

# Newton's Method for Multiobjective Optimization

J. Fliege\*  
L.M. Graña Drummond†  
B.F. Svaiter‡

February 23, 2008

## Abstract

We propose an extension of Newton's Method for unconstrained multi-objective optimization (multicriteria optimization). The method does not scalarize the original vector optimization problem, i.e. we do not make use of any of the classical techniques that transform a multiobjective problem into a family of standard optimization problems. Neither ordering information nor weighting factors for the different objective functions need to be known. The objective functions are assumed to be twice continuously differentiable. Under these hypotheses, the method, as in the classical case, is locally superlinear convergent to optimal points. Again as in the scalar case, if the second derivatives are Lipschitz continuous, the rate of convergence is quadratic. This is the first time that a method for an optimization problem with an objective function with a partially ordered vector space as a codomain is considered and convergence results of this order are provided.

Our convergence analysis uses a Kantorovich-like technique. As a byproduct, existence of optima is obtained under semi-local assumptions.

**Keywords:** Multicriteria optimization, multi-objective programming, Pareto points, Newton's method.

## 1 Introduction

In multicriteria optimization, several objective functions have to be minimized simultaneously. Usually, no single point will minimize all given objective functions at once, and so the concept of optimality has to be replaced by the concept of *Pareto-optimality* or *efficiency*. A point is called Pareto-optimal or

---

\*J.Flidge@soton.ac.uk, School of Mathematics, University of Southampton, Southampton SO17 1BJ, U.K.

†lmgd@facc.ufrj.br, Faculdade de Ciências Contábeis, Universidade Federal do Rio de Janeiro - FACC-UFRJ, Pasteur 250, Praia Vermelha, Rio de Janeiro, RJ, 22.290-240, Brazil

‡benar@impa.br, Instituto de Matemática Pura e Aplicada - IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ, 22460-320, Brasil

efficient, if there does not exist a different point with the same or smaller objective function values, such that there is a decrease in at least one objective function value. Applications for this type of problem can be found in engineering [14, 28] (especially truss optimization [8, 29, 40], design [27, 20, 38], space exploration [34, 41]), statistics [5], management science [15, 2, 22, 35, 42, 1], environmental analysis [31, 16, 17], cancer treatment planning [25], etc.

One of the main solution strategies for multicriteria optimization problems is the scalarization approach, first introduced by Geoffrion [21]. Here, one or several parameterized single-objective (i. e., classical) optimization problems are solved, resulting in a corresponding number of Pareto-optimal points. The disadvantage of this approach is that the choice of the parameters is not known in advance, leaving the modeler and the decision-maker with the burden of choosing them. Only recently, adaptive scalarization techniques, where scalarization parameters are chosen automatically during the course of the algorithm such that a certain quality of approximation is maintained have been proposed [6, 23, 13, 18]. Still other techniques, working only in the bicriteria case, can be viewed as choosing a fixed grid in a particular parameter space [7, 8]

Multicriteria optimization algorithms that do not scalarize have recently been developed (see, e. g., [4, 3] for an overview on the subject). Some of these techniques are extensions of scalar optimization algorithms (notably the steepest descent algorithm [19, 37, 11, 12] with at most linear convergence), while others borrow heavily from ideas developed in heuristic optimization [30, 33]. For the latter, no convergence proofs are known, and empirical results show that convergence generally is, as in the scalar case, quite slow [43]. Other parameter-free multicriteria optimization techniques use an ordering of the different criteria, i. e., an ordering of importance of the components of the objective function vector. In this case, the ordering has to be prespecified. Moreover, the corresponding optimization process is usually augmented by an interactive procedure [32], adding an additional burden to the task of the decision maker.

In this paper, we propose a parameter-free optimization method for computing a point satisfying a certain (natural) first-order necessary condition for multicriteria optimization. Neither ordering information nor weighting factors for the different objective functions is assumed to be known. The rate of convergence is at least superlinear for twice continuously differentiable functions and quadratic in case the second derivatives are Lipschitz continuous. In this respect, Newton's method for the scalar case is exactly mimicked.

The outline of this paper is as follows. Section 2 establishes the problem considered and the necessary notation. Section 3 introduces a first-order optimality condition for multiobjective optimization and derives a direction search program based on this. Such program is equivalent to an optimization problem with linear objective and convex quadratic constraints. In Section 3, we establish the algorithm under consideration, Sections 5 and 6 contain the convergence results: superlinear convergence is discussed in Section 5, while quadratic convergence is discussed in Section 6. Numerical results presented in Section 7 show the applicability of our method.

## 2 Notation

Denote by  $\mathbb{R}$  the set of real numbers, by  $\mathbb{R}_+$  the set of non-negative real numbers, and by  $\mathbb{R}_{++}$  the set of strictly positive real numbers. Assume that  $U \subset \mathbb{R}^n$  is an open set and

$$F : U \longrightarrow \mathbb{R}^m$$

is a given function. The problem is to find an *efficient* point or *Pareto optimum* of  $F$ , i. e., a point  $x^* \in U$  such that

$$\nexists y \in U, \quad F(y) \leq F(x^*) \text{ and } F(y) \neq F(x^*),$$

where the inequality sign  $\leq$  between vectors is to be understood in a componentwise sense. In effect, we are employing the partial order induced by  $\mathbb{R}_+^m = \mathbb{R}_+ \times \cdots \times \mathbb{R}_+$  (the non-negative orthant or Paretian cone of  $\mathbb{R}^m$ ) defined by

$$F(y) \leq F(z) \iff F(z) - F(y) \in \mathbb{R}_+^m,$$

and we are searching for minimal points induced by such partial order. A point  $x^*$  is *weakly efficient* or *weak Pareto optimum* if

$$\nexists y \in U, \quad F(y) < F(x^*),$$

where the vector strict inequality  $F(y) < F(x^*)$  is to be understood componentwise too. This relation is induced by  $\mathbb{R}_{++}^m$ , the interior of the Paretian cone ( $F(y) < F(z)$  if, and only if,  $F(z) - F(y) \in \mathbb{R}_{++}^m$ ).

A point  $x^* \in U$  is *locally efficient* (respectively *locally weakly efficient*) if there is a neighborhood  $V \subseteq U$  of  $x^*$  such that the point  $x^*$  is efficient (respectively weakly efficient) for  $F$  restricted to  $V$ .

Locally efficient points are also called *local Pareto optimal*, and locally weakly efficient points are also called *local weak Pareto optimal*. Note that if  $U$  is convex and  $F$  is  $\mathbb{R}_+^m$ -convex (i. e., if  $F$  is componentwise-convex), then each local Pareto optimal point is globally Pareto optimal. Clearly, every locally efficient point is locally weakly efficient.

Throughout the paper, unless explicitly mentioned, we will assume that

$$F \in C^2(U, \mathbb{R}^m),$$

i. e.,  $F$  is twice continuously differentiable on  $U$ . For  $x \in U$ , denote by  $DF(x) \in \mathbb{R}^{m \times n}$  the Jacobian of the function  $F$  at  $x$ , by  $\nabla F_j(x) \in \mathbb{R}^n$  the gradient of the function  $F_j$  at  $x$  and by  $\nabla^2 F_j(x) \in \mathbb{R}^{n \times n}$  the Hessian of  $F_j$  at  $x$ . The range, or image space, of a matrix  $M \in \mathbb{R}^{m \times n}$  will be denoted by  $R(M)$  and  $I \in \mathbb{R}^{n \times n}$  will denote the unit matrix. For two matrices  $A, B \in \mathbb{R}^{n \times n}$ ,  $B \leq A$  ( $B < A$ ) will mean  $A - B$  positive semidefinite (definite). Unless explicitly mentioned, we will also assume that

$$\nabla^2 F_j(x) > 0, \quad \forall x \in U, \quad j = 1, \dots, m,$$

which means that  $\nabla^2 F_j(x)$  is positive definite for all  $x \in U$  and  $j = 1, \dots, m$ . Under this assumption,  $F$  is  $\mathbb{R}^m$ -convex on each convex subset of  $U$ .

In what follows, the Euclidean norm in  $\mathbb{R}^n$  will be denoted by  $\|\cdot\|$ , and  $B[x, r]$  denotes the ball of radius  $r$  with center  $x \in \mathbb{R}^n$ . We will use the same notation  $\|\cdot\|$  for the induced operator norms on the corresponding matrix spaces.

### 3 The Newton Method

sec:direction

We start by introducing a necessary (but in general not sufficient) condition for Pareto optimality. We say that a point  $\bar{x} \in U$  is *critical* (or *stationary*) for  $F$  if

$$R(DF(\bar{x})) \cap (-\mathbb{R}_{++}^m) = \emptyset. \quad (1)$$

station

This notion of criticality has already been used in [19] to define a parameter-free steepest descent algorithm for multiobjective optimization. Clearly, if  $\bar{x}$  is critical for  $F$ , then for all  $s \in \mathbb{R}^n$  there exists  $j_0 = j_0(s) \in \{1, \dots, m\}$  such that

$$\nabla F_{j_0}(\bar{x})^T s \geq 0. \quad (2)$$

nondescent

Note that if  $x \in U$  is noncritical, then there exists  $s \in \mathbb{R}^m$  such that  $\nabla F_j(x)^T s < 0$  for all  $j = 1, \dots, m$ . As  $F$  is continuously differentiable,

$$\lim_{t \rightarrow 0} \frac{F_j(x + ts) - F_j(x)}{t} = \nabla F_j(x)^T s < 0, \quad j = 1, 2, \dots, m.$$

So  $s$  is a *descent direction* for  $F$  at  $x$ , i. e., there exists  $t_0 > 0$  such that

$$F(x + ts) < F(x) \quad \text{for all } t \in (0, t_0]. \quad (3)$$

descent\_direct

Finally, observe that for  $m = 1$  (scalar optimization), (1) reduces to the classical “gradient-equal-zero” condition.

Efficiency and criticality are related as follows.

t1

**Theorem 1.** *Assume that  $F \in C^1(U, \mathbb{R}^m)$ .*

1. *If  $\bar{x}$  is locally weak Pareto-optimal, then  $\bar{x}$  is a critical point for  $F$ .*
2. *If  $U$  is convex,  $F$  is  $\mathbb{R}^m$ -convex and  $\bar{x} \in U$  is critical for  $F$ , then  $\bar{x}$  is weak Pareto-optimal.*
3. *If  $U$  is convex,  $F \in C^2(U, \mathbb{R}^m)$ ,  $\nabla^2 F_j(x) > 0$  for all  $j \in \{1, \dots, m\}$  and all  $x \in U$ , and if  $\bar{x} \in U$  is critical for  $F$ , then  $\bar{x}$  is Pareto-optimal.*

*Proof.* Assume that  $\bar{x}$  is weakly efficient. If  $\bar{x}$  is non-stationary, then (1) does not hold and so there exist  $t_0 > 0$  and  $s \in \mathbb{R}^n$  for which (3) holds, in contradiction with the weak efficiency of  $\bar{x}$ . So item 1 is true.

