Technische Universität Ilmenau
Institut für Mathematik

Preprint No. M **18/04**

A trust region algorithm for heterogeneous multiobjective optimization

Jana Thomann and Gabriele Eichfelder

Februar 2018

URN: urn:nbn:de:gbv:ilm1-2018200043

Impressum:
Hrsg.: Leiter des Instituts für Mathematik
Weimarer Straße 25
98693 Ilmenau
Tel.: +49 3677 69-3621
Fax: +49 3677 69-3270
http://www.tu-ilmenau.de/math/

ilmedia

A Trust Region Algorithm for Heterogeneous Multiobjective Optimization *

Jana Thomann †and Gabriele Eichfelder ‡

28.02.2018

Abstract

This paper presents a new trust region method for multiobjective heterogeneous optimization problems. One of the objective functions is an expensive black-box function, for example given by a time-consuming simulation. For this function derivative information cannot be used and the computation of function values involves high computational effort. The other objective functions are given analytically and derivatives can easily be computed. The method uses the basic trust region approach by restricting the computations in every iteration to a local area and replacing the objective functions by suitable models. The search direction is generated in the image space by using local ideal points. It is proved that the presented algorithm converges to a Pareto critical point. Numerical results are presented and compared to another algorithm.

Key Words: multiobjective optimization, trust region method, derivative-free algorithm, heterogeneous optimization, Pareto critical point

Mathematics subject classifications (MSC 2010): 90C29, 90C56, 90C30

1 Introduction

Multiobjective optimization problems can be found in various fields, such as engineering, medicine, economics or finance [30, 15, 32, 1] where several conflicting objectives are optimized. An additional difficulty can arise if some of the objectives are not given analytically, but are a black box because they are the result of an experiment or a simulation run. This can include a long evaluation time for every function value and hence the number of function evaluations needs to be reduced. Black box functions can be smooth functions, that is derivatives do exist, but are not available with reasonable efforts. Hence using

*This work was funded by DFG under no. GRK 1567.

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

‡Institute for Mathematics, Technische Universität Ilmenau, Po 10 05 65, D-98684 Ilmenau, Germany, Gabriele.Eichfelder@tu-ilmenau.de

derivative information should be avoided and therefore many solution methods from the literature [14, 19, 20, 21] are not applicable.

In this paper we focus on smooth multiobjective optimization problems with so-called heterogeneous functions, i.e. the objective functions differ in certain aspects affecting the optimization process. There are different kinds of heterogeneity and various reasons why it can occur, this is discussed in [23, p.125f]. The heterogeneity considered in this paper is the different amount of information available for the functions and the computation time. For one of the objectives the function values are only obtained with high computational effort and derivatives are not available with reasonable effort. Such a function can be, for instance, a computationally expensive black-box function, not given analytically, but only by a time-consuming simulation. The other functions are given analytically and derivatives are easily available. These functions will be called cheap in contrast to the expensive function. Such multiobjective optimization problems with heterogeneous and expensive black box functions arise for example in engineering or medicine [33, 18, 31]. For instance in Lorentz force velocimetry [33] the task 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. While the first objective is an analytically given function, in general the second one can only be determined by a time-consuming simulation. According to [23, p.124] heterogeneous problems with expensive functions also occur in imaging techniques in interventional radiology [18]. Whereas one objective is the sum of squared differences and therefore analytically given, the other objective is described by physical models for fluids and diffusion processes given by an implicit differential equation.

In the literature there are a lot of solution methods for multiobjective optimization problems and one common approach is scalarization, that is combining the objectives to obtain a scalar-valued function and optimize this surrogate problem with known methods for scalar optimization problems. Among numerous scalarization approaches, e.g. [12, 16, 27], the weighted sum approach is a commonly known and used method. Every objective is assigned a positive weight - a scalar constant - and the weighted sum of all objectives is optimized. A problem for this approach and also for every scalarization technique is that whenever one of the objectives is an expensive function, the high computational effort 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 cannot exploit heterogeneity of objective functions and therefore neglect some information.

Other methods for multiobjective optimization problems, like the generalized steepest descent method [14, 20] or the generalized Newton method [19] need derivative information and are therefore not applicable to heterogeneous problems where the derivatives are not available with reasonable efforts. Approximating the derivatives is not an option due to the expensive black-box functions. Either the obtained approximation would not be viable or too many function evaluations would be necessary.

However, there are also derivative-free methods in multiobjective optimization and a very common approach, both in scalar and multiobjective optimization, is direct search [2, 10, 11]. This approach only needs function values and there are several versions and realizations such as the basic DMS [11] or BIMADS [3] for biobjective bound constrained problems where the structure of the objective functions is absent or unreliable. A disadvantage of these methods is the fact that the performance deteriorates if the number

of variables increases [26]. 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.
Another approach on which derivative-free methods are based on is the trust region method [6, 7, 8, 9, 10]. There are also multiobjective realizations of this approach [29, 36]. 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 properties are documented in the literature. A basic generalization of such a method to multiobjective problems based on derivative information is given in [36]. They proof convergence to a Pareto critical point using a characterization of such points that is also used in multiobjective descent theory [14, 20]. The needed assumptions are derived from the scalar version of trust region approaches and the convergence analysis follows the strategy and structure of the proof from the basic scalar approach [8] 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 presented here where using derivative information of the expensive function shall be avoided.
Unlike this in [29] a trust region algorithm is presented for biobjective expensive problems where derivative information is absent for both objectives. The algorithm uses a scalarization 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.
So far there are no solution methods for heterogeneous multiobjective problems that can exploit the differences of the objective functions. This paper will present a new trust region method that can regard heterogeneity. Like [36] we use the idea of generalizing the trust region approach to a multiobjective problem, but our algorithm differs in computing the descent direction and not needing the gradient of the objectives. The search direction is computed in the image space by using a local ideal point. 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, we can use the same strategy to prove convergence to a Pareto critical point as [36] also using the characterization of such points from [20]. Since we also follow closely the basic scalar idea of trust region methods, the convergence analysis is also related to that in the scalar case [8].
The paper is organized as follows. The basic theory is presented in section 2 followed by the description of the multiobjective trust region method in section 3 and the convergence analysis in section 4. Numerical details and modifications for the implementation of the Algorithm are discussed in section 5, experimental results are in section 6 and the conclusions follow in section 7.

2 Problem statement and basic definitions

The optimization problem considered in this paper is described by

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

with $f(x) = (f_1(x), ..., f_q(x))^\top$. The objective functions $f_i : \mathbb{R}^n \to \mathbb{R}$ are assumed to be twice continuously differentiable for all $i = 1, ..., q$ and $\max_{i=1,...,q} f_i(x)$ is assumed to be bounded from below. The function f_1 is a so-called expensive function, which is not given analytically but only by a time-consuming simulation. The simulation only gives function values, derivative information is not available with reasonable effort and therefore not used. The other objective functions $f_i, i = 2, ..., q$, are so-called cheap functions, which are analytically given, easy to compute and derivatives are easily available. For defining solutions of (MOP) we use the optimality concept for multiobjective optimization problems according to [25].

Definition 2.1 *A point $\overline{x} \in \mathbb{R}^n$ is called efficient (solution) for (MOP) (or Pareto optimal), if there exists no point $x \in \mathbb{R}^n$ satisfying $f_i(x) \le f_i(\overline{x})$ for all $i \in \{1, ..., q\}$ and $f(x) \neq f(\overline{x})$.*
A point $\overline{x} \in \mathbb{R}^n$ is called weakly efficient (solution) for (MOP) (or weakly Pareto optimal), if there exists no point $x \in \mathbb{R}^n$ satisfying $f_i(x) < f_i(\overline{x})$ for all $i \in \{1, ..., q\}$.
These concepts can be restricted to local areas. Accordingly, a point $\overline{x} \in \mathbb{R}^n$ is called locally (weak) efficient for (MOP) if there exists a neighborhood $U \subset \mathbb{R}^n$ with $\overline{x} \in U$ such that $\overline{x}$ is (weakly) efficient for (MOP) in U.

Obviously every efficient point is weakly efficient. The following concept [20] gives a necessary condition for weak efficiency.

Definition 2.2 *Let $f = (f_1, ..., f_q)$ be totally differentiable at a point $\overline{x} \in \mathbb{R}^n$. This point is called Pareto critical for (MOP), if for every vector $d \in \mathbb{R}^n$ there exists an index $j \in \{1, ..., q\}$ such that $\nabla_x f_j(\overline{x})^\top d \ge 0$ holds.*

This concept is a generalization of the stationarity notion for scalar optimization problems. Consider such a scalar problem by setting $q = 1$ for (MOP) and let $\overline{x} \in \mathbb{R}^n$ be a Pareto critical point according to the above definition. Then it holds $\nabla_x f(\overline{x})^\top d \ge 0$ for all $d \in \mathbb{R}^n$. Hence it holds $\nabla_x f(\overline{x}) = 0_n$ and the standard stationarity notion for the scalar valued case is obtained.
The following lemma shows that Pareto criticality is a necessary condition for locally weak efficiency, see for example [20, 25].

Lemma 2.1 *If $\bar{x} \in \mathbb{R}^n$ is locally weak efficient for (MOP), then it is Pareto critical for (MOP).*

The following lemma gives a characterization of Pareto critical points and comes from multiobjective descent methods [14, 19, 20].

Lemma 2.2 *Let $f_i : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable functions for all $i = 1, ..., q$. For the function*

$$\omega(x) := -\min_{\|d\| \le 1} \max_{i=1,...,q} \nabla_x f_i(x)^\top d \tag{1}$$

the following statements hold.

(i) The mapping $x \mapsto \omega(x)$ is continuous.

(ii) It holds $\omega(x) \ge 0$ 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 (1) have some helpful properties.

Lemma 2.3 *Let $x \in \mathbb{R}^n$ be an arbitrary but fixed point and let d_ω denote a solution of the optimization problem stated in (1).*

(i) If x is not Pareto critical for (MOP) then d_ω is a descent direction for (MOP) at the point x, i.e. there exists a scalar $t_0 > 0$ such that it holds $f_i(x + t\, d_\omega) < f_i(x)$ for all $t \in (0, t_0]$ and for all $i \in \{1, ..., q\}$.

