

A TRUST-REGION ALGORITHM FOR HETEROGENEOUS MULTIOBJECTIVE OPTIMIZATION*

JANA THOMANN[†] AND GABRIELE EICHFELDER[†]

Abstract. This paper presents a trust-region method for multiobjective heterogeneous optimization problems. One of the objective functions is an expensive black-box function, given, for example, 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. The algorithm generates a sequence of iterates. It is proved that any limit point is Pareto critical. Numerical results are presented and compared to two other algorithms.

Key words. multiobjective optimization, trust-region method, derivative-free algorithm, heterogeneous optimization, Pareto critical point

AMS subject classifications. 90C29, 90C56, 90C30

DOI. 10.1137/18M1173277

1. Introduction. Multiobjective optimization problems can be found in fields such as engineering, medicine, economics, or finance [38, 16, 40, 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 computational effort. Hence, use of derivative information should be avoided and therefore many solution methods from the literature [15, 20, 21, 22] 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 [26, pp. 125ff]. 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 [41, 19, 39]. For instance, in Lorentz force velocimetry [41] the task is to find an optimal design of a magnet that minimizes

*Received by the editors February 28, 2018; accepted for publication (in revised form) January 2, 2019; published electronically April 3, 2019.

<http://www.siam.org/journals/siopt/29-2/M117327.html>

Funding: This work was funded by DFG under grant GRK 1567.

[†]Institute for Mathematics, Technische Universität Ilmenau, Ilmenau, Germany (jana.thomann@tu-ilmenau.de, gabriele.eichfelder@tu-ilmenau.de, <https://www.tu-ilmenau.de/en/mmor/>).

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 [26, p. 124], heterogeneous problems with expensive functions also occur in imaging techniques in interventional radiology [19]. 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 methods for scalar optimization problems. Among numerous scalarization approaches, e.g., [13, 17, 32], 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 that is easy and quick to compute, this has no impact. Hence, such scalarization methods cannot exploit the heterogeneity of objective functions and therefore neglect some information.

Other methods for multiobjective optimization problems, such as the generalized steepest descent method [15, 21] or the generalized Newton method [20], need derivative information and are therefore not applicable to heterogeneous problems where the derivatives are not available with reasonable effort. 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, in both scalar and multiobjective optimization, is direct search [2, 11, 12]. This approach only needs function values, and there are several versions and realizations, such as the basic direct multisearch (DMS) [12] or biobjective mesh adaptive direct search (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 [31]. 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 the information given (namely, the derivative information of the cheap functions) is used during the optimization process.

This also applies to evolutionary or genetic algorithms [14, 6]. These are heuristic methods with the search strategy of sampling by changing or recombining former points. This can cause high computational effort when expensive functions are considered. Furthermore, such 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, the approach of surrogate models and meta-modeling is also often used [30, 29, 35]. The model for the objective function is built during the optimization process, based on the results in the iterations, and used in the subsequent search. This approach is well suited for expensive functions, but not required for analytically given functions.

Another approach on which derivative-free methods are based is the trust-region method [7, 8, 9, 10, 11, 33]. There are also multiobjective realizations of this approach

[36, 44]. Trust-region methods were 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 by Villacorta, Oliveira, and Soubeyran [44]. They prove convergence to a Pareto critical point using a characterization of such points that is also used in multiobjective descent theory [15, 21]. The used assumptions are derived from the scalar version of the trust-region approaches, and the convergence analysis follows the strategy and structure of the proof from the basic scalar approach [9] 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 the use of derivative information of the expensive function should be avoided.

Unlike this approach, a trust-region algorithm is presented in [36] 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.

Another multiobjective trust-region approach is presented in [5]. Following the steepest descent method [21] and the Newton method [20], a multiobjective descent direction is computed. This needs derivative information of first and second order, and therefore it is not suitable for heterogeneous optimization problems. Carrizo, Lotito, and Maciel [5] prove convergence to a point satisfying a necessary condition for optimality.

So far, there are no solution methods for heterogeneous multiobjective problems that can exploit the differences in the objective functions. This paper presents a new trust-region method that can account for heterogeneity and gives a theoretical foundation for further modifications using this heterogeneity.

Like [44], 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 local ideal points. 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 modification. Still, we can use the same strategy as [44] to prove convergence to a Pareto critical point by also using the characterization of such points from [21]. 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 [9].

The algorithm presented in this paper—applied once—generates one Pareto critical point. Different starting points result, in general, in different Pareto critical points. Generally, the aim for multiobjective optimization problems is to approximate the whole set of efficient points. If expensive functions are involved, this is often unrealistic due to the high computational effort. In this case, computing one efficient point is already a benefit. Indeed, in applications, the aim is often to improve the function values within a maximum number of permitted function evaluations. The method presented is useful and suitable for such problems due to the search strategy.

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 given 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

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

with $f(x) = (f_1(x), \dots, f_q(x))^\top$. The objective functions $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ are assumed to be twice continuously differentiable for all $i = 1, 2, \dots, q$, and $\max_{i=1, \dots, q} f_i(x)$ is assumed to be bounded from below. The function f_1 is 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 , $i = 2, \dots, q$, are so-called cheap functions that are given analytically, easy to compute, and easily yield function values and derivatives. For defining solutions of (MOP) we use the optimality concept for multiobjective optimization problems according to [28].

DEFINITION 2.1. A point $\bar{x} \in \mathbb{R}^n$ is called (an) efficient (solution) for (MOP) (or Pareto optimal) if there exists no point $x \in \mathbb{R}^n$ satisfying $f_i(x) \leq f_i(\bar{x})$ for all $i \in \{1, 2, \dots, q\}$ and $f(x) \neq f(\bar{x})$. The image of an efficient point is called nondominated.

A point $\bar{x} \in \mathbb{R}^n$ is called (a) weakly efficient (solution) for (MOP) (or weakly Pareto optimal) if there exists no point $x \in \mathbb{R}^n$ satisfying $f_i(x) < f_i(\bar{x})$ for all $i \in \{1, 2, \dots, q\}$.

These concepts can be restricted to local areas. Accordingly, a point $\bar{x} \in \mathbb{R}^n$ is called locally (weak) efficient for (MOP) if there exists a neighborhood $U \subset \mathbb{R}^n$ with $\bar{x} \in U$ such that \bar{x} is (weakly) efficient for (MOP) in U .

Obviously every efficient point is weakly efficient. The following concept [21] gives a necessary condition for weak efficiency.

DEFINITION 2.2. Let $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, 2, \dots, q$, be continuously differentiable. 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 $j \in \{1, 2, \dots, q\}$ such that $\nabla_x f_j(\bar{x})^\top d \geq 0$ 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 $\bar{x} \in \mathbb{R}^n$ be a Pareto critical point according to the above definition. Then it holds that $\nabla_x f(\bar{x})^\top d \geq 0$ for all $d \in \mathbb{R}^n$. Hence, $\nabla_x f(\bar{x}) = 0_n$ holds, 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, [21, 28].

LEMMA 2.3. 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 [15, 20, 21].

LEMMA 2.4. Let $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ be continuously differentiable functions for all $i = 1, 2, \dots, q$. For the function

$$(1) \quad \omega(x) := - \min_{\|d\| \leq 1} \max_{i=1, \dots, q} \nabla_x f_i(x)^\top d$$

the following statements hold.

- (i) The mapping $x \mapsto \omega(x)$ is continuous.
- (ii) It holds that $\omega(x) \geq 0$ for all $x \in \mathbb{R}^n$.
- (iii) A point $x \in \mathbb{R}^n$ is Pareto critical for (MOP) if and only if $\omega(x) = 0$ holds.

LEMMA 2.5. 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 $f_i(x + t d_\omega) < f_i(x)$ holds for all $t \in (0, t_0]$ and for all $i \in \{1, 2, \dots, q\}$.
- (ii) There exist scalars $\alpha_i \in [0, 1]$ for $i \in \{1, 2, \dots, q\}$ with $\sum_{i=1}^q \alpha_i = 1$ and $\mu \geq 0$ such that $d_\omega = -\mu \sum_{i=1}^q \alpha_i \nabla_x f_i(x)$ holds. If x is not Pareto critical for (MOP), it holds that $\|d\| = 1$. If x is Pareto critical for (MOP), it holds that $d_\omega = \sum_{i=1}^q \alpha_i \nabla_x f_i(x) = 0$. Furthermore, $\omega(x) \leq \|\sum_{i=1}^q \alpha_i \nabla_x f_i(x)\|$ holds.

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

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

Let (t_ω, d_ω) denote a solution of (2) and initially let x not be 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 the following holds:

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

If x is Pareto critical, then the zero vector is a solution of (2) and the KKT conditions imply the existence of constants $\alpha_i \in [0, 1]$, $i \in \{1, 2, \dots, q\}$, 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), $-t_\omega = \omega(x)$ holds. This implies $\nabla_x f_i(x)^\top d_\omega \leq t_\omega$ for all $i \in \{1, 2, \dots, q\}$, 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 that

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

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

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 that $a_i \leq b_i$ for all $i \in \{1, \dots, n\}$.

