Representation of the Pareto front for heterogeneous multi-objective optimization[*]

Jana Thomann[†] and Gabriele Eichfelder[‡]

August 26, 2019

Abstract

Optimization problems with multiple objectives which are expensive, i. e. where function evaluations are time consuming, are difficult to solve. Finding at least one locally optimal solution is already a difficult task. In case only one of the objective functions is expensive while the others are cheap, for instance analytically given, this can be used in the optimization procedure. Using a trust-region approach and the Tammer-Weidner-functional for finding descent directions, in [19] an algorithm was proposed which makes use of the heterogeneity of the objective functions. In this paper, we present three heuristic approaches which allow to find additional optimal solutions of the multiobjective optimization problem and by that representations at least of parts of the Pareto front. We present the related theoretical results as well as numerical results on some test instances.

Key Words: multi-objective optimization, heterogeneous optimization, Tammer-Weidner-functional, trust region algorithm

1 Introduction

In multi-objective optimization one studies optimization problems with several competing objectives. An optimal solution for such problems is a feasible point such that there exists no other feasible solution which satisfies all objective functions at least equally good and which is better w.r.t. at least one objective function. Based on this optimality concept, in general an infinite number of optimal solutions, called efficient, exist. The images of these efficient points in the image space are denoted as nondominated, and the set of nondominated points are also known as Pareto front. This set is then also in general not a singleton. Determining even just representations of this Pareto front is known to be a difficult task.

[†]Institute for Mathematics, Technische Universität Ilmenau, Po 10 05 65, D-98684 Ilmenau, Germany, jana.thomann@tu-ilmenau.de

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

In many applications multi-objective optimization problems arise with expensive black-box functions, i.e. one or several objective functions are not given analytically, but for example by a time-consuming simulation. Solving such problems is a challenge as it is not possible to evaluate the expensive objective functions too often. In case only some of the objective functions are expensive, but the others are cheap, we call such problems heterogeneous, and this heterogeneity can be used in a numerical approach.

The trust region algorithm presented in [19, 18, 20] and called MHT is designed to consider such multi-objective heterogeneous optimization problems where one of the objectives is an expensive black-box function. The simulation only gives function values. Derivative information is not available with reasonable effort and therefore not used. The other objective functions are so-called cheap functions which are analytically given, easy to compute, and derivatives are easily available. The general optimization problem is given by

$$\min_{x \in \mathbb{R}^n} f(x) \qquad\qquad (MOP)$$

with $f \colon \mathbb{R}^n \to \mathbb{R}^q$ and $f(x) = (f_1(x), \dots, f_q(x))^\top$. The objective functions $f_i \colon \mathbb{R}^n \to \mathbb{R}$ are assumed to be twice continuously differentiable for all $i = 1, 2, \dots, q$ and $\max_{i=1,\dots,q} f_i(x)$ is assumed to be bounded from below. The function f_1 is an expensive function, and the objective functions $f_i, i = 2, 3, \dots, q$, are assumed to be cheap functions. While the problem is stated here as an unconstrained problem, box constraints can also be handled, see [18]. Based on a trust-region approach a sequence of points is generated. Thereby, the descent directions are calculated by finding a point within the trust region which is as close as possible to a local ideal point. This leads to single-objective subproblems, which have to be solved, which have the structure of the well-known Tammer-Weidner functional [11].

However, the algorithm MHT proposed in [19, 20] calculates only one efficient solution, to be more concrete, accumulation points of the generated sequence of iteration points satisfy a necessary optimality criteria to be a locally optimal solution. The method is not designed to find a representation of the Pareto front. It is the topic of this paper to find representations of the Pareto front, at least of parts of the Pareto front, starting from single solutions generated by MHT . Thereby, the heterogeneity of the objective functions is used.

In section 2 the basic definitions and the algorithm MHT are stated. Three heuristic approaches based on MHT are described in section 3 which all aim on finding representations (of parts) of the Pareto front. The results of numerical tests are summarized in section 4. The results presented here are all based on [18].

2 Basic Definitions and Algorithm MHT

We first shortly recall the necessary definitions used in this paper. Then we present the algorithm MHT on which we base our approaches for a representation of the Pareto front. For the basic concepts in vector optimization see for instance the book [12], or the book chapters [7, 8].

A point $\overline{x} \in \mathbb{R}^n$ is called efficient for (MOP) if there exists no point $x \in \mathbb{R}^n$ with

$f_i(x) \leq f_i(\overline{x})$ for all $i \in \{1, 2, \ldots, q\}$ and $f(x) \neq f(\overline{x})$. The image of an efficient point is called nondominated. The set of all nondominated points is also known as Pareto front. For theoretical purposes we also need a weaker concept: a point $\overline{x} \in \mathbb{R}^n$ is called weakly efficient for (MOP) if there exists no point $x \in \mathbb{R}^n$ with $f_i(x) < f_i(\overline{x})$ for all $i \in \{1, 2, \ldots, q\}$. Both concepts can be restricted to local areas. Accordingly, a point $\overline{x} \in \mathbb{R}^n$ is called locally (weakly) efficient for (MOP) if there exists a neighborhood $U \subset \mathbb{R}^n$ with $\overline{x} \in U$ such that $\overline{x}$ is (weakly) efficient for (MOP) in U.

Obviously every efficient point is weakly efficient. The above concepts use the componentwise partial ordering in the image space $\mathbb{R}^q$. In the following we will use the inequality relations $<$ and $\leq$ for vectors in such a componentwise manner. Hence, for $a, b \in \mathbb{R}^n$ we write $a \leq b$ if it holds $a_i \leq b_i$ for all $i \in \{1, 2, \ldots, n\}$, and the same for $<$.

The following concept gives a necessary condition for weak efficiency, see for example [10].

Definition 2.1 *Let $f_i \colon \mathbb{R}^n \to \mathbb{R}, i = 1, 2, \ldots, q$, be continuously differentiable functions. A point $\overline{x} \in \mathbb{R}^n$ is called Pareto critical for (MOP), if for every vector $d \in \mathbb{R}^n$ there exists an index $j \in \{1, 2, \ldots, q\}$ such that it holds $\nabla f_j(\overline{x})^\top d \geq 0$.*

This concept is a generalization of the stationarity notion for scalar optimization problems, i.e. $\nabla f(\overline{x}) = 0_n$ for $q = 1$. The following lemma shows that Pareto criticality is a necessary condition for local weak efficiency, see for example [9, 12].

Lemma 2.2 *Let $f_i \colon \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, be continuously differentiable functions. If $\overline{x} \in \mathbb{R}^n$ is locally weakly efficient for (MOP), then it is Pareto critical for (MOP).*

The following lemma gives a characterization of Pareto critical points and comes from multi-objective descent methods [4, 9, 10]. This characterization is a main tool in the convergence proof of MHT in [19].

Lemma 2.3 *[10, Lem. 3] Let $f_i \colon \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, be continuously differentiable functions. For the function*

$$\omega(x) := -\min_{\|d\| \leq 1} \max_{i=1,\ldots,q} \nabla f_i(x)^\top d \tag{1}$$

the following statements hold.

(i) ω is continuous.

(ii) It holds $\omega(x) \geq 0$ for all $x \in \mathbb{R}^n$.

(iii) A point $x \in \mathbb{R}^n$ is Pareto critical for (MOP) if and only if it holds $\omega(x) = 0$.

The multi-objective heterogeneous trust region algorithm MHT is based on the trust region approach and generalizes it to multi-objective problems. Following the trust region concept, the computations are restricted to a local area in every iteration and the functions are replaced by local models. The search direction is generated in the image space and local ideal points of the model functions are used. The local area called trust region is defined as

$$B_k = \left\{ x \in \mathbb{R}^n \mid \|x - x^k\| \leq \delta_k \right\} \tag{2}$$

3

with x^k the current iteration point, $\delta_k > 0$ the radius and $\|\cdot\|$ a norm. We consider $\|\cdot\| := \|\cdot\|_2$ the Euclidean norm and a compatible matrix norm.

Within each trust region, the original functions are replaced by suitable local model functions $m_i^k, i = 1, 2, \ldots, q$, defined in B_k. For the analytically given functions $f_i, i = 2, 3, \ldots, q$, the Taylor model of second order is used since it is standardly used in trust region methods. It requires derivative information and can therefore not be used for the expensive function f_1. Instead, an interpolation model based on Lagrange polynomials is used. The interpolation points are supposed to fulfill the criterion of poisedness [2, 3] to guarantee a good interpolation model. For the implementation, function evaluations for f_1 can be saved by reusing former interpolation points if possible.

As a search direction the local ideal point $p^k = (p_1^k, p_2^k, \ldots, p_q^k)^\top$ defined by

$$p_i^k = \min_{x \in B_k} m_i^k(x) \tag{3}$$

for $i = 1, 2, \ldots, q$ is used. The trial point x^{k+}, the candidate for the next iteration point, is computed by moving into the search direction determined by the point p^k. This is realized by solving the auxiliary problem

$$\min \left\{ t \in \mathbb{R} \mid f(x^k) + t\, r^k - m^k(x) \in \mathbb{R}_+^q, x \in B_k \right\} \tag{TW}$$

with

$$r^k := f(x^k) - p^k \in \mathbb{R}_+^q$$

and p^k defined in (3). Hence, we minimize here the well-known Tammer-Weidner-functional [11, 14] which has many important properties, see for instance [17] and which turns out to be very useful in many theoretical and numerical approaches to multi-objective optimization. It is also the base of a very general scalarization in multiobjective optimization and covers several other scalarizations as special cases, see [5] and the recent review [1].

Let $(t^{k+}, x^{k+})^\top$ be a solution of (TW). If x^k is not weakly efficient for $\min_{x \in B_k} m^k(x)$ with $m^k(x) = (m_1^k(x), \ldots, m_q^k(x))^\top$, it can be proved that it holds $t^{k+} \in [-1, 0)$, see [19, Lem.3.3]. Whether x^{k+} is chosen as next iteration point or not is decided by the trial point acceptance test. For this purpose, the auxiliary functions

$$\phi(x) := \max_{i=1,\ldots,q} f_i(x) \text{ and } \phi_m^k(x) := \max_{i=1,\ldots,q} m_i^k(x) \tag{4}$$

are considered and the quotient

$$\rho_\phi^k := \frac{\phi(x^k) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})} \tag{5}$$

is computed. If it holds $\rho_\phi^k \geq \eta$, $\eta \in [0, 1]$, the trial point generates a descent for at least one objective function and it is accepted as next iteration point. The resulting trust region algorithm is stated in Algorithm 1 according to [19, Alg. 3.1].

For obtaining theoretical results for this algorithm, several assumptions are required which are listed in Appendix A. Most of them are analogous to the assumptions used in trust region approaches and in methods for expensive functions. Differences are caused by the heterogeneity of the objective functions and the search direction. The following theorem states the main result about the convergence behavior of MHT.

Algorithm 1 MHT: multi-objective heterogeneous trust region algorithm

---

   Input: Functions $f_1, f_2, \ldots, f_q$, initial point x^0, initial trust region radius δ_0, parameters $0 < \eta_1 \leq \eta_2 < 1$, $0 < \gamma_1 \leq \gamma_2 < 1$

   Step 0: Initialization

      Set $k = 0$ and compute initial model functions m_i^k for $i = 1, 2, \ldots, q$.

   Step 1: Ideal point

      Compute $p^k = (p_1^k, p_2^k \ldots, p_q^k)^\top$ by $p_i^k = \min_{x \in B_k} m_i^k(x)$ for $i = 1, 2, \ldots, q$.

   Step 2: Trial point

      If $r_i^k = f_i(x^k) - p_i^k > 0$ holds for all $i \in \{1, 2, \ldots, q\}$, compute $(t^{k+}, x^{k+})^\top$ by solving (TW) : $\min \left\{ t \in \mathbb{R} \mid f(x^k) + t\, r^k - m^k(x) \in \mathbb{R}_+^q, x \in B_k \right\}$.
      Otherwise, set $(t^{k+}, x^{k+})^\top = (0, x^k)$.

   Step 3: Trial point acceptance test

      If $t^{k+} = 0$ or $\phi_m^k(x^k) - \phi_m^k(x^{k+}) = 0$, set $\rho_\phi^k = 0$.

      Otherwise, compute $f_i(x^{k+})$, $i = 1, 2, \ldots, q$, and $\rho_\phi^k = \frac{\phi(x^k) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})}$.

      If $\rho_\phi^k \geq \eta_1$, set $x^{k+1} = x^{k+}$. Otherwise, set $x^{k+1} = x^k$.

   Step 4: Trust region update

      Set $\delta_{k+1} \in \begin{cases} [\gamma_1 \delta_k, \gamma_2 \delta_k] & \text{, if } \rho_\phi^k < \eta_1 \\ [\gamma_2 \delta_k, \delta_k] & \text{, if } \eta_1 \leq \rho_\phi^k < \eta_2 \\ [\delta_k, \infty) & \text{, if } \rho_\phi^k \geq \eta_2 \end{cases}$.

   Step 5: Model update

      Compute new model m_i^{k+1} for $i = 1, 2, \ldots, q$, set $k = k + 1$ and go to **Step 1**.

Theorem 2.4 *[19, Th. 4.21] Suppose Assumptions A.1 to A.11 hold. Then MHT (Algorithm 1) produces a sequence of iterates $\{x^k\}_k$ with $\lim_{k \to \infty} \omega(x^k) = 0$. If the sequence $\{x_k\}_k$ has accumulation points, then all these points are Pareto critical for (MOP).*

Next to the standard version of MHT as presented above, also modifications are of interest. In standard trust region methods a second order Taylor model for analytically given functions is used for making the subproblems easier to solve. For the heterogeneous problems, the evaluation of the expensive functions is such time consuming that this is not so important any more. For that reason, instead of surrogate models for the analytically given functions f_i, $i = 2, 3, \ldots, q$, it can be an advantage to directly use these functions themselves in MHT.

Assumption 2.5 *Let $m_i^k \equiv f_i$ be for all $i \in \{2, 3, \ldots, q\}$ and for all $k \in \mathbb{N}$.*

It can be proved that Theorem 2.4 still holds if Assumption 2.5 is assumed to hold, see [18, Sec. 4.6.4]. Moreover, if Assumption 2.5 holds, the constraints of (TW) imply $f_2(x^k) \leq f_2(x^{k+})$ for all iterations $k \in \mathbb{N}$.

