TECHNISCHE UNIVERSITÄT
ILMENAU
Faculty of Mathematics and Natural Sciences
Mathematical Methods of Operations Research

A Trust Region Approach for Multi-Objective Heterogeneous Optimization

Doctoral Thesis in partial fulfillment of the requirements
for the degree of Dr. rer. nat.

by Jana Thomann

1. Reviewer: Prof. Dr.rer.nat. habil. G. Eichfelder
2. Reviewer: Prof. Dr.rer.nat. habil. J. Fliege

Day of submission: November 27, 2018
Day of defense: March 04, 2019

urn:nbn:de:gbv:ilm1-2019000059

Zusammenfassung

In dieser Arbeit wird ein „Trust-Region“ Algorithmus für multikriterielle Optimierungsprobleme mit heterogenen Zielfunktionen vorgestellt. Eine der Zielfunktionen ist eine teure Black-Box-Funktion. Sie ist nicht analytisch gegeben, sondern beispielsweise durch eine Simulation. Für diese Funktion wird angenommen, dass die Berechnung von Funktionswerten zeitaufwändig ist und die Ableitungen nicht mit vertretbarem numerischen Aufwand berechnet werden können. Des Weiteren wird vorausgesetzt, dass die anderen Zielfunktionen analytisch gegeben sind und die Berechnung von Funktionswerten und Ableitungen mit geringem numerischen Aufwand verbunden ist.

Es wird ein grundlegender Algorithmus für derartige Optimierungsprobleme vorgestellt. Der Ansatz ist iterativ und nutzt lokale Modellfunktionen und eine im Bildraum definierte Suchrichtung. Der Algorithmus erzeugt eine Folge von Iterationspunkten. Es wird bewiesen, dass der Häufungspunkt dieser Folge ein notwendiges lokales Optimalitätskriterium erfüllt. Darüber hinaus werden verschiedene Modifikationen dieses Algorithmus vorgestellt, welche die Heterogenität der Zielfunktionen weiter nutzen und teilweise mehr als einen Punkt als Ausgabe erzeugen.

Des Weiteren werden Ergebnisse von numerischen Tests mit der Grundversion und einigen Modifikationen des Algorithmus präsentiert und diskutiert. Sie bestätigen die theoretischen Resultate und zeigen die Nützlichkeit der Verfahren. Der grundlegende Algorithmus wurde außerdem auf ein Anwendungsproblem der Fluiddynamik angewandt. Die zugehörigen Ergebnisse werden präsentiert und im Rahmen des Anwendungsproblems interpretiert.

Abstract

This thesis presents a trust region approach for multi-objective optimization problems with heterogeneous objective functions. One of the objective functions is an expensive black-box function, not given analytically, but for example by a simulation. Computing function values is assumed to be time-consuming and derivative information is not available with reasonable effort. The other objective functions are assumed to be given analytically and function evaluations and derivatives are easily available with low numerical effort.

A basic algorithm for such optimization problems is presented. It is an iterative approach using local model functions and a search direction which is defined in the image space. The algorithm generates a sequence of iteration points. It is proved that the accumulation point of this sequence fulfills a necessary condition for local optimality. Moreover, several modifications of the basic algorithm are presented that make more use of the heterogeneity of the objective functions and partly produce several points as output.

Numerical results for the basic algorithm and several modifications are presented and discussed. They confirm the theoretical findings and show the usefulness of the approaches. Moreover, an application-motivated optimization problem from fluid dynamics is considered and the results are presented and interpreted according to the application.

Table of Contents

1 Introduction

Multi-objective optimization problems can be found in various fields, such as engineering, medicine, economics or finance [Ste+08; Eic08; TC07; AM11] where several conflicting objectives are optimized. If some of the objectives are not given analytically, but are black-box functions, an additional difficulty arises. This is for example the case if one objective function is the result of an experiment or a simulation. Black-box functions can be smooth functions and derivatives can exist, but they are not available with reasonable computational effort.

In this thesis we focus on smooth multi-objective optimization problems with heterogeneous functions, i.e. the objective functions differ in certain aspects affecting the optimization process. Different types of heterogeneity can occur, for example in scaling, in basic properties such as smoothness or the presence of plateaus, in the evaluation time or in the theoretical and practical difficulty for optimizing the individual functions. An overview of different aspects of heterogeneity and optimization problems in which they occur is, for instance, given in [Gre+15, p.125f].

The heterogeneity considered in this thesis is the different amount of information available for the functions and the computation time. One of the objective functions is not given analytically, the function values are only obtained with high computational effort and derivatives are not available with reasonable numerical effort. Such a function can be, for instance, a black-box function given by a time-consuming simulation. We will refer to this function as expensive function. The other objective functions are given analytically and derivatives are easily available. These functions will be called cheap in contrast to the expensive function.

Such multi-objective optimization problems with heterogeneous and expensive black-box functions arise for example in engineering or medicine [TPE14; Fla14; Str+08]. In Lorentz force velocimetry [TPE14], a contactless flow measurement technique, the aim is to find an optimal design of a magnet which minimizes the weight of the magnet

and maximizes the induced Lorentz force of the magnet. The first objective is an analytically given function, but the second objective can in general only be determined by a time-consuming simulation.
Furthermore and according to [Gre+15, p.124], heterogeneous problems with expensive functions also occur in imaging techniques in interventional radiology [Fla14]. One objective function is given analytically as the sum of squared differences. The second objective function is described by physical models for fluids and diffusion processes and is given by an implicit differential equation.

1.1 Literature Overview

In the literature there are many solution methods for multi-objective optimization problems and also for problems with expensive objective functions. An overview of approaches for expensive single-objective optimization problems, i.e. optimization problems with one objective function, is for example given in [CSV09b; AH17].
The overall aim when considering expensive functions is saving function evaluations and one possible approach is parallelizing the function evaluations. However, this is not possible for all optimization problems, for example if the expensive function is a simulation and there are temporal dependencies. Moreover, if given by a simulation, the time required to compute one function value can vary and does not need to be constant. In general, parallelization can only be realized to a certain extent. Computational resources and of course the computation time itself of one function evaluation lead to an overall time limit for optimization procedures with expensive functions. Therefore, it is important to use a suitable solution method.
A commonly used approach to solve multi-objective optimization problems is scalarization [DD98; LPV05; Eic09], that is reformulating the multi-objective optimization problem to a single-objective optimization problem mostly by combining the objective functions into a scalar-valued function as in the weighted sum approach. The single-objective problem is then considered and solved as a surrogate to the multi-objective problem. The obtained solutions are connected to the solutions of the original problem. However, a problem for most scalarization techniques is that whenever one of the objective functions is an expensive function, the scalarized function will also be an ex-

pensive function and the high computational effort possibly linked to only one of the objectives affects the whole method. If there is an analytically given function which is easy and quick to compute this has no impact. Hence, such scalarization methods are not well suited for heterogeneous optimization problems.
Other methods for multi-objective optimization problems, like the generalized steepest descent method [FS00; DS05] or the generalized Newton method [FDS09] require derivative information and are therefore not applicable to heterogeneous problems where the derivatives are not available with reasonable numerical effort for all objective functions. Approximating the derivatives is not a reasonable option for expensive black-box functions. Either the obtained approximation would not be reliable or too many function evaluations would be required.
However, there are also derivative-free methods in multi-objective optimization and a very common approach, both in single- and multi-objective optimization, is direct search [AD06; CSV09b; Cus+11]. This approach is based only on function values.
Several versions and realizations exist, such as the basic DMS (Direct MultiSearch) [Cus+11] for multi-objective optimization problems in general or BIMADS [ASZ08]. The latter is designed for bi-objective box constrained problems where the structure of the objective functions is absent or unreliable. This includes simulation given black-box functions as they are considered in this thesis. The algorithm is based on the direct search method MADS (mesh adaptive direct search) [AD06]. BIMADS solves a series of scalarizations of the given bi-objective optimization problem. The series of auxiliary optimization problems is constructed such that the entire set of optimal points, referred to as Pareto front, is attempted to be approximated, even if it is nonconvex or disconnected.
A disadvantage of direct search methods is the fact that the performance deteriorates if the number of variables increases [KLT03]. However, the main drawback when applying such methods to heterogeneous problems is again that the expensive function would 'dominate' the procedure. The heterogeneity is not considered and not all information given is used during the optimization process, namely the derivative information of the cheap functions.
This also applies to evolutionary or genetic algorithms [Deb01; CVL02]. These are heuristic methods with the search strategy of sampling by changing or recombining previously used points. In general, the sampling strategy can cause a high number of function evaluations and thus a high computational effort if expensive functions are

considered. Furthermore, evolutionary algorithms do not include any intrinsic measures of distance to convergence such as a step-length and therefore there is no clear stopping criterion.

In connection with evolutionary algorithms and expensive optimization problems also the approach of surrogate models and meta-modeling is often used [JSW98; KN08; Roy+18]. The model for the objective function is built during the optimization process based on the results in the iterations and the model is used in the subsequent search. This approach is well suited for expensive functions, but not required for analytically given functions. An overview of model-based evolutionary algorithms is for example given in [Chu+17] and an overview of model-based non-evolutionary algorithms in [Tab+15].

Another approach on which derivative-free methods are based on is the trust region method [Mor83; CGT88; CGT89; CGT00; CSV09a; CSV09b]. There are also multi-objective realizations of this approach [RK14; VOS14; CLM16]. Trust region methods are not initially designed for expensive functions but can easily be adapted to them. It is an efficient and flexible approach for which many theoretical results are documented in the literature.

A basic generalization of the trust region approach to multi-objective problems is given in [VOS14]. It is based on derivative information. The authors prove that the sequence of iterates converges to a point fulfilling a necessary optimality criterion referred to as Pareto criticality. The proof is based on a characterization of this necessary optimality criterion from multi-objective descent theory [FS00; DS05]. The required assumptions are derived from the single-objective version of trust region approaches and the convergence analysis follows the strategy and structure of the proof from the basic approach [CGT00] closely. However, this method needs derivative information and in the nonsmooth case the Clarke subdifferential is used. Hence, this approach is not suitable for the heterogeneous problems considered in this thesis where using derivative information of the expensive function shall be avoided.

The same applies to the trust region approach presented in [CLM16]. It follows the Newton method [FDS09] to compute a multi-objective descent direction and requires derivative information of first and second order. It is proved that the algorithm converges to a Pareto critical point.

A trust region method for bi-objective expensive problems where derivative information is absent for both objectives is presented in [RK14]. The algorithm uses a scalar-

ization technique and approximates the Pareto front. The authors prove convergence to a Pareto critical point. This algorithm is applicable to heterogeneous problems but would again neglect some information given for the cheap functions.

1.2 Contribution of this Thesis

So far, there are no solution methods for heterogeneous multi-objective problems that explicitly consider the differences of the objective functions. This thesis presents a trust region method that can regard heterogeneity and gives a theoretical foundation for further modifications using this heterogeneity. Several modifications that exploit the heterogeneity further are presented.
Like [VOS14; RK14; CLM16] the main algorithm uses the idea of generalizing the trust region approach to a multi-objective problem. It is an iterative approach which considers surrogate model functions for the objective functions in local areas in every iteration. However, it differs in various aspects from the approaches from the literature. Besides the aspect of heterogeneity, a new aspect is that the search direction is computed in the image space by using local ideal points, i.e. individual minima of the model functions in local areas.
The differences in the determination of the search direction affect the convergence analysis such that it is not transferable from other trust region approaches without significant modifications. Still, it can be proved that the sequence of iterates converges to a point fulfilling the necessary optimality criterion of Pareto criticality. Like in [VOS14], a concept from multi-objective descent algorithms [FS00; DS05] is used for the theoretical considerations which are based on the theoretical considerations from the single-objective trust region methods [CGT00; CSV09b].
The idea of this trust region approach including the convergence results is firstly presented in [TE18a]. At the time of the submission of this thesis, it is under second review in a peer-reviewed journal and only available as preprint. Moreover, a summary of the numerical results is available in [TE18b].

1.3 Structure

This thesis is structured as follows. In Chapter 2 the required basic concepts from multi-objective optimization are given. In Chapter 3 the general idea of trust region methods for single-objective optimization problems is explained and an overview of model functions in this context is given.
A multi-objective trust region method based on the single-objective trust region approach and referred to as MHT is presented in Chapter 4. Besides a description of this method, the chapter includes theoretical results leading to the main statement that the sequence of iterates generated by MHT converges to a point fulfilling a necessary optimality criterion.
Furthermore, several minor modifications of MHT are presented in this chapter. In Chapter 5 three heuristic approaches are presented to extend MHT and to make more use of the heterogeneity of the objective functions.
The basic version of the algorithm from Chapter 4 and the modifications from Chapter 5 have been implemented and tested with test problems. The results are presented in Chapter 6. In Chapter 7 the results of applying MHT to an application-motivated optimization problem from fluid dynamics are presented. In Chapter 8 the findings are summarized and an outlook is given.

2 Mathematical Background

In this thesis unconstrained multi-objective optimization problems given by

$$\min_{x \in \mathbb{R}^n} f(x) = \min_{x \in \mathbb{R}^n} \left(f_1(x), f_2(x), \ldots, f_q(x)\right)^\top \tag{MOP}$$

with $f : \mathbb{R}^n \to \mathbb{R}^q$, $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and twice continuously differentiable functions $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, are considered.
The objective function f_1 is assumed to be an expensive black-box function which is not given analytically. Computing function values is assumed to be time-consuming and derivative information is not available with reasonable numerical effort. The objective functions $f_2, f_3, \ldots, f_q$ are assumed to be analytically given functions for which function values and derivative information is easily available and quick to compute. We will refer to f_1 as expensive function and, in contrast, to the other objective functions as cheap functions.

2.1 Efficiency

For the theoretical considerations also constrained optimization problems are considered. Hence, we present the basic definitions for the general optimization problem

$$\min_{x \in \Omega} \left(f_1(x), f_2(x), \ldots, f_q(x)\right)^\top \tag{MOP_c}$$

with $f : \mathbb{R}^n \to \mathbb{R}^q$, $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$, and the constraint set $\Omega \subseteq \mathbb{R}^n$.
An introduction to multi-objective optimization can for example be found in [Mie98; Ehr05; Jah11].

Definition 2.1 A point $\bar{x} \in \Omega$ is called *efficient (solution)* for (MOP_c) (or *Pareto optimal*), if there exists no point $x \in \Omega$ fulfilling

$$f_i(x) \leq f_i(\bar{x}) \text{ for all } i \in \{1, 2, \ldots, q\} \text{ and } f(x) \neq f(\bar{x}).$$

The image of an efficient point $f(\bar{x})$ is called *nondominated*. The set of all nondominated points is called *Pareto front*. The images of all points that are not efficient are called *dominated*. A point $\bar{x} \in \Omega$ is called *weakly efficient (solution)* for (MOP_c) (or *weakly Pareto optimal*), if there exists no point $x \in \Omega$ fulfilling

$$f_i(x) < f_i(\bar{x}) \text{ for all } i \in \{1, 2, \ldots, q\}.$$

The image of a weakly efficient point $f(\bar{x})$ is called *weakly nondominated*.

Apparently, every efficient point is also weakly efficient. The two concepts are illustrated in the image space for a bi-objective optimization problem (MOP_c) $(q = 2)$ in Figure 2.1. It shows an efficient point $\bar{x}^1$ and a weakly efficient point $\bar{x}^2$.

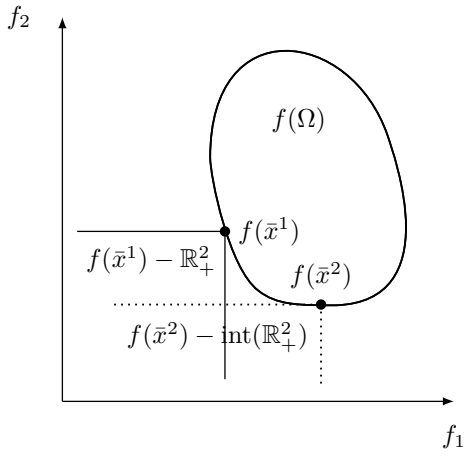

Figure 2.1 – Nondominated and weakly nondominated points

Both concepts of efficiency can be restricted to local areas.

Definition 2.2 A point $\bar{x} \in \Omega$ is called *locally (weakly) efficient* for (MOP_c) if there exists a neighborhood $U(\bar{x})$ of $\bar{x}$ such that $\bar{x}$ is (weakly) efficient for the optimization problem $\min \{f(x) \mid x \in \Omega \cap U(\bar{x})\}$.

Similarly to nondominated and dominated points of the optimization problem (MOP_c), we define these terms regarding a finite set of points. For that purpose and throughout this thesis, we will use the inequality relations $\leq$ and $<$ for vectors in a componentwise manner, i.e. we write $a \leq b$ ($a < b$) for $a, b \in \mathbb{R}^q$ if it holds $a_i \leq b_i$ ($a_i < b_i$) for all $i \in \{1, 2, \ldots, q\}$.

Definition 2.3 Let $Y = \{y^1, y^2, \ldots, y^k\} \subset \mathbb{R}^q$ be a finite set of points. A point $y^i \in Y$, $i \in \{1, 2, \ldots, k\}$, is called *dominated* in Y, if there exists a point $y^j \in Y$, $j \in \{1, 2, \ldots, k\} \setminus \{i\}$, such that it holds $y^j \leq y^i$. A point $y^i \in Y$, $i \in \{1, 2, \ldots, k\}$, is called *nondominated* in Y, if there exists no point in $Y \setminus \{y^i\}$ dominating it.

Furthermore, we define the sets $\mathbb{R}^q_+ := \{x \in \mathbb{R}^q \mid x_i \geq 0 \text{ for all } i = 1, 2, \ldots, q\}$ and $\mathbb{R}^q_{++} := \{x \in \mathbb{R}^q \mid x_i > 0 \text{ for all } i = 1, 2, \ldots, q\}$.

2.2 Multi-Objective Descent and Pareto Criticality

The concept of descent directions from single-objective optimization can be transferred to multi-objective optimization problems, see for example [FS00; DS05].

Definition 2.4 Let $f : \mathbb{R}^n \to \mathbb{R}^q$ be with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$. A vector $d \in \mathbb{R}^n$ is called a *descent direction* for the function f at $x \in \mathbb{R}^n$ if there exists a scalar $t_0 > 0$ with $f_i(x + td) < f_i(x)$ for all $t \in (0, t_0]$ and for all $i \in \{1, 2, \ldots, q\}$.

The following lemmas give a sufficient and a necessary condition for descent directions. For a better understanding of the context we give short proofs, see also [Tho15].

Lemma 2.5 *Let $f : \mathbb{R}^n \to \mathbb{R}^q$ be with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, continuously differentiable functions. Furthermore, let $x, d \in \mathbb{R}^n$ be given with $\nabla f_i(x)^\top d < 0$ for all $i \in \{1, 2, \ldots, q\}$. Then d is a descent direction for f at x.*

Proof. Let $x, d \in \mathbb{R}^n$ be given with $\nabla f_i(x)^\top d < 0$ for all $i \in \{1, 2, \ldots, q\}$. Since f_i is continuously differentiable for all $i \in \{1, 2, \ldots, q\}$, it holds

$$0 > \nabla f_i(x)^\top d = \lim_{t \to 0+} \frac{f_i(x + td) - f_i(x)}{t} \text{ for all } i \in \{1, 2, \ldots, q\}.$$

Thus, there exists a scalar $t_0 > 0$ with $f_i(x + td) - f_i(x) < 0$ for all $t \in (0, t_0]$ and for all $i \in \{1, 2, \ldots, q\}$. This implies $f_i(x + td) < f_i(x)$ for all $t \in (0, t_0]$ and for all $i \in \{1, 2, \ldots, q\}$. Therefore, d is a descent direction for f at x. □

Lemma 2.6 *Let $f : \mathbb{R}^n \to \mathbb{R}^q$ be with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, continuously differentiable functions. Furthermore, let $d \in \mathbb{R}^n$ be a descent direction for f at $x \in \mathbb{R}^n$. Then it holds $\nabla f_i(x)^\top d \leq 0$ for all $i \in \{1, 2, \ldots, q\}$.*

Proof. Let $d \in \mathbb{R}^n$ be a descent direction for f at $x \in \mathbb{R}^n$. Then there exists a scalar $t_0 > 0$ with $f_i(x + td) < f_i(x)$ for all $t \in (0, t_0]$ and for all $i \in \{1, 2, \ldots, q\}$. Thus, it holds $f_i(x+td) - f_i(x) < 0$ for all $t \in (0, t_0]$ and for all $i \in \{1, 2, \ldots, q\}$ and therefore

$$\nabla f_i(x)^\top d = \lim_{t \to 0+} \frac{f_i(x + td) - f_i(x)}{t} \leq 0.$$

□

The concept of Pareto criticality is connected to descent directions and gives a necessary condition for local weak efficiency. Although derivative information is not available for all objective functions of the optimization problems considered in this thesis, the concept of Pareto criticality is helpful for the theoretical considerations.

Definition 2.7 Let $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, 2, \ldots, q$, be continuously differentiable functions. A point $\bar{x} \in \mathbb{R}^n$ 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

$$\nabla f_j(\bar{x})^\top d \geq 0.$$

This concept is a generalization of the stationarity notion for single-objective optimization problems. Setting $q = 1$ in (MOP) gives a single-objective optimization problem. Let $\bar{x} \in \mathbb{R}^n$ be a Pareto critical point according to the definition above. Then

it holds $\nabla f(\bar{x})^\top d \geq 0$ for all $d \in \mathbb{R}^n$. This implies $\nabla f(\bar{x}) = 0_n$ and the standard stationarity notion for the single-objective case is obtained.
Minimality of individual functions, efficiency, criticality and descent directions are connected. These relations are summarized in the following theorem. A proof and further discussion can for example be found in [Tho15, Thm. 2.2.17].

Theorem 2.8 *Let $f : \mathbb{R}^n \to \mathbb{R}^q$ be with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, continuously differentiable functions. Let $\bar{x} \in \mathbb{R}^n$ be given. The following statements hold.*

(i) If $\bar{x}$ is not Pareto critical for (MOP), there exists a descent direction for f at $\bar{x}$.

(ii) If there exists a descent direction for f at $\bar{x}$, then $\bar{x}$ is not weakly efficient for (MOP).

(iii) If there exists an index $i \in \{1, 2, \ldots, q\}$, such that $\bar{x}$ is minimal for f_i, then $\bar{x}$ is Pareto critical for (MOP).

(iv) If there exists no vector $d \in \mathbb{R}^n$ satisfying the necessary condition of Lemma 2.6, then $\bar{x}$ is Pareto critical for (MOP).

(v) If there exists no vector $d \in \mathbb{R}^n$ satisfying the sufficient condition by Lemma 2.5, then $\bar{x}$ is Pareto critical for (MOP).

The following lemma shows that Pareto criticality is a necessary condition for local weak efficiency, see for example [FDS09; Jah11]. For a better understanding of the context we give a short proof.

Lemma 2.9 *Let $f : \mathbb{R}^n \to \mathbb{R}^q$ be with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, continuously differentiable functions. If $\bar{x} \in \mathbb{R}^n$ is locally weakly efficient for (MOP), it is Pareto critical for (MOP).*

Proof. Suppose the vector $\bar{x} \in \mathbb{R}^n$ is locally weakly efficient for (MOP), but not Pareto critical. Then there exists a neighborhood $U(\bar{x})$ of $\bar{x}$ such that $\bar{x}$ is weakly efficient for $\min_{x \in U(\bar{x})} f(x)$. According to Theorem 2.8 (i), there exists a descent direction d for f at $\bar{x}$, i.e. there exists a scalar $t_0 > 0$ such that it holds

$$f_i(\bar{x} + td) - f_i(\bar{x}) < 0$$

for all $t \in (0, t_0]$ and for all $i \in \{1, 2, \ldots, q\}$. Then there exists a scalar $t_1 \in (0, t_0]$ such that it holds $y := \bar{x} + t_1 d \in U(\bar{x})$. Furthermore, it holds $f_i(y) < f_i(\bar{x})$ for all $i \in \{1, 2, \ldots, q\}$ which contradicts $\bar{x}$ being weakly efficient for $\min_{x \in U(\bar{x})} f(x)$. □

The following auxiliary function characterizes Pareto criticality. It is defined by an auxiliary optimization problem used in the literature about multi-objective descent directions [FS00]. Similar approaches that can also characterize Pareto criticality can be found for example in [DS05; FDS09]. Throughout this thesis, we consider the euclidean norm $\|\cdot\| := \|\cdot\|_2$ as vector norm and as compatible matrix norm the Frobenius norm $\|\cdot\|_F$.

Definition 2.10 Let $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, be continuously differentiable functions. We define the function $\omega : \mathbb{R}^n \to \mathbb{R}$ by

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

The characterization of Pareto critical points by means of the function ω is given in Lemma 2.11. It is formulated according to [FS00, Lem. 3], the proof can be found there.

Lemma 2.11 *Let $f : \mathbb{R}^n \to \mathbb{R}^q$ be with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, continuously differentiable functions. For the function ω from Definition 2.10 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 solutions of the optimization problem in Definition 2.10 have some useful properties.

Lemma 2.12 *Let $f : \mathbb{R}^n \to \mathbb{R}^q$ be with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, continuously differentiable functions. Furthermore, let $\bar{x} \in \mathbb{R}^n$ be not Pareto critical for (MOP) and let d_ω denote a minimal solution of the optimization problem*

$$-\min_{\|d\| \leq 1} \max_{i=1,\ldots,q} \nabla f_i(\bar{x})^\top d \tag{2.1}$$

given by ω from Definition 2.10. The following statements hold.

(i) The vector d_ω is a descent direction for f at $\bar{x}$.

(ii) There exist scalars $\alpha_i \in [0,1]$, $i = 1, 2, \ldots, q$, with $\sum_{i=1}^{q} \alpha_i = 1$ and $\mu \geq 0$ such that it holds

$$d_\omega = -\mu \sum_{i=1}^{q} \alpha_i \nabla f_i(\bar{x}) \text{ and } \|d_\omega\| = 1.$$

Furthermore, it holds

$$\omega(\bar{x}) \leq \left\| \sum_{i=1}^{q} \alpha_i \nabla f_i(\bar{x}) \right\|.$$

Proof. Let $\bar{x}$ be not Pareto critical for (MOP). To prove statement (i), let d_ω denote a minimal solution of (2.1). According to Lemma 2.11, it holds $\omega(\bar{x}) > 0$. This implies $\max_{i=1,\ldots,q} \nabla f_i(\bar{x})^\top d_\omega < 0$. Thus, the preconditions of Lemma 2.5 are fulfilled and it follows that d_ω is a descent direction for f at $\bar{x}$.
To prove statement (ii), consider an equivalent reformulation of (2.1) given by

$$-\min \left\{ t \in \mathbb{R} \mid \nabla f_i(\bar{x})^\top d \leq t \text{ for all } i = 1, 2, \ldots, q \text{ and } \|d\| \leq 1 \right\}. \tag{2.2}$$

Let (t_ω, d_ω) denote a minimal solution of (2.2). According to Lemma 2.11, it holds $\omega(\bar{x}) > 0$ and therefore $d_\omega \neq 0_n$. In the following, we consider the KKT conditions for the optimization problem (2.2). In this context, the Lagrange function is defined by

$$L(t, d, \alpha, \lambda) := t + \sum_{i=1}^{q} \alpha_i (\nabla f_i(\bar{x})^\top d - t) + \lambda (d^\top d - 1)$$

with $\alpha \in \mathbb{R}^q$ and $\lambda \in \mathbb{R}$. Note that the MFCQ is fulfilled in (t_ω, d_ω). Then according to the KKT conditions, there exists a vector $(\bar{\alpha}, \bar{\lambda}) \in \mathbb{R}_+^{q+1}$ with

$$\begin{aligned} &\nabla_{(t,d)} L(t_\omega, d_\omega, \bar{\alpha}, \bar{\lambda}) = 0_{1+n}, \quad \bar{\lambda}(d_\omega^\top d_\omega - 1) = 0 \quad \text{and} \\ &\bar{\alpha}_i (\nabla f_i(\bar{x})^\top d_\omega - t_\omega) = 0 \text{ for all } i \in \{1, 2, \ldots, q\}. \end{aligned}$$

This implies

$$1 - \sum_{i=1}^{q} \bar{\alpha}_i = 0 \quad \text{and} \quad \sum_{i=1}^{q} \bar{\alpha}_i \nabla f_i(\bar{x}) + 2\bar{\lambda} d_\omega = 0_n. \tag{2.3}$$

Consequently, it holds $\bar{\alpha}_i \in [0,1]$ and $\sum_{i=1}^{q} \bar{\alpha}_i = 1$. Suppose it holds $\bar{\lambda} = 0$, then it follows

$$\sum_{i=1}^{q} \bar{\alpha}_i \nabla f_i(\bar{x}) = 0_n.$$

Due to the complementary conditions it holds $\bar{\alpha}_i \nabla f_i(\bar{x})^\top d_\omega = \bar{\alpha}_i t_\omega$ for all indices $i \in \{1, 2, \ldots, q\}$ which implies

$$0 = \sum_{i=1}^{q} \bar{\alpha}_i \nabla f_i(\bar{x})^\top d_\omega = \sum_{i=1}^{q} \bar{\alpha}_i t_\omega = t_\omega.$$

Due to the definition of ω it holds $\omega(\bar{x}) = -t_\omega = 0$. This is a contradiction to $\omega(\bar{x}) > 0$ and $\bar{x}$ being Pareto critical for (MOP). Thus, it holds $\bar{\lambda} \neq 0$ and it follows from the complementary condition $\|d_\omega\| = 1$. This implies due to (2.3)

$$d_\omega = -\mu \sum_{i=1}^{q} \bar{\alpha}_i \nabla f_i(\bar{x}) \text{ with } \mu := \frac{1}{2\bar{\lambda}} = \frac{1}{\left\| \sum_{i=1}^{q} \bar{\alpha}_i \nabla f_i(\bar{x}) \right\|} \tag{2.4}$$

Since it holds $\omega(\bar{x}) = -t_\omega$, it follows furthermore

$$t_\omega = -\omega(\bar{x}) = \max_{i=1,\ldots,q} \nabla f_i(\bar{x})^\top d_\omega \geq \nabla f_j(\bar{x})^\top d_\omega$$

for all $j \in \{1, 2, \ldots, q\}$. From (2.3) and (2.4) it follows

$$\begin{aligned} \omega(\bar{x}) &= -t_\omega = -\sum_{i=1}^{q} \bar{\alpha}_i t_\omega \\ &\leq -\sum_{i=1}^{q} \bar{\alpha}_i \nabla f_i(\bar{x})^\top d_\omega = \mu \left\| \sum_{i=1}^{q} \bar{\alpha}_i \nabla f_i(\bar{x}) \right\|^2 = \left\| \sum_{i=1}^{q} \bar{\alpha}_i \nabla f_i(\bar{x}) \right\|. \end{aligned}$$

□

Moreover, Pareto critical points can be characterized as given in Lemma 2.13. This is based on ideas presented in [DS15] and follows from Gordan's theorem [Gor73]. The proof can be found in [Tho15].

Lemma 2.13 *Let $f : \mathbb{R}^n \to \mathbb{R}^q$ be with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, continuously differentiable functions. A vector $x \in \mathbb{R}^n$ is Pareto critical for (MOP) if and only if there exists a vector $\lambda \in \mathbb{R}^q_+ \setminus \{0_q\}$ such that it holds $\sum_{i=1}^q \lambda_i \nabla f_i(x) = 0_n$.*

2.3 Convexity

Convexity is an important concept in the context of optimization methods since it gives information about the connection of local and global optimality. The concepts of this section result in special cases for the theoretical results of the multi-objective solution method presented in Chapter 4. The following basic definitions conform to [Mie98; Jah11].

Definition 2.14 dummy

(i) A nonempty set $S \subseteq \mathbb{R}^n$ is called *convex* if it holds $\lambda x + (1 - \lambda)y \in S$ for all $\lambda \in [0, 1]$ and for all $x, y \in S$.

(ii) Let $S \subseteq \mathbb{R}^n$ be a nonempty, convex set. A function $g : S \to \mathbb{R}$ is called *convex* if it holds $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$ for all $\lambda \in [0, 1]$ and for all $x, y \in S$. It is called *strictly convex* if it holds $g(\lambda x + (1 - \lambda)y) < \lambda g(x) + (1 - \lambda)g(y)$ for all $\lambda \in (0, 1)$ and for all $x, y \in S$ with $x \neq y$.

(iii) A function $f : \mathbb{R}^n \to \mathbb{R}^q$ with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ is called *(strictly) convex* or *(strictly) $\mathbb{R}^q_+$-convex* if the functions f_i are (strictly) convex for all $i \in \{1, 2, \ldots, q\}$.

(iv) The multi-objective optimization problem (MOP_c) is called *convex* if all the objective functions and the feasible set are convex.

If a function is continuously differentiable, convexity can be characterized using the gradient. The proof can be found for example in [GK99].

Lemma 2.15 *Let $S \subseteq \mathbb{R}^n$ be a nonempty, convex set and $g : S \to \mathbb{R}$ a continuously differentiable function. The following statements hold.*

(i) The function g is convex on S if and only if it holds $g(y) + \nabla g(y)^\top (x - y) \leq g(x)$ for all $x, y \in S$.

(ii) The function g is strictly convex on S if and only if it holds $g(y) + \nabla g(y)^\top (x - y) < g(x)$ for all $x, y \in S$ with $x \neq y$.

These characterizations of convexity for scalar-valued functions are helpful to describe the connection between Pareto critical and (weakly) efficient points of the multi-objective problem (MOP). If $f : \mathbb{R}^n \to \mathbb{R}^q$, the objective function of (MOP), is $\mathbb{R}^q_+$-convex, the Pareto critical points of (MOP) coincide with the weakly efficient points. If even strict $\mathbb{R}^q_+$-convexity of f is given, the Pareto critical points coincide with the efficient points. This is stated in Theorem 2.16 and can also be found in [FDS09, Thm. 3.1].

Theorem 2.16 *Let $f : \mathbb{R}^n \to \mathbb{R}^q$ be with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, continuously differentiable functions. The following statements hold.*

(i) Let f be $\mathbb{R}^q_+$-convex. The vector $\bar{x}$ is weakly efficient for (MOP) if and only if $\bar{x}$ is Pareto critical for (MOP).

(ii) Let f be strictly $\mathbb{R}^q_+$-convex. The vector $\bar{x}$ is efficient for (MOP) if and only if $\bar{x}$ is Pareto critical for (MOP).

Proof. To prove statement (i), let f be $\mathbb{R}^q_+$-convex. Every weakly efficient point for (MOP) is also locally weakly efficient and thus, according to Lemma 2.9, Pareto critical. Now let $\bar{x} \in \mathbb{R}^n$ be Pareto critical for (MOP) and $x \in \mathbb{R}^n$ an arbitrary point. Since f is $\mathbb{R}^q_+$-convex, it follows from Lemma 2.15

$$f_i(\bar{x}) + \nabla f_i(\bar{x})^\top (x - \bar{x}) \leq f_i(x) \tag{2.5}$$

for all $i \in \{1, 2, \ldots, q\}$. Since $\bar{x} \in \mathbb{R}^n$ is Pareto critical for (MOP), there exists an index $j \in \{1, 2, \ldots, q\}$ such that it holds $\nabla f_j(\bar{x})^\top (x - \bar{x}) \geq 0$. This implies together with (2.5)

$$f_j(\bar{x}) \leq f_j(\bar{x}) + \nabla f_j(\bar{x})^\top (x - \bar{x}) \leq f_j(x).$$

Thus, there exists no vector $y \in \mathbb{R}^n$ fulfilling $f_i(y) < f_i(\bar{x})$ for all $i \in \{1, 2, \ldots, q\}$ and $\bar{x}$ is weakly efficient for (MOP).

To prove (ii), let f be strictly $\mathbb{R}_+^q$-convex. If $\bar{x}$ is efficient for (MOP), it is also weakly efficient and according to (i) Pareto critical. Let $\bar{x}$ be Pareto critical for (MOP) and let $x \in \mathbb{R}^n \setminus \{\bar{x}\}$ be an arbitrary point. Analogously to the proof of statement (i) there exists an index $j \in \{1, 2, \ldots, q\}$ such that it holds

$$f_j(\bar{x}) \leq f_j(\bar{x}) + \nabla f_j(\bar{x})^\top (x - \bar{x}) < f_j(x)$$

due to Lemma 2.15. Thus, there exists no vector $x \in \mathbb{R}^n \setminus \{\bar{x}\}$ such that it holds $f(x) \leq f(\bar{x})$. Consequently, $\bar{x}$ is efficient for (MOP). □

Definition 2.17 Let $S \subseteq \mathbb{R}^n$ be a nonempty, convex set. If $g : S \to \mathbb{R}$ is continuously differentiable, it is called *pseudoconvex* on S if $\nabla g(y)^\top (x - y) \geq 0 \Rightarrow g(y) \leq g(x)$ holds for all $x, y \in S$.

It can be proved that every convex function is also pseudoconvex. In general, the reverse statement does not hold. The following theorem gives the connection between Pareto criticality and (weak) efficiency under the precondition of pseudoconvexity. The proof can for example be found in [Jah11].

Theorem 2.18 *Let $f : \mathbb{R}^n \to \mathbb{R}^q$ be with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$, continuously differentiable functions. If the functions f_i, $i = 1, 2, \ldots, q$, are pseudoconvex on $\mathbb{R}^n$, every Pareto critical point of (MOP) is weakly efficient for (MOP).*

These concepts of convexity and the connection of Pareto critical and weakly efficient points are used for special cases for the the theoretical results of the multi-objective solution method presented in Chapter 4.

3 Trust Region Concept

The multi-objective trust region algorithm for heterogeneous optimization problems presented in this thesis is based on the general trust region concept for single-objective optimization problems. Following [CGT00; CSV09b], this chapter gives an overview of the basic trust region approach including the main aspects for the theoretical results.

3.1 Basic Method

The basic trust region concept is formulated for unconstrained single-objective optimization problems of the form

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

with a twice continuously differentiable objective function $f : \mathbb{R}^n \to \mathbb{R}$ bounded from below. It is an iterative method which approximates the function f by a suitable model function $m^k : \mathbb{R}^n \to \mathbb{R}$ in every iteration $k \in \mathbb{N}$. A common choice are quadratic models, see also Section 3.2, and furthermore, only local model functions are required. In every iteration k the model functions are considered in a local area called trust region and defined by

$$B_k := B\left(x^k, \delta_k\right) = \left\{x \in \mathbb{R}^n \mid \left\|x - x^k\right\| \leq \delta_k\right\} \tag{3.1}$$

using the current iteration point x^k, the so-called trust region radius $\delta_k > 0$ and the euclidean norm $\|\cdot\| = \|\cdot\|_2$. The Manhattan or infinity norm can also be used and the norm can even be made dependent of the iteration k. Further information about the choice of different norms and model functions can be found for example in [CGT00]. On the one hand, the trust region restricts the local area for the model functions and therefore it helps to guarantee reliability of the model functions. On the other hand

and as outlined in the following, the computations in every iteration k are restricted to this area and therefore it functions also as a step size control.
Instead of the original function f the local model function m^k is considered in every iteration $k \in \mathbb{N}$ and minimized restricted to the trust region B_k. Solving the surrogate problem

$$\min_{x \in B_k} m^k(x)$$

is called the trust region subproblem. The model functions are supposed to be easier in the sense of solving this optimization problem. Since m^k represents the original function except for an error, it is in general not necessary to solve the trust region subproblem exactly. An approximate solution already provides a good candidate for the next iteration point, the so-called trial point x^{k+}.
As outlined in [CGT00; CSV09b], the steepest descent idea can be used for this purpose by computing the so-called Cauchy point. It provides a decrease for the model function in the trust region and can be obtained with reasonable effort for quadratic models, see [CGT00; CSV09b]. Quadratic models are most commonly used in the trust region approach and various approaches exist in the literature to solve such trust region subproblems, for example the generalized Lanczos method [Gou+99; CGT00] or methods based on the conjugate gradient method [HS52; Toi81; Ste83; CGT00]. General information about the solvability of trust region subproblems is for example given in [Mor83; BM13].
For the convergence results, it is required that the trial point x^{k+} provides a sufficient decrease for the model function in terms of the gradient $\nabla m^k(x^k)$. We will come back to that later when we shortly describe the convergence analysis.
The trial point shall not only provide a sufficient decrease for the model function in the trust region, but also for the original function f. Thus, having obtained the trial point, the function value $f(x^{k+})$ is computed and compared to the value predicted by the model. This is realized by the quotient

$$\rho_k := \frac{f(x^k) - f(x^{k+})}{m^k(x^k) - m^k(x^{k+})}. \tag{3.2}$$

Note that the denominator is always positive due to the computation of the trial point x^{k+}. Hence, whenever $\rho_k > 0$ holds, the trial point also provides a decrease for the original function.

The trial point is accepted as next iteration point if $\rho_k \geq \eta_1$ holds with $\eta_1 > 0$ a given constant. In this case, the prediction of the model was good and the trial point x^{k+} provides a sufficient decrease for f. Otherwise, the approximation by the model was not accurate enough and the trial point is not accepted. Based on the value of ρ_k also the trust region B_k is updated for the next iteration. A general description of the trust region approach including the trust region update is given in Algorithm 1.

Algorithm 1 Basic single-objective trust region algorithm

Input: Function $f : \mathbb{R}^n \to \mathbb{R}$, 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 $f(x^0)$.

Step 1: Model definition

Compute model m^k in $B_k = \left\{x \in \mathbb{R}^n \mid \left\|x - x^k\right\| \leq \delta_k\right\}$.

Step 2: Step calculation

Compute step s^k that sufficiently reduces the model m^k such that $x^{k+} = x^k + s^k \in B_k$.

Step 3: Trial point acceptance test

Compute $f(x^{k+})$ and $\rho^k = \frac{f(x^k) - f(x^{k+})}{m^k(x^k) - m^k(x^{k+})}$.

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

Step 4: Trust region update

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

Set $k = k + 1$ and go to **Step 1**.

The choice of the constants η_1, η_2, γ_1 and γ_2 can of course depend on the optimization problem. According to [CGT00], reasonable values are $\eta_1 = 0.01$ or $\eta_1 = 0.05$, $\eta_2 = 0.9$ and $\gamma_1 = \gamma_2 = 0.5$.

The procedure of Algorithm 1 is illustrated schematically in Figure 3.1 for a function $f : \mathbb{R} \to \mathbb{R}$. The left figure shows the trust region B_k, the iteration point x^k and the trial point x^{k+} which is accepted as next iteration point. The right figure shows the next iteration where a new trial point x^{k+1+} in the trust region B_{k+1} is computed.

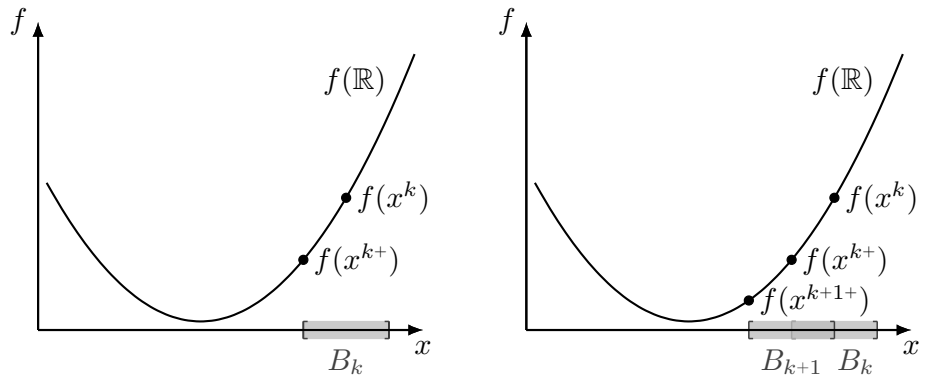

Figure 3.1 – Schematical procedure of Algorithm 1

The trust region approach in Algorithm 1 is formulated according to the literature without a stopping criterion. For the implementation, one option is the norm of the gradient. Hence, the algorithm stops if

$$\left\|\nabla f(x^k)\right\| \leq \varepsilon_1$$

holds for a suitable constant $\varepsilon_1 > 0$. Since this is an indicator for stationarity, it is a reasonable choice and commonly used. However, such a test is only possible if derivative information is available with reasonable numerical effort. Otherwise and if the value $\nabla m^k(x^k)$ is reliable, it can be used as a surrogate. Another option for a stopping criterion free of these concerns is

$$\delta_k \leq \varepsilon_2 \left\|x^k\right\|$$

with a suitable constant $\varepsilon_2 > 0$. This is reasonable as a stopping criterion since the convergence results show that the trust region radius converges to zero if a stationary point is approached. These stopping criteria including further numerical details are discussed in [CGT00].

In every trust region approach the iterations are classified according to their outcome. An iteration $k \in \mathbb{N}$ is called successful, if $\rho_k \geq \eta_1$ holds and thus if the trial point is accepted as next iteration point, that is it holds $x^{k+1} = x^{k+}$. The set of indices of all successful iterations is denoted by

$$\mathcal{S} \coloneqq \{k \in \mathbb{N} | \rho_k \geq \eta_1\}.$$

Similarly, the set of iterations

$$\mathcal{V} := \{k \in \mathbb{N} | \rho_k \geq \eta_2\}$$

is defined as the set of very successful iterations. Apparently, it holds $\mathcal{V} \subseteq \mathcal{S}$. Besides, all indices not contained in $\mathcal{S}$ are called unsuccessful. In these iterations the trial point x^{k+} is discarded and the next iteration point is defined as $x^{k+1} = x^k$.
The convergence analysis as for example in [CGT00; CSV09b] shows that the trust region approach is a first-order method, i.e. convergence to a stationary point can be proved. Since the single-objective method is only the base for the multi-objective method presented in Chapter 4, we do not present the convergence results here.
However, the assumptions for the multi-objective method are derived from the assumptions of the single-objective approach and the convergence analysis transfers the main aspects of the convergence analysis from the single-objective approach. Therefore, based on [CGT00; CSV09b], we give a small overview of the necessary assumptions and the main components of the convergence results in the single-objective approach and outline the ideas.
Note that we consider the Frobenius norm $\|\cdot\|_F$ as matrix norm which is compatible with the Euclidean norm which is chosen as vector norm. The necessary assumptions are the following.

(1) The objective function $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable on $\mathbb{R}^n$.

(2) f is bounded from below, i.e. there exists a constant $\kappa_1 \in \mathbb{R}$ such that $f(x) \geq \kappa_1$ holds for all $x \in \mathbb{R}^n$.

(3) The Hessian of f is uniformly bounded, i.e. there exists a constant $\kappa_2 > 0$ such that $\|\nabla^2 f(x)\|_F \leq \kappa_2$ holds for all $x \in \mathbb{R}^n$.

(4) The model function $m^k : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable for all $k \in \mathbb{N}$.

(5) The model functions are exact at the current iteration point, i.e. it holds $m^k(x^k) = f(x^k)$ for all $k \in \mathbb{N}$.

(6a) The gradient of the model function is exact at the current iteration point, i.e. it holds $\nabla m^k(x^k) = \nabla f(x^k)$ for all $k \in \mathbb{N}$, or

(6b) there exists a constant $\kappa_3 \geq 0$ such that $\left\|\nabla f(x^k) - \nabla m^k(x^k)\right\| \leq \kappa_3 \left\|\nabla m^k(x^k)\right\|$ holds for all $k \in \mathbb{N}$.

(7) The Hessian of the model is bounded in the trust region, i.e. there exists a constant $\kappa_4 \geq 1$ such that $\left\|\nabla^2 m^k(x)\right\|_F \leq \kappa_4 - 1$ holds for all $x \in B_k$ and $k \in \mathbb{N}$.

(8) There exists a constant $\kappa_5 \in (0,1)$ such that it holds
$m^k(x^k) - m^k(x^{k+}) \geq \kappa_5 \left\|\nabla m^k(x^k)\right\| \min\left\{\frac{\left\|\nabla m^k(x^k)\right\|}{\beta_k}, \delta_k\right\}$
for all $k \in \mathbb{N}$ with $\beta_k := 1 + \max_{x \in B_k} \left\|\nabla^2 m^k(x)\right\|_F$.

Similar assumptions to (1)-(8) are also required for the multi-objective trust region method presented in Chapter 4, see Assumptions A.1 to A.7 and A.11 in Section 4.5. Some further assumptions are needed there due to the multi-objective framework.
Assumption (8) is particularly important in the trust region scheme. It is referred to as sufficient decrease condition since it ensures that the trial point provides a sufficient decrease for the model function. An analogous assumption is also formulated in the multi-objective case, see Assumption A.11 in Section 4.5. This assumption and the lemmas justifying it is one of the main aspects of the convergence analysis of both single- and multi-objective trust region methods.
Based on the assumptions (1)-(8), convergence of first order can be proved. Since we want to outline the idea behind the main steps of the convergence analysis in the single-objective case, we do not formulate them as lemmas, but only list them with headwords. For all of them the assumptions (1)-(8) are assumed to hold.

(I) Accuracy of the model function in the trial point:
$\left|f(x^{k+}) - m^k(x^{k+})\right| \leq \kappa_5 \delta_k^2$ for all $k \in \mathbb{N}$; $\kappa_5 > 0$ suitable constant

(II) Existence of successful iterations if the current iteration point is non-stationary:
$\left\|\nabla m^k(x^k)\right\| \geq \varepsilon_1$ and $\delta < \varepsilon_2 \Rightarrow k \in \mathcal{V}$; $\varepsilon_1, \varepsilon_2 > 0$ suitable constants

(III) Convergence to stationary point in case of finitely many successful iterations:
$|\mathcal{S}|$ finite $\Rightarrow x^k = \bar{x}$ for all k sufficiently large, $\bar{x}$ stationary point

Analogous results and further auxiliary results due to the search strategy are used in the convergence analysis for the multi-objective trust region method in Section 4.5.
Let $\left\{x^k\right\}$ be a sequence of iterates obtained by a trust region method as described in Algorithm 1. Given the assumptions (1)-(8) and the main results (I)-(III), it can be proved – with further auxiliary results – that $\lim_{k\to\infty} \left\|\nabla f(x^k)\right\| = 0$ holds. Thus, if

the sequence of iterates has accumulation points, every of these points is a stationary point. Under stronger assumptions, also second order convergence can be proved, see for example [CGT00].

3.2 Model Functions

As described in the former section, in every trust region approach the functions are replaced by local model functions. Most commonly, quadratic models are used given by

$$m(x) = m(y) + g(y)^\top (x - y) + \frac{1}{2} (x - y)^\top H (x - y) \tag{3.3}$$

with $m(y) = f(y)$ for a fixed point $y \in \mathbb{R}^n$, $g(y) = \nabla f(y)$ and H a symmetric approximation of the Hessian $\nabla^2 f(y)$. This is also referred to as first-order Taylor model, see [CSV09b]. In the framework of Algorithm 1, the point y is chosen as the current iteration point x^k.
Since the computed model functions are only local approximations, it is important to ensure a good local accuracy of the model function. This is commonly defined in the following way, see also [CGT00, Section 9.1].

Definition 3.1 Consider a function $f : \mathbb{R}^n \to \mathbb{R}$ and the local area $B = B(\tilde{x}, \delta) = \{x \in \mathbb{R}^n | \, \|x - \tilde{x}\| \leq \delta\}$ for a point $\tilde{x} \in \mathbb{R}^n$ and a constant $\delta > 0$. A function $m : \mathbb{R}^n \to \mathbb{R}$ is called a *valid model* for the function f in B, if it holds

$$|f(x) - m(x)| \leq \kappa \delta^2$$

for all $x \in B(\tilde{x}, \delta)$ and some constant $\kappa > 0$ independent of x.

If derivative information of second order is available for the function f, the second-order Taylor model can be used, see also [CSV09b].

Definition 3.2 For a twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and a point $y \in \mathbb{R}^n$ with gradient $\nabla f(y)$ and Hessian $\nabla^2 f(y)$ the *(second-order) Taylor model* is defined by

$$m(x) = m_T(x; f, y) \coloneqq f(y) + \nabla f(y)^\top (x - y) + \frac{1}{2} (x - y)^\top \nabla^2 f(y) (x - y) \,. \tag{3.4}$$

The trust region approach is very flexible and due to the model approach it is also well suited for optimization problems where derivative information is absent or not reliable. A derivative-free formulation of the trust region idea is described in [CSV09b; AH17].
If derivative information is not available, polynomial interpolation is an appropriate choice for the model functions. The trust region algorithm presented in Chapter 4 uses a special type of interpolation models. Hereafter, a short overview of these models is given conforming to [CSV09b]. More detailed information can be found there.

3.2.1 Polynomial Interpolation

Consider $\mathcal{P}_n^d$ the space of polynomials of degree less than or equal to d in $\mathbb{R}^n$. Let p be the dimension of this space. For $d = 1$, i.e. $\mathcal{P}_n^d$ contains only linear polynomials, it holds

$$p = n + 1$$

and for $d = 2$, that is $\mathcal{P}_n^d$ contains linear and quadratic polynomials, it holds

$$p = \frac{(n+1)(n+2)}{2}.$$

Now consider a basis $\psi = \{\psi_1, \psi_2, \ldots, \psi_p\}$ of $\mathcal{P}_n^d$, then every polynomial $m(x) \in \mathcal{P}_n^d$ can be expressed as

$$m(x) = \sum_{j=1}^{p} \alpha_j \psi_j(x) \tag{3.5}$$

with $\alpha \in \mathbb{R}^p$ some suitable coefficients. Of course different bases can be considerd. The simplest polynomial basis is the monomial basis defined by

$$\bar{\psi} \coloneqq \left\{\bar{\psi}_1, \bar{\psi}_2, \ldots, \bar{\psi}_p\right\} = \left\{1, x_1, x_2, \ldots, x_n, \frac{x_1^2}{2}, x_1 x_2, \ldots, \frac{x_{n-1}^{d-1} x_n}{(d-1)!}, \frac{x_n^d}{d!}\right\}. \tag{3.6}$$

In this thesis we consider the basis of so-called Lagrange polynomials. They are most commonly used in multivariate polynomial interpolation literature to define poisedness – a quality criterion for interpolation points.

Definition 3.3 Given a set of interpolation points $Y = \{y^1, y^2, \ldots, y^p\} \subset \mathbb{R}^n$, a basis of p polynomials $l_i, i = 1, 2, \ldots, p$, in $\mathcal{P}_n^d$ is called a *basis of Lagrange polynomials* if it holds

$$l_i(y^j) = \begin{cases} 1 & , \text{ if } i = j \\ 0 & , \text{ else} \end{cases}.$$

For $n = 1$ the Lagrange polynomials $l_i, i = 1, 2, \ldots, p$, can be computed by

$$l_i(x) = \prod_{\substack{j=1 \\ j \neq i}}^{p} \frac{x - y^j}{y^i - y^j}$$

for all $i = 1, 2, \ldots, p$. For higher dimensions it is more complex; a general method to compute the Lagrange polynomials is presented in Algorithm 2 later in this section.

A polynomial m interpolates the function f at a given point y if $m(y) = f(y)$ holds. Let $Y = \{y^1, y^2, \ldots, y^p\} \subset \mathbb{R}^n$ be a set of interpolation points and let $m \in \mathcal{P}_n^d$ denote a polynomial that interpolates a given function $f : \mathbb{R}^n \to \mathbb{R}$ at the points in Y. The conditions

$$m(y^i) = f(y^i) \text{ for all } i = 1, 2, \ldots, p$$

are called interpolation conditions. Considering the general formulation of an interpolating polynomial given by (3.5), the interpolating polynomial m is determined if the coefficients α_i are determined for all $i \in \{1, 2, \ldots, p\}$. In order to determine these coefficients, the interpolation conditions can be used:

$$m(y^i) = \sum_{j=1}^{p} \alpha_j \psi_j(y^j) = f(y^i), i = 1, 2, \ldots, p. \tag{3.7}$$

The conditions (3.7) form a linear system in terms of the interpolation coefficients which can be written in matrix form as

$$M(\psi, Y)\alpha = f(Y) \tag{3.8}$$

with $(M(\psi, Y))_{ij} = \psi_j(y^i)$, $\alpha = (\alpha_1, \ldots, \alpha_p)^\top$ and $f(Y) = (f(y^1), \ldots, f(y^p))^\top$. This system of equations has a unique solution if the matrix $M(\psi, Y)$ is nonsingular.
Interpolation using Lagrange polynomials as it is presented in this and the follow-

ing subsection will be used in the trust region method presented Chapter 4. There, the model depends in particular on the current iteration point and the current trust region which is the reason for defining the interpolation model based on Lagrange polynomials as follows.

Definition 3.4 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function and $Y = \{y^1, y^2, \dots, y^p\} \subset \mathbb{R}^n$ a set of interpolation points with $B \coloneqq B\,(y^1, \delta) = \{y \in \mathbb{R}^n | \, \|y - y^1\| \leq \delta\}, \delta > 0$ and $y^j \in B$ for all $j = 1, 2, \dots, p$. The interpolation polynomial using the basis of Lagrange polynomials $l_i, i = 1, 2, \dots, p$, is defined by

$$m(x) = m_L(x; f, y^1, B) \coloneqq \sum_{i=1}^{p} f(y^i) l_i(x) \tag{3.9}$$

and is referred to as *Lagrange model*.

Usually, quadratic models are used for the trust region approach, i.e. it holds $d = 2$ and the Lagrange polynomials are chosen from $\mathcal{P}_n^2$. As outlined in the beginning of this section, computing one quadratic interpolation model requires $p = (n + 1)(n + 2)/2$ function values. However, for high dimensional expensive functions this amount of function evaluations can cause too much numerical effort. In this case, linear models $(d = 1)$ can be used. The Lagrange models are then linear interpolation functions from $\mathcal{P}_n^1$. The theory and the algorithms presented in the following apply for the general case of $d \in \mathbb{N}$.
However, for the algorithm presented in this thesis quadratic model functions are used by default and linear models are only an option for a modification according to the user. Further information about the use of linear models can for example be found in Section 4.6.1.

3.2.2 Choice of Interpolation Points

An important aspect in modeling using interpolation is the choice of the interpolation points. The quality of the model depends on the quality of the interpolation points. Following [CSV09b] we use the concept of poisedness.

Definition 3.5 The set $Y = \{y^1, y^2, \dots, y^p\}$ is *poised* for polynomial interpolation in $\mathbb{R}^n$ if the corresponding matrix $M(\psi, Y)$, defined in (3.8), is nonsingular for some basis ψ in $\mathcal{P}_n^d$.

To clarify the meaning of poisedness, let the matrix $M(\psi, Y)$ be singular. Then there exits a vector $\gamma \in \mathbb{R}^p \setminus \{0\}$ with $M(\psi, Y)\gamma = 0$ and it holds $\sum_{i=1}^p \gamma_i \psi_i(y) = 0$ for all $y \in Y$. Thus, there exists a polynomial of degree at most d given by $m(x) \coloneqq \sum_{j=1}^p \gamma_j \psi_j(x)$ and it holds $m(y) = 0$ for all $y \in Y$.
Thus, if Y shall form a poised set, this means that the interpolation points must not lie on a line in case linear interpolation models are considered. If quadratic interpolation models are considered and Y shall form a poised set, the interpolation points must not lie on any quadratic surface in $\mathbb{R}^n$, see also [CGT00; CSV09b; AH17].
In the literature, poised sets are also referred to as d-unisolvent sets. It can be proved that if $M(\psi, Y)$ is nonsingular for some basis ψ, then it is nonsingular for any basis of $\mathcal{P}_n^d$. If the interpolation points form a poised set, the interpolating polynomial m exists and is unique. This is stated in Lemma 3.6; the proof can for example be found in [CSV09b, Lem. 3.2].

Lemma 3.6 *Given a function $f : \mathbb{R}^n \to \mathbb{R}$ and a poised set $Y \subset \mathbb{R}^n$, the interpolating polynomial m defined by* (3.5) *exists and is unique.*

Lemma 3.6 holds for a general basis. If the set of interpolation points is poised, it can be proved that the Lagrange polynomials exist, are uniquely defined and actually form a basis of $\mathcal{P}_n^d$. Hence, Lemma 3.6 holds in particular for the basis of the Lagrange polynomials. This is given in Lemmas 3.7 and 3.8 and can be found in detail and with proof in [CSV09b, Lem. 3.4, Lem. 3.5].

Lemma 3.7 *Let $Y \subset \mathbb{R}^n$ be a poised set. Then the basis of Lagrange polynomials from Definition* 3.3 *exists and is uniquely defined.*

Lemma 3.8 *For any function $f : \mathbb{R}^n \to \mathbb{R}$ and any poised set $Y = \{y^1, y^2, \dots, y^p\} \subset \mathbb{R}^n$, the unique polynomial m that interpolates f on Y can be expressed as*

$$m(x) = \sum_{i=1}^{p} f(y^i) l_i(x),$$

where $\{l_1, l_2, \dots, l_p\}$ is the basis of Lagrange polynomials for Y.

Furthermore, from the basic statement in Lemma 3.6 the following statement about exactness of the Lagrange model for quadratic functions follows.

Lemma 3.9 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic function, that is $f \in \mathcal{P}_n^2$. Moreover, let $Y = \{y^1, y^2, \dots, y^p\} \subset \mathbb{R}^n$ be a poised set of interpolation points and let the Lagrange model m_L be given by $m_L(x) = \sum_{i=1}^p f(y^i) l_i(x)$ according to Definition 3.4 with $l_i \in \mathcal{P}_n^2$, $i \in \{1, 2, \dots, q\}$ the Lagrange polynomials. Then the Lagrange model is exact, that is it holds $f \equiv m_L$.*

Proof. According to Lemma 3.7, the Lagrange polynomials form a basis of $\mathcal{P}_n^2$ since Y is a poised set. Consequently, there exist scalars $c_i \in \mathbb{R}$ for $i \in \{1, 2, \dots, p\}$ such that it holds $f(x) = \sum_{i=1}^p c_i l_i(x)$ for all $x \in \mathbb{R}^n$. Due to the definition of the Lagrange polynomials it holds

$$f(y^j) = \sum_{i=1}^{p} c_i l_i(y^j) = c_j$$

for all $j \in \{1, 2, \dots, p\}$. Then the definition of m_L implies $f \equiv m_L$. □

Let $Y = \{y^1, y^2, \dots, y^p\}$ be a poised set of interpolation points with p the dimension of $\mathcal{P}_n^d$. The Lagrange polynomials can be computed by Algorithm 2 [CSV09b, Alg. 6.1].

Algorithm 2 Computing Lagrange polynomials

Initialization: Polynomials $\bar{\psi}_i$ of monomial basis (3.6) as initial polynomials $l_i, i = 1, 2, \dots, p$, set of interpolation points Y, $|Y| = p$
for $i = 1, 2, \dots, p$ **do**
  Point selection: Compute $j_i = \underset{i \le j \le p}{\operatorname{argmax}} |l_i(y^j)|$.
    If $|l_i(y^{j_i})| = 0$, then STOP (Y is not poised).
    Otherwise, exchange points y^i and y^{j_i} in set Y.
  Normalization: $l_i(x) = l_i(x)/l_i(y^i)$
  Orthogonalization: $l_j(x) = l_j(x) - l_j(y^i) l_i(x)$ for $j = 1, 2, \dots, p, j \neq i$
end for

Algorithm 2 terminates prematurely if the set of interpolation points is not poised. Algorithm 3 [CSV09b, Alg. 6.2] describes a method to complete a nonpoised set to a poised set using the Lagrange polynomials.
As input a set of interpolation points Y and a closed ball B are required. The set Y can

contain one point or several points. The algorithm completes this set to a poised set and computes new points if necessary. These points are chosen from B. In the trust region framework B is defined as the trust region $B := B_k$.

Algorithm 3 Completing a nonpoised set by Lagrange polynomials

Initialization: Polynomials $\bar{\psi}_i$ of monomial basis (3.6) as initial polynomials $l_i, i = 1, 2, \ldots, p$, initial set of interpolation points Y, $|Y| = p_{ini}$, closed ball B
for $i = 1, 2, \ldots, p$ **do**
Point selection: Compute $j_i = \underset{i \leq j \leq p_{ini}}{\operatorname{argmax}} |l_i(y^j)|$.
If $|l_i(y^{j_i})| > 0$ and $i \leq p_{ini}$, exchange points y^i and y^{j_i} in set Y.
Otherwise, compute (or recompute if $i \leq p_{ini}$) y^i as

$$y^i \in \underset{x \in B}{\operatorname{argmax}} |l_i(x)|.$$

Normalization: $l_i(x) = l_i(x)/l_i(y^i)$
Orthogonalization: $l_j(x) = l_j(x) - l_j(y^i) l_i(x)$ for $j = 1, 2, \ldots, p, j \neq i$
end for

Example 3.10 Figure 3.2 illustrates two examples for the choice of interpolation points based on [Lau18, example 3.4]. The considered function is $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x) = 0.5x_1^3 + x_2^2 - 10x_1 - 100$. Two sets of interpolation points X_1 and X_2 with the corresponding Lagrange models $m^1, m^2 : \mathbb{R}^2 \to \mathbb{R}$ from Definition 3.4 are considered. In the left figure the model m^1 based on the interpolation points

$$X_1 = \left\{(16, 37)^\top, (30, 35)^\top, (33, 21)^\top, (27, 7)^\top, (12, 9)^\top, (9, 25)^\top\right\}$$

is illustrated. The points nearly lie on a circle. The right figure shows the model function m^2 resulting from the set of interpolation points

$$X_2 = \left\{(21.00, 22.00)^\top, (41.00, 22.00)^\top, (21.00, 42.00)^\top, (31.00, 22.77)^\top, (38.01, 32.52)^\top, (29.42, 40.14)^\top\right\}.$$

These points form a poised set and are computed by Algorithm 3 with $(21.00, 22.00)^\top$ as input. Both sets of interpolation points cover a similar area. However, the interpolation model m^2 based on the poised set X_2 gives a better local approximation of f.

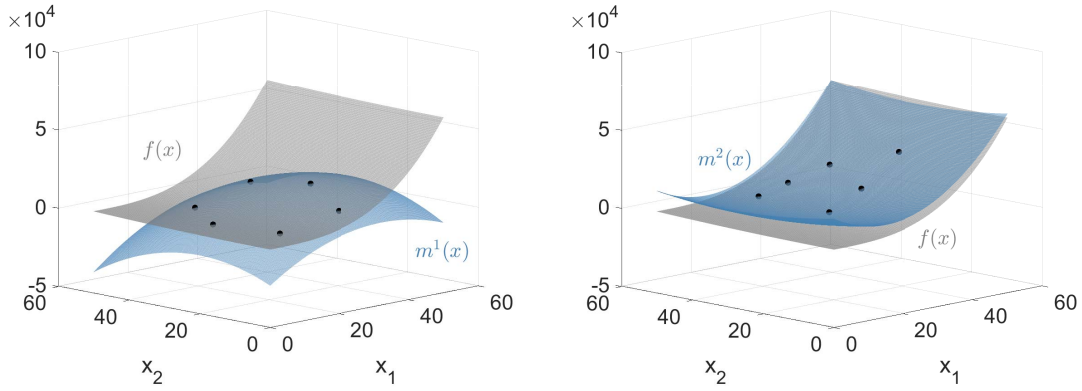

Figure 3.2 – Model functions m^1 and m^2 based on interpolation points X_1 (left) and X_2 (right)

The effects of different choices of interpolation points is discussed more detailed in [Lau18]. Further information is also given in [CSV09b, Section 3.3].

