Journal of Convex Analysis
Volume 24 (2017), No. 4, 1117–1142

Parametric Semidifferentiability of Minimax of Lagrangians: Averaged Adjoint Approach

Michel C. Delfour*
Centre de Recherches Mathématiques and Dép. de Mathématiques et de Statistique, Université de Montréal, CP 6128, succ. Centre-ville, Montréal, Canada H3C 3J7
delfour@crm.umontreal.ca

Kevin Sturm
Fakultät für Mathematik, Universität Duisburg-Essen, Thea-Leymann-Straße 9, 45127 Essen, Germany
kevin.sturm@uni-due.de

Received: January 20, 2016
Revised manuscript received: April 14, 2016
Accepted: May 30, 2016

A standard approach to the *minimization* of a *constrained objective function* in the presence of *equality constraints* in Mathematical Programming or of a *state equation* in Control Theory is to introduce Lagrange multipliers or an *adjoint state.* The initial minimization problem is equivalent to the minimax of the associated Lagrangian.

In this paper, the targeted application is the computation of shape and topological derivatives (Sensitivity Analysis) where the control variable belongs to a metric space of geometries. In that context, it is natural to consider the right-hand side differentiability of the minimax of a Lagrangian with respect to a positive parameter as a first step towards the computation of semidifferentials and differentials on a metric space.

By using the new notion of *averaged adjoint* introduced by K. Sturm [31] [32], the minimax problem need not be related to a saddle point as in Correa and Seeger [6] and the so-called *dual problem* need not make sense. We extend his results from the single valued case to the case where the solutions of the state/averaged adjoint state equations are not unique. In such a case, a non-differentiability can occur even if the functions at hand are infinitely differentiable as was illustrated by the *seesaw problem* in the seminal paper of J. M. Danskin [7] in 1966.

Keywords: Minimax, Lagrangian, sensitivity analysis, semidifferential, averaged adjoint, shape, topological derivatives, optimal control, mathematical programming.

2010 Mathematics Subject Classification: 49K20, 49K27, 49K35, 49K40, 49Q10, 49Q12.

*The research of the first author was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

ISSN 0944-6532 / $ 2.50

1. Introduction

Theorems on the parametric differentiability of a minimax with or without saddle point have a long history. J. M. Danskin [7] in 1966 and [8] in 1967 studied such problems and, as a first step, he considered the differentiability of an extremum with respect to parameters. He pointed out that, when the solution of the extremum problem is not unique, a non-differentiability can occur and that only a semidifferential may be available even if all the functions at hand are infinitely differentiable. He was followed by Dem'janov (cf. Dem'janov and Malozemov [25]) in 1968. V. F. Dem'janov [23]) also studied the differentiation of the maximin function with or without the saddle point assumption in 1968 and published a book [24] on that topic in 1974. Extensions to infinite dimensional problems were given in the thesis of B. Lemaire [30] in 1970. The reader is referred to the guided tour [3] in 1998 and the important book [4] in 2000 by J. F. Bonnans and A. Shapiro for a more recent extensive treatment of such problems with applications to nonlinear programming and optimal control.

In 1986 Delfour and Zolésio [16] introduced a theorem on the differentiability of a minimax in the context of optimal control and extended it in 1987 to shape semidifferentiability in [17]. This theorem that didn't assume the existence of a saddle point could not be applied directly to the *primal problem* since some of the assumptions were not satisfied. Yet, by adding a term to the Lagrangian, the assumptions were satisfied for the *dual problem* and the theorem could be applied to the initial minimax problem. This technique was put aside when Correa and Seeger [6] published in 1985 a direct theorem on the differentiability of the saddle point $g(z)$ of a functional $G(z, x, y)$ where $z \in Z$, Z a locally convex space, and $(x, y) \in X \times Y$, X and Y two Hausdorff topological spaces. Since spaces of shapes and domains are not locally convex spaces, Delfour and Zolésio [19] in 1988 reformulated the hypotheses to make them readily applicable to the computation of the shape derivative with respect to a velocity field. Some of those theorems were sharpened by Delfour and Morgan [14, Theorem 3] in 1992 and extended to ε-solutions in [15] in 1994.

Recently, K. Sturm [31] [32] introduced a novel approach to the differentiability of a minimax *without a saddle point assumption* for the Lagrangian. Its originality is to replace, within the hypotheses and the proof, the *standard adjoint state* by an *averaged adjoint state*. An important consequence of this approach is to be able to directly deal with many non-convex objective functions and non-linear state equations.

As illustrated by the *seesaw problem* of J. M. Danskin [7], even if all the functions at hand are infinitely differentiable, a non-differentiability can arise from the fact that the solutions of the *state equation* or the *adjoint state equation* are not unique.

In this paper we relax the assumption in [31, Thm. 3.2] and [32, Thm. 3.1] that the sets of solutions be singletons to the multivalued case that was considered in [19] and [14]. We also relax the global differentiability assumption with respect to the parameter to a local one. Several sets of hypotheses will be presented for the existence of the right-hand side derivative of the minimax with respect to a positive parameter followed by a discussion of its expression. The targeted applications are *shape* and *topological sensitivity analysis* but the material is also pertinent to optimal control and mathematical programming for which a very rich literature is available. Among technical applications of the multivalued case are objective functions that depends on the solution of partial differential equations with a non-homogeneous Dirichlet boundary condition (cf. Delfour and Zolésio [21, 22]).

2. Abstract framework via the averaged adjoint state

2.1. Minimax formulation in the computation of (semi-)differentials

In order to motivate the general approach that will be followed, we quote the following example of a state constrained objective function which is detailed in [31, p. 64].

Let Ω be a bounded open domain in $\mathbb{R}^N$, $N \geq 1$, and $a \in L^2(\Omega)$ be the *control variable* to which is associated the *state* $u = u(a)$ solution of partial differential equation

$$-\Delta u(a) + \rho(u(a)) = a \text{ in } \Omega, \quad u(a) = 0 \text{ on } \partial\Omega,$$

where ρ is a continuously differentiable, bounded, and non-decreasing function. In weak form the *state* $u = u(a) \in H_0^1(\Omega)$ is solution of the variational *state equation*

$$\int_\Omega \nabla u \cdot \nabla \psi + \rho(u)\,\psi - a\,\psi\,dx = 0, \quad \forall \psi \in H_0^1(\Omega), \tag{1}$$

where $x \cdot y$ denotes the inner product of x and y in $\mathbb{R}^N$. Given a target function $g \in L^2(\Omega)$, associate with $u(a)$ the *state constrained objective function*

$$f(a) \stackrel{\text{def}}{=} \int_\Omega |u(a) - g|^2 + \alpha\,|a|^2\,dx, \quad \alpha > 0,$$

to be minimized over $L^2(\Omega)$. By introducing the Lagrangian, we get an unconstrained minimax expression of $f(a)$:

$$G(a, \varphi, \psi) \stackrel{\text{def}}{=} \int_\Omega |\varphi - g|^2 + \alpha\,|a|^2 + \nabla\varphi \cdot \nabla\psi + \rho(\varphi)\,\psi - a\,\psi\,dx,$$

$$f(a) = \inf_{\varphi \in H^1(\Omega)} \sup_{\psi \in H^1(\Omega)} G(a, \varphi, \psi).$$

Under appropriate conditions, the minimax formulation yields a necessary optimality condition in terms of a coupled system of state and adjoint state equations.

If we are only interested in a descent method, we can obtain the semidifferential of $f(a)$ by a similar minimax formulation. Given the direction $b \in L^2(\Omega)$, we want to compute $df(a;b) = \lim_{t\searrow 0}(f(a+tb) - f(a))/t$. The *state* $u^t \in H_0^1(\Omega)$ at $t \geq 0$ is solution of

$$\int_\Omega \nabla u^t \cdot \nabla\psi + \rho(u^t)\,\psi - (a+tb)\,\psi\,dx = 0, \quad \forall \psi \in H_0^1(\Omega), \tag{2}$$

and the *associated Lagrangian* is

$$L(t,\varphi,\psi) \stackrel{\text{def}}{=} \int_\Omega |\varphi - g|^2 + \alpha\,|a+tb|^2 + \nabla\varphi \cdot \nabla\psi + \rho(\varphi)\,\psi - (a+tb)\,\psi\,dx.$$

It is readily seen that

$$g(t) \stackrel{\text{def}}{=} f(a+tb) = \inf_{\varphi \in H_0^1(\Omega)} \sup_{\psi \in H_0^1(\Omega)} L(t,\varphi,\psi),$$

$$dg(0) \stackrel{\text{def}}{=} \lim_{t\searrow 0}(g(t) - g(0))/t = df(a;b).$$

So it is sufficient to have a theorem for the existence and expression of the right-hand side t-derivative of a parametrized minimax at $t = 0$ as a first step toward the computation of a (semi) differential with respect to the control variable a.

2.2. Further motivation: Metric spaces of shapes and geometries

When compared to *optimal control* of partial differential equations, *shape and topological optimization* are characterized by the fact that the design or control variable is the domain on which the partial differential equation is defined. The reader is referred to Delfour and Zolésio [22] for the *Courant metrics* and metrics constructed from set-parametrized functions and to the recent tutorial paper of Delfour [11] that introduces new complete metric spaces. It also contains constructions of equivalent metrics that make the spaces of distance and oriented distance functions Abelian groups that contain both the empty set and the hold-all in their completion.

The so-called *shape and topological derivatives* turn out to be differentials and semi-differentials defined on complete metric spaces with at best a group structure. For the *shape derivative* the domain is perturbed by a family of diffeomorphisms generated by velocity fields and this yields differentials. For the *topological derivative*, the domain is perturbed by dilations of rectifiable sets (see Federer [27]) or sets of positive reach (see Federer [26]). Here the Minkowski content (which is obtained as the limit of the dilatation when the dilatation parameter goes to zero) is either a Hausdorff or a Radon measure. In general, we only get semidifferentials. The reader is referred to [12] for more details.