Besides, it is possible to use a stricter trial point acceptance test in step 3 of MHT. Instead of ρ_ϕ^k, the quotients

$$\rho_i^k = \frac{f_i(x^k) - f_i(x^{k+})}{m_i^k(x^k) - m_i^k(x^{k+})} \quad \text{for} \quad i = 1, 2, \ldots, q \tag{6}$$

can be considered. The trial point x^{k+} is accepted as next iteration point x^{k+1} if it holds $\rho_i^k \geq \eta_1$ for all $i \in \{1, 2, \ldots, q\}$. In this case, the trial point provides a descent for every objective function. This realization of the trial point acceptance test is for example used in [16] and it is stricter than the trial point acceptance test used in MHT:

Lemma 2.6 *[18, Lem. 4.23] Let $k \in \mathbb{N}$ be an arbitrary index, $\eta_1 \in (0, 1)$ a constant and let Assumption A.4 hold. Furthermore, let ρ_ϕ^k and ρ_i^k, $i = 1, 2, \ldots, q$, be defined as in (5) and (6). If $\rho_i^k \geq \eta_1$ holds for all $i \in \{1, 2, \ldots, q\}$, then it holds $\rho_\phi^k \geq \eta_1$.*

The influence of the trial point acceptance test and its effects are discussed in detail in [18] where also numerical results are presented for the different versions of the acceptance test. It can be of interest to use this stricter test from an application point of view, see [15], as it guarantees a descent in all objective function in each iteration.

Furthermore, it is possible to modify the search direction of MHT and replace p^k by a lower bound

$$\tilde{p}_i^k \leq p_i^k \text{ for all } i \in \{1, 2, \ldots, q\}. \tag{7}$$

It is proved in [18, Sec. 4.6.2] that the theoretical results of MHT still hold if this modification is done, i.e. Theorem 2.4 holds.

3   Representation of Pareto Front

The algorithm MHT is designed to compute one Pareto critical point for (MOP). This is reasonable since due to the high numerical effort associated with the expensive function f_1, approximating the whole set of efficient points is often not realizable. However, it is possible to obtain several Pareto critical points by exploiting the heterogeneity of the objective functions further. In the next subsections, we present three heuristic approaches for this which are motivated by ideas for bi-objective optimization problems. Thus, we consider in the following the general bi-objective optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) = \min_{x \in \mathbb{R}^n} (f_1(x), f_2(x)) \tag{BOP}$$

with $f \colon \mathbb{R}^n \to \mathbb{R}^2$, i.e. (MOP) with $q = 2$. We also describe shortly for every approach the applicability for optimization problems (MOP) with $q \geq 3$. Moreover, we discuss for every modification if the convergence results of MHT from Theorem 2.4 can be transferred. The aim is to find a representation, at least in local areas, of the Pareto front. Thereby, we aim on spreading the additional points on the Pareto front, and it is not the aim to use any continuation method, i.e. it is not the aim to find points too close to the already found Pareto critical points.

In subsections 3.1 and 3.2 two approaches are presented which are combined in subsection 3.3. The approaches are based on the assumption that the objective function f_2 is analytically given and - compared to the expensive function f_1 - minimizing or maximizing it, even globally or w.r.t. some constraints as on a trust region, can be realized with reasonable numerical effort.

For that purpose we impose Assumption 3.1. Unlike the standard version of MHT where the function $\phi \colon \mathbb{R}^n \to \mathbb{R}$ with $\phi(x) = \max_{i=1,2} f_i(x)$ is assumed to be bounded from below, see Assumption A.2, it is necessary to assume f_2 being bounded from below.

Assumption 3.1 *Let the cheap objective function $f_2 : \mathbb{R}^n \to \mathbb{R}$ of (BOP) be bounded from below and let $\min_{x \in \mathbb{R}^n} f_2(x) = \inf_{x \in \mathbb{R}^n} f_2(x)$.*

3.1 Spreading via Individual Minima

The first approach starts after one initial run of MHT with the thereby generated point $\bar{x} \in \mathbb{R}^n$. According to Theorem 2.4, this point is Pareto critical for (BOP). Starting from $\bar{x}$, further Pareto critical points are computed by spreading in a certain direction. The overall search direction is given by the function value of a global individual minimum of the cheap function f_2 which is assumed to exist according to Assumption 3.1. Moreover, minima of f_2 in certain areas are also considered.

Let $x_{glob} \in \mathbb{R}^n$ be a global individual minimum of the function f_2 defined by

$$x_{glob} \in \operatorname*{argmin}_{x \in \mathbb{R}^n} f_2(x). \tag{8}$$

It follows immediately from the definition of weak efficiency that x_{glob} is weakly efficient for (BOP). If f_2 is convex, then a local solution method can be applied to obtain x_{glob}. Otherwise, a global solution method is required.

Moreover, we consider a closed ball around $\bar{x}$ with radius $\delta > 0$ defined by

$$B_{loc} := B\left(\bar{x}, \bar{\delta}\right) = \left\{x \in \mathbb{R}^n \mid \|\bar{x} - x\| \leq \bar{\delta}\right\}. \tag{9}$$

Such balls have already be considered within the trust region method MHT. An individual minimum of function f_2 in this area is then defined by

$$x_{loc} \in \operatorname*{argmin}_{x \in B_{loc}} f_2(x). \tag{10}$$

Although no general statement about the connection of x_{loc} and Pareto critical or weakly efficient points can be made, Lemma 3.2 states some special cases. They will be used for formulating an algorithm later in this section.

Lemma 3.2 *Let $\bar{x} \in \mathbb{R}^n$ be Pareto critical for (BOP) and let x_{glob}, B_{loc}, x_{loc} be defined as in (8)–(10).*

(i) If $x_{loc} \in \operatorname{int} B_{loc}$, i.e. $\|\bar{x} - x_{loc}\| < \bar{\delta}$, then x_{loc} is locally weakly efficient for (BOP).

(ii) If $f_2\left(x_{loc}\right) = f_2\left(x_{glob}\right)$, then x_{loc} is weakly efficient for (BOP).

Proof. First, let $x_{loc} \in \operatorname{int} B_{loc}$. Then there exists a neighborhood $U(x_{loc}) \subseteq B_{loc}$ of x_{loc} and it holds $f_2\left(x_{loc}\right) \leq f_2(x)$ for all $x \in U(x_{loc})$. Thus, there exists no vector $x \in U(x_{loc})$ such that it holds $f_i(x) < f_i(x_{loc})$ for $i = 1, 2$ and statement (i) follows. Statement (ii) follows from the definition of x_{glob} and weak efficiency. $\qquad\square$

The idea of the heuristic approach of spreading is to minimize the cheap objective function f_2 on moving local areas defined analogously to B_{loc} and thereby to generate good starting points for further runs of MHT. The idea is illustrated schematically in Figure 1.
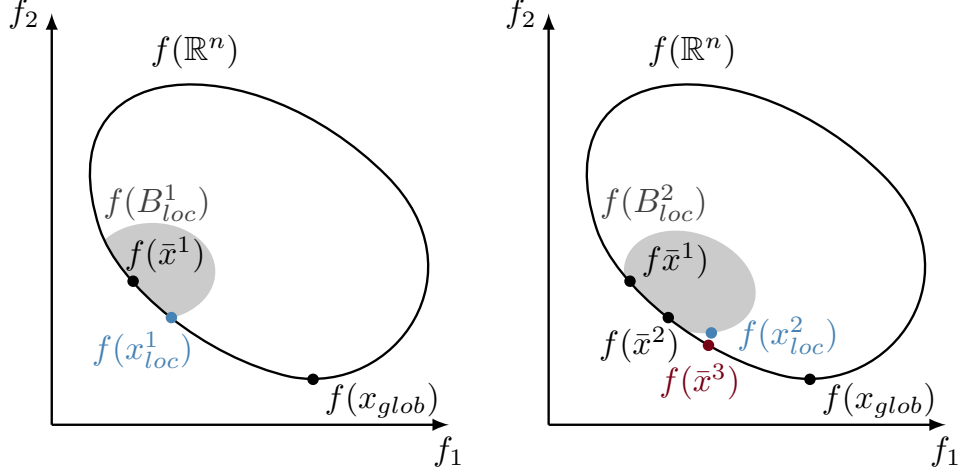
7

Figure 1: Illustration of spreading approach using individual minima of f_2

For the first step of the spreading approach, let $\bar{x}^1 \in \mathbb{R}^n$ be the result of an initial run of MHT and therefore Pareto critical for (BOP). For the illustration in Figure 1 it is depicted as an efficient point. Now, the cheap function f_2 is minimized in

$$B_{loc}^1 = B\left(\bar{x}^1, \bar{\delta}\right) = \left\{x \in \mathbb{R}^n \mid \left\|\bar{x}^1 - x\right\| \leq \bar{\delta}\right\}.$$

If the resulting point $x_{loc}^1 \in \arg\min\{f_2(x) \mid x \in B_{loc}^1\}$ lies on the boundary of B_{loc}^1 and if it holds $f_2(x_{loc}^1) \neq f_2(x_{glob})$, it is used as a starting point for a new run of MHT. Let $\bar{x}^2 \in \mathbb{R}^n$ be the point obtained by MHT, then it is a Pareto critical point for (BOP).

Otherwise, that is it holds $x_{loc}^1 \in \text{int} B_{loc}^1$ or $f_2(x_{loc}^1) = f_2(x_{glob})$, the point x_{loc}^1 is locally weakly efficient for (BOP), see Lemma 3.2, and we set $\bar{x}^2 = x_{loc}^1$. Thus, in both cases a Pareto critical point $\bar{x}^2$ is obtained. The left figure of Figure 1 illustrates the case that $f(x_{loc}^1)$ can be a nondominated point.

The same strategy can be applied again, this time to $\bar{x}^2$ by minimizing f_2 in

$$B_{loc}^2 := B\left(\bar{x}^2, \bar{\delta}\right) = \left\{x \in \mathbb{R}^n \mid \left\|\bar{x}^2 - x\right\| \leq \bar{\delta}\right\}.$$

Again, the resulting point $x_{loc}^2 \in \arg\min\{f_2(x) \mid x \in B_{loc}^2\}$ can be locally weakly efficient or a good starting point for another run of MHT. The latter case is depicted in the right figure of Figure 1. The image of the point $\bar{x}^3 \in \mathbb{R}^n$ obtained by MHT is marked red.

This spreading strategy can be iterated; the approach is described in Algorithm 2 and referred to as MHT$_{\text{spread}}$. As input, a radius $\delta_0 > 0$, the spreading distance, and a starting point x^0 are required. The point x^0 can be the result of a run of MHT or any Pareto critical point obtained otherwise. The constant δ_0 functions as a step size control since it defines the size of the areas B_{loc}^k, $k \in \mathbb{N}$, in which the individual minima of f_2 are computed.

Remark 3.3 *Let $k \in \mathbb{N}$ be an iteration of MHT$_{spread}$. It is possible that it holds $x^{k+1} = x^k$ and $f_2(x^{k+1}) > f_2(x_{glob})$, i.e. x^k is a local minimum of f_2. To avoid an infinite loop with $x^j = x^k$ for all $j \geq k + 1$, the value of δ is increased, i.e. the considered local area*

Algorithm 2 $\text{MHT}_{\text{spread}}$: Generating further Pareto critical points

---

    Input: Functions f_1 (expensive), f_2 (cheap), spreading distance δ_0, Pareto critical point x^0, e.g. as result of MHT.

    Compute $x_{glob} \in \text{argmin}_{x \in \mathbb{R}^n} f_2(x)$, set $X = \{x^0, x_{glob}\}$, $\delta = \delta_0$ and $k = 0$.

    while $f_2(x^k) > f_2(x_{glob})$ **do**

        Compute $x_{loc} \in \text{argmin}_{x \in B_{loc}^k} f_2(x)$ with $B_{loc}^k = \left\{ x \in \mathbb{R}^n \mid \left\| x^k - x \right\| \leq \delta \right\}$.

        if $\left\| x^k - x_{loc} \right\| = \delta$ and $f_2(x_{loc}) \neq f_2(x_{glob})$ **then**

            Execute MHT with x_{loc} and δ: $x^{k+1} = \text{MHT}(x_{loc}, \delta)$.

        else

            Set $x^{k+1} = x_{loc}$.

            if $\left\| x^{k+1} - x^k \right\| = 0$ **then**

                Set $\delta = 2\delta$.

            else

                Set $\delta = \delta_0$.

            end if

        end if

        Set $X = X \cup \left\{ x^{k+1} \right\}$ and $k = k + 1$.

    end while

    Output: X

---

is increased. Furthermore, the subproblems $\min_{x \in B_{loc}^j} f_2(x)$ need to be solved by a global method for every $j \geq k + 1$. This is assumed to be realizable with acceptable numerical effort compared to the numerical effort associated with the expensive function f_1. The radius δ is increased until a point y is obtained with either $y \neq x^k$ or $f_2(y) = f_2(x_{glob})$.

    The strategy of computing the individual minima in $\text{MHT}_{\text{spread}}$ is iterated until $f_2(y) = f_2(x_{glob})$ holds for a point y obtained during the spreading procedure, i.e. until a global minimum of f_2 is reached. According to Remark 3.3, an infinite loop with $x^k = x^{k+j}$ for $j \geq k$ is not possible. $\text{MHT}_{\text{spread}}$ is guaranteed to terminate after finitely many iterations if it holds

$$f_2(x^{i+1}) < f_2(x^i) \text{ for all } i = 1, 2, 3 \ldots, \tag{11}$$

i.e. if the distance to the value $f_2(x_{glob})$ decreases in every iteration. This can be guaranteed by using the strict version of the acceptance test in MHT or by setting $m_2^k \equiv f_2$ for all iterations $k \in \mathbb{N}$ in MHT.

    For every run of MHT in $\text{MHT}_{\text{spread}}$ new function evaluations for the expensive function f_1 are caused since reliable model functions are required. Potential savings can be in reusing previous interpolation points as described for MHT in section 2. Thereby and by the choice of the starting points for the runs of MHT it is expected that every further run of MHT needs only few function evaluations.

    All points generated by the heuristic search method $\text{MHT}_{\text{spread}}$ are Pareto critical for (BOP):

Lemma 3.4 *Suppose Assumptions A.1 to A.11 hold. Then every point of the set X produced by MHT_{spread} (Algorithm 2) is Pareto critical for (BOP).*

9

Proof. The proof follows from Theorem 2.4 and Lemmas 3.2 and 2.2. □

For the implementation of $\text{MHT}_{\text{spread}}$ the cheap function f_2 is not replaced by a model function, i.e. Assumption 2.5 holds. In this case, Lemma 3.4 still holds since Theorem 2.4 still holds, see [18, Sec. 4.6.4, 5.1].

The spreading technique of $\text{MHT}_{\text{spread}}$ can be transferred to multi-objective unconstrained optimization problems (MOP) with $q \geq 3$ objective functions. For this purpose, one of the cheap functions f_i, $i \in \{2, 3, \ldots, q\}$, needs to be chosen. $\text{MHT}_{\text{spread}}$ is then applied with f_i defined as the function realizing the spreading approach. The other cheap functions f_j, $j \in \{2, 3, \ldots, q\} \setminus \{i\}$, are fixed. By this, only one of the cheap functions is used for this spreading approach. To make use of all cheap functions, $\text{MHT}_{\text{spread}}$ needs to be executed with all functions f_i, $i = 2, 3, \ldots, q$, separately. Since these runs are independent from each other, they could be parallelized.

