

Methods for Many-Objective Optimization: An Analysis

Ioannis Giagkiozis, and Peter J. Fleming

Department of Automatic Control and Systems Engineering

The University of Sheffield

Research Report No. 1030

November 2012

Abstract—Decomposition-based methods are often cited as the solution to problems related with many-objective optimization. Decomposition-based methods employ a scalarizing function to reduce a many-objective problem into a set of single objective problems, which upon solution yields a good approximation of the set of optimal solutions. This set is commonly referred to as Pareto front. In this work we explore the implications of using decomposition-based methods over Pareto-based methods from a probabilistic point of view. Namely, we investigate whether there is an advantage of using a decomposition-based method, for example using the Chebyshev scalarizing function, over Pareto-based methods.

Index Terms—Many-Objective optimization, Chebychev decomposition, Pareto-based methods, Decomposition-based methods

I. INTRODUCTION

A traditional classification of optimisation problems has been their separation into linear and nonlinear problems [1]. However, this ignores a large body of research in convex programming, which is a special class of methods that apply to a subclass of nonlinear optimisation problems, for which a solution can be obtained with relative ease, even for relatively large scale problems [2]. Therefore a more relevant distinction of optimisation problems would be their classification based on convexity. This is so because this distinction will radically impact the approach used in solving such problems and the expected quality of the produced solutions. In the fortunate case that a problem can be expressed in a convex form, then for all practical purposes it can almost be considered as solved. Furthermore the solution obtained for convex problems is guaranteed to be the global optimum, however it is not necessarily unique. Namely, a solution obtained for a convex program is the global minimum (maximum) for a minimisation (maximisation) problem. Also such a solution can be obtained quite efficiently [2]. These facts strongly motivate the additional effort required to attempt to identify whether a problem is convex. However the process of identifying a convex problem is non-trivial and is further complicated by the fact that different formulations can be more difficult to solve [2].

For nonconvex problems, guarantees about the obtained solution can only be given when an exhaustive search is performed. That is, only if the entire domain of definition of the objective function is explored. Naturally such a task can very easily become unmanageable. However once the fact that a problem is nonconvex is established, there are several metaheuristics¹ that can be employed to obtain a solution. Some examples of metaheuristics, often referred to as evolutionary algorithms (EAs) in the literature are, genetic algorithms (GAs) [3], [4], evolution strategies (ES) [5], differential evolution (DE) [6] particle swarm optimisation (PSO) [7], [8] and others [9]–[11].

Although a solution produced by any of the aforementioned methods will most likely be suboptimal, metaheuristics perform *well* in practice. Meaning that compared to the alternative of using random search [12], [13], which has the property of asymptotical convergence [14], EAs in practice, converge *faster* to the neighbourhood of optimal solutions for a number of problems [15]–[18]. Of course, this does not imply that EAs are superior to random search for all problems. The implication is that if *domain knowledge* is exploited then EAs can be a very effective [19], especially in light of the fact that even convex problems turn into nonconvex at the slightest *provocation*. An example of this phenomenon is seen in machine learning, where kernel-based learning algorithms that have a *shallow* architecture, namely a single layer of kernels for which the weights are to be determined, are linear in the parameters and produce convex problems. However, such architectures seem to be inefficient for certain tasks while *deep* architectures, namely architectures with several layers of kernels, can learn more complex tasks more efficiently, however the estimation of their parameters (learning) is non-convex [20]. Such examples serve as feedback to practitioners not to become overly dependent on a particular method, but instead, carefully investigate the nature of the problem so as to select the *optimal* approach for its solution, a process that is nonconvex in itself.

Another important classification is the separation of optimisation problems into single-objective, that is problems where the objective function is a mapping of the type $f : C^m \rightarrow C$, and, multi-objective whereby the objective function mapping

The authors are with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK, S1 3JD.
E-mail: i.giakiozis@sheffield.ac.uk
URL: <http://ioannis-giakiozis.staff.shef.ac.uk>

¹An algorithmic framework of heuristic optimisation algorithms. Heuristic is the Greek word for *search*. Metaheuristic algorithms commonly have a stochastic (Greek word for *random*) component.

has the following form² $f : C^n \rightarrow C^k$. This classification is important because in the latter case, there is no obvious or unique way to induce a complete ordering for nonconvex [21, pp. 113], [1, pp. 61] as well as for convex problems [2, pp. 174]. Without order a direction of search cannot be established. This is a well known issue in multi-objective optimisation and has been addressed to varying degrees of success by several researchers in the field of mathematical programming [21]–[23] and more recently in evolutionary computation [4], [24], [25]. In general there are two approaches employed to resolve this issue: Pareto-based and decomposition-based methods. In both methodologies and assuming, the *a posteriori* preference articulation paradigm [1, pp. 63] is employed, the relative importance of the objectives is unknown. In the case that preference information is given by the decision maker (DM), then using a decomposition method to combine the scalar objective functions can be used, see Section IV. An alternative is to distill the preference information given by the decision maker in a utility function, however this requires extensive knowledge of the problem structure and does not guarantee that its solutions will be Pareto optimal [21, pp. 62]. Pareto-based methods use the Pareto-dominance relations [1] to induce partial ordering in the objective space.

Decomposition-based methods have been predominantly used in mathematical programming [1], [2]. These methods use scalarising functions to decompose a multi-objective problem into several single objective subproblems. These subproblems are defined with the help of weighting vectors. The weighting vectors are k -dimensional vectors with positive³ components whose sum is equal to one. The *location* on the Pareto front that each subproblem tends to converge, depends on the choice of weighting vectors and the scalarizing function of choice. Therefore, to control the final distribution of solutions on the Pareto front the set of weighting vectors must be carefully selected for the entire Pareto front to be covered.

Many-objective problems are an important subclass of multi-objective problems that have more than 3 objectives, with applications in control and aerospace [9], [26]. However, as many authors point out, Pareto-based evolutionary algorithms are facing difficulties for many-objective problems for a number of reasons. For example, for increasing number of dimensions the number of incomparable solutions dominates the population, hence the selection pressure is massively reduced which leads to poor convergence rate to the Pareto front [27]. A proposed solution is to increase the population size [28], however this approach is of limited use since it is easy to overrun the available computational resources for even a 10 objective problem. To see this consider the fact that to obtain a Pareto optimal set for a 10-objective problem, that has comparable quality with the resulting Pareto set of a 2-objective problem and a population size of 100, the size of the set must be approximately 10^{10} [29]. This obviates the fact that *throwing* more resources at the problem is not an

option. Another problem that Pareto-based methods face for many objectives is that it is unclear how to preserve diversity in the solutions. This problem hints toward the fact that perhaps in many-objective problems we should redefine the objectives of a *posteriori* preference articulation based algorithms to a smaller region of the Pareto front.

Some authors allege that the solution is to use decomposition-based algorithms since they scale well for large population sizes and seem to have better convergence rate compared with Pareto-based algorithms [28], a view that seems to be adopted by several authors [30]–[33] and as can be seen by the number of publications based on the MOEA/D algorithm introduced in [34]. However, although the experimental results seem to support this view, the difference of decomposition-based with Pareto-based algorithms is not impressive if relative performance is to be considered. Additionally, decomposition-based methods have their fair share of difficulties. For instance, a straightforward method to distribute the solutions on the Pareto front seems elusive to obtain for decomposition-based methods. This deficiency stems from the fact that it is not straightforward to select the weighting vectors and the scalarizing function as most results available in the literature apply only to convex optimization problems [1], [21], [35]. There is however one scalarizing function, namely the Chebyshev scalarizing function, that can be used for nonconvex problems as well since the produced solutions will at least be weakly Pareto optimal⁴ and there is a theorem that states that all Pareto optimal solutions can be obtained for some weighting vector [1, pp. 99]. Perhaps this is the reason for the increased use of this scalarizing function in the literature, see for example [34], [36]–[38].

To date there is no theoretical evidence to support the above-mentioned view. Namely, that decomposition-based methods are superior to Pareto-based methods for many-objective problems. Some studies have appeared in the literature, for example [29], [39] but the assumption is that the objective function is unimodal, i.e. convex or quasi-convex. This assumption limits the scope of these works since evolutionary algorithms (EAs) are applied to nonconvex problems, that is problems that classical optimization methods fail or are inefficient. In this work we study the difficulties that EAs face in many-objective problems and explore the difference of Pareto-based and dominance-based methods for this class of problems. Our prior assumptions about the problem structure are much more relaxed and realistic compared with [29].

The main contributions of this work can be summarized as follows:

- The effect of Pareto dominance methods is studied from a theoretical perspective and an explanation of the difficulties experienced by several Pareto-based algorithms is presented.
- Decomposition-based methods are also studied and their relation to dominance methods is clarified. A major result is that methods based on the Chebyshev scalarizing function are equivalent to methods based on Pareto-dominance under certain assumptions that are usually

²In this work we can assume $C = \mathbb{R}$, since we only consider problems with real variables.

³On some occasions this constraint is relaxed to allow for non-negative weighting vector components, i.e. zeros are allowed in the weighting vector.

⁴See Section II for definition.

trivially met in decomposition-based algorithms.

- Lastly, given some prior information about the Pareto front geometry the *optimal* scalarizing function is identified. Optimal in the sense that with this scalarizing function the probability of finding a better solution, given a starting point \mathbf{z}_c , will have a slower rate of decrease compared to other scalarizing functions and at the same time similar guarantees provided by the Chebyshev scalarizing functions can be given.

The remainder of this paper is structured as follows. In Section II fundamental concepts pertaining to many-objective problems and set ordering relations are introduced. In Section III we discuss Pareto-based methods and explore the effect of dominance relations in many-objective problems. Furthermore in Section IV we perform a similar analysis as was conducted for Pareto-based methods, for a popular class of decomposition methods based on the weighted metrics scalarizing functions. In Section V we show that similar assurances as the ones provided by the Chebyshev scalarizing function can be given for an ℓ_p -norm based decomposition function with $p < \infty$. Furthermore, in Section VI we reflect on the consequences of the presented results in this work and present contexts in which our results can be used constructively to improve algorithms tackling many-objective problems. Lastly in Section VII, this work is summarized and concluded.

II. FUNDAMENTAL CONCEPTS AND DEFINITIONS

A multi-objective optimisation problem is defined as:

$$\begin{aligned} \min_{\mathbf{x}} \mathbf{F}(\mathbf{x}) &= (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_k(\mathbf{x})) \\ \text{subject to } \mathbf{x} &\in S, \end{aligned} \quad (1)$$

where k is the number of scalar objective functions and \mathbf{x} is the decision vector with a domain of definition $S \subseteq \mathbb{R}^n$, while Z is the objective space and is the forward image⁵ of S under the mapping \mathbf{F} . When the number of objectives, k , is more than 3 then the problem defined by (1) is referred to as many-objective. The issue with many-objective problems⁶ is that a complete ordering cannot be defined without the help of a decision maker (DM). This in turn creates difficulties for evolutionary algorithms (EAs) that use Pareto-dominance methods for fitness assignment. Pareto-dominance relations define a partial ordering in objective space, therefore enabling the comparison of solutions. Pareto-dominance relations are usually denoted by the \preceq binary relation. Also, for any two vectors $\mathbf{a}, \mathbf{b} \in Z$, the expression $\mathbf{a} \preceq \mathbf{b}$ is interpreted as: solution \mathbf{a} dominates \mathbf{b} in the context of a minimization problem. The aforementioned relation holds when all the elements of \mathbf{a} are smaller or equal to the corresponding elements in \mathbf{b} and at least one element is strictly smaller. Another family of methods used for fitness assignment is based on scalarizing functions, usually referred to as decomposition methods. In decomposition-based fitness assignment, the objective function is weighted and aggregated using a scalarising function, thus transforming (1) to a single-objective problem. However, it is

common for problems with more than one objective that a family of solutions are generated. This family of solutions, in the absence of prior information, should be representative of the entire Pareto front, whenever possible. The Pareto front is the minimal set of the set of feasible solutions in objective space.

