Parametric semidifferentiability of minimax of Lagrangians: averaged adjoint state approach*

Michel C. Delfour[†] Kevin Sturm[‡]

Centre de recherches mathématiques,
Université de Montréal,
Montréal, Canada

Rapport CRM-3351

Montréal, le 9 novembre 2015

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

[†]Centre de recherches mathématiques and Département de mathématiques et de statistique, Université de Montréal, CP 6128, succ. Centre-ville, Montéal (Qc), Canada H3C 3J7 (delfour@crm.umontreal.ca).

[‡]Universität Duisburg-Essen, Fakultät für Mathematik, Thea-Leymann-straße 9, D-45127 Essen, Germany (kevin.sturm@uni-due.de).

Abstract

A standard approach to the *minimization* of an *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*, that is, some form of linear penalization of the equality constraints or the state equation. The *multipliers* and the *adjoint state variable* plays the same role. The initial minimization problem is equivalent to the minimax of the associated Lagrangian. We consider the semidifferentiability of the minimax of a Lagrangian with respect to a positive parameter as a first step towards the computation of semidifferentials and differentials. By using the new notion of *averaged adjoint equation* introduced by K. Sturm [30] [31], the minimax problem need not be related to a saddle point as in Correa and Seeger [4] 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 usually occurs and only a semidifferential is available even if the functions at hand are infinitely differentiable as was illustrated in the seminal paper of J. M. Danskin [5] in 1966. The results have broad applications in Mathematical Programming, Control Theory, and Shape/Topological Optimization.
Keywords: ***Minimax, Lagrangian, sensitivity analysis, semidifferentiability, averaged adjoint, Shape and Topological Optimization, Optimal Control, Mathematical Programming.***
AMS(MOS) subject classifications. 49K20, 49K27, 49K35, 49K40, 49Q10, 49Q12.

1 Introduction

By parametric differentiability or semidifferentiability, we mean differentiability or semidifferentiability with respect to a set of parameters. Theorems on the parametric differentiability of a minimax with or without saddle point have a long history. J. M. Danskin [5] in 1966 and [6] 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 usually occurs and only a semidifferential is available even if the functions at hand are infinitely differentiable. He was followed by Dem'janov (cf. Dem'janov and Malozemov [23]) in 1968. V. F. Dem'janov [21]) also studied the differentiation of the maximin function with or without the saddle point assumption in 1968 and published a book [22] on that topic in 1974. Extensions to infinite dimensional problems were given in the thesis of B. Lemaire [28] in 1970.

In 1986 Delfour and Zolésio [13] introduced a theorem on the differentiability of a minimax in the context of optimal control and extended it in 1987 to shape semidifferentiability in [14]. 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 [4] 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 [16] 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 [11, Theorem 3] in 1992 and extended to ε-solutions in [12] in 1994. In [17] an interesting penalization method was introduced where the state is solution of a variational inequality (thermal radiator).

Recently, K. Sturm [30] [31] 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 equation* by an *averaged adjoint state equation.* 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 [5], 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 [30, Thm. 3.2] and [31, Thm. 3.1] that the sets of solutions be singletons to the multi-valued case that was considered in [16] and [11]. We also relax the global differentiability assumptions to local ones. 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 main applications are to control and shape sensitivity analysis as well as to problems in mathematical programming. 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 [19, 20]).

Throughout the paper, we use the terminology *state equation* and *adjoint state equation.* Yet, it applies to problems in Mathematical Programming where the state equation is replaced by *equality constraints* that characterize the *set of feasible solutions* and the adjoint state by the *Lagrange multipliers.*

2 Lagrangian and averaged adjoint state equation

2.1 Abstract framework to include metric spaces of shapes and geometries

The results will be presented in an abstract framework with ready-to-check hypotheses. This set up of the parametric minimax problem is necessary and justified by the fact that, in general, the space of *parameters* will not be a locally convex topological space. This situation typically occurs in the computation of the so-called *shape and topological derivatives* which turn out to be semi-differentials or differentials defined on complete metric spaces with at best a group structure.[1] The reader is referred to Delfour and Zolésio [20] for the *Courant metrics* and metrics constructed from set-parametrized functions and to the recent tutorial paper of Delfour [9] 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 ccntain both the empty set and the hold-all in their completion.

2.2 Definitions and problem formulation

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

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

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*

$$t \mapsto g(t) \stackrel{\text{def}}{=} \inf_{x\in X} \sup_{y\in Y} G(t, x, y) : [0, \tau] \to \mathbb{R}\,. \tag{1}$$

Notation 1. When the limits exist and X is open we shall use the following compact notation:

$$dg(0) \stackrel{\text{def}}{=} \lim_{t\searrow 0} \frac{g(t) - g(0)}{t} \tag{2}$$

$$d_tG(0, x, y) \stackrel{\text{def}}{=} \lim_{t\searrow 0} \frac{G(t, x, y) - G(0, x, y)}{t} \tag{3}$$

$$v \in X, \quad d_xG(t, x, y; v) \stackrel{\text{def}}{=} \lim_{\theta\searrow 0} \frac{G(t, x + \theta v, y) - G(t, x, y)}{\theta} \tag{4}$$

$$w \in Y, \quad d_yG(t, x, y; w) \stackrel{\text{def}}{=} \lim_{\theta\to 0} \frac{G(t, x, y + \theta w) - G(t, x, y)}{\theta}. \tag{5}$$

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

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

$$\forall w \in Y, \quad d_yG(t, x, y; w) = G(t, x, w) - G(t, x, 0). \tag{6}$$

The variational equation

$$\text{to find } x^t \in X \text{ such that for all } \varphi \in Y,\ d_yG(t, x^t, 0; \varphi) = 0 \tag{7}$$

[1] The notion of differential on an Abelian group was studed by Fréchet [27] in 1948.