4 Multi-Objective Trust Region Method

In this chapter an algorithm for unconstrained multi-objective optimization problems with heterogeneous objective functions is presented. The idea of this algorithm including the convergence results from Section 4.5 are firstly presented in [TE18a].
The algorithm uses the basic trust region scheme from Chapter 3 and transfers it to multi-objective optimization problems similar to [RK14; VOS14]. The new aspects of this solution method are the search direction which is defined in the image space and the fact that heterogeneous objective functions can be considered. We consider the optimization problem

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

with $f(x) = (f_1(x), f_2(x), \ldots, f_q(x))^\top$ and twice continuously differentiable functions $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, q$. Let f_1 be an expensive black-box function which is not given analytically. It is assumed that computing function values is time-consuming and that derivative information is not available with reasonable numerical effort. Thus, derivatives are not used for f_1 in the optimization method presented hereafter. Let $f_2, f_3, \ldots, f_q$, be cheap, analytically given functions for which function values and derivatives can easily be computed.
The basic version of the multi-objective heterogeneous trust region algorithm MHT is presented in Section 4.1. The individual steps of the algorithm are explained in detail in Sections 4.2 to 4.4. The convergence analysis and theoretical results are given in Section 4.5; it is proved that the accumulation point of the generated sequence of iterates is a Pareto critical point. For this purpose, several assumptions are needed. They are introduced in this chapter when required and listed for an overview in Appendix A. In Section 4.6 some minor modifications of MHT are presented for which the convergence results are transferable.
The algorithm can be modified to handle box constraints. However, since the algorithm is constructed for unconstrained problems, the convergence results cannot be

transferred. The algorithmic realization of box constraints and the theoretical results that hold are presented in Section 4.7.

4.1 Algorithm MHT

MHT is an iterative method and following the basic trust region approach from Chapter 3 in every iteration $k \in \mathbb{N}$ the computations are restricted to the local area

$$B_k = \left\{ x \in \mathbb{R}^n \middle| \left\| x - x^k \right\| \leq \delta_k \right\} \tag{4.1}$$

called trust region. It is defined by the current iteration point x^k, the trust region radius $\delta_k > 0$ and the Euclidean norm $\|\cdot\| = \|\cdot\|_2$. Conforming to the trust region approach, the following assumption is used.

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

This is a strong assumption for the black-box function f_1. However, such an assumption is required and commonly used in methods for expensive functions if a first-order convergence proof is given, see also [AH17].
In every iteration k the objective functions $f_i, i = 1, 2, \ldots, q$, are replaced by local quadratic models $m_i^k : \mathbb{R}^n \to \mathbb{R}$ which are supposed to fulfill the equality

$$f_i(x^k) = m_i^k(x^k) \tag{4.2}$$

for all $i \in \{1, 2, \ldots, q\}$, i.e. they are supposed to be exact at the current iteration point. Further details about the model functions are described in Section 4.2. The model functions are considered as surrogates to the original functions. In every iteration $k \in \mathbb{N}$ the trust region subproblem

$$\min_{x \in B_k} m^k(x) \tag{MOP_m^k}$$

with $m^k(x) = (m_1^k(x), m_2^k(x), \ldots, m_q^k(x))^\top$ is considered. The model functions are supposed to be easier than the original functions in the sense of solving the trust region subproblem. This surrogate optimization problem and the local model functions are used to compute a sufficient decrease for the original optimization problem. As a search direction the local ideal point $p^k = (p_1^k, p_2^k, \ldots, p_q^k)^\top$ defined by

$$p_i^k = \min_{x \in B_k} m_i^k(x) \tag{4.3}$$

for $i = 1, 2, \ldots, q$ is used. These subproblems need to be solved in every iteration. They are classic trust region subproblems since a quadratic function is minimized over a trust region. Information of how these subproblems are solved in the literature can be found in Section 3.1. However, since (4.3) defines a quadratic optimization problem, any quadratic solver can be used.

The ideal point p^k gives a direction for decreasing the model functions and, depending on the quality of the model functions, also the original functions. The aim is to move as far as possible – as far as the trust region allows – into the direction of p^k. A similar idea is used in methods based on reference points, see for example [Mie98].

The trust region functions not only as a guarantee that the models are reliable approximations of the original functions, but also as a step size control. Moving towards the ideal point is realized by the Pascoletti-Serafini scalarization [PS84] given by

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

with $r^k \coloneqq f(x^k) - p^k \in \mathbb{R}^q_+$ and p^k the ideal point defined in (4.3). This auxiliary problem is also known as Tammer-Weidner functional [GW90] and is further discussed in Section 4.3. Note that it holds $f(x^k) = m^k(x^k)$ in every iteration k due to (4.2) which implies $r^k \in \mathbb{R}^q_+$.

Solving the auxiliary problem (PS), we obtain the trial point x^{k+}, a candidate for the next iteration point. Figure 4.1 illustrates the idea for a bi-objective optimization problem ((MOP) with $q = 2$). The trial point given by a minimal solution $\left(t^{k+}, x^{k+}\right)$ of (PS) is marked black. To clarify the idea of the search direction and the effect of (PS), the cone $C_1 \coloneqq m^k(x^k) - t^{k+}\, r^k - \mathbb{R}^2_+$ given by the minimal solution is depicted as well as the cone $C_2 \coloneqq m^k(x^k) - t\, r^k - \mathbb{R}^2_+$ given by a feasible scalar $t > t^{k+}$.

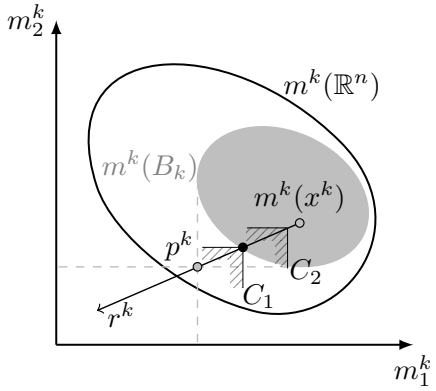

Figure 4.1 – Pascoletti-Serafini problem (PS) in iteration k

Analogously to the single-objective trust region method of Chapter 3 the trial point x^{k+} is only accepted as next iteration point if a condition describing the improvement of the function values is fulfilled. We use the same approach as [VOS14] defining the functions

$$\phi(x) \coloneqq \max_{i=1,\dots,q} f_i(x) \text{ and } \phi_m^k(x) \coloneqq \max_{i=1,\dots,q} m_i^k(x) \tag{4.4}$$

to examine if

$$\rho_\phi^k \coloneqq \frac{\phi(x^k) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})} \tag{4.5}$$

is bigger than a given positive constant. In this case, there is a guaranteed descent in at least one component. A detailed discussion of this multi-objective condition for the trial point acceptance test is given in Section 4.4.

The Multi-objective Heterogeneous Trust region algorithm (MHT) is formulated in Algorithm 4. It describes a trust region approach in which heterogeneity of the objective functions can be considered and which differs from other trust region methods by the computation of the search direction.

As input a starting point, some parameters and the objective functions are required. Note that f_1 is still an expensive black-box function and $f_2, f_3, \dots, f_q$ are analytically given, cheap functions. The choice of the parameters η_1, η_2, γ_1 and γ_2 can of course be problem-dependent, but according to [CGT00] reasonable values are $\eta_1 = 0.01, \eta_2 = 0.9$ and $\gamma_1 = \gamma_2 = \frac{1}{2}$. If $r_i^k = 0$ holds for an index $i \in \{1, 2, \dots, q\}$, x^k is weakly

efficient for (MOP_m^k), see also Section 4.3. In this case, (PS) is not considered in step 2 of MHT, but the trial point is defined as $x^{k+} = x^k$.

Algorithm 4 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 (PS) : $\min \left\{ t \in \mathbb{R} \mid f(x^k) + t\, r^k - m^k(x) \in \mathbb{R}^q_+, x \in B_k \right\}$.

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

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

Step 5: Model update

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

Algorithm 4 includes no stopping criterion. This is not required for the theoretical results in Section 4.5. However, for the implementation a suitable stopping criterion is required. Different options are presented and discussed in Section 6.1 along with other numerical details for the implementation of MHT. Moreover, the update rule for the trust region radius in Step 4 of MHT is formulated in a general way using constants γ_1, γ_2 and intervals. This formulation conforms to the literature, see for example [CGT00; VOS14], and is sufficient for the theoretical considerations. The numerical realization is described in Section 6.1.

Figure 4.2 gives a schematical overview of the procedure in the image space for a bi-objective optimization problem, i.e. (MOP) with $q = 2$. Figures a) and b) show the computations in iteration k. Starting from the current iteration point x^k, the local ideal point p^k is computed. The trial point x^{k+} is computed by moving into this direction

using the auxiliary problem (PS). We assume the trial point acceptance test gives a positive result and the trial point is accepted, i.e. it holds $x^{k+1} = x^{k+}$ as depicted in figure b).

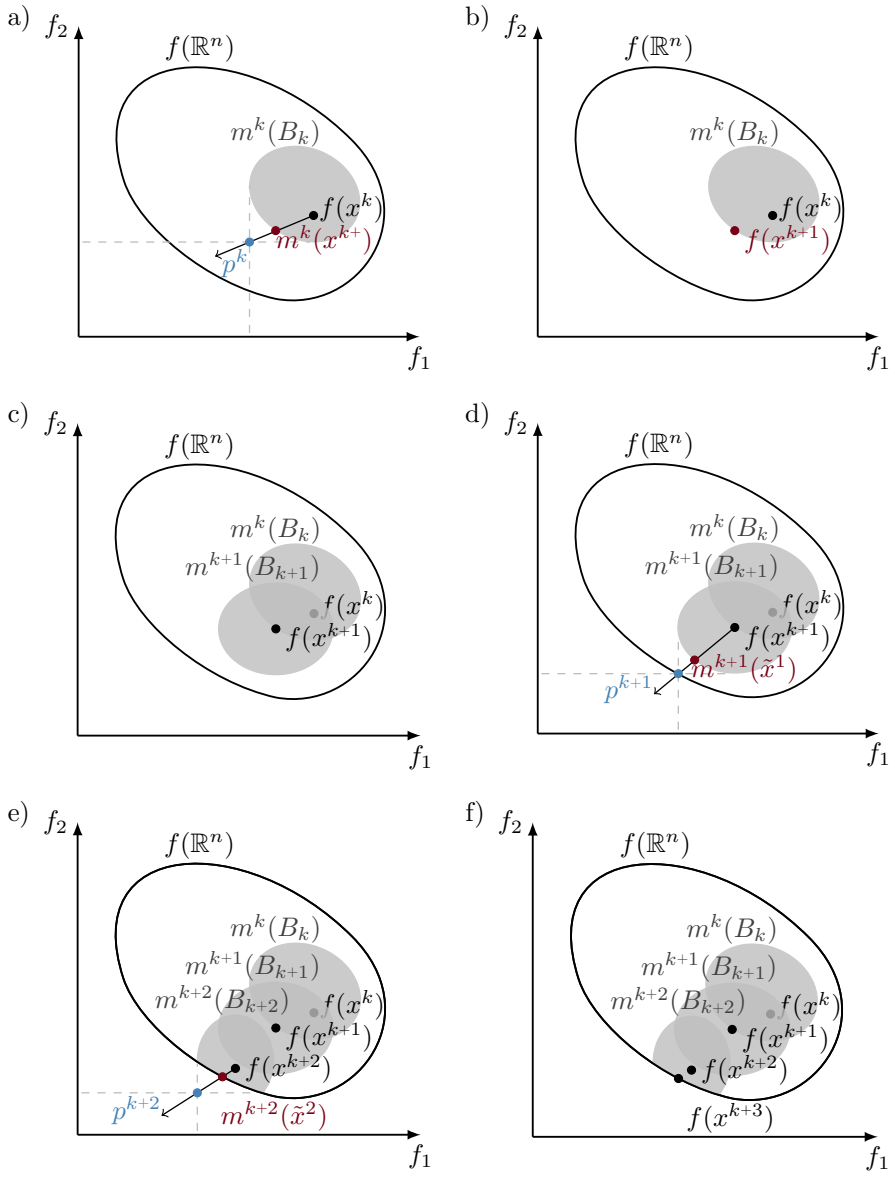

Figure 4.2 – Sketch of algorithm MHT : computing local ideal points and using (PS) to compute trial points

Consequently, the next trust region is built around x^{k+1} which is illustrated in the image space in figure c). In this new trust region again the local ideal point p^{k+1} is computed and the next trial point denoted by $\tilde{x}^1$ is obtained by moving into the di-

rection of p^{k+1} as far as possible. This is illustrated in figure d). Again, the trial point acceptance test decides if this point is suitable as next iteration point. We assume that $\tilde{x}^1$ is accepted.
Figure e) shows the computations in iteration $k+2$ with the trial point denoted as $\tilde{x}^2$ which is accepted as next iteration point in figure f). This point is efficient and in the subsequent iterations the trust region radius will reduce and the iteration point will not change any more.

4.2 Model Functions

In the trust region approach most commonly quadratic models are used, see also Section 3.2. For these, derivative information is required. Due to Assumption A.1 the objective functions are twice continuously differentiable and for the cheap, analytically given functions derivatives can be computed easily. Therefore, we choose the (second-order) Taylor model for the functions f_i, $i = 2, 3, \ldots, q$, given by

$$\begin{aligned} m_i^k(x) &= m_T(x; f_i, x^k) \\ &= f_i(x^k) + \nabla f_i(x^k)^\top \left(x - x^k\right) + \frac{1}{2} \left(x - x^k\right)^\top \nabla^2 f_i(x^k) \left(x - x^k\right) \end{aligned} \tag{4.6}$$

from Definition 3.2. Since for the expensive function derivative information is not available with reasonable effort, such a model cannot be used. To obtain a quadratic model as well, we use interpolation based on quadratic Lagrangian polynomials. According to Definition 3.4, the model function m_1^k is then defined by

$$m_1^k(x) = m_L(x; f_1, y^1, B_k) = \sum_{i=1}^{p} f_1(y^i) l_i(x) \tag{4.7}$$

with interpolation points $Y = \{y^1, y^2, \ldots, y^p\}$, $y^1 = x^k$ and the Lagrange polynomials $l_i \in \mathcal{P}_n^2$, $i = 1, 2, \ldots, p$, from Definition 3.3. The interpolation points are not chosen randomly from the trust region, but are computed such that they satisfy the quality criterion of poisedness, see Definition 3.5. They are computed by Algorithm 3 presented in Section 3.2.2 which also computes the related Lagrange polynomials.
For the implementation, Algorithm 3 is slightly modified to save function evaluations.

All points x for which $f_1(x)$ has already been computed, are stored in a list. If in the point selection step of Algorithm 3 a new point needs to be computed, a point from this list is used if possible. Only if this is not possible, i.e. the points are not situated in the trust region or violate the criterion of poisedness, a new point is computed.
As described in Section 4.1, the model functions need to be exact at the current iteration point x^k and fulfill $f_i(x^k) = m_i^k(x^k)$ for all $i = 1, 2, \ldots, q$ and for all $k \in \mathbb{N}$, see (4.2). This is required for the theoretical results in Section 4.5. The model functions for the cheap functions are recomputed in every iteration since it results only in low numerical effort. Due to the definition of the Taylor model, see (4.6), it holds $f_i(x^k) = m_i^k(x^k)$ for all $i = 2, 3, \ldots, q$ and for all $k \in \mathbb{N}$.
To ensure $f_1(x^k) = m_1^k(x^k)$ for all $k \in \mathbb{N}$ in the theoretical considerations, the model function m_1^k also needs to be updated in every iteration. Since x^k is always chosen as interpolation point, the equality is then fulfilled. However, for the implementation of MHT this is weakened. To save function evaluations, m_1^k is only updated if necessary. Otherwise, the previous model function is reused. Whether a model update is necessary or not, is decided by the outcome of the trial point acceptance test which is described in detail in Section 4.4. If it holds

$$\rho_\phi^k = \frac{\phi(x^k) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})} < \eta_1,$$

with $\phi(x) = \max_{i=1,\ldots,q} f_i(x)$ and $\phi_m^k(x) = \max_{i=1,\ldots,q} m_i^k(x)$ from (4.4) and (4.5), the trial point is not accepted since the model m^k is not accurate enough. In this case the model for f_1 is also recomputed. Otherwise, the previous model m_1^k is reused in iteration $k+1$. If the model is updated, former interpolation points are reused if possible as already described.
If none of the already evaluated points can be reused, $(n+1)(n+2)/2$ new interpolation points – including the current iteration point which is always contained in Y – are required. This can be problematic for higher dimensions. Thus, we suggest to use linear models based on linear Lagrange polynomials $l_i \in \mathcal{P}_n^1, i = 1, 2, \ldots, p$, for higher dimensions which require only $n+1$ interpolation points. The convergence proof also holds when using these models, detailed information about the induced changes is given in Section 4.6.1.
Another option for building models in the trust region scheme are radial basis functions (RBFs), see for example [Buh03; WS11].

4.3 Computing the Trial Point

The search for a sufficient decrease in the function values is realized by computing the ideal point $p^k = (p_1^k, p_2^k, \dots, p_q^k)^\top$ defined by $p_i^k = \min_{x \in B_k} m_i^k(x)$ for all $i = 1, 2, \dots, q$, see (4.3). The auxiliary optimization problem (PS) defined by

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

with $r^k = f(x^k) - p^k \in \mathbb{R}^q_+$ realizes moving into this search direction. (PS) is a scalarization approach to solve multi-objective optimization problems in general. However, in the context of this thesis it is not used as a scalarization method as such, but instead it is used as an auxiliary tool to move as far as possible into the search direction. The notation in this thesis follows the formulation of the Pascoletti-Serafini scalarization in [Eic08]. If x^k is not Pareto critical for the optimization problem

$$\min_{x \in \mathbb{R}^n} m^k(x), \qquad (MOP^k_{m,\mathbb{R}^n})$$

then $r_i^k > 0$ holds for all $i \in \{1, 2, \dots, q\}$. This follows directly from Theorem 2.8 (iii), but to illustrate the connections, we outline it shortly. Note that the model functions are assumed to be exact at the current iteration point, see (4.2), i.e. it holds $f(x^k) = m^k(x^k)$ in every iteration $k \in \mathbb{N}$.

Remark 4.1 Let $k \in \mathbb{N}$ be with x^k not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. According to Lemma 2.9, x^k is not locally weakly efficient for $(MOP^k_{m,\mathbb{R}^n})$. Since it holds $x^k \in \text{int} B_k$, x^k is not weakly efficient for $\min_{x \in B_k} m^k(x)$ referred to as (MOP^k_m). This implies $m_i^k(x^k) > \min_{x \in B_k} m_i^k(x) = p_i^k$ for all $i \in \{1, 2, \dots, q\}$. Since it holds $f(x^k) = m^k(x^k)$ in every iteration $k \in \mathbb{N}$, see (4.2), this implies $r_i^k = f_i(x^k) - p_i^k > 0$ for all $i \in \{1, 2, \dots, q\}$.

Remark 4.2 Note that $\mathbb{R}^q_+$ is a closed convex cone with $\text{int}\mathbb{R}^q_+ = \mathbb{R}^q_{++} \neq \emptyset$. If $r_i^k > 0$ holds for all $i \in \{1, 2, \dots, q\}$, i.e. $r^k \in \mathbb{R}^q_{++}$, then the optimization problem (PS) has feasible points and there exists a minimal solution of (PS). This is given in a more general context for example in [Göp+03; Eic08]. Thus, in step 3 of MHT the auxiliary problem (PS) is only considered if a minimal solution exists.

If $r_i^k = 0$ holds for an index $i \in \{1, 2, \ldots, q\}$, it follows $m_i^k(x^k) = p_i^k$. Thus, the current iteration point x^k is weakly efficient for (MOP_m^k). In this case, (PS) is not considered in step 2 of MHT, but $t^{k+} = 0$ and $x^{k+} = x^k$ is set directly instead.
Moreover, the auxiliary problem (PS) is related to a weighted Chebyshev distance.

Remark 4.3 The auxiliary problem (PS) minimizes, in case $r^k \in \mathbb{R}^q_{++}$, the weighted Chebyshev distance between the set $m^k(B_k)$ and the point p^k with weights $w_i = 1/r_i^k$ for $i \in \{1, 2, \ldots, q\}$. To conclude this, let $r^k \in \mathbb{R}^q_{++}$. Then (PS) can be reformulated as follows.

$$
\begin{aligned}
&\min \left\{ t \in \mathbb{R} \mid f(x^k) + t\, r^k - m^k(x) \in \mathbb{R}^q_+, x \in B_k \right\} \\
=& \min \left\{ t \in \mathbb{R} \mid f(x^k) + t\, (f(x^k) - p^k) + p^k - p^k - m^k(x) \geq 0_q, x \in B_k \right\} \\
=& \min \left\{ t \in \mathbb{R} \mid (1+t) r^k \geq m^k(x) - p^k, x \in B_k \right\} \\
=& \min \left\{ t \in \mathbb{R} \mid 1 + t \geq \frac{m_i^k(x) - p_i^k}{r_i^k} \text{ for all } i = 1, 2, \ldots, q, x \in B_k \right\} \\
=& \min \left\{ t \in \mathbb{R} \mid 1 + t \geq \max_{i=1,\ldots,q} \frac{m_i^k(x) - p_i^k}{r_i^k}, x \in B_k \right\} \\
=& \min \left\{ s - 1 \in \mathbb{R} \mid s \geq \max_{i=1,\ldots,q} \frac{m_i^k(x) - p_i^k}{r_i^k}, x \in B_k \right\} \\
=& \min_{x \in B_k} \max_{i=1,\ldots,q} \frac{1}{r_i^k} \left| m_i^k(x) - p_i^k \right| \\
=& \min_{x \in B_k} \max_{i=1,\ldots,q} \left\| m^k(x) - p^k \right\|_{(w,\infty)}
\end{aligned}
$$

Thus, (PS) minimizes the weighted Chebyshev distance between $m^k(B_k)$ and p^k with weights $w_i = 1/r_i^k$ for $i \in \{1, 2, \ldots, q\}$. This is illustrated in Figure 4.3.

The following Lemma gives general information about the minimal solutions of (PS).

Lemma 4.4 *Let $(\bar{t}, \bar{x}) \in \mathbb{R}^{1+n}$ be a minimal solution of (PS). Then it holds $\bar{t} \leq 0$ and $m^k(\bar{x}) \leq m^k(x^k)$.*

Proof. According to (4.2) it holds $f(x^k) = m^k(x^k)$ for all $k \in \mathbb{N}$. This implies

$$
f(x^k) + 0 \cdot r^k - m^k(x^k) = m^k(x^k) + 0 \cdot r^k - m^k(x^k) = 0_q
$$

and therefore $(0, x^k)$ is feasible for (PS) and it follows $\bar{t} \leq 0$. Since it holds $r^k \geq 0$ in all iterations $k \in \mathbb{N}$, it follows

$$m^k(x^k) - m^k(\bar{x}) \geq -\bar{t}\, r^k \geq 0_q.$$

□

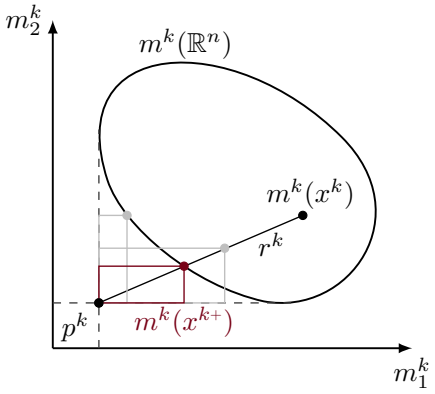

Figure 4.3 – Connection between (PS) and weighted Chebyshev distance

If the current iteration point x^k is not weakly efficient for (MOP_m^k), the statement from Lemma 4.4 can be specified.

Lemma 4.5 *Let x^k be not weakly efficient for (MOP_m^k). For every minimal solution $(\bar{t}, \bar{x}) \in \mathbb{R}^{1+n}$ of (PS) it holds $\bar{t} \in [-1, 0)$.*

Proof. Let $(\bar{t}, \bar{x}) \in \mathbb{R}^{1+n}$ be a minimal solution of (PS). According to Lemma 4.4, it holds $\bar{t} \leq 0$. Due to x^k being not weakly efficient for $\min_{x \in B_k} m^k(x)$ there exists a point $\tilde{x} \in B_k$ with $m_i^k(\tilde{x}) < m_i^k(x^k)$ for all $i \in \{1, 2, \ldots, q\}$. Since it holds $f(x^k) = m^k(x^k)$ due to (4.2), this implies

$$r_i^k = f_i(x^k) - p_i^k = m_i^k(x^k) - \min_{x \in B_k} m_i^k(x) > 0_q$$

for all $i \in \{1, 2, \ldots, q\}$. Consequently, there exists a scalar $t < 0$ with

$$m^k(x^k) + t\, r^k - m^k(\tilde{x}) \geq 0_q,$$

i.e. $(t, \tilde{x})$ is feasible for (PS) and it holds $\bar{t} \leq t < 0$. Now suppose $\bar{t} := -1 - s < -1$ with $s > 0$. Due to the constraints of (PS) it holds

$$0_q \leq m^k(x^k) + (-1 - s)\left(m^k(x^k) - p^k\right) - m^k(\bar{x}) = p^k - s\, r^k - m^k(\bar{x})$$

and it follows $p^k - m^k(\bar{x}) \geq s\, r^k$. Again, due to x^k being not weakly efficient and thus $r^k > 0_q$, it follows $p^k > m^k(\bar{x})$ which contradicts the definition of p^k. Consequently, it holds $\bar{t} \in [-1, 0)$. □

The optimization problem (PS) has some useful properties which can be found in a more general framework and with proof in [Eic08, Th. 2.1]

Lemma 4.6

(i) If $(\bar{t}, \bar{x}) \in \mathbb{R}^{1+n}$ is a minimal solution of (PS), then $\bar{x}$ is weakly efficient for (MOP_m^k).

(ii) If $(\bar{t}, \bar{x}) \in \mathbb{R}^{1+n}$ is a local minimal solution of (PS), then $\bar{x}$ is locally weakly efficient for (MOP_m^k).

(iii) If $x^k \in \mathbb{R}^n$ is weakly efficient for (MOP_m^k) and $r^k \in \mathbb{R}^q_{++}$, then $(0, x^k)$ is a minimal solution of (PS).

(iv) If $x^k \in \mathbb{R}^n$ is efficient for (MOP_m^k) and $r^k \neq 0_q$, then $(0, x^k)$ is a minimal solution of (PS).

(v) If $x^k \in \mathbb{R}^n$ is locally weakly efficient for (MOP_m^k) and $r^k \in \mathbb{R}^q_{++}$, then $(0, x^k)$ is a local minimal solution of (PS).

(vi) If $x^k \in \mathbb{R}^n$ is locally efficient for (MOP_m^k) and $r^k \neq 0_q$, then $(0, x^k)$ is a local minimal solution of (PS).

4.4 Trial Point Acceptance Test

Step 3 of MHT is the trial point acceptance test in which it is decided if x^{k+} is accepted as next iteration point. For this purpose, the function values $m_i^k(x^{k+})$, $i = 1, 2, \ldots, q$, of the model functions are compared to the function values $f_i(x^{k+})$, $i = 1, 2, \ldots, q$, of the original functions, i.e. the prediction of the model functions is compared to the

actual behavior of the original functions.
In the single-objective trust region approach this is realized by considering the quotient $(g(x^k) - g(x^{k+}))/(m_g(x^k) - m_g(x^{k+}))$ for $g : \mathbb{R}^n \to \mathbb{R}$ being a scalar valued function and $m_g : \mathbb{R}^n \to \mathbb{R}$ its model function. If this quotient is bigger than a given nonnegative constant, the trial point is accepted, see also (3.2) in Section 3.1. This criterion can be transferred to multi-objective trust region approaches by applying it to every function individually, that is considering the quotients

$$\rho_i^k = \frac{f_i(x^k) - f_i(x^{k+})}{m_i^k(x^k) - m_i^k(x^{k+})} \text{ for } i = 1, 2, \ldots, q. \tag{4.8}$$

If all quotients are bigger than a given nonnegative constant, the trial point is accepted. This trial point acceptance test is used for example in the trust region approach from [RK14]. We use a more general approach which can also be found in [VOS14] and consider the quotient

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

with the functions $\phi(x) = \max_{i=1,\ldots,q} f_i(x)$ and $\phi_m^k(x) = \max_{i=1,\ldots,q} m_i^k(x)$ defined in (4.4) and (4.5). The trial point x^{k+} is accepted if $\rho_\phi^k \geq \eta_1$ holds with $\eta_1 > 0$ from MHT. As described in the following, this guarantees only a descent for at least one objective function. Using the acceptance test with ρ_i^k, $i = 1, 2, \ldots, q$, guarantees a descent for every objective function. Thus, the latter acceptance criterion is stricter which is also explained in Section 4.6.3.
Numerical results that illustrate the difference of the acceptance tests are presented in Section 6.4.4. The difference is schematically illustrated in Figure 4.4. Two areas are depicted: the gray shaded area includes the images of all points that would be accepted by the strict version of the acceptance test, as for example $\hat{x}$. This area is a subset of the area contoured by dashed lines. This bigger region contains the images of those points that would be accepted by the trial point acceptance test defined by ρ_ϕ^k.

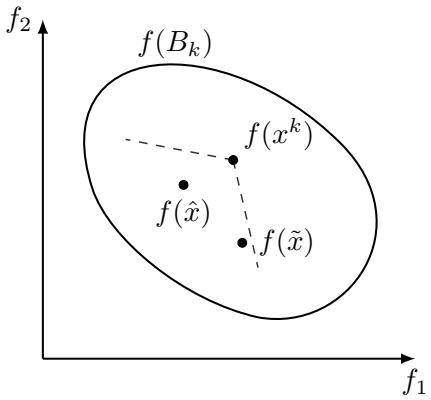

Figure 4.4 – Trial point acceptance test with ρ_ϕ^k and ρ_i^k, $i = 1, 2, \ldots, q$

4.4.1 General Properties

Due to the determining of x^{k+} by (PS) it holds $m^k(x^{k+}) \leq m^k(x^k)$ in all iterations $k \in \mathbb{N}$, see Lemma 4.4. Therefore, it holds

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq 0 \tag{4.9}$$

for all $k \in \mathbb{N}$. Thus, if $\rho_\phi^k < 0$ holds, it follows $\phi(x^k) - \phi(x^{k+}) < 0$. If x^k is not weakly efficient for (MOP_m^k), it holds according to Lemma 4.5 $t^{k+} < 0$. Moreover, it holds $r^k \in \mathbb{R}_{++}^q$, see also the line of argument in Remark 4.1. Then the constraint of (PS) implies $m^k(x^k) - m^k(x^{k+}) \geq -t^{k+}\, r^k > 0$ and it follows

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) > 0.$$

4.4.2 Acceptance of Trial Point

If $\rho_\phi^k \geq \eta_1 > 0$ holds, the trial point x^{k+} is accepted as next iteration point. Together with (4.9) this implies $\phi(x^k) - \phi(x^{k+}) > 0$. Let

$$\phi(x^k) = \max_{i=1,\ldots,q} f_i(x^k) = f_{i_1}(x^k) \tag{4.10}$$

$$\phi(x^{k+}) = \max_{i=1,\ldots,q} f_i(x^{k+}) = f_{i_2}(x^{k+}) \tag{4.11}$$

be with $i_1, i_2 \in \{1, 2, \ldots, q\}$. Thus, it holds $\phi(x^{k+}) = f_{i_2}(x^{k+}) \geq f_{i_1}(x^{k+})$ and it follows

$$0 < \phi(x^k) - \phi(x^{k+}) = f_{i_1}(x^k) - f_{i_2}(x^{k+}) \leq f_{i_1}(x^k) - f_{i_1}(x^{k+}).$$

Therefore, it holds $f_{i_1}(x^{k+}) < f_{i_1}(x^k)$ and the trial point x^{k+} guarantees a descent for at least one objective function of (MOP). Of course this includes the case that x^{k+} generates a descent for all objective functions, see also the schematical illustration in Figure 4.4.
The trial point is accepted if ρ_ϕ^k is bigger than a strictly positive constant η_1. Choosing $\eta_1 = 0$ means that any decrease is accepted, even if it is only marginal. This also applies to the single-objective trust region approach and to the stricter version of the trial point acceptance test. Considering the trust region approach in general, the constant η_1 is often chosen equal to zero for theoretical considerations, e.g. in the multi-objective trust region approach presented in [RK14]. Choosing η_1 strictly positive defines which decrease is interpreted as sufficient to accept the trial point.

4.4.3 Discarding the Trial Point

If $\rho_\phi^k < \eta_1$ holds, the trial point x^{k+} is not accepted as next iteration point. The trust region radius is reduced and the model functions are updated for the next iteration. In the following we describe the cases that can occur. If $0 < \rho_\phi^k < \eta_1$ holds, the trial point provides a decrease for at least one objective function, see Section 4.4.2, but this decrease is interpreted as not sufficient.
Now let $\rho_\phi^k < 0$ hold. Note that $\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq 0$ always holds, see (4.9). Due to $\phi(x^k) = \max_{i=1,\ldots,q} f_i(x^k) = f_{i_1}(x^k) \geq f_{i_2}(x^k)$, see (4.10) and (4.11), it follows

$$0 > \phi(x^k) - \phi(x^{k+}) = f_{i_1}(x^k) - f_{i_2}(x^{k+}) \geq f_{i_2}(x^k) - f_{i_2}(x^{k+}).$$

This implies $f_{i_2}(x^{k+}) > f_{i_2}(x^k)$ and therefore an increase for at least one objective function of (MOP).
Now assume $\rho_\phi^k = 0$. According to step 3 of MHT, this implies

$$t^{k+} = 0, \tag{4.12}$$

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) = 0 \text{ or} \tag{4.13}$$
$$\phi(x^k) - \phi(x^{k+}) = 0. \tag{4.14}$$

If (4.14) holds, it follows according to (4.10) and (4.11)

$$\phi(x^{k+}) = f_{i_2}(x^{k+}) = f_{i_1}(x^k) \geq f_{i_2}(x^k).$$

If this inequality is strict, the trial point provides an increase for at least one of the objective functions.
If (4.12) is fulfilled, then either it holds $r^k = 0_q$ or $(0, x^k)$ is a minimal solution of (PS). In both cases, x^k is weakly efficient for (MOP_m^k), in the first case due to the definition of r^k and in the second case due to Lemma 4.6. By setting $\rho_\phi^k = 0$, the trust region radius is reduced for the next iteration and the model functions are updated.
If x^k is also weakly efficient for $\min_{x \in B_k} f(x)$, then the trial point acceptance test in the subsequent iterations will have the same result as in iteration k. Thus, the trust region radius will converge to zero.
If x^k is not weakly efficient for $\min_{x \in B_k} f(x)$, at least one of the model functions m_i^k, $i \in \{1, 2, \ldots, q\}$, was not reliable in iteration k. The accuracy of the model functions is assumed to be connected to the size of the trust region, see Definition 3.1 in Section 3.2 and Lemma 4.8 and Assumption A.8 in Section 4.5. The smaller the trust region, the more accurate are the model functions supposed to be. Thus, by reducing the size of the trust region, the model functions are assumed to be more accurate in the next iteration. Then there will be a subsequent iteration in which the model functions are accurate enough to obtain a trial point providing a sufficient decrease.
Now let (4.13) hold. Reducing the trust region radius and updating the model function in this case is motivated by Assumption A.11 from Section 4.5. It is explained in detail there and a lemma justifying it is proved. Since it clarifies the trial point acceptance test, we anticipate it here. This assumption ensures

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

with $\omega(x)$ from Definition 2.10, $\kappa_\phi \in (0, 1)$ and $\beta_\phi^k > 0$. According to Lemma 2.11, it holds $\omega(x) = 0$ if and only if the point x is Pareto critical for (MOP) and according to

Lemma 2.9 Pareto criticality is a necessary condition for local weak efficiency. Therefore, the difference $\phi_m^k(x^k) - \phi_m^k(x^{k+})$ has a strictly positive lower bound as long as the point x^k is not Pareto critical and therefore not locally weakly efficient for (MOP). If $\phi_m^k(x^k) - \phi_m^k(x^{k+}) = 0$ holds, the lower bound of Assumption A.11 is only satisfied if $\omega(x^k) = 0$, that is x^k is Pareto critical for (MOP). Thus, in this case it is reasonable to reduce the trust region radius and update the model function to confirm the behavior of the model function.

4.5 Convergence Analysis of MHT

In this section we present a convergence analysis for the basic formulation of MHT as given in Section 4.1. The results can be transferred for several modifications of MHT. This is described in Section 4.6.
We prove that the sequence of iterates generated by MHT converges to a Pareto critical point of (MOP), i.e. it fulfills a necessary condition for local weak efficiency. For the theoretical results, assumptions on the original and the model functions are necessary. All these assumptions are connected to the commonly used assumptions in single-objective trust region approaches [CGT00; CSV09b; The11], see also Section 3.1, and in multi-objective trust region approaches [RK14; VOS14; CLM16]. They are introduced in this section when necessary and listed for an overview in Appendix A.1.
As already mentioned in Section 4.1, see Assumption A.1, all objective functions are assumed to be twice continuously differentiable. Note that, as explained there, this is a strong assumption for expensive black-box functions. Nevertheless, it is commonly used to prove first order convergence of methods for black-box functions, see e.g. [AH17].

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

For the convergence analysis of the single-objective trust region approach the objective function is assumed to be bounded from below, see also Section 3.1. This would

be transferred directly to the multi-objective case by assuming that all objective functions $f_i, i = 1, 2, \ldots, q$, are bounded from below. This is for example used in the multi-objective trust region approaches in [VOS14; CLM16]. In our case, the following assumption is sufficient.

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.

Since in the iterations of MHT the model functions are considered instead of the original functions, the surrogate optimization problem

$$\min_{x \in \mathbb{R}^n} m^k(x) \qquad (MOP^k_{m,\mathbb{R}^n})$$

also needs to be considered in the convergence analysis. The assumptions necessary for the model functions are analogous to the assumptions in single- and the multi-objective trust region versions. Although we have specified the choice of the model functions in Section 4.2 to the quadratic interpolation model based on Lagrange polynomials for the expensive function f_1 and the second order Taylor model for the cheap functions $f_2, f_3, \ldots, f_q$, we formulate these assumptions in a general way to clarify the connection to other trust region approaches.

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

This assumption is obviously fulfilled for the quadratic interpolation model m_1^k and the Taylor models $m_2^k, m_3^k, \ldots, m_q^k$. In trust region methods, the more general assumption of twice continuously differentiable model functions is also often used, see for example [CGT00].

Assumption A.4 The model functions m_i^k are exact at the current iteration point, that is it holds

$$m_i^k(x^k) = f_i(x^k) \text{ for all } i \in \{1, 2, \ldots, q\} \text{ and } k \in \mathbb{N}.$$

This holds true for every interpolation model which uses x^k as interpolation point and also for the Taylor model. Thus, Assumption A.4 is fulfilled for all model functions used in MHT. Since for the cheap functions derivative information is available, the

derivatives are supposed to fulfill the following condition. In the context of this thesis, it is fulfilled since the model functions $m_2^k, m_3^k, \ldots, m_q^k$ are chosen as the Taylor model.

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

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

Such an assumption is used for all objective functions in the multi-objective trust region approach in [CLM16]. In the basic single-objective version of the trust region approach different types of assumptions for the gradient are used, the above is one of them, see for example [CGT00]. In the following, when we introduce further assumptions connected to the gradient in the context of this thesis, we will also outline the connection to other commonly used assumptions in the basic trust region approach. Note that we consider the Frobenius norm $\|\cdot\|_F$ as matrix norm since it is compatible with the Euclidean norm which is chosen as vector norm.

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 $\kappa_{\text{uhf}_i} > 1$ fulfilling

$$\left\|\nabla^2 f_i(x)\right\|_F \leq \kappa_{\text{uhf}_i} - 1$$

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

Remark 4.7 From Assumptions A.1 and A.6 it follows that the function ω from Definition 2.10 is uniformly continuous, see also [VOS14]. To clarify this connection, let $i \in \{1, 2, \ldots, q\}$ be an arbitrary but fixed index. Since the objective functions are twice continuously differentiable and the Euclidean norm and the Frobenius norm are compatible, it follows from the mean value theorem applied to ∇f_i and Assumption A.6

$$\|\nabla f_i(x) - \nabla f_i(y)\| \leq \left(\sup_{\xi \in [x,y]} \left\|\nabla^2 f_i(\xi)\right\|_F\right) \|x - y\| \leq (\kappa_{\text{uhf}_i} - 1) \|x - y\|$$

for $x, y \in \mathbb{R}^n$ with $x \leq y$ without loss of generality. Hence, $\nabla f_i : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with the Lipschitz constant $L_i := \kappa_{\text{uhf}_i} - 1 > 0$ for all $i \in \{1, 2, \ldots, q\}$.

Supposed ω is not uniformly continuous, then there exists a scalar $\varepsilon > 0$ such that for all $\delta > 0$ there exist $x, y \in \mathbb{R}^n$ such that it holds $\|x - y\| < \delta$ and $|\omega(x) - \omega(y)| \geq \varepsilon$. Moreover, let

$$\omega(x) = -\min_{\|d\| \leq 1} \max_{i=1,\dots,q} \nabla f_i(x)^\top d = -\max_{i=1,\dots,q} \nabla f_i(x)^\top \bar{d} \quad \text{and}$$
$$\max_{i=1,\dots,q} \nabla f_i(y)^\top \bar{d} = \nabla f_j(y)^\top \bar{d}$$

be with $\bar{d} \in \mathbb{R}^n$, $\left\|\bar{d}\right\| \leq 1$ and $j \in \{1, 2, \dots, q\}$. We choose $\delta = \varepsilon / L_j$. Without loss of generality let $|\omega(x) - \omega(y)| = \omega(x) - \omega(y)$ be. Then it follows with the inequality of Cauchy-Schwarz and the Lipschitz continuity of ∇f_j

$$\begin{aligned}
\varepsilon &\leq |\omega(x) - \omega(y)| = \omega(x) - \omega(y) \\
&= -\max_{i=1,\dots,q} \nabla f_i(x)^\top \bar{d} + \min_{\|d\| \leq 1} \max_{i=1,\dots,q} \nabla f_i(y)^\top d \\
&\leq -\max_{i=1,\dots,q} \nabla f_i(x)^\top \bar{d} + \nabla f_j(y)^\top \bar{d} \\
&\leq -\nabla f_j(x)^\top \bar{d} + \nabla f_j(y)^\top \bar{d} \\
&\leq \|-\nabla f_j(x) + \nabla f_j(y)\| \left\|\bar{d}\right\| \leq L_j \|x - y\| < L_j \delta = \varepsilon
\end{aligned}$$

which is a contradiction. Consequently, ω is uniformly continuous.

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

$$\left\|\nabla^2 m_i^k(x)\right\|_F \leq \kappa_{\text{uhm}_i} - 1$$

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

The assumptions on the Hessians of the model functions and the original functions are analogous to the single-objective trust region approach [CGT00; CSV09b] and are also used in multi-objective trust region approaches [RK14; VOS14; CLM16].
Furthermore, as in every model-based solution method it is important to ensure a good local accuracy of the model functions in every iteration. For this purpose, we use the common notion of validity from Definition 3.1. In the trust region approach in general,

validity is assumed for the model functions. In our context we can even prove this for the models of the cheap functions.

Lemma 4.8 *Suppose Assumptions* A.1 *and* A.3 *to* A.7 *hold. In every iteration $k \in \mathbb{N}$ the model m_i^k is valid for f_i in B_k for all $i \in \{2, 3, \ldots, q\}$, that is it holds*

$$\left| f_i(x) - m_i^k(x) \right| \leq \kappa_{\mathrm{em}_i} \delta_k^2$$

for all $x \in B_k$ with $\kappa_{\mathrm{em}_i} := \max\left\{\kappa_{\mathrm{uhf}_i}, \kappa_{\mathrm{uhm}_i}\right\} - 1 > 0$. The index '$\mathrm{em}_i$' stands for error of the model function m_i^k.

Proof. According to Assumption A.1, the functions $f_i, i = 2, 3, \ldots, q$, are twice continuously differentiable. It follows from Taylor's theorem for every $h \in \mathbb{R}^n$ with $\|h\| \leq \delta_k$,

$$f_i(x^k + h) = f_i(x^k) + \nabla f_i(x^k)^\top h + \frac{1}{2} h^\top \nabla^2 f_i(\xi_i^k) h$$

with $\xi_{i_j}^k \in \left[x_j^k, x_j^k + h\right]$ for $j \in \{1, 2, \ldots, n\}$ and for $i \in \{2, 3, \ldots, q\}$. Since according to Assumption A.3 the model functions m_i^k, $i = 2, 3, \ldots, q$, are quadratic functions, it holds

$$m_i^k(x^k + h) = m_i^k(x^k) + \nabla m_i^k(x^k)^\top h + \frac{1}{2} h^\top \nabla^2 m_i^k(x^k) h$$

for every $h \in \mathbb{R}^n$ with $\|h\| \leq \delta_k$ and for all $i \in \{2, 3, \ldots, q\}$. Moreover, it holds due to Assumptions A.4 and A.5 $m_i^k(x^k) = f_i(x^k)$ and $\nabla m_i^k(x^k) = \nabla f_i(x^k)$ for all $i \in \{2, 3, \ldots, q\}$. Every vector $x \in B_k$ is given by $x = x^k + h$ with a suitable vector h with $\|h\| \leq \delta_k$. Using the triangle inequality it follows for all $x \in B_k$ and for all $i \in \{2, 3, \ldots, q\}$

$$\begin{aligned}
\left| f_i(x) - m_i^k(x) \right| &= \left| \frac{1}{2} h^\top \left(\nabla^2 f_i(\xi_i^k) - \nabla^2 m_i^k(x^k) \right) h \right| \\
&\leq \frac{1}{2} \|h\|^2 \left(\left\| \nabla^2 f_i(\xi_i^k) \right\| + \left\| \nabla^2 m_i^k(x^k) \right\| \right) \\
&\leq \delta_k^2 \left(\max\left\{\kappa_{\mathrm{uhf}_i}, \kappa_{\mathrm{uhm}_i}\right\} - 1 \right)
\end{aligned}$$

with the constants κ_{uhf_i} and κ_{uhm_i} from Assumptions A.6 and A.7. Then the statement of the lemma follows by defining $\kappa_{\mathrm{em}_i} := \max\left\{\kappa_{\mathrm{uhf}_i}, \kappa_{\mathrm{uhm}_i}\right\} - 1 > 0$ for all $i \in \{2, 3, \ldots, q\}$. □

For the expensive function such a result is not provable, thus and like in the standard trust region approach we assume validity.

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 $\kappa_{\text{em}_1} > 0$ independent of k such that it holds for all $x \in B_k$

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

The accuracy of the model is also reflected in the gradients. For the cheap functions $m_i^k, i \in \{2, 3, \ldots, q\}$, we impose Assumption A.5 analogous to the single-objective approach which states $\nabla m_i^k(x^k) = \nabla f_i(x^k)$ for all iterations $k \in \mathbb{N}$ and is always fulfilled since the Taylor model is considered.
For the expensive function f_1 the following Lemma holds regarding the gradient. Such a statement is also proved in standard trust region approaches and can be found for example in [CGT00]. Due to the problem-dependent constants we give a short proof. If the gradient of a model function fulfills the inequality of this lemma and is valid according to Definition 3.1, then the model is also referred to as fully linear in the literature, see for example [CSV09b; AH17]. These conditions are also used in the multi-objective trust region approach presented in [RK14].

Lemma 4.9 *Suppose Assumptions* A.1, A.3, A.4 *and* A.6 *to* A.8 *hold. Then there exists a constant* $\kappa_{\text{eg}} > 0$ *such that it holds*

$$\left\|\nabla f_1(x^k) - \nabla m_1^k(x^k)\right\| \leq \kappa_{\text{eg}} \delta_k.$$

for all $k \in \mathbb{N}$. *The index 'eg' stands for error of gradient.*

Proof. Analogous to the proof of Lemma 4.8 and similar to [CGT00, Th. 9.1.1] it follows by using Taylor's theorem, Assumptions A.1, A.3, A.4 and A.6 to A.8 and the triangle inequality

$$\begin{aligned}\left|\left(\nabla f_1(x^k) - \nabla m_1^k(x^k)\right)^\top h\right| &\leq \left|f_1(x) - m_1^k(x)\right| + \frac{1}{2}\|h\|^2 \left\|\nabla^2 f_1(\xi^k) - \nabla^2 m_1^k(x^k)\right\| \\ &\leq \kappa_{\text{em}_1}\delta_k^2 + \left(\max\left\{\kappa_{\text{uhf}_1}, \kappa_{\text{uhm}_1}\right\} - 1\right)\delta_k^2\end{aligned}$$

for every $h \in \mathbb{R}^n$ with $\|h\| \leq \delta_k$, $x \coloneqq x^k + h \in B_k$ and a suitable intermediate vector $\xi^k \in \left[x^k, x^k + h\right] \coloneqq \left\{x^k + t\, h \mid t \in [0,1]\right\}$. Defining

$$h \coloneqq \delta_k \frac{\nabla f_1(x^k) - \nabla m_1^k(x^k)}{\left\|\nabla f_1(x^k) - \nabla m_1^k(x^k)\right\|}$$

the statement of the lemma follows with the constant $\kappa_{\text{eg}} \coloneqq \kappa_{\text{em}_1} + \max\left\{\kappa_{\text{uhf}_1}, \kappa_{\text{uhm}_1}\right\} - 1 > 0$. □

This lemma guarantees that whenever the trust region radius is small enough, the gradient $\nabla m_1^k(x^k)$ of the model is a good approximation for the gradient $\nabla f_1(x^k)$ of the original function. In addition to this result, the approximation of the gradient of the expensive function in the current iteration point x^k shall be accurate enough to ensure reliability whenever Pareto critical points are approached. Such points are characterized by the function $\omega(x) = -\min_{\|d\| \leq 1} \max_{i=1,\dots,q} \nabla f_i(x)^\top d$ from Definition 2.10. Analogously we define

$$\omega_m(x) \coloneqq -\min_{\|d\| \leq 1} \max_{i=1,\dots,q} \nabla m_i^k(x)^\top d \tag{4.15}$$

for the model functions. The function ω_m is an auxiliary function used to deduce a sufficient decrease condition in terms of ω necessary for the convergence analysis. Such a condition is a main component of the convergence analysis of every trust region approach, both in single- and multi-objective versions, see also Section 3.1. As stated in Lemma 2.11, the function ω characterizes Pareto critical points of multi-objective optimization problems. Obviously, this result also holds for the function ω_m using the model functions m_i^k, $i \in \{1, 2, \dots, q\}$, and the corresponding optimization problem $(MOP^k_{m,\mathbb{R}^n})$.

Assumption A.9 gives a connection of Pareto critical points of $(MOP^k_{m,\mathbb{R}^n})$ with the model functions and (MOP) with the original functions. Pareto criticality can be considered as a generalization of the notion of stationarity for single-objective problems to multi-objective problems, see also Section 2.2. Thus, Assumption A.9 is analogous to another assumption for the gradients of the model functions used in single-objective trust region approaches. Besides the equality of the gradients at the current iteration

point, see Assumption A.5, or an upper bound for the difference of the gradient of the model and the original function, see Lemma 4.9, an assumption of the form

$$\left\|\nabla g(x^k) - \nabla m_g(x^k)\right\| \leq \kappa \left\|\nabla m_g(x^k)\right\|$$

is also used, see for example [CGT00], where $g : \mathbb{R}^n \to \mathbb{R}$ is a scalar-valued function, $m_g : \mathbb{R}^n \to \mathbb{R}$ its model function and $\kappa \geq 0$ a suitable constant. Instead of using such an assumption for every objective function $f_i, i = 1, 2, \ldots, q$, we assume the following multi-objective condition using the auxiliary functions ω and ω_m.

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

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

This assumption ensures that whenever the iteration point x^k is Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$ or close to such a point, this is also satisfied for the original optimization problem (MOP). This property and a connection to (PS) is stated in the following Lemma. It is connected to Lemmas 4.4 and 4.5 from Section 4.3.

Lemma 4.10 *Suppose Assumptions* A.1*,* A.3*,* A.4 *and* A.9 *holds. Let $k \in \mathbb{N}$ be an iteration index of MHT and x^k not Pareto critical for (MOP). Then the following statements hold.*

(i) x^k is not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$.

(ii) For the minimal solution $(t^{k+}, x^{k+}) \in \mathbb{R}^{1+n}$ of (PS) computed in iteration k of MHT it holds $t^{k+} < 0$.

Proof. To prove item (i) assume that x^k is Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. According to Lemma 2.11, this is equivalent to $\omega_m(x^k) = 0$. Since x^k is not Pareto critical for (MOP), it holds $\omega(x^k) > 0$, also according to Lemma 2.11. Then Assumption A.9 implies

$$0 < \omega(x^k) = \left|\omega_m(x^k) - \omega(x^k)\right| \leq \kappa_\omega \, \omega_m(x^k) = 0$$

which is a contradiction. Thus, the initial assumption is false and x^k is not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$.
To prove statement (ii), we use the finding from (i) that x^k is not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. This implies $r^k_i > 0$ for all $i \in \{1, 2, \ldots, q\}$ and therefore, there exists

a minimal solution of (PS), see Remarks 4.1 and 4.2 in Section 4.3. Moreover, it follows from Lemma 2.9 that x^k is not locally weakly efficient for $(MOP^k_{m,\mathbb{R}^n})$, that is there exists a vector $x \in U(x^k) \subseteq B_k$ with $U(x^k)$ a suitable neighborhood of x^k and $m^k_i(x) < m^k_i(x^k)$ for all $i \in \{1, 2, \ldots, q\}$.
Since it holds $r^k > 0_q$, there exists a scalar $t < 0$ such that it holds $m^k(x^k) + t\, r^k - m^k(x) \geq 0$. According to Assumption A.4, it holds $f(x^k) = m^k(x^k)$ and therefore (t, x) is feasible for (PS). This implies $t^{k+} < 0$ for the minimal solution (t^{k+}, x^{k+}) of (PS) computed in iteration k. □

The convergence proof in this section is based on the characterization of Pareto critical points by the function ω given in Definition 2.10. It will be proved that MHT produces a sequence of points with ω converging to zero. For this purpose, a sufficient decrease condition for the function $\phi^k_m : \mathbb{R}^n \to \mathbb{R}$ defined in (4.4) by

$$\phi^k_m(x) = \max_{i=1,\ldots,q} m^k_i(x)$$

is necessary. Such a sufficient decrease condition is commonly used in trust region approaches, both in single- and multi-objective versions [CGT00; CSV09b; RK14; VOS14]. It is based on the idea of minimizing along a descent direction, either for the individual functions or in the multi-objective way. As stated in Lemma 2.12, the minimal solution of the optimization problem in ω provides a descent direction.
In the single-objective approach [CGT00; CSV09b] a backtracking strategy is used to obtain the trial point x^{k+}. Instead of minimizing the function along the steepest descent direction exactly, the Armijo linesearch is used to approximate it. An analogous strategy, but transferred to the multi-objective case by using the function ω, is used in [VOS14]. In [RK14] the objectives are considered individually in addition to a scalarization and therefore several trial points are computed. They are compared to the results of minimizing along the steepest descent directions of the individual functions. Each trial point is assumed to provide a sufficient decrease for the corresponding function compared to this point.
The method presented in this thesis does not use derivative information for the expensive function and also does not consider the functions individually or a scalarized problem as a surrogate, but computes a direction for decreasing the function values in the image space by the ideal point. Therefore, the reasoning for a sufficient decrease

condition differs from the literature. Still, we can use the strategy of comparing the trial point to the result of minimizing along a multi-objective descent direction. For this purpose an assumption regarding the auxiliary problem (PS) given by

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

is necessary which is prepared and justified by the following lemma.

Lemma 4.11 *Suppose Assumptions* A.3 *and* A.7 *hold. Let* $r^k = m^k(x^k) - p^k$ *be the search direction of* (PS) *defined by the ideal points* $p_i^k = \min_{x \in B_k} m_i^k(x)$ *for* $i = 1, 2, \ldots, q$. *In every iteration* $k \in \mathbb{N}$ *with* x^k *being not Pareto critical for* $(MOP^k_{m,\mathbb{R}^n})$ *it holds for every* $i \in \{1, 2, \ldots, q\}$

$$\frac{1}{2} \left\| \nabla m_i^k(x^k) \right\| \min \left\{ \frac{\left\| \nabla m_i^k(x^k) \right\|}{\beta_i^k}, \delta_k \right\} < r_i^k \leq \delta_k \left\| \nabla m_i^k(x^k) \right\| + \frac{1}{2} \delta_k^2 \left(\kappa_{\mathrm{uhm}_i} - 1 \right)$$

with $\beta_i^k \coloneqq 1 + \left\| \nabla^2 m_i^k(x^k) \right\|_F$ *and* $\kappa_{\mathrm{uhm}_i} > 1$ *from Assumption* A.7.

Proof. Let $k \in \mathbb{N}$ be an iteration with x^k not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. It follows from Theorem 2.8 that x^k is not minimal for m_i^k for all $i \in \{1, 2, \ldots, q\}$. Thus, it holds $\nabla m_i^k(x^k) \neq 0_n$ for all $i \in \{1, 2, \ldots, q\}$. Now let $i \in \{1, 2, \ldots, q\}$ be arbitrary, but fixed. Consider the normed steepest descent direction for m_i^k at x^k defined by $d_{sdi} \coloneqq -\nabla m_i^k(x^k) / \left\| \nabla m_i^k(x^k) \right\|$. From Taylor's theorem, Assumption A.3 and the Cauchy Schwarz inequality it follows

$$\begin{aligned}
r_i^k &= m_i^k(x^k) - \min_{x \in B_k} m_i^k(x) \geq m_i^k(x^k) - \min_{|t| \leq \delta_k} m_i^k(x^k + t\, d_{sdi}) \\
&= m_i^k(x^k) - \min_{|t| \leq \delta_k} \left(m_i^k(x^k) + t \nabla m_i^k(x^k)^\top d_{sdi} + \frac{1}{2} t^2 d_{sdi}^\top \nabla^2 m_i^k(x^k) d_{sdi} \right) \\
&= \max_{|t| \leq \delta_k} \left(-t \nabla m_i^k(x^k)^\top d_{sdi} - \frac{1}{2} t^2 d_{sdi}^\top \nabla^2 m_i^k(x^k) d_{sdi} \right) \\
&\geq \max_{|t| \leq \delta_k} \left(t \nabla m_i^k(x^k)^\top \frac{\nabla m_i^k(x^k)}{\left\| \nabla m_i^k(x^k) \right\|} - \frac{1}{2} t^2 \left\| d_{sdi} \right\|^2 \left\| \nabla^2 m_i^k(x^k) \right\|_F \right) \\
&> \max_{|t| \leq \delta_k} \left(t \left\| \nabla m_i^k(x^k) \right\| - \frac{1}{2} t^2 \beta_i^k \right)
\end{aligned}$$

with $\beta_i^k = 1 + \left\|\nabla^2 m_i^k(x^k)\right\|_F$. The possible candidates for the solution of the above maximization problem are $t_1 = \left\|\nabla m_i^k(x^k)\right\| / \beta_i^k$ and $t_2 = \delta_k$ if $t_1 > \delta_k$. If the maximum is obtained for t_1, it follows

$$r_i^k > \frac{\left\|\nabla m_i^k(x^k)\right\|^2}{\beta_i^k} - \frac{1}{2}\frac{\left\|\nabla m_i^k(x^k)\right\|^2}{\beta_i^k} = \frac{1}{2}\frac{\left\|\nabla m_i^k(x^k)\right\|^2}{\beta_i^k}. \tag{4.16}$$

If the maximum is obtained for t_2, it follows due to $\delta_k < t_1 = \left\|\nabla m_i^k(x^k)\right\| / \beta_i^k$

$$\begin{aligned} r_i^k &> \delta_k \left\|\nabla m_i^k(x^k)\right\| - \frac{1}{2}\delta_k^2 \beta_i^k \\ &> \delta_k \left\|\nabla m_i^k(x^k)\right\| - \frac{1}{2}\delta_k \frac{\left\|\nabla m_i^k(x^k)\right\|}{\beta_i^k} \beta_i^k \\ &= \frac{1}{2}\delta_k \left\|\nabla m_i^k(x^k)\right\|. \end{aligned} \tag{4.17}$$

Thus, it holds

$$\begin{aligned} r_i^k &> \min\left\{\frac{1}{2}\frac{\left\|\nabla m_i^k(x^k)\right\|^2}{\beta_i^k}, \frac{1}{2}\left\|\nabla m_i^k(x^k)\right\| \delta_k\right\} \\ &= \frac{1}{2}\left\|\nabla m_i^k(x^k)\right\| \min\left\{\frac{\left\|\nabla m_i^k(x^k)\right\|}{\beta_i^k}, \delta_k\right\} \end{aligned} \tag{4.18}$$

for all $i \in \{1, 2, \ldots, q\}$ and the lower bound of the lemma holds. For the upper bound let again $i \in \{1, 2, \ldots, q\}$ be an arbitrary, but fixed index. Furthermore, let $\min_{x \in B_k} m_i^k(x) = m_i^k(\tilde{x})$ with $\tilde{x} := x^k + t\,d$, $|t| \leq \delta_k$ and $\|d\| = 1$. From Taylor's theorem and the Cauchy Schwarz inequality it follows

$$\begin{aligned} r_i^k &= m_i^k(x^k) - \min_{x \in B_k} m_i^k(x) = m_i^k(x^k) - m_i^k(\tilde{x}) \\ &= -t\nabla m_i^k(x^k)^\top d - \frac{1}{2}t^2 d^\top \nabla^2 m_i^k(x^k) d \\ &\leq |t| \left\|\nabla m_i^k(x^k)\right\| \|d\| + \frac{1}{2}t^2 \|d\|^2 \left\|\nabla^2 m_i^k(x^k)\right\|_F \\ &\leq \delta_k \left\|\nabla m_i^k(x^k)\right\| + \frac{1}{2}\delta_k^2 \left\|\nabla^2 m_i^k(x^k)\right\|_F. \end{aligned}$$

According to Assumption A.7, it holds $\left\|\nabla^2 m_i^k(x^k)\right\|_F \leq \kappa_{\text{uhm}_i} - 1$ with $\kappa_{\text{uhm}_i} > 1$. This implies

$$r_i^k \leq \delta_k \left\|\nabla m_i^k(x^k)\right\| + \frac{1}{2}\delta_k^2 \left(\kappa_{\text{uhm}_i} - 1\right)$$

for all $i \in \{1, 2, \ldots, q\}$. □

As stated in Remark 4.1, it holds $r^k \in \mathbb{R}^q_{++}$ if x^k is not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. Then according to the preceding lemma, the following assumption on the search direction r^k is reasonable. It states that r^k is neither too flat nor too steep.

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 $(MOP^k_{m,\mathbb{R}^n})$

$$\frac{\min\limits_{i=1,\ldots,q} r_i^k}{\max\limits_{j=1,\ldots,q} r_j^k} \geq \kappa_r. \tag{4.19}$$

To formulate a sufficient decrease condition for the iterates of MHT, consider

$$d_\omega \in \operatorname*{argmin}_{\|d\|\leq 1} \max_{i=1,\ldots,q} \nabla m_i^k(x^k)^\top d, \tag{4.20}$$

a solution of the optimization problem given by $\omega_m(x^k)$ from (4.15). If x^k is not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$, then according to Lemma 2.12 d_ω is a descent direction for m^k at the current iteration point x^k. Therefore, it will provide a descent also in the trust region B_k. Furthermore, there exist scalars $\alpha_i \in [0, 1], i \in \{1, 2, \ldots, q\}$, with $\sum_{i=1}^q \alpha_i = 1$ and $\mu \geq 0$ such that

$$d_\omega = -\mu \sum_{i=1}^q \alpha_i \nabla m_i^k(x^k) \tag{4.21}$$

holds with $\|d_\omega\| = 1$. Now we consider the auxiliary function $g(x) = \sum_{i=1}^q \alpha_i m_i^k(x^k)$ and minimize g along its normed steepest descent direction d_ω starting from x^k.

Lemma 4.12 *Suppose Assumption* A.3 *holds. Let $k \in \mathbb{N}$ be an iteration with x^k not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. Let $g : \mathbb{R}^n \to \mathbb{R}$ be the quadratic function defined by $g(x) := \sum_{i=1}^q \alpha_i m_i^k(x)$ with constants $\alpha_i \geq 0, i \in \{1, 2, \ldots, q\}$, from* (4.21). *Furthermore, define*

x_c *by* $g(x_c) \coloneqq \min_{|t|\leq\delta_k} g(x^k + t\,d)$ *with* $d \coloneqq -\nabla g(x^k)/\left\|\nabla g(x^k)\right\|$ *and set* $\beta_g^k \coloneqq 1 + \left\|\nabla^2 g(x^k)\right\|_F$. *Then it holds*

$$g(x^k) - g(x_c) \geq \frac{1}{2}\left\|\nabla g(x^k)\right\| \min\left\{\frac{\left\|\nabla g(x^k)\right\|}{\beta_g^k}, \delta_k\right\}. \tag{4.22}$$

Proof. The normed steepest descent direction for g at x^k is given in (4.21) by $d_\omega = -\nabla g(x^k)/\left\|\nabla g(x^k)\right\|$. According to Assumption A.3, all model functions m_i^k, $i = 1, 2, \ldots, q$, are quadratic and it follows from Taylor's theorem

$$g(x^k + t\,d_\omega) = g(x^k) + t\nabla g(x^k)d_\omega + \frac{1}{2}t^2 d_\omega^\top \nabla^2 g(x^k) d_\omega$$

for every $t \in \mathbb{R}$. The Cauchy Schwarz inequality implies together with calculations and estimations analogous to (4.16)–(4.18) in the proof of Lemma 4.11

$$\begin{aligned}
g(x^k) - g(x_c) &= g(x^k) - \min_{|t|\leq\delta_k} g(x^k + t\,d_\omega) \\
&= \max_{|t|\leq\delta_k}\left(-t\nabla g(x^k)^\top d_\omega - \frac{1}{2}t^2 d_\omega^\top \nabla^2 g(x^k) d_\omega\right) \\
&\geq \max_{|t|\leq\delta_k}\left(t\left\|\nabla g(x^k)\right\| - \frac{1}{2}t^2\beta_g^k\right) \\
&\geq \min\left\{\frac{1}{2}\frac{\left\|\nabla g(x^k)\right\|^2}{\beta_g^k}, \frac{1}{2}\left\|\nabla g(x^k)\right\|\delta_k\right\} \\
&= \frac{1}{2}\left\|\nabla g(x^k)\right\| \min\left\{\frac{\left\|\nabla g(x^k)\right\|}{\beta_g^k}, \delta_k\right\}
\end{aligned}$$

with $\beta_g^k = \left\|\nabla^2 g(x^k)\right\|_F + 1 > 0$. □

With these findings a first decrease condition for the iteration points of MHT can be formulated.

