A Trust-Region Method for Unconstrained Multiobjective Problems with Applications in Satisficing Processes

Kely D. V. Villacorta[*]
Department of Scientific Computing, CI
Federal University of Paraíba, Brazil
kelydvv@ci.ufpb.br

Paulo Roberto Oliveira[†]
Department of Systems Engineering and Computer Sciences, COPPE
Federal University of Rio de Janeiro, Brazil
poliveir@cos.ufrj.br

Antoine Soubeyran
Aix-Marseille University (Aix-Marseille School of Economics)
CNRS & EHESS, France
antoine.soubeyran@gmail.com

May 19, 2013

Abstract

Multiobjective optimization has a significant number of real life applications. For this reason, in this paper, we consider the problem of finding Pareto critical points for unconstrained multiobjective problems and present a trust-region method to solve it. Under certain assumptions, which are derived in a very natural way from assumptions used by Conn et al. [1] to establish convergence results of the scalar trust-region method, we prove that our trust-region method generates a sequence which converges in the Pareto critical way, this means that our generalized marginal function, which generalizes the norm of the gradient for the multiobjective case, converges to zero. In the last section of this paper, we give an application to satisficing processes in Behavioral Sciences. Multiobjective trust-region methods appear to be remarkable specimens of much more abstract satisficing processes, based on "variational rationality" concepts. One of their important merits is to allow for efficient computations. This is a striking result in Behavioral Sciences.

[*]This work was partially supported by CNPq - Brasil.
[†]This author's research was partially supported by CNPq

1 Introduction.

In this work, we extend the traditional scalar trust-region method to find Pareto critical points for unconstrained multiobjective problem, where each component, of our objective function, is twice continuously differentiable.

Before we begin our study, it is appropriate to give a brief survey and description of the trust-region methods.

The concept of this kind of method has matured over 60 years, but just from 1970 onwards, the subject was taken on actively by the research community, resulting in a proliferation of results and ideas. The last decade has seen considerable progress in this theory for scalar-valued problems, also they enjoy of a good reputation based on its remarkable confidence numbers, as well as a solid and complete theory of convergence, see e.g. Ahookhosh and Amini [2], Bastin et al. [3], Conn et al. [1], Erway and Gill [4], Gardašević Filipović [5], Ji et al. [6], Yu et al. [7] and their references. In trust-region methods are considered a model m_k of the objective function and an "adequate" trust-region, which is a neighborhood of the current iteration. This neighborhood is often represented by a ball in some norm, the radius Δ_k is "updated" from iteration to iteration, in relation to how well m_k approximates to the objective function in the "trial point".

In the scalar case, when there is only one objective, under mild assumptions, one can prove that trust-region algorithms produce a sequence of iterates for which the norm of the gradient, evaluated at this points, converges to zero. This kind of proofs are often called first-order convergence proofs. These proofs are global convergence proofs because they do not assume that there is an iterate which is sufficiently close to a critical point, but only assume that all iterates lie in a region in which the objective function is bounded below. Unfortunately, it is not usually possible to produce global proofs of convergence of the sequence of iterates, without additional assumptions.

One of the main reasons to work in an extension of this method is that although trust-region methods provide an efficient approach to solve scalar non-convex optimization problem, few studies are found in the literature that apply this theory to multiobjective and/or vector optimization, see e.g. Peng et al. [8], Yao et al. [9], Xi and Shi [10].

On the other hand, multiobjective trust-region methods appear to be remarkable specimens of much more abstract satisficing processes, based on "variational rationality" concepts. In behavioral sciences, Simon [11] spends his whole life to convince economists that agents do not always optimize. He first proposes a bounded rationality approach which explains why agents do not optimize, a critic of the substantive optimization approach widely, if not exclusively, used in economics. Simon [11] says that agents cannot optimize because they have limited capabilities to perceive, compare, judge, estimate, and explore new alternatives, and limited stocks of resources, time, money and energy to spend enough efforts to set several conflicting goals, strive these goals and revise them. A bounded rational agent will try to improve, knowing more or less precisely that improving is costly. Then, he must spend costs to know the costs to improve and so on.... To be perfectly rational it will require to enter in an infinite costs regression....This is not an economizing process !

But, if agents do not optimize what do they do? Simon [11] is the father of the satisficing concept which represents a procedural approach of bounded rationality where agents set goals they find satisficing, search until they reach them, and stop there. While researchers globally agree with his view, they consider his formalization as very incomplete. What does mean that a goal is satisficing ("improving enough", but "what is enough?"), how agents set these goals, why they stop as soon they reach them, because a satisficing goal is not free of un-met needs? Using his recent "variational rationality" approach of stability and change theories [12, 13] has been able to provide, as an application, a more complete and dynamic model of satisficing, which gives an answer to all these questions. This approach has generated a lot of applications in Applied Mathematics (variational analysis, optimization) concerning proximal algorithms, local search algorithms, convergence to a Nash equilibrium, dual equilibrium problems, variational inequalities [14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25]. However this variational model, being very general, does not provide computational aspects. This is a weakness. In the last section we remedy to this important point. Because trust-region methods are efficient computational methods to find an optimum, we provide, among very few, a specific multi criteria satisficing model which works at the computational level.

The outline of this paper is as follows. In Section 2, we recall some results to be used throughout the paper. In section 3, we present a trust-region method. In section 4, we establish the assumption of a "sufficient reduction of the model function", and show when it is satisfied. In section 5, it is proved that the error between the scalarization function value and its model approximation is bounded. In section 6, it is proved that the entire sequence generated by the method converges to a Pareto critical point of the problem. In section 7, on applications, we want to show how multiobjective trust-region methods provide a striking and complete specific model of much more abstract satisficing processes. Finally, in the last section we give our conclusions.

2 Preliminaries

Since we are working with multiobjective optimization, we consider the Paretian cone, $\mathbb{R}^p_+$, to induce a partial order $\preceq$ in $\mathbb{R}^p$. In other words, given $y, y' \in \mathbb{R}^p$, $y \preceq y'$ $(y \prec y')$ means $y_i \leq y'_i$ $(y_i < y'_i)$ for $i = 1, \ldots, p$. In this paper, we find Pareto critical points for the following unconstrained multiobjective problem:

$$(MOP) \qquad \min\{F(x) : x \in \mathbb{R}^n\},$$

where $F(x) = (F_1(x) \ldots, F_p(x))^T$, $F_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, p$ and each F_i is twice continuously differentiable. In other words, we present a trust-region method which finds $x^* \in \mathbb{R}^n$ such that

$$\forall d \in \mathbb{R}^n,\ \exists i_0 = i_0(d) \in \{1, \ldots, p\} : \langle \nabla F_{i_0}(x^*), d \rangle \geq 0.$$

Note that in the scalar case, $p = 1$, we recover the first-order condition "gradient equal to zero".

In order to establish our trust-region algorithm, denoted by TRMPA, it is necessary to define some auxiliary functions and review some properties about them. From now on, I will denote the set of indices $\{1, \ldots, p\}$.

We start by defining the marginal function $\omega : \mathbb{R}^n \to \mathbb{R}$, associated to our problem, given by

$$\omega(x) := -\min_{\|d\|\leq 1} \left(\max_{i \in I} \left\{ \langle \nabla F_i(x), d \rangle \right\} \right), \tag{1}$$

Observe that when $p = 1$, the scalar case, then the optimal solution and value are given by $d^*(x) = -\frac{\nabla F(x)}{\|\nabla F(x)\|}$ and $\omega(x) = \|\nabla F(x)\|$, respectively. On the other hand, Fliege and Svaiter [26] show some proprieties of $\omega(x)$ related to the concept of Pareto critical points. In order to remember these properties we denote the set solution of (1) by $\mathscr{D}(x)$.

Lemma 2.1. [26, Lemma 3]

i) $\omega(x) \geq 0$, *for all* $x \in \mathbb{R}^n$*;*

ii) If x is Pareto critical of (MOP), then $0 \in \mathscr{D}(x)$ and $\omega(x) = 0$;

iii) If x is not Pareto critical of (MOP), then $\omega(x) > 0$ and for any $d \in \mathscr{D}(x)$ we have that

$$\langle \nabla F_j(x), d \rangle \leq \max_{i \in I} \{ \langle \nabla F_i(x), d \rangle \} < 0, \quad \forall\, j \in I,$$

i.e., d is a descent direction of (MOP);

iv) The application $x \mapsto \omega(x)$ is continuous;

v) If x^k converges to $\bar{x}$, $d_k \in \mathscr{D}(x^k)$ and d_k converges to $\bar{d}$, then $\bar{d} \in \mathscr{D}(\bar{x})$.

Note that function ω is essential for the validity of our extension because it will play the role of the norm of the gradient, besides it gives us information about Pareto criticality.

Now we consider another scalar function $\phi : \mathbb{R}^n \to \mathbb{R}$, given by

$$\phi(x) := \max_{i \in I} F_i(x).$$

Note that the main reason to consider ϕ is that the following problem

$$\min_{x \in \mathbb{R}^n} \phi(x),$$

is a strict scalar representation of (MOP).

