SIAM J. Optim.
Vol. 30, No. 1, pp. 319–348

EXISTENCE OF LAGRANGE MULTIPLIERS UNDER GÂTEAUX DIFFERENTIABLE DATA WITH APPLICATIONS TO STOCHASTIC OPTIMAL CONTROL PROBLEMS*

A. JOURANI† AND F. J. SILVA‡

Abstract. The main objective of this work is to study the existence of Lagrange multipliers for infinite dimensional problems under Gâteaux differentiability assumptions on the data. Our investigation follows two main steps: the proof of the existence of Lagrange multipliers under a calmness assumption on the constraints and the study of sufficient conditions, which only use the Gâteaux derivative of the function defining the constraint, that ensure this assumption. We apply the abstract results to show directly the existence of Lagrange multipliers of two classes of standard stochastic optimal control problems.

Key words. Lagrange multipliers, Gâteaux differentiability, calmness, metric regularity, optimality conditions, stochastic optimal control problems

AMS subject classifications. 65K10, 49K27, 93E20

DOI. 10.1137/18M1223411

1. Introduction.

Consider the following optimization problem:

$$\min\{f(x)\,;\;\; g(x)\in D\}, \tag{1.1}$$

where $f : X \to \mathbb{R}$ and $g : X \to Y$ are (for simplicity of the exposition) differentiable mappings, X and Y are Banach spaces, and $D \subseteq Y$ is nonempty. In the case where Y is finite dimensional the following result holds for any closed set D: if x_0 is a local solution to (P), then there exist $\lambda \geq 0$ and $y^* \in N(D, g(x_0))$ such that

$$(\lambda, y^*) \neq (0,0), \tag{1.2}$$

$$\lambda f'(x_0) + y^* \circ g'(x_0) = 0. \tag{1.3}$$

Here $N(D, g(x_0))$ denotes some normal cone to D at $g(x_0)$ (say, for instance, the Clarke normal cone, the approximate normal cone, etc.).

The following example proposed by Brokate in [8, section 2] shows that the previous result is no longer true in the infinite dimensional case.

Example 1. Let $X = Y = \ell^2$ be the Hilbert space of square summable real sequences. Denote by $(e_k)_{k\geq 1}$ the canonical orthonormal base of ℓ^2 and consider the operator $A : \ell^2 \to \ell^2$ defined by

$$A\left(\sum_{i\geq 1} x_i e_i\right) = \sum_{i\geq 1} 2^{1-i} x_i e_i.$$

*Received by the editors October 29, 2018; accepted for publication (in revised form) September 30, 2019; published electronically February 6, 2020.
https://doi.org/10.1137/18M1223411

†Institut de Mathématiques de Bourgogne, UMR 5584, CNRS, Université Bourgogne Franche-Comté, F-21000, Dijon, France (Abderrahim.Jourani@u-bourgogne.fr).
‡Institut de recherche XLIM-DMI, UMR-CNRS 7252, Université de Limoges, 87060 Limoges, France (francisco.silva@unilim.fr).

It is easy to check that A is injective but not surjective and that the image of A, denoted by $\mathrm{Im}(A)$, is a proper dense subspace of ℓ^2. As a consequence, the adjoint operator A^* is injective but not surjective. Now, let $x^* \in \ell^2 \setminus \mathrm{Im}(A^*)$ and consider the function $f : X \to \mathbb{R}$ defined by $f(x) = \langle x^*, x\rangle$, for all $x \in X$, where $\langle \cdot, \cdot \rangle$ denotes the standard scalar product in ℓ^2. Set $g := A$ and $D := \{0\}$. Using this data for problem (1.1), 0 is the only feasible point, and, hence, 0 is the solution of this problem. Using that $x^* \in \ell^2 \setminus \mathrm{Im}(A^*)$, we easily check that there is no $(\lambda, y^*) \neq (0, 0)$ satisfying (1.3).

In infinite dimension, most of the authors have assumed that D is a closed convex cone with a nonempty interior or that $D = D_1 \times \{0\}$, where D_1 is a closed convex cone with a nonempty interior and $\{0\} \subseteq \mathbb{R}^n$ (see [15, 20, 28, 37, 47] and references therein). The first result which gives a condition for the validity of (1.2)–(1.3) in the case where D is closed is due to Jourani and Thibault [31], where it is assumed that the system $g(x) \in D$ is *metrically regular* (see [12, 27] and the references therein for a systematic study of this property). This condition is expressed metrically in terms of g and D and implies that λ can be taken different from zero. In [30], it is shown that relations (1.2)–(1.3) subsist in the case where f is vector-valued and D is epi-Lipschitz-like in the sense of Borwein (see [6]). In [32, 33], the authors gave general conditions ensuring (1.2) and (1.3). More precisely, let x_0 be a local solution to problem (P), and suppose that f and g are locally Lipschitz mappings at x_0, with g strongly compactly Lipschitz at x_0 (see [31]). Denote by $\partial_A d(u, D)$ the approximate subdifferential of $d(\cdot, D)$ at u (see [24, 25]), and assume the existence of a locally compact cone $K^* \subseteq Y^*$ and a neighborhood V of $g(x_0)$ such that

$$\partial_A d(u, D) \subseteq K^* \quad \forall\, u \in V \cap D,$$

or, equivalently (see [26]), D is compactly epi-Lipschitzian in the sense of Borwein and Strojwas [7]. Then, there exist $\lambda \geq 0$ and $y^* \in \mathbb{R}_+ \partial_A d(g(x_0), D)$, with $(\lambda, y^*) \neq (0, 0)$, such that

$$\lambda \partial_A f(x_0) + \partial_A (y^* \circ g)(x_0) \ni 0.$$

In order to ensure the existence of Lagrange multipliers (i.e., $\lambda \neq 0$ in (1.3)), several qualification conditions have been considered in the literature, including such classical ones as the Slater condition and the Mangasarian–Fromovitz condition. In this paper, we are interested in the existence of Lagrange multipliers for problem (1.1), where the problem is nonconvex, the data is Gâteaux differentiable, and the set D is a closed set. These multipliers are obtained in Theorems 3.1 and 3.3 under the so-called *calmness* condition (introduced in [45] as pseudo–upper-Lipschitz continuity), which is a kind of constraint qualification, and which is implied by the aforementioned notion of metric regularity.

Several sufficient conditions for calmness of the constraint system have been considered in the literature. They are given by using boundary qualification conditions (see [23, 22, 21] and the references therein) or (directional) coderivative conditions (see [16, 18, 17, 19] and the references therein). Concerning the stronger notion of metric regularity of the constraint system, besides the work [31], mentioned above, and which provides some criteria to ensure this property by means of suitable approximations, several sufficient conditions already exist in the literature. They are given either in the dual space via the notion of coderivative (see the monograph by Mordukhovich [38, 39] and the references therein) or in the primal space by using the notion of tangency (see the monograph by Aubin and Frankowska [1] and the references therein).

In this article we propose a new qualification condition, written in terms of the Gâteaux derivative of the function defining the constraint, which ensures that the constraint system is metrically regular around a nominal point. Inspired by the work by Ekeland [14], Theorems 4.2 and 4.6 establish the metric regularity property of the constraint system under Gâteaux differentiability assumptions only. As in [14], the proofs of these results do not rely on any iteration scheme.

Our main motivation to prove the existence of Lagrange multipliers under Gâteaux differentiability assumptions on the data arises from stochastic optimal theory. The first application of our results deals with first order necessary optimality conditions for stochastic optimal control problems in continuous time. Following the functional framework proposed by Backhoff and Silva in [2], our abstract results allow us to provide a direct proof of the existence of Lagrange multipliers at any local solution of the control problem. Having in mind the importance of Lagrange multipliers in sensitivity analysis in optimization theory (see, e.g., [5, Chapter 4]), the study of their existence in stochastic control plays an important role. As pointed out in [2], the main difficulty in deriving the existence result, from standard variational principles, is that the smoothness of the equality constraint that defines the dynamics of the controlled diffusion process is difficult to check. The results in sections 3 and 4, which assume only Gâteaux differentiability of the mapping that defines the constraints and a uniform surjectivity property of the Gâteaux derivative in a neighborhood of the optimal solution, allow us to avoid this issue. Moreover, by means of the identification between Lagrange multipliers and adjoint states, proved in [2], we recover the weak Pontryagin minimum principle proved first in [3] by using different techniques. A detailed discussion and extensions of this result are provided in section 5.1.

In the second application, we consider a discrete time stochastic optimal control problem where the randomness is modeled by a multiplicative independent noise. As in the continuous time case, the main difficulty in applying standard abstract Lagrange multiplier results comes from the functional equation defining the controlled trajectory. By introducing a suitable functional framework for the optimization problem and using our abstract results, we are able to prove in a rather straightforward manner the validity of the optimality system obtained in [35] under more general assumptions than those imposed in that article (see Remark 6.5(i)).

The paper is organized as follows. In the next section we set up the notation and recall some standard results in nonsmooth analysis. In section 3, we establish the existence of Lagrange multipliers for problem (1.1) under the calmness assumption. Next, in section 4, we provide sufficient conditions, in terms of the Gâteaux derivative of g, for the metric regularity of the constraint system. Finally, in sections 5 and 6, we apply these abstract results to the stochastic control problems described in the previous paragraphs.

2. Notation and preliminaries.

Throughout the paper, $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are (real) Banach spaces. The dual spaces of X and Y are denoted by X^* and Y^*, respectively, and for $x^* \in X^*$, $h \in X$ we set $\langle x^*, h\rangle_X := x^*(h)$. Given $r > 0$ and $x \in X$, we denote by $B_X(x, r) := \{x' \in X \,;\, \|x' - x\|_X \leq r\}$ the closed ball of radius r centered at x. For $A \subseteq X$ we denote by $\mathrm{cl}(A)$ and $\mathrm{int}(A)$ its closure and its topological interior, respectively.

Let us recall some basic notions in nonsmooth analysis (see, e.g., [9, 5, 38] for a detailed account of the theory). Given a locally Lipschitz function $\varphi : X \to \mathbb{R}$, the directional derivative $\varphi^\circ(x; h)$ of φ at x in the direction $h \in X$ and the subdifferential

$\partial_C\varphi(x)$ of φ at x are both defined in the sense of *Clarke* as

$$\varphi^\circ(x;h) := \limsup_{y\to x,\tau\downarrow 0} \frac{\varphi(y+\tau h)-\varphi(y)}{\tau},$$
$$\partial_C\varphi(x) := \{x^*\in X^*\ ;\ \langle x^*,h\rangle_X \le \varphi^\circ(x;h)\ \ \forall\, h\in X\}\,.$$

Note that for all $x\in X$, $\varphi^\circ(x;\cdot): X\to\mathbb{R}$ is well defined, positively homogeneous, subadditive, and Lipschitz continuous, and it satisfies that $\varphi^\circ(x;0)=0$. This implies that $\varphi^\circ(x;\cdot)$ is the support function of $\partial_C\varphi(x)$, which is a nonempty, weak*-compact, and convex set (see [9, Proposition 2.1.2]). Given a nonempty set $A\subseteq X$, we denote by $d_A(\cdot) := \inf_{x\in A}\|\cdot - x\|_X$ the distance to the A function. Given $x\in \mathrm{cl}(A)$, the *Clarke's* tangent cone is defined as

$$T_A(x) := \left\{h\in X:\quad \lim_{y\to x,\ y\in A,\ \tau\to 0^+} \frac{d_A(y+\tau h)}{\tau} = 0\right\}.$$

If $x\notin \mathrm{cl}(A)$, we set $T_A(x) := \emptyset$. If $x\in\mathrm{cl}(A)$, we have that $h\in T_A(x)$ iff for every sequence (x_n) such that $x_n\in A$, $x_n\to x$, and $\tau_n\to 0^+$ there exists a sequence $h_n\to h$ such that $x_n+\tau_n h_n\in A$ for all n large enough. The *Clarke's* normal cone to A at x is defined as $N_A(x)=T_A(x)^0$, where for a given cone $\mathcal{K}$ we denote by $\mathcal{K}^0$ its negative polar cone, defined as

$$\mathcal{K}^0 := \{x^*\in X^*:\quad \langle x^*,h\rangle_X\le 0\ \ \forall\, h\in\mathcal{K}\}.$$

We have (see, e.g., [9, Proposition 2.4.2])

$$N_A(x) = w^*\text{-cl}\left(\textstyle\bigcup_{\lambda\ge 0}\lambda\partial_C d_A(x)\right), \tag{2.1}$$

where w^*-cl denotes the weak-star closure in X^*. The *adjacent* (or *Ursescu*) tangent cone to A at $x\in\mathrm{cl}(A)$ is defined by

$$\mathcal{T}(A,x) = \left\{h\in X:\quad \lim_{\tau\to 0^+}\frac{d_A(x+\tau h)}{\tau} = 0\right\}.$$

We set $\mathcal{T}(A,x) := \emptyset$ if $x\notin\mathrm{cl}(A)$. By definition, if $x\in\mathrm{cl}(A)$, then $h\in\mathcal{T}(A,x)$ iff for any sequence $\tau_n\to 0^+$ there exists a sequence $h_n\to h$ such that $x+\tau_n h_n\in A$ for all n sufficiently large.

Finally, the *contingent* (or *Bouligand*) tangent cone to A at $x\in\mathrm{cl}(A)$ is defined as

$$K(A,x) := \left\{h\in X:\quad d_A^-(x;h)=0\right\},$$

where $d_A^-(x;h)$ is the lower Dini directional derivative of d_A at x in the direction h, that is,

$$d_A^-(x;h) := \liminf_{\tau\to 0^+}\frac{d_A(x+\tau h)}{\tau}.$$

We set $K(A,x):=\emptyset$ if $x\notin\mathrm{cl}(A)$. By definition, if $x\in\mathrm{cl}(A)$, then $h\in K(A,x)$ iff there exist sequences $\tau_n\to 0^+$ and $h_n\to h$ such that $x+\tau_n h_n\in A$ for n sufficiently large. Note that

$$T_A(x)\subseteq\mathcal{T}(A,x)\subseteq K(A,x).$$

If A is convex, then the previous tangent cones coincide. In the general case these cones are closed, they differ, and only $T_A(x)$ is guaranteed to be convex.

We say that A is *tangentially regular* at x if

$$K(A,x) = \mathcal{T}(A,x). \tag{2.2}$$

For later use, we state the following result, whose proof can be easily deduced from the previous definitions.