To prove item 2, take any  $x \in U$ . Since  $\bar{x}$  is critical, (2) holds for  $s = x - \bar{x}$  and some  $j_0$ . Using now the convexity of  $F_{j_0}$ , we have

$$F_{j_0}(x) \geq F_{j_0}(\bar{x}) + \nabla F_{j_0}(\bar{x})^T (x - \bar{x}) \geq F_{j_0}(\bar{x}) \quad (4)$$

aux1

and so,  $\bar{x}$  is weakly efficient for  $F$ .

To prove item 3, take again any  $x \in U$ . Since we are assuming that  $\bar{x}$  is critical, by the same reasoning of item 2, there exists some  $j_0$  for which (4) holds. Note that  $F_{j_0}$  is strictly convex. Therefore, if  $x \neq \bar{x}$ , the first inequality in (4) is strict and so  $\bar{x}$  is efficient.  $\square$

Note that even when  $F$  is  $\mathbb{R}_+^m$ -convex (as in item 2), criticality does not imply Pareto optimality. Consider, for instance, the case  $n = 1$ ,  $m = 2$ ,  $U = \mathbb{R}$  and  $F(t) = (1, t)$ . In this example, every  $t \in \mathbb{R}$  is critical but there are no Pareto optima for  $F$ . Only under some assumption of *strict* convexity, as in item 3, the set of Pareto optima coincides with the set of stationary points.

We now proceed by defining the Newton direction for the multiobjective problem under consideration. As in the classical one-criterion case, the Newton direction will be a solution to a suitably defined problem involving quadratic approximations of the objective functions  $F_j$ . Moreover, again as in the scalar case, in a critical point, the Newton step will be  $0 \in \mathbb{R}^n$ .

For  $x \in U$ , we define  $s(x)$ , the *Newton direction* at  $x$ , as the optimal solution of

$$\begin{cases} \min & \max_{j=1, \dots, m} \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s. \\ \text{s.t.} & s \in \mathbb{R}^n \end{cases} \quad (5) \quad \boxed{\text{eq:prob.newton}}$$

First of all, observe that problem (5) has always a unique minimizer, since the functions  $\nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s$  are strongly convex for  $j = 1, \dots, m$ . Also note that for  $m = 1$ , the direction  $s(x)$  is the “classical” Newton direction for scalar optimization.

The optimal value of problem (5) will be denoted by  $\theta(x)$ . Hence,

$$\theta(x) = \inf_{s \in \mathbb{R}^n} \max_{j=1, \dots, m} \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s, \quad (6) \quad \boxed{\text{eq:def.medida}}$$

and

$$s(x) = \arg \min_{s \in \mathbb{R}^n} \max_{j=1, \dots, m} \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s. \quad (7) \quad \boxed{\text{eq:def.step}}$$

Here, we are approximating

$$\max_{j=1, \dots, m} F_j(x + s) - F_j(x)$$

by the maximum of the local quadratic models at  $x$  of each  $F_j$ . Although (5) is a non-smooth problem, it can be framed as a convex quadratic optimization problem and so, it can be effectively solved. Indeed, (5) is equivalent to

$$\begin{cases} \min & g(t, s) = t \\ \text{s.t.} & \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s - t \leq 0 \quad (1 \leq j \leq m) \\ & (t, s) \in \mathbb{R} \times \mathbb{R}^n \end{cases} \quad (8) \quad \boxed{\text{prob_equiv}}$$

The Lagrangian of this problem is

$$L((t, s), \lambda) = t + \sum_{j=1}^m \lambda_j \left( \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s - t \right). \quad (9) \quad \boxed{\text{lagr}}$$

Direct calculation of the Karush-Kuhn-Tucker conditions yields

$$\sum_{j=1}^m \lambda_j = 1, \quad (10) \quad \boxed{\text{lag\_t}}$$

$$\sum_{j=1}^m \lambda_j \left( \nabla F_j(x) + \nabla^2 F_j(x) s \right) = 0, \quad (11) \quad \boxed{\text{lag\_s}}$$

$$\nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s \leq t \quad (1 \leq j \leq m), \quad (12) \quad \boxed{\text{prim\_feas}}$$

$$\lambda_j \geq 0 \quad (1 \leq j \leq m), \quad (13) \quad \boxed{\text{dual\_feas}}$$

$$\lambda_j \left( \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s - t \right) = 0 \quad (1 \leq j \leq m). \quad (14) \quad \boxed{\text{complement}}$$

Problem (8) has a unique solution,  $(\theta(x), s(x))$ . As this is a convex problem and has a Slater point (e.g.,  $(1, 0)$ ), there exists a KKT multiplier  $\lambda = \lambda(x)$ , which, together with  $s = s(x)$  and  $t = \theta(x)$ , satisfies conditions (10)–(14). In particular, from (11) we obtain

$$s(x) = - \left[ \sum_{j=1}^m \lambda_j(x) \nabla^2 F_j(x) \right]^{-1} \sum_{j=1}^m \lambda_j(x) \nabla F_j(x). \quad (15) \quad \boxed{\text{Newton\_dir}}$$

So the Newton direction defined in this paper is a Newton direction for a standard scalar optimization problem, implicitly induced by weighting the given objective functions by the (non-negative) *a priori* unknown KKT multipliers. As a consequence, the standard weighting factors [21], well known in multi-objective programming, do show up in our approach, albeit *a posteriori* and implicitly. In particular, it is not necessary to fix such weights in advance; every point  $x \in U$  defines such weights by way of the KKT multipliers in the corresponding direction search program.

Existence of the KKT multipliers for the convex problem (8) implies that there is no duality gap, and so

$$\begin{aligned} \theta(x) &= \sup_{\lambda \geq 0} \inf_{s \in \mathbb{R}^n} L((t, s), \lambda) \\ &= \sup_{\substack{\lambda \geq 0 \\ \sum \lambda_j = 1}} \inf_{s \in \mathbb{R}^n} \sum_{j=1}^m \lambda_j \left( \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s \right). \end{aligned} \quad (16) \quad \boxed{\text{no\_gap1,5}}$$

Let us now study some properties of function  $\theta$  and analyze its relation with  $s(x)$  and stationarity of  $x$ .

**bas1** **Lemma 2.** *Under our general assumptions, we have:*

1. For any  $x \in U$ ,  $\theta(x) \leq 0$ .
2. The following conditions are equivalent.

(a) The point  $x$  is noncritical.

(b)  $\theta(x) < 0$ ,

(c)  $s(x) \neq 0$ ,

3. The function  $\theta : U \rightarrow \mathbb{R}$ , given by (6), is continuous.

*Proof.* Note that, by (6),  $\theta(x) \leq \max_{j=1, \dots, m} \nabla F_j(x)^T 0 + \frac{1}{2} 0^T \nabla^2 F_j(x) 0 = 0$ , so item 1 holds.

Let us now prove the equivalences of item 2. First, assume that (a) holds; that is to say,  $R(DF(x)) \cap (-\mathbb{R}_{++}^m) \neq \emptyset$ , which in turn means that there exists  $\tilde{s} \in \mathbb{R}^n$  such that  $\nabla F_j(x)^T \tilde{s} < 0$  for all  $j = 1, \dots, m$ . So, using (6), for all  $t > 0$ , we have

$$\begin{aligned} \theta(x) &\leq \max_{j=1, \dots, m} \nabla F_j(x)^T t \tilde{s} + \frac{1}{2} t \tilde{s}^T \nabla^2 F_j(x) t \tilde{s} \\ &= t \max_{j=1, \dots, m} \nabla F_j(x)^T \tilde{s} + \frac{t}{2} \tilde{s}^T \nabla^2 F_j(x) \tilde{s}. \end{aligned}$$

Therefore, for  $t > 0$  small enough the right hand side of the above inequality is negative and (b) holds.

To prove that (b) implies (c), recall that  $\theta(x)$  is the optimal value of problem (5) and so, being negative, the solution of that problem cannot be  $s(x) = 0$ .

Finally, let us see that (c) implies (a). Since all  $\nabla^2 F_j(x)$  are positive definite matrices, by virtue of (6)-(7), for all  $j \in \{1, \dots, m\}$ , we have that

$$\nabla F_j(x)^T s(x) < \nabla F_j(x)^T s(x) + \frac{1}{2} s(x)^T \nabla^2 F_j(x)^T s(x) = \theta(x) \leq 0,$$

where the last inequality is a consequence of item 1. Hence,  $DF(x)s(x) \in (-\mathbb{R}_{++}^m)$ , and  $x$  is noncritical, i.e., (a) holds.

We now prove item 3. It suffices to show continuity of  $\theta$  in a fixed but arbitrary compact set  $W \subset U$ . First observe that, in view of item 1, for any  $y \in U$ , we have

$$s(y) \nabla^2 F_j(x)^T s(y) \leq -\nabla F_j(y)^T s(y), \quad \text{for all } j = 1, \dots, m. \quad (17)$$

desig1

Since  $F$  is twice continuously differentiable and all Hessians are positive definite everywhere, its' eigenvalues are uniformly bounded away from zero on  $W$ , so there exists  $K$  and  $\hat{\lambda} > 0$ , such that

$$K = \max_{y \in W, j=1, \dots, m} \|\nabla F_j(y)\|$$

and

$$\hat{\lambda} = \min_{x \in W, \|u\|=1} u^T \nabla^2 F_j(x) u \quad (1 \leq j \leq m).$$

So, combining (17) with the two above equations and using Cauchy-Schwartz inequality, we conclude that

$$\hat{\lambda} \|s(y)\|^2 \leq \|\nabla F_j(y)\| \|s(y)\| \leq K \|s(y)\|,$$

for all  $y \in W$  and all  $j \in \{1, \dots, m\}$ . Therefore

$$\|s(y)\| \leq \frac{1}{\bar{\lambda}} K \text{ for all } y \in W, \quad (18) \quad \boxed{\text{newton\_bound}}$$

i. e.,  $s(y)$ , the Newton directions are uniformly bounded on  $W$ .

For  $x \in W$  and  $j \in \{1, \dots, m\}$ , define

$$\begin{aligned} \varphi_{x,j} : W &\rightarrow \mathbb{R}, \\ z &\mapsto \nabla F_j(z)^T s(x) + \frac{1}{2} s(x)^T \nabla^2 F_j(z) s(x). \end{aligned}$$

The family  $\{\varphi_{x,j}\}_{x \in W, j=1, \dots, m}$  is equicontinuous. Therefore, the family

$$\{\Phi_x = \max_{j=1, \dots, m} \varphi_{x,j}\}_{x \in W}$$

is also equicontinuous.

