A quadratically convergent Newton method for vector optimization

L. M. Graña Drummond[*] F. M.P. Raupp[†] B. F. Svaiter[‡]

Abstract

We propose a Newton method for solving smooth unconstrained vector optimization problems under partial orders induced by general closed convex pointed cones. The method extends the one proposed by Fliege, Graña Drummond and Svaiter for multicriteria, which in turn is an extension of the classical Newton method for scalar optimization. The steplength is chosen by means of an Armijo-like rule, guaranteeing an objective value decrease at each iteration. Under standard assumptions, we establish superlinear convergence to an efficient point. Additionally, as in the scalar case, assuming Lipschitz continuity of the second derivative of the objective vector-valued function, we prove q-quadratic convergence.

Keywords: vector optimization; efficient points; convexity with respect to cones; Newton method

[*]Faculdade de Administração e Ciências Contábeis, Universidade Federal do Rio de Janeiro, FACC-UFRJ, email: lmgd@facc.ufrj.br. This author was partially supported by FAPERJ/CNPq through PRONEX-Optimization and by CNPq through Projeto Universal 473818/2007-8.

[†]Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro - DEI-PUC-Rio, Brazil, e-mail: fraupp@puc-rio.br. This author was partially supported by FAPERJ/CNPq through PRONEX-Optimization and by CNPq through Projeto Universal 473818/2007-8.

[‡]Instituto de Matemática Pura e Aplicada - IMPA, Brazil, e-mail: benar@impa.br. This author was partially supported by CNPq Grant 301200/93-9(RN) and by FAPERJ/CNPq through PRONEX-Optimization.

1 Introduction

The canonical order in the real line can be characterized by means of inclusion on $\mathbb{R}_+$, that is $a \leq b$ if $b - a \in \mathbb{R}_+$. Extensions of this order to $\mathbb{R}^m$ can be derived using inclusions on a closed (pointed) convex cone K as follows: $u \preceq v$ if $v - u \in K$. Classical minimization of scalar-valued functions can be also extended to vector-valued functions, replacing the requirement of finding a minimum by finding minimal elements in these (partial) orders.

If the cone under consideration is $\mathbb{R}^m_+$, points whose image are minimal are called *Pareto optimal (or efficient)* points, and minimization with respect to this order is called *multicriteria optimization.* For other cones, we have the so-called vector optimization problems and the optimal elements are know as *efficient points.* Many practical problems are modeled as minimization of vector-valued functions with respect to orders induced by $\mathbb{R}^m_+$ and also other (than $\mathbb{R}^m_+$) cones [1, 2].

Multicriteria optimization problems are traditionally solved by two different approaches: scalarization and ordering techniques ([14], [12]). The scalarization approach enables the computation of efficient (Pareto) or weakly efficient solutions by formulating single objective optimization problems (with or without constraints) and choosing "good" parameters in advance. Other scalarization techniques that are free of establishing parameters can compute a discrete approximation of the optimal set. The solution strategies that use ordering techniques require a prior ranking specification for the objectives.

A new type of strategy for solving multicriteria optimization problems, that does not require parameter nor ordering information, is constituted by a class of "descent" methods. The very first of these descent methods, proposed by Fliege and Svaiter in [6], is the multiobjective steepest descent method. The second one is the Newton method for multicriteria, recently proposed by Fliege, Graña Drummond and Svaiter in [7].

For vector optimization, i.e., when the partial order is induced by a general closed convex pointed cone with nonempty interior, besides the classical procedures (see [11]), we also have some "descent" methods. For unconstrained problems, Graña Drummond and Svaiter [10] extended the steepest descent method. Later on, Graña Drummond and Iusem [9] and Fukuda and Graña Drummond [8] studied a projected gradient method for vector optimization problems with decision variables restricted to closed convex sets. An extension of the proximal point method for vector optimization was studied by Bonnel, Iusem and Svaiter in [3].

In this work, the Newton method together with a globalization strategy for multicriteria optimization is extended to vector optimization. For the convergence analysis, strong convexity of the vector-valued objective function respect to the ordering cone is required. The proposed method performs a full Newton step if the decrement of the objective function is at least a fraction of the predicted decrement, otherwise, it performs a damped Newton step using an Armijo-like rule for choosing the steplength. We prove global superlinear convergence to an efficient point whenever the starting point is in a compact level set. If, additionally, the second derivative of the objective function is Lispchitz continuous, then the generated sequence converges q-quadratically, as in the classical scalar-valued case.

It is worth to notice that, in the Paretian case, proper implementation of the method, with randomly chosen initial points, allow us to retrieve a good approximation of the efficient set (see Section 7 in [7]).

2 Notation and auxiliary results

Throughout this work, $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\mathbb{R}^n$ and $\| \cdot \|$ stands for the Euclidean norm, i.e., $\|x\| = \sqrt{\langle x, x \rangle}$ for $x \in \mathbb{R}^n$.

Let $K \subset \mathbb{R}^m$ be a closed convex pointed (i.e., $K \cap (-K) = \{0\}$) cone, with nonempty interior. The partial order induced in $\mathbb{R}^m$ by K is defined as

$$u \preceq v \quad \text{(alternatively, } v \succeq u) \quad \text{if and only if} \quad v - u \in K.$$

We will also consider the following stronger relation

$$u \prec v \quad \text{(alternatively, } v \succ u) \quad \text{if and only if} \quad v - u \in \text{int}(K).$$

Consequently, $u \npreceq v$ ($u \nprec v$) means that $v - u \notin K$ ($v - u \notin \text{int}(K)$).

The *(positive) polar* cone of $K \subseteq \mathbb{R}^m$ is the set $K^* := \{w \in \mathbb{R}^m \mid \langle v, w \rangle \geq 0 \ \ \forall v \in K\}$. Since $K \subset \mathbb{R}^m$ is a closed convex cone, it follows that $K = K^{**}$ (see Theorem 14.1 in [13]). For our purposes, a *subbase* of K^* is a compact set $C \subset K^* \setminus \{0\}$ such that the cone generated by the convex hull of C coincides with K^*, i.e., $0 \notin C$ and $\text{cone}(\text{conv}(C)) = K^*$. As $K = K^{**}$, we can use C to effectively describe the cone and its interior:

$$K = \{v \in \mathbb{R}^m \mid \langle v, w \rangle \geq 0 \ \ \forall\, w \in C\}, \tag{1}$$

$$\text{int}(K) = \{v \in \mathbb{R}^m \mid \langle v, w \rangle > 0 \ \ \forall\, w \in C\}. \tag{2}$$

Here, we are interested in finding a minimizer (with respect to the partial order induced by K) of the problem

$$\begin{array}{ll} \min & F(x) \\ \text{s.t.} & x \in U, \end{array} \tag{3}$$

where $F : U \longrightarrow \mathbb{R}^m$ and $U \subset \mathbb{R}^n$ is an open convex set. This means that we seek for an *efficient* point for F in U, i.e., a point $x^* \in U$ such that

$$\nexists\, x \in U, \quad F(x) \preceq F(x^*) \;\text{ and }\; F(x) \neq F(x^*).$$

Recall that a point $x^* \in U$ is *weakly efficient* if

$$\nexists\, x \in U, \quad F(x) \prec F(x^*).$$

Clearly, every efficient point is weakly efficient.

