Proximity measures based on KKT points for constrained multi-objective optimization

Gabriele Eichfelder[*], Leo Warnow[*]

August 19, 2020

Abstract

An important aspect of optimization algorithms, for instance evolutionary algorithms, are termination criteria that measure the proximity of the found solution to the optimal solution set. A frequently used approach is the numerical verification of necessary optimality conditions such as the Karush-Kuhn-Tucker (KKT) conditions. In this paper, we present a proximity measure which characterizes the violation of the KKT conditions. It can be computed easily and is continuous in every efficient solution. Hence, it can be used as an indicator for the proximity of a certain point to the set of efficient (Edgeworth-Pareto-minimal) solutions and is well suited for algorithmic use due to its continuity properties. This is especially useful within evolutionary algorithms for candidate selection and termination, which we also illustrate numerically for some test problems.

Key Words: Multiobjective Optimization, KKT Approximation, Proximity Measure

Mathematics subject classifications (MSC 2010): 90C26, 90C29, 90C46, 90C59

1 Introduction

In applications, one often has to deal with not only one but multiple objectives at the same time. This leads to multi-objective optimization problems. Then it is the aim to find globally optimal solutions, called efficient solutions, for such optimization problems using optimization algorithms, e.g., evolutionary algorithms as proposed in [6].
Especially evolutionary algorithms are often considered to be able to overcome regions with only locally efficient solutions and to generate points close to the global Pareto front, i.e., close to the image set of all globally efficient solutions. An important aspect in such algorithms is then the question when the algorithm can finally be stopped as one is sufficiently close to the Pareto front. For that decision, for instance in [7] and [10], proximity measures for termination have been examined by Deb, Dutta and co-authors. We follow this line of research for multi-objective problems, and we do this without the detour of scalarization.

---

[*]Institute for Mathematics, Technische Universität Ilmenau, Po 10 05 65, D-98684 Ilmenau, Germany, {gabriele.eichfelder,leo.warnow}@tu-ilmenau.de

A necessary condition for being efficient is to satisfy necessary optimality conditions, at least under certain constraint qualifications. This can also be used, based on the results of this paper, to evaluate the proximity of the generated points to the Pareto front: a necessary condition for being close to the Pareto front is that certain necessary optimality conditions are satisfied at least approximately.

In this paper we present two proximity measures which characterize the approximate fulfillment of the so-called Karush-Kuhn-Tucker (KKT) conditions. They can be used for candidate selection and, what is more, as a termination criterion for evolutionary algorithms, similar as proposed in [7] and [10]. In particular, proximity measures can be used in addition to already implemented techniques within evolutionary algorithms. We will show that the presented proximity measures are continuous in every efficient solution. Thereby we only assume the objective and constraint functions to be continuously differentiable and that some constraint qualifications hold. The continuity implies that when the algorithm produces points which come closer and closer to the Pareto front then also the values of the measure decrease continuously to zero.

Of course, proximity measures can also be used for deterministic algorithms. Within iterative approaches, an abort has to be done after a finite number of iterations. Hence, an approach to check whether the solution found by the algorithm is actually an efficient solution or at least an approximately efficient solution is needed in this case as well.

Finding exact KKT points can be a hard task, in particular when using computer algorithms. A common approach to handle this problem is to relax the KKT conditions and we will also do this within this paper. Within the last decade, several concepts for such a relaxation have been presented. An often followed approach is to provide a sequence of points that satisfy some relaxed KKT conditions and to show convergence of this sequence towards a KKT point.

Such a relaxation are the Approximate-KKT conditions, which were presented in 2011 in [3] for single-objective optimization problems. Those are satisfied for a feasible point if there exists a sequence of Lagrange multipliers and not necessarily feasible points such that the KKT error decreases to 0 in the limit of this sequence, which is also called AKKT sequence. This concept was extended to multi-objective optimization problems in [14] and [12].

The idea to use a relaxed version of KKT points to define proximity measures (also called error measures) appears more recently in [10] for single-objective optimization problems. There, the relaxation of KKT points are so-called ε-KKT and modified ε-KKT points. Their definition does no longer rely on a sequence of points but only on a single feasible point. The authors propose to use a proximity measure which is based on this relaxation and hence, based on the KKT error. A possibility to use this approach also for multi-objective optimization is to first scalarize the vector-valued optimization problem to a single-objective optimization problem, which was discussed in [7], [2], and [1]. Within this paper, we avoid the detour of a scalarization. The concept of (modified) ε-KKT points has also been extended to multi-objective optimization problems without using scalarization in [9] and [18], but no proximity measure was derived so far.

In all these papers the authors only show that an AKKT sequence or a convergent sequence of modified ε-KKT points with $\varepsilon \to 0$ indeed converges towards a KKT point. However, in practice, for instance within an evolutionary algorithm, one generates a sequence for which one does not know whether it consists of modified ε-KKT points with $\varepsilon \to 0$ or is an AKKT sequence. Still, one wants to know if the points of the sequence come close to a KKT point. With the results of this paper we give a necessary

condition for that. Thus, if the proximity measure is too large, the points of the generated sequence are for sure not close to an efficient solution and the algorithm should not be stopped.

Thereby, it is important that the proximity measure is continuous in the efficient solutions. Otherwise, the value of the measure could be large for points being arbitrarily close to an efficient solution and could drop to zero only in the efficient solution itself. So while in the literature for AKKT points and (modified) ε-KKT points it is shown that a sequence with certain properties converges to a KKT point, we will show that a sequence that converges to an efficient solution (which is a KKT point under constraint qualifications) has certain properties. Thus, for our new proximity measures in this paper we will show that they converge (and decrease to zero) for any such sequence.

As a consequence, the value of the presented proximity measures can be used as an indicator for non-efficiency as a point with positive value for the proximity measure is definitely not an efficient solution. Hence, for evolutionary algorithms, a small value of the proximity measure could be added as a criterion within candidate selection or as an additional termination criterion. In other words, we will show that a small value of the proximity measure is a necessary condition for being close to an efficient solution.

For the remaining part of this paper we start in Section 2 with some notations and definitions as well as the problem formulation and we recall basic results to necessary optimality conditions. In Section 3, we briefly discuss the concept of modified ε-KKT points from [10] and [18]. Then, we introduce our new proximity measures for multi-objective optimization problems. Finally, in Section 4, some numerical results for the proposed proximity measure for candidate selection are presented.

2   Notations and Basic Definitions

Within this paper, for a positive integer $n \in \mathbb{N}$ we use the notation $[n] := \{1, \ldots, n\}$. For a differentiable function $f \colon \mathbb{R}^n \to \mathbb{R}^m$ the Jacobian matrix of f at $x \in \mathbb{R}^n$ is $Df(x)$. Now let $x, \tilde{x} \in \mathbb{R}^n$. Then $\leq$ and $<$ are meant component-wise, i.e.

$$x \leq \tilde{x} \Leftrightarrow x_i \leq \tilde{x}_i \text{ for all } i \in [n],$$
$$x < \tilde{x} \Leftrightarrow x_i < \tilde{x}_i \text{ for all } i \in [n].$$

We focus on multi-objective optimization problems with inequality constraints. We denote by $f_i \colon \mathbb{R}^n \to \mathbb{R}$, $i \in [m]$ the objective functions and by $g_j \colon \mathbb{R}^n \to \mathbb{R}$, $j \in [p]$ the constraint functions. We also write $f = (f_1, \ldots, f_m)$ and $g = (g_1, \ldots, g_p)$. The multi-objective optimization problem of this paper is then defined by

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g(x) \leq 0. \tag{CMOP}$$

We denote the feasible set by $S = \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ and assume that it is a nonempty set. Moreover, for $x \in \mathbb{R}^n$ the active index set is $I(x) := \{j \in [p] \mid g_j(x) = 0\}$. All functions f_i, g_j for $i \in [m], j \in [p]$ are assumed to be continuously differentiable. Optimality for (CMOP) is defined as follows.

Definition 2.1 *A point $\bar{x} \in S$ is called an efficient or an Edgeworth-Pareto-minimal solution for* (CMOP) *if there exists no $x \in S$ with*

$$f_i(x) \leq f_i(\bar{x}) \text{ for all } i \in [m],$$
$$f_j(x) < f_j(\bar{x}) \text{ for at least one } j \in [m].$$

A point $\bar{x} \in S$ is called a weakly efficient or a weakly Edgeworth-Pareto-minimal solution for (CMOP) *if there exists no $x \in S$ with*

$$f_i(x) < f_i(\bar{x}) \text{ for all } i \in [m].$$

Every efficient solution $\bar{x} \in S$ is also weakly efficient. If all objective functions f_i, $i \in [m]$ are strictly convex and the set S is convex, then every weakly efficient solution $\bar{x} \in S$ is also an efficient solution.

In unconstrained single-objective optimization, i.e., for $S = \mathbb{R}^n$ and $m = 1$, a well known necessary optimality condition is $\nabla f(\bar{x}) = 0$. For necessary optimality conditions in constrained optimization, i.e., for the KKT conditions, it is known that a constraint qualification has to hold for an optimal solution $\bar{x}$ to guarantee that the necessary optimality conditions are satisfied. Several constraint qualifications are used in the literature. We use here the Abadie Constraint Qualification (Abadie CQ) as it is quite general and as it is implied by other well-known constraint qualifications as the LICQ or MFCQ as explained below.

Thus, we shortly recall the definition of the contingent cone as well as the linearized contingent cone to the set S at some point $\bar{x} \in \mathrm{cl}(S)$. So let $S \subseteq \mathbb{R}^n$ be a nonempty set as defined above and $\bar{x} \in \mathrm{cl}(S)$. Then the contingent cone at S in $\bar{x}$ is given as

$$T(S, \bar{x}) := \left\{ d \in \mathbb{R}^n \;\middle|\; \exists (x^i)_{\mathbb{N}} \subseteq S, (\lambda_i)_{\mathbb{N}} \subseteq \mathrm{int}(\mathbb{R}_+) \colon \bar{x} = \lim_{i \to \infty} x^i, \; d = \lim_{i \to \infty} \lambda_i(x^i - \bar{x}) \right\}.$$

Moreover, the linearized (contingent) cone is given by

$$T^{\mathrm{lin}}(S, \bar{x}) := \left\{ d \in \mathbb{R}^n \;\middle|\; \nabla g_j(\bar{x})^\top d \leq 0 \; \forall j \in I(\bar{x}) \right\}.$$

It always holds $T(S, \bar{x}) \subseteq T^{\mathrm{lin}}(S, \bar{x})$. We say that the Abadie CQ holds for some $\bar{x} \in S$ if

$$T(S, \bar{x}) = T^{\mathrm{lin}}(S, \bar{x}).$$

In case of affine linear constraint functions $g_j, j \in [p]$ the Abadie CQ is satisfied for all $x \in S$. Often stronger constraint qualifications than the Abadie CQ are used as those are easier to verify. One of them is the Mangasarian-Fromovitz Constraint Qualification (MFCQ). We say that the MFCQ is satisfied for some $\bar{x} \in S$ if there exists a direction $d \in \mathbb{R}^n$ such that $\nabla g_j(\bar{x})^\top d < 0$ for all $j \in I(\bar{x})$. Another well known constraint qualification is the Linear Independence Constraint Qualification (LICQ) which is satisfied for some $\bar{x} \in S$ if for all $j \in I(\bar{x})$ the gradients $\nabla g_j(\bar{x})$ are linearly independent. For these constraint qualifications it holds

$$\text{LICQ} \Rightarrow \text{MFCQ} \Rightarrow \text{Abadie CQ}.$$

Finally, if all constraint functions $g_j, j \in [p]$ are convex, another useful constraint qualification is Slater's Constraint Qualification (Slater's CQ). It is satisfied if there exists some $x^* \in S$ such that $g(x^*) < 0$. If Slater's CQ is satisfied, then the Abadie CQ holds for all feasible points $\bar{x} \in S$. In this paper, in general neither the objective functions $f_i, i \in [m]$ nor the constraint functions $g_j, j \in [p]$ are assumed to be convex.

