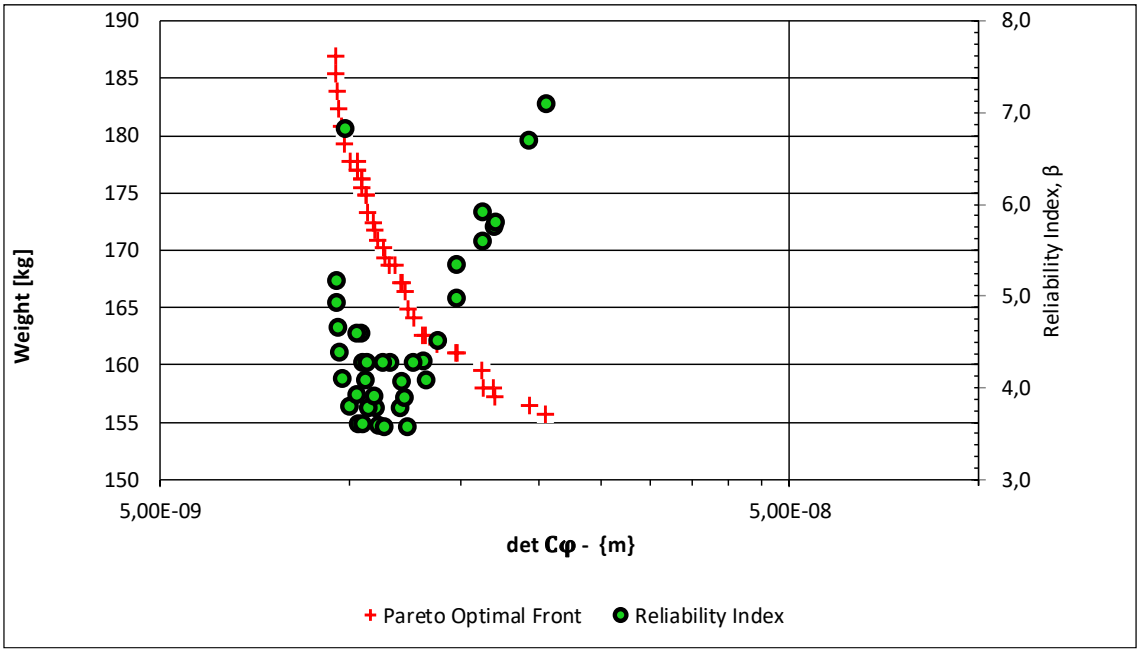

Figure A.2: Distribution of the reliability index, along the Pareto front, obtained with the uncertainty set {**m**}.