(ii) There exist scalars $\alpha_i \in [0, 1]$ for $i \in \{1, ..., q\}$ with $\sum_{i=1}^q \alpha_i = 1$ and $\mu \geq 0$ such that it holds $d_\omega = -\mu \sum_{i=1}^q \alpha_i \nabla_x f_i(x)$. If x is not Pareto critical for (MOP) it holds $\|d\| = 1$. If x is Pareto critical for (MOP) it holds $d_\omega = \sum_{i=1}^q \alpha_i \nabla_x f_i(x) = 0$. Furthermore it holds $\omega(x) \leq \|\sum_{i=1}^q \alpha_i \nabla_x f_i(x)\|$.

Proof. Statement (i) follows from the definition of Pareto criticality and descent directions. To prove statement (ii) reformulate (1) to

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

Let (t_ω, d_ω) denote a solution of (2) and firstly let x be not Pareto critical for (MOP). Then it follows from the KKT conditions, that there exist scalars $\alpha_i \in [0, 1]$ with $\sum_{i=1}^q \alpha_i = 1$ and $\mu \geq 0$ such that it holds

$$d_\omega = -\mu \sum_{i=1}^q \alpha_i \nabla_x f_i(x) \text{ with } \mu = \frac{1}{\|\sum_{i=1}^q \alpha_i \nabla_x f_i(x)\|} \text{ and } \|d\| = 1. \tag{3}$$

If x is Pareto critical, then the zero vector is a solution of (2) and the KKT conditions imply the existence of constants $\alpha_i \in [0, 1], i \in \{1, ..., q\}$, with $\sum_{i=1}^q \alpha_i = 1$ and $\sum_{i=1}^q \alpha_i \nabla_x f_i(x) = 0$.

Furthermore let (t_ω, d_ω) be a solution of (2). As it is an equivalent reformulation of (1) it holds $-t_\omega = \omega(x)$. This implies $\nabla_x f_i(x)^\top d_\omega \leq t_\omega$ for all $i \in \{1, ..., q\}$ and therefore

$$\omega(x) = -t_\omega = -\sum_{i=1}^q \alpha_i t_\omega \leq -\sum_{i=1}^q \alpha_i \nabla_x f_i(x)^\top d_\omega.$$

If x is not Pareto critical for (MOP), then (3) holds and it follows

$$\omega(x) \leq -\sum_{i=1}^q \alpha_i \nabla_x f_i(x)^\top d_\omega = \mu \left\| \sum_{i=1}^q \alpha_i \nabla_x f_i(x) \right\|^2 = \left\| \sum_{i=1}^q \alpha_i \nabla_x f_i(x) \right\|.$$

If x is Pareto critical for (MOP) it holds $\sum_{i=1}^q \alpha_i \nabla_x f_i(x) = 0$ and $\omega(x) = 0$ and the above inequality is also fulfilled. □

In the following we will use the inequality relations $<$ and $\leq$ for vectors in a componentwise manner. For $a, b \in \mathbb{R}^n$ we write $a \leq b$ if it holds $a_i \leq b_i$ for all $i \in \{1, ..., n\}$.

3 Algorithm description

The basic trust region concept [8, 10] is constructed for unconstrained scalar optimization problems with a twice continuously differentiable objective function bounded from below. It is an iterative method which approximates the function by suitable models in every iteration. These models are supposed to be easier than the original function and are used to compute a sufficient decrease. Furthermore the model and the computations are restricted to a local area in every iteration. This area is called trust region and is defined by

$$B_k := B\left(x^k, \delta_k\right) = \left\{x \in \mathbb{R}^n \mid \left\|x - x^k\right\| \leq \delta_k\right\} \tag{4}$$

using the current iteration point x^k, the so-called trust region radius $\delta_k > 0$ and the euclidean norm $\|\cdot\| := \|\cdot\|_2$. Further information about the choice of other norms can be found in [8]. Now consider a multiobjective optimization problem of the form of (MOP) with f_1 being an expensive, simulation-given function. The multiobjective method presented in this paper is an iterative approach as well and in every iteration $k \in \mathbb{N}$ each objective function f_i with $i \in \{1, ..., q\}$ is replaced by a suitable quadratic model $m_i^k : \mathbb{R}^n \to \mathbb{R}$ which satisfies the interpolation condition

$$f_i(x^k) = m_i^k(x^k), \tag{5}$$

see subsection 3.1 for detailed information. As a surrogate for (MOP) the problem

$$\min_{x \in \mathbb{R}^n} m^k(x) \tag{$MOPm$}$$

is considered in every iteration k. Furthermore the computations are restricted to a local area, the trust region B_k as defined in (4). The search for a sufficient decrease in the function values is realized by computing the ideal point $p^k = (p_1^k, ..., p_q^k)^\top$ defined by $p_i^k = \min_{x \in B_k} m_i^k(x)$ for all $i = 1, ..., q$. These subproblems need to be solved in every iteration. However, they are only quadratic problems with simple constraints and therefore any quadratic solver can be used. Also a trust-region approach is possible, see for example [4] for solving trust region subproblems. The ideal point p^k gives a direction for decreasing the model functions and, depending on the quality of the approximations, 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. The trust region functions not only as a guarantee that the models are good enough approximations, but also as a step size control. Moving towards the ideal point is realized by the Pascoletti-Serafini scalarization [28] given by

$$\begin{array}{ll} \min & t \\ \text{s.t} & f(x^k) + t\, r^k - m^k(x) \in \mathbb{R}^q_+ \\ & t \in \mathbb{R} \\ & x \in B_k \end{array} \tag{PS}$$

with $r^k := f(x^k) - p^k \in \mathbb{R}^q_+$, p^k the ideal point of m^k in B_k and $m^k = (m_1^k, ..., m_q^k)^\top$ the model functions. This scalarization is also known as Tammer-Weidner functional [22]. Note that it holds $f(x^k) = m^k(x^k)$ in every iteration k due to the interpolation conditions (5). The problem (PS) minimizes, in case $r^k \in int\, \mathbb{R}^q_+$, the weighted Chebyshev distance between the set $m^k(B_k)$ and the point $f(x^k)$ with weights $w_i = 1/r_i^k$ for $i \in \{1, ..., q\}$. Solving (PS) we obtain the trial point x^{k+}, a candidate for the next iteration point. Figure 1 illustrates the idea in the biobjective case with $q = 2$ and $\left(t^{k+}, x^{k+}\right)$ being the solution of (PS). The image of the trial point x^{k+} is marked black.

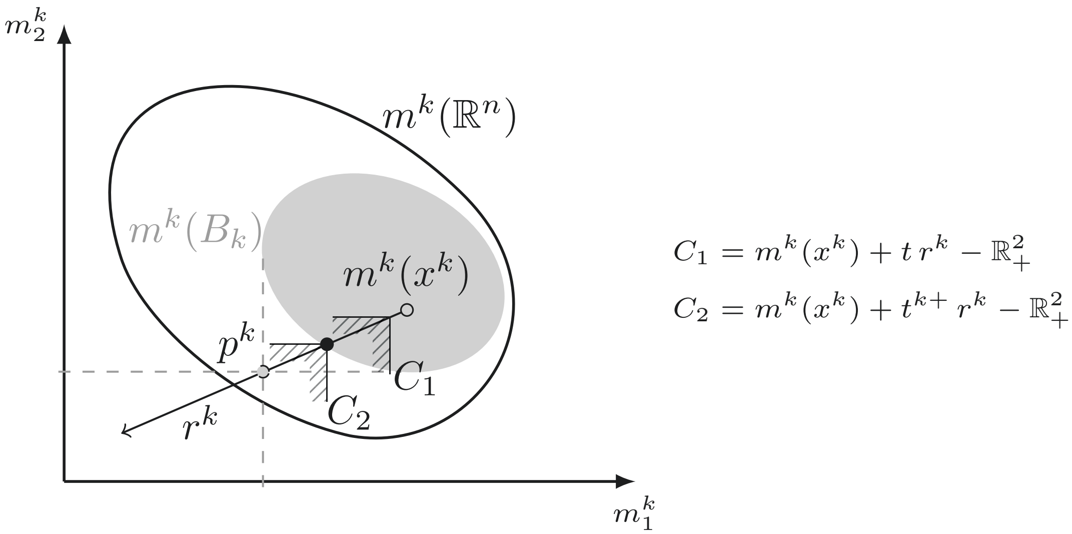

Figure 1: Pascoletti-Serafini scalarization (PS)

Analogously to the scalar trust region method [8, 10] the trial point x^{k+} is only accepted as next iteration point if a condition describing the improvement of the function values is met. We use the same approach as [36] defining the functions

$$\phi(x) \coloneqq \max_{i=1,\dots,q} f_i(x) \text{ and } \phi_m^k(x) \coloneqq \max_{i=1,\dots,q} m_i^k(x) \tag{6}$$

to examine if

$$\rho_\phi^k \coloneqq \frac{\phi(x^k) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})} \tag{7}$$

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 multiobjective condition for the trial point acceptance test can be found in subsection 3.3.

The trust region algorithm for heterogeneous multiobjective problems TRAHM is formulated in Algorithm 1. It describes a new trust region approach which differs from the previously known methods by the computation of the search direction. In TRAHM the direction is determined in the image space by using the local ideal points of the model functions. As input a starting point, some parameters and the objective functions are needed, whereby f_1 is expensive and f_i are cheap for all $i \in \{2, ..., q\}$. Hence also the used model functions differ, which is explained in detail in subsection 3.1.

Algorithm 1 TRAHM

Input: functions f_i, $i = 1, ..., q$, initial point x_0, 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

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

Step 1: Ideal Point

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

Step 2: Trial Point

Compute $(t^{k+}, x^{k+})^\top$ by solving (PS)

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

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

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

Step 4: Trust Region Update

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

Step 5: Model Update

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

The choice of the parameters η_1, η_2, γ_1 and γ_2 can of course be problem-dependent, but according to [8] reasonable values are $\eta_1 = 0.01, \eta_2 = 0.9$ and $\gamma_1 = \gamma_2 = \frac{1}{2}$.