3. Algorithm description. The basic trust-region concept [9, 11] is constructed for unconstrained scalar optimization problems with a twice continuously differentiable objective function bounded from below. It is an iterative method that 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 the trust region and is defined by

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

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 [9]. Now consider a multiobjective optimization problem of the form in (MOP), with f_1 being an expensive, simulation-given function. The multiobjective method presented in this paper is also an iterative approach, and in every iteration $k \in \mathbb{N}$, each objective function f_i with $i \in \{1, 2, \dots, q\}$ is replaced by a suitable quadratic model $m_i^k : \mathbb{R}^n \rightarrow \mathbb{R}$ that satisfies the interpolation condition

$$(5) \quad f_i(x^k) = m_i^k(x^k);$$

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

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

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 local ideal point $p^k = (p_1^k, \dots, p_q^k)^\top$ defined by $p_i^k = \min_{x \in B_k} m_i^k(x)$ for all $i = 1, 2, \dots, q$ is used as a search direction. Moving along this direction as far as possible, as described in detail in subsection 3.2, gives a candidate for the next iteration point, the so-called trial point x^{k+} . As in the scalar trust-region method [9, 11], the trial point x^{k+} is only accepted as the next iteration point if a condition describing the improvement of the function values is met; this can be found in subsection 3.3. For this purpose, the auxiliary functions

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

are defined. The trust-region and the model functions are adapted according to the outcome of this trial point acceptance test.

The multiobjective heterogeneous trust-region algorithm (MHT) is formulated in Algorithm 1. It describes a new trust-region approach, which differs from known methods by the computation of the search direction. This direction is determined in the image space by using the local ideal points of the model functions. A starting point, some parameters, and the objective functions are needed as inputs, whereby f_1 is expensive and f_i are cheap for all $i \in \{2, 3, \dots, q\}$. The choice of the parameters η_1 , η_2 , γ_1 , and γ_2 can of course be problem dependent, but according to [9] 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 replace all functions, including the analytically given functions, by quadratic models, which is standard in trust-region methods. A quadratic model $m : \mathbb{R}^n \rightarrow \mathbb{R}$ for a function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is given by

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

Algorithm 1 MHT.

Input: functions f_i , $i = 1, 2, \dots, q$, 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 $i = 1, 2, \dots, q$.

Step 1: Ideal point

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

Step 2: Trial point

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

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

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 $f_i(x^{k+})$, $i = 1, 2, \dots, q$, 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

Define models m_i^{k+1} for $i = 1, 2, \dots, q$, set $k = k + 1$, and go to **Step 1**.

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 , $i = 2, \dots, q$, 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

$$(7) \quad m_T(x; f_i, x^k) := f_i(x^k) + \nabla_x f_i(x^k)^\top (x - x^k) + \frac{1}{2}(x - x^k)^\top \nabla_{xx} f_i(x^k)(x - x^k)$$

in every iteration $k \in \mathbb{N}$ using the current iteration point x^k ($i = 2, \dots, q$). For such models it always holds that $\nabla_x m_i^k(x^k) = \nabla_x f_i(x^k)$. 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 $m_1 : \mathbb{R}^n \rightarrow \mathbb{R}$ for the expensive function f_1 , let \mathcal{P}_n^2 denote the space of polynomials of degree less than or equal to 2 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 $\psi = \{\psi_1, \dots, \psi_p\}$ 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 $Y = \{y^1, y^2, \dots, y^p\} \subset \mathbb{R}^n$ 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 $i = 1, 2, \dots, q$. For the basis ψ we choose the basis of quadratic Lagrange polynomials $l_i \in \mathcal{P}_n^2$, $i = 1, \dots, p$, defined by

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

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 $Y_k = \{y^1, y^2, \dots, y^p\} \subset B_k$ 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 poisedness; see [11, 43]. Since Lagrange polynomials are not only compatible with this concept but commonly used for measuring poisedness, they are chosen as an interpolation basis here. Another option for building models in the trust-region scheme is radial basis functions. This option is described for scalar trust-region methods in [45].

The models for the cheap functions are recomputed in every iteration, but the model for f_1 is only updated if necessary to save function evaluations. Otherwise, the former model function is reused. To decide if a model update is necessary, different options can be considered, and the specific realization can be found in subsection 5.4.

3.2. Computing the trial point. The search for a sufficient decrease in the function values is realized by computing the ideal point $p^k = (p_1^k, \dots, p_q^k)^\top$ defined by $p_i^k = \min_{x \in B_k} m_i^k(x)$ for all $i = 1, 2, \dots, q$. 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. A trust-region approach is also 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, the original functions. The aim is to move as far as possible—as far as the trust region allows—in 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 [34] given by

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

with $r^k := f(x^k) - p^k \in \mathbb{R}_+^q$, p^k the ideal point of m^k in B_k , and $m^k = (m_1^k, \dots, m_q^k)^\top$ the model functions. This scalarization is also known as the Tammer–Weidner functional [25]. Note that $f(x^k) = m^k(x^k)$ holds in every iteration k due to the interpolation conditions (5). It is easy to see that the direction r^k is not the zero vector as long as x^k is not Pareto critical for $(MOPm)$.

Remark 3.1. Let x^k be not Pareto critical for $(MOPm)$. By Lemma 2.3, x^k is not locally weakly efficient for $(MOPm)$ and, as $x^k \in \text{int } B_k$, also not weakly efficient for $\min_{x \in B_k} m^k(x)$. Thus, x^k cannot be an individual minimum of one of the functions m_i^k , $i \in \{i = 1, 2, \dots, q\}$, on B_k ; hence, for the direction r^k of (PS) , $r_i^k = m_i^k(x^k) - \min_{x \in B_k} m_i^k(x) > 0$ holds for all $i \in \{i = 1, 2, \dots, q\}$.

Furthermore, in the case when $r^k \in \text{int } \mathbb{R}_+^q$, the problem (PS) minimizes 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 $i \in \{1, 2, \dots, q\}$. 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 (t^{k+}, x^{k+}) being the solution of (PS) . The image of the trial point x^{k+} is marked by the black dot.

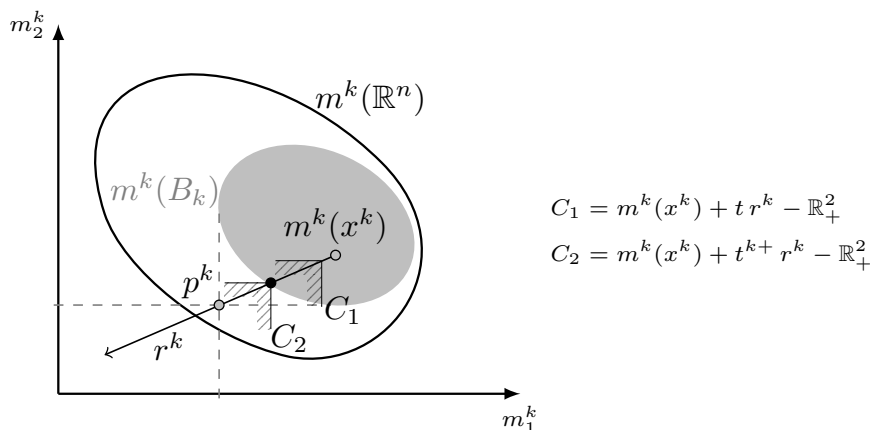


FIG. 1. Pascoletti-Serafini scalarization (PS) .

The optimization problem (PS) has some useful properties, which can be found in detail and with its proof in [16, Thm. 2.1].

LEMMA 3.2.

- (i) If (\bar{t}, \bar{x}) is a minimal solution of (PS) , then we have that \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 we have that \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 $r^k \in \text{int } \mathbb{R}_+^q$, then we have that $(0, \bar{x})$ is a minimal solution of (PS) .

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

LEMMA 3.3. 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 that $\bar{t} \in [-1, 0)$.

Proof. Let (\bar{t}, \bar{x}) be a minimal solution of (PS) . Since $(0, x^k)$ is always feasible for (PS) , $\bar{t} \leq 0$ holds. Due to x^k not being 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 $\bar{t} < 0$ holds.

Now suppose $\bar{t} := -1 - s < -1$ with $s > 0$. Resulting from the constraints of (PS) , it holds that $p^k - m^k(\bar{x}) \geq s r^k$. Again due to x^k not being weakly efficient, and thus $r^k > 0$, it follows that $p^k > m^k(\bar{x})$, which contradicts the definition of p^k . Consequently, $\bar{t} \in [-1, 0)$ holds. \square

3.3. Trial point acceptance test. Following the basic idea of trust-region methods the trial point x^{k+} is only accepted as the next iteration point if a condition describing the improvement of the function values is satisfied. In the scalar version we test whether for a scalar-valued function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and its model function

$m^k : \mathbb{R}^n \rightarrow \mathbb{R}$ the quotient $(f(x^k) - f(x^{k+})) / (m^k(x^k) - m^k(x^{k+}))$ is bigger than a given nonnegative constant. To transfer this criterion to the multiobjective case we choose the same approach as Villacorta, Oliveira, and Soubeyran in [44]: combining the objective functions in auxiliary functions given by

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

as defined in (6). The trial point is accepted if the quotient

$$(8) \quad \rho_\phi^k := \frac{\phi(x^k) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})}$$

is bigger than a given nonnegative constant. In this case there is a guaranteed descent in at least one component, as outlined in the following. If all functions are decreasing, this is also sufficient, but the formulated criterion is weaker and thus more points are accepted and fewer iterations are unsuccessful.

Due to the way in which x^{k+} was determined, it always holds that $\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq 0$. 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 3.3. Together with the definition of the trial point, it follows that $\phi_m^k(x^k) - \phi_m^k(x^{k+}) > 0$ as long as x^k is not weakly efficient.

Suppose $\rho_\phi^k > 0$ holds, which implies $\phi(x^k) - \phi(x^{k+}) > 0$. Then there exist indices $i, j \in \{1, 2, \dots, q\}$ such that $0 < f_i(x^k) - f_j(x^{k+}) \leq f_i(x^k) - f_i(x^{k+})$ holds. Therefore, the trial point x^{k+} guarantees a descent in at least one component of f . In MHT x^{k+} is accepted if ρ_ϕ^k is bigger than a strictly positive constant η_1 , not only to ensure a decrease in at least one component but to guarantee that this decrease is “sufficient.”