The mapping F is said *convex respect to* K (or simply K*-convex*) on U if

$$F(\alpha x + (1-\alpha)y) \preceq \alpha F(x) + (1-\alpha)F(y),$$

for all $x, y \in U$ and all $\alpha \in (0,1)$.

If the above inequality holds strictly, F is called *strictly convex respect to* K (or simply *strictly* K*-convex*) on U. Finally, we will say that the function F is *strongly* K*-convex* on U if there exists $\hat{u} \in \operatorname{int}(K)$ such that

$$F(\alpha x + (1-\alpha)y) \preceq \alpha F(x) + (1-\alpha)F(y) - \frac{1}{2}\alpha(1-\alpha)\|x-y\|^2\hat{u} \qquad (4)$$

for all $x, y \in U$ and all $\alpha \in (0,1)$. Note that $\hat{u}$ plays the roll of the "modulus of strong K-Convexity" of the mapping F. Clearly, as in the scalar case, strong K-convexity implies strict K-convexity, which, in turn, implies K-convexity.

Define for each $w \in \mathbb{R}^m$, $\Psi_w : U \to \mathbb{R}$,

$$\Psi_w(x) := \langle w, F(x)\rangle. \qquad (5)$$

From now on, let C be a fixed but arbitrary subbase of K^*. Clearly, (strict) K-convexity of F is equivalent to (strict) convexity of Ψ_w for all $w \in C$. We will see that strong K-convexity of F on U is equivalent to (uniform) strong convexity of Ψ_w on U for all $w \in C$.

Lemma 2.1. *The mapping $F : U \to \mathbb{R}^m$ is strongly K-convex if and only if there exists $\rho > 0$ such that $\Psi_w : U \to \mathbb{R}$, given by* (5)*, is strongly convex on U with modulus ρ for all $w \in C$.*

Proof. First assume that F is strongly K-convex. Take $w \in C$, $x, y \in U$ and $\alpha \in (0,1)$. From (4) we obtain

$$F(\alpha x + (1-\alpha)y) - \alpha F(x) - (1-\alpha)F(y) + \frac{1}{2}\alpha(1-\alpha)\|x-y\|^2\hat{u} \in -K,$$

which by (1) and the properties of the inner product, gives us

$$\langle w, F(\alpha x+(1-\alpha)y)\rangle \leq \alpha\langle w, F(x)\rangle+(1-\alpha)\langle w, F(y)\rangle-\frac{\rho_w}{2}\alpha(1-\alpha)\|x-y\|^2, \tag{6}$$

where $\rho_w := \langle w, \hat{u}\rangle > 0$. From the compactness of C and the Bolzano-Weierstrass Theorem, there exists $\bar{w} \in C$ such that $\rho := \langle \bar{w}, \hat{u}\rangle \leq \langle w, \hat{u}\rangle = \rho_w$ for all $w \in C$. And so, we can write inequality (6) with $\rho_w = \rho$, i.e., all Ψ_w are strongly convex with the same modulus ρ.

To prove the converse, assume that (6) holds with $\rho_w = \rho$. Trivially, in view of the compactness of C, there exists $\hat{u} \in \mathrm{int}(K)$ such that

$$\langle w, \hat{u}\rangle < \rho \text{ for all } w \in C. \tag{7}$$

Combining this inequality with (6), we get

$$\langle w, F(\alpha x + (1-\alpha)y) - \alpha F(x) - (1-\alpha)F(y) + \frac{1}{2}\alpha(1-\alpha)\|x-y\|^2\hat{u}\rangle \leq 0,$$

for all $w \in C$, which, by virtue of (1), means that F is strongly K-convex. □

From the last proposition, we can see that for F twice continuously differentiable, strong K-convexity is equivalent to uniform boundedness (away from zero) of the eigenvalues of all Hessians $\nabla^2\Psi_w$.

Corollary 2.2. *Assume that F is twice continuously differentiable. Then, F is strongly K-convex if and only if there exists $\rho > 0$ such that*

$$\lambda_{\min}(\nabla^2\Psi_w(x)) \geq \rho \text{ for all } x \in U \text{ and all } w \in C,$$

where Ψ_w is given by (5) and $\lambda_{\min} : \mathbb{R}^{n\times n} \to \mathbb{R}$ denotes the smallest-eigenvalue function.

Proof. Each Ψ_w is $\mathcal{C}^2$. Therefore, ρ-strong convexity of Ψ_w is equivalent to $\nabla^2\Psi_w(x) - \rho I$ being positive-semidefinite for all $x \in U$, which, in turn, is equivalent to $\lambda_{\min}(\nabla^2\Psi_w(x)) \geq \rho$ for all x in U. To complete the proof, combine this observation with Lemma 2.1. □

Now, for F just continuously differentiable, we extend the classical stationarity condition "gradient-equal-zero" to the vector-valued case. As in [10], we say that a point $\bar{x}$ is *stationary* or *critical* for F (with respect to K) if

$$R(JF(\bar{x})) \cap (-\mathrm{int}(K)) = \emptyset, \tag{8}$$

where $JF(\bar{x})$ denotes the Jacobian of F at $\bar{x}$ and $R(JF(\bar{x}))$ is the image of the linear operator $JF(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^m$. Note that $\bar{x}$ is stationary if, and only if, for all $s \in \mathbb{R}^n$ we have $JF(\bar{x})s \notin -\mathrm{int}(K)$, which, from (2), is equivalent to

$$\langle \bar{w}, JF(\bar{x})s \rangle \geq 0 \text{ for some } \bar{w} = \bar{w}(s) \in C. \tag{9}$$

On the other hand, a point $\bar{x}$ is not critical if there exists $s \in \mathbb{R}^n$, such that $JF(\bar{x})s \in -\mathrm{int}(K)$. Or, from (2), for $\bar{x}$ nonstationary, there exists s satisfying

$$\langle w, JF(\bar{x})s \rangle < 0 \ \ \forall\, w \in C. \tag{10}$$

Recall that s is a *descent direction* if there exists $\bar{t} > 0$ such that

$$F(\bar{x} + ts) \prec F(\bar{x}) \ \forall\, t \in (0, \bar{t}]. \tag{11}$$

Now observe that if $JF(\bar{x})s \prec 0$, then s is a descent direction. Indeed,

$$F(\bar{x} + ts) = F(\bar{x}) + tJF(\bar{x})s + \varepsilon(t) = F(\bar{x}) + t\left(JF(\bar{x})s + \frac{\varepsilon(t)}{t}\right),$$

where $\varepsilon(t)/t \to 0$ whenever $t \to 0$. So, since $JF(\bar{x})s \prec 0$, there exists $\bar{t} > 0$ sufficiently small such that $JF(\bar{x})s + \varepsilon(t)/t \prec 0$ for all $t \in (0, \bar{t}]$, and so (11) holds.