3.2 Image Space Split

Another approach to exploit the heterogeneity of the objective functions further and to compute further Pareto critical points is to use the cheap function f_2 to split the image space. Distinct search areas can be generated - in the image space - in which a modified version of MHT is applied. For all methods in this section f_2 is not replaced by a model function, that is Assumption 2.5 is fulfilled and it holds $m_2^k \equiv f_2$ for all $k \in \mathbb{N}$. As mentioned above, the convergence results of MHT and therefore Theorem 2.4 still hold for this modification.

In subsection 3.2.1 we describe how lower bounds defined in the image space can be handled in MHT. Due to these constraints the convergence results for MHT cannot be transferred. In subsection 3.2.2 we give an overview of the theoretical results that hold for this approach. In subsection 3.2.3 an approach is presented to restrict the search area in the image space after one initial run of MHT. Next, subsection 3.2.4 introduces an approach for an upfront image space split. Finally, the choice of suitable starting points is described in subsection 3.2.5.

3.2.1 MHT with Lower Bound in the Image Space

In the following we discuss how a lower bound in the image space for the cheap function f_2 of the form

$$f_2(x) \geq C \tag{12}$$

with $C \in \mathbb{R}$ in the bi-objective optimization problem (BOP) can be handled by MHT. Integrating such a nonlinear constraint poses of course additional numerical effort. However, the numerical effort is assumed to be acceptable in comparison with the expensive function f_1. To ensure that the constraint (12) is reasonable, we impose Assumption 3.5.

Assumption 3.5 *Let Assumption 3.1 hold and let $C \in \mathbb{R}$ be a constant with $C \geq \min_{x \in \mathbb{R}^n} f_2(x)$. If $\sup_{x \in \mathbb{R}^n} f_2(x) < \infty$ holds, we suppose furthermore $C \leq \sup_{x \in \mathbb{R}^n} f_2(x)$.*

Trivially, if it holds $C < \min_{x \in \mathbb{R}^n} f_2(x)$, the standard version of MHT can be used. To consider (12), several modifications are necessary in MHT. The resulting algorithm is formulated in Algorithm 3.

10

Including the lower bound $f_2(x) \geq C$ into MHT results in an additional constraint for computing the ideal point p^k in step 1 of the algorithm and a modified problem for determining the descent direction in step 2. All other steps in MHT remain unchanged. Let $k \in \mathbb{N}$ be an iteration index. The modified ideal point $\tilde{p}^k = (\tilde{p}_1^k, \tilde{p}_2^k)^\top$ is defined by

$$\tilde{p}_1^k := \min \left\{ m_1^k(x) \mid x \in B_k \right\} = p_1^k, \tag{13}$$

$$\tilde{p}_2^k := \min \left\{ f_2(x) \mid x \in B_k \text{ and } f_2(x) \geq C \right\} = \max \left\{ p_2^k, C \right\}. \tag{14}$$

It is not necessary to integrate the additional constraint $f_2(x) \geq C$ for $\tilde{p}_1^k$ since it is defined by the function m_1^k. Furthermore, the aim is to be as close as possible to the original version of MHT and therefore $\tilde{p}_1^k$ is defined as p_1^k. If $C \leq p_2^k = \min_{x \in B_k} f_2(x)$ holds, then it follows $\tilde{p}_2^k = p_2^k$ and $\tilde{p}^k = p^k$. If $C > \min_{x \in B_k} f_2(x)$ holds, i.e. C is defined as a non-trivial lower bound, the additional constraint induced by (12) affects the ideal point and therefore also the computation of the trial point.

The modified version of the problem (TW) is given by

$$\min \left\{ t \in \mathbb{R} \mid f(x^k) + t\tilde{r}^k - m^k(x) \in \mathbb{R}_+^2, x \in B_k, f_2(x) \geq C \right\} \tag{$\widetilde{TW}_C$}$$

with $\tilde{r} = f(x^k) - \tilde{p}^k \in \mathbb{R}_+^2$ and $\tilde{p}^k$ as defined in (13) and (14). Hence, we make use of the Tammer-Weidner-functional also in this modification. The case of a nontrivial lower bound C is illustrated schematically in Figure 2. The search area restricted in the image space is depicted as gray shaded area. The figure illustrates the different search directions r^k and $\tilde{r}^k$ obtained by the different ideal points p^k and $\tilde{p}^k$.
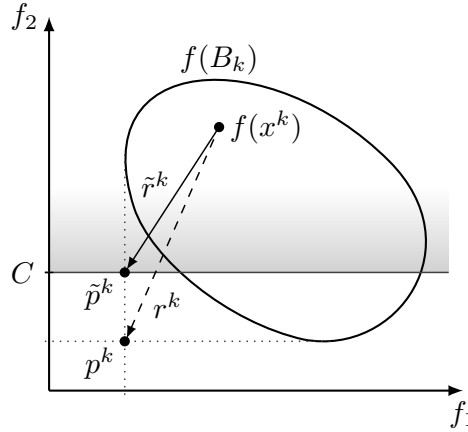

Figure 2: Restricted search area and modified ideal point

The constraint $f_2(x) \geq C$ needs to be integrated explicitly into $(\widetilde{TW}_C)$ to ensure $f_2(\bar{x}) \geq C$ for all minimal solutions $(\bar{t}, \bar{x})^\top \in \mathbb{R}^{1+n}$ of $(\widetilde{TW}_C)$. It is not sufficient to adapt only the determination of the ideal point as described in (13) and (14), see e.g. [18, Ex. 5.5]. The algorithm resulting from including this additional constraint into MHT is given in Algorithm 3 and referred to as MHT$_{\text{lb}}$.

As in MHT for the implementation of MHT$_{\text{lb}}$ the model function m_1^k is updated in iteration k only if necessary. If it holds $\rho_\phi^k < \eta_1$ in the trial point acceptance test, the

---

Algorithm 3 MHT_{lb}: MHT with lower bound in image space

Input: Functions f_1 (expensive), f_2 (cheap), lower bound $C \in \mathbb{R}$ with $C \geq \min_{x \in \mathbb{R}^n} f_2(x)$, initial point x^0 with $f_2(x^0) \geq C$, initial trust region radius δ_0, parameters $0 < \eta_1 \leq \eta_2 < 1$, $0 < \gamma_1 \leq \gamma_2 < 1$

Step 0: Initialization

Set $k = 0$, compute initial model m_1^k and set $m_2^k \equiv f_2$.

Step 1: Ideal point

Compute $\tilde{p}^k = (\tilde{p}_1^k, \tilde{p}_2^k)^\top$ and $p^k = (p_1^k, p_2^k)^\top$:

$\tilde{p}_1^k = p_1^k = \min\{m_1^k(x) \mid x \in B_k\}$,

$p_2^k = \min\{f_2(x) \mid x \in B_k\}$,

$\tilde{p}_2^k = \max\{p_2^k, C\}$

and set $\tilde{r}^k = f(x^k) - \tilde{p}^k$ and $r^k = f(x^k) - p^k$.

Step 2: Trial point

If $r_i^k > 0$ holds for $i = 1, 2$, compute $(t^{k+}, x^{k+})^\top$ by solving $(\widetilde{TW}_C)$:

$\min\{t \in \mathbb{R} \mid f(x^k) + t\tilde{r}^k - m^k(x) \in \mathbb{R}_+^2, x \in B_k, f_2(x) \geq C\}$.

Otherwise, set $(t^{k+}, x^{k+})^\top = (0, x^k)$.

Step 3: Trial point acceptance test

If $t^{k+} = 0$ or $\phi_m^k(x^k) - \phi_m^k(x^{k+}) = 0$, set $\rho_\phi^k = 0$.

Otherwise, compute $f_1(x^{k+})$ and $\rho_\phi^k = \frac{\phi(x^k) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})}$.

If $\rho_\phi^k \geq \eta_1$, set $x^{k+1} = x^{k+}$, else set $x^{k+1} = x^k$.

Step 4: Trust region update

$$\text{Set } \delta_{k+1} \in \begin{cases} [\gamma_1\delta_k, \gamma_2\delta_k] & , \text{ if } \rho_\phi^k < \eta_1 \\ [\gamma_2\delta_k, \delta_k] & , \text{ if } \eta_1 \leq \rho_\phi^k < \eta_2 . \\ [\delta_k, \infty) & , \text{ if } \rho_\phi^k \geq \eta_2 \end{cases}$$

Step 5: Model update

Compute new model m_1^{k+1}, set $m_2^{k+1} \equiv f_2$, set $k = k + 1$ and go to **Step 1**.

---

model function m_1^k is recomputed, otherwise the old model is reused in the next iteration. Note that for the computation of the interpolation points for m_1^k the additional constraint $f_2(x) \geq C$ is not considered; the interpolation points are still chosen from the whole trust region B_k in every iteration $k \in \mathbb{N}$.

Moreover, note that the criterion for considering $(\widetilde{TW}_C)$ in step 2 is the same as in step 2 of MHT. Thus, it is only tested if it holds $\tilde{r}_1^k = r_1^k = 0$. In this case, it follows $f_1(x^k) = \tilde{p}_1^k = p_1^k$ and x^k is weakly efficient for $\min_{x \in B_k} m^k(x)$. Analogous to MHT, $(\widetilde{TW}_C)$ is not considered in this case in step 2 of MHT_{lb}. Instead, $t^{k+} = 0$ and $x^{k+} = x^k$ is set directly.

In addition to that, if $r_i^k > 0$ holds for $i = 1, 2$, $(\widetilde{TW}_C)$ is considered. Thus, $(\widetilde{TW}_C)$ is also considered if it holds $\tilde{r}_1^k > 0$ and $\tilde{r}_2^k = 0$ since the latter does not imply that x^k is weakly efficient for $\min_{x \in B_k} m^k(x)$. Instead, it can be possible to obtain a trial point which provides a decrease for m_1^k, see Lemma 3.8 in subsection 3.2.2.

3.2.2 Theoretical Results

The convergence results for MHT are not transferable for $\mathrm{MHT_{lb}}$ due to the additional constraint, i.e. Theorem 2.4 does not hold. Still, some properties can be proved which show that the approach is reasonable. We impose Assumptions 2.5 and 3.5, i.e. it holds $m_2^k \equiv f_2$ for all iterations $k \in \mathbb{N}$ and f_2 is bounded from below and a constant $C \in \mathbb{R}$ is given with $\min_{x \in \mathbb{R}^n} f_2(x) \leq C$. Two aspects prevent the applicability of the theoretical results for MHT: the modified ideal point $\tilde{p}^k$ and the additional constraint $f_2(x) \geq C$ in the problem $(\widetilde{TW}_C)$. For the modified version of the problem $(\widetilde{TW}_C)$ analogous results to the results for (TW) hold, see [19, Lem. 3.2,3.3]. We summarize the most important results in Lemma 3.6. For this purpose, we define for $k \in \mathbb{N}$ the optimization problem

$$\min \left\{ m^k(x) \mid x \in B_k \text{ and } f_2(x) \geq C \right\}. \qquad (BOP_{m,C}^k)$$

Lemma 3.6 *Let Assumptions 2.5, 3.5 and A.4 hold and let $k \in \mathbb{N}$ be an iteration of MHT_{lb}. Consider the problem $(\widetilde{TW}_C)$ with $\tilde{r}^k = f(x^k) - \tilde{p}^k \in \mathbb{R}_+^2$ and $\tilde{p}^k$ as defined in (13) and (14). Let $(\bar{t}, \bar{x})^\top \in \mathbb{R}^{1+n}$ be a minimal solution of $(\widetilde{TW}_C)$.*

(i) It holds $\bar{t} \leq 0$.

(ii) It holds $m_1^k(\bar{x}) \leq m_1^k(x^k)$ and $f_2(\bar{x}) \leq f_2(x^k)$.

(iii) If x^k is not weakly efficient for $(BOP_{m,C}^k)$, then it holds $\tilde{r}_i^k > 0$ for $i = 1, 2$ and $\bar{t} \in [-1, 0)$.

(iv) If x^k is weakly efficient for $(BOP_{m,C}^k)$ and $\tilde{r}^k \in int(\mathbb{R}_+^2)$, then $(0, x^k)$ is a minimal solution of $(\widetilde{TW}_C)$.

(v) If x^k is efficient for $(BOP_{m,C}^k)$ and $\tilde{r}^k \neq 0$, then $(0, x^k)$ is a minimal solution of $(\widetilde{TW}_C)$.

Proof. Statement (i) follows since $(0, x^k)^\top$ is feasible for $(\widetilde{TW}_C)$. Since $\tilde{r}^k \in \mathbb{R}_+^2$ holds for all $k \in \mathbb{N}$, it follows $m_i^k(x^k) - m_i^k(\bar{x}) \geq -\bar{t} \tilde{r}_i^k \geq 0$ for $i = 1, 2$. This implies together with Assumption 2.5 statement (ii). To prove statement (iii), let x^k be not weakly efficient for $(BOP_{m,C}^k)$. Then there exists a vector $x \in B_k$ with $f_2(x) \geq C$ such that it holds $m_i^k(x) < m_i^k(x^k)$ for $i = 1, 2$. This implies $\tilde{r}_i^k > 0$ for $i = 1, 2$. The rest of the proof is analogous to the proof of [19, Lem.3.3]. Statements (iv) and (v) follow from [6, Th. 2.1]. $\qquad \square$

This lemma shows that if the current iteration point x^k is not weakly efficient for the trust region subproblem $(BOP_{m,C}^k)$, a minimal solution $(t^{k+}, x^{k+})^\top \in \mathbb{R}^{1+n}$ of $(\widetilde{TW}_C)$ with $t^{k+} < 0$ is computed in step 2 of $\mathrm{MHT_{lb}}$. Thus, it also holds $x^{k+} \neq x^k$. If the approximation of the model function m_1^k is accurate enough, this trial point is accepted as next iteration point and it holds $x^{k+1} \neq x^k$. However, no general statement about the outcome of the trial point acceptance test in step 3 of $\mathrm{MHT_{lb}}$ can be made.

Besides these general properties of $(\widetilde{TW}_C)$ which are analogous to the properties of (TW) of the original version of MHT, in some iterations the modified ideal point $\tilde{p}^k$ can

be replaced by a vector y with $y \leq p^k$, i.e. a lower bound for the original ideal point. This surrogate vector is constructed in Lemma 3.7. In MHT, this modification is permitted and does not affect the convergence results. Therefore, for some iterations, a stronger connection to the original algorithm MHT respectively a modification of it which does not affect the convergence results can be made since the only difference then is the additional constraint $f_2(x) \geq C$ in $(\widetilde{TW}_C)$.

