On the performance degradation of dominance-based evolutionary algorithms in many-objective optimization

Thiago Santos and Ricardo H. C. Takahashi *Member, IEEE,*

***Abstract*—In the last decade, it has become apparent that the performance of Pareto-dominance based evolutionary multiobjective optimization algorithms degrades as the number of objective functions of the problem, given by n, grows. This performance degradation has been the subject of several studies in the last years, but the exact mechanism behind this phenomenon has not been fully understood yet. This paper presents an analytical study of this phenomenon under problems with continuous variables, by a simple setup of quadratic objective functions with spherical contour curves and a symmetrical arrangement of the function minima location. Within such a setup, some analytical formulae are derived to describe the probability of the optimization progress as a function of the distance λ to the exact Pareto-set. A main conclusion is stated about the nature and structure of the performance degradation phenomenon in many-objective problems: when a current solution reaches a λ that is an order of magnitude smaller than the length of the Pareto-set, the probability of finding a new point that dominates the current one is given by a power law function of λ with exponent $(n-1)$. The dimension of the space of decision variables has no influence on that exponent. Those results give support to a discussion about some general directions that are currently under consideration within the research community.**

***Index Terms*—Multiobjective optimization; evolutionary computation; many-objective problems.**

I. INTRODUCTION

The field of Evolutionary Multiobjective Optimization (EMO) experienced a fast development in the 1990's, when some algorithms reached a widespread recognition, due to their favorable properties exhibited in academic studies and their successful application to real problems [1]–[3]. Those algorithms shared a common feature: they were based on the *Pareto-dominance* selection principle, where the main selection pressure drives algorithms to the optimal solution set by assigning higher fitness to the non-dominated solutions and lower fitness to the dominated ones. However, most of the research studies and the application scenarios considered were conducted on problems with two or three objectives only.

The first mention to the difficulties that dominance-based evolutionary multiobjective algorithms would find in problems with a number of objectives higher than usual, caused by the growth of the proportion of non-dominated solutions in population, seems to appear in [1], in 1995. However, that reference did not provide a specific study about the theme. The early experimental studies that revealed some evidence that a growing number of objectives could cause difficulties in the convergence of EMO algorithms were published in 2001. An attempt to develop some formulae for the expected progress speed of mutation operations as a function of the number of objectives was presented in [4], within the framework of Evolution Strategies. However, those formulae were found to be not solvable analytically. Some numerical experiments were conducted in that work, showing that the proportion of solutions in population that are not dominated by any other one increases with the number of objectives.

In the same year, the concept of *dominance resistant solutions* was proposed in [5] in order to designate points located far away from the Pareto-set that attract the population of EMO algorithms due to the low probability of generation of other solutions that dominate them. That work showed that dominance resistant solutions arise in some classes of multiobjective optimization problems, possibly leading EMO algorithms to fail in finding the true Pareto-optimal solutions. That reference also showed that *block-separable problems*, a class of problems that are defined by the concatenation of smaller problems, will present such dominance resistant solutions. In 2002, a set of test problems for EMO algorithms was proposed in [6], including some problems that are endowed with dominance resistant solutions. That reference also indicated that the difficulty in solving those problems increases with the dimension of the objective space, in this way connecting the phenomenon of dominance resistance to the number of objectives in the problem.

In 2003, a study about the effect of increasing the number of objective functions on the performance of existing EMO algorithms was presented in [7]. In that work, it was verified empirically that for a number of objectives higher than three, the main EMO algorithms presented a significant degradation of performance. Those problems with four or more objectives were named the *many-objective* problems.

In 2007, a detailed experimental evaluation of the dominance resistance phenomenon in many-objective evolutionary algorithms was presented in [8]. Those experiments indicated that the proportion of individuals in a population that are not dominated by individuals in the next population is greater for a greater number of objectives, with this proportion growing fast along the iterations of the evolutionary algorithm. The same behavior was found in algorithms based on different variation

Thiago Santos is with the Department of Mathematics, Universidade Federal de Ouro Preto, Ouro Preto, MG, Brazil, e-mail: santostf@iceb.ufop.br.

Ricardo H. C. Takahashi is with Department of Mathematics, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil, email: taka@mat.ufmg.br.

Manuscript received xx; revised xx.

(recombination and mutation) operators.

In the years that followed, the pursuit of the reasons behind that performance degradation and the search for new approaches that could alleviate this degradation have become a major theme of research. A large number of works were developed for dealing with those issues, as reported in the recent survey papers [9], [10].

This paper presents an analytical study of the performance degradation phenomenon in Pareto-dominance based search in many-objective optimization problems. A simple and fully symmetric analytical problem composed of quadratic objective functions is studied, and analytical formulae for the probability of solution enhancement as a function of the distance from the solution to the Pareto-set are derived. The big picture that emerges from this analysis is: (i) For any number of objectives, the task of finding an estimate of a Pareto-optimal solution with an accuracy of the order of the length of the Pareto-set[1] is not a hard task. (ii) The task of finding accurate estimates of Pareto-optimal solutions becomes increasingly harder as the required accuracy increases, for any number of objectives $n \geq 2$. Specifically, the probability of success of a "small" mutation asymptotically reaches a power law relationship with the distance from the point to the Pareto-set. As the number of objectives increases the problem becomes more difficult because higher exponents in the power law, associated with greater values of n, lead to higher rates of decay in the probability of success. Those conclusions explain to a large extent the phenomenon described in [1], [4], [5], [7], [8]. A preliminary study in those directions was presented in [11], considering only the case $n = 2$.

This paper is organized as follows. The Section II states the general formulation of the multiobjective optimization problems and of the Pareto-dominance search that are considered. Section III presents the development of the analytical asymptotic formulae that describe the probability of finding a direction in which there exists a point that dominates the current solution, in a special problem that is stated in order to allow such an analysis. Those power law formulae constitute the main results in this paper. Section IV presents some simulation studies that examine the transition between the domains of validity of the formula that describes the situation "far from" the Pareto-set and the formula that describes that situation "close to" the Pareto-set. In the same section, simulations show that the same power law still holds in more general configurations of the problem. The Section V discusses in more detail the two stages in which a dominance-based algorithm converges to a Pareto-set, as described by the proposed formulae. A discussion about the consequences of the proposed tools for the analysis of the behavior of some decomposition-based algorithms that constitute a significant part of the current research efforts on many-objective optimization is presented in Section VI. The final conclusions are stated in Section VII.

It is important to mention that this work focuses on many-objective problems in the context of continuous decision variables only. Although a difficulty of the same kind arises in combinatorial problems (see for instance [12]–[14]), the analysis of such problems will require different analysis tools which are not in the scope of this study.

[1]The length of the Pareto-set is defined as the maximum distance between two points belonging to that set.

II. Multi-Objective Optimization

In a finite-dimensional continuous-variable multi-criteria decision problem, a decision variable $\mathbf{x}$ should be chosen from a set $\Omega \subseteq \mathbb{R}^N$, according to n criteria functions $f_i : \Omega \mapsto \mathbb{R}$. Let $\mathbf{x}_1 \in \Omega$ and $\mathbf{x}_2 \in \Omega$. It is assumed, by convention, that $\mathbf{x}_1$ is better than $\mathbf{x}_2$ in criterion f_i if $f_i(\mathbf{x}_1) < f_i(\mathbf{x}_2)$. As the problem deals with n different criteria, the following relational operators are defined, in order to compare vectors. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, then:

$$\begin{array}{lcl} \mathbf{u} \preceq \mathbf{v} & \Leftrightarrow & u_i \leq v_i,\ i = 1, \ldots, n \\ \mathbf{u} \neq \mathbf{v} & \Leftrightarrow & \exists i \in \{1, \ldots, n\} : u_i \neq v_i \\ \mathbf{u} \prec \mathbf{v} & \Leftrightarrow & \mathbf{u} \preceq \mathbf{v} \text{ and } \mathbf{u} \neq \mathbf{v} \end{array}$$

Considering two candidate points $\mathbf{x}_1, \mathbf{x}_2 \in \Omega$, if $\mathbf{f}(\mathbf{x}_1) \prec \mathbf{f}(\mathbf{x}_2)$ then a rational choice between $\mathbf{x}_1$ and $\mathbf{x}_2$ would impose the choice of $\mathbf{x}_1$. This motivates the terminology: $\mathbf{x}_1$ *dominates* $\mathbf{x}_2$ if $\mathbf{f}(\mathbf{x}_1) \prec \mathbf{f}(\mathbf{x}_2)$. It is possible that, given $\mathbf{f}(\mathbf{x}_1)$ and $\mathbf{f}(\mathbf{x}_2)$, neither $\mathbf{f}(\mathbf{x}_1) \prec \mathbf{f}(\mathbf{x}_2)$ nor $\mathbf{f}(\mathbf{x}_2) \prec \mathbf{f}(\mathbf{x}_1)$ occurs. This means that the relation $\prec$ defines a *strict partial order* for $\mathbb{R}^n$ (notice that a strict partial order is not reflexive and is antisymmetric).

The same argument of rationality, when extended to the decision problem over the set Ω, leads to the choice of the solutions $\mathbf{x} \in \Omega$ which are not dominated by any other solution in Ω: those solutions are the *minimal* solutions in Ω, considering the strict partial order $\prec$. A *multi-objective optimization* problem is defined as the problem of finding such minimal solutions:

$$\begin{array}{l} \min \mathbf{f}(\mathbf{x}) \ = \ (f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_n(\mathbf{x})) \\ \text{subject to: } \ \mathbf{x} \in \Omega \end{array} \tag{1}$$

Formally, the solution set of this problem, denoted by $\mathcal{P}$, is defined by:

$$\mathcal{P} = \{\mathbf{x} \in \Omega \mid \ \nexists \bar{\mathbf{x}} \in \Omega \text{ such that } \mathbf{f}(\bar{\mathbf{x}}) \prec \mathbf{f}(\mathbf{x})\} \tag{2}$$

The set $\mathcal{P}$ is usually called the *Pareto-set* or the *non-dominated solution set* of the problem.

