The self regulation problem as an inexact steepest descent method for multicriteria optimization

G. C. Bento[*]   Da Cruz Neto, J. X. [†]   P. R. Oliveira [‡]   A. Soubeyran [§]

Abstract In this paper, we study an inexact steepest descent method, with Armijo's rule, for multicriteria optimization. The sequence generated by the method is guaranteed to be well-defined. Assuming quasi-convexity of the multicriteria function we prove full convergence of the sequence to a critical Pareto point. As an application, this paper offers a model of self regulation in Psychology, using a recent variational rationality approach.

Keywords. Steepest descent · Pareto optimality · Multicriteria optimization · Quasi-Fejér convergence · Quasi-convexity.

1   Introduction

The steepest descent method with Armijo's rule for real continuously differentiable optimization problem (see, for instance, Burachik et al. [8]), generates a sequence such that any accumulation point of it, if any, is critical for the objective function. This was later generalized for multicriteria optimization by Fliege and Svaiter [11], namely, whenever the objective function is a vectorial function. The full convergence result for real optimization problem was assured when the solution set of the problem be non-empty and the objective function is convex (see Burachik et al. [8]) or, more generally, a quasi-convex function (see Kiwiel

[*]The author was supported in part by CNPq Grant 473756/2009-9 and PROCAD/NF. IME-Universidade Federal de Goiás, Goiânia-GO 74001-970, BR (Email: glaydston@mat.ufg.br)

[†]DM, Universidade Federal do Piauí, Teresina, PI 64049-500, BR (Email: jxavier@ufpi.br). This author was partially supported by CNPq GRANT 302011/2011-0 and PRONEX–Optimization(FAPERJ/CNPq)

[‡]This author was supported in part by CNPq. COPPE/Sistemas-Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21945-970, BR (Email: poliveir@cos.ufrj.br).

[§]GREQAM-AMSE, Aix-Marseille University, antoine.soubeyran@gmail.com

and Murty [19], Bello Cruz and Lucâmbio Pérez [4]). This result has been generalized for convex vectorial optimization by Graña Drummond and Svaiter [14] (see also Graña Drummond and Iusem [15], Fukuda and Graña Drummond [12]) and, in the quasi-convex case, for multicriteria optimization by Bento et al. [6] (see also Bello Cruz et al.[5]). For extensions of other scalar optimization methods to the vectorial setting see, for instance, [7, 10, 9] and references therein.

As far as we know, Bento et al. [6] presented the first result of full convergence of the exact steepest descent method, with Armijo's rule, for quasi-convex multicriteria optimization, which includes contributions within Euclidean and Riemannian context; see also Bello Cruz et al. [5]. In the present paper, we study the method proposed by Fliege and Svaiter [11], which is the inexact version of the method presented in [6]. In this method are admitted relative errors on the search directions, more precisely, an approximation of the exact search direction is computed at each iteration. In this paper, we proved full convergence of the sequence generated by this inexact method to a critical Pareto point associated to quasi-convex multicriteria optimization problems. In particular, we proved full convergence of the sequence to a weak Pareto optimal point in the case that the objective function is pseudo-convex.

The organization of our paper is as follows: In Section 2, we present the self regulation problem in the context of Psychology. In Section 3, the multicriteria problem, the first order optimality condition for it and some basic definitions are presented. In Section 4, the inexact steepest descent method for finding one solution of multicriteria problems is stated and the well-definedness of the sequence generated for it is established. In Section 5, a partial convergence result for continuous differentiability multicriteria optimization is presented without any additional assumption on the objective function. Moreover, assuming that the objective function be quasi-convex and the Riemannian manifold has non-negative curvature, a full convergence result is presented. Finally, Section 6 offers a "distal-proximal" model of self regulation in Psychology, using a recent variational rationality approach ([26, 27, 28, 29]) which modelizes behaviors as an approach, or avoidance, course poursuit between "desired, or undesired enough" ends, and "feasible enough" means.

2   The Self Regulation Problem

In this section devoted to applications we direct the attention of the reader to the very important "multiple goals" self regulation problem in Behavioral sciences. We show the strong link between: i) our paper which extends the steepest descent methods of Fliege and Svaiter [11] to the quasiconvex case in multicri-

teria optimization and ii) the "variational rationality" approach of the "theories of change" of Soubeyran [26, 27, 28, 29].Change problems consider "why, how, and when" it is worthwhile to move from a bad or not so good situation $x \in X$ to a better one $y \in X$ (known or unknown), the limit case of full rationality being an optimizing one, the case of bounded rationality being a better one (in a lot of different formulations, depending of the context). The variational rationality approach examines two polar kinds of "change problems", choice and transformation problems: i) adaptive choice problems like the "choosing the context to choose" problem (the formation of consideration sets), and ii) transformation problems like creation and destruction, invention, innovation, the evolution of institutions, dynamics interactions (dynamic games), health, behavioral, organizational and cultural changes, ... in Economics, Decision theory, Management, Psychology, Artificial Intelligence, Philosophy, Sociology, Applied Mathematics (Variational Analysis, Optimization and Variational Inequalities). In this "variational context" our present paper shows how setting joint distal and proximal goals greatly help to reach a distal goal. It offers an "aspiration driven local search proximal algorithm". This variational approach emphasizes, each step of the process, two main variational principles (among others, in more general settings): a "satisficing-but not too much sacrificing" principle and a "worthwhile to change" principle. Because the state space of situations is the Euclidian space $X = \mathbb{R}^n$, changes $u = y - x$ from a given situation x to a hopefully better situation y can be characterized by their directions $v \in X$ and their depth (length) $t > 0$. In this context where $u = tv$ these two variational principles specialize to,

i) the choice, each step, of a "satisficing but not too much sacrificing" direction (a directional "satisficing-but not too much sacrificing" principle);

ii) the choice, each step, of a "worthwhile change" (a "worthwhile to change" step length principle).

2.1 Self regulation problems

Self regulation considers the systematic activities (efforts) made to direct thoughts, feelings, and actions, towards the attainment of one's goals (Zimmerman [32]). A goal is a conscious or unconscious mental representation of some future end (to approach or to avoid), more or less distant, abstract (concrete), vague (precise), desirable and feasible.

Goals can be more or less desirable and more or less feasible. Related to the desirability aspects are conscious or not, vague or concrete, distal or proximal, long term or short term, extrinsic or intrinsic, set by others or oneself, individual or collective, learning or performance oriented contents, high or low in

commitment .... Related to the feasibility aspects are importance, priority, urgency, direction, intensity, difficulty, measurability,

Self regulation have two aspects. The positive side of self regulation considers purposive processes where agents engage in goal-directed actions. It examines goal setting, goal striving and goal pursuit processes.

- Goal setting is the mental process of moving from the consideration of distal goals to the formation of more proximal goals. Distal goals are desired future ends (visions, imagined desired futures), either promotion aspirations (like ideals, fantasies, dreams, wishes, hopes and challenges) or prevention aspirations (like oughts and obligations). They represent desirable but quite irrealistic distal and vague ends (higher order goals). Proximal goals can be wants, intentions, task goals, i.e much more feasible but less desirable intermediate ends (sub goals).

- Goal striving (goal implementation) examines the transition phase between setting a distal goal and reaching it.

- Goal pursuit (goal revision) focuses on the final phase, after reaching the given goal or failing to reach it. It examines the role of feedbacks (self evaluations of successes and failures, including the revision of causal attributions and self efficacy beliefs, see Tolli and Schmidt [30]) in order to revise goals.

The negative side of self regulation considers what an agent must refrain to do instead of what he must do to set and attain some given goal. This negative aspect of self regulation is named self control (overriding of one action tendency in order to attain an other goal). It considers self regulation failures like lack of vision, the inability to transform irrealistic aspirations into intentions and realistic proximal goals (preparation to action problems), procrastination and inertia (starting problems), interruptions, distractions, temptations, lack of feedbacks, lack of interest, perseverance and motivation (on the track problems) and goal disengagement (ends problems).

Our paper considers only the positive aspect of self regulation. It focuses on proximal goal setting activities, examines some aspects of goal revision activities, and renounce to consider goal striving activities.

2.2   Setting proximal goals

The Michael Jordan "step by step" principle: The famous basketball player Michael Jordan wrote the following about goal setting in his book (Jordan and Miller [18]), "I approach everything step by step .... I had always set short-term goals. As I look back, each one of the steps or successes led to the next one. When I got cut from the varsity team as a sophomore in high school, I learned something. I knew I

never wanted to feel that bad again .... So I set a goal of becoming a starter on the varsity. That's what I focused on all summer. When I worked on my game, that's what I thought about. When it happened, I set another goal, a reasonable, manageable goal that I could realistically achieve if I worked hard enough .... I guess I approached it with the end in mind. I knew exactly where I wanted to go, and I focused on getting there. As I reached those goals, they built on one another. I gained a little confidence every time I came through ....

