UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

Un Algoritmo Evolutivo Multi-Objetivo basado en Mutación por Rankeo y una Metodología para la Comparación del Desempeño de Optimizadores Estocásticos

T E S I S

Que para obtener el grado de

Doctor en Ciencias (Computación)

Presenta

Juan Arturo Herrera Ortiz

Directora de la Tesis

Dra. Katya Rodríguez Vázquez

Ciudad de México 2011

NATIONAL AUTONOMOUS UNIVERSITY OF MEXICO

POSTGRADUATE PROGRAM IN COMPUTER SCIENCES AND ENGINEERING

A Multi-Objective Evolutionary Algorithm based on Rank Mutation and a Performance Comparison Methodology to Stochastic Optimizers

T H E S I S

As the fulfilment of the requirement for the degree of

Ph.D. in Computer Sciences

by

Juan Arturo Herrera Ortiz

Advisor

Dr. Katya Rodríguez Vázquez

Mexico City 2011

Heureux l'homme occupé

Heureux l'homme, occupé de l'éternel destin,
qui, tel qu'un voyageur qui part de grand matin,
se réveille, l'esprit rempli de rêverie,
et, dès l'aube du jour, se met à lire et prie!
À mesure qu'il lit, le jour vient lentement
et se fait dans son âme ainsi qu'au firmament.
Il voit distinctement, à cette clarté blême,
des choses dans sa chambre et d'autres en lui-même;
tout dort dans la maison ; il est seul, il le croit;
et, cependant, fermant leur bouche de leur doigt,
derrière lui, tandis que l'extase l'enivre,
les anges souriants se penchent sur son livre.

Victor Hugo
Les contemplations

Dedico esta tesis a Papá, Mamá y Kena, por su amor incondicional…

Abstract

Evolutionary Algorithms have proved to be well suited for optimization problems with multiple conflicting objectives. In this PhD thesis a new Multi-Objective Evolutionary Algorithm called RankMOEA is proposed; it involves the design of innovative niching and ranking-mutation procedures which avoid the need of parameters definition and are compliant with search structure space; such procedures outperform traditional diversity-preservation mechanisms under spread-hardness situations.

Several quality indicators have been proposed in Evolutionary Multi-Objective Optimization literature and some studies have been performed in order to evaluate their inferential power. However, such inferential power becomes restricted at the time of dealing with approximations to the Pareto-optimal front with similar convergence, therefore one will be interested in how well such approximations achieve one or more of the multi-objective evaluation goals (convergence, uniformity and spread) by means of which quality differences can be inferred. Most of the existent quality indicators have been conceived in the scope of such goals, therefore it will be helpful to use them in order to untie incomparable approximation to the Pareto-optimal front. Although, a study of how appropriate the quality indicators measure what they claim to assess has not been performed. In this PhD thesis is presented a summarized review and an empirical taxonomy framework based on multi-objective evaluation goals of most of the quality indicators found in literature (about 38 indicators). Two additional contributions are reported: a new quality indicator to measure spread within the approximation to the Pareto-optimal front and a methodology to compare performance of stochastic multi-objective optimizers.

Additionally, RankMOEA is applied to approximate the Pareto Front of a Dynamic Principal-Agent model with discrete actions posed in a Multi-Objective Optimization framework, cutting edge modelling that allows to consider more powerful assumptions than those used in the traditional single-objective optimization approach. Within this new framework a set of feasible contracts is described, while other similar studies focus only on one single contract; hence a better economic analysis can be accomplished by characterizing contracts in the trade-off surface. RankMOEA performance is compared with those of other state of the art Multi-Objective Evolutionary Algorithms using the comparison methodology developed, the results suggest that RankMOEA is very effective in sampling from along the entire Pareto-optimal front and distributing the generated solutions over the trade-off surface, this by showing better spread and minor convergence error.

Agradecimientos

Mi profunda gratitud a la Dra. Katya Rodríguez Vázquez por haber dirigido esta tesis. Agradezco la oportunidad de haber recibido sus enseñanzas y conocimientos, pero sobre todo su amistad.

Mi sincero aprecio a los miembros del comité tutorial de esta tesis: Dr. Christopher Stephens Stevens, Dr. Fernando Arámbula Cosío, Dr. Carlos Artemio Coello Coello y Dr. Arturo Hernández Aguirre. Particularmente deseo agradecer sus valiosas sugerencias y críticas, las cuales hicieron del proceso una experiencia intelectual retadora y reconfortante.

A mi familia adoptiva: Abue, Tio Arturo, Isi y Daniel; porque me brindaron el calor y amor de familia estando lejos de casa.

A mis dos queridos amigos David Palma e Ignacio Palmeros. David, porque las circunstancias no nos permitieron realizar lo que algún día soñamos, pero aun así estuviste físicamente al inicio de esta aventura y estas en mi corazón al final de ella. Ignacio, porque los cafés, las interminables charlas y las gomitas nos dan nuevas perspectivas de vida, para nuestra grata sorpresa las expectativas siempre son excedidas, excepto la genialidad en nuestra amistad.

A todas las demás personas que de alguna u otra forma tuvimos el placer de convivir durante mi estadía en el IIMAS, la Facultad de Ciencias y el Centro de Idiomas de la UNAM, gracias.

Y al final, el más importante, pues él es el principio y el fin, sus ojos velan por la ciencia. Gracias Dios por la vida y la salud, por lo que me diste y retuviste, por la sabiduría que es tuya hoy y siempre.

Contents

Figures

Page

Tables

List of Symbols

Ω	decision space
$\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$	decision vectors in Ω
Λ	objective space
$\boldsymbol{F}$	vector of objective functions
f_i	i-th objective function in $\boldsymbol{F}$
k	number of objective functions
$\boldsymbol{F}(\boldsymbol{x}), \boldsymbol{F}(\boldsymbol{y}), \boldsymbol{F}(\boldsymbol{z})$	images of corresponding decision vectors in Λ
$\boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x})$	$\boldsymbol{F}(\boldsymbol{y})$ dominates $\boldsymbol{F}(\boldsymbol{x})$ according to Pareto optimal concept
$\boldsymbol{F}(\boldsymbol{y}) \prec\prec \boldsymbol{F}(\boldsymbol{x})$	$\boldsymbol{F}(\boldsymbol{y})$ strictly dominates $\boldsymbol{F}(\boldsymbol{x})$ according to Pareto optimal concept
$\boldsymbol{F}(\boldsymbol{y}) \preccurlyeq \boldsymbol{F}(\boldsymbol{x})$	$\boldsymbol{F}(\boldsymbol{y})$ weakly dominates $\boldsymbol{F}(\boldsymbol{x})$ according to Pareto optimal concept
PS^*	Pareto Optimal set, $PS^* = \{\boldsymbol{x} \in \Omega : \nexists y \in \Omega, \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x})\}$
PF^*	Pareto Optimal Front, $PF^* = \{\boldsymbol{F}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \dots, f_k(\boldsymbol{x})] : \boldsymbol{x} \in PS^*\}$
$P(t)$	population of individuals in a EA in generation t
P_{size}	MOEA's population size
Φ	mapping function that encapsulates the decoding algorithm to derive the decision vector
$\boldsymbol{\chi}, \boldsymbol{\zeta}, \boldsymbol{\varsigma}$	individuals in a defined population, $\boldsymbol{x} = \Phi(\boldsymbol{\chi})$, $\boldsymbol{y} = \Phi(\boldsymbol{\zeta})$ and $\boldsymbol{z} = \Phi(\boldsymbol{\varsigma})$
$P^*_{known}(t)$	off-line population that stores all non-dominated individuals found up to generation t
$PS_{known}(t)$	mapping of $P^*_{known}(t)$ into Ω, $\Phi(\boldsymbol{\chi}) \in PS_{known}(t) : \boldsymbol{\chi} \in P^*_{known}(t)$
$PF^*_{known}(t)$	projection of $PS_{known}(t)$ into Λ, $\boldsymbol{F}(\Phi(\boldsymbol{\chi})) \in PF^*_{known}(t) : \Phi(\boldsymbol{\chi}) \in PS_{known}(t)$
PF^*_{known}	set of objective vectors that constitute the outcome of a stochastic Multi-Objective optimizer, i.e. $PF^*_{known} = PF^*_{known}(t)$ when t reaches its maximum possible value
R, X, Υ	arbitrary sets of objective vectors
R^*	reference set of objective vectors that can be PF^*, PF^*_{known} from another MOEA, or an arbitrary set R defined by the user
$\boldsymbol{F}(\boldsymbol{x}^i) \in PF^*_{known}$	projection into Λ of the i-th decision vector, i-th objective vector in PF^*_{known}
$\lvert\cdot\rvert$	cardinality of a set
$\lVert\cdot\rVert$	an appropriate L^p norm (e.g. taxicab norm, Euclidean norm, infinity norm)

List of Publications

"Pareto Frontier Exploration using Genetic Algorithms to Agent-Principal Dynamic Model with Discrete Actions". Article published by Centre for Research and Teaching in Economics. May 2010. México, D.F.

"Amount and Type of Information: A new GA-Hardness Taxonomy", Genetic and Evolutionary Computation Conference, GECCO'2010. Article published in ACM, ISBN 978-1-4503-0072-8, pp. 835-836. Portland, Or., EEUU.

"A RankMOEA to Approach the Pareto Front of a Dynamic Principal-Agent Model", Genetic and Evolutionary Computation Conference, GECCO'2011. Article published in ACM, ISBN 978-1-4503-0557-0, pp. 785-792. Dublin, Irland.

"Autonomous Robot Navigation based on the Evolutionary Multiobjective-Optimization of Potential Fields". Article accepted to be published in the Journal of Engineering Optimization edited by Taylor & Francis Group.

"Pareto Frontier of a Dynamic Principal-Agent Model with Discrete Actions: An Evolutionary Multi-Objective Approach". Article under revision in the Journal of Computational Economics.

"Pareto Frontier of a Dynamic Principal-Agent Model with Continuous Actions: An Evolutionary Multi-Objective Approach". Article to be send to a Journal of Economics.

Chapter 1 Introduction

[*Where the world ceases to be the scene of our personal hopes and wishes, where we face it as free beings admiring, asking and observing, there we enter the realm of Art and Science…*]

Albert Einstein

1.1 MULTI-OBJECTIVE OPTIMIZATION

Innumerable situations in the real world involve in natural way problems with multiple criteria or objectives to be optimized, such picture emerges frequently in scientific and engineering areas. *Multi-Objectives Optimization* (MO), also called Multi-Criteria Optimization, studies the process of simultaneously taking optimal decisions in the presence of trade-offs between two or more conflicting objectives. This mathematical discipline was conceived in the middle of last century, starting from principles proposed by Koopmans [1] and Kuhn & Tucker [2], whose inspiration was based on Edgeworth [3] and Pareto's [4] earlier works. MO is considerably more elaborate than the classical optimization approach, where the decision of an optimal point is trivial.

Maximizing profit and minimizing the cost of a product, maximizing performance and minimizing fuel consumption of a vehicle, and minimizing weight while maximizing the strength of a particular component are examples of *Multi-Objective Optimization Problems* (MOPs). Since MO implies to optimize conflicting objectives subject to certain constraints, most of the time it is impossible to determine a unique solution. Hence, MOPs are characterized by a set of alternative optimal solutions that must be considered as equivalents given the lack of information about relevance of one objective with regard to the others; such set of solutions is discriminated based on dominance relations which entail a pre-order structure in a multidimensional objective function space.

In MO a space for decision variables and a space for their objective functions evaluation are considered. In real valued functions, those two spaces are related by a mapping $\boldsymbol{F}: \Re^n \rightarrow \Re^k$. It is assumed that a solution to the MOP can be defined in terms of a decision vector $\boldsymbol{x} = [x_1, x_2, \dots, x_n]^T$ in the decision space $\Re^n$. The set of imposed constraints defines a feasible region $\Omega \subseteq \Re^n$ in the decision space along with its corresponding image $\Lambda \subseteq \Re^k$ on the objective space, which involves the evaluation of every point in Ω with the $k \geq 2$ conflicting objective functions $\boldsymbol{F}(\boldsymbol{x}) = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \dots, f_k(\boldsymbol{x})]$ that constitute the MOP, see Figure 1.1.

By definition, there is a possibly infinite set of optimal solutions which are found at the frontier of Λ and are called the Pareto Optimal Front (PF^*), while their corresponding decision variables values in Ω are called the Pareto Optimal Set (PS^*). A solution $\boldsymbol{x}$ in PS^* is Pareto optimal (also called non-dominated with respect to Ω), which means that there is no other solution $\boldsymbol{y} \in \Omega$ for which $\boldsymbol{F}(\boldsymbol{y})$ dominates $\boldsymbol{F}(\boldsymbol{x})$

(denoted by $\boldsymbol{F(y) \prec F(x)}$). $\boldsymbol{F(y)}$ is said to dominate $\boldsymbol{F(x)}$ if and only if $\boldsymbol{y}$ improves any objective to optimize with respect to $\boldsymbol{x}$ without inducing some simultaneous deterioration in at least another objective, e.g. assuming only minimization $\boldsymbol{F(y)}$ is partially less than $\boldsymbol{F(x)}$, i.e., $\forall f_i \in \boldsymbol{F}: f_i(\boldsymbol{y}) \leq f_i(\boldsymbol{x}) \wedge \exists f_i \in \boldsymbol{F}: f_i(\boldsymbol{y}) < f_i(\boldsymbol{x})$.

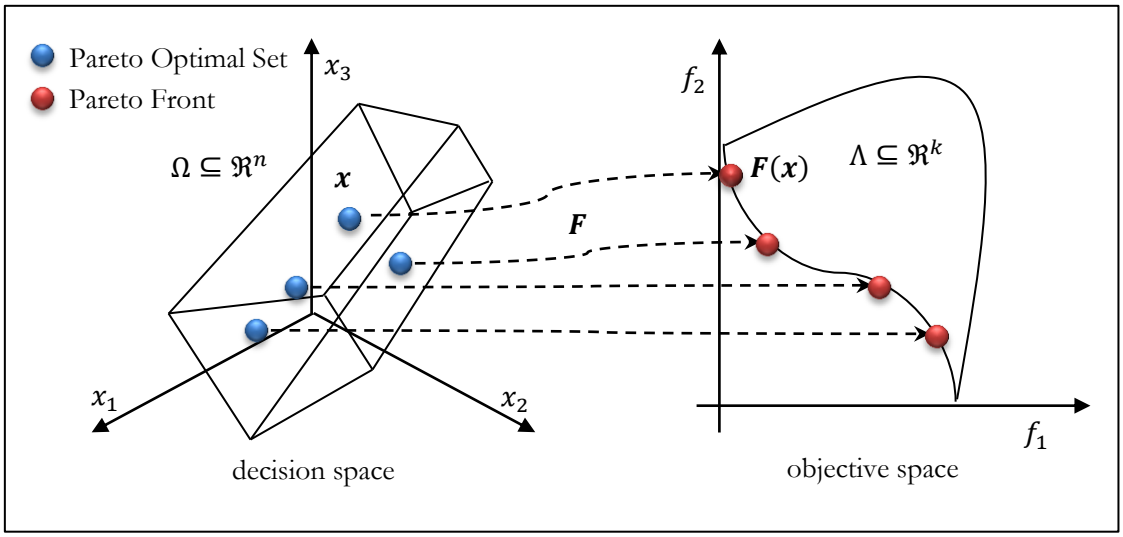

Figure 1.1 Decision and objective spaces in MO. A solution parameterization $\boldsymbol{x}$ is mapped by a vector function $\boldsymbol{F}$ into a vector in the objective space.

In mathematical terms, without loose of generality, a MOP can be written as:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{F(x)} \\ \text{s.t.} \quad & \boldsymbol{x} \in \Omega \end{aligned} \tag{1.1}$$

Since each Pareto optimal solution represents a different compromise among objectives, finding different Pareto optimal solutions implies finding the structure of the trade-off surface involved in the MOP.

1.2 Evolutionary Optimization

Evolutionary Algorithms (EAs) are stochastic methods of search often applied to optimization [5]. As the history of the field suggests there are many variants of EAs. The common underlying idea behind all these techniques is an evolutionary analogy of the "survival of the fittest" which takes its inspiration from the

modern evolutionary synthesis, where natural selection can be seen as a learning process which a long the time generates fittest individuals to survive in a defined environment.

An EA maintains a set of individuals $P(t)$ or population of strings at each stage or generation t, each string $\boldsymbol{\chi} \in P(t)$ is also called chromosome and encodes a candidate solution to the problem's domain (classically encoded in a binary string or real vector, but now in almost any conceivable representation). A mapping function Φ encapsulates the decoding algorithm to derive the decision vector $\boldsymbol{x} = \Phi(\boldsymbol{\chi})$ from $\boldsymbol{\chi}$. As in nature, every individual has a fitness value associated to its performance in the environment, which is defined as an abstract measure of the maximizing quality function of the problem. $P(t)$ is evolved from generation t to generation $t + 1$ attempting to adapt itself to the difficulties of the environment by varying and selecting the genetic material of the individuals (notion of inheritance), environment pressure causes natural selection, thus a rise in the fitness of the population is induced. Selection weeds out poor candidate solutions in generation t by favouring individuals according to their fitness, desirably fittest individuals are chosen to seed a pool of parents $P'(t)$. Genetic material variation from parents is used to generate offspring, which compete based on their fitness with their parents for a place in generation $t + 1$. This cycle of birth/death influenced by fitness is iterated until a candidate solution with sufficient quality is found or a previously set computational limit is reached.

The combined application of genetic material variation and selection commonly leads to improving fitness values in consecutive generations, thus biasing solutions towards promising regions of the search space. It is easy to see such process as if the evolution is optimizing, or at least approximating, by approaching optimal values closer and closer over its course. Alternatively, evolution is often seen as a process of adaptation; from this perspective, the fitness is not seen as an objective function to be optimized but as an expression of environmental requirements, matching these requirements more closely implies an increased viability reflected in a higher number of offspring. The evolutionary process makes the population adapt to the environment better and better.

The ability of EAs to maintain a diverse set of candidate solutions not only provides a way to escape from local optimum, but a way to cope with large and discontinuous search spaces. Besides, if several copies of a good solution can be maintained, it provides a natural and robust way of dealing with problems where there is noise or uncertainty associated with the assignment of fitness to candidate solutions.

Genetic Algorithms (GAs) are one of the best known approaches in EAs, their abstraction of the modern evolutionary synthesis is at the level of individuals [6], thus mating and mutation are involved as methods of genetic material variation, with a pre-eminence of mating over mutation. Selected parents $P'(t)$ are mated among them by swapping parts of their genetic material producing the intermediate generation $P''(t)$; this mechanism accelerates search process by exploitation of the gathered information. Mutation provides diversity by performing a small random variation to a single element of $P''(t)$, i.e. exploration of new regions in the search space, the offspring achieved after mutation is stored in $P'''(t)$.

1.3 EVOLUTIONARY MULTI-OBJECTIVE OPTIMIZATION

In recent years, several stochastic search strategies have been developed and adapted in order to deal with MO, because most of the time the complexity of the underlying MOPs prevents close solution methods from being applicable [7], since generating PS^* can be computationally expensive and often infeasible. Such strategies find Pareto Front approximations based on multiple execution of their optimizer algorithm. It is in the late 1960s, when Rosenberg suggested with a study in his PhD thesis [8] to apply EAs to MO, that the area known as *Evolutionary Multi-Objective Optimization* (EMO) was born.

EAs viability to deal with MO is related to their population approach that suits well to find multiple solutions in the same algorithm execution, their diversity-preservation mechanism that can be exploited to keep heterogeneous candidate solutions, their ability to deal with search and multiple decision simultaneously, their implicit parallelism, among other some well-known intrinsic EAs advantages. Aforesaid features constitute EMO as a reliable methodology to achieve two ideal goals of MO: attaining good convergence to PF^* and maintaining the distribution of the Pareto Front approximation as diverse as possible.

Interest in EMO has considerably increased during the last two decades, when many *Multi-Objective Evolutionary Algorithms* (MOEAs) have been proposed. MOEAs constitute a promising approach to deal with real world MOPs [9], even they usually do not guarantee to identify optimal trade-offs, but to find good assessments, i.e., sets of solutions whose objective vectors are not too far away from PF^*.

One of the major heuristic experimental goals within EMO research is to compare well-engineered MOEAs in terms of efficiency and effectiveness as regards selected MOPs through the use of appropriate

metrics. These metrics are an essential part of a successful experimental methodology able to characterize accurately the performance of different algorithms. A trade-off between efficiency and effectiveness is always present in heuristic approaches. How to measure efficiency in the sense of computational effort has been widely studied along Computer Sciences history. Contrariwise, how to measure effectiveness in the sense of the accuracy and convergence of experimental outcomes is still an open problem in EMO.

1.4 STATEMENT OF THE PROBLEM

MOEAs have evolved by reformulating and improving some of their inherent elements as fitness assignment, diversity-preservation mechanism and elitism, a continuous improvement process is pursued. When solving real MOPs, good assortment of the Pareto Front approximation (PF^*_{known}, outcome of the MOEA) is preferred since wider variety of heterogeneous solutions could give a better sight of the trade-off surface. Thus, better spread and dispersion of the candidate solutions in PF^*_{known} with good convergence towards PF^* will contribute with important information of the approximated structure of PF^*, giving rise to better informed decision making process by choosing the solutions in PF^*_{known} that best meet with compromises among objectives.

Therefore, the design of a MOEA that involves a robust diversity-preservation mechanism with non-parameter definition compliant with search structure space and, consequently, able to achieve PF^*_{known}s with low convergence error and good spread and dispersion is desirable. Furthermore, how to compare the performance of stochastic multi-objective optimizers is not clear at this time, since several effectiveness metrics (referred as quality indicators hereinafter) have been designed. Thus, a statistically confidence methodology which may discriminate and involve a suitable subset of quality indicators in order to assess MOEAs outcomes compliant with dominance relations between non-dominated sets could provide a first attempt to the problem of measuring outcomes effectiveness in EMO.

1.4.A How to improve the diversity of MOEAs?

Diversity-preservation mechanisms impulse divergence in tangential direction to the promising regions discovered by the MOEA, this through probability selection bias in less conglomerated regions. Most of the designed diversity mechanisms in EMO require parameters specification or are unable to deal with

incommensurable objectives or were not designed to be compliant with the search space. Hence, three premises are considered in order to design a new MOEA which emphasizes spread and dispersion of PF^*_{known} and preserves equilibrium between exploitation and exploration:

- Since Pareto dominance rules sort candidate solutions in a certain order according to their proximity to the frontier of Λ, some advantage can be taken from such arrangement by intensifying exploration in candidate solutions far from the frontier and reducing exploration in candidate solutions close to the frontier. This assuming that the first type of solutions does not have much information about PF^* structure and need more effort to achieve a good performance.
- The structure of the search is defined by Ω and not by Λ, thus, diversity preservation mechanisms should work better in Ω if they are compliant with Ω structure. Hence, exploitation of the information could be successful by mating nearby candidate solutions in Ω since such process is less disruptive.
- In most of the cases, after a certain threshold in the evolutionary process of MOEAs, the number of non-dominated solutions grows rapidly, thus reduced mutation in solutions closer to the frontier of Λ that are less conglomerated in Ω should improve performance by controlling exploration and preserving the emphasized exploitation in such regions.

1.4.B How to measure outcomes quality of MOEAs?

A wide number of quality indicators (quality features of the found non-dominated solutions set or PF^*_{known} expressed in a quantitative way) have been proposed in EMO literature along history, even though it is not obvious which of such indicators must be used in practice [10] [11]. Zitzler *et al* [12] propose that the number of indicators to use should be proportional to the number of objectives to be optimized; however, a wide stock of indicators does not guarantee a precise and detailed PF^*_{known} description. In addition, several studies have been performed in order to discriminate quality indicators by their inferential power, though, such inferential power becomes restricted at the time of dealing with PF^*_{known}'s with similar convergence, therefore one will be interested in how well PF^*_{known} achieves one or more of the MO evaluation goals (convergence, uniformity and spread) by means of which quality differences can be inferred. Most of the existent quality indicators have been conceived in the scope of such goals, therefore it will be helpful to use them in order to untie incomparable PF^*_{known}'s.

Actually, several papers have included analysis from some indicators [10] [12] [13] [14] [15] [16] [17] [18]. Although, a study of how accurately the quality indicators measure what they claim to assess has not been performed (in this case one or more MO evaluation goals), an important concern at the time to select the most adequate quality indicator(s). Thus, a summarized review and an empirical taxonomy framework based on multi-objective evaluation goals of most of the quality indicators found in literature should be accomplished in order to compose a methodology to compare performance of stochastic multi-objective optimizers.

1.5 DISSERTATION CONTRIBUTIONS

The main contributions of this thesis to achieve the degree of PhD in Computer Sciences are:

- A new efficient and effective MOEA called RankMOEA, which includes a robust diversity-preservation mechanism with non-parameters definition compliant with search space structure and able to accomplish good performance over spread-harness situations.
- A methodology to compare the quality of outcomes of MOEAs, covering the analysis of most of the quality indicators known at the present, for this purpose the following sub-targets are achieved:
  - An empirical taxonomy framework of quality indicators based on MO evaluation goals accuracy, i.e. how well the quality indicator measures the MO evaluation goals, including important features (some have already been reported for a few indicators but not studied for all of them e.g. set dependence, evaluated characteristics, monotony, relativity, computational complexity), attempting to develop a guide to choose suitable quality indicators according to experimental goals.
  - A new quality indicator to measure spread within PF^*_{known}s, achieving a more accurate assessment since former indicators offer a superior bound or are susceptible to convergence error and uniformity variation.
  - A methodology to quantify the outcomes quality of stochastic multi-objective optimizers and compare their performance, conceived as statistically confident and compliant with dominance relations between non-dominated sets, using only a suitable subset of quality indicators that fulfil requirements according to MO evaluation goals.

- Additionally, RankMOEA is applied to approximate the Pareto Front of a Dynamic Principal-Agent model with Discrete Actions posed in a multi-objective optimization framework, cutting edge modelling that allows to consider more powerful assumptions than those used in the traditional single objective optimization approach. Within this new framework a set of feasible contracts is described, while others similar studies only focus on one single contract. The results achieved with RankMOEA show better spread and minor error than those obtained by already well-known MOEAs, allowing to perform better economic analysis by characterizing contracts in the trade-off surface.

1.6 DISSERTATION OUTLINE

The remainder of this PhD thesis is organized as follows. Chapter 2 includes a description of the key concepts in EMO, a brief overview of most relevant state of the art MOEAs and a summary of most of the MO quality indicators found in the literature. The proposed MOEA, called RankMOEA, which includes a new diversity-preservation mechanism with non-parameters definition is described in detail and tested over spread-harness situations in Chapter 3. A complete study of the summarized MO quality indicators with regard to the outperformance relations and the MO evaluation goals is shown in Chapter 4, also an experimental methodology to compare stochastic multi-objective optimizers is proposed and tested. Chapter 5 presents the performance comparison of RankMOEA versus some well-known MOEAs over various theoretical MOPs. Chapter 6 introduces the Dynamic Principal-Agent problem as a MOP, shows the outstanding PF^*_{known} achieved with RankMOEA and presents some conclusions that can be deduced from the analysis of the achieved results. Finally, conclusions and future work are drawn in Chapter 7.

Chapter 2 Review of the Literature

[*It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is the most adaptable to change*…]

Charles Darwin

2.1 INTRODUCTION

In order to propose alternatives solution to the two main goals posed in this PhD thesis: improve MOEAs diversity and compose a methodology to compare performance of stochastic multi-objective optimizers, this chapter presents the review of the state of the art concerning to key issues in EMO, the idea behind some well-known successful MOEAs and a summarized review of most of the quality indicators found in literature (38 indicators).

2.2 KEYS ISSUES IN EVOLUTIONARY MULTI-OBJECTIVE OPTIMIZATION

Zitzler [19] reformulates, in a general way, the optimization aspiration pursued when optimizing MOPs based on three goals:

- The distance of the resulting PF^*_{known} to PF^* should be minimized.
- The distribution of PF^*_{known} should be uniform in most cases.
- The spread of PF^*_{known} should be maximized, i.e. for each objective a wide range of values should be covered by the non-dominated solutions.

According to Zitzler, in the intention of achieving the aforesaid goals, two major problems must be addressed when an EA is applied to MO:

- How to accomplish fitness assignment and selection, respectively, in order to guide the search towards PS^*.
- How to maintain a diverse population in order to prevent premature convergence and achieve a well distributed and well spread PF^*_{known}.

In the following, a categorization of general techniques which deal with these issues is presented; the modified usage of elitism is also included, given that its notion interacts with both situations.

2.2.A Fitness Assignment and Selection

Within EMO, the way in which fitness is assigned and selection is performed is modified in order to deal with several objectives simultaneously. Three types of fitness assignment and selection are distinguished in the state of the art [9] [19]: aggregation-based, objective-based and Pareto dominance-based.

Aggregation-Based Fitness Assignment

The objectives to be optimized are combined into one single linear or nonlinear parameterized function; the parameters of such function are not changed for different optimization runs, but systematically varied during the same run. The potential bias towards convex portions of PF^* may restrict the effectiveness of this approach.

Objective-Based Fitness Assignment

The most suitable sequence of the objective(s) to be optimized is chosen during the selection phase, i.e., only one subset of objectives is optimized at the time by the entire population or by portions of the entire population. E.g. initially only the most important objective is optimized, as the population evolves more objectives are consecutively considered in the process according to their pre-eminence [20]; the mating pool (selected parents) is filled with equal portions according to the distinct objectives [21]. This approach may have bias towards extreme solutions and be sensitive to non-convex PF^*.

Pareto Dominance-Based Fitness Assignment

Every individual in the population is ranked according to the Pareto dominance concept, the rank of an individual determines its fitness value where it is clearly related to the whole population, contrariwise to the abovementioned techniques where the raw fitness of an individual is calculated independently of the other individuals. Several distinctive rules to rank have been conceived along EMO history, the first one was proposed by Goldberg [5], whose idea is to assign rank one to all non-dominated individuals in the population and temporarily remove them from the population, then, the next non-dominated individuals are assigned rank two and also temporarily remove them from the population, then, the next non-dominated individuals are assigned rank three and so forth (see Figure 2.1 a). Equation (2.1) describes this recursive Pareto rank rule where $\boldsymbol{x} = \Phi(\boldsymbol{\chi})$ and $\boldsymbol{y} = \Phi(\boldsymbol{\zeta})$.

$$rank_g(\boldsymbol{\chi}, t) = \begin{cases} 1 & \text{iff } \nexists \boldsymbol{\zeta} \in P(t): \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x}) \\ \max\limits_{\boldsymbol{\zeta} \in P(t): \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x})} \left[rank_g(\boldsymbol{\zeta}, t) \right] + 1 & \text{otherwise} \end{cases} \quad (2.1)$$

Fonseca & Fleming [22] proposed a Pareto rank rule where every individual has a rank equivalent to the number of individuals that dominate it increased by one, thus non-dominated individuals have rank one (see Figure 2.1 b). This Pareto rank rule is defined in Equation (2.2).

$$rank_{ff}(\boldsymbol{\chi}, t) = |\{\boldsymbol{\zeta} \in P(t): \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x})\}| + 1 \quad (2.2)$$

In 1998, an innovative Pareto rank rule which involves an off-line population $P^*_{known}(t)$ that stores all non-dominated individuals found up to generation t was proposed by Zitzler & Thiele [23]. The rank for each individual in $P(t)$ is computed with the strength of the individuals in $P^*_{known}(t)$ that dominate it, the strength of an individual $\boldsymbol{\zeta}$ in $P^*_{known}(t)$ is proportional to the number of population members it dominates, $strength'_{\boldsymbol{\zeta}} = |\{\boldsymbol{\varsigma}: \boldsymbol{\varsigma} \in P(t) \wedge \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{z})\}|$ where $\boldsymbol{z} = \Phi(\boldsymbol{\varsigma})$ (see Figure 2.1 c). Equation (2.3) describes this Pareto rank rule.

$$rank_{zt}(\boldsymbol{\chi}, t) = \sum_{\boldsymbol{\zeta} \in P^*_{known}(t): \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x})} \frac{strength'_{\boldsymbol{\zeta}}}{|P(t)| + 1} + 1 \quad (2.3)$$

Van Veldhuizen [24] proposed in his PhD thesis a simpler Pareto rank rule defined in Equation (2.4), here non-dominated individuals get rank zero and dominated individuals get rank one (see Figure 2.1 d).

$$rank_{vv}(\boldsymbol{\chi}, t) = \begin{cases} 0 & \text{iff } |\{\boldsymbol{\zeta} \in P(t): \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x})\}| = 0 \\ 1 & \text{otherwise} \end{cases} \quad (2.4)$$

Later, Zitzler & Thiele [25] improved their previously proposed Pareto rank rule in order to avoid the situation that individuals dominated by the same $P^*_{known}(t)$ members have identical fitness values, within this new approach for each individual dominating and dominated individuals are taken into account (see Figure 2.1 e). In detail, the rank of an individual in $P(t)$ is composed by a redefined strength value $strength''_{\boldsymbol{\zeta}} = |\{\boldsymbol{\varsigma}: \boldsymbol{\varsigma} \in \{P(t) + P^*_{known}(t)\} \wedge \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{z})\}|$ of individuals in $P(t)$ and $P^*_{known}(t)$ that dominate it and a density estimate incorporated as the inverse of the distance in the objective space to the k-nearest neighbour in $P(t)$ and $P^*_{known}(t)$. As a common setting, they proposed to use k as the square-root of the sample size. Equation (2.5) describes this Pareto rank rule, where two is added in the denominator of the density estimate to ensure that its value is greater than zero and less than one.

$$rank_{zt2}(\chi, t) = \sum_{\zeta \in \{P(t)+P^*_{known}(t)\}: \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x})} strength''_{\zeta} + \frac{1}{\sigma^{\chi}_{k-nearest} + 2} \quad (2.5)$$

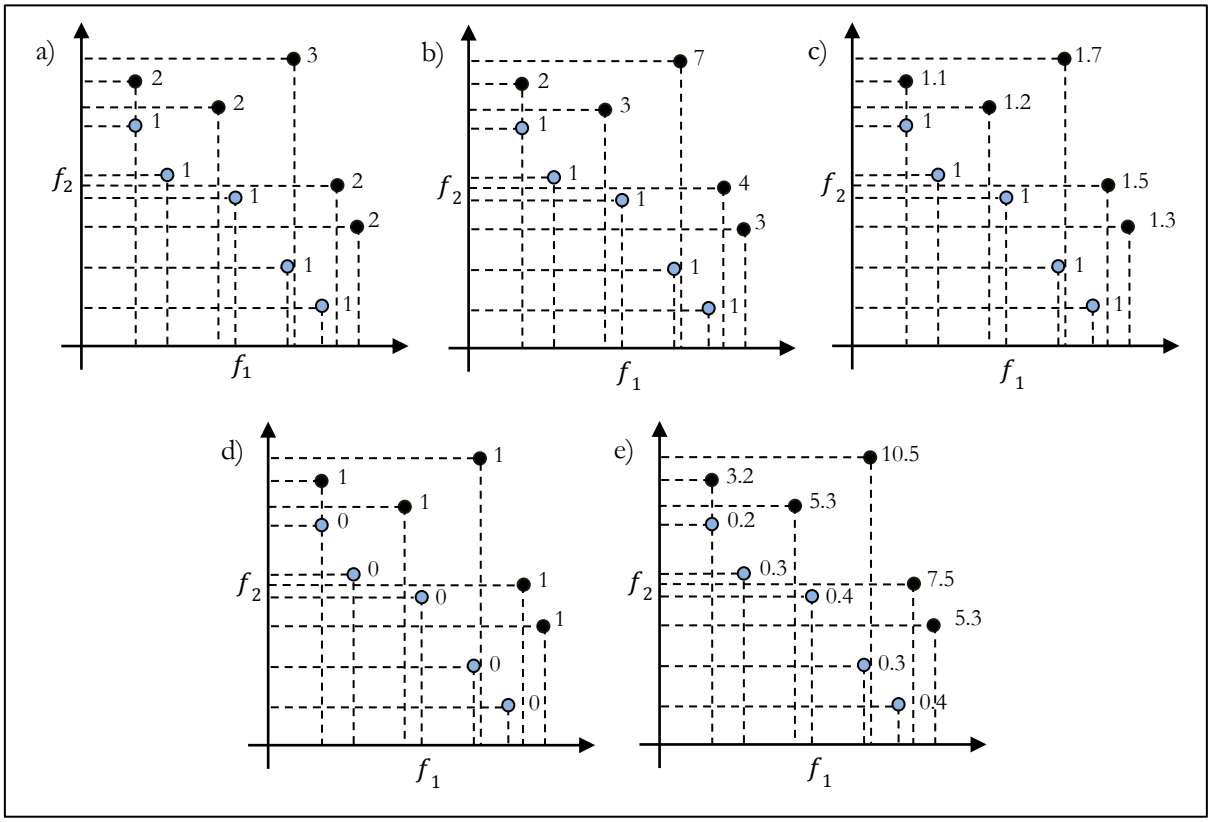

Figure 2.1 Graphic examples of Pareto rank rules: a) Goldberg b) Fonseca & Fleming c) Zitzler & Thiele first version d) Van Veldhuizen e) Zitzler & Thiele improved version.

2.2.B Elitism

Elitism or an elitist strategy is a mechanism which ensures that the chromosome(s) of the most highly fit member(s) of $P(t)$ are passed on to the next generation without being altered in order to prevent losing them due to sampling effects or operators disruption. In EMO, elitism is extended to the concept of an offline population $P^*_{known}(t)$, which stores all non-dominated solutions found up to epoch t. The set of decision vectors decoded from $P^*_{known}(t)$ is $PS_{known}(t)$, and its corresponding image in Λ is $PF^*_{known}(t)$. Since generally PF^* is an infinite set, the maximum size of $P^*_{known}(t)$ must be taken in consideration by physical memory restrictions, thus most of the time its growth is controlled, e.g. using

mechanisms of clustering or truncation. Two classes of elitism can be distinguished within the EMO approach: isolated and interactive.

Isolated Elitism

$P^*_{known}(t)$ can act only as a repository unit to store non-dominated individuals, it is updated along the evolutionary process.

Interactive Elitism

$P^*_{known}(t)$ can be interactive, which means that besides being updated it can cooperate in the evolutionary process by selecting new parents from it to generation $t+1$.

2.2.C Diversity Preservation Mechanisms

A simple EA tends to converge towards a single solution and often losses solutions due to three effects: selection pressure, selection noise and operator disruption [23]. To overcome this problem, which is known as genetic drift, several methods called diversity-preservation mechanisms have been developed, the ones most frequently used in EMO are briefly summarized here. A diversity-preservation mechanism attempts to impulse divergence in tangential direction to the promising regions discovered by the MOEA, this through probability selection bias towards less conglomerated regions.

Fitness Sharing

Fitness sharing is a niching mechanism which was proposed by Goldberg & Richardson [26]; a niche describes the relational position of a species or population in its ecosystem to each other, i.e., how an organism or population responds to the distribution of resources and competitors and how it in turn alters those same factors. In EMO, individuals in the same niche have presumably similar features and, as in nature, they have to share available resources; thus the fitness value of a certain individual is more degraded as more individuals are located in the same niche; such idea allows to maintain stable subpopulations (niches) providing additional selective pressure. The niche size σ_{share} defines a neighbourhood in terms of distance between individuals and can be measured in the genotype $\|\boldsymbol{\chi}, \boldsymbol{\zeta}\|$, the phenotype $\|\boldsymbol{x}, \boldsymbol{y}\|$ or the objective space $\|\boldsymbol{F}(\boldsymbol{x}), \boldsymbol{F}(\boldsymbol{y})\|$ (see Figure 2.2 a). Mathematically, the shared fitness $sf_{\boldsymbol{\chi}}$ of an individual $\boldsymbol{\chi}$ is equal to its raw fitness divided by its niche count, see Equation (2.6).

$$sf_{\chi} = \frac{fitness_{\chi}}{\sum_{\zeta \in P(t)} sh(\|\chi, \zeta\|)} \tag{2.6}$$

An individual's niche count is a measure of how saturated is its niche, it is computed as the sum of sharing function values $sh(\|\chi, \zeta\|)$ between itself and the individuals in $P(t)$, sharing functions commonly used are of the form shown in Equation (2.7) where α_{share} regulates the shape of the function $sh(\|\chi, \zeta\|)$.

$$sh(\|\chi, \zeta\|) = \begin{cases} 1 - \left(\frac{\|\chi, \zeta\|}{\sigma_{share}}\right)^{\alpha_{share}} & \text{if } \|\chi, \zeta\| < \sigma_{share} \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

Restricted Mating

Restricted mating [22] is a mechanism which ideally limits mating to individuals with similar genetic material, thus it is expected that offspring with similar features to their parents can be generated. It is preferred to mate individuals within the same niche since it may avoid the formation of lethal individuals and therefore improve the online performance by controlling diversity (see Figure 2.2 b). In a homologous way to fitness sharing, the parameter of niche size σ_{mate} should be defined.