Take  $\varepsilon > 0$ ; there exists  $\delta > 0$  such that

$$\forall y, z \in W, \|y - z\| < \delta \implies |\Phi_x(y) - \Phi_x(z)| < \varepsilon \quad \forall x \in W.$$

Hence, for  $\|y - z\| < \delta$ ,

$$\begin{aligned} \theta(z) &\leq \nabla F_j(z)^T s(y) + \frac{1}{2} s(y)^T \nabla^2 F_j(z) s(y) \\ &= \Phi_y(z) \\ &\leq \Phi_y(y) + |\Phi_y(z) - \Phi_y(y)| \\ &< \theta(y) + \varepsilon, \end{aligned}$$

i. e.,  $\theta(z) - \theta(y) < \varepsilon$ . Interchanging the roles of  $z$  and  $y$ , we conclude that  $|\theta(z) - \theta(y)| < \varepsilon$ .  $\square$

For the scalar case ( $m = 1$ )  $F : U \rightarrow \mathbb{R}$ , at a non-stationary point  $x \in U$ , the classical Armijo-rule for the Newton search direction  $s(x)$  is

$$F(x + ts(x)) \leq F(x) + \beta t s(x)^T \nabla F(x),$$

with  $\beta \in (0, 1)$ . To accept a full Newton step close to a local optimum where  $\nabla^2 F > 0$ , one must choose  $\beta \in (0, 1/2)$ , see [10]. Note that in this setting ( $m = 1$ ),

$$\theta(x) = \frac{1}{2} s(x)^T \nabla F(x).$$

So, we can rewrite the Armijo rule as

$$F(x + ts(x)) \leq F(x) + \sigma t \theta(x),$$

with the choice  $\sigma = 2\beta \in (0, 1)$  allowing full Newton steps to be accepted close to a local optimum where  $\nabla^2 F > 0$ . The above inequality, interpreted componentwise, will be our criterion for accepting a step in the multiobjective Newton direction.

cor:stepdefined

**Corollary 3.** *If  $x \in U$  is a noncritical point for  $F$ , then for any  $0 < \sigma < 1$  there exists  $\bar{t} \in (0, 1]$  such that*

$$x + ts(x) \in U \quad \text{and} \quad F_j(x + ts(x)) \leq F_j(x) + \sigma t \theta(x)$$

hold for all  $t \in [0, \bar{t}]$  and  $j \in \{1, \dots, m\}$ .

*Proof.* Since  $U$  is an open set and  $x \in U$ , there exists  $0 < t_1 \leq 1$  such that  $x + ts(x) \in U$  for  $t \in [0, t_1]$ . Therefore, for  $t \in [0, t_1]$

$$F_j(x + ts(x)) = F_j(x) + t \nabla F_j(x)^T s(x) + o_j(t), \quad j = 1, \dots, m$$

where  $\lim_{t \rightarrow 0^+} o_j(t)/t = 0$ . As  $\nabla F_j(x)^T s(x) \leq \theta(x)$ , we conclude that for  $t \in [0, t_1]$ ,  $j = 1, \dots, m$

$$\begin{aligned} F_j(x + ts(x)) &\leq F_j(x) + t \theta(x) + o_j(t) \\ &= F_j(x) + t \sigma \theta(x) + t \left[ (1 - \sigma) \theta(x) + \frac{o_j(t)}{t} \right] \end{aligned}$$

Now observe that, since  $x$  is noncritical,  $\theta(x) < 0$  (Lemma 2, item 2) and so, for  $t \in [0, t_1]$  small enough, the last term at the right hand side of the above equations is non-positive.  $\square$

Now we sketch the Newton Algorithm for Multicriteria. At each step, at a nonstationary point, we minimize the maximum of all local models as in (5) to obtain the Newton step (7), which is a descent direction. After that, the step length is determined by means of an Armijo-like rule (see Corollary 3) coupled with a backtracking procedure. Under suitable local assumptions, full Newton steps are always accepted and the generated sequence converges superlinear (or quadratically) to a local solution.

Formally, the algorithm for finding a Pareto point is the following.

#### Newton Algorithm for Multicriteria

1. (*Initialization*) Choose  $x_0 \in U$ ,  $0 < \sigma < 1$ , set  $k := 0$  and define  $J = \{1/2^n \mid n = 0, 1, 2, \dots\}$ .
2. (*Main loop*)
  - (a) Solve the direction search program (5) to obtain  $s(x_k)$  and  $\theta(x_k)$  as in (7) and (6).
  - (b) If  $\theta(x_k) = 0$ , then stop. Else, proceed to the line search, step 2. (c).
  - (c) (*Line Search*) Choose  $t_k$  as the largest  $t \in J$  such that

$$x_k + ts(x_k) \in U,$$

$$F_j(x_k + ts(x_k)) \leq F_j(x_k) + \sigma t \theta(x_k), \quad j = 1, \dots, m.$$

- (d) (*Update*) Define

$$x_{k+1} = x_k + t_k s(x_k)$$

and set  $k := k + 1$ . Go to step 2. (a).

Welldefinedness of the algorithm follows directly from Corollary 3. Note that this is an  $\mathbb{R}^m$ -decreasing method, i. e., the objective values always keep on going downhill in the componentwise partial order. Indeed, for  $x_k$  nonstationary for  $F$ , due to item 2 of Lemma 2, we have that  $\theta(x_k) < 0$ , so from (c)-(d) of the main loop we see that  $F(x_{k+1}) < F(x_k)$ .

## 4 Auxiliary Technical Results

sec:auxil-thechn-results

In order to analyze convergence properties of the Newton Method for Multiobjective Optimization, we need some technical results, which are presented in the sequel.

A basic feature of the Newton Method for scalar minimization and equations is to use respectively quadratic and linear approximations. In the next lemma we give estimations on the error of approximating  $F$ ,  $\nabla F_j$  by its quadratic and respectively linear local models. Its proof is provided for the sake of completeness.

lm:error

**Lemma 4.** *Suppose that  $V \subset U$  is a convex set. Let  $\varepsilon > 0$  and  $\delta > 0$  be such that for any  $x, y \in V$ , with  $\|y - x\| < \delta$ , then*

$$\|\nabla^2 F_j(y) - \nabla^2 F_j(x)\| < \varepsilon, \quad (j = 1, \dots, m) \quad (19) \quad \text{hyp\_lem3}$$

follows.

*Under this assumption, for any  $x, y \in V$  such that  $\|y - x\| < \delta$  we have that*

$$\|\nabla F_j(y) - [\nabla F_j(x) + \nabla^2 F_j(x)(y - x)]\| < \varepsilon \|y - x\|, \quad (20) \quad \text{res1}$$

and

$$\left\| F_j(y) - \left[ F_j(x) + \nabla F_j(x)^T (y - x) + \frac{1}{2} (y - x)^T \nabla^2 F_j(x) (y - x) \right] \right\| < \frac{\varepsilon}{2} \|y - x\|^2 \quad (21) \quad \text{res2}$$

hold for  $j = 1, \dots, m$ .

*If  $\nabla^2 F_j$  is Lipschitz continuous on  $V$  with constant  $L$  for  $j = 1, \dots, m$ , then*

$$\|\nabla F_j(y) - [\nabla F_j(x) + \nabla^2 F_j(x)(y - x)]\| < \frac{L}{2} \|y - x\|^2 \quad (22) \quad \text{res3}$$

holds for  $j = 1, \dots, m$ .

*Proof.* Since  $D\nabla F_j(\cdot) = \nabla^2 F_j(\cdot)$ , we have

$$\nabla F_j(y) - \nabla F_j(x) - \nabla^2 F_j(x)(y - x) = \int_0^1 [\nabla^2 F_j(x + t(y - x)) - \nabla^2 F_j(x)]^T (y - x) dt.$$

Therefore, since  $\|x + t(y - x) - x\| < t\delta$  for  $0 < t < 1$ ,

$$\begin{aligned} \|\nabla F_j(y) - \nabla F_j(x) - \nabla^2 F_j(x)(y - x)\| &\leq \int_0^1 \|\nabla^2 F_j(x + t(y - x)) - \nabla^2 F_j(x)\| \|y - x\| dt \\ &< \varepsilon \|y - x\|, \end{aligned} \quad (23) \quad \text{partial\_res1}$$

that is to say (20) is true. Note that, under the Lipschitz continuity assumption, the integrand in (23) is bounded by  $tL\|y-x\|^2$  and so (22) is true.

Now we prove (21). We know that there exists  $\zeta \in (0, 1)$  such that

$$F_j(y) - F_j(x) - \nabla F_j(x)^T(y-x) = \frac{1}{2}(y-x)^T \nabla^2 F_j(x + \zeta(y-x))(y-x). \quad (24) \quad \boxed{\text{aux}}$$

Subtracting  $(y-x)^T \nabla^2 F_j(x)(y-x)$  on both sides of (24), using Cauchy-Schwartz inequality and the fact that  $\|x + \zeta(y-x) - x\| < \delta$ , we get

$$|F_j(y) - F_j(x) - \nabla F_j(x)^T(y-x) - (y-x)^T \nabla^2 F_j(x)(y-x)| < \frac{\varepsilon}{2}\|y-x\|^2,$$

and the proof is complete.  $\square$

The following auxiliary results provide estimates on the length of the Newton direction  $s(x)$  and on  $\theta(x)$ , the optimal value of the direction search program (5).

$\boxed{\text{theta.s}}$  **Lemma 5.** *Take  $x \in U$  and let  $a, b \in \mathbb{R}$  be such that  $0 < a \leq b$ . If*

$$aI \leq \nabla^2 F_j(x) \leq bI, \quad j = 1, \dots, m,$$

then

$$\frac{a}{2}\|s(x)\|^2 \leq |\theta(x)| \leq \frac{b}{2}\|s(x)\|^2.$$

*Proof.* Recall that  $\lambda_j(x)$  is the Lagrange multiplier for constraint  $j$  of problem (8) ( $j = 1, \dots, m$ ), fulfilling (10)–(15), and define

$$H := \sum_{j=1}^m \lambda_j(x) \nabla^2 F_j(x), \quad v := \sum_{j=1}^m \lambda_j(x) \nabla F_j(x). \quad (25) \quad \boxed{\text{matrixH}}$$

From the assumptions and (15), we have

$$aI \leq H \leq bI, \quad s(x) = -H^{-1}v,$$

and

$$\begin{aligned} \theta(x) &= v^T s(x) + \frac{1}{2} s(x)^T H s(x) \\ &= (-H s(x))^T s(x) + \frac{1}{2} s(x)^T H s(x) \\ &= -\frac{1}{2} s(x)^T H s(x) \end{aligned} \quad (26) \quad \boxed{\text{equal}}$$

Combining (25) and (26), the conclusion follows, since, by item 1 of Lemma 2,  $\theta(x) \leq 0$ .  $\square$

**thetax2**

**Lemma 6.** Take  $x \in U$ ,  $0 < a$ . If

$$aI \leq \nabla^2 F_j(x), \quad j = 1, \dots, m$$

then

$$|\theta(x)| \leq \frac{1}{2a} \left\| \sum_{j=1}^m \lambda_j \nabla F_j(x) \right\|^2$$

for all  $\lambda_j \geq 0$  ( $j = 1, \dots, m$ ), with  $\sum_{j=1}^m \lambda_j = 1$ .