Goal hierarchies and goal proximity: the Bandura dual "proximal-distal" self regulation principle: Bandura [3] argued that people possess multiple systems of goals, hierarchically arranged from proximal goals to extreme distal goals. Goal proximity defines "how far goals are conceptualized into the future". A goal hierarchy interconnects at least three levels of goals: peak goals (higher order goals, like visions, dreams, fantasies, aspirations, ideals, wishes, hopes), distal goals (challenges), and task goals.... A subset of task goals can be subordinate to distal goals which can be subordinate to peak goals. Hence, the proximal goal distinction is relative to the interconnected network of goals, other goal's providing the temporal context). The main point to be emphasized is that distal goals and proximal goals serve different and complementary conative functions (connected to cognition, affect and motivation) related to goal difficulty, goal commitment, psychological distance....

i) distal goals define desired ends (enduring aspirations) that attract individuals;

ii) proximal goals regulate immediate conative functions, which provide the ways to find and follow a path of step by step changes moving from the initial situation to approach the desired end or avoid an undesirable end. In this context it is important to distinguish task goals and strategies. The former defines what is to be accomplished, and the later defines how it is to be accomplished (Wood and Bandura [31]).

3 The Multicriteria Problem

In this section, we present the multicriteria problem, the first order optimality condition for it and some basic definitions.

Let $I := \{1, \ldots, m\}$, $\mathbb{R}_+^m = \{x \in \mathbb{R}^m : x_i \geq 0, j \in I\}$ and $\mathbb{R}_{++}^m = \{x \in \mathbb{R}^m : x_j > 0, j \in I\}$. For $x, y \in \mathbb{R}_+^m$, $y \succeq x$ (or $x \preceq y$) means that $y - x \in \mathbb{R}_+^m$ and $y \succ x$ (or $x \prec y$) means that $y - x \in \mathbb{R}_{++}^m$.

Given a continuously differentiable vector function $F : \mathbb{R}^n \to \mathbb{R}^m$, we consider the problem of finding a

optimum Pareto point of F, i.e., a point $x^* \in \mathbb{R}^n$ such that there exists no other $x \in \mathbb{R}^n$ with $F(x) \preceq F(x^*)$ and $F(x) \neq F(x^*)$. We denote this unconstrained problem as

$$\min_{x \in \mathbb{R}^n} F(x). \tag{1}$$

Let F be given by $F(x) := (f_1(x), \ldots, f_m(x))$. We denote the jacobian of F by

$$JF(x) := (\nabla f_1(x), \ldots, \nabla f_m(x)), \qquad x \in \mathbb{R}^n,$$

and the image of the jacobian of F at a point $x \in \mathbb{R}^n$ by

$$\mathrm{Im}(JF(x)) := \{JF(x)v = (\langle \nabla f_1(x), v \rangle, \ldots, \langle \nabla f_m(x), v \rangle) : v \in \mathbb{R}^n\}.$$

Using the above equality, the first-order optimality condition for Problem 1 (see, for instance, [11]) is stated as

$$x \in \mathbb{R}^n, \qquad \mathrm{Im}(JF(x)) \cap (-\mathbb{R}^m_{++}) = \emptyset. \tag{2}$$

Note that the condition in (2) generalizes, to multicriteria optimization, the classical condition "gradient equals zero" for the real-valued case.

In general, (2) is necessary, but not sufficient, for optimality. A point of $\mathbb{R}^n$ satisfying (2) is called *critical Pareto point*.

4   Inexact Steepest Descent Methods for Multicriteria Problems

In this section, we state the inexact steepest descent methods for solving multicriteria problems admitting relative errors in the search directions, more precisely, an approximation of the exact search direction is computed at each iteration, as considered in [11, 14, 12].

Let $x \in \mathbb{R}^n$ be a point which is not critical Pareto point. Then there exists a direction $v \in \mathbb{R}^n$ satisfying

$$JF(x)v \in -\mathbb{R}^m_{++},$$

that is, $JF(x)v \prec 0$. In this case, v is called a *descent direction* for F at x.

For each $x \in \mathbb{R}^n$, we consider the following unconstrained optimization problem in $\mathbb{R}^n$

$$\min_{v \in \mathbb{R}^n} \left\{ \max_{i \in I} \langle \nabla f_i(x), v \rangle + (1/2)\|v\|^2 \right\}, \qquad I := \{1, \ldots, m\}. \tag{3}$$

Lemma 4.1. *The following statements hold:*

i) The unconstrained optimization problem in (3) has only one solution. Moreover, the vector v is the solution of the problem in (3) if and only if there exists $\alpha_i \geq 0$, $i \in I(x,v)$, such that

$$v = - \sum_{i \in I(x,v)} \alpha_i \nabla f_i(x), \qquad \sum_{i \in I(x,v)} \alpha_i = 1,$$

where $I(x,v) := \{i \in I : \langle \nabla f_i(x), v \rangle = \max_{i \in I} \langle \nabla f_i(x), v \rangle \}$;

ii) If x is critical Pareto point of F and v denotes the solution of the problem in (3), then $v = 0$ and the optimal value associated to v is equal to zero;

iii) If $x \in \mathbb{R}^n$ is not a critical Pareto point of F and v is the solution of the problem in (3), then $v \neq 0$ and

$$\max_{i \in I} \langle \nabla f_i(x), v \rangle + (1/2)\|v\|^2 < 0.$$

In particular, v is a descent direction for F at x.

Proof. The proof of the item i can be found in [6]. For the proof of the remaining items, see, for example, [11]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Remark 4.1. *From the item i of Lemma 4.1 we note that the solution of the minimization problem (3) is of the form:*

$$v = -JF(x)^t w, \quad w = (\alpha_1, \ldots, \alpha_m) \in \mathbb{R}_+^m, \quad \|w\|_1 = 1 \quad (\text{sum norm in } \mathbb{R}^m),$$

with $\alpha_i = 0$ for $i \in I \setminus I(x,v)$. In other words, if $S := \{e_i \in \mathbb{R}^m : i \in I\}$ (set of the elements of the canonical base of Euclidean space $\mathbb{R}^m$), then w is an element of the convex hull of $S(x,v)$, where

$$S(x,v) := \{\bar{u} \in S : \langle \bar{u}, JF(x)v \rangle = \max_{u \in S} \langle u, JF(x)v \rangle \}. \qquad\qquad (4)$$

Note that the minimization problem (3) may be rewritten as follows:

$$\min_{v \in \mathbb{R}^n} \left\{ \max_{u \in S} \langle u, JF(x)v \rangle + (1/2)\|v\|^2 \right\} = \min_{v \in \mathbb{R}^n} \left\{ \max_{u \in S} \langle JF(x)^t u, v \rangle + (1/2)\|v\|^2 \right\}.$$

In view of the previous lemma and (3), we define the steepest descent direction function for F as follows.

Definition 4.1. *The steepest descent direction function for F is defined as*

$$\mathbb{R}^n \ni x \longmapsto v(x) := argmin_{v \in \mathbb{R}^n} \left\{ \max_{i \in I} \langle \nabla f_i(x), v \rangle + (1/2)\|v\|^2 \right\} \in \mathbb{R}^n.$$

Remark 4.2. *This definition was proposed in* [11]. *Note that, from the item i of* Lemma 4.1 *it follows that the steepest descent direction for vector functions becomes the steepest descent direction when* $m = 1$.

The optimal value associated to $v(x)$ will be denoted by $\alpha(x)$. Note that the function

$$\mathbb{R}^n \ni x \longmapsto \max_{i \in I} \langle \nabla f_i(x), v \rangle + (1/2)\|v\|^2 \in \mathbb{R},$$

is strongly convex with modulus $1/2$ and

$$0 \in \partial \left(\max_{i \in I} \langle \nabla f_i(x), \, . \, \rangle + 1/2\|.\|^2 \right)(v(x)).$$

So, for all $v \in \mathbb{R}^n$,

$$\max_{i \in I} \langle \nabla f_i(x), v \rangle + (1/2)\|v\|^2 - \alpha(x) \geq 1/2\|v - v(x)\|^2. \tag{5}$$

Lemma 4.2. *The steepest descent direction function for* F, $\mathbb{R}^n \ni x \mapsto v(x) \in \mathbb{R}^n$, *is continuous. In particular, the function* $\mathbb{R}^n \ni x \mapsto \alpha(x) \in \mathbb{R}$ *is also continuous.*

Proof. See [6] for the proof of the first part. The second part is a immediate consequence of the first. $\quad\square$

Definition 4.2. *Let* $\sigma \in [0, 1)$. *A vector* $v \in \mathbb{R}^n$ *is say be a* $\sigma-$*approximate steepest descent direction at* x *for* F *if*

$$\max_{1 \leq i \leq m} \langle \nabla f_i(x), v \rangle + 1/2\|v\|^2 \leq (1 - \sigma)\alpha(x).$$

Note that the exact steepest descent direction at x is a σ-approximate steepest descent direction for F with $\sigma = 0$. As a immediate consequence of Lemma 4.1 together with last definition, it is possible to prove the following:

Lemma 4.3. *Given* $x \in \mathbb{R}^n$,

    a) $v = 0$ *is a* σ*-approximate steepest descent direction at* x *if, only if,* x *is a critical Pareto point;*

    b) if x *is not a critical Pareto point and* v *is a* σ*-approximate steepest descent direction at* x, *then* v *is a descent direction for* F.

