



**HAL**  
open science

# Sequential convex programming for non-linear stochastic optimal control

Riccardo Bonalli, Thomas Lew, Marco Pavone

► **To cite this version:**

Riccardo Bonalli, Thomas Lew, Marco Pavone. Sequential convex programming for non-linear stochastic optimal control. *ESAIM: Control, Optimisation and Calculus of Variations*, 2022, 28, pp.64. 10.1051/cocv/2022060 . hal-03823441

**HAL Id: hal-03823441**

**<https://hal.science/hal-03823441>**

Submitted on 20 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## SEQUENTIAL CONVEX PROGRAMMING FOR NON-LINEAR STOCHASTIC OPTIMAL CONTROL\*

RICCARDO BONALLI<sup>1,\*\*</sup>, THOMAS LEW<sup>2</sup> AND MARCO PAVONE<sup>2</sup>

**Abstract.** This work introduces a sequential convex programming framework for non-linear, finite-dimensional stochastic optimal control, where uncertainties are modeled by a multidimensional Wiener process. We prove that any accumulation point of the sequence of iterates generated by sequential convex programming is a candidate locally-optimal solution for the original problem in the sense of the stochastic Pontryagin Maximum Principle. Moreover, we provide sufficient conditions for the existence of at least one such accumulation point. We then leverage these properties to design a practical numerical method for solving non-linear stochastic optimal control problems based on a deterministic transcription of stochastic sequential convex programming.

**Mathematics Subject Classification.** 49K40, 65C30, 93E20.

Received November 9, 2021. Accepted September 17, 2022.

### 1. INTRODUCTION

Over the past few decades, the applied control community has devoted increasing attention toward the optimal control of stochastic systems. The general formulation of this problem entails steering a dynamical system from an initial configuration to a final configuration while optimizing some prescribed performance criterion (*e.g.*, minimizing some given cost) and satisfying constraints. This dynamical system is also subject to uncertainties which are modeled by Wiener processes and may come from unmodeled and/or unpredicted behaviors, as well as from measurement errors. Active development of new theoretical and numerical methods to address this problem continues, and the subject already has a rich literature.

We can classify the existing works into two main categories.

The first category consists of contributions that focus on Linear Convex Problems (LCPs), *i.e.*, whose dynamics are linear and whose costs are convex in both state and control variables. An important class of LCPs is given by Linear Quadratic Problems (LQPs) in which costs are quadratic in both state and control variables and for which the analysis of optimal solutions may be reduced to the study of an algebraic relation known as Stochastic Riccati Equation (SRE) [8, 42, 44, 49]. Efficient algorithmic frameworks have been devised to numerically solve LCPs, ranging from local search [27] and dual representations [28, 46], to deterministic-equivalent reformulations [4, 6], among others. In the special case of LQPs, those techniques may be further improved

---

\*This research was supported by the National Science Foundation under the CPS program (grant #1931815).

*Keywords and phrases:* Nonlinear stochastic optimal control, sequential convex programming, convergence of Pontryagin extremal-numerical deterministic reformulation.

<sup>1</sup> Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190 Gif-sur-Yvette, France.

<sup>2</sup> Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305, USA.

\*\* Corresponding author: [riccardo.bonalli@centralesupelec.fr](mailto:riccardo.bonalli@centralesupelec.fr)

by combining SRE theory with semidefinite programming [45, 53], finite-dimensional approximation [5, 19], or chaos expansion [33].

The second category of works deals with problems that do not enjoy any specific regularity, allowing non-linear dynamics or non-convex (therefore non-quadratic) costs. Throughout this paper, we call these Non-Linear Problems (NLPs). It is unquestionable that NLPs have so far received less attention from the community than LCPs, especially since the analysis of the former is usually more involved. Similar to the deterministic case, there are two main theoretical tools that have been developed to analyze NLPs: stochastic Dynamic Programming (DP) [3, 35] and the stochastic Pontryagin Maximum Principle (PMP) [31, 41, 43] (an extensive survey of generalizations of DP and the PMP may be found in [54]). In the case of LQPs, one can show that DP and the PMP lead to SRE [54]. DP provides optimal policies through the solution of a partial differential equation, whereas the necessary conditions for optimality offered by the PMP allow one to set up a two-point boundary value problem which returns candidate locally-optimal solutions when solved. Both methods only lead to analytical solutions in a few cases, and they can involve complex numerical challenges (the stochastic setting is even more problematic than the deterministic one, the latter being better understood for a wide range of problems, see, *e.g.*, [11, 14, 51]). This has fostered the investigation of more tractable approaches to solve NLPs such as Monte Carlo simulation [24, 47], Markov chain discretization [29, 30], and deterministic (though non-equivalent) reformulations [2, 4], among others. Importantly, many of the aforementioned approaches, *e.g.*, [29], are often based on some sort of approximation of the original formulation and thus offer powerful alternatives to DP and the PMP, especially since they are more numerically tractable and can be shown to converge to policies satisfying DP or the stochastic PMP.

In this paper, our objective is to lay the theoretical foundations to leverage Sequential Convex Programming (SCP) for the purposes of computing candidate optimal solutions for a specific class of NLPs. SCP is among the most well-known and earliest approximation techniques for deterministic non-linear optimal control and, to the best of our knowledge, such an approach has not been extended to stochastic settings yet. The simplest SCP scheme (which we consider in this work) consists of successively linearizing any non-linear terms in the dynamics and any non-convex functions in the cost and seeking a solution to the original formulation by solving a sequence of LCPs. This approach leads to two desirable properties, which jointly are instrumental to the design of efficient numerical schemes. First, one can rely on the many efficient techniques and associated software libraries that have been devised to solve LCPs (or LQPs depending on the shape of the original NLP). Second, as we will show in this paper, when this iterative process converges, it returns a strategy that satisfies the PMP related to the original NLP, *i.e.*, a candidate optimum for the original formulation. Unlike existing methods such as [29], which introduce approximated formulations that are still non-linear, our approach offers the main advantage of requiring the solution to LCPs only. Specifically, we identify three key contributions:

1. We introduce and analyze a new framework to compute candidate optimal solutions for finite-horizon, finite-dimensional non-linear stochastic optimal control problems. with control-affine dynamics and uncontrolled diffusion. This hinges on the basic principle of SCP, *i.e.*, iteratively solving a sequence of LCPs that stem from successive linear approximations of the original problem.
2. Through a meticulous study of the continuity of the stochastic Pontryagin cones of variations with respect to linearization, we prove that any accumulation point of the sequence of iterates generated by SCP is a strategy satisfying the PMP related to the original formulation. In addition, by leveraging additional assumptions we prove that any such accumulation point always exists, which in turn provides a “weak” guarantee of success for the method.
3. Through an explicit example, we show how to leverage the properties offered by this framework to better understand what approximations may be adopted for the design of efficient numerical schemes for NLPs, although we leave the theoretical investigation of the approximation error as future direction.

The paper is organized as follows. Section 2 introduces notation and preliminary results and defines the stochastic optimal control problem of interest. In Section 3, we introduce the framework of stochastic SCP and the stochastic PMP, and we state our main result of convergence. For the sake of clarity, in Section 4 we retrace the proof of the stochastic PMP and introduce all the necessary technicalities we need to prove our main result

of convergence, though we recall that the stochastic PMP is a well-established result and Section 4 should not be understood as part of our main contribution. In Section 6, we show how to leverage our analysis to design a practical numerical method to solve non-linear stochastic optimal control problems, and we provide numerical experiments. Finally, Section 7 provides concluding remarks and perspectives on future directions.

## 2. STOCHASTIC OPTIMAL CONTROL SETTING

Let  $(\Omega, \mathcal{G}, P)$  be a second-countable probability space and let  $B_t = (B_t^1, \dots, B_t^d)$  be a  $d$ -dimensional Brownian motion with continuous sample paths starting at zero and whose filtration  $\mathcal{F} \triangleq (\mathcal{F}_t)_{t \geq 0} = (\sigma(B_s : 0 \leq s \leq t))_{t \geq 0}$  is complete. We consider processes that are defined within bounded time intervals. Hence, for every  $n \in \mathbb{N}$ ,  $\ell \geq 2$  and maximal time  $T \in \mathbb{R}_+$ , we introduce the space  $L_{\mathcal{F}}^{\ell}([0, T] \times \Omega; \mathbb{R}^n)$  of progressive processes  $x : [0, T] \times \Omega \rightarrow \mathbb{R}^n$  (with respect to the filtration  $\mathcal{F}$ ) such that  $\mathbb{E} \left[ \int_0^T \|x(s, \omega)\|^{\ell} ds \right] < \infty$ , where  $\|\cdot\|$  is the Euclidean norm. In this setting, for every  $x \in L_{\mathcal{F}}^{\ell}([0, T] \times \Omega; \mathbb{R}^n)$  and  $i = 1, \dots, d$ , the Itô integral of  $x$  with respect to  $B^i$  is the continuous, bounded in  $L^2$ , and  $n$ -dimensional martingale in  $[0, T]$  (with respect to the filtration  $\mathcal{F}$ ) that starts at zero, denoted  $\int_0^{\cdot} x(s) dB_s^i : [0, T] \times \Omega \rightarrow \mathbb{R}^n$ . For  $\ell \geq 2$ , we denote by  $L_{\mathcal{F}}^{\ell}(\Omega; C([0, T]; \mathbb{R}^n))$  the space of  $\mathcal{F}$ -adapted processes  $x : [0, T] \times \Omega \rightarrow \mathbb{R}^n$  that have continuous sample paths and satisfy  $\mathbb{E} \left[ \sup_{s \in [0, T]} \|x(s, \omega)\|^{\ell} \right] < \infty$ .

### 2.1. Stochastic differential equations

From now on, we fix two integers  $n, m \in \mathbb{N}$ , a maximal time  $T \in \mathbb{R}_+$ , and a compact, convex subset  $U \subseteq \mathbb{R}^m$ . Although in this work we consider differential equations steered by deterministic controls (see Sect. 2.2), for the sake of generality, we introduce stochastic dynamics which depend on either deterministic or stochastic controls. Specifically, we denote by  $\mathcal{U}$  the set of admissible controls and consider either deterministic controls  $\mathcal{U} = L^2([0, T]; U)$  or stochastic controls  $\mathcal{U} = L_{\mathcal{F}}^2([0, T] \times \Omega; U)$ . Note that since  $U$  is compact, admissible controls are almost everywhere, or a.e. for brevity (and additionally almost surely, or a.s. for brevity) bounded. We are given continuous mappings  $b_i : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $i = 0, \dots, m$  and  $\sigma_j : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $j = 1, \dots, d$  which are at least  $C^2$  with respect to the variable  $x$ . For a given  $u \in \mathcal{U}$ , we consider dynamical systems modeled through the following forward stochastic differential equation with uncontrolled diffusion

$$\begin{aligned} dx(t) &= b(t, u(t), x(t)) dt + \sigma(t, x(t)) dB_t, \quad x(0) = x^0 \\ &\triangleq \left( b_0(t, x(t)) + \sum_{i=1}^m u^i(t) b_i(t, x(t)) \right) dt + \sum_{j=1}^d \sigma_j(t, x(t)) dB_t^j \end{aligned} \quad (2.1)$$

where we assume that the fixed initial condition satisfies  $x^0 \in L_{\mathcal{F}_0}^{\ell}(\Omega; \mathbb{R}^n)$ , for every  $\ell \geq 2$  (for instance, this holds when  $x^0$  is a deterministic vector of  $\mathbb{R}^n$ ).

The procedure developed in this work is based on the following linearization of (2.1). For  $\ell \geq 2$ , let  $v \in \mathcal{U}$  and  $y \in L_{\mathcal{F}}^{\ell}(\Omega; C([0, T]; \mathbb{R}^n))$ . For a given  $u \in \mathcal{U}$ , we define the linearization of (2.1) around  $(v, y)$  to be the following well-defined forward stochastic differential equation with uncontrolled diffusion

$$\begin{aligned} dx(t) &= b_{v,y}(t, u(t), x(t)) dt + \sigma_y(t, x(t)) dB_t, \quad x(0) = x^0 \\ &\triangleq \left( b_0(t, y(t)) + \frac{\partial b_0}{\partial x}(t, y(t))(x(t) - y(t)) + \sum_{i=1}^m \left( u^i(t) b_i(t, y(t)) + v^i(t) \frac{\partial b_i}{\partial x}(t, y(t))(x(t) - y(t)) \right) \right) dt \\ &+ \sum_{j=1}^d \left( \sigma_j(t, y(t)) + \frac{\partial \sigma_j}{\partial x}(t, y(t))(x(t) - y(t)) \right) dB_t^j. \end{aligned} \quad (2.2)$$

We require the solutions to (2.1) and to (2.2) to be bounded in expectation uniformly with respect to  $u$ ,  $v$  and  $y$ . For this, we consider the following (standard) assumption:

(A<sub>1</sub>) Functions  $b_i$ ,  $i = 0, \dots, m$ ,  $\sigma_j$ ,  $j = 1, \dots, d$ , have compact supports in  $[0, T] \times \mathbb{R}^n$ .

Under (A<sub>1</sub>), for every  $\ell \geq 2$  and every  $u \in \mathcal{U}$ , the stochastic equation (2.1) has a unique (up to stochastic indistinguishability) solution  $x_u \in L_{\mathcal{F}}^{\ell}(\Omega; C([0, T]; \mathbb{R}^n))$ , whereas for every  $u, v \in \mathcal{U}$  and  $y \in L_{\mathcal{F}}^{\ell}(\Omega; C([0, T]; \mathbb{R}^n))$ , the stochastic equation (2.2) has a unique (up to stochastic indistinguishability) solution  $x_{u,v,y} \in L_{\mathcal{F}}^{\ell}(\Omega; C([0, T]; \mathbb{R}^n))$  (see, e.g., [17, 32]), and the following technical result holds, which will be used in the proof of our convergence result, with proof given in the appendix (see Appendix A.1).

**Lemma 2.1** (Uniform boundness and continuity with respect to control inputs). *Fix  $\ell \geq 2$ , and for  $u \in \mathcal{U}$ , let  $x_u$  denote the solution to (2.1), whereas for  $u, v \in \mathcal{U}$  and  $y \in L_{\mathcal{F}}^{\ell}(\Omega; C([0, T]; \mathbb{R}^n))$ , let  $x_{u,v,y}$  denote the solution to (2.2). Under (A<sub>1</sub>), there exists a constant  $C \geq 0$  which does not depend on  $u$ ,  $v$ , or  $y$  such that it holds that*

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \|x_u(t)\|^{\ell} + \sup_{t \in [0, T]} \|x_{u,v,y}(t)\|^{\ell} \right] \leq C,$$

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \|x_{u_1}(t) - x_{u_2}(t)\|^{\ell} + \sup_{t \in [0, T]} \|x_{u_1,v,y}(t) - x_{u_2,v,y}(t)\|^{\ell} \right] \leq C \mathbb{E} \left[ \left( \int_0^T \|u_1(s) - u_2(s)\| ds \right)^{\ell} \right], u_1, u_2 \in \mathcal{U}.$$

## 2.2. Stochastic optimal control problem

Given  $q \in \mathbb{N}$ , we consider continuous mappings  $g : \mathbb{R}^n \rightarrow \mathbb{R}^q$ ,  $G : \mathbb{R}^m \rightarrow \mathbb{R}$ ,  $H : \mathbb{R}^n \rightarrow \mathbb{R}$ , and  $L : \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$  with  $L(t, x, u) \triangleq L_0(t, x) + \sum_{i=1}^m u^i L_i(t, x)$ . We require  $g$ ,  $H$ , and  $L_i$ ,  $i = 0, \dots, m$ , to be at least  $C^2$  with respect to the variable  $x$ , and we require  $G$ ,  $H$  to be convex. In particular,  $G$  is Lipschitz when restricted to the compact and convex set  $U$ . We focus on finite-horizon, finite-dimensional non-linear stochastic Optimal Control Problems (OCP) with control-affine dynamics and uncontrolled diffusion, of the form

$$\begin{cases} \min_{u \in \mathcal{U}} \mathbb{E} \left[ \int_0^T f^0(s, u(s), x(s)) ds \right] \triangleq \mathbb{E} \left[ \int_0^T \left( G(u(s)) + H(x(s)) + L(s, u(s), x(s)) \right) ds \right] \\ dx(t) = b(t, u(t), x(t)) dt + \sigma(t, x(t)) dB_t, \quad x(0) = x^0, \quad \mathbb{E}[g(x(T))] = 0 \end{cases}$$

where we optimize over deterministic controls  $u \in \mathcal{U} = L^2([0, T]; U)$ . We adopt the (fairly mild) assumption: (A<sub>2</sub>) Mappings  $g$ ,  $H$ , and  $L_i$ ,  $i = 0, \dots, m$ , either are affine-in-state or have compact supports in  $\mathbb{R}^n$  and in  $[0, T] \times \mathbb{R}^n$ , respectively.

Our choice of optimizing over deterministic controls as opposed to stochastic controls is motivated by practical considerations. Specifically, in several applications of interest ranging from aerospace to robotics, it is often advantageous to compute and implement simpler deterministic controls at higher control rates, to be able to quickly react to external disturbances, unmodeled dynamical effects, and changes in the cost function (e.g., moving obstacles that a robot should avoid in real-time). Moreover, in cases where a feedback controller is accounted for, a common and efficient approach entails decomposing the stochastic control into a state-dependent feedback term, plus a nominal deterministic control trajectory to be optimized for, which is equivalent to adopting deterministic controls (see, e.g., [26, 34]). Nevertheless, for the sake of completeness and generality, in Section 5 we introduce appropriate conditions under which our method may extend to stochastic controls; we also analyze possible extensions to the case of free-final-time optimal control problems.

Many applications of interest often involve state constraints. In this case, to make sure the procedure developed in this work still applies, every such constraint needs to be considered in expectation and penalized within the cost of OCP (for example by including those contributions in  $H$  or  $L$  through some penalization function).

In future work, we plan to extend our method to more general settings, whereby for instance state constraints are enforced through more accurate chance constraints (see Sect. 7 for a thorough discussion).

### 3. STOCHASTIC SEQUENTIAL CONVEX PROGRAMMING

We propose the following framework to solve OCP, based on the classical SCP methodology. Starting from some initial guesses of control  $u_0 \in \mathcal{U}$  and trajectory  $x_0 \in L_{\mathcal{F}}^{\ell}(\Omega; C([0, T]; \mathbb{R}^n))$ ,  $\ell \geq 2$ , we inductively define a sequence of stochastic linear-convex problems whose dynamics and costs stem from successive linearizations of the mappings  $b$ ,  $\sigma$ , and  $L$ , and we successively solve those problems while updating user-defined parameters. The convergence of the method generally depends on a good choice of the initial guess  $(u_0, x_0)$  and of updates in each iteration to *trust-region constraints*. These constraints are added to make the successive linearizations of OCP well-posed. Below, we detail this procedure.

#### 3.1. The method

At iteration  $k + 1 \in \mathbb{N}$ , by denoting

$$f_{v,y}^0(s, u, x) \triangleq G(u) + H(x) + L(s, u, y) + \frac{\partial L}{\partial x}(s, v, y)(x - y), \quad (3.1)$$

we define the following stochastic Linearized Optimal Control Problem ( $\text{LOCP}_{k+1}^{\Delta}$ )

$$\left\{ \begin{array}{l} \min_{u \in \mathcal{U}} \mathbb{E} \left[ \int_0^T f_{k+1}^0(s, u(s), x(s)) \, ds \right] \triangleq \mathbb{E} \left[ \int_0^T f_{u_k, x_k}^0(s, u(s), x(s)) \, ds \right] \\ dx(t) = b_{k+1}(t, u(t), x(t)) \, dt + \sigma_{k+1}(t, x(t)) \, dB_t, \quad x(0) = x^0 \\ \quad \triangleq b_{u_k, x_k}(t, u(t), x(t)) \, dt + \sigma_{x_k}(t, x(t)) \, dB_t \\ \mathbb{E} [g_{k+1}(x(T))] \triangleq \mathbb{E} \left[ g(x_k(T)) + \frac{\partial g}{\partial x}(x_k(T))(x(T) - x_k(T)) \right] = 0 \\ \int_0^T \mathbb{E} [\|x(s) - x_k(s)\|^2] \, ds \leq \Delta_{k+1} \end{array} \right.$$

where we optimize over deterministic controls  $u \in \mathcal{U} = L^2([0, T]; U)$ . The tuple  $(u_k, x_k) \in \mathcal{U} \times L_{\mathcal{F}}^{\ell}(\Omega; C([0, T]; \mathbb{R}^n))$  is defined inductively and denotes a solution to  $(\text{LOCP}_k^{\Delta})$ .

