Metric regularity under Gâteaux differentiability with applications to optimization and stochastic optimal control problems

A. Jourani∗  F. J. Silva†

October 30, 2018

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 recover in a direct manner the optimality systems associated to two types of standard stochastic optimal control problems.

Keywords: Lagrange multipliers, Gâteaux differentiability, calmness, metric regularity, optimality conditions, stochastic optimal control problems.

1 Introduction

Consider the following optimization problem

\[ \min \{ f(x) : g(x) \in D \}, \]

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 are \( \lambda \geq 0 \) and \( y^* \in N(D, g(x_0)) \) such that

\[ (\lambda, y^*) \neq (0, 0), \]

\[ \lambda f'(x_0) + y^* \circ g'(x_0) = 0. \]

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 [7, 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 \( \ell^2 \) and consider the operator \( A : \ell^2 \to \ell^2 \) defined by

\[ A \left( \sum_{i \geq 1} x_i e_i \right) = \sum_{i \geq 1} 2^{1-i} x_i e_i. \]
It is easy to check that $A$ is injective but not surjective and that the image of $A$, denoted by $\text{Im}(A)$, is a proper dense subspace of $\ell^2$. As a consequence, the adjoint operator $A^*$ is injective but not surjective. Now, let $x^* \in \ell^2 \setminus \text{Im}(A^*)$ and consider $f = x^*$, $g = A$ and $D = \{0\}$ as the data for problem (1.1). Then 0 is the only feasible point and, hence, the solution of this problem. Moreover, since $x^* \in \ell^2 \setminus \text{Im}(A^*)$, we easily check that there is no $(\lambda, y^*) \neq (0, 0)$ satisfying (1.4).

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 $\{0\} \subset \mathbb{R}^n$ (see [12, 13, 15, 29, 31] and references therein). The first result which gives a condition for the validity of (1.2) in the case where $D$ is closed is due to Jourani and Thibault [21], where it is assumed that the system $g(x) \in D$ is metrically regular (see [5, 7] 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 $\lambda$ can be taken different from zero. In [19] it is shown that relations (1.2) and (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 [22, 23], 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 [21]). Denote by $\partial_A d(u, D)$ the approximate subdifferential of $d(\cdot, D)$ at $u$ (see [14, 15]) and assume the existence of a locally compact cone $K^* \subset Y^*$ and a neighbourhood $V$ of $g(x_0)$ such that

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

or equivalently (see [16]), $D$ is compactly epi-Lipschitzian in the sense of Borwein-Strojwas [4]. Then there exists $\lambda \geq 0$ and $y^* \in \mathbb{R}_+ \partial_A d(g(x_0), D)$, with $(\lambda, y^*) \neq (0, 0)$, such that

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

Now, 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 the classical ones as Slater condition, Mangasarian-Fromovitz condition and so on. 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 Theorem 3.1 and in Theorem 3.2 under the so-called calmness condition which is a kind of constraint qualification. Inspired by the work by Ekeland [11], our main results (Theorem 4.1 and Theorem 4.2) establish the metric regularity property for the constraint system under Gâteaux differentiability assumptions only. We point out that the proofs of these results do not rely on any iteration scheme.

The first application of our results are first order necessary optimality conditions for stochastic optimal control problems in continuous time. Following the functional framework proposed by Backhoff and Silva in [1], our abstract results allow us to recover a weak version of the general stochastic Pontryagin’s maximum principle, proved in [29], under rather general assumptions (see our Remark 5.1 (ii)). As pointed out in [1], the main difficulty in deriving this 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. Our abstract results, which assume only Gâteaux differentiability of the mapping that defines the constraints and a uniform surjectivity property of the Gâteaux derivative in a neighbourhood of the optimal solution, allow us to avoid this issue and to establish the existence of Lagrange multipliers and, as a consequence of the characterization of these multipliers studied in [1], the validity of a weak version of Pontryagin’s principle. We point that this result is not new and is weaker than the one proved in [29], which, however, needs strong assumptions on the second order derivatives of the data. On the other hand, the proof presented here is new, short and clarifies the role of the adjoint states as Lagrange multipliers when the stochastic control problem is formulated in the correct functional framework.

In the second application, we consider a discrete time stochastic optimal control problem where the randomness is modelled by a multiplicative independent noise. As in the
continuous time case, the main difficulty to apply standard abstract Lagrange multiplier results comes from the functional equation defining the controlled trajectory. By considering 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 [25] under more general assumptions than those imposed in that article.

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 (which is a stronger property than its calmness).

Finally, in Sections 5 and 6 we apply these abstract results to the stochastic control problems described in the previous paragraphs.

## 2 Notations and preliminaries

In all 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 \( h \in X \) we set \( \langle x^*, h \rangle_X := x^*(h) \), with an analogous notation for the duality paring between \( Y^* \) and \( Y \). Given \( r > 0 \) and \( x \in X \) we denote \( B_X(x, r) := \{ x' \in X : \| x' - x \|_X \leq r \} \) the closed ball of radius \( r \) centered at \( x \). For \( A \subseteq X \) we denote by \( \text{cl}(A) \) and \( \text{int}(A) \) its closure and its topological interior, respectively.

Let us recall some basic notions in nonsmooth analysis (see e.g. \[8\] [27] for a detailed account of the theory). Given a locally Lipschitz function \( \varphi : X \to \mathbb{R} \), the directional derivative \( \varphi^0(x; h) \) of \( \varphi \) at \( x \) in the direction \( h \in X \) and the subdifferential \( \partial_C \varphi(x) \) of \( \varphi \) at \( x \) are both defined in sense of Clarke as

\[
\varphi^0(x; h) := \limsup_{y \to x, \tau \downarrow 0} \frac{\varphi(y + \tau h) - \varphi(y)}{\tau},
\]

\[
\partial_C \varphi(x) := \{ x^* \in X^* : \langle x^*, h \rangle_X \leq \varphi^0(x; h) \ \forall \ h \in X \}.
\]

Note that for all \( x \in X \), \( \varphi^0(x; \cdot) : X \to \mathbb{R} \) is well-defined, positively homogeneous, subadditive, Lipschitz continuous and satisfies that \( \varphi^0(x; 0) = 0 \). This implies that \( \varphi^0(x; \cdot) \) is the support function of \( \partial_C \varphi(x) \), which is a nonempty, weak* compact and convex set (see [8] Proposition 2.1.2). Given a nonempty set \( A \subseteq X \), we denote by \( d_A(\cdot) := \inf_{x \in A} \| \langle \cdot \rangle - x \| \) the distance to \( A \) function. Given \( x \in \text{cl}(A) \), the Clarke’s tangent cone is defined as

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

If \( x \notin \text{cl}(A) \) we set \( T_A(x) := \emptyset \). If \( x \in \text{cl}(A) \), we have that \( h \in T_A(x) \) iff for every sequences \( (x_n) \) 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 \( K \) we denote by \( K^0 \) its negative polar cone, defined as

\[
K^0 := \{ x^* \in X^* : \langle x^*, h \rangle_X \leq 0 \ \forall \ h \in K \}.
\]

We have (see e.g. \[8\] Proposition 2.4.2)

\[
N_A(x) = w^*-\text{cl} \left( \bigcup_{\lambda \geq 0} \lambda \partial_C d_A(x) \right),
\]

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

\[
\tau(A, x) = \left\{ h \in X : \lim_{\tau \to 0^+} \frac{d_A(x + \tau h)}{\tau} = 0 \right\}.
\]
We set $\mathcal{T}(A,x) := \emptyset$ if $x \notin \text{cl}(A)$. By definition, if $x \in \text{cl}(A)$ 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 $x \in \text{cl}(A)$ is defined as

$$K(A,x) := \{h \in X : d^-_A(x;h) = 0\},$$

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 $x \notin \text{cl}(A)$. By definition, if $x \in \text{cl}(A)$ 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

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

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

**Lemma 2.1** Let $A \subset X$ and $B \subset Y$ 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) $K(A \times B, (x_0,y_0)) \subset K(A,x_0) \times K(B,y_0)$. 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 $k \in Y$, $d^-_{A \times B}((x_0,y_0),(h,k)) \leq d^-_A(x_0,h) + d^-_B(y_0,k)$.

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

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

### 3 Lagrange multipliers for optimization problems under Gâteaux differentiability assumptions on the data

This section is concerned with necessary optimality conditions or existence of Lagrange multipliers associated to local solutions of optimization problems of the form

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

where $f : X \mapsto \mathbb{R} \cup \{+\infty\}$ is function, $g : X \mapsto Y$ 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 $x_0$.

(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

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

(H$_g$) $g$ is Gâteaux differentiable at $x_0$.\pagebreak
If $(H_g)$ holds true, we will denote by $Dg(x_0) : X \to Y$ the Gâteaux derivative and by $D^*g(x_0) : Y^* \to X^*$ its adjoint operator. Similar notations will be used for the Gâteaux derivative of $f$ if $(H_f)$ holds.