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

$$E(t) \overset{\text{def}}{=} \left\{x^t \in X : \forall \varphi \in Y,\ d_y G(t, x^t, 0; \varphi) = 0\right\} \tag{8}$$

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

$$X(t) \overset{\text{def}}{=} \left\{x^t \in X : \sup_{y \in Y} G(t, x^t, y) = g(t)\right\}. \tag{9}$$

Lemma 1 (Constrained infimum and minimax)**.** (i) *In general,*

$$\inf_{x \in X} \sup_{y \in Y} G(t, x, y) = \inf_{x \in E(t)} G(t, x, 0). \tag{10}$$

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

(iii) *Assume that $E(t) \neq \varnothing$. Then there exists $x^t \in E(t)$ such that*

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

if and only $x^t \in X(t)$. In particular, $X(t) \subset E(t)$.

Proof. (i) Since $y \mapsto G(t, x, y)$ is affine, for all (t, x, y)

$$G(t, x, y) = G(t, x, 0) + d_y G(t, x, 0; y)$$

$$\sup_{y \in Y} G(t, x, y) = G(t, x, 0) + \sup_{y \in Y} d_y G(t, x, 0; y) = \begin{cases} G(t, x, 0), & \text{if } x \in E(t) \\ +\infty, & \text{if } x \in X \backslash E(t) \end{cases}$$

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

(ii) If $g(t) = +\infty$, then for all $x \in X$, $\sup_{y \in Y} G(t, x, y) = +\infty$ and

$$+\infty = G(t, x, 0) + \sup_{y \in Y} d_y G(t, x, 0; y) \quad \Rightarrow \sup_{y \in Y} d_y G(t, x, 0; y) = +\infty$$

$$\Rightarrow \exists y \in Y \text{ such that } d_y G(t, x, 0; y) > 0 \quad \Rightarrow \forall x \in X,\ x \notin E(t)$$

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 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 $y \in Y$, $d_y G(t, x^t, 0; y) = 0$, $G(t, x^t, y) = G(t, x^t, 0)$, and

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

From part (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, since $E(t) \neq \varnothing$, we get from part (ii) that $g(t) < +\infty$. If $X(t) \neq \varnothing$, for every $x^t \in X(t)$,

$$-\infty < G(t,x^t,0) \le \sup_{y\in Y} G(t,x^t,y) = g(t) = \inf_{x\in E(t)} G(t,x,0) \le G(t,x^t,0)$$

and $g(t)$ is finite and equal to $G(t,x^t,0)$. But

$$\begin{aligned} g(t) = \sup_{y\in Y} G(t,x^t,y) &= G(t,x^t,0) + \sup_{y\in Y} d_yG(t,x,0;y) \\ \Rightarrow\ 0 = g(t) - G(t,x^t,0) &= \sup_{y\in Y} d_yG(t,x^t,0;y). \end{aligned}$$

Since $y \mapsto d_yG(t,x^t,0;y)$ is linear, for all $y\in Y$, $d_yG(t,x^t,0;y) = 0$ and $x^t \in E(t)$, that is $X(t) \subset E(t)$. □

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

$$s \mapsto G(t, x^0 + s(x^t - x^0), y) : [0,1] \to \mathbb{R} \tag{11}$$

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,

$$G(t,x^t,y) = G(t,x^0,y) + \int_0^1 d_xG(t, x^0 + s(x^t - x^0), y; x^t - x^0)\, ds. \tag{12}$$

(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 1. Given $x^0 \in X(0)$ and $x^t \in X(t)$, the *averaged adjoint state equation* is defined as follows:[2]

$$\begin{aligned} &\text{to find } y^t \in Y \text{ such that for all } \varphi \in X, \\ &\int_0^1 d_xG(t, x^0 + s(x^t - x^0), y^t; \varphi)\, ds = 0. \end{aligned} \tag{13}$$

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

$$Y(0,x^0) \overset{\text{def}}{=} \left\{ p \in Y : \forall \varphi \in X,\ d_xG(0,x^0,p;\varphi) = 0 \right\}. \tag{14}$$

[2]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.

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 $y^t \in Y(t, x^0, x^t)$,

$$\boxed{g(t) = G(t, x^t, y^t) = G(t, x^0, y^t).} \tag{15}$$

This is readily seen by substituting y^t into equation (12).

We prove three theorems for the existence and the characterization of

$$dg(0) \stackrel{\text{def}}{=} \lim_{t \searrow 0} \frac{g(t) - g(0)}{t}. \tag{16}$$

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

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

The framed 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 2 with weaker hypotheses yields: there exists $\hat{x}^0 \in X(0)$ and $\hat{y}^0 \in Y(0, \hat{x}^0)$ such that

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

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.

(iii) Theorem 4 with another set of weaker hypotheses yields: there exist $\hat{x}^0 \in X(0)$ and $\hat{y}^0 \in Y(0, \hat{x}^0)$ such that

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

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.

We shall see in section 2.6 that, when $X(0) = \{\hat{x}^0\}$ and $Y(0, \hat{x}^0)$ are singletons, the hypotheses of the three theorems coincides.

2.3 First theorem.

This first theorem establishes the existence of $dg(0)$ and gives an expression independent of the choice of the pairs $(\hat{x}^0, \hat{y}^0)$ solution of the minimax problem.