Reinitialization

Reinitialization of the whole or parts of $P(t)$ after a certain number of generations or whenever the search stagnates is a way to prevent premature convergence that was proposed by Fonseca & Fleming [27].

Clustering

Zitzler & Thiele [23] proposed to involve a hierarchical clustering as a mechanism to maintain diversity. In their approach the offspring in generation t and the individuals in $P^*_{known}(t)$ compete for a place in generation $t + 1$, thus a selection of survivors compliant with their good distribution in the objective space should be used. Initially every individual is seen as a cluster, distance between clusters is computed in Λ, the two clusters with the minimum distance between them are combined into a single cluster, this process continues until there are as many clusters as individuals that can be preserved to generation $t + 1$. The centroid in every cluster is selected as the individual to be preserved. The main drawback of this approach is that tends to delete individuals whose projection in $PF^*_{known}(t)$ is an exterior solution.

Grid Mapping

A mechanism that attempts to maintain a uniform sampling of the solutions over PF^* using a grid mapping was suggested by Knowles & Corne [28]. This procedure recursively divides Λ into hypercubes, so it is possible to compute the density of every hypercube as the number of individuals that it contains (see Figure 2.2 c), thus an individual with high density in the hypercube that contain it has less probability to survive when $P^*_{known}(t)$ is redefined. A priori knowledge of the geometric structure of PF^* is necessary in order to define the most appropriate number of divisions in every dimension of Λ.

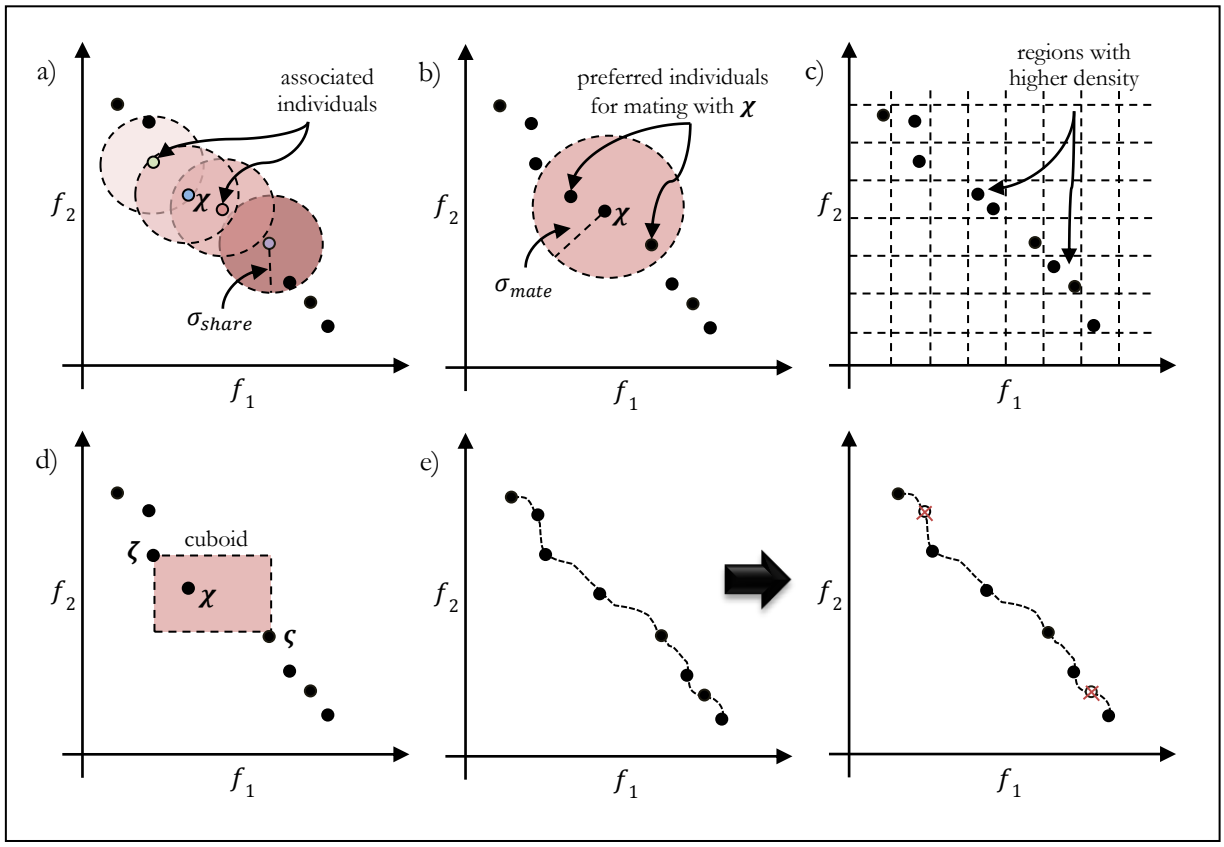

Figure 2.2 Graphic examples of diversity-preservation mechanisms: a) Fitness sharing b) Restricted Mating c) Grid Mapping d) Crowding e) Truncation.

Crowding

Crowding, proposed by Deb et al [29], was conceived with the idea of non-parameter definition; this mechanism estimates the density of the projection in Λ of an individual by taking the average distance of the two points on either side of such projection along each of the objectives. This quantity, called

crowding distance, serves as an estimate of the size of the largest cuboid enclosing the aforesaid individual without including any other individual in $P(t)$. The higher the crowding distance of an individual, the higher the disaggregation of the points surrounding its projection in Λ, whereas lower the crowding distance of an individual, higher the concentration of the points surrounding its projection in Λ (see Figure 2.2 d). The crowding distance can be used to bias selection in the evolutionary process towards a uniformly spread out PF^*_{known}.

Truncation

In order to improve the clustering approach, Zitzler & Thiele [23] proposed to use a truncation mechanism when $P^*_{known}(t)$ exceeds its limit size. The idea is to maintain a highly representative sampling of $P^*_{known}(t)$ by deleting iteratively an individual of $P^*_{known}(t)$ until its size is suitable. The individual to be deleted is the one with the minor distance to its closest neighbour in Λ, in case of a tie, the distance to the second closest neighbour is considered and so forth (see Figure 2.2 e).

2.3 SOME WELL-KNOWN MULTI-OBJECTIVE EVOLUTIONARY ALGORITHMS

Six MOEAs whose diversity-preservation mechanisms and features have had significant contributions to EMO research are briefly summarized in the following. Such MOEAs have been chosen with the purpose of studying their advantages and drawbacks in order to propose a new MOEA in this PhD thesis. Table 2.1 draws the most relevant features (evolutionary approach, fitness assignment, diversity-preservation mechanism, parents' selection and elitism) of every MOEA here described.

VEGA

The *Vector Evaluated Genetic Algorithm* (VEGA) was proposed by Schaffer [21], it is a GA with objective-based fitness assignment and generational elitism, i.e., no offline population is used. VEGA modifies the selection process in order to favour survival of the best individuals in every objective and those that are better than the average in more than one objective. $P'(t)$ is generated by shuffling k subpopulations, every subpopulation is formed by $|P(t)|/k$ individuals chosen with regard to one of the k objectives to be optimized. VEGA evolves $P(t)$ until build a suitable set of solutions to the MOP. Although some serious drawbacks are known, as instability to search in concave PF^*s and inability to create middling individuals, this algorithm has been used as a strong point of reference up to now.

MOGA

The *Multi-Objective Genetic Algorithm* (MOGA) was proposed by Fonseca & Fleming [27] [30], it is a GA with Pareto dominance-based fitness assignment and isolated elitism. MOGA was innovative with its fitness assignment process by attempting to avoid the genetic drift effect. Fitness assignment is performed with Fonseca & Fleming's Pareto rank rule using an interpolation function (usually linear but no necessary). Then fitness sharing is applied in order to maintain a controlled selective pressure, the niche count is computed in Λ with an infinity norm $\|\cdot\|_\infty$, thus it is expected that $P(t)$ may evolve with a uniform distribution. In order to generate $P'(t)$, restricted mating and reinitialization are utilized.

NPGA

The *Niched Pareto Genetic Algorithm* (NPGA) was proposed by Horn *et al* [31], it is a GA with Pareto dominance-based fitness assignment and generational elitism. NPGA enhances selection process by using a modified version of tournament selection with replacement called tournament by Pareto dominance. In tournament by Pareto dominance two candidates to tournament and a comparison subset $subP(t) \subset P(t)$ are randomly chosen, $|subP(t)| \in \{1,2,..,|P(t)|-2\}$, every candidate is compared with $subP(t)$, if any of both is dominated by $subP(t)$ and the other is not, the one that is non-dominated is the winner., contrariwise if both are dominated or non-dominated a fitness sharing modification called equivalence class sharing is used, here the winner is the candidate whose niche count in the phenotypic space is lower.

NSGA-II

The *Non-dominated Sorting Genetic Algorithm II* (NSGA-II) was proposed by Deb *et al* [29], it is a GA with Pareto dominance-based fitness assignment and interactive elitism. NSGA-II was one of the first attempts to avoid parameter definition; the main idea is to create layers of individuals (also called sub-Pareto Fronts approximations) according to its dominance using Goldberg's Pareto rank rule and crowding. $P'(t+1)$ is generated from $P^*_{known}(t)$ and $P'''(t)$ using a binary tournament with replacement considering in a hierarchical pre-eminence the rank and crowding distance. NSGA-II outstanding efficiency is because of reducing multiple objectives optimization to one single criterion using the non-dominated sorting.

SPEA2

The *Strength Pareto Evolutionary Algorithm 2* (SPEA2) was proposed by Zitzler *et al* [25], it is an EA with Pareto dominance-based fitness assignment and interactive elitism. SPEA2 is inspired in immune systems approach; it uses $P^*_{known}(t)$ to evaluate $P(t)$ fitness with Zitzler & Thiele's improved Pareto rank rule,

while $P^*_{known}(t)$ is updated with non-dominated individuals in $P^*_{known}(t-1)$ and $P(t-1)$. Given that $P^*_{known}(t)$ should always keep the same size that $P(t)$, two situations can occur: if $P^*_{known}(t)$ exceeds such size a truncation process is used to achieve a reduced representation, if $P^*_{known}(t)$ lacks of such size it is complemented with dominated individuals. In order to generate $P'(t)$, a binary tournament with replacement is performed over $P^*_{known}(t)$; non exceptional modifications over mating or mutation are used.

Table 2.1. Summarized features of VEGA, MOGA, NPGA, NSGA-II, SPEA2 and PAES.

MOEA	Evolutionary Approach	Fitness assignment	Diversity-preservation mechanism	Parents' Selection	Elitism	Offline population growth
VEGA	Genetic Algorithm	Objective-based	–	Cooperative subpopulations	–	–
MOGA	Genetic Algorithm	Fonseca & Fleming's Pareto rank rule	Fitness sharing, restricted mating and reinitialization	Proportional selection	Isolated	No limit
NPGA	Genetic Algorithm	Pareto rank rule	Equivalence class sharing	Tournament by Pareto dominance	–	–
NSGA-II	Genetic Algorithm	Goldberg's Pareto rank rule	Crowding	Binary tournament with replacement from offline population	Interactive	Limited by non-dominated sorting
SPEA2	Evolutionary Algorithm	Zitzler & Thiele's improved Pareto rank rule	Truncation	Binary tournament with replacement from offline population	Interactive	Limited by truncation
PAES	Evolutionary Strategy	–	Grid mapping	–	Interactive	Limited by sorted inclusion

PAES

The *Pareto Archived Evolution Strategy* (PAES) was proposed by Knowles & Corne [28], it is an Evolutionary Strategy with Pareto dominance-based fitness assignment and interactive elitism. PAES is conceived under a reproductive scheme $(1+1)$, i.e., a unique parent generates a unique offspring, thus only mutation can be conceptualized. $P^*_{known}(t)$ is used as an historical record of comparison versus every offspring. Once

an offspring is generated by mutation, $P^*_{known}(t)$ update is performed if such offspring is non-dominated by the parent or by any individual in $P^*_{known}(t)$, when $P^*_{known}(t)$ exceeds its maximum size, grid mapping is used to eliminate an individual within the region of higher density. The offspring becomes the parent in the next generation only if it dominates the parent or it is located in a region with lower density according to grid mapping in the current generation.

2.4 QUALITY INDICATORS IN EVOLUTIONARY MULTI-OBJECTIVE OPTIMIZATION

This section summarizes contributions accomplished by Srinivas & Deb [32], Schott [33], Gong *et al* [34], Esbensen & Kuh [35], Fonseca & Fleming [36], Van Veldhuizen [24], Van Veldhuizen & Lamont [37] [13], Zitzler [38], Zitzler & Thiele [39] [40], Zitzler *et al* [12], Hansen & Jaszkiewicz [41], Wu & Azarm [14], Knowles *et al* [18], Deb & Jain [10], Czyzak & Jaszkiewicz [42], Leung & Wang [43], Meng *et al* [44], Lizárraga *et al* [45] and Li & Zheng [46]. Indicators acronyms are modified in some cases to simplify its future categorization. Due to indicators work over $PF^*_{known}(t)$ or PF^*_{known}, in this section the term solution refers to a point in any of this sets.

Spacing Distribution (SD)

Srinivas & Deb [32] developed a measurement schema to know how well distributed each solution is over the non-dominated region. They propose to divide the non-dominated region in q subspaces, expressing distribution horizon from solutions as in Equation (2.8).

$$SD \triangleq \left(\sum_{i=1}^{q+1} \left(\frac{n_i - \bar{n}_i}{\sigma_i} \right)^2 \right)^{1/2} \tag{2.8}$$

where $\bar{n}_i$ is the expected number of non-dominated solutions for the *i*-th subspace, n_i the number of solutions which are non-dominated within the *i*-th subspace, and σ_i the standard deviation of the expected number of non-dominated solutions for the *i*-th subspace defined in Equation (2.9).

$$\sigma_i = \bar{n}_i \left(1 - \frac{\bar{n}_i}{\left| PF^*_{known} \right|} \right) \tag{2.9}$$

The (q+1)–th subspace represents the dominated region, thus $\bar{n}_{q+1} = 0$. Based on a statistic study [32] demonstrated σ_{q+1} to be defined as in Equation (2.10).

$$\sigma_{q+1} = \sum_{i=1}^{q} \sigma_i^2 \tag{2.10}$$

Distribution is ideal if there are $\bar{n}_i$ solutions in each subspace $\leftrightarrow SD = 0$.

Extended Spacing Efficiency (ESE)

Schott [33] proposed a dispersion indicator based on distance variation among nearby solutions from PF^*_{known} called Spacing Efficiency (SE). Afterward, Gong *et al* [34] generalized SE to k–dimensions, amplifying the idea by proposing progressive integration of solutions cluster with minor distance. SE extension is defined in Equations (2.11) and (2.12).

$$ESE \triangleq \left(\frac{1}{\left|PF^*_{known}\right| - 1} \cdot \sum_{i=1}^{\left|PF^*_{known}\right|} \left[\bar{d} - ESE'_i\right]^2 \right)^{1/2} \tag{2.11}$$

$$ESE'_i = \min_{\substack{\boldsymbol{F}(\boldsymbol{x}) \in PF^*_{known,i-1} \\ \boldsymbol{F}(\boldsymbol{y}) \in \{PF^*_{known} - PF^*_{known,i-1}\}}} \sum_{j=1}^{k} \left| f_j(\boldsymbol{x}) - f_j(\boldsymbol{y}) \right| \tag{2.12}$$

where $PF^*_{known,i-1} \subseteq PF^*_{known} : i = 1,2,\ldots,|PF^*_{known}|$, $PF^*_{known,i} = PF^*_{known,i-1} \cup \{\boldsymbol{F}(\boldsymbol{y})\}$, by definition $PF^*_{known,0}$ contains the first solution in PF^*_{known} and $\bar{d}$ is the minimum average distance among every two solutions in PF^*_{known} described in Equation (2.13).

$$\bar{d} = \frac{\sum_{i=1}^{\left|PF^*_{known}\right|} ESE'_i}{\left|PF^*_{known}\right|} \tag{2.13}$$

An equidistance spacing among every PF^*_{known} point occurs $\leftrightarrow ESE = 0$.

Weighted-Sum Aggregation (WSA)

Esbensen & Kuh [35] proposed an indicator based on the creation of linear combinations with certain probability distribution; each linear combination denotes *Decision Maker* (*DM*) possible preferences. For each linear combination, every vector solution from PF^*_{known} is evaluated with the purpose of calculating

the minimum weighted-sum, the PF^*_{known} quality is obtained as the average of the minimum weighted-sums, as in Equation (2.14).

$$WSA \triangleq \frac{\sum_{j=1}^{|LC|}\left[\min_{\boldsymbol{F}(\boldsymbol{x})\in PF^*_{known}} lc_j\left(\boldsymbol{F}(\boldsymbol{x})\right)\right]}{|LC|} \tag{2.14}$$

where $lc_j \in LC$ is the *j*-th linear combination.

Performance Assessment through Attainment Surfaces (PAAS)

Fonseca & Fleming [36] developed a non-parametric statistical procedure to quantitatively measure the relative performance among different multi-criterion optimizers. Each multi-criterion optimizer is executed n times, storing in each iteration the achieved PF^*_{known}. From each PF^*_{known} is possible to divide Λ into two regions:

- objective vectors whose correspondent decision vectors are not dominated by at least an element of PF^*_{known}, and
- objective vectors whose correspondent decision vectors are dominated by at least an element of PF^*_{known}.

Such limit function adjusts achieved objective vectors, named as attainment surface. An attainment surface combines information about solutions convergence and dispersion (see Figure 2.3 a). When attainment surfaces from each multi-criterion optimizer execution are overlapped in the same graphic, it is possible to visualize (see Figure 2.3 b):

- an area from Λ which contains the never achieved vectors in any optimizer execution, lower left section,
- an area from Λ which contains the always achieved and improved vectors in any optimizer execution, upper right section, and
- an area from Λ which contains achieved vectors in some optimizer executions, section enclosed between extreme attainment surfaces. This area can be subdivided in smaller ones based on executions percentage with achieved objective vectors.

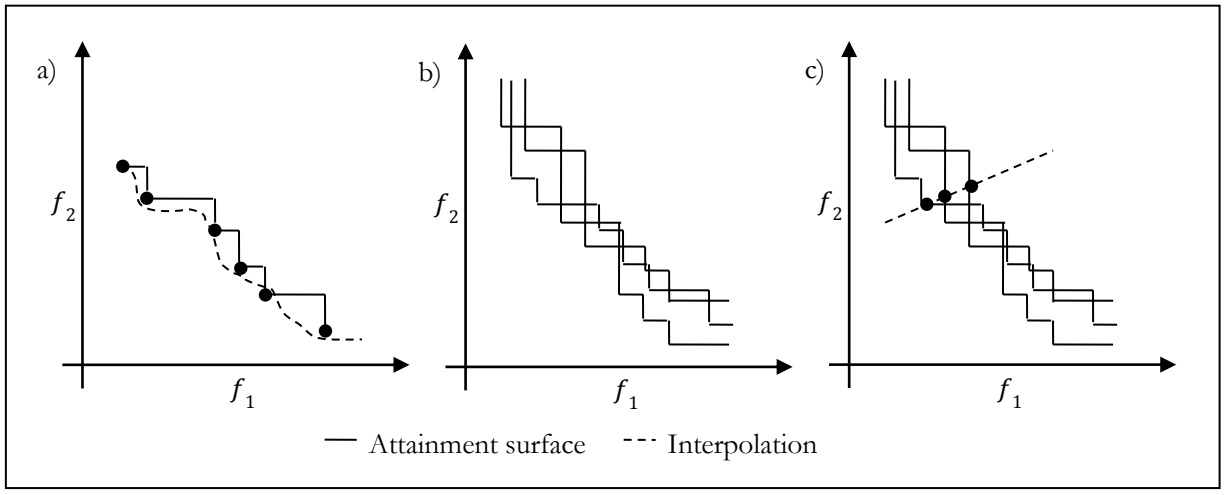

Figure 2.3 Performance Assessment through Attainment Surfaces in bi-objective Λ: a) comparison between interpolation and the attainment surface from solutions set, b) overlapping of three attainment surfaces, c) attainment mid-surface construction from the intersection between an attainment surfaces set and a straight line which denotes solutions order in the sample.

When n executions are considered, it is possible to compute attainment mid-surfaces from using auxiliary straight lines, diagonal to axes and in the same direction as criteria improvement, and the sample of its intersections with the achieved attainment surfaces (see Figure 2.3 c). It is possible to evaluate samples represented by the mid-surface attainment involving convergence and dispersion statistically detailed. Such process gives feasible estimations to support comparison among optimizers in terms of the best performance optimizer; however, its disadvantages are not enough clarity to express how different the performance is and the fact that it was designed to bi-dimensional spaces that is why its visualization in more complex spaces is difficult.

Found Ratio of Pareto Front (FRPF)

Error Ratio (ER) was a first attempt to model this quality indicator. Gong *et al* [34] extended this definition to be able to use any set R^* (where R^* can be: PF^*, PF^*_{known} from another MOEA, or an arbitrary set R defined by the user), measuring the percentage of PF^*_{known} solutions which form the absolute Pareto Front in relation to PF^*_{known} and R^*, see Equation (2.15).

$$FRPF \triangleq \frac{|\{\boldsymbol{F}(\boldsymbol{x}) \in PF^*_{known} : \nexists \boldsymbol{F}(\boldsymbol{y}) \in R^*, \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x})\}|}{|PF^*_{known}|} \quad (\mathbf{2.15})$$

where $\forall \boldsymbol{F}(\boldsymbol{x}) \in PF^*_{known} \exists \boldsymbol{F}(\boldsymbol{y}) \in R^* : \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x}) \leftrightarrow FRPF = 0$ and $\forall \boldsymbol{F}(\boldsymbol{x}) \in PF^*_{known} \nexists \boldsymbol{F}(\boldsymbol{y}) \in R^* : \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x}) \leftrightarrow FRPF = 1$.

Generational Distance (GD_t) and Generational Convergence Velocity (GCV)

Van Veldhuizen & Lamont [13] [24] [37] introduced GD_t as the average distance between PF^*_{known} and R^* in the *t*-th MOEA generation; i.e. the deviation from PF^*_{known} with regard to R^* in a specific evolution time, see Equation (2.16). They also modified the definition of progress measure proposed by Bäck [47] in order to define GCV from GD_t, as in Equation (2.17).

$$GD_t \triangleq \frac{\left(\sum_{\boldsymbol{F}(\boldsymbol{x}) \in PF^*_{known}} \left[\min_{\boldsymbol{F}(\boldsymbol{y}) \in R^*} \|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{y})\|_2^2\right]\right)^{1/2}}{\left|PF^*_{known}(t)\right|} \tag{2.16}$$

$$GCV \triangleq ln\left(\frac{GD_1}{GD_t}\right)^{1/2} \tag{2.17}$$

$PF^*_{known} \subseteq R^* \leftrightarrow GD = 0$.

Overall Non-dominated Vector Generation (ONVG) and Overall Non-dominated Vector Generation Ratio (ONVGR)

Schott [33] proposed to measure the total amount of non-dominated vectors found during MOEA execution, as in Equation (2.18). Later on, this indicator was extended by Van Veldhuizen & Lamont [13] as the percentage of non-dominated solutions, see Equation (2.19).

$$ONVG \triangleq |PF^*_{known}| \tag{2.18}$$

$$ONVGR \triangleq \frac{|PF^*_{known}|}{|R^*|} \tag{2.19}$$

$PF^*_{known} = R^* \leftrightarrow ONVGR = 1$.

Generational Non-dominated Vector Generation (GNVG)

In general, a MOEA adds at each evolutionary step non-dominated solutions found during current generation $PF^*_{current}$ to $PF^*_{known}(t)$. Van Veldhuizen [24] proposed to monitor the amount of non-dominated solutions along evolutionary process as quality measure, Equation (2.20).

$$GNVG \triangleq |PF^*_{current}(t)| \tag{2.20}$$

where $PF^*_{current}(t)$ represents non-dominated solutions found during the *t*-th MOEA's generation.

Hypervolume (H)

Zitzler & Thiele published in [40] a PF^*_{known} error measure in reference to lower boundaries (worst value) from each objective: $f_i^{inf}, i = 1,2,\dots,k$. H is obtained from covering hypervolumes merging formed by each solution in PF^*_{known} as in Equation (2.21).

$$H \triangleq \left\{ \bigcup_i a_i : \boldsymbol{F}(\boldsymbol{x}^i) \in PF^*_{known} \right\} \tag{2.21}$$

where a_i is the covering hypervolume for the objective vector corresponding to *i*-th decision vector (see Figure 2.4 a).

Coverage of Two Sets (CTS)

Zitzler & Thiele proposed in [39] [40] a relative covering comparison between two sets, based in the number of dominated solutions in both sets. Let $\mathrm{X}, \Upsilon \subseteq \Lambda$ be two decision vectors sets to be compared. Mapping ordered pairs (X, Υ) into the interval $[0,1]$ can be mathematically expressed as Equation (2.22).

$$CTS(\mathrm{X}, \Upsilon) \triangleq \frac{|\{\boldsymbol{F}(\boldsymbol{y}) \in \Upsilon : \exists \boldsymbol{F}(\boldsymbol{x}) \in \mathrm{X}, \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x})\}|}{|\Upsilon|} \tag{2.22}$$

In other words, previous equation computes the percentage of non-dominated elements in the second objective vectors set by elements from first set. By definition $CTS = 1$ if every objective vector in X is dominated by Υ. $\exists \boldsymbol{F}(\boldsymbol{x}) \in \mathrm{X} : \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x}) \forall \boldsymbol{F}(\boldsymbol{y}) \in \Upsilon \rightarrow CTS = 1$, otherwise $\nexists \boldsymbol{F}(\boldsymbol{x}) \in \mathrm{X} : \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x}) \forall \boldsymbol{F}(\boldsymbol{y}) \in \Upsilon \rightarrow CTS = 0$. Both $CTS(\mathrm{X}, \Upsilon)$ and $CTS(\Upsilon, \mathrm{X})$ must be considered, since $\mathrm{X} \cap \Upsilon$ is not necessarily the empty set. CTS disadvantage, just like $PAAS$, is its inability to express how different the performance is.

Coverage Difference of Two Sets (CDTS)

Zitzler defined in [19] a new indicator attempting to solve CTS inability to express differences in performance magnitude. Let $\mathrm{X}, \Upsilon \subseteq \Lambda$ be two decision vectors sets to be compared, the size of the space weakly dominated by X but not by Υ is computed in Equation (2.23).

$$CDTS(\mathrm{X}, \Upsilon) \triangleq H(\mathrm{X} + \Upsilon) - H(\Upsilon) \tag{2.23}$$

It is said that $\boldsymbol{F}(\boldsymbol{x})$ strictly dominates $\boldsymbol{F}(\boldsymbol{y})$ (denoted by $\boldsymbol{F}(\boldsymbol{x}) \prec\prec \boldsymbol{F}(\boldsymbol{y})$) if and only if $\boldsymbol{x}$ improves all objectives to optimize with respect to $\boldsymbol{y}$, e.g. assuming only minimization $\boldsymbol{F}(\boldsymbol{x})$ is less than $\boldsymbol{F}(\boldsymbol{y})$, i.e., $\forall f_i \in \boldsymbol{F}\colon f_i(\boldsymbol{x}) < f_i(\boldsymbol{y})$. $\mathrm{X} \prec\prec \Upsilon$ if and only if $\forall \boldsymbol{F}(\boldsymbol{y}) \in \Upsilon \exists \boldsymbol{F}(\boldsymbol{x}) \in \mathrm{X}\colon \boldsymbol{F}(\boldsymbol{x}) \prec\prec \boldsymbol{F}(\boldsymbol{y})$. Thus $\mathrm{X} \prec\prec \Upsilon \leftrightarrow CDTS(\mathrm{X}, \Upsilon) = 0 \wedge CDTS(\Upsilon, \mathrm{X}) > 0$.

Probability of Superiority (PS)

Hansen & Jaszkiewicz [41] proposed to statistically measure how much better a solutions set is over another by computing integration over a utility functions set FU, see Equation (2.24).

$$PS(\mathrm{X}, \Upsilon, FU, p) \triangleq \int_{fu \in FU} C(\mathrm{X}, \Upsilon, g) p(fu)\, dfu \tag{2.24}$$

where the performance comparative function between two solutions sets X and Υ is given by Equations (2.25) and (2.26).

$$C(\mathrm{X}, \Upsilon, fu) = \begin{cases} 1 & \text{if } fu^*(\mathrm{X}) > fu^*(\Upsilon) \\ 0.5 & \text{if } fu^*(\mathrm{X}) = fu^*(\Upsilon) \\ 0 & \text{if } fu^*(\mathrm{X}) < fu^*(\Upsilon) \end{cases} \tag{2.25}$$

$$fu^*(\mathrm{X}) = \max_{\boldsymbol{F}(\boldsymbol{x}) \in \mathrm{X}} fu\big(\boldsymbol{F}(\boldsymbol{x})\big) \tag{2.26}$$

$p(fu)$ is an intensity function which expresses the probability of the utility function fu, $fu \in FU\colon \mathbb{R}^k \rightarrow \mathbb{R}$ is DM's preferences model which relates each point from the objective space with an utility value, it is assumed that DM task is to maximize utility. Therefore, it is possible to assure that PS states the DM's preference rate over a solutions set in reference to another. $PS(\mathrm{X}, \Upsilon, FU, p) = 1 - PS(\Upsilon, \mathrm{X}, FU, p)$. [41] also suggested to use Equation (2.27).

$$PS_R(\mathrm{X}, FU, p) = PS(\mathrm{X}, R, FU, p) \tag{2.27}$$

such that R is an arbitrary reference set common to every set to be evaluated. Nevertheless, measurements produced by PS will greatly depend on the defined R set. In order to eliminate such influence, it is useful to consider more than one R set. PS drawbacks are the number of fu's needed, their respective intensity functions, computational cost and the fact that it only performs comparisons in an homologous way to CTS. Two final generalizations of $C(\mathrm{X}, \mathrm{Y}, fu)$ were presented as alternatives in [41]: 1) more than only two solutions sets and 2) a group of reference sets.

Expected Degree of Superiority (EDS) and Expected Proportion of Superiority (EPS)

In order to solve the PS conflict, Hansen & Jaszkiewicz [41] defined in a homologous way to $CDTS$, two indicators to express magnitude of difference in the superiority level from a solutions set over another. The first one measures the expected degree of superiority, while the second one, anticipates that in certain cases is more significant to measure the percentage of the best profit values instead of the discrepancies. Mathematically, these two quality indicators can be expressed as in Equations (2.28) and (2.29).

$$EDS(\mathrm{X}, \mathrm{Y}, FU, p) \triangleq \int_{fu \in FU} \big(fu^*(\mathrm{X}) - fu^*(\mathrm{Y})\big) p(fu) dfu \tag{2.28}$$

$$EPS(\mathrm{X}, \mathrm{Y}, FU, p) \triangleq \int_{fu \in FU} \frac{fu^*(\mathrm{Y}) - fu^*(\mathrm{X})}{fu^*(\mathrm{Y})} p(fu) dfu \tag{2.29}$$

$EDS(\mathrm{X}, \mathrm{Y}, FU, p) = -EDS(\mathrm{Y}, \mathrm{X}, FU, p)$. In an analogous way to PS_R: $EDS_R(\mathrm{X}, FU, p) = EDS(R, \mathrm{X}, FU, p)$ and $EPS_R(\mathrm{X}, FU, p) = EPS(\mathrm{X}, R, FU, p)$.

Maximum Pareto Front Error (MPFE)

This indicator specifies the line possessing maximum error with regard to PF^*_{known} containing each vector in PF^*, based on the computation of the maximum distance between each objective vector in PF^*_{known} and the closest element from the corresponding PF^* [9] [24]. Initially, it was defined for bi-objective problems and can be mathematically expressed as in Equation (2.30).

$$MPFE \triangleq \max_{\boldsymbol{F}(\boldsymbol{x})\in PF^*} \left[\min_{\boldsymbol{F}(\boldsymbol{y})\in PF^*_{known}} [|f_1(\boldsymbol{x}) - f_1(\boldsymbol{y})|^2 + |f_2(\boldsymbol{x}) - f_2(\boldsymbol{y})|^2] \right]^{1/2} \quad (2.30)$$

However, its extension to *k*–dimensions is hypothetically possible. $PF^*_{known} \subseteq PF^* \leftrightarrow MPFE = 0$.

Non-dominated Vector Addition (NVA)

It defines the amount of non-dominated vectors from cardinal difference between PF^*_{known}s in different generations [24], see Equation (2.31).

$$NVA \triangleq |PF^*_{known}(t)| - |PF^*_{known}(t-1)| \quad (2.31)$$

where $PF^*_{known}(t)$ represents non-dominated solutions found until *t*-th MOEA's generation.

Average Distance to Pareto Front (ADPF)

Zitzler *et al* presented in [39] a function to measure the average distance from PF^*_{known} to PF^* using the Equation (2.32).

$$ADPF \triangleq \frac{\sum_{\boldsymbol{F}(\boldsymbol{x})\in PF^*_{known}} \left[\min_{\boldsymbol{F}(\boldsymbol{y})\in PF^*} \|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{y})\| \right]}{|PF^*_{known}|} \quad (2.32)$$

Distribution of the Found Pareto Front (DFPF)

Zitzler *et al* [39] proposed to measure PF^*_{known} uniformity as combination of its distribution and its cardinality with Equation (2.33).

$$DFPF \triangleq \frac{\sum_{\boldsymbol{F}(\boldsymbol{x})\in PF^*_{known}} |\{\boldsymbol{F}(\boldsymbol{y}) \in PF^*_{known} : \|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{y})\| > \sigma\}|}{|PF^*_{known}| - 1} \quad (2.33)$$

where σ is a neighbourhood parameter. High uniformity exist $\leftrightarrow DFPF = |PF^*_{known}|$ and high swarming $\leftrightarrow DFPF = 0$.

Extent of the Found Pareto Front (EFPF)

The last indicator proposed by Zitzler *et al* [39] takes in count PF^*_{known} spread and is defined in Equation (2.34).

$$EFPF \triangleq \left(\sum_{i=1}^{k} \left[\max_{\boldsymbol{F}(\boldsymbol{x}),\boldsymbol{F}(\boldsymbol{y}) \in PF^*_{known}} \|f_i(\boldsymbol{x}) - f_i(\boldsymbol{y})\| \right] \right)^{1/2} \tag{2.34}$$

Hypervolume Difference (HD)

Wu & Azarm [14] suggested a performance measure based on covering hypervolumes difference between two solutions subsets, they defined Equation (2.35).

$$HD \triangleq space\left(H(R^*) - H(PF^*_{known})\right) \tag{2.35}$$

where $space(\ \ )$ is a scaling function which maps the objective space into $[0,1]$, (see Figure 2.4 b). Knowles *et al* [18] proposed a similar indicator but without scaling.

Overall Pareto Spread (OPS) and kth Objective Pareto Spread (OPS$_k$)

Wu & Azarm [14] defined two indicators to quantify how widely dispersed PF^*_{known} is in Λ when objectives are considered as a whole. The first one, OPS, provides a global sight of PF^*_{known} uniformity, see Equation (2.36); the second one, OPS_k, quantitatively measures solutions range with regard to each objective in an individual way, see Equation (2.37).

$$OPS \triangleq \frac{H_{ex}(PF^*_{known})}{H_{gb}} \tag{2.36}$$

$$OPS_k \triangleq \left| \max_{\boldsymbol{F}(\boldsymbol{x}) \in PF^*_{known}} f_k(\boldsymbol{x}) - \min_{\boldsymbol{F}(\boldsymbol{y}) \in PF^*_{known}} f_k(\boldsymbol{y}) \right| \tag{2.37}$$

where $H_{ex}(PF^*_{known})$ defines the hypervolume formed by extreme points from PF^*_{known} and H_{gb} is the hypervolume delimited by upper and lower values of each objective (see Figure 2.4 c).

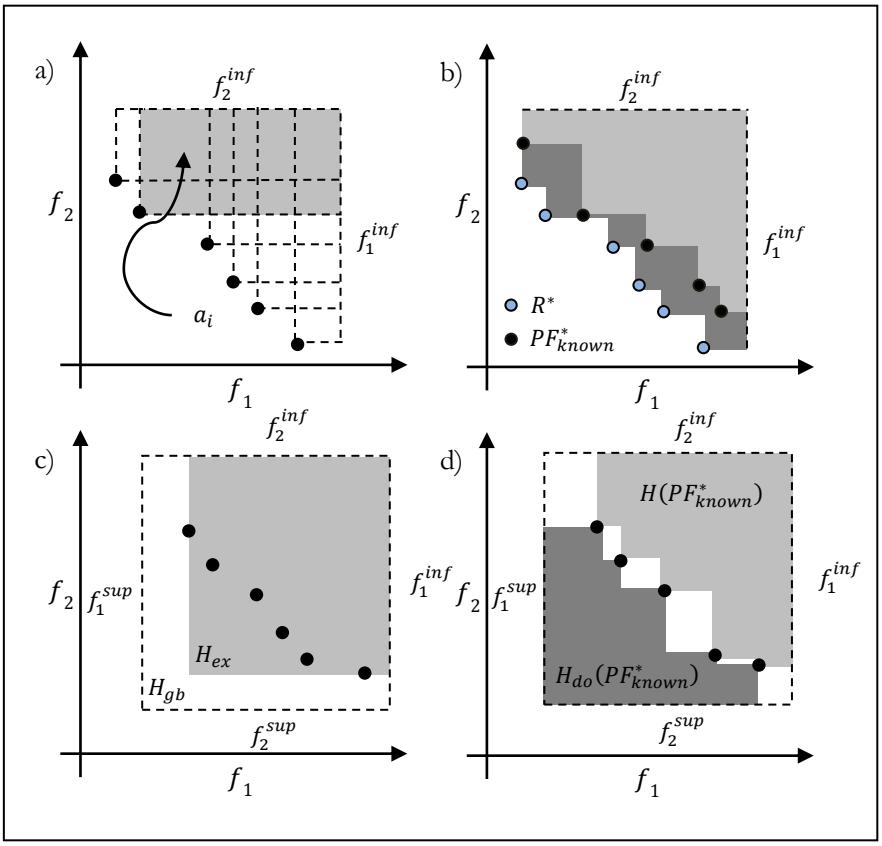

Figure 2.4 Bi-dimensional illustration of a) covering Hypervolumes for objective vectors b) *HD* between two objective vectors sets c) *OPS* and d) *AOPF*.

Accuracy of the Observed Pareto Front (AOPF)

Wu & Azarm [14] introduced $AOPF$ as the inverse of the frontier approximation to PF^*_{known}, see Equation (2.38).

$$AOPF \triangleq \left(1 - space\left(H(PF^*_{known})\right) - space\left(H_{do}(PF^*_{known})\right)\right)^{-1} \quad (2.38)$$

where H_{do} is the superior covering hypervolumes merging of dominated solutions from PF^*_{known} (see Figure 2.4 d). $AOPF$ computes the percentage of the area in white inside H_{gb}, it is easy to see that PF^*_{known} solutions better distributed and closer to superior objectives values will let $AOPF \to \infty$.

Number of Distinct Choices (NDC_μ)

High $|PF^*_{known}|$ does not necessary imply more options to the MOP, some solutions could be too close, i.e. variations between them can be indistinguishable. In a strict way, only strongly distinguishable solutions should be counted. Let $\mu \in [0,1]$ be a defined k–dimensional division value, it is possible to divide the scaled k–dimensional Λ in $(\mu^k)^{-1}$ grids, each grid is a k–dimensional hypercube, $T_\mu(q)$ is specified as the indifferent region, i.e. the space where any two solutions are considered as indistinguishable. Wu & Azarm [14] defined NDC_μ as in Equation (2.39).

$$NDC_\mu \triangleq \sum_{l_k=1}^{1/\mu} \dots \sum_{l_2=1}^{1/\mu} \sum_{l_1=1}^{1/\mu} NT_\mu\left(q_{l_1,l_2,\dots,l_k}, PF^*_{known}\right) \tag{2.39}$$

$$NT_\mu\left(q_{l_1,l_2,\dots,l_k}, PF^*_{known}\right) = \begin{cases} 0 & \text{if } \nexists \boldsymbol{F}(\boldsymbol{x}) \in PF^*_{known} : \boldsymbol{F}(\boldsymbol{x}) \in T_\mu\left(q_{l_1,l_2,\dots,l_k}\right) \\ 1 & \text{otherwise} \end{cases} \tag{2.40}$$

where $q_{l_1,l_2,\dots,l_k}$ is the grid identifier to be explored.