For $x \in \mathbb{R}^n, \eta \in \mathbb{R}^m$, and $\lambda \in \mathbb{R}^p$ the following conditions

$$\sum_{i=1}^m \eta_i \nabla f_i(x) + \sum_{j=1}^p \lambda_j \nabla g_j(x) = 0, \qquad \text{(KKT1)}$$

$$g(x) \leq 0, \qquad \text{(KKT2)}$$

$$\lambda \geq 0, \qquad \text{(KKT3)}$$

$$\lambda^\top g(x) = 0, \qquad \text{(KKT4)}$$

$$\eta \geq 0, \qquad \text{(KKT5)}$$

$$\sum_{i=1}^m \eta_i = 1 \qquad \text{(KKT6)}$$

are called Karush-Kuhn-Tucker (KKT) conditions. If a tuple (x, η, λ) satisfies those six conditions, it is called a KKT point for (CMOP) and η and λ are called multipliers. Under the Abadie CQ (or any other of the stronger constraint qualifications as given above) we get the following necessary optimality condition for (CMOP), see [16, Satz 2.35].

Theorem 2.2 *Let $\bar{x} \in S$ be a weakly efficient solution for* (CMOP) *and let the Abadie CQ hold at $\bar{x}$. Then there exist multipliers $\bar{\eta} \in \mathbb{R}_+^m$ and $\bar{\lambda} \in \mathbb{R}_+^p$ such that $(\bar{x}, \bar{\eta}, \bar{\lambda})$ is a KKT point.*

This result can also be found in the literature by using the so-called Cottle Constraint Qualification in [19, Corollary 3.2.10] and by using the Kuhn-Tucker Constraint Qualification in [11, Theorem 3.21] and [21, Theorem 3.5.3].

Note that no convexity is required for the necessary optimality condition given in Theorem 2.2. One obtains similar necessary optimality conditions by making use of a weighted sum scalarization of (CMOP). However, then convexity is needed: in case of convex objective and constraint functions, for any weakly efficient solution $\bar{x} \in S$ there exist weights $\eta_i \in \mathbb{R}_+$, $i \in [m]$, which also satisfy (KKT6), such that $\bar{x}$ also minimizes

$$\min_{x \in S} \sum_{i=1}^m \eta_i f_i(x). \qquad (2.1)$$

The necessary optimality conditions to (2.1) are similar to the KKT conditions for (CMOP) above, but they only hold in case of convexity. We illustrate the difference with the next example.

Example 2.3 *The point $\bar{x} := (1/\sqrt{2}, 1/\sqrt{2})^\top$ is an efficient solution for the non-convex multi-objective optimization problem*

$$\min\{(x_1, x_2)^\top \mid 1 - (x_1^2 + x_2^2) \leq 0, \ -x_1 \leq 0, \ -x_2 \leq 0\}.$$

There exist no weights $w_1, w_2 \geq 0$, $w_1 + w_2 = 1$, such that $\bar{x}$ minimizes the weighted-sum scalarization of this multi-objective optimization problem, i.e., the problem

$$\min\{w_1 x_1 + w_2 x_2 \mid 1 - (x_1^2 + x_2^2) \leq 0, \ -x_1 \leq 0, \ -x_2 \leq 0\}.$$

However, there exist multipliers $\bar{\eta} \in \mathbb{R}_+^2$ and $\bar{\lambda} \in \mathbb{R}_+^3$ such that $(\bar{x}, \bar{\eta}, \bar{\lambda})$ is a KKT point: For

$$\bar{\eta}_1 = \bar{\eta}_2 = \frac{1}{2}, \qquad \bar{\lambda}_1 = \frac{1}{2\sqrt{2}}, \qquad \bar{\lambda}_2 = \bar{\lambda}_3 = 0$$

the KKT conditions for (CMOP) *are satisfied. This illustrates that for the KKT conditions for* (CMOP) *no convexity assumption is required.*

In the special case of unconstrained multi-objective optimization problems the KKT conditions reduce to

$$\sum_{i=1}^{m} \eta_i \nabla f_i(x) = 0, \tag{KKT1'}$$

$$\eta \geq 0, \tag{KKT5}$$

$$\sum_{i=1}^{m} \eta_i = 1. \tag{KKT6}$$

In this particular setting, $x \in \mathbb{R}^n$ is also called Pareto critical if there exists $\eta \in \mathbb{R}^m$ such that (x, η) satisfies (KKT1'), (KKT5), and (KKT6), see [13].

Only for giving a sufficient optimality condition for (CMOP) we need assumptions on the convexity of the problem. Recall that a continuously differentiable function $f_i \colon \mathbb{R}^n \to \mathbb{R}$ is called pseudoconvex if it holds for all $x, y \in \mathbb{R}^n$

$$\nabla f_i(x)^\top (y - x) \geq 0 \;\Rightarrow\; f_i(y) \geq f_i(x).$$

It is called quasiconvex if it holds for all $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$

$$f_i(\lambda x + (1 - \lambda)y) \leq \max\{f_i(x), f_i(y)\}.$$

Every continuously differentiable convex function is also pseudoconvex and quasiconvex. Using these convexity concepts, we obtain a sufficient optimality condition which can also be found in [17, Corollary 7.24].

Lemma 2.4 *Let $\bar{x} \in S$, f_i be pseudoconvex for all $i \in [m]$ and g_j be quasiconvex for all $j \in I(\bar{x})$. If there exist multipliers $\bar{\eta} \in \mathbb{R}_+^m$ and $\bar{\lambda} \in \mathbb{R}_+^p$ such that $(\bar{x}, \bar{\eta}, \bar{\lambda})$ is a KKT point, then $\bar{x}$ is weakly efficient for (CMOP). In case the functions f_i, $i \in [m]$ are even strictly convex, this is sufficient for $\bar{x}$ to be efficient for (CMOP).*

Again, we want to point out that for the remaining part of this paper there are no convexity assumptions concerning the functions $f_i, i \in [m]$ and $g_j, j \in [p]$ unless otherwise stated. Finally, we recall a basic result on the convergence of sequences which we need in the proof of our main result later on.

Lemma 2.5 *Let a sequence $(x^i)_\mathbb{N} \subseteq \mathbb{R}^n$ and a point $\bar{x} \in \mathbb{R}^n$ be given. If for all subsequences $(x^{i_k})_{k \in \mathbb{N}}$ of $(x^i)_\mathbb{N}$ there exists a subsubsequence $(x^{i_{k_l}})_{l \in \mathbb{N}}$ such that*

$$\lim_{l \to \infty} x^{i_{k_l}} = \bar{x},$$

then the original sequence $(x^i)_\mathbb{N}$ converges to $\bar{x}$ as well, i.e. $\lim_{i \to \infty} x^i = \bar{x}$.

Proof. Assume that $(x^i)_\mathbb{N}$ does not converge to $\bar{x}$. Then there exists an $\varepsilon > 0$ such that for all $k \in \mathbb{N}$ there exists an $i_k > k$ with $\|x^{i_k} - \bar{x}\| \geq \varepsilon$. This implies that the subsequence $(x^{i_k})_{k \in \mathbb{N}}$ of $(x^i)_\mathbb{N}$ cannot have any subsubsequence $(x^{i_{k_l}})_{l \in \mathbb{N}}$ with

$$\lim_{l \to \infty} x^{i_{k_l}} = \bar{x}.$$

Hence, the assumption does not hold and $(x^i)_\mathbb{N}$ converges to $\bar{x}$. $\qquad\square$

3 Proximity Measures

As motivated in the introduction, we want to provide a proximity measure which characterizes the fulfillment of a necessary optimality condition. Moreover, it should be easy to compute and provide good numerical properties as we want to use it with optimization algorithms, e.g., evolutionary algorithms. First, we present the desired properties of such a proximity measure in the next definition.

Definition 3.1 *A function* $\omega\colon \mathbb{R}^n \to \mathbb{R}$ *is called a proximity measure if for every efficient solution* $\bar{x} \in S$ *of* (CMOP) *in which the Abadie CQ holds and every sequence* $(x^i)_{\mathbb{N}} \subseteq \mathbb{R}^n$ *with*

$$\lim_{i \to \infty} x^i = \bar{x}$$

the following three properties are satisfied:

(PM1) $\omega(x) \geq 0$ *for all* $x \in \mathbb{R}^n$,

(PM2) $\omega(\bar{x}) = 0$,

(PM3) $\lim_{i \to \infty} \omega(x^i) = \omega(\bar{x})$.

While properties (PM1) and (PM2) are quite easy to realize, property (PM3) is more challenging. It ensures that a proximity measure ω is continuous at least in every efficient solution $\bar{x} \in S$ of (CMOP) in which the Abadie CQ holds. Hence, we can expect a proximity measure to have small values locally around every efficient solution. This makes it suitable for applications, e.g., for termination or candidate selection in evolutionary algorithms. Convergence statements also appear in [10] and [18]. However, these statements only hold for certain sequences $(x^i)_{\mathbb{N}}$ or rely on stronger assumptions, e.g., convexity of the objective and constraint functions. Whenever a function ω is a proximity measure in the sense of Definition 3.1, then property (PM3) holds for any sequence $(x^i)_{\mathbb{N}}$ (converging towards an efficient solution $\bar{x}$ in which the Abadie CQ holds).

