On the multiplier rules

Joël Blot

ON THE MULTIPLIER RULES

JOËL BLOT

Abstract. We establish new results of first-order necessary conditions of optimality for finite-dimensional problems with inequality constraints and for problems with equality and inequality constraints, in the form of John's theorem and in the form of Karush-Kuhn-Tucker's theorem. In comparison with existing results we weaken assumptions of continuity and of differentiability.

Key Words: Multiplier rule, Karush-Kuhn-Tucker theorem.
M.S.C. 2010: 90C30, 49K99.

1. Introduction

We consider first-order necessary conditions of optimality for finite-dimensional problems under inequality constraints and under equality and inequality constraints.

Let Ω be a nonempty open subset of $\mathbb{R}^n$, let $f_i : \Omega \to \mathbb{R}$ (when $i \in \{0, ..., m\}$) be functions, let $\phi : \Omega \to \mathbb{R}$, $g_i : \Omega \to \mathbb{R}$ (when $i \in \{1, ..., p\}$) and $h_j : \Omega \to \mathbb{R}$ (when $j \in \{1, ..., q\}$) be functions. With these elements, we build the two following problems:

$$(\mathcal{I}) \begin{cases} \text{Maximize} & f_0(x) \\ \text{when} & x \in \Omega \\ \text{and when} & \forall i \in \{1, ..., m\}, f_i(x) \geq 0, \end{cases}$$

and

$$(\mathcal{M}) \begin{cases} \text{Maximize} & \phi(x) \\ \text{when} & x \in \Omega \\ \text{when} & \forall i \in \{1, ..., p\}, g_i(x) \geq 0 \\ \text{and when} & \forall j \in \{1, ..., q\}, h_j(x) = 0. \end{cases}$$

We provide necessary conditions of optimality under the form of Fritz John's conditions and under the form of Karush-Kuhn-Tucker's conditions. Our aim is to weaken the assumptions which permit to obtain such results. We can delete certain conditions of continuity and we can replace certain conditions of Fréchet-differentiability by conditions of Gâteaux-differentiability.

The Farkas-Minkowski Theorem is one of the main tools that we use to establish the result for the problem $(\mathcal{I})$ with inequality constraints. A (local) theorem of implicit function permits to transform (locally) a problem with equality and inequality constraints (like $(\mathcal{M})$) into a problem with only inequality constraints (like $(\mathcal{I})$); it is why the Implicit Function Theorem of Halkin is one of the main tools to establish our result for $(\mathcal{M})$.

These results are usual when we assume that all the functions are continuously Fréchet-differentiable on a neighborhood of $\hat{x}$, ([1], Chapter 3, Scetion 3.2)), ([13],

Date: September 3, 2014.

Chapitre 13, Section 2), [9], [3], [14], [11], ([17], Chapter 11). In [8], Halkin gives a multiplier rule only using the continuity on a neighborhood of $\hat{x}$ and the Fréchet-differentiability at $\hat{x}$ of the functions. His proof uses his implicit function theorem (Theorem 2.2). In ([12], Chapter 24, Section 24.7) Michel gives another proof of the result of Halkin without to use an implicit function theorem but nevertheless using the Fixed Point Theorem of Brouwer. The proof of Michel is also explained in ([4], Appendix B). In [15] we find a result for $(\mathcal{I})$ with only the Fréchet-differentiability of the functions f_i at $\hat{x}$.

There exist several works on the multiplier rules for locally Lipschitzian functions which are obtained by using the Clarke Calculus [6]. After a famous theorem of Rademacher on the Lebesgue-almost everywhere Fréchet-differentiability of a locally Lipschtzian mapping, and since the Clarke-gradient is a upper semi-continuous correspondence, we can say that the locally Lipschitzian generalize the continuously Fréchet-differentiable mappings. Note that a mapping which is only Fréchet-differentiable (even all over a naighborhood of a point) is not necessarily locally Lipschtzian and a locally Lipschitzian mapping is not necessarily Fréchet-differentiable at a given point. And so there exist two different ways for the generalisation of the multiplier rules of the continuously differentiable setting: the locally Lipschitzian setting, and the (only) Fréchet-differentiable (or differentiability in a weaker sense than this one of Fréchet) setting. Our paper belongs to the second way.

Now we briefly describe the contents of the paper. In Section 2 we precise our notation and we recall two important tools. In Section 3 we state the new results for $(\mathcal{I})$ and for $(\mathcal{M})$. In Section 4 we prove the theorem of necessary condition of optimality for $(\mathcal{I})$, and in Section 5, we prove the theorem of necessary condition of optimality for $(\mathcal{M})$.