We recall that system (3.2) and the weak differentiability assumptions $(H_f)$ hold and that system (3.3) as in $(H_f)$ and (3.3), respectively. Then,

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

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

In particular, there exists $y^* \in Y^*$, with $\|y^*\|_{Y^*} \leq K_f a$, 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 $\|y^*\|_{Y^*} \leq K_f a$, 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 $(H_f)$, by [8, Proposition 2.4.3] we have that $x_0$ is a local minimum of

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

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

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

(3.5)

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

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

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 $h \mapsto Df(x_0)h + \|Dg(x_0)h\|_Y$ 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 [8, Proposition 2.4.3] again, we obtain the existence of \( s > 0 \) such that

\[
    f(x) + Kf a\|g(x)\|_Y + K_f(1 + K_g a)dc(x) \geq f(x_0) \quad \forall x \in B_X(x_0, s).
\]

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

\[
    Df(x_0)h + Kf a\|Dg(x_0)h\|_Y + K_f(1 + K_g a)d_C(x_0; h) \geq 0.
\]  (3.6)

Noting that \( d_C(x_0; h) \leq d_C^g(x_0; h) \), we obtain

\[
    Df(x_0)h + Kf a\|Dg(x_0)h\|_Y + K_f(1 + K_g a)d_C(x_0; h) \geq 0 \quad \forall h \in X,
\]

or equivalently the convex function

\[
    h \in X \mapsto Df(x_0)h + Kf a\|Dg(x_0)h\|_Y + K_f(1 + K_g a)d_C(x_0, h)
\]

attains its minimum at \( h = 0 \). Using that \( \partial_c d_C^g(x_0, \cdot)(0) = \partial_c dc(x_0) \) and (2.1), the (convex) subdifferential calculus produces a \( y^* \in Y^* \), with \( \|y^*\|_Y \leq K_f a \), such that

\[
    -Df(x_0) - D^*g(x_0)y^* \in K_f(1 + K_g a)\partial d_C(x_0) \subset N_C(x_0).
\]

So that assertion (ii) follows. Finally, inequality (3.6) yields

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

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

\[
    \begin{aligned}
    \min & \quad f(x) \\
    \text{s.t.} & \quad g(x) \in D, \quad x \in C,
    \end{aligned}
\]  (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

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

Problem (3.7) can be rephrased as follows

\[
    \begin{aligned}
    \min & \quad \tilde{f}(x, y) \\
    \text{s.t.} & \quad \tilde{g}(x, y) = 0, \quad (x, y) \in C \times D,
    \end{aligned}
\]  (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 (3.10). In the following result, we transfer the calmness property of system (3.8) to that of system

\[
    \begin{aligned}
    \text{Find} & \quad (x, y) \in C \times D, \quad \tilde{g}(x, y) = 0,
    \end{aligned}
\]  (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 3.1 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 \( \text{(3.8)} \) is calm at \( x_0 \in g^{-1}(D) \cap C \).
(ii) The system \( \text{(3.11)} \) 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 \( \text{(3.8)} \) 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

\[
d_{g^{-1}(D) \cap C}(x) \leq ad_D(g(x)) \quad \forall x \in B_X(x_0, 3s) \cap C,
\]

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 \( u \in B(x_0, 3s) \cap (g^{-1}(D) \cap C) \). Thus,

\[
\|g(x) - g(u)\| \leq K_g \|x - u\|.
\]

We have

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

and using the triangle inequality, we get

\[
\|y - g(u)\| + \|x - u\| \leq \|y - g(x)\| + \|g(x) - g(u)\| + \|x - u\|
\]

\[
\leq \|y - g(x)\| + (1 + K_g)\|x - u\|
\]

\[
\leq \|y - g(x)\| + (1 + K_g)d_{g^{-1}(D) \cap C}(x) + t(1 + K_g)
\]

\[
\leq \|y - g(x)\| + (1 + K_g)a\|y - g(x)\| + t(1 + K_g)
\]

\[
\leq (1 + a(1 + K_g))\|y - g(x)\| + t(1 + K_g)
\]

\[
= (1 + a(1 + K_g))\|\tilde{g}(x, y)\| + t(1 + K_g).
\]

As \( t \) is arbitrary, relation \( \text{(3.12)} \) yields

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

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

\[
d_{g^{-1}(0) \cap (C \times D)}(x, y) \geq d_{g^{-1}(D) \cap C}(x).
\]

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

Theorem 3.2 Let \( x_0 \) be a local solution to problem \( \text{(3.7)} \) and suppose that system \( \text{(3.8)} \) is calm at \( x_0 \). Suppose that \((H_f)\) and \((H_g)\) hold and that \( g \) is locally Lipschitz around \( x_0 \). Then

(i) There exists \( y^* \in N_D(g(x_0)) \), with \( \|y^*\|_Y \leq K_f(1 + a(1 + K_g)) \) (where \( K_f, K_g \) and \( a \) are as in \((H_f),(H_g)\) and \( \text{(3.9)} \), 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 \( \|y^*\|_Y \leq K_f(1 + a(1 + K_g)) \) and

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

**Proof.** Since \( x_0 \) solves (3.10) locally, \((x_0, g(x_0))\) is a local solution to problem (3.10). Using that the constant \( a \) satisfies (4.1), the proof of Lemma 3.1 shows that the calmness constant associated to system (3.10) 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

\[
\sup_{y \in A} \sup_{v \in B_Y} \inf_{u \in B_X} \|v\|_Y \leq \alpha.
\]

The main result of our article is the following.

**4 Metric regularity under Gâteaux differentiability**

In this section, we first provide a sufficient condition for a stronger property than the calmness of system (3.2), namely its metric regularity (see [9, 17] 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)). Throughout this section, we fix a point \( x_0 \) and all \( r > 0 \).

Let \( y \in A \) and \( r > 0 \), we set \( B_A(y, r) := B_y(y, r) \). Throughout this section, we fix a point \( x_0 \) and all \( r > 0 \).

**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 right-inverses exist because (4.1) implies that \( Dg(x) \) is surjective). Then, assumption (4.1) can be rephrased in terms of \( G \) as follows

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

The main result of our article is the following.

**Theorem 4.1** Suppose that (Hg) and (H_eq) hold 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

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

we have

\[
d_{g^{-1}(y) \cap C}(x) \leq \alpha\|g(x) - y\|_Y.
\]

**Proof.** The proof is inspired from [11]. 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

\[
h(u) := \|g(u) - y\|_Y.
\]
Let \( \beta > \alpha \) be such that \( 0 < h(x) = \|g(x) - y\|_{Y} < \frac{\beta}{\beta + \alpha} \). 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 [10, Theorem 1.1]) gives the existence of \( \bar{u} \in B_{C}(x_0, r) \) such that
\[
h(\bar{u}) \leq h(x), \tag{4.3}
\]
\[
\|\bar{u} - x\|_{X} \leq \beta h(x), \tag{4.4}
\]
\[
h(\bar{u}) \leq h(u) + \frac{1}{\beta} \|\bar{u} - u\|_{X} \quad \forall u \in B_{C}(x_0, r). \tag{4.5}
\]
Inequality (4.3) and the choice of \( x \) and \( \beta \) imply that
\[
\|\bar{u} - x\|_{X} < r_{2} \quad \text{and so} \quad \|\bar{u} - x\|_{X} \leq \|\bar{u} - x\|_{X} + \|x - x_{0}\|_{X} < r_{2} + r_{1} = r. \tag{4.6}
\]
Claim: we have that \( y = g(\bar{u}) \). Let us assume for a moment that the claim is true. By (4.2), we obtain
\[
d_{g^{-1}(y) \cap C}(x) \leq \beta \|g(x) - y\|_{Y},
\]
and, as \( \beta > \alpha \) is arbitrary, we get that (4.2) holds true.
It remains to prove the claim. Suppose the contrary and define
\[
w = \frac{y - g(\bar{u})}{\|y - g(\bar{u})\|_{Y}}.
\]
Since \( \bar{u} \in B_{C}(x_0, r) \), assumption \((H_{eq})\) implies the existence of \( v \in B_{K(C, \alpha)}(0, \alpha) \) such that
\[
w = Dg(\bar{u}) v.
\]
Since \( v \in B_{K(C, \alpha)}(0, \alpha) \), there exist sequences \( \tau_{n} \to 0^{+} \) and \( v_{n} \to v \) such that
\[
u_{n} := \bar{u} + \tau_{n} v_{n} \in C \quad \text{for} \quad n \quad \text{sufficiently large}.
\]
We may write \( u_{n} = \bar{u} + \tau_{n} v + o(\tau_{n}) \in C \), where \( \lim_{n \to +\infty} \frac{o(\tau_{n})}{\tau_{n}} = 0 \). Note that the second inequality in (4.2) implies that \( u_{n} \in B_{C}(x_0, r) \) for \( n \) sufficiently large. Now, using inequality (4.5), we get
\[
h(\bar{u}) \leq h(u_{n}) + \frac{1}{\beta} \|\tau_{n} v + o(\tau_{n})\|_{X}. \tag{4.7}
\]
On the other hand, since \( g \) is Gâteaux differentiable at \( \bar{u} \), we have
\[
g(u_{n}) = g(\bar{u}) + \tau_{n} Dg(\bar{u}) v + \tau_{n} \varepsilon(\tau_{n}), \quad \text{where} \quad \lim_{n \to +\infty} \varepsilon(\tau_{n}) = 0,
\]
which, combined with (4.4), ensures that
\[
\left\| g(\bar{u}) - y + \tau_{n} Dg(\bar{u}) v + \tau_{n} \varepsilon(\tau_{n}) \right\|_{Y} - \left\| g(\bar{u}) - y \right\|_{Y} \geq -\frac{1}{\beta} \left\| v + o(\tau_{n}) \right\|_{X}.
\]
Since
\[
\lim_{n \to +\infty} \left\| g(\bar{u}) - y + \tau_{n} Dg(\bar{u}) v \right\|_{Y} - \left\| g(\bar{u}) - y \right\|_{Y} = \max_{y^{*} \in \partial\|y - g(\bar{u})\|} \langle g^{*}, Dg(\bar{u}) v \rangle_{Y},
\]
we get the existence of \( y_{v}^{*} \in \partial\|y - g(\bar{u})\| \), such that
\[
-1 = \langle y_{v}^{*}, w \rangle_{Y} = \langle y_{v}^{*}, Dg(\bar{u}) v \rangle_{Y} \geq -\frac{1}{\beta} \|v\|_{X} \geq -\frac{\alpha}{\beta}. \tag{4.8}
\]
where the first equality follows from the fact that we are assuming that \( g(\bar{u}) \neq y \) and the standard relation
\[
y^*_0 \in \partial \cdot ||y (g(\bar{u}) - y)\| \implies \|y^*_0\|_Y = 1 \quad \text{and} \quad \langle y^*_0, g(\bar{u}) - y \rangle_Y = \|g(\bar{u}) - y\|_Y.
\]
Since (4.8) contradicts \( \alpha < \beta \), the claim follows.  

