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

Gabriele Eichfelder∗, Leo Warnow∗

December 17, 2019

Abstract

A central requirement for solving optimization problems numerically with the help of computer algorithms is the verification of the optimality of a found solution. A frequently used approach to meet this requirement 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 KKT conditions, can easily be computed, and moreover 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, which we also illustrate numerically for some common test problems.

Key Words: Multiobjective Optimization, KKT Approximation, Proximity Measure

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

1 Introduction

In applications, often one has to deal with not only one but multiple objectives at the same time. This leads to multi-objective optimization problems. A main goal is now to find globally optimal solutions for such optimization problems using computer algorithms, e.g. evolutionary algorithms as proposed in [6].

A central requirement is then to make sure that solutions found by these algorithms are actually optimal solutions or at least approximately optimal solutions. One way to do that is to consider necessary optimality conditions under certain regularity assumptions. Although these conditions cannot guarantee that a given point is an optimal solution, it is known that if they are not satisfied it is not an optimal solution. In this paper we present two proximity measures which characterize the fulfillment of the so-called Karush-Kuhn-Tucker (KKT) optimality conditions. They can be used for candidate selection and also as a termination criterion for evolutionary algorithms. Moreover, we will show that the presented proximity measures are continuous in every efficient solution only assuming continuously differentiable objective and constraint functions as well as regularity.

∗Institute for Mathematics, Technische Universität Ilmenau, Po 10 05 65, D-98684 Ilmenau, Germany, {gabriele.eichfelder,leo.warnow}@tu-ilmenau.de
Proximity measures can not only be used for evolutionary, but also for deterministic algorithms. However, the practical implementation and use of such algorithms on a computer system has some limitations that one should keep in mind. On the one hand, these systems only work with limited accuracy. On the other hand, one would like to get an output of the algorithm after finite time. Especially with 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 optimal solution or at least an approximately optimal 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. Within the last decade, several concepts on how to do this have been presented. A main idea used in these concepts, is to provide a sequence of points that satisfy some relaxed KKT conditions and to show convergence of this sequence towards a KKT point.

A possible 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 [12] and [11].

The idea to use a relaxed version of KKT points to define proximity measures (also called error measures) does appear more recently in [10] for single-objective optimization problems. There, the relaxation of KKT points are so-called $\varepsilon$-KKT and modified $\varepsilon$-KKT points. Their definition does no longer rely on a sequence of points but only on a single feasible point. The authors then propose to use a proximity measure which is based on this relaxation and hence, based on the KKT error. One approach to use this idea for multiobjective optimization is by scalarization of the objective function, which was discussed in [7], [2], and [1]. The concept of (modified) $\varepsilon$-KKT points has also been extended to multi-objective optimization problems without using scalarization in [9] and [16], 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 $\varepsilon$-KKT points with $\varepsilon \to 0$ does indeed converge towards a KKT point. In terms of proximity measures this means that only for certain sequences convergence is shown. Having evolutionary algorithms in mind, it would be more interesting to know whether a proximity measure is continuous concerning any sequence of points converging to an efficient solution, which is much stronger. This is what we will show for our new proximity measures in this paper. In fact, we will present two proximity measures that are continuous at least in every 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 recap regularity assumptions as well as necessary optimality criteria. We then briefly discuss in Section 3 the concept of modified $\varepsilon$-KKT points from [10] and [16]. Then, we introduce our new proximity measures for multiobjective optimization problems. Finally in Section 4 some numerical results for using the proposed proximity measure for candidate selection for instance in evolutionary algorithms are presented.
2 Notations and Definitions

Within this paper, for a positive integer \( n \in \mathbb{N} \) the set of all integers in range from 1 to \( n \) is denoted by \( [n] := \{1, \ldots, n\} \). For a differentiable function \( f : \mathbb{R}^n \to \mathbb{R}^m \) the Jacobian matrix of \( f \) at \( x \in \mathbb{R}^n \) is \( Df(x) \). Now let \( x, \bar{x} \in \mathbb{R}^n \). Then relations are meant to be read component-wise as

\[
\begin{align*}
x \leq \bar{x} & \iff x_i \leq \bar{x}_i \text{ for all } i \in [n], \\
x < \bar{x} & \iff x_i < \bar{x}_i \text{ for all } i \in [n].
\end{align*}
\]

In this paper we will focus on multi-objective optimization problems with inequality constraints. For this reason, denote by \( f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m \) the objective function as well as by \( g = (g_1, \ldots, g_p) : \mathbb{R}^n \to \mathbb{R}^p \) a given constraint function. This leads to the problem

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

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. As the different objective functions are usually competing with each other, in general it is not possible to find some feasible point that minimizes all objectives at the same time. So some more suitable terms of optimality are needed.

