On the Influence of the Number of Objectives on the Hardness of a Multiobjective Optimization Problem

Oliver Schütze, Adriana Lara, and Carlos A. Coello Coello, *Senior Member, IEEE*

***Abstract*—In this paper, we study the influence of the number of objectives of a continuous multiobjective optimization problem on its hardness for evolution strategies which is of particular interest for many-objective optimization problems. To be more precise, we measure the hardness in terms of the evolution (or convergence) of the population toward the set of interest, the Pareto set. Previous related studies consider mainly the number of nondominated individuals within a population which greatly improved the understanding of the problem and has led to possible remedies. However, in certain cases this ansatz is not sophisticated enough to understand all phenomena, and can even be misleading. In this paper, we suggest alternatively to consider the probability to improve the situation of the population which can, to a certain extent, be measured by the sizes of the descent cones. As an example, we make some qualitative considerations on a general class of uni-modal test problems and conjecture that these problems get harder by adding an objective, but that this difference is practically not significant, and we support this by some empirical studies. Further, we address the scalability in the number of objectives observed in the literature. That is, we try to extract the challenges for the treatment of many-objective problems for evolution strategies based on our observations and use them to explain recent advances in this field.**

AQ:1 ***Index Terms*—XXX, XXX, XXX.**

I. Introduction

EVOLUTIONARY algorithms for the numerical treatment of multiobjective optimization problems (MOPs) have been studied intensively during the last few years (see [11], [8] and references therein). Typically, few objectives (i.e., mainly two or three) are being investigated resulting in a variety of very efficient algorithms. The consideration of many (i.e., more than three) objectives, however, is a relatively young field and is yet not studied thoroughly enough. With this paper, we want to contribute to this field by looking at the influence of the number k of objectives in a continuous MOP on the hardness of the problem. To be more precise, we try to understand the behavior of the evolution with respect to k by looking at the descent cones of the individuals of the populations. The resulting analysis is of qualitative nature; however, it can for instance be used to disprove a common belief, namely that the addition of an objective makes a problem per se harder. Further, the new ansatz can be used to explain recent advances in the field of evolutionary many-objective optimization, and is thus hopefully helpful for designers of evolutionary algorithms aimed to deal with such problems.

When investigating continuous MOPs with respect to k, two facts have to be considered: 1) the solution set, the so-called Pareto set, forms typically a $(k-1)$-dimensional set [22], and 2) the problem gets harder the more local solutions it contains and the smaller the basin of attraction for the global solutions are since then the chance increases that a population can get stuck in locally optimal regions. The choice of k has thus, by 1), a direct influence on the dimension of the Pareto set, and hence, also on the hardness of the problem. If, for instance, $N_2 = 100$ points are chosen to obtain a "sufficient" representation of a solution set for $k = 2$ in the Hausdorff sense (which is a typical value in the literature), in principle the practically intractable amount of $N_{15} = 100^{14} = 10^{28}$ elements is required to obtain the same approximation quality for $k = 15$. Even if the lower bound of $N_2 = 2$ elements is used to "represent" the Pareto set for $k = 2$, still $N_{15} = 16\,384$ elements are needed to obtain the same (low) approximation quality for $k = 15$ (see also [51] for a related discussion on the required number of comparisons with respect to k). As a possible remedy, one can in certain cases try to reduce the number of objectives (e.g., [6], [15], [27]) since in practice it may happen that several objectives are correlated. Since we are interested in the influence of k we will not follow that approach. Another more practical remedy researchers dealing with evolutionary many-objective optimization have chosen is to bound the population/archive size to a moderate (and hence tractable) number for all values of k (say, $N = 100$). We will, in the following, consider that scenario and will restrict ourselves to investigate the evolution of these N individuals toward the Pareto set. That is, we will only consider the convergence of the individuals and will leave out the (very important) question of the distribution of the limit population since this is still an open problem. It has to be noted that by using the descent cones only the convergence (in terms of the semi-distance *dist*) of the population toward the set of interest can be understood. Further important aspects are not treated here. As discussed

Received July 6, 2009; revised December 29, 2009, March 31, 2010, and June 10, 2010; accepted July 6, 2010. The work of A. Lara was supported by CONACyT through a scholarship to pursue graduate studies at the Department of Computer Science, CINVESTAV-IPN. The work of C. A. Coello Coello was supported by CONACyT, under Project 103570.

The authors are with the Department of Computer Science, CINVESTAV-IPN, Mexico City, D.F. 07300, Mexico (e-mail: schuetze@cs.cinvestav.mx; alara@computacion.cs.cinvestav.mx; ccoello@cs.cinvestav.mx).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TEVC.2010.2064321

above, an approximation in the Hausdorff sense has strong limitations with respect to the value of k; however, there are further interesting metrics for the treatment of many-objective problems such as the set coverage metric or the hypervolume metric [55], as considered in [29] and [3], respectively. To understand the evolution of the populations with respect to these metrics, a (sole) consideration of the descent cones does not seem to be adequate.

While the choice of k has a direct influence on the dimension of the solution set the relation to 2) is rather indirect. On the one hand, an additional objective certainly increases the chance that more locally optimal solutions exist since every local solution of each objective is also a local solution of the MOP (see the Appendix). Hence, every multi-modal objective makes the problem harder as it is the case for the DTLZ test problems [16] which are often considered in the context of the evaluation of many-objective evolutionary algorithms. On the other hand, this increase of hardness comes rather from the multi-modality of the model than from the additional objective and can be "substituted" by increasing the multi-modality of the already existing objectives. However, it is a common belief that more objectives make a MOP harder (e.g., [11], [18], [20], [23]) which has an impact on the design in particular of evolution strategies for the treatment of many-objective optimization problems. As reason for this behavior it is sometimes argued that the number of incomparable solutions increases if further objectives are added to a problem (empirically studied, e.g., in [26], [29], [35], and [42], and proven in [54]), and thus, that the evolution of the populations toward the Pareto sets is slowed down.

The aim of this paper is to investigate the influence of the hardness of a problem for an evolutionary search procedure with respect to k. Instead of looking at the number of nondominated solutions within a population, we will focus on the ability of the populations to evolve toward the Pareto sets. Since there is a certain relation between the probability to (locally) improve an individual x by the generational operators and the size of the descent cone at x, we will use and adapt some considerations from [7] of the sizes of the cones in order to try to explain the behavior of the evolution. To handle 1), we will restrict the population size to a fixed value as discussed above, and to avoid the problem described in 2), we will concentrate on uni-modal models. We will argue that a MOP (theoretically) indeed gets harder when adding an objective, but that this difference is—at least for uni-modal models and under an additional assumption on the evolutionary algorithm—not significant, and demonstrate this empirically on three examples. Further on, we will address the treatment of general models where such a scalability has been observed by many researchers so far. Based on our considerations we try to extract the challenges for many-objective evolutionary algorithms and give an attempt to explain recent advances in this field in light of the new insight. A critical discussion on the influence of k for discrete MOPs can be found in [5], but the study presented in this paper seems to be the first one for continuous models. Since our ansatz is using descent cones, the conclusions we draw are restricted to continuous models. Similar explanations for combinatorial problems do not seem to exist.

The remainder of this paper is organized as follows. Section II gives the required background for the understanding of the sequel. In Section III, we investigate a class of uni-modal test functions analytically and empirically with respect to the influence of the number of objectives to the hardness of the problem. In Section IV, we discuss our results and give an attempt to explain recent advances in the field of evolutionary many-objective optimization. Finally, we draw some conclusions in Section V.

II. Background

In the following, we consider continuous MOPs which are of the following form:

$$\min_{x\in Q}\{F(x)\} \qquad \text{(MOP)}$$

where $Q \subset \mathbb{R}^n$ is the domain and the function F is defined as the vector of the objective functions

$$F : Q \to \mathbb{R}^k \qquad F(x) = (f_1(x), \ldots , f_k(x))$$

and where each objective $f_i : Q \to \mathbb{R}$ is continuous. The optimality of a MOP is defined by the concept of *dominance* [40].

Definition 2.1:

1) Let $v, w \in \mathbb{R}^k$. Then the vector v is *less than* w ($v <_p w$), if $v_i < w_i$ for all $i \in \{1, \ldots , k\}$. The relation $\leq_p$ is defined analogously.
2) A vector $y \in \mathbb{R}^n$ is *dominated* by a vector $x \in \mathbb{R}^n$ ($x \prec y$) with respect to (MOP) if $F(x) \leq_p F(y)$ and $F(x) \neq F(y)$, else y is called non-dominated by x.
3) A point $x \in Q$ is called *(Pareto) optimal* or a *Pareto point* if there is no $y \in Q$ which dominates x.

The set of all Pareto optimal solutions is called the *Pareto set*, and is denoted by P_Q. The image $F(P_Q)$ of the Pareto set is called the *Pareto front*. If required, we will denote the Pareto set of a particular MOP by $P_Q(MOP)$ to avoid confusion. In case all the objectives of the MOP are differentiable, the following famous theorem of Kuhn and Tucker [36] states a necessary condition for Pareto optimality for unconstrained MOPs.

Theorem 2.2: Let x^* be a Pareto point of (MOP), then there exists a vector $\alpha \in \mathbb{R}^k$ with $\alpha_i \geq 0, i = 1, \ldots , k$, and $\sum_{i=1}^k \alpha_i = 1$ such that

$$\sum_{i=1}^{k} \alpha_i \nabla f_i(x^*) = 0. \qquad (1)$$

The theorem claims that the vector of zeros can be written as a convex combination of the gradients of the objectives at every Pareto point. Obviously, (1) does not state a sufficient condition for Pareto optimality. On the other hand, points satisfying (1) are certainly "Pareto candidates."

Definition 2.3: A point $x \in \mathbb{R}^n$ is called a *Karush–Kuhn–Tucker point*[1] (KKT–point) if there exist scalars $\alpha_1, \ldots , \alpha_k \geq 0$ such that $\sum_{i=1}^k \alpha_i = 1$ and that (1) is satisfied.

[1]Named after the works of Karush [28], and Kuhn and Tucker [36].

Next, we define some distances between points as well as between different sets.

Definition 2.4: Let $u, v \in \mathbb{R}^n$ and $A, B \subset \mathbb{R}^n$. The maximum norm distance d_∞, the semi-distance $\text{dist}(\cdot,\cdot)$ and the *Hausdorff distance* $d_H(\cdot,\cdot)$ are defined as follows:

1) $d_\infty(u, v) := \max\limits_{i=1,\ldots,n} |u_i - v_i|$;
2) $dist(u, A) := \inf\limits_{v\in A} d_\infty(u, v)$;
3) $dist(B, A) := \sup\limits_{u\in B} dist(u, A)$;
4) $d_H(A, B) := \max\{dist(A, B), dist(B, A)\}$.

As discussed above, we are in particular interested in the convergence of the archive entries toward the set of interest. In case of the Pareto front, it is

$$\text{dist}(F(A_l), F(P_Q)) \tag{2}$$

where $A_l = \{a_1, \ldots, a_m\}$ is the archive in generation l. Since *dist* (and thus also d_H) is sensitive to outliers which is a potential drawback when measuring the solution of stochastic algorithms one can use instead the *generational distance* (GD, see [52]) which measures the average distance of the elements of A_l to the Pareto front

$$GD(A_l) := \frac{1}{m}\sqrt{\sum_{i=1}^{l} dist(F(a_i), F(P_Q))^2}. \tag{3}$$

The Pareto sets of the test functions considered in the following are given by simplexes which are defined as follows.

Definition 2.5: Let $v_1, \ldots, v_k \subset R^n$, $n \geq k$, be given. The set

$$S(v_1, \ldots, v_k) := \left\{\sum_{i=1}^{k} \lambda_i v_i \ :\ \lambda \in [0,1]^k, \text{ and } \sum_{i=1}^{k} \lambda_i = 1\right\} \tag{4}$$

is called the $(k-1)$-simplex of $v_1, \ldots, v_k$.

A hyperplane $H = H(\tilde{x}, \eta)$ in n-dimensional space is defined by a point $\tilde{x} \in H$ and a normal vector $\eta \in \mathbb{R}\backslash\{0\}$, that is

$$H(\tilde{x}, \eta) = \{x \in \mathbb{R}^n \ :\ \langle x - \tilde{x}, \eta\rangle = 0\} \tag{5}$$

where $\langle\cdot,\cdot\rangle$ defines the standard scalar product. The point $p(x)$ which is closest to H is given by

$$p(x) = x - \frac{\langle x - \tilde{x}, \eta\rangle}{\langle \eta, \eta\rangle}\eta. \tag{6}$$

III. Investigation of a Class of Uni-Modal Models

A. A Class of Test Problems with Simplicial Pareto Sets

Here, we construct a set of quadratic (and hence uni-modal) test functions where the Pareto sets are given by simplexes which eases the computation of the distance of a point to the Pareto set and front. The resulting models we consider are slight variants of the P^* problems introduced in [34] tailored to our needs.

1) *Construction:* First we construct the base problem. Given points $a_1, \ldots, a_k \in \mathbb{R}^n$, we define the MOP as follows:

$$\begin{aligned} &\min F : \mathbb{R}^n \to \mathbb{R}^k \\ &f_i(x) = \|x - a_i\|_2^2 = \sum_{j=1}^{n} (x_j - a_{i,j})^2 \end{aligned} \tag{7}$$

where $a_{i,j}$ denotes the jth entry of a given vector a_i. The Pareto set of the problem defined by (7) [in short MOP(7)] is given by the simplex spanned by the k minimizers a_i.

Proposition 3.1: $P_Q(MOP\ (7)) = S(a_1, \ldots, a_k)$.

Proof: It is $\nabla f_i(x) = 2(x - a_i)$. Let $x \in S(a_1, \ldots, a_k)$, i.e., there exist scalars $\lambda_1, \ldots, \lambda_k \geq 0$ with $\sum_{i=1}^k \lambda_i = 1$ such that $x = \sum_{i=1}^k \lambda_i a_i$. Then

$$\begin{aligned} \sum_{i=1}^{k} \lambda_i \nabla f_i(x) &= \sum_{i=1}^{k} \lambda_i 2(x - a_i) = 2(x\underbrace{\sum_{i=1}^{k}\lambda_i}_{=1} - \sum_{i=1}^{k}\lambda_i a_i) \\ &= 2\left(x - \sum_{i=1}^{k}\lambda_i a_i\right) = 0. \end{aligned} \tag{8}$$

The claim follows since MOP (7) is strictly convex, and thus, the Pareto set is equal to the set of Karush–Kuhn–Tucker (KKT) points. ■

The problem is quadratic and unconstrained. Note that for the special case $n = 1$, $k = 2$, $a_1 = 0$, and $a_2 = 1$ the MOP (7) coincides with the well-known problem of Schaffer [45]. The authors of [34] propose to locate all the minima a_i on an Euclidean plane which results in a 2-D Pareto set. In order, e.g., to obtain a $(k-1)$-dimensional object, the volume of $S(a_1, \ldots, a_k)$ has to be positive, i.e., the $k-1$ difference vectors $a_2 - a_1, \ldots, a_k - a_1$ have to be linearly independent.

In the following, we use Proposition 1 to construct constrained problems with variable dimension of the solution set. For this, we will use hyperplanes. Given a hyperplane $H = H(\tilde{x}, \eta)$, there exists for every point $x \in \mathbb{R}^n$ a $\lambda = \lambda(x) \in \mathbb{R}$ such that

$$x - p(x) = \lambda\eta \tag{9}$$

which can be used to divide the space $\mathbb{R}^n$ as follows. Let $j \in \{1, \ldots, n\}$ such that $\eta_j \neq 0$, then we define

$$\begin{aligned} &g_H : \mathbb{R}^n \to \mathbb{R} \\ &g_H(x) = \frac{x_j - p(x)_j}{\eta_j} \end{aligned} \tag{10}$$

and the constrained MOP is

$$\begin{aligned} &min\ F(x) \\ &s.t.\ g_H(x) \leq 0 \end{aligned} \tag{11}$$

where F is as defined in (7). Thus, the domain is given by $Q = \{x \in \mathbb{R} : g_H(x) \leq 0\}$. Constrained problems can now be constructed by using MOP (7) and placing the a_i's at the boundary of Q. The following result shows how further constrained MOPs can be generated with different dimensions of the Pareto set (see also Fig. 1). Further on, we give one such example.