Lemma 3.7 *Let Assumptions 2.5 and 3.5 hold and let $C > \min_{x \in \mathbb{R}^n} f_2(x)$. Furthermore, let $k \in \mathbb{N}$ be an iteration of MHT_{lb} with $\tilde{r}_i^k > 0$ for $i = 1, 2$. We define*

$$\hat{p}^k := f(x^k) + \hat{\lambda} \tilde{r}^k \text{ with } \hat{\lambda} := -\frac{f_2(x^k) - p_2^k}{f_2(x^k) - \tilde{p}_2^k} \qquad (15)$$

with p^k and $\tilde{p}^k$ the ideal points of (TW) and $(\widetilde{TW}_C)$. Moreover, we define the problem

$$\min \left\{ t \in \mathbb{R} \mid f(x^k) + t\,\hat{r}^k - m^k(x) \in \mathbb{R}_+^2, x \in B_k, f(x) \geq C \right\} \qquad (T\hat{W})$$

with $\hat{r}^k := f(x^k) - \hat{p}^k$. The following statements hold:

(i) It holds $\hat{\lambda} \leq -1$, $\hat{p}^k \leq p^k$, and $\hat{r}^k = -\hat{\lambda} \tilde{r}^k$.

(ii) Let $(\bar{t}, \bar{x}) \in \mathbb{R}^{1+n}$ be a minimal solution of $(\widetilde{TW}_C)$. Then $(-\bar{t}/\hat{\lambda}, \bar{x})$ is a minimal solution of $(T\hat{W})$.

(iii) Let $(\bar{t}, \bar{x}) \in \mathbb{R}^{1+n}$ be a minimal solution of $(T\hat{W})$. Then $(-\hat{\lambda}\,\bar{t}, \bar{x})$ is a minimal solution of $(\widetilde{TW}_C)$.

The construction of the vector $\hat{p}^k$ out of p^k and $\tilde{p}^k$ is illustrated schematically in Figure 3. The search area generated by $f_2(x) \geq C$ is depicted as gray shaded area.
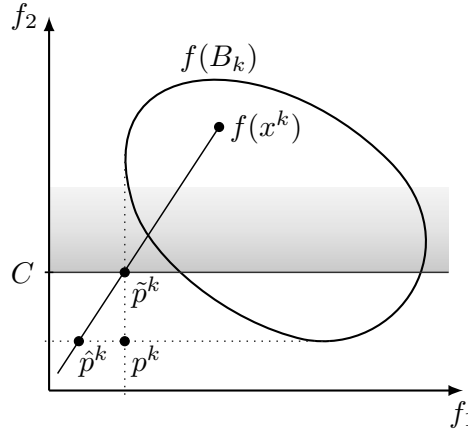

Figure 3: Modified ideal points $\hat{p}^k$ and $\tilde{p}^k$

14

Proof. Let $k \in \mathbb{N}$ be an arbitrary iteration. To prove item (i), note that it holds $\tilde{p}_2^k \geq p_2^k$. Together with the precondition $\tilde{r}_2^k > 0$ this implies $0 < f_2(x^k) - \tilde{p}_2^k \leq f_2(x^k) - p_2^k$ and therefore it holds $\hat{\lambda} \leq -1$. Due to the definition of $\hat{p}^k$ in (15) it holds

$$\hat{p}_1^k = f_1(x^k) + \hat{\lambda}\left(f_1(x^k) - \tilde{p}_1^k\right) \leq f_1(x^k) - \left(f_1(x^k) - \tilde{p}_1^k\right) = \tilde{p}_1^k = p_1^k \text{ and}$$

$$\hat{p}_2^k = f_2(x^k) - \frac{f_2(x^k) - p_2^k}{f_2(x^k) - \tilde{p}_2^k}\left(f_2(x^k) - \tilde{p}_2^k\right) = p_2^k.$$

It follows $\hat{p}^k \leq p^k$ and, furthermore, it holds

$$\hat{r}^k = f(x^k) - \hat{p}^k = f(x^k) - \left(f(x^k) + \hat{\lambda}\tilde{r}^k\right) = -\hat{\lambda}\tilde{r}^k.$$

To prove statement (ii), let $(\bar{t}, \bar{x}) \in \mathbb{R}^{1+n}$ be a minimal solution of $(\widetilde{TW}_C)$. Then it holds $\bar{x} \in B_k$ and $f(\bar{x}) \geq C$, and together with (i) it follows

$$0 \leq f(x^k) + \bar{t}\tilde{r}^k - m^k(\bar{x}) = f(x^k) - \frac{\bar{t}}{\hat{\lambda}}\hat{r}^k - m^k(\bar{x}).$$

Thus, $(-\bar{t}/\hat{\lambda}, \bar{x})$ is feasible for $(\hat{T}W)$. Note that it holds $\hat{\lambda} \leq -1$. Assume there exists a vector (t, x) feasible for $(\hat{T}W)$ with $t < -\bar{t}/\hat{\lambda}$. Then it holds again due to (i)

$$0 \leq f(x^k) + t\hat{r}^k - m^k(x) = f(x^k) - t\hat{\lambda}\tilde{r}^k - m^k(x).$$

and $(-t\hat{\lambda}, x)$ is feasible for $(\widetilde{TW}_C)$ with $-t\hat{\lambda} < \bar{t}$. This contradicts $(\bar{t}, \bar{x})$ being a minimal solution of $(\widetilde{TW}_C)$. Therefore, $(-\bar{t}/\hat{\lambda}, \bar{x})$ is a minimal solution of $(\hat{T}W)$. Statement (iii) follows analogously to statement (ii). $\qquad\square$

Given Lemma 3.7, in some iterations it is possible to circumvent the difficulty caused by the modified trial point. Still, the additional constraint $f_2(x) \geq C$ needs to be considered in $(\widetilde{TW}_C)$. Therefore, the second aspect preventing the convergence results from being transferred is still given.

The most important precondition of Lemma 3.7 is $\tilde{r}_2^k > 0$. This is fulfilled if it holds

$$f_2(x^k) > \tilde{p}_2^k = \min\left\{f_2(x) \mid x \in B_k, f_2(x) \geq C\right\} = \max\left\{p_2^k, C\right\},$$

i.e. if it holds $f_2(x^k) > C$ and x^k is not an individual minimum of f_2 in B_k. In this case and if $r_1^k > 0$, the vector $\tilde{p}^k$ can be replaced by $\hat{p}^k$.

If $\tilde{r}_2^k = 0$ holds, then either the constraint $f_2(x) \geq C$ is active or it holds $f_2(x^k) = p_2^k$. The latter implies x^k being weakly efficient for $\min_{x \in B_k} m^k(x)$. In this case, it holds $r_2^k = 0$ and the trial point is defined as $x^{k+} = x^k$, see step 2 of MHT_{lb}. Thus, in this case the additional constraint $f_2(x) \geq C$ has no effect.

If $f_2(x^k) = C$ holds in iteration k, i.e. the additional constraint is active, some general properties for the subsequent iterations can be proved:

Lemma 3.8 *Let Assumptions 2.5, 3.5 and A.4 and $C > \min_{x \in \mathbb{R}^n} f_2(x)$ hold. Let $k \in \mathbb{N}$ be the first iteration of MHT_{lb} (Algorithm 3) with $f_2(x^k) = C$. Moreover, let $r_i^k > 0$ hold for $i = 1, 2$ and let $(t^{k+}, x^{k+}) \in \mathbb{R}^{1+n}$ be the minimal solution of $(\widetilde{TW}_C)$ computed in iteration k. The following statements hold:*

(i) *For every feasible vector $(t, x)^\top \in \mathbb{R}^{1+n}$ of $(\widetilde{TW}_C)$ it holds $f_2(x) = C$.*

(ii) *It holds $f_2(x^l) = f_2(x^k) = C$ for all $l \geq k$.*

(iii) *If there exists a vector $x \in B_k$ with*

$$f_2(x) = C \quad and \quad m_1^k(x) < m_1^k(x^k), \tag{16}$$

then it holds $t^{k+} < 0$ and $x^{k+} \neq x^k$.

(iv) *If for all vectors $x \in B_k \setminus \{x^k\}$ it holds*

$$f_2(x) > f_2(x^k) = C \quad or \quad m_1^k(x) \geq m_1^k(x^k), \tag{17}$$

then it holds $x^{k+1} = x^k$.

Proof. Consider the problem $(\widetilde{TW}_C)$ with $\tilde{r}^k = f(x^k) - \tilde{p}^k$ and $\tilde{p}^k$ defined in (13) and (14). Since it holds $f_2(x^k) = C$, this implies $\tilde{p}_2^k = C$ and $\tilde{r}_2^k = 0$. According to Assumption 2.5, it holds $m_2^k \equiv f_2$ for all $k \in \mathbb{N}$. Thus, it holds for every feasible vector $(t, x)^\top \in \mathbb{R}^{1+n}$ of $(\widetilde{TW}_C)$

$$0 \leq f_2(x^k) + t\,\tilde{r}_2^k - m_2^k(x) = C - f_2(x).$$

Since it also holds $f_2(x) \geq C$, this implies $f_2(x) = C$.

To prove statement (ii), note that x^{k+1} is either defined as x^k or as the trial point x^{k+}. According to statement (i), it holds $f_2(x^{k+}) = C$. Thus, in both cases $f_2(x^{k+1}) = C$ follows. With this line of argument statement (ii) follows by induction.

To prove statement (iii), let $x \in B_k$ be with $f_2(x) = C$ and $m_1^k(x) < m_1^k(x^k)$. According to the proof of (i), it holds $\tilde{r}_2^k = 0$ and $f_2(x^k) + t\,\tilde{r}_2^k - m_2^k(x) \geq 0$ for all $t \in \mathbb{R}$. According to Assumption A.4, it holds

$$f_1(x^k) - m_1^k(x) = m_1^k(x^k) - m_1^k(x) > 0.$$

Since $\tilde{r}_1^k > 0$ holds, there exists a vector $t \in (-\infty, 0)$ such that (t, x) is feasible for $(\widetilde{TW}_C)$. Then it follows $t^{k+} < 0$ and $(0, x^k)^\top$ cannot be a minimal solution of $(\widetilde{TW}_C)$. This implies $x^{k+} \neq x^k$.

Now let the preconditions of statement (iv) be fulfilled. Note that $(0, x^k)$ is always feasible for $(\widetilde{TW}_C)$. Now let $x \in B_k \setminus \{x^k\}$ be arbitrarily chosen. If the first part of (17) is fulfilled then by (i) there exists no $t \in \mathbb{R}$ such that (t, x) is feasible. If the second part of (17) is fulfilled, it holds $m_1^k(x^k) - m_1^k(x) \leq 0$. Since $\tilde{r}_1^k > 0$ holds, it follows $t \geq 0$ for all (t, x) feasible for $(\widetilde{TW}_C)$. This implies $t^{k+} = 0$ for the minimal value in both cases. According to the trial point acceptance test in step 3 of MHT_{lb}, it holds $x^{k+1} = x^k$ in both cases. $\square$

This lemma shows that whenever the additional constraint $f_2(x) \geq C$ is active for an iteration $k \in \mathbb{N}$, it is active for all subsequent iterations $l \geq k$. However, this does not imply $x^l = x^k$ for all $l \geq k$. In some cases, a trial point x^{k+} distinct from x^k can still be obtained by $(\widetilde{TW}_C)$. Whether it is accepted as next iteration point depends on the quality of the model function and the trial point acceptance test.

Furthermore, in the case (iv) of Lemma 3.8 it holds $t^{k+} = 0$ and therefore $\rho_\phi^k = 0$. In this case, the model function m_1^k is updated for iteration $k + 1$ and a new model m_1^{k+1} is computed. Therefore, the precondition of (iv) does not need to be fulfilled in iteration $k + 1$. Thus, no general statement is possible on whether the iteration point will change in subsequent iterations, since it depends on the quality of the model function for f_1.

For the considerations in the following subsections also constraints of the form $f_2(x) \leq D$ with $D \in \mathbb{R}$ are considered. Analogous to MHT$_{\text{lb}}$, a modification of MHT could be formulated that handles such a constraint. However, this is not necessary since it is fulfilled for all iteration points of MHT if the starting point x^0 fulfills it and if Assumption 2.5 holds, i.e. $f_2 \equiv m_2^k$ for all $k \in \mathbb{N}$. This is stated in the following two lemmata.

Lemma 3.9 *Consider MHT (Algorithm 1) applied to (BOP) and let Assumptions 3.1, 2.5 and A.4 hold. Let $D \in \mathbb{R}$ be a scalar with $\min_{x \in \mathbb{R}^n} f_2(x) \leq D$ and let $x^0 \in \mathbb{R}^n$ be the starting point of MHT with $f_2(x^0) \leq D$. Then it holds $f_2(x^k) \leq D$ for all iterations $k \in \mathbb{N}$.*

Proof. The statement follows by induction from Assumption 2.5, the constraints of (TW) and from the fact that it holds $t^{k+} \leq 0$ for every solution $(t^{k+}, x^{k+})^\top$ of (TW). □

Lemma 3.10 *Consider MHT$_{\text{lb}}$ (Algorithm 3) applied to (BOP) and let Assumptions 2.5, 3.5 and A.4 hold. Let $D \in \mathbb{R}$ be a scalar with $\min_{x \in \mathbb{R}^n} f_2(x) \leq D$ and $C \leq D$ with C from Assumption 3.5. Moreover, let $(t^{k+}, x^{k+})^\top \in \mathbb{R}^{1+n}$ be the minimal solution of $(\widetilde{TW}_C)$ for every iteration $k \in \mathbb{N}$ with $r_i^k > 0$, $i = 1, 2$. Let $x^0 \in \mathbb{R}^n$ be the starting point of MHT$_{\text{lb}}$ with $f_2(x^0) \leq D$. Then it holds $f_2(x^k) \leq D$ for all iterations $k \in \mathbb{N}$.*

Proof. The statement follows by induction from Assumption 2.5, the constraints of $(\widetilde{TW}_C)$ and Lemma 3.6. □

3.2.3 Subsequent Image Space Split

Now we use the above results to generate further Pareto critical points. Let $\bar{x} \in \mathbb{R}^n$ be a Pareto critical point for (BOP). The aim is to find additional points which are not too close to the current point $f(\bar{x})$, but which spread over the Pareto front. MHT resp. MHT$_{\text{lb}}$ can be used to compute such Pareto critical points avoiding the area around $f(\bar{x})$. The idea is to subsequently split the image space into two search areas defined by

$$A^+ := \{x \in \mathbb{R}^n \mid f_2(x) \geq f_2(\bar{x}) + \varepsilon\} \text{ and}$$
$$A^- := \{x \in \mathbb{R}^n \mid f_2(x) \leq f_2(\bar{x}) - \varepsilon\} \tag{18}$$

with $\varepsilon > 0$ a suitable positive constant. The two search areas A^+ and A^- are schematically illustrated in Figure 4. They are depicted as gray shaded areas.