Lemma 4.13 *Suppose Assumptions* A.1, A.3, A.4, A.7, A.9 *and* A.10 *hold. Let $k \in \mathbb{N}$ be an iteration of MHT and let $x^{k+} \in \mathbb{R}^n$ be the trial point computed in this iteration with the associated step size $t^{k+} \in \mathbb{R}$. Moreover, let $\phi_m^k(x) = \max_{i=1,\dots,q} m_i^k(x)$ be defined as in* (4.4) *and $\beta_\phi^k := \max_{i=1,\dots,q} \left\|\nabla^2 m_i^k(x^k)\right\|_F + 1$. Then there exists a constant $\tilde{\kappa}_\phi \in (0,1)$ independent of k and for each $k \in \mathbb{N}$ an index $j = j(k) \in \mathbb{N}$ such that it holds*

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq \left(\frac{1}{2}\right)^j \tilde{\kappa}_\phi \omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta_\phi^k}, \delta_k\right\}. \tag{4.23}$$

Proof. Firstly, let x^k be not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. Then according to Lemma 2.9, x^k is not locally weakly efficient for $(MOP^k_{m,\mathbb{R}^n})$ and therefore not weakly efficient for (MOP^k_m). It follows from Remark 4.1 that it holds

$$r_i^k = m_i^k(x^k) - \min_{x \in B_k} m_i^k(x) > 0 \text{ for all } i \in \{1, 2, \dots, q\}. \tag{4.24}$$

Thus, $(t^{k+}, x^{k+}) \in \mathbb{R}^{1+n}$ is computed in step 3 of MHT as a minimal solution of the auxiliary problem

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

Note that, as outlined in Remark 4.2, a minimal solution exists. According to Lemma 4.4, it holds

$$t^{k+} \leq 0 \text{ and } m^k(x^k) \geq m^k(x^{k+}). \tag{4.25}$$

Since x^k is not Pareto critical, $t^{k+} \in [-1, 0)$ follows from Lemma 4.5. This implies together with (4.25) and the constraints of (PS)

$$m_i^k(x^k) - m_i^k(x^{k+}) \geq -t^{k+} r_i^k > 0 \tag{4.26}$$

for all $i \in \{1, 2, \dots, q\}$. Together with the definition of the function ϕ_m^k it follows

$$-t^{k+} = |t^{k+}| \leq \frac{m_i^k(x^k) - m_i^k(x^{k+})}{r_i^k} \leq \frac{\phi_m^k(x^k) - m_i^k(x^{k+})}{\min\limits_{j=1,\dots,q} r_j^k} \tag{4.27}$$

for all $i \in \{1, 2, \dots, q\}$. Let $d_\omega \in \operatorname{argmin}_{\|d\| \leq 1} \max_{i=1,\dots,q} \nabla m_i^k(x^k)^\top d$ be a solution of the optimization problem given by $\omega_m(x^k)$ from (4.15). Then according to

Lemma 2.12(ii), there exist scalars $\alpha_i \in [0,1], i \in \{1,2,\dots,q\}$, with $\sum_{i=1}^q \alpha_i = 1$ and $\mu \geq 0$ such that $\|d_\omega\| = 1$ and (4.21) holds, that is $d_\omega = -\mu \sum_{i=1}^q \alpha_i \nabla m_i^k(x^k)$. For the resulting function $g(x) = \sum_{i=1}^q \alpha_i m_i^k(x)$ and the point x_c defined by $x_c \in \operatorname{argmin}_{|t|\leq \delta_k} g(x^k + t\, d_\omega)$ Lemma 4.12 holds and thus, it holds

$$g(x^k) - g(x_c) \geq \frac{1}{2} \left\|\nabla g(x^k)\right\| \min\left\{ \frac{\left\|\nabla g(x^k)\right\|}{\beta_g^k}, \delta_k \right\} \tag{4.28}$$

with $\beta_g^k = \left\|\nabla^2 g(x^k)\right\|_F + 1$. For this parameter it holds due to Assumption A.7

$$\beta_g^k = \left\|\nabla^2 g(x^k)\right\|_F + 1 \leq \sum_{i=1}^q \alpha_i \left\|\nabla^2 m_i^k(x^k)\right\|_F + 1 \leq \max_{i=1,\dots,q} \left\|\nabla^2 m_i^k(x^k)\right\|_F + 1 = \beta_\phi^k$$

which implies with (4.28)

$$g(x^k) - g(x_c) \geq \frac{1}{2} \left\|\nabla g(x^k)\right\| \min\left\{ \frac{\left\|\nabla g(x^k)\right\|}{\beta_\phi^k}, \delta_k \right\} \geq 0. \tag{4.29}$$

Due to the definition of the vector x_c there exists a scalar $|t_c| \leq \delta_k$ such that $x_c = x^k + t_c\, d_\omega$ holds. According to Lemma 2.12(i), d_ω is a descent direction for m^k at x^k. Thus, there exists a scalar $t_0 > 0$ such that it holds $m_i^k(x^k + t\, d_\omega) < m_i^k(x^k)$ for all $t \in (0, t_0]$ and for all $i \in \{1,2,\dots,q\}$. Note that $r_i^k > 0$ holds for all $i \in \{1,2,\dots,q\}$, see (4.24). We define

$$\tilde{x} := \begin{cases} x_c & \text{, if } |t_c| \leq t_0 \\ x^k + t_0\, d_\omega & \text{, else} \end{cases} \quad \text{and} \quad \tilde{t} := \max_{i=1,\dots,q} -\frac{m_i^k(x^k) - m_i^k(\tilde{x})}{r_i^k}.$$

It holds $\tilde{x} \in B_k$, $m_i^k(x^k) - m_i^k(\tilde{x}) > 0$ for all $i \in \{1,2,\dots,q\}$ and $\tilde{t} < 0$. Moreover, it holds for all $i \in \{1,2,\dots,q\}$

$$m_i^k(x^k) + \tilde{t}\, r_i^k - m_i^k(\tilde{x}) \geq m_i^k(x^k) - \frac{m_i^k(x^k) - m_i^k(\tilde{x})}{r_i^k} r_i^k - m_i^k(\tilde{x}) = 0.$$

This implies that $(\tilde{t}, \tilde{x})$ is feasible for (PS). Due to the definition of $\tilde{t}$ it follows

$$|\tilde{t}| = \min_{i=1,\dots,q} \frac{m_i^k(x^k) - m_i^k(\tilde{x})}{r_i^k} \geq \frac{\min\limits_{i=1,\dots,q} \left(m_i^k(x^k) - m_i^k(\tilde{x})\right)}{\max\limits_{i=1,\dots,q} r_i^k}. \tag{4.30}$$

Due to t^{k+} being the minimal value of (PS) it holds $\tilde{t} \geq t^{k+}$. Since $\tilde{t} < 0$ and $t^{k+} \in [-1, 0)$ holds, it follows $|\tilde{t}| \leq |t^{k+}|$. Together with (4.27) and (4.30) it follows

$$\frac{\min\limits_{j=1,\dots,q} \left(m_j^k(x^k) - m_j^k(\tilde{x})\right)}{\max\limits_{j=1,\dots,q} r_j^k} \leq |\tilde{t}| \leq |t^{k+}| \leq \frac{\phi_m^k(x^k) - m_i^k(x^{k+})}{\min\limits_{j=1,\dots,q} r_j^k} \tag{4.31}$$

for all $i \in \{1, 2, \dots, q\}$. From Assumption A.10 it follows $\min_{j=1,\dots,q} r_j^k / \max_{j=1,\dots,q} r_j^k \geq \kappa_r$. By choosing $i \in \{1, 2, \dots, q\}$ in (4.31) such that $m_i^k(x^{k+}) = \phi_m^k(x^{k+})$ holds, it follows

$$\begin{aligned}
\phi_m^k(x^k) - \phi_m^k(x^{k+}) &= \phi_m^k(x^k) - m_i^k(x^{k+}) \\
&\geq \frac{\min\limits_{j=1,\dots,q} r_j^k}{\max\limits_{j=1,\dots,q} r_j^k} \min_{j=1,\dots,q} \left(m_j^k(x^k) - m_j^k(\tilde{x})\right) \\
&\geq \kappa_r \min_{j=1,\dots,q} \left(m_j^k(x^k) - m_j^k(\tilde{x})\right).
\end{aligned} \tag{4.32}$$

Since it holds $g(x_c) \leq g(\tilde{x})$, $\sum_{i=1}^q \alpha_i = 1$ and $m_i^k(x^k) - m_i^k(\tilde{x}) > 0$ for all $i \in \{1, 2, \dots, q\}$, it follows

$$\begin{aligned}
g(x^k) - g(x_c) \geq g(x^k) - g(\tilde{x}) &= \sum_{i=1}^{q} \alpha_i \left(m_i^k(x^k) - m_i^k(\tilde{x})\right) \\
&\geq \min_{i=1,\dots,q} \left(m_i^k(x^k) - m_i^k(\tilde{x})\right) > 0.
\end{aligned}$$

Now consider the inequality given in (4.29). Either it holds

$$\min_{i=1,\dots,q} \left(m_i^k(x^k) - m_i^k(\tilde{x})\right) \geq \frac{1}{2} \left\|\nabla g(x^k)\right\| \min \left\{ \frac{\left\|\nabla g(x^k)\right\|}{\beta_\phi^k}, \delta_k \right\} \text{ or}$$

$$\min_{i=1,\dots,q} \left(m_i^k(x^k) - m_i^k(\tilde{x})\right) < \frac{1}{2} \left\|\nabla g(x^k)\right\| \min \left\{ \frac{\left\|\nabla g(x^k)\right\|}{\beta_\phi^k}, \delta_k \right\}.$$

Thus, there exists an index $j = j(k) \in \mathbb{N}$ such that it holds

$$\min_{i=1,\dots,q} \left(m_i^k(x^k) - m_i^k(\tilde{x})\right) \geq \left(\frac{1}{2}\right)^j \left\|\nabla g(x^k)\right\| \min\left\{\frac{\left\|\nabla g(x^k)\right\|}{\beta_\phi^k}, \delta_k\right\}$$

and therefore it follows from (4.32) and the definition of g

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq \kappa_r \left(\frac{1}{2}\right)^j \left\|\sum_{i=1}^q \alpha_i \nabla m_i^k(x^k)\right\| \min\left\{\frac{\left\|\sum_{i=1}^q \alpha_i \nabla m_i^k(x^k)\right\|}{\beta_\phi^k}, \delta_k\right\}$$

for every iteration $k \in \mathbb{N}$ with x^k not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. Moreover, it holds $\omega_m(x^k) \leq \left\|\sum_{i=1}^q \alpha_i \nabla m_i^k(x^k)\right\|$, see Lemma 2.12, and thus

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq \kappa_r \left(\frac{1}{2}\right)^j \omega_m(x^k) \min\left\{\frac{\omega_m(x^k)}{\beta_\phi^k}, \delta_k\right\} \tag{4.33}$$

if x^k is not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. If x^k is Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$, it holds according to Lemma 2.11 $\omega_m(x^k) = 0$. According to Lemma 4.4, it holds $m_i^k(x^k) - m_i^k(x^{k+}) \geq 0$ for all $i \in \{1, 2, \dots, q\}$ which implies due to the definition of ϕ_m^k

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) = m_i^k(x^k) - m_j^k(x^{k+}) \geq m_j^k(x^k) - m_j^k(x^{k+}) \geq 0$$

with $i, j \in \{1, 2, \dots, q\}$ suitable indices. Thus, (4.33) is also satisfied if x^k is not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. From Assumption A.9 it follows $\omega(x^k) - \omega_m(x^k) \leq \kappa_\omega \omega_m(x^k)$ for all $k \in \mathbb{N}$ with $\kappa_\omega > 0$. This implies

$$\omega_m(x^k) \geq \frac{1}{1+\kappa_\omega} \omega(x^k)$$

for all $k \in \mathbb{N}$. Then it holds for every iteration $k \in \mathbb{N}$

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq \kappa_r \left(\frac{1}{2}\right)^j \frac{1}{1+\kappa_\omega} \omega(x^k) \min\left\{\frac{1}{1+\kappa_\omega} \frac{\omega(x^k)}{\beta_\phi^k}, \delta_k\right\}.$$

Define $\tilde{\kappa}_\phi := \kappa_r/(1+\kappa_\omega)^2$. Since it holds $\kappa_r \in (0,1]$ and $1/(1+\kappa_\omega) \in (0,1)$, it holds $\tilde{\kappa}_\phi \in (0,1)$. Furthermore, it follows $\delta_k > \delta_k/(1+\kappa_\omega)$ for all $k \in \mathbb{N}$. This implies

$$\begin{aligned}\phi_m^k(x^k) - \phi_m^k(x^{k+}) &\geq \frac{\kappa_r}{1+\kappa_\omega}\left(\frac{1}{2}\right)^j \omega(x^k) \min\left\{\frac{1}{1+\kappa_\omega}\frac{\omega(x^k)}{\beta_\phi^k}, \frac{1}{1+\kappa_\omega}\delta_k\right\} \\ &= \tilde{\kappa}_\phi \left(\frac{1}{2}\right)^j \omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta_\phi^k}, \delta_k\right\}\end{aligned}$$

for all $k \in \mathbb{N}$. □

This Lemma gives a decrease condition for the trial point x^{k+} obtained by MHT in terms of a lower bound for the difference $\phi_m^k(x^k) - \phi_m^k(x^{k+})$. This lower bound is strictly positive if x^k is not Pareto critical for $(MOP_{m,\mathbb{R}^n}^k)$. In this case, the trial point ensures a decrease for at least one objective function. Thus, the following assumption is reasonable to ensure a sufficient decrease in every iteration.

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 $\beta_\phi^k = \max_{i=1,\dots,q} \left\|\nabla^2 m_i^k(x^k)\right\|_F + 1$.

This lower bound on the difference $\phi_m^k(x^k) - \phi_m^k(x^{k+})$ is essential for the convergence analysis and formulates a sufficient decrease. In every trust region approach, e.g. [CGT00; VOS14], such an assumption is used and following this general approach we proved as well a motivation for the sufficient decrease assumption.
Provided Assumption A.11, the remaining of the convergence analysis of MHT follows the single-objective trust region methods [CGT00; CSV09b] closely. Consequently, it is also similar to the convergence analysis of the multi-objective trust region method in [VOS14] which is based on the single-objective considerations. The structure of the proof is transferable – with some modifications due to the differences in the methods – and convergence to a Pareto critical point of (MOP) can be proved for MHT.

Remark 4.14 Due to Assumption A.7 it holds in every iteration $k \in \mathbb{N}$ for the constant β_ϕ^k from Assumption A.11

$$\beta_\phi^k = \max_{i=1,\dots,q} \left\| \nabla^2 m_i^k(x^k) \right\|_F + 1 \leq \max_{i=1,\dots,q} \kappa_{\text{uhm}_i}.$$

Lemma 4.15 *Suppose Assumptions* A.1 *and* A.3 *to* A.8 *hold, then it holds*

$$\left| \phi(x^{k+}) - \phi_m^k(x^{k+}) \right| \leq \kappa_{\text{em}} \delta_k^2$$

in every iteration $k \in \mathbb{N}$ *with* $\kappa_{\text{em}} \coloneqq \max_{i=1,\dots,q} \kappa_{\text{em}_i} > 0$ *and the corresponding constants from Lemma* 4.8 *and Assumption* A.8*.*

Proof. For the difference on the left-hand side it holds

$$\left| \phi(x^{k+}) - \phi_m^k(x^{k+}) \right| = \begin{cases} \left| f_i(x^{k+}) - m_i^k(x^{k+}) \right| & (i) \\ \left| f_i(x^{k+}) - m_j^k(x^{k+}) \right| & (ii) \end{cases}$$

with indices $i, j \in \{1, 2, \dots, q\}$ and $i \neq j$. In case (i) it follows $\left| \phi(x^{k+}) - \phi_m^k(x^{k+}) \right| \leq \kappa_{\text{em}_i} \delta_k^2$ due to $x^{k+} \in B_k$, Lemma 4.8 and Assumption A.8. Now consider case (ii) and assume $f_i(x^{k+}) - m_j^k(x^{k+}) > 0$. Due to the definition of ϕ, Lemma 4.8, Assumption A.8 and $x^{k+} \in B_k$ it holds

$$\left| \phi(x^{k+}) - \phi_m^k(x^{k+}) \right| \leq \left| f_i(x^{k+}) - m_i^k(x^{k+}) \right| \leq \kappa_{\text{em}_i} \delta_k^2.$$

Now assume $f_i(x^{k+}) - m_j^k(x^{k+}) < 0$. Then it holds again according to the definition of ϕ, Lemma 4.8, Assumption A.8 and $x^{k+} \in B_k$

$$\left| \phi(x^{k+}) - \phi_m^k(x^{k+}) \right| = - \left(f_i(x^{k+}) - m_j^k(x^{k+}) \right) \leq -f_j(x^{k+}) + m_j^k(x^{k+}) \leq \kappa_{\text{em}_j} \delta_k^2.$$

This implies $\left| \phi(x^{k+}) - \phi_m^k(x^{k+}) \right| \leq \max_{i=1,\dots,q} \kappa_{\text{em}_i} \delta_k^2$. □

Analogously to the single-objective trust region approach, the iterations of MHT are classified into successful, very successful and unsuccessful iterations, see also Sec-

tion 3.1. This classification uses the constants $0 < \eta_1 \leq \eta_2 < 1$ from the algorithm description in Section 4.1. The set of indices of all successful iterations is defined by

$$\mathcal{S} = \left\{ k \in \mathbb{N} \,\middle|\, \rho_\phi^k = \frac{\phi(x^k) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})} \geq \eta_1 \right\}. \tag{4.34}$$

For every successful iteration the trial point is accepted as next iteration point, i.e. it holds $x^{k+1} = x^{k+}$ for all $k \in \mathcal{S}$. The classification of iterations is specified further and the set of indices

$$\mathcal{V} = \left\{ k \in \mathbb{N} \mid \rho_\phi^k \geq \eta_2 \right\} \subseteq \mathcal{S} \tag{4.35}$$

denotes the set of very successful iterations. All iterations k not contained in $\mathcal{S}$ and therefore with $\rho_\phi^k < \eta_1$ are called unsuccessful. In these iterations the trial point is discarded and it holds $x^{k+1} = x^k$ for all $k \notin \mathcal{S}$. With this classification of iterations the following lemma illustrates the behavior of MHT for non-Pareto-critical iteration points.

Lemma 4.16 *Let $k \in \mathbb{N}$ be an iteration and suppose Assumptions* A.1 *and* A.3 *to* A.11 *hold. Suppose furthermore that x^k is not Pareto critical for* (MOP) *and*

$$\delta_k \leq \frac{\kappa_\phi (1 - \eta_2) \omega(x^k)}{\kappa_{\mathrm{e}}} \tag{4.36}$$

with $\kappa_{\mathrm{e}} \coloneqq \max_{i=1,\dots,q} \max\left\{\kappa_{\mathrm{em}_i}, \kappa_{\mathrm{uhm}_i}\right\} > 0$ and $\kappa_\phi \in (0,1)$ from Assumption A.11. *Then it holds $k \in \mathcal{V}$, that is iteration k is very successful, and $\delta_{k+1} \geq \delta_k$.*

Proof. Consider the non-Pareto critical point x^k and the corresponding iteration k. According to Lemma 2.11, it holds $\omega(x^k) > 0$ and due to $\eta_2, \kappa_\phi \in (0,1)$ it holds $\kappa_\phi(1 - \eta_2) < 1$. By (4.36), the definition of κ_{e} and Remark 4.14 it follows

$$\delta_k \leq \frac{\kappa_\phi (1 - \eta_2) \omega(x^k)}{\kappa_{\mathrm{e}}} < \frac{\omega(x^k)}{\kappa_{\mathrm{e}}} \leq \frac{\omega(x^k)}{\max_{i=1,\dots,q} \kappa_{\mathrm{uhm}_i}} \leq \frac{\omega(x^k)}{\beta_\phi^k}. \tag{4.37}$$

Consequently, it follows with Assumption A.11

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq \kappa_\phi \omega(x^k) \min\left\{ \frac{\omega(x^k)}{\beta_\phi^k}, \delta_k \right\} = \kappa_\phi \omega(x^k) \delta_k > 0. \tag{4.38}$$

Since x^k is not Pareto critical for (MOP), it follows from Lemma 4.10 that x^k is not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. Due to Assumption A.4 and Theorem 2.8 it holds $f_i(x^k) = m_i^k(x^k)$ and $m_i^k(x^k) > p_i^k$ for all $i \in \{1, 2, \ldots, q\}$. Thus, in step 2 of MHT the trial point is computed by solving (PS).
Now consider the trial point acceptance test in step 3 of MHT. According to Lemma 4.10, it holds $t^{k+} < 0$. Furthermore, it holds $\phi_m^k(x^k) - \phi_m^k(x^{k+}) > 0$ and therefore ρ_ϕ^k is computed by $\rho_\phi^k = \big(\phi(x^k) - \phi(x^{k+})\big) \,/\, \big(\phi_m^k(x^k) - \phi_m^k(x^{k+})\big)$. Due to Assumption A.4 it holds $\phi_m^k(x^k) = \phi(x^k)$. From Lemma 4.15, (4.38), the definition of κ_{e} and (4.36) it follows

$$\begin{aligned}\left|\rho_\phi^k - 1\right| &= \left|\frac{\phi(x^k) - \phi(x^{k+}) - \phi_m^k(x^k) + \phi_m^k(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})}\right| = \left|\frac{\phi_m^k(x^{k+}) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})}\right| \\ &\leq \frac{\delta_k^2 \max\limits_{i=1,\ldots,q} \kappa_{\mathrm{em}_i}}{\kappa_\phi \omega(x^k)\delta_k} \leq \frac{\delta_k \kappa_{\mathrm{e}}}{\kappa_\phi \omega(x^k)} \leq 1 - \eta_2.\end{aligned}$$

This implies

$$1 - \rho_\phi^k \leq \left|\rho_\phi^k - 1\right| \leq 1 - \eta_2$$

and therefore it holds $\rho_\phi^k \geq \eta_2$ and $k \in \mathcal{V}$. According to the trust region update in step 4 of MHT, it holds for the new trust region radius $\delta_{k+1} \geq \delta_k$. □

The next lemma shows that if the function ω is bounded from below by a positive constant for all iterations, i.e. the iteration points are not Pareto critical for (MOP), then the trust region radius is bounded from below by a positive constant in all iterations.

Lemma 4.17 *Suppose Assumptions* A.1 *and* A.3 *to* A.11 *hold. Suppose furthermore that there exists a constant* $\kappa_{\mathrm{lb}\omega} > 0$ *such that* $\omega(x^k) \geq \kappa_{\mathrm{lb}\omega}$ *holds for every iteration* $k \in \mathbb{N}$. *Then there exists a constant* $\kappa_{\mathrm{lb}\delta} > 0$ *such that* $\delta_k \geq \kappa_{\mathrm{lb}\delta}$ *holds for all* $k \in \mathbb{N}$.

Proof. Assume that for every $\kappa > 0$ there exists an index $k \in \mathbb{N}$ with $\delta_k < \kappa$. Consider

$$\kappa := \frac{\gamma_1 \kappa_\phi \kappa_{\mathrm{lb}\omega}(1 - \eta_2)}{\kappa_{\mathrm{e}}}$$

with the constants $\gamma_1 \in (0, 1)$ from MHT and $\kappa_\phi, \kappa_{\mathrm{e}}$ defined in Assumption A.11 and Lemma 4.16. Let k_0 be the first iteration with $\delta_{k_0} < \kappa$. Then it holds $\delta_{k_0} < \delta_{k_0 - 1}$ and

according to the trust region update in step 4 of MHT it holds $\gamma_1 \delta_{k_0-1} \leq \delta_{k_0}$. This implies

$$\delta_{k_0-1} \leq \frac{\delta_{k_0}}{\gamma_1} < \frac{\kappa}{\gamma_1} = \frac{\kappa_\phi \kappa_{\text{lb}\omega}(1-\eta_2)}{\kappa_{\text{e}}} \leq \frac{\kappa_\phi \omega(x^{k_0-1})(1-\eta_2)}{\kappa_{\text{e}}}.$$

Due to the assumption $\omega(x^{k_0-1}) \geq \kappa_{\text{lb}\omega} > 0$ and Lemma 2.11 x^{k_0-1} is not Pareto critical for (MOP). Therefore, the preconditions of Lemma 4.16 are fulfilled and it holds $k_0 - 1 \in \mathcal{V}$ and $\delta_{k_0-1} \leq \delta_{k_0}$. This contradicts $\delta_{k_0} < \delta_{k_0-1}$ and therefore the initial assumption. □

With the preceding results it can be proved that in case of finitely many successful iterations MHT converges to a Pareto critical point.

Lemma 4.18 *Suppose Assumptions* A.1 *and* A.3 *to* A.11 *hold and MHT (Algorithm* 4*) has only finitely many successful iterations $k \in \mathcal{S} = \{k \in \mathbb{N} \mid \rho_\phi^k \geq \eta_1\}$. Then there exists an index $j \in \mathbb{N}$ such that it holds $x^i = x^j$ for all $i > j$ and x^j is a Pareto critical point for (MOP).*

Proof. Let k_0 be the index of the last successful iteration. Then all subsequent iterations are unsuccessful, i.e. it holds $\rho_\phi^k < \eta_1$ for all $k > k_0$. Step 3 of MHT ensures $x^{k_0+j} = x^{k_0+1}$ for all $j > 1$. Since all iterations are unsuccessful for sufficiently large $k \in \mathbb{N}$, the choice of the constants $0 < \gamma_1 \leq \gamma_2 < 1$ and the trust region update in step 4 imply $\lim_{k\to\infty} \delta_k = 0$.
Assume that x^{k_0+1} is not Pareto critical for (MOP). Then Lemma 4.16 implies that there exists a successful iteration whose index is larger than k_0. This is a contradiction to k_0 being the last successful iteration. Hence, x^{k_0+1} is Pareto critical for (MOP). □

Now we consider the case that MHT has infinitely many successful iterations.

Lemma 4.19 *Suppose Assumptions* A.1 *to* A.11 *hold and MHT (Algorithm* 4*) has infinitely many successful iterations $k \in \mathcal{S}$. Then it holds*

$$\liminf_{k\to\infty} \omega(x^k) = 0.$$

Proof. Suppose it holds $\liminf_{k\to\infty} \omega(x^k) \neq 0$. Then without loss of generality there exists a sequence $\{\omega(x^k)\}_k$ and a constant $\varepsilon > 0$ with $\omega(x^k) \geq \varepsilon$ for all $k \in \mathbb{N}$.

According to Lemma 4.17, there exists a constant $\kappa_{\text{lb}\delta} > 0$ such that $\delta_k \geq \kappa_{\text{lb}\delta}$ holds for all $k \in \mathbb{N}$. From Remark 4.14 it follows

$$\beta_\phi^k \leq \max_{i=1,\dots,q} \kappa_{\text{uhm}_i} \leq \max_{i=1,\dots,q} \left\{\kappa_{\text{uhm}_i}, \kappa_{\text{em}_i}\right\} = \kappa_{\text{e}}$$

for every iteration $k \in \mathbb{N}$ given the constants κ_{uhm_i}, κ_{em_i} and κ_{e} from Assumptions A.7 and A.8 and Lemmas 4.8 and 4.16. Consider a successful iteration $k \in \mathcal{S}$ and let (t^{k+}, x^{k+}) be the minimal solution of (PS) computed in iteration k. Then it holds $\rho_\phi^k \geq \eta_1$ and it follows from Assumption A.11

$$\begin{aligned}
\phi(x^k) - \phi(x^{k+}) &\geq \eta_1 \left(\phi_m^k(x^k) - \phi_m^k(x^{k+})\right) \\
&\geq \eta_1 \kappa_\phi \omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta_\phi^k}, \delta_k\right\} \\
&\geq \eta_1 \kappa_\phi \varepsilon \min\left\{\frac{\varepsilon}{\kappa_{\text{e}}}, \kappa_{\text{lb}\delta}\right\}.
\end{aligned}$$

For every successful iteration it holds $x^{k+1} = x^{k+}$. Thus, summing over all successful iterations gives

$$\phi(x^0) - \phi(x^{k+1}) = \sum_{j=0, j\in\mathcal{S}}^{k} \phi(x^j) - \phi(x^{j+1}) \geq \sigma_k \eta_1 \kappa_\phi \varepsilon \min\left\{\frac{\varepsilon}{\kappa_{\text{e}}}, \kappa_{\text{lb}\delta}\right\}$$

with σ_k being the number of successful iterations up to iteration k. Since there are infinitely many such iterations in $\mathcal{S}$, it holds $\lim_{k\to\infty} \sigma_k = \infty$. Hence, the difference between $\phi(x^0)$ and $\phi(x^{k+1})$ is unbounded. This is a contradiction to Assumption A.2 that ϕ is bounded from below. Consequently, the initial assumption is false and it holds $\liminf_{k\to\infty} \omega(x^k) = 0$. □

The following theorem is the main result about convergence of MHT. It states that the algorithm produces a sequence of iterates with ω converging to zero. According to Lemma 2.11, this characterizes Pareto criticality.

Theorem 4.20 *Suppose Assumptions* A.1 *to* A.11 *hold. Then MHT (Algorithm* 4*) produces a sequence of iterates* $\left\{x^k\right\}_k$ *with*

$$\lim_{k\to\infty} \omega(x^k) = 0.$$

If the sequence $\left\{x^k\right\}_k$ *has accumulation points, then all these points are Pareto critical for* (MOP).

Proof. If MHT has only finitely many successful iterations $k \in \mathcal{S}$, then according to Lemma 4.18 the sequence of iterates $\left\{x^k\right\}_k$ converges to a Pareto critical point of (MOP). By Lemma 2.11 it follows $\lim_{k\to\infty} \omega(x^k) = 0$.
Now consider the case if there are infinitely many successful iterations $k \in \mathcal{S}$. According to Lemma 4.19 it holds $\liminf_{k\to\infty} \omega(x^k) = 0$. Hence, it remains to show $\limsup_{k\to\infty} \omega(x^k) = 0$. For that purpose, we consider at first only the set of successful iterations $\mathcal{S}$. Assume that there exists a subsequence $\{t_j\}_j \subset \mathcal{S}$ with

$$\omega(x^{t_j}) \geq 2\varepsilon > 0 \tag{4.39}$$

for some constant $\varepsilon > 0$ and for all j. By Lemma 4.19 it follows that for all t_j there exists a first successful iteration $l_j > t_j$ fulfilling $\omega(x^{l_j+1}) < \varepsilon$. Then there is another subsequence indexed by $\{l_j\}_j$ such that it holds

$$\omega(x^k) \geq \varepsilon \text{ for } t_j \leq k \leq l_j \text{ and } \omega(x^{l_j+1}) < \varepsilon. \tag{4.40}$$

Consider the subsequence whose indices are contained in the set

$$\mathcal{K} \coloneqq \{k \in \mathcal{S} \mid \exists\, j \in \mathbb{N} : t_j \leq k \leq l_j\} \subseteq \mathcal{S},$$

where t_j and l_j belong to the two subsequences defined above. For every successful iteration $k \in \mathcal{S}$ it holds $\rho_\phi^k \geq \eta_1$ and $x^{k+1} = x^{k+}$. The definition of ρ_ϕ^k, the fact $\mathcal{K} \subseteq \mathcal{S}$, Assumption A.11, Remark 4.14 and (4.40) imply for $k \in \mathcal{K}$

$$\begin{aligned}\phi(x^k) - \phi(x^{k+1}) &\geq \eta_1 \left(\phi_m^k(x^k) - \phi_m^k(x^{k+1})\right) \\ &\geq \eta_1 \kappa_\phi \omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta_\phi^k}, \delta_k\right\} \\ &\geq \eta_1 \kappa_\phi \varepsilon \min\left\{\frac{\varepsilon}{\kappa_{\mathrm{e}}}, \delta_k\right\}. \end{aligned} \tag{4.41}$$

Due to the definition of the trial point it holds $m_i^k(x^{k+}) \leq m_i^k(x^k)$ for all $i \in \{1, 2, \ldots, q\}$ which implies $\phi(x^{k+1}) = \phi(x^{k+}) \leq \phi(x^k)$ for all $k \in \mathcal{K}$. Thus, the sequence $\left\{\phi(x^k)\right\}_k$ is monotonically decreasing. Furthermore, due to Assumption A.2 it is bounded from below. Hence, $\left\{\phi(x^k)\right\}_k$ is convergent and it holds $\lim_{k\to\infty} \phi(x^k) - \phi(x^{k+1}) = 0$ which implies

$$\lim_{k\to\infty, k\in\mathcal{K}} \delta_k = 0.$$

Thus, the second term dominates the minimum in (4.41) and it holds for $k \in \mathcal{K}$ sufficiently large

$$\delta_k \leq \frac{1}{\eta_1 \kappa_\phi \varepsilon} \left(\phi(x^k) - \phi(x^{k+1})\right).$$

Note that it holds $x^{k+1} = x^k$ for every unsuccessful iteration $k \notin \mathcal{S}$. Consequently, it follows for j sufficiently large

$$\left\|x^{t_j} - x^{l_j+1}\right\| \leq \sum_{i=t_j, i\in\mathcal{K}}^{l_j} \left\|x^i - x^{i+1}\right\| \leq \sum_{i=t_j, i\in\mathcal{K}}^{l_j} \delta_i \leq \frac{1}{\eta_1 \kappa_\phi \varepsilon} \left(\phi(x^{t_j}) - \phi(x^{l_j+1})\right).$$

Again, because the sequence $\left\{\phi(x^k)\right\}_k$ is monotonically decreasing and bounded from below, it holds

$$\lim_{j\to\infty} \frac{1}{\eta_1 \kappa_\phi \varepsilon} \left(\phi(x^{t_j}) - \phi(x^{l_j+1})\right) = 0$$

which implies

$$\lim_{j\to\infty} \left\|x^{t_j} - x^{l_j+1}\right\| = 0.$$

Since ω is uniformly continuous due to Assumption A.6, see Remark 4.7, it follows

$$\lim_{j\to\infty} \left|\omega(x^{t_j}) - \omega(x^{l_j+1})\right| = 0.$$

This is a contradiction to the definition of the sequences $\{t_j\}_j$ and $\{l_j\}_j$ in (4.40) which implies with (4.39) that $\left|\omega(x^{t_j}) - \omega(x^{l_j+1})\right| \geq \varepsilon$ holds. Consequently, no subsequence satisfying (4.39) can exist, i.e.

$$\begin{aligned} &\text{there exists no } \varepsilon > 0 \text{ and no subsequence } \{t_j\}_j \subset \mathcal{S} \\ &\text{with } \omega(x^{t_j}) \geq 2\varepsilon > 0 \text{ for all } j \in \mathbb{N}. \end{aligned} \tag{4.42}$$

Let $\mathcal{I} = \{1, 2, 3, \ldots\}$ be the set of indices of all iterations of MHT. Since it holds $\mathcal{S} \subset \mathcal{I}$ and $|\mathcal{S}| = \infty$, it follows $|\mathcal{I}| = \infty$. Assume there exists a constant $\varepsilon > 0$ and a subsequence $\{l_j\}_j \subset \mathcal{I}$ with $\omega(x^{l_j}) \geq 2\varepsilon > 0$ for all $j \in \mathbb{N}$. At first we consider the case $|\mathcal{I} \setminus \mathcal{S}| < \infty$, i.e. there exist only finitely many unsuccessful iterations. Then there exists a largest index $j^{\max} \coloneqq \max\{j \in \mathbb{N} \mid l_j \in \mathcal{I} \setminus \mathcal{S}\}$. This implies the existence of a sequence $\{l_j\}_{j, j>j^{\max}} \subset \mathcal{S}$ with $\omega(x^{l_j}) \geq 2\varepsilon > 0$ for all $j \in \mathbb{N}$, $j > j^{\max}$. This contradicts (4.42) and therefore it follows $\limsup_{k\to\infty} \omega(x^k) = 0$. This implies together with Lemma 4.19 $\lim_{k\to\infty} \omega(x^k) = 0$.
Now consider the case $|\mathcal{I} \setminus \mathcal{S}| = \infty$. According to (4.42) there can only exist finitely many indices $j \in \mathbb{N}$ such that $l_j \in \mathcal{S}$ holds. Thus, we consider the subsequence

$$\left\{\tilde{l}_j\right\}_j \subset \{l_j\}_j \text{ with } \left\{\tilde{l}_j\right\}_j \subset \mathcal{I} \setminus \mathcal{S} \text{ and } \omega(x^{\tilde{l}_j}) \geq 2\varepsilon \text{ for all } j \in \mathbb{N}.$$

Note that for every unsuccessful iteration $\tilde{l}_j \in \mathcal{I} \setminus \mathcal{S}$ it holds $x^{\tilde{l}_j+1} = x^{\tilde{l}_j}$. Since we consider the case $|\mathcal{S}| = \infty$, this implies that for every index $\tilde{l}_j \in \mathcal{I} \setminus \mathcal{S}$ there exists a smallest subsequent successful iteration $z = z(\tilde{l}_j) \coloneqq \min\left\{k \in \mathcal{S} \mid k > \tilde{l}_j\right\}$. Note that it holds $x^z = x^i$ for all $i \in \mathcal{I}$ with $\tilde{l}_j \leq i < z$. Therefore, there exists a subsequence $\left\{z(\tilde{l}_j)\right\}_j \subset \mathcal{S}$ with

$$\omega(x^{z(\tilde{l}_j)}) = \omega(x^{\tilde{l}_j}) \geq 2\varepsilon > 0$$

for all $j \in \mathbb{N}$. This contradicts (4.42) and it follows $\limsup_{k\to\infty} \omega(x^k) = 0$ and again, due to Lemma 4.19, $\lim_{k\to\infty} \omega(x^k) = 0$.
Let $\bar{x}$ be an accumulation point of the sequence $\left\{x^k\right\}_k$ produced by MHT and assume

that it is not Pareto critical for (MOP). Then according to Lemma 2.11, it holds $\omega(\bar{x}) > 0$. This is a contradiction to $\lim_{k\to\infty} \omega(x^k) = 0$. Thus, every accumulation point of $\{x^k\}_k$ is Pareto critical for (MOP). □

According to the basic relations between convexity and efficiency presented in Section 2.3, the following special cases hold for Theorem 4.20.

Lemma 4.21 *Suppose Assumptions* A.1 *to* A.11 *hold. Let $\{x^k\}_k$ be a sequence with accumulation points generated by MHT (Algorithm* 4*). Then for every accumulation point $\bar{x}$ the following statements hold.*

(i) If f is $\mathbb{R}^q_+$-convex or if the functions f_i, $i = 1, 2, \ldots, q$, are pseudoconvex on $\mathbb{R}^n$, then $\bar{x}$ is weakly efficient for (MOP).

(ii) If f is strictly $\mathbb{R}^q_+$-convex, then $\bar{x}$ is efficient for (MOP).

Proof. All results follow immediately from Theorems 2.16, 2.18 and 4.20. □

4.6 Minor Modifications of MHT

MHT as presented in Section 4.1 is a basic version of a multi-objective trust region approach for heterogeneous functions. Several aspects of the algorithm, such as the choice of the model functions or the realization of the trial point acceptance test can be modified.

Besides, it is possible to apply MHT to optimization problems with only expensive functions. Assumptions A.1 to A.3 and A.5 to A.11 are still required. Since no cheap function exists, Assumption A.4 needs to be omitted and the statement in Lemma 4.8 about the validity of the model functions $m_2^k, m_3^k, \ldots, m_q^k$ for all $k \in \mathbb{N}$ does not hold any more. Instead, Assumption A.8 is now also required for $m_2^k, m_3^k, \ldots, m_q^k$ to ensure the validity of all model functions. However, such assumptions are commonly used in trust region methods, see for example [CGT00; AH17]. With these modifications in the assumptions the results from Section 4.5 and especially the main result about convergence given in Theorem 4.20 still hold.

Hereafter, we present further minor modifications of MHT for which the convergence results from Section 4.5 transfer.

4.6.1 Linear Model for Expensive Function

In MHT the model function for the expensive function f_1 is defined as interpolation model using quadratic polynomials. Computing such a quadratic model function needs $(n+1)(n+2)/2$ function evaluations, where n is the dimension of the domain. Linear models, i.e. interpolation models based on linear polynomials, only need $n+1$ function evaluations. Table 4.1 gives an overview of how many function evaluations are required to compute one model function (quadratic/linear interpolation with Lagrange polynomials) depending on the dimension of the domain.

n	2	3	4	5	10	20	30	40	50
quadratic model	6	10	15	21	66	231	496	861	1326
linear model	3	4	5	6	11	21	31	41	51

Table 4.1 – Function evaluations for computing an interpolation model

However, in MHT not in every iteration a new model is built with a whole new set of interpolation points. The model is only updated if necessary and former interpolation points are reused if possible. See also Sections 3.2 and 4.2 for further information about the model functions and updating them.

Quadratic models are most commonly used in the trust region concept, see for example [CGT00; CSV09b], but linear models are mentioned as an alternative. They are an option to save function evaluations, for example when considering optimization problems with higher dimensions, see also [CGT00]. Trust region approaches using linear model functions are for example presented in [The11; Car+18]. The trust region approach presented in [The11] is designed for expensive scalar-valued optimization problems and is used as a comparison method for the numerical tests of MHT in Chapter 6.

First numerical tests show that linear models can save function evaluations for the expensive function f_1 if optimization problems of higher dimensions are considered. For the purpose of these tests, MHT was applied to three test problems (Test Problems P.16, P.19 and P.34 from Appendix A.2). These test problems are scalable regarding the dimension n of the domain. We considered the values $n = 4, 5, 10, 20, 30$ for the tests to compare MHT with linear and quadratic model functions for the expensive function. For every test problem various starting points were considered.

Table 4.2 summarizes the results of the tests to compare linear and quadratic model functions. The table lists the range (R) and the mean values (M) of function evaluations required until MHT terminates. In most cases it terminates with an efficient or weakly efficient point. Only in some cases it terminates because the maximum number of allowed function evaluations is reached which was set to 2000.

n	4	5	10	20	30
linear (R)	10-2000	12-2000	15-2000	46-621	39-2000
quadratic (R)	21-125	28-276	132-737	462-2000	499-2000
linear (M)	276.76	337.70	246.50	204.90	439.41
quadratic (M)	40.64	69.00	222.92	1000.28	1484.86

Table 4.2 – Range (R) and mean value (M) of function evaluations for MHT with linear and quadratic model

The first row of the table shows a high variation in the number of required function evaluations for the linear model. The high numbers of 2000 function evaluations occurred only for individual instances due to the starting points. This is confirmed by the mean values listed in the third row.
The table shows that it is reasonable to use a linear model in MHT if optimization problems with higher dimensions are considered. Comparing the mean values in the third and the fourth row indicates that up to dimension 10 linear models require on average more function evaluations than quadratic models. These first results also indicate that from dimension 10 function evaluations can be saved by using linear model functions.
However, a linear model is less accurate than a quadratic model and therefore it needs to be updated more often. Thus, it is not guaranteed that less function evaluations are required when using these simpler model functions. This is illustrated in Table 4.2 for lower dimensions n of the domain. Of course the location of the starting point also influences the number of function evaluations required until MHT terminates. Further results of numerical tests with the scalable test problems are stated in Section 6.4.3.
Since the topic of this thesis is not to study the behavior of the model functions, no extensive tests on this subject were executed. Further information about different choices of model functions in the trust region framework can for example be found in [CGT00;

CSV09b; WS11; AH17]. Based on the results of the first numerical tests, in the implementation of MHT linear interpolation models are used for the test instances with dimension $n \geq 10$. For $n < 10$, quadratic interpolation models are used.
For the convergence analysis of MHT in Section 4.5, the model functions are assumed to be quadratic in every iteration $k \in \mathbb{N}$, see Assumption A.3. Consequently, if a linear model is chosen for f_1, Assumption A.3 needs to be omitted for m_1^k. Nevertheless, if Assumptions A.1, A.2 and A.4 to A.11 and Assumption A.3 only for $m_2^k, m_3^k, \ldots, m_q^k$ are imposed and m_1^k is a linear interpolation model, the results from Section 4.5 still hold.
Lemmas 4.8 and 4.10 hold unchanged. Since m_1^k is a linear function, it holds $\nabla^2 m_1^k(x) = 0_{n,n} \in \mathbb{R}^{n \times n}$ for all $k \in \mathbb{N}$, i.e. the Hessian is the zero matrix. This needs to be regarded in Lemma 4.9, but the proof of the statement follows analogously.
Furthermore, using a linear model m_1^k implies $\beta_1^k = 1 + \left\|\nabla^2 m_1^k(x^k)\right\|_F = 1$ for all $k \in \mathbb{N}$. This constant β_1^k is firstly introduced in Lemma 4.11 and used throughout the subsequent lemmas in Section 4.5. Regarding this and adapting the assumptions as explained above, Lemmas 4.12, 4.13 and 4.15 to 4.19 and Remark 4.14 still hold and the main result about convergence given in Theorem 4.20 also holds.
With the same line of argument and some further changes in the assumptions the results from Section 4.5 would also be transferable if for the cheap functions $f_2, f_3, \ldots, f_q$ linear interpolation models are used. However, this is not reasonable. The computational effort to obtain quadratic models for the cheap functions is assumed to be low or at least acceptable, which is why these more accurate models should be used.

4.6.2 Alternative to Ideal Point

The algorithm MHT uses a search direction defined in the image space. It is given by local ideal points p^k with $p_i^k = \min_{x \in B_k} m_i^k(x)$ for $i = 1, 2, \ldots, q$. However, this search direction can be modified, also with user-given information, such that the convergence results from Section 4.5 still hold.
Instead of computing the individual minima of the functions $m_i^k, i \in \{1, 2, \ldots, q\}$, in B_k in every iteration $k \in \mathbb{N}$, a lower bound is also sufficient. The ideal point p^k can be replaced by a point $\tilde{p}^k \in \mathbb{R}^n$ with

$$\tilde{p}_i^k \leq p_i^k \text{ for all } i \in \{1, 2, \ldots, q\}\,. \tag{4.43}$$

The surrogate vector $\tilde{p}^k$ depends on the index k and can therefore be adapted from iteration to iteration. It can also be defined independently of the iteration index as a lower bound for the functions f_i, $i = 1, 2, \ldots, q$, i.e. it can be defined as $\tilde{p} = \tilde{p}^k$ for all $k \in \mathbb{N}$ with $\tilde{p}_i \leq \min_{x \in \mathbb{R}^n} f_i(x)$ for $i = 1, 2, \ldots, q$.
Figure 4.5 illustrates the alternative choices for the ideal point p^k in iteration $k \in \mathbb{N}$. All points from the gray shaded area can be used as surrogate vector $\tilde{p}^k$.

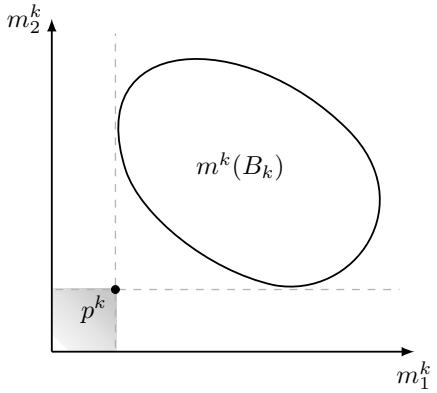

Figure 4.5 – Lower bound for local ideal point p^k

This modification of MHT is useful for applications. In some applications the user has additional information about the optimization problem, a preference for the solution or a desired result that may be unrealizable. This kind of information can be included in MHT by replacing the ideal point by this user-given desired point $\hat{p}$. If $\hat{p} \leq p^k$ holds for all $k \in \mathbb{N}$, i.e. if it is ensured that $\hat{p}$ provides a lower bound, it can be used for all iterations $k \in \mathbb{N}$ of the algorithm. If it is not ensured that it is a lower bound, it could still be used in the first iterations. If the algorithm reaches a point with the function values of $\hat{p}$, the ideal point p^k could be computed and used again.
For the theoretical considerations we use the general formulation of $\tilde{p}^k \in \mathbb{R}^n$ depending on the iteration index $k \in \mathbb{N}$. With few small changes, the convergence results from Section 4.5 can be transferred if p^k is replaced by a vector $\tilde{p}^k$ fulfilling (4.43). Before explaining the necessary changes, we give an overview of some general properties for the iterations $k \in \mathbb{N}$ of MHT (Algorithm 4) if p^k is surrogated by $\tilde{p}^k$. For this purpose, we define the related Pascoletti-Serafini problem by

$$\min \left\{ t \in \mathbb{R} \mid f(x^k) + t\,\tilde{r}^k - m^k(x) \in \mathbb{R}^q_+, x \in B_k \right\} \tag{$\widetilde{PS}_1$}$$

with $\tilde{r}^k := f(x^k) - \tilde{p}^k$ and $\tilde{p}^k \leq p^k$. Note that, analogous to Remark 4.2 and according to [Göp+03; Eic08], there exists a minimal solution for ($\widetilde{PS}_1$) if $\tilde{r}^k \in \mathbb{R}^q_{++}$ holds. This is for example fulfilled if $\tilde{p}^k < p^k$ holds, i.e. if the surrogate ideal point is a strict lower bound for the original ideal point. Thus, in step 2 of MHT it needs to be tested if $\tilde{r}^k_i > 0$ holds for all $i \in \{1, 2, \ldots, q\}$.

Lemma 4.22 *Let Assumption* A.4 *hold. Consider MHT (Algorithm* 4*) with p^k surrogated by $\tilde{p}^k$ with $\tilde{p}^k \leq p^k$ and the resulting Pascoletti-Serafini problem ($\widetilde{PS}_1$) with $\tilde{r}^k = f(x^k) - \tilde{p}^k$ in all iterations $k \in \mathbb{N}$. The following statements hold.*

(i) It holds $0_q \leq r^k \leq \tilde{r}^k$ with $r^k = f(x^k) - p^k$ from (PS) for all iterations $k \in \mathbb{N}$.

(ii) For every minimal solution $(\bar{t}, \bar{x})$ of ($\widetilde{PS}_1$) it holds $\bar{t} \leq 0$.

(iii) Let $k \in \mathbb{N}$ be an iteration such that x^k is not weakly efficient for (MOP^k_m). For every minimal solution $(\bar{t}, \bar{x})$ of ($\widetilde{PS}_1$) it holds $\bar{t} \in [-1, 0)$.

(iv) For the trial point x^{k+} in MHT it holds $m^k(x^k) \geq m^k(x^{k+})$ and $\phi^k_m(x^k) - \phi^k_m(x^{k+}) \geq 0$ in all iterations $k \in \mathbb{N}$.

(v) Let $k \in \mathbb{N}$ be an iteration such that x^k is not weakly efficient for (MOP^k_m). Then it holds $m^k(x^k) > m^k(x^{k+})$ and $\phi^k_m(x^k) - \phi^k_m(x^{k+}) > 0$.

Proof. Statement (i) follows directly from the definition of $\tilde{r}^k$ and the inequality $\tilde{p}^k \leq p^k$. Statement (ii) follows analogously to Lemma 4.4 since $(0, x^k)$ is always feasible for ($\widetilde{PS}_1$). Statement (iii) follows analogously to Lemma 4.5 since it holds $r^k \leq \tilde{r}^k$ according to (i).
To prove statement (iv) let (t^{k+}, x^{k+}) be the minimal solution of ($\widetilde{PS}_1$) computed in iteration $k \in \mathbb{N}$ of MHT. Thus, together with (i) and (ii) it follows

$$f(x^k) - m^k(x^{k+}) \geq -t^{k+}\,\tilde{r}^k \geq 0_q.$$

Since the model functions are exact at the current iteration point, see Assumption A.4, it follows $m^k(x^k) = f(x^k) \geq m^k(x^{k+})$. Furthermore, this implies

$$\phi(x^k) - \phi^k_m(x^{k+}) = \phi^k_m(x^k) - \phi^k_m(x^{k+}) \geq 0.$$

Statement (v) follows from (iii), (iv) and the fact that $r_i^k > 0$ holds for all $i \in \{1, 2, \ldots, q\}$ if x^k is not weakly efficient for (MOP_m^k), see also the proof of Lemma 4.5. □

Lemmas 4.8 and 4.9 from Section 4.5 hold with the same assumptions and without any changes. Statement (i) of Lemma 4.10 also holds unchanged, statement (ii) follows from Lemma 4.22 (iii). Lemma 4.11 gives upper and lower bounds for the vector $r^k = f(x^k) - p^k$, $k \in \mathbb{N}$, with the original ideal point p^k to justify Assumption A.10. Since it holds $\tilde{r}^k \geq r^k$ for all $k \in \mathbb{N}$, the lower bounds from Lemma 4.11 also hold. To ensure upper bounds and to justify an analogous assumption to Assumption A.10, the following assumption for the surrogate point $\tilde{p}^k$ is added.

Assumption A.12 There exist constants $\tilde{p}_i \in \mathbb{R}$, $i = 1, 2, \ldots, q$, such that it holds $\tilde{p}^k \geq \tilde{p}_i$ for all $i \in \{1, 2, \ldots, q\}$ and for all $k \in \mathbb{N}$.

Analogously to Assumption A.10, we impose the following assumption for $\tilde{r}^k$.

Assumption A.13 There exists a constant $\kappa_{\tilde{r}} \in (0, 1]$ such that it holds for every iteration $k \in \mathbb{N}$ with x^k not Pareto critical for $(MOP_{m,\mathbb{R}^n}^k)$

$$\frac{\min\limits_{i=1,\ldots,q} \tilde{r}_i^k}{\max\limits_{j=1,\ldots,q} \tilde{r}_j^k} \geq \kappa_{\tilde{r}}.$$

With Assumptions A.1 to A.9 complemented by Assumption A.12 and Assumption A.10 replaced by Assumption A.13, the further results stated in Lemmas 4.12 and 4.13 follow analogously. Thus, Assumption A.11 can be justified analogously and the remaining auxiliary results given by Remark 4.14 and Lemmas 4.15 to 4.19 follow analogously under the stated assumptions and by using $\tilde{r}^k \geq r^k$. Thus, the main result about convergence given in Theorem 4.20 transfers.

4.6.3 Strict Version of Trial Point Acceptance Test

In every iteration of the basic version of MHT, see Algorithm 4, a trial point x^{k+} is computed as a candidate for the next iteration point. In step 3 of the algorithm, referred to as the trial point acceptance test, it is decided if x^{k+} is accepted or discarded as next iteration point. This depends on the value of the quotient

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

defined in (4.5) with $\phi(x) = \max_{i=1,\dots,q} f_i(x)$ and $\phi_m^k(x) = \max_{i=1,\dots,q} m_i^k(x)$ from (4.4) in Section 4.1. If $\rho_\phi^k \geq \eta_1$ holds with $\eta_1 \in (0,1)$, x^{k+} is accepted and $x^{k+1} = x^{k+}$ is set. Otherwise, the next iteration point is defined as $x^{k+1} = x^k$. This is described in detail in Section 4.4.

The acceptance test is part of every trust region approach and this multi-objective realization using ρ_ϕ^k is for example also used in [VOS14]. An alternative to considering ρ_ϕ^k is transferring the acceptance test from the single-objective trust region approach directly, that is considering quotients

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

as given in (4.8) in Section 4.4. 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 indices $i \in \{1, 2, \dots, q\}$. 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 [RK14] and it is stricter than the trial point acceptance test used in MHT. This is proved in Lemma 4.23 and illustrated in Figure 4.4 in Section 4.4 on page 46. The influence of the trial point acceptance test and its effects are also discussed in Section 6.4.4 where numerical results for MHT are presented. Lemma 4.23 states the connection between the two versions of the trial point acceptance test.

Lemma 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*

$$\rho_\phi^k = \frac{\phi(x^k) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})} \quad \textit{and} \quad \rho_i^k = \frac{f_i(x^k) - f_i(x^{k+})}{m_i^k(x^k) - m_i^k(x^{k+})}$$

for all $i \in \{1, 2, \ldots, q\}$ *be defined as in* (4.5) *and* (4.8) *with* $\phi(x) = \max_{i=1,\ldots,q} f_i(x)$ *and* $\phi_m^k(x) = \max_{i=1,\ldots,q} m_i^k(x)$ *from* (4.4). *If* $\rho_i^k \geq \eta_1$ *holds for all* $i \in \{1, 2, \ldots, q\}$, *then it holds* $\rho_\phi^k \geq \eta_1$.

Proof. Let $\rho_i^k \geq \eta_1$ hold for all $i \in \{1, 2, \ldots, q\}$. According to Assumption A.4, it holds $f(x^k) = m^k(x^k)$ for all $k \in \mathbb{N}$. This implies together with the definition of ρ_i^k

$$f_i(x^k) - f_i(x^{k+}) \geq \eta_1 \left(m_i^k(x^k) - m_i^k(x^{k+})\right) = \eta_1 \left(f_i(x^k) - m_i^k(x^{k+})\right)$$

for all $i \in \{1, 2, \ldots, q\}$. This is equivalent to

$$(1 - \eta_1)\, f_i(x^k) \geq f_i(x^{k+}) - \eta_1 m_i^k(x^{k+})$$

for all $i \in \{1, 2, \ldots, q\}$. Since it holds $\eta_1 \in (0, 1)$, $f_i(x^k) \leq \phi(x^k)$ and $m_i^k(x^{k+}) \leq \phi_m^k(x^{k+})$ for all $i \in \{1, 2, \ldots, q\}$, it follows

$$(1 - \eta_1)\, \phi(x^k) \geq f_i(x^{k+}) - \eta_1 m_i^k(x^{k+}) \geq f_i(x^{k+}) - \eta_1 \phi_m^k(x^{k+})$$

for all $i \in \{1, 2, \ldots, q\}$. Let $j \in \{1, 2, \ldots, q\}$ be the index with $f_j(x^{k+}) = \phi(x^{k+})$. Then it holds

$$(1 - \eta_1)\, \phi(x^k) \geq f_j(x^{k+}) - \eta_1 \phi_m^k(x^{k+}) = \phi(x^{k+}) - \eta_1 \phi_m^k(x^{k+}).$$

According to Assumption A.4, it holds $\phi(x^k) = \phi_m^k(x^k)$ and it follows

$$\phi(x^k) - \phi(x^{k+}) \geq \eta_1 \left(\phi(x^k) - \phi_m^k(x^{k+})\right) = \eta_1 \left(\phi_m^k(x^k) - \phi_m^k(x^{k+})\right).$$

From the definition of ρ_ϕ^k it then follows $\rho_\phi^k \geq \eta_1$. □

Since the version of the trial point acceptance test using ρ_i^k, $i = 1, 2, \ldots, q$, is stricter than the version using ρ_ϕ^k in MHT, it is possible that less iterations are successful when using ρ_i^k, i.e. in less iterations the trial point is accepted.
As illustrated schematically in Figure 4.4 on page 46, in MHT trial points can be accepted that decrease one objective function, but increase another objective. This can be negligible and even a benefit since in this way small yet acceptable discrepancies of the model functions can be ignored and do not generate new expensive function

evaluations by updating the model function first. Thus, the less strict acceptance test can save function evaluations.

Results of numerical tests are presented in detail in Section 6.4.4. They illustrate that the stricter version of the acceptance test can cause more function evaluations than the less strict version. However, the numerical tests also show that in most of the test instances no difference occurred since the trial points provided a decrease for all objective functions.

Nevertheless, for some applications, e.g. the application described in Chapter 7, it can be useful to force a descent for each objective function f_i, $i \in \{1, 2, \ldots, q\}$. Otherwise, as illustrated there, it is possible that the trial point acceptance test defined by ρ_ϕ^k generates further iteration points although the current iteration point is a Pareto critical point. This can cause a higher amount of function evaluations for the expensive function. For using the strict version of the trial point acceptance test, steps 3 and 4 of MHT (Algorithm 4) need to be modified. The necessary modifications are stated in Algorithm 5. With this strict version of the trial point acceptance test the conver-

Algorithm 5 Modifications of MHT to use strict trial point acceptance test

Step 3: Trial point acceptance test (strict)

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

Otherwise, compute $f_i(x^{k+})$, $i = 1, 2, \ldots, q$, and

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

Set $\tilde{\rho}^k = \min_{i=1,\ldots,q} \rho_i^k$.

If $\tilde{\rho}^k \geq \eta_1$, set $x^{k+1} = x^{k+}$, otherwise set $x^{k+1} = x^k$.

Step 4: Trust region update

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

gence results for MHT from Section 4.5 cannot be transferred directly. Given Assumptions A.1 to A.10, all results from Section 4.5 up to Lemma 4.12 hold unchanged. However, the sufficient decrease condition for the function ϕ_m^k given in Assumption A.11 needs to be replaced by an analogous assumption for the model functions m_i^k for all $i \in \{1, 2, \ldots, q\}$.

This surrogate assumption for Assumption A.11 is stated in Assumption A.14. Given this sufficient decrease condition in addition to Assumptions A.1 to A.10, the remain-

ing results of Section 4.5 can be proved analogously. The surrogate Assumption A.14 is justified by Lemma 4.24 which is analogous to Lemma 4.13. Although the proof of Lemma 4.24 is mostly analogous to the proof of Lemma 4.13, we state it here to clarify the connections and differences to the standard version of MHT and its trial point acceptance test. Note that $(MOP^k_{m,\mathbb{R}^n})$ describes the optimization problem $\min_{x\in\mathbb{R}^n} m^k(x)$ with $k \in \mathbb{N}$ an iteration index of MHT.

Lemma 4.24 *Suppose Assumptions* A.1, A.3, A.4, A.7, A.9 *and* A.10 *hold. Let $k \in \mathbb{N}$ be an iteration of MHT and let $x^{k+} \in \mathbb{R}^n$ be the trial point computed in iteration k with the associated step size $t^{k+} \in \mathbb{R}$. Let $\beta^k_\phi = \max_{i=1,\dots,q} \left\|\nabla^2 m^k_i(x^k)\right\|_F + 1$ be. Then there exists a constant $\tilde{\kappa}_\phi \in (0,1)$ independent of k and for each $k \in \mathbb{N}$ an index $j = j(k) \in \mathbb{N}$ such that it holds*

$$m^k_i(x^k) - m^k_i(x^{k+}) \geq \left(\frac{1}{2}\right)^j \tilde{\kappa}\, \omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta^k_\phi}, \delta_k\right\} \text{ for all } i \in \{1,2,\dots,q\}.$$

Proof. Firstly, let x^k be not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$ and $(t^{k+}, x^{k+}) \in \mathbb{R}^{1+n}$ the minimal solution of (PS) given by $\min\left\{t \in \mathbb{R} \mid f(x^k) + t\, r^k - m^k(x) \in \mathbb{R}^q_+, x \in B_k\right\}$ computed in iteration k. Then according to Lemma 4.5 and Remark 4.1, it holds $t^{k+} \in [-1, 0)$ and $r^k_i = f_i(x^k) - p^k_i > 0$ for all $i \in \{1,2,\dots,q\}$. Due to the constraints of (PS) and Assumption A.4 it holds

$$m^k_i(x^k) - m^k_i(x^{k+}) \geq -t^{k+} r^k_i > 0 \tag{4.44}$$

for all $i \in \{1,2,\dots,q\}$. This implies

$$-t^{k+} = |t^{k+}| \leq \frac{m^k_i(x^k) - m^k_i(x^{k+})}{r^k_i} \tag{4.45}$$

for all $i \in \{1,2,\dots,q\}$. Let d_ω be given by $d_\omega \in \operatorname{argmin}_{\|d\|\leq 1} \max_{i=1,\dots,q} \nabla m^k_i(x^k)^\top d$. Then according to Lemma 2.12(ii), there exist scalars $\alpha_i \in [0,1], i \in \{1,2,\dots,q\}$, with $\sum_{i=1}^q \alpha_i = 1$ and $\mu \geq 0$ such that $\|d_\omega\| = 1$ and $d_\omega = -\mu \sum_{i=1}^q \alpha_i \nabla m^k_i(x^k)$ holds. For

the resulting function $g(x) = \sum_{i=1}^{q} \alpha_i m_i^k(x)$ and the corresponding point x_c defined by $x_c \in \operatorname{argmin}_{|t| \leq \delta_k} g(x^k + t\, d_\omega)$ Lemma 4.12 holds and therefore it holds

$$g(x^k) - g(x_c) \geq \frac{1}{2} \left\| \nabla g(x^k) \right\| \min \left\{ \frac{\left\| \nabla g(x^k) \right\|}{\beta_g^k}, \delta_k \right\} \geq 0. \tag{4.46}$$

Analogously to the proof of Lemma 4.13 it follows

$$g(x^k) - g(x_c) \geq \frac{1}{2} \left\| \nabla g(x^k) \right\| \min \left\{ \frac{\left\| \nabla g(x^k) \right\|}{\beta_\phi^k}, \delta_k \right\}. \tag{4.47}$$

Also analogously to the proof of Lemma 4.13, we obtain a feasible tuple $(\tilde{t}, \tilde{x}) \in \mathbb{R}^{1+n}$ for (PS) with $\tilde{x} := x^k + t\, d_\omega$, $|t| \leq \delta_k$, $m_i^k(x^k) - m_i^k(\tilde{x}) > 0$ for all $i \in \{1, 2, \ldots, q\}$ and

$$\tilde{t} := \max_{i=1,\ldots,q} -\frac{m_i^k(x^k) - m_i^k(\tilde{x})}{r_i^k} = -\min_{i=1,\ldots,q} \frac{m_i^k(x^k) - m_i^k(\tilde{x})}{r_i^k} < 0.$$

Due to t^{k+} being the minimal value of (PS) it holds again $|\tilde{t}| \leq |t^{k+}|$ which implies together with (4.45)

$$\frac{\min\limits_{j=1,\ldots,q} \left(m_j^k(x^k) - m_j^k(\tilde{x}) \right)}{\max\limits_{j=1,\ldots,q} r_j^k} \leq |\tilde{t}| \leq |t^{k+}| \leq \frac{m_i^k(x^k) - m_i^k(x^{k+})}{\min\limits_{j=1,\ldots,q} r_j^k} \tag{4.48}$$

for all $i \in \{1, 2, \ldots, q\}$. From Assumption A.10 it follows $\min_{j=1,\ldots,q} r_j^k / \max_{j=1,\ldots,q} r_j^k \geq \kappa_r$. Therefore, it follows

$$m_i^k(x^k) - m_i^k(x^{k+}) \geq \kappa_r \min_{j=1,\ldots,q} \left(m_j^k(x^k) - m_j^k(\tilde{x}) \right). \tag{4.49}$$

for all $i \in \{1, 2, \ldots, q\}$. Since it holds $g(x_c) \leq g(\tilde{x})$, $\sum_{i=1}^{q} \alpha_i = 1$ and $m_i^k(x^k) - m_i^k(\tilde{x}) > 0$ for all $i \in \{1, 2, \ldots, q\}$, it follows analogously to the proof of Lemma 4.13 due to the definition of g

$$g(x^k) - g(x_c) \geq g(x^k) - g(\tilde{x}) \geq \min_{i=1,\ldots,q} \left(m_i^k(x^k) - m_i^k(\tilde{x}) \right) > 0.$$

Now consider the inequality given in (4.47). Either it holds

$$\min_{i=1,\dots,q} \left(m_i^k(x^k) - m_i^k(\tilde{x})\right) \geq \frac{1}{2} \left\|\nabla g(x^k)\right\| \min\left\{\frac{\left\|\nabla g(x^k)\right\|}{\beta_\phi^k}, \delta_k\right\} \text{ or}$$

$$\min_{i=1,\dots,q} \left(m_i^k(x^k) - m_i^k(\tilde{x})\right) < \frac{1}{2} \left\|\nabla g(x^k)\right\| \min\left\{\frac{\left\|\nabla g(x^k)\right\|}{\beta_\phi^k}, \delta_k\right\}.$$

Thus, there exists an index $j = j(k) \in \mathbb{N}$ such that it holds

$$\min_{i=1,\dots,q} \left(m_i^k(x^k) - m_i^k(\tilde{x})\right) \geq \left(\frac{1}{2}\right)^j \left\|\nabla g(x^k)\right\| \min\left\{\frac{\left\|\nabla g(x^k)\right\|}{\beta_\phi^k}, \delta_k\right\}$$

and therefore it follows from (4.49) and the definition of g

$$m_i^k(x^k) - m_i^k(x^{k+}) \geq \kappa_r \left(\frac{1}{2}\right)^j \left\|\sum_{i=1}^q \alpha_i \nabla m_i^k(x^k)\right\| \min\left\{\frac{\left\|\sum_{i=1}^q \alpha_i \nabla m_i^k(x^k)\right\|}{\beta_\phi^k}, \delta_k\right\}$$

for all $i \in \{1, 2, \dots, q\}$ and for every iteration $k \in \mathbb{N}$ with x^k not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. Moreover, it holds according to Lemma 2.12 $\omega_m(x^k) \leq \left\|\sum_{i=1}^q \alpha_i \nabla m_i^k(x^k)\right\|$ and thus

$$m_i^k(x^k) - m_i^k(x^{k+}) \geq \kappa_r \left(\frac{1}{2}\right)^j \omega_m(x^k) \min\left\{\frac{\omega_m(x^k)}{\beta_\phi^k}, \delta_k\right\} \tag{4.50}$$

if x^k is not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$. If x^k is Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$, then it holds $\omega_m(x^k) = 0$. Due to x^{k+} being a solution of (PS) it holds $m_i^k(x^k) - m_i^k(x^{k+}) \geq 0$ for all $i \in \{1, 2, \dots, q\}$. Thus, (4.50) is also satisfied if x^k is not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$.