Cluster (CL_μ)

According to Wu & Azarm [14], cluster phenomenon cannot be correctly interpreted by NDC_μ. Then they proposed Equation (2.41), an additional indicator able to achieve such phenomenon:

$$CL_\mu \triangleq \frac{|PF^*_{known}|}{NDC_\mu(PF^*_{known})} \tag{2.41}$$

A PF^*_{known} posseses good uniformity if every solution is found in different $T_\mu(q) \leftrightarrow CL_\mu = 1$. The PF^*_{known} clustering index is greater than 1 as CL_μ is.

Running Metric for Convergence (RMC) and Running Metric for Diversity (RMD)

Deb & Jain [10] proposed the use of two indicators during MOEA evolutionary process, arguing that information about how a MOEA achieves final population has not been commonly analysed. The first execution indicator is RMC, which is defined in Equation (2.42) and computes a ratio from minimum normalized Euclidian distance between each element in the $PF^*_{current}(t)$ and an arbitrary reference decision vectors set R to be compared on $PF^*_{current}(t)$ cardinality. The second indicator, RMD defined in

Equation (2.43), evaluates $PF^*_{current}(t)$ uniformity, i.e., each solution vector is projected into a hyperplane from the objective hyperspace, one dimension is reduced during this process. The hyperplane is divided into a mesh of h grids, each grid is a (k–1)–dimensional hyperarea.

$$RMC_t \triangleq \frac{\sum_{\boldsymbol{F}(\boldsymbol{x})\in PF^*_{current}} \left[\min_{\boldsymbol{F}(\boldsymbol{y})\in R}\left[\sum_{i=1}^{k}\left|\frac{f_i(\boldsymbol{x})-f_i(\boldsymbol{y})}{f_i^{sup}-f_i^{inf}}\right|\right]\right]^{1/2}}{|PF^*_{current}(t)|} \tag{2.42}$$

normalized with $RMC_t \triangleq \frac{RMC_t}{RMC_0}$

$$RMD_t \triangleq \frac{\sum_{l_{k-1}=1}^{h}\cdots\sum_{l_2=1}^{h}\sum_{l_1=1}^{h} neighborhood\left(h\left(q_{l_1,l_2,\ldots,l_{k-1}},t\right)\right)}{\sum_{l_{k-1}=1}^{h}\cdots\sum_{l_2=1}^{h}\sum_{l_1=1}^{h} neighborhood\left(H\left(q_{l_1,l_2,\ldots,l_{k-1}}\right)\right)} \tag{2.43}$$

where $q_{l_1,l_2,\ldots,l_{k-1}}$ is the grid identifier to be explored and

$$h\left(q_{l_1,l_2,\ldots,l_{k-1}},t\right)=\begin{cases}0 & \text{if } H\left(q_{l_1,l_2,\ldots,l_{k-1}}\right)=0 \vee \\ & \nexists \boldsymbol{F}(\boldsymbol{x})\in PF^*_{current,PF^*_{known}}(t): \boldsymbol{F}(\boldsymbol{x})\in T_\mu\left(q_{l_1,l_2,\ldots,l_{k-1}}\right)\\ 1 & \text{otherwise}\end{cases} \tag{2.44}$$

$$H\left(q_{l_1,l_2,\ldots,l_{k-1}}\right)=\begin{cases}0 & \text{if } \nexists \boldsymbol{F}(\boldsymbol{x})\in PF^*_{known}: \boldsymbol{F}(\boldsymbol{x})\in T_\mu\left(q_{l_1,l_2,\ldots,l_{k-1}}\right)\\ 1 & \text{otherwise}\end{cases} \tag{2.45}$$

$PF^*_{current,PF^*_{known}}(t)$ represents non-dominated solutions from $PF^*_{current}(t)$ by PF^*_{known}, $neighborhood\left(g\left(q_{l_1,l_2,\ldots,l_{k-1}}\right)\right)$ provides an evaluation by analysing $g\left(q_{l_1,l_2,\ldots,l_{k-1}}\right)$ values in the $q_{l_1,l_2,\ldots,l_{k-1}}$ neighbourhood.

Average Distance (AD) and Worst Distance (WD)

Czyzak & Jaszkiewicz [42] proposed indicators to measure convergence error based on distance between solutions from two decision vectors sets. Let $\mathrm{X}, \mathrm{Y} \subseteq \Lambda$ be two decision vectors sets to be compared. The first proposal measures the average distance from solutions in set Y with regard to the closest solution in set X, while the second one evaluates the worst distance instead of the average, see Equations (2.46) and (2.47) respectively.

$$AD \triangleq \frac{\sum_{\boldsymbol{F}(\boldsymbol{y})\in \Upsilon} \min_{\boldsymbol{F}(\boldsymbol{x})\in \mathrm{X}} \left[\max_{i=1,2,\dots,k} [\rho_i | f_i(\boldsymbol{y}) - f_i(\boldsymbol{x})|] \right]}{|V|} \tag{2.46}$$

$$WD \triangleq \max_{\boldsymbol{F}(\boldsymbol{y})\in \Upsilon} \left[\min_{\boldsymbol{F}(\boldsymbol{x})\in \mathrm{X}} \left[\max_{i=1,2,\dots,k} [\rho_i | f_i(\boldsymbol{y}) - f_i(\boldsymbol{x})|] \right] \right] \tag{2.47}$$

where a weighted vector $\mathrm{P} = [\rho_1, \rho_2, \dots, \rho_k]$ is defined in advance $\rho_i = \left| \min_{\boldsymbol{F}(\boldsymbol{y})\in \Upsilon} f_i(\boldsymbol{y}) - \max_{\boldsymbol{F}(\boldsymbol{y})\in \Upsilon} f_i(\boldsymbol{y}) \right|^{-1}$.

Indicator $\boldsymbol{\varepsilon}$ ($\boldsymbol{I_\varepsilon}$)

Knowles *et al* [18] and Zitzler *et al* [12] suggested a quality indicator ε which compares two solutions sets $\mathrm{X}, \Upsilon \subseteq \Lambda$. Two versions of such indicator exists: multiplicative I_ε and additive $I_{\varepsilon+}$. The ε indicator computes the minimum value ε by which each $\boldsymbol{F}(\boldsymbol{y}) \in \Upsilon$ is modifed in order to make it worse than any $\boldsymbol{F}(\boldsymbol{x}) \in \mathrm{X}$. The multiplicative version is defined in Equation (2.48).

$$I_\varepsilon \triangleq \inf_{\varepsilon\in\mathbb{R}} \{\exists \boldsymbol{F}(\boldsymbol{x}) \in \mathrm{X}: \boldsymbol{F}(\boldsymbol{x}) \preccurlyeq_\varepsilon \boldsymbol{F}(\boldsymbol{y}) \forall \boldsymbol{F}(\boldsymbol{y}) \in \Upsilon\} = \max_{\boldsymbol{F}(\boldsymbol{y})\in \Upsilon} \left[\min_{\boldsymbol{F}(\boldsymbol{x})\in \mathrm{X}} \left[\max_{i=1,2,\dots,k} \frac{f_i(\boldsymbol{x})}{f_i(\boldsymbol{y})} \right] \right] \tag{2.48}$$

where $\boldsymbol{F}(\boldsymbol{x}) \preccurlyeq_\varepsilon \boldsymbol{F}(\boldsymbol{y}) \leftrightarrow f_i(\boldsymbol{x}) \leq \varepsilon f_i(\boldsymbol{y}) \forall i = 1,2,\dots,k$. It is said that $\boldsymbol{F}(\boldsymbol{x})$ weakly dominates $\boldsymbol{F}(\boldsymbol{y})$ (denoted by $\boldsymbol{F}(\boldsymbol{x}) \preccurlyeq \boldsymbol{F}(\boldsymbol{y})$) if and only if $\boldsymbol{F}(\boldsymbol{x}) \prec \boldsymbol{F}(\boldsymbol{y})$ or $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{F}(\boldsymbol{y})$, e.g. assuming only minimization $\forall f_i \in \boldsymbol{F}: f_i(\boldsymbol{x}) \leq f_i(\boldsymbol{y})$. X is better than Υ (denoted by $\mathrm{X} \lhd \Upsilon$) if and only if $\forall \boldsymbol{F}(\boldsymbol{y}) \in \Upsilon \exists \boldsymbol{F}(\boldsymbol{x}) \in \mathrm{X}: \boldsymbol{F}(\boldsymbol{x}) \preccurlyeq \boldsymbol{F}(\boldsymbol{y}) \wedge \mathrm{X} \neq \Upsilon$. Thus $\mathrm{X} \lhd \Upsilon \leftrightarrow I_\varepsilon > 1$, $\Upsilon \subseteq \mathrm{X} \leftrightarrow I_\varepsilon = 1$ and $\mathrm{X} \rhd \Upsilon \leftrightarrow I_\varepsilon < 1$. The additive version is defined in Equation (2.49).

$$\begin{aligned} I_{\varepsilon+} &\triangleq \inf_{\varepsilon\in\mathbb{R}} \{\exists \boldsymbol{F}(\boldsymbol{x}) \in \mathrm{X}: \boldsymbol{F}(\boldsymbol{x}) \preccurlyeq_{\varepsilon+} \boldsymbol{F}(\boldsymbol{y}) \forall \boldsymbol{F}(\boldsymbol{y}) \in \Upsilon\} \\ &= \max_{\boldsymbol{F}(\boldsymbol{y})\in \Upsilon} \left[\min_{\boldsymbol{F}(\boldsymbol{x})\in \mathrm{X}} \left[\max_{i=1,2,\dots,k} [f_i(\boldsymbol{x}) - f_i(\boldsymbol{y})] \right] \right] \end{aligned} \tag{2.49}$$

where $\boldsymbol{F}(\boldsymbol{x}) \preccurlyeq_{\varepsilon+} \boldsymbol{F}(\boldsymbol{y}) \leftrightarrow f_i(\boldsymbol{x}) \leq \varepsilon + f_i(\boldsymbol{y}) \forall i = 1,2,\dots,k$, $\mathrm{X} \lhd \Upsilon \leftrightarrow I_{\varepsilon+} > 0$, $\Upsilon \subseteq \mathrm{X} \leftrightarrow I_{\varepsilon+} = 0$ and $\mathrm{X} \rhd \Upsilon \leftrightarrow I_{\varepsilon+} < 0$.

U-measure

Leung & Wang [43] proposed an indicator to measure uniformity by computing the discrepancy among the distance between neighbours of each solution in every f_i. First, domains in the Pareto frontier (f_i^{inf} and f_i^{sup}) are determined. Next, to every f_i abstraction of $\boldsymbol{F}$ two extreme solutions (computed by f_i^{inf} and f_i^{sup}) are added. Then for every f_i abstraction and every solution two nearest neighbours are found (one in every f_i direction), for every extreme solution the nearest neighbour is found. Then for every solution and extreme solution at every f_i abstraction, the distance d_{ij} between its neighbours is computed in $\boldsymbol{F}$ and valuated using Equation (2.50).

$$U - measure \triangleq \frac{1}{D}\sum_{i=1}^{k}\sum_{j=|PF_{known}^{*}|+2}\left|\frac{d_{ij}}{d_{ideal}} - 1\right| \tag{2.50}$$

where $D = k(|PF_{known}^{*}| + 2),\ j > |PF_{known}^{*}|$ are the extreme solutions at f_i whose d_{ij} are computed by the distance between them and their closest neighbour and modified by adding d_{ij} non-extreme solutions average, $d_{ideal} = \sum_{i=1}^{k}\sum_{j=|PF_{known}^{*}|+2} d_{ij}/D$.

Uniformity of the Pareto Optimal Set (UPOS)

Meng *et al* [44] suggested a quality indicator to measure uniformity in order to improve Schott [33] works. $UPOS$ is defined in Equation (2.51).

$$UPOS \triangleq \left(\frac{1}{\left|PF_{known}^{*}\right| - 1}\cdot \sum_{\boldsymbol{F}(\boldsymbol{x}^i)\in PF_{known}^{*}} \left[1 - \mathrm{inv}\left(d_{\boldsymbol{x}^i}, \bar{d}\right)\right]^2\right)^{1/2} \tag{2.51}$$

$$d_{\boldsymbol{x}^i} = \min_{\boldsymbol{F}(\boldsymbol{y})\in\{PF_{known}^{*} - \boldsymbol{F}(\boldsymbol{x}^i)\}}\left(\sum_{j=1}^{k}\left[f_j(\boldsymbol{x}^i) - f_j(\boldsymbol{y})\right]^2\right)^{1/2} \tag{2.52}$$

$$\mathrm{inv}\left(d_{\boldsymbol{x}^i}, \bar{d}\right) = \begin{cases} d_{\boldsymbol{x}^i}/\bar{d} & \text{if } d_{\boldsymbol{x}^i} > \bar{d} \\ \bar{d}/d_{\boldsymbol{x}^i} & \text{otherwise} \end{cases} \tag{2.53}$$

where $\bar{d}$ is the minimum average distance among every $\boldsymbol{F}(\boldsymbol{x}) \in PF_{known}^{*}$. If $UPOS$ gets an equal value for two PF_{known}^{*}s, then each PF_{known}^{*} is reduced by the union of the two closest solutions and $UPOS$ is recalculated.

Well-Extended of the Pareto Optimal Set (W-EPOS)

Meng *et al* [44] also attempted to measure PF^*_{known} well-extent by computing for every solution in a R^* (which may be the PF^* or the non-dominated solutions achieved by several runs) the average distance to the closest solution in PF^*_{known}, it is mathematically expressed in Equation (2.54).

$$W - EPOS \triangleq \frac{\sum_{\boldsymbol{F}(\boldsymbol{y})\in R^*} \min_{\boldsymbol{F}(\boldsymbol{x})\in PF^*_{known}} \|\boldsymbol{F}(\boldsymbol{x}) - \boldsymbol{F}(\boldsymbol{y})\|}{|R^*|} \tag{2.54}$$

G-metric

Lizárraga *et al* [45] proposed $G - metric$, a *n*-ary indicator that ranks PF^*_{known}s based on their convergence and dispersion. Initially every vector in every PF^*_{known} must be normalized using the maximum and minimum value of the non-dominated solutions product of the union: $\bigcup_{i=1}^{n} \mathrm{X}_i$. Then two components are computed for every X_i and combined to create a number that represents its relative performance respect to the others $\mathrm{X}_1, \dots, \mathrm{X}_{i-1}, \mathrm{X}_{i+1}, \dots \mathrm{X}_n$. The first component is calculated by means of the outperformance relation O_C [41], classifying every X_i according to the partial order that O_C gives by levels, e.g. the first level includes those X_i so that $\nexists \mathrm{X}_j, j = 1,2, \dots, i-1, i+1, \dots, n: \mathrm{X}_j O_C \mathrm{X}_i$. The second component is calculated from the zone of influence of every $\boldsymbol{F}(\boldsymbol{x}) \in \mathrm{X}_i$, computing regions of integration for every solution with a radius υ, avoiding intersecting zones.

Spread Assessment (SA)

Li & Zheng [46] suggested SA to quantify how widely PF^*_{known} spreads over Λ. The idea is to compute the hypervolume product (with $f_i^{inf} = 0$) of every $BS_i: i = 1,2, \dots, k$, set of boundary solutions in the *i*-th PF^*_{known} projection $\in \mathbb{R}^{k-1}$, where such projection is on $\{f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k\}$. The average value of f_i is computed for every projection: $w_i = \sum_{\boldsymbol{F}(\boldsymbol{x})\in PF^*_{known}} f_i(\boldsymbol{x}) / |PF^*_{known}|$. The total assessment of SA is computed using Equation (2.55).

$$SA \triangleq \left(\prod_{i=1}^{k} \frac{H(BS_i)}{\prod_{j=1, j\neq i}^{k} |w_i|} \right)^{1/k} \tag{2.55}$$

Chapter 3 RankMOEA

[*Small minds are concerned with the extraordinary, great minds with the ordinary…*]

Blaise Pascal

3.1 INTRODUCTION

Diversity-preservation mechanisms impulse divergence in tangential direction to the promising regions discovered by the MOEA, this through probability selection bias in less conglomerated regions. Most of the designed diversity mechanism in EMO require the specification of parameters, or are unable to deal with incommensurable objectives, or were not designed to be compliant with the search space. Hence, three premises are considered in order to design a new MOEA (called RankMOEA) which emphasizes spread and dispersion of PF^*_{known} and preserves equilibrium between exploitation and exploration:

- Since Pareto dominance rules sort candidate solutions in a certain order according to their proximity to the frontier of Λ, some advantage can be taken from such arrangement by intensifying exploration in candidate solutions far from the frontier and reducing exploration in candidate solutions close to the frontier. This assuming that the first type of solutions does not have much information about PF^* structure and need more effort to achieve a good performance.
- The structure of the search is defined by Ω and not by Λ, thus, diversity preservation mechanisms should work better in Ω if they are compliant with Ω structure. Hence, exploitation of the information could be successful by mating nearby candidate solutions in Ω since such process is less disruptive.
- In most of the cases, after a certain threshold in the evolutionary process of MOEAs, the number of non-dominated solutions grows rapidly, thus reduced mutation in solutions closer to the frontier of Λ that are less conglomerated in Ω should improve performance by controlling exploration and preserving the emphasized exploitation in such regions.

The first and third premises are related with exploration, recent works [48] [49] have shown the advantage of using exploration applied with probabilities that depend on the fitness rank of a genotype or phenotype in single objective problems, such approach has shown to be a robust alternative since it avoids some questions of mutation rates tuning without having to introduce an explicit encoded self-adaptation mechanism. Thus, the ideas behind RankMOEA are motivated by appealing to previous theoretical analysis [48] that show how different landscapes and population states require different mutation rates to dynamically optimize the balance between exploration and exploitation.

3.2 RANKMOEA DESCRIPTION

RankMOEA extends the rank mutation presented in [48] [49] to the MO framework, overcoming the mutation fine tuning drawbacks and promoting a controlled diversity according to Pareto dominance and the degree of conglomerate. RankMOEA is described in Figure 3.1.

RankMOEA **( )**

1 $t \leftarrow 1$
2 **random initialization** of each individual $\boldsymbol{\chi} \in P(t)$
3 $\forall \boldsymbol{\chi} \in P(t)$ **evaluate** $f_i(\Phi(\boldsymbol{\chi}))\forall f_i \in \boldsymbol{F}$
4 $ranking_{\boldsymbol{\chi}\in P(t)} \leftarrow rank_g(\boldsymbol{\chi}, P(t)), \forall \boldsymbol{\chi} \in P(t)$
5 $P^*_{known}(t) \leftarrow nondominated(P(t))$
6 **do while** $t <$ stopping criterion
7 $P'(t) \leftarrow selection(P(t))$
8 $P''(t) \leftarrow mst\_niching(P'(t), P(t))$
9 $P'''(t) \leftarrow rank\_mutation(P''(t))$
10 $\forall \boldsymbol{\varsigma} \in P'''(t)$ **evaluate** $f_i(\Phi(\boldsymbol{\varsigma}))\forall f_i \in \boldsymbol{F}$
11 $ranking_{\boldsymbol{\zeta}\in P'''(t)} \leftarrow rank_g(\boldsymbol{\varsigma}, P'''(t)), \forall \boldsymbol{\varsigma} \in P'''(t)$
12 $P^*_{known}(t+1) \leftarrow nondominated(\{P^*_{known}(t) \cup P'''(t)\})$
13 **if** $|P^*_{known}(t+1)| > P_{size}$
14 $P^*_{known}(t+1) \leftarrow truncation(P^*_{known}(t+1), P_{size})$
15 $P(t+1) \leftarrow P^*_{known}(t+1)$
16 **else**
17 **sort** $\{P'''(t) - P^*_{known}(t+1)\}$ **by** $ranking_{\boldsymbol{\varsigma}\in\{P'''(t)-P^*_{known}(t+1)\}}$ **ascendant**
18 $P(t+1) \leftarrow \{P^*_{known}(t+1) \cup \{P'''(t) - P^*_{known}(t+1)\}[1: P_{size} - |P^*_{known}(t+1)|]\}$
19 **end if**
20 $t \leftarrow t+1$
21 **end do**

Figure 3.1 RankMOEA.

First, a set of P_{size} individuals are initialized as the early population $\boldsymbol{\chi} \in P(1)$ and evaluated in the set of k objectives to be optimized (lines 1 to 3). Then, Goldberg's ranking is used to sort non-dominated and dominated individuals (line 4):

$$rank_g(\chi, P(t)) = \begin{cases} 1 & \text{iff } \nexists \zeta \in P(t): \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x}) \\ \max\limits_{\zeta \in P(t): \boldsymbol{F}(\boldsymbol{y}) \prec \boldsymbol{F}(\boldsymbol{x})} [rank_g(\zeta, P(t))] + 1 & \text{otherwise} \end{cases} \quad (3.1)$$

By definition, non-dominated individuals in $P(t)$ have ranking value of 1, thus individuals closer to such non-dominated individuals in Λ have lower ranking values. Goldberg's ranking was preferred since it allows smoother ranking landscapes of Pareto domination. RankMOEA uses an interactive online file $PF^*_{known}(t)$ to store continuously its approximation to PF^* (line 5). During the evolution process only $P_{size}/2$ of the parents $P(t)$ are chosen (line 7) by the selection procedure.

Mates of the $P_{size}/2$ parents are chosen using a minimum spanning tree niching which works over the phenotypic space (line 8). This mechanism builds a minimum spanning tree in Ω including all individuals in $P(t)$, distance in Ω is computed normalizing every phenotypic feature which allows to handle incommensurable variables. In this approach, niches are not isolated elements, moreover they are elements partially coupled by the tree structure (see left side of Figure 3.2). Since each individual in the minimum spanning tree can be connected with more than one individual, every $\chi \in P(t)$ is weighted with Equation (3.2).

$$mst(\chi) = \frac{1}{rank_g(\chi, t) + \left(1 - \frac{1}{mst\_arity(\Phi(\chi))}\right)} \quad (3.2)$$

where $mst\_arity(\Phi(\chi))$ counts the number of decision vectors (produced by mapped individuals) connected to $\Phi(\chi)$ in the minimum spanning tree. So individuals with lower Goldberg ranking value and lower arity (conglomerate measure) in Ω will accomplish a higher value of $mst(\chi)$, a hierarchical preference of ranking over arity is denoted.

In order to select the mates of the $P_{size}/2$ parents, a stochastic selection process (e.g. stochastic universal selection) can be used with $mst(\chi)$ as the desirability of selection, including all the neighbours of the parent in the minimum spanning tree, then parents will be mated with less conglomerated individuals whose projection in Λ is closer to PF^*_{known}. It is important to observe that there is no need to define a proximity value. This procedure can be performed using Chazelle's algorithm [50] which is based on the soft heap, the most asymptotically efficient known structure to find the minimum spanning tree. Its running time is $O(\lambda\, \alpha(\lambda, m))$, where λ is the number of edges and α is the classical functional inverse of

the Ackermann function. The function α grows extremely slowly, so that for all practical purposes it may be considered a constant no greater than 4; thus Chazelle's algorithm takes very close to linear time.

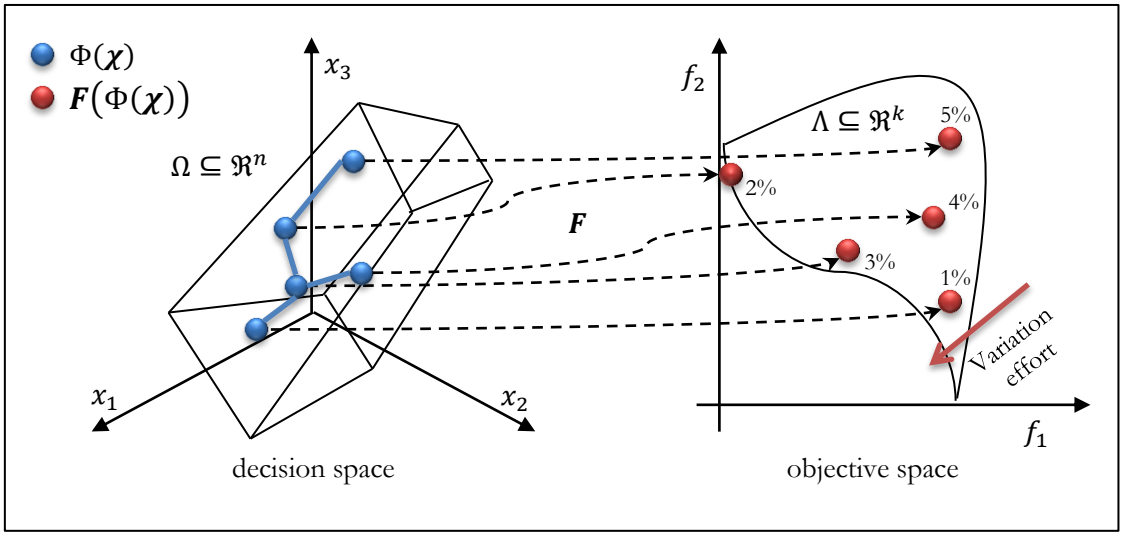

Figure 3.2 Minimum spanning tree niching and ranking mutation.

The proposed rank mutation considers pre-order, the intrinsic inconvenient of MOPs. Rank mutation (line 9) consists of the definition of a mutation rate range and the assignment of a uniformly distributed mutation rate to individuals according to their inherited $mst(\chi)$ value, i.e. individuals with lower Goldberg ranking value and lower arity will get a lower mutation rate and individuals with higher Goldberg ranking value and higher arity will get a higher mutation rate, denoting tight exploration in the neighbourhood of individuals closer to PF^* and widespread exploration in the neighbourhood of individuals farther from PF^* (see right side of Figure 3.2).

When the entire population is non-dominated, tight exploration is performed in the neighbourhood of individuals with lower arity and widespread exploration in the neighbourhood of individuals with higher arity. The mutation rate range will be specified by a minimum and maximum mutation rates, p_{min} and p_{max} respectively, and divided into P_{size} steps to generate the deterministic rule of choosing the mutation rate. So the mutation rate of the *i*-th individual in $P''(t)$ sorted in descendant order according to $mst(\chi)$ is $p_{min} + i \cdot (p_{max} - p_{min})/(P_{size} - 1)$. According to [48] a natural range to cover any eventuality is $p_{min} = 0$ and $p_{max} = 1 - 1/l$, where l is individual length (when working with binary representation), however if there is knowledge of the vicinity of the optimum, and the population is in the vicinity, then a

lower p_{max} may be appropriate. Mutation range remains fixed during entire evolution. Since mutation only requires to sort individuals in order to assign the mutation rate, this step has a complexity of $O(P_{size} \log P_{size})$.

Thereafter, RankMOEA evaluates the offspring in the set of k objectives and ranks them with Goldberg's ranking (lines 10 and 11). $P^*_{known}(t)$ is updated with new non-dominated individuals (line 12). Finally, if $P^*_{known}(t)$ size is larger than P_{size}, a truncation process proposed in [23] is used to reduce its size (lines 14 and 15) and the $P(t+1)$ is constituted by such reduction; else $P(t+1)$ is constituted by $P^*_{known}(t)$ and a controlled insertion using ranking based selection of the best offspring that were not already included in $P^*_{known}(t)$ (lines 17 and 18).

The speed performance of RankMOEA is ruled by the mutation process, therefore the computational complexity of RankMOEA can be calculated as $O(P_{size} \log P_{size})$, which makes it a fast algorithm, worthy to compete with other state of the art MOEAs.

3.3 TESTING RANKMOEA

In the following tests some well-known MOEAs (VEGA, MOGA, NPGA, NSGA-II, SPEA2 and PAES) were used to compare the performance of RankMOEA. The seven MOEAs used binary-coded chromosomes, one point crossover and bit-wise mutation. VEGA, MOGA and RankMOEA were tested using the stochastic universal sampling, PAES its natural reproductive scheme, while NPGA, NSGA-II and SPEA2 were tested using their tournament selection operator. MOGA's restricted mating, NPGA's equivalence class sharing, NSGA-II's crowding, SPEA2's k-nearest neighbour, PAES's grid mapping and RankMOEA's minimum spanning tree niching were implemented in the phenotypic space.

The mating rates used for the seven MOEAs were: 70%, 80% and 90%. The mutation rates used for VEGA, MOGA, NPGA, NSGA-II and SPEA2 were 1%, 2%, 3%, 4%, 5% and 6%, whereas for RankMOEA p_{min} was set to 0% and p_{max} to 6%. A precision of 0.001 was set for each variable in the phenotype. The seven algorithms were run 30 times with each mating-mutation configuration, the average behavior of each configuration was assessed using a version of $G - metric$ [45] to work in Ω, the n-ary quality indicator that ranks PF^*_{known}s based on the their attained dispersion and convergence.

3.3.A Spread-hardness Test

The first test was designed to examine the robustness of the diversity-preservation mechanisms of the seven MOEAs by finding a good diversity of the solutions in Ω. A function with three reference points $\vec{z}_i: i = 1,2,3$ in a bidimensional Ω was defined; the idea is to minimize the distance to such reference points, i.e., $\min F(\vec{x}) = [f_1(\vec{x}), f_2(\vec{x}), f_3(\vec{x})]$ with $f_i(\vec{x}) = \|\vec{z}_i - \vec{x}\|_2$, where $\vec{z}_1 = (2,1)$, $\vec{z}_2 = (3,5)$, $\vec{z}_3 = (4,1)$ and subject to $x_j \in [0,6]: j = 1,2$. It is clear that PS is formed by all the points located within the triangle constituted by the three reference points. It is expected that a diversity-preservation mechanism with good performance and compliant with Ω structure will achieve a quasi-uniform spread. A population and $PF^*_{known}(t)$ size of 50 individuals with 20,000 objective function evaluations were considered. Figure 3.3 shows the best approximation to PS achieved by the best run of the best mating-mutation configuration of each MOEA according to the average of $G - metric$ over the 30 runs.

The best mating-mutation configuration of MOGA, NSGA-II, SPEA2, PAES and RankMOEA found enough non-dominated solutions to complete $PF^*_{known}(t)$ until its boundary size, while best mating-mutation configuration of VEGA and NPGA only found 20.9 and 41.3 non-dominated solutions in average respectively. By sorting MOEAs' performance according to the average of $G - metric$ over the 30 runs of their best mating-mutation configuration, the following order is accomplished: RankMOEA, SPEA2, NSGA-II, PAES, MOGA, NPGA and VEGA; where clearly VEGA achieves the worst distribution and RankMOEA the best distribution.

3.3.B Complicated Pareto Set Test

The second test was performed using the UF4 problem from the CEC'09 contest [51], a MOP with complicated PS which demonstrated to be a very hard problem even for the best algorithms that participated in MO contest in CEC'09. Figure 3.4 shows the best PF^*_{known} achieved by the best run of the best mating-mutation configuration of each MOEA according to the average of $G - metric$ over the 30 runs. A population and $PF^*_{known}(t)$ size of 100 individuals with 300,000 objective function evaluations were considered. The best mating-mutation configuration of NSGA-II, SPEA2 and RankMOEA found enough non-dominated solutions to complete $PF^*_{known}(t)$ until its boundary size, while best mating-

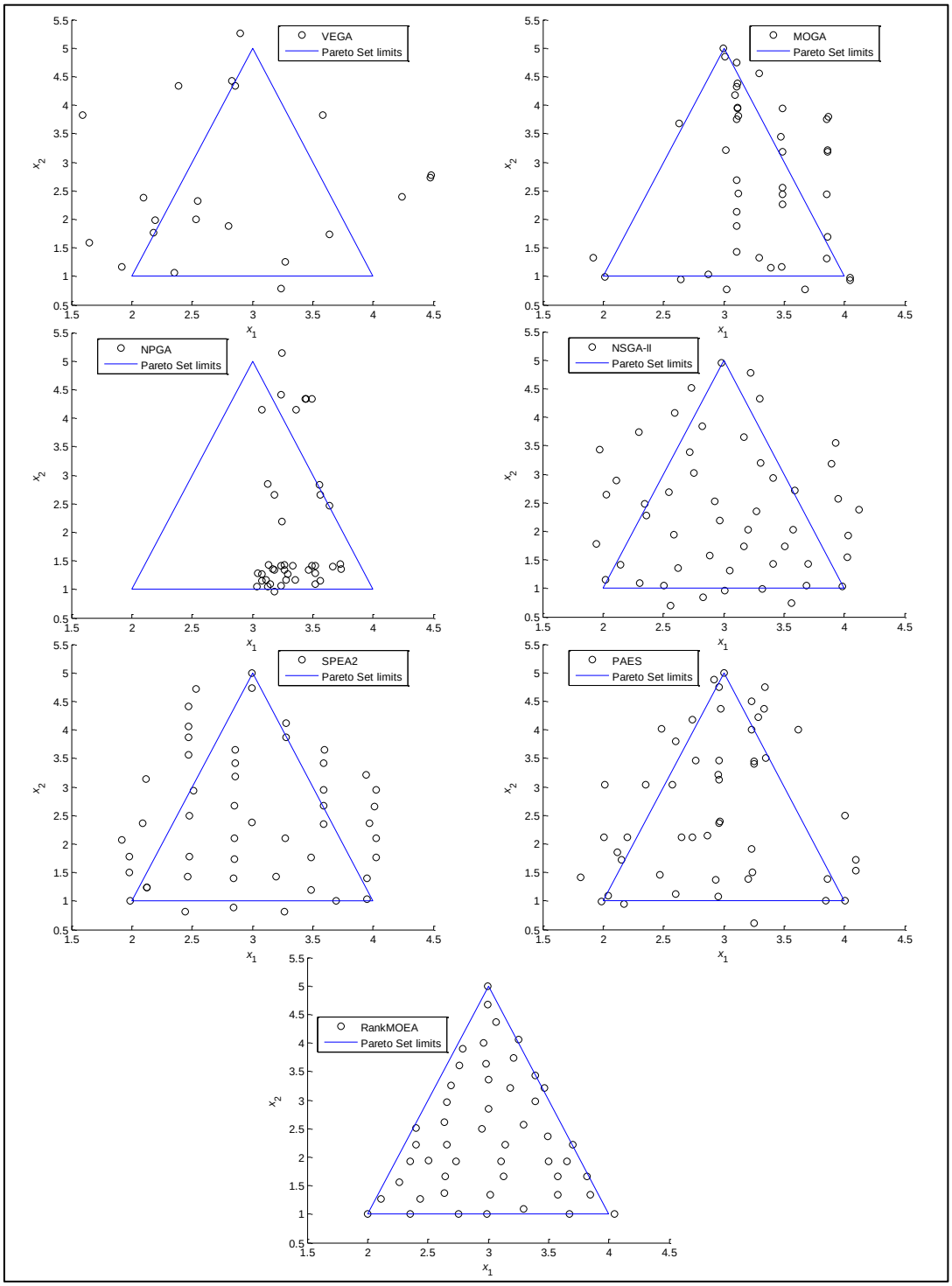

Figure 3.3 VEGA, MOGA, NPGA, NSGA-II, SPEA2, PAES and RankMOEA best approximation to Spread-hardness test.

mutation configuration of VEGA, MOGA, NPGA and PAES only found 28.1, 98.8, 30.4 and 92.9 non-dominated solutions in average respectively. By sorting MOEAs' performance according to the average of $G - metric$ over the 30 runs of their best mating-mutation configuration, the following order is accomplished: RankMOEA, NSGA-II, SPEA2, MOGA, PAES, NPGA and VEGA. RankMOEA achieves the best spread and the lowest convergence error, followed by NSGA-II with worse spread and by SPEA2 with worse convergence error; on other hand MOGA, PAES, NPGA and VEGA show poor performance. In Chapter 5 several additional experiments with different MOPs are shown.

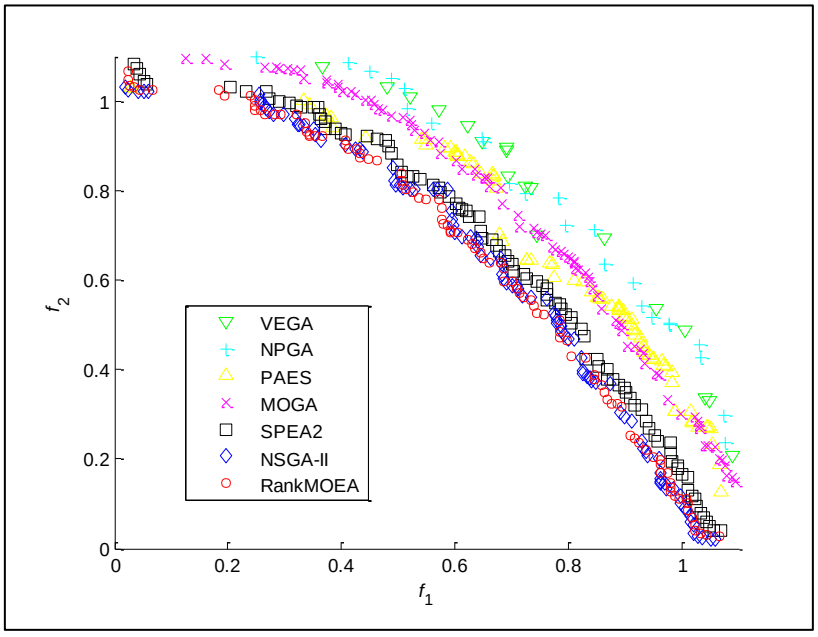

Figure 3.4 VEGA, MOGA, NPGA, NSGA-II, SPEA2, PAES and RankMOEA best approximation to Complicated Pareto Set test.

Chapter 4 Inferential Power of Quality Indicators in Evolutionary Multi-Objective Optimization

[*Common sense is not so common…*]

Voltaire

4.1 INTRODUCTION

Several quality indicators classifications have been proposed in EMO field attempting to described dissimilarities among them and discriminate which ones should be used in practice. A very early quality indicators classification divides them in: *unary indicators*, which take a single PF^*_{known} as argument and assign a real number that reflects a quality aspect; *binary indicators*, which take two PF^*_{known}s as arguments and assign them a real number that reflects the relative quality of the first one with regard to the other one; *n-ary indicators*, which, in analogous way to binary indicators, analyse more than two PF^*_{known}s at the same time; and a last branch which assess performance through *attainment surfaces*. Despite this classification and indicators variety, quality indicators advantages and drawbacks are not clear, hence Hansen & Jaszkiewicz [41] proposed a first attempt to inferential power quantification by means of three "outperformance relations":

- *Complete outperformance*. Let X and Y be two different PF^*_{known}s, X *completely outperforms* Y ($\mathrm{X}\, O_C\, \mathrm{Y}$) if $\forall F(\boldsymbol{y}) \in \mathrm{Y}, \exists F(\boldsymbol{x}) \in \mathrm{X}: F(\boldsymbol{x}) \prec F(\boldsymbol{y})$.
- *Strong outperformance*. X *strongly outperforms* Y ($\mathrm{X}\, O_S\, \mathrm{Y}$) if $\forall F(\boldsymbol{y}) \in \mathrm{Y},\ \exists F(\boldsymbol{x}) \in \mathrm{X}: F(\boldsymbol{x}) \prec F(\boldsymbol{y}) \vee F(\boldsymbol{x}) = F(\boldsymbol{y})$ and $\exists F(\boldsymbol{z}), F(\boldsymbol{z}'): F(\boldsymbol{z}) \in \mathrm{X}, F(\boldsymbol{z}') \in \mathrm{Y}$ such $F(\boldsymbol{z}) \prec F(\boldsymbol{z}')$.
- *Weak outperformance*. X *weakly outperforms* Y ($\mathrm{X}\, O_W\, \mathrm{Y}$) if $\forall F(\boldsymbol{y}) \in \mathrm{Y},\ \exists F(\boldsymbol{x}) \in \mathrm{X}: F(\boldsymbol{x}) \prec F(\boldsymbol{y}) \vee F(\boldsymbol{x}) = F(\boldsymbol{y})$ and $\exists F(\boldsymbol{z}) \in \mathrm{Y}$ such $F(\boldsymbol{z}) \notin \mathrm{X}$.

It is clear that $O_C \subset O_S \subset O_W$. Additionally, Hansen & Jaszkiewicz [41] studied whether certain quality indicators were compatible with each outperformance relation. In order to enhance Hansen & Jaszkiewicz [41] study, Zitzler *et al* [12] provided a rigorous analysis of quality indicators inference power limitations, aiming to define what statements can be made on the basis of the information provided by quality indicators.