On the other hand, with the goal of dribbling the problem of working with functions not necessarily differentiable neither convex, it is necessary to introduce the Clarke directional derivative and the Clarke subdifferential.

Definition 2.1. *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be locally Lipschitz around the point $x \in \mathbb{R}^n$. Then, the Clarke directional derivative is the function*

$$d \in \mathbb{R}^n \mapsto f^\circ(x,d) = \lim_{t \to 0+} \sup_{y \to x} \frac{f(y+td) - f(x)}{t},$$

and the Clarke subdifferential of f at x is the nonempty compact convex set

$$\partial_\circ f(x) = \{ v \in \mathbb{R}^n : \langle v, d \rangle \leq f^\circ(x,d) \text{ for all } d \in \mathbb{R}^n \}$$

Lemma 2.2. [27, Lemma 2.1] *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitzian function. Then*

i) $f^\circ(x;d) = \max\{\langle \xi, d\rangle : \xi \in \partial_\circ f(x)\}$.

ii) The set-valued mapping $\partial_\circ f(\cdot)$ is locally bounded and upper-semicontinuous.

iii) Let $f(x) = \max\{f_l(x) : l \in \Lambda\}$, where the index set $\Lambda \subset \mathbb{R}$ is a compact topological space (e.g. a finite set in discrete topology). If each f_l is continuously differentiable at x, then

$$\partial_\circ f(x) = \operatorname{conv}\{\nabla f_l(x) : l \in \Lambda(x)\},$$

where $\Lambda(x) = \{l \in \Lambda : f_l(x) = f(x)\}$.

Using this lemma, it is easy to prove that the critical set of the function ϕ is contained in the Pareto critical set of F, i.e., $0 \in \partial_\circ \phi(x^*)$ implies

$$\forall d \in \mathbb{R}^n,\ \exists i_0 = i_0(d) \in I : \langle \nabla F_{i_0}(x^*), d\rangle \geq 0.$$

Remark 2.1. *We must note that we can assume without loss of generality that F_i is lower bounded, for all $i \in I$. If this does not hold, then the objective function $F(x)$ can be replaced by $(\exp(F_1(x)),...,\exp(F_p(x)))$ and instead of (MOP) we can consider the following unconstrained multiobjective problem:*

$$(MOP_1) \qquad \min\left\{\Big(\exp\big(F_1(x)\big),\ldots,\exp\big(F_p(x)\big)\Big)^T : x \in \mathbb{R}^n\right\}.$$

It is not difficult to check that the Pareto critical sets of (MOP) and (MOP_1) are the same.

Remark 2.2. *By the above remark it is immediately clear that*

$$-\infty < \inf\{\phi(x) : x \in \mathbb{R}^n\}.$$

3 A Trust-Region Method

The following method, denoted by "TRMP", is based on the basic algorithm of trust-region to unconstrained non-convex scalar optimization problems, presented by Conn et al. [1].

Before presenting our algorithm, we have to define the following local approximation of ϕ:

$$m(x,H,d) := \max_{i\in I}\{F_i(x) + \langle \nabla F_i(x), d\rangle\} + \frac{1}{2}\langle d, Hd\rangle, \quad \text{with} \quad H \in \mathbb{S}^{n\times n}.$$

Remark 3.1. *Although ϕ and m are not necessarily differentiable neither convex functions, from their definitions and* Lemma 2.2*, we have, given $x \in \mathbb{R}^n$*

$$\partial_\circ \phi(x) = \operatorname{conv}\{\nabla F_i(x) : i \in I(x)\}$$

where $I(x) = \{j \in I : \phi(x) = F_j(x)\}$; *and given* $d \in \mathbb{R}^n$, $x \in \mathbb{R}^n$ *and* $H \in \mathbb{S}^{n\times n}$ *fixed, we get*

$$\partial_\circ^d m(x,H,d) = \mathrm{conv}\{\nabla F_i(x) : i \in J(d)\} + Hd$$

where $J(d) = \{j \in I : m(x,H,d) = F_i(x) + \langle \nabla F_i(x), d\rangle\}$ *and* $\partial_\circ^d m(x,H,d)$ *denote the Clarke subdifferential of* m *with respect to* d.

So, the function m *at* $(x,H,0)$ *is a "good" first-order approximation of* ϕ *at* x, *because we have*

$$m(x,H,0) = \phi(x) \quad \textit{and} \quad \partial_\circ^d m(x,H,0) = \partial_\circ \phi(x).$$

Since in the $k-$th iteration, x^k, Δ_k and H_k are given, we establish that

$$m_k(\cdot) := m(x^k, H_k, \cdot)$$

and the trust-region with respect to $k-$th iteration is denoted by

$$\mathscr{B}_k := \{d \in \mathbb{R}^n : \|d\| \le \Delta_k\}.$$

TRMP Algorithm.

Step 0. *Data:*

$$0 < \eta_1 \le \eta_2 < 1 \quad \textit{and} \quad 0 < \gamma_1 \le \gamma_2 < 1. \tag{2}$$

Step 1. *INITIALIZATION: Choose* x^0, H_0 *and* Δ_0. *Set* $k := 0$.

Step 2. *STEP CALCULATION: Compute a step* $d_k \in \mathscr{B}_k$ *that "sufficiently reduces" the function* m_k.

Step 3. *ACCEPTANCE OF THE TRIAL POINT: Compute* $\phi(x^k + d_k)$ *and define*

$$\rho_k = \frac{\phi(x^k) - \phi(x^k + d_k)}{m_k(0) - m_k(d_k)}. \tag{3}$$

If $\rho_k \ge \eta_1$, *then define* $x^{k+1} = x^k + d_k$; *otherwise define* $x^{k+1} = x^k$.

Step 4. *TRUST-REGION UPDATE: Set*

$$\Delta_{k+1} \in \begin{cases} [\Delta_k, \infty[, & \text{if } \rho_k \ge \eta_2; \\ [\gamma_2\Delta_k, \Delta_k], & \text{if } \eta_1 \le \rho_k < \eta_2; \\ [\gamma_1\Delta_k, \gamma_2\Delta_k], & \text{if } \rho_k < \eta_1; \end{cases} \tag{4}$$

Update H_k *to* H_{k+1}. *Set* $k = k+1$ *and go to* **Step 1**.

4 Pareto Cauchy, Optimal and Approximate local Solutions

A crucial point, in this algorithm, is to determine, in **Step 2**, the step d_k which "sufficiently reduces" the function m_k within the trust-region $\mathscr{B}_k$, because this reduction can or cannot guarantee global convergence. In the following subsections, we derive a formal definition of this property, which is derived, in a natural way, from the idea given in the scalar case, see e.g. Conn et al. [1], Nocedal and Wright [28].

Pareto Cauchy Point Reasoning in the same way as in the scalar trust-region methods, one of the simplest possible strategies for reducing the model within the trust-region is to examine the behavior of the model along the steepest descent belonging to $\mathscr{D}(x)$, the set solution of (1), within the trust-region $\mathscr{B}_k$. In other words, given $d_k^* \in \mathscr{D}(x^k)$, we calculate the scalar $\alpha_k \geq 0$ such that

$$\min_{\alpha \geq 0} \{m_k(\alpha d_k^*) : \alpha d_k^* \in \mathscr{B}_k\}, \tag{5}$$

and define the Pareto Cauchy point as $d_k^C := \alpha_k d_k^*$.

Lemma 4.1. *The Pareto Cauchy Point d_k^C satisfies*

$$m_k(0) - m_k(d_k^C) \geq \frac{1}{2}\omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta_k}, \Delta_k\right\},$$

where $\omega(x)$ is defined by (1) *and*

$$\beta_k = 1 + \|H_k\|. \tag{6}$$

Proof. We first note that, $d_k^* \in \mathscr{D}(x^k)$ implies that $\|d_k^*\| \leq 1$, which in turn implies that, for all $\alpha \in [0, \Delta_k]$, $\alpha d_k^* \in \mathscr{B}_k$. Thus the problem (5) is equivalent to

$$\max_{0 \leq \alpha \leq \Delta_k} \{m_k(0) - m_k(\alpha d_k^*)\}. \tag{7}$$

On the other hand, for all $\alpha \geq 0$, from m_k at αd_k^* it follows that

$$m_k(0) - m_k(\alpha d_k^*) \quad \geq \quad -\alpha \max_{i \in I}\{\langle \nabla F_i(x^k), d_k^* \rangle\} - \frac{1}{2}\alpha^2 \langle d_k^*, H_k d_k^* \rangle$$

Thus, from the definition of ω at x^k, i.e.,

$$\omega(x^k) = -\max_{i \in I}\{\langle \nabla F_i(x^k), d_k^* \rangle\}, \tag{8}$$

the Cauchy-Schwarz inequality, $\|d_k^*\| \leq 1$ and (6),

$$m_k(0) - m_k(\alpha d_k^*) \quad \geq \quad \alpha \omega(x^k) - \frac{1}{2}\alpha^2 \beta_k.$$