The previous result extends the following inverse function theorem result, proved first in [11 Theorem 2] in the case \( C = X \).

**Corollary 4.1** Suppose that the assumptions of Theorem 4.1 are satisfied. Then,
\[
d_{g^{-1}(y) \cap C}(x_0) \leq \varepsilon \|y\|_Y \quad \forall \ y \in Y, \text{with} \ \|y\|_Y < \frac{r}{\alpha}. \quad (4.9)
\]
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
\[
\|x - x_0\|_X < r, \quad \|x - x_0\|_X \leq \beta \|y\|_Y. \quad (4.10)
\]

**Proof.** By Theorem 4.1 in order to prove (4.9) it suffices to choose \( \varepsilon > 0 \) such that \( (x_0, y) \in D_{\varepsilon - \varepsilon', C} \), which is possible because of the strict inequality in (4.9). It remains to prove that (4.10) holds for \( \beta > \alpha \) and \( \|y\|_Y < r/\alpha \). In this case, the first inequality in (4.9) becomes strict and we get the existence of \( x_0 \in g^{-1}(y) \cap C \) such that the second inequality in (4.10) holds true.  

Since there exists \( \varepsilon > 0 \) such that \( \|y\|_Y \leq (r - \varepsilon)/\alpha \) then the first inequality in (4.10) holds for \( x_0 \) provided that \( \alpha < \beta < \alpha r/(r - \varepsilon) \). If \( \beta \geq \alpha r/(r - \varepsilon) \) then (4.10) holds for \( x_0 \) with \( \beta' = \alpha r/(r - \varepsilon) \) and so \( \|x_0 - y\|_Y \leq \beta' \|y\|_Y \). The result follows.

Now, we study the corresponding metric regularity property for system (3.5). We consider the following qualification condition:

\[ (H_{cq}^\prime) \quad \text{there exist} \ \alpha_1, \alpha_2 > 0 \quad \text{and} \ \ r > 0 \quad \text{such that} \ \text{g is continuous and Gâteaux differentiable on} \ B_C(x_0, r) \ \text{and} \]
\[
B_{Y}(0, 1) \subset Dg(x)\{B_{K(C,x)}(0, \alpha_1)\} - B_{K(D,y)}(0, \alpha_2) \quad \forall (x, y) \in B_{C \times D}(x_0, g(x_0)), r). \quad (4.11)
\]

**Theorem 4.2** Suppose that (H), (H_q) and (H_{cq}^\prime) 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
\[
d_{g^{-1}(D+y)\cap C}(x) \leq \alpha d_{B_D(g(x_0), r_1)}(g(x) - y). \quad (4.12)
\]

**Proof.** Using that at least one of the sets \( C \) and \( D \) is convex, for all \( (x', y') \in C \times D \) we have
\[
B_{K(C,x')}((0, 0), \alpha_1) \times B_{K(D,y')}((0, 0), \alpha_2) \subseteq B_{K(C \times D, (x', y'))}((0, 0), \alpha).
\]
Therefore, defining \( \bar{g} : X \times Y \to Y \) as \( \bar{g}(x, z) := g(x) - z \), condition (4.11) implies that
\[
B_{Y}(0, 1) \subseteq D\bar{g}(x', y')\left[B_{K(C \times D, (x', y'))}((0, 0), \alpha)\right] \quad \forall (x', y') \in B_{C \times D}(x_0, g(x_0)), r). \quad (4.12)
\]
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
\[
\|g(x) - y - z_{\varepsilon}\|_Y \leq d_{B_D(g(x_0), r_1)}(g(x) - y) + \varepsilon < \frac{r_2}{\alpha}. \quad (4.13)
\]
By (4.12), we can apply Theorem 4.1 to \(\tilde{g}\) and deduce that
\[
d_{\tilde{g}^{-1}(y) \cap (C \times D)}(x, z_c) \leq \alpha\|g(x) - z_c - y\| \leq \alpha d_{B_D(g(x_0), r_1)}(g(x) - y) + \alpha \varepsilon. \tag{4.14}
\]
Finally, since \((x', z') \in \tilde{g}^{-1}(y) \cap (C \times D)\), if \(x' \in C, z' \in D\) and \(g(x') - y = z'\), we get that
\[
d_{\tilde{g}^{-1}(D+y) \cap C}(x) \leq d_{\tilde{g}^{-1}(y) \cap (C \times D)}(x, z_c). \tag{4.15}
\]
Since \(\varepsilon\) is arbitrary, the result follows from (4.14)-4.15 \(\blacksquare\)

We can ask if we can replace the assumption \((H')\) by the following one

\((H'_{cq})\) 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
\[
B_{Y}(0, 1) \subset Dg(x)(B_{K(C, x)}(0, \alpha_1)) - B_{K(D, g(x))}(0, \alpha_2) \quad \forall \ x \in B_{g^{-1}(D) \cap C}(x_0, r). \tag{4.16}
\]

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
\[
D = \{(x, y) \in \mathbb{R}^2 : |y - x| or \ |x - 0, x^2 + (y + 2)^2 = 4\},
\]
(see Figure 1) and take \(g\) be the identity function in \(\mathbb{R}^2\). Then \(C \cap D = \emptyset\), \(g^{-1}(C \cap D) = \emptyset\), \(K(C, (0, 0)) = \mathbb{R}_+ \times \{0\}\) and \(K(D, (0, 0)) = \{(x, x) : x \in \mathbb{R}\} \cup (\mathbb{R}_+ \times \{0\})\). Thus,
\[
B_{R^2}(0, 1) \subset B_{K(C, (0, 0))}(0, 2) - B_{K(D, (0, 0))}(0, 2).
\]
Similarly, we have that (4.16) holds true and it is easy to check that (4.11) does not hold.

We will show that there is no \(a > 0\) such that
\[
d_{g^{-1}(C \cap D)}(u) \leq ad(g(u), D) \quad \text{for} \ u \in C \text{ near } 0.
\]
Indeed, for \(x > 0\) and \(x^2 + (y + 1)^2 = 1\), with \((x, y)\) near \((0, 0)\), we have
\[
d_{g^{-1}(C \cap D)}(x, y) = \sqrt{x^2 + y^2} \quad \text{and} \quad d(g(x, y), D) \leq 2 - \sqrt{4 - (x^2 + y^2)}
\]
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 near to \((0, 0)\).

## 5 Application to stochastic optimal control in continuous time

Let \(T > 0\) and consider a filtered probability space \((\Omega, \mathcal{F}, \mathbb{P})\), on which a \(d\)-dimensional \((d \in \mathbb{N}^*)\) Brownian motion \(W(t)\) is defined. We suppose that \(\mathcal{F} = \{\mathcal{F}_t\}_{0 \leq t \leq T}\) is the natural filtration, augmented by all \(\mathbb{P}\)-null sets in \(\mathcal{F}\), associated to \(W(t)\). The filtration \(\mathcal{F}\) is right-continuous, i.e. \(\mathcal{F}_t = \cap_{s \leq t} \mathcal{F}_s\) (see [20] 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
\[
(L^2_{W_{\mathbb{P}}})^n := \{v \in L^2(\Omega; L^2([0, T]; \mathbb{R}^n)) ; (\omega, t) \mapsto v(\omega, t) := v(\omega)(t) \text{ is progressively measurable}\}.
\]
Concerning the terms defining the cost functions \( \ell \)

The functions \( a.s. \) in \((\omega,t)\) with the scalar product

where each \( b \) are given. In what follows we use the notation \( \| \cdot \| \)

For \( \sigma \) we will denote by \( \psi \) the gradient of \( x \) and \( \sigma \) is real valued. The columns of \( \sigma \) are written \( \sigma_j \) for \( j = 1, \ldots, d \).