A binary relation $R (\leq, \preceq)$ on a set C is said to be a **partial ordering** [40, pp. 7] if,

- R is reflexive: xRx for every $x \in C$.
- R is transitive: if xRy and $yRz \implies xRz$.
- R is antisymmetric, namely if xRy and $yRx \implies x = y$.

If a partial ordering is defined on a set C , then this set is said to be **partially ordered** or a **poset**. For a partially ordered set, it may happen that the relation R does not hold for all elements of the set, so for a binary relation⁷ \leq the following is a possibility: for $x, y \in C$, $x \not\leq y$ and $y \not\leq x$, in which case the elements x, y are incomparable - this is exactly the reason why such a relation is called *partial* ordering for the set C . For example one way to extend the \leq relation from \mathbb{R} to \mathbb{R}^2 is to define it as the application of the common \leq relation to the elements of the vectors in \mathbb{R}^2 , namely, $(x_1, x_2) \leq (y_1, y_2) \iff x_1 \leq y_1 \text{ and } x_2 \leq y_2$. Then this relation is a partial ordering for the set $C = \mathbb{R}^2$, that is for $x = (2, 2), y = (3, 4), z = (6, 5)$, $x \leq y$, also since $x \leq y$ and $y \leq z \implies (2, 2) \leq (6, 5) = x \leq z$. Furthermore, a binary relation $R (<, <)$ on a set C is said to be a **strict partial ordering** [41] if,

- R is irreflexive: $\neg(xRx)$ for every $x \in C$.
- R is transitive: if xRy and $yRz \implies xRz$.
- R is asymmetric, namely if $xRy \implies \neg(yRx)$.

If there exist no incomparable elements in a set, C , under the binary relation R , then C is said to be **ordered** (equivalently, **linearly** or **simply** ordered) and R is said to be a **complete ordering** on the set C , or equivalently **linear** or **simple** ordering.

A set $C \subseteq \mathbb{R}^n$ is *convex* if for any $\mathbf{x}, \mathbf{y} \in C$ and any $\theta \in [0, 1]$,

$$\theta\mathbf{x} + (1 - \theta)\mathbf{y} \in C. \quad (2)$$

A set C is a *cone* if $\theta\mathbf{x} \in C$ for all $\mathbf{x} \in C$ and $\theta \geq 0$. A cone $C \subset \mathbb{R}^n$ is called a *proper cone* if it is convex, closed, *pointed* and has nonempty interior. A cone is *pointed* if it contains no line, for example the cone, $C = \{(x, f(x)) \in \mathbb{R}^2 : f(x) > 0\}$, is not pointed since it contains an infinite number of lines: $f(x) = c, \forall x \in \mathbb{R}$, for any $c \in \mathbb{R}_+$. An example of a pointed cone is the nonnegative orthant \mathbb{R}_+^n , which is also a proper cone. Proper cones play a significant role in inducing partial ordering and are strongly related to the concept of Pareto optimality as will become clear in the next section.

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is said to be convex if the domain of definition of f , denoted as $\mathbf{dom} f$, is a convex set and $\forall \mathbf{x}, \mathbf{y} \in \mathbf{dom} f$ and $\theta \in [0, 1]$ we have,

$$f(\theta\mathbf{x} + (1 - \theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1 - \theta)f(\mathbf{y}). \quad (3)$$

⁵Namely, $\mathbf{F} : S \rightarrow Z$.

⁶As well as multi-objective problems

⁷Note that the relation, \leq , has a more abstract meaning and is not its usual definition. It is after all only a symbol.

A function is strictly convex if the inequality in (3) is strict. Accordingly a function is concave if $-f$ is convex. A more interesting definition of convex and concave functions is formulated with the aid of the *epigraph* of a function, see Appendix II-A. For an applications driven exposition on convex analysis and optimization the reader is referred to [2], while a more theoretical perspective can be found in [42], [43] and of course [44].

A. Epigraph

The *epigraph* of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined as:

$$\mathbf{epi} f = \{(\mathbf{x}, t) : \mathbf{x} \in \mathbf{dom} f, t \in \mathbb{R}, f(\mathbf{x}) \leq t\}, \quad (4)$$

consequently $\mathbf{epi} f \subset \mathbb{R}^{n+1}$. If the epigraph of a function is a convex set then the function is convex and vice versa. The *hypograph* of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, meaning *below the graph*, is defined as,

$$\mathbf{hypo} f = \{(\mathbf{x}, t) : \mathbf{x} \in \mathbf{dom} f, t \in \mathbb{R}, f(\mathbf{x}) \geq t\}. \quad (5)$$

If a function is concave, its hypograph is a convex set. In general a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ with a convex domain of definition is:

- Convex, if and only if $\mathbf{epi} f$ is a convex set. If in addition $\mathbf{hypo} f$ is nonconvex then, f is strictly convex.
- Concave, if and only if $\mathbf{hypo} f$ is a convex set. If in addition $\mathbf{epi} f$ is nonconvex then, f is strictly concave.
- Convex and concave, if both $\mathbf{epi} f$ and $\mathbf{hypo} f$ are convex. A concave and convex function is affine.
- Nonconvex, if both $\mathbf{epi} f$ and $\mathbf{hypo} f$ are nonconvex.

B. Pareto Front Geometry

Assuming that the Pareto front can be represented by a piecewise continuous function, $g : \mathbb{R}^{k-1} \rightarrow \mathbb{R}$ and k the number of objectives, then there are three types of *geometries* and combinations thereof, that the PF can have. Namely the function, g , can have parts that are convex, concave, or affine. We refer to a Pareto front as,

- Convex, if $\mathbf{epi} g$ is a convex set.
- Concave, if $\mathbf{hypo} g$ is a convex set.
- Affine, if both $\mathbf{epi} g$ and $\mathbf{hypo} g$ are convex.
- Discontinuous, if $\mathbf{dom} g$ is nonconvex or g is discontinuous.
- Partially convex, if g is convex over a convex subset of $\mathbf{dom} g$.
- Partially concave, if g is concave over a convex subset of $\mathbf{dom} g$.
- Partially affine, if g is convex and concave over a convex subset of $\mathbf{dom} g$.
- Piecewise convex, if g partially convex over all convex subsets of $\mathbf{dom} g$.
- Piecewise concave, if g partially concave over all convex subsets of $\mathbf{dom} g$.
- Piecewise affine, if g partially affine over all convex subsets of $\mathbf{dom} g$.

III. PARETO METHODS AND DERIVATIVES

A. Overview

In mathematical programming, the Pareto dominance relations are mainly used for theoretical purposes. However, in evolutionary computation they are heavily used in fitness assignment. Fitness assignment has a similar function to the negative gradient in gradient search - it indicates a promising direction of search. Therefore if such a direction is unavailable to the EA, then continuation of the search becomes increasingly more difficult as there is no indication that *better* solutions are being generated. This type of difficulty that EAs face in many-objective problems is described as *loss of selective pressure* in the EA literature [45].

If the relative importance of the objectives is unspecified, one way to partially order the objective vectors, $\mathbf{z} \in Z$, is to use the Pareto⁸ dominance relations, originally introduced by Edgeworth [46] and further studied by the economist Vilfredo Pareto [47]. A more general way to define dominance relations is by using generalised inequalities (\prec , \preceq) and the help of a proper cone K . A commonly used cone for this is the non-negative orthant, \mathbb{R}_+^k . So, for $K = \mathbb{R}_+^k$ and $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$, $\mathbf{a} \prec_K \mathbf{b}$ is true when⁹ $\mathbf{b} - \mathbf{a} \in \mathbf{int} K$, and, $\mathbf{a} \preceq_K \mathbf{b}$ when $\mathbf{b} - \mathbf{a} \in K$. However, since the non-negative orthant is almost always used to define generalised inequalities the subscript, K , is usually omitted. This notational convention is adopted in this work, so a subscript in generalised inequalities will be used only when the proper cone, K , is other than the non-negative orthant or the meaning is unclear from the context.

Specifically, in a minimisation context, a decision vector $\tilde{\mathbf{x}} \in S$ is said to be **Pareto optimal** if there is no other decision vector $\mathbf{x} \in S$ such that $f_i(\mathbf{x}) \leq f_i(\tilde{\mathbf{x}})$, for all i , and, $f_i(\mathbf{x}) < f_i(\tilde{\mathbf{x}})$ for at least one $i = 1, \dots, k$. Namely there exists no other decision vector that maps to a clearly superior objective vector. Similarly, a decision vector $\tilde{\mathbf{x}} \in S$ is said to be **weakly Pareto optimal** if there is no other decision vector $\mathbf{x} \in S$ such that $f_i(\mathbf{x}) < f_i(\tilde{\mathbf{x}})$ for all $i = 1, \dots, k$. Furthermore, a decision vector $\tilde{\mathbf{x}} \in S$ is said to **Pareto-dominate** a decision vector \mathbf{x} iff $f_i(\tilde{\mathbf{x}}) \leq f_i(\mathbf{x})$, $\forall i \in \{1, 2, \dots, k\}$ and $f_i(\tilde{\mathbf{x}}) < f_i(\mathbf{x})$, for at least one $i \in \{1, 2, \dots, k\}$ then $\tilde{\mathbf{x}} \preceq \mathbf{x}$. So, in terms of generalised inequalities, if $\mathbf{F}(\tilde{\mathbf{x}}) \preceq \mathbf{F}(\mathbf{x})$ and $\mathbf{F}(\tilde{\mathbf{x}}) \neq \mathbf{F}(\mathbf{x})$, then $\tilde{\mathbf{x}} \preceq \mathbf{x}$. Also, a decision vector $\tilde{\mathbf{x}} \in S$ is said to **strictly dominate**, in the Pareto sense, a decision vector \mathbf{x} iff $f_i(\tilde{\mathbf{x}}) < f_i(\mathbf{x})$, $\forall i \in \{1, 2, \dots, k\}$ then $\tilde{\mathbf{x}} \prec \mathbf{x}$. That is, if $\mathbf{F}(\tilde{\mathbf{x}}) \prec \mathbf{F}(\mathbf{x})$, then $\tilde{\mathbf{x}} \prec \mathbf{x}$. It should be noted at this point, that when \prec , \preceq are used in decision space, their meaning is mostly symbolic and is used to reflect the dominance relations in the objective space. For example, let $\mathbf{x}_1 = (0, 0, 0, 0)$, $\mathbf{x}_2 = (3, 3, 3, 3)$ and $f(\mathbf{x}_1) = (4, 4)$, $f(\mathbf{x}_2) = (1, 1)$. Clearly, for $K_s = \mathbb{R}_+^4$, $\mathbf{x}_1 \prec_{K_s} \mathbf{x}_2$, however according to the above definition of strict dominance it should be $\mathbf{x}_2 \prec \mathbf{x}_1$, because $\mathbf{F}(\mathbf{x}_2) \prec_{K_z} \mathbf{F}(\mathbf{x}_1)$ for $K_z = \mathbb{R}_+^2$. This can happen because the decision vectors are implicitly ordered according to their forward image in objective space, where the usual partial ordering, induced by the cone K_z is employed. Lastly, the

⁸Referred to as Edgeworth-Pareto dominance relations by some authors.

⁹The notation $\mathbf{int} K$ is used to denote the *interior* of the set K , in this case the cone K .