In the absence of a general theory, the approach that has been privileged is to formulate the hypotheses in terms of sequences since the *control variable* might only belong to a metric space. It is a choice at the level of the verification of the hypotheses which is complemented by resorting to the many notions and tools available in the literature (see, for instance, [4]). The results will be presented in an abstract framework with ready-to-check but not necessarily easy-to-check hypotheses since they implicitly require checking some compactness or boundedness conditions.

2.3. Definitions and problem formulation

For the purposes of the paper, a *Lagrangian* is a function of the form

$$(t, x, y) \mapsto G(t, x, y) \colon [0, \tau] \times X \times Y \to \mathbb{R}, \quad \tau > 0,$$

where Y is a *vector space*, X is a non empty subset of a vector space, and $y \mapsto G(t, x, y)$ is *affine*. Associate with the *parameter* t the *parametrized minimax*

$$t \mapsto g(t) \overset{\text{def}}{=} \inf_{x \in X} \sup_{y \in Y} G(t, x, y) \colon [0, \tau] \to \mathbb{R}. \tag{3}$$

Definition 2.1. When the limits exist and X is open we use the following compact notation:

$$dg(0) \overset{\text{def}}{=} \lim_{t \searrow 0} \frac{g(t) - g(0)}{t} \tag{4}$$

$$d_t G(0, x, y) \overset{\text{def}}{=} \lim_{t \searrow 0} \frac{G(t, x, y) - G(0, x, y)}{t} \tag{5}$$

$$v \in X, \quad d_x G(t, x, y; v) \overset{\text{def}}{=} \lim_{\theta \searrow 0} \frac{G(t, x + \theta v, y) - G(t, x, y)}{\theta} \tag{6}$$

$$w \in Y, \quad d_y G(t, x, y; w) \overset{\text{def}}{=} \lim_{\theta \to 0} \frac{G(t, x, y + \theta w) - G(t, x, y)}{\theta}. \tag{7}$$

The notation $t \searrow 0$ and $\theta \searrow 0$ means that t and θ go to 0 by strictly positive values.

Since $G(t, x, y)$ is affine in y, for all $(t, x, y) \in [0, \tau] \times X \times Y$,

$$\forall w \in Y, \quad d_y G(t, x, y; w) = G(t, x, w) - G(t, x, 0). \tag{8}$$

The variational equation

$$\text{find } x^t \in X \text{ such that for all } \varphi \in Y,\ d_y G(t, x^t, 0; \varphi) = 0 \tag{9}$$

will be referred to as the *state equation* and the set of solutions

$$E(t) \overset{\text{def}}{=} \left\{ x^t \in X : \forall \varphi \in Y,\ d_y G(t, x^t, 0; \varphi) = 0 \right\} \tag{10}$$

as the *states* at $t \geq 0$. Finally, let

$$X(t) \overset{\text{def}}{=} \left\{ x^t \in X : \sup_{y \in Y} G(t, x^t, y) = g(t) \right\}. \tag{11}$$

Lemma 2.2 (Constrained infimum and minimax).

(i)
$$\inf_{x\in X}\sup_{y\in Y} G(t,x,y)=\inf_{x\in E(t)} G(t,x,0). \tag{12}$$

(ii) *The minimax $g(t)=+\infty$ if and only if $E(t)=\varnothing$. In that case $X(t)=X$.*

(iii) *If $E(t)\neq\varnothing$, then*
$$X(t)=\{x^t\in E(t):G(t,x^t,0)=\inf_{x\in E(t)} G(t,x,0)\} \tag{13}$$
and $g(t)<+\infty$.

(iv) *$X(t)\cap E(t)\neq\varnothing$ if and only if*
$$\{x^t\in E(t):G(t,x^t,0)=\inf_{x\in E(t)} G(t,x,0)\}\neq\varnothing \tag{14}$$
if and only if $E(t)\neq\varnothing$ and $X(t)\neq\varnothing$.

Proof. (i) Since $y\mapsto G(t,x,y)$ is affine, for all (t,x,y)
$$G(t,x,y)=G(t,x,0)+d_yG(t,x,0;y)$$
$$\sup_{y\in Y} G(t,x,y)=G(t,x,0)+\sup_{y\in Y} d_yG(t,x,0;y)=\begin{cases}G(t,x,0), & \text{if } x\in E(t)\\ +\infty, & \text{if } x\in X\backslash E(t)\end{cases}$$
$$\Rightarrow \inf_{x\in X}\sup_{y\in Y} G(t,x,y)=\inf_{x\in E(t)} G(t,x,0).$$

(ii) If $g(t)=+\infty$, then for all $x\in X$, $\sup_{y\in Y} G(t,x,y)=+\infty$ and
$$+\infty=G(t,x,0)+\sup_{y\in Y} d_yG(t,x,0;y) \quad\Rightarrow \sup_{y\in Y} d_yG(t,x,0;y)=+\infty$$
$$\Rightarrow \exists y\in Y \text{ such that } d_yG(t,x,0;y)>0 \quad\Rightarrow \forall x\in X,\ x\notin E(t)$$
and $E(t)=E(t)\cap X=\varnothing$. Conversely, from (i) and by definition of the infimum,
$$E(t)=\varnothing \quad\Rightarrow g(t)=\inf_{x\in E(t)} G(t,x,0)=+\infty.$$

(iii) If $E(t)\neq\varnothing$ and $x^t\in X(t)$, from part (i)
$$G(t,x^t,0)\le\sup_{y\in Y} G(t,x^t,y)=g(t)=\inf_{x\in E(t)} G(t,x,0)<+\infty.$$

If $x^t\notin E(t)$, $\sup_{y\in Y} d_yG(t,x^t,0;y)=+\infty$ and
$$g(t)=\sup_{y\in Y} G(t,x^t,y)=G(t,x^t,0)+\sup_{y\in Y} d_yG(t,x,0;y)=+\infty$$

which contradict the fact that $g(t)$ is finite. So $X(t) \subset E(t)$ and

$$G(t,x^t,0) \le \sup_{y\in Y} G(t,x^t,y) = g(t) = \inf_{x\in E(t)} G(t,x,0) \le G(t,x^t,0)$$
$$\Rightarrow\ x^t \in E(t), \quad g(t) = G(t,x^t,0) = \inf_{x\in E(t)} G(t,x,0).$$

Conversely, if there exists $x^t \in E(t)$ such that

$$G(t,x^t,0) = \inf_{x\in E(t)} G(t,x,0),$$

then, for all $y \in Y$, $d_yG(t,x^t,0;y) = 0$, $G(t,x^t,y) = G(t,x^t,0)$, and from (i)

$$\sup_{y\in Y} G(t,x^t,y) = G(t,x^t,0) = \inf_{x\in E(t)} G(t,x,0) = \inf_{x\in X}\sup_{y\in Y} G(t,x,y)$$

and $x^t \in X(t)$.

(iv) If there exists $x^t \in E(t)$ such that

$$G(t,x^t,0) = \inf_{x\in E(t)} G(t,x,0),$$

then, for all $y \in Y$, $d_yG(t,x^t,0;y) = 0$, $G(t,x^t,y) = G(t,x^t,0)$, and from (i)

$$\sup_{y\in Y} G(t,x^t,y) = G(t,x^t,0) = \inf_{x\in E(t)} G(t,x,0) = \inf_{x\in X}\sup_{y\in Y} G(t,x,y)$$

and $x^t \in X(t)$. Conversely, for any $x^t \in E(t)\cap X(t)$,

$$G(t,x^t,0) \le \sup_{y\in Y} G(t,x^t,y) = \inf_{x\in X}\sup_{y\in Y} G(t,x,y)$$
$$\inf_{x\in X}\sup_{y\in Y} G(t,x,y) \le \inf_{x\in E(t)}\sup_{y\in Y} G(t,x,y) = \inf_{x\in E(t)}\sup_{y\in Y} G(t,x,0) + d_yG(t,0;y)$$
$$= \inf_{x\in E(t)} G(t,x,0) \le G(t,x^t,0)$$

and $x^t \in E(t)$ is a minimizer of $G(t,x,0)$ with respect to $E(t)$.

Finally, $E(t)\cap X(t) \neq \varnothing$ implies $E(t) \neq \varnothing$ and $X(t) \neq \varnothing$. Conversely, $E(t) \neq \varnothing$ and $X(t) \neq \varnothing$ implies $E(t)\cap X(t) \neq \varnothing$ from part (iii). □

Hypothesis (H0).

Let X be a vector space.

(i) For all $t \in [0,\tau]$, $x^0 \in X(0)$, $x^t \in X(t)$, and $y \in Y$, the function

$$s \mapsto G(t,x^0 + s(x^t - x^0), y)\colon [0,1] \to \mathbb{R} \tag{15}$$

is absolutely continuous. This implies that, for almost all s, the derivative exists and is equal to $d_xG(t,x^0 + s(x^t - x^0), y; x^t - x^0)$ and that it is the integral of its derivative. In particular,

$$G(t,x^t,y) = G(t,x^0,y) + \int_0^1 d_xG(t,x^0 + s(x^t - x^0), y; x^t - x^0)\, ds. \tag{16}$$

(ii) For all $t \in [0,\tau]$, $x^0 \in X(0)$, $x^t \in X(t)$, $y \in Y$, $\varphi \in X$, and almost all $s \in (0,1)$, $d_xG(t, x^0 + s(x^t - x^0), y; \varphi)$ exists and the function $s \mapsto d_xG(t, x^0 + s(x^t - x^0), y; \varphi)$ belongs to $L^1(0,1)$.

Definition 2.3. Given $x^0 \in X(0)$ and $x^t \in X(t)$, the *averaged adjoint state equation* is defined as follows[1]: find $y^t \in Y$ such that for all $\varphi \in X$,

$$\int_0^1 d_xG(t, x^0 + s(x^t - x^0), y^t; \varphi)\, ds = 0. \tag{17}$$

The set of solutions will be denoted $Y(t, x^0, x^t)$. At $t = 0$, $Y(0, x^0, x^0)$ reduces to the usual set of *adjoint states* associated with x^0:

$$Y(0, x^0) \overset{\text{def}}{=} \left\{p \in Y : \forall \varphi \in X,\, d_xG(0, x^0, p; \varphi) = 0\right\}. \tag{18}$$

An important consequence of the introduction of the averaged adjoint state is the following identity: for all $x^0 \in X(0)$, $x^t \in X(t)$, and $y^t \in Y(t, x^0, x^t)$,

$$g(t) = G(t, x^t, y^t) = G(t, x^0, y^t). \tag{19}$$

This is readily seen by substituting y^t into equation (16).