For \( \psi = \ell, \Phi, \sigma \) we will denote by \( \nabla_x \psi \) the gradient of \( \psi \) w.r.t. to \( x \). We will also use the notation \( \nabla_x \psi \) to denote, respectively, the Jacobians of \( b \) and \( \sigma \) w.r.t. \( x \). Similar notations will be using 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 \( \sigma \) we will assume

(A1) For \( \psi = b^i, \sigma^j \) we have:

(i) \( \psi \) 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 \((x,u) \rightarrow \psi(\omega,t,x,u)\) belongs to \( C^1(\mathbb{R}^n \times \mathbb{R}^m)\), the application \((\omega,t) \in \Omega \times [0,T] \rightarrow \psi(\omega,t,\cdot,\cdot)\) is progressively measurable and there exists \( c_1 > 0 \) and \( \rho_1 \in L^{2,2}_{\mathbf{P}} \) such that almost surely (a.s.) in \((\omega,t)\)

\[
|\psi(\omega,t,x,u)| \leq c_1 (\rho_1(\omega,t) + |x| + |u|),
\]

\[
|\nabla_x \psi(\omega,t,x,u)| + |\nabla_u \psi(\omega,t,x,u)| \leq c_1. \tag{5.1}
\]

Concerning the terms defining the cost functions \( \ell \) and \( \Phi \) we will assume

(A2) The functions \( \ell \) and \( \Phi \) 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. \((\omega, t)\) the maps \((x, u) \mapsto \ell(\omega, t, x, u)\) and \(x \mapsto \Phi(\omega, x)\) are \(C^1\). The application \((\omega, t) \in \Omega \times [0, T] \mapsto \ell(\omega, t, \cdot, \cdot) \in C^1(\mathbb{R}^n \times \mathbb{R}^n)\) is progressively measurable. In addition, there exists \(c_2 > 0\), \(\rho_2 \in L^2_{\mathbb{F}}\) and \(\rho_3 \in L^2(\Omega, \mathcal{F}_T)\) such that almost surely in \((\omega, t)\) we have

\[
\begin{align*}
|\ell(\omega, t, x, u)| &\leq c_2 (\rho_2(\omega, t) + |x|^2 + |u|^2), \\
|\nabla_x \ell(\omega, t, x, u)| + |\nabla_u \ell(\omega, t, x, u)| &\leq c_2 (\rho_2(\omega, t) + |x| + |u|), \\
|\Phi(\omega, x)| &\leq c_2 (\rho_3(\omega) + |x|^2),
\end{align*}
\]

(5.2)

The previous assumptions are rather general and cover the case of linear quadratic problems (see e.g. [22, 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 [1] and we focus first in writing the SDE constraint in the form of an equality constraint in a suitable function space.

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

\[
I(x_0, x_1, x_2)(\cdot) := x_0 + \int_0^\cdot x_1(s)ds + \sum_{j=1}^d \int_0^\cdot x_2^j(s)dW^j(s).
\]

(5.3)

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

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

(5.4)

we have that \(\mathcal{I}^n\) is a Hilbert space, which, since \(I\) is injective (see [1, Lemma 2.1]), can be identified with \(\mathbb{R}^n \times (L^2_{\mathbb{F}})^n \times (L^2_{\mathbb{F}})^{n \times d}\). 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

\[
x(t) = x_0 + \int_0^t b(s, x(s), u(s))ds + \int_0^t \sigma(s, x(s), u(s))dW(s) \quad \forall t \in [0, T],
\]

(5.5)

It is well known that under (A1) equation (5.5) admits a unique solution \(x \in \mathcal{I}^n\) (see e.g. [24, Chapter 5]). It is also known that \(\mathbb{E} \left( \sup_{t \in [0, T]} |x(t)|^2 \right)\) is finite (see e.g. [1, Lemma 2.2]). A 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 \(u \in (L^2_{\mathbb{F}})^m\), the solution \(x \in \mathcal{I}^n\) satisfies

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

(5.6)

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

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

\[
|x(s)|^2 \leq 3 \left( |x_0|^2 + s \int_0^s |b(s', x(s'), u(s'))|^2ds' + \int_0^s |\sigma(s', x(s'), u(s'))dW(s')|^2 \right).
\]

By the linear growth condition in [53] and the fact that \(x \in \mathcal{I}^n\) and \(u \in (L^2_{\mathbb{F}})^m\), we have that \(\sigma(\cdot, x(\cdot), u(\cdot)) \in (L^2_{\mathbb{F}})^{n \times d}\) and so, for each \(j = 1, \ldots, d\), the \(\mathbb{R}^n\)-valued
Proof. Note that for any \((\sigma^j(s'), x(s'), u(s'))dW^j(s')\) 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 \(\sigma\) with respect to \(x\) in \((5.8)\) imply that
\[
g(t) \leq 3 \left[|x(0)|^2 + 2T \mathbb{E} \left( \int_0^t |b(s, x(s), u(s))|^2 \, ds \right) + 4T \mathbb{E} \left( \int_0^t |\sigma(s, x(s), u(s))|^2 \, ds \right) \right] + 4E \left( \int_0^t |\sigma(s, 0, u(s))|^2 + 2c_1^2|x(s)|^2 \, ds \right)
\leq a + b \int_0^t g(s) \, ds,
\]
where
\[
a = \max\{24, 6T\} \left[|x(0)|^2 + 2T \mathbb{E} \left( \int_0^t |b(s, 0, u(s))|^2 \, ds \right) + 4T \mathbb{E} \left( \int_0^t |\sigma(s, 0, u(s))|^2 \, ds \right) \right],
\]
and \(b = 6c_1^2 \max\{T, 4d\}\). The result then follows from Gronwall’s Lemma. ■

Remark 5.1 Estimates of the form \((5.4)\) can be easily extended to any power \(p > 1\) by using in the previous proof the Burkholder-Davis-Gundy inequality (see e.g. [25]) instead of Doob’s inequality.

Now, let us consider the application \(g : \mathcal{I}^n \times (L^2_{\mathbb{P}})^m \to \mathbb{R}^n\) defined by
\[
g(x, u)(\cdot) := \tilde{x}_0 + \int_0^{(\cdot)} b(s, x(s), u(s)) \, ds + \int_0^{(\cdot)} \sigma(s, x(s), u(s)) \, dW(s) - x(\cdot), \quad (5.7)
\]
which defines the SDE constraint in \((SP)\) by imposing \(g(x, u) = 0\). Consider also the application \(f : \mathcal{I}^n \times (L^2_{\mathbb{P}})^m \to \mathbb{R}\) defined by
\[
f(x, u) := \mathbb{E} \left( \int_0^T f(t, x(t), u(t)) \, dt + \Phi(x(T)) \right),
\]
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
\[
\inf f(x, u) \text{ subject to } g(x, u) = 0, \quad u \in \mathcal{U}. \quad (SP)
\]

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

We begin by studying some properties of \(g\). The following result is proved in the appendix in \(\Xi\). For the sake of completeness we provide here a short proof.

Lemma 5.2 Under \((A1)\) the mapping \(g\) is Lipschitz continuous and Gâteaux differentiable. Its Gâteaux derivative \(Dg(x, u) : \mathcal{I}^n \times (L^2_{\mathbb{P}})^m \to \mathcal{I}^n\) is given by
\[
Dg(x, u)(z, v)(\cdot) = \int_0^{(\cdot)} \left[ b_x(t, x(t), u(t))z(t) + b_u(t, x(t), u(t))v(t) \right] \, dt + \sum_{j=1}^d \int_0^{(\cdot)} \left[ \sigma_{x_j}(t, x(t), u(t))z(t) + \sigma_{u_j}(t, x(t), u(t))v(t) \right] \, dW^j(t)
\]
for all \((z, v) \in \mathcal{I}^n \times (L^2_{\mathbb{P}})^m\).