Each problem  $\text{LOCP}_{k+1}^{\Delta}$  consists of linearizing OCP around the *solution at the previous iteration*  $(u_k, x_k)$ , starting from  $(u_0, x_0)$ . To avoid misguidance due to high linearization error, we must restrict the search for optimal solutions for  $\text{LOCP}_{k+1}^{\Delta}$  to neighborhoods of  $(u_k, x_k)$ . This is achieved by the final constraints listed in  $\text{LOCP}_{k+1}^{\Delta}$ , referred to as trust-region constraints, where the constant  $\Delta_{k+1} \geq 0$  is the trust-region radius. No such constraint is enforced on controls, since those appear linearly in  $b$ ,  $\sigma$ , and  $L$ . To ensure well-posedness of the sequence  $(\text{LOCP}_k^{\Delta})_{k \in \mathbb{N}}$ , we require  $\text{LOCP}_{k+1}^{\Delta}$  has a solution, for which we consider the following assumption:

(A<sub>3</sub>) For every  $k \in \mathbb{N}$ , problem  $\text{LOCP}_{k+1}^{\Delta}$  is feasible.

**Proposition 3.1** (Existence of solutions of  $\text{LOCP}_{k+1}^{\Delta}$ ). *Under (A<sub>3</sub>),  $\text{LOCP}_{k+1}^{\Delta}$  has a solution for every  $k \in \mathbb{N}$ .*

*Proof.* The proof is based on the argument developed to prove ([54], Thm. 5.2, Chap. 2). Specifically, let  $k \in \mathbb{N}$  and  $\{(u_{k+1}^{\alpha}, x_{k+1}^{\alpha})\}_{\alpha \in \mathbb{N}} \in \mathcal{U} \times L_{\mathcal{F}}^{\ell}(\Omega; C([0, T]; \mathbb{R}^n))$  be a minimizing sequence for  $\text{LOCP}_{k+1}^{\Delta}$ .

The sequence  $\{u_{k+1}^{\alpha}\}_{\alpha \in \mathbb{N}}$  is uniformly bounded in  $L^2([0, T]; \mathbb{R}^m)$ , and thus we may assume the existence of  $\tilde{u}_{k+1} \in L^2([0, T]; \mathbb{R}^m)$  such that  $\{u_{k+1}^{\alpha}\}_{\alpha \in \mathbb{N}}$  converges, up to some subsequence, to  $\tilde{u}_{k+1}$  for the weak topology

of  $L^2([0, T]; \mathbb{R}^m)$ . Moreover, the compactness and convexity of  $U$  yield  $\tilde{u}_{k+1} \in \mathcal{U}$ . Finally, by Mazur's theorem there exists a sequence of convex combinations

$$\tilde{u}_{k+1}^\alpha \triangleq \sum_{\beta \geq 1} c_{\alpha, \beta}^{k+1} u_{k+1}^{\alpha+\beta}, \quad c_{\alpha, \beta}^{k+1} \geq 0, \quad \sum_{\beta \geq 1} c_{\alpha, \beta}^{k+1} = 1,$$

such that  $\{\tilde{u}_{k+1}^\alpha\}_{\alpha \in \mathbb{N}}$  converges to  $\tilde{u}_{k+1}$  for the strong topology of  $L^2([0, T]; \mathbb{R}^m)$  (although, since controls are deterministic and the diffusion is uncontrolled, weak convergence of controls would suffice in our setting).

Thanks to Lemma 2.1, the sequence of trajectories  $\{x_{\tilde{u}_{k+1}^\alpha, u_k, x_k}\}_{\alpha \in \mathbb{N}} \subseteq L_{\mathcal{F}}^\ell(\Omega; C([0, T]; \mathbb{R}^n))$  converges to the trajectory  $\tilde{x}_{k+1} \triangleq x_{\tilde{u}_{k+1}, u_k, x_k} \in L_{\mathcal{F}}^\ell(\Omega; C([0, T]; \mathbb{R}^n))$  for the strong topology of  $L_{\mathcal{F}}^\ell(\Omega; C([0, T]; \mathbb{R}^n))$ . At this step, for every  $\alpha \in \mathbb{N}$ , thanks to the linearity of (2.2) we obtain that

$$x_{\tilde{u}_{k+1}^\alpha, u_k, x_k} \equiv \sum_{\beta \geq 1} c_{\alpha, \beta}^{k+1} x_{u_{k+1}^{\alpha+\beta}, u_k, x_k} = \sum_{\beta \geq 1} c_{\alpha, \beta}^{k+1} x_{k+1}^{\alpha+\beta},$$

and therefore the linearity of the function  $g_{k+1}$  yields

$$\mathbb{E} \left[ g_{k+1}(x_{\tilde{u}_{k+1}^\alpha, u_k, x_k}(T)) \right] = \sum_{\beta \geq 1} c_{\alpha, \beta}^{k+1} \mathbb{E} \left[ g_{k+1}(x_{k+1}^{\alpha+\beta}(T)) \right] = 0,$$

whereas the convexity of the norm yields

$$\left( \int_0^T \mathbb{E} \left[ \|x_{\tilde{u}_{k+1}^\alpha, u_k, x_k}(s) - x_k(s)\|^2 \right] ds \right)^{\frac{1}{2}} \leq \sum_{\beta \geq 1} c_{\alpha, \beta}^{k+1} \left( \int_0^T \mathbb{E} \left[ \|x_{k+1}^{\alpha+\beta}(s) - x_k(s)\|^2 \right] ds \right)^{\frac{1}{2}} \leq \Delta_{k+1}^{\frac{1}{2}}.$$

In turn, passing to the limit for  $\alpha \rightarrow \infty$ , we infer that the tuple  $(\tilde{u}_{k+1}, \tilde{x}_{k+1}) \in \mathcal{U} \times L_{\mathcal{F}}^\ell(\Omega; C([0, T]; \mathbb{R}^n))$  is admissible for  $\text{LOCP}_{k+1}^\Delta$ . In addition, the convexity of the function  $f_{k+1}^0$  allows us to compute

$$\begin{aligned} \mathbb{E} \left[ \int_0^T f_{k+1}^0(s, \tilde{u}_{k+1}(s), \tilde{x}_{k+1}(s)) ds \right] &\leq \lim_{\alpha \rightarrow \infty} \sum_{\beta \geq 1} c_{\alpha, \beta}^{k+1} \mathbb{E} \left[ \int_0^T f_{k+1}^0(s, u_{k+1}^{\alpha+\beta}(s), x_{k+1}^{\alpha+\beta}(s)) ds \right] \\ &= \min_{u \in \mathcal{U}} \mathbb{E} \left[ \int_0^T f_{k+1}^0(s, u(s), x(s)) ds \right], \end{aligned}$$

from which we conclude that  $(\tilde{u}_{k+1}, \tilde{x}_{k+1}) \in \mathcal{U} \times L_{\mathcal{F}}^\ell(\Omega; C([0, T]; \mathbb{R}^n))$  is a solution of  $\text{LOCP}_{k+1}^\Delta$ .  $\square$

**Remark 3.2.** Although one finds empirically that  $(A_3)$  is often satisfied in practice (see for instance results in Sect. 6), it is generally difficult to derive sufficient conditions for this assumption to hold true a priori. In particular, to the best of our knowledge, only stochastic dynamics with constant coefficients and with specific regularity conditions on the stochastic diffusion yield controllability [52]. In deterministic settings, some works obtain feasibility of each convex subproblem, though only locally around the unknown solution of the original formulation by assuming second-order regularity conditions, which however can not be checked a priori [20, 21]. Thus, even SCP-based schemes in simpler deterministic settings must often either explicitly assume the feasibility of each convex subproblem, or must modify the linearized dynamics by infusing additional slack controls to force feasibility [37]. Motivated by these remarks, we leave the investigation for sufficient conditions for the validity of  $(A_3)$  as a direction of future research, which are out of the scope of this work which focuses on the properties

of accumulations points of the sequence generated by SCP. Note that  $(A_3)$  is satisfied when  $g = 0$ , *i.e.*, no final constraints are imposed: a problem that remains generally relevant though computationally difficult to solve.

Under assumptions  $(A_1)$ – $(A_3)$ , the method consists of iteratively solving the aforementioned linearized problems through the update of the sequence of trust-region radii, producing a sequence of tuples  $(u_k, x_k)_{k \in \mathbb{N}}$  such that for each  $k \in \mathbb{N}$ ,  $(u_{k+1}, x_{k+1})$  solves  $\text{LOCP}_{k+1}^\Delta$ . The user may often steer this procedure to convergence (with respect to appropriate topologies) by adequately selecting an initial guess  $(u_0, x_0)$  and an update rule for  $(\Delta_k)_{k \in \mathbb{N}}$ <sup>1</sup>, and appropriate choices will be described in Section 6. Assuming that an accumulation point for  $(u_k, x_k)_{k \in \mathbb{N}}$  can be found (whose existence will be discussed shortly), our objective consists of proving that this is a candidate locally-optimal solution to the original formulation OCP. Specifically, we show that any accumulation point for  $(u_k, x_k)_{k \in \mathbb{N}}$  satisfies the stochastic PMP related to OCP. To develop such analysis, we require the absence of state constraints, and in particular of trust-region constraints.

### 3.2. Stochastic Pontryagin Maximum Principle

In this section, we recall the statement of the PMP which provides classical first-order necessary conditions for optimality, upon which we will establish our main result. For the sake of clarity, we introduce the PMP related to OCP and the PMP related to each convexified problem  $\text{LOCP}_k^\Delta$  separately.

#### 3.2.1. PMP related to OCP

For every  $p \in \mathbb{R}^n$ ,  $p^0 \in \mathbb{R}$ , and  $q = (q_1, \dots, q_d) \in \mathbb{R}^{n \times d}$  define the *Hamiltonian* (same notation as in (2.1))

$$H(s, u, x, p, p^0, q) \triangleq p^\top b(s, x, u) + p^0 f^0(s, x, u) + \sum_{i=1}^d q_i^\top \sigma_i(s, x).$$

**Theorem 3.3** (Stochastic Pontryagin Maximum Principle for OCP [54]). *Let  $(u, x)$  be a locally-optimal solution to OCP. There exist  $p \in L_{\mathcal{F}}^2(\Omega; C([0, T]; \mathbb{R}^n))$ , a tuple  $(\mathbf{p}, p^0)$ , where  $\mathbf{p} \in \mathbb{R}^d$  and  $p^0 \leq 0$  are constant, and  $q = (q_1, \dots, q_d) \in L_{\mathcal{F}}^2([0, T] \times \Omega; \mathbb{R}^{n \times d})$  such that the following relations are satisfied:*

1. *Non-Triviality Condition:*  $(\mathbf{p}, p^0) \neq 0$ .
2. *Adjoint Equation:*

$$dp(t) = -\frac{\partial H}{\partial x}(t, u(t), x(t), p(t), p^0, q(t)) dt + q(t) dB_t, \quad p(T) = \mathbb{E} \left[ \frac{\partial g}{\partial x}(x(T)) \right]^\top \mathbf{p} \in \mathbb{R}^n.$$

3. *Maximality Condition:*

$$u(t) = \arg \max_{v \in U} \mathbb{E} \left[ H(t, v, x(t), p(t), p^0, q(t)) \right], \quad \text{a.e.}$$

The quantity  $(u, \mathbf{p}, p^0)$  uniquely determines  $x$ ,  $p$ , and  $q$  and is called *extremal* for OCP (associated with the tuple  $(u, x, p, \mathbf{p}, p^0, q)$ , or simply with  $(u, x)$ ). An extremal  $(u, \mathbf{p}, p^0)$  is called *normal* if  $p^0 \neq 0$ .

Albeit final conditions are specified instead of initial conditions for  $p$ , it turns out that processes satisfying backward stochastic differential equations are adapted with respect to the filtration  $\mathcal{F}$  (see, *e.g.*, [17, 54]), which makes the adjoint equation well-posed. Although conditions for optimality for stochastic optimal control problems are usually developed when considering stochastic controls only, the proof of Theorem 3.3 (and of Thm. 3.4) readily follows from classical arguments (*e.g.*, see [54], Chap. 3.6). Nevertheless, to prove our main

<sup>1</sup>When state constraint penalization is adopted, SCP procedures may also consider update rules for *penalization weights*, and those must be provided together with update rules for the trust-region radii. Further details can be found in [9, 40].



result, we rely on a proof of Theorem 3.3 (and of Thm. 3.4) which stems from implicit-function-theorem-type results (see Sects. 4.1 and 4.2). Thus, in Section 4.1, we provide a new proof of Theorem 3.3 (more precisely, of Thm. 3.4) which follows from the original idea developed by Pontryagin and his group (see, *e.g.*, [1, 23, 43]), though the latter result should not be understood as part of our main contribution.

### 3.2.2. PMP related to $LOCP_k^\Delta$

For every  $k \geq 1$ ,  $p \in \mathbb{R}^n$ ,  $p^0, p^1 \in \mathbb{R}$ , and  $q = (q_1, \dots, q_d) \in \mathbb{R}^{n \times d}$ , define the *Hamiltonian* (same notation as in (2.2))

$$H_k(s, u, x, p, p^0, p^1, q) \triangleq p^\top b_k(s, x, u) + p^0 f_k^0(s, x, u) + p^1 \|x - x_{k-1}(s)\|^2 + \sum_{i=1}^d q_i^\top (\sigma_k)_i(s, x).$$

**Theorem 3.4** (Weak Stochastic Pontryagin Maximum Principle for  $LOCP_k^\Delta$  [54]). *Let  $(u_k, x_k)$  be a locally-optimal solution to  $LOCP_k^\Delta$ . There exists  $p_k \in L_{\mathcal{F}}^2(\Omega; C([0, T]; \mathbb{R}^n))$ , a tuple  $(\mathbf{p}_k, p_k^0, p_k^1)$ , where  $\mathbf{p}_k \in \mathbb{R}^q$ ,  $p_k^0 \leq 0$ , and  $p_k^1 \in \mathbb{R}$  are constant, and  $q_k = ((q_k)_1, \dots, (q_k)_d) \in L_{\mathcal{F}}^2([0, T] \times \Omega; \mathbb{R}^{n \times d})$  such that the following relations are satisfied:*

1. *Non-Triviality Condition:*  $(\mathbf{p}_k, p_k^0, p_k^1) \neq 0$ .
2. *Adjoint Equation:*

$$dp_k(t) = -\frac{\partial H_k}{\partial x}(t, u_k(t), x_k(t), p_k(t), p_k^0, p_k^1, q_k(t)) dt + q_k(t) dB_t, \quad p_k(T) = \mathbb{E} \left[ \frac{\partial g_k}{\partial x}(x_k(T)) \right]^\top \mathbf{p}_k \in \mathbb{R}^n.$$

3. *Maximality Condition:*

$$u_k(t) = \arg \max_{v \in U} \mathbb{E} \left[ H_k(t, v, x_k(t), p_k(t), p_k^0, p_k^1, q_k(t)) \right], \text{ a.e.}$$

The quantity  $(u_k, \mathbf{p}_k, p_k^0, p_k^1)$  uniquely determines  $x_k$ ,  $p_k$ , and  $q_k$  and is called *extremal* for  $LOCP_k^\Delta$  (associated with  $(u_k, x_k, p_k, \mathbf{p}_k, p_k^0, p_k^1, q_k)$ , or simply with  $(u_k, x_k)$ ). An extremal  $(u_k, \mathbf{p}_k, p_k^0, p_k^1)$  is called *normal* if  $p_k^0 \neq 0$ .

**Remark 3.5.** By introducing the new variable

$$y(t) \triangleq \int_0^t \|x(s) - x_k(s)\|^2 ds, \tag{3.2}$$

the trust-region constraint in  $LOCP_{k+1}^\Delta$  can be rewritten as  $\mathbb{E}[y(T) - \Delta_{k+1}] \leq 0$ . Thus, by leveraging this transformation,  $LOCP_{k+1}^\Delta$  may be reformulated as a standard stochastic optimal control problem with final inequality constraints. Nevertheless, although the conditions listed in Theorem 3.3 are essentially sharp, the statement of Theorem 3.4 may be strengthened as follows. If  $(u_{k+1}, x_{k+1}, \mathbf{p}_{k+1}, p_{k+1}^0, p_{k+1}^1, q_{k+1})$  is an extremal for  $LOCP_{k+1}^\Delta$ , one can additionally prove that  $p_{k+1}^1 \leq 0$  (*e.g.*, see [22]; note that in [22] the multipliers have opposite signs because a different convention is adopted) and

$$p_{k+1}^1 \mathbb{E} \left[ \int_0^T \|x_{k+1}(s) - x_k(s)\|^2 ds - \Delta_{k+1} \right] = 0$$

(the latter is known as *slack condition*), which motivates the choice “weak stochastic PMP for  $\text{LOCP}_{k+1}^\Delta$ ” as name for Theorem 3.4. Nevertheless, since the trust-region constraints do not appear in the original problem, we do not need to leverage these latter additional conditions on  $p_{k+1}^1$  to prove our claims, *i.e.*, Theorem 3.4 suffices to establish the aforementioned properties of accumulation points for SCP when applied to solve OCP.

### 3.3. Main results

Our contribution is twofold. First, under very mild assumptions, we prove that any accumulation point of the sequence of iterates generated by SCP is a candidate locally-optimal solution for OCP. Specifically, we prove that any accumulation point of the sequence  $(u_k, x_k, \mathbf{p}_k, p_k^0, p_k^1, q_k)_{k \in \mathbb{N}}$  generated by SCP, where  $(u_k, \mathbf{p}_k, p_k^0, p_k^1)$  is an extremal for  $\text{LOCP}_k^\Delta$  associated with  $(u_k, x_k)$  in the sense of Theorem 3.3, is an extremal for OCP in the sense of Theorem 3.4 (see Thm. 3.6). Although the optimization community is aware that establishing the convergence of the sequence generated by SCP is generally difficult (see, *e.g.*, [36, 39]), one can often prove optimality-related properties of its accumulation points, a much natural property which justifies the use of SCP (see, *e.g.*, [36]). Second, by strengthening our original assumptions, we prove existence of at least one accumulation point of the aforementioned sequence  $(u_k, x_k, \mathbf{p}_k, p_k^0, p_k^1, q_k)_{k \in \mathbb{N}}$  generated by SCP (see Thm. 3.9). Below, we organize these claims in two more precise statements.

#### 3.3.1. Properties of accumulation points for SCP

**Theorem 3.6** (Properties of Accumulation Points for SCP). *Assume that  $(A_1)$ – $(A_3)$  hold and that SCP generates a sequence  $(\Delta_k, u_k, x_k)_{k \in \mathbb{N}}$  such that  $(\Delta_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}_+ \setminus \{0\}$  converges to zero, and for every  $k \geq 1$ , the tuple  $(u_k, x_k)$  locally solves  $\text{LOCP}_k^\Delta$ . For every  $k \geq 1$ , letting  $(u_k, \mathbf{p}_k, p_k^0, p_k^1)$  be an extremal associated with  $(u_k, x_k)$  for  $\text{LOCP}_k^\Delta$  (whose existence is ensured by Thm. 3.4), assume the following Accumulation Condition holds:*

(AC) *Up to some subsequence,  $(u_k, \mathbf{p}_k, p_k^0, p_k^1)$  converges to some  $(u, \mathbf{p}, p^0, p^1) \in L^2([0, T]; \mathbb{R}^m) \times \mathbb{R}^{q+2}$  for the weak topology of  $L^2([0, T]; \mathbb{R}^m) \times \mathbb{R}^{q+2}$ .*

*If  $(\mathbf{p}, p^0) \neq 0$ , then  $(u, \mathbf{p}, p^0)$  is an extremal for OCP associated with  $(u, x_u)$ .*

The guarantees offered by Theorem 3.6 read as follows. Under  $(A_1)$ – $(A_3)$  and by selecting a *shrinking-to-zero* sequence of trust-region radii, if iteratively solving problems  $\text{LOCP}_k^\Delta$  returns a sequence of strategies whose extremals satisfy (AC) with a non-trivial multiplier  $(\mathbf{p}, p^0) \neq 0$ , then SCP finds a candidate (local) solution to OCP. Theorem 3.6 extends classical results on the well-posedness of SCP (see, *e.g.*, [36]) from deterministic to stochastic settings. In particular, the requirement  $(\mathbf{p}, p^0) \neq 0$  in Theorem 3.6 is natural and has an equivalent in deterministic settings, playing the role of some sort of *qualification condition* (see, *e.g.*, [36], Thm. 3.4). Note that the requirement  $(\mathbf{p}, p^0) \neq 0$  can be easily numerically checked (see our discussion after Thm. 3.9).