Next lemma establishes the degree of proximity between an approximate direction v and the exact direction $v(x)$, in terms of the optimal value $\alpha(x)$.

Lemma 4.4. *Let* $\sigma \in [0, 1)$. *If* $v \in \mathbb{R}^n$ *is a* $\sigma-$*approximate steepest descent direction at* x, *then*

$$\|v - v(x)\|^2 \leq 2\sigma|\alpha(x)|.$$

Proof. The proof follows from (5) combined with Definition 4.2. See [14]. □

A particular class of σ- approximate steepest descent directions for F at x is given by the directions $v \in \mathbb{R}^n$ which are *scalarization compatible*, i.e., such that there exists $\tilde{w} \in \text{conv}S$ with

$$v = -JF(x)^t \tilde{w}. \tag{6}$$

Note that $\tilde{w}$ determines a scalar function $g(x) := \langle \tilde{w}, F(x) \rangle$ whose steepest descent direction coincides with v, which justifies the name previously attributed to the direction v; see [14] for a good discussion.

Next proposition establishes a sufficient condition for v, given as in (6), to be a σ-approximate steepest descent direction for F at x.

Proposition 4.1. *Let $\sigma \in [0, 1)$ and v as in (6). If*

$$\max_{i \in I} \langle \nabla f_i(x), v \rangle \leq -(1 - \sigma/2)\|v\|^2,$$

or equivalently,

$$\max_{u \in S} \langle JF(x)^t u, v \rangle \leq -(1 - \sigma/2)\|v\|^2, \tag{7}$$

then v is a σ-approximate steepest descent direction for F at x.

Proof. See [14]. □

From Remark 4.1 we note that, for each $x \in \mathbb{R}^n$, the steepest descent direction for F at x, $v(x)$, is scalarization compatible. Next lemma tell us that $v(x)$ satisfies the sufficient condition of the last proposition with $\sigma = 0$ and, hence, that such condition is natural.

Lemma 4.5. *The following statements hold:*

i) $\alpha(x) = -(1/2)\|v(x)\|^2$;

ii) $\max_{u \in S} \langle JF(x)^t u, v(x) \rangle = -\|v(x)\|^2$.

Proof. In order to prove the item i note that

$$\alpha(x) = \max_{u \in S} \langle JF(x)^t u, v(x) \rangle + (1/2)\|v(x)\|^2. \tag{8}$$

Moreover, from Remark 4.1, we have

$$v(x) = -JF(x)^t w, \qquad w \in \text{conv}S(v(x)), \qquad S(v(x)) := S(x, v(x)), \tag{9}$$

9

where conv$S(v(x))$ denotes the convex hull of $S(v(x))$. So, combining (8) and (9) with the definition of $S(v(x))$, we get

$$\alpha(x) = \langle JF(x)^t \bar{u}, -JF(x)^t w \rangle + (1/2)\|JF(x)^t w\|^2, \qquad \bar{u} \in S(v(x)).$$

Hence,

$$\alpha(x) = \langle JF(x)^t w, -JF(x)^t w \rangle + (1/2)\|JF(x)^t w\|^2,$$

from where it follows the item i. The item ii is an immediate consequence of the item i combined with (8). $\qquad\square$

The *inexact steepest descent method with the Armijo rule* for solving the unconstrained optimization problem (1) is as follows:

Method 4.1 (Inexact steepest descent method with Armijo rule)**.**

INITIALIZATION. *Take $\beta \in (0, 1)$ and $x^0 \in \mathbb{R}^n$. Set $k = 0$.*

STOP CRITERION. *If x^k is a critical Pareto point STOP. Otherwise.*

ITERATIVE STEP. *Compute a σ- approximate steepest descent direction v^k for F at x^k and the steplength $t_k \in]0, 1]$ as follows:*

$$t_k := \max\left\{ 2^{-j} : j \in \mathbb{N}, \; F\left(x^k + 2^{-j} v^k\right) \preceq F(x^k) + \beta 2^{-j} \, JF(x^k) v^k \right\}, \tag{10}$$

and set

$$x^{k+1} := x^k + t_k v^k, \tag{11}$$

and GOTO STOP CRITERION.

Remark 4.3. *The previous method was proposed by Fliege and Svaiter [11] and becomes the classical steepest descent method when $m = 1$. Other variants of Method 4.1 can be found in [14, 15, 12].*

Next proposition ensures that the sequence generated by the Method 4.1 is well-defined.

Proposition 4.2. *The sequence $\{x^k\}$ generated by the steepest descent method with Armijo rule is well-defined.*

Proof. The proof follows from the item ii of Lemma 4.3 combined with the fact that F is continuously differentiable. See [11] for more details. $\qquad\square$

5 Convergence Analysis

In this section, we present a partial convergence result without any additional assumption on F besides the continuous differentiability. In the sequel, assuming quasi-convexity of F and following the ideas of [14, 6], we extend the full convergence result presented in [14] for quasi-convex multicriteria optimization. It can be immediately seen that, if Method 4.1 terminates after a finite number of iterations, then it terminates at a critical Pareto point. From now on, we will assume that $\{x^k\}$, $\{v^k\}$ and $\{t_k\}$ are infinite sequences generated by Method 4.1.

To simplify the notation, in what follows we will utilize the scalar function $\varphi : \mathbb{R}^m \to \mathbb{R}$ defined as follows:

$$\varphi(y) = \max_{i \in I} \langle y, e_i \rangle, \qquad I = \{1, \ldots, m\},$$

where $\{e_i\} \subset \mathbb{R}^m$ is the canonical base of the space $\mathbb{R}^m$. It is easy to see that the following properties of the function φ hold:

$$\varphi(x + y) \leq \varphi(x) + \varphi(y), \qquad \varphi(tx) = t\varphi(x), \qquad x, y \in \mathbb{R}^m, \quad t \geq 0. \tag{12}$$

$$x \preceq y \quad \Rightarrow \quad \varphi(x) \leq \varphi(y), \qquad x, y \in \mathbb{R}^m. \tag{13}$$

5.1 Partial Convergence Result

The following theorem shows that if F is continuously differentiable then the sequence of the functional values of the sequence $\{x^k\}$, $\{F(x^k)\}$, is monotonously decreasing and the accumulation points of $\{x^k\}$ are critical Pareto points. The proof of the next theorem can be found partly in [11] and [14]. We chose to present a proof within this paper.

Theorem 5.1. *The following statements hold:*

i) $\{F(x^k)\}$ *is decreasing;*

ii) *If* $\{x^k\}$ *has accumulation point, then* $\{t_k^2 \|v^k\|^2\}$ *is a summable sequence and*

$$\lim_{k \to +\infty} t_k \|v^k\|^2 = 0; \tag{14}$$

iii) *Each accumulation point of the sequence* $\{x^k\}$, *if any, is a critical Pareto point.*

Proof. The iterative step in Method 4.1 implies that

$$F(x^{k+1}) \preceq F(x^k) + \beta t_k JF(x^k)v^k, \qquad x^{k+1} = x^k + t_k v^k, \qquad k = 0, 1, \dots. \tag{15}$$

Since $\{x^k\}$ is an infinite sequence, for all k, x^k is not a critical Pareto point of F. Thus, the item i follows from the item ii of Lemma 4.3 combined with the last vector inequality.