A. Pareto-Dominance Search

An abstraction of an iteration of an Evolutionary Algorithm may be stated as:

- Consider a current set, in iteration k, of tentative solutions $\{\mathbf{x}_1^k, \mathbf{x}_2^k, \ldots, \mathbf{x}_p^k\}$.
- A perturbed set of tentative solutions is generated as perturbations of the current solutions:

$$\bar{\mathbf{x}}_i^k = \mathbf{x}_i^k + \boldsymbol{\zeta}_i$$

  If the perturbation $\boldsymbol{\zeta}_i$ is not conditioned by any other solution $\mathbf{x}_j^k$, the operation is usually called a *mutation*, while if $\boldsymbol{\zeta}_i$ is conditioned by another solution $\mathbf{x}_j^k$, the operation is usually called a *crossover*.

- The perturbed set of solutions $\{\bar{\mathbf{x}}_1^k, \bar{\mathbf{x}}_2^k, \ldots, \bar{\mathbf{x}}_p^k\}$ is compared with the current set $\{\mathbf{x}_1^k, \mathbf{x}_2^k, \ldots, \mathbf{x}_p^k\}$, and a *selection* is performed in order to choose a new set of solutions $\{\mathbf{x}_1^{k+1}, \mathbf{x}_2^{k+1}, \ldots, \mathbf{x}_p^{k+1}\}$.

Evolutionary Multiobjective Optimization Algorithms (EMOA) are Evolutionary Algorithms which are intended to deliver a set of samples of the Pareto-set $\mathcal{P}$ of a multiobjective optimization problem. It is said that the EMOA performs a *Pareto-dominance search* if the selection, in the third step above, is based on the dominance relation:

- If $\mathbf{f}(\mathbf{x}_i) \prec \mathbf{f}(\mathbf{x}_j)$ then choose $\mathbf{x}_i$;
- Otherwise, choose between $\mathbf{x}_i$ and $\mathbf{x}_j$ taking into account a diversity criterion.

This abstraction will be considered in this work. Specifically, the difficulty for the operation $\bar{\mathbf{x}} = \mathbf{x} + \boldsymbol{\zeta}$ to generate a new tentative solution $\bar{\mathbf{x}}$ such that $\mathbf{f}(\bar{\mathbf{x}}) \prec \mathbf{f}(\mathbf{x})$ will be studied here.

III. Analytical Tools for the Study of Pareto-Dominance Search

In this section, some analytical tools are developed in order to allow a detailed analysis of the difficulty associated to Pareto-dominance search in many-objective optimization. The general strategy to be followed in this section is:

- A very simple, scalable and symmetric problem is stated, first considering the number of objectives equal to the number of dimensions of the decision space, i.e. $n = N$. This simplification will be removed later. Such a problem is endowed with a particular geometric structure which allows the construction of tight bounds for the probability of generation of a solution that dominates a previous solution, as a perturbation (mutation) of this solution.
- Some points $\mathbf{x}_d$, located in a particular direction, are chosen in order to become the basis for the analysis of the process of generation of new points that dominate it.
- The region $\mathcal{D}(\mathbf{x}_d)$ which contains the solutions that dominate $\mathbf{x}_d$ is determined analytically. A bounding cone $\mathcal{C}_d$ which is tangent to $\mathcal{D}(\mathbf{x}_d)$ is determined, and a spherical cone $\mathcal{C}_k$ which contains $\mathcal{C}_d$ is found.
- The hyperspherical cap $H(\mathbf{x}_d, \delta)$ which is the intersection between the cone $\mathcal{C}_k$ and a ball with center in $\mathbf{x}_d$ and radius δ is determined. The hyperspherical simplex-cap $X(\mathbf{x}_d, \delta)$ which is the intersection between the cone $\mathcal{C}_d$ and a ball with center in $\mathbf{x}_d$ and radius δ is also determined.
- An upper bound $p_n(d)$ for the probability of generation of a solution which dominates $\mathbf{x}_d$ is calculated as the ratio between the areas of the hyperspherical cap $H(\mathbf{x}_d, \delta)$ and the whole ball. An asymptotic expression for a tighter bound $p_n^*(d)$ for the same probability is then calculated, using an asymptotic relation between $X(\mathbf{x}_d, \delta)$ and $H(\mathbf{x}_d, \delta)$. Those expressions for the probability show an explicit dependency with the number of objectives and with the distance from $\mathbf{x}_d$ to the Pareto-set.

All analysis is conducted in the decision variable space. Those steps are described in the next sub-sections.

A. Analysis-Oriented Problem Formulation

In the order to analyze the specific effect of increasing the number of objectives in evolutionary multiobjective optimization, consider the problem of minimization of the following set of functions:

$$f_i(\mathbf{x}) = \|\mathbf{x} - \mathbf{e}_i\|^2 \quad , \quad i = 1, \ldots, n \tag{3}$$

where $\mathbf{e}_i \in \mathbb{R}^n$, $i = 1, \cdots, n$ is the i-th canonical basis vector (the vector with all coordinates equal to zero except the i-th one, which is equal to 1), and $\| \cdot \|$ stands for the Euclidean norm of the argument vector. This problem presents the following characteristics:

(i) The functions are convex;
(ii) The functions' contour sets present spherical symmetry around their points of minima, which are located in $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$,
(iii) The Pareto-set is the $(n-1)$-dimensional simplex with vertices in $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$.[2]

The feature (i) means that anywhere in the decision variable space, the local information provided by the samples of the objective functions will globally indicate the correct direction in which the Pareto-optimal solutions are located. Features (ii) and (iii) will be employed here in the construction of an analytical characterization of the problem difficulty along some space directions.

The analysis of the behavior of Pareto-dominance based search on this simple problem has the purpose of allowing the identification of the structural difficulties related only to the growth of the number of objectives, removing other possibly interfering factors such as multi-modality, ill-conditioning, deceptive behavior, and so forth.

An analysis of a multiobjective problem similar to (3) was suggested in [16]. However, in that reference the analysis was only qualitative, and no conclusive results were obtained.

B. Dominating Solutions and Bounding Cones

1) Dominating Solutions: Denote by $S_{\mathbf{v},\delta} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{v}\| \leq \delta\}$ the closed ball with center in $\mathbf{v}$ and radius δ, and by $\bar{S}_{\mathbf{v},\delta} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{v}\| = \delta\}$ the sphere which is the boundary of $S_{\mathbf{v},\delta}$. Consider a point $\mathbf{x}_d \in \mathbb{R}^n$ such that $\mathbf{x}_d = (d, d, \ldots, d)$. This point is equally distant from the individual minima of functions f_i, with those distances given by

$$r = \|\mathbf{x}_d - \mathbf{e}_i\| = \sqrt{(n-1)d^2 + (d-1)^2}.$$

The contour sets of the functions $f_i(\mathbf{x})$, defined by $\bar{L}_{i,\omega} = \{\mathbf{x} \mid f_i(\mathbf{x}) = \omega^2\}$, are given by $\bar{L}_{i,\omega} = \bar{S}_{\mathbf{e}_i,\omega}$, while the sub-level sets defined by $L_{i,\omega} = \{\mathbf{x} \mid f_i(\mathbf{x}) \leq \omega^2\}$ are given by $L_{i,\omega} = S_{\mathbf{e}_i,\omega}$. Therefore, the set of points which dominate the point $\mathbf{x}_d$, denoted by $\mathcal{D}(\mathbf{x}_d)$, is given by:

$$\mathcal{D}(\mathbf{x}_d) = S_{\mathbf{e}_1,r} \cap S_{\mathbf{e}_2,r} \cap \ldots S_{\mathbf{e}_n,r} \tag{4}$$

[2] The reader can check this condition by noticing that the Kuhn-Tucker conditions for efficiency [15] are satisfied within this simplex, and do not hold outside it.

2) Distance to the Pareto-Set: The point $\mathbf{x}_d^* = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ belongs to the Pareto-set, and it is the Pareto-optimal point which is nearest to the point $\mathbf{x}_d$. The distance from $\mathbf{x}_d$ to $\mathbf{x}_d^*$ is denoted by $\lambda = \|\mathbf{x}_d^* - \mathbf{x}_d\|$.

3) Tangent Cone: Let $\mathcal{C}_d$ be the convex cone with vertex in $\mathbf{x}_d$ and boundaries defined by the n planes M_i which are tangent to the surfaces $\bar{S}_{\mathbf{e}_1,r}, \ldots, \bar{S}_{\mathbf{e}_n,r}$ on $\mathbf{x}_d$, such that $\mathcal{C}_d \supset \mathcal{D}(\mathbf{x}_d)$.

4) Spherical Cone: Consider the function

$$F_i(\mathbf{x}) = f_i(\mathbf{x}) - \|\mathbf{x}_d - \mathbf{e}_i\|^2, \tag{5}$$

It should be noticed that $\bar{S}_{\mathbf{e}_i,r} = F_i^{-1}(0)$. The gradient of $F_i(\mathbf{x})$ is given by

$$\nabla F_i(\mathbf{x}_d) = \nabla f_i(\mathbf{x}_d) = (\underbrace{2d, \ldots, 2d}_{1,\ldots,i-1}, 2(d-1), \underbrace{2d, \ldots, 2d}_{i+1,\ldots,n})$$

The tangent plane to $\bar{S}_{\mathbf{e}_i,r}$ at $\mathbf{x}_d$ (plane M_i) is the orthogonal complement of the line determined by vector $\nabla F_i(\mathbf{x}_d)$. Define the vector $\mathbf{k}_j$ as

$$\mathbf{k}_j = (\underbrace{d, \ldots, d}_{1,\ldots,j-1}, -((n-1)d-1), \underbrace{d, \ldots, d}_{j+1,\ldots,n}) \tag{6}$$

Clearly, $\nabla F_i(\mathbf{x}_d) \cdot \mathbf{k}_j^T = 0 \;\forall\; i \neq j$. Therefore:

$$\mathbf{k}_j \in \bigcap_{i \neq j} M_i$$

Define now a cone $\mathcal{C}_k$ with vertex on $\mathbf{x}_d$ and with spherical section, such that its boundary surface $\bar{C}_k$ contains the lines passing on $\mathbf{x}_d$ and with directions given by $\mathbf{k}_j, j = 1, \ldots, n$. This means that the cone $\mathcal{C}_d$ is inscribed in cone $\mathcal{C}_k$. Clearly:

$$\mathcal{C}_k \supset \mathcal{C}_d \supset \mathcal{D}(\mathbf{x}_d)$$

The Figure 1 shows a representation of cones $\mathcal{C}_d$ and $\mathcal{C}_k$ for the case $n = 3$.

