

Trust region globalization strategy for the nonconvex unconstrained multiobjective optimization problem

Gabriel A. Carrizo¹ · Pablo A. Lotito² ·
María C. Maciel¹

Received: 19 June 2013 / Accepted: 22 October 2015 / Published online: 30 October 2015
© Springer-Verlag Berlin Heidelberg and Mathematical Optimization Society 2015

Abstract A trust-region-based algorithm for the nonconvex unconstrained multiobjective optimization problem is considered. It is a generalization of the algorithm proposed by Fliege et al. (SIAM J Optim 20:602–626, 2009), for the convex problem. Similarly to the scalar case, at each iteration a subproblem is solved and the step needs to be evaluated. Therefore, the notions of decrease condition and of predicted reduction are adapted to the vectorial case. A rule to update the trust region radius is introduced. Under differentiability assumptions, the algorithm converges to points satisfying a necessary condition for Pareto points and, in the convex case, to a Pareto points satisfying necessary and sufficient conditions. Furthermore, it is proved that the algorithm displays a q -quadratic rate of convergence. The global behavior of the algorithm is shown in the numerical experience reported.

Keywords Multiobjective optimization · Trust region · Newton method · Convergence

Mathematics Subject Classification 90C29 · 65K05 · 49M37

This work has been partially supported by Universidad Nacional del Sur, Project 24/L082.

✉ Gabriel A. Carrizo
gabrielanibal@gmail.com

Pablo A. Lotito
plotito@exe.unicen.edu.ar

María C. Maciel
immaci@criba.edu.ar

¹ Departamento de Matemática, Universidad Nacional del Sur, Av. Alem 1253,
8000 Bahía Blanca, Argentina

² Universidad Nacional del Centro, Tandil, Argentina

1 Introduction

In this work, let us consider the multiobjective optimization problem where a set a objective functions must be minimized simultaneously

$$\min_{x \in \mathbb{R}^n} F(x), \quad F : \mathbb{R}^n \rightarrow \mathbb{R}^m.$$

Since there is not a single point which minimize all the functions simultaneously the concept of optimality is established in terms of *Pareto optimality* or *efficiency*. Let us recall that a point is said to be Pareto optimal or 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 of this type of problems can be found in different areas. For instance [14, 17, 18] to mention some of them.

Among the strategies for multiobjective optimization problems we can mention the scalarization approach [11, 12]. The idea is to solve one or several parameterized single-objective problems, resulting in a corresponding number of Pareto optimal points. One of the drawbacks of this approach is the choice of the parameters that are not known in advance (see [9, sect. 7]).

Algorithms that do not scalarize the problem have recently been developed. Some of these techniques are extensions of well known scalar algorithms [9, 10], while others take ideas developed in heuristic optimization [13]. For the latter, no convergence proofs are known, and empirical results show that convergence generally is, as in the scalar case, quite slow [20].

In this paper, we focus on some a priori parameter-free optimizations problems developed in [9, 10]. In such methods there is a scalarization procedure that does not set parameters before solving the problem. The main idea of it consists in replacing, at each iteration, each objective function with a quadratic model for it, as in the scalar Newton's method. With these local models they minimize the max-scalarization of the variations of the quadratic approximations and obtain a descent direction for the convex case. For globalization purposes [9, 10] use a line search strategy and obtain convergence to Pareto points in the convex case. In the present work we consider a trust region algorithm for the nonconvex case by using some ideas of these works.

According to our knowledge there is not too much activity on this subject. In 2006, Ashry [2] converts the multiobjective optimization problem in a scalar one and uses a TR-algorithm for solving the general nonlinear programming problem. He uses an algorithm based on the Byrd theory for equality constrained minimization. Recently Qu et al. [16] have developed a TR-algorithm for the unconstrained vector optimization case which has some points in common with ours. For instance, they apparently were inspired by the work of Fliege et. al. [9] like us, but they focus the discussion on the non differentiable case that is not our case. Finally, it is also interesting the work of Villacorta et. al. [19] where they propose a TR-algorithm for the same problem with a different subproblem and a different way to evaluate the step.

This work is organized as follows: the next section is devoted to state the optimality condition, the subproblem involved and some properties about its solution. The algorithm is presented in the third section and the fourth one is dedicated to its global

convergence properties. The fifth section is dedicated to the analysis of the local q -quadratic rate of convergence of the algorithm. Numerical tests are presented in the sixth section and in the last one some conclusions are established.

2 The subproblem and the optimality conditions

The problem under discussion consists to find an efficient point or Pareto optimum of $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, i.e. a point $x^* \in \mathbb{R}^n$ such that

$$\nexists y \in \mathbb{R}^n, F(y) \leq F(x^*), F(y) \neq F(x^*),$$

where the inequality sign \leq between vectors must be understood in the componentwise sense. This means that one is seeking to minimize F in the partial order induced by the positive orthant \mathbb{R}_+^m .

A point x^* is weakly efficient or weak Pareto optimum if

$$\nexists y \in \mathbb{R}^n, F(y) < F(x^*),$$

where the vector strict inequality $F(y) < F(x^*)$ has to be understood componentwise too.

Locally efficient points are also called local Pareto optimal, and locally weakly efficient points are also called local weak Pareto optimal. Note that if 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 this paper, unless explicitly mentioned, we will assume that $F \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^m)$, i.e., F is twice continuously differentiable on \mathbb{R}^n , $F(x) = (f_1(x), \dots, f_m(x))^T$. For $x \in \mathbb{R}^n$, denote by $\nabla F(x) \in \mathbb{R}^{m \times n}$ the Jacobian matrix of the vectorial function F at x , by $\nabla f_j(x) \in \mathbb{R}^n$ the gradient vector of the scalar function f_j at x , and by $\nabla^2 f_j(x) \in \mathbb{R}^{n \times n}$ the Hessian matrix of f_j at x .

Denoting the column space of the Jacobian matrix by $\mathcal{R}(\nabla F(x))$, a point is $x^* \in \mathbb{R}^n$ is *critical* for F if

$$\mathcal{R}(\nabla F(x)) \cap (-\mathbb{R}_{++}^m) = \emptyset.$$

This is necessary condition for Pareto optimality [15].

Clearly, if x^* is critical for F , then for all $s \in \mathbb{R}^n$ there exists an index $j_0 = j_0(s) \in \{1, \dots, m\}$ such that

$$\nabla f_{j_0}(x^*)^T s \geq 0.$$

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

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

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

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

The next proposition due to Fliege et. al. [9] establishes the relationship between the properties of being a critical and an optimal point.

Proposition 1 *Let F be continuously differentiable on \mathbb{R}^n , i.e. $F \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m)$.*

- (i) *If x^* is locally weak Pareto optimal, then x^* is a critical point for F .*
- (ii) *If F is \mathbb{R}^m -convex and x^* is critical for F , then x^* is weak Pareto optimal.*
- (iii) *If $F \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^m)$, and the Hessians matrices are positive definite for all x and for all j , and if x^* is critical point for F , then x^* is Pareto optimal.*

With the purpose to obtain a descent direction for F at x , in [10] the problem

$$\begin{aligned} \min_{s.t.} \quad & \alpha + \frac{1}{2} \|s\|^2 \\ & (\nabla F(x)s)_j \leq \alpha, \quad j = 1, \dots, m, \end{aligned} \tag{1}$$

is solved. In this problem α is an auxiliar variable and s is used in order to obtain a step for the iterations. If a point x is a critical point then the minimum of (1) is 0 (see Lemma 1 of [10]). In this way the authors obtain a steepest descent direction.

In [9] a Newton direction is defined as the solution of

$$\min_{s \in \mathbb{R}^n} \max_{j=1, \dots, m} \left[\nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s \right]. \tag{2}$$

Because the problem (2) is not differentiable, the following problem is introduced

$$\begin{aligned} \min_{s.t.} \quad & t \\ & \nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s - t \leq 0, \quad j = 1, \dots, m, \\ & (t, s) \in \mathbb{R} \times \mathbb{R}^n, \end{aligned}$$

which is differentiable and has the same solution as (2). In [9] the convex case is considered and the direction s , solution of the problem, is a descent direction. Unfortunately, in the nonconvex case can not be guaranteed that the direction s given by this subproblem will be a descent direction, It can be seen in the next example.

Example 1 Consider $F(x_1, x_2) = \left[e^{x_1-1} + e^{x_2-1}, \frac{x_1^2-x_2^2}{2} - 10x_1 + x_2 \right]^T$. Clearly f_1 is convex and f_2 is not. For this, the associated subproblem (2) at the point $x = (1, 1)$ is

$$\begin{aligned} \min_{s.t.} \quad & t \\ & \frac{1}{2} s^T \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} s + [-10 \ 1] s - t \leq 0, \\ & \frac{1}{2} s^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} s + [1 \ 1] s - t \leq 0. \end{aligned}$$

The approximate solution is $t = -0.6024$ and $s = (-0.1284, -1.1883)^T$. This direction is a descent direction for f_1 but is not for f_2 .

Therefore, to overcome this difficulty in the nonconvex case, to obtain a descent direction we add the linear inequality constraints $\nabla f_j(x)^T s - t \leq 0, j = 1, \dots, m$,

$$\begin{aligned} \min_{s,t} \quad & t \\ \text{s.t.} \quad & \nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s - t \leq 0, \quad j = 1, \dots, m, \\ & \nabla f_j(x)^T s - t \leq 0, \quad j = 1, \dots, m, \\ & (t, s) \in \mathbb{R} \times \mathbb{R}^n. \end{aligned}$$

The resulting problem could be not bounded, so we introduce a trust region, $\|s\| \leq \Delta$:

$$\begin{aligned} \min_{s,t} \quad & t \\ \text{s.t.} \quad & \nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s - t \leq 0, \quad j = 1, \dots, m, \\ & \nabla f_j(x)^T s - t \leq 0, \quad j = 1, \dots, m, \\ & \|s\| \leq \Delta, \\ & (t, s) \in \mathbb{R} \times \mathbb{R}^n, \end{aligned} \quad (3)$$

where $\Delta > 0$.

Remark 1 The trust region is applied only to the original variables of the problem $s \in \mathbb{R}^n$ and not to the variable t which has been introduced only to avoid the nondifferentiability of the problem.

The problem (3) is the key in our development, so we will analyze some of its properties. Let us define $\theta(x)$ as the optimal value of problem (3) and $s(x)$ as the value of s in an optimal solution of (3).