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

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

It is called weakly efficient or weakly Edgeworth-Pareto-minimal for (CMOP) if there is no \( x \in S \) with

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

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 regularity condition has to hold for an optimal solution \( \bar{x} \) to actually satisfy the necessary optimality conditions. For instance the Abadie Constraint Qualification is such a regularity condition. Thus, we shortly recall the definition of the contingent cone as well as the linearized contingent cone to the set \( S \) at some point \( \bar{x} \). So let \( S \subseteq \mathbb{R}^n \) be a nonempty set as defined above and \( \bar{x} \in \text{cl}(S) \). Then the contingent cone at \( S \) in \( \bar{x} \) is given as

\[
T(S, \bar{x}) := \left\{ d \in \mathbb{R}^n \mid \exists (x^i)_i \subseteq S, (\lambda_i)_i \subseteq \text{int}(\mathbb{R}_+) : \bar{x} = \lim_{i \to \infty} x^i, d = \lim_{i \to \infty} \lambda_i(x^i - \bar{x}) \right\}.
\]

Moreover the linearized (contingent) cone is given by

\[
T^{\text{lin}}(S, \bar{x}) := \left\{ d \in \mathbb{R}^n \mid \nabla g_j(\bar{x})^\top d \leq 0 \ \forall j \in I(\bar{x}) \right\}.
\]

It always holds \( T(S, \bar{x}) \subseteq T^{\text{lin}}(S, \bar{x}) \). Moreover, we say that the Abadie Constraint Qualification (Abadie CQ) holds for some \( \bar{x} \in S \) if

\[
T(S, \bar{x}) = T^{\text{lin}}(S, \bar{x}).
\]
In case of affine linear constraint functions \( g_j, j \in [p] \) the Abadie CQ is satisfied for every \( x \in S \). But often stronger regularity assumptions than the Abadie CQ are used as these are easier to check. 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

\[
\nabla g_j(\bar{x})^\top d < 0 \text{ for all } j \in I(\bar{x}).
\]

Another well known criterion 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. Moreover, for those 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. It is satisfied, if there exists some \( x^* \in S \) such that

\[
g(x^*) < 0.
\]

Moreover, if Slater’s Constraint Qualification is satisfied then the Abadie CQ holds for every feasible point \( \bar{x} \in S \).

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

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

\[
g(x) \leq 0, \quad (\text{KKT2})
\]

\[
\lambda \geq 0, \quad (\text{KKT3})
\]

\[
\lambda^\top g(x) = 0, \quad (\text{KKT4})
\]

\[
\eta \geq 0, \quad (\text{KKT5})
\]

\[
\sum_{i=1}^{m} \eta_i = 1 \quad (\text{KKT6})
\]

are called Karush-Kuhn-Tucker (KKT) conditions. If a tuple \((x, \eta, \lambda)\) satisfies those six conditions it is called a KKT point for \((\text{CMOP})\) and \(\eta\) and \(\lambda\) are called Lagrange multipliers. Under certain regularity assumptions this can now be used to gain a necessary optimality condition for \((\text{CMOP})\).

**Theorem 2.2** ([14, Satz 2.35]) Let \( \bar{x} \in S \) be weakly efficient for \((\text{CMOP})\) and let the Abadie CQ hold at \( \bar{x} \). Then there exist Lagrange 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.