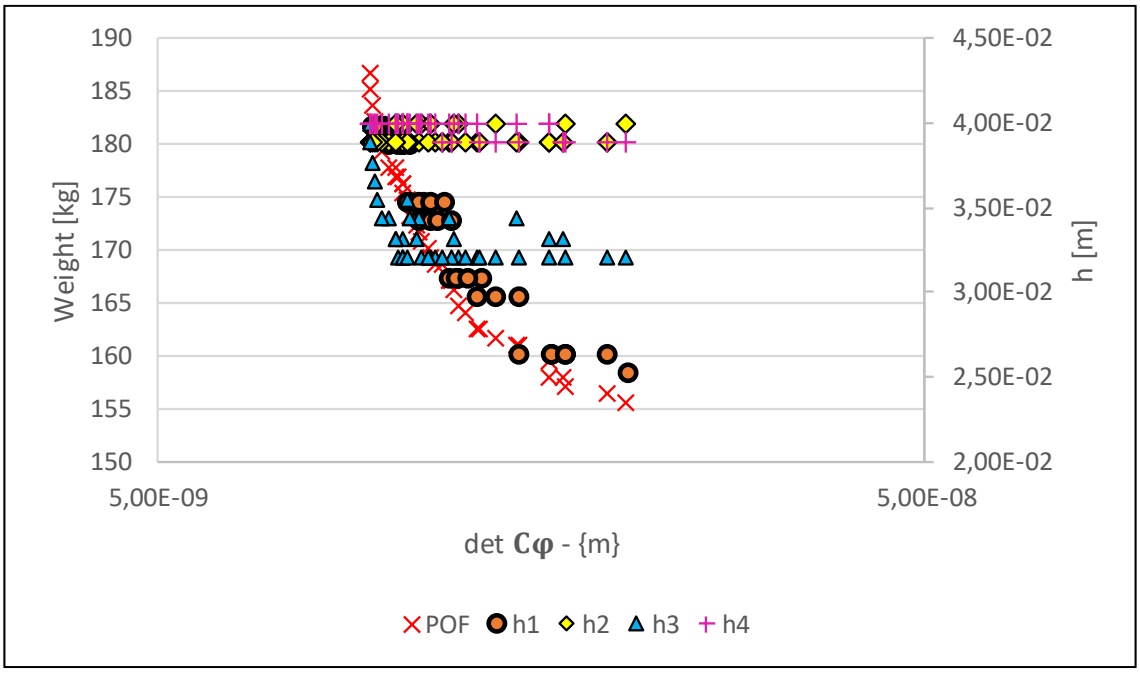

Figure A.3: Distribution of the thickness variables, along the Pareto front, obtained with the uncertainty set {**m**}.

A.2 Uncertainty set {m, h}

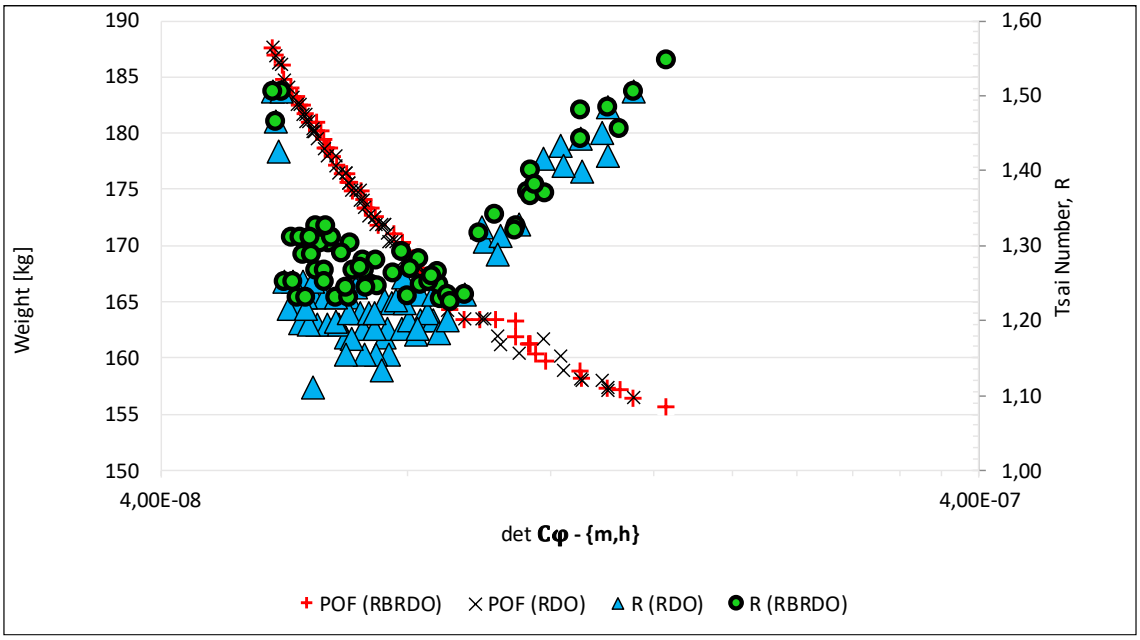

Figure A.4: Distribution of the critical Tsai number, along the Pareto front, obtained with the uncertainty set {**m**, **h**}.