Proposition 3.2: Let $H = H(\tilde{x}, \eta)$ be a hyperplane and $a_1, \ldots, a_k \in \mathbb{R}^n$ such that

$$a_1, \ldots, a_l \in H \quad l \leq k \tag{12}$$

and

$$\begin{aligned} g_H(a_i) &> 0 \quad i = l+1, \ldots, k \\ p(a_i) &\in S(a_1, \ldots, a_l) \quad i = l+1, \ldots, k. \end{aligned} \tag{13}$$

Then, the Pareto set of MOP (11) is given by

$$P_Q(MOP\ (11)) = S(a_1, \ldots, a_l). \tag{14}$$

Proof: By Proposition 1, it is clear that: 1) $S(a_1, \ldots, a_l) \subset P_Q$, and 2) none of the points $x \in \mathbb{R}$ with $g_H(x) < 0$, i.e., the points where g_H is inactive, is Pareto optimal [else 0 can be expressed as a convex combination of the objectives' gradients, but this was prevented by the first assumption in (13)]. It remains to show that $H \backslash S(a_1, \ldots, a_l)$ is not contained in P_Q. For $x \in H \backslash S(a_1, \ldots, a_l)$ choose $z \in S(a_1, \ldots, a_l)$ such that

$$z \in \operatorname{argmin}_{s \in S(a_1, \ldots, a_l)} \|x - s\|_2. \tag{15}$$

Since $S(a_1, \ldots, a_l)$ is a convex set and $x \notin S(a_1, \ldots, a_l)$ it follows that

$$\|s - z\|_2 < \|s - x\|_2 \quad \forall s \in S(a_1, \ldots, a_l). \tag{16}$$

Since (16) holds for a_i, $i = 1, \ldots, l$, it follows that $f_i(z) < f_i(x)$, $i = 1, \ldots, l$. Further, by the same argument on $p(a_i)$, $i = l+1, \ldots, k$, and Pythagoras

$$\begin{aligned} &x \in H \ \Rightarrow\ \|a_i - x\|_2^2 = \|a_i - p(a_i)\|_2^2 + \|p(a_i) - x\|_2^2 \\ &i = l+1, \ldots, k \end{aligned} \tag{17}$$

it follows that also $f_i(z) < f_i(x)$, $i = l+1, \ldots, k$, and thus, that $F(z) < F(x)$, which implies that $x \notin P_Q$ which concludes the proof. ■

If for instance $H = H(e_1, \eta)$ is chosen as

$$\eta = (\underbrace{-1, \ldots, -1}_{k}, \underbrace{0, \ldots, 0}_{n-k})^T \tag{18}$$

and $a_i = e_i$, $i = 1, \ldots, k$, then $a_i \in H$, $i = 1, \ldots, k$ ($p(a_i) = a_i$) and thus, $P_Q = S(e_i, \ldots, e_k)$. The dimension of the solution set can be reduced by one if choosing, e.g., $a_i = e_i$, $i = 1, \ldots, k-1$, $a_k = 0$, and $H = H(e_1, \eta)$ with

$$\eta = (\underbrace{-1, \ldots, -1}_{k-1}, \underbrace{0, \ldots, 0}_{n-k+1})^T. \tag{19}$$

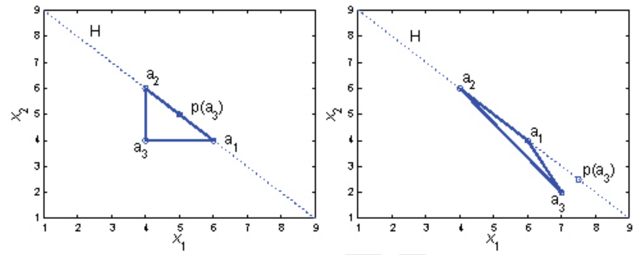

Fig. 1. Two examples where the facet of a 3-simplex is included in the hyperplane. Left: the Pareto set of MOP (11) is given by $S(a_1, a_2)$ since $p(a_3) \in S(a_1, a_2)$. Right: $p(a_3) \notin S(a_1, a_2)$, and thus, the Pareto set is not equal to the facet $S(a_1, a_2)$.

It is $p(a_k) = \frac{-1}{k-1}\eta \in S(a_1, \ldots, a_{k-1})$ [using the weights $\alpha_i = 1/(k-1)$] and $g_H(a_k) = 1/(k-1) > 0$, and thus, it follows by Proposition 2 that $P_Q = S(a_1, \ldots, a_{k-1})$.

Continuing in a similar manner, the dimension of the Pareto set can be reduced. The extreme situation—i.e., that P_Q consists of one single solution—can, e.g., be obtained as follows: set $a_1 = e_1$, and $a_i = \lambda_i e_i$, $\lambda_i < 1$, for $i = 2, \ldots, k$, and $H = H(e_1, \eta)$ with $\eta = (-1, 0, \ldots, 0)^T$. Then, it is $p(a_i) = e_1$ and $g(a_i) = 1 - \lambda_i > 0$ for $i = 2, \ldots, k$, and thus, $P_Q = \{e_1\}$.

2) *Test Problems:* Based on the above observations, we propose two test functions which are used to investigate the hardness of a MOP with respect to the number of objectives.

a) *PS1:* Given vectors $a_1, \ldots, a_k \in \mathbb{R}^n$, $n \geq k$, we define the first test problem PS1 as in (7). For the a_i's we suggest choosing $a_i = e_i$, and as domain $Q = [-10, 10]$. By Proposition 1 it follows that

$$P_Q(PS1) = S(e_1, \ldots, e_k). \tag{20}$$

b) *PS2:* Here we define a constrained model where the dimension of the Pareto set can be chosen between 0 and $k-1$, where k is the number of objectives: given a number $1 \leq l \leq k$, we define PS2(l) as follows. Let $H = H(e_1, \eta)$ with

$$\eta = (\underbrace{-1, \ldots, -1}_{l}, \underbrace{0, \ldots, 0}_{n-l}) \tag{21}$$

let g_H as in (10), and F as in (7), where $a_i = e_i$, $i = 1, \ldots, l$, and $a_j = -\frac{1}{l}\eta + \frac{j-l}{l}\eta$, $j = l+1, \ldots, k$. Then PS2(l) reads as follows:

$$\begin{aligned} &\min F(x) \\ &\text{s.t. } x_i \in [-10, 10]^n \ i = 1, \ldots, n \\ &\quad g_H(x) \leq 0. \end{aligned} \tag{22}$$

Due to the discussion in the previous subsection it is

$$P_Q(PS2(l)) = S(e_1, \ldots, e_l) \tag{23}$$

i.e., a l-simplex which is located within the boundary of the domain. The characteristic of this model is that the Pareto set of PS2(l) for k_1 objectives [denoted by $PS2_{k_1}(l)$] is equal to the Pareto set of $PS2_{k_2}(l)$, where k_1 and k_2 are any numbers larger than or equal to l.

B. Hardness of the PS Problems with Respect to k

In the following, we investigate the hardness of the PS test problems by some (non-rigorous) theoretical considerations and by empirical studies.

1) *Qualitative Considerations:* In the following, we consider the PS test problems for general locations of the minima a_i. If further assumptions are required, we will mention them.

For our considerations, we use the descent cones to investigate the hardness of a problem. Given a MOP with s objectives the descent cone at a point $x \in Q$ is given by (e.g., [4])

$$D(f_1, \ldots, f_s, x) = \{v \in \mathbb{R}^n \backslash \{0\} \ : \ \langle \nabla f_i(x), v \rangle < 0 \\ \forall\, i = 1, \ldots, s.\} \tag{24}$$

$D(f_1, \ldots, f_s, x)$ is the set of all directions in which dominating points can be found, i.e., for each $v \in D(f_1, \ldots, f_s, x)$ there exists a (possibly small) $t \in \mathbb{R}_+$ such that $F(x + tv) <_p F(x)$. There exists a certain relation of the size of the descent cone to the probability to (locally) improve the value of x by the generational operators of a MOEA. For the mutation operator, the relation is proportional when assuming the existence of a suitable or small step size control (i.e., the value of t for the offspring $o := x + tv$). For the most common crossover strategies (e.g., SBX [12]) such a relation still holds; however, the success rate is here in addition depending on the location of the parents. Hence, one can say that a small descent cone results in a small probability of finding a better, i.e., dominating, solution near to x, and large descent cones in turn lead to a larger improvement possibility.

Assume we are given $l + 1$ objectives of the form defined in (7), which are entirely determined by the choice of the a_i's and assume further that $a_{l+1} \notin S(a_1, \ldots, a_l)$. Clearly, $D(f_1, \ldots, f_{l+1}, x)$, i.e., the descent cone for the $(l + 1)$-objective problem is a subset of $D(f_1, \ldots, f_l, x)$, i.e., the according descent cone for the MOP consisting of the first l objectives. The equality of both cones holds if $-\nabla f_{l+1}(x)$ is "between" the vectors $-\nabla f_i(x),\ i = 1, \ldots, l$. Since for the PS problems it is $\nabla f_i(x) = 2(x - a_i)$ (i.e., the steepest descent $-\nabla f_i(x)$ points to the minimizer of f_i at every point $x \in Q$) we have

$$D(f_1, \ldots, f_{l+1}, x) = D(f_1, \ldots, f_l, x) \Leftrightarrow \exists \lambda_1, \ldots, \lambda_l \geq 0 : \\ a_{l+1} - x = \sum_{i=1}^{l} \lambda_i (a_i - x). \tag{25}$$

Thus, a necessary condition for the equality of the cones is that $a_{l+1} - x \in \operatorname{span}\{a_1 - x, \ldots, a_l - x\}$ by which it follows that the set of points $x \in Q$ which satisfies (25) is maximal l-dimensional (and thus a zero set in Q). To be more precise, for every point x which is not included in the affine subspace

$$A := \operatorname{span}\{a_1, \ldots, a_l\} + \left\{ \frac{-a_{l+1}}{\sum_{i=1}^{l} \alpha_i - 1} \right\} \tag{26}$$

where $\alpha \in \mathbb{R}^l$ such that $a_{l+1} - x = \sum_{i=1}^{l} \alpha_i (a_i - x)$ [note that since $a_{l+1} \notin S(a_1, \ldots, a_l)$ it is $\sum_{i=1}^{l} \alpha_i \neq 1$, and hence, (26) is well defined], the equality of the cones does not hold. Hence, picking a randomly chosen point $x_0 \in Q$ the probability is one that $D(f_1, \ldots, f_{l+1}, x_0)$ of the $(l + 1)$-objective problem is a proper subset of the cone $D(f_1, \ldots, f_l, x_0)$ of the related "reduced" l-objective problem. This result is in accord with the observation made in [54] that the number of incomparable solutions generally increases with an increasing number of objectives.

Thus, it can be said that—from a theoretical point of view—the PS problems get harder with increasing number of objectives. Since (25) can in principle be applied to any set of gradients, the statement holds for general MOPs. On the other hand, this (point-wise) observation is of qualitative nature and gives no statement about the quantity of the difference which is needed to judge the hardness of a problem for a given evolutionary search procedure with respect to k. The following qualitative considerations,[2] however, question the common belief that the addition of further objectives makes a given MOP per se harder.

Assume we are given MOP1 which consists of the objectives $f_1, \ldots, f_k$ of the form defined in (7) and MOP2 which contains the same k objectives as in MOP1 plus the l objectives $f_{k+1}, \ldots, f_{k+l}$. If the initial population P_0 is chosen at random from the domain Q, it can be assumed that most of its individuals $p \in P_0$ are "far away" from both Pareto sets (note that under the reasonable assumption $n > k + l$ both sets $S(e_1, \ldots, e_k)$ and $S(e_1, \ldots, e_{k+l})$ are zero sets in Q). Thus, the vectors $\{p - a_i\}_{i=1,\ldots,s}$ for such an individual p point nearly in the same direction, and this holds for $s = k$ as well as for $s = k + l$. One way to see this is that if a sequence of points is chosen with unbounded increasing distance to all the minima a_i, both simplexes $S(a_1, \ldots, a_k)$ and $S(a_1, \ldots, a_{k+l})$ shrink in the limit down to a point, and hence, both descent cones $D(f_1, \ldots, f_k, p)$ and $D(f_1, \ldots, f_{k+l}, p)$ form the same half space as the cones $D(f_i, p)$, $i \in 1, \ldots, k + l$, for single-objective optimization. This implies that it can be expected that also for finite distances the descent cones $D(f_1, \ldots, f_k, p)$ and $D(f_1, \ldots, f_{k+l}, p)$ are nearly equal (and large), and thus, that the evolution of the populations should be nearly equal for both problems MOP1 and MOP2. The situation will change after a small number of generations: due to the sizes of the descent cones there is a high chance for improvement, and thus, it can be expected that the sequence of populations performs a certain evolution toward the Pareto set. If so, it cannot be expected any more that the cones have similar sizes. Since MOP2 contains more objectives it is more likely that $D(f_1, \ldots, f_{k+l}, p)$ is smaller than $D(f_1, \ldots, f_k, p)$ for an element p of the current population. [Compare to the theorem of Kuhn and Tucker: if, for instance, two gradients point in opposite directions then the associated cone defined by (24) is empty. By continuity of F, the descent cones near to KKT points are hence small.] However, this is mainly due to the geometry of multiobjective optimization since the Pareto set of MOP2 is indeed larger [$P_Q(MOP2)$ is $(k+l-1)$-dimensional while $P_Q(MOP1)$ is $(k-1)$-dimensional]. Thus, the evolution has to terminate earlier for MOP2 resulting in smaller cones compared to MOP1. Another point—and this one cannot be explained by looking at the descent cones—is

[2]Here we adapt some observations made in [7] to the present context.

one population-based aspect of MOEAs, namely that single "good" solutions—i.e., solutions which are "near" to the Pareto set—can pull the entire population to the set of interest. Using the dimensionality of the different Pareto sets, it can be argued that the chance to find a "good" solution is higher for MOP2 than for MOP1. Hence, using the dimensionality, the argumentation of the influence of k can be turned: under the above assumption (which we will refer to as the *pulling assumption* in the sequel and which will be discussed in more detail in Section IV) and the additional assumption that the population/archive size is fixed and equal for both MOPs it is rather likely that MOP2 is the easiest model in terms of convergence [i.e., when considering $dist(A_l, P_Q)$].

Concluding, it can be said that by adding an objective in a PS model (or other models), the resulting MOP gets indeed "harder" from a theoretical point of view, but it is ad hoc unclear if the amount is indeed significant since some considerations argue against it. However, the above analysis covers only the extreme situations (points which are either far away or near to P_Q) and is only of qualitative nature. To elucidate this problem sufficiently, empirical studies seem to be required which we will do in the following.

2) *Empirical Studies:* As mentioned before, we are in particular interested in the evolution (or convergence) of the populations toward the set of interest. For this, we use the generational distance defined in (3) and a variant of this indicator which we propose in the following.

Given a population $A = \{a_1, \ldots, a_l\}$, GD measures the average distance of the elements of A to the Pareto front. Since the dimension of the vectors $F(a_i)$ varies with the number of objectives, one may argue that for a comparison which includes different number of objectives GD is not well suited. Thus, we propose here a variant of GD, namely

$$GD_x(A) := \frac{1}{l}\sqrt{\sum_{i=1}^{l} dist(a_i, P_Q)^2} \tag{27}$$

which is analog to GD but measures the averaged distance of A to the Pareto *set*, i.e., in parameter space. Hereby, the distance of a point $a \in A$ to the Pareto set and its image to the Pareto front are given by

$$\begin{aligned} dist(a, P_Q) &= \min_{p \in P_Q} \|a - p\|_2 \\ dist(F(a), F(P_Q)) &= \min_{p \in P_Q} \|F(a) - F(p)\|_2. \end{aligned} \tag{28}$$

These are single-objective optimization problems (SOPs) with n-dimensional parameter space. In case $P_Q = S := S(a_1, \ldots, a_k)$ as for our test problems, (28) can be written as

$$\begin{aligned} dist(a, S) &= \min_{\alpha \in S} \left\| a - \sum_{i=1}^{k} \alpha_i a_i \right\|_2 \\ dist(F(a), F(S)) &= \min_{\alpha \in S} \left\| F(a) - F(\sum_{i=1}^{k} \alpha_i a_i) \right\|_2. \end{aligned} \tag{29}$$

Since the SOPs in (29) are convex problems (domain and objective are convex) with k free parameters, it can easily be solved with standard mathematical techniques (note that in the context of scalar optimization, a problem is noted as small if the dimension of the parameter space is less than 10 000, which is definitely beyond the scope of many-objective optimization).

We have chosen to take NSGA-II [14] for our empirical studies since this algorithm was shown to scale badly with increasing number of objectives for certain models (e.g., [53]). Additionally, we have made (but do not display) analog computations with SPEA2 [56] which confirmed the results shown below.

Figs. 2–5 show some numerical results obtained by NSGA-II for PS1 and PS2 [using $l = k$, denoted here by $\mathrm{PS2}_k$(k) to avoid confusion] and for different numbers k of objectives. In all examples, we have used parameter dimension $n = 30$, population size $N_p = 100$, and the probabilities $p_c = 0.85$ and $p_m = 0.05$ for crossover and mutation, respectively. The initial population P_0 has been chosen randomly from $I := [9, 10]^{30}$, since by the above discussion for every point $x \in I$ the descent cone $D(f_1, \ldots, f_{k+l}, x)$ of the $(k+l)$-objective problem is a proper subset of the cone $D(f_1, \ldots, f_k, x)$ of the reduced problem (analog empirical studies where P_0 has been chosen randomly from $Q_i, i = 1, 2$, however, have led to the same results). For both the unconstrained and the constrained case as well as for a measurement in parameter and image space (GD_x and GD, respectively) the same behavior can be observed: in the large scale, i.e., when considering all 500 generations, the evolution of the populations is basically the same (note that there is a difference of 12 objectives). When zooming into the figures, little differences appear, and as anticipated, the values of GD_x and GD get (little) larger with increasing number of objectives (note the difference of the values with the initial values of GD and GD_x, respectively).

Whereas the results can be explained to a certain extent by the above considerations, a sole consideration of the number of nondominated solutions in a population may be misleading in this example. Fig. 6 shows the (averaged) number of nondominated solutions for the PS1 problems within the populations found by NSGA-II, and here, the differences are significant. For instance, for $k = 3$ there are about 90% of dominated solutions after 100 generations (and about 50% of dominated solutions after 200 generations) while for $k \geq 10$ practically all members of a population are mutually nondominating after about 100 generations. Hence, by only looking at these values one could have come to the conclusion that the problem gets clearly harder with increasing k which cannot be confirmed by our studies.