Next to the three properties from Definition 3.1, we also aim to provide a non-trivial proximity measure, i.e., not to choose $\omega \equiv 0$, as such a proximity measure would provide no additional information for optimization algorithms. In the introduction of [7], the authors presented an example for a naively defined candidate $\hat{\omega}$ for a proximity measure based on the KKT conditions to motivate their further examinations. For all $x \in S$ this is

$$\hat{\omega}(x) := \min\left\{ \left\| \sum_{i=1}^m \eta_i \nabla f_i(x) + \sum_{j \in I(x)} \lambda_j \nabla g_j(x) \right\| \in \mathbb{R}_+ \;\middle|\; \eta \in \mathbb{R}_+^m, \lambda \in \mathbb{R}_+^p, \sum_{i=1}^m \eta_i = 1 \right\}.$$

We shortly recall the numerical example from [7] which illustrates that this function does not satisfy property (PM3) in general. This also shows that more effort is needed to find a suitable proximity measure, i.e., a function that is not only based on KKT conditions but also satisfies the properties from Definition 3.1.

Example 3.2 *We consider the constrained multi-objective optimization problem*

$$\min_{x \in \mathbb{R}^2} f(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} := \begin{pmatrix} x_1 \\ \dfrac{1 + x_2}{1 - (x_1 - 0.5)^2} \end{pmatrix} \quad s.t. \quad 0 \leq x_1, x_2 \leq 1. \quad \text{(P1)}$$

7

The set of efficient solutions for this problem is

$$\mathcal{E} := \left\{ x \in \mathbb{R}^2 \,\middle|\, x_1 \in [0, 0.5], x_2 = 0 \right\}. \tag{3.1}$$

A sequence of feasible points $(x^\alpha) \subseteq S$ is given by

$$x^\alpha = \begin{pmatrix} 0.2 \\ \alpha \end{pmatrix} \text{ for } \alpha \in [0, 1].$$

Using (3.1), we derive that the only efficient solution in this sequence is $x^0 = (0.2, 0)^\top$, in which also the Abadie CQ is satisfied. With the Euclidean norm in the definition of $\hat{\omega}$ we obtain

$$\hat{\omega}(x^\alpha) \geq 0.2 \text{ for all } \alpha \in (0, 1). \tag{3.2}$$

Hence, there exists a sequence $(x^\alpha) \subseteq S$ and an efficient solution x^0 such that $\lim\limits_{\alpha \to 0} x^\alpha = x^0$. But for the function $\hat{\omega}$ by (3.2) it holds

$$\lim_{\alpha \to 0} \hat{\omega}(x^\alpha) \neq 0 = \hat{\omega}(x^0)$$

so that property (PM3) is not satisfied, and $\tilde{\omega}$ is not a proximity measure.
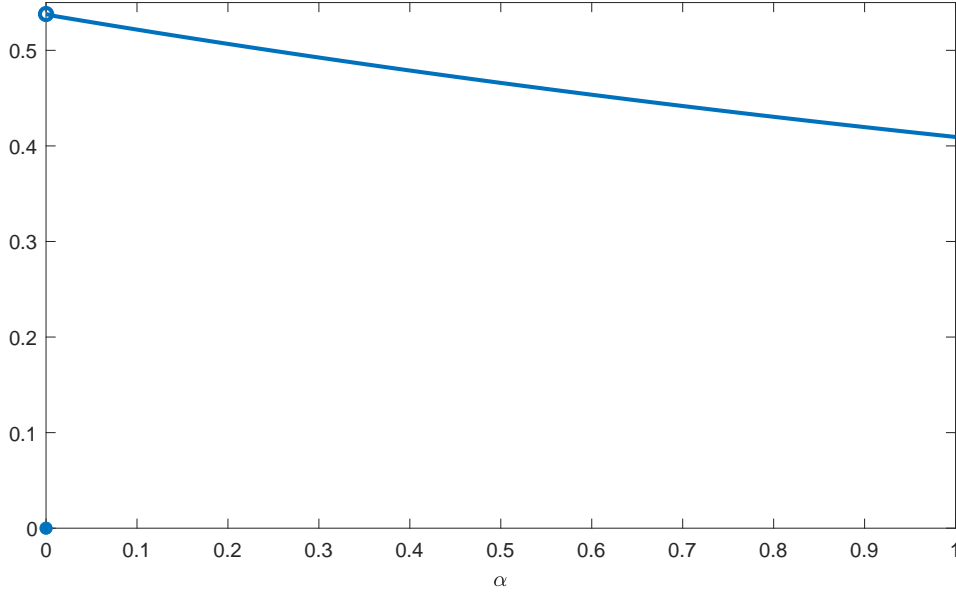

Figure 1: Function $\alpha \mapsto \hat{\omega}(x^\alpha)$

In Figure 1, it can be seen that $\hat{\omega}$ is not continuous at the efficient solution $x^0 \in S$ of (P1). This is exactly what we want to prevent and why we introduced property (PM3) in Definition 3.1. It is also important to notice that the function values $\hat{\omega}(x^\alpha), \alpha \neq 0$ are monotonically increasing for α decreasing to 0. Hence, the closer we get to the efficient solution x^0, the higher the function value gets, whereas a proximity measure should return values close to zero.

In [7] the authors used a scalarization approach to further investigate proximity measures for (CMOP). Moreover, they did not check the properties (PM1), (PM2), and (PM3). In the remaining part of this section, we first recall a proximity measure from

the literature for single-objective optimization problems. After that, in Section 3.2, we introduce a new proximity measure for the multi-objective problem (CMOP) based on the results from the single-objective case. This new measure is not based on a scalarization approach. Finally, another new proximity measure with simpler structure will be introduced in Section 3.3. That proximity measure is especially useful for optimization algorithms as it can be computed by solving a linear optimization problem only.

3.1 Single-Objective Case

For $m = 1$, the problem (CMOP) is a constrained single-objective optimization problem which we denote by (CSOP). In the single-objective case, it is not common to use the term efficient solution for an optimal solution, but to use the term minimal solution. For the remaining part of Section 3.1 we follow this convention.

As already mentioned in the introduction, in general it is a hard task to compute exact KKT points. Therefore, different relaxations have been presented in the literature. One of these relaxations are the so-called modified ε-KKT points. This concept was introduced in [10]. Those modified ε-KKT points can be seen as a relaxation of KKT points where a small deviation concerning the KKT conditions is allowed. In [10] the objective and constraint functions have not been assumed to be differentiable but only Lipschitz continuous, and the definitions used Clarke's subdifferential, see [4]. With the assumptions of our paper, this leads to the following definition.

Definition 3.3 *Let $x \in S$ be a feasible point for (CSOP) and $\varepsilon > 0$. If there exist $\hat{x} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}_+^p$ with*

(i) $\|\hat{x} - x\| \leq \sqrt{\varepsilon}$,

(ii) $\left\| \nabla f(\hat{x}) + \sum\limits_{j=1}^{p} \lambda_j \nabla g_j(\hat{x}) \right\| \leq \sqrt{\varepsilon}$,

(iii) $\sum\limits_{j=1}^{p} \lambda_j g_j(x) \geq -\varepsilon$,

then x is called a modified ε-KKT point for (CSOP).

Based on this concept, the authors from [10] introduced a candidate for a proximity measure that we present in the following definition.

Definition 3.4 *A function $\tilde{\omega} \colon \mathbb{R}^n \to \mathbb{R}$ based on modified ε-KKT points is given as*

$$\tilde{\omega}(x) := \inf \left\{ \varepsilon \in \mathbb{R}_+ \; \middle| \; \begin{array}{l} \|\hat{x} - x\| \leq \sqrt{\varepsilon}, \\[2mm] \left\| \nabla f(\hat{x}) + \sum\limits_{j=1}^{p} \lambda_j \nabla g_j(\hat{x}) \right\| \leq \sqrt{\varepsilon}, \\[2mm] \sum\limits_{j=1}^{p} \lambda_j g_j(x) \geq -\varepsilon, \\[2mm] g_j(x) \leq \varepsilon \text{ for all } j \in [p], \\[1mm] \lambda \in \mathbb{R}_+^p, \; \hat{x} \in \mathbb{R}^n \end{array} \right\}$$

for all $x \in \mathbb{R}^n$.

The proximity measure in its formulation in [10] is only defined on the feasible set S. To match our Definition 3.1 of proximity measures, we added the constraints $g_j(x) \leq \varepsilon$, $j \in [p]$. First, we show that properties (PM1) and (PM2) are satisfied for $\tilde{\omega}$.

Lemma 3.5 *The function $\tilde{\omega}$ satisfies (PM1) and (PM2).*

Proof. By definition we have $\tilde{\omega}(x) \geq 0$ for all $x \in \mathbb{R}^n$ and hence, property (PM1) holds. Now let $\bar{x} \in S$ be a minimal solution for (CSOP) in which the Abadie CQ holds. Then property (PM2) is satisfied by Theorem 2.2. $\qquad \square$

As already mentioned, our focus in this paper is on property (PM3). While it has not been examined in [10], in the following theorem we show that this property does indeed hold for $\tilde{\omega}$.

Theorem 3.6 *The function $\tilde{\omega}$ satisfies property (PM3).*

Proof. Let $\bar{x} \in S$ be a minimal solution for (CSOP) in which the Abadie CQ holds and $(x^i)_{\mathbb{N}} \subseteq \mathbb{R}^n$ a sequence of points with $\lim_{i \to \infty} x^i = \bar{x}$. We are interested in the sequence $(\tilde{\omega}(x^i))_{\mathbb{N}}$ and aim to apply Lemma 2.5. Let $(x^{i_k})_{k \in \mathbb{N}}$ be a subsequence of $(x^i)_{\mathbb{N}}$ which will be denoted by $(x^p)_{\mathbb{N}}$, i.e., $(\tilde{\omega}(x^p))_{\mathbb{N}}$ is a subsequence of $(\tilde{\omega}(x^i))_{\mathbb{N}}$. We now construct a subsequence of $(x^p)_{\mathbb{N}}$, and hence of $(\tilde{\omega}(x^p))_{\mathbb{N}}$, and show that $\tilde{\omega}$ converges to 0 on this subsequence.