When taking a look back at the KKT conditions, they may appear to be related to a weighted sum scalarization of \((\text{CMOP})\), in particular regarding \((\text{KKT1})\) with the term

\[
\sum_{i=1}^{m} \eta_i \nabla f_i(x).
\]

But, this is not the case, which can be seen as no convexity is required for the statement in Theorem 2.2. If the objective functions \( f_i, i \in [m] \) and all constraint functions \( g_j, j \in [p] \) do meet some special convexity requirements, we obtain a sufficient optimality condition.
Lemma 2.3 ([15, Corollary 7.24]) Let $\bar{x} \in S$ be a feasible point, $f_i$ pseudoconvex for all $i \in [m]$ and $g_j$ quasiconvex for all $j \in I(\bar{x})$. If there exist Lagrange 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).

Note that every continuously differentiable convex function is also pseudoconvex and quasiconvex and that the results do hold for such convex functions as well.

3 Proximity Measures

As motivated right at the beginning of this paper, we want to provide a new 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 within computer algorithms, e.g. evolutionary algorithms. First, we collect the properties such a proximity measure should satisfy within the next definition.

Definition 3.1 A function $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ is called a proximity measure if for every efficient point $\bar{x} \in S$ of (CMOP) in which the Abadie CQ holds, and every sequence $(x^i)_N \subseteq \mathbb{R}^n$ with

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

the following three properties are satisfied:

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

(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 a stronger requirement and also the property we will focus on in this paper. It ensures that a proximity measure $\omega$ is continuous at least in every efficient solution $\bar{x} \in S$ of (CMOP). Hence, we can expect a proximity measure to have values close to zero locally around every efficient solution. This makes it suitable for applications, e.g. for candidate selection in evolutionary algorithms. It should be mentioned that convergence statements do appear in [10] and [16] as well. However, these statements do only hold for certain sequences $(x^i)_N$ or rely on stronger assumptions, e.g. convexity of the objective and constraint functions. Whenever a function $\omega$ is a proximity measure in the sense of Definition 3.1, then property (PM3) holds for any sequence $(x^i)_N$ (converging towards an optimal solution $\bar{x}$).

Besides the three properties from Definition 3.1, one should of course 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 information to use within computer algorithms. In [7] the authors presented an example for a naively defined candidate $\hat{\omega}$ for a proximity measure, which is based on the KKT conditions, to motivate their further examinations. For every $x \in S$ this is

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

We will now present the numerical example as used in [7]. It demonstrates that in general, this function is not continuous in every efficient solution $\bar{x} \in S$ of (CMOP).
and hence, does not satisfy property (PM3). 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 Consider the constrained multi-objective optimization problem

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

The set of efficient points for this problem is

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

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

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

Using (3.1), we know that the only efficient point in this sequence is \(x^0 = (0.2, 0)^\top\). With the Euclidean norm, we obtain

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

So there exists a sequence \((x^\alpha) \subseteq S\) and an efficient point \(x^0\) such that \(\lim_{\alpha \to 0} x^\alpha = x^0\) but for the function \(\hat{\omega}\) by (3.2) it is

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

so that property (PM3) is not satisfied, and \(\hat{\omega}\) is no proximity measure.

In Figure 1, it can be seen that there is a discontinuity for \(\hat{\omega}\) 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 \( \alpha \to 0 \). Hence, the closer we get to the efficient point \( x^0 \), the higher gets the function value, 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 ideas from the single-objective case. This new measure will not be using any scalarization approach. Finally, another new proximity measure with simpler structure will be introduced in Section 3.3. That proximity measure is especially useful for computer algorithms as it is a lot easier to compute.

3.1 Single-Objective Case

For \( m = 1 \), the problem (CMOP) turns into a constrained single-objective optimization problem. This will be denoted by (CSOP). As already mentioned in the introduction, in general it is a hard task to compute exact KKT points and therefore different relaxations were presented in the literature. One of these relaxations are the so-called modified \( \varepsilon \)-KKT points. This concept was introduced in [10]. Those modified \( \varepsilon \)-KKT points can be seen as a relaxation of KKT points where a small deviation concerning the KKT conditions is allowed. As in the paper mentioned above the objective and constraint functions were not assumed to be differentiable but only Lipschitz continuous, the definitions relied on such concepts as 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 \( \hat{x} \in \mathbb{R}^n, \lambda \in \mathbb{R}^p_+ \) with