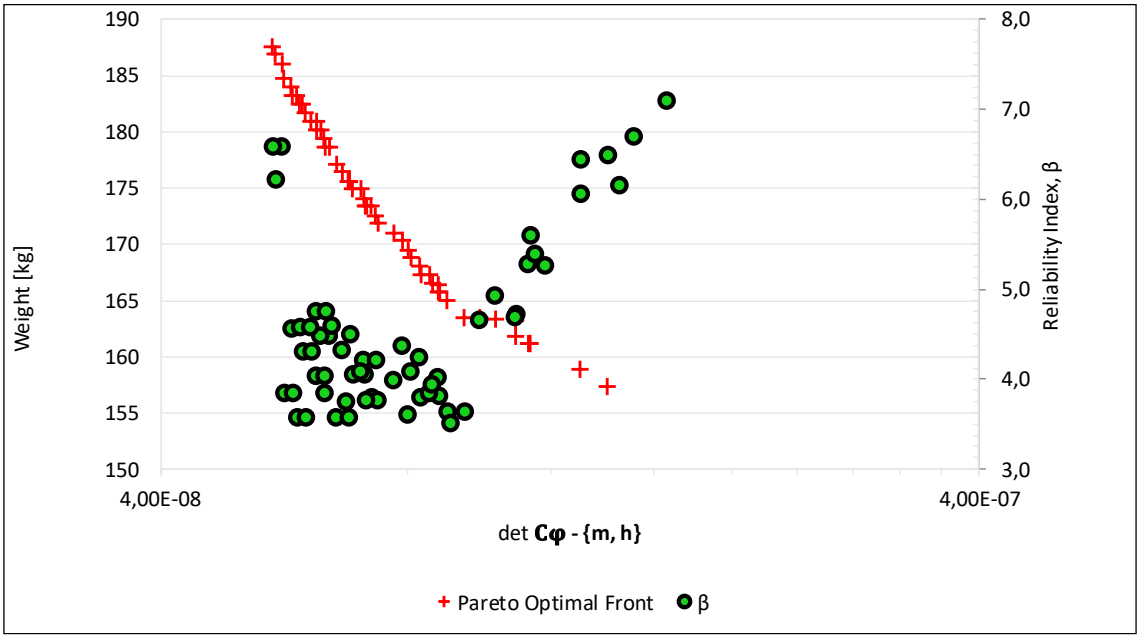

Figure A.5: Distribution of the reliability index, along the Pareto front, obtained with the uncertainty set {**m**, **h**}.

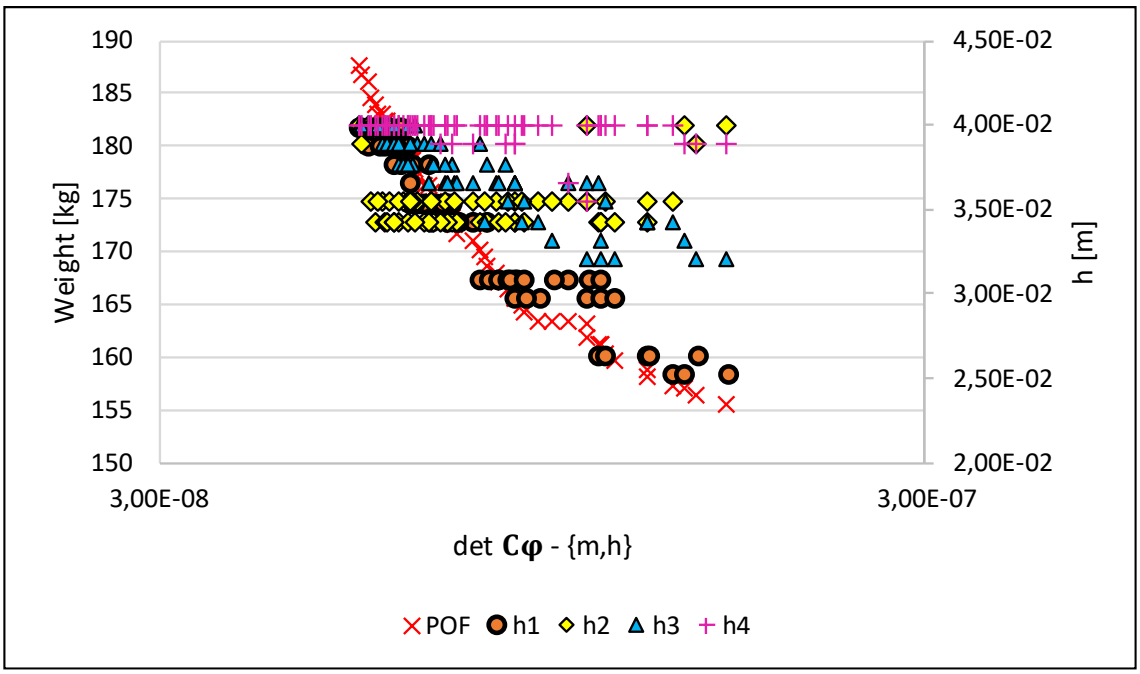

Figure A.6: Distribution of the thickness variables, along the Pareto front, obtained with the uncertainty set {**m**, **h**}.