2.4. Summary of the results

We present three theorems for the existence and the characterization of

$$dg(0) \overset{\text{def}}{=} \lim_{t \searrow 0} \frac{g(t) - g(0)}{t}. \tag{20}$$

Note that when the set $X(0)$ is not a singleton, a right-hand side derivative in 0 is often the best that can be expected. The three cases give the existence of $dg(0)$ but differ in its *explicit expression*.

(i) Theorem 4.1 yields: there exists $\hat{x}^0 \in X(0)$ and $\hat{y}^0 \in Y(0, \hat{x}^0)$ such that

$$dg(0) = d_tG(0, \hat{x}^0, \hat{y}^0) = \sup_{y \in Y(0,\hat{x}^0)} d_tG(0, \hat{x}^0, y) = \inf_{x \in X(0)} \sup_{y \in Y(0,x)} d_tG(0, x, y).$$

The expression of $dg(0)$ does not depend on a specific choice of $\hat{x}^0 \in X(0)$ and $\hat{y}^0 \in Y(0, \hat{x}^0)$.

(ii) Theorem 4.2 with weaker hypotheses (H4) yields: there exists $\hat{x}^0 \in X(0)$ and $\hat{y}^0 \in Y(0, \hat{x}^0)$ such that

$$dg(0) = d_tG(0, \hat{x}^0, \hat{y}^0) = \sup_{y \in Y(0,\hat{x}^0)} d_tG(0, \hat{x}^0, y).$$

The explicit expression of $dg(0)$ depends on the choice of some $\hat{x}^0 \in X(0)$, except when $X(0)$ is a singleton (Corollary 4.3).

[1]We adopt the terminology *state* for x and *adjoint state* for the Lagrange multiplier y. It comes from *Control Theory.* We could also speak of an *averaged Lagrange multiplier* or *dual variable* but the dual problem is usually irrelevant in the non-convex case.

(iii) Theorem 4.4 merging hypotheses (H3) and (H4) yields: there exist $\hat{x}^0 \in X(0)$ and $\hat{y}^0 \in Y(0, \hat{x}^0)$ such that

$$dg(0) = d_t G(0, \hat{x}^0, \hat{y}^0).$$

The explicit expression of $dg(0)$ depends on the choice of some pair $(\hat{x}^0, \hat{y}^0)$, except when $X(0) = \{\hat{x}^0\}$ and $Y(0, \hat{x}^0) = \{\hat{y}^0\}$ are singletons (Theorem 4.5 which is a local version of the one originally given in [31, 32]).

We shall see in section 4.4 that, when $X(0) = \{\hat{x}^0\}$ and $Y(0, \hat{x}^0)$ are singletons, the hypotheses of the three theorems coincides.

3. Singleton case revisited: New condition and new term

Prior to dealing with the multivalued case, it is useful to briefly revisit the singleton case ([31], [32, Thm. 3.1]). In the process we introduce a new condition (H3), a slight relaxation of (H2) to a local t-differentiability condition at $t = 0$, and a new term in the conclusion.

Theorem 3.1. *Consider the Lagrangian functional*

$$(t, x, y) \mapsto G(t, x, y) \colon [0, \tau] \times X \times Y \to \mathbb{R}, \quad \tau > 0,$$

where X and Y are vector spaces and the function $y \mapsto G(t, x, y)$ is affine. Let (H0) *and the following hypotheses be satisfied:*

(H1) *for all $t \in [0, \tau]$, $g(t)$ is finite, $X(t) = \{x^t\}$ and $Y(t, x^0, x^t) = \{y^t\}$ are singletons;*

(H2) *$d_t G(0, x^0, y^0)$ exists;*

(H3) *the following limit exists*

$$R(0, x^0, y^0) \overset{\text{def}}{=} \lim_{t \searrow 0} d_y G\left(t, x^0, 0; \frac{y^t - y^0)}{t}\right). \tag{21}$$

Then, $dg(0)$ exists and $dg(0) = d_t G(0, x^0, y^0) + R(0, x^0, y^0)$.

Proof. Since $g(t)$ is finite, $E(t) \neq \varnothing$. Moreover, since $X(t) \neq \varnothing$, $X(t)$ is the set of minimizers of $G(t, x, 0)$ over $E(t)$. Recalling that $g(t) = G(t, x^t, y^t) = G(t, x^0, y^t)$, we conclude

$$\begin{aligned}
g(t) - g(0) &= G(t, x^0, y^t) - G(0, x^0, y^0) \\
&= G(t, x^0, y^0) + d_y G(t, x^0, 0; y^t - y^0) - G(0, x^0, y^0)
\end{aligned}$$

$$\Rightarrow \ (g(t) - g(0))/t = d_y G\left(t, x^0, 0; (y^t - y^0)/t\right) + (G(t, x^0, y^0) - G(0, x^0, y^0))/t$$

$$\begin{aligned}
\Rightarrow \ dg(0) &= \lim_{t \searrow 0} d_y G\left(t, x^0, 0; (y^t - y^0)/t\right) + d_t G(0, x^0, y^0) \\
&= R(0, x^0, y^0) + d_t G(0, x^0, y^0)
\end{aligned}$$

from hypotheses (H2) and (H3). $\square$

Remark 3.2. In [31, 32] it was further assumed that $d_tG(t,x^0,y)$ exists for all $t\in[0,\tau]$ and $y\in Y$. Hence, there exists $\theta_t\in(0,1)$ such that

$$\begin{aligned} G(t,x^0,y^t)-G(0,x^0,y^0) &= G(0,x^0,y^t)+t\,d_tG(\theta_t t,x^0,y^t)-G(0,x^0,y^0)\\ &= d_yG(0,x^0,0;y^t-y^0)+t\,d_tG(\theta_t t,x^0,y^t)\\ &= t\,d_tG(\theta_t t,x^0,y^t)\end{aligned}$$

since $d_yG(0,x^0,0;y^t-y^0)=0$. This led to the following choice of hypothesis (H3) in [32, Thm. 3.1, p. 2024]:

$$\lim_{s\searrow 0,\ t\searrow 0} d_tG(s,x^0,y^t)=d_tG(0,x^0,y^0) \tag{22}$$

which implies that $R(0,x^0,y^0)=0$ in condition (21). So assuming $R(0,x^0,y^0)=0$ in Theorem 3.1 is a weaker form of condition (22) leading to the same conclusion: $dg(0)=d_tG(0,x^0,y^0)$.

4. Multivalued case

4.1. Main theorem

This first theorem establishes the existence of $dg(0)$ and of a pair $(\hat{x}^0,\hat{y}^0)$ which is a solution of a new minimax problem for $d_tG(t,x,y)$ over $\{(x,y): x\in X(0),\ y\in Y(0,x)\}$.

Theorem 4.1. *Consider the Lagrangian*

$$(t,x,y)\mapsto G(t,x,y)\colon [0,\tau]\times X\times Y\to\mathbb{R},\quad \tau>0,$$

where X and Y are vector spaces, the function $y\mapsto G(t,x,y)$ is affine. Let (H0) *and the following hypotheses be satisfied:*

(H1) *for all $t\in[0,\tau]$, $X(t)\neq\varnothing$, $g(t)$ is finite, and for all $x^t\in X(t)$ and $x^0\in X(0)$, $Y(t,x^0,x^t)\neq\varnothing$;*

(H2) *for all x in $X(0)$ and $y\in Y(0,x)$, $d_tG(0,x,y)$ exists;*

(H3) *for each sequence $t_n\to 0$, $0<t_n\le\tau$, there exists $x^0\in X(0)$ such that for all $y^0\in Y(0,x^0)$, there exist a subsequence $\{t_{n_k}\}$ of $\{t_n\}$, $x^{t_{n_k}}\in X(t_{n_k})$, and $y^{t_{n_k}}\in Y(t_{n_k},x^0,x^{t_{n_k}})$ such that*

$$\liminf_{k\to\infty} d_yG\left(t_{n_k},x^0,0;\frac{y^{t_{n_k}}-y^0}{t_{n_k}}\right)\ge 0\,; \tag{23}$$

(H4) *for each sequence $t_n\to 0$, $0<t_n\le\tau$ and all $x^0\in X(0)$, there exist $y^0\in Y(0,x^0)$, a subsequence $\{t_{n_k}\}$ of $\{t_n\}$, $x^{t_{n_k}}\in X(t_{n_k})$, and $y^{t_{n_k}}\in Y(t_{n_k},x^0,x^{t_{n_k}})$ such that*

$$\limsup_{k\to\infty} d_yG\left(t_{n_k},x^0,0;\frac{y^{t_{n_k}}-y^0}{t_{n_k}}\right)\le 0. \tag{24}$$

Then, $dg(0)$ exists and there exist $\hat{x}^0 \in X(0)$ and $\hat{y}^0 \in Y(0, \hat{x}^0)$ such that

$$dg(0) = d_t G(0, \hat{x}^0, \hat{y}^0) = \sup_{y \in Y(0,\hat{x}^0)} d_t G(0, \hat{x}^0, y) = \inf_{x \in X(0)} \sup_{y \in Y(0,x)} d_t G(0, x, y).$$

Proof. (i) Since $g(t)$ is finite, $E(t) \neq \varnothing$. Moreover, since $X(t) \neq \varnothing$, $X(t)$ is the set of minimizers of $G(t, x, 0)$ over $E(t)$. By Hypothesis (H1) and Lemma 2.2,

$$\forall x^t \in X(t),\, \forall y \in Y, \quad g(t) = G(t, x^t, y)$$

and from identity (19), for all $x^0 \in X(0)$ and all $x^t \in X(t)$,

$$\forall y^t \in Y(t, x^0, x^t), \quad g(t) = G(t, x^0, y^t).$$

The differential quotient can now be written as follows: for all $x^0 \in X(0)$, $y^0 \in Y(0, x^0)$, and $\forall y^t \in Y(t, x^0, x^t)$,

$$\begin{aligned} \frac{g(t) - g(0)}{t} &= \frac{G(t, x^0, y^t) - G(0, x^0, y^0)}{t} \\ &= d_y G\left(t, x^0, 0; \frac{y^t - y^0}{t}\right) + \frac{G(t, x^0, y^0) - G(0, x^0, y^0)}{t}. \end{aligned}$$

At this juncture, introduce the liminf and limsup of the differential quotient

$$\underline{d}g(0) \overset{\text{def}}{=} \liminf_{t \searrow 0} \frac{g(t) - g(0)}{t}, \quad \overline{d}g(0) \overset{\text{def}}{=} \limsup_{t \searrow 0} \frac{g(t) - g(0)}{t}.$$

We show that they exist and are equal.