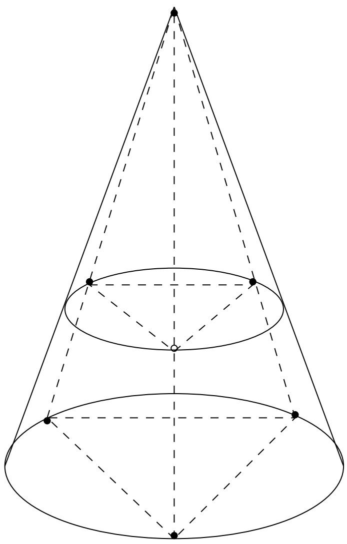

Fig. 1. Representation of the cones $\mathcal{C}_d$ and $\mathcal{C}_k$, for the case $n = 3$. In this case, $\mathcal{C}_k$ is the cone with circular section and $\mathcal{C}_d$ is the cone with triangular section inscribed in $\mathcal{C}_k$.

C. Intersection Cone-Ball

1) Internal Angle: Let the internal angle of the cone $\mathcal{C}_k$ (which is also the angle between any $\mathbf{k}_j$ and $\mathbf{x}_d$) be denoted by θ. An expression for θ comes from

$$\begin{aligned} \cos(\theta) &= \frac{\mathbf{k}_j \cdot \mathbf{x}_d^T}{\|\mathbf{k}_j\|\|\mathbf{x}_d\|} \\ &= \frac{(n-1)d^2 - d^2(n-1) + d}{\sqrt{(n-1)d^2 + ((n-1)d-1)^2}\sqrt{nd^2}} \\ &= \frac{1}{\sqrt{n}} \frac{1}{\sqrt{(n-1)d^2 + ((n-1)d-1)^2}} \end{aligned} \tag{7}$$

The following statements come directly from (7):

- If $d \to +\infty$ then $\theta \to \pi/2$;
- If $d = 1/n$ then $\theta = 0$.

When $d = 1/n$, the point $\mathbf{x}_d$ becomes exactly on the Pareto-set.

2) Hyperspherical Cap and Hyperspherical Simplex-Cap: The hyperspherical cap obtained from the intersection of sets $\bar{S}_{\mathbf{x}_d,\delta}$ and $\mathcal{C}_k$ is denoted by $H(\mathbf{x}_d, \delta)$. The intersection between the sets $\bar{S}_{\mathbf{x}_d,\delta}$ and $\mathcal{C}_d$ (a hyperspherical simplex-cap) is denoted by $X(\mathbf{x}_d, \delta)$.

Figure 2 shows a representation of the entities described previously in this section for the case $n = 2$.

D. Probability of Generation of a Dominating Solution

Define a random variable vector $\boldsymbol{\zeta} \in \mathbb{R}^n$ with a probability density function $\phi(\boldsymbol{\zeta})$ which obeys

$$\phi(\boldsymbol{\zeta}) = \phi_\rho \;\forall\; \boldsymbol{\zeta} \in \bar{S}_{0,\rho}. \tag{8}$$

This means that $\phi(\boldsymbol{\zeta})$ presents a spherical symmetry ($\phi(\boldsymbol{\zeta})$ is equal to a constant ϕ_ρ for all $\boldsymbol{\zeta}$ on a given sphere of radius ρ), and by consequence $\boldsymbol{\zeta}$ has a uniform probability of being in any space direction. Now, consider a point $\mathbf{x}_z$ which is obtained by a perturbation of the point $\mathbf{x}_d$ by ζ:

$$\mathbf{x}_z = \mathbf{x}_d + \boldsymbol{\zeta} \tag{9}$$

The probability of $\mathbf{x}_z$ to belong to the cone $\mathcal{C}_k$ is given by:

$$P(\mathbf{x}_z \in \mathcal{C}_k) = p_n(d) = \frac{A(H(\mathbf{x}_d, \delta))}{A(\bar{S}_{\mathbf{x}_d,\delta})}, \tag{10}$$

where $A(\cdot)$ represents the area of the argument hyper-surface. The notation $p_n(d)$ is intended to let explicit the dependence of this probability with the variable d, given a problem dimension n. The probability of $\mathbf{x}_z$ to belong to the cone $\mathcal{C}_d$ is given by:

$$P(\mathbf{x}_z \in \mathcal{C}_d) = p_n^*(d) = \frac{A(X(\mathbf{x}_d))}{A(\bar{S}_{\mathbf{x}_d,\delta})}, \tag{11}$$

The notation $p_n^*(d)$ is intended to highlight that this probability is the object of interest in this paper, as long as it constitutes a tight bound for the probability of success of a perturbation, in the search for dominating solutions.

The following change of variable is performed on the functions $p_n(d)$ and $p_n^*(d)$:

$$\begin{aligned} \tilde{p}_n(\lambda) &= p_n\left(\frac{1}{n} + \frac{\lambda}{\sqrt{n}}\right) \\ \tilde{p}_n^*(\lambda) &= p_n^*\left(\frac{1}{n} + \frac{\lambda}{\sqrt{n}}\right) \end{aligned} \tag{12}$$

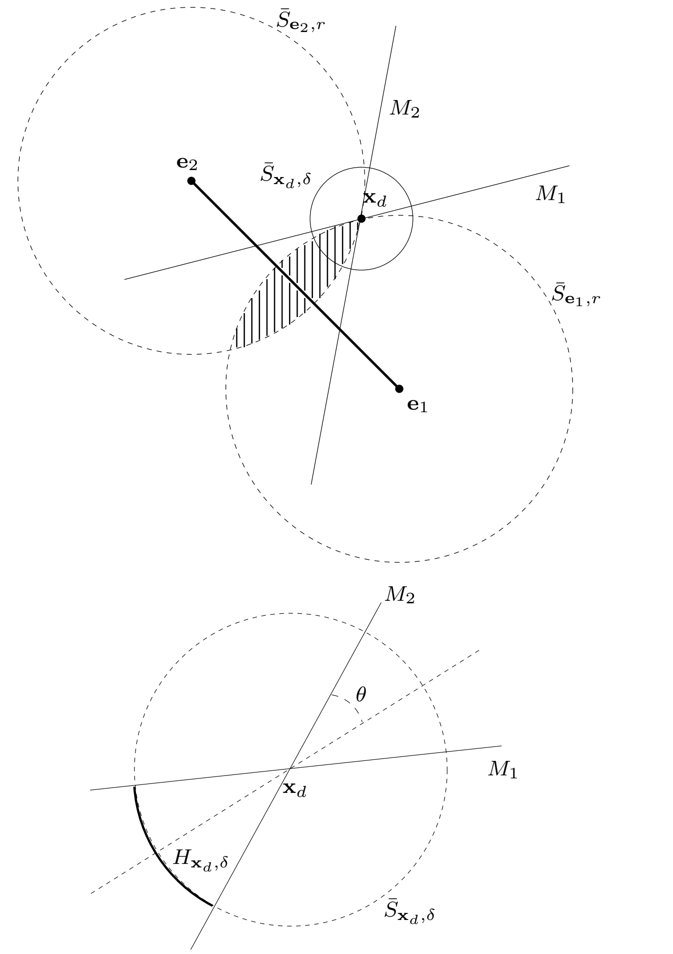

Fig. 2. Top: A problem with two variables and two objectives is represented. The Pareto-set is the line segment between the points $\mathbf{e}_1$ and $\mathbf{e}_2$. For a given point $\mathbf{x}_d$, the level surfaces $\bar{S}_{\mathbf{e}_1,r}$ and $\bar{S}_{\mathbf{e}_2,r}$ are the circles with center in $\mathbf{e}_1$ and $\mathbf{e}_2$, passing on $\mathbf{x}_d$. The tangent lines to $\bar{S}_{\mathbf{e}_1,r}$ and $\bar{S}_{\mathbf{e}_2,r}$, respectively M_1 and M_2, define the cone $\mathcal{C}_d$ which, in the two-dimension case, is equivalent to the cone $\mathcal{C}_k$. The circle $\bar{S}_{\mathbf{x}_d,\delta}$, with center on $\mathbf{x}_d$ and radius δ is also represented. The set of points that dominate $\mathbf{x}_d$, $\mathcal{D}(\mathbf{x}_d)$, is represented by the shaded region. Bottom: A zoom on the circle $\bar{S}_{\mathbf{x}_d,\delta}$ is presented, showing the angle θ and the hyperspherical cap $H(\mathbf{x}_d, \delta)$.

The function $\tilde{p}_n(\lambda)$ represents the probability of $\mathbf{x}_z = \mathbf{x}_d + \boldsymbol{\zeta}$ to belong to the cone $\mathcal{C}_k$, and the function $\tilde{p}_n^*(\lambda)$ represents the probability of $\mathbf{x}_z = \mathbf{x}_d + \boldsymbol{\zeta}$ to belong to the cone $\mathcal{C}_d$. In both cases, the probabilities are now parametrized in terms of λ, the distance from $\mathbf{x}_d$ to the Pareto-set.

The following relations hold:

$$\tilde{p}_n(\lambda) = P(\mathbf{x}_z \in \mathcal{C}_k) > \tilde{p}_n^*(\lambda) = P(\mathbf{x}_z \in \mathcal{C}_d) > P(\mathbf{x}_z \in \mathcal{D}(\mathbf{x}_d)) \tag{13}$$

1) Expression of $p_n(d)$: Now, an analytical description of the function $p_n(d)$ is pursued. In the reference [17], the authors suggest a way to calculate $A(\bar{S}_{\mathbf{x}_d,\delta})$ for any $n > 1$.

Proposition 1 ([17]): For any $n > 1$,

$$A(\bar{S}_{\mathbf{x}_d,\delta}) = \frac{2\pi^{n/2}\delta^{n-1}}{\Gamma(n/2)},$$

where Γ is the gamma function.

In the reference [18], the authors determine a relationship between $A(\bar{S}_{\mathbf{x}_d,\delta})$ and $A(H(\mathbf{x}_d, \delta))$. This relationship is presented next.