3.1 Model functions

In basic trust region methods quadratic models are most commonly used to replace the original functions. The subproblem of minimizing the model function can then be solved by quadratic methods. Hence in our algorithm we also replace the functions by quadratic models, even the cheap functions which are analytically available. A quadratic model $m : \mathbb{R}^n \to \mathbb{R}$ for a function $g : \mathbb{R}^n \to \mathbb{R}$ is given by

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

with $m(y) = g(y)$ for a fixed point $y \in \mathbb{R}^n$ and H a symmetric approximation to $\nabla_{xx} g(y)$. This is only possible if the function is twice continuously differentiable and the derivative information is available. Since this is the case for the cheap functions $f_i, i = 2, ..., q$, in our context, we use the so-called Taylor model $m_i^k(x) = m_T(x; f_i, x^k)$. It is a quadratic model defined by

$$m_T(x; f_i, x^k) := f_i(x^k) + \nabla_x f_i(x^k)^\top \left(x - x^k\right) + \frac{1}{2} \left(x - x^k\right)^\top \nabla_{xx} f_i(x^k) \left(x - x^k\right) \quad (8)$$

in every iteration $k \in \mathbb{N}$ using the current iteration point x^k $(i = 2, ..., q)$. For such models it always holds $\nabla_x m_i^k(x^k) = \nabla_x f_i(x^k)$. However, this kind of model cannot be used for the expensive function due to the high computational effort this would entail. To obtain a quadratic model as well we use interpolation based on quadratic Lagrangian polynomials. To build such a model $m_1 : \mathbb{R}^n \to \mathbb{R}$ for the expensive function f_1 let $\mathcal{P}_n^2$ denote the space of polynomials of degree less than or equal to two in $\mathbb{R}^n$. It is known that the dimension p of this space is given by $p = (n+1)(n+2)/2$. Given a basis $\psi = \{\psi_1, ..., \psi_p\}$ of $\mathcal{P}_n^2$, every polynomial $g \in \mathcal{P}_n^2$ is defined as $g(x) = \sum_{i=1}^p \alpha_i \psi_i(x)$ with $\alpha \in \mathbb{R}^p$ some suitable coefficients. For the interpolation of the expensive function f_1 let $Y = \{y^1, y^2, ..., y^p\} \subset \mathbb{R}^n$ be a set of interpolation points for which the interpolation conditions

$$m_1(y^i) = f_1(y^i)$$

are required to hold true for all $i = 1, ..., q$. For the basis ψ we choose the basis of quadratic Lagrange polynomials $l_i \in \mathcal{P}_n^2, i = 1, ..., p$, defined by

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

Hence the expensive function f_1 is replaced in every iteration $k \in \mathbb{N}$ by the model

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

with a set of interpolation points $Y_k = \{y^1, y^2, ..., y^p\} \subset B_k$ from the current trust region and $x^k \in Y_k$. The interpolation points are not randomly chosen from the trust region but are computed such that they satisfy a quality criterion called well poisedness. This concept will not be explained here but can be found in detail in [10]. Since Lagrange polynomials are not only compatible with this concept, but most commonly used for measuring well poisedness, they are chosen as an interpolation basis here.
Another option for building models in the trust region scheme are radial basis functions (RBFs). This is described for scalar trust region methods in [37].

3.2 Computing the trial point

For computing the trial point x^{k+} in step 2 of TRAHM the auxiliary optimization problem (PS) is used given by

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

Due to the interpolation conditions it holds $f(x^k) = m^k(x^k)$ in every iteration $k \in \mathbb{N}$.

Remark 3.1 *Let x^k be not Pareto-critical for $(MOPm)$. According to Lemma 2.1 x^k is not locally weakly efficient for $(MOPm)$ and, as $x^k \in \operatorname{int} B_k$, also not weakly efficient for $\min_{x \in B_k} m^k(x)$. Thus, x^k cannot be an individiual minimum of one of the functions m_i^k, $i \in \{1, ..., q\}$, on B_k, hence for the direction r^k of (PS) it holds $r_i^k = m_i^k(x^k) - \min_{x \in B_k} m_i^k(x) > 0$ for all $i \in \{1, ..., q\}$.*

The optimization problem (PS) has some useful properties, which can be found in detail and with proof in [15, Th. 2.1].

Lemma 3.1

(i) If $(\bar{t},\bar{x})$ is a minimal solution of (PS) then $\bar{x}$ is weakly efficient for $\min_{x\in B_k} m^k(x)$.

(ii) If $(\bar{t},\bar{x})$ is a local minimal solution of (PS) then $\bar{x}$ is locally weakly efficient for $\min_{x\in B_k} m^k(x)$.

(iii) If $\bar{x}$ is a weakly efficient solution for $\min_{x\in B_k} m^k(x)$ and $r^k \in \operatorname{int}\mathbb{R}^q_+$, then $(0,\bar{x})$ is a minimal solution of (PS).

Another property of (PS) is stated in the following lemma.

Lemma 3.2 *Let x^k be not weakly efficient for $\min_{x\in B_k} m^k(x)$. For every minimal solution $(\bar{t},\bar{x})$ of (PS) it holds $\bar{t}\in[-1,0)$.*

Proof. Let $(\bar{t},\bar{x})$ be a minimal solution of (PS). Since $(0,x^k)$ is always feasible for (PS), it holds $\bar{t}\le 0$. 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 $m^k(\tilde{x}) < m^k(x^k)$. This also implies $r^k = m^k(x^k) - \min_{x\in B_k} m^k(x) > 0$. Then there exists a scalar $t>0$ with $m^k(x^k) - t\,r^k - m^k(\tilde{x}) > 0$. Hence $(-t,\tilde{x})$ is feasible for (PS) and it holds $\bar{t}<0$.
Now suppose $\bar{t} := -1-s < -1$ with $s>0$. Resulting from the constraints of (PS) it holds $p^k - m^k(\bar{x}) \ge s\,r^k$. Again due to x^k being not weakly efficient and thus $r^k>0$ it follows $p^k > m^k(\bar{x})$ which contradicts the definition of p^k. Consequently, it holds $\bar{t}\in[-1,0)$. □

3.3 Trial point acceptance test

Step 3 of TRAHM is the trial point acceptance test which uses the quotient $\rho^k_\phi = (\phi(x^k) - \phi(x^{k+}))/(\phi^k_m(x^k) - \phi^k_m(x^{k+}))$ with the functions $\phi(x) = \max_{i=1,\dots,q} f_i(x)$ and $\phi^k_m(x) = \max_{i=1,\dots,q} m^k_i(x)$ from (6). Due to the determining of x^{k+} it always holds $\phi^k_m(x^k) - \phi^k_m(x^{k+}) \ge 0$. Furthermore, as long as x^k is not weakly efficient for $\min_{x\in B_k} m^k(x)$ there exists a point $\tilde{x}\in B_k$ with $m^k(\tilde{x}) < m^k(x^k)$, see also the reasoning in the proof of Lemma 3.2. Together with the definition of the trial point it follows $\phi^k_m(x^k) - \phi^k_m(x^{k+}) > 0$ as long as x^k is not weakly efficient.
Supposed it holds $\rho^k_\phi > 0$ which implies $\phi(x^k) - \phi(x^{k+}) > 0$. Then there exist indices $i,j\in\{1,\dots,q\}$ such that $0 < f_i(x^k) - f_j(x^{k+}) \le f_i(x^k) - f_i(x^{k+})$ holds. Therefore the trial point x^{k+} guarantees a descent in at least one component of f. In TRAHM x^{k+} is accepted if ρ^k_ϕ is bigger than a strictly positive constant η_1 to assure not only a decrease in at least one component but to guarantee that this decrease is "sufficient".
In the case $\rho^k_\phi < 0$ there exist indices $i,j\in\{1,\dots,q\}$ with $0 > f_i(x^k) - f_j(x^{k+}) \ge f_j(x^k) - f_j(x^{k+})$. This implies an increase in at least one component of f. Hence the trial point is not accepted as next iteration point.
Now assume $\rho^k_\phi = 0$. This implies $t^{k+} = 0$, $\phi^k_m(x^k) - \phi^k_m(x^{k+}) = 0$ or $\phi(x^k) - \phi(x^{k+}) = 0$. If it holds $t^{k+} = 0$, then according to Lemma 3.1 (i) x^k is a weakly efficient point for $\min_{x\in B_k} m^k(x)$. If the model is a good approximation to the original function, x^k is a locally weak efficient point for (MOP). By setting $\rho^k_\phi = 0$ in this case the trust region radius will be reduced and the model will be updated to affirm the model information. If the model was reliable the trust region will also shrink in the next iterations and therefore the

radius will converge to zero. If the model was not reliable then there will be a subsequent iteration in which the trial point produces a sufficient decrease.
If it holds $\phi_m^k(x^k) - \phi_m^k(x^{k+}) = 0$ there exist indices $i, j \in \{1, ..., q\}$ fulfilling $m_j^k(x^k) \leq m_i^k(x^k) = m_j^k(x^{k+}) \geq m_i^k(x^{k+})$, so either there is no decrease in at least one component or the points x^k and x^{k+} are incomparable. In this case the trial point is rejected and the trust region radius is reduced. The same line of argument, but for the original functions, applies if $\phi(x^k) - \phi(x^{k+}) = 0$ holds.
For the convergence analysis in section 4 some assumptions are needed and will be explained there in detail. We want to anticipate Assumption 4.6 here because it clarifies the trial point acceptance test. This assumption ensures a sufficient decrease in every iteration of the form of

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

with $\omega(x)$ from (1), $\kappa_\phi \in (0, 1)$ and $\beta_k^\phi > 0$. Due to Lemma 2.2 it holds $\omega(x) = 0$ if and only if the point x is Pareto critical for (MOP) and according to Lemma 2.1 Pareto criticality is a necessary condition for local weak efficiency. If it holds $\phi_m^k(x^k) - \phi_m^k(x^{k+}) = 0$ this bound implies $\omega(x^k) = 0$. This gives another reason for setting ρ_ϕ^k equal to zero if $\phi_m^k(x^k) - \phi_m^k(x^{k+}) = 0$ holds.