(ii) There exists a sequence $t_n \to 0$, $0 < t_n \leq \tau$, such that $(g(t_n) - g(0))/t_n \to \underline{d}g(0)$. By Hypothesis (H3), there exists $\hat{x}^0 \in X(0)$ such that for all $y^0 \in Y(0, \hat{x}^0)$, there exists a subsequence $\{t_{n_k}\}$ of $\{t_n\}$, there exists $x^{t_{n_k}} \in X(t_{n_k})$ and $y^{t_{n_k}} \in Y(t_{n_k}, \hat{x}^0, x^{t_{n_k}})$ such that

$$\begin{aligned} \underline{d}g(0) &\geq \liminf_{k \to \infty} d_y G\left(t_{n_k}, x^0, 0; \frac{y^{t_{n_k}} - y^0}{t_{n_k}}\right) + \lim_{k \to \infty} \frac{G(t_{n_k}, x^0, y^0) - G(0, x^0, y^0)}{t_{n_k}} \\ &\geq 0 + d_t G(0, \hat{x}^0, y^0). \end{aligned}$$

Therefore: $\exists \hat{x}^0 \in X(0)$ such that $\forall y^0 \in Y(0, \hat{x}^0)$, $\underline{d}g(0) \geq d_t G(0, \hat{x}^0, y^0)$

$\Rightarrow \exists \hat{x}^0 \in X(0)$ such that $\forall y^0 \in Y(0, \hat{x}^0)$, $\underline{d}g(0) \geq d_t G(0, \hat{x}^0, y^0)$

$\Rightarrow \exists \hat{x}^0 \in X(0)$ such that $\underline{d}g(0) \geq \sup_{y \in Y(0,\hat{x}^0)} d_t G(0, \hat{x}^0, y)$

$$\Rightarrow \underline{d}g(0) \geq \sup_{y \in Y(0,\hat{x}^0)} d_t G(0, \hat{x}^0, y) \geq \inf_{x \in X(0)} \sup_{y \in Y(0,x)} d_t G(0, x, y). \tag{25}$$

(iii) There exists a sequence $t_n \to 0$, $0 < t_n \le \tau$, such that

$$\lim_{n\to\infty} (g(t_n) - g(0))/t_n = \overline{d}g(0).$$

By Hypothesis (H4), for all $x^0 \in X(0)$ there exist $y^0 \in Y(0, x^0)$, a subsequence $\{t_{n_k}\}$ of $\{t_n\}$, $x^{t_{n_k}} \in X(t_{n_k})$, and $y^{t_{n_k}} \in Y(t_{n_k}, x^0, x^{t_{n_k}})$ such that

$$\begin{aligned}\overline{d}g(0) &\le \limsup_{k\to\infty} d_y G\left(t_{n_k}, x^0, 0; \frac{y^{t_{n_k}} - y^0}{t_{n_k}}\right) + \lim_{k\to\infty} \frac{G(t_{n_k}, x^0, y^0) - G(0, x^0, y^0)}{t_{n_k}} \\ &\le 0 + d_t G(0, x^0, y^0).\end{aligned}$$

Therefore, for all $x^0 \in X(0)$ there exists $y^0 \in Y(0, x^0)$ such that $\overline{d}g(0) \le d_t G(0, x^0, y^0)$. Since for all $x^0 \in X(0)$ there exists $y^0 \in Y(0, x^0)$ such that

$$\begin{aligned}\overline{d}g(0) \le d_t G(0, x^0, y^0) \quad &\Rightarrow \overline{d}g(0) \le d_t G(0, x^0, y^0) \le \sup_{y\in Y(0,x^0)} d_t G(0, x^0, y) \\ &\Rightarrow \forall x^0 \in X(0), \quad \overline{d}g(0) \le \sup_{y\in Y(0,x^0)} d_t G(0, x^0, y),\end{aligned}$$

and we can take the infimum of the right-hand side over all $x^0 \in X(0)$,

$$\overline{d}g(0) \le \inf_{x^0\in X(0)} \sup_{y\in Y(0,x^0)} d_t G(0, x^0, y). \tag{26}$$

(iv) Combining (26) and (25), there exists $\hat{x}^0 \in X(0)$ such that

$$\underline{d}g(0) \ge \sup_{y\in Y(0,\hat{x}^0)} d_t G(0, \hat{x}^0, y) \ge \inf_{x\in X(0)} \sup_{y\in Y(0,x)} d_t G(0, x, y) \ge \overline{d}g(0) \ge \underline{d}g(0).$$

Therefore, $dg(0)$ exists and there exists $\hat{x}^0 \in X(0)$ such that

$$dg(0) = \sup_{y\in Y(0,\hat{x}^0)} d_t G(0, \hat{x}^0, y) = \inf_{x\in X(0)} \sup_{y\in Y(0,x)} d_t G(0, x, y). \tag{27}$$

But we can get more. From part (iii), we have shown that

$$\exists \hat{x}^0 \in X(0), \forall y \in Y(0, \hat{x}^0), \quad \underline{d}g(0) \ge d_t G(0, \hat{x}^0, y). \tag{28}$$

From part (iii), we have shown that

$$\forall x^0 \in X(0), \exists y^0 \in Y(0, x^0), \quad \overline{d}g(0) \le d_t G(0, x^0, y^0).$$

In particular, for $\hat{x}^0$, there exists $\hat{y}^0 \in Y(0, \hat{x}^0)$ such that $\overline{d}g(0) \le d_t G(0, \hat{x}^0, \hat{y}^0)$. Taking $y = \hat{y}^0$ in (28), $\underline{d}g(0) \ge d_t G(0, \hat{x}^0, \hat{y}^0)$ and

$$\underline{d}g(0) \ge d_t G(0, \hat{x}^0, \hat{y}^0) \ge \overline{d}g(0) \quad \Rightarrow \quad dg(0) = d_t G(0, \hat{x}^0, \hat{y}^0)$$

and the conclusion of the theorem. □

4.2. Theorem 4.2 and its specialization to $X(0)$ a singleton

We now weaken Hypothesis (H4). We still get the existence of $dg(0)$ but its expression depends on the choice of $\hat{x}^0 \in X(0)$ unless $X(0)$ is a singleton.

Theorem 4.2. *Consider the Lagrangian*

$$(t,x,y) \mapsto G(t,x,y)\colon [0,\tau] \times X \times Y \to \mathbb{R}, \quad \tau > 0,$$

where X and Y are vector spaces, the function $y \mapsto G(t,x,y)$ is affine. Let (H0) *and the following hypotheses be satisfied:*

(H1) *for all $t \in [0,\tau]$, $X(t) \neq \varnothing$ and $g(t)$ is finite, and for all $x^t \in X(t)$ and $x^0 \in X(0)$, $Y(t,x^0,x^t) \neq \varnothing$;*

(H2) *for all x in $X(0)$ and all $y \in Y(0,x)$, $d_tG(0,x,y)$ exists;*

(H3) *for each sequence $t_n \to 0$, $0 < t_n \leq \tau$, there exists $\hat{x}^0 \in X(0)$ such that for all $y^0 \in Y(0,\hat{x}^0)$, there exist a subsequence $\{t_{n_k}\}$ of $\{t_n\}$, $x^{t_{n_k}} \in X(t_{n_k})$, and $y^{t_{n_k}} \in Y(t_{n_k},\hat{x}^0,x^{t_{n_k}})$ such that*

$$\liminf_{k\to\infty} d_yG\left(t_{n_k},\hat{x}^0,0;\frac{y^{t_{n_k}}-y^0}{t_{n_k}}\right) \geq 0; \tag{29}$$

(H4) *and there exist $\hat{y}^0 \in Y(0,\hat{x}^0)$, a subsequence $\{t_{n_k}\}$ of $\{t_n\}$, $x^{t_{n_k}} \in X(t_{n_k})$, and $y^{t_{n_k}} \in Y(t_{n_k},\hat{x}^0,x^{t_{n_k}})$ such that*

$$\limsup_{k\to\infty} d_yG\left(t_{n_k},\hat{x}^0,0;\frac{y^{t_{n_k}}-\hat{y}^0}{t_{n_k}}\right) \leq 0. \tag{30}$$

Then, $dg(0)$ exists and there exists $\hat{x}^0 \in X(0)$ and $\hat{y}^0 \in Y(0,\hat{x}^0)$ such that

$$dg(0) = d_tG(0,\hat{x}^0,\hat{y}^0) = \sup_{y\in Y(0,\hat{x}^0)} d_tG(0,\hat{x}^0,y).$$

If $X(0)$ is a singleton, the above hypotheses are equivalent to the ones of Theorem 4.1 *and the expression of $dg(0)$ is intrinsic.*

Proof. Similar to the proof of Theorem 4.1 with obvious changes. □

For convenience, we rewrite Theorem 4.2 below in case $X(0)$ is a singleton.