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

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

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

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

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

**Definition 3.4** A function \( \hat{\omega} : \mathbb{R}^n \to \mathbb{R} \) based on modified \( \varepsilon \)-KKT points is given as

\[
\hat{\omega}(x) := \min \left\{ \varepsilon \in \mathbb{R}_+ : \exists \lambda \in \mathbb{R}^p_+, \exists \hat{x} \in \mathbb{R}^n : \| \hat{x} - x \| \leq \sqrt{\varepsilon}, \left\| \nabla f(\hat{x}) + \sum_{j=1}^{p} \lambda_j \nabla g_j(\hat{x}) \right\| \leq \sqrt{\varepsilon}, \sum_{j=1}^{p} \lambda_j g_j(x) \geq -\varepsilon, g_j(x) \leq \varepsilon \text{ for all } j \in [p] \right\}
\]
for all \( x \in \mathbb{R}^n \).

It should be mentioned that the proximity measure in its formulation from [10] is only declared on the feasible set \( S \). To match our Definition 3.1 of proximity measures, we just added the restriction \( g_j(x) \leq \varepsilon \) for all \( j \in [p] \). 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.** Just by its 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. \( \square \)

As already mentioned, our focus in this paper is on property \((PM3)\) which is the strongest of all three properties for a proximity measure. 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 \( (x^n)_{n} \subseteq \mathbb{R}^n \) a sequence of points with \( \lim_{n \to \infty} x^n = \bar{x} \). Further, let \( (x^p)_{k \in \mathbb{N}} \) be a subsequence of \( (x^n)_{n} \). We will now construct a subsequence of \( (x^p)_{n} \) and show that \( \tilde{\omega} \) converges to 0 on this subsequence.

As \( \bar{x} \) is a minimal solution for (CSOP), we know from Theorem 2.2 that there exists a \( \bar{\lambda} \in \mathbb{R}^p_+ \) such that \((\bar{x}, \bar{\lambda})\) is a KKT point. Hence, \((KKT1), (KKT2),\) and \((KKT4)\) hold which leads to

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

(3.3)

Now let \( \varepsilon > 0 \). All constraint 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 a \( \delta^1_{\varepsilon} > 0 \) such that for every \( x \in \mathbb{R}^n \) with \( \|x - \bar{x}\| \leq \delta^1_{\varepsilon} \) it holds

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

(3.4)

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

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

(3.5)

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

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

(3.6)

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

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

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

(3.7)

\[
g(x) \leq \varepsilon \quad \text{for all } j \in [p].
\]
Moreover, there exists a $p_\varepsilon \in \mathbb{N}$ 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 all restrictions of the optimization problem for $\tilde{\omega}(x^p)$ as given in Definition 3.4 are satisfied with $\tilde{x}^p = x^p$ and $\lambda = \bar{\lambda}$. This implies that $\tilde{\omega}(x^p) \leq \varepsilon$.

Now let $(\varepsilon_n)_n \subseteq \text{int}(\mathbb{R}_+)$ be a monotonically decreasing sequence with

$$\lim_{n \to \infty} \varepsilon_n = 0.$$ 

Then for all $n \in \mathbb{N}$ there exists a $p_n = p_{\varepsilon_n} \in \mathbb{N}$ such that

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

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

$$0 \leq \tilde{\omega}(x^{p_n}) \leq \varepsilon_n$$

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 $(\tilde{\omega}(x^{i}))_{k \in \mathbb{N}}$ of $(\tilde{\omega}(x^i))_n$ there exists a subsequence $(\tilde{\omega}(x^{i_k}))_{k \in \mathbb{N}}$ (described by $(\tilde{\omega}(x^{p_n}))_{n \in \mathbb{N}}$ above) with

$$\lim_{l \to \infty} \tilde{\omega}(x^{i_k}) = 0 = \tilde{\omega}(\bar{x}). \quad (3.8)$$

Finally, we use that in case we have a sequence $(\omega^i)_n \subseteq \mathbb{R}$ and an $\bar{\omega} \in \mathbb{R}$ such that for all subsequences $(\omega^{i_k})_{k \in \mathbb{N}}$ there exists a subsubsequence $(\omega^{i_{k_l}})_{l \in \mathbb{N}}$ with

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

then the original sequence $(\omega^i)_n$ converges to $\bar{\omega}$ as well. Together with (3.8) this implies

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

$\square$

### 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 ideas from the single-objective case from Section 3.1. While no proximity measures were presented in [16], the authors there did generalize the concept of modified $\varepsilon$-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

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

# 9
(iii) \( \sum_{j=1}^{p} \lambda_j g_j(x) \geq -\epsilon, \)

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

then \( x \) is called a modified \( \varepsilon \)-KKT point for \((\text{CMOP})\).