In our next proposition, we will establish some relations between stationarity and optimality.

Proposition 2.3. *Let $C \subseteq \mathbb{R}^m$ be a subbase of the cone K^* and assume that $F : U \to \mathbb{R}^n$ is continuously differentiable.*

1. *If $\bar{x} \in U$ is weakly efficient, then $\bar{x}$ is a critical point for F.*
2. *If F is K-convex and $\bar{x} \in U$ is critical for F, then $\bar{x}$ is weakly efficient.*
3. *If F is strictly K-convex and $\bar{x} \in U$ is critical for F, then $\bar{x}$ is efficient.*

Proof. Assume that $\bar{x}$ is weakly efficient. If $\bar{x}$ is not critical, then (8) does not hold, so, as we saw above, there exists $s \in \mathbb{R}^n$, a descent direction for F at $\bar{x}$, i.e., an s such that (11) holds, in contradiction with the weak efficiency of $\bar{x}$.

To prove item 2, take any $x \in U \setminus \{\bar{x}\}$. Since $\bar{x}$ is critical, there exists $\bar{w} = \bar{w}(x)$ such that (9) holds for $s = x - \bar{x}$. Using the convexity of $\Psi_{\bar{w}} = \langle \bar{w}, F(\cdot) \rangle$, the fact that $\nabla_x \langle \bar{w}, F(\bar{x}) \rangle^T (x - \bar{x}) = \langle \bar{w}, JF(x)(x - \bar{x}) \rangle$ and (9), we obtain

$$\langle \bar{w}, F(x) \rangle \geq \langle \bar{w}, F(\bar{x}) \rangle + \langle \bar{w}, JF(\bar{x})(x - \bar{x}) \rangle \geq \langle \bar{w}, F(\bar{x}) \rangle. \tag{12}$$

Therefore, there does not exist $x \in U$ such that $\langle \bar{w}, F(x)\rangle < \langle \bar{w}, F(\bar{x})\rangle$, which, by virtue of (2), means that $\bar{x}$ is weakly efficient for F.

Finally, to prove item 3, observe that, since $\bar{x}$ is critical and F is strictly K-convex, as in item 2, we see that there exists some $\bar{w} \in C$ for which the first inequality in (12) is strict. So, by (1), $\bar{x}$ is efficient. □

3 Newton direction for vector optimization

From now on, we will assume that F is strongly K-convex and twice continuously differentiable on the open convex set U. Therefore $\Psi_w(\cdot) = \langle w, F(\cdot)\rangle$ is a strongly convex $\mathcal{C}^2$ scalar-valued function for all $w \in C$.

Define, for each $w \in \mathbb{R}^m$, $\psi_w : U \times \mathbb{R}^n \to \mathbb{R}$,

$$\begin{aligned}\psi_w(x,s) :=& \nabla_x \Psi_w(x)^T s + \frac{1}{2} s^T \nabla_x^2 \Psi_w(x) s \\ =& \nabla_x \langle w, F(x)\rangle^T s + \frac{1}{2} s^T \nabla_x^2 \langle w, F(x)\rangle s \qquad (13)\end{aligned}$$

and let $s(x)$, the *Newton direction* at x, be

$$\begin{aligned}s(x) :=& \arg \min_{s\in\mathbb{R}^n} \max_{w\in C} \quad \psi_w(x,s) \\ =& \arg \min_{s\in\mathbb{R}^n} \max_{w\in C} \{\nabla_x \langle w, F(x)\rangle^T s + \frac{1}{2} s^T \nabla_x^2 \langle w, F(x)\rangle s\} \qquad (14)\end{aligned}$$

By Corollary 2.2, $\psi_w(x,s)$ is ρ-strongly convex in s, for any $x \in U$ and all $w \in C$. Therefore $\max_{w\in C} \psi_w(x,s)$ is also ρ-strongly convex in s and so, $s(x)$ is well-defined.

As in [10], the subbase C allows us to define a *gauge function* for K, $\varphi_C : \mathbb{R}^m \to \mathbb{R}$,

$$\varphi_C(u) = \sup_{w\in C} \langle w, u\rangle.$$

Let q_x be the second order approximation of *the variation* of F at $x \in U$, that is,

$$q_x(s) = DF(x)s + \frac{1}{2} D^2 F(x)[s,s].$$

The Newton direction the at x can also be defined as

$$s(x) := \arg \min_{s\in\mathbb{R}^n} \varphi_C(q_x(s)).$$

If $K = \mathbb{R}^m_+$ and $C = \{e^1, \ldots, e^m\}$, where e^i is the i-th canonical vector of $\mathbb{R}^m$, then $s(x)$ given by (14) becomes the Newton direction for multicriteria defined in [7]. Moreover, if $m = 1$, $K = \mathbb{R}_+$ and $C = \{1\}$, $s(x)$ is the classical Newton direction for scalar-valued optimization problems.

Denote by $\theta(x)$ the optimal value of the minimization problem in (14), i.e.,

$$\begin{aligned}\theta(x) &:= \min_{s\in\mathbb{R}^n} \max_{w\in C} \{\langle w, JF(x)s\rangle + \frac{1}{2}s^T\nabla_x^2\langle w, F(x)\rangle s\} \\ &= \max_{w\in C} \{\langle w, JF(x)s(x)\rangle + \frac{1}{2}s(x)^T\nabla_x^2\langle w, F(x)\rangle s(x)\}. \qquad (15)\end{aligned}$$

Let us now present some basic results relating stationarity of a given point x to its Newton step $s(x)$ and $\theta(x)$.

Proposition 3.1. *Let $\theta : U \longrightarrow \mathbb{R}$ and $s : U \longrightarrow \mathbb{R}^n$ be given by (15) and (14), respectively. Under our general assumptions, we have:*

1. *$\theta(x) \leq 0$ for all $x \in U$.*
2. *The following conditions are equivalent.*
   - *(a) The point x is nonstationary.*
   - *(b) $\theta(x) < 0$.*
   - *(c) $s(x) \neq 0$.*
   - *(d) $s(x)$ is a descent direction.*

Proof. By the first equality in (15) and (13), $\theta(x) \leq \max_{w\in C} \psi_w(x, 0) = 0$, so item 1 holds.