It is important to note that $s(x)$ could not be a function because the problem (3) is not convex, therefore has no unique solution. The relation among a critical point x^* , the Newton step $s(x^*)$ and the value $\theta(x^*)$ is established in the next lemma,

Lemma 1 *Let $\theta(x)$ be as above, then:*

- (a) *For any $x \in \mathbb{R}^n$, $\theta(x)$ is well defined, continuous and $\theta(x) \leq 0$.*
- (b) *$s(x)$ is bounded.*
- (c) *The following statements:*
 - (i) *x^* is not a critical point.*
 - (ii) *$\theta(x^*) < 0$.*
 - (iii) *$s(x^*) \neq 0$.*

satisfy (i) \Leftrightarrow (ii) \Rightarrow (iii). Moreover, if the vectorial function F is \mathbb{R}^m -convex, then (i), (ii) and (iii) are equivalents.

Proof It's clear that

$$\theta(x) = \min_{\|s\| \leq \Delta} \max_{j=1, \dots, m} \left\{ \nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s, \nabla f_j(x)^T s \right\}$$

is well defined because it is the maximum of continuous functions over a compact set, the feasible set of the subproblem. Then $\theta(x)$ is well defined.

Due to $(t, s) = (0, \mathbf{0})$ is feasible a point, from (3), $\theta(x) \leq 0$.

Now, we prove the continuity of $\theta(x)$. As any point in \mathbb{R}^n is in the interior of a compact subset of \mathbb{R}^n , it is enough to show continuity of θ on an arbitrary compact set $W \subseteq \mathbb{R}^n$. For each $x \in W$ and $j = 1, \dots, m$, we define

$$\phi_{x,j}(z) = \nabla f_j(z)^T s(x) + \frac{1}{2} s(x)^T \nabla^2 f_j(z) s(x)$$

and

$$\psi_{x,j}(z) = \nabla f_j(z)^T s(x).$$

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

$$\Phi_x(z) = \max_{j=1, \dots, m} \{\phi_{x,j}(z), \psi_{x,j}(z)\},$$

where the maximum is over $\phi_{x,1}(z), \dots, \phi_{x,m}(z), \psi_{x,1}(z), \dots, \psi_{x,m}(z)$, is equicontinuous too.

Given $\epsilon > 0$; there exists $\delta > 0$ such that $\forall y, z \in W$

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

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

$$\begin{aligned} \theta(z) &\leq \max_{j=1, \dots, m} \left\{ \nabla f_j(z)^T s(y) + \frac{1}{2} s(y)^T \nabla^2 f_j(z) s(y), \nabla f_j(z)^T s(y) \right\} \\ &= \Phi_y(z) \\ &\leq \Phi_y(y) + |\Phi_y(z) - \Phi_y(y)| \\ &< \theta(y) + \epsilon. \end{aligned}$$

i.e. $\theta(z) - \theta(y) < \epsilon$. Interchanging the roles of z and y , we conclude that $|\theta(y) - \theta(z)| < \epsilon$. Therefore (a) has been proved.

The values of $s(x)$ are clearly bounded because of the trust region.

To prove the implications of (c):

(i) \Rightarrow (ii) If x^* is not critical

$$\mathcal{R}(\nabla F(x^*)) \cap (-\mathbb{R}_{++}^m) \neq \emptyset,$$

so $\exists \tilde{s} \in \mathbb{R}^n$ such that $\nabla f_j(x^*)^T \tilde{s} < 0$ for all $j = 1, \dots, m$.

Then there exists $\tilde{t} < 0$ such that

$$\nabla f_j(x^*)^T \tilde{s} - \tilde{t} < 0 \quad j = 1, \dots, m.$$

If $\tilde{s}^T \nabla^2 f_j(x^*) \tilde{s} \leq 0$ for all $j = 1, \dots, m$, then

$$\nabla f_j(x^*)^T \tilde{s} + \frac{1}{2} \tilde{s}^T \nabla^2 f_j(x^*) \tilde{s} < \tilde{t} < 0. \quad (4)$$

If $\tilde{s}^T \nabla^2 f_j(x^*) \tilde{s} > 0$ for some $j = 1, \dots, m$, then choosing $l\tilde{s}$, with l small enough, we have that there exists $\hat{t} < 0$

$$\nabla f_j(x^*)^T \tilde{s} l + \frac{1}{2} l^2 \tilde{s}^T \nabla^2 f_j(x^*) \tilde{s} < \hat{t} < 0, \quad j = 1, \dots, m. \quad (5)$$

Then, for (4), (5) and the fact that s is a descent direction for F at x^* , $\theta(x^*) < 0$.
(ii) \Rightarrow (i) If $\theta(x^*) < 0$ then $\exists(t, s)$, feasible for (3), with $x = x^*$, such that

$$\nabla f_j(x^*)^T s \leq t < 0 \quad j = 1, \dots, m,$$

so $\mathcal{R}(\nabla F(x^*)) \cap (-\mathbb{R}_{++}^m) \neq \emptyset$, i.e. x^* is not critical.

(ii) \Rightarrow (iii) Assume that $s(x^*) = 0$, then

$$\begin{aligned} \theta(x^*) &= \min_{s.t.} t \\ &\quad \nabla f_j(x^*)^T s + \frac{1}{2} s^T \nabla^2 f_j(x^*) s - t \leq 0, \quad j = 1, \dots, m, \\ &\quad \nabla f_j(x^*)^T s - t \leq 0, \quad j = 1, \dots, m, \\ &\quad \|s\| \leq \Delta, \\ &\quad (t, s) \in \mathbb{R} \times \mathbb{R}^n \\ &= \min_{s.t.} t \\ &\quad -t \leq 0, t \in \mathbb{R} \\ &= 0. \end{aligned}$$

This contradicts $\theta(x^*) < 0$. Then $s(x^*) \neq 0$.

Finally, assuming that F is \mathbb{R}^m -convex we prove (iii) \Rightarrow (i). Due to $\nabla^2 f_j(x^*)$ are positive definite for $j = 1, \dots, m$, and $s = s(x^*) \neq 0$ we have

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

for $j = 1, \dots, m$, then $\mathcal{R}(\nabla F(\bar{x})) \cap (-\mathbb{R}_{++}^m) \neq \emptyset$, therefore, x^* is not critical. \square

The relation (iii) \Rightarrow (ii) is not true in the nonconvex case, as it can be seen in the next example.

Example 2 Consider $\min_{s.t.} [\cos(x_1) + e^{x_2} - x_2 \quad -\cos(x_2) - e^{x_1} - x_1]^T$.

The associate TR-subproblem (3) at the point $x = (0 \ 0)^T$ is

$$\begin{aligned} &\min_{s.t.} t \\ &[0 \ 0]s + \frac{1}{2}s^T \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} s - t \leq 0, \\ &[0 \ 0]s - t \leq 0, \\ &\|s\| \leq \Delta, \\ &(t, s) \in \mathbb{R} \times \mathbb{R}^2. \end{aligned}$$

Clearly, the solutions are $t = 0$ and $s \in \mathbb{R}^2$ such that $|s_2| \leq |s_1|$. Thus, there are solutions with $s \neq 0$ and $t = 0$.

3 The trust region algorithm

Algorithms which use trust regions establish a quadratic model around the current iterate and later solve a subproblem. This subproblem consists in finding the minimizer of the quadratic model in a compact set, the trust region. We will propose an algorithm defining a quadratic model for each objective function:

$$q_k^j(s) = \nabla f_j(x_k)^T s + \frac{1}{2}s^T \nabla^2 f_j(x_k) s.$$

In contrast to the scalar case we do not minimize the quadratic models but the maximum of them in a descent direction. In each iterate x_k , the next algorithm solve the trust region subproblem (3), whose solution allows us to define the next iterate x_{k+1} . The trial step s_k is obtained like the component s from the solution of the subproblem (3) in $x = x_k$. In the steps 3 and 4 of the next algorithm a criterion for evaluating and accepting the step is presented, and, in the step 5, a rule for updating the trust region radius is established.

Algorithm 1 Set $x_0 \in \mathbb{R}^n$, $\Delta_0 > 0$, $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 < \gamma_2 < 1$, $tol > 0$ and $k = 0$

1. Evaluate $\nabla f_j(x_k)$, $\nabla^2 f_j(x_k)$, $j = 1, \dots, m$, and solve the subproblem (3) with $x = x_k$ and $\Delta = \Delta_k$, in order to obtain (t_k, s_k) .
2. If $|t_k| < tol$, stop.
3. Set

$$\rho_k^j = \frac{f_j(x_k) - f_j(x_k + s_k)}{q_k^j(0) - q_k^j(s_k)}.$$

4. If $\rho_k^j \geq \eta_1 \ \forall j$, $x_{k+1} = x_k + s_k$; otherwise $x_{k+1} = x_k$.

5. Update the trust region radius:

$$\Delta_{k+1} \in \begin{cases} (\Delta_k, \infty) & \text{if } \eta_2 \leq \rho_k^j \forall j. \\ (\gamma_2 \Delta_k, \Delta_k] & \text{if } \eta_1 \leq \rho_k^j \forall j \text{ and } \exists l \text{ such that } \rho_k^l < \eta_2. \\ [\gamma_1 \Delta_k, \gamma_2 \Delta_k] & \text{if } \exists l \text{ such that } \rho_k^l \leq \eta_1. \end{cases}$$

6. $k \leftarrow k + 1$.

In scalar trust region algorithms, at each iteration, the trial step is evaluated using the agreement between the functional and the model reduction around x_k , (i.e: ρ_k). In our algorithm m quotients: ρ_k^j , $j = 1, \dots, m$, one for each objective function, were used. We said that the k -th iteration is *very successful* if $\rho_k^j \geq \eta_2$ for all $j = 1, \dots, m$, and, in such case, we increase the trust region radius, i.e. $\Delta_{k+1} \in (\Delta_k, \infty)$. We say that the k -th iteration is *successful* if $\rho_k^j \geq \eta_1$ for all $j = 1, \dots, m$, and there are some index l such that $\rho_k^l \leq \eta_2$, in such case, we update the new trust region radius taking $\Delta_{k+1} \in (\gamma_2 \Delta_k, \Delta_k]$. In other case, we said that the k -th iteration is *unsuccessful* and we reduce the trust region radius, i.e. $\Delta_{k+1} \in [\gamma_1 \Delta_k, \gamma_2 \Delta_k]$.