*Proof.* Let  $\lambda_j \geq 0$  ( $j = 1, \dots, m$ ) with  $\sum_{j=1}^m \lambda_j = 1$  be given. Define  $w := \sum_{j=1}^m \lambda_j \nabla F_j(x)$ . Then, by (16),

$$\theta(x) \geq \inf_s \sum_{j=1}^m \lambda_j \left( \nabla F_j(x)^T s + \frac{1}{2} s^T \nabla^2 F_j(x) s \right) \quad (27)$$

$$= \inf_s \left( w^T s + \sum_{j=1}^m \lambda_j \frac{1}{2} s^T \nabla^2 F_j(x) s \right)$$

$$\geq \inf_s \left( w^T s + \frac{1}{2} a \|s\|^2 \right) = -\frac{\|w\|^2}{2a}, \quad (28)$$

**theta\_grad**

because  $s \mapsto w^T s + \frac{1}{2} a \|s\|^2$  is a strongly convex function, so its minimum is achieved at the unique point where its gradient vanishes, i. e., at  $s$  such that  $w + as = 0$ . The conclusion follows from (28), since, by item 1 of Lemma 2,  $\theta(x) \leq 0$ . □

## 5 Superlinear Convergence

**sec:convergence**

The main theoretical results of this paper are presented in this section. First we give sufficient conditions for local superlinear convergence.

**th:basic**

**Theorem 7.** Denote by  $\{x_k\}_k$  a sequence generated by the algorithm proposed in Section 3. Suppose that  $V \subset U$ ,  $0 < \sigma < 1$ ,  $a, b, r, \delta, \varepsilon > 0$  and

- (a)  $aI \leq \nabla^2 F_j(x) \leq bI$  for all  $x \in V$ ,  $j = 1, \dots, m$ ,
- (b)  $\|\nabla^2 F_j(x) - \nabla^2 F_j(y)\| \leq \varepsilon$  for all  $x, y \in V$  with  $\|x - y\| < \delta$ ,
- (c)  $\varepsilon/a \leq 1 - \sigma$ ,
- (d)  $B[x_0, r] \subset V$ ,
- (e)  $\|s(x_0)\| \leq \min\{\delta, r(1 - \varepsilon/a)\}$ .

Then, for all  $k$ , we have that:

1.  $\|x_k - x_0\| \leq \|s(x_0)\| \frac{1 - (\varepsilon/a)^k}{1 - \varepsilon/a}$
2.  $\|s(x_k)\| \leq \|s(x_0)\| (\varepsilon/a)^k$
3.  $t_k = 1$
4.  $\|s(x_{k+1})\| \leq \|s(x_k)\| (\varepsilon/a)$ .

Moreover, the sequence  $\{x_k\}_k$  converges to some locally Pareto-optimal point  $\bar{x} \in \mathbb{R}^n$  with

$$\|\bar{x} - x_0\| \leq \frac{\|s(x_0)\|}{1 - \varepsilon/a} \leq r. \quad (29) \quad \boxed{\text{eq:item4}}$$

The convergence rate of  $\{x_k\}_k$  is superlinear.

*Proof.* First we show that if items 1 and 2 hold for some  $k$ , then items 3 and 4 also hold for that  $k$ .

From the triangle inequality, item 1 and item 2, we have

$$\|(x_k + s(x_k)) - x_0\| \leq \|s(x_0)\| \frac{1 - (\varepsilon/a)^{k+1}}{1 - \varepsilon/a} \quad (30) \quad \boxed{\text{provis0}}$$

Hence, from assumptions (e) and (c), we get

$$x_k, x_k + s(x_k) \in B[x_0, r] \quad \text{and} \quad \|(x_k + s(x_k)) - x_k\| \leq \delta. \quad (31) \quad \boxed{\text{provis}}$$

Using now assumption (b) and Lemma 4 we conclude that, for  $j = 1, \dots, m$ ,

$$\begin{aligned} F_j(x_k + s(x_k)) &\leq F_j(x_k) + \nabla F_j(x_k)^T s(x_k) + \frac{1}{2} s(x_k)^T \nabla^2 F_j(x_k) s(x_k) \\ &\quad + \frac{\varepsilon}{2} \|s(x_k)\|^2 \\ &\leq F_j(x_k) + \theta(x_k) + \frac{\varepsilon}{2} \|s(x_k)\|^2 \\ &= F_j(x_k) + \sigma \theta(x_k) + (1 - \sigma) \theta(x_k) + \frac{\varepsilon}{2} \|s(x_k)\|^2 \end{aligned}$$

Since, from Lemma 2,  $\theta(x_k) \leq 0$ , using assumption (a) and Lemma 5 we obtain

$$(1 - \sigma) \theta(x_k) + \frac{\varepsilon}{2} \|s(x_k)\|^2 \leq (\varepsilon - a(1 - \sigma)) \frac{\|s(x_k)\|^2}{2} \leq 0,$$

where the last equality follows from assumption (c). Combining the above inequalities we conclude that the Armijo-like rule (condition (c) of the iterative step of the method) holds with no need of a single backtracking, i. e., item 3 holds:

$$t_k = 1.$$

Therefore, from (31) we get

$$x_{k+1} = x_k + s(x_k), \quad x_k, x_{k+1} \in B[x_0, r], \quad \|x_{k+1} - x_k\| \leq \delta. \quad (32) \quad \boxed{\text{provis\_induct}}$$

Now define

$$\tilde{v}_{k+1} = \sum_{j=1}^m \lambda_j(x_k) \nabla F_j(x_{k+1}). \quad (33) \quad \boxed{\text{def:vk}}$$

Using (10), (13) and Lemma 6, we have

$$|\theta(x_{k+1})| \leq \frac{1}{2a} \|\tilde{v}_{k+1}\|^2. \quad (34) \quad \boxed{\text{eq:usefull}}$$

To estimate  $\|\tilde{v}_{k+1}\|$ , define for  $x \in U$

$$G_k(x) := \sum_{j=1}^m \lambda_j(x_k) F_j(x), \quad (35) \quad \boxed{\text{def:Gk}}$$

where  $\lambda_j(x_k)$  are the KKT multipliers from (5) for  $x = x_k$ . Observe that

$$\tilde{v}_{k+1} = \nabla G_k(x_{k+1}) = \nabla G_k(x_k + s(x_k)).$$

Direct calculation yields

$$\nabla G_k(x) = \sum_{j=1}^m \lambda_j(x_k) \nabla F_j(x) \quad \text{and} \quad \nabla^2 G_k(x) = \sum_{j=1}^m \lambda_j(x_k) \nabla^2 F_j(x),$$

so, according to (15),

$$s(x_k) = - [\nabla^2 G_k(x_k)]^{-1} \nabla G_k(x_k). \quad (36) \quad \boxed{\text{eq:s.and.Gk}}$$

Note that by assumption (b),

$$\|\nabla^2 G_k(y) - \nabla^2 G_k(x)\| \leq \varepsilon \quad \forall x, y \in V, \quad \|y - x\| \leq \delta,$$

which allows us to apply Lemma 4. With this, the estimate

$$\|\nabla G_k(x_k + s(x_k)) - [\nabla G_k(x_k) + \nabla^2 G_k(x_k) s(x_k)]\| \leq \varepsilon \|s(x_k)\|. \quad (37) \quad \boxed{\text{eq:Gkestimate}}$$

follows. Since (36) is equivalent to  $\nabla G_k(x_k) + \nabla^2 G_k(x_k) s(x_k) = 0$ , from (37) we arrive at

$$\|\tilde{v}_{k+1}\| = \|\nabla G_k(x_{k+1})\| \leq \varepsilon \|s(x_k)\|, \quad (38) \quad \boxed{\text{eq:whatisgoingonhere}}$$

The combination of (38) with (34) leads us to

$$|\theta(x_{k+1})| \leq \frac{(\varepsilon \|s(x_k)\|)^2}{2a},$$

and so, from assumption (a) and an application of Lemma 5, we arrive at

$$\frac{a}{2} \|s(x_{k+1})\|^2 \leq \frac{(\varepsilon \|s(x_k)\|)^2}{2a}.$$

Therefore,

$$\|s(x_{k+1})\| \leq \|s(x_k)\| \frac{\varepsilon}{a},$$

and item 4 also holds.

Now we will prove by induction that items 1 and 2 hold for all  $k$ . For  $k = 0$ , using assumption (c), we see that they hold trivially. If item 1 and 2 hold for some  $k$ , then, as we already saw, items 3 and 4 also hold for such  $k$  and this implies that  $x_{k+1} = x_k + s(x_k)$ , so item 1 for  $k + 1$  is true, by virtue of (30). Item 2 for  $k + 1$  follows from item 4, which, as we have already seen, is true under our inductive hypotheses.

As items 1 and 2 hold for all  $k$ , items 3 and 4 also hold for all  $k$ . In particular, (32) holds for all  $k$ . So, for all  $k$  and  $j$  with  $k > j$ ,

$$\|x_k - x_j\| \leq \sum_{i=j+1}^k \|x_i - x_{i-1}\| = \sum_{i=j+1}^k \|s(x_{i-1})\|.$$

Whence, from item 2,

$$\|x_k - x_j\| \leq \sum_{i=j}^{k-1} \|s(x_i)\| \leq \|s(x_0)\| \sum_{i=j}^{k-1} (\varepsilon/a)^i.$$

From assumption (c),  $\varepsilon/a \in (0, 1)$ . Thus,  $\{x_k\}$  is a Cauchy sequence, and there exists  $\bar{x} \in \mathbb{R}^n$ , such that

$$\lim_{k \rightarrow \infty} x_k = \bar{x}. \quad (39) \quad \boxed{\text{limit}}$$

Moreover,  $\{\|s(x_k)\|\}$  converges to 0, so, from assumption (a) and Lemma 5,  $\theta(x_k) \rightarrow 0$  for  $k \rightarrow \infty$ . Therefore, combining (39) with the continuity of  $\theta$  (item 3 of Lemma 2), we see that  $\theta(\bar{x}) = 0$ . So, from item 2 of Lemma 2,  $\bar{x}$  is stationary for  $F$  and, in view of item 3 of Theorem 1, we conclude that  $\bar{x}$  is locally Pareto-optimal.

To prove superlinear convergence, define

$$r_k = \|s(x_0)\| \frac{(\varepsilon/a)^k}{1 - \varepsilon/a}, \quad \delta_k = \|s(x_0)\| (\varepsilon/a)^k, \quad k = 0, 1, \dots$$

Using the triangle inequality, item 1, assumptions (e) and (d), we conclude that

$$B[x_k, r_k] \subset B[x_0, r] \subset V. \quad (40) \quad \boxed{\text{auxiliar}}$$

Take any  $\tau > 0$  and define

$$\hat{\varepsilon} := \min \left\{ a \frac{\tau}{1 + 2\tau}, \varepsilon \right\}.$$

For  $k$  large enough

$$\|\nabla^2 F_j(x) - \nabla^2 F_j(y)\| \leq \hat{\varepsilon} \quad \forall x, y \in B[x_k, r_k], \|x - y\| \leq \delta_k. \quad (41) \quad \boxed{\text{auxb}}$$

Hence, assumptions (a)–(e) are satisfied for  $\hat{\varepsilon}$ ,  $\hat{r} = r_k$ ,  $\hat{\delta} = \delta_k$  and  $\hat{x}_0 = x_k$ . Indeed, (a) and (c) follow from the fact that (a) and (c) hold for  $\varepsilon$ ,  $a$  and  $\sigma$ ; (b) and (d) are just (41) and (40) and (e) follows from the definitions of  $\hat{\delta}$ ,  $\hat{r}$  and  $\hat{\varepsilon}$ .