Zitzler *et al* [12] first important contribution was the formal definition of comparison method, conceived from a boolean function which combines one or more quality indicators in order to offer measures interpretation. Let $E: \mathbb{R}^\omega \times \mathbb{R}^\omega \to \{false, true\}$ be a function mapping vectors with length ω of real numbers to boolean values, Λ the feasible solutions space, I_i an m-ary quality indicator: $I_i: \Lambda^m \to \mathbb{R}$, and $I = (I_1, I_2, \dots, I_\omega)$ a combination of quality indicators. A comparison method that only considers unary indicators is defined as $C_{I,E}(\mathrm{X}, \mathrm{Y}) = E\big(I(\mathrm{X}), I(\mathrm{Y})\big)$, meanwhile one that only considers binary indicators is

defined as: $C_{I,E}(\mathrm{X},\mathrm{Y}) = E\big(I(\mathrm{X},\mathrm{Y}), I(\mathrm{Y},\mathrm{X})\big)$, where $I(\mathrm{X}') = \big(I_1(\mathrm{X}'), I_2(\mathrm{X}'), \dots, I_\omega(\mathrm{X}')\big)$ and $I(\mathrm{X}',\mathrm{Y}') = \big(I_1(\mathrm{X}',\mathrm{Y}'), I_2(\mathrm{X}',\mathrm{Y}'), \dots, I_\omega(\mathrm{X}',\mathrm{Y}')\big)$, $\mathrm{X}',\mathrm{Y}' \in \Lambda$. The second important contribution of Zitzler *et al* [12] was coupling comparison method concept with dominance relations, i.e., the agreement with the most general notions in terms of dominance relations between PF^*_{known}s (see Table 4.1) under two conditions. Let $\bowtie$ be a dominance relation between PF^*_{known}s:

- *Compatibility*. The comparison method $C_{I,E}$ is denoted as $\bowtie$-compatible, if either for any $\mathrm{X},\mathrm{Y} \in \Lambda$: $C_{I,E}(\mathrm{X},\mathrm{Y}) \Rightarrow \mathrm{X} \bowtie \mathrm{Y}$ or $C_{I,E}(\mathrm{X},\mathrm{Y}) \Rightarrow \mathrm{Y} \bowtie \mathrm{X}$, sufficient condition.
- *Completeness*. The comparison method $C_{I,E}$ is denoted as $\bowtie$-complete if either for any $\mathrm{X},\mathrm{Y} \in \Lambda$: $\mathrm{X} \bowtie \mathrm{Y} \Rightarrow C_{I,E}(\mathrm{X},\mathrm{Y})$ or $\mathrm{Y} \bowtie \mathrm{X} \Rightarrow C_{I,E}(\mathrm{X},\mathrm{Y})$, necessary condition.

Table 4.1. Dominance relations between sets of non-dominated solutions.

Dominance Relation		**Description**
X strictly dominates Y	$\mathrm{X} \prec\prec \mathrm{Y}$	$\forall F(\boldsymbol{y}) \in \mathrm{Y} \exists F(\boldsymbol{x}) \in \mathrm{X}: F(\boldsymbol{x}) \prec\prec F(\boldsymbol{y})$
X dominates Y	$\mathrm{X} \prec \mathrm{Y}$	$\forall F(\boldsymbol{y}) \in \mathrm{Y} \exists F(\boldsymbol{x}) \in \mathrm{X}: F(\boldsymbol{x}) \prec F(\boldsymbol{y})$
X is better than Y	$\mathrm{X} \lhd \mathrm{Y}$	$\forall F(\boldsymbol{y}) \in \mathrm{Y} \exists F(\boldsymbol{x}) \in \mathrm{X}: F(\boldsymbol{x}) \preccurlyeq F(\boldsymbol{y}) \wedge \mathrm{X} \neq \mathrm{Y}$
X weakly dominates Y	$\mathrm{X} \preccurlyeq \mathrm{Y}$	$\forall F(\boldsymbol{y}) \in \mathrm{Y} \exists F(\boldsymbol{x}) \in \mathrm{X}: F(\boldsymbol{x}) \preccurlyeq F(\boldsymbol{y})$
X and Y are incomparable	$\mathrm{X} \parallel \mathrm{Y}$	neither $\mathrm{X} \prec \mathrm{Y}$ nor $\mathrm{Y} \prec \mathrm{X}$

Despite that Zitzler *et al* [12] gave a proof to verify a theorem that states that in the general case, it is not possible to create a compatible and complete unary comparison method, Lizárraga *et al* [52] recently demonstrated that under practical conditions, the afore mentioned theorem does not hold, giving two possibilities: to find another demonstration for theorem or to demonstrate that a compatible and complete unary comparison method is possible in practice.

In spite of such deductions, a compatible and complete comparison method is restricted in its inference power, as it is unable to distinguish among PF^*_{known}s with incomparable dominance between them but with features that clearly make one better than other one in a none preference objective space. Because in many cases when none of the compared PF^*_{known}s improves the other within dominance relations, one will be interested in whether there are MO evaluation goals by means of which quality differences can be

inferred. Such features have been already defined [39] [34] in order to compute quality in a quantitative way (PF^*_{known}s accuracy, tendency and concordance):

- *Convergence*, it describes how well and how fast PF^*_{known} progress towards R^* (where R^* can be: the real Pareto front PF^*, PF^*_{known} from another MOEA, or an arbitrary set R defined by the user), i.e. how close PF^*_{known} is from R^* (error) and how quick PF^*_{known} approximates to R^* in relation to the evolution process (speed).
- *Uniformity*, it describes how appropriate PF^*_{known} distribution is, meaning the relative distance among solutions; most of the time a homogeneous dispersion is ideal.
- *Spread*, it describes how appropriate PF^*_{known} extension is; a wider PF^*_{known} involves more options.

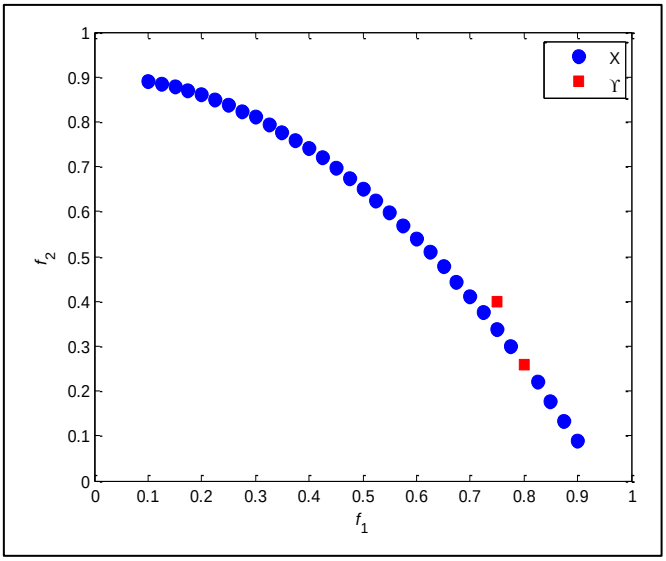

Figure 4.1 Two incomparable approximation to the Pareto Front by a compatible and complete comparison method, X is clearly preferable over Y by MO evaluation goals.

Uniformity and spread imply broader solution choices. Convergence requires a search towards PF^*, while uniformity and spread require a search along PF^*, thus convergence could be seen as orthogonal to uniformity and spread. These MO evaluation goals allow to establish certain judgments when PF^*_{known}s have similar closeness to PF^*. Figure 4.1 shows an example of two Pareto fronts indistinguishable by dominance relations (or a compatible and complete comparison method) but clearly distinguishable by MO evaluation goals. Using MO evaluation goals to untie incomparable PF^*_{known}s requires an appropriated knowledge of indicators accuracy with regard to evaluation goals, though, as it was

mentioned before, such study has not been performed. The present empirical framework is proposed as a guide to know how well an indicator measures what it claims to assess, e.g. each one of the MO evaluation goals.

This chapter presents a new indicator to measure spread, the experimental design and conditions that attempt to classify quality indicators according to the MO evaluation goals, a detailed analysis and discussion of achieved results, and also the description and test of a proposed methodology to compare stochastic multi-objective optimizers.

4.2 A NEW INDICATOR: AVERAGE SPREAD OF THE FOUND PARETO FRONT (ASFPF)

Since some quality indicators assess only a superior bound with similar PF^*, and others are susceptible to uniformity (as it will be seen in the experiments section), we propose a new quality indicator that attempts to achieve a more accurate measure of spread independent of PF^*_{known} uniformity and/or convergence. *Average Spread of the Found Pareto Front* ($ASFPF$) is an improved quality indicator inspired in Lizárraga *et al* [45] and Li & Zheng [46] works, conceived as an *n*-ary indicator. $ASFPF$ is computed as follows:

1. To avoid convergence dependence, PF^*_{known}s are classified according to the partial order that O_C gives by levels as in $G - metric$. Let X_{lp} be the p-th PF^*_{known} in the dominance level l, where $\mathrm{X}_{lp} O_C \mathrm{X}_{l'q} \forall l > l'$.

2. k sets of BS_i are obtained from every X_{lp} as in SA. BS_i is the set of boundary solutions in the i-th X_{lp} projection $\in \mathbb{R}^{k-1}$ where only $\{f_1, \dots, f_{i-1}, f_{i+1}, \dots, f_k\}$ are involved.

3. BS_i computation can be seen as the non-dominated solutions choice of X_{lp} in the projection subspace i by minimizing and maximizing dominance, $BS_i = ND_{min}(\mathrm{X}^i_{lp}) \cup ND_{max}(\mathrm{X}^i_{lp})$. Due to different number of solutions can be expected from both process according to X_{lp} distribution, it is necessary to normalize the amount of boundary solutions implicated in order to avoid uniformity dependence. It is achieved by computing the minimum number of solutions in the dominance process of minimization and maximization of every BS_i of X_{lp} by level, by calculating $\alpha_{li} = \min_{\forall p} |ND_{min}(\mathrm{X}^i_{lp})|$ and $\beta_{li} = \min_{\forall p} |ND_{max}(\mathrm{X}^i_{lp})|$, and restricting $ND_{min}(\mathrm{X}^i_{lp})$ and

$ND_{max}(X_{lp}^{i})$ cardinality to such bounds by truncation process [25], obtaining BS_i', which are new BS_i versions keeping original spread but with homogeneous cardinality in the projection i and the dominance level l.

4. Every BS_i' is normalized within the subspace defined by minimum and maximum projection $\in \mathbb{R}^{k-1}$ computed only with BS_i' from the same projection i and the same dominance level l.

5. The BS_i spread is calculated as the size of the space covered by BS_i', measured by Hypervolume indicator, over BS_i' centroid. The total assessment result $ASFPF$ for X_{lp} is computed as in Equation (4.1).

$$ASFPF \triangleq \left(\prod_{i=1}^{k} \frac{H(BS_i')}{\sum_{\boldsymbol{F}(\boldsymbol{x}) \in BS_i'} f_i(\boldsymbol{x}) / |BS_i'|} \right)^{1/k} \quad (4.1)$$

When comparing PF_{known}^{*}s, the PF_{known}^{*} with a larger $ASFPF$ value has a wider spread.

4.3 EXPERIMENT DESIGN

The aim of this section is to describe the experiments performed to study quality indicators previously shown within MO evaluation goals (convergence, uniformity and spread) in order to perform an empirical taxonomy framework, this through analysing the accuracy and stability of the achieved quality assessment in different PF^{*}s.

4.3.A Experiment Goal

Four experiments are performed with the aim of visualizing the effectiveness of quality indicators. The first experiment attempts to detect if indicators performance is affected by PF_{known}^{*} shape. The second experiment evaluates if indicators performance is affected by PF_{known}^{*} relative position. The third experiment pursues to determine indicators performance with regard to MO evaluation goals. The fourth experiment evaluates how robust are the indicators that measure spread.

4.3.B Experiment Description

Experiment 1

Two PF^*_{known}s are used to evaluate quality indicators sensitivity shape: $A_{concave}$ and B_{convex} with the same convergence, uniformity, spread and number of solutions but different convexity (see Figure 4.2 a) [53].

Experiment 2

Five PF^*_{known}s are used to evaluate quality indicators sensitivity to PF^*_{known} relative position with regard to PF^*: C_1, C_2, C_3, C_4 and C_5 with the same convergence, uniformity, spread and number of solutions but they are located in different zones in relation to PF^* (see Figure 4.2 b) [53].

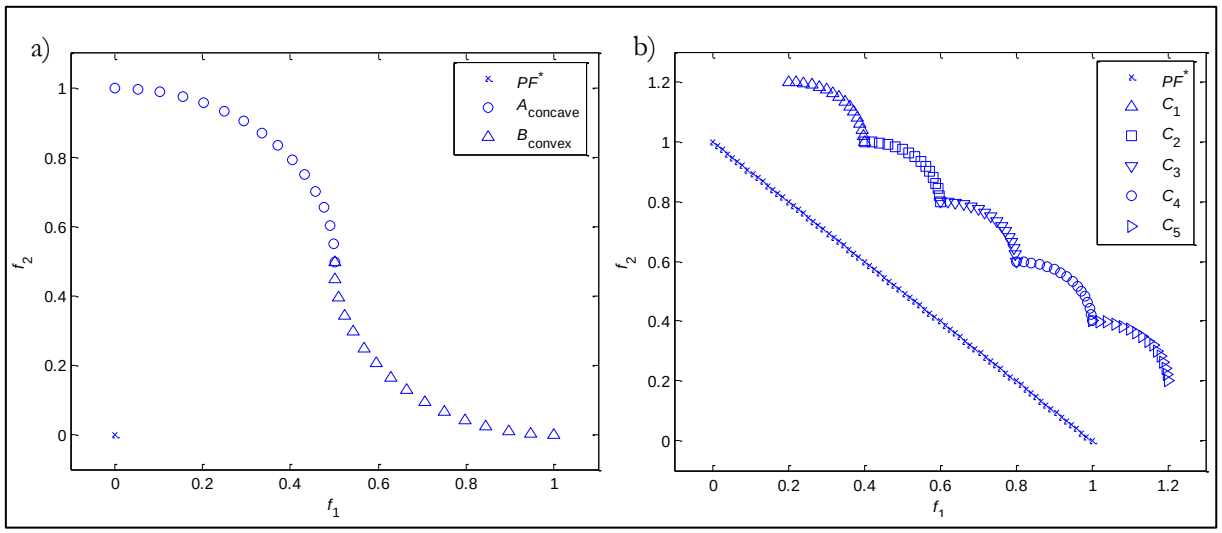

Figure 4.2 Pareto Fronts used in a) Experiment 1 and b) Experiment 2.

Experiment 3

Several synthetic PF^*_{known}s with representative characteristics with regard to MO evaluation goals are generated. Thus starting in a PF_{ini} (initial PF^*) with specific characteristics, several artificial PF^*_{known}s are constructed through PF_{ini} degeneration on each MO evaluation goal (see Figure 4.3 a). Such process can be viewed as the construction of artificial PF^*_{known}s cube formed by circumscribed cubes, where each inner cube represents a synthetic PF^*_{known}, the origin contains PF_{ini} and as it moves away, MO evaluation goals are deteriorated (see Figure 4.3 b). Any PF^*_{known} located at the same distance from the origin with

regard to an axis, has the same value in the evaluation goal related to such axis. PF_{ini}s with different shape (concave, concave-convex and disjoined) in two and three dimensions are used (see Figure 4.4), normalizing them within an unitary square or cube. Synthetic PF^*_{known}s construction can be contextualized as three nested procedures: convergence error increase, uniformity loss and spread deterioration, hence such procedures can be viewed as creating synthetic PF^*_{known}s vectors parallel to error axis. Procedures to degenerate MO evaluation goals are explained as follows. Let $PF_{\varphi,\iota,\psi}$ be a synthetic PF^*_{known} with spread deterioration level φ, uniformity loss level ι and convergence error level ψ. As PF_{ini} has the best feasible convergence, uniformity and spread: $PF_{ini} = PF_{0,0,0}$.

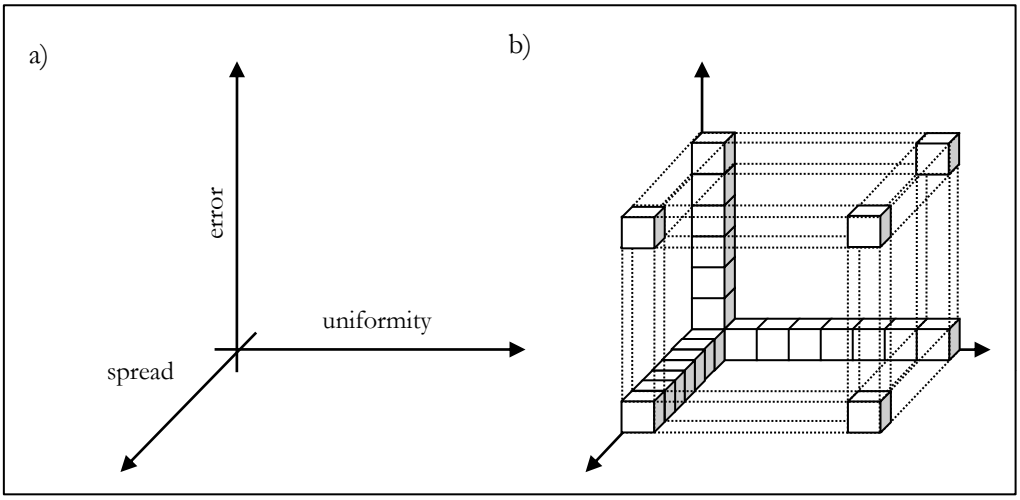

Figure 4.3 Synthetic Pareto Fronts generation: a) describes evaluation criteria dimensions b) represents the total set of synthetic Pareto Fronts generated.

Spread is degenerated by $PF_{\varphi,0,0} = PF_{ini} - FBS_{\varphi}$, where FBS_{φ} is the forbidden boundary subspace at the spread deterioration level φ. By definition $FBS_0 = \{\emptyset\}$, $FBS_{\varphi} = \left\{\boldsymbol{F}(\boldsymbol{x}) \in PF_{ini} \middle| f_i(\boldsymbol{x}) \leq \min_{PF_{ini}} f_i(\boldsymbol{x}) + d_{sd}(\varphi, i) \vee f_i(\boldsymbol{x}) \geq \max_{PF_{ini}} f_i(\boldsymbol{x}) - d_{sd}(\varphi, i), \forall f_i \in \boldsymbol{F}\right\}$, where $d_{sd}(\varphi, i) = \varphi/(2 \cdot edge_{size} + 1) \cdot \left[\max_{PF_{ini}} f_i(\boldsymbol{x}) - \min_{PF_{ini}} f_i(\boldsymbol{x})\right]$, $0 \leq \varphi \leq edge_{size}$ and $edge_{size}$ is the amount of circumscribed cubes by goal axis, gathering ${edge_{size}}^3$ synthetic PF^*_{known}s (see Figure 4.3 b). Spread is degenerated by gradually deleting solutions sections at the boundaries of PF_{ini} (see Figure 4.5 a).

Uniformity is lost by $PF_{\varphi,\iota+1,0} = PF_{\varphi,\iota,0} - FIS_{\varphi,\iota}$, where $FIS_{\varphi,\iota}$ is the forbidden inner subspace at the uniformity loss level ι within $PF_{\varphi,0,0}$, by definition $FIS_{\varphi,0} = \{\emptyset\}$, $FIS_{\varphi,\iota} = \left\{\boldsymbol{F}(\boldsymbol{x}) \in PF_{ini} \middle| f_i(\boldsymbol{x}) \geq\right.$

$\max_{PF_{\varphi,0,0}} f_i(\boldsymbol{x}) - d_{un}(\varphi,\iota,i) \wedge f_i(\boldsymbol{x}) < \max_{PF_{\varphi,0,0}} f_i(\boldsymbol{x}), \forall i = 2,\dots,k\Big\}$, where $d_{un}(\varphi,\iota,i) = \iota/(edge_{size}+1) \cdot \Big[\max_{PF_{\varphi,0,0}} f_i(\boldsymbol{x}) - \min_{PF_{\varphi,0,0}} f_i(\boldsymbol{x})\Big]$. Uniformity is lost by gradually deleting solutions within a subspace located at the extreme of $PF_{\varphi,0,0}$ keeping boundary points to guarantee controlled spread degeneration (see Figure 4.5 a).

Since this experiment pursues to find out how well quality indicators measure each MO evaluation goal in independent way, a "good uniformity" is kept despite spread degeneration, "good uniformity" in the sense of an appropriated (equidistant) distribution within the $PF_{\varphi,\iota,0}$ boundary solutions progressive degenerated, not in the sense of an appropriated distribution according to PF_{ini}. This approach will allow to determine indicators sensibility to spread and uniformity correlation. It is easy to observe that the cardinality of synthetic PF^*_{known}s decrease according to spread and uniformity deterioration, which is why PF_{ini} with cardinality large enough to allow synthetic PF^*_{known}s generation with at least $edge_{size}+2$ was used. In a similar way, aiming to avoid comparative inequality by having PF^*_{known}s with low spread-uniformity deterioration level and high cardinality in one hand, and in the other one PF^*_{known}s with high spread-uniformity deterioration level and low cardinality, a truncation process [25] was performed before starting incremental error phase, guarantying artificial PF^*_{known}s with homogenous cardinality. As PF_{ini} was created using a known function, every solution in any $PF_{\varphi,\iota,\psi}$ is equidistance over the hyper-curve to its closest neighbours, except for the outer and inner boundary solutions in $PF_{\varphi,\iota,\psi}$ (it is important to remark that such distance is measured over the hyper-curve and not as Euclidian distance).

Finally, the convergence error is increased by $PF_{\varphi,\iota,\psi+1} = \{PF^1_{\varphi,\iota,\psi} \cup PF^2_{\varphi,\iota,\psi} \cup \dots \cup PF^l_{\varphi,\iota,\psi} \cup \dots \cup PF^L_{\varphi,\iota,\psi}\}$, where $PF^l_{\varphi,\iota,\psi} \subseteq PF_{\varphi,\iota,\psi}$ and $\left\{PF^j_{\varphi,\iota,\psi} \cap PF^l_{\varphi,\iota,\psi}\right\}_{j,l=1,2,\dots,L \wedge j \neq l} = \emptyset$, such sets are created by inclusion of vectors belonging to $PF_{\varphi,\iota,\psi}$ according to a lexicographic sorting where the sorting index will associate every $\boldsymbol{F}(\boldsymbol{x}) \in PF_{\varphi,\iota,\psi}$ to an specific $PF^l_{\varphi,\iota,\psi}$, preserving homogeneous cardinality in every $PF^l_{\varphi,\iota,\psi}$. At every ψ, convergence error level is augmented as $\forall \boldsymbol{F}(\boldsymbol{x}) \in PF^{(\psi-1)\bmod L}_{\varphi,\iota,\psi}: f_i(\boldsymbol{x}) = f_i(\boldsymbol{x}) + \delta_{error}, \forall f_i \in \boldsymbol{F}$, where the increasing rate $\delta_{error} = \inf\{\min_{\boldsymbol{F}(\boldsymbol{x}),\boldsymbol{F}(\boldsymbol{y}) \in PF_{ini}} |f_i(\boldsymbol{x}) - f_i(\boldsymbol{y})|, \forall f_i \in \boldsymbol{F}\}$. Note than despite of alternate error increase, initial lexicographic sorting is not affected due to δ_{error} definition. Convergence error is increased by incrementing a specific $PF^l_{\varphi,\iota,\psi}$ at every iteration process

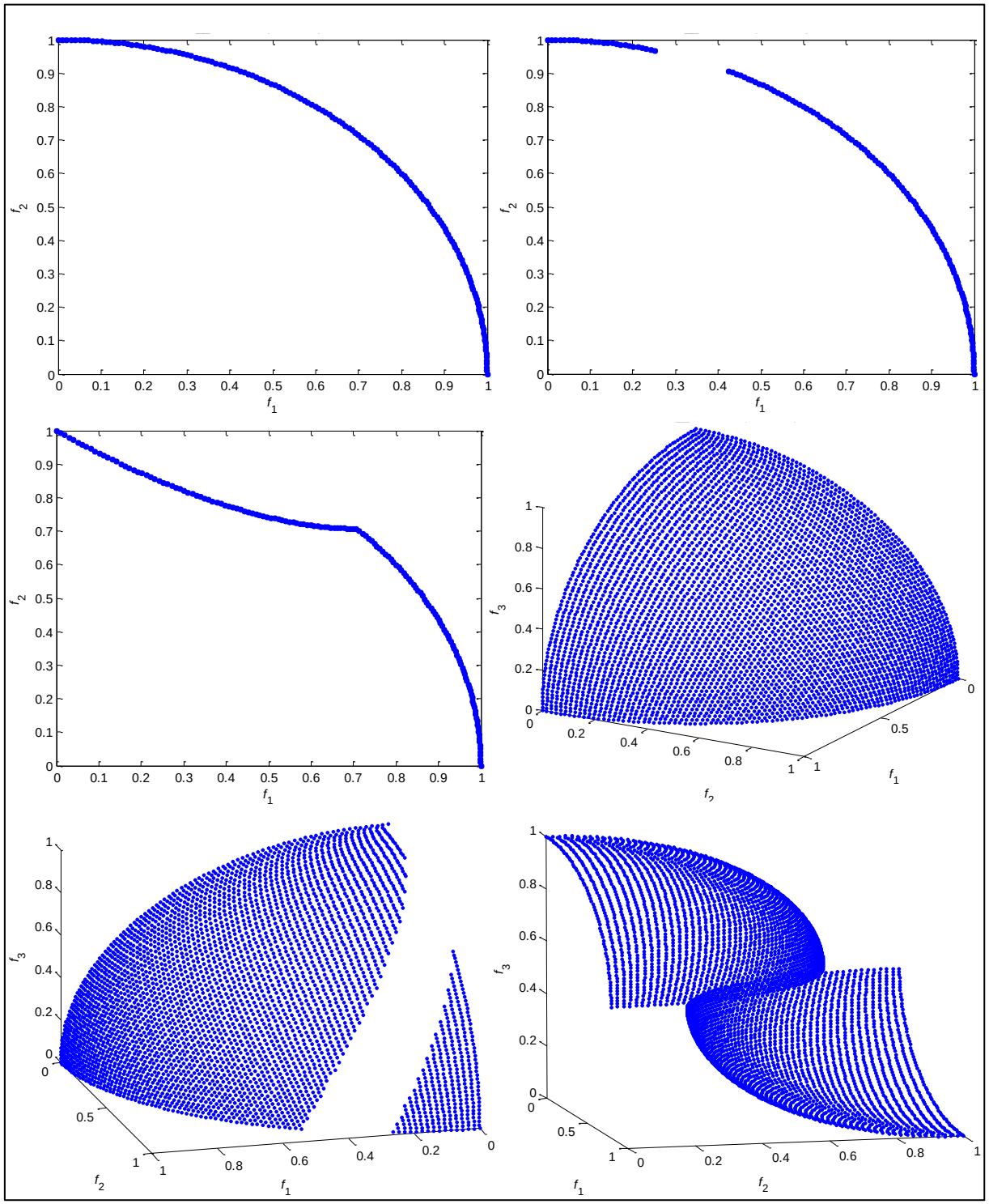

Figure 4.4 Experiment 3: PF_{inf}s used with different shape (concave, disjoined and concave-convex) in two and three dimensions.

with δ_{error} (see Figure 4.5 b), such error expression was conceived instead of a simple progressive deterioration attempting to include performance improvement relations described in [41]. For this experiment two configurations are used: $edge_{size} = 1000$ and $L = 10$ to visualize a detailed performance panorama, and $edge_{size} = 100$ and $L = 2$ to visualize a general performance panorama.

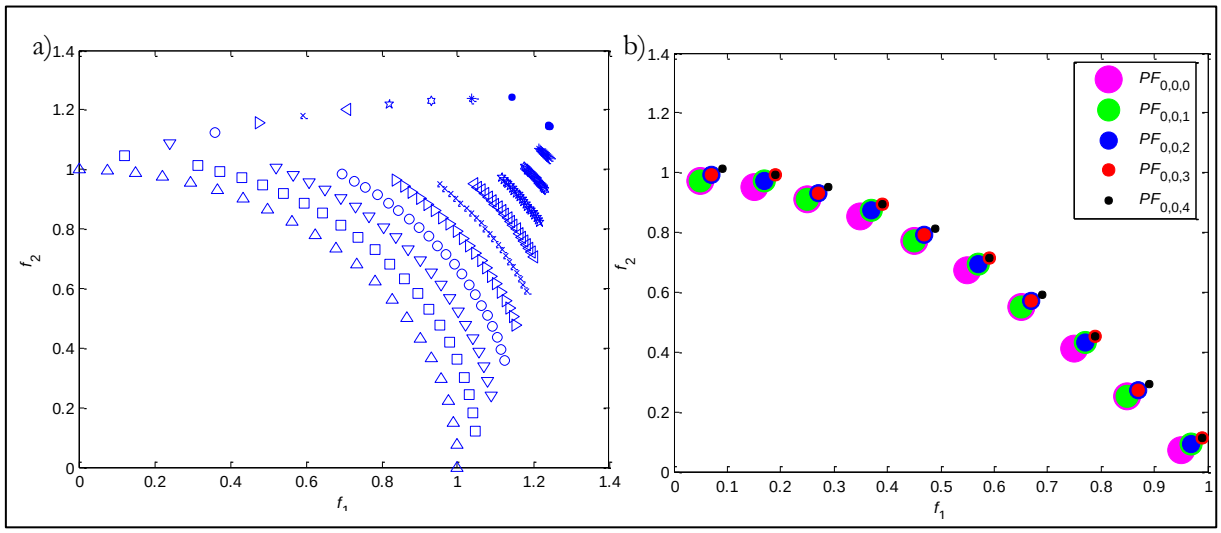

Figure 4.5 Graphic illustration of synthetic Pareto Fronts degeneration, lowest curve represents PF_{ini}: a) progressive error degeneration is showed aiming to clearly view extension and dispersion degeneration b) error alternate increase with L=2 and *edge_size*=4.

Experiment 4

Four PF^*_{known}s are used to evaluate quality indicators robustness in measuring PF^*_{known} spread: D_1, D_2, D_3 and D_4 with the same convergence, number of solutions and extreme solutions in each axis, equidistance solutions but different spread (see Figure 4.6). Even Experiment 3 is able to determine if quality indicators measure spread with no influence of convergence or uniformity, some indicators can only get a superior bound for some PF^*s family, thus this experiment is added with the aim to show $ASFPF$ superior performance over other spread indicators.

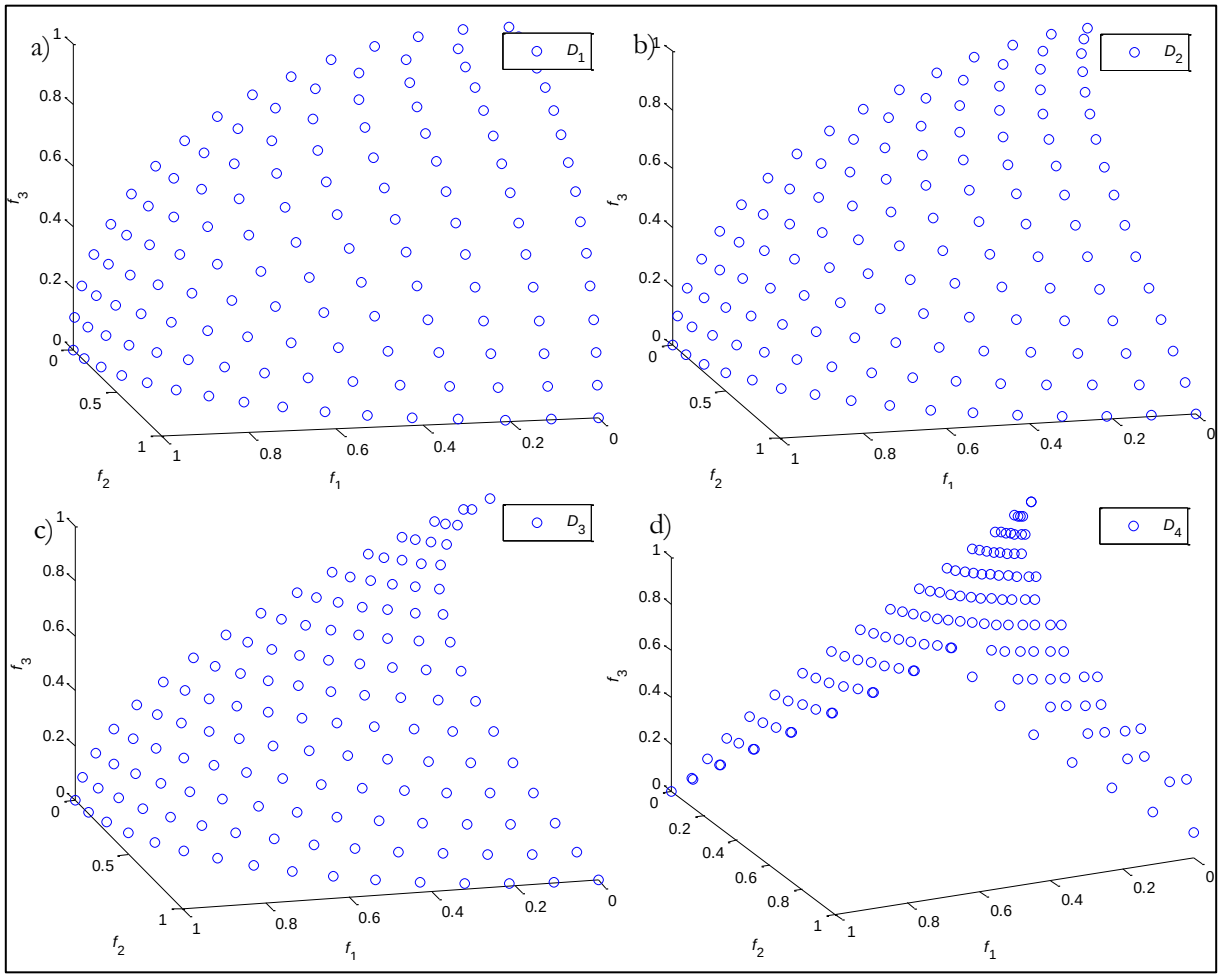

Figure 4.6 Pareto Fronts used in Experiment 4.

4.3.C Experiment Conditions

In the case of quality indicators configuration, for those which involve weighted vectors or utility vectors, a uniform distribution $[0,1]$ is used to create them using a proportional intensity function for each vector produced; those indicators which use utility functions are tested with functions sets of the same type (e.g. maximum aggregation sum and Chebyshev distance). $\|\cdot\|_2$ is used for every indicator that uses a non-predefined norm. $PAAS$ and OPS_k indicators are not included in experimentation due to their graphics as values of utility are difficult to interpret. Concerning to indicators which measure convergence speed (GCV, $GNVG$, NVA and RMC), they are not evaluated through synthetic PF^*_{known}s, rather a MOEA

applied to a specific MOP is used. Experiment 3 only involves quality indicators that measure spread. Finally, neither $ONVG$ nor $ONVGR$ are included in Experiment 4 due to their evaluated characteristics are not compatible with the MO evaluation goals, since PF^*_{known}s with same indicator value could have very different MO evaluation goals levels.

4.4 RESULTS AND DISCUSSION

4.4.A Results of Experiments 1 and 2

Results of Experiment 1 and 2 are shown in Table 4.2, and summarized in Table 4.4 columns 9 and 10. A quality indicator independent of PF^*_{known} convexity gets the same value for $A_{concave}$ and B_{convex}, meanwhile, a quality indicator independent of PF^*_{known} relative position gets the same value for every $C_i: 1 \leq i \leq 5$. Quality indicators that show convexity and relative position independence are: SD, ESE, $FRPF$, $ONVG$, $ONVGR$, CTS, $DFPF$, $EFPF$, NDC_μ, CL_μ, $G - metric$ and $ASFPF$.

4.4.B Results of Experiment 3

As a result of evaluating quality indicators with synthetic PF^*_{known}s, three *behaviour graphics* were obtained by every PF_{ini} by every experiment configuration for each indicator in a three-dimensional space: uniformity-convergence, spread-convergence and spread-uniformity. In every case, the axis z represents quality indicator estimation, whose value is computed as the average of the synthetic PF^*_{known}s contained in the parallel vector to the MO evaluation goal absent in the graphic. Goal axis scale of behaviour graphics is denoted by 1 as the best synthetic PF^*_{known} and 0 as the worst synthetic PF^*_{known} in the related MO evaluation goal. Due to PF_{ini}s with different shape-scalability were used and that most of the indicators show normal changes in their behaviour according to PF_{ini} type, graphics of 3D-concave PF_{ini} are included and indicated only in situations where different PF_{ini}s shape clear up extra information. The same applies to experiment configuration, most of the time a general panorama is visualized; only when detailed panorama gives extra information such graphic is shown and explained. Every behaviour graphic

Table 4.2. Results of Experiment 1 and 2.

Quality Indicator	$A_{concave}$	B_{convex}	C_1	C_2	C_3	C_4	C_5
SD	10	10	10	10	10	10	10
ESE	0.0120	0.0120	0.0068	0.0068	0.0068	0.0068	0.0068
WSA	0.3762	0	0.452	0.451	0.4	0.2	0
$FRPF$	0	0	0	0	0	0	0
GD	0.2189	0.1812	0.2708	0.2310	0.2161	0.2310	0.2708
$ONVG$	17	17	12	12	12	12	12
$ONVGR$	1.7	1.7	1.2	1.2	1.2	1.2	1.2
H	0.2972	0.4390	0.1669	0.2866	0.3266	0.2866	0.1669
CTS	1	1	1	1	1	1	1
$CDTS$	0.7027	0.5609	0.8330	0.7133	0.6733	0.7133	0.8330
PS	0.6772	0.3781	1	0.9845	0.7564	0.9845	1
ED	-0.002	-0.003	0.001	-0.002	-0.003	-0.002	0.001
EPS	0.001	0.002	0.001	0.002	0.003	0.002	0.001
$MPFE$	0.8413	0.7868	0.9083	0.8495	0.8495	0.8495	0.9083
$ADPF$	0.8977	0.7377	0.9363	0.7995	0.7486	0.7995	0.9363
$DFPF$	15	15	6.5454	6.5454	6.5454	6.5454	6.5454
$EFPF$	0.9995	0.9995	0.6321	0.6321	0.6321	0.6321	0.6321
HD	0.7027	0.5609	0.8330	0.7133	0.6733	0.7133	0.8330
OPS	0.4990	0.4990	0.1996	0.3193	0.3592	0.3193	0.1996
$AOPF$	3.8130	3.8130	1.5567	2.4822	3.0958	2.4822	1.5567
NDC_{μ}	17	17	12	12	12	12	12
CL_{μ}	0.5882	0.5882	0.8333	0.8333	0.8333	0.8333	0.8333
RMD	6.321	6.982	5.269	5.257	5.789	5.257	5.269
AD	0.501	0.501	0.801	0.601	0.5521	0.601	0.801
WD	0.501	0.501	0.801	0.601	0.5521	0.601	0.801
I_{ε}	0.4995	0.4995	0.7992	0.5994	0.4685	0.5994	0.7992
$I_{\varepsilon+}$	-0.504	-0.504	-0.207	-0.405	-0.534	-0.405	-0.207
$U-measure$	1.2623	1.2623	1.3044	1.2841	1.2748	1.2841	1.3044
$UPOS$	4.5244	3.6043	1.1142	1.2032	3.6983	1.2032	1.1142
$W-EPOS$	0.0416	0.0364	0.0687	0.0601	0.0601	0.0601	0.0687
$G-metric$	0.7831	0.7831	0.6624	0.6624	0.6624	0.6624	0.6624
SA	1.7093	2.4200	1.7622	1.8849	1.9015	1.8849	1.7622
$ASFPF$	6	6	18	18	18	18	18

is interpreted according to optimum value of quality indicator (PS, EDS and EPS imply opt in the evaluation set FU). For facilitating discussion understanding use Table 4.4.

After analysing every indicator behaviour three possible views in MO evaluation goals were found: 1) the indicator measures only one MO evaluation goal; 2) the indicator measures more than one MO evaluation goal considering them with the same pre-eminence; and 3) the indicator measures more than one MO evaluation goal considering them with different pre-eminence level; such views are determined by observing slope of behaviour graphics, e.g. largest slope implies highest pre-eminence. With Pareto compliant indicators conception [18], indicators classification by MO evaluation goals is not complete, thus overlapping classes are possible, that is why through experimentation three types of assessments are defined. Indicators characterization under situations 1 and 2, and the indicator connection with the MO evaluation goal of highest hierarchy in the situation 3 will be called *strong assessment*. In the other hand, MO evaluation goals whose relation with the indicators is not the highest hierarchy will be called *weak* and *weakest assessment* according to their measurement hierarchy. Monotony and relativity of quality indicators are also analysed; (weak) monotony is defined as the fact of given a non-dominated set, adding a non-dominated point improves (does not degrade) its evaluation; meanwhile (weak) relativity is related to PF^* evaluations as (non)-uniquely optimal [15]. Results of Experiment 3 and discussion of the whole experimentation are presented in the following.