Proposition 2 ([18]): For any $n > 1$,

$$A(H(\mathbf{x}_d, \delta)) = (1/2)A(\bar{S}_{\mathbf{x}_d,\delta}) I_{\sin^2(\theta)}\left(\frac{n-1}{2}, \frac{1}{2}\right),$$

where $I_z(p, q)$ is the regularized incomplete beta function.

The function $I_z(p, q)$ is defined by the incomplete beta function $B(z; p, q)$ and the (complete) beta function $B(p, q)$ as follows

$$I_z(p, q) = \frac{B(z; p, q)}{B(p, q)}, \tag{14}$$

where

$$\begin{aligned} B(z; p, q) &= \int_0^z u^{p-1}(1-u)^{q-1} du \\ B(p, q) &= \int_0^1 u^{p-1}(1-u)^{q-1} du \end{aligned}$$

Now, by using the two previous results, a description of the function $p_n(d)$ is obtained.

Lemma 1: The function $p_n(d)$ defined in (10) *can be calculated as*

$$p_n(d) = \frac{1}{2} I_{\sin^2(\theta)}\left(\frac{n-1}{2}, \frac{1}{2}\right), \tag{15}$$

where θ is given by

$$\theta = \arccos\left(\frac{1}{\sqrt{n}} \frac{1}{\sqrt{(n-1)d^2 + ((n-1)d-1)^2}}\right).$$

Some special cases of the $I_{\sin^2(\theta)}\left(\frac{n-1}{2}, \frac{1}{2}\right)$ which were displayed in [19] are summarized in the Table I.

TABLE I
SPECIAL CASES OF THE $p_n(d)$.

n	$p_n(d)$
2	θ/π
3	$(1-\cos(\theta))/2$
4	$(2\theta - \sin(2\theta))/2\pi$
5	$(1-(3/2)\cos(\theta)+(1/2)\cos^3(\theta))/2$

The function $p_n(d)$ represents the probability of $\mathbf{x}_z = \mathbf{x}_d + \boldsymbol{\zeta}$ to belong to the cone $\mathcal{C}_k$, which is an upper bound for the probability of $\mathbf{x}_z$ to dominate $\mathbf{x}_d$.

2) Asymptotic Expression for $p_n^*(d)$: An asymptotic expression for $p_n^*(d)$, which is valid when $\mathbf{x}_d$ becomes close to the Pareto-set, is derived now.

Lemma 2: The following relation holds:

$$\mathcal{A}_n = \lim_{\lambda \to 0} \frac{A(X(d))}{A(H(\mathbf{x}_d, \delta))} = \frac{\sqrt{n+1}\ \Gamma\left(\frac{n}{2}+1\right)}{n!\ \pi^{n/2} \left(\sqrt{\frac{n}{n+1}}\right)^n} \tag{16}$$

Proof: Let $T(\mathbf{x}_d, \delta)$ denote the tangent plane to the spherical cap $H(\mathbf{x}_d, \delta)$ in its center point (this plane is normal to the vector $\mathbf{x}_d$). Then, notice that $\theta \to 0$ as $\lambda \to 0$. Clearly:

$$\lim_{\lambda \to 0} H(\mathbf{x}_d, \delta) = T(\mathbf{x}_d, \delta) \cap \mathcal{C}_k$$

Since $\mathcal{C}_k$ is a spherical cone and $\mathcal{C}_d$ is a polyhedral cone inscribed in $\mathcal{C}_k$, then:

- A section of $\mathcal{C}_k$ by a hyperplane normal to the direction $\mathbf{x}_d$ is an $(n-1)$-dimensional hypersphere, denoted by $\mathcal{S}^{n-1}$.
- A section of $\mathcal{C}_d$ by the same hyperplane is an $(n-1)$-dimensional simplex inscribed in that hypersphere, denoted by $\mathcal{X}^{n-1}$.
- Given n, the ratio between the volume of $\mathcal{X}^{n-1}$ and the volume of $\mathcal{S}^{n-1}$ is a constant $\mathcal{A}_n$, for any internal angle θ.

Now, noticing that the volume of a unitary ball of dimension $n-1$ is given by:

$$\mathrm{Vol}\,(\mathcal{S}^{n-1}) = \frac{\pi^{n/2} \left(\sqrt{\frac{n}{n+1}}\right)^n}{\Gamma\left(\frac{n}{2}+1\right)}$$

and the volume of its inscribed regular simplex is given by:

$$\mathrm{Vol}\,(\mathcal{X}^{n-1}) = \frac{\sqrt{n+1}}{n!}$$

the expression of constant $\mathcal{A}_n$ becomes the expression (16).

The main result of this paper is stated now, as a combination of lemmas 1 and 2:

Theorem A: The probability $\tilde{p}_n^*(\lambda)$ is such that:

(i) $\lim_{\lambda \to 0^+} \tilde{p}_n^*(\lambda) = \mathcal{A}_{n-1}\, \tilde{p}_n(\lambda)$;

(ii) $\lim_{\lambda \to \infty} \tilde{p}_n^*(\lambda) = 1/2$.

E. The case $N \geq n$

Now consider a more general situation in which the dimension of the decision space, N, may be greater than or equal to the dimension of the objective space, n. Let the decision variable vector $\mathbf{x}$ be decomposed into two sub-vectors $\mathbf{x}_l \in \mathbb{R}^n$ and $\mathbf{x}_h \in \mathbb{R}^{N-n}$:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_l \\ \mathbf{x}_h \end{bmatrix} \quad , \quad \mathbf{x}_l = \begin{bmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix} \quad , \quad \mathbf{x}_h = \begin{bmatrix} x_{n+1} \\ \vdots \\ x_N \end{bmatrix} \tag{17}$$

In terms of subspaces:

$$\begin{aligned} &\mathbf{x} \in \mathcal{X} = \mathbb{R}^N \quad , \quad \mathbf{x}_l \in \mathcal{X}_l \quad , \quad \mathbf{x}_h \in \mathcal{X}_h \\ &\mathcal{X} = \mathcal{X}_l \oplus \mathcal{X}_h \end{aligned} \tag{18}$$

The same n objective functions $f_i(\mathbf{x})$ are considered:

$$f_i(\mathbf{x}) = \|\mathbf{x} - \mathbf{e}_i\|^2 \quad , \quad i = 1, \ldots, n \tag{19}$$

Now $\mathbf{e}_i \in \mathbb{R}^N$, with $i = 1, \cdots, n$, for $n \leq N$, still means the i-th canonical basis vector. The Pareto-set of problem (19), when projected onto the space $\mathbb{R}^n$, corresponds to the same Pareto-set of problem (3). Each objective function in expression (19) may be written as:

$$\begin{aligned} &f_i(\mathbf{x}) = f_i^l(\mathbf{x}_l) + f_i^h(\mathbf{x}_h) \\ &f_i^l(\mathbf{x}_l) = x_1^2 + \ldots + (x_i - 1)^2 + \ldots + x_n^2 \\ &f_i^h(\mathbf{x}_h) = f^h(\mathbf{x}_h) = x_{n+1}^2 + \ldots + x_N^2 \end{aligned} \tag{20}$$

With those entities, it becomes possible to describe the progression of a Pareto-dominance based evolutionary multiobjective optimization algorithm in the task of minimization of functions (19), as follows:

- Let the algorithm start on a point that does not belong neither to $\mathcal{X}_l$ nor to $\mathcal{X}_h$, for instance the point $\mathbf{x}_0$ given by:

$$\mathbf{x}_0 = \begin{bmatrix} \mathbf{x}_0^l \\ \mathbf{x}_0^h \end{bmatrix}$$

- Consider the tentative point $\bar{\mathbf{x}} = \mathbf{x}_0 + \boldsymbol{\zeta}$. Decompose $\bar{\mathbf{x}}$ as:

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{\mathbf{x}}_l \\ \bar{\mathbf{x}}_h \end{bmatrix}$$

  such that $\bar{\mathbf{x}}_l \in \mathcal{X}_l$ and $\bar{\mathbf{x}}_h \in \mathcal{X}_h$.
- For small perturbations $\boldsymbol{\zeta}$, the probability of $\bar{\mathbf{x}}_l$ dominating $\mathbf{x}_0^l$ concerning the objective function $\mathbf{f}^l(\mathbf{x}_l)$ is given by $\tilde{p}_n^*(\lambda)$, which is described by Theorem A, case (i), in the interesting case of $\mathbf{x}_0$ close to the Pareto-set. On the other hand, the probability of $\bar{\mathbf{x}}_h$ dominating $\mathbf{x}_0^h$ concerning the objective function $\mathbf{f}^h(\mathbf{x}_h)$ is equal to 0.5, no matter the number of space dimensions N and the values of the coordinates of $\mathbf{x}_h$.
- In this situation of $\mathbf{x}_0$ close to the Pareto-set, the following outcomes of mutations may occur:
  - (a) There will be situations in which $\mathbf{f}^l(\bar{\mathbf{x}}_l) \prec \mathbf{f}^l(\mathbf{x}_0^l)$, which will occur with probability $\lim_{\lambda \to 0^+} \tilde{p}_n^*(\lambda) = \mathcal{A}_{n-1}\, \tilde{p}_n(\lambda)$.
  - (b) There will be situations in which $\mathbf{f}^h(\bar{\mathbf{x}}_h) \prec \mathbf{f}^h(\mathbf{x}_0^h)$, which will occur with probability 0.5.
- Therefore, the probability of a small perturbation $\boldsymbol{\zeta}$ to produce a new point $\bar{\mathbf{x}} = \mathbf{x}_0 + \boldsymbol{\zeta}$ which satisfies both (a) and (b) is given by:

$$p_N(\lambda) = \frac{1}{2} \mathcal{A}_{n-1}\, \tilde{p}_n(\lambda) \tag{21}$$

  However, it is possible that a $\bar{\mathbf{x}}$ that dominates $\mathbf{x}_0$ may occur, in some cases, without the simultaneous satisfaction of conditions (a) and (b), because the objective functions are composed of sums of functions f_i^l and f^h, as shown in (20), which means that an increment in a function may be compensated by a decrement in another one, still leading to a decrement in the sum. Therefore, the expression (21) represents a lower bound on the probability of generation of a $\bar{\mathbf{x}}$ that dominates $\mathbf{x}_0$.