LEMMA 2.1. *Let $A \subseteq X$ and $B \subseteq Y$ be closed sets, and let $x_0 \in A$ and $y_0 \in B$. The space $X \times Y$ is endowed with the product norm, that is, $\|(x,y)\|_{X\times Y} = \|x\|_X + \|y\|_Y$. Then*

(i) *$K(A\times B,(x_0,y_0)) \subseteq K(A,x_0) \times K(B,y_0)$. The equality holds whenever A is tangentially regular at x_0 or B is tangentially regular at y_0.*

(ii) *For all $h \in X$ and $k \in Y$, $d^-_{A\times B}((x_0,y_0),(h,k)) \le d^-_A(x_0,h) + d^0_B(y_0,k)$.*

(iii) *If A is tangentially regular at x_0 or B is tangentially regular at y_0, then for all $h \in X$ and $k \in Y$,*

$$d^-_{A\times B}((x_0,y_0),(h,k)) \le d_{K(A,x_0)}(h) + d_{K(B,y_0)}(k).$$

Recall that $g : X \to Y$ is said to be *Gâteaux differentiable* at $x_0 \in X$ (see, e.g., [9, section 2.2]) if there exists a bounded linear operator[1] $Dg(x_0) : X \to Y$ such that

$$\lim_{\tau\downarrow 0} \frac{g(x_0+\tau h) - g(x_0)}{\tau} = Dg(x_0)h \quad \forall\, h \in X.$$

To finish this section, we state the following lemma, first proved in [43, Lemma 1]. For the sake of completeness, we provide a short proof based on the separation theorem.

LEMMA 2.2. *Let $A \subseteq X$ be a closed convex set, and let $s > 0$ and $r \in]0,s[$. Then the following implication holds:*

$$B_X(0,s) \subseteq A + B_X(0,r) \Longrightarrow B_X(0,s-r) \subseteq A.$$

Proof. Let $x \in B_X(0,s-r)$. Suppose that $x \notin A$. Then, by a separation theorem, there exist $x^* \in X^*$, with $\|x^*\|_{X^*} = 1$, and $\alpha \in \mathbb{R}$ such that

$$\langle x^*, x\rangle_X > \alpha \ge \langle x^*, u\rangle_X \quad \forall\, u \in A.$$

By assumption, for all $z \in B_X(0,s)$ there exists $b \in B_X(0,r)$ such that $z + b \in A$. Thus,

$$\langle x^*, x\rangle_X > \alpha \ge \langle x^*, z+b\rangle_X \ge \langle x^*, z\rangle_X - r,$$

and, hence,

$$s - r \ge \langle x^*, x\rangle_X > \alpha \ge s - r.$$

This contradiction completes the proof of the lemma. ∎

3. Lagrange multipliers for optimization problems under Gâteaux differentiability assumptions on the data. This section is concerned with the existence of Lagrange multipliers associated to local solutions of optimization problems of the form

$$\begin{cases} \min & f(x) \\ \text{s.t.} & g(x) = 0, \quad x \in C, \end{cases} \tag{3.1}$$

[1] Some authors drop the linearity requirement in this definition.

where $f : X \to \mathbb{R} \cup \{+\infty\}$ is a function, $g : X \to Y$ is a mapping from a (real) Banach space $(X, \|\cdot\|_X)$ to a (real) Banach space $(Y, \|\cdot\|_Y)$, and C is a nonempty closed subset of X.

Suppose that x_0 is a local solution to problem (3.1). Let us state now our basic assumptions that will allow us to establish first order optimality conditions at x_0.

($\mathbf{H_f}$) f is Gâteaux differentiable at x_0 and locally Lipschitz around x_0 with constant $K_f > 0$; that is, there exists $r > 0$ such that

$$|f(x) - f(x')| \le K_f \|x - x'\|_X \quad \forall\, x, x' \in B_X(x_0, r).$$

($\mathbf{H_g}$) g is Gâteaux differentiable at x_0.

If ($\mathbf{H_g}$) holds true, we will denote by $D^*g(x_0) : Y^* \to X^*$ the adjoint operator of $Dg(x_0)$.

We recall that the system

$$x \in C \quad \text{and} \quad g(x) = 0 \tag{3.2}$$

is said to be *calm* at $x_0 \in g^{-1}(0) \cap C$ if there exist $a > 0$ and $s > 0$ such that

$$d_{g^{-1}(0)\cap C}(x) \le a\|g(x)\|_Y \qquad \forall\, x \in B_X(x_0, s) \cap C. \tag{3.3}$$

The following result gives existence of Lagrange multipliers for problem (3.1) under the calmness condition (3.3) and the weak differentiability assumptions ($\mathbf{H_f}$)–($\mathbf{H_g}$).

THEOREM 3.1. *Suppose that* ($\mathbf{H_f}$)–($\mathbf{H_g}$) *hold and that system* (3.2) *is calm at* x_0. *Let* K_f *and* a *be as in* ($\mathbf{H_f}$) *and* (3.3), *respectively. Then*

(i) *if* $x_0 \in \operatorname{int}(C)$, *then there exists* $y^* \in Y^*$, *with* $\|y^*\|_{Y^*} \le K_f a$, *such that*

$$Df(x_0) + D^*g(x_0)y^* = 0;$$

(ii) *if* g *is locally Lipschitz around* x_0 *with constant* $K_g > 0$, *then*

$$Df(x_0)h + K_f a\|Dg(x_0)h\|_Y + K_f(1 + K_g a)d_C^-(x_0; h) \ge 0 \quad \forall\, h \in X. \tag{3.4}$$

In particular, there exists $y^* \in Y^*$, *with* $\|y^*\|_{Y^*} \le K_f a$, *such that*

$$0 \in Df(x_0) + D^*g(x_0)y^* + N_C(x_0).$$

If, in addition, $K(C, x_0)$ *is convex, then there exists* $y^* \in Y^*$, *with* $\|y^*\|_{Y^*} \le K_f a$, *such that*

$$0 \in Df(x_0) + D^*g(x_0)y^* + (K(C, x_0))^0.$$

Proof. Since x_0 is a local solution of problem (3.1) and f satisfies ($\mathbf{H_f}$), by [9, Proposition 2.4.3] we have that x_0 is a local minimum of

$$X \ni x \mapsto f(x) + K_f d_{g^{-1}(0)\cap C}(x) \in \mathbb{R} \cup \{+\infty\}.$$

Using the calmness assumption of system (3.2), we get that x_0 is a local solution to

$$\min \;\; f(x) + K_f a\|g(x)\|_Y \quad \text{s.t.} \quad x \in C. \tag{3.5}$$

Now, let us prove assertion (i). Since $x_0 \in \operatorname{int}(C)$, there exists $s > 0$ such that

$$f(x) + K_f a\|g(x)\|_Y \ge f(x_0) \quad \forall\, x \in B_X(x_0, s).$$

Let $h \in X$ be arbitrary, and choose $\tau > 0$ small enough such that $x_0 + \tau h \in B_X(x_0, s)$. Then

$$\frac{f(x_0 + \tau h) - f(x_0)}{\tau} + K_f a \left\| \frac{g(x_0 + \tau h) - g(x_0)}{\tau} \right\|_Y \geq 0.$$

Using that f and g are Gâteaux differentiable at x_0, we get

$$Df(x_0)h + K_f a \|Dg(x_0)h\|_Y \geq 0.$$

This means that the convex function $h \mapsto Df(x_0)h + K_f a \|Dg(x_0)h\|_Y$ attains its minimum at $h = 0$. Thus, the (convex) subdifferential calculus produces a $y^* \in Y^*$, with $\|y^*\|_{Y^*} \leq K_f a$, such that

$$Df(x_0) + D^* g(x_0) y^* = 0.$$

In order to prove assertion (ii), note that since x_0 solves locally (3.5) and f and g are locally Lipschitz at x_0, by using [9, Proposition 2.4.3] again, we obtain the existence of $s > 0$ such that

$$f(x) + K_f a \|g(x)\|_Y + K_f(1 + K_g a) d_C(x) \geq f(x_0) \quad \forall\, x \in B_X(x_0, s).$$

Let $h \in X$ be arbitrary, and choose a sequence $\tau_n \to 0^+$ such that

$$d_C^-(x_0; h) = \lim_{n \to +\infty} \frac{d_C(x_0 + \tau_n h)}{\tau_n}.$$

Then, using the Gâteaux differentiability of f and g, we get

$$Df(x_0)h + K_f a \|Dg(x_0)h\|_Y + K_f(1 + K_g a) d_C^-(x_0; h) \geq 0. \tag{3.6}$$

Noting that $d_C^-(x_0; h) \leq d_C^\circ(x; h)$, we obtain

$$Df(x_0)h + K_f a \|Dg(x_0)h\|_Y + K_f(1 + K_g a) d_C^\circ(x_0; h) \geq 0 \quad \forall\, h \in X,$$

or, equivalently, the convex function

$$X \ni h \mapsto Df(x_0)h + K_f a \|Dg(x_0)h\|_Y + K_f(1 + K_g a) d_C^\circ(x_0, h) \in \mathbb{R}$$

attains its minimum at $h = 0$. Using that $\partial_C d_C^\circ(x_0, \cdot)(0) = \partial_C d_C(x_0)$ and (2.1), the (convex) subdifferential calculus produces a $y^* \in Y^*$, with $\|y^*\|_{Y^*} \leq K_f a$, such that

$$-Df(x_0) - D^* g(x_0) y^* \in K_f(1 + K_g a) \partial d_C(x_0) \subseteq N_C(x_0),$$

so that assertion (ii) follows.

Finally, inequality (3.6) yields

$$Df(x_0)h + K_f a \|Dg(x_0)h\|_Y \geq 0 \quad \forall\, h \in K(C, x_0).$$

Thus, if $K(C, x_0)$ is convex, the last assertion in (ii) follows from the convex subdifferential calculus. ∎

Now consider the following optimization problem:

$$\begin{cases} \min & f(x) \\ \text{s.t.} & g(x) \in D, \quad x \in C, \end{cases} \tag{3.7}$$

and the system

$$\text{Find} \quad x \in C, \quad g(x) \in D. \tag{3.8}$$

System (3.8) is said to be calm at $x_0 \in g^{-1}(D) \cap C$ if there exist $a > 0$ and $s > 0$ such that

$$d_{g^{-1}(D)\cap C}(x) \leq a d_D(g(x)) \quad \forall\, x \in B_X(x_0, s) \cap C. \tag{3.9}$$

Problem (3.7) can be rephrased as follows:

$$\begin{cases} \min & \tilde{f}(x,y) \\ \text{s.t.} & \tilde{g}(x,y) = 0, \quad (x,y) \in C \times D, \end{cases} \tag{3.10}$$

where $\tilde{f}(x,y) = f(x)$ and $\tilde{g}(x,y) = g(x) - y$. Therefore, (3.7) can be written in the form (3.1). In the following result, we transfer the calmness property of system (3.8) to that of system

$$\text{Find} \quad (x,y) \in C \times D, \quad \tilde{g}(x,y) = 0, \tag{3.11}$$

where the product space $X \times Y$ is endowed with the norm given by the sum of the norms in X and Y.

LEMMA 3.2. *Suppose that g is locally Lipschitz around x_0, and set $y_0 := g(x_0)$. Then the following assertions are equivalent:*
(i) *The system* (3.8) *is calm at $x_0 \in g^{-1}(D) \cap C$.*
(ii) *The system* (3.11) *is calm at $(x_0, y_0) \in C \times D$.*

Proof. For notational convenience, we omit the subscripts for the norms $\|\cdot\|_X$ and $\|\cdot\|_Y$. (i) $\Rightarrow$ (ii): Since the system (3.8) is calm at $x_0 \in g^{-1}(D) \cap C$ and g is locally Lipschitz around x_0, there exist $a > 0$, $s > 0$, and $K_g > 0$ such that

$$d_{g^{-1}(D)\cap C}(x) \leq a d_D(g(x)) \quad \forall\, x \in B_X(x_0, 3s) \cap C,$$

and

$$\|g(x) - g(x')\| \leq K_g \|x - x'\| \quad \forall\, x,\ x' \in B_X(x_0, 3s).$$

Let $(x,y) \in B((x_0,y_0), s) \cap (C \times D)$. For all $t \in]0, s[$ there exists $u \in g^{-1}(D) \cap C$ such that

$$\|x - u\| \leq d_{g^{-1}(D)\cap C}(x) + t \leq \|x - x_0\| + t \leq 2s,$$

and this asserts that $u \in B(x_0, 3s) \cap \left(g^{-1}(D) \cap C\right)$. Thus,

$$\|g(x) - g(u)\| \leq K_g \|x - u\|.$$

We have

$$d_{\tilde{g}^{-1}(0)\cap(C\times D)}(x,y) = \inf_{v \in C \cap g^{-1}(D)} [\|x - v\| + \|y - g(v)\|] \leq \|x - u\| + \|y - g(u)\|, \tag{3.12}$$

and, using the triangle inequality, we get

$$\begin{aligned}\|y-g(u)\|+\|x-u\| &\le \|y-g(x)\|+\|g(x)-g(u)\|+\|x-u\|\\ &\le \|y-g(x)\|+(1+K_g)\|x-u\|\\ &\le \|y-g(x)\|+(1+K_g)d_{g^{-1}(D)\cap C}(x)+t(1+K_g)\\ &\le \|y-g(x)\|+(1+K_g)a\|y-g(x)\|+t(1+K_g)\\ &\le (1+a(1+K_g))\|y-g(x)\|+t(1+K_g)\\ &= (1+a(1+K_g))\|\tilde g(x,y)\|+t(1+K_g).\end{aligned}$$

As t is arbitrary, relation (3.12) yields

$$\forall\ (x,y)\in B((x_0,y_0),s)\cap(C\times D),\quad d_{\tilde g^{-1}(0)\cap(C\times D)}(x,y)\le (1+a(1+K_g))\|\tilde g(x,y)\|,$$

which implies that (ii) holds. The implication (ii) $\Rightarrow$ (i) is obvious since the following inequality holds true for all $x\in X$ and $y\in Y$:

$$d_{\tilde g^{-1}(0)\cap(C\times D)}(x,y)\ge d_{g^{-1}(D)\cap C}(x). \qquad \square$$

The following theorem, which is a consequence of Theorem 3.1, Lemma 2.1, and Lemma 3.2, gives the existence of Lagrange multipliers for problem (3.7) under the calmness condition and the weak differentiability assumptions $(\mathbf{H_f})$–$(\mathbf{H_g})$.