Observe that, from item 1 with  $j$ ,  $\hat{x}_0$  and  $\hat{\varepsilon}$  instead of  $k$ ,  $x_0$  and  $\varepsilon$ , respectively, we get

$$\|x_j - x_k\| \leq \|s(x_k)\| \frac{1 - (\hat{\varepsilon}/a)^k}{1 - \hat{\varepsilon}/a}.$$

So, letting  $j \rightarrow \infty$ , we have

$$\|\bar{x} - x_k\| \leq \|s(x_k)\| \frac{1}{1 - \hat{\varepsilon}/a}.$$

So

$$\|\bar{x} - x_{k+1}\| \leq \|s(x_{k+1})\| \frac{1}{1 - \hat{\varepsilon}/a} \leq \|s(x_k)\| \frac{\hat{\varepsilon}/a}{1 - \hat{\varepsilon}/a}, \quad (42) \quad \boxed{\text{bound}}$$

where the last inequality follows from item 4, with  $\hat{\varepsilon}$  instead of  $\varepsilon$ .

Using the triangle inequality, the definition of  $x_{k+1}$  and (42), we get

$$\begin{aligned} \|\bar{x} - x_k\| &\geq \|x_{k+1} - x_k\| - \|\bar{x} - x_{k+1}\| \\ &\geq \|s(x_k)\| - \|s(x_k)\| \frac{\hat{\varepsilon}/a}{1 - \hat{\varepsilon}/a} \\ &= \|s(x_k)\| \frac{1 - 2\hat{\varepsilon}/a}{1 - \hat{\varepsilon}/a} \end{aligned} \quad (43) \quad \boxed{\text{triang}}$$

In view of the definition of  $\hat{\varepsilon}$ , we have  $1 - 2\hat{\varepsilon}/a > 0$ , and from (43) and (42) we arrive at

$$\|\bar{x} - x_{k+1}\| \leq \|\bar{x} - x_k\| \frac{\hat{\varepsilon}/a}{1 - 2\hat{\varepsilon}/a}$$

which, combined with the definition of  $\hat{\varepsilon}$  yields

$$\|\bar{x} - x_{k+1}\| \leq \tau \|\bar{x} - x_k\|.$$

As  $\tau > 0$  is arbitrary, we conclude that  $\{x_k\}$  converges  $q$ -superlinearly to  $\bar{x}$ .  $\square$

Some comments are in order. Theorem 7 makes no assumption on the existence of Pareto optima for  $F$ . Instead, it shows that under some regularity conditions there exists at least a locally efficient point in a vicinity of the starting point, i. e. in a  $B[x_0, r]$ . Moreover, it shows that the whole sequence remains in that vicinity and converges superlinearly to that solution. Note also that, under these regularity assumptions, in general the limit point is not an isolated local optimum.

In standard local analysis of classical Newton's Method, close enough to a solution, all starting points produce sequences which remain in a vicinity of this solution and converge to it. In opposition to the scalar case, in Multicriteria Newton's Method, just some of these features are preserved: close enough to a solution, all starting points generate sequences which remain in a neighborhood of this optimum and converge to *some* Pareto optimal solution in that vicinity.

Now observe that, under our general assumptions (continuity and positive definiteness of all Hessians everywhere in  $U$ , etc.), if we know that there exists a locally Pareto optimal point  $\hat{x}$ , then, by item 1 of Theorem 1,  $\hat{x}$  is

critical for  $F$ , so, in view of item 2 of Lemma 2,  $\theta(\hat{x}) = 0$ . On the other hand, if  $V$  is a compact vicinity of  $\hat{x}$ , we have  $aI \leq \nabla^2 F_j(x) \leq bI$  for all  $j \in \{1, \dots, m\}$  and all  $x \in V$  (here,  $0 < a = \min_{j=1, \dots, m, x \in V, \|u\|=1} u^T \nabla^2 F(x) u \leq \max_{j=1, \dots, m, x \in V, \|u\|=1} u^T \nabla^2 F(x) u = b$ ). Since  $\theta$  is continuous (Lemma 2, item 3) and  $\theta(\hat{x}) = 0$ , there exists  $r > 0$  such that  $|\theta(x)|$  is small enough for all  $x \in B[\hat{x}, r]$ , so that, by virtue of Lemma 5,  $\|s(x)\|$  will also be small in that ball. In other words, assumptions (a)-(e) will hold in this situation. Therefore, we have the following result.

**cor9** **Corollary 8.** *If  $\hat{x}$  is a locally Pareto-optimal point for  $F$ , then there exists  $r > 0$  such that, for any  $x_0 \in B[\hat{x}, r] \subset U$ , the algorithm generates a sequence which converges superlinearly to some locally efficient point  $\bar{x}$ .*

Suppose that  $F$  has a compact level set  $\Gamma_z = \{x \in U : F(x) \leq z\}$ , with  $z \in \mathbb{R}^m$ . If  $x_0 \in \Gamma_z$ , then, as  $\{F(x_k)\}$  is  $\mathbb{R}^m$ -decreasing, the whole sequence  $\{x_k\}$  will remain in this compact level set. Furthermore, boundedness of the sequence  $\{F(x_k)\}$  combined with step 2. (c) of the algorithm enforces

$$\lim_{k \rightarrow \infty} t_k \theta(x_k) = 0.$$

Let  $\bar{x}$  be the limit of a subsequence of  $\{x_k\}$ . A uniform version of Corollary 3 around  $\bar{x}$  shows that the corresponding subsequence of  $\{\theta(x_k)\}$  must converge to 0. So, in view of Lemma 2,  $\bar{x}$  is a local optimum. Now, using Corollary 8, we conclude that the whole sequence converges to an optimum. We therefore have the following corollary.

**cor10** **Corollary 9.** *If  $x_0 \in U$  is taken in a compact level set of  $F$ , then the algorithm generates from  $x_0$  a sequence which converges superlinear to some locally efficient point  $\bar{x}$ .*

## 6 Quadratic Convergence

subsec:quadratic

The additional assumption of Lipschitz continuity of the Hessians  $\nabla^2 F_j$  ( $j = 1, \dots, m$ ) guarantees a  $q$ -quadratic converge rate of the algorithm, as we will now prove.

th:quadratic

**Theorem 10.** *Suppose that, in addition to all assumptions of Theorem 7, we have that  $\nabla^2 F_j$  is Lipschitz continuous on  $V$  with Lipschitz constant  $L$  ( $j = 1, \dots, m$ ).*

*Take  $\zeta \in (0, 1/2)$ . Then, there exists  $k_0$  such that for all  $k \geq k_0$  it holds that  $\tau_k := (L/a)\|s(x_k)\| < \zeta$  and*

$$\|\bar{x} - x_{k+1}\| \leq \frac{L}{a} \frac{(1 - \tau_k)}{(1 - 2\tau_k)^2} \|\bar{x} - x_k\|^2 \leq \frac{L}{a} \frac{(1 - \zeta)}{(1 - 2\zeta)^2} \|\bar{x} - x_k\|^2 \quad (44)$$

quad\_conclusion

where  $\bar{x}$  is the Pareto-optimal limit of  $\{x_k\}$ .

*The sequence  $\{x_k\}_k$  converges  $q$ -quadratically to  $\bar{x}$ .*

*Proof.* Due to item 2 of Theorem 7, it is clear that  $\tau_k = (L/a)\|s(x_k)\| < \zeta$ , for  $k$  large enough, say  $k \geq k_0$ . We follow the proof of Theorem 7 by defining vectors  $\tilde{v}_{k+1}$  as in (33) and functions  $G_k$  as in (35). As all  $\nabla^2 F_j$  are Lipschitz continuous,  $\nabla^2 G_k$  is also Lipschitz continuous. Therefore, using Lemma 4, we can refine (38) by way of (37) to

$$\|\tilde{v}_{k+1}\| \leq L\|s(x_k)\|^2,$$

and (34) leads to

$$|\theta(x_{k+1})| \leq \frac{1}{2a}L^2\|s(x_k)\|^4.$$

Hence, from Lemma 5, we get the inequality

$$(a/2)\|s(x_{k+1})\|^2 \leq \frac{1}{2a}L^2\|s(x_k)\|^4,$$

which, in turn, gives us

$$\|s(x_{k+1})\| \leq \frac{L}{a}\|s(x_k)\|^2. \quad (45) \quad \boxed{\text{quad\_Newt}}$$

Let  $i \geq k \geq k_0$ . Then, from the triangle inequality and the fact that we are taking full Newton steps, we get

$$\|x_i - x_{k+1}\| \leq \sum_{j=k+2}^i \|x_j - x_{j-1}\| = \sum_{j=k+2}^i \|s(x_{j-1})\|.$$

So, letting  $i \rightarrow \infty$  and using the convergence result of Theorem 7, we obtain

$$\|\bar{x} - x_{k+1}\| \leq \sum_{j=k+1}^{\infty} \|s(x_j)\|. \quad (46) \quad \boxed{\text{provis\_bound}}$$

Whence, combining (45) and (46) and using that  $\tau_k := (L/a)\|s(x_k)\| < 1/2$ , we get

$$\begin{aligned} \|\bar{x} - x_{k+1}\| &\leq (L/a)\|s(x_k)\|^2 + (L/a)^3\|s(x_k)\|^4 + (L/a)^7\|s(x_k)\|^8 + \dots \\ &= (L/a)\|s(x_k)\|^2(1 + \tau_k^2 + \tau_k^6 + \dots) \\ &\leq (L/a)\|s(x_k)\|^2 \sum_{j=0}^{\infty} \tau_k^j \\ &= (L/a)\|s(x_k)\|^2 \frac{1}{1 - \tau_k}. \end{aligned} \quad (47) \quad \boxed{\text{provis\_Newt2}}$$

Hence, from the triangle inequality, we get

$$\begin{aligned} \|\bar{x} - x_k\| &\geq \|x_{k+1} - x_k\| - \|\bar{x} - x_{k+1}\| \\ &\geq \|s(x_k)\| - \|s(x_k)\|^2 \frac{L/a}{1 - \tau_k} \\ &= \|s(x_k)\| \frac{1 - 2\tau_k}{1 - \tau_k}, \end{aligned} \quad (48) \quad \boxed{\text{triang\_ineq}}$$

where the last equality is a consequence of the definition of  $\tau_k$ . Now, from (47) and (48) we obtain

$$\|\bar{x} - x_{k+1}\| \leq (L/a) \left[ \|\bar{x} - x_k\| \frac{1 - \tau_k}{1 - 2\tau_k} \right]^2 \frac{1}{1 - \tau_k} = \|\bar{x} - x_k\|^2 \frac{(L/a)(1 - \tau_k)}{(1 - 2\tau_k)^2}.$$

The last inequality of (44) follows trivially.  $\square$

Under the Lipschitz assumption on all Hessians, we have the following  $q$ -quadratic versions of Corollaries 8 and 9.