As $\bar{x}$ is a minimal solution for (CSOP), by Theorem 2.2 there exists $\bar{\lambda} \in \mathbb{R}_+^p$ such that $(\bar{x}, 1, \bar{\lambda})$ is a KKT point. Hence, (KKT1),(KKT2), and (KKT4) hold which implies that

$$\nabla f(\bar{x}) + \sum_{j=1}^p \bar{\lambda}_j \nabla g_j(\bar{x}) = 0, \qquad \sum_{j=1}^p \bar{\lambda}_j g_j(\bar{x}) = 0, \qquad g(\bar{x}) \leq 0. \qquad (3.3)$$

Now let $\varepsilon > 0$. All functions $g_j, j \in [p]$ and the objective function f are continuously differentiable. Hence, every composition of those continuous functions and their continuous derivatives is continuous itself. So there exists $\delta_\varepsilon^1 > 0$ such that for all $x \in \mathbb{R}^n$ with $\|x - \bar{x}\| \leq \delta_\varepsilon^1$ it holds that

$$\left\| \nabla f(x) + \sum_{j=1}^p \bar{\lambda}_j \nabla g_j(x) - \left(\nabla f(\bar{x}) + \sum_{j=1}^p \bar{\lambda}_j \nabla g_j(\bar{x}) \right) \right\| \leq \sqrt{\varepsilon}. \qquad (3.4)$$

Moreover, there exists $\delta_\varepsilon^2 > 0$ such that for all $x \in \mathbb{R}^n$ with $\|x - \bar{x}\| \leq \delta_\varepsilon^2$ it holds that

$$\left| \sum_{j=1}^p \bar{\lambda}_j g_j(x) - \sum_{j=1}^p \bar{\lambda}_j g_j(\bar{x}) \right| \leq \varepsilon. \qquad (3.5)$$

Finally, there exists $\delta_\varepsilon^3 > 0$ such that for all $x \in \mathbb{R}^n$ with $\|x - \bar{x}\| \leq \delta_\varepsilon^3$ it holds that

$$|g_j(x) - g_j(\bar{x})| \leq \varepsilon \text{ for all } j \in [p]. \qquad (3.6)$$

Now define $\delta_\varepsilon := \min\{\delta_\varepsilon^1, \delta_\varepsilon^2, \delta_\varepsilon^3\} > 0$. Then for all $x \in \mathbb{R}^n$ with $\|x - \bar{x}\| \leq \delta_\varepsilon$ by (3.4), (3.5), (3.6), and (3.3) it holds that

$$\left\| \nabla f(x) + \sum_{j=1}^p \bar{\lambda}_j \nabla g_j(x) \right\| \leq \sqrt{\varepsilon},$$

$$\sum_{j=1}^p \bar{\lambda}_j g_j(x) \geq -\varepsilon, \qquad (3.7)$$

$$g_j(x) \leq \varepsilon \text{ for all } j \in [p].$$

10

As $(x^p)_\mathbb{N}$ converges to $\bar{x}$, there exists $p_\varepsilon \in \mathbb{N}$ with $\|x^p - \bar{x}\| \leq \delta_\varepsilon$ for all $p \geq p_\varepsilon$. Hence, for all $p \geq p_\varepsilon$ we obtain from (3.7) that ε, $\hat{x}^p = x^p$, and $\lambda = \bar{\lambda}$ is feasible for the optimization problem for $\tilde{\omega}(x^p)$ as given in Definition 3.4. This implies that $\tilde{\omega}(x^p) \leq \varepsilon$. Now let $(\varepsilon_n)_\mathbb{N} \subseteq \text{int}(\mathbb{R}_+)$ be a monotonically decreasing sequence with $\lim_{n \to \infty} \varepsilon_n = 0$. Then for all $n \in \mathbb{N}$ there exists $p_n = p_{\varepsilon_n} \in \mathbb{N}$ such that

$$\tilde{\omega}(x^{p_n}) \leq \varepsilon_n.$$

Moreover, without loss of generality, we can assume that for all $n \in \mathbb{N}$ it holds that $p_{n+1} > p_n$. This is just a result of what was discussed above. So there exists a subsequence $(x^{p_n})_{n \in \mathbb{N}}$ of $(x^p)_{p \in \mathbb{N}} = (x^{i_k})_{k \in \mathbb{N}}$ with

$$0 \leq \tilde{\omega}(x^{p_n}) \leq \varepsilon_n \text{ for all } n \in \mathbb{N}.$$

As $\varepsilon_n \to 0$ and using the squeeze theorem and (PM2) this leads to

$$\lim_{n \to \infty} \tilde{\omega}(x^{p_n}) = 0 = \tilde{\omega}(\bar{x}).$$

Overall, for every subsequence $(\tilde{\omega}(x^{i_k}))_{k \in \mathbb{N}} = (\tilde{\omega}(x^p))_\mathbb{N}$ of $(\tilde{\omega}(x^i))_\mathbb{N}$ there exists a sub-subsequence $(\tilde{\omega}(x^{i_{k_l}}))_{l \in \mathbb{N}}$ (described by $(\tilde{\omega}(x^{p_n}))_{n \in \mathbb{N}}$ above) with

$$\lim_{l \to \infty} \tilde{\omega}(x^{i_{k_l}}) = 0 = \tilde{\omega}(\bar{x}). \tag{3.8}$$

Finally, with Lemma 2.5

$$\lim_{i \to \infty} \tilde{\omega}(x^i) = 0 = \tilde{\omega}(\bar{x}).$$

$\square$

In the following, we give a brief comparison of our results to those in [10]. Then, in the next section, we extend the results to the case with multiple objectives. The main difference is that in [10] not the proximity measure $\tilde{\omega}$ was examined but (modified) ε-KKT points and sequences of (modified) ε-KKT points. One of their main results is [10, Theorem 3.6]. For that let $(\varepsilon_i)_{i \in \mathbb{N}} \subseteq \mathbb{R}_+$ be a sequence which converges to 0 and let $(x^i)_\mathbb{N} \subseteq S$ be a sequence of feasible points which converges to $\bar{x}$. The theorem states that if in $\bar{x} \in \mathbb{R}^n$ a certain CQ holds and, most of all, if x^i is a modified ε_i-KKT point for all $i \in \mathbb{N}$, then $\bar{x}$ is a KKT point.

This result says that a convergent sequence of modified ε-KKT points with ε decreasing to 0 converges to a KKT point. Instead, we have shown that for any sequence of (not necessarily feasible) points that converges to a minimal solution (in which a CQ holds and which is thus a KKT point) the value of $\tilde{\omega}$ has to converge to 0. For the feasible points of that sequence, this implies that they are a sequence of modified ε-KKT points with ε decreasing to 0.

For evolutionary algorithms that are used to solve (CSOP), we usually expect them to generate a sequence $(x^i)_\mathbb{N} \subseteq \mathbb{R}^n$ that converges to a minimal solution $\bar{x} \in S$. For such algorithms it is important to have a criterion to decide when to terminate. In other words one needs a criterion to decide whether the current point $x^i, i \in \mathbb{N}$ should be improved or not. We have shown that a small value of $\tilde{\omega}$ is at least a necessary condition for being close to a minimal solution. Hence, if $\tilde{\omega}(x^i)$ is not small then x^i should possibly be improved. As a consequence, a small value of $\tilde{\omega}$ can be used as termination criterion. This is not the case regarding the result in [10] as it already

11

assumes that $(x^i)_\mathbb{N} \subseteq \mathbb{R}^n$ is a sequence of modified ε-KKT points with ε decreasing to 0.

Another main result of [10] is their Theorem 3.5 which states that when all functions $f, g_j, j \in [p]$ are convex, Slater's CQ holds, and $x \in S$ is feasible and satisfies

$$f(x) \leq \inf_{x' \in S} f(x') + \varepsilon,$$

then x is a modified ε-KKT point.

This result can be used to show that for any minimal solution $\bar{x} \in S$ of (CSOP) and any sequence $(x^i)_\mathbb{N} \subseteq S$ of feasible points converging to $\bar{x}$, it holds

$$\lim_{i \to \infty} \tilde{\omega}(x^i) = \tilde{\omega}(\bar{x}).$$

This is closely related to (PM3) from Definition 3.4. However, for our result no convexity assumptions are needed.

3.2   A First Approach for Multi-Objective Problems

A possible approach to find proximity measures for the multi-objective optimization problem (CMOP) (with $m \geq 2$) is to generalize the results from the single-objective case from Section 3.1. While no proximity measures were presented in [18], the authors there did generalize the concept of modified ε-KKT points from Definition 3.3 as follows.

Definition 3.7 *Let $x \in S$ be a feasible point for* (CMOP) *and $\varepsilon > 0$. If there exist $\hat{x} \in \mathbb{R}^n, \eta \in \mathbb{R}^m_+$, and $\lambda \in \mathbb{R}^p_+$ with*

(i) $\|\hat{x} - x\| \leq \sqrt{\varepsilon}$,

(ii) $\left\| \sum\limits_{i=1}^m \eta_i \nabla f_i(\hat{x}) + \sum\limits_{j=1}^p \lambda_j \nabla g_j(\hat{x}) \right\| \leq \sqrt{\varepsilon}$,

(iii) $\sum\limits_{j=1}^p \lambda_j g_j(x) \geq -\varepsilon$,

(iv) $\sum\limits_{i=1}^m \eta_i = 1$,

then x is called a modified ε-KKT point for (CMOP).

Again, we adapted the definition to fit our assumption of differentiability. In [18, Theorem 3.4] it was shown that for a sequence $(x^i)_\mathbb{N}$ of modified ε-KKT points with ε decreasing to 0 and $\lim_{i \to \infty} x^i = \bar{x}$, the point $\bar{x}$ is a KKT point. As for instance evolutionary algorithms generate an arbitrary sequence of points, for which it is typically not guaranteed that the points are modified ε-KKT points, this result is not necessarily usable for such applications. A more applicable result when having evolutionary algorithms in mind is presented in [9]. There, a relation between so-called weakly ε-efficient solutions and modified ε-KKT points was given. For $\varepsilon > 0$, a feasible point $\bar{x} \in S$ is called weakly ε-efficient solution for (CMOP) with respect to $d \in \text{int}(\mathbb{R}^m_+), \|d\| = 1$ if there is no $x \in S$ with

$$f(x) + \varepsilon d < f(\bar{x}).$$

The following theorem is an adaption of [9, Theorem 3.7], where the result was proven in a more general setting.

12

Theorem 3.8 *Consider* (CMOP) *with convex functions* $f_i, i \in [m]$, $g_j, j \in [p]$ *and let Slater's CQ be satisfied, i.e., there exists* $x^* \in S$ *with* $g(x^*) < 0$. *Then every weakly* ε-*efficient solution for* (CMOP) *with respect to* $d \in int(\mathbb{R}_+^m)$ *is also a modified* ε-*KKT point.*

Weakly ε-efficient solutions $\bar{x} \in S$ have images $f(\bar{x})$ which are close to the image of the set of all weakly efficient solutions for (CMOP). By Theorem 3.8, a necessary condition (in case of Slater's CQ and convexity) for this 'ε-closeness', i.e., for $\bar{x}$ to be weakly ε-efficient, is that $\bar{x}$ is a modified ε-KKT point. This can be used as a selection or termination criterion in evolutionary algorithms. Such a relation as in Theorem 3.8 was also shown for the single-objective case in [10, Theorem 3.5], see the discussion on page 12. The downside of this result is that it requires convexity of the functions.
We now introduce a new proximity measure for which we need no convexity assumption to prove (PM1)-(PM3). Thereby we generalize the concept which was used in Definition 3.4 for the single-objective case.