A.3 Uncertainty set {**m**, $\boldsymbol{\theta}$}

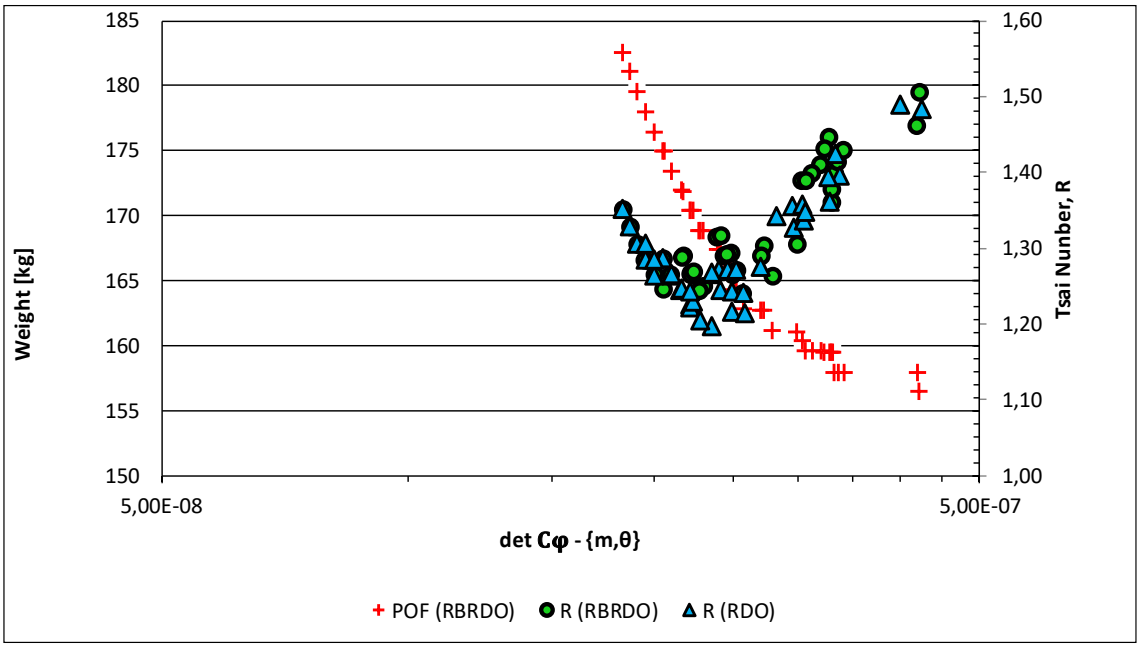

Figure A.7: Distribution of the critical Tsai number, along the Pareto front, obtained with the uncertainty set {**m**, **θ**}.

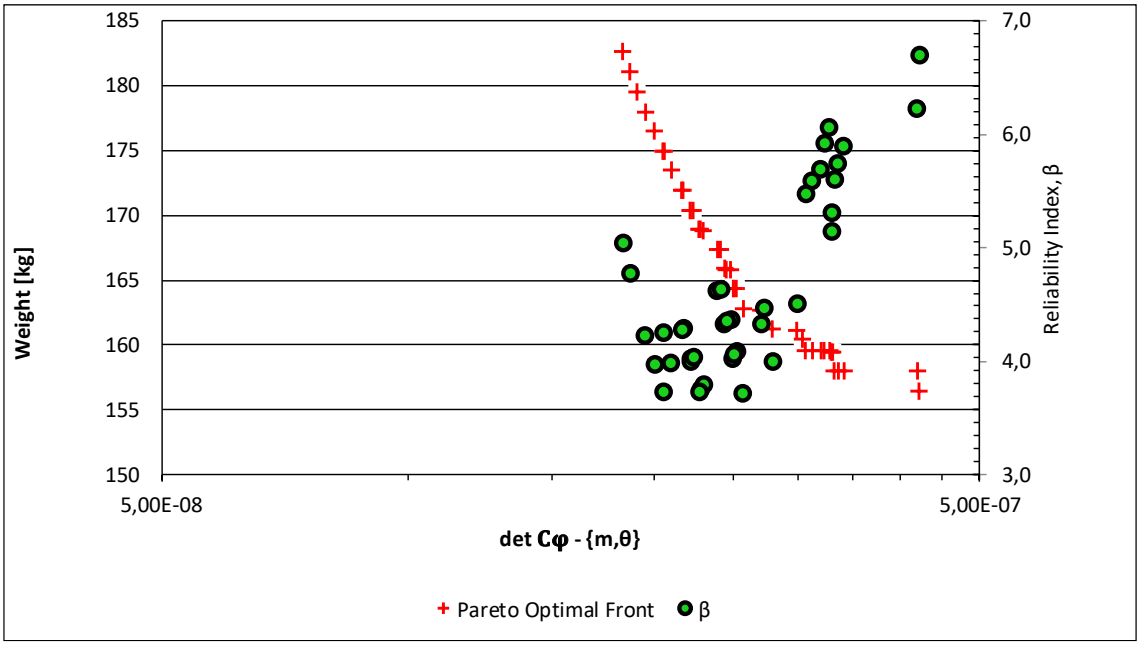

Figure A.8: Distribution of the reliability index, along the Pareto front, obtained with the uncertainty set {**m**, **θ**}.