Quality Indicators measuring Convergence error

Indicators that only assess convergence error are: $FRPF$, GD, CTS, $ADPF$, I_ε and $I_{\varepsilon+}$. $FRPF$ measures the non-dominated solutions proportion from PF^*_{known} with regard to R^*, hence not relevant information is provided when the PF^*_{known} does not include any non-dominated solutions with regard to R^*, it is seen in behaviour graphics as convergence error increases (see Figure 4.7 a and b). Furthermore $FRPF$ measurement could be highly subjective due to defining R^* is not a trivial task. As $FRPF$ is based on ER definition, it is not monotonic but weak relative [15]. GD measures effectively convergence error as it is seen in behaviour graphics (see Figure 4.7 c and d), however according to [15] this indicator is not monotonic but weak relative.

CTS was measured by $CTS(PF_{\varphi,\iota,\psi}, PF^*)$, i.e. solutions percentage in PF^* that dominate solutions in the synthetic $PF_{\varphi,\iota,\psi}$. CTS behaviour graphics show its capability to measure convergence but its inability to measure uniformity and spread (see Figure 4.7 e and f); CTS behaviour is disruptive in relation to

uniformity and spread, since a better achievement of such goals produces that more solutions in PF^* dominate $PF_{\varphi,\iota,\psi}$ letting $CTS \to 1$. In its unary version CTS violates relativity, given two subsets, the first one a PF^* subset and the second an arbitrary non-Pareto optimal subset but with wider spread than the PF^* subset; CTS will favor the second subset over the first. Otherwise CTS exhibits weak monotony, due to adding solutions to a non-dominated subset could not improve CTS evaluation.

$ADPF$ measures convergence error (see Figure 4.7 g and h), slight variation in top of behaviour graphics is explained by uniformity-spread degeneration. $ADPF$ is not monotonic but weak relative for the same reasons that GD is. I_ε and $I_{\varepsilon+}$ show total independence from uniformity and spread, computing effectively convergence error by performing uniformity-convergence and spread-convergence monotonic behaviour graphics with regard to convergence error. Such performance is due to I_ε and $I_{\varepsilon+}$ compute the infimum (see Figure 4.7 i, j, k and l). Unary versions of both indicators are weak relative and weak monotonic, since any PF^* subset has an optimal value and the addition of solutions to a non-dominated subset may not improve their value.

Quality Indicator measuring Spread

$EFPF$ and $ASFPF$ are the only indicators formulated able to isolate spread estimation from convergence and uniformity; however as it will be shown in Experiment 4, $EFPF$ only computes a spread superior bound, meanwhile $ASFPF$ achieves a more accurate measurement. Only $EFPF$ graphics are shown (see Figure 4.8 a and b) given that both indicators get exactly the same behaviour in this experiment. Due to adding a non-dominated solution does not degrade quality indicator and a PF^*_{known} with wider extension in the objective space than PF^* does not imply better performance, $EFPF$ and $ASFPF$ are both weak monotonic but not relative.

Quality Indicators measuring Convergence error and Spread

Indicators that assess convergence error and spread are: WSA, PS, EDS, EPS, OPS, RMD and SA. WSA measures convergence and spread (see Figure 4.8 c, d and e), though the hierarchy level of both goals is highly dependent of the distribution used to generate LC; the main drawback of this indicator is that the quality assessment by linear combination is defined only by the worst solution. As WSA is LC dependent it is neither monotonic nor relative.

PS, EDS and EPS perform estimations over FU. Consequently, according to utility functions type the efficiency of measurement and computed MO evaluation goals may vary. In this experiment aggregative addition functions and Chebyshev functions are used, the first one allows to compute convergence error and spread (see Figure 4.8 f, g, h, l, Figure 4.9 a, b, f, g and h), meanwhile the second one allows to compute convergence error and misunderstand spread (see Figure 4.8 i, j, k, Figure 4.9 c, d, e, i, j and k). EDS and EPS complement PS by expressing how much the difference between solutions sets is, computing discrepancy of expected degree of superiority and the expected proportion of superiority. Due to a wide FU variety is required, these indicators are in general computationally expensive. As PS, EDS and EPS are FU dependent, they are neither monotonic nor relative.

OPS denotes hierarchical distinction between spread and convergence. Giving a major preeminence to spread, such fact is induced because $H_{ex}(\ )$ is defined by boundary solutions, influencing mainly OPS sensitivity to PF^*_{known} spread (see Figure 4.9 l, Figure 4.10 a and b). OPS is weak monotonic and weak relative, because of adding solutions to a non-dominated subset could not improve OPS, and a nonempty PF^* subset could have an optimal OPS. RMD was implemented with neighbourhood analysis schema suggested by the author, whose definition is by itself very subjective. As it is seen in behaviour graphics (see Figure 4.10 c and d), RMD measures spread over convergence and fails to measure uniformity; it is shown when spread is close to the maximum. Due to neighbourhood definition and RMD susceptibility to PF^*_{known} position, it is weak monotonic but not relative.

SA measures correctly convergence over spread and erroneously uniformity. This is due to 1) using H within SA, 2) as uniformity is reduced, the average value of f_i computed for every projection (w_i) is affected, producing a bias towards the region with the highest solutions concentration (see Figure 4.10 e, f and g), making it disturbed by uniformity variation. SA exhibits weak monotony and weak relativity, because adding non-boundary solutions to a non-dominated subset does not improve SA, and a PF^* subset could achieve SA optimal value.

Quality Indicators measuring Convergence error and Uniformity

Indicators that assess convergence error and uniformity are: $MPFE$, AD, WD and $W-EPOS$. $MPFE$ measures the maximum error between R^* and PF^*_{known}, as it is shown in behaviour graphics (see Figure 4.10 h and i), $MPFE$ measures convergence error and uniformity with a hierarchical distinction favoring

convergence; however spread estimation is misunderstood, such fact is deduced by not monotony in its measurements as in Figure 4.10 j, and explained by $MPFE$ susceptibility to PF^*_{known}s location.

AD and WD were designed aiming to offer information about divergence between PF^*_{known} and R^*, measuring the average and the major separation respectively. Due to synthetic PF^*_{known}s continuous degeneration in MO evaluation goals and artificial equidistance solutions, behaviour graphics from both indicators are equal in the scope of Experiment 3. However in a non-controlled environment, WD decrease could be a more fortuitous because it computes the worst distance. Even AD and WD were created to measure only convergence, they measure uniformity with a minor hierarchy and misunderstand spread; such behaviour is more evident for a concave-convex PF^*_{known} (see Figure 4.10 k, l and Figure 4.11 a). This performance is due to average and worst distance are computed from every solution in R^* to the closest solution in PF^*_{known}, as PF^*_{known} spread decreases some solutions in R^* will compute a bigger distance to the closest solution in PF^*_{known}. As $MPFE$, $W-EPOS$ measures convergence error over uniformity (see Figure 4.11 b and c); misunderstanding spread (see Figure 4.11 d), fact explained by their susceptibility to PF^*_{known}s location. Due to $MPFE$, AD, WD and $W-EPOS$ have similar conception they share weak relativity and their absence of monotony. They are not monotonic, because given a non-dominated set, the addition of a non-dominated solution with a large enough convergence error could degrades evaluation; and exhibit weak relativity, since any P^* subset has an optimal quality indicator value.

Quality Indicators measuring Uniformity and Spread

Indicators that assess uniformity and spread are: SD, ESE, $AOPF$, NDC_μ, CL_μ, $DFPF$, $U-$*measure* and $UPOS$. SD measures spacing between PF^*_{known} solutions within indifference regions. In the experiment indifference regions as those defined by NDC_μ and CL_μ were used. Due to regions definition is quite subjective, as soon as PF^*_{known} drives enough away from R^*, SD measurement becomes useless (see Figure 4.11 e and f). SD with a maximum convergence error minor to the size of the indifference region is showed in Figure 4.11 g and h in order to visualize SD instability.

ESE computes spacing between solutions using Euclidian distance, even ESE measures correctly uniformity, spread is improperly computed (see Figure 4.11 i, j and k). It is mainly caused by ESE sensitivity to uniformity-spread correlation, widespread synthetic PF^*_{known}s will allow a wider uniformity variation impacting ESE assessment. Besides, ESE is unstable before disjoined PF^*_{known}s (see Figure 4.11

l, behaviour graphic achieved with a disjoined PF_{ini}). To avoid its degeneration, distance exclusion between limit points in PF^* formed from two or more Pareto curves is needed. Due to PF^* could be disjoined, ESE is neither monotonic nor relatively.

$AOPF$ measures properly uniformity and spread (see Figure 4.12 c); nevertheless it is unstable with regard to convergence (see Figure 4.12 a and b) due to its measurement is highly dependent on PF^*_{known} location within normalized objective space, i.e. the hypervolume computed by boundary solutions varies as PF^*_{known} position does. Figure 4.12 d shows two PF^*_{known}s with the same uniformity and spread but different position generating dissimilar $AOPF$ measurement. As $AOPF$ is based on H, it is monotonic but not relative, since its PF^*_{known} location dependence.

NDC_μ and CL_μ measure uniformity and spread using a grid of hypercubes to define indifferent regions. Both indicators show roughness in their behaviour graphics (see Figure 4.12 e, f, g, h, i and j) due to the grid concept they utilize and PF^*_{known}s location, conceiving a conditioned measurement and giving raise to no reflexivity. The principal difference between NDC_μ and CL_μ, is CL_μ capacity to give higher priority to uniformity over spread. In the other hand, $DFPF$ is not susceptible to PF^*_{known}s location, as a result of indifferent regions conceptualization as hyperspheres, allowing to achieve smoother behaviour graphics (see Figure 4.12 k, l and Figure 4.13 a) and as a consequence, a better inference power about uniformity.

SD, NDC_μ, CL_μ and $DFPF$ compute their assessment considering only solutions located in the same indifference region. Thus indifference regions definition is very important to these indicators characterization, since in an environment with a very small or very large indifference region none measure could be achieved. Such indicators are highly susceptible to indifference space resolution, factor defined by q, μ or σ. SD, NDC_μ, CL_μ and $DFPF$ are weak monotonic because adding solutions to a non-dominated subset could not improve indicator value. NDC_μ and CL_μ are not relative due to a non-dominated subset far from PF^* with wider extension and more solutions could get a better value than PF^*. Due to SD uses indifferent regions fixed to R^* and that $DFPF$ can be conceived under an appropriated definition of resolution factor, both indicators are weak relative, since non-dominated subsets could achieve at most the same assessment than PF^*.

Finally, $U-measure$ and $UPOS$ measure uniformity and spread too; as $U-measure$ sets a slightly preference of spread over uniformity (see Figure 4.13 b, c and d), $UPOS$ does in a stronger way of uniformity over

spread (see Figure 4.13 e, f and g). $UPOS$ improves ESE by both, inv( ) function and PF^*_{known} reduction in case of tide. Due to both indicators are sensible to disjoined PF^*_{known}s, they are neither monotonic nor relatively.

Quality Indicators attempting to measure Convergence Error, Uniformity and Spread

Indicators that assess convergence error, uniformity and spread are: H, HD, $CDTS$ and $G-metric$. On the basis of experimentation none indicator is able to measure with the same hierarchy the three MO evaluation goals. H (see Figure 4.13 h, i and j) and HD (see Figure 4.13 k, l and Figure 4.14 a) are able to measure convergence error, spread and uniformity; convergence error with the highest level, spread with medium level and uniformity with the lowest level. Due to by experiment definition R^* dominates weakly any solution in the synthetic PF^*_{known}s, $CDTS$ has the same performance that HD. As H, HD and $CDTS$ are based on hypervolume computation they are monotonic and relative.

$G-metric$ measures in hierarchical order convergence error, spread and uniformity (see Figure 4.14 b, c and d). This indicator is monotonic and relative and overcomes a drawback common in H, HD and $CDTS$: their inability to distinguish between PF^*_{known}s with the different level of complete outperformance, since $G-metric$ does not allow to a nondominate subset get a better assessment than another if the first subset does not outperformance the second one. Despite, a new disadvantage rises with $G-metric$ projection, Figure 4.14 e shows a concave-convex PF^*_{known} with equidistance solutions. Figure 4.14 f is Figure 4.14 e projection; such graphic supports $G-metric$ susceptibility to mix up uniformity in relation to PF^*_{known} shape, since center solutions in the projection are closer than the remaining solutions.

Quality Indicators measuring Convergence speed

Indicators that assess MOEAs convergence speed according to its evolution through generations are: $GNVG$, NVA, GCV and RMC. $GNVG$ and NVA are independent from any other MO evaluation goal; regardless, they provide poor information since they describe in a limited sense PF^*_{known} robustness. $GNVG$ measures effectiveness of evolutionary process, meanwhile NVA measures the evolutionary steps certainty given through evolutionary process (see Figure 4.14 h). In the other hand, GCV and RMC get the speed ratio from $PF^*_{current}(t)$ approximation to PF^*_{known} (see Figure 4.14 g).

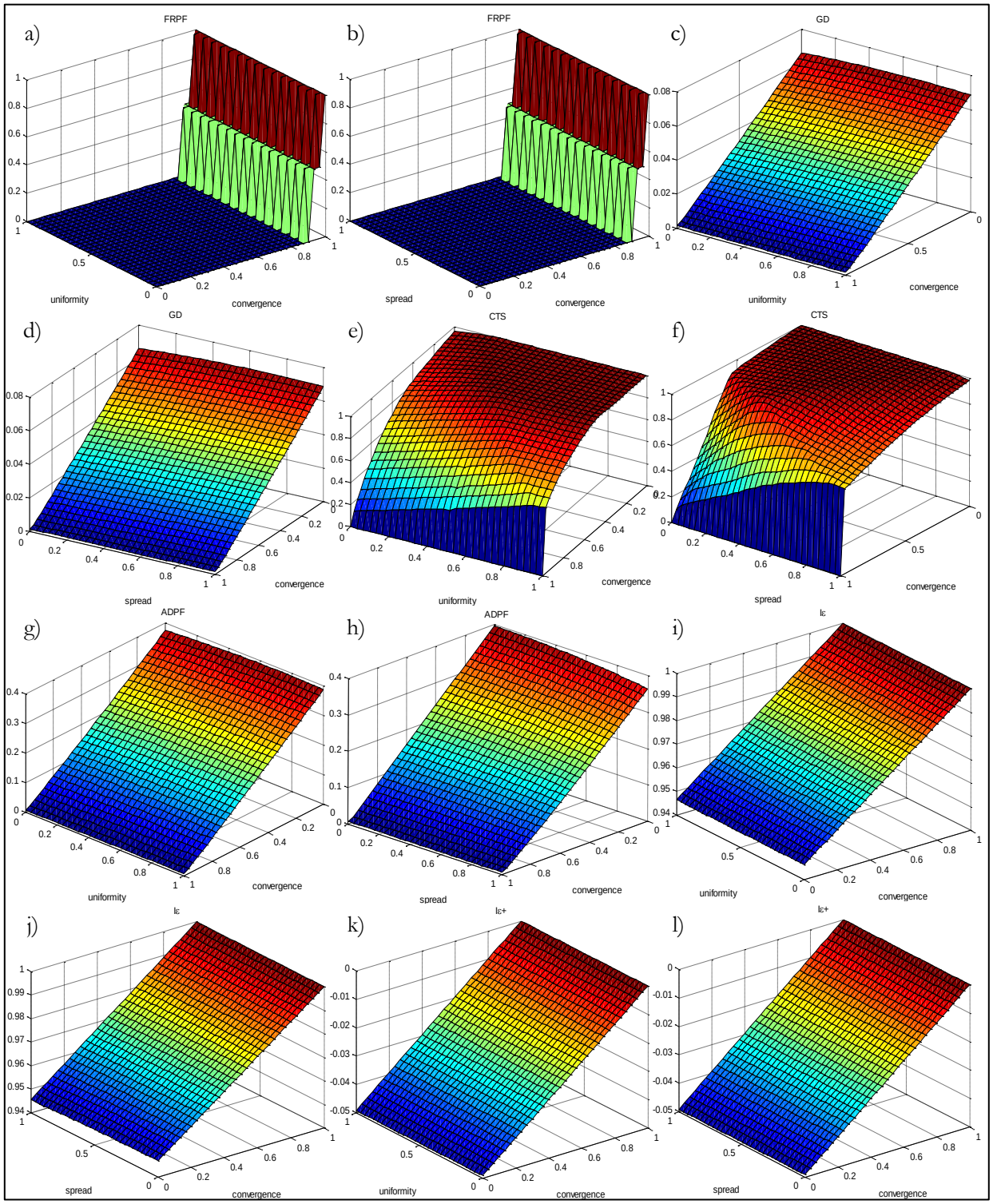

Figure 4.7 Behaviour graphics: *FRPF* a) and b); *GD* c) and d); *CTS* e) and f); *ADPF* g) and h); I_{ε} i) and j); $I_{\varepsilon+}$ k) and l).

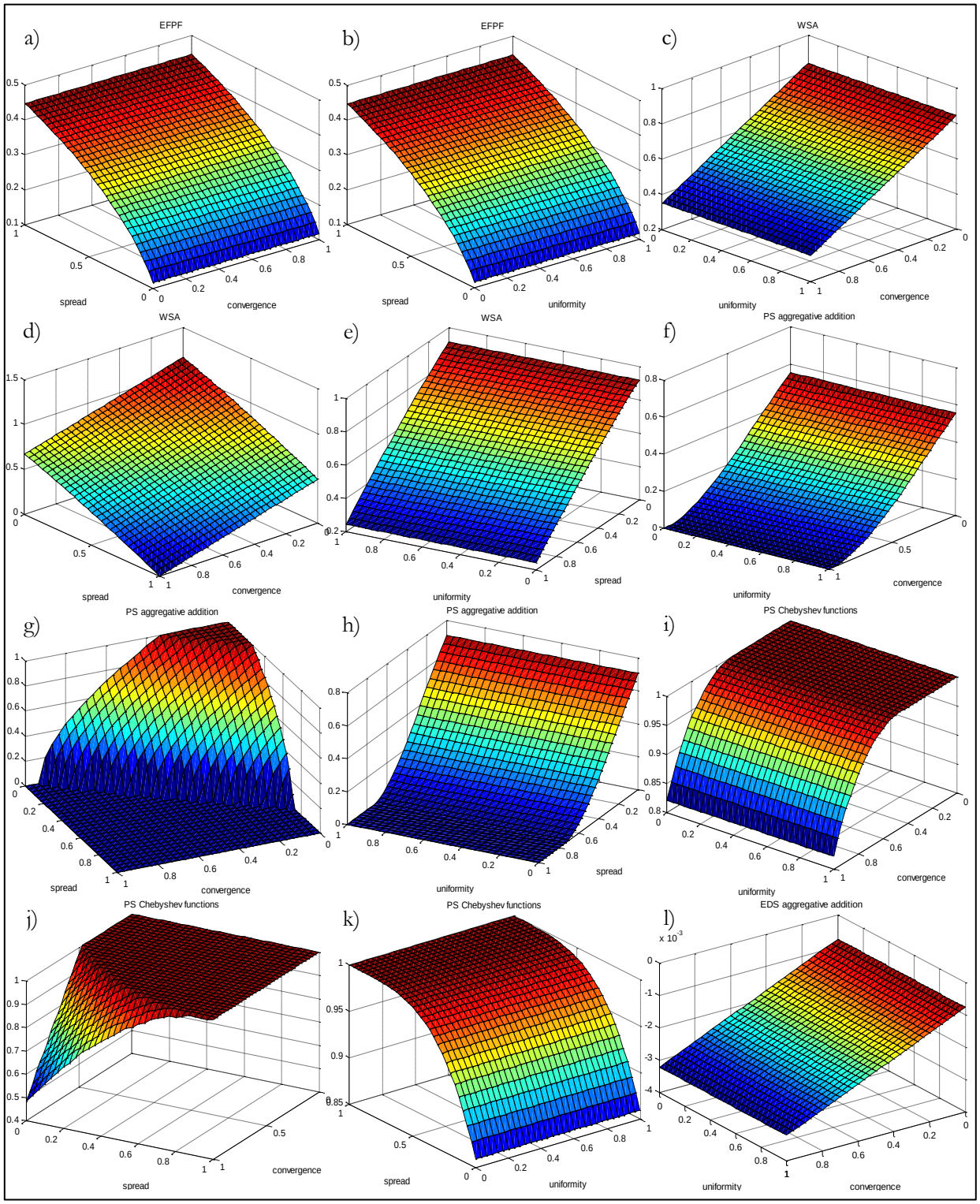

Figure 4.8 Behaviour graphics: *EFPF* a) and b); WSA c), d) and e); *PS* f), g), h), i), j) and k); *EDS* l).

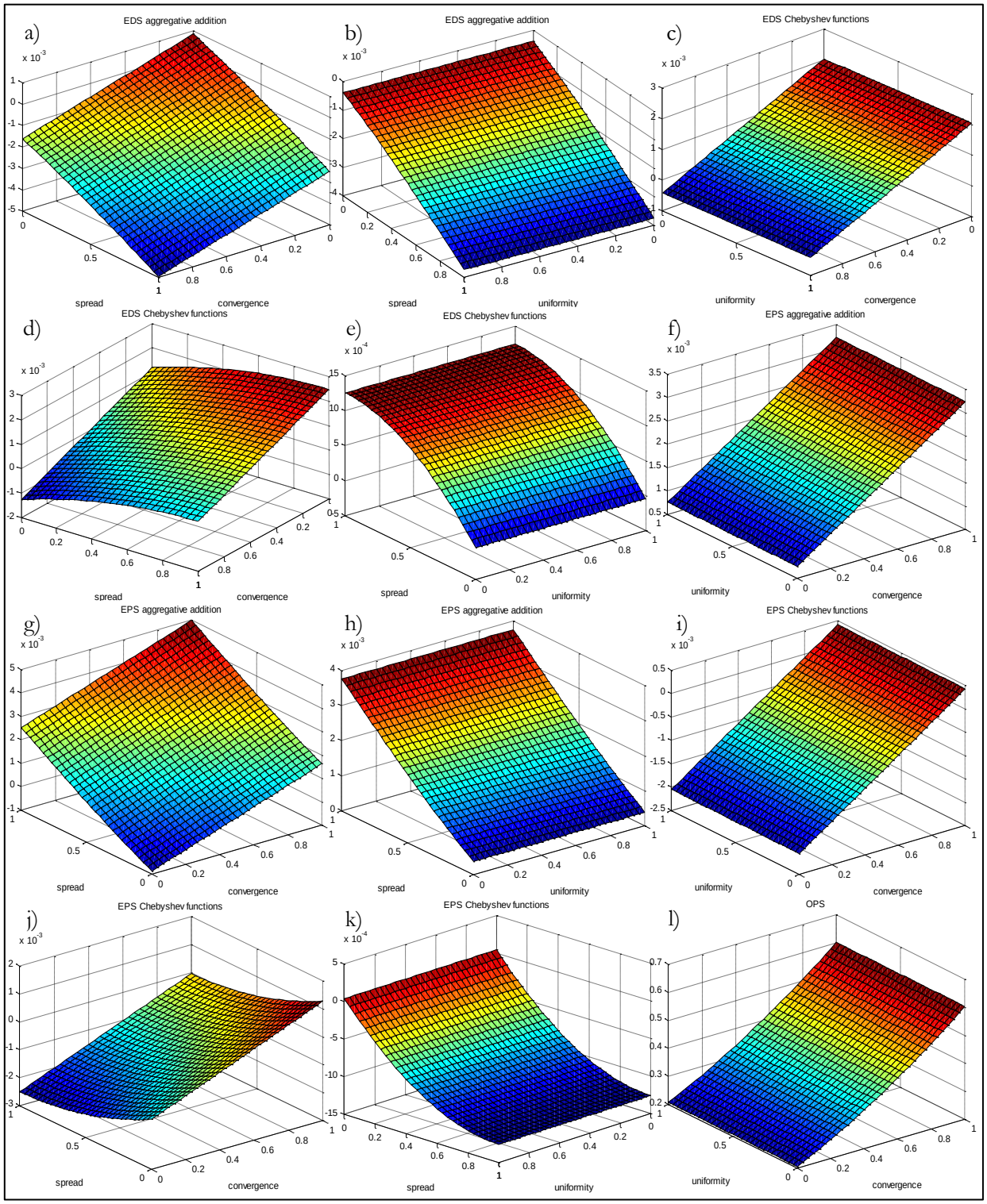

Figure 4.9 Behaviour graphics: *EDS* a), b), c), d) and e); *EPS* f), g), h), i), j) and k); *OPS* l).

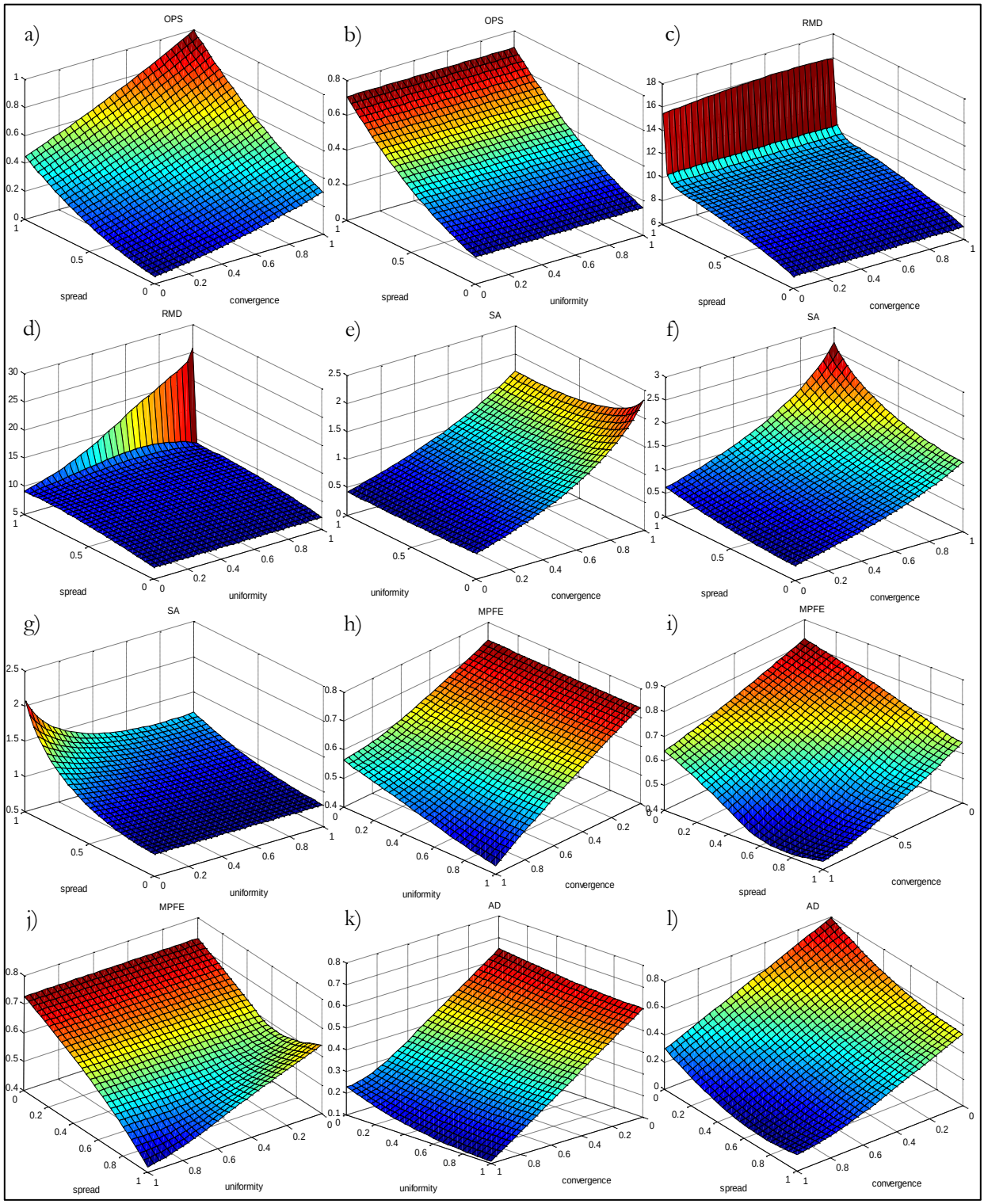

Figure 4.10 Behaviour graphics: *OPS* a) and b); *RMD* c) and d); *SA* e), f) and g); *MPFE* h), i) and j); *AD* k) and l).

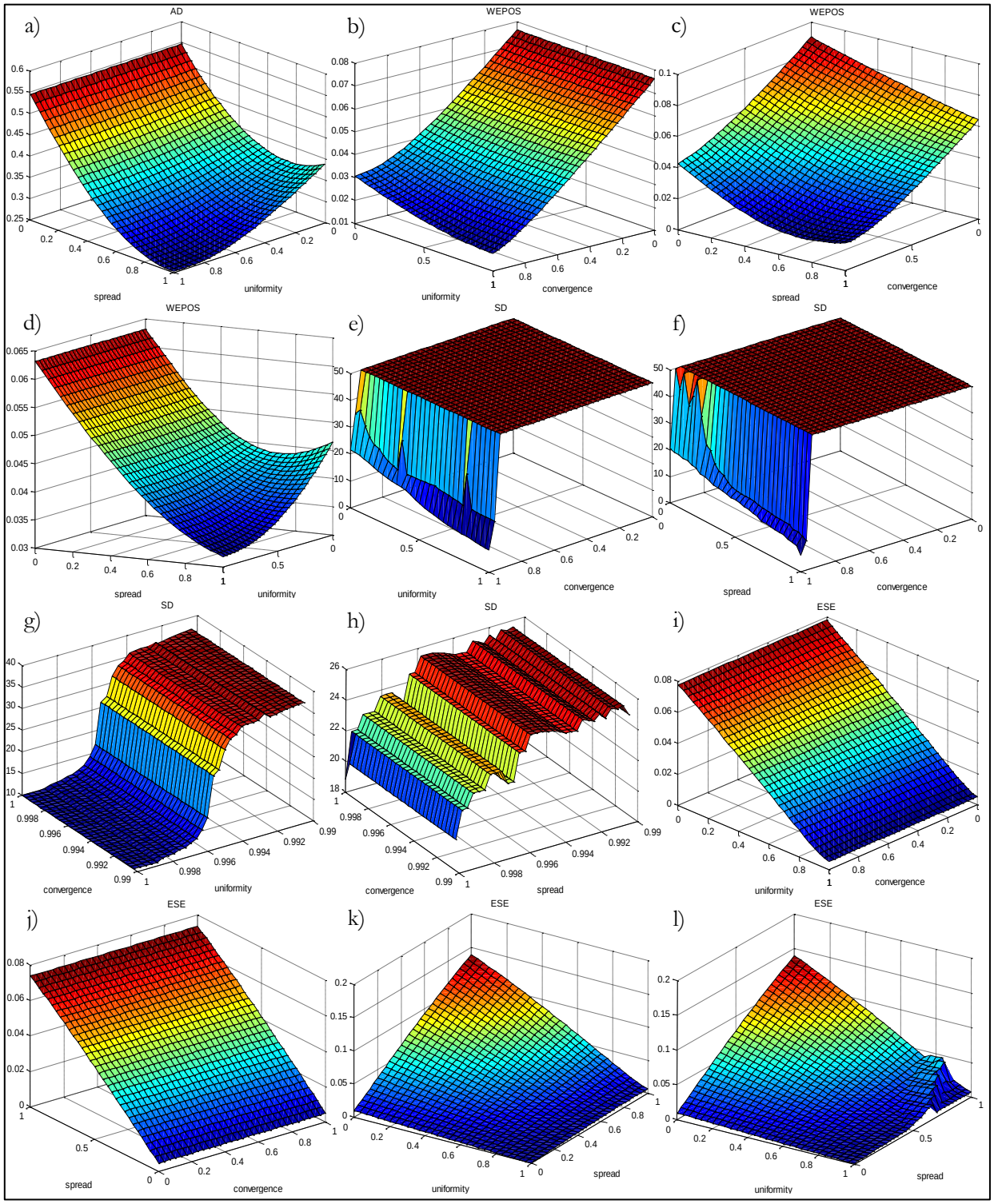

Figure 4.11 Behaviour graphics: *AD* a); *WEPOS* b), c) and d); *SD* e), f), g) and h); *ESE* i), j), k) and l).

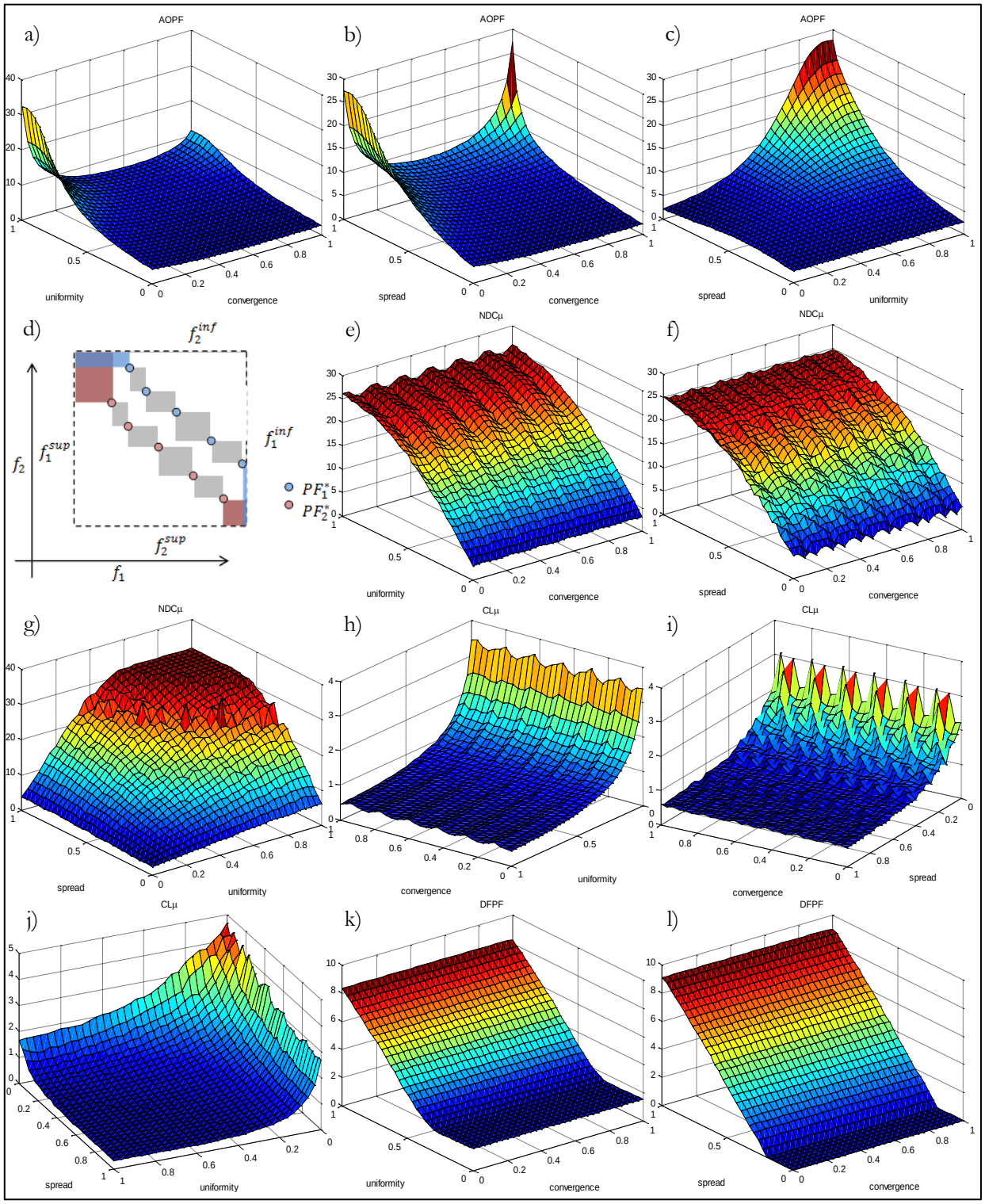

Figure 4.12 Behaviour graphics: *AOPF* a), b), c) and d); NDC_{μ} e), f) and g); CL_{μ} h), i) and j); *DFPF* k) and l).

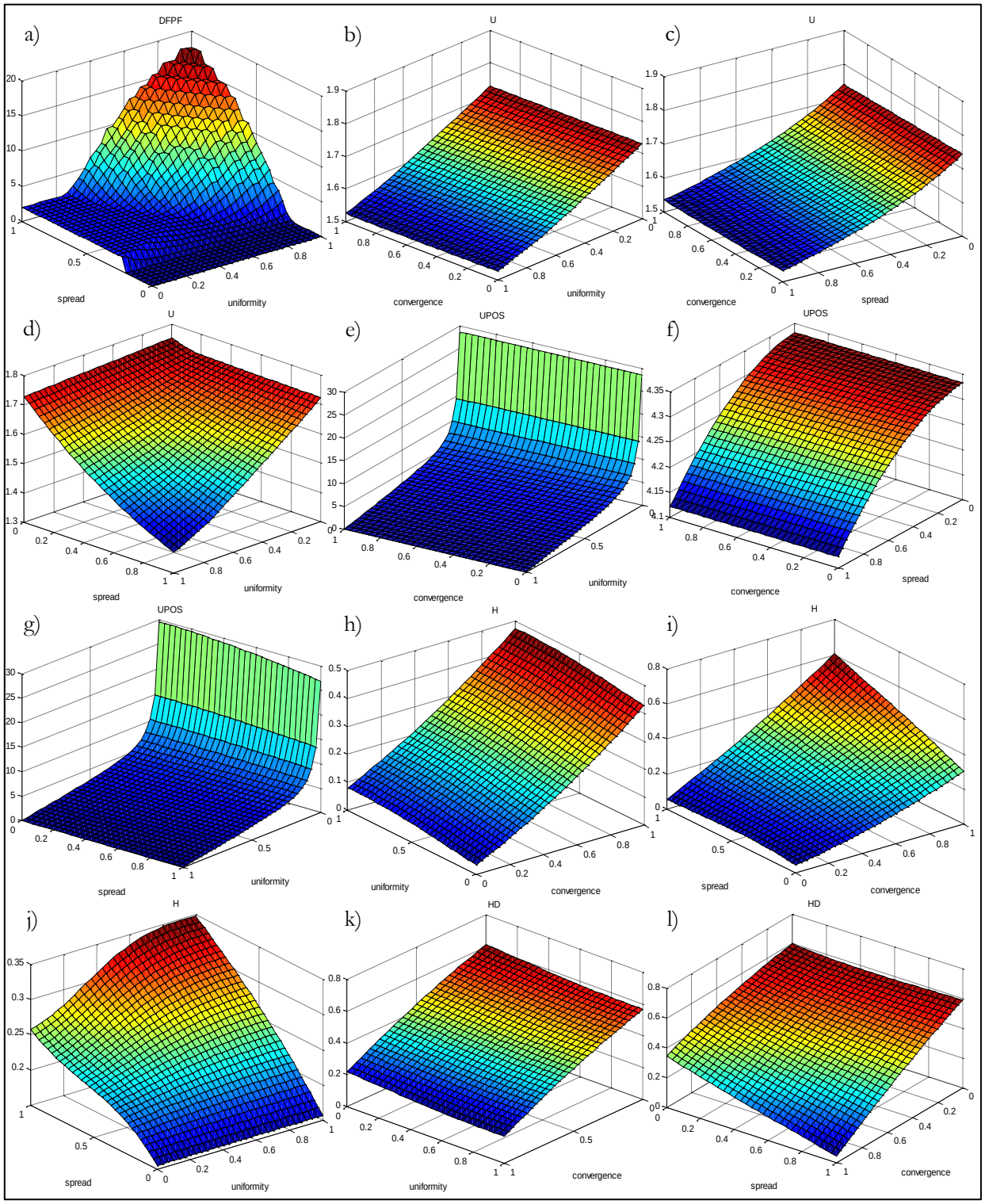

Figure 4.13 Behaviour graphics: *DFPF* a); *U-measure* b), c) and d); *UPOS* e), f) and g); *H* h), i) and j); *HD* k) and l).