Theorem 1. *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 $t \in [0,\tau]$, $X(t) \neq \varnothing$, $g(t)$ 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 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 $x^0 \in X(0)$ 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*

$$\liminf_{k\to\infty} \frac{G(t_{n_k}, x^0, y^{t_{n_k}}) - G(0,x^0,y^0)}{t_{n_k}} \geq d_tG(0,x^0,y^0);$$

(H4) *for each sequence $t_n \to 0$, $0 < t_n \leq \tau$ 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*

$$\limsup_{k\to\infty} \frac{G(t, x^0, y^{t_{n_k}}) - G(0,x^0,y^0)}{t_{n_k}} \leq d_tG(0,x^0,y^0).$$

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

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

Proof. (i) By Hypothesis (H1) and Lemma 1,

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

and from identity (15), 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 $\forall y^t \in Y(t,x^0,x^t)$,

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

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

$$\underline{d}g(0) \overset{\text{def}}{=} \liminf_{t\searrow 0} \frac{g(t)-g(0)}{t} \quad \text{and} \quad \overline{d}g(0) \overset{\text{def}}{=} \limsup_{t\searrow 0} \frac{g(t)-g(0)}{t}. \tag{17}$$

It is sufficient to show that they are equal to get the existence of $dg(0)$.

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

$$\frac{g(t_n) - g(0)}{t_n} \to \underline{d}g(0).$$

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

$$\liminf_{k\to\infty} \frac{G(t_{n_k}, \hat{x}^0, y^{t_{n_k}}) - G(0, \hat{x}^0, y^0)}{t_{n_k}} \ge d_t G(0, \hat{x}^0, y^0).$$

Therefore, there exists $\hat{x}^0 \in X(0)$ such that for all $y^0 \in Y(0, \hat{x}^0)$

$$\begin{aligned}
&\underline{d}g(0) \ge d_t G(0, \hat{x}^0, y^0) \\
&\Rightarrow \exists \hat{x}^0 \in X(0) \text{ such that } \forall y^0 \in Y(0, \hat{x}^0), \quad \underline{d}g(0) \ge d_t G(0, \hat{x}^0, y^0) \\
&\Rightarrow \exists \hat{x}^0 \in X(0) \text{ such that } \quad \underline{d}g(0) \ge \sup_{y\in Y(0,\hat{x}^0)} d_t G(0, \hat{x}^0, y)
\end{aligned}$$

$$\Rightarrow \boxed{\underline{d}g(0) \ge \sup_{y\in Y(0,\hat{x}^0)} d_t G(0, \hat{x}^0, y) \ge \inf_{x\in X(0)} \sup_{y\in Y(0,x)} d_t G(0, x, y).} \tag{18}$$

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

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

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

$$\lim_{k\to\infty} \frac{g(t_{n_k}) - g(0)}{t_{n_k}} = \limsup_{k\to\infty} \frac{G(t_{n_k}, x^0, y^{t_{n_k}}) - G(0, x^0, y^0)}{t_{n_k}} \le d_t G(0, x^0, y^0).$$

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 $x^0 \in X(0)$,

$$\boxed{\overline{d}g(0) \le \inf_{x^0\in X(0)} \sup_{y\in Y(0,x^0)} d_t G(0, x^0, y).} \tag{19}$$

(iv) Combining (19) and (18), there exists $\hat{x}^0 \in X(0)$ such that

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

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

$$\boxed{dg(0) = \sup_{y\in Y(0,\hat{x}^0)} d_tG(0,\hat{x}^0,y) = \inf_{x\in X(0)} \sup_{y\in Y(0,x)} d_tG(0,x,y).} \tag{20}$$

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

$$\exists \hat{x}^0 \in X(0),\ \forall y \in Y(0,\hat{x}^0), \quad \underline{d}g(0) \geq d_tG(0,\hat{x}^0,y). \tag{21}$$

From part (ii), we have shown that

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

In particular, for $\hat{x}^0$, there exists $\hat{y}^0 \in Y(0,\hat{x}^0)$ such that

$$\overline{d}g(0) \leq d_tG(0,\hat{x}^0,\hat{y}^0).$$

Taking $y = \hat{y}^0$ in (21), $\underline{d}g(0) \geq d_tG(0,\hat{x}^0,\hat{y}^0)$ and

$$\underline{d}g(0) \geq d_tG(0,\hat{x}^0,\hat{y}^0) \geq \overline{d}g(0) \quad \Rightarrow \boxed{dg(0) = d_tG(0,\hat{x}^0,\hat{y}^0)}$$

and the conclusion of the theorem. □

2.4 Second theorem.

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 2. *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 $t \in [0,\tau]$, $X(t) \neq \varnothing$ 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;*

for each sequence $t_n \to 0$, $0 < t_n \leq \tau$, there exists $\hat{x}^0 \in X(0)$ such that

(H3) *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*

$$\liminf_{k\to\infty} \frac{G(t_{n_k},\hat{x}^0,y^{t_{n_k}}) - G(0,\hat{x}^0,y^0)}{t_{n_k}} \geq d_tG(0,\hat{x}^0,y^0)$$

(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*

$$\limsup_{k\to\infty} \frac{G(t_{n_k},\hat{x}^0,y^{t_{n_k}}) - G(0,\hat{x}^0,\hat{y}^0)}{t_{n_k}} \leq d_t G(0,\hat{x}^0,\hat{y}^0).$$

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*

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

When $X(0)$ *is a singleton, the above hypotheses are equivalent to the ones of Theorem* 1 *and the expression of* $dg(0)$ *is intrinsic.*

Proof. Similar to the proof of Theorem 1 with obvious changes. □

We rewrite Theorem 2 below when $X(0)$ is a singleton.