Corollary 4.3. *Consider the Lagrangian*

$$(t,x,y) \mapsto G(t,x,y)\colon [0,\tau] \times X \times Y \to \mathbb{R}, \quad \tau > 0,$$

where X and Y are vector spaces, the function $y \mapsto G(t,x,y)$ is affine. Let (H0) *and the following hypotheses be satisfied:*

(H1) *for all $t \in [0,\tau]$, $X(t)$ is not empty and $g(t)$ is finite, $X(0) = \{\hat{x}^0\}$ is a singleton and for all $x^t \in X(t)$, $Y(t,\hat{x}^0,x^t) \neq \varnothing$;*

(H2) *for all* $y \in Y(0, \hat{x}^0)$, $d_t G(0, \hat{x}^0, y)$ *exists;*

(H3) *for each sequence* $t_n \to 0$, $0 < t_n \leq \tau$, *and all* $y^0 \in Y(0, \hat{x}^0)$, *there exist a subsequence* $\{t_{n_k}\}$ *of* $\{t_n\}$, $x^{t_{n_k}} \in X(t_{n_k})$, *and* $y^{t_{n_k}} \in Y(t_{n_k}, \hat{x}^0, x^{t_{n_k}})$ *such that*

$$\liminf_{k\to\infty} d_y G\left(t_{n_k}, \hat{x}^0, 0; \frac{y^{t_{n_k}} - y^0}{t_{n_k}}\right) \geq 0; \tag{31}$$

(H4) *for each sequence* $t_n \to 0$, $0 < t_n \leq \tau$, *there exist* $\hat{y}^0 \in Y(0, \hat{x}^0)$, *a subsequence* $\{t_{n_k}\}$ *of* $\{t_n\}$, $x^{t_{n_k}} \in X(t_{n_k})$, *and* $y^{t_{n_k}} \in Y(t_{n_k}, \hat{x}^0, x^{t_{n_k}})$ *such that*

$$\limsup_{k\to\infty} d_y G\left(t_{n_k}, \hat{x}^0, 0; \frac{y^{t_{n_k}} - \hat{y}^0}{t_{n_k}}\right) \leq 0. \tag{32}$$

Then, $dg(0)$ *exists and there exists* $\hat{y}^0 \in Y(0, \hat{x}^0)$ *such that*

$$dg(0) = d_t G(0, \hat{x}^0, \hat{y}^0) = \sup_{y \in Y(0,\hat{x}^0)} d_t G(0, \hat{x}^0, y).$$

4.3. Theorem 4.4 merging Hypotheses (H3) and (H4)

From Hypothesis (H3), Hypothesis (H4) in Theorem 4.2 can be rewritten as

Assumption (A1): For each sequence $t_n \to 0$, $0 < t_n \leq \tau$, there exists $\hat{x}^0 \in X(0)$ and $\hat{y}^0 \in Y(0, \hat{x}^0)$, a subsequence $\{t_{n_k}\}$ of $\{t_n\}$, $x^{t_{n_k}} \in X(t_{n_k})$, and $y^{t_{n_k}} \in Y(t_{n_k}, \hat{x}^0, x^{t_{n_k}})$ such that

$$\lim_{k\to\infty} d_y G\left(t_{n_k}, \hat{x}^0, 0; \frac{y^{t_{n_k}} - \hat{y}^0}{t_{n_k}}\right) = 0. \tag{33}$$

When $X(0)$ and $Y(0, \hat{x}^0)$ are not singletons, this condition alone without (H3) is too weak, since it will yield a pair (x^0, y^0) for the sequence associated with $\underline{d}g(0)$ and a (possibly different) pair for the sequence associated with $\overline{d}g(0)$. To get the same pair, the order of the first two quantifiers must be changed as follows:

Assumption (A2): There exist $\hat{x}^0 \in X(0)$ and $\hat{y}^0 \in Y(0, \hat{x}^0)$ such that for each sequence $t_n \to 0$, $0 < t_n \leq \tau$, there exist a subsequence $\{t_{n_k}\}$ of $\{t_n\}$, $x^{t_{n_k}} \in X(t_{n_k})$, and $y^{t_{n_k}} \in Y(t_{n_k}, \hat{x}^0, x^{t_{n_k}})$ such that

$$\lim_{k\to\infty} d_y G\left(t_{n_k}, \hat{x}^0, 0; \frac{y^{t_{n_k}} - \hat{y}^0}{t_{n_k}}\right) = 0. \tag{34}$$

In fact, we don't need subsequences and we shall use the following formulation:

(H3') *There exist* $\hat{x}^0 \in X(0)$ *and* $\hat{y}^0 \in Y(0, \hat{x}^0)$ *such that for each sequence* $t_n \to 0$, $0 < t_n \le \tau$, *there exist* $x^{t_n} \in X(t_n)$ *and* $y^{t_n} \in Y(t_n, \hat{x}^0, x^{t_n})$ *such that*

$$\lim_{n\to\infty} d_y G\left(t_n, \hat{x}^0, 0; \frac{y^{t_n} - \hat{y}^0}{t_n}\right) = 0. \tag{35}$$

With this somewhat *stronger hypothesis*, we get the existence of $dg(0)$ and of a pair $(\hat{x}^0, \hat{y}^0)$ such that $dg(0) = d_t G(0, \hat{x}^0, \hat{y}^0)$.

Theorem 4.4. *Consider the Lagrangian*

$$(t, x, y) \mapsto G(t, x, y) : [0, \tau] \times X \times Y \to \mathbb{R}, \quad \tau > 0,$$

where X *and* Y *are vector spaces, the function* $y \mapsto G(t, x, y)$ *is affine. Let* (H0) *and the following hypotheses be satisfied:*

(H1) *for all* t *in* $[0, \tau]$, $X(t)$ *is not empty and* $g(t)$ *is finite, and for all* $x^t \in X(t)$ *and* $x^0 \in X(0)$, $Y(t, x^0, x^t)$ *is not empty;*

(H2) *for all* x *in* $X(0)$ *and all* $y \in Y(0, x)$, $d_t G(0, x, y)$ *exists;*

(H3') there exist $\hat{x}^0 \in X(0)$ *and* $\hat{y}^0 \in Y(0, \hat{x}^0)$ *such that for each sequence* $t_n \to 0$, $0 < t_n \le \tau$, *there exist* $x^{t_n} \in X(t_n)$ *and* $y^{t_n} \in Y(t_n, \hat{x}^0, x^{t_n})$ *such that*

$$\lim_{n\to\infty} d_y G\left(t_n, \hat{x}^0, 0; \frac{y^{t_n} - \hat{y}^0}{t_n}\right) = 0. \tag{36}$$

Then, $dg(0)$ *exists and there exist* $\hat{x}^0 \in X(0)$ *and* $\hat{y}^0 \in Y(0, \hat{x}^0)$ *such that*

$$dg(0) = d_t G(0, \hat{x}^0, \hat{y}^0).$$

When the sets $X(0) = \{x^0\}$ *and* $Y(0, x^0) = \{y^0\}$ *are both singletons, the above hypotheses are equivalent to the ones of Theorems* 4.1 *and* 4.2.

Proof. Similar to the proof of Theorem 4.1 with obvious changes. □

4.4. Theorem 4.5: Specialization of Theorem 4.4 to the case when $X(0)$ and $Y(0, x^0)$ are singletons

The hypotheses of all the theorems in this section coincide when $X(0)$ and $Y(0, \hat{x}^0)$ are singletons. The next theorem differs slightly from Theorem 3.1 since, for $t > 0$, the set $X(t)$ and, for each $x^t \in X(t)$, the sets $Y(t, \hat{x}^0, x^t)$ need not be singletons.

Theorem 4.5. *Consider the Lagrangian functional*

$$(t, x, y) \mapsto G(t, x, y) : [0, \tau] \times X \times Y \to \mathbb{R}, \quad \tau > 0,$$

where X *and* Y *are vector spaces, the function* $y \mapsto G(t, x, y)$ *is affine. Let* (H0) *and the following hypotheses be satisfied:*

(H1) *for all* $t \in [0,\tau]$*,* $X(t) \neq \varnothing$ *and* $g(t)$ *is finite,* $X(0) = \{\hat{x}^0\}$ *is a singleton, for all* $x^t \in X(t)$*,* $Y(t, \hat{x}^0, x^t) \neq \varnothing$*, and* $Y(0, \hat{x}^0) = \{\hat{y}^0\}$ *is a singleton;*

(H2) $d_t G(0, \hat{x}^0, \hat{y}^0)$ *exists;*

(H3') *for each sequence* $t_n \to 0$*,* $0 < t_n \leq \tau$*, there exist* $x^{t_n} \in X(t_n)$ *and* $y^{t_n} \in Y(t_n, \hat{x}^0, x^{t_n})$ *such that*

$$\lim_{n\to\infty} d_y G\left(t_n, \hat{x}^0, 0; \frac{y^{t_n} - \hat{y}^0}{t_n}\right) = 0. \tag{37}$$

Then, $dg(0)$ *exists and* $dg(0) = d_t G(0, \hat{x}^0, \hat{y}^0)$.

Proof. Similar to the proof of Theorem 4.1 with obvious changes. □

5. Finite Dimensional Quadratic Programming Examples.

"Despite the fact that curvature of the objective function is constant and that constraints are linear, quadratic problems exhibit all basic difficulties you may encounter in global optimization..." (I. M. Bomze [2]). So, finite dimensional quadratic programming examples will be used *as toy problems* to illustrate the main theorems. What motivated this choice is that complete necessary and sufficient optimality conditions are available by going to second order optimality conditions.

For infinite dimensional examples (a semi-linear problem, an electrical impedance tomography problem, a model for distortion compensation in elasticity, a quasi-linear problem describing electro-magnetic fields, along with numerical simulations), the reader is referred to K. Sturm [31, Chapter 5] and [32]. A classical convex quadratic example is also provided by the *Stokes equations* for the velocity of the fluid where the Lagrange multiplier for the incompressibility constraint is the *pressure* which is specified up to a constant.

5.1. Quadratic objective function and affine constraints

Let Q be an $n \times n$ symmetric matrix and B be an $m \times n$ matrix, q an n-vector, and b an m-vector. Consider the problem

$$\inf_{x \in E} f(x), \quad f(x) \stackrel{\text{def}}{=} \frac{1}{2} Qx \cdot x - q \cdot x, \quad E \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n : Bx = b\}, \tag{38}$$

where $\cdot$ denotes the inner product in the appropriate Euclidean space. The set E is an affine (convex) subset of $\mathbb{R}^n$.