Proof. Note that for any \((x, u_1), (y, u_2) \in \mathcal{I}^n \times (L^2_{\mathbb{P}})^m\) we have
\[
\|g(x, u_1)(\cdot) - g(y, u_2)(\cdot)\|_{\mathcal{I}^n}^2
= |x_0 - y_0|^2 + \mathbb{E} \left( \int_0^T \left[ b(t, x(t), u_1(t)) - b(t, y(t), u_2(t)) \right] + y_1(t) - x_1(t) \right|^2 \, dt
+ \sum_{j=1}^d \mathbb{E} \left( \int_0^T \left[ \sigma(t, x(t), u_1(t)) - \sigma(t, y(t), u_2(t)) + y_2(t) - x_2(t) \right] \right|^2 \, dt),
\]

14
which, by the Lipschitz assumption in (5.1), is bounded by
\[
    c \left[ \|x - y\|^2_{\mathcal{F}^n} + \mathbb{E} \left( \int_0^T |x(t) - y(t)|^2 dt \right) + \mathbb{E} \left( \int_0^T |u^1(t) - u^2(t)|^2 dt \right) \right],
\]
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
\[
    \mathbb{E} \left( \int_0^T |x(t) - y(t)|^2 dt \right) \leq c' \|x - y\|^2_{\mathcal{F}^n},
\]
from which the Lipschitz property of \( g \) easily follows. Now, for \( j = 1, \ldots, d \) let us set
\[
    Db(t, x, u)(z, v) = b_x(t, x, u)z + b_u(t, x, u)v, \quad D\sigma^j(t, x, u)(z, v) = \sigma^j_x(t, x, u)z + \sigma^j_u(t, x, u)v
\]
and define
\[
    I_1 := \mathbb{E} \left( \int_0^T \left[ b(t, x(t) + \tau z(t), u(t) + \tau v(t)) - b(t, x(t), u(t)) \right](z(t), v(t)) \right)^2 dt, \\
    I_2 := \mathbb{E} \left( \int_0^T \sigma^j(t, x(t) + \tau z(t), u(t) + \tau v(t)) - \sigma^j(t, x(t), u(t)) \right)(z(t), v(t)) \right)^2 dt.
\]
By the Lipschitz property of \( b \) and \( \sigma \) in (5.1) and the dominated convergence theorem, we get that \( I_1 \) and \( I_2 \) tend to 0 as \( \tau \downarrow 0 \). This implies that
\[
    (x, u) \in \mathcal{T} \times (L^2_{\mathcal{F}})^m \rightarrow \int_0^1 b(s, x(s), u(s))ds + \int_0^1 \sigma(s, x(s), u(s))dW(s) \in \mathcal{T}
\]
is directionally differentiable with directional derivative
\[
    (z, v) \in \mathcal{T} \times (L^2_{\mathcal{F}})^m \rightarrow \int_0^1 Db(t, x(t), u(t))(z(t), v(t))dt \\
    + \sum_{j=1}^d \int_0^1 D\sigma^j(t, x(t), u(t))(z(t), v(t))dt.
\]
The continuity of the linear application above follows easily from the bounds in the second relation in (5.1). Finally, since \( (x, u) \in \mathcal{T} \times (L^2_{\mathcal{F}})^m \rightarrow x \in \mathcal{T} \) is \( C^\infty \) with derivative \( (z, v) \in \mathcal{T} \times (L^2_{\mathcal{F}})^m \rightarrow z \in \mathcal{T} \), we obtain (5.8).  

The previous lemma yields the following result

**Lemma 5.3** For every \((x, u) \in \mathcal{T} \times (L^2_{\mathcal{F}})^m\) and \(\delta \in \mathcal{T}\), there exists a unique \(z \in \mathcal{T}\) such that \(Dg(x, u)(z, 0) = \delta\). Moreover, there exists a constant \(c > 0\), independent of \((x, u, z, \delta)\), such that \(\|z\|_{\mathcal{T}} \leq c\|\delta\|_{\mathcal{T}}\).

**Proof.** By Lemma 5.2 we have that \(Dg(x, u)(z, 0) = \delta\) is equivalent to the SDE
\[
    d\hat{z} = [b_x(t, x(t), u(t))z(t) - \hat{z}] dt + [\sigma_x(t, x(t), u(t))z(t) - \hat{z}] dW(t), \\
    \hat{z}(0) = -\delta_0.
\]
The existence and uniqueness of a solution \(z\) of this equation is well-known (see e.g. [24] Chapter 5). Moreover, using that \(\|b_x\|_\infty \leq c_1\) and \(\|\sigma_x\|_\infty \leq c_1\), Lemma 5.1 implies the existence of a constant \(c > 0\), independent of \((x, u, z, \delta)\), such that
\[
    \|z\|_{\mathcal{T}} \leq c \left[ \|\delta_0\|^2 + \mathbb{E} \left( \int_0^T |\delta_1|^2 dt \right) + \mathbb{E} \left( \int_0^T |\delta_2|^2 dt \right) \right].
\]
The result follows.  

As a consequence of the last two lemmas and Theorem 4.1 \(g\) satisfies (4.1) with \(C := \mathcal{T} \times \mathcal{V}\) and \(\alpha = c\), where \(\mathcal{V}\) is any closed set of \((L^2_{\mathcal{F}})^m\). Therefore, the following result holds true.
Corollary 5.1 For any closed set $\mathcal{V} \subset (L^2_y)^n$, we have
\[
d((x, u), g^{-1}(y) \cap (T^n \times \mathcal{V})) \leq c\|g(x, u) - y\| \quad \forall \ x, \ y \in T^n \text{ and } u \in \mathcal{V}.
\]
Now, we consider the properties of the cost functional $f$.

Lemma 5.4 The function $f$ is locally Lipschitz and Gâteaux differentiable, with
\[
Df(x, u)(z, v) = \mathbb{E}\left(\int_0^T [\ell_x(t, x(t), u(t))z(t) + \ell_u(t, x(t), u(t))v(t)] \, dt\right)
+ \mathbb{E}\left(D\Phi(x(T))z(T)\right).
\]

Proof. For $\tau \in [0, 1]$, set $x_{\tau} := x_1 + (x_2 - x_1)$, $u_{\tau} := u_1 + \tau(u_2 - u_1)$, $\delta x = x_2 - x_1$ and $\delta u = u_2 - u_1$. We have that
\[
|f(x_2, u_2) - f(x_1, u_1)| \leq \mathbb{E}\left(\int_0^T \left| D\ell(t, x_{\tau}(t), u_{\tau}(t))(\delta x(t), \delta u(t)) \right| \, d\tau dt\right)
+ \mathbb{E}\left(\int_0^1 |D\Phi(x_{\tau}(T))\delta x(T)| \, d\tau \right).
\]
By the second assumption in (5.2), we can find $c > 0$ such that
\[
|D\ell(t, x_{\tau}(t), u_{\tau}(t))(\delta x(t), \delta u(t))| \leq c(1 + |x_{\tau}(t)| + |u_{\tau}(t)|)(|\delta x(t)| + |\delta u(t)|)
\]
\[
\leq c(1 + |x_1(t)| + |\delta x(t)| + |u_1(t)| + |\delta u(t)|)(|\delta x(t)| + |\delta u(t)|),
\]
which, by the Cauchy-Schwarz inequality, implies that
\[
\left[\mathbb{E}\left(\int_0^T \left| D\ell(t, x_{\tau}(t), u_{\tau}(t))(\delta x(t), \delta u(t)) \right| \, d\tau dt\right)\right]^2 \leq

\leq c' \mathbb{E}\left(\int_0^1 (1 + |x_1(t)|^2 + |\delta x(t)|^2 + |u_1(t)|^2 + |\delta u(t)|^2) \, dt\right) \left(\|\delta x\|_{2, 2}^2 + \|\delta u\|_{2, 2}^2\right).
\]
Analogously, there exists $c'' > 0$ such that
\[
\left[\mathbb{E}\left(\int_0^1 |D\Phi(x_{\tau}(T))\delta x(T)| \, d\tau\right)\right]^2 \leq c'' \mathbb{E}\left(1 + |x_1(T)|^2 + |\delta x(T)|^2\right) \mathbb{E}\left(|\delta x(T)|^2\right),
\]
from which the local Lipschitz property for $f$ follows. Now, we prove the formula for the directional derivative. Consider the term
\[
\mathbb{E}\left(\int_0^T \left[\ell(t, x(t) + \tau z(t), u(t) + \tau v(t)) - \ell(t, x(t) + \tau z(t), u(t) + \tau v(t)) - D\ell(t, x(t), u(t))\right] \, dt\right).
\]
Since $\ell$ is Gâteaux differentiable, the expression inside the integral converges to zero pointwise. Now, writing the ratio inside the integral in integral form, if $\tau < 1$, we have
\[
\int_0^1 D\ell(t, x_{\tau}(t), u_{\tau}(t))(z(t), v(t)) \, d\tau \leq c(1 + |x(t)| + |z(t)| + |u(t)| + |v(t)|)(|z(t)| + |v(t)|),
\]
where $x_{\tau} = x + \tau z$ and $u_{\tau} = u + \tau v$. 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 $\tau \downarrow 0$,
\[
\mathbb{E}\left(\Phi(x(T) + \tau z(T)) - \Phi(x(T)) \right) / \tau - D\Phi(x(T))z(T) \to 0.
\]
Formula (5.9) follows. $\blacksquare$

As 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 5.3 and Corollary 5.4 we can prove now a weak version of the stochastic Pontryagin's minimum principle (see [29] and Remark 5.1 ii) below).
Theorem 5.1 Suppose that \((\tilde{x}, \tilde{u})\) is a local solution of problem \((SP)\), then there exists \(\bar{p} \in \mathcal{I}^n\) and \(\bar{q} \in (L^2_\mathbb{F})^{n \times d}\) such that
\[
\bar{p}(\cdot) = \nabla_x \Phi(\tilde{x}(T)) + \int_0^T \nabla_x H(s, \tilde{x}(s), \bar{p}(s), \bar{q}(s), \tilde{u}(s)) ds - \int_0^T \bar{q}(s) dW(s),
\]
and
\[
\mathbb{E} \left( \int_0^T \nabla_u H(t, \tilde{x}(t), \bar{p}(t), \bar{q}(t), \tilde{u}(t)) \cdot v(t) dt \right) \geq 0 \quad \text{for all } v \in T_{\mathcal{U}}(\tilde{u}).
\]
(5.11)

If, in addition, \(K(\mathcal{U}, \tilde{u})\) is convex, then the second relation in (5.11) is valid for all \(v \in K(\mathcal{U}, \tilde{u})\).