Theorem 3. *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,* $X(0) = \{\hat{x}^0\}$ *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*

$$\liminf_{k\to\infty} \frac{G(t_{n_k},\hat{x}^0,y^{t_{n_k}}) - G(0,\hat{x}^0,y^0)}{t_{n_k}} \geq d_t G(0,\hat{x}^0,y^0);$$

(H4) *for each sequence* $t_n \to 0$, $0 < t_n \leq \tau$, *there exist* $y^0 \in Y(0,\hat{x}^0)$, *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*

$$\limsup_{k\to\infty} \frac{G(t,\hat{x}^0,y^{t_{n_k}}) - G(0,\hat{x}^0,y^0)}{t_{n_k}} \leq d_t G(0,\hat{x}^0,y^0).$$

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

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

2.5 Third theorem merging Hypotheses (H3) and (H4).

From Hypothesis (H3), Hypothesis (H4) in Theorem 2 can be rewritten as

Assumption 1. For each sequence $t_n \to 0$, $0 < t_n \le \tau$, 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

$$\lim_{k\to\infty} \frac{G(t_{n_k}, \hat{x}^0, y^{t_{n_k}}) - G(0, \hat{x}^0, \hat{y}^0)}{t_{n_k}} = d_t G(0, \hat{x}^0, \hat{y}^0).$$

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.

Assumption 2. 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 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

$$\lim_{k\to\infty} \frac{G(t_{n_k}, \hat{x}^0, y^{t_{n_k}}) - G(0, \hat{x}^0, \hat{y}^0)}{t_{n_k}} = d_t G(0, \hat{x}^0, \hat{y}^0).$$

With this somewhat *stronger hypothesis*, we get *the less precise expression* so far, that is, an existence theorem with an explicit expression of $dg(0)$ that depends on the choice of some $(\hat{x}^0, \hat{y}^0)$.

Theorem 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;*

(H34) *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 exists a subsequence $\{t_{n_k}\}$ of $\{t_n\}$ such that there exist $x^{t_{n_k}} \in X(t_{n_k})$, $y^{t_{n_k}} \in Y(t_{n_k}, \hat{x}^0, x^{t_{n_k}})$ such that*

$$\lim_{k\to\infty} \frac{G(t_{n_k}, \hat{x}^0, y^{t_{n_k}}) - G(0, \hat{x}^0, \hat{y}^0)}{t_{n_k}} = d_t G(0, \hat{x}^0, \hat{y}^0).$$

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 1 *and* 2.

Proof. Similar to the proof of Theorem 1 with obvious changes. □

2.6 Specialization to $X(0)$ and $Y(0, x^0)$ singletons.

The hypotheses of the previous theorems all coincide when $X(0)$ and $Y(0, \hat{x}^0)$ are singletons. The following theorem is a version of the theorem of K. Sturm [30] [31, Thm. 3.1] with only a local differentiability condition.

Theorem 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;*

(H34) *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*

$$\lim_{k \to \infty} \frac{G(t_{n_k}, \hat{x}^0, y^{t_{n_k}}) - G(0, \hat{x}^0, \hat{y}^0)}{t_{n_k}} = d_t G(0, \hat{x}^0, \hat{y}^0).$$

Then, $dg(0)$ exists and

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

Proof. Similar to the proof of Theorem 1 with obvious changes. □

Remark 1. Note that, 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.

Remark 2. When, for all t in $[0, \tau]$, the sets $X(t) = \{x^t\}$ and $Y(t, \hat{x}^0, x^t) = \{y^t\}$ are both singletons, it would be sufficient to prove that $x^t \to \hat{x}^0$ in some topology on X (for instance, strong continuity) and that for all sequences $t_n \to 0$, $0 < t_n \leq \tau$, there is a subsequence of $\{y^{t_n}\}$ that converges to $\hat{y}^0$ in some topology on Y (for instance, the weak topology of Y). Then all that would be required to conclude is the continuity of the function

$$(t, y) \mapsto d_t G(t, \hat{x}^0, y) : [0, \tau[\times Y \to \mathbb{R} \tag{22}$$

in $(t, y) = (0, \hat{y}^0)$.

2.7 Saddle point case

For the sake of comparison and completeness, we give a slightly relaxed version of the theorem of Correa and Seeger[4] for a minimax with saddle point. Consider a function (not necessarily a Lagrangian)

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

for some $\tau > 0$ and non-empty sets X and Y. For each t in $[0, \tau]$ define

$$(23)\quad g(t) \stackrel{\text{def}}{=} \inf_{x\in X} \sup_{y\in Y} G(t,x,y), \quad X(t) \stackrel{\text{def}}{=} \left\{ x^t \in X : \sup_{y\in Y} G(t,x^t,y) = g(t) \right\}.$$

Similarly, define

$$(24)\quad h(t) \stackrel{\text{def}}{=} \sup_{y\in Y} \inf_{x\in X} G(t,x,y), \quad Y(t) \stackrel{\text{def}}{=} \left\{ y^t \in Y : \inf_{x\in X} G(t,x,y^t) = h(t) \right\}.$$

When X and Y are not empty, we always have the inequality $h(t) \leq g(t)$. To complete the notation, introduce the set of *saddle points*

$$(25)\quad S(t) = \left\{ (x,y) \in X \times Y : \sup_{y'\in Y} G(t,x,y') = G(t,x,y) = \inf_{x'\in X} G(t,x',y) \right\}$$

which may be empty. When X and Y are not empty, we have the following necessary and sufficient condition: $S(t)$ non-empty if and only if $X(t) \times Y(t)$ non-empty and $g(t) = h(t)$. In such a case, $S(t) = X(t) \times Y(t)$.

The following theorem relaxes and localizes the hypotheses of Theorem 2.1 in Correa and Seeger [4] and puts it in a form easier to compare with the previous theorems.

