

Gradient-Based Multiobjective Optimization with Uncertainties

Sebastian Peitz* and Michael Dellnitz*

*Department of Mathematics, Paderborn University, Warburger Str. 100, D-33098 Paderborn

April 6, 2017

Abstract

In this article we develop a gradient-based algorithm for the solution of multiobjective optimization problems with uncertainties. To this end, an additional condition is derived for the descent direction in order to account for inaccuracies in the gradients and then incorporated in a subdivision algorithm for the computation of global solutions to multiobjective optimization problems. Convergence to a superset of the Pareto set is proved and an upper bound for the maximal distance to the set of substationary points is given. Besides the applicability to problems with uncertainties, the algorithm is developed with the intention to use it in combination with model order reduction techniques in order to efficiently solve PDE-constrained multiobjective optimization problems.

1 Introduction

In many applications from industry and economy, one is interested in simultaneously optimizing several criteria. For example, in transportation one wants to reach a destination as fast as possible while minimizing the energy consumption. This example illustrates that in general, the different objectives contradict each other. Therefore, the task of computing the set of optimal compromises between the conflicting objectives, the so-called *Pareto set*, arises. This leads to a multiobjective optimization problem (MOP). Based on the knowledge of the Pareto set, a *decision maker* can use this information either for improved system design or for changing parameters during operation, as a reaction on external influences or changes in the system state itself.

Multiobjective optimization is an active area of research. Different methods exist to address MOPs, e.g. deterministic approaches [Mie99, Ehr05], where ideas from scalar optimization theory are extended to the multiobjective situation. In many cases, the resulting solution method involves solving multiple scalar optimization problems consecutively. Continuation methods make use of the fact that under certain smoothness assumptions, the Pareto set is a manifold that can be approximated by continuation methods known from dynamical systems theory [Hil01]. Another prominent

approach is based on evolutionary algorithms [CVL02], where the underlying idea is to evolve an entire set of solutions (population) during the optimization process. Set oriented methods provide an alternative deterministic approach to the solution of MOPs. Utilizing subdivision techniques (cf. [DSH05, Jah06, SWOBD13]), the desired Pareto set is approximated by a nested sequence of increasingly refined box coverings. In the latter approach, gradients are evaluated to determine a descent direction for all objectives.

Many solution approaches are gradient-free since the benefit of derivatives is less established than in single objective optimization [Bos12]. Exceptions are scalarization methods since the algorithms used in this context mainly stem from scalar optimization. Moreover, in so-called *memetic algorithms* [NCM12], evolutionary algorithms are combined with local search strategies, where gradients are utilized (see e.g. [LSCS10, SML⁺16, SASL16]). Finally, several authors also develop gradient-based methods for MOPs directly. In [FS00, SSW02, Des12], algorithms are developed where a single descent direction is computed in which all objectives decrease. In [Bos12], a method is presented by which the entire set of descent directions can be determined, an extension of Newton's method to MOPs with quadratic convergence is presented in [FDS09].

Many real world problems possess uncertainties for various reasons such as the ignorance of the exact underlying dynamical system or unknown material properties. The use of reduced order models in order to decrease the computational effort also introduces errors which can often be quantified [TV09]. A similar concept exists in the context of evolutionary computation, where *Surrogate-assisted evolutionary computation* (see e.g. [Jin11] for a survey) is often applied in situations where the exact model is too costly to solve. In this setting, the set of *almost Pareto optimal points* has to be computed (see also [Whi86], where the idea of ϵ efficiency was introduced). Many researchers are investigating problems related to uncertainty quantification and several authors have addressed multiobjective optimization problems with uncertainties. In [Hug01] and in [Tei01], probabilistic approaches to multiobjective optimization problems with uncertainties were derived independently. In [DG05, DMM05, BZ06, SM08], evolutionary algorithms were developed for problems with uncertain or noisy data in order to compute robust approximations of the Pareto set. In [SCTT08], stochastic search was used, cell mapping techniques were applied in [HSS13] and in [EW07], the weighted sum method was extended to uncertainties.

In this article, we present extensions to the gradient-free and gradient-based global subdivision algorithms for unconstrained MOPs developed in [DSH05] which take into account inexactness in both the function values and the gradients. The algorithms compute a set of solutions which is a superset of the Pareto set with an upper bound for the distance to the Pareto set. The remainder of the article is organized in the following manner. In Section 2, we give a short introduction to multiobjective optimization in general and gradient-based descent directions for MOPs. In Section 3, a descent direction for all objectives under inexact gradient information is developed and an upper bound for the distance to a Pareto optimal point is given. In Section 4, the subdivision algorithm presented in [DSH05] is extended to inexact function and gradient values before we present our results in Section 5 and draw a conclusion in Section 6.

2 Multiobjective Optimization

Consider the continuous, unconstrained multiobjective optimization problem

$$\min_{x \in \mathbb{R}^n} F(x) = \min_{x \in \mathbb{R}^n} \begin{pmatrix} f_1(x) \\ \vdots \\ f_k(x) \end{pmatrix}, \quad (\text{MOP})$$

where $F : \mathbb{R}^n \rightarrow \mathbb{R}^k$ is a vector valued objective function with continuously differentiable objective functions $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, k$. The space of the parameters x is called the *decision space* and the function F is a mapping to the k -dimensional *objective space*. In contrast to single objective optimization problems, there exists no total order of the objective function values in \mathbb{R}^k , $k \geq 2$ (unless the objectives are not conflicting). Therefore, the comparison of values is defined in the following way [Mie99]:

Definition 2.1. Let $v, w \in \mathbb{R}^k$. The vector v is less than w ($v <_p w$), if $v_i < w_i$ for all $i \in \{1, \dots, k\}$. The relation \leq_p is defined in an analogous way.

A consequence of the lack of a total order is that we cannot expect to find isolated optimal points. Instead, the solution of (MOP) is the set of optimal compromises, the so-called *Pareto set* named after Vilfredo Pareto:

Definition 2.2.

- (a) A point $x^* \in \mathbb{R}^n$ dominates a point $x \in \mathbb{R}^n$, if $F(x^*) \leq_p F(x)$ and $F(x^*) \neq F(x)$.
- (b) A point $x^* \in \mathbb{R}^n$ is called (globally) Pareto optimal if there exists no point $x \in \mathbb{R}^n$ dominating x^* . The image $F(x^*)$ of a (globally) Pareto optimal point x^* is called a (globally) Pareto optimal value.
- (c) The set of non-dominated points is called the Pareto set \mathcal{P}_S , its image the Pareto front \mathcal{P}_F .

Consequently, for each solution that is contained in the Pareto set, one can only improve one objective by accepting a trade-off in at least one other objective. That is, roughly speaking, in a two-dimensional problem, we are interested in finding the “lower left” boundary of the reachable set in objective space (cf. Figure 1 (b)). A more detailed introduction to multiobjective optimization can be found in e.g. [Mie99, Ehr05].

Similar to single objective optimization, a necessary condition for optimality is based on the gradients of the objective functions. In the multiobjective situation, the corresponding Karush-Kuhn-Tucker (KKT) condition is as follows:

Theorem 2.3 ([KT51]). Let x^* be a Pareto point of (MOP). Then, there exist nonnegative scalars $\alpha_1, \dots, \alpha_k \geq 0$ such that

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

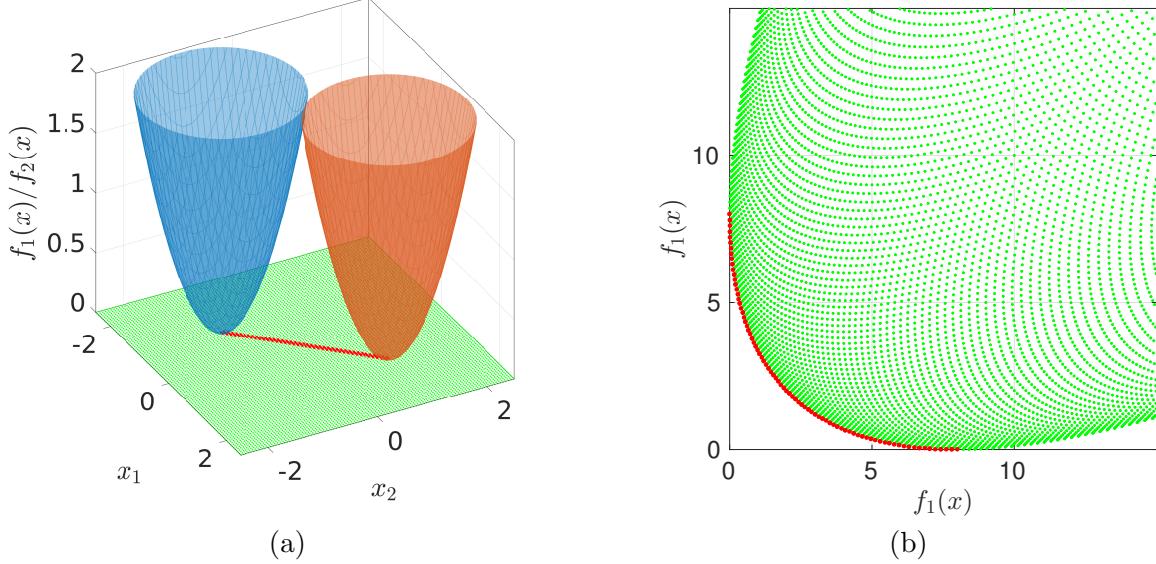


Figure 1: The red lines depict the Pareto set (a) and Pareto front (b) of an exemplary multi-objective optimization problem (two paraboloids) of the form (MOP) with $n = 2$ and $k = 2$.

Observe that (1) is only a necessary condition for a point x^* to be a Pareto point and the set of points satisfying (1) is called the set of *substationary points* $\mathcal{P}_{S,\text{sub}}$. Under additional smoothness assumptions (see [Hil01]) $\mathcal{P}_{S,\text{sub}}$ is locally a $k - 1$ -dimensional manifold. Obviously $\mathcal{P}_{S,\text{sub}}$ is a superset of the Pareto set \mathcal{P}_S .

If $x \notin \mathcal{P}_{S,\text{sub}}$ then the KKT conditions can be utilized in order to identify a descent direction $q(x)$ for which all objectives are non-increasing, i.e.:

$$-\nabla f_i(x) \cdot q(x) \geq 0, \quad i = 1, \dots, k. \quad (2)$$

One way to compute a descent direction satisfying (2) is to solve the following auxiliary optimization problem [SSW02]:

$$\min_{\alpha \in \mathbb{R}^k} \left\{ \left\| \sum_{i=1}^k \alpha_i \nabla f_i(x) \right\|_2^2 \mid \alpha_i \geq 0, \quad i = 1, \dots, k, \sum_{i=1}^k \alpha_i = 1 \right\}. \quad (\text{QOP})$$

Using (QOP), we obtain the following result:

Theorem 2.4 ([SSW02]). Define $q : \mathbb{R}^n \rightarrow \mathbb{R}^n$ by

$$q(x) = - \sum_{i=1}^k \widehat{\alpha}_i \nabla f_i(x), \quad (3)$$

where $\hat{\alpha}$ is a solution of (QOP). Then either $q(x) = 0$ and x satisfies (1), or $q(x)$ is a descent direction for all objectives $f_1(x), \dots, f_k(x)$ in x . Moreover, $q(x)$ is locally Lipschitz continuous.