In order to show the equivalences in item 2, first assume that (a) holds, that is, there exists $\tilde{s} \in \mathbb{R}^n$ such that $JF(x)\tilde{s} \in -\text{int}(K)$, which means that $\langle w, JF(x)\tilde{s}\rangle < 0$ for all $w \in C$ (see (10)). From (15), for $s = t\tilde{s}$ with $t > 0$, we get

$$\theta(x) \leq t \max_{w\in C} \{\langle w, JF(x)\tilde{s}\rangle + \frac{t}{2}\tilde{s}^T\nabla_x^2\langle w, F(x)\rangle\tilde{s}\}.$$

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

To prove that (b) implies (c), note that if $\theta(x) < 0$ then, in view of the second equality in (15), $s(x) \neq 0$.

Suppose that (c) holds. Since $f_x(s) = \max_{w\in C} \psi_w(x, s)$ is strongly convex and $f_x(0) = 0$, if $s(x)$, the minimizer of f_x, is different form 0, then $\theta(x)$, the minimal value of f_x must be strictly smaller than 0. Therefore, for any $w \in C$

$$\langle w, JF(x)s(x)\rangle < \langle w, JF(x)s(x)\rangle + \frac{1}{2}s(x)^T\nabla_x^2\langle w, F(x)\rangle s(x) \leq \theta(x) < 0$$

where the first inequality follows form the fact that $\nabla_x^2\langle w, F(x)\rangle$ is positive definite for any $w \in C$. Using the above inequality and the fact that (10) implies (11), we conclude that item (d) holds. Item (d) trivially implies item (a). □

The next lemma supplies an implicit definition of $s(x)$, which will be used for deriving bounds for $\|s(x)\|$. These bounds will allow us to prove continuity of $s(\cdot)$, $\theta(\cdot)$ and will be used to establish the rate of convergence of the Newton method.

Given $x \in U$ and $s \in \mathbb{R}^n$, we define the set of *active* w's for $\psi_{(\cdot)}(x, s)$ as

$$W(x, s) := \{\bar{w} \in C \mid \psi_{\bar{w}}(x, s) = \max_{w \in C} \psi_w(x, s)\}, \tag{16}$$

where ψ_w is given by (13).

Lemma 3.2. *For any $x \in U$, there exists an integer $r = r(x)$ and*

$$w^i = w^i(x) \in W(x, s(x)), \quad \alpha_i = \alpha_i(x) \in [0, 1] \qquad (1 \leq i \leq r)$$

such that

$$s(x) = -\left[\sum_{i=1}^{r} \alpha_i \nabla_x^2 \langle w^i, F(x)\rangle\right]^{-1} \sum_{i=1}^{r} \alpha_i \langle w^i, JF(x)\rangle, \tag{17}$$

with $\sum_{i=1}^{r} \alpha_i = 1$

Proof. To simplify the notation, for a fixed $x \in U$ and ψ_w given by (13), let us define

$$h : C \times \mathbb{R}^n \to \mathbb{R}, \quad h(w, s) = \psi_w(x, s),$$

and

$$f : \mathbb{R}^n \to \mathbb{R}, \quad f(s) = \max_{w \in C} h(w, s).$$

Since C is compact, h is continuous, $h(w, \cdot)$ is a continuously differentiable convex function for all $w \in C$ and $\nabla_s h(\cdot, s)$ is continuous on C for all $s \in \mathbb{R}^n$, using Danskin's Theorem (see Proposition 4.5.1, pp. 245-247, [4]), we conclude that the subdifferential of $f(\cdot)$ at s is the convex hull of all s-gradients $\nabla_s h(w, s) = \nabla_s \psi_w(x, s)$ on active w's. Since $s(x)$ is the minimizer of f, $0 \in \partial f(s(x))$. Hence there exists $r = r(x)$, $w^i = w^i(x) \in W(x, s(x))$ and $\alpha_i = \alpha_i(x) \in [0,1]$ $(1 \le i \le r)$ such that

$$0 = \sum_{i=1}^{r} \alpha_i \left(\nabla_x \langle w^i, F(x) \rangle + \nabla_x^2 \langle w^i, F(x) \rangle s(x) \right), \tag{18}$$

where $\sum_{i=1}^{r} \alpha_i = 1$. To end the proof, use the above equation and note that, since F is strongly K-convex, $\nabla_x^2 \langle w^i, F(x) \rangle$ is positive definite for $i = 1, \dots, r$. □

Note that, according to the above lemma, the Newton direction (14) is a classical Newton direction for a scalarized problem with weighting parameter $\sum \alpha_i w^i$ *implicitly defined* at the point x. This characterization allows us to obtain an upper bound for the norm of the Newton direction.

Corollary 3.3. *For any $x \in U$, we have*

$$\|s(x)\| \le \frac{\kappa}{\rho} \|JF(x)\| \quad \text{and} \quad \frac{\rho}{2} \|s(x)\|^2 \le -\theta(x),$$

where $\kappa = \sup_{w \in C} \|w\|$.

Proof. For simplifying the proof, define

$$\tilde{w} = \sum_{i=1}^{r} \alpha_i w^i,$$

where r, α_i, w^i $(i = 1, \dots, r)$ are given by Lemma 3.2. Using Corollary 2.2,

$$\sum_{i=1}^{r} \alpha_i \nabla_x^2 \langle w^i, F(x) \rangle = \nabla_x^2 \langle \tilde{w}, F(x) \rangle \ge \rho I.$$

From (17) and the above relation, it follows that

$$\|s(x)\| \le \frac{\kappa}{\rho} \|JF(x)\|.$$

Let us now prove the second inequality. Using again Lemma 3.2, we have

$$s(x) = - \left[\nabla_x^2 \langle \tilde{w}, F(x) \rangle \right]^{-1} \nabla_x \langle \tilde{w}, F(x) \rangle$$

and

$$
\begin{aligned}
\theta(x) =& \sum_{i=1}^{r} \alpha_i \psi_{w^i}(x, s(x)) \\
=& \sum_{i=1}^{r} \alpha_i \left[\nabla_x \langle w^i, F(x) \rangle^T s(x) + \frac{1}{2} s(x)^T \nabla_x^2 \langle w^i, F(x) \rangle s(x) \right] \\
=& \nabla_x \langle \tilde{w}, F(x) \rangle^T s(x) + \frac{1}{2} s(x)^T \nabla_x^2 \langle \tilde{w}, F(x) \rangle s(x).
\end{aligned}
$$

Combining all the previous equations, we conclude that

$$
\theta(x) = -\frac{1}{2} s(x)^T \nabla_x^2 \langle \tilde{w}, F(x) \rangle s(x) \leq -\frac{\rho}{2} \|s(x)\|^2,
$$

which completes the proof. □

4 Additional properties of Newton direction

In this section we establish the continuity of $\theta(\cdot)$ and $s(\cdot)$ as functions defined on U. We also analyze the welldefinedness of an Armijo-like rule for choosing the steplengths along the Newton direction.