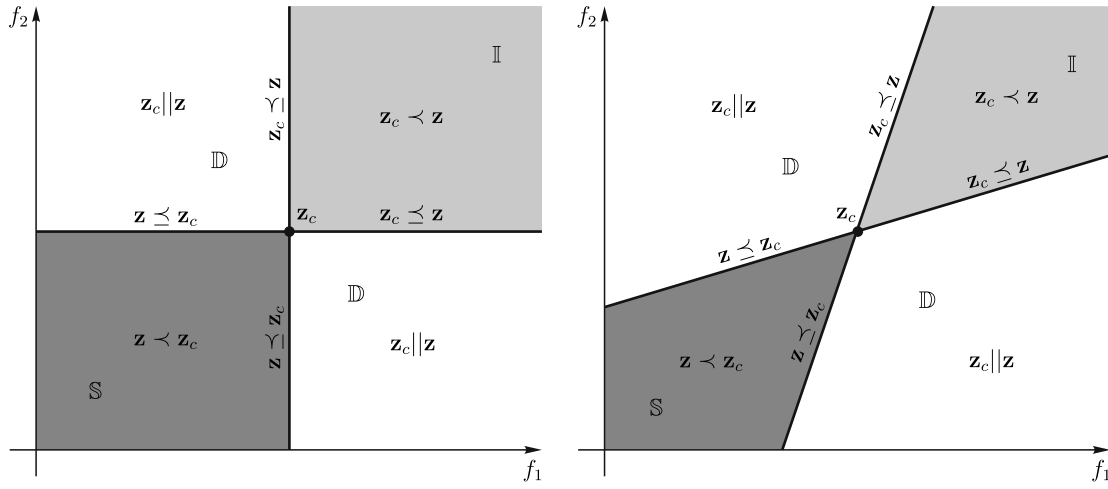


Fig. 1. **Left:** Dominance relations defined by a cone $K = \mathbb{R}_+^k$, in this instance $K = \mathbb{R}_+^2$. **Right:** Dominance relations defined by an *acute* proper cone $K = \{z : \sum_{i=1}^k \theta_i z_i, \theta_i \geq 0\}$ with z_i forming an acute angle with z_j for all $i \neq j$. The region, \mathbb{S} contains superior solutions to z_c and the region \mathbb{II} inferior solutions, while \mathbb{D} is the union of all the regions in, Z , that contain incomparable solutions to z_c . A mnemonic for the notation of the regions in the above figures is *Superior* for \mathbb{S} , *Inferior* for \mathbb{II} and *non-Dominated* for the \mathbb{D} regions respectively.

ordering induced by the binary relations \prec, \preceq is called *partial* because of the following possibility: $x, y \in Z$ but $x \not\preceq y$ and $y \not\preceq x$, in which case the vectors x, y are said to be **incomparable**. For example, the vectors $x = (3, 2, 1)$ and $y = (1, 2, 3)$ are incomparable. Dominance relations induced by two different proper cones are depicted in Fig. (1). Pareto-dominance relations and dominance relations imposed by the set $K = \mathbb{R}_+^n \setminus \{0\}$ are equivalent [1, pp. 24]. Notice that the 0 element is removed from K , this means that Pareto-dominance relations are not reflexive, i.e. $x \preceq_K x$ does not hold as, $x - x = 0 \notin K$.

Most multi-objective problem solvers attempt to identify a set of Pareto optimal solutions, this set is a subset of the **Pareto optimal set** (PS) which is also referred to as **Pareto front** (PF). The Pareto optimal set is defined as follows: $\mathcal{P} = \{z : \nexists \bar{z} \preceq z, \forall \bar{z} \in Z\}$, namely, it is the set of objective vectors that are not dominated by any objective vector in the feasible objective space. The decision vectors whose forward image under the objective function is the set, \mathcal{P} , are also referred to as the Pareto set and are denoted as \mathcal{D} , namely $\mathbf{F} : \mathcal{D} \rightarrow \mathcal{P}$. That is, the decision space is implicitly ordered according to the partial ordering applied to the objective space.

Algorithms based on Pareto dominance based methods for tackling multi-objective (and many-objective) problems have several difficulties to overcome. For instance a well distributed Pareto front is not guaranteed simply by using dominance relations. One answer to this problem has been presented in [48], where the authors introduce ε -dominance. In essence, ε -dominance creates a strict partial ordering based on the set $K_\varepsilon = \mathbb{R}_+^k + \varepsilon$. This type of dominance relation is useful to maintain well distributed Pareto optimal solutions [48], however it cannot escape the deficiency that Pareto-based methods face in many-objective problems [27]. Another very interesting approach introduced in [49], termed cone ε -dominance, uses the union of an *acute* proper cone and the set $K_\varepsilon = \mathbb{R}_+^k + \varepsilon$ to define a relation that is a partial ordering. Namely, they use ε -dominance in combination with an *acute*

cone, $K_\alpha = \{z : \sum_{i=1}^k \theta_i z_i, \theta_i \geq 0\}$ with z_i forming an acute angle with z_j for all $i \neq j$. Therefore, cone ε -dominance is defined with the help of the set $K_c = K_\varepsilon \cup K_\alpha$. The partial ordering induced by an *acute* cone is shown in Fig. (1). The motivation for the introduction of this type of Pareto dominance is that the diversity of produced Pareto optimal solutions seems to be better. Namely, cone ε -dominance promotes a good spread of solutions across the entire Pareto front, and their distribution seems to be more uniform. However, the problem reported in [27] persists for this type of dominance as well. In fact, since the regions where solutions become non-comparable are larger in cone ε -dominance it is expected that the number of non-dominated solutions increase more rapidly, compared to the Pareto dominance using $K = \mathbb{R}_+^k$, see Fig. (1).

B. Bias in the Objective Function

In the following sections of this work we assume that the objective function is *unbiased* or that it is not *biased* towards the Pareto front. This term is related to what the authors of the WFG¹⁰ toolkit [50] refer to as *bias* in the objective function. An objective function is considered to be *unbiased* when for decision vectors that are uniformly distributed in S the resulting distribution in objective space is also uniform, or close to uniform [50]. In this work we employ the same notion of *bias*, however we also provide a definition which should clarify the underlying assumptions of the statements: “an objective function has no bias”, or “an objective function is biased toward the Pareto front” etc. Specifically, let \mathcal{X} be an independent random deviate distributed according to, $\mathcal{U}(S)$, namely a uniform distribution in the feasible decision space, then we say that the objective function, \mathbf{F} , has no bias if,

$$\mathbf{F}(\mathcal{U}(S)) = \mathcal{Z} \sim \mathcal{U}(Z). \quad (6)$$

¹⁰Walking Fish Group. The WFG toolkit can be used to create scalable test problems in objective and decision space.

In other words, a uniformly sampled decision space maps to a uniformly sampled objective space, or at least approximately so. In some fortunate cases, the objective function is biased towards the Pareto front, namely,

$$\underbrace{\int \dots \int}_k \int_B h(z_1, \dots, z_k) dz_1 \dots dz_k > \mathbb{P}_{\mathcal{U}}(\mathcal{Z} \in B), \quad (7)$$

$$B = \{\mathbf{z} : \inf\{\|\mathbf{z} - \mathbf{z}_p\|\} \leq r, \mathbf{z}_p \in \mathcal{P}, \mathbf{z} \in \mathcal{Z}\},$$

where h , is the probability density function (pdf) of the objective space and B is the set of all feasible objective vectors with distance r or less from the Pareto front. So if the probability (the integral in (7)) to obtain a solution in B is larger for, h , than the uniform distribution then we say that the objective function is biased towards the Pareto front. If (7) holds for B^c , the complement of B , then we say that the objective function is biased away from the Pareto front. Similar definitions for bias toward any other region in objective space are trivial to define by simply changing the definition of the set B .

C. Pareto Dominance for Many-Objective Problems

In [27] the authors provide empirical results in an attempt to explain the reason for the *poor* performance of Pareto dominance-based algorithms applied to many-objective problems. The main argument is that the ratio of non-dominated (incomparable) individuals to the size of the population is approaching 1, meaning that almost the entire population is non-dominated, therefore the algorithms' selection mechanism is provided with no useful information. In what follows we elaborate further on this argument and prove that this behaviour is to be expected in many-objective problems and we reveal, to an extent, the underlying cause for such difficulties.

Consider the simplest multi-objective case, namely a 2-objective problem. Every point in objective space defines 4 regions shown in Fig. (1), (i) a region that contains solutions that are clearly better denoted as \mathbb{S} , (ii) a region that contains solutions that are clearly worse, \mathbb{I} , and (iii, iv) two regions where the solutions are incomparable to the point in question, \mathbb{D} . For 3-objective problems there are 8 such regions (2^3), however there is only 1 region which contains clearly better solutions and 1 region with clearly worse solutions. So, there are $2^3 - 2 = 6$ regions that contain solutions incomparable to the point in question. In general the following is true, for k -dimensional problems, there is always 1 region with clearly better solutions, 1 region with clearly worse solutions and $2^k - 2$ regions containing incomparable solutions. Furthermore, assuming that there is no bias towards any of these regions in the problem (objective function), the probability that a solution is generated in any one of these regions by a stochastic process (algorithm) is proportional to the volume of these regions divided by the volume of the entire feasible set in objective space¹¹, \mathcal{Z} . However, for increasing number of dimensions, the likelihood that a solution will be generated within the region \mathbb{S} , becomes almost insignificant the closer the point is to the

Pareto front. For example, for 10 dimensions there are 1 024 regions, hence for the above problem, the probability for a solution to be generated, that dominates the current point, is approximately $p = 1/1\,024 = 9.76 \cdot 10^{-4}$ if the point in question is exactly in the *middle* of a feasible objective space. To contrast this, the probability that a non-comparable solution is generated is $1\,022/1\,024 \approx 0.99$.

Although the assumption that the problem has no bias seems to limit the generality of the above argument, this is not entirely true. To illustrate this let us consider the relative *directions* of bias in the objective function in the context of optimization. These bias can be: (i) towards the Pareto front, namely it is easier to obtain solutions near the PF than in any other region, (ii) towards the region containing clearly worse solutions, and (iii) towards any region or regions containing incomparable solutions. Only in case (i) the solution of the optimisation problem becomes *easier* compared with the unbiased version. However this favourable scenario is seldom encountered in practice. So by assuming no bias in the objective function, all the probabilities that we calculate are in the worst case upper bounds on the probabilities of obtaining solutions in the set \mathbb{S} . In other words, the probabilities reported in this work represent the *best* attainable probability with respect to the location of an objective vector. We elaborate further on this point in Section VI.

To better appreciate and understand the reasons for the apparent difficulties that many-objective optimization algorithms face with such problems, we frame the aforementioned example on a more concrete basis. Assume that the objective space, \mathcal{Z} , is bounded from above by a hyperplane as shown in Fig. (2), specifically the upper bound is the set of points $M_{\mathcal{P}} = \{\mathbf{z} : \sum_{i=1}^k z_i = M, z_i \geq 0\}$. The reasons for selecting a feasible objective region with this particular geometry will become clear in what follows. Also, let the Pareto front be a $(k-1)$ -simplex, namely Pareto optimal objective vectors are part of the set $L_{\mathcal{P}} = \{\mathbf{z} : \sum_{i=1}^k z_i = L, z_i \geq 0\}$, obviously we have to select $L < M$ for minimization problems as $L > M$ would imply $\mathcal{Z} = \{\emptyset\}$. If we also assume that the problem has no bias, then given an objective vector, $\mathbf{z}_c \in \mathcal{Z}$, it would be possible to calculate the probability of obtaining a better solution for any point in the objective space. This information can be useful in many ways, we elaborate on those in Section VI.