The search for another Pareto critical point in the area A^+ can – at least heuristically – be realized by applying MHT$_{\text{lb}}$ with the lower bound $C = f_2(\bar{x}) + \varepsilon$. The problem of finding a well located starting point for executing MHT$_{\text{lb}}$ is addressed in subsection 3.2.5. To realize a search for further Pareto critical points in the area A^-, MHT can be applied with a starting point $x^0 \in A^-$. According to Lemma 3.9, it then holds $f_2(x^k) \leq f_2(\bar{x}) - \varepsilon$ for all iterations $k \geq 1$ in MHT. Thus, the additional constraint induced by the search area A^- does not need to be handled explicitly, but is inherently fulfilled for all iteration
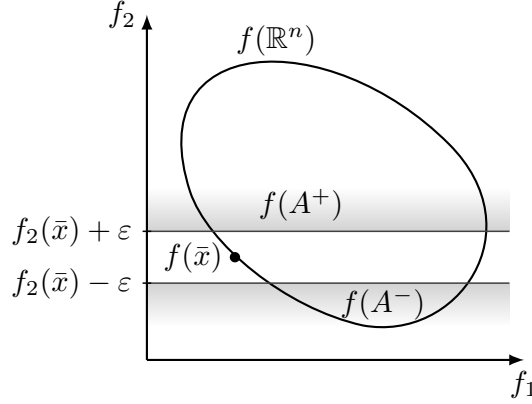
Figure 4: Search areas A^+ and A^- in the image space

points x^k produced by MHT. Figure 4 also shows that a so-called taboo-region is generated around $f(\bar{x})$. This area is ignored for runs with MHT_{lb} respectively MHT in A^+ and A^-. By varying the size of the constant ε the size of the taboo-region can be varied.

The approach of subsequently splitting the image space after one run of MHT is also applicable for optimization problems (MOP) with $q \geq 3$ objective functions. Let again $\bar{x} \in \mathbb{R}^n$ be the result of an initial run of MHT. It is possible to choose one function f_j, $j \in \{2, 3, \ldots, q\}$, and define the search areas A^+ and A^- as in (18) with this function f_j. Alternatively, it is possible to consider all cheap functions. For this purpose, define areas A_i, $i \in \{2, 3, \ldots, q\}$, either as

$$A_i = A_i^+ := \{x \in \mathbb{R}^n \mid f_i(x) \geq f_i(\bar{x}) + \varepsilon\} \text{ or}$$
$$A_i = A_i^- := \{x \in \mathbb{R}^n \mid f_i(x) \leq f_i(\bar{x}) - \varepsilon\}$$

with $\varepsilon > 0$ a suitable positive constant. The set $A := \bigcap_{i=2}^{q} A_i$ is then the search area for a new run of MHT_{lb}. In case $A_i = A_i^-$ for all $i \in \{2, 3, \ldots, q\}$, MHT can be applied with a starting point $x^0 \in A$. Analogously to Lemma 3.9, it can be proved that then the upper bounds do not need to be considered explicitly, but are fulfilled for every iteration point of MHT. If it holds $A_i = A_i^+$ for some indices $i \in \{2, 3, \ldots, q\}$, the induced lower bounds need to be integrated into the algorithm analogous to MHT_{lb}.

3.2.4 Upfront Image Space Split

It is also possible to split the image space into several disjoint search areas before applying any version of MHT. For defining the search areas, we assume that f_2 is bounded from below and from above. In this case we define $f_2^{\min}$ and $f_2^{\max}$ by

$$f_2^{\min} := \min_{x \in \mathbb{R}^n} f_2(x) \quad \text{and} \quad f_2^{\max} := \max_{x \in \mathbb{R}^n} f_2(x). \tag{19}$$

Again, it needs to be regarded that these optimization problems are global problems. If f_2 is not bounded from above, $f_2^{\max}$ can be defined as $f_2^{\max} := E \in \mathbb{R}$ with $E > f_2^{\min}$. $f_2^{\min}$ and $f_2^{\max}$ give an upper and a lower bound for the search area in the image space. A

number p of distinct search areas A_j with $j = 1, 2, \ldots, p$ can then be obtained by defining

$$A_j := \{x \in \mathbb{R}^n \mid a_{j-1} \leq f_2(x) \leq a_j\} \tag{20}$$

$$\text{with } a_0 = f_2^{\min} \text{ and } a_j = f_2^{\min} + j \frac{f_2^{\max} - f_2^{\min}}{p} \tag{21}$$

This is illustrated schematically in Figure 5 for different values of p. Figure 5 indicates that not in every search area A_j, $j = 1, 2, \ldots, p$, a weakly efficient point needs to exist.
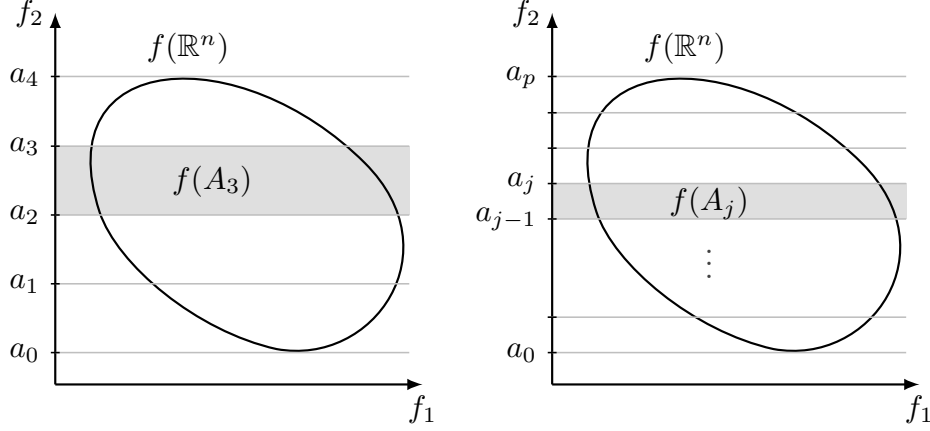

Figure 5: Image space splits with different values for p

To realize the splitting idea, the optimization problems

$$\min_{x \in A_j} f(x) = \min_{x \in A_j} (f_1(x), f_2(x))^\top \tag{BOP_j}$$

for $j = 1, 2, \ldots, p$ need to be considered. Thus, MHT_{lb} is executed for every optimization problem (BOP_j), $j = 1, 2, \ldots, p$. The location of the starting point for MHT_{lb} is important; an approach to generate well located starting points for the areas A_j, $j \in \{1, 2, \ldots, p\}$, is presented in subsection 3.2.5. The numerical tests confirm that the splitting approach described in this subsection is a useful heuristic.

MHT_{lb} needs as input a lower bound C. It is given by the set A_j as $C = C_j = a_{j-1}$ for all $j \in \{1, 2, \ldots, p\}$. Note that the upper bounds given in A_j do not need to be included explicitly if $x^0 \in A_j$ holds, see Lemmas 3.9 and 3.10. Moreover, MHT can be applied for $j = 1$, since the lower bound $a_0 = f_2^{\min}$ is trivially fulfilled.

As the examples for the image space splits depicted in Figure 5 illustrate, not in every area A_j, $j \in \{1, 2, \ldots, p\}$, a weakly efficient point for (BOP) needs to exist. Also, not in every area A_j a Pareto critical point for (BOP) needs to exist. To avoid unnecessary function evaluations and runs of MHT_{lb} in areas not containing Pareto critical points for (BOP), a heuristic strategy is used in our implementation. The runs of MHT_{lb} are executed consecutively and the search areas are considered in the order $A_1, A_2, \ldots, A_p$. This is illustrated in Figure 6 with $B_i = f(A_i), i = 1, 2, \ldots, p$.

At first, (BOP_1) and A_1 are considered. Since the constraint $f_2(x) \geq a_0$ in A_1 is trivially fulfilled, MHT can be applied to (BOP_1). Let $\bar{x}^1 \in \mathbb{R}^n$ be the result of this
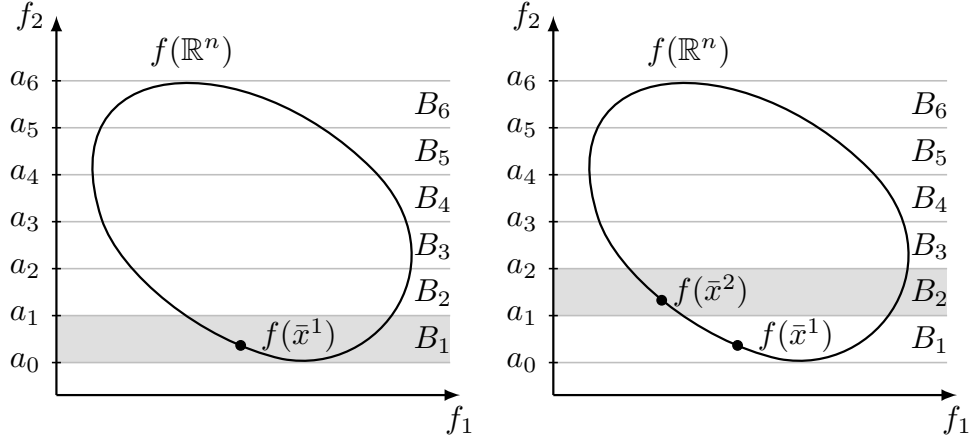
19

Figure 6: MHT resp. MHT_{lb} applied in A_1 and A_2

run. The search areas A_j are then considered in ascending order of j. Thus, in the subsequent runs search areas with increasing distance to a_0, the global minimal value of f_2, are considered. For all search areas A_j with $j \geq 2$ MHT_{lb} needs to be used to handle the lower bound given in A_j. Let $\bar{x}^j$ be the point obtained by MHT_{lb} applied to (BOP_j) and A_j with $j \in \{2, 3, \ldots, p\}$. It is not guaranteed that $\bar{x}^j$, $j \in \{2, 3, \ldots, p\}$, is Pareto critical for (BOP), see subsection 3.2.2. It is possible to start MHT with $\bar{x}^j$ as starting point to obtain a Pareto critical point of (BOP). However, for the implementation of the splitting approach this is not done in order to save function evaluations.

After every run of MHT_{lb} in a search area A_j, $j \in \{2, 3, \ldots, p\}$, it is tested if the computed solution $\bar{x}^j$ is dominated by the previous solutions $\bar{x}^l$, $l = 1, 2, \ldots, j - 1$, generated in the search areas A_l. In this case, the splitting procedure is terminated. This is illustrated schematically in Figure 7. We assume that no further nondominated points will be computed when considering the search areas $A_{j+1}, A_{j+2}, \ldots, A_p$ located further from the global minimum of the cheap function. This is only a heuristic strategy and gives no
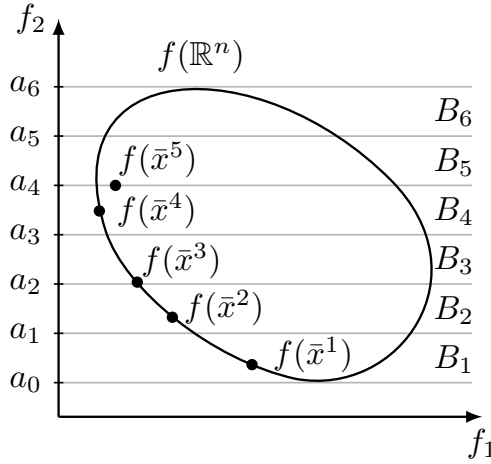

Figure 7: Heuristic strategy to terminate splitting approach

20

guarantee to only terminate the splitting approach if no Pareto critical points exist in the remaining search areas. However, in the numerical tests it worked well.

The algorithm realizing the image space split with consecutive runs of MHT_{lb} and the heuristic stopping criterion is described in Algorithm 4 and referred to as $\text{MHT}_{\text{split}}$.

Algorithm 4 $\text{MHT}_{\text{split}}$: Image space split

Input: Functions f_1 (expensive) f_2 (cheap), number p of search areas, initial trust region radius δ_0, values for the parameters $0 < \eta_1 \leq \eta_2 < 1$, $0 < \gamma_1 \leq \gamma_2 < 1$

Step 0: Initialization

   Compute $f_2^{\max} = \max_{x \in \mathbb{R}^n} f_2(x)$ and
   $x^0 \in \text{argmin}_{x \in \mathbb{R}^n} f_2(x)$ and set $f_2^{\min} = \min_{x \in \mathbb{R}^n} f_2(x)$.
   Set $X = \{x^0\}$ and $k = 1$.

Step 1: Search areas

   Define $a_i = f_2^{\min} + i \frac{f_2^{\max} - f_2^{\min}}{p}$ for $i = 0, 1, \ldots, p$.
   Define search areas $A_i = \{x \in \mathbb{R}^n \mid a_{i-1} \leq f_2(x) \leq a_i\}$ for $i = 1, 2, \ldots, p$.
   Compute starting points $x_i^0 \in A_i$ for $i = 1, 2, \ldots, p$ (subsection 3.2.5).

Step 2: Consecutive runs of MHT_{lb} (Algorithm 3)

   Compute $x^k = \text{MHT}_{\text{lb}}(x_k^0, C)$ with $C = a_{k-1}$.
   If it holds $f(x^k) \geq f(x^j)$ for an index $j \in \{1, 2, \ldots, k-1\}$, STOP.
   Otherwise, set $X = X \cup \{x^k\}$, $k = k + 1$ and go to Step 2.

Output: X

It is possible to apply this heuristic approach also for optimization problems (MOP) with $q \geq 3$. For this purpose, one of the cheap functions f_i, $i \in \{2, 3, \ldots, q\}$, needs to be chosen to apply $\text{MHT}_{\text{split}}$.

3.2.5   Choice of Starting Points

For the splitting approaches from subsections 3.2.3 and 3.2.4 suitable starting points are required. Let f_2 be bounded from below. We consider the general unconstrained bi-objective optimization problem (BOP) and the general search area

$$A := \{x \in \mathbb{R}^n \mid C \leq f_2(x) \leq D\}$$

with $C, D \in \mathbb{R}$, $\min_{x \in \mathbb{R}^n} f_2(x) \leq C \leq D$.

The aim is not only to guarantee feasibility, but also to compute a well located starting point. As outlined in Lemmas 3.9 and 3.10 in subsection 3.2.2, the constraint $f_2(x) \leq D$ does not complicate the computations in MHT respectively MHT_{lb}, only the constraint $f_2(x) \geq C$ changes the computations in the iterations. If $f_2(x^0) = C$ holds, it is possible that MHT_{lb} stops with x^0 also if it is not a Pareto critical point for (BOP), see subsection 3.2.2.

For the approaches of splitting the image space it is assumed that it holds $m_2^k \equiv f_2$ for all iterations $k \in \mathbb{N}$, see Assumption 2.5. Thus, it holds $f_2(x^{k+1}) \leq f_2(x^k)$ for all $k \in \mathbb{N}$. Considering the distance $d(k) := f_2(x^k) - C$ for an iteration $k \in \mathbb{N}$, it follows $d(k+1) \leq d(k)$ for all $k \in \mathbb{N}$. Due to the definition of the set A it holds $d(k) \leq D - C$ for

all $k \in \mathbb{N}$. We assume that the bigger the value of $d(0)$, i.e. the distance for the starting point, the more iterations in MHT_{lb} can be executed until it terminates.