Definition 3.9 *A function* $\omega\colon \mathbb{R}^n \to \mathbb{R}$ *based on modified* ε-*KKT points is given as*

$$\omega(x) := \inf \left\{ \varepsilon \in \mathbb{R}_+ \left| \begin{array}{l} \|\hat{x} - x\| \le \sqrt{\varepsilon}, \\[2mm] \left\| \sum_{i=1}^m \eta_i \nabla f_i(\hat{x}) + \sum_{j=1}^p \lambda_j \nabla g_j(\hat{x}) \right\| \le \sqrt{\varepsilon}, \\[2mm] \sum_{j=1}^p \lambda_j g_j(x) \ge -\varepsilon, \ \sum_{i=1}^m \eta_i = 1, \\[2mm] g_j(x) \le \varepsilon \text{ for all } j \in [p], \\[1mm] \eta \in \mathbb{R}_+^m, \ \lambda \in \mathbb{R}_+^p, \ \hat{x} \in \mathbb{R}^n \end{array} \right. \right\}$$

for all $x \in \mathbb{R}^n$.

We show that this function ω is indeed a proximity measure. We first state that properties (PM1) and (PM2) hold.

Lemma 3.10 *The function* ω *satisfies (PM1) and (PM2).*

Proof. By definition we have $\omega(x) \ge 0$ for all $x \in \mathbb{R}^n$ and hence, property (PM1) holds. Now let $\bar{x} \in S$ be an efficient solution for (CMOP) in which the Abadie CQ holds. Then property (PM2) is satisfied by Theorem 2.2. $\square$

In addition to property (PM2), we can also show a stronger relation between the zeros of ω and (exact) KKT points of (CMOP).

Lemma 3.11 *For any* $x \in \mathbb{R}^n$ *it holds that* $\omega(x) = 0$ *if and only if there exist* $\eta \in \mathbb{R}_+^m$ *and* $\lambda \in \mathbb{R}_+^p$ *such that* (x, η, λ) *is a KKT point.*

Proof. Let $x \in \mathbb{R}^n$ with $\omega(x) = 0$. This is the case if and only if $x \in S$ and there exist $\eta \in \mathbb{R}_+^m, \lambda \in \mathbb{R}_+^p$ with

$$\sum_{i=1}^m \eta_i \nabla f_i(x) + \sum_{j=1}^p \lambda_j \nabla g_j(x) = 0, \qquad \sum_{j=1}^p \lambda_j g_j(x) = 0, \qquad \sum_{i=1}^m \eta_i = 1,$$

i.e., (x, η, λ) is a KKT point. $\square$

Finally, we show that property (PM3) is satisfied for ω as well. This implies that ω is indeed a proximity measure.

Theorem 3.12 *The function ω satisfies (PM3).*

Proof. Let $\bar{x} \in S$ be an efficient solution for (CMOP) in which the Abadie CQ holds, $(x^i)_\mathbb{N} \subseteq \mathbb{R}^n$ a sequence with $\lim_{i \to \infty} x^i = \bar{x}$ and $(x^{i_k})_{k \in \mathbb{N}}$ a subsequence of $(x^i)_\mathbb{N}$ which will be denoted by $(x^p)_\mathbb{N}$. As in the proof of Theorem 3.6, we construct a subsequence of $(x^p)_\mathbb{N}$ and show that ω converges to 0 on this subsequence.

As $\bar{x}$ is efficient for (CMOP), there exist $\bar{\eta} \in \mathbb{R}^m_+$ and $\bar{\lambda} \in \mathbb{R}^p_+$ such that $(\bar{x}, \bar{\eta}, \bar{\lambda})$ is a KKT point by Theorem 2.2. Hence, (KKT1), (KKT2), (KKT4), and (KKT6) hold which implies that

$$\sum_{i=1}^m \bar{\eta}_i \nabla f_i(\bar{x}) + \sum_{j=1}^p \bar{\lambda}_j \nabla g_j(\bar{x}) = 0, \ \sum_{j=1}^p \bar{\lambda}_j g_j(\bar{x}) = 0, \ \sum_{i=1}^m \bar{\eta}_i = 1, \ g(\bar{x}) \leq 0. \qquad (3.9)$$

Now let $\varepsilon > 0$. All functions $f_i, i \in [m]$ and $g_j, j \in [p]$ are continuously differentiable. Thus, every composition of these continuous functions and their continuous derivatives is continuous itself. So there exists $\delta^1_\varepsilon > 0$ such that for all $x \in \mathbb{R}^n$ with $\|x - \bar{x}\| \leq \delta^1_\varepsilon$ it holds that

$$\left\| \sum_{i=1}^m \bar{\eta}_i \nabla f_i(x) + \sum_{j=1}^p \bar{\lambda}_j \nabla g_j(x) - \left(\sum_{i=1}^m \bar{\eta}_i \nabla f_i(\bar{x}) + \sum_{j=1}^p \bar{\lambda}_j \nabla g_j(\bar{x}) \right) \right\| \leq \sqrt{\varepsilon}. \qquad (3.10)$$

Moreover, there exists $\delta^2_\varepsilon > 0$ such that for all $x \in \mathbb{R}^n$ with $\|x - \bar{x}\| \leq \delta^2_\varepsilon$ it holds that

$$\left| \sum_{j=1}^p \bar{\lambda}_j g_j(x) - \sum_{j=1}^p \bar{\lambda}_j g_j(\bar{x}) \right| \leq \varepsilon. \qquad (3.11)$$

Finally, there exists $\delta^3_\varepsilon > 0$ such that for all $x \in \mathbb{R}^n$ with $\|x - \bar{x}\| \leq \delta^3_\varepsilon$ it holds that

$$|g_j(x) - g_j(\bar{x})| \leq \varepsilon \text{ for all } j \in [p]. \qquad (3.12)$$

Now define $\delta_\varepsilon := \min\{\delta^1_\varepsilon, \delta^2_\varepsilon, \delta^3_\varepsilon\} > 0$. Then for all $x \in \mathbb{R}^n$ with $\|x - \bar{x}\| \leq \delta_\varepsilon$ by (3.10), (3.11), (3.12), and (3.9) it holds that

$$\begin{aligned} \left\| \sum_{i=1}^m \bar{\eta}_i \nabla f_i(x) + \sum_{j=1}^p \bar{\lambda}_j \nabla g_j(x) \right\| &\leq \sqrt{\varepsilon}, \\ \sum_{j=1}^p \bar{\lambda}_j g_j(x) &\geq -\varepsilon, \\ \sum_{i=1}^m \bar{\eta}_i &= 1, \\ g_j(x) &\leq \varepsilon \text{ for all } j \in [p]. \end{aligned} \qquad (3.13)$$

As $(x^p)_\mathbb{N}$ converges to $\bar{x}$, there exists $p_\varepsilon \in \mathbb{N}$ with $\|x^p - \bar{x}\| \leq \delta_\varepsilon$ for all $p \geq p_\varepsilon$. Hence, for all $p \geq p_\varepsilon$ we obtain from (3.13) that ε, $\hat{x}^p = x^p$, $\eta = \bar{\eta}$, and $\lambda = \bar{\lambda}$ define a feasible point for the optimization problem for $\omega(x^p)$ as given in Definition 3.9, which implies $\omega(x^p) \leq \varepsilon$. Now by taking a sequence $(\varepsilon_n)_\mathbb{N} \subseteq \text{int}(\mathbb{R}_+)$ with $\lim_{n \to \infty} \varepsilon_n = 0$ as in the proof of Theorem 3.6, we can construct a subsequence $(x^{p_n})_{n \in \mathbb{N}}$ of $(x^p)_{p \in \mathbb{N}} = (x^{i_k})_{k \in \mathbb{N}}$

with $0 \leq \omega(x^{p_n}) \leq \varepsilon_n$ for all $n \in \mathbb{N}$. With the same arguments as in the proof of Theorem 3.6, we obtain

$$\lim_{i \to \infty} \omega(x^i) = 0 = \omega(\bar{x})$$

and thus ω satisfies property (PM3). □

At the beginning of this section, in Example 3.2, a naively definition of a candidate for a proximity measure was presented. This candidate function $\hat{\omega}$ was not a proximity measure. In particular, it was not continuous in every efficient solution in which the Abadie CQ holds. We now reconsider the problem (P1) from Example 3.2 and illustrate the advantages of ω compared to $\hat{\omega}$.

Example 3.13 *We consider again the constrained multi-objective optimization problem (P1) from Example 3.2 and the sequence of feasible points $(x^\alpha) \subseteq S$.*
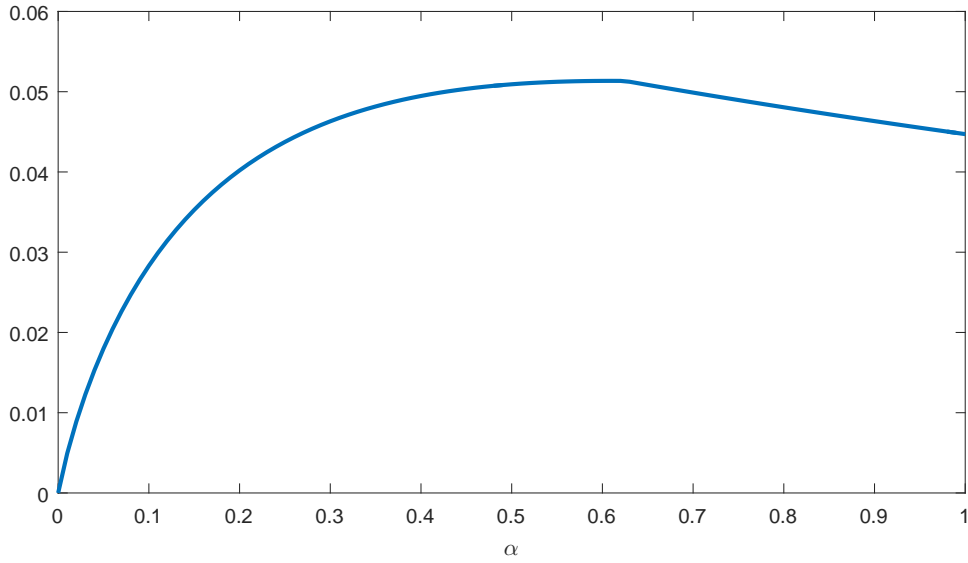
Figure 2: Function $\alpha \mapsto \omega(x^\alpha)$

The values for $\omega(x^\alpha)$ are shown in Figure 2. The function ω is continuous in x^0, what corresponds to $\alpha = 0$. Moreover, for $\alpha < 0.6$ the function values $\omega(x^\alpha)$ are monotonically decreasing to zero.

3.3 An Easy to Compute Proximity Measure

Although ω is a proximity measure in the sense of Definition 3.1, it is not really suited for computation. In general, the functions $\hat{x} \mapsto Dg(\hat{x})$ and $\hat{x} \mapsto Df(\hat{x})$ are nonlinear and even nonconvex. This makes solving the optimization problem to compute $\omega(x)$ for $x \in \mathbb{R}^n$ a hard task.