Proposition 4.1. *The functions $\theta : U \longrightarrow \mathbb{R}$ and $s : U \longrightarrow \mathbb{R}^n$, given by (15) and (14) respectively, are continuous.*

Proof. Take $x, x' \in U$ and let $s = s(x)$, $s' = s(x')$. For any $w \in C$,

$$
\begin{aligned}
\theta(x') \leq& \langle w, JF(x')s \rangle + \frac{1}{2} s^T \nabla_x^2 \langle w, F(x') \rangle s \\
=& \langle w, JF(x)s \rangle + \frac{1}{2} s^T \nabla_x^2 \langle w, F(x) \rangle s + \\
& + \langle w, (JF(x') - JF(x))s \rangle + \frac{1}{2} s^T \left(\nabla_x^2 \langle w, F(x') \rangle - \nabla_x^2 \langle w, F(x) \rangle \right) s \\
\leq& \theta(x) + \kappa \|s\| \|JF(x') - JF(x)\| + \frac{\kappa}{2} \|s\|^2 \|D^2F(x) - D^2F(x')\|,
\end{aligned}
$$

where D^2F stands for the second derivative of F and κ is defined in Corollary 3.3. Therefore, using the above inequality and Corollary 3.3 we conclude that

$$
\begin{aligned}
\theta(x') - \theta(x) \leq& \frac{\kappa^2}{\rho} \|JF(x)\| \|JF(x') - JF(x)\| + \\
& \frac{\kappa^3}{2\rho^2} \|JF(x)\|^2 \|D^2F(x) - D^2F(x')\|.
\end{aligned}
$$

Since the above inequality also holds interchanging x and x', and F is $\mathcal{C}^2$, the function $\theta(\cdot)$ is continuous.

Finally, let us prove the continuity of $s(\cdot)$. By the same reasoning as above,

$$\begin{aligned}
\theta(x) =& \max_{w\in C} \langle w, JF(x)s\rangle + \frac{1}{2}s^T\nabla_x^2\langle w, F(x)\rangle s \\
=& \max_{w\in C} \Big[\langle w, JF(x')s\rangle + \frac{1}{2}s^T\nabla_x^2\langle w, F(x')\rangle s + \\
&+\langle w, (JF(x) - JF(x'))s\rangle + \frac{1}{2}s^T\Big(\nabla_x^2\langle w, F(x)\rangle - \nabla_x^2\langle w, F(x')\rangle\Big)s\Big] \\
\geq& \max_{w\in C} \langle w, JF(x')s\rangle + \frac{1}{2}s^T\nabla_x^2\langle w, F(x')\rangle s - \\
&- \kappa\|s\|\|JF(x') - JF(x)\| - \frac{\kappa}{2}\|s\|^2\|D^2F(x) - D^2F(x')\| \\
=& f_{x'}(s) - \kappa\|s\|\|JF(x') - JF(x)\| - \frac{\kappa}{2}\|s\|^2\|D^2F(x) - D^2F(x')\|,
\end{aligned}$$

where $f_{x'}(s) = \max_{w\in C}\psi_w(x', s)$. Since $f_{x'}$ is ρ-strongly convex, it is minimized by s' (that is $0 \in \partial_s f_{x'}(s')$) and $f_{x'}(s') = \theta(x')$, we have

$$f_{x'}(s) \geq \theta(x') + \frac{\rho}{2}\|s - s'\|^2.$$

Combining the two above inequalities, we obtain

$$\begin{aligned}
\theta(x) \geq& \theta(x') + \frac{\rho}{2}\|s - s'\|^2 - \kappa\|s\|\|JF(x') - JF(x)\| \\
&- \frac{\kappa}{2}\|s\|^2\|D^2F(x) - D^2F(x')\|.
\end{aligned}$$

Since $\theta(\cdot)$ is continuous and F is $\mathcal{C}^2$, taking the limit $x' \to x$ at the right hand side of the above inequality, we conclude that $s' \to s$, and so, $s(\cdot)$ is continuous. □

For $m = 1$, recall the classical Armijo rule for the Newton direction $s(x)$ at a noncritical point for a scalar function $F : U \to \mathbb{R}$:

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

with $\beta \in (0, \frac{1}{2})$. Note that, in this case, since $\nabla F(x) + \nabla^2 F(x)s(x) = 0$, $\theta(x) = (1/2)\nabla F(x)^T s(x)$, so the Armijo rule can be rewritten as

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

where $\sigma := 2\beta \in (0,1)$.

For $m > 1$, the corresponding Armijo rule for a vector-valued function F and $s(x)$ will be an extension of the above inequality. Now we show the welldefinedness of a backtracking procedure for implementing the Armijo-like rule for an arbitrary s such that $JF(x)s \in -\mathrm{int}(K)$.

Proposition 4.2 ([10, Proposition 2.1])**.** *Consider $x \in U$ and $s \in \mathbb{R}^n$ with $JF(x)s \prec 0$, then, for any $0 < \sigma < 1$ there exists $0 < \bar{t} < 1$ such that*

$$x + ts \in U \quad \text{and} \quad F(x+ts) \prec F(x) + \sigma t JF(x)s$$

hold for all $t \in (0, \bar{t}]$ and all $w \in C$.

Note that in the last proposition we just needed to have continuous differentiability. Under our general hypotheses, i.e., for F twice continuously differentiable and strongly convex, we show now the welldefinedness of the backtracking procedure for the Armijo-like rule with the Newton direction.

Corollary 4.3. *If $x \in U$ is a noncritical point for F, then for $0 < \sigma < 1$, there exists $0 < \bar{t} < 1$ such that*

$$x + ts(x) \in U \quad \text{and} \quad \langle w, F(x+ts(x))\rangle < \langle w, F(x)\rangle + \sigma t \theta(x)$$

hold for all $w \in C$ and all $t \in (0, \bar{t}]$.

Proof. Using Proposition 3.1, item 2, Proposition 4.2 with $s = s(x)$, (15) and (2), we have, for all $w \in C$ and all $t \in (0, \bar{t}]$, that $x + ts(x) \in U$ and

$$\langle w, F(x+ts(x))\rangle < \langle w, F(x)\rangle + \sigma t \langle w, JF(x)s(x)\rangle \leq \langle w, F(x)\rangle + \sigma t \theta(x).$$

□

5 Newton method for vector optimization

We are now ready to define the Newton method for vector optimization. Besides presenting the algorithm, in this section we begin the study of its convergence. First we proof that all accumulation points are efficient and, afterwards, under reasonable hypotheses, we establish its (global) convergence to optimal points.