Proof. Lemma 5.2 and Lemma 5.4 imply that \(g\) and \(f\) satisfy the assumptions \((\mathcal{H}_g)\) and \((\mathcal{H}_f)\), respectively. Since Corollary 5.1 implies that \(g\) is calm at \((\tilde{x}, \tilde{u})\), Theorem 5.1 yields the existence of a Lagrange multiplier \(\lambda \in \mathcal{I}^n\) such that
\[
0 \in Df(\tilde{x}, \tilde{u}) + Dg(\tilde{x}, \tilde{u})^* \lambda + \{0\} \times N_{\mathcal{U}}(\tilde{u}),
\]
which can be written as
\[
D_x f(\tilde{x}, \tilde{u}) + D_u g(\tilde{x}, \tilde{u})^* \lambda = 0,
\]
\[
\langle D_u f(\tilde{x}, \tilde{u}) + D_u g(\tilde{x}, \tilde{u})^* \lambda, v \rangle_{(L^2_\mathbb{F})^m} \geq 0 \quad \forall \; v \in T_{\mathcal{U}}(\tilde{u}).
\]
(5.12)

Setting \(\bar{p} = \lambda_1\) and \(\bar{q} = \lambda_2\), [1] Theorem 3.12 implies that the first and second relations in (5.11) are equivalent to the corresponding relations in (5.12). Finally, if \(K(\mathcal{U}, \tilde{u})\) is convex, by Theorem 5.3 we have
\[
0 \in Df(\tilde{x}, \tilde{u}) + Dg(\tilde{x}, \tilde{u})^* \lambda + \{0\} \times K(\mathcal{U}, \tilde{u})^0.
\]
Reasoning as before, we have that the second relation in (5.11) is valid for all \(v \in K(\mathcal{U}, \tilde{u})\). The result follows. $\blacksquare$

5.1 Comments and extensions

Let us provide some comments on the previous result.

(i) As pointed out in [1], 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 \((SP)\). The results presented in Section 5 and in Section 4 allow us to overcome this difficulty.

(ii) It is possible to prove Theorem 5.1 by following a different strategy that does not involve the Lagrange multiplier theory. In order to simplify the discussion, we suppose that no constraints are imposed on the controls, i.e. \(\mathcal{U} = (L^2_\mathbb{F})^m\), and refer the interested reader to [1] for the detailed presentation in the general case. Assumption \((A1)\) implies that for each \(u \in (L^2_\mathbb{F})^m\), the equation \(g(x, u) = 0\) admits a unique solution \(x[u] \in \mathcal{I}^n\). As a consequence, problem \((SP)\) can be rewritten as the unconstrained optimization problem
\[
\inf \left\{ J(u) := f(x[u], u) \; ; \; \text{s.t. } u \in (L^2_\mathbb{F})^m \right\}.
\]
\((SP')\)

If \(\tilde{u}\) is a local solution of \((SP')\), then it is possible to provide a first order expansion of \(v \in (L^2_\mathbb{F})^m \to J(\tilde{u} + v)\) if \(v\) is progressively-measurable and essentially bounded. By defining \((\bar{p}, \bar{q})\) by the first relation in (5.11) (which can be justified by the results in [2]), the aforementioned expansion of \(J\) implies that the second relation in (5.11) holds for every essentially bounded \(v\) and so, by a density argument, for every \(v \in (L^2_\mathbb{F})^m\). Even if this approach provides another proof of Theorem 5.1, the latter is considerably more technical than the one presented in this article and does not provide the explicit relation between \(\bar{p}\) and \(\bar{q}\) and the Lagrange multiplier \(\lambda\) associated to the SDE defining the controlled trajectories.
(iii) In the particular case of pointwise control constraints
\[ U := \{ u \in (L_{F}^{2,2})^{m} : u(\omega, t) \in U \ a.s. \}, \]
where \( U \subseteq \mathbb{R}^{m} \) is a nonempty closed set, a result stronger than Theorem 5.1 has been shown in [28]. In this paper, the author shows that a variation of the Hamiltonian \( H \), which involves an additional pair of adjoint processes, is almost surely pointwisely minimized at \( \hat{u}(\omega, t) \). In this result, no regularity assumptions on the data with respect to \( u \) are imposed. On the other hand, stronger assumptions with respect to the dependence on the state variable \( x \) are assumed (which involve strong requirements on the second order derivatives of \( \ell, \Phi, b \) and \( \sigma \)).

(iv) A straightforward extension of Theorem 5.1 is the case where the initial point \( \hat{x}_{0} \) is also a decision variable. More precisely, let \( X_{0} \subseteq \mathbb{R}^{n} \) be a closed set and consider the following extension of problem \((SP)\)
\[
\inf_{x, \hat{x}, u} \mathbb{E} \left( \int_{0}^{T} \ell(\omega, t, x(t), u(t))dt + \Phi(\omega, x(T)) \right) \quad \text{s.t.} \quad \begin{aligned}
    dx(t) &= b(\omega, t, x(t), u(t))dt + \sigma(\omega, t, x(t), u(t))dW(t) & t \in (0, T), \\
    x(0) &= \hat{x}_{0} \in X_{0}, \\
    u &\in U.
\end{aligned} \tag{SP'}
\]
Then, this problem can be written in the abstract form
\[
\inf f(x, u) \text{ subject to } \tilde{g}(x, u) \in \mathcal{I}^{n} \times X_{0}, \quad u \in U, \quad \text{(SP')} \]
where
\[
\tilde{g}(x, u) := \left( x(0) + \int_{0}^{(\cdot)} b(s, x(s), u(s))ds + \int_{0}^{(\cdot)} \sigma(s, x(s), u(s))dW(s) - x(\cdot), x(0) \right).
\]
Suppose that \((\hat{x}, \hat{u}) \in \mathcal{I}^{n} \times U\) 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 5.13), it is easy to check that \((H'_{eq})\) is satisfied at \((\hat{x}, \hat{u})\) (with \( C = \mathcal{I}^{n} \times U \) and \( D = \mathcal{T}^{n} \times X_{0} \)). Thus, by Theorem 4.2, Theorem 5.2 and reasoning as in the proof of Theorem 5.1, we obtain the existence of \( \tilde{p} \in \mathcal{I}^{n} \) and \( \tilde{q} \in (L_{F}^{2,2})^{n \times d} \) such that
\[
\tilde{p}(\cdot) = \nabla_{x} \Phi(\hat{x}(T)) + \int_{0}^{T} \nabla_{x} H(s, \hat{x}(s), \tilde{p}(s), \tilde{q}(s), \hat{u}(s))ds - \int_{0}^{T} \tilde{q}(s)dW(s),
\]
\[
-\tilde{p}(0) \in N_{X_{0}}(\hat{x}(0)), \quad \text{and} \quad \mathbb{E} \left( \int_{0}^{T} \nabla_{x} H(t, \hat{x}(t), \tilde{p}(t), \tilde{q}(t), \hat{u}(t)) \cdot v(t)dt \right) \geq 0 \quad \text{for all} \quad v \in T_{U}(\hat{u}). \tag{5.13}
\]
(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 \((\hat{x}, \hat{u})\) in order to ensure that \((H'_{eq})\) holds. We refer the reader to [H] for a more detailed discussion on this matter. The case of final pointwise constraints having the form \( x(\omega, T) \in X_{T} \), for some closed set \( X_{T} \subseteq \mathbb{R}^{n} \), 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 \( w_{1}, \ldots, w_{N} \) be \( N \) independent \( \mathbb{R}^{d} \)-valued random variables
defined in \((\Omega, \mathcal{F}, \mathbb{P})\) such that for all \(k = 1, \ldots, N\) the coordinates of \(w_k = (w^1_k, \ldots, w^d_k)\) are independent and satisfy
\[
\mathbb{E}(w^i_k) = 0, \quad \mathbb{E}(|w^i_k|^2) = 1.
\]
Define \(w_0 := 0\) and for \(k = 0, \ldots, N\) set \(\mathcal{F}_k := \sigma(w_0, \ldots, w_k)\), the sigma-algebra generated by \(w_0, \ldots, w_k\), and
\[
L^2_{\mathcal{F}_k} := \{ y \in L^2(\Omega) : y \text{ is } \mathcal{F}_k \text{ measurable} \}.
\]
Let \(U \subseteq \Pi_{k=0}^{N-1}(L^2_{\mathcal{F}_k})^m\) be a non-empty closed set. In this section we consider the following discrete-time stochastic optimal control problem (see [25])
\[
\begin{aligned}
\inf \mathbb{E} & \left( \sum_{k=0}^{N-1} \ell(k, x_k, u_k) + \Phi(x_N) \right) \\
\text{s.t.} & \quad x_{k+1} = b(k, x_k, u_k) + \sigma(k, x_k, u_k)w_{k+1} \quad k = 0, \ldots, N-1 \\
\text{and} & \quad x_0 = x_0 \in \mathbb{R}^n, \quad u \in U,
\end{aligned}
\tag{SP_d}
\]
where, denoting \([0:N-1] := \{0, \ldots, N-1\}\), \(\ell : [0:N-1] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}\), \(\Phi : \mathbb{R}^n \to \mathbb{R}\), \(b : [0:N-1] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n\) and \(\sigma : [0:N-1] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n \times d}\) are Borel measurable functions. Denoting \(\sigma^j (j = 1, \ldots, d)\) the \(j\)th column of \(\sigma\), for \(\psi = b, \sigma\) we suppose that \(\psi \in \mathcal{C}^1\) with respect to \((x, u)\) and the existence of \(c_1 > 0\) such that for all \(k \in [0:N-1]\)
\[
\begin{align*}
|\psi(x, u)| & \leq c_1 (1 + |x| + |u|), \\
|\psi_x(x, u)| + |\psi_u(x, u)| & \leq c_1.
\end{align*}
\tag{6.1}
\]
Similarly, in the remainder of this section we will assume that there exists \(c_2 > 0\) such that for all \(k \in [0:N-1]\)
\[
\begin{align*}
|\ell(k, x, u)| & \leq c_2 (1 + |x| + |u|)^2, \\
|\ell_x(k, x, u)| + |\ell_u(k, x, u)| & \leq c_2 (1 + |x| + |u|), \\
|\Phi(x)| & \leq c_2 (1 + |x|)^2, \quad |\Phi_x(x)| \leq c_2 (1 + |x|).
\end{align*}
\tag{6.2}
\]
As in Section 3 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
\[
X_k := \left\{ y_{k-1}^0 + \sum_{i=1}^d y_{k-1}^i w_{k-1}^i : y_{k-1}^i \in \left(L^2_{\mathcal{F}_{k-1}}\right)^n \quad \forall i = 0, \ldots, d \right\}.
\]
Endowed with the scalar product
\[
\langle x, x' \rangle_{X_k} := \mathbb{E} \left( \sum_{i=0}^d y_{k-1}^i \cdot z_{k-1}^i \right) \quad \forall x = y_{k-1}^0 + \sum_{i=1}^d y_{k-1}^i w_{k-1}^i, \quad x' = z_{k-1}^0 + \sum_{i=1}^d z_{k-1}^i w_{k-1}^i,
\]
the following elementary result shows that \(X_k\) is a Hilbert space.