2. Notation and recall

First we precise the used notions of differentiability. Let E and F be two real normed spaces, let Ω be a nonempty open subset of E, $f : \Omega \to F$ be a mapping and let $x \in \Omega$ and $v \in E$. When it exists, the directional derivative of f at x in the direction of v is $\vec{D}f(x;v) := \frac{d}{dt}_{|t=0} f(x+tv)$. When $\vec{D}f(x;v)$ exists for all $v \in E$ and when $[v \mapsto \vec{D}f(x;v)]$ is linear continuous, we say that f is Gâteaux-differentiable at x; its Gâteaux-differential at x is $D_G f(x) \in \mathcal{L}(E,F)$ (the vector space of the linear continuous mappings from E into F) defined by $D_G f(x).v := \vec{D}f(x;v)$. The mapping f is Fréchet-differentiable at x when there exists $Df(x) \in \mathcal{L}(E,F)$ (so-called the Fréchet-differential of f at x) and a mapping $\rho : \Omega - x \to F$ such that $\lim_{v\to 0} \rho(v) = 0$ and $f(x+v) = f(x) + Df(x).v + \|v\|\rho(v)$ for all $v \in \Omega - x$. When f is Fréchet-differentiable at x then f is Gâteaux-differentiable at x, and $D_G f(x) = Df(x)$. When $E = E_1 \times E_2$, when $k \in \{1,2\}$, $D_k f(x)$ (respectively $D_{G,k} f(x)$) denotes the partial Fréchet (respectively Gâteaux)-differential of f at x with respect to the k-th variable. For all these notions we refer to the books ([1], Chapter 2, Section 2.2) and ([7], Chapter 4, sections 4.1, 4.2).

$\mathbb{N}$ denotes the set of the non negative integer numbers, $\mathbb{N}_* : \mathbb{N}\backslash\{0\}$, $\mathbb{R}$ denotes the set of the real numbers and $\mathbb{R}_+$ denotes the set of the non negative real numbers. When $n \in \mathbb{N}_*$, we write $\mathbb{R}^{n*} := \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ the dual space.

We recall the Farkas-Minkowski Theorem.

Theorem 2.1. *Let $m, n \in \mathbb{N}_*$, φ_1, ..., $\varphi_m \in \mathbb{R}^{n*}$, and $a \in \mathbb{R}^{n*}$. The two following assertions are equivalent.*

(i) *For all $x \in \mathbb{R}^n$, $(\forall i \in \{i, ..., m\}, \varphi_i.x \geq 0) \Longrightarrow (a.x \geq 0)$.*
(ii) *There exists λ^1, ..., $\lambda^m \in \mathbb{R}_+$ such that $a = \sum_{1 \leq i \leq m} \lambda^i \varphi_i$.*

A complete proof of this result is given in ([16], Chapter, Sections 4.14-4.19) and in ([10], Chapter 2, Sections 2.5, 2.6). This result is present in many books like, for example ([13], Chapter 13, Section 2), ([17], p. 176), ([2], p. 164). A main difficulty of the proof of this theorem is the closedness of a finitely generated convex cone; a difficulty which is not ever well treated.

A second fundamental tool that we recall is the Implicit Function Theorem of Halkin for the Fréchet-differentiable mappings which are not necessarily continuously Fréchet-differentiable.

Theorem 2.2. *Let X, Y, Z be three real finite-dimensional mormed vector spaces, let $A \subset X \times Y$ be a nonempty open subset, let $f : A \to Z$ be a mapping, and let $(\bar{x}, \bar{y}) \in A$. We assume that the following conditions are fulfilled.*

(i) *$f(\bar{x}, \bar{y}) = 0$.*
(ii) *f is continuous on a neighborhood of $(\bar{x}, \bar{y})$*
(iii) *f is Fréchet-differentiable at V and the partial Fréchet-differential $D_2 f(\bar{x}, \bar{y})$ is bijective.*

Then there exist a neighborhood U of $\bar{x}$ in X, a neighborhood V of $\bar{y}$ in Y such that $U \times V \subset A$, and a mapping $\psi : U \to V$ which satisfy the following conditions.