Since it may be argued that for different values of k a comparison for the above models is not completely fair (in addition to the difference of $F(a)$ described above there is the difference in the dimension of the Pareto sets), we consider $\mathrm{PS2}_k$(l) for a fixed value of l but with different values of k. To be more precise, we consider $\mathrm{PS2}_k(k)$ and $\mathrm{PS2}_{k+1}(k)$. The reason is that in both cases, i.e., for the k-objective model $\mathrm{PS2}_k(k)$ as well as for the $(k+1)$-objective model $\mathrm{PS2}_{k+1}(k)$, the Pareto set is given by $S(e_1, \ldots, e_k)$. That is, in this case at least GD_x can be assumed to be completely fair for a

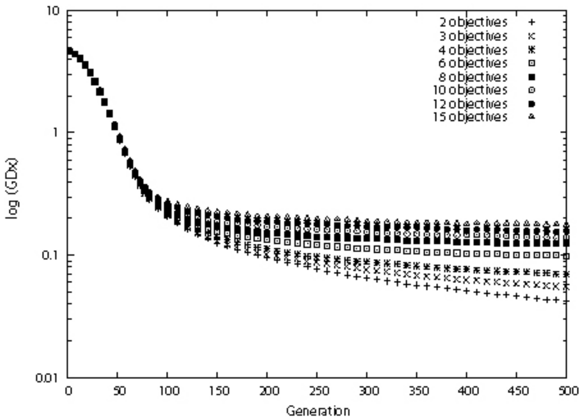

Fig. 2. Numerical results of NSGA-II for PS1 for k = 2, 3, 4, 6, 8, 10, 12, 15 objectives. The results are in parameter space [log(GD$_x$)] and averaged over 50 independent runs. Compare to Table 1.

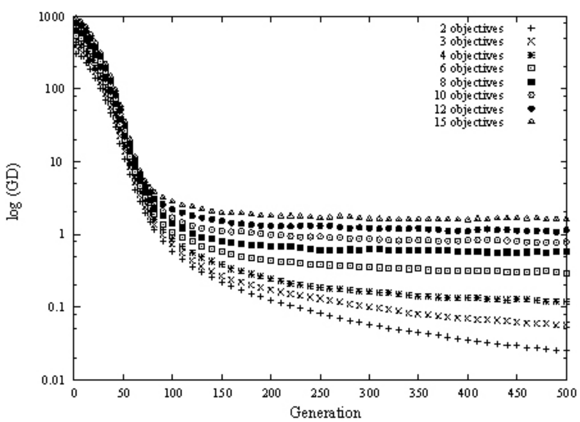

Fig. 3. Numerical results of NSGA-II for PS1 for k = 2, 3, 4, 6, 8, 10, 12, 15 objectives. The results are in objective space [log(GD)] and averaged over 50 independent runs. Compare to Table 2.

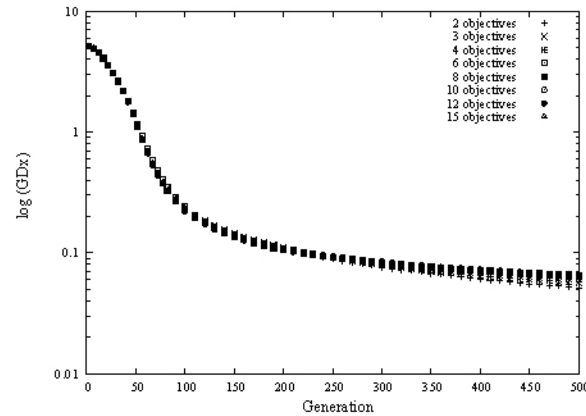

Fig. 4. Numerical results of NSGA-II for PS2$_k$(k) for k = 2, 3, 4, 6, 8, 10, 12, 15 objectives. The results are in parameter space [log(GD$_x$)] and averaged over 50 independent runs.

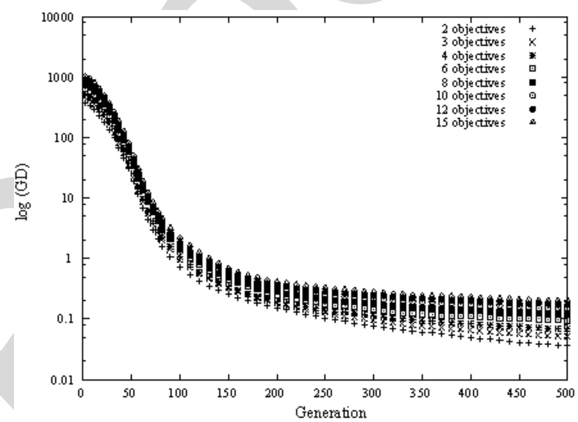

Fig. 5. Numerical results of NSGA-II for PS2$_k$(k) for k = 2, 3, 4, 6, 8, 10, 12, 15 objectives. The results are in objective space [log(GD)] and averaged over 50 independent runs.

comparison. Figs. 7 and 8 show such comparisons for values of k between 3 and 14, where we have chosen the same setting as in the previous study. Also here, small differences in the performances can be observed, but it is certainly not justified to talk about different orders of magnitude.

IV. Discussion and an Attempt to Explain Recent Advances

In the previous section, we have investigated a particular class of uni-modal MOPs with respect to the influence of the number of objectives on the hardness of the problem. Putting theoretical and empirical observations together we can conclude that by adding an objective to a given MOP the problem does per se not get harder by a significant amount, at least not on the (easy) class of models under consideration. However, such a scalability has been observed by many researchers on other, more complex, models. The question which now naturally arises is how this can be put together, i.e., if the observations made above can also be helpful for the design of algorithms for general many-objective models. In the following, we hazard to guess the sources of difficulties when dealing with many-objective problems, and try to explain recent advances in the field of evolutionary many-objective optimization in light of our discussion.

Based on the above considerations, three influential factors for the efficient numerical treatment of many-objective optimization problems with evolutionary algorithms regardless of the particular choice of the algorithm seem to be:

1) the pulling assumption as described in Section III-B1;
2) the probability to improve an individual;
3) the multi-modality of the MOP.

Problems 1) and 2) are to a certain extent in the hands of the algorithm designer, whereas problem 3) is given to him/her (or is possibly a modeling problem).

Much research has been done so far to improve the pulling property [i.e., problem 1)]. In case a population consists only of nondominated solutions and the generational operators produce further nondominated candidates the question arises which point to keep and which one to discard in order to converge toward the Pareto set. Since not all these nondominated solutions have the same distance to the solution set one can laxly say that *"some nondominated points are better than others"* [9]. The quest for those points has led so far to a variety of substitute distance assignments in NSGA-II

TABLE I
NUMERICAL RESULTS OF NSGA-II FOR PS1 FOR k = 3, 4, 6, 8, 10, 12, 15 OBJECTIVES

	Number of Generations					
k	50	100	200	300	400	500
2	8.57E−001	2.07E−001	9.54E−002	6.48E−002	5.10E−002	4.24E−002
3	8.74E−001	2.13E−001	1.02E−001	7.62E−002	6.28E−002	5.53E−002
4	8.85E−001	2.19E−001	1.12E−001	8.83E−002	7.67E−002	6.98E−002
6	8.79E−001	2.22E−001	1.33E−001	1.13E−001	1.03E−001	9.84E−002
8	8.78E−001	2.38E−001	1.52E−001	1.37E−001	1.28E−001	1.24E−001
10	8.94E−001	2.43E−001	1.72E−001	1.55E−001	1.46E−001	1.39E−001
12	9.18E−001	2.60E−001	1.87E−001	1.72E−001	1.60E−001	1.58E−001
15	9.27E−001	2.72E−001	2.06E−001	1.88E−001	1.79E−001	1.76E−001

The results are in parameter space (GD_x) and averaged over 50 independent runs (compare to Fig. 2).

TABLE II
NUMERICAL RESULTS OF NSGA-II FOR PS1 FOR k = 3, 4, 6, 8, 10, 12, 15 OBJECTIVES

	Number of Generations					
k	50	100	200	300	400	500
2	1.06E+001	5.86E−001	1.24E−001	5.65E−002	3.53E−002	2.48E−002
3	1.36E+001	7.24E−001	1.71E−001	9.96E−002	6.94E−002	5.61E−002
4	1.63E+001	8.56E−001	2.44E−001	1.62E−001	1.34E−001	1.17E−001
6	1.97E+001	1.06E+000	4.33E−001	3.49E−001	3.11E−001	2.93E−001
8	2.27E+001	1.41E+000	6.71E−001	6.22E−001	5.76E−001	5.75E−001
10	2.67E+001	1.69E+000	9.84E−001	8.64E−001	8.30E−001	7.82E−001
12	3.07E+001	2.20E+000	1.30E+000	1.22E+000	1.10E+000	1.14E+000
15	3.52E+001	5.38E+000	1.78E+000	1.61E+000	1.56E+000	1.58E+000

The results are in objective space (GD) and averaged over 50 independent runs (compare to Fig. 3).

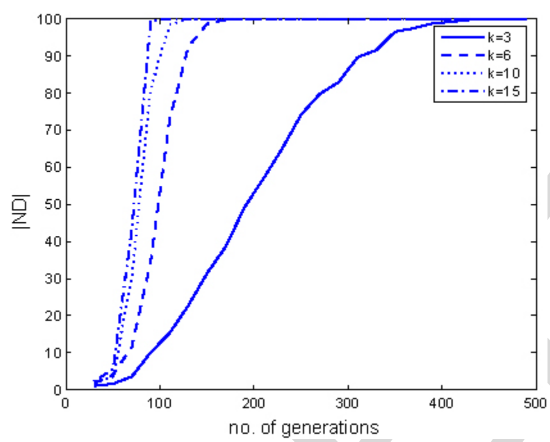

Fig. 6. Number of nondominated points ($|ND|$) during the run of NSGA-II for different values of k for the PS1 problems with population size 100 (averaged over 20 test runs).

(e.g., [2], [9], [35], [41], [50]). All these methods were able to outperform its base MOEA on scalable benchmark models (such as the DTLZ models). Though these results are all satisfying from the practical point of view, however, none of them ensures convergence toward the set of interest. It is known that in NSGA-II cycling (see [21]) or deterioration can occur which prevents that a predescribed "limit set" is reached resulting in a certain lack of efficiency, at least from the theoretical point of view [38]. Due to the dimensionality, the problem of defining a suitable limit set is getting more important with increasing value of k which would ease the evaluation of the newly developed strategies.

In multiobjective particle swarm optimization (MOPSO) algorithms, the pulling property is closely related to the choice of the guidance mechanism which has been addressed in [34] and [39] for many-objective problems.

To downsize problem 2), several remedies have been proposed so far which all lead to an augmentation of the descent cones of the related auxiliary models. One way to increase the improvement probability (while reducing the multi-modality of the problem) is to consider instead of the given k-objective problem a sequence of lower objective problems. For instance, the methods MSOPS [24] and RSO [25] are based on aggregation functions to find Pareto optimal solutions. Another approach is to use "space partitioning" [1], [2], i.e., partitioning the objective space into subspaces and performing one or several generations of the evolutionary search in each subspace. In both cases, the descent cone of the auxiliary model at a point x is typically larger than the original problem, and in the case of space partitioning the number of local minima is typically fewer (see the Appendix). The latter is not always true when using an aggregation function f_a since this depends on the choice of f_a as well as on the original model (see [31] for a counterexample).

For these approaches it holds that the speed of convergence gets improved, but, in turn, problems arise concerning the diversity maintenance. In particular, it may happen that not every Pareto point can be reached by the auxiliary problems which leads to a bias of the approaches. The potential drawbacks of aggregation functions are known (e.g., [11]), the reason for a potential bias when using space partitioning is because the union of the Pareto sets of all subproblems does typically not

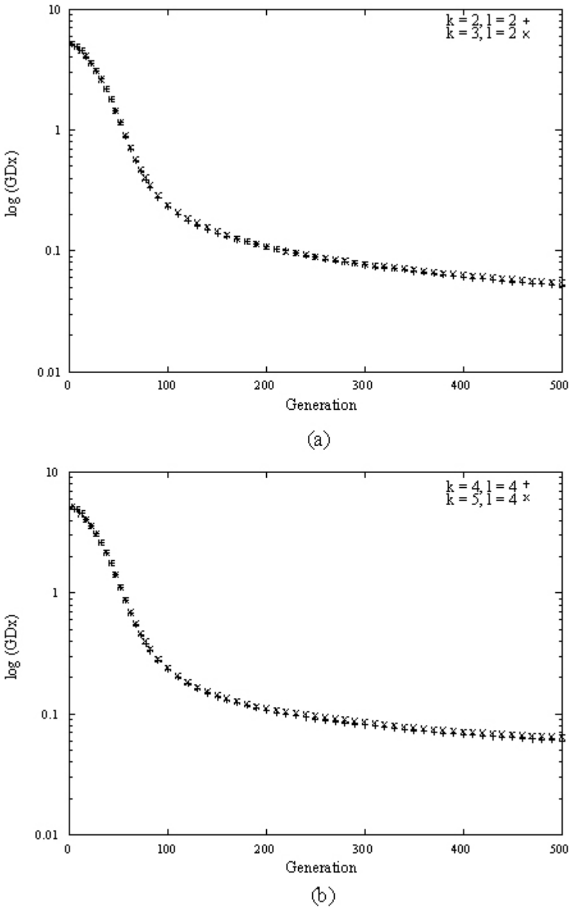

Fig. 7. Numerical results of NSGA-II for $PS2_k(k)$ and $PS2_{k+1}(k)$ for $k = 2$ and 4. The plots show the number of generations vs. $\log(GD_x)$. The results are averaged over 50 independent runs. (a) $PS2_2(2)$ and $PS2_3(2)$. (b) $PS2_4(4)$ and $PS2_5(4)$.

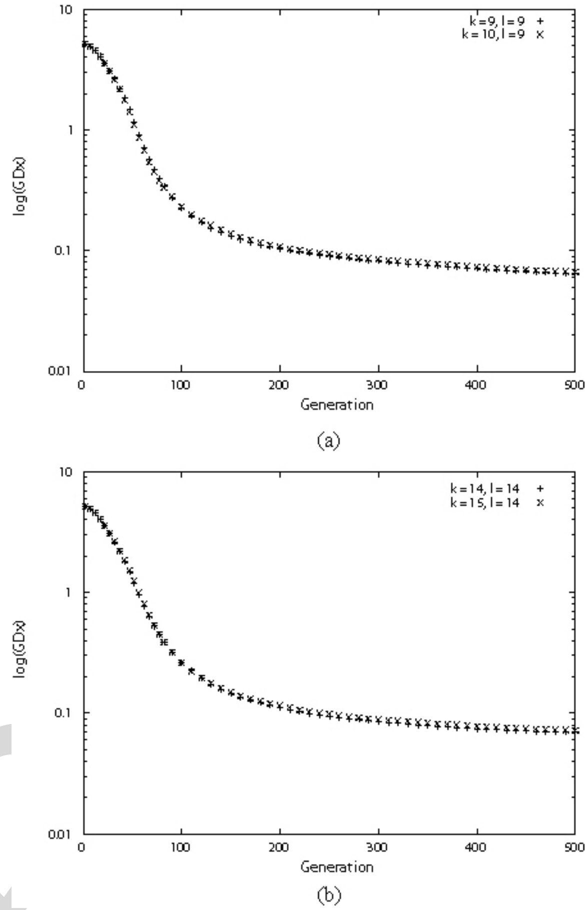

Fig. 8. Numerical results of NSGA-II for $PS2_k(k)$ and $PS2_{k+1}(k)$ for $k = 9$ and 14. The plots show the number of generations vs. $\log(GD_x)$. The results are averaged over 50 independent runs. (a) $PS2_9(9)$ and $PS2_{10}(9)$. (b) $PS2_{14}(14)$ and $PS2_{15}(14)$.

form the Pareto set of the "full" MOP. For instance, when choosing the PS1 problem with minimizers a_1, a_2, and a_3 (for the objectives f_1 to f_3, respectively) such that the volume of $S(a_1, a_2, a_3)$ is positive, then the union of the Pareto sets of all bi-objective subproblems (f_1, f_2), (f_1, f_3), and (f_2, f_3) is $S(a_1, a_2) \cup S(a_1, a_3) \cup S(a_2, a_3)$, i.e., is equal to the boundary of the "complete" Pareto set $S(a_1, a_2, a_3)$, but no interior point is included.

Another way to increase the improvement probability is to modify the Pareto dominance relation. Clearly, a larger dominance cone (defined in objective space) is related to a larger descent cone [defined in parameter space, see (24)] which in turn increases the probability to find a "better" solution as discussed above. The usage of such modified dominance cones within MOEAs can be found in [44], and in [17], [32], [33] fuzzifications of the Pareto dominance relation can be found which by its relaxation similarly influences the size of the dominance cones. Also for these methods, problems in diversity maintenance have been reported.

One aspect so far disregarded by researchers—but probably worth exploring—is the ability of memetic strategies to improve the performance of many-objective optimization problems. On the one hand, mathematical programming techniques (e.g., [4], [19]) allow—if gradient information is at hand—to compute a descent direction at every given non optimal point regardless of the size of the descent cone nor of the value of k, and hence it can be argued that the probability for improvement is one. On the other hand, the use of gradient information within a memetic strategy results in a certain additional cost [37] and in case the model is highly multi-modal [i.e., problem 3)] the effect of the local search on the overall performance is questionable.