In the case when $\rho_\phi^k < 0$ there exist indices $i, j \in \{1, 2, \dots, q\}$ with $0 > f_i(x^k) - f_j(x^{k+}) \geq f_j(x^k) - f_j(x^{k+})$. This implies an increase in at least one component of f . Hence, the trial point is not accepted as the next iteration point.

Now assume $\rho_\phi^k = 0$. This implies $t^{k+} = 0$, $\phi_m^k(x^k) - \phi_m^k(x^{k+}) = 0$, or $\phi(x^k) - \phi(x^{k+}) = 0$. If $t^{k+} = 0$ holds, then according to Lemma 3.2(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 $\rho_\phi^k = 0$ in this case, the trust-region radius will be reduced and the model will be updated to affirm the model information. If the model is reliable, the trust region will also shrink in the next iterations and therefore the radius will converge to zero. If the model is not reliable, then there will be a subsequent iteration in which the trial point produces a sufficient decrease.

If $\phi_m^k(x^k) - \phi_m^k(x^{k+}) = 0$, there exist indices $i, j \in \{1, 2, \dots, q\}$ satisfying

$$m_j^k(x^k) \leq m_i^k(x^k) = m_j^k(x^{k+}) \geq m_i^k(x^{k+}).$$

Thus, 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 anticipate Assumption 4.14 here because it clarifies the trial point acceptance test. This assumption ensures a sufficient decrease in every

iteration of the form

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

with $\omega(x)$ from (1), $\kappa_\phi \in (0, 1)$, and $\beta_k^\phi > 0$. According to Lemma 2.4, $\omega(x) = 0$ holds if and only if the point x is Pareto critical for (MOP) , and according to Lemma 2.3 Pareto criticality is a necessary condition for local weak efficiency. If $\phi_m^k(x^k) - \phi_m^k(x^{k+}) = 0$ holds, 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 convergence proof for MHT to a Pareto critical point of the optimization problem (MOP) is presented, and some assumptions on the original and the model functions are needed for these results. All these assumptions are connected to the commonly used assumptions in the scalar trust region and derivative-free optimization context [9, 11, 42] or in multiobjective trust-region methods [36, 44]. After stating these basic assumptions, we give some standard statements about the accuracy of the model functions for f_1, \dots, f_q . Then, some auxiliary lemmas follow to motivate a sufficient decrease assumption for the trial point. Such an assumption is commonly used in the convergence analysis of trust-region methods and formulated according to the setting of this paper in Assumption 4.14. After obtaining this essential condition, which is used throughout the subsequent results, another basic statement (Lemma 4.16) about the accuracy of the models, this time for the functions ϕ and ϕ_m , follows.

Then, the main statements about convergence, consisting of several lemmas, are given. Initially, it is shown in Lemma 4.17 that whenever the current iteration point is not Pareto critical for (MOP) and the trust-region radius is small enough, the trial point generates a sufficient decrease and is accepted. Additionally, Lemma 4.18 shows that as long as no Pareto critical point is being approached the radius will not converge to zero. Then, the main result follows and it is proved that MHT produces a sequence of iterates that—if it exists—has a Pareto critical accumulation point. This result is first shown for a special case regarding the iterates in Lemma 4.19 and then in the general case (see Lemma 4.20 and Theorem 4.21).

As stated within the problem description in section 2, the functions f_i are assumed to be twice continuously differentiable for all $i \in \{1, 2, \dots, q\}$, and $\phi(x) = \max_{i=1, \dots, q} f_i(x)$ is assumed to be bounded from below. Furthermore, for every index $i \in \{1, 2, \dots, q\}$ and for every iteration $k \in \mathbb{N}$ the model functions m_i^k are assumed to be quadratic functions and therefore they are twice continuously differentiable. The model is assumed to be exact at the current iteration point x^k , that is, it holds that

$$(9) \quad m^k(x^k) = f(x^k)$$

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

$$(10) \quad \nabla_x m_i^k(x^k) = \nabla_x f_i(x^k)$$

for all $i \in \{2, 3, \dots, q\}$ 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, we need some further assumptions that are also commonly used in the trust-region concept. Furthermore, we give some basic statements about the accuracy of the model functions that are needed in the convergence results. In addition, a matrix norm compatible with the vector norm used is necessary. Since we use the Euclidean norm, we consider the Frobenius norm as the matrix norm.

Assumption 4.1. For every index $i \in \{1, 2, \dots, q\}$, the Hessian of the function f_i is uniformly bounded, that is, there exists a constant $\kappa_{\text{uhfi}} > 1$ satisfying

$$\|\nabla_{xx} f_i(x)\| \leq \kappa_{\text{uhfi}} - 1$$

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

Remark 4.2. Assumption 4.1, together with the mean value theorem, implies that the functions $\nabla_x f_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ are Lipschitz continuous for all $i = 1, 2, \dots, q$. It follows that the function ω defined in (1) is uniformly continuous; see also [44].

Assumption 4.3. For every index $i \in \{1, 2, \dots, q\}$, 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 $\kappa_{\text{uhmi}} > 1$ independent of k satisfying

$$\|\nabla_{xx} m_i^k(x)\| \leq \kappa_{\text{uhmi}} - 1$$

for all $x \in B_k$. The index “uhmi” stands for upper bound on the Hessian of m_i^k .

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, which can be found, for example, in [9].

DEFINITION 4.4. Let $i \in \{1, 2, \dots, q\}$ and $k \in \mathbb{N}$ be indices. A model function $m_i^k : \mathbb{R}^n \rightarrow \mathbb{R}$ is called valid for the function $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ in the trust region $B_k = \{x \in \mathbb{R}^n \mid \|x - x^k\| \leq \delta_k\}$ if there exists a constant $\kappa_{\text{cndi}} > 0$ such that

$$|f_i(x) - m_i^k(x)| \leq \kappa_{\text{cndi}} \delta_k^2$$

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 4.5. Suppose Assumptions 4.1 and 4.3 hold. In every iteration $k \in \mathbb{N}$, the model m_i^k is valid for f_i in B_k for all $i \in \{2, 3, \dots, q\}$, that is,

$$|f_i(x) - m_i^k(x)| \leq \kappa_{\text{cndi}} \delta_k^2$$

holds for all $x \in B_k$ and $\kappa_{\text{cndi}} := \max\{\kappa_{\text{uhfi}}, \kappa_{\text{uhmi}}\} - 1 > 0$.

Proof. Since the functions f_i are twice continuously differentiable, it follows from Taylor’s theorem for every $h \in \mathbb{R}^n$ with $\|h\| \leq \delta_k$ that

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

with $\xi_{ij}^k \in [x_j^k, x_j^k + h]$ for $j \in \{1, \dots, n\}$ and $i \in \{2, 3, \dots, q\}$. Since the model functions m_i^k are quadratic functions, it holds that

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

for every $h \in \mathbb{R}^n$ with $\|h\| \leq \delta_k$ and for all indices $i \in \{2, 3, \dots, q\}$. Moreover, $\nabla_x m_i^k(x^k) = \nabla_x f_i(x^k)$ holds for all $i \in \{2, 3, \dots, q\}$ according to (10), which is given for the Taylor model (7). Using the triangle inequality it follows for every $x \in B_k$ that

$$|f_i(x) - m_i^k(x)| \leq \frac{1}{2} \|h\|^2 (\|\nabla_{xx} f_i(\xi_i^k)\| + \|\nabla_{xx} m_i^k(x^k)\|) \leq \delta_k^2 (\max\{\kappa_{\text{uhfi}}, \kappa_{\text{uhmi}}\} - 1)$$

with the constants κ_{uhfi} and κ_{uhmi} from Assumptions 4.1 and 4.3. Then the statement of the lemma holds for $\kappa_{\text{cndi}} := \max\{\kappa_{\text{uhfi}}, \kappa_{\text{uhmi}}\} - 1 > 0$. \square

For the expensive function, such a result is not provable in this way, and as in the standard trust-region approach, we assume validity.

Assumption 4.6. 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 $\kappa_{\text{cnd1}} > 0$ independent of k such that, for all $x \in B_k$, it holds that

$$|f_1(x) - m_1^k(x)| \leq \kappa_{\text{cnd1}} \delta_k^2.$$

The accuracy of the model is also reflected in the gradients. For the cheap functions m_i^k , $i \in \{2, 3, \dots, q\}$, 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 (7). 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 [9]. Since the constants are problem dependent, we give a short proof.