Then, from (7) and the above inequality, we get

$$\max_{0 \leq \alpha \leq \Delta_k} \{m_k(0) - m_k(\alpha d_k^*)\} \quad \geq \quad \max_{0 \leq \alpha \leq \Delta_k} \left\{\alpha \omega(x^k) - \frac{1}{2}\alpha^2 \beta_k\right\}.$$

In others words,

$$m_k(0) - m_k(d_k^C) \quad \geq \quad \max_{0 \leq \alpha \leq \Delta_k} \left\{\alpha \omega(x^k) - \frac{1}{2}\alpha^2 \beta_k\right\}.$$

Now, we will analyze the right side of this inequality. Since the function $\alpha \mapsto \alpha\omega(x^k) - \frac{1}{2}\alpha^2\beta_k$ in concave and non-negative, it needs to analyze two possibilities for the maximum above. Before, we denote the unconstrained maximizer of this function by α^*, so

$$\alpha^* = \frac{\omega(x^k)}{\beta_k} \geq 0$$

with maximum value equal to

$$\frac{1}{2}\frac{\omega(x^k)^2}{\beta_k} \geq 0.$$

a) If $0 \leq \alpha^* < \Delta_k$, then

$$\max_{0\leq\alpha\leq\Delta_k} \left\{ \alpha\omega(x^k) - \frac{1}{2}\alpha^2\beta_k \right\} = \frac{1}{2}\frac{\omega(x^k)^2}{\beta_k}.$$

b) If $\alpha^* \geq \Delta_k$, then

$$\max_{0\leq\alpha\leq\Delta_k} \left\{ \alpha\omega(x^k) - \frac{1}{2}\alpha^2\beta_k \right\} = \Delta_k\omega(x^k) - \frac{1}{2}\Delta_k^2\beta_k.$$

But, $\alpha^* \geq \Delta_k$ also implies that

$$\Delta_k\omega(x^k) - \frac{1}{2}\Delta_k^2\beta_k \geq \frac{1}{2}\Delta_k\omega(x^k).$$

Consequently

$$\begin{aligned} m_k(0) - m_k(d_k^C) &\geq \min\left\{ \frac{1}{2}\frac{\omega(x^k)^2}{\beta_k}, \frac{1}{2}\Delta_k\omega(x^k) \right\} \\ &= \frac{1}{2}\omega(x^k)\min\left\{ \frac{\omega(x^k)}{\beta_k}, \Delta_k \right\}. \end{aligned}$$

□

For classical optimization, we have $p = 1$. In this case, the natural choice is the steepest descent direction $d_k^* = \frac{\nabla F(x^k)}{\|\nabla F(x^k)\|}$, and we have thus recovered an elementary result in trust-region theory, see, e.g., [1, Theorem 6.3.1] and [28, Lemma 4.5].

Corollary 4.1. *Consider the scalar problem*

$$\min_{d\in\mathscr{B}_k} m_k(d). \tag{9}$$

Then every optimal solution d_k of this problem verifies the following inequality

$$m_k(0) - m_k(d_k) \geq \frac{1}{2}\omega(x^k)\min\left\{ \frac{\omega(x^k)}{\beta_k}, \Delta_k \right\}.$$

Proof. As d_k is a solution of the problem (9), it follows that

$$m_k(0) - m_k(d_k) \geq m_k(0) - m_k(d), \quad \forall d \in \mathscr{B}_k.$$

So, the result follows from the fact that $d_k^C \in \mathscr{B}_k$. □

Remark 4.1. *The result of this corollary was first established in* Xi and Shi [10], *but the minimum in* (1) *is considered on $\mathscr{B}_k$ and not on unit ball, as our case.*

Backtracking to get approximate solutions Although Lemma 4.1 and Corollary 4.1 are sufficient for establishing convergence results of the sequence generated by TRMPA. To obtain exact minimizers of the problems (5) and (9), respectively, could be a difficult task. Furthermore, in practice, it is not desirable to have a high computational cost in obtaining the new iterate, inexact algorithms must be considered.

In these conditions, it is interesting to note that following the idea of Armijo's linesearch, [see 29], it is possible to consider a *backtracking* technique on some direction $d \in \mathscr{D}(x^k)$ (since all these directions are descent directions, if x^k is not Pareto critical), in order to obtain a length step which will provide a "good" decrease of the value function m_k. More precisely, determine the smallest non-negative integer $j = j^*$ such that the direction

$$d_k^*(j) = (\kappa_{\text{bck}})^j \Delta_k d_k^* \tag{10}$$

satisfies the following condition

$$m_k(d_k^*(j)) \leq m_k(0) - \kappa_{\text{lsi}}(\kappa_{\text{bck}})^j \Delta_k \omega(x^k), \tag{11}$$

where d_k^* verifies (8), $\kappa_{\text{lsi}} \in]0, \frac{1}{2}[$ and $\kappa_{\text{bck}} \in]0,1[$ are constants. As it is expected in the scalar case, $p = 1$, inequality (11) is reduced to the Armijo's linesearch, in the direction $-\nabla F(x^k)$. Then, defining this direction by

$$d_k^{\text{apx}} := d_k^*(j^*),$$

we establish the following result.

Lemma 4.2. *The approximate direction d_k^{apx} is well defined in the sense that j^* is finite. Moreover,*

$$m_k(0) - m_k(d_k^{\text{apx}}) \geq \kappa_{\text{dpa}} \omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta_k}, \Delta_k\right\}, \tag{12}$$

where $\kappa_{\text{dpa}} \in]0,1[$ is a constant independent of k.

Proof. We first consider the case where condition (11) is violated for some j, i.e.,

$$m_k(d_k^*(j)) > m_k(0) - \kappa_{\text{lsi}}(\kappa_{\text{bck}})^j \Delta_k \omega(x^k). \tag{13}$$

Now, from definition of m_k, we see that

$$\begin{aligned} m_k(d_k^*(j)) &= \max_{i \in I}\{F_i(x^k) + \langle \nabla F_i(x^k), d_k^*(j)\rangle\} + \frac{1}{2}\langle d_k^*(j), H_k d_k^*(j)\rangle \\ &\leq m_k(0) + (\kappa_{\text{bck}})^j \Delta_k \max_{i \in I}\{\langle \nabla F_i(x^k), d_k^*\rangle\} + \frac{1}{2}((\kappa_{\text{bck}})^j \Delta_k)^2 \langle d_k^*, H_k d_k^*\rangle \\ &= m_k(0) - (\kappa_{\text{bck}})^j \Delta_k \omega(x^k) + \frac{1}{2}((\kappa_{\text{bck}})^j \Delta_k)^2 \langle d_k^*, H_k d_k^*\rangle. \end{aligned} \tag{14}$$

As $\|H_k\| \leq \beta_k$ and $\|d_k^*\| \leq 1$, we obtain from Cauchy-Schwarz's inequality that

$$\langle d_k^*, H_k d_k^*\rangle \leq \beta_k. \tag{15}$$

Now, combining (15), (14) and (13), we get

$$(\kappa_{\text{bck}})^j > \frac{2(1-\kappa_{\text{lsi}})\omega(x^k)}{\Delta_k \beta_k}.$$

However, $\kappa_{\text{bck}} < 1$. Therefore there must be some j^* finite such that

$$(\kappa_{\text{bck}})^{j^*} < \frac{2(1-\kappa_{\text{lsi}})\omega(x^k)}{\Delta_k \beta_k}, \tag{16}$$

for which (11) is verified. Then, the approximate direction d_k^{apx} is well-defined, and we get from (10) and (11)

$$m_k(0) - m_k(d_k^{\text{apx}}) \geq \kappa_{\text{lsi}}(\kappa_{\text{bck}})^{j^*} \Delta_k \omega(x^k). \tag{17}$$

As $j^* \geq 1$ and since it is the smallest j which guarantees (16), we may deduce that

$$(\kappa_{\text{bck}})^{j^*} = \kappa_{\text{bck}}(\kappa_{\text{bck}})^{j^*-1} \geq 2\kappa_{\text{bck}}(1-\kappa_{\text{lsi}})\frac{\omega(x^k)}{\Delta_k \beta_k},$$

which together with (17) imply that

$$m_k(0) - m_k(d_k^{\text{apx}}) \geq 2\kappa_{\text{bck}}\kappa_{\text{lsi}}(1-\kappa_{\text{lsi}})\frac{\omega(x^k)^2}{\beta_k}.$$

On other hand, if $j^* = 0$ then (17) reduces to

$$m_k(0) - m_k(d_k^{\text{apx}}) \geq \kappa_{\text{lsi}} \Delta_k \omega(x^k). \tag{18}$$

Combining (17) and (18), it follows that (12) is verified with

$$\kappa_{\text{dpa}} = \min\{\kappa_{\text{lsi}}, 2\kappa_{\text{lsi}}\kappa_{\text{bck}}(1-\kappa_{\text{lsi}})\} < 1.$$

□

Then, in light of Lemma 4.1, Corollary 4.1 and Lemma 4.2, it is acceptable to require that "sufficiently reduction"of the function m_k is given by the following assumptions.