**Corollary 11.** *If all  $\nabla^2 F_j$  are Lipschitz continuous and  $\hat{x}$  is Pareto optimal, then there exists an  $r > 0$  such that, for all  $x_0 \in B[x, r] \subset U$ , the algorithm generates a sequence which converges quadratically to some efficient point  $\bar{x}$ .*

**Corollary 12.** *If  $x_0 \in U$  is taken in a compact level set of  $F$  on which all  $\nabla^2 F_j$  are Lipschitz continuous, then the algorithm generates from  $x_0$  a sequence which converges  $q$ -quadratically to some Pareto optimal point  $\bar{x}$ .*

## 7 Numerical Results

sec: numerical

A MATLAB prototype implementation of the method described in the last sections was tested on various problems from the literature. All tests were executed within Matlab V7.2.0 (R2006a). The implementation uses the stopping criterion

$$\theta(x_k) > -\delta$$

for some prespecified parameter  $\delta > 0$  in order to stop at the point  $x_k$ . Box constraints of the form  $L \leq x \leq U$  for the original multiobjective problem at hand are handled by augmenting the direction search program (8) that is being solved in each step of our implementation by an additional box constraint of the form

$$L - x \leq s \leq U - x.$$

Note that values of  $L_i = -\infty$  or  $S_i = +\infty$  can be used in our implementation.

For all test cases considered, we used the value of  $\delta = 5 \times \mathbf{eps}$  for the stopping criterion. Here,  $\mathbf{eps}$  denotes the machine precision given. For the line search,  $\sigma = 0.1$  was used. The maximum number of iterations was set to 500.

Table 1 specifies the main characteristics of the problems investigated. We use a three-letter abbreviation of author names to indicate the origin of a problem instance, followed by a number to indicate the number of the problem in the corresponding reference (i. e. "JOS1" for problem no. 1 from Jin, Olhofer, and Sendhoff [26], "ZDT6" for problem no. 6 from Zitzler, Deb, and Thiele [43], etc.). Many problems from the literature (i.e. JOS1, JOS4, ZDT1–ZDT6) are constructed in such a way that the dimension  $n$  of the variable space can be chosen by the user. We have done so and considered various choices of  $n$  in our experiments; the corresponding problem names are augmented by letters a–h (i.e. "JOS1a" to "JOS1h"), see Table 1 for details. Moreover, we vary lower

and upper bounds to investigate the effect of different starting points on the number of iterations and function evaluations (problem instances DD1a–DD1d and JOS1e–JOSh).

Note that all cases except the last three problem types, LTDZ, TR1, and FDS, are bicriteria problems, i. e.  $m = 2$  holds. The latter three problems are three-criteria. We do not consider problem no. 5 from [43] (ZDT5), since this problem is a discrete multiobjective optimization problem. While we have  $F \in C^\infty$  in the interior of the box constraints for each problem considered, only the problems JOS1a–JOS1h, ZDT1a, ZDT1b, and FDS are convex, so we are mainly concerned with the amount of work (iterations, function evaluations) it takes in finding local Pareto points by using our local search method.

All problems were solved 200 times using starting points from a uniform random distribution between a lower bound  $U \in \mathbb{R}^n$  and an upper bound  $L \in \mathbb{R}^n$  as specified in Table 1, column 3 and 4. These bounds also define the box constraints for each problem considered. Average number of iterations and average number of function evaluations are reported in the two right-most columns of Table 1.

We proceed by discussing our findings some further by taking a detailed look at some of the test problems and the corresponding results.

Problem JOS1 is a simple convex quadratic test problem used by various authors to benchmark algorithms. As it can be seen, this problem, solved for various dimensions  $n$ , poses no challenges to the method proposed here, and the number of function evaluations shows that the simple Armijo-like step size rule employed here is successful either immediately or after one backtracking step. The same holds for the nonquadratic problems JOS4, ZDT1–ZDT4, and TR1. Extending the feasible region (see JOS1e–JOSh and DD1a–DD1d) seems to indicate, quite naturally, that the method is robust with respect to the choice of the starting point for convex problems, but not necessarily so for nonconvex problems.

Problem Hil is a non-convex, albeit smooth multicriteria problem. As it can be seen from Figure 2, our method is able to generate a rough outline of the set of efficient points (cmp. [24]), but, as it has to be expected, it also gets stuck in locally efficient points. Similar results hold for the other nonconvex problems considered.

Likewise, Figure 3 shows clearly that, given a reasonable number of starting points, the method is able to identify the set of local, nonglobal Pareto points as well as the set of global Pareto points for problem PNR.

On the other hand, it is clear that our method does not perform well on problems SD and ZDT6 as on the other problems considered. Here, either the simple linesearch of Armijo-type employed fails due to the strong curvature of the problems under consideration [39, 43], or the nonconvexity of the problem leads to nonconvex direction search programs (8), and the local optimization routine employed to solve such programs cannot cope.

Results for a three-criteria problem, problem LTDZ, are visualized in Figure 4. Note that for this problem, the image of the set of efficient points is on the boundary of a convex set in  $\mathbb{R}^3$ , albeit the problem in itself is not convex.

Name	$n$	$U^T$	$L^T$	Source	iter	feval
Deb	2	[1/10, 1]	[1, 1]	[9]	5.49	32.13
Hil	2	[0, ..., 0]	[5, ..., 5]	[24]	3.34	8.27
DD1a	5	-[1, ..., 1]	[1, ..., 1]	[8]	3.51	7.63
DD1b	5	-[5, ..., 5]	[5, ..., 5]	[8]	11.12	16.33
DD1c	5	-10[1, ..., 1]	10[1, ..., 1]	[8]	71.78	77.77
DD1d	5	-20[1, ..., 1]	20[1, ..., 1]	[8]	202.645	206.28
KW2	2	[-3, -3]	[3, 3]	[29]	8.06	13.05
JOS1a	50	-[2, ..., 2]	[2, ..., 2]	[26, Ex. 1]	2.01	3.01
JOS1b	100	-[2, ..., 2]	[2, ..., 2]	[26, Ex. 1]	2.00	3.00
JOS1c	200	-[2, ..., 2]	[2, ..., 2]	[26, Ex. 1]	2.00	3.00
JOS1d	500	-[2, ..., 2]	[2, ..., 2]	[26, Ex. 1]	2.00	3.00
JOS1e	100	-[10, ..., 10]	[10, ..., 10]	[26, Ex. 1]	2.26	3.26
JOS1f	100	-[50, ..., 50]	[50, ..., 50]	[26, Ex. 1]	2.44	3.44
JOS1g	100	-100[1, ..., 1]	100[1, ..., 1]	[26, Ex. 1]	2.21	3.21
JOS1h	200	-100[1, ..., 1]	100[1, ..., 1]	[26, Ex. 1]	2.25	3.25
JOS4a	50	[1, ..., 1]/100	[1, ..., 1]	[26, Ex. 4]	2.00	2.00
JOS4b	100	[1, ..., 1]/100	[1, ..., 1]	[26, Ex. 4]	2.00	2.00
JOS4c	200	[1, ..., 1]/100	[1, ..., 1]	[26, Ex. 4]	2.00	2.00
PNR	2	-[2, ..., 2]	[2, ..., 2]	[36]	4.11	5.95
SD	4	[1, $\sqrt{2}$ , $\sqrt{2}$ , 1]	[3, ..., 3]	[39]	28.08	277.85
ZDT1a	100	[1, ..., 1]/100	[1, ..., 1]	[43, Ex. 1]	2.00	2.00
ZDT1b	200	[1, ..., 1]/100	[1, ..., 1]	[43, Ex. 1]	2.00	2.00
ZDT2a	50	[0, ..., 0]	[1, ..., 1]	[43, Ex. 2]	2.00	2.00
ZDT2b	100	[0, ..., 0]	[1, ..., 1]	[43, Ex. 2]	2.00	2.00
ZDT3a	50	[1, ..., 1]/100	[1, ..., 1]	[43, Ex. 3]	2.18	2.56
ZDT3b	100	[1, ..., 1]/100	[1, ..., 1]	[43, Ex. 3]	2.21	2.57
ZDT4a	50	-[0.01, 5, ..., 5]	[1, 5, ..., 5]	[43, Ex. 4]	2.07	2.50
ZDT4b	100	-[0.01, 5, ..., 5]	[1, 5, ..., 5]	[43, Ex. 4]	2.06	2.48
ZDT6a	3	[0, 0, 0]	[1, 1, 1]	[43, Ex. 6]	10.36	97.86
ZDT6b	10	[0, ..., 0]	[1, ..., 1]	[43, Ex. 6]	7.36	68.46
LTDZ	3	[0, 0, 0]	[1, 1, 1]	[30]	2.46	3.23
TR1	3	[0, 0, 0]	[1, 1, 1]	[40]	2.83	2.83
FDSa	5	[-2, ..., 2]	[2, ..., 2]	(49)-(51)	8.39	17.04
FDSb	10	[-2, ..., 2]	[2, ..., 2]	(49)-(51)	14.67	26.49
FDSc	50	[-2, ..., 2]	[2, ..., 2]	(49)-(51)	44.54	42.31
FSDc	100	[-2, ..., 2]	[2, ..., 2]	(49)-(51)	424.88	433.15
FDSd	200	[-2, ..., 2]	[2, ..., 2]	(49)-(51)	381.20	382.50

tbl:problems

Figure 1: Main data for the problem instances considered as well as average number of iterations (iter) and average number of function evaluations (feval). The last five problems considered are three-criteria problems. Note that the number of gradient and Hessian evaluations is equal to the number of iterations.