Now, given a point in objective space, \mathbf{z}_c where the subscript is an abbreviation for *current point*, we can calculate the probability of obtaining a better solution using the following relation,

$$\mathbb{P}(\mathbf{z} \in \mathbb{S} | \mathbf{z}_c) = \frac{\mathcal{V}_{\mathbb{S}}(\mathbf{z}_c)}{\mathcal{V}_{\mathcal{Z}}}, \quad (8)$$

where, $\mathcal{V}_{\mathbb{S}}(\mathbf{z}_c) = \mathcal{V}_{\mathcal{P}}(\mathbf{z}_c)$, for Pareto-based methods, $\mathcal{V}_{\mathcal{Z}}$ is the *volume* of the feasible objective space which is equal to the volume of the slab in between $M_{\mathcal{P}}$, $L_{\mathcal{P}}$ and the positive orthant \mathbb{R}_+^k , see Fig. (2). Additionally, $\mathbb{P}(\mathbf{z} \in \mathbb{S} | \mathbf{z}_c)$, is the probability of finding a better objective vector, \mathbf{z}_n , given the objective vector \mathbf{z}_c . The expression in (8) is valid only for problems whose objective function would produce objective vectors uniformly distributed, or nearly so, given a set of uniformly distributed decision vectors. For biased problems knowledge of

¹¹We assume that the feasible objective set is bounded.

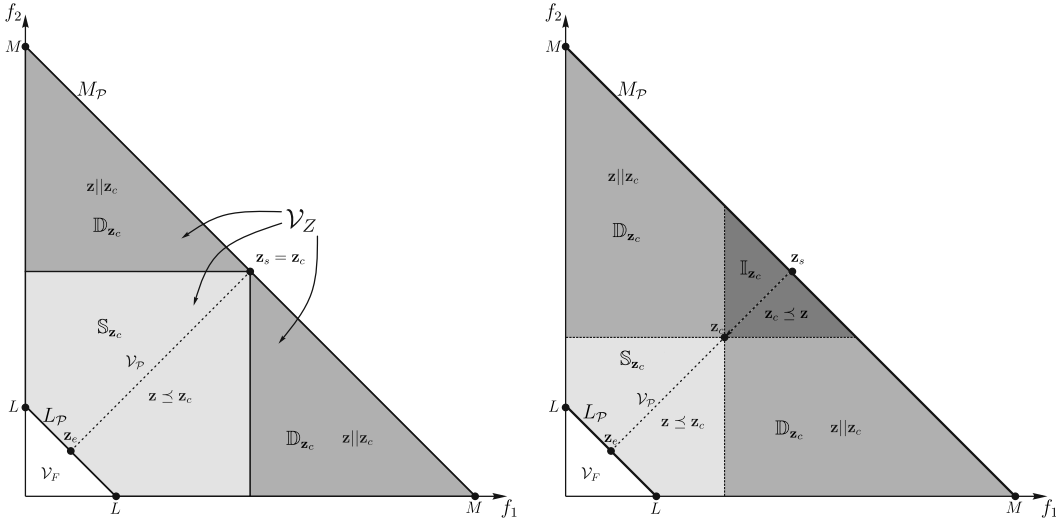


Fig. 2. Trajectory for the experiment described in Section III-D comparing decomposition and Pareto-based methods. M_P is the upper bound of the feasible objective space while L_P is the Pareto front and the lower bound of the feasible objective space. Also V_F is the volume below the Pareto front and V_Z is the volume of the feasible objective space, while V_P is the volume of the region containing superior solutions to the current solution z_c . Lastly, z_s and z_e are the starting and target objective vectors, with z_e being Pareto optimal. The **left** figure illustrates the aforementioned quantities for $z_c = z_s$ and the **right** figure illustrates how the above quantities change as z_c moves towards z_e along the $(z_e - z_s)$ direction. The results can be seen in Fig. (3).

the exact probability density function in objective space would be necessary so that we can *weigh* the integrals. However, as we mentioned above, in all but the most trivial problems the bias will be towards the Pareto front, otherwise it will be away from it, and so (8) will still describe a useful quantity, namely the upper bound of the probability of finding a better solution, assuming that there is no bias towards the Pareto front.

The volume of the region containing clearly better solutions, $V_P(z_c)$, for Pareto dominance or cone dominance using an ordering cone $K = \mathbb{R}_+^k$ is,

$$V_P(z) = \prod_{i=1}^k z_i - V_F, \quad (9)$$

where V_F is the volume of the non-dominated region beneath the Pareto front, which is the volume beneath the simplex, L_P . The $(k-1)$ -simplex corresponds to a Pareto front with affine geometry and V_F is calculated as,

$$V_L = \frac{\det [\mathbf{v}_1 \cdots \mathbf{v}_k]}{\Gamma(k+1)}. \quad (10)$$

Here, \mathbf{v}_i , are the vertices that the Pareto front intersects with the axes. The vectors, \mathbf{v}_i for the Pareto front are equal to $\mathbf{v}_i = L \cdot \mathbf{e}_i$, where \mathbf{e}_i is a vector of zeros and its i^{th} element is equal to one. Furthermore, the volume beneath the hyperplane M_P , V_M , is calculated using (10) and $\mathbf{v}_i = M \cdot \mathbf{e}_i$. Once V_M and V_L have been evaluated, the volume of the entire feasible objective space is calculated as,

$$V_Z = V_M - V_L. \quad (11)$$

Also the volume of the non-dominated region for ε -dominance is simply,

$$V_{P_\varepsilon}(z) = \prod_{i=1}^k (z_i - \varepsilon) - V_F, \quad (12)$$

assuming that the same ε value is used for every objective. If different values for ε are used it is trivial to modify (12). The volume of the non-dominated region for cone ε -dominance [49] is much more involved to calculate exactly, however, given that its defining set is the intersection of a proper cone and the set $\mathbb{R}^k + \varepsilon$ it stands to reason that its volume, V_{K_ε} , will be within,

$$V_{P_\varepsilon} \leq V_{K_\varepsilon} \leq V_P, \quad (13)$$

depending on the selected acute cone.

D. Experiment

Using (8)-(10) and a trajectory in objective space we can explore the change in the probability to obtain a solution in \mathbb{S} from a current point, z_c . So, assume we start from a point that is on the upper bound of the objective space, $z_s \in M_P$, and a target point on the Pareto front z_e , the question is how likely is to find a *better* solution with respect to any point on the trajectory with direction $z_e - z_s$, see Fig. (2). The trajectory that the point z_c follows can be seen in Fig. (2) is simply the line segment between z_s and z_e . This information for Pareto dominance methods will give us a basis for comparison with other methods for inducing a partial order in the objective space and should illuminate any differences. The steps involved, for this and for decomposition-based methods described in Section IV, as follows:

- Set $z_c = z_s$. Subsequently we divide the line segment from z_s to z_e into $N-1$ segments, thus from start to end there are N points $z_c[i] = z_s + (z_e - z_s) \frac{i}{N}$ and $i = 0, \dots, N-1$, see Fig. (2).
- For every $z_c[i]$ we calculate (8). The results are shown in Fig. (3)-(a) for Pareto-dominance.

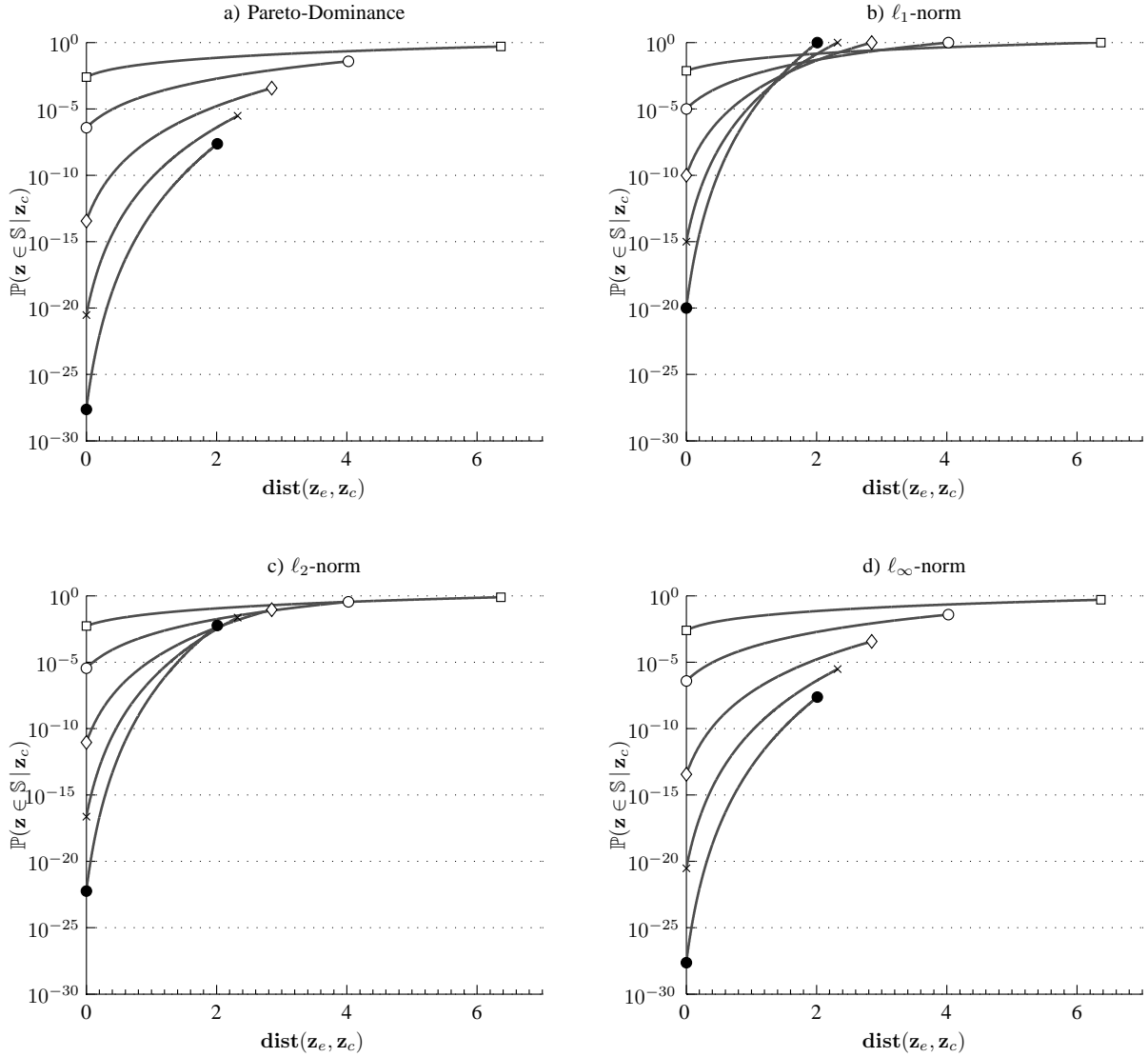


Fig. 3. Probability to find a better solution to \mathbf{z}_c , $\mathbb{P}(\mathbf{z} \in \mathbb{S} | \mathbf{z}_c)$, as a function of the Euclidean distance of the solution \mathbf{z}_c to \mathbf{z}_e , denoted by $\text{dist}(\mathbf{z}_e, \mathbf{z}_c)$, for different number of objectives (see Fig. (2)). Here $\{\square, \circ, \diamond, \times, \bullet\}$ correspond to $k = \{2, 5, 10, 15, 20\}$ objectives respectively.

IV. DECOMPOSITION METHODS

A. Overview

An alternative for defining a partial order in objective space, and of course an implicit partial order in decision space, can be found in decomposition methods. As mentioned in Section I, these methods employ a scalarizing function to aggregate all the objectives into a single scalar objective function. To obtain different Pareto optimal points, a set of weighting vectors can be used which would result in a set of single objective *subproblems*. This is the reason why such methods are called *decomposition-based*, it is because the employed strategy is to decompose a complex problem into a set of *simpler* ones. Simpler in this context does not necessarily mean easier to solve, it means that it is straightforward to apply standard EAs to the resulting subproblems.

The family of scalarizing functions that we focus our attention in this work, is the weighted metrics method [1, pp.