For quadratic objective functions (not necessarily convex on $\mathbb{R}^n$) with affine equality or inequality constraints, the theorem of Frank and Wolfe [29] (see, for instance, [10, Thm. 7.10, p. 223]) says that, for a *feasible problem*, either there exists a minimizer or the infimum is $-\infty$.

If Q is positive semidefinite on $\mathbb{R}^n$, the problem is convex and the necessary and sufficient conditions for the existence is: there exists $\hat{x} \in \mathbb{R}^n$ such that

$$B\hat{x} = b \quad \text{and} \quad \forall x \text{ such that } Bx = b, \quad (Q\hat{x} - q) \cdot (x - \hat{x}) \geq 0.$$

This is equivalent to

$$\forall x \in \ker B, \quad (Q\hat{x} - q) \cdot x = 0 \quad \iff \quad Q\hat{x} - q \in (\ker B)^\perp$$

or, since $(\ker B)^\perp = \operatorname{Im} B^\top$,[2]

$$\exists (\hat{x}, \hat{p}) \in \mathbb{R}^n \times \mathbb{R}^m \text{ such that } \begin{bmatrix} Q & B^\top \\ B & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} q \\ b \end{bmatrix}. \tag{39}$$

Equivalently, $(\hat{x}, \hat{p})$ is a saddle point of the convex-concave Lagrangian

$$G(x, y) \stackrel{\text{def}}{=} \frac{1}{2} Qx \cdot x - q \cdot x + y \cdot (Bx - b). \tag{40}$$

I. Babuska [1] in 1972 and F. Brezzi [5] in 1974 gave necessary and sufficient conditions for the existence of a unique minimizer in the convex case (Q positive semi-definite on $\mathbb{R}^n$) that only requires $Qx \cdot x > 0$ for all $x \neq 0$ in $\ker B$.

When Q is not semidefinite positive on $\mathbb{R}^n$, the objective function is not convex on $\mathbb{R}^n$, there is no saddle point, the dual problem cannot be used, and condition (39) is only necessary. The following theorem gives general necessary and sufficient conditions (see, for instance, [4], [10, Thm. 7.11, p. 227]).

Theorem 5.1. (i) *Problem* (38) *has a minimizer in* $\mathbb{R}^n$ *if and only if* [3]

$$\forall x \in \ker B, \quad Qx \cdot x \geq 0, \quad b \in \operatorname{Im} B, \ \textit{and} \ \ q \in \operatorname{Im}(Q|_E) + \operatorname{Im} B^\top. \tag{41}$$

(ii) *Problem* (38) *has a unique minimizer in* $\mathbb{R}^n$ *if and only if*

$$\exists \beta > 0, \ \forall x \in \ker B, \quad Qx \cdot x \geq \beta \, \|x\|^2 \quad \textit{and} \quad b \in \operatorname{Im} B. \tag{42}$$

(iii) *Problem* (38) *has a unique minimizer in* $\mathbb{R}^n$ *and the associated Lagrange multiplier* $p \in \mathbb{R}^m$ *solution of* (39) *is unique if and only if*

$$\exists \beta > 0, \ \forall x \in \ker B, \quad Qx \cdot x \geq \beta \, \|x\|^2 \quad \textit{and} \quad B \textit{ is surjective.} \tag{43}$$

Corollary 5.2. (i) *Under condition* (41)*, for all* $(q, b) \in (\operatorname{Im}(Q|_E) + \operatorname{Im} B^\top) \times \operatorname{Im} B$*, problem* (38) *has a minimizer and the minimizer and the associated Lagrange multiplier are unique up to an element of* $\ker \mathcal{Q}$*, where*

$$\mathcal{Q} \stackrel{\text{def}}{=} \begin{bmatrix} Q & B^\top \\ B & 0 \end{bmatrix}. \tag{44}$$

[2] The transpose of a matrix M is denoted $M^\top$ and the orthogonal of a set $U \subset \mathbb{R}^n$ is defined as $U^\perp = \{y \in \mathbb{R}^n : \forall x \in U, \, y \cdot x = 0\}$.

[3] $\operatorname{Im}(Q|_E) = \{Qx : x \in E\} = \{Qx : x \in \mathbb{R}^n \text{ and } Bx = b\}$.

(ii) *Under condition* (42)*,*

$$\ker \begin{bmatrix} Q & B^\top \\ B & 0 \end{bmatrix} = \{0\} \times \ker B^\top, \quad \operatorname{Im} \begin{bmatrix} Q & B^\top \\ B & 0 \end{bmatrix} = \mathbb{R}^n \times \operatorname{Im} B.$$

and, for all $(q, b) \in \mathbb{R}^n \times \operatorname{Im} B$*, problem* (38) *has a unique minimizer and the associated Lagrange multiplier* p *is unique up to an element of* $\ker B^\top$*.*

(iii) *Under condition* (43)*, the matrix* $\mathcal{Q}$ *is invertible and for all* $(q, b) \in \mathbb{R}^n \times \mathbb{R}^m$*, problem* (38) *has a unique minimizer in* $\mathbb{R}^n$ *and a unique Lagrange multiplier* p *.*

Example 5.3. As an example of case (iii), pick

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad q \in \mathbb{R}.$$

The problem is not convex and we do not have a saddle point. See also the example of sections 5.3 and 5.4.

Example 5.4 (Mixed finite element schemes for Stokes equations). In the approximation of the solution of the *Stokes equations*, the matrix B is associated with the discretization of the divergence operator (see Fortin-Mghazli [28]). The strict positivity of Q on $\ker B$ is essential to construct mixed finite element schemes. For applications of Theorem 5.1 to the computation of shape gradients for mixed finite element formulations see Delfour-Mghazli-Fortin [13].

5.2. Analysis of the perturbed problems

Given $t \geq 0$, let $Q(t)$ be an $n \times n$ symmetric matrix function and $B(t)$ be an $m \times n$ matrix, $q(t)$ an n-vector and $b(t)$ an m-vector functions. Given $t \geq 0$, consider the family of *perturbed problems* near $t = 0$

$$\inf_{\substack{x \in \mathbb{R}^n \\ B(t)x = b(t)}} \frac{1}{2} Q(t)x \cdot x - q(t) \cdot x. \tag{45}$$

The objective is to start from one of the necessary and sufficient conditions of Theorem 5.1 at 0 and construct perturbed problems that would preserve that local condition at $t = 0$ for small $t > 0$. As can be readily seen, case (i) is not obvious. However, in cases (ii) and (iii) we have the following results that do not require any condition on $b(t)$ and $q(t)$.

Lemma 5.5. *Assume that* $Q(t)$ *and* $B(t)$ *are continuous at* $t = 0$*.*

(i) *Assume that*

$$\exists \beta > 0, \forall x \in \ker B(0), \quad Q(0)x \cdot x \geq \beta \|x\|^2. \tag{46}$$

Then, there exist $\tau > 0$ *and* $\beta' > 0$ *such that, for all* t*,* $0 \leq t \leq \tau$*,*

$$\forall x \in \ker B(t), \quad Q(t)x \cdot x \geq \beta' \|x\|^2. \tag{47}$$

(ii) *Assume that*

$$\exists \beta > 0,\ \forall x \in \ker B(0),\ Q(0)x \cdot x \geq \beta \|x\|^2 \text{ and } \operatorname{Im} B(0) = \mathbb{R}^m. \tag{48}$$

Then, there exist $\tau > 0$ and $\beta' > 0$ such that, for all t, $0 \leq t \leq \tau$,

$$\forall x \in \ker B(t), \quad Q(t)x \cdot x \geq \beta' \|x\|^2, \quad \operatorname{Im} B(t) = \mathbb{R}^m, \tag{49}$$

and the symmetric matrix

$$\mathcal{Q}(t) \stackrel{\text{def}}{=} \begin{bmatrix} Q(t) & B(t)^\top \\ B(t) & 0 \end{bmatrix} \tag{50}$$

is invertible. In particular, problem (45) *has a unique minimizer x^t and a unique Lagrange multiplier p^t solution of the optimality system*

$$\begin{bmatrix} Q(t) & B(t)^\top \\ B(t) & 0 \end{bmatrix} \begin{bmatrix} x^t \\ p^t \end{bmatrix} = \begin{bmatrix} q(t) \\ b(t) \end{bmatrix}. \tag{51}$$

Remark 5.6. Case (iii) of Theorem 5.1 does not require any condition on $q(t)$ and $b(t)$, but case (ii) requires that $b(t) \in \operatorname{Im} B(t)$.

Proof. (i) By contradiction: for all $n \geq 1$, there exist $0 < t_n < \tau$ and $0 \neq x^{t_n} \in \ker B(t_n)$ such that

$$Q(t_n)x^{t_n} \cdot x^{t_n} < \frac{1}{n}\|x^{t_n}\|^2 \quad \Rightarrow Q(t_n)\frac{x^{t_n}}{\|x^{t_n}\|} \cdot \frac{x^{t_n}}{\|x^{t_n}\|} < \frac{1}{n}.$$

By compactness of the unit sphere, there exists x^*, $\|x^*\| = 1$, and a subsequence, still denoted $\{x^{t_n}\}$, such that $x^{t_n}/\|x^{t_n}\| \to x^*$,

$$Q(t_n)\frac{x^{t_n}}{\|x^{t_n}\|} \cdot \frac{x^{t_n}}{\|x^{t_n}\|} \to Q(0)x^* \cdot x^* \leq 0 \quad \text{and} \quad 0 = B(t_n)\frac{x^{t_n}}{\|x^{t_n}\|} \to B(0)x^*.$$

So $x^* \in \ker B(0)$ and we get the contradiction $0 < \beta \leq Q(0)x^* \cdot x^* \leq 0$.