#### 3.3.2. Existence of accumulation points for SCP

Assumptions  $(A_1)$ – $(A_3)$  together with some minor requirements are sufficient to establish that any accumulation point for the sequence of iterates  $(u_k, x_k)_{k \in \mathbb{N}}$  satisfies the stochastic PMP related to OCP (Thm. 3.6). We can additionally infer the existence of accumulation points, *i.e.*, (AC) in Theorem 3.6 holds true, if some more structure on the data defining OCP is assumed. Importantly, the validity of (AC) endows stochastic SCP with the additional guarantee that at least one accumulation point (with respect to weak topologies) exists, which in turn provides a “weak” guarantee of success for the method *via* the result of Theorem 3.6 (“weak” because convergence is satisfied up to some subsequence). We introduce the following technical condition:

(A<sub>4</sub>) The mapping  $G : \mathbb{R}^{m+1} \rightarrow \mathbb{R}$  is given by  $G(t, u) = u^\top \mathbb{U}(t)u$ , where each  $\mathbb{U}(t) \in \mathbb{R}^{m \times m}$  is symmetric definite positive and the mapping  $t \mapsto \mathbb{U}(t)^{-1}$  is continuous.

The use we make of  $(A_4)$  is essentially contained in the following result.

**Lemma 3.7.** *Under  $(A_4)$ , for every  $k \in \mathbb{N}$ , every normal extremal  $(u_{k+1}, \mathbf{p}_{k+1}, p_{k+1}^0, p_{k+1}^1)$  for  $\text{LOCP}_{k+1}^\Delta$ , i.e., for which  $p_{k+1}^0 \neq 0$ , is such that the corresponding control  $u_{k+1}$  is time-continuous.*

*Proof.* Fix  $k \in \mathbb{N}$ , and let  $(u_{k+1}, \mathbf{p}_{k+1}, p_{k+1}^0, p_{k+1}^1)$  be a normal extremal for  $\text{LOCP}_{k+1}^\Delta$ . Due to  $p_{k+1}^0 \neq 0$ , the convexity of  $U \subseteq \mathbb{R}^m$  and the maximality condition in Theorem 3.4 yield

$$u_{k+1}(s) = \begin{cases} \frac{1}{2p_{k+1}^0} \mathbb{U}^{-1}(s) \gamma_k(s) & \text{if } \frac{1}{2p_{k+1}^0} \mathbb{U}^{-1}(s) \gamma_k(s) \in U, \\ \text{Proj}_U \left( \frac{1}{2p_{k+1}^0} \mathbb{U}^{-1}(s) \gamma_k(s) \right) & \text{if } \frac{1}{2p_{k+1}^0} \mathbb{U}^{-1}(s) \gamma_k(s) \notin U, \end{cases}$$

where we denote  $\gamma_k(s) \triangleq ((\gamma_k(s))_1, \dots, (\gamma_k(s))_m)$  with  $(\gamma_k(s))_i \triangleq \mathbb{E}[p_{k+1}^\top(s) b_i(s, x_k(s)) + p_{k+1}^0 L_i(s, x_k(s))]$ , whereas  $\text{Proj}_U : \mathbb{R}^m \rightarrow \mathbb{R}^m$  denotes the projection over the convex set  $U$ . The claim readily follows once we prove the mappings  $t \in [0, T] \mapsto \gamma_k(s)$  are continuous, given that  $t \in [0, T] \mapsto \mathbb{U}^{-1}(s)$  and  $v \mapsto \text{Proj}_U(v)$  are continuous. We only prove the continuity of  $t \in [0, T] \mapsto \mathbb{E}[p_k(t)]$ , given that the continuity of  $t \in [0, T] \mapsto \gamma_k(s)$  can be proved by leveraging similar arguments, and  $(A_1)$  and  $(A_2)$ . For this, thanks to  $(A_1)$ ,  $(A_2)$ , Lemma 2.1, Theorem 3.4, and Hölder and Burkholder–Davis–Gundy inequalities, for every  $0 \leq s < t \leq T$ , we obtain that

$$\begin{aligned} & \mathbb{E}[\|p_k(t) - p_k(s)\|] \\ & \leq 2\mathbb{E} \left[ \left\| \int_s^t \left( p_k^\top(r) \frac{\partial f_k}{\partial x}(r, x_k(r), u_k(r)) + p_k^0 \frac{\partial f_k^0}{\partial x}(r, x_k(r), u_k(r)) + 2p_k^1(x_k(r) - x_{k-1}(r)) \right. \right. \right. \\ & \quad \left. \left. \left. + \sum_{i=1}^d (q_k)_i^\top(r) \frac{\partial (\sigma_k)_i}{\partial x}(r, x_k(r)) \right) dr \right\| \right] + 2\mathbb{E} \left[ \left\| \int_s^t q_k(r) dB_r \right\| \right] \\ & \leq C \left( p_k^0 + p_k^1 + \mathbb{E} \left[ \sup_{t \in [0, T]} \|x_k(t)\|^2 \right]^{\frac{1}{2}} + \mathbb{E} \left[ \sup_{t \in [0, T]} \|p_k(t)\|^2 \right]^{\frac{1}{2}} + \mathbb{E} \left[ \int_0^T \|q_k(t)\|^2 dt \right]^{\frac{1}{2}} \right) (t-s)^{\frac{1}{2}}, \end{aligned}$$

for some appropriate constant  $C \geq 0$ , and the conclusion follows.  $\square$

**Remark 3.8.** Through Lemma 3.7,  $(A_4)$  becomes crucial to ensure the validity of (AC) in Theorem 3.6. In particular, the works [10, 12, 13, 25], which analyze continuity properties of extremals with respect to appropriate deformations of some deterministic optimal control problems and which inspired our work, show that the time-continuity of optimal controls  $u_k$  for each  $\text{LOCP}_k^\Delta$  represents a requirement which is not easy to relax (in particular, see the counterexample in [13], Sect. 2.3), especially in the presence of trust-region constraints. Importantly, motivated by regularity results in deterministic optimal control settings (see, e.g., [48], Thm. 3.2), we reckon that more generic mappings  $G$  might yield Lemma 3.7, especially when optimizing over deterministic controls, although we leave the investigation of more general sufficient conditions for the time-continuity of optimal controls  $u_k$  for each  $\text{LOCP}_k^\Delta$  as a future research direction, in that it is out of the scope of this work which again focuses on the properties of accumulation points of the sequence generated by SCP.

**Theorem 3.9** (Existence of Accumulation Points for SCP). *Assume that  $(A_1)$ – $(A_4)$  hold and that SCP generates a sequence  $(\Delta_k, u_k, x_k)_{k \in \mathbb{N}}$  such that  $(\Delta_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}_+ \setminus \{0\}$  converges to zero, and for every  $k \geq 1$ , the tuple  $(u_k, x_k)$  locally solves  $\text{LOCP}_k^\Delta$ . For every  $k \geq 1$ , let  $(u_k, \mathbf{p}_k, p_k^0, p_k^1)$  be an extremal associated with  $(u_k, x_k)$  for  $\text{LOCP}_k^\Delta$  (whose existence is ensured by Thm. 3.4). If  $p_k^0 \neq 0$  for every  $k \geq 1$ , then (AC) in Theorem 3.6 holds true.*

In addition, if  $(u, \mathbf{p}, p^0)$  denotes the extremal for OCP associated with some  $(u, x, p, \mathbf{p}, p^0, q)$ , which is provided by Theorem 3.6, the following convergence holds, up to some subsequence, for every  $\ell \geq 2$  when  $k \rightarrow \infty$ :

$$\|(\mathbf{p}_k, p_k^0) - (\mathbf{p}, p^0)\| + \mathbb{E} \left[ \sup_{s \in [0, T]} \|x_k(s) - x(s)\|^\ell + \sup_{s \in [0, T]} \|p_k(s) - p(s)\|^2 + \int_0^T \|q_k(s) - q(s)\|^2 ds \right] \rightarrow 0. \quad (3.3)$$

The main takeaway of Theorem 3.9 is that an accumulation point for stochastic SCP always exists, which in turn is a candidate (local) solution to OCP due to Theorem 3.6, as soon as appropriate qualification-type conditions are satisfied. In particular, similarly to the condition  $(\mathbf{p}, p^0) \neq 0$  in Theorem 3.6, the requirement  $p_k^0 \neq 0$  in Theorem 3.9 plays the role of an additional qualification condition, but at the level of each subproblem  $\text{LOCP}_k^\Delta$ . Note that the latter is a *generic* property in deterministic optimal control (see, e.g., [18]). Finally, the requirement  $(\mathbf{p}, p^0) \neq 0$  in Theorem 3.6 from which (local) optimality of the strategy found by SCP stems can be numerically checked thanks to (3.3), as soon as the multipliers  $(\mathbf{p}_k, p_k^0)$  are accessible through SCP iterations.

### 3.3.3. Insights to speed up the convergence of SCP

In Theorem 3.9, there are also insightful statements concerning the convergence of Pontryagin extremals. Let us outline how those statements may be leveraged to speed up convergence.

For this, adopt the notation of Theorem 3.3 and assume that we are in the situation where applying the maximality condition of the PMP to problem OCP leads to smooth candidate optimal controls, as functions of the variables  $x$ ,  $p^0$ , and  $p$  (be aware that this might not be straightforward to obtain). We are then in a position to define two-point boundary value problems to solve OCP, also known as *shooting methods*, for which the decision variables become  $p^0$ ,  $p(T)$ , and  $q$ . In particular, the core of the method consists of iteratively choosing  $(p^0, p(T), q)$  and making the adjoint equation evolve until some given final condition is met (see, e.g., [7, 16] for a more detailed explanation of shooting methods). In the context of deterministic optimal control, when convergence is achieved, shooting methods terminate quite fast (at least quadratically). However, here the bottlenecks are: (1) to deal with the presence of the variable  $q$  and (2) to find a good guess for the initial value of  $(p^0, p(T), q)$  to make the whole procedure converge. In the setting of Theorem 3.9, a valid option to design well-posed shooting methods is as follows. With the notation and assumptions of Theorem 3.9, up to some subsequence it holds that  $(p_k^0, p_k(T), q_k) \rightarrow (p^0, p(T), q)$  (with respect to appropriate topologies) as  $k \rightarrow \infty$ . Therefore, assuming we have access to Pontryagin extremals along iterations and given some large enough iteration  $k$ , we can fix  $q = q_k$  and initialize with  $(p_k^0, p_k(T))$  a shooting method for OCP that operates on the finite-dimensional variable  $(p^0, p(T)) \in \mathbb{R}^{q+1}$ . If successful, this strategy would speed up the convergence of the entire numerical scheme, though we leave its investigation as a future research direction.

## 4. PROOF OF THE MAIN RESULTS

Since the statement of Theorem 3.6 is in particular contained in Theorem 3.9, it is sufficient to prove Theorem 3.9 only. We split this proof into three main steps. First, we retrace the proof of the stochastic PMP to introduce necessary notation and expressions. In addition, we leverage this step to provide novel insight on how to prove the stochastic PMP by following the lines of the original work of Pontryagin and his group (see, e.g., [1, 23, 43]), a proof that we could not find in the stochastic literature. Second, we show the convergence of trajectories and controls, together with the convergence of variational inequalities (see Sect. 5.2.2 for a definition). The latter represents the cornerstone of the proof and paves the way for the final step, which consists of proving the convergence of the Pontryagin extremals. For the sake of clarity and brevity and without loss of generality, we carry out the proof in the case of scalar Brownian motion, i.e., we assume  $d = 1$ . Moreover, for any  $x \in \mathbb{R}^n$  with  $n \in \mathbb{N}$ , we adopt the notation  $\tilde{x} \triangleq (x, x^{n+1}) \in \mathbb{R}^{n+1}$ .

#### 4.1. Main steps of the proof of the stochastic maximum principle

For the sake of clarity and brevity, we retrace the proof of Theorem 3.3 only. The proof of Theorem 3.4 follows from a straightforward modification of the steps we provide below, by introducing the additional final constraint  $\mathbb{E}[y(T) - \Delta_{k+1}] \leq 0$  via (3.2). In particular, we highlight those modifications below and in Section 4.2.2.

##### 4.1.1. Linear stochastic differential equations

Define the stochastic matrices  $A(t) \triangleq \frac{\partial(b, f^0)}{\partial x}(t, u(t), x(t))$  and  $D(t) \triangleq \frac{\partial(\sigma, 0)}{\partial x}(t, u(t), x(t))$ . For any time  $r \in [0, T]$  and any bounded initial condition  $\tilde{\xi}_r \in L^2_{\mathcal{F}_r}(\Omega; \mathbb{R}^{n+1})$ , the following problem

$$\begin{cases} dz(t) = A(t)z(t) dt + D(t)z(t) dB_t \\ z(s) = 0, \quad s \in [0, r), \quad z(r) = \tilde{\xi}_r \end{cases} \quad (4.1)$$

is well-posed [54]. Its unique solution is the  $\mathcal{F}$ -adapted with right-continuous sample paths process  $z : [0, T] \times \Omega \rightarrow \mathbb{R}^{n+1} : (t, \omega) \mapsto \mathbb{1}_{[r, T]}(t)\phi(t, \omega)\psi(r, \omega)\xi_r(\omega)$ , where the matrix-valued  $\mathcal{F}$ -adapted with continuous sample paths processes  $\phi$  and  $\psi$  satisfy

$$\begin{cases} d\phi(t) = A(t)\phi(t)dt + D(t)\phi(t)dB_t \\ \phi(0) = I, \end{cases} \quad \begin{cases} d\psi(t) = -\psi(t)(A(t) - D(t)^2)dt - \psi(t)D(t)dB_t \\ \psi(0) = I, \end{cases} \quad (4.2)$$

respectively. In particular, a straightforward application of the Itô formula shows that  $\phi(t)\psi(t) = \psi(t)\phi(t) = I$ , and therefore  $\psi(t) = \phi(t)^{-1}$ , for every  $t \in [0, T]$ .

##### 4.1.2. Needle-like variations and end-point mapping

One way to prove the PMP comes from the analysis of specific variations called *needle-like variations* on a mapping called the *end-point mapping*. Those concepts are introduced below in the context of optimization over deterministic controls (see Sect. 5 for the generalization of this argument to stochastic controls).

Given an integer  $j \in \mathbb{N}$ , fix  $j$  times  $0 < t_1 < \dots < t_j < T$  which are Lebesgue points for  $u$ , and fix  $j$  random variables  $u_1, \dots, u_j$  such that  $u_i \in U$ . For fixed scalars  $0 \leq \eta_i < t_{i+1} - t_i$ ,  $i = 1, \dots, j-1$ , and  $0 \leq \eta_j < T - t_j$ , the needle-like variation  $\pi = \{t_i, \eta_i, u_i\}_{i=1, \dots, j}$  of the control  $u$  is defined to be the admissible control  $u_\pi(t) = u_i$  if  $t \in [t_i, t_i + \eta_i]$  and  $u_\pi(t) = u(t)$  otherwise. Denote by  $\tilde{x}_v$  the solution related to an admissible control  $v$  of the augmented system

$$\begin{cases} dx(t) = b(t, v(t), x(t)) dt + \sigma(t, x(t)) dB_t, & x(0) = x^0 \\ dx^{n+1}(t) = f^0(t, v(t), x(t)) dt, & x^{n+1}(0) = 0 \end{cases} \quad (4.3)$$

and define the mapping  $\tilde{g} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{q+1} : \tilde{x} \mapsto (g(x), x^{n+1})$ . For every fixed time  $t \in (t_j, T]$ , by denoting  $\delta_t \triangleq \min\{t_{i+1} - t_i, t - t_j : i = 1, \dots, j-1\} > 0$ , the end-point mapping at time  $t$  is defined to be the function

$$\begin{aligned} F_t^j : \mathcal{C}_t^j \triangleq B_{\delta_t}^j(0) \cap \mathbb{R}_+^j &\rightarrow \mathbb{R}^{q+1} \\ (\eta_1, \dots, \eta_j) &\mapsto \mathbb{E}[\tilde{g}(\tilde{x}_{u_\pi}(t)) - \tilde{g}(\tilde{x}_u(t))] \end{aligned} \quad (4.4)$$

where  $B_\rho^j$  is the open ball in  $\mathbb{R}^j$  of radius  $\rho > 0$ . Due to Lemma 2.1, it is not difficult to see that  $F_t^j$  is Lipschitz (see also the argument developed to prove Lem. 4.2). In addition, this mapping may be Gateaux differentiated

at zero along admissible directions of the cone  $\mathcal{C}_t^j$ . For this, denote  $\tilde{b} = (b^\top, f^0)^\top$ ,  $\tilde{\sigma} = (\sigma^\top, 0)^\top$  and let  $z_{t_i, u_i}$  be the unique solution to (4.1) with  $\xi_{t_i} = \tilde{b}(t_i, u_i, x(t_i)) - \tilde{b}(t_i, u(t_i), x(t_i))$ .

**Remark 4.1.** For the proof of Theorem 3.4, the only change compared to the proof of Theorem 3.3 which is required up to this point consists of replacing the function  $\tilde{g} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{q+1}$  which defines the end-point mapping (4.4) by the function

$$(\tilde{g}_k, g_k^{n+2}) : \mathbb{R}^{n+2} \rightarrow \mathbb{R}^{q+2} : (\tilde{x}, y) \mapsto (g_k(x), x^{n+1}, y).$$

Note that this change is consistent since we will effectively make use of the mapping  $F_t^j$  at the time  $t = T$  only.

**Lemma 4.2** (Stochastic needle-like variation formula). *Let  $(\eta_1, \dots, \eta_j) \in \mathcal{C}_t^j$ . For any  $t > t_j$ , it holds that*

$$\left\| \mathbb{E} \left[ \tilde{g}(\tilde{x}_{u_\pi}(t)) - \tilde{g}(\tilde{x}_u(t)) - \sum_{i=1}^j \eta_i \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(t)) z_{t_i, u_i}(t) \right] \right\| = o \left( \sum_{i=1}^j \eta_i \right).$$

The proof of this result is technical (it requires an intense use of stochastic inequalities) but not difficult. We provide an extensive proof of Lemma 4.2 in the appendix in a more general context (see also Sect. 5).

#### 4.1.3. Variational inequalities

The main step in the proof of the PMP goes by contradiction, leveraging Lemma 4.2. To this end, for every  $j \in \mathbb{N}$ , define the linear mapping

$$dF_T^j : \mathbb{R}_+^j \rightarrow \mathbb{R}^{q+1} : (\eta_1, \dots, \eta_j) \mapsto \sum_{i=1}^j \eta_i \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(T)) z_{t_i, u_i}(T) \right],$$

which due to Lemma 4.2, satisfies

$$\lim_{\alpha > 0, \alpha \rightarrow 0} \frac{F_T^j(\alpha \eta)}{\alpha} = dF_T^j(\eta),$$

for every  $\eta \in \mathbb{R}_+^j$ . Finally, consider the closed, convex cone of  $\mathbb{R}^{q+1}$  given by

$$K \triangleq \text{Cl} \left( \text{Cone} \left\{ \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(T)) z_{t_i, u_i}(T) \right] : \text{for } u_i \in U \text{ and } t_i \in (0, T) \text{ is Lebesgue for } u \right\} \right).$$

If  $K = \mathbb{R}^{q+1}$ , it would hold  $dF_T^j(\mathbb{R}_+^j) = K = \mathbb{R}^{q+1}$ , and by applying ([1], Lem. 12.1), one would find that the origin is an interior point of  $F_T^j(\mathcal{C}_T^j)$ . This would imply that  $(u, x)$  cannot be optimal for OCP, which gives a contradiction.