IV. Simulation Studies

Three issues are investigated now. The first one is the range of validity of the asymptotic expression for $\tilde{p}_n^*$. It is important to know what is the range of λ in which the asymptotic expression presented in Theorem A provides a good approximation for $\tilde{p}_n^*$. The second issue is the effect of the point $\mathbf{x}$ which is to be enhanced to be outside the specific direction $\mathbf{x}_d$. It is relevant to understand to what extent the analysis performed on the line $\mathbf{x}_d$ can be extended to points in other directions. The third issue is related to the validity of the lower bound stated in expression (21) for the probability of generation of a solution that dominates the current one in the case $N > n$. These three questions are difficult to deal analytically; therefore they will be addressed via simulation studies in this section.

A. Range of Validity of Asymptotic Expression

In order to evaluate the range of validity of the asymptotic formulation for the probability $\tilde{p}_n^*(\lambda)$, a series of simulation studies was conducted here. In the simulations, for each $n \geq 2$ and for 50 values of λ logarithmically distributed in the interval $[10^{-3}, 10^1]$, points $\mathbf{x}_z = \mathbf{x}_d + \boldsymbol{\zeta}$ are generated with $\boldsymbol{\zeta}$ normally distributed, with zero mean and standard deviation $\sigma = \lambda \times 10^{-3}$. With such a standard deviation, all points become close to $\mathbf{x}_d$, which makes the gap between $\mathcal{D}(\mathbf{x}_d)$ and $\mathcal{C}_d$ to become small. Such a standard deviation also guarantees that no point is generated on the other side of the Pareto-set. For each n and each λ, a total number of 10^8 realizations of $\mathbf{x}_z$ are generated, and the proportion of elements that dominate $\mathbf{x}_d$ is then observed.

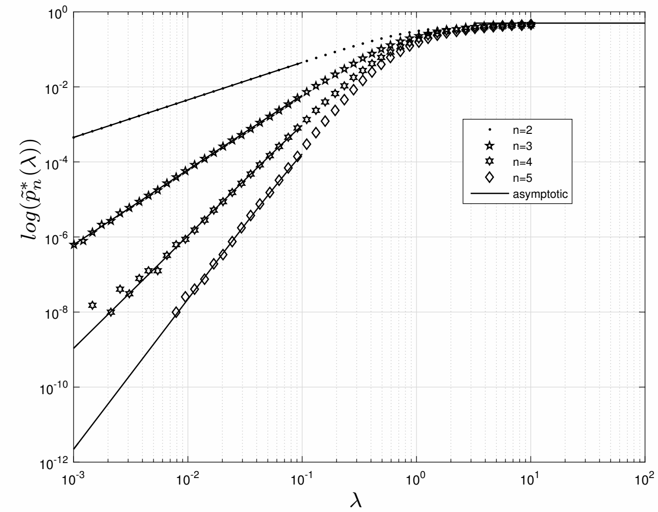

Fig. 3. In logarithmic coordinates, a simulation of the Pareto-dominance relation for different values of n.

The Figure 3 presents the simulation results superimposed to the asymptotic values of $\tilde{p}_n^*(\lambda)$ that are furnished by Theorem A. It can be noticed that in the intervals $\lambda < 10^{-1}$ and $\lambda > 10$, the simulation results are in good agreement with the predicted values. Only in the interval $10^{-1} < \lambda < 10$ the actual values of $\tilde{p}_n^*(\lambda)$ are not accurately predicted by the asymptotic formulae.

TABLE II
Parameters of power law equation (22) for different values of n, calculated for $\lambda = 1 \times 10^{-4}$ and $\lambda = 5 \times 10^{-4}$.

n	c_n	γ_n
2	0.4502	1.0000
3	0.6202	2.0000
4	1.0807	3.0000
5	2.2207	4.0005

The most noticeable feature that can be observed in the behavior of $\tilde{p}_n^*(\lambda)$ is the presence, for all values of n, of a knee nearby the position $\lambda = 1$ (or $\lambda = 10^0$), which separates the graphic of $\tilde{p}_n^*(\lambda)$ in three distinct regions:

- For $\lambda > 10$, $\tilde{p}_n(\lambda) \approx 0.5$, which means that it is uniformly easy to find an $\mathbf{x}_z$ which dominates $\mathbf{x}_d$ when $\mathbf{x}_d$ is located in this region;
- For $\lambda < 0.1$, the graphic of $\tilde{p}_n(\lambda)$ approaches a line, in logarithmic scale, which indicates a power law behavior;
- There is a region of transition between the former regions (the knee), for $0.1 \leq \lambda \leq 10$.

The power law that holds for $\lambda < 0.1$ should be given by:

$$\tilde{p}_n^*(\lambda) = c_n \, \lambda^{\gamma_n} \tag{22}$$

The values of parameters c_n and γ_n, for $n = \{2, 3, 4, 5\}$, calculated with the values provided by Theorem A for $\lambda = 1 \times 10^{-4}$ and $\lambda = 5 \times 10^{-4}$ are shown in Table II.

It is noticeable that the estimated power law exponent γ_n seems to be equal to $(n-1)$. The small deviation in γ_5 may be attributed to the numerical errors that may arise in the computation of a fraction of two very large numbers which appear in the asymptotic expression of Theorem A as n grows.

B. Arbitrary Points

When a point $\mathbf{x}$ is nearby the Pareto-set $\mathcal{P}$, there are two possible situations: the projection of the point into the $\mathcal{P}$ may be on the relative interior of the Pareto-set, or that projection may be on the relative boundary of $\mathcal{P}$. In the second case, the probability of enhancement would become much greater than in the case of points whose projection fall on the relative interior of the Pareto-set – this situation approximates the case in which the points to be determined are Pareto-optimal solutions of reduced-order problems involving a subset of the objective functions. It is a well-known fact that this is much easier than the determination of solutions of the full multiobjective problem.

Clearly, the relevant situation for the generation of solutions that cover the Pareto-set $\mathcal{P}$ is that of points whose projection fall in the relative interior of $\mathcal{P}$. In order to consider points that represent this situation, the following experiment was performed for the dimensions $n = 2$, $n = 3$, $n = 4$ and $n = 5$, in each case considering $\lambda = 0.1$, $\lambda = 0.05$ and $\lambda = 0.01$:

(i) A random convex combination $\mathbf{x}_p$ of the extremal points of the Pareto-set is generated with uniform probability distribution of the combination coefficients. The point $\mathbf{x}_p$ generated in this way is a Pareto-set point. A unitary vector $\mathbf{u}_d$ normal to the Pareto-set is determined.

TABLE III
SIMULATED PROBABILITY OF ENHANCEMENT OF A POINT SITUATED AT A DISTANCE λ FROM THE PARETO-SET, IN 100 EXPERIMENTS, FOR DIFFERENT VALUES OF n. p_{05}: 5% PERCENTILE, p_{50}: 50% PERCENTILE, p_{95}: 95% PERCENTILE.

	n	p_{05}	p_{50}	p_{95}	Theo. A
$\lambda = 0.1$	2	4.45E-2	5.10E-2	1.34E-1	4.47E-2
	3	6.12E-3	8.56E-3	1.90E-2	5.94E-3
	4	1.09E-3	1.75E-3	4.72E-3	9.43E-4
	5	2.40E-4	4.43E-4	1.18E-3	1.63E-4
$\lambda = 0.05$	2	2.24E-2	2.44E-2	9.44E-2	2.25E-2
	3	1.59E-3	2.36E-3	1.02E-2	1.53E-3
	4	1.45E-4	2.61E-4	1.33E-3	1.30E-4
	5	1.58E-5	3.46E-5	1.51E-4	1.28E-5
$\lambda = 0.01$	2	4.45E-3	4.98E-3	2.44E-2	4.50E-3
	3	5.84E-5	9.56E-5	4.13E-4	6.20E-5
	4	1.03E-6	1.78E-6	1.22E-5	1.08E-6
	5	2.50E-8	5.50E-8	3.86E-7	2.21E-8

(ii) A point $\mathbf{x}$ is chosen, such that $\mathbf{x} = \mathbf{x}_p + \lambda \mathbf{u}_d$.

(iii) Random points $\mathbf{x}_z = \mathbf{x} + \boldsymbol{\zeta}$ are generated with $\boldsymbol{\zeta}$ normally distributed, with zero mean and standard deviation $\sigma = \lambda \times 10^{-3}$. A total number of N realizations of $\mathbf{x}_z$ are generated, and the proportion of elements that dominate $\mathbf{x}$ is then observed. The value of N is chosen such that the expected number of $\mathbf{x}_z$ points which dominate $\mathbf{x}$ is at least 30, for each value of n, according to the asymptotic estimate provided by Theorem A.

The experiment was repeated 100 times for each combination of values of n and λ. The Table III summarizes the results of those experiments.

The data on Table III suggest that the asymptotic value of $\tilde{p}_n^*(\lambda)$ provided by Theorem A corresponds to the smallest value observed, in each case. This means that the "most difficult" direction for finding an enhancement is exactly $\mathbf{x}_d$. In all cases, 50% of the points have a probability of being enhanced that is less than twice the asymptotic value of $\tilde{p}_n^*(\lambda)$. Also in all cases, 95% of the points have a probability of being enhanced which is up to 10 times the asymptotic value of $\tilde{p}_n^*(\lambda)$.

This means that, although the asymptotic value of $\tilde{p}_n^*(\lambda)$ cannot be interpreted as a specific probability of enhancement of a point which is situated at a distance λ to the Pareto-set, this value can be used as a predictor for the order of magnitude of such a probability of enhancement.

A more insightful analysis is obtained from the observation of the parameters of the power law model which are estimated from the data displayed in Table III. Using the values of $\tilde{p}_n^*(\lambda)$ which are estimated for $\lambda = 0.01$ and $\lambda = 0.05$, different power law models are obtained for the 5%, 50% and 95% percentiles of $\tilde{p}_n^*$. The resulting values of parameters are presented in Table IV.

The data in Table IV suggests that the multiplicative constant c_n presents large variations when considering points in different relative positions to the Pareto-set, although still within one order of magnitude of variation. The power law exponent γ_n presents a more stable behavior, remaining nearby the value $n - 1$ for all samples that were considered in the simulation. This means that the rate in which the problem becomes more difficult as the distance to the Pareto-set decreases is approximately the same, for a given n, no matter the value of that distance, for almost all relative positions between the point to be enhanced and the Pareto-set.