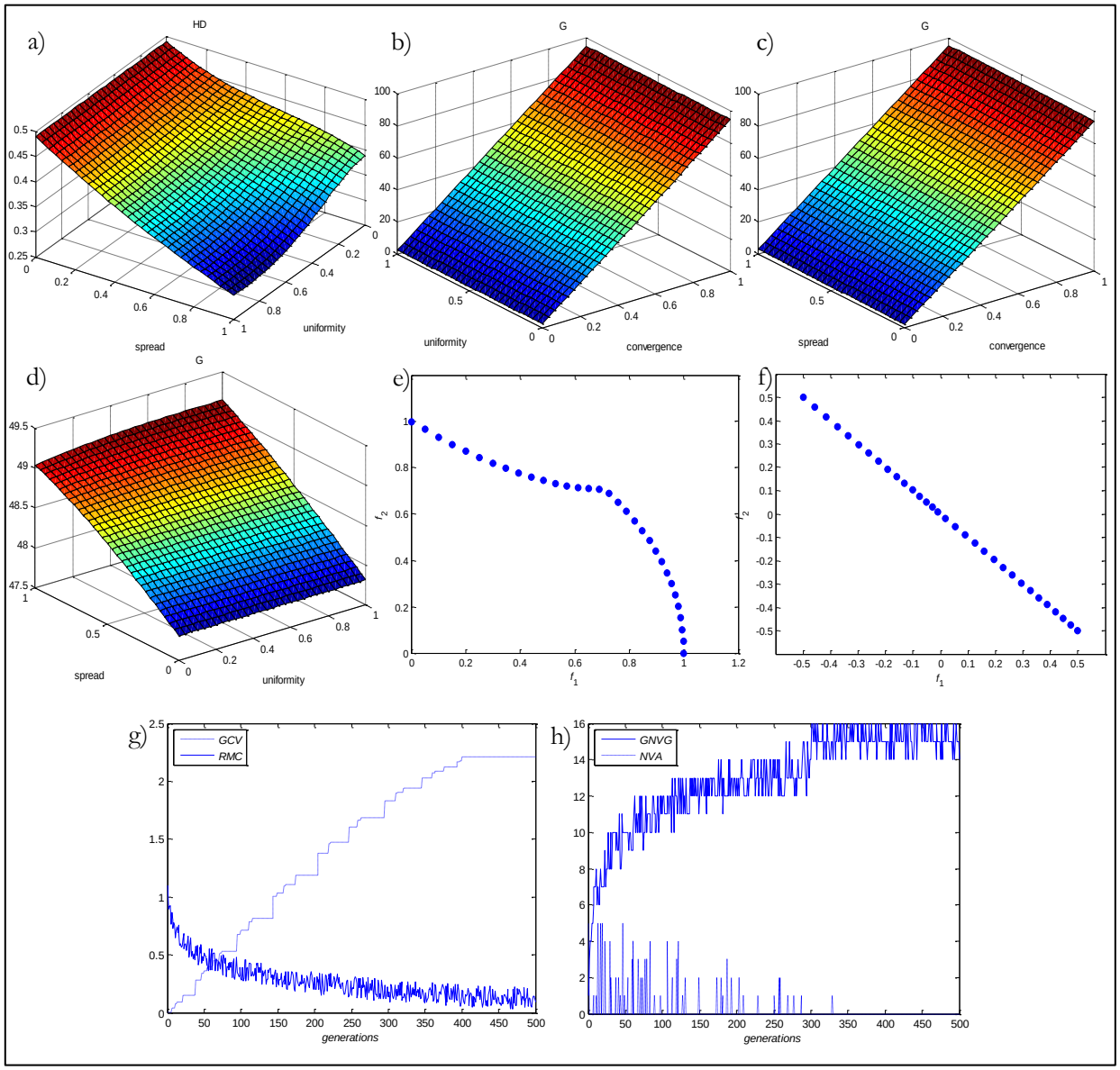

Figure 4.14 Behaviour graphics: *HD* a); *G* b), c) and d); e) concave-convex Pareto Front and f) its projection. Performance from quality indicators measuring convergence speed: g) *GCV* and *RMC*, h) *GNVG* and *NVA*.

4.4.C Results of Experiment 4

Results of Experiment 4 are shown in Table 4.3. This experiment validates $ASFPF$ robustness, since it is the only indicator able to distinguish among the four different PF^*_{known}s and order them from D_1 to D_4.

4.4.D Quality Indicators Summary

Table 4.4 summarizes relevant characteristics from each of the thirty-eight quality indicators reviewed and $ASFPF$ (proposed in this research). For each quality indicator information about type, reference set requirement, MO evaluation goals achieved, evaluated characteristic (cardinality gets its index value from solution sets or solution subsets elements counting to be evaluated; distance is founded in the use of norms to measure length; volume calculates solution quality based on dominance hypervolume), sensitivity

Table 4.3. Results of Experiment 4.

Quality Indicator	D_1	D_2	D_3	D_4
SD	10	10	10	10
WSA	0.0001	0.0001	0.0001	0.0001
GD	0.0857	0.0857	0.0857	0.0857
H	0.0270	0.0178	0.0441	0.0578
$CDTS$	0.9729	0.9821	0.9558	0.9421
PS	1	1	1	1
EDS	-0.001	-0.001	-0.001	-0.001
EPS	0.001	0.001	0.001	0.001
$ADPF$	1.0045	1.0048	1.0050	1.0051
$DFPF$	76	76	76	75.955
$EFPF$	1.7311	1.7311	1.7311	1.7296
HD	0.9729	0.9821	0.9558	0.9421
OPS	0.9970	0.9970	0.9970	0.9970
$AOPF$	1.0461	1.0566	1.0813	1.0965
NDC_{μ}	76	76	76	66
CL_{μ}	0.0735	0.0735	0.0735	0.0793
RMD	21	21	21	21
$U-measure$	0.6594	0.5974	0.5928	0.7315
$UPOS$	0.0509	0.0443	0.5745	0.5034
$G-metric$	6.9134	6.9134	6.7960	6.0926
SA	5.0597	4.9180	4.8200	6.9181
$ASFPF$	12.9310	10.3346	8.1016	3.1871

Table 4.4. Taxonomy of studied Quality Indicators.

Quality Indicator	Type	Needs R^*	MO evaluation goals: Convergence Error	MO evaluation goals: Convergence Speed	MO evaluation goals: Uniformity	MO evaluation goals: Spread	Evaluated Characteristic	Independent to Convexity	Independent to Relative Position	Additional Parameters	Monotony	Relativity	Range	Optimum Value	Computational complexity
SD	u				✓	✓	c	✓	✓		*	*	$[0,\infty)$	0	$O(k/2\mu \cdot \lvert PF^*_{known}\rvert)$
ESE	u				✓		d	✓	✓				$[0,\infty)$	0	$O(\lvert PF^*_{known}\rvert^2)$
WSA	u		✓			✓	d			LC			$[0,\infty)$	0	$O(\lvert PF^*_{known}\rvert \cdot \lvert FU\rvert)$
$PAAS$	–		✓		✓		c	–	–		–	–	–	–	–
$FRPF$	u	✓	✓				c	✓	✓			*	[0,1]	1	$O(\lvert PF^*_{known}\rvert \cdot \lvert PF^*\rvert)$
GD	u	✓	✓				d					*	$[0,\infty)$	0	$O(\lvert PF^*_{known}\rvert \cdot \lvert R^*\rvert)$
GCV	u	✓	✓	✓			d	–	–			*	$[0,\infty)$	–	$O(\lvert PF^*_{known}\rvert \cdot \lvert R^*\rvert)$
$ONVG$	u						c	✓	✓		–	–	$[0,\infty)$	∞	$O(\lvert PF^*_{known}\rvert)$
$ONVGR$	u	✓					c	✓	✓		–	–	[0,1]	1	$O(\lvert PF^*_{known}\rvert)$
$GNVG$	u			✓			c	–	–		–	–	$[0,\infty)$	∞	$O(\lvert PF^*_{current}(t)\rvert)$
H	u		✓		**	*	v			f_i^{inf}	✓	✓	$[0,\infty)$	∞	$O\left(\lvert PF^*_{known}\rvert^{k/2}\right)$
CTS	b	✓	✓				c	✓	✓		*		[0,1]	0	$O(\lvert U\rvert \cdot \lvert V\rvert)$
$CDTS$	b	✓	✓		**	*	v			f_i^{inf}	✓	✓	$[0,\infty)$	0	$O\left((\lvert U\rvert + \lvert V\rvert)^{k/2}\right)$
PS	b	✓	✓			*	d			FU			[0,1]	opt PS	$O(\lvert U\rvert \cdot \lvert FU\rvert)$
EDS	b	✓	✓			*	d			FU			$(-\infty,\infty)$	opt EDS	$O(\lvert U\rvert \cdot \lvert FU\rvert)$
EPS	b	✓	✓			*	d			FU			$(-\infty,\infty)$	opt $-EPS$	$O(\lvert U\rvert \cdot \lvert FU\rvert)$
$MPFE$	u	✓	✓		*		d					*	$[0,\infty)$	0	$O(\lvert U\rvert \cdot \lvert V\rvert)$
NVA	u			✓			c	–	–		–	–	$[0,\infty)$	∞	$O(\lvert PF^*_{known}(t)\rvert)$
$ADPF$	u	✓	✓				d					*	$[0,\infty)$	0	$O(\lvert PF^*_{known}\rvert \cdot \lvert PF^*\rvert)$
$DFPF$	u				✓	*	d	✓	✓	σ	*	*	$[0,\infty]$	∞	$O(\lvert PF^*_{known}\rvert^2)$
$EFPF$	u					✓	d	✓	✓		*		$[0,\infty)$	∞	$O(\lvert PF^*_{known}\rvert^2)$
HD	b	✓	✓		**	*	v			f_i^{inf}	✓	✓	[0,1]	0	$O\left(\lvert PF^*_{known}\rvert^{k/2}\right)$
OPS	u		*			✓	v	✓		f_i^{inf}, f_i^{sup}	*	*	[0,1]	1	$O\left(\lvert PF^*_{kown}\rvert^{k/2}\right)$
OPS_k	u					✓	d	–	–		–	–	$[0,\infty)$	∞	$O(\lvert PF^*_{known}\rvert)$
$AOPF$	u				✓	✓	v	✓		f_i^{inf}, f_i^{sup}	✓		$[0,\infty)$	∞	$O\left(\lvert PF^*_{known}\rvert^{k/2}\right)$
NDC_μ	u				✓	✓	c	✓	✓	μ	*		$[0,\infty)$	∞	$O(k/2\mu \cdot \lvert PF^*_{known}\rvert)$
CL_μ	u				✓	*	c	✓	✓	μ	*		$[1,\infty]$	1	$O(k/2\mu \cdot \lvert PF^*_{known}\rvert)$
RMC	u	✓	✓	✓			d	–	–		–	–	$[0,\infty)$	0	$O(\lvert PF^*_{current}(t)\rvert \cdot \lvert R\rvert)$
RMD	u		*			✓	c			h, nh^{**}	*		$[0,\infty)$	∞	$O(h \cdot \lvert PF^*_{known}\rvert)$
AD	u	✓	✓		*		d	✓				*	[0,1]	0	$O(\lvert U\rvert \cdot \lvert V\rvert)$
WD	u	✓	✓		*		d	✓				*	[0,1]	0	$O(\lvert U\rvert \cdot \lvert V\rvert)$
I_ε	b	✓	✓				d	✓			*	*	$(0,\infty)$	∞	$O(\lvert U\rvert \cdot \lvert V\rvert)$
$I_{\varepsilon+}$	b	✓	✓				d	✓			*	*	$(-\infty,\infty)$	∞	$O(\lvert U\rvert \cdot \lvert V\rvert)$
$U-measure$	u				*	✓	d	✓		f_i^{inf}, f_i^{sup}			$[0,\infty)$	0	$O(\lvert PF^*_{known}\rvert^2)$
$UPOS$	u				✓	*	d						$(1,\infty)$	1	$O(\lvert PF^*_{known}\rvert^2)$
$W-EPOS$	u	✓	✓		*		d					*	$[0,\infty)$	0	$O(\lvert PF^*_{known}\rvert \cdot \lvert PF^*\rvert)$
$G-metric$	n		✓		**	*	d	✓	✓	*radius* υ	✓	✓	$[0,\infty)$	∞	$O\left(n \cdot \max_{\{U_1,..U_i,...U_n\}} \lvert U_i\rvert^2\right)$
SA	u		✓			*	v				*	*	$(0,\infty)$	∞	$O\left(\lvert PF^*_{known}\rvert^{k/2}\right)$
$ASFPF$	n					✓	v	✓	✓		*		$[0,\infty)$	∞	$O\left(\lvert PF^*_{known}\rvert^{k/2}\right)$

to convexity, sensitivity to relative position, additional parameters, monotony, relativity, value ranges, optimum value and computational complexity is given. First column specifies quality indicator acronym. Second column denotes type of indicator (*u* = unary, *b* = binary and *n* = n-ary). Third column indicates whether the indicator needs any type of R^*. From the fourth column to the seventh MO evaluation goals are related (✓ = strong assessment, * = weak assessment, ** = weakest assessment). Eighth column describes what kind of evaluated characteristic use the indicator (*d* = distance, *c* = cardinality and *v* = volume). Ninth column indicates whether the indicator is sensible or not to PF^*_{known} convexity. Tenth column indicates whether the indicator is sensible or not to the relative position of the PF^*_{known}. Eleventh column shows indicators parameters that need be specified by the user (nh^{**} = analysis of neighborhood). Twelfth and thirteenth columns specify whether the indicator exhibits monotony or relativity (✓ = monotony/relativity, * = weak monotony/weak relativity). Fourteenth and fifteenth columns detail the range of possible values achieved by the quality indicator and the best value that the non-dominated set under assessment could achieve. Sixteenth column describes computational complexity estimated for the quality indicator.

4.5 A PERFORMANCE COMPARISON METHODOLOGY TO STOCHASTIC MULTI-OBJECTIVE OPTIMIZERS

In EMO, how to evaluate PF^*_{known} quality that different MOEAs generate is still an open problem. Based on that fact, it is suggested in this PhD thesis an approach to solve such problem. The proposed methodology summarizes PF^*_{known}s achieved on certain number of MOEA runs to accomplish an average PF^*_{known} achieved by the MOEA, then it compares MOEAs results in the context of [12] and [41] relations and finally untied incomparable front using MO evaluation goal, i.e., it is proposed to combine attainment surfaces, a binary comparison method compatible and complete, and unary indicators in order to discriminate among PF^*_{known}s quality. The methodology is shown below:

Step 1: For each $\text{MOEA}_i \in \Gamma: i = 1,2,\ldots,\varrho$ to be compared, use its S runs to compute a summary attainment surface. Since such process can take a long time [54], it is suggested to use an approximation method as the one presented in [54], where an approximated summary attainment surface is computed in

polynomial time with regard to k and $\max_{s=1,2,..,S}\left|PF^{*}_{known,i,s}\right|$ with a fixed total number of sampling lines, $PF^{*}_{known,i,s}$ is the PF^{*}_{known} achieved by MOEA_i in its s-th run.

Step 2: Use a binary comparison method $\lhd$-compatible and $\lhd$-complete in order to sort summary attainment surfaces according to $\lhd$-relation. $C_{CTS,E'}(\mathrm{X},\mathrm{Y})$ and $C_{I_\varepsilon,E''}(\mathrm{X},\mathrm{Y})$ are $\lhd$-compatible and $\lhd$-complete comparison methods computationally not too expensive that could be used, where $E' \coloneqq (CTS(\mathrm{X},\mathrm{Y}) = 1 \wedge CTS(\mathrm{Y},\mathrm{X}) < 1)$ and $E'' \coloneqq (I_\varepsilon(\mathrm{X},\mathrm{Y}) \leq 1 \wedge I_\varepsilon(\mathrm{Y},\mathrm{X}) > 1)$ [12]. Step 1 is included in this methodology since it reduces Step 2 complexity from $\varrho(\varrho-1)s^2$ to $\varrho(\varrho-1)$.

Step 3: Calculate MO evaluation goals mean and variance using unary indicators and R^* over the S runs of each MOEA_i; R^* is formed as the total Pareto Front taking in count every run from every MOEA. On the basis of this research it is suggested to use: I_ε or $I_{\varepsilon+}$ to assess convergence error, $DFPF$ to assess uniformity and $ASFPF$ to assess spread.

Step 4: Rank MO evaluation goals using a statistical measure, e.g. ε' presented in [48] and shown in Equation (4.2).

$$\varepsilon'_{i,j} = 10\left(\frac{\mu_{I,\mathrm{MOEA}_i} - \mu_{I,\mathrm{MOEA}_j}}{\sqrt{\sigma^2_{I,\mathrm{MOEA}_i} + \sigma^2_{I,\mathrm{MOEA}_j}}}\right) \tag{4.2}$$

where the numerator is the difference in average of indicator I between MOEA_i and MOEA_j, and the denominator is the variance difference. If we assume a normal distribution then, $\varepsilon'_{i,j} > 2$ corresponds to a 95% confidence interval which will take to mean that it is statistically significant that MOEA_j is leading to better I values than MOEA_i. The rank process of every indicator I can be performed with Equation (4.3).

$$rank_I(\mathrm{MOEA}_i) = \begin{cases} 0 & \text{if } \nexists \mathrm{MOEA}_j \in \Gamma \left| \varepsilon'_{j,i} > 2 \right. \\ \max\limits_{\mathrm{MOEA}_j \in \Gamma : \varepsilon'_{j,i} > 2} \left[rank_I(\mathrm{MOEA}_j) \right] + 1 & \text{otherwise} \end{cases} \tag{4.3}$$

Step 5: Combine binary comparison method assessment and MO evaluation goals rank using a lexicographical approach, giving the highest hierarchy to binary comparison method and then to MO evaluation goals in some predefined order. In this PhD thesis it is suggested to use the following lexicographical order: 1) binary comparison method, 2) convergence error, 3) uniformity and 4) spread.

Detailed analysis in each of the unary quality indicators is possible; however discrimination among their measurements constitutes itself a MOP. So we propose to combine the MO quality indicators in order to discriminate among PF^*_{known}s quality. Since linear combination of the quality indicators could smooth differences and hide trade-offs when there is no an absolute winner-MOEA in all indicators, we propose a lexicographic combination of their preference according to expectations of the user, hence providing a suitable framework of analysis from the point of view of the user. It is easy to see that the MOEA with the best performance according to the defined order of the MO quality indicators will have the highest ranking value.

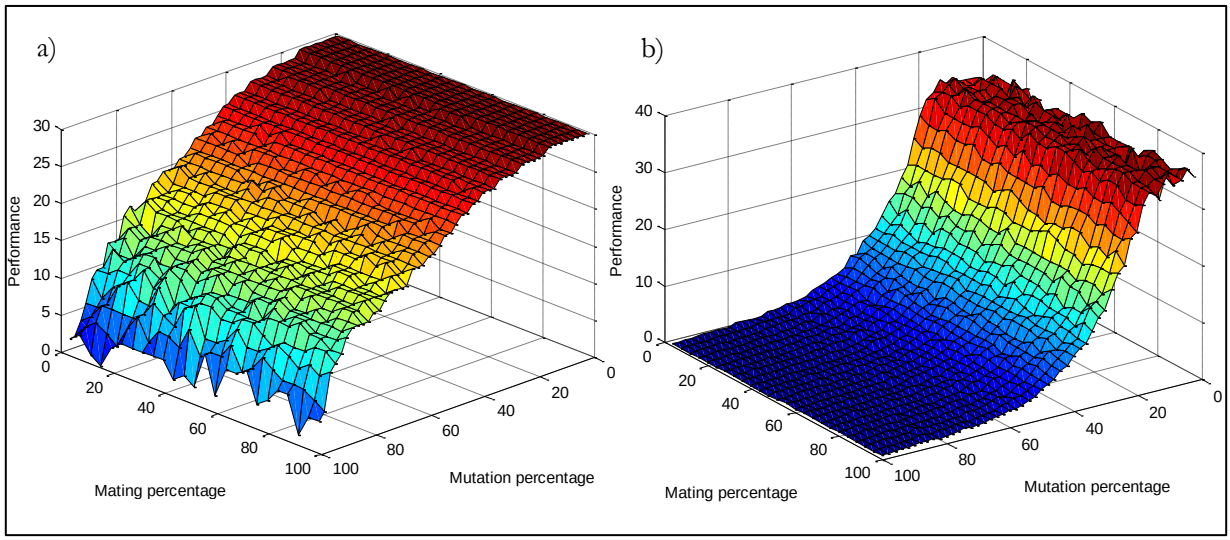

Figure 4.15 NSGA-II performance measured using the proposed methodology to compare stochastic multi-objective optimizers, mutation and mating percentages are varied from 1% to 99%. a) Performance achieved using only steps 1 and 2, b) Performance achieved using the five steps.

This suggested methodology is clearly statistically confidence and compliant with dominance relations between non-dominated sets, showing a superior inference power. In order to test correctness of the proposed methodology, it is applied to measure NSGA-II performance over Kursawe's MOP [9]. The performance landscape generated by using different combinations of mating and mutation is showed in Figure 4.15, for each mating-mutation combination 50 runs of NSGA-II were executed, a higher value in the graphic indicates better performance. Figure 4.15 a shows the performance achieved using only steps 1 and 2, it is evident how a binary comparison method $\lhd$-compatible and $\lhd$-complete is limited in its inference power, since results produced with mutation between 1% and 15% are indistinguishable among them. On other hand, Figure 4.15 b shows the performance achieved using the five steps, amplifying

inference power and suggesting than the best NSGA-II outcomes in Kursawe's MOP can be achieved within the intersection of mating from 60% to 95% and mutation from 7% to 11%.

Chapter 5 Approaching the Pareto Front of Theoretical MOPs

[*Everything is theoretically impossible, until it is done…*]

Robert Heinlein

5.1 INTRODUCTION

In order to broaden RankMOEA's assessment several theoretical MOPs suggested in EMO literature were included to test RankMOEA's performance. Due to the results presented in Section 3.3, besides RankMOEA, only NSGA-II and SPEA2 were included in this comparison. The same algorithmic specifications described in Section 3.3 were used. A precision of 0.0001 was required for each variable in the phenotype, the size of the binary chromosomes varies according to MOP specifications, a population and $P^*_{known}(t)$ size of 100 individuals with 100,000 objective function evaluations were considered. The three algorithms were run 50 times with different mating and mutation rates combination, the comparison methodology described in Section 4.5. was utilized to show MOEAs' performance.

The mating rates used for NSGA-II, SPEA2 and RankMOEA were: 50%, 60%, 70%, 80% and 90%. The mutation rates used for NSGA-II and SPEA2 were 1%, 2%, 3%, 4%, 5% and 6%, whereas for RankMOEA p_{min} was set to 0% and p_{max} to 6%. In the following sections test results over different benchmarks of unconstrained numeric MOPs are presented.

5.2 BASIC BENCHMARK OF MOPS

Fonseca2

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\right] \\ & f_1(\boldsymbol{x}) = 1 - e^{-\sum_{i=1}^{n}\left(x - \frac{1}{\sqrt{n}}\right)^2} \\ & f_2(\boldsymbol{x}) = 1 - e^{-\sum_{i=1}^{n}\left(x_i + \frac{1}{\sqrt{n}}\right)^2} \\ & x_i \in [-4,4];\ i = 1,2,3 \end{aligned} \tag{5.1}$$

Fonseca2 [9] has a connected and symmetric PS^* and a connected and concave PF^* curve, which are plotted in Figure 5.1 a and b. The performance of the three MOEAs in Fonseca2 using the comparison process is plotted in Figure 5.1 c, d and e. The NSGA-II and SPEA2's performance seems to be highly affected by the mutation rate. The best PF^*_{known}s achieved by SPEA2 and RankMOEA with their best mutation-mating configuration are plotted in Figure 5.1 f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is better than NSGA-II. It is possible to see how RankMOEA achieves a slightly better uniformity than SPEA2 in this simple test problem.

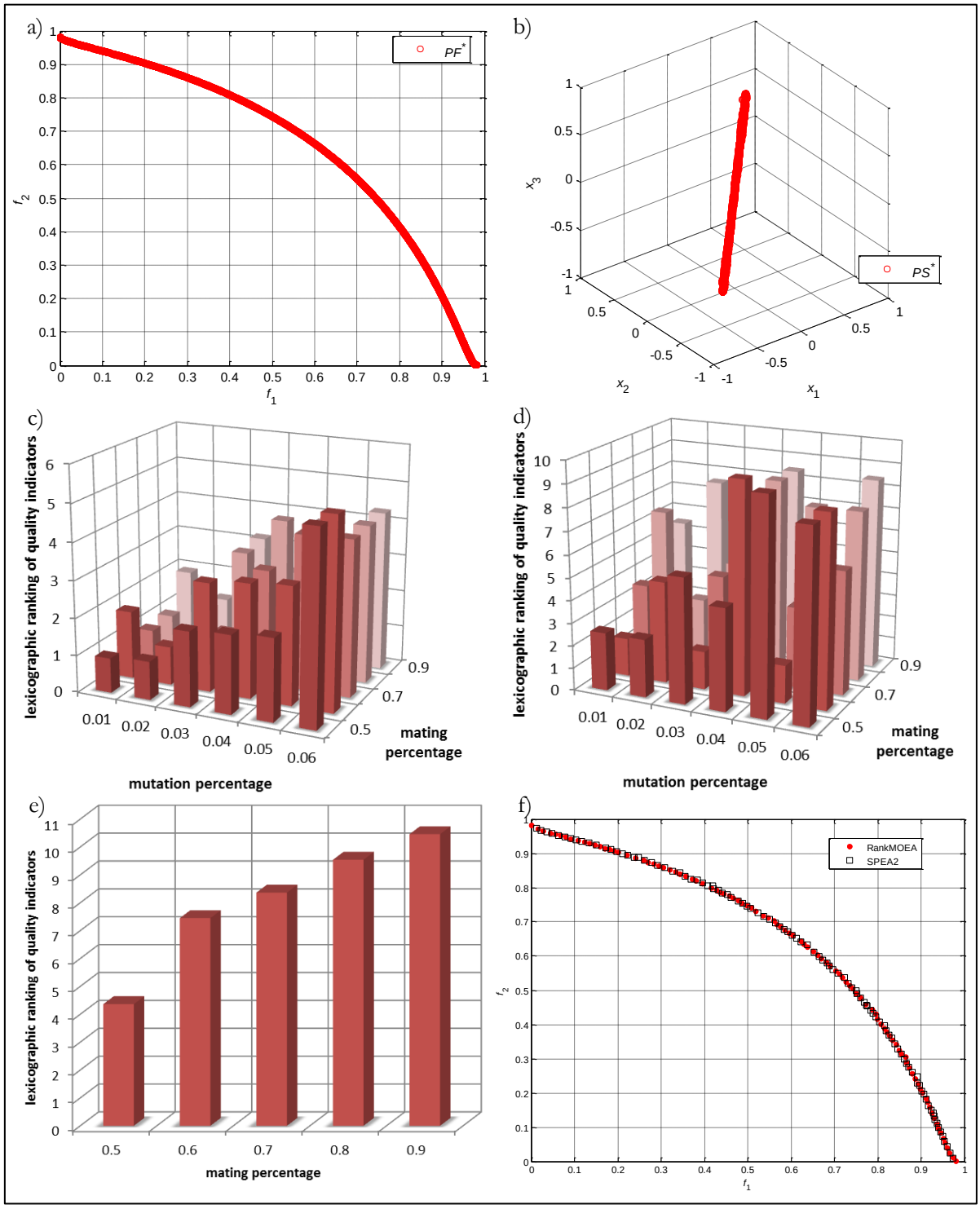

Figure 5.1 Fonseca2: a) *PF**, b) *PS**, c) NSGA-II performance, d) SPEA2 performance, e) RankMOEA performance and f) SPEA2 and RankMOEA's best achieved outcome.

Kursawe

$$\begin{gathered}\min_{\boldsymbol{x}} \qquad \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\right] \\ f_1(\boldsymbol{x}) = -\sum_{i=1}^{n-1} 10e^{-0.2\cdot\sqrt{x_i^2+x_{i+1}^2}} - 10e^{-0.2\cdot|x_n|} \\ f_2(\boldsymbol{x}) = \sum_{i=1}^{n} (|x_i|^a + 5\sin^b x_i) \\ x_i \in [-5,5];\ i = 1,2,3; a = 0.8; b = 3\end{gathered} \tag{5.2}$$

Kursawe [9] has a disconnected and symmetric PS^* and disconnected and concave PF^* curves, which are plotted in Figure 5.2 a and b. The performance of the three MOEAs in Kursawe using the comparison process is plotted in Figure 5.2 c, d and e. The NSGA-II's performance seems to be clearly affected my mutation rate, while SPEA2's performance attaints a more stable performance landscape. The best PF^*_{known}s achieved by SPEA2 and RankMOEA with their best mutation-mating configuration are plotted in Figure 5.2 f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is better than NSGA-II. SPEA2 and RankMOEA achieve similar PF^*_{known}s.

Poloni

$$\begin{gathered}\min_{\boldsymbol{x}} \qquad \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\right] \\ f_1(\boldsymbol{x}) = -1 - (A_1 - B_1)^2 - (A_2 - B_2)^2 \\ {}_2(\boldsymbol{x}) = -(x_1 + 3)^2 - (x_2 + 1)^2 \\ x_i \in [-\pi, \pi];\ i = 1,2 \\ A_1 = 0.5\sin 1 - 2\cos 1 + \sin 2 - 1.5\cos 2 \\ A_2 = 1.5\sin 1 - \cos 1 + 2\sin 2 - 0.5\cos 2 \\ B_1 = 0.5\sin x_1 - 2\cos x_1 + \sin x_2 - 1.5\cos x_2 \\ B_2 = 1.5\sin x_1 - \cos x_1 + 2\sin x_2 - 0.5\cos x_2\end{gathered} \tag{5.3}$$

Poloni [9] has a disconnected PS^* and disconnected and concave PF^* curves, which are plotted in Figure 5.3 a and b. The performance of the three MOEAs in Poloni using the comparison process is plotted in Figure 5.3 c, d and e. SPEA2 seems to be very sensitive to mutation rate changes, whereas NSGA-II is more robust to such changes achieving its best performance with medium mutation-mating rates. The best PF^*_{known}s achieved by NSGA-II and RankMOEA with their best mutation-mating configuration are plotted in Figure 5.3 f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is better than SPEA2. RankMOEA clearly outperforms NSGA-II uniformity.

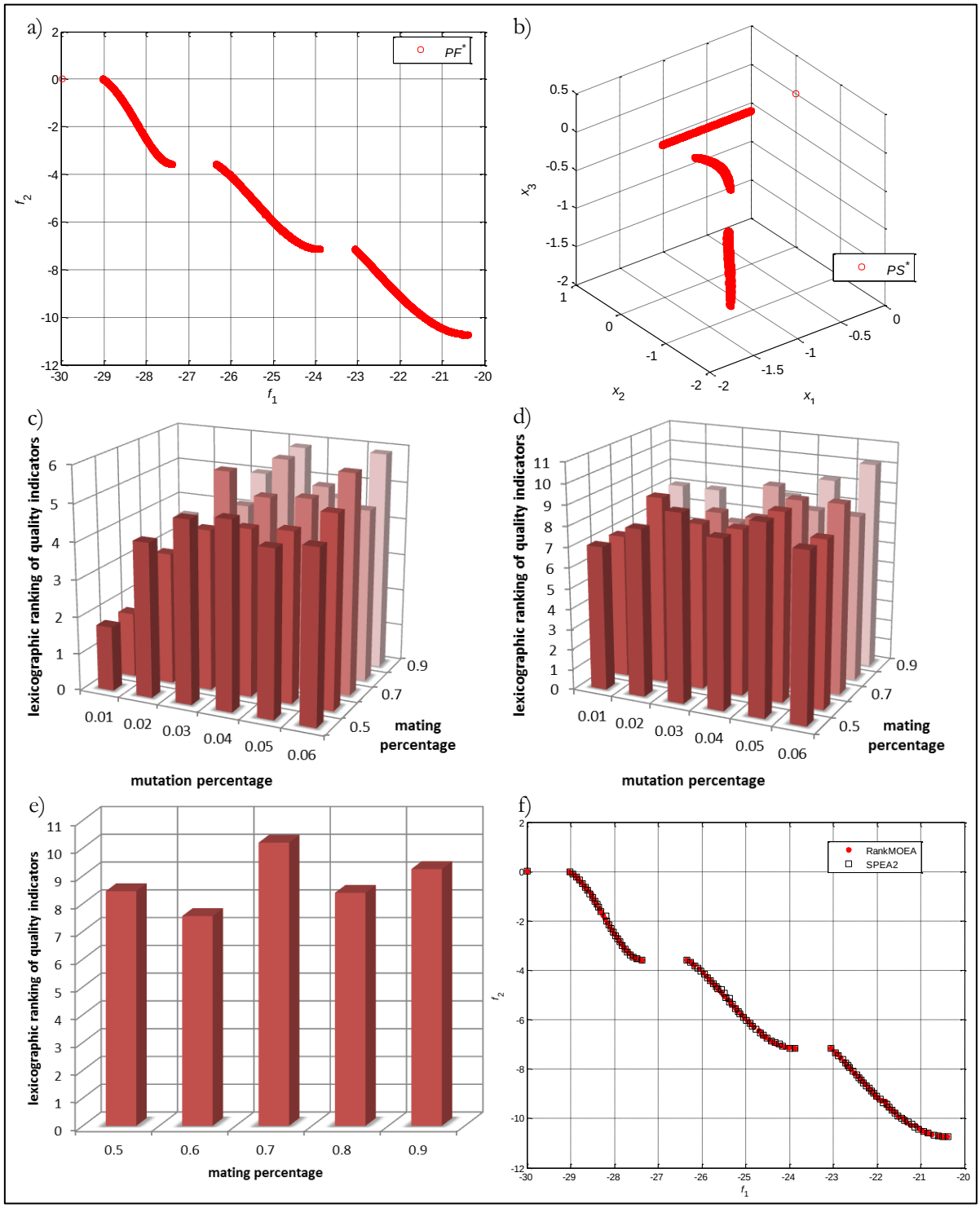

Figure 5.2 Kursawe: a) *PF**, b) *PS**, c) NSGA-II performance, d) SPEA2 performance, e) RankMOEA performance and f) SPEA2 and RankMOEA's best achieved outcome.

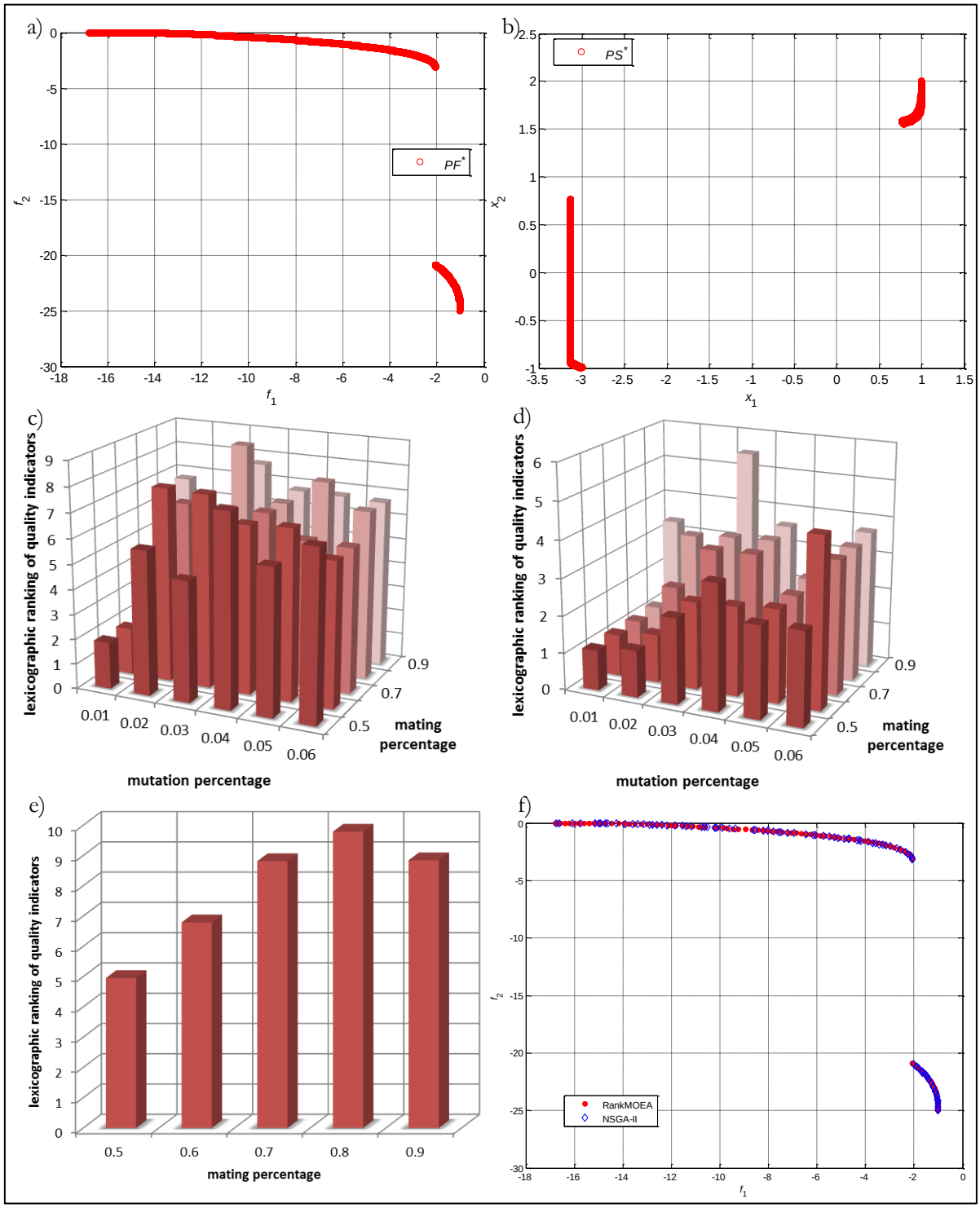

Figure 5.3 Poloni: a) *PF**, b) *PS**, c) NSGA-II performance, d) SPEA2 performance, e) RankMOEA performance and f) NSGA-II and RankMOEA's best achieved outcome.

Viennet1

$$\min_{\boldsymbol{x}} \quad \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), f_3(\boldsymbol{x})\right]$$
$$f_1(\boldsymbol{x}) = \frac{x_1^2 + x_2^2}{2} + \sin(x_1^2 + x_2^2)$$
$$f_2(\boldsymbol{x}) = \frac{(3x_1 - 2x_2 + 4)^2}{8} + \frac{(x_1 - x_2 + 1)^2}{27} + 15 \tag{5.4}$$
$$f_3(\boldsymbol{x}) = \frac{1}{x_1^2 + x_2^2 + 1} - 1.1e^{-x_1^2 - x_2^2}$$
$$x_i \in [-30,30];\ i = 1,2$$

Viennet1 [9] has a connected and symmetric PS^* and a connected and convex PF^* surface, which are plotted in Figure 5.4 a and b. The performance of the three MOEAs in Viennet1 using the comparison process is plotted in Figure 5.4 c, d and e. The NSGA-II and SPEA2's performance seems to be highly affected by mutation rate. The best PF^*_{known}s achieved by NSGA-II and RankMOEA with their best mutation-mating configuration are plotted in Figure 5.4 f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is better than SPEA2. RankMOEA clearly achieves a better spread in Λ than NSGA-II.