In [53], it has been reported that ϵ-MOEA [13] copes well with many-objective optimization problems, even on highly multi-modal models. ϵ-MOEA is a steady state MOEA equipped with an archiving strategy which is based on the concept of ϵ-dominance and guarantees under certain assumptions convergence toward a finite size representation of the Pareto set [43], [38]. The good behavior of ϵ-MOEA with respect to k can partly be explained by our considerations: elements of the archive are only replaced by dominated solutions, i.e., a good solution will not be discarded due to any distance assignment, but only due to the existence of a better one. Hence, such solutions cannot be discarded by mistake which certainly helps to pull the population toward the Pareto set. Further, by the use of the archiving strategy proposed in [38] the descent cone is enlarged: the objective space gets divided into a grid of boxes, whose size can be adjusted by the size of ϵ. Every solution of the archive has to be located in a different box (i.e., every archive entry is associated with a box of the grid). The dominance relation is now enlarged since only nondominated boxes are allowed. Hence, ϵ-MOEA has mechanisms to cope with problems 1) and 2). Further good results can hence in principle be expected with related algorithms such as PAES [10] or PESA [30], or with any MOEA which is equipped with an archive which converges toward a finite size representation of the set of interest (e.g., [46]–[49]). The problem—at least when using archivers based on ϵ-dominance—is certainly the proper choice of ϵ: as discussed in [9], if the value of ϵ is too small, the archive sizes become intractable, and for large values of ϵ the limit archive set basically consists of a set of randomly selected (but not close by) points from the Pareto set.

In summary, it can be said that yet a variety of promising approaches exist in terms of their ability to converge toward the Pareto set P_Q. The distribution, however, is still an open problem. For this, a clear definition of the optimal distribution of the (few) individuals $a \in A$ is still missing but required to evaluate the finite size approximation of P_Q beyond convergence in the sense of $dist(A, P_Q)$ or $dist(F(A), F(P_Q))$.

V. Conclusion

In this paper, we have investigated the influence of the number k of objectives in a MOP on the hardness of the problem when solving it by evolution strategies. For this, we have utilized the descent cones which can be used to measure the probability to improve a solution by the generational operators. Though these considerations are of qualitative nature and can hardly be quantified, they help to a certain extent to understand the behavior of the population's evolution with respect to k. As an example, we have considered a class of uni-modal test functions and have investigated the resulting models qualitatively and empirically. Qualitative studies based on the descent cones led to the conclusion that, on the one hand, the addition of an objective makes the problem indeed harder, but, on the other hand, it can be argued that the difference is not significant, which is, later on, empirically validated. That is, it can be argued that the addition of an objective to a MOP does not make the problem per se harder.

In contrast to this, many researchers have so far observed a certain scalability in the hardness of the problem with respect to k, albeit for more complex models. Based on our considerations on the uni-modal models we have tried to identify the challenges which have to be mastered by evolution strategies for general models: the ability to keep "good" solutions in order to pull the population toward the set of interest, the probability to improve an individual, and the multi-modality of the MOP. This together with the qualitative discussions in Section III-B can be used to a certain extent to explain recent advances in the field of evolutionary many-objective optimization.

We hope that this new insight into the geometry of multiobjective optimization may help researchers in the field of evolutionary computation for further developments of efficient specialized algorithms, particularly when dealing with many-objective problems.

Appendix

The following little discussion shows that by adding objectives to a given MOP the set of local minima cannot get smaller but rather gets bigger which we argue by the set of KKT points (note that every local minimizer is a KKT point). Let a MOP be given consisting of the k objectives $(f_1, \ldots, f_k)$ (denote by MOP1). Further, let an extended model be given by the objectives $(f_1, \ldots, f_k, f_{k+1})$ [denote by (MOP2)] where the first k objectives are identical in MOP1 and MOP2. For simplicity, we assume that all objectives are defined on the same domain $Q \subset \mathbb{R}^n$. Define

$$KKT(i) := \{x \in Q \ : \ x \text{ is KKT point of MOPi}\} \quad i = 1, 2. \tag{30}$$

The following consideration shows that $KKT(1)$ is a subset of $KKT(2)$: let $x \in KKT(1)$, i.e., there exists a convex weight $\alpha \in \mathbb{R}^k$ (i.e., $\alpha_i \geq 0$ for all $i = 1, \ldots, k$ and $\sum_{i=1}^k \alpha_i = 1$) such that $\sum_{i=1}^k \alpha_i \nabla f_i(x) = 0$. Since $\tilde{\alpha} = (\alpha, 0) \in \mathbb{R}^{k+1}$ is also a convex weight and $\sum_{i=1}^{k+1} \tilde{\alpha}_i \nabla f_i(x) = 0$ it follows that x is also included in $KKT(2)$.

Further, it is for instance every substationary point of f_{k+1} (i.e., a point $x \in Q$ which satisfies $\nabla f_{k+1}(x) = 0$) also included in $KKT(2)$ [for this, choose $\alpha = (0, \ldots, 0, 1)$], and hence, $KKT(2)$ is typically a strict superset of $KKT(1)$. Furthermore, it can be shown that under certain additional assumptions the set $KKT(2)$ forms a k-dimensional object while $KKT(1)$ is $(k-1)$-dimensional [22].

References

[1] H. E. Aguirre and K. Tanaka, "Many-objective optimization by space partitioning and adaptive epsilon-ranking on MNK-landscapes," in *Proc. 4th Int. Conf. EMO*, 2009, pp. 407–422.

[2] H. E. Aguirre and K. Tanaka, "Space partitioning with adaptive epsilon-ranking and substitute distance assignments: A comparative study on many-objective MNK-landscapes," in *Proc. GECCO*, 2009, pp. 547–554.

[3] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler, "Articulating user preferences in many-objective problems by sampling the weighted hypervolume," in *Proc. 11th Annu. Conf. GECCO*, 2009, pp. 555–562.

[4] P. A. N. Bosman and E. D. de Jong, "Exploiting gradient information in numerical multiobjective evolutionary optimization," in *Proc. GECCO*, vol. 1. Jun. 2005, pp. 755–762.

[5] D. Brockhoff, T. Friedrich, N. Hebbinghaus, and C. Klein, "On the effects of adding objectives to plateau functions," *IEEE Trans. Evol. Comput.*, vol. 13, no. 3, pp. 591–603, Jun. 2009.

[6] D. Brockhoff and E. Zitzler, "Are all objectives necessary? On dimensionality reduction in evolutionary multiobjective optimization," in *Proc. 9th Int. Conf. PPSN*, 2006, pp. 533–542.

[7] M. Brown and R. E. Smith, "Directed multiobjective optimization," *Int. J. Comput. Syst. Signals*, vol. 6, no. 1, pp. 3–17, 2005.

[8] C. A. Coello Coello, G. B. Lamont, and D. A. Van Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd ed. New York: Springer, Sep. 2007.

[9] D. Corne and J. D. Knowles, "Techniques for highly multiobjective optimization: Some nondominated points are better than others," in *Proc. 9th Annu. Conf. GECCO*, 2007, pp. 773–780.

[10] D. W. Corne, J. D. Knowles, and M. J. Oates, "The Pareto-envelope based selection algorithm for multiobjective optimization," in *Proc. 6th Int. Conf. PPSN*, 2000, pp. 869–878.

[11] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. Chichester, U.K.: Wiley, 2001.

[12] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," *Complex Syst.*, vol. 9, no. 2, pp. 115–148, 1995.

[13] K. Deb, M. Mohan, and S. Mishra, "Evaluating the ϵ-dominated based multiobjective evolutionary algorithm for a quick computation of Pareto-optimal solutions," *Evol. Comput.*, vol. 13, no. 4, pp. 501–525, 2005.

[14] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.

[15] K. Deb and D. K. Saxena, "Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems," in *Proc. IEEE CEC*, 2006, pp. 3353–3360.

[16] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable test problems for evolutionary multiobjective optimization," in *Evolutionary Multiobjective Optimization. Theoretical Advances and Applications*, A. Abraham, L. Jain, and R. Goldberg, Eds. New York: Springer, 2005, pp. 105–145.

[17] M. Farina and P. Amato, "A fuzzy definition of 'optimality' for many-criteria optimization problems," *IEEE Trans. Syst., Man, Cybern.*, vol. 34, no. 3, pp. 315–326, May 2004.

[18] P. Fleming, R. C. Purshouse, and R. J. Lygoe, "Many-objective optimization: An engineering design perspective," in *Proc. 3rd Int. Conf. EMO*, LNCS 3410. Mar. 2005, pp. 14–32.

[19] J. Fliege and B. F. Svaiter, "Steepest descent methods for multicriteria optimization," *Math. Methods Oper. Res.*, vol. 51, no. 3, pp. 479–494, 2000.

[20] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multiobjective optimization," *Evol. Comput.*, vol. 3, no. 1, pp. 1–16, 1995.

[21] T. Hanne, "On the convergence of multiobjective evolutionary algorithms," *Eur. J. Oper. Res.*, vol. 117, no. 3, pp. 553–564, 1999.

[22] C. Hillermeier, *Nonlinear Multiobjective Optimization - A Generalized Homotopy Approach*. Boston, MA: Birkhäuser, 2001.

[23] J. Horn, "Multicriterion decision making," in *Handbook of Evolutionary Computation*. Oxford, U.K.: Oxford University Press, 1997, pp. 1–9.

[24] E. J. Hughes, "Multiple single objective Pareto sampling," in *Proc. IEEE CEC*, 2003, pp. 2678–2684.

[25] E. J. Hughes, "Evolutionary many-objective optimization: Many ones or one many?" in *Proc. IEEE CEC*, 2005, pp. 222–227.

[26] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization: A short review," in *Proc. IEEE CEC*, 2008, pp. 2419–2426.

[27] A. López Jaimes, C. A. Coello Coello, and J. E. Urías Barrientos, "Online objective reduction to deal with many-objective problems," in *Proc. 5th Int. Conf. EMO*, LNCS 5467. 2009, pp. 423–437.

[28] W. E. Karush, "Minima of functions of several variables with inequalities as side conditions," Ph.D. dissertation, Univ. Chicago, Chicago, IL, 1939. AQ:3

[29] J. Knowles and D. Corne, "Quantifying the effects of objective space dimension in evolutionary multiobjective optimization," in *Proc. 4th Int. Conf. EMO*, LNCS 4403. Mar. 2007, pp. 757–771.

[30] J. D. Knowles and D. W. Corne, "Approximating the nondominated front using Pareto archived evolution strategy," *Evol. Comput.*, vol. 8, no. 2, pp. 149–172, 2000.

[31] J. D. Knowles, R. A. Watson, and D. W. Corne, "Reducing local optima in single-objective problems by multi-objectivization," in *Proc. 1st Int. Conf. EMO*, 2001, pp. 269–283.

[32] M. Köppen and R. Vicente-Garcia, "A fuzzy scheme for the ranking of multivariate data and its application," in *Proc. IEEE Annu. Meeting Fuzzy Information*, 2004, pp. 140–145.

[33] M. Köppen, R. Vicente-Garcia, and B. Nickolay, "Fuzzy Pareto-dominance and its application in evolutionary multiobjective optimization," in *Proc. 3rd Int. Conf. EMO*, LNCS 3410. 2005, pp. 399–412.

[34] M. Köppen and K. Yoshida, "Many-objective particle swarm optimization by gradual leader selection," in *Proc. 8th ICANNGA*, 2007, pp. 323–331.

[35] M. Köppen and K. Yoshida, "Substitute distance assignment in NSGA-II for handling many-objective optimization problems," in *Proc. 4th Int. Conf. EMO*, LNCS 4403. 2007, pp. 727–741.

[36] H. Kuhn and A. Tucker, "Nonlinear programming," in *Proc. 2nd Berkeley Symp. Math. Stat. Probab.*, 1951, pp. 481–492.

[37] A. Lara, G. Sanchez, C. A. Coello Coello, and O. Schütze, "HCS: A new local search strategy for memetic multiobjective evolutionary algorithms," *IEEE Trans. Evol. Comput.*, vol. 14, no. 1, pp. 112–132, Feb. 2010.

[38] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multiobjective optimization," *Evol. Comput.*, vol. 10, no. 3, pp. 263–282, 2002.

[39] S. Mostaghim and H. Schmeck, "Distance based ranking in many-objective particle swarm optimization," in *Proc. 10th Int. Conf. PPSN*, 2008, pp. 753–762.

[40] V. Pareto, *Manual of Political Economy*. New York: MacMillan, 1971.

[41] K. Praditwong and X. Yao, "How well do multiobjective evolutionary algorithms scale to large problems?" in *Proc. IEEE CEC*, 2007, pp. 3959–3966.

[42] R. C. Purshouse and P. J. Fleming, "Evolutionary many-objective optimization: An exploratory analysis," in *Proc. CEC*, 2003, pp. 2066–2073.

[43] G. Rudolph and A. Agapie, "On a multiobjective evolutionary algorithm and its convergence to the Pareto set," in *Proc. CEC*, 2000, pp. 1010–1016.

[44] H. Sato, H. E. Aguirre, and K. Tanaka, "Controlling dominance area of solutions and its impact on the performance of MOEAs," in *Proc. 4th Int. Conf. EMO*, LNCS 4403. 2007, pp. 5–20.

[45] D. J. Schaffer, "Multiple objective optimization with vector evaluated genetic algorithms," in *Proc. 1st Int. Conf. Genet. Algorithms*, 1985, pp. 93–100.

[46] O. Schütze, C. A. Coello Coello, E. Tantar, and E.-G. Talbi, "Computing finite size representations of the set of approximate solutions of an MOP with stochastic search algorithms," in *Proc. 10th Annu. Conf. GECCO*, 2008, pp. 713–720.

[47] O. Schütze, M. Laumanns, C. A. Coello Coello, M. Dellnitz, and E.-G. Talbi, "Convergence of stochastic search algorithms to finite size Pareto set approximations," *J. Global Optimization*, vol. 41, no. 4, pp. 559–577, 2008.

[48] O. Schütze, M. Laumanns, E. Tantar, C. A. Coello Coello, and E.-G. Talbi. "Convergence of stochastic search algorithms to gap-free Pareto front approximations," in *Proc. GECCO*, 2007, pp. 892–901.

[49] O. Schütze, M. Laumanns, E. Tantar, C. A. Coello Coello, and E.-G. Talbi, "Computing gap free Pareto front approximations with stochastic search algorithms," *Evol. Comput.*, vol. 18, no. 1, pp. 65–96, 2010.

[50] H. K. Singh, A. Isaacs, T. Ray, and W. Smith, "A study on the performance of substitute distance based approaches for evolutionary many objective optimization," in *Proc. 7th Int. Conf. SEAL*, 2008, pp. 401–410.

[51] O. Teytaud, "On the hardness of offline multiobjective optimization," *Evol. Comput.*, vol. 15, no. 4, pp. 475–491, 2007.

[52] D. A. Van Veldhuizen, "Multiobjective evolutionary algorithms: Classifications, analyses, and new innovations," Ph.D. dissertation, Dept. Electr. Comput. Eng., Graduate School Eng., Air Force Instit. Technol., Wright-Patterson AFB, OH, May 1999.

[53] T. Wagner, N. Beume, and B. Naujoks, "Pareto-, aggregation-, and indicator-based methods in many-objective optimization," in *Proc. 3rd Int. Conf. EMO*, LNCS 4403. 2007, pp. 742–756.

[54] P. Winkler, "Random orders," *Order*, vol. 1, no. 4, pp. 317–331, Dec. 1985.

[55] E. Zitzler, "Evolutionary algorithms for multiobjective optimization: Methods and applications," Ph.D. dissertation, ETH Zurich, Zurich, Switzerland, 1999.

[56] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," in *Proc. Evol. Methods Des., Optimization Control Applicat. Ind. Problems (EUROGEN)*, 2002, pp. 95–100.

Oliver Schütze received the Diploma degree in mathematics from the University of Bayreuth, Bayreuth, Germany, and the Ph.D. degree in natural sciences from the University of Paderborn, Paderborn, Germany, in 1999 and 2004, respectively.

He is currently a Professor (CINVESTAV-3A Researcher) with the Department of Computer Science, CINVESTAV-IPN, Mexico City, Mexico.

Adriana Lara received the B.S. degree in physics and mathematics from the National Polytechnic Institute of Mexico, Mexico City, Mexico, in 2001, and the M.S. degree in electrical and computer engineering from CINVESTAV-IPN, Mexico City, in 2003, where she is currently working toward the Ph.D. degree in computer sciences from the Department of Computer Science.

She was an Associate Professor with the National Polytechnic Institute of Mexico, Mexico City. Her current research interests include multiobjective optimization, gradient-based techniques, and hybrid algorithms.

Carlos A. Coello Coello (SM'XX) received the B.S. degree in civil engineering from the Universidad Autónoma de Chiapas, Chiapas, Mexico, and the M.S. and Ph.D. degrees in computer science from Tulane University, New Orleans, LA, in 1991, 1993, and 1996, respectively. AQ:4

He is currently a Professor (CINVESTAV-3E Researcher) with the Department of Computer Science, CINVESTAV-IPN, Mexico City, Mexico. He has authored and co-authored over 250 technical papers and several book chapters. He has also co-authored the book *Evolutionary Algorithms for Solving Multi-Objective Problems* (Berlin, Germany: Springer, 2007, 2nd ed.). His major research interests include evolutionary multiobjective optimization and constraint-handling techniques for evolutionary algorithms.

Dr. Coello Coello is currently an Associate Editor of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION and serves on the editorial board of 14 other international journals. He also chairs the IEEE Computationally Intelligent Society Task Force on Multiobjective Evolutionary Algorithms. He received the 2007 National Research Award from the Mexican Academy of Sciences in the area of Exact Sciences. He is a member of the ACM, Sigma Xi, and the Mexican Academy of Sciences.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ:1= Please provide index terms for this paper.
AQ:2= Please provide the expanded form of "MOEA."
AQ:3= Please provide department name in Refs. [28], [55].
AQ:4= Please provide membership year of "Coello Coello."