The argument above (together with an application of the separation plane theorem) provides the existence of a non-zero vector denoted  $\tilde{\mathbf{p}} = (\mathbf{p}^\top, \mathbf{p}^0) \in \mathbb{R}^{q+1}$  such that the following variational inequality holds

$$\tilde{\mathbf{p}}^\top \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(T)) z_{r, v}(T) \right] \leq 0, \quad r \in [0, T] \text{ is Lebesgue for } u, v \in U. \quad (4.5)$$

**Remark 4.3.** For the proof of Theorem 3.4, the variational inequalities (4.5) are replaced by the following upgraded variational inequalities which hold for  $(u_{k+1}, x_{k+1})$ , solution to  $\text{LOCP}_{k+1}^\Delta$ :

$$\begin{pmatrix} \tilde{\mathbf{p}}_{k+1} \\ p_{k+1}^1 \end{pmatrix}^\top \mathbb{E} \left[ \begin{pmatrix} \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_k(T)) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_{r,v}^k(T) \\ (z_{r,v}^k)^{n+2}(T) \end{pmatrix} \right] \leq 0, \quad (4.6)$$

$r \in [0, T]$  is Lebesgue for  $u_k, v \in U$ ,

where each  $((z_{r,v}^k)^\top, (z_{r,v}^k)^{n+2})^\top$  solves an equation similar to (4.1) (we report this new equation in Sect. 4.2.2).

#### 4.1.4. Conclusion of the proof of the stochastic maximum principle

The conditions of the PMP are derived by working out the variational inequality (4.5) and finding expressions of some appropriate conditional expectations. The main details are developed below in the context of optimization over deterministic controls (see Sect. 5 for the generalization to stochastic controls).

First, by appropriately developing solutions to (4.1), (4.5) can be rewritten as

$$\mathbb{E} \left[ \left( \tilde{\mathbf{p}}^\top \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(T)) \phi(T) \psi(r) \right)^\top \left( \begin{pmatrix} b \\ f^0 \end{pmatrix}(r, v, x(r)) - \begin{pmatrix} b \\ f^0 \end{pmatrix}(r, u(r), x(r)) \right) \right] \leq 0$$

for every  $r \in [0, T]$  Lebesgue point for  $u$  and every  $v \in U$ . Second, again from the structure of (4.1), it can be readily checked that by denoting

$$p(t) \triangleq \left( \tilde{\mathbf{p}}^\top \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(T)) \phi(T) \middle| \mathcal{F}_t \right] \psi(t) \right)_{1, \dots, n}, \quad p^0 \triangleq \left( \tilde{\mathbf{p}}^\top \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(T)) \phi(T) \psi(t) \right)_{n+1}, \quad (4.7)$$

the quantity  $p^0$  is constant in  $[0, T]$  (in addition, its negativity can be shown through a standard reformulation of problem OCP, as done in [1], Sect. 12.4). Notice that the stochastic process  $p : [0, T] \times \Omega \rightarrow \mathbb{R}^n$  is by definition  $\mathcal{F}$ -adapted. The quantities so far introduced allow one to reformulate the inequality above as

$$\mathbb{E} \left[ p(t)^\top \left( b(t, u(t), x(t)) - b(t, v, x(t)) \right) + p^0 \left( f^0(t, u(t), x(t)) - f^0(t, v, x(t)) \right) \right] \geq 0$$

for every  $t \in [0, T]$  Lebesgue point for  $u$  and  $v \in U$ , from which we infer the maximality condition of the PMP.

It remains to show the existence of the process  $q \in L_{\mathcal{F}}(\Omega; L^2([0, T]; \mathbb{R}^n))$ , the continuity of the sample paths of the process  $p$ , and the validity of the adjoint equation. For this, remark that, due to Jensen inequality and Lemma 2.1, the martingale  $\left( \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(T)) \phi(T) \middle| \mathcal{F}_t \right] \right)_{t \in [0, T]}$  is bounded in  $L^2$ . Hence, the martingale representation theorem provides the existence of a process  $\mu \in L_{\mathcal{F}}^2([0, T] \times \Omega; \mathbb{R}^{(q+1) \times (n+1)})$  such that  $\mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(T)) \phi(T) \middle| \mathcal{F}_t \right] = \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(T)) \phi(T) \middle| \mathcal{F}_0 \right] + \int_0^t \mu(s) dB_s \triangleq N + \chi(t)$ , where  $N \in \mathbb{R}^{(q+1) \times (n+1)}$  is a constant matrix. The definition in (4.7) immediately gives that the sample paths of the process  $p$  are continuous. Next, an application of Itô formula (component-wise) readily shows that the product  $\chi\psi$  satisfies, for  $t \in [0, T]$ ,

$$\begin{aligned} (\chi\psi)(t) &= \left( \int_0^t \mu(s) dB_s \right) \psi(t) = \int_0^t \mu(s) \psi(s) dB_s - \int_0^t \mu(s) \psi(s) D(s) ds \\ &\quad - \int_0^t \chi(s) \psi(s) (A(s) - D(s)^2) ds - \int_0^t \chi(s) \psi(s) D(s) dB_s. \end{aligned}$$

Denoting  $q(t) \triangleq \left( \tilde{\mathbf{p}}^\top \left( \mu(t)\psi(t) - (N\psi(t) + \chi(t)\psi(t))D(t) \right) \right)_{1,\dots,n}$ , the computations above readily give the adjoint equation of the PMP. Those computations may also be leveraged to show  $p \in L^2_{\mathcal{F}}(\Omega; C([0, T]; \mathbb{R}^n))$  and  $q \in L^2_{\mathcal{F}}([0, T] \times \Omega; \mathbb{R}^n)$  (see, e.g., [54], Sect. 7.2).

## 4.2. Proof of the convergence result

Here we enter the core of the proof of Theorem 3.6. The convergence of trajectories and controls is addressed first. We devote the last two sections to the convergence of variational inequalities and Pontryagin extremals. For the sake of clarity and brevity, we only consider free-final-time problems. From now on, we implicitly assume  $(A_1)$ – $(A_3)$ .

### 4.2.1. Convergence of controls and trajectories

Due to  $(A_1)$ – $(A_3)$ , there exists a sequence of tuples  $(u_k, x_k)_{k \in \mathbb{N}}$  such that for every  $k \in \mathbb{N}$ , the tuple  $(u_{k+1}, x_{k+1})$  solves  $\text{LOCP}_{k+1}^\Delta$ . In what follows, we implicitly adopt the reformulation of each problem convexified problem  $\text{LOCP}_{k+1}^\Delta$ , which consist of adding  $\mathbb{E}[y(T) - \Delta_{k+1}] \leq 0$  to the final constraints through the new variable  $y$  in (3.2). If  $u \in \mathcal{U}$  denotes an admissible control for OCP that fulfills the conditions of Theorem 3.6, we denote by  $\tilde{x} : [0, T] \times \Omega \rightarrow \mathbb{R}^{n+1}$  the  $\mathcal{F}$ -adapted with continuous sample paths process solution to the augmented system (4.3) related to OCP with control  $u$ . The following holds.

**Lemma 4.4** (Convergence of trajectories). *Assume that the sequence  $(\Delta_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}_+$  converges to zero. If the sequence  $(u_k)_{k \in \mathbb{N}}$  converges to  $u$  for the weak topology of  $L^2$ , then  $\mathbb{E} \left[ \sup_{t \in [0, T]} \|x_k(t) - x(t)\|^\ell \right] \rightarrow 0$ , as  $k \rightarrow \infty$ , for every  $\ell \geq 2$ .*

*Proof.* For every  $t \in [0, T]$ , we have (below,  $C \geq 0$  represents some often overloaded appropriate constant)

$$\begin{aligned} \|x_{k+1}(t) - x(t)\|^\ell &\leq C \left\| \int_0^t \left( b_0(s, x_k(s)) - b_0(s, x(s)) \right) ds \right\|^\ell \\ &+ C \sum_{i=1}^m \left\| \int_0^t \left( u_{k+1}^i(s) b_i(s, x_k(s)) - u^i(s) b_i(s, x(s)) \right) ds \right\|^\ell \\ &+ C \left\| \int_0^t \left( \frac{\partial b_0}{\partial x}(s, x_k(s)) + \sum_{i=1}^m u_k^i(s) \frac{\partial b_i}{\partial x}(s, x_k(s)) \right) (x_{k+1}(s) - x_k(s)) ds \right\|^\ell \\ &+ C \left\| \int_0^t \left( \sigma(s, x_k(s)) - \sigma(s, x(s)) + \frac{\partial \sigma}{\partial x}(s, x_k(s))(x_{k+1}(s) - x_k(s)) \right) dB_s \right\|^\ell \end{aligned} \quad (4.8)$$

Now, we take expectations. For the last term, we compute

$$\begin{aligned} &\mathbb{E} \left[ \left\| \int_0^t \left( \sigma(s, x_k(s)) - \sigma(s, x(s)) + \frac{\partial \sigma}{\partial x}(s, x_k(s))(x_{k+1}(s) - x_k(s)) \right) dB_s \right\|^\ell \right] \\ &\leq C \left( \int_0^t \mathbb{E} \left[ \sup_{r \in [0, s]} \|x_{k+1}(r) - x(r)\|^\ell \right] ds + \mathbb{E} \left[ \int_0^T \|x_{k+1}(s) - x_k(s)\|^{1+(\ell-1)} ds \right] \right) \\ &\leq C \int_0^t \mathbb{E} \left[ \sup_{r \in [0, s]} \|x_{k+1}(r) - x(r)\|^\ell \right] ds + C \left( \int_0^T \mathbb{E} [\|x_{k+1}(s) - x_k(s)\|^2] ds \right)^{\frac{1}{2}} \left( \sup_{k \in \mathbb{N}} \mathbb{E} \left[ \sup_{s \in [0, T]} \|x_k(s)\|^{2(\ell-1)} \right] \right)^{\frac{1}{2}} \\ &\leq C \left( \int_0^t \mathbb{E} \left[ \sup_{r \in [0, s]} \|x_{k+1}(r) - x(r)\|^\ell \right] ds + \Delta_{k+1} \right) \end{aligned}$$



due to Hölder and Burkholder–Davis–Gundy inequalities, and the last inequality comes from Lemma 2.1. Similar computations can be carried out for the first and third terms of (4.8).

To handle the second term of (4.8), we proceed as follows. We see that since  $\mathbb{E} \left[ \int_0^T \|b_i(s, x(s))\|^2 dt \right] < \infty$  implies  $\int_0^T \|b_i(s, x(s))\|^2 dt < \infty$ ,  $\mathcal{G}$ -almost surely, for every fixed  $i = 1, \dots, m$  and  $t \in [0, T]$ , the convergence of the sequence of controls for the weak topology of  $L^2$  entails that  $\left\| \int_0^t (u_k^i(s) - u^i(s)) b_i(s, x(s)) ds \right\| \rightarrow 0$ ,  $\mathcal{G}$ -almost surely as  $k \rightarrow \infty$ . In addition,  $(A_1)$  gives  $\left\| \int_a^b (u_k^i(s) - u^i(s)) b_i(s, x(s)) ds \right\| \leq C|b - a|$  for every  $a, b \in [0, T]$  and  $\mathcal{G}$ -almost surely. Hence, Lemma 3.4 of [50] and the dominated convergence theorem finally provide that

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \left\| \int_0^t (u_k^i(s) - u^i(s)) b_i(s, x(s)) ds \right\|^{\ell'} \right] \rightarrow 0, \quad k \rightarrow \infty.$$

It is easy to conclude the proof by applying a routine Grönwall inequality argument.  $\square$

The sought-after convergence of trajectories is a consequence of Lemma 4.4 when the conditions of Theorem 3.6 are met. In addition, limiting points for the sequences of controls fulfilling the conditions of Theorem 3.6 always exist up to some subsequence. Indeed, in this case, the set of admissible controls  $\mathcal{U}$  is closed and convex for the strong topology of  $L^2$ . Hence it is closed for the weak topology of  $L^2$ . Therefore, since  $(u_k)_{k \in \mathbb{N}}$  is bounded in  $L^2$ , there exists  $u \in \mathcal{U}$  such that, up to some subsequence,  $(u_k)_{k \in \mathbb{N}}$  weakly converges to  $u$  for the weak topology of  $L^2$ . It remains to show that the process  $x$  is feasible for OCP. For this, we compute

$$\begin{aligned} \left| \mathbb{E}[g(x(T))] \right| &\leq \mathbb{E} \left[ \|g(x(T)) - g(x_k(T))\| \right] + \mathbb{E} \left[ \left\| \frac{\partial g}{\partial x}(x_k(T))(x_{k+1}(T) - x_k(T)) \right\| \right] \\ &\leq C \left( \mathbb{E} \left[ \|x(T) - x_k(T)\|^2 \right]^{\frac{1}{2}} + \mathbb{E} \left[ \|x_{k+1}(T) - x(T)\|^2 \right]^{\frac{1}{2}} \right) \rightarrow 0 \end{aligned}$$

due to Lemma 4.4 and the dominated convergence theorem (here  $C \geq 0$  is a constant coming from  $(A_2)$ , and we use the continuity of the sample paths of  $x$ ).

#### 4.2.2. Convergence of variational inequalities

We start with a crucial result on linear stochastic differential equations. Recall the notation of Section 4.1.

**Lemma 4.5** (Convergence of variational inequalities). *Fix  $\ell \geq 2$  and consider sequences of times  $(r_k)_{k \in \mathbb{N}} \subseteq [0, T]$  and of uniformly bounded variables  $(\tilde{\xi}_k)_{k \in \mathbb{N}}$  such that for every  $k \in \mathbb{N}$ ,  $\tilde{\xi}_k \in L^2_{\mathcal{F}_{r_k}}(\Omega; \mathbb{R}^{n+1})$ . Assume that  $r_k \rightarrow r$  with  $r_k \leq r$  for  $k \in \mathbb{N}$  and  $\tilde{\xi}_k \xrightarrow{L^\ell} \tilde{\xi} \in L^2_{\mathcal{F}_r}(\Omega; \mathbb{R}^{n+1})$  with  $\tilde{\xi}$  bounded. Denote  $\tilde{w}_{k+1}, \tilde{w}$  the stochastic process solutions, respectively, to*

$$\begin{cases} dw(t) = \left( \frac{\partial b_0}{\partial x}(t, x_k(t)) + \sum_{i=1}^m u_k^i(t) \frac{\partial b_i}{\partial x}(t, x_k(t)) \right) w(t) dt + \frac{\partial \sigma}{\partial x}(t, x_k(t)) w(t) dB_t \\ dw^{n+1}(t) = \left( \frac{\partial H}{\partial x}(x_{k+1}(t)) + \frac{\partial L}{\partial x}(t, u_k(t), x_k(t)) \right) w(t) dt \\ \tilde{w}(t) = 0, \quad t \in [0, r_k), \quad \tilde{w}(r_k) = \tilde{\xi}_{k+1}, \end{cases}$$

$$\begin{cases} dw(t) = \left( \frac{\partial b_0}{\partial x}(t, x(t)) + \sum_{i=1}^m u^i(t) \frac{\partial b_i}{\partial x}(t, x(t)) \right) w(t) dt + \frac{\partial \sigma}{\partial x}(t, x(t)) w(t) dB_t \\ dw^{n+1}(t) = \left( \frac{\partial H}{\partial x}(x(t)) + \frac{\partial L}{\partial x}(t, u(t), x(t)) \right) w(t) dt \\ \tilde{w}(t) = 0, t \in [0, r], \quad \tilde{w}(r) = \tilde{\xi}. \end{cases}$$

Under the assumptions of Lemma 4.4, it holds that  $\mathbb{E} \left[ \sup_{t \in [r, T]} \|\tilde{w}_k(t) - \tilde{w}(t)\|^\ell \right] \rightarrow 0$  for  $k \rightarrow \infty$ .

*Proof.* From  $r_k \leq r$ ,  $k \in \mathbb{N}$ , for  $t \in [r, T]$  we have (below,  $C \geq 0$  is a constant)

$$\begin{aligned} \|\tilde{w}_{k+1}(t) - \tilde{w}(t)\|^\ell &\leq C \|\tilde{\xi}_{k+1} - \tilde{\xi}\|^\ell \\ &+ C \left\| \int_{r_k}^t \left( \frac{\partial H}{\partial x}(x_{k+1}(s)) w_{k+1}(s) - \frac{\partial H}{\partial x}(x(s)) w(s) \right) ds \right\|^\ell \\ &+ C \left\| \int_{r_k}^t \left( \frac{\partial b_0}{\partial x}(s, x_k(s)) w_{k+1}(s) - \frac{\partial b_0}{\partial x}(s, x(s)) w(s) \right) ds \right\|^\ell \\ &+ C \left\| \int_0^t \mathbb{1}_{[r_k, T]}(t) \left( \frac{\partial \sigma}{\partial x}(s, x_k(s)) w_{k+1}(s) - \frac{\partial \sigma}{\partial x}(s, x(s)) w(s) \right) dB_s \right\|^\ell \\ &+ C \left\| \int_{r_k}^t \left( \frac{\partial L}{\partial x}(t, u_k(t), x_k(t)) w_{k+1}(s) - \frac{\partial L}{\partial x}(t, u(t), x(t)) w(s) \right) ds \right\|^\ell \\ &+ C \sum_{i=1}^m \left\| \int_{r_k}^t \left( u_k^i(s) \frac{\partial b_i}{\partial x}(s, x_k(s)) w_{k+1}(s) - u^i(s) \frac{\partial b_i}{\partial x}(s, x(s)) w(s) \right) ds \right\|^\ell. \end{aligned} \quad (4.9)$$

We consider the expectation of the fourth term on the right-hand side (a similar argument can be developed for the first three terms, which are omitted in the interest of clarity and brevity). The Burkholder–Davis–Gundy and Hölder inequalities give

$$\begin{aligned} &\mathbb{E} \left[ \left\| \int_0^t \mathbb{1}_{[r_k, T]}(t) \left( \frac{\partial \sigma}{\partial x}(s, x_k(s)) w_{k+1}(s) - \frac{\partial \sigma}{\partial x}(s, x(s)) w(s) \right) dB_s \right\|^\ell \right] \\ &\leq C|r - r_k| + C \int_r^t \mathbb{E} \left[ \sup_{s' \in [0, s]} \|w_{k+1}(s') - w(s')\|^\ell \right] ds + C \mathbb{E} \left[ \sup_{s \in [0, T]} \|w(s)\|^{2\ell} \right] \mathbb{E} \left[ \sup_{s \in [0, T]} \|x_k(s) - x(s)\|^{2\ell} \right], \end{aligned}$$

where the last inequality holds due to Lemma 4.4 and because Lemma 2.1 may be readily extended to  $\tilde{w}$ .

