Journal of Convex Analysis
Volume 33 (2026), No. 3&4, 817–840

A Boosted Proximal Point Method for Difference of Convex Functions in Multiobjective Optimization and the Growth of Multiproduct Firms

J. X. Cruz Neto, J. O. Lopes, R. V. G. Serra, J. C. O. Souza
Dept. of Mathematics, Federal University of Piauí, Teresina, Brazil
jxavier@ufpi.edu.br, jurandir@ufpi.edu.br, rayserra@ufpi.edu.br, joaocos.mat@ufpi.edu.br

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

Dedicated to the memory of Hedy Attouch.

Received: January 30, 2025
Accepted: July 24, 2025

This paper has two sides: a mathematical side and a behavioral side. In the first part, we consider the unconstrained multiobjective optimization problem of finding Pareto critical points of difference of convex functions. We study a method that minimizes at each iteration a convex approximation instead of the (non-convex) objective function constrained to a possible non-convex set which imposes the method decreases, at each step, each objective function. In the second part, as an application, we propose a variational rationality perspective to multiproduct firm problem that wants to improve progressively the profits of all its sub-units by adjusting their production levels. This approach is completly new in management sciences/economics. It opens the door to a variational rational theory of the multiproduct firm.

Keywords: Multiobjective programming, proximal point method, DC function, variational rationality, multiproduct firm.

2020 Mathematics Subject Classification: 65K05, 65K10, 90C26, 47N10, 58E17.

1. Introduction

The multiobjective optimization problem involves simultaneously minimizing two or more objective functions. More precisely, consider the vector-valued function $F : \mathbb{R}^n \to \mathbb{R}^m$ defined by $F(x) := (F_1(x), \ldots, F_m(x))$. Then, the unconstrained multiobjective optimization problem is denoted as

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

This research was supported in part by CNPq grant 302156/2022-4 and 315937/2023-8. The project leading to this publication has received funding from the French government under the "France 2023" investment plan managed by the French National Research Agency (reference: ANR-17-EURE-0020) and from Excellence Initiative of Aix-Marseille University-A*MIDEX.

ISSN 0944-6532 / $ 2.50

In recent years, significant efforts have been dedicated to extending and adapting scalar optimization methods to the multiobjective or vector optimization setting. This approach was first introduced by Fliege and Svaiter in [24], where a steepest descent method for multiobjective optimization was proposed and analyzed. Following the seminal work [24], numerous studies have addressed the extension of scalar optimization methods to the multiobjective context; see for instance, the survey by Fukuda and Drummond in [20].

The work of Bonnel, Iusem, and Svaiter [14] was the first to consider an extension of the proximal point method to the multiobjective optimization context. If we restrict the analysis to the finite-dimensional multiobjective setting, this method generates a sequence that satisfies

$$x^{k+1} \in \operatorname{argmin}_w \left\{ F(x) + \frac{\lambda_k}{2} ||x - x^k||^2 \varepsilon^k \ : \ x \in \Omega_k \right\},$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathbb{R}^m_+$-convex (or equivalently, each F_j is a convex function, for $j = 1, \ldots, m$), $\{\lambda_k\} \subset \mathbb{R}_{++}$, $\{\varepsilon^k\} \subset \mathbb{R}^m_{++}$ (see Section 2 for notation) such that $\|\varepsilon^k\| = 1$ and $\Omega_k = \{x \in \mathbb{R}^n \ : \ F(x) \preceq F(x^k)\}$. They proved that, in the $\mathbb{R}^m_+$-convex case, finding a Pareto optimal solution of (1) is equivalent to minimizing the scalar function $x \mapsto \langle z, F(x) \rangle$ for $z \in \mathbb{R}^m$ in the positive polar cone; see [14] for more details. Following [14], several works have been published studying the proximal point method in the multiobjective context, improving and extending the results from the scalar setting. A comprehensive survey on this topic can be found in Grad [23].

This paper focuses on using the proximal point method to solve problem (1), assuming that the objective function has a special difference of convex (DC) structure, i.e.,

$$F(x) := G(x) - H(x),$$

where $G, H : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable and $\mathbb{R}^m_+$-convex. The first study to extend the proximal point method to the multiobjective setting with vector-valued DC functions was done by Ji et al. [30]. However, their method does not satisfy the descent property $F(x^{k+1}) \preceq F(x^k)$, for all $k \in \mathbb{N}$, unlike the method proposed by Bonnel et al. [14] for the $\mathbb{R}^m_+$-convex case. In Bento et al. [10], a proximal point method preserving this descent property is considered for vector-valued DC functions; see also Souza et al. [53].

Recently, Aragon Artacho et al. [4, 5] introduced a boosted DC algorithm (BDCA) designed to solve scalar DC problems. This method enhances the convergence efficiency of the traditional DC algorithm (DCA), which was specifically developed for scalar DC functions; see [56]. While DCA is inherently a descent method, BDCA improves upon it by computing a descent direction based on the point generated by DCA and incorporating a line search to take larger decreasing steps. This approach results in a more significant reduction in the objective value per iteration compared to DCA. This method was extended to the proximal point method in [3] in the scalar case.

The aim of this paper is to introduce a boosted version of the proximal point method, as discussed in [3], adapted for the context of multiobjective DC optimization. This approach improves the computational performance of existing methods presented in

[10, 30, 53]. By incorporating the acceleration technique proposed in [4, 5], we aim to achieve greater efficiency in solving multiobjective DC optimization problems. As an application, we apply this framework to a DC vectorial optimization problem, focusing on a multiproduct firm that progressively improve its vectorial profit by adjusting production levels.

This paper is organized as follows. In Section 2, we present fundamental concepts and results in multiobjective optimization. In Section 3, we introduce the boosted proximal point method and establish its well-definition. In Section 4, we state and prove the convergence results of the proposed method. Section 5 provides iteration-complexity bounds for the method. Finally, in Section 6, we propose an application to a multiproduct firm.

2. Multiobjective optimization

In this section, we establish some basic definitions and properties of multiobjective optimization which can be found, for instance, in Jahn [29] and Luc [36]. We state and prove some results that will allow us to define the algorithm and study its convergence properties.

Let $\mathbb{R}^m$ be the m-dimensional Euclidean space. In order to define the partial order in $\mathbb{R}^m$, we consider

$$\mathbb{R}^m_+ = \{x \in \mathbb{R}^m \ : \ x_j \geq 0, \ j \in \mathcal{I}\} \quad \text{and} \quad \mathbb{R}^m_{++} = \{x \in \mathbb{R}^m \ : \ x_j > 0, \ j \in \mathcal{I}\},$$

with $\mathcal{I} = \{1, \ldots, m\}$. The partial order "$\preceq$" induced by the Paretian cone $\mathbb{R}^m_+$ is defined by $y \preceq z$ (or $z \succeq y$) if $z - y \in \mathbb{R}^m_+$, and its associated strict relation $\prec$ holds when $z - y \in \mathbb{R}^m_{++}$, that is, $y \prec z$ (or $z \succ y$). In the following, we state some notion about the optimality condition for a multiobjective optimization problem. Moreover, since typically lacks a single point that minimizes all functions simultaneously, we use the concept of Pareto optimality to characterize a solution, as defined below.

Definition 2.1. A point $\bar{x} \in \mathbb{R}^n$ is said to be a *weakly Pareto optimal* point for (1) if there does not exist another $x \in \mathbb{R}^n$ such that $F(x) \prec F(\bar{x})$.

Now, consider a vector function denoted by $F(x) := (F_1(x), \ldots, F_m(x))^\top$. The Jacobian $m \times n$ matrix of F at $x \in \mathbb{R}^n$ is defined by

$$JF(x) := (\nabla F_1(x), \ldots, \nabla F_m(x))^\top .$$

In the multiobjective setting, the well known condition for optimality (necessary, but, in general, nonsufficient) for problem (1) is the notion of Pareto critical point. Briefly, a point, $\bar{x} \in \mathbb{R}^n$ is

$$(-\mathbb{R}^m_{++}) \cap \operatorname{Im}(JF(\bar{x})) = \varnothing. \tag{2}$$

A point x is *critical* if it satisfies (2). Therefore, if a point x is not critical, there exists a direction $v \in \mathbb{R}^n$ satisfying

$$JF(x)v \prec 0. \tag{3}$$

It is a well known fact that such v is a descent direction for the objective F if v satisfies (3), in this case there exists $\bar{t} > 0$ such that

$$F(x + tv) \prec F(x) \text{ for all } t \in (0, \bar{t}).$$

If v is a descent direction at x, we say that $t > 0$ satisfies the "Armijo-like" rule for $\beta \in (0, 1)$ if

$$F(x + tv) \preceq F(x) + \beta t JF(x)v \text{ for all } t \in (0, \bar{t}).$$

Proposition 2.2. *Let $\beta \in (0, 1)$. If $JF(x)v \prec 0$, then there exists $\bar{t} > 0$ such that*

$$F(x + tv) \prec F(x) + \beta t JF(x)v \textit{ for all } t \in (0, \bar{t}).$$

Proof. See [24]. □

For a vector function $F : \mathbb{R}^n \to \mathbb{R}^m$ we say that F is $\mathbb{R}^m_+$*-convex* if and only if, for every $x, y \in \mathbb{R}^n$ and $\lambda \in (0, 1)$

$$F(\lambda x + (1 - \lambda)y) \preceq \lambda F(x) + (1 - \lambda)F(y).$$

Note that this concept is equivalent to component-wise convexity.

Proposition 2.3. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a differentiable function. Then, F is a convex function if, only if, for every $x, y \in \mathbb{R}^n$*

$$JF(x)(y - x) \preceq F(y) - F(x).$$

Proof. See [12]. □

Proposition 2.4. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a convex differentiable function. Then, $\bar{x}$ is a weak Pareto optimal point of F if, only if,*

$$(-\mathbb{R}^m_{++}) \cap Im(JF(\bar{x})) = \varnothing.$$

Proof. See [12]. □

A scalar-valued function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is *locally Lipschitz at a point $x \in \mathbb{R}^n$* if there exists a neighborhood U of this point and some real number $L > 0$ such that

$$|\varphi(y) - \varphi(y')| \leq L||y - y'||, \quad \forall y, y' \in U.$$

A function φ is *locally Lipschitz* when it is locally Lipschitz at all points.