Assumption 1. *For all k,*

$$m_k(0) - m_k(d_k) \geq \kappa_{\text{dasm}}\omega(x^k)\min\left\{\frac{\omega(x^k)}{\beta_k}, \Delta_k\right\},$$

for some constant $\kappa_{\text{dasm}} \in]0,1[$, where $\omega(x^k)$ and β_k are defined by (1) *and* (6)*, respectively.*

Remark 4.2. *As noted earlier, in the scalar case we have $\omega(x^k) = \|\nabla F(x^k)\|$, and thereby recover the Hypothesis A.A.1 required by* Conn et al. in [1].

5 How good is the model Function?

We are now able to prove that the error between the function value ϕ and its approximation m_k at x^k is bounded. First, we need some assumptions to establish the next result as well as the convergence results of TRMPA. In this way, we make the following assumptions:

Assumption 2. *There exists a positive constant $\kappa_{\text{lb}F}$ such that, for all $x \in \mathbb{R}^n$,*

$$\|\nabla^2 F_i(x)\| \leq \kappa_{\text{lb}F}, \quad \textit{for all} \quad i \in I.$$

Assumption 3. *The matrix H_k is uniformly bounded, that is, there exists a constant $\kappa_{\text{ufh}m} \geq 1$ such that:*

$$\|H_k\| \leq \kappa_{\text{ufh}m} - 1, \quad \textit{for all } k.$$

Note that these assumptions are a natural way of extending the assumptions used in Conn et al. [1] to establish convergence results of the scalar trust-region method. Although, in the scalar case, there are works with weaker assumptions [see 30, 2, and their respective references], there are also recent works considering the same assumptions as [1], see [3]. Moreover, observe that our convergence analysis extends naturally to the case $p = 1$ in [1].

Remark 5.1. *Note that assumption 2 implies that ∇F_i is Lipschitz continuous, for each $i \in I$, which in turn implies the uniform continuity of ω.*

Proposition 5.1. *Suppose that* Assumptions 2 and 3 *are verified. Then*

$$|\phi(x^k + d_k) - m_k(d_k)| \leq \kappa_{\text{lsh}} \Delta_k^2. \tag{19}$$

where

$$\kappa_{\text{lsh}} := \max\{\kappa_{\text{lb}F}, \kappa_{\text{sah}m}\}$$

Proof. F_i is twice continuously differentiable, thereby by the mean value theorem, for each $i \in I$, there exists $\alpha_i \in [0,1]$:

$$F_i(x^k + d_k) = F_i(x^k) + \langle \nabla F_i(x^k), d_k \rangle + \frac{1}{2}\langle d_k, \nabla^2 F_i(\xi_k^i) d_k \rangle, \tag{20}$$

where $\xi_k^i = x^k + \alpha_i d_k$. Consequently,

$$\max_{i \in I}\{F_i(x^k + d_k)\} \quad \leq \quad \max_{i \in I}\left\{F_i(x^k) + \langle \nabla F_i(x^k), d_k \rangle\right\} + \max_{i \in I}\left\{\frac{1}{2}\langle d_k, \nabla^2 F_i(\xi_k^i) d_k \rangle\right\}.$$

Now, adding and subtracting $\frac{1}{2}\langle d_k, H_k d_k \rangle$ in the right-side of above inequality, it gets

$$\phi(x^k + d_k) - m_k(d_k) \quad \leq \quad \max_{i \in I}\left\{\frac{1}{2}\langle d_k, (\nabla^2 F_i(\xi_k^i) - H_k) d_k \rangle\right\}. \tag{21}$$

[1] "lb$F$" stands for "lower bound on the objective function".
[2] "ufh$m$" stands for "upper bound on the objective function's Hessian".

On the other hand, (20) can also be written as

$$F_i(x^k+d_k) - \frac{1}{2}\langle d_k, \nabla^2 F_i(\xi_k^i) d_k\rangle = F_i(x^k) + \langle \nabla F_i(x^k), d_k\rangle$$

for each $i \in I$. Then

$$\phi(x^k+d_k) \quad \geq \quad \max_{i\in I}\left\{F_i(x^k) + \langle \nabla F_i(x^k), d_k\rangle\right\} - \max_{i\in I}\left\{-\frac{1}{2}\langle d_k, \nabla^2 F_i(\xi_k^i) d_k\rangle\right\},$$

or equivalently,

$$\phi(x^k+d_k) \quad \geq \quad \max_{i\in I}\left\{F_i(x^k) + \langle \nabla F_i(x^k), d_k\rangle\right\} + \min_{i\in I}\left\{\frac{1}{2}\langle d_k, \nabla^2 F_i(\xi_k^i) d_k\rangle\right\}.$$

Again, adding and subtracting $\frac{1}{2}\langle d_k, H_k d_k\rangle$ in the right-side of this inequality imply that

$$\phi(x^k+d_k) - m_k(d_k) \geq \min_{i\in I}\left\{\frac{1}{2}\langle d_k, (\nabla^2 F_i(\xi_k^i) - H_k) d_k\rangle\right\}. \tag{22}$$

So, combining (21), (22) and using the fact $\|d_k\| \leq \Delta_k$ for all $k \in \mathbb{N}$, we have

$$\begin{aligned}
\left|\phi(x^k+d_k) - m_k(d_k)\right| \quad &\leq \quad \max\left\{\begin{array}{c} \left|\min_{i\in I}\left\{\frac{1}{2}\langle d_k, (\nabla^2 F_i(\xi_k^i) - H_k) d_k\rangle\right\}\right|, \\ \\ \left|\max_{i\in I}\left\{\frac{1}{2}\langle d_k, (\nabla^2 F_i(\xi_k^i) - H_k) d_k\rangle\right\}\right| \end{array}\right\} \\
&\leq \quad \max_{i\in I}\left|\frac{1}{2}\langle d_k, (\nabla^2 F_i(\xi_k^i) - H_k) d_k\rangle\right| \\
&\leq \quad \frac{1}{2}\left(\max_{i\in I}\|\nabla^2 F_i(\xi_k^i)\| + \|H_k\|\right)\Delta_k^2.
\end{aligned}$$

Therefore, the assumptions 2 and 3 imply that

$$|\phi(x^k+d_k) - m_k(d_k)| \quad \leq \quad \kappa_{\mathrm{lsh}}\Delta_k^2.$$

where

$$\kappa_{\mathrm{lsh}} := \max\{\kappa_{\mathrm{lb}F}, \kappa_{\mathrm{ufh}m}\}.$$

□

6 Convergence to Pareto Critical Points

We will show that the TRMP algorithm globally converges to a Pareto critical point, in the sense that $\omega(x^k)$ converges to zero. In order to establish these results, we define the set of indices of the *successful iterations* $\mathscr{S}$ denoted and defined by

$$\mathscr{S} = \{k \in \mathbb{N} : \rho_k \geq \eta_1\}.$$

Similarly, the set of indices of *very successful* iterations $\mathscr{V}$ is denoted and defined by

$$\mathscr{V} = \{k \in \mathbb{N} : \rho_k \geq \eta_2\}.$$

Proposition 6.1. *Suppose that* Assumptions 1, 2 and 3 *hold. If x^k is not Pareto critical and*

$$\Delta_k \leq \frac{\kappa_{\mathrm{das}m}(1-\eta_2)\omega(x^k)}{\kappa_{\mathrm{lsh}}}, \tag{23}$$

then the iteration k belongs to $\mathscr{V}$ and

$$\Delta_{k+1} \geq \Delta_k. \tag{24}$$

Proof. Observe first that condition $\eta_2,\ \kappa_{\mathrm{dpa}} \in]0,1[$ implies that

$$\kappa_{\mathrm{das}m}(1-\eta_2) < 1.$$

Thus condition (23) and $\beta_k = 1 + \|H_k\|$ imply that

$$\Delta_k < \frac{\omega(x^k)}{\beta_k}. \tag{25}$$

As a consequence, Assumption 1 implies directly that

$$m_k(0) - m_k(d_k) \geq \kappa_{\mathrm{das}m}\omega(x^k)\min\left\{\frac{\omega(x^k)}{\beta_k}, \Delta_k\right\} = \kappa_{\mathrm{das}m}\omega(x^k)\Delta_k.$$

In these conditions, from (19), above inequality and (25) we deduce that

$$|\rho_k - 1| = \left|\frac{\phi(x^k+d_k) - m_k(d_k)}{m_k(0) - m_k(d_k)}\right| \leq \frac{\kappa_{\mathrm{lsh}}}{\kappa_{\mathrm{das}m}\omega(x^k)}\Delta_k \leq 1 - \eta_2.$$

Therefore, $\rho_k \geq \eta_2$, shows that an iteration is very successful. Furthermore, the updating of Δ_k, (4), ensures that (24) holds. □

We may now prove that the trust-region radius cannot become too small as long as a *Pareto critical point* is not approached. This property is obtained as a consequence of the above proposition. The technique of the proof is similar to [1, Theorem 6.4.3]. It also indicates how small the trust-region radius has to be relative to $\omega(x^k)$ in order to guarantee the success of iteration.