Finally, due to the fact that  $L$  is affine with respect to the control variable, we may handle the fifth and the sixth terms in (4.9) by combining the argument above with the final steps in the proof of Lemma 4.4, and a routine Grönwall inequality argument provides the conclusion.  $\square$

Consider  $(u_k)_{k \in \mathbb{N}} \subseteq \mathcal{U}$  which converges to  $u \in \mathcal{U}$  for the weak convergence of  $L^2$ . By assuming  $(A_4)$ , we may denote by  $\mathcal{L} \subseteq [0, T]$  the full Lebesgue-measure subset such that  $r \in \mathcal{L}$  if and only if  $r$  is a Lebesgue point for  $u$  and  $r \notin \cup_{k \in \mathbb{N}} \mathcal{D}_{k+1}$  (we use the notation introduced with  $(A_4)$ ). We prove the existence of a non-zero vector  $\tilde{\mathbf{p}} \in \mathbb{R}^{q+1}$  whose last component is non-positive such that for every  $r \in \mathcal{L}$ ,  $v \in U$ ,

$$\tilde{\mathbf{p}}^\top \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}(T)) z_{r, v}(T) \right] \leq 0, \quad (4.10)$$

where the  $\mathcal{F}$ -adapted with continuous sample paths stochastic process  $z_{r,v} : [0, T] \times \Omega \rightarrow \mathbb{R}^{n+1}$  solves (4.1) with

$$A(t) = \begin{pmatrix} \frac{\partial b_0}{\partial x}(t, x(t)) + \sum_{i=1}^m u^i(t) \frac{\partial b_i}{\partial x}(t, x(t)) \\ \frac{\partial H}{\partial x}(x(t)) + \frac{\partial L}{\partial x}(t, u(t), x(t)) \end{pmatrix}, \quad D(t) = \begin{pmatrix} \frac{\partial \sigma}{\partial x}(t, x(t)) \\ 0 \end{pmatrix},$$

$$\tilde{\xi}_r = \tilde{\xi}_{r,v} \triangleq \tilde{b}(r, v, x(r)) - \tilde{b}(r, u(r), x(r)),$$

where we denote  $\tilde{b} = (b^\top, f^0)^\top$  – we will use this notation from now on.

For this, due to (4.6), for every  $k \in \mathbb{N}$ , the optimality of  $(u_{k+1}, x_{k+1})$  for  $\text{LOCP}_{k+1}^\Delta$  provides a non-zero vector  $(\tilde{p}_{k+1}^\top, p_{k+1}^1)^\top \in \mathbb{R}^{q+2}$  whose second to last component  $p_{k+1}^0 \leq 0$  is non-zero by assumption, so that for  $r \in \mathcal{L}$  and  $v \in U$ , it holds that

$$\begin{pmatrix} \tilde{p}_{k+1} \\ p_{k+1}^1 \end{pmatrix}^\top \mathbb{E} \left[ \begin{pmatrix} \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_k(T)) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_{r,v}^{k+1}(T) \\ (z_{r,v}^{k+1})^{n+2}(T) \end{pmatrix} \right] \leq 0,$$

where with the notation

$$A_{k+1}(t) = \begin{pmatrix} \frac{\partial b_0}{\partial x}(t, x_k(t)) + \sum_{i=1}^m u_k^i(t) \frac{\partial b_i}{\partial x}(t, x_k(t)) & 0 \\ \frac{\partial H}{\partial x}(x_{k+1}(t)) + \frac{\partial L}{\partial x}(t, u_k(t), x_k(t)) & 0 \\ 2(x_{k+1}(t) - x_k(t)) & 0 \end{pmatrix}, \quad D_{k+1}(t) = \begin{pmatrix} \frac{\partial \sigma}{\partial x}(t, x_k(t)) & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

the  $\mathcal{F}$ -adapted with continuous sample paths stochastic process  $((z_{r,v}^k)^\top, (z_{r,v}^k)^{n+2})^\top : [0, T] \times \Omega \rightarrow \mathbb{R}^{n+2}$  solves a higher dimensional version of (4.1) with

$$A = A_{k+1}, \quad D = D_{k+1}, \quad \text{and initial condition } \begin{pmatrix} \tilde{\xi}_{r,v}^{k+1} \\ 0 \end{pmatrix} \triangleq \begin{pmatrix} \tilde{b}_{k+1}(r, v, x_{k+1}(r)) - \tilde{b}_{k+1}(r, u_{k+1}(r), x_{k+1}(r)) \\ 0 \end{pmatrix},$$

where again we denote  $\tilde{b}_{k+1} = (b_{k+1}^\top, f_{k+1}^0)^\top$ .

Now, fix  $r \in \mathcal{L}$  and  $v \in U$ . The following comes from combining Lemma 3.7 with Lemma 3.11 of [13].

**Lemma 4.6** (Pointwise convergence on controls). *Under  $(A_4)$ , there exists  $(r_k)_{k \in \mathbb{N}} \subseteq (0, r)$  such that for every  $k \in \mathbb{N}$ ,  $r_k$  is a Lebesgue point for  $u_k$ , and  $r_k \rightarrow r$ ,  $u_k(r_k) \rightarrow u(r)$  as  $k \rightarrow \infty$ .*

If  $(r_k)_{k \in \mathbb{N}} \subseteq (0, r)$  denotes the sequence given by Lemma 4.6, we define  $\tilde{\xi}_k = \tilde{\xi}_{r_k, v}^k$  and  $\tilde{\xi} = \tilde{\xi}_{r, v}$ . Straightforward computations give (below,  $C \geq 0$  is a constant)

$$\begin{aligned} \mathbb{E} \left[ \|\tilde{\xi}_{k+1} - \tilde{\xi}\|^2 \right] &\leq C \left( \|u_{k+1}(r_{k+1}) - u(r)\|^2 + |r_{k+1} - r|^2 \right. \\ &\quad \left. + \mathbb{E} \left[ \|x(r_{k+1}) - x(r)\|^2 \right] + \mathbb{E} \left[ \sup_{s \in [0, T]} \|x_k(s) - x(s)\|^2 \right] \right), \end{aligned} \quad (4.11)$$

from which  $\mathbb{E} \left[ \|\tilde{\xi}_{k+1} - \tilde{\xi}\|^2 \right] \rightarrow 0$  for  $k \rightarrow \infty$  (Lem. 4.4 and 4.6). Therefore, from Lemma 4.5 we infer that

$$\mathbb{E} \left[ \sup_{t \in [r, T]} \|z_{r_k, v}^k(t) - z_{r, v}(t)\|^2 \right] \rightarrow 0, \quad k \rightarrow \infty,$$

which, together with  $\Delta_k \rightarrow 0$ , readily yields

$$\mathbb{E} \left[ |(z_{r_k, v}^k)^{n+2}(T)| \right] \rightarrow 0, \quad k \rightarrow \infty.$$

At this step, we point out that the variational inequalities in (4.5) and in (4.6) still hold if we take multipliers of norm one. Specifically, we may assume that  $\|(\tilde{\mathbf{p}}_{k+1}^\top, p_{k+1}^1)^\top\| = 1$  for every  $k \in \mathbb{N}$ . Therefore, up to some subsequence, there exists a vector  $(\tilde{\mathbf{p}}^\top, p^1)^\top = (\mathbf{p}^\top, p^0, p^1)^\top \in \mathbb{R}^{q+1}$  such that  $(\tilde{\mathbf{p}}_k^\top, p_k^1)^\top \rightarrow (\tilde{\mathbf{p}}^\top, p^1)^\top$  for  $k \rightarrow \infty$  and satisfying  $(\tilde{\mathbf{p}}^\top, p^1)^\top \neq 0$ ,  $p^0 \leq 0$ . We use this remark to conclude as follows. The definition of  $\tilde{g}$  and the Hölder inequality give ( $C \geq 0$  is a constant)

$$\begin{aligned} \tilde{\mathbf{p}}^\top \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}(T)) z_{r, v}(T) \right] &\leq \left| \tilde{\mathbf{p}}^\top \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}(T)) z_{r, v}(T) \right] - \begin{pmatrix} \tilde{\mathbf{p}}_{k+1}^\top \\ p_{k+1}^1 \end{pmatrix}^\top \mathbb{E} \left[ \begin{pmatrix} \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_k(T)) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_{r_{k+1}, v}^{k+1}(T) \\ (z_{r_{k+1}, v}^{k+1})^{n+2}(T) \end{pmatrix} \right] \right| \\ &\leq C \left( \|\tilde{\mathbf{p}} - \tilde{\mathbf{p}}_{k+1}\| + p_{k+1}^1 \mathbb{E} \left[ |(z_{r_k, v}^k)^{n+2}(T)| \right] \right. \\ &\quad \left. + \mathbb{E} \left[ \|z_{r_{k+1}, v}^{k+1}(T) - z_{r, v}(T)\|^2 \right]^{\frac{1}{2}} + \mathbb{E} \left[ \|z_{r, v}(T)\|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \sup_{s \in [0, T]} \|x_k(s) - x(s)\|^2 \right]^{\frac{1}{2}} \right), \end{aligned}$$

and in this case, (4.10) follows from Lemma 4.4 and the convergences obtained above. From what we showed in Section 4.1.4, this latter inequality yields that  $(u, \mathbf{p}, p^0)$  is extremal for OCP as soon as  $(\mathbf{p}^\top, p^0)^\top \neq 0$ .

Finally, we turn to the case for which the sequence  $(u_k)_{k \in \mathbb{N}} \subseteq \mathcal{U}$  converges to  $u \in \mathcal{U}$  for the strong topology of  $L^2$ , but without assuming  $(A_4)$ . For this, fix  $v \in U$  and define the stochastic processes  $\tilde{\xi}_k(s) = \tilde{\xi}_{s, v}^k$  and  $\tilde{\xi}(s) = \tilde{\xi}_{s, v}$ , where  $s \in [0, T]$ . Similar computations to the ones developed to compute the bound (4.11) provide that  $\int_0^T \mathbb{E} \left[ \|\tilde{\xi}_k(s) - \tilde{\xi}(s)\|^2 \right] ds \rightarrow 0$  as  $k \rightarrow \infty$ , and therefore up to some subsequence, the quantity  $\mathbb{E} \left[ \|\tilde{\xi}_k(s) - \tilde{\xi}(s)\|^2 \right]$  converges to zero for  $k \rightarrow 0$  a.e. in  $[0, T]$ . By taking countable intersections of sets of Lebesgue points (one for each control  $u_k$ , for all  $k \in \mathbb{N}$ ), it follows that the argument above can be iterated exactly in the same manner (*via* Lem. 4.5), leading to the same conclusion.

#### 4.2.3. Convergence of multipliers and conclusion

By applying the construction developed in Section 4.1.4 to the variational inequality (4.10), we retrieve a tuple  $(p, \mathbf{p}, p^0, q)$ , where  $\mathbf{p} \in \mathbb{R}^q$ ,  $p^0 \leq 0$  are constant,  $p \in L^2_{\mathcal{F}}(\Omega; C^0([0, T]; \mathbb{R}^n))$  and  $q \in L^2_{\mathcal{F}}([0, T] \times \Omega; \mathbb{R}^n)$ , such that, as soon as  $(\mathbf{p}^\top, p^0)^\top \neq 0$ , the tuple  $(u, \mathbf{p}, p^0, q)$  is extremal for OCP associated with  $(u, x, p, \mathbf{p}, p^0, q)$  satisfying conditions 1., 2., and 3. of Theorem 3.3. To conclude, it only remains to prove that

$$\mathbb{E} \left[ \sup_{s \in [0, T]} \|p_k(s) - p(s)\|^2 \right] \rightarrow 0, \quad \int_0^T \mathbb{E} \left[ \|q_k(s) - q(s)\|^2 \right] ds \rightarrow 0, \quad k \rightarrow \infty. \quad (4.12)$$

Here, each tuple  $(u_k, \mathbf{p}_k, p_k^0)$  is the extremal of  $\text{LOCP}_k^\Delta$  associated with the tuple  $(u_k, x_k, p^k, \mathbf{p}_k, p_k^0, q_k)$ , where  $(p_k, q_k)$  solves the adjoint equation of Theorem 3.4. Since, under the assumption  $(\mathbf{p}^\top, p^0)^\top \neq 0$ , the multipliers

$p_k^1$ ,  $k \in \mathbb{N}$ , and  $p^1$  do not play any role, for the sake of clarity and brevity of notation, and without loss of generality, in what follows we implicitly assume  $p_k^1 = p^1 = 0$ , for every  $k \in \mathbb{N}$ .

Let us start with the first convergence of (4.12). For this, fix  $k \in \mathbb{N}$  and consider the process

$$\left( \mathbb{E} \left[ \sup_{s \in [0, T]} \left\| \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_k(T)) \phi_{k+1}(T) \psi_{k+1}(s) - \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}(T)) \phi(T) \psi(s) \right\| \middle| \mathcal{F}_t \right] \right)_{t \in [0, T]},$$

where  $\phi_{k+1}$ ,  $\psi_{k+1}$  solve (4.2) with matrices  $A_{k+1}$ ,  $D_{k+1}$  from which, due to  $p_k^1 = p^1 = 0$ , for every  $k \in \mathbb{N}$ , we remove the last column and row, whereas  $\phi$ ,  $\psi$  solve (4.2) with matrices  $A$ ,  $D$ , those matrices being defined above. Due to a straightforward extension of Lemma 2.1 to equations (4.2), this process is a martingale, bounded in  $L^2$ . Hence, the martingale representation theorem allows us to infer that this process is a martingale with continuous sample paths, and Doob and Jensen inequalities give

$$\begin{aligned} & \mathbb{E} \left[ \sup_{t \in [0, T]} \mathbb{E} \left[ \sup_{s \in [0, T]} \left\| \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_k(T)) \phi_{k+1}(T) \psi_{k+1}(s) - \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}(T)) \phi(T) \psi(s) \right\| \middle| \mathcal{F}_t \right]^2 \right] \\ & \leq 4 \mathbb{E} \left[ \sup_{s \in [0, T]} \left\| \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_k(T)) \phi_{k+1}(T) \psi_{k+1}(s) - \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}(T)) \phi(T) \psi(s) \right\|^2 \right], \end{aligned} \quad (4.13)$$

which holds for every  $k \in \mathbb{N}$ . By combining (4.13) with (4.7), we compute

$$\begin{aligned} & \mathbb{E} \left[ \sup_{s \in [0, T]} \|p_{k+1}(s) - p(s)\|^2 \right] \leq C \mathbb{E} \left[ \sup_{t \in [0, T]} \left\| \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}(t_f)) \phi(T) \psi(s) \right\|^2 \right] \|\tilde{\mathbf{p}}_{k+1} - \tilde{\mathbf{p}}\|^2 \\ & + C \mathbb{E} \left[ \sup_{s \in [0, T]} \left\| \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_k(T)) \phi_{k+1}(T) \psi_{k+1}(s) - \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}(T)) \phi(T) \psi(s) \right\|^2 \right], \end{aligned}$$

where  $C \geq 0$  is a constant. Up to some subsequence, the first term on the right-hand side converges to zero. Moreover, the definition of  $\tilde{g}$  and Hölder inequality give

$$\begin{aligned} & \mathbb{E} \left[ \sup_{s \in [0, T]} \left\| \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_k(T)) \phi_{k+1}(T) \psi_{k+1}(s) - \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}(T)) \phi(T) \psi(s) \right\|^2 \right] \\ & \leq C \mathbb{E} \left[ \|\phi_{k+1}(T)\|^4 \right]^{\frac{1}{2}} \mathbb{E} \left[ \sup_{s \in [0, T]} \|\psi_{k+1}(s) - \psi(s)\|^4 \right]^{\frac{1}{2}} \\ & + C \left( \mathbb{E} \left[ \sup_{s \in [0, T]} \|\psi(s)\|^4 \right]^{\frac{1}{2}} \mathbb{E} \left[ \sup_{s \in [0, T]} \|\phi_{k+1}(s) - \phi(s)\|^4 \right]^{\frac{1}{2}} + \mathbb{E} \left[ \sup_{s \in [0, T]} \|\phi(T) \psi(s)\|^4 \right]^{\frac{1}{2}} \mathbb{E} \left[ \sup_{s \in [0, T]} \|x_k(s) - x(s)\|^4 \right]^{\frac{1}{2}} \right), \end{aligned}$$

and a straightforward extension of Lemma 4.5 to equations (4.2) entails that all the terms on the right-hand side tend to zero. The convergence of  $(p_k)_{k \in \mathbb{N}}$  is proved.

It remains to prove the second convergence of (4.12). For this, we apply Itô formula to  $\|p_{k+1}(t) - p(t)\|^2$ , which due to (4.7), (4.13) and Lemma 2.1 extended to (4.2) gives (below,  $C \geq 0$  is a constant)

$$\begin{aligned} & \mathbb{E} \left[ \|p_{k+1}(0) - p(0)\|^2 \right] + \int_0^T \mathbb{E} \left[ \|q_{k+1}(s) - q(s)\|^2 \right] ds = \mathbb{E} \left[ \|p_{k+1}(T) - p(T)\|^2 \right] \\ & + 2 \mathbb{E} \left[ \int_0^T \left( p_{k+1}(s) - p(s) \right)^\top \left( p(s)^\top_{k+1} \frac{\partial b_{k+1}}{\partial x}(s, u_{k+1}(s), x_{k+1}(s)) - p(s)^\top \frac{\partial b}{\partial x}(s, u(s), x(s)) \right) ds \right] \end{aligned}$$

$$\begin{aligned}
 & + 2\mathbb{E} \left[ \int_0^T \left( p_{k+1}(s) - p(s) \right)^\top \left( p_{k+1}^0 \frac{\partial f_{k+1}^0}{\partial x}(s, u_{k+1}(s), x_{k+1}(s)) - p^0 \frac{\partial f^0}{\partial x}(s, u(s), x(s)) \right) ds \right] \\
 & + 2\mathbb{E} \left[ \int_0^T \left( p_{k+1}(s) - p(s) \right)^\top \left( q(s)_{k+1}^\top \frac{\partial \sigma_{k+1}}{\partial x}(s, x_{k+1}(s)) - q(s)^\top \frac{\partial \sigma}{\partial x}(s, x(s)) \right) ds \right] \\
 & \leq C\mathbb{E} \left[ \sup_{s \in [0, T]} \|p_{k+1}(s) - p(s)\|^2 \right]^{\frac{1}{2}} \left( \int_0^T \mathbb{E} [\|q_{k+1}(s) - q(s)\|^2] ds \right)^{\frac{1}{2}} \\
 & \quad + C\mathbb{E} \left[ \sup_{s \in [0, T]} \|p_{k+1}(s) - p(s)\|^2 \right]^{\frac{1}{2}} \left( 1 + \left( \int_0^T \mathbb{E} [\|q(s)\|^2] ds \right)^{\frac{1}{2}} \right).
 \end{aligned}$$

The conclusion finally follows from Young's inequality and the convergence of  $(p_k)_{k \in \mathbb{N}}$ .

## 5. EXTENSION TO PROBLEMS WITH FREE FINAL TIME AND STOCHASTIC CONTROLS

The accumulation properties of SCP can be extended to OCPs with free final time and whereby optimization is undertaken over stochastic controls. Specifically, in this section we investigate how results from Theorem 3.6 may be extended to finite-horizon, finite-dimensional non-linear stochastic General Optimal Control Problems (GOCP) having control-affine dynamics and uncontrolled diffusion, of the form

$$\begin{cases} \min_{u, t_f} \mathbb{E} \left[ \int_0^{t_f} f^0(s, u(s), x(s)) ds \right] \triangleq \mathbb{E} \left[ \int_0^{t_f} \left( G(u(s)) + H(x(s)) + L(s, u(s), x(s)) \right) ds \right] \\ dx(t) = b(t, u(t), x(t)) dt + \sigma(t, x(t)) dB_t, \quad x(0) = x^0, \quad \mathbb{E} [g(x(t_f))] = 0, \end{cases}$$

where the final time  $0 \leq t_f \leq T$  may be free or not (here,  $T > 0$  is some fixed maximal time), and we optimize over controls  $u \in \mathcal{U}$  which are either deterministic, *i.e.*,  $\mathcal{U} = L^2([0, T]; U)$ , or stochastic, *i.e.*,  $\mathcal{U} = L^2_{\mathcal{F}}([0, T] \times \Omega; U)$ . Solutions to GOCP will be denoted by  $(t_f, u, x) \in [0, T] \times \mathcal{U} \times L^{\ell}_{\mathcal{F}}(\Omega; C([0, T]; \mathbb{R}^n))$ , for  $\ell \geq 2$ .