The next result presents a necessary condition for a point x^* to be a solution of the problem

$$\min_w \{F(x) : \; x \in S\}, \tag{4}$$

where $F(x) = (F_1(x), F_2(x), \cdots, F_m(x)) : \mathbb{R}^n \to \mathbb{R}^m$ and

$$S = \{x \in \mathbb{R}^n : s_j(x) \leq 0, \quad j \in \mathcal{I}\} \quad \text{with} \quad s_j : \mathbb{R}^n \to \mathbb{R}. \tag{5}$$

Lemma 2.5. *Let $w \in \mathbb{R}^m_+ \backslash \{0\}$ and assume that $S \subset \mathbb{R}^n$ is a nonempty and closed set. If $-JF(x)^\top w \in N_S(x)$, then x is a Pareto critical point of F.*

Proof. See Bento at al. [11]. □

Theorem 2.6. *Let S be a set as in* (5)*, and the functions $F_j, s_j : \mathbb{R}^n \to \mathbb{R}$, $j \in \mathcal{I}$, are locally Lipschitz. If $x^* \in \mathbb{R}^n$ is a weakly Pareto solution of* (4)*, then there exist real numbers $u_j \geq 0$, $v_j \geq 0$, with $j \in \mathcal{I}$ such that*

$$0 \in \sum_{j\in\mathcal{I}} u_j \partial F_j(x^*) + \sum_{j\in\mathcal{I}} v_j \partial s_j(x^*),$$

$$\textit{with} \quad \sum_{j\in\mathcal{I}} (u_j + v_j) = 1 \quad \textit{and} \quad v_j s_j(x^*) = 0, \quad j \in \mathcal{I}.$$

Proof. The proof follows from Minami [37, Theorem 3.1]. □

Remark 2.7. Theorem 2.6 follows the formulation established by Minami [37, Theorem 3.1], which is stated for locally Lipschitz continuous functions and employs the Clarke subdifferential. Nevertheless, in the present work, all functions involved are assumed to be continuously differentiable. As a consequence, the Clarke subdifferential reduces to the classical gradient, and the optimality condition in Theorem 2.6 can be equivalently expressed in terms of gradients.

3. Proximal point method

If we restrict the analysis to the finite dimensional multiobjective setting, the method proposed in [14] generates a sequence satisfying

$$x^{k+1} \in \operatorname{argmin}_w \left\{ F(x) + \frac{\lambda_k}{2} ||x - x^k||^2 \varepsilon^k \;:\; x \in \Omega_k \right\},$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is $\mathbb{R}^m_+$-convex, $\{\lambda_k\} \subset \mathbb{R}_{++}$, $\{\varepsilon^k\} \subset \mathbb{R}^m_+$ and $\Omega_k = \{x \in \mathbb{R}^n : F(x) \preceq F(x^k)\}$. Note that the constrained set Ω_k requires that the algorithm is a descent method in the sense of $F(x^{k+1}) \preceq F(x^k)$.

From now on, we assume that the vector-valued function $F : \mathbb{R}^n \to \mathbb{R}^m$ is a multiobjective DC function, that is, it can be written as $F(x) = G(x) - H(x)$, where G and H are $\mathbb{R}^m_+$-convex and differentiable functions. Take the auxiliary sequences $\{\lambda_k\} \subset \mathbb{R}_{++}$ such that $0 < a \leq \lambda_k \leq b$ and $\{\varepsilon^k\} \subset \mathbb{R}^m_{++}$ such that $\varepsilon^k := e = (1, 1, \ldots, 1)$ for all k. In this case, the accelerated proximal point method generates a sequence $\{x^k\}$ as follows:

Algorithm 1

1: Choose $\beta \in (0, 1)$ and $x^0 \in \mathbb{R}^n$. Set $k := 0$.
2: If x^k is a Pareto critical point, stop. Otherwise,
3: find $y^k \in \mathbb{R}^n$ such that

$$y^k \in \arg\min_w \big\{ G(x) - JH(x^k)(x - x^k) + \frac{\lambda_k}{2} \|x - x^k\|^2 e \;:\; x \in \Omega_k \big\}, \tag{6}$$

where $\Omega_k = \{x \in \mathbb{R}^n \;:\; F(x) \preceq F(x^k)\}$.
4: If $y^k = x^k$, stop. Otherwise,
5: define $d^k = y^k - x^k$ and find $t_k > 0$ such that

$$F(y^k + t_k d^k) \preceq F(y^k) + \beta t_k JF(x^k) d^k. \tag{7}$$

6: Set $x^{k+1} := y^k + t_k d^k, k := k + 1$ and return to 2.

Remark 3.1. It is worth to mention that Algorithm 1 satisfies $y^k \in \Omega_k$, for all $k \geq 0$, i.e.,

$$F(y^k) \preceq F(x^k).$$

If $y^k = x^k$, from optimality condition of (6), we have that x^k is a Pareto critical point. As mentioned in Huang and Yang [28], the vector functions

$$F(\cdot) \quad \text{and} \quad e^{F(\cdot)} := (e^{f_1(\cdot)}, \ldots, e^{f_m(\cdot)})$$

have the same set of weakly Pareto points, where e^α denotes the exponential map valued at $\alpha \in \mathbb{R}$. This result can be easily extended to the Pareto critical setting. Hence, concerning Pareto critical points, we can assume without loss of generality that $F \succ 0$. On the other hand, any non-negative DC function admits a non-negative DC decomposition; see [26]. Therefore, we will assume throughout the paper that $G \succeq 0$.

Proposition 3.2. *The Algorithm 1 is well defined.*

Proof. The starting point $x^0 \in \mathbb{R}^n$ is chosen in the initialization step. Assuming in iteration k that x^k is given. Denote by

$$F_k(x) := G(x) - JH(x^k)(x - x^k) + \frac{\lambda_k}{2}\|x - x^k\|^2 e.$$

Note that $x^k \in \Omega_k$ which implies that $F_k(\Omega_k)$ is nonempty. It is straightforward to check that $F_k(x) \succ 0$ and $F_k(\Omega_k)$ is closed. Since the cone $\mathbb{R}^m_+$ is Daniell, it follows from [36, Lemma 3.5] that $F_k(\Omega_k)$ is $\mathbb{R}^m_+$-complete. Thus, from [36, Theorem 3.3] there exists

$$y^k \in \arg\min_w\{F_k(x) \; : \; x \in \Omega_k\}.$$

Let $d^k = y^k - x^k$, we will show that $d^k \neq 0$ is a descent direction of F at y^k. Indeed, for all $k \in \mathbb{N}$, by Theorem 2.6 there exist $u^k, v^k \in \mathbb{R}^m_+$ with $||u^k + v^k||_1 = 1$ such that

$$[JG(y^k) - JH(x^k)]^\top u^k + \lambda_k(y^k - x^k)||u^k||_1 + JF(y^k)^\top v^k = 0. \tag{8}$$

Note that $$JF(y^k)^T v^k = v_1^k \nabla F_1(y^k) + \ldots + v_n^k \nabla F_m(y^k)$$

then $$\langle JF(y^k)^T v^k, d^k\rangle = v_1^k \langle \nabla F_1(y^k), d^k\rangle + \ldots v_m^k \langle \nabla F_m(y^k), d^k\rangle.$$

Now, we have for $i = 1, \ldots, m$ that

$$v_i^k \langle \nabla F_i(y^k), d^k\rangle = -u_i^k \langle \nabla G_i(y^k) - \nabla H_i(x^k), d^k\rangle - \lambda_k ||u^k||_1 \|d^k\|^2,$$

and hence,

$$v_i^k \langle \nabla F_i(y^k), d^k\rangle = -u_i^k \langle \nabla F_i(y^k), d^k\rangle - u_i^k \langle \nabla H_i(y^k) - \nabla H_i(x^k), d^k\rangle - \lambda_k ||u^k||_1 \|d^k\|^2.$$

Using the fact that ∇H_i is monotone for $i = 1, \ldots, m$ we obtain

$$(v_i^k + u_i^k)\langle \nabla F_i(y^k), d^k\rangle \leqslant -\lambda_k ||u^k||_1 \|d^k\|^2 < 0,$$

since $u^k, v^k \succeq 0$ and $\sum_{j\in\mathcal{I}}(u_j + v_j) = 1$, follow the result, i.e. $JF(y^k)d^k \prec 0, \; \forall \; k$.

Now, by Proposition 2.4 there exists t_k satisfying (7) such that

$$F(y^k + t_k d^k) \preceq F(y^k) + \beta t_k JF(y^k) d^k.$$

Thus, define $x^{k+1} = y^k + t_k d^k$. □

From now on, $\{x^k\}$, $\{y^k\}$ and $\{\lambda_k\}$ denote the sequences considered in Algorithm 1.

Remark 3.3. Note that $\{u^k\}$ and $\{v^k\}$ are bounded sequences. As mentioned by Bolte et al. [13, Remark 1], $\partial g_j, \partial h_j$ is bounded on compact sets. So, we have that $\{JG(y^k)\}$ and $\{JH(x^k)\}$ are bounded sequences as long as $\{x^k\}$ and $\{y^k\}$ are bounded. □

As a consequence of the previous proposition, we obtain the following stopping rule for Algorithm 1.