It is important to note that, due to $\theta(x) \leq 0$, if x_k is not a critical point, the direction s_k obtained by the subproblem (3) is a descent direction for F at x_k . Moreover, the stopping criteria established in the 2nd step is suitable due to the continuity of $\theta(x)$. Furthermore, we have shown that the subproblem (3) is well defined.

To establish the well-definedness of the algorithm it is necessary to choose the step s_k among the vectors s such that $(\theta(x_k), s)$ be a solution of (3). After that, $q_k^j(0) - q_k^j(s_k) > 0$ for all $j = 1, \dots, m$, has to be proved at every noncritical x_k and it is necessary to prove that Algorithm 1 will accept a step in a finite number of inner iterations. This will be discussed in the Remarks 6 and 7.

4 Global convergence analysis

The global convergence analysis shares characteristics from the analysis developed for the scalar case and the hypothesis also are similar, see [4]. Assuming second order information, we will suppose uniformly boundness of the first and second order derivatives of F :

(A1) $F \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^m)$.

(A2) The Hessian matrices are uniformly bounded:

$$\left\| \nabla^2 f_j(x) \right\| \leq \kappa_H \quad \forall x \in \mathbb{R}^n, \quad \text{for } j = 1, \dots, m.$$

(A3) $f_j(x)$ is lower bounded for $j = 1, \dots, m$.

(A4) Exists $\Delta_{\max} > 0$ such that

$$\Delta_k \leq \Delta_{\max}, \quad \forall k \geq 0.$$

(A5) Exists $\kappa_G > 0$ such that

$$\|\nabla f_j(x)\| \leq \kappa_G \quad \forall x \in \mathbb{R}^n, \text{ for } j = 1, \dots, m.$$

Remark 2 The hypothesis (A3) can be assumed without loss of generality. See [19].

Remark 3 The hypothesis (A4) is an algorithmic assumption: it allows us to have an bound for all quadratics models $q_k^j, j = 1, \dots, m$. This assumption is necessary in order to the ratios $\rho_k^j, j = 1, \dots, m$, result to be well defined.

Remark 4 The hypothesis (A2) and (A5) might seem too strong but if we consider that the level sets of the functions $f_j, j = 1, \dots, m$, are bounded, this requirement is automatically satisfied because the continuity of ∇f_j and $\nabla^2 f_j, j = 1, \dots, m$, and the fact that our method is a descent method [4].

In the scalar case, the strategy consists in obtaining that the reduction produced by a step be at least a fraction of the reduction produced by the Cauchy step. In the vectorial case, since it does not have a Cauchy step with the optimal length for all the functionals, the strategy consists in estimating the optimal reduction for each of the models with a descent direction different of the steepest descent for it. The bound for the reduction predicted by a quadratical model q_k^j at x_k in a descent direction for F is given by the next lemma.

Lemma 2 *Let $v \in \mathbb{R}^n$ a descent direction for F at x_k , a noncritical point. For each $j = 1, \dots, m$, there exists $\bar{t} > 0$, such that $\|\bar{t}v\| \leq \Delta_k$ and*

$$q_k^j(\bar{t}v) \leq q_k^j(tv), \quad \forall t > 0 \text{ such that } \|tv\| \leq \Delta_k. \tag{6}$$

Furthermore

$$q_k^j(0) - q_k^j(\bar{t}v) \geq -\frac{1}{2} \frac{\nabla f_j(x_k)^T v}{\|v\|} \min \left\{ -\frac{\nabla f_j(x_k)^T v}{\|v\| \kappa_H}, \Delta_k \right\}.$$

Proof The existence of $\bar{t} > 0$ is straightforward. Let us consider the function $\phi(t) : [0, \Delta_k/\|v\|] \rightarrow \mathbb{R}$ defined as

$$\phi(t) = q_k^j(tv).$$

The function ϕ is continuous in a compact set, $[0, \Delta_k/\|v\|]$, then there exists $\bar{t} \in [0, \Delta_k/\|v\|]$ such that

$$q_k^j(\bar{t}v) \leq q_k^j(tv), \quad \forall t > 0 \text{ such that } \|tv\| \leq \Delta_k.$$

Furthermore, $\bar{t} > 0$ because $\phi'(0) = \nabla f_j(x_k)^T v < 0$, due to v is a descent direction for f_j at x_k .

In the second part, for each quadratic model let us consider the cases in which the curvature along the line defined by v is positive or not. If v is a direction such that

$$v^T \nabla^2 f_j(x_k) v > 0, \quad (7)$$

the model has a minimizer on the line tv :

$$t^* = \frac{-\nabla f_j(x_k)^T v}{v^T \nabla^2 f_j(x_k) v} > 0.$$

Then, if the minimizer is inside the trust region ($\|t^*v\| \leq \Delta_k$), we took $\bar{t} = t^*$, then

$$\begin{aligned} q_k^j(\bar{t}v) &= \frac{-\nabla f_j(x_k)^T v}{v^T \nabla^2 f_j(x_k) v} \nabla f_j(x_k)^T v + \frac{1}{2} \left(\frac{-\nabla f_j(x_k)^T v}{v^T \nabla^2 f_j(x_k) v} \right)^2 v^T \nabla^2 f_j(x_k) v \\ &= -\frac{1}{2} \frac{(\nabla f_j(x_k)^T v)^2}{v^T \nabla^2 f_j(x_k) v}. \end{aligned}$$

Now, using the hypothesis (A2) and the fact $q_k^j(0) = 0$, we obtain

$$q_k^j(0) - q_k^j(\bar{t}v) = \frac{1}{2} \frac{(\nabla f_j(x_k)^T v)^2}{v^T \nabla^2 f_j(x_k) v} \geq \frac{1}{2} \frac{(\nabla f_j(x_k)^T v)^2}{\|v\|^2 \kappa_H}. \quad (8)$$

On the other hand, if $\|t^*v\| > \Delta_k$, the minimizer along the line tv is outside the trust region, then \bar{t} is defined as $\bar{t} = \frac{\Delta_k}{\|v\|}$, then

$$q_k^j(\bar{t}v) = \frac{\Delta_k}{\|v\|} \nabla f_j(x_k)^T v + \frac{1}{2} \left(\frac{\Delta_k}{\|v\|} \right)^2 v^T \nabla^2 f_j(x_k) v. \quad (9)$$

From $\|t^*v\| > \Delta_k$, we have $\frac{-\nabla f_j(x_k)^T v}{v^T \nabla^2 f_j(x_k) v} \|v\| = \|t^*v\| > \Delta_k = \bar{t} \|v\|$ and, for (7),

$$\frac{\nabla f_j(x_k)^T v}{\bar{t}} < -v^T \nabla^2 f_j(x_k) v.$$

Therefore, using the last inequality, (9), the hypothesis and the fact $\|\bar{t}v\| = \Delta_k$,

$$\begin{aligned} q_k^j(0) - q_k^j(\bar{t}v) &> -\frac{\Delta_k}{\|v\|} \nabla f_j(x_k)^T v + \frac{1}{2} \left(\frac{\Delta_k}{\|v\|} \right)^2 \frac{\nabla f_j(x_k)^T v}{\bar{t}} \\ &= -\frac{\Delta_k}{\|v\|} \nabla f_j(x_k)^T v + \frac{1}{2} \frac{\Delta_k}{\|v\|} \nabla f_j(x_k)^T v \\ &= -\frac{1}{2} \frac{\Delta_k}{\|v\|} \nabla f_j(x_k)^T v. \end{aligned} \quad (10)$$

If it happens that $v^T \nabla^2 f_j(x_k) v \leq 0$, the minimum value of the model in the trust region is reached at its border, then we take $\bar{t} = \frac{\Delta_k}{\|v\|}$ and

$$q_k^j(\bar{t}v) = \frac{\Delta_k}{\|v\|} \nabla f_j(x_k)^T v + \frac{1}{2} \frac{\Delta_k^2}{\|v\|^2} v^T \nabla^2 f_j(x_k) v \leq \frac{\Delta_k}{\|v\|} \nabla f_j(x_k)^T v. \tag{11}$$

Then, since v is a descent direction for F at x_k ,

$$q_k^j(0) - q_k^j(\bar{t}v) \geq -\frac{\Delta_k}{\|v\|} \nabla f_j(x_k)^T v \geq -\frac{1}{2} \frac{\Delta_k}{\|v\|} \nabla f_j(x_k)^T v, \tag{12}$$

where the first inequality due to (11).

Summarizing, for (8), (10), (12), and using again that v is a descent direction for F at x_k ,

$$q_k^j(0) - q_k^j(\bar{t}v) \geq -\frac{1}{2} \frac{\nabla f_j(x_k)^T v}{\|v\|} \min \left\{ -\frac{\nabla f_j(x_k)^T v}{\|v\| \kappa_H}, \Delta_k \right\}.$$

□

It is important to note that the previous lemma is true for each $f_j, j = 1, \dots, m$, because v is a descent direction for F at x_k , but, naturally, at first glance, the value of \bar{t} will be different for each function $f_j, j = 1, \dots, m$.

Remark 5 The reduction obtained in the quadratic models q_k^j using the direction s_k given by the solution of the subproblem (3) can be bounded below using the bound given by the previous lemma. For that, let us observe $q_k^j(s_k) \leq t_k$, for $j = 1, \dots, m$, for the definition of the subproblem (3). If Algorithm 1 does not terminate at the second step, $-t_k = |t_k| \geq tol$

$$q_k^j(0) - q_k^j(s_k) \geq -t_k \geq tol, \text{ for } j = 1, \dots, m. \tag{13}$$

On the other hand, the assumptions (A2), (A4) and (A5) allow us to obtain the bound

$$\begin{aligned} q_k^j(0) - q_k^j(\bar{t}s_k) &= -\bar{t} \nabla f_j(x_k)^T s_k - \frac{1}{2} \bar{t}^2 s_k^T \nabla^2 f_j(x_k) s_k \\ &\leq \|\nabla f_j(x_k)\| \|\bar{t}s_k\| + \frac{1}{2} \|\bar{t}s_k\|^2 \kappa_H \\ &\leq \kappa_G \Delta_{\max} + \frac{1}{2} \Delta_{\max}^2 \kappa_H, \end{aligned} \tag{14}$$

where $\bar{t} = \bar{t}_j$ satisfies (6).