4 Convergence

In the following a convegence proof for TRAHM to a Pareto critical point of the optimization problem (MOP) is presented and for these results some assumptions on the original and the model functions are needed. All these assumptions are connected to the commonly used assumptions in the scalar trust region and derivative-free optimization context [8, 10, 34] or in multiobjective trust region methods [29, 36]. As stated within the problem description in section 2, the functions f_i are assumed to be twice continuously differentiable for all $i \in \{1, ..., q\}$ and $\phi(x) = \max_{i=1,...,q} f_i(x)$ is assumed to be bounded from below. Furthermore, for every index $i \in \{1, ..., q\}$ and for every iteration $k \in \mathbb{N}$ the model functions m_i^k are assumed to be quadratic and twice continuously differentiable functions. The model is assumed to be exact in the current iteration point x^k, that is it holds

$$m^k(x^k) = f(x^k) \tag{9}$$

in every iteration $k \in \mathbb{N}$. This holds true for every interpolation model which uses x^k as interpolation point and also for the model functions presented in subsection 3.1. For the cheap functions also the gradients shall coincide in the current iteration point, that is it holds

$$\nabla_x m_i^k(x^k) = \nabla_x f_i(x^k) \tag{10}$$

for all $i \in \{2, ..., q\}$ and for all $k \in \mathbb{N}$. This is fulfilled for the Taylor model, which is used for the cheap functions as explained in subsection 3.1. These general assumptions will be used throughout the convergence analysis in this section. In addition to these basic assumptions some further assumptions are necessary. Besides, a matrix norm compatible with the used vector norm is necessary. Since we use the Euclidean norm, we consider the Frobenius norm as matrix norm.

Assumption 4.1 *For every index $i \in \{1, ..., q\}$ the Hessian of the function f_i is uniformly bounded, that is there exists a constant $\kappa_{\mathrm{uh}f_i} > 1$ fulfilling*

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

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

Remark 4.1 *Assumption 4.1 together with the mean value theorem implies that the functions $\nabla_x f_i : \mathbb{R}^n \to \mathbb{R}^n$ are Lipschitz continuous for all $i = 1, ..., q$. It follows that the function ω defined in (1) is uniformly continuous, see also [36].*

Assumption 4.2 *For every index $i \in \{1, ..., q\}$ the Hessian of the model function m_i^k is uniformly bounded for all iterations $k \in \mathbb{N}$, that is there exists a constant $\kappa_{\mathrm{uh}m_i} > 1$ independent of k fulfilling*

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

for all $x \in B_k$. The index 'uhm_i' stands for upper bound on the Hessian of m_i.

Furthermore as in every model-based solution method it is important to assure a good local accuracy of the model functions in every iteration. For this purpose we use the common notion of validity which can be found for example in [8].

Definition 4.1 *Let $i \in \{1, ..., q\}$ and $k \in \mathbb{N}$ be indices. A model function $m_i^k : \mathbb{R}^n \to \mathbb{R}$ is called valid for the function $f_i : \mathbb{R}^n \to \mathbb{R}$ in the trust region $B_k = \left\{x \in \mathbb{R}^n \mid \left\|x - x^k\right\| \leq \delta_k\right\}$, if there exists a constant $\kappa_{\mathrm{cnd}i} > 0$ such that*

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

holds for all $x \in B_k$. The index 'cnd' stands for conditional error.

Generally, in the trust region approach validity is assumed for the models. In our context we can even prove this for the models of the cheap functions.

Lemma 4.1 *Suppose Assumptions 4.1 and 4.2 hold. In every iteration $k \in \mathbb{N}$ the model m_i^k is valid for f_i in B_k for all $i \in \{2, ..., q\}$, that is it holds*

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

for all $x \in B_k$ and $\kappa_{\mathrm{cnd}i} := \max\left\{\kappa_{\mathrm{uh}f_i}, \kappa_{\mathrm{uh}m_i}\right\} - 1 > 0$.

Proof. Due to the functions f_i being twice continuously differentiable it follows from Taylor's theorem for every $h \in \mathbb{R}^n$ with $\|h\| \leq \delta_k$,

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

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

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

for every $h \in \mathbb{R}^n$ with $\|h\| \leq \delta_k$ and for all indices $i \in \{2, ..., q\}$. Moreover it holds $\nabla_x m_i^k(x^k) = \nabla_x f_i(x^k)$ for all $i \in \{2, ..., q\}$ due to (10) which is given for the Taylor model (8). Using the triangle inequality it follows for every $x \in B_k$

$$\left|f_i(x) - m_i^k(x)\right| \leq \frac{1}{2} \|h\|^2 \left(\left\|\nabla_{xx} f_i(\xi_i^k)\right\| + \left\|\nabla_{xx} m_i^k(x^k)\right\|\right) \leq \delta_k^2 \left(\max\left\{\kappa_{\mathrm{uh}f_i}, \kappa_{\mathrm{uh}m_i}\right\} - 1\right)$$

with the constants $\kappa_{\mathrm{uh}f_i}$ and $\kappa_{\mathrm{uh}m_i}$ from Assumptions 4.1 and 4.2. Then the statement of the lemma holds for $\kappa_{\mathrm{cnd}i} := \max\left\{\kappa_{\mathrm{uh}f_i}, \kappa_{\mathrm{uh}m_i}\right\} - 1 > 0$. □

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

Assumption 4.3 *In every iteration $k \in \mathbb{N}$ the model m_1^k is valid for the function f_1 in B_k, that is there exists a constant $\kappa_{\mathrm{cnd}1} > 0$ independent of k such that it holds for all $x \in B_k$*

$$\left|f_1(x) - m_1^k(x)\right| \leq \kappa_{\mathrm{cnd}1} \delta_k^2.$$

The accuracy of the model is also reflected in the gradients. For the cheap functions $m_i^k, i \in \{2, ..., q\}$, the equality $\nabla_x m_i^k(x^k) = \nabla_x f_i(x^k)$ is required for all iterations $k \in \mathbb{N}$, see (10). This is fulfilled in our context as we use the Taylor model (8). 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 [8]. Due to the problem-dependent constants we give a short proof.

Lemma 4.2 *Suppose Assumptions 4.1, 4.2 and 4.3 hold. Then there exists a constant $\kappa_{\mathrm{eg}} > 0$ such that it holds*

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

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

Proof. Analogous to Lemma 4.1 and similar to [8, Th. 9.1.1] it follows by using Taylor's theorem, (9) and the triangle inequality

$$\begin{aligned}\left|\left(\nabla_x f_1(x^k) - \nabla_x m_1^k(x^k)\right)^\top h\right| &\leq \left|f_1(x) - m_1^k(x)\right| + \frac{1}{2}\|h\|^2 \left\|\nabla_{xx} f_1(\xi^k) - \nabla_{xx} m_1^k(x^k)\right\| \\ &\leq \kappa_{\mathrm{cnd}1}\delta_k^2 + \max\left\{\kappa_{\mathrm{uh}f_1} - 1, \kappa_{\mathrm{uh}m_1} - 1\right\}\delta_k^2\end{aligned}$$

for every $h \in \mathbb{R}^n$ with $\|h\| \leq \delta_k$ and $x := x^k + h \in B_k$. It holds $\xi_i^k \in \left[x_i^k, x_i^k + h\right]$ for $i \in \{1, ..., n\}$ and the constants $\kappa_{\mathrm{uh}f_1}$, $\kappa_{\mathrm{uh}m_1}$ and $\kappa_{\mathrm{cnd}1}$ are from Assumptions 4.1, 4.2 and 4.3. Setting $h := \delta_k \frac{\nabla_x f_1(x^k) - \nabla_x m_1^k(x^k)}{\left\|\nabla_x f_1(x^k) - \nabla_x m_1^k(x^k)\right\|}$ the statement of the Lemma follows with the constant $\kappa_{\mathrm{eg}} := \kappa_{\mathrm{cnd}1} + \max\left\{\kappa_{\mathrm{uh}f_1}, \kappa_{\mathrm{uh}m_1}\right\} - 1 > 0$. □

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

$$\omega_m(x) \coloneqq -\min_{\|d\|\leq 1} \max_{i=1,\dots,q} \nabla_x m_i^k(x)^\top d \tag{11}$$

for the model functions.

Assumption 4.4 *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 ($MOPm$) or close to such a point, this is also satisfied for the original optimization problem (MOP). The convergence proof in this section is based on the characterization of Pareto critical points by the function ω. It will be proved that TRAHM produces a sequence of iterates with ω converging to zero. For this purpose, a sufficient decrease condition for the iteration points is necessary. Such a sufficient decrease condition is commonly used in trust region approaches, both in scalar and multiobjective versions [8, 10, 29, 36]. It is based on the idea of minimizing along a descent direction, either for the individual functions or in the multiobjective way given by the function ω.

In the scalar approach [8, 10] 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 multiobjective case by using the function ω, is used in [36]. In [29] 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 paper 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 literature. Still we can use the strategy of comparing the trial point to the result of minimizing along a multiobjective descent direction. For this purpose an assumption regarding the optimization problem (PS) given by $\min\left\{t \in \mathbb{R} \mid m^k(x^k) + t\, r^k - m^k(x) \in \mathbb{R}^q_+, x \in B_k\right\}$ is necessary which is prepared by the following lemma.

Lemma 4.3 *Suppose Assumption 4.2 holds. 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 $i = 1, \dots, q$. In every iteration $k \in \mathbb{N}$ with x^k being not Pareto critical for ($MOPm$) it holds for every $i \in \{1, \dots, q\}$*

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

with $\beta_i^k \coloneqq 1 + \left\|\nabla_{xx} m_i^k(x^k)\right\|$ *and* $\kappa_{\mathrm{uhm}_i} > 1$ *from* Assumption 4.2.