The introduction of $\hat{x}$ in the definition of ω is a result of the weaker assumptions used in [18] and [9] (as well as in [10] for the single-objective case). All functions have not been assumed to be necessarily continuously differentiable but only locally Lipschitz continuous. Hence, the gradients of those functions do not necessarily exist in all feasible points. This is why the authors used the generalized gradient as presented by Clarke in [4]. However, the generalized gradient is not really suited for computation and a proximity measure should thus avoid using it. This is why $\hat{x}$ was introduced.

15

By Rademacher's theorem every Lipschitz continuous function is differentiable almost everywhere. As all functions $f_i, i \in [m]$ and $g_j, j \in [p]$ are at least locally Lipschitz, this motivates that near to $x \in S$ there should be a $\hat{x} \in \mathbb{R}^n$ where those functions are differentiable and the generalized gradients could be replaced by gradients as in Definition 3.7.

However, in our paper, the functions are assumed to be continuously differentiable. Thus, it is possible to evaluate the gradients at every $\hat{x} \in \mathbb{R}^n$ and especially at $\hat{x} = x$. We will take this idea of fixing $\hat{x} = x$ as a starting point to define a new relaxation of KKT points. This approach was already briefly mentioned in [7] for single-objective optimization but without any further examination. The new relaxation of KKT points for (CMOP) which we introduce next is called simplified ε-KKT points.

Definition 3.14 *Let $x \in S$ be a feasible point for* (CMOP) *and $\varepsilon > 0$. If there exist $\eta \in \mathbb{R}_+^m$ and $\lambda \in \mathbb{R}_+^p$ with*

(i) $\left\| \sum_{i=1}^m \eta_i \nabla f_i(x) + \sum_{j=1}^p \lambda_j \nabla g_j(x) \right\|_\infty \leq \varepsilon,$

(ii) $\sum_{j=1}^p \lambda_j g_j(x) \geq -\varepsilon,$

(iii) $\sum_{i=1}^m \eta_i = 1,$

then x is called a simplified ε-KKT point for (CMOP).

The term 'simplified' is chosen as it is a simplified version of Definition 3.7, and as the corresponding proximity measure, which we will introduce shortly, can be computed easier. Compared to Definition 3.7 we have not only removed $\hat{x}$ from the definition but also replaced $\sqrt{\varepsilon}$ by ε and fixed the norm to the maximum norm in (i). As a consequence, in our new proximity measure, the function value at x only relies on the evaluation of g, Df, and Dg at x itself and the optimization problem to compute the value of this proximity measure can easily be formulated as a linear optimization problem.

Definition 3.15 *Define a function $\omega_s \colon \mathbb{R}^n \to \mathbb{R}$ based on simplified ε-KKT points by*

$$\omega_s(x) := \min \left\{ \varepsilon \in \mathbb{R}_+ \left|
\begin{array}{l}
\left\| \sum_{i=1}^m \eta_i \nabla f_i(x) + \sum_{j=1}^p \lambda_j \nabla g_j(x) \right\|_\infty \leq \varepsilon, \\
\sum_{j=1}^p \lambda_j g_j(x) \geq -\varepsilon, \ \sum_{i=1}^m \eta_i = 1, \\
g_j(x) \leq \varepsilon \text{ for all } j \in [p], \\
\eta \in \mathbb{R}_+^m, \ \lambda \in \mathbb{R}_+^p
\end{array}
\right. \right\}$$

for all $x \in \mathbb{R}^n$.

We can show that ω_s is indeed a proximity measure. This can be done analogously to the proofs of Lemma 3.10 and Theorem 3.12. In particular, fixing the norm to the maximum norm and replacing $\sqrt{\varepsilon}$ by ε has no effect concerning the proofs which rely on continuity statements and which already used $\hat{x} = x$. Also the characterization of exact KKT points from Lemma 3.11 holds for ω_s. We summarize these results in the following theorem.

Theorem 3.16 *The function ω_s is a proximity measure. Moreover, for any $x \in \mathbb{R}^n$ it holds that $\omega_s(x) = 0$ if and only if there exist $\eta \in \mathbb{R}^m_+$ and $\lambda \in \mathbb{R}^p_+$ such that (x, η, λ) is a KKT point.*

Not all results for modified ε-KKT points can easily be transferred to simplified ε-KKT points. For instance, the proof of Theorem 3.8 which can be found in [9] relies on Ekeland's variational principle. Hence, the relation cannot directly be extended as we would have to ensure $x = \hat{x}$ for that case. Whether such a statement can be found or not is an open question.

4 Numerical Results

In [10] the authors investigated the behavior of their proximity measure (which we also presented as $\tilde{\omega}$ in Definition 3.4) for single-objective optimization problems for iterates of the evolutionary algorithm RGA. They observed that the function value of their proximity measure decreased throughout the iterations for several test instances. As a result, the authors proposed to use their proximity measure as a termination criterion for evolutionary algorithms.

While this could also be done with the new proximity measures ω and ω_s which we introduced in this paper for multi-objective optimization problems (CMOP), we focus on another illustration of the measures, in particular of ω_s. In the previous section it was already discussed why this proximity measure is more suited for numerical evaluation and use with optimization algorithms: it can be calculated by solving a linear optimization problem only.

For the following examples we have generated k points equidistantly distributed in the preimage space and then computed the value of the proximity measure ω_s in MATLAB using `linprog`. If the computed value was below a specified limit of $\alpha > 0$, those points were selected as possible efficient solutions (also called solution candidates). The set of those points will be denoted by $\mathcal{C}$ in this section. Moreover, the set of efficient solutions will be denoted by $\mathcal{E}$. For visualization the default `parabula` colormap of MATLAB was used. Hence, dark blue corresponds to a value of ω_s close to 0. Then, as the value rises up, the color turns green and finally yellow for the highest value of ω_s that was reached within the generated discretization of the preimage space. The sets $\mathcal{C}$ and $f(\mathcal{C})$ are illustrated by red triangles. The generated data and the MATLAB files are available from the authors on reasonable request.

Test instance 1 The first example is convex and taken from [5].

$$\min_{x \in \mathbb{R}^2} f(x) := \begin{pmatrix} x_1^2 + x_2^2 \\ (x_1 - 5)^2 + (x_2 - 5)^2 \end{pmatrix} \tag{BK1}$$
$$\text{s.t.} \quad x_1, x_2 \in [-5, 10].$$

The set of efficient solutions for this multi-objective optimization problem is

$$\mathcal{E} := \left\{ x \in \mathbb{R}^2 \mid x_1 = x_2 \in [0, 5] \right\}.$$

For computation in MATLAB a total of $k = (64 + 1)^2 = 4225$ points were generated equidistantly distributed in $S = [-5, 10]^2$. A number of $|\mathcal{C}| = 21$ points lead to a value of ω_s lower or equal to $\alpha = 0.001$. In particular, for the set $\mathcal{C}$ delivered by MATLAB it holds $\mathcal{C} \subseteq \mathcal{E}$. The result is shown in Figure 3a and Figure 3b.

(a) Results for $k = 4225, \alpha = 0.001$ in the image space

(b) Results for $k = 4225, \alpha = 0.001$ in the preimage space
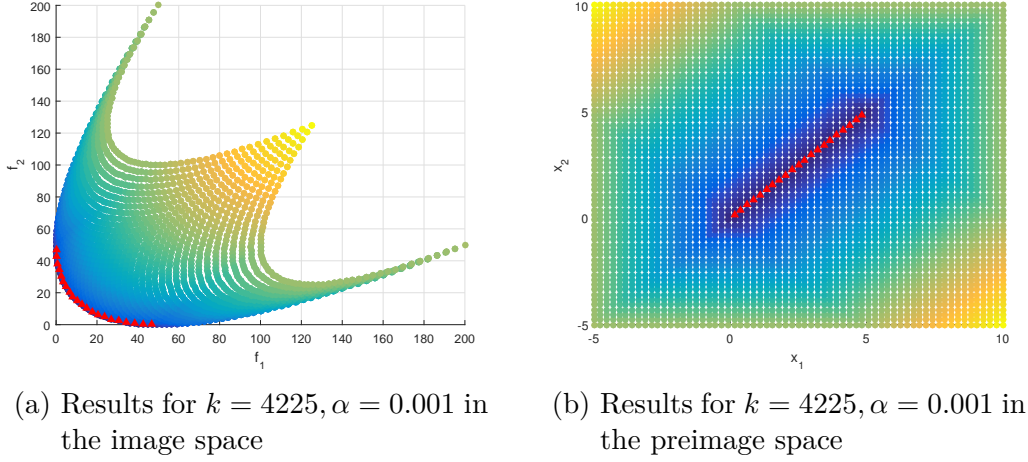
Figure 3: Numerical results for (BK1)

As (BK1) is a convex problem and the Abadie CQ holds for all $x \in S$ due to the linear constraints, the KKT conditions are not only a necessary optimality condition (see Theorem 2.2) but also sufficient by Lemma 2.4. Moreover, it was already discussed that $\omega_s(x) = 0$ if and only if there exist $\eta \in \mathbb{R}^m_+$ and $\lambda \in \mathbb{R}^p_+$ such that (x, η, λ) is a KKT point. This implies that for (BK1) $\omega_s(x) = 0$ if and only if $x \in S$ is a weakly efficient solution for this problem. Thus, ω_s is well suited for characterizing weakly efficient solutions for (BK1) and also as a termination criterion for optimization algorithms.

Test instance 2 In [22] Srinivas and Deb presented the following non-convex problem with convex but nonlinear constraint functions.

$$
\begin{aligned}
\min_{x \in \mathbb{R}^2} f(x) &:= \begin{pmatrix} 2 + (x_1 - 2)^2 + (x_2 - 1)^2 \\ 9x_1 - (x_2 - 1)^2 \end{pmatrix} \\
\text{s.t.} \quad g_1(x) &:= x_1^2 + x_2^2 - 225 \leq 0, \\
g_2(x) &:= x_1 - 3x_2 + 10 \leq 0, \\
x_1, x_2 &\in [-20, 20]
\end{aligned}
\tag{SRN}
$$

The set of efficient solutions for this problem is presented in [8] as

$$
\mathcal{E} := \left\{ x \in \mathbb{R}^2 \;\middle|\; x_1 = -2.5, x_2 \in [2.50, 14.79] \right\}.
$$

For the computation with MATLAB, a total of $k = (64 + 1)^2 = 4225$ points were generated equidistantly distributed in $X = [-20, 20]^2$. The results can be seen in Figure 4a and Figure 4b.

In the figures, it may look as if the Pareto Front $f(\mathcal{E})$ is larger than the representation which is covered by the candidates $f(\mathcal{C})$. This is not the case, as there are a lot of infeasible points within the set X. For clarifying this, in Figures 4c and 4d only feasible points and the corresponding images are shown.