Theorem 6. *Let the sets X and Y, the real number $\tau > 0$, and the function*

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

be given. Assume that the following assumptions hold:

(H1) *$S(0)$ is non empty*[3] *and $g(t) = h(t)$ for all $t \in [0,\tau]$;*

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

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

$$(26)\quad \forall y \in Y(0), \quad \liminf_{k\to\infty} \frac{G(t_{n_k}, x_{n_k}, y) - G(0, x_{n_k}, y)}{t_{n_k}} \geq d_t G(0, x^0, y);$$

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

$$(27)\quad \forall x \in X(0), \quad \limsup_{k\to\infty} \frac{G(t_{n_k}, x, y_{n_k}) - G(0, x, y_{n_k})}{t_{n_k}} \leq d_t G(0, x, y^0).$$

[3] Correa and Seeger [4] use the assumption that $g(t) = h(t)$, $t \in [0,\tau]$ and $X(0) \times Y(0)$ non empty which implies that $S(0) = X(0) \times Y(0)$. Conversely, $S(0)$ non empty implies $X(0) \times Y(0)$ non empty and $g(0) = h(0)$. This assumption does not require that $G(0,x,y)$ be convex-concave in (x,y).

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

$$dg(0) = \sup_{y \in Y(0)} d_t G(0, x^0, y) = d_t G(0, x^0, y^0) = \inf_{x \in X(0)} d_t G(0, x, y^0) \tag{28}$$

and (x^0, y^0) is a saddle point of $d_t G(0, x, y)$ on $X(0) \times Y(0)$.

Proof. (i) Consider the liminf and limsup of the differential quotient of g:

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

We first establish upper and lower bounds on the differential quotient and show that $\underline{d}g(0) = \overline{d}g(0)$.

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

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

By Hypothesis (H3), there exist $x^0 \in X(0)$, a subsequence $\{t_{n_k}\}$ of $\{t_n\}$, and $x_{n_k} \in X(t_{n_k})$ such that

$$\forall y \in Y(0), \quad \liminf_{k \to \infty} \frac{G(t_{n_k}, x_{n_k}, y) - G(0, x_{n_k}, y)}{t_{n_k}} \ge d_t G(0, x^0, y).$$

By definition, since $x_{n_k} \in X(t_{n_k})$ and $y \in Y(0)$

$$\begin{aligned}
g(t_{n_k}) - g(0) = g(t_{n_k}) - h(0) &= \sup_{y' \in Y} G(t_{n_k}, x_{n_k}, y') - \inf_{x' \in X} G(0, x', y) \\
&\ge G(t_{n_k}, x_{n_k}, y) - G(0, x_{n_k} y) \\
\Rightarrow \underline{d}g(0) = \lim_{k \to \infty} \frac{g(t_{n_k}) - g(0)}{t_{n_k}} &= \liminf_{k \to \infty} \frac{g(t_{n_k}) - g(0)}{t_{n_k}} \\
&= \liminf_{k \to \infty} \frac{G(t_{n_k}, x_{n_k}, y) - G(0, x_{n_k}, y)}{t_{n_k}} \\
&\ge d_t G(0, x^0, y).
\end{aligned}$$

Therefore,

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

and

$$\inf_{x \in X(0)} \sup_{y \in Y(0)} d_t G(0, x, y) \le \sup_{y \in Y(0)} d_t G(0, x^0, y) \le \underline{d}g(0). \tag{29}$$

By a dual argument and assumption (H4) we also obtain

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

$$\overline{d}g(0) \le \inf_{x \in X(0)} d_t G(0, x, y^0) \le \sup_{y \in Y(0)} \inf_{x \in X(0)} d_t G(0, x, y), \tag{30}$$

and necessarily

$$\inf_{x\in X(0)} \sup_{y\in Y(0)} d_tG(0,x,y) = \underline{d}g(0) = \overline{d}g(0) = \sup_{y\in Y(0)} \inf_{x\in X(0)} d_tG(0,x,y).$$

In particular, from (29) and (30),

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

and (x^0,y^0) is a saddle point of $d_tG(0,\cdot,\cdot)$. □

3 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 to illustrate the main theorems.

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 [30, Chapter 5] and [31]. 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.

3.1 Quadratic objective function and affine state equation (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

$$\text{(31)} \qquad \inf_{x\in E} f(x), \quad f(x) \overset{\text{def}}{=} \frac{1}{2} Qx\cdot x - q\cdot x, \quad E \overset{\text{def}}{=} \{x\in\mathbb{R}^n : Bx=b\},$$

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 [26] (see, for instance, [8, 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

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

or, since $(\ker B)^\perp = \operatorname{Im} B^\top$,[4]

$$\exists (\hat{x}, \hat{p}) \in \mathbb{R}^n \times \mathbb{R}^m \text{ such that } \begin{bmatrix} Q & B^\top \\ B & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} q \\ b \end{bmatrix}. \tag{32}$$

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

$$G(x, y) \stackrel{\text{def}}{=} \frac{1}{2} Qx \cdot x - q \cdot x + y \cdot (Bx - b). \tag{33}$$

I. Babuska [1] in 1972 and F. Brezzi [3] 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 (32) is only necessary. The following theorem gives general necessary and sufficient conditions (see, for instance, [8, Thm. 7.11, p. 227]).

Theorem 7. (i) *Problem* (31) *has a minimizer in* $\mathbb{R}^n$ *if and only if* [5]

$$\forall x \in \ker B, \quad Qx \cdot x \geq 0, \quad b \in \operatorname{Im} B, \quad \text{and} \quad q \in \operatorname{Im}(Q|_E) + \operatorname{Im} B^\top. \tag{34}$$

(ii) *Problem* (31) *has a unique minimizer in* $\mathbb{R}^n$ *if and only if*

$$\exists \beta > 0, \forall x \in \ker B, \quad Qx \cdot x \geq \beta \|x\|^2 \quad \text{and} \quad b \in \operatorname{Im} B. \tag{35}$$