Newton algorithm for vector optimization

0. Choose C a subbase for K^*, $\sigma \in (0,1)$, $x^0 \in U$, set $k \leftarrow 0$ and define $J = \{1/2^j \mid j = 0, 1, 2, \ldots\}$.

1. Compute $s^k := s(x^k)$ by solving

$$\min_{s\in\mathbb{R}^n} \max_{w\in C} \{\nabla_x \langle w, F(x^k)\rangle^T s + \frac{1}{2} s^T \nabla_x^2 \langle w, F(x^k)\rangle s\}.$$

2. If $s^k = 0$, then stop. Otherwise, continue with **3.**

3. Choose t_k as the largest $t \in J$ such that

$$x^{k+1} = x^k + ts^k \in U,$$
$$\langle w, F(x^k + ts^k)\rangle \le \langle w, F(x^k)\rangle + \sigma t\theta(x^k) \quad \forall\, w \in C.$$

4. Define $x^{k+1} = x^k + t_k s^k$, set $k \leftarrow k + 1$, and go to **1.**

If at step 2 the stopping condition is not satisfied, that is $s^k = s(x^k) \neq 0$, then, existence of t_k is guaranteed by Corollary 4.3. Hence, the above algorithm is well defined.

Note that whenever the Newton algorithm for vector optimization has finite termination, the last iterate is a stationary point, and therefore an optimum for (3). So, the relevant case for the convergence analysis is the one in which the algorithm generates an infinite sequence. In view of this consideration, from now we assume that $\{x^k\}$, $\{s^k\}$ and $\{t_k\}$ are *infinite* sequences generated by the Newton algorithm for vector optimization.

First, we show that the sequence $\{F(x^k)\}$ is K-decreasing and give an estimation for $\langle w, F(x^0) - F(x^{k+1})\rangle$.

Proposition 5.1. *For all* k

1. $\theta(x^k) < 0$.
2. $F(x^{k+1}) \prec F(x^k)$.
3. $\sum_{i=0}^{k} t_i|\theta(x^i)| \le \sigma^{-1}\langle w, F(x^0) - F(x^{k+1})\rangle$ *for any* $w \in C$.

Proof. Item 1 follows from the assumption that $s(x_k) = s_k \neq 0$ for any k and item 2 of Proposition 3.1.

From step 3 and item 1 we conclude that for any i,

$$\langle w, F(x^{i+1})\rangle \le \langle w, F(x^i)\rangle + \sigma t_i\theta(x^i) < \langle w, F(x^i)\rangle \qquad \forall w \in C.$$

So, item 2 follows directly from the above inequalities and (2), while item 3 is obtained by adding the first inequality for $i = 1, \ldots, k$. □

Now we prove that every accumulation point, if any, of the sequence $\{x^k\}$ is efficient.

Proposition 5.2. *If* $\bar{x} \in U$ *is an accumulation point of* $\{x^k\}$*, then* $\theta(\bar{x}) = 0$ *and* $\bar{x}$ *is efficient.*

Proof. Suppose that $\bar{x} \in U$ is an accumulation point of $\{x^k\}$. This means that there exists a subsequence $\{x^{k_j}\}$ such that

$$x^{k_j} \to \bar{x}, \qquad \text{as } j \to \infty.$$

Using item 3 of Proposition 5.1 with $k = k_j$ and taking the limit $j \to \infty$ we conclude that

$$\sum_{i=0}^{\infty} t_i |\theta(x^i)| < \infty.$$

Therefore, $\lim_{k\to\infty} t_k\theta(x^k) = 0$ and, in particular,

$$\lim_{j\to\infty} t_{k_j}\ \theta(x^{k_j}) = 0. \tag{19}$$

Suppose that $\bar{x}$ is nonstationary, which, by Proposition 3.1, is equivalent to

$$\theta(\bar{x}) < 0 \qquad \text{and} \qquad \bar{s} = s(\bar{x}) \neq 0. \tag{20}$$

Define

$$g : \mathbb{R}^m \to \mathbb{R}, \qquad g(u) = \max_{w\in C}\langle w, u\rangle.$$

Using Corollary 4.3 we conclude that there exists an integer q such that

$$\bar{x} + 2^{-q}\bar{s} \in U \quad \text{and} \quad g(F(\bar{x} + 2^{-q}\bar{s}) - F(\bar{x})) < \sigma 2^{-q}\theta(\bar{x}).$$

Since $s(\cdot)$, $\theta(\cdot)$ and $g(\cdot)$ are continuous in their respective domains,

$$\lim_{j\to\infty} s^{k_j} = \bar{s}, \qquad \lim_{j\to\infty} \theta(x^{k_j}) = \theta(\bar{x}) < 0 \tag{21}$$

and, for j large enough

$$x^{k_j} + 2^{-q}s^{k_j} \in U \quad \text{and} \quad g(F(x^{k_j} + 2^{-q}s^{k_j}) - F(x^{k_j})) < \sigma 2^{-q}\theta(x^{k_j}),$$

which, in view of the definition of g and step 3 of the algorithm, implies that $t_{k_j} \geq 2^{-q}$ for j large enough. Hence, taking into account the second limit in (21) we conclude that $\liminf_{j\to\infty} t_{k_j}|\theta(x^{k_j})| > 0$, in contradiction with (19). Therefore, (20) is not true and, by Proposition 3.1, $\theta(\bar{x}) = 0$ and $\bar{x}$ is stationary. Whence, in view of item 3 of Proposition 2.3, $\bar{x}$ is efficient. □

Theorem 5.3. *If x^0 belongs to a compact level set of F, then $\{x^k\}$ converges to an efficient point $\bar{x} \in U$.*

Proof. Let Γ_0 be the $F(x^0)$-level set of F, that is,

$$\Gamma_0 = \{x \in U \mid F(x) \preceq F(x^0)\}.$$

By Proposition 5.1, the sequence $\{F(x^k)\}$ is K-decreasing, so we have $x^k \in \Gamma_0$ for all k. Therefore $\{x^k\}$ is bounded, all its accumulation points belong to $\Gamma_0 \subset U$ and, using also Proposition 5.2, we conclude that all accumulation points of $\{x^k\}$ are efficient. Moreover, since $\{F(x^k)\}$ is K-decreasing, all these accumulation points of $\{x^k\}$ have the same objective value. As U is convex and F is strongly (an therefore strictly) K-convex, there exists just one accumulation point of $\{x^k\}$. □

6 Convergence rate

In this section we analyze the convergence rate of the sequence $\{x^k\}$ generated by the Newton method for vector optimization. First, we provide a bound for $\theta(x^{k+1})$ based on data at the former iterate x^k. Then we show that for k large enough, full Newton steps are performed, that is to say, $t_k = 1$. Then, using this result we prove superlinear convergence. In the end of this section, we prove quadratic convergence under additional regularity assumptions.