Proof. Let $i \in \{1, ..., q\}$ denote an index and $k \in \mathbb{N}$ an iteration with x^k being not Pareto critical for (*MOPm*). By Lemma 2.1 it follows $\nabla_x m_i^k(x^k) \neq 0$. Consider the normed steepest descent direction for m_i^k in x^k defined by $d_{sdi} \coloneqq -\left(\nabla_x m_i^k(x^k)\right) / \left(\left\|\nabla_x m_i^k(x^k)\right\|\right)$. From Taylor's theorem and the Cauchy Schwarz inequality it follows

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

with $\beta_i^k = 1 + \left\|\nabla_{xx} m_i^k(x^k)\right\|$. The possible candidates for the solution of the above maximization problem are $t_1 = \left\|\nabla_x m_i^k(x^k)\right\| / \beta_i^k$ and $t_2 = \delta_k$ if $t_1 > \delta_k$. By calculating the function values for these candidates it follows

$$r_i^k > \min \left\{ \frac{1}{2} \frac{\left\|\nabla_x m_i^k(x^k)\right\|^2}{\beta_i^k}, \delta_k \left\|\nabla_x m_i^k(x^k)\right\| - \frac{1}{2} \delta_k^2 \beta_i^k \right\}. \tag{12}$$

The second term is obtained if it holds $\delta_k < t_1$. Thus, by estimating it the lower bound of the lemma follows by

$$r_i^k > \min \left\{ \frac{1}{2} \frac{\left\|\nabla_x m_i^k(x^k)\right\|^2}{\beta_i^k}, \frac{1}{2} \left\|\nabla_x m_i^k(x^k)\right\| \delta_k \right\}. \tag{13}$$

For the upper bound let $\min_{x \in B_k} m_i^k(x) = m_i^k(\tilde{x})$ with $\tilde{x} \coloneqq x^k + t\, d$, $|t| \leq \delta_k$ and $\|d\| = 1$. From Taylor's theorem and the Cauchy Schwarz inequality it follows

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

This implies with Assumption 4.2

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

for every $i \in \{1, ..., q\}$. □

As stated in Remark 3.1 it holds $r^k > 0$ as long as x^k is not Pareto critical for (*MOPm*). Then according to the lemma above the following assumption on the search direction r^k is reasonable which means that r^k is neither too flat nor too steep.

Assumption 4.5 *There exists a constant $\kappa_r \in (0,1]$ such that it holds for every iteration $k \in \mathbb{N}$ with x^k being not Pareto critical for (MOPm)*

$$\frac{\min\limits_{i=1,\dots,q} r_i^k}{\max\limits_{i=1,\dots,q} r_i^k} \geq \kappa_r. \tag{14}$$

To formulate a sufficient decrease condition for the iterates of TRAHM consider

$$d_\omega \in \underset{\|d\|\leq 1}{\operatorname{argmin}} \max_{i=1,\dots,q} \nabla_x m_i^k(x^k)^\top d \tag{15}$$

a solution of (11). If x^k is not a Pareto critical point for (MOPm), then according to Lemma 2.3 applied to (11) d_ω is a descent direction for the multiobjective problem (MOPm) at the current iteration point x^k. Therefore it will provide a descent also in the trust region B_k. Furthermore there exist scalars $\alpha_i \in [0,1], i \in \{1,\dots,q\}$, with $\sum_{i=1}^q \alpha_i = 1$ and $\mu \geq 0$ such that

$$d_\omega = -\mu \sum_{i=1}^q \alpha_i \nabla_x m_i^k(x^k) \tag{16}$$

holds with $\|d_\omega\| = 1$. Now 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 4.4 *Let $k \in \mathbb{N}$ be an iteration with x^k not being Pareto critical for (MOPm). 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 $\alpha_i \geq 0, i \in \{1,\dots,q\}$ from (16). Furthermore define x_c by $g(x_c) := \min_{|t|\leq\delta_k} g(x^k + t\,d)$ with $d := -\nabla_x g(x^k)/\left\|\nabla_x g(x^k)\right\|$ and set $\beta_g^k := 1 + \left\|\nabla_{xx} g(x^k)\right\|$. Then it holds*

$$g(x^k) - g(x_c) \geq \frac{1}{2}\left\|\nabla_x g(x^k)\right\| \min\left\{\frac{\left\|\nabla_x g(x^k)\right\|}{\beta_g^k}, \delta_k\right\}. \tag{17}$$

Proof. The normed steepest descent direction for g at x^k is given by $d_\omega = -\nabla_x g(x^k)/\left\|\nabla_x g(x^k)\right\|$ defined in (16). Since all model functions are quadratic it follows from Taylor's theorem

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

for every $t \in \mathbb{R}$. Define $\beta_g^k := \left\|\nabla_{xx} g(x^k)\right\| + 1 > 0$. The Cauchy Schwarz inequality implies together with calculations and estimations analogous to (12) and (13) in the proof of Lemma 4.3

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

which gives the inequality of the lemma. □

Remark 4.2 *If x^k is Pareto critical for (MOPm) no steepest descent for the function g in Lemma 4.4 exists. In this case we set $x_c = x^k$ and due to $\nabla_x g(x^k) = 0$ the inequality (17) still holds.*

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

Lemma 4.5 *Suppose Assumptions 4.2 , 4.4 and 4.5 hold. Let x^{k+} be the solution of (PS) and let $\phi_m^k(x) = \max_{i=1,...,q} m_i^k(x)$ be defined as in (6). Furthermore define $\beta_\phi^k := \max_{i=1,...,q} \left\|\nabla_{xx} m_i^k(x^k)\right\| + 1$. 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*

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq \left(\frac{1}{2}\right)^j \tilde{\kappa}_\phi \omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta_\phi^k}, \delta_k\right\}. \tag{18}$$

Proof. Let $(t^{k+}, x^{k+}) \in \mathbb{R}^{1+n}$ be the solution of the auxiliary problem (PS) given by $\min\left\{t \in \mathbb{R} \mid f(x^k) + t\, r^k - m^k(x) \in \mathbb{R}_+^q, x \in B_k\right\}$. Firstly, let x^k be not Pareto critical for (MOPm). Then according to Lemma 3.2 and Remark 3.1 it holds $t^{k+} \in [-1, 0)$ and $r^k > 0$ defined by $r_i^k = m_i^k(x^k) - \min_{x \in B_k} m_i^k(x)$ for $i \in \{1, ..., q\}$. Due to the constraints of (PS) it holds

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

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

$$-t^{k+} = |t^{k+}| \leq \frac{m_i^k(x^k) - m_i^k(x^{k+})}{r_i^k} \leq \frac{\phi_m^k(x^k) - m_i^k(x^{k+})}{\min\limits_{j=1,...,q} r_j^k} \tag{19}$$

for all $i \in \{1, ..., q\}$. Let $d_\omega \in \operatorname{argmin}_{\|d\| \leq 1} \max_{i=1,...,q} \nabla_x m_i^k(x^k)^\top d$ be a solution of the optimization problem from (11). Then according to Lemma 2.3(ii) applied to (11) there exist scalars $\alpha_i \in [0,1], i \in \{1, ..., q\}$, with $\sum_{i=1}^q \alpha_i = 1$ and $\mu \geq 0$ such that $\|d_\omega\| = 1$ and (16) holds, that is $d_\omega = -\mu \sum_{i=1}^q \alpha_i \nabla_x m_i^k(x^k)$. For the resulting function $g(x) = \sum_{i=1}^q \alpha_i m_i^k(x)$ and the corresponding point $x_c = x^k + \tau\, d_\omega$ with $|\tau| \leq \delta_k$ Lemma 4.4 and therefore (17) holds. Furthermore it holds for β_g^k from Lemma 4.4

$$\beta_g^k = \left\|\nabla_{xx} g(x^k)\right\| + 1 \leq \sum_{i=1}^q \alpha_i \left\|\nabla_{xx} m_i^k(x^k)\right\| + 1 \leq \max_{i=1,...,q} \left\|\nabla_{xx} m_i^k(x^k)\right\| + 1 = \beta_\phi^k$$

which implies with (17) from Lemma 4.4

$$g(x^k) - g(x_c) \geq \frac{1}{2} \left\|\nabla_x g(x^k)\right\| \min\left\{\frac{\left\|\nabla_x g(x^k)\right\|}{\beta_\phi^k}, \delta_k\right\}. \tag{20}$$

Due to $x_c \in B_k$ and d_ω being a descent direction for (MOPm), see Lemma 2.3(i) for (11), there exists a scalar t such that (t, x_c) is feasible for (PS). According to [17] there exists a smallest scalar t_c such that (t_c, x_c) is feasible for (PS) and it follows

$$-t_c = |t_c| = \min_{i=1,...,q} \frac{m_i^k(x^k) - m_i^k(x_c)}{r_i^k} \geq \frac{\min\limits_{i=1,...,q} \left(m_i^k(x^k) - m_i^k(x_c)\right)}{\max\limits_{i=1,...,q} r_i^k}. \tag{21}$$

Due to t^{k+} being the minimal value of (PS) it holds $|t_c| \leq |t^{k+}|$ which implies together with (19) for the index i with $m_i^k(x^{k+}) = \phi_m^k(x^{k+})$, (21) and Assumption 4.5

$$\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq \kappa_r \min_{i=1,\dots,q} \left(m_i^k(x^k) - m_i^k(x_c)\right). \tag{22}$$

Since it holds $\sum_{i=1}^q \alpha_i = 1$ and (t_c, x_c) is feasible for (PS) it follows for the function g defined in Lemma 4.4

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

This inequality together with (20) implies the existence of an index $j \in \mathbb{N}$ such that

$$\min_{i=1,\dots,q} \left(m_i^k(x^k) - m_i^k(x_c)\right) \geq \left(\frac{1}{2}\right)^j \left\|\nabla_x g(x^k)\right\| \min\left\{\frac{\left\|\nabla_x g(x^k)\right\|}{\beta_\phi^k}, \delta_k\right\} \tag{23}$$

holds and therefore it follows from (22) and the definition of g

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

for every iteration $k \in \mathbb{N}$ with x^k being not Pareto critical. If x^k is Pareto critical for $(MOPm)$, then it holds $\omega_m(x^k) = 0$ and the solution of (11) is $d_\omega = 0$. Therefore it holds $\sum_{i=1}^q \alpha_i \nabla_x m_i^k(x^k) = 0$, see Lemma 2.3(ii). Due to x^{k+} being the solution of (PS) it holds $\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq 0$ and the above inequality is also satisfied.
Furthermore, it holds according to Lemma 2.3(ii) $\omega_m(x^k) \leq \left\|\sum_{i=1}^q \alpha_i \nabla_x m_i^k(x^k)\right\|$ and from Assumption 4.4 it follows

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

with $1/(1+\kappa_\omega) \in (0,1)$. Then it holds for every iteration $k \in \mathbb{N}$

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

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

This Lemma gives a decrease condition for the trial point x^{k+} obtained by TRAHM 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 as long as x^k is not Pareto critical for (MOP) and therefore ensures a decrease in this case. Thus, the following assumption is reasonable to ensure a sufficient decrease in every iteration.