(ii) From part (i) and the fact that the set of invertible $(n+m) \times (n+m)$ matrices is an open subset of the set of all $(n+m) \times (n+m)$ matrices. □

5.3. Application of Theorem 4.1

It is easy to construct a simple non-convex example where $X(0)$ and $Y(0)$ are not singletons. Given two (scalar) directions u and v, consider for each $t \geq 0$

the following perturbations of $b(0)$ and $Q(0)$:

$$B(t) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad b(t) = \begin{bmatrix} tu \\ 0 \\ 0 \end{bmatrix}, \quad Q(t) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & (1+tv) & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad q(t) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

$$G(t,x,y) = \frac{1}{2} Q(t)x \cdot x - q(t) \cdot x + y \cdot [B(t)x - b(t)]$$

$$G(t,(x_1,x_2,x_3),(y_1,y_2,y_3)) = \frac{1+tv}{2}(x_2)^2 - x_2 + y_1 [x_2 - tu] - \frac{1}{2}(x_3)^2 + y_3 x_3$$

$$d_t G(t,x,y) = \frac{1}{2}(x_2)^2 v - y_1 u.$$

At $t = 0$, there is no saddle point and the *dual problem* does not help since

$$\sup_{y \in \mathbb{R}^3} \inf_{x \in \mathbb{R}^3} G(0,x,y) = -\infty < 0 = \inf_{x \in \mathbb{R}^3} \sup_{y \in \mathbb{R}^3} G(0,x,y) \tag{52}$$

and the duality gap is infinite. It is readily seen that $\operatorname{Im} B(t) = \mathbb{R} \times \{0\} \times \mathbb{R}$, $\ker B(t) = \{(a,0,0) : a \in \mathbb{R}\}$, and

$$\forall \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} \in \ker B(t), \qquad \begin{bmatrix} 0 & 0 & 0 \\ 0 & (1+tv) & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} = 0.$$

As for the second set of conditions

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & (1+tv) & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} tu \\ 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} a \\ tu \\ 0 \end{bmatrix}, \ a \in \mathbb{R} \ \text{ and } \begin{bmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \end{bmatrix} = \begin{bmatrix} c \\ 1-(1+t\,v)tu \\ 0 \end{bmatrix}, \ c \in \mathbb{R}.$$

Hence, the conditions of Theorem 5.1 (i) are satisfied and we have existence but not uniqueness of the pair of solutions $(x(t), p(t))$.

The hypotheses of Theorem 4.1 (i) are satisfied. For the averaged adjoint $y(t)$:

$$Q(t)\frac{x(t)+x(0)}{2}+B(t)^\top y(t)=q(t)$$

$$\begin{bmatrix}0 & 0 & 0\\ 0 & (1+tv) & 0\\ 0 & 0 & -1\end{bmatrix}\begin{bmatrix}a\\ tu/2\\ 0\end{bmatrix}+\begin{bmatrix}0&0&0\\1&0&0\\0&0&1\end{bmatrix}\begin{bmatrix}y_1(t)\\y_2(t)\\y_3(t)\end{bmatrix}=\begin{bmatrix}0\\1\\0\end{bmatrix}$$

$$\begin{matrix}(1+tv)t\,u/2+y_1(t)=1\\ y_3(t)=0\end{matrix}\quad\Rightarrow\quad\begin{bmatrix}y_1(t)\\y_2(t)\\y_3(t)\end{bmatrix}=\begin{bmatrix}1-(1+tv)tu/2\\ c\\ 0\end{bmatrix}.$$

So, $Y(t,x(0),x(t))$ is independent of $(x(0),x(t))\in X(0)\times X(t)$:

$$X(t)=\left\{\begin{bmatrix}a\\tv\\0\end{bmatrix}:a\in\mathbb{R}\right\},\quad Y(t,x(0),x(t))=\left\{\begin{bmatrix}1-(1+tv)tu/2\\ c\\ 0\end{bmatrix}:c\in\mathbb{R}\right\}$$

and

$$\begin{aligned}dg(0)&=\inf_{\begin{bmatrix}x_1(0)\\x_2(0)\\x_3(0)\end{bmatrix}=\begin{bmatrix}a\\0\\0\end{bmatrix},\ a\in\mathbb{R}}\ \sup_{\begin{bmatrix}y_1(0)\\y_2(0)\\y_3(0)\end{bmatrix}=\begin{bmatrix}1\\c\\0\end{bmatrix},\ c\in\mathbb{R}}\ \frac{1}{2}\,x_2(0)^2\,v-y_1(0)\,u\\ &=\inf_{\begin{bmatrix}x_1(0)\\x_2(0)\\x_3(0)\end{bmatrix}=\begin{bmatrix}a\\0\\0\end{bmatrix},\ a\in\mathbb{R}}\ \frac{1}{2}\,x_2(0)^2v-u=-u.\end{aligned}$$

The $dg(0)$ arises from the perturbation $b(t)=(tu,0,0)$ of $b(0)=(0,0,0)$; the perturbation $1+tv$ of the (2.2)-entry 1 in the matrix $Q(0)$ has no effect. Note that the resulting semidifferential $(u,v)\mapsto -u:\mathbb{R}^2\to\mathbb{R}$ is linear which means Gateaux differentiability. An additional argument would be required to assert that it is Fréchet differentiable and that the chain rule is applicable to its composition with a Hadamard semidifferentiable function.

5.4. Application of Theorem 4.5: $X(0)=\{x^0\}$ and $Y(0,x^0)$ both singletons

Given $t\geq 0$, let $Q(t)$ be an $n\times n$ symmetric matrix function and $B(t)$ be an $m\times n$ matrix, $q(t)$ an n-vector and $b(t)$ an m-vector functions. Given $t\geq 0$, consider the *perturbed problems* near $t=0$

$$\inf_{\substack{x\in\mathbb{R}^n\\ B(t)x=b(t)}}\ \frac{1}{2}\,Q(t)x\cdot x-q(t)\cdot x. \tag{53}$$

If $Q(t)$ is semidefinite positive on $\mathbb{R}^n$, the problem is convex-concave. Instead, we assume that the hypotheses of Theorem 5.1 (iii) are satisfied at $t = 0$. Therefore, $B(0)$ is surjective and $Q(0)x \cdot x \geq \beta \|x\|^2$ for all $x \in \ker B(0)$ for some $\beta > 0$. If $n \geq m$, $\ker B(0) \neq \mathbb{R}^n$ and $Q(0)$ is not necessarily positive definite on $\mathbb{R}^n$. As a result, the problem is no longer convex and the Theorem of Correa and Seeger [6] cannot be applied.

We now show that Theorem 4.5 applies providing a simple finite dimensional example of the pertinence of the averaged adjoint. We first study the perturbed problems.

Theorem 5.7. *Assume that the right-hand side derivatives $Q' = Q'(0^+)$, $B' = B'(0^+)$, $q' = q'(0^+)$, and $b' = b'(0^+)$ exist in $t = 0^+$ and that the hypotheses of Theorem* 5.1 (iii) *are satisfied at $t = 0$.*

(i) *The hypotheses of Theorem* 5.1 (iii) *are satisfied for $t \geq 0$ sufficiently small: the symmetric matrix $\mathcal{Q}(t)$ defined in* (50) *is invertible and*

$$\forall x \in \ker B(t),\, x \neq 0, \quad Q(t)x \cdot x > 0. \tag{54}$$

In particular, problem (53) *has a unique minimizer x^t and the associated multiplier p^t solution of the optimality system*

$$\begin{bmatrix} Q(t) & B(t)^\top \\ B(t) & 0 \end{bmatrix} \begin{bmatrix} x^t \\ p^t \end{bmatrix} = \begin{bmatrix} q(t) \\ b(t) \end{bmatrix} \tag{55}$$

is unique.

(ii) *The pair (x^t, p^t) is differentiable from the right and the derivatives (x', p') are unique solutions of*

$$\begin{bmatrix} Q(0) & B(0)^\top \\ B(0) & 0 \end{bmatrix} \begin{bmatrix} x' \\ p' \end{bmatrix} = - \begin{bmatrix} Q' & B'^\top \\ B' & 0 \end{bmatrix} \begin{bmatrix} x^0 \\ p^0 \end{bmatrix} + \begin{bmatrix} q' \\ b' \end{bmatrix}, \tag{56}$$

where the pair (x^0, p^0) is the solution of

$$\begin{bmatrix} Q(0) & B(0)^\top \\ B(0) & 0 \end{bmatrix} \begin{bmatrix} x^0 \\ p^0 \end{bmatrix} = \begin{bmatrix} q(0) \\ b(0) \end{bmatrix}. \tag{57}$$

Proof. From Theorem 5.1 (iii), Lemma 5.5, and Corollary 5.2 (iii). □

The associated *Lagrangian* is

$$G(t, x, y) \stackrel{\text{def}}{=} \frac{1}{2} Q(t)x \cdot x - q(t) \cdot x + y \cdot (B(t)x - b(t)). \tag{58}$$

Since it is quadratic, we have all the differentiability we need

$$d_t G(0, x, y) = \frac{1}{2}\, Q'x \cdot x - q' \cdot x + y \cdot (B'x - b') \tag{59}$$

$$d_y G(t, x, y; \psi) = \psi \cdot (B(t)x - b(t)) \tag{60}$$

$$d_x G(t, x, y; \phi) = Q(t)x \cdot \phi - q(t) \cdot \phi + y \cdot B(t)\phi. \tag{61}$$

Given $x^0 \in X(0)$, $x^t \in X(t)$, and $\varphi \in \mathbb{R}^n$, the average adjoint equation is given by

$$\forall \varphi \in \mathbb{R}^n, \quad \int_0^1 d_x G(t, x^0 + s(x^t - x^0), y^t; \varphi)\, ds = 0\,. \tag{62}$$

From the previous expressions

$$\begin{aligned} d_x G(t, x^0{+}s(x^t - x^0), y^t; \varphi) &= \\ &= Q(t)(x^0 + s(x^t - x^0)) \cdot \varphi - q(t) \cdot \varphi + y^t \cdot B(t)\varphi \end{aligned} \tag{63}$$

and upon integration

$$\begin{aligned} \int_0^1 d_x G(t, x^0{+}s(x^t - x^0), y^t; \varphi)\, ds &= \\ &= \left[Q(t) \frac{x^t + x^0}{2} - q(t) + B(t)^\top y^t \right] \cdot \varphi = 0. \end{aligned} \tag{64}$$