LEMMA 4.7. *Suppose Assumptions 4.1, 4.3, and 4.6 hold. Then there exists a constant $\kappa_{\text{eg}} > 0$ such that*

$$\|\nabla_x f_1(x^k) - \nabla_x m_1^k(x^k)\| \leq \kappa_{\text{eg}} \delta_k$$

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

Proof. Analogously to Lemma 4.5 and similarly to [9, Thm. 9.1.1], it follows by using Taylor’s theorem, (9), and the triangle inequality that

$$\begin{aligned} |(\nabla_x f_1(x^k) - \nabla_x m_1^k(x^k))^\top h| &\leq |f_1(x) - m_1^k(x)| + \frac{1}{2} \|h\|^2 \|\nabla_{xx} f_1(\xi^k) - \nabla_{xx} m_1^k(x^k)\| \\ &\leq \kappa_{\text{cnd1}} \delta_k^2 + \max\{\kappa_{\text{uhf}_1} - 1, \kappa_{\text{uhm}_1} - 1\} \delta_k^2 \end{aligned}$$

for every $h \in \mathbb{R}^n$ with $\|h\| \leq \delta_k$ and $x := x^k + h \in B_k$. It holds that $\xi_i^k \in [x_i^k, x_i^k + h]$ for $i \in \{1, \dots, n\}$ and the constants κ_{uhf_1} , κ_{uhm_1} , and κ_{cnd1} are from Assumptions 4.1, 4.3, and 4.6. Setting $h := \delta_k \frac{\nabla_x f_1(x^k) - \nabla_x m_1^k(x^k)}{\|\nabla_x f_1(x^k) - \nabla_x m_1^k(x^k)\|}$, the statement of the lemma follows with the constant $\kappa_{\text{eg}} := \kappa_{\text{cnd1}} + \max\{\kappa_{\text{uhf}_1}, \kappa_{\text{uhm}_1}\} - 1 > 0$. \square

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 is supposed to 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

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

for the model functions.

Assumption 4.8. There exists a constant $\kappa_\omega > 0$ such that, for every iteration $k \in \mathbb{N}$,

$$|\omega_m(x^k) - \omega(x^k)| \leq \kappa_\omega \omega_m(x^k)$$

holds.

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 MHT 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 both scalar and multiobjective versions of trust-region approaches [9, 11, 36, 44]. 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 [9, 11] 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 line search is used to approximate it. An analogous strategy transferred to the multiobjective case by using the function ω is used in [44]. In [36] the objectives are considered individually in addition to a scalarization, and therefore several trial points are computed. These 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 does not consider the functions individually or a surrogate scalarized problem, but computes a direction for decreasing the function values in the image space by the ideal point. Therefore, our 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 multiobjective descent direction. For this purpose, an assumption regarding the optimization problem (PS) given by $\min\{t \in \mathbb{R} \mid m^k(x^k) + t r^k - m^k(x) \in \mathbb{R}_+^q, x \in B_k\}$ is necessary, which is prepared by the following lemma. This lemma, together with its concluding assumption and the subsequent lemma, leads to the sufficient decrease condition in Lemma 4.13. This condition is then generalized in Assumption 4.14 and is essential for the convergence result.

LEMMA 4.9. *Suppose Assumption 4.3 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 $i = 1, 2, \dots, q$. In every iteration $k \in \mathbb{N}$ with x^k not Pareto critical for (MOPm), for every $i \in \{1, 2, \dots, q\}$, it holds that*

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

with $\beta_i^k := 1 + \|\nabla_{xx} m_i^k(x^k)\|$ and $\kappa_{\text{uhmi}} > 1$ from Assumption 4.3.

Proof. Let $i \in \{1, 2, \dots, q\}$ denote an index and $k \in \mathbb{N}$ denote an iteration with x^k not Pareto critical for $(MOPm)$. By Lemma 2.3 it follows that $\nabla_x m_i^k(x^k) \neq 0$. Consider the normed steepest descent direction for m_i^k in x^k defined by $d_{\text{sdi}} := -(\nabla_x m_i^k(x^k)) / (\|\nabla_x m_i^k(x^k)\|)$. From Taylor's theorem and the Cauchy–Schwarz inequality it follows that

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

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

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

The second term is obtained if $\delta_k < t_1$ holds. Thus, by estimating the second term of (12), we obtain

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

For the upper bound 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 that

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

This implies, with Assumption 4.3, that

$$r_i^k \leq \delta_k \|\nabla_x m_i^k(x^k)\| + \frac{1}{2} \delta_k^2 (\kappa_{\text{uhmi}} - 1)$$

for every $i \in \{1, 2, \dots, q\}$. \square

As stated in Remark 3.1 $r^k > 0$ holds 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 4.10. There exists a constant $\kappa_r \in (0, 1]$ such that, for every iteration $k \in \mathbb{N}$ with x^k not Pareto critical for $(MOPm)$, it holds that

$$(14) \quad \frac{\min_{i=1, \dots, q} r_i^k}{\max_{i=1, \dots, q} r_i^k} \geq \kappa_r.$$

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

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

a solution of (11). If x^k is not a Pareto critical point for $(MOPm)$, then, according to Lemma 2.5 applied to (11), d_ω is a descent direction for the multiobjective problem $(MOPm)$ at the current iteration point x^k . Therefore, it will also provide a descent in the trust region B_k . Furthermore, 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

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

and $\|d_\omega\| = 1$. 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 4.11. *Let $k \in \mathbb{N}$ be an iteration with x^k not Pareto critical for $(MOPm)$. Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be the quadratic function defined by $g(x) := \sum_{i=1}^q \alpha_i m_i^k(x)$ with constants $\alpha_i \geq 0$, $i \in \{1, 2, \dots, q\}$ from (16). Furthermore, define x_c by $g(x_c) := \min_{|t| \leq \delta_k} g(x^k + t d)$ with $d := -\nabla_x g(x^k) / \|\nabla_x g(x^k)\|$ and set $\beta_g^k := 1 + \|\nabla_{xx} g(x^k)\|$. Then it holds that*

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

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

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

for every $t \in \mathbb{R}$. Define $\beta_g^k := \|\nabla_{xx} g(x^k)\| + 1 > 0$. Together with calculations and estimations analogous to (12) and (13) in the proof of Lemma 4.9, the Cauchy–Schwarz inequality implies

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

which gives the inequality in the lemma. \square

REMARK 4.12. If x^k is Pareto critical for $(MOPm)$, no steepest descent for the function g in Lemma 4.11 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 MHT can be formulated.

LEMMA 4.13. Suppose Assumptions 4.3, 4.8, and 4.10 hold. Let x^{k+} be the solution of (PS) and let $\phi_m^k(x) = \max_{i=1,\dots,q} m_i^k(x)$ be defined as in (6). Furthermore, define $\beta_\phi^k := \max_{i=1,\dots,q} \|\nabla_{xx} m_i^k(x^k)\| + 1$. Then there exist 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

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

holds.

Proof. Let $(t^{k+}, x^{k+}) \in \mathbb{R}^{1+n}$ be the solution of the auxiliary problem (PS) given by $\min\{t \in \mathbb{R} \mid f(x^k) + t r^k - m^k(x) \in \mathbb{R}_+^q, x \in B_k\}$. First, let x^k be not Pareto critical for (MOPm). Then, according to Lemma 3.3 and Remark 3.1, it holds that $t^{k+} \in [-1, 0]$ and $r^k > 0$ as defined by $r_i^k = m_i^k(x^k) - \min_{x \in B_k} m_i^k(x)$ for $i \in \{1, 2, \dots, q\}$. Due to the constraints of (PS) it holds that

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

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

$$(19) \quad -t^{k+} = |t^{k+}| \leq \frac{m_i^k(x^k) - m_i^k(x^{k+})}{r_i^k} \leq \frac{\phi_m^k(x^k) - m_i^k(x^{k+})}{\min_{j=1,\dots,q} r_j^k}$$

for all $i \in \{1, 2, \dots, q\}$. Let $d_\omega \in \operatorname{argmin}_{\|d\| \leq 1} \max_{i=1,\dots,q} \nabla_x m_i^k(x^k)^\top d$ be a solution of the optimization problem from (11). Then, according to Lemma 2.5(ii) applied to (11), 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 (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 4.11 holds, and therefore (17) holds. Furthermore, it holds for β_g^k from Lemma 4.11 that

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

which, with (17) from Lemma 4.11, implies

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

Due to $x_c \in B_k$ and d_ω being a descent direction for (MOPm) (see Lemma 2.5(i) for (11)), there exists a scalar t such that (t, x_c) is feasible for (PS). According to [18], there exists a smallest scalar t_c such that (t_c, x_c) is feasible for (PS), and it follows that

$$(21) \quad -t_c = |t_c| = \min_{i=1,\dots,q} \frac{m_i^k(x^k) - m_i^k(x_c)}{r_i^k} \geq \frac{\min_{i=1,\dots,q} (m_i^k(x^k) - m_i^k(x_c))}{\max_{i=1,\dots,q} r_i^k}.$$

Because t^{k+} is the minimal value of (PS), $|t_c| \leq |t^{k+}|$ holds, which, together with (19) for the index i with $m_i^k(x^{k+}) = \phi_m^k(x^{k+})$, (21), and Assumption 4.10, implies

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

Since it holds that $\sum_{i=1}^q \alpha_i = 1$ and (t_c, x_c) is feasible for (PS) , it follows for the function g defined in Lemma 4.11 that

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

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

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

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

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

for every iteration $k \in \mathbb{N}$, with x^k not Pareto critical. If x^k is Pareto critical for $(MOPm)$, then $\omega_m(x^k) = 0$ holds and the solution of (11) is $d_\omega = 0$. Therefore, it holds that $\sum_{i=1}^q \alpha_i \nabla_x m_i^k(x^k) = 0$; see Lemma 2.5(ii). Since x^{k+} is the solution of (PS) , $\phi_m^k(x^k) - \phi_m^k(x^{k+}) \geq 0$ holds, and the above inequality is also satisfied.

Furthermore, according to Lemma 2.5(ii), $\omega_m(x^k) \leq \|\sum_{i=1}^q \alpha_i \nabla_x m_i^k(x^k)\|$ holds, and from Assumption 4.8 it follows that

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

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

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

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

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 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 4.14. There exists a constant $\kappa_\phi \in (0, 1)$ such that, for every iteration $k \in \mathbb{N}$,

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

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

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., [9, 44], such an assumption is used, and following this general approach we have also proved a motivation for the sufficient decrease assumption. Given Assumption 4.14, the remainder of the convergence analysis of MHT follows the scalar

trust-region methods [9, 11] closely. Consequently, it is also similar to the convergence analysis of the multiobjective trust-region method in [44], 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 MHT.