So, using (13) and (14), we obtain

$$\frac{q_k^j(0) - q_k^j(s_k)}{q_k^j(0) - q_k^j(\bar{t}s_k)} \geq \frac{tol}{\kappa_G \Delta_{\max} + \frac{1}{2} \Delta_{\max}^2 \kappa_H} = \beta.$$

Using Lemma 2, we can claim that for every x_k and s_k generated by Algorithm 1, we have

$$q_k^j(0) - q_k^j(s_k) \geq -\beta \frac{1}{2} \frac{\nabla f_j(x_k)^T s}{\|s\|} \min \left\{ -\frac{\nabla f_j(x_k)^T s}{\|s\| \kappa_H}, \Delta_k \right\} > 0, \quad (15)$$

for $j = 1, \dots, m$, except when x_k is a critical point.

Remark 6 It is important to observe that the bound (15) is greater than zero then the ratios ρ_k^j are well defined.

The next result establishes the conditions to accept a step and enlarge the trust region radius.

Lemma 3 Assuming (A2), if s_k is solution of (3) such that $|\nabla f_j(x_k)^T s_k| \neq 0$ and

$$\Delta_k \leq \frac{\beta |\nabla f_j(x_k)^T s_k| (1 - \eta_2)}{2\kappa_H \|s_k\|}, \quad (16)$$

for all $j = 1, \dots, m$, then the k -th iteration is successful and $\Delta_{k+1} \geq \Delta_k$.

Proof From the fact that $0 < \eta_2 < 1$, $\frac{1}{2}(1 - \eta_2) < 1$, and, for (16), we have

$$\Delta_k < \frac{\beta |\nabla f_j(x_k)^T s_k|}{\kappa_H \|s_k\|}.$$

Then, it is possible to rewrite the bound for the decreased observed in the models (15):

$$q_k^j(0) - q_k^j(s_k) \geq -\beta \frac{1}{2} \frac{\nabla f_j(x_k)^T s_k}{\|s_k\|} \Delta_k. \quad (17)$$

On the other hand

$$|\rho_k^j - 1| = \left| \frac{f_j(x_k) - f_j(x_k + s_k)}{-q_k^j(s_k)} - 1 \right| = \left| \frac{f_j(x_k) - f_j(x_k + s_k) + q_k^j(s_k)}{-q_k^j(s_k)} \right|.$$

Expanding f_j by Taylor series around x_k ; there exists $\xi = x_k + \lambda s_k$, with $\lambda \in (0, 1)$ so that

$$\begin{aligned} |\rho_k^j - 1| &= \left| \frac{f_j(x_k) - f_j(x_k) - \nabla f_j(x_k)^T s_k - \frac{1}{2} s_k^T \nabla^2 f_j(\xi) s_k + q_k^j(s_k)}{-q_k^j(s_k)} \right| \\ &= \left| \frac{-\frac{1}{2} s_k^T \nabla^2 f_j(\xi) s_k + \frac{1}{2} s_k^T \nabla^2 f_j(x_k) s_k}{-q_k^j(s_k)} \right| \\ &= \frac{1}{2} \frac{|s_k^T (\nabla^2 f_j(\xi) - \nabla^2 f_j(x_k)) s_k|}{|-q_k^j(s_k)|}. \end{aligned}$$

Finally, by the Cauchy-Schwartz inequality:

$$\left| \rho_k^j - 1 \right| \leq \frac{1}{2} \frac{\|s_k\|^2 \|\nabla^2 f_j(\xi) - \nabla^2 f_j(x_k)\|}{\left| q_k^j(0) - q_k^j(s_k) \right|}.$$

The numerator can be bounded by using the assumption (A2) and since s_k is a descent direction and it is inside of the trust region, by (17):

$$\left| \rho_k^j - 1 \right| \leq \frac{1}{2} \frac{\Delta_k^2 2\kappa_H}{\left| q_k^j(0) - q_k^j(s_k) \right|} \leq \frac{\Delta_k^2 \kappa_H}{-\beta \frac{1}{2} \frac{\nabla f_j(x_k)^T s_k}{\|s_k\|} \Delta_k},$$

then by using again the fact that s_k is a descent direction it is possible to write

$$\begin{aligned} \left| \rho_k^j - 1 \right| &\leq \frac{2\Delta_k \|s_k\| \kappa_H}{-\beta \nabla f_j(x_k)^T s_k} = \frac{2\Delta_k \|s_k\| \kappa_H}{\beta \left| \nabla f_j(x_k)^T s_k \right|} \\ &< 1 - \eta_2, \end{aligned}$$

where the last inequality due to (16). Therefore we obtain $\rho_k^j > \eta_2$ for $j = 1, \dots, m$, then, considering the 5th. step of Algorithm 1, the step is accepted and the trust radius is increased, except in case that $\Delta_k = \Delta_{\max}$, so $\Delta_{k+1} \geq \Delta_k$. □

Remark 7 From this lemma we can claim that a step will be accepted in a finite number of inner iterations; Let's observe that if x_k is not critical, then $\left\{ \frac{|\nabla f_j(x_k)^T s_k|}{\|s_k\|} : j = 1, \dots, m \right\}$, will be bounded below by $\min_{j=1, \dots, m} \{ \|\nabla f_j(x_k)\| \} > 0$. Therefore, after a finite numbers of inner iterations, in which the step is rejected and the trust radius reduced, the inequality (16) will be satisfied and the step accepted.

Lemma 4 *Assuming that there exists $\sigma > 0$ such that $\frac{|\nabla f_j(x_k)^T s_k|}{\|s_k\|} > \sigma$ for all indices k and j , then there exists Δ_{\min} so that $\Delta_k > \Delta_{\min}$ for all k .*

Proof Let us suppose a subsequence of trust region radius $\{\Delta_{k_i}\}$ such that $\Delta_{k_i} \rightarrow 0$ where the k th iterate is the first one such that

$$\Delta_{k+1} \leq \frac{\gamma_1 \sigma \beta (1 - \eta_2)}{2\kappa_H}. \tag{18}$$

Since k is the the first index satisfying the inequality $\gamma_1 \Delta_k \leq \Delta_{k+1}$, we have

$$\Delta_k \leq \frac{\sigma \beta (1 - \eta_2)}{2\kappa_H} \leq \frac{1 - \eta_2}{2\kappa_H} \frac{\beta \left| \nabla f_j(x_k)^T s_k \right|}{\|s_k\|},$$

where the second inequality due to the hypothesis of this lemma. Then, for Lemma 3, the previous iteration was very successful and the radius must increase, which is a

contradiction with the assumption that k is the first index that satisfies (18). Therefore there does not exist k which satisfies (18) so that

$$\Delta_{k+1} > \frac{\gamma_1 \sigma \beta (1 - \eta_2)}{2\kappa_H} = \Delta_{\min} \quad \forall k.$$

□

Theorem 1 *Assuming there are a finite number of successful iterations, then $x_k = x^*$ for all k large enough and x^* is a critical point.*

Proof Following the algorithm $x^* = x_k = x_{k+l}$ for all $l > 0$ where k is the last index of a successful iterate then, since the next iterations are rejected $\Delta_k \rightarrow 0$.

If for all $j = 1, \dots, n$, $|\nabla f_j(x_k)^T s_k| \neq 0$, by Lemma 4, some future iterate must be successful, contradicting the fact that x_k was the last successful iterate, hence $|\nabla f_j(x_k)^T s_k| = 0$, and this implies that x_k is a critical point. □

Theorem 2 *Let $\{x_k\}$ and $\{s_k\}$ generated by Algorithm 1, then*

$$\liminf_{k \rightarrow \infty} \max_{j=1, \dots, m} \left\{ \frac{|\nabla f_j(x_k)^T s_k|}{|s_k|} \right\} = 0.$$

Proof Let us suppose that exists ϵ , such that $\forall k$ there exist $j \in \{1, \dots, m\}$ such that

$$\frac{|\nabla f_j(x_k)^T s_k|}{\|s_k\|} \geq \epsilon. \quad (19)$$

Let \mathcal{S} be the index set of successful iterations, then, for $k \in \mathcal{S}$,

$$\begin{aligned} f_j(x_k) - f_j(x_{k+1}) &\geq \eta_1 \left(q_k^j(0) - q_k^j(s_k) \right) \\ &\geq \eta_1 \beta \left(-\frac{1}{2} \frac{\nabla f_j(x_k)^T s_k}{\|s_k\|} \min \left\{ -\frac{\nabla f_j(x_k)^T s_k}{\|s_k\| \kappa_H}, \Delta_k \right\} \right), \end{aligned}$$

where the first inequality due to the iteration was successful and the second to the bound for the expected descent (15). Then, due to s_k being a descent direction and by (19)

$$\begin{aligned} f_j(x_k) - f_j(x_{k+1}) &\geq \eta_1 \epsilon \beta \frac{1}{2} \min \left\{ -\frac{\nabla f_j(x_k)^T s_k}{\|s_k\| \kappa_H}, \Delta_k \right\} \\ &\geq \eta_1 \epsilon \beta \frac{1}{2} \min \left\{ \frac{\epsilon}{\kappa_H}, \Delta_{\min} \right\}, \end{aligned}$$

for $k \in \mathcal{S}$.

Then, summing over all the successful iterations up to the k -th index,

$$\begin{aligned} f_j(x_0) - f_j(x_{k+1}) &= \sum_{l=0, l \in \mathcal{S}}^k (f_j(x_l) - f_j(x_{l+1})) \\ &\geq \sum_{l=0, l \in \mathcal{S}}^k \eta_1 \epsilon \beta \frac{1}{2} \min \left\{ \frac{\epsilon}{\kappa_H}, \Delta_{\min} \right\} \\ &\geq \sigma_k \eta_1 \epsilon \beta \frac{1}{2} \min \left\{ \frac{\epsilon}{\kappa_H}, \Delta_{\min} \right\}, \end{aligned}$$

where σ_k is the number of successful iterations up to the k -th. Then if

$$\lim_{k \rightarrow \infty, k \in \mathcal{S}} \sigma_k < +\infty,$$

the Theorem 1 implies the first result. If

$$\lim_{k \rightarrow \infty} \sigma_k = \infty,$$

the difference $f_j(x_0) - f_j(x_{k+1})$ goes to infinite for (19) and Lemma 4, and f_j will not be bounded contradicting (A3). Therefore does not exist $\epsilon > 0$ that satisfies (19) for some $j \in \{1, \dots, m\}$. Then

$$\liminf_{k \rightarrow \infty} \max_{j=1, \dots, m} \left\{ \frac{|\nabla f_j(x_k)^T s_k|}{|s_k|} \right\} = 0.$$

□

Finally, if we assume bounded level sets for the function F we can claim the next corollary.