END OF ALL QUERIES

On the Influence of the Number of Objectives on the Hardness of a Multiobjective Optimization Problem

Oliver Schütze, Adriana Lara, and Carlos A. Coello Coello, *Senior Member, IEEE*

***Abstract*—In this paper, we study the influence of the number of objectives of a continuous multiobjective optimization problem on its hardness for evolution strategies which is of particular interest for many-objective optimization problems. To be more precise, we measure the hardness in terms of the evolution (or convergence) of the population toward the set of interest, the Pareto set. Previous related studies consider mainly the number of nondominated individuals within a population which greatly improved the understanding of the problem and has led to possible remedies. However, in certain cases this ansatz is not sophisticated enough to understand all phenomena, and can even be misleading. In this paper, we suggest alternatively to consider the probability to improve the situation of the population which can, to a certain extent, be measured by the sizes of the descent cones. As an example, we make some qualitative considerations on a general class of uni-modal test problems and conjecture that these problems get harder by adding an objective, but that this difference is practically not significant, and we support this by some empirical studies. Further, we address the scalability in the number of objectives observed in the literature. That is, we try to extract the challenges for the treatment of many-objective problems for evolution strategies based on our observations and use them to explain recent advances in this field.**

AQ:1 ***Index Terms*—XXX, XXX, XXX.**

I. Introduction

EVOLUTIONARY algorithms for the numerical treatment of multiobjective optimization problems (MOPs) have been studied intensively during the last few years (see [11], [8] and references therein). Typically, few objectives (i.e., mainly two or three) are being investigated resulting in a variety of very efficient algorithms. The consideration of many (i.e., more than three) objectives, however, is a relatively young field and is yet not studied thoroughly enough. With this paper, we want to contribute to this field by looking at the influence of the number k of objectives in a continuous MOP on the hardness of the problem. To be more precise, we try to understand the behavior of the evolution with respect to k by looking at the descent cones of the individuals of the populations. The resulting analysis is of qualitative nature; however, it can for instance be used to disprove a common belief, namely that the addition of an objective makes a problem per se harder. Further, the new ansatz can be used to explain recent advances in the field of evolutionary many-objective optimization, and is thus hopefully helpful for designers of evolutionary algorithms aimed to deal with such problems.

When investigating continuous MOPs with respect to k, two facts have to be considered: 1) the solution set, the so-called Pareto set, forms typically a $(k-1)$-dimensional set [22], and 2) the problem gets harder the more local solutions it contains and the smaller the basin of attraction for the global solutions are since then the chance increases that a population can get stuck in locally optimal regions. The choice of k has thus, by 1), a direct influence on the dimension of the Pareto set, and hence, also on the hardness of the problem. If, for instance, $N_2 = 100$ points are chosen to obtain a "sufficient" representation of a solution set for $k = 2$ in the Hausdorff sense (which is a typical value in the literature), in principle the practically intractable amount of $N_{15} = 100^{14} = 10^{28}$ elements is required to obtain the same approximation quality for $k = 15$. Even if the lower bound of $N_2 = 2$ elements is used to "represent" the Pareto set for $k = 2$, still $N_{15} = 16\,384$ elements are needed to obtain the same (low) approximation quality for $k = 15$ (see also [51] for a related discussion on the required number of comparisons with respect to k). As a possible remedy, one can in certain cases try to reduce the number of objectives (e.g., [6], [15], [27]) since in practice it may happen that several objectives are correlated. Since we are interested in the influence of k we will not follow that approach. Another more practical remedy researchers dealing with evolutionary many-objective optimization have chosen is to bound the population/archive size to a moderate (and hence tractable) number for all values of k (say, $N = 100$). We will, in the following, consider that scenario and will restrict ourselves to investigate the evolution of these N individuals toward the Pareto set. That is, we will only consider the convergence of the individuals and will leave out the (very important) question of the distribution of the limit population since this is still an open problem. It has to be noted that by using the descent cones only the convergence (in terms of the semi-distance *dist*) of the population toward the set of interest can be understood. Further important aspects are not treated here. As discussed

Received July 6, 2009; revised December 29, 2009, March 31, 2010, and June 10, 2010; accepted July 6, 2010. The work of A. Lara was supported by CONACyT through a scholarship to pursue graduate studies at the Department of Computer Science, CINVESTAV-IPN. The work of C. A. Coello Coello was supported by CONACyT, under Project 103570.

The authors are with the Department of Computer Science, CINVESTAV-IPN, Mexico City, D.F. 07300, Mexico (e-mail: schuetze@cs.cinvestav.mx; alara@computacion.cs.cinvestav.mx; ccoello@cs.cinvestav.mx).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TEVC.2010.2064321

1089-778X/$26.00 © 2010 IEEE

above, an approximation in the Hausdorff sense has strong limitations with respect to the value of k; however, there are further interesting metrics for the treatment of many-objective problems such as the set coverage metric or the hypervolume metric [55], as considered in [29] and [3], respectively. To understand the evolution of the populations with respect to these metrics, a (sole) consideration of the descent cones does not seem to be adequate.

While the choice of k has a direct influence on the dimension of the solution set the relation to 2) is rather indirect. On the one hand, an additional objective certainly increases the chance that more locally optimal solutions exist since every local solution of each objective is also a local solution of the MOP (see the Appendix). Hence, every multi-modal objective makes the problem harder as it is the case for the DTLZ test problems [16] which are often considered in the context of the evaluation of many-objective evolutionary algorithms. On the other hand, this increase of hardness comes rather from the multi-modality of the model than from the additional objective and can be "substituted" by increasing the multi-modality of the already existing objectives. However, it is a common belief that more objectives make a MOP harder (e.g., [11], [18], [20], [23]) which has an impact on the design in particular of evolution strategies for the treatment of many-objective optimization problems. As reason for this behavior it is sometimes argued that the number of incomparable solutions increases if further objectives are added to a problem (empirically studied, e.g., in [26], [29], [35], and [42], and proven in [54]), and thus, that the evolution of the populations toward the Pareto sets is slowed down.

The aim of this paper is to investigate the influence of the hardness of a problem for an evolutionary search procedure with respect to k. Instead of looking at the number of nondominated solutions within a population, we will focus on the ability of the populations to evolve toward the Pareto sets. Since there is a certain relation between the probability to (locally) improve an individual x by the generational operators and the size of the descent cone at x, we will use and adapt some considerations from [7] of the sizes of the cones in order to try to explain the behavior of the evolution. To handle 1), we will restrict the population size to a fixed value as discussed above, and to avoid the problem described in 2), we will concentrate on uni-modal models. We will argue that a MOP (theoretically) indeed gets harder when adding an objective, but that this difference is—at least for uni-modal models and under an additional assumption on the evolutionary algorithm—not significant, and demonstrate this empirically on three examples. Further on, we will address the treatment of general models where such a scalability has been observed by many researchers so far. Based on our considerations we try to extract the challenges for many-objective evolutionary algorithms and give an attempt to explain recent advances in this field in light of the new insight. A critical discussion on the influence of k for discrete MOPs can be found in [5], but the study presented in this paper seems to be the first one for continuous models. Since our ansatz is using descent cones, the conclusions we draw are restricted to continuous models. Similar explanations for combinatorial problems do not seem to exist.

The remainder of this paper is organized as follows. Section II gives the required background for the understanding of the sequel. In Section III, we investigate a class of uni-modal test functions analytically and empirically with respect to the influence of the number of objectives to the hardness of the problem. In Section IV, we discuss our results and give an attempt to explain recent advances in the field of evolutionary many-objective optimization. Finally, we draw some conclusions in Section V.

II. Background

In the following, we consider continuous MOPs which are of the following form:

$$\min_{x \in Q}\{F(x)\} \qquad \text{(MOP)}$$

where $Q \subset \mathbb{R}^n$ is the domain and the function F is defined as the vector of the objective functions

$$F : Q \to \mathbb{R}^k \qquad F(x) = (f_1(x), \ldots, f_k(x))$$

and where each objective $f_i : Q \to \mathbb{R}$ is continuous. The optimality of a MOP is defined by the concept of *dominance* [40].

Definition 2.1:

1) Let $v, w \in \mathbb{R}^k$. Then the vector v is *less than* w ($v <_p w$), if $v_i < w_i$ for all $i \in \{1, \ldots, k\}$. The relation $\leq_p$ is defined analogously.
2) A vector $y \in \mathbb{R}^n$ is *dominated* by a vector $x \in \mathbb{R}^n$ ($x \prec y$) with respect to (MOP) if $F(x) \leq_p F(y)$ and $F(x) \neq F(y)$, else y is called non-dominated by x.
3) A point $x \in Q$ is called *(Pareto) optimal* or a *Pareto point* if there is no $y \in Q$ which dominates x.

The set of all Pareto optimal solutions is called the *Pareto set*, and is denoted by P_Q. The image $F(P_Q)$ of the Pareto set is called the *Pareto front*. If required, we will denote the Pareto set of a particular MOP by $P_Q(MOP)$ to avoid confusion. In case all the objectives of the MOP are differentiable, the following famous theorem of Kuhn and Tucker [36] states a necessary condition for Pareto optimality for unconstrained MOPs.

Theorem 2.2: Let x^* be a Pareto point of (MOP), then there exists a vector $\alpha \in \mathbb{R}^k$ with $\alpha_i \geq 0$, $i = 1, \ldots, k$, and $\sum_{i=1}^k \alpha_i = 1$ such that

$$\sum_{i=1}^{k} \alpha_i \nabla f_i(x^*) = 0. \qquad (1)$$

The theorem claims that the vector of zeros can be written as a convex combination of the gradients of the objectives at every Pareto point. Obviously, (1) does not state a sufficient condition for Pareto optimality. On the other hand, points satisfying (1) are certainly "Pareto candidates."

Definition 2.3: A point $x \in \mathbb{R}^n$ is called a *Karush–Kuhn–Tucker point*[1] (KKT–point) if there exist scalars $\alpha_1, \ldots, \alpha_k \geq 0$ such that $\sum_{i=1}^k \alpha_i = 1$ and that (1) is satisfied.

[1] Named after the works of Karush [28], and Kuhn and Tucker [36].

Next, we define some distances between points as well as between different sets.

Definition 2.4: Let $u, v \in \mathbb{R}^n$ and $A, B \subset \mathbb{R}^n$. The maximum norm distance d_∞, the semi-distance $\mathrm{dist}(\cdot,\cdot)$ and the *Hausdorff distance* $d_H(\cdot,\cdot)$ are defined as follows:

1) $d_\infty(u, v) := \max_{i=1,\ldots,n} |u_i - v_i|$;

2) $dist(u, A) := \inf_{v \in A} d_\infty(u, v)$;

3) $dist(B, A) := \sup_{u \in B} dist(u, A)$;

4) $d_H(A, B) := \max\{dist(A, B), dist(B, A)\}$.

As discussed above, we are in particular interested in the convergence of the archive entries toward the set of interest. In case of the Pareto front, it is

$$\mathrm{dist}(F(A_l), F(P_Q)) \tag{2}$$

where $A_l = \{a_1, \ldots, a_m\}$ is the archive in generation l. Since *dist* (and thus also d_H) is sensitive to outliers which is a potential drawback when measuring the solution of stochastic algorithms one can use instead the *generational distance* (GD, see [52]) which measures the average distance of the elements of A_l to the Pareto front

$$GD(A_l) := \frac{1}{m}\sqrt{\sum_{i=1}^{l} dist(F(a_i), F(P_Q))^2}. \tag{3}$$

The Pareto sets of the test functions considered in the following are given by simplexes which are defined as follows.

Definition 2.5: Let $v_1, \ldots, v_k \subset R^n$, $n \geq k$, be given. The set

$$S(v_1, \ldots, v_k) := \left\{ \sum_{i=1}^{k} \lambda_i v_i \;:\; \lambda \in [0,1]^k, \text{ and } \sum_{i=1}^{k} \lambda_i = 1 \right\} \tag{4}$$

is called the $(k-1)$-simplex of $v_1, \ldots, v_k$.

A hyperplane $H = H(\tilde{x}, \eta)$ in n-dimensional space is defined by a point $\tilde{x} \in H$ and a normal vector $\eta \in \mathbb{R}\backslash\{0\}$, that is

$$H(\tilde{x}, \eta) = \{x \in \mathbb{R}^n \;:\; \langle x - \tilde{x}, \eta\rangle = 0\} \tag{5}$$

where $\langle\cdot,\cdot\rangle$ defines the standard scalar product. The point $p(x)$ which is closest to H is given by

$$p(x) = x - \frac{\langle x - \tilde{x}, \eta\rangle}{\langle \eta, \eta\rangle}\eta. \tag{6}$$

III. Investigation of a Class of Uni-Modal Models

A. *A Class of Test Problems with Simplicial Pareto Sets*

Here, we construct a set of quadratic (and hence uni-modal) test functions where the Pareto sets are given by simplexes which eases the computation of the distance of a point to the Pareto set and front. The resulting models we consider are slight variants of the P^* problems introduced in [34] tailored to our needs.

1) *Construction:* First we construct the base problem. Given points $a_1, \ldots, a_k \in \mathbb{R}^n$, we define the MOP as follows:

$$\begin{aligned} &\min F : \mathbb{R}^n \to \mathbb{R}^k \\ &f_i(x) = \|x - a_i\|_2^2 = \sum_{j=1}^{n} (x_j - a_{i,j})^2 \end{aligned} \tag{7}$$

where $a_{i,j}$ denotes the jth entry of a given vector a_i. The Pareto set of the problem defined by (7) [in short MOP(7)] is given by the simplex spanned by the k minimizers a_i.

Proposition 3.1: $P_Q(MOP\ (7)) = S(a_1, \ldots, a_k)$.

Proof: It is $\nabla f_i(x) = 2(x - a_i)$. Let $x \in S(a_1, \ldots, a_k)$, i.e., there exist scalars $\lambda_1, \ldots, \lambda_k \geq 0$ with $\sum_{i=1}^{k} \lambda_i = 1$ such that $x = \sum_{i=1}^{k} \lambda_i a_i$. Then

$$\begin{aligned} \sum_{i=1}^{k} \lambda_i \nabla f_i(x) &= \sum_{i=1}^{k} \lambda_i 2(x - a_i) = 2(x \underbrace{\sum_{i=1}^{k} \lambda_i}_{=1} - \sum_{i=1}^{k} \lambda_i a_i) \\ &= 2\left(x - \sum_{i=1}^{k} \lambda_i a_i\right) = 0. \end{aligned} \tag{8}$$

The claim follows since MOP (7) is strictly convex, and thus, the Pareto set is equal to the set of Karush–Kuhn–Tucker (KKT) points. ∎

The problem is quadratic and unconstrained. Note that for the special case $n = 1$, $k = 2$, $a_1 = 0$, and $a_2 = 1$ the MOP (7) coincides with the well-known problem of Schaffer [45]. The authors of [34] propose to locate all the minima a_i on an Euclidean plane which results in a 2-D Pareto set. In order, e.g., to obtain a $(k-1)$-dimensional object, the volume of $S(a_1, \ldots, a_k)$ has to be positive, i.e., the $k-1$ difference vectors $a_2 - a_1, \ldots, a_k - a_1$ have to be linearly independent.

In the following, we use Proposition 1 to construct constrained problems with variable dimension of the solution set. For this, we will use hyperplanes. Given a hyperplane $H = H(\tilde{x}, \eta)$, there exists for every point $x \in \mathbb{R}^n$ a $\lambda = \lambda(x) \in \mathbb{R}$ such that

$$x - p(x) = \lambda\eta \tag{9}$$

which can be used to divide the space $\mathbb{R}^n$ as follows. Let $j \in \{1, \ldots, n\}$ such that $\eta_j \neq 0$, then we define

$$\begin{aligned} &g_H : \mathbb{R}^n \to \mathbb{R} \\ &g_H(x) = \frac{x_j - p(x)_j}{\eta_j} \end{aligned} \tag{10}$$

and the constrained MOP is

$$\begin{aligned} &min\ F(x) \\ &s.t.\ g_H(x) \leq 0 \end{aligned} \tag{11}$$

where F is as defined in (7). Thus, the domain is given by $Q = \{x \in \mathbb{R} : g_H(x) \leq 0\}$. Constrained problems can now be constructed by using MOP (7) and placing the a_i's at the boundary of Q. The following result shows how further constrained MOPs can be generated with different dimensions of the Pareto set (see also Fig. 1). Further on, we give one such example.