Then it follows analogously to the proof of Lemma 4.13 from Assumption A.9

$$m_i^k(x^k) - m_i^k(x^{k+}) \geq \tilde{\kappa} \left(\frac{1}{2}\right)^j \omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta_\phi^k}, \delta_k\right\}.$$

with $\tilde{\kappa} \coloneqq \kappa_r/(1+\kappa_\omega)^2 \in (0,1)$. □

Given this result, it is reasonable to impose Assumption A.14 which is the analogous formulation of Assumption A.11 for the functions m_i^k, $i = 1, 2, \ldots, q$.

Assumption A.14 There exists a constant $\kappa_\phi \in (0, 1)$ such that it holds for every iteration $k \in \mathbb{N}$ and for all $i \in \{1, 2, \ldots, q\}$

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

with $\beta_\phi^k = \max_{i=1,\ldots,q} \left\|\nabla^2 m_i^k(x^k)\right\|_F + 1$.

Replacing Assumption A.11 by Assumption A.14, the remaining results from Section 4.5, i.e. Remark 4.14 up to Theorem 4.20, can be transferred. For this purpose, some general aspects need to be regarded.
Whenever ρ_ϕ^k would be considered, the quotients ρ_i^k, $i = 1, 2, \ldots, q$, need to be considered instead respectively $\tilde{\rho}^k = \min_{i=1,\ldots,q} \rho_i^k$ as defined in Algorithm 5. Furthermore, the sets $\mathcal{S} = \left\{k \in \mathbb{N} \middle| \rho_\phi^k \geq \eta_1\right\}$ of successful iterations and $\mathcal{V} = \left\{k \in \mathbb{N} \mid \rho_\phi^k \geq \eta_2\right\} \subseteq \mathcal{S}$ of very successful iterations from (4.34) and (4.35) need to be redefined. Instead of $\mathcal{S}$ and $\mathcal{V}$, the sets

$$\tilde{\mathcal{S}} \coloneqq \left\{k \in \mathbb{N} \mid \rho_i^k \geq \eta_1 \text{ for all } i = 1, 2, \ldots, q\right\} \subseteq \mathcal{S} \text{ and}$$
$$\tilde{\mathcal{V}} \coloneqq \left\{k \in \mathbb{N} \mid \rho_i^k \geq \eta_2 \text{ for all } i = 1, 2, \ldots, q\right\} \subseteq \mathcal{V}$$

need to be considered. The relations $\tilde{\mathcal{S}} \subseteq \mathcal{S}$ and $\tilde{\mathcal{V}} \subseteq \mathcal{V}$ follow directly from Lemma 4.23. Consequently, it is possible that less iterations are successful, i.e. that in less iterations the trial point x^{k+} is accepted as next iteration point x^{k+1}.
The modified acceptance test also influences the trust region update in step 4 of MHT, see Algorithm 5. Since it is possible that less iterations are classified as successful, it is possible that the radius of the trust region will be reduced more often. However, this is not apparent in the theoretical results from Section 4.5 and does not need to be considered explicitly.
With the modifications of the assumptions as explained above, Remark 4.14 still holds. Lemma 4.15 is not required any more, since the validity of the model functions m_i^k, $i = 1, 2, \ldots, q$, given in Lemma 4.8 and Assumption A.8 can be applied directly to the individual functions. This needs to be considered for the proof of Lemma 4.16 for

which also ρ_ϕ^k needs to be replaced by ρ_i^k, $i = 1, 2, \ldots, q$. Lemmas 4.17 and 4.18 can be proved analogously, only for the latter $\mathcal{S}$ needs to be replaced by $\tilde{\mathcal{S}}$.
For Lemma 4.19 and the main result Theorem 4.20 $\mathcal{S}$ is replaced by $\tilde{\mathcal{S}}$, ρ_ϕ^k is replaced by ρ_i^k, $i = 1, 2, \ldots, q$, but the auxiliary function ϕ is still considered to transfer the proofs. Due to replacing Assumption A.11 by Assumption A.14 some estimations need to be adapted. Let $k \in \tilde{\mathcal{S}}$ hold and $i, j \in \{1, 2, \ldots, q\}$ be the indices such that it holds $\phi(x^k) = f_i(x^k)$ and $\phi(x^{k+}) = f_j(x^{k+})$. Due to the definition of the function ϕ it holds

$$\phi(x^k) - \phi(x^{k+}) = f_i(x^k) - f_j(x^{k+}) \geq f_j(x^k) - f_j(x^{k+})$$

and due to $k \in \tilde{\mathcal{S}}$, i.e. $\rho_i^k \geq \eta_1$ for all $i \in \{1, 2, \ldots, q\}$, it follows

$$\phi(x^k) - \phi(x^{k+}) \geq f_j(x^k) - f_j(x^{k+}) \geq \eta_1 \left(m_j^k(x^k) - m_j^k(x^{k+}) \right).$$

Now the lower bound given by Assumption A.14 can be applied. Thus, whenever a successful iteration is considered and the lower bound from Assumption A.11 would be applied to obtain a lower bound for the difference $\phi(x^k) - \phi(x^{k+})$, the above estimations are inserted and the lower bound given by Assumption A.14 can be applied. The rest of the proofs of Lemma 4.19 and Theorem 4.20 is analogous. Consequently, the main result about convergence given in Theorem 4.20 also holds for MHT with the strict version of the trial point acceptance test.

4.6.4 Exact Model Functions for the Cheap Functions

The aim of this thesis is to develop a solution method for heterogeneous problems and the heterogeneity can be exploited further by using the cheap functions itselves instead of surrogate models. We will outline in this subsection that the results from Section 4.5 transfer if Assumption A.15 holds.

Assumption A.15 Let the model functions of the cheap functions be exact, that is it holds $m_i^k \equiv f_i$ for all $i \in \{2, 3, \ldots, q\}$ and for all $k \in \mathbb{N}$.

Assumption A.3 states that all model functions m_i^k, $i = 1, 2, \ldots, q$, are quadratic functions. Quadratic models are most commonly used in the trust region framework since

then the trust region subproblems of minimizing the model functions in the trust region can be solved easily, see also Section 3.1. By replacing a quadratic model function by the possibly non-quadratic original function solving the trust region subproblem becomes more difficult. However, in comparison with the numerical effort associated with the expensive function we assume the additional numerical effort caused by this modification to be acceptable.
Moreover, for the theoretical results it is sufficient to assume that the model functions are twice continuously differentiable. In case of exact model functions for the cheap functions, this is already given in Assumption A.1. Thus, to transfer the results from Section 4.5, Assumption A.3 needs to be replaced by Assumption A.16

Assumption A.16 The model function $m_1^k : \mathbb{R}^n \to \mathbb{R}$ is quadratic for all $k \in \mathbb{N}$.

Given Assumption A.15, both Assumptions A.4 and A.7 for $i = 2, 3, \ldots, q$ and Assumption A.5 and Lemma 4.8 from Section 4.5 are trivially fulfilled. Considering all other results from Section 4.5, only at one point an important change is required. An essential part of the convergence considerations are the numbers

$$\beta_i^k = \left\|\nabla^2 m_i^k(x^k)\right\|_F + 1, i \in \{1, 2, \ldots, q\}$$

defined in Lemma 4.11 and the related numbers $\beta_g^k = \left\|\nabla^2 g(x^k)\right\|_F + 1$ from Lemma 4.12 and $\beta_\phi^k = \max_{i=1,\ldots,q} \left\|\nabla^2 m_i^k(x^k)\right\|_F + 1$ from Lemma 4.13. These numbers are defined in the context of Taylor approximations of second order which are exact for quadratic functions. When replacing the quadratic model functions m_i^k, $i = 2, 3, \ldots, q$, by the possibly non-quadratic original functions f_i, a suitable intermediate point is required when applying Taylor's theorem. For this purpose and without loss of generality, let $x \in \mathbb{R}^n$ be with $x \leq x^k$. Then it holds

$$f_i(x) = f_i(x^k) + \nabla f_i(x^k)^\top (x - x^k) + \frac{1}{2}(x - x^k)^\top \nabla^2 f_i(\xi^k)(x - x^k)$$

with $\xi^k \in [x, x^k] = \left\{x + t(x^k - x) \mid t \in [0, 1]\right\}$. This also needs to be considered when replacing the function $g(x) = \sum_{i=1}^q \alpha_i m_i^k(x)$ with $\alpha_i \geq 0$ for all $i \in \{1, 2, \ldots, q\}$

from Lemma 4.12 by a Taylor approximation of second order. Again, without loss of generality let $x \in \mathbb{R}^n$ with $x \leq x^k$ hold. Analogously it follows

$$g(x) = g(x^k) + \nabla g(x^k)^\top (x - x^k) + \frac{1}{2}(x - x^k)^\top \nabla^2 g(\zeta^k)(x - x^k)$$

with a suitable intermediate point $\zeta^k \in [x, x^k]$. Regarding these changes of the Taylor approximations of second order in Lemmas 4.11 to 4.13, the numbers β_i^k, $i = 2, 3, \ldots, q$, β_g^k and β_ϕ^k need to be redefined in the following way:

- $\beta_i^k := \max\limits_{x \in B_k} \|\nabla^2 f_i(x)\|_F + 1$ for $i \in \{2, 3, \ldots, q\}$ (first defined in Lemma 4.11)
- $\beta_g^k := \max\limits_{x \in B_k} \|\nabla^2 g(x)\|_F + 1$ (first defined in Lemma 4.12)
- $\beta_\phi^k := \max\limits_{i=1,\ldots,q} \max\limits_{x \in B_k} \left\|\nabla^2 m_i^k(x)\right\|_F + 1$ (first defined in Lemma 4.13)

In the single-objective version of trust region algorithms [CGT00] analogous definitions are used standardly. According to Assumptions A.6 and A.7, the Hessians of the functions $f_2, f_3, \ldots, f_q$ and the model function m_1^k, $k \in \mathbb{N}$, are uniformly bounded and therefore the constants defined above are well defined. The estimations based on the Taylor approximations of second order in Lemmas 4.11 to 4.13 then all follow the scheme

$$\left\|\nabla^2 f_i(\xi^k)\right\|_F + 1 \leq \max_{x \in B_k} \left\|\nabla^2 f_i(x)\right\|_F + 1 = \beta_i^k,$$

$i \in \{1, 2, \ldots, q\}$, analogously for the functions g and m_i^k, $i = 1, 2, \ldots, q$, with an intermediate point ζ^k. Thus, Lemmas 4.11 to 4.13 hold with these redefined numbers. Furthermore, Remark 4.14 also holds for the new definition of β_ϕ^k. All remaining results given by Lemmas 4.15 to 4.19 and Theorem 4.20 from Section 4.5 then follow analogously.

Thus, using Assumptions A.1, A.2, A.4 to A.10 and A.15, replacing Assumption A.3 by Assumption A.16 and redefining the numbers β_i^k, $i \in \{2, 3, \ldots, q\}$, β_g^k and β_ϕ^k as described above, the main result about convergence of MHT given in Theorem 4.20 holds. Moreover, the function values for the cheap functions are non-increasing in all iterations as explained in Remark 4.25.

Remark 4.25 Suppose Assumption A.15 holds. In every iteration $k \in \mathbb{N}$ of MHT the subsequent iteration point x^{k+1} is either defined as x^k or as the trial point x^{k+} resulting from the Pascoletti-Serafini problem (PS). In the latter case, Assumption A.15 and Lemma 4.4 from Section 4.3 imply $f_i(x^{k+}) \leq f_i(x^k)$ for all $i \in \{2, 3, \ldots, q\}$. Thus, it holds $f_i(x^{k+1}) \leq f_i(x^k)$ for all $i \in \{2, 3, \ldots, q\}$ and for all $k \in \mathbb{N}$.

Using exact model functions for the cheap objective function, it is possible that less iterations and therefore less function evaluations for f_1 are required in MHT. However, first numerical tests show only a small difference in the number of function evaluations. They are presented and discussed in Section 6.4.5.

4.7 Box Constrained Optimization Problems

The algorithm MHT presented in this thesis is formulated for unconstrained multi-objective optimization problems. It is not suitable for multi-objective optimization problems with equality or inequality constraints. However, box constraints can be handled by the algorithm. Though, the convergence analysis from Section 4.5 is not transferable since it is adjusted to the framework of unconstrained problems. In this section we consider box constrained optimization problem given by

$$\min_{x \in \Omega} \left(f_1(x), f_2(x), \ldots, f_q(x)\right)^\top \tag{MOP_b}$$

with the constraint set $\Omega \coloneqq \{x \in \mathbb{R}^n \mid lb \leq x \leq ub\}$ and the lower and upper bounds $lb \in (\mathbb{R} \cup \{-\infty\})^n$, $ub \in (\mathbb{R} \cup \{\infty\})^n$, $lb \leq ub$. The box constraints need to be regarded in two parts of MHT: the computation of the ideal point and the Pascoletti-Serafini problem. Let $k \in \mathbb{N}$ be an iteration of MHT. The ideal point p^k is modified to

$$\tilde{p}_i^k = \min_{x \in B_k \cap \Omega} m_i^k(x) \tag{4.51}$$

for all $i \in \{1, 2, \ldots, q\}$ and the Pascoletti-Serafini problem (PS) is changed to

$$\min \left\{t \in \mathbb{R} \mid f(x^k) + t\,\tilde{r}^k - m^k(x) \in \mathbb{R}^q_+, x \in B_k \cap \Omega\right\} \tag{$\widetilde{PS}_2$}$$

with $\tilde{r}^k = f(x^k) - \tilde{p}^k$. Integrating the box constraints into the Pascoletti-Serafini problem ensures feasibility for all iteration points x^k, $k \in \mathbb{N}$. (PS) – without the modifications of above – is a nonlinear optimization problem and adding the box constraints contained in Ω does not further complicate solving this problem.
In the single-objective trust region approach, the trust region subproblem is minimizing a quadratic function in the trust region. As outlined in Section 3.1, there are many efficient approaches in literature to solve these trust region subproblems. If box constraints are integrated, the subproblem is more difficult and these methods cannot be applied any more. This is the reason why in standard trust region approaches projection methods are used to integrate box constraints, see for example [CGT88; CGT00]. The idea of projection methods is to firstly neglect the box constraints when solving the subproblem. If the obtained point is violating any of the box constraints it is projected back to the feasible set, mostly to $\partial\Omega$. For further trust region approaches for box constrained optimization problems see for example [FMS94; CL96].
The computation of the ideal point in the standard version of MHT, i.e. $\min_{x \in B_k} m_i^k$, $i \in \{1, 2, \ldots, q\}$, $k \in \mathbb{N}$, without further constraints, is such a standard trust region subproblem. However, for the implementation we did not use projection methods for this purpose, but integrated the box constraints directly as given in (4.51). Since projection methods are not reasonable for solving $(\widetilde{PS}_2)$, we integrated them directly both in $(\widetilde{PS}_2)$ and in the computation of the ideal point. Moreover, the additional numerical effort for computing $\tilde{p}^k$ by (4.51) is assumed to be negligible in comparison with the numerical effort associated with the expensive function f_1.
Thus, box constraints do not pose a large algorithmic difficulty in the context of heterogeneous optimization problems as considered in this thesis and can be handled by the algorithm MHT. However, the convergence results from Section 4.5 do not transfer. Nevertheless, some general properties of $(\widetilde{PS}_2)$ can be proved analogous to the properties of (PS) and a connection to the unconstrained case and Pareto critical points can be made. For this purpose, we define the optimization problem

$$\min_{x \in B_k \cap \Omega} m^k(x) \qquad (MOP^k_{m,b})$$

The following general properties hold for $(\widetilde{PS}_2)$ and are analogous to Lemmas 4.4 and 4.5. The proofs are also analogous when integrating the box constraints by replacing B_k by $B_k \cap \Omega$ and by considering $(\widetilde{PS}_2)$ and $\tilde{p}^k$.

Lemma 4.26 *Let $(\bar{t}, \bar{x})$ be a minimal solution of of $(\widetilde{PS}_2)$.*

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

(ii) If x^k is not weakly efficient for $(MOP^k_{m,b})$, then it holds $\bar{t} \in [-1, 0)$.

Analogous to Lemma 4.6 and according to [Eic08, Th. 2.1], the following general properties hold for $(\widetilde{PS}_2)$.

Lemma 4.27

(i) If $(\bar{t}, \bar{x}) \in \mathbb{R}^{1+n}$ is a minimal solution of $(\widetilde{PS}_2)$, then $\bar{x}$ is weakly efficient for $(MOP^k_{m,b})$.

(ii) If $x^k \in \mathbb{R}^n$ is weakly efficient for $(MOP^k_{m,b})$ and $\tilde{r}^k \in \mathbb{R}^q_{++}$, then $(0, x^k)$ is a minimal solution of $(\widetilde{PS}_2)$.

(iii) If $x^k \in \mathbb{R}^n$ is efficient for $(MOP^k_{m,b})$ and $\tilde{r}^k \neq 0_q$, then $(0, x^k)$ is a minimal solution of $(\widetilde{PS}_2)$.

In [FV16; GPD18] the concept of Pareto criticality is extended to constrained multi-objective optimization problems with equality and inequality constraints. Following these approaches, we transfer the concept of Pareto criticality to box constrained optimization problems. Furthermore, we give a short overview of which theoretical results hold for the modified version of MHT and why the convergence analysis from Section 4.5 does not transfer.

In the unconstrained case, a vector $\bar{x} \in \mathbb{R}^n$ is called Pareto critical for (MOP) if for all vectors $d \in \mathbb{R}^n$ there exits an index $i \in \{1, 2, \ldots, q\}$ such that it holds $\nabla f_i(\bar{x})^\top d \geq 0$, see Definition 2.7 in Section 2.2. According to Lemma 2.9, it is a necessary condition for local weak efficiency.

To transfer this notion to constrained optimization problems, the functions describing the constraints are assumed to be continuously differentiable. This holds for box constraints and following the definition of Pareto criticality from [FV16; GPD18], we formulate it for box constrained optimization problems in Definition 4.28.

To avoid confusion due to the same terminology, we will refer to Pareto critical points for box constrained problems as stationary points. Stationary points as defined in Definition 4.28 also fulfill a necessary condition of first order for optimality; this is stated

in Lemma 4.29. The proof can be found for example in [GPD18, Lem. 3.8] and is analogous to the proof of Lemma 2.9 in Section 2.2 which gives the statement in the unconstrained case. Nevertheless, we give a short proof to illustrate the connection to the unconstrained case.

Definition 4.28 Let $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, 2, \ldots, q$, be continuously differentiable functions. Let $\bar{x} \in \Omega$ be a feasible vector for the box constrained optimization problem (MOP_b) and define the set

$$L(\bar{x}) := \{d \in \mathbb{R}^n \mid d_i \geq 0, \text{ if } \bar{x}_i = lb_i, d_i \leq 0, \text{ if } \bar{x}_i = ub_i, i = 1, 2, \ldots, q\}. \quad (4.52)$$

The vector $\bar{x}$ is called a *stationary point* for (MOP_b) if for all $d \in L(\bar{x})$ there exists an index $i \in \{1, 2, \ldots, q\}$ such that it holds

$$\nabla f_i(\bar{x})^\top d \geq 0.$$

Lemma 4.29 *Let $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, 2, \ldots, q$, be continuously differentiable functions. If $\bar{x} \in \Omega$ is locally weakly efficient for (MOP_b), it is a stationary point for (MOP_b).*

Proof. Suppose the vector $\bar{x} \in \Omega$ is locally weakly efficient for (MOP_b), but not a stationary point. Then there exists a neighborhood $U(\bar{x})$ of $\bar{x}$ such that $\bar{x}$ is weakly efficient for $\min\{f(x) \mid x \in \Omega \cap U(\bar{x})\}$. Since $\bar{x}$ is not a stationary point for (MOP_b), there exists a vector $d \in L(\bar{x})$ such that it holds $\nabla f_i(\bar{x})^\top d < 0$ for all $i \in \{1, 2, \ldots, q\}$. This implies

$$0 > \nabla f_i(\bar{x})^\top d = \lim_{t \to 0} \frac{f_i(\bar{x} + td) - f_i(\bar{x})}{t}$$

for all $i \in \{1, 2, \ldots, q\}$. Consequently, there exists a scalar $t_0 > 0$ such that it holds

$$f_i(\bar{x} + td) - f_i(\bar{x}) < 0$$

for all $t \in (0, t_0]$ and for all $i \in \{1, 2, \ldots, q\}$. If $\bar{x}_i = lb_i$ holds for an index $i \in \{1, 2, \ldots, q\}$, then $d \in L(\bar{x})$ implies $d_i \geq 0$. If $\bar{x}_i = ub_i$ holds for an index $i \in \{1, 2, \ldots, q\}$, then $d \in L(\bar{x})$ implies $d_i \leq 0$. Thus, there exists a scalar $t_1 \in (0, t_0]$ such that it holds $y := \bar{x} + t_1 d \in U(\bar{x}) \cap \Omega$. Furthermore, it holds $f_i(y) < f_i(\bar{x})$ for all $i \in \{1, 2, \ldots, q\}$ which contradicts $\bar{x}$ being locally weakly efficient for (MOP_b). □

From the previous results it follows that if the current iteration point x^k of MHT is not a stationary point for $\min_{x\in\Omega} m^k(x)$, then in step 2 of MHT a trial point $x^{k+} \neq x^k$ is computed. This is stated in Remark 4.30. Whether x^{k+} is accepted as next iteration point depends on the accuracy of the model functions.

Remark 4.30 Let $k \in \mathbb{N}$ be an iteration of MHT. If x^k is not a stationary point for $\min_{x\in\Omega} m^k(x)$, it follows from Lemma 4.29 that x^k is not locally weakly efficient for $\min_{x\in\Omega} m^k(x)$. Since it holds $x^k \in B_k$, x^k is also not locally weakly efficient for $(MOP^k_{m,b})$.
Let $(\bar{t}, \bar{x})$ be a minimal solution of $(\widetilde{PS}_2)$. Then according to Lemma 4.26 it holds $\bar{t} \in [-1, 0)$. This implies that $(0, x^k)$ is not a minimal solution of $(\widetilde{PS}_2)$. Since the trial point x^{k+} results from a minimal solution of $(\widetilde{PS}_2)$, it follows $x^{k+} \neq x^k$.

The notion of Pareto criticality for box constrained optimization problems is connected to Pareto criticality in the unconstrained case. This is stated in Lemma 4.31.

Lemma 4.31 *Let $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, 2, \ldots, q$, be continuously differentiable functions. Consider the unconstrained multi-objective optimization problem (MOP) and the box constrained problem (MOP_b). If a vector $\bar{x} \in \Omega$ is Pareto critical for (MOP), then it is a stationary point for (MOP_b).*

Proof. The statement follows directly from $L(\bar{x}) \subseteq \mathbb{R}^n$. □

As described in Section 2.2, the auxiliary function $\omega : \mathbb{R}^n \to \mathbb{R}$ defined by

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

see Definition 2.10, characterizes Pareto criticality for the unconstrained optimization problem (MOP). As stated in Lemma 2.11, a vector $\bar{x} \in \mathbb{R}^n$ is Pareto critical for (MOP) if and only if it holds $\omega(\bar{x}) = 0$. Therefore, Lemma 4.31 implies that whenever a vector $\bar{x}$ is feasible for (MOP_b) and it holds $\omega(\bar{x}) = 0$, then $\bar{x}$ is a stationary point for (MOP_b).
According to [GPD18] which is based on [FS00], stationary points of (MOP_b) can be characterized in a similar way by considering the auxiliary optimization problem

$$\alpha(x) \coloneqq \min_{d\in L(x), \|d\|\leq 1} \max_{i=1,\ldots,q} \nabla f_i(x)^\top d + \frac{1}{2}\|d\|^2.$$

According to [GPD18, Rem.3.14] a vector $x \in \Omega$ is a stationary point for (MOP_b) if and only if it holds $\alpha(x) = 0$. If x is not a stationary point for (MOP_b), then it holds $\alpha(x) < 0$. Due to the active set strategy the solution d and the optimal value $\alpha(x)$ depend in general not continuously on x. A function analogous to ω, i.e. α without the additional summand, can be defined by

$$\omega_b(x) := -\min_{d \in L(x), \|d\| \leq 1} \max_{i=1,\dots,q} \nabla f_i(x)^\top d. \tag{4.53}$$

By this, stationary points can also be characterized as given in Lemma 4.32. However, analogous to the auxiliary optimization problem from [GPD18], ω_b does in general not depend continuously on x.

Lemma 4.32 *Let $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, 2, \dots, q$, be continuously differentiable functions. Let $\omega_b(x)$ be defined as in* (4.53)*. The following statements hold.*

(i) It holds $\omega_b(x) \geq 0$ for all $x \in \Omega$.

(ii) A vector $x \in \Omega$ is a stationary point for (MOP_b) if and only if it holds $\omega_b(x) = 0$.

Proof. Let $x \in \Omega$ be an arbitrary point and let

$$\omega_b(x) = -\min_{d \in L(x), \|d\| \leq 1} \max_{i=1,\dots,q} \nabla f_i(x)^\top d = -\max_{i=1,\dots,q} \nabla f_i(x)^\top \bar{d}$$

with $\bar{d} \in L(x) = \{d \in \mathbb{R}^n \mid d_i \geq 0, \text{ if } x_i = lb_i, d_i \leq 0, \text{ if } x_i = ub_i, i = 1, 2, \dots, q\}$. Since it holds $0_n \in L(x)$ for all $x \in \Omega$, it follows $\omega_b(x) \geq 0$ and statement (i) holds. To prove statement (ii), let firstly $x \in \Omega$ be a stationary point for (MOP_b). Then there exists an index $j \in \{1, 2, \dots, q\}$ such that it holds $\nabla f_j(x)^\top \bar{d} \geq 0$. This implies $\max_{i=1,\dots,q} \nabla f_i(x)^\top \bar{d} \geq 0$ and $\omega_b(x) \leq 0$. Since it holds $\omega_b(x) \geq 0$ for all $x \in \Omega$, it follows $\omega_b(x) = 0$.
Now let x be not a stationary point for (MOP_b). Then there exists a vector $d^1 \in L(x)$ such that it holds $\nabla f_i(x)^\top d^1 < 0$ for all indices $i \in \{1, 2, \dots, q\}$. Consider the vector $d^2 := d^1 / \|d^1\|$. Since it holds $d^1 \in L(x)$, it follows $d^2 \in L(x)$. Furthermore, it holds $\nabla f_i(x)^\top d^2 < 0$ for all $i \in \{1, 2, \dots, q\}$. This implies

$$\omega_b(x) = -\min_{d \in L(x), \|d\| \leq 1} \max_{i=1,\dots,q} \nabla f_i(x)^\top d \geq -\max_{i=1,\dots,q} \nabla f_i(x)^\top d^2 > 0$$

and statement (ii) follows. □

Remark 4.33 From the definition of ω and ω_b, the fact $L(x) \subseteq \mathbb{R}^n$ and from Lemma 4.31 it follows directly $\omega(x) \geq \omega_b(x)$ for all $x \in \Omega$.

Lemma 4.32 and Remark 4.33 imply that if the algorithm MHT including the box constraints of Ω produces a sequence of iteration points $\left\{x^k\right\}_k$ with $\omega(x^k)$ converging to zero, then $\omega_b(x^k)$ also converges to zero. However, in case of box constraints it is not guaranteed that MHT produces a sequence of iterates with $\omega(x^k)$ converging to zero. In the convergence analysis of MHT in Section 4.5 uniform continuity of ω is required for the main results of Lemma 4.19 and Theorem 4.20. Since ω_b is in general not continuous, it is also in general not uniformly continuous. Thus, the main result of the convergence analysis for MHT given by Theorem 4.20 cannot be transferred to box constrained optimization problems.

5 Heuristic Approaches Based on MHT to Generate Several Pareto Critical Points

The algorithm MHT is designed to compute one Pareto critical point for the unconstrained multi-objective optimization problem (MOP). This is reasonable since for heterogeneous optimization problems with expensive functions the aim cannot be to approximate the whole set of efficient points. Due to the high numerical effort associated with the expensive function this is not realizable. However, the aim can be - after obtaining one Pareto critical point - to explore the area around it further and to compute further Pareto critical points. In some applications this is of interest, for example because the obtained point is not preferred by the user or it is of interest to invest some more time to gain further insight into the optimization problem.
One option is to start MHT with a different starting point. The numerical tests in Section 6.4 show that in general MHT computes different points if different starting points are used. Alternative approaches that make more use of the heterogeneity of the objective functions are presented in this chapter. They are heuristic approaches based on MHT. Since the approaches are motivated by ideas for bi-objective optimization problems, we consider in the following optimization problems of the form

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

with $f : \mathbb{R}^n \to \mathbb{R}^2$, i.e. ($MOP$) with $q = 2$. We also shortly discuss 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 given in Section 4.5 can be transferred.
In Sections 5.1 and 5.2 two heuristic approaches are presented that start after one run of MHT. They 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

can be realized with reasonable numerical effort. These two methods are combined in Section 5.3 and form a heuristic approach to approximate the set of Pareto critical points.
For that purpose we impose Assumption A.17. Unlike the standard version of MHT where the function $\phi : \mathbb{R}^n \to \mathbb{R}$ 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 A.17 Let the cheap objective function $f_2 : \mathbb{R}^n \to \mathbb{R}$ of (BOP) be bounded from below and let $\min_{x\in\mathbb{R}^n} f_2(x) = \inf_{x\in\mathbb{R}^n} f_2(x)$.

5.1 Spreading via Individual Minima

The idea of the approach presented in this section is to make more use of the cheap function f_2 and the fact that individual minima can be computed with low numerical effort. The approach is heuristically motivated and the numerical results presented in Section 6.5.1 confirm the usefulness. It starts after one initial run of MHT with the thereby generated point $\bar{x} \in \mathbb{R}^n$. According to Theorem 4.20, this point is Pareto critical for (BOP). Starting from this point, further Pareto critical points are computed spreading in a certain direction.
The overall search direction is given by the global individual minimum of the cheap function f_2 which is assumed to exist according to Assumption A.17. 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

$$x_{glob} \in \operatorname*{argmin}_{x\in\mathbb{R}^n} f_2(x). \tag{5.1}$$

It follows immediately from the definition of weak efficiency, see Definition 2.1, 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. Solving single-objective minimization problems - locally or globally - has its own difficulties, but shall not be the topic of this thesis. Different approaches can be found in the literature, see for example [HP95; HT96; Ber99; JM04]. However, compared with the numerical

effort associated with evaluating the expensive function f_1 we assume the numerical effort to solve any minimization problem for the analytically given function f_2, either unconstrained or in a certain area, to be acceptable. We consider a closed ball around $\bar{x}$ with radius $\delta > 0$ defined by

$$B_{loc} := B\left(\bar{x}, \bar{\delta}\right) = \left\{x \in \mathbb{R}^n \mid \|\bar{x} - x\| \leq \bar{\delta}\right\}. \tag{5.2}$$

An individual minimum of function f_2 in this area is then defined by

$$x_{loc} \in \underset{x \in B_{loc}}{\operatorname{argmin}} f_2(x). \tag{5.3}$$

No general statement can be made about weak efficiency for such a point. However, since it holds $\bar{x} \in B_{loc}$ for the Pareto critical point $\bar{x}$, x_{loc} is expected to be close to the set of Pareto critical points if $\bar{\delta}$ is not chosen too large.
This is illustrated schematically in the image space in Figure 5.1. Let $\bar{x}^1, \bar{x}^2 \in \mathbb{R}^n$ be the resulting points from two different runs of MHT and B^1_{loc}, B^2_{loc} the according local areas as in (5.2). The images of these areas are depicted as gray shaded areas. Both figures show the images of the global individual minimum x_{glob} of the function f_2. The left side of Figure 5.1 shows the image of an individual minimum x^1_{loc} of f_2 in B^1_{loc} which is efficient for (BOP), whereas the right side of Figure 5.1 shows the image of an individual minimum x^2_{loc} of f_2 in the area B^2_{loc} which is not efficient for (BOP) and may also not be Pareto critical for (BOP).

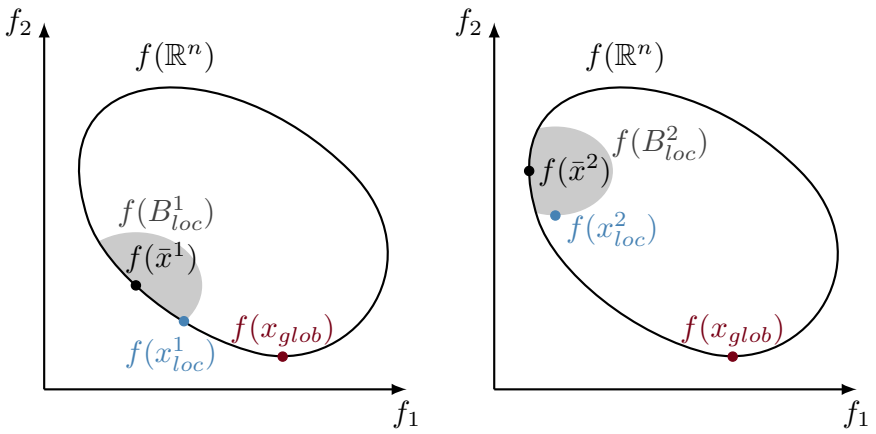

Figure 5.1 – Individual minima of cheap function f_2

Although no general statement about the connection of x_{loc} and Pareto critical or weakly efficient points can be made, Lemma 5.1 states some special cases. They will be used for formulating an algorithm later in this section.

Lemma 5.1 *Let $\bar{x} \in \mathbb{R}^n$ be the result of the initial run of MHT. Furthermore, let x_{glob}, B_{loc} and x_{loc} be defined as in* (5.1)–(5.3).

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

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

Proof. To prove statement (i), let $x_{loc} \in \text{int}B_{loc}$. Then there exists a neighborhood $U(x_{loc}) \subseteq B_{loc}$ of x_{loc} and it holds $f_2(x_{loc}) \leq f_2(x)$ 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 $i = 1, 2$. This implies that x_{loc} is locally weakly efficient for (BOP).
Statement (ii) follows directly from the definition of x_{glob} and the definition of weak efficiency.

□

The idea of the approach in this section is to minimize the cheap objective function f_2 on moving local areas defined analogously to B_{loc} and generate good starting points for further runs of MHT. This spreading approach is illustrated schematically in Figure 5.2 and formulated as algorithm in Algorithm 6.
Let $\bar{x}^1 = \bar{x} \in \mathbb{R}^n$ be the result of the initial run of MHT and therefore Pareto critical for (BOP). For the illustration in Figure 5.2 it is depicted as an efficient point. As explained above, in the first step of the spreading approach the cheap function f_2 is minimized in B^1_{loc} with

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

see (5.2). If the resulting point $x^1_{loc} \in \text{argmin}\{f_2(x) \mid x \in B^1_{loc}\}$ lies on the boundary of B^1_{loc} and if it holds $f_2(x^1_{loc}) \neq f_2(x_{glob})$, 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), see Theorem 4.20.

Otherwise, that is it holds $x^1_{loc} \in \text{int}B^1_{loc}$ or $f_2(x^1_{loc}) = f_2(x_{glob})$, the point x^1_{loc} is locally weakly efficient for (BOP), see Lemma 5.1, and we set $\bar{x}^2 = x^1_{loc}$. Thus, in both cases a Pareto critical point $\bar{x}^2$ is obtained. The left figure of Figure 5.2 illustrates the case that $f(x^1_{loc})$ can be a nondominated point. The same strategy can be applied again, this time to $\bar{x}^2$ by minimizing f_2 in B^2_{loc}, the closed ball around $\bar{x}^2$ with radius $\bar{\delta}$ given by

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

Again, the resulting point $x^2_{loc} \in \operatorname{argmin}\{f_2(x) \mid x \in B^2_{loc}\}$ can be a locally weakly efficient point or a good starting point for another run of MHT. The latter case is depicted in the right figure of Figure 5.2. The image of the point $\bar{x}^3 \in \mathbb{R}^n$ obtained by MHT is marked red.

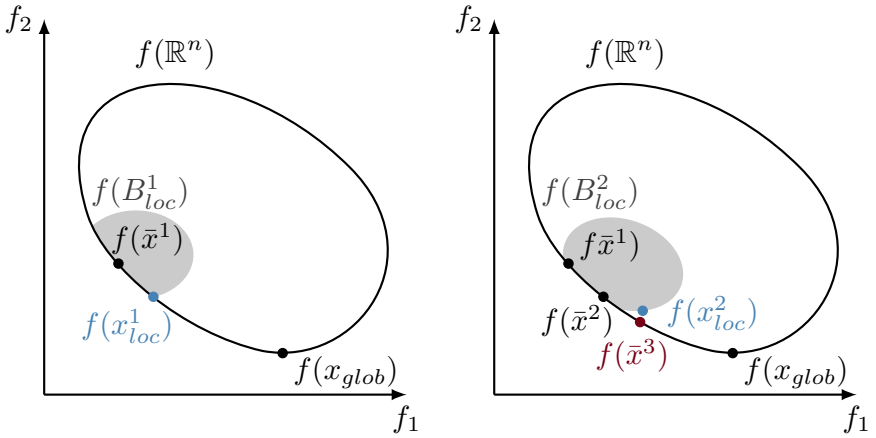

Figure 5.2 – Illustration of spreading approach using individual minima of f_2

This spreading strategy can be iterated; the approach is described in Algorithm 6 and referred to as $\text{MHT}_{\text{spread}}$. 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^k_{loc}, $k \in \mathbb{N}$, in which the individual minima of f_2 are computed.

Remark 5.2 Let $k \in \mathbb{N}$ be an iteration of $\text{MHT}_{\text{spread}}$. 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 is increased. Furthermore, the subproblems $\min_{x \in B^j_{loc}} f_2(x)$

Algorithm 6 $\text{MHT}_{\text{spread}}$: 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

need to be solved by a global method for every $j \geq k + 1$. As already mentioned, we assume this 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 $\text{MHT}_{\text{spread}}$ (Algorithm 6) is iterated until $f_2(y) = f_2(x_{glob})$ holds for a point y obtained during the spreading procedure, i.e. until the global minimum of f_2 is reached. It is obtained within a finite number of iterations in the cases described in Remark 5.3.
It is also possible to determine a maximum number of function evaluations and include this as a stopping criterion.

Remark 5.3 Let $k \in \mathbb{N}$ be an iteration of $\text{MHT}_{\text{spread}}$. According to Remark 5.2, an infinite loop with $x^k = x^{k+j}$ for $j \geq k$ is not possible. Hence, $\text{MHT}_{\text{spread}}$ is guaranteed to terminate after finitely many iterations if it holds

$$f_2(x^{i+1}) \leq f_2(x^i) \text{ for all } i = 1, 2, 3 \ldots, \tag{5.4}$$

i.e. if the distance to the value $f_2(x_{glob})$ decreases in every iteration. In $\text{MHT}_{\text{spread}}$ as it is formulated in Algorithm 6 the general version of MHT is used. The trial point acceptance test does not guarantee that the obtained point provides a decrease for both f_1 and f_2 compared to the starting point, see Section 4.4. Thus, (5.4) is not guaranteed to hold.
By using the strict version of the acceptance test as described in Section 4.6.3 or setting $m_2^k \equiv f_2$ for all iterations $k \in \mathbb{N}$ in MHT, see Section 4.6.4, (5.4) holds. As outlined in these sections, the convergence results of MHT also hold for these modifications.

The aim in this chapter is to make more use of the heterogeneity of the objective functions. Therefore, $\text{MHT}_{\text{spread}}$ was implemented for the numerical tests with the modification of MHT from Section 4.6.4, i.e. $m_2^k \equiv f_2$ is set in all iterations $k \in \mathbb{N}$.
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 4.2. Whenever the model function m_1^k needs to be updated and new interpolation points are computed by the subroutine given in Algorithm 3 in Section 3.2.2, previously evaluated points are reused if possible. 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.
It is stated in Lemma 5.4 that all points generated by the heuristic search method $\text{MHT}_{\text{spread}}$ are Pareto critical for (BOP).

Lemma 5.4 *Suppose Assumptions* A.1 *to* A.11 *hold. Then every point of the set X produced by $\text{MHT}_{\text{spread}}$ (Algorithm* 6*) is Pareto critical for* (BOP).

Proof. Let $X = \{x^0, x^1, \ldots, x^p\}$ be the output of $\text{MHT}_{\text{spread}}$ with $p \geq 1$ and consider $x^k \in X$ with $k \in \{0, 1, \ldots, p\}$. Then three cases can occur: x^k is the result of a run of MHT, it holds $f_2(x^k) = f_2(x_{glob})$ or it holds $x^k \in \operatorname{argmin}_{x \in B_{loc}^{k-1}} f_2(x)$ and $x^k \in \operatorname{int} B_{loc}^{k-1}$ with $B_{loc}^{k-1} = \left\{x \in \mathbb{R}^n \mid \left\|x^{k-1} - x\right\| \leq \delta\right\}$ and δ as defined in Algorithm 6.
In the first case it follows from the main convergence result about MHT, Theorem 4.20 in Section 4.5, that x^k is Pareto critical for (BOP). In the second case it follows from Lemma 5.1 (ii) that x^k is weakly efficient for (BOP). In the third case it follows from Lemma 5.1 (i) that x^k is locally weakly efficient for (BOP). Since Pareto criticality is a

necessary condition for local weak efficiency, see Lemma 2.9 in Section 2.2, it follows that x^k is Pareto critical for (BOP) in all three cases. □

For the implementation of $\text{MHT}_{\text{spread}}$ the modification from Section 4.6.4 is used. As outlined there, the convergence results of MHT transfer and the main result given in Theorem 4.20 holds. Thus, Lemma 5.4 also holds in this case. The numerical results are presented in Section 6.5.1.
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, it is possible to execute $\text{MHT}_{\text{spread}}$ with all functions f_i, $i = 2, 3, \ldots, q$, separately. Since these runs are independent from each other, they could be parallelized. More sophisticated strategies to include all cheap functions are subject to future research.

5.2 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 A.15 is fulfilled and it holds $m_2^k \equiv f_2$ for all $k \in \mathbb{N}$, see also Section 4.6.4. As outlined there, the convergence results of MHT from Section 4.5 still hold for this modification.
In Section 5.2.1 we describe how lower bounds defined in the image space can be handled in MHT. Due to these constraints the convergence results for MHT from Section 4.5 cannot be transferred. In Section 5.2.2 we give an overview of the theoretical results that hold for this approach. In Section 5.2.3 an approach is presented to restrict the search area in the image space after one initial run of MHT. By this, two new

search areas are generated in which either MHT or its modification from Section 5.2.1 is applied. This approach is generalized in Section 5.2.4 to splitting the image space in several distinct areas. How suitable starting points can be chosen in the context of these image space splits is described in Section 5.2.5.

5.2.1 MHT with Lower Bound in the Image Space

MHT applied to bi-objective optimization problems (BOP) can be modified to handle a lower bound in the image space for the cheap function f_2 of the form

$$f_2(x) \geq C \tag{5.5}$$

with $C \in \mathbb{R}$. Integrating such a nonlinear constraint poses of course additional numerical effort. However, the numerical effort is assumed to be acceptable at least in comparison with the expensive function f_1 since f_2 is an analytically given, cheap function. Since the function f_1 is assumed to be expensive, it is not reasonable to include an analogous constraint for f_1. To ensure that the constraint (5.5) is reasonable, we impose Assumption A.18.

Assumption A.18 Let Assumption A.17 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 $C \leq \sup_{x\in\mathbb{R}^n} f_2(x)$.

We exclude the cases of choosing the lower bound C as $C < \min_{x\in\mathbb{R}^n} f_2(x)$ since then the standard version of MHT can be used. This is also the case for $C = \min_{x\in\mathbb{R}^n} f_2(x)$. Still, we include this choice of C, because it will be considered in Section 5.2.4 where the image space is split into disjoint areas in advance of applying any version of MHT. Hereafter, we describe the necessary modifications of MHT to handle the constraint $f_2(x) \geq C$ from (5.5) and formulate the resulting algorithm in Algorithm 7.
The modifications affect the applicability of the convergence results. This is addressed in Section 5.2.2; the convergence analysis is not transferable, but some theoretical results can be proved. However, it is a useful heuristic approach for separate runs of MHT respectively a modified version of it. This is outlined in Sections 5.2.3 and 5.2.4.

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 Pascoletti-Serafini problem in step 2. All other steps in MHT remain unchanged. Let $k \in \mathbb{N}$ be an iteration index. The modified ideal point $\tilde{p}^k = \left(\tilde{p}_1^k, \tilde{p}_2^k\right)^\top$ is defined by

$$\tilde{p}_1^k := \min\left\{m_1^k(x) \mid x \in B_k\right\} = p_1^k, \tag{5.6}$$

$$\tilde{p}_2^k := \min\left\{f_2(x) \mid x \in B_k \text{ and } f_2(x) \geq C\right\} = \max\left\{p_2^k, C\right\}. \tag{5.7}$$

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 (5.5) affects the ideal point and therefore also the computation of the trial point. The modified version of the Pascoletti-Serafini problem (PS) is given by

$$\min\left\{t \in \mathbb{R} \mid f(x^k) + t\,\tilde{r}^k - m^k(x) \in \mathbb{R}^2_+, x \in B_k, f_2(x) \geq C\right\} \tag{$\widetilde{PS}_3$}$$

with $\tilde{r} = f(x^k) - \tilde{p}^k \in \mathbb{R}^2_+$ and $\tilde{p}^k$ as defined in (5.6) and (5.7). The case of a nontrivial lower bound C is illustrated schematically in Figure 5.3. 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 $\tilde{p}^k$.

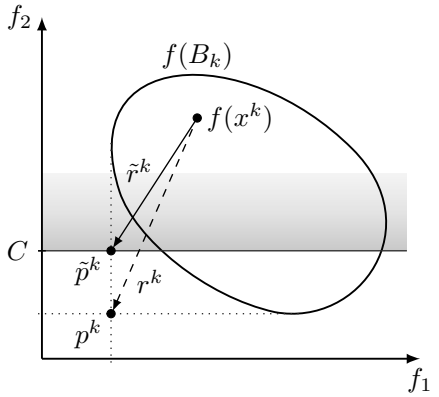

Figure 5.3 – Restricted search area and modified ideal point

The constraint $f_2(x) \geq C$ needs to be integrated explicitly into $(\widetilde{PS}_3)$ to ensure $f_2(\bar{x}) \geq C$ for all minmal solutions $(\bar{t}, \bar{x})^\top \in \mathbb{R}^{1+n}$ of $(\widetilde{PS}_3)$. It is not sufficient to adapt only the determination of the ideal point as described in (5.6) and (5.7). This is illustrated by the following example.

Example 5.5 Let $k \in \mathbb{N}$, $m_1^k \equiv f_1$ and $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ with $f_1(x) = (x+4)^2$ and $f_2(x) = x^2$. Let $x^k = 3$, $\delta_k = 2$ and $C = 4$. Then it holds $B_k = \{x \in \mathbb{R} \mid |x-3| \leq 2\}$, $f(x^k) = m^k(x^k) = (49, 9)^\top$ and

$$\begin{aligned}\tilde{p}_1^k &= \min_{|x-3|\leq 2} (x+4)^2 = 25\\ \tilde{p}_2^k &= \min_{|x-3|\leq 2, x^2 \geq 4} x^2 = 4 = C.\end{aligned}$$

Consider $(\widetilde{PS}_3)$ without the constraint $f_2(x) \geq C$. Then $(\bar{t}, \bar{x})^\top = (-1, 1)^\top$ is feasible since it holds $\bar{x} \in B_k$ and

$$\begin{aligned}f_1(x^k) + \bar{t}\,\tilde{r}_1^k - f_2(\bar{x}) &= 49 - (49 - 25) - 25 = 0,\\ f_2(x^k) + \bar{t}\,\tilde{r}_2^k - f_2(\bar{x}) &= 9 - (9 - 4) - 1 = 3 \geq 0.\end{aligned}$$

For $(\widetilde{PS}_3)$ analogous results to Lemmas 4.4 to 4.6 hold which are given in Lemma 5.6 in Section 5.2.2. There it is stated that for every minimal solution $(\hat{t}, \hat{x})^\top \in \mathbb{R}^{1+n}$ of $(\widetilde{PS}_3)$ it holds $\hat{t} \leq 0$ and if x^k is not weakly efficient for $\min \{m^k(x) \mid x \in B_k, f_2(x) \geq C\}$, then it holds $\hat{t} \in [-1, 0)$. This also holds if the additional constraint $f_2(x) \geq C$ is omitted.
In this example x^k is not weakly efficient for $\min \{m^k(x) \mid x \in B_k, f_2(x) \geq C\}$, since it holds for example $f(\tilde{x}) < f(x^k)$ for $\tilde{x} = 2 \in B_k$ with $f_2(\tilde{x}) = 4 = C$. Hence, $(\bar{t}, \bar{x})^\top = (-1, 1)^\top$ with $f_2(\bar{x}) = 1 < C$ is a minimal solution of $(\widetilde{PS}_3)$ if the constraint $f_2(x) \geq C$ is omitted.

Example 5.5 illustrates that the additional constraint $f_2(x) \geq C$ needs to be integrated explicity into $(\widetilde{PS}_3)$. The algorithm resulting from including this additional constraint into MHT is given in Algorithm 7 and referred to as MHT_{lb}.

Algorithm 7 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 $p^k = (p_1^k, p_2^k)^\top$:
$\tilde{p}_1^k = p_1^k = \min\left\{m_1^k(x) \mid x \in B_k\right\}$,
$\tilde{p}_2^k = \min\left\{f_2(x) \mid x \in B_k, f_2(x) \geq C\right\}$,
$p_2^k = \min\left\{f_2(x) \mid x \in B_k\right\}$
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 $(\widetilde{PS}_3)$:
$\min\left\{t \in \mathbb{R} \mid f(x^k) + t\,\tilde{r}^k - m^k(x) \in \mathbb{R}^2_+, x \in B_k, f_2(x) \geq C\right\}$.
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**.

For the implementation of MHT_{lb} the model function m_1^k is updated in iteration k only if necessary. Analogously to MHT and as described in Section 4.2, this is decided by the outcome of the trial point acceptance test. If $\rho_\phi^k < \eta_1$ holds, the model function m_1^k is recomputed, otherwise the old model is reused in the next iteration.

If the model function is recomputed, the same approach to save function evaluations is used as for the original version of MHT described in Section 4.2. Whenever the model function m_1^k is updated and new interpolation points are computed by the subroutine given in Algorithm 3 in Section 3.2.2, previously evaluated points are reused if pos-

sible. The additional constraint $f_2(x) \geq C$ is not considered in this subroutine; the interpolation points are still chosen from the whole trust region B_k in every iteration $k \in \mathbb{N}$.

Note that the criterion for considering $(\widetilde{PS}_3)$ 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, $(\widetilde{PS}_3)$ is not considered in this case in step 2 of MHT_{lb}. Instead, $t^{k+} = 0$ and $x^{k+} = x^k$ is set directly.

If $r_i^k > 0$ holds for $i = 1, 2$, $(\widetilde{PS}_3)$ is considered. Thus, $(\widetilde{PS}_3)$ 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 5.8 in Section 5.2.2.

5.2.2 Theoretical Results

The convergence results for MHT from Section 4.5 are not transferable for MHT_{lb} due to the additional constraint. In this section we describe the properties that hold in every iteration.

We impose Assumption A.18, i.e. the function f_2 is bounded from below and a constant $C \in \mathbb{R}$ is given with $\min_{x \in \mathbb{R}^n} f_2(x) \leq C$. This constant is considered for the additional constraint $f_2(x) \geq C$ in MHT_{lb}. Furthermore, Assumption A.15 is assumed to be fulfilled, that is it holds $m_2^k \equiv f_2$ for all iterations $k \in \mathbb{N}$.

If $C \leq \min_{x \in \mathbb{R}^n} f_2(x)$ holds, no actual additional constraint for f_2 is introduced and MHT and MHT_{lb} are identical except for the model functions. In this case, the convergence transfers from MHT, see Section 4.6.4. According to Theorem 4.20, the point obtained by MHT is Pareto critical for (BOP). If $C > \min_{x \in \mathbb{R}^n} f_2(x)$ holds, no general statement can be made about the point obtained by MHT_{lb}.

Two aspects prevent the applicability of the results from Section 4.5: the modified ideal point $\tilde{p}^k$ and the additional constraint $f_2(x) \geq C$ in the Pascoletti-Serafini problem $(\widetilde{PS}_3)$. As illustrated in Example 5.5 in the previous section, the latter is necessary to ensure $f_2(x^k) \geq C$ for all $k \in \mathbb{N}$.

For the modified version of the Pascoletti-Serafini problem $(\widetilde{PS}_3)$ analogous results to Remark 4.1 and Lemmas 4.4 to 4.6 for (PS) from Section 4.3 hold. We summarize

the most important results in Lemma 5.6. For this purpose, we define for $k \in \mathbb{N}$ the optimization problem

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

Lemma 5.6 *Let Assumptions* A.4, A.15 *and* A.18 *hold and let $k \in \mathbb{N}$ be an iteration of MHT$_{lb}$. Consider the Pascoletti-Serafini problem*

$$\min \left\{ t \in \mathbb{R} \mid f(x^k) + t\,\tilde{r}^k - m^k(x) \in \mathbb{R}^2_+, x \in B_k, f_2(x) \geq C \right\} \tag{$\widetilde{PS}_3$}$$

with $\tilde{r} = f(x^k) - \tilde{p}^k \in \mathbb{R}^2_+$ and $\tilde{p}^k$ with $\tilde{p}^k_1 = \min \left\{ m^k_1(x) \mid x \in B_k \right\} = p^k_1$ and $\tilde{p}^k_2 = \min \left\{ f_2(x) \mid x \in B_k \text{ and } f_2(x) \geq C \right\}$ as defined in (5.6) *and* (5.7). *Let $(\bar{t}, \bar{x})^\top \in \mathbb{R}^{1+n}$ be a minimal solution of ($\widetilde{PS}_3$). The following statements hold.*

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

(ii) It holds $m^k_1(\bar{x}) \leq m^k_1(x^k)$ and $f_2(\bar{x}) \leq f_2(x^k)$.

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

(iv) If x^k is weakly efficient for ($BOP^k_{m,C}$) and $\tilde{r}^k \in \mathbb{R}^2_{++}$, then $(0, x^k)$ is a minimal solution of ($\widetilde{PS}_3$).

(v) If x^k is efficient for ($BOP^k_{m,C}$) and $\tilde{r}^k \neq 0_2$, then $(0, x^k)$ is a minimal solution of ($\widetilde{PS}_3$).

Proof. To prove statement (i), note that due to Assumption A.4, i.e. $f(x^k) = m^k(x^k)$ for all iterations $k \in \mathbb{N}$, it holds $f(x^k) + 0 \cdot \tilde{r}^k - m^k(x^k) = 0_2$. Thus, the vector $(0, x^k)^\top$ is feasible for ($\widetilde{PS}_3$) and it follows $\bar{t} \leq 0$.
Since $\tilde{r}^k \in \mathbb{R}^2_+$ holds for all $k \in \mathbb{N}$, it follows $m^k_i(x^k) - m^k_i(\bar{x}) \geq \bar{t}\,\tilde{r}^k_i \geq 0$ for $i = 1, 2$. This implies together with Assumption A.15, i.e. $f_2 \equiv m^k_2$ for all $k \in \mathbb{N}$, statement (ii).

To prove statement (iii), let x^k be not weakly efficient for ($BOP^k_{m,C}$). Then there exists a vector $x \in B_k$ with $f_2(x) \geq C$ such that it holds $m^k_i(x) < m^k_i(x^k)$ for $i = 1, 2$. This implies $\tilde{r}^k_i > 0$ for $i = 1, 2$. The rest of the proof is analogous to the proof of Lemma 4.5.
Analogous to Lemma 4.6, statements (iv) and (v) follow from [Eic08, Th. 2.1]. □

This lemma shows that if the current iteration point x^k is not weakly efficient for the trust region subproblem $(BOP^k_{m,C})$, a minimal solution $(t^{k+}, x^{k+})^\top \in \mathbb{R}^{1+n}$ of $(\widetilde{PS}_3)$ with $t^{k+} < 0$ is computed in step 2 of 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 MHT_{lb} can be made.
Besides these general properties of $(\widetilde{PS}_3)$ which are analogous to (PS) 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 5.7.
As outlined in Section 4.6.2, replacing the ideal point in MHT by a vector $y \leq p^k$ 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 $(\widetilde{PS}_3)$.

Lemma 5.7 *Let Assumptions* A.15 *and* A.18 *hold and let* $C > \min_{x \in \mathbb{R}^n} f_2(x)$ *be. Furthermore, let* $k \in \mathbb{N}$ *be an iteration of MHT*$_{lb}$ *with* $\tilde{r}_i^k > 0$ *for* $i = 1, 2$. *We define*

$$\hat{p}^k := f(x^k) + \hat{\lambda}\tilde{r}^k \text{ with } \hat{\lambda} := -\frac{f_2(x^k) - p_2^k}{f_2(x^k) - \tilde{p}_2^k} \tag{5.8}$$

with p^k *and* $\tilde{p}^k$ *the ideal points of* (PS) *and* $(\widetilde{PS}_3)$. *Moreover, we define the Pascoletti-Serafini problem*

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

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* $(\widetilde{PS}_3)$. *Then* $(-\bar{t}/\hat{\lambda}, \bar{x})$ *is a minimal solution of* $(\hat{PS})$.

(iii) Let $(\bar{t}, \bar{x}) \in \mathbb{R}^{1+n}$ *be a minimal solution of* $(\hat{PS})$. *Then* $(-\hat{\lambda}\,\bar{t}, \bar{x})$ *is a minimal solution of* $(\widetilde{PS}_3)$.

The construction of the vector $\hat{p}^k$ out of p^k and $\tilde{p}^k$ is illustrated schematically in Figure 5.4. *The search area generated by $f_2(x) \geq C$ is depicted as gray shaded area.*

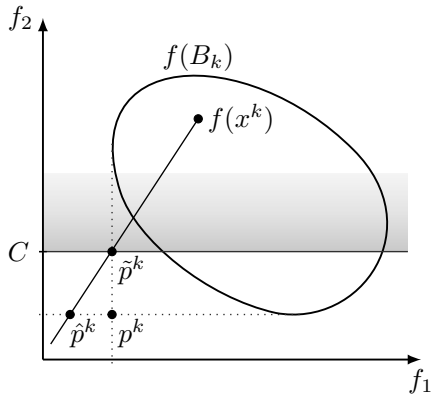

Figure 5.4 – Modified ideal points $\hat{p}^k$ and $\tilde{p}^k$

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 (5.8) 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 $(\widetilde{PS}_3)$. Then it holds $\bar{x} \in B_k$ and $f(\bar{x}) \geq C$ and together with (i) it follows

$$0_2 \leq f(x^k) + \bar{t}\,\tilde{r}^k - m^k(\bar{x}) = f(x^k) - \frac{\bar{t}}{\hat{\lambda}}\,\hat{r}^k - m^k(\bar{x}).$$

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

$$0_2 \leq f(x^k) + t\,\hat{r}^k - m^k(x) = f(x^k) - t\,\hat{\lambda}\,\tilde{r}^k - m^k(x).$$

and $(-t\,\hat{\lambda}, x)$ is feasible for $(\widetilde{PS}_3)$ with $-t\,\hat{\lambda} < \bar{t}$. This contradicts $(\bar{t}, \bar{x})$ being a minimal solution of $(\widetilde{PS}_3)$. Therefore, $(-\bar{t}/\hat{\lambda}, \bar{x})$ is a minimal solution of $(\hat{PS})$. Statement (iii) follows analogously to statement (ii). □

Given Lemma 5.7, 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 $(\widetilde{PS}_3)$. Therefore, the second aspect preventing the convergence results from being transferred is still given.
The most important precondition of Lemma 5.7 is $\tilde{r}_2^k > 0$. This is fulfilled if it holds

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

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. They are given in Lemma 5.8.

Lemma 5.8 *Let Assumptions* A.4, A.15 *and* A.18 *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 7) 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* $(\widetilde{PS}_3)$ *computed in iteration* k*. The following statements hold.*

(i) For every feasible vector $(t, x)^\top \in \mathbb{R}^{1+n}$ *of* $(\widetilde{PS}_3)$ *it holds* $f_2(x) = C$*.*

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

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

$$f_2(x) = C \quad \text{and} \quad m_1^k(x) < m_1^k(x^k), \tag{5.9}$$

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*

$$f_2(x) > f_2(x^k) = C \quad \textit{or} \quad m_1^k(x) \geq m_1^k(x^k), \tag{5.10}$$

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

Proof. Consider the Pascoletti-Serafini problem $(\widetilde{PS}_3)$ given by

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

with $\tilde{r}^k = f(x^k) - \tilde{p}^k$, $\tilde{p}_1^k = p_1^k$ and $\tilde{p}_2^k = \min\{f_2(x) \mid x \in \mathbb{R}^n, f_2(x) \geq C\}$. Since it holds $f_2(x^k) = C$, this implies $\tilde{p}_2^k = C$ and $\tilde{r}_2^k = 0$. According to Assumption A.15, 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 $(\widetilde{PS}_3)$

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

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 it holds

$$f_1(x^k) + t\,\tilde{r}_1^k - m_1^k(x) \geq 0$$

and (t, x) is feasible for $(\widetilde{PS}_3)$. Then it holds $t^{k+} < 0$ and $(0, x^k)^\top$ cannot be a minimal solution of $(\widetilde{PS}_3)$. Thus, it holds $x^{k+} \neq x^k$.
Now let the preconditions of statement (iv) be fulfilled. Note that $(0, x^k)$ is always feasible for $(\widetilde{PS}_3)$. If the first part of (5.10) is fulfilled, i.e. there exists no vector $x \in B_k \setminus \{x^k\}$ with $f_2(x) = C$, then it follows $(t^{k+}, x^{k+}) = (0, x^k)$ from statement (i). If the second part of (5.10) is fulfilled, it holds $m_1^k(x^k) - m_1^k(x) \leq 0$ for all vectors

$x \in B_k \setminus \{x^k\}$. Since $\tilde{r}_1^k > 0$ holds, it follows $t \geq 0$ for all (t, x) feasible for $(\widetilde{PS}_3)$. This implies $t^{k+} = 0$ for the minimal value.
Both cases imply $t^{k+} = 0$. According to the trial point acceptance test in step 3 of MHT_{lb}, it holds $x^{k+1} = x^k$ in both cases. □

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, obtaining an iteration point $x^k \in \mathbb{R}^n$ with $f_2(x^k) = C$ does not necessarily 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 $(\widetilde{PS}_3)$. If 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 5.8 it holds $t^{k+} = 0$ and therefore $\rho_\phi^k = 0$. Hence, iteration k is unsuccessful. 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 if 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 Sections 5.2.3 and 5.2.4 also constraints of the form $f_2(x) \leq D$ with $D \in \mathbb{R}$ are considered. Analogous to MHT_{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 the model function m_2^k is exact, i.e. if it holds $f_2 \equiv m_2^k$ for all $k \in \mathbb{N}$ (Assumption A.15). This is stated in Lemma 5.9 and the analogous result for MHT_{lb} is stated in Lemma 5.10.

Lemma 5.9 *Consider MHT (Algorithm 4) applied to* (BOP) *and let Assumptions* A.4, A.15 *and* A.17 *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. We prove the statement by induction. The base case is trivially fulfilled since it holds $f_2(x^0) \leq D$. Let $f_2(x^k) \leq D$ hold for $k \geq 1$. The next iteration point is either defined as $x^{k+1} = x^k$ or as $x^{k+1} = x^{k+}$ with x^{k+} the trial point obtained by (PS). In the first case, it follows $f_2(x^{k+1}) \leq D$ directly.
In the latter case it holds $r_i^k > 0$ for $i = 1, 2$ due to the definition of step 2 in MHT. Note that in this case a minimal solution for (PS) exists, see Remark 4.2. Let $\left(t^{k+}, x^{k+}\right)^\top \in$

$\mathbb{R}^{1+n}$ be the minimal solution of (PS) computed in step 2. From the constraints of (PS) it follows

$$f(x^k) + t^{k+} r^k - m^k(x^{k+}) \in \mathbb{R}^2_+$$

According to Lemma 4.4, it holds $t^{k+} \leq 0$ and according to Assumption A.15 it holds $m_2^l \equiv f_2$ for all iterations $l \in \mathbb{N}$. This implies

$$f_2(x^k) - f_2(x^{k+}) = f_2(x^k) - m_2^k(x^{k+}) \geq -t^{k+} r_2^k \geq 0. \tag{5.11}$$

Thus, it holds $f_2(x^{k+1}) = f_2(x^{k+}) \leq f_2(x^k) \leq D$. □

Lemma 5.10 *Consider MHT$_{lb}$ applied to* (BOP) *and let Assumptions* A.4, A.15 *and* A.18 *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* A.18. *Moreover, let* $\left(t^{k+}, x^{k+}\right)^\top \in \mathbb{R}^{1+n}$ *be the minimal solution of* $(\widetilde{PS}_3)$ *for every iteration* $k \in \mathbb{N}$ *with* $r^k \in \mathbb{R}^2_{++}$. *Let* $x^0 \in \mathbb{R}^n$ *be the starting point of MHT$_{lb}$ with* $f_2(x^0) \leq D$. *Then it holds* $f_2(x^k) \leq D$ *for all iterations* $k \in \mathbb{N}$.

Proof. Note that it holds $\tilde{r}^k \in \mathbb{R}^2_+$ in all iterations $k \in \mathbb{N}$. Furthermore, it holds $\bar{t} \leq 0$ for every minimal solution $(\bar{t}, \bar{x})^\top \in \mathbb{R}^{1+n}$ of $(\widetilde{PS}_3)$, see Lemma 5.6. The proof then follows analogously to the proof of Lemma 5.9. □

5.2.3 Subsequent Image Space Split

Let $\bar{x} \in \mathbb{R}^n$ be a Pareto critical point for (BOP) obtained for example by the basic version of MHT. In this subsection we describe a heuristic approach of computing another Pareto critical point avoiding the area around $f(\bar{x})$. The idea is to subsequently split the image space into two search areas defined by

$$\begin{aligned} A^+ &:= \{x \in \mathbb{R}^n \mid f_2(x) \geq f_2(\bar{x}) + \varepsilon\} \text{ and} \\ A^- &:= \{x \in \mathbb{R}^n \mid f_2(x) \leq f_2(\bar{x}) - \varepsilon\} \end{aligned} \tag{5.12}$$

with $\varepsilon > 0$ a suitable positive constant. The two search areas A^+ and A^- are schematically illustrated in Figure 5.5. They are depicted as gray shaded areas.