97] defined as:

$$\min_{\mathbf{x}} \left(\sum_{i=1}^k w_i |f_i(\mathbf{x}) - z_i^*|^p \right)^{\frac{1}{p}}, \quad (14)$$

where, w_i are the weighting coefficients, $w_i \geq 0$ for all $i = 1, \dots, k$, and $\sum_{i=1}^k w_i = 1$, also $p \in (0, \infty)$. The vector $\mathbf{z}^* = (z_1, \dots, z_k)$, is called the *ideal vector* and is defined as $\mathbf{z}^* = (\inf_{\mathbf{x}} \{f_1(\mathbf{x})\}, \dots, \inf_{\mathbf{x}} \{f_k(\mathbf{x})\})$. For the purpose of this work we will assume that $\mathbf{z}^* = (0, \dots, 0)$, which means that (14) can be rewritten as,

$$\min_{\mathbf{x}} \left(\sum_{i=1}^k w_i f_i(\mathbf{x})^p \right)^{\frac{1}{p}}. \quad (15)$$

Notice that we are allowed to remove the absolute value while maintaining the equivalency relation between (14) and (15), since, $\mathbf{z}^* = (0, \dots, 0)$, implies that $\mathbf{z} \in \mathbb{R}_+^k$. The formulation shown in (15) obviates the relationship of the weighted metrics scalarizing function with the weighting method and the

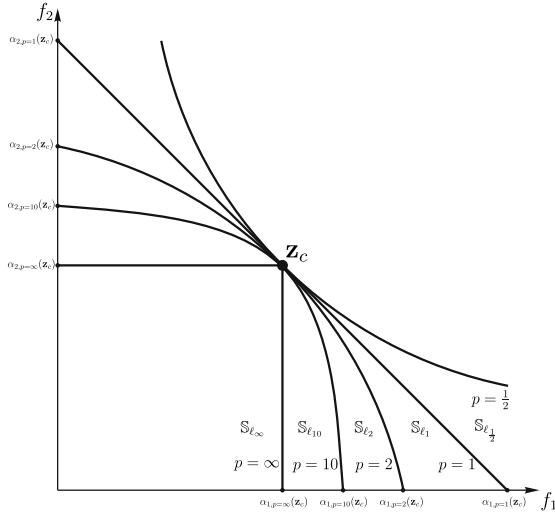


Fig. 4. The curves in the left figure represent the boundary of solutions that will be perceived as clearly better with respect to the corresponding p -norm. A geometric (although not entirely true) explanation as to why the Chebyshev scalarizing function ($p = \infty$) guarantees the generation of Pareto optimal solutions is seen to the right. In effect Chebyshev scalarizing function creates something that resembles a *asymptotically stable equilibrium*. This seems to be the case for any p -norm with $p > 1$.

Chebyshev decomposition. Namely, for $p = 1$ we obtain the weighting method [1, pp. 78],

$$\min_{\mathbf{x}} \sum_{i=1}^k w_i f_i(\mathbf{x}), \quad (16)$$

while for $p = \infty$ we obtain the Chebyshev scalarizing function,

$$\min_{\mathbf{x}} (\max\{w_1 f_1(\mathbf{x}), \dots, w_k f_k(\mathbf{x})\}). \quad (17)$$

A derivation of (17) from (15), is included in Appendix A for completeness. It should be noted that the assumption that the ideal vector is equal to the zero vector also implies that the objective function is bounded from below. In extension if the ideal vector is known and is not zero a change of *variables* in the objective function would be sufficient to meet our assumption. For example, for $\mathbf{z}^* = (-2, 4)$, it suffices to change the objective function, \mathbf{F} , with $\tilde{\mathbf{F}}(\mathbf{x}) = (f_1(\mathbf{x}) + 2, f_2(\mathbf{x}) - 4)$.

Although all norms are *equivalent*, in the sense that for every norm in a finite dimensional space multiplicative constants can be found relating two norms [2, pp. 636], their effect in an optimization problem can be significantly different, depending on the intricacies of the problem. For example, for $p = \infty$, namely the Chebyshev scalarizing function, there exist theoretical results stating that the solutions of (17) will be at least weakly Pareto optimal for any weighting vector $\mathbf{w} \in \mathbb{R}_+^k$ and that any Pareto optimal solution can be obtained for some weighting vector [1, pp. 99]. The interest of the MOEA community with respect to this particular norm is that the previous statement holds for nonconvex problems as well. Note that this does not imply that there is a guarantee that the algorithm will be able to find a Pareto optimal solution for a nonconvex problem, rather the statement refers to the equivalency of the two problems. In other words, assuming

that the selected algorithm is able to solve the problem defined in (17) then the solution will be at least a weakly Pareto optimal, and that all the Pareto optimal solutions can be obtained for some weighting vector. Such a result does not exist for $p < \infty$. In Section V we show that, given some prior information, it is possible to find a norm other than infinity with the same properties mentioned above. Namely, the ability of the a scalarized problem to converge to a weakly Pareto optimal solution for every weighting vector $\mathbf{w} \succ \mathbf{0}$ and that all Pareto optimal solutions can be reached.

However, it is not obvious as to why a norm, other than the ℓ_∞ -norm that is employed in the Chebyshev scalarizing function, would be more useful for decomposing a many-objective problem. For this reason we extend the experiment conducted for Pareto-based methods to decomposition-based methods that employ (15) as the scalarizing function to decompose a many-objective problem and study the effects that different values of p have on the resulting subproblems, see Section IV-B.

B. Decomposition Methods for Many-Objective Problems

The difference between scalarizing functions and the various forms of dominance relations discussed in Section III, is that the former define a complete ordering in the objective space, namely for a subproblem defined as,

$$g_p(\mathbf{x}) = \left(\sum_{i=1}^k w_i |f_i(\mathbf{x}) - z_i^*|^p \right)^{\frac{1}{p}}, \quad (18)$$

then for any two decision vectors $\mathbf{x}, \tilde{\mathbf{x}} \in Z$, only one of the following relations obtains,

$$g_p(\tilde{\mathbf{x}}) < g_p(\mathbf{x}), \text{ or } g_p(\tilde{\mathbf{x}}) = g_p(\mathbf{x}), \text{ or } g_p(\tilde{\mathbf{x}}) > g_p(\mathbf{x}),$$

given that the two decision vectors are applied to the same subproblem. Namely, regions containing incomparable solutions (regions \mathbb{D} in Fig. (1)) are eliminated, and depending on the ℓ_p -norm used in (15), parts of the \mathbb{D} regions are absorbed by the region containing inferior solutions, \mathbb{I} , and the region containing clearly better solutions, \mathbb{S} . This phenomenon has the potential to reduce the rate of decrease of the probability that a better solution is generated as the current solution approaches the optimal point, see Fig. (3)-(b-d). A better solution in this context is a solution that yields a lower value for the selected scalarizing function. In turn, this can reduce algorithm stagnation due to large number of non-dominated solutions as is the case in Pareto-based methods [27]. To see this consider a scenario in which the weighted sum method is used. In this scenario the weighting vector represents the normal of a hyperplane that divides the feasible objective space in two partitions. One, a region containing better solutions, \mathbb{S}_{ℓ_1} , and one with worse solutions, \mathbb{I}_{ℓ_1} , shown in Fig. (4). Solutions above the hyperplane are considered to be *worse* while solutions below the hyperplane are taken to be *better* with respect to the particular subproblem. Therefore, since the volume of the \mathbb{S} region is larger comparatively to dominance-based methods, it would be easier for the algorithm to identify solutions that are somewhat closer to the front with respect to

the currently best objective vector. However we have made a concession here, as the new solution may not Pareto-dominate the previous best solution. We will return to this issue in Section V and Section VI.

To explore how decomposition-based methods relate to Pareto-based methods, we must be able to calculate (8) for every $p = (0, \infty]$. The volume of the feasible objective space is calculated in the same way as in (11), while the volume of the \mathbb{S} region for $p = (0, \infty)$ is calculated as:

$$\mathcal{V}_{\mathbb{S}_{\ell_p}}(\mathbf{z}) = \frac{\left(\Gamma\left(1 + \frac{1}{p}\right)\right)^k}{\Gamma\left(\frac{k}{p} + 1\right)} \cdot \prod_{i=1}^k \alpha_i(\mathbf{z}) - \mathcal{V}_F, \quad (19)$$

which is essentially the volume of the positive orthant of a hyperellipsoid calculated as seen in [51]. The factors $\alpha_i(\mathbf{z})$ represent the distance of the ideal vector from the intersection of the ellipsoid with the positive axis of the i^{th} objective, shown in Fig. (4) and are calculated as,

$$\alpha_i(\mathbf{z}) = \left(\frac{\sum_{m=1}^k w_m z_m^p}{w_i} \right)^{\frac{1}{p}}, \quad (20)$$

see [51]. Since for the special case that $p = \infty$,

$$\lim_{p \rightarrow \infty} \frac{\left(\Gamma\left(1 + \frac{1}{p}\right)\right)^k}{\Gamma\left(\frac{k}{p} + 1\right)} = 1, \quad (21)$$

the volume of the \mathbb{S} region becomes,

$$\mathcal{V}_{\mathbb{S}_{\ell_\infty}}(\mathbf{z}) = \alpha_1(\mathbf{z}) \dots \alpha_k(\mathbf{z}) - \mathcal{V}_F, \quad (22)$$

and,

$$\alpha_i(\mathbf{z}) = \frac{\max\{w_1 z_1, \dots, w_k z_k\}}{w_i}. \quad (23)$$

Furthermore, to replicate the selected trajectory described in Section III-C and shown in Fig. (2), the weighting vector is set to $\mathbf{w} = \frac{1}{k} \cdot (1, \dots, 1)$ ascribing equal importance to all objectives so the resulting subproblem will tend to follow this trajectory and converge to the point \mathbf{z}_e . For this particular weighting vector (22) becomes,

$$\begin{aligned} \max\{w_1 z_1, \dots, w_k z_k\} &= w_m z_m, \\ \mathcal{V}_{\mathbb{S}_{\ell_\infty}}(\mathbf{z}) &= \frac{\left(\frac{1}{k}\right)^k z_m^k}{\left(\frac{1}{k}\right)^k} - \mathcal{V}_F = z_m^k - \mathcal{V}_F. \end{aligned} \quad (24)$$

However, as can be seen in Fig. (2), all points in the trajectory from \mathbf{z}_s to \mathbf{z}_e have $z_1 = z_2 = \dots = z_k$, hence $z_m = z_i$ for all $i = 1, \dots, k$, thus (24) can be calculated for any point on the trajectory.

As seen in Fig. (3)-(a-d), the probability to find a better solution as \mathbf{z}_c approaches the optimal solution \mathbf{z}_e decreases more rapidly for the Chebyshev scalarizing function and Pareto-based methods when compared to scalarizing functions employing the ℓ_1 -, ℓ_2 -norm. However, the results for the Chebyshev scalarizing function are remarkably similar to the Pareto-based method. In fact, for this trajectory, the two are identical, see (9) and (24). This interesting result means that Pareto-based methods and decomposition-based methods using

the Chebyshev scalarizing function are identical in the sense that,

$$\mathcal{V}_{\mathbb{S}_{\ell_\infty}} = \mathcal{V}_P. \quad (25)$$

This result is quite intriguing given the increased number of reports showing decomposition-based algorithms outperforming their Pareto-based counterparts for many-objective problems (cite some). However, we have only shown that the above equality holds for one particular trajectory and not necessarily for every possible trajectory towards any point on the Pareto front. We claim that (25) holds for an entire family of trajectories and that these particular trajectories are the ones that both decomposition and dominance-based algorithms attempt to follow in their approach towards the PF.