Corollary 3.4. *Let $k_0 \in \mathbb{N}$ be such that $u^{k_0} = 0$. Then, y^{k_0} is a Pareto critical point of F.*

Proof. If there exists $k_0 \in \mathbb{N}$ such that $u^{k_0} = 0$, then from (8) it implies that $JF(y^{k_0})^T v^{k_0} = 0$. On the other hand, since $||u^k + v^k||_1 = 1$, we can say that $v^{k_0} \in \mathbb{R}^m_+ \backslash \{0\}$. The desired result follows by using Lemma 2.5 with $C = \mathbb{R}^n$, $w = v^{k_0}$ and $x = y^{k_0}$. □

4. Convergence analysis

In this section, we analyze the convergence of Algorithm 1. Note that, if there exists k_0 such that $x^{k_0} = y^{k_0}$ or $u^{k_0} = 0$, then y^{k_0} is a Pareto critical point of F. Throughout this paper, we suppose that $x^k \neq y^k$ and $u^k \neq 0$ for all $k \geq 0$.

Proposition 4.1. *The following properties hold:*
(i) $\{F(x^k)\}$ *is decreasing;* (ii) $\lim_{k\to+\infty} \|y^k - x^k\| = 0$.

Proof. (i) Follows from step 5 of the algorithm.

(ii) From the definition of the algorithm we have that y^k, for each $k \in \mathbb{N}$, is an optimal solution of the problem

$$\min\nolimits_w \left\{ G(x) - JH(x^k)(x - x^k) + \frac{\lambda_k}{2}||x - x^k||^2 e \, : \, x \in \Omega_k \right\}.$$

This implies

$$\max_{1\leq j\leq m} \left\{ g_j(x^k) - g_j(y^k) + \langle \nabla h_j(x^k), y^k - x^k \rangle - \frac{\lambda_k}{2}||y^k - x^k||^2 \right\} \geq 0.$$

Take $j_0(k) = j_0 \in \{1, \ldots, m\}$ the index where the maximum in last inequality is attained. Then, from the lower boundedness assumption of $\{\lambda_k\}$, we have

$$\frac{a}{2}||x^k - y^k||^2 \leq g_{j_0}(x^k) - g_{j_0}(y^k) + \langle \nabla h_j(x^k), y^k - x^k \rangle. \quad (9)$$

By convexity of h_{j_0}, we have

$$\langle \nabla h_{j_0}(x^k), y^k - x^k \rangle \leq h_{j_0}(y^k) - h_{j_0}(x^k). \quad (10)$$

Combining (9) and (10), we obtain

$$\frac{a}{2}||x^k - y^k||^2 \leq F_{j_0}(x^k) - F_{j_0}(y^k) \leq F_{j_0}(x^k) - F_{j_0}(x^{k+1}) \leq \max_{1\leq i\leq m}\{F_i(x^k) - F_i(x^{k+1})\}.$$

Since $F(x^{k+1}) \preceq F(y^k) \preceq F(x^k)$ and $F \succ 0$, we obtain that the right-hand side of (9) converges to 0 as $k \to +\infty$. Hence, $||x^k - y^k|| \to 0$ as $k \to +\infty$. □

Theorem 4.2. *Every cluster point of $\{x^k\}$, if any, is a Pareto critical point of F.*

Proof. Let $\hat{x}$ be a cluster point of $\{x^k\}$ and consider a subsequence $\{x^{k_l}\}$ of $\{x^k\}$ converging to $\hat{x}$ and by Proposition 3.2 $\{y^{k_l}\}$ of $\{y^k\}$ also converges to $\hat{x}$. Since $\{x^{k_l}\}$ is convergent and $\{\lambda_k\}$ is bounded, it follows from Remark 3.3 that $\{JG(y^k)\}$, $\{JH(x^k)\}$, $\{u^k\}$ and $\{v^k\}$ are bounded. Thus, we can assume without loss of generality that $JG(y^{k_l}) \to JG\hat{x}$, $JH(x^{k_l}) \to JH\hat{x}$, $u^{k_l} \to \hat{u}$ and $v^{k_l} \to \hat{v}$(we may extract other subsequences if necessary). From (8) we have

$$[JG(y^{k_l}) - JH(x^{k_l})]^\top u^{k_l} + \lambda_{k_l}(y^{k_l} - x^{k_l})||u^{k_l}||_1 + JF(y^{k_l})^\top v^{k_l} = 0. \tag{11}$$

Set $\gamma_{k_l} = \lambda_{k_l}||u^{k_l}||_1$. Note that $\{\gamma_k\}$, is bounded, because

$$0 < \gamma_k = \lambda_k||u^k||_1 \leq \lambda_k \leq b.$$

This fact combined with Proposition 4.1 implies that $\{\gamma_{k_l}(y^{k_l} - x^{k_l})\}$ converges to 0 as $l \to +\infty$. Thus, taking the limit as $l \to +\infty$ in (11), we have

$$[(JG - JH)(\hat{x})]^\top \hat{u} + JF(\hat{x})^\top \hat{v} = JF(\hat{x})^\top(\hat{u} + \hat{v}) = 0.$$

Note that $\hat{u}, \hat{v} \in \mathbb{R}^m_+$, $||\hat{u} + \hat{v}||_1 = 1$ and $JF(\hat{x})^\top(\hat{u} + \hat{v}) = 0$. Therefore, by Lemma 2.5 $\hat{x}$ is Pareto critical of F. □

4.1. Convergence analysis: convex case

Throughout this section we will assume that F is $\mathbb{R}^m_+$-convex, in the sense that $H = 0$. We will consider that the set below is non-empty

$$U := \{x \in \mathbb{R}^n | \ F(x) \preceq F(y^k), \forall k\}.$$

Proposition 4.3. *For all $\bar{x} \in U$, the following conditions hold:*

(a) $\|y^k - \bar{x}\| \leq \|x^k - \bar{x}\|$;

(b) $\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 + (t_k)^2||d^k||^2$;

(c) $\sum_{k=1}^{\infty}(t_k)^2||d^k||^2 < \infty$.

Proof. (a) Let $\bar{x} \in U$, we have $F(y^k) - F(\bar{x}) \succeq 0$ for all k. By convexity of F and the fact that $\bar{x} \in U$, we obtain

$$JF(y^k)(y^k - \bar{x}) \succeq F(y^k) - F(\bar{x}) \succeq 0. \tag{12}$$

From (8), we obtain that $JF(y^k)^\top u^k + \gamma_k(y^k - x^k) + JF(y^k)^\top v^k = 0$, which implies

$$JF(y^k)^\top(u^k + v^k) = -\gamma_k d^k, \tag{13}$$

where $\gamma_k = \lambda_k||u^k||_1$ and $d^k = y^k - x^k$.

Through simple manipulations and (13), we have

$$\begin{aligned}||x^k - \bar{x}||^2 &= ||x^k - y^k||^2 + ||y^k - \bar{x}||^2 + 2\langle x^k - y^k, y^k - \bar{x}\rangle \\ &= ||x^k - y^k||^2 + ||y^k - \bar{x}||^2 + 2(\gamma_k)^{-1}\langle -\gamma_k d^k, y^k - \bar{x}\rangle \\ &= ||x^k - y^k||^2 + ||y^k - \bar{x}||^2 + 2(\gamma_k)^{-1}\langle u^k + v^k, JF(y^k)(y^k - \bar{x})\rangle.\end{aligned}$$

From (12), we have that $2(\gamma_k)^{-1}\langle u^k + v^k, JF(y^k)(y^k - \bar{x})\rangle \geq 0$. In conclusion we have $||y^k - \bar{x}|| \leq ||x^k - \bar{x}||$.

(b) Using the fact that $x^{k+1} = y^k + t_k d^k$, we get

$$\begin{aligned}||x^{k+1} - \bar{x}||^2 &= ||x^{k+1} - y^k||^2 + ||y^k - \bar{x}||^2 + 2\langle x^{k+1} - y^k, y^k - \bar{x}\rangle \\ &= (t_k)^2||d^k||^2 + ||y^k - \bar{x}||^2 + 2t_k\langle d^k, y^k - \bar{x}\rangle \\ &= (t_k)^2||d^k||^2 + ||y^k - \bar{x}||^2 - 2(\gamma_k)^{-1}t_k\langle (u^k + v^k), JF(y^k)(y^k - \bar{x})\rangle.\end{aligned}$$

Using again (12), we obtain $||x^{k+1} - \bar{x}||^2 \leq ||x^k - \bar{x}||^2 + (t_k)^2||d^k||^2$.

(c) According to the steps of Algorithm 1 and Remark 3.1, we have

$$-\beta t_k JF(y^k)d^k \preceq F(x^k) - F(x^{k+1}).$$

Thus, $$-\beta t_k\langle JF(y^k)d^k, (u^k + v^k)\rangle \leq \langle F(x^k) - F(x^{k+1}), (u^k + v^k)\rangle.$$

The fact that $\langle JF(y^k)d^k, (u^k + v^k)\rangle = \langle d^k, JF(y^k)^\top(u^k + v^k)\rangle$ and (13), we get

$$\beta t_k(\gamma)^{-1}||d^k||^2 \leq \langle F(x^k) - F(x^{k+1}), (u^k + v^k)\rangle.$$

Therefore, by Cauchy-Schwartz inequality and equivalence of norms, we obtain

$$t_k||d^k||^2 \leq \frac{b}{\beta}\sum_{i=1}^{m}(F_i(x^k) - F_i(x^{k+1})).$$

Now, adding up from $k = 0$ to $k = N$, we get

$$\sum_{k=0}^{N} t_k||d^k||^2 \leq \frac{b}{\beta}\sum_{i=1}^{m}(F_i(x^0) - F_i(x^{N+1})) \leq \frac{b}{\beta}||F(x^0) - F(\bar{x})||_1. \tag{14}$$

Since the last inequality holds for any nonnegative integer N , we conclude that $\sum_{k=0}^{\infty} t_k||d^k||^2 < \infty$, and as a consequence $\sum_{k=0}^{\infty}(t_k)^2||d^k||^2 < \infty$. □

Before we give the main result of this section, let us recall that a sequence $\{z^k\}$ is said to be quasi-Fejér convergent to a non-empty set $U \subset \mathbb{R}^n$ if, for every $u \in U$, there exist a sequence $\{\epsilon_k\} \subset \mathbb{R}_+$ such that $\sum \epsilon_k < \infty$ and

$$||z^{k+1} - u|| \leq ||z^k - u|| + \epsilon_k, \quad \forall k = 0, 1, 2, \cdots .$$

The following result is well known and its proof is elementary.