THEOREM 3.3. *Let x_0 be a local solution to problem* (3.7), *and suppose that system* (3.8) *is calm at x_0. Suppose that $(\mathbf{H_f})$ and $(\mathbf{H_g})$ hold and that g is locally Lipschitz around x_0. Then the following hold:*

(i) *There exists $y^*\in N_D(g(x_0))$, with $\|y^*\|_{Y^*}\le K_f(1+a(1+K_g))$ (where K_f, K_g, and a are as in $(\mathbf{H_f})$, $(\mathbf{H_g})$, and* (3.9), *respectively), such that*

$$-Df(x_0)-D^*g(x_0)y^*\in N_C(x_0).$$

(ii) *Moreover, if $K(C,x_0)$ and $K(D,g(x_0))$ are convex and C is tangentially regular at x_0 or D is tangentially regular at $g(x_0)$, then there exists $y^*\in (K(D,g(x_0)))^0$ such that $\|y^*\|_{Y^*}\le K_f(1+a(1+K_g))$ and*

$$0\in Df(x_0)+D^*g(x_0)y^*+(K(C,x_0))^0.$$

Proof. Since x_0 solves (3.7) locally, $(x_0,g(x_0))$ is a local solution to problem (3.10). Using that the constant a satisfies (3.9), the proof of Lemma 3.2 shows that the calmness constant associated to system (3.11) is given by $(1+a(1+K_g))$. Applying the second assertion in Theorem 3.1(ii) to problem (3.10) yields the first assertion (i). In order to prove assertion (ii), note that (3.4) implies that

$$Df(x_0)h+K_f(1+a(1+K_g))\|Dg(x_0)h-k\|_Y\ge 0\quad \forall\ (h,k)\in K(C\times D,(x_0,g(x_0))).$$

By Lemma 2.1 we have that $K(C\times D,(x_0,g(x_0)))=K(C,x_0)\times K(D,g(x_0))$, which is a convex set. The result then follows from standard convex analysis calculus. $\square$

4. Sufficient conditions for calmness under Gâteaux differentiability. In this section, we first provide sufficient conditions for a stronger property than the calmness of the system (3.2), namely its metric regularity (see [12, 27] and the references therein). Then, and as in the previous section, we deduce the corresponding sufficient condition for system (3.8) by reducing it to an instance of system (3.2) (see (3.11)). Let us stress the fact that the aforementioned conditions only involve Gâteaux differentiability assumptions on the function g.

In the remainder of this article, given a subset A of a real Banach space $(Z, \|\cdot\|_Z)$, $y \in A$, and $r > 0$, we set $B_A(y, r) := B_Z(y, r) \cap A$. Let us recall that the system (3.2) is *metrically regular* at $x_0 \in g^{-1}(0) \cap C$ if there exist $\alpha > 0$ and $r > 0$ such that

$$d_{g^{-1}(y)\cap C}(x) \leq \alpha \|g(x) - y\|_Y \quad \forall\, x \in B_C(x_0, r), \forall\, y \in B_Y(0, r).$$

From the very definition, it follows that the previous notion is stronger than the calmness property of system (3.2) (see, e.g., [21] for a more detailed discussion of this subject).

Let us fix a point $x_0 \in g^{-1}(0) \cap C$. We consider the following *constraint qualification condition* on a neighborhood of x_0.

$(\mathbf{H_{cq}})$ there exist $\alpha > 0$ and $r > 0$ such that g is continuous and Gâteaux differentiable on $B_C(x_0, r)$ and

$$B_Y(0, 1) \subseteq Dg(x)\big(B_{K(C,x)}(0, \alpha)\big) \quad \forall\, x \in B_C(x_0, r). \tag{4.1}$$

Remark 4.1. For each $x \in B(x_0, r)$ consider a right-inverse $G(x) : Y \rightrightarrows X$ of $Dg(x)$, i.e., $Dg(x)G(x)y = \{y\}$ for all $y \in Y$ (we know that such a right-inverse exists because (4.1) implies that $Dg(x)$ is surjective). Then, assumption (4.1) can be rephrased in terms of G as follows:

$$\sup_{x \in B_X(x_0, r),\, y \in B_Y(0,1)} \; \inf_{v \in G(x)y \cap K(C,x)} \|v\|_X \leq \alpha.$$

In the following result, we provide a sufficient condition for the metric regularity of system (3.2) at $x_0 \in g^{-1}(0) \cap C$ under $(\mathbf{H_{cq}})$.

THEOREM 4.2. *Suppose that* $(\mathbf{H_{cq}})$ *holds true, and let* $\alpha > 0$ *and* $r > 0$ *be such that* (4.1) *is satisfied. Then, for all* $r_1 > 0$ *and* $r_2 > 0$*, with* $r_1 + r_2 = r$*, and all*

$$(x, y) \in D_{r_1, r_2} := \left\{ (u, v) \in B_C(x_0, r_1) \times Y : \|g(u) - v\|_Y < \frac{r_2}{\alpha} \right\},$$

we have

$$d_{g^{-1}(y)\cap C}(x) \leq \alpha \|g(x) - y\|_Y. \tag{4.2}$$

Proof. The proof is inspired from [14]. Fix $(x, y) \in D_{r_1, r_2}$. If $y = g(x)$, then (4.2) is trivial, so let us assume that $y \neq g(x)$. Consider the function $h : X \to \mathbb{R}$ defined as

$$h(u) := \|g(u) - y\|_Y. \tag{4.3}$$

Let $\beta > \alpha$ be such that $0 < h(x) = \|g(x) - y\|_Y < \frac{r_2}{\beta}$. As h is continuous and bounded from below on the closed set $B_C(x_0, r)$ and, evidently,

$$h(x) \leq \inf_{x' \in B_C(x_0, r)} h(x') + h(x),$$

Ekeland's variational principle (see [13, Theorem 1.1]) gives the existence of $\bar{u} \in B_C(x_0, r)$ such that

$$h(\bar{u}) \le h(x), \tag{4.4}$$

$$\|\bar{u} - x\|_X \le \beta h(x), \tag{4.5}$$

$$h(\bar{u}) \le h(u) + \frac{1}{\beta}\|\bar{u} - u\|_X \quad \forall\, u \in B_C(x_0, r). \tag{4.6}$$

Inequality (4.5) and the choice of x and β imply that

$$\|\bar{u} - x\|_X < r_2 \quad \text{and so} \quad \|\bar{u} - x_0\|_X \le \|\bar{u} - x\|_X + \|x - x_0\|_X < r_2 + r_1 = r. \tag{4.7}$$

Claim: We have that $y = g(\bar{u})$. Let us assume for a moment that the claim is true. By (4.5), we obtain

$$d_{g^{-1}(y)\cap C}(x) \le \beta \|g(x) - y\|_Y,$$

and, as $\beta > \alpha$ is arbitrary, we get that (4.2) holds true.

It remains to prove the claim. Suppose the contrary and define

$$w = \frac{y - g(\bar{u})}{\|y - g(\bar{u})\|_Y}. \tag{4.8}$$

Since $\bar{u} \in B_C(x_0, r)$, assumption $(\mathbf{H_{cq}})$ implies the existence of $v \in B_{K(C,\bar{u})}(0, \alpha)$ such that

$$w = Dg(\bar{u})v.$$

Since $v \in B_{K(C,\bar{u})}(0, \alpha)$, there exist sequences $\tau_n \to 0^+$ and $v_n \to v$ such that

$$u_n := \bar{u} + \tau_n v_n \in C \text{ for } n \text{ sufficiently large.}$$

We may write $u_n = \bar{u} + \tau_n v + o(\tau_n) \in C$, where $\lim_{n\to+\infty} o(\tau_n)/\tau_n = 0$. Note that the second inequality in (4.7) implies that $u_n \in B_C(x_0, r)$ for n sufficiently large. Now, using inequality (4.6), we get

$$h(\bar{u}) \le h(u_n) + \frac{1}{\beta}\|\tau_n v + o(\tau_n)\|_X. \tag{4.9}$$

On the other hand, since g is Gâteaux differentiable at $\bar{u}$, we have

$$g(u_n) = g(\bar{u}) + \tau_n Dg(\bar{u})v + \tau_n \varepsilon(\tau_n), \text{ where } \lim_{n\to+\infty} \varepsilon(\tau_n) = 0, \tag{4.10}$$

which, combined with (4.9), ensures that

$$\frac{\|g(\bar{u}) - y + \tau_n Dg(\bar{u})v + \tau_n\varepsilon(\tau_n)\|_Y - \|g(\bar{u}) - y\|_Y}{\tau_n} \ge -\frac{1}{\beta}\left\| v + \frac{o(\tau_n)}{\tau_n}\right\|_X. \tag{4.11}$$

Since

$$\lim_{n\to+\infty} \frac{\|g(\bar{u}) - y + \tau_n Dg(\bar{u})v\|_Y - \|g(\bar{u}) - y\|_Y}{\tau_n} = \max_{y^* \in \partial\|\cdot\|_Y(g(\bar{u})-y)} \langle y^*, Dg(\bar{u})v\rangle_Y,$$

we get the existence of $y_v^* \in \partial\|\cdot\|_Y(g(\bar{u}) - y)$, such that

$$-1 = \langle y_v^*, w\rangle_Y = \langle y_v^*, Dg(\bar{u})v\rangle_Y \ge -\frac{1}{\beta}\|v\|_X \ge -\frac{\alpha}{\beta}, \tag{4.12}$$

where the first equality follows from the fact that we are assuming that $g(\bar{u}) \neq y$ and the standard relation

$$y_v^* \in \partial\|\cdot\|_Y(g(\bar{u})-y) \Leftrightarrow \|y_v^*\|_{Y^*}=1 \text{ and } \langle y_v^*, g(\bar{u})-y\rangle_Y = \|g(\bar{u})-y\|_Y.$$

Since (4.12) contradicts $\alpha < \beta$, the claim follows. □

The previous result extends the following inverse function theorem result, proved first in [14, Theorem 2] in the case $C = X$.

Corollary 4.3. *Suppose that the assumptions of Theorem* 4.2 *are satisfied. Then*

$$d_{g^{-1}(y)\cap C}(x_0) \leq \alpha\|y\|_Y \quad \forall\, y \in Y, \text{ with } \|y\|_Y < \frac{r}{\alpha}. \tag{4.13}$$

Consequently, for all $y \in Y$, with $\|y\|_Y < \frac{r}{\alpha}$, and for all $\beta > \alpha$ there exists $x \in g^{-1}(y) \cap C$ such that

$$\|x - x_0\|_X < r, \quad \|x-x_0\|_X \leq \beta\|y\|_Y. \tag{4.14}$$

Proof. By Theorem 4.2, in order to prove (4.13) it suffices to choose $\varepsilon > 0$ such that $(x_0, y) \in D_{\varepsilon, r-\varepsilon}$, which is possible because of the strict inequality in (4.13). It remains to prove that (4.14) holds for $\beta > \alpha$ and $\|y\|_Y < r/\alpha$. In this case, the first inequality in (4.13) becomes strict, and we get the existence of $x_\beta \in g^{-1}(y) \cap C$ such that the second inequality in (4.14) holds true.

Since there exists $\varepsilon > 0$ such that $\|y\|_Y \leq (r-\varepsilon)/\alpha$, the first inequality in (4.14) holds for x_β provided that $\alpha < \beta < \alpha r/(r-\varepsilon)$. If $\beta \geq \alpha r/(r-\varepsilon)$, then (4.14) holds for $x_{\beta'}$ with $\beta' \in]\alpha, \alpha r/(r-\varepsilon)[$ and so $\|x_{\beta'} - x_0\|_X \leq \beta'\|y\|_Y \leq \beta\|y\|_Y$. The result follows. □

By taking a closer look at the proof of Theorem 4.2, we see that (4.2) holds under alternative assumptions involving the notion of *strict differentiability* of g, which is much stronger than its Gâteaux differentiability. Let us recall that g is strictly differentiable at $x_0 \in C$ with respect to (w.r.t.) C if

$$\lim_{\substack{x\to x_0,\ x'\to x_0 \\ x\neq x',\ x,\, x'\in C}} \frac{g(x)-g(x')-Dg(x_0)(x-x')}{\|x-x'\|_X} = 0$$

(see, e.g., [38, Definition 1.13]). In this framework, we can replace condition $(\mathbf{H_{cq}})$ by the following:

$(\mathbf{H^1_{cq}})$ There exist $\alpha > 0$ and $r > 0$ such that

$$B_Y(0,1) \subseteq Dg(x_0)\big(B_{K(C,x)}(0,\alpha)\big) \quad \forall\, x \in B_C(x_0, r). \tag{4.15}$$

We obtain the following theorem, whose proof is similar to that of Theorem 4.2.

Theorem 4.4. *Suppose that g is strictly differentiable at x_0 and condition* $(\mathbf{H^1_{cq}})$ *holds. Then the system* (3.2) *is metrically regular at x_0.*

Proof. Since g is strictly differentiable at x_0 w.r.t. C, for all $\varepsilon \in (0, 1/\alpha)$ there exists $\delta > 0$ such that

$$x, x' \in B_C(x_0,\delta) \Rightarrow \|g(x)-g(x')-Dg(x_0)(x-x')\|_Y \leq \varepsilon\|x-x'\|_X. \tag{4.16}$$

Let us fix $\varepsilon \in (0, 1/\alpha)$ and $\delta > 0$ such that (4.16) holds. We may assume that $\delta \leq r$, where r is as in $(\mathbf{H^1_{cq}})$. Let $\delta_1 > 0$ and $\delta_2 > 0$ be such that $\delta_1 + \delta_2 = \delta$, and let

$$(x,y) \in \left\{(u,v) \in B_C(x_0,\delta_1) \times Y : \|g(u)-v\|_Y < \left(\frac{1-\varepsilon\alpha}{\alpha}\right)\delta_2\right\}.$$

As in the proof of Theorem 4.2, define $h : X \to \mathbb{R}$ by (4.3). If $h(x) = 0$, then there is nothing to prove. Therefore, let us assume that $h(x) \neq 0$ and pick $\beta \in (\frac{\alpha}{1-\varepsilon\alpha}, \frac{\delta_2}{h(x)})$. Then, arguing exactly as in the proof of Theorem 4.2, we get the existence of $\bar{u} \in B_C(x_0, \delta)$ such that (4.4), (4.5), and (4.6) hold with r replaced by δ. As a consequence of (4.5) and the choice of β, we get that $\|\bar{u} - x_0\|_X < \delta$. Let us prove that $g(\bar{u}) = y$. If this is not the case, then defining w by (4.8), assumption $(\mathbf{H^1_{cq}})$ yields the existence of $v \in K(C, \bar{u})$, with $\|v\|_Y \le \alpha$, such that $Dg(x_0)v = w$. Since $v \in B_{K(C,\bar{u})}(0, \alpha)$, we can find sequences $\tau_n \to 0^+$ and $v_n \to v$ such that $u_n := \bar{u} + \tau_n v_n = \bar{u} + \tau_n v + o(\tau_n) \in C$ for n sufficiently large. Therefore, by (4.16), we get

$$\|g(u_n) - g(\bar{u}) - Dg(x_0)(u_n - \bar{u})\|_Y \le \varepsilon\tau_n \|v + o(\tau_n)/\tau_n\|_X \text{ for } n \text{ large enough.}$$

Using the previous inequality, (4.6) yields

$$\begin{aligned}
\frac{-1}{\beta}\big\|v + o(\tau_n)/\tau_n\big\|_X &\le \frac{\|g(u_n) - y\|_Y - \|g(\bar{u}) - y\|_Y}{\tau_n} \\
&\le \frac{\|g(u_n) - g(\bar{u}) - Dg(x_0)(u_n - \bar{u})\|_Y + \|g(\bar{u}) + Dg(x_0)(u_n - \bar{u}) - y\|_Y}{\tau_n} \\
&\quad - \frac{\|g(\bar{u}) - y\|_Y}{\tau_n} \\
&\le \varepsilon\|v + o(\tau_n)/\tau_n\|_X + \frac{\|g(\bar{u}) - y + \tau_n Dg(x_0)v\|_Y - \|g(\bar{u}) - y\|_Y}{\tau_n} + o(\tau_n)/\tau_n.
\end{aligned}$$