Proposition 6.2. *Suppose that* Assumptions 1, 2 and 3 *hold. Suppose furthermore that there exists a constant $\kappa_{\mathrm{li}\omega} > 0$ such that $\omega(x^k) \geq \kappa_{\mathrm{li}\omega}$ for all k. Then there is a constant $\kappa_{\mathrm{li}\Delta} > 0$ such that*

$$\Delta_k \geq \kappa_{\mathrm{li}\Delta}, \quad \textit{for all} \quad k.$$

Proof. Assume, for the purpose of deriving a contradiction, that for each $\kappa > 0$ there exists $\hat{k}$ such that

$$\Delta_{\hat{k}} < \kappa.$$

In particular, we consider

$$\kappa = \frac{\gamma_1 \kappa_{\mathrm{das}m}\kappa_{\mathrm{li}\omega}(1-\eta_2)}{\kappa_{\mathrm{lsh}}},$$

and let $\hat{k}$ be the first iteration such that

$$\Delta_{\hat{k}} < \frac{\gamma_1 \kappa_{\text{das}m} \kappa_{\text{li}\omega}(1-\eta_2)}{\kappa_{\text{lsh}}}. \tag{26}$$

On the other hand, by updating the trust-region, (4), we get

$$\gamma_1 \Delta_{\hat{k}-1} \quad \leq \quad \Delta_{\hat{k}}.$$

So, these two inequalities imply

$$\Delta_{\hat{k}-1} < \frac{\kappa_{\text{das}m} \kappa_{\text{li}\omega}(1-\eta_2)}{\kappa_{\text{lsh}}}. \tag{27}$$

Thus the assumption on $\omega(x^k)$ and (27) imply that (23) is verified. Then iteration $\hat{k}-1$ is *very successful* and

$$\Delta_{\hat{k}-1} \leq \Delta_{\hat{k}}.$$

But this contradicts the fact that iteration $\hat{k}$ is the first such that (26) holds, and our assumption is therefore impossible. □

Using the preceding two results and if there are only finitely many successful iterations, we can stablish the Pareto criticality of the unique accumulation point.

Proposition 6.3. *Suppose that* Assumptions 1, 2 and 3 *hold. Furthermore, suppose that, given* x^0*, there are only finitely many successful iterations. Then* $x^k = x^*$ *for all sufficiently large* k *and* x^* *is a Pareto critical point.*

Proof. From definition of TRMPA we ensure that $x^* = x^{k_0+1} = x^{k_0+j}$ for all $j \in \mathbb{N}$, where k_0 is the index of the last successful iterate. Moreover, since all iterations are unsuccessful for sufficiently large k, (2) – (4) imply that the sequence Δ_k converges to zero, as k go to $+\infty$. Therefore, if x^{k_0+1} is not Pareto critical point, Proposition 6.1 implies that there must be a successful iteration whose index is larger than k_0, which is impossible. Hence, $x^{k_0+1} = x^*$ is a Pareto critical point. □

Proposition 6.4. *Suppose that* Assumptions 1, 2 and 3 *hold. Then*

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

Proof. Suppose, for the purpose of establishing a contradiction, that for all k,

$$\omega(x^k) \geq \varepsilon$$

for some $\varepsilon > 0$. By Proposition 6.2 there exists $\kappa_{\text{li}\Delta} > 0$ such that

$$\Delta_k \geq \kappa_{\text{li}\Delta}, \quad \forall k.$$

Now consider a successful iteration with index k. The fact that $k \in \mathscr{S}$, together with Assumption 1, give

$$\beta_k \leq \kappa_{\text{sah}m}$$

and above inequality implies

$$\phi(x^k) - \phi(x^{k+1}) \geq \eta_1(m_k(0) - m_k(d_k)) \geq \kappa_{\text{dasm}}\varepsilon\eta_1 \min\left\{\frac{\varepsilon}{\kappa_{\text{sah}m}}, \kappa_{\text{li}\Delta}\right\}.$$

Summing now over all successful iterations from 0 to k we obtain

$$\phi(x^0) - \phi(x^{k+1}) = \sum_{\substack{j=0 \\ k\in\mathscr{S}}}^{k} (\phi(x^j) - \phi(x^{j+1})) \geq \sigma_k \kappa_{\text{dasm}}\varepsilon\eta_1 \min\left\{\frac{\varepsilon}{\kappa_{\text{sah}m}}, \kappa_{\text{li}\Delta}\right\},$$

where σ_k is the number of successful iterations up to iteration k. But since there are infinitely many such iterations, we have that

$$\lim_{k\to\infty} \sigma_k = +\infty,$$

and the difference between $\phi(x^0)$ and $\phi(x^{k+1})$ is unbounded, which clearly contradicts the fact that $\phi(x)$ is bounded below, commented on Remark 2.2. Therefore the result follows. □

In the scalar case, where $p = 1$, this proof and the following ones are known as global convergence proofs, because it is not necessary to assume that some x^k is sufficiently close to a critical point.

Remark 6.1. *We must note that, if the sequence of iterates is bounded, then Proposition 6.4 implies that there is at least one accumulation point, of this sequence, such that it is a Pareto critical point. Now, we prove the strongest result.*

Theorem 6.1. *Suppose that* Assumptions 1, 2 and 3 *hold. Then*

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

Proof. Suppose, for the purpose of deriving a contradiction, that there exists a subsequence of successful iterates, with index $\{t_j\} \subseteq \mathscr{S}$ such that

$$\omega(x^{t_j}) \geq 2\varepsilon > 0 \tag{28}$$

for some $\varepsilon > 0$ and for all j. Then Proposition 6.4 ensures, for each t_j, the existence of a first successful iteration $l(t_j) > t_j$ such that $\omega(x^{l(t_j)}) < \varepsilon$. Denoting $l_j = l(t_j)$, we thus obtain that there is another subsequence of $\mathscr{S}$ indexed by $\{l_j\}$ such that

$$\omega(x^k) \geq \varepsilon \quad \text{for} \quad t_j \leq k < l_j \quad \text{and} \quad \omega(x^{l_j}) < \varepsilon. \tag{29}$$

Now we analyze the subsequence of iterations whose indices are in

$$\mathscr{K} := \{k \in \mathscr{S} : t_j \leq k < l_j\}$$

where t_j and l_j belong to the two subsequences defined above.

The Assumption 1, the fact of $\mathcal{K} \subseteq \mathcal{S}$ and (29) imply that, for $k \in \mathcal{K}$,

$$\phi(x^k) - \phi(x^{k+1}) \geq \eta_1(m_k(0) - m_k(d_k)) \geq \kappa_{\text{dasm}}\varepsilon\eta_1 \min\left\{\frac{\varepsilon}{\kappa_{\text{sah}m}}, \Delta_k\right\}. \tag{30}$$

But, the sequence $\{\phi(x^k)\}$ is monotonically decreasing and bounded below, by Remark 2.2, thereby convergent. Therefore, the left-side of (30) must converge to zero, as k go to $+\infty$. So, we get

$$\lim_{\substack{k\to\infty \\ k\in\mathcal{K}}} \Delta_k = 0.$$

As a consequence, from minimum of (30) we obtain that, for $k \in \mathcal{K}$ sufficiently large,

$$\Delta_k \leq \frac{1}{\kappa_{\text{dasm}}\varepsilon\eta_1}\left(\phi(x^k) - \phi(x^{k+1})\right).$$

Then we deduce from this bound that, for j sufficiently large,

$$\|x^{t_j} - x^{l_j}\| \leq \sum_{\substack{i=t_j \\ i\in\mathcal{K}}}^{l_j-1} \|x^i - x^{i+1}\| \leq \sum_{\substack{i=t_j \\ i\in\mathcal{K}}}^{l_j-1} \Delta_i \leq \frac{1}{\kappa_{\text{dasm}}\varepsilon\eta_1}\left(\phi(x^{t_j}) - \phi(x^{l_j})\right).$$

Again, from lower bounded of ϕ and the monotonicity of sequence $\{\phi(x^k)\}$ the right-side of inequality converges to zero, and therefore we get that

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

Thereby, by uniform continuity of ω, Remark 5.1, we thus deduce that

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

However, this is impossible because of the definitions of sequences $\{t_j\}$ and $\{l_j\}$. It follows that

$$|\omega(x^{t_j}) - \omega(x^{l_j})| \geq \varepsilon.$$

Hence, no subsequence satisfying (28) can exists, and the theorem is proved. □

Remark 6.2. *This theorem implies that if the sequence $\{x^k\}$ has accumulation points, then all of them are Pareto critical points. The proof of this affirmation follows directly by contradiction, i.e., suppose that there is an accumulation point such that it is not a Pareto critical point, and use Lemma 2.1 (iii). Unfortunately, it is not usually possible to produce global proofs of convergence for the sequence of iterates, without additional assumptions.*

Remark 6.3. *In this work we do not consider the updating of the matrix H_k. In the scalar case it is performed by an approximation of the Hessian matrix $\nabla^2 F(x^k)$, for example using the technique of Quasi-Newton BFGS update. For more details see* [1]. *Yigui and Qian in* [27] *also had a way to get an update of H_k which is symmetric and positive definite. However, it depends on both the exact solution and the Lagrange multiplier associated with it for problem* (9), *which is not a realistic assumption from the point of view of implementation.*

7 Multiobjective Trust-Region Methods as Computationable Satisficing Processes

In this last section we give an application to satisficing processes in Behavioral Sciences, using the general variational approach of stability and change theories [12, 13] as a benchmark. However this qualitative approach has been silent on computational aspects (a more quantitative view). Multiobjective trust-region methods (TRM) remedy to this weakness. We will show that they are a specific but very important case which allows efficient computations. For this purpose we will offer a parallel between variational concepts and key ingredients of the trust-region methods.