Most of the $|\mathcal{C}| = 25$ candidates presented by the algorithm are efficient solutions for (SRN). However, there are 5 candidates which are actually not belonging to $\mathcal{E}$. An approach to improve this could be to reduce the acceptance bound α, but the result remains unchanged even for $\alpha = 1 \cdot 10^{-8}$. In case one decreases α further, the set $\mathcal{C}$ gets smaller and some gaps appear in the preimage space (and as a consequence in the
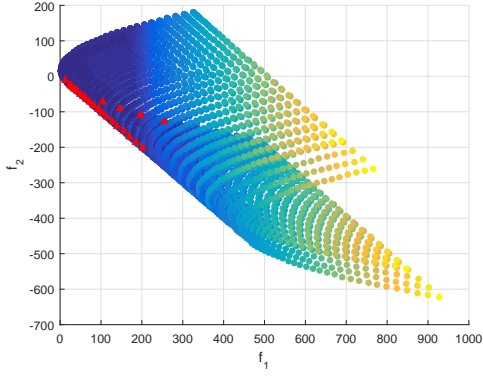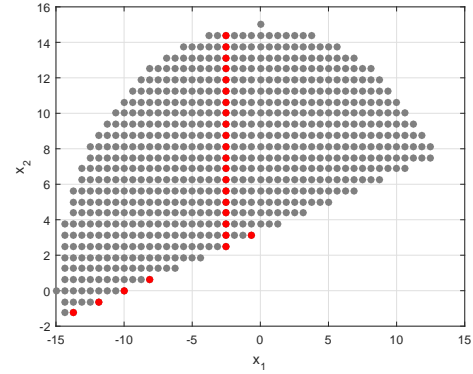
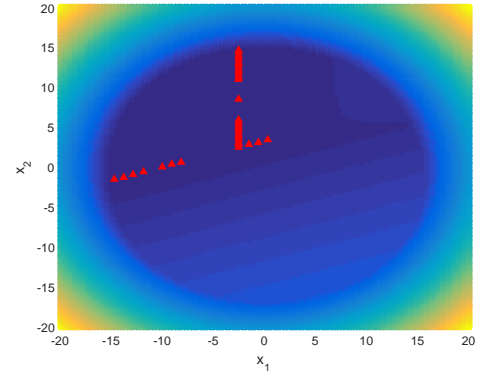(a) Results for $k = 4225, \alpha = 0.001$ in the image space
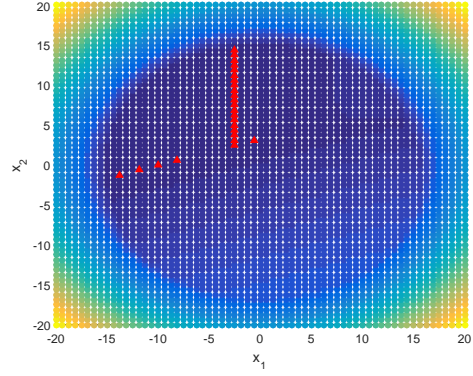
(b) Results for $k = 4225, \alpha = 0.001$ in the preimage space
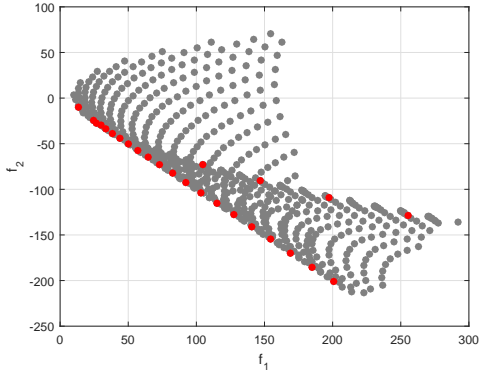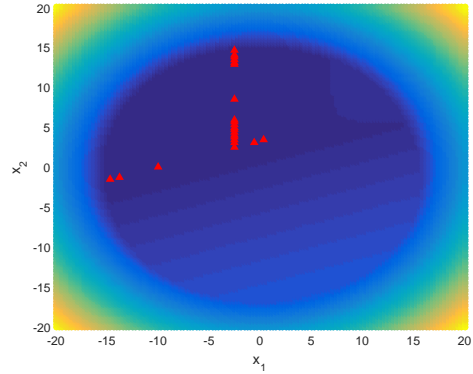
(c) Images of feasible points for $k = 4225, \alpha = 0.001$

(d) Feasible points for $k = 4225, \alpha = 0.001$

(e) Results for $k = 16641, \alpha = 1 \cdot 10^{-9}$ in the preimage space

(f) Results for $k = 16641, \alpha = 1 \cdot 10^{-12}$ in the preimage space

Figure 4: Numerical results for (SRN)

image space as well). In terms of quality, the result remains the same as the set $\mathcal{C}$ still contains some points that do not belong to $\mathcal{E}$. The effect is the same when choosing a finer discretization. This is shown in Figures 4e and 4f for $k = (128 + 1)^2 = 16641$ equidistantly distributed points in X.

The reason why there are points in $\mathcal{C}$ that are not belonging to $\mathcal{E}$ is just that these points (approximately) satisfy the KKT conditions without being efficient. Recall that the KKT conditions are just a necessary optimality condition and that they are sufficient

only in case of convexity, see Lemma 2.4.

Due to the specific structure of $\mathcal{E}$ for this test instance, we examined the influence of not discretizing equidistantly but randomly. The results using 16641 points are given in Figure 5 for two different values of α. First, we chose $\alpha = 0.001$ as this worked very well for (BK1) and (SRN) with equidistantly distributed points in the preimage space, see Figures 3 and 4b. However, for (SRN) with randomly distributed points, only a single solution candidate is found by MATLAB for $\alpha = 0.001$ and this is $x^c = (-2.3746, 2.5611)^\top$, see Figure 5a. When increasing α, the set $\mathcal{C}$ starts to get larger. For instance, Figure 5b shows the results for $\alpha = 0.01$. Comparing the results to those seen in Figure 4b, the structure of the sets of solution candidates is quite similar. This is exactly what we could expect due to the continuity of ω_s in every efficient solution. This specific run shows that the tolerance α should be chosen carefully. If it is too small like in this case for $\alpha = 0.001$, only few solution candidates will be found. On the other hand, choosing a large α can result in solution candidates that are not close to the set of efficient solutions $\mathcal{E}$ at all.

(a) Results for $k = 16641, \alpha = 0.001$ in the preimage space

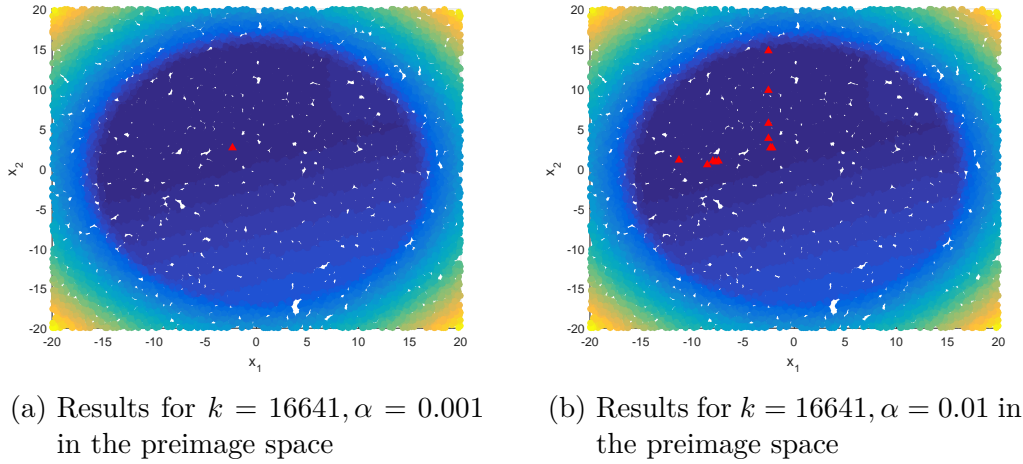(b) Results for $k = 16641, \alpha = 0.01$ in the preimage space

Figure 5: Numerical results for (SRN) with 16641 randomly distributed points

Test instance 3 This test instance by Osyczka and Kundu is taken from [20]. It has a larger number of $n = 6$ optimization variables and a larger number of constraints as well. The objective function $f \colon \mathbb{R}^6 \to \mathbb{R}^2$ is given as

$$f(x) := \begin{pmatrix} -(25(x_1 - 2)^2 + (x_2 - 2)^2 + (x_3 - 1)^2 + (x_4 - 4)^2 + (x_5 - 1)^2) \\ x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 \end{pmatrix} \quad \forall x \in \mathbb{R}^6$$

and the optimization problem is

$$\min_{x \in \mathbb{R}^6} f(x) \quad \text{s.t.} \quad \begin{aligned} g_1(x) &:= -x_1 - x_2 + 2 \le 0, \\ g_2(x) &:= x_1 + x_2 - 6 \le 0, \\ g_3(x) &:= -x_1 + x_2 - 2 \le 0, \\ g_4(x) &:= x_1 - 3x_2 - 2 \le 0, \\ g_5(x) &:= (x_3 - 3)^2 + x_4 - 4 \le 0, \\ g_6(x) &:= -(x_5 - 3)^2 - x_6 + 4 \le 0, \\ x_1, x_2 &, x_6 \in [0, 10], x_3, x_5 \in [1, 5], x_4 \in [0, 6]. \end{aligned} \quad \text{(OSY)}$$

A characterization of the set $\mathcal{E}$ for this problem can be found in [8]. As the set $X := [0, 10]^2 \times [1, 5] \times [0, 6] \times [1, 5] \times [0, 10]$ is quite huge compared to $\mathcal{E}$, we decided to consider
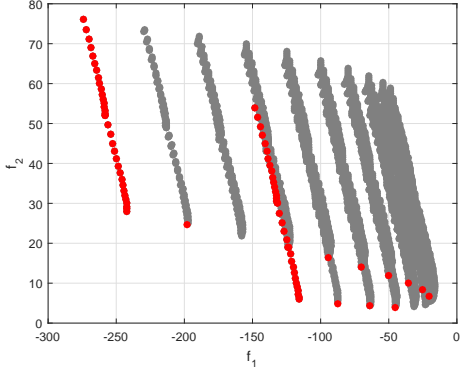
$\hat{X} := [0,5] \times [0,2] \times [1,5] \times \{0\} \times [1,5] \times \{0\}$. For this set we still have $\mathcal{E} \subseteq \hat{X}$. A total of $k = (16+1)^4 = 83521$ points were generated equidistantly distributed in $\hat{X}$. The MATLAB implementation found 70 candidates with a value of ω_s less or equal to $\alpha = 0.001$. The result can be seen in Figure 6a.
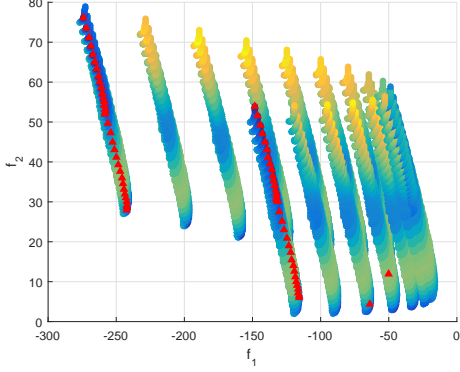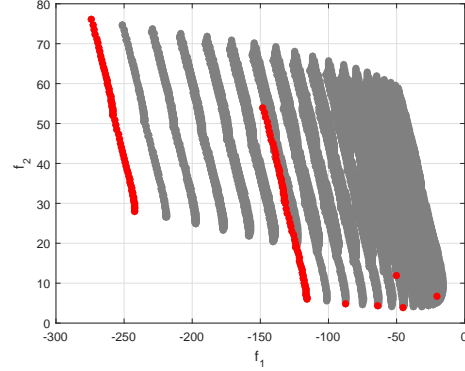