Assumption 4.6 *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 $\beta_\phi^k = \max_{i=1,\dots,q} \left\|\nabla_{xx} m_i^k(x^k)\right\| + 1$.

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. [8, 36], such an assumption is used and following this general approach we proved as well a motivation for the sufficient decrease assumption. Provided Assumption 4.6, the remaining of the convergence analysis of TRAHM follows the scalar trust region methods [8, 10] closely. Consequently it is also similar to the convergence analysis of the multiobjective trust region method in [36], which is based on the scalar 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 TRAHM.

Remark 4.3 *Due to Assumption 4.2 it holds in every iteration* $k \in \mathbb{N}$ *for the constant* β_k^ϕ *from Assumption 4.6*

$$\beta_k^\phi = \max_{i=1,\dots q} \left\|\nabla_{xx} m_i^k(x^k)\right\| + 1 \leq \max_{i=1,\dots q} \kappa_{\mathrm{uhm}_i}.$$

Lemma 4.6 *Suppose Assumptions 4.1, 4.2 and 4.3 hold, then it holds*

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

in every iteration $k \in \mathbb{N}$ *with* $\kappa_{cnd} := \max_{i=1,\dots,q} \kappa_{\mathrm{cnd}i} > 0$ *and the corresponding constants from Lemma 4.1 and Assumption 4.3.*

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 $i,j \in \{1,\dots,q\}$ and $i \neq j$. In case (i) it follows $\left|\phi(x^{k+}) - \phi_m^k(x^{k+})\right| \leq \kappa_{\mathrm{cnd}i}\delta_k^2$ due to $x^{k+} \in B_k$, Lemma 4.1 and Assumption 4.3. Now consider case (ii) and assume $f_i(x^{k+}) - m_j^k(x^{k+}) > 0$. Due to the definition of ϕ, Lemma 4.1, Assumption 4.3 and $x^{k+} \in B_k$ it holds $\left|\phi(x^{k+}) - \phi_m^k(x^{k+})\right| \leq \left|f_i(x^{k+}) - m_i^k(x^{k+})\right| \leq \kappa_{\mathrm{cnd}i}\delta_k^2$. Next assume $f_i(x^{k+}) - m_j^k(x^{k+}) < 0$. Then it holds again according to the definition of ϕ, Lemma 4.1, Assumption 4.3 and $x^{k+} \in B_k$

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

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

In the following every point x^{k+1} is given by TRAHM as a result of iteration $k \in \mathbb{N}$. Either the trial point is accepted and it holds $x^{k+1} = x^{k+}$ or it is discarded and $x^{k+1} = x^k$.

For the further considerations the iterations of TRAHM are classified according to their outcome using the constants $0 < \eta_1 \leq \eta_2 < 1$ from the description of the algorithm in section 3. An iteration is called successful, if it holds $\rho_k \geq \eta_1$ and the set of indices of all successful iterations is denoted by

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

Similarly the set of indices

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

denotes the set of very successful iterations and all iterations k with $\rho_\phi^k < \eta_1$ are called unsuccessful. With this classification of iterations the following Lemma illustrates the behavior of TRAHM for non-Pareto critical iteration points.

Lemma 4.7 *Let $k \in \mathbb{N}$ be an iteration and suppose Assumptions 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6 hold. Suppose furthermore that x^k is not Pareto critical for (MOP) and*

$$\delta_k \leq \frac{\kappa_\phi(1-\eta_2)\omega(x^k)}{\kappa_{\mathrm{e}}} \tag{24}$$

with $\kappa_{\mathrm{e}} := \max_{i=1,\ldots,q} \max\{\kappa_{\mathrm{cnd}i}, \kappa_{\mathrm{uhm}_i}\} > 0$ and $\kappa_\phi \in (0,1)$ from Assumption 4.6. 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 2.2 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 (24), the definition of κ_{e} and Remark 4.3 it follows

$$\delta_k \leq \frac{\kappa_\phi(1-\eta_2)\omega(x^k)}{\kappa_{\mathrm{e}}} < \frac{\omega(x^k)}{\kappa_{\mathrm{e}}} \leq \frac{\omega(x^k)}{\max\limits_{i=1,\ldots,q} \kappa_{\mathrm{uhm}_i}} \leq \frac{\omega(x^k)}{\beta_k^\phi}. \tag{25}$$

According to Assumption 4.6 it holds

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

Now consider $\rho_\phi^k = \left(\phi(x^k) - \phi(x^{k+})\right) / \left(\phi_m^k(x^k) - \phi_m^k(x^{k+})\right)$ the trial point acceptance quotient defined in (7). Due to the interpolation condition (9) it holds $\phi_m^k(x^k) = \phi(x^k)$ and from Lemma 4.6, the definition of κ_{e} and (24) it follows

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

This implies $\rho_\phi^k \geq \eta_2$ and therefore $k \in \mathcal{V}$. According to the trust region update in step 4 of TRAHM in section 3 it holds for the new trust region radius $\delta_{k+1} \geq \delta_k$. □

The next lemma shows that whenever the function ω is strictly positive, so is the trust region radius. Hence as long as no Pareto critical point is being approached the trust region radius is bounded from below by a strictly positive constant.

Lemma 4.8 *Suppose Assumptions 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6 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 TRAHM and $\kappa_\phi, \kappa_{\mathrm{e}}$ defined in Assumption 4.6 and Lemma 4.7. 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 TRAHM it holds $\gamma_1 \delta_{k_0-1} \leq \delta_{k_0}$. These two inequalities imply

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

Because of the assumption on $\omega(x^{k_0-1})$ and Lemma 2.2 x^{k_0-1} is not Pareto critical for (*MOP*). Therefore the preconditions of Lemma 4.7 are satisfied 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 preceeding results it can be proved that in case of finitely many successful iterations TRAHM converges to a Pareto critical point.

Lemma 4.9 *Suppose Assumptions 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6 hold and TRAHM 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 $x^k = x^{k+1}$ for all $k \geq j$ 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. $\rho_\phi^k < \eta_1$ for all $k > k_0$. Step 3 of TRAHM ensures $x^{k_0+1} = x^{k_0+j}$ for all $j \in \mathbb{N}$. 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 a Pareto critical point for (*MOP*). Then Lemma 4.7 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 TRAHM has infinitely many successful iterations.

Lemma 4.10 *Suppose Assumptions 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6 hold and TRAHM 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 $\{\omega(x^k)\}$ and a constant $\varepsilon > 0$ with $\omega(x^k) \geq \varepsilon$ for all $k \in \mathbb{N}$. According to Lemma 4.8 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 4.3 it follows

$$\beta_k^\phi \leq \max_{i=1,\ldots,q} \kappa_{\text{uh}m_i} \leq \max_{i=1,\ldots,q} \{\kappa_{\text{uh}m_i}, \kappa_{\text{cnd}i}\} = \kappa_{\text{e}}$$

for every iteration $k \in \mathbb{N}$ given the constants $\kappa_{\text{uh}m_i}, \kappa_{\text{cnd}i}$ and κ_{e} from Assumption 4.2, Lemma 4.1, Assumption 4.3 and Lemma 4.7. Consider a successful iteration $k \in \mathcal{S}$. Then it holds $\rho_\phi^k \geq \eta_1$ and it follows from Assumption 4.6

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

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 the general assumption 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 TRAHM. It shows that the algorithm produces a sequence of iterates with ω converging to zero. According to Lemma 2.2 this characterizes Pareto criticality.

Theorem 4.1 *Suppose Assumptions 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6 hold. Then TRAHM produces a sequence of iterates $\{x^k\}$ with*

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

If the sequence $\{x_k\}$ has accumulation points, then every of these points is a Pareto critical point for (MOP).