(a) *$\psi(\bar{x}) = \bar{y}$*
(b) *For all $x \in U$, $f(x, \psi(x)) = 0$*
(c) *ψ is Fréchet-differentiable at $\bar{x}$ and $D\psi(\bar{x}) = -D_2 f(\bar{x}, \bar{y})^{-1} \circ D_1 f(\bar{x}, \bar{y})$.*

This result is proven in [8]. Its proof uses the Fixed Point Theorem of Brouwer. The electronic paper of Border [5] is very useful to understand the role of each assumption of the theorem. Halkin does not use an open subset A; his function is defined on $X \times Y$. But it is easy to adapt his result. Since ψ is Fréchet-differentiable at $\bar{x}$, ψ is continuous at $\bar{x}$ and then we can consider a neighborhood of $\bar{y}$ and a neighborhood U of $\bar{x}$ such that $\psi(U) \subset V$ and such that $U \times V \subset A$.

3. The main results

For the problem $(\mathcal{I})$ we state the following result.

Theorem 3.1. *Let $\hat{x}$ be a solution of $(\mathcal{I})$. We assume that the following assumptions are fulfilled.*

(i) *For all $i \in \{1, ..., m\}$, f_i is Gâteaux-differentiable at $\hat{x}$.*
(ii) *For all $i \in \{1, ..., m\}$, f_i is lower semicontinuous at $\hat{x}$ when $f_i(\hat{x}) > 0$.*

Then there exist λ^0,..., $\lambda^m \in \mathbb{R}_+$ such that the following conditions hold.

(a) *$(\lambda^0, ..., \lambda^m) \neq (0, ..., 0)$.*
(b) *For all $i \in \{1, ..., m\}$, $\lambda^i f_i(\hat{x}) = 0$.*
(c) *$\sum_{0 \leq i \leq m} \lambda^i D_G f_i(\hat{x}) = 0$.*

If, in addition, we assume that the following assumption is fulfilled,

(iii) *There exists* $w \in \mathbb{R}^n$ *such that, for all* $i \in \{1, ..., m\}$, $D_G f_i(\hat{x}).w > 0$ *when* $f_i(\hat{x}) = 0$,

then we can take $\lambda^0 = 1$.

The notion of lower semicontinuity is the classical one; see for instance [2] (p.74). For the problem $(\mathcal{M})$, we state the following result.

Theorem 3.2. *Let* $\hat{x}$ *be a solution of* $(\mathcal{M})$. *We assume that the following assumptions are fulfilled.*

(i) ϕ *is Fréchet-differentiable at* $\hat{x}$.
(ii) *For all* $i \in \{1, ..., p\}$, g_i *is Fréchet-differentiable at* $\hat{x}$ *when* $g_i(\hat{x}) = 0$.
(iii) *For all* $i \in \{1, ..., p\}$, g_i *is lower semicontinuous at* $\hat{x}$ *and Gâteaux-differentiable at* $\hat{x}$ *when* $g_i(\hat{x}) > 0$.
(iv) *For all* $j \in \{1, ..., q\}$, h_j *is continuous on a neighborhood of* $\hat{x}$ *and Fréchet-differentiable at* $\hat{x}$.

Then there exist λ^0, λ^1,..., $\lambda^p \in \mathbb{R}_+$ *and* μ^1,..., $\mu^q \in \mathbb{R}$ *such the following conditions are satisfied.*

(a) $(\lambda^0, \lambda^1, ..., \lambda^p, \mu^1, ..., \mu^q) \neq (, 0, ..., 0)$.
(b) *for all* $i \in \{1, ..., p\}$, $\lambda^i g_i(\hat{x}) = 0$.
(c) $\lambda^0 D\phi(\hat{x}) + \sum_{1 \leq i \leq p} \lambda^i D_G g_i(\hat{x}) + \sum_{1 \leq j \leq q} \mu^j Dh_j(\hat{x}) = 0$.

Moreover, under the additional assumption

(v) $Dh_1(\hat{x})$,..., $Dh_q(\hat{x})$ *are linearly independent,*

we can take

(d) $(\lambda^0, \lambda^1, ..., \lambda^p) \neq (0, 0, ..., 0)$.

Furthermore, under (v) and under the additional assumption

(vi) *There exists* $w \in \bigcap_{1 \leq j \leq q} \mathrm{Ker} Dh_j(\hat{x})$ *such that, for all* $i \in \{1, ..., p\}$, $Dg_i(\hat{x}).w > 0$ *when* $g_i(\hat{x}) = 0$,

we can take

(e) $\lambda^0 = 1$.