Viennet2

$$\min_{\boldsymbol{x}} \quad \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), f_3(\boldsymbol{x})\right]$$
$$f_1(\boldsymbol{x}) = \frac{(x_1 - 2)^2}{2} + \frac{(x_2 + 1)^2}{13} + 3$$
$$f_2(\boldsymbol{x}) = \frac{(x_1 + x_2 - 3)^2}{36} + \frac{(-x_1 + x_2 + 2)^2}{8} - 17 \tag{5.5}$$
$$f_3(\boldsymbol{x}) = \frac{(x_1 + 2x_2 - 1)^2}{175} + \frac{(2x_2 - x_1)^2}{17} - 13$$
$$x_i \in [-40,40];\ i = 1,2$$

Viennet2 [9] has a connected PS^* and a connected and concave PF^* surface, which are plotted in Figure 5.5 a and b. The performance of the three MOEAs in Viennet2 using the comparison process is plotted in Figure 5.5 c, d and e. The SPEA2's performance seems to be affected by mutation rate, whereas NSGA-II attains a more flat performance landscape. The best PF^*_{known}s achieved by SPEA2 and RankMOEA with their best mutation-mating configuration are plotted in Figure 5.5 f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is better than NSGA-II. RankMOEA

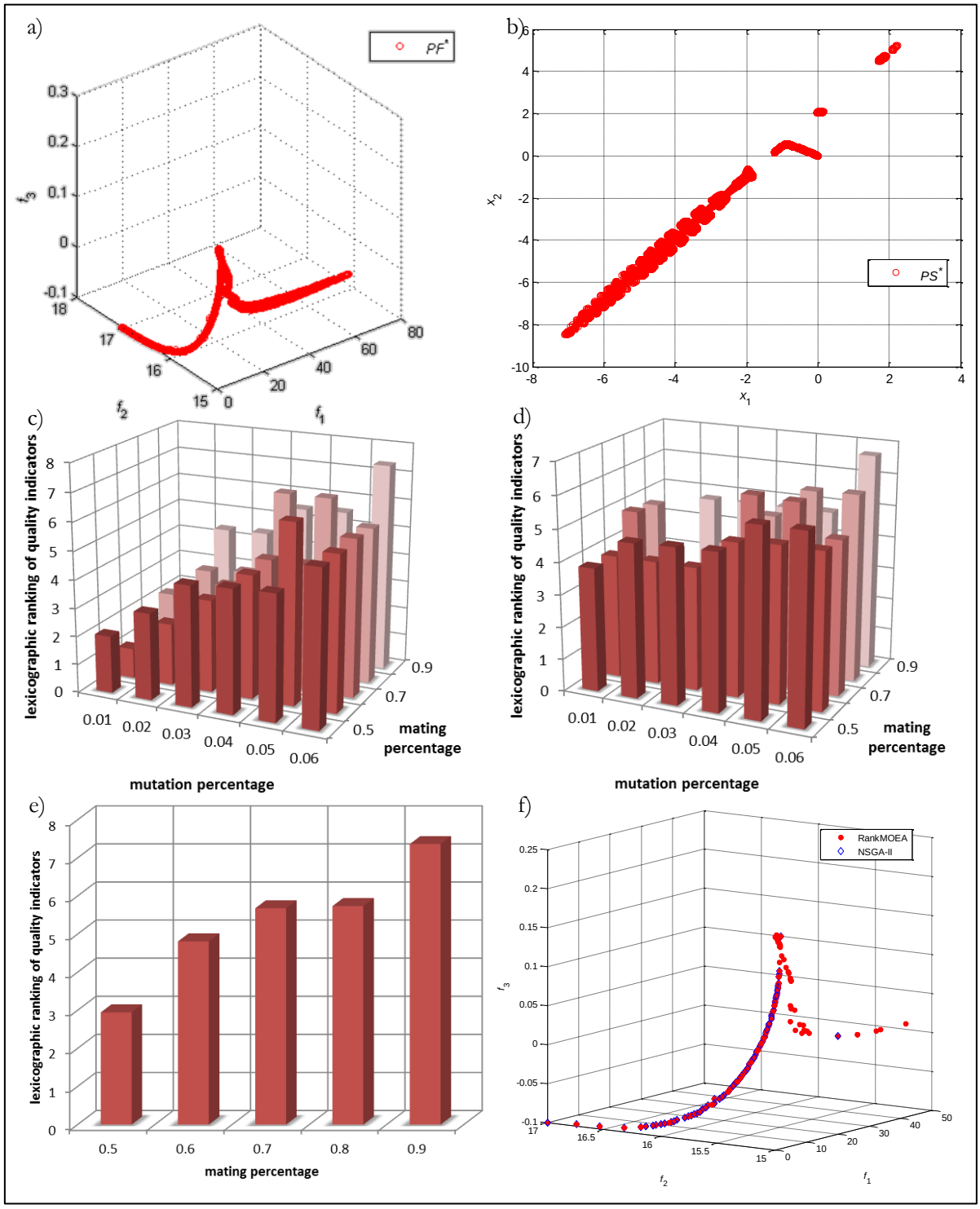

Figure 5.4 Viennet1: a) *PF**, b) *PS**, c) NSGA-II performance, d) SPEA2 performance, e) RankMOEA performance and f) NSGA-II and RankMOEA's best achieved outcome.

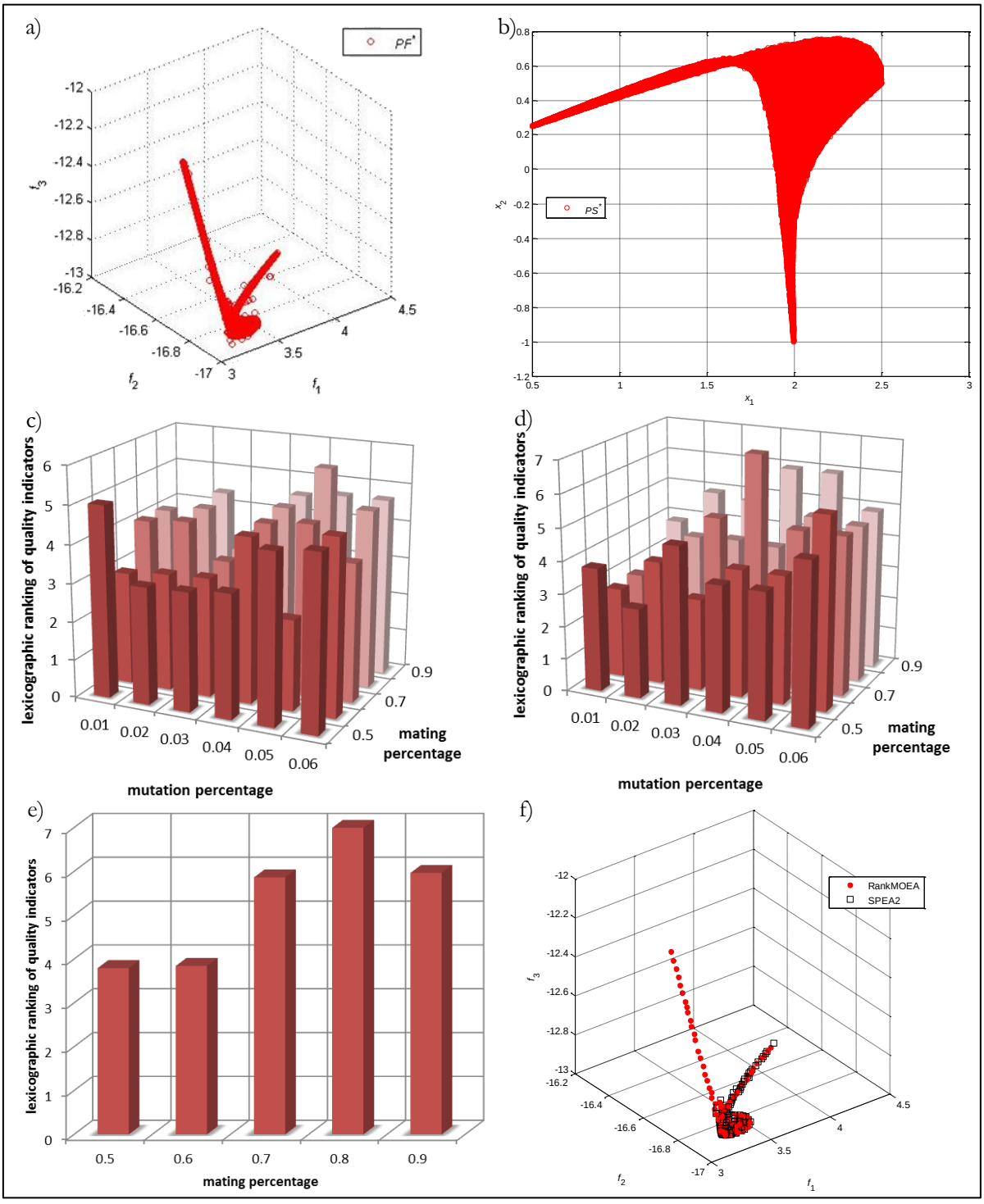

Figure 5.5 Viennet2: a) *PF**, b) *PS**, c) NSGA-II performance, d) SPEA2 performance, e) RankMOEA performance and f) SPEA2 and RankMOEA's best achieved outcome.

clearly achieves a better spread in Λ than SPEA2.

5.3 OKA's MOPs

OKA1

$$\begin{gathered}\min_{\boldsymbol{x}} \quad \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}'), f_2(\boldsymbol{x}')\right] \\ f_1(\boldsymbol{x}') = x_1' \\ f_2(\boldsymbol{x}') = \sqrt{2\pi} - \sqrt{|x_1'|} + 2\sqrt[3]{|x_2' - 3\cos x_1' - 3|} \\ x_1' = x_1 \cos {}^{\pi}/_{12} - x_2 \sin {}^{\pi}/_{12} \\ x_2' = x_1 \sin {}^{\pi}/_{12} + x_2 \cos {}^{\pi}/_{12} \\ x_1 \in \left[6\sin {}^{\pi}/_{12}, 6\sin {}^{\pi}/_{12} + 2\pi\cos {}^{\pi}/_{12}\right] \\ x_2 \in \left[-2\pi \sin {}^{\pi}/_{12}, 6\cos {}^{\pi}/_{12}\right]\end{gathered} \tag{5.6}$$

OKA1 [55] has a connected PS^* and a PF^* curve no piecewise linear in parameter space, which are plotted in Figure 5.6 a and b. The performance of the three MOEAs in OKA1 using the comparison process is plotted in Figure 5.6 c, d and e. The SPEA2's performance is clearly affected by mutation rate, whereas NSGA-II's performance is less susceptible to such variations. The best PF^*_{known}s achieved by SPEA2 and RankMOEA with their best mutation-mating configuration are plotted in Figure 5.6 f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is better than NSGA-II. RankMOEA outcome clearly outperforms SPEA2 in uniformity and spread.

OKA2

$$\begin{gathered}\min_{\boldsymbol{x}} \quad \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\right] \\ f_1(\boldsymbol{x}) = x_1 \\ f_2(\boldsymbol{x}) = 1 - \frac{(x_1 + \pi)^2}{4\pi^2} + \sqrt[3]{|x_2 - 5\cos x_1|} + \sqrt[3]{|x_3 - 5\sin x_1|} \\ x_1 \in [-\pi, \pi] \quad x_2, x_3 \in [-5,5]\end{gathered} \tag{5.7}$$

OKA2 [55] has a connected PS^* and a PF^* curve no piecewise linear in parameter space, which are plotted in Figure 5.7 a and b. The performance of the three MOEAs in OKA2 using the comparison process is plotted in Figure 5.7 c, d and e. The best PF^*_{known}s achieved by NSGA-II and RankMOEA

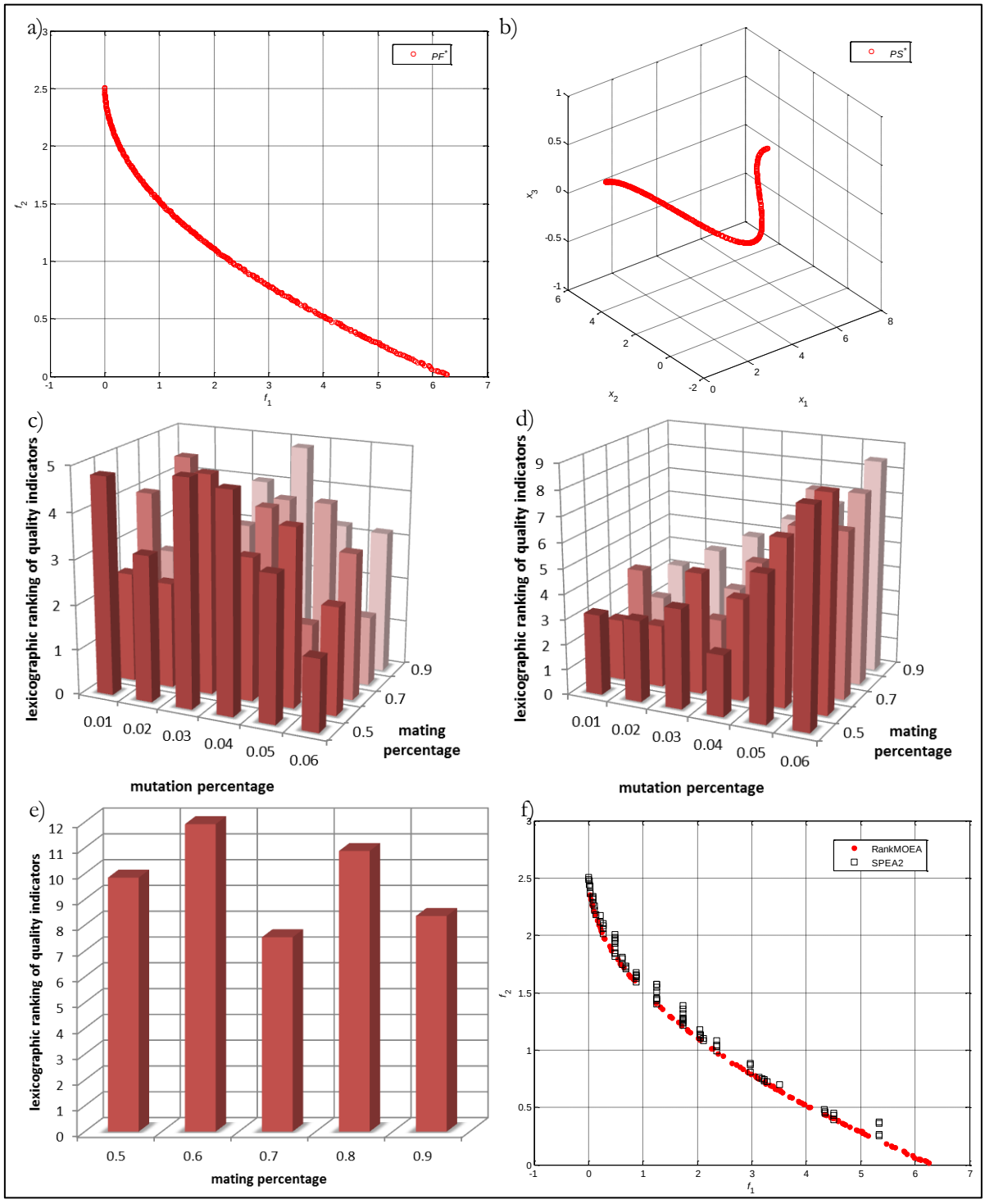

Figure 5.6 OKA1: a) *PF**, b) *PS**, c) NSGA-II performance, d) SPEA2 performance, e) RankMOEA performance and f) SPEA2 and RankMOEA's best achieved outcome.

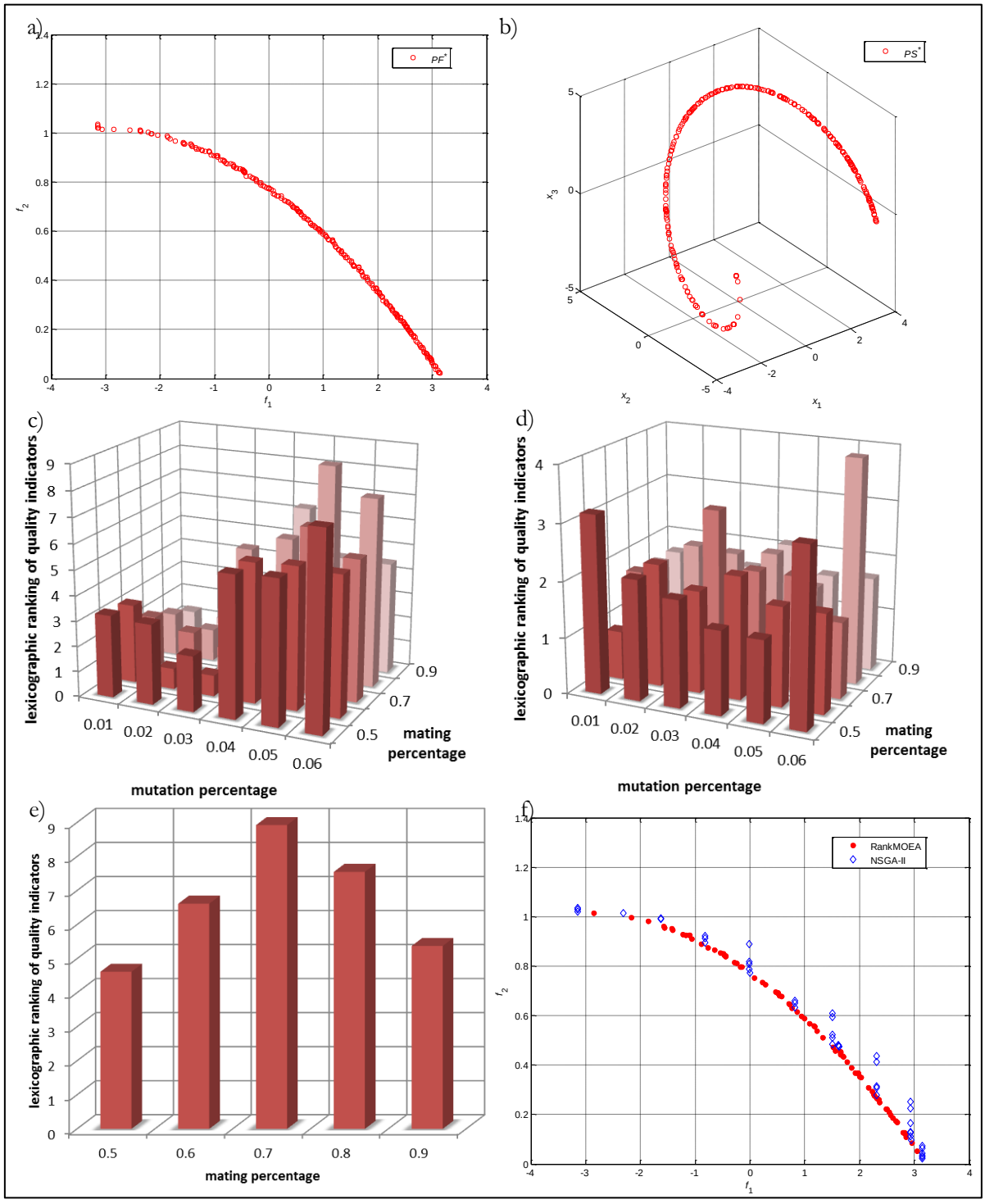

Figure 5.7 OKA2: a) *PF**, b) *PS**, c) NSGA-II performance, d) SPEA2 performance, e) RankMOEA performance and f) NSGA-II and RankMOEA's best achieved outcome.

with their best mutation-mating configuration are plotted in Figure 5.7 f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is better than SPEA2. RankMOEA outcome clearly outperforms NSGA-II in uniformity.

5.4 DTLZ's MOPs

DTLZ1

$$\begin{aligned}
\min_{\boldsymbol{x}} \quad & \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \dots, f_k(\boldsymbol{x})\right] \\
& f_1(\boldsymbol{x}) = \frac{1}{2} x_1 x_2 \cdots x_{k-1}\left(1 + g(\boldsymbol{x}')\right) \\
& f_2(\boldsymbol{x}) = \frac{1}{2} x_1 x_2 \cdots (1 - x_{k-1})\left(1 + g(\boldsymbol{x}')\right) \\
& \vdots \\
& f_{k-1}(\boldsymbol{x}) = \frac{1}{2} x_1 (1 - x_2)\left(1 + g(\boldsymbol{x}')\right) \\
& f_k(\boldsymbol{x}) = \frac{1}{2}(1 - x_1)\left(1 + g(\boldsymbol{x}')\right) \\
& g(\boldsymbol{x}') = 100\left[|\boldsymbol{x}'| + \sum_{x_i \in \boldsymbol{x}'} (x_i - 0.5)^2 - \cos\left(20\pi(x_i - 0.5)\right)\right] \\
& x_i \in [0,1];\ i = 1,2,\dots,n
\end{aligned} \tag{5.8}$$

For all DTLZ MOPs discussed in this chapter we adopted k=3. DTLZ1 [56] is a k-objective MOP with a linear PF^* surface, the functional $g(\boldsymbol{x}')$ requires $|\boldsymbol{x}'| = \varpi$ variables. It is recommended to use $\varpi = 5$, the total number of variables is $n = k + \varpi - 1$. A good sample of PF^* surface is plotted in Figure 5.8 a. The performance of the three MOEAs in DTLZ1 using the comparison process is plotted in Figure 5.8 b, c and d. The NSGA-II and SPEA2's performance is clearly affected by mutation rate. The best PF^*_{known}s achieved by SPEA2 and RankMOEA with their best mutation-mating configuration are plotted in Figure 5.8 e and f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is better than NSGA-II. SPEA2 and RankMOEA achieve similar outcomes, however in this particular problem SPEA2 achieves a slightly better spread than RankMOEA does.

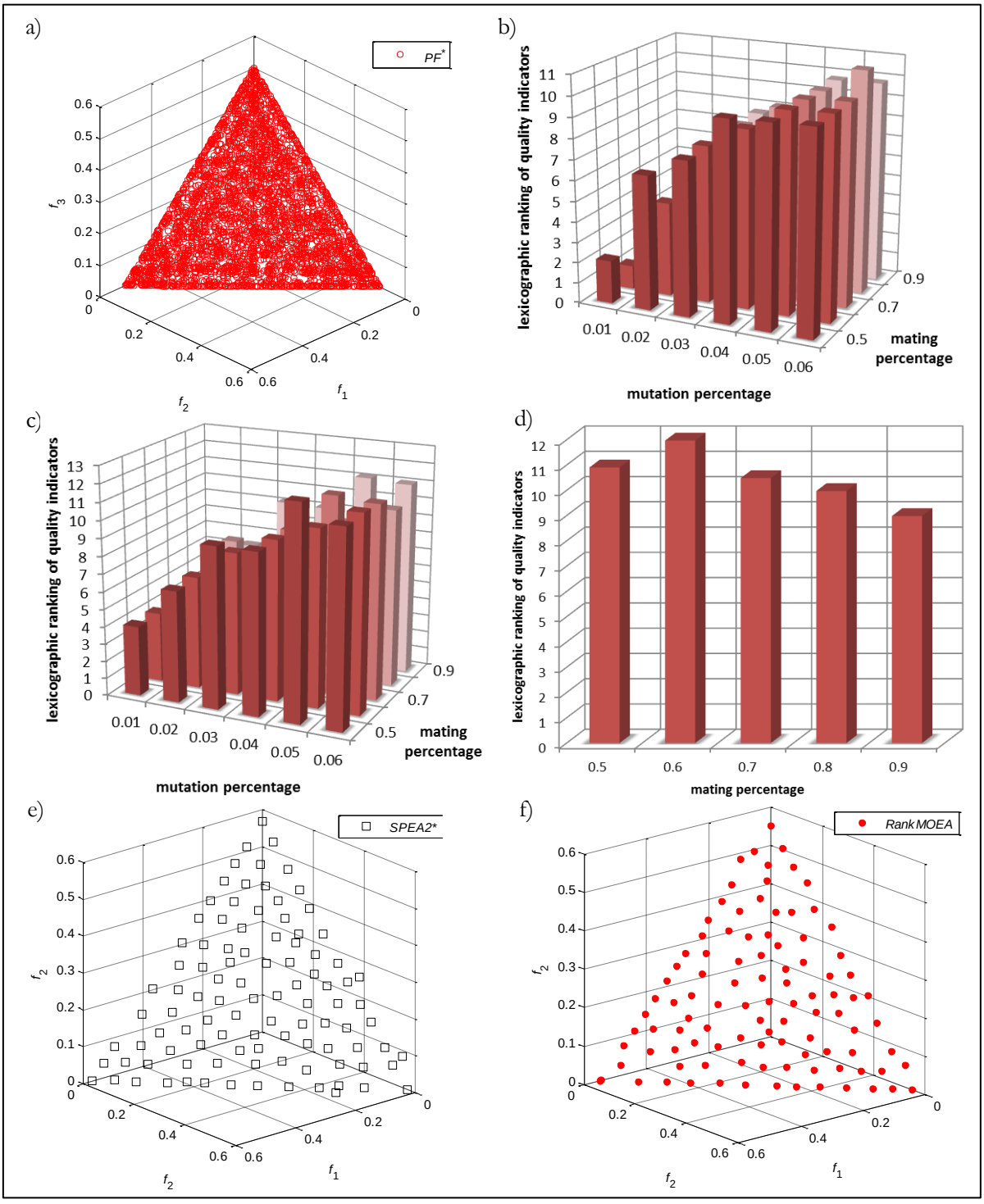

Figure 5.8 DTLZ1: a) *PF**, b) NSGA-II performance, c) SPEA2 performance, d) RankMOEA performance, e) SPEA2's best achieved outcome and f) RankMOEA's best achieved outcome.

DTLZ2

$$\begin{aligned}\min_{\boldsymbol{x}} \quad & \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \dots, f_k(\boldsymbol{x})\right] \\ & f_1(\boldsymbol{x}) = \left(1 + g(\boldsymbol{x}')\right) \cos\left(\frac{\pi x_1}{2}\right) \cdots \cos\left(\frac{\pi x_{k-1}}{2}\right) \\ & f_2(\boldsymbol{x}) = \left(1 + g(\boldsymbol{x}')\right) \cos\left(\frac{\pi x_1}{2}\right) \cdots \sin\left(\frac{\pi x_{k-1}}{2}\right) \\ & \vdots \\ & f_k(\boldsymbol{x}) = \left(1 + g(\boldsymbol{x}')\right) \sin\left(\frac{\pi x_1}{2}\right) \\ & g(\boldsymbol{x}') = \sum_{x_i \in \boldsymbol{x}'} (x_i - 0.5)^2 \\ & x_i \in [0,1];\ i = 1,2,\dots,n\end{aligned} \tag{5.9}$$

DTLZ2 [56] is a k-objective MOP with a spherical PF^* surface, the functional $g(\boldsymbol{x}')$ requires $|\boldsymbol{x}'| = \varpi$ variables. It is recommended to use $\varpi = 10$, the total number of variables is $n = k + \varpi - 1$. A good sample of PF^* surface is plotted in Figure 5.9 a. The performance of the three MOEAs in DTLZ2 using the comparison process is plotted in Figure 5.9 b, c and d. The NSGA-II, SPEA2 and RankMOEA's performance is clearly affected by mutation rate. The best PF^*_{known}s achieved by SPEA2 and RankMOEA with their best mutation-mating configuration are plotted in Figure 5.9 e and f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is better than NSGA-II. RankMOEA achieves minor error and better uniformity than SPEA2.

DTLZ3

$$\begin{aligned}\min_{\boldsymbol{x}} \quad & \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \dots, f_k(\boldsymbol{x})\right] \\ & f_1(\boldsymbol{x}) = \left(1 + g(\boldsymbol{x}')\right) \cos\left(\frac{\pi x_1}{2}\right) \cdots \cos\left(\frac{\pi x_{k-1}}{2}\right) \\ & f_2(\boldsymbol{x}) = \left(1 + g(\boldsymbol{x}')\right) \cos\left(\frac{\pi x_1}{2}\right) \cdots \sin\left(\frac{\pi x_{k-1}}{2}\right) \\ & \vdots \\ & f_k(\boldsymbol{x}) = \left(1 + g(\boldsymbol{x}')\right) \sin\left(\frac{\pi x_1}{2}\right) \\ & g(\boldsymbol{x}') = 100\left[|\boldsymbol{x}'| + \sum_{x_i \in \boldsymbol{x}'} (x_i - 0.5)^2 - \cos\big(20\pi(x_i - 0.5)\big)\right] \\ & x_i \in [0,1];\ i = 1,2,\dots,n\end{aligned} \tag{5.10}$$

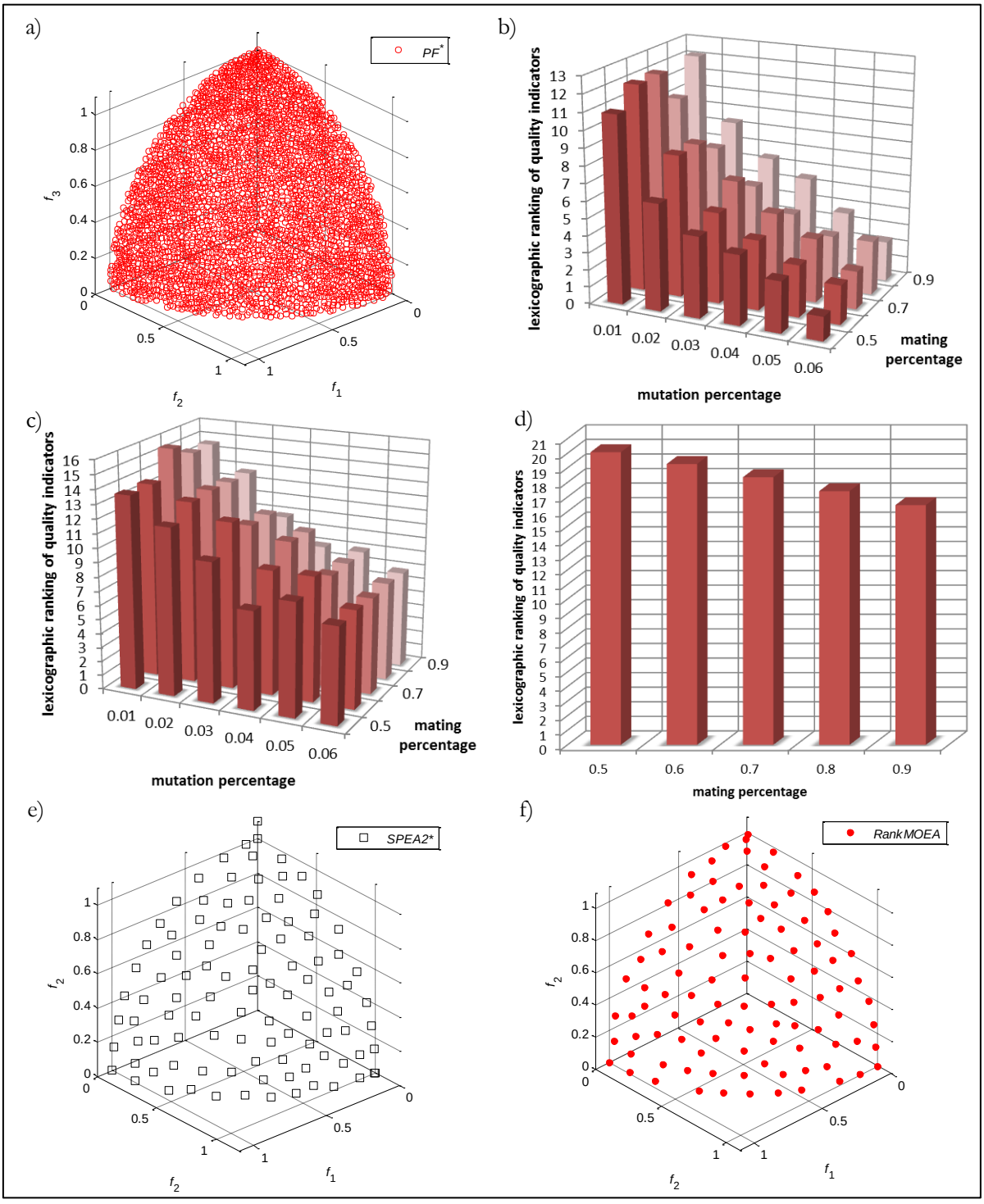

Figure 5.9 DTLZ2: a) *PF**, b) NSGA-II performance, c) SPEA2 performance, d) RankMOEA performance, e) SPEA2's best achieved outcome and f) RankMOEA's best achieved outcome.

DTLZ3 [56] is a k-objective MOP with a concave PF^* surface, the functional $g(\boldsymbol{x}')$ requires $|\boldsymbol{x}'| = \varpi$ variables. It is recommended to use $\varpi = 10$, the total number of variables is $n = k + \varpi - 1$. A good sample of PF^* surface is plotted in Figure 5.10 a. The performance of the three MOEAs in DTLZ3 using the comparison process is plotted in Figure 5.10 b, c and d. The NSGA-II and SPEA2's performance is clearly affected by mutation rate. The best PF^*_{known}s achieved by SPEA2 and RankMOEA with their best mutation-mating configuration are plotted in Figure 5.10 e and f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is better than NSGA-II. RankMOEA achieves minor error, better spread and uniformity than SPEA2.

DTLZ4

$$\begin{aligned}
\min_{\boldsymbol{x}} \quad & \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \dots, f_k(\boldsymbol{x})\right] \\
& f_1(\boldsymbol{x}) = \left(1 + g(\boldsymbol{x}')\right) \cos\left(\frac{\pi x_1^{\alpha}}{2}\right) \cdots \cos\left(\frac{\pi x_{k-1}^{\alpha}}{2}\right) \\
& f_2(\boldsymbol{x}) = \left(1 + g(\boldsymbol{x}')\right) \cos\left(\frac{\pi x_1^{\alpha}}{2}\right) \cdots \sin\left(\frac{\pi x_{k-1}^{\alpha}}{2}\right) \\
& \qquad \vdots \\
& f_k(\boldsymbol{x}) = \left(1 + g(\boldsymbol{x}')\right) \sin\left(\frac{\pi x_1^{\alpha}}{2}\right) \\
& g(\boldsymbol{x}') = \sum_{x_i \in \boldsymbol{x}'} (x_i - 0.5)^2 \\
& x_i \in [0,1];\ i = 1,2,\dots,n
\end{aligned} \tag{5.11}$$

DTLZ4 [56] is a k-objective MOP with a concave and separable PF^* surface, the functional $g(\boldsymbol{x}')$ requires $|\boldsymbol{x}'| = \varpi$ variables. It is recommended to use $\alpha = 100$ and $\varpi = 10$, the total number of variables is $n = k + \varpi - 1$. A good sample of PF^* surface is plotted in Figure 5.11 a. The performance of the three MOEAs in DTLZ4 using the comparison process is plotted in Figure 5.11 b, c and d. The NSGA-II and SPEA2's performance is clearly affected by mutation rate. The best PF^*_{known}s achieved by SPEA2 and RankMOEA with their best mutation-mating configuration are plotted in Figure 5.11 e and f, in order to facilitate visualization only such MOEAs are included in the graphic since their performance is slightly better than NSGA-II. RankMOEA achieves minor error than SPEA2.

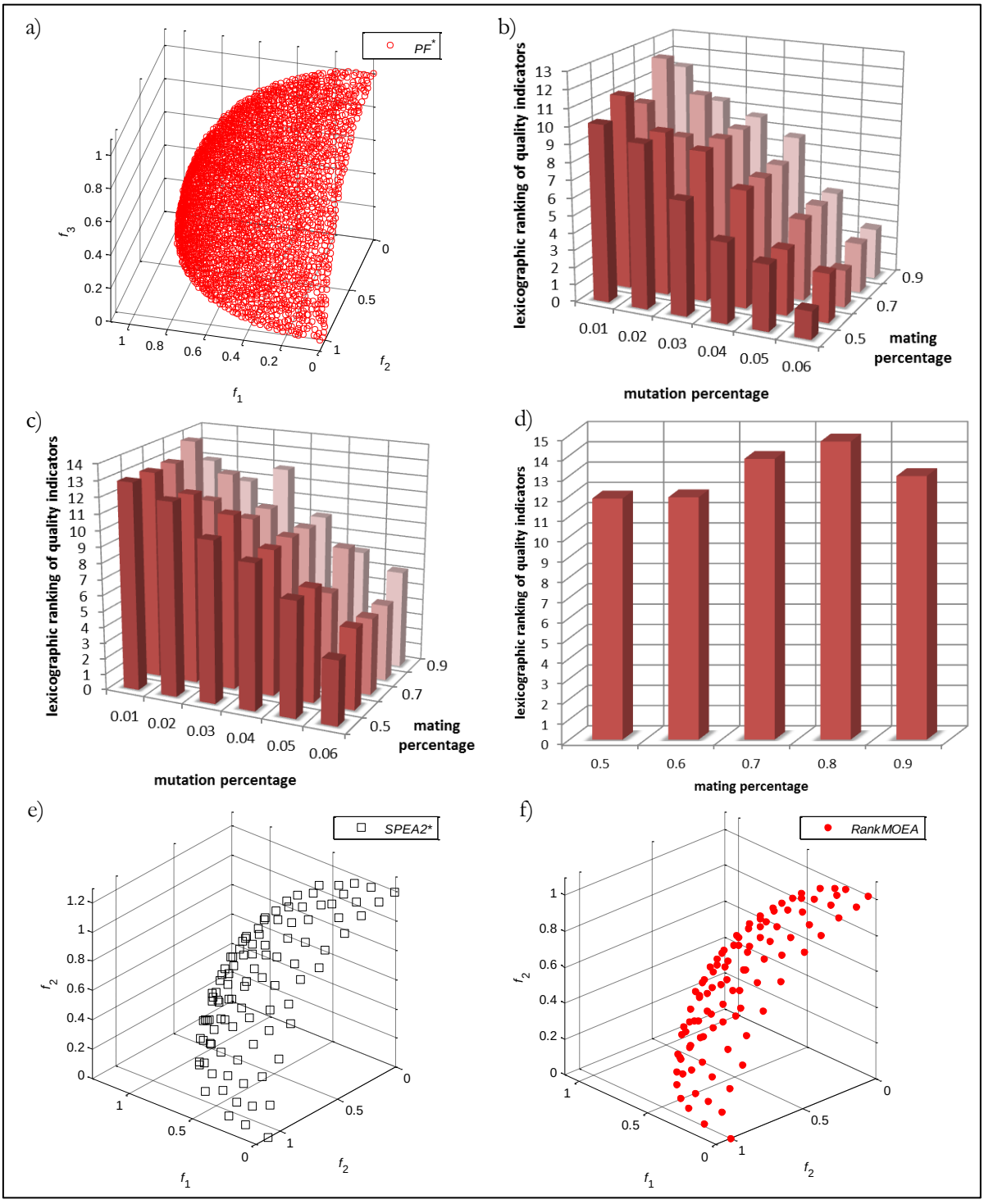

Figure 5.10 DTLZ3: a) *PF**, b) NSGA-II performance, c) SPEA2 performance, d) RankMOEA performance, e) SPEA2's best achieved outcome and f) RankMOEA's best achieved outcome.

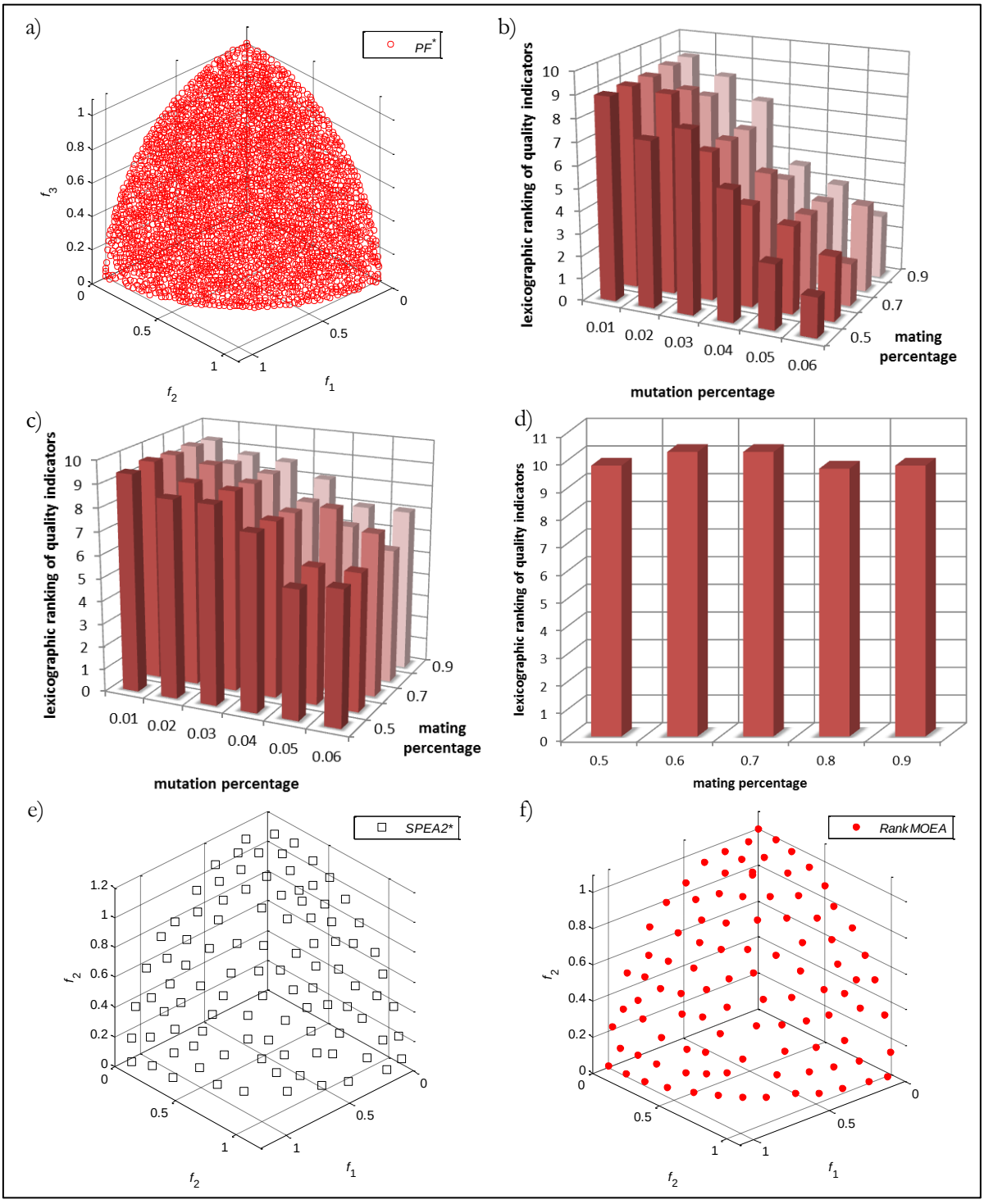

Figure 5.11 DTLZ4: a) *PF**, b) NSGA-II performance, c) SPEA2 performance, d) RankMOEA performance, e) SPEA2's best achieved outcome and f) RankMOEA's best achieved outcome.