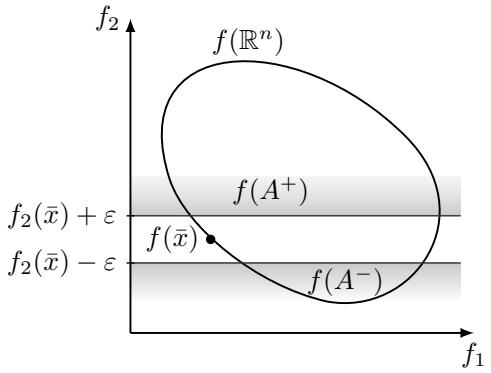

Figure 5.5 – Search areas A^+ and A^-

The search for another Pareto critical point in the area A^+ can – at least heuristically – be realized by applying MHT_{lb} with the lower bound $C = f_2(\bar{x}) + \varepsilon$. The problem of finding a well located starting point for executing MHT_{lb} is addressed in Section 5.2.5. However, it is not guaranteed that the point $\hat{x}$ resulting from MHT_{lb} is Pareto critical for (BOP), see Section 5.2.2. Of course the original version of MHT could be started with $\hat{x}$ as starting point. This will either show that $\hat{x}$ is Pareto critical or produce a Pareto critical point.

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 5.9 in Section 5.2.2, 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 points x^k produced by MHT. According to Theorem 4.20 in Section 4.5, the point resulting from MHT is Pareto critical for (BOP).

Figure 5.5 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 (5.12) 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$_{\text{lb}}$. In case $A_i = A_i^-$ for all $i \in \{2, 3, \ldots, q\}$, MHT can be applied with a starting point $x^0 \in A$. As outlined in Lemma 5.9 in Section 5.2.2, the upper bounds contained in $A_i = A_i^-$, $i \in \{2, 3, \ldots, q\}$, do not need to be considered explicitly, but are fulfilled for every iteration point of MHT if it holds $x^0 \in A$ for the starting point. 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$_{\text{lb}}$.
This is only a first approach to transfer the splitting idea to optimization problems (MOP) with $q \geq 3$ objective functions. More sophisticated approaches to define a search area respectively a taboo-region in case of more than two objective functions is subject to future research.

5.2.4 Upfront Image Space Split

In this section we describe the approach of splitting the image space not after one initial run of MHT, but at the beginning and before MHT or any modification of it is applied. Furthermore, various search areas are considered instead of only two as described in Section 5.2.3. In those areas MHT$_{\text{lb}}$ is then applied with suitable lower bounds. 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

$$f_2^{\min} := \min_{x \in \mathbb{R}^n} f_2(x) \quad \text{and} \quad f_2^{\max} := \max_{x \in \mathbb{R}^n} f_2(x). \tag{5.13}$$

If f_2 is not bounded from above, $f_2^{\max}$ can be defined as $f_2^{\max} := E$ with $E \in \mathbb{R}$ and $E > f_2^{\min}$. The optimization problems in (5.13) are global optimization problems. Again and as in Section 5.1, if f_2 is convex (resp. concave), a local optimization method can be applied and otherwise, a global method is required, see for example [HP95;

HT96]. Since the objective function f_2 of the unconstrained problem (BOP) is a cheap function, we assume that these optimization problems can be solved with reasonable numerical effort, at least in comparison with the numerical effort associated with the expensive function f_1.
$f_2^{\min}$ 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

$$A_j := \{x \in \mathbb{R}^n \mid a_{j-1} \leq f_2(x) \leq a_j\} \tag{5.14}$$

$$\text{with } a_0 = f_2^{\min} \text{ and } a_j = f_2^{\min} + j\frac{f_2^{\max} - f_2^{\min}}{p} \tag{5.15}$$

for $j = 1, 2, \ldots, p$. This is illustrated schematically in Figure 5.6 for different values of p. Figure 5.6 indicates that not in every search area A_j, $j = 1, 2, \ldots, p$, a weakly efficient point needs to exist.

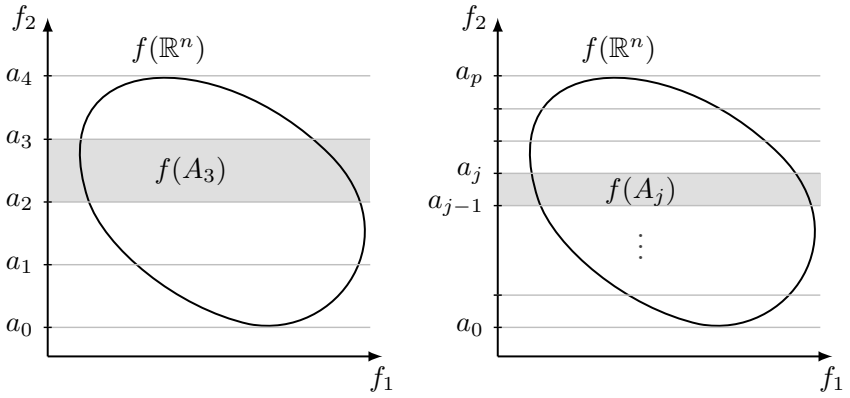

Figure 5.6 – Image space splits with different values for p

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 Section 5.2.5. The numerical tests confirm that the splitting approach described in this subsection is a useful heuristic.

Besides a starting point, MHT_{lb} needs as input a lower bound C. This constant C is given by the set A_j as $C = C_j = a_{j-1}$ for all $j \in \{1, 2, \ldots, p\}$. If $j = 1$ holds, the lower bound in A_1 is $a_0 = f_2^{\min}$. Thus, it is a trivial lower bound and can be omitted. For $j = 1$ MHT can be applied, since the second constraint in A_1 given by $f_2(x) \leq a_1$ does not need to be included explicitly. If a starting point $x^0 \in A_1$ is chosen, $f_2(x^k) \leq a_1$ is fulfilled in all iterations $k \in \mathbb{N}$, see Lemma 5.9 in Section 5.2.2. This also applies to MHT_{lb}. As outlined in Lemma 5.10 in Section 5.2.2, an upper bound of the form $f_2(x) \leq a_j$ for $j \in \{1, 2, \ldots, p\}$ is inherent and does not need to be included explicitly if the starting point fulfills it.

The different runs of MHT_{lb} with A_j and C_j are independent from each other and could be applied parallel for all $j = 1, 2, \ldots, p$. However, parallelization techniques are not the main subject of this thesis. Furthermore, as the examples for the image space splits depicted in Figure 5.6 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 the 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 schematically illustrated in Figure 5.7 with $B_i = f(A_i), i = 1, 2, \ldots, p$.

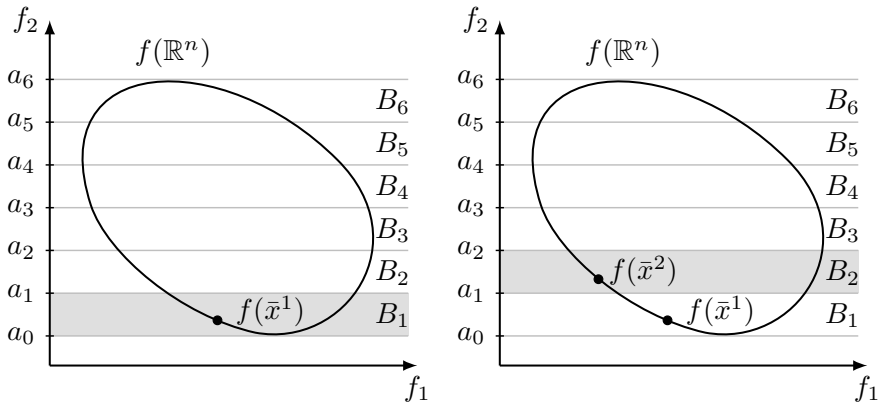

Figure 5.7 – MHT resp. MHT_{lb} applied in A_1 and A_2

At first, (BOP_1) and A_1 are considered. As outlined above, the constraint $f_2(x) \geq a_0$ in A_1 is trivially fulfilled since it holds $a_0 = f_2^{\min}$. Thus, MHT can be applied to

(BOP_1). Let $\bar{x}^1 \in \mathbb{R}^n$ be the result of this run. Then according to Theorem 4.20 in Section 4.5, $\bar{x}^1$ is Pareto critical for (BOP). 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 value of the global minimum 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 Section 5.2.2. 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 5.8. 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.

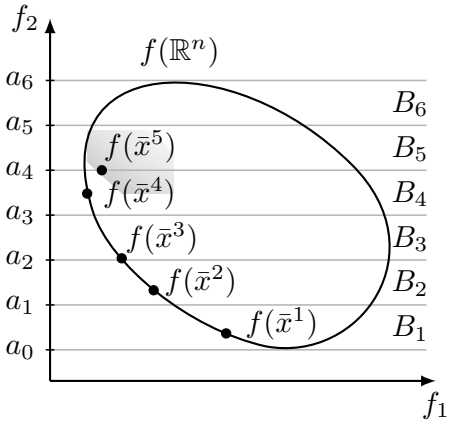

Figure 5.8 – Heuristic strategy to terminate splitting approach

This is only a heuristic strategy and gives no guarantee to only terminate the splitting approach if no Pareto critical points exist in the remaining search areas. This is illustrated by an example in Section 6.5.2 where the heuristic test fails and further runs could find further Pareto critical points.
However, the numerical results show that this happens only in individual cases among the considered test problems. In general, the heuristic test to terminate the splitting

approach and to save function evaluations is reasonable and works well. Different examples from the numerical results are presented and discussed in Section 6.5.2. The algorithm realizing the image space split with consecutive runs of MHT_{lb} and the heuristic stopping criterion is described in Algorithm 8 and referred to as $\text{MHT}_{\text{split}}$.

Algorithm 8 $\text{MHT}_{\text{split}}$: 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
$x^0 \in \operatorname{argmin}_{x \in \mathbb{R}^n} f_2(x)$ 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 $i = 1, 2, \ldots, p$ (Section 5.2.5).

Step 2: Consecutive runs of MHT_{lb} (Algorithm 7)

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 $\text{MHT}_{\text{split}}$. More sophisticated approaches to include all cheap functions in a splitting approach are subject to future research.

5.2.5 Choice of Starting Points

In this section we present a heuristic approach to generate well located starting points for the splitting approaches from Sections 5.2.3 and 5.2.4. 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 $C, D \in \mathbb{R}$, $\min_{x \in \mathbb{R}^n} f_2(x) \leq C \leq D$. If the only aim is to guarantee feasibility, a starting point $x^0 \in A$ can be defined as

$$x^0 \in \operatorname*{argmin}_{x \in A} g(x)$$

with $g : \mathbb{R}^n \to \mathbb{R}$ any function that can easily be minimized locally, e.g. the cheap function f_2 or a linear function or a constant function. For a general overview of the feasibility problem see for example [Chi08]. For the splitting approach not only feasibility is important, but also that the starting point is located well.
As outlined in Lemmas 5.9 and 5.10 in Section 5.2.2, 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 Section 5.2.2.
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 A.15. Thus, it holds $f_2(x^{k+1}) \leq f_2(x^k)$ for all $k \in \mathbb{N}$, see also Remark 4.25 in Section 4.6.4 or Lemma 5.6 in Section 5.2.2.
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

$$\min \{f_2(x) \mid x \in A \text{ and } f_2(x) \geq D - \varepsilon\} \tag{5.16}$$

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 (5.16), 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 (5.16), 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 (5.16). The numerical results presented in Section 6.5.2 show that this approach of defining starting points is useful and works well for the majority of the considered test problems.

5.3 Combination of Image Space Split and Spreading

The modifications of MHT from the former subsections can be combined to a heuristic approach to approximate the set of Pareto critical points of the unconstrained optimization problem (BOP). At first, the image space split from Section 5.2.4 is applied. The algorithm $\text{MHT}_{\text{split}}$ is applied and the image space is split into p areas $A_1, A_2, \ldots, A_p$. Let

$$\bar{X} \coloneqq \left\{\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^k\right\} \tag{5.17}$$

with $k \leq p$ be the set of points resulting from $\text{MHT}_{\text{split}}$. As described in Section 5.2.4 $\text{MHT}_{\text{split}}$ 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 5.9. 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 $\text{MHT}_{\text{split}}$.

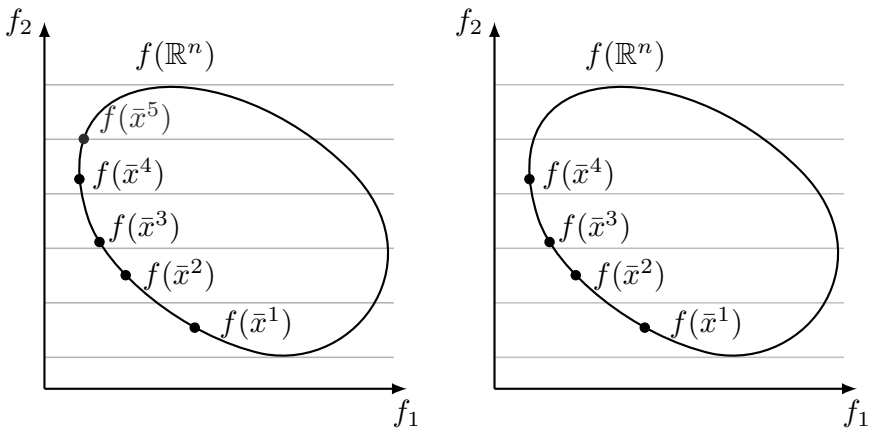

Figure 5.9 – Step 1: Image space split with $\text{MHT}_{\text{split}}$

For all points in $\bar{X}$ the spreading technique described in Section 5.1 based on individual minima of the cheap functions is then used to compute further Pareto critical points. We use a slightly modified version of $\text{MHT}_{\text{spread}}$ (Algorithm 6) given in Algorithm 9 and referred to as $\text{MHT}^{\text{b}}_{\text{spread}}$.

As in $\text{MHT}_{\text{spread}}$, the cheap objective function f_2 is minimized subsequently on moving local areas. However, instead of doing so until the value of the global minimum $f_2(x_{glob}) = \min_{x\in\mathbb{R}^n} f_2(x)$ 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 9 $\text{MHT}^{\text{b}}_{\text{spread}}$: $\text{MHT}_{\text{spread}}$ (Algorithm 6) 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 $x_{glob} \in \text{argmin}_{x\in\mathbb{R}^n} f_2(x)$, set $X = \{x^0\}$, $\delta = \delta_0$ and $k = 1$.
while $f_2(x^k) > b$ **do**
  Compute $x_{loc} \in \text{argmin}_{x\in B^k_{loc}} f_2(x)$ with $B^k_{loc} = \left\{x \in \mathbb{R}^n \mid \left\|x^k - x\right\| \leq \delta\right\}$.
  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

Note that in the while-loop it is possible to obtain a point x^k with $f_2(x^k) < b$ either as $x^k = x_{loc} \in \text{argmin}_{x\in B^k_{loc}} f_2(x)$ or as $x^k = \text{MHT}(x_{loc}, \delta)$. Nevertheless, this point is contained in the set X which is the output of $\text{MHT}^{\text{b}}_{\text{spread}}$ since it is found to be a Pareto critical point for (BOP), see Lemma 5.1 in Section 5.1. This is also stated in Lemma 5.11 which holds analogously to Lemma 5.4 from Section 5.1. The proof is analogous to Lemma 5.4 and therefore omitted.

Lemma 5.11 *Suppose Assumptions* A.1 *to* A.11 *hold. Consider* MHT^{b}_{spread} *(Algorithm* 9*) 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) $\text{MHT}^{\text{b}}_{\text{spread}}$ is applied for all points contained in $\bar{X}$, the output of $\text{MHT}_{\text{split}}$ defined in

(5.17). 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 $\text{MHT}^{\text{b}}_{\text{spread}}$. The lower bound b for termination is chosen as

$$b = b_j = f_2(\bar{x}^{j-1}) \tag{5.18}$$

given by the preceding point from $\bar{X}$. For $\bar{x}^1$ $\text{MHT}_{\text{spread}}$ is applied. Applying $\text{MHT}^{\text{b}}_{\text{spread}}$ respectively $\text{MHT}_{\text{spread}}$ to the points of $\bar{X}$ is illustrated schematically Figure 5.10. All points obtained by the spreading approach are marked gray. The left figure shows the result of $\text{MHT}^{\text{b}}_{\text{spread}}$ with $x^0 = \bar{x}^2$ and $b = f_2(\bar{x}^1)$ and the right figure the result for all points of $\bar{X}$.

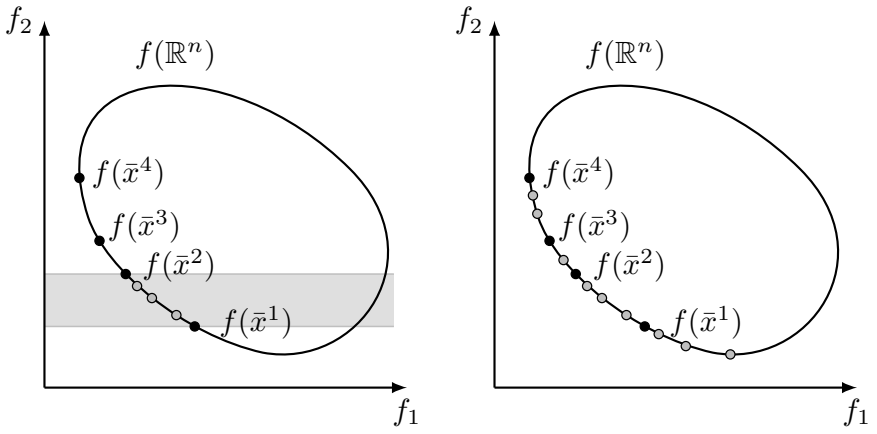

Figure 5.10 – Step 2: Spreading with $\text{MHT}^{\text{b}}_{\text{spread}}$ resp. $\text{MHT}_{\text{spread}}$

The complete procedure combining the described steps and various runs of MHT with different modifications is given in Algorithm 10 and referred to as $\text{MHT}_{\text{combi}}$.
The output of $\text{MHT}_{\text{combi}}$ are the sets $\bar{X}$ and Y. According to Lemma 5.11, all points contained in $Y \setminus \bar{X}$ are Pareto critical for (BOP). The set Y is expected to be a densification of the set $\bar{X}$, that is Y contains more points and the distance between them reduces compared to the points in $\bar{X}$. The numerical results presented in Section 6.5.3

confirm the usefulness of this approach.
Since Pareto criticality is only a necessary condition for local weak efficiency, it is possible that the points obtained by $\text{MHT}_{\text{combi}}$ contained in the set $Z := \bar{X} \cup Y$ are not all nondominated in this set. This is also illustrated in the numerical results.
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 $\text{MHT}_{\text{spread}}$ from Section 5.1. For the numerical tests the extraction of the nondominated points 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 or the Jahn-Graef-Younes method, see [You93; JR06].

Algorithm 10 $\text{MHT}_{\text{combi}}$: 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 $\text{MHT}_{\text{split}}$
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 $\text{MHT}^{\text{b}}_{\text{spread}}$ and $\text{MHT}_{\text{spread}}$
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: $\bar{X}, Y$

6 Numerical Tests

In this chapter the results of numerical tests of MHT and its modifications are presented. At first, some numerical details including the stopping criterion and the trust region update are presented in Section 6.1. In Section 6.2 the set of test problems is described. They are listed in detail in Appendix A.2. To evaluate and compare the results of MHT, two comparison methods are chosen which are described in Section 6.3. In Section 6.4 the test results for MHT and in Section 6.5 the test result for the heuristic approaches $\text{MHT}_{\text{spread}}$, $\text{MHT}_{\text{split}}$ and $\text{MHT}_{\text{combi}}$ are presented.

6.1 Numerical Details

6.1.1 Stopping Criterion

The general formulation of the solution method MHT in Section 4.1 includes no stopping criterion. This is not required for the theoretical considerations about the convergence behavior of MHT in Section 4.5. However, for the implementation a suitable stopping criterion is required. In the following, we present and discuss several options. In the end, we give a summary of the criteria used for the implementation.
Since the objective function f_1 is expensive regarding the evaluation time, it is reasonable to determine a maximum number of allowed function evaluations n_{max} for f_1 and to stop the algorithm if this number is reached.
It is proved in the convergence analysis of MHT in Section 4.5 that the trust region radius converges to zero when the iterates converge to a Pareto critical point of the unconstrained multi-objective optimization problem (MOP). MHT is designed to reduce the trust region radius whenever no sufficient decrease is possible with the current model functions. Additionally, Lemma 4.16 from Section 4.5 ensures that whenever

the current iteration point x^k is not Pareto critical for (MOP) and the trust region radius falls below a fraction of $\omega(x^k)$, the radius will not decrease in the next iteration. Additionally, according to Lemma 4.17 from Section 4.5, the trust region radius is bounded from below if x^k is not Pareto critical for (MOP).
Thus, if the trust region radius is smaller than a suitable constant $\varepsilon_{tr} > 0$, the algorithm can stop. Therefore, it is reasonable to terminate the algorithm if it holds

$$\delta_k \leq \varepsilon_{tr}. \tag{6.1}$$

This is analogous to a stopping criterion commonly used in the single-objective trust region approach, see Section 3.1. Moreover, it is suggested in [CGT00], for numerical reasons, to connect this bound for δ_k to the norm of the current iteration point x^k. A stopping criterion is then given by $\delta_k < \varepsilon \left\|x^k\right\|$. For the implementation of MHT the simpler form from (6.1) is used since no numerical issues occurred.
Another reasonable option for a stopping criterion is using an indicator for Pareto criticality. Such a stopping criterion would correspond to a query $\left\|\nabla g(x^k)\right\| = 0$ for a scalar-valued function $g : \mathbb{R}^n \to \mathbb{R}$ and a single-objective trust region approach for g. Besides the test if the trust region is small enough, such a query is commonly suggested and used as stopping criterion in the single-objective trust region versions. This is also mentioned in Section 3.1; further details can be found in [CGT00]. To transfer this to the multi-objective trust region approach, the function

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

from Definition 2.10 can be used. As proved in Section 2.2, this function characterizes Pareto criticality, the corresponding property to the single-objective notion of stationarity. A vector x is Pareto critical for (MOP) if and only if it holds $\omega(x) = 0$. However, for the expensive, simulation-given function f_1 derivative information is not available with reasonable numerical effort and therefore $\nabla f_1(x^k)$ cannot be computed. As a surrogate $\nabla m_1^k(x^k)$ could be used. Consequently, it could be tested if it holds

$$-\min_{\|d\| \leq 1} \max \left\{ \nabla m_1^k(x^k)^\top d, \left\{ \max_{i=2,3,\dots,q} \nabla f_i(x^k)^\top d \right\} \right\} = 0$$

in every iteration $k \in \mathbb{N}$. However, the model function m_1^k is an interpolation model using quadratic or linear Lagrange polynomials. Since derivative information is not available, the model function is based only on function values. Therefore, the derivative information of the model is not reliable and should not be used for computing the function value of ω and verifying a stopping criterion. Hence, this option for a stopping criterion is not included in the numerical realization of MHT.
Another indicator for Pareto criticality is the step size t^{k+} of the Pascoletti-Serafini problem (PS) given by

$$\min \left\{ t \in \mathbb{R} \mid f(x^k) + t(f(x^k) - p^k) - m^k(x) \in \mathbb{R}^q_+, x \in B_k \right\}$$

from Section 4.1. According to Lemmas 4.5 and 4.6, it holds $t^{k+} \in [-1, 0)$ if x^k is not weakly efficient for (MOP_m^k) and if $(0, x^k)$ is a solution of (PS), i.e. it holds $t^{k+} = 0$, then x^k is weakly efficient for (MOP_m^k). Since it always holds $x^k \in \text{int}B_k$, this implies that x^k is locally weakly efficient and therefore Pareto critical for $\min_{x \in \mathbb{R}^n} m^k(x)$ if $t^{k+} = 0$ holds, see also Lemma 2.9 in Section 2.2.
The reliability of the model functions is characterized in terms of validity, see Definition 3.1. A model function $m_i^k, i \in \{1, 2, \ldots, q\}$, is called valid for the original function f_i in the trust region B_k, if it holds

$$\left| f_i(x) - m_i^k(x) \right| \leq \kappa \delta_k^2$$

for all $x \in B_k$ with a suitable constant $\kappa > 0$. Thus, the validity of the model functions depends on the size of the trust region. The smaller the region, the more accurate the approximation of the model functions. As stated in Lemma 4.8, validity of the models for the cheap functions can even be proved in every iteration $k \in \mathbb{N}$. For the model function m_1^k, validity is assumed in every iteration, see Assumption A.8. Therefore, it is reasonable to use the step size t^{k+} as a stopping criterion if the trust region in iteration k is small enough. If it holds

$$\left| t^{k+} \right| \leq \varepsilon_{PS} \text{ and } \delta_k \leq \delta_{PS}$$

with $\varepsilon_{PS} > 0$ and $\delta_{PS} > 0$ suitable constants, the algorithm can terminate. For the implementation of MHT a combination of the presented stopping criteria is used. If

one of the conditions (i) - (iii) is fulfilled, MHT – or its modification MHT_{lb} presented in Section 5.2.1 – terminates.

(i) $n_k \geq n_{max}$ with n_k the number of function evaluations of f_1 up to iteration k and n_{max} the maximum number of allowed function evaluations for f_1

(ii) $\delta_k \leq \varepsilon_{tr}$ with $\varepsilon_{tr} > 0$ a suitable constant

(iii) $\left|t^{k+}\right| \leq \varepsilon_{PS}$ and $\delta_k \leq \delta_{PS}$ with $\varepsilon_{PS} > 0$ and $\delta_{PS} > 0$

If statement (i) holds, it is possible that the algorithm stops without having reached a Pareto critical point. In this case, the search strategy and the trial point acceptance test of MHT respectively MHT_{lb} ensure a decrease for at least one objective function of MHT_{lb}, see Section 4.4.
For the implementation of MHT and its modifications the parameters from the conditions (i)-(iii) are chosen as

$$n_{max} = 2000, \quad \varepsilon_{tr} = 0.1, \quad \varepsilon_{PS} = 0.001.$$

Note that for testing MHT not only unconstrained test problems are considered, but also box constrained test problems. Let $lb, ub \in \mathbb{R}^n$ be the lower and upper bounds of the optimization problem. The value δ_{PS} is defined as

$$\delta_{PS} = \max_{i=1,\dots,n} \frac{ub_i - lb_i}{10} \quad \text{or} \quad \delta_{PS} = 1$$

if it holds $lb_i = -\infty$ or $ub_i = \infty$ for an index $i \in \{1, 2, \dots, n\}$. As explained in Section 4.7, box constraints can be handled by MHT, yet the convergence results do not transfer. Nevertheless, we use the same stopping criteria. The numerical results show that for the considered test problems this is unproblematic.
Moreover, stopping criterion (iii) is justified analogously to the unconstrained case due to the properties of the adjusted Pascoletti-Serafini problem. These are given in Lemma 4.26 and are analogous to the properties of the Pascoletti-Serafini problem for unconstrained problems, see Lemmas 4.5 and 4.6.

6.1.2 Trust Region Update

Following the single-objective trust region approach from [CGT00; CSV09b] and the multi-objective trust region approaches from [VOS14; RK14], the trust region update is realized in general as

$$\delta_{k+1} = \begin{cases} \frac{1}{2}\delta_k & , \text{ if } \rho_\phi^k < \eta_1 \\ \delta_k & , \text{ if } \eta_1 \leq \rho_\phi^k < \eta_2 \\ 2\delta_k & , \text{ if } \rho_\phi^k \geq \eta_2 \end{cases} .$$

Thus, the input parameters γ_1 and γ_2 for MHT (Algorithm 4) are chosen as $\gamma_1 = \gamma_2 = 0.5$. To save further function evaluations, the realization of the trust region update is slightly modified with regard to the stopping criterion (iii) from Section 6.1.1.

During the iterations the trust region radius δ_k can become very large. If the first condition of the stopping criterion (iii) is fulfilled, that is it holds $|t^{k+}| \leq \varepsilon_{PS}$ with $\varepsilon_{PS} > 0$ for an iteration $k \in \mathbb{N}$, the trial point provides no sufficient decrease. If the second condition of (iii) is not fulfilled, this results in decreasing the trust region radius and updating the model function. For this purpose, new function evaluations are required since most likely, due to the smaller radius, not all previously used interpolation points are situated in the new trust region.

In the subsequent iterations two cases can occur. Either in one of the iterations $j > k$ the trial point provides a sufficient decrease and it holds $|t^{j+}| > \varepsilon_{PS}$ or it holds $|t^{j+}| \leq \varepsilon_{PS}$ for all subsequent iterations $j > k$. In the latter case, the procedure of reducing the radius continues until the stopping criterion (iii) is fulfilled, i.e. until the trust region radius is small enough and it holds $\delta_j \leq \delta_{PS}$ with $\delta_{PS} > 0$ from Section 6.1.1.

In all of these iterations new function evaluations would be required. To reduce the number of evaluations, the trust region radius is set to δ_{PS} in one step if it holds $|t^{k+}| \leq \varepsilon_{PS}$ for an iteration $k \in \mathbb{N}$. The radius is reduced to the value of δ_{PS} since this is the largest size of a trust region in which a local model function is assumed to be reliable to apply the stopping criterion (iii).

The model function is updated and the trial point and the acceptance test are executed. Either the more accurate model confirms the previous behavior and the stopping criterion (iii) from Section 6.1.1 is fulfilled, i.e. it holds $|t^{j+}| \leq \varepsilon_{PS}$ and $\delta_j \leq \delta_{PS}$ for

$j = k + 1$, or a new trial point is computed which provides a sufficient decrease. Thus, the trust region update in step 4 of MHT (Algorithm 4) can be summarized as follows. The parameters η_1 and η_2 are chosen for the implementation as $\eta_1 = 0.001$ and $\eta_2 = 0.9$.

Algorithm 11 Step 4 of MHT: Trust Region Update

If $|t^{k+}| \leq \varepsilon_{PS}$ and $\delta_k > \delta_{PS}$, set $\delta_k = \delta_{PS}$.

Otherwise, set $\delta_{k+1} = \begin{cases} \frac{1}{2}\delta_k & , \text{ if } \rho_\phi^k < \eta_1 \\ \delta_k & , \text{ if } \eta_1 \leq \rho_\phi^k < \eta_2 \\ 2\delta_k & , \text{ if } \rho_\phi^k \geq \eta_2 \end{cases}$.

6.1.3 Further Numerical Details

The algorithms MHT, $\text{MHT}_{\text{spread}}$, $\text{MHT}_{\text{split}}$ and $\text{MHT}_{\text{combi}}$ are implemented as introduced in Chapters 4 and 5 in MATLAB (version 2017a) with the following numerical adaptions and specifications.

The computation of the interpolation points for the Lagrange models is implemented according to Algorithm 3 in Section 3.2. As outlined in Section 4.2, former interpolation points are reused if possible to save function evaluations.

The subproblems of computing the ideal point p^k and solving the Pascoletti-Serafini problem are realized by using FMINCON with the following specifications: the step tolerance 10^{-10}, the scaling option and the algorithm 'SQP' if it holds $n < 10$ for the dimension of the domain respectively the algorithm 'interior-point' if it holds $n \geq 10$.

To evaluate the results of the numerical tests and to compare MHT to other solution methods described in Section 6.3, the number of required function evaluations is considered. Furthermore, the distance to the Pareto front is used to evaluate the result of any algorithm considered hereinafter. If the distance falls below a problem dependent fraction, the optimization problem is considered as solved. The set of efficient points can be computed with reasonable numerical effort for every test problem described in Section 6.2.

Moreover, all queries if any scalar is equal to zero is realized by testing if the scalar is smaller than a suitable positive constant.

6.2 Test Problems

The following set of 78 test problems is based on test problems from the literature for general multi-objective and derivative-free algorithms [HM79; Sch85; Kur91; FF95; FF98; Deb99; Vel99; CL99; ZDT00; JOS01; Deb+02; CVL02; Hub+06; Cus+11] and complemented with some self-chosen problems. All considered problems are test problems and do not involve an actual expensive function. For the test problems the efficient points can be computed which is necessary to compare the results of the algorithm to them. For evaluating the results one of the functions is declared as expensive and the amount of function evaluations for this function is counted.
Among the test examples are quadratic and nonquadratic functions, convex and nonconvex problems, either unconstrained or with box constraints. Table 6.1 shows an overview of all 78 considered test problems with information about the dimension of the domain (n), the constraints and the convexity of the problem. It also includes information about the geometry of the Pareto front (convex, nonconvex, disconnencted). We say the Pareto front PF is convex (nonconvex) if the set $PF + \mathbb{R}^2_+$ is convex (nonconvex). Besides, the table contains in its last column which of the objective functions is declared as expensive for the test runs of MHT. If there are significant differences regarding the difficulty of the functions, the more difficult function is declared as expensive.
Some of the test examples are scalable and different values for n are considered. They are listed in the table. For every test problem 10 randomly generated, but fixed, starting points are used. All test problems are listed in Appendix A.2 including the considered starting points. One test instance is defined as one test problem with one starting point. In total, 780 test instances have been considered, among them 340 convex and 440 nonconvex instances.

name	n	constraints	convexity	PF	exp.
bi-objective test problems (q=2)					
BK1 [Hub+06; Cus+11]	2	box	conv.	conv.	f_1
CL1 [CL99; Cus+11]	4	box	nonc.	conv.	f_1
Deb41 [Deb99; Cus+11]	2	box	conv.	conv.	f_2
Deb53 [Deb99; Cus+11]	2	unc.	nonc.	nonc.	f_2
Deb513 [Deb99; Cus+11]	2	box	nonc.	discon.	f_2
Deb521b [Deb99; Cus+11]	2	box	nonc.	nonc.	f_2
DG01 [Hub+06; Cus+11]	1	box	nonc.	conv.	f_1
DTLZ1 [Deb+02; Cus+11]	2	box	nonc.	conv.	f_1
ex005 [HM79; Cus+11]	2	box	nonc.	nonc.	f_2
Far1 [Hub+06; Cus+11]	2	box	nonc.	nonc.	f_1
FF [FF95]	2,3,4,5	box	nonc.	nonc.	f_1
Fonseca [FF98]	2	box	nonc.	nonc.	f_1
IM1 [Hub+06; Cus+11]	2	box	nonc.	conv.	f_1
Jin1 [JOS01; Cus+11]	2,3,4,5,10, 20,30,40,50	box	conv.	conv.	f_1
Jin2 [JOS01; Cus+11]	2,3,4,5	box	nonc.	conv.	f_2
Jin3 [JOS01; Cus+11]	2,3,4,5	box	nonc.	nonc.	f_2
Jin4 [JOS01; Cus+11]	2,3,4,5	box	nonc.	nonc.	f_2
JOS3 [Hub+06]	3	box	conv.	conv.	f_1
Kursawe [Kur91; Cus+11]	3	box	nonc.	discon.	f_1
Laumanns [Hub+06]	2	box	conv.	conv.	f_1
LE1 [Hub+06; Cus+11]	2	box	nonc.	nonc.	f_1
Lis [CVL02; Hub+06]	2	box	nonc.	nonc.	f_1
lovison1 [Cus+11; Lov11]	2	box, unc.	conv.	conv.	f_1
lovison2 [Cus+11; Lov11]	2	box, unc.	nonc.	nonc.	f_2

lovison3 [Cus+11; Lov11]	2	box, unc.	nonc.	conv.	f_1
lovison4 [Cus+11; Lov11]	2	box, unc.	nonc.	nonc.	f_1
MOP1 [Hub+06; Cus+11]	1	unc.	conv.	conv.	f_1
Schaffer2 [Sch85]	1	box	nonc.	discon.	f_1
T1	2	unc.	conv.	conv.	f_1
T2	2	unc.	nonc.	nonc.	f_2
T3	2	box	conv.	conv.	f_2
T4	2,3,4,5,10, 20,30,40,50	box	conv.	conv.	f_1
T5	2	box	conv.	conv.	f_1
T6	2	box	conv.	conv.	f_1
T7	3	box	conv.	conv.	f_1
VU1 [Hub+06; Cus+11]	2	box	nonc.	conv.	f_1
VU2 [Hub+06; Cus+11]	2	box	conv.	conv.	f_2
ZDT1 [ZDT00; Cus+11]	4	box	nonc.	conv.	f_2
ZDT2 [ZDT00; Cus+11]	4	box	nonc.	nonc.	f_2
ZDT3 [ZDT00; Cus+11]	4	box	nonc.	discon.	f_2
ZDT4 [ZDT00; Cus+11]	2	box	nonc.	conv.	f_2
ZDT6 [ZDT00; Cus+11]	4	box	nonc.	nonc.	f_2
test problems with q=3 objective functions					
FES2 [Hub+06; Cus+11]	10	box	nonc.	nonc.	f_3
IKK1 [Hub+06; Cus+11]	2	box	conv.	conv.	f_1
T8	3	box	conv.	conv.	f_3
ZLT1 [Hub+06; Cus+11]	4	box	conv.	conv.	f_3

Table 6.1 – Test Problems

6.3 Comparison Methods

The solution method MHT is compared with two other methods. On the one hand, since MHT computes only one Pareto critical point and does not approximate the set of efficient points, the weighted sum approach with equal weights is used and the solution method EFOS (Expensive Function Optimization Solver) [The11] is applied to it with the predefined standard parameters.
EFOS is a solution method for expensive, simulation-based single-objective optimization problems also using the trust region approach and local model functions. The purpose of this solution method also is to save computation time and reduce the number of function evaluations. As a stopping criterion the gradients of the model functions are used in combination with a validity criterion for the models. For convex multi-objective optimization problems every efficient point can be computed by a weighted sum of the objectives with suitable weights. For nonconvex problems only a subset of the efficient points can be computed. This needs to be regarded when comparing the results.
On the other hand, and to circumvent the disadvantages of the weighted sum approach, the multi-objective method DMS [Cus+11] is used as a comparative method. It is a direct search approach and therefore derivative-free and suitable for expensive functions. It approximates the whole set of efficient points, but offers also the option to compute only one efficient point. We used the latter option with the predefined standard parameters. However, for the numerical results in Section 6.5.4 DMS is applied with the option to approximate the whole Pareto front since it is compared to the modifications of MHT from Chapter 5 which aim to compute several Pareto critical points.
As a stopping criterion DMS uses the maximum number of function evaluations given by the user, i.e. 2000 in our numerical tests, and the step size for the search step. If the step size is lower than the predefined value (10^{-3}), DMS stops. Furthermore, DMS includes a method to compute starting points on its own which is chosen by default. To use the starting point the user passes as input to the algorithm, the parameter 'list' in the parameter file needs to be changed from the predefined value 3 to 0.
Of course also the way of the implementation of the algorithms influences the performance for the test problems. In the currently available implementation of EFOS often internal errors occur and runs are terminated without having computed a solution.

6.4 Numerical Results of MHT

Hereafter, we present some selected test instances to illustrate the procedure of MHT and to compare the results to the methods DMS and EFOS. The presentation is based on the summary of results given in [TE18b] and arranged in subsections according to convex, nonconvex and scalable test problems. Furthermore, the influence of the model function and the trial point acceptance test is discussed. In the last subsection, we present performance profiles including all considered test instances. All values given for variables, function values or parameters are rounded to two decimals.
Among the test problems both unconstrained and box constrained problems exist. Furthermore, not all considered algorithms contain a guarantee that the obtained point fulfills a necessary optimality condition. Thus and to avoid confusion due to terminology, we will refer to the points obtained as output by any algorithm as 'solutions'.
MHT is designed for unconstrained problems, still we used also box-constrained optimization problems for the tests. It occurred for some test instances that the iteration points reached the boundary of the box and stayed there. However, in most of the instances, this is unproblematic since the obtained point is (weakly) efficient for the considered optimization problem.

6.4.1 Convex Test Problems

At first we consider the quadratic, convex test problem (BK1) from [Hub+06; Cus+11], see also Test Problem P.1 in Appendix A.2, given by

$$\min_{x\in\Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x\in[-5,10]^2} \begin{pmatrix} x_1^2 + x_2^2 \\ (x_1-5)^2 + (x_2-5)^2 \end{pmatrix} \tag{BK1}$$

to illustrate the procedure of MHT. For this test problem function f_1 is declared as expensive function. For all instances of this test problem MHT and EFOS compute efficient points, DMS only for most of the instances. EFOS computes for different starting points always the same efficient point, whereas MHT and DMS generate different efficient points. MHT needs 12-13 expensive function evaluations and therefore significantly less than EFOS (57-73) and DMS (41-61).

Figure 6.1 shows one test result for MHT (domain top left, image space top right), EFOS (domain middle left, image space middle right) and DMS (domain bottom left, image space bottom right). The domain resp. image set is represented by gray points, the starting point is marked black and the solutions are marked orange. The starting point is $x^0 = (9.42, 9.46)^\top$ and the solutions computed by MHT, EFOS and DMS are $\bar{x}_{MHT} = (4.34, 4.34)^\top$, $\bar{x}_{EFOS} = (2.50, 2.50)^\top$, $\bar{x}_{DMS} = (4.92, 4.96)^\top$.
For MHT the iteration points are marked black and connected by a dotted line and the interpolation points that are evaluated to compute the model functions are marked as unfilled circles. In the domain the trust regions are depicted as gray shaded, transparent circles (the more areas overlap, the darker the gray shade). The representation of the domain is restricted to the feasible set if possible. For EFOS and DMS it is not possible to distinguish between iteration points and further evaluated points during the iterations. Thus, all points evaluated for computing the solution are marked as unfilled circles and only the starting point and the solution are highlighted as for MHT.
For quadratic functions the quadratic interpolation model used for the expensive function in MHT is exact. Thus, the model built in the beginning of the algorithm is reused in all following iterations and the trust region radius increases from iteration to iteration. Only in the last iteration, when it is checked if a stopping criterion is fulfilled, the model function is recomputed in a local area.
This is visible in the top left part of Figure 6.1, where the interpolation points are situated in the first and in the last trust region. The interpolation points are also close to the iteration points in the image space which the top right figure shows. Both figures illustrate the local search strategy of MHT.
EFOS computes more points than MHT and they are more spread both in the image space and in the domain. During the run even infeasible points are generated. This can be seen only in the domain (middle left) since for illustrative reasons we used the same range in the image space for all figures on the right.
DMS also computes more points than MHT, but they are not spread in the domain as for EFOS, but accumulate in a local area. Apart from this, the bottom left figure illustrates the search along the coordinate directions. In the last iterations the step size decreases.
Figure 6.2 illustrates how the choice of the starting point affects MHT, DMS and EFOS. The randomly generated starting points which are listed for Test Problem P.1 in Appendix A.2 are marked as unfilled circles and the solutions obtained by the algorithms

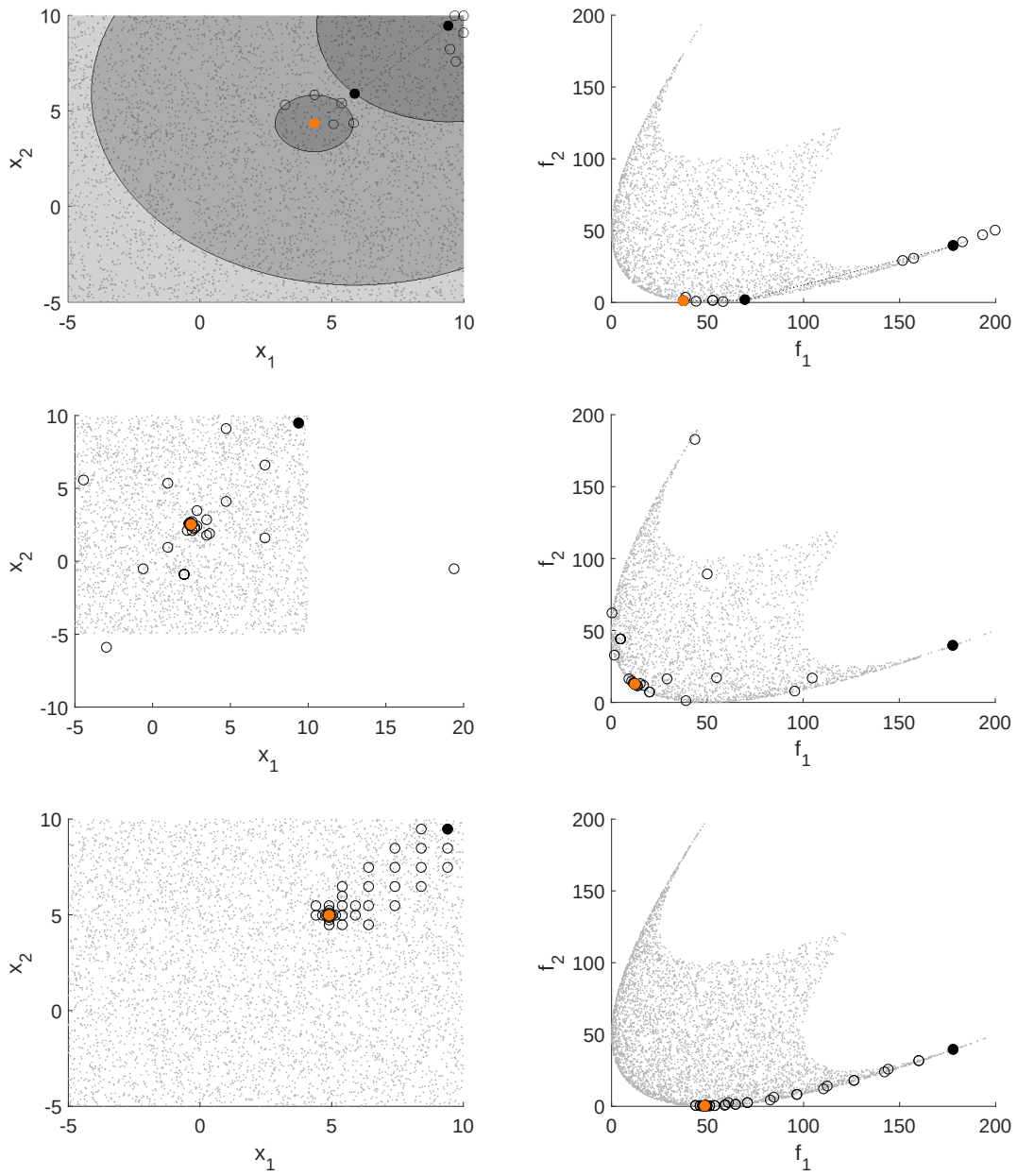

Figure 6.1 – MHT (top), EFOS (middle) and DMS (bottom) applied to (BK1)

are marked black. The figure shows that MHT (top left) and DMS (bottom) compute different solutions for different starting points. The points obtained by MHT are all nondominated points, whereas not all of the points obtained by DMS are nondominated, but some of them have a large distance to the Pareto front. EFOS (top right) generates the same nondominated point for all considered starting points. A reason

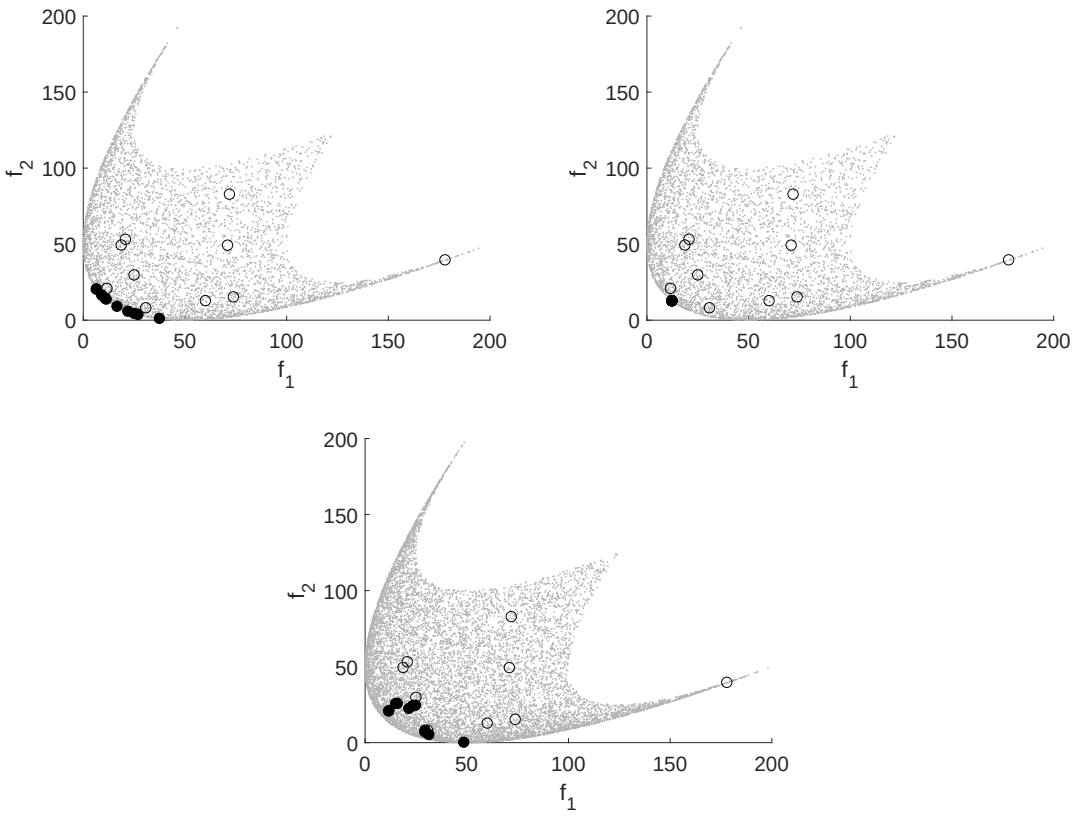

Figure 6.2 – MHT (top left), EFOS (top right) and DMS (bottom) applied to (BK1) with different starting points

for this is the weighted sum approach.

As second test example we consider the convex, but not quadratic problem (T6) given by

$$\min_{x\in\Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x\in(0,100]^2} \begin{pmatrix} -\ln(x_1) - \ln(x_2) \\ x_1^2 + x_2 \end{pmatrix}, \tag{T6}$$

see also Test Problem P.36 in Appendix A.2. For all test instances of this optimization problem EFOS is prematurely canceled due to an internal error. Consequently, we only show results for DMS and MHT.

The as expensive declared function f_1 is not quadratic and therefore the interpolation model used in MHT is not exact. Though, the algorithm reuses former interpolation points if possible which is illustrated for one instance, see the top left (domain) and top right image (image space) in Figure 6.3. The bottom left and right

image show the result for DMS in the domain and the image space. The starting point is $x^0 = (61.69, 50.70)^\top$ and the solutions computed by MHT and DMS are $\bar{x}_{MHT} = (47.48, 100)^\top$, $\bar{x}_{DMS} = (61.69, 50.70)^\top$.

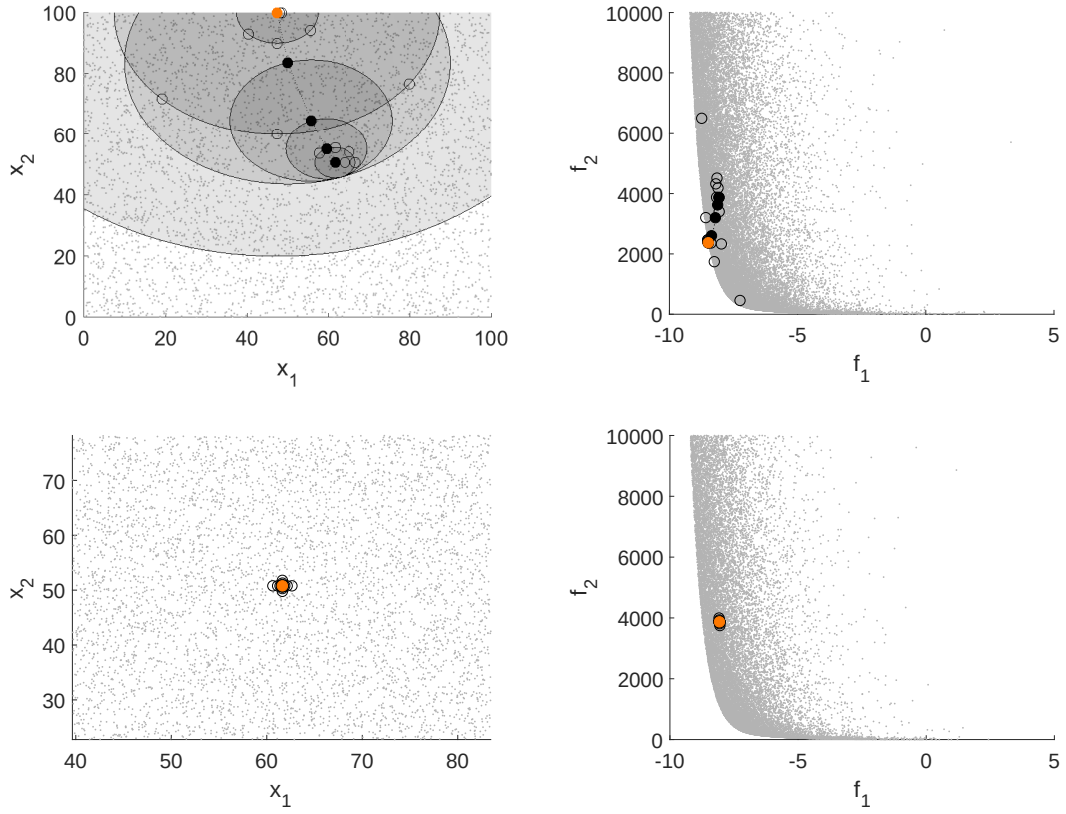

Figure 6.3 – MHT (top) and DMS (bottom) applied to (T6)

For this test instance the iterates of MHT move towards an efficient point with few interpolation points that are mostly close to the iteration points. As the top left figure shows, the model function is only updated in some iterations and already evaluated points are reused. The point $\bar{x}_{MHT}$ lies on the boundary of Ω. Although MHT is not designed for handling constraints, it stops legitimately since no sufficient decrease is found in the last trust region and $\bar{x}_{MHT}$ is efficient for (T6). In total, 19 function evaluations of f_1 are required.

In contrast, DMS terminates after 41 function evaluations with the starting point which is not close to an efficient point. The results of runs with randomly generated starting points (listed for Test Problem P.36 in Appendix A.2) are depicted in Figure 6.4. MHT generates nondominated points from different areas of the Pareto front using

8-114 function evaluations. DMS terminates for all starting points with points close to them. This is illustrated in the right figure as all the unfilled starting points are overlapped by the filled points given as output by DMS. All of these points are computed within 41 function evaluations and most of them are not close to the Pareto front.

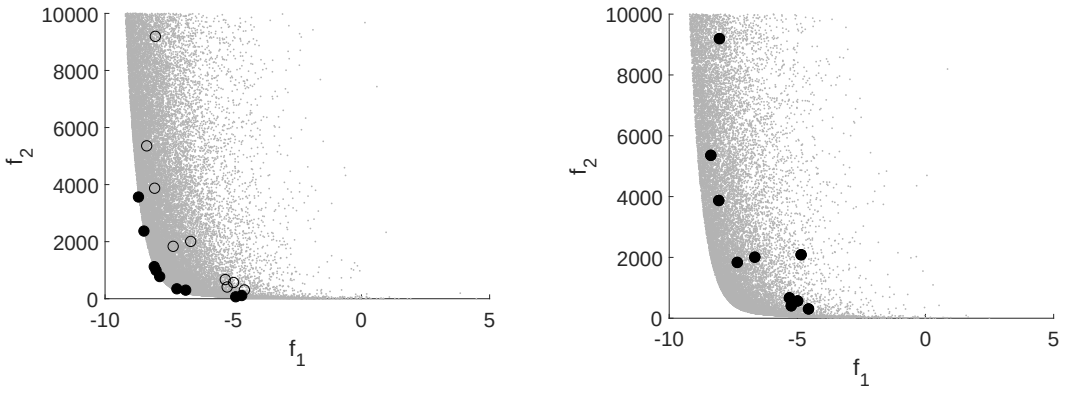

Figure 6.4 – MHT (top left), EFOS (top right) and DMS (bottom) applied to (T6) with different starting points

Another convex problem is (T7), but with a three-dimensional domain, given by

$$\min_{x\in\Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x\in[0,30]^3} \begin{pmatrix} \sum_{i=1}^{3} x_i^4 + \sum_{i=1}^{3} x_i^3 \\ \sum_{i=1}^{3} x_i \end{pmatrix} \tag{T7}$$

and with f_1 declared as expensive function, see also Test Problem P.37 in Appendix A.2. The unique efficient point for this optimization problem is $\bar{x} = (0,0,0)^\top$ with the function values $f_1(\bar{x}) = f_2(\bar{x}) = 0$. For all considered starting points all three algorithms compute $\bar{x}$ respectively a point with vanishing distance to it. This is shown exemplarily for the starting point $x^0 = (23.77, 28.78, 19.67)^\top$ in Figure 6.5.
The range of function evaluations for all instances is similar for MHT (29-57) and EFOS (9-75). Compared to DMS with 206-348 evaluations, both MHT and EFOS save function evaluations. The reason for this large difference is the direct search approach of DMS. DMS produces dense points in the image space during the search steps. This is illustrated in the bottom of Figure 6.5. Again, EFOS computes also infeasible points during the runs which is also the case for the instance depicted in Figure 6.5.

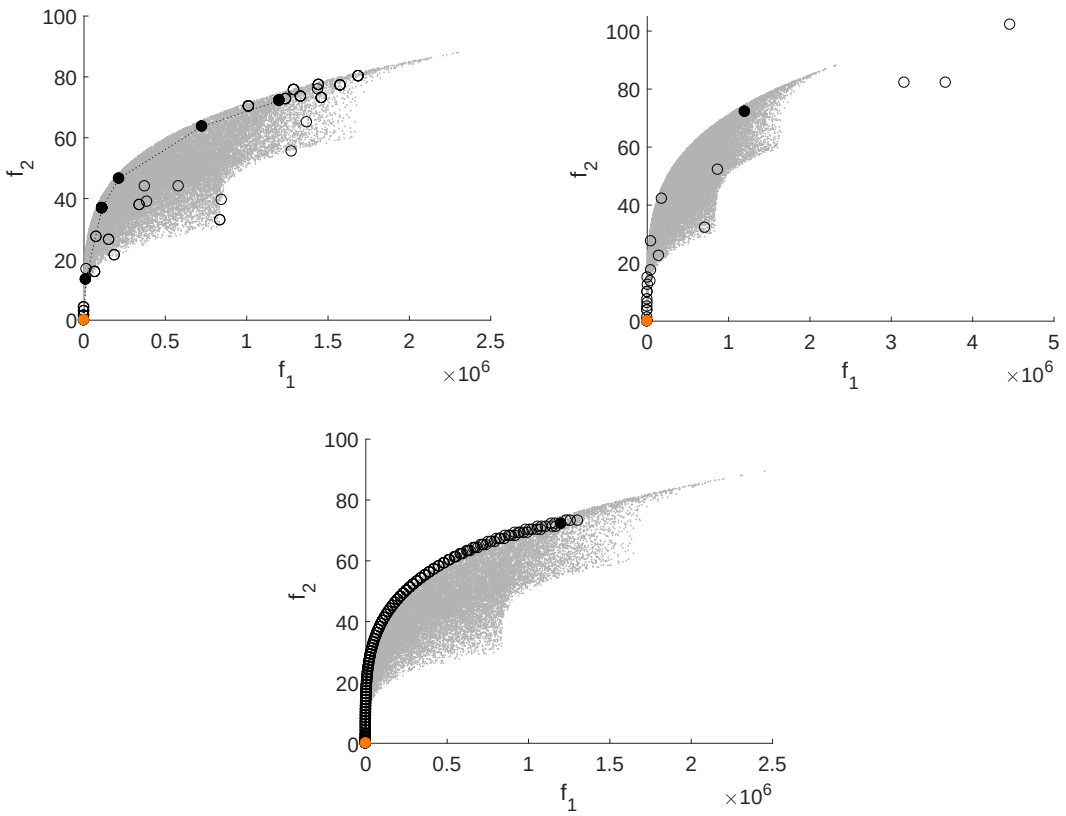

Figure 6.5 – MHT (top left), EFOS (top right) and DMS (bottom) applied to (T7)

6.4.2 Nonconvex Test Problems

MHT is a local method and for convex optimization problems local and global optimality is identical. In general, this does not apply to nonconvex optimization problems. Still, the algorithm performs well for most of the nonconvex test problems. Exemplarily, we consider the results of three nonconvex test problems. The first test problem is (Deb513) from [Deb99; Cus+11] defined by

$$\min_{x\in\Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x\in[0,1]^2} \begin{pmatrix} x_1 \\ g(x)h(x) \end{pmatrix} \tag{Deb513}$$

with $g(x) = 1+10x_2$, $h(x) = 1-(x_1/g(x))^2-(x_1/g(x))\sin(8\pi x_1)$, see also Test Problem P.5 in Appendix A.2. For this test problem f_2 is declared as expensive function and besides the nonconvexity, the Pareto front is disconnected. All three algorithms are

capable of computing a nondominated point for all instances of this test problem. This is shown in Figure 6.6, where all considered starting points (listed in Appendix A.2) are marked as unfilled points and the computed efficient points are marked as filled black points.

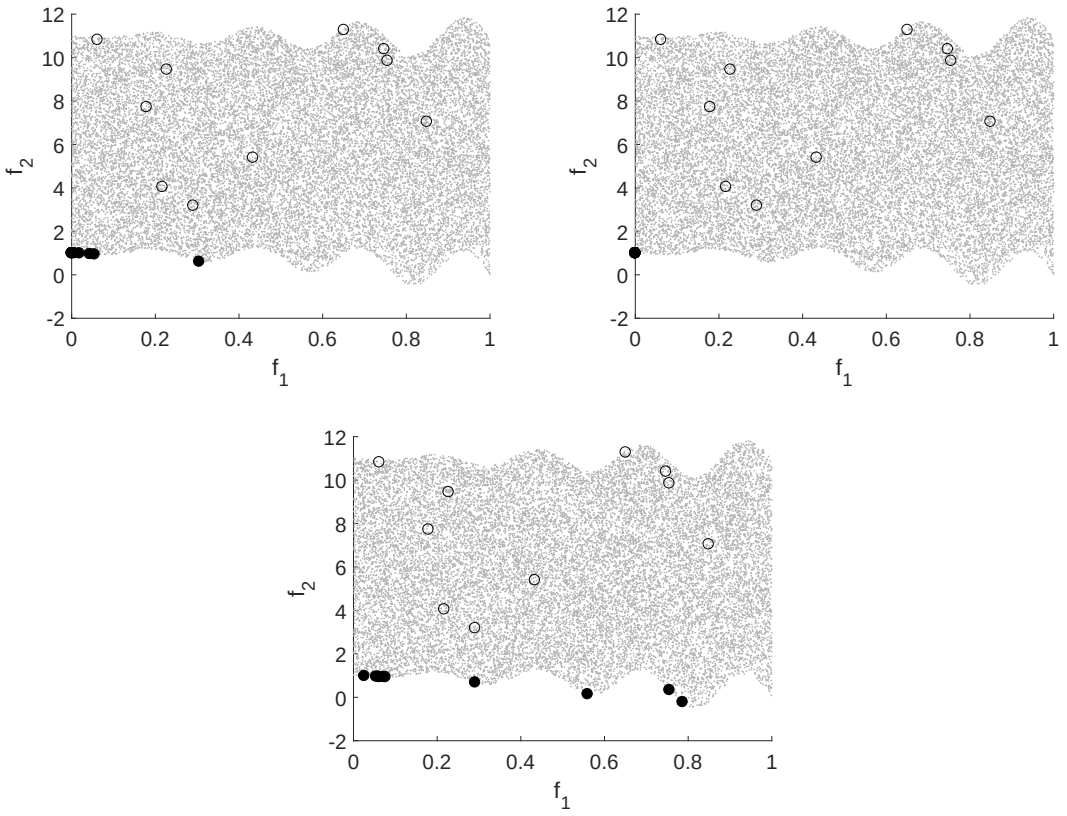

Figure 6.6 – MHT (top left), EFOS (top right) and DMS (bottom) applied to (Deb513) with different starting points

Furthermore, Figure 6.6 shows that again MHT (top left) and DMS (bottom) generate different nondominated points whereas EFOS (top right) – presumably due to the weighted sum approach – computes the same nondominated point for all considered starting points.

The different search strategies are recognizable in Figure 6.7 (domain left, image space right). The starting point is $x^0 = (0.43, 0.40)^\top$ and the results of MHT, EFOS and DMS are $\bar{x}_{MHT} = (0.05, 0.00)^\top$, $\bar{x}_{EFOS} = (0.00, 0.00)^\top$ and $\bar{x}_{DMS} = (0.06, 0.00)^\top$.

EFOS (middle) computes the individual minimum of function f_1 with 7 function evaluations. The points that are evaluated during the run are mostly infeasible and situated

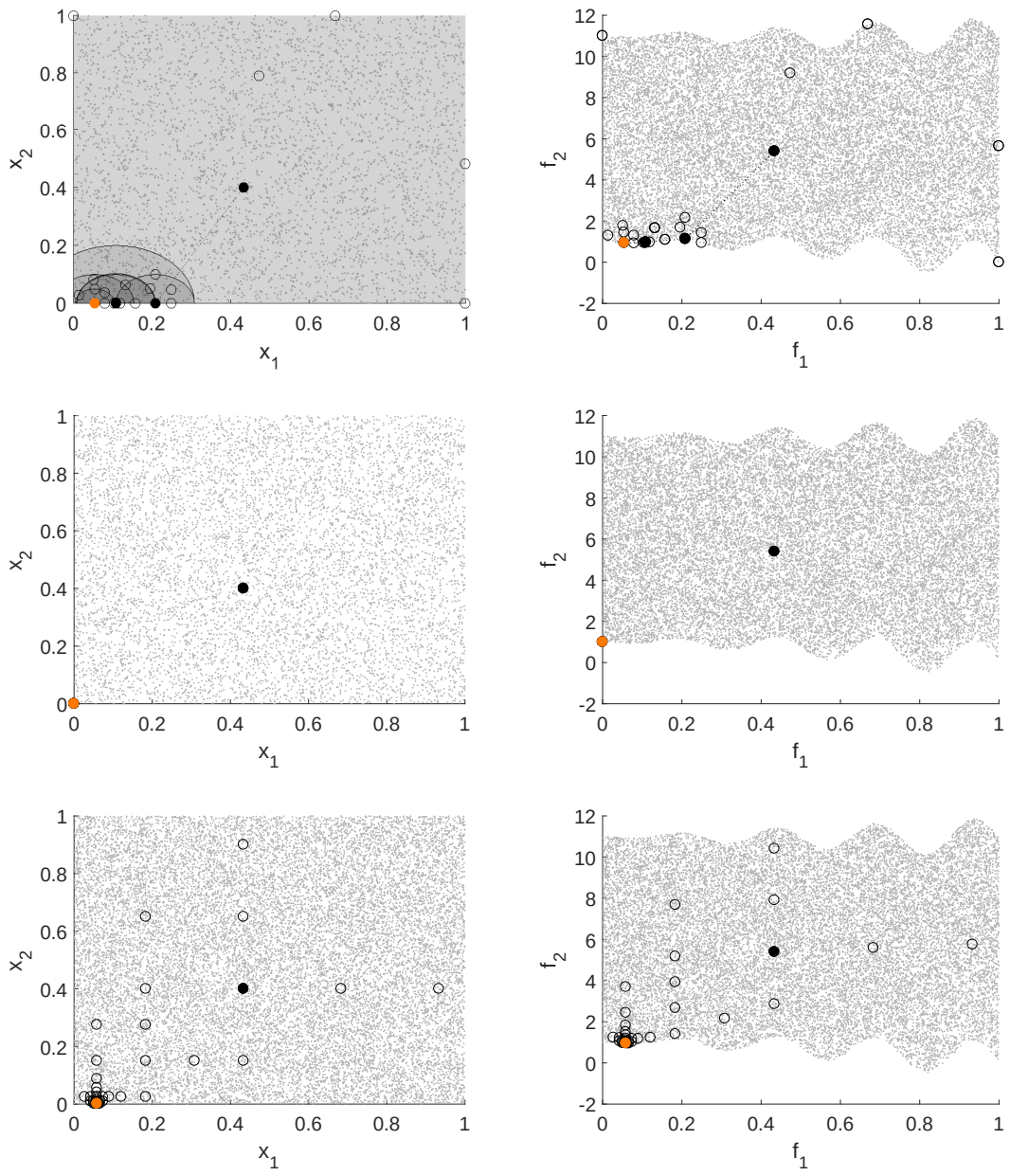

Figure 6.7 – MHT (top), EFOS (middle) and DMS (bottom) applied to (Deb513)

far outside the pictured area. We zoomed in for reasons of illustration. For MHT (25 function evaluations) the top figures illustrate the local search behavior. Only for the initial model the interpolation points are spread broader both in the domain and in the image space, yet in the further iterations the interpolation points are close to the iteration points.