Proposition 4.4. *Let $U \subset \mathbb{R}^n$ be a non-empty set and $\{z^k\}$ be a quasi-Fejér convergent sequence to U. Then, $\{z^k\}$ is bounded. Moreover, if a cluster point z of $\{z^k\}$ belongs to U, then $\{z^k\}$ converges to z.*

Theorem 4.5. *The sequence $\{x^k\}$ converges to a Pareto critical point of F.*

Proof. By Proposition 4.3 the sequence $\{x^k\}$ is quasi-Fejér convergent to U. Hence, $\{x^k\}$ is bounded and all its cluster points belongs U, because $F(x^{k+1}) \preceq F(x^k)$. It follows that $\{x^k\}$ converges; thus by Theorem (4.2) the proof is complete. □

5. Iteration-complexity analysis

In this section we present an iteration-complexity bound, for this from now on we will consider $F_i : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable and with a gradient Lipschitz continuous with constant $L_i > 0, i = 1, 2, \cdots, m$.

Lemma 5.1. *In Algorithm 1 the stepsize always satisfies*

$$t_k \geq t_{min} = \min\left\{\frac{(1-\beta)}{2L_{max}}, 1\right\},$$

where $L_{max} = \max\{L_1, L_2, \cdots, L_m\}$ and $\beta \in (0,1)$ is the parameter of the sufficient decrease condition (7).

Proof. See Fliege at al. [19, Lemma 3.1]. □

Remark 5.2. It is interesting to note that, for $F_i : \mathbb{R}^n \to \mathbb{R}$ continuously differentiable and with a gradient Lipschitz continuous, the analysis of method with an Armijo line search can be considered constant stepsizes $t_k = t_{min}$ for all $k \geq 0$; therefore, for the sake of simplicity, this will be the case in this section. □

Proposition 5.3. *Let $\{x^k\}$ and $\{y^k\}$ be the sequences generated accelerated proximal point method with stepsizes $t_k = t_{min}$, for all $k \geq 0$. Then, for every $\bar{x} \in U$, there hold*

(a) $\gamma_k \left(\|y^k - x^k\|^2 + \|\bar{x} - y^k\|^2 - \|\bar{x} - x^k\|^2\right) \leq 2||F(\bar{x}) - F(y^k)||_1$;

(b) $\gamma_k \left(\|\bar{x} - x^{k+1}\|^2 - (t_{min})^2\|d^k\|^2 - \|\bar{x} - y^k\|^2\right) \leq 2||F(\bar{x}) - F(y^k)||_1$,

where $\gamma_k = \lambda_k||u^k||_1$.

Proof. It follows immediately from the proof of Proposition 4.3. □

Theorem 5.4. *Let $\{x^k\}$ and $\{d^k\}$ be the sequences generated accelerated proximal point method with stepsizes $t_k = t_{min}$, for all $k \geq 0$. Then, for every $N \in \mathbb{N}$, there holds*

$$\min\{||d^k||;\ k = 0, 1, 2, \cdots, N\} \leq \frac{b\sqrt{m}}{\beta t_{min}} \frac{\sqrt{F^{max}(x^0) - F^{min}(\bar{x})}}{\sqrt{N+1}},$$

where $F^{max}(x^0) = \max\limits_{1\leq i\leq m}\{F_i(x^0)\}$ and $F^{min}(\bar{x}) = \min\limits_{1\leq i\leq m}\{F_1(\bar{x}), \cdots, F_m(\bar{x})\}$.

Proof. From (14) and the fact that $t_k = t_{min}$, we get

$$\sum_{k=0}^{N} ||d^k||^2 \leq \frac{b}{\beta t_{min}} \sum_{i=1}^{m}(F_i(x^0) - F_i(\bar{x})) \leq \frac{bm}{\beta t_{min}}(F^{max}(x^0) - F^{min}(\bar{x})).$$

Hence, we have

$$(N+1)\min\{||d^k||^2;\ k=0,1,2,\cdots,N\} \le \frac{b}{\beta t_{min}}(F^{max}(x^0)-F^{min}(\bar{x})),$$

which proves the assertion. □

Theorem 5.5. *Let $\{x^k\}$ and $\{d^k\}$ be the sequences generated accelerated proximal point method with stepsizes $t_k = t_{min}$, for all $k \ge 0$. Then, for every $N \in \mathbb{N}$, there holds*

$$||F(x^{N+1})-F(\bar{x})||_1 \le \frac{b}{2}\left[\frac{||x^0-\bar{x}||^2+\frac{bm}{\beta}(F^{max}(x^0)-F^{min}(\bar{x}))}{N+1}\right],$$

where $F^{max}(x^0) = \max\limits_{1\le i\le m}\{F_i(x^0)\}$ *and* $F^{min}(\bar{x}) = \min\limits_{1\le i\le m}\{F_i(\bar{x})\}$.

Proof. It follows from the Proposition 5.3(b) that

$$2\sum_{i=1}^{m}(F(y^k)-F(\bar{x})) \le \gamma_k\left[\|\bar{x}-x^k\|^2-\|\bar{x}-x^{k+1}\|^2+(t_{min})^2\|d^k\|^2\right].$$

Now, adding up from $k=0$ to $k=N$ and using the facts:

$$F(x^{k+1}) \preceq F(y^k),\ \gamma_k \le b \ \text{ and } \ \sum_{k=0}^{N} t_{min}||d^k||^2 \le \frac{b}{\beta}\sum_{i=1}^{m}(F(x^0)-F(\bar{x})),$$

we get

$$2(N+1)||F(x^{N+1})-F(\bar{x})||_1 \le b\left[\|\bar{x}-x^0\|^2+\frac{bm}{\beta}(F^{max}(x^0)-F^{min}(\bar{x}))\right].$$

Therefore, we obtain the desired result. □

6. Application to the multiproduct firm: a variational rationality perspective

In this last section, we apply the mathematical part of this paper relative to the DC vectorial optimization problem. We consider a (MPF) multiproduct firm where different production units produce different goods or products. Each component of the MPF vectorial profit is the profit of a monoproduct firm. It is the difference between its concave benefit and its concave cost of production. The needs/desires of the multiproduct firm are to grow, i.e., to improve, each period (hence progressively), his vectorial profit, i.e., the profit of each production unit. This is done through changing its vectorial level of production. This represents a specific example of the very general (multiple) needs/desires satisfaction problem considered by the recent (VR) variational rationality approach of stay and change human dynamics. See Soubeyran [52] for a recent synthetic view of this approach. For earlier references see Soubeyran [45, 46, 47, 48, 49, 50, 51]. This general problem is relative to an individual, a group, a firm, an organization, or interacting agents (games).

6.1. The vectorial profit of a multi-product firm

Consider a multiproduct firm which produces n final goods $i \in I = \{1, 2, ..., n\}$. It groups n components (individual firms) $i \in I = \{1, 2, ..., n\}$. The individual firm i produces the quantity $q_i \in \mathbb{R}_+$ of the product i, given that ρ_i is the unit cost of production of good i and p_i is the unit price of the product $i \in I$. As we will see later the price p_i, and the unit cost of production ρ_i can depend of the quantity q_i. Then, if the firm i produces q_i units of product i, its profit is $\pi_i(q_i) = p_i q_i - \rho_i q_i$. It is the difference between the benefits $b_i(q_i) = p_i q_i$ and the costs of production $c_i(q_i) = \rho_i q_i$. The vectorial profit (= net value) of the multiproduct firm is $\Pi(q) = B(q) - C(q)$, where the vectorial benefit is $B(q) = (b_i(q_i), i \in I)$ and the vectorial cost of production is $C(q) = (c_i(q_i), i \in I)$.

Remark 6.1. In the behavioral part of this paper (this last section) we consider the maximization of the vectorial profit function

$$\Pi(.) : q \in X = \mathbb{R}^n_+ \longmapsto \Pi(q) \in Z = \mathbb{R}^n.$$

In the mathematical part of the paper we considered the minimization of the vectorial objective $F(.) : q \in X = \mathbb{R}^n \longmapsto F(q) \in Z = \mathbb{R}^m$, where $F(q) = G(q) - H(q)$. The mathematical notation is $x = q$.

6.2. Monopolistic competition and economies of scale/learning by doing: concave benefits and concave costs of production

Suppose that each firm i of the multiproduct firm (MPF) is a monopoly on its own final market i where it sells the product i. Then, the price of product i is a decreasing function of the quantity produced and sold q_i. For example the inverse demand is $p_i(q_i) = \alpha_i - \beta_i q_i \geq 0$ with $0 \leq q_i \leq \overline{q}_i = \alpha_i/\beta_i$ and $\alpha_i, \beta_i \in \mathbb{R}_{++}$. Then, the benefit of the firm i is $b_i(q_i) = p_i(q_i)q_i = (\alpha_i - \beta_i q_i)q_i$. It defines a concave benefit function.

Suppose also that the unit cost of production $\rho_i(q_i)$ of firm i decreases with the quantity produced q_i. For example, $\rho_i(q_i) = \gamma_i - \delta_i q_i \geq 0$, with $0 \leq q_i \leq \overline{\overline{q}}_i = \gamma_i/\delta_i$ and $\gamma_i, \delta_i \in \mathbb{R}_{++}$. Then, the cost of production of the firm i is

$$c_i(q_i) = \rho_i(q_i)q_i = (\gamma_i - \delta_i q_i)q_i.$$

It defines a concave cost of production function. See [8].

Thus, the profit of firm i is $\pi_i(q_i) = p_i(q_i)q_i - \rho_i(q_i)q_i = [p_i(q_i) - \rho_i(q_i)]\, q_i$. That is, $\pi_i(q_i) = [(\alpha_i - \gamma_i) - (\beta_i - \delta_i)q_i]\, q_i$. A non negative profit excludes $\alpha_i - \gamma_i < 0$ and $\beta_i - \delta_i \geq 0$. The profit function $\pi_i(.) : q_i \in \mathbb{R}_+ \longmapsto \pi_i(q_i) \in \mathbb{R}$ can be concave, linear, or convex, given the signs of $\alpha_i - \gamma_i$ and $\beta_i - \delta_i$. Hence this vectorial profit includes concave, convex and linear components.