1) The static satisficing process Suppose that an agent has a choice problem to solve, defined by a subset of available and known alternatives $F \subset X$ and a scalar utility function $g(\cdot) : x \in X \longrightarrow g(x) \in \mathbb{R}$. The agent does not know the utility or satisfaction level $g(x)$ of each alternative. To know each of them, he will have to pick it and estimate its utility, spending some effort to do that. If he optimizes he will try to find some alternative $x^* \in F$ such that $g(x^*) \geq g(x)$, for all $x \in F$. The process used to find such an optimal alternative remains unclear

The static model of satisficing [11] is the following. If the subset of available alternatives is too big, he faces too much choices (choice overload). Then, Simon [11] proposes that he will set a given utility level $\widetilde{g} \in \mathbb{R}$ (but how?) that he considers as satisficing (but why?) and tries to find (but how?) some alternative $\widetilde{x} \in F$ such that $g(\widetilde{x}) \geq \widetilde{g}$. which he will consider as satisficing. But why? This is because he can feel some residual frustration to do not have improved more, i.e., to have set a too low satisficing level $\widetilde{g}$). In order to satisfice, Simon [11] supposes that the agent uses a "sequential rejection-acceptance" principle. This is a two periods model with search and exploitation. First he picks alternatives in a given order (a sequential search process), rejects all alternatives as long as they do not satisfice and accepts the first alternative which satisfices. Then, he exploits his choice. Notice carefully that Simon [11] and his followers wrongly choose the term "aspiration level" instead of the right term "satisficing level" for $\widetilde{g}$. This has introduced a lot of confusion in this literature, because, in Psychology [31, 32] an aspiration level is defined as a distal goal $\widehat{g}$ (dream, vision, wish) which is very different from a satisficing level $\widetilde{g} \leq \widehat{g}$ which is a proximal goal.

2) From satisfaction levels to unmet needs Now we come back to our present paper. Consider a set X of available alternatives $x \in X = \Re^n$ and an agent, say a consumer, who considers p criteria $i \in I$ to evaluate the satisfaction provided by choosing and, then, consuming the bundle of goods $x = (x_1, x_2, \ldots, x_n) \in X$. Let $G_i(x) \in \mathbb{R}$ be his satisfaction level relative to criteria $i \in I$ and $G(x) = (G_1(x), G_2(x), \ldots, G_p(x))$ be his vector of satisficing levels relative to alternative x. His multicriteria satisfaction function is $G(\cdot) : x \in X \longrightarrow G(x) \in \mathbb{R}^p$. Let us suppose that, for each criteria, his satisfaction level $G_i(\cdot)$ is bounded above. Then, relative to each criteria i, $\overline{G}_i = \sup\{G_i(x) : x \in X\} < +\infty$ represents a global aspiration level. The vector of aspirations levels will be $\overline{G} = (\overline{G}_1, \overline{G}_2, \ldots, \overline{G}_p)$. In this case $F(x) = \overline{G} - G(x) \geq 0$ models the unmet needs the agent tries to minimize, as it is the case in our paper. If, in a repeated choice problem, the consumer passes from alternative x to alternative y, his variation is $y-x=d$ and if, in period k, he passes from x^k to x^{k+1}, his variation will be $x^{k+1}-x^k=d_k$.

Let $\psi(x)=\min\{G_i(x):i\in I\}$ be the lowest satisfaction level and

$$\phi(x)=\max\{F_i(x):i\in I\}$$

the highest unmet need with respect to all criteria. In the paper the consumer goal is to reach a Pareto critical point of the vectorial minimization problem $\min\{F(x):x\in X\}$. Pareto critical points concern marginal satisfaction functions or marginal unmet needs functions $\omega(x)$ (see 2.1).

We use as a scalarization function, the lowest satisfaction level

$$\psi(x)=\min\{G_i(x):i\in I\}$$

or the highest unmet need

$$\phi(x)=\max\{F_i(x):i\in I\}$$

with respect to all criteria. This scalar function does not require any weighting process. Then, it escapes to big trade off problems which are time consuming, and a source of a lot of emotions (ex ante stress and hesitations, ex post regrets, etc.).

3) Trust-regions as consideration sets Consider a repeated choice problem. In his variational approach [12, 13] supposes that, at period k, the agent starts from somewhere, a chosen alternative $x^k\in X$. Before trying to choose a new alternative $x=x^k$, the agent will choose the context to choose this new alternative, a consideration set $\mathscr{C}(x^k)\subset X$. This allow him to escape to " too much choices". This implies that, in this first period, the agent will choose to ignore all the alternatives outside $\mathscr{C}(x^k)$. The given satisfaction level $G(x^k)\in\mathbb{R}^p$ is his lower frame (the present statu quo).

In Management Sciences, the literature on consideration sets abounds ([33, 34], see also [35], for a very good survey). Within each period k, the formation of the current consideration set $\mathscr{C}(x^k)$ for a given class of products is a categorization process which follows several stages,

1. Initial stage: available brands (total set)
2. Awareness stage : $\longrightarrow$(unaware brands, aware brands)
3. Processing stage :$\longrightarrow$ (unprocessed brands or foggy brands, processed brands)
4. Consideration stage: $\longrightarrow$(hold brands, reject brands, evoked brands)
5. Preference stage: $\longrightarrow$(other evoked brands, first choice)

In Applied Mathematics, trust-region methods (TRM) consider, each step, trust-regions as balls of variable radius $\mathscr{B}_k=\{d\in X:\|d\|\leq\Delta_k\}$, where $\Delta_k>0$ is the radius of this ball. In our application to satisficing processes we want to show that these methods offer a specific but important model for the choice and the formation of consideration sets (enlargement, shrinking, no change) driven by goal improvement. This TRM model of consideration sets seems to be important because it can give efficient computations of consideration sets.

4) Trust-region models: a way to quantify the degree of seriousness of a consideration process [12, 13] emphasizes that one of the main lessons coming from the literature on consideration sets is that, each period, the consideration process has two major stages: first the "not immediate rejection" of a subset of alternatives which defines the awareness set (the extensive aspect, each period, of the consideration process), ii) a more or less serious consideration of each alternative inside this awareness set (its intensive aspect). The consideration process is extensive if the awareness set contains a lot of alternatives. It is intensive if each alternative the agent is aware of is considered very seriously. The unprocessed (or foggy) set, the hold set, the rejected set, the evoked set, the first choice and second choice set offer different degree of seriousness.

Then, the problem of the formation of a consideration set has two sides:

i) which alternatives the agent is aware off? and how he becomes aware of them? The answer to these two first questions is either internally (using retrievial memory) or externally (from incidental exposure and salience of some stimuli, or intentional search using world of mouth communication, magazines, TV channels, books, etc.)

ii) how seriously the agent will consider each alternative he is aware off? and how he can perform a serious enough consideration of some of them. The answer to the second question is: the agent must perform a lot of consideration activities related to motivation, cognition and affect, like perception, search, exploration, estimation, evaluation, goal constructs, using categorization, approximations, heuristics, scaling, and framing processes, ...The more seriously he will perform each of these activities, the more serious will be the consideration process. The more criteria he will consider, the finer the scales he will use to estimate the degree of satisfaction of each criteria, the less "non compensatory" rules (crude rejection rules) he will use to eliminate some alternatives, the more compensatory rules he will use to solve trade off between conflicting criteria, the more seriously he will consider alternatives in his awareness set. This will also depend of the use of sequential considerations or parallel considerations, comparing all alternatives for each criteria, or evaluating each alternative with respect to all criteria or some of them. The formation of a consideration set requires insertions of new alternatives, exclusions and conservation of old alternatives of the previous consideration set. This is a reference dependent process.