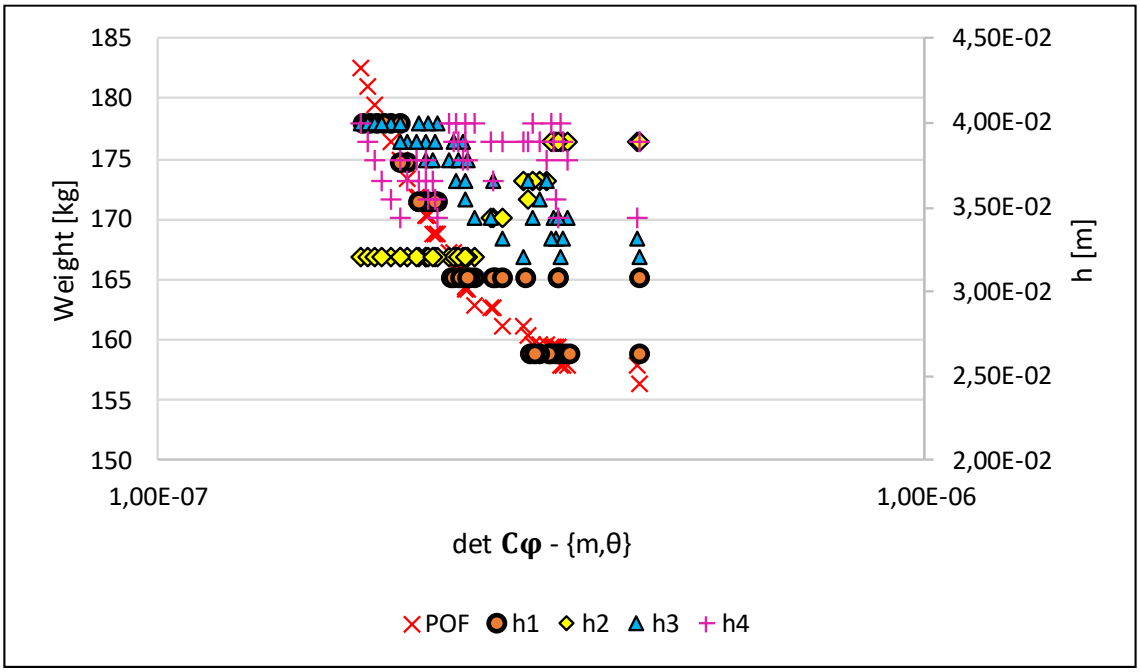

Figure A.9: Distribution of the thickness variables, along the Pareto front, obtained with the uncertainty set {**m**, **θ**}.