Corollary 1 *Let $\{x_k\}$ be generated by Algorithm 1, and assume that the level sets of the functions $f_j, j = 1, \dots, m$, are bounded, then $\{x_k\}$ converges to a Pareto critical point.*

Proof Set $A = \{x \in \mathbb{R}^n : f_j(x) \leq f_j(x_0), j = 1, \dots, m\}$. Due to the fact that the level sets of $f_j, j = 1, \dots, m$, are bounded, A is bounded. Furthermore, the set A is closed because $f_j, j = 1, \dots, m$, are continuous functions. Then A is compact. Therefore, the sequence $\{x_k\}$ generated by Algorithm 1 has an acumulation point x^* . Later, due to the Theorem 2, x^* is a critical point. □

5 The convex case and the local convergence analysis

In the convex case, obviously it is possible to obtain better results: at first glance the subproblem involves a convex feasible set, which makes it easier. Furthermore the theoretical results for the multiobjective problem are stronger.

In the local discussion we need to add the following assumption:

(A6) There exists $U \subseteq \mathbb{R}^n$ such that the sequence generated by Algorithm 1 $\{x_k\} \subset U$ and there exists $a > 0$ such that

$$a \|v\|^2 \leq v^T \nabla^2 f_j(x) v, \quad j = 1, \dots, m,$$

for all $x \in U$ and $v \in \mathbb{R}^n$.

This condition, clearly, implies the local strict convexity of the quadratical models q_k^j of the subproblem (3) for $j = 1, \dots, m$.

For solving nonlinear programs there are several optimality conditions [3] based on the Lagrangian function associated to the problem and many of them are established for the differentiable problem, so we will consider the problem

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & \nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s - t \leq 0, \quad j = 1, \dots, m \\ & \nabla f_j(x)^T s - t \leq 0, \quad j = 1, \dots, m \\ & \frac{\|s\|^2}{2} - \frac{\Delta^2}{2} \leq 0, \\ & (t, s) \in \mathbb{R} \times \mathbb{R}^n, \end{aligned} \quad (20)$$

which is differentiable and equivalent to (3).

For the subproblem (20) the associated Lagrangian function is

$$\begin{aligned} \mathcal{L}(t, s, \lambda^1, \lambda^2, \lambda) = & t + \sum_{j=1}^m \lambda_j^1 \left(\nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s - t \right) \\ & + \sum_{j=1}^m \lambda_j^2 \left(\nabla f_j(x)^T s - t \right) + \lambda \left(\frac{\|s\|^2}{2} - \frac{\Delta^2}{2} \right), \end{aligned} \quad (21)$$

with $\lambda^1, \lambda^2 \in \mathbb{R}^n, \lambda \in \mathbb{R}$.

Due to the convexity of the quadratical models q_k^j , the subproblem (20) satisfies the Slater constraint qualification [3], then, for the optimal value of (20) there exist the multipliers $\lambda_j^1, \lambda_j^2, \lambda$, with $j = 1, \dots, m$, satisfying the first order necessary conditions: The gradient of the Lagrangian function vanishes, the point is feasible and satisfies the complementary condition for the multipliers.

Remark 8 If x is not a critical point, from Lemma 1, $s(x) \neq 0$, and, from the strict convexity of the functionals f_j , $j = 1, \dots, m$, the Hessian matrices $\nabla^2 f_j$, $j = 1, \dots, m$, are positive definite. Then the linear constraints never will be active and, from the complementarity condition, the multipliers $\lambda_j^2 = 0$, for all $j = 1, \dots, m$.

The relationship between the variables and the multipliers occurs from the previous remark and the fact that the gradient vector of Lagrangian function vanishes at the optimal solution of (20).

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \sum_{j=1}^m \lambda_j^1 \left(\begin{bmatrix} 0 & 0 \\ 0 & \nabla^2 f_j(x) \end{bmatrix} \begin{bmatrix} 0 \\ s \end{bmatrix} + \begin{bmatrix} 0 \\ \nabla f_j(x) \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) + \lambda \begin{bmatrix} 0 \\ s \end{bmatrix} = 0,$$

which it can be written

$$1 - \sum_{j=1}^m \lambda_j^1 = 0,$$

$$\sum_{j=1}^m \lambda_j^1 \nabla^2 f_j(x) s + \sum_{j=1}^m \lambda_j^1 \nabla f_j(x) + \lambda s = 0.$$

Since $\nabla^2 f_j(x), j = 1, \dots, m$ are positive definite for all $x \in U$, after arithmetic calculations, we obtain

$$s = - \left[\sum_{j=1}^m \lambda_j^1 \nabla^2 f_j(x) + \lambda I \right]^{-1} \sum_{j=1}^m \lambda_j^1 \nabla f_j(x). \tag{22}$$

Therefore, we have obtained expressions of s and t as function of x, λ and λ^1 .

Remark 9 Let us observe that for obtaining (22) is enough that $\lambda_j^1 > 0$ for some $j = 1, \dots, m$, but this equation is valid if and only if $\lambda_j^2 = 0$ for all $j = 1, \dots, m$. Assuming strict convexity of $f_j(x)$ we can ensure that $\lambda_j^2 = 0$ for all $j = 1, \dots, m$. If the strict convexity assumption is dropped, it is not possible to ensure the uniqueness of the multipliers $\lambda_j^1, j = 1, \dots, m$, and the solution as well.

In order to discuss the rate of convergence of Algorithm 1 we establish some technical lemmas.

Lemma 5 *Let us assume that for any $\epsilon, \delta > 0$ such that $\|x - y\| < \delta$ then*

$$\left\| \nabla^2 f_j(x) - \nabla^2 f_j(y) \right\| < \epsilon \quad j = 1, \dots, m.$$

Then $\forall x, y$ such that $\|x - y\| < \delta$ the following inequalities hold

$$\left\| \nabla f_j(y) - \left(\nabla f_j(x) + \nabla^2 f_j(x)(y - x) \right) \right\| \leq \epsilon \|y - x\|,$$

and

$$\left| f_j(y) - \left(f_j(x) + \nabla f_j(x)^T (y - x) + \frac{1}{2} (y - x)^T \nabla^2 f_j(x) (y - x) \right) \right| \leq \frac{\epsilon}{2} \|y - x\|^2,$$

for all $j = 1, \dots, m$. If $\nabla^2 f_j$ is Lipschitz continuous with constant L for $j = 1, \dots, m$, then

$$\left\| \nabla f_j(y) - \left(\nabla f_j(x) + \nabla^2 f_j(x)(y - x) \right) \right\| \leq \frac{L}{2} \|y - x\|^2$$

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

Proof This is easy applying Lemmas 4.1.10 and 4.1.12 from [8]. \square

The next lemma establishes the relationship between the absolute value of the solution of the subproblem, $|\theta(x)|$, and the size of the step, $\|s(x)\|$.

Lemma 6 *Let $U \subseteq \mathbb{R}^n$ and $a > 0$ satisfying (A6), then*

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

Proof We should analyze two different cases according to the trust region constraint being active or not. In both cases, the term associated to the trust region in the Lagrangian function (21) vanishes, for this reason we develop only the case when the trust region is inactive. The another case is analogous.

Let us consider the case $\|s\| < \Delta$. The complementarity conditions and Remark 8 implies:

$$\begin{aligned} \mathcal{L}(t, s, \lambda^1, \lambda^2, \lambda) &= t + \sum_{j=1}^m \lambda_j^1 \left(\nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s - t \right) \\ &= t. \end{aligned}$$

From (22) and $\lambda = 0$, we obtain

$$\sum_{j=1}^m \lambda_j^1 \nabla f_j(x) = - \left[\sum_{j=1}^m \lambda_j^1 \nabla^2 f_j(x) \right] s.$$

Then, using the last two equalities,

$$\begin{aligned} \mathcal{L}(t, s, \lambda^1, \lambda^2, \lambda) &= \left(- \left[\sum_{j=1}^m \lambda_j^1 \nabla^2 f_j(x) \right] s \right)^T s + \frac{1}{2} \sum_{j=1}^m \lambda_j^1 s^T \nabla^2 f_j(x) s \\ &= -\frac{1}{2} \sum_{j=1}^m \lambda_j^1 s^T \nabla^2 f_j(x) s \\ &= t, \end{aligned} \tag{23}$$

where the equality (23) due to $\nabla^2 f_j(x)$ is symmetric, because $f_j \in \mathcal{C}^2(\mathbb{R}^n)$ for every $j = 1, \dots, m$.

Then $t = -\frac{1}{2} s^T \left(\sum_{j=1}^m \lambda_j^1 \nabla^2 f_j \right) s < 0$, therefore, using the assumption (A6),

$$|t| \geq \frac{1}{2} s^T \left(\sum_{j=1}^m \lambda_j^1 \nabla^2 f_j \right) s \geq \frac{a}{2} \|s\|^2.$$

\square

The following lemma gives a bound to the subproblem in term of the gradients of the objective functions.

Lemma 7 *Let $x \in U$ and $a > 0$ satisfying (A6), then*

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

for every $\lambda_j \geq 0, j = 1, \dots, m$, such that $\sum_{j=1}^m \lambda_j = 1$, and for all $x \in U$.