TABLE IV
VALUES OF THE POWER LAW PARAMETERS γ_n AND c_n, FOR DIFFERENT VALUES OF n, ESTIMATED FOR THE DATA PRESENTED IN TABLE III.

n	γ_{05}	γ_{50}	γ_{95}	γ: Theo. A
2	1.0042	0.9874	0.8406	1.0000
3	2.0530	1.9922	1.9924	2.0000
4	3.0739	3.0992	2.9150	3.0000
5	4.0069	4.0041	3.7089	4.0005

n	c_{05}	c_{50}	c_{95}	c: Theo. A
2	0.4536	0.4699	1.1713	0.4502
3	0.7454	0.9221	3.9885	0.6202
4	1.4472	2.8102	8.2479	1.0807
5	2.5810	5.6036	10.1002	2.2207

By noticing that the value of γ_n in the 95% percentile becomes smaller than $(n-1)$ but stays always nearer to this value than to $(n-2)$, it may be conjectured that the points nearby the edges of the Pareto-set will present a behavior which takes part of the smaller dimension of that edge. In other words, the search for points on $(n-2)$-dimensional edges of the Pareto-set is similar to a multi-objective optimization with $(n-1)$ objectives, which will be performed with an exponent $\gamma = n - 2$. The points which are on the $(n-1)$-dimensional Pareto-set, in positions nearby those $(n-2)$-dimensional edges, will be pursued with a γ value which is smaller than $(n-1)$, but which is lower-bounded by $(n-2)$. Of course, as the points to be found approach an $(n-i)$-dimensional edge, the value of γ reduces to $(n-i)$, up to the case in which the n single-objective solutions (which are zero-dimensional objects) are pursued, which can be performed with $\gamma = 0$, without any performance degradation. This observation is exploited in the reference [20], which performs searches on the corners of the Pareto-set of many-objective problems.

C. The case $N > n$

Finally, in order to examine the situation in which the number of dimensions of the decision variable space is greater than the number of dimensions of the objective space. An experiment was conducted considering the problem (19) with $n = 4$ and values of N in the range $N \in \{4, 5, 6, 7, 8, 9\}$. Points $\mathbf{x}_z = \mathbf{x}_d + \boldsymbol{\zeta}$ were generated for a $\mathbf{x}_d = \mathbf{x}_d^* + d.\mathbf{v}$, with $\mathbf{v} = (1, \ldots, 1)$ and $d \in \{10, 1, 0.1, 0.01, 0.01\}$. $\boldsymbol{\zeta}$ was normally distributed, with zero mean and standard deviation $\sigma = 10^{-4}$. For each N, a total number of 10^8 realizations of $\mathbf{x}_z$ were generated, and the proportion of elements that dominate $\mathbf{x}_d$ was observed.

The results of this experiment are shown in Table V. It is noticeable that, no matter the value of N, the empirical probability of $\mathbf{x}_z \prec \mathbf{x}_d$ remains almost the same for the same d. It seems that, considering the power law (22), the exponent γ_n does not change for different values of N, and only the constant c_n varies. It is also noticeable that the probability of domination is slightly higher for greater values of N.

TABLE V
EMPIRICAL PROBABILITIES OF $\mathbf{x}_z \prec \mathbf{x}_d$, FOR $n = 4$ AND DIFFERENT VALUES OF N AND d.

	N					
d	4	5	6	7	8	9
10	4.79E-1	4.82E-1	4.83E-1	4.85E-1	4.86E-1	4.86E-1
1	3.09E-1	3.27E-1	3.41E-1	3.51E-1	3.60E-1	3.68E-1
0.1	7.28E-3	9.79E-3	1.24E-2	1.51E-2	1.78E-2	2.06E-2
0.01	8.84E-6	1.22E-5	1.61E-5	1.98E-5	2.55E-5	2.87E-5
0.001	1.00E-8	1.40E-8	1.40E-8	2.90E-8	1.60E-8	3.10E-8

In order to examine the meaning of the slight growth in the probability of generation of a new point that dominates the current one for greater values of N, another experiment was performed. A sequence $\{\mathbf{x}\}_k$ was generated such that:

$$\bar{\mathbf{x}} = \mathbf{x}_k + \boldsymbol{\zeta}$$

$$\begin{array}{l} \text{if } \bar{\mathbf{x}} \prec \mathbf{x}_k \\ \text{then } \mathbf{x}_{k+1} \leftarrow \bar{\mathbf{x}} \\ \text{else } \mathbf{x}_{k+1} \leftarrow \mathbf{x}_k \end{array}$$

This iteration was run for $n = 4$, and $N \in \{4, 9\}$. $\boldsymbol{\zeta}$ was normally distributed, with zero mean and standard deviation $\sigma = 10^{-4}$, and 3×10^8 realizations of $\mathbf{x}_z$ were generated for each value of N. The distances $d_l(k)$ and $d_h(k)$ were defined as the distance from $\mathbf{x}_k$ to the Pareto-set when the variables are projected respectively into the subspaces $\mathcal{X}_l$ and $\mathcal{X}_h$. Those distances are shown in Figure 4.

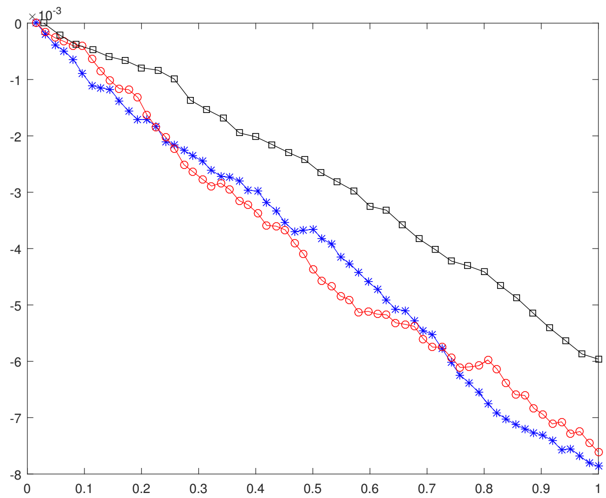

Fig. 4. The line with □ indicates the distance $d_l(k)$ for $N = 4$. The line with ∘ indicates the distance $d_l(k)$ for $N = 9$, and the line with ∗ indicates the distance $d_h(k)$ for $N = 9$. The vertical axis represents the distance in relation to the initial point of the sequence. The horizontal axis represents the ordering of the successful steps (the non-successful trials are not represented in the graphics).

Consistently with the higher probability of generation of a dominating point for greater values of N, as shown in Table V, the Figure 4 shows that $d_l(k)$ decreases slightly faster for $N = 9$ than for $N = 4$. However, for $N = 9$, both $d_l(k)$ and $d_h(k)$ have a non-monotonic behavior: there are some points in which one of those distances increases, associated with a decrease in the other one.

V. PROGRESSION OF PARETO-DOMINANCE SEARCH

In all the analysis to be developed in this section it should be noticed that, in the multiobjective optimization problem considered here, the functions are scaled such that the distance between any two points of minimum (points in which an objective function reaches its minimum value) is equal to $\sqrt{2} \approx 1.414$. Therefore, when a distance λ from a point $\mathbf{x}_d$ to the Pareto-set is mentioned: if this distance is greater than 10, this means that the order of magnitude of that distance is greater than the order of magnitude of the length of the Pareto-set, while if this distance is less than 0.1, this means that the order of magnitude of that distance is smaller than the order of magnitude of the length of the Pareto-set.

The general picture that emerges from the study developed in this work is: Pareto-dominance based multiobjective optimization algorithms will present a two-stage convergence to the Pareto-set. In the first stage, a set of solutions initially far from the Pareto-set will converge to a region $\mathcal{R}$ which contains the Pareto-set $\mathcal{P}$. The region $\mathcal{R}$ includes points which are distant from the Pareto-set up to distances of the order of the length of the Pareto-set itself, in the following sense:

$$\max_{\mathbf{x}_r \in \mathcal{R}, \mathbf{x}_p \in \mathcal{P}} \|\mathbf{x}_r - \mathbf{x}_p\| \approx \max_{\mathbf{x}_{p1}, \mathbf{x}_{p2} \in \mathcal{P}} \|\mathbf{x}_{p1} - \mathbf{x}_{p2}\| \tag{23}$$

In this expression, the notation (≈) means “is of the same order of magnitude as”. The right side of expression (23) is the “length” of the Pareto-set. This first stage is expected to be performed efficiently, with a high success rate in the generation of solutions that dominate the current ones. In the second stage the solutions, now located in $\mathcal{R}$, will converge to a vicinity of the Pareto-set $\mathcal{P}$:

$$\|\mathbf{x} - \bar{\mathbf{x}}_{\mathcal{P}}\| < \epsilon \tag{24}$$

in which $\mathbf{x} \in \mathcal{R}$ is a current solution, and $\bar{\mathbf{x}}_{\mathcal{P}} \in \mathcal{P}$ is its projection in the set $\mathcal{P}$. As ϵ becomes smaller, it becomes more difficult to achieve new solutions that enhance the current solution $\mathbf{x}$. The growth of the difficulty becomes faster for a greater number of objectives n, as described by the power law (22).

Then, an analysis of the progression of a basic Pareto-domination based evolutionary algorithm can be stated as follows:

New tentative solution points that dominate $\mathbf{x}_d$ are easily found when $\mathbf{x}_d$ is up to the distance of 1 from the Pareto-set, with a success rate of approximately $1 : 2$ (for each two attempts, one dominating point is found). This occurs no matter what is the number of objectives. This behavior may be interpreted as: in locations far from the Pareto-set, the objectives become mostly “non-conflicting”, i.e., the directions in which they are enhanced almost coincide, for all functions. In this region of $\lambda > 1$, the multiobjective optimization problem behaves like a single-objective optimization problem. Therefore, it is always easy to find the location of the Pareto-set, up to a resolution of the order of its length.