Thus, letting $n \to \infty$, we get the existence of $y_v^* \in \partial\|\cdot\|_Y(g(\bar{u}) - y)$ such that

$$-\frac{\|v\|_X}{\beta} \le \varepsilon\|v\|_X + \langle y_v^*, w\rangle_Y = \varepsilon\|v\|_X - 1.$$

Using that $\|v\|_X \le \alpha$, the previous inequality yields

$$-\frac{\alpha}{\beta} \le \varepsilon\alpha - 1,$$

contradicting $\beta > \alpha/(1 - \varepsilon\alpha)$. Thus, $g(\bar{u}) = y$, and, hence, by (4.5) and the fact that $\beta \in (\alpha/(1 - \varepsilon\alpha), \delta_2/h(x))$ is arbitrary, we get

$$d_{g^{-1}(y)\cap C}(x) \le \frac{\alpha}{1 - \varepsilon\alpha}\|g(x) - y\|_Y.$$

The proof is then completed. ∎

Remark 4.5. In the convex case, that is, C is convex, condition $(\mathbf{H^1_{cq}})$ is equivalent to the following Robinson constraint qualification [11]:

$(\mathbf{H^R_{cq}})$ g is strictly differentiable at x_0 and there exists $\alpha > 0$ such that

$$B_Y(0,1) \subseteq Dg(x_0)\big(B_{(C-x_0)}(0,\alpha)\big). \tag{4.17}$$

Indeed, from Theorem 4.4 and the equivalence between metric regularity and the Robinson constraint qualification (4.17) (see [11]), it is enough to show that (4.17) implies $(\mathbf{H^1_{cq}})$. Let $1 > \varepsilon > 0$ be such that $\varepsilon\|Dg(x_0)\| < 1$ and $x \in B_C(x_0, \varepsilon)$. By (4.17), for all $y \in B_Y(0, 1)$ there exists $c \in B_C(x_0, \alpha)$ such that

$$y = Dg(x_0)(c - x_0) = Dg(x_0)(c - x) + Dg(x_0)(x - x_0),$$

and, hence, $y \in Dg(x_0)\big((C-x)\cap B_X(0,\alpha+\varepsilon)\big)+\varepsilon\|Dg(x_0)\|B_Y(0,1)$. Thus,

$$B_Y(0,1)\subseteq Dg(x_0)\big((C-x)\cap B_X(0,\alpha+\varepsilon)\big)+\varepsilon\|Dg(x_0)\|B_Y(0,1),$$

and, hence, by Lemma 2.2,

$$B_Y(0,1-\varepsilon\|Dg(x_0)\|)\subseteq \mathrm{cl}\Big(Dg(x_0)\big((C-x)\cap B_X(0,\alpha+\varepsilon)\big)\Big).$$

This implies (see [44, Lemma 1] or [11, Lemma 2])

$$B_Y\left(0,\frac{1-\varepsilon\|Dg(x_0)\|}{2}\right)\subseteq Dg(x_0)\big((C-x)\cap B_X(0,\alpha+\varepsilon)\big).$$

Now, it suffices to conclude by remarking that, since C is convex, $C-x\subseteq K(C,x)$.

We study now the metric regularity property for system (3.8) by assuming that g is Gâteaux differentiable. We consider the following qualification condition:

($\mathbf{H'_{cq}}$) There exist $\alpha_1,\alpha_2>0$ and $r>0$ such that g is continuous and Gâteaux differentiable on $B_C(x_0,r)$ and

$$B_Y(0,1)\subseteq Dg(x)\big(B_{K(C,x)}(0,\alpha_1)\big)-B_{K(D,y)}(0,\alpha_2) \quad \forall\,(x,y)\in B_{C\times D}((x_0,g(x_0)),r). \tag{4.18}$$

THEOREM 4.6. *Suppose that* ($\mathbf{H_g}$) *and* ($\mathbf{H'_{cq}}$) *hold true and that at least one of the sets C and D is convex. Denote $\alpha=\max\{\alpha_1,\alpha_2\}$. Then, for all $r_1>0$ and $r_2>0$, with $r_1+r_2=r$, and all*

$$(x,y)\in D_{r_1,r_2}:=\left\{(u,v)\in B_C(x_0,r_1)\times Y:\, d_{B_D(g(x_0),r_1)}(g(u)-v)<\frac{r_2}{\alpha}\right\},$$

we have

$$d_{g^{-1}(D+y)\cap C}(x)\le \alpha d_{B_D(g(x_0),r_1)}(g(x)-y).$$

Proof. Using that at least one of the sets C and D is convex, for all $(x',y')\in C\times D$ we have

$$B_{K(C,x')}(0,\alpha_1)\times B_{K(D,y')}(0,\alpha_2)\subseteq B_{K(C\times D,(x',y'))}((0,0),\alpha).$$

Therefore, defining $\tilde g: X\times Y\to Y$ as $\tilde g(x,z):=g(x)-z$, condition (4.18) implies that

$$B_Y(0,1)\subseteq D\tilde g(x',y')\left[B_{K(C\times D,(x',y'))}((0,0),\alpha)\right] \quad \forall\,(x',y')\in B_{C\times D}((x_0,g(x_0)),r). \tag{4.19}$$

Now, let $(x,y)\in D_{r_1,r_2}$ and $\varepsilon>0$ be such that $d_{B_D(g(x_0),r_1)}(g(x)-y)+\varepsilon<\frac{r_2}{\alpha}$. Then, there exists $z_\varepsilon\in B_D(g(x_0),r_1)$ such that

$$\|g(x)-y-z_\varepsilon\|_Y\le d_{B_D(g(x_0),r_1)}(g(x)-y)+\varepsilon<\frac{r_2}{\alpha}. \tag{4.20}$$

By (4.19), we can apply Theorem 4.2 to $\tilde g$ and deduce that

$$d_{\tilde g^{-1}(y)\cap(C\times D)}(x,z_\varepsilon)\le \alpha\|g(x)-z_\varepsilon-y\|_Y\le \alpha d_{B_D(g(x_0),r_1)}(g(x)-y)+\alpha\varepsilon. \tag{4.21}$$

Finally, since $(x',z')\in\tilde g^{-1}(y)\cap(C\times D)$ iff $x'\in C$, $z'\in D$, and $g(x')-y=z'$, we get that

$$d_{g^{-1}(D+y)\cap C}(x)\le d_{\tilde g^{-1}(y)\cap(C\times D)}(x,z_\varepsilon). \tag{4.22}$$

Since ε is arbitrary, the result follows from (4.21)–(4.22). ∎

We can ask if we can replace the assumption ($\mathbf{H'_{cq}}$) by the following one:

($\mathbf{H''_{cq}}$) There exist $\alpha_1, \alpha_2 > 0$ and $r > 0$ such that g is continuous and Gâteaux differentiable on $B_C(x_0, r)$ and

$$\begin{aligned} &B_Y(0,1) \subseteq Dg(x)\big(B_{K(C,x)}(0,\alpha_1)\big) - B_{K(D,g(x))}(0,\alpha_2) \\ &\forall\, x \in B_{g^{-1}(D)\cap C}(x_0, r). \end{aligned} \tag{4.23}$$

As the following example shows, the answer is negative.

Example 2. Let C and D be closed sets in $\mathbb{R}^2$ defined by

$$C = \{(x,y) \in \mathbb{R}^2 : x \geq 0,\, x^2 + (y+1)^2 = 1\}$$

and

$$D = \{(x,y) \in \mathbb{R}^2 : [y = x] \text{ or } [x \geq 0,\, x^2 + (y+2)^2 = 4]\}$$

(see Figure 1), and take g to be the identity function in $\mathbb{R}^2$. Then $C \cap D = \{0\}$, $g^{-1}(C \cap D) = \{0\}$, $K(C,(0,0)) = \mathbb{R}_+ \times \{0\}$, and $K(D,(0,0)) = \{(x,x) : x \in \mathbb{R}\} \cup (\mathbb{R}_+ \times \{0\})$. Thus,

$$B_{\mathbb{R}^2}(0,1) \subseteq B_{K(C,(0,0))}(0,2) - B_{K(D,(0,0))}(0,2).$$

Similarly, we have that (4.23) holds true, and it is easy to check that (4.18) does not hold. We will show that there is no $a > 0$ such that

$$d_{g^{-1}(C\cap D)}(u) \leq a d(g(u), D) \quad \text{ for } u \in C \text{ near } 0.$$

Indeed, for $x > 0$ and $x^2 + (y+1)^2 = 1$, with (x,y) near $(0,0)$, we have

$$d_{g^{-1}(C\cap D)}(x,y) = \sqrt{x^2+y^2} \text{ and } d(g(x,y), D) \leq 2 - \sqrt{4 - (x^2+y^2)},$$

and the inequality

$$\sqrt{x^2+y^2} \leq a(2 - \sqrt{2 - (x^2+y^2)}) \approx a\frac{x^2+y^2}{4}$$

is never satisfied when (x,y) is sufficiently close to $(0,0)$.

5. Application to stochastic optimal control in continuous time. Let $T > 0$, and consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, on which a d-dimensional ($d \in \mathbb{N}^*$) Brownian motion $W(\cdot)$ is defined. We suppose that $\mathbb{F} = \{\mathcal{F}_t\}_{0\leq t\leq T}$ is the natural filtration, augmented by all $\mathbb{P}$-null sets in $\mathcal{F}$, associated to $W(\cdot)$. The filtration $\mathbb{F}$ is right-continuous, i.e., $\mathcal{F}_t = \cap_{t<u\leq T}\mathcal{F}_u$ (see [42, Chapter I, Theorem 31]).

Recall that a stochastic process $v : \Omega \times [0,T] \to \mathbb{R}^n$ is *progressively measurable* w.r.t. $\mathbb{F}$ if for all $t \in [0,T]$ the application $\Omega \times [0,t] \ni (s,\omega) \mapsto v(\omega,s) \in \mathbb{R}^n$ is $\mathcal{F}_t \times \mathcal{B}([0,t])$ measurable (here $\mathcal{B}([0,t])$ denotes the set of Borel sets in $[0,T]$). Let us define the space

$$\begin{aligned} (L^{2,2}_{\mathbb{F}})^n := \quad &\big\{v \in L^2\left(\Omega; L^2\left([0,T];\mathbb{R}^n\right)\right);\ (\omega,t) \mapsto v(\omega,t) := v(\omega)(t) \\ &\qquad\qquad\qquad\qquad \text{is progressively measurable}\big\}. \end{aligned}$$

When $n = 1$ we will simply denote $L^{2,2}_{\mathbb{F}} := (L^{2,2}_{\mathbb{F}})^1$. It is easy to see that $(L^{2,2}_{\mathbb{F}})^n$, endowed with the scalar product

$$\langle v_1, v_2\rangle_{L^{2,2}} := \mathbb{E}\left(\int_0^T v_1(t)\cdot v_2(t)\mathrm{d}t\right),$$

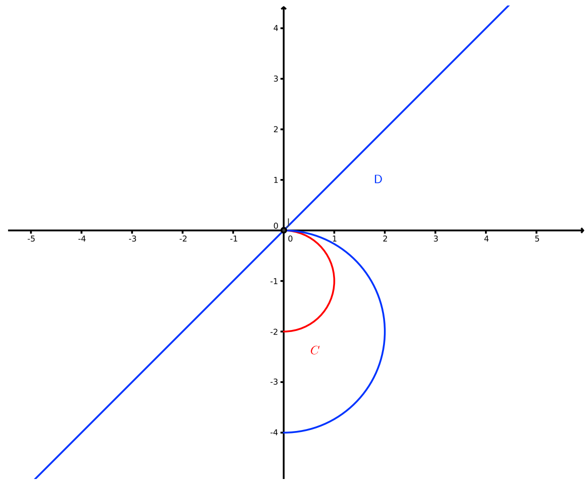

FIG. 1. *Sets C and D in Example* 2.

is a Hilbert space. We denote by $\|\cdot\|_{2,2} := \langle\cdot,\cdot\rangle_{L^{2,2}}^{\frac{1}{2}}$ the associated Hilbertian norm.

In this section we consider the stochastic optimal control problem

$$\left.\begin{aligned} &\inf_{x,u}\ \mathbb{E}\left(\int_0^T \ell(\omega,t,x(t),u(t))\mathrm{d}t + \Phi(\omega,x(T))\right)\\ &\text{s.t.}\quad \mathrm{d}x(t) = b(\omega,t,x(t),u(t))\mathrm{d}t + \sigma(\omega,t,x(t),u(t))\mathrm{d}W(t),\quad t\in(0,T),\\ &\qquad\ x(0)=\hat{x}_0,\\ &\qquad\ u\in\mathcal{U}, \end{aligned}\right\} \qquad (SP)$$

where $\mathcal{U}$ is a nonempty, closed subset of $(L^{2,2}_{\mathbb{F}})^m$ and $b:\Omega\times[0,T]\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^n$, $\sigma:\Omega\times[0,T]\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}^{n\times d}$, $\ell:\Omega\times[0,T]\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$, $\Phi:\Omega\times\mathbb{R}^n\to\mathbb{R}$, and $\hat{x}_0\in\mathbb{R}^n$ are given. In what follows we use the notation $b=(b^i)_{1\leq i\leq n}$ and $\sigma=(\sigma^{ij})_{1\leq i\leq n,\,1\leq j\leq d}$, where each b^i and σ^{ij} is real-valued. The columns of σ are written σ^j for $j=1,\ldots,d$. For $\psi=\ell,\ \Phi,\ b^j,\ \sigma^{ij}$ we will denote by $\nabla_x\psi$ the gradient of ψ w.r.t. x. We will also use the notation b_x and σ_x^j to denote, respectively, the Jacobians of b and σ^j w.r.t. x. Similar notation will be used when differentiating w.r.t. u.

In order to make problem (SP) meaningful, we need to impose some assumptions on the data. Concerning the terms defining the dynamics b and σ we will assume the following:

(A1) For $\psi=b^j,\ \sigma^{ij}$ we have

(i) ψ is $\mathcal{F}_T\otimes\mathcal{B}([0,T]\times\mathbb{R}^n\times\mathbb{R}^m)$-measurable.