Lemma 6.1. *For any k ,there exists $\tilde{w}^k \in \mathrm{conv}(W(x^k, s^k))$ such that*

$$\nabla_x \Psi_{\tilde{w}^k}(x^k) + \nabla_x^2 \Psi_{\tilde{w}^k}(x^k) s^k = 0 \tag{22}$$

and

$$\theta(x^{k+1}) \geq -\frac{1}{2\rho} \left\| \nabla_x \Psi_{\tilde{w}_k}(x^{k+1}) \right\|^2, \tag{23}$$

where $W(x^k, s^k)$ and Ψ_w are given given by (16) *and* (5)*, respectively.*

Proof. First use Lemma 3.2 with $x = x^k$ and let $r = r(x^k)$, $w^i = w^i(x^k)$ and $\alpha_i = \alpha_i(x^k)$ for $i = 1, \dots, r$. Defining

$$\tilde{w}^k = \sum_{i=1}^{r} \alpha_i w^i,$$

and using (18) we conclude that (22) holds for such $\tilde{w}^k$. Moreover,

$$\begin{aligned}\theta(x^{k+1}) &\geq \min_s \sum_{i=1}^r \alpha_i \left[\nabla_x \langle w^i, F(x^{k+1})\rangle^T s + \frac{1}{2} s^T \nabla_x^2 \langle w^i, F(x^{k+1})\rangle s\right] \\ &= \min_s \nabla_x \langle \tilde{w}^k, F(x^{k+1})\rangle^T s + \frac{1}{2} s^T \nabla_x^2 \langle \tilde{w}^k, F(x^{k+1})\rangle s \\ &\geq \min_s \nabla_x \langle \tilde{w}^k, F(x^{k+1})\rangle^T s + \frac{\rho}{2}\|s\|^2,\end{aligned}$$

where the second inequality follows from Corollary 2.2 and the definition of $\tilde{w}^k$. The second part of the lemma follows trivially from the above inequality. □

Recall that, under the assumption of Theorem 5.3, the sequence $\{x^k\}$ converges to an efficient solution.

Theorem 6.2. *Suppose that x^0 belongs to a compact level set of F. Then $\{x^k\}$ converges to an efficient solution $\bar{x}$. Moreover, $t_k = 1$ for k large enough and the convergence of $\{x^k\}$ to $\bar{x}$ is superlinear.*

Proof. Convergence of $\{x^k\}$ to an efficient solution $\bar{x}$ was established in Theorem 5.3. By Proposition 2.3 item 1 and Proposition 3.1 item 2, we know that $\bar{x}$ is critical and

$$s(\bar{x}) = 0. \tag{24}$$

For any $\varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that

$$B(\bar{x}, \delta_\varepsilon) \subset U, \qquad \|D^2F(x) - D^2F(y)\| < \varepsilon \quad \forall x, y \in B(\bar{x}, \delta_\varepsilon).$$

Therefore, for any $w \in \mathbb{R}^m$,

$$\|\nabla_x^2 \Psi_w(x) - \nabla_x^2 \Psi_w(y)\| < \varepsilon \|w\| \quad \forall x, y \in B(\bar{x}, \delta_\varepsilon) \subset U. \tag{25}$$

As $\{x^k\}$ converges to $\bar{x}$, using (24) and the continuity of $s(\cdot)$ (Proposition 4.1, we conclude that $\{s^k\}$ converges to 0. Therefore there exists $k_\varepsilon \in \mathbb{N}$ such that for $k \geq k_\varepsilon$ we have

$$x^k, x^k + s^k \in B(\bar{x}, \delta_\varepsilon)$$

and using (25) for estimating the integral remainder on Taylor's second order expansion at x^k for Ψ_w, given by (5), we conclude that for any $w \in \mathbb{R}^m$

$$\Psi_w(x^k + s^k) \leq \Psi_w(x^k) + \nabla_x \Psi_w(x^k)^T s^k + \frac{1}{2}(s^k)^T \nabla_x^2 \Psi_w(x^k) s^k + \frac{\varepsilon \|w\|}{2}\|s^k\|^2.$$

Taking in the above inequality $w \in C$ and using Corollary 3.3, we obtain, after simple algebraic manipulations,

$$\begin{aligned}\langle w, F(x^k + s^k)\rangle - \langle w, F(x^k)\rangle \leq & \psi_w(x^k, s^k) + \frac{\varepsilon\kappa}{2}\|s^k\|^2 \\ \leq & \theta(x^k) + \frac{\varepsilon\kappa}{2}\|s^k\|^2 \\ \leq & \sigma\theta(x^k) + (1-\sigma)\theta(x^k) + \frac{\varepsilon\kappa}{2}\|s^k\|^2 \\ \leq & \sigma\theta(x^k) + [\varepsilon\kappa - (1-\sigma)\rho]\frac{\|s^k\|^2}{2}.\end{aligned}$$

The above inequality shows that if $\varepsilon < (1-\sigma)\rho/\kappa$, then $t_k = 1$ is accepted for $k \geq k_\varepsilon$.

Suppose that

$$\varepsilon < (1-\sigma)\rho/\kappa.$$

Using the first part of Lemma 6.1 and (25) for estimating the integral remainder on Taylor's first order expansion at x^k for $\Psi_{\tilde{w}^k}$, given by (5) (with $w = \tilde{w}^k$), in $x^k + s^k$, we conclude that for any $k \geq k_\varepsilon$ it holds

$$\begin{aligned}\|\nabla_x \Psi_{\tilde{w}^k}(x^{k+1})\| = & \|\nabla_x \Psi_{\tilde{w}^k}(x^k + s^k)\| \\ = & \left\|\nabla_x \Psi_{\tilde{w}^k}(x^k + s^k) - \left[\nabla_x \Psi_{\tilde{w}^k}(x^k) + \nabla_x^2 \Psi_{\tilde{w}^k}(x^k) s^k\right]\right\| \\ \leq & \varepsilon\kappa\|s^k\|,\end{aligned}$$

which, combined with the second part of Lemma 6.1 shows that

$$|\theta(x^{k+1})| \leq \frac{1}{2\rho}\left[\varepsilon\kappa\|s^k\|\right]^2.$$