Remark 2.5. As in the classical case of scalar optimization there exist in general infinitely many valid descent directions $q(x)$ and using the result from Theorem 2.4 yields one particular direction. As already stated in the introduction, there are alternative ways to compute such a direction, see e.g. [FS00] for the computation of a single direction or [Bos12], where the entire set of descent directions is determined.

3 Multiobjective Optimization with Inexact Gradients

Suppose now that we only have approximations $\tilde{f}_i(x), \nabla \tilde{f}_i(x)$ of the objectives $f_i(x)$ and their gradients $\nabla f_i(x)$, $i = 1, \dots, k$, respectively. To be more precise we assume that

$$\tilde{f}_i(x) = f_i(x) + \bar{\xi}_i, \quad \|\tilde{f}_i(x) - f_i(x)\|_2 = \|\bar{\xi}_i\|_2 \leq \xi_i, \quad (4)$$

$$\nabla \tilde{f}_i(x) = \nabla f_i(x) + \bar{\epsilon}_i, \quad \|\nabla \tilde{f}_i(x) - \nabla f_i(x)\|_2 = \|\bar{\epsilon}_i\|_2 \leq \epsilon_i, \quad (5)$$

where the upper bounds ξ_i, ϵ_i are given. In the following, when computing descent directions we will assume that

$$\epsilon_i \leq \|\nabla f_i(x)\|_2 \quad \text{for all } x,$$

since otherwise, we are already in the vicinity of the set of stationary points as will be shown in Lemma 3.2. Using (5) we can derive an upper bound for the angle between the exact and the inexact gradient by elementary geometrical considerations (cf. Figure 2 (a)):

$$\sphericalangle(\nabla f_i(x), \nabla \tilde{f}_i(x)) = \arcsin\left(\frac{\|\bar{\epsilon}_i\|_2}{\|\nabla f_i(x)\|_2}\right) \leq \arcsin\left(\frac{\epsilon_i}{\|\nabla f_i(x)\|_2}\right) =: \varphi_i. \quad (6)$$

Here we denote the largest possible angle (i.e. the “worst case”) by φ_i . Based on this angle, one can easily define a condition for the inexact gradients such that the exact gradients could satisfy (1) for the first time. This is precisely the case if each inexact gradient deviates from the hyperplane defined by (1) at most by φ_i , see Figure 2 (b). This motivates the definition of an *inexact descent direction*:

Definition 3.1. A direction analog to (3) but based on inexact gradients, i.e.

$$q_u(x) = - \sum_{i=1}^k \hat{\alpha}_i \nabla \tilde{f}_i(x) \quad \text{with } \hat{\alpha}_i \geq 0 \text{ for } i = 1, \dots, k \text{ and } \sum_{i=1}^k \hat{\alpha}_i = 1, \quad (7)$$

is called inexact descent direction.

We can prove an upper bound for the norm of $q_u(x^*)$ when x^* satisfies the KKT condition of the exact problem:

Lemma 3.2. Consider the multiobjective optimization problem (MOP) with inexact gradient information according to (5). Let x^* be a point satisfying the KKT conditions (1) for the exact problem. Then, the inexact descent direction $q_u(x^*)$ is bounded by

$$\|q_u(x^*)\|_2 \leq \|\epsilon\|_\infty.$$

Proof. Since x^* satisfies (1), we have $\sum_{i=1}^k \hat{\alpha}_i \nabla f_i(x^*) = 0$. Consequently,

$$\sum_{i=1}^k \hat{\alpha}_i \nabla \tilde{f}_i(x) = \sum_{i=1}^k \hat{\alpha}_i (\nabla f_i(x) + \bar{\epsilon}_i) = \sum_{i=1}^k \hat{\alpha}_i \bar{\epsilon}_i,$$

and thus,

$$\begin{aligned} \left\| \sum_{i=1}^k \hat{\alpha}_i \nabla \tilde{f}_i(x) \right\|_2 &= \left\| \sum_{i=1}^k \hat{\alpha}_i \bar{\epsilon}_i \right\|_2 \leq \sum_{i=1}^k \hat{\alpha}_i \|\bar{\epsilon}_i\|_2 \leq \sum_{i=1}^k \hat{\alpha}_i \epsilon_i \\ &\leq \sum_{i=1}^k \hat{\alpha}_i \|\epsilon\|_\infty = \|\epsilon\|_\infty \sum_{i=1}^k \hat{\alpha}_i = \|\epsilon\|_\infty. \end{aligned}$$

□

A consequence of Lemma 3.2 is that in the presence of inexactness, we cannot compute the set of points satisfying (1) exactly. At best, we can compute the set of points determined by $\|q_u(x)\|_2 \leq \|\epsilon\|_\infty$. In the following section we will derive a criterion for the inexact descent direction $q_u(x)$ which guarantees that it is also a descent direction for the exact problem when x is sufficiently far away from the set of substationaly points.

3.1 Descent Directions in the Presence of Inexactness

The set of valid descent directions for (MOP) is a cone defined by the intersection of all half-spaces orthogonal to the gradients $\nabla f_1(x), \dots, \nabla f_k(x)$ (cf. Figure 3 (a)), i.e. it consists of all directions $q(x)$ satisfying

$$\sphericalangle(q(x), -\nabla f_i(x)) \leq \frac{\pi}{2}, \quad i = 1, \dots, k. \quad (8)$$

This fact is well known from scalar optimization theory, see e.g. [NW06]. Observe that here we allow directions also to be valid for which all objectives are at least non-decreasing. In terms of the angle $\gamma_i \in [0, \pi/2]$ between the descent direction $q(x)$ and the hyperplane orthogonal to the gradient of the i^{th} objective, this can be expressed as

$$\gamma_i = \frac{\pi}{2} - \arccos \left(\frac{q(x) \cdot (-\nabla f_i(x))}{\|q(x)\|_2 \cdot \|\nabla f_i(x)\|_2} \right) \geq 0, \quad i = 1, \dots, k. \quad (9)$$

Remark 3.3. Note that γ_i is equal to $\pi/2$ when $q(x) = \nabla f_i(x)$ and approaches zero when a point x approaches the set of substationaly points, i.e.

$$\lim_{\|q(x)\|_2 \rightarrow 0} \gamma_i \rightarrow 0 \quad \text{for } i = 1, \dots, k.$$

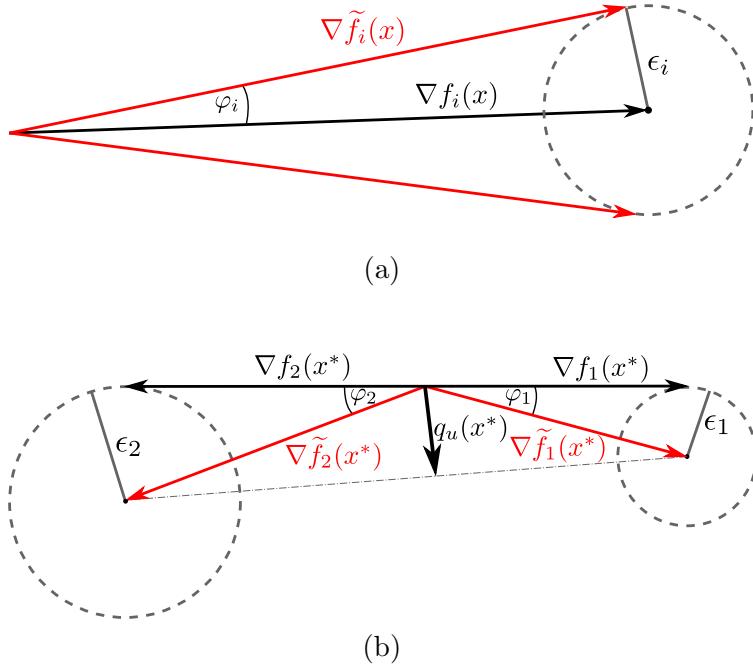


Figure 2: (a) Maximal angle between the exact and the inexact gradient in dependence on the error ϵ_i . (b) Maximal angle between the inexact gradients in the situation where x^* satisfies (1) ($\pi - (\varphi_1 + \varphi_2)$ in the 2D case).

We call the set of all descent directions satisfying (8) the *exact cone* \mathcal{Q} . If we consider inexact gradients according to (5) then \mathcal{Q} is reduced to the *inexact cone* \mathcal{Q}_u by the respective upper bounds for the angular deviation φ_i :

$$\gamma_i \geq \varphi_i \geq 0, \quad i = 1, \dots, k. \quad (10)$$

This means that we require the angle between the descent direction $q(x)$ and the i^{th} hyperplane to be at least as large as the maximum deviation between the exact and the inexact gradient (cf. Figure 3 (b)). Thus, if an inexact descent direction $q_u(x)$ satisfies (10), then it is also a descent direction for the exact problem.

We would like to derive an algorithm by which we can determine an inexact descent direction $q_u(x)$ in such a way that it is also valid for the exact problem. To this end, we derive an additional criterion which is equivalent to (10). Concretely, we prove the following lemma:

Lemma 3.4. *Consider the multiobjective optimization problem (MOP) with inexact gradient information according to (5). Let $q_u(x)$ be an inexact descent direction according to Definition 3.1.*

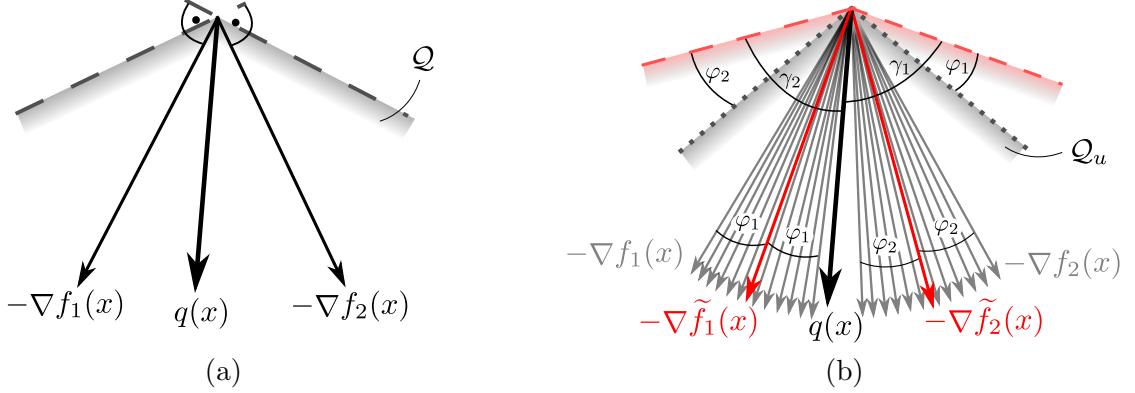


Figure 3: (a) Set of valid descent directions (the exact cone \mathcal{Q} bounded by the dashed lines) determined by the intersection of half-spaces defined by the negative gradients. (b) Reduction of the set of valid descent directions in dependence on the errors ϵ_i (the inexact cone \mathcal{Q}_u bounded by the dotted lines). The gray vectors represent the set of possible values of the exact gradients $\nabla f_i(x)$ and the inexact cone \mathcal{Q}_u is defined by the “most aligned” (here the uppermost) realizations of $\nabla f_i(x)$.