(ii) For almost all (a.a.) $(\omega,t)\in\Omega\times[0,T]$ the mapping $(x,u)\to\psi(\omega,t,x,u)$ belongs to $C^1(\mathbb{R}^n\times\mathbb{R}^m)$, the application $(\omega,t)\in\Omega\times[0,T]\to\psi(\omega,t,\cdot,\cdot)\in C^1(\mathbb{R}^n\times\mathbb{R}^m)$ is progressively measurable, and there exist $c_1>0$ and $\rho_1\in L^{2,2}_{\mathbb{F}}$ such that almost surely (a.s.) in (ω,t)

$$\left\{\begin{aligned} |\psi(\omega,t,x,u)| &\leq c_1\left(\rho_1(\omega,t)+|x|+|u|\right),\\ |\nabla_x\psi(\omega,t,x,u)|+|\nabla_u\psi(\omega,t,x,u)| &\leq c_1. \end{aligned}\right. \tag{5.1}$$

Concerning the terms defining the cost functions ℓ and Φ we will assume the following:

(A2) The functions ℓ and Φ are, respectively, $\mathcal{F}_T\otimes\mathcal{B}([0,T]\times\mathbb{R}^n\times\mathbb{R}^m)$ and $\mathcal{F}_T\otimes\mathcal{B}(\mathbb{R}^n)$ measurable. Moreover, for a.a. (ω,t) the maps $(x,u)\to\ell(\omega,t,x,u)$ and $x\to\Phi(\omega,x)$ are C^1. The application $(\omega,t)\in\Omega\times[0,T]\to\ell(\omega,t,\cdot,\cdot)\in C^1(\mathbb{R}^n\times\mathbb{R}^m)$ is progressively measurable. In addition, there exist $c_2>0$, $\rho_2\in L^{2,2}_{\mathbb{F}}$, and $\rho_3\in L^2(\Omega,\mathcal{F}_T)$ such that a.s. in (ω,t) we have

$$\left\{\begin{array}{c} |\ell(\omega,t,x,u)|\leq c_2\left(\rho_2(\omega,t)+|x|^2+|u|^2\right), \\ |\nabla_x\ell(\omega,t,x,u)|+|\nabla_u\ell(\omega,t,x,u)|\leq c_2\left(\rho_2(\omega,t)+|x|+|u|\right), \\ |\Phi(\omega,x)|\leq c_2\left(\rho_3(\omega)+|x|^2\right),\ |\nabla_x\Phi(\omega,x)|\leq c_2\left(\rho_3(\omega)+|x|\right). \end{array}\right. \tag{5.2}$$

The previous assumptions are rather general and cover the case of linear quadratic problems (see, e.g., [46, Chapter 3 and Chapter 6]).

Our aim now is to provide a functional framework for problem (SP) that will allow us to apply the abstract results in the previous sections to derive a first order optimality condition at a local solution. We proceed as in [2], and we focus first on writing the SDE constraint in the form of an equality constraint in a suitable function space.

Let us consider the mapping $I:\mathbb{R}^n\times(L^{2,2}_{\mathbb{F}})^n\times(L^{2,2}_{\mathbb{F}})^{n\times d}\to(L^{2,2}_{\mathbb{F}})^n$ defined by

$$I(x_0,x_1,x_2)(\cdot):=x_0+\int_0^{(\cdot)}x_1(s)\mathrm{d}s+\sum_{j=1}^d\int_0^{(\cdot)}x_2^j(s)\mathrm{d}W^j(s). \tag{5.3}$$

Standard results in Itô's stochastic calculus theory imply that I is well defined. Consider the *Itô space* $\mathcal{I}^n:=I(\mathbb{R}^n\times(L^{2,2}_{\mathbb{F}})^n\times(L^{2,2}_{\mathbb{F}})^{n\times d})$. Endowed with the scalar product

$$\langle x,y\rangle_{\mathcal{I}^n}:=x_0\cdot y_0+\mathbb{E}\left(\int_0^T x_1(t)\cdot y_1(t)\mathrm{d}t\right)+\sum_{j=1}^d\mathbb{E}\left(\int_0^T x_2^j(t)\cdot y_2^j(t)\mathrm{d}t\right), \tag{5.4}$$

we have that $\mathcal{I}^n$ is a Hilbert space, which, since I is injective (see [2, Lemma 2.1]), can be identified with $\mathbb{R}^n\times(L^{2,2}_{\mathbb{F}})^n\times(L^{2,2}_{\mathbb{F}})^{n\times d}$. Let us denote by $\|\cdot\|_{\mathcal{I}^n}:=\langle\cdot,\cdot\rangle_{\mathcal{I}^n}^{\frac{1}{2}}$ the associated Hilbertian norm.

Recall that by definition $x\in\mathcal{I}^n$ solves the controlled SDE in (SP) iff

$$x(t)=x_0+\int_0^t b(s,x(s),u(s))\mathrm{d}s+\int_0^t\sigma(s,x(s),u(s))\mathrm{d}W(s)\quad\forall\, t\in[0,T]. \tag{5.5}$$

It is well known that under **(A1)** (5.5) admits a unique solution $x\in\mathcal{I}^n$ (see, e.g., [34, Chapter 5]). It is also known that $\mathbb{E}(\sup_{t\in[0,T]}|x(t)|^2)$ is finite (see, e.g., [2, Lemma 2.2]). More precise information is given by the following lemma, whose proof is by now standard. We provide here the details of the proof since we need to obtain explicit expressions for the involved constants.

LEMMA 5.1. *For all $t\in[0,T]$ and $u\in(L^{2,2}_{\mathbb{F}})^m$, the solution $x\in\mathcal{I}^n$ satisfies*

$$\mathbb{E}\left(\sup_{s\in[0,t]}|x(s)|^2\right)=c\left[|x_0|^2+\mathbb{E}\left(\int_0^t|b(s,0,u)|^2\mathrm{d}s\right)+\mathbb{E}\left(\int_0^t|\sigma(s,0,u)|^2\mathrm{d}s\right)\right], \tag{5.6}$$

where $c=\max\{24,6T\}e^{6Tc_1^2\max\{T,4d\}}$.

Proof. Using the inequality $(a_1+a_2+a_3)^2 \le 3(a_1^2+a_2^2+a_3^2)$ for all a_1, a_2, and a_3 in $\mathbb{R}$ and Jensen's inequality, for all $0 \le s \le t \le T$ expression (5.5) yields

$$|x(s)|^2 \le 3\left(|x_0|^2 + s\int_0^s |b(s',x(s'),u(s'))|^2 \mathrm{d}s' + \left|\int_0^s \sigma(s',x(s'),u(s'))\mathrm{d}W(s')\right|^2\right).$$

By the linear growth condition in (5.1) and the fact that $x \in \mathcal{I}^n$ and $u \in (L^{2,2}_{\mathbb{F}})^m$, we have that $\sigma(\cdot,x(\cdot),u(\cdot)) \in (L^{2,2}_{\mathbb{F}})^{n\times d}$, and so, for each $j=1,\dots,d$, the $\mathbb{R}^n$-valued process $[0,T] \ni s \mapsto \int_0^s \sigma^j(s',x(s'),u(s'))\mathrm{d}W^j(s')$ is a martingale. Thus, defining $g(t) := \mathbb{E}(\sup_{s\in[0,t]}|x(s)|^2)$, Doob's inequality and the Lipschitz property of b and σ w.r.t. x in (5.1) imply that

$$\begin{aligned} g(t) &\le 3\left[|x_0|^2 + T\mathbb{E}\left(\int_0^t |b(s,x(s),u(s))|^2\mathrm{d}s\right) + 4\mathbb{E}\left(\int_0^t |\sigma(s,x(s),u(s))|^2\mathrm{d}s\right)\right] \\ &\le 3\left[|x_0|^2 + 2T\mathbb{E}\left(\int_0^t \left[|b(s,0,u(s))|^2 + c_1^2|x(s)|^2\right]\right)\mathrm{d}s\right. \\ &\qquad \left. +8\mathbb{E}\left(\int_0^t \left[|\sigma(s,0,u(s))|^2 + dc_1^2|x(s)|^2\right]\mathrm{d}s\right)\right] \\ &\le a_1 + a_2\int_0^t g(s)\mathrm{d}s, \end{aligned}$$

where

$$a_1 = \max\{24,6T\}\left[|x_0|^2 + \mathbb{E}\left(\int_0^t |b(s,0,u(s))|^2\mathrm{d}s\right) + \mathbb{E}\left(\int_0^t |\sigma(s,0,u(s))|^2\mathrm{d}s\right)\right],$$

and $a_2 = 6c_1^2\max\{T,4d\}$. The result then follows from Gronwall's lemma. □

Remark 5.2. Estimates of the form (5.6) can be easily extended to any power $p>1$ by using in the previous proof the Burkholder–Davis–Gundy inequality (see, e.g., [40]) instead of Doob's inequality.

Now, let us consider the application $g : \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \to \mathcal{I}^n$ defined by

$$g(x,u)(\cdot) := \hat{x}_0 + \int_0^{(\cdot)} b(s,x(s),u(s))\mathrm{d}s + \int_0^{(\cdot)} \sigma(s,x(s),u(s))\mathrm{d}W(s) - x(\cdot), \tag{5.7}$$

which defines the SDE constraint in (SP) by imposing $g(x,u)=0$. Consider also the application $f : \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \to \mathbb{R}$ defined by

$$f(x,u) := \mathbb{E}\left(\int_0^T \ell(t,x(t),u(t))\mathrm{d}t + \Phi(x(T))\right),$$

which describes the cost functional in (SP). Assumption **(A2)** implies that f is well defined. Problem (SP) can thus be rewritten in the following abstract form:

$$\inf \;\; f(x,u) \;\text{ subject to }\; g(x,u)=0, \quad u \in \mathcal{U}. \tag{SP}$$

We proceed now to verify that f and g satisfy the assumptions considered in section 3, when the underlying space given by $X := \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$.

We begin by studying some properties of g. The following result is proved in the appendix in [2]. For the sake of completeness we provide here a short proof.

LEMMA 5.3. *Under* **(A1)** *the mapping g is Lipschitz continuous and Gâteaux differentiable. Its Gâteaux derivative $Dg(x,u) : \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \to \mathcal{I}^n$ is given by*

$$
\begin{aligned}
Dg(x,u)(z,v)(\cdot) \quad = \quad & \int_0^{(\cdot)} \left[b_x(t,x(t),u(t))z(t) + b_u(t,x(t),u(t))v(t)\right] \mathrm{d}t \\
& + \sum_{j=1}^d \int_0^{(\cdot)} \left[\sigma_x^j(t,x(t),u(t))z(t) + \sigma_u^j(t,x(t),u(t))v(t)\right] \mathrm{d}W^j(t) \\
& - z(\cdot)
\end{aligned} \tag{5.8}
$$

for all $(z,v) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$.

Proof. Note that for any (x,u_1), $(y,u_2) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$ we have

$$
\begin{aligned}
& \|g(x,u_1)(\cdot) - g(y,u_2)(\cdot)\|^2_{\mathcal{I}^n} \\
= & |x_0 - y_0|^2 + \mathbb{E}\left(\int_0^T \left|b(t,x(t),u^1(t)) - b(t,y(t),u^2(t)) + y_1(t) - x_1(t)\right|^2 \mathrm{d}t\right) \\
& + \sum_{j=1}^d \mathbb{E}\left(\int_0^T \left|\sigma^j(t,x(t),u^1(t)) - \sigma^j(t,y(t),u^2(t)) + y_2^j(t) - x_2^j(t)\right|^2 \mathrm{d}t\right),
\end{aligned}
$$

which, by the Lipschitz assumption in (5.1), is bounded by

$$
c\left[\|x-y\|^2_{\mathcal{I}^n} + \mathbb{E}\left(\int_0^T |x(t)-y(t)|^2 \mathrm{d}t\right) + \mathbb{E}\left(\int_0^T |u^1(t)-u^2(t)|^2 \mathrm{d}t\right)\right]
$$

for some constant $c > 0$. Now, as in the proof of Lemma 5.1, by Jensen's and Doob's inequalities we easily get the existence of a constant $c' > 0$ such that

$$
\mathbb{E}\left(\int_0^T |x(t)-y(t)|^2 \mathrm{d}t\right) \le c' \|x-y\|^2_{\mathcal{I}^n},
$$

from which the Lipschitz property of g easily follows. Now, for $j = 1, \ldots, d$ let us set

$$
Db(t,x,u)(z,v) = b_x(t,x,u)z + b_u(t,x,u)v, \quad D\sigma^j(t,x,u)(z,v) = \sigma_x^j(t,x,u)z + \sigma_u^j(t,x,u)v
$$

and define

$$
\begin{aligned}
I_1 &:= \mathbb{E}\left(\int_0^T \left[\frac{b(t,x(t)+\tau z(t),u(t)+\tau v(t)) - b(t,x(t),u(t))}{\tau} - Db(t,x(t),u(t))(z(t),v(t))\right]^2 \mathrm{d}t\right), \\
I_2^j &:= \mathbb{E}\left(\int_0^T \left[\frac{\sigma^j(t,x(t)+\tau z(t),u(t)+\tau v(t)) - \sigma^j(t,x(t),u(t))}{\tau} - D\sigma^j(t,x(t),u(t))(z(t),v(t))\right]^2 \mathrm{d}t\right).
\end{aligned}
$$

By the Lipschitz property of b and σ in (5.1) and the dominated convergence theorem, we get that I_1 and I_2^j tend to 0 as $\tau \downarrow 0$. This implies that

$$
\mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \ni (x,u) \mapsto \int_0^{(\cdot)} b(s,x(s),u(s))\mathrm{d}s + \int_0^{(\cdot)} \sigma(s,x(s),u(s))\mathrm{d}W(s) \in \mathcal{I}^n
$$

is directionally differentiable with directional derivative

$$
\begin{aligned}
\mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \ni (z,v) \mapsto \quad & \int_0^{(\cdot)} Db(t,x(t),u(t))(z(t),v(t))\mathrm{d}t \\
& + \sum_{j=1}^d \int_0^{(\cdot)} D\sigma^j(t,x(t),u(t))(z(t),v(t))\mathrm{d}t.
\end{aligned}
$$

The continuity of the linear application above follows easily from the bounds in the second relation in (5.1). Finally, since $\mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \ni (x,u) \mapsto x \in \mathcal{I}^n$ is C^∞ with derivative given by $\mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m \ni (z,v) \mapsto z \in \mathcal{I}^n$, we obtain (5.8). ∎

The previous lemma yields the following result.

Lemma 5.4. *For every $(x,u) \in \mathcal{I}^n \times (L^{2,2}_{\mathbb{F}})^m$ and $\delta \in \mathcal{I}^n$, there exists a unique $z \in \mathcal{I}^n$ such that $Dg(x,u)(z,0) = \delta$. Moreover, there exists a constant $c > 0$, independent of (x,u,z,δ), such that $\|z\|_{\mathcal{I}^n} \le c\|\delta\|_{\mathcal{I}^n}$.*

Proof. By Lemma 5.3, we have that $Dg(x,u)(z,0) = \delta$ is equivalent to the SDE

$$\begin{aligned} \mathrm{d}z &= [b_x(t,x(t),u(t))z(t) - \delta_1]\,\mathrm{d}t + [\sigma_x(t,x(t),u(t))z(t) - \delta_2]\,\mathrm{d}W(t),\\ z(0) &= -\delta_0. \end{aligned}$$

The existence and uniquenes of a solution z of this equation is well known (see, e.g., [34, Chapter 5]). Moreover, using that $\|b_x\|_\infty \le c_1$ and $\|\sigma_x\|_\infty \le c_1$, Lemma 5.1 implies the existence of a constant $c > 0$, independent of (x,u,z,δ), such that

$$\|z\|_{\mathcal{I}^n} \le c\left[|\delta_0|^2 + \mathbb{E}\left(\int_0^T |\delta_1|^2 \mathrm{d}t\right) + \mathbb{E}\left(\int_0^T |\delta_2|^2 \mathrm{d}t\right)\right].$$

The result follows. ∎

As a consequence of the last two lemmas and Theorem 4.2, g satisfies (4.1) with $C := \mathcal{I}^n \times \mathcal{V}$ and $\alpha = c$, where $\mathcal{V}$ is any closed set of $(L^{2,2}_{\mathbb{F}})^m$. Therefore, the following result holds true.