Proof For this lemma, it is useful to consider the dual problem of (3)

$$\max_{(t,s)} \left\{ t + \sum_{j=1}^m \lambda_j^1 (\nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s - t) + \sum_{j=1}^m \lambda_j^2 (\nabla f_j(x)^T s - t) + \lambda \left(\frac{\|s\|^2}{2} - \frac{\Delta^2}{2} \right) \right\}.$$

For (A6), we have that all the constraints of (20) are convex and the Slater constraint qualification is satisfied, therefore, there is not duality gap. It allows us to ensure that the optimal value of the primal problem $\theta(x)$ is the optimal value of the dual problem

$$\theta(x) = \max_{(t,s)} \left\{ t + \sum_{j=1}^m \lambda_j^1 (\nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s - t) + \sum_{j=1}^m \lambda_j^2 (\nabla f_j(x)^T s - t) + \lambda \left(\frac{\|s\|^2}{2} - \frac{\Delta^2}{2} \right) \right\}.$$

If we add the constraint $\sum_{j=1}^m \lambda_j^1 = 1, \lambda_j^1, \lambda \geq 0$ and $\lambda_j^2 = 0$, we have

$$\theta(x) \geq \max_{s,a} \left\{ \inf_{(t,s)} \left\{ t + \sum_{j=1}^m \lambda_j^1 (\nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s - t) + \lambda \left(\frac{\|s\|^2}{2} - \frac{\Delta^2}{2} \right) \right\} \right. \\ \left. \lambda_j^1, \lambda \geq 0, \sum_{j=1}^m \lambda_j^1 = 1. \right\}$$

Now, using the constraint that we impose and the infimum properties we have

$$\theta(x) = \sup_{\substack{\lambda_j^1, \lambda \geq 0, \\ \sum_{j=1}^m \lambda_j^1 = 1}} \inf_s \left\{ \sum_{j=1}^m \lambda_j^1 \left(\nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s \right) + \lambda \left(\frac{\|s\|^2}{2} - \frac{\Delta^2}{2} \right) \right\} \\ \geq \sup_{\substack{\lambda_j^1, \lambda \geq 0, \\ \sum_{j=1}^m \lambda_j^1 = 1}} \left\{ \inf_s \left\{ \sum_{j=1}^m \lambda_j^1 \left(\nabla f_j(x)^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s \right) \right\} \right. \\ \left. + \lambda \inf_s \left\{ \frac{\|s\|^2}{2} - \frac{\Delta^2}{2} \right\} \right\}.$$

The minimum in the second term is obtained taking $s = 0$,

$$\theta(x) \geq \sup_{\substack{\lambda_j^1, \lambda \geq 0, \\ \sum_{j=1}^m \lambda_j^1 = 1}} \left\{ \inf_s \left\{ \sum_{j=1}^m \lambda_j^1 (\nabla f_j(x))^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s \right\} - \lambda \frac{\Delta^2}{2} \right\}$$

Because $-\frac{\Delta^2}{2} < 0$, the supremum must be $\lambda = 0$, then

$$\theta(x) \geq \sup_{\lambda_j^1 \geq 0, \sum_{j=1}^m \lambda_j^1 = 1} \inf_s \left\{ \sum_{j=1}^m \lambda_j^1 (\nabla f_j(x))^T s + \frac{1}{2} s^T \nabla^2 f_j(x) s \right\}$$

Using the hypothesis (A6),

$$\theta(x) \geq \sup_{\lambda_j^1 \geq 0, \sum_{j=1}^m \lambda_j^1 = 1} \inf_s \sum_{j=1}^m \lambda_j^1 \frac{1}{2} a \|s\|^2 + \sum_{j=1}^m \lambda_j^1 \nabla f_j(x)^T s. \quad (24)$$

Note that the last infimum is, actually, a minimum at the zero of the gradient, because is taken over a convex function. Then, we obtain the gradient and find such value of s

$$\begin{aligned} \sum_{j=1}^m \lambda_j^1 a s + \sum_{j=1}^m \lambda_j^1 \nabla f_j &= 0, \\ s &= -\frac{\sum_{j=1}^m \lambda_j^1 \nabla f_j}{\sum_{j=1}^m \lambda_j^1 a}. \end{aligned}$$

Replacing it in (24)

$$\begin{aligned} \theta(x) &\geq \sup_{\substack{\lambda_j^1 \geq 0, \\ \sum_{j=1}^m \lambda_j^1 = 1}} \sum_{j=1}^m \lambda_j^1 \frac{1}{2} a \left\| \frac{\sum_{j=1}^m \lambda_j^1 \nabla f_j(x)}{\sum_{j=1}^m \lambda_j^1 a} \right\|^2 \\ &\quad - \left(\sum_{j=1}^m \lambda_j^1 \nabla f_j(x) \right)^T \frac{\sum_{j=1}^m \lambda_j^1 \nabla f_j(x)}{\sum_{j=1}^m \lambda_j^1 a} \\ &= \sup_{\substack{\lambda_j^1 \geq 0, \\ \sum_{j=1}^m \lambda_j^1 = 1}} -\frac{1}{2} \frac{\left\| \sum_{j=1}^m \lambda_j^1 \nabla f_j(x) \right\|^2}{\sum_{j=1}^m \lambda_j^1 a} \\ &\geq -\frac{1}{2} \frac{\left\| \sum_{j=1}^m \lambda_j^1 \nabla f_j(x) \right\|^2}{a}, \end{aligned}$$

where the last inequality is due to $\sum_{j=1}^m \lambda_j^1 = 1$.

Finally, taking in account that $\theta(x) \leq 0$ and with $\lambda_j^1 = \lambda_j$, we obtain

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

$\forall \lambda_j \geq 0, j = 1, \dots, m$, such that $\sum_{j=1}^m \lambda_j = 1$. □

Remark 10 The bound giving by the previous lemma is valid not only for the multipliers associated to the subproblem but also to any convex combination.

Lemma 8 Let $U \subseteq \mathbb{R}^n, \hat{x}$ and $\hat{x}_+ = \hat{x} + s(\hat{x}), a, r, \delta, \epsilon > 0$ such that

- (a) $U \subseteq \mathbb{R}^n$ and $a > 0$ satisfies (A6),
- (b) $B(\hat{x}, r) \subset U$,
- (c) $\|\nabla^2 f_j(x) - \nabla^2 f_j(y)\| \leq \epsilon \forall x, y \in B(\hat{x}, r)$ such that $\|x - y\| < \delta$,
- (d) $\|s(\hat{x}_+)\| < \min\{\delta, r\}$,

then

$$\|s(\hat{x}_+)\| \leq \frac{\epsilon}{a} \|\hat{s}\|.$$

Moreover, if $\nabla^2 f_j$ are Lipschitz continuous with constant L in $B(\hat{x}, r)$:

$$\|\nabla^2 f_j(x) - \nabla^2 f_j(y)\| \leq L \|x - y\| \quad \forall x, y \in B(\hat{x}, r),$$

then

$$\|s(\hat{x}_+)\| \leq \frac{L}{2a} \|\hat{s}\|^2.$$

Proof Let $\hat{\lambda}_j^1, \hat{\lambda}_j^2, \hat{\lambda} > 0, j = 1, \dots, m$, the multipliers associated to the subproblem (3) with $x = \hat{x}$.

From the convexity of the functionals $f_j, j = 1, \dots, m$, and since the trust region constraint is inactive, the multipliers associated to them are all zero but $\hat{\lambda}_j^1$, then $\sum_{j=1}^m \hat{\lambda}_j^1 = 1$.

The Lemma 7 allows to write

$$|\theta(\hat{x}_+)| \leq \frac{1}{2a} \left\| \sum_{j=1}^m \hat{\lambda}_j^1 \nabla f_j(\hat{x}_+) \right\|^2. \tag{25}$$

On the other hand, let us define

$$G(x) = \sum_{j=1}^m \hat{\lambda}_j^1 f_j(x),$$

then

$$\nabla G(\widehat{x}_+) = \nabla G(\widehat{x} + s(\widehat{x})) = \sum_{j=1}^m \widehat{\lambda}_j^1 \nabla f_j(\widehat{x} + s(\widehat{x})),$$

and also

$$\nabla^2 G(\widehat{x}) = \sum_{j=1}^m \widehat{\lambda}_j^1 \nabla^2 f_j(\widehat{x}). \quad (26)$$

From (22) with $x = \widehat{x}$ and the fact $\lambda = 0$, we can write

$$\widehat{s} = s(\widehat{x}) = - \left[\sum_{j=1}^m \widehat{\lambda}_j^1 \nabla^2 f_j(\widehat{x}) \right]^{-1} \sum_{j=1}^m \widehat{\lambda}_j^1 \nabla f_j(\widehat{x}) = -\nabla^2 G(\widehat{x})^{-1} \nabla G(\widehat{x}).$$

Since $\nabla^2 G$ is a convex combination of the $\nabla^2 f_j$, $j = 1, \dots, m$, it inherits the uniform continuity, therefore Lemma 5 can be used having in mind the last equation,

$$\left\| \nabla G(\widehat{x} + \widehat{s}) - \left[\nabla G(\widehat{x}) + \nabla^2 G(\widehat{x}) \widehat{s} \right] \right\| \leq \epsilon \|\widehat{s}\|,$$

thus, by (26),

$$\|\nabla G(\widehat{x}_+)\| \leq \epsilon \|\widehat{s}\|.$$

Now, relating the last inequality with (25) and Lemma 6, we have

$$\frac{a}{2} \|\widehat{s}_+\|^2 \leq |\theta(\widehat{x}_+)| \leq \frac{1}{2a} (\epsilon \|\widehat{s}\|)^2,$$

then

$$\|\widehat{s}_+\| \leq \frac{\epsilon}{a} \|\widehat{s}\|.$$

If in addition $\nabla^2 f_j$, $j = 1, \dots, m$, are Lipschitz continuous with constant L , also $\nabla^2 G$ is Lipschitz continuous with the same constant. Then, once again Lemma 5 and (25) allow us to write,

$$\begin{aligned} |\theta(\widehat{x}_+)| &\leq \frac{1}{2a} \|\nabla G(\widehat{x}_+)\|^2 = \frac{1}{2a} \left\| \nabla G(\widehat{x} + \widehat{s}) - \left[\nabla G(\widehat{x}) + \nabla^2 G(\widehat{x}) \widehat{s} \right] \right\|^2 \\ &\leq \frac{1}{4a} L^2 \|\widehat{s}\|^4, \end{aligned}$$

where the inequality follows from (26).

Then, using Lemma 6,

$$\frac{a}{2} \|\widehat{s}_+\|^2 \leq \frac{1}{2a} \frac{L^2}{4} \|\widehat{s}\|^4,$$

meaning

$$\|\widehat{s}_+\| \leq \frac{L}{2a} \|\widehat{s}\|^2.$$

□

The next theorem guarantees the convergence to an optimal Pareto point in a neighborhood of any starting point. No assuming the existence of such a point is an interesting characteristic of the result which has the flavor of the Kantorovich’s theorem for the scalar case.