Therefore, we define x^0 as a minimal solution of the optimization problem

$$\min \left\{ f_2(x) \mid x \in A \text{ and } f_2(x) \geq D - \varepsilon \right\} \qquad (22)$$

with $\varepsilon > 0$ a suitable positive constant. This generates a point x^0 with $d(0) = f_2(x^0) - C \geq D - C - \varepsilon$, i.e. with a large value for $d(0)$. Instead of minimizing f_2 in (22), it is possible to choose any function $g : \mathbb{R}^n \to \mathbb{R}$ which is easy to minimize locally, e.g. a constant function. This would generate a feasible point in A as well. By choosing f_2 in (22), x^0 minimizes one of the objective functions of (BOP) in the search area $A \cap \{x \in \mathbb{R}^n \mid f_2(x) \geq D - \varepsilon\}$. The hope is that the obtained point x^0 is closer to the set of Pareto critical points than by using any arbitrary function g in (22). The numerical experiments presented in section 4 confirm the usefulness of this approach.

3.3 Combination of Image Space Split and Spreading

The splitting and spreading techniques can be combined to a heuristic approach to approximate the set of Pareto critical points of (BOP). At first, $\text{MHT}_{\text{split}}$ from subsection 3.2.4 is applied and the image space is split into p areas $A_1, A_2, \ldots, A_p$. Let

$$\bar{X} := \left\{ \bar{x}^1, \bar{x}^2, \ldots, \bar{x}^k \right\} \qquad (23)$$

with $k \leq p$ be the set of points resulting from $\text{MHT}_{\text{split}}$. $\text{MHT}_{\text{split}}$ uses a heuristic stopping criterion to save function evaluations. As a result, it is possible that not all areas $A_1, A_2, \ldots, A_p$ are considered, but only the areas $A_1, A_2, \ldots, A_k$ up to an index $k \leq p$. This is illustrated schematically for $p = 5$ in Figure 8. The left figure shows all obtained points until the splitting approach is terminated. The point $\bar{x}^5$ is dominated by $\bar{x}^4$ and therefore not included in the list $\bar{X} = \{\bar{x}^1, \bar{x}^2, \bar{x}^3, \bar{x}^4\}$ given as output by $\text{MHT}_{\text{split}}$.
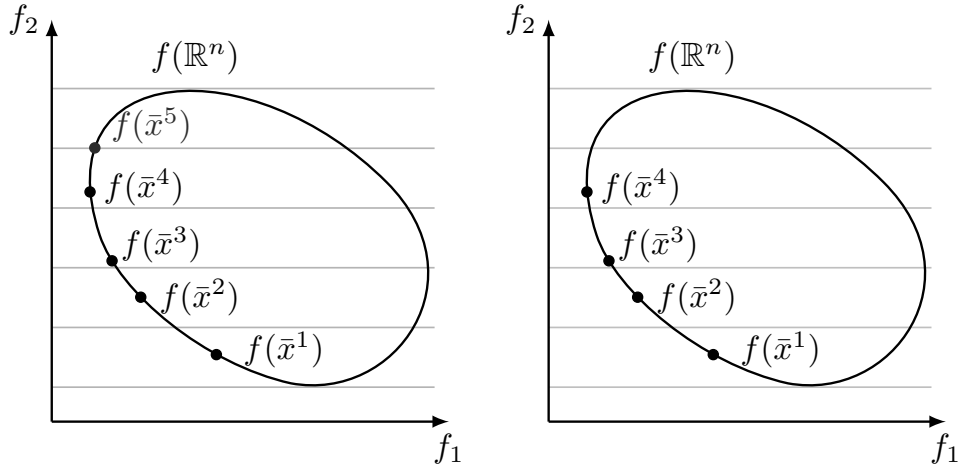

Figure 8: Step 1: Image space split with $\text{MHT}_{\text{split}}$

For all points in $\bar{X}$ the spreading technique from subsection 3.1 is then used to compute further Pareto critical points. We use a slightly modified version of $\text{MHT}_{\text{spread}}$ given in

Algorithm 5 and referred to as $\mathrm{MHT}^{\mathrm{b}}_{\mathrm{spread}}$. As in $\mathrm{MHT}_{\mathrm{spread}}$, the cheap objective function f_2 is minimized subsequently on moving local areas. However, instead of doing so until the global minimal value $f_2(x_{glob}) = \min_{x \in \mathbb{R}^n} f_2(x)$ is reached, a lower bound b with $b \geq f_2(x_{glob})$ is used. If it holds $f_2(y) = b$ for a point y obtained by the spreading approach, it is terminated.

Algorithm 5 $\mathrm{MHT}^{\mathrm{b}}_{\mathrm{spread}}$: $\mathrm{MHT}_{\mathrm{spread}}$ (Algorithm 2) on bounded area

---

Input: Functions f_1 (expensive), f_2 (cheap), spreading distance δ_0, starting point x^0, lower bound $b \in \mathbb{R}$ with $\min_{x \in \mathbb{R}^n} f_2(x) \leq b \leq f_2(x^0)$

Compute $x_{glob} \in \mathrm{argmin}_{x \in \mathbb{R}^n} f_2(x)$, set $X = \{x^0\}$, $\delta = \delta_0$ and $k = 1$.

while $f_2(x^k) > b$ **do**

    Compute $x_{loc} \in \mathrm{argmin}_{x \in B^k_{loc}} f_2(x)$ with $B^k_{loc} = \{x \in \mathbb{R}^n \mid \|x^k - x\| \leq \delta\}$.

    if $\|x^k - x_{loc}\| = \delta$ and $f_2(x_{loc}) \neq f_2(x_{glob})$ **then**

        Execute MHT with x_{loc} and δ: $x^{k+1} = \mathrm{MHT}(x_{loc}, \delta)$.

    else

        Set $x^{k+1} = x_{loc}$.

        if $\|x^{k+1} - x^k\| = 0$ **then**

            Set $\delta = 2\delta$.

        else

            Set $\delta = \delta_0$.

        end if

    end if

    Set $X = X \cup \{x^{k+1}\}$ and $k = k + 1$.

end while

Output: X

Note that in the while-loop it is possible to obtain a point x^k with $f_2(x^k) < b$ either as $x^k = x_{loc} \in \mathrm{argmin}_{x \in B^k_{loc}} f_2(x)$ or as $x^k = \mathrm{MHT}(x_{loc}, \delta)$. Nevertheless, this point is contained in the set X which is the output of $\mathrm{MHT}^{\mathrm{b}}_{\mathrm{spread}}$ since it is found to be a Pareto critical point for (BOP), see Lemma 3.2 in subsection 3.1. This is also stated in the following lemma which holds analogously to Lemma 3.4 from subsection 3.1. The proof is analogous and therefore omitted.

Lemma 3.11 *Suppose Assumptions A.1 to A.11 hold. Consider MHT^b_{spread} (Algorithm 5) with the starting point x^0. Let X be the output of MHT^b_{spread}. Then every point contained in the set $X \setminus \{x^0\}$ is Pareto critical for (BOP).*

In the overall procedure to approximate the set of Pareto critical points of (BOP) $\mathrm{MHT}^{\mathrm{b}}_{\mathrm{spread}}$ is applied for all points contained in $\bar{X}$, the output of $\mathrm{MHT}_{\mathrm{split}}$ from (23). For this purpose and without loss of generality, let the points in $\bar{X}$ be sorted according to function f_2 such that it holds

$$f_2(\bar{x}^1) \leq f_2(\bar{x}^2) \leq \ldots \leq f_2(\bar{x}^k).$$

Let $\bar{x}^j$ with $j \in \{2, 3, \ldots, k\}$ be the starting point of $\mathrm{MHT}^{\mathrm{b}}_{\mathrm{spread}}$. The lower bound b for termination is chosen as

$$b = b_j = f_2(\bar{x}^{j-1}) \tag{24}$$

given by the preceding point from $\bar{X}$. For $\bar{x}^1$ MHT$_{\text{spread}}$ is applied. Applying MHT$^{\text{b}}_{\text{spread}}$ respectively MHT$_{\text{spread}}$ to the points of $\bar{X}$ is illustrated schematically in Figure 9. All points obtained by the spreading approach are marked gray. The left figure shows the result of MHT$^{\text{b}}_{\text{spread}}$ with $x^0 = \bar{x}^2$ and $b = f_2(\bar{x}^1)$ and the right figure the result for all points of $\bar{X}$.
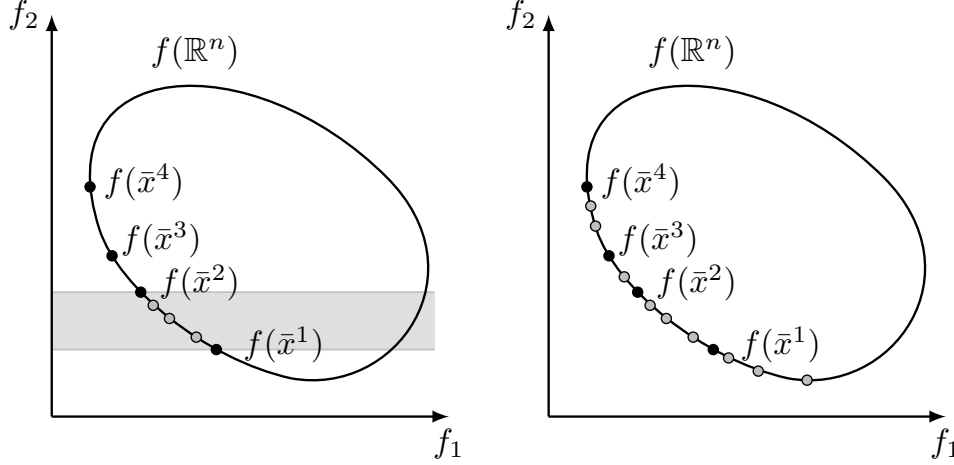

Figure 9: Step 2: Spreading with MHT$^{\text{b}}_{\text{spread}}$ resp. MHT$_{\text{spread}}$

The complete procedure combining the described steps and various runs of MHT with different modifications is given in Algorithm 6 and referred to as MHT$_{\text{combi}}$. The output of MHT$_{\text{combi}}$ are the sets $\bar{X}$ and Y. According to Lemma 3.11, all points contained in $Y \setminus \bar{X}$ are Pareto critical for (BOP). Since Pareto criticality is only a necessary condition for local weak efficiency, it is possible that the points obtained by MHT$_{\text{combi}}$ contained in the set $Z := \bar{X} \cup Y$ are not all nondominated in this set. Thus, it is reasonable to include a method to extract the nondominated points from Z. This also applies for the set of points obtained as output of the heuristic approach MHT$_{\text{spread}}$ from subsection 3.1. For the implementation this is realized by a pairwise comparison since the obtained sets of points are small. However, more sophisticated methods to find the nondominated points in a finite set are for example the Graef-Younes method, see [13].

4   Results of Numerical Tests

The heuristic approaches from section 3 have been implemented in Matlab R2017a and tested on 20 test problems from [19, Tab. 1], [20]: BK1, CL1, IM1, T1, LE1, Schaffer2, VU1, FF (n=2,3,4,5), Jin1 (n=2,3,5), lovison1 (constrained and unconstrained), lovison4 (constrained and unconstrained) and T4 (n=2,3). For reasons of comparison, the test problems do not contain expensive black box functions, but are all given analytically. For evaluating the results, one of the objectives is defined as expensive function and the number of function evaluations is counted. We summarize the findings and present individual results exemplarily. For more details on numerical tests for the method MHT and the test instances see [19, 20].

Algorithm 6 MHT$_{\text{combi}}$: Heuristic approximation of the set of Pareto critical points
---
    Input: Functions f_1 (expensive), f_2 (cheap), number p of search areas, spreading distance δ_0, values for the parameters $0 < \eta_1 \leq \eta_2 < 1$, $0 < \gamma_1 \leq \gamma_2 < 1$
    Step 1: Image space split with MHT$_{\text{split}}$
        Compute $\bar{X} = \left\{\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^k\right\} = \text{MHT}_{\text{split}}\ (p)$.
    Step 2: Sorting $\bar{X}$
        Sort $\bar{X}$ such that $f_2(\bar{x}^i) \leq f_2(\bar{x}^{i+1})$ holds for all $i = 1, 2, \ldots, k - 1$.
    Step 3: Spreading with MHT$^{\mathbf{b}}_{\text{spread}}$ and MHT$_{\text{spread}}$
        Compute $Y_1 = \text{MHT}_{\text{spread}}\ (\bar{x}^1)$.
        Compute $Y_i = \text{MHT}^{\text{b}}_{\text{spread}}\ (\bar{x}^i, f_2(\bar{x}^{i-1}))$ for $i = 2, 3, \ldots, k$.
        Set $Y = \bigcup_{i=1}^{k} Y_i$.
    Output: $\bar{X}, Y$

4.1 Results of Spreading

The results of MHT$_{\text{spread}}$ depend on the location of the first Pareto critical point $\bar{x}$ computed by an initial run of MHT and on the spreading distance δ_0. The closer $f_2(\bar{x})$ is to the global minimal value of f_2, the less points can be computed. The larger the spreading distance δ_0, the bigger is the distance between the computed points and the less points can be computed. To illustrate this, we consider at first the convex test problem

$$\min_{x \in [-5,10]^2} \left(\begin{array}{c} x_1^2 + x_2^2 \\ (x_1 - 5)^2 + (x_2 - 5)^2 \end{array} \right) \tag{BK1}$$

with f_1 declared as expensive function.

    Figure 10 shows the result of MHT$_{\text{spread}}$ with the spreading distance $\delta = 0.5$. The left figure shows the domain and the right figure the image space. The starting point $x^0 = (2.38, 7.39)^\top$ is marked blue, the point resulting from the initial run of MHT is marked orange and the points resulting from the spreading approach are marked black. All points that are evaluated during the procedure are depicted as unfilled circles. In total, MHT$_{\text{spread}}$ requires 35 function evaluations to compute 6 efficient points.

    The influence of the initial run of MHT is illustrated by comparing Figure 10 to Figure 11. The latter shows the result of MHT$_{\text{spread}}$ with the same spreading distance $\delta = 0.5$, but with a different starting point $x^0 = (7.65, -3.69)^\top$. In this run, 9 efficient points are computed with 56 function evaluations.

    The influence of different values for the spreading distance is illustrated for the non-convex test problem

$$\min_{x \in [-5,10]^2} \left(\begin{array}{c} \sqrt[8]{x_1^2 + x_2^2} \\ \sqrt[4]{(x_1 - 0.5)^2 + (x_2 - 0.5)^2} \end{array} \right) \tag{LE1}$$

with f_1 declared as expensive function. Figure 12 shows four runs of MHT$_{\text{spread}}$ with the same starting point $x^0 = (1.47, 8.66)^\top$ but with different spreading distances δ.