In the experiments presented in this Chapter, NSGA-II and SPEA2 showed to be more susceptible to mutation rate changes than mating rate changes in most of the cases; this suggests that mutation rate is a key feature to maintain equilibrium between exploration and exploitation in order to be compliant with the structure of the search space. Even RankMOEA is not the absolute MOEA-winner in every problem that was presented in this Chapter (or can be conceivable), the attained results in the above subset of problems suggest that RankMOEA is more robust to deal with changes in the structure of search space of the problems. Thus, using a robust technique, as RankMOEA, could be an important advantage in the case of time-consuming scenarios where the resources needed to tune a super specialized algorithm whose performance could overcome RankMOEA's performance is not feasible.

Chapter 6 Approaching the Pareto Front of a Dynamic Principal-Agent Model

[*The eyes of the Lord preserve knowledge…*]

Proverbs 22:12a

6.1 INTRODUCTION

The Principal-Agent problem is a political science and economics well known problem which analyses a situation of asymmetric information where a risk neutral Principal delegates tasks to a risk averse Agent, i.e., it treats the difficulties that arise under asymmetric information conditions when a Principal hires an Agent, such as the problem that the two may not have the same interests, while the Principal is, presumably, hiring the Agent to pursue the interests of the former.

Asymmetric information, a specific aspect of imperfect information in markets, arises when one individual to an economic decision has different information to that of another, i.e. the Principal cannot observe the effort level that the Agent chooses, due to monitoring the Agent is too costly for the Principal.

Thus, the idea behind the Principal-Agent problem is to try to align the interests of the Agent in solidarity with those of the Principal using a compensation plan, though the existence of uncertainty in the production process makes the design of the Agent's compensation plan a non-trivial problem [57]. In theory, and assuming perfect competition, both parties to an exchange would be acting for their own interests but would also be aware of the basis on which the other was operating, the resulting exchange would benefit both parties to an equal degree (see Figure 6.1).

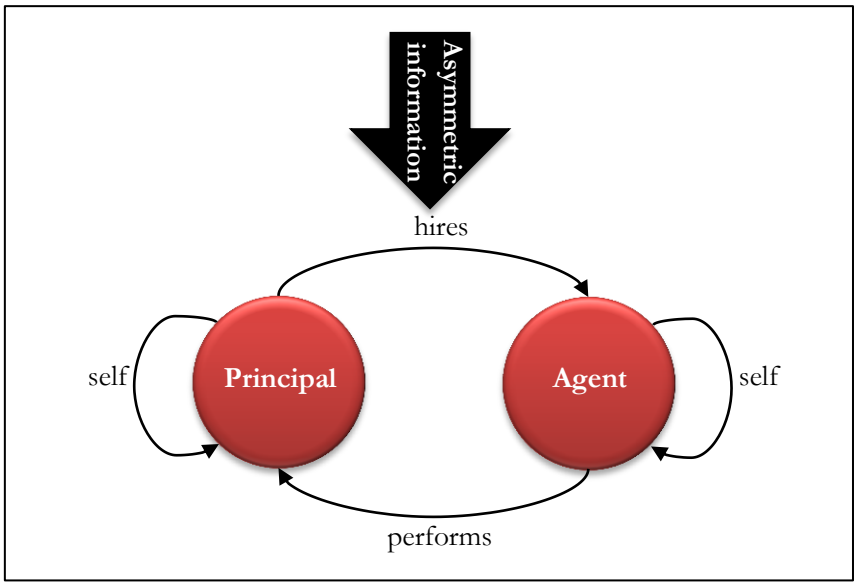

Figure 6.1 Principal-Agent's problem interaction.

A dynamic problem can be modelled when the Principal-Agent relationship is recurrent, i.e., the relationship goes on for infinite periods [58] [59]. In such context, the Agent's compensation plan has two components: present and future compensation. Both components of the Agent's compensation plan aim to link the Agent's wealth with the Principal's wealth. In the Dynamic Principal-Agent problem, the Principal maximizes his discounted expected utility subject to two fundamental constraints: the participation constraint (i.e. the contractual relationship should be accepted by the Agent), and the incentive compatibility constraint (i.e. the level of effort implemented by the Principal in every period should be chosen by the Agent given the unobservability of his effort).

6.2 A MULTI-OBJECTIVE APPROACH

The contractual arrangement between the Principal and the Agent affects how the economic surplus is divided and the sheer magnitude of such surplus. Hence, characterizing a Pareto Front where the Principal and the Agent have different levels of bargaining power is an interesting exercise to shed light into how the economic surplus is affected by those different contractual arrangements. Demougin & Helm [60] analyse this problem using a static Principal-Agent model where the Principal and the Agent are risk neutral. They found that the same set of contracts emerge by varying the Agent's reservation utility in the frame of the Principal-Agent model, by varying the discount factor in the Rubinstein game [61], or by varying the bargaining power coefficient in the Nash bargaining game. These authors obtain a Pareto Frontier that is concave, thus the previous equivalence result is not surprising; however, they find that the variation in the ratio of the bargaining power of individuals affects the outcome of the negotiation.

Given that in MO, each Pareto optimal solution represents a different compromise among objectives, finding different Pareto optimal solutions implies finding the structure of the trade-off surface involved in the problem. Thus, since the Principal-Agent model represents a situation of conflict of interests, the characterization of its PF^* will allow:

- to consider diverse contractual arrangements between the Principal and the Agent in which they have dissimilar levels of bargaining power,
- to achieve a better insight on how the creation of economic surplus is affected by the diverse contractual arrangements, and

- to obtain a better idea of how the conflict of interest and asymmetry of information between the Principal and the Agent affect the creation of economy surplus.

The difference with regard to the same exercise in a static model is that the contracts derived from the dynamic model have two components (present and future compensations), and thus the interrelation of these two components of managerial compensation with all the variables mentioned above can be analysed. Given the difficulty of obtaining analytical results with dynamic Principal-Agent models, close solutions methods are not applicable, hence it is common in the literature to numerically approximate the optimal contracts, see e.g. Wang [59] and Fernandes & Phelan [62]. Now, if we envision the dynamic Principal-Agent as a MOP in which both the Principal and the Agent's expected discounted utilities are maximized subject to the usual constraints, then the proposed RankMOEA can be used to approximate the dynamic Principal-Agent problem's PF^*.

6.3 MATHEMATICAL MODEL

In order to reconceive the Dynamic Principal-Agent model in a MO framework, two objectives are considered: maximize the Principal's discounted expected utility U, and maximize the Agent's discounted expected utility V. The Dynamic Principal-Agent model is about choosing an action plan, a compensation plan for each level of production and a future utility plan such that U and V are simultaneously maximized. It can be stated as in Equations (6.1).

$$\max_{a(V),w(y,V),\tilde{V}(y,V)} \{U, V\} \tag{6.1}$$

As in Wang [59], it is considered an infinite horizon Principal-Agent model where time is discrete and denoted by $t = 0,1,2,\ldots$, and $t = 0$ is the initial period contract. There is only one perishable consumption good, which is consumed by both individuals. The Principal and the Agent are assumed to be risk neutral and risk averse, respectively.

The Agent's expected utility at time t is $v_t\big(a(V), w(y,\bar{V})\big)$, it is assumed to be closed, strictly increasing and concave with respect to the current compensation $w(y,\bar{V})$, and strictly decreasing with respect to action $a(V)$; where $w(y,\bar{V})$ is consumption or salary at the current period, and $a(V)$ is the action chosen

by the Agent, which is not observed by the Principal. The Principal's expected utility at time t is $u_t(y, w(y, \bar{V}))$, it associates the realization of the production activity output y and the current compensation $w(y, \bar{V})$. Thus, the Principal and Agent's discounted expected utility by t can be modeled as in Equations (6.2) and (6.3).

$$U_t = \sum_{y \in Y} f(y; a(V)) u_t(y, w(y, \bar{V})) + \beta U_{t+1}\left(\tilde{V}(y, \bar{V})\right) \tag{6.2}$$

$$V_t = \sum_{y \in Y} f(y; a(V)) v_t(a(V), w(y, \bar{V})) + \beta \tilde{V}_{t+1}(y, \bar{V}) \tag{6.3}$$

where $f(y; a(V))$ is the probability function that associates action $a(V)$ and the output y, $\bar{V}$ is the Agent's reservation utility, $\tilde{V}(y, \bar{V})$ is the state variables for tomorrow on, and β is the discount factor. This Dynamic Principal-Agent model with Discrete Actions can be represented as a Bellman equation given [58] methodology. This model is subject to the participation constraint,

$$\sum_{y \in Y} f(y; a(V)) v(a(V), w(y, \bar{V})) + \beta \tilde{V}(y, \bar{V}) = \bar{V} \tag{6.4}$$

to the fact that the actions are in the space of feasible actions,

$$a(\bar{V}) \in A \tag{6.5}$$

where A is the action set, to the inability of temporal borrow,

$$0 \leq w(y, \bar{V}) \leq y \quad \forall y \in Y \tag{6.6}$$

where Y is the output set, and to the fact that the future compensation is in the feasible space.

$$\bar{V}(y, \boldsymbol{V}) \in V \quad \forall y \in Y \tag{6.7}$$

6.4 APPROXIMATING THE PRINCIPAL AGENT MODEL

6.4.A Finding the Optimal Contract: a Numerical Example

The same functional forms and parameters of Wang [59] are used. In particular, the Agent's expected utility function is assumed to be exponential, i.e. $v_t\big(a(V), w(y,\bar{V})\big) = -e^{\gamma\left(a(V)-\alpha w(y,\bar{V})\right)}$ because the Agent is risk averse, where $\gamma > 0$ is the coefficient of absolute risk aversion and $\alpha > 0$ measures the relative cost for the Agent of exercising a unit of effort. On the other hand, the Principal's expected utility is $u_t\big(y, w(y,\bar{V})\big) = y - w(y,\bar{V})$ because risk neutral is assumed.

For the standard model $\gamma = \alpha = 1$ and two feasible action levels $A = \{a_L = 0.1, a_H = 0.2\}$ are assumed, i.e., the Agent can choose either to shirk or to work. Hence, $a_H > a_L$ indicates that shirking is less costly than working. Also, it is assumed that there are two levels of output: low or high $Y = \{y_L = 0.4, y_H = 0.8\}$, and the probability function that associates effort and output is defined as in Equation (6.8).

$$\begin{aligned} f(y_L; a_L) &= f(y_H; a_H) = 2/3 \\ f(y_L; a_H) &= f(y_H; a_L) = 1/3 \end{aligned} \tag{6.8}$$

These probabilities capture the idea that the more diligently the Agent works, the greater the likelihood of the realization of the high output level. Finally, the Principal and the Agent's common discount factor was set to $\beta = 0.96$.

The numerical solution of the Bellman equation is $\{U, V, \hat{a}(V), w(y_H,\bar{V}), w(y_L,\bar{V}), \tilde{V}(y_H,\bar{V}), \tilde{V}(y_L,\bar{V})\}$ where $\hat{a}(V)$ is the optimal action. Given a finite horizon, the chromosome of the individuals in the population is characterized by 3 substrings of length N, where N is the number of periods of time an individual lives, i.e. the length of each chromosome is $3N$. The first substring indicates the history of actions of the individual, the second and third one show the history of compensations conditional on a high or low output level respectively. Therefore, the phenotype of an individual is defined as in Equation (6.9).

$$[a_1(V), a_2(V), \dots, a_N(V), w_1(y_H, \bar{V}), w_2(y_H, \bar{V}), \dots, w_N(y_H, \bar{V}), w_1(y_L, \bar{V}), w_2(y_L, \bar{V}), \dots, w_N(y_L, \bar{V}) \quad (6.9)$$

In order to compute U and V a backward induction must be used [63]. The number of periods in the Agent's life-span was set as $N = 70$.

6.4.B Experimental Results

In this test, besides NSGA-II, SPEA2 and RankMOEA, MOGA was included as an inferior bound. The same algorithmic specifications described in Section 3.3 were used. A precision of 0.0001 was required for each variable in the phenotype, thus binary chromosomes of 1820 bits were used, a population and $P^*_{known}(t)$ size of 200 individuals with 100,000 objective function evaluations were considered. The four algorithms were run 50 times with different mating and mutation rates combination, the comparison methodology described in Section 4.5. was utilized to show MOEAs' performance.

The mating rates used for MOGA, NSGA-II, SPEA2 and RankMOEA were: 40%, 50%, 60%, 70%, 80% and 90%. The mutation rates used for MOGA, NSGA-II and SPEA2 were 1%, 2%, 3%, 4%, 5%, 6%, 7% and 8%, whereas for RankMOEA p_{min} was set to 0% and p_{max} to 8%. For the four algorithms, constraints were handled with the idea of superiority of feasible points proposed in [64]. The performance of the four MOEAs in the Dynamic Principal-Agent model with Discrete Actions using the comparison process is plotted in Figure 6.2.

MOEA's configuration with performance ranking values lower than 2 did not achieved feasible solutions by violating some constraints. MOGA shows a bad performance, since only four configurations achieved feasible solutions but with very low ranking value, i.e. poor convergence and spread, besides mutation percentage seems to affect the performance of MOGA in an erratic way.

Mutation percentage seems to have an important role in the performance of NSGA-II and SPEA2, since lower mutations rates allow to achieve a better performance. In NSGA-II higher values of mating rate seem to offer a better PF^*_{known}, while in SPEA2, medium values of mating rate subjugated to a low mutation rate is clearly a key to achieve better PF^*_{known}s. Both algorithms have analogous average behavior over all the combinations of mutation and mating rates. About RankMOEA's performance, lower values of mating rates allow to achieve a better performance, even better than those obtained by

NSGA-II and SPEA2. A remarkable fact of RankMOEA's performance contrary to the other three MOEAs, is that it always achieves feasible solutions. Even worst approximations of RankMOEA are comparable to best approximations of NSGA-II and SPEA2.

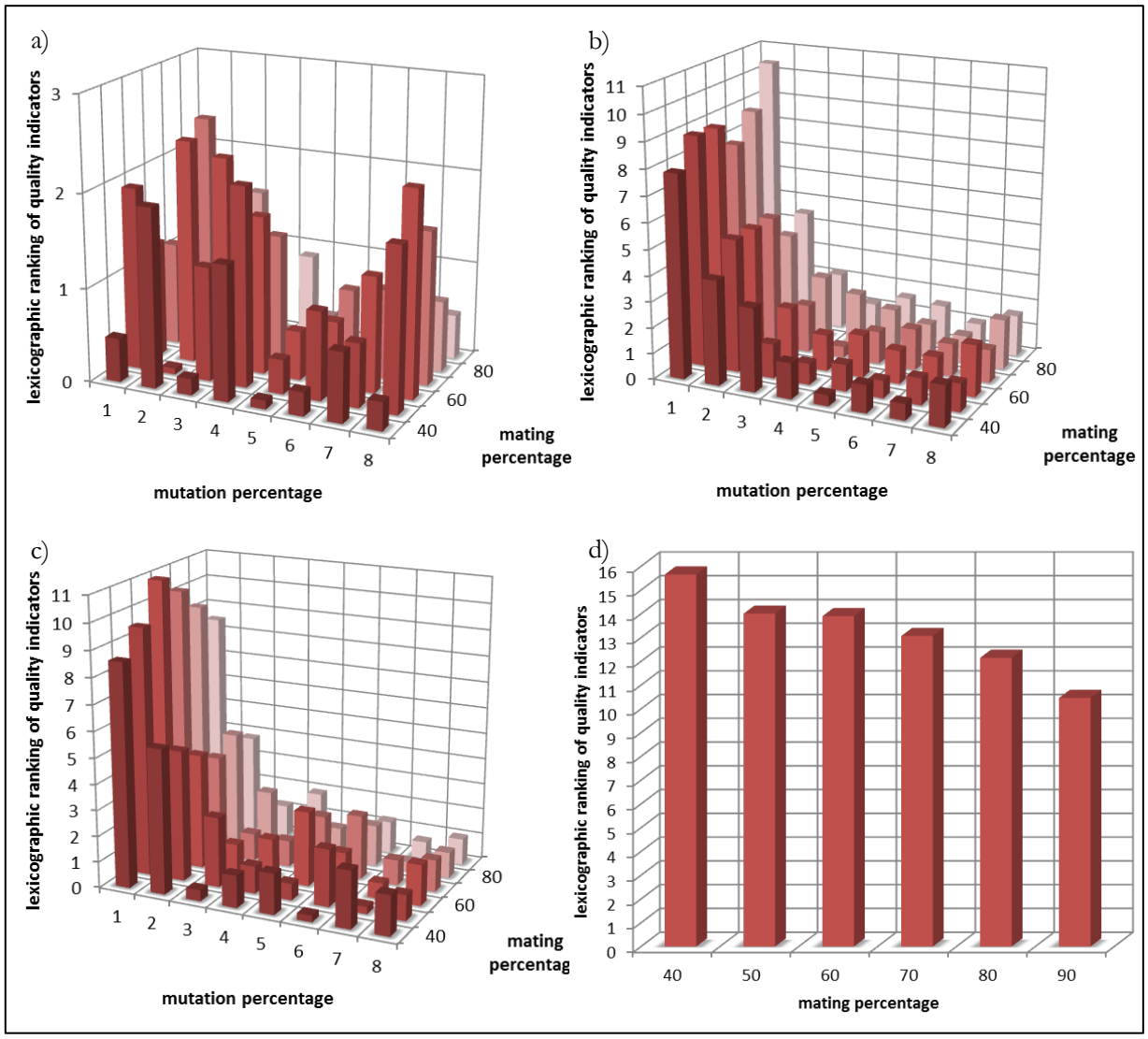

Figure 6.2 Performance in Principal-Agent model: a) MOGA, b) NSGA-II, c) SPEA2 and d) RankMOEA.

In order to have a better idea of MOEAs performance, the best PF^{*}_{known} achieved by every MOEA with its best mutation-mating configuration is plotted in Figure 6.3. The success of the proposed comparison process is confirmed by the correct classification of the quality of the achieved MOEA's outcomes. RankMOEA clearly enhances the convergence and spread achieved by MOGA, NSGA-II and SPEA2.

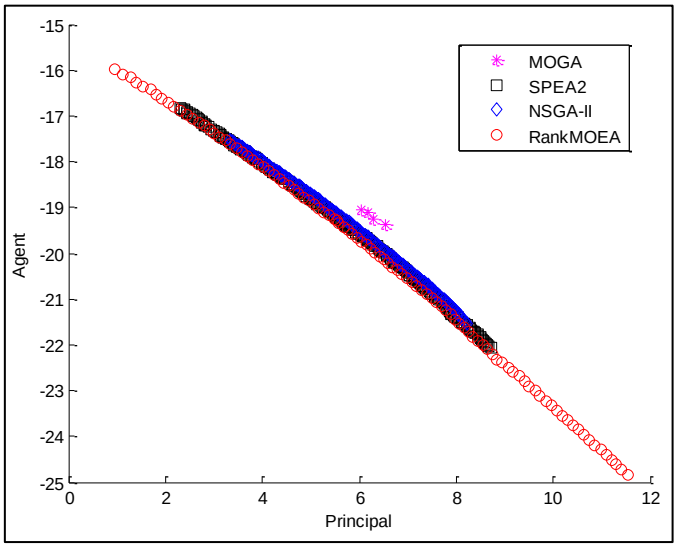

Figure 6.3 Best Pareto Front approximations achieved by the best configuration of every MOEA tested in the Dynamic Principal-Agent model with Discrete Actions.

6.4.C Analysis of the Achieved Approximation to the Pareto Front

As a result, a concave PF^* is numerically approximated, which is consequence of the information asymmetry between the Principal and the Agent, see Figure 6.3. As contracts vary in the trade-off surface towards those that are more advantageous to the Agent, it is observed the prevalence of compensation plans in which the Principal assumes most of the risk of the productive activity. When the Principal and the Agent are more patient, both obtain higher values of their discounted expected utilities, which generates a higher level of economic surplus. The Agent faces lower variability in future compensation when it is costlier for him to exert an additional effort unit. In Figure 6.4 the current compensation schedules of the most advantageous contract for the Principal (PC) and the most advantageous contract for the Agent (AC) can be observed over the periods of time. Low and high salaries of AC are higher than those of PC, moreover, in most of the cases the low level of the salary for AC is higher than the high level of the salary for PC. Note that the low salary schedules of these two contracts do not vary, i.e., both PC and AC provide incentives to the Agent through variability in the high levels of salary.

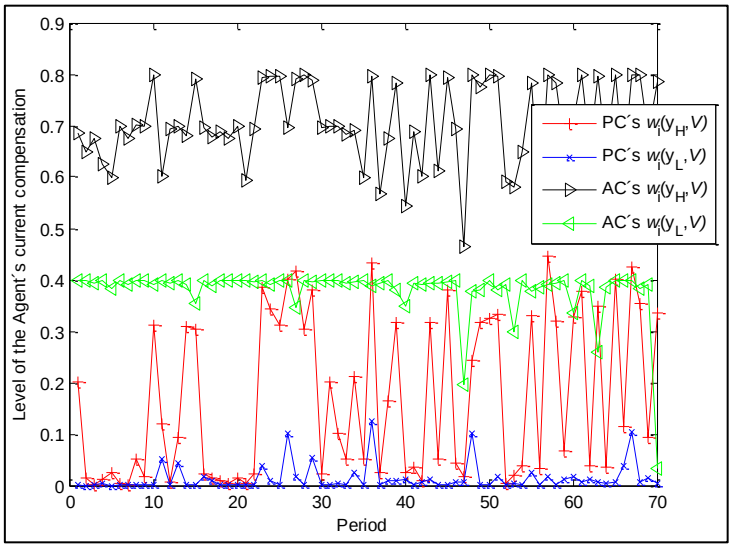

Figure 6.4 Agent's current compensation for PC and AC.

Chapter 7 Conclusions and Future Work

[*The beginning of wisdom is this: Get wisdom. Though it cost all you have, get understanding. Cherish her, and she will exalt you; embrace her, and she will honour you. She will give you a garland to grace your head and present you with a glorious crown…*]

Proverbs 4:7-9

7.1 CONCLUSIONS

In this thesis we have proposed a new efficient and effective MOEA called RankMOEA, which was designed using a minimum spanning tree niching and a ranking-mutation procedure. The new diversity-preservation mechanism involved does not need extra parameters to work and is compliant with the structure of the search space. RankMOEA outperforms traditional diversity-preservation mechanisms under spread-hardness situations, showing good spread and lower convergence error compared with other state of the art MOEAs. RankMOEA was tested with benchmarks of theoretical MOPs observing in most of the cases an outstanding performance.

An empirical taxonomy framework of quality indicators based on MO evaluation goals accuracy, including important features as computation complexity, monotony, relative, sensitivity to shape and position is presented. Table 4.4 is the result of such analysis (discussed in Chapter 4), attempting to be a helpful guide for EMO researchers at the time of choosing suitable quality indicators according to experimental goals, since the wide variety of published quality indicators. Besides, a new quality indicator to measure spread called Average Spread of the Found Pareto Front was proposed; overcoming general spread indicators drawbacks and offering a more accurate assessment by being not sensible to uniformity, showing a robust behaviour.

Within the scope that a unique indicator cannot completely describe the Pareto Front quality and that besides outperformance relations, MO evaluation goals may be helpful to untie non-dominated sets, an alternative methodology to compare performance of stochastic multi-criterion optimizers statistically confident and compliant with dominance relations between non-dominated sets was proposed and tested. Such methodology is accurate, reliable, consistent and adjustable with regard to the included lexicographical order.

In addition, RankMOEA is applied to approximate the Pareto Front of the Dynamic Principal-Agent model with Discrete Actions. The results achieved with RankMOEA show better spread and minor error than those obtained by some well-known MOEAs, allowing to perform better economic analysis by characterizing contracts in the trade-off surface. The achieved approximation of the Pareto Front allows to observe different compensation plans at different levels of bargaining power of the Agent and the Principal, and how the different contractual arrangements affect the generation of economic surplus.

7.2 FUTURE WORK

RankMOEA was designed with mechanisms that are compliant with search space structure, in spite of that more tests about its robustness and how it can be affected by MOPs with many objectives could be performed in the future. Also a combination of RankMOEA with an objective reduction technique could provide a good solution to MO with many objectives.

Parallelization of RankMOEA using GPGPU maybe an interesting future work in order to execute demanding computing tasks faster, thus achieving good results in shorter time.

RankMOEA was also tested within an autonomous robot navigation system and within a feature selection procedure in data mining; even though these works were not included in this document. The application of RankMOEA to the solution of other real MOPs is likely to produce significant improved results

Bibliography

[1] Koopmans , T.C. Analysis of Production as an Efficient Combination of Activities. *In T.C. Koopmans, editor, Activity Analysis of Production and Allocation*, no. 13 in Cowles Commission Monographs, pp. 33-97. John Wiley & Sons, New York, 1951.

[2] Kuhn, H.W. and A.W. Tucker. *Non-Linear Programming*, in 2nd Berkely Symp. on Mathematical Statistics and Probability (Ed. Neyman), University of California Press, Berkeley, 1951.

[3] Edgeworth, F.Y. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*, London: C. Kegan Paul and Co., 1881.

[4] Pareto, V. *Cours d'Economie Politique*. Lausanne, Switzerland: F. Rouge and Co., 1896.

[5] Goldberg, D.E. *Genetic Algorithms in Search*, Optimization and Machine Learning. Addison-Wesley Publishing Company, Reading, Massachusetts, 1989.

[6] H.J. Holland, Adaptation in Natural and Artificial Systems, *Ann arbor, MI: Univ. of Michigan Press*, 1975.

[7] K. Miettinen, *Nonlinear Multiobjective Optimization*. New York, NY, USA: Kluwer Academic Publishers, 1999.

[8] R. S. Rosenberg. Simulation of genetic populations with biochemical properties. PhD thesis, University of Michigan, Ann Arbor, Michigan, USA, 1967.

[9] C.A. Coello, D.A. Van Veldhuizen and G.B. Lamont, *Evolutionary Algorithms for Solving Multi-Objective Problems*. New York, NY, USA: Kluwer Academic Publishers, 2007.

[10] K. Deb and S. Jain, "Running Performance Metrics for Evolutionary Multi-Objective Optimization," Genetic Algorithms Laboratory, Department of Mechanical Engineering, Indian Institute of Technology, Kanpur, India, Tech. Rep. KanGAL 2002004, 2002.

[11] K. Deb, Multi-Objective Optimization Using Evolutionary Algorithms. New York, NY, USA: John Wiley & Sons, 2001.

[12] E. Zitzler, L. Thiele, M. Laumanns, C.M. Fonseca and V. Grunert, "Performance Assessment of Multiobjective Optimizers: An Analysis and Review," *IEEE Transactions on Evolutionary Computation*, Vol. 7, no. 2, pp. 117-132, April, 2003.

[13] D.V. Veldhuizen and G.B. Lamont, "On Measuring Multiobjective Evolutionary Algorithm Performance," *Proc. the Congress on Evolutionary Computation* (*CEC'2000*), July 16-19, 2000, La Jolla Marriott, San Diego, CA, USA. Piscataway, NJ, USA: IEEE Press, vol. 1, pp. 204-211, 2000.

[14] J. Wu and S. Azarm, "Metrics for Quality Assessment of a Multiobjective Design Optimization Solution Set," *Transactions of the ASME, J. Mechanical Design*, Vol. 123, pp. 18-25, 2001.

[15] J.D. Knowles and D.W. Corne, "On Metrics for Comparing Non-Dominated Sets," *Proc. the Congress on Evolutionary Computation* (*CEC'2002*), May 12-17, 2002, Honolulu, HI, USA. Piscataway, NJ, USA: IEEE Press, vol. 1, pp. 711-716, 2002.

[16] J.D. Knowles, "Local-Search and Hybrid Evolutionary Algorithms for Pareto Optimization," Ph.D. dissertation, Department of Computer Science, The University of Reading, Reading, Berkshire. UK, 2002.

[17] T. Okabe, Y. Jin and B. Sendhoff, "A Critical Survey of Performance Indices for Multi-Objective Optimisation," *Proc. the Congress on Evolutionary Computation* (*CEC'2003*), Dec, 8-12 2003, Canberra, Australia. Piscataway, NJ, USA: IEEE Press, vol. 2, pp. 878-885, 2003.

[18] J.D. Knowles, L. Thiele and E. Zitzler, "A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers," Computer Engineering and Comunication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, Tech. Rep. 214, 2006.

[19] Eckart Zitzler, *Evolutionary Algorithms for Multiobjective Optimization: Methods and Applications*, PhD thesis, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, November 1999.

[20] Fourman, M.P. Compaction of Symbolic Layout using Genetic Algorithms, In *Genetic Algorithms and their Applications: Proceedings of the First International Conference on Genetic Algorithms*, pages 141-153. Lawrence Erlbaum, 1985.

[21] Schaffer, D. Multiple Objective Optimization with Vector-Evaluated Genetic Algorithms. *Int' 1st Conf. on Genetic Algorithms and their Applications*, Carnegie-Mellon Univ., Pittsburgh, pp. 93-100, 1985.

[22] Fonseca, C. and P.J. Fleming. *Genetic Algorithms for Multiobjective Optimization: Formulation, discussion and Generalization*, In Stephanie Forrest editor, Proc. 5th International Conference on Genetic Algorithms, pp. 416-423, San Mateo, California, 1993.

[23] Zitzler, E. and L. Thiele. *An Evolutionary Algorithm for Multiobjective Optimization: The Strength Pareto Approach*, tech. report 43, Computer Engineering and Comunication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, May 1998.

[24] Van Veldhuizen, D. *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. PhD thesis, Department of Electrical and Computer Engineering, Graduate School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, 1999.

[25] Zitzler, E., M. Laumanns and L. Thiele. *SPEA2: Improving the Strength Pareto Evolutionary Algorithm*, Technical Report 103, Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH) Zurich, Gloriastrasse 35, CH-8092 Zurich, Switzerland, May 2001.

[26] Goldberg, D. and Richardson, J. (1987). Genetic algorithms with sharing for multimodalfunction optimization. In Grefenstette, J., editor, Proceedings of the 2nd International Conference on Genetic Algorithms and their Applications, pages 41-49, Cambridge, MA. Lawrence Erlbaum Associates.

[27] Fonseca, C. and P. Fleming. Multiobjective optimization and multiple constraint handling with evolutionary algorithms. II. Application example. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 28(1): 38-47, January 1998.

[28] Knowles, J.D. and David W. Corne. *The Pareto Archived Evolution Strategy. A New Baseline Algorithm for Multiobjective Optimisation*, In 1999 Congress on Evolutionary Computation, pp. 95-105, Washington, D.C., July 1999.

[29] Deb, K., S. Agrawal, A. Pratab, and T. Meyarivan, *A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II*, tech.repot. KanGAL-200001, Indian Institute of Technology, Kanpur, India, 2000.

[30] Carlos M. Fonseca and Peter J. Fleming, Multiobjective Optimization and Multiple Constraint Handling with Evolutionary Algorithms – Part I: A unified formulation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, 28(1):26–37, 1998.

[31] Jeffrey Horn, Nicholas Nafpliotis and David E. Goldberg, "A Niched Pareto Genetic Algorithm for Multiobjective Optimization," *Proc. 5th IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, vol. 1, pp. 82-87, Piscataway, New Jersey, June 1994.

[32] N. Srinivas and K. Deb, "Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms," *Evolutionary Computation*, Vol. 2, no. 3, pp. 221–248, June, 1994.

[33] J.R. Schott, "Fault Tolerant Design Using Single and Multicriteria Genetic Algorithm Optimization," M.S. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1995.

[34] M. Gong, L. Jiao, H. Durillo, R. Shang and B. Lu, "Performance Assessment of an Artificial Immune System Multiobjective Optimizer by Two Improved Metrics," Proc. the Genetic and Evolutionary Computation Conf. (GECCO 2005), June 25-29, 2005, Washington, D.C., USA. New York, NY, USA: ACM Press, vol. 1, pp. 373-374, 2005.

[35] H. Esbensen and E.S. Kuh, "Design Space Exploration Using the Genetic Algorithm," *IEEE International Symposium on Circuits and Systems* (*ISCAS'96*), May 12-15, 1996, Atlanta, GA, USA. Piscataway, NJ, USA: IEEE Press, vol. 4, pp. 500–503, 1996.

[36] C.M. Fonseca and P.J. Fleming, "On the Performance Assessment and Comparison of Stochastic rMultiobjective Optimizers," *Proc. Parallel Problem Solving from Nature* (*PPSN IV*), Sep 22-26, 1996, Berlin, Germany. Berlin, Germany: Lecture Notes in Computer Science, Springer-Verlag, vol. 1411, pp. 584–593, 1996.

[37] Veldhuizen, D.V., Lamont, G.B.: Evolutionary Computation and Convergence to a Pareto Front. Conf. Genetic Programming, pp. 221-228. Stanford Bookstore (1998)

[38] E. Zitzler, "Evolutionary Algorithms for Multiobjective Optimization, Methods and Applications," Ph.D. dissertation, Computer Engineering and Comunication Networks Lab (TIK), Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 1999.

[39] E. Zitzler, K. Deb and L. Thiele, "Comparison of Multiobjective Evolutionary Algorithms: Empirical Results," IEEE Transactions on Evolutionary Computation, Vol. 8, no. 2, pp. 173-195, June, 2000.

[40] E. Zitzler and L. Thiele, "Multiobjective Optimization Using Evolutionary Algorithms – A Comparative Study," *Proc. Parallel Problem Solving from Nature* (*PPSN V*), Sep 27-30, 1998, Amsterdam, Netherlands. Berlin, Germany: Lecture Notes in Computer Science, Springer-Verlag, vol. 1498, pp. 292-301, 1998.

[41] M.P. Hansen and A. Jaszkiewicz, "Evaluating the Quality of Approximations to the Non-dominated Set," Technical University of Denmark, Lyngby, Denmark, Tech. Rep. IMM-REP-1998-7, 1998.

[42] P. Czyzak and A. Jaszkiewicz, "Pareto Simulated Annealing - A Metaheuristic for Multiple-Objective Combinatorial Optimization," J. Multi-Criteria Decision Analysis, Vol. 7, pp. 34-37, 1998.

[43] Y. Leung and Y. Wang, "U-measure: a Quality Measure for Multiobjective Programming," *IEEE Transactions on Systems, Man, and Cybernetics: Part A-Systems and Humans*, Vol. 33, no. 3, pp. 337-343, May, 2003.

[44] H. Meng, X. Zhang and S. Liu, "New Quality Measures for Multiobjective Programming," *Proc. the 1st Int'l Conf. Advances in Natural Computation* (*ICNC 2005*), Aug 27-29, 2005, Changsha, China. Berlin, Germany: Lecture Notes in Computer Science, Springer-Verlag, vol. 3611, pp. 1044-1048, 2005.

[45] G. Lizárraga, A. Hernández and S. Botello "G–Metric: an M–ary Quality Indicator for the Evaluation of Non–dominated Sets," *Proc. the Genetic and Evolutionary Computation Conf.* (*GECCO 2007*), July 12-16, 2006, Atlanta, GA, USA. New York, NY, USA: ACM Press, vol. 1, pp. 665-672, 2007.

[46] M. Li and J. Zheng, "Spread Assessment for Evolutionary Multi-Objective Optimization," *Proc. Evolutionary Multi-Criterion Optimization*, *5th Int'l Conf.* (*EMO'09*), April 7-10, 2009, Nantes, France. Berlin, Germany: Lecture Notes in Computer Science, Springer-Verlag, vol. 5467, pp. 216-230, 2009.

[47] T. Bäck, Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms. Oxford, UK: Oxford University Press, 1996.

[48] J. Cervantes and C. Stephens, "Limitations of Existing Mutation Rate Heuristics and How a Rank GA Overcomes Them," IEEE Transactions on Evolutionary Computation, Vol. 3, no. 2, pp. 369-397, April, 2009.

[49] Cervantes, J. and C.R. Stephens. A Rank Proportional Generic Genetic Algorithm. Lecture Series on Computer and Computational Sciences, vol. 7, pp. 71-74. Brill Academic Publishers. 2006.

[50] Chazelle, B. 2000. *A minimum spanning tree algorithm with inverse-Ackermann type complexity*. J. ACM. 2000, vol. 47, no6, pp. 1028-1047.

[51] Zhang, Q., Zhou, A., Zhao, S.Z., Suganthan, P.N. and Liu, W, Multiobjective Optimization Test Instances for the CEC 2009 Special Session and Competition, tech.report CES-887, University of Essex and Nanyang Technological University, 2008

[52] G. Lizárraga, A. Hernández and S. Botello "On the Possibility to Create a Compatible-Complete Unary Comparison Method for Evolutionary Multiobjective Algorithms," *Proc. the Genetic and Evolutionary Computation Conf.* (*GECCO 2007*), July 12-16, 2006, Atlanta, GA, USA. New York, NY, USA: ACM Press, vol. 1, pp. 759-760, 2007.

[53] G. Lizárraga, A. Hernández and S. Botello, "A Set of Test Cases for Performance Measures in Multiobjective Optimization," 7th Mexican International Conference on Artificial Intelligence (MICAI 2008), Oct 27-31, 2008, México DF, México. Berlin, Germany: Lecture Notes in Computer Science, Springer-Verlag, vol. 5317, pp. 429-239, 2008.

[54] J.D. Knowles, "A Summary-Attainment-Surface Plotting Method for Visualizing the Performance of Stochastic Multiobjective Optimizers," 5th International Conference on Intelligent Systems

Design and Applications (ISDA'05), Sep 8-10, 2005, Washington, DC, USA. Piscataway, NJ, USA: IEEE Computer Society, pp. 552-557, 2005.

[55] T. Okabe, Y. Jin, M. Olhofer, and B. Sendhoff, "On test functions for evolutionary multi-objective optimization," in Parallel Problem Solving from Nature – PPSN VIII : 8th International Conference Proceedings. Springer, 2004, pp. 792–802.

[56] Deb, K., Thiele, L., Laumanns, M., and Zitzler E. Scalable Multi-Objective Optimization Test Problems. Proc. Congress on Evolutionary Computation 2002 (CEC '02). IEEE Press, pp. 825 – 830.

[57] Holmstrom, B. Moral hazard and observability. Bell Journal of Economics. 10. 74.91. 1979.

[58] Spear, S. and S. Srivastava. On repeated moral hazard with discounting. Review of Economic Studies. 54. 599-617. 1987.

[59] Wang, C. Incentives, CEO Compensation, and Shareholder Wealth in a Dynamic Agency Model. Journal of Economic Theory. 76. 72-105. 1997.

[60] Demougin, D., and C. Helm. Moral Hazard and Bargaining Power. German Economic Review. 7(4), 463-470. 2006.

[61] Rubinstein, A. Perfect Equilibrium in a Bargaining Model. Econometrica. 50. 97-109. 1982.

[62] Fernandes, A. and C. Phelan. A recursive formulation for repeated agency with history dependence. Journal of Economic Theory, vol. 91, pp. 223-247. 2000.

[63] Judd, K. L. 1998. Numerical methods in economics. The MIT Press.

[64] Deb, K.. An Efficient Constraint Handling Method for Genetic Algorithms. Comput. Methods Appl. Mech. Eng. Vol. 186, pp. 311-338. 2000.