A.4 Uncertainty set {m, h, θ}

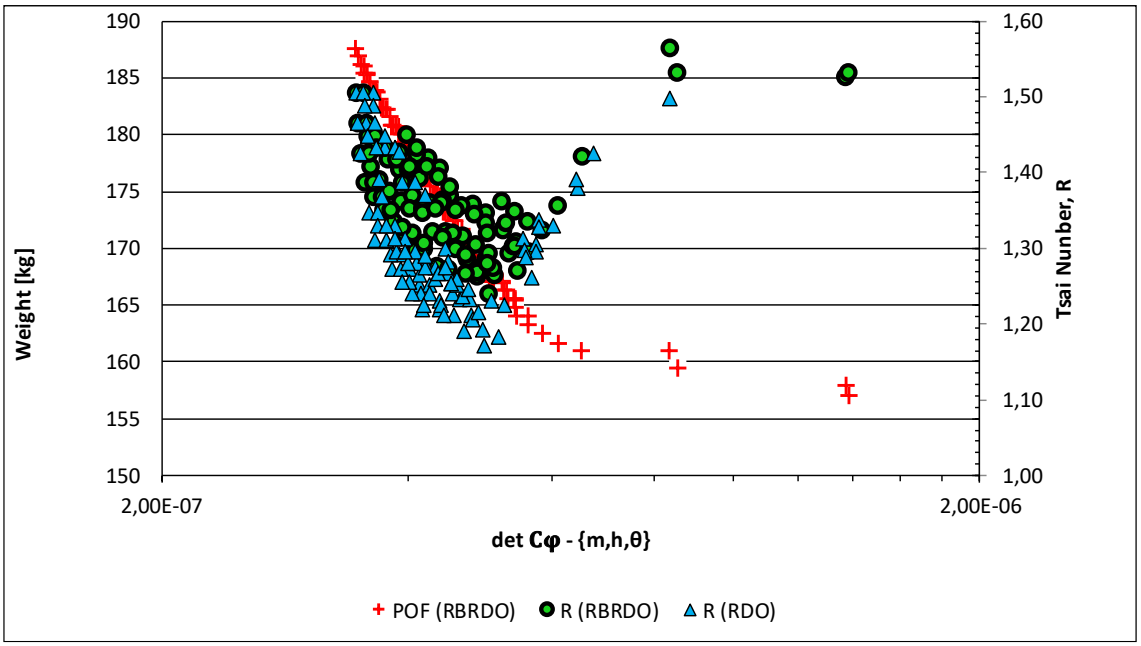

Figure A.10: Distribution of the critical Tsai number, along the Pareto front, obtained with the uncertainty set {**m**, , **h**, **θ**}.

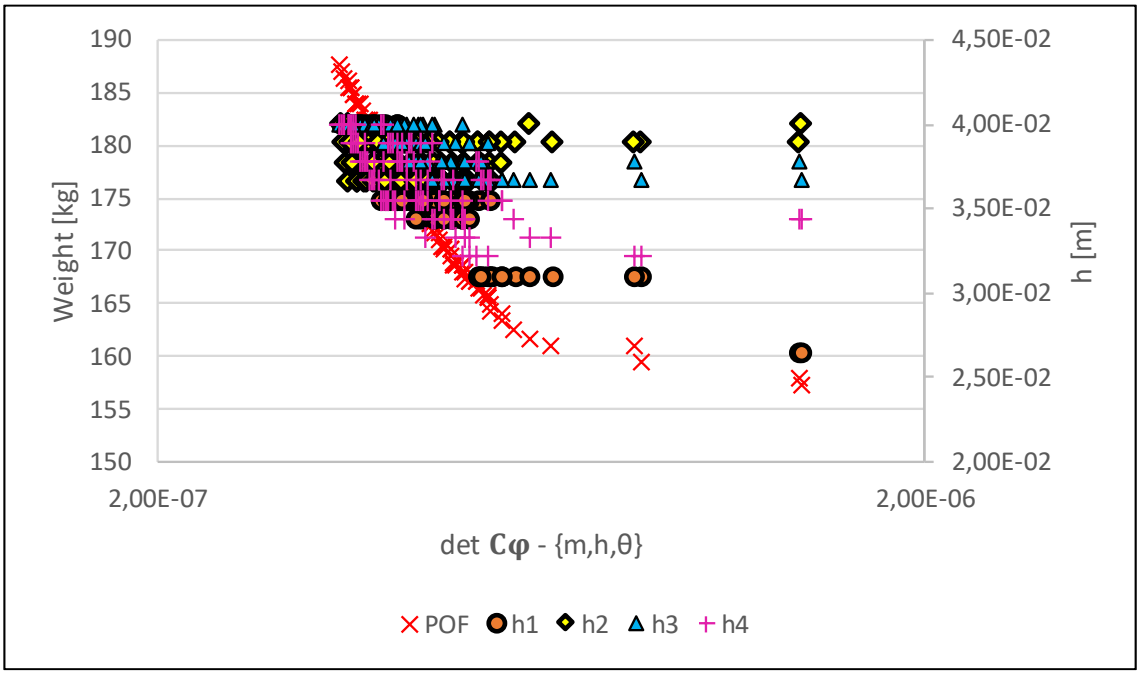

Figure A.11: Distribution of the thickness variables, along the Pareto front, obtained with the uncertainty set {**m**, **h**, **θ**}.