We assume $\|q_u(x)\|_2 \neq 0$, $\|\nabla \tilde{f}_i(x)\|_2 \neq 0$, $i = 1, \dots, k$. Then (10) is equivalent to

$$\hat{\alpha}_i \geq \frac{1}{\|\nabla \tilde{f}_i(x)\|_2^2} \left(\|q_u(x)\|_2 \epsilon_i - \sum_{\substack{j=1 \\ j \neq i}}^k \hat{\alpha}_j (\nabla \tilde{f}_j(x) \cdot \nabla \tilde{f}_i(x)) \right), \quad i = 1, \dots, k. \quad (11)$$

In particular, $q_u(x)$ is a descent direction for all objective functions $f_i(x)$ if (11) is satisfied.

Proof. Inserting the expression for γ_i in (9) into (10) yields

$$\frac{q_u(x) \cdot (-\nabla \tilde{f}_i(x))}{\|q_u(x)\|_2 \cdot \|\nabla \tilde{f}_i(x)\|_2} \geq \cos \left(\frac{\pi}{2} - \varphi_i \right) = \sin(\varphi_i) = \frac{\epsilon_i}{\|\nabla \tilde{f}_i(x)\|_2}. \quad (12)$$

Using the definition of $q_u(x)$ this is equivalent to

$$\begin{aligned}
\frac{\left(\sum_{j=1}^k \hat{\alpha}_j \nabla \tilde{f}_j(x)\right) \cdot \nabla \tilde{f}_i(x)}{\|q_u(x)\|_2 \|\nabla \tilde{f}_i(x)\|_2} &= \frac{\hat{\alpha}_i \|\nabla \tilde{f}_i(x)\|_2^2 + \sum_{j=1, j \neq i}^k \hat{\alpha}_j \nabla \tilde{f}_j(x) \cdot \nabla \tilde{f}_i(x)}{\|q_u(x)\|_2 \|\nabla \tilde{f}_i(x)\|_2} \\
&= \frac{\|\nabla \tilde{f}_i(x)\|_2}{\|q_u(x)\|_2} \hat{\alpha}_i + \frac{\sum_{j=1, j \neq i}^k \hat{\alpha}_j \nabla \tilde{f}_j(x) \cdot \nabla \tilde{f}_i(x)}{\|q_u(x)\|_2 \|\nabla \tilde{f}_i(x)\|_2} \\
&\geq \frac{\epsilon_i}{\|\nabla \tilde{f}_i(x)\|_2} \\
\iff \hat{\alpha}_i &\geq \frac{1}{\|\nabla \tilde{f}_i(x)\|_2^2} \left(\|q_u(x)\|_2 \epsilon_i - \sum_{\substack{j=1 \\ j \neq i}}^k \hat{\alpha}_j (\nabla \tilde{f}_j(x) \cdot \nabla \tilde{f}_i(x)) \right).
\end{aligned}$$

□

Remark 3.5. By setting $\epsilon_i = 0$ (i.e. $\tilde{f}_i(x) = f_i(x)$ for $i = 1, \dots, k$) in (11) and performing some elemental manipulations, we again obtain the condition (2) for an exact descent direction:

$$\begin{aligned}
\hat{\alpha}_i &\geq \frac{1}{\|\nabla f_i(x)\|_2^2} \left(- \sum_{\substack{j=1 \\ j \neq i}}^k \hat{\alpha}_j (\nabla f_j(x) \cdot \nabla f_i(x)) \right) \\
\Leftrightarrow \hat{\alpha}_i (\nabla f_i(x) \cdot \nabla f_i(x)) &\geq \left(- \sum_{\substack{j=1 \\ j \neq i}}^k \hat{\alpha}_j (\nabla f_j(x) \cdot \nabla f_i(x)) \right) \\
\Leftrightarrow -\nabla f_i(x) \cdot q(x) &\geq 0.
\end{aligned}$$

The condition (11) can be interpreted as a lower bound for the “impact” of a particular gradient on the descent direction induced by $\hat{\alpha}_i$. The larger the error ϵ_i , the higher the impact of the corresponding gradient needs to be in order to increase the angle between the descent direction and the hyperplane normal to the gradient. The closer a point x is to the Pareto front (i.e. for small values of $\|q(x)\|_2$), the more confined the region of possible descent directions becomes. Hence, the inaccuracies gain influence until it is no longer possible to guarantee the existence of a descent direction for every objective. This is the case when the sum over the lower bounds from (11) exceeds one as shown in part (a) of the following result:

Theorem 3.6. Consider the multiobjective optimization problem (MOP) and suppose that the assumptions in Lemma 3.4 hold. Let

$$\hat{\alpha}_{min,i} = \frac{1}{\|\nabla \tilde{f}_i(x)\|_2^2} \left(\|q_u(x)\|_2 \epsilon_i - \sum_{\substack{j=1 \\ j \neq i}}^k \hat{\alpha}_j (\nabla \tilde{f}_j(x) \cdot \nabla \tilde{f}_i(x)) \right), \quad i = 1, \dots, k.$$

Then the following statements are valid:

(a) If $\sum_{i=1}^k \widehat{\alpha}_{min,i} > 1$ then $\mathcal{Q}_u = \emptyset$ (see Figure 3 (b)), and therefore it cannot be guaranteed that there is a descent direction for all objective functions.

(b) All points x with $\sum_{i=1}^k \widehat{\alpha}_{min,i} = 1$ are contained in the set

$$\mathcal{P}_{S,\epsilon} = \left\{ x \in \mathbb{R}^n \mid \left\| \sum_{i=1}^k \widehat{\alpha}_i \nabla f_i(x) \right\|_2 \leq 2\|\epsilon\|_\infty \right\}. \quad (13)$$

Proof. For part (a), suppose that we have a descent direction $q_u(x)$ for which $\sum_{i=1}^k \widehat{\alpha}_{min,i} > 1$. Then by (11)

$$\begin{aligned} & \sum_{i=1}^k \left(\frac{\|q_u(x)\|_2}{\|\nabla \tilde{f}_i(x)\|_2^2} \epsilon_i - \frac{\sum_{j=1, j \neq i}^k \widehat{\alpha}_j \nabla \tilde{f}_j(x) \cdot \nabla \tilde{f}_i(x)}{\|\nabla \tilde{f}_i(x)\|_2^2} \right) > 1 \\ \Leftrightarrow & \sum_{i=1}^k \left(\frac{\|q_u(x)\|_2}{\|\nabla \tilde{f}_i(x)\|_2^2} \epsilon_i - \frac{(-q_u(x) - \widehat{\alpha}_i \nabla \tilde{f}_i(x)) \cdot \nabla \tilde{f}_i(x)}{\|\nabla \tilde{f}_i(x)\|_2^2} \right) > 1 \\ \Leftrightarrow & \sum_{i=1}^k \left(\frac{\|q_u(x)\|_2}{\|\nabla \tilde{f}_i(x)\|_2^2} \epsilon_i - \frac{-q_u(x) \cdot \nabla \tilde{f}_i(x)}{\|\nabla \tilde{f}_i(x)\|_2^2} + \widehat{\alpha}_i \right) > 1 \\ \Leftrightarrow & \sum_{i=1}^k \left(\frac{\|q_u(x)\|_2}{\|\nabla \tilde{f}_i(x)\|_2^2} \epsilon_i - \frac{-q_u(x) \cdot \nabla \tilde{f}_i(x)}{\|\nabla \tilde{f}_i(x)\|_2^2} \right) > 1 - \sum_{i=1}^k \widehat{\alpha}_i = 0 \\ \Leftrightarrow & \sum_{i=1}^k \left(\frac{\epsilon_i}{\|\nabla \tilde{f}_i(x)\|_2} - \frac{-q_u(x) \cdot \nabla \tilde{f}_i(x)}{\|q_u(x)\|_2 \cdot \|\nabla \tilde{f}_i(x)\|_2} \right) > 0 \\ \Leftrightarrow & \sum_{i=1}^k \sin \varphi_i > \sum_{i=1}^k \sin \gamma_i. \end{aligned}$$

Since $\varphi_i, \gamma_i \in [0, \pi/2]$ for $i = 1, \dots, k$, it follows that $\varphi_i > \gamma_i$ for at least one $i \in \{1, \dots, k\}$. This is a contradiction to (10) yielding $\mathcal{Q}_u = \emptyset$.

For part (b), we repeat the calculation from part (a) with the distinction that $\sum_{i=1}^k \widehat{\alpha}_{min,i} = 1$, and obtain $\sum_{i=1}^k \sin \varphi_i = \sum_{i=1}^k \sin \gamma_i$. This implies $\varphi_i = \gamma_i$ for $i = 1, \dots, k$, i.e. the set of descent directions is reduced to a single valid direction. This is a situation similar to the one described in Lemma 3.2. Having a single valid descent direction results in the fact that there is now a possible realization of the gradients $\nabla f_i(x)$ such that each one is orthogonal to $q_u(x)$. In this situation, x would satisfy (1) and hence, $\|q_u(x)\|_2 \leq \|\epsilon\|_\infty$ which leads to

$$\begin{aligned} \left\| \sum_{i=1}^k \widehat{\alpha}_i \nabla f_i(x) \right\|_2 &= \left\| \sum_{i=1}^k \widehat{\alpha}_i (\nabla \tilde{f}_i(x) + (-\bar{\epsilon}_i)) \right\|_2 = \left\| q_u(x) + \sum_{i=1}^k \widehat{\alpha}_i (-\bar{\epsilon}_i) \right\|_2 \\ &\leq \|q_u(x)\|_2 + \left\| \sum_{i=1}^k \widehat{\alpha}_i \bar{\epsilon}_i \right\|_2 \leq \|q_u(x)\|_2 + \left\| \sum_{i=1}^k \bar{\epsilon}_i \right\|_2 \leq 2\|\epsilon\|_\infty. \end{aligned}$$

□

Corollary 3.7. *If $\sum_{i=1}^k \hat{\alpha}_{min,i} > 1$, the angle of the cone spanned by the exact gradients $\nabla f_i(x)$ is larger than for $\sum_{i=1}^k \hat{\alpha}_{min,i} \leq 1$ (cf. Figure 4 (b)). Moreover, if $\|q_u(x)\|_2$ is monotonically decreasing for decreasing distances of x to the set of stationary points $\mathcal{P}_{S,sub}$, then the result from Theorem 3.6 (b) holds for $\sum_{i=1}^k \hat{\alpha}_{min,i} \geq 1$.*

The results from Theorem 3.6 are visualized in Figure 4. In the situation where the inexactness in the gradients $\nabla \tilde{f}_i(x)$ permits the exact gradients $\nabla f_i(x)$ to satisfy the KKT condition (1), $\mathcal{Q}_u = \emptyset$ and a descent direction can no longer be computed.