By adopting the notation given in (3.1), the stochastic Linearized General Optimal Control Problem (LGOCP $_{k+1}^{\Delta}$ ) at iteration  $k \in \mathbb{N}$  may be defined accordingly as

$$\begin{cases} \min_{u, t_f} \mathbb{E} \left[ \int_0^{t_f} f_{k+1}^0(s, u(s), x(s)) ds \right] \triangleq \mathbb{E} \left[ \int_0^{t_f} f_{u_k, x_k}^0(s, u(s), x(s)) ds \right] \\ dx(t) = b_{k+1}(t, u(t), x(t)) dt + \sigma_{k+1}(t, x(t)) dB_t, \quad x(0) = x^0 \\ \quad \triangleq b_{u_k, x_k}(t, u(t), x(t)) dt + \sigma_{x_k}(t, x(t)) dB_t \\ \mathbb{E} [g_{k+1}(x(t_f))] \triangleq \mathbb{E} \left[ g(x_k(t_f^k)) + \frac{\partial g}{\partial x}(x_k(t_f^k))(x(t_f) - x_k(t_f^k)) \right] = 0 \\ \int_0^T \mathbb{E} [\|x(s) - x_k(s)\|^2] ds \leq \Delta_{k+1}, \quad |t_f - t_f^k| \leq \Delta_{k+1}, \end{cases}$$

whose solutions are denoted by  $(t_f^{k+1}, u_{k+1}, x_{k+1}) \in [0, T] \times \mathcal{U} \times L^{\ell}_{\mathcal{F}}(\Omega; C([0, T]; \mathbb{R}^n))$ , for  $\ell \geq 2$ . In particular, trust-region constraints are now imposed on the variable  $t_f$  as well to derive convergence guarantees.

We point out that the results we present in this section are not to be considered as part of the main contribution, but they rather aim at providing insights for the future design of efficient numerical methods

for stochastic optimal control problems which consider free final time and stochastic admissible controls. In particular, extending SCP in the presence of free final time and stochastic controls requires introducing additional assumptions which might seem demanding. We leave the investigation of the validity of this framework under sharper assumptions as a future direction of research.

### 5.1. Refined assumptions and extended result of convergence

The presence of the free final time hinders the well-posedness of the stochastic PMP when applied to each  $\text{LGOCP}_{k+1}^\Delta$ , *i.e.*, Theorem 3.3, and in turn the validity of Theorem 3.6 in such more general setting. To overcome this issue, we need to tighten assumptions  $(A_1)$ – $(A_3)$ . Specifically, although  $(A_1)$  remains unchanged, assumptions  $(A_2)$ – $(A_3)$  are replaced by the followings, respectively:

$(A'_2)$  Mappings  $g$ ,  $H$ , and  $L_i$ ,  $i = 0, \dots, m$ , either are affine-in-state or have compact supports in  $\mathbb{R}^n$  and in  $[0, T] \times \mathbb{R}^n$ , respectively. In the case of free final time,  $g$  is affine.

$(A'_3)$  For every  $k \in \mathbb{N}$ ,  $\text{LGOCP}_{k+1}^\Delta$  is feasible. In addition, in the case of free final time, for every  $k \in \mathbb{N}$ , any optimal control  $u_k$  for  $\text{LGOCP}_k^\Delta$  has continuous sample paths at the optimal final time  $t_f^{k+1}$  for  $\text{LGOCP}_{k+1}^\Delta$ .

Assumption  $(A'_3)$  plays a crucial role to provide the existence of pointwise necessary conditions for optimality for the linearized problems  $\text{LGOCP}_k^\Delta$ , thus enabling classical formulations of the stochastic PMP. However, we do recognize  $(A'_3)$  might be demanding and, as future research direction, we propose to relax it by leveraging integral-type necessary conditions for optimality, which are best suited to deal with optimal control problems which show explicit discontinuous dependence on time (*e.g.*, [15]). Unfortunately, when optimization over stochastic controls is adopted, successfully leveraging weakly converging subsequences of controls to show “weak” guarantee of success for SCP, *i.e.*, the equivalent of Theorem 3.9, becomes more challenging, in which case only Theorem 3.6 still holds, though under appropriate modifications. We leave proving success guarantees for SCP when optimizing over stochastic controls with weaker assumptions as a future direction of research.

Extending the stochastic PMP to GOCP and  $\text{LGOCP}_{k+1}^\Delta$  comes by assuming  $(A_1)$ ,  $(A'_2)$ , and  $(A'_3)$ . In particular, under those assumptions we have the following extension of Theorem 3.3 and of Theorem 3.4:

**Theorem 5.1** (Stochastic Pontryagin Maximum Principle for GOCP). *Let  $(t_f, u, x)$  be a locally-optimal solution to GOCP. There exist  $p \in L^2_{\mathcal{F}}(\Omega; C([0, T]; \mathbb{R}^n))$  and a tuple  $(\mathbf{p}, p^0, q)$ , where  $\mathbf{p} \in \mathbb{R}^q$  and  $p^0 \leq 0$  are constant, and  $q = (q_1, \dots, q_d) \in L^2_{\mathcal{F}}([0, T] \times \Omega; \mathbb{R}^{n \times d})$  such that the following relations are satisfied:*

1. *Non-Triviality Condition:*  $(\mathbf{p}, p^0) \neq 0$ .
2. *Adjoint Equation:*

$$dp(t) = -\frac{\partial H}{\partial x}(t, u(t), x(t), p(t), p^0, q(t)) dt + q(t) dB_t, \quad p(t_f) = \mathbb{E} \left[ \frac{\partial g}{\partial x}(x(t_f)) \right]^\top \mathbf{p} \in \mathbb{R}^n.$$

3. *Maximality Condition:*

$$u(t) = \arg \max_{v \in U} \mathbb{E} \left[ H(t, v, x(t), p(t), p^0, q(t)) \right], \quad \text{a.e. (deterministic controls)}$$

$$u(t) = \arg \max_{v \in U} H(t, v, x(t), p(t), p^0, q(t)), \quad \text{a.e., a.s. (stochastic controls)}$$

4. *Transversality Condition:* if the final time is free

$$\max_{v \in U} \mathbb{E} \left[ H(t_f, v, x(t_f), p(t_f), p^0, q(t_f)) \right] \geq 0 \quad (\text{deterministic controls})$$

$$\mathbb{E} \left[ \max_{v \in U} H(t_f, v, x(t_f), p(t_f), p^0, q(t_f)) \right] \geq 0 \text{ (stochastic controls)}$$

where equalities hold in the case  $t_f < T$ .

The quantity  $(t_f, u, \mathbf{p}, p^0)$  uniquely determines  $x, p,$  and  $q$  and is called *extremal* for GOCP (associated with the tuple  $(t_f, u, x, p, \mathbf{p}, p^0, q)$ , or simply with  $(t_f, u, x)$ ). An extremal  $(t_f, u, \mathbf{p}, p^0)$  is called *normal* if  $p^0 \neq 0$ .

**Theorem 5.2** (Weak Stochastic Pontryagin Maximum Principle for LGOCP $_k^\Delta$ ). *Let  $(t_f^k, u_k, x_k)$  be a locally-optimal solution to LGOCP $_k^\Delta$ . There exist  $p_k \in L^2_{\mathcal{F}}(\Omega; C([0, T]; \mathbb{R}^n))$ , a tuple  $(\mathbf{p}_k, p_k^0, p_k^1)$ , where  $\mathbf{p}_k \in \mathbb{R}^q, p_k^0 \leq 0$ , and  $p_k^1 \in \mathbb{R}$  are constant, and  $q_k = ((q_k)_1, \dots, (q_k)_d) \in L^2_{\mathcal{F}}([0, T] \times \Omega; \mathbb{R}^{n \times d})$  such that the following relations are satisfied:*

1. *Non-Triviality Condition:*  $(\mathbf{p}_k, p_k^0, p_k^1) \neq 0$ .
2. *Adjoint Equation:*

$$dp_k(t) = -\frac{\partial H_k}{\partial x}(t, u_k(t), x_k(t), p_k(t), p_k^0, p_k^1, q_k(t)) dt + q_k(t) dB_t, \quad p_k(t_f^k) = \mathbb{E} \left[ \frac{\partial g_k}{\partial x}(x_k(t_f^k)) \right]^\top \mathbf{p}_k \in \mathbb{R}^n.$$

3. *Maximality Condition:*

$$u_k(t) = \arg \max_{v \in U} \mathbb{E} \left[ H_k(t, v, x_k(t), p_k(t), p_k^0, p_k^1, q_k(t)) \right], \text{ a.e. (deterministic controls)}$$

$$u_k(t) = \arg \max_{v \in U} H_k(t, v, x_k(t), p_k(t), p_k^0, p_k^1, q_k(t)), \text{ a.e., a.s. (stochastic controls)}$$

4. *Transversality Condition: if the final time is free*

$$\max_{v \in U} \mathbb{E} \left[ H_k(t_f^k, v, x_k(t_f^k), p_k(t_f^k), p_k^0, p_k^1, q_k(t_f^k)) \right] \geq 0 \text{ (deterministic controls)}$$

$$\mathbb{E} \left[ \max_{v \in U} H_k(t_f^k, v, x_k(t_f^k), p_k(t_f^k), p_k^0, p_k^1, q_k(t_f^k)) \right] \geq 0 \text{ (stochastic controls)}$$

where equalities hold in the case  $t_f^k < T$ .

The quantity  $(t_f^k, u_k, \mathbf{p}_k, p_k^0, p_k^1)$  uniquely determines  $x_k, p_k,$  and  $q_k$  and is called *extremal* for LOCP $_k^\Delta$  (associated with  $(t_f^k, u_k, x_k, p_k, \mathbf{p}_k, p_k^0, p_k^1, q_k)$ , or  $(t_f^k, u_k, x_k)$ ). An extremal  $(t_f^k, u_k, \mathbf{p}_k, p_k^0, p_k^1)$  is called *normal* if  $p_k^0 \neq 0$ .

Extending the stochastic PMP to the new linearized problems LGOCP $_k^\Delta$ , and in particular the new transversality condition 4., additionally requires to assume  $(A'_3)$  together with  $(A_1)$  and  $(A'_2)$ . Although the proof of this result for fixed-final-time problems is well-established (see [54], Chap. 3), we could not find any published proof of Theorem 5.1 when the final time is free. Therefore, we provide its proof in Section 5.2. The proof of Theorem 5.2 is achieved similarly, thus for the sake of clarity and brevity, we avoid reporting any detail concerning the latter. Thanks to Theorem 5.1, we can extend the convergence of SCP as follows:

**Theorem 5.3** (Generalized Properties of Accumulation Points for SCP). *Assume that  $(A_1)$ ,  $(A'_2)$ , and  $(A'_3)$  hold and that SCP generates a sequence  $(\Delta_k, t_f^k, u_k, x_k)_{k \in \mathbb{N}}$  such that  $(\Delta_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}_+ \setminus \{0\}$  converges to zero, and for every  $k \geq 1$ , the tuple  $(t_f^k, u_k, x_k)$  locally solves LGOCP $_k^\Delta$ . For every  $k \geq 1$ , letting  $(t_f^k, u_k, \mathbf{p}_k, p_k^0, p_k^1)$  be an extremal associated with  $(t_f^k, u_k, x_k)$  for LGOCP $_k^\Delta$  (whose existence is ensured by Theorem 5.2), assume the following Accumulation Condition holds:*



(AC) *Up to some subsequence,  $(t_f^k, u_k, \mathbf{p}_k, p_k^0, p_k^1)$  converges to some  $(u, \mathbf{p}, p^0, p^1) \in L^2([0, T]; \mathbb{R}^m) \times \mathbb{R}^{q+2}$  for the strong topology of  $L^2_{\mathcal{F}}([0, T] \times \Omega; \mathbb{R}^m) \times \mathbb{R}^{q+2}$ .*

*If  $(\mathbf{p}, p^0) \neq 0$ , then  $(t_f^k, u, \mathbf{p}, p^0)$  is an extremal for OCP associated with  $(t_f^k, u, x_u)$ .*

The guarantees offered by Theorem 5.3 read similarly to what we have explained in Section 3.3. One sees that the computations provided in Section 4.2 for the proof of Theorem 3.6 straightforwardly generalize as soon as weak convergence of controls in  $L^2([0, T]; \mathbb{R}^m)$  is replaced with strong convergence of controls in  $L^2_{\mathcal{F}}([0, T] \times \Omega; \mathbb{R}^m)$  (actually, the proofs become even simpler), and may be endorsed to prove Theorem 5.3, provided that we are capable to extend the proof of Theorems 3.3 and 3.4 in the case of free final time and stochastic controls. Therefore, to conclude, in the next section we develop the necessary technical details which enable proving Theorems 5.1 and 5.2 through the machinery developed in Section 4.2. Specifically, since the proofs of Theorems 5.1 and 5.2 are similar, for the sake of clarity and brevity we only provide details for the proof of Theorem 5.1.

## 5.2. Proof of the extension for the stochastic Pontryagin Maximum Principle

Before getting started, we need to introduce the notion of a *Lebesgue point* for a stochastic control  $u \in L^2_{\mathcal{F}}([0, T] \times \Omega; \mathbb{R}^m)$ . For this, we adopt the theory of Bochner integrals, showing that  $L^2_{\mathcal{F}}([0, T] \times \Omega; \mathbb{R}^m) \subseteq L^2([0, T]; L^2_{\mathcal{G}}(\Omega; \mathbb{R}^m))$ , where the latter is the space of Bochner integrable mappings  $u : [0, T] \rightarrow L^2_{\mathcal{G}}(\Omega; \mathbb{R}^m)$ . Let  $u \in L^2_{\mathcal{F}}([0, T] \times \Omega; \mathbb{R}^m)$ . First of all, by definition we see that for almost every  $t \in [0, T]$ , it holds that  $\mathbb{E}[\|u(t)\|^2] < \infty$ , and thus this control is well-defined as a mapping  $u : [0, T] \rightarrow L^2_{\mathcal{G}}(\Omega; \mathbb{R}^m)$ . Since  $\Omega$  is second-countable,  $L^2_{\mathcal{G}}(\Omega; \mathbb{R}^m)$  is separable, and therefore the claim follows from the Pettis measurability theorem once we prove that  $u : [0, T] \rightarrow L^2_{\mathcal{G}}(\Omega; \mathbb{R}^m)$  is strongly measurable with respect to the Lebesgue measure of  $[0, T]$ . For this, it is sufficient to show that for every  $A \in \mathcal{B}([0, T]) \otimes \mathcal{G}$  and  $\alpha \in L^2_{\mathcal{G}}(\Omega; \mathbb{R}^m)$ , the mapping  $t \mapsto \mathbb{E}[\mathbb{1}_A(t, \omega)\alpha(\omega)]$  is Lebesgue measurable. By fixing  $\alpha \in L^2_{\mathcal{G}}(\Omega; \mathbb{R}^m)$ , this can be achieved by proving that the family  $\{A \in \mathcal{B}([0, T]) \otimes \mathcal{G} : t \mapsto \mathbb{E}[\mathbb{1}_A(t, \omega)\alpha(\omega)] \text{ is Lebesgue measurable}\}$  is a monotone class and then using standard monotone class arguments (the details are left to the reader). At this step, the Lebesgue differentiation theorem provides that for almost every  $t \in [0, T]$ , the following relations hold:

$$\lim_{\eta \rightarrow 0} \frac{1}{\eta} \int_t^{t+\eta} \mathbb{E}[\|u(s) - u(t)\|] ds = 0, \quad \lim_{\eta \rightarrow 0} \frac{1}{\eta} \int_t^{t+\eta} \mathbb{E}[\|u(s) - u(t)\|^2] ds = 0.$$

Such a time  $t \in [0, T]$  is called Lebesgue point for the control  $u \in L^2_{\mathcal{F}}([0, T] \times \Omega; \mathbb{R}^m)$ .

### 5.2.1. Modified needle-like variations and end-point mapping

We may extend the concept of needle-like variations and of end-point mapping previously introduced in Section 4.1.2 to the setting of free final time and of stochastic controls as follows.

Given an integer  $j \in \mathbb{N}$ , fix  $j$  times  $0 < t_1 < \dots < t_j < t_f$  which are Lebesgue points for  $u$ , and fix  $j$  random variables  $u_1, \dots, u_j$  such that  $u_i \in L^2_{\mathcal{F}_{t_i}}(\Omega; U)$ . For fixed scalars  $0 \leq \eta_i < t_{i+1} - t_i$ ,  $i = 1, \dots, j-1$ , and  $0 \leq \eta_j < t_f - t_j$ , the needle-like variation  $\pi = \{t_i, \eta_i, u_i\}_{i=1, \dots, j}$  of the control  $u$  is defined to be the admissible control  $u_{\pi}(t) = u_i$  if  $t \in [t_i, t_i + \eta_i]$  and  $u_{\pi}(t) = u(t)$  otherwise. Denote by  $\tilde{x}_v$  the solution related to an admissible control  $v$  of the augmented system

$$\begin{cases} dx(t) = b(t, v(t), x(t)) dt + \sigma(t, x(t)) dB_t, & x(0) = x^0 \\ dx^{n+1}(t) = f^0(t, v(t), x(t)) dt, & x^{n+1}(0) = 0 \end{cases}$$

and define the mapping  $\tilde{g} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{q+1} : \tilde{x} \mapsto (g(x), x^{n+1})$ . For every fixed time  $t \in (t_j, t_f]$ , by denoting  $\delta_t \triangleq \min\{t_{i+1} - t_i, t - t_j, T - t : i = 1, \dots, j-1\} > 0$ , the end-point mapping at time  $t$  is defined to be

$$F_t^j : \mathcal{C}_t^j \triangleq B_{\delta_t}^{j+1}(0) \cap (\mathbb{R} \times \mathbb{R}_+^j) \rightarrow \mathbb{R}^{q+1} \\ (\delta, \eta_1, \dots, \eta_j) \mapsto \mathbb{E}[\tilde{g}(\tilde{x}_{u_\pi}(t + \delta))] - \mathbb{E}[\tilde{g}(\tilde{x}_u(t))]$$

where  $B_\rho^{j+1}$  is the open ball in  $\mathbb{R}^{j+1}$  of radius  $\rho > 0$ . Variations on the variable  $\delta$  are necessary only if free-final-time problems are considered, in which case  $(A'_2)$ , and in particular the fact that  $g$  is an affine function, play a crucial role for computations. Due to Lemma 2.1 (and to  $(A'_2)$  in the case of free-final-time problems), it is not difficult to see that  $F_t^j$  is Lipschitz (see also the argument developed to prove Lem. 5.4). In addition, this mapping may be Gateaux differentiated at zero along admissible directions of the cone  $\mathcal{C}_t^j$ . For this, denote  $\tilde{b} = (b^\top, f^0)^\top$ ,  $\tilde{\sigma} = (\sigma^\top, 0)^\top$  and let  $z_{t_i, u_i}$  be the unique solution to (4.1) with  $\xi_{t_i} = \tilde{b}(t_i, u_i, x(t_i)) - \tilde{b}(t_i, u(t_i), x(t_i))$ .

**Lemma 5.4** (Generalized stochastic needle-like variation formula). *Let  $(\delta, \eta_1, \dots, \eta_j) \in \mathcal{C}_t^j$  and assume  $(A_2)$  holds (in particular,  $\tilde{g}$  is an affine function when  $\delta \neq 0$ ). If  $t > t_j$  is a Lebesgue point for  $u$ , then it holds that*

$$\left\| \mathbb{E} \left[ \tilde{g}(\tilde{x}_{u_\pi}(t + \delta)) - \tilde{g}(\tilde{x}_u(t)) - \delta \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(t)) \tilde{b}(t, u(t), x_u(t)) - \sum_{i=1}^j \eta_i \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(t)) z_{t_i, u_i}(t) \right] \right\| = o \left( \delta + \sum_{i=1}^j \eta_i \right).$$

As for Lemma 4.2, we provide an extensive proof of Lemma 5.4 in the appendix.

### 5.2.2. Variational inequalities and conclusion

Similarly to what we have argued in Section 5.2.2, the main step in the proof of the stochastic PMP with free final time and stochastic control goes by contradiction, leveraging Lemma 5.4. For this, from now on we assume that when free, the final time is a Lebesgue point for the optimal control. Otherwise, one may proceed by mimicking the argument developed in Section 7.3 of [23].