Finally, the *state* x^t and *averaged adjoint state* y^t are given by the equations

$$\left\{ \begin{aligned} Q(t)\frac{x^t + x^0}{2} + B(t)^\top y^t - q(t) = 0 \\ B(t)x^t - b(t) = 0 \end{aligned} \right\} \quad \left\{ \begin{aligned} Q(0)x^0 + B(0)^\top y^0 - q(0) = 0 \\ B(0)x^0 - b(0) = 0 \end{aligned} \right\} \tag{65}$$

Theorem 5.8. *Given $t \geq 0$, let $Q(t)$ be an $n \times n$ symmetric matrix function and $B(t)$ be an $m \times n$ matrix, $q(t)$ an n-vector and $b(t)$ an m-vector functions. Assume that the right-hand-side derivatives $Q' = Q'(0^+)$, $B' = B'(0^+)$, $q' = q'(0^+)$, and $b' = b'(0^+)$ exist in $t = 0^+$ and that the hypotheses of Theorem* 5.1 (iii) *are satisfied at $t = 0$: $B(0)$ is surjective and there exists $\beta > 0$ such that $Q(0)x \cdot x \geq \beta \|x\|^2$ for all $x \in \ker B(0)$.*

(i) *For t sufficiently small, the minimization problem* (53) *has a unique minimizer x^t and the associated multiplier p^t is unique.*

(ii) *The coupled equations* (65) *of the state and averaged adjoint state has a unique solution (x^t, y^t) and the pair (x^t, y^t) is differentiable from the right. The derivative (x', y') of the pair (x^t, y^t) is the unique solution of*

$$\begin{bmatrix} Q(0) & B(0)^\top \\ B(0) & 0 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = - \begin{bmatrix} Q' & B'^\top \\ B' & 0 \end{bmatrix} \begin{bmatrix} x^0 \\ y^0 \end{bmatrix} + \begin{bmatrix} q' \\ b' \end{bmatrix}, \tag{66}$$

where (x^0, p^0) is the solution of

$$\begin{bmatrix} Q(0) & B(0)^\top \\ B(0) & 0 \end{bmatrix} \begin{bmatrix} x^0 \\ y^0 \end{bmatrix} = \begin{bmatrix} q(0) \\ b(0) \end{bmatrix}. \tag{67}$$

(iii) *The function $g(t)$ is right differentiable at $t = 0$ and*

$$dg(0) = \frac{1}{2} Q' x^0 \cdot x^0 - q' \cdot x^0 + y^0 \cdot [B' x^0 - b'], \tag{68}$$

where (x^0, y^0) is the solution of (67).

Proof. (i) From Theorem 5.7.

(ii) The equations (65) for (x^t, y^t) can be rewritten in the form

$$\begin{bmatrix} Q(t) & B(t)^\top \\ B(t) & 0 \end{bmatrix} \begin{bmatrix} x^t/2 \\ y^t \end{bmatrix} = \begin{bmatrix} q(t) - Q(t)x^0/2 \\ b(t)/2 \end{bmatrix}, \quad \begin{bmatrix} Q(0) & B(0)^\top \\ B(0) & 0 \end{bmatrix} \begin{bmatrix} x^0 \\ y^0 \end{bmatrix} = \begin{bmatrix} q(0) \\ b(0) \end{bmatrix}.$$

In view of the invertibility of the matrix appearing in the last set of equations, the pair (x^t, y^t) is unique and $x^t \to x^0$ and $y^t \to y^0$. By direct computation of the differential quotients, the right-hand side derivatives (x', y') exist and are solution of

$$\begin{bmatrix} Q(0) & B(0)^\top \\ B(0) & 0 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = - \begin{bmatrix} Q' & B'^\top \\ B' & 0 \end{bmatrix} \begin{bmatrix} x^0 \\ y^0 \end{bmatrix} + \begin{bmatrix} q' \\ b' \end{bmatrix}.$$

(iii) By direct computation, all hypotheses of Theorem 4.5 are satisfied. □

References

[1] I. Babuska: The finite element method with Lagrange multipliers, Numer. Math. 20 (1972) 179–192.

[2] I. M. Bomze: Global optimization: a quadratic programming perspective, in: Nonlinear Optimization, G. Di Pillo and F. Schoen (eds.), Springer-Verlag, Berlin et al. (2010) 1–54.

[3] J. F. Bonnans, A. Shapiro: Optimization problems with perturbations, a guided tour, SIAM Review 40 (1998) 202–227.

[4] J. F. Bonnans, A. Shapiro: Perturbation Analysis of Optimization Problems, Springer, New York (2000).

[5] F. Brezzi: On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers, R.A.I.R.O. 8(R2) (1974) 129–151.

[6] R. Correa, A. Seeger: Directional derivatives of a minimax function, Nonlinear Anal. Theory Methods and Appl. 9 (1985) 13–22.

[7] J. M. Danskin: The theory of max-min, with applications, SIAM J. Applied Mathematics 14(4) (1966) 641–644.

[8] J. M. Danskin: The Theory of Max-min and its Application to Weapons Allocation Problems, Econometrics and Operations Research, Vol. V, Springer, New York (1967).

[9] M. C. Delfour: Shape derivatives and differentiability of min max, in: Shape Optimization and Free Boundaries (Montreal, PQ, 1990), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. 380, Kluwer Academic Publishers, Dordrecht (1992) 35–111.

[10] M. C. Delfour: Introduction to Optimization and Semidifferential Calculus, MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, Philadelphia (2012).

[11] M. C. Delfour: Metric spaces of shapes and geometries from set parametrized functions, in: New Trends in Shape Optimization, A. Pratelli and G. Leugering (eds.), Int. Ser. Num. Math. 166, Birkhäuser, Basel (2015) 57–101.

[12] M. C. Delfour: Differentials and semidifferentials for metric spaces of shapes and geometries, in: System Modeling and Optimization (CSMO 2015), L. Bociu, J. A. Desideri and A. Habbal (eds.), AICT Series, Springer-Verlag, Berlin et al. (2017), to appear.

[13] M. C. Delfour, Z. Mghazli, J.-P. Zolésio: Computation of shape gradients for mixed finite element formulations, in: Partial Differential Equations Methods in Control and Shape Analysis, G. Da Prato and J. P. Zolésio (eds.), Marcel Dekker, New York (1997) 77–93.

[14] M. C. Delfour, J. Morgan: A complement to the differentiability of saddle points and min-max, Optimal Control Appl. Methods 13(1) (1992) 87–94.

[15] M. C. Delfour, J. Morgan: One-sided derivative of minmax and saddle points with respect to a parameter, Optimization 31(4) (1994) 343–358.

[16] M. C. Delfour, J. P. Zolésio: Dérivation d'un min max et application à la dérivation par rapport au contrôle d'une observation non différentiable de l'état, C. R. Acad. Sci. Paris Sér. Math. 302 (1986) 571–574.

[17] M. C. Delfour, J. P. Zolésio: Differentiability of a min max and application to optimal control and design problems, I, II, in: Control Problems for Systems Described by Partial Differential Equations and Applications (Gainesville 1986), I. Lasiecka and R. Triggiani (eds.), Lecture Notes in Control and Inform. Sci. 97, Springer-Verlag, Berlin (1987) 204–219, 220–229.

[18] M. C. Delfour, J. P. Zolésio: Further development in shape sensitivity analysis via penalization method, in: Boundary Control and Boundary Variations (Nice 1986), Lecture Notes in Comput. Sci. 100, Springer-Verlag, Berlin (1988) 153–191.

[19] M. C. Delfour, J. P. Zolésio: Shape sensitivity analysis via min max differentiability, SIAM J. Control Optimization 26 (1988) 834–862.

[20] M. C. Delfour, J. P. Zolésio: Analyse des problèmes de forme par la dérivation des minimax (French), Ann. Inst. H. Poincaré Anal. Non Linéaire 6 (1989), suppl., 211–227.

[21] M. C. Delfour, J. P. Zolésio: Velocity method and Lagrangian formulation for the computation of the shape Hessian, SIAM J. Control Optimization 29 (1991) 1414–1442.

[22] M. C. Delfour, J. P. Zolésio: Shapes and geometries: metrics, analysis, differential calculus, and optimization, SIAM Series on Advances in Design and Control, Society for Industrial and Applied Mathematics, Philadelphia (2011).

[23] V. F. Dem'janov: Differentiation of the maximin function I, II (Russian), Z. Vyčisl. Mat. i Mat. Fiz. 8 (1968) 1186–1195 and ibid. 9 (1969) 55–67.

[24] V. F. Dem'janov: Minimax: Directional Differentiation (Russian), Izdat. Leningrad Gos. Univ., Leningrad (1974).

[25] V. F. Dem'janov, V. N. Malozemov: Introduction to the Minimax, Translated from the Russian by D. Louvish, Halsted Press, New York et al. (1974).

[26] H. Federer: Curvature measures, Trans. Amer. Math. Soc. 93 (1959) 418–419.

[27] H. Federer: Geometric Measure Theory, Grundlehren der mathematischen Wissenschaften 153, Springer, New York (1969).

[28] M. Fortin, Z. Mghazli: Analyse d'un élément mixte pour le problème de Stokes I: Résultats généraux, II: Construction et estimations d'erreur (French), Numer. Math. 62(2) (1992) 149–160, 161–188.

[29] M. Frank, P. Wolfe: An algorithm for quadratic programming, Naval Res. Logist. Quart. 3 (1956) 95–110.

[30] B. Lemaire: Problèmes Min-max et Applications au Contrôle Optimal de Systèmes Gouvernés par des Équations aux Dérivées Partielles Linéaires, Thèse de Doctorat d'État, Université de Montpellier, Montpellier (1970).

[31] K. Sturm: On Shape Optimization with Non-linear Partial Differential Equations, Doctoral Thesis, Technische Universität Berlin (2014).

[32] K. Sturm: Minimax Lagrangian approach to the differentiability of non-linear PDE constrained shape functions without saddle point assumption, SIAM J. Control Optimization 53 (2015) 2017–2039.