When $\mathbf{x}_d$ enters a region in which its distance to the Pareto-set is smaller than 1, the further enhancements become exponentially more difficult as the point approaches the Pareto-set, for any number of objectives $n \geq 2$, according to the power law (22). This occurs even in the cases of $n = 2$ and $n = 3$, which are not usually considered to give rise to "many-objective problems". The enhancement of a solution at a distance of $\lambda = 0.1$, in the case of $n = 3$, is performed with a success rate of approximately $6 : 10^3$. In this way, it is still practical to generate solutions that are "reasonably nearby" the Pareto-set, perhaps situated into a distance of 10^{-2} of it. Further enhancements in such solutions would be performed, in the case $n = 3$, at a success rate of $1 : 10^4$, which begins to become computationally hard. This explains the need, which has been pointed out in several references [21], [22], of a local search mechanism when solutions with higher precision are required in Pareto-dominance based multiobjective optimization, even for a small number of objective functions.

As n grows, the power law exponent γ_n also grows, according to the relation $\gamma_n = n - 1$, which holds for the computational search of most of the points belonging to the Pareto-set. In the case $n = 4$, which is the first situation considered usually as a "many-objective optimization", when the distance of $\mathbf{x}_d$ to the Pareto-set is smaller than 1, the enhancement of solutions which already have a precision of 10^{-3} would be performed with a success rate as low as $1 : 10^9$. For $n = 5$, the enhancement of solutions with such a precision would be performed with a success rate as low as $2 : 10^{12}$. Those numbers explain why the optimization of multiobjective problems with more than 4 objectives has been found to be a hard problem, in the context of Pareto-dominance methods.

VI. Discussion: Decomposition-based Algorithms

The difficulty associated to Pareto-domination schemes for solution ranking in algorithms for many-objective optimization was described earlier in this paper as the phenomenon of fast decrease of the relative size of the space region which contains points that dominate current solutions. The probability of generation of any such a point rapidly approaches zero, leading those algorithms to stagnate when the search reaches regions in which there is conflict between the (many) objectives. Analytical formulae were derived, showing a precise description of this phenomenon for isotropic mutation operators, in an artificial setting of a problem that was built with several symmetries that allowed the analysis. However, in more general situations, the same phenomenon will also occur, although not obeying exactly the same formulae. A very important observation is: the deep cause of the difficulty lies in the collapse of the size of the region that contains dominating solutions, and not in the particular structure of the isotropic mutation. Any stochastic search operator (non-isotropic mutation operators, crossover operators, etc) would face the same problem, if it is committed to finding new points that dominate current ones.

The empirical observation of this fact has driven research efforts to solve this problem into the general direction of developing solution ranking methods that do not rely on Pareto-dominance in order to induce a convergence pressure towards the Pareto-set. One of the main ideas that were pursued was to employ *decomposition-based algorithms*. Those algorithms try to solve multiobjective problems by decomposing them into several scalar problems. Since each resulting optimization problem is a single-optimization problem, it should not be affected by the difficulties related to many-objective optimization. A strong motivation of this approach was inherited from the classical nonlinear programming theory of multiobjective optimization [15], which showed that some formulations of scalar optimization problems that aggregate the different objectives are able to guarantee that: (i) any solution of the scalar problem is a Pareto-optimal solution; and (ii) any Pareto-optimal solution may be found by an instance of the scalar problem. In the discussion that follows, the conjunction of those properties will be referred as the *Pareto-equivalence property*.

Even before the subject of many-objective optimization received so much attention, there were attempts to use decomposition methods in evolutionary multiobjective optimization, which resulted in important evolutionary algorithms, such as the MOEA/D [23]. In recent years, a large number of works has been devoted to the development of decomposition-based algorithms in order to deal with many-objective optimization problems. However, it is noticeable that some scalarization procedures that were employed successfully in the traditional setting of few objectives were not found to be good choices in the many-objective setting, while one scalarization procedure that was not so popular before became the most used one in the new setting. In this section, the analytical tools that were developed in this paper are applied in order to provide an interpretation of those issues.

1) Weighted sum:

$$\min F(\mathbf{x}) = \sum_{i=1}^{n} \omega_i f_i(\mathbf{x}) \tag{25}$$

In this case, the problem behaves as a single-objective problem, in which the evolutionary algorithm finds an enhancing point with probability 0.5, provided that the new tentative point is generated sufficiently near the current point. This is due to the fact that a weighted sum of smooth functions is a smooth function and therefore its contour curve passing on a point $\mathbf{x}_0$ will have a tangent plane on that point which will represent an asymptotic local approximation of the contour curve on $\mathbf{x}_0$. Therefore, any direction from $\mathbf{x}_0$ pointing into the semi-space which (at least locally) contains the contour curves will be an enhancing direction for $F(\mathbf{x})$ on $\mathbf{x}_0$. The main objections to this approach are its lack of ability to find points on non-convex portions of the Pareto-set (the Pareto-equivalence property does not hold) and its difficulty in relating the values of the weights and the region of the Pareto-set to be sampled. The initial proposition of the MOEA/D algorithm [23] proposed the weighted sum scalarization as one of its three recommended decomposition methods. However, perhaps due to that objections, this scalarization procedure has not reached popularity in subsequent works, neither in the multiobjective setting, nor in the many-objective one. In the recent literature on many-objective optimization, the weighted sum approach appears in [24], as one among four different scalarization methods.

2) Weighted Tchebyschev norm:

$$\min F(\mathbf{x}) = \max_i \ \omega_i(f_i(\mathbf{x}) - f_i^{ref}) \tag{26}$$

In this expression, $\mathbf{f}^{ref} = (f_1^{ref}, f_2^{ref}, \ldots, f_n^{ref})$ stands for a reference point in the objective space. This scalarization procedure was also used within MOEA/D [23]. This scalar problem is endowed with the Pareto-equivalence property. Perhaps due to this reason, this scalarization scheme became very popular in applications of MOEA/D. However, in the recent research on many-objective optimization, this procedure has not achieved popularity.

This phenomenon can be explained using the tools proposed here. In the case of the weighted Tchebyschev norm, in a small vicinity of $\mathbf{x}$, the region in which $F(\mathbf{x})$ decreases corresponds to the interior of the contour curve of the function $f_j(\mathbf{x})$, for the j for which occurs the maximum value of $\omega_j(f_j(\mathbf{x}) - f_j^{ref})$. This means that, if such a maximizing j is unique, a tangent plane to a contour curve of $f_j(\mathbf{x})$ will define a whole semi-space of enhancing directions for $F(\mathbf{x})$ as in the case of the weighted sum of objectives. However, it should be recalled that the region in which all the functions $f_i(\mathbf{x})$ become enhanced simultaneously will shrink as the number of objectives grows, as shown in this paper. This means that the enhancement of $f_j(\mathbf{x})$ will tend to occur with a degradation of other objective functions $f_i(\mathbf{x})$. So, on each step of the search procedure, the new point $\mathbf{x}^{k+1}$ will tend to be closer to the curve $\mathcal{F}$ defined by

$$\mathcal{F} = \Big\{ \mathbf{x} \mid \omega_1(f_1(\mathbf{x}) - f_1^{ref}) = \omega_2(f_2(\mathbf{x}) - f_2^{ref}) = \ldots = \omega_n(f_n(\mathbf{x}) - f_n^{ref}) \Big\} \tag{27}$$

than its predecessor $\mathbf{x}^k$. Once on that curve, any new enhancement on the scalar function $F(\mathbf{x})$ will occur only by finding new points that dominate the current one – in this way falling back into the same problem of Pareto-dominance search. This process of the search sequence converging to the curve $\mathcal{F}$ is quite similar to the well-known Maratos effect that may appear when classical nonlinear programming line search algorithms are applied to non-smooth problems [25]. The search sequences in those line search algorithms tend to be attracted to the region in which the function is non-smooth, becoming trapped in that region.

In this way, it is expected that the minimization of a scalar Tchebyschev decomposition as shown in (26) goes farther than a pure Pareto-dominance search, because the algorithm presents some progress towards the Pareto-set while the sequence approaches the curve $\mathcal{F}$. However, it is possible that the search sequence converges to $\mathcal{F}$ before reaching the Pareto-set, becoming trapped on that curve.

3) Weighted 2-norm:

$$\min F(\mathbf{x}) = \sqrt{\sum_{i=1}^{n} \omega_i^2 (f_i(\mathbf{x}) - f_i^{ref})^2} \tag{28}$$

The problem (28) is traditionally known as the *goal programming* formulation. Similarly to the case of weighted sum scalarization, the problem here behaves as a single-objective smooth problem, provided that the objective functions are smooth. The geometric reasoning here is the same as in the case of weighted sum: again, the scalar function $F(\mathbf{x})$ is a smooth function, and a tangent plane to a contour curve will locally define a whole semi-space of enhancing directions. In this case, differently from the case of weighted sum, it becomes possible to manage the reference point f^{ref}, in order to find any point on the Pareto-set. However, the formulation does not present the Pareto-equivalence property, since non Pareto-optimal points can be generated. However, the main difficulty of formulation (28) is related to the choice of the suitable locations for the reference points.

It is noticeable that this kind of decomposition strategy was not that popular in decomposition algorithms in the classical multiobjective setting, with few objectives. However, in the context of many-objective problems, this approach has been used in several recent references. For instance, [24] employs this scalarization procedure directly as one of the ranking methods to be used within its algorithm. The references [26] and [27] employ this scalarization procedure within a two-step ranking routine, in which a goal programming ranking subproblem is solved after the candidate solutions are assigned to their respective reference points. The references [28], [29] also employ weighted 2-norms in order to rank solutions.

4) Goal-attainment formulation:

$$\begin{aligned} &\min F(\mathbf{x}) = \alpha \\ &\text{s.t. } f_i(\mathbf{x}) \le f_i^{ref} + \alpha\, \omega_i \end{aligned} \tag{29}$$

This is another classical formulation of a scalarization for a multiobjective problem [15], which also complies with the Pareto-equivalence property. This formulation was employed within some evolutionary multiobjective optimization algorithms (see, for instance, [22]). In MOEA/D [23], a slightly different formulation was employed, as follows:

$$\begin{aligned} &\min F(\mathbf{x}) = \alpha \\ &\text{s.t. } f_i(\mathbf{x}) = f_i^{ref} + \alpha\, \omega_i \end{aligned} \tag{30}$$

This formulation is called the *Boundary Intersection Approach* in [23]. It is easy to verify that this formulation does not present the Pareto-equivalence property, since it may deliver output points that are not Pareto-optimal. In order to manage the equality constraint, the reference [23] proposed the usage of the following penalty approach:

$$\min F(\mathbf{x}) = d_1(\mathbf{x}) + \theta d_2(\mathbf{x}) \tag{31}$$

in which:

$$\begin{aligned} d_1(\mathbf{x}) &= \frac{\|(\mathbf{f}(\mathbf{x}) - \mathbf{f}^{ref}) \cdot \boldsymbol{\omega}\|}{\|\boldsymbol{\omega}\|} \\ d_2(\mathbf{x}) &= \|\mathbf{f}(\mathbf{x}) - (\mathbf{f}^{ref} - d_1 \boldsymbol{\omega})\| \end{aligned} \tag{32}$$

This formulation was called the *Penalty-based Boundary Intersection*, or PBI, in [23]. This function was used as the third recommended scalarization function in that reference.