Theorem 3 *Let be $\{x_k\}$ the sequence generated by Algorithm 1, and let be $U \subseteq \mathbb{R}^m$, $a, r, \delta, \epsilon > 0$ such that*

- (a) $U \subseteq \mathbb{R}^n$ and $a > 0$ satisfies the assumption (A6),
- (b) $\|\nabla^2 f_j(x) - \nabla^2 f_j(y)\| \leq \epsilon, \forall x, y \in U$, such that $\|x - y\| < \delta$,
- (c) $\frac{\epsilon}{a} < 1 - \eta_2$, where η_2 is the parameter to consider very successful an iteration in Algorithm 1,
- (d) $B(x_0, r) \subseteq U$,
- (e) $\|s(x_0)\| \leq \min\{\delta, r(1 - \frac{\epsilon}{a})\}$,

then

- (i) $\|x_k - x_0\| \leq \|s(x_0)\| \frac{1 - (\frac{\epsilon}{a})^k}{1 - \frac{\epsilon}{a}} \forall k \geq 0$.
- (ii) $\|s(x_k)\| \leq \|s(x_0)\| (\frac{\epsilon}{a})^k \forall k \geq 0$.
- (iii) $\Delta_{k+1} \geq \Delta_k \forall k \geq 0$.
- (iv) $\|s(x_{k+1})\| \leq \|s(x_k)\| (\frac{\epsilon}{a}) \forall k \geq 0$.

Moreover the sequence $\{x_k\}$ converges to some local Pareto point.

Proof First we show that if items (i) and (ii) hold for some k , then items (iii) and (iv) also hold for that k .

Let us assume that (i) and (ii) hold for k and together with the triangular inequality,

$$\begin{aligned} \|x_{k+1} - x_0\| &= \|x_k + s(x_k) - x_0\| \leq \|x_k - x_0\| + \|s(x_k)\| \\ &\leq \|s(x_0)\| \frac{1 - (\frac{\epsilon}{a})^k}{1 - \frac{\epsilon}{a}} + \|s(x_0)\| \left(\frac{\epsilon}{a}\right)^k \\ &= \|s(x_0)\| \frac{1 - (\frac{\epsilon}{a})^{k+1}}{1 - \frac{\epsilon}{a}}. \end{aligned}$$

Then, by the hypothesis (c) and, $0 < \eta_2 < 1$, we have $0 \leq \frac{\epsilon}{a} \leq 1$, therefore, by (e)

$$\|x_{k+1} - x_0\| \leq r \left(1 - \frac{\epsilon}{a}\right) \frac{1 - \left(\frac{\epsilon}{a}\right)^{k+1}}{1 - \frac{\epsilon}{a}} = r \left(1 - \left(\frac{\epsilon}{a}\right)^{k+1}\right) \leq r. \quad (27)$$

In the same way, using (e) and (i),

$$\|x_k - x_0\| \leq r. \quad (28)$$

Moreover, from (ii) and the definition of x_{k+1} (4th step of Algorithm 1)

$$\|x_{k+1} - x_k\| = \|s(x_k)\| \leq \|s(x_0)\| \left(\frac{\epsilon}{a}\right)^k \leq \|s(x_0)\| \leq \delta,$$

where the last inequality is due to (e).

Then, by (27) and (28), $x_k, x_k + s(x_k) \in B(x_0, r)$ and $\|x_{k+1} - x_k\| \leq \delta$.

Let us prove that (iii) holds, proving that ρ_k^j is greater than η_2 for $j = 1, \dots, m$.

Using Lemma 5

$$\begin{aligned} f_j(x_k + s_k) - f_j(x_k) &\leq \nabla f_j(x_k)^T s_k + \frac{1}{2} s_k^T \nabla^2 f_j(x_k) s_k + \frac{\epsilon}{2} \|s_k\|^2 \\ &= q_k^j(s_k) + \frac{\epsilon}{2} \|s_k\|^2 \\ &= \eta_1 q_k^j(s_k) + (1 - \eta_1) q_k^j(s_k) + \frac{\epsilon}{2} \|s_k\|^2, \end{aligned} \quad (29)$$

where, naturally, $s_k = s(x_k)$.

Let us check that the sum of the two last terms is not positive, because the definition of $\theta(x_k)$ and from Lemmas 5 and 7,

$$\begin{aligned} (1 - \eta_2) q_k^j(s_k) + \frac{\epsilon}{2} \|s_k\|^2 &\leq (1 - \eta_2) \theta(x_k) + \frac{\epsilon}{2} \|s_k\|^2 \\ &\leq - (1 - \eta_2) \frac{a}{2} \|s_k\|^2 + \frac{\epsilon}{2} \|s_k\|^2 \\ &= a \frac{\|s_k\|^2}{2} \left(\frac{\epsilon}{a} - (1 - \eta_2)\right) \leq 0, \end{aligned}$$

where the last inequality is consequence of (c).

Then (29) becomes, $f_j(x_k + s_k) - f_j(x_k) \leq \eta_2 q_k^j(s_k)$. Since $q_k^j(s_k) < 0$ and $q_k^j(0) = 0$,

$$\frac{f_j(x_k) - f_j(x_k + s_k)}{q_k^j(0) - q_k^j(s_k)} \geq \eta_2,$$

thus

$$\Delta_{k+1} \geq \Delta_k.$$

The item (iv) is true from Lemma 8 and from the fact that, satisfied (iii), the step is accepted.

Now let us prove by the induction principle the items (i) and (ii) for all $k \geq 0$.

Clearly they hold for $k = 0$ by the assumptions. Let us suppose that they are valid up to k . Since (iii) holds, $x_{k+1} = x_k + s_k$, then

$$\begin{aligned} \|x_{k+1} - x_0\| &\leq \|x_k - x_0\| + \|s_k\| \leq \|s(x_0)\| \frac{1 - \left(\frac{\epsilon}{a}\right)^k}{1 - \frac{\epsilon}{a}} + \|s(x_0)\| \left(\frac{\epsilon}{a}\right)^k \\ &= \|s(x_0)\| \frac{1 - \left(\frac{\epsilon}{a}\right)^{k+1}}{1 - \frac{\epsilon}{a}}, \end{aligned}$$

therefore (i) holds for $k + 1$.

Using (iv) y (ii) for k we obtain

$$\|s_{k+1}\| \leq \|s_k\| \frac{\epsilon}{a} \leq \|s_0\| \left(\frac{\epsilon}{a}\right)^{k+1},$$

then (ii) is true for $k + 1$, and since (iii) and (iv) are derived from (i) and (ii), the four statements hold for k .

Now, from (i) and (ii),

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| = \sum_{k=0}^{\infty} \|s_k\| \leq \|s_0\| \sum_{k=0}^{\infty} \left(\frac{\epsilon}{a}\right)^k < \infty,$$

then the sequence $\{x_k\}$ is a Cauchy sequence, then there exists x^* such that $x_k \rightarrow x^*$.

Since $\|s_k\| \rightarrow 0$ then $\theta(x_k) \rightarrow 0$, by Lemma 5, $q_k^j(s_k) \leq \theta(x_k) \leq 0$, and by continuity of $\theta(x)$, $\theta(x^*) = 0$. Thus x^* is a local optimal Pareto point, because of Lemma 5 and the Theorem 3.1 of [9]. □

From the result and considering the Lipschitz continuity condition of the Hessian matrices of the objective functions, the local q -quadratic rate of convergence is obtained.

Theorem 4 *Under the assumptions of Theorem 3 and assuming that $\nabla^2 f_j$ is Lipschitz continuous with constant L in U , let*

$$\tau_k = \left(\frac{L}{2a}\right) \|s_k\|$$

and $\xi \in (0, \frac{1}{2})$ then there exists k_0 such that $\forall k \geq k_0$, $\tau_k < \xi$ and

$$\|x^* - x_{k+1}\| \leq \frac{L}{a} \frac{1 - \tau_k}{(1 - 2\tau_k)^2} \|x^* - x_k\|^2 \leq \frac{L}{a} \frac{1 - \xi}{(1 - 2\xi)^2} \|x^* - x_k\|^2,$$

where x^* is an accumulation point of $\{x_k\}$ and a local optimal Pareto point.

Proof From Theorem 3, we have $\|s_k\| \leq \|s_0\| \left(\frac{\epsilon}{a}\right)^k$, then

$$\tau_k \leq \frac{L}{2a} \|s_0\| \left(\frac{\epsilon}{a}\right)^k,$$

so, because $\frac{\epsilon}{a} < 1$, there exists k_0 such that $\forall k \geq k_0$

$$\tau_k < \xi, \tag{30}$$

From Lemma 8

$$\|s_{k+1}\| \leq \frac{L}{2a} \|s_k\|^2. \tag{31}$$

Let $i > k \geq k_0$ then

$$\|x_i - x_{k+1}\| \leq \sum_{j=k+2}^i \|x_j - x_{j-1}\| \leq \sum_{j=k+2}^i \|s_{j-1}\|,$$

and, if i goes to ∞ ,

$$\|x^* - x_{k+1}\| \leq \sum_{j=k+1}^{\infty} \|s_j\|.$$

Applying the inequality (31)

$$\|x^* - x_{k+1}\| \leq \frac{L}{2a} \|s_k\|^2 \sum_{j=0}^{\infty} \left(\frac{L}{2a} \|s_k\|\right)^{2j} \leq \frac{L}{2a} \|s_k\|^2 \sum_{j=0}^{\infty} (\tau_k)^{2j}.$$

Since (30), $\tau_k < \frac{1}{2}$, a convergent geometric series is obtained, thus

$$\|x^* - x_{k+1}\| \leq \frac{L}{2a} \|s_k\|^2 \sum_{j=0}^{\infty} \tau_k^j = \frac{L}{2a} \|s_k\|^2 \frac{1}{1 - \tau_k}. \tag{32}$$