Proposition 3.2: Let $H = H(\tilde{x}, \eta)$ be a hyperplane and $a_1, \dots, a_k \in \mathbb{R}^n$ such that

$$a_1, \dots, a_l \in H \quad l \le k \tag{12}$$

and

$$\begin{aligned} g_H(a_i) > 0 \quad & i = l+1, \dots, k \\ p(a_i) \in S(a_1, \dots, a_l) \quad & i = l+1, \dots, k. \end{aligned} \tag{13}$$

Then, the Pareto set of MOP (11) is given by

$$P_Q(MOP\ (11)) = S(a_1, \dots, a_l). \tag{14}$$

Proof: By Proposition 1, it is clear that: 1) $S(a_1, \dots, a_l) \subset P_Q$, and 2) none of the points $x \in \mathbb{R}$ with $g_H(x) < 0$, i.e., the points where g_H is inactive, is Pareto optimal [else 0 can be expressed as a convex combination of the objectives' gradients, but this was prevented by the first assumption in (13)]. It remains to show that $H \backslash S(a_1, \dots, a_l)$ is not contained in P_Q. For $x \in H \backslash S(a_1, \dots, a_l)$ choose $z \in S(a_1, \dots, a_l)$ such that

$$z \in \text{argmin}_{s \in S(a_1, \dots, a_l)} \|x - s\|_2. \tag{15}$$

Since $S(a_1, \dots, a_l)$ is a convex set and $x \notin S(a_1, \dots, a_l)$ it follows that

$$\|s - z\|_2 < \|s - x\|_2 \quad \forall s \in S(a_1, \dots, a_l). \tag{16}$$

Since (16) holds for a_i, $i = 1, \dots, l$, it follows that $f_i(z) < f_i(x)$, $i = 1, \dots, l$. Further, by the same argument on $p(a_i)$, $i = l+1, \dots, k$, and Pythagoras

$$\begin{aligned} & x \in H \;\Rightarrow\; \|a_i - x\|_2^2 = \|a_i - p(a_i)\|_2^2 + \|p(a_i) - x\|_2^2 \\ & i = l+1, \dots, k \end{aligned} \tag{17}$$

it follows that also $f_i(z) < f_i(x)$, $i = l+1, \dots, k$, and thus, that $F(z) < F(x)$, which implies that $x \notin P_Q$ which concludes the proof. ∎

If for instance $H = H(e_1, \eta)$ is chosen as

$$\eta = (\underbrace{-1, \dots, -1}_{k}, \underbrace{0, \dots, 0}_{n-k})^T \tag{18}$$

and $a_i = e_i$, $i = 1, \dots, k$, then $a_i \in H$, $i = 1, \dots, k$ ($p(a_i) = a_i$) and thus, $P_Q = S(e_i, \dots, e_k)$. The dimension of the solution set can be reduced by one if choosing, e.g., $a_i = e_i$, $i = 1, \dots, k-1$, $a_k = 0$, and $H = H(e_1, \eta)$ with

$$\eta = (\underbrace{-1, \dots, -1}_{k-1}, \underbrace{0, \dots, 0}_{n-k+1})^T. \tag{19}$$

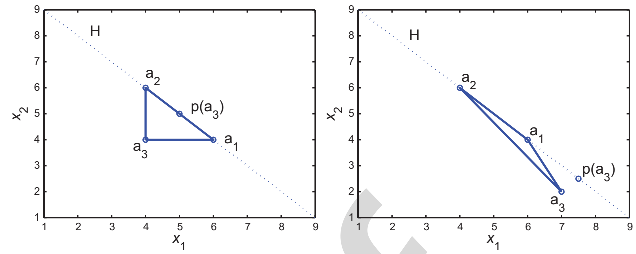

Fig. 1. Two examples where the facet of a 3-simplex is included in the hyperplane. Left: the Pareto set of MOP (11) is given by $S(a_1, a_2)$ since $p(a_3) \in S(a_1, a_2)$. Right: $p(a_3) \notin S(a_1, a_2)$, and thus, the Pareto set is not equal to the facet $S(a_1, a_2)$.

It is $p(a_k) = \frac{-1}{k-1}\eta \in S(a_1, \dots, a_{k-1})$ [using the weights $\alpha_i = 1/(k-1)$] and $g_H(a_k) = 1/(k-1) > 0$, and thus, it follows by Proposition 2 that $P_Q = S(a_1, \dots, a_{k-1})$.

Continuing in a similar manner, the dimension of the Pareto set can be reduced. The extreme situation—i.e., that P_Q consists of one single solution—can, e.g., be obtained as follows: set $a_1 = e_1$, and $a_i = \lambda_i e_i$, $\lambda_i < 1$, for $i = 2, \dots, k$, and $H = H(e_1, \eta)$ with $\eta = (-1, 0, \dots, 0)^T$. Then, it is $p(a_i) = e_1$ and $g(a_i) = 1 - \lambda_i > 0$ for $i = 2, \dots, k$, and thus, $P_Q = \{e_1\}$.

2) *Test Problems:* Based on the above observations, we propose two test functions which are used to investigate the hardness of a MOP with respect to the number of objectives.

a) *PS1:* Given vectors $a_1, \dots, a_k \in \mathbb{R}^n$, $n \ge k$, we define the first test problem PS1 as in (7). For the a_i's we suggest choosing $a_i = e_i$, and as domain $Q = [-10, 10]$. By Proposition 1 it follows that

$$P_Q(PS1) = S(e_1, \dots, e_k). \tag{20}$$

b) *PS2:* Here we define a constrained model where the dimension of the Pareto set can be chosen between 0 and $k-1$, where k is the number of objectives: given a number $1 \le l \le k$, we define PS2(l) as follows. Let $H = H(e_1, \eta)$ with

$$\eta = (\underbrace{-1, \dots, -1}_{l}, \underbrace{0, \dots, 0}_{n-l}) \tag{21}$$

let g_H as in (10), and F as in (7), where $a_i = e_i$, $i = 1, \dots, l$, and $a_j = -\frac{1}{l}\eta + \frac{j-l}{l}\eta$, $j = l+1, \dots, k$. Then PS2(l) reads as follows:

$$\begin{aligned} & \min F(x) \\ & \text{s.t. } x_i \in [-10, 10]^n \; i = 1, \dots, n \\ & \quad g_H(x) \le 0. \end{aligned} \tag{22}$$

Due to the discussion in the previous subsection it is

$$P_Q(PS2(l)) = S(e_1, \dots, e_l) \tag{23}$$

i.e., a l-simplex which is located within the boundary of the domain. The characteristic of this model is that the Pareto set of PS2(l) for k_1 objectives [denoted by PS2$_{k_1}$(l)] is equal to the Pareto set of PS2$_{k_2}$(l), where k_1 and k_2 are any numbers larger than or equal to l.

B. Hardness of the PS Problems with Respect to k

In the following, we investigate the hardness of the PS test problems by some (non-rigorous) theoretical considerations and by empirical studies.

1) *Qualitative Considerations:* In the following, we consider the PS test problems for general locations of the minima a_i. If further assumptions are required, we will mention them.

For our considerations, we use the descent cones to investigate the hardness of a problem. Given a MOP with s objectives the descent cone at a point $x \in Q$ is given by (e.g., [4])

$$D(f_1, \dots, f_s, x) = \{v \in \mathbb{R}^n \backslash \{0\} \ : \ \langle \nabla f_i(x), v \rangle < 0 \\ \forall\, i = 1, \dots, s.\} \tag{24}$$

$D(f_1, \dots, f_s, x)$ is the set of all directions in which dominating points can be found, i.e., for each $v \in D(f_1, \dots, f_s, x)$ there exists a (possibly small) $t \in \mathbb{R}_+$ such that $F(x + tv) <_p F(x)$. There exists a certain relation of the size of the descent cone to the probability to (locally) improve the value of x AQ:2 by the generational operators of a MOEA. For the mutation operator, the relation is proportional when assuming the existence of a suitable or small step size control (i.e., the value of t for the offspring $o := x + tv$). For the most common crossover strategies (e.g., SBX [12]) such a relation still holds; however, the success rate is here in addition depending on the location of the parents. Hence, one can say that a small descent cone results in a small probability of finding a better, i.e., dominating, solution near to x, and large descent cones in turn lead to a larger improvement possibility.

Assume we are given $l + 1$ objectives of the form defined in (7), which are entirely determined by the choice of the a_i's and assume further that $a_{l+1} \notin S(a_1, \dots, a_l)$. Clearly, $D(f_1, \dots, f_{l+1}, x)$, i.e., the descent cone for the $(l + 1)$-objective problem is a subset of $D(f_1, \dots, f_l, x)$, i.e., the according descent cone for the MOP consisting of the first l objectives. The equality of both cones holds if $-\nabla f_{l+1}(x)$ is "between" the vectors $-\nabla f_i(x),\ i = 1, \dots, l$. Since for the PS problems it is $\nabla f_i(x) = 2(x - a_i)$ (i.e., the steepest descent $-\nabla f_i(x)$ points to the minimizer of f_i at every point $x \in Q$) we have

$$D(f_1, \dots, f_{l+1}, x) = D(f_1, \dots, f_l, x) \Leftrightarrow \exists \lambda_1, \dots, \lambda_l \geq 0 : \\ a_{l+1} - x = \sum_{i=1}^{l} \lambda_i (a_i - x). \tag{25}$$

Thus, a necessary condition for the equality of the cones is that $a_{l+1} - x \in \operatorname{span}\{a_1 - x, \dots, a_l - x\}$ by which it follows that the set of points $x \in Q$ which satisfies (25) is maximal l-dimensional (and thus a zero set in Q). To be more precise, for every point x which is not included in the affine subspace

$$A := \operatorname{span}\{a_1, \dots, a_l\} + \left\{ \frac{-a_{l+1}}{\sum_{i=1}^{l} \alpha_i - 1} \right\} \tag{26}$$

where $\alpha \in \mathbb{R}^l$ such that $a_{l+1} - x = \sum_{i=1}^{l} \alpha_i (a_i - x)$ [note that since $a_{l+1} \notin S(a_1, \dots, a_l)$ it is $\sum_{i=1}^{l} \alpha_i \neq 1$, and hence, (26) is well defined], the equality of the cones does not hold. Hence, picking a randomly chosen point $x_0 \in Q$ the probability is one that $D(f_1, \dots, f_{l+1}, x_0)$ of the $(l + 1)$-objective problem is a proper subset of the cone $D(f_1, \dots, f_l, x_0)$ of the related "reduced" l-objective problem. This result is in accord with the observation made in [54] that the number of incomparable solutions generally increases with an increasing number of objectives.

Thus, it can be said that—from a theoretical point of view—the PS problems get harder with increasing number of objectives. Since (25) can in principle be applied to any set of gradients, the statement holds for general MOPs. On the other hand, this (point-wise) observation is of qualitative nature and gives no statement about the quantity of the difference which is needed to judge the hardness of a problem for a given evolutionary search procedure with respect to k. The following qualitative considerations,[2] however, question the common belief that the addition of further objectives makes a given MOP per se harder.

Assume we are given MOP1 which consists of the objectives $f_1, \dots, f_k$ of the form defined in (7) and MOP2 which contains the same k objectives as in MOP1 plus the l objectives $f_{k+1}, \dots, f_{k+l}$. If the initial population P_0 is chosen at random from the domain Q, it can be assumed that most of its individuals $p \in P_0$ are "far away" from both Pareto sets (note that under the reasonable assumption $n > k+l$ both sets $S(e_1, \dots, e_k)$ and $S(e_1, \dots, e_{k+l})$ are zero sets in Q). Thus, the vectors $\{p - a_i\}_{i=1,\dots,s}$ for such an individual p point nearly in the same direction, and this holds for $s = k$ as well as for $s = k + l$. One way to see this is that if a sequence of points is chosen with unbounded increasing distance to all the minima a_i, both simplexes $S(a_1, \dots, a_k)$ and $S(a_1, \dots, a_{k+l})$ shrink in the limit down to a point, and hence, both descent cones $D(f_1, \dots, f_k, p)$ and $D(f_1, \dots, f_{k+l}, p)$ form the same half space as the cones $D(f_i, p),\ i \in 1, \dots, k + l$, for single-objective optimization. This implies that it can be expected that also for finite distances the descent cones $D(f_1, \dots, f_k, p)$ and $D(f_1, \dots, f_{k+l}, p)$ are nearly equal (and large), and thus, that the evolution of the populations should be nearly equal for both problems MOP1 and MOP2. The situation will change after a small number of generations: due to the sizes of the descent cones there is a high chance for improvement, and thus, it can be expected that the sequence of populations performs a certain evolution toward the Pareto set. If so, it cannot be expected any more that the cones have similar sizes. Since MOP2 contains more objectives it is more likely that $D(f_1, \dots, f_{k+l}, p)$ is smaller than $D(f_1, \dots, f_k, p)$ for an element p of the current population. [Compare to the theorem of Kuhn and Tucker: if, for instance, two gradients point in opposite directions then the associated cone defined by (24) is empty. By continuity of F, the descent cones near to KKT points are hence small.] However, this is mainly due to the geometry of multiobjective optimization since the Pareto set of MOP2 is indeed larger [$P_Q(MOP2)$ is $(k+l-1)$-dimensional while $P_Q(MOP1)$ is $(k-1)$-dimensional]. Thus, the evolution has to terminate earlier for MOP2 resulting in smaller cones compared to MOP1. Another point—and this one cannot be explained by looking at the descent cones—is

[2]Here we adapt some observations made in [7] to the present context.

one population-based aspect of MOEAs, namely that single "good" solutions—i.e., solutions which are "near" to the Pareto set—can pull the entire population to the set of interest. Using the dimensionality of the different Pareto sets, it can be argued that the chance to find a "good" solution is higher for MOP2 than for MOP1. Hence, using the dimensionality, the argumentation of the influence of k can be turned: under the above assumption (which we will refer to as the *pulling assumption* in the sequel and which will be discussed in more detail in Section IV) and the additional assumption that the population/archive size is fixed and equal for both MOPs it is rather likely that MOP2 is the easiest model in terms of convergence [i.e., when considering $dist(A_l, P_Q)$].

Concluding, it can be said that by adding an objective in a PS model (or other models), the resulting MOP gets indeed "harder" from a theoretical point of view, but it is ad hoc unclear if the amount is indeed significant since some considerations argue against it. However, the above analysis covers only the extreme situations (points which are either far away or near to P_Q) and is only of qualitative nature. To elucidate this problem sufficiently, empirical studies seem to be required which we will do in the following.

2) *Empirical Studies:* As mentioned before, we are in particular interested in the evolution (or convergence) of the populations toward the set of interest. For this, we use the generational distance defined in (3) and a variant of this indicator which we propose in the following.

Given a population $A = \{a_1, \ldots, a_l\}$, GD measures the average distance of the elements of A to the Pareto front. Since the dimension of the vectors $F(a_i)$ varies with the number of objectives, one may argue that for a comparison which includes different number of objectives GD is not well suited. Thus, we propose here a variant of GD, namely

$$GD_x(A) := \frac{1}{l}\sqrt{\sum_{i=1}^{l} dist(a_i, P_Q)^2} \tag{27}$$

which is analog to GD but measures the averaged distance of A to the Pareto *set*, i.e., in parameter space. Hereby, the distance of a point $a \in A$ to the Pareto set and its image to the Pareto front are given by

$$\begin{aligned} dist(a, P_Q) &= \min_{p\in P_Q} \|a - p\|_2 \\ dist(F(a), F(P_Q)) &= \min_{p\in P_Q} \|F(a) - F(p)\|_2. \end{aligned} \tag{28}$$

These are single-objective optimization problems (SOPs) with n-dimensional parameter space. In case $P_Q = S := S(a_1, \ldots, a_k)$ as for our test problems, (28) can be written as

$$\begin{aligned} dist(a, S) &= \min_{\alpha\in S} \left\| a - \sum_{i=1}^{k} \alpha_i a_i \right\|_2 \\ dist(F(a), F(S)) &= \min_{\alpha\in S} \left\| F(a) - F(\sum_{i=1}^{k} \alpha_i a_i) \right\|_2. \end{aligned} \tag{29}$$

Since the SOPs in (29) are convex problems (domain and objective are convex) with k free parameters, it can easily be solved with standard mathematical techniques (note that in the context of scalar optimization, a problem is noted as small if the dimension of the parameter space is less than 10 000, which is definitely beyond the scope of many-objective optimization).

We have chosen to take NSGA-II [14] for our empirical studies since this algorithm was shown to scale badly with increasing number of objectives for certain models (e.g., [53]). Additionally, we have made (but do not display) analog computations with SPEA2 [56] which confirmed the results shown below.

Figs. 2–5 show some numerical results obtained by NSGA-II for PS1 and PS2 [using $l = k$, denoted here by $PS2_k$(k) to avoid confusion] and for different numbers k of objectives. In all examples, we have used parameter dimension $n = 30$, population size $N_p = 100$, and the probabilities $p_c = 0.85$ and $p_m = 0.05$ for crossover and mutation, respectively. The initial population P_0 has been chosen randomly from $I := [9, 10]^{30}$, since by the above discussion for every point $x \in I$ the descent cone $D(f_1, \ldots, f_{k+l}, x)$ of the $(k + l)$-objective problem is a proper subset of the cone $D(f_1, \ldots, f_k, x)$ of the reduced problem (analog empirical studies where P_0 has been chosen randomly from $Q_i, i = 1, 2$, however, have led to the same results). For both the unconstrained and the constrained case as well as for a measurement in parameter and image space (GD_x and GD, respectively) the same behavior can be observed: in the large scale, i.e., when considering all 500 generations, the evolution of the populations is basically the same (note that there is a difference of 12 objectives). When zooming into the figures, little differences appear, and as anticipated, the values of GD_x and GD get (little) larger with increasing number of objectives (note the difference of the values with the initial values of GD and GD_x, respectively).

Whereas the results can be explained to a certain extent by the above considerations, a sole consideration of the number of nondominated solutions in a population may be misleading in this example. Fig. 6 shows the (averaged) number of nondominated solutions for the PS1 problems within the populations found by NSGA-II, and here, the differences are significant. For instance, for $k = 3$ there are about 90% of dominated solutions after 100 generations (and about 50% of dominated solutions after 200 generations) while for $k \geq 10$ practically all members of a population are mutually nondominating after about 100 generations. Hence, by only looking at these values one could have come to the conclusion that the problem gets clearly harder with increasing k which cannot be confirmed by our studies.