Remark 4.15. Because of Assumption 4.3, in every iteration $k \in \mathbb{N}$, for the constant β_k^ϕ from Assumption 4.14 it holds that

$$\beta_k^\phi = \max_{i=1,\dots,q} \|\nabla_{xx} m_i^k(x^k)\| + 1 \leq \max_{i=1,\dots,q} \kappa_{\text{uhmi}}.$$

For the trial point x^{k+} and the auxiliary functions $\phi = \max_{i=1,\dots,q} f_i(x)$ and $\phi_m^k = \max_{i=1,\dots,q} m_i^k(x)$ as given in (6), the following error bound, which depends on the trust-region radius, can be proved. It is crucial for the convergence proof and used either directly or indirectly in the following results.

LEMMA 4.16. *Suppose Assumptions 4.1, 4.3, and 4.6 hold. Then*

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

in every iteration $k \in \mathbb{N}$, with $\kappa_{\text{cnd}} := \max_{i=1,\dots,q} \kappa_{\text{cnd}i} > 0$ and the corresponding constants from Lemma 4.5 and Assumption 4.6.

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

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

with indices $i, j \in \{1, 2, \dots, q\}$, $i \neq j$. In case (i) it follows that $|\phi(x^{k+}) - \phi_m^k(x^{k+})| \leq \kappa_{\text{cnd}i} \delta_k^2$ due to $x^{k+} \in B_k$, Lemma 4.5, and Assumption 4.6. Now consider case (ii) and assume $f_i(x^{k+}) - m_j^k(x^{k+}) > 0$. By the definition of ϕ , Lemma 4.5, Assumption 4.6, and $x^{k+} \in B_k$ it holds that $|\phi(x^{k+}) - \phi_m^k(x^{k+})| \leq |f_i(x^{k+}) - m_i^k(x^{k+})| \leq \kappa_{\text{cnd}i} \delta_k^2$. Next assume $f_i(x^{k+}) - m_j^k(x^{k+}) < 0$. Then, again according to the definition of ϕ , Lemma 4.5, Assumption 4.6, and $x^{k+} \in B_k$, it holds that

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

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

In the following every point x^{k+1} is given by MHT as a result of iteration $k \in \mathbb{N}$. Either the trial point is accepted and $x^{k+1} = x^{k+}$ holds, or it is discarded and $x^{k+1} = x^k$. For further consideration, the iterations of MHT 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 $\rho_k \geq \eta_1$ holds, and the set of indices of all successful iterations is denoted by

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

Similarly, the set of indices

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

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 two lemmas illustrate the behavior of MHT for non-Pareto critical iteration points.

LEMMA 4.17. *Let $k \in \mathbb{N}$ be an iteration and suppose Assumptions 4.1, 4.3, 4.6, 4.8, 4.10, and 4.14 hold. Suppose furthermore that x^k is not Pareto critical for (MOP) and*

$$(24) \quad \delta_k \leq \frac{\kappa_\phi(1 - \eta_2)\omega(x^k)}{\kappa_e},$$

with $\kappa_e := \max_{i=1, \dots, q} \max\{\kappa_{\text{cndi}}, \kappa_{\text{uhmi}}\} > 0$ and $\kappa_\phi \in (0, 1)$ from Assumption 4.14. Then $k \in \mathcal{V}$ holds, 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 2.4, $\omega(x^k) > 0$ holds and, because $\eta_2, \kappa_\phi \in (0, 1)$, $\kappa_\phi(1 - \eta_2) < 1$ holds. By (24), the definition of κ_e , and Remark 4.15 it follows that

$$(25) \quad \delta_k \leq \frac{\kappa_\phi(1 - \eta_2)\omega(x^k)}{\kappa_e} < \frac{\omega(x^k)}{\kappa_e} \leq \frac{\omega(x^k)}{\max_{i=1, \dots, q} \kappa_{\text{uhmi}}} \leq \frac{\omega(x^k)}{\beta_k^\phi}.$$

According to Assumption 4.14, the following 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 $\rho_\phi^k = (\phi(x^k) - \phi(x^{k+})) / (\phi_m^k(x^k) - \phi_m^k(x^{k+}))$, the trial point acceptance quotient defined in (8). By the interpolation condition (9), $\phi_m^k(x^k) = \phi(x^k)$ holds, and from Lemma 4.16, the definition of κ_e , and (24) it follows that

$$|\rho_\phi^k - 1| = \left| \frac{\phi_m^k(x^{k+}) - \phi(x^{k+})}{\phi_m^k(x^k) - \phi_m^k(x^{k+})} \right| \leq \frac{\delta_k \max_{i=1, \dots, q} \kappa_{\text{cndi}}}{\kappa_\phi \omega(x^k)} \leq \frac{\delta_k \kappa_e}{\kappa_\phi \omega(x^k)} \leq 1 - \eta_2.$$

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

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 4.18. *Suppose Assumptions 4.1, 4.3, 4.6, 4.8, 4.10, and 4.14 hold. Suppose furthermore that there exists a constant $\kappa_{\text{lb}\omega} > 0$ such that $\omega(x^k) \geq \kappa_{\text{lb}\omega}$ holds for every iteration $k \in \mathbb{N}$. Then there exists a constant $\kappa_{\text{lb}\delta} > 0$ such that $\delta_k \geq \kappa_{\text{lb}\delta}$ holds for all $k \in \mathbb{N}$.*

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

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

with the constants $\gamma_1 \in (0, 1)$ from MHT and κ_ϕ, κ_e defined in Assumption 4.14 and Lemma 4.17. Let k_0 be the first iteration with $\delta_{k_0} < \kappa$. Then it holds that

$\delta_{k_0} < \delta_{k_0-1}$ and according to the trust-region update in Step 4 of MHT it holds that $\gamma_1 \delta_{k_0-1} \leq \delta_{k_0}$. These two inequalities imply

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

Because of the assumption on $\omega(x^{k_0-1})$ and Lemma 2.4, x^{k_0-1} is not Pareto critical for (MOP). Therefore, the preconditions of Lemma 4.17 are satisfied and it holds that $k_0 - 1 \in \mathcal{V}$ and $\delta_{k_0-1} \leq \delta_{k_0}$. This contradicts $\delta_{k_0} < \delta_{k_0-1}$ and thus the initial assumption. \square

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

LEMMA 4.19. *Suppose Assumptions 4.1, 4.3, 4.6, 4.8, 4.10, and 4.14 hold and MHT 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 $x^k = x^{k+1}$ holds 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 MHT 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 $\lim_{k \rightarrow \infty} \delta_k = 0$.

Assume that x^{k_0+1} is not a Pareto critical point for (MOP). Then Lemma 4.17 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). \square

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

LEMMA 4.20. *Suppose Assumptions 4.1, 4.3, 4.6, 4.8, 4.10, and 4.14 hold and MHT has infinitely many successful iterations $k \in \mathcal{S}$. Then it holds that*

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

Proof. Suppose it holds that $\liminf_{k \rightarrow \infty} \omega(x^k) \neq 0$. Then without loss of generality there exist 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 4.18, there exists a constant $\kappa_{lb\delta} > 0$ such that $\delta_k \geq \kappa_{lb\delta}$ holds for all $k \in \mathbb{N}$. From Remark 4.15 it follows that

$$\beta_k^\phi \leq \max_{i=1, \dots, q} \kappa_{uhmi} \leq \max_{i=1, \dots, q} \{\kappa_{uhmi}, \kappa_{cndi}\} = \kappa_e$$

for every iteration $k \in \mathbb{N}$ given the constants κ_{uhmi} , κ_{cndi} , and κ_e from Assumption 4.3, Lemma 4.5, Assumption 4.6, and Lemma 4.17. Consider a successful iteration $k \in \mathcal{S}$. Then it holds that $\rho_\phi^k \geq \eta_1$ and it follows from Assumption 4.14 that

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

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

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

with σ_k being the number of successful iterations up to iteration k . Since there are infinitely many such iterations in \mathcal{S} , $\lim_{k \rightarrow \infty} \sigma_k = \infty$ holds. 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 $\liminf_{k \rightarrow \infty} \omega(x^k) = 0$ holds. \square

The following theorem is the main result about convergence of MHT. It shows that the algorithm produces a sequence of iterates with ω converging to zero. This characterizes Pareto criticality, a necessary condition for weak efficiency; see Lemmas 2.3 and 2.4.

THEOREM 4.21. *Suppose Assumptions 4.1, 4.3, 4.6, 4.8, 4.10, and 4.14 hold. Then MHT produces a sequence of iterates $\{x^k\}$ with*

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

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

Proof. If MHT has only finitely many successful iterations $k \in \mathcal{S}$, then according to Lemma 4.19 the sequence of iterates $\{x^k\}$ converges to a Pareto critical point of (MOP). By Lemma 2.4 it follows that $\lim_{k \rightarrow \infty} \omega(x^k) = 0$.