A more complex formulation can be $p_i(q_i) = \alpha_i - \beta_i q_i^{\lambda_i}$ and $\rho_i(q_i) = \gamma_i - \delta_i q_i^{\mu_i}$ with $\lambda_i, \mu_i \in \mathbb{R}_{++}$.

Another formulation could have been: $p_i(q_i) = \lambda_i q_i^{-\alpha_i}$ and $c_i(q_i) = \mu_i q_i^{-\beta_i}$. Then, $\pi_i = \lambda_i q_i^{1-\alpha_i} - \mu_i q_i^{1-\beta_i}$ is DC if $0 < \alpha_i, \beta_i < 1$. Then, $\Pi_i \geq 0$ if $\lambda_i q_i^{-\alpha_i} \geq \mu_i q_i^{-\beta_i}$, i.e., $y_i^{\beta_i - \alpha_i} \geq \mu_i/\lambda_i$. That is, q_i must be small or big depending of $\beta_i - \alpha_i < 0$, or > 0.

The literature on economies of scale presents two main circumstances leading to DC functions: economies of scale and learning by doing. Because DC optimization is so important and because this is not clearly done in the DC optimization literature, we take the time to justify its utilization in economics/management sciences.

6.2.1. Concave production costs: Economies of scale

For useful references and formalisations on this topic, see Korgaonker [32], Teece [57], Bayón et al. [8] and Koca et al. [31].

Economies of scale in production

This happens when the production cost on a per-unit basis declines as the output increases, i.e., when production at a larger scale (more output) can be achieved at a lower unit cost (i.e, with economies or savings). Hence, such a compagny does things more efficiently with increasing size (= larger quantity produced). Economies of scale can occur in the four types of transformation processes of a firm: manufacture, transport, supply, and service. For K. Carey, in Encyclopedia of Health Economics (2014), "Economies of scale refer to the notion that average cost falls as the firm expands. Conversely, diseconomies of scale occur when expansion incurs increasing average costs. From a technical standpoint, a measure of economies of scale is equivalent to the ratio of marginal to average costs. This is because if cost at the margin is lower than average cost, then average cost will fall with increased output". Usual sources of economies of scale are purchasing (bulk buying of materials through long-term contracts), managerial (increasing the specialization of managers), financial (obtaining lower-interest charges when borrowing from banks and having access to a greater range of financial instruments), marketing (spreading the cost of advertising over a greater range of output in media markets), and technological (taking advantage of returns to scale in the production function).

Multiproduct firms with economies of scale

In the multiproduct context, we can contrast two distinct economies of scale concepts, i.e., product specific economies of scale and ray scale economies. Multiproduct firms (MPF) can use common inputs in the production process of several subunits (individual firms of the MPF). Given their greater scale, they can spread their fixed costs over a greater volume of services, potentially reducing their average cost per unit. This creates barriers to entry for potential competitors that protect their profitability. They also benefit for access to many different markets and use supertankers to transport their goods. Furthermore, as seen before, by buying a large number of products at once, they could negotiate a lower price per unit than their competitors (bulk purchasing advantages).

6.2.2. Concave production costs: learning by doing.

The learning curve effect (learning by doing) is not about economies of scale. "It is a human phenomenon that occurs because of the fact that people get quicker at performing repetitive tasks once they have been doing them for a while. The first time a new process is performed, the workers are unfamiliar with it since the process is untried. As the process is repeated, however, the workers become more familiar with it and better at performing it" (see *The learning rate and learning effect*, ACCA Think Ahead, 2024). This learning by doing effect occurs when costs of doing (= costs of using means) decrease with repetition, i.e., the more we do. With repetition, as the process is learned, tasks require less time and resources, the more they are performed. The learning curve was first described by psychologist Hermann Ebbinghaus in 1885 and is used as a way to measure production efficiency and to forecast costs. That is, "the learning curve is a visual representation of how long it takes to acquire new skills or knowledge. In business, the slope of the learning curve represents the rate in which learning new skills translates into cost savings for a company. A learning curve is usually described with a percentage that identifies the rate of improvement. The steeper the slope of the learning curve, the higher the cost savings per unit of output"; see Julia Kagan, June 10, 2024 in *What is a learning curve? Formula, calculation, and example*, Investopedia.

6.3. The growth of a multi-product firm: improving its vectorial profits progressively

6.3.1. The growth of firms

It can be defined in terms of revenue generation, value addition, and expansion in terms of volume of the business in order to achieve different objectives, including increasing sales, maximising profits or increasing market share. At the theoretical level, "enterprise growth can be identified in four theoretical perspectives: the resource-based perspective, the motivation perspective, the strategic adaptation perspective and the configuration perspective"; see [16]. Furthermore, "strategies for firm growth vary in terms of their degrees of novelty, uncertainty and synergy. Modes of firm growth include replication (growth by 'more of the same'), diversification and internationalization. Growth strategies can be implemented using organic growth or through acquisitions. Desire to grow is a necessary but insufficient condition for growth – what also counts is the availability of growth opportunities"; see [17]. In this context, the growth of a monopolistic multi-product (MPF) firm is improving, progressively, the profitability of all its units of production. The MPF can, step by step, economize and gain in efficiency to approach its production possibility frontier until reaching a weak Pareto solution. In this way, the product mix changes with drops and adds and materializes in an improving way to the goal (Goldberg et al. [22]). For the list of determinants of firm growth, see Gupta et al. [25].

6.3.2. A short reminder of the VR approach in the context of the growth of a multiproduct firm

To save space, we will make this reminder as short as possible.

Improving vectorial profits decreases vectorial frustration gaps

In the context of the growth of a multiproduct firm, the (VR) variational rationality approach of stay and change human dynamics can start as follows. For a general panorama, see Soubeyran [52], and for other references see

`https://sites.google.com/view/antoine-soubeyran`.

Suppose that, a given period, the MPF is able to produce the vectorial quantity $q = (q_i, i \in I)$ of the different final goods $i \in I$ and choose to produce q. Then, within the period, it earns the vectorial profit $\Pi(q) = (\pi_i(q_i), i \in I)$. Suppose also that, at the end of the period, it compares this realized vectorial profit with a desired vectorial profit (= vectorial goal standard) Π^*. In consequence, if $\Pi(q^k) < \Pi^*$ is lower than Π^*, the vectorial frustration gap of the MPF will be $F(q^k) = \Pi^* - \Pi(q^k) > 0$. Thus, in the next period, the producer can want to improve his vectorial profit from $\Pi(q^k)$ to $\Pi(q^{k+1}) \geq \Pi(q^k)$ to reduce his vectorial frustration gap from $F(q^k)$ to $F(q^{k+1}) = \Pi^* - \Pi(q^{k+1}) < F(q^k)$.

The payoff (net value) of a move

To reduce this frustration gap requires to do a move $m = (q^k, q^{k+1})$ that goes from having produced the vectorial quantity q^k, to end in producing the vectorial quantity q^{k+1}. This move must be broken in two different stages,

(i) first stage (a translocation) : if we are not able to produce the vectorial quantity q^{k+1}, we must move from having produced the vectorial quantity q^k, to end in becoming able to produce the vectorial quantity $q^{k+1} \neq q^k$ and,

(ii) second stage (execution): being able, now, to effectively produce the vectorial quantity q^{k+1}, produce it.

Both stages are costly. A main lesson of the VR approach is that the vectorial translocation cost $\mathbb{C}(q^k, q^{k+1}) \in X = \mathbb{R}^n_+$ differs radically from the vectorial execution cost $C(q^{k+1}) \in X = \mathbb{R}^n_+$, which is the second component of the vectorial profit $\Pi(q^{k+1}) = B(q^{k+1}) - C(q^{k+1})$. Then, the vectorial net value of doing the move $m = (q^k, q^{k+1})$, is $P(q^k, q^{k+1}) = \Pi(q^{k+1}) - \xi_{k+1}\mathbb{C}(q^k, q^{k+1}) \in \mathbb{R}^n$, where $\xi_{k+1} \in \mathbb{R}_{++}$ is the weight put on the vectorial translocation cost $\mathbb{C}(q^k, q^{k+1})$.

Examples of translocation costs

For example, for each product i, from one period to the next,

- the cost of becoming able to produce more than before of the product i is, $\mathbb{C}_i(q_i^k, q_i^{k+1}) = \mathbb{C}_i(q_i^k, q_i^k) + \rho_i^+(q_i^{k+1} - q_i^k)$ if the add is $q_i^{k+1} - q_i^k \geq 0$;
- the cost of becoming able to produce less than before of the product i is, $\mathbb{C}_i(q_i^k, q_i^{k+1}) = \mathbb{C}_i(q_i^{k+1}, q_i^{k+1}) + \rho_i^{\searrow}(q_i^k - q_i^{k+1})$ if the drop is $q_i^k - q_i^{k+1} > 0$,

where $\rho_i^+(\rho_i^{\searrow}) \in \mathbb{R}_{++}$ is the unit cost of becoming able to produce more (less) of the product i. Then, the net value of the move $m = (q^k, q^{k+1})$ is

$$P(q^k, q^{k+1}) = \Pi(q^{k+1}) - \xi_{k+1}\mathbb{C}(q^k, q^{k+1}) \in \mathbb{R}^n.$$

The net value of the stay $\sigma = (q^k, q^k)$ is $P(q^k, q^k) = \Pi(q^k) - \xi_{k+1}\mathbb{C}(q^k, q^k) \in \mathbb{R}^n$.