Remark 3.3. *The assumption (iii) is generally called the Mangarasian-Fromowitz's condition. In (*[13]*, p. 289) the author associates this condition at a work of Abadie in 1965 (it is difficult to find the reference). In (*[17]*, p. 197) we find a catalog of the variations of this condition due to Cottle, Zandwill, Kuhn and Tucker, and Abadie.*

In comparison with the Halkin's multiplier rule, for problem $(\mathcal{I})$ we have deleted the assumptions of local continuity on a neighborhood of $\hat{x}$ of the f_i and we have replaced their Fréchet-differentiability by their Gâteaux-differentiability, and for problem $(\mathcal{M})$, we have deleted the assumptions of local continuity on ϕ and on the g_i. In comparison with the result of [15] for problem $(\mathcal{I})$, we have replaced the Fréchet-differentiability of the f_i by their Gâteaux-differentiability. Note that the Gâteaux-differentiability of a mapping at a point does not imply the continuity of this mapping at this point.

4. Proof of Theorem 3.1

Doing a change of index, we can assume that $\{1, ..., e\} = \{i \in \{1, ..., m\} : f_i(\hat{x}) = 0\}$. If $f_i(\hat{x}) > 0$ for all $i \in \{1, ..., m\}$, the using the lower semicontinuity of (ii), there exists an open neighborhood of $\hat{x}$ on which $\hat{x}$ maximizes f_0 (without constraints). Then using (i) we obtain $D_G f_0(\hat{x}) = 0$, and we conclude by taking $\lambda^0 := 1$ and $\lambda^i := 0$ for all $i \in \{1, ..., m\}$. And so, for the sequel of the proof we assume that $1 \leq e \leq p$.

4.1. **Proof of (a), (b), (c).** Ever using (ii) we can assert that there exists an open neighborhood $\Omega_1 \subset \Omega$ of $\hat{x}$ such that, for all $x \in \Omega_1$ and for all $i \in \{e + 1, ..., m\}$, $f_i(x) > 0$ when $e < m$. When $e = m$ we simply take $\Omega_1 := \Omega$. Then $\hat{x}$ is a solution of the following problem.

$$(\mathcal{P}) \begin{cases} \text{Maximize} & f_0(x) \\ \text{when} & x \in \Omega_1 \\ \text{and when} & \forall i \in \{1, ..., e\},\ f_i(x) \geq 0. \end{cases}$$

For all $k \in \{0, ..., e\}$ we introduce the set

$$A_k := \{v \in \mathbb{R}^n : \forall i \in \{k, ..., e\}, D_G f_i(\hat{x}).v > 0\}. \tag{4.1}$$

We want to prove that $A_0 = \emptyset$. To realize that, we proceed by contradiction; we assume that $A_0 \neq \emptyset$, and so there exists $w \in \mathbb{R}^n$ such that $D_G f_i(\hat{x}).w > 0$ for all $i \in \{0, ..., e\}$. Since Ω_1 is open, there exists $\theta_* \in (0, +\infty)$ such that $\hat{x} + \theta w \in \Omega_1$ for all $\theta \in [0, \theta_*]$. After (i), for all $i \in \{0, ..., e\}$, the function $\sigma_i : [0, \theta_*] \to \mathbb{R}$, defined by $\sigma_i(\theta) := f_i(\hat{x} + \theta w)$, is differentiable at 0, and its derivative is $\sigma_i'(0) = D_G f_i(\hat{x}).w$. The differentiability of σ_i at 0 implies the existence of a function $\rho_i : [0, \theta_*] \to \mathbb{R}$ such that $\lim_{\theta \to 0} \rho_i(\theta) = 0$ and such that $\sigma_i(\theta) = \sigma_i(0) + \sigma_i'(0)\theta + \theta\rho_i(\theta)$ for all $\theta \in [0, \theta_*]$. Translating this last equality we obtain $f_i(\hat{x} + \theta w) = f_i(\hat{x}) + \theta(D_G f_i(\hat{x}).w + \rho_i(\theta))$. Since $D_G f_i(\hat{x}).w > 0$ and since $\lim_{\theta \to 0} \rho_i(\theta) = 0$, we obtain the existence of $\theta_i \in (0,_t heta_*]$ such that $D_G f_i(\hat{x}).w + \rho_i(\theta) > 0$ for all $\theta \in (0, \theta_i]$. Setting $\hat{\theta} := \min\{\theta_i : i \in \{0, ..., e\}$ we obtain that $f_i(\hat{x} + \theta w) > f_i((\hat{x})$ for all $\theta \in (0, \hat{\theta}]$ and for all $i \in \{0, ..., e\}$. Then using $i \in \{1, ..., e\}$, this last relation ensures that $\hat{x} + \theta w$ is admissible for $(\mathcal{P})$ when $\theta \in (0, \hat{\theta}]$, and using this last relation when $i = 0$ we obtain $f_0(\hat{x} + \theta w) > f_0((\hat{x})$ when $\theta \in (0, \hat{\theta}]$, that is impossible since $\hat{x}$ is a solution of $(\mathcal{P})$. And so the reasoning by contradiction is complete, and we have proven