Suppose now that $\{x^k\}$ has an accumulation point $\bar{x} \in \mathbb{R}^n$ and let $\{x^{k_s}\}$ be a subsequence of $\{x^k\}$ such that $\lim_{s \to +\infty} x^{k_s} = \bar{x}$. Since F is continuous and $\lim_{s \to +\infty} x^{k_s} = \bar{x}$ we have $\lim_{s \to +\infty} F(x^{k_s}) = F(\bar{x})$. So, taking into account that $\{F(x^k)\}$ is a decreasing sequence and has $F(\bar{x})$ as an accumulation point, it is easy to conclude that the whole sequence $\{F(x^k)\}$ converges to $F(\bar{x})$. So, from the definition of the function φ, we conclude that $\{\varphi(F(x^k))\}$ converges to $\varphi(F(\bar{x}))$ and, in particular,

$$\varphi(F(\bar{x})) \le \varphi(F(x^k)), \qquad k = 0, 1, \dots. \tag{16}$$

From (15), (12), (13) and definition of v^k, we obtain

$$\varphi(F(x^{k+1})) \le \varphi(F(x^k)) + \beta t_k \left((1 - \sigma)\alpha(x^k) - (1/2)\|v^k\|^2\right), \qquad k = 0, 1 \dots,$$

or, equivalently,

$$\varphi(F(x^{k+1})) - \varphi(F(x^k)) \le \beta \left((1 - \sigma)t_k\alpha(x^k) - (1/2)t_k\|v^k\|^2\right), \qquad k = 0, 1 \dots. \tag{17}$$

Adding the last inequality from $k = 0$ to n and taking into account that $|\alpha(x^k)| = -\alpha(x^k)$, we have

$$\varphi(F(p^{n+1})) - \varphi(F(p^0)) \le -\beta \sum_{k=0}^{n} \left[(1 - \sigma)t_k|\alpha(x^k)| + (1/2)t_k\|v^k\|^2\right].$$

Thus, because $\beta \in (0, 1)$ and $\varphi(F(\bar{x})) \le \varphi(F(p^{n+1}))$ (see (16)), from the last inequality, we get

$$\sum_{k=0}^{n} \left[(1 - \sigma)t_k|\alpha(x^k)| + (1/2)t_k\|v^k\|^2\right] \le \frac{\varphi(F(p^0)) - \varphi(F(\bar{x}))}{\beta}, \qquad n \ge 0.$$

But this tell us that (recall that $\sigma \in [0, 1)$)

$$\sum_{k=0}^{+\infty} t_k|\alpha(x^k)| < +\infty \qquad \text{and} \qquad \sum_{k=0}^{+\infty} t_k\|v^k\|^2 < +\infty, \tag{18}$$

from which follows the second part of the item ii. The first part of the item ii follows from last inequality in (18) together with the fact that $t_k \in (0, 1]$.

We assume initially that $\bar{x}$ is an accumulation point of the sequence $\{x^k\}$ and that $\{x^{k_s}\}$ is a subsequence of $\{x^k\}$ converging to $\bar{x}$. From Lemma 4.2, we may conclude that $\{v(x^{k_s})\}$ and $\{\alpha_{x^{k_s}}\}$ converge, respectively, to $v(\bar{x})$ and $\alpha_{\bar{x}}$. In particular, from Lemma 4.4, it follows that $\{v^{k_s}\}$ is bounded and, hence, has a convergent subsequence. Moreover, the sequence $\{t_k\} \subset]0, 1]$ also has an accumulation point $\bar{t} \in [0, 1]$. We assume, without loss of generality, that $\{t_{k_s}\}$ converges to $\bar{t}$ and $\{v^{k_s}\}$ converges to some $\bar{v}$. From the equality (14), it follows that

$$\lim_{s \to +\infty} t_{k_s} \|v^{k_s}\|^2 = 0. \tag{19}$$

We have two possibilities to consider:

a) $\bar{t} > 0$;

b) $\bar{t} = 0$.

Assume that item **a** holds. Then, from (19), it follows that $\bar{v} = 0$. On the other hand, from the Definition 4.2 of v^k, we obtain

$$\max_{1 \leq i \leq m} \langle \nabla f_i(x^{k_s}), v^{k_s} \rangle + 1/2 \|v^{k_s}\|^2 \leq (1 - \sigma)\alpha(x^{k_s}), \qquad s = 0, 1 \dots.$$

Letting s go to $+\infty$ in above inequality, it follows that $\bar{v} = 0$ is a σ-approximation steepest descent method for F at $\bar{x}$ and, from the item i of Lemma 4.3, we conclude that $\bar{x}$ is a critical Pareto point of F.

Now, assume that item **b** holds true. Since v^{k_s} is a σ-approximation steepest descent method for F at x^{k_s} and $\{x^{k_s}\}$ is not a critical Pareto point, we have

$$\max_{i \in I} \langle \nabla f_i(x^{k_s}), v^{k_s} \rangle \leq \max_{i \in I} \langle \nabla f_i(x^{k_s}), v^{k_s} \rangle + (1/2)\|v^{k_s}\|^2 < (1 - \sigma)\alpha(x^{k_s}) < 0,$$

where the last inequality is a consequence of the item iii of Lemma 4.1. Hence, letting s go to $+\infty$ in the last inequalities and using that $\{v^{k_s}\}$ converges to $\bar{v}$, we obtain

$$\max_{i \in I} \langle \nabla f_i(\bar{x}), v(\bar{x}) \rangle \leq (1 - \sigma)\alpha(\bar{x}) \leq 0. \tag{20}$$

Take $r \in \mathbb{N}$. Since $\{t_{k_s}\}$ converges to $\bar{t} = 0$, we conclude that if s is large enough,

$$t_{k_s} < 2^{-r}.$$

From (10) this means that the Armijo condition (15) is not satisfied for $t = 2^{-r}$, i.e.,

$$F(x^k + 2^{-j}v^{k_s}) \npreceq F(x^{k_s}) + \beta 2^{-r} JF(x^{k_s})v^{k_s},$$

13

which means that there exists at least one $i_0 \in I$ such that

$$f_{i_0}(x^{k_s} + 2^{-r}v^{k_s}) > f_{i_0}(x^{k_s}) + \beta 2^{-r} \langle \nabla f_{i_0}(x^{k_s}), v^{k_s} \rangle.$$

Letting s go to $+\infty$ in the above inequality, taking into account that ∇f_{i_0} and exp are continuous and using that $\{v^{k_s}\}$ converges to $\bar{v}$, we obtain

$$f_{i_0}(\bar{x} + 2^{-r}v(\bar{x})) \geq f_{i_0}(\bar{x}) + \beta 2^{-r} \langle \nabla f_{i_0}(\bar{x}), v(\bar{x}) \rangle.$$

The last inequality is equivalent to

$$\frac{f_{i_0}(\bar{x} + 2^{-r}v(\bar{x})) - f_{i_0}(\bar{x})}{2^{-r}} \geq \beta \langle \nabla f_{i_0}(\bar{x}), v(\bar{x}) \rangle,$$

which, letting r go to $+\infty$ and assuming that $0 < \beta < 1$, yields $\langle \nabla f_{i_0}(\bar{x}), v(\bar{x}) \rangle \geq 0$. Hence,

$$\max_{i \in I} \langle \nabla f_i(\bar{x}), v(\bar{x}) \rangle \geq 0.$$

Combining the last inequality with (20) and taking into account that $\sigma \in [0, 1)$, we have

$$\alpha(\bar{x}) = 0.$$

Therefore, from the item iii of Lemma 4.1 it follows that $\bar{x}$ is a critical Pareto point of F, and the proof is concluded. $\qquad\square$

Remark 5.1. *If the sequence $\{x^k\}$ begins in a bounded level set, for example, if*

$$L_F(F(p_0)) := \{x \in \mathbb{R}^n : F(x) \preceq F(p_0)\},$$

is a bounded set, then, since F is a continuous function, $L_F(F(p_0))$ is a compact set. So, item i of Theorem 5.1 implies that $\{x^k\} \subset L_F(F(p_0))$ and consequently $\{x^k\}$ is bounded. In particular, $\{x^k\}$ has at least one accumulation point.

5.2 Full Convergence

In this section, under the quasi-convexity assumption on F, full convergence of the steepest descent method is obtained.