A worthwhile move (change rather than stay)

The MPF will want to improve the net value of the MPF when moving from producing q^k to produce q^{k+1}. That is, if $P(q^k, q^{k+1}) \geq P(q^k, q^k)$. This is equivalent to the vectorial inequality

$$\Pi(q^{k+1}) - \Pi(q^k) \geq \xi_{k+1}\left[\mathbb{C}(q^k, q^{k+1}) - \mathbb{C}(q^k, q^k)\right].$$

This inequality defines a worthwhile move $m = (q^k, q^{k+1})$, where

$$A(q^{k+1}/q^k) = \Pi(q^{k+1}) - \Pi(q^k) \quad \text{and} \quad I(q^{k+1}/q^k) = \mathbb{C}(q^k, q^{k+1}) - \mathbb{C}(q^k, q^k)$$

represent vectorial advantages and inconveniences to do the change $m = (q^k, q^{k+1})$ rather than the stay $\sigma = (q^k, q^k)$.

This worthwhile improving process satisfies

$$q^{k+1} \in \Omega_k = \Omega(q^k) = \left\{q \in X,\ \Pi(q) \geq \Pi(q^k)\right\}.$$

That is, we have the succession of profit improvements

$$\Pi(q^{k+1}) \geq \Pi(q^k) \geq .... \geq \Pi(q^0) \geq 0$$

where we suppose (a viability condition) that $\Pi(q^0) \geq 0$. In this way, the MPF can hope, at the end, to reach a Pareto solution where it cannot increase the profit of one sub-unit firm without lowering the profit of another sub-unit firm. This process is a way to get efficiency gains.

6.3.3. Proximal vectorial optimization: mathematical perturbations as approximative costs to move

One of the main finding of the (VR) variational rationality approach of stay and change human dynamics is that the perturbation term of an optimization algorithm can be seen as a cost to move. Let us see that, as expected, it can be the same in the vectorial context.

From the minimization to the maximization of a vectorial payoff to move

The Bonnel et al. [14] multiobjectives (vectorial) proximal minimization method is

$$q^{k+1} \in \arg\min \left\{F(q) + (\lambda_k/2)\left\|q - q^k\right\|^2 \varepsilon^k, q \in \Omega_k\right\},$$

where $\Omega_k = \left\{q \in X = \mathbb{R}^n,\ F(q) \leq F(q^k)\right\}$ and $F(.) : \mathbb{R}^n \longmapsto \mathbb{R}^m$ is -$\mathbb{R}^m_+$ convex, and $\varepsilon^k \in \mathbb{R}^m_{++}$.

As seen before, in the context of the VR approach, if a MPF produces the vectorial quantities q^k, the vector $F(q^k) = \Pi^* - \Pi(q^k) > 0$, represents the vectorial frustration gap of this MPF. Thus, the equality

$$F(q) + (\lambda_k/2)\left\|q - q^k\right\|^2 \varepsilon = \Pi^* - \left[\Pi(q^k) - (\lambda_k/2)\left\|q - q^k\right\|^2 \varepsilon^k\right] = \Pi^* - P(q^k, q)$$

shows that minimizing the frustration payoff $\Phi(q^k, q) = F(q) + (\lambda_k/2)\left\|q - q^k\right\|^2 \varepsilon^k$ of the move $m = (q^k, q)$ is equivalent to maximize the vectorial payoff

$$P(q^k, q) = \Pi(q^k) - (\lambda_k/2)\left\|q - q^k\right\|^2$$

of the move m.

Then, the vectorial Bonnel et al. [14] minimization problem is equivalent to the maximization of the vectorial payoff of the MPF,

$$q^{k+1} \in \arg\max \left\{\Pi(q) - (\lambda_k/2)\left\|q - q^k\right\|^2 \varepsilon^k, q \in \Omega_k\right\}.$$

Approximative costs to do a move

Consider the previous behavioral costs to move, where, the cost to become able to produce more than before is $\mathbb{C}_i(q_i^k, q_i^{k+1}) = \mathbb{C}_i(q_i^k, q_i^k) + \rho_i^+(q_i^{k+1} - q_i^k)$ if the add is $q_i^{k+1} - q_i^k \geq 0$, while the cost to become able to produce less than before is $\mathbb{C}_i(q_i^k, q_i^{k+1}) = \mathbb{C}_i(q_i^{k+1}, q_i^{k+1}) + \rho_i^{\searrow}(q_i^k - q_i^{k+1})$ if the drop is $q_i^{k+1} - q_i^k < 0$.

In contrast, the Bonnel et al. [14] vectorial perturbation term is

$$\mathbb{C}(q^k, q) = \mathbb{C}(q, q) + \left\|q - q^k\right\|^2 \varepsilon^k = \left\|q - q^k\right\|^2 \varepsilon^k,$$

with $\mathbb{C}(q, q) = 0$ for all $q \in X$. That is, we have $\mathbb{C}_i(q_i^k, q_i) = \varepsilon_i^k (q_i - q_i^k)^2$. Then, if, for example, $\rho_i^{\succ} \leq \rho_i^+$, take as an approximation, $\rho_i^{\succ} \leq \varepsilon_i^k \leq \rho_i^+$.

Finding: A behavioral reinterpretation of the perturbation terms in optimization theory

Our previous observations show that the mathematical perturbation term $\|q–q^k\|^2\varepsilon^k$ of a vectorial optimization program can be seen as an approximative vectorial behavioral costs to move. This is one of the main early finding of the VR approach (Soubeyran [45, 46]): in an optimizing algorithm, adding a perturbation term is an instance of modeling costs to move in psychology/behavioral sciences. This shows the power of the VR approach to provide a strong connexion between variational analysis/optimization theory and a unifying VR mathematical modeling of psychology/behavioral sciences.

6.4. Mathematical acceleration processes: mathematical couplings

Accelerating the growth of a MPF is a major topic in management sciences and economics. Thus, the consideration of acceleration methods in optimization theory seem very attractive for these two disciplines. Thanks to the VR approach, a close link can be done between mathematical acceleration methods (mathematical couplings) and behavioral acceleration methods (behavioral couplings) revealed for the first time by the VR approach in behavioral sciences (Soubeyran [52]).

Acceleration methods in optimization theory are mainly coupling methods

In optimization theory acceleration methods consider, each period, a succession of two steps (a coupling), where their combinaison helps to decrease enough the objective function. Their choice changes with the properties of the objective function F or Π. That is, when, each period, in a not accelerated optimization method, a step decreases not enough the objective, add an other step to decrease it more. This is a way to divide (the difficulty) to conquer (solve the problem) more quickly. Making two small improving steps each period can be better than doing one. However, in contrast to the gradient algorithm ([21]) this can lead to transient overshooting problems and non monotone dynamics.

Let us list the main couplings, initiated by Polyak and Nesterov.

(1) The famous Polyak's heavy ball method [43] couples, each period, a gradient step before a momentum step when the objective function is smooth.

(2) The celebrated Nesterov's accelerated gradient descent method [38] couples a gradient step after a momentum step. This helps to limit overshooting. Nesterov gave two other variants of this coupling [40]. In his different publications, he supposed that the objective function can be smooth, strongly convex, both smooth and strongly convex, convex, differentiable or not, ... See also [6]. Other coupling methods has been given to better understand these two, a priori, rather mysterious coupling methods, essentially to understand, (i) what is

the adhoc momentum step [55] and, (ii) what is the intuition behind the method of estimate sequences [42] that drives the proof presented in [38, 40]. Notice that the non differentiability aspect that is present in almost any application makes the famous Nesterov acceleration method even more complicated. See [39].

Let us cite, for example, two papers that try to give a more intuitive presentation of the Nesterov approach.

(3) [2] primal-dual approach couples a gradient step with a mirror descent step. The mirror descent step remaining a bit mysterious with the use of a Bregman distance; even if a mirror step can be seen as a dual averaging step in the dual space.

(4) [1] recent coupling method combines two proximal steps using each period, a lower bound and an upper bound of the objective function. These two steps remaining however a bit mysterious with their use of looking ahead evaluations of the objective function.

Other interpretations of acceleration

They refer to,

(5) Geometric interpretations [15, 18] that use quadratic balls approximation functions to approximate (as lower and upper bounds) the objective function.

(6) A dynamical system perspective [33, 54, 58].

(7) A control view, i.e., a dissipative system approach [7, 27, 34, 59] building and using elegant potential functions (inspired by different energy functions in physic). What make these potential functions adhoc is the mysterious choice of the way to weight potential and kinetic energy functions.

(8) A numerical optimization perspective that accelerates a sequence of iterates in numerical optimization of differential equations [44]. In this context optimization methods can be seen as instances of integration methods.

Acceleration methods for DC functions

In this context where the objective is the difference between two convex functions,

- [41] couples, each period, a Nesterov's momentum step with a DC step that minimizes a convex approximation of the objective function which is equal to the difference between the first convex component and a linear approximation of the second component.

- [4, 5] coupled, each period, a DC step with an Armijo step.

- [35] minimizes a quadratic approximation of the DC objective function, using for that purpose a Nesterov's acceleration process with two steps: a momentum step followed by a gradient step (FISTA, see [9]). That is, the authors apply a Nesterov acceleration method to the minimization of a quadratic approximation of the DC objective.

6.5. When a mathematical DC acceleration coupling is an approximate behavioral DC acceleration coupling

As seen before, acceleration methods in optimization use different couplings to speed up the convergence of an algorithm. The VR approach has recently shown the very new result that the same is true in psychology. In this coupling context, the application part of this paper shows how the DC acceleration coupling proposed in the mathematical part of this paper models a specific instance of a behavioral multiobjective DC acceleration coupling where asymmetric costs to move (asymmetric distances) are modeled by a symmetric perturbation function (the symmetric Euclidian distance).

6.5.1. VR behavioral acceleration couplings

The VR approach of psychology/behavioral sciences recently highlights the importance of behavioral couplings to accelerate the satisfaction of multiple needs/desires. More precisely, Soubeyran [52],

(i) showed that to satisfy our needs/desires we must move from having done/ stop to do something (a task, an activity, a bundle of activities) to become able to do, start to do and do the same, and/or, other things that will help to satisfy these recurrent and changing needs/desires;

(ii) defined two main basic moves, worthwhile and satisficing moves, where,

(a) a move is worthwhile if it is improving and not too costly (see above a formalization for the multiproduct firm). That is, if advantages to move are high enough compared to inconveniences to move;

(b) a move is satisficing if it improves enough, i.e., makes enough progress – makes a sufficient ascent, i.e., goes fast enough.