Besides, it holds $\bar{x}_{MHT} \in \partial\Omega$. The top left figure shows that the second iteration point is already situated on the boundary. This instance illustrates that although MHT is not designed for box constrained problems, it does not terminate as soon as the boundary is reached. The top right figure shows that these interpolation points are not (weakly) efficient for (Deb513). Thus, MHT behaves correctly and after some further steps on the boundary in which the model function for f_2 is updated it terminates with an efficient point.

The direct search approach of DMS (38 function evaluations) is visible in the bottom figures. It illustrates the search along the axes. DMS and MHT need a similar amount of function evaluations, yet both need significantly more than EFOS. As second nonconvex test problem we consider (Jin2) with $n = 4$ from [JOS01; Cus+11] defined by

$$\min_{x\in\Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x\in[0,1]^4} \begin{pmatrix} x_1 \\ g(x)\left(1 - \sqrt{\frac{x_1}{g(x)}}\right) \end{pmatrix} \tag{Jin2}$$

with $g(x) = 1 + 3\sum_{i=2}^{4} x_i$ and f_2 declared as expensive, see also Test Problem P.17 in Appendix A.2. For all considered starting points EFOS did not compute a solution, but stopped with an internal error. DMS and MHT compute efficient points for all instances of this problem. This is illustrated in Figure 6.8 which shows the results of MHT (left) and DMS (right) in the image space for randomly chosen starting points. The points computed by DMS are better distributed than the points computed by MHT. No clear statement can be made which algorithm performs better regarding the required function evaluations (DMS 92-126, MHT 30-169).

In all runs MHT needs many function evaluations in the end of the procedure. Due to the nonconvexity and the local search strategy the high number of function evaluations is required to ensure a stopping criterion being fulfilled. This is exemplarily shown for one run in Figure 6.9. The starting point is $x^0 = (0.81, 0.69, 0.34, 0.89)^\top$ and the computed solutions are $\bar{x}_{MHT} = (0.2146, 0.00, 0.00, 0.00)^\top$ and $\bar{x}_{DMS} = (0.81, 0.00, 0.00, 0.00)^\top$.

Note that computing one model function for the expensive function requires $p = (n+1)(n+2)/2$ function evaluations, see Section 3.2. For this test problem it holds $n = 4$ and $p = 15$. MHT needs in total 155 function evaluations. As the left figure illustrates, the interpolation points for MHT accumulate close to the orange marked

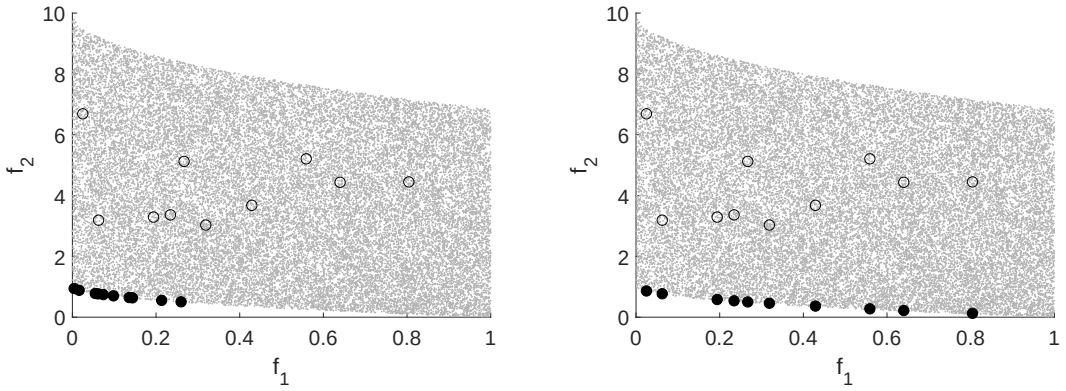

Figure 6.8 – MHT (left) and DMS (right) applied to (Jin2) with different starting points

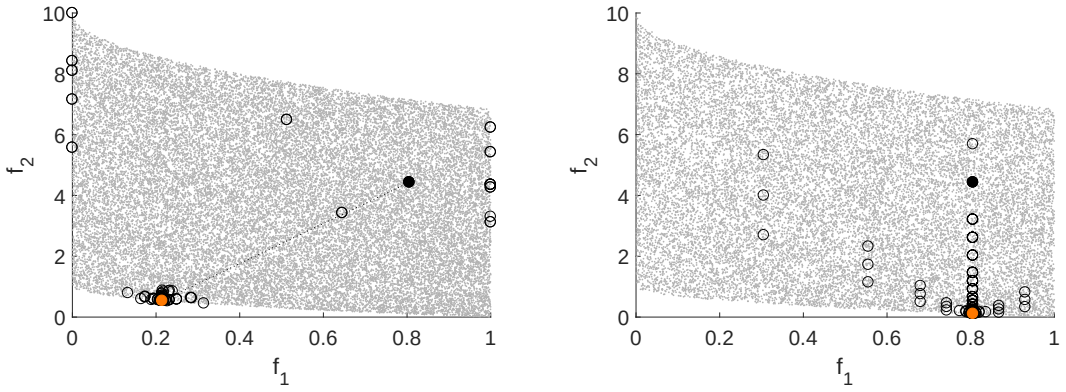

Figure 6.9 – MHT (left) and DMS (right) applied to (Jin2)

point $f(\bar{x}_{MHT})$ and many function evaluations are required in the last iterations until a stopping criterion is fulfilled. DMS needs 112 function evaluations; the right figure illustrates again the coordinate search strategy of DMS.

As last nonconvex test problem we consider (FF) with $n = 3$ from [FF95], see also Test Problem P.12 in Appendix A.2, given by

$$\min_{x\in\Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x\in[-4,4]^n} \begin{pmatrix} 1 - \exp\left(-\sum_{i=1}^{n}\left(x_i - \frac{1}{\sqrt{n}}\right)^2\right) \\ 1 - \exp\left(-\sum_{i=1}^{n}\left(x_i + \frac{1}{\sqrt{n}}\right)^2\right) \end{pmatrix} \tag{FF}$$

with f_1 declared as expensive function. Figure 6.10 shows the result of MHT for (FF) with the starting point $x^0 = (0.70, -3.39, -1.95)^\top$ (marked black) for which the ob-

tained point $\bar{x} = (-0.06, -1.21, -0.84)^\top$ with $f(\bar{x}) = (1.00, 0.52)^\top$ (marked orange) is not weakly efficient for (FF), but fulfills a necessary condition for weak efficiency.

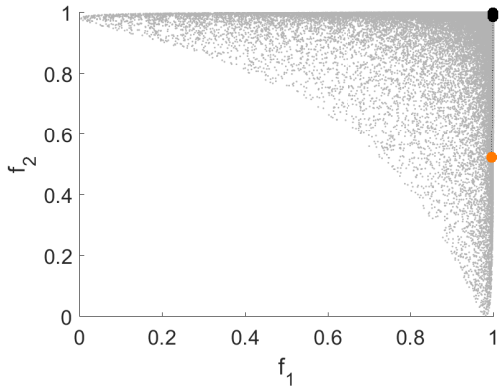

Figure 6.10 – MHT applied to (FF)

Note that the constraints of (FF) are not active in $\bar{x}$. MHT terminates after few iterations since the stopping criterion (iii) from Section 6.1.1 is fulfilled, i.e. the step size $\bar{t}$ of the Pascelotti-Serafini problem is small enough ($\bar{t} = 3.32 \cdot 10^{-7}$) and the trust region is small enough ($\bar{\delta} = 0.8$). As outlined in Section 6.1.1 and according to Lemma 4.26 from Section 4.7, this implies that the model function $\bar{m}$ of the last iteration is reliable in its trust region $\bar{B}$ and that $\bar{x}$ is weakly efficient for the optimization problem $\min \{\bar{m}(x) \mid x \in \Omega \cap \bar{B}\}$.

Furthermore, since the constraints are not active in $\bar{x}$, it holds $\omega_b(\bar{x}) = \omega(\bar{x})$ for the functions ω, ω_b characterizing necessary optimality conditions in the unconstrained case respectively the box constrained case (Pareto critical/stationary points), see Sections 2.2 and 4.7. As stated there, x is a stationary (Pareto critical) point for the considered optimization problem if and only if it holds $\omega_b(\bar{x}) = 0$ ($\omega(\bar{x}) = 0$).

Although these functions are not used as stopping criteria for the implementation of MHT as explained in Section 6.1.1, they can be considered in this case to examine if the termination of MHT was correct. Since for all test problems the functions are given analytically, the values of the auxiliary functions ω, ω_b in $\bar{x}$ can be computed. It holds $\omega_b(\bar{x}) = \omega(\bar{x}) = 8.26 \cdot 10^{-3}$. This confirms that $\bar{x}$ fulfills a necessary condition for weak efficiency for (FF) and that MHT terminates correctly.

6.4.3 Scalable Test Problems

Test problems with different dimensions n for the domain have been considered and runs for MHT with two scalable test problems up to dimension 50 have been executed. As expected, the numerical effort rises when the dimension rises. The higher the dimension is, the more function evaluations are required to compute a model.
If in every iteration of MHT a complete new set of interpolation points would be computed, a total number of $(n+1)(n+2)/2$ function evaluations would be required for the expensive function in every iteration. However, the model is not updated in every iteration, but only if necessary and if it is updated, former interpolation points are reused if possible, see also Section 4.2. Thus, the number of new function evaluations is kept to a minimum.
Furthermore, we suggest in Section 4.6.1 to use a linear interpolation model for higher dimensions. It requires only $n+1$ interpolation points and therefore the number of function evaluations also reduces. Though, it is possible that a linear model needs to be updated more often since it is less accurate. In the numerical tests linear models are used for all instances with dimension $n \geq 10$. To illustrate the behavior of MHT with rising dimension and the effect of linear models, we consider exemplarily the scalable test problem (Jin1) from [JOS01; Cus+11] defined by

$$\min_{x\in\Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x\in[0,1]^n} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_i^2 \\ \frac{1}{n}\sum_{i=1}^{n} (x_i-2)^2 \end{pmatrix} \tag{Jin1}$$

with f_1 declared as expensive, see also Test Problem P.16 in Appendix A.2. It has been tested with MHT for $n \in \{2,3,4,5,10,20,30,40,50\}$. Figure 6.11 shows results of MHT applied to (Jin1) with $n=5$ on the left and $n=10$ on the right.
For $n=10$ a linear model is used which as expected worsens the predictions of the model functions. Since the trial point acceptance test in MHT does not demand a strict decrease for every objective function, but only for at least one objective function, points can be accepted that increase one of the objective functions. This is the case for some iteration points of the instance depicted in the right of Figure 6.11.
Further information about the effects of the acceptance test and the comparison to a stricter version can be found in Sections 4.4 and 4.6.3. Results of numerical tests are

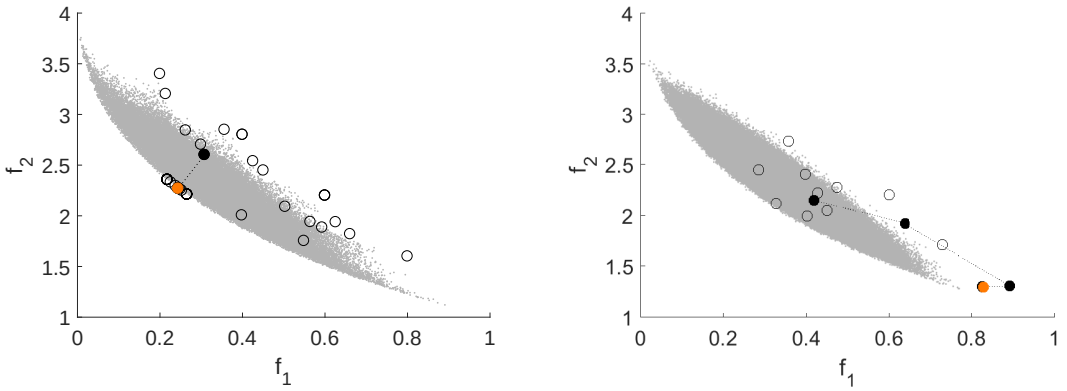

Figure 6.11 – MHT applied to (Jin1) with $n = 5$ (left) and $n = 10$ (right)

presented in Section 6.4.4. Figure 6.12 shows the results for DMS applied to (Jin1) with $n = 5$ and $n = 10$ with the same starting points as used in the runs depicted in Figure 6.11. MHT needs 42 function evaluations for $n = 5$ (21 for $n = 10$) and therefore significantly less than DMS which needs 78 evaluations (152 for $n = 10$).
Even though DMS computes many function values, it explores only the area close to the starting point. It terminates with a point close to the starting point but with a significant distance to the Pareto front.

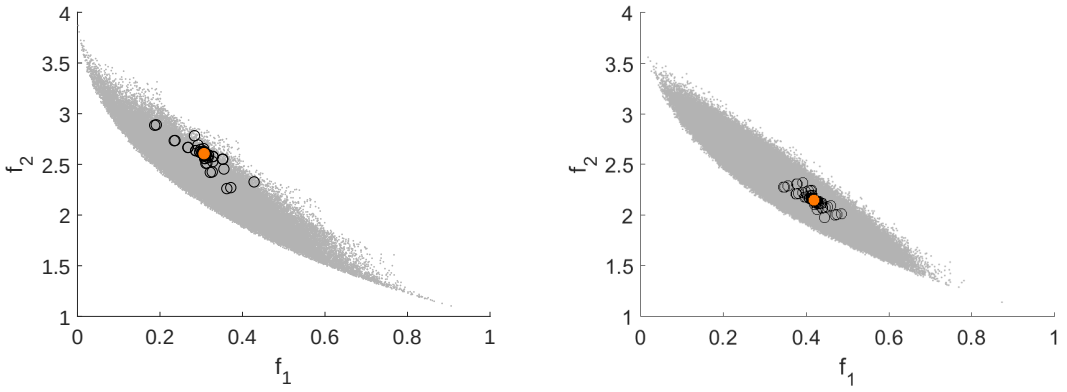

Figure 6.12 – DMS applied to (Jin1) with $n = 5$ (left) and $n = 10$ (right)

In the following, we give a short overview of the number of function evaluations required by MHT and DMS for the scalable test problems (Jin1) and (T4), see Test Problem P.34 in Appendix A.2. Table 6.2 gives an overview of the range (R) and mean value (M) of the number of function evaluations for MHT.

n	2	3	4	5	10	20	30	40	50
Jin1 (R)	11-12	20-32	15-46	42-82	15-667	46-621	39-322	63-244	109-316
Jin1 (M)	11.40	24.30	29.90	52.40	107.30	174.40	120.20	121.30	165.00
T4 (R)	12-14	21-22	31-32	43-44	68-206	252-433	399-561	609-1013	1023-1459
T4 (M)	13.20	21.40	31.40	43.70	152.10	338.00	483.70	794.40	1246.90

Table 6.2 – Function evaluations (range R and mean value M) per dimension for MHT

Table 6.2 shows that there exist instances for which a higher dimension requires less function evaluations than a lower dimension, e.g. dimension 5 and 10. This also occurred for the instances of (Jin1) presented above. The reasons are the choice of starting points and the different kinds of model functions.
For $n \geq 10$, linear model functions are used for the as expensive declared objective function instead of quadratic model functions. In general, this saves function evaluations as discussed in Section 4.6.1, but there exist individual instances for which the reverse relation holds. However, the general tendency of rising function evaluations with rising dimension is apparent in Table 6.2.
The comparison method DMS from [Cus+11] is a direct search approach. For the general direct search approach $2n$ function evaluations are required in every iteration. For DMS as it is implemented and freely available, Table 6.3 shows the range (R) and mean value (M) of function evaluations required for the two scalable test problems (Jin1) and (T4).

n	2	3	4	5	10	20	30	40	50
Jin1 (R)	30-35	47-52	63-69	78-85	152-165	307-328	469-488	622-648	767-818
Jin1 (M)	33.10	49.80	66.20	81.70	160.70	321.40	479.70	637.20	800.50
T4 (R)	54-152	71-213	117-424	181-510	706-1128	1849-2000	2000	2000	2000
T4 (M)	83.80	133.20	212.10	303.50	834.10	1998.60	2000	2000	2000

Table 6.3 – Function evaluations (range R and mean value M) per dimension for DMS

As stated in Section 6.3, for the runs with DMS the maximum number of allowed function evaluations is set to 2000. This does not suffice for some instances of (T4) as Table 6.3 shows. For these instances DMS terminates because the maximum number of function evaluations is reached, but without having computed a weakly efficient point.

If further function evaluations are allowed, DMS will probably solve these instances. Tables 6.2 and 6.3 give a first impression of how the dimension n influences MHT in comparison to DMS. As expected, the direct search approach needs significantly more function evaluations with rising dimension than MHT.
The mean values of function evaluations per dimension for all instances of the scalable test problems Test Problem P.12 (FF), Test Problems P.16 to P.19 (Jin1-Jin4) and Test Problem P.34 (T4), see Appendix A.2, are listed in Table 6.4.

n	2	3	4	5	10	20	30	40	50
MHT	16.09	39.97	72.63	109.22	129.70	256.20	301.95	457.85	705.95
DMS	46.40	79.00	122.38	174.52	834.10	1153.20	1239.80	1318.60	1400.30

Table 6.4 – Mean value of function evaluations per dimension for all scalable problems for MHT and DMS

This indicates that, compared to the direct search approach of DMS, the algorithm MHT introduced in this thesis can save function evaluations in higher dimensions. However, it is important to note that DMS does not make use of any derivative information, also not of the cheap function. Therefore, it must be expected that DMS requires in general more function evaluations than any method that uses such information.

6.4.4 Versions of Trial Point Acceptance Test

The acceptance test in MHT does not guarantee a descent for every objective function. It is only guaranteed that at least one of the function values is decreasing. This is outlined in detail in Section 4.3. However, a stricter version of the trial point acceptance test can be formulated and the convergence results still hold, see Section 4.6.3. Numerical tests with the 740 bi-objective test instances described in Section 6.2 show that in most instances there is no difference between the two versions of the trial point acceptance test and the trial points provide a decrease for every objective function. Only for few instances a difference caused by the acceptance tests occurred. To illustrate these differences, we present some test results.
Figure 6.13 shows the result of MHT for one instance of the test problem (FF) with $n = 2$ (Test Problem P.12 in Appendix A.2, already considered in Section 6.4.2 with

$n = 3$), $x^0 = (-2.67, 1.84)^\top$) with the standard version of the acceptance test (left figure, solution $\bar{x}^1 = (-0.71, -0.70)^\top$ marked orange) and the strict acceptance test (right figure, solution $\bar{x}^2 = (1.78, -0.43)^\top$ marked orange).

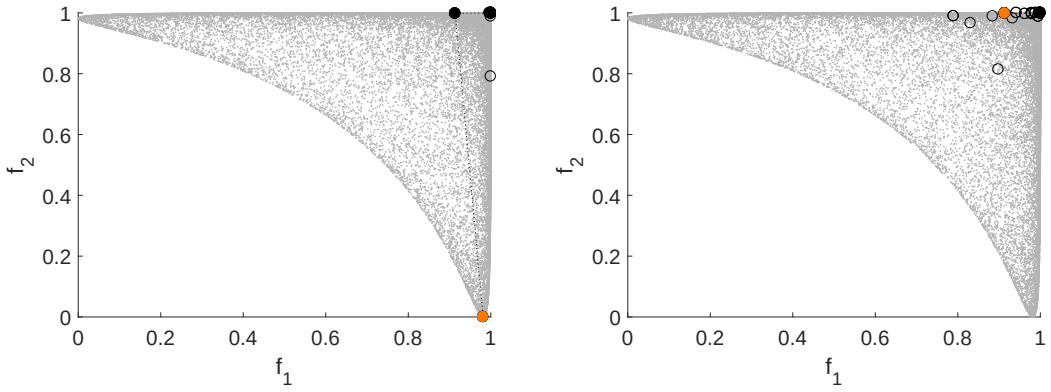

Figure 6.13 – MHT applied to (FF) ($n = 2$) with standard trial point acceptance test (left) and strict version (right)

MHT with the standard version of the acceptance test generates an efficient point after 22 function evaluations. The effect of the acceptance test is visible for the last iteration in the left of Figure 6.13. The function value of f_1 increases only slightly, yet the decrease for f_2 is significant. The obtained point $\bar{x}^1$ is efficient for (FF).

In contrast, MHT using the strict acceptance test terminates with the point $\bar{x}^2$ which is not efficient for (FF). However, the auxiliary functions ω, ω_b indicate that $\bar{x}^2$ is a stationary point for (FF) ($\omega_b(\bar{x}^2) = \omega(\bar{x}^2) = 9.50 \cdot 10^{-3}$). Nevertheless, this run of MHT requires more function evaluations (30). Thus, the weaker version of the acceptance test performs better for this instance.

There are also instances for which MHT generates similar points with both versions of the acceptance test, but the standard acceptance test requires significantly less function evaluations. An example for this is illustrated in Figure 6.14 (starting point $x^0 = (2.04, -1.33)^\top$). It depicts the result of MHT with the standard trial point acceptance test on the left ($\bar{x}^1 = (-0.50, -0.47)^\top$) and with the strict version on the right ($\bar{x}^2 = (-0.44, -0.55)^\top$).

Both runs generate points with similar values in the domain and the image space, but MHT needs 23 function evaluations with the standard acceptance test and 40 with the strict version. Thus, the stricter version of the trial point acceptance test causes more model updates. This is also visible in Figure 6.14, since more unfilled circles which

mark the points evaluated during the procedure are depicted in the right figure. However, it is not guaranteed that the less strict acceptance test requires always less function evaluations. There are also instances for which MHT with the strict version of the acceptance test requires the same amount of function evaluations or less. Based on the results with the considered test instances, no statement can be made which of the acceptance tests is more appropriate, also since no difference occurred for most test instances.
However, for applications it can be reasonable to use the strict version of the trial point acceptance test. This is for example the case for the application presented in Chapter 7. Due to the problem formulation and the long computation time for function values it is better to guarantee a descent for every objective function in every iteration.

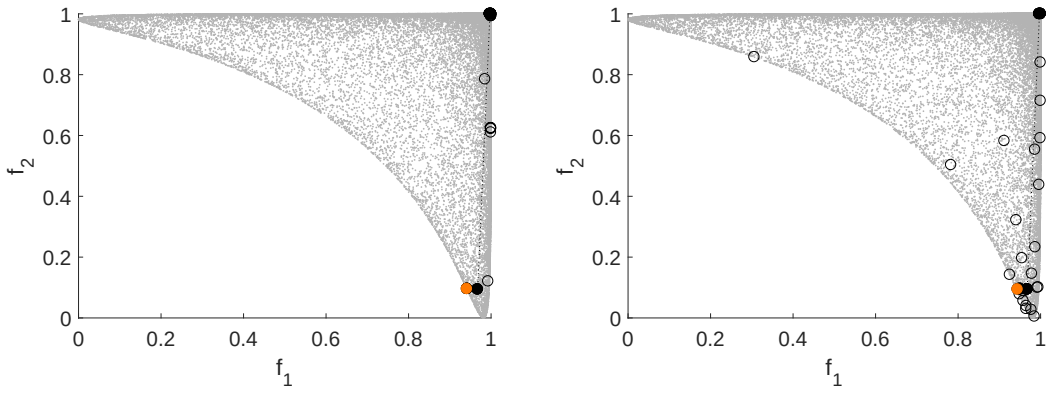

Figure 6.14 – MHT applied to (FF) ($n = 2$) with standard trial point acceptance test (left) and strict version (right)

6.4.5 Exact Model Function for Cheap Functions

Instead of using a model function for the cheap functions, it is also possible to use the cheap functions themselves, see Section 4.6.4. One would expect that this saves function evaluations for the expensive function. However, first numerical tests show only a small difference, at least for the considered test problems.
For many of the test problems listed in Appendix A.2 the objective functions declared as cheap are quadratic or linear. For such functions the model function is exact since it is chosen as the quadratic Taylor model. Thus, for the numerical tests in this section

we consider test problems for which the cheap function is not quadratic or linear. We consider the following 14 test problems with the starting points given in Appendix A.2: Test Problems P.2, P.4, P.7, P.8, P.10, P.13, P.21, P.23, P.24 and P.45 and Test Problem P.12 with $n = 2, 3, 4, 5$.

Table 6.5 lists for every test problem the range (R) and the mean value (M) of function evaluations for the expensive function until MHT terminates. The values for MHT without a model function for the cheap function are given in columns two and three. The values for MHT with a model function for the cheap function are given in columns four and five. The last row gives the range and the mean value over all considered instances. It shows that for the considered instances there is in general no difference between replacing the cheap function by a model function or not.

	model (R)	model (M)	no model (R)	no model (M)
Test Problem P.2	17-30	20.90	29-141	44.10
Test Problem P.4	11-36	19.30	12-38	19.10
Test Problem P.7	4-31	11.90	4-97	19.40
Test Problem P.8	8-23	13.50	7-39	15.20
Test Problem P.10	10-28	17.90	9-24	16.10
Test Problem P.12, $n = 2$	15-38	25.33	12-73	30.89
Test Problem P.12, $n = 3$	28-87	56.22	28-159	70.40
Test Problem P.12, $n = 4$	42-143	88.30	56-184	96.00
Test Problem P.12, $n = 5$	61-193	128.90	58-193	108.13
Test Problem P.13	12-36	22.00	13-38	23.11
Test Problem P.21	42-100	67.40	14-52	23.90
Test Problem P.23	18-54	38.20	10-42	20.90
Test Problem P.24	35-61	46.40	11-199	42.90
Test Problem P.45	29-84	36.50	29-54	43.17
all	4-193	42.34	4-199	40.95

Table 6.5 – Function evaluations for MHT without and with model function for the cheap function

However, for individual instances differences occur as the individual values for the test problems show. As expected, it is possible to save function evaluations by using the original function instead of a model. This is for example shown by the results of Test Problem P.12 ($n = 5$) or Test Problem P.23. In these cases, the exact model causes less iterations and thus less function evaluations.
The reverse result is also possible, but it only occurred for individual instances, i.e. only for some starting points of some test problems, e.g. for Test Problem P.2 or Test Problem P.12 ($n = 3$). The reasons are mostly the size of the trust region radius δ and the step size t_{PS} for the Pascoletti-Serafini problem. For some instances the value of t_{PS} is small in subsequent iterations, yet either it provides a small but sufficient decrease or does not provide a sufficient decrease but is too large for the stopping criterion described in Section 6.1.1. Both cases result in more iterations and thus more function evaluations for the expensive function.
Furthermore, the search direction differs significantly for some instances due to different values of the ideal points. This can affect the number of function evaluations positively or negatively. The negative effect is illustrated for one instance of (FF) respectively Test Problem P.12 with $n = 3$ and the starting point $x^0 = (-2.67, 1.84, 1.96)^\top$ in Figure 6.15. For this test problem function f_1 is declared as expensive.

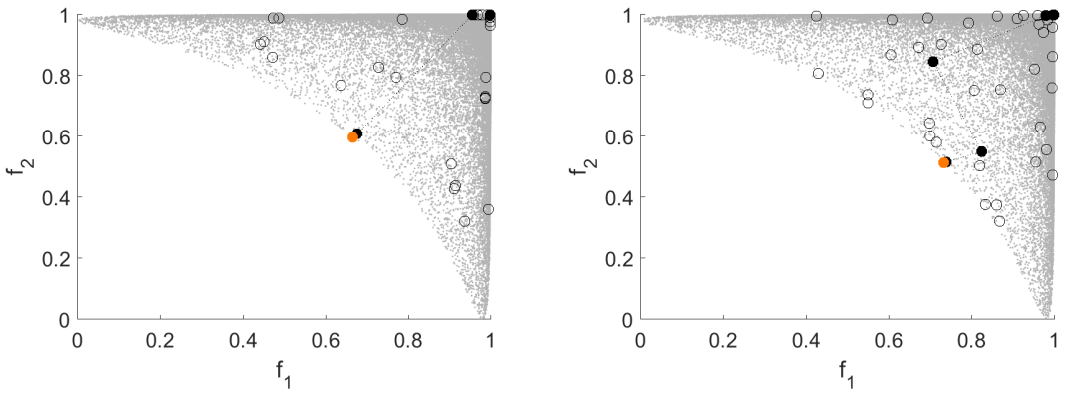

Figure 6.15 – MHT applied to Test Problem P.12 ($n = 3$) with model function for cheap function (left) and without model function (right)

The iteration points are marked black, the solution orange and the interpolation points as unfilled circles. The result in the image space for MHT in its standard version using a surrogate model for f_2 is depicted in the left figure. 65 function evaluations are required to compute the point $\bar{x}^2 = (0.00, -0.05, -0.03)^\top$. The result of MHT with

$m_2^k \equiv f_2$ for all iterations $k \in \mathbb{N}$ is depicted in the right figure. 118 function evaluations are required to compute the point $\bar{x}^1 = (-0.08, -0.09, -0.09)^\top$. The figure illustrates that the search directions differ and that in case $m_2^k \equiv f_2$ (right) more iterations are required which causes more function evaluations as the distribution of the interpolation points show. Moreover, this instance shows the influence of the trial point acceptance test which does not guarantee a descent for each component. This is discussed in detail in Sections 4.6.3 and 6.4.4.

The positive effect of using the cheap function itself instead of a model function is illustrated in Figure 6.16 for the nonconvex test problem (LE1), see also Test Problem P.23 in Appendix A.2. It is defined by

$$\min_{x \in \Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x \in [-5,10]^2} \begin{pmatrix} \sqrt[8]{x_1^2 + x_2^2} \\ \sqrt[4]{(x_1 - 0.5)^2 + (x_2 - 0.5)^2} \end{pmatrix} \tag{LE1}$$

with f_1 declared as expensive function. The starting point is $x^0 = (-3.40, 9.43)^\top$, the left figure shows the result of MHT ($\bar{x}^2 = (0.02, -0.06)^\top$) and the right figure the result of MHT with $m_2^k \equiv f_2$ for all iterations $k \in \mathbb{N}$ ($\bar{x}^1 = (0.44, 0.38)^\top$).

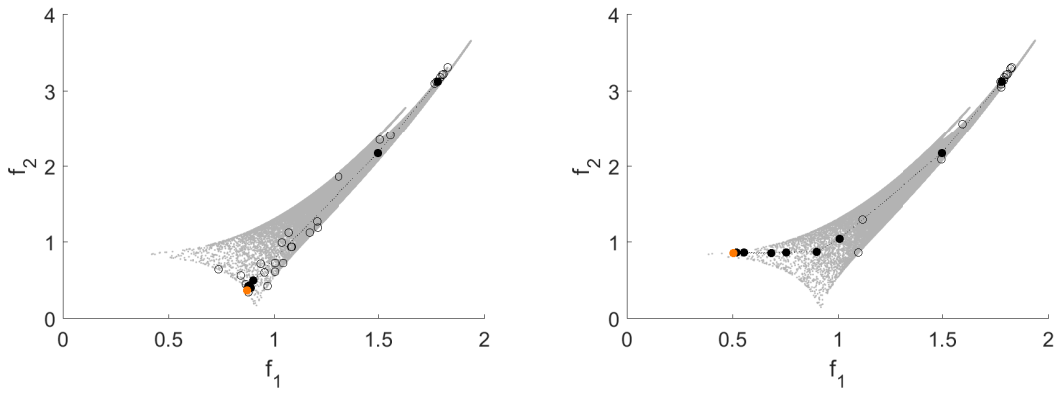

Figure 6.16 – MHT applied to (LE1) with model function for cheap function (left) and without model function (right)

For this instance, using no model for the cheap function also causes more iterations. However, this does not cause more function evaluations. The model of f_1 is still reliable and a sufficient decrease can be generated in most iterations. MHT without model function for f_2 requires 22 function evaluations and therefore less than MHT with

model for f_2 which requires 38 function evaluations. Moreover, the search directions in the iterations differ significantly and the computed solutions (marked orange) are from different areas of the Pareto front.

6.4.6 Performance Profiles

In this section a general overview of the outcome of the test runs with MHT compared to the methods DMS and EFOS is presented. For classifying the test runs as successful or not successful the distance to the Pareto front is used. If it falls below a problem dependent constant, the test run for an instance is classified as solved. To compare the performance of the algorithms, the number of function evaluations for the as expensive declared function is counted until the algorithm terminates. Figure 6.17 shows a performance profile for all 340 convex test instances in full range on the left and zoomed in on the right. One instance is defined as one test problem from Appendix A.2 with one of the starting points listed there.

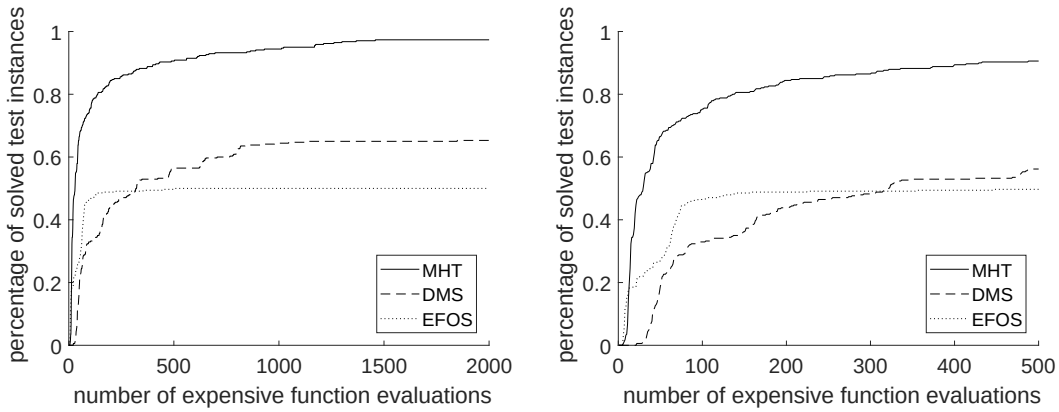

Figure 6.17 – Performance Profile for MHT, DMS and EFOS for 340 convex instances

If up to 500 function evaluations are allowed for the expensive function, DMS and EFOS behave similar. With further function evaluations DMS is capable of solving further test instances whereas EFOS stagnates and cannot solve more instances. In general, Figure 6.17 shows that MHT needs less function evaluations than EFOS and DMS to solve the convex test problems.

MHT solves 331 of 340 instances with at most 1459 function evaluations of the expensive function. For the instances that are not classified as solved either the distance to the Pareto front was too large, MHT stopped due to an internal error or the maximum number of function evaluations did not suffice.
The high number of 1459 function evaluations is due to the high dimensional test instances that are included. If considering only test instances up to dimension 10, all convex instances (except two instances for which MHT stopped due to an internal error) are solved by MHT after 667 function evaluations as the performance profile in Figure 6.18 shows (full range left, zoomed in right). With the same amount of function evaluations (667) DMS solves 66.54% and EFOS 63.46 % of the convex test problems up to dimension 10.

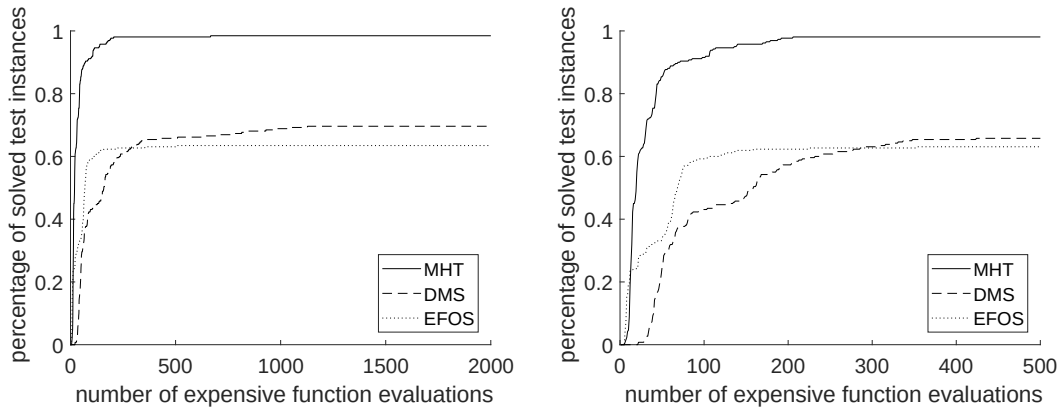

Figure 6.18 – Performance Profile for MHT, DMS and EFOS for 260 convex instances, $n \leq 10$

As already noted, MHT is a local method and guarantees only a necessary condition for local weak efficiency. Additionally, this is only proved for unconstrained problems. MHT can handle box constraints, but the convergence results do not transfer, see Section 4.7. Furthermore, for nonconvex optimization problems local optimality is in general not synonymous to global optimality. The point obtained by MHT does not necessarily have a small distance to the Pareto front, it can also be only a local weakly efficient point or it can only fulfill a necessary condition for local weak efficiency. This must be considered when evaluating the results of MHT for the nonconvex test problems.
Thus, for a performance profile over all considered test examples we do not only use the distance to the Pareto front to classify test instances as solved, but complement

it with a measure for local optimality. The auxiliary function $\omega : \mathbb{R}^n \to \mathbb{R}$ given in Definition 2.10 in Section 2.2 and defined by

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

for continuously differentiable functions $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \dots, q$, characterizes Pareto criticality which is a necessary condition for local weak efficiency for unconstrained optimization problems (MOP). According to Lemma 2.11 from Section 2.2, a point $x \in \mathbb{R}^n$ is Pareto critical for (MOP) if and only if it holds $\omega(x) = 0$.
Furthermore, the auxiliary function $\omega_b : \mathbb{R}^n \to \mathbb{R}$ given in (4.53) in Section 4.7 and defined by

$$\omega_b(x) = -\min_{d \in L(x), \|d\| \leq 1} \max_{i=1,\dots,q} \nabla f_i(x)^\top d$$

with $L(\bar{x}) = \{d \in \mathbb{R}^n \mid d_i \geq 0, \text{ if } x_i = lb_i, d_i \leq 0, \text{ if } x_i = ub_i, i = 1, 2, \dots, q\}$ characterizes stationary points, the analogous condition to Pareto criticality for box constrained optimization problems with $lb \in (\mathbb{R} \cup \{-\infty\})^n$, $ub \in (\mathbb{R} \cup \{\infty\})^n$ the lower and upper bounds. It is a necessary condition for local weak efficiency and a point $x \in \mathbb{R}^n$ is a stationary point for a box constrained optimization problem if and only if it holds $\omega_b(x) = 0$, see Lemmas 4.29 and 4.32 in Section 4.7.
Consequently, given $\bar{x}$ the solution generated by one of the considered algorithms (MHT, DMS, EFOS), we classify a test instance as solved if either the distance of $f(\bar{x})$ to the Pareto front is small enough or if it holds $\omega(\bar{x}) \leq \varepsilon$ for unconstrained optimization problems respectively $\omega_b(\bar{x}) \leq \varepsilon$ for box constrained optimization problems. We chose $\varepsilon = 0.1$ for the data analysis.
Using these classifications, the performance profile in Figure 6.19 shows the percentage of solved test instances depending on the required function evaluations for MHT, DMS and EFOS. All 780 test instances are considered for this performance profile. The full range is shown on the left and on the right the performance profile is zoomed in to 500 function evaluations.
Figure 6.19 illustrates that by applying MHT 95.90% of all test instances are solved. Within the same number of function evaluations (1459) EFOS solves 57.56% and DMS solves 82.95% of all considered test instances. Thus, MHT solves more test instances than the comparison methods and requires less function evaluations.

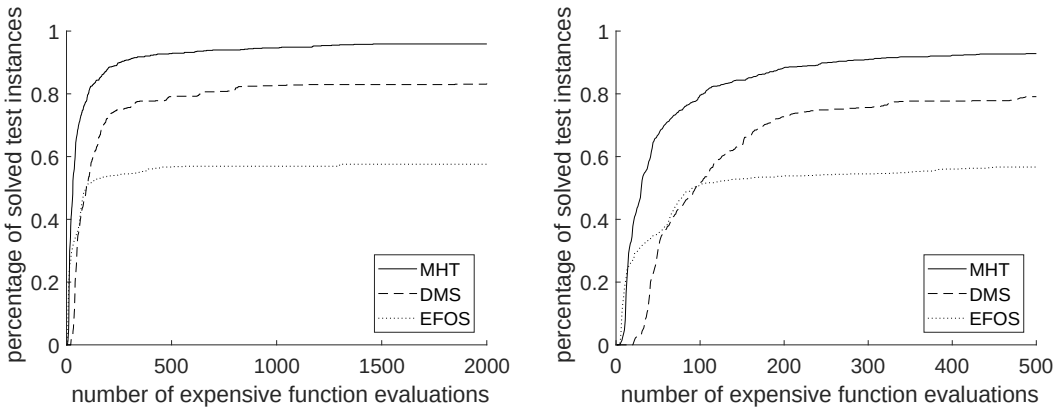

Figure 6.19 – Performance Profile for MHT, DMS and EFOS for all 780 instances

6.5 Numerical Results of Heuristic Approaches Based on MHT

In this section we present results for the heuristic approaches $\text{MHT}_{\text{spread}}$, $\text{MHT}_{\text{split}}$ and $\text{MHT}_{\text{combi}}$ from Chapter 5. For these approaches not all of the test problems listed in Section 6.2 and Appendix A.2 have been considered. Only first numerical test have been executed with the following subset of 20 test problems: Test Problems P.1, P.2, P.15, P.23, P.30, P.31 and P.39, Test Problem P.12 (n=2,3,4,5), Test Problem P.16 (n=2,3,5), Test Problem P.25 (constrained and unconstrained), Test Problem P.28 (constrained and unconstrained) and Test Problem P.34 (n=2,3).

In the following subsections, we present results exemplarily for two test problems and discuss the influence of the parameters (splitting number and spreading distance) observed for the considered instances. At first, the three approaches $\text{MHT}_{\text{spread}}$, $\text{MHT}_{\text{split}}$ and $\text{MHT}_{\text{combi}}$ are considered individually in Sections 6.5.1 to 6.5.3 and in Section 6.5.4 the approaches are compared based on the test results.

Moreover and as explained in the end of Section 5.3, it is reasonable to test the sets of points that are the output of $\text{MHT}_{\text{spread}}$ and $\text{MHT}_{\text{combi}}$ for dominated points. For the presentation of the results hereafter, the obtained points are compared pairwise to delete dominated points from the set of solutions. Nevertheless, to illustrate the behavior of the different algorithms, all obtained points are integrated in the figures.

6.5.1 Spreading via Individual Minima

The spreading technique from Section 5.1 starts with an initial run of the basic version of MHT. Starting from the resulting point $\bar{x} \in \Omega$, the cheap function is subsequently minimized in moving local areas. The resulting points are either recognized as (locally) weakly efficient or chosen as starting points for MHT. To illustrate this spreading technique, we consider again the convex test problem (BK1) (Test Problem P.1 in Appendix A.2) given by

$$\min_{x \in \Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x \in [-5,10]^2} \begin{pmatrix} x_1^2 + x_2^2 \\ (x_1 - 5)^2 + (x_2 - 5)^2 \end{pmatrix} \tag{BK1}$$

with f_1 declared as expensive function. Figure 6.20 shows the result of $\text{MHT}_{\text{spread}}$ 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, $\text{MHT}_{\text{spread}}$ requires 35 function evaluations to compute 6 efficient points.

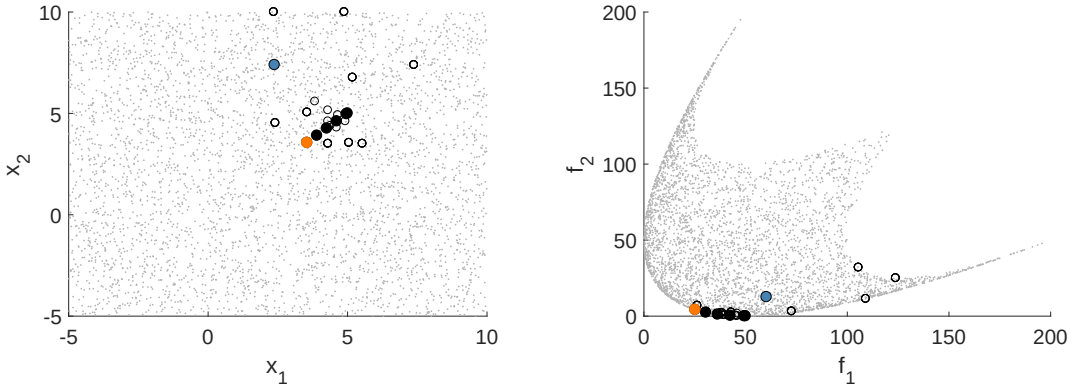

Figure 6.20 – $\text{MHT}_{\text{spread}}$ applied to (BK1) (domain left, image space right)

The result of the spreading approach depends on the spreading distance and on the result $\bar{x}$ of the initial run of MHT and therefore on the starting point x^0. The closer $\bar{x}$ is to the global minimum of the cheap function, in this example function f_2, the less points can be computed. Additionally, the larger the spreading distance δ, the bigger

is the distance between the computed points and the less points can be computed. The influence of the initial run of MHT is illustrated by comparing Figure 6.20 to Figure 6.21. The latter shows the result of $\text{MHT}_{\text{spread}}$ 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.

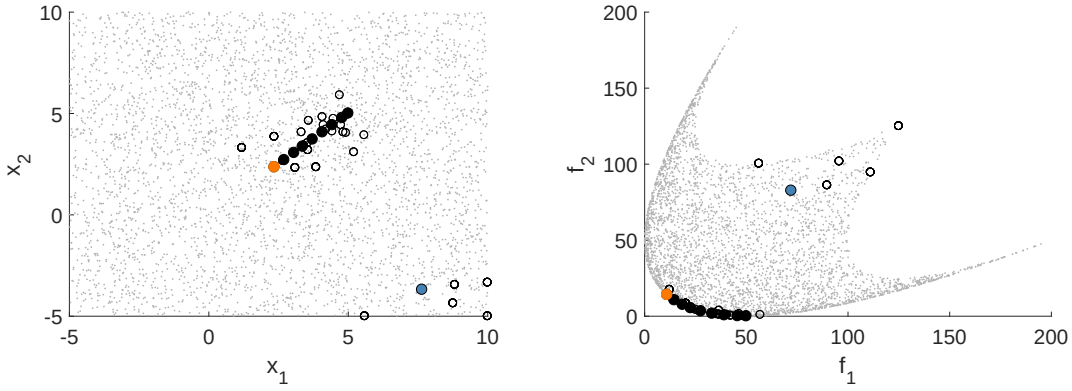

Figure 6.21 – $\text{MHT}_{\text{spread}}$ applied to (BK1) (domain left, image space right)

Table 6.6 lists the number of function evaluations per obtained efficient point for these two runs of $\text{MHT}_{\text{spread}}$ with (BK1). The first row lists the numbers of function evaluations of the first run depicted in Figure 6.20. The second row lists the numbers of function evaluations for the run depicted in Figure 6.21.

Solution of $\text{MHT}_{\text{spread}}$	1	2	3	4	5	6	7	8	9
$x^0 = (2.38, 7.39)^\top$, Figure 6.20	12	7	2	12	1	1	-	-	-
$x^0 = (7.645, -3.69)^\top$, Figure 6.21	13	5	2	3	5	10	13	4	1

Table 6.6 – Function evaluations per efficient point of (BK1) obtained by $\text{MHT}_{\text{spread}}$ with $\delta = 0.5$ and different starting points x^0

This table shows that different amounts of function evaluations are needed per efficient point produced by MHT. In some cases only few function evaluations are required, either because former interpolation points can be reused or because the starting point for MHT is already located well. Additionally, the last listed solution always needs only one function evaluation. This point is always the global minimum of the cheap function. According to Lemma 5.1 (ii), it is weakly efficient and therefore always added to the list of solutions.

The influence of different values for the spreading distance is illustrated for the non-convex test problem (LE1), see also Test Problem P.23 in Appendix A.2. This optimization problem was already considered in Section 6.4.5. It is defined as

$$\min_{x\in\Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x\in[-5,10]^2} \begin{pmatrix} \sqrt[8]{x_1^2 + x_2^2} \\ \sqrt[4]{(x_1 - 0.5)^2 + (x_2 - 0.5)^2} \end{pmatrix} \tag{LE1}$$

with f_1 declared as expensive function. Figure 6.22 shows four runs of $\text{MHT}_{\text{spread}}$ with the same starting point $x^0 = (1.47, 8.66)^\top$ but with different spreading distances δ.

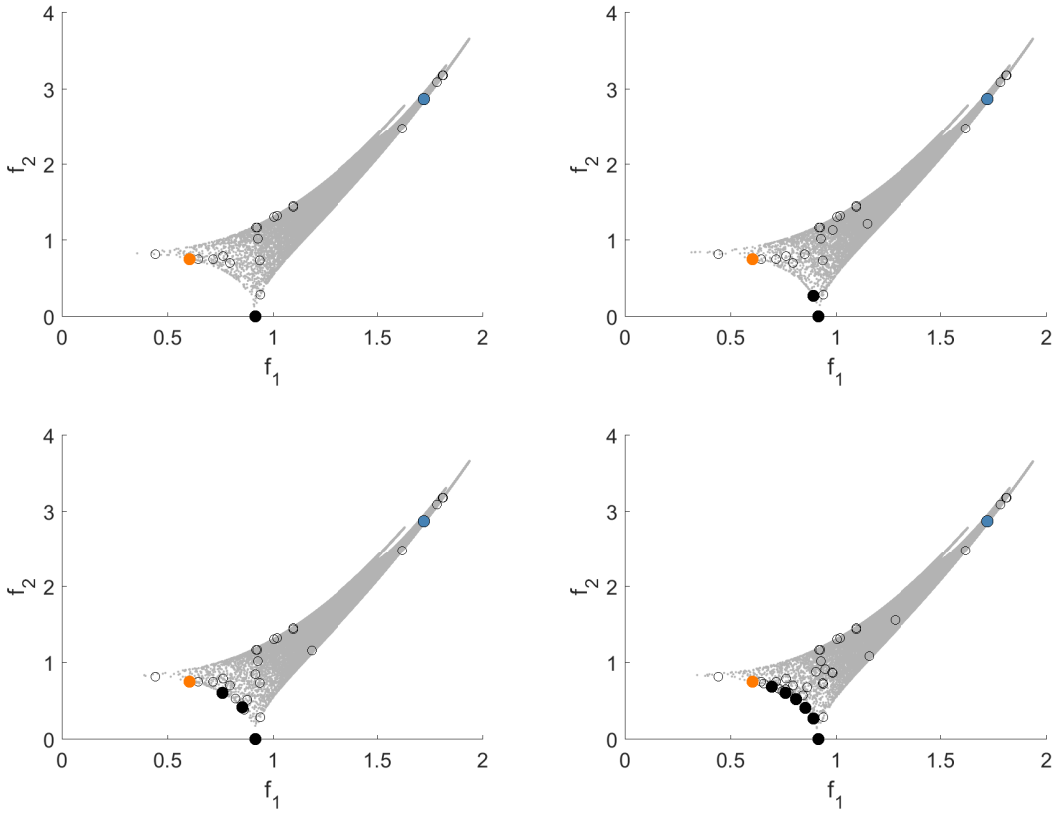

Figure 6.22 – $\text{MHT}_{\text{spread}}$ applied to (LE1) with spreading distances $\delta = 1$ (top left), $\delta = 0.5$ (top right), $\delta = 0.2$ (bottom left), $\delta = 0.1$ (bottom right)

The top left figure 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 minimum of f_2 and a point very close to it are computed. They cannot be distinguished in the

figure. Since the global minimum of f_2 defines the stopping criterion of $\text{MHT}_{\text{spread}}$, it is computed in every run of $\text{MHT}_{\text{spread}}$. 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, four 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 minimum 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).
Table 6.7 gives an overview of the number of function evaluations required to obtain the efficient points that are the output of $\text{MHT}_{\text{spread}}$ applied to (LE1) with the different spreading distances, but the same starting point.
As for the first test example, the table shows that in the subsequent runs of MHT done during the procedure, former interpolation points are reused and by this, the overall number of function evaluations is reduced. Furthermore, the initial run of MHT needs always 25 function evaluations. Since the starting point is equal for all four runs and only the spreading distance differs, the result of this initial run is always the same point. This is also illustrated in Figure 6.22 since the orange marked point is identical in all four figures.

Solution of $\text{MHT}_{\text{spread}}$	1	2	3	4	5	6	7	8
$\delta = 1$, Figure 6.22 (top left)	25	1	1	-	-	-	-	-
$\delta = 0.5$, Figure 6.22 (top right)	25	7	1	1	-	-	-	-
$\delta = 0.2$, Figure 6.22 (bottom left)	25	10	3	1	1	-	-	-
$\delta = 0.1$, Figure 6.22 (bottom right)	25	17	10	7	7	2	1	1

Table 6.7 – Function evaluations per efficient point of (LE1) obtained by $\text{MHT}_{\text{spread}}$ with $x^0 = (1.4712, 8.6597)^\top$ and different spreading distances δ

Moreover, this spreading technique illustrates that MHT can also produce dominated points as output. For this purpose, we still consider the test problem (LE1). Figure 6.23 shows the result of $\text{MHT}_{\text{spread}}$ with spreading distance $\delta = 0.5$ and the starting point

$x^0 = (-4.93, 6.62)^\top$ (domain left, image space right). In total, 41 function evaluations are required for this run. The result of the initial run of MHT is $\bar{x} = (1.11, -0.55)^\top$ with the function values $f_1(\bar{x}) = 1.06$ and $f_2(\bar{x}) = 1.10$, marked orange in the figures. As the right figure illustrates, $\bar{x}$ is not efficient for (LE1) and furthermore, the black points resulting from the spreading approach, dominate it.
Moreover, this point is not a stationary point for (LE1). The box constraints are inactive for $\bar{x}$. Thus, it holds $\omega_b(\bar{x}) = \omega(\bar{x})$ for the functions ω, ω_b characterizing necessary optimality conditions in the unconstrained case respectively the box constrained case (Pareto critical/stationary points), see Sections 2.2 and 4.7. If and only if the function value is equal to zero, the point fulfills a necessary condition for local weak efficiency. It holds $\omega_b(\bar{x}) = \omega(\bar{x}) = 0.21$. Thus, $\bar{x}$ is not a stationary point for (LE1) and MHT terminated incorrectly. Nevertheless, the further computed points which are marked black are efficient points.

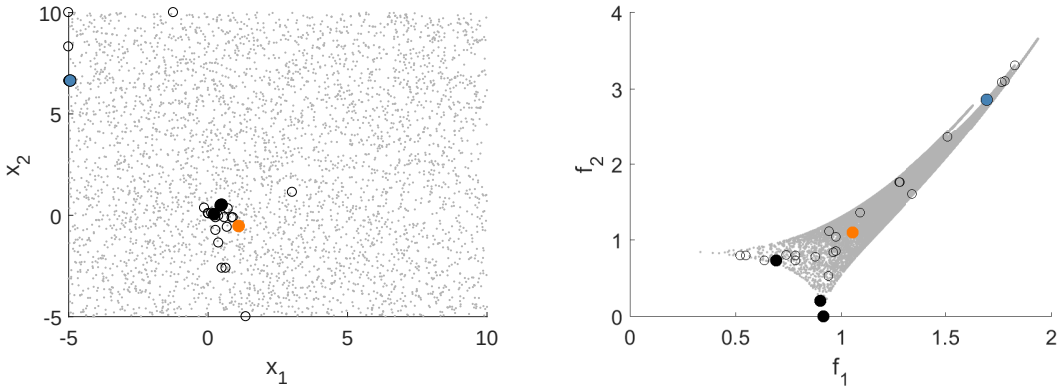

Figure 6.23 – $\text{MHT}_{\text{spread}}$ applied to (LE1): Detecting dominated point generated by MHT

6.5.2 Image Space Split

The spreading technique described in Section 5.2.4 splits the image space in disjoint search regions A_j, $j = 1, 2, \ldots, p$. They are defined by

$$\begin{aligned} A_j &= \{x \in \Omega \mid a_{j-1} \leq f_2(x) \leq a_j\} \quad \text{with} \\ a_0 &= f_2^{\min}, \quad a_j = f_2^{\min} + j\frac{f_2^{\max} - f_2^{\min}}{p}, \\ f_2^{\min} &= \min_{x \in \Omega} f_2(x), \quad f_2^{\max} = \max_{x \in \Omega} f_2(x) \end{aligned}$$

for $j = 1, 2, \ldots, p$ and with Ω either defined as $\mathbb{R}^n$ or by the box constraints of the considered test problems given in Appendix A.2. After splitting the image space, MHT_{lb} is applied on the areas A_j for all $j = 1, 2, \ldots, p$. The starting points for runs with MHT_{lb} are computed by the approach presented in Section 5.2.5. Thus, the starting point $x^{0,j}$ for the search area A_j is defined by

$$x^{0,j} \in \operatorname{argmin} \{f_2(x) \mid x \in \Omega \text{ and } f_2(x) \geq a_j - \varepsilon\}$$

with $\varepsilon = 0.001$. The difficulty of applying $\text{MHT}_{\text{split}}$ lies in choosing an appropriate number p of split regions. Moreover, not in every of these regions a (weakly) efficient point needs to exist. This is discussed in Section 5.2.4. As outlined there, the runs of MHT_{lb} are executed consecutively and the search areas are considered in ascending order $A_1, A_2, \ldots, A_p$. Thus, the distance to the global minimum of f_2 increases with every new run of MHT_{lb}.

After every run of MHT_{lb} it is tested if the computed solution is dominated by the previously computed solutions. In this case, the splitting approach is terminated. It is assumed that no further nondominated points will be computed when considering the remaining search areas.

However, this does not always apply as for example for (Deb513) (Test Problem P.5 in Appendix A.2) which has a discontinuous Pareto front. This optimization problem was already considered in Section 6.4.2, see page 147; f_2 is declared as expensive function. The result of an image space split with $p = 10$ is illustrated in Figure 6.24. The three orange marked points are the result of the runs of MHT_{lb} in the search areas A_1, A_2 and A_3. The unfilled circles mark the points that are evaluated during the procedure.

The splitting approach is terminated after the third area since the thereby computed point is dominated by the previous solutions. As Figure 6.24 shows, there exist further nondominated points in the areas $A_4, A_5, \ldots, A_{10}$.
Thus, the result of the heuristic test to terminate the splitting approach is disadvantageous in this case. However, two efficient points are computed by this approach. Moreover, this is an individual case and in the majority of the considered instances the heuristic termination criterion is reasonable and works well.

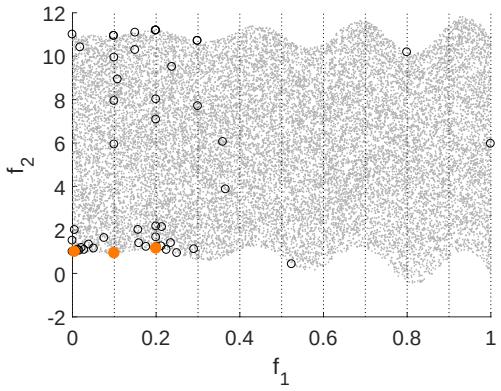

Figure 6.24 – Image space split for (Deb513) with $p = 10$ regions

To illustrate the splitting approach of $\text{MHT}_{\text{split}}$, we consider the quadratic, convex test problem (BK1) (Test Problem P.1, see also page 166) and the nonconvex problem (LE1) (Test Problem P.23, see also page 168). In general, the higher the splitting number p is, the more efficient points are computed and the more function evaluations are required. Still, interpolation points are reused if possible. As outlined in Section 5.2, the computation of the interpolation points is not restricted to the search areas A_j, $j = 1, 2, \ldots, p$. They are chosen from the entire feasible set Ω.
For the test problem (BK1) we consider the three numbers of split regions $p = 4, p = 8$ and $p = 22$. The result of $\text{MHT}_{\text{split}}$ for $p = 4$ is depicted in Figure 6.25. 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_{lb} produces a nondominated point. Moreover, the left figure shows that MHT_{lb} is not executed in the last search region. After the run in the third search area, the

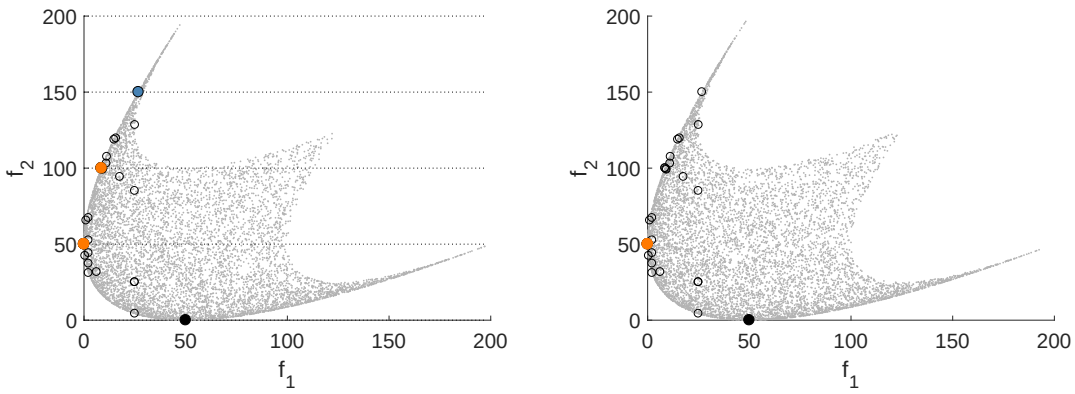

Figure 6.25 – $\text{MHT}_{\text{split}}$ applied to (BK1) with $p = 4$ regions: all computed solutions (left) and nondominated solutions (right)

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 last search area.

Two of the three starting points computed for the search areas are not visible in the left figure. They are close to the orange marked points with which MHT_{lb} respectively MHT terminates. Thus, these runs required only few function evaluations. Overall, 25 function evaluations are required to compute two efficient points.

This result of the splitting approach shows that if the number p of areas is too small, only one point is computed in addition to the global minimum of the cheap function. The global minimum is always computed for the image space split since it is one component to define the search areas. It is weakly efficient, see Lemma 5.1 (ii) in Section 5.1 and thus always added to the list of solutions given as output of $\text{MHT}_{\text{split}}$. The results of $\text{MHT}_{\text{split}}$ for the larger splitting number $p = 8$ are depicted in Figure 6.26.

The behavior is similar to the run with $p = 4$ depicted in Figure 6.25. For the test run with $p = 8$, three efficient points are computed after 30 function evaluations. Again, not every run of MHT_{lb} produces a nondominated point. The splitting procedure is terminated after the fourth search area since the therein computed solution is dominated by the previously computed solutions.

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 6.27 shows the result of the splitting approach with $p = 22$ search areas. The domain is depicted in the left

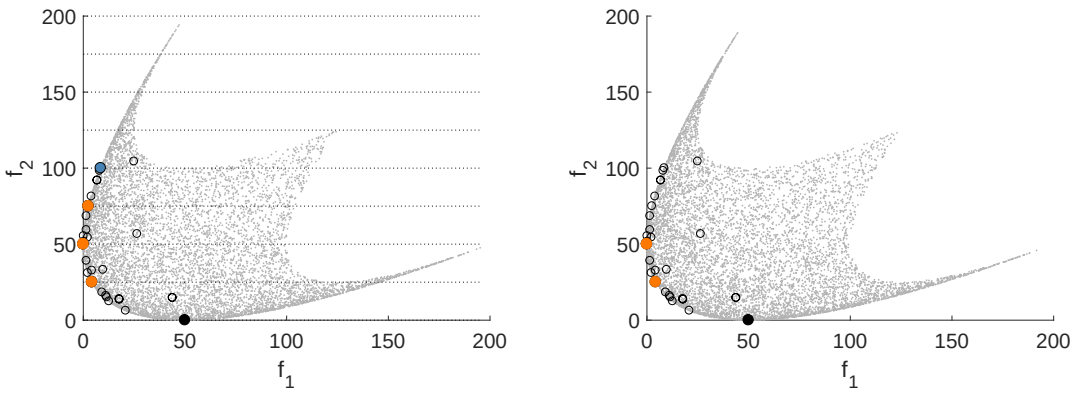

Figure 6.26 – $\text{MHT}_{\text{split}}$ applied to (BK1) with $p = 8$ regions: all computed solutions (left) and nondominated solutions (right)

figure and the image space in the right figure. In this test of $\text{MHT}_{\text{split}}$, 7 efficient points are computed with 46 function evaluations.
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.

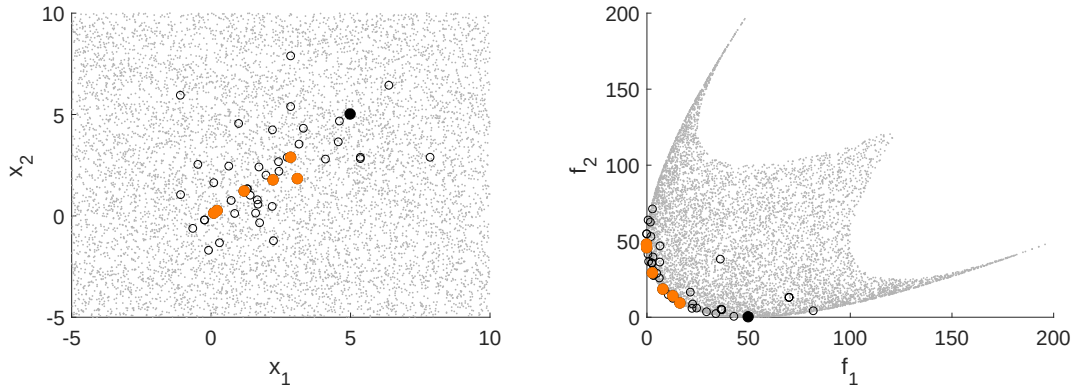

Figure 6.27 – $\text{MHT}_{\text{split}}$ applied to (BK1) with $p = 22$ (domain left, image space right)

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 6.8 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 6.8 – $\text{MHT}_{\text{split}}$: 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.