Definition 5.1. *Let $H : \mathbb{R}^n \to \mathbb{R}^m$ be a vectorial function.*

14

i) H is called convex iff for every $x, y \in \mathbb{R}^n$, the following holds:

$$H((1-t)x + ty) \preceq (1-t)H(x) + tH(y), \qquad t \in [0,1];$$

ii) H is called quasi-convex iff for every $x, y \in \mathbb{R}^n$, the following holds:

$$H((1-t)x + ty) \preceq \max\{H(x), H(y)\}, \qquad t \in [0,1],$$

where the maximum is considered coordinate by coordinate;

iii) H is called pseudo-convex iff H is differentiable and, for every $x, y \in \mathbb{R}^n$, the following holds:

$$JH(x)(y-x) \not\prec 0 \quad \Rightarrow \quad H(y) \not\prec H(x).$$

Remark 5.2. *For the two first above definitions see* Definition 6.2 *and* Corollary 6.6 *of* [23], *pages* 29 *and* 31, *respectively. For the third definition see Definition* 9.2.3 *of* [13], *page* 274. *Note that H is convex (resp. quasi-convex) iff, H is componentwise convex (resp. quasi-convex) . On the other hand, H componentwise pseudo-convex is a sufficient condition , but not necessary for H to be pseudo-convex; see Theorem* 9.2.3 *of* [13], *page* 274 *and Remark* 5.3. *It is immediate from above definitions that if H is convex then it is quasi-convex (the reciprocal is clearly false). If H is differentiable, convexity of H implies that for every $x, y \in \mathbb{R}^n$,*

$$JH(x)(y-x) \preceq H(y) - H(x), \tag{21}$$

from which we may conclude that H is pseudo-convex. It is easy to obtain an example showing that the reciprocal is false.

Next proposition provides a characterization for differentiable quasi-convex functions.

Proposition 5.1. *Let $H : \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable function. Then, H is a quasi-convex function if, only if, for every $x, y \in \mathbb{R}^n$, it holds*

$$H(y) \prec H(x) \quad \Rightarrow \quad JH(x)(y-x) \preceq 0.$$

Proof. Let us assume that, for every pair of points $x, y \in \mathbb{R}^n$, it holds

$$H(y) \prec H(x) \quad \Rightarrow \quad JH(x)(x-y) \preceq 0. \tag{22}$$

Take $\tilde{x}, \tilde{y} \in \mathbb{R}^n$ and assume that holds

$$H(\tilde{y}) \prec H((1 - t)\tilde{x} + t\tilde{y}), \qquad t \in [0, 1).$$

Using (22) with $y = \tilde{y}$ and $x = (1 - t)\tilde{x} + t\tilde{y}$, we obtain $(1 - t)JH((1 - t)\tilde{x} + t\tilde{y})(\tilde{x} - \tilde{y}) \preceq 0$, which implies

$$\frac{d}{dt}h_i((1 - t)\tilde{x} + t\tilde{y}) = \langle \nabla h_i((1 - t)\tilde{x} + t\tilde{y}), \tilde{y} - \tilde{x} \rangle \leq 0, \qquad i \in \{1, \ldots, m\},$$

where $h_1, \ldots, h_m$ represent the coordinate functions of H. But this implies that

$$h_i((1 - t)\tilde{x} + t\tilde{y}) \leq h_i(\tilde{x}), \quad i \in \{1, \ldots, m\},$$

and, hence, that

$$H((1 - t)\tilde{x} + t\tilde{y}) \preceq H(\tilde{p}) = \max\{H(\tilde{x}), H(\tilde{y})\},$$

which proves the first part of the proposition. The proof of the second part follows immediately from the definition of quasi-convexity combined with differentiability of H; see [6] for more details. $\qquad\square$

From the previous proposition follows immediately that pseudo-convex functions are quasi-convex. The reciprocal is naturally false. Next proposition provides a sufficient condition for a differentiable quasi-convex function to be pseudo-convex.

Definition 5.2. *A point $x^* \in \mathbb{R}^n$ is a weak optimal Pareto point of F iff there is no $x \in \mathbb{R}^n$ with $F(x) \prec F(x^*)$.*

Proposition 5.2. *Let $H : \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable quasi-convex function. If each critical Pareto point of H is a weak Pareto optimal point, then H is a pseudo-convex function.*

Proof. Take $y \in \mathbb{R}^n$. Since that, by hypothesis, each critical Pareto point is an optimal weak Pareto, if y is critical Pareto we have nothing to do. Let us suppose that y is not a critical Pareto point. Then, there exists $v \in \mathbb{R}^n$ such that

$$JH(y)v \prec 0. \tag{23}$$

Let us assume, by contradiction, that H is not pseudo-convex. In this case, there exists $\tilde{x} \in \mathbb{R}^n$ such that $H(\tilde{x}) \prec H(y)$, with

$$JH(y)(\tilde{x} - y) \nprec 0. \tag{24}$$

16

From (23) and (24), it follows that

$$JH(y)(\tilde{x} - y) - \beta JH(y)v \not\preceq 0, \qquad \beta > 0. \tag{25}$$

Now, since $H(\tilde{x}) \prec H(y)$, from the continuity of H there exists $\delta > 0$ such that $H(z) \prec H(y)$ for all $z \in B(\tilde{x}, \delta)$ (ball with center in $\tilde{x}$ and ray δ). In particular, $H(\tilde{x} - (\delta/2)(v/\|v\|)) \prec H(y)$ and, because H is quasi-convex, we obtain

$$JH(y)\left(\tilde{x} - (\delta/2)(v/\|v\|) - y\right) \preceq 0.$$

But this tell us that with $\beta = \delta/(2\|v\|)$, we have

$$JH(y)(\tilde{x} - y) - \beta JH(y)v \preceq 0,$$

which is a contradiction with (25), and the resulted is proved. $\square$

Remark 5.3. *Consider the following vectorial function $H : \mathbb{R} \to \mathbb{R}^2$ given by $H(t) = (t, -t^3/3)$. Note that H is not compenentwise pseudo-convex because $h_2(t) := -t^3/3$ is not pseudo-convex. However, since H is quasi-convex and each critical Pareto point of H is weak Pareto optimal point for H, from last proposition, it follows that H is pseudo-convex.*

We know that criticality is a necessary, but not sufficient, condition for optimality. In [6] the authors proved that, under convexity of the vectorial function F, criticality is equivalent to the weak optimality. Next we prove that the equivalence still happens if F is just pseudo-convex.

Proposition 5.3. *Let $H : \mathbb{R}^n \to \mathbb{R}^m$ be a pseudo-convex function. Then, $x \in \mathbb{R}^n$ is a critical Pareto point of H, i.e.,*

$$Im(\nabla H(x)) \cap (-\mathbb{R}^m_{++}) = \emptyset,$$

iff x is a weak optimal Pareto point of H.

Proof. Let us suppose that x is a critical Pareto point of H. Assume by contradiction that x be not a weak optimal Pareto point of H, i.e., that there exists $\tilde{x} \in \mathbb{R}^n$ such that

$$H(\tilde{x}) \prec H(x). \tag{26}$$

As H is pseudo-convex, then (26) implies that

$$JH(x)(\tilde{x} - x) \prec 0.$$

17

But this contradicts the fact of x being a critical Pareto point of H, and the first part is concluded. The second part is a simple consequence of the fact that F is differentiable with the definitions of critical Pareto point and weak optimal Pareto point. For more details, see [6]. $\qquad\square$

Definition 5.3. *A sequence $\{z^k\} \subset M$ is quasi-Fejér convergent to a nonempty set U iff, for all $z \in U$, there exists a sequence $\{\epsilon_k\} \subset \mathbb{R}_+$ such that*

$$\sum_{k=0}^{+\infty} \epsilon_k < +\infty, \qquad \|z^{k+1} - z\|^2 \leq \|z^k - z\|^2 + \epsilon_k, \qquad k = 0, 1, \ldots.$$

In next lemma we recall the theorem known as quasi-Fejér convergence.

Lemma 5.1. *Let $U \subset \mathbb{R}^n$ be a nonempty set and $\{z^k\} \subset \mathbb{R}^n$ a quasi-Fejér convergent sequence. Then, $\{z^k\}$ is bounded. Moreover, if an accumulation point $\bar{z}$ of $\{z^k\}$ belongs to U, then the whole sequence $\{z^k\}$ converges to $\bar{z}$ as k goes to $+\infty$.*

Proof. See Burachik et al. [8]. $\qquad\square$

Consider the following set

$$U := \{x \in \mathbb{R}^n : F(x) \preceq F(x^k), \ \ k = 0, 1, \ldots\}. \tag{27}$$

In general, the above set may be an empty set. To guarantee that U is nonempty, an additional assumption on the sequence $\{x^k\}$ is needed. In the next remark we give such a condition.

Remark 5.4. *If the sequence $\{x^k\}$ has an accumulation point, then U is nonempty. Indeed, let $\bar{x}$ be an accumulation point of the sequence $\{x^k\}$. Then, there exists a subsequence $\{p^{k_j}\}$ of $\{x^k\}$ which converges to $\bar{x}$. Since F is continuous $\{F(x^k)\}$ has $F(\bar{x})$ as an accumulation point. Hence, using $\{F(x^k)\}$ as a decreasing sequence (see* item i *of Theorem 5.1) the usual arguments easily show that the whole sequence $\{F(x^k)\}$ converges to $F(\bar{x})$ and the following relation holds*

$$F(\bar{x}) \preceq F(x^k), \qquad k = 0, 1, \ldots,$$

which implies that $\bar{x} \in U$, i.e., $U \neq \emptyset$.