    The top left figure in Figure 12 shows the result for $\delta = 1$. In this case, the spreading distance is too large and besides the result of the initial run of MHT only the global minimal value of f_2 and a point very close to it are computed. They cannot be distinguished in the
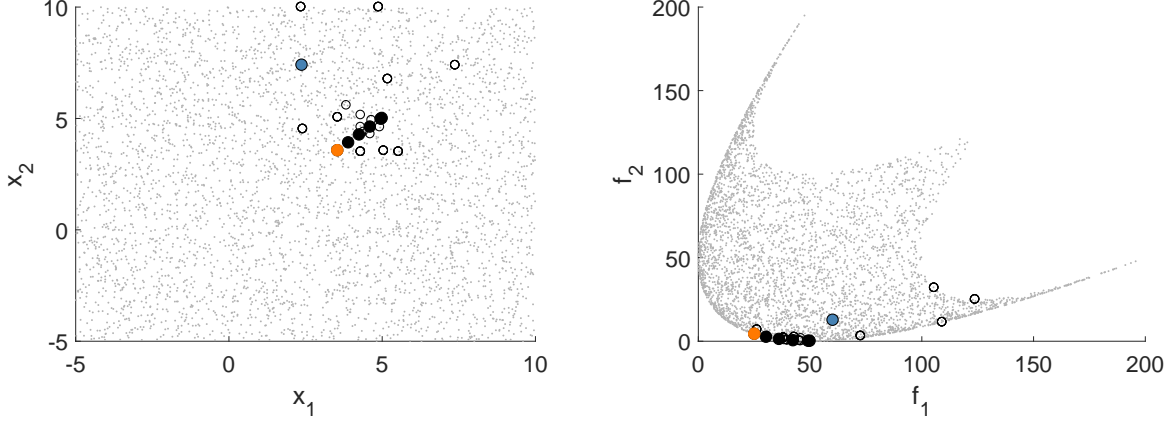
Figure 10: MHT$_{\text{spread}}$ applied to (BK1) (domain left, image space right)

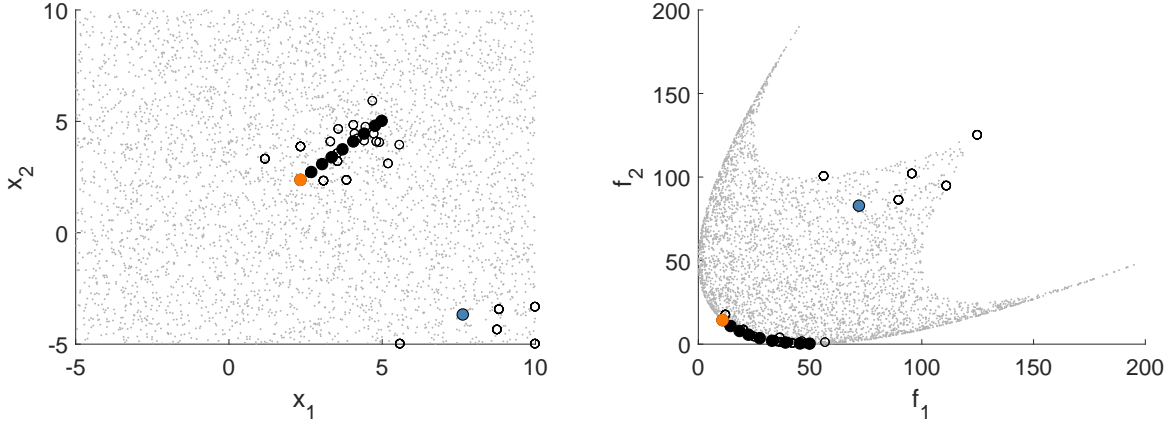Figure 11: MHT$_{\text{spread}}$ applied to (BK1) (domain left, image space right)

figure. Since the global minimal value of f_2 defines the stopping criterion of MHT$_{\text{spread}}$, it is computed in every run of MHT$_{\text{spread}}$. In total, 27 function evaluations of f_1 are required to compute three efficient points. Using the smaller spreading distance $\delta = 0.5$ as depicted in the top right figure, 4 efficient points are computed with a total amount of 34 function evaluations. Again and as for all runs with (LE1), one point is very close to the global minimal value of f_2 such that they cannot be distinguished in the figure. By choosing smaller spreading distances, the number of efficient points increases. This is illustrated in the bottom figures. The left figure shows the result for $\delta = 0.2$ (5 efficient points, 40 function evaluations) and the right figure the result for $\delta = 0.1$ (8 efficient points, 70 function evaluations).

4.2 Results of Image Space Split

The difficulty of applying MHT$_{\text{split}}$ lies in choosing an appropriate number p of split regions. In general, the higher the splitting number p is, the more efficient points are computed and the more function evaluations are required. However, this does not always apply.
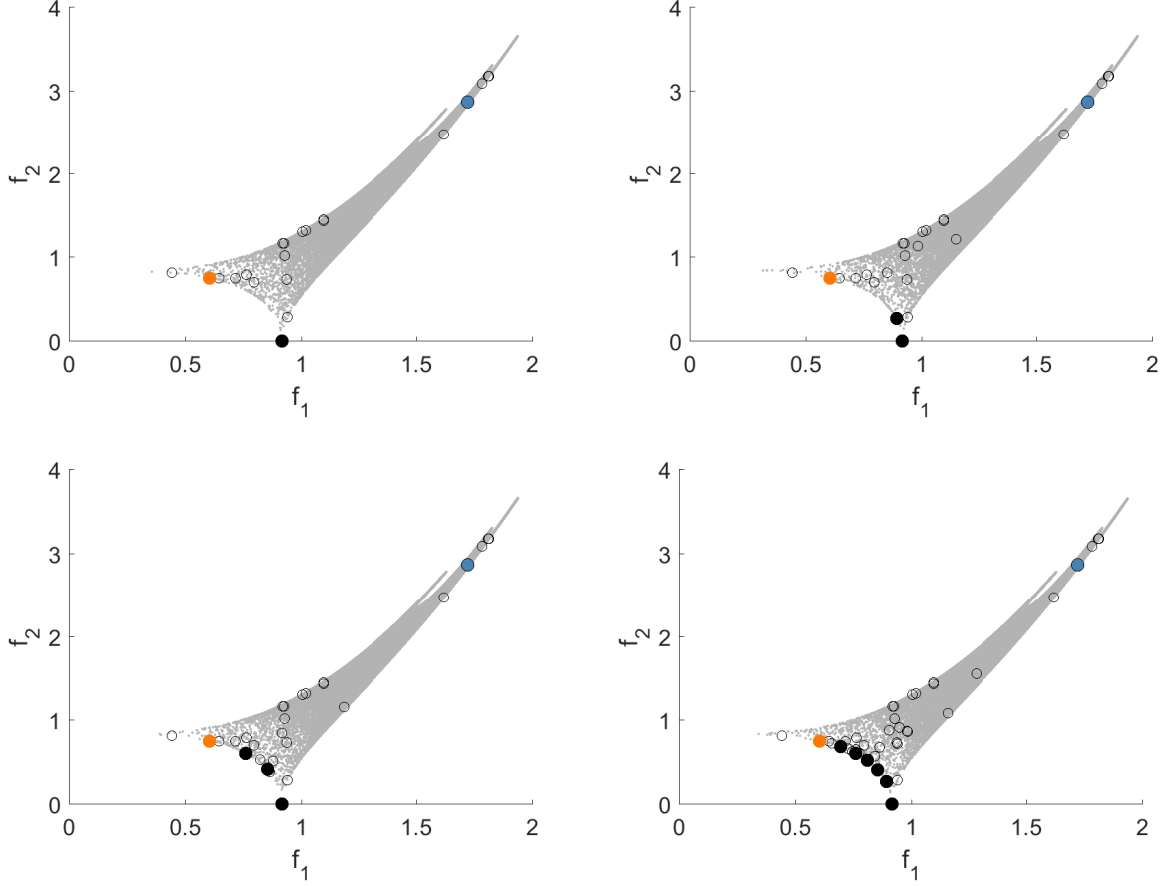
Figure 12: MHT$_{\text{spread}}$ applied to (LE1) with spreading distances $\delta = 1$ (top left), $\delta = 0.5$ (top right), $\delta = 0.2$ (bottom left), $\delta = 0.1$ (bottom right)

To illustrate this, we consider again the test problem (BK1). The result of MHT$_{\text{split}}$ with $p = 8$ is depicted in Figure 13. The starting points are marked blue and the computed solutions orange. The global minimum of the cheap function f_2 is marked black. All points evaluated during the procedure are depicted as unfilled circles. The left figure shows all obtained solutions and the right figure the solutions that remain after the dominated points are deleted. This illustrates that not every run of MHT$_{\text{lb}}$ produces a nondominated point. Moreover, the left figure shows that MHT$_{\text{lb}}$ is not executed in every search region. After the run in the fourth search area, the splitting approach is terminated since the obtained point in this run is dominated by the previously computed solution. By this, function evaluations are saved since no (weakly) efficient points are contained in the remaining search areas. In total, three efficient points are computed after 30 function evaluations.

By choosing larger values for p, more search regions are considered and thus also more efficient points are computed. However, the more search areas are considered, the more the number of function evaluations increases. Figure 14 shows the result of the splitting approach with $p = 22$ search areas. The domain is depicted in the left figure and the image space in the right figure. In this test of MHT$_{\text{split}}$, 7 efficient points are computed with 46 function evaluations.
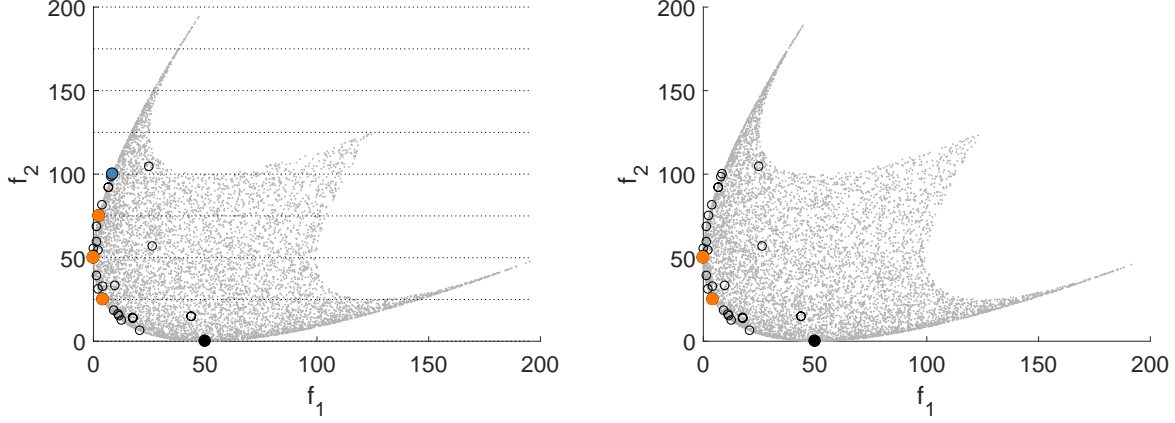
Figure 13: MHT$_\text{split}$ applied to (BK1) with $p = 8$ regions: all computed solutions (left) and nondominated solutions (right)
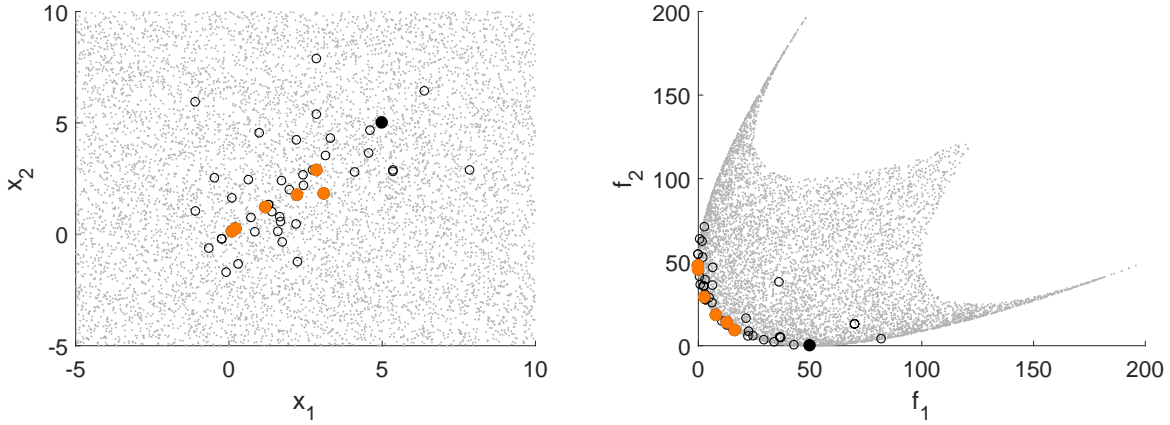

Figure 14: MHT$_\text{split}$ applied to (BK1) with $p = 22$ (domain left, image space right)

Besides, these test runs illustrate that the splitting approach does in general not compute points that are equally distributed on the Pareto front. In contrast, the distance both in the domain and in the image space can vary. By choosing higher numbers of p, the distance can be reduced. However, an equal distribution cannot be guaranteed. Furthermore, a higher number of p also correlates with a higher number of function evaluations. The general tendency of increasing function evaluations is confirmed by Table 1 which gives an overview of the number n_1 of obtained efficient points and the number n_2 of required function evaluations for different values of p.

p	2	4	6	8	10	12	14	16	18	20	22
n_1	2	3	3	4	5	4	4	5	3	4	7
n_2	21	25	31	30	42	34	38	37	26	49	46

Table 1: MHT$_\text{split}$: Numbers n_1 of efficient points and n_2 of function evaluations for (BK1) for different splitting numbers p

For some instances, the number of function evaluations or efficient points also decreases even though p increases. The reason is the location of the search areas and the position of the starting points. Not all search areas contain (weakly) efficient points and some starting points are already close to efficient points.

4.3 Results of the Combination of Splitting and Spreading

The difficulty for the combination of splitting and spreading is to choose appropriate values for the splitting number p and the spreading distance δ. As illustrated in subsection 4.2, the higher the number p is, the more solutions are computed by $\text{MHT}_{\text{split}}$. However, this correlates with the number of function evaluations. In general, the more search regions are considered, the more function evaluations are required. As illustrated in subsection 4.1, the smaller the value of δ is chosen, the more solutions are computed by $\text{MHT}_{\text{spread}}$ and the more function evaluations are required. In the following, we give an overview of the interaction of these two parameters and the quality of the approximation of the Pareto front obtained by $\text{MHT}_{\text{combi}}$.

In $\text{MHT}_{\text{combi}}$ the p runs of MHT_{lb} due to the image space split are executed first. They are executed consecutively and the splitting approach is terminated if one of the computed solutions is dominated by the others. Thus, it is possible that not all search areas are considered. Then, $\text{MHT}_{\text{spread}}$ is applied for all the solutions obtained by $\text{MHT}_{\text{split}}$.