To illustrate $\text{MHT}_{\text{split}}$ also for nonconvex optimization problems, we consider the test problem (LE1) (Test Problem P.23 in Appendix A.2) with different numbers of search areas. Figure 6.28 illustrates $\text{MHT}_{\text{split}}$ for (LE1) with $p = 6$. As in the runs with (BK1), MHT_{lb} is not executed in every search area, but only in the first two.

The left figure shows that the two solutions obtained by MHT_{lb} are situated close to each other on the boundary of the search areas. Both are located close to the Pareto front, yet one dominates the other. The right figure shows the efficient points which remain after the dominated point is deleted. These two points are the output of $\text{MHT}_{\text{split}}$. In total, 38 function evaluations are required.

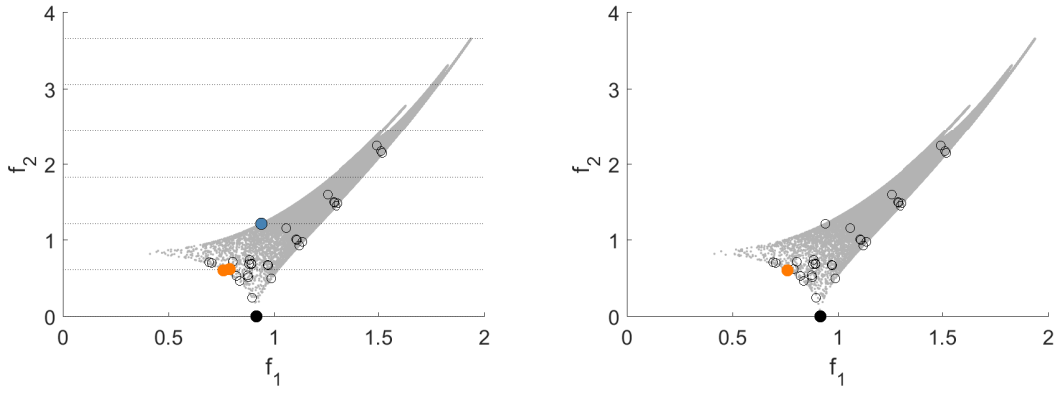

Figure 6.28 – $\text{MHT}_{\text{split}}$ applied to (LE1) with $p = 6$ regions: all computed solutions (left) and nondominated solutions (right)

By choosing a larger number p, more efficient points can be computed. Figure 6.29 shows the result of $\text{MHT}_{\text{split}}$ with $p = 10$. With a total amount of 54 function evaluations, 4 efficient points are computed. Again, MHT_{lb} is not applied in all of the 10

search areas, but the splitting approach is terminated after the fourth search area. The point produced in the last run of MHT_{lb} is dominated by the previous computed orange marked points. This can be seen in the left image of Figure 6.29.

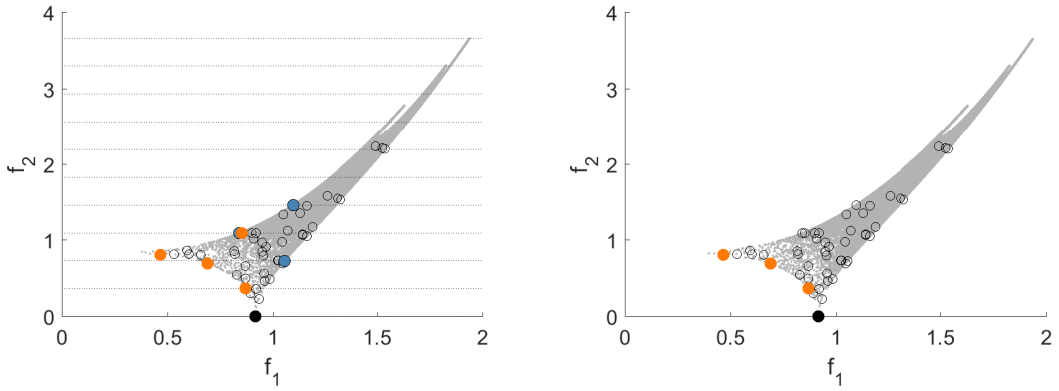

Figure 6.29 – $\text{MHT}_{\text{split}}$ applied to (LE1) with $p = 10$ regions: all computed solutions (left) and nondominated solutions (right)

Table 6.9 gives an overview of the number of obtained efficient points n_1 and the required function evaluations n_2 when considering different values for p. Similarly to Table 6.8, the general tendency is that for higher values of p more efficient points are computed and more function evaluations are required. Again, this does not apply to all instances since the location of the search areas and the starting points can also cause the reverse correlation for individual instances.

p	2	4	6	8	10	12	14	16	18	20	30	40
n_1	2	3	2	3	4	4	4	4	5	6	8	9
n_2	35	52	46	43	54	48	24	39	33	41	50	50

Table 6.9 – $\text{MHT}_{\text{split}}$: Numbers n_1 of efficient points and n_2 of function evaluations for (LE1) for different splitting numbers p

6.5.3 Combination of Image Space Split and Spreading

The heuristic approximation of the Pareto front as introduced in Section 5.3 consists of an image space split with $\text{MHT}_{\text{split}}$ and separate runs of MHT_{lb} and the spreading technique of $\text{MHT}_{\text{spread}}$. Thus, a number p of search areas for the image space split and a spreading distance δ need to be chosen. The difficulty for this approach is to choose appropriate values for these parameters.
As illustrated in Section 6.5.2, 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 Section 6.5.1, 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 $\text{MHT}_{\text{combi}}$.
As outlined in Section 5.3, for the implementation of $\text{MHT}_{\text{combi}}$ 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 $\text{MHT}_{\text{split}}$. It is important to note that the computation of the interpolation points is not restricted to the search areas of the splitting approach. They are chosen from the whole feasible set Ω.
To illustrate the procedure of $\text{MHT}_{\text{combi}}$, we consider again the two test problems (BK1) and (LE1) (Test Problems P.1 and P.23 in Appendix A.2). The effects of different choices of the parameters p and δ are illustrated.
The result of $\text{MHT}_{\text{combi}}$ applied to (BK1) with $p = 8$ and $\delta = 1$ is shown in the image space in Figure 6.30. The points obtained by MHT_{lb} due to the image space split are marked orange and the starting points for the search areas are marked blue. The points resulting from the spreading with $\text{MHT}^{\text{b}}_{\text{spread}}$ are marked black. All points evaluated during the procedure are depicted as unfilled circles. The boundaries of the search regions are represented as dotted lines. The left figure shows the image space split and all obtained solutions of MHT_{lb} and $\text{MHT}^{\text{b}}_{\text{spread}}$ and the right figure the nondominated points among them. In total, 10 efficient points are computed and 122 function evaluations are required.

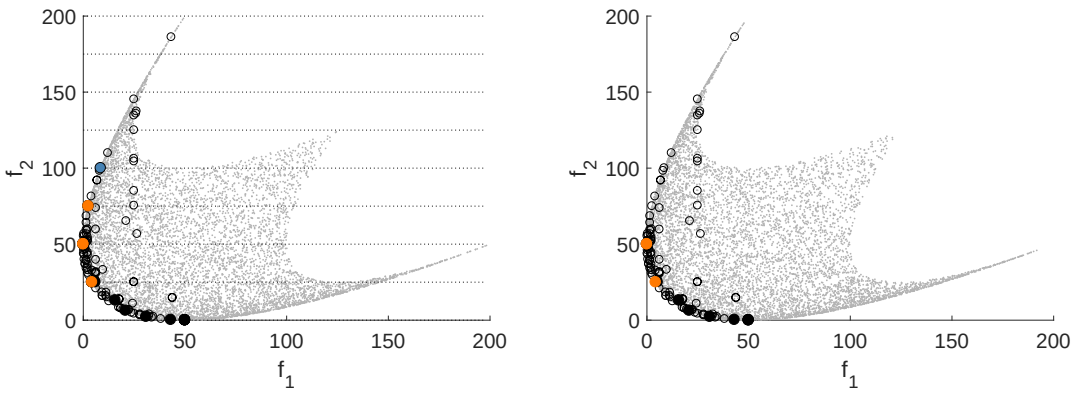

Figure 6.30 – $\text{MHT}_{\text{combi}}$ applied to (BK1) with $p = 8$ and $\delta = 1$

With the same spreading distance of $\delta = 1$, but with $p = 14$ search areas for the image space split, 13 efficient points are computed requiring 94 function evaluations. Although more search regions are considered and more efficient points are computed, less function evaluations are required. An overview of the number of obtained efficient points n_1 and required function evaluations n_2 for (BK1) with different values for the parameters p and δ is given in Table 6.10. The entries of the cells are given in the order n_1/n_2.

δ \ p	4	6	8	10	12	14	16	18	20
1	9/108	11/79	10/122	14/99	12/81	13/94	17/93	13/79	18/109
0.5	14/121	14/96	13/133	17/125	18/121	17/106	21/120	17/91	22/112

Table 6.10 – $\text{MHT}_{\text{combi}}$: Number of efficient points / number of function evaluations for (BK1) depending on splitting number p and spreading distance δ

It shows that the general and expected tendency of an increasing number of function evaluations and efficient points when increasing p or decreasing δ does not always apply. This occurred also for the splitting respectively spreading approach applied separately, see Sections 6.5.1 and 6.5.2. The main reason for this is the computation of the model functions. Interpolation points are reused whenever possible. If for any choice of parameters less function evaluations are required, the main reason is that either less model functions are computed or many interpolation points are reused.

Additionally, the position of the starting point for each run of MHT_{lb} is an important factor. If it is already close to a (weakly) efficient point, MHT_{lb} terminates after few iterations and function evaluations. Furthermore, the higher the value of p is, the more likely is the existence of search areas that do not contain (weakly) efficient points and are not considered during the splitting approach. Thus, a higher number of p does not necessarily imply a higher number of function evaluations.
This is for example illustrated by the test problem (BK1), the fixed spreading distance $\delta = 0.5$ and varying values for p. For $p = 8$, in total 13 efficient points are computed using 133 function evaluations, whereas 17 efficient points are computed with 106 function evaluations for $p = 14$.
The result of the latter run is shown in Figure 6.31. The left figure shows the domain and the right figure the image space. The efficient points obtained by the image space split are marked orange and the results from the spreading are marked black. All other evaluated points are marked as unfilled circles. The obtained nondominated points generate a good approximation of the Pareto front.

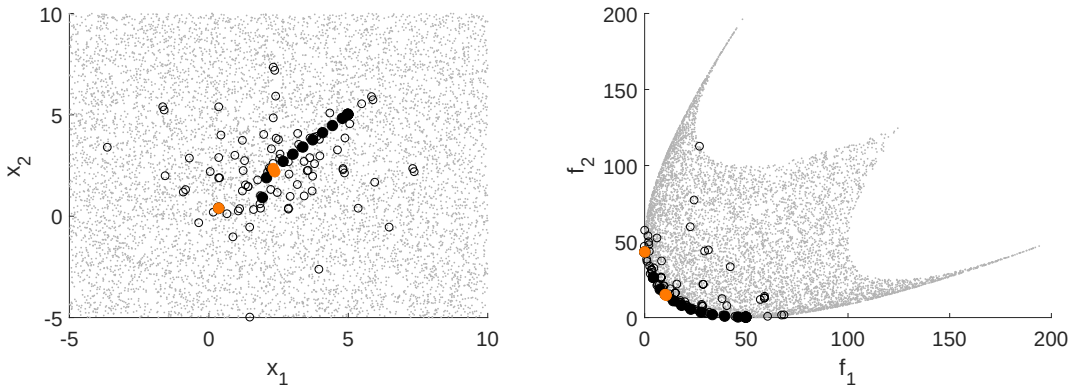

Figure 6.31 – $\text{MHT}_{\text{combi}}$ applied to (BK1) with $p = 14$ and $\delta = 0.5$

For the nonconvex test problem (LE1) the results are similar. Figure 6.32 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.

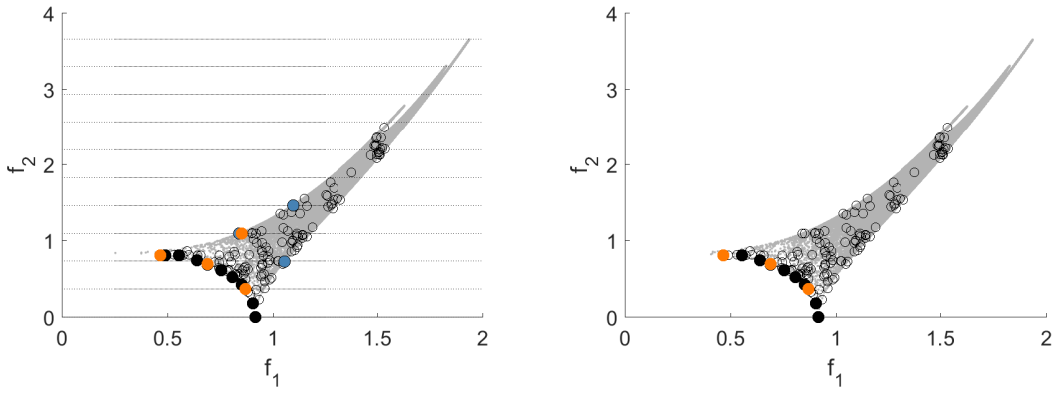

Figure 6.32 – $\text{MHT}_{\text{combi}}$ applied to (LE1) with $p = 10$ and $\delta = 0.1$

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 6.11 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 6.11 – $\text{MHT}_{\text{combi}}$: Number of efficient points / number of function evaluations for (LE1) depending on splitting number p and spreading distance δ

These results confirm the findings from the test runs with (BK1) 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 6.11. 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 $\text{MHT}_{\text{combi}}$.

This is exemplarily illustrated for two runs of $\text{MHT}_{\text{combi}}$ in Figure 6.33. 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.

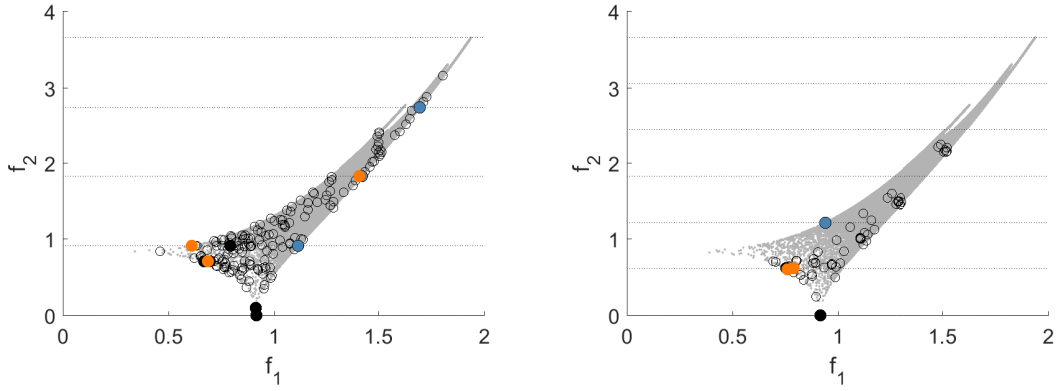

Figure 6.33 – $\text{MHT}_{\text{combi}}$ applied to (LE1) with $\delta = 0.5$ and $p = 4$ (left) and $p = 6$ (right)

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 MHT_{lb}.

Although there exist nondominated points in the second search area for $p = 6$, 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 MHT_{lb} terminates.

Additionally, Figure 6.33 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 $\text{MHT}_{\text{spread}}$ – are depicted in the figures.

6.5.4 Comparison of Heuristic Approaches

We have considered 20 test problems of the 78 test problems described in Section 6.2 and Appendix A.2 to test the different approaches of $\text{MHT}_{\text{spread}}$, $\text{MHT}_{\text{split}}$ and $\text{MHT}_{\text{combi}}$. Comparing the results of these first numerical tests, no general statement can be made which approach gives the best results. The numerical tests show that the difficulty lies in choosing appropriate values for the parameters p for the splitting approach and δ for the spreading approach.

As in the previous subsections, the approaches are exemplarily compared on the basis of the convex test problem (BK1) and the nonconvex test problem (LE1). Tables 6.12 and 6.13 give an overview of the number of obtained efficient points n_1 and required function evaluations n_2 for (BK1) and (LE1) (Test Problems P.1 and P.23 in Appendix A.2). The entries of the cells are given in the order n_1/n_2.

The first row of both tables contains the results of the splitting approach with $\text{MHT}_{\text{split}}$ for different values of p. The first column of both tables contains the results of the spreading approach with $\text{MHT}_{\text{spread}}$ for different values of δ. Only for this approach different starting points are considered which is why in the first column ranges are given.

The remaining cells contain the results of the combination of both splitting and spreading with $\text{MHT}_{\text{combi}}$ for different values of p and δ. The first cell is empty in both tables. It would describe the results of no splitting and no spreading, i.e. the result of the basic trust region algorithm MHT. The results are discussed in detail in Section 6.4. They are not included here since MHT computes only one solution and the aim of this subsection is to compare the approaches of computing several solutions.

Tables 6.12 and 6.13 illustrate together with the findings from Section 6.5.1 that the spreading approach of $\text{MHT}_{\text{spread}}$ is probably the best choice if only 2-3 solutions shall be computed. For the two considered test problems, it mostly outperforms the approach of only splitting the image space given by $\text{MHT}_{\text{split}}$. However, if a higher number of solutions shall be obtained, a general statement about which approach will obtain the best results cannot be derived. For such a statement more extensive tests are required.

As already discussed in the previous subsections, the general and expected correlation of an increasing number of efficient points and function evaluations if p is increased or δ is decreased does not always apply.

δ \ p	none	4	6	8	10	12	14	16	18	20
none		3/25	3/31	4/30	5/42	4/34	4/38	5/37	3/26	4/49
1	3-6/ 15-48	9/ 108	11/ 79	10/ 122	14/ 99	12/ 81	13/ 94	17/ 93	13/ 79	18/ 109
0.5	3-10/ 25-181	14/ 121	14/ 96	13/ 133	17/ 125	18/ 121	17/ 106	21/ 120	17/ 91	22/ 112
0.2	6-24/ 21-205	24/ 178	26/ 150	21/ 155	28/ 179	29/ 165	27/ 148	32/ 168	23/ 125	32/ 176

Table 6.12 – Number of efficient points and function evaluations for (BK1) depending on split number p and spreading distance δ

δ \ p	none	2	4	6	8	10
none		2/35	3/52	2/46	3/43	4/54
0.5	2-3/13-102	3/71	5/177	4/90	6/136	5/137
0.2	2-6/13-102	6/89	7/184	5/92	7/136	7/145
0.1	2-10/13-106	8/97	10/202	8/104	9/140	10/144

Table 6.13 – Number of efficient points and function evaluations for (LE1) depending on split number p and spreading distance δ

The main reasons are the reuse of former interpolation points, the starting points for the splitting approach, for $\text{MHT}_{\text{combi}}$ the location of the points obtained by the splitting and for $\text{MHT}_{\text{spread}}$ the location of the solution obtained by the initial run of MHT. To compare and classify these results of the various modifications of MHT, we consider again the direct search approach of DMS, yet this time with the option of computing several efficient points.

DMS applied to the test problem (BK1) computes 5-54 efficient points with 21-200 function evaluations and applied to (LE1) it computes 1-50 efficient points with 23-202 function evaluations. This is illustrated in Figure 6.34 which shows the number of obtained efficient points depending on the number of required function evaluations. The results of the test instances for DMS are marked as '$\diamond$' and the results of $\text{MHT}_{\text{spread}}$, $\text{MHT}_{\text{split}}$ and $\text{MHT}_{\text{combi}}$ with different values for the parameters δ and p are marked

as '○'. In the figures these algorithms are collectively referred to as MHT (heuristics). The results for the test problem (BK1) are depicted in the left figure and the results for the test problem (LE1) are depicted in the right figure. For the sake of completeness, all instances listed in Tables 6.12 and 6.13 are included in Figure 6.34.

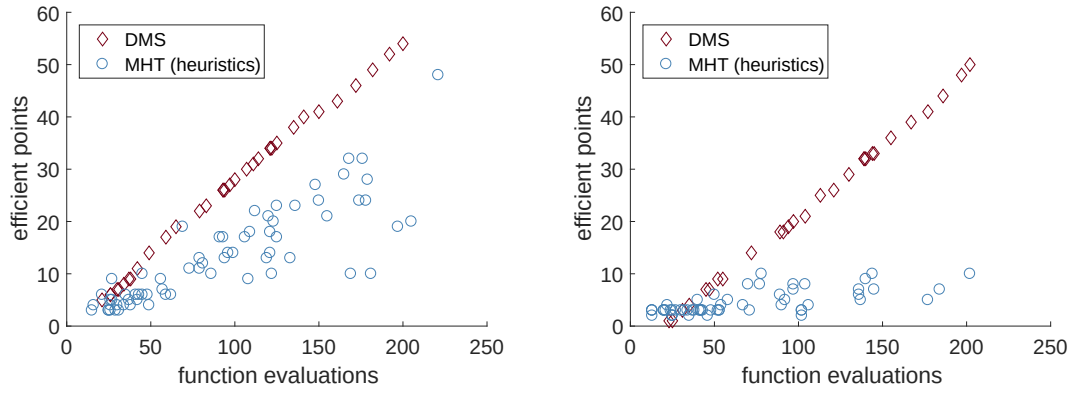

Figure 6.34 – Comparison of DMS and MHT (heuristics) for (BK1) (left) and (LE1) (right)

Since no strict correlation between the choice of the parameters p and δ and the different approaches of mere splitting, mere spreading or the combination of both can be constituted, no clear tendency is visible for the modifications of MHT. In contrast, for one number of function evaluations various values of obtained solutions is possible and vice versa.

These two test problems exemplarily confirm the results from Section 6.4, especially from Section 6.4.6 where performance profiles for MHT are presented. MHT is an approach to compute one solution based on a local search strategy and the presented modifications share its main properties. According to the results in Section 6.4 based on 780 test instances, MHT computes one solution in general faster than DMS. This is also illustrated in Figure 6.34 since the first marked points for DMS have a higher number of function evaluations than the first marked points for the modifications of MHT.

Considering the test problem (BK1), the first efficient points are computed after 15 function evaluations if a modification of MHT is used, whereas 21 function evaluations are required until DMS generates the first efficient points. A similar result is obtained for the test problem (LE1), see also Table 6.13. Using a modification of MHT, the first efficient points are obtained after 13 function evaluations, whereas DMS requires

23 function evaluations until the first efficient point is computed. These first numerical tests indicate that, if only about 20-40 function evaluations are allowed for the expensive function, the heuristic approaches based on MHT compute more efficient points than DMS. Beyond that, DMS outperforms $\text{MHT}_{\text{spread}}$, $\text{MHT}_{\text{split}}$ and $\text{MHT}_{\text{combi}}$. For these two test problems, DMS computes more efficient points if higher numbers of function evaluations are allowed.

However, not only the number of efficient points is an important aspect to evaluate the results, but also their location on the Pareto front and the distance between them. The modifications of MHT are as well as DMS capable of computing points from different regions of the Pareto front and capable of approximating the Pareto front. Though, they differ in the distance between the efficient points.

For DMS the distance decreases arbitrarily. In contrast, the spreading distance δ which is an input parameter for $\text{MHT}_{\text{spread}}$ and $\text{MHT}_{\text{combi}}$ controls the distance between the obtained points. This is also a reason why for higher numbers of function evaluations less efficient points are the output of these algorithms if compared to DMS. Though, a smaller amount of obtained efficient points does not generally mean a worse approximation of the Pareto front.

To illustrate this, we consider again the two test problems (LE1) and (BK1). Firstly, we consider (LE1) and the run of $\text{MHT}_{\text{combi}}$ with $p = 10$ and $\delta = 0.1$ from Section 6.5.3 again, see also Figure 6.32 on page 180. The Pareto front is approximated well by 10 points which require 144 function evaluations in $\text{MHT}_{\text{combi}}$. With the same amount of function evaluations DMS computes 33 efficient points. Yet, the distance between these points is very small both in the domain and in the image space.

This is illustrated in Figure 6.35. It shows the result of $\text{MHT}_{\text{combi}}$ in the top figures and for DMS in the bottom figures. The domain is depicted on the left and the image space on the right. All obtained efficient points are marked black and all other evaluated points are marked as unfilled circles. Both algorithms give a good approximation of the Pareto front, but the points generated by DMS have a smaller distance to each other.

A similar result is obtained if comparing DMS and $\text{MHT}_{\text{combi}}$ for the test problem (BK1). Exemplary results are depicted in Figure 6.36, whereas the top figures show the result of $\text{MHT}_{\text{combi}}$ with $p = 20$ and $\delta = 0.5$ and the bottom figures the result of DMS. $\text{MHT}_{\text{combi}}$ requires 112 function evaluations to compute 22 efficient points. With the same amount of function evaluations, DMS computes 31 efficient points. However, as

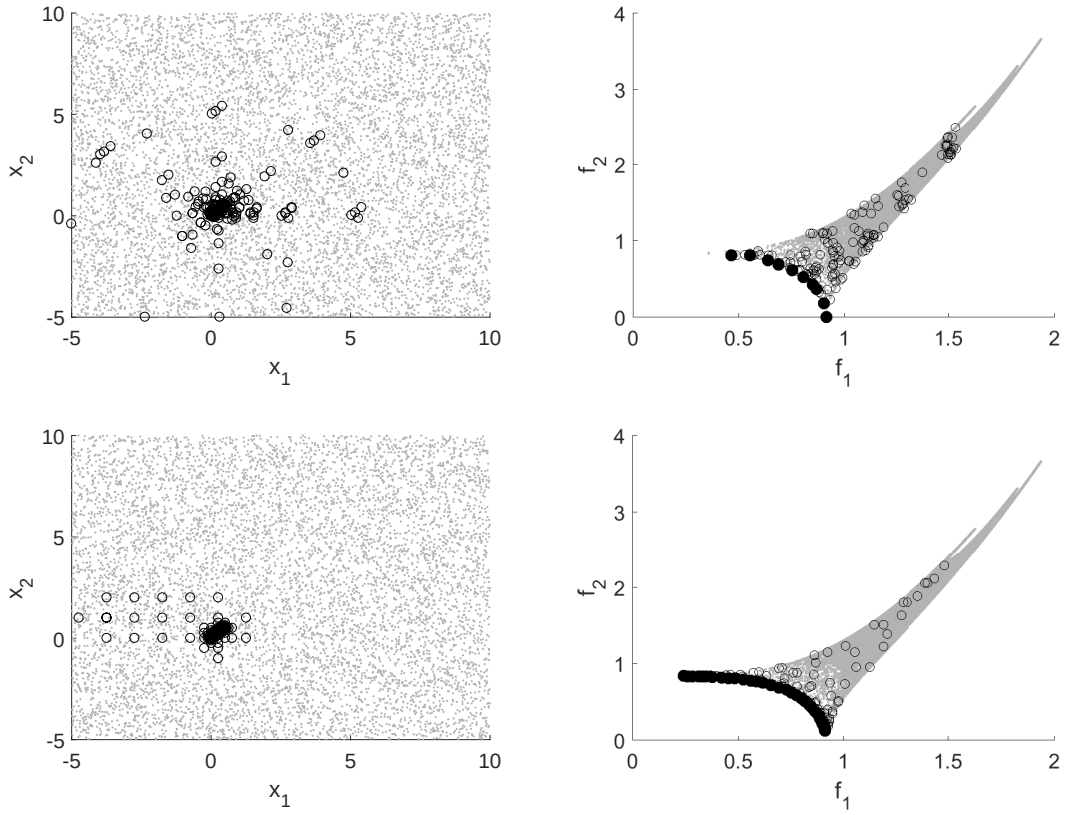

Figure 6.35 – $\mathrm{MHT}_{\mathrm{combi}}$ (domain top left, image space top right) and DMS (domain bottom left, image space bottom right) applied to (LE1)

the figures illustrate, the points obtained by DMS are closer to each other both in the domain and in the image space.

Furthermore, for an appropriate comparison it is important to take the underlying techniques of the algorithms into account. DMS is a direct search approach that, essentially, computes the nondominated points among all evaluated points. Thereby, the evaluated points are chosen by the direct search strategy which compares function values to find directions in which a descent is possible. It does not include any guarantee that the obtained points fulfill a necessary optimality condition.

In contrast, MHT and $\mathrm{MHT}_{\mathrm{spread}}$ guarantee that a necessary condition for weak efficiency is fulfilled in case unconstrained optimization problems are considered. This is also given for a subset of solutions obtained by $\mathrm{MHT}_{\mathrm{combi}}$. One main aspect that ensures this optimality condition are the local model functions. They provide local approximations of the original functions. Such model information is not used in the

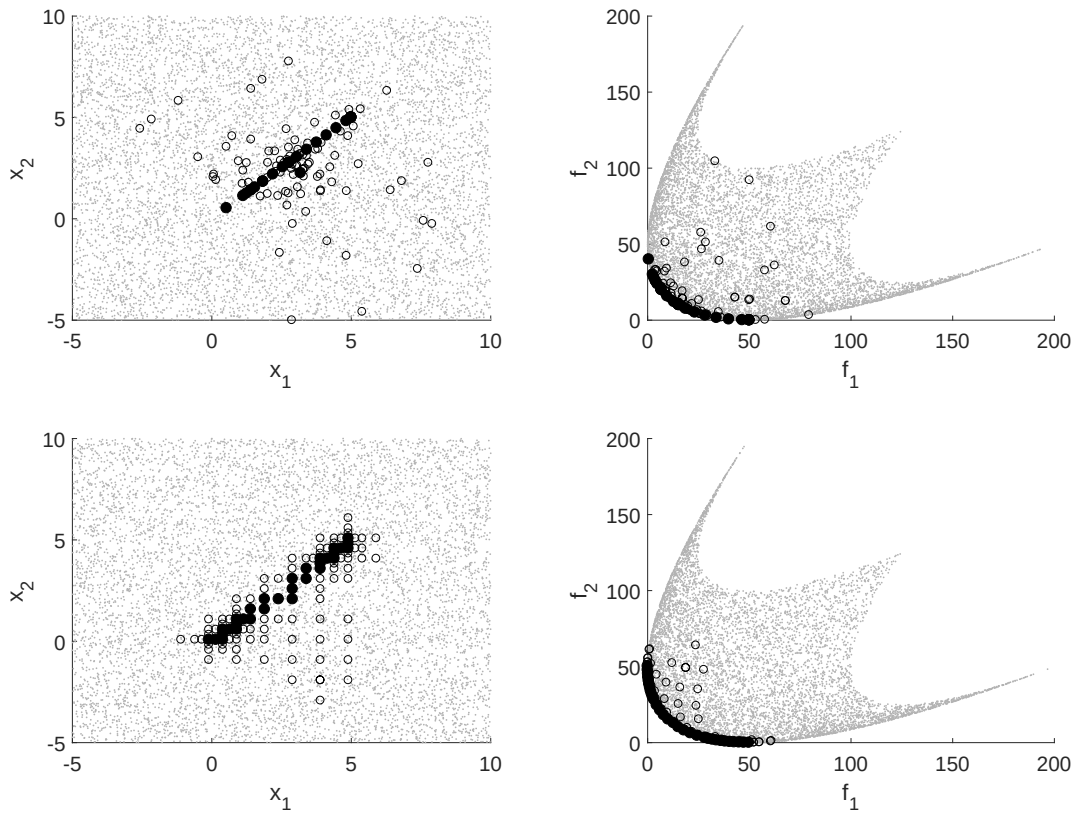

Figure 6.36 – $\mathrm{MHT}_{\mathrm{combi}}$ (domain top left, image space top right) and DMS (domain bottom left, image space bottom right) applied to (BK1)

direct search approach. As these numerical results illustrate, computing and updating reliable model functions can cause more function evaluations, especially if more than one point is obtained.

7 Application: Mixing of Liquid Metals

In this chapter we present an application for the algorithm MHT from the field of fluid dynamics. Here, a liquid metal melt is stirred by a time-dependent magnetic field in order to achieve a homogeneous distribution of particles.
Such flows are motivated by applications in metallurgy where magnetic fields are commonly used to generate a stirring motion inside liquid metal melts. These are usually hot and chemically aggressive and, therefore, cannot be stirred by conventional contact-based methods. The stirring motion is required to enhance mixing processes of additives during the solidification process of the liquid metal since the quality of the final ingot depends on the homogeneous distribution of these additives. The theoretical background of fluid dynamics can for example be found in [Dav01].
The theoretical background for the flow and the mixing problem is described in Section 7.1. The resulting bi-objective optimization problem which has two simulation-given functions is described in Section 7.2 and the results of applying MHT to it are presented in Section 7.3.

7.1 Theoretical Background

Note that in fluid dynamics, it is common to use non-dimensional quantities, i.e. to introduce reference and characteristic values. This is done here, too. The problem of mixing a liquid metal to obtain a homogeneous distribution of additives by using a magnetic field is defined as follows.
A permanent magnet is mounted beneath a rectangular container that is filled with a liquid metal. The thickness L of the liquid metal layer is used as characteristic length. The container has the dimensions of $3L \times 3L \times L$. The rectangular permanent magnet has the dimensions of $L \times L \times 0.5L$. The gap between the surface of the magnet and

the bottom of the liquid metal is $0.25L$. The particles of the additives are randomly seeded within a subsection measuring $1L \times 1L \times 0.5L$. The subsection is located in the center of the liquid metal layer, see Figure 7.1.

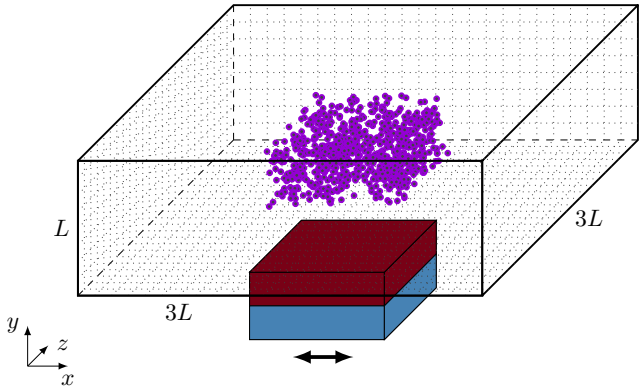

Figure 7.1 – Mixing liquid metals: sketch of problem

As the permanent magnet begins to perform harmonic oscillations, eddy currents are induced within the fluid, i.e. the liquid metal, that give rise to an electromagnetic force, the Lorentz force. It acts as a body force on the flow, see also [Dav01], and drives the stirring motion. Note that the fluid is non-magnetic and the particles representing the additives are also non-magnetic. Furthermore, for the theoretical considerations they are assumed to be mass-less. Hence, the trajectories of the particles are governed solely by the induced fluid flow.

The characteristic velocity is defined as $u := \omega A$, where ω is the angular frequency of the oscillation, and A is the amplitude. The characteristic time is defined as the period length of one oscillation, i.e. $T := 1/f = 2\pi/\omega$, where f is the frequency of the oscillation. The fluid has the kinematic viscosity ν, the electrical conductivity σ, and the densityρ. The scale for the magnetic field is B. The electromagnetic part is anaologous to [Pri+16]. Using the above introduced scales, the physical model of the flow (i.e. the non-dimensional Navier stokes equations) depends on three reference values, see also [TK91]:

$$\beta = \frac{fL^2}{\nu}, \quad \mathrm{KC} = 2\pi A \quad \text{and} \quad \mathrm{Ha} = BL\left(\frac{\sigma}{\rho\nu}\right)^{0.5}. \tag{7.1}$$

Note that these reference values are standard in fluid dynamics, whereas KC is the Keulegan-Carpenter number, Ha the Hartmann number and β is a frequency parameter obtained by the ratio of the Reynolds number and KC. If a constant experimental condition is assumed, i.e. the geometrical constraints and the properties of the fluid are constant, these values define variables that can be used to formulate an optimization problem: β represents the frequency of the oscillation, KC the amplitude and Ha the magnetic field strength.

7.2 Optimization Problem

Based on the explanations in the previous section, the problem of mixing additives into liquid metals can be described as a bi-objective optimization problem with a three-dimensional domain. The first objective is to achieve a good mixing ratio, that is a preferably homogeneous distribution of particles of the additives in the entire liquid metal. A second and competing objective is to reduce the energy expenditure necessary for the oscillation of the magnet.
The domain is defined by the three parameters β, KC and Ha as defined in (7.1). Thus, the variables for the optimization problem are

$$x = (x_1, x_2, x_3)^\top := (\beta, \mathrm{KC}, \mathrm{Ha})^\top.$$

These three variables are used as input for a simulation which describes the behavior of the flow and the mixing process. It defines the objective functions which are described hereafter. The simulation code is not publicly available. It was developed as part of a research project of the department of fluid dynamics, see also [KZB11] .
For the numerical results, $N_p = 1000$ particles are randomly seeded in the above-mentioned subdomain of the container and shall be mixed into the fluid. This random distribution is fixed for all considered simulation runs for reasons of comparison. The mixing process in the simulation is executed for a fixed duration t_{max} to guarantee a fair comparison between each realization given by the values of the variables x.
The simulation gives as output the final positions of the particles and based on this distribution a mixing ratio can be computed. For this purpose, a grid is used. In total, 144 grid cells are used and the array of cells is given as $M = (M_0, M_1, \ldots, M_{k-1})^\top$

with $k = 144$. The value of k is fixed for all computations. The first objective function f_1 quantifying the mixing ratio is defined by

$$f_1(x) = \xi(x) = \frac{1}{k}\sqrt{\sum_{i=0}^{k-1}\left(\frac{N_p}{k} - N_p^i(x)\right)^2} \tag{7.2}$$

with $N_p = 1000$ the number of particles in the entire container and $N_p^i(x)$ the number of particles in the cell M_i, $i = 0, 1, \ldots, 143$, resulting from the mixing process with x as configuration. The smaller the value of ξ is, the better is the mixing ratio and the distribution of the particles in the container. The ideal distribution of the particles is given by $N_p^i = N_p/k$ for all $i = 0, 1, \ldots, k-1$, i.e. if the same amount of particles is contained in every cell M_i, $i = 0, 1, \ldots, 143$. If this distribution is obtained for a vector $\tilde{x}$, it holds $f_1(\tilde{x}) = 0$.
The second objective is to minimize the energy expenditure for moving the magnet, i.e. the conducted work. It is defined by

$$f_2(x) = W(x) = \frac{\mathrm{Ha}^2\mathrm{Re}}{t_{max}} \int_0^{t_{max}} \Delta x_{mag}(t)\, \langle f_{\mathrm{x}}(t)\rangle_V \; dt. \tag{7.3}$$

with $\langle f_{\mathrm{x}}(t)\rangle_V$ the x-component of the Lorentz force acting on the permanent magnet averaged over the volume V of the fluid, $\Delta x_{mag}(t)$ the displacement of the permanent magnet and t_{max} the duration of the mixing in the simulation. For reasons of comparison, the value is multiplied by $\mathrm{Ha}^2\mathrm{Re}/t_{max}$.
The resulting bi-objective optimization problem is given by

$$\min_{x\in\Omega} \left(f_1(x), f_2(x)\right)^\top \tag{MOP_{mix}}$$

with

- the objective functions $f_1, f_2 : \mathbb{R}^3 \to \mathbb{R}$ given by (7.2) and (7.3),
- the variables $x = (x_1, x_2, x_3)^\top = (\beta, \mathrm{KC}, \mathrm{Ha})^\top \in \mathbb{R}^3$ given by (7.1) and
- the constraint set $\Omega = \{x \in \mathbb{R}^n \mid lb \leq x \leq ub\} \subset \mathbb{R}^3$ with lower bounds $lb = (100, 1, 10)^\top \in \mathbb{R}^3$ and upper bounds $ub = (1000, 5, 40)^\top \in \mathbb{R}^3$.

It holds $f_1(x) \geq 0$ and $f_2(x) \geq 0$ for all $x \in \Omega$. Besides, the values of both objective functions are computed based on the output of the simulation. The time required for

one simulation run varies, durations of 5 minutes up to one hour occurred. Thus, both objective functions can be classified as expensive functions. However, both expensive functions are the result of the same simulation, therefore one simulation run provides both function values for a given point $x \in \mathbb{R}^3$.
In general, it is expected that the more energy is applied to move the magnets, i.e. the higher the function value of f_2 is, the better is the distribution of the particles, i.e. the lower is the function value of f_1. This is verified by the numerical results presented in the following section.

7.3 Results of MHT

The bi-objective optimization problem (MOP_{mix}) has two expensive, simulation-given objective functions and is therefore not a heterogeneous optimization problem. Nevertheless, we applied MHT to see how it works although it is not especially designed for such an optimization problem. The following adaptions have been made.
For both objective functions quadratic interpolation models based on Lagrange polynomials as presented in Section 4.2 are used. The interpolation points are chosen such that they fulfill the quality criterion of poisedness which is only based on the distribution of the points in the domain, see also Sections 3.2 and 4.2. One simulation run provides the function values of both objective functions and to reduce the number of simulation runs, the same set of interpolation points can be used for both functions. Consequently, for this optimization problem the numerical effort for computing the model functions of two expensive functions is equal to the numerical effort of computing the model function of one expensive function.
For applying MHT we scaled the variables since the feasibility range differs in magnitude of up to 10^2. For general information about scaling the variables of an optimization problem see for example [Mor83] and for scaling in a trust region framework see for example [CGT00]. In the context of (MOP_{mix}), the model functions in MHT are based only on function values and do not use derivative information. Thus and according to [CGT00], the 'scaled' model functions are computed by using the scaled values of the variables and the original function values of f_1 and f_2. No further modifications

are necessary. Thus, scaling the variables is unproblematic and easy to realize in this context.
The parameters for MHT and the trust region update is realized as described in Section 6.1 with $\varepsilon_{tr} = 0.001$ and $\delta_{PS} = 0.5$. The maximum number of allowed simulation runs was set to $n_{max} = 100$. Since no preference for the values of the starting point was existent, MHT was applied to (MOP_{mix}) with starting points from different areas of the domain. Interesting results arose that provided further insight into the optimization problem of mixing liquid metals. Furthermore, these results confirmed the behavior of MHT observed for the test problems in Chapter 6.
The MHT runs required 36 to 84 simulation runs and the stopping criteria always indicated that a stationary point was obtained. We present two selected results. The values of the variables and function values are rounded to two decimals. To compare the results of MHT, we consider again the direct search approach DMS as in Chapter 6.
Figure 7.2 illustrates the results of one run of MHT. The starting point x^0, the point $\bar{x}^1$ obtained as output of the algorithm and further values are:

- $x^0 = (710, 3, 40)^\top$ with $f_1(x^0) = 8.83$ and $f_2(x^0) = 338.49$,
- $\bar{x}^1 = (433.57, 1, 32.86)^\top$ with $f_1(\bar{x}^1) = 15.89$ and $f_2(\bar{x}^1) = 15.24$,
- 59 simulation runs in 10 iterations.

The domain is depicted on the left and the image space on the right of Figure 7.2. The starting point is marked blue, $\bar{x}^1$ and $f(\bar{x}^1)$ are marked orange, the iteration points are marked black and connected by dotted lines and all further evaluated points are marked gray.
Note that it holds $\bar{x}^1_2 = lb_2$, i.e. one of the box constraints is active. MHT is designed for unconstrained optimization problems and as outlined in Section 4.7 the convergence results do not hold if box constrained optimization problems are considered. Therefore, it is not guaranteed that the obtained point $\bar{x}^1$ fulfills a necessary condition for local optimality. However, MHT terminated because the step size of the Pascoletti-Serafini problem ($\bar{t}^1 = -6.32 \cdot 10^{-6}$) and the trust region radius ($\bar{\delta}^1 = 0.5$) were small enough, i.e. it is supposed that no sufficient decrease exists in the last trust region, see also Section 6.1.1 where the stopping criteria are explained.
The right figure shows that for some iteration points the function values of f_1 increase. The reason is the trial point acceptance test. It guarantees a descent for at least one of

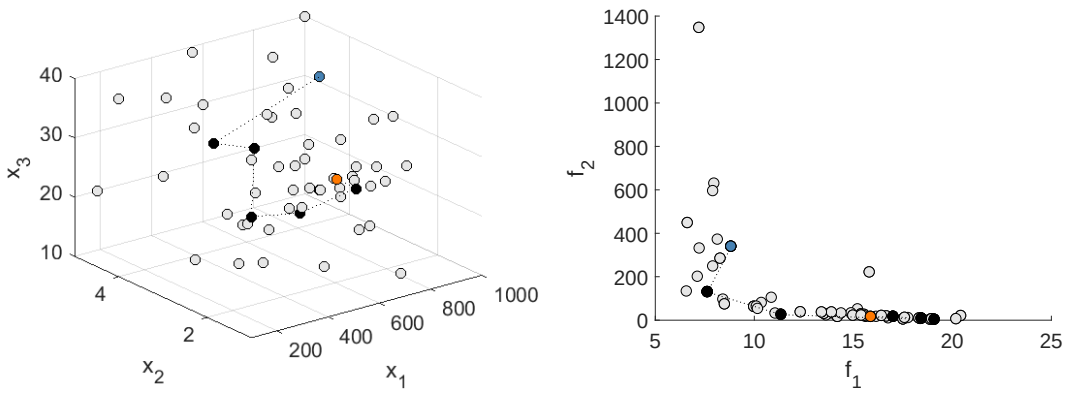

Figure 7.2 – MHT applied to (MOP_{mix}) with $x^0 = (710, 3, 40)^\top$

the objective functions in every iteration. Thus, it is possible that iteration points are accepted that increase one of the objective functions, see also Section 4.4 for a detailed description.

This occurred for several runs of MHT with (MOP_{mix}) and therefore we also considered MHT with the strict version of the trial point acceptance test as described in Section 4.6.3. Using the starting point x^0, but applying MHT with the strict acceptance test, the algorithm terminates with

- $\bar{x}^2 = (199.27, 3.88, 35.24)^\top$ with $f_1(\bar{x}^2) = 7.04$, $f_2(\bar{x}^2) = 106.84$ and
- 70 simulation runs in 15 iterations.

It terminated since the step size of the Pascoletti-Serafini problem ($\bar{t}^2 = -2.05 \cdot 10^{-14}$) and the trust region radius ($\bar{\delta}^2 = 0.5$) were small enough. Thus, again it can be supposed that no sufficient decrease exists in the last trust region. However, MHT with the strict acceptance test requires more iterations and more simulation runs than MHT with the standard acceptance test. The results are illustrated in Figure 7.3.

The first iteration point is identical for both runs, but the subsequent iteration points differ. Whereas MHT with the standard acceptance test computes various further iteration points with increasing values for f_1 (see Figure 7.2), MHT with the strict acceptance test computes only one further iteration point which is given as output.

Nevertheless, more iterations are executed and most of them are unsuccessful, i.e. the trial point is not accepted. In these iterations the model functions are updated and new interpolation points are required. This is illustrated in Figure 7.3 since many of

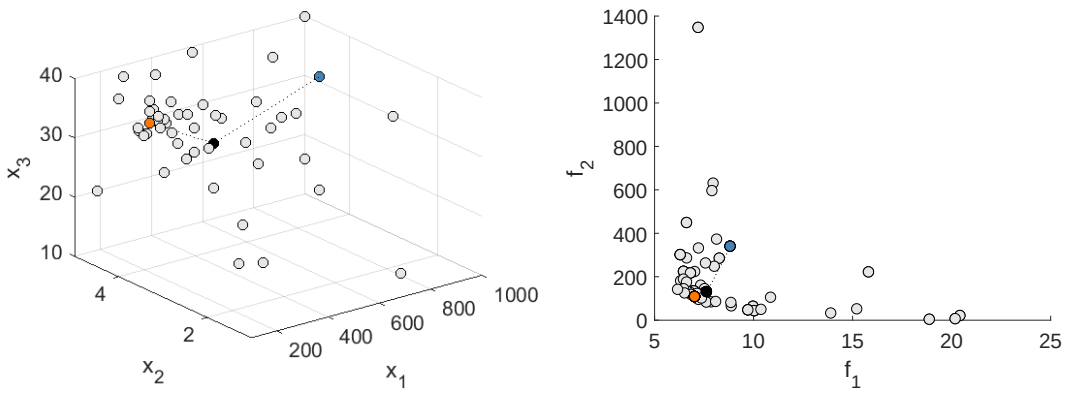

Figure 7.3 – MHT (strict acceptance test) applied to (MOP_{mix}) with $x^0 = (710, 3, 40)^\top$

the gray marked point accumulate around $\bar{x}^2$ respectively $f(\bar{x}^2)$ (both marked orange). Figure 7.4 illustrates the distribution of the particles with the configuration of the mixing process given by $\bar{x}^1$ (left figure) and $\bar{x}^2$ (right figure). The square marks the area in which the particles are initially situated and the purple marked points show the distribution of the particles after the mixing process with the two different configurations obtained by MHT.

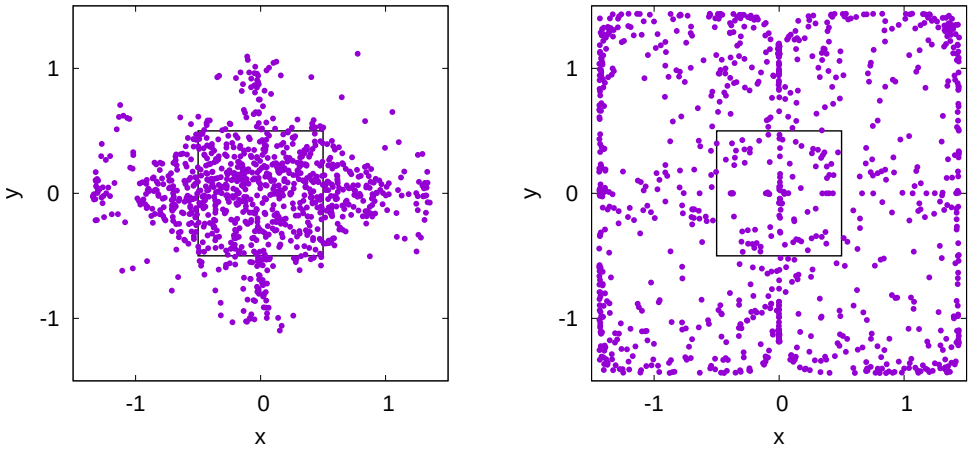

Figure 7.4 – Distribution of the particles after mixing with the configuration given by $\bar{x}^1$ (left, MHT standard) and $\bar{x}^2$ (right, MHT strict)

The illustrations in Figure 7.4 are only a top view of the container with the liquid metal. Nevertheless, they illustrate that the configuration obtained by MHT with the strict acceptance test (right figure) significantly enhances the mixing ratio in comparison to

the configuration obtained by MHT with the standard acceptance test (left figure).
To compare the results of MHT, DMS was applied to (MOP_{mix}) with $x^0 = (710, 3, 40)^\top$. The resulting values are

- $\bar{x}^3 = (709.43, 2.99, 40)^\top$ with $f_1(\bar{x}^3) = 7.31$, $f_2(\bar{x}^3) = 198.81$ and
- 70 simulation runs

Thus, DMS requires the same amount of simulation runs as MHT with the strict acceptance test, but generates a point which differs significantly in the values of the variables and the function value of f_2. Furthermore, it is close to the starting point. The results are illustrated in Figure 7.5. The point $\bar{x}^3$ is not visible in the left figure since many points situated closely to it are generated during the procedure.

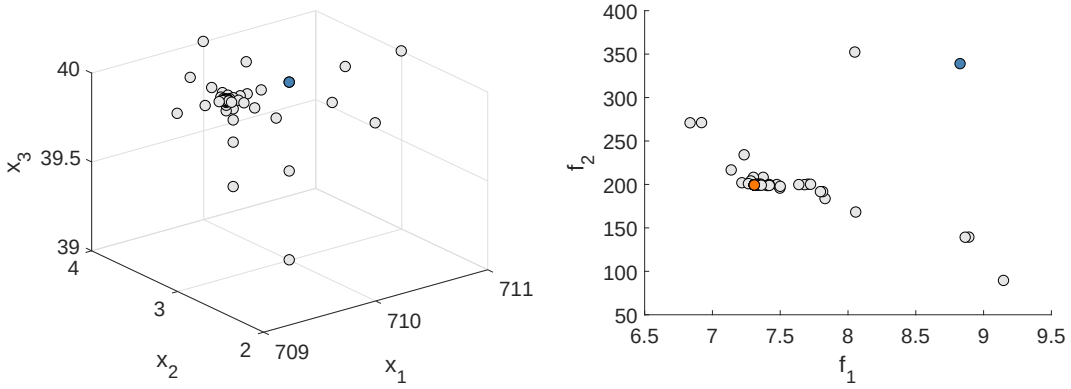

Figure 7.5 – DMS applied to (MOP_{mix}) with $x^0 = (710, 3, 40)^\top$

The distribution of the particles after the mixing process is illustrated in Figure 7.6. It is similar to the result of the mixing with the configuration given by $\bar{x}^2$ (right part of Figure 7.4). Comparing the function values $f(\bar{x}^2) = (7.04, 106.84)^\top$ and $f(\bar{x}^3) = (7.31, 198.81)^\top$, it follows that $\bar{x}^2$ dominates $\bar{x}^3$.
The objective function f_2 quantifies the energy amount necessary for the mixing process. Therefore, the configuration given by $\bar{x}^2$ requires significantly less energy to obtain a similar distribution of the particles. Thus, when comparing only these two results, $\bar{x}^2$ and therefore MHT with the strict acceptance test generates the preferable configuration for the mixing.
Another result which also shows the influence of the trial point acceptance test is illustrated in Figure 7.7. This run was canceled externally before MHT terminated due

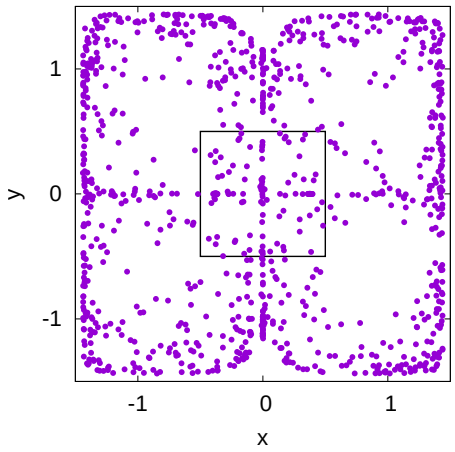

Figure 7.6 – Distribution of the particles after mixing with the configuration given by $\bar{x}^3$ (DMS)

to problems with the software. MHT was restarted with the last obtained iteration point from the previous run. We present the merged results. The starting point $\tilde{x}^0$, the point $\bar{x}^4$ obtained as output of the algorithm and further values are:

- $\tilde{x}^0 = (200, 1.5, 11)^\top$ with $f_1(\tilde{x}^0) = 20.32$ and $f_2(\tilde{x}^0) = 1.96$
- $\bar{x}^4 = (100, 1.98, 38.08)^\top$ with $f_1(\bar{x}^4) = 14.95$ and $f_2(\bar{x}^4) = 13.86$
- 81 simulation runs (37 in the first part and 44 in the second part) in 26 iterations

The number of function evaluations is not fully representable since it is possible that already evaluated points have been lost due to the external termination of the algorithm.

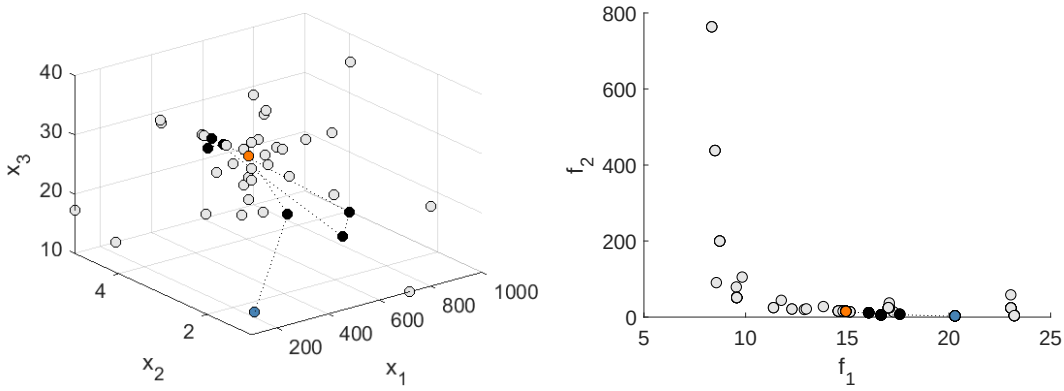

Figure 7.7 – MHT applied to (MOP_{mix}) with $\tilde{x}^0 = (200, 1.5, 11)^\top$

In the end of the second run, MHT terminated since the step size of the Pascoletti-Serafini problem was small enough ($\bar{t}^4 = -3.43 \cdot 10^{-34}$) and the trust region was small enough ($\bar{\delta}^4 = 0.5$). Thus, it is assumed that the model functions are reliable and no sufficient decrease can be generated in the last trust region.
A comparison run with MHT with the strict version of the trial point acceptance test terminated with the starting point $\tilde{x}^0$ with the same stopping criterion. It required 84 simulation runs and therefore again more than the run of MHT with the standard acceptance test.
Figure 7.8 illustrates the distribution of the particles after the mixing process with the configuration given by $\tilde{x}^0$ (left) and $\bar{x}^4$ (right). They show that the configuration obtained by MHT enhances the mixing ratio compared to the starting configuration. Moreover, it confirms the expected connection between energy input and mixing ratio, that a lower energy expenditure generates a less homogeneous distribution of the particles.

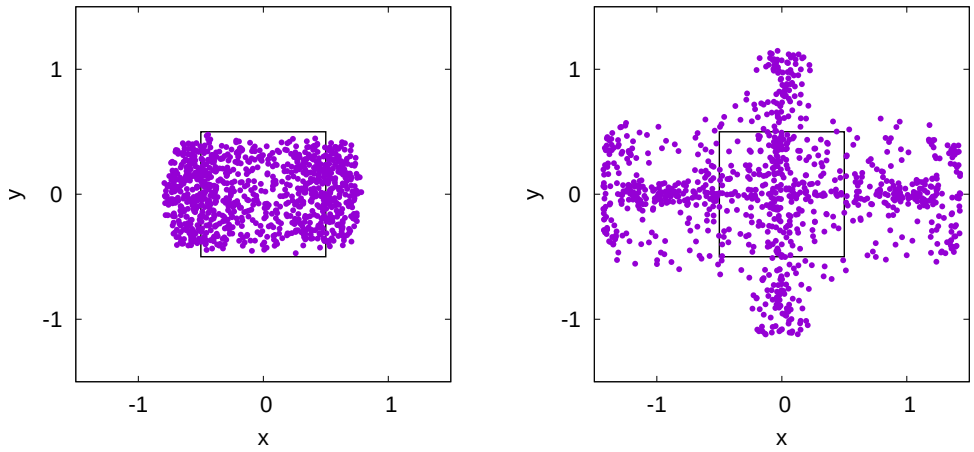

Figure 7.8 – Distribution of the particles after mixing with the configuration given by $\tilde{x}^0$ (left) and $\bar{x}^4$ (right)

DMS applied to (MOP_{mix}) with the starting point $\tilde{x}^0 = (200, 1.5, 11)^\top$ terminates with the point $\bar{x}^5 = (199.98, 1.5, 11)^\top$ very close to it. However, it requires only 68 simulation runs and therefore less than both runs of MHT.
The results of applying MHT to (MOP_{mix}) confirmed the theoretical results and the numerical results of the test problems. The algorithm was capable of computing stationary points also for an application-motivated optimization problem.
Furthermore, the results show that it is reasonable to use the strict version of the trial

point acceptance test to avoid computing further iteration points although the current iteration point is already a stationary point. Nevertheless, the strict trial point acceptance test does not guarantee that the algorithm requires less function evaluations. In contrast, it also occurred that unsuccessful iterations accumulated during the end of the algorithm. Thus, the iteration point did not change any more, but many additional function evaluations were required to ensure the local accuracy of the model functions.

It needs to be noted that most of the points generated by MHT were not reasonable from the perspective of the application. Mostly, the mixing ratio was considered to be insufficient or the amount of the energy expense too low, i.e. the value of f_1 too high or the value of f_2 too low. Thus, there was a 'hidden' constraint which was not formulated for the optimization problem since it was not apparent before applying MHT. One possibility to realize this constraint could be to integrate a lower bound for f_2. With such a constraint, MHT is not applicable and a different type of algorithm is required.

The optimization problem (MOP_{mix}) has changed several times by modifying fixed parameters and settings during applying MHT. We presented only results for the final formulation of (MOP_{mix}) as described in Sections 7.1 and 7.2. However, for all considered variants of (MOP_{mix}) the behavior was similar and mostly the starting points were classified as stationary points by MHT or close to such points. This was an unexpected finding and therefore the results of MHT provided further insight into the mixing problem as formulated in Section 7.2.

8 Summary and Conclusion

In this thesis a trust region approach for unconstrained multi-objective optimization problems is presented. In comparison to similar approaches from literature, the new aspect is that heterogeneity can be considered and that the search direction is defined in the image space. The basic version of the algorithm MHT is presented in Chapter 4 and it is proved that the accumulation point of the produced sequence of iterates is a Pareto critical point, i.e. it fulfills a necessary condition for local optimality. The convergence proof cannot be transferred directly from other trust region approaches. Due to the different model functions, the aspect of heterogeneity and the definition of the search direction, several aspects from convergence considerations from single- and multi-objective trust region approaches and multi-objective descent methods are used.

In the basic version of the algorithm presented in Chapter 4 the heterogeneity of the objective functions is only considered in the different choice of the model functions. However, several small modifications can be made, for example using the cheap functions themselves instead of model functions to exploit the heterogeneity further. Moreover, heuristic strategies based on MHT are presented in Chapter 5 to exploit the heterogeneity further and to compute several Pareto critical points.

The numerical tests of MHT and its modifications confirm the theoretical findings and the usefulness of the approaches. The optimization problem from fluid dynamics presented in Chapter 7 also shows that MHT can compute stationary points for application-motivated optimization problems.

For the numerical tests also box constrained optimization problems have been considered. Box constraints can be integrated in MHT, but not all of the theoretical results from Section 4.5 hold in this case. Especially the main statement about producing a point fulfilling a necessary condition for local optimality cannot be proved. The theoretical results that hold are given in Section 4.7. Possibly, the algorithm can be adapted

in a way such that a similar convergence result can be proved by considering the approach from [GPD18]. There, a descent method including constraints is presented and the theoretical background was partly used in Section 4.7 to prove some theoretical results for MHT including box constraints. Besides that, general methods for trust region approaches to integrate constraints can be considered, see for example [CGT88; FMS94; CL96; CGT00].

For the heuristic methods only first tests have been executed. An extensive parameter study would be necessary to make general statements about which spreading distance or splitting number or the combination of both can be assumed to provide the best result. Of course further modifications or different ways of implementation could also improve the results. The current implementation of the heuristics is only a first realization.

Moreover, other starting points for the splitting approach might improve the results, another reasonable option could for example be the nondominated points among all evaluated points. These points could also be integrated into the splitting approach and into MHT. Instead of only one iteration point, a list of points could be considered given by these points.

Furthermore, the presented heuristics from Chapter 5 are only formulated for bi-objective optimization problems since they are motivated by considerations for such problems. As outlined in the chapter about the heuristics, all of these approaches could be applied to optimization problems with $q \geq 3$ objective functions by choosing one of the cheap functions for the heuristic approach. More sophisticated approaches are subject to future research.

A Appendix

A.1 Assumptions

Assumptions A.1 to A.11 are for the basic version of MHT presented in Chapter 4 and Assumptions A.12 to A.18 for the modifications of MHT presented in Section 4.6 and Chapter 5.

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

$$m_i^k(x^k) = f_i(x^k) \text{ for all } i \in \{1, 2, \ldots, q\} \text{ and } k \in \mathbb{N}.$$

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 $\kappa_{\text{uhf}_i} > 1$ fulfilling

$$\left\|\nabla^2 f_i(x)\right\|_F \leq \kappa_{\text{uhf}_i} - 1$$

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 $\kappa_{\text{uhm}_i} > 1$ independent of k fulfilling

$$\left\|\nabla^2 m_i^k(x)\right\|_F \leq \kappa_{\text{uhm}_i} - 1$$

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 $\kappa_{\text{em}_1} > 0$ independent of k such that it holds for all $x \in B_k$

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

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 ($MOP^k_{m,\mathbb{R}^n}$)

$$\frac{\min\limits_{i=1,\ldots,q} r_i^k}{\max\limits_{j=1,\ldots,q} r_j^k} \geq \kappa_r. \tag{4.19}$$

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 $\beta_\phi^k = \max_{i=1,\ldots,q} \left\|\nabla^2 m_i^k(x^k)\right\|_F + 1$.

The following two assumptions are necessary for replacing the ideal point by a lower bound as described in Section 4.6.2. In this context, Assumption A.10 needs to be replace by Assumption A.13

Assumption A.12 There exist constants $\tilde{p}_i \in \mathbb{R}$, $i = 1, 2, \ldots, q$, such that it holds $\tilde{p}^k \geq \tilde{p}_i$ for all $i \in \{1, 2, \ldots, q\}$ and for all $k \in \mathbb{N}$.

Assumption A.13 There exists a constant $\kappa_{\tilde{r}} \in (0, 1]$ such that it holds for every iteration $k \in \mathbb{N}$ with x^k not Pareto critical for $(MOP^k_{m,\mathbb{R}^n})$

$$\frac{\min\limits_{i=1,\ldots,q} \tilde{r}_i^k}{\max\limits_{j=1,\ldots,q} \tilde{r}_j^k} \geq \kappa_{\tilde{r}}.$$

If MHT is considered with the strict trial point acceptance test described in Section 4.6.3, Assumption A.11 needs to be replaced by Assumption A.14

Assumption A.14 There exists a constant $\kappa_\phi \in (0, 1)$ such that it holds for every iteration $k \in \mathbb{N}$ and for all $i \in \{1, 2, \ldots, q\}$

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

with $\beta_\phi^k = \max_{i=1,\ldots,q} \left\|\nabla^2 m_i^k(x^k)\right\|_F + 1$.

The following assumptions are necessary for MHT if the cheap objective functions $f_2, f_3, \ldots, f_q$ are not replaced by model functions as described in Section 4.6.4. In this context, Assumption A.3 needs to be replaced by Assumption A.16

Assumption A.15 Let the model functions of the cheap functions be exact, that is it holds $m_i^k \equiv f_i$ for all $i \in \{2, 3, \ldots, q\}$ and for all $k \in \mathbb{N}$.

Assumption A.16 The model function $m_1^k : \mathbb{R}^n \to \mathbb{R}$ is quadratic for all $k \in \mathbb{N}$.

The following assumption is necessary for the heuristic approaches based on MHT from Chapter 5.

Assumption A.17 Let the cheap objective function $f_2 : \mathbb{R}^n \to \mathbb{R}$ of (BOP) be bounded from below and let $\min_{x\in\mathbb{R}^n} f_2(x) = \inf_{x\in\mathbb{R}^n} f_2(x)$.

The following assumption is necessary for the image space split in Section 5.2.

Assumption A.18 Let Assumption A.17 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 $C \leq \sup_{x \in \mathbb{R}^n} f_2(x)$.

A.2 Test Functions

In the following we give a detailed description of the test problems listed in Table 6.1 in Section 6.2 with the size of the domain (n), the number of objective functions (q), the objective functions, the information which function is declared as expensive, the box constraints (Ω, if existent) the used starting points (given as a $n \times 10$ matrix) and references. For the computations the starting points were used with four decimals. For reasons of space, we list them here for some test examples with only two decimals.