Once again, the subgradients from the original definition are just gradients due to the assumption that all objective and constraint functions are continuously differentiable. In [16, Theorem 3.4] it was shown that for a sequence \( (x^i) \) of modified \( \varepsilon \)-KKT points with \( \varepsilon \) 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, this result is not necessarily usable for such applications. A more helpful result when having evolutionary algorithms in mind is presented in [9]. There, a relation between so-called weakly \( \varepsilon \)-efficient solutions and modified \( \varepsilon \)-KKT points was given. For \( \varepsilon > 0 \), a feasible point \( \bar{x} \in S \) is called weakly \( \varepsilon \)-efficient for \((\text{CMOP})\) with respect to \( d \in \text{int}(\mathbb{R}_+^m) \), \( \|d\| = 1 \) if there is no \( x \in S \) with \( f(x) + \varepsilon d < f(\bar{x}) \).

**Theorem 3.8** ([9, Theorem 3.7]) Consider \((\text{CMOP})\) with convex functions \( f_i, g_j \) for all \( i \in [m], j \in [p] \) and let Slater's constraint qualification be satisfied, i.e. there exists \( x^* \in S \) with \( g(x^*) < 0 \). Then every weakly \( \varepsilon \)-efficient point for \((\text{CMOP})\) with respect to \( d \in \text{int}(\mathbb{R}_+^m) \) is also a modified \( \varepsilon \)-KKT point.

Weakly \( \varepsilon \)-efficient points \( \bar{x} \in S \) have images \( f(\bar{x}) \) which are close to the image of the set of all weakly efficient solutions for \((\text{CMOP})\). By Theorem 3.8, a necessary condition for this '\( \varepsilon \)-closeness', i.e. for \( \bar{x} \) to be weakly \( \varepsilon \)-efficient, is that \( \bar{x} \) is a modified \( \varepsilon \)-KKT point. This can be used for instance as a selection 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]. The downside of this result is that it prerequires convexity of the objective and constraint functions which is quite restrictive. We now introduce a new proximity measure following the idea which was used in Definition 3.4 for the single-objective case.

**Definition 3.9** A function \( \omega: \mathbb{R}^n \to \mathbb{R} \) based on modified \( \varepsilon \)-KKT points is given as

\[
\omega(x) := \min_{\varepsilon \in \mathbb{R}_+} \left\{ \begin{array}{ll}
\exists \eta \in \mathbb{R}_+^m, \exists \lambda \in \mathbb{R}_+^p, \exists \hat{x} \in \mathbb{R}^n: & \|\hat{x} - x\| \leq \sqrt{\varepsilon}, \\
\sum_{i=1}^{m} \eta_i \nabla f_i(\hat{x}) + \sum_{j=1}^{p} \lambda_j \nabla g_j(\hat{x}) & \leq \sqrt{\varepsilon}, \\
\sum_{j=1}^{p} \lambda_j g_j(x) & \geq -\epsilon, \sum_{i=1}^{m} \eta_i = 1, \\
g_j(x) & \leq \epsilon \text{ for all } j \in [p]
\end{array} \right\}
\]

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

We show that this function \( \omega \) is indeed a proximity measure. We first state that properties \([\text{PM1}]\) and \([\text{PM2}]\) hold.

**Lemma 3.10** The function \( \omega \) satisfies \([\text{PM1}]\) and \([\text{PM2}]\).
Proof. Just by its definition we have $\omega(x) \geq 0$ 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. Hence, (KKT1), (KKT2), (KKT4) and (KKT6) hold.