To illustrate the procedure of $\text{MHT}_{\text{combi}}$, we consider again the test problem (LE1). Figure 15 shows the result of $\text{MHT}_{\text{combi}}$ with $p = 10$ and $\delta = 0.1$. The left figure shows the splitting areas and all resulting points from the splitting approach. The starting points are marked blue and the obtained solutions are marked orange. All evaluated points are marked as unfilled circles. The black marked points are the result of the spreading with $\text{MHT}_{\text{spread}}$. The right figure shows the remaining points after the dominated points among the solutions are deleted.
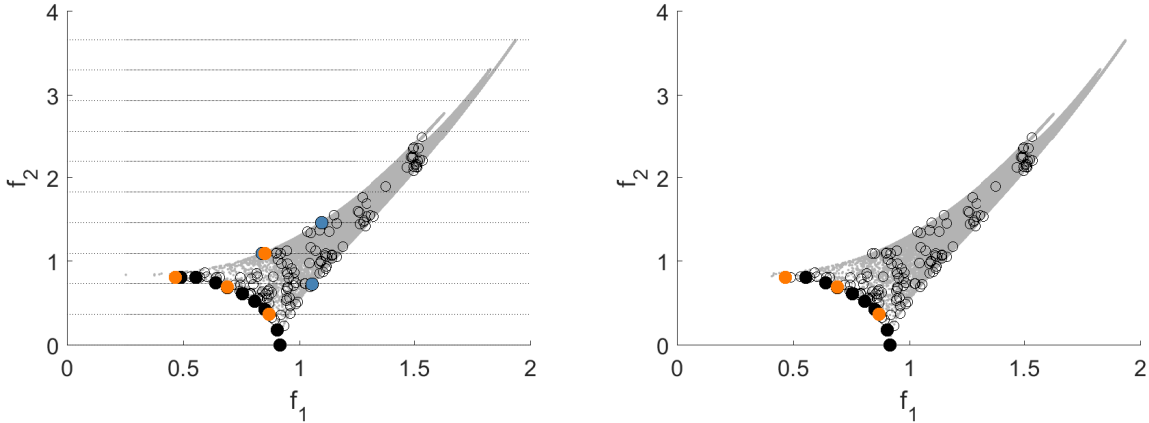

Figure 15: $\text{MHT}_{\text{combi}}$ applied to (LE1) with $p = 10$ and $\delta = 0.1$

With this choice of parameters, a good approximation of the Pareto front is computed. In total, 10 efficient points are computed and 144 function evaluations are required. Table 2 gives an overview of the number of obtained efficient points n_1 and required function

evaluations n_2 for (LE1) with different values for the parameters p and δ. The entries of the cells are given in the order n_1/n_2.

δ \ p	2	4	6	8	10
0.5	3/71	5/177	4/90	6/136	5/137
0.2	6/89	7/184	5/92	7/136	7/145
0.1	8/97	10/202	8/104	9/140	10/144

Table 2: MHT$_{\text{combi}}$: Number of efficient points / number of function evaluations for (LE1) depending on splitting number p and spreading distance δ

These results show that the expected correlation between the values of the parameters p and δ and the number of computed efficient points and required function evaluations does not always apply. A higher number of split regions can also cause a lower number of efficient points or a lower number of function evaluations, as for example for $p = 4$ and $p = 6$ in Table 2. The reasons are the splitting procedure, the location of the search areas and the location of the therein computed solutions. The latter influences, together with the spreading distance δ, the number of efficient points that are computed by MHT$_{\text{combi}}$.

This is exemplarily illustrated for two runs of MHT$_{\text{combi}}$ in Figure 16. The left figure shows the result of splitting and spreading for $p = 4$ and $\delta = 0.5$ and on the right the result for $p = 6$ and $\delta = 0.5$. The bounds of the search areas are depicted as dotted lines, the results of the splitting approach are marked orange and the results of the spreading approach are marked black. For illustration reasons, the dominated points among the solutions of the image space split are also depicted.
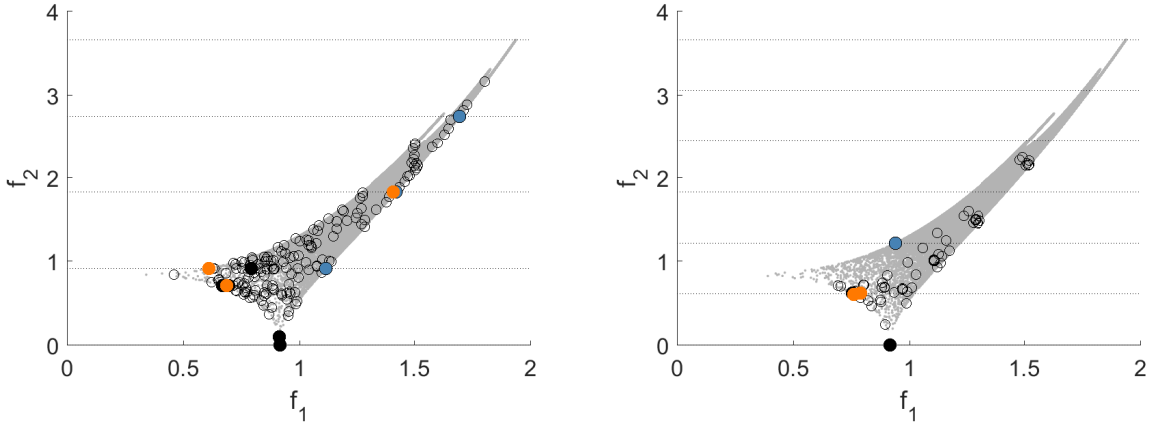

Figure 16: MHT$_{\text{combi}}$ applied to (LE1) with $\delta = 0.5$ and $p = 4$ (left) and $p = 6$ (right)

As the two figures illustrate, for $p = 4$ only one search area contains nondominated points, whereas for $p = 6$ two search areas contain nondominated points. Moreover, for $p = 4$ three of four search areas are considered until the splitting approach is terminated because the latest computed solution is dominated by the other solutions. The splitting approach with $p = 6$ is already terminated after two search areas since the second computed solution is dominated by the first. This also illustrates the local search character of

MHT and the influence of the lower bound in the image space used in $\mathrm{MHT_{lb}}$. Although there exist nondominated points in the second search area for $p = 6$, $\mathrm{MHT_{lb}}$ computes a dominated point which is situated on the boundary of the search area. Thus, it is a weakly efficient point for the auxiliary optimization problem including this lower bound and it is reasonable that $\mathrm{MHT_{lb}}$ terminates.

Additionally, Figure 16 shows that the spreading parameter $\delta = 0.5$ is not a good choice for the test problem (LE1) since only few black marked points – the results of the spreading with $\mathrm{MHT_{spread}}$ – are depicted in the figures.

5 Conclusions

This paper presents three approaches based on the algorithm MHT from [19] that make additional use of the heterogeneity of the objective functions and compute several Pareto critical points. They are based on ideas for bi-objective optimization problems and on the assumption that minimizing or maximizing the analytically given functions on certain areas is realizable with acceptable numerical effort. The heuristic approaches have been implemented and first numerical tests with 20 test problems have been executed. These first tests show the usefulness and applicability of the approaches. However, no general statement can be made which approach gives the best results. The difficulty lies in choosing appropriate values for the parameters p for the splitting approach and δ for the spreading approach. An extensive parameter study is necessary to obtain reliable recommendations for the choice of the parameters for a given specific optimization problem. The presented heuristic approaches are formulated for bi-objective optimization problems, but can be transferred to optimization problems with three or more objective functions. More sophisticated strategies to include all cheap functions for the spreading idea are subject to future research.

A Assumptions

Assumptions A.1 to A.11 as required for Theorem 2.4.

Assumption A.1 *The objective functions $f_i : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable for all $i \in \{1, 2, \ldots, q\}$.*

Assumption A.2 *The function $\phi : \mathbb{R}^n \to \mathbb{R}$ with $\phi(x) = \max_{i=1,2,\ldots,q} f_i(x)$ is bounded from below.*

Assumption A.3 *The model functions $m_i^k : \mathbb{R}^n \to \mathbb{R}$ are quadratic for all $k \in \mathbb{N}$ and $i \in \{1, 2, \ldots, q\}$.*

Assumption A.4 *The model functions m_i^k are exact at the current iteration point, that is it holds*
$$m_i^k(x^k) = f_i(x^k) \text{ for all } i \in \{1, 2, \ldots, q\} \text{ and } k \in \mathbb{N}.$$

31

Assumption A.5 *The gradients of the model functions $m_i^k, i \in \{2, 3, \ldots, q\}$, for the cheap functions f_i coincide with the original gradients in the current iteration point, that is it holds*

$$\nabla m_i^k(x^k) = \nabla f_i(x^k) \text{ for all } i \in \{2, 3, \ldots, q\} \text{ and for all } k \in \mathbb{N}.$$

Assumption A.6 *For every function f_i, $i \in \{1, 2, \ldots, q\}$, the Hessian of f_i is uniformly bounded, that is there exists a constant $\kappa_{\mathrm{uhf}_i} > 1$ fulfilling*

$$\|\nabla_{xx} f_i(x)\|_F \leq \kappa_{\mathrm{uhf}_i} - 1$$

for all $x \in \mathbb{R}^n$. The index 'uhf$_i$' stands for upper bound on the Hessian of f_i.

Assumption A.7 *For every function m_i^k, $i \in \{1, 2, \ldots, q\}$, the Hessian of m_i^k is uniformly bounded for all iterations $k \in \mathbb{N}$, that is there exists a constant $\kappa_{\mathrm{uhm}_i} > 1$ independent of k fulfilling*

$$\left\|\nabla_{xx} m_i^k(x)\right\|_F \leq \kappa_{\mathrm{uhm}_i} - 1$$

for all $x \in B_k$. The index 'uhm$_i$' stands for upper bound on the Hessian of m_i^k.

Assumption A.8 *In every iteration $k \in \mathbb{N}$ the model m_1^k is valid for the function f_1 in B_k, that is there exists a constant $\kappa_{\mathrm{em}_1} > 0$ independent of k such that it holds for all $x \in B_k$*

$$\left| f_1(x) - m_1^k(x) \right| \leq \kappa_{\mathrm{em}_1} \delta_k^2.$$

Assumption A.9 *There exists a constant $\kappa_\omega > 0$ such that it holds for every iteration $k \in \mathbb{N}$*

$$\left| \omega_m(x^k) - \omega(x^k) \right| \leq \kappa_\omega \, \omega_m(x^k).$$

Assumption A.10 *There exists a constant $\kappa_r \in (0, 1]$ such that it holds for every iteration $k \in \mathbb{N}$ with x^k not Pareto critical for $\min_{x \in \mathbb{R}^n} m^k(x)$*

$$\frac{\min\limits_{i=1,\ldots,q} r_i^k}{\max\limits_{j=1,\ldots,q} r_j^k} \geq \kappa_r. \tag{25}$$

Assumption A.11 *There exists a constant $\kappa_\phi \in (0, 1)$ such that it holds for every iteration $k \in \mathbb{N}$*

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq \kappa_\phi \omega(x^k) \min\left\{ \frac{\omega(x^k)}{\beta_\phi^k}, \delta_k \right\}$$

with $\beta_\phi^k = \max_{i=1,\ldots,q} \left\|\nabla_{xx} m_i^k(x^k)\right\|_F + 1$.

References

[1] G. Bouza, E. Quintana, and C. Tammer. A unified characterization of nonlinear scalarizing functionals in optimization. *Vietnam Journal of Mathematics*, 31 pages, 2019.

[2] A. Conn, N. Gould, and P. Toint. *Trust-Region Methods.* MPS-SIAM Series on Optimization, 2000.

[3] A. Conn, K. Scheinberg, and L. Vicente. *Introduction to Derivative-Free Optimization.* MPS-SIAM Series on Optimization, 2009.

[4] L. Drummond and B. F. Svaiter. A steepest descent method for vector optimization. *J. Comput. Appl. Math.*, 175:395–414, 2005.

[5] G. Eichfelder. Scalarizations for adaptively solving multi-objective optimization problems. *Computational Optimization and Applications*, 44(2):249–273, 2009.

[6] G. Eichfelder. *Adaptive Scalarization Methods in Multiobjective Optimization.* Springer, Heidelberg 2008.

[7] G. Eichfelder and J. Jahn. *Vector optimization problems and their solution concepts* Recent Developments in Vector Optimization, Ansari, Q.H., Yao, J.-C. (Eds.), pages 1–27. Springer, Heidelberg 2012.

[8] G. Eichfelder and M. Pilecka. *Ordering Structures and Their Applications*, In: Applications of Nonlinear Analysis, Rassias, Themistocles M. (ed.), pages 256–304. Springer, Heidelberg 2018.

[9] J. Fliege, L. G. Drummond, and B. Svaiter. Newton's method for multiobjective optimization. *SIAM J. Optim.*, 20:602–626, 2009.

[10] J. Fliege and B. F. Svaiter. Steepest descent methods for multicriteria optimization. *Math. Methods of Oper. Res.*, 51:479–494, 2000.

[11] C. Gerth (Tammer) and P. Weidner. Nonconvex separation theorems and some applications in vector optimization. *J. Optim. Theory Appl.*, 67(2):297–320, 1990.

[12] J. Jahn. *Vector Optimization.* Springer, Heidelberg 2011.

[13] J. Jahn and U. Rathje. Graef-younes method with backward iteration. In K. Küfer, H. Rommelfanger, C. Tammer, and K. Winkler, editors, *Multicriteria Decision Making and Fuzzy Systems - Theory, Methods and Applications*, pages 75–81. Shaker Verlag, 2006.

[14] A.A. Khan, C. Tammer, and C. Zălinescu. *Nonconvex Separation Theorems*, Chapter in: Set-valued Optimization, pages 213–248. Springer Heidelberg 2015

[15] S. Prinz, J. Thomann, G. Eichfelder, T. Boeck, and J. Schumacher. Multiobjective optimization of the mixing of particles in a liquid metal layer by electromagnetic stirring. Technical report, arxiv, 2019.

[16] J.-H. Ryu and S. Kim. A derivative-free trust-region method for biobjective optimization. *SIAM J. Optim.*, 24:334–362, 2014.

[17] C. Tammer and C. Zălinescu. Lipschitz properties of the scalarization function and applications. *Optimization*, 59:305–319, 2010.

[18] J. Thomann. *A Trust Region Approach for Multi-Objective Heterogeneous Optimization*. Dissertation, Technische Universität Ilmenau, Germany, 2019.

[19] J. Thomann and G. Eichfelder. A trust region algorithm for heterogeneous multi-objective optimization. *SIAM J. Optim.*, 29(2):1017–1047, 2019.

[20] J. Thomann and G. Eichfelder. Numerical results for the multi-objective trust region algorithm MHT. *Data in Brief*, 25, Doi 10.1016/j.dib.2019.104103, 2019.