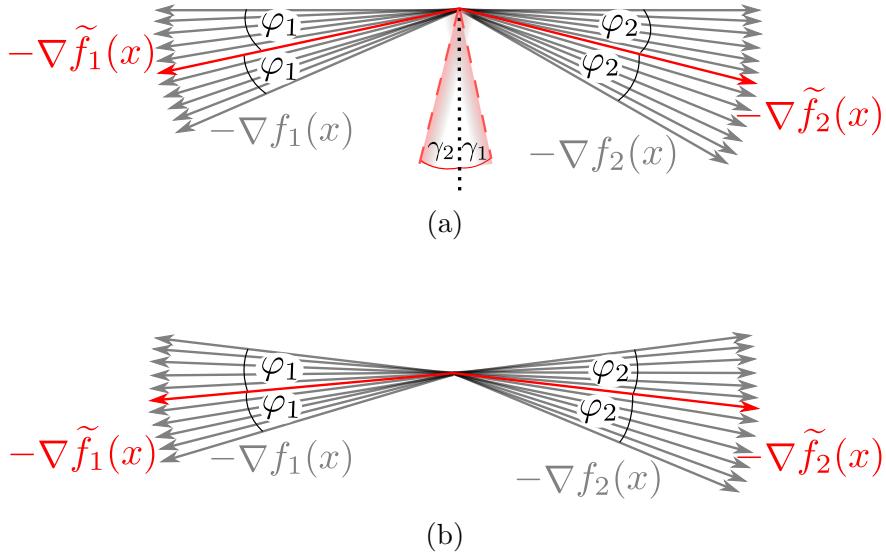


Figure 4: (a) Situation when a valid descent direction can no longer be computed. (b) Situation where $\sum_{i=1}^k \hat{\alpha}_{min,i} > 1$.

In order to numerically compute a valid descent direction, one can compute a direction by solving (QOP) and consequently verifying that $q_u(x)$ is indeed descending for all objectives by checking condition (11). If this is not the case, we can solve (QOP) again with adjusted lower bounds for $\hat{\alpha}$ (cf. Algorithm 1). Alternatively, one can compute the entire set of descent directions [Bos12] and chose a direction from this set which satisfies (10) or (11), respectively. In the examples we investigated, we observed that $q_u(x)$ is equal to $q(x)$ for the majority of points x . This is likely due to the fact that by solving (QOP), we obtain a steepest descent like direction which very often is relatively far away from the descent cone's boundaries. Close to the set of substationary points, we observe updates of the lower bounds $\hat{\alpha}_{min,i}$, but in many cases, the line search strategy directly leads to points where $\mathcal{Q}_u = \emptyset$ without requiring adjusted bounds.

The obtained descent direction can now be utilized to efficiently compute a point which is approximately Pareto optimal, utilizing the advantages of gradient-based methods. In order to compute the entire Pareto set for MOPs with inexact gradient information, we will combine the above result with the algorithm presented in [DSH05] in the next section.

Algorithm 1 (Descent direction for inexact gradients)

Require: Inexact gradients $\nabla \tilde{f}_i(x)$, error bounds ϵ_i and lower bounds $\hat{\alpha}_{min,i} = 0$, $i = 1, \dots, k$

```

1: loop
2:   Compute  $\hat{\alpha}_i$  by solving (QOP) with  $\hat{\alpha}_i \in [\hat{\alpha}_{min,i}, 1]$ ,  $i = 1, \dots, k$ 
3:   Evaluate the condition (11) and update  $\hat{\alpha}_{min,i}$ 
4:   if  $\sum_{i=1}^k \hat{\alpha}_{min,i} \geq 1$  then
5:      $q_u(x) = 0$  ( $Q_u = \emptyset$ )
6:     STOP
7:   else if  $\hat{\alpha}_i \geq \hat{\alpha}_{min,i}$ ,  $i = 1, \dots, k$  then
8:      $q_u(x) = -\sum_{i=1}^k \hat{\alpha}_i \nabla \tilde{f}_i(x)$ 
9:     STOP
10:  end if
11: end loop

```

4 Subdivision Algorithm

Using the result of Theorem 3.6 we can construct a global subdivision algorithm which computes a nested sequence of increasingly refined box coverings of the entire set of *almost substationary points* $\mathcal{P}_{S,\epsilon}$ (Equation (13)). Besides globality, a benefit of this technique is that it can easily be applied to higher dimensions whereas especially geometric approaches struggle with a larger number of objectives. The computational cost, however, increases exponentially with the dimension of the Pareto set such that in practice, we are restricted to a moderate number of objectives. In the following, we recall the subdivision algorithm for the solution of multiobjective optimization problems with exact information before we proceed with our extension to inexact information. For details we refer to [DSH05].

4.1 Subdivision Algorithm with Exact Gradients

In order to apply the subdivision algorithm to a multiobjective optimization problem, we first formulate a descent step of the optimization procedure using a line search approach, i.e.

$$x_{j+1} = g(x_j) = x_j + h_j q_j(x_j), \quad (14)$$

where $q_j(x_j)$ is the descent direction according to (3) and h_j is an appropriately chosen step length (e.g. according to the Armijo rule [NW06] for all objectives). The subdivision algorithm was initially developed in the context of dynamical systems [DH97] in order to compute the global attractor of a dynamical system g relative to a set Q , i.e. the set A_Q such that $g(A_Q) = A_Q$. Using a multilevel subdivision scheme, the following algorithm yields an outer approximation of A_Q in the form of a sequence of sets $\mathcal{B}_0, \mathcal{B}_1, \dots$, where each \mathcal{B}_s is a subset of \mathcal{B}_{s-1} and consists of finitely many subsets B (from now on referred to as *boxes*) of Q covering the relative global attractor A_Q . For each set \mathcal{B}_s , we define a box diameter

$$\text{diam}(\mathcal{B}_s) = \max_{B \in \mathcal{B}_s} \text{diam}(B)$$

which tends to zero for $s \rightarrow \infty$ within Algorithm 2. Interpreting (14) as a dynamical system, the

Algorithm 2 (Subdivision algorithm)

Let \mathcal{B}_0 be an initial collection of finitely many subsets of the compact set Q such that $\bigcup_{B \in \mathcal{B}_0} B = Q$. Then, \mathcal{B}_s is inductively obtained from \mathcal{B}_{s-1} in two steps:

- (i) **Subdivision.** Construct from \mathcal{B}_{s-1} a new collection of subsets $\widehat{\mathcal{B}}_s$ such that

$$\bigcup_{B \in \widehat{\mathcal{B}}_s} B = \bigcup_{B \in \mathcal{B}_{s-1}} B,$$

$$\text{diam}(\widehat{\mathcal{B}}_s) = \theta_s \text{diam}(\mathcal{B}_{s-1}), \quad 0 < \theta_{\min} \leq \theta_s \leq \theta_{\max} < 1.$$

- (ii) **Selection.** Define the new collection \mathcal{B}_s by

$$\mathcal{B}_s = \left\{ B \in \widehat{\mathcal{B}}_s \mid \exists \widehat{B} \in \widehat{\mathcal{B}}_s \text{ such that } g^{-1}(B) \cap \widehat{B} \neq \emptyset \right\}.$$

attractor is the set of points for which $q(x_j) = 0$, i.e. the set of points satisfying the KKT conditions. As a first step, we can prove that each accumulation point of the system is a substational point for (MOP):

Theorem 4.1 ([DSH05]). *Suppose that x^* is an accumulation point of the sequence $(x_j)_{j=0,1,\dots}$ created by (14). Then, x^* is a substational point of (MOP).*

Using Algorithm 2, we can now compute an outer approximation of the attractor of the dynamical system (14) which contains all points satisfying (1). The attractor of a dynamical system is always connected, which is not necessarily the case for $\mathcal{P}_{S,\text{sub}}$. In this situation, the attractor is a superset of $\mathcal{P}_{S,\text{sub}}$. However, if $\mathcal{P}_{S,\text{sub}}$ is bounded and connected, it coincides with the attractor of (14), which is stated in the following theorem:

Theorem 4.2 ([DSH05]). *Suppose that the set $\mathcal{P}_{S,\text{sub}}$ of points $x \in \mathbb{R}^n$ satisfying (1) is bounded and connected. Let Q be a compact neighborhood of $\mathcal{P}_{S,\text{sub}}$. Then, an application of Algorithm 2 to Q with respect to the iteration scheme (14) leads to a sequence of coverings \mathcal{B}_s which converges to the entire set $\mathcal{P}_{S,\text{sub}}$, that is,*

$$d_h(\mathcal{P}_{S,\text{sub}}, \mathcal{B}_s) \rightarrow 0, \quad \text{for } s = 0, 1, 2, \dots,$$

where d_h denotes the Hausdorff distance.

The concept of the subdivision algorithm is illustrated in Figure 5, where in addition a numerical realization has been introduced (cf. Section 4.2.1 for details).

4.2 Subdivision Algorithm with Inexact Function and Gradient Values

In this section, we combine the results from [DSH05] with the results from Section 3 in order to devise an algorithm for the approximation of the set of substational points of (MOP) with

inexact function and gradient information. So far, we have only considered inexactness in the descent direction where we only need to consider errors in the gradients. For the computation of a descent step, we further require a step length strategy [NW06] where we additionally need to consider errors in the function values. For this purpose, we extend the concept of *non-dominance* (2.2) to inexact function values:

Definition 4.3. Consider the multiobjective optimization problem (MOP), where the objective functions $f_i(x)$, $i = 1, \dots, k$, are only known approximately according to (4). Then

- (a) a point $x^* \in \mathbb{R}^n$ confidently dominates a point $x \in \mathbb{R}^n$, if $\tilde{f}_i(x^*) + \xi_i \leq \tilde{f}_i(x) - \xi_i$ for $i = 1, \dots, k$ and $\tilde{f}_i(x^*) + \xi_i < \tilde{f}_i(x) - \xi_i$ for at least one $i \in 1, \dots, k$.
- (b) a set $\mathcal{B}^* \subset \mathbb{R}^n$ confidently dominates a set $\mathcal{B} \subset \mathbb{R}^n$ if for every point $x \in \mathcal{B}$ there exists at least one point $x^* \in \mathcal{B}^*$ dominating x .
- (c) The set of almost non-dominated points which is a superset of the Pareto set \mathcal{P}_S is defined as:

$$\mathcal{P}_{S,\xi} = \left\{ x^* \in \mathbb{R}^n \mid \nexists x \in \mathbb{R}^n \text{ with } \tilde{f}_i(x) + \xi_i \leq \tilde{f}_i(x^*) - \xi_i, i = 1, \dots, k \right\}. \quad (15)$$

Note that the same definition was also introduced in [SVC09] in order to increase the number of almost Pareto optimal points and thus, the number of possible options for a decision maker.