$$A_0 = \emptyset. \tag{4.2}$$

When $A_e = \emptyset$ there is not any $v \in \mathbb{R}^n$ such that $D_G f_e(\hat{x}).v > 0$, that implies that $D_G f_e(\hat{x}) = 0$. Then taking $\lambda^e := 1$ and $\lambda^i := 0$ when $i \in \{0, ..., m\} \setminus \{e\}$, we obtain the conclusions (a), (b), (c). And so we have proven

$$A_e = \emptyset \Longrightarrow ((a), (b), (c) \ \text{hold}). \tag{4.3}$$

Now we assume that $A_e \neq \emptyset$. Since we have $A_0 = \emptyset$ after (4.2) and $A_i \subset A_{i+1}$ we can define

$$k := \min\{i \in \{1, ..., e\} : A_i \neq \emptyset\}. \tag{4.4}$$

Note that $A_k \neq \emptyset$ and that $A_{k-1} = \emptyset$. We consider the following problem

$$(\mathcal{Q}) \begin{cases} \text{Maximize} & D_G f_{k-1}(\hat{x}).v \\ \text{when} & v \in \mathbb{R}^n \\ \text{and when} & \forall i \in \{k, ..., e\},\ \ D_G f_i(\hat{x}).v \geq 0. \end{cases}$$

We want to prove that 0 is a solution of $(\mathcal{Q})$. To do that, we proceed by contradiction; we assume that there exists $y \in \mathbb{R}^n$ such that $(\forall i \in \{k, ..., e\},\ D_G f_i(\hat{x}).y \geq 0)$ and $D_G f_{k-1}(\hat{x}).y > 0 = D_G f_{k-1}(\hat{x}).0$. Since $A_k \neq \emptyset$, there exists $z \in \mathbb{R}^n$ such that $D_G f_i(\hat{x}).z > 0$ when $i \in \{k, ..., e\}$. We cannot have $D_G f_{k-1}(\hat{x}).z > 0$ since $A_{k-1} = \emptyset$. Therefore we have $D_G f_{k-1}(\hat{x}).z \leq 0$. If $D_G f_{k-1}(\hat{x}).z < 0$ we choose ϵ such that $0 < \epsilon < \frac{D_G f_{k-1}(\hat{x}).y}{D_G f_{k-1}(\hat{x}).z}$. Then we have $D_G f_{k-1}(\hat{x}).y + \epsilon D_G f_{k-1}(\hat{x}).z > 0$. If $D_G f_{k-1}(\hat{x}).z = 0$ we arbitrarily choose $\epsilon \in (0, +\infty)$ and we have also $D_G f_{k-1}(\hat{x}).y + \epsilon D_G f_{k-1}(\hat{x}).z > 0$. We set $u_\epsilon := y + \epsilon z$, and we note that $D_G f_{k-1}(\hat{x}).u_\epsilon = D_G f_{k-1}(\hat{x}).y + \epsilon D_G f_{k-1}(\hat{x}).z > 0$. Furthermore, when $i \in \{k, ..., e\}$, we have $D_G f_i(\hat{x}).u_\epsilon = D_G f_i(\hat{x}).y + \epsilon D_G f_i(\hat{x}).z > 0$ since the three terms are positive. Therefore we have $u_\epsilon \in A_{k-1}$ that is impossible since $A_{k-1} = \emptyset$. And so the reasoning by contradiction is complete, and we have proven

$$A_e \neq \emptyset \Longrightarrow (0 \text{ solves } (\mathcal{Q})). \tag{4.5}$$

Since 0 solves $(\mathcal{Q})$, we have, for all $v \in \mathbb{R}^n$,

$$(\forall i \in \{k, ..., e\},\ \ D_G f_i(\hat{x}).v \geq 0) \Longrightarrow (D_G f_{k-1}(\hat{x}).v \geq 0).$$

Then we use Theorem 2.1 that ensures the existence of α^k,..., $\alpha^e \in \mathbb{R}_+$ such that $D_G f_{k-1}(\hat{x}) + \sum_{k \leq i \leq e} \alpha^i D_G f_i(\hat{x}) = 0$. We set

$$\lambda^i := \begin{cases} 0 & \text{if} \quad i \in \{0, ..., k-2\} \\ 1 & \text{if} \quad i = k-1 \\ \alpha^i & \text{if} \quad i \in \{k, ..., e\} \\ 0 & \text{if} \quad i \in \{e+1, ..., m\}, \end{cases}$$

and we obtain

$$A_e \neq \emptyset \Longrightarrow ((a), (b), (c) \text{ hold}). \tag{4.6}$$

Then, with (4.3) and (4.6) the conclusions (a), (b), (c) are proven.