**Lemma 6.1** For every \((y_{k-1}^0, y_{k-1}^1, \ldots, y_{k-1}^d) \in \left(L^2_{\mathcal{F}_{k-1}}\right)^n \times \left(L^2_{\mathcal{F}_{k-1}}\right)^{n \times d}\) we have
\[
\mathbb{E} \left( \left( y_{k-1}^0 + \sum_{i=1}^d y_{k-1}^i w_{k-1}^i \right)^2 \right) = \sum_{i=0}^d \mathbb{E} \left( |y_{k-1}^i|^2 \right).
\tag{6.3}
\]
As a consequence, for every \(k \in [1:N]\) the linear operator
\[
I : \left(L^2_{\mathcal{F}_{k-1}}\right)^n \times \left(L^2_{\mathcal{F}_{k-1}}\right)^{n \times d} \to X_k
\]
deﬁned as
\[
I(y_{k-1}^0, y_{k-1}^1, \ldots, y_{k-1}^d) := y_{k-1}^0 + \sum_{i=1}^d y_{k-1}^i w_{k-1}^i,
\]
is a bijection.
Proof. Relation (6.3) follows directly from the relations
\[
\mathbb{E} (y^0_{k-1} \cdot y^i_{k-1} w^j_k) = \mathbb{E} (y^0_{k-1} \cdot y^i_{k-1} \mathbb{E} (w^j_k | \mathcal{F}_{k-1})) = 0 \quad \forall \ i \in [1 : d],
\]
\[
\mathbb{E} (y^i_{k-1} \cdot y^j_{k-1} w^i_k w^j_k) = \mathbb{E} (y^i_{k-1} \cdot y^j_{k-1} \mathbb{E} (w^i_k w^j_k | \mathcal{F}_{k-1})) = \begin{cases} \mathbb{E} (|y^i_{k-1}|^2) & \text{if } i = j, \\ 0 & \text{otherwise} \end{cases}
\]
By definition of \( X_k \) we only need to show that \( I \) is injective. But this is clear because if
\[
I(y^0_{k-1}, y^1_{k-1}, \ldots, y^d_{k-1}) = 0,
\]
then (6.3) implies that \( \mathbb{E} (|y^i_{k-1}|^2) = 0 \) and so \( y^i_{k-1} = 0 \) a.e. for all \( i \in [0 : d] \).

Define \( g : \Pi^N_{k=0} X_k \times \Pi^{N-1}_{k=0} (L^2_{x_k})^m \to \Pi^N_{k=0} X_k \) as
\[
g_0(x, u) := x_0 - x_0,
\]
\[
g_{k+1}(x, u) := b(k, x_k, u_k) + \sigma(k, x_k, u_k)w_{k+1} - x_{k+1} \quad \forall \ k = 0, \ldots, N-1,
\]
and \( f : \Pi^N_{k=0} X_k \times \Pi^{N-1}_{k=0} (L^2_{x_k})^m \to \mathbb{R} \) 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
\[
\inf f(x, u) \text{ subject to } g(x, u) = 0, \quad u \in \mathcal{U}.
\]
As in the previous section, we prove now that if we set \( X := \Pi^N_{k=0} X_k \times \Pi^{N-1}_{k=0} (L^2_{x_k})^m \), then under our assumptions the mappings \( f \) and \( b \) 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 \( Dg(x, u)(z, v) = (Dg_0(x, u)(z, v), \ldots, Dg_N(x, u)(z, v)) \), where
\[
Dg_0(x, u)(z, v) = -z_0,
\]
\[
Dg_{k+1}(x, u)(z, v) = b(x, u)(k, x_k, u_k)(z_k, v_k) + \sum_{i=1}^d \sigma^i_{(x, u)}(k, x_k, u_k)(z_k, v_k)w^i_{k+1} - z_{k+1},
\]
for all \( k = 0, \ldots, N-1 \).
(ii) The mapping \( f \) is locally Lipschitz and Gâteaux differentiable, with
\[
Df(x, u)(z, v) = \mathbb{E} \left( \sum_{k=0}^{N-1} \ell_{(x, u)}(k, x_k, u_k)(z_k, v_k) + D\Phi(x_N)z_N \right),
\]
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 \( k = 0, \ldots, N-1 \),
\[
\|g_{k+1}(x^1, u^1) - g_{k+1}(x^2, u^2)\|^2_{\mathcal{X}_{k+1}} \\
= \mathbb{E} \left( |b(k, x^1_k, u^1_k) - b(k, x^2_k, u^2_k)|^2 + \sum_{i=1}^d |\sigma^i(k, x^1_k, u^1_k) - \sigma^i(k, x^2_k, u^2_k)|^2 \right) \\
\leq c \mathbb{E} \left( |x^1_k - x^2_k|^2 + |u^1_k - u^2_k|^2 \right) = c \left( \|x^1_k - x^2_k\|^2_{\mathcal{X}_k} + \|u^1_k - u^2_k\|^2_{L^2_{x_k}} \right),
\]
Proof. Noting that (6.6). Relation (6.7) follows directly from (6.6) and Theorem 4.1. where \( \bar{V} \subseteq \mathbb{R}^n \) that \( z, v \rightarrow Dg(x,u)(z, v) \) 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 \( Dg(x,u)(z, v) = 0 \). Moreover, there exists \( c > 0 \), independent of \((x,u,z,\delta)\), such that

\[
\sum_{k=0}^N \|z_k\|_{X_k} \leq c \sum_{k=0}^N \|\delta_k\|_{X_k}.
\]

(6.6)

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

\[
d((x,u), \mathcal{V}) \leq c \|\delta\|_{X_k} \quad \forall (x, u) \in X, \ y \in \Pi_{k=0}^N X_k.
\]

(6.7)

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

\[
z_0 = -\delta_0, \\
z_{k+1} = b_k(k, x_k, u_k)z_k + \sum_{i=1}^d \sigma^i_k(k, x_k, u_k)z_kw_{k+1}^i - \delta_{k+1} \quad \forall k = 0, \ldots, N-1.
\]

Noting that

\[
\|z_{k+1}\|_{X_{k+1}} = E(|z_{k+1}|^2) \leq (d+2)^2 \left[ c^2 E(|z_k|^2) + c^2 \sum_{i=1}^d E(|z_k|^2|w_{k+1}^i|^2) + |\delta_{k+1}|^2 \right],
\]

\[
\leq (d+2)^2 c E(|z_k|^2) + (d+2)E(|\delta_{k+1}|^2),
\]

\[
\leq \tilde{c} E(|z_k|^2) + E(|\delta_{k+1}|^2),
\]

\[
\leq (N+1)\tilde{c}^{N+1} \sum_{k=0}^N E(|\delta_k|^2),
\]

\[
= (N+1)\tilde{c}^{N+1} \sum_{k=0}^N \|\delta_k\|^2_{X_k},
\]

where \( \tilde{c} := (d+2)^2(c^2 + 1) > 1 \) and the last equality is a consequence of (6.3). This proves (6.6). Relation (6.7) follows directly from (6.6) and Theorem 6.1. Let us define the Hamiltonian \( H : [0 : N-1] \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{m} \to \mathbb{R} \) 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 (SPd).