As a consequence of the inexactness in the function values and the gradients, the approximated set is a superset of the Pareto set. Depending on the errors ξ_i and ϵ_i , $i = 1, \dots, k$, in the function values and in the gradients, respectively, each point is either contained in the set $\mathcal{P}_{S,\xi}$ (15) or in the set $\mathcal{P}_{S,\epsilon}$ (13). Based on these considerations, we introduce an inexact dynamical system similar to (14):

$$x_{j+1} = x_j + h_j p_j, \quad (16)$$

where the direction p_j is computed using Algorithm 1 ($p_j = q_u(x_j)$) and the step length h_j is determined by a modified Armijo rule [NW06] such that $\tilde{f}_i(x_j + h_j p_j) + \xi_i \leq \tilde{f}_i(x_{j-1}) + c_1 h_j p_j^\top \nabla \tilde{f}_i(x_{j-1})$. If the errors ξ and ϵ are zero, the computed set is reduced to the set of substationary points. In this situation $\mathcal{P}_{S,\xi} = \mathcal{P}_S$ and $\sum_{i=1}^k \hat{\alpha}_{min,i} = 0$, hence $p_j = q(x_j)$. Convergence of the dynamical system (16) to an *approximately substationary* point is investigated in the following theorem:

Theorem 4.4. Consider the multiobjective optimization problem (MOP) with inexact objective functions and inexact gradients according to (4) and (5), respectively. Suppose that x^* is an accumulation point of the sequence $(x_j)_{j=0,1,\dots}$ created by (16). Then,

- (a) $x^* \in \mathcal{P}_{S,\epsilon,\xi} = \mathcal{P}_{S,\epsilon} \cup \mathcal{P}_{S,\xi}$ where $\mathcal{P}_{S,\epsilon}$ and $\mathcal{P}_{S,\xi}$ are defined according to (13) and (15), respectively.
- (b) If $\xi_i = 0$, $\epsilon_i = 0$, $i = 1, \dots, k$, x^* is a substationary point of (MOP).

Proof. (a) For a point x_j created by the sequence (16) one of the following statements is true:

- | | |
|---|---|
| i) $x_j \in \mathcal{P}_{S,\epsilon} \quad \wedge \quad x_j \notin \mathcal{P}_{S,\xi}$ | ii) $x_j \notin \mathcal{P}_{S,\epsilon} \quad \wedge \quad x_j \in \mathcal{P}_{S,\xi}$ |
| iii) $x_j \in \mathcal{P}_{S,\epsilon} \quad \wedge \quad x_j \in \mathcal{P}_{S,\xi}$ | iv) $x_j \notin \mathcal{P}_{S,\epsilon} \quad \wedge \quad x_j \notin \mathcal{P}_{S,\xi}$ |

In case i) $x_j \in \mathcal{P}_{S,\epsilon}$ which means that the gradients $\nabla \tilde{f}_i(x_j)$, $i = 1, \dots, k$, approximately satisfy the KKT conditions. We obtain $\sum_{i=1}^k \hat{\alpha}_{min,i} = 1$, i.e. the set of valid descent directions is empty ($\mathcal{Q}_u = \emptyset$, cf. Theorem 3.6). Consequently, $p_j = 0$ and the point x_j is an accumulation point of the sequence (16). In case ii) the inaccuracies in the function values $\tilde{f}_i(x_j)$, $i = 1, \dots, k$ prohibit a guaranteed decrease for all objectives. According to the modified Armijo rule $h_j = 0$ such that x_j is an accumulation point. In case iii) both $p_j = 0$ and $h_j = 0$. In case iv) we have $p_j = q(x_j)$. If for any $j \in \{0, 1, \dots\}$, $x_j \in \mathcal{P}_{S,\epsilon}$ or $x_j \in \mathcal{P}_{S,\xi}$, we are in one of the cases i) to iii) and x_j is an accumulation point. Otherwise, we obtain a descent direction such that the sequence (16) converges to a substationary point $x^* \in \mathcal{P}_{S,sub} \subseteq \mathcal{P}_{S,\epsilon}$ of (MOP) which is proved in [DSH05].

For part (b), we obtain $\hat{\alpha}_{min,i} = 0$ by setting the errors ϵ_i , $i = 1, \dots, k$, to zero and hence, the descent direction is $p_j = q(x_j)$ (cf. Algorithm 1). When $\xi_i = 0$, $i = 1, \dots, k$, the modified Armijo rule becomes the standard Armijo rule for multiple objectives. Consequently, the problem is reduced to the case with exact function and gradient values (case iv) in part (a)). \square

Following along the lines of [DSH05], we can use this result in order to prove convergence of the subdivision algorithm with inexact values:

Theorem 4.5. *Suppose that the set $\mathcal{P}_{S,\epsilon,\xi} = \mathcal{P}_{S,\epsilon} \cup \mathcal{P}_{S,\xi}$ where $\mathcal{P}_{S,\epsilon}$ and $\mathcal{P}_{S,\xi}$ are defined according to (13) and (15), respectively, is bounded and connected. Let Q be a compact neighborhood of $\mathcal{P}_{S,\epsilon,\xi}$. Then, an application of Algorithm 2 to Q with respect to the iteration scheme (16) leads to a sequence of coverings \mathcal{B}_s which is a subset of $\mathcal{P}_{S,\epsilon,\xi}$ and a superset of the set $\mathcal{P}_{S,sub}$ of substationary points $x \in \mathbb{R}^n$ of (MOP), that is,*

$$\mathcal{P}_{S,sub} \subset \mathcal{B}_s \subset \mathcal{P}_{S,\epsilon,\xi}.$$

Consequently, if the errors tend towards zero, we observe

$$\lim_{\epsilon_i, \xi_i \rightarrow 0, i=1, \dots, k} d_h(\mathcal{B}_s, \mathcal{P}_{S,\epsilon,\xi}) = d_h(\mathcal{P}_{S,sub}, \mathcal{B}_s) = 0.$$

4.2.1 Numerical Realization of the Selection Step

In this section, we briefly describe the numerical realization of Algorithm 2. For details, we refer to [DSH05]. The elements $B \in \mathcal{B}_s$ are n -dimensional boxes. In the selection step, each box is represented by a prescribed number of sample points at which the dynamical system (16) is evaluated according to Algorithm 3 (see Figure 5 (a) for an illustration). Then, we evaluate which boxes the sample points are mapped into and eliminate all “empty” boxes, i.e. boxes which do not possess a preimage within \mathcal{B}_s (Figure 5 (b)). The remaining boxes are subdivided and we proceed

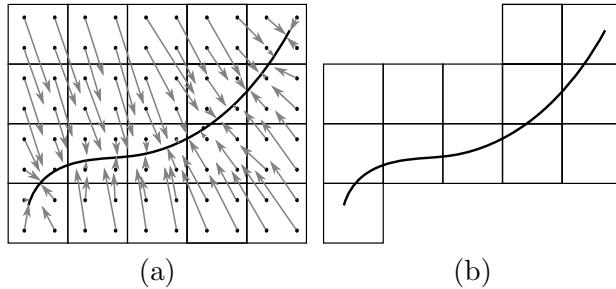


Figure 5: Global subdivision algorithm – selection step. (a) Evaluation of the dynamical system (14). (b) All boxes that do not possess a preimage within the collection are discarded.

with the next elimination step until a certain stopping criterion is met (e.g. a prescribed number of subdivision steps).

Algorithm 3 (Descent step under inexactness)

Require: Initial point x_0 , error bounds ξ_i and ϵ_i , $i = 1, \dots, k$, constant $0 < c_1 < 1$

- 1: Compute $\tilde{f}_i(x_0)$ and $\nabla \tilde{f}_i(x_0)$, $i = 1, \dots, k$
 - 2: Compute a direction p according to Algorithm 1
 - 3: Compute a step length h that satisfies the modified Armijo rule
 $\tilde{f}_i(x_0 + hp) + \xi_i \leq f_i(x_0) - \xi_i + c_1 hp^\top \nabla \tilde{f}_i(x_0)$, e.g. via backtracking [NW06]
 - 4: Compute $x_{step} = x_0 + hp$
-

4.3 Gradient Free Realization

In many applications, gradients are unknown or difficult to compute. In this case, we can use a gradient-free alternative of Algorithm 2 which is called the *Sampling algorithm* in [DSH05]. Algorithm 4 also consists of a subdivision and a selection step with the difference that the selection step is a non-dominance test. Hence, we compare all sample points and eliminate all boxes that contain only dominated points. This way, it is also possible to easily include constraints. In the presence of inequality constraints, for example, we eliminate all boxes for which all sample points violate the constraints and then consequently perform the non-dominance test on the remaining boxes. Equality constraints are simply modeled by introducing two inequality constraints. Finally, a combination of both the gradient-based and the gradient-free algorithm can be applied in order to speed up convergence or to reduce the gradient-based algorithm to the computation of the Pareto set \mathcal{P}_S instead of the set of substational points $\mathcal{P}_{S,sub}$.

Considering errors in the function values, we can use the sampling algorithm to compute the superset $\mathcal{P}_{S,\xi}$ of the global Pareto set \mathcal{P}_S . In the limit of vanishing errors, this is again reduce to the exact Pareto set:

$$\lim_{\xi_i \rightarrow 0, i=1, \dots, k} d_h(\mathcal{P}_{S,\xi}, \mathcal{P}_S) = 0.$$

Algorithm 4 (Sampling algorithm)

Let \mathcal{B}_0 be an initial collection of finitely many subsets of the compact set Q such that $\bigcup_{B \in \mathcal{B}_0} B = Q$. Then, \mathcal{B}_s is inductively obtained from \mathcal{B}_{s-1} in two steps:

- (i) Subdivision. Construct from \mathcal{B}_{s-1} a new collection of subsets $\widehat{\mathcal{B}}_s$ such that

$$\begin{aligned}\bigcup_{B \in \widehat{\mathcal{B}}_s} B &= \bigcup_{B \in \mathcal{B}_{s-1}} B, \\ \text{diam}(\widehat{\mathcal{B}}_s) &= \theta_s \text{diam}(\mathcal{B}_{s-1}), \quad 0 < \theta_{\min} \leq \theta_s \leq \theta_{\max} < 1.\end{aligned}$$

- (ii) Selection. Define the new collection \mathcal{B}_s by

$$\mathcal{B}_s = \left\{ B \in \widehat{\mathcal{B}}_s \mid \nexists \widehat{B} \in \widehat{\mathcal{B}}_s \text{ such that } \widehat{B} \text{ confidently dominates } B \right\}.$$

5 Results

In this section, we illustrate the results from Sections 3 and 4 using three examples. To this end, we add random perturbations to the respective model such that (4) and (5) hold. We start with a two dimensional example function $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ for two paraboloids:

$$\min_{x \in \mathbb{R}^2} F(x) = \min_{x \in \mathbb{R}^2} \begin{pmatrix} (x_1 - 1)^2 + (x_2 - 1)^4 \\ (x_1 + 1)^2 + (x_2 + 1)^2 \end{pmatrix}. \quad (17)$$

In Figure 6, the box covering of the Pareto set and the corresponding Pareto front obtained with Algorithm 2 are shown without errors and with $\xi = (0, 0)^\top$, $\epsilon = (0.1, 0.1)^\top$ (Figure 6 (a)) and $\xi = (0, 0)^\top$, $\epsilon = (0.0, 0.2)^\top$ (Figure 6 (b)), respectively. The background in (a) and (b) is colored according to the norm of the optimality condition (1), obtained by solving (QOP), and the white line indicates the upper bound of the error (13). We see that in (a), the box covering is close to the error bound whereas it is less sharp in (b). Consequently, the error estimate is more accurate when the errors are of comparable size in all gradients. The Pareto fronts corresponding to (a) and (b) are shown in (c) and (d). We see that the difference between the Pareto front of the exact solution (red) and the Pareto front of the inexact solution (green) is relatively small but that additional points are computed at the boundary of the front, i.e. close to the individual minima $F_1(x) = 0$ and $F_2(x) = 0$.