Since it may be argued that for different values of k a comparison for the above models is not completely fair (in addition to the difference of $F(a)$ described above there is the difference in the dimension of the Pareto sets), we consider $PS2_k$(l) for a fixed value of l but with different values of k. To be more precise, we consider $PS2_k(k)$ and $PS2_{k+1}(k)$. The reason is that in both cases, i.e., for the k-objective model $PS2_k(k)$ as well as for the $(k + 1)$-objective model $PS2_{k+1}(k)$, the Pareto set is given by $S(e_1, \ldots, e_k)$. That is, in this case at least GD_x can be assumed to be completely fair for a

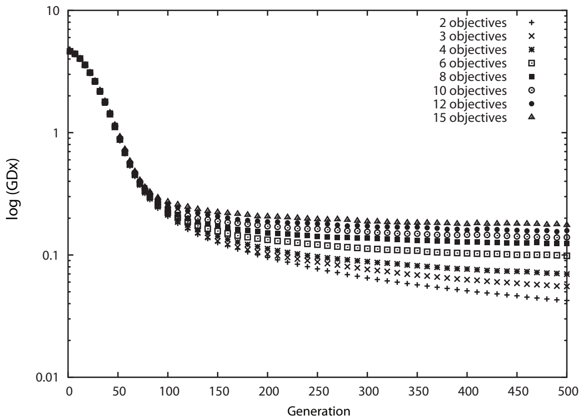

Fig. 2. Numerical results of NSGA-II for PS1 for k = 2, 3, 4, 6, 8, 10, 12, 15 objectives. The results are in parameter space [log(GD$_x$)] and averaged over 50 independent runs. Compare to Table 1.

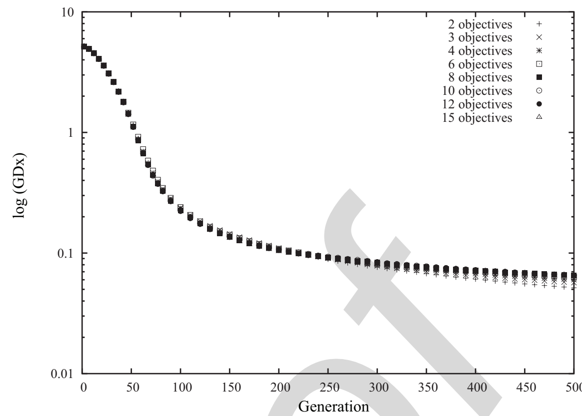

Fig. 4. Numerical results of NSGA-II for PS2$_k$(k) for k = 2, 3, 4, 6, 8, 10, 12, 15 objectives. The results are in parameter space [log(GD$_x$)] and averaged over 50 independent runs.

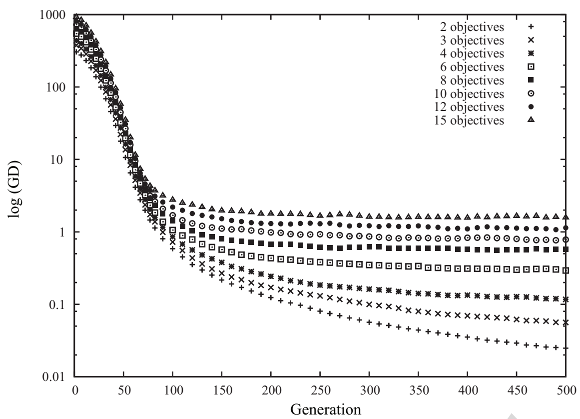

Fig. 3. Numerical results of NSGA-II for PS1 for k = 2, 3, 4, 6, 8, 10, 12, 15 objectives. The results are in objective space [log(GD)] and averaged over 50 independent runs. Compare to Table 2.

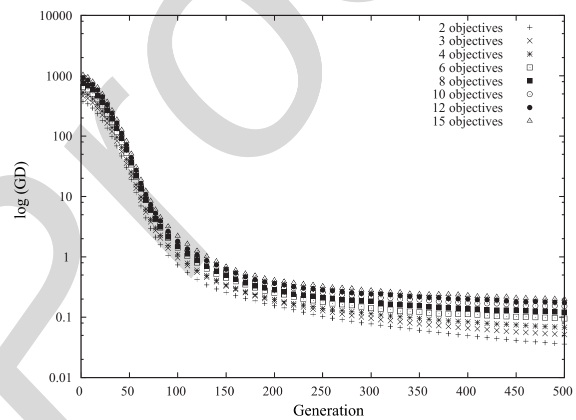

Fig. 5. Numerical results of NSGA-II for PS2$_k$(k) for k = 2, 3, 4, 6, 8, 10, 12, 15 objectives. The results are in objective space [log(GD)] and averaged over 50 independent runs.

comparison. Figs. 7 and 8 show such comparisons for values of k between 3 and 14, where we have chosen the same setting as in the previous study. Also here, small differences in the performances can be observed, but it is certainly not justified to talk about different orders of magnitude.

IV. Discussion and an Attempt to Explain Recent Advances

In the previous section, we have investigated a particular class of uni-modal MOPs with respect to the influence of the number of objectives on the hardness of the problem. Putting theoretical and empirical observations together we can conclude that by adding an objective to a given MOP the problem does per se not get harder by a significant amount, at least not on the (easy) class of models under consideration. However, such a scalability has been observed by many researchers on other, more complex, models. The question which now naturally arises is how this can be put together, i.e., if the observations made above can also be helpful for the design of algorithms for general many-objective models. In the following, we hazard to guess the sources of difficulties when dealing with many-objective problems, and try to explain recent advances in the field of evolutionary many-objective optimization in light of our discussion.

Based on the above considerations, three influential factors for the efficient numerical treatment of many-objective optimization problems with evolutionary algorithms regardless of the particular choice of the algorithm seem to be:

1) the pulling assumption as described in Section III-B1;
2) the probability to improve an individual;
3) the multi-modality of the MOP.

Problems 1) and 2) are to a certain extent in the hands of the algorithm designer, whereas problem 3) is given to him/her (or is possibly a modeling problem).

Much research has been done so far to improve the pulling property [i.e., problem 1)]. In case a population consists only of nondominated solutions and the generational operators produce further nondominated candidates the question arises which point to keep and which one to discard in order to converge toward the Pareto set. Since not all these nondominated solutions have the same distance to the solution set one can laxly say that *"some nondominated points are better than others"* [9]. The quest for those points has led so far to a variety of substitute distance assignments in NSGA-II

TABLE I

NUMERICAL RESULTS OF NSGA-II FOR PS1 FOR k = 3, 4, 6, 8, 10, 12, 15 OBJECTIVES

	Number of Generations					
k	50	100	200	300	400	500
2	8.57E−001	2.07E−001	9.54E−002	6.48E−002	5.10E−002	4.24E−002
3	8.74E−001	2.13E−001	1.02E−001	7.62E−002	6.28E−002	5.53E−002
4	8.85E−001	2.19E−001	1.12E−001	8.83E−002	7.67E−002	6.98E−002
6	8.79E−001	2.22E−001	1.33E−001	1.13E−001	1.03E−001	9.84E−002
8	8.78E−001	2.38E−001	1.52E−001	1.37E−001	1.28E−001	1.24E−001
10	8.94E−001	2.43E−001	1.72E−001	1.55E−001	1.46E−001	1.39E−001
12	9.18E−001	2.60E−001	1.87E−001	1.72E−001	1.60E−001	1.58E−001
15	9.27E−001	2.72E−001	2.06E−001	1.88E−001	1.79E−001	1.76E−001

The results are in parameter space (GD_x) and averaged over 50 independent runs (compare to Fig. 2).

TABLE II

NUMERICAL RESULTS OF NSGA-II FOR PS1 FOR k = 3, 4, 6, 8, 10, 12, 15 OBJECTIVES

	Number of Generations					
k	50	100	200	300	400	500
2	1.06E+001	5.86E−001	1.24E−001	5.65E−002	3.53E−002	2.48E−002
3	1.36E+001	7.24E−001	1.71E−001	9.96E−002	6.94E−002	5.61E−002
4	1.63E+001	8.56E−001	2.44E−001	1.62E−001	1.34E−001	1.17E−001
6	1.97E+001	1.06E+000	4.33E−001	3.49E−001	3.11E−001	2.93E−001
8	2.27E+001	1.41E+000	6.71E−001	6.22E−001	5.76E−001	5.75E−001
10	2.67E+001	1.69E+000	9.84E−001	8.64E−001	8.30E−001	7.82E−001
12	3.07E+001	2.20E+000	1.30E+000	1.22E+000	1.10E+000	1.14E+000
15	3.52E+001	5.38E+000	1.78E+000	1.61E+000	1.56E+000	1.58E+000

The results are in objective space (GD) and averaged over 50 independent runs (compare to Fig. 3).

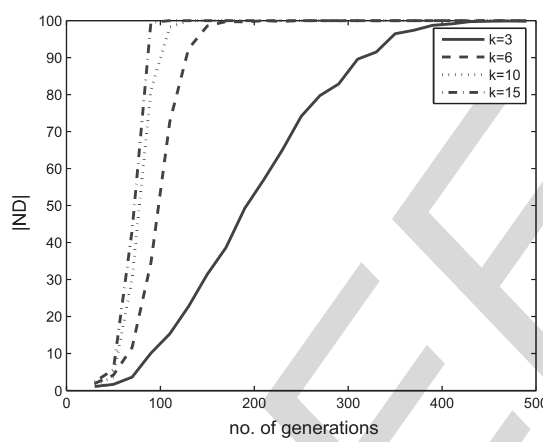

Fig. 6. Number of nondominated points ($|ND|$) during the run of NSGA-II for different values of k for the PS1 problems with population size 100 (averaged over 20 test runs).

(e.g., [2], [9], [35], [41], [50]). All these methods were able to outperform its base MOEA on scalable benchmark models (such as the DTLZ models). Though these results are all satisfying from the practical point of view, however, none of them ensures convergence toward the set of interest. It is known that in NSGA-II cycling (see [21]) or deterioration can occur which prevents that a predescribed "limit set" is reached resulting in a certain lack of efficiency, at least from the theoretical point of view [38]. Due to the dimensionality, the problem of defining a suitable limit set is getting more important with increasing value of k which would ease the evaluation of the newly developed strategies.

In multiobjective particle swarm optimization (MOPSO) algorithms, the pulling property is closely related to the choice of the guidance mechanism which has been addressed in [34] and [39] for many-objective problems.

To downsize problem 2), several remedies have been proposed so far which all lead to an augmentation of the descent cones of the related auxiliary models. One way to increase the improvement probability (while reducing the multi-modality of the problem) is to consider instead of the given k-objective problem a sequence of lower objective problems. For instance, the methods MSOPS [24] and RSO [25] are based on aggregation functions to find Pareto optimal solutions. Another approach is to use "space partitioning" [1], [2], i.e., partitioning the objective space into subspaces and performing one or several generations of the evolutionary search in each subspace. In both cases, the descent cone of the auxiliary model at a point x is typically larger than the original problem, and in the case of space partitioning the number of local minima is typically fewer (see the Appendix). The latter is not always true when using an aggregation function f_a since this depends on the choice of f_a as well as on the original model (see [31] for a counterexample).

For these approaches it holds that the speed of convergence gets improved, but, in turn, problems arise concerning the diversity maintenance. In particular, it may happen that not every Pareto point can be reached by the auxiliary problems which leads to a bias of the approaches. The potential drawbacks of aggregation functions are known (e.g., [11]), the reason for a potential bias when using space partitioning is because the union of the Pareto sets of all subproblems does typically not

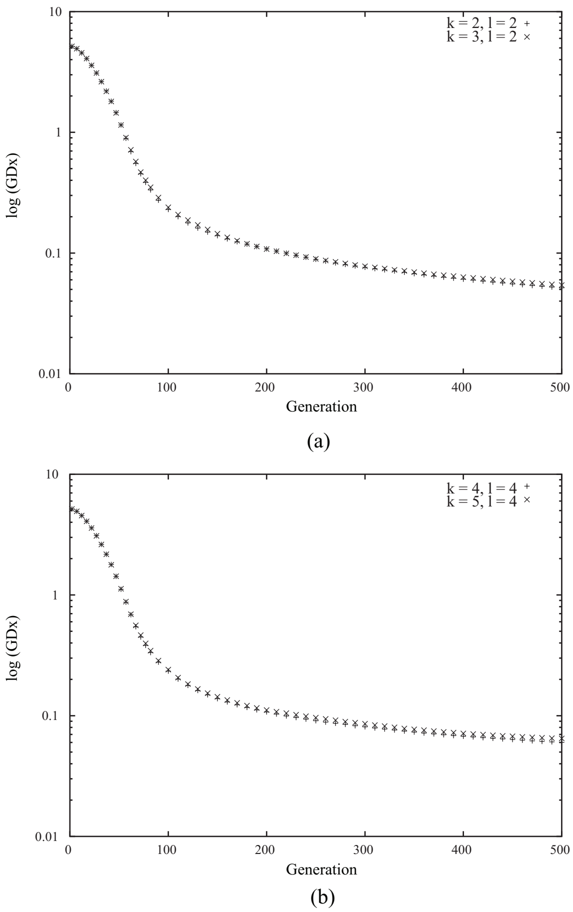

Fig. 7. Numerical results of NSGA-II for $PS2_k$(k) and $PS2_{k+1}$(k) for $k = 2$ and 4. The plots show the number of generations vs. $\log(GD_x)$. The results are averaged over 50 independent runs. (a) $PS2_2(2)$ and $PS2_3(2)$. (b) $PS2_4(4)$ and $PS2_5(4)$.

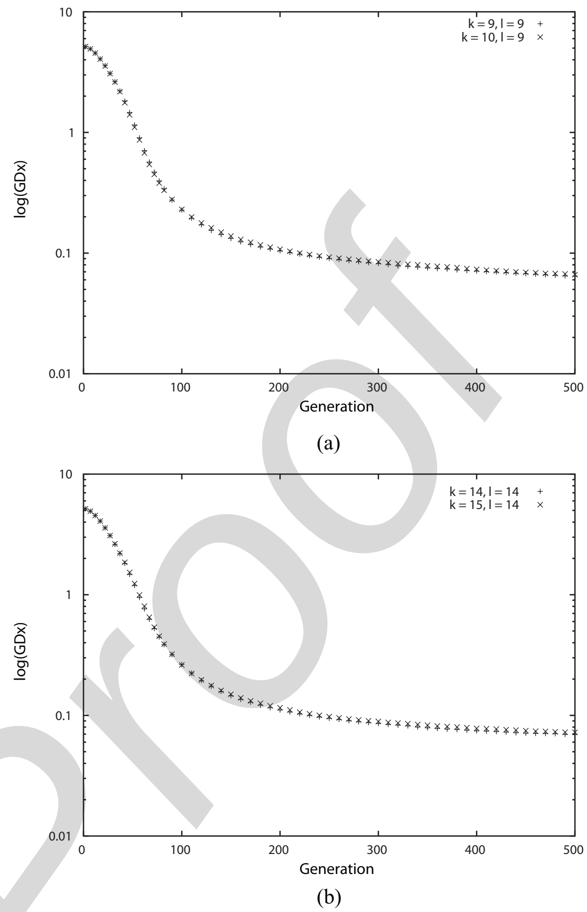

Fig. 8. Numerical results of NSGA-II for $PS2_k$(k) and $PS2_{k+1}$(k) for $k = 9$ and 14. The plots show the number of generations vs. $\log(GD_x)$. The results are averaged over 50 independent runs. (a) $PS2_9(9)$ and $PS2_{10}(9)$. (b) $PS2_{14}(14)$ and $PS2_{15}(14)$.

form the Pareto set of the "full" MOP. For instance, when choosing the PS1 problem with minimizers a_1, a_2, and a_3 (for the objectives f_1 to f_3, respectively) such that the volume of $S(a_1, a_2, a_3)$ is positive, then the union of the Pareto sets of all bi-objective subproblems (f_1, f_2), (f_1, f_3), and (f_2, f_3) is $S(a_1, a_2) \cup S(a_1, a_3) \cup S(a_2, a_3)$, i.e., is equal to the boundary of the "complete" Pareto set $S(a_1, a_2, a_3)$, but no interior point is included.

Another way to increase the improvement probability is to modify the Pareto dominance relation. Clearly, a larger dominance cone (defined in objective space) is related to a larger descent cone [defined in parameter space, see (24)] which in turn increases the probability to find a "better" solution as discussed above. The usage of such modified dominance cones within MOEAs can be found in [44], and in [17], [32], [33] fuzzifications of the Pareto dominance relation can be found which by its relaxation similarly influences the size of the dominance cones. Also for these methods, problems in diversity maintenance have been reported.

One aspect so far disregarded by researchers—but probably worth exploring—is the ability of memetic strategies to improve the performance of many-objective optimization problems. On the one hand, mathematical programming techniques (e.g., [4], [19]) allow—if gradient information is at hand—to compute a descent direction at every given non optimal point regardless of the size of the descent cone nor of the value of k, and hence it can be argued that the probability for improvement is one. On the other hand, the use of gradient information within a memetic strategy results in a certain additional cost [37] and in case the model is highly multi-modal [i.e., problem 3)] the effect of the local search on the overall performance is questionable.