Proof. If TRAHM has only finitely many successful iterations $k \in \mathcal{S}$, then according to Lemma 4.9 the sequence of iterates $\{x^k\}$ converges to a Pareto critical point of (MOP). By Lemma 2.2 it follows $\lim_{k\to\infty}\omega(x^k) = 0$.
Now consider the case if there are infinitely many successful iterations $k \in \mathcal{S}$. Assume that there exists a subsequence of successful iterates $\{t_j\} \subset \mathcal{S}$ with

$$\omega(x^{t_j}) \geq 2\varepsilon > 0 \tag{26}$$

for some constant $\varepsilon > 0$ and for all j. By Lemma 4.10 it follows that for all t_j there exists a first successful iteration $l_j > t_j$ satisfying $\omega(x^{l_j}) < \varepsilon$. Then there is another subsequence indexed by $\{l_j\}$ such that

$$\omega(x^k) \geq \varepsilon \text{ for } t_j \leq k < l_j \text{ and } \omega(x^{l_j}) < \varepsilon. \tag{27}$$

Consider the subsequence whose indices are in $\mathcal{K} := \{k \in \mathcal{S} \mid \exists\, j \in \mathbb{N} : t_j \leq k < l_j\} \subseteq \mathcal{S}$, where t_j and l_j belong to the two subsequences defined above. For every successful iteration 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 4.6, Remark 4.3 and (27) imply for $k \in \mathcal{K}$

$$\phi(x^k) - \phi(x^{k+1}) \geq \eta_1 \left(\phi_m^k(x^k) - \phi_m^k(x^{k+1})\right) \geq \eta_1 \kappa_\phi \varepsilon \min\left\{\frac{\varepsilon}{\kappa_{\mathrm{e}}}, \delta_k\right\}. \tag{28}$$

The sequence $\left\{\phi(x^k)\right\}$ is monotonically decreasing and bounded from below. Hence $\left\{\phi(x^k)\right\}$ 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 (28) and it holds for $k \in \mathcal{K}$ sufficiently large

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

and consequently for j sufficiently large

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

Again, because the sequence $\left\{\phi(x^k)\right\}$ is monotonically decreasing and bounded from below it holds $\lim_{j\to\infty} \frac{1}{\eta_1\kappa_\phi\varepsilon} \left(\phi(x^{t_j}) - \phi(x^{l_j})\right) = 0$ and thus

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

Since ω is uniformly continuous due to Assumption 4.1, see Remark 4.1, it follows

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

This is a contradiction to the definition of the sequences $\{t_j\}$ and $\{l_j\}$ in (27) which implies with (26) that $\left|\omega(x^{t_j}) - \omega(x^{l_j})\right| \geq \varepsilon$ holds. Consequently, no subsequence satisfying (26) can exist and it holds $\lim_{k\to\infty} \omega(x^k) = 0$.
Let $\bar{x}$ be an accumulation point of the sequence $\{x_k\}$ produced by TRAHM and assume that it is not a Pareto critical point for (MOP). Then according to Lemma 2.2 it holds $\omega(\bar{x}) > 0$. This is a contradiction to $\lim_{k\to\infty} \omega(x^k) = 0$ and hence every accumulation point of $\{x_k\}$ is Pareto critical for (MOP). □

The convergence result can also be proved if all the objectives are expensive functions. Of course assumptions like Assumption 4.3 are then needed for all functions.

5 Numerical details and modifications of the algorithm

The Algorithm as presented in section 3 is formulated for the theoretical considerations in section 4. For the numerical realization some modifications can be made.

5.1 Stopping criterion

For the implementation of TRAHM a suitable stopping criterion is needed. Since one of the objectives is expensive regarding the evaluation time it is reasonable to set a maximum number of allowed function evaluations and stop the algorithm when this number is reached.
Furthermore TRAHM is designed to reduce the trust region radius whenever there is no sufficient decrease possible with the current model functions. Moreover Lemma 4.7 assures that whenever the current iteration point x^k is not Pareto critical 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 4.8, the trust region radius is bounded from below as long as x^k is not a Pareto critical point. Hence if the trust region radius is small enough in terms of being smaller than a suitable constant $\varepsilon_{tr} > 0$ the algorithm can stop.
The Pascoletti-Serafini scalarization (PS) is used to compute the search direction in every iteration. According to Lemma 3.2 the solution t^{k+} of (PS) is strictly negative as long as x^k is not weakly efficient for $(MOPm)$. Thus, if the models are reliable approximations the algorithm can stop if t^{k+} is equal to zero.

5.2 Trust region update

The update rule for the trust region radius in TRAHM uses the general formulation from the literature, see for example [8, 36]. For the implementation it is specified as

$$\delta_{k+1} = \begin{cases} \frac{1}{2}\delta_k & \rho_\phi^k < \eta_1 \\ \delta_k & \eta_1 \leq \rho_\phi^k < \eta_2 \\ 2\delta_k & \rho_\phi^k \geq \eta_2 \end{cases} .$$

5.3 User-given information

The presented method is only a local method producing one efficient solution of several efficient solutions of the considered multiobjective problem. Primarily, the ideal point determines the outcome of TRAHM. Instead of computing the individual minima of the model functions $m_i^k, i \in \{1, ..., q\}$, in every iteration $k \in \mathbb{N}$, a strictly lower bound is also sufficient. All the findings about TRAHM presented in section 4 also hold if the ideal point p^k is replaced by a point $\tilde{p}$ with $\tilde{p}_i < p_i^k = \min_{x \in B_k} m_i^k(x)$ for all $i \in \{1, ..., q\}$. In this context user-given information can be included. In some applications the user has additional information about the optimization problem, such as a 'working solution'. Furthermore in most applications there is a preference for the solution or a desired result that may be unrealizable. Yet, this kind of information can be included in TRAHM by replacing the ideal point by this user-given desired point $\tilde{p}$.

5.4 Saving computation time

In methods for expensive black box optimization the main strategy for saving computation time is reducing the number of expensive function evaluations. In the context of this paper one option is to update the model for f_1 only in iterations in which the approximation by the model was too poor, that is it holds $\rho_\phi^k < \eta_1$. Otherwise the old model of iteration k is reused in the next iteration. Furthermore in every iteration interpolation points need to be computed for the model of the expensive function f_1. Instead of recomputing them in every iteration every point situated in the current trust region and not violating the quality criterion of well poisedness can be reused.

5.5 Constrained optimization problems

TRAHM is only formulated for unconstrained problems, but box constraints can easily be added without affecting the method. The subproblems of computing the ideal point in step 1 of TRAHM can still be solved quickly even with box constraints. For including such constraints into the computation of the trial points it is possible to use projection methods as suggested in basic trust region methods, see [8]. For the implementation we do not use this approach but include the box constraints into the Pascoletti Serafini scalarization. Still, these auxiliary optimization problems are easy to solve.

6 Experimental results

TRAHM has been implemented in Matlab (version 2017a) with the modifications and stopping criteria described in section 5 and tested for several biobjective problems. All considered problems are test problems and do not involve an actual expensive function. Furthermore the test problems are both self-chosen and from the literature [24, 13, 35, 5]. Among them are quadratic and nonquadratic functions, convex and nonconvex problems, either unconstrained or with box constraints.

Since TRAHM computes only one solution and does not approximate the set of Pareto critical points, the results are compared to the weighted sum scalarization of (MOP) with equal weights $w_1 = w_2 = 0.5$. For convex problems every efficient point can be computed by a weighted sum with suitable weights. For nonconvex problems only a subset of the efficient points can be computed, but still the test results can be used to compare the needed amount of function evaluations.

To solve the surrogate scalar problem the algorithm EFOS (Expensive Function Optimization Solver) [34] is used. This is a solution method for expensive, simulation-based scalar optimization problems which also uses the trust region approach. The purpose of this solution method also is to save computation time and reduce the number of function evaluations. Thus, it is very well suited as a comparative method. As a stopping criterion a criticality measure using the gradients of the model functions is applied in conjunction with a validity criterion for the models.

Both algorithms were tested on ten convex problems and eight nonconvex problems with different starting points. For all convex test problems TRAHM is successful and produces a Pareto critical point, mostly an efficient point. Even for most of the nonconvex problems

a Pareto critical point, often an efficient point, is computed.
As expected due to the weighted sum approach EFOS often computes the individual minima of the objective functions or the exact tradeoff between the two functions, whereas TRAHM is capable of finding points from different areas of the Pareto front. However, a more important aspect is the number of function evaluations and TRAHM often needs significantly less function evaluations than EFOS. In general the test results show that TRAHM can save computation time.
Exemplarily, we want to present some test results for two selected test problems. The first test example is a self-chosen convex problem defined by

$$\min_{x\in\mathbb{R}^n} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x\in\mathbb{R}^n} \begin{pmatrix} \frac{1}{2}x_1^2 + x_2^2 - 10x_1 - 100 \\ x_1^2 + \frac{1}{2}x_2^2 - 10x_2 - 100 \end{pmatrix}. \tag{A}$$

For all starting points of this test example both algorithms produce efficient points. TRAHM needs 9-11 function evaluations and therefore significantly less than EFOS which needs 13-78. Figure 2 shows one result for TRAHM on the left-hand side and for EFOS on the right-hand side. The starting point is marked orange and the solution is marked yellow. The image set is represented by scattered gray points and the evaluated points are marked black. This example illustrates the effect of the trust region concept used in

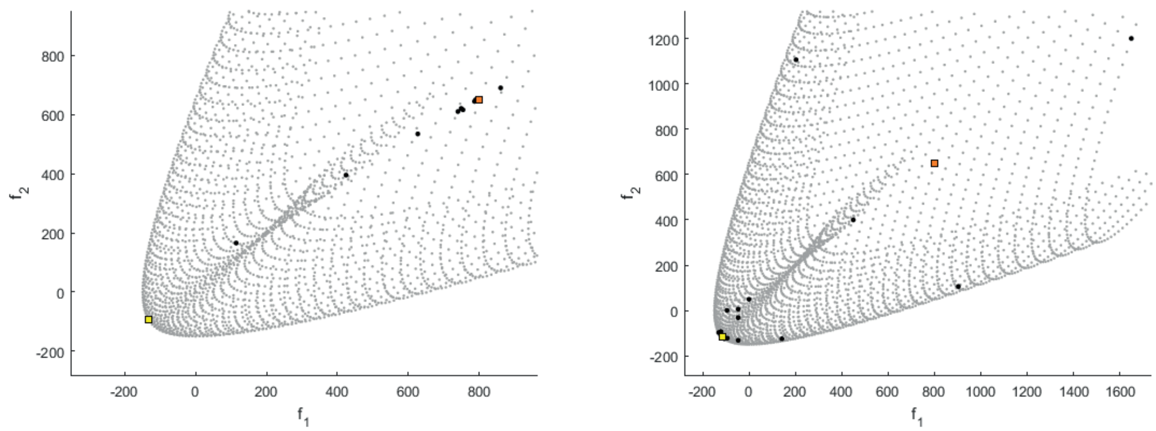

Figure 2: Test run for TRAHM and EFOS for test problem (A)

TRAHM. For EFOS, the evaluated points are spread in the image space and more evaluations are needed. By contrast, the local trust regions in TRAHM have the effect that the model is only built on points near to the current iteration point. Furthermore this figure shows that in TRAHM the model was not actualized very often, but could be reused for several iterations. It also illustrates the trust region as a step size control since in the last iterations the step size increased. Apparently the model was reliable also in bigger and shifted trust regions.
To illustrate how good TRAHM also works for nonconvex problems we want to consider the test problem "Lis" [5] defined by

$$\min_{x\in\Omega} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x\in[-5,10]\times[-5,10]} \begin{pmatrix} \sqrt[8]{x_1^2 + x_2^2} \\ \sqrt[8]{(x_1-0.5)^2 + (x_2-0.5)^2} \end{pmatrix}. \tag{B}$$

For all starting points of this test example both algorithms produce Pareto critical points, mostly even efficient points. Since (B) is a nonconvex problem the weighted sum approach with EFOS gives as expected a point close to the individual minimum of f_2 as a solution. Still, TRAHM clearly saves function evaluations by using 10-34 evaluations in the different test runs, whereas EFOS needs 77-101. In Figure 3 again the result of one test run for TRAHM is shown on the left-hand side and for EFOS on the right-hand side. In this example the starting point marked orange is already close to the Pareto front. Figure 3 illustrates this for the image space, but it is also reflected in the domain. EFOS needs evaluations from nearly the whole domain and image space, whereas TRAHM does not spread the evaluations that much.