Now consider the case when 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

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

for some constant $\varepsilon > 0$ and for all j . By Lemma 4.20 it follows that for all t_j there exists a first successful iteration $l_j > t_j$ satisfying $\omega(x^{l_j+1}) < \varepsilon$. Then there is another subsequence, indexed by $\{l_j\}$, such that

$$(27) \quad \omega(x^k) \geq \varepsilon \text{ for } t_j \leq k \leq l_j \text{ and } \omega(x^{l_j+1}) < \varepsilon.$$

Consider the subsequence whose indices are in $\mathcal{K} := \{k \in \mathcal{S} \mid \exists j \in \mathbb{N} : t_j \leq k \leq l_j\} \subseteq \mathcal{S}$, where t_j and l_j belong to the two subsequences defined above. For every successful iteration $k \in \mathcal{S}$ it holds that $\rho_\phi^k \geq \eta_1$ and $x^{k+1} = x^{k+}$. The definition of ρ_ϕ^k , the fact that $\mathcal{K} \subseteq \mathcal{S}$, Assumption 4.14, Remark 4.15, and (27) imply, for $k \in \mathcal{K}$, that

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

The sequence $\{\phi(x^k)\}$ is monotonically decreasing and bounded from below. Hence, $\{\phi(x^k)\}$ is convergent and $\lim_{k \rightarrow \infty} \phi(x^k) - \phi(x^{k+1}) = 0$ holds, which implies

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

Thus, the second term dominates the minimum in (28), and for $k \in \mathcal{K}$ sufficiently large it holds that

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

and consequently, for j sufficiently large,

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

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

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

Since ω is uniformly continuous due to Assumption 4.1 (see Remark 4.2), it follows that

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

This contradicts the definition of the sequences $\{t_j\}$ and $\{l_j\}$ in (27), which implies with (26) that $|\omega(x^{t_j}) - \omega(x^{l_j+1})| \geq \varepsilon$ holds.

Consequently, no subsequence of successful iterations satisfying (26) can exist. This fact can also be used to show that no subsequence $\{t_j\}$ of unsuccessful iterations with (26) can exist. Otherwise, a subsequence $\{z_j\} \subset \mathcal{S}$ can be constructed by defining $z_j := \min\{k \in \mathcal{S} \mid k > t_j\}$, which leads to a contradiction due to $\omega(x^{z_j}) = \omega(x^{t_j}) \geq 2\varepsilon$ for all $j \in \mathbb{N}$. By these findings, it can be proved that there exists no subsequence $\{t_j\}$ with $\omega(x^{t_j}) \geq \varepsilon > 0$ for a constant $\varepsilon > 0$ and for all $j \in \mathbb{N}$. Together with Lemma 4.20, this implies $\lim_{k \rightarrow \infty} \omega(x^k) = 0$.

Let \bar{x} be an accumulation point of the sequence $\{x^k\}$ produced by MHT and assume that it is not a Pareto critical point for (MOP). Then, according to Lemma 2.4, $\omega(\bar{x}) > 0$ holds. This is a contradiction of $\lim_{k \rightarrow \infty} \omega(x^k) = 0$, and hence every accumulation point of $\{x^k\}$ is Pareto critical for (MOP). \square

The convergence result can also be proved if all the objectives are expensive functions. Of course, assumptions such as Assumption 4.6 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. A suitable stopping criterion is needed for the implementation of MHT. Since one of the objectives regarding the evaluation time is expensive, it is reasonable to set a maximum number of permitted function evaluations and stop the algorithm when this number is reached. Thus, it is possible that the algorithm stops without having reached a Pareto critical point. Still, the algorithm ensures an improvement compared to the starting point due to the search strategy.

Furthermore, MHT is designed to reduce the trust-region radius whenever there is no sufficient decrease possible with the current model functions. Additionally,

Lemma 4.17 ensures 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 4.18, 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. Additionally, if the functions are sufficiently smooth, they can be approximated well by quadratic functions in small local areas. Therefore, the model functions are reliable when the trust-region radius is small and it is reasonable to stop the algorithm.

The Pascoletti–Serafini scalarization (PS) is used to compute the search direction in every iteration. According to Lemma 3.3, 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 MHT uses the general formulation from the literature; see, for example, [9, 44]. For the implementation it is specified as $\delta_{k+1} = \delta_k/2$ if $\rho_\phi^k < \eta_1$, $\delta_{k+1} = \delta_k$ if $\eta_1 \leq \rho_\phi^k < \eta_2$, and $\delta_{k+1} = 2\delta_k$ if $\rho_\phi^k \geq \eta_2$ holds.

5.3. User-given information. The search direction in MHT is defined by local ideal points. Instead of computing the individual minima of the model functions m_i^k , $i \in \{1, 2, \dots, q\}$, in every iteration $k \in \mathbb{N}$, a strictly lower bound is also sufficient. All the findings about MHT 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 $i \in \{1, 2, \dots, q\}$. 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 MHT by replacing the ideal point by this user-given desired point \tilde{p} .

5.4. Saving computation time. As outlined in section 3, the model for the expensive function f_1 is not updated in every iteration, but only when necessary. This is decided by the result of the trial point acceptance test. If $\rho_\phi^k < \eta_1$ holds, the model is not accurate enough and is recomputed. Otherwise, the old model is reused in the next iteration. If the model is updated, interpolation points need to be computed. Not all of them are recomputed, but every point situated in the current trust region and not violating the quality criterion of poisedness is reused.

5.5. Constrained optimization problems. MHT 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 MHT can still be solved quickly even with box constraints. To include such constraints in the computation of the trial points it is possible to use projection methods as suggested in basic trust-region methods; see [9]. For the implementation we do not use this approach, but include the box constraints in the Pascoletti–Serafini scalarization. Still, these auxiliary optimization problems are easy to solve.

6. Numerical results. MHT was implemented in MATLAB (version 2017a) with the modifications and stopping criteria described in section 5 and the parameters $\eta_1 = 0.001$, $\eta_2 = 0.9$. It was tested for several multiobjective problems with two or three objective functions. All the problems considered were test problems and did not involve an actual expensive function. For evaluating the results, one of the

functions was declared expensive and the number of function evaluations for this function was counted. The advantage of this is that the correct solutions are known for the test instances. Furthermore, the test problems were both self-chosen (“T”) and taken from the literature [12, 27, 23, 24, 37]. Table 1 gives an overview of all 78 test examples used, showing information about the dimension of the domain (n), the constraints, the convexity of the problem, the geometry of the Pareto front, and the set of all nondominated points (convex, nonconvex, disconnected). For scalable problems, n is denoted by “var”, that is, $n = 2, 3, 4, 5$, and for Jin1 and T4, additionally $n = 10, 20, 30, 40, 50$.

TABLE 1
Test problems.

Name	n	Constr.	Convex	PF	Name	n	Constr.	Convex	PF
Biobjective test problems ($q = 2$)					lovison2	2	box, unc.	nonc.	nonc.
BK1	2	box	conv.	conv.	lovison3	2	box, unc.	nonc.	conv.
CL1	4	box	nonc.	conv.	lovison4	2	box, unc.	nonc.	nonc.
Deb41	2	box	conv.	conv.	MOP1	1	unc.	conv.	conv.
Deb53	2	unc.	nonc.	nonc.	Schaffer2	1	box	nonc.	discon.
Deb513	2	box	nonc.	discon.	T1	2	unc.	conv.	conv.
Deb521b	2	box	nonc.	nonc.	T2	2	unc.	nonc.	nonc.
DG01	1	box	nonc.	conv.	T3	2	box	conv.	conv.
DTLZ1	2	box	nonc.	conv.	T4	var	box	conv.	conv.
ex005	2	box	nonc.	nonc.	T6	2	box	conv.	conv.
Far1	2	box	nonc.	nonc.	T7	2	box	conv.	conv.
FF	var	box	nonc.	nonc.	T10	3	box	conv.	conv.
Fonseca	2	box	nonc.	nonc.	VU1	2	box	nonc.	conv.
IM1	2	box	nonc.	conv.	VU2	2	box	conv.	conv.
Jin1	var	box	conv.	conv.	ZDT1	4	box	nonc.	conv.
Jin2	var	box	nonc.	conv.	ZDT2	4	box	nonc.	nonc.
Jin3	var	box	nonc.	nonc.	ZDT3	4	box	nonc.	discon.
Jin4	var	box	nonc.	nonc.	ZDT4	2	box	nonc.	conv.
JOS3	3	box	conv.	conv.	ZDT6	4	box	nonc.	nonc.
Kursawe	3	box	nonc.	discon.	Test problems with $q = 3$ objective functions				
Laumanns	2	box	conv.	conv.	FES2	10	box	nonc.	nonc.
LE1	2	box	nonc.	nonc.	IKK1	2	box	conv.	conv.
Lis	2	box	nonc.	nonc.	T8	3	box	conv.	conv.
lovison1	2	box, unc.	conv.	conv.	ZLT1	4	box	conv.	conv.

To compare the results of MHT we use two methods. On the one hand, since MHT computes only one Pareto critical point and does not approximate the set of efficient points, we use the weighted sum approach with equal weights and apply EFOS (expensive function optimization solver) [42] to it with the predefined standard parameters. It is a solution method for expensive, simulation-based scalar optimization problems also using the trust-region approach. As the stopping criterion, a criticality measure using the gradients of the model functions is applied in conjunction with a validity criterion for the models. For convex multiobjective 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 taken into account when comparing the results.

On the other hand, and to circumvent the disadvantages of the weighted sum approach, the multiobjective method DMS [12] 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 also offers the option to compute only one efficient point. We used the latter option with the predefined standard parameters varying the maximum number of function evaluations. As a stopping criterion DMS uses a maximum number of function evaluations given by the user and the step size for the search step. If the step size is lower than the predefined value, DMS stops. Of course, the way the algorithms are implemented influences the performance for a given test problem. In the currently available implementation of EFOS, internal errors often occur.