4.2. **Proof of (d).** The assumption (iii) means that $A_1 \neq \emptyset$, and by (4.2) we know that $A_0 = \emptyset$. Proceeding like in the proof of (4.5) we prove that 0 is a solution of the following problem

$$\begin{cases} \text{Maximize} & D_G f_0(\hat{x}).v \\ \text{when} & v \in \mathbb{R}^n \\ \text{and when} & \forall i \in \{1, ..., e\},\ \ D_G f_i(\hat{x}).v \geq 0. \end{cases}$$

Then using Theorem 2.1, there exist α^1,..., $\alpha \in \mathbb{R}_+$ such that

$$D_G f_0(\hat{x}) + \sum_{1 \leq i \leq e} \alpha^i D_G f_i(\hat{x}) = 0.$$

We conclude by setting

$$\lambda^i := \begin{cases} 1 & \text{if} \quad i = 0 \\ \alpha^i & \text{if} \quad i \in \{1, ..., e\} \\ 0 & \text{if} \quad i \in \{e+1, ..., m\}. \end{cases}$$

And so the proof of Theorem 3.1 is complete.

Remark 4.1. *The use of the sets A_k comes from the book of Alexeev-Tihomirov-Fomin* [1]*, end the proof of formula (4.6) is similar to their proof (p. 247-248). The use of the set A_0 is yet done in* [8].

5. Proof of Theorem 3.2

We split this proof in seven steps.

5.1. **First step : a first simple case.** If $Dh_1(\hat{x})$, ..., $Dh_q(\hat{x})$ are linearly dependent, there exist μ^1,..., $\mu^q \in \mathbb{R}$ such that $(\mu^1, ..., \mu^q) \neq (0, ..., 0)$ and such that $\sum_{1\leq j\leq q} \mu^j Dh_j(\hat{x}) = 0$. Then it suffices to take $\lambda^i = 0$ for all $i \in \{0, ..., p\}$ to obtain the conclusions (a), (b), (c).

Now in the sequel of the proof **we assume that the assumption (v) is fulfilled**.

5.2. **Second step : To delete the non satured inequality constraints.** Doing a change of index, we can assume that $\{1, ..., e\} := \{i \in \{1, ..., p\} : g_i(\hat{x}) = 0\}$. Using the lower semicontinuity at $\hat{x}$ of the g_i when $i \in \{e+1, ..., p\}$, we can say that there exists an open neighborhood Ω_1 of $\hat{x}$ in Ω such that $g_i(x) > 0$ when $x \in \Omega_1$ and when $i \in \{e + 1, ..., p\}$. And so $\hat{x}$ is a solution of the following problem

$$(\mathcal{M}_1)\begin{cases} \text{Maximize} & \phi(x) \\ \text{when} & x \in \Omega_1 \\ \text{when} & \forall i \in \{1, ..., e\},\ \ g_i(x) \geq 0 \\ \text{and when} & \forall j \in \{1, ..., q\},\ \ h_j(x) = 0. \end{cases}$$

5.3. **To delete the equality constraints.** We consider the mapping $h : \Omega_1 \to \mathbb{R}^q$ defined by $h(x) := (h_1(x), ..., h_q(x))$. Under (iv) and (v), h continuous on a neighborhood of $\hat{x}$, and it is Fréchet-differentiable at $\hat{x}$ with $Dh(\hat{x})$ onto.

We set $E_1 := \text{Ker}Dh(\hat{x})$ and we take a vector subspace of $\mathbb{R}^n$ such that $E_1 \oplus E_2 = \mathbb{R}^n$. And we can do the assimilitation $\mathbb{R}^n = E_1 \times E_2$. We set $(\hat{x}_1, \hat{x}_2) := \hat{x} \in E_1 \times E_2$. Then the partial differential $D_2h(\hat{x})$ is an isomorphism from E_2 onto $\mathbb{R}^q$. Now we can use Theorem 2.2 and assert that there exist a neighborhood U_1 of $\hat{x}_1$ in E_1, a neighborhood U_2 of $\hat{x}_2$ in E_2, and a mapping $\psi : U_1 \to U_2$ such that $\psi(\hat{x}_1) = \hat{x}_2$, $h(x_1, \psi(x_1)) = 0$ for all $x_1 \in U_1$, and such that ψ is Fréchet-differentiable at $\hat{x}_1$ with $D\psi(\hat{x}_1) = -D_2h(\hat{x})^{-1} \circ D_1h(\hat{x}) = 0$ since $D_1h(\hat{x}) = Dh(\hat{x})_{|_{E_1}} = 0$.