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

**Lemma 3.11** For any $x \in \mathbb{R}^n$ it is $\omega(x) = 0$ if and only if there exist $\eta \in \mathbb{R}^m_+$ and $\lambda \in \mathbb{R}^p_+$ such that $(x, \eta, \lambda)$ 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 $\eta \in \mathbb{R}^m_+, \lambda \in \mathbb{R}^p_+$ such that

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

i.e. $(x, \eta, \lambda)$ is a KKT point.

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

**Theorem 3.12** The function $\omega$ satisfies (PM3).

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

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

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

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

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

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

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

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

$$|g_j(x) - g_j(\bar{x})| \leq \varepsilon \text{ for all } j \in [p]. \quad (3.12)$$
Now define $\delta_\varepsilon := \min\{\delta_1^\varepsilon, \delta_2^\varepsilon, \delta_3^\varepsilon\} > 0$. Then for every $x \in \mathbb{R}^n$ with $\|x - \bar{x}\| \leq \delta_\varepsilon$ by (3.10), (3.11), (3.12) and (3.9) it is

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

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

$$\sum_{i=1}^m \bar{\eta}_i = 1,$$

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

(3.13)

Moreover, there exists a $p_\varepsilon \in \mathbb{N}$ 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 all restrictions of the optimization problem for $\omega(x^p)$ as given in Definition 3.9 are satisfied with $x^p = x^p, \eta = \bar{\eta}$, and $\lambda = \bar{\lambda}$. This implies that $\omega(x^p) \leq \varepsilon$.

Now by taking a sequence $(\varepsilon_n)_{n\in\mathbb{N}} \subseteq \text{int}(\mathbb{R}_+)$ 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 $(x^p)_{p\in\mathbb{N}} = (x^k)_{k\in\mathbb{N}}$ with

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

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 $\omega$ 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 no proximity measure. In particular, it was not continuous in every efficient solution. We will now reconsider the problem (P1) from Example 3.2 and demonstrate the advantages of $\omega$ 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 $(x^n) \subseteq S$.

The values for $\omega(x^n)$ are shown in Figure 2. Compared to the results in Figure 1 from Example 3.2, there is no longer a discontinuity at $\alpha = 0$. Moreover, for $\alpha < 0.6$ the function values $\omega(x^n)$ are now monotonically decreasing to zero.

### 3.3 An Easy to Compute Proximity Measure

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

Moreover, it is important to notice that the introduction of $\hat{x}$ is a result of the weaker assumptions used in [16] and [9] (as well as in [10] for the single-objective case). All
objective and constraint functions were not assumed to be continuously differentiable in those papers. Instead they were only assumed to be locally Lipschitz continuous. Hence, the gradients of those functions do not necessarily exist in all feasible points. This is why the authors made use of the generalized gradient as presented by Clarke in [4]. But, 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. By Rademacher’s theorem every Lipschitz continuous function is differentiable almost everywhere. As all objective functions \( f_i, i \in [m] \) and all constraint functions \( g_j, j \in [p] \) 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 objective and constraint 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 idea was very shortly mentioned in [7] for single-objective optimization but without any further examination. The new relaxation of KKT points for \( \text{(CMOP)} \) which we will introduce are the so-called simplified \( \varepsilon \)-KKT points.

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

\[
\begin{align*}
(i) & \quad \left\| \sum_{i=1}^{m} \eta_i \nabla f_i(x) + \sum_{j=1}^{p} \lambda_j \nabla g_j(x) \right\|_\infty \leq \varepsilon, \\
(ii) & \quad \sum_{j=1}^{p} \lambda_j g_j(x) \geq -\varepsilon, \\
(iii) & \quad \sum_{i=1}^{m} \eta_i = 1,
\end{align*}
\]

then \( x \) is called a simplified \( \varepsilon \)-KKT point for \( \text{(CMOP)} \).
On the one hand, the term ‘simplified’ is motivated by the simplification of Definition 3.7 by fixing \( \hat{x} = x \). On the other hand, it is motivated by the simpler computation of the corresponding error measure, which we will introduce shortly. Compared to Definition 3.7, we have not only removed \( \hat{x} \) from the definition but also replaced \( \sqrt{\varepsilon} \) by \( \varepsilon \) and fixed the norm to the maximum norm in (i). As a result, 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 \( \omega_s : \mathbb{R}^n \rightarrow \mathbb{R} \) based on simplified \( \varepsilon \)-KKT points by