Test Problem P.1 (BK1) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [-5, 10]^2$,

$$f_1(x) = x_1^2 + x_2^2, \qquad f_2(x) = (x_1 - 5)^2 + (x_2 - 5)^2$$

from [Hub+06; Cus+11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} 5.11 & 9.42 & 7.65 & 3.90 & 4.01 & 2.38 & 5.00 & 0.77 & -1.15 & 2.75 \\ 2.20 & 9.46 & -3.69 & -1.92 & -2.21 & 7.39 & -0.43 & 3.35 & 8.36 & 8.15 \end{pmatrix}$$

Test Problem P.2 (CL1) (MOP_c) with $n = 4$, $q = 2$, constraint set $\Omega = \left\{x \in \mathbb{R}^4 \mid x_1, x_4 \in [1, 3], x_2, x_3 \in \left[\sqrt{2}, 3\right]\right\}$,

$$\begin{aligned} f_1(x) &= 400x_1 + 200\sqrt{2}x_2 + 200\sqrt{x_3} + 200x_4, \\ f_2(x) &= \frac{2}{100x_1} + \frac{2\sqrt{2}}{100x_2} - \frac{2\sqrt{2}}{100x_3} + \frac{2}{100x_4} \end{aligned}$$

from [CL99; Cus+11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} 1.8220 & 1.1312 & 2.6947 & 1.4721 & 1.1132 & 2.6820 & 1.1932 & 1.3662 & 1.1280 & 2.1255 \\ 2.9884 & 1.7300 & 2.7943 & 2.5214 & 1.7719 & 2.3658 & 1.4447 & 2.5453 & 2.2443 & 2.5737 \\ 2.2570 & 2.7878 & 1.9005 & 1.4231 & 1.7135 & 2.8591 & 2.8339 & 2.7468 & 1.8231 & 1.7452 \\ 2.7545 & 1.6104 & 2.3046 & 1.4470 & 1.0701 & 2.2887 & 2.0789 & 2.9922 & 1.8385 & 2.9020 \end{pmatrix}$$

Test Problem P.3 (Deb41) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = \{x \in \mathbb{R}^2 \mid x_1 \in [0.1, 1], x_2 \in [0, 1]\}$,

$$f_1(x) = x_1, \qquad f_2(x) = \frac{g(x)}{x_1},$$
$$g(x) = 2 - \exp\left(-\left(\frac{x_2 - 0.2}{0.004}\right)^2\right) - 0.8\exp\left(-\left(\frac{x_2 - 0.6}{0.4}\right)^2\right)$$

from [Deb99; Cus+11] with f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.5887 & 0.5219 & 0.3560 & 0.9164 & 0.4109 & 0.5864 & 0.1670 & 0.4993 & 0.7219 & 0.1593 \\ 0.5807 & 0.6207 & 0.5997 & 0.3355 & 0.8156 & 0.3910 & 0.8813 & 0.0042 & 0.8176 & 0.3456 \end{pmatrix}$$

Test Problem P.4 (Deb53) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [0, 1]^2$,

$$f_1(x) = 1 - \exp(-4x_1)\sin(5\pi x_1)^4, \qquad f_2(x) = g(x)h(x),$$
$$g(x) = \begin{cases} 4 - 3\exp\left(-\left(\frac{x_2 - 0.2}{0.02}\right)^2\right) & , \text{ if } x_2 \leq 0.4 \\ 4 - 2\exp\left(-\left(\frac{x_2 - 0.7}{0.2}\right)^2\right) & , \text{ else} \end{cases},$$
$$h(x) = \begin{cases} 1 - \left(\frac{f_1(x)}{\beta g(x)}\right)^\alpha & , \text{ if } f_1(x) \leq \beta g(x) \\ 0 & , \text{ else} \end{cases}$$

from [Deb99; Cus+11] with $\beta = 1, \alpha = 4$, f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.3073 & 0.2955 & 0.2491 & 0.8225 & 0.2515 & 0.4047 & 0.6532 & 0.9010 & 0.9796 & 0.5402 \\ 0.0406 & 0.9243 & 0.3215 & 0.0500 & 0.2402 & 0.0734 & 0.6027 & 0.4889 & 0.8841 & 0.2093 \end{pmatrix}$$

Test Problem P.5 (Deb513) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [0, 1]^2$

$$f_1(x) = x_1, \qquad f_2(x) = g(x)h(x),$$
$$g(x) = 1 + 10x_2, \quad h(x) = 1 - \left(\frac{x_1}{g(x)}\right)^\alpha - \left(\frac{x_1}{g(x)}\right)\sin(2\pi p x_1)$$

from [Deb99; Cus+11] with $\alpha = 2, p = 4$, f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.6506 & 0.4335 & 0.7470 & 0.0616 & 0.2168 & 0.7550 & 0.2273 & 0.8490 & 0.1790 & 0.2906 \\ 0.9923 & 0.4001 & 0.9388 & 0.9883 & 0.2901 & 0.9005 & 0.8332 & 0.6657 & 0.6555 & 0.2458 \end{pmatrix}$$

Test Problem P.6 (Deb521b) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [0, 1]^2$

$$\begin{aligned} f_1(x) &= x_1, \qquad f_2(x) = g(x)h(x), \\ g(x) &= 1 + x_2^\gamma, \qquad h(x) = 1 - \left(\frac{x_1}{g(x)}\right)^2 \end{aligned}$$

from [Deb99; Cus+11] with $\gamma = 1$, f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.6506 & 0.4335 & 0.7470 & 0.0616 & 0.2168 & 0.7550 & 0.2273 & 0.8489 & 0.1790 & 0.2906 \\ 0.9923 & 0.4001 & 0.9388 & 0.9883 & 0.2901 & 0.9005 & 0.8332 & 0.6657 & 0.6555 & 0.2458 \end{pmatrix}$$

Test Problem P.7 (DG01) (MOP_c) with $n = 1$, $q = 2$, constraint set $\Omega = [-10, 13]$,

$$f_1(x) = \sin(x), \qquad f_2(x) = \sin(x + 0.7)$$

from [Hub+06; Cus+11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} -7.43 & 11.92 & 8.38 & -4.04 & 1.51 & -8.80 & -0.48 & -3.61 & -3.71 & 11.26 \end{pmatrix}$$

Test Problem P.8 (DLTZ1) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [0, 1]^2$,

$$\begin{aligned} f_1(x) &= 0.5x_1(1 + g(x)), \qquad f_2(x) = 0.5(1 - x_1)(1 + g(x)), \\ g(x) &= 100\left(1 + \left((x_2 - 0.5)^2 - \cos(20\pi(x_2 - 0.5))\right)\right) \end{aligned}$$

from [Deb+02; Cus+11] with $k = 1$, f_1 declared as expensive and the starting points

$$\begin{pmatrix} 0.2608 & 0.7504 & 0.6298 & 0.3355 & 0.1815 & 0.4296 & 0.9756 & 0.6740 & 0.7429 & 0.1462 \\ 0.3662 & 0.0158 & 0.2438 & 0.8471 & 0.9039 & 0.3286 & 0.0888 & 0.6109 & 0.2120 & 0.4630 \end{pmatrix}$$

Test Problem P.9 (ex005) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = \{x \in \mathbb{R}^2 \mid x_1 \in [-1, 2], x_2 \in [1, 2]\}$,

$$f_1(x) = x_1^2 - x_2^2, \qquad f_2(x) = \frac{x_1}{x_2}$$

from [HM79; Cus+11] with f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.09 & 0.01 & -0.86 & -0.13 & -0.65 & -0.91 & -0.63 & -0.20 & -0.69 & 0.35 \\ 1.18 & 1.87 & 1.77 & 1.47 & 1.42 & 1.91 & 1.30 & 1.47 & 1.84 & 1.18 \end{pmatrix}$$

Test Problem P.10 (Far1) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [-1, 1]^2$,

$$\begin{aligned} f_1(x) = &- 2\exp(15(-(x_1 - 0.1)^2 - x_2^2)) - \exp(20(-(x_1 - 0.6)^2 - (x_2 - 0.6)^2)) \\ &+ \exp(20(-(x_1 + 0.6)^2 - (x_2 - 0.6)^2)) + \exp(20(-(x_1 - 0.6)^2 - (x_2 + 0.6)^2)) \\ &+ \exp(20(-(x_1 + 0.6)^2 - (x_2 + 0.6)^2)) \\ f_2(x) = &2\exp(20(-x_1^2 - x_2^2)) + \exp(20(-(x_1 - 0.4)^2 - (x_2 - 0.6)^2)) \\ &- \exp(20(-(x_1 + 0.5)^2 - (x_2 - 0.7)^2)) - \exp(20(-(x_1 - 0.5)^2 - (x_2 + 0.7)^2)) \\ &+ \exp(20(-(x_1 + 0.4)^2 - (x_2 + 0.8)^2)) \end{aligned}$$

from [Hub+06; Cus+11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} -0.24 & -0.86 & 0.75 & -0.81 & 0.04 & 0.04 & 0.07 & -0.00 & 0.10 & -0.61 \\ -0.18 & -0.59 & -0.64 & 0.33 & 0.21 & -0.93 & 0.97 & -0.51 & -0.47 & 0.58 \end{pmatrix}$$

Test Problem P.11 (FES2) (MOP_c) with $n = 10, q = 3$, constraint set $\Omega = [0, 1]^{10}$,

$$\begin{aligned} f_1(x) &= \sum_{i=1}^{n} (x_i - 0.5\cos(10\pi i/n) - 0.5)^2 \\ f_2(x) &= \sum_{i=1}^{n} \left| x_i - \sin^2(i-1)\cos^2(i-1) \right|^{0.5} \\ f_3(x) &= \sum_{i=1}^{n} |x_i - 0.25\cos(i-1)\cos(2i-2) - 0.5|^{0.5} \end{aligned}$$

from [Hub+06; Cus+11] with f_3 declared as expensive and the starting points

$$\begin{pmatrix} 0.2718 & 0.7504 & 0.6298 & 0.3355 & 0.1815 & 0.4296 & 0.9756 & 0.6740 & 0.7429 & 0.1462 \\ 0.3662 & 0.0258 & 0.2438 & 0.8471 & 0.9139 & 0.3286 & 0.1888 & 0.6109 & 0.2120 & 0.4630 \\ 0.8113 & 0.0334 & 0.9902 & 0.4293 & 0.4609 & 0.6665 & 0.3115 & 0.3971 & 0.7920 & 0.7029 \\ 0.7461 & 0.5742 & 0.6209 & 0.1995 & 0.9002 & 0.0980 & 0.2954 & 0.8980 & 0.7774 & 0.1758 \\ 0.9244 & 0.8414 & 0.1932 & 0.4648 & 0.2929 & 0.6805 & 0.2328 & 0.6825 & 0.7279 & 0.0917 \\ 0.5746 & 0.2220 & 0.0886 & 0.1913 & 0.9952 & 0.0034 & 0.3167 & 0.6999 & 0.2553 & 0.3135 \\ 0.2940 & 0.5776 & 0.4261 & 0.5287 & 0.9195 & 0.0380 & 0.4288 & 0.1106 & 0.2265 & 0.1646 \\ 0.4627 & 0.3460 & 0.7655 & 0.7986 & 0.6363 & 0.2556 & 0.0215 & 0.8495 & 0.6748 & 0.0205 \\ 0.3347 & 0.7659 & 0.5160 & 0.0194 & 0.3200 & 0.4587 & 0.5545 & 0.5180 & 0.6672 & 0.5616 \\ 0.7129 & 0.9708 & 0.9160 & 0.3834 & 0.2898 & 0.2754 & 0.7786 & 0.3878 & 0.8520 & 0.8518 \end{pmatrix}$$

Test Problem P.12 (FF) (MOP_c) with $n \in \{2, 3, 4, 5\}$, $q = 2$, constraint set $\Omega = [-4, 4]^n$,

$$f_1(x) = 1 - \exp\left(-\sum_{i=1}^{n}\left(x_i - \frac{1}{\sqrt{n}}\right)^2\right)$$
$$f_2(x) = 1 - \exp\left(-\sum_{i=1}^{n}\left(x_i + \frac{1}{\sqrt{n}}\right)^2\right)$$

from [FF95] with f_1 declared as expensive and the starting points for $n = 2$

$$\begin{pmatrix} 1.63 & 0.42 & -2.67 & 3.69 & 0.70 & 1.06 & -0.91 & 2.04 & -0.92 & -1.97 \\ -1.87 & 3.29 & 1.84 & 1.40 & -3.39 & -4.00 & -1.76 & -1.33 & -2.21 & 3.51 \end{pmatrix}$$

Test Problem P.13 (Fonseca) (MOP_c) with $n = 2, q = 2$, constraint set $\Omega = [-4, 4]^2$,

$$f_1(x) = 1 - \exp(-(x_1 - 1)^2 - (x_2 + 1)^2)$$
$$f_2(x) = 1 - \exp(-(x_1 + 1)^2 - (x_2 - 1)^2)$$

from [FF98] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} 3.72 & -3.09 & -2.09 & -2.22 & 1.21 & 3.54 & 3.03 & -2.99 & -3.44 & -3.68 \\ -1.19 & 1.99 & 3.64 & -0.79 & 1.60 & 0.12 & 0.08 & -3.26 & 0.81 & -3.91 \end{pmatrix}$$

Test Problem P.14 (IKK1) (MOP_c) with $n = 2, q = 3$, constraint set $\Omega = [-50, 50]^2$,

$$f_1(x) = x_1^2, \qquad f_2(x) = (x_1 - 20)^2, \qquad f_3(x) = x_2^2$$

from [Hub+06; Cus+11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} -42.09 & 31.81 & -31.79 & 34.78 & -26.46 & -24.63 & 34.09 & -25.46 & -29.13 & 27.43 \\ -19.91 & 28.12 & 39.50 & -31.29 & 33.61 & 36.84 & 23.41 & -8.45 & -2.93 & 14.10 \end{pmatrix}$$

Test Problem P.15 (IM1) (MOP_c) with $n = 2$, $q = 2$, constraint set
$\Omega = \{x \in \mathbb{R}^2 \mid x_1 \in [1, 4], x_2 \in [1, 2]\}$,

$$f_1(x) = 2\sqrt{x_1}, \qquad f_2(x) = x_1(1 - x_2) + 5$$

from [Hub+06; Cus+11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} 3.2670 & 3.7229 & 1.8244 & 2.0959 & 3.3221 & 2.9111 & 1.1437 & 1.8765 & 2.6472 & 3.7113 \\ 1.7063 & 1.9860 & 1.3932 & 1.5626 & 1.2897 & 1.2399 & 1.7155 & 1.5171 & 1.0256 & 1.9058 \end{pmatrix}$$

Test Problem P.16 (Jin1) (MOP_c) with $n \in \{2, 3, 4, 5, 10, 20, 30, 40, 50\}$, $q = 2$, constraint set $\Omega = [0, 1]^n$,

$$f_1(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \qquad f_2(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - 2)^2$$

from [JOS01; Cus+11] with f_1 declared as expensive and the starting points for $n = 2$

$$\begin{pmatrix} 0.2200 & 0.2682 & 0.8054 & 0.0638 & 0.1951 & 0.4301 & 0.0260 & 0.6411 & 0.0255 & 0.5701 \\ 0.3273 & 0.9832 & 0.6893 & 0.2257 & 0.3218 & 0.3591 & 0.8288 & 0.3008 & 0.5965 & 0.5174 \end{pmatrix}$$

Test Problem P.17 (Jin2) (MOP_c) with $n \in \{2, 3, 4, 5\}$, $q = 2$, constraint set $\Omega = [0, 1]^n$,

$$f_1(x) = x_1, \qquad f_2(x) = g(x)\left(1 - \sqrt{\frac{x_1}{g(x)}}\right),$$
$$g(x) = 1 + \frac{9}{n-1}\left(\sum_{i=2}^{n} x_i\right)$$

from [JOS01; Cus+11] with f_2 declared as expensive and the starting points for $n = 2$

$$\begin{pmatrix} 0.2200 & 0.2682 & 0.8054 & 0.0638 & 0.1951 & 0.4301 & 0.0360 & 0.6411 & 0.0255 & 0.5601 \\ 0.3273 & 0.9832 & 0.6893 & 0.2257 & 0.3218 & 0.3591 & 0.8288 & 0.3008 & 0.5965 & 0.5174 \end{pmatrix}$$

Test Problem P.18 (Jin3) (MOP_c) with $n \in \{2, 3, 4, 5\}$, $q = 2$, constraint set $\Omega = [0, 1]^n$,

$$f_1(x) = x_1, \qquad f_2(x) = 1 + \frac{9}{n-1} \sum_{i=2}^{n} x_i$$

from [JOS01; Cus+11] with f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.2200 & 0.2682 & 0.8054 & 0.0638 & 0.1951 & 0.4301 & 0.0360 & 0.6411 & 0.0255 & 0.5701 \\ 0.3273 & 0.9832 & 0.6893 & 0.2257 & 0.3218 & 0.3591 & 0.8288 & 0.3008 & 0.5965 & 0.5174 \end{pmatrix}$$

Test Problem P.19 (Jin4) (MOP_c) with $n \in \{2, 3, 4, 5\}$, $q = 2$, constraint set $\Omega = [0, 1]^n$,

$$f_1(x) = x_1, \qquad f_2(x) = g(x) \left(1 - \left(\frac{x_1}{g(x)} \right)^{0.25} - \left(\frac{x_1}{g(x)} \right)^{4} \right),$$
$$g(x) = 1 + \frac{9}{n-1} \left(\sum_{i=2}^{n} x_i \right)$$

from [JOS01; Cus+11] with f_2 declared as expensive and the starting points for $n = 2$

$$\begin{pmatrix} 0.2200 & 0.2682 & 0.8054 & 0.0638 & 0.1951 & 0.4301 & 0.0360 & 0.6411 & 0.0255 & 0.5601 \\ 0.3273 & 0.9832 & 0.6893 & 0.2257 & 0.3218 & 0.3591 & 0.8288 & 0.3008 & 0.5965 & 0.5174 \end{pmatrix}$$

Test Problem P.20 (JOS3) (MOP_c) with $n = 3$, $q = 2$, constraint set $\Omega = [0, 1]^3$,

$$f_1(x) = \sum_{i=1}^{n} \frac{x_i^2}{n}, \qquad f_2(x) = \sum_{i=1}^{n} \frac{(x_i - 2)^2}{n}$$

from [Hub+06] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} 0.8407 & 0.2435 & 0.1966 & 0.4733 & 0.5853 & 0.2858 & 0.3804 & 0.0540 & 0.9340 & 0.4694 \\ 0.2543 & 0.9293 & 0.2511 & 0.3517 & 0.5497 & 0.7572 & 0.5678 & 0.5308 & 0.1299 & 0.0119 \\ 0.8143 & 0.3500 & 0.6160 & 0.8308 & 0.9172 & 0.7537 & 0.0759 & 0.7792 & 0.5688 & 0.3371 \end{pmatrix}$$

Test Problem P.21 (Kursawe) (MOP_c) with $n = 3, q = 2$, constraint set $\Omega = [-5, 5]^n$,

$$f_1(x) = \sum_{i=1}^{n-1}(-10\exp(-0.2\sqrt{x_i^2 + x_{i+1}^2}))$$
$$f_2(x) = \sum_{i=1}^{n}(|x_i|^{0.8} + 5\sin(x_i)^3)$$

from [Kur91; Cus+11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} 0.38 & -1.37 & 0.09 & 0.16 & 1.91 & 4.88 & 2.35 & 2.08 & 1.26 & -0.55 \\ -0.72 & -0.12 & -0.53 & 1.19 & 3.64 & 1.63 & 1.64 & 1.29 & -4.14 & -0.05 \\ -1.00 & -2.35 & 4.05 & 3.01 & -3.68 & -3.20 & -4.97 & 4.47 & -3.67 & 0.78 \end{pmatrix}$$

Test Problem P.22 (Laumanns) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [-50, 50]^2$,

$$f_1(x) = x_1^2 + x_2^2, \qquad f_2(x) = (x_1 + 2)^2 + x_2^2$$

from [Hub+06] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} -33.78 & -18.88 & -33.44 & -23.70 & 18.92 & -4.95 & -27.10 & -34.76 & 3.83 & -42.18 \\ 29.43 & 2.85 & 10.20 & 15.41 & 24.82 & -41.62 & 41.33 & 32.58 & 49.61 & -5.73 \end{pmatrix}$$

Test Problem P.23 (LE1) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [-5, 10]^2$,

$$f_1(x) = \sqrt[8]{x_1^2 + x_2^2}, \qquad f_2(x) = \sqrt[4]{(x_1 - 0.5)^2 + (x_2 - 0.5)^2}$$

from [Hub+06; Cus+11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} -3.4002 & -4.9305 & 0.4910 & -3.7335 & 4.0213 & 1.4712 & -2.2723 & -2.8169 & 8.0394 & 3.2479 \\ 9.4285 & 6.6237 & 0.5142 & 0.9967 & 4.1553 & 8.6597 & -1.0430 & -2.9590 & 3.6956 & -2.8257 \end{pmatrix}$$

Test Problem P.24 (Lis) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [-5, 10]^2$,

$$f_1(x) = \sqrt[8]{x_1^2 + x_2^2}, \qquad f_2(x) = \sqrt[8]{(x_1 - 0.5)^2 + (x_2 - 0.5)^2}$$

from [CVL02; Hub+06] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} 7.2799 & 6.9341 & -0.1400 & -0.5411 & 6.3488 & 4.5612 & 0.9712 & 8.8727 & 9.7414 & 5.1394 \\ 2.1038 & -3.3389 & 9.5712 & 1.9713 & 5.0128 & -4.3926 & 4.5116 & 5.6561 & 2.4254 & 6.2769 \end{pmatrix}$$

Test Problem P.25 (lovison1) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [0, 3]^2$ or $\Omega = \mathbb{R}^2$,

$$f_1(x) = 1.05x_1^2 + 0.98x_2^2, \qquad f_2(x) = 0.99(x_1 - 3)^2 + 1.03(x_2 - 2.5)^2$$

from [Cus+11; Lov11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} 1.0664 & 1.7012 & 1.5035 & 2.3280 & 1.9678 & 2.2750 & 2.3757 & 0.0491 & 2.1827 & 1.8831 \\ 2.4960 & 0.4918 & 2.9481 & 2.4247 & 1.8908 & 2.1709 & 0.7280 & 2.0835 & 1.8225 & 0.0382 \end{pmatrix}$$

Test Problem P.26 (lovison2) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = \{x \in \mathbb{R}^2 \mid x_1 \in [-0.5, 0], x_2 \in [-0.5, 0.5]\}$ or $\Omega = \mathbb{R}^n$,

$$f_1(x) = x_2, \qquad f_2(x) = -\frac{x_2 - x_1^3}{x_1 + 1}$$

from [Cus+11; Lov11] with f_2 declared as expensive and the starting points

$$\begin{pmatrix} -0.01 & -0.04 & -0.26 & -0.20 & -0.35 & -0.12 & -0.33 & -0.38 & -0.31 & -0.47 \\ 0.09 & -0.23 & -0.36 & 0.15 & 0.42 & 0.42 & -0.17 & 0.43 & -0.08 & 0.09 \end{pmatrix}$$

Test Problem P.27 (lovison3) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = \{x \in \mathbb{R}^2 \mid x_1 \in [0, 6], x_2 \in [-4, 4]\}$ or $\Omega = \mathbb{R}^n$,

$$f_1(x) = x_1^2 + x_2^2, \qquad f_2(x) = (x_1 - 6)^2 - (x_2 + 0.3)^2$$

from [Cus+11; Lov11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} 1.3158 & 5.8931 & 5.8811 & 1.4557 & 4.5694 & 0.2416 & 3.4611 & 4.2050 & 3.1811 & 2.9385 \\ -2.0845 & -0.9588 & -2.2810 & 0.5870 & 3.8155 & -2.8964 & -3.1304 & 0.7960 & -0.2692 & 3.2582 \end{pmatrix}$$

Test Problem P.28 (lovison4) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = \{x \in \mathbb{R}^2 \mid x_1 \in [0, 6], x_2 \in [-1, 1]\}$ or $\Omega = \mathbb{R}^n$,

$$f_1(x) = x_1^2 + x_2^2 + 4(\exp(-(x_1 + 2)^2 - x_2^2) + \exp(-(x_1 - 2)^2 - x_2^2)),$$
$$f_2(x) = (x_1 - 6)^2 + (x_2 + 0.5)^2$$

from [Cus+11; Lov11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} [4.1310 & 0.0971 & 1.1050 & 4.7880 & 5.3449 & 1.7134 & 4.7925 & 5.2571 & 1.1353 & 2.4022 \\ 0.3978 & 0.3016 & -0.7318 & 0.4300 & 0.8027 & 0.2935 & -0.7054 & -0.3587 & -0.0336 & 0.5774 \end{pmatrix}$$

Test Problem P.29 (MOP1) (MOP) with $n = 1$, $q = 2$,

$$f_1(x) = x^2, \qquad f_2(x) = (x - 2)^2$$

from [Hub+06; Cus+11] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} -8.27 & -45.03 & 40.27 & 44.48 & -0.91 & -1.07 & -16.23 & 40.01 & -13.08 & -38.88 \end{pmatrix}$$

Test Problem P.30 (Schaffer2) (MOP_c) with $n = 1$, $q = 2$, constraint set $\Omega = [-5, 10]$,

$$f_1(x) = \begin{cases} -x & , \text{ if } x \leq 1 \\ x - 2 & , \text{ if } 1 < x \leq 3 \\ 4 - x & , \text{ if } 3 < x \leq 4 \\ x - 4 & , \text{ if } x > 4 \end{cases}, \qquad f_2(x) = (x - 5)^2$$

from [Sch85] with f_1 declared as expensive and the starting points

$$\begin{pmatrix} 6.7038 & 0.8461 & -1.3746 & 1.0587 & -3.5532 & -3.0204 & 9.1308 & 9.3420 & 3.6281 & -4.1033 \end{pmatrix}$$

Test Problem P.31 (T1) (MOP) with $n = 2$, $q = 2$,

$$f_1(x) = \frac{1}{2}x_1^2 + x_2^2 - 10x_1 - 100, \qquad f_2(x) = x_1^2 + \frac{1}{2}x_2^2 - 10x_2 - 100,$$

f_1 declared as expensive. and the starting points

$$\begin{pmatrix} -34.24 & 45.72 & 30.03 & -7.82 & 29.22 & 15.57 & 34.91 & 17.87 & 24.31 & 15.55 \\ 47.06 & -1.46 & -35.81 & 41.57 & 45.95 & -46.43 & 43.40 & 25.77 & -10.78 & -32.88 \end{pmatrix}$$

Test Problem P.32 (T2) (MOP) with $n = 2$, $q = 2$,

$$f_1(x) = \sin x_2, \qquad f_2(x) = 1 - \exp\left(-\left(x_1 - \frac{1}{\sqrt{2}}\right)^2 - \left(x_2 - \frac{1}{\sqrt{2}}\right)^2\right),$$

f_2 declared as expensive. and the starting points

$$\begin{pmatrix} 20.60 & -22.31 & -40.29 & -0.40 & 45.02 & -6.13 & -0.01 & -31.31 & 0.51 & 0.90 \\ -46.82 & -45.38 & 32.35 & 1.00 & -46.56 & -11.84 & -0.51 & -1.02 & 1.00 & 1.51 \end{pmatrix}$$

Test Problem P.33 (T3) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = [-2, 2]^2$,

$$f_1(x) = x_1 + 2, \qquad f_2(x) = x_1 - 2 + x_2,$$

f_2 declared as expensive and the starting points

$$\begin{pmatrix} 2.5763 & -0.0793 & -0.3193 & 0.0511 & 1.9058 & 0.8659 & 1.8695 & -0.8956 & 2.2557 & 0.7349 \\ 1.6543 & -0.3848 & -1.1619 & 0.0646 & 1.7690 & -0.7283 & 0.1970 & 2.6340 & 0.3009 & 0.5223 \end{pmatrix}$$

Test Problem P.34 (T4) (MOP_c) with $n \in \{2, 3, 4, 5, 10, 20, 30, 40, 50\}$, $q = 2$, constraint set $\Omega = [-10, 10]^n$,

$$f_1(x) = \sum_{i=1}^{n-1} x_i^2 + 2, \qquad f_2(x) = \sum_{i=1}^{n} x_i - 2,$$

f_1 declared as expensive and the starting points for $n = 2$

$$\begin{pmatrix} 8.2401 & -5.8961 & 0.0861 & 3.7051 & -0.4188 & 2.4755 & -9.0298 & 9.8070 & 2.5461 & -8.1104 \\ 2.7963 & 9.4970 & -4.3330 & -8.8077 & 0.7261 & 6.2651 & 3.0539 & -0.4131 & 5.1145 & -2.6839 \end{pmatrix}$$

Test Problem P.35 (T5) (MOP_c) with $n = 2$, $q = 2$, constraint set $\Omega = (0, 30] \times [0, 30]$,

$$f_1(x) = x_1 \ln(x_1) + x_2^2, \qquad f_2(x) = x_1^2 + x_2^4,$$

f_1 declared as expensive and the starting points

$$\begin{pmatrix} 25.9816 & 4.2130 & 7.7573 & 2.1431 & 10.4985 & 22.1156 & 18.7311 & 29.6945 & 25.6892 & 14.5115 \\ 2.6608 & 19.5013 & 24.4573 & 8.8787 & 6.5516 & 9.8276 & 9.1069 & 23.0979 & 25.2217 & 3.8922 \end{pmatrix}$$

Test Problem P.36 (T6) (MOP_c) with $n = 2, q = 2$, constraint set $\Omega = (0, 100]^2$,

$$f_1(x) = -\ln(x_1) - \ln(x_2), \qquad f_2(x) = x_1^2 + x_2,$$

f_1 declared as expensive and the starting points

$$\begin{pmatrix} 9.5896 & 95.6433 & 61.6860 & 23.3495 & 44.4905 & 5.8716 & 72.6812 & 25.4441 & 42.2296 & 44.4367 \\ 9.4335 & 20.8035 & 50.6963 & 6.1688 & 2.7671 & 52.5858 & 58.6400 & 7.8104 & 35.9188 & 17.2381 \end{pmatrix}$$

Test Problem P.37 (T7) (MOP_c) with $n = 3, q = 2$, constraint set $\Omega = [0, 30]^3$,

$$f_1(x) = \sum_{i=1}^{n} x_i^4 + \sum_{i=1}^{n} x_i^3, \qquad f_2(x) = \sum_{i=1}^{n} x_i,$$

f_1 declared as expensive and the starting points

$$\begin{pmatrix} 24.4417 & 27.4013 & 8.3549 & 28.9467 & 28.7150 & 4.2566 & 23.7662 & 1.0714 & 20.3621 & 11.7668 \\ 27.1738 & 18.9708 & 16.4064 & 4.7284 & 14.5613 & 12.6528 & 28.7848 & 25.4739 & 22.7322 & 19.6643 \\ 3.8096 & 2.9262 & 28.7252 & 29.1178 & 24.0084 & 27.4721 & 19.6722 & 28.0198 & 22.2940 & 5.1356 \end{pmatrix}$$

Test Problem P.38 (T8) (MOP_c) with $n = 3, q = 3$, constraint set
$\Omega = (0, 10] \times [0, 10] \times [0, 10]$

$$f_1(x) = \sum_{i=1}^{n} x_i^3, \qquad f_2(x) = \sum_{i=1}^{n-1} (x_i - 4)^2 + x_n^2$$
$$f_3(x) = -\ln(x_1) + 5 \sum_{i=2}^{n} x_i^2,$$

f_3 declared as expensive and the starting points

$$\begin{pmatrix} 5.5838 & 0.9029 & 3.0902 & 9.7538 & 1.1705 & 7.5247 & 2.2173 & 7.4561 & 5.6927 & 8.3627 \\ 7.2176 & 8.0288 & 8.5613 & 6.3541 & 7.8634 & 7.5924 & 9.9288 & 2.5898 & 9.6926 & 4.2902 \\ 4.6193 & 0.3624 & 0.5026 & 2.9378 & 0.1012 & 5.8483 & 2.7427 & 4.1388 & 9.2055 & 3.9705 \end{pmatrix}$$

Test Problem P.39 (VU1) (MOP_c) with $n = 2, q = 2$, constraint set $\Omega = [-3, 3]^2$,

$$f_1(x) = \frac{1}{x_1^2 + x_2^2 + 1}, \qquad f_2(x) = x_1^2 + 3x_2^2 + 1$$

from [Hub+06; Cus+11] f_1 declared as expensive and the starting points

$$\begin{pmatrix} 0.57 & 0.62 & -1.67 & -1.22 & -0.46 & -2.49 & 1.81 & 2.57 & -0.07 & -1.58 \\ -1.43 & 1.27 & -2.30 & -1.09 & 0.05 & -1.43 & -2.82 & 1.38 & 0.47 & -0.25 \end{pmatrix}$$

Test Problem P.40 (VU2) (MOP_c) with $n = 2, q = 2$, constraint set $\Omega = [-3, 3]^2$,

$$f_1(x) = x_1 + x_2 + 1, \qquad f_2(x) = x_1^2 + 2x_2 - 1$$

from [Hub+06; Cus+11] with f_2 declared as expensive and the starting points

$$\begin{pmatrix} 2.5763 & -0.0793 & -0.3193 & 0.0511 & 1.9058 & 0.8659 & 1.8695 & -0.8956 & 2.2557 & 0.7349 \\ 1.6543 & -0.3848 & -1.1619 & 0.0646 & 1.7690 & -0.7283 & 0.1970 & 2.6340 & 0.3009 & 0.5223 \end{pmatrix}$$

Test Problem P.41 (ZDT1) (MOP_c) with $n = 4, q = 2$, constraint set $\Omega = [0, 1]^4$,

$$\begin{aligned} &f_1(x) = x_1, \qquad f_2(x) = g(x)h(x), \\ &g(x) = 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i, \qquad h(x) = 1 - \sqrt{\frac{x_1}{g(x)}} \end{aligned}$$

from [ZDT00; Cus+11] with f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.8147 & 0.6324 & 0.9575 & 0.9572 & 0.4418 & 0.6557 & 0.6787 & 0.6555 & 0.2769 & 0.6948 \\ 0.9058 & 0.0975 & 0.9649 & 0.4854 & 0.8157 & 0.0457 & 0.7587 & 0.1712 & 0.1462 & 0.3171 \\ 0.1270 & 0.2785 & 0.1576 & 0.8003 & 0.7922 & 0.8491 & 0.5721 & 0.7060 & 0.0871 & 0.9702 \\ 0.9134 & 0.5469 & 0.9706 & 0.1419 & 0.9595 & 0.9340 & 0.3192 & 0.0318 & 0.8235 & 0.0985 \end{pmatrix}$$

Test Problem P.42 (ZDT2) (MOP_c) with $n = 4, q = 2$, constraint set $\Omega = [0, 1]^4$,

$$\begin{aligned} &f_1(x) = x_1, \qquad f_2(x) = g(x)h(x), \\ &g(x) = 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i, \qquad h(x) = 1 - \left(\frac{x_1}{g(x)}\right) \end{aligned}$$

from [ZDT00; Cus+11] with f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.4387 & 0.1869 & 0.7094 & 0.6551 & 0.9597 & 0.7513 & 0.8909 & 0.1493 & 0.8143 & 0.1966 \\ 0.3816 & 0.4898 & 0.7547 & 0.1626 & 0.3404 & 0.2551 & 0.9593 & 0.2575 & 0.2435 & 0.2511 \\ 0.7655 & 0.4456 & 0.2760 & 0.1190 & 0.5853 & 0.5060 & 0.5472 & 0.8407 & 0.9293 & 0.6160 \\ 0.7952 & 0.6463 & 0.6797 & 0.4984 & 0.2238 & 0.6991 & 0.1386 & 0.2543 & 0.3500 & 0.4733 \end{pmatrix}$$

Test Problem P.43 (ZDT3) (MOP_c) with $n = 4, q = 2$, constraint set $\Omega = [0, 1]^4$,

$$f_1(x) = x_1, \qquad f_2(x) = g(x)h(x),$$
$$g(x) = 1 + \frac{9}{n-1}\sum_{i=2}^{n} x_i, \qquad h(x) = 1 - \sqrt{\frac{x_1}{g(x)}} - \frac{x_1}{g(x)}\sin(10\pi x_1)$$

from [ZDT00; Cus+11] with f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.3517 & 0.9172 & 0.3804 & 0.5408 & 0.5588 & 0.1622 & 0.1656 & 0.6992 & 0.5390 & 0.5383 \\ 0.8508 & 0.2858 & 0.5678 & 0.7802 & 0.4994 & 0.7543 & 0.6220 & 0.7882 & 0.9233 & 0.9961 \\ 0.5853 & 0.7572 & 0.0769 & 0.9360 & 0.0219 & 0.3312 & 0.2830 & 0.4605 & 0.2574 & 0.0882 \\ 0.5497 & 0.7537 & 0.0540 & 0.1199 & 0.3371 & 0.5285 & 0.6641 & 0.1938 & 0.8248 & 0.4427 \end{pmatrix}$$

Test Problem P.44 (ZDT4) (MOP_c) with $n = 4, q = 2$, constraint set $\Omega = \{x \in \mathbb{R}^4 \mid x_1 \in [0, 1], x_2 \in [-5, 5]\}$,

$$f_1(x) = x_1, \qquad f_2(x) = g(x)h(x),$$
$$g(x) = 1 + 10(n-1) + \sum_{i=2}^{n}(x_i^2 - 10\cos(4\pi x_i)), \qquad h(x) = 1 - \sqrt{\frac{x_1}{g(x)}}$$

from [ZDT00; Cus+11] with f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.1425 & 0.3476 & 0.9365 & 0.7182 & 0.6418 & 0.7587 & 0.5276 & 0.9532 & 0.4829 & 0.5519 \\ 2.6474 & 2.3220 & -3.9110 & 4.2052 & -1.9606 & 4.2151 & 4.8861 & -3.5517 & -1.9149 & -2.5197 \end{pmatrix}$$

Test Problem P.45 (ZDT6) (MOP_c) with $n = 4, q = 2$, constraint set $\Omega = [0, 1]^4$,

$$f_1(x) = 1 - \exp(-4x_1)\sin(6\pi x_1)^6, \qquad f_2(x) = g(x)h(x),$$

$$g(x) = 1 + 9\left(\frac{1}{n-1}\sum_{i=2}^{n} x_i\right)^{0.25}, \qquad h(x) = 1 - \left(\frac{f_1(x)}{g(x)}\right)^2$$

from [ZDT00; Cus+11] with f_2 declared as expensive and the starting points

$$\begin{pmatrix} 0.1067 & 0.8173 & 0.2599 & 0.1818 & 0.8793 & 0.8530 & 0.4018 & 0.1839 & 0.9027 & 0.3377 \\ 0.9619 & 0.8687 & 0.8001 & 0.2638 & 0.5797 & 0.6221 & 0.0860 & 0.2400 & 0.9448 & 0.9001 \\ 0.0046 & 0.0844 & 0.4314 & 0.1455 & 0.5499 & 0.3510 & 0.2399 & 0.4173 & 0.4909 & 0.3692 \\ 0.7749 & 0.3998 & 0.9106 & 0.1361 & 0.1450 & 0.5132 & 0.1233 & 0.0497 & 0.4793 & 0.1112 \end{pmatrix}$$

Test Problem P.46 (ZLT1) (MOP_c) with $n = 4, q = 3$, constraint set $\Omega = [-1000, 1000]^4$,

$$f_1(x) = (x_1 - 1)^2 + \sum_{i=2}^{n} x_i^2, \qquad f_2(x) = (x_2 - 1)^2 + \sum_{i=1, i\neq 2}^{n} x_i^2,$$
$$f_3(x) = (x_3 - 1)^2 + \sum_{i=1, i\neq 3}^{n} x_i^2$$

from [Hub+06; Cus+11] with f_3 declared as expensive and the starting points

$$\begin{pmatrix} -479.91 & -887.75 & -354.17 & 395.99 & -206.12 & 464.09 & 489.24 & -377.91 & 726.57 & 393.62 \\ 569.06 & -552.98 & -949.71 & 827.02 & 319.87 & 231.57 & 445.98 & -159.59 & -510.01 & 185.94 \\ -661.59 & -909.44 & 465.65 & -232.91 & 645.90 & 510.38 & 713.44 & -824.52 & -132.80 & -71.62 \\ -451.48 & 501.98 & -54.43 & -513.78 & 244.73 & -431.07 & 837.64 & 119.74 & -729.25 & -517.69 \end{pmatrix}$$

List of Nomenclature

Descent direction	Direction in which function values of all objective functions are decreasing (Definition 2.4)
Efficient point	Optimal point for multi-objective optimization problem (Definition 2.1)
Pareto critical point	Point fulfilling a necessary condition for local weak efficiency for unconstrained optimization problems (Definition 2.7)
Pareto front	Set of images of all efficient points (Definition 2.1)
Stationary point	Point fulfilling a necessary condition for local weak efficiency for box-constrained optimization problems, analogous to Pareto critical point (Definition 4.28)
Weakly efficient point	Weakly optimal point for multi-objective optimization problem (Definition 2.1)

List of Abbreviations

DMS	Direct MultiSearch for multi-objective optimization, direct search approach from [Cus+11]
EFOS	Expensive Function Optimization Solver, trust region approach from [The11]
MHT	Multi-objective Heterogeneous Trust region algorithm (Algorithm 4)
$\mathbf{MHT_{combi}}$	Multi-objective Heterogeneous Trust region algorithm with combination of splitting and spreading (Algorithm 10)
$\mathbf{MHT_{lb}}$	Multi-objective Heterogeneous Trust region algorithm with lower bound in image space (Algorithm 7)
$\mathbf{MHT_{split}}$	Multi-objective Heterogeneous Trust region algorithm with image space split (Algorithm 8)
$\mathbf{MHT_{spread}}$	Multi-objective Heterogeneous Trust region algorithm with spreading (Algorithm 6)
$\mathbf{MHT^{b}_{spread}}$	Multi-objective Heterogeneous Trust region algorithm with spreading on bounded area (Algorithm 9)

List of Symbols

$\|\cdot\|$	Vector norm: euclidean norm $\|\cdot\| = \|\cdot\|_2$
$\|\cdot\|_F$	Matrix norm: Frobenius norm (compatible with euclidean norm)
$\mathbb{R}^q_+$	$\mathbb{R}^q_+ = \{x \in \mathbb{R} \mid x_i \geq 0 \text{ for all } i = 1, 2, \ldots, q\}$
$\mathbb{R}^q_{++}$	$\mathbb{R}^q_{++} = \{x \in \mathbb{R} \mid x_i > 0 \text{ for all } i = 1, 2, \ldots, q\}$
$\nabla f(x)$	Gradient of function f at x
$\nabla^2 f(x)$	Hessian matrix of function f at x
$f(A)$	Image set of function f under A, $f(A) = \{f(x) \mid x \in A\}$
$\text{int}A$	Interior of set A, $\text{int}A = \{x \in A \mid \exists\, U \subseteq A \text{ open with } x \in U\}$
∂A	Boundary of set A, $\partial A = A \setminus \text{int}A$
$\mathcal{S}$	Set of successful iterations
$\mathcal{V}$	Set of very successful iterations
0_q	Zero vector in $\mathbb{R}^q$, $0_q = (0, 0, \ldots, 0)^\top \in \mathbb{R}^q$

List of Tables

List of Figures

List of Algorithms

Bibliography

Primary Sources

[AD06] C. Audet and J.E. Dennis. "Mesh Adaptive Direct Search Algorithms for Constrained Optimization". In: *SIAM J. Optim.* 17 (1 2006), pp. 188–217. DOI: 10.1137/040603371 (cit. on p. 3).

[AH17] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2017. DOI: 10.1007/978-3-319-68913-5 (cit. on pp. 2, 26, 29, 34, 49, 54, 75, 78).

[AM11] K.P. Anagnostopoulos and G. Mamanis. "Multiobjective evolutionary algorithms for complex portfolio optimization problems". In: *Comput. Manag. Sci.* 8 (2011), pp. 259–279. DOI: 10.1007/s10287-009-0113-8 (cit. on p. 1).

[ASZ08] C. Audet, G. Savard, and W. Zgal. "Multiobjective Optimization through a Series of Single-Objective Formulations". In: *SIAM J. Optim.* 19 (1 2008), pp. 188–210. DOI: 10.1137/060677513 (cit. on p. 3).

[Ber99] D. P. Bertsekas. *Nonlinear programming*. Athena Scientific, 1999 (cit. on p. 100).

[BM13] D. Bienstock and A. Michalka. "Polynomial Solvability of Variants of the Trust-Region Subproblem". In: *Proceedings of the 25th Annual CM-SIAM Symposium on Discrete Algorithms*. 2013, pp. 380–390 (cit. on p. 20).

[Buh03] M. Buhmann. *Radial Basis Functions: Theory and Implementations*. Cambridge University Press, 2003. DOI: 10.1017/CBO9780511543241 (cit. on p. 40).

[Car+18] C. Cartis, L. Roberts, J. Fiala, B. Marteau, and K. Scheinberg. "Optimization with expensive and uncertain data (challenges and improvements)". In: *EUROPT Workshop on Advances in Continuous Optimization*. 2018 (cit. on p. 76).

[CGT00] A.R. Conn, N.I.M. Gould, and P.L. Toint. *Trust-Region Methods*. MPS-SIAM Series on Optimization, 2000. DOI: 10.1137/1.9780898719857 (cit. on pp. 4 sq., 19–23, 25, 29, 36 sq., 49–52, 54, 56 sq., 66, 75 sqq., 91, 93, 132, 135, 193, 202).

[CGT88] A.R. Conn, N.I.M. Gould, and P.L. Toint. "Global Convergence of a Class of Trust Region Algorithms for Optimization with Simple Bounds". In: *SIAM J. Numer. Anal.* 25 (2 1988), pp. 433–460. DOI: 10.1137/0725029 (cit. on pp. 4, 93, 202).

[CGT89] A.R. Conn, N.I.M. Gould, and P.L. Toint. "Correction to the Paper on Global Convergence of a Class of Trust Region Algorithms for Optimization with Simple Bounds". In: *SIAM J. Numer. Anal.* 26 (3 1989), pp. 764–767 (cit. on p. 4).

[Chi08] J. W. Chinneck. *Feasibility and Infeasibility in Optimization:Algorithms and Computational Methods*. Springer US, 2008. DOI: 10.1007/978-0-387-74932-7 (cit. on p. 125).

[Chu+17] T. Chugh, K. Sindhya, J. Hakanen, and K. Miettinen. "A survey on handling computationally expensive multiobjective optimization problems with evolutionary algorithms". In: *Soft Computing* (2017), pp. 1–30. DOI: 10.1007/s00500-017-2965-0 (cit. on p. 4).

[CL96] T.F. Coleman and Y. Li. "An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds". In: *SIAM J. Optim.* 6 (2 1996), pp. 418–445. DOI: 10.1137/0806023 (cit. on pp. 93, 202).

[CL99] F.Y. Cheng and X.S. Li. "Generalized Center Method for Multiobjective Engineering Optimization". In: *Eng. Opt.* 31 (1999), pp. 641–661. DOI: 10.1080/03052159908941390 (cit. on pp. 137 sq., IV).

[CLM16] G. A. Carrizo, P. A. Lotito, and M. C. Maciel. "Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem". In: *Math. Program.* 159 (2016), pp. 339–369. DOI: 10.1007/s10107-015-0962-6 (cit. on pp. 4 sq., 49–52).

[CSV09a] A.R. Conn, K. Scheinberg, and L.N. Vicente. "Global Convergence of General Derivative-Free Trust-Region Algorithms to First- and Second-Order Critical Points". In: *SIAM J. Optim.* 20 (1 2009), pp. 387–415. DOI: 10.1137/060673424 (cit. on p. 4).

[CSV09b] A.R. Conn, K. Scheinberg, and L.N. Vicente. *Introduction to Derivative-Free Optimization*. MPS-SIAM Series on Optimization, 2009. DOI: 10.1137/1.9780898718768 (cit. on pp. 2–5, 19 sq., 23, 25 sq., 28 sqq., 32, 49, 52, 54, 57, 66, 76, 78, 135).

[Cus+11] A.L. Custodio, J.F.A. Madeira, A.I.F. Vaz, and L.N. Vicente. "Direct Multisearch for Multiobjective Optimization". In: *SIAM J. Optim.* 21 (3 2011), pp. 1109–1140. DOI: 10.1137/10079731X (cit. on pp. 3, 137–141, 147, 150, 153, 155, IV–XIII, XVI sqq., XXII).

[CVL02] C.A.C. Coello, D.A. Van Veldhuizen, and G.B. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic/Plenum Publishers, 2002. DOI: 10.1007/978-0-387-36797-2 (cit. on pp. 3, 137 sq., XII).

[Dav01] P. A. Davidson. *An introduction to magnetohydrodynamics*. Cambridge University Press, 2001. DOI: 10.1017/CBO9780511626333 (cit. on pp. 189 sq.).

[DD98] I. Das and J. E. Dennis. "Normal-Boundary Intersection: a New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems". In: *SIAM J. Optim.* 8 (3 1998), pp. 631–657. DOI: 10.1137/S1052623496307510 (cit. on p. 2).

[Deb+02] K. Deb, L. Thiele, M. Laumanns, and E. Zitzler. "Scalable multi-objective optimization test problems". In: *Proceedings of the 2002 Congress on Evo-*

lutionary Computation. CEC'02. Vol. 1. 2002, pp. 825–830. DOI: 10.1109/CEC.2002.1007032 (cit. on pp. 137 sq., VI).

[Deb01] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, Inc., 2001 (cit. on p. 3).

[Deb99] K. Deb. "Multiobjective Genetic Algorithms: Problem Difficulties and Construction of Test Problems". In: *Evolutionary Computation* 7 (3 1999), pp. 205–230. DOI: 10.1162/evco.1999.7.3.205 (cit. on pp. 137 sq., 147, V sq.).

[DS05] L.M.G. Drummond and B. F. Svaiter. "A steepest descent method for vector optimization". In: *J. Comput. Appl. Math.* 175 (2005), pp. 395–414. DOI: 10.1016/j.cam.2004.06.018 (cit. on pp. 3 sqq., 9, 12).

[DS15] L.G. Drummond and B.F. Svaiter. "A general framework for multicriteria descent methods". In: *Conference EURO*. 2015 (cit. on p. 14).

[Ehr05] M. Ehrgott. *Multicriteria Optimization*. 2nd ed. Springer Berlin Heidelberg, 2005. DOI: 10.1007/3-540-27659-9 (cit. on p. 7).

[Eic08] G. Eichfelder. *Adaptive Scalarization Methods in Multiobjective Optimization*. Springer, 2008. DOI: 10.1007/978-3-540-79159-1 (cit. on pp. 1, 41, 44, 80, 94, 112).

[Eic09] G. Eichfelder. "An Adaptive Scalarization Method in Multiobjective Optimization". In: *SIAM J. Optim.* 19 (4 2009), pp. 1694–1718. DOI: 10.1137/060672029 (cit. on p. 2).

[FDS09] J. Fliege, L.M. Graña Drummond, and B.F. Svaiter. "Newton's method for multiobjective optimization". In: *SIAM J. Optim.* 20 (2 2009), pp. 602–626. DOI: 10.1137/08071692X (cit. on pp. 3 sq., 11 sq., 16).

[FF95] C.M. Fonseca and P.J. Fleming. "An Overview of Evolutionary Algorithms in Multiobjective Optimization". In: *Evolutionary Computation* 3 (1 1995), pp. 1–16. DOI: 10.1162/evco.1995.3.1.1 (cit. on pp. 137 sq., 151, VIII).

[FF98] C.M. Fonseca and P.J. Fleming. “Multiobjective optimization and multiple constraint handling with evolutionary algorithms. I. A unified formulation”. In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 28 (1 1998), pp. 26–37. DOI: 10.1109/3468.650319 (cit. on pp. 137 sq., VIII).

[Fla14] B. Flach. “Rekonstruktion mit Vorinformation für die Niedrigstdosis-CT-Fluoroskopie in der Interventionellen Radiologie”. Dissertation. Friedrich-Alexander-Universität Erlangen Nürnberg, 2014 (cit. on pp. 1 sq.).

[FMS94] A. Friedlander, J.M. Martinez, and S.A. Santos. “A new trust region algorithm for bound constrained minimization”. In: *Applied Mathematics and Optimization* 30 (3 1994), pp. 235–266. DOI: 10.1007/BF01183013 (cit. on pp. 93, 202).

[FS00] J. Fliege and B. F. Svaiter. “Steepest Descent Methods for Multicriteria Optimization”. In: *Math. Methods of Oper. Res.* 51 (3 2000), pp. 479–494. DOI: 10.1007/s001860000043 (cit. on pp. 3 sqq., 9, 12, 96).

[FV16] J. Fliege and A.I.F. Vaz. “A Method for Constrained Multiobjective Optimization Based on SQP Techniques”. In: *SIAM J. Optim.* 26 (4 2016), pp. 2091–2119. DOI: 10.1137/15M1016424 (cit. on p. 94).

[GK99] C. Geiger and C. Kanzow. *Numerische Verfahren zur Lösung unrestringierter Optimierungsverfahren*. Springer, 1999. DOI: 10.1007/978-3-642-58582-1 (cit. on p. 15).

[Göp+03] A. Göpfert, H. Riahi, C. Tammer, and C. Zalinescu. *Variational Methods in Partially Ordered Spaces*. CMS Books in Mathematics. Springer-Verlag New York, 2003. DOI: 10.1007/b97568 (cit. on pp. 41, 80).

[Gor73] P. Gordan. “Über die Auflösung linearer Gleichungen mit reellen Coefficienten”. In: *Mathematische Annalen* 6 (1873), pp. 23–28 (cit. on p. 14).

[Gou+99] N. I. M.. Gould, S. Lucidi, M. Roma, and P.L. Toint. "Solving the Trust-Region Subproblem Using the Lanczos Method". In: *SIAM J. Optim.* 9 (2 1999), pp. 504–525. DOI: 10.1137/S1052623497322735 (cit. on p. 20).

[GPD18] B. Gebken, S. Peitz, and M. Dellnitz. "A Descent Method for Equality and Inequality Constrained Multiobjective Optimization Problems". arXiv: 17 12. 03005v2 [math.OC]. 2018-10-01. (Visited on 2018-10-20) (cit. on pp. 94–97, 202).

[Gre+15] S. Greco, K. Klamroth, J.D. Knowles, and G. Rudolph. "Understanding Complexity in Multiobjective Optimization (Dagstuhl Seminar 15031)". In: *Dagstuhl Reports* 5 (1 2015) (cit. on pp. 1 sq.).

[GW90] C. Gerth (Tammer) and P. Weidner. "Nonconvex Separation Theorems and Some Applications in Vector Optimization". In: *J. Optim. Theory Appl.* 67.2 (1990), pp. 297–320. DOI: 10.1007/BF00940478 (cit. on p. 35).

[HM79] C.-L. Hwang and A.S.MD. Masud. *Multiple Objective Decision Making—Methods and Applications: A State-of-the-Art Survey*. Vol. 164. Lecture Notes in Economics and Mathematical Systems. Springer Berlin Heidelberg, 1979. DOI: 10.1007/978-3-642-45511-7 (cit. on pp. 137 sq., VII).

[HP95] R. Horst and P. M. Pardalos. *Handbook of Global Optimization*. Springer US, 1995. DOI: 10.1007/978-1-4615-2025-2 (cit. on pp. 100, 120).

[HS52] M. R. Hestenes and E. Stiefel. "Methods of Conjugate Gradients for Solving Linear Systems". In: *Journal of Research of the National Bureau of Standards* 49.6 (1952), pp. 409–436 (cit. on p. 20).

[HT96] R. Horst and H. Tuy. *Global Optimization*. Deterministic Approaches. Springer-Verlag Berlin Heidelberg, 1996. DOI: 10.1007/978-3-662-03199-5 (cit. on pp. 100, 121).

[Hub+06] S. Huband, P. Hingston, L. Barone, and L. While. "A Review of Multiobjective Test Problems and a Scalable Test Problem Toolkit". In: *IEEE Trans.*

Evolutionary Computation 10 (5 2006), pp. 477–506. DOI: 10.1109/TEVC.2005.861417 (cit. on pp. 137 sqq., 141, IV, VI–IX, XI sqq., XVI, XVIII).

[Jah11] J. Jahn. *Vector Optimization*. Springer, 2011. DOI: 10.1007/978-3-642-17005-8 (cit. on pp. 7, 11, 15, 17).

[JM04] M.C. Joshi and K.M. Moudgalya. *Optimization: Theory and Practice*. Alpha Science International Limited, 2004 (cit. on p. 100).

[JOS01] Y. Jin, M. Olhofer, and B. Sendhoff. "Dynamic Weighted Aggregation for Evolutionary Multi-objective Optimization: Why Does It Work and How?" In: *Proceedings of the 3rd Annual Conference on Genetic and Evolutionary Computation*. GECCO'01. Morgan Kaufmann Publishers Inc., 2001, pp. 1042–1049 (cit. on pp. 137 sq., 150, 153, IX sq.).

[JR06] J. Jahn and U. Rathje. "Graef-Younes Method with Backward Iteration". In: *Multicriteria Decision Making and Fuzzy Systems - Theory, Methods and Applications*. (Aachen). Ed. by K.H. Küfer, H. Rommelfanger, C. Tammer, and K. Winkler. Shaker Verlag, 2006, pp. 75–81 (cit. on p. 129).

[JSW98] D. Jones, M. Schonlau, and W. Welch. "Efficient Global Optimization of Expensive Black-Box Functions". In: *Journal of Global Optimization* 13 (4 1998), pp. 455–492. DOI: 10.1023/A:1008306431147 (cit. on p. 4).

[KLT03] T.G. Kolda, R.M. Lewis, and V. Torczon. "Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods". In: *SIAM Rev.* 45 (3 2003), pp. 385–482. DOI: 10.1137/S003614450242889 (cit. on p. 3).

[KN08] J. Knowles and H. Nakayama. "Multiobjective Optimization. Interactive and Evolutionary Approaches". In: ed. by J. Branke, K. Deb, K. Miettinen, and R. Slowinski. Springer, 2008. Chap. Meta-Modeling in Multiobjective Optimization, pp. 258–284. DOI: 10.1007/978-3-540-88908-3 (cit. on p. 4).

[Kur91] F. Kursawe. "A Variant of Evolution Strategies for Vector Optimization". In: *Proceedings of the 1st Workshop on Parallel Problem Solving from Nature*.

PPSN I. Berlin, Heidelberg: Springer, 1991, pp. 193–197 (cit. on pp. 137 sq., XI).

[KZB11] D. Krasnov, O. Zikanov, and T. Boeck. "Comparative study of finite difference approaches in simulation of magnetohydrodynamic turbulence at low magnetic Reynolds number". In: *Computers & Fluids* 50 (1 2011), pp. 46–59. DOI: 10.1016/j.compfluid.2011.06.015 (cit. on p. 191).

[Lov11] A. Lovison. "Singular Continuation: Generating Piecewise Linear Approximations to Pareto Sets via Global Analysis". In: *SIAM J. Optim.* 21 (2 2011), pp. 463–490. DOI: 10.1137/100784746 (cit. on pp. 138 sq., XII sq.).

[LPV05] D.T. Luc, T. Q. Phong, and M. Volle. "Scalarizing Functions for Generating the Weakly Efficient Solution Set in Convex Multiobjective Problems". In: *SIAM J. Optim.* 15 (4 2005), pp. 987–1001. DOI: 10.1137/040603097 (cit. on p. 2).

[Mie98] K.M. Miettinen. *Nonlinear Multiobjective Optimization.* Boston, USA: Springer US, 1998. DOI: 10.1007/978-1-4615-5563-6 (cit. on pp. 7, 15, 35).

[Mor83] J. J. Moré. "Recent Developments in Algorithms and Software for Trust Region Methods". In: *Mathematical Programming: The State of the Art.* Ed. by A. Bachem, M. Groetschel, and B. Korte. Springer, Berlin, Heidelberg. 1983, pp. 258–287. DOI: 10.1007/978-3-642-68874-4_11 (cit. on pp. 4, 20, 193).

[Pri+16] S. Prinz, V. Bandaru, Y. Kolesnikov, D. Krasnov, and T. Boeck. "Numerical simulations of magnetohydrodynamic flows driven by a moving permanent magnet". In: *Physical Review Fluids* 1 (4 2016). DOI: 10.1103/PhysRevFluids.1.043601 (cit. on p. 190).

[PS84] A. Pascoletti and P. Serafini. "Scalarizing Vector Optimization Problems". In: *J. Optim. Theory Appl.* 42 (4 1984), pp. 499–524. DOI: 10.1007/BF00934564 (cit. on p. 35).

[RK14] J.-H. Ryu and S. Kim. "A Derivative-free Trust-Region Method for Biobjective Optimization". In: *SIAM J. Optim.* 24 (1 2014), pp. 334–362. DOI: 10.1137/120864738 (cit. on pp. 4 sq., 33, 45, 47, 49, 52, 54, 57, 82, 135).

[Roy+18] P. C. Roy, J. Blank, R. Hussein, and K. Deb. "Trust-region Based Algorithms with Low-budget for Multi-objective Optimization". In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. GECCO '18. ACM Press, 2018, pp. 195–196. DOI: 10.1145/3205651.3205727 (cit. on p. 4).

[Sch85] J.D. Schaffer. "Multiple Objective Optimization with Vector Evaluated Genetic Algorithms". In: *Proceedings of the 1st International Conference on Genetic Algorithms*. L. Erlbaum Associates Inc., 1985, pp. 93–100 (cit. on pp. 137, 139, XIII).

[Ste+08] T. Stewart et al. "Multiobjective Optimization. Interactive and Evolutionary Approaches". In: ed. by J. Branke, K. Deb, K. Miettinen, and R. Slowinski. Springer, 2008. Chap. Real-World Applications of Multiobjective Optimization, pp. 285–327. DOI: 10.1007/978-3-540-88908-3 (cit. on p. 1).

[Ste83] T. Steihaug. "The Conjugate Gradient Method and Trust Regions in Large Scale Optimization". In: *SIAM J. Numer. Anal.* 20 (3 1983), pp. 626–637. DOI: 10.1137/0720042 (cit. on p. 20).

[Str+08] F. Strauss, J.-A. Desideri, R. Duvigneau, and V. Heuveline. "multiobjective optimization in hydrodynamic stability control". In: *[Research Report] RR-6608, INRIA* (2008). <inria-00309693> (cit. on p. 1).

[Tab+15] M. Tabatabaei, J. Hakanen, M. Hartikainen, K. Miettinen, and K. Sindhya. "A survey on handling computationally expensive multiobjective optimization problems using surrogates: non-nature inspired methods". In: *Structural and Multidisciplinary Optimization* 52 (1 2015), pp. 1–25. DOI: 10.1007/s00158-015-1226-z (cit. on p. 4).

[TC07] M.G.C. Tapia and C.A.C. Coello. "Applications of Multi-Objective Evolutionary Algorithms in Economics and Finance: A Survey". In: *proceedings*

of the IEEE Congress on Evolutionary Computation. 2007, pp. 532–539. DOI: 10.1109/CEC.2007.4424516 (cit. on p. 1).

[The11] A. Thekale. "Trust-Region Methods for Simulation Based Nonlinear Optimization". Dissertation. Friedrich-Alexander-Universität Erlangen Nürnberg, 2011 (cit. on pp. 49, 76, 140, XXII).

[Tho15] J. Thomann. "Decomposed descent methods in multiobjective optimization". Master thesis. Technische Universität Ilmenau, 2015 (cit. on pp. 9, 11, 14).

[TK91] A. W. Troesch and S. K. Kim. "Hydrodynamic forces acting on cylinders oscillating at small amplitudes". In: *Journal of Fluids and Structures* 5 (1 1991), pp. 113–126. DOI: 10.1016/0889-9746(91)80014-5 (cit. on p. 190).

[Toi81] P. Toint. "Towards an Efficient Sparsity Exploiting Newton Method for Minimization". In: *Sparse Matrices and Their Uses*. Ed. by I. S. Duff. London: Academic Press, 1981, pp. 57–88 (cit. on p. 20).

[TPE14] D. Terzijska, M. Porcelli, and G. Eichfelder. "Multi-objective optimization in the Lorentz force velocimetry framework". In: *Book of digest & program*. OIPE, International Workshop on Optimization and Inverse Problems in Electromagnetism 13, 2014. Chap. Book of digests & program, pp. 81–82 (cit. on p. 1).

[Vel99] D.A. Van Veldhuizen. "Multiobjective Evolutionary Algorithms: Classifications, Analyses and New Innovations". Dissertation. Wright Patterson AFB, OH, USA: Air Force Institute of Technology, 1999 (cit. on p. 137).

[VOS14] K.D.V. Villacorta, P.R. Oliveira, and A. Soubeyran. "A Trust-Region Method for Unconstrained Multiobjective Problems with Applications in Satisficing Processes". In: *J. Optim. Theory Appl.* 160 (3 2014), pp. 865–889. DOI: 10.1007/s10957-013-0392-7 (cit. on pp. 4 sq., 33, 36 sq., 45, 49–52, 57, 66, 82, 135).

[WS11] S.M. Wild and C. Shoemaker. “Global Convergence of Radial Basis Functon Trust Region Derivative-Free Algorithms”. In: *SIAM J. Optim.* 21 (3 2011), pp. 761–781. DOI: 10.1137/09074927X (cit. on pp. 40, 78).

[You93] Y. M. Younes. “Studies on discrete vector optimization”. Dissertation. University of Demiatta, 1993 (cit. on p. 129).

[ZDT00] E. Zitzler, K. Deb, and L. Thiele. “Comparison of multiobjective evolutionary algorithms: empirical results”. In: *Evol. Comput.* 8 (2 2000), pp. 173–195. DOI: 10.1162/106365600568202 (cit. on pp. 137, 139, XVI sqq.).

Publications

The following papers resulted from the work on this thesis. [TE18a] was submitted to a peer-reviewed journal. At the moment of submission of this thesis it is under second review of this journal and only available as preprint.

[TE18a] J. Thomann and G. Eichfelder. "A Trust Region Algorithm for Heterogeneous Multi-objective Optimization". In: *Preprint-Series of the Institute for Mathematics, Technische Universität Ilmenau, Germany* (2018). URL: http://www.optimization-online.org/DB_HTML/2018/03/6495.html (cit. on pp. 5, 33, XLI).

[TE18b] J. Thomann and G. Eichfelder. "Numerical Results for the Multi-objective Trust Region Algorithm MHT". 2018-08. URL: http://www.optimization-online.org/DB_HTML/2018/08/6768.html (cit. on pp. 5, 141).

Student Thesis

The following bachelor thesis was supervised in the context of this thesis. It is not available publicly. If desired, please query the author of this thesis for a copy.

[Lau18] L. Laudenberg. "Modellfunktionen in der multikriteriellen Optimierung". Bachelor thesis. Technische Universität Ilmenau, 2018-08-21 (cit. on pp. 31 sq.).

Eidesstattliche Erklärung

Ich versichere, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.
Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.
Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer Prüfungsbehörde vorgelegt.
Ich bin darauf hingewiesen worden, dass die Unrichtigkeit der vorstehenden Erklärung als Täuschungsversuch bewertet wird und gemäß § 7 Abs. 10 der Promotionsordnung den Abbruch des Promotionsverfahrens zur Folge hat.

Jana Thomann
Ilmenau, 27.11.2018