Before the general case is examined, consider a subproblem defined by the following weighting vector,

$$\begin{aligned} w_i &= \frac{2i}{k(k+1)}, \\ \mathbf{w} &= \left(\frac{2}{k(k+1)}, \dots, \frac{2k}{k(k+1)} \right) \end{aligned} \quad (26)$$

furthermore the selected starting point \mathbf{z}_s , is,

$$\mathbf{z}_s = M \cdot \left(\frac{k(k+1)}{2}, \dots, \frac{k(k+1)}{2k} \right) \quad (27)$$

and the end solution, \mathbf{z}_e , that is the target Pareto optimal solution is,

$$\mathbf{z}_e = L \cdot \left(\frac{k(k+1)}{2}, \dots, \frac{k(k+1)}{2k} \right). \quad (28)$$

Therefore any solution on the trajectory will be,

$$\begin{aligned} C &\in [L, M], \\ \mathbf{z}_c &= C \cdot \left(\frac{k(k+1)}{2}, \dots, \frac{k(k+1)}{2k} \right). \end{aligned} \quad (29)$$

Given (26)-(29),

$$\mathcal{V}_P(\mathbf{z}_c) = \prod_{i=1}^k z_i - \mathcal{V}_F = \frac{(Ck(k+1))^k}{2^k k!} - \mathcal{V}_F. \quad (30)$$

To calculate the volume of \mathbb{S} for the Chebyshev scalarizing function we need to find what is the maximum element. However, a closer look at (26) and (29) will reveal that this element is simply C , hence,

$$\mathcal{V}_{\mathbb{S}_{\ell_\infty}}(\mathbf{z}_c) = \prod_{i=1}^k \alpha_i - \mathcal{V}_F = \frac{(Ck(k+1))^k}{2^k k!} - \mathcal{V}_F, \quad (31)$$

so the relation $\mathcal{V}_{\mathbb{S}_{\ell_\infty}}(\mathbf{z}_c) = \mathcal{V}_P(\mathbf{z}_c)$ holds for this weighting vector as well. At this point we need to justify the assumption that a solution will attempt to follow the trajectory defined by the weighting vector in (26), since it appears to be artificial. For this we refer to the work by Ballesterio [52] where he refers to this trajectory as *well-balanced baskets* due to the relation,

$$w_1 z_1 = w_2 z_2 = \dots = w_k z_k, \quad (32)$$

for a solution $\mathbf{z} \in Z$. This essentially describes the *action* of the scalarizing function on the objective vector, which is to minimize the largest deviation in the given ℓ_p -norm. This is

most easily observed in the ℓ_∞ -norm used by the Chebyshev decomposition whereby only the largest deviation is taken into account thus reeling the solution toward the *balanced* trajectory. For example given an objective vector $\mathbf{z} = (1, 1.1, 1)$ and a weighting vector¹² $\mathbf{w} = (0.\overline{33}, 0.\overline{33}, 0.\overline{33})$, the ℓ_∞ -norm will attempt to minimize the second component of, \mathbf{z} , simply because $\max\{0.33 \cdot 1, 0.33 \cdot 1.1, 0.33 \cdot 1\} = 0.3667$. By this reasoning, when the ℓ_∞ -norm is used in a minimization problem, the *focus* of the algorithm will be to maintain the Hadamard product $\mathbf{w} \circ \mathbf{z}$ as close as possible to the vector $M \cdot \mathbf{1}$ while attempting to reduce M . By changing the weighting vector, this *equilibrium* that the Chebyshev scalarizing function is attempting to maintain, changes, so a different trajectory is followed, which of course converges to a different Pareto optimal point if the optimization algorithm is successful. Well that trajectory is found by finding the objective vector that *sends* the weighting vector \mathbf{w} to the unit vector.

Therefore for any given weighting vector,

$$\mathbf{w} = \left(\frac{c_1}{s}, \dots, \frac{c_k}{s} \right), \quad s = \sum_{i=1}^k c_i, \quad (33)$$

the balanced trajectory for the ℓ_p -norm with $p = (0, \infty)$ is the set of points given by,

$$\mathbf{z}_c = C \cdot \left(\left(\frac{s}{c_1} \right)^{\frac{1}{p}}, \dots, \left(\frac{s}{c_k} \right)^{\frac{1}{p}} \right), \quad s = \sum_{i=1}^k c_i, \quad c_i \in \mathbb{R}_+, \quad (34)$$

and for the ℓ_∞ -norm by,

$$\mathbf{z}_c = C \cdot \left(\frac{s}{c_1}, \dots, \frac{s}{c_k} \right), \quad s = \sum_{i=1}^k c_i, \quad c_i \in \mathbb{R}_+, \quad (35)$$

and therefore since,

$$\begin{aligned} \mathcal{V}_{\ell_\infty}(\mathbf{z}_c) &= \frac{(\max\{w_1 z_{c,1}, \dots, w_k z_{c,k}\})^k}{\prod_{i=1}^k w_i} - \mathcal{V}_F \\ &= \frac{C^k}{\prod_{i=1}^k c_i} - \mathcal{V}_F = \frac{(Cs)^k}{\prod_{i=1}^k c_i} - \mathcal{V}_F \end{aligned} \quad (36)$$

and,

$$\mathcal{V}_P(\mathbf{z}_c) = \prod_{i=1}^k k z_{c,i} - \mathcal{V}_F = \frac{(Cs)^k}{\prod_{i=1}^k c_i} - \mathcal{V}_F, \quad (37)$$

meaning that $\mathcal{V}_P(\mathbf{z}_c) = \mathcal{V}_{\ell_\infty}(\mathbf{z}_c)$ whenever the objective vectors are allowed to follow the balanced trajectory that the Chebyshev scalarizing function attempts to follow. Notice that this results also hold for for biased objective functions.

¹²An over-line a number denotes infinite repetition of the digits below, e.g. $0.\overline{33} = 0.33\dots$

It follows that for objective vectors following a balanced trajectory,

$$\mathcal{V}_{\ell_1} > \mathcal{V}_{\ell_2} > \dots > \mathcal{V}_{\ell_\infty} = \mathcal{V}_P. \quad (38)$$

A proof for (38) is given in Appendix A-A. Meaning that,

$$\begin{aligned} \mathbb{P}_{\ell_1}(\mathbf{z} \in \mathbb{S}_{\ell_1} | \mathbf{z}_c) &> \mathbb{P}_{\ell_2}(\mathbf{z} \in \mathbb{S}_{\ell_2} | \mathbf{z}_c) > \dots \\ &> \mathbb{P}_{\ell_\infty}(\mathbf{z} \in \mathbb{S}_{\ell_\infty} | \mathbf{z}_c) = \mathbb{P}_P(\mathbf{z} \in \mathbb{S} | \mathbf{z}_c), \end{aligned} \quad (39)$$

where $\mathbf{z} \in Z$ and \mathbb{S}_{ℓ_p} is the region containing *better* solutions according to the ℓ_p -norm version of the scalarizing function and $\mathbb{P}_{\ell_p}(\mathbf{z} \in \mathbb{S}_{\ell_p})$ is the probability of finding a better solution in \mathbb{S}_{ℓ_p} given that the current best solution is \mathbf{z}_c . The result in (39) can be read directly from Fig. (4).

V. SCALARIZATION AND STABILITY OF THE EQUIVALENT PROBLEM

The results in the previous section must be interpreted with care since (39) does not imply in any way that by using a scalarizing function based on a norm with $p < \infty$, all the Pareto optimal solutions will be *reachable*. However it does imply that by using a scalarizing function with p small, there is a better chance in finding better solutions with respect to that norm. Nevertheless, we require Pareto optimal solutions and not just any solutions that are closer to the front in some ℓ_p -norm, which means that if we cannot ensure that the subproblems are able to converge to Pareto optimal solutions and that all Pareto optimal solutions will be obtainable, the importance of (39) would be limited to the fact that Pareto-dominance methods are *equivalent* to decomposition-based methods that employ the Chebyshev scalarizing function. Equivalent in the sense that for an objective vector following a well balanced trajectory the probability to obtain a solution dominating the current solution is the same in both methods.

To understand the tradeoff between using a dominance-based method versus a decomposition-based method let us consider the effect of a scalarizing function to the objective space. A scalarizing function projects the entire objective space onto a line¹³, therefore some regions that contain incomparable solutions in the Pareto sense, now become solutions that are either better or worse for the particular subproblem. Therefore, a major difference between decomposition-based and Pareto-based algorithms is that the former provide unambiguous information about the quality of the produced solutions at every iteration while the latter cannot always guarantee such information because the likelihood of generating incomparable solutions is high for problems with many objectives [27]. However it is easy to reduce the above argument into a *zugzwang*¹⁴ between Pareto-based methods and decomposition-based methods. This is accomplished by the simple observation that the *clearly better* regions in the Chebyshev scalarizing function ($p = \infty$ in Fig. (4)) are

¹³In this work a segment of a ray, since the objective space is bounded.

¹⁴A chess terminology whereby the player whose turn is to play will be at a disadvantage, while had it not been his turn the outcome of the game would be unknown. The parallel in the context of this work is that both decomposition and Pareto-based methods are accompanied by a disadvantage, hence choice (a *move*) by the analyst leads to, different, but unfavorable results in both cases.

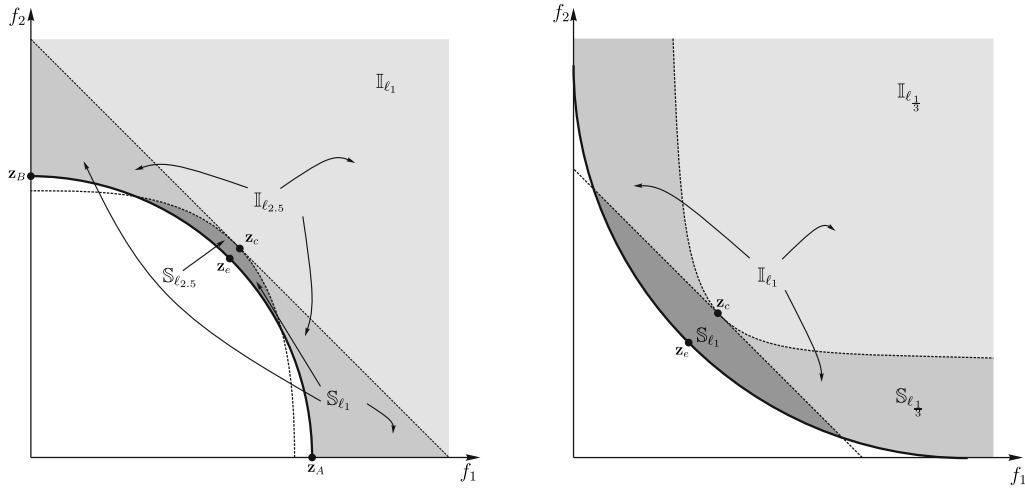


Fig. 5. Stable and unstable scalarizing functions.

identical to the regions generated by Pareto dominance based methods (Fig. (1)), while the incomparable and clearly worse regions in Pareto-based methods are mapped to *clearly worse* regions by the Chebyshev scalarizing function. Namely, if we require a decomposition method that can guarantee the generation of Pareto optimal solutions, then, we have to use the Chebyshev scalarizing function but in so doing we give up the favourable convergence rates¹⁵ achieved when using, for example the weighted sum method, and vice versa. In general there are two competing trends:

- As $p \rightarrow 0$, the probability of finding a better solution with respect to the ℓ_p -norm increases, hence it is less likely that the algorithm stagnates due to its inability to find direction of search. Additionally, it becomes increasingly more difficult to obtain all Pareto optimal solutions.
- However, as $p \rightarrow \infty$, we can obtain more Pareto optimal solutions on the Pareto front, but the probability to find a better solution with respect to the norm defined by p is also decreasing. In the limit, namely for $p = \infty$, we obtain the Chebyshev scalarizing function that guarantees that we will be able to find all Pareto optimal solutions for some weighting vector \mathbf{w} but this scalarizing function is equivalent with Pareto-dominance methods.