Assumption 1. Each v^k of the sequence $\{v^k\}$ is a scalarization compatible, i.e., exists a sequence $\{w^k\} \subset \text{conv} S$ such that
$$v^k = -JF(x^k)^t w^k, \qquad k = 0, 1, \ldots.$$

18

As was observed in Section 4, this assumption holds if $v^k = v(x^k)$, i.e., if v^k is the exact steepest descent direction at x^k. We observe that the Assumption 1 also was used in [14] for proving the full convergence of the sequence generated for the Method in the case that F is convex. From now on, we will assume that the Assumption 1 holds true.

In next lemma we present the main result of this section. It is fundamental to the proof of the global convergence result of the sequence $\{x^k\}$.

Lemma 5.2. *Suppose that F is quasi-convex and U, defined in (27), is nonempty. Then, for all $\tilde{x} \in U$, the following inequality is true:*

$$\|x^{k+1} - \tilde{x}\|^2 \le \|x^k - \tilde{x}\|^2 + t_k^2 \|v^k\|^2.$$

Proof. Consider the hinge $\left(\overline{x^k \tilde{x}}, \overline{x^k x^{k+1}}, \alpha\right)$, where $\overline{x^k \tilde{x}}$ is the segment joining x^k to $\tilde{x}$; $\overline{x^k x^{k+1}}$ is the segment joining x^k to x^{k+1} and $\alpha = \angle(\tilde{x} - x^k, v^k)$. By the law of cosines, we have

$$\|x^{k+1} - \tilde{x}\|^2 = \|x^k - \tilde{x}\|^2 + t_k^2 \|v^k\|^2 - 2t_k \|x^k - \tilde{x}\| \|v^k\| \cos \alpha, \qquad k = 0, 1, \ldots.$$

Thus, taking into account that $\cos(\pi - \alpha) = -\cos \alpha$ and $\langle -v^k, \tilde{x} - x^k \rangle = \|v^k\| \|x^k - \tilde{x}\| \cos(\pi - \alpha)$, the above equality becomes

$$\|x^{k+1} - \tilde{x}\|^2 = \|x^k - \tilde{x}\|^2 + t_k^2 \|v^k\|^2 + 2t_k \langle -v^k, \tilde{x} - x^k \rangle, \qquad k = 0, 1, \ldots.$$

On the other hand, from Assumption 1, there exists $w^k \in \text{conv} S$ such that

$$v^k = -JF(x^k)^t w^k, \qquad k = 0, 1, \ldots.$$

Hence, the last equality yields

$$\|x^{k+1} - \tilde{x}\| = \|x^k - \tilde{x}\|^2 + t_k^2 \|v^k\|^2 + 2t_k \langle JF(x^k)^t w^k, \tilde{x} - x^k \rangle, \quad k = 0, 1, \ldots,$$

from which, we obtain

$$\|x^{k+1} - \tilde{x}\|^2 = \|x^k - \tilde{x}\|^2 + t_k^2 \|v^k\|^2 + 2t_k \langle w^k, JF(x^k)(\tilde{x} - x^k) \rangle, \quad k = 0, 1, \ldots. \tag{28}$$

Since F is quasi-convex and $\tilde{x} \in U$, from Proposition 5.1 with $H = F$, $x = x^k$ and $y = \tilde{x}$, we have

$$JF(x^k)(\tilde{x} - x^k) \preceq 0, \qquad k = 0, 1, \ldots.$$

So, because $w^k \in \text{conv} S$, we get

$$\langle w^k, JF(x^k)(\tilde{x} - x^k) \rangle \le 0, \quad k = 0, 1, \ldots. \tag{29}$$

Therefore, by combining (28) with (29), the lemma proceeds. $\qquad \square$

19

Proposition 5.4. *If F is quasi-convex, $\mathbb{R}^n$ has non-negative curvature and U, defined in (27), is a nonempty set, then the sequence $\{x^k\}$ is quasi-Fejér convergent to U.*

Proof. The resulted follows from the item ii of Theorem 5.1 and Lemma 5.2 combined with Definition 5.3. □

Theorem 5.2. *Suppose that F is quasi-convex, and U, as defined in (27), is a nonempty set. Then, the sequence $\{x^k\}$ converges to a critical Pareto point of F.*

Proof. From Proposition 5.4, $\{x^k\}$ is Fejér convergent to U. Thus, Lemma 5.1 guarantees that $\{x^k\}$ is bounded and, hence, has an accumulation point $\bar{x} \in \mathbb{R}^n$. Thus, from Remark 5.4, we conclude that $\bar{x} \in U$ and, hence, that the whole sequence $\{x^k\}$ converges to $\bar{x}$ as k goes to $+\infty$ (see Lemma 5.1). The conclusion of the proof is a consequence of item iii of Theorem 5.1. □

Corollary 5.1. *If F is pseudo-convex, $\mathbb{R}^n$ has non-negative curvature and U, as defined in (27), is a nonempty set, then the sequence $\{x^k\}$ converges to a weak optimal Pareto point of F.*

Proof. Since F is pseudo-convex, and in particular quasi-convex, the corollary is a consequence of the previous theorem and Proposition 5.3. □

6 Variational Rationality: Inexact Proximal Algorithms as Self Regulation Problems

In this section, we consider an endless unsatisfied man, who, instead to renounce, aspires, and partially satisfice, using worthwhile changes.

6.1 Variational rationality

1) The course between unsatisfied needs, aspirations, and satisfaction levels

   Variational rationality (Soubeyran [26, 27, 28, 29]) is a purposive and dynamic approach of behaviors. It modelizes the course pursuit between desired ends and feasible means. It is a theory of the endless unsatisfied man, who, given a lot of unsatisfied needs, both renounces to satisfy some of them and aspires to satisfice some others. Let us summarize some of the main points of this conative approach of behaviors, based on cognition (knowledges), motivation (desires) and affect (feelings and emotions),

i) the agent, focusing his attention on the unsatisfied needs he has chosen to satisfice, considers desired ends. He forms aspirations (distal goals). Setting aspirations is a way to know what he really wants among all his wishes, without considering if they are realistic or not.

ii) then, the agent starts to consider feasible means (defined as the means he must find, build, gather and learn how to use);

iii) given the difficulty to gather such feasible means, the agent chooses to partially satisfice his aspirations;

Then, the agent self regulates all his goal oriented activities:

iv) goal setting, setting proximal goals is a way for him to divide the difficulty, to better know what he can really do. This allows him to balance between "desired enough" ends and "feasible enough" means;

v) goal striving represents the path (way of doing, strategy) the agent chooses to follow and the obstacles he must overcome to attain his successive proximal goals and partially satisfice;

vi) goal pursuit is the revision of his goals, using feedbacks coming from successes and failures.

This variational approach is progressive (adaptive). The step by step joint formation of distal (global) and proximal (local) goals and related actions is a process including a lot of interactions, tatonnements, adjustments,..., driven by inexact perceptions, evaluations and judgments.

Among several variational principles, three of them are worth mentioning in the present paper,

- the "satisficing with not too much sacrificing" principe and the "worthwhile to change principle" (Soubeyran [26, 27])

- the "tension reduction-tension production" principle (Soubeyran [28, 29]).