Therefore

$$\begin{aligned}\|x^* - x_k\| &\geq \|x_{k+1} - x_k\| - \|x^* - x_{k+1}\| \geq \|s_k\| - \|s_k\|^2 \frac{\frac{L}{2a}}{1 - \tau_k} \\ &= \|s_k\| \frac{1 - 2\tau_k}{1 - \tau_k},\end{aligned}$$

where the last equality occurs because the definition of τ_k ; then, because $\tau_k < \frac{1}{2}$,

$$\|s_k\| \leq \|x^* - x_k\| \frac{1 - \tau_k}{1 - 2\tau_k}.$$

Connecting the last inequality and (32)

$$\begin{aligned}\|x^* - x_{k+1}\| &\leq \frac{L}{2a} \|s_k\|^2 \frac{1}{1 - \tau_k} \\ &= \frac{L}{2a} \|x^* - x_k\|^2 \frac{1 - \tau_k}{(1 - 2\tau_k)^2} \\ &\leq \frac{L}{2a} \|x^* - x_k\|^2 \frac{1 - \xi}{(1 - 2\xi)^2},\end{aligned}$$

because $\phi(t) = \frac{1-t}{(1-2t)^2}$ is monotonically increasing (its derivative is positive) in $(0, \frac{1}{2})$. This proves the local q -quadratic convergence of Algorithm 1. \square

6 Numerical results

In order to exhibit the behavior of the algorithm we consider a set of ten problem from the literature [6,7,9], one of this with variable size [9] and a new nonconvex problem proposed in [5]. Each problem has been considered with different starting points generated at random. The points have been clustered in three regions: the small region is the set $[-1, 1] \times \cdots \times [-1, 1] \subset \mathbb{R}^n$, the medium region is $[-10, 10] \times \cdots \times [-10, 10] \subset \mathbb{R}^n$, and the big region is $[-100, 100] \times \cdots \times [-100, 100] \subset \mathbb{R}^n$; and the problems that originally were box constrained were tested with random points in that region too. The algorithm has been coded in FORTRAN 95 in double precision using GFortran compiler. The code has been executed on a personal computer with a processor INTEL Core I3 with 3 Gb of memory RAM using Ubuntu 10.04 Linux operation system.

We establish the maximum number of iterations as $maxiter = 500$, and the tolerance for the stopping criterion as $\varepsilon = 10^{-8}$. The parameters in the algorithm are chosen as follows: $\eta_1 = 10^{-1}$, $\eta_2 = 0.9$. The trust region subproblem has been solved by means the routine ALGENCAN [1].

To show the behavior of Algorithm 1 we compare it with the procedure proposed by Qu, Goh and Lian [16] and the proposed by Villacorta, Oliveira and Souberyran [19], using the problem proposed in [9]. For the algorithm proposed in

Table 1 Average of iterations needed for convergence in each region

Prob.	It. small reg.	It. medium reg.	It. big reg.	It. original reg.	Source
bk1	2.0	3.6	6.4	2.7	[6]
dg01	2.4	2.2	18.8	2.1	[6]
ex005	18.8	21.1	28.8	16.9	[6]
le1	7.9	17.1	20.2	12.7	[6]
mlf1	1.4	1.8	1.2	1.4	[6]
sk2	5.1	6.9	5.7	6.1	[6]
SCH	1.4	3.3	5.3	8.7	[7]
FON	5.7	1.0	1.0	3.2	[7]
POL	6.9	7.1	7.9	6.3	[7]
GAC	2.8	3.5	5.9	–	[5]

Table 2 FGS problem [9]

Size	Algorithm [19]		Algorithm [16]		Algorithm 1	
	It.	Succes.	It.	Succes.	It.	Succes.
5	39.8	10	4.52	10	3.73	10
10	95.67	6	5.35	10	3.71	10
50	11.83	6	7.75	10	4.16	10
100	4.20	5	9.17	10	4.03	10
200	–	0	10.26	10	5.57	10

[19] many features, such as the parameters to update the trust region radius and to evaluate the steps for acceptances as well as the election of the matrix for the second order term, are not specified and should be chosen by the user. We consider $B_k = (\nabla^2 f_1(x_k) + \nabla^2 f_2(x_k) + \nabla^2 f_3(x_k)) / 3$. The parameters to evaluate the steps and to update the trust region radius were established the same as our algorithm in order to run both methods under the same conditions. The algorithm of [16] was coded with the parameters suggested by the authors.

To visualize the whole behavior of the algorithm we present tables for each cluster points. In Table 1 we show the average of iterations needed for each problem in each region and the source of the problem. In Table 2, we present the number of iterations for the problem proposed in [9] for Algorithm 1, [16] and [19]. In this table we can observe a better performance of Algorithm 1 than the algorithm of [16], in a similar way that Newton's method is better than secant methods. The poor performance of the algorithm of [19] is attributed to the fact that the subproblem can provide long steps even when the current iterate is near to a Pareto critical point.

Moreover, in Fig. 1, we can observe the performance profile with the problems clustered by the region where the starting point was located. Considering this graphic, it is possible to observe that Algorithm 1 had a good performance independently of the starting point. The good performance of the some of test problems for initial points in the small region is due to have their Pareto frontier near zero.

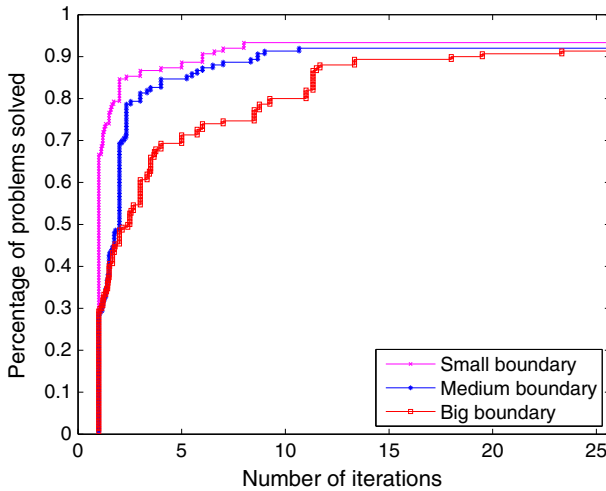


Fig. 1 Performance profile of Algorithm 1 according with the location of the starting points

7 Conclusions

A trust region algorithm for the nonconvex unconstrained multiobjective optimization problem has been presented. From a theoretical point of view the proposed algorithm exhibits a similar behavior to the scalar case. We remark that the notion of fraction of Cauchy decrease has been adapted to the multiobjective problem. The algorithm provides a sequence which has at least a subsequence converging to a weak Pareto point. Moreover, under convexity assumptions the limit point of this subsequence is a Pareto point. In this case, the locally q -quadratic rate of convergence of the algorithm is proved.

An interesting local convergence result was obtained, where the information is related to the starting point, not to the solution. In that sense it is possible to consider Theorem 3 as similar to Kantorovich's theorem [8].

From a practical point of view, the algorithm shows a good behavior in a set of problems. With respect to the problem FGS we observe that our algorithm behave better than the proposed by Qu, Goh and Lian [16] and, according with ours test, we can not establish a good comparison with the algorithm [19].

In the future, following this line of research we consider to implement an algorithm where the Hessian matrices be updated via a secant approximation. In order to obtain q -superlinear rate of convergence we are trying to introduce a Dennis-More type condition. On the other side we believe that the algorithm can be easily extended if nonlinear constraints are added to the problem.

Acknowledgments The authors are grateful to the anonymous reviewers, whose comments improved this work.

References

1. Andreani, R., Birgin, E., Martínez, J., Schuverdt, M.: On augmented Lagrangian methods with general lower-level constraints. *SIAM J. Optim.* **18**(4), 1286–1309 (2008)
2. Ashry, G.A.: On globally convergent multi-objective optimization. *Appl. Math. Comput.* **183**, 209–216 (2006)
3. Bertsekas, D.P.: *Nonlinear Programming*, 2nd edn. Athenas Scientific, Belmont, MA (1999)
4. Conn, A., Gould, N.I.M., Toint, P.: *Trust-Region Methods*. SIAM-MPS, Philadelphia, Pennsylvania (2000)
5. Carrizo, G.A.: *Estrategia de región de confianza para problemas multiobjetivo no convexos*. Universidad Nacional del Sur, Argentina (2013)
6. Custodio, A.L., Madeira, J.F.A., Vaz, A.I.F., Vicente, L.N.: Direct multisearch for multiobjective optimization. *SIAM J. Optim.* **21**, 1109–1140 (2011)
7. Deb, K., Agrawal, S., Pratap, A., Meyarivan, T.: A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *IEEE Trans. Evol. Comput.* **6**(2), 182–197 (2002)
8. Dennis, J.E., Schnabel, R.B.: *Numerical Methods for Unconstrained Optimization and Nonlinear Systems*. Prentice-Hall, Englewood Cliffs, New Jersey (1983)
9. Fliege, J., Drummond, L.M.Graña, Svaiter, B.F.: Newton's method for multiobjective optimization. *SIAM J. Optim.* **20**, 602–626 (2009)
10. Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. *Math. Methods Oper. Res.* **51**(3), 479–494 (2000)
11. Fischer, A., Shukia, P.K.: A Levenberg–Marquardt algorithm for unconstrained multicriteria optimization. *Oper. Res. Lett.* **175**, 395–414 (2005)
12. Geoffrion, A.M.: Proper efficiency and the theory of vector maximization. *J. Optim. Theory Appl.* **22**, 618–630 (1968)
13. Laumanns, M., Thiele, L., Deb, K., Zitzler, E.: Combining convergence and diversity in evolutionary multiobjective optimization. *Evol. Comput.* **10**, 263–282 (2002)
14. Lobato, F.S., Steffen, J.V.: Multi-objective optimization firefly algorithm applied to (bio)chemical engineering system design. *Am. J. Appl. Math. Stat.* **1**, 110–116 (2013)
15. Luc, D.T.: *Theory of Vector Optimization*. Springer, Berlin (1989)
16. Qu, S., Goh, M., Lian, B.: Trust region methods for solving multiobjective optimisation. *Optim. Methods Softw.* **28**(4), 796–811 (2013)
17. Ravanbakhsh, A., Franchini, S.: Multiobjective optimization applied to structural sizing of low cost university-class microsatellite projects. *Acta Astronaut.* **79**, 212–220 (2012)
18. Steuer, R.E., Na, P.: Multiple criteria decision making combined with finance: a categorized bibliographic study. *Eur. J. Oper. Res.* **150**(3), 496–515 (2003)
19. Villacorta, K.D., Oliveira, P.R., Souberyan, A.: A trust region method for unconstrained multiphjective problems with applications in satisficing processes. *J. Optim. Theory Appl.* **160**, 865–889 (2014)
20. Zitzler, E., Deb, K., Thiele, L.: Comparison of multiobjective evolutionary algorithms: empirical results. *Evol. Comput.* **8**, 173–195 (2000)