Using the above inequality and the second inequality in Corollary 3.3, we conclude that for $k \geq k_\varepsilon$ we have

$$\|x^{k+1} - x^{k+2}\| = \|s^{k+1}\| \leq \frac{\varepsilon\kappa}{\rho}\|s^k\| = \frac{\varepsilon\kappa}{\rho}\|x^k - x^{k+1}\|.$$

Therefore, if $k \geq k_\varepsilon$ and $j \geq 1$ then

$$\|x^{k+j} - x^{k+j+1}\| \leq \left(\frac{\varepsilon\kappa}{\rho}\right)^j \|x^k - x^{k+1}\|. \qquad (26)$$

To prove superlinear convergence, take $0 < \tau < 1$ and define

$$\bar{\varepsilon} = \min\left\{1 - \sigma, \frac{\tau}{1 + 2\tau}\right\}\frac{\rho}{\kappa} \qquad (27)$$

If $\varepsilon < \bar{\varepsilon}$ and $k \geq k_\varepsilon$, using (26) and the convergence of $\{x^i\}$ to $\bar{x}$, we get

$$\begin{aligned}\|\bar{x} - x^{k+1}\| &\leq \sum_{j=1}^{\infty} \|x^{k+j} - x^{k+j+1}\| \\ &\leq \sum_{j=1}^{\infty} \left(\frac{\tau}{1+2\tau}\right)^j \|x^k - x^{k+1}\| = \frac{\tau}{1+\tau}\|x^k - x^{k+1}\|.\end{aligned}$$

Hence,

$$\|\bar{x} - x^k\| \geq \|x^k - x^{k+1}\| - \|x^{k+1} - \bar{x}\| \geq \frac{1}{1+\tau}\|x^k - x^{k+1}\|.$$

Combining the two above inequalities we conclude that, if $\varepsilon < \bar{\varepsilon}$ and $k \geq k_\varepsilon$ then

$$\frac{\|\bar{x} - x^{k+1}\|}{\|\bar{x} - x^k\|} \leq \tau$$

and, since τ was arbitrary in $(0,1)$, the proof is complete. □

Recall that in "classical" Newton method for minimizing a scalar convex $\mathcal{C}^2$ function, the proof of quadratic convergence requires Lipschitz continuity of the objective function's second derivative. Likewise, under the assumption of Lipschitz continuity of D^2F on U, Newton method for vector optimization also converges quadratically.

Theorem 6.3. *Suppose that x^0 belongs to a compact level set of F and D^2F is Lipschitz continuous on U, then $\{x^k\}$ converges quadratically to an efficient solution $\bar{x}$.*

Proof. By Theorem 6.2, $\{x^k\}$ converges superlinearly to an efficient point $\bar{x} \in U$ and $t_k = 1$ for k large enough.

Let L be the Lipschitz constant for D^2F. Then, for $\tilde{w}^k$ and κ as in Lemma 6.1 and Corollary 3.3, respectively, we have that $\nabla^2\Psi_{\tilde{w}^k}$ is κL Lipschitz continuous. Therefore, for k large enough,

$$t_k = 1, \qquad x^{k+1} - x^k = s^k, \quad \text{and} \quad \|\nabla_x \Psi_{\tilde{w}^k}(x^{k+1})\| \leq \frac{\kappa L}{2}\|s^k\|^2, \tag{28}$$

where the last inequality follows from the second equality, Taylor's development of $\nabla_x\Psi_{\tilde{w}^k}$ and (22). Using the above inequality and (23), we obtain

$$-\theta(x^{k+1}) \leq \frac{1}{2\rho}\left[\frac{\kappa L}{2}\|s^k\|^2\right]^2,$$

which, combined with Corollary 3.3 yields

$$\|s^{k+1}\| \leq \frac{\kappa L}{2\rho}\|s^k\|^2. \tag{29}$$

Take $\tau \in (0,1)$. Since $\{x^k\}$ converges superlinearly to $\bar{x}$, there exists N such that for $k \geq N$ we have (28), (29) and

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

Therefore, from the triangle inequality, for $\ell \geq N$ we obtain

$$(1-\tau)\|\bar{x} - x^\ell\| \leq \|x^\ell - x^{\ell+1}\| \leq (1+\tau)\|\bar{x} - x^\ell\|. \tag{30}$$

Whence, using the above second inequality for $\ell = k \geq N$ and the second equality in (28), we have

$$\|s^k\| \leq (1+\tau)\|\bar{x} - x^k\|.$$

while using the first inequality in (30) for $\ell = k+1$ and the second equality in (28) (for $k+1$ instead of k), we obtain

$$(1-\tau)\|\bar{x} - x^{k+1}\| \leq \|s^{k+1}\|.$$

Quadratic convergence of $\{x^k\}$ to $\bar{x}$ follows from the two above inequalities and (29). □

References

[1] C.D. Aliprantis, M. Florenzano, V.F. Martins da Rocha, R. Tourky, *General equilibrium in infinite security markets*, Journal of Mathematical Economics, 40 (2004), pp. 683-699.

[2] C.D. Aliprantis, M. Florenzano, R. Tourky, *General equilibrium analysis in ordered topological vector spaces*, Journal of Mathematical Economics, 40, 3-4 (2004), pp. 247-269.

[3] H. Bonnel, A.N. Iusem and B.F. Svaiter, *Proximal methods in vector optimization*, SIAM Journal of Optimization, 15 (2005), pp. 953-970.

[4] D. Bertsekas, *Convex Analysis and Optimization*, Athena Scientific, Belmont, 2003.

[5] J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, 1996.

[6] J. Fliege and B.F. Svaiter, *Steepest descent methods for multicriteria optimization*, Math. Meth. Oper. Res., 51, 3, (2000), pp. 479-494.

[7] J. Fliege, L.M. Graña Drummond and B.F. Svaiter, *Newton's method for mutiobjective optimization*, SIAM J. Opt., 20, 2 (2209), pp. 602-626.

[8] E.H. Fukuda and L.M. Graña Drummond, *On the convergence of the projected gradient method for vector optimization*, to appear in Optimization.

[9] L.M. Graña Drummond and A.N. Iusem, *A projected gradient method for vector optimization problems*, Computational Optimization and Applications, 28, 1 (2004), pp. 5-29.

[10] L.M. Graña Drummond and B.F. Svaiter, *A steepest descent method for vector optimization*, J. Comput. Appl. Math., 175, 2 (2005), pp. 395-414.

[11] J. Jahn, *Vector optimization*, Springer, Berlin, 2004.

[12] K.M. Miettinen, *Nonlinear multiobjective optimization* Kluwe Academic Publishers, Norwell, 1999.

[13] R.T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, 1972.

[14] H. Sawaragi and T. Tanino, *Theory of multiobjective optimization*, Academic Press, Orlando, 1985.