\[
\omega_s(x) := \min \left\{ \varepsilon \in \mathbb{R}_+^+ : \exists \eta \in \mathbb{R}_+^m, \exists \lambda \in \mathbb{R}_+^p : \right.
\]

\[
\left. \left\| \sum_{i=1}^m \eta_i \nabla f_i(x) + \sum_{j=1}^p \lambda_j \nabla g_j(x) \right\|_{\infty} \leq \varepsilon, \quad \sum_{j=1}^p \lambda_j g_j(x) \geq -\varepsilon, \quad \sum_{i=1}^m \eta_i = 1, \quad g_j(x) \leq \varepsilon \text{ for all } j \in [p] \right\}
\]

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

It can again be shown that our new function \( \omega_s \) 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 \( \varepsilon \) 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 \( \omega_s \). We summarize these results in the following theorem.

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

However, not all results for modified \( \varepsilon \)-KKT points can be easily transferred for simplified \( \varepsilon \)-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 easily 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 \( \omega \) and \( \omega_s \), which we introduced in this paper for multi-objective optimization problems (CMOP), we focus on another application purpose. As mentioned early on, proximity measures can also
be used within evolutionary algorithms as a selection criterion, i.e. to find candidates for efficient solutions. This will be illustrated in this section. For this, we will use the proximity measure $\omega_s$. In the previous section it was already discussed why this proximity measure is well suited for numerical evaluation and use within computer algorithms: it can be calculated by solving a linear problem only.

For the following examples we have generated $k$ points distributed in the preimage space and then computed the value of the proximity measure $\omega_s$ 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 $C$ in this section. Moreover, the set of efficient solutions will be denoted by $E$. For visualization the default parabula colormap of MATLAB was used. Hence, dark blue corresponds to a value of $\omega_s$ close to 0. Then, as the value rises up, the color turns green and finally yellow for the highest value of $\omega_s$ that was reached within the generated discretization of the preimage space. The sets $C$ and $f(C)$ are illustrated by red triangles.

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

$$\min_{x \in \mathbb{R}^2} f(x) := \left( x_1^2 + x_2^2 \right) (x_1 - 5)^2 + (x_2 - 5)^2 \quad \text{(BK1)}$$

s.t. $x_1, x_2 \in [-5, 10]$.

The set of efficient solutions for this problem is

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

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 $|C|=21$ points lead to a value of $\omega_s$ lower or equal to $\alpha = 0.001$. In particular, for the approximation $C$ delivered by MATLAB it holds $C \subseteq E$. The result is shown in Figure 3a and Figure 3b.

![Figure 3](image-url)

(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 2.3. 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 $(x, \eta, \lambda)$ 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, $\omega_s$ is well suited for characterization of weakly efficient points for (BK1) and also as a termination criterion for computer algorithms.

**Test instance 2** In [19] Srinivas and Deb presented the following non-convex problem with convex but nonlinear constraint functions.

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

(SRN)

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

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

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. Looking at these figures, it may seem that the Pareto Front $f(\mathcal{E})$ is larger than the representation which is covered by the candidates the MATLAB implementation provides. This is not the case, as there are a lot of infeasible points within the set $X$. Therefore, in Figures 4c and 4d only feasible points and the corresponding image are shown. It is not surprising that all solution candidates are indeed feasible. 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 $\mathcal{E}$. One idea to change this could be to reduce the acceptance limit $\alpha$, but the result remains unchanged even for $\alpha = 1 \cdot 10^{-8}$. If then further decreasing $\alpha$, the set $\mathcal{C}$ gets smaller and some gaps appear in the preimage space (and as a consequence in the image space as well). But, in terms of quality the result remains the same as the set $\tilde{\mathcal{C}}$ 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 2.3.

Due to the specific structure of $\mathcal{E}$ for this test instance, another question is what the results look like if the preimage space is not discretized equally but using a random discretization. The results for this approach using 16641 points can be seen in Figure 5 for two different values of $\alpha$. 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 $\alpha$, the set $\mathcal{C}$ starts to get bigger. 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
Figure 4: Numerical results for (SRN)

is exactly what we could expect due to the continuity of $\omega_s$ in every efficient solution. This specific run shows that the tolerance $\alpha$ 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 $\alpha$ can result in solution candidates that are not close to the set of efficient solutions $\mathcal{E}$ at all.