Corollary 5.5. *For any closed set $\mathcal{V} \subseteq (L^{2,2}_{\mathbb{F}})^m$, we have*

$$d_{g^{-1}(y)\cap(\mathcal{I}^n\times\mathcal{V})}(x,u) \le c\|g(x,u) - y\|^2_{\mathcal{I}^n} \quad \forall\, x,\ y \in \mathcal{I}^n \ \text{ and } u \in \mathcal{V}.$$

Now, we consider the properties of the cost functional f.

Lemma 5.6. *The function f is locally Lipschitz and Gâteaux differentiable, with*

$$\begin{aligned} Df(x,u)(z,v) &= \mathbb{E}\left(\int_0^T [\ell_x(t,x(t),u(t))z(t) + \ell_u(t,x(t),u(t))v(t)]\,\mathrm{d}t\right)\\ &\quad + \mathbb{E}\left(D\Phi(x(T))z(T)\right). \end{aligned} \tag{5.9}$$

Proof. For $\tau \in [0,1]$, set $x_\tau := x_1 + \tau(x_2 - x_1)$, $u_\tau := u_1 + \tau(u_2 - u_1)$, $\delta x = x_2 - x_1$, and $\delta u = u_2 - u_1$. We have that

$$\begin{aligned} |f(x_2,u_2) - f(x_1,u_1)| \le\ & \mathbb{E}\left(\int_0^T \int_0^1 |D\ell(t,x_\tau(t),u_\tau(t))(\delta x(t),\delta u(t))|\,\mathrm{d}\tau\mathrm{d}t\right)\\ & + \mathbb{E}\left(\int_0^1 |D\Phi(x_\tau(T))\delta x(T)|\,\mathrm{d}\tau\right). \end{aligned}$$

By the second assumption in (5.2) we can find $c > 0$ such that

$$\begin{aligned} |D\ell(t,x_\tau(t),u_\tau(t))(\delta x(t),\delta u(t))| &\le c(1 + |x_\tau(t)| + |u_\tau(t)|)(|\delta x(t)| + |\delta u(t)|)\\ &\le c(1 + |x_1(t)| + |\delta x(t)| + |u_1(t)| + |\delta u(t)|)(|\delta x(t)| + |\delta u(t)|), \end{aligned}$$

which, by the Cauchy–Schwarz inequality, implies that

$$\begin{aligned} &\left[\mathbb{E}\left(\int_0^T \int_0^1 |D\ell(t,x_\tau(t),u_\tau(t))(\delta x(t),\delta u(t))|\,\mathrm{d}\tau\mathrm{d}t\right)\right]^2\\ &\le c'\mathbb{E}\left(\int_0^T (1 + |x_1(t)|^2 + |\delta x(t)|^2 + |u_1(t)|^2 + |\delta u(t)|^2)\mathrm{d}t\right)(\|\delta x\|^2_{2,2} + \|\delta u\|^2_{2,2}). \end{aligned}$$

Analogously, there exists $c'' > 0$ such that

$$\left[\mathbb{E}\left(\int_0^1 |D\Phi(x_\tau(T))\delta x(T)|\,\mathrm{d}\tau\right)\right]^2 \leq c''\mathbb{E}\left(1+|x_1(T)|^2+|\delta x(T)|^2\right)\mathbb{E}\left(|\delta x(T)|^2\right),$$

from which the local Lipschitz property for f follows. Now, we prove the formula for the directional derivative. Consider the term

$$\mathbb{E}\left(\int_0^T \left[\frac{\ell(t,x(t)+\tau z(t),u(t)+\tau v(t))-\ell(t,x(t)+\tau z(t),u(t)+\tau v(t))}{\tau} - D\ell(t,x(t),u(t))\right]\mathrm{d}t\right). \tag{5.10}$$

Since ℓ is Gâteaux differentiable, the expression inside the integral converges to zero pointwisely. Now, writing the ratio inside the integral in integral form, if $\tau < 1$, we have

$$\int_0^1 D\ell(t,x_{\gamma\tau}(t),u_{\gamma\tau}(t))(z(t),v(t))\mathrm{d}\gamma \leq c(1+|x(t)|+|z(t)|+|u(t)|+|v(t)|)(|z(t)|+|v(t)|),$$

where $x_{\gamma\tau} := x+\gamma\tau z$ and $u_{\gamma\tau} := u+\gamma\tau v$. The term $D\ell(t,x(t),u(t))$ is dominated by $c(1+|x(t)|+|u(t)|)$, and thus we can pass to the limit to obtain that the term in (5.10) tends to 0 as $\tau \downarrow 0$. Analogously, as $\tau \downarrow 0$,

$$\mathbb{E}\left(\frac{\Phi(x(T)+\tau z(T))-\Phi(x(T))}{\tau} - D\Phi(x(T))z(T)\right) \to 0.$$

Formula (5.9) follows. ◻

As is customary in optimal control theory, it is convenient to introduce the Hamiltonian $H : \Omega\times]0,T[\times\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^{n\times d}\times\mathbb{R}^m \to \mathbb{R}$ defined as

$$H(\omega,t,x,p,q,u) := \ell(\omega,t,y,u) + p\cdot b(\omega,t,x,u) + \sum_{i=1}^d q^i\cdot\sigma^i(\omega,t,x,u).$$

With the help of Theorem 3.1 and Corollary 5.5 we prove in the next result the existence of Lagrange multipliers for problem (SP), and, as a consequence, we recover a *weak version* of the stochastic Pontryagin minimum principle first proved in [3] (see relations (5.12)–(5.13) below).

THEOREM 5.7. *Suppose that $(\bar{x},\bar{u})$ is a local solution of problem (SP); then there exists a Lagrange multiplier $\lambda \in \mathcal{I}^n$ such that*

$$0 \in Df(\bar{x},\bar{u}) + Dg(\bar{x},\bar{u})^*\lambda + \{0\}\times N_{\mathcal{U}}(\bar{u}). \tag{5.11}$$

In particular, defining $\bar{p} := \lambda_1$, $\bar{q} := \lambda_2$, we have that $\bar{p}\in\mathcal{I}^n$, $\bar{q}\in(L^{2,2}_{\mathbb{F}})^{n\times d}$, and the following relations hold true:

$$\bar{p}(\cdot) = \nabla_x\Phi(\bar{x}(T)) + \int_{(\cdot)}^T \nabla_x H(s,\bar{x}(s),\bar{p}(s),\bar{q}(s),\bar{u}(s))\mathrm{d}s - \int_{(\cdot)}^T \bar{q}(s)\mathrm{d}W(s), \tag{5.12}$$

$$\mathbb{E}\left(\int_0^T \nabla_u H(t,\bar{x}(t),\bar{p}(t),\bar{q}(t),\bar{u}(t))\cdot v(t)\mathrm{d}t\right) \geq 0 \qquad \forall\, v\in T_{\mathcal{U}}(\bar{u}). \tag{5.13}$$

If, in addition, $K(\mathcal{U},\bar{u})$ is convex, then (5.13) is valid for all $v\in K(\mathcal{U},\bar{u})$.

Proof. Lemmas 5.3 and 5.6 imply that g and f satisfy the assumptions ($\mathbf{H_g}$) and ($\mathbf{H_f}$), respectively. Since Corollary 5.5 implies that the constraint system in (SP) is calm at $(\bar{x},\bar{u})$, the existence of $\lambda \in \mathcal{I}^n$ satisfying (5.11) follows directly from Theorem 3.1. Noticing that (5.11) can be written as

$$\begin{array}{rcl} D_x f(\bar{x},\bar{u}) + D_x g(\bar{x},\bar{u})^*\lambda & = & 0, \\ \langle D_u f(\bar{x},\bar{u}) + D_u g(\bar{x},\bar{u})^*\lambda, v\rangle_{(L_{\mathbb{F}}^{2,2})^m} & \geq & 0 \quad \forall\, v \in T_{\mathcal{U}}(\bar{u}), \end{array} \tag{5.14}$$

Theorem 3.12 in [2] directly yields relations (5.12)–(5.13). Finally, if $K(\mathcal{U},\bar{u})$ is convex, by Theorem 3.1(ii) we have

$$0 \in Df(\bar{x},\bar{u}) + Dg(\bar{x},\bar{u})^*\lambda + \{0\} \times K(\mathcal{U},\bar{u})^0.$$

Reasoning as before, we get that (5.13) is valid for all $v \in K(\mathcal{U},\bar{u})$. The result follows. ∎

5.1. Comments and extensions. Let us provide some comments on the previous result.

(i) As pointed out in [2], it is not clear that in general the function g defined in (5.7) is C^1. Therefore, standard Lagrange multiplier results, in infinite dimensions, are not directly applicable to problem (SP). Nevertheless, we have shown that the function g, which determines the constraint system in (SP), satisfies the weak regularity assumptions introduced in section 4, which are sufficient to guarantee the metric regularity of the aforementioned constraint system and, hence, the existence of a Lagrange multiplier associated to a local solution. Let us stress the fact that the verification of ($\mathbf{H_{cq}}$) is rather simple in this case, because it amounts to checking the stability of solutions to linear SDEs under random additive perturbations of the right-hand side (see Lemma 5.4).

(ii) As mentioned in the introduction, the main contribution of this section is the simple justification of the existence of a Lagrange multiplier associated to the SDE constraint $g(x,u) = 0$. It is reasonable to conjecture that this fact plays an essential role in a rigorous sensitivity analysis for problem (SP) under general perturbations of b and σ in the controlled SDE. Indeed, it is well known in optimization theory that Lagrange multipliers are central in the study of the sensitivity of the optimal cost functional under perturbations of the data (see, e.g., [9, section 6.5], [5, Chapter 4] and the references therein). In the case of deterministic optimal control problems, the literature is also very rich (see, e.g., [10], [29, Chapter 2], and the bibliographic notes in [5, section 7.5]). On the other hand, to the best of our knowledge, there exist only a few results on the sensitivity analysis for stochastic optimal control problems. We refer the reader to [2] for convex problems and functional random perturbations of the dynamics and to [36] for a class of nonconvex problems and finite dimensional perturbations.

Notice that the results in sections 3 and 4, and the identification between Lagrange multipliers and adjoint states [2], allowed us to give a direct proof of the weak version of Pontryagin's minimum principle (5.12)–(5.13). This result was first proved in [3] by writing the state x as a function of u, using that for each u there exists a unique solution $x[u]$ of $g(x,u) = 0$, and expanding the cost function in terms of u. This approach is useful to establish (5.12)–(5.13) but hides the importance of the pair (p,q) as a Lagrange multiplier associated to the SDE constraint.

(iii) In the particular case of pointwise control constraints

$$\mathcal{U} := \{u \in (L^{2,2}_{\mathbb{F}})^m \ ; \ u(\omega,t) \in U \ \text{ a.s}\},$$

where $U \subseteq \mathbb{R}^m$ is a nonempty closed set, a result stronger than Theorem 5.7 was shown in [41]. In this paper, the author shows that a modified Hamiltonian H, which involves an additional pair of adjoint processes, is a.s. pointwisely minimized at $\bar{u}(\omega,t)$. In this result, which is the stochastic analogue of the classical Pontryagin minimum principle, no regularity assumptions on the data w.r.t. u are imposed. On the other hand, stronger assumptions w.r.t. the dependence on the state variable x are assumed (which involve requirements on the second order derivatives of ℓ, Φ, b, and σ).

(iv) A straightforward extension of Theorem 5.7 is the case where the initial point $\hat{x}_0$ is also a decision variable. More precisely, let $\mathcal{X}_0 \subseteq \mathbb{R}^n$ be a closed set, and consider the following extension of problem (SP):

$$\left.\begin{aligned}
&\inf_{x,\hat{x}_0,u} \ \mathbb{E}\left(\int_0^T \ell(\omega,t,x(t),u(t))\mathrm{d}t + \Phi(\omega,x(T))\right)\\
&\text{s.t.} \quad \mathrm{d}x(t) = b(\omega,t,x(t),u(t))\mathrm{d}t + \sigma(\omega,t,x(t),u(t))\mathrm{d}W(t), \quad t\in(0,T),\\
&\qquad x(0) = \hat{x}_0 \in \mathcal{X}_0,\\
&\qquad u \in \mathcal{U}.
\end{aligned}\right\} \qquad (SP')$$

Then, this problem can be written in the abstract form

$$\inf \ \ f(x,u) \ \text{ subject to } \ \tilde{g}(x,u) \in \mathcal{I}^n \times \mathcal{X}_0, \quad u \in \mathcal{U}, \qquad (SP')$$

where

$$\tilde{g}(x,u) := \left(x(0) + \int_0^{(\cdot)} b(s,x(s),u(s))\mathrm{d}s + \int_0^{(\cdot)} \sigma(s,x(s),u(s))\mathrm{d}W(s) - x(\cdot), x(0) \right).$$

Suppose that $(\bar{x},\bar{u}) \in \mathcal{I}^n \times \mathcal{U}$ is a local solution to (SP'), and assume that **(A1)–(A2)** hold true. Using the surjectivity property of the derivative of the first coordinate of $\tilde{g}$ (as in Lemma 5.4), it is easy to check that (4.18) in $(\mathbf{H}'_{\mathbf{cq}})$ is satisfied at $(\bar{x},\bar{u})$ (with $C = \mathcal{I}^n \times \mathcal{U}$ and $D = \mathcal{I}^n \times \mathcal{X}_0$). Thus, by Theorems 4.6 and 3.3, and reasoning as in the proof of Theorem 5.7, we obtain the existence of $\bar{p} \in \mathcal{I}^n$ and $\bar{q} \in (L^{2,2}_{\mathbb{F}})^{n\times d}$ such that

(5.15)
$$\begin{aligned}
&\bar{p}(\cdot) = \nabla_x \Phi(\bar{x}(T)) + \int_{(\cdot)}^T \nabla_x H(s,\bar{x}(s),\bar{p}(s),\bar{q}(s),\bar{u}(s))\mathrm{d}s - \int_{(\cdot)}^T \bar{q}(s)\mathrm{d}W(s),\\
&-\bar{p}(0) \in N_{\mathcal{X}_0}(\bar{x}(0)),\\
&\text{and} \quad \mathbb{E}\left(\int_0^T \nabla_u H(t,\bar{x}(t),\bar{p}(t),\bar{q}(t),\bar{u}(t))\cdot v(t)\mathrm{d}t\right) \geq 0 \ \text{ for all } \ v \in T_{\mathcal{U}}(\bar{u}).
\end{aligned}$$

(v) Another easy extension is the case where finitely many final constraints on the state, in expectation form, are added to problem (SP). In this case, a qualification condition has to be imposed on the local solution $(\bar{x},\bar{u})$ in order to ensure that $(\mathbf{H}'_{\mathbf{cq}})$ holds. We refer the reader to [2] for a more detailed discussion on this matter. The case of final pointwise constraints having the form $x(\omega,T) \in \mathcal{X}_T$, for some closed set $\mathcal{X}_T \subseteq \mathbb{R}^n$, and with probability one, remains as an interesting open problem.