(iii) *Problem* (31) *has a unique minimizer in* $\mathbb{R}^n$ *and the associated Lagrange multiplier* $p \in \mathbb{R}^m$ *solution of* (32) *is unique if and only if*

$$\exists \beta > 0, \forall x \in \ker B, \quad Qx \cdot x \geq \beta \|x\|^2 \quad \text{and} \quad B \text{ is surjective.} \tag{36}$$

Proof. (i) If $\hat{x} \in E$ is a minimizer, then $b \in \operatorname{Im} B$ and for all $x \in E$,

$$0 \leq f(x) - f(\hat{x}) = (Q\hat{x} - q) \cdot (x - \hat{x}) + \frac{1}{2} Q(x - \hat{x}) \cdot (x - \hat{x}).$$

It is necessary that

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

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

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

From this condition, for all $x \in E$,

$$0 \le f(x) - f(\hat{x}) = \frac{1}{2} Q(x - \hat{x}) \cdot (x - \hat{x}).$$

Since $x - \hat{x} \in \ker B$, we get the first condition (34)

$$f(x) - f(\hat{x}) \ge (Q\hat{x} - q) \cdot (x - \hat{x}) \cdot (x - \hat{x}).$$

Finally, since $Q\hat{x} - q \in (\ker B)^\perp$ and $(\ker B)^\perp = \operatorname{Im}(B^T)$,

$$\exists (\hat{x}, \hat{p}) \in \mathbb{R}^n \times \mathbb{R}^m \text{ such that } \begin{bmatrix} Q & B^\top \\ B & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{p} \end{bmatrix} = \begin{bmatrix} q \\ b \end{bmatrix} \tag{37}$$

and $q \in \operatorname{Im}(Q|_E) + \operatorname{Im} B^\top$.

Conversely, for all pairs $(q, b) \in (\operatorname{Im}(Q|_E) + \operatorname{Im} B^\top) \times \operatorname{Im} B$,

$$\exists (\hat{x}, \hat{p}) \in \mathbb{R}^n \times \mathbb{R}^m \text{ such that } \begin{bmatrix} Q & B^\top \\ B & 0 \end{bmatrix} \begin{bmatrix} x \\ p \end{bmatrix} = \begin{bmatrix} q \\ b \end{bmatrix}$$
$$\Rightarrow \forall x \in E, \quad \nabla f(\hat{x}) \cdot (x - \hat{x}) = 0$$

and, from the first condition (34)

$$f(x) - f(\hat{x}) = \underbrace{(Q\hat{x} - q) \cdot (x - \hat{x})}_{=0} + \frac{1}{2} \underbrace{Q(x - \hat{x}) \cdot (x - \hat{x})}_{\ge 0} \ge 0.$$

Therefore, problem (31) has a minimizer $\hat{x}$ and the pair $(\hat{x}, \hat{p})$ is unique up to an element of

$$\ker \begin{bmatrix} Q & B^\top \\ B & 0 \end{bmatrix}. \tag{38}$$

(ii) Under condition (35), $\ker B \cap \ker Q = \{0\}$ and $(\operatorname{Im} B^\top + \operatorname{Im} Q) = \mathbb{R}^n$.
(iii) Under condition (36), $\ker B \cap \ker Q = \{0\}$, $(\operatorname{Im} B^\top + \operatorname{Im} Q) = \mathbb{R}^n$, and $\operatorname{Im} B = \mathbb{R}^m$. □

Corollary 1. (i) *Under condition* (34)*, for all* $(q, b) \in (\operatorname{Im}(Q|_E) + \operatorname{Im} B^\top) \times \operatorname{Im} B$*, problem* (31) *has a minimizer and the minimizer and the associated Lagrange multiplier are unique up to an element of* (38)*.*

(ii) *Under condition* (35),

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

and, for all $(q, b) \in \mathbb{R}^n \times \operatorname{Im} B$*, problem* (31) *has a unique minimizer and the associated Lagrange multiplier* p *is unique up to an element of* $\ker B^\top$*.*

(iii) *Under condition* (36)*, the matrix*

$$\begin{bmatrix} Q & B^\top \\ B & 0 \end{bmatrix} \tag{39}$$

is invertible and for all $(q,b) \in \mathbb{R}^n \times \mathbb{R}^m$*, problem* (31) *has a unique minimizer in* $\mathbb{R}^n$ *and the associated Lagrange multiplier* p *is also unique.*

Example 1. 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 3.3 and 3.4.

Example 2 (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 [24, 25]). The strict positivity of Q on $\ker B$ is essential to construct mixed finite element schemes. For applications of Theorem 6 to the computation of shape gradients for mixed finite element formulations see Delfour-Mghazli-Fortin [10].

3.2 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 $t = 0$

$$\inf_{\substack{x \in \mathbb{R}^n \\ B(t)x = b(t)}} \frac{1}{2} Q(t)x \cdot x - q(t) \cdot x. \tag{40}$$

The objective is to start from one of the necessary and sufficient conditions of Theorem 7 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 2. *Assume that* $Q(t)$ *and* $B(t)$ *are continuous at* $t = 0$*.*

(i) *Assume that*

$$\exists \beta > 0, \forall x \in \ker B(0), \quad Q(0)x \cdot x \geq \beta \|x\|^2. \tag{41}$$

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

$$\forall x \in \ker B(t), \quad Q(t)x \cdot x \geq \beta' \|x\|^2. \tag{42}$$

(ii) *Assume that*

(43)
$$\exists \beta > 0, \forall x \in \ker B(0), \quad Q(0)x \cdot x \geq \beta \|x\|^2 \quad \textit{and} \quad \operatorname{Im} B(0) = \mathbb{R}^m .$$

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

$$\forall x \in \ker B(t), \quad Q(t)x \cdot x \geq \beta' \|x\|^2, \quad \operatorname{Im} B(t) = \mathbb{R}^m, \tag{44}$$

and the symmetric matrix

$$\begin{bmatrix} Q(t) & B(t)^\top \\ B(t) & 0 \end{bmatrix} \tag{45}$$

is invertible. In particular, problem (40) has a unique minimizer x^t and the corresponding Lagrange multiplier p^t solution of the optimality system

$$\begin{bmatrix} Q(t) & B(t)^\top \\ B(t) & 0 \end{bmatrix} \begin{bmatrix} x^t \\ p^t \end{bmatrix} = \begin{bmatrix} q(t) \\ b(t) \end{bmatrix} \tag{46}$$

is unique.