The three algorithms were tested on the 78 test problems listed in Table 1 with a set of randomly generated but fixed starting points. One test instance is defined as one test problem with one fixed starting point. In total 802 test instances are considered: 348 convex instances and 454 nonconvex instances.

To classify the test runs as successful or unsuccessful, 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 declared expensive function is counted until the algorithm stops. In Figure 2 a performance profile for the full range is depicted on the left and zoomed in for up to 500 function evaluations on the right.

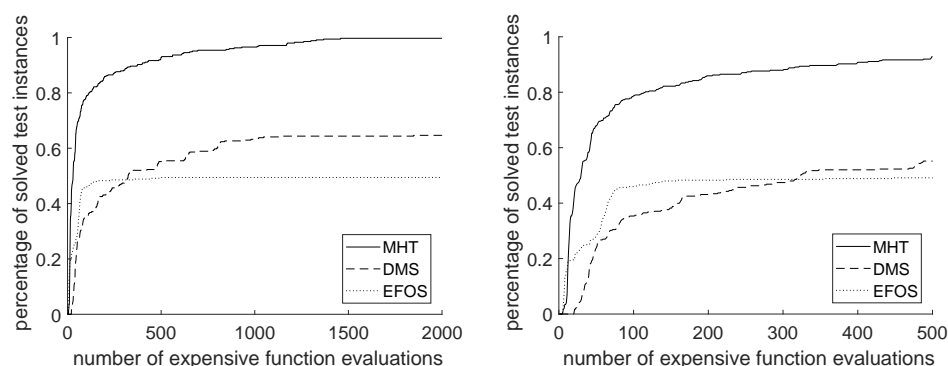


FIG. 2. Performance profiles for MHT, DMS, and EFOS for 348 convex instances.

Figure 2 shows that MHT is able to solve all 348 convex test instances within at most 1459 expensive function evaluations. This high number is due to the high-dimensional test instances included. If we consider only test instances up to dimension 10, all convex instances are solved by MHT after 667 expensive function evaluations. With the same number of function evaluations DMS solves 64.93%, and EFOS 62.31%, of the convex test problems up to dimension 10. Figure 2 also shows that if up to 480 function evaluations are allowed for the expensive function, DMS and EFOS behave similarly. With further function evaluations DMS is capable of solving further test instances, whereas EFOS stagnates and cannot solve more instances. Figure 2 shows that, in general, MHT needs fewer function evaluations than EFOS and DMS to solve the convex test problems.

Since MHT is a local method it will generally produce locally efficient points. Thus, for nonconvex problems where local and global efficiency are in general not identical, the distance to the Pareto front is not reliable for classifying a test instance as solved and therefore a performance profile like that above is not appropriate for the nonconvex problems. However, using this performance measure, MHT solves 87% of the nonconvex instances after 443 expensive function evaluations. With the same number of evaluations DMS is capable of solving 87.67%, and EFOS 50.88%, of the nonconvex instances.

Looking into the specific test results, we found that EFOS often computes individual minima of the objective functions or the exact tradeoff between the two functions. This is not surprising and a reason for this is the formulation by the weighted sum for the nonconvex problems. By contrast, DMS and MHT are capable of finding points from different areas of the Pareto front. However, a more important aspect is the number of function evaluations, and MHT often needs significantly fewer function

evaluations than EFOS and DMS. In general, the test results show that MHT can save computational effort.

Exemplarily, we present test results for two selected test problems. To compare the results, we use the number of function evaluations for the expensive function. The first test problem is the self-chosen convex problem (T1) defined by

$$(T1) \quad \min_{x \in \mathbb{R}^n} \begin{pmatrix} f_1(x) \\ f_2(x) \end{pmatrix} = \min_{x \in \mathbb{R}^n} \begin{pmatrix} \frac{1}{2}x_1^2 + x_2^2 - 10x_1 - 100 \\ x_1^2 + \frac{1}{2}x_2^2 - 10x_2 - 100 \end{pmatrix},$$

with f_1 declared the expensive function. For all 10 test instances, i.e., (T1) with 10 different starting points, MHT and EFOS produce efficient points, while DMS produces efficient points for most instances. MHT needs 11–14 function evaluations for f_1 , which is significantly less than EFOS (13–78) and DMS (41–131). Figure 3 shows a result for MHT on the left and for EFOS on the right. The image set is represented by scattered gray points; the starting point and—for MHT—the iteration points are marked in black. The solution is marked in dark gray and all other points that are evaluated during the iterations are marked as unfilled circles.

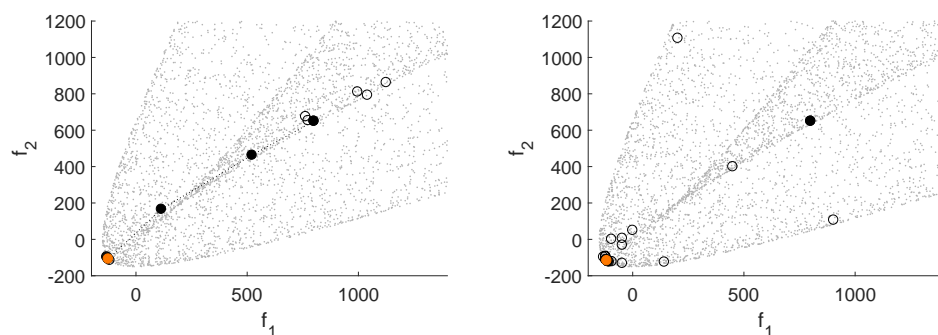


FIG. 3. Test run for MHT (left) and EFOS (right) for test problem (T1).

This example illustrates the effect of the local search strategy used in MHT. For EFOS, the evaluated points are spread over the image space and more evaluations are needed. By contrast, the local trust regions in MHT have the effect that the model is only built on points near to the current iteration point. Furthermore, this figure shows that in MHT the model was not updated very often, but could be reused in several iterations since it was also reliable in larger and shifted trust regions.

To illustrate that MHT can also work well for nonconvex problems, even though it is only a local method, we consider the test problem (Lis) [27, 6] defined by

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

with f_1 declared the expensive function. For all 12 starting points of this test example, MHT produces Pareto critical points, mostly even efficient points. Since (Lis) is a nonconvex problem, the weighted sum approach with EFOS gives, as expected, points close to an individual minimum as solutions. MHT and DMS need similar numbers of function evaluations (MHT: 19–68, DMS: 41–70) and always less than EFOS (77–219).

Both MHT and DMS generate different nondominated points. In Figure 4 the result of one test run for MHT is shown on the left, and one for DMS on the right. It shows that for MHT the iteration points move, along with the interpolation points in local areas around them, to a nondominated point. DMS computes the individual minimum of function f_1 within 67 function evaluations. MHT needs slightly fewer function evaluations (60) to generate an efficient point.

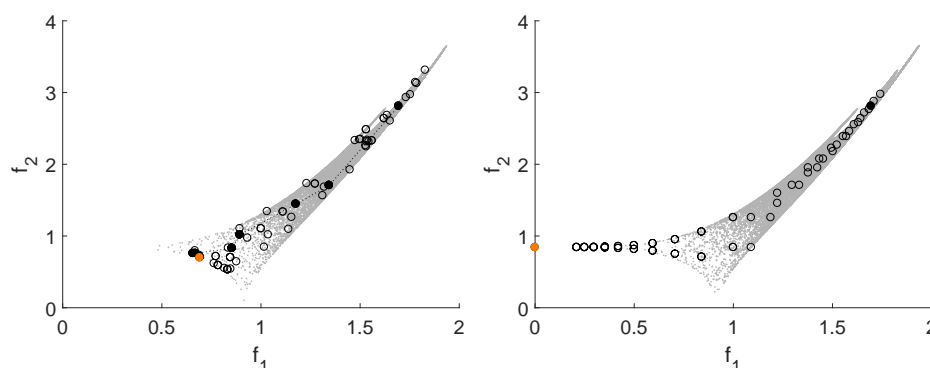


FIG. 4. Test run for MHT (left) and DMS (right) for test problem (Lis).

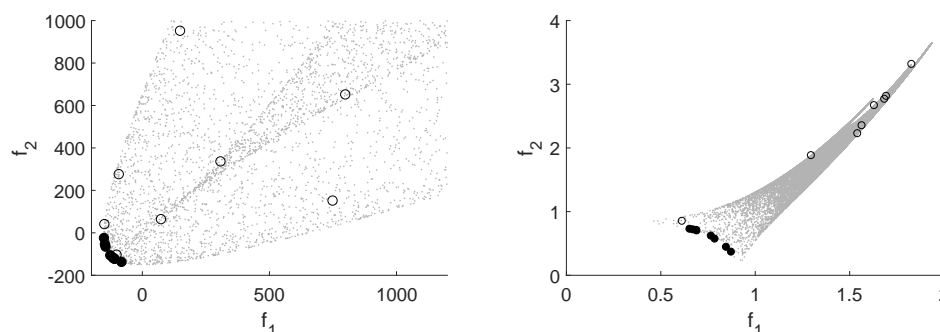


FIG. 5. Multistart approach for MHT for test problems (T1) (left) and (Lis) (right).

We have considered test examples with dimensions up to 50 and, as expected, with rising dimension, more function evaluations are needed. As outlined in subsection 3.1, computing one quadratic model for the expensive function needs $(n+1)(n+2)/2$ interpolation points, which is problematic for higher dimensions. For such problems we suggest using linear interpolation models that need only $n+1$ interpolation points. The convergence proof transfers.