To modelize these two sides, [12, 13] defined, not only a consideration set, but also the degree of seriousness of a consideration process within this set. Then, they have mixed in a new way the theory of consideration sets with the "disconfirmation bias theory" of consumer satisfaction [36] which makes the distinction between the ex ante (before usage) expected satisfaction level $\widehat{G}(x)$ and the ex post (after usage) satisfaction level $G(x)$. Satisfaction $G(x)$, as a post-usage phenomenon, is purely experiential. It is related to some performance usually reported on an objective scale bounded by good and bad levels of performance". The difference $\delta(x) = \widehat{G}(x) - G(x) = \widehat{F}(x) - F(x) \in \mathbb{R}^p$ is the disconfirmation bias, the difference between what was expected and what was observed, the expectation-performance discrepancy being better than or worse than expected with regard to a product or service . It represents an error of estimation, i.e., the degree of accuracy of the consideration process.

The smaller its norm $\delta[\mathscr{C}(x)] = \sup\{\|\delta(y)\| : y \in \mathscr{C}(x)\} < +\infty$, the more serious is the consideration process within $\mathscr{C}(x)$. Then, the higher the degree of seriousness, the lower will be the disconfirmation bias index.

Trust-region methods define, each step, a model function. In this paper, when the agent considers his unmet needs $F(x)$, and the related scalarization function ϕ the local approximation function (trust-region model) is

$$m(x,H,d) = \max_{i\in I}\{F_i(x) + \langle \nabla F_i(x), d\rangle\} + \tfrac{1}{2}\langle d, Hd\rangle \text{ with } H \in S^{n\times m}.$$

If the agent considers his satisfaction levels $G(x)$ and the related scalarization function ψ, the local approximation function is

$$n(x,H,d) = \min_{i\in I}\{G_i(x) + \langle \nabla G_i(x), d\rangle\} + \tfrac{1}{2}\langle d, Hd\rangle.$$

Proposition 5.1 determines "how good this approximation is". Under Assumptions 2 and 3 it shows that its degree of accuracy is proportional to the square of the trust-region ball radius. Hence this trust-region model defines in a precise way the disconfirmation bias $\left|\phi(x^k+d_k) - m_k(d_k)\right| = \left|\psi(x^k+d_k) - n_k(d_k)\right|$. The trust-region approach tells us that the lower is the square of the trust-region radius, the higher is the degree of seriousness of the consideration process, $\left|\phi(x^k+d_k) - m_k(d_k)\right| = \left|\psi(x^k+d_k) - n_k(d_k)\right| \leq k_{lsh}\Delta_k^2$.

5) Costs of consideration Consideration activities (perception, exploration, search, judgment, estimation, evaluation, etc.) are costly and the more serious the consideration process, the more costly it will be. The variational theory of change [12, 13] offers a detailed theory of consideration costs and more generally of costs to change that will not be reproduced here. The comparison of trust-region methods with the variational approach suggests (see the next section) that expected consideration costs $\mathscr{K}\left[\mathscr{C}(x^k)\right]$ differ when changes $\Delta_k > 0$ are low or large:

i) if changes are low, i.e., $\Delta_k < \frac{\omega(x^k)}{\beta_k}$, then expected consideration costs

$$\mathscr{K}\left[\mathscr{C}(x^k)\right] = \omega(x^k)\Delta_k$$

are proportional to marginal changes $\omega(x^k)$ and the factor of proportionality is the radius Δ_k of the trust-region $\mathscr{B}_k = \mathscr{C}(x^k)$.

ii) if changes are high, i.e., $\Delta_k > \frac{\omega(x^k)}{\beta_k}$, then expected consideration costs

$$\mathscr{K}\left[\mathscr{C}(x^k)\right] = (\frac{1}{\beta_k})\omega^2(x^k)$$

are proportional to the square of the marginal change, and the factor of proportionality is $\frac{1}{2}$.

The marginal change is

$$\omega(x) = -\min_{\|d\| \leq 1} \left\{ \max_{i \in I} \{ \langle \nabla F_i(x), d \rangle \} \right\}$$

or

$$\omega(x) = \max_{\|d\| \leq 1} \left\{ \min_{i \in I} \{ \langle \nabla G_i(x), d \rangle \} \right\}.$$

Lemma 2.1 gives several properties of this marginal change function which is non-negative. When the consumer uses one criteria to calibrate his unmet needs or his satisfaction level $(p = 1)$, his marginal change $\omega(x) = \|\nabla F(x)\|$ or $\omega(x) = \|GF(x)\|$ is the norm of the gradient of the unmet needs or the norm of the gradient of the satisfaction level.

Trust-region methods show that, for a given cost of consideration, the smaller the consideration set, the smaller will be the disconfirmation bias index.

6) "Sufficient decrease" conditions as "worthwhile changes" One of the main hypothesis of the variational rationality approach [12, 13] is that agents follow worthwhile changes where expected advantages to change are higher than some proportion of expected costs to change. In the context of trust-region methods, this assumption is a "sufficient decrease" condition of the model function $m_k(\cdot)$. Let us see why. In our multi-objective trust-region method, expected advantages to change are $m_k(0) - m_k(d_k)$ or $n_k(d_k) - n_k(0)$, expected costs to changes are $\mathscr{K}\left[\mathscr{C}(x^k)\right] = \omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta_k}, \Delta_k\right\}$ and the proportion is $\xi = \frac{1}{2}$. This is exactly the step condition (4.33), $m_k(0) - m_k(d_k) \geq \kappa_{\text{das}m}\omega(x^k) \min\left\{\frac{\omega(x^k)}{\beta_k}, \Delta_k\right\}$.

This means that either,

i) changes are large, $\omega(x^k) < \beta_k \Delta_k$, i.e.,

$$\Delta_k > \frac{\omega(x^k)}{\beta_k}, \text{ and } m_k(0) - m_k(d_k) \geq \left(\frac{\kappa_{\text{das}m}}{\beta_k}\right)\omega^2(x^k),$$

or

ii) changes are low, $\omega(x^k) > \beta_k \Delta_k$, i.e.,

$$\Delta_k < \frac{\omega(x^k)}{\beta_k}, \text{ and } m_k(0) - m_k(d_k) \geq \kappa_{\text{das}m}\omega(x^k)\Delta_k.$$

Such worthwhile changes exist. Lemma 4.1 provides an optimal solution example where $\kappa_{\text{das}m} = \frac{1}{2}$ and Lemma 4.2 provides a backtracking example where $\kappa_{\text{das}m} = \kappa_{\text{dpa}}$.

Cauchy points are well known points which provide sufficient decrease (see also the Dogled method and the Two dimensional subspace minimization methods, Nocedal and Wright [28, chapter 4]).

7) Acceptance criteria: setting a "satisficing enough" gap The problem is now to determine how an agent can set a satisficing level. The idea will be that " he will try to improve enough ". Hence the definition of satisficing as "improving enough". "But what is enough?" The variational approach proposes two ways to modelize "what is enough", using different lower and upper frames: i) to set a satisficing gap, with reference to a statu quo, ii) to fill some portion of some aspiration gap [12, 13]. Trust-region methods choose the first solution. Let us see how this work. Suppose that the agent has a scalar satisfaction level $g(\cdot) : z \in X \longmapsto g(z) \in \mathbb{R}$, and has yet experienced a given alternative $x \in X$. The statu quo is $g(x)$. Then, to "improve enough" with respect to the statu quo means to set a satisficing gap $\varepsilon(x, \cdot) > 0$, to define the reference dependent satisficing level $\widetilde{g}(x, \cdot) = g(x) + \varepsilon(x, \cdot)$ and to try to find $y \in X$ such that $g(y) \geq \widetilde{g}(x, \cdot)$, where the dot $(\cdot)$ represents some variable. In the present paper, $g(x) = \psi(x)$ is the scalarization function. Trust-region methods define a satisficing gap $\varepsilon(x, \cdot)$ as follows. They define an acceptance rule such that, at period k, the trial point x^{k+1} is accepted if it provides a "worthwhile enough change". The agent computes $\psi(x^k + d_k)$.Then, either

i) $x^{k+1} = x^k + d_k$ if $\psi(x^k + d_k) - \psi(x^k) \geq \rho_k\big(n_k(d_k) - n_k(0)\big)$ with $\rho_k \geq \eta_1$,

ii) otherwise, $x^{k+1} = x^k$.

This rule can be seen as a "satisficing enough" rule: set the satisficing level $\widetilde{g}(x, \cdot) = \widetilde{\psi}(x^k, d_k) = \psi(x^k) + \rho_k\big(n_k(d_k) - n_k(0)\big)$ such that $\rho_k \geq \eta_1$ (a "satisficing enough" condition)

Then, the satisficing gap is $\varepsilon(x^k, d_k) = \rho_k\big(n_k(d_k) - n_k(0)\big)$.