So the question is: is there a way that a scalarizing function can be used with p relatively small while preserving the guarantees that the Chebyshev function provides? The answer is affirmative for many-objective problems whose Pareto front geometry is continuous (see Section II). Specifically, a Pareto front can be described with the following parametrization,

$$f_1^{p_1} + f_2^{p_2} + \dots + f_k^{p_k} = C, \quad (40)$$

where $p_i > 0$ for all i and C is a positive constant. For simplicity we assume that $f_i \geq 0$. We claim that if the weighted metrics scalarizing function is used with $p = \max\{p_1, \dots, p_k\}$, then this scalarization will have the same guarantees as the Chebyshev function, given that our estimate of $\max\{p_1, \dots, p_k\}$ is correct and that the objective function is

continuous. The reason for this is illustrated in Fig. (5). To see this, consider that when \mathbf{z}_c reaches \mathbf{z}_e in Fig. (5), the volume of the region \mathbb{S}_{ℓ_1} is still positive, meaning that according to the ℓ_1 -norm there are still better solutions to the current solution. Continuing on the same line of reasoning, the solution \mathbf{z}_c will either converge to \mathbf{z}_A or \mathbf{z}_B since at these two locations there is no way that the ℓ_1 -norm to be improved. This result follows directly from (38) and the results in [51] for calculating the volume in (40), it follows that,

$$\lim_{\mathbf{z}_c \rightarrow \mathbf{z}_e} (\mathcal{V}_{\mathcal{P}_z} - \mathcal{V}_{\ell_p}) \leq 0, \quad (41)$$

when $p > \max\{p_i\}$. In which case we say that the scalarization is *stable* while if $p < \max\{p_i\}$ the scalarization is *unstable* and we have,

$$\lim_{\mathbf{z}_c \rightarrow \mathbf{z}_e} (\mathcal{V}_{\ell_p} - \mathcal{V}_{\mathcal{P}_z}) > 0. \quad (42)$$

Stability in terms of scalarizations is taken to mean the following:

- A subproblem of a many-objective problem is a **stable scalarization** if for a given weighting vector $\mathbf{w} \succ \mathbf{0}$, it is able to converge to a Pareto optimal solution $\mathbf{z}_e = (z_1, \dots, z_k)$, with $z_i > 0$ for every $i = 1, \dots, k$.
- Conversely, a subproblem is an **unstable scalarization** if for a given weighting vector $\mathbf{w} \succ \mathbf{0}$, it converges to a Pareto optimal solution \mathbf{z}_e with $z_i = 0$ for at least one $i \in \{1, \dots, k\}$.

Therefore if the Pareto front geometry is known and it can be expressed in terms of (40), then we can select the ℓ_p -norm that will have the maximum probability to produce better solutions while preserving the guarantee that the final population will be (weakly) Pareto optimal and that all the Pareto optimal solutions will be obtainable for some weighting vector.

VI. DISCUSSION

By calculating the probability to find a better solution, we have essentially turned the problem of extending a many-objective optimization algorithm into a functional optimization

¹⁵Or more correctly the potential for favourable convergence rates.

problem. Namely, the question that can now be posed is: “what is the *optimal* ℓ_p -norm for the scalarization and trajectory for an objective vector?”. By optimal trajectory we mean the trajectory in objective space that will present the least *resistance* to our optimization algorithm while simultaneously moving towards a Pareto optimal solutions as fast as possible. This question although very interesting, it has either a trivial answer: a straight line, or for biased problems we would need to have knowledge of the probability density function in objective space, something which in general is unavailable even for test problems. Therefore, we use a balanced trajectory, since this is in accord with the scalarizing functions, in the sense that this is the path that they tend to follow. Using this we investigated how the probability to obtain better solutions varies as a function of the distance of the current best solution and the sought for Pareto optimal solution. We found that this probability is largest the smaller the ℓ_p -norm is, with respect to p . This information can be used to reduce the difficulty of many-objective problems, to some extent.

However, we cannot simply use the smallest norm that is numerically feasible since with decreasing p the ability of a scalarizing function to converge to a particular point of the Pareto front is also reduced, hence, a concession must be made. Although, if the Pareto front is continuous and can be described in a parametric way (see (40)), an optimal value, p^* , can be obtained for which the decrease of the probability of finding a better solution is minimal while the ability of the scalarizing function of finding every Pareto optimal solution is retained. The optimal value of p , separates the family of scalarizing functions into two subclasses. First, values of $p < p^*$ produce *unstable* scalarizing functions and $p > p^*$ result in *stable* scalarizing functions. Here stability refers to the ability of the scalarizing function to converge to any point on the Pareto front, while instability refers to the opposite.

A way to *convexify* the Pareto front, and thus allowing for an arbitrarily *small*¹⁶ ℓ_p -norm to be used, has been proposed in [53]. Essentially, what the author of the aforementioned work suggested is that the objective function is raised to a power until the Pareto front becomes convex. Although this suggestion may seem intriguing, the effect of such a non-linear operation to the objective function would be, among other things, introducing bias in the objective function and potentially making the problem more difficult to solve in the case that the initial set of scalar objective functions are non-convex.

VII. CONCLUSION

Based on the results in Section III and Section IV we have seen that under mild conditions the Chebyshev function is identical to Pareto-dominance methods. Identical in the sense that, for a solution following a balanced trajectory, the reduction of probability to find a better solution is identical for both methods. This curious fact suggests that the decomposition-based methods are actually not *better* compared with Pareto-based methods. But if that is so, how can the results observed by several researchers for many-objective

problems be justified? Given the fact that the reported results are only *slightly* better in [30]–[33] our hypothesis is that the difference is simply due to the ease with which a constant direction of search in objective space can be maintained in decomposition-based methods, while the same is very difficult to achieve with Pareto-based methods. A good example of this behaviour is seen in MOGLS¹⁷ [54] when compared with MOEA/D in [34]. In the aforementioned work MOGLS was outperformed by MOEA/D, and as the authors note, one reason was that MOGLS generated different weighting vectors on every iteration. This amounts to an attempt to identify the entire Pareto front, but also means that the direction of search in objective space is not constant as is the case for MOEA/D. The same problem is present in Pareto-based methods, however there is no clear way for this situation to be remedied.

The results in this work show that:

- Pareto-dominance methods and the Chebyshev scalarizing function are equivalent, in the sense that neither method in itself, has better probability to find *superior* solutions. In fact the aforementioned probabilities are the same.
- Given some prior information about the problem, namely the geometry of the Pareto front, we can find the *optimal* scalarizing function. Optimal in this context means that using the above scalarizing function all Pareto optimal solutions will be obtainable for some weighting vector, and that, the probability of obtaining a better solution, with respect to the particular scalarizing function, decreases more *slowly* compared to all other scalarizing functions (and Pareto-dominance methods) that can provide the same guarantee of finding all Pareto optimal solutions.
- Using generalized decomposition (gD) [55], [56] in conjunction with the results in this work, the required weighting vectors for obtaining Pareto optimal solutions in specific locations on the Pareto front, can be identified for any ℓ_p -norm.

Some of the mentioned benefits apply only when we are able to identify the Pareto front geometry prior to obtaining Pareto optimal solutions. We have identified a solution to this problem and preliminary results seem very promising.

APPENDIX A

GAMMA FUNCTION DEFINITION AND NORM VOLUMES

The Γ function¹⁸ is defined as:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt. \quad (43)$$

For $x \in \mathbb{N}$,

$$\Gamma(x+1) = x!. \quad (44)$$

The **psi** or **digamma** function is [57],

$$\begin{aligned} \psi(x) &= \frac{d}{dx} ((\ln \Gamma(x))) \\ &= -\gamma - \frac{1}{x} + \sum_{n=1}^{\infty} \left(\frac{1}{n} - \frac{1}{x+n} \right). \end{aligned} \quad (45)$$

¹⁷Multi-Objective Genetic Local Search.

¹⁸Pronounced *gamma*.

¹⁶Small here refers to p .

The term γ is Euler's constant, see [57] for its definition.

Lemma A.1. For a vector, $\mathbf{x} \in \mathbb{R}^n$,

$$\lim_{p \rightarrow \infty} \|\mathbf{x}\|_p = \max\{|x_1|, \dots, |x_n|\}.$$

Proof: Let, $|x_i| = \max\{|x_1|, \dots, |x_n|\}$ and $x_i \neq 0$, then

$$(|x_i|^p)^{\frac{1}{p}} \leq (|x_1|^p + \dots + |x_n|^p)^{\frac{1}{p}} \leq (n|x_i|^p)^{\frac{1}{p}} \Rightarrow$$

$$1 \leq \left(\frac{|x_1|^p}{|x_i|^p} + \dots + \frac{|x_n|^p}{|x_i|^p} \right)^{\frac{1}{p}} \leq n^{\frac{1}{p}}$$

and since $\lim_{p \rightarrow \infty} n^{\frac{1}{p}} = 1$, it follows that,

$$\lim_{p \rightarrow \infty} \left(\frac{|x_1|^p}{|x_i|^p} + \dots + \frac{|x_n|^p}{|x_i|^p} \right)^{\frac{1}{p}} = 1 \Rightarrow$$

$$\lim_{p \rightarrow \infty} (|x_1|^p + \dots + |x_n|^p)^{\frac{1}{p}} = |x_i|.$$

In the case that $|x_i| = 0$ it follows that the vector $\mathbf{x} = \mathbf{0}$, hence, by the very definition of norms $\|\mathbf{x}\| = 0$. ■

A. Norm Volumes

We want to show that,

$$\mathcal{V}_{\mathbf{S}_{\ell_1}} > \mathcal{V}_{\mathbf{S}_{\ell_2}} > \dots > \mathcal{V}_{\mathbf{S}_{\ell_\infty}} = \mathcal{V}_p, \quad (46)$$

for a balanced trajectory. The term \mathcal{V}_F is omitted as it is constant and independent of the scalarizing function or the dominance method employed.

Proof: To show that (46) obtains, it would suffice to show that,

$$\mathcal{V}_{\mathbf{S}_{\ell_p}} > \mathcal{V}_{\mathbf{S}_{\ell_{p+h}}}, \quad (47)$$

for $h > 0$ and that,

$$\mathcal{V}_{\mathbf{S}_{\ell_p}} > \mathcal{V}_{\mathbf{S}_{\ell_\infty}} = \mathcal{V}_p, \quad (48)$$

for any $p \in (1, \infty)$. We have from (33), (34) and (35) that (19) expands to,

$$\mathcal{V}_{\mathbf{S}_{\ell_p}}(\mathbf{z}) = \frac{\left(\sum_{i=1}^k w_i z_i^p \right)^{\frac{k}{p}} \left(\Gamma \left(1 + \frac{1}{p} \right) \right)^k}{\left(\prod_{i=1}^k w_i \right)^{\frac{1}{p}} \Gamma \left(\frac{k}{p} + 1 \right)}$$

$$= \frac{(k)^{\frac{k}{p}} \left(\Gamma \left(1 + \frac{1}{p} \right) \right)^k}{\left(\prod_{i=1}^k w_i \right)^{\frac{1}{p}} \Gamma \left(\frac{k}{p} + 1 \right)}$$

furthermore for $M = \frac{1}{\prod_{i=1}^k w_i}$ with $w_i > 0$ for all $i = 1, \dots, k$ we have,

$$g(p) = \ln \left(\frac{(k)^{\frac{k}{p}} \left(\Gamma \left(1 + \frac{1}{p} \right) \right)^k}{\left(\prod_{i=1}^k w_i \right)^{\frac{1}{p}} \Gamma \left(\frac{k}{p} + 1 \right)} \right) \quad (50)$$

whose partial derivative with respect to p is,

$$g'(p) = -\frac{1}{p^2} \ln(M) - \frac{k}{p^2} \ln(k)$$

$$- \frac{k}{p^2} \psi \left(1 + \frac{1}{p} \right) + \frac{k}{p^2} \psi \left(1 + \frac{k}{p} \right) \quad (51)$$

where $\psi(\cdot)$ is the digamma function [57]. Since (51) is strictly negative it follows that $\mathcal{V}_{\mathbf{S}_{\ell_p}}$ is a monotone decreasing function of p . ■