Assume that  $(A_1)$  and  $(A'_2)$  hold. For every  $j \in \mathbb{N}$ , define the linear mapping

$$dF_{t_f}^j(\delta, \eta) = \delta \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(t_f)) \tilde{b}(t_f, u(t_f), x_u(t_f)) \right] + \sum_{i=1}^j \eta_i \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(t_f)) z_{t_i, u_i}(t_f) \right],$$

which due to Lemma 5.4, satisfies

$$\lim_{\alpha > 0, \alpha \rightarrow 0} \frac{F_{t_f}^j(\alpha(\delta, \eta))}{\alpha} = dF_{t_f}^j(\delta, \eta),$$

for every  $(\delta, \eta) \in \mathbb{R} \times \mathbb{R}_+^j$ . Finally, consider the closed, convex cone of  $\mathbb{R}^{q+1}$  given by

$$K \triangleq \text{Cl} \left( \text{Cone} \left\{ \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(t_f)) z_{t_i, u_i}(t_f) \right], \pm \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(t_f)) \tilde{b}(t_f, u(t_f), x_u(t_f)) \right] \right. \right. \\ \left. \left. \text{for } u_i \in U \text{ and } t_i \in (0, t_f) \text{ is Lebesgue for } u \right\} \right).$$

If  $K = \mathbb{R}^{q+1}$ , it would hold  $dF_{t_f}^j(\mathbb{R} \times \mathbb{R}_+^j) = K = \mathbb{R}^{q+1}$ , and by Lemma 12.1 of [1], one would find that the origin is an interior point of  $F_{t_f}^j(\mathcal{C}_{t_f}^j)$ . This would imply that  $(t_f, u, x)$  is not optimal for OCP, a contradiction.

The argument above (together with an application of the separation plane theorem) provides the existence of a non-zero vector denoted  $\tilde{\mathbf{p}} = (\mathbf{p}^\top, \mathbf{p}^0) \in \mathbb{R}^{q+1}$  such that the following variational inequalities hold

$$\begin{cases} \tilde{\mathbf{p}}^\top \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(t_f)) \tilde{b}(t_f, u(t_f), x_u(t_f)) \right] = 0 \\ \tilde{\mathbf{p}}^\top \mathbb{E} \left[ \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(t_f)) z_{r,v}(t_f) \right] \leq 0, \quad r \in [0, t_f] \text{ is Lebesgue for } u, v \in L^2_{\mathcal{F}_r}(\Omega; U). \end{cases} \quad (5.1)$$

In the case of deterministic controls, the random variables  $v \in L^2_{\mathcal{F}_r}(\Omega; U)$  in (5.1) are replaced by deterministic vectors  $v \in U$ . Moreover, when  $t_f = T$ , only negative variations on the final time are allowed. Hence in this case, the first equality of (5.1) actually becomes a greater-or-equal-to-zero inequality (see also [23], Chap. 7).

The rest of the proof remains unchanged, *i.e.*, it follows the argument of Section 4.1.4 verbatim, and the proof of Theorem 5.3 may be readily inferred by the computations of Section 4.2.

## 6. AN EXAMPLE NUMERICAL SCHEME

Although the procedure detailed previously provides methodological steps to tackle OCP through successive linearizations, numerically solving LCPs that depend on stochastic coefficients still remains a challenge. In this last section, under appropriate assumptions we propose an approximate, though very practical, numerical scheme to effectively solve each subproblem  $\text{LOCP}_k^\Delta$ . We stress the fact that our main goal is not the development of an ultimate algorithm, but rather consists of demonstrating how one may leverage the theoretical insights provided by Theorem 3.6 to design efficient strategies to practically solve OCP.

### 6.1. A simplified context

The proposed approach relies on a specific shape of the cost and the dynamics of OCP. Specifically, we consider OCPs with deterministic admissible controls  $u \in \mathcal{U} = L^2([0, T]; U)$  (over a fixed time horizon  $[0, T]$ ) only and whose cost functions  $f^0$  are such that  $H = 0$ . Moreover, we assume the state variable is given by two components  $(x, z) \in \mathbb{R}^{n_x + n_z}$  for  $n_x, n_z \in \mathbb{N}$ , satisfying the following system of forward stochastic differential equations ( $b^x$  and  $b^z$  are accordingly defined as in (2.1))

$$\begin{cases} dx(t) = b^x(t, u(t), x(t), z(t)) dt + \sigma(t, z(t)) dB_t, & x(0) = x^0 \\ dz(t) = b^z(t, u(t), z(t)) dt, & z(0) = z^0 \in \mathbb{R}^{n_z}. \end{cases} \quad (6.1)$$

In particular, any  $\mathcal{F}$ -adapted process  $(x, z)$  solution to (6.1) with continuous sample paths for a given control  $u \in \mathcal{U}$  is such that  $z$  is deterministic. For the sake of clarity and to avoid cumbersome notation, from now on we assume  $g = 0$  and that  $b^x$  does not explicitly depend on  $z$ . This is clearly done without loss of generality.

### 6.2. The proposed approach

With the assumptions adopted previously, we see that the diffusion in the dynamics of OCP is now forced to be deterministic. This fact is at the root of our method, which mimics the procedure proposed in [4]. Specifically, we transcribe every stochastic subproblem  $\text{LOCP}_k^\Delta$  into a deterministic and convex optimal control problem, whereby the variables are the mean and the covariance of the solution  $x_k$  to  $\text{LOCP}_k^\Delta$ . The main advantage in doing so is that deterministic reformulations of the subproblems  $\text{LOCP}_k^\Delta$  can be efficiently solved *via* off-the-shelf convex solvers. Unlike [4], here we rely on some *upstream information* for the design of this numerical scheme which is entailed by Theorem 3.6 as follows.

By recalling the notation introduced in the previous sections, we denote  $\mu(t) \triangleq \mathbb{E}[x(t)]$ ,  $\Sigma(t) \triangleq \mathbb{E}[(x(t) - \mu(t))(x(t) - \mu(t))^\top]$ , and for  $k \in \mathbb{N}$ ,  $\mu_k(t) \triangleq \mathbb{E}[x_k(t)]$  and  $\Sigma_k(t) \triangleq \mathbb{E}[(x_k(t) - \mu_k(t))(x_k(t) - \mu_k(t))^\top]$ . Heuristically,

assuming that solutions to  $\text{LOCP}_k^\Delta$  have *small variance*, i.e.,  $\|\text{tr } \Sigma_k\|_{L^2} \ll 1$ , one can compute the linearization of  $b^x$ ,  $b^z$ ,  $\sigma$ , and  $L$  at  $\mu_k$  rather than at  $x_k$ . In doing so, for the cost of  $\text{LOCP}_{k+1}^\Delta$  we obtain the following approximation (the notation goes accordingly as in (3.1))

$$\mathbb{E} \left[ \int_0^T f_{k+1}^0(s, u(s), x(s), z(s)) ds \right] \approx \int_0^T f_{u_k, \mu_k}^0(s, u(s), \mu(s), z(s)) ds, \quad (6.2)$$

whereas for the dynamics of  $\text{LOCP}_{k+1}^\Delta$  (the notation goes accordingly as in (2.2))

$$\begin{cases} dx(t) \approx b_{u_k, \mu_k}^x(t, u(t), x(t)) dt + \sigma_{z_k}(t, z(t)) dB_t \\ \quad = \left( \mathcal{A}_{k+1}(t)x(t) + \mathcal{B}_{k+1}(t, u(t)) \right) dt + \mathcal{C}_{k+1}(t, z(t)) dB_t \\ dz(t) \approx b_{u_k, z_k}^z(t, u(t), z(t)) dt = \left( \mathcal{D}_{k+1}(t)z(t) + \mathcal{E}_{k+1}(t, u(t)) \right) dt \end{cases} \quad (6.3)$$

where  $\mathcal{A}_{k+1}(t) \in \mathbb{R}^{n_x \times n_x}$ ,  $\mathcal{B}_{k+1}(t, u) \in \mathbb{R}^{n_x}$ ,  $\mathcal{C}_{k+1}(t, z) \in \mathbb{R}^{n_x}$ ,  $\mathcal{D}_{k+1}(t) \in \mathbb{R}^{n_z \times n_z}$ , and  $\mathcal{E}_{k+1}(t, u) \in \mathbb{R}^{n_z}$  are deterministic and with  $\mathcal{C}_{k+1}(t, z)$  affine in  $z$ . Accordingly, by introducing  $\mu^{e_k}(t) \triangleq \mu(t) - \mu_k(t)$  and  $\Sigma^{e_k}(t) \triangleq \mathbb{E}[(x(t) - x_k(t) - \mu(t) + \mu_k(t))(x(t) - x_k(t) - \mu(t) + \mu_k(t))^\top]$ , slightly tighter trust-region constraints are

$$\int_0^T \text{tr } \Sigma^{e_k}(t) dt + \int_0^T \|\mu^{e_k}(t)\|^2 dt \leq \Delta_{k+1}, \quad |t_f^{k+1} - t_f^k| \leq \Delta_{k+1}. \quad (6.4)$$

At this point, as all coefficients are deterministic, solutions to (6.3) are Gaussian processes whose dynamics take the form (see, e.g., [38])

$$\begin{cases} \dot{\mu}(t) = \mathcal{A}_{k+1}(t)\mu(t) + \mathcal{B}_{k+1}(t, u(t)), & \dot{z}(t) = \mathcal{D}_{k+1}(t)z(t) + \mathcal{E}_{k+1}(t, u(t)) \\ \dot{\Sigma}(t) = \mathcal{A}_{k+1}(t)\Sigma(t) + \Sigma(t)\mathcal{A}_{k+1}(t)^\top + \mathcal{C}_{k+1}(t, z(t))\mathcal{C}_{k+1}(t, z(t))^\top. \end{cases} \quad (6.5)$$

The system above is not linear because of  $\mathcal{C}_{k+1}(t, z(t))\mathcal{C}_{k+1}(t, z(t))^\top$ . Nevertheless, we may call upon the convergences of Theorem 3.6, and in particular the convergence of the sequence  $(z_k)_{k \in \mathbb{N}}$  to some deterministic curve  $z$ , to replace (6.5) with

$$\begin{cases} \dot{\mu}(t) = \mathcal{A}_{k+1}(t)\mu(t) + \mathcal{B}_{k+1}(t, u(t)), & \dot{z}(t) = \mathcal{D}_{k+1}(t)z(t) + \mathcal{E}_{k+1}(t, u(t)) \\ \dot{\Sigma}(t) = \mathcal{A}_{k+1}(t)\Sigma(t) + \Sigma(t)\mathcal{A}_{k+1}(t)^\top + \mathcal{C}_{k+1}(t, z_k(t))\mathcal{C}_{k+1}(t, z_k(t))^\top. \end{cases} \quad (6.6)$$

In conclusion, we may heuristically replace every  $\text{LOCP}_{k+1}^\Delta$  with the deterministic convex optimal control problem whose dynamics are (6.6), whose trust-region constraints are (6.4), whose variables are  $\mu$  and  $\Sigma$  (and additionally  $\mu^{e_k}$  and  $\Sigma^{e_k}$ ), and whose cost consists of replacing (6.2) with

$$\int_0^T \left( f_{u_k, \mu_k}^0(s, u(s), \mu(s), z(s)) + \text{tr } \Sigma(t) \right) ds$$

to force solutions to have small variances, which in turn justifies the whole approach.

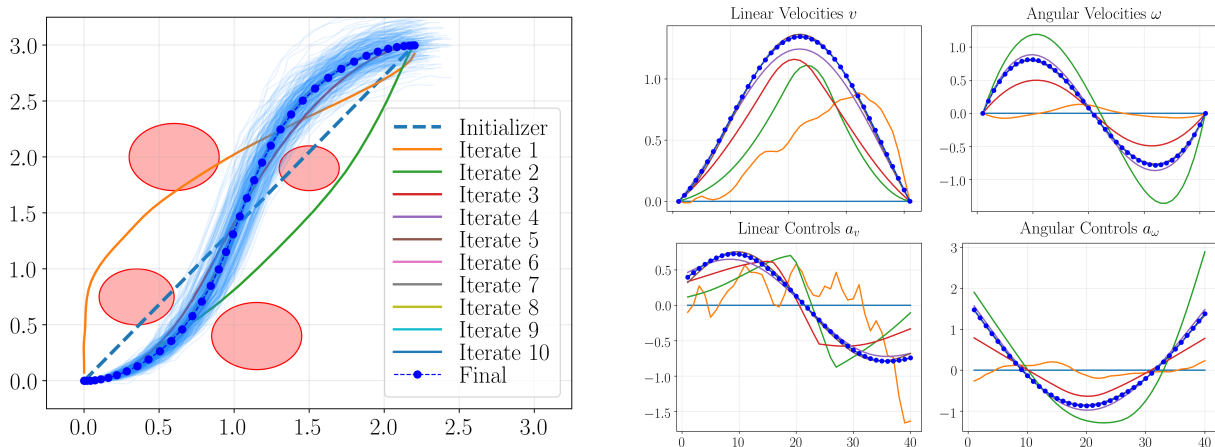


FIGURE 1. Left: solution of each SCP iteration and 100 sample paths of the resulting final trajectory. Right: velocities and control inputs at each SCP iteration.

### 6.3. Uncertain car trajectory planning problem

We now provide numerical experiments. We consider the problem of planning the trajectory of a car whose state and control inputs are  $x = (r_x, r_y, \theta, v, \omega)$  and  $u = (a_v, a_\omega)$  and whose dynamics are

$$b(t, x, u) = (v \cos(\theta), v \sin(\theta), \omega, a_v, a_\omega), \quad \sigma(t, x) = \text{diag}([\alpha^2 \omega v, \alpha^2 \omega v, \beta^2 \omega v, 0, 0]),$$

where  $\alpha^2 = 0.1$  and  $\beta^2 = 0.01$  quantify the effect of slip. The evolution of  $(v, \omega)$  is deterministic: given actuator commands, the change in velocity is known exactly, but uncertainty in positional variables persists. This model is motivated by driving applications where the vehicle may drift during tight turns and suffer from important positional uncertainty, but limits its acceleration to avoid wheel slip. By defining  $z = (v, \omega)$ , this problem setting matches the one presented in the previous section. We consider minimizing control effort  $G(u(s)) = \|u(s)\|_2^2$ . The initial state  $x^0$  is fixed, and the final state constraint consists of reaching the goal  $x^f$  in expectation such that  $g(x(t_f)) = x(t_f) - x^f$ . Further, we consider cylindrical obstacles of radius  $\delta_o > 0$  centered at  $r_o \in \mathbb{R}^2$  for which we define the potential function  $c_o : \mathbb{R}^n \rightarrow \mathbb{R}$  as  $c_o(x) = \|r - r_o\|^2 - (\delta_o + \varepsilon)^2$  if  $\|r - r_o\| < (\delta_o + \varepsilon)$  and 0 otherwise, where  $r = [r_x, r_y] \subset x$  and  $\varepsilon = 0.1$  is an additional clearance. We penalize obstacle avoidance constraint violation directly within the cost, defining  $L(s, x(s), u(s)) = L_0(s, x(s)) = \lambda c_o(x(s))$ , with  $\lambda = 500$  (note that  $H = 0$  in this setting, following Sect. 6.1).

### 6.4. Results

We set  $x(0) = [0, \dots, 0]$ ,  $x^f = [2.2, 3, 0, 0, 0]$ , and  $t_f = 5$  s, and we set up four obstacles. We discretize (6.6) using a forward Euler scheme with  $N = 41$  discretization nodes, set  $\Delta_0 = 100$ , set  $\Delta_{k+1} = 0.99\Delta_k$  at each SCP iteration, and use IPOPT to solve each convexified problem. We check convergence of SCP by verifying that  $\int_0^{t_f} \|u_{k+1} - u_k\|^2(s) + \|u_k - u_{k-1}\|^2(s) ds \leq 10^{-3}$ . Although our method is initialized with an unfeasible straight-line initial trajectory, SCP converges in 10 iterations, with a trajectory avoiding the obstacles in expectation. Indeed, after evaluating  $10^4$  sample paths of the system, only 6.14% of the trajectories intersect with obstacles. We also verify that  $p_k^0 \neq 0$  at every SCP iteration, and at convergence, so that  $(\mathbf{p}, p^0) \neq 0$ . Thus, Theorems 3.6 and 3.9 apply and the solution generated by SCP is an extremal for OCP. We visualize our results in Figure 1 and release our implementation at <https://github.com/StanfordASL/stochasticSCP>.

## 7. CONCLUSION AND PERSPECTIVES

In this paper we introduced and analyzed convergence properties for sequential convex programming for non-linear stochastic optimal control, from which we derived a practical numerical framework to solve non-linear stochastic optimal control problems.

Future work may consider extending this analysis to tackle more general problem formulations, *e.g.*, risk measures as costs and state (chance) constraints. In this context, some preliminary results using SCP only exist for discrete time problem formulations [34]. However, tackling continuous time formulations will require more sophisticated necessary conditions for optimality. In addition, we plan to better investigate the setting of free final time and stochastic admissible controls, thus devising sharper assumptions and improving the convergence result. Finally, we plan to further leverage our theoretical insights to design new and more efficient numerical schemes for non-linear stochastic optimal control.

## APPENDIX A

### A.1 Proof of Lemma 2.1

*Proof of Lemma 2.1.* For the sake of clarity in the notation, we only consider the case for which  $d = 1$ , the case with multivariate Brownian motion being similar.

Let us start with the first inequality. For this, by denoting  $x = x_{u,v,y}$ , for every  $t \in [0, T]$  we compute (below,  $C \geq 0$  is an appropriate constant)

$$\begin{aligned} & \mathbb{E} \left[ \sup_{s \in [0, t]} \|x(s)\|^\ell \right] \leq C \mathbb{E} \left[ \|x^0\|^\ell \right] + C \mathbb{E} \left[ \sup_{r \in [0, t]} \left\| \int_0^r \sigma(s, y(s)) \, dB_s \right\|^\ell \right] \\ & + C \mathbb{E} \left[ \sup_{r \in [0, t]} \left\| \int_0^r \left( b_0(s, y(s)) + \sum_{i=1}^m u^i(s) b_i(s, y(s)) \right) \, ds \right\|^p \right] \\ & + C \mathbb{E} \left[ \sup_{r \in [0, t]} \left\| \int_0^r \left( \frac{\partial b_0}{\partial x}(s, y(s)) + \sum_{i=1}^m v^i(s) \frac{\partial b_i}{\partial x}(s, y(s)) \right) (x(s) - y(s)) \, ds \right\|^\ell \right] \\ & + C \mathbb{E} \left[ \sup_{r \in [0, t]} \left\| \int_0^r \frac{\partial \sigma}{\partial x}(s, y(s)) (x(s) - y(s)) \, dB_s \right\|^\ell \right]. \end{aligned}$$

For the last term, by denoting  $S_{y,\sigma} \triangleq \left\{ (s, \omega) \in [0, T] \times \Omega : (s, y(s, \omega)) \in \text{supp } \sigma \right\}$ , Burkholder–Davis–Gundy, Hölder, and Young inequalities give

$$\begin{aligned} & \mathbb{E} \left[ \sup_{r \in [0, t]} \left\| \int_0^r \frac{\partial \sigma}{\partial x}(s, y(s)) (x(s) - y(s)) \, dB_s \right\|^p \right] \\ & \leq C \left( \int_0^t \mathbb{E} \left[ \sup_{s \in [0, r]} \|x(r)\|^p \right] \, ds + \int_{S_{y,\sigma}} \left\| \frac{\partial \sigma}{\partial x}(s, y(s)) \right\|^p \|y(s)\|^p \, d(s \times P) \right) \\ & \leq C \left( 1 + \int_0^t \mathbb{E} \left[ \sup_{s \in [0, r]} \|x(r)\|^p \right] \, ds \right). \end{aligned}$$

Similar computations apply to the other terms and when considering solutions to (2.1). Therefore, we conclude from a routine Grönwall inequality argument.