We define $f_0 : U_1 \to \mathbb{R}$ by setting $f_0(x_1) := \phi(x_1, \psi(x_1))$, and $f_i : U_1 \to \mathbb{R}$ by setting $f_i(x_1) := g_i(x_1, \psi(x_1))$ for all $i \in \{1, ..., e\}$. Since $\hat{x}$ is a solution of $(\mathcal{M}_1)$, $\hat{x}_1$ is a solution of the following problem without equality constraints

$$(\mathcal{R})\begin{cases} \text{Maximize} & f_0(x_1) \\ \text{when} & x_1 \in U_1 \\ \text{and when} & \forall i \in \{1, ..., e\};\ \ f_i(x_1) \geq 0. \end{cases}$$

5.4. **Fourth step : To use Theorem 3.1.** Since ψ is Fréchet-differentiable at $\hat{x}_1$, the mapping $[x_1 \mapsto (x_1, \psi(x_1))]$ is Fréchet-differentiable at $\hat{x}_1$, and using (i) and (ii), we obtain that f_i is Fréchet-differentiable (and therefore Gâteaux-differentiable) at $\hat{x}_1$, for all $i \in \{0, ..., e\}$. Note that $f_i(\hat{x}_1) = 0$ for all $i \in \{1, ..., e\}$. Consequently we can use Theorem 3.1 on $(\mathcal{R})$ that permits us to ensure the existence of λ^0,λ^1,..., $\lambda^e \in \mathbb{R}_+$ such that

$$(\lambda^0, \lambda^1, ..., \lambda^e) \neq (0, 0, ..., 0) \tag{5.1}$$

$$\forall i \in \{1, ..., e\},\ \ \lambda^i f_i(\hat{x}_1) = 0 \tag{5.2}$$

$$\sum_{0\leq i\leq e} \lambda^i D_G f_i(\hat{x}_1) = 0. \tag{5.3}$$

5.5. **The proof of (a), (b), (c).** Since $D_G f_0(\hat{x}_1) = Df_0(\hat{x}_1) = D_1\phi(\hat{x}) + D_2\phi(\hat{x}) \circ D\psi((\hat{x}_1) = D_1\phi(\hat{x})$ since $D\psi(\hat{x}_1) = 0$,$D_G f_i(\hat{x}_1) = Df_i(\hat{x}_1) = D_1 g_i(\hat{x}) + D_2 g_i(\hat{x}) \circ D\psi(\hat{x}_1) = D_1 g_i(\hat{x})$, for all $i \in \{1, ..., e\}$, the formula (5.3) implies

$$\lambda^0 D_1\phi(\hat{x}) + \sum_{1\leq i\leq e} \lambda^i D_1 g_i(\hat{x}) = 0. \tag{5.4}$$

We set

$$M := -(\lambda^0 D_2\phi(\hat{x}) + \sum_{1\leq i\leq e} \lambda^i D_2 g_i(\hat{x})) \circ D_2 h(\hat{x})^{-1}) \in \mathbb{R}^{q*}. \tag{5.5}$$

Then we have

$$\lambda^0 D_2\phi(\hat{x}) + \sum_{1\leq i\leq e} \lambda^i D_2 g_i(\hat{x}) + M \circ D_2 h(\hat{x}) = 0.$$

Denoting by μ^1,..., $\mu^q \in \mathbb{R}$ the coordinates of M in the canonical basis of $\mathbb{R}^{q*}$, we obtain

$$\lambda^0 D_2\phi(\hat{x}) + \sum_{1\leq i\leq e} \lambda^i D_2 g_i(\hat{x}) + \sum_{1\leq j\leq q} \mu^j D_2 h_j(\hat{x}) = 0. \tag{5.6}$$

Since $E_1 = \mathrm{Ker} Dh(\hat{x}) = \bigcap_{1\leq j\leq q} \mathrm{Ker} Dh_j(\hat{x})$, we have $D_1 h_j(\hat{x}) = Dh(\hat{x})_{|_{E_1}} = 0$ for all j, from (5.4) we obtain

$$\lambda^0 D_1\phi(\hat{x}) + \sum_{1\leq i\leq e} \lambda^i D_1 g_i(\hat{x}) + \sum_{1\leq j\leq q} \mu^j D_1 h_j(\hat{x}) = 0. \tag{5.7}$$