In the numerical realization, we approximate each box by an equidistant grid with two points in each direction, i.e. by four sample points in total. This results in a total number of $\approx 50,000$ function evaluations for the exact problem. The number of boxes is much higher for the inexact solution, in this case by a factor of ≈ 8 . This is not surprising since the equality condition (1) is now replaced by the inequality condition (13). Hence, the approximated set is no longer a $k - 1$ -dimensional object. All sets $B \in \mathcal{B}_s$ that satisfy the inequality condition (13) are not discarded in the selection step. Consequently, at a certain box size, the number of boxes increases exponentially

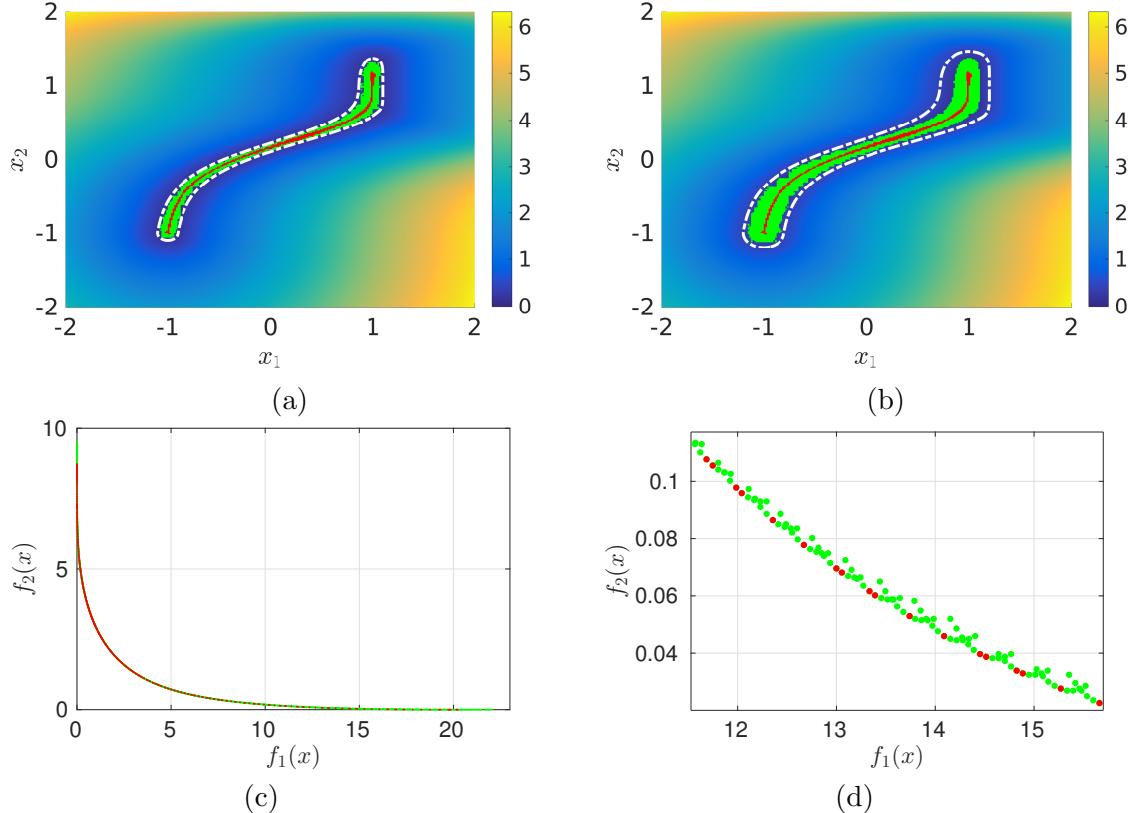


Figure 6: (a) Box covering of the Pareto set of problem (17) after 16 subdivision steps ($\text{diam}(\mathcal{B}_0) = 4$, $\text{diam}(\mathcal{B}_{20}) = 1/2^6$). The solution without errors ($\mathcal{P}_{S,sub}$) is shown in red, the solution with $\epsilon = (0.1, 0.1)^\top$, $\xi = (0, 0)^\top$ ($\mathcal{P}_{S,\epsilon}$) is shown in green. The background color represents the norm of the optimality condition (1) and the white line is the iso-curve $\|q(x)\|_2 = 2\|\epsilon\|_\infty = 0.2$, i.e. the upper bound of the error. (b) Analog to (a) but with $\epsilon = (0, 0.2)^\top$ and the iso-curve $\|q(x)\|_2 = 0.4$. (c)–(d) The Pareto fronts corresponding to (a). The points are the images of the box centers (color coding as in (a) and (b)).

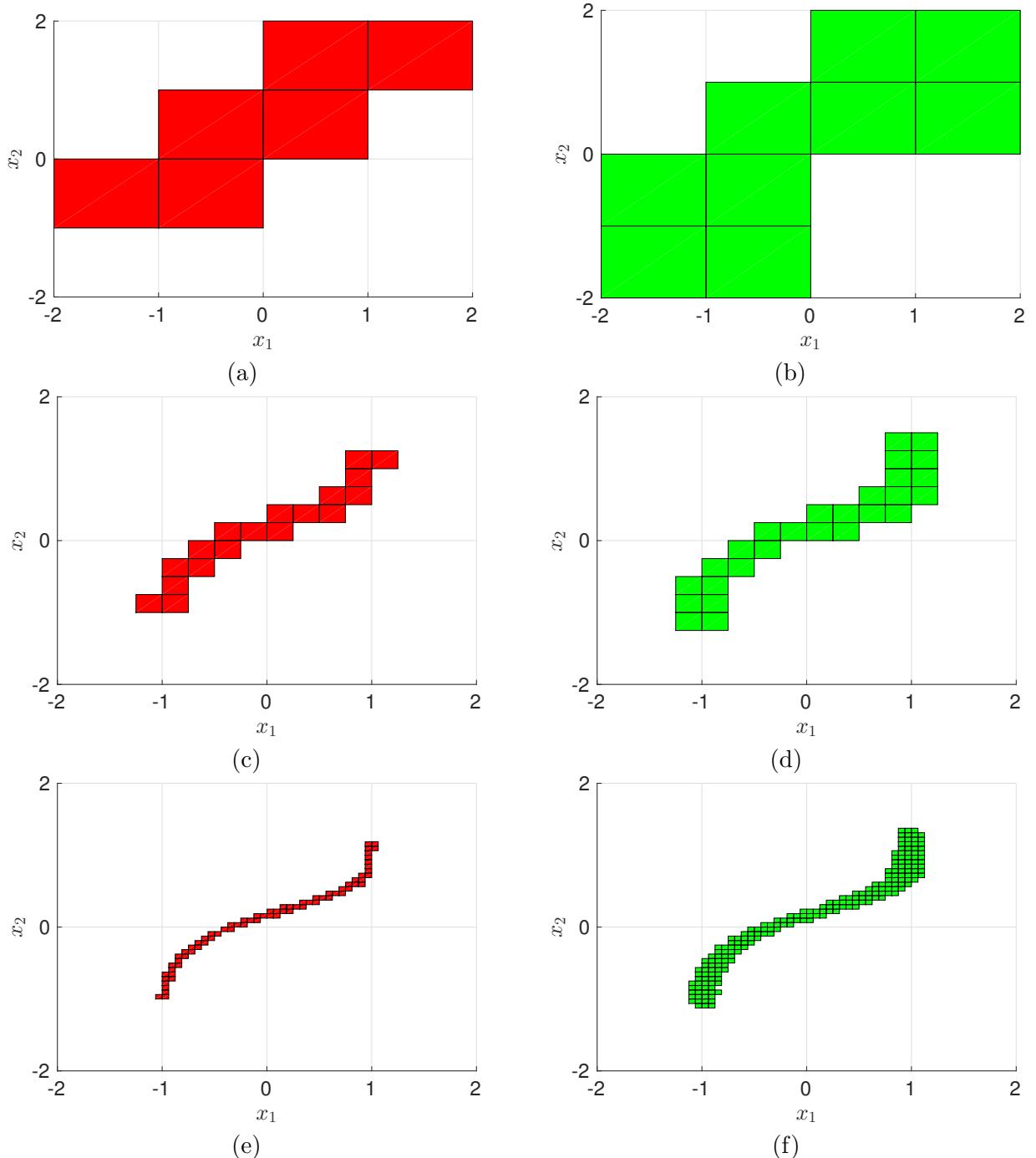


Figure 7: Comparison of the inexact and the exact solution of problem (17) after 4, 8 and 12 subdivision steps.

with decreasing diameter $\text{diam}(B)$ (cf. Figure 10 (b)). The result is an increased computational effort for later iterations. This is visualized in Figure 7, where the solutions at different stages of the subdivision algorithm are compared. A significant difference between the solutions can only be observed in later stages (Figure 7 (e) and (f)). For this reason, an adaptive strategy needs to be developed where boxes satisfying (13) remain within the box collection but are no longer considered in the subdivision algorithm.

As a second example, we consider the function $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$:

$$\min_{x \in \mathbb{R}^3} F(x) = \min_{x \in \mathbb{R}^3} \begin{pmatrix} (x_1 - 1)^4 + (x_2 - 1)^2 + (x_3 - 1)^2 \\ (x_1 + 1)^2 + (x_2 + 1)^4 + (x_3 + 1)^2 \\ (x_1 - 1)^2 + (x_2 + 1)^2 + (x_3 - 1)^4 \end{pmatrix}. \quad (18)$$

The observed behavior is very similar to the two-dimensional case, cf. Figure 8, where in (b) the box covering for the inexact problem is shown as well as the iso-surface $\|q(x)\|_2 = 2\|\epsilon\|_\infty$. One can see that the box covering lies inside this iso-surface except for small parts of some boxes. This is due to the finite box size and the fact that at least one sample point is mapped into the box itself. For smaller box radii, this artifact does no longer occur.