A.5 Uncertainty set {m, h, θ, f}

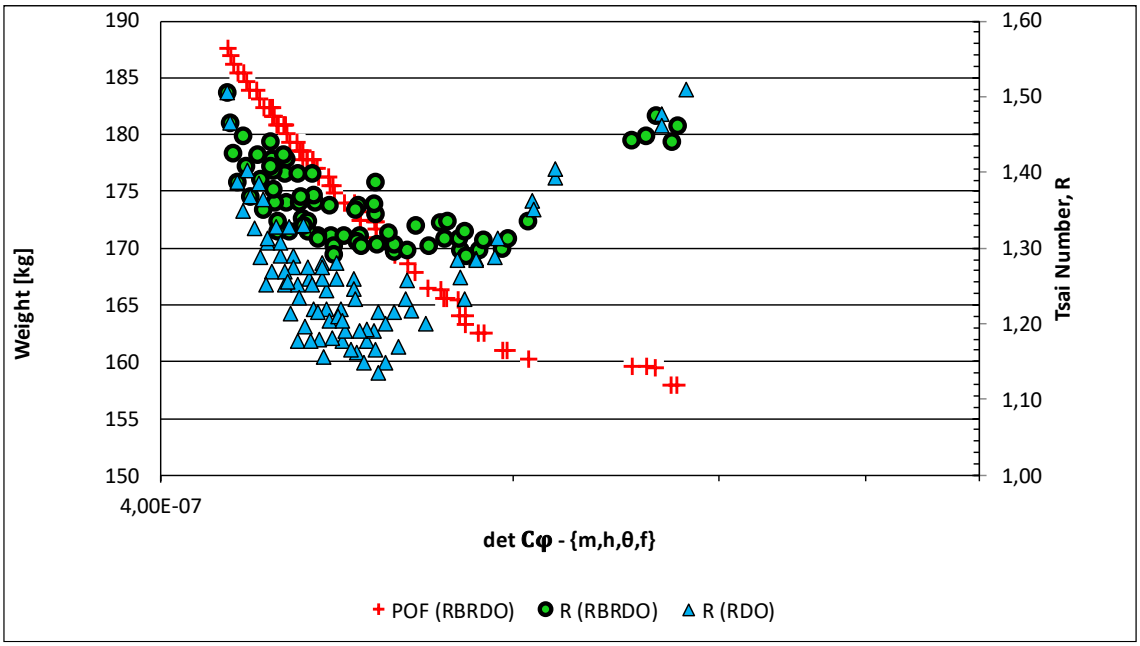

Figure A.12: Distribution of the critical Tsai number, along the Pareto front, obtained with the uncertainty set {**m**, , **h**, **θ**, **f**}.

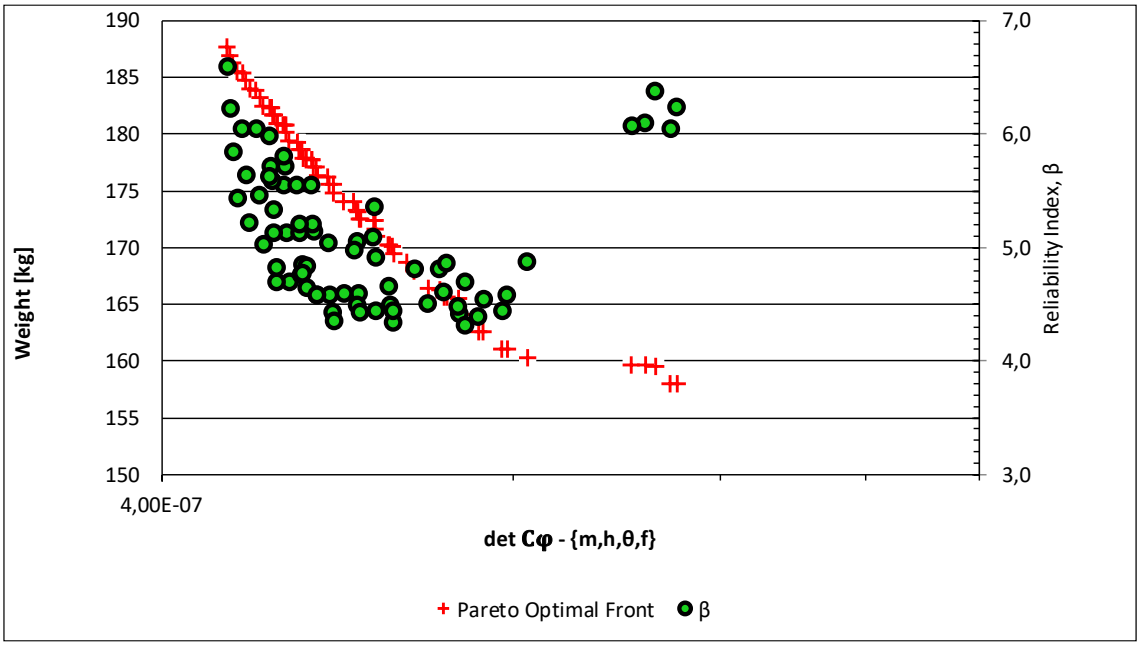

Figure A.13: Distribution of the reliability index, along the Pareto front, obtained with the uncertainty set {**m**, **h**, **θ**, **f**}.