2) Unsatisfied needs, aspirations, satisfaction and satisfying levels, and aspiration gaps

In the specific case of this paper let us modelize the main motivational concepts of the variational approach of Soubeyran [26, 27]. They include,

a) **The map of unsatisfied needs (the needs system).** An agent has two ways to perceive, judge and estimate a situation, either in term of unsatisfaction or in term of satisfaction. Usually an agent deals with a lot of unsatisfied needs which depend of his present situation $x \in X$. Let $I = \{1, 2, \ldots, m\}$ be the list of different potential needs. The perceived unsatisfied needs functions are $0 \leq n_i(x) < +\infty$ be the strength of each perceived need $i \in I$ for each situation $x \in X$. Let $N : X \to \mathbb{R}^m$, given by

$$N(x) = (n_1(x), n_2(x), \ldots, n_i(x), \ldots, n_m(x)),$$

be the map of unsatisfied needs in this situation. These needs can be rather vague and abstract. Hull [17] and Murray [24] give an extensive list of different needs.

b) **The map of aspiration gaps.** As soon as the agent chooses to do not renounce to satisfy, at least partially, all these unsatisfied needs, they become, in this present situation, aspiration gaps $a_i(x), i \in I$, although in general, the perceived aspiration gaps $a_i(x) \geq 0, i \in I$ are lower than perceived unsatisfied needs: $0 \leq a_i(x) \leq n_i(x), i \in I$ because agents usually aspire to fill no more than their unsatisfied needs.

c) **The map of aspiration levels (desirable ends, or the distal goal system).** Let us denote by $\overline{g}_i(x), i \in I$, the aspiration levels for $x \in X$. They represent still vague, abstract, and non committed higher order goals (visions, ideals, aspirations, fantasies, dreams, wishes, hopes, wants, and desires). These aspirations levels represent desirable (but perhaps irrealistic) ends. Lewin [21] defines aspiration levels as desirable ends, some being irrealistic in a near future, and others not.

d) **The map of satisfaction levels (the experienced utility system).** Most of the time unsatisfied needs are partially satisfied. Let $G : X \to \mathbb{R}^m$, $G(x) = (g_1(x), g_2(x), .., g_i(x), .., g_m(x))$, be the map of present satisfaction levels (or outcomes) in the present situation x, where $g_i(x) \leq \overline{g}_i(x), i \in I$, i.e, the levels at which all needs are partially satisfied.

e) **The map of discrepancies (the drive system).** The differences between aspiration levels and satisfaction levels define more precisely aspiration gaps $a_i(x) = \overline{g}_i(x) - g_i(x) \geq 0, i \in I$ which are non negative. We will assume in this paper that aspirations gaps are equals to unsatisfied needs, i.e, they represent the discrepancies

$$f_i(x) = n_i(x) = a_i(x) = \overline{g}_i(x) - g_i(x) \geq 0, i \in I, \qquad x \in X,$$

because, usually, agents aspire to satisfy their perceived unsatisfied needs, even if, in a second stage, they have not the intention to satisfy all of them. The crude perception of these gaps generates feelings and emotions, the so called drives, Hull [17].

We consider here satisfaction levels $g_i(x), i \in I$, instead of discrepancies. Moreover, for simplification, we consider all aspiration levels as constant, i.e, $\overline{g}_i(x) = \overline{g}_i < +\infty, i \in I, x \in X$. Then,

$$f_i(x) = \overline{g}_i - g_i(x) \geq 0, \quad \text{and} \quad g_x(v) = -f_x(v), \qquad i \in I, \ x \in X.$$

The main problem is to know how the agent sets all these levels, step by step (progressively).

3) Feasible means and the "goal system"

A "goal system" (Kruglanski et al.[20]) comprises i) a cognitive network of mental representations of goals and means which are structurally interconnected by several more or less strong cognitive links, ii) in the short run a subset of limited available resources (physiological, material, mental, social means) because means are scarce and difficult to obtain, iii) an allocation process where goals compete for the use of these limited available resources, iv) a motivational process of goal setting, goal commitment, goal striving and goal pursuit (using affective feedback engendered in response to success and failure outcomes, goal revision including persistence of pursuit, means substitution and the management of goal-conflict). In our specific case the goal system is the satisfaction map

$$X \ni x \longmapsto G(x) \in R^m.$$

Available means are identied to the situation $x \in X$. These means can represent actions, resources and capabilities "to be able to do" them (see Soubeyran [26, 27]). The fact that outcomes compete for restricted means can be modelized as follows: we decompose the given x into the sum $x = x_1 + x_2 + ... + x_i + ... + x_k$ where $x_i \in X$ is the bundle of means allocated to goal i and $g_i = g_i(x_i)$ is the level of satisfaction of this objective.

6.2 The proximal "satisficing-but not too much sacrificing" principle

The local evaluation of marginal satisfaction levels of change. Starting from the situation $x \in X$, let $v \in X$ be a direction of change, $t > 0$ be the intensity of change, $u = y - x = tv$ be the change and $G_x(v) = JG(x)v$ be the vector of marginal satisfaction levels of change. The related differents needs may have different degree of importance and urgency and, each step, the agent must weight each of them to define priorities. This task (solving trade off) is not easy and must be done progressively. Define $g_x : \mathbb{R}^n \to \mathbb{R}$, given by

$$g_x(v) := \min_{i \in I} \langle \nabla g_i(x), v \rangle = \min_{i \in I} (JG(x)v)_i, i \in I, \qquad i \in I,$$

the marginal satisfaction function. It represents the minimum of the different marginal satisfaction levels $(JG(x)v)_i, i \in I$. The consideration of this marginal satisfaction function avoids to choose weights for each marginal satisfaction level and to have to adapt them each step.

Taking care of exploration costs. Let situations like $x \in X$ represents means which generate the vector of satisfaction levels $G(x)$. The agent, in situation $x \in X$, considers (explores) new situations $y = x + tv, t > 0, v \in X$. This global exploration process is costly. Let us define the local consideration costs

(search costs, exploration costs) $c_x(v) = (1/2) \|v\|^2 \geq 0$. The choice of a quadratic function modelizes the case where local exploration is not too coslty: consideration costs $c_x(v) \geq 0$ are large "in the large" and small "in the small". Notice that, while the agent considers feasible directions of change over the whole state space, he takes care of consideration costs (a local aspect).

The local search of directions of aspirations. In general (Soubeyran [26, 27]) the proximal payoff balances desired ends and feasible means. In this paper the proximal payoff $l_x(v) = g_x(v) - c_x(v)$ balances the marginal satisfaction levels and the costs to consider them (local exploration costs). Since

$$l_x(v) \geq 0 \Longleftrightarrow g_x(v) \geq c_x(v),$$

we will say that it is "worthwhile to explore in direction v, starting from x", because the marginal satisfaction level $g_x(v)$ in this direction is higher than the costs $c_x(v)$ to be able to consider them. For each $x \in X$ consider the local search proximal problem: find a direction of change $v(x) \in X$ such that $l_x(v(x)) = \sup\{l_x(v), v \in X\}$. Let $\bar{l}_x = \sup\{l_x(v), v \in X\}$ be the optimal proximal payoff function at x and $v(x) = \arg\max\{l_x(v), v \in X\} \in X$ be the unique optimal direction of change, starting from x. Then, $\bar{l}_x = l_x(v(x))$. From Lemma 4.1 and Lemma 4.2, it follows that

1) If $x \in X$ is Pareto critical, then $v(x) = 0 \in X$ and $\bar{l}_x = 0$.

2) If $x \in X$ is not Pareto critical, then $\bar{l}_x > 0$ and $g_x(v) \geq g_x(v(x)) > c_x(v(x)), i \in I$.

3) The mappings $X \ni x \mapsto v(x) \in X$ and $X \ni x \mapsto \bar{l}_x \in R$ are continuous.

Starting from x and using variational rationality concepts, the optimal direction of change $v(x)$ defines the unique direction of aspiration and the optimal proximal payoff $\bar{l}_x = l_x(v(x))$ defines the proximal aspiration level (net of consideration costs). From Lemma 4.1 and Lemma 4.2, it follows that

1) If $x \in X$ is Pareto critical, then the direction of aspiration and the proximal aspiration level are zero.

2) If $x \in X$ is not Pareto critical, then, their is a strictly positive direction of aspiration and the proximal aspiration level is strictly positive.

3) direction of aspirations and proximal aspiration levels are continuous

The local determination of a local satisficing direction of change: If $x \in X$ is not Pareto critical,

i) set a local (net) satisficing level of change $\widetilde{l}_x = \sigma \bar{l}_x, 0 < \sigma \leq 1$ which is positive and strictly lower than the local aspiration level of change $\bar{l}_x > 0$. As a "variational rationality" concept (Soubeyran [26, 27]), this local satisficing level of change $\widetilde{l}_x$ is situational dependent (it changes with $x \in X$). Simon [25], the father of the satisficing concept, defines an invariant satisficing level without any reference to an aspiration level of change $\bar{l}_x$.

ii) Then, the agent will try to find a direction $v^\S \in X$ such that $l_x(v^\S) \geq \widetilde{l}_x$. This means that the satisficing direction of change $v^\S$ "improves enough" with respect to the aspiration direction of change $v(x)$, including exploration costs. In variational term such a direction not only satisfices (Simon [25]) but even more, it balances satisficing ("improving enough") marginal satisfactions to change $n_x(v)$ with some sacrifices to change $c_x(v)$, because the net satisfaction level $l_x(v^\S)$ is higher than the (net) satisficing level. This is a local version of the variational "sacrificing with not too much sacrificing" principle (Soubeyran [26, 27]). This is equivalent to say that the satisficing direction of change $v^\S \in X$ is an inexact solution of the local search proximal problem. In this context proximal goals are local aspiration levels of change and satisficing levels of change.