Finally, we consider an example where the Pareto set is disconnected. This is an example from production and was introduced in [SSW02]. We want to minimize the failure of a product which consists of n components. The probability of failing is modeled individually for each component and depends on the additional cost x :

$$\begin{aligned} p_1(x) &= 0.01 \exp(-(x_1/20)^{2.5}), \\ p_2(x) &= 0.01 \exp(-(x_2/20)^{2.5}), \\ p_j(x) &= 0.01 \exp(-x_1/15), \quad j = 3, \dots, n. \end{aligned}$$

The resulting MOP is thus to minimize the failure and the additional cost at the same time:

$$\min_{x \in \mathbb{R}^n} F(x) = \min_{x \in \mathbb{R}^n} \left(\frac{\sum_{j=1}^n x_j}{1 - \sum_{j=1}^n (1 - p_j(x))} \right). \quad (19)$$

We now assume that the additional cost x is subject to some uncertainty, e.g. due to varying prices, and set $|\tilde{x}_i - x_i| < 0.01$ for $i = 1, \dots, n$. Using this, we can estimate the error bounds within the initial box $\mathcal{B}_0 = [0, 40]^n$ and obtain $\xi = (0.05, 2 \cdot 10^{-5})^\top$ and $\epsilon = (0, 8 \cdot 10^{-7})^\top$.

Since the set of substationary points is disconnected, we here utilize the Sampling Algorithm 4. For $n = 5$, the resulting Pareto sets \mathcal{P}_S and $\mathcal{P}_{S,\xi}$ are depicted in Figures ?? (a) and (b), respectively. Due to the small gradient of the objectives, this results in a significantly increased number of boxes by a factor of ≈ 300 (cf. Figure 10 (b)).

The quality of the solution can be measured using the Hausdorff distance $d_h(\mathcal{P}_S, \mathcal{P}_{S,\xi})$ [RW98]. This is depicted in Figure 10 (a), where the distance between the exact and the inexact solution is shown for all subdivision steps for the three examples above. We see that the distance reaches an almost constant value in the later stages. This distance is directly influenced by the upper bounds ϵ and ξ , respectively. However, it cannot simply be controlled by introducing a bound on the error in the objectives or gradients since it obviously depends on the objective functions. Hence, in order to limit the error in the decision space, further assumptions on the objectives have to be made.

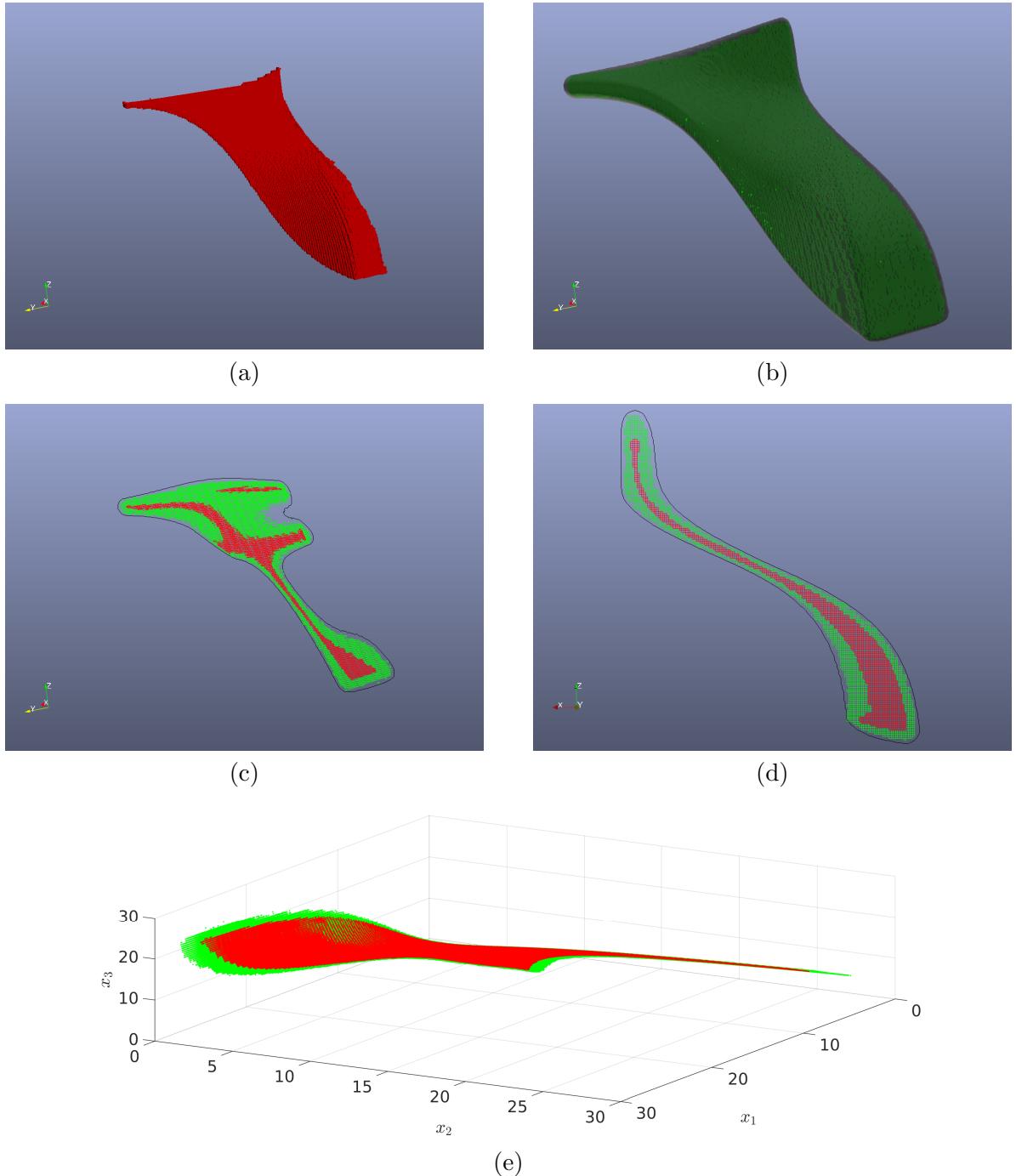


Figure 8: Box coverings of the Pareto set (exact solution: red, inexact solution: green) of problem (18) after 24 subdivision steps ($\text{diam}(\mathcal{B}_0) = 4$, $\text{diam}(\mathcal{B}_{24}) = 1/2^6$). (a) Solution without errors. (b) Solution with $\xi = (0, 0, 0)^\top$ and $\epsilon_i = (0.1, 0.1, 0.1)^\top$. The iso-surface ($\|q(x)\|_2 = 2\|\epsilon\|_\infty = 0.2$) is the upper bound of the error. (c)–(d) Two-dimensional cut planes through the Pareto set. The point of view in (c) is as in (a), (b) and (d) is a cut plane parallel to the x_2 plane at $x_2 = -0.9$. (e) The corresponding Pareto fronts.

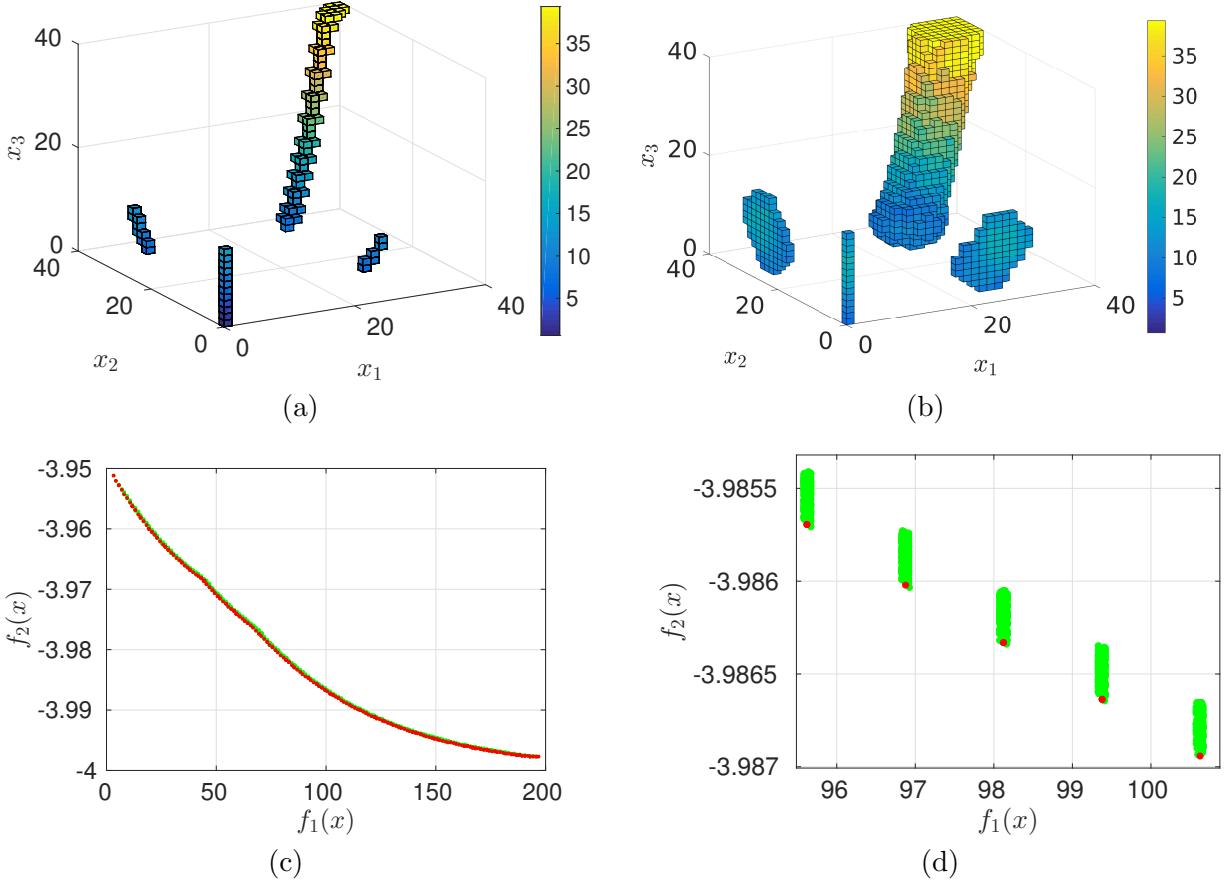


Figure 9: (a) Projection of the box covering of \mathcal{P}_S for problem (19) with $n = 5$ after 25 subdivision steps ($\text{diam}(\mathcal{B}_0) = 40$, $\text{diam}(\mathcal{B}_{25}) = 1.25$). The coloring represents the fourth component x_4 . (b) Box covering of $\mathcal{P}_{S,\xi}$ with inexact data \tilde{x} with $|\tilde{x}_i - x_i| < 0.01$ for $i = 1, \dots, 5$. (c)–(d) The Pareto fronts corresponding to (a) and (b) in green and red, respectively. The points are the images of the box centers.