Remark 3. Case (iii) of Theorem 7 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. □

3.3 Application of Theorem 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 $Q(0)$

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

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

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

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

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

$$\sup_{y \in \mathbb{R}^3} \inf_{x \in \mathbb{R}^3} G(0,x,y) = -\infty < 0 = \inf_{x \in \mathbb{R}^3} \sup_{y \in \mathbb{R}^3} G(0,x,y) \tag{47}$$

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

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

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

Hence, the conditions of Theorem 7 (i) are satisfied and we have existence but not uniqueness of the pair of solutions $(x(t), p(t))$.

The hypotheses of Theorem 1 (i) are satisfied. For the averaged adjoint $y(t)$:

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

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

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

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

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

$$= \inf_{\begin{bmatrix} x_1(0)\\ x_2(0)\\ x_3(0)\end{bmatrix} = \begin{bmatrix} a\\ 0\\ 0\end{bmatrix},\, a\in\mathbb{R}} \frac{1}{2} x_2(0)^2 v - u = -u.$$

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 (Hadamard) differentiable and that the chain rule is applicable to its composition with a Hadamard semidifferentiable function.

3.4 Application of Theorem 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 $t = 0$

$$\inf_{\substack{x\in\mathbb{R}^n\\ B(t)x=b(t)}} \frac{1}{2} Q(t)x \cdot x - q(t) \cdot x. \tag{48}$$

If $Q(t)$ is semidefinite positive on $\mathbb{R}^n$, the problem is convex-concave. Instead, we assume that the hypotheses of Theorem 7 (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 6 (Correa and Seeger) cannot be applied.

We now show that Theorem 5 applies providing a simple finite dimensional example of the pertinence of the averaged adjoint. We first study the perturbed problems.

Theorem 8. *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 7 (iii) are satisfied at $t = 0$.*

(i) *The hypotheses of Theorem 7* (iii) *are satisfied for $t \geq 0$ sufficiently small: the symmetric matrix*

$$\begin{bmatrix} Q(t) & B(t)^\top \\ B(t) & 0 \end{bmatrix} \tag{49}$$

is invertible and

$$\forall x \in \ker B(t),\, x \neq 0, \quad Q(t)x \cdot x > 0. \tag{50}$$

In particular, problem (48) has a unique minimizer x^t *and the associated multiplier* p^t *solution of the optimality system*

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

is unique.

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

$$\begin{bmatrix} Q(0) & B(0)^\top \\ B(0) & 0 \end{bmatrix} \begin{bmatrix} x' \\ p' \end{bmatrix} = -\begin{bmatrix} Q' & B'^\top \\ B' & 0 \end{bmatrix} \begin{bmatrix} x^0 \\ p^0 \end{bmatrix} + \begin{bmatrix} q' \\ b' \end{bmatrix}, \tag{52}$$

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

$$\begin{bmatrix} Q(0) & B(0)^\top \\ B(0) & 0 \end{bmatrix} \begin{bmatrix} x^0 \\ p^0 \end{bmatrix} = \begin{bmatrix} q(0) \\ b(0) \end{bmatrix}. \tag{53}$$

Proof. (i) From Theprem 7 (iii), Lemma **??**, and Corollary 1 (iii).
(ii) Obvious. □

The associated *Lagrangian* is

$$G(t, x, y) \stackrel{\text{def}}{=} \frac{1}{2} Q(t)x \cdot x - q(t) \cdot x + y \cdot (B(t)x - b(t)). \tag{54}$$

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

$$d_t G(0, x, y) = \frac{1}{2} Q'x \cdot x - q' \cdot x + y \cdot (B'x - b') \tag{55}$$

$$d_y G(t, x, y; \psi) = \psi \cdot (B(t)x - b(t)) \tag{56}$$

$$d_x G(t, x, y; \phi) = Q(t)x \cdot \phi - q(t) \cdot \phi + y \cdot B(t)\phi. \tag{57}$$

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

$$\forall \varphi \in \mathbb{R}^n, \int_0^1 d_x G(t, x^0 + s(x^t - x^0), y^t; \varphi)\, ds = 0 \tag{58}$$

From the previous expressions

$$\begin{aligned} &d_x G(t, x^0 + s(x^t - x^0), y^t; \varphi) \\ &= Q(t)(x^0 + s(x^t - x^0)) \cdot \varphi - q(t) \cdot \varphi + y^t \cdot B(t)\varphi \end{aligned} \tag{59}$$

and upon integration

$$\begin{aligned}(60)\qquad &\int_0^1 d_x G(t, x^0 + s(x^t - x^0), y^t; \varphi)\, ds \\ &= \left[Q(t)\frac{x^t + x^0}{2} - q(t) + B(t)^\top y^t\right] \cdot \varphi = 0.\end{aligned}$$

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

$$(61)\qquad \boxed{\begin{array}{l} Q(t)\dfrac{x^t + x^0}{2} + B(t)^\top y^t - q(t) = 0 \\ B(t)x^t - b(t) = 0 \end{array}} \quad \boxed{\begin{array}{l} Q(0)x^0 + B(0)^\top y^0 - q(0) = 0 \\ B(0)x^0 - b(0) = 0 \end{array}}$$