**Test instance 3** This test instance by Osyczka and Kundu is taken from [18]. It has a larger number of $n = 6$ optimization variables and a larger number of constraints as
A characterization of the set \( X := \{0 \} \times [1, 5] \times [0, 6] \times [1, 5] \times [0, 10] \) is quite huge compared to \( E \), 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 \( E \subseteq \hat{X} \). A total of \( k = (16 + 1)^4 = 83521 \) points were generated equally distributed in \( \hat{X} \). The MATLAB implementation found 70 candidates with a value of \( \omega \) less or equal to \( \alpha = 0.001 \). The result can be seen in Figure 6a.

For a better characterization of the set \( E \) of all efficient points 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 is \( \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3 \subseteq \mathcal{E} \). Considering the (discretized) image \( f(S) \) in Figure 6a, \( \mathcal{E}_1 \) 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 \( f_1 \approx -120 \). Looking at Figure 6a it might seem that actually none of the elements within \( C \) is an efficient solution. But,
one should always keep in mind that 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 \( C \) does contain efficient and locally efficient solutions. In particular, out of the 69 points in \( C \) there are 17 within \( E_1 \), 17 within \( E_2 \) and 11 within \( E_3 \). In fact these are all of the 83521 generated points that belong to the sets \( E_1, E_2 \) and \( E_3 \). Also the point \( x^c := (0.625, 1.375, 1, 0, 1, 0) \in C \) is part of the set of all efficient points \( E \).

However, there are also 24 points which do not belong to \( E \). There is one outlier which can also be seen in Figure 6a. But there are also some locally efficient solutions which can be seen as an extension of \( E_3 \). In particular, these are

\[
\begin{align*}
C_1 & := \{(0, 2, \beta, 0, 1, 0) \in S \mid \beta \in \{3.75, 4.00, 4.25, 4.50, 4.75, 5.00\}\} \quad \text{and} \\
C_2 & := \left\{(0, 2, \beta, 0, 5, 0) \in S \mid \beta \in \left\{1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 3.25, 3.50, 3.75, 4.00, 4.25, 4.50, 4.75, 5.00\right\}\right\}.
\end{align*}
\]

In the image space, \( f(E_3) \) 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 \( C_2 \). The points in between belong to \( C_1 \).

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, \( \alpha \) 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, e.g. to choose $k = (32 + 1)^4 = 1185921$ points equally 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. But we also included some more points which are not approximately efficient solutions within $C$.

5 Conclusions

We presented two new proximity measures $\omega$ and $\omega_s$ for (CMOP) in the sense of Definition 3.1. The proximity measure $\omega$ is a generalization of the proximity measure presented in [10] for the single-objective case by using the results from [9] and [16]. One drawback of this approach is that $\omega$ can be hard to compute. Thus it is not really suited for use within computer algorithms. This is why we introduced the proximity measure $\omega_s$. Compared to $\omega$, 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 $\omega_s$ 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, $\omega_s$ is well suited for numerical applications. In particular, we have illustrated its capabilities as an additional criterion for candidate selection in evolutionary algorithms. It could now be argued that the computation of $\omega_s(x)$ relies on solving an optimization problem and hence, all the problems mentioned in the introduction, e.g. limited accuracy, are critical aspects as well. While such limitations should always be taken into account, the computation of $\omega_s$ relies on solving a linear optimization problem. For linear optimization problems, exact solvers such as SoPlex (see [13]) are available and can handle even numerically troublesome problems.

It should also be mentioned that at no point we had to assume $m \geq 2$ for the dimension of the image space. Hence, all results do still hold for single-objective optimization problems. Moreover, the case of unconstrained problems is also contained as a special case in the used setting. As a result, our new proximity measure $\omega_s$ can be used for a large class of optimization problems. In particular, its numerical benefits remain for those problem classes. That motivates to use $\omega_s$ as an additional criterion within the corresponding algorithms as well.

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

[12] 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.

[13] 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.

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

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

[16] P. Keskarwani, 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.

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

[18] 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.

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