From (5.6) and (5.7) we obtain

$$\lambda^0 D\phi(\hat{x}) + \sum_{1\leq i\leq e} \lambda^i Dg_i(\hat{x}) + \sum_{1\leq j\leq q} \mu^j Dh_j(\hat{x}) = 0. \tag{5.8}$$

We set $\lambda^i := 0$ when $i \in \{e+1, ..., p\}$, and so (5.8) implies (c). With (5.1) we obtain (a), and with (5.2) we obtain (b). And so the proof of (a), (b), (c) is complete.

5.6. **The proof of (d).** The relation (5.3) provides the conclusion (d).

5.7. **The proof of (e).** When $i \in \{1, ..., e\}$, we have yet seen that $Df_i(\hat{x}_1) = D_1 g_i(\hat{x} = Dg_i(\hat{x})_{|_{E_1}}$. And so the translation of the assumption (vi) gives

$$\exists w \in E_1 \text{ s.t. } \forall i \in \{1, ..., e\}, \ \ Df_i(\hat{x}_1).w > 0.$$

That permits us to use the last assertion of Theorem 3.1 on $(\mathcal{R})$ to ensure that we can choose $\lambda^0 = 1$.
Then the proof of Theorem 3.2 is complete.

Remark 5.1. *We see in this proof that the assumption of Fréchet-diffferentiability of the h_j is used to can apply the Implicit Function of Halkin. The assumption of Fréchet-diffferentiability of ϕ and of the g_i for which the associated constraint is satured is used to obtain the differentiability when we compose them with h_j (to obtain the differentiability of the f_i). The Hadamard-differentiability is sufficient to do that, but in finite-dimensional spaces, the Hadamard-differentiability and the Fréchet-differentiability coincide ([7], p. 266).*

References

[1] V.M. Alexeev, V.M. Tihomirov, and S.V. Fomin, Commande optimale, French edition, MIR, Moscow, 1982.

[2] C. Berge, Topological spaces, English edition, Dover Publications, Inc., Mineola, New York, 1997.

[3] Ş.I. Birbil, J.B.G. Frenk, and G.J. Still, *An elementary proof of the Fritz John and Karush-Kuhn-Tucker conditions in nonlinear programming*, European J. of Oper. Res., 2006.

[4] J. Blot and N. Hayek, Infinite-horizon optimal control in the discrete-time framework, Springer, New York, 2014.

[5] K.C. Border, *Notes on the implicit function theorem*, California Institut of Technology, Division of the Humanities and Social Sciences, www.hss.calthch.edu/ kcb/Notes/ITF.pdf.

[6] F.H. Clarke, , Optimization and Nonsmooth Analysis, Wiley-Interscience, New York,1983.

[7] T.M. Flett, Differential analysis, Cambridge University Press, Cambridge, 1980.

[8] H. Halkin, *Implicit functions and optimization problems without continuous differentiability of the data*, SIAM J. Control, **12**(2), 1974, 229-236.

[9] J.-B. Hiriart-Urruty, L'optimisation, P.U.F., Paris, 1996.

[10] V.G. Karmanov, Programmation mathématique, French edition, MIR, Moscow, 1977.

[11] O.L Mangasarian and S. Fromowitz, *The Fritz John necessary conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl. **17**, 1967, 37-47.

[12] P. Michel, Cours de mathématiques pour économistes, Second edition, Economica, Paris, 1989.

[13] R. Pallu de la Barrière, Cours d'automatique théorique, Dunod, Paris, 1966.

[14] O. Stein, *On Karush-Kuhn-Tucker points for a smoothing method in semi-infinite optimization*, Comput. Math. **24**(6), 2006, 719-732.

[15] B.H. Pourciau, *Modern multiplier rules*, Amer. Math. Monthly **87**(6), 1980, 433-457.

[16] J. van Tiel, Convex analysis, John Wiley and Sons, Chichester, 1984.

[17] M. Truchon, Théorie de l'optimisation statique et différentiable, Gaëtan Morin, Chicoutimi, 1987.

Joël Blot: Laboratoire SAMM EA4543,
Université Paris 1 Panthéon-Sorbonne, centre P.M.F.,
90 rue de Tolbiac, 75634 Paris cedex 13, France.
E-mail address: blot@univ-paris1.fr