Theorem 9. *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 7 (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* (48) *has a unique minimizer x^t and the associated multiplier p^t is unique.*

(ii) *The coupled equations* (61) *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*

$$(62)\qquad \begin{bmatrix} Q(0) & B(0)^\top \\ B(0) & 0 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = -\begin{bmatrix} Q' & B'^\top \\ B' & 0 \end{bmatrix} \begin{bmatrix} x^0 \\ y^0 \end{bmatrix} + \begin{bmatrix} q' \\ b' \end{bmatrix},$$

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

$$(63)\qquad \begin{bmatrix} Q(0) & B(0)^\top \\ B(0) & 0 \end{bmatrix} \begin{bmatrix} x^0 \\ y^0 \end{bmatrix} = \begin{bmatrix} q(0) \\ b(0) \end{bmatrix}.$$

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

$$(64)\qquad dg(0) = \frac{1}{2}Q'x^0 \cdot x^0 - q' \cdot x^0 + y^0 \cdot [B'x^0 - b'],$$

where (x^0, y^0) is the solution of (63).

Proof. (i) From Theorem 8.

(ii) The equations (61) 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 $y^t \to y^0$. It is easy to show, by direct computation of the differential quotients, that 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, it is easy to check that all hypotheses of Theorem 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., pp, 1–54, Springer-Verlag, Berlin Heidelberg 2010.

[3] 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.

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

[5] J. M. Danskin, *The theory of max-min, with applications*, SIAM J. on Applied Mathematics **14**, no. 4 (1966), 641–644.

[6] J. M. Danskin, *The theory of max-min and its application to weapons allocation problems*, Econometrics and Operations Research, Vol. V, Springer-Verlag New York 1967.

[7] M. C. Delfour, *Shape derivatives and differentiability of min max*, in “Shape optimization and free boundaries (Montreal, PQ, 1990),” pp. 35–111, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 380, Kluwer Acad. Publ., Dordrecht 1992.

[8] M. C. Delfour, *Introduction to optimization and semidifferential calculus*, MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, Philadelphia, USA 2012.

[9] M. C. Delfour, *Metrics spaces of shapes and geometries from set parametrized functions*, in “New Trends in Shape Optimization”, A. Pratelli and G. Leugering, eds., International Series of Numerical Mathematics, Birkhäuser Verlag (Proceedings of the ERC Workshop “New Trends in Shape Optimization”, Universität Erlangen-Nürnberg, September 23–27, 2013), to appear.

[10] M. C. Delfour, Z. Mghazli, and 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, pp. 77–93, Marcel Dekker, New York 1997.

[11] M. C. Delfour and J. Morgan, *A complement to the differentiability of saddle points and min-max*, Optimal Control Appl. Methods **13** (1992), no. 1, 87–94.

[12] M. C. Delfour and J. Morgan, *One-sided derivative of minmax and saddle points with respect to a parameter*, Optimization **31** (1994), no. 4, 343–358.

[13] M. C. Delfour and 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.

[14] M. C. Delfour and 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, Fla., 1986), I. Lasiecka and R. Triggiani, eds., 204–219 and 220–229, Lecture Notes in Control and Inform. Sci., 97, Springer, Berlin, 1987.

[15] M. C. Delfour and J. P. Zolésio, *Further development in shape sensitivity analysis via penalization method*, in "Boundary control and boundary variations" (Nice, 1986), 153–191, Lecture Notes in Comput. Sci., 100, Springer, Berlin 1988.

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

[17] M. C. Delfour and J. P. Zolésio, *Shape sensitivity analysis via a penalization method*, Annali di Matematica Pura ed Applicata **CLI** (1988), 179–212.

[18] M. C. Delfour and J. P. Zolésio, *Analyse des problèmes de forme par la drivation des minimax*, (French) [Analysis of shape problems by minimax differentiation], Ann. Inst. H. Poincaré Anal. Non Linéaire 6 (1989), suppl., 211–227.

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

[20] M. C. Delfour and 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, USA, 2011.

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

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

[23] V. F. Dem'janov and V. N. Malozemov, *Introduction to the Minimax*, Translated from the Russian by D. Louvish, Halsted Press [John Wiley & Sons], New York-Toronto, Ont.; Israel Program for Scientific Translations, Jerusalem-London, 1974. Translated in German: Einführung in Minimax-probleme, Academische Verlagsgesellsch. Geest & Portig, Leipzig, 1975.

[24] M. Fortin and Z. Mghazli, *Analyse d'un élément mixte pour le problème de Stokes. I. Résultats généraux*, (French) [Analysis of a mixed element for the Stokes problem. I. General results], Numer. Math. **62** (1992), 149–160.

[25] M. Fortin and Z. Mghazli, *Analyse d'un élément mixte pour le problème de Stokes. II. Construction et estimations d'erreur* (French) [Analysis of a mixed element for the Stokes problem. II. Construction and error estimates], Numer. Math. **62** (1992), no. 2, 161–188.

[26] M. Frank and P. Wolfe, *An Algorithm for Quadratic Programming*, Naval Res. Log. Quart. **3** (1956), 95–110.

[27] M. Fréchet, *La notion de différentielle sur un groupe Abélien*, Portugaliae Math. **7** (1948). 59–72.

[28] 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, France, 1970.

[29] K. Sturm, *Lagrange method in shape optimization for non-linear partial differential equations: A material derivative free approach*, Preprint No, 1817, WIAS, Berlin 2013.

[30] K. Sturm, *On shape optimization with non-linear partial differential equations*, Doctoral thesis, Technische Universiltät of Berlin, Germany 2014.

[31] K. Sturm, *Minimax Lagrangian approach to the differentiability of non-linear PDE constrained shape functions without saddle point assumption*, accepted for publication in SIAM J. on Control and Optim., May 2015.