Furthermore, the algorithm as introduced in section 3 generates one Pareto critical point in one run. Due to the search strategy and as the numerical tests confirm, different starting points generally result in different Pareto critical points. This is shown exemplarily in Figure 5 for the test problems (T1) and (Lis). The unfilled points are randomly chosen starting points and the nondominated points obtained

are marked in black. Further test results with a detailed discussion and performance profiles for all problems considered, including the nonconvex ones, can be found in [43].

7. Conclusions. The multiobjective method MHT for heterogeneous functions considers in general one expensive function. It can also be modified to consider only expensive functions; the convergence proof is transferable. Even then, the presented approach is new due to the search direction. One has to be aware of the fact that the method presented is only a local method. Therefore, for highly nonlinear problems one can only expect to find locally efficient points.

In future work modifications for MHT will be developed to exploit the heterogeneity of the objective functions further. With regard to practical applications, the user-given information already described to some extent in subsection 5.3 will be considered in future work. We are currently working on further modifications to make more use of the cheap functions. It is possible to replace the model functions for the cheap functions by the functions themselves. The convergence proof is transferable. Another approach for modifying MHT consists of including some constraints for the cheap functions or using multistarts with strategically chosen starting points to generate approximations of the set of efficient points. Initial theoretical examinations and numerical tests are promising. For all these purposes, heuristic strategies based on the method presented in this paper are currently being explored and examined.

Acknowledgments. The authors wish to thank Prof. Dr. J. Fliege (University of Southampton) for valuable discussions, as well as the two anonymous referees for their helpful and constructive remarks.

REFERENCES

- [1] K. ANAGNOSTOPOULOS AND G. MAMANIS, *Multiobjective evolutionary algorithms for complex portfolio optimization problems*, Comput. Manag. Sci., 8 (2011), pp. 259–279.
- [2] C. AUDET AND J. DENNIS, JR., *Mesh adaptive direct search algorithms for constrained optimization*, SIAM J. Optim., 17 (2006), pp. 188–217.
- [3] C. AUDET, G. SAVARD, AND W. ZGAL, *Multiobjective optimization through a series of single-objective formulations*, SIAM J. Optim., 19 (2008), pp. 188–210.
- [4] D. BIENSTOCK AND A. MICHALKA, *Polynomial solvability of variants of the trust-region subproblem*, in Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, PA, 2013, pp. 380–390.
- [5] G. A. CARRIZO, P. A. LOTITO, AND M. C. MACIEL, *Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem*, Math. Program., 159 (2016), pp. 339–369.
- [6] C. COELLO, D. V. VELDHIJZEN, AND G. LAMONT, *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer Academic/Plenum, 2002.
- [7] A. CONN, N. GOULD, AND PH. TOINT, *Global convergence of a class of trust region algorithms for optimization with simple bounds*, SIAM J. Numer. Anal., 25 (1988), pp. 433–460.
- [8] A. CONN, N. GOULD, AND PH. 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 (1989), pp. 764–767.
- [9] A. CONN, N. GOULD, AND P. TOINT, *Trust-Region Methods*, MOS-SIAM Ser. Optim., SIAM, Philadelphia, PA, 2000.
- [10] 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 (2009), pp. 387–415.
- [11] A. CONN, K. SCHEINBERG, AND L. VICENTE, *Introduction to Derivative-Free Optimization*, MOS-SIAM Ser. Optim., SIAM, Philadelphia, PA, 2009.
- [12] A. CUSTODIO, J. MADEIRA, A. VAZ, AND L. VICENTE, *Direct multisearch for multiobjective optimization*, SIAM J. Optim., 21 (2011), pp. 1109–1140.
- [13] 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 (1998),

- pp. 631–657.
- [14] K. DEB, *Multi-objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, Chichester, 2001.
 - [15] L. DRUMMOND AND B. SVAITER, *A steepest descent method for vector optimization*, J. Comput. Appl. Math., 175 (2005), pp. 395–414.
 - [16] G. EICHFELDER, *Adaptive Scalarization Methods in Multiobjective Optimization*, Springer, Berlin, 2008.
 - [17] G. EICHFELDER, *An adaptive scalarization method in multiobjective optimization*, SIAM J. Optim., 19 (2009), pp. 1694–1718.
 - [18] G. EICHFELDER, *Variable Ordering Structures in Vector Optimization*, Vector Optimization, Springer, Berlin, 2014.
 - [19] B. FLACH, *Rekonstruktion mit Vorinformation für die Niedrigstdosis-CT-Fluoroskopie in der Interventionellen Radiologie*, Ph.D. thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2014.
 - [20] J. FLIEGE, L. DRUMMOND, AND B. SVAITER, *Newton's method for multiobjective optimization*, SIAM J. Optim., 20 (2009), pp. 602–626.
 - [21] J. FLIEGE AND B. SVAITER, *Steepest descent methods for multicriteria optimization*, Math. Methods Oper. Res., 51 (2000), pp. 479–494.
 - [22] J. FLIEGE AND A. VAZ, *A method for constrained multiobjective optimization based on SQP techniques*, SIAM J. Optim., 26 (2016), pp. 2091–2119.
 - [23] C. FONSECA AND P. FLEMING, *An overview of evolutionary algorithms in multiobjective optimization*, Evol. Comput., 3 (1995), pp. 1–16.
 - [24] C. FONSECA AND P. FLEMING, *Multiobjective optimization and multiple constraint handling with evolutionary algorithms. I. A unified formulation*, IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, 28 (1998), pp. 26–37.
 - [25] C. GERTH AND P. WEIDNER, *Nonconvex separation theorems and some applications in vector optimization*, J. Optim. Theory Appl., 67 (1990), pp. 297–320.
 - [26] S. GRECO, K. KLAMROTH, J. KNOWLES, AND G. RUDOLPH, *Understanding complexity in multiobjective optimization*, Dagstuhl Rep., 5(1) (2015), pp. 96–163.
 - [27] S. HUBAND, P. HINGSTON, L. BARONE, AND L. WHILE, *A review of multiobjective test problems and a scalable test problem toolkit*, IEEE Trans. Evol. Comput., 10 (2006), pp. 477–506.
 - [28] J. JAHN, *Vector Optimization*, Springer, Berlin, 2011.
 - [29] D. JONES, M. SCHONLAU, AND W. WELCH, *Efficient global optimization of expensive black-box functions*, J. Global Optim., 13 (1998), pp. 455–492.
 - [30] J. KNOWLES AND H. NAKAYAMA, *Meta-modeling in multiobjective optimization*, in Multiobjective Optimization: Interactive and Evolutionary Approaches, J. Branke, K. Deb, K. Miettinen, and R. Słowiński, eds., Lecture Notes in Comput. Sci. 5252, Springer, Cham, 2008, pp. 258–284.
 - [31] T. KOLDA, R. LEWIS, AND V. TORCZON, *Optimization by direct search: New perspectives on some classical and modern methods*, SIAM Rev., 45 (2003), pp. 385–482.
 - [32] 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 (2005), pp. 987–1001.
 - [33] J. J. MORÉ, *Recent developments in algorithms and software for trust region methods*, in Mathematical Programming: The State of the Art, A. Bachem, M. Groetschel, and B. Korte, eds., Springer, Berlin, 1983, pp. 258–287.
 - [34] A. PASCOLETTI AND P. SERAFINI, *Scalarizing vector optimization problems*, J. Optim. Theory Appl., 42 (1984), pp. 499–524.
 - [35] 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, ACM Press, New York, 2018, pp. 195–196.
 - [36] J.-H. RYU AND S. KIM, *A derivative-free trust-region method for biobjective optimization*, SIAM J. Optim., 24 (2014), pp. 334–362.
 - [37] J. SCHAFFER, *Multiple objective optimization with vector evaluated genetic algorithms*, in Proceedings of the 1st International Conference on Genetic Algorithms, L. Erlbaum Associates, Hillsdale, NJ, 1985, pp. 93–100.
 - [38] T. STEWART, O. BANDTE, H. BRAUN, N. CHAKRABORTI, M. EHRGOTT, M. GÖBELT, Y. JIN, H. NAKAYAMA, S. POLES AND D. DI STEFANO, *Real-world applications of multiobjective optimization*, in Multiobjective Optimization: Interactive and Evolutionary Approaches, J. Branke, K. Deb, K. Miettinen, and R. Słowiński, eds., Lecture Notes in Comput. Sci. 5252, Springer, Cham, 2008, pp. 285–327.
 - [39] F. STRAUSS, J.-A. DESIDERI, R. DUVIGNEAU, AND V. HEUVELINE, *Multiobjective Optimization*

- In Hydrodynamic Stability Control*, Research report RR-6608, INRIA, 2008.
- [40] 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, IEEE Press, Piscataway, NJ, 2007, pp. 532–539.
 - [41] D. TERZIJSKA, M. PORCELLI, AND G. EICHFELDER, *Multi-objective optimization in the Lorentz force velocimetry framework*, in Book of Digests and Program, 13th International Workshop on Optimization and Inverse Problems in Electromagnetism, Delft, 2014, pp. 81–82.
 - [42] A. THEKALE, *Trust-Region Methods for Simulation Based Nonlinear Optimization*, Ph.D. thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, 2011.
 - [43] J. THOMANN AND G. EICHFELDER, *Numerical Results for the Multi-objective Trust Region Algorithm MHT*, preprint, Technische Universität Ilmenau, http://www.optimization-online.org/DB_HTML/2018/08/6768.html, 2018.
 - [44] 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 (2014), pp. 865–889.
 - [45] S. WILD AND C. SHOEMAKER, *Global convergence of radial basis function trust region derivative-free algorithms*, SIAM J. Optim., 21 (2011), pp. 761–781.