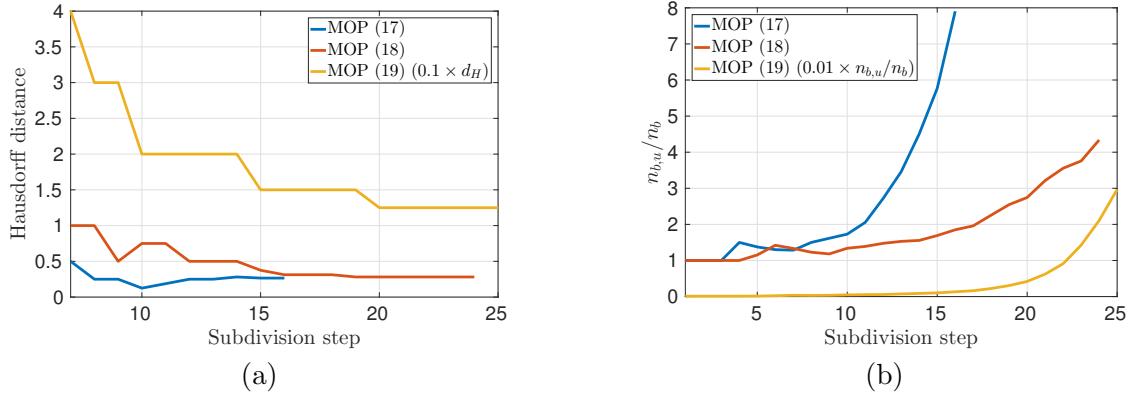


Figure 10: (a) Hausdorff distance $d_h(\mathcal{P}_S, \mathcal{P}_{S,\xi})$ for the three examples above. (b) The corresponding ratio of box numbers between the inexact solution ($n_{b,u}$) and the exact solution (n_b).

6 Conclusion

In this article, we present an extension to the subdivision algorithms developed in [DSH05] for MOPs with uncertainties in the form of inexact function and gradient information. An additional condition for a descent direction is derived in order to account for the inaccuracies in the gradients. Convergence of the extended subdivision algorithm to a superset of the Pareto set is proved and an upper bound for the maximal distance to the set of substational points is given. When taking into account errors, the number of boxes in the covering of the Pareto set increases, especially for later iterations, causing larger computational effort. For this reason, an adaptive strategy needs to be developed where boxes approximately satisfying the KKT conditions (according to (13)) remain within the box collection but are no longer considered in the subdivision algorithm. Furthermore, we intend to extend the approach to constrained MOPs in the future. A comparison to other methods, especially memetic algorithms, would be interesting to investigate the numerical efficiency of the presented method. Finally, we intend to combine the subdivision algorithm with model order reduction techniques in order to solve multiobjective optimal control problems constrained by PDEs (see e.g. [BBV16] for the solution of bicriterial MOPs with the *reference point* approach or [POBD15] for multiobjective optimal control of the Navier-Stokes equations). For the development of reliable and robust multiobjective optimization algorithms, error estimates for reduced order models [TV09] need to be combined with concepts for inexact function values and gradients. The idea is to set upper bounds for the errors ξ and ϵ and thereby define the corresponding accuracy requirements for the reduced order model.

Acknowledgement: This work is supported by the Priority Programme SPP 1962 “Non-smooth and Complementarity-based Distributed Parameter Systems” of the German Research Foundation (DFG).

References

- [BBV16] S. Banholzer, D. Beermann, and S. Volkwein. POD-Based Bicriterial Optimal Control by the Reference Point Method. In *2nd IFAC Workshop on Control of Systems Governed by Partial Differential Equations*, pages 210–215, 2016.
- [Bos12] P. A. N. Bosman. On Gradients and Hybrid Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, 16(1):51–69, 2012.
- [BZ06] M. Basseur and E. Zitzler. Handling uncertainty in indicator-based multiobjective optimization. *International Journal of Computational Intelligence Research*, 2(3):255–272, 2006.
- [CVL02] C. A. Coello Coello, D. A. Van Veldhuizen, and G. B. Lamont. *Evolutionary algorithms for solving multi-objective problems*, volume 242. Kluwer Academic New York, 2002.
- [Des12] J.-A. Desideri. Multiple-Gradient Descent Algorithm for Multiobjective Optimization. In J. Berhardsteiner, editor, *European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS)*, 2012.
- [DG05] K. Deb and H. Gupta. Searching for Robust Pareto-Optimal Solutions in Multi-objective Optimization. In C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, editors, *Evolutionary Multi-Criterion Optimization*, pages 150–164. Springer Berlin Heidelberg, 2005.
- [DH97] M. Dellnitz and A. Hohmann. A subdivision algorithm for the computation of unstable manifolds and global attractors. *Numerische Mathematik*, 75(3):293–317, 1997.
- [DMM05] K. Deb, M. Mohan, and S. Mishra. Evaluating the ϵ -Domination Based Multi-Objective Evolutionary Algorithm for a Quick Computation of Pareto-Optimal Solutions. *Evolutionary Computation*, 13(4):501–525, 2005.
- [DSH05] M. Dellnitz, O. Schütze, and T. Hestermeyer. Covering Pareto Sets by Multilevel Subdivision Techniques. *Journal of Optimization Theory and Applications*, 124(1):113–136, 2005.
- [Ehr05] M. Ehrgott. *Multicriteria optimization*. Springer Berlin Heidelberg New York, 2005.
- [EW07] A. Engau and M. M. Wiecek. Generating ϵ -efficient solutions in multiobjective programming. *European Journal of Operational Research*, 177(3):1566–1579, 2007.
- [FDS09] J. Fliege, L. M. G. Drummond, and B. F. Svaiter. Newton’s Method for Multiobjective Optimization. *SIAM Journal on Optimization*, 20(2):602–626, 2009.
- [FS00] J. Fliege and B. F. Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51(3):479–494, 2000.

- [Hil01] C. Hillermeier. *Nonlinear Multiobjective Optimization - A Generalized Homotopy Approach*. Birkhäuser, 2001.
- [HSS13] C. Hernández, J.-Q. Sun, and O. Schütze. *Computing the Set of Approximate Solutions of a Multi-objective Optimization Problem by Means of Cell Mapping Techniques*, pages 171–188. Springer International Publishing, Heidelberg, 2013.
- [Hug01] E. J. Hughes. Evolutionary Multi-objective Ranking with Uncertainty and Noise. In E. Zitzler, L. Thiele, K. Deb, C. A. Coello Coello, and D. Corne, editors, *Evolutionary Multi-Criterion Optimization*, pages 329–343. Springer Berlin Heidelberg, 2001.
- [Jah06] J. Jahn. Multiobjective Search Algorithm with Subdivision Technique. *Computational Optimization and Applications*, 35(2):161–175, 2006.
- [Jin11] Y. Jin. Surrogate-assisted evolutionary computation: Recent advances and future challenges. *Swarm and Evolutionary Computation*, 1(2):61–70, 2011.
- [KT51] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Proceedings of the 2nd Berkeley Symposium on Mathematical and Statsitical Probability*, pages 481–492. University of California Press, 1951.
- [LSCS10] A. Lara, G. Sanchez, C. A. Coello Coello, and O. Schütze. HCS: A New Local Search Strategy for Memetic Multiobjective Evolutionary Algorithms. In *IEEE Transactions on Evolutionary Computation*, volume 14, pages 112–132. IEEE, 2010.
- [Mie99] K. Miettinen. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, 1999.
- [NCM12] F. Neri, C. Cotta, and P. Moscato. *Handbook of memetic algorithms*, volume 379. Springer, 2012.
- [NW06] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Science & Business Media, 2006.
- [POBD15] S. Peitz, S. Ober-Blöbaum, and M. Dellnitz. Multiobjective Optimal Control Methods for Fluid Flow Using Model Order Reduction. *arXiv preprint: <http://arxiv.org/abs/1510.05819>*, 2015.
- [RW98] R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317. Springer Berlin Heidelberg, 1998.
- [SASL16] O. Schütze, S. Alvarado, C. Segura, and R. Landa. Gradient subspace approximation: a direct search method for memetic computing. *Soft Computing*, 2016.
- [SCTT08] O. Schütze, C. A. Coello Coello, E. Tantar, and E.-G. Talbi. Computing the Set of Approximate Solutions of an MOP with Stochastic Search Algorithms. In *Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation*, pages 713–720. ACM, 2008.

- [SM08] A. Singh and B. S. Minsker. Uncertainty-based multiobjective optimization of ground-water remediation design. *Water Resources Research*, 44(2), 2008.
- [SML⁺16] O. Schütze, A. Martín, A. Lara, S. Alvarado, E. Salinas, and C. A. Coello Coello. The directed search method for multi-objective memetic algorithms. *Computational Optimization and Applications*, 63(2):305–332, 2016.
- [SSW02] S. Schäffler, R. Schultz, and K. Weinzierl. Stochastic method for the solution of unconstrained vector optimization problems. *Journal of Optimization Theory and Applications*, 114(1):209–222, 2002.
- [SVC09] O. Schütze, M. Vasile, and C. A. Coello Coello. Computing the Set of epsilon-efficient Solutions in Multi-Objective Space Mission Design. *Journal of Aerospace Computing, Information, and Communication*, 8(3):53–70, 2009.
- [SWOBD13] O. Schütze, K. Witting, S. Ober-Blöbaum, and M. Dellnitz. Set Oriented Methods for the Numerical Treatment of Multiobjective Optimization Problems. In E. Tantar, A.-A. Tantar, P. Bouvry, P. Del Moral, P. Legrand, C. A. Coello Coello, and O. Schütze, editors, *EVOLVE- A Bridge between Probability, Set Oriented Numerics and Evolutionary Computation*, volume 447 of *Studies in Computational Intelligence*, pages 187–219. Springer Berlin Heidelberg, 2013.
- [Tei01] J. Teich. Pareto-Front Exploration with Uncertain Objectives. In E. Zitzler, L. Thiele, K. Deb, C. A. Coello Coello, and D. Corne, editors, *Evolutionary Multi-Criterion Optimization*, pages 314–328. Springer Berlin Heidelberg, 2001.
- [TV09] F. Tröltzsch and S. Volkwein. POD a-posteriori error estimates for linear-quadratic optimal control problems. *Computational Optimization and Applications*, 44:83–115, 2009.
- [Whi86] D. J. White. Epsilon efficiency. *Journal of Optimization Theory and Applications*, 49(2):319–337, 1986.