(a) Results for $k = 83521, \alpha = 0.001$ in the image space
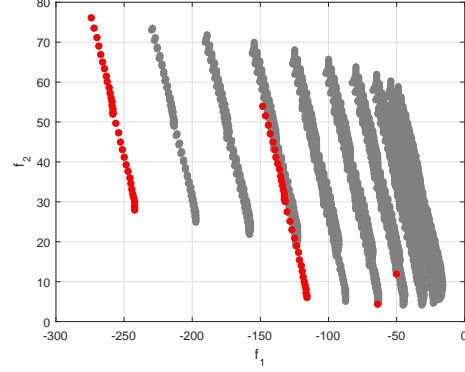
(b) Images of feasible points for $k = 83521, \alpha = 0.001$

(c) Images of feasible points for $k = 83521, \alpha = 0.1$

(d) Images of feasible points for $k = 1185921, \alpha = 0.001$

Figure 6: Numerical results for (OSY)

For a better characterization of the set $\mathcal{E}$ of all efficient solutions we set

$$\mathcal{E}_1 := \{(5,1,\beta,0,5,0) \in S \mid \beta \in [1,5]\},$$
$$\mathcal{E}_2 := \{(5,1,\beta,0,1,0) \in S \mid \beta \in [1,5]\},$$
$$\mathcal{E}_3 := \{(0,2,\beta,0,1,0) \in S \mid \beta \in [1,3.73]\}.$$

For those sets it holds that $\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3 \subseteq \mathcal{E}$. Considering the (discretized) image $f(S)$ in Figure 6b, $\mathcal{E}_1$ contains all efficient solutions belonging to the upper half of the left 'stroke' and $\mathcal{E}_2$ the efficient solutions belonging to its lower half. The set $\mathcal{E}_3$ contains all efficient solutions belonging to the stroke at about $f_1 \approx -120$. In Figure 6a, it might look as if actually none of the elements within $\mathcal{C}$ is an efficient solution. However, the images of infeasible points are also included in that figure. For this reason, in Figure 6b only the images of feasible points are shown, and it can be seen that $\mathcal{C}$ contains efficient and locally efficient solutions. In particular, out of the 69 points in $\mathcal{C}$ there are 17 within $\mathcal{E}_1$, 17 within $\mathcal{E}_2$ and 11 within $\mathcal{E}_3$. These are all of the 83521 generated points

that belong to the sets $\mathcal{E}_1, \mathcal{E}_2,$ and $\mathcal{E}_3$. Also the point $x^c := (0.625, 1.375, 1, 0, 1, 0) \in \mathcal{C}$ is part of the set of all efficient solutions $\mathcal{E}$.

Nevertheless, there are 24 points which do not belong to $\mathcal{E}$. There is one outlier which is easy to spot in Figure 6a. In addition, there are also some locally efficient solutions which can be seen as an extension of $\mathcal{E}_3$. In particular, these are

$$\mathcal{C}_1 := \{(0, 2, \beta, 0, 1, 0) \in S \mid \beta \in \{3.75, 4.00, 4.25, 4.50, 4.75, 5.00\}\} \text{ and}$$

$$\mathcal{C}_2 := \left\{ (0, 2, \beta, 0, 5, 0) \in S \mid \beta \in \left\{ \begin{array}{l} 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, \\ 3.25, 3.50, 3.75, 4.00, 4.25, 4.50, 4.75, 5.00 \end{array} \right\} \right\}.$$

In the image space, $f(\mathcal{E}_3)$ is the lower part of the line at about $f_1 \approx -120$ rising up to $f_2 \approx 18$. The upper end of this line with $f_2 \geq 30$ is the image of $\mathcal{C}_2$. The points in between belong to $\mathcal{C}_1$.

In a next step, we aimed to obtain those points at the bottom of the 'strokes' which are indeed images of efficient solutions. As a first approach, α was increased to $\alpha = 0.1$. The result is shown in Figure 6c. Another idea was to keep $\alpha = 0.001$ and increase the fineness of the discretization, in particular to choose $k = (32 + 1)^4 = 1185921$ points equidistantly distributed in the preimage space. This leads to the results which can be seen in Figure 6d. In both cases some of the (efficient) bottom tips were found. However, we also included more points which are not approximately efficient solutions within $\mathcal{C}$.

5   Conclusions

We presented two new proximity measures ω and ω_s for (CMOP) in the sense of Definition 3.1. The proximity measure ω is a generalization of the proximity measure presented in [10] for the single-objective case by using the results from [9] and [18]. One drawback of this approach is that ω can be hard to compute. Thus, it is not really suited for use within optimization algorithms. This is why we introduced the proximity measure ω_s. Compared to ω, the computation of $\omega_s(x)$ for some $x \in \mathbb{R}^n$ only relies on a single evaluation of $g, Df,$ and Dg. Moreover, it only requires solving a linear optimization problem, which makes its computation a lot faster.

In addition, the ability of ω_s to characterize the proximity of a certain point $x \in \mathbb{R}^n$ to the set of efficient solutions for (CMOP) was demonstrated in the previous section. As a result, ω_s is well suited for numerical applications. In particular, we have illustrated its capabilities as an additional criterion for candidate selection or termination in evolutionary algorithms.

It could now be argued that the computation of ω_s relies on solving an optimization problem and hence, all the problems mentioned in the introduction as limited accuracy are critical aspects as well. While such limitations should always be taken into account, the computation of ω_s relies on solving a linear optimization problem. For linear optimization problems, exact solvers such as SoPlex (see [15]) are available and can handle even numerically troublesome problems.

At no point we had to assume $m \geq 2$ for the dimension of the image space. Hence, all results hold still for single-objective optimization problems. Moreover, the case of unconstrained problems is also contained as a special case. As a consequence, our proximity measure ω_s can be used for a large class of optimization problems.

References

[1] M. Abouhawwash and M. A. Jameel, Evolutionary multi-objective optimization using Benson's Karush-Kuhn-Tucker proximity measure, in Lecture Notes in Computer Science, vol. 11411, Springer, 2019, pp. 27–38.

[2] M. Abouhawwash, H. Seada, and K. Deb, Towards faster convergence of evolutionary multi-criterion optimization algorithms using Karush Kuhn Tucker optimality based local search, Computers & Operations Research, 79 (2017), pp. 331–346.

[3] R. Andreani, G. Haeser, and J. M. Martínez, On sequential optimality conditions for smooth constrained optimization, Optimization, 60 (2011), pp. 627–641.

[4] F. H. Clarke, Optimization and Nonsmooth Analysis, vol. 5 of Classics in applied mathematics, Society for Industrial and Applied Mathematics, 1990.

[5] A. L. Custódio, J. F. A. Madeira, A. I. F. Vaz, and L. N. Vicente, Direct multisearch for multiobjective optimization, SIAM Journal on Optimization, 21 (2011), pp. 1109–1140.

[6] K. Deb, Multi-objective optimization using evolutionary algorithms, John Wiley & Sons, 2001.

[7] K. Deb, M. Abouhawwash, and J. Dutta, An optimality theory based proximity measure for evolutionary multi-objective and many-objective optimization, in Lecture Notes in Computer Science, vol. 9019, Springer, 2015.

[8] K. Deb, A. Pratap, and T. Meyarivan, Constrained test problems for multi-objective evolutionary optimization, in Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2001, pp. 284–298.

[9] M. Durea, J. Dutta, and C. Tammer, Stability properties of KKT points in vector optimization, Report No. 18, Martin-Luther-Universität Halle-Wittenberg, (2009).

[10] J. Dutta, K. Deb, R. Tulshyan, and R. Arora, Approximate KKT points and a proximity measure for termination, Journal of Global Optimization, 56 (2012), pp. 1463–1499.

[11] M. Ehrgott, Multicriteria optimization, Springer, Berlin New York, 2005.

[12] M. Feng and S. Li, An approximate strong KKT condition for multiobjective optimization, TOP, 26 (2018), pp. 489–509.

[13] B. Gebken, S. Peitz, and M. Dellnitz, A descent method for equality and inequality constrained multiobjective optimization problems, in Numerical and Evolutionary Optimization – NEO 2017, Springer International Publishing, 2018, pp. 29–61.

[14] G. Giorgi, B. Jiménez, and V. Novo, Approximate Karush–Kuhn–Tucker condition in multiobjective optimization, Journal of Optimization Theory and Applications, 171 (2016), pp. 70–89.

[15] A. M. Gleixner, D. E. Steffy, and K. Wolter, Improving the accuracy of linear programming solvers with iterative refinement, in Proceedings of the 37th International Symposium on Symbolic and Algebraic Computation - ISSAC '12, ACM Press, 2012.

[16] A. Göpfert and R. Nehse, Vektoroptimierung: Theorie, Verfahren und Anwendungen, Teubner, 1990.

[17] J. Jahn, Vector Optimization, Springer, 2004.

[18] P. Kesarwani, P. K. Shukla, J. Dutta, and K. Deb, Approximations for Pareto and proper Pareto solutions and their KKT conditions. `http://www.optimization-online.org/DB_HTML/2018/10/6845.html`, 2019.

[19] K. Miettinen, Nonlinear Multiobjective Optimization, Springer, 1998.

[20] A. Osyczka and S. Kundu, A new method to solve generalized multicriteria optimization problems using the simple genetic algorithm, Structural Optimization, 10 (1995), pp. 94–99.

[21] Y. Sawaragi, Theory of multiobjective optimization, Academic Press, Orlando, 1985.

[22] N. Srinivas and K. Deb, Multiobjective optimization using nondominated sorting in genetic algorithms, Evolutionary Computation, 2 (1994), pp. 221–248.