6. Application to a class of stochastic control problems in discrete time. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and, as in the previous section, denote by $\mathbb{E}$ the expectation under $\mathbb{P}$. Let $w_1, \dots, w_N$ be N independent $\mathbb{R}^d$-valued random variables defined in $(\Omega, \mathcal{F}, \mathbb{P})$ such that for all $k = 1, \dots, N$ the coordinates of $w_k = (w_k^1, \dots, w_k^d)$ are independent and satisfy

$$\mathbb{E}(w_k^i) = 0, \quad \mathbb{E}(|w_k^i|^2) = 1.$$

Define $w_0 := 0$ and for $k = 0, \dots, N$ set $\mathcal{F}_k := \sigma(w_0, \dots, w_k)$, the sigma-algebra generated by $w_0, \dots, w_k$, and

$$L^2_{\mathcal{F}_k} := \left\{ y \in L^2(\Omega)\ ;\ y \text{ is } \mathcal{F}_k \text{ measurable} \right\}.$$

Let $\mathcal{U} \subseteq \Pi_{k=0}^{N-1} (L^2_{\mathcal{F}_k})^m$ be a nonempty closed set. In this section we consider the following discrete time stochastic optimal control problem (see [35]):

$$\left.\begin{aligned} &\inf\ \mathbb{E}\left(\textstyle\sum_{k=0}^{N-1} \ell(k, x_k, u_k) + \Phi(x_N) \right) \\ &\text{s.t.} \quad x_{k+1} = b(k, x_k, u_k) + \sigma(k, x_k, u_k) w_{k+1}, \qquad k = 0, \dots, N-1, \\ &\qquad\ \ x_0 = \hat{x}_0 \in \mathbb{R}^n, \\ &\qquad\ \ x \in \Pi_{k=0}^{N} (L^2_{\mathcal{F}_k})^n, \quad u \in \mathcal{U}, \end{aligned}\right\} \qquad (SP_d)$$

where, denoting $[0 : N-1] := \{0, \dots, N-1\}$, $\ell : [0 : N-1] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, $\Phi : \mathbb{R}^n \to \mathbb{R}$, $b : [0 : N-1] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, and $\sigma : [0 : N-1] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n \times d}$ are Borel measurable functions. Denoting σ^j $(j = 1, \dots, d)$ the jth column of σ, for $\psi = b$, σ^j we suppose that ψ is $\mathcal{C}^1$ w.r.t. (x, u) and the existence of $c_1 > 0$ such that for all $k \in [0 : N-1]$

$$\left\{ \begin{array}{c} |\psi(k, x, u)| \le c_1 (1 + |x| + |u|), \\ |\psi_x(k, x, u)| + |\psi_u(k, x, u)| \le c_1. \end{array} \right. \tag{6.1}$$

Similarly, in the remainder of this section we will assume that there exists $c_2 > 0$ such that for all $k \in [0 : N-1]$

$$\left\{ \begin{array}{c} |\ell(k, x, u)| \le c_2 (1 + |x| + |u|)^2, \\ |\ell_x(k, x, u)| + |\ell_u(k, x, u)| \le c_2 (1 + |x| + |u|), \\ |\Phi(x)| \le c_2 (1 + |x|)^2, \ |\Phi_x(x)| \le c_2 (1 + |x|). \end{array} \right. \tag{6.2}$$

As in section 5 we introduce now a Hilbert space for the state x which is suitable for the application of the results in sections 3 and 4. Set $X_0 = \mathbb{R}^n$, and, given $k \in [1 : N]$, define

$$X_k := \left\{ y^0_{k-1} + \sum_{i=1}^{d} y^i_{k-1} w^i_k\ ;\ y^i_{k-1} \in \left(L^2_{\mathcal{F}_{k-1}} \right)^n \quad \forall\, i = 0, \dots, d \right\}.$$

Endowed with the scalar product

$$\langle x, x' \rangle_{X_k} := \mathbb{E}\left(\sum_{i=0}^{d} y^i_{k-1} \cdot z^i_{k-1} \right) \quad \forall\, x = y^0_{k-1} + \sum_{i=1}^{d} y^i_{k-1} w^i_k, \quad x' = z^0_{k-1} + \sum_{i=1}^{d} z^i_{k-1} w^i_k,$$

the following elementary result shows that X_k is a Hilbert space.

LEMMA 6.1. *For every* $(y_{k-1}^0, y_{k-1}^1, \dots, y_{k-1}^d) \in (L^2_{\mathcal{F}_{k-1}})^n \times (L^2_{\mathcal{F}_{k-1}})^{n\times d}$ *we have*

$$\mathbb{E}\left(\left|y_{k-1}^0 + \sum_{i=1}^d y_{k-1}^i w_k^i\right|^2\right) = \sum_{i=0}^d \mathbb{E}\left(|y_{k-1}^i|^2\right). \tag{6.3}$$

As a consequence, for every $k \in [1:N]$ *the linear operator* $I : (L^2_{\mathcal{F}_{k-1}})^n \times (L^2_{\mathcal{F}_{k-1}})^{n\times d} \to X_k$*, defined as*

$$I(y_{k-1}^0, y_{k-1}^1, \dots, y_{k-1}^d) := y_{k-1}^0 + \sum_{i=1}^d y_{k-1}^i w_k^i,$$

is a bijection.

Proof. Relation (6.3) follows directly from the relations

$$\begin{aligned}
&\mathbb{E}\left(y_{k-1}^0 \cdot y_{k-1}^i w_k^i\right) = \mathbb{E}\left(y_{k-1}^0 \cdot y_{k-1}^i \mathbb{E}\left(w_k^i|\mathcal{F}_{k-1}\right)\right) = 0 \quad \forall\, i \in [1:d],\\
&\mathbb{E}\left(y_{k-1}^i \cdot y_{k-1}^j w_k^j w_k^i\right) = \mathbb{E}\left(y_{k-1}^i \cdot y_{k-1}^j \mathbb{E}\left(w_k^j w_k^i|\mathcal{F}_{k-1}\right)\right) = \begin{cases} \mathbb{E}\left(|y_{k-1}^i|^2\right) & \text{if } i=j,\\ 0 & \text{otherwise.}\end{cases}
\end{aligned}$$

By definition of X_k we only need to show that I is injective. But this is clear because if

$$I(y_{k-1}^0, y_{k-1}^1, \dots, y_{k-1}^d) = 0,$$

then (6.3) implies that $\mathbb{E}\left(|y_{k-1}^i|^2\right) = 0$ and so $y_{k-1}^i = 0$ almost everywhere for all $i \in [0:d]$. ∎

Define $g : \Pi_{k=0}^N X_k \times \Pi_{k=0}^{N-1}\left(L^2_{\mathcal{F}_k}\right)^m \to \Pi_{k=0}^N X_k$ as

$$\begin{aligned}
g_0(x,u) &:= \hat{x}_0 - x_0,\\
g_{k+1}(x,u) &:= b(k,x_k,u_k) + \sigma(k,x_k,u_k)w_{k+1} - x_{k+1} \qquad \forall\, k = 0,\dots,N-1,
\end{aligned}$$

and $f : \Pi_{k=0}^N X_k \times \Pi_{k=0}^{N-1}\left(L^2_{\mathcal{F}_k}\right)^m \to \mathbb{R}$ as

$$f(x,u) := \mathbb{E}\left(\sum_{k=0}^{N-1} \ell(k,x_k,u_k) + \Phi(x_N)\right).$$

Under these notations, problem (SP_d) can be rephrased as

$$\inf \quad f(x,u) \ \text{ subject to } \ g(x,u) = 0, \quad u \in \mathcal{U}. \tag{SP_d}$$

As in the previous section, we prove now that if we set $X := \Pi_{k=0}^N X_k \times \Pi_{k=0}^{N-1}\left(L^2_{\mathcal{F}_k}\right)^m$, then under our assumptions the mappings f and g satisfy the assumptions in section 3.

LEMMA 6.2. *The following assertions hold true:*

(i) *The mapping* g *is Lipschitz and Gâteaux differentiable. For* $(x,u), (z,v) \in X$ *the directional derivative of* g *at* (x,u) *in the direction* (z,v) *is given by* $Dg(x,u)(z,v) = (Dg_0(x,u)(z,v), \dots, Dg_N(x,u)(z,v))$*, where*

$$\begin{aligned}
Dg_0(x,u)(z,v) &= -z_0,\\
Dg_{k+1}(x,u)(z,v) &= b_{(x,u)}(k,x_k,u_k)(z_k,v_k)\\
&\quad + \textstyle\sum_{i=1}^d \sigma^i_{(x,u)}(k,x_k,u_k)(z_k,v_k)w^i_{k+1} - z_{k+1}
\end{aligned} \tag{6.4}$$

for all $k = 0, \ldots, N-1$.

(ii) *The mapping* f *is locally Lipschitz and Gâteaux differentiable, with*

$$Df(x,u)(z,v) = \mathbb{E}\left(\sum_{k=0}^{N-1} \ell_{(x,u)}(k, x_k, u_k)(z_k, v_k) + D\Phi(x_N) z_N\right) \tag{6.5}$$

for all (x,u), $(z,v) \in X$.

Proof. We only prove assertion (i) since the proof of (ii) is analogous. By the second relation in assumption (6.1), there exists $c > 0$ such that for all $k = 0, \ldots, N-1$,

$$\begin{aligned}
&\|g_{k+1}(x^1,u^1) - g_{k+1}(x^2,u^2)\|^2_{X_{k+1}} \\
&= \mathbb{E}\left(|b(k,x^1_k,u^1_k) - b(k,x^2_k,u^2_k)|^2 + \textstyle\sum_{i=1}^d |\sigma^i(k,x^1_k,u^1_k) - \sigma^i(k,x^2_k,u^2_k)|^2\right) \\
&\leq c\mathbb{E}\left(|x^1_k - x^2_k|^2 + |u^1_k - u^2_k|^2\right) = c\left(\|x^1_k - x^2_k\|^2_{X_k} + \|u^1_k - u^2_k\|^2_{L^2_{\mathcal{F}_k}}\right),
\end{aligned}$$

where the last equality follows from (6.3). The Lipschitz continuity of g easily follows. Now, for $\psi = b, \sigma^i$ $(i = 1, \ldots, d)$ we have

$$\mathbb{E}\left(\tfrac{\psi(k,x_k+\tau z_k,u_k+\tau v_k) - \psi(k,x_k,u_k)}{\tau} - \psi_x(k,x_k,u_k)z_k - \psi_u(k,x_k,u_k)v_k\right)^2 \to 0,$$

by the Lipschitz continuity of $\psi(k,\cdot,\cdot)$ and the Lebesgue dominated convergence theorem. The continuity of the linear mapping $(z,v) \to Dg(x,u)(z,v)$ follows easily from (6.4), assumption (6.1), and the isometry (6.3). ∎

As a corollary of the first assertion in the previous lemma, we obtain the following result.

LEMMA 6.3. *For every* $(x,u) \in X$ *and* $\delta \in \Pi_{k=0}^N X_k$ *there exists a unique* $z \in \Pi_{k=0}^N X_k$ *such that* $Dg(x,u)(z,0) = \delta$. *Moreover, there exists* $c > 0$, *independent of* (x,u,z,δ), *such that*

$$\sum_{k=0}^N \|z_k\|_{X_k} \leq c \sum_{k=0}^N \|\delta_k\|_{X_k}. \tag{6.6}$$

In particular, for every closed set $\mathcal{V} \subseteq \Pi_{k=0}^{N-1}\left(L^2_{\mathcal{F}_k}\right)^m$ *we have that*

$$d\left((x,u), g^{-1}(y) \cap \left(\Pi_{k=0}^N X_k \cap \mathcal{V}\right)\right) \leq c \quad \forall\, (x,u) \in X, \;\; y \in \Pi_{k=0}^N X_k. \tag{6.7}$$

Proof. The unique $z \in \Pi_{k=0}^N X_k$ such that $Dg(x,u)(z,0) = \delta$ is given recursively by

$$\begin{aligned}
z_0 &= -\delta_0, \\
z_{k+1} &= b_x(k,x_k,u_k)z_k + \textstyle\sum_{i=1}^d \sigma^i_x(k,x_k,u_k) z_k w^i_{k+1} - \delta_{k+1} \qquad \forall\, k = 0, \ldots, N-1.
\end{aligned}$$

Note that

$$\begin{aligned}
\|z_{k+1}\|^2_{X_{k+1}} = \mathbb{E}\left(|z_{k+1}|^2\right) &\leq (d+2)\left[c_1^2\mathbb{E}\left(|z_k|^2\right) + c_1^2 \textstyle\sum_{i=1}^d \mathbb{E}\left(|z_k|^2 (w^i_{k+1})^2\right) + |\delta_{k+1}|^2\right] \\
&\leq (d+2)^2 c_1^2 \mathbb{E}\left(|z_k|^2\right) + (d+2)\mathbb{E}(|\delta_{k+1}|^2) \\
&\leq \bar{c}\left[\mathbb{E}\left(|z_k|^2\right) + \mathbb{E}(|\delta_{k+1}|^2)\right] \\
&\leq (N+1)\bar{c}^{N+1} \textstyle\sum_{k=0}^N \mathbb{E}\left(|\delta_k|^2\right) \\
&= (N+1)\bar{c}^{N+1} \textstyle\sum_{k=0}^N \|\delta_k\|^2_{X_k},
\end{aligned}$$

where $\bar{c} := (d+2)^2(c_1^2+1) > 1$ and the last equality is a consequence of (6.3). This proves (6.6). Relation (6.7) follows directly from (6.6) and Theorem 4.2. ∎

Let us define the Hamiltonian $H : [0:N-1] \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{n\times d} \times \mathbb{R}^m \to \mathbb{R}$ by

$$H(k,x,p,q,u) := \ell(k,x,u) + p \cdot b(k,x,u) + \sum_{i=1}^{d} q^i \cdot \sigma^i(k,x,u).$$

We have now all the elements to establish the optimality system for problem (SP_d).