In [53], it has been reported that ϵ-MOEA [13] copes well with many-objective optimization problems, even on highly multi-modal models. ϵ-MOEA is a steady state MOEA equipped with an archiving strategy which is based on the concept of ϵ-dominance and guarantees under certain assumptions convergence toward a finite size representation of the Pareto set [43], [38]. The good behavior of ϵ-MOEA with respect to k can partly be explained by our considerations: elements of the archive are only replaced by dominated solutions, i.e., a good solution will not be discarded due to any distance assignment, but only due to the existence of a better one. Hence, such solutions cannot be discarded by mistake which certainly helps to pull the population toward the Pareto set. Further, by the use of the archiving strategy proposed in [38] the descent cone is enlarged: the objective space gets divided into a grid of boxes, whose size can be adjusted by the size of ϵ. Every solution of the archive has to be located in a different box (i.e., every archive entry is associated with a box of the grid). The dominance relation is now enlarged since only nondominated boxes are allowed. Hence, ϵ-MOEA has mechanisms to cope with problems 1) and 2). Further good results can hence in principle be expected with related algorithms such as PAES [10] or PESA [30], or with any MOEA which is equipped with an archive which converges toward a finite size representation of the set of interest (e.g., [46]–[49]). The problem—at least when using archivers based on ϵ-dominance—is certainly the proper choice of ϵ: as discussed in [9], if the value of ϵ is too small, the archive sizes become intractable, and for large values of ϵ the limit archive set basically consists of a set of randomly selected (but not close by) points from the Pareto set.

In summary, it can be said that yet a variety of promising approaches exist in terms of their ability to converge toward the Pareto set P_Q. The distribution, however, is still an open problem. For this, a clear definition of the optimal distribution of the (few) individuals $a \in A$ is still missing but required to evaluate the finite size approximation of P_Q beyond convergence in the sense of $dist(A, P_Q)$ or $dist(F(A), F(P_Q))$.

V. Conclusion

In this paper, we have investigated the influence of the number k of objectives in a MOP on the hardness of the problem when solving it by evolution strategies. For this, we have utilized the descent cones which can be used to measure the probability to improve a solution by the generational operators. Though these considerations are of qualitative nature and can hardly be quantified, they help to a certain extent to understand the behavior of the population's evolution with respect to k. As an example, we have considered a class of uni-modal test functions and have investigated the resulting models qualitatively and empirically. Qualitative studies based on the descent cones led to the conclusion that, on the one hand, the addition of an objective makes the problem indeed harder, but, on the other hand, it can be argued that the difference is not significant, which is, later on, empirically validated. That is, it can be argued that the addition of an objective to a MOP does not make the problem per se harder.

In contrast to this, many researchers have so far observed a certain scalability in the hardness of the problem with respect to k, albeit for more complex models. Based on our considerations on the uni-modal models we have tried to identify the challenges which have to be mastered by evolution strategies for general models: the ability to keep "good" solutions in order to pull the population toward the set of interest, the probability to improve an individual, and the multi-modality of the MOP. This together with the qualitative discussions in Section III-B can be used to a certain extent to explain recent advances in the field of evolutionary many-objective optimization.

We hope that this new insight into the geometry of multiobjective optimization may help researchers in the field of evolutionary computation for further developments of efficient specialized algorithms, particularly when dealing with many-objective problems.

Appendix

The following little discussion shows that by adding objectives to a given MOP the set of local minima cannot get smaller but rather gets bigger which we argue by the set of KKT points (note that every local minimizer is a KKT point). Let a MOP be given consisting of the k objectives $(f_1, \ldots, f_k)$ (denote by MOP1). Further, let an extended model be given by the objectives $(f_1, \ldots, f_k, f_{k+1})$ [denote by (MOP2)] where the first k objectives are identical in MOP1 and MOP2. For simplicity, we assume that all objectives are defined on the same domain $Q \subset \mathbb{R}^n$. Define

$$KKT(i) := \{x \in Q \;:\; x \text{ is KKT point of MOPi}\} \quad i = 1, 2. \tag{30}$$

The following consideration shows that $KKT(1)$ is a subset of $KKT(2)$: let $x \in KKT(1)$, i.e., there exists a convex weight $\alpha \in \mathbb{R}^k$ (i.e., $\alpha_i \geq 0$ for all $i = 1, \ldots, k$ and $\sum_{i=1}^k \alpha_i = 1$) such that $\sum_{i=1}^k \alpha_i \nabla f_i(x) = 0$. Since $\tilde{\alpha} = (\alpha, 0) \in \mathbb{R}^{k+1}$ is also a convex weight and $\sum_{i=1}^{k+1} \tilde{\alpha}_i \nabla f_i(x) = 0$ it follows that x is also included in $KKT(2)$.

Further, it is for instance every substationary point of f_{k+1} (i.e., a point $x \in Q$ which satisfies $\nabla f_{k+1}(x) = 0$) also included in $KKT(2)$ [for this, choose $\alpha = (0, \ldots, 0, 1)$], and hence, $KKT(2)$ is typically a strict superset of $KKT(1)$. Furthermore, it can be shown that under certain additional assumptions the set $KKT(2)$ forms a k-dimensional object while $KKT(1)$ is $(k-1)$-dimensional [22].

References

[1] H. E. Aguirre and K. Tanaka, "Many-objective optimization by space partitioning and adaptive epsilon-ranking on MNK-landscapes," in *Proc. 4th Int. Conf. EMO*, 2009, pp. 407–422.

[2] H. E. Aguirre and K. Tanaka, "Space partitioning with adaptive epsilon-ranking and substitute distance assignments: A comparative study on many-objective MNK-landscapes," in *Proc. GECCO*, 2009, pp. 547–554.

[3] A. Auger, J. Bader, D. Brockhoff, and E. Zitzler, "Articulating user preferences in many-objective problems by sampling the weighted hypervolume," in *Proc. 11th Annu. Conf. GECCO*, 2009, pp. 555–562.

[4] P. A. N. Bosman and E. D. de Jong, "Exploiting gradient information in numerical multiobjective evolutionary optimization," in *Proc. GECCO*, vol. 1. Jun. 2005, pp. 755–762.

[5] D. Brockhoff, T. Friedrich, N. Hebbinghaus, and C. Klein, "On the effects of adding objectives to plateau functions," *IEEE Trans. Evol. Comput.*, vol. 13, no. 3, pp. 591–603, Jun. 2009.

[6] D. Brockhoff and E. Zitzler, "Are all objectives necessary? On dimensionality reduction in evolutionary multiobjective optimization," in *Proc. 9th Int. Conf. PPSN*, 2006, pp. 533–542.

[7] M. Brown and R. E. Smith, "Directed multiobjective optimization," *Int. J. Comput. Syst. Signals*, vol. 6, no. 1, pp. 3–17, 2005.

[8] C. A. Coello Coello, G. B. Lamont, and D. A. Van Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems*, 2nd ed. New York: Springer, Sep. 2007.

[9] D. Corne and J. D. Knowles, "Techniques for highly multiobjective optimization: Some nondominated points are better than others," in *Proc. 9th Annu. Conf. GECCO*, 2007, pp. 773–780.

[10] D. W. Corne, J. D. Knowles, and M. J. Oates, "The Pareto-envelope based selection algorithm for multiobjective optimization," in *Proc. 6th Int. Conf. PPSN*, 2000, pp. 869–878.
[11] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. Chichester, U.K.: Wiley, 2001.
[12] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," *Complex Syst.*, vol. 9, no. 2, pp. 115–148, 1995.
[13] K. Deb, M. Mohan, and S. Mishra, "Evaluating the ϵ-dominated based multiobjective evolutionary algorithm for a quick computation of Pareto-optimal solutions," *Evol. Comput.*, vol. 13, no. 4, pp. 501–525, 2005.
[14] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.
[15] K. Deb and D. K. Saxena, "Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multi-objective optimization problems," in *Proc. IEEE CEC*, 2006, pp. 3353–3360.
[16] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler, "Scalable test problems for evolutionary multiobjective optimization," in *Evolutionary Multiobjective Optimization. Theoretical Advances and Applications*, A. Abraham, L. Jain, and R. Goldberg, Eds. New York: Springer, 2005, pp. 105–145.
[17] M. Farina and P. Amato, "A fuzzy definition of 'optimality' for many-criteria optimization problems," *IEEE Trans. Syst., Man, Cybern.*, vol. 34, no. 3, pp. 315–326, May 2004.
[18] P. Fleming, R. C. Purshouse, and R. J. Lygoe, "Many-objective optimization: An engineering design perspective," in *Proc. 3rd Int. Conf. EMO*, LNCS 3410. Mar. 2005, pp. 14–32.
[19] J. Fliege and B. F. Svaiter, "Steepest descent methods for multicriteria optimization," *Math. Methods Oper. Res.*, vol. 51, no. 3, pp. 479–494, 2000.
[20] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multiobjective optimization," *Evol. Comput.*, vol. 3, no. 1, pp. 1–16, 1995.
[21] T. Hanne, "On the convergence of multiobjective evolutionary algorithms," *Eur. J. Oper. Res.*, vol. 117, no. 3, pp. 553–564, 1999.
[22] C. Hillermeier, *Nonlinear Multiobjective Optimization - A Generalized Homotopy Approach*. Boston, MA: Birkhäuser, 2001.
[23] J. Horn, "Multicriterion decision making," in *Handbook of Evolutionary Computation*. Oxford, U.K.: Oxford University Press, 1997, pp. 1–9.
[24] E. J. Hughes, "Multiple single objective Pareto sampling," in *Proc. IEEE CEC*, 2003, pp. 2678–2684.
[25] E. J. Hughes, "Evolutionary many-objective optimization: Many ones or one many?" in *Proc. IEEE CEC*, 2005, pp. 222–227.
[26] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization: A short review," in *Proc. IEEE CEC*, 2008, pp. 2419–2426.
[27] A. López Jaimes, C. A. Coello Coello, and J. E. Urías Barrientos, "Online objective reduction to deal with many-objective problems," in *Proc. 5th Int. Conf. EMO*, LNCS 5467. 2009, pp. 423–437.
[28] W. E. Karush, "Minima of functions of several variables with inequalities as side conditions," Ph.D. dissertation, Univ. Chicago, Chicago, IL, 1939.
AQ:3
[29] J. Knowles and D. Corne, "Quantifying the effects of objective space dimension in evolutionary multiobjective optimization," in *Proc. 4th Int. Conf. EMO*, LNCS 4403. Mar. 2007, pp. 757–771.
[30] J. D. Knowles and D. W. Corne, "Approximating the nondominated front using Pareto archived evolution strategy," *Evol. Comput.*, vol. 8, no. 2, pp. 149–172, 2000.
[31] J. D. Knowles, R. A. Watson, and D. W. Corne, "Reducing local optima in single-objective problems by multi-objectivization," in *Proc. 1st Int. Conf. EMO*, 2001, pp. 269–283.
[32] M. Köppen and R. Vicente-Garcia, "A fuzzy scheme for the ranking of multivariate data and its application," in *Proc. IEEE Annu. Meeting Fuzzy Information*, 2004, pp. 140–145.
[33] M. Köppen, R. Vicente-Garcia, and B. Nickolay, "Fuzzy Pareto-dominance and its application in evolutionary multiobjective optimization," in *Proc. 3rd Int. Conf. EMO*, LNCS 3410. 2005, pp. 399–412.
[34] M. Köppen and K. Yoshida, "Many-objective particle swarm optimization by gradual leader selection," in *Proc. 8th ICANNGA*, 2007, pp. 323–331.
[35] M. Köppen and K. Yoshida, "Substitute distance assignment in NSGA-II for handling many-objective optimization problems," in *Proc. 4th Int. Conf. EMO*, LNCS 4403. 2007, pp. 727–741.
[36] H. Kuhn and A. Tucker, "Nonlinear programming," in *Proc. 2nd Berkeley Symp. Math. Stat. Probab.*, 1951, pp. 481–492.
[37] A. Lara, G. Sanchez, C. A. Coello Coello, and O. Schütze, "HCS: A new local search strategy for memetic multiobjective evolutionary algorithms," *IEEE Trans. Evol. Comput.*, vol. 14, no. 1, pp. 112–132, Feb. 2010.
[38] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler, "Combining convergence and diversity in evolutionary multiobjective optimization," *Evol. Comput.*, vol. 10, no. 3, pp. 263–282, 2002.
[39] S. Mostaghim and H. Schmeck, "Distance based ranking in many-objective particle swarm optimization," in *Proc. 10th Int. Conf. PPSN*, 2008, pp. 753–762.
[40] V. Pareto, *Manual of Political Economy*. New York: MacMillan, 1971.
[41] K. Praditwong and X. Yao, "How well do multiobjective evolutionary algorithms scale to large problems?" in *Proc. IEEE CEC*, 2007, pp. 3959–3966.
[42] R. C. Purshouse and P. J. Fleming, "Evolutionary many-objective optimization: An exploratory analysis," in *Proc. CEC*, 2003, pp. 2066–2073.
[43] G. Rudolph and A. Agapie, "On a multiobjective evolutionary algorithm and its convergence to the Pareto set," in *Proc. CEC*, 2000, pp. 1010–1016.
[44] H. Sato, H. E. Aguirre, and K. Tanaka, "Controlling dominance area of solutions and its impact on the performance of MOEAs," in *Proc. 4th Int. Conf. EMO*, LNCS 4403. 2007, pp. 5–20.
[45] D. J. Schaffer, "Multiple objective optimization with vector evaluated genetic algorithms," in *Proc. 1st Int. Conf. Genet. Algorithms*, 1985, pp. 93–100.
[46] O. Schütze, C. A. Coello Coello, E. Tantar, and E.-G. Talbi, "Computing finite size representations of the set of approximate solutions of an MOP with stochastic search algorithms," in *Proc. 10th Annu. Conf. GECCO*, 2008, pp. 713–720.
[47] O. Schütze, M. Laumanns, C. A. Coello Coello, M. Dellnitz, and E.-G. Talbi, "Convergence of stochastic search algorithms to finite size Pareto set approximations," *J. Global Optimization*, vol. 41, no. 4, pp. 559–577, 2008.
[48] O. Schütze, M. Laumanns, E. Tantar, C. A. Coello Coello, and E.-G. Talbi. "Convergence of stochastic search algorithms to gap-free Pareto front approximations," in *Proc. GECCO*, 2007, pp. 892–901.
[49] O. Schütze, M. Laumanns, E. Tantar, C. A. Coello Coello, and E.-G. Talbi, "Computing gap free Pareto front approximations with stochastic search algorithms," *Evol. Comput.*, vol. 18, no. 1, pp. 65–96, 2010.
[50] H. K. Singh, A. Isaacs, T. Ray, and W. Smith, "A study on the performance of substitute distance based approaches for evolutionary many objective optimization," in *Proc. 7th Int. Conf. SEAL*, 2008, pp. 401–410.
[51] O. Teytaud, "On the hardness of offline multiobjective optimization," *Evol. Comput.*, vol. 15, no. 4, pp. 475–491, 2007.
[52] D. A. Van Veldhuizen, "Multiobjective evolutionary algorithms: Classifications, analyses, and new innovations," Ph.D. dissertation, Dept. Electr. Comput. Eng., Graduate School Eng., Air Force Instit. Technol., Wright-Patterson AFB, OH, May 1999.
[53] T. Wagner, N. Beume, and B. Naujoks, "Pareto-, aggregation-, and indicator-based methods in many-objective optimization," in *Proc. 3rd Int. Conf. EMO*, LNCS 4403. 2007, pp. 742–756.
[54] P. Winkler, "Random orders," *Order*, vol. 1, no. 4, pp. 317–331, Dec. 1985.
[55] E. Zitzler, "Evolutionary algorithms for multiobjective optimization: Methods and applications," Ph.D. dissertation, ETH Zurich, Zurich, Switzerland, 1999.
[56] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," in *Proc. Evol. Methods Des., Optimization Control Applicat. Ind. Problems (EUROGEN)*, 2002, pp. 95–100.

Oliver Schütze received the Diploma degree in mathematics from the University of Bayreuth, Bayreuth, Germany, and the Ph.D. degree in natural sciences from the University of Paderborn, Paderborn, Germany, in 1999 and 2004, respectively.

He is currently a Professor (CINVESTAV-3A Researcher) with the Department of Computer Science, CINVESTAV-IPN, Mexico City, Mexico.

Adriana Lara received the B.S. degree in physics and mathematics from the National Polytechnic Institute of Mexico, Mexico City, Mexico, in 2001, and the M.S. degree in electrical and computer engineering from CINVESTAV-IPN, Mexico City, in 2003, where she is currently working toward the Ph.D. degree in computer sciences from the Department of Computer Science.

She was an Associate Professor with the National Polytechnic Institute of Mexico, Mexico City. Her current research interests include multiobjective optimization, gradient-based techniques, and hybrid algorithms.

Carlos A. Coello Coello (SM'XX) received the B.S. degree in civil engineering from the Universidad Autónoma de Chiapas, Chiapas, Mexico, and the M.S. and Ph.D. degrees in computer science from Tulane University, New Orleans, LA, in 1991, 1993, and 1996, respectively. AQ:4

He is currently a Professor (CINVESTAV-3E Researcher) with the Department of Computer Science, CINVESTAV-IPN, Mexico City, Mexico. He has authored and co-authored over 250 technical papers and several book chapters. He has also co-authored the book *Evolutionary Algorithms for Solving Multi-Objective Problems* (Berlin, Germany: Springer, 2007, 2nd ed.). His major research interests include evolutionary multiobjective optimization and constraint-handling techniques for evolutionary algorithms.

Dr. Coello Coello is currently an Associate Editor of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION and serves on the editorial board of 14 other international journals. He also chairs the IEEE Computationally Intelligent Society Task Force on Multiobjective Evolutionary Algorithms. He received the 2007 National Research Award from the Mexican Academy of Sciences in the area of Exact Sciences. He is a member of the ACM, Sigma Xi, and the Mexican Academy of Sciences.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

AQ:1= Please provide index terms for this paper.
AQ:2= Please provide the expanded form of "MOEA."
AQ:3= Please provide department name in Refs. [28], [55].
AQ:4= Please provide membership year of "Coello Coello."

END OF ALL QUERIES