(iii) showed on an example (loosing weight) as well as in the context of a gradient ascent, how the co-existence or succession of these two kinds of moves can lead to the satisfaction of our needs/desires. This is a striking result that highlights a strong correspondance between the main principles that drive psychology and a part of behavioral sciences, and variational analysis/optimizing algorithms. This paves the way to "algorithmic psychology".

More precisely, a main finding of the VR approach is to exhibit a close relationships between behavioral couplings and optimizing couplings in relation to the acceleration of optimizing algorithms. In this context a proximal step is an example of a worthwhile move and an Armijo step is an instance of a satisficing move, because it makes enough progress.

The principle of acceleration is this: each stage of the algorithm (period) decomposes the move $m = (x^k, x^{k+1})$ in two steps: $x^k \curvearrowright y \curvearrowright x^{k+1}$. For a multiproduct firm, it decomposes the move $m = (q^k, q^{k+1}) = (x^k, x^{k+1})$ in two steps: $(q^k, q) = (x^k, y)$ and $(q, q^{k+1}) = (y, x^{k+1})$, where the mathematical notations for the successive steps are $x^k \curvearrowright y \curvearrowright x^{k+1}$ and the behavioral notations are

$$q^k = x^k \curvearrowright q = y \curvearrowright q^{k+1} = x^{k+1}.$$

6.5.2. This paper: Building an approximate worthwhile step and a sufficient ascent step

The present paper couples, each period, an approximate vectorial proximal step (hence, a vectorial worthwhile step) with a sufficient descent or ascent vectorial Armijo step. This works as follow.

Build an approximate worthwhile first step

In this first stage $(q^k, q) = (x^k, y)$ of a move $m = (q^k, q^{k+1}) = (x^k, x^{k+1})$, we approximate the payoff $P(x^k, y) = \Pi(y) - (\lambda_k/2)\mathbb{C}(x^k, y)$ to do the step (x^k, y). More precisely, to get an approximate worthwhile first step, the mathematical part of this paper:

(A) approximates the DC profit $\Pi(y) = B(y) - C(y)$ around x^k by the following expression: $\Pi_k(y) = B(y) - \left[C(x^k) + JC(x^k)(y - x^k)\right]$. That is, it leaves unchanged the benefit function $B(y)$ and approximates around x^k the cost $C(y)$ to produce y through the linearization $C(y) = C(x^k) + JC(x^k)(y - x^k)$. In the mathematical part of the paper the constant term $C(x^k)$ is hidden when the problem is to choose y in an approximate optimal way.

(B) approximates the costs $\mathbb{C}(x^k, y)$ by the following formulation:
$\mathbb{C}_k(x^k, y) = \mathbb{C}(x^k, x^k) + (\lambda_k/2)\left\|y - x^k\right\|^2 \varepsilon^k$.
That is, if $\mathbb{C}(x^k, x^k) = 0$, we get $\mathbb{C}_k(x^k, y) = (\lambda_k/2)\left\|y - x^k\right\|^2 \varepsilon^k$.

(C) and finally approximates the behavioral payoff
$$P(x^k, y) = \Pi(y) - (\lambda_k/2)\mathbb{C}(x^k, y)$$
derived from doing the move (x^k, y) by the following approximate payoff:
$$P_k(x^k, y) = \Pi_k(y) - (\lambda_k/2)\mathbb{C}_k(x^k, y) = \Pi_k(y) - (\lambda_k/2)\left\|y - x^k\right\|^2 \varepsilon^k.$$

The resolution of this first stage provides an approximate worthwhile move $(q^k, q) = (x^k, y)$.

Build a satisficing (improving enough) step

Let $y^k = x^k + d^k$ be the end of the first step. The end of the second step of the move $m = (x^k, x^{k+1})$ is $x^{k+1} = y^k + t_k d^k$. An improvement $\Pi(y^k + t_k d^k) - \Pi(y^k)$ of the vectorial profit is sufficient if it exists $t_k > 0$ and $\beta \in]0, 1[$ such that

$\Pi(y^k + t_k d^k) - \Pi(y^k) \geq \beta t_k JC(x^k) d^k$. This vectorial Armijo step represents an instance of a satisficing step (y^k, x^{k+1}) put in front by the VR approach.

6.5.3. Interpretations of the results

The accelerated DC algorithm performs a succession of worthwhile moves

Consider the vectorial profit Π instead of the vectorial frustration F. Given that, each period, the MPF improves its vectorial profit, i.e., $\Pi(x^{k+1}) - \Pi(x^k) \geq 0$ for all k, we have

$$\begin{aligned}\left\|\Pi(x^{k+1}) - \Pi(x^k)\right\|_1 &= \min_{1\leq i\leq m} \left\{ \left|\Pi_i(x^{k+1}) - \Pi_i(x^k)\right|\right\} \\ &= \min_{1\leq i\leq m} \left\{ \Pi_i(x^{k+1}) - \Pi_i(x^k)\right\}.\end{aligned}$$

Then, Proposition 4.3 shows that $\left\|\Pi(x^{k+1})-\Pi(x^k)\right\|_1 \geq (\beta/\lambda_2)t_k \left\|y^k-x^k\right\|^2$ (see the proof).

Given that $x^{k+1}=y^k+t_k(y^k-x^k)=x^k+(y^k-x^k)+t_k(y^k-x^k)=x^k+(1+t_k)(y^k-x^k)$ we get

$y^k-x^k=(x^{k+1}-x^k)/(1+t_k)$. This gives

$$\left\|\Pi(x^{k+1})-\Pi(x^k)\right\|_1 \geq \mu_k \left\|x^{k+1}-x^k\right\|^2,$$

with $\mu_k=(\beta/\lambda_2)t_k/(1+t_k)^2$. Then, coming back to the terminology of the VR approach (to go straight, see Soubeyran, 2024), if $A(q^{k+1}/q^k)=\Pi(q^{k+1})-\Pi(q^k)$ and $I(q^{k+1}/q^k)=\mathbb{C}(q^k,q^{k+1})-\mathbb{C}(q^k,q^k)$ represent vectorial advantages and inconveniences to do the change $m=(q^k,q^{k+1})$ rather than to do the stay $\sigma=(q^k,q^k)$, the scalar overall motivation and resistance to move are

$$M(q^{k+1}/q^k)=U\left[A(q^{k+1}/q^k)\right]=\left\|\Pi(x^{k+1})-\Pi(x^k)\right\|_1$$

and $\quad R(q^{k+1}/q^k)=D\left[I(q^{k+1}/q^k)\right]=\left[\mathbb{C}(q^k,q^{k+1})-\mathbb{C}(q^k,q^k)\right]^2=\mathbb{C}(q^k,q^{k+1})^2$

if $\mathbb{C}(q^k,q^k)=0$. Then, the expression $\left\|\Pi(x^{k+1})-\Pi(x^k)\right\|_1 \geq \mu_k \left\|x^{k+1}-x^k\right\|^2$ is $M(q^{k+1}/q^k)\geq \mu_k R(q^{k+1}/q^k)$. This shows that the move $m=(x^k,x^{k+1})$ is a worthwhile move because motivation to move is high enough with respect to resistance to move. Hence the striking interdisciplinary mathematical and behavioral result.

No weights: protecting the poor

Our acceleration algorithm uses no weight to define the importance given to the profit of each individual firm i of the MPF. In contrast, it considers the scalarization function

$$\max_{1\leq i\leq m}\left\{\ \left|\Pi_i(x^{k+1})-\Pi_i(x^k)\right|\right\}=\max_{1\leq i\leq m}\left\{\ \Pi_i(x^{k+1})-\Pi_i(x^k)\right\}.$$

This shows that the MPF protects the profit of its least profitable individual subunit firm. This is a specific example, within a group of individuals, of the desire to protect the poors of this group. It is a way to benefit from a strong diversification effect.

Acknowledgments. The authors are grateful to the the Editors allowing to celebrate Hedy Attouch memory. In particular, his enthusiasm and help when Antoine proposed to him to work on one of the initial ideas of the VR approach: to introduce costs to move as perturbation terms in optimizing algorithms to make a link between psychology/social sciences and variational analysis.

References

[1] K. Ahn, S. Sra: *Understanding Nesterov's acceleration via proximal point method*, in: Proc. 5th SIAM Symposium on Simplicity in Algorithms (SOSA January 2022), K. Bringmann et al. (eds.), SIAM, Philadelphia (2022) 117–130.

[2] Z. Allen-Zhu, L. Orecchia: *Linear coupling: An ultimate unification of gradient and mirror descent*, in: *8th Innovations in Theoretical Computer Science Conference* (ITCS 2017), C. H. Papadimitriou et al. (eds.), Leibniz Zentrum für Informatik, Proceedings in Informatics 67 (2017), art.no. 3, 22p.

[3] Y. T. Almeida, J. X. Cruz Neto, P. R. Oliveira, J. C. O. Souza: *A modified proximal point method for DC functions on Hadamard manifolds*, Comput. Optim. Appl.76 (2020) 649–673.

[4] F. J. Aragon Artacho, R. M. T. Fleming, P. T. Vuong: *Accelerating the DC algorithm for smooth functions*, Math. Programming, Ser. B 169 (2018) 95–118.

[5] F. J. Aragon Artacho, P. T. Vuong: *The boosted difference of convex functions algorithm for nonsmooth functions*, SIAM J. Optim. 30 (2020) 980–1006.

[6] A. Auslender, M. Teboulle: *Interior gradient and proximal methods for convex and conic optimization*, SIAM J. Optim. 16/3 (2006) 697–725.

[7] N. Bansal, A. Gupta: *Potential-function proofs for gradient methods*, Theory Comput. 15/4 (2019) 1–32.