Remark 6.1. *In term of variational rationality, the Fliege-Swaiter [11] steepest descent method appears to be an "aspiration driven local search proximal algorithm". In situation $x \in X$, the distal goal is the aspiration level of change $\bar{l}_x$ and the proximal goal is the satisficing level $\widetilde{l}_x$.*

6.3   The proximal "worthwhile to change" principle and goal difficulty

Inertia matters because to be able to change from some situation $x \in X$ to a new improving situation $y \in X$ is costly. As variational concepts, there are two kinds of "costs to be able to change" (Soubeyran [26, 27]): i) consideration costs (perception, exploration, search and evaluation costs . . .), ii) capability costs to change $C(x, y)$, i.e, the costs to be able to change (to be able to stop to use old means, to be able to use again old means, and to be able to imagine, find, build, gather, and learn how to use new means). Means can be capabilities (competences, skills), ingredients and resources. In the present paper consideration costs are $c_x(v) = (1/2) \|v\|^2$ and capability costs to change are $C(x, y) = K\left[tJG(x)(y - x)\right] = tJG(x)(y - x)$, with $t > 0$. This formulation, specific to the present paper, means that costs to change from x to y increase with the difficulty to change, modelized here as the vector of gradients $\Lambda(x, y) = tJG(x)(y - x)$, including the step length of change $t > 0$. Then, the second variational principle tells us that it is "worthwhile to change" from x to y if advantages to change $A(x, y) = G(y) - G(x)$ are higher than some proportion, $\beta > 0$, of costs to change, i.e $G(y) - G(x) \geq \beta C(x, y)$ where $\beta > 0$ is a rate of tolerance which calibrates how the change (transition) $x \curvearrowright y$ is acceptable. More generally (Soubeyran [26, 27]) it is "worthwhile to change" from x to y if motivations to change $M(x, y) = U\left[A(x, y)\right]$ are higher than some proportion, $\beta > 0$, of resistances to change $R(x, y) = \Lambda\left[C(x, y)\right]$ where $U(.)$ and $\Lambda(.)$ are the experienced utility and desutility of advantages and costs to change.

25

The variationel concept of "worthwhile changes" is related to the famous Lindblom [22] "muddling through" economizing principle where agents make small steps (incremental changes, choosing the step size in our context) and successive limited comparisons (balancing pro and cons).

Remark 6.2. *Our paper considers quasi-convex (or quasi-concave) payoffs. In term of variational rationality, this case is very interesting, because it allows large flat portions which can be very costly to explore (quadratic exploration costs are "large in the large"). Hence, in this case, convergence is a very nice result.*

6.4 Local exploration traps

The goal of this paper has been to give conditions of convergence of a path of change towards a Pareto optimum in a multicriteria optimization setting. The variational concept of a behavioral trap (Soubeyran [26, 27]) appears in this context at the local level of the consideration (say exploration) process. More precisely, we will say that $x^* \in X$ is a local exploration trap if

$$\bar{l}_{x^*} = 0 \iff l_{x^*}(v) = g_{x^*}(v) - c_{x^*}(v) \leq 0, \qquad v \in X.$$

This means that, locally, it is not worthwhile to explore, because, whatever the direction of change $v \in X$, marginal advantages to change $n_{x^*}(v)$ are lower than local exploration costs to change $c_{x^*}(v)$. Lemma 4.1 shows that if $x^* \in X$ is Pareto critical, then $x^* \in X$ is a local exploration trap.

7 Final Remarks

We proved full convergence of the sequence generated by this inexact method to a critical Pareto point associated to quasi-convex multicriteria optimization problems. We also show a striking result, i.e, the strong connexion of such an inexact proximal algorithm with the self regulation problem in Psychology. Further researches can be made in this direction.

References

1. Bandura, A. (1977a). Social learning theory. Englewood Cliffs, NJ: Prentice-Hall.

2. Bandura, A. (1977b). Self-efficacy: toward a unifying theory of behavioral change. Psychological Review, 84, 191-215.

3. Bandura, A. (1977c). Self-efficacy: The exercise of control. (1st ed.). New York: W.H. Freeman & Company.

4. Bello Cruz, J. Y. & Lucambio Pérez, L. R. (2010). Convergence of a projected gradient method variant for quasiconvex objectives. Nonlinear Analysis, 9, 2917-2922.

5. Bello Cruz, J.Y., Lucambio Pérez, L.R. & Melo, J.G. (2011). Convergence of the projected gradient method for quasiconvex multiobjective optimization. Nonlinear Anal., 74, 5268-5273.

6. Bento, G. C., Ferreira, O. P. & Oliveira, P. R. (2012). Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds. J. Optim. Theory Appl., 154 (1), 88-107.

7. Bonnel, H., Iusem, A. N. & Svaiter, B. F. (2005). Proximal methods in vector optimization. SIAM J. Optim., 15 (4), 953-970.

8. Burachik, R., Graña Drummond, L. M., Iusem, A. N. & Svaiter, B. F. (1995). Full convergence of the steepest descent method with inexact line searches. Optimization, 32 (2), 137-146.

9. Ceng, L. C. Mordukhovich, B. S. & Yao, J. C. (2010). Hybrid Approximate Proximal Method with Auxiliary Variational Inequality for Vector Optimization. J. Optim .Theory Appl., 146, 267-303.

10. Fliege, J., Graña Drummond, L. M. & Svaiter, B. F. (2009). Newton's method for multiobjective optimization. SIAM J. Optim. 20 (2), 602-626.

11. Fliege, J. & Svaiter, B. F. (2000). Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res., 51 (3), 479-494.

12. Fukuda, E. H. & Graña Drummond, L. M. (2011). Inexact projected gradient method for vector optimization. Available in $http : //www.optimization - online.org/DB_F ILE/2011/06/3062.pdf$, 1-23.

13. Goh, C. J. & Yang, X. Q. (2002). Duality in optimization and variational inequalities, New York: Taylor & Francis.

14. Graña Drummond, L. M. & Svaiter, B. F. (2005). A steepest descent method for vector optimization. J. Comput. Appl. Math. 175 (2), 395-414.

15. Graña Drummond, L. M. & Iusem, A. N. (2004). A projected gradient method for vector optimization problems. Comput. Optim. Appl. 28 (1), 5-29.

16. Hiriart-Urruty, J.B. & Lemaréchal, C. (1993). Convex analysis and minimization algorithms I and II. Berlin: Springer-Verlag.

17. Hull, C. (1935). The conflicting psychologies of learning: a way out, Psychological Review, 42, 491-516.

18. Jordan, M. & Miller, S. (1994). I can't accept not trying: Michael Jordan on the pursuit of excellence. (1st ed.). Harper San Francisco, a division of Harper Collins Publishers.

19. Kiwiel, K. C. & Murty, K. (1996). Convergence of the steepest descent method for minimizing quasiconvex functions. J. Optim. Theory Appl., 89 (1) 221-226.

20. Kruglanski, A., Shah, J., Fishbach, A., Friedman, R., Chun, W. & Sleeth-Keppler, D. (2002). A theory of goal systems, Chapter in M. P. Zanna (Ed.) Advances in Experimental Social Psychology, San Diego: Academic Press. 34, 331-378.

21. Lewin, K. (1951). Field theory in social science. New York: Harper & Row.

22. Lindblom, C. (1959). The science of "muddling through". Public Administration Review, 19, 79-88.

23. Luc, D. T. (1989). Theory of vector optimization. Lecture Notes in Economics and Mathematical Systems, 319, Berlin: Springer.

24. Murray, H. (1938). Explorations in personality. New York: Oxford University Press.

25. Simon, H. (1955). A behavioral model of rational choice. The Quarterly Journal of Economics, 69 (1), 99-118.

26. Soubeyran, A. (2009). Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors. Mimeo, Unpublished results.

27. Soubeyran, A. (2010). Variational rationality and the unsatisfied man: routines and the course pursuit between aspirations, capabilities and beliefs. Mimeo, Unpublished results.

28. Soubeyran, A. (2012a). Worthwhile to change course pursuits and behavioral traps, Unpublished results.

29. Soubeyran, A. (2012b). Variational rationality: a course pursuit between desired ends and feasible means, Unpublished results.

30. Tolli, A. & Schmidt, A. (2008). The role of feedback, causal attributions, and self-efficacy in goal revision. Journal of Applied Psychology, 93 (3), 692-701.

31. Wood, R. & Bandura, A. (1989). Social cognitive theory of organizational management. Academy of Management Review, 10, 361-384.

32. Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective, in: M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of self-regulation, San Diego: Academic Press, 13-39.