**Theorem 6.1** Suppose that \((\bar{x}, \bar{u})\) is a local solution to (SPd). 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

\[
p_{k-1}^i = E(\nabla_x H(k, \bar{x}_k, p_k, q_k, \bar{u}_k)|_{\mathcal{F}_{k-1}}) \quad \forall k \in [1 : N-1],
\]

\[
g_{k-1}^i = E(\nabla_q H(k, \bar{x}_k, p_k, q_k, \bar{u}_k)w_k^i|_{\mathcal{F}_{k-1}}) \quad \forall k \in [1 : N-1], i \in [1 : d],
\]

(6.8)

\[
q_{N-1}^i = E(\nabla q H(\bar{x}_N)|_{\mathcal{F}_{N-1}}) \quad \forall i \in [1 : d],
\]

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

(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 6.1 there exists \( \lambda \in \Pi_{k=0}^{N} X_{k} \) such that

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

from which we deduce that for all \( z = (z_{0}, \ldots, z_{N}) \in \Pi_{k=0}^{N} X_{k} \)

\[
\begin{align*}
D_{xk} f(\bar{x}, \bar{u}) z_{k} + \sum_{j=0}^{N} (\lambda_{j}, D_{xk} g_{j}(\bar{x}, \bar{u}) z_{k}) X_{j} &= 0 \quad \forall k = 0, \ldots, N, \\
D_{uk} f(\bar{x}, \bar{u}) v + \sum_{k=1}^{N} (\lambda_{k}, D_{uk} g_{k}(\bar{x}, \bar{u}) v) X_{k} &\geq 0 \quad \forall v \in T_{U(\bar{u})}.
\end{align*}
\]

(6.10)

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

\[
\begin{align*}
E(\ell_{k}(\bar{x}_{k}, \bar{u}_{k}) z_{k}) + \left( \lambda_{k+1}, b_{k}(\bar{x}_{k}, \bar{u}_{k}) z_{k} + \sum_{i=1}^{d} w_{k+1}^{i} \sigma_{k}^{i}(k, \bar{x}_{k}, \bar{u}_{k}) z_{k} \right) \big|_{X_{k+1}} \\
E(\Phi(\bar{x}_{N}) z_{N}) &= \langle \lambda_{k}, z_{k} \rangle \big|_{X_{k}},
\end{align*}
\]

(6.11)

Setting

\[
z_{k} = y_{k-1}^{0} + \sum_{i=1}^{d} g_{k-1}^{i} w_{k}^{i} \in X_{k}, \quad \lambda_{k} = p_{k-1} + \sum_{i=1}^{d} q_{k-1}^{i} w_{k}^{i} \in X_{k},
\]

relation (6.11) yields

\[
\begin{align*}
E(\nabla_{x} H(k+1, \bar{x}_{k+1}, p_{k+1}, q_{k+1}, \bar{u}_{k+1}) \cdot z_{k}) &= E \left( p_{k-1} \cdot y_{k-1}^{0} + \sum_{i=1}^{d} q_{k-1}^{i} \cdot y_{k-1}^{0} \right), \\
E(\nabla \Phi(\bar{x}_{N}) \cdot z_{N}) &= E \left( p_{N-1} \cdot y_{N-1}^{0} + \sum_{i=1}^{d} q_{N-1}^{i} \cdot y_{N-1}^{0} \right).
\end{align*}
\]

(6.12)

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

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

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

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

\[
\langle \lambda_{k+1}, D_{uk} g_{k}(\bar{x}, \bar{u}) v \rangle \big|_{X_{k+1}} = E \left( p_{k} \cdot b_{k}(k, \bar{x}_{k}, \bar{u}_{k}) v_{k} + \sum_{i=1}^{d} q_{k}^{i} \cdot \sigma_{k}^{i}(k, \bar{x}_{k}, \bar{u}_{k}) v_{k} \right),
\]

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

**Remark 6.1** (i) The optimality system (6.8)-(6.9) has been first shown in [25] under more restrictive assumptions on \( \ell, \Phi, f, \sigma \) (see [25] Equation (9)) and the control constraint set \( U \) (see [25] 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 3), it is easy to extend the results in this section to the case where the initial state \( \bar{x}_{0} \) is a decision variable subject to the constraint \( \bar{x}_{0} \in X_{0} \), where \( X_{0} \) is a closed subset of \( \mathbb{R}^{n} \). In this case, the optimality system is as in Theorem 6.1 with the additional constraint on the adjoint state (called transversality condition) \( -p_{0} \in N_{X_{0}}(\bar{x}_{0}) \).

**References**

[1] J. Backhoff and F. J. Silva, *Sensitivity results in stochastic optimal control: A Lagrangian perspective*, ESAIM: COCV, 23, (2017), 39-70.
[2] J.-M. Bismut, *Linear quadratic optimal stochastic control with random coefficients*, SIAM J. Control Optimization 14, (1976), 419-444.

[3] J. F. Bonnans and A. Shapiro, *Perturbations analysis of optimization problems*, Springer Series in Operations Research, Springer-Verlag, Berlin, 2006.

[4] J.P. Bonnans F. J. Silva, *First and second order necessary conditions for stochastic optimal control problems*, Applied Mathematics and Optimization 65, (2012), 403-439.

[5] J. M. Borwein and H. M. Strojwas, Tangential approximations, Nonlinear Anal., 9 (1985), 1347-1366.

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

[7] M. Brokate, A regularity condition for optimization in Banach spaces : Counter examples, Appl. Math. Optim., 6 (1980), 189-192.

[8] F. H. Clarke, Optimization and Nonsmooth Analysis. Wiley-Interscience, New York, 1983. Republished as vol. 5 of *Classics in Applied Mathematics*, SIAM, 1990.

[9] A. Dontchev and R.T. Rockafellar, Implicit functions and solutions mappings, Springer Monographs in Mathematics, Springer, Dordrecht, 2009.

[10] I. Ekeland, On the variational principle, Journal of Mathematical Analysis and Applications, 47 (2), (1974), 324–353.

[11] I. Ekeland, An inverse function theorem in Fréchet spaces, Ann. Inst. H. Poincaré Anal. Non Linéaire, 28 (1) (2011) 91–105.

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

[13] B.M. Glover and B.D. Craven, A Fritz John optimality condition using the approximate subdifferential, J. Optim. Th. Appl., 82 (1994), 253-265.

[14] A. D. Ioffe, Approximate subdifferentials and applications 2 : Functions on locally convex spaces, Mathematica, 33 (1986), 111-128.

[15] A. D. Ioffe, Approximate subdifferentials and applications 3 : Metric theory, Mathematica, 36 (1989), 1-38.

[16] A.D. Ioffe, Codirectional compactness, metric regularity and subdifferential calculus, Constructive, Experimental and Nonlinear Analysis (M. Théra ed.), Canadian Math. Soc. Proc. Conferences Series 27 (2000), 123-163, Amer. Math. Soc. Providence.

[17] A.D. Ioffe, Variational analysis of regular mappings, Springer Monographs in Mathematics, Springer, Cham, 2017.

[18] Y. Ishizuka, Optimality conditions for directionally differentiable multiobjective programming problems, J. Optim. Th. Appl., 72 (1992) 91-112.

[19] A. Jourani, Qualification conditions for multivalued functions in Banach spaces with applications to nonsmooth vector optimization problems, Math. Prog., 66 (1994), 1-23.

[20] A. Jourani, *The role of locally compact cones in nonsmooth analysis*, Communications Appl. Nonlinear. Anal., 5 (1998), 1-35.

[21] A. Jourani and L. Thibault, Approximations and metric regularity in mathematical programming in Banach spaces, Math. Oper. Res., 18 (1993), 390-401.
[22] A. Jourani and L. Thibault, Metric regularity for strongly compactly Lipschitzian mappings, Nonlinear Anal. Th. Meth. Appl. 24 (1995), 229-240.

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

[24] I. Karatzas and S.E. Shreve, Brownian Motion and Stochastic Calculus, Second Edition Springer-Verlag, New York, 1991.

[25] X. Lin and W. Zhang, A Maximum Principle for Optimal Control of Discrete-Time Stochastic Systems With Multiplicative Noise, IEEE Transactions on Automatic Control, 60 (2015), 1121-1126.

[26] M. Minami, Weak Pareto-optimal necessary conditions in a nondifferentiable multi-objective program on a Banach space, J. Optim. Th. Appl., 41 (1983) 451-461.

[27] B.S. Mordukhovich, Variational analysis and generalized differentiation. I, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 330, Springer-Verlag, Berlin, 2006.

[28] 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), 539–567.

[29] S. Peng, A General Stochastic Maximum Principle for Optimal Control Problems, SIAM J. Control Optim., 28, (1990), 966-979.

[30] P. Protter, Stochastic Integration and Differential Equations, Second Edition, Springer-Verlag, Berlin, 2005.

[31] P.L. Yu, Multicriteria decision making: Concepts, Techniques and Extensions (Plenum Press, New York, 1985).

[32] J. Yong and X.Y. Zhou, Stochastic controls, Hamiltonian systems and HJB equations, Springer-Verlag, New York, Berlin, 2000.