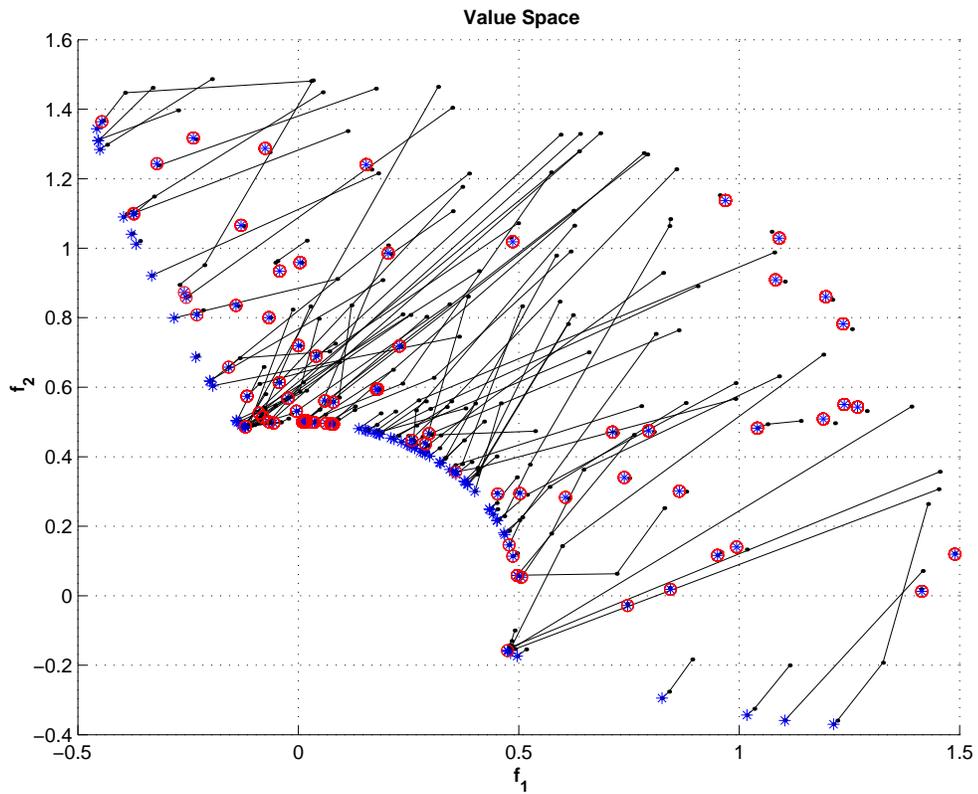


fig:prob2

Figure 2: Value space of Problem No. 2 for 150 random starting points. Circled points indicate local efficient points that are not globally efficient.

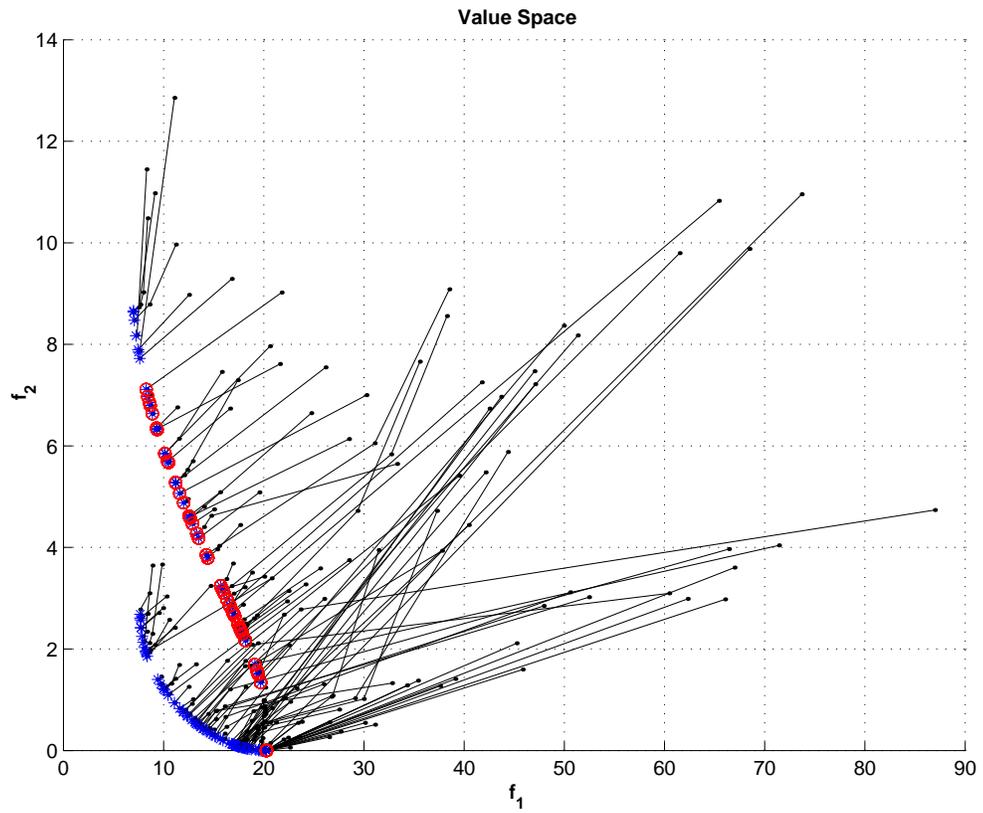
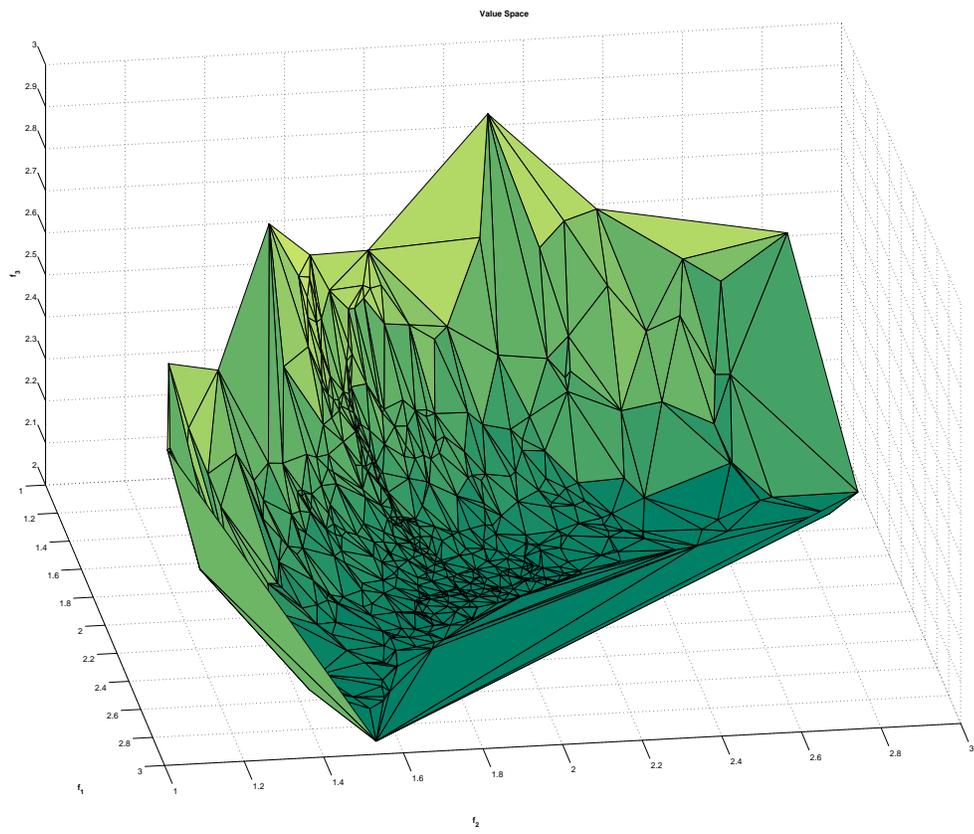


fig:prob6

Figure 3: Value space of Problem PNR for 150 random starting points. Circled points indicate local efficient points that are not globally efficient.



`fig:probltdz`

Figure 4: Delaunay triangulation of the approximation of the set of efficient points in value space for Problem LTDZ, as generated by Newton's Method. 200 random starting points have been used to generate a pointwise approximation to the set of efficient points, from which the Delaunay triangulation as shown has been constructed.

Finally, the algorithm was tested on problem FDS, a problem of variable dimension  $n$  defined by

$$F_1(x) := \frac{1}{n^2} \sum_{k=1}^n k(x_k - k)^4, \quad (49) \quad \boxed{\text{eq:crit1}}$$

$$F_2(x) := \exp\left(\sum_{k=1}^n x_k/n\right) + \|x\|_2^2, \quad (50)$$

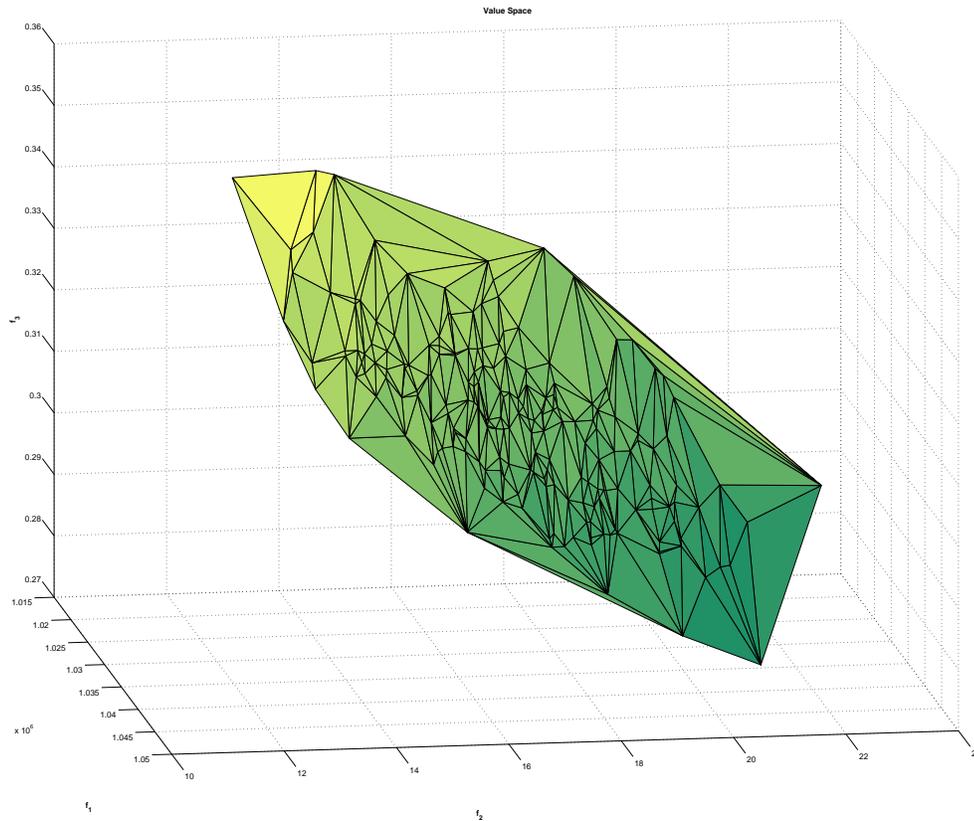
$$F_3(x) := \frac{1}{n(n+1)} \sum_{k=1}^n k(n-k+1)e^{-x_k}. \quad (51) \quad \boxed{\text{eq:crit3}}$$

These objectives have been designed as a convex problem with three criteria whose numerical difficulty is sharply increasing in the dimension  $n$ . A visualization of the image of the set of efficient points, together with a Delaunay triangulation of the corresponding image points in  $\mathbb{R}^3$  can be found in Figure 5. As it can be seen in Table 1, the implementation works well if the dimension of the problem is not too high. However, for problem FDS<sub>c</sub> resp. FDS<sub>d</sub>, 166 resp. 126 of the 200 starting points considered resulted in an iteration history that reached the maximum number of iterations allowed.