[8] L. Bayón, J. A. Otero, M. M. Ruiz, P. M. Suárez, C. Tasis: *The profit maximization problem in economies of scale*, J. Comput. Appl. Math. 236/12 (2012) 3065–3072.

[9] A. Beck, M. Teboulle: *A fast iterative shrinkage-thresholding algorithm*, SIAM J. Imaging Sci. 2/1 (2009) 183–202.

[10] G. C. Bento, S. D. P. Bitar, J. X. Cruz Neto, A. Soubeyran, J. C. D. Souza: *A proximal point method for difference of convex functions in multi-objective optimization with application to group dynamic problems*, Comput. Optim. Appl. 75 (2020) 263–290.

[11] G. C. Bento, J. X. Cruz Neto, G. López, A. Soubeyran, J. C. O. Souza: *The proximal point method for locally Lipschitz functions in multi-objective optimization with application to the compromise problem*, SIAM J. Optim. 28/2 (2018) 1104–1120.

[12] G. C. Bento, J. X. Cruz Neto, P. R. Oliveira, A. Soubeyran: *The self regulation problem as an inexact steepest descent method for multicriteria optimization*, Eur. J. Oper. Res. 235 (2014) 494–502.

[13] J. Bolte, A. Danilidis, A. Lewis, M. Shiota: *Clarke critical values of subanalytic Lipschitz continuous functions*, Ann. Polon. Math. 87 (2005) 13–25.

[14] H. Bonnel, A. N. Iusem, B. F. Svaiter: *Proximal methods in vector optimization*, SIAM J. Optim. 15 (2005) 953–970.

[15] J. S. Bubeck, Y. Tat Lee, M. Singh: *A geometric alternative to Nesterov's accelerated gradient descent*, arXiv: 1506.08187 (2015).

[16] A. Coad: *The Growth of Firms: A Survey of Theories and Empirical Evidence*, New Perspectives on the Modern Corporation, Elgar, Cheltenham (2009).

[17] A. Coad: *Strategies for firm growth*, in: *Palgrave Encyclopedia of Strategic Management*, M. Augier et al. (eds.), Palgrave Macmillan, London (2016), 6p.

[18] D. Drusvyatskiy, M. Fazel, S. Roy: *An optimal first order method based on optimal quadratic averaging*, SIAM J. Optim. 28/1 (2018) 251–271.

[19] J. Fliege, I. F. Vaz, L. N. Vicente: *Complexity of gradient descent for multi-objective optimization*, Optim. Methods Software 34/5 (2019) 949–959.

[20] E. H. Fukuda, L. M. G. Drummond: *A survey on multiobjective descent methods*, Pesquisa Operacional 34/3 (2014) 585–620.

[21] A. Garivier: *Machine learning 9: Convex optimization for machine learning*, Master 2 Computer Science, Institut de Mathématiques, Toulouse (2019).

[22] P. K. Goldberg, A. K. Khandelwal, N. Pavcnik, P. Topalova: *Multiproduct firms and product turnover in the developing world: Evidence from India*, Rev. Econ. Stat. 92/4 (2010) 1042–1049.

[23] S. M. Grad: *A survey on proximal point type algorithms for solving vector optimization problems*, in: *Splitting Algorithms, Modern Operator Theory, and Applications*, Workshop at Oaxaca, Mexico, September 2017, H. Bauschke et al. (eds.), Springer, Cham (2019) 269–308.

[24] L. M. Graña Drummond, B. F. Svaiter: *A steepest descent method for vector optimization*, J. Comput. Appl. Math. 175 (2005) 395–414.

[25] P. D. Gupta, S. Samapti Guha, S. S. Krishnaswami: *Firm growth and its determinants*, J. Innovation Entrepreneurship 2 (2013), art.no. 15.

[26] J. B. Hiriart-Urruty: *Generalized differentiabity, duality and optimization for problems dealing with difference of convex functions*, in: *Convexity and Duality in Optimization*, J. Ponstein (ed.), Lecture Notes in Economics and Mathematical Systems 256 (1986) 37–70.

[27] B. Hu, L. Lessard: *Dissipativity theory for Nesterov's accelerated method*, in: Proc. 34th Int. Conf. Machine Learning, Vol. 70, (2017) 1549–1557.

[28] X. X. Huang, X. Q. Yang: *Duality for multi-objective optimization via nonlinear Lagrangian functions*, J. Optim. Theory Appl. 120 (2004) 111–112.

[29] J. Jahn: *Vector Optimization: Theory, Applications and Extensions*, Springer, Berlin (2004).

[30] Y. Ji, M. Goh, R. De Souza: *Proximal point algorithms for multi-criteria optimization with the difference of convex objective functions*, J. Optim. Theory Appl. 169 (2016) 280–289.

[31] E. Koca, Y. Yaman, S. Akturk: *Lot sizing with piecewise concave production costs*, Dept. Industrial Engineering, Bilkent University, Ankara (2013).

[32] M. G. Korgaonker: *Production smoothing under piecewise concave costs, capacity constraints and nondecreasing requirements*, Management Sci. 24/3 (1977) 302–311.

[33] M. Laborde, A. Oberman: *A Lyapunov analysis for accelerated gradient methods: from deterministic to stochastic case*, in: Proc. 23rd Int. Conf. Artificial Intelligence and Statistics (AISTATS 2020), Palermo, Italy 2020, PMLR Vol. 108 (2020) 10p.

[34] L. Lessard, B. Recht, A. Packard: *Analysis and design of optimization algorithms via integral quadratic constraints*, SIAM J. Optimization 26/1 (2016) 57–95.

[35] T. Liu, A. Takeda: *An inexact successive quadratic approximation method for a class of difference of convex optimization problems*, Comput. Optim. Appl. 82 (2022) 141–173.

[36] D. T. Luc: *Theory of Vector Optimization*, Lecture Notes in Economics and Mathematical Systems 319, Springer, Berlin (1989).

[37] M. Minami: *Weak Pareto-optimal necessary conditions in a nondifferentiable multiobjective program on a Banach space*, J. Optim. Theory Appl. 41 (1983) 451–461.

[38] Y. Nesterov: *A method for solving a convex programming problem with convergence rate* $O(1/K^2)$, Sov. Math. Dokl. 27 (1983) 372–367.

[39] Y. Nesterov: *Smooth minimization of non-smooth functions*, Math. Programming, Ser. A 103 (2005) 127–152.

[40] Y. Nesterov: *Lectures on Convex Optimization*, 2nd ed., Springer Optimization and Its Applications Vol. 137, Springer, Cham (2018).

[41] D. N. Phan, H. Minh Le, H. A. Le Thi: *Accelerated difference of convex functions algorithm and its application to sparse binary logistic regression*, in: Proc. 27th Int. Joint Conf. Artificial Intelligence (IJCAI-18), (2018) 1369–1375.

[42] S. Pokutta: *Cheat sheet: Acceleration from first principles*, www.pokutta.com (2019).

[43] B. T. Polyak: *Some methods of speeding up the convergence of iteration methods*, USSR Comput. Math. Math. Phys. 4/5 (1964) 1–17.

[44] A. Scieur: *Acceleration in Optimization. Optimization and Control*, Université Paris Sciences et Lettres (2018).

[45] A. Soubeyran: *Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors*, GREQAM, Aix-Marseille University, preprint (2009).

[46] A. Soubeyran: *Variational rationality and the unsatisfied man: routines and the course pursuit between aspirations, capabilities and beliefs*, GREQAM, Aix-Marseille University, preprint (2010).

[47] A. Soubeyran: *Variational rationality: towards a grand theory of motivation driven by worthwhile moves*, AMSE, Aix-Marseille University, preprint (2021).

[48] A. Soubeyran: *Variational rationality: the concepts of motivation and motivational force*, AMSE, Aix-Marseille University, preprint (2021).

[49] A. Soubeyran: *Variational rationality. The resolution of goal conficts via stop and go approach-avoidance dynamics*, AMSE, Aix-Marseille University, preprint (2021)

[50] A. Soubeyran: *Variational rationality. A general theory of moving goals and intentions as satisficing worthwhile moves*, AMSE, Aix-Marseille University, preprint (2021).

[51] A. Soubeyran: *Variational rationality. Self regulation success as a succession of worthwhile moves that make sufficient progress*, AMSE, Aix-Marseille University, working paper, HAL deposit (2022).

[52] A. Soubeyran: *Variational rationality: finding cheap and quick enough ways of motion to satisfice our recurrent and changing needs/desires*, AMSE, Aix-Marseille University, working paper, HAL deposit (2024).

[53] J. C. O. Souza, J. X. Cruz Neto, J. O. Lopes, R. C. M. Silva, S. D. B. Bitar: *The proximal point method with a vectorial Bregman regularization in multiobjective DC programming*, Optimization 71/1 (2020) 263–284.

[54] W. Su, S. Boyd, E. J. Candes: *A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights*, J. Machine Learning Res. 17 (2016), art.no. 153, 43p.

[55] I. Sutskever, J. Martens, G. Dahl, G. Hinton: *On the importance of initialization and momentum in deep learning*, in: Proc. 30th Int. Conf. Machine Learning, PMLR 28/3 (2013) 1139–1147.

[56] P. D. Tao, E. B. Souad: *Algorithms for solving a class of nonconvex optimization problems. Methods of subgradients*, in: *Fermat Days 85: Mathematics for Optimization*, Toulouse 1985, North-Holland Math. Stud. 129 (1986) 249–271.

[57] D. J. Teece: *Toward an economic theory of the multiproduct firm*, in: *Economics Meets Sociology in Strategic Management*, J. A. C. Baum et al. (eds.), Advances in Strategic Management Vol. 17, Emerald Group, Bingley (2000) 29–53.

[58] A. Wibisono, A. C. Wilson, M. I. Jordan: *A variational perspective on accelerated methods in optimization*, Proc. Nat. Acad. Sci. USA 113/47 (2016) E7351–E7358.

[59] A. C. Wilson, B. Recht, M. I. Jordan: *A Lyapunov analysis of momentum methods in optimization*, Mach. Learn. Res. 22 (2021), art.no. 113, 34p.