It is interesting to notice that the goal attainment scalarization in its original formulation (29) has a geometry that is equivalent to the geometry of the weighted Tchebyschev norm minimization, leading to an equivalent behavior. This means that such a formulation would present the same difficulties than the scalarization via weighted Tchebyschev norm, in dealing with many-objective problems.

The PBI approach, on the other hand, presents a penalty factor θ which can modulate the algorithm behavior. In any case, except for $\theta = 0$, the contour curves of the function $F(\mathbf{x})$ will be non-smooth, with a vertex that will lie on the reference direction. For $\theta = 1$, the behavior of PBI will be identical to the behavior of the weighted Tchebyschev norm. For $0 < \theta \leq 1$, the region that contains solutions that enhance the PBI of a previous solution will be greater than in the case of the weighted Tchebyschev norm, while for $\theta > 1$ this region will be smaller. In any case, it is expected that the Maratos effect that was predicted for the weighted Tchebyschev norm will also occur for the PBI solution ranking, with the solution sequence being attracted to the reference line, without further convergence to the exact Pareto-set after the convergence to that line.

The PBI function inspired the decomposition employed in the references [26], [27], [29], which employed a hierarchical composition of d_1 and d_2, and it was used again in its original format in [30].

VII. Conclusion

This paper presented an analytical study of the phenomenon of performance degradation of Pareto-dominance based search algorithms in many-objective problems with continuous variables. Within a simple setup of n quadratic objective functions with spherical contour curves and a symmetrical arrangement of the function minima location, some analytical formulae that describe the probability of progress of the optimization process as a function of the distance λ to the exact Pareto-set were derived.

The main conclusion that was obtained about the nature and structure of the performance degradation phenomenon in many-objective problems can be stated as the observation that the search occurs in two distinct phases:

(i) When the current tentative solution is at a distance λ greater than the length of the Pareto-set, the probability of success in enhancing the current solution is high;

(ii) When the current solution reaches a λ which has its order of magnitude smaller than that of the length of the Pareto-set, the probability of finding a new point that dominates the current one is given by a power law function of λ with exponent $(n-1)$.

(iii) The dimension of the decision variable space, N, has no influence on that exponent, influencing only the multiplicative coefficient of the power law.

Those conclusions were derived considering isotropic mutation operators, in an artificial setting of a problem that was built with several symmetries that allowed the analysis. However, the same phenomenon will occur in more general situations of many nonlinear and smooth objective functions, although not obeying exactly the same formulae. It is important to notice that the deep cause of the convergence difficulty of Pareto-based ranking schemes lies in the collapse of the size of the region that contains dominating solutions, and not in the particular structure of the isotropic mutation. Any stochastic search operator (non-isotropic mutation operators, crossover operators, etc) would face the same problem in the task of finding new points that dominate current ones.

The results obtained here provide a theoretical framework that articulates most of the previous studies about the structure of many-objective optimization problems, as well as explains the causes of success or failure of algorithmic schemes proposed in the last few years.

VIII. Acknowledgments

The authors would like to thank Sebastian Xavier, adjunct professor at Federal University of Ouro Preto, for his valuable suggestions in preparing this article.

This work was supported by the Brazilian agencies CNPq, CAPES and FAPEMIG. This research was also supported by a Marie Curie International Research Staff Exchange Scheme Fellowship within the 7th European Community Framework Programme.

References

[1] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multiobjective optimization," *Evolutionary Computation*, vol. 7, no. 3, pp. 205–230, 1995.

[2] E. Zitzler, L. Thiele, and K. Deb, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evolutionary Computation*, vol. 8, no. 2, pp. 173–195, 2000.

[3] C. A. C. Coello, "An updated survey of GA-based multiobjective optimization techniques," *ACM Computing Surveys*, vol. 32, no. 2, 2000.

[4] T. Hanne, "Global multiobjective optimization with evolutionary algorithms: Selection mechanisms and mutation control," in *Proceedings of the 1st International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, ser. LNCS, no. 1993. Faro, Portugal: Springer, 2001, pp. 197–212.

[5] K. Ikeda, H. Kita, and S. Kobayashi, "Failure of Pareto-based MOEAs: Does non-dominated really mean near to optimal?" in *Proceedings of the 2001 IEEE Congress on Evolutionary Computation (CEC 2001)*, 2001, pp. 957–962.

[6] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable multi-objective optimization test problems," in *Proceedings of the 2002 IEEE Congress on Evolutionary Computation (CEC 2002)*. Honolulu, HI: IEEE, 2002, pp. 825–830.

[7] V. Khare, X. Yao, and K. Deb, "Performance scaling of multi-objective evolutionary algorithms," in *Proceedings of the 2nd International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003)*, Faro, Portugal, 2003, pp. 376–390.

[8] R. C. Purshouse and P. J. Fleming, "On the evolutionary optimization of many conflicting objectives," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 770–784, 2007.

[9] C. von Lucken, B. Baran, and C. Brizuela, "A survey on multi-objective evolutionary algorithms for many-objective problems," *Computational Optimization and Applications*, vol. 58, pp. 707–756, 2014.

[10] B. Li, J. Li, K. Tang, and X. Yao, "Many-objective evolutionary algorithms: a survey," *ACM Computing Surveys*, vol. 48, no. 1, 2015, article 13.

[11] R. H. C. Takahashi, E. G. Carrano, and E. F. Wanner, "On a stochastic differential equation approach for multiobjective optimization up to Pareto-criticality," in *Proceedings of the 6th International Conference on Evolutionary Multi-criterion Optimization (EMO 2011)*, Ouro Preto, Brazil, 2011, pp. 61–75.

[12] H. Aguirre and K. Tanaka, "A hybrid scalarization and adaptive epsilon-ranking strategy for many-objective optimization," in *Proceedings of the 11th International Conference on Parallel Problem Solving from Nature (PPSN XI)*, Krakow, Poland, 2010, pp. 11–20.

[13] H. Ishibuchi, Y. Hitotsuyanagi, H. Ohyanagi, and Y. Nojima, "Effects of the existence of highly correlated objectives on the behavior of MOEA/D," in *Proceedings of the 6th International Conference on Evolutionary Multi-criterion Optimization (EMO 2011)*, Ouro Preto, Brazil, 2011, pp. 166–181.

[14] H. Ishibuchi, N. Akedo, and Y. Nojima, "Behavior of multiobjective evolutionary algorithms on many-objective knapsack problems," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 2, pp. 264–283, 2015.

[15] V. Chankong and Y. Y. Haimes, *Multiobjective Decision Making: Theory and Methodology*. Elsevier, 1983.

[16] O. Schuetze, A. Lara, and C. A. Coello Coello, "On the influence of the number of objectives on the hardness of a multiobjective optimization problem," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 4, pp. 444–455, 2011.

[17] P. Loskot and N. C. Beaulieu, "On monotonicity of the hypersphere volume and area," *Journal of Geometry*, vol. 87, pp. pp. 96–98, 2007.

[18] S. Li, "Concise formulas for the area and volume of a hyperspherical cap," *Asian Journal of Mathematics and Statistics*, vol. 4, no. 1, pp. 66–70, 2011.

[19] G. E. Andrews, R. Askey, and R. Roy, *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2011, ch. Special Functions.

[20] H. K. Singh, A. Isaacs, and T. Ray, "A Pareto corner search evolutionary algorithm and dimensionality reduction in many-objective optimization problems," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 4, pp. 539–556, 2011.

[21] D. Sharma, A. Kumar, K. Deb, and K. Sindhya, "Hybridization of SBX based NSGA-II and sequential quadratic programming for solving multi-objective optimization problems," in *Proceedings of the 2007 IEEE Congress on Evolutionary Computation (CEC 2007)*, Singapore, Singapore, 2007, pp. 3003–3010.

[22] E. F. Wanner, F. G. Guimares, R. H. C. Takahashi, and P. J. Fleming, "Local search with quadratic approximations into memetic algorithms for optimization with multiple criteria," *Evolutionary Computation*, vol. 16, no. 2, pp. 185–224, 2008.

[23] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.

[24] H. Karshenas, R. Santana, C. Bielza, and P. Larranaga, "Multiobjective estimation of distribution algorithm based on joint modeling of objectives and variables," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 519–542, 2014.

[25] E. R. Panier and A. L. Tits, "Avoiding the Maratos effect by means of a nonmonotone line search I. General constrained problems," *SIAM Journal on Numerical Analysis*, vol. 28, no. 4, pp. 1183–1195, 1991.

[26] H. Jain and K. Deb, "An improved adaptive approach for elitist nondominated sorting genetic algorithm for many-objective optimization," in *Proceedings of the 7th International Conference on Evolutionary Multiobjective Optimization (EMO 2013)*, Sheffield, UK, 2013, pp. 307–321.

[27] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, 2014.

[28] I. Giagkiozis, R. C. Purshouse, and P. J. Fleming, "Generalized decomposition," in *Proceedings of the 7th International Conference on Evolutionary Multi-Criterion Optimization (EMO 2013)*, Sheffield, UK, 2013, pp. 428–442.

[29] M. Asafuddoula, T. Ray, and R. Sarker, "A decomposition-based evolutionary algorithm for many objective optimization," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 3, pp. 445–460, 2015.

[30] K. Li, K. Deb, Q. Zhang, and S. Kwong, "An evolutionary many-objective optimization algorithm based on dominance and decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 19, no. 5, pp. 694–716, 2015.