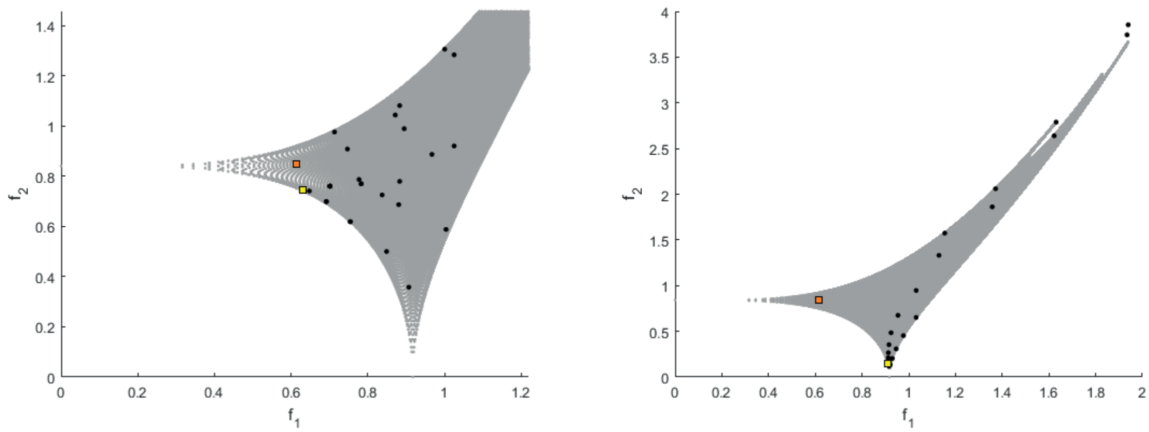

Figure 3: Test run for TRAHM and EFOS for test problem (B)

7 Conclusions

This paper presents the new multiobjective trust region me-thod TRAHM. The objective functions can be heterogeneous and besides analytically given functions also expensive, simulation-given functions can be considered. The algorithm can also be modified to consider only expensive functions. Even then the presented approach is new due to the direction that is used to compute the next iteration point. This direction is computed by using local ideal points and an auxiliary optimization problem for the model functions. Thus, starting from the initial point a new point is generated - if possible - that decreases every function value of the models. The connection to the original functions is made by suitable validity assumptions on the model functions. Generally, no derivative information is necessary and hence the algorithm is well suited for expensive functions for which derivative information is absent or not available with reasonable effort. Since different model functions are used for different types of objective functions, e.g. simulation-given or analytically given functions, heterogeneity of the objective functions can easily be considered.
It is proved that the sequence of iterates $\{x^k\}$ produced by TRAHM converges to a Pareto

critical point in terms of $\omega(x^k)$ converging to zero. The test runs confirm the theoretical findings by showing that TRAHM computes Pareto critical points. Mostly, these points are even efficient points. Furthermore TRAHM is capable of computing points from different parts of the Pareto front. The numerical experiments also show that TRAHM can significantly save computation time due to the local trust regions and the updating of the model functions only if necessary.

In future work modifications for TRAHM will be developed to exploit the heterogeneity of the objective functions even more. With regard to practical applications also user-given information already described to some extent in subsection 5.3 shall be regarded. So far TRAHM computes only one solution. The aim also is to spread the points during the computations and to obtain several Pareto critical points in consideration of the computational effort. For all these purposes some heuristic strategies will be considered which make use of the cheap functions to a larger extent.

References

[1] K. Anagnostopoulos and G. Mamanis. Multiobjective evolutionary algorithms for complex portfolio optimization problems. *Comput. Manag. Sci.*, 8:259–279, 2011.

[2] C. Audet and J. Dennis. Mesh adaptive direct search algorithms for constrained optimization. *SIAM J. Optim.*, 17(1):188–217, 2006.

[3] C. Audet, G. Savard, and W. Zgal. Multiobjective optimization through a series of single-objective formulations. *SIAM J. Optim.*, 19(1):188–210, 2008.

[4] 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*, pages 380–390, 2013.

[5] C. Coello, D. V. Veldhuizen, and G. Lamont. *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic/Plenum Publishers, 2002.

[6] A. Conn, N. Gould, and P. Toint. Global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM J. Numer. Anal.*, 25(2):433–460, 1988.

[7] A. Conn, N. Gould, and P. Toint. Correction to the paper on global convergence of a class of trust region algorithms for optimization with simple bounds. *SIAM J. Numer. Anal.*, 26(3):764–767, 1989.

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

[9] A. Conn, K. Scheinberg, and L. Vicente. Global convergence of general derivative-free trust-region algorithms to first- and second-order critical points. *SIAM J. Optim.*, 20(1):387–415, 2009.

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

[11] A. Custodio, J. Madeira, A. Vaz, and L. Vicente. Direct multisearch for multiobjective optimization. *SIAM J. Optim.*, 21(3):1109–1140, 2011.

[12] I. Das and J. E. Dennis. Normal-boundary intersection: a new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM J. Optim.*, 8(3):631–657, 1998.

[13] K. Deb. Multiobjective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation*, 7(3):205–230, 1999.

[14] L. Drummond and B. Svaiter. A steepest descent method for vector optimization. In *J. Comput. Appl. Math.*, volume 175, pages 395–414, 2005.

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

[16] G. Eichfelder. An adaptive scalarization method in multiobjective optimization. *SIAM J. Optim.*, 19(4):1694–1718, 2009.

[17] G. Eichfelder. *Variable Ordering Structures in Vector Optimization.* Vector Optimization. Springer, 2014.

[18] B. Flach. *Rekonstruktion mit Vorinformation für die Niedrigstdosis-CT-Fluoroskopie in der Interventionellen Radiologie.* PhD thesis, Friedrich-Alexander-Universität Erlangen Nürnberg, 2014.

[19] J. Fliege, L. Drummond, and B. Svaiter. Newton‘s method for multiobjective optimization. *SIAM J. Optim.*, 20(2):602–626, 2009.

[20] J. Fliege and B. Svaiter. Steepest descent methods for multicriteria optimization. *Math. Methods of Oper. Res.*, 51(3):479–494, 2000.

[21] J. Fliege and A. Vaz. A method for constrained multiobjective optimization based on sqp techniques. *SIAM J. Optim.*, 26(4):2091–2119, 2016.

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

[23] S. Greco, K. Klamroth, J. Knowles, and G. Rudolph. Understanding complexity in multiobjective optimization (dagstuhl seminar 15031). *Dagstuhl Reports*, 5(1), 2015.

[24] S. Huband, P. Hingston, L. Barone, and L. While. A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Trans. Evolutionary Computation*, 10(5):477–506, 2006.

[25] J. Jahn. *Vector Optimization.* Springer, 2011.

[26] T. Kolda, R. Lewis, and V. Torczon. Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Rev.*, 45(3):385–482, 2003.

[27] D. T. Luc, T. Q. Phong, and M. Volle. Scalarizing functions for generating the weakly efficient solution set in convex multiobjective problems. *SIAM J. Optim.*, 15(4):987–1001, 2005.

[28] A. Pascoletti and P. Serafini. Scalarizing vector optimization problems. *J. Optim. Theory Appl.*, 42(4):499–524, April 1984.

[29] J.-H. Ryu and S. Kim. A derivative-free trust-region method for biobjective optimization. *SIAM J. Optim.*, 24(1):334–362, 2014.

[30] T. Stewart et al. *Multiobjective Optimization. Interactive and Evolutionary Approaches*, chapter Real-World Applications of Multiobjective Optimization, pages 285–327. Springer, 2008.

[31] F. Strauss, J.-A. Desideri, R. Duvigneau, and V. Heuveline. multiobjective optimization in hydrodynamic stability control. *[Research Report] RR-6608, INRIA*, 2008. <inria-00309693>.

[32] M. Tapia and C. Coello. Applications of multi-objective evolutionary algorithms in economics and finance: A survey. In *proceedings of the IEEE Congress on Evolutionary Computation*, pages 532–539, 2007.

[33] D. Terzijska, M. Porcelli, and G. Eichfelder. Multi-objective optimization in the lorentz force velocimetry framework. In *Book of digest & program*, pages 81–82. OIPE, International Workshop on Optimization and Inverse Problems in Electromagnetism 13, 2014.

[34] A. Thekale. *Trust-Region Methods for Simulation Based Nonlinear Optimization*. PhD thesis, Friedrich-Alexander-Universität Erlangen Nürnberg, 2011.

[35] D. V. Veldhuizen. *Multiobjective Evolutionary Algorithms: Classifications, Analyses and New Innovations*. PhD thesis, Air Force Institute of Technology, Wright Patterson AFB, OH, USA, 1999.

[36] K. Villacorta, P. Oliveira, and A. Soubeyran. A trust-region method for unconstrained multiobjective problems with applications in satisficing processes. *J. Optim. Theory Appl.*, 160:865–889, 2014.

[37] S. Wild and C. Shoemaker. Global convergence of radial basis functon trust region derivative-free algorithms. *SIAM J. Optim.*, 21(3):761–781, 2011.