The acceptance criteria means that the change is worthwhile not only for the model function but also for the scalarization function,

$\psi(x^k + d_k) - \psi(x^k) \geq \rho_k\big(n_k(d_k) - n_k(0)\big) \geq \eta_1 \kappa_{\text{dasm}} \mathscr{K}\left[\mathscr{C}(x^k)\right]$ with $\rho_k \geq \eta_1$

8) To update the trust-region: the course of a satisficing process The trust-region method supposes that the updating rule is the following:

$\Delta_{k+1} \in [\Delta_k, +\infty[$ if $\rho_k \geq \eta_2$, i.e., increase the radius if the change is "very satisficing" (very successful in the TRM parlance, $\rho_k \geq \eta_2$)

$\Delta_{k+1} \in [\gamma_2 \Delta_k, \Delta_k]$ if $\eta_2 \geq \rho_k > \eta_1$, i.e., decrease "not too much" the radius if the change is "satisficing enough" (successful in the TRM parlance, $\eta_2 \geq \rho_k > \eta_1$)

$\Delta_{k+1} \in [\gamma_1 \Delta_k, \gamma_2 \Delta_k]$ if $\eta_1 > \rho_k$, i.e., decrease "very much" the radius if the change is " not satisficing enough" ($\eta_1 > \rho_k$)

for $0 < \eta_1 \leq \eta_2 < 1$ and $0 < \gamma_1 \leq \gamma_2 < 1$.

These conditions show that the satisficing process is adaptive, bracketing the choice process to divide the difficulty.

9) Global convergence It depends on the approximation solution which obtains at least a fixed positive fraction of the Cauchy point decrease in the model function $m(\cdot)$ [28, Chapter 4, section 4.2]. In behavioral terms, global convergence depends on the fact that the agent follows worthwhile changes and satisfies each step. Under Assumptions 1, 2 and 3 Proposition 6.1 shows that if x^k is not Pareto critical and if the radius of the trust-region is less than some portion of the marginal change, then the iterate is very satisficing, and will increase. Proposition 6.2 shows that if, along a path of changes, marginal changes are larger than some given constant, then, radius of all trust-regions will be also larger than another given constant. Proposition 6.3 and theorem 6.1 prove that stationarity at a Pareto critical point is obtained in a finite number of iterations (when there are finitely many successful iterations) and converges to zero of the marginal change.

8 Concluding Remarks

This paper has given an extension of trust-region methods for the multiobjective case. In this context, we get "Pareto critical convergence", which means that $\omega(x^k)$ converges to zero. If the sequence generated, by our TRMP, has accumulation points, then all them are Pareto critical points. Also, it has been shown that such methods offer a striking computable model of satisficing processes in Behavioral Sciences, using a recent variational approach where the new concept of "worthwhile changes" plays a key role.

References

[1] Conn, A.R., Gould, N.I.M., Toint, P.L.: Trust-region methods. MPS-SIAM Series on Optimization, Philadelphia, PA (2000)

[2] Ahookhosh, M., Amini, K.: A nonmonotone trust region method with adaptive radius for unconstrained optimization problems. Comput. Math. Appl. **60**(3), 411–422 (2010)

[3] Bastin, F., Malmedy, V., Mouffe M. Toint, P.L., Tomanos, D.: A retrospective trust-region method for unconstrained optimization. Math. Program. **123**(2, Ser. A), 395–418 (2010)

[4] Erway, J.B., Gill, P.E.: A subspace minimization method for the trust-region step. SIAM J. Optim. **20**(3), 1439–1461 (2009)

[5] Gardašević Filipović, M.: A trust region method using subgradient for minimizing a nondifferentiable function. Yugosl. J. Oper. Res. **19**(2), 249–262 (2009)

[6] Ji, Y., Li, Y., Zhang, K., Zhang, X.: A new nonmonotone trust-region method of conic model for solving unconstrained optimization. J. Comput. Appl. Math. **233**(8), 1746–1754 (2010)

[7] Yu, Z., Zhang, W., Lin, J.: A trust region algorithm with memory for equality constrained optimization. Numer. Funct. Anal. Optim. **29**(5-6), 717–734 (2008)

[8] Peng, Y.H., Shi, B.C., Yao, S.B.: Nonmonotone trust region algorithm for linearly constrained multiobjective programming. J. Huazhong Univ. Sci. Technol. Nat. Sci. **31**(7), 113–114 (2003)

[9] Yao, S.B., Shi, B.C., Peng, Y.H.: Nonmonotone trust region algorithms for multiobjective programming with linear constraints. Math. Appl. (Wuhan) **15**(suppl.), 55–59 (2002)

[10] Xi, H., Shi, B.C.: A trust region method for multiobjective programming without constraints. Math. Appl. (Wuhan) **13**(3), 67–69 (2000)

[11] Simon, H.: A behavioral model of rational choice. Quart. J. Econom. **69**(1), 99 –118 (1955)

[12] Soubeyran, A.: Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors. Mimeo (2009)

[13] Soubeyran, A.: Variational rationality and the unsatisfied man: routines and the course pursuit between aspirations, capabilities and beliefs. Mimeo (2010)

[14] Attouch, H., Soubeyran, A.: Inertia and reactivity in decision making as cognitive variational inequalities. J. Convex Anal. **13**(2), 207–224 (2006)

[15] Attouch, H., Soubeyran, A.: Local search proximal algorithms as decision dynamics with costs to move. Set-Valued Var. Anal. **19**(1), 157–177 (2010)

[16] Martinez-Legaz, J.E., Soubeyran, A.: A tabu search scheme for abstract problems, with applications to the computation of fixed points. J. Math. Anal. Appl. **338**(1), 620–627 (2007)

[17] Attouch, H., Redont, P., Soubeyran, A.: A new class of alternating proximal minimization algorithms with costs-to-move. SIAM J. Optim. **18**(3), 1061 (2007)

[18] Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Alternating proximal algorithms for weakly coupled convex minimization problems. applications to dynamical games and pde's. J. Convex Anal. **15**(3), 485–506 (2008)

[19] Attouch, H., Bolte, J., Redont, P., Soubeyran.A: Proximal alternating minimization and projection methods for nonconvex problems. an approach based on the kurdyka- lojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)

[20] Souza, S., Oliviera P.and Cruz Neto, J., Soubeyran, A.: A proximal point algorithm with separable bregman distances for quasiconvex optimization over the nonnegative orthant. European J. Oper. Res. **201**(2), 365–376 (2010)

[21] Luc, T.D., Sarabi, E., Soubeyran, A.: Existence of solutions in variational relations problems without convexity. J. Math. Anal. Appl. **364**(2), 544–555 (2010)

[22] Moreno, F., Oliveira, P., Sou beyran, A.: proximal algorithm with quasi distance. application to habits formation. Optimization (2011)

[23] Flores Bazan, F., Luc, T.D., Soubeyran, A.: Maximal elements under reference-dependent preferences with applications to behavioral traps and games. to appear in J. Optim. Theory. Appl. (2012)

[24] Godal, O., Flam, S., Soubeyran, A.: Gradient differences and bilateral barters. to appear in Optimization Manuscript ID: 679940 (2012)

[25] Cruz Neto, J., P., O., Soares, P., Soubeyran, A.: Learning how to play nash and alternating minimization method for structured nonconvex problems on riemannian manifolds. Accepted for publication. J. Convex Anal. (2012)

[26] Fliege, J., Svaiter, B.F.: Steepest descent methods for multicriteria optimization. Math. Methods Oper. Res. **51**(3), 479–494 (2000)

[27] Yigui, O., Qian, Z.: A nonmonotonic trust region algorithm for a class of semi-infinite minimax programming. Appl. Math. Comput. **215**(2), 474–480 (2009)

[28] Nocedal, J., Wright, S.: Numerical optimization. Springer, New York (1999)

[29] Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, Belmont, MA (1995)

[30] Shi, Z.J., Guo, J.H.: A new trust region method for unconstrained optimization. J. Comput. Appl. Math. **213**(2) 509–520 (2008)

[31] Lewin, K., Dembo, T., Festinger, L., Sears, P.: Level of aspiration. In Personality and the behavior disorders. New York: Ronald Press (1994)

[32] Bandura, A., Schunk, D.: Cultivating competence, self efficacy, and intrinsic interest through proximal self motivation. J. Pers. Soc. Psychol. **41**, 586-598 (1981)

[33] Brisoux, J.: Le phénomène des ensembles évoqués: une étude empirique des dimensions contenu et taille. Thèse, Université Laval, (1995).

[34] Brisoux, J. Laroche, M. : Evoked set formation and composition: an empirical investigation under a routinized response behavior situation. Advances in Consumer Research **8**, ed. Kent Monroe, Ann Arbor, Michigan, Association for Consumer Research, 357–361 (1981)

[35] Jolivot, A.: Thirty years of research on consideration set: a state of the art. Série “Recherche”. W.P. **502**, Institut d’Administration des Entreprises, Clos Guiot, 13540 Puyricard, France (1997)

[36] Oliver, R.: Satisfaction: a behavioral perspective on the consumer. McGraw-Hill, New York, NY (2011)