REFERENCES

- [1] K. Miettinen, *Nonlinear Multiobjective Optimization*. Springer, 1999, vol. 12.
- [2] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [3] J. Holland, "Outline for a logical theory of adaptive systems," *Journal of the ACM (JACM)*, vol. 9, no. 3, pp. 297–314, 1962.
- [4] D. Goldberg and J. Holland, "Genetic Algorithms and Machine Learning," *Machine Learning*, vol. 3, no. 2, pp. 95–99, 1988.
- [5] I. Rechenberg, "Cybernetic Solution Path of An Experimental Problem," 1965.
- [6] R. Storn and K. Price, "Differential Evolution - A Simple and Efficient Adaptive Scheme for Global Optimization Over Continuous Spaces," *International Computer Science Institute-Publications TR*, 1995.
- [7] R. Eberhart and J. Kennedy, "A New Optimizer Using Particle Swarm Theory," in *International Symposium on Micro Machine and Human Science*. IEEE, 1995, pp. 39–43.
- [8] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," in *IEEE International Conference on Neural Networks*, vol. 4. IEEE, 1995, pp. 1942–1948.
- [9] P. Fleming and R. Purshouse, "Evolutionary Algorithms in Control Systems Engineering: A Survey," *Control Engineering Practice*, vol. 10, no. 11, pp. 1223–1241, 2002.
- [10] M. Hauschild and M. Pelikan, "A survey of estimation of distribution algorithms," 2011.
- [11] A. Zhou, B. Qu, H. Li, S. Zhao, P. Suganthan, and Q. Zhang, "Multiobjective evolutionary algorithms: A survey of the state-of-the-art," *Swarm and Evolutionary Computation*, 2011.
- [12] D. Karnopp, "Random search techniques for optimization problems," *Automatica*, vol. 1, no. 2-3, pp. 111–121, 1963.
- [13] F. Solis, "Minimization by random search techniques," *Mathematics of operations research*, pp. 19–30, 1981.
- [14] D. Wolpert and W. Macready, "No Free Lunch Theorems for Optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.
- [15] E. Zitzler and L. Thiele, "Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.
- [16] E. Zitzler, K. Deb, and L. Thiele, "Comparison of Multiobjective Evolutionary Algorithms: Empirical Results," *Evolutionary Computation*, vol. 8, no. 2, pp. 173–195, 2000.
- [17] R. Purshouse, C. Jalbă, and P. Fleming, "Preference-Driven Co-Evolutionary Algorithms Show Promise for Many-Objective Optimisation," in *Evolutionary Multi-Criterion Optimization*. Springer, 2011, pp. 136–150.
- [18] R. Wang, R. Purshouse, and P. Fleming, "Preference-Inspired Co-Evolutionary Algorithms for Many-objective Optimisation," *IEEE Transactions on Evolutionary Computation*, vol. PP, no. 99, p. 1, 2012.
- [19] N. Radcliffe and P. Surry, "Fundamental Limitations on Search Algorithms: Evolutionary Computing in Perspective," in *Computer Science Today*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 1995, vol. 1000, pp. 275–291, 10.1007/BFb0015249.
- [20] Y. Bengio and Y. LeCun, "Scaling Learning Algorithms Towards AI," *Large-Scale Kernel Machines*, vol. 34, 2007.
- [21] C. Vira and Y. Haimes, *Multiobjective Decision Making: Theory and Methodology*. North-Holland, 1983, no. 8.
- [22] H. Kuhn and A. Tucker, "Nonlinear Programming," in *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, vol. 1. University of California Press, Berkeley, 1951, pp. 481–492.
- [23] A. M. Geoffrion, "Solving Bicriterion Mathematical Programs," *Operations Research*, vol. 15, no. 1, pp. pp. 39–54, 1967. [Online]. Available: <http://www.jstor.org/stable/168510>
- [24] J. Schaffer, "Multiple Objective Optimization with Vector Evaluated Genetic Algorithms," in *Conference on Genetic Algorithms*. L. Erlbaum Associates Inc., 1985, pp. 93–100.
- [25] C. Fonseca and P. Fleming, "Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion and Generalization," in *Conference on Genetic Algorithms*, vol. 423, 1993, pp. 416–423.
- [26] E. Hughes, "Radar Waveform Optimisation as a Many-Objective Application Benchmark," in *Evolutionary Multi-Criterion Optimization*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2007, vol. 4403, pp. 700–714.
- [27] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary Many-Objective Optimization: A Short Review," in *IEEE Congress on Evolutionary Computation*, june 2008, pp. 2419–2426.

- [28] H. Ishibuchi, Y. Sakane, N. Tsukamoto, and Y. Nojima, "Evolutionary Many-Objective Optimization by NSGA-II and MOEA/D with Large Populations," in *IEEE International Conference on Systems, Man and Cybernetics*, oct. 2009, pp. 1758–1763.
- [29] O. Schutze, A. Lara, and C. 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, aug. 2011.
- [30] E. Hughes, "MSOPS-II: A General-Purpose Many-Objective Optimiser," in *IEEE Congress on Evolutionary Computation*, sept. 2007, pp. 3944–3951.
- [31] D. Hadka and P. Reed, "Diagnostic Assessment of Search Controls and Failure Modes in Many - Objective Evolutionary Optimization," *Evolutionary Computation*, 2011.
- [32] H. Ishibuchi, N. Akedo, H. Ohyanagi, and Y. Nojima, "Behavior of EMO Algorithms on Many-Objective Optimization Problems with Correlated Objectives," in *IEEE Congress on Evolutionary Computation*, june 2011, pp. 1465–1472.
- [33] K. Deb and H. Jain, "Handling Many-Objective Problems Using an Improved NSGA-II Procedure," in *IEEE Congress on Evolutionary Computation*, june 2012, pp. 1–8.
- [34] 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.
- [35] K. Miettinen and M. Mäkelä, "On Scalarizing Functions in Multiobjective Optimization," *OR Spectrum*, vol. 24, no. 2, pp. 193–213, 2002.
- [36] P. Chang, S. Chen, Q. Zhang, and J. Lin, "MOEA/D for flowshop scheduling problems," in *IEEE Congress on Evolutionary Computation*. IEEE, 2008, pp. 1433–1438.
- [37] H. Li and Q. Zhang, "Multiobjective Optimization Problems with Complicated Pareto Sets, MOEA/D and NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 2, pp. 284–302, 2009.
- [38] Y. yan Tan, Y. chang Jiao, H. Li, and X. kuan Wang, "MOEA/D + uniform design: A new version of MOEA/D for optimization problems with many objectives," *Computers & Operations Research*, no. 0, 2012.
- [39] R. Takahashi, E. Carrano, and E. Wanner, "On a Stochastic Differential Equation Approach for Multiobjective Optimization up to Pareto-Criticality," in *Evolutionary Multi-Criterion Optimization*, ser. Lecture Notes in Computer Science. Springer Berlin, 2011, vol. 6576, pp. 61–75.
- [40] A. Kolmogorov and S. Fomin, *Introductory Real Analysis*. Dover Publications, 1975.
- [41] J. Lin, "Maximal vectors and multi-objective optimization," *Journal of Optimization Theory and Applications*, vol. 18, no. 1, pp. 41–64, 1976.
- [42] D. Bertsekas, A. Nedi, A. Ozdaglar *et al.*, "Convex Analysis and Optimization," 2003.
- [43] D. Bertsekas, *Convex Optimization Theory*. Athena Scientific, 2009.
- [44] R. Rockafellar, *Convex Analysis*. Princeton University Press, 1970, vol. 28.
- [45] E. J. Hughes, "Evolutionary Many-Objective Optimisation: Many Once or One Many?" in *IEEE Congress on Evolutionary Computation*, vol. 1, 2005, pp. 222–227.
- [46] F. Edgeworth, *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. CK Paul, 1881, no. 10.
- [47] V. Pareto, "Cours D'Économie Politique," 1896.
- [48] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining Convergence and Diversity in Evolutionary Multiobjective Optimization," *Evolutionary Computation*, vol. 10, no. 3, pp. 263–282, 2002.
- [49] L. Batista, F. Campelo, F. Guimarães, and J. Ramírez, "Pareto Cone ϵ -Dominance: Improving Convergence and Diversity in Multiobjective Evolutionary Algorithms," in *Evolutionary Multi-Criterion Optimization*. Springer, 2011, pp. 76–90.
- [50] S. Huband, P. Hingston, L. Barone, and L. While, "A Review of Multiobjective Test Problems and A Scalable Test Problem Toolkit," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 5, pp. 477–506, 2006.
- [51] X. Wang, "Volumes of Generalized Unit Balls," *Mathematics Magazine*, vol. 78, no. 5, pp. 390–395, dec. 2005.
- [52] E. Ballestero, "Selecting the CP Metric: A Risk Aversion Approach," *European Journal of Operational Research*, vol. 97, no. 3, pp. 593–596, 1997.
- [53] D. Li, "Convexification of a Noninferior Frontier," *Journal of Optimization Theory and Applications*, vol. 88, pp. 177–196, 1996. [Online]. Available: <http://dx.doi.org/10.1007/BF02192028>
- [54] A. Jaskiewicz, "On the Performance of Multiple-Objective Genetic Local Search on the 0/1 Knapsack Problem - A comparative Experiment," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 4, pp. 402–412, 2002.
- [55] I. Giagkiozis, R. Purshouse, and P. Fleming, "Generalized decomposition and cross entropy methods for many-objective optimization," Department of Automatic Control and Systems Engineering, The University of Sheffield, Research Report No. 1029, November 2012.
- [56] —, "Generalized Decomposition," in *Evolutionary Multi-Criterion Optimization*, ser. Lecture Notes in Computer Science. Springer Berlin, 2013.
- [57] P. Sebah and X. Gourdon, "Introduction to the Gamma Function," 2002.



Ioannis Giagkiozis received the B.Sc. degree from TEI of Thessaloniki, Thessaloniki, Greece in 2009. He then obtained the M.Sc. degree in Control and Systems Engineering with Distinction from the University of Sheffield, Sheffield, U.K. in 2010, for which he was awarded the Nicholson Prize for most outstanding student. He joined the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, as a Research Associate in 2011 and is currently working towards a Ph.D. degree in multiobjective evolutionary algorithms.

gorithms.

His research interests are in many-objective optimization, estimation of distribution algorithms and applied convex optimization.



Peter J. Fleming received the B.Sc. and Ph.D. degrees from The Queen's University, Belfast, U.K. He joined the University of Sheffield as Professor of Industrial Systems and Control in 1991, having previously been with Syracuse University, NY, NASA, Langley, VA and the University of Wales, Bangor, U.K. He was founder of the Rolls-Royce University Technology Centre and was Director from 1993–2012. He was the head of Automatic Control and Systems Engineering from 1993 to 1999, Director of Research (Engineering) from 2001 to 2003, and

Pro Vice-Chancellor for External Relations from 2003 to 2008. His control and systems engineering research interests include multicriteria decision making, optimization, grid computing, and industrial applications of modeling, monitoring, and control. He has over 400 research publications, including six books, and his research interests have led to the development of close links with a variety of industries in sectors such as aerospace, power generation, food processing, pharmaceuticals, and manufacturing.

Prof. Fleming is a Fellow of the Royal Academy of Engineering, both a Fellow of, and Adviser to, the International Federation of Automatic Control, a Fellow of the Institution of Electrical Engineers, a Fellow of the Institute of Measurement and Control, and the Editor-in-Chief of the International Journal of Systems Science.