THEOREM 6.4. *Suppose that $(\bar{x},\bar{u})$ is a local solution to (SP_d). Then, there exist $p \in \Pi_{k=0}^{N-1}(L^2_{\mathcal{F}_k})^n$, $q \in \Pi_{k=0}^{N-1}(L^2_{\mathcal{F}_k})^{n\times d}$ such that*

$$\begin{array}{rcll} p_{k-1} &=& \mathbb{E}\left(\nabla_x H(k,\bar{x}_k,p_k,q_k,\bar{u}_k)|\mathcal{F}_{k-1}\right) & \forall\, k \in [1:N-1], \\ q^i_{k-1} &=& \mathbb{E}\left(\nabla_x H(k,\bar{x}_k,p_k,q_k,\bar{u}_k)w^i_k|\mathcal{F}_{k-1}\right) & \forall\, k \in [1:N-1],\ i \in [1:d], \\ p_{N-1} &=& \mathbb{E}\left(\nabla\Phi(\bar{x}_N)|\mathcal{F}_{N-1}\right), & \\ q^i_{N-1} &=& \mathbb{E}\left(\nabla\Phi(\bar{x}_N)w^i_N|\mathcal{F}_{N-1}\right) & \forall\, i \in [1:d], \end{array} \tag{6.8}$$

and

$$\mathbb{E}\left(\sum_{k=1}^{N-1} \nabla_u H(k,\bar{x}_k,\bar{p}_k,\bar{q}_k,\bar{u}_k)\cdot v_k\right) \geq 0 \quad \forall\, v \in T_{\mathcal{U}}(\bar{u}). \tag{6.9}$$

If in addition $K(\mathcal{U},\bar{u})$ is convex, then (6.9) holds for all $v \in K(\mathcal{U},\bar{u})$.

Proof. By Lemma 6.2 and Theorem 3.1 there exists $\lambda \in \Pi_{k=0}^{N} X_k$ such that

$$(0,0) \in Df(\bar{x},\bar{u}) + Dg(\bar{x},\bar{u})^*\lambda + \{0\}\times N_{\mathcal{U}}(\bar{u}),$$

from which we deduce that for all $z = (z_0,\dots,z_N) \in \Pi_{k=0}^{N} X_k$

$$\begin{array}{l} D_{x_k} f(\bar{x},\bar{u})z_k + \sum_{j=0}^{N}\langle \lambda_j, D_{x_k} g_j(\bar{x},\bar{u})z_k\rangle_{X_j} = 0 \quad \forall\, k = 0,\dots,N, \\ D_u f(\bar{x},\bar{u})v + \sum_{k=1}^{N}\langle \lambda_k, D_u g_k(\bar{x},\bar{u})v\rangle_{X_k} \geq 0 \quad \forall v \in T_{\mathcal{U}}(\bar{u}). \end{array} \tag{6.10}$$

Lemma 6.2 and the first equation in (6.10) imply that for all $k = 1,\dots,N-1$

$$\begin{array}{c} \mathbb{E}\left(\ell_x(k,\bar{x}_k,\bar{u}_k)z_k\right) + \left\langle \lambda_{k+1}, b_x(k,\bar{x}_k,\bar{u}_k)z_k + \sum_{i=1}^{d} w^i_{k+1}\sigma^i_x(k,\bar{x}_k,\bar{u}_k)z_k \right\rangle_{X_{k+1}} \\ = \langle \lambda_k, z_k\rangle_{X_k}, \\ \mathbb{E}\left(\Phi_x(\bar{x}_N)z_N\right) = \langle \lambda_N, z_N\rangle_{X_N}. \end{array} \tag{6.11}$$

Setting

$$z_k = y^0_{k-1} + \sum_{i=1}^{d} y^i_{k-1}w^i_k \in X_k, \quad \lambda_k = p_{k-1} + \sum_{i=1}^{d} q^i_{k-1}w^i_k \in X_k,$$

relation (6.11) yields

$$\begin{array}{c} \mathbb{E}\left(\nabla_x H(k+1,\bar{x}_{k+1},p_{k+1},q_{k+1},\bar{u}_{k+1})\cdot z_k\right) = \mathbb{E}\left(p_{k-1}\cdot y^0_{k-1} + \sum_{i=1}^{d} q^i_{k-1}\cdot y^i_{k-1}\right), \\ \mathbb{E}\left(\nabla\Phi(\bar{x}_N)\cdot z_N\right) = \mathbb{E}\left(p_{N-1}\cdot y^0_{N-1} + \sum_{i=1}^{d} q^i_{N-1}\cdot y^i_{N-1}\right). \end{array} \tag{6.12}$$

Taking $y_{k-1}^i = 0$ for all $i \in [1:d]$ the first equation in (6.12) gives

$$\mathbb{E}\left(\nabla_x H(k+1, \bar{x}_{k+1}, p_{k+1}, q_{k+1}, \bar{u}_{k+1}) \cdot y_{k-1}^0\right) = \mathbb{E}\left(p_{k-1} \cdot y_{k-1}^0\right),$$

and so, since $y_{k-1}^0 \in L^2_{\mathcal{F}_{k-1}}$ is arbitrary, by definition of conditional expectation w.r.t. $\mathcal{F}_{k-1}$, the first equality in (6.8) follows. Similarly, fixing $\bar{i} \in [1:d]$ and letting $y_{k-1}^i = 0$ for all $i \in [0:d] \setminus \{\bar{j}\}$, we obtain the second relation (6.8) for $i = \bar{i}$. The last two relations in (6.8) follow by an analogous argument.

Finally, since for all $k = 0, \ldots, N-1$,

$$\langle \lambda_{k+1}, D_u g_{k+1}(\bar{x}, \bar{u})v \rangle_{X_{k+1}} = \mathbb{E}\left(p_k \cdot b_u(k, \bar{x}_k, \bar{u}_k)v_k + \sum_{i=1}^d q_k^i \cdot \sigma_u^i(k, \bar{x}_k, \bar{u}_k)v_k\right),$$

relation (6.9) follows directly from the second relation in (6.10) and Lemma 6.2(ii). If $K(\mathcal{U}, \bar{u})$ is convex, then Theorem 3.1(ii) ensures that the second relation in (6.10) holds for all $v \in K(\mathcal{U}, \bar{u})$, from which the last assertion of the theorem easily follows. $\square$

Remark 6.5. (i) The optimality system (6.8)–(6.9) was first shown in [35] under more restrictive assumptions on ℓ, Φ, f, σ (see Assumption 1 in [35]) and the control constraint set $\mathcal{U}$ (see [35, section 3]). The results in sections 3 and 4 allow us to prove a more general result in a quite direct manner.

(ii) Similarly to the continuous case (see section 5), it is easy to extend the results in this section to the case where the initial state $\hat{x}_0$ is a decision variable subject to the constraint $\hat{x}_0 \in \mathcal{X}_0$, where $\mathcal{X}_0$ is a closed subset of $\mathbb{R}^n$. In this case, the optimality system is as in Theorem 6.4 with the additional constraint on the adjoint state (called the transversality condition) $-p_0 \in N_{\mathcal{X}_0}(\hat{x}_0)$.

(iii) Note that if $\{w_1, \ldots, w_N\}$ corresponds to a sequence of normalized increments of a d-dimensional Brownian motion on a time grid in $[0,T]$ ($T > 0$), then, by suitably redefining b and σ, problem (SP_d) can be seen as an Euler discretization of (SP). It is well known that, under general assumptions, the optimal cost of (SP_d) converges to the optimal cost of (SP), provided that the maximum time step of the grid tends to zero (see, e.g., [4]). Therefore, the analysis of the optimality system in Theorem 6.4 can be useful even if one is interested in solving (SP).

REFERENCES

[1] J. P. Aubin and H. Frankowska, *Set-Valued Analysis*, Systems Control Found. Appl. 2, Birkhäuser Boston, Boston, 1990.

[2] J. Backhoff and F. J. Silva, *Sensitivity results in stochastic optimal control: A Lagrangian perspective*, ESAIM Control Optim. Calc. Var., 23 (2017), pp. 39–70.

[3] A. Bensoussan, *Lectures on stochastic control*, in Nonlinear Filtering and Stochastic Control (Cortona, 1981), Lecture Notes in Math. 972, Springer, Berlin, New York, 1982, pp. 1–62.

[4] J. F. Bonnans, J. Gianatti, and F. J. Silva, *On the time discretization of stochastic optimal control problems: The dynamic programming approach*, ESAIM Control Optim. Calc. Var., 25 (2019), 63.

[5] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Ser. Oper. Res., Springer-Verlag, New York, 2000.

[6] J. M. Borwein, *Epi-Lipschitz-like sets in Banach spaces: Theorems and examples*, Nonlinear Anal., 11 (1987), pp. 1207–1217.

[7] J. M. Borwein and H. M. Strojwas, *Tangential approximations*, Nonlinear Anal., 9 (1985), pp. 1347–1366.

[8] M. Brokate, *A regularity condition for optimization in Banach spaces: Counterexamples*, Appl. Math. Optim., 6 (1980), pp. 189–192.

[9] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983; Classics Appl. Math. 5, SIAM, Philadelphia, 1990, https://doi.org/10.1137/1.9781611971309.

[10] F. H. Clarke and P. D. Loewen, *The value function in optimal control: Sensitivity, controllability, and time-optimality*, SIAM J. Control Optim., 24 (1986), pp. 243–263, https://doi.org/10.1137/0324014.

[11] R. Cominetti, *Metric regularity, tangent sets, and second-order optimality conditions*, Appl. Math. Optim., 21 (1990), pp. 265–287.

[12] A. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer Monogr. Math., Springer, Dordrecht, The Netherlands, 2009.

[13] I. Ekeland, *On the variational principle*, J. Math. Anal. Appl., 47 (1974), pp. 324–353.

[14] I. Ekeland, *An inverse function theorem in Fréchet spaces*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 28 (2011), pp. 91–105.

[15] B. El Abdouni and L. Thibault, *Lagrange multipliers for Pareto non-smooth programming problems in Banach spaces*, Optimization, 26 (1992), pp. 277–285.

[16] H. Gfrerer, *First order and second order characterizations of metric subregularity and calmness of constraint set mappings*, SIAM J. Optim., 21 (2011), pp. 1439–1474, https://doi.org/10.1137/100813415.

[17] H. Gfrerer, *On directional metric regularity, subregularity and optimality conditions for non-smooth mathematical programming*, Set-Valued Var. Anal., 21 (2013), pp. 151–176.

[18] H. Gfrerer, *On directional metric subregularity and second-order optimality conditions for a class of nonsmooth mathematical programs*, SIAM J. Optim., 23 (2013), pp. 632–665, https://doi.org/10.1137/120891216.

[19] H. Gfrerer, *On metric pseudo-(sub) regularity of multifunctions and optimality conditions for degenerated mathematical programming*, Set-Valued Var. Anal., 22 (2014), pp. 79–115.

[20] B. M. Glover and B. D. Craven, *A Fritz John optimality condition using the approximate subdifferential*, J. Optim. Theory Appl., 82 (1994), pp. 253–265.

[21] R. Henrion and A. Jourani, *Subdifferential conditions for calmness of convex constraints*, SIAM J. Optim., 13 (2002), pp. 520–534, https://doi.org/10.1137/S1052623401386071.

[22] R. Henrion, A. Jourani, and J. Outrata, *On the calmness of a class of multifunctions*, SIAM J. Optim., 13 (2002), pp. 603–618, https://doi.org/10.1137/S1052623401395553.

[23] R. Henrion and J. Outrata, *A subdifferential criterion for calmness of multifunctions*, J. Math. Anal. Appl., 258 (2001), pp. 110–130.

[24] A. D. Ioffe, *Approximate subdifferentials and applications* 2: *Functions on locally convex spaces*, Mathematika, 33 (1986), pp. 111–128.

[25] A. D. Ioffe, *Approximate subdifferentials and applications* 3: *Metric theory*, Mathematika, 36 (1989), pp. 1–38.

[26] A. D. Ioffe, *Codirectional compactness, metric regularity and subdifferential calculus*, in Constructive, Experimental, and Nonlinear Analysis, M. Théra, ed., CRC Math. Model. Ser. 27, CRC, Boca Raton, FL, 2000.

[27] A. D. Ioffe, *Variational Analysis of Regular Mappings*, Springer Monogr. Math., Springer, Cham, 2017.

[28] Y. Ishizuka, *Optimality conditions for directionally differentiable multiobjective programming problems*, J. Optim. Theory Appl., 72 (1992), pp. 91–112.

[29] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, Adv. Des. Control 15, SIAM, Philadelphia, 2008, https://doi.org/10.1137/1.9780898718614.

[30] A. Jourani, *Qualification conditions for multivalued functions in Banach spaces with applications to nonsmooth vector optimization problems*, Math. Program., 66 (1994), pp. 1–23.

[31] A. Jourani and L. Thibault, *Approximations and metric regularity in mathematical programming in Banach spaces*, Math. Oper. Res., 18 (1993), pp. 390–401.

[32] A. Jourani and L. Thibault, *Metric regularity for strongly compactly Lipschitzian mappings*, Nonlinear Anal., 24 (1995), pp. 229–240.

[33] A. Jourani and L. Thibault, *Verifiable conditions for openness and metric regularity in Banach spaces*, Trans. Amer. Math. Soc., 347 (1995), pp. 1255–1268.

[34] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer-Verlag, New York, 1991.

[35] X. Lin and W. Zhang, *A maximum principle for optimal control of discrete-time stochastic systems with multiplicative noise*, IEEE Trans. Automat. Control, 60 (2015), pp. 1121–1126.

[36] P. Loewen, *Parameter sensitivity in stochastic optimal control*, Stochastics, 22 (1987), pp. 1–40.

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

[38] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation.* I, Grundlehren Math. Wiss. 330, Springer-Verlag, Berlin, 2006.

[39] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation.* II, Grundlehren Math. Wiss. 331, Springer-Verlag, Berlin, 2006.

[40] L. Mou and J. Yong, *A variational formula for stochastic controls and some applications*, Pure Appl. Math. Q., 3 (2, Special Issue: In honor of Leon Simon. Part 1) (2007), pp. 539–567.

[41] S. Peng, *A general stochastic maximum principle for optimal control problems*, SIAM J. Control Optim., 28 (1990), pp. 966–979, https://doi.org/10.1137/0328054.

[42] P. Protter, *Stochastic Integration and Differential Equations*, 2nd ed., Springer-Verlag, Heidelberg, 2005.

[43] H. Rådström, *An embedding theorem for spaces of convex sets*, Proc. Amer. Math. Soc., 3 (1952), pp. 165–169.

[44] S. Robinson, *Regularity and stability for convex multivalued functions*, Math. Oper. Res., 1 (1976), pp. 130–143.

[45] J. J. Ye and X. Y. Ye, *Necessary optimality conditions for optimization problems with variational inequality constraints*, Math. Oper. Res., 22 (1997), pp. 977–997.

[46] J. Yong and X. Y. Zhou, *Stochastic Controls, Hamiltonian Systems and HJB Equations*, Springer-Verlag, New York, Berlin, 2000.

[47] P. L. Yu, *Multicriteria Decision Making: Concepts, Techniques and Extensions*, Plenum Press, New York, 1985.