## 8 Conclusions

With respect to theoretical results obtained, Newton's method for multiobjective optimization behaves exactly as its counterpart for single-criterion optimization: if all functions involved are twice continuous differentiable and strictly convex, the sequence provided by the method converges superlinearly to a solution of the given problem, and full Armijo-steps will be chosen in a neighbourhood of such solution. Moreover, quadratic convergence holds in case the second derivatives are Lipschitz-continuous.

Finally, we note that the numerical performance of the Newton method for multicriteria optimization presented here mimics the performance of Newton's method for the standard single-criterion case: problems whose objective functions display a weak or moderate amount of curvature pose no challenge to the method at hand, and in such cases the method is fairly robust with respect to the dimension of the problem and the starting point chosen. Moreover, the numerical results presented indicate that the method works also well in the non-convex case, although, of course, no guarantee can be given that the computed Pareto points are locally optimal instead of globally optimal. Our implementation, however, experiences difficulties when employed on problems exhibiting a high degree of curvature. This doubtlessly so due to the rather simple line-search mechanism employed here. This, as well as an adaptation of Newton's method for constrained multiobjective problems is subject of further research.



**fig:problfds**

Figure 5: Delaunay triangulation of the approximation of the set of efficient points in value space for Problem FDS, as generated by Newton's Method. 200 random starting points have been used to generate a pointwise approximation to the set of efficient points, from which the Delaunay triangulation as shown has been constructed.

## References

- Adams:99** [1] Warren Adams, Marie Coffin, Peter Kiessler, K. B. Kulasekera, Matthew J. Saltzmann, and Margaret Wiecek. An optimization-based framework for affordable science and technology decisions at the office of naval research. Clemson report, Clemson University, Clemson, South Carolina, October 16 1999.
- Bagchi:89** [2] U. Bagchi. Simultaneous minimization of mean and variation of flow time and waiting time in single machine systems. *Operations Research*, 37:118–125, 1989.
- BDMS:06** [3] Jürgen Branke, Kalyanmoy Deb, Kaisa Miettinen, and Roman Slowinski, editors. *Practical Approaches to Multi-Objective Optimization*. Dagstuhl Seminar, 2006. Seminar 06501. Forthcoming.
- BDMS:04** [4] Jürgen Branke, Kalyanmoy Deb, Kaisa Miettinen, and Ralph E. Steuer, editors. *Practical Approaches to Multi-Objective Optimization*. Dagstuhl Seminar, 2004. Seminar 04461. Electronically available at <http://drops.dagstuhl.de/portals/index.php?semnr=04461>.
- CF:98** [5] E[milio] Carrizosa and J. B. G. Frenk. Dominating sets for convex functions with some applications. *Journal of Optimization Theory and Applications*, 96:281–295, 1998.
- HeWeFl:05** [6] Bernd Weyers Christoph Heermann and Jörg Fliege. A new adaptive algorithm for convex quadratic multicriteria optimization. In Jürgen Branke and Kaisa Miettinen, editors, *Practical Approaches to Multi-Objective Optimization*, Schloß Dagstuhl, Germany, 2005. Dagstuhl Seminar Proceedings. Electronically available from <http://drops.dagstuhl.de/opus/volltexte/2005/238/>.
- Das:97a** [7] I. Das and J. E. Dennis. A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multi-criteria optimization problems. *Structural Optimization*, 14:63–69, 1997.
- DaDe:98** [8] Indraneel Das and J. E. Dennis. Normal-Boundary Intersection: A New Method for Generating Pareto Optimal Points in Nonlinear Multicriteria Optimization Problems. *SIAM Journal on Optimization*, 8(3):631–657, 1998.
- Deb:99** [9] Kalyanmoy Deb. Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *Evolutionary Computation*, 7:205–230, 1999.
- DennisSchnabel:96** [10] J. E. Dennis and R. B. Schnabel. *Numerical methods for Unconstrained Nonlinear Equations*. SIAM, Philadelphia, 1996.

- DrIu:04** [11] L. M. Grana Drummond and A. N. Iusem. A projected gradient method for vector optimization problems. *Computational Optimization and Applications*, 28:5–29, 2004.
- DrSv:05** [12] L. M. Grana Drummond and B. F. Svaiter. A steepest descent method for vector optimization. *Journal of Computational and Applied Mathematics*, 175:395–414, 2005.
- Eichfelder:06** [13] Gabriele Eichfelder. *Parameter Controlled Solutions of Nonlinear Multi-Objective Optimization Problems (in german)*. PhD thesis, Department of Mathematics, University of Erlangen-Nürnberg, 2006.
- eschenauer:90** [14] H. Eschenauer, J. Koski, and A. Osyczka. *Multicriteria Design Optimization*. Springer, Berlin, 1990.
- evans:84** [15] G. Evans. Overview of techniques for solving multiobjective mathematical programs. *Management Science*, 30:1268–1282, 1984.
- Fliege:99** [16] Jörg Fliege. Project report *Research in Brussels: Atmospheric Dispersion Simulation*. Report BEIF/116, Vakgroep Beleidsinformatica, Department of Management Informatics, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussel, Belgium, October 1999. Also published by the Region Bruxelles-Capital, Belgium. Available electronically via <http://www.mathematik.uni-dortmund.de/user/lxs/fliege/olaf/>.
- Fliege:01** [17] Jörg Fliege. OLAF — a general modeling system to evaluate and optimize the location of an air polluting facility. *ORSpektrum*, 23(1):117–136, 2001.
- Fliege:06** [18] Jörg Fliege. An efficient interior-point method for convex multicriteria optimization problems. *Mathematics of Operations Research*, 31:825–845, November 2006.
- SvFl:00** [19] Jörg Fliege and Benar Fux Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51:479–494, 2000.
- FuDiwekar:04** [20] Yan Fu and Urmila M. Diwekar. An efficient sampling approach to multiobjective optimization. *Annals of Operations Research*, 132:109–134, November 2004.
- geoffrion:68** [21] Arthur M. Geoffrion. Proper Efficiency and the Theory of Vector Maximization. *Journal of Optimization Theory and Applications*, 22:618–630, 1968.
- Graveletal:92** [22] M. Gravel, I. M. Martel, R. Madeau, W. Price, and R. Tremblay. A multicriterion view of optimal resource allocation in job-shop production. *European Journal of Operation Research*, 61:230–244, 1992.
- Heseler:05** [23] Andree Heseler. *Warmstart-Strategies in Parametric Optimization with Application in Multicriteria Optimization (in german)*. PhD thesis, Department of Mathematics, University of Dortmund, Germany, 2005.

- Hillermeier:01** [24] Claus Hillermeier. *Nonlinear Multiobjective Optimization: a Generalized Homotopy Approach*, volume 25 of *International Series of Numerical Mathematics*. Birkäuser Verlag, Basel — Boston — Berlin, 2001.
- HuJa:00** [25] Angelika Hutterer and Johannes Jahn. Optimization of the location of antennas for treatment planning in hyperthermia. Preprint 265, Institut für Angewandte Mathematik, Universität Erlangen-Nürnberg, Martensstraße 3, D-91058 Erlangen, June 15 2000.
- JOS:01** [26] Yaochu Jin, Markus Olhofer, and Bernhard Sendhoff. Dynamic weighted aggregation for evolutionary multi-objective optimization: Why does it work and how? In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1042–1049, 2001.
- JuJaKi:97** [27] A. Jüschke, J. Jahn, and A. Kirsch. A bicriterial optimization problem of antenna design. *Computational Optimization and Applications*, 7:261–276, 1997.
- Kasperska:04** [28] Renata Kasperska. Bi-criteria Optimization Of Open Cross-Section Of The Thin-Walled Beams With Flat Flanges. Presentation at 75th Annual Scientific Conference of the Gesellschaft für Angewandte Mathematik und Mechanik e. V., Dresden, Germany, March 2004.
- KW:05** [29] I. Y. Kim and O. L. de Weck. Adaptive weighted sum method for bi-objective optimization: Pareto front generation. *Structured Multidisciplinary Optimization*, 29:149–158, 2005.
- LTDZ:02** [30] M. Laumanns, L. Thiele, K. Deb, and E. Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary Computation*, 10:263–282, 2002.
- LeVaWe:92** [31] T. M. Leschine, H. Wallenius, and W. A. Verdini. Interactive multiobjective analysis and assimilative capacity-based ocean disposal decisions. *European Journal of Operation Research*, 56:278–289, 1992.
- miettinen:95** [32] K[aisa] Miettinen and M[arko] M. Mäkelä. Interactive bundle-based method for nondifferentiable multiobjective optimization: Nimbus. *Optimization*, 34:231–246, 1995.
- MDS:06** [33] Sanaz Mostaghim, Jürgen Branke, and Hartmut Schmeck. Multi-objective particle swarm optimization on computer grids. Technical report. 502, Institut AIFB, University of Karlsruhe (TH), December 2006.
- PSVZ:03** [34] G. Palermo, C. Silvano, S. Valsecchi, and V. Zaccaria. A system-level methodology for fast multi-objective design space exploration. In *Proceedings of the 13th ACM Great Lakes symposium on VLSI*, pages 92–95, New York, 2003. ACM.

- Prabuddhaetal:92** [35] D. Prabuddha, J. B. Gosh, and C. E. Wells. On the minimization of completion time variance with a bicriteria extension. *Operations Research*, 40:1148–1155, 1992.
- PNR:01** [36] Mike Preuss, Boris Naujoks, and Guenter Rudolph. Pareto set and emoa behavior for simple multimodal multiobjective functions. Unpublished manuscript., 2001.
- Recchioni:03** [37] Maria Cristina Recchioni. A path following method for box-constrained multiobjective optimization with applications to goal programming problems. *Mathematical Methods of Operations Research*, 58:69–85, September 2003.
- ShanWang:05** [38] Songqing Shan and G. Gary Wang. An efficient pareto set identification approach for multiobjective optimization on black-box functions. *Journal of Mechanical Design*, 127:866–874, September 2005.
- SD:92** [39] W. Stadtler and J. Dauer. *Multicriteria optimization in engineering: A tutorial and survey*. American Institute of Aeronautics and Astronautics, 1992.
- TR:07** [40] Ravindra V Tappeta and John E. Renaud. Interactive multiobjective optimization procedure. Accepted for publication in *AIAA Journal*, 2007.
- Tavana:04** [41] Madjid Tavana. A subjective assessment of alternative mission architectures for the human exploration of Mars at NASA using multicriteria decision making. *Computers and Operations Research*, 31:1147–1164, 2004.
- white:98** [42] D[ouglas] J[ohn] White. Epsilon-dominating solutions in mean-variance portfolio analysis. *European Journal of Operational Research*, 105:457–466, 1998.
- ZDT:00** [43] Eckart Zitzler, Kalyanmoy Deb, and Lothar Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8:173–195, April 2000.