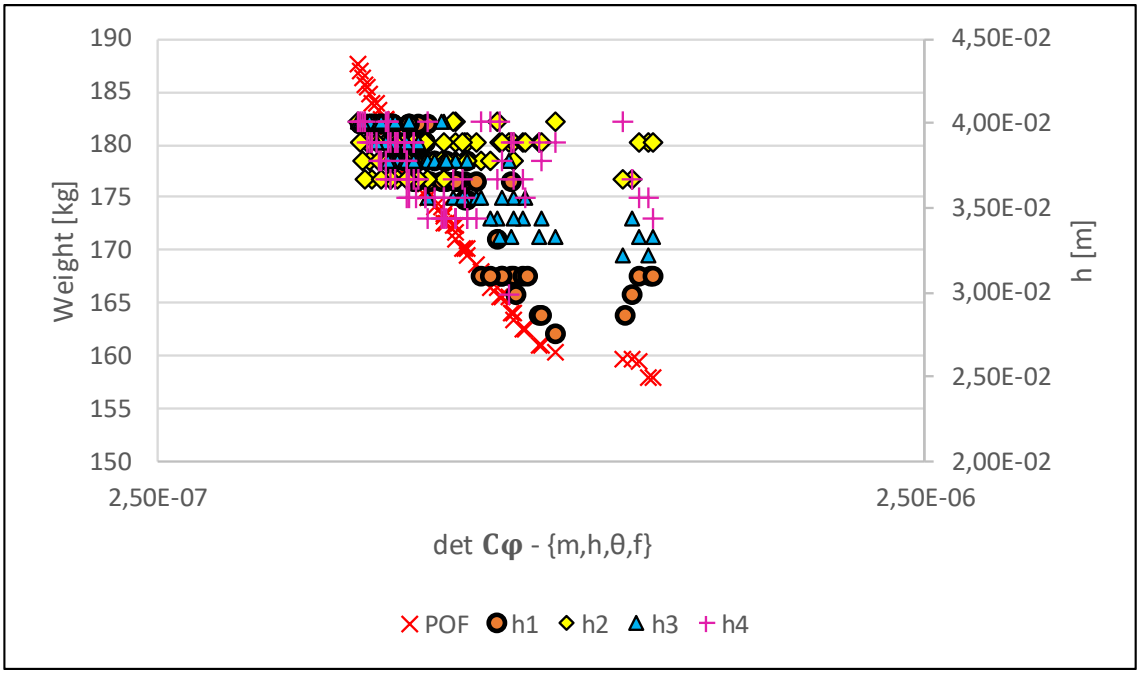

Figure A.14: Distribution of the thickness variables, along the Pareto front, obtained with the uncertainty set {**m**, **h**, **θ**, **f**}.

APPENDIX B

B.1 Boundary conditions of cylindrical shell (Figure 7.1)

	u_x	u_y	u_z	u_{xx}	u_{yy}
Clamped side	**1**	**1**	**1**	**1**	**1**
Lateral sides	**0**	**0**	**0**	**0**	**0**
Side $\overline{AB}$	**0**	**1**	**0**	**0**	**1**

B.2 Boundary conditions of fuselage-like shell (Figure 7.23)

	u_x	u_y	u_z	u_{xx}	u_{yy}
Points A, B	**1**	**1**	**1**	**0**	**0**
Side $\overline{AB}$	**1**	**0**	**1**	**0**	**0**
Points C, D	**1**	**1**	**0**	**0**	**0**
Side $\overline{AC}, \overline{BD}$	**0**	**1**	**0**	**0**	**0**
Side $\overline{CD}$	**1**	**0**	**0**	**0**	**0**

u_x: displacement along the x-axis

u_y: displacement along the y-axis

u_z: displacement along the z-axis

u_{xx}: rotation around the x-axis

u_{yy}: rotation around the y-axis

0: unrestrained degree-of-freedom

1: restrained degree-of-freedom