Let us prove the second inequality of the lemma. For  $t \in [0, T]$ , we compute

$$\begin{aligned} \mathbb{E} \left[ \sup_{s \in [0, t]} \|x_{u_1}(s) - x_{u_2}(s)\|^p \right] &\leq C \mathbb{E} \left[ \sup_{s \in [0, t]} \left\| \int_0^r (b_0(s, x_{u_1}(s)) - b_0(s, x_{u_2}(s))) \, ds \right\|^p \right] \\ &+ C \sum_{i=1}^m \mathbb{E} \left[ \sup_{s \in [0, t]} \left\| \int_0^r (u_1^i(s) b_i(s, x_{u_1}(s)) - u_2^i(s) b_i(s, x_{u_2}(s))) \, ds \right\|^p \right] \\ &+ C \mathbb{E} \left[ \sup_{s \in [0, t]} \left\| \int_0^r (\sigma(s, x_{u_1}(s)) - \sigma(s, x_{u_2}(s))) \, dB_s \right\|^p \right]. \end{aligned}$$

For the second term on the right-hand side, for  $i = 1, \dots, m$  Hölder inequality gives

$$\begin{aligned} &\mathbb{E} \left[ \sup_{s \in [0, t]} \left\| \int_0^r (u_1^i(s) b_i(s, x_{u_1}(s)) - u_2^i(s) b_i(s, x_{u_2}(s))) \, ds \right\|^p \right] \\ &\leq C \left( \int_0^t \mathbb{E} \left[ \sup_{r \in [0, s]} \|x_{u_1}(r) - x_{u_2}(r)\|^p \right] ds + \mathbb{E} \left[ \left( \int_0^T \|u_1(s) - u_2(s)\| \, ds \right)^p \right] \right), \end{aligned}$$

and similar computations hold for the remaining terms and when considering solutions to (2.1). Again, we conclude by a Grönwall inequality argument.  $\square$

## A.2 Proof of Lemma 4.2 and of Lemma 5.4

The proof of Lemma 4.2 immediately follows from the following preliminary result.

**Lemma A.1** (Stochastic needle-like variation formula - no free final time). *Let  $(\eta_1, \dots, \eta_j) \in \text{Pr}_{\mathbb{R}_+^j}(\mathcal{C}_t^j)$  (in particular, no  $(A'_2)$  or any assumption on  $t > t_j$  are required). For  $\varepsilon \in [0, t - t_j]$ , uniformly for  $\delta \in [-\varepsilon, T - t]$ ,*

$$\mathbb{E} \left[ \left\| \tilde{g}(\tilde{x}_{u_\pi}(t + \delta)) - \tilde{g}(\tilde{x}_u(t + \delta)) - \sum_{i=1}^j \eta_i \frac{\partial \tilde{g}}{\partial \tilde{x}}(\tilde{x}_u(t + \delta)) z_{t_i, u_i}(t + \delta) \right\|^2 \right]^{\frac{1}{2}} = o \left( \sum_{i=1}^j \eta_i \right).$$

*Proof.* We only consider the case  $j = 1$ , being the most general case  $j > 1$  done by adopting a classical induction argument (see, e.g., [1]). We only need to prove that

$$\mathbb{E} \left[ \|\tilde{x}_{u_\pi}(t + \delta) - \tilde{x}_u(t + \delta) - \eta_1 z_{t_1, u_1}(t + \delta)\|^2 \right] = o(\eta_1^2) \tag{A.1}$$

uniformly for every  $\delta \in [-\varepsilon, T - t]$ . For this, first we remark that  $\tilde{x}_{u_\pi}(t) = \tilde{x}_u(t)$  for  $t \in [0, t_1]$ . Since  $t \in (t_j, t_f]$  and  $\varepsilon \geq 0$  are fixed and we take the limit  $\eta_1 \rightarrow 0$ , we may assume that  $\eta_1 + t_1 < t - \varepsilon$ . Therefore, without loss of generality, we replace  $t + \delta$  with  $t$ , assuming  $t \geq t_1 + \eta_1$  uniformly. We have (below,  $C \geq 0$  denotes a constant)

$$\begin{aligned} \mathbb{E} \left[ \|\tilde{x}_{u_\pi}(t) - \tilde{x}_u(t) - \eta_1 z_{t_1, u_1}(t)\|^2 \right] &\leq C \mathbb{E} \left[ \left\| \int_{t_1}^{t_1 + \eta_1} \eta_1 A(s) z_{t_1, u_1}(s) \, ds \right\|^2 \right] \\ &+ C \mathbb{E} \left[ \left\| \int_{t_1 + \eta_1}^t (\tilde{b}(s, u_\pi(s), x_{u_\pi}(s)) - \tilde{b}(s, u(s), x_u(s)) - \eta_1 A(s) z_{t_1, u_1}(s)) \, ds \right\|^2 \right] \\ &+ C \mathbb{E} \left[ \left\| \int_0^t \mathbb{1}_{(t_1, T]}(s) (\tilde{\sigma}(s, x_{u_\pi}(s)) - \tilde{\sigma}(s, x_u(s)) - \eta_1 D(s) z_{t_1, u_1}(s)) \, dB_s \right\|^2 \right] \end{aligned}$$

$$+ C\mathbb{E} \left[ \left\| \int_{t_1}^{t_1+\eta_1} \left( \tilde{b}(s, u_\pi(s), x_{u_\pi}(s)) - \tilde{b}(s, u(s), x_u(s)) \right) ds - \eta_1 z_{t_1, u_1}(t_1) \right\|^2 \right].$$

Let us analyze those integrals separately. Starting with the last one, we have

$$\begin{aligned} & \mathbb{E} \left[ \left\| \int_{t_1}^{t_1+\eta_1} \left( \tilde{b}(s, u_\pi(s), x_{u_\pi}(s)) - \tilde{b}(s, u(s), x_u(s)) \right) ds - \eta_1 z_{t_1, u_1}(t_1) \right\|^2 \right] \\ & \leq C\mathbb{E} \left[ \left\| \int_{t_1}^{t_1+\eta_1} \left( \tilde{b}(s, u_1, x_{u_\pi}(s)) - \tilde{b}(s, u_1, x_u(s)) \right) ds \right\|^2 \right] \\ & \quad + C\mathbb{E} \left[ \left\| \int_{t_1}^{t_1+\eta_1} \left( \tilde{b}(s, u_1, x_u(s)) - \tilde{b}(s, u(s), x_u(s)) \right) ds - \eta_1 z_{t_1, u_1}(t_1) \right\|^2 \right]. \end{aligned}$$

Lemma 2.1 immediately gives that the first term on the right-hand side is  $o(\eta_1^2)$ . In addition, from Hölder inequality, it follows that

$$\begin{aligned} & \frac{1}{\eta_1^2} \mathbb{E} \left[ \left\| \int_{t_1}^{t_1+\eta_1} \left( \tilde{b}(s, u_1, x_u(s)) - \tilde{b}(s, u(s), x_u(s)) \right) ds - \eta_1 z_{t_1, u_1}(t_1) \right\|^2 \right] \\ & \leq \frac{C}{\eta_1} \int_{t_1}^{t_1+\eta_1} \mathbb{E} \left[ \left\| \tilde{b}(s, u_1, x_u(s)) - \tilde{b}(t_1, u_1, x_u(t_1)) \right\|^2 \right] ds \\ & \quad + \frac{C}{\eta_1} \int_{t_1}^{t_1+\eta_1} \mathbb{E} \left[ \left\| \tilde{b}(s, u(s), x_u(s)) - \tilde{b}(t_1, u(t_1), x_u(t_1)) \right\|^2 \right] ds. \end{aligned}$$

Since  $t_1$  is a Lebesgue point for  $u$ , the two terms above go to zero as  $\eta_1 \rightarrow 0$ . Next, by the Burkholder–Davis–Gundy inequality and a Taylor development, we have

$$\begin{aligned} & \mathbb{E} \left[ \left\| \int_0^t \mathbb{1}_{(t_1, T]}(s) \left( \tilde{\sigma}(s, x_{u_\pi}(s)) - \tilde{\sigma}(s, x_u(s)) - \eta_1 D(s) z_{t_1, u_1}(s) \right) dB_s \right\|^2 \right] \\ & \leq C\mathbb{E} \left[ \int_{t_1}^t \|D(s) (\tilde{x}_{u_\pi}(s) - \tilde{x}_u(s) - \eta_1 z_{t_1, u_1}(s))\|^2 ds \right] \\ & \quad + C\mathbb{E} \left[ \int_{t_1}^t \left\| \int_0^1 \theta \frac{\partial^2 \tilde{\sigma}}{\partial \tilde{x}^2} (s, \theta \tilde{x}_u(s) + (1-\theta)(\tilde{x}_{u_\pi} - \tilde{x}_u)(s)) (\tilde{x}_{u_\pi} - \tilde{x}_u)^2(s) d\theta \right\|^2 ds \right] \\ & = C \int_{t_1}^t \mathbb{E} \left[ \|\tilde{x}_{u_\pi}(s) - \tilde{x}_u(s) - \eta_1 z_{t_1, u_1}(s)\|^2 \right] ds + o(\eta_1^2), \end{aligned}$$

from Lemma 2.1. Similar estimates hold for the remaining terms in the first inequality of the proof. Summarizing, there exists a constant  $C \geq 0$  such that

$$\mathbb{E} \left[ \|\tilde{x}_{u_\pi}(t) - \tilde{x}_u(t) - \eta_1 z_{t_1, u_1}(t)\|^2 \right] \leq o(\eta_1^2) + C \int_{t_1}^t \mathbb{E} \left[ \|\tilde{x}_{u_\pi}(s) - \tilde{x}_u(s) - \eta_1 z_{t_1, u_1}(s)\|^2 \right] ds$$

and the conclusion follows from a routine Grönwall inequality argument.  $\square$

*Proof of Lemma 5.4.* We may assume  $\delta \neq 0$  and  $(A'_2)$ . Hence,  $\tilde{g}$  is an affine function, and in the rest of this proof we denote  $\tilde{g}(\tilde{x}) = M\tilde{x} + d$ , where  $M \in \mathbb{R}^{(q+1) \times (n+1)}$  and  $d \in \mathbb{R}^{q+1}$ .



Developing, we have

$$\begin{aligned} & \left\| \mathbb{E} \left[ \tilde{g}(\tilde{x}_{u_\pi}(t + \delta)) - \tilde{g}(\tilde{x}_u(t)) - \delta M \tilde{b}(t, u(t), x_u(t)) - \sum_{i=1}^j \eta_i A z_{t_i, u_i}(t) \right] \right\| \\ & \leq \left\| \mathbb{E} \left[ \tilde{g}(\tilde{x}_u(t + \delta)) - \tilde{g}(\tilde{x}_u(t)) - \delta M \tilde{b}(t, u(t), x_u(t)) \right] \right\| \\ & \quad + \left\| \mathbb{E} \left[ \tilde{g}(\tilde{x}_{u_\pi}(t + \delta)) - \tilde{g}(\tilde{x}_u(t + \delta)) - \sum_{i=1}^j \eta_i M z_{t_i, u_i}(t + \delta) \right] \right\| + \sum_{i=1}^j \eta_i \|M\| \left\| \mathbb{E} \left[ z_{t_i, u_i}(t + \delta) - z_{t_i, u_i}(t) \right] \right\|. \end{aligned}$$

From Lemma A.1, the second term on the right-hand side is  $o\left(\sum_{i=1}^j \eta_i\right)$ . For the last summand, from the property of the stochastic integral, we have

$$\begin{aligned} & \left\| \mathbb{E} \left[ z_{t_i, u_i}(t + \delta) - z_{t_i, u_i}(t) \right] \right\| \leq \left\| \mathbb{E} \left[ \int_t^{t+\delta} A(s) z_{t_i, u_i}(s) ds \right] \right\| \\ & \quad + \left\| \mathbb{E} \left[ \int_0^{t+\delta} \mathbb{1}_{[t, T]}(s) D(s) z_{t_i, u_i}(s) dB_s \right] \right\| \leq C \mathbb{E} \left[ \sup_{s \in [0, T]} \|z_{t_i, u_i}(s)\| \right] \delta, \end{aligned}$$

where the last term is  $o\left(\delta + \sum_{i=1}^j \eta_i\right)$ . It is worth pointing out the importance for  $g$  being affine to provide the last claim. Finally, for the remaining term, we apply Itô formula to each coordinate  $h = 1, \dots, q + 1$ , obtaining

$$\begin{aligned} \left[ \tilde{g}(\tilde{x}_u(t + \delta)) - \tilde{g}(\tilde{x}_u(t)) - \delta M \tilde{b}(t, u(t), x_u(t)) \right]_h &= \sum_{k=1}^{n+1} M_{hk} \int_0^{t+\delta} \mathbb{1}_{[t, T]}(s) \tilde{\sigma}_k(s, x_u(s)) dB_s \\ & \quad + \sum_{k=1}^{n+1} M_{hk} \left( \int_t^{t+\delta} \tilde{b}_k(s, u(s), x_u(s)) ds - \delta \tilde{b}_k(t, u(t), x_u(t)) \right), \end{aligned}$$

and therefore

$$\begin{aligned} & \frac{1}{\delta + \sum_{i=1}^j \eta_i} \left\| \mathbb{E} \left[ \tilde{g}(\tilde{x}_u(t + \delta)) - \tilde{g}(\tilde{x}_u(t)) - \delta M \tilde{b}(t, u(t), x_u(t)) \right] \right\| \\ & \leq \frac{\|M\|}{\delta} \int_t^{t+\delta} \mathbb{E} \left[ \left\| \tilde{b}_k(s, u(s), x_u(s)) - \tilde{b}_k(t, u(t), x_u(t)) \right\| \right] ds. \end{aligned}$$

As  $t$  is Lebesgue for  $u$ , this last quantity goes to zero for  $\delta \rightarrow 0$ , and we conclude.  $\square$

## REFERENCES

- [1] A. Agrachev and Y. Sachkov, Vol. 87 of Control theory from the geometric viewpoint. Springer Science & Business Media (2013).
- [2] M. Annunziato and A. Borzi, A Fokker–Planck control framework for multidimensional stochastic processes. *J. Comput. Appl. Math.* **237** (2013) 487–507.
- [3] R. Bellman, Dynamic Programming. Princeton Univ. Press, Princeton, New Jersey (1957).
- [4] Berret, Bastien, and F. Jean, Stochastic optimal open-loop control as a theory of force and impedance planning via muscle co-contraction. *PLoS Comput. Biol.* **16** (2020) e1007414.
- [5] D. Bertsimas and D.B. Brown, Constrained stochastic LQC: a tractable approach. *IEEE Trans Autom. Control* **52** (2007) 1826–1841.
- [6] C. Bes and S. Sethi, Solution of a class of stochastic linear-convex control problems using deterministic equivalents. *J. Optim. Theory Appl.* **62** (1989) 17–27.
- [7] J. Betts, Survey of numerical methods for trajectory optimization, *J. Guid. Control Dyn.* **21** (1998) 193–207.

- [8] J.-M. Bismut, Linear quadratic optimal stochastic control with random coefficients. *SIAM J. Control Optim.* **14** (1976) 419–444.
- [9] P.T. Boggs and J.W. Tolle, Sequential quadratic programming. *Acta Numer.* **4** (1995) 1–51.
- [10] R. Bonalli, Optimal control of aerospace systems with control-state constraints and delays, Ph.D. thesis. Sorbonne Université (2018).
- [11] R. Bonalli, B. Hérisse and E. Trélat, Analytical initialization of a continuation-based indirect method for optimal control of endo-atmospheric launch vehicle systems, in IFAC World Congress (2017).
- [12] R. Bonalli, B. Hérisse and E. Trélat, Solving Optimal Control Problems for Delayed Control-Affine Systems with Quadratic Cost by Numerical Continuation, in American Control Conference (2017).
- [13] R. Bonalli, B. Hérisse and E. Trélat, Continuity of Pontryagin extremals with respect to delays in nonlinear optimal control. *SIAM J. Control Optim.* **57** (2019) 1440–1466.
- [14] R. Bonalli, B. Hérisse and E. Trélat, Optimal control of endo-atmospheric launch vehicle systems: geometric and computational issues. *IEEE Trans. Autom. Control* **65** (2019) 2418–2433.
- [15] L. Bourdin and E. Trélat, Pontryagin maximum principle for finite dimensional nonlinear optimal control problem on time scales. *SIAM J. Control Optim.* **51** (2013) 3781–3813.
- [16] A. Bryson, Applied optimal control: optimization, estimation and control. CRC Press (1975).
- [17] R. Carmona, vol. 1 of Lectures on BSDEs, stochastic control, and stochastic differential games with financial applications. SIAM (2016).
- [18] Y. Chitour, F. Jean and E. Trélat, Singular trajectories of control-ane systems. *SIAM J. Control Optim.* **47** (2008) 1078–1095.
- [19] Damm, Tobias, H. Mena, and T. Stillfjord, Numerical solution of the finite horizon stochastic linear quadratic control problem. *Numer. Linear Algebra Appl.* **24** (2017) e2091.
- [20] M. Diehl and F. Messerer, Local convergence of generalized Gauss-Newton and sequential convex programming, in *Conference on Decision and Control* (2019) .
- [21] Q.T. Dinh and M. Diehl, Local convergence of sequential convex programming for nonconvex optimization, in Recent Advances in *Optimization and its Applications in Engineering*. Springer (2010) 93–102.
- [22] H. Frankowska, H. Zhang and X. Zhang, Stochastic optimal control problems with control and initial-final states constraints. *SIAM J. Control Optim.* **56** (2018) 1823–1855.
- [23] R. Gamkrelidze, Principles of optimal control theory, vol. 7. Springer Science & Business Media (2013).
- [24] E. Gobet, Monte-Carlo methods and stochastic processes: from linear to non-linear. CRC Press (2016).
- [25] T. Haberkorn and E. Trélat, Convergence results for smooth regularizations of hybrid nonlinear optimal control problems. *SIAM J. Control Optim.* **49** (2011) 1498–1522.
- [26] O. Kazuhide, M. Goldshtein and P. Tsiotras, Optimal covariance control for stochastic systems under chance constraints. *IEEE Control Syst. Lett.* **2** (2018) 266–271.
- [27] P. Kleindorfer and K. Glover, Linear convex stochastic optimal control with applications in production planning. *IEEE Trans. Autom. Control* **18** (1973) 56–59.
- [28] D. Kuhn, W. Wiesemann and A. Georghiou, Primal and dual linear decision rules in stochastic and robust optimization. *Math. Program.* **130** (2011) 177–209.
- [29] H.J. Kushner, Numerical methods for stochastic control problems in continuous time. *SIAM J. Control Optim.* **28** (1990) 999–1048.
- [30] H.J. Kushner and L.F. Martins, Numerical methods for stochastic singular control problems. *SIAM J. Control Optim.* **29** (1991) 1443–1475.
- [31] H.J. Kushner and F.C. Schweppe, A maximum principle for stochastic control systems. *J. Math. Anal. Appl.* **8** (1964) 287–302.
- [32] J.-F. Le Gall, vol. 274 of *Brownian motion, martingales, and stochastic calculus*. Springer (2016).
- [33] T. Levajković, H. Mena and L.-M. Pfurtscheller, Solving stochastic LQR problems by polynomial chaos. *IEEE Control Syst. Lett.* **2** (2018) 641–646.
- [34] T. Lew, R. Bonalli and M. Pavone, Chance-constrained sequential convex programming for robust trajectory optimization, in European Control Conference (2020).
- [35] P.-L. Lions, Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations, Part I. *Commun. Partial Differ. Equ.* **8** (1983) 1101–1174.
- [36] Z. Lu, Sequential Convex Programming Methods for A Class of Structured Nonlinear Programming. Tech. rep. (2013).
- [37] Y. Mao, D. Dueri, M. Szmuk and B. Açikmeşe, Successive convexification of non-convex optimal control problems with state constraints. *IFAC-PapersOnLine* **50** (2017) 4063–4069.
- [38] P.S. Maybeck, Stochastic models, estimation, and control. Academic Press (1982).
- [39] J. Nocedal and S. Wright, Numerical Optimization. Springer (1999).
- [40] F. Palacios-Gomez, L. Lasdon and M. Engquist, Nonlinear optimization by successive linear programming. *Manag. Sci.* **28** (1982) 1106–1120.
- [41] S. Peng, A general stochastic maximum principle for optimal control problems. *SIAM J. Control Optim.* **28** (1990) 966–979.
- [42] S. Peng, Stochastic Hamilton–Jacobi–Bellman equations. *SIAM J. Control Optim.* **30** (1992) 284–304.
- [43] L. Pontryagin, Mathematical theory of optimal processes. Routledge (2018).
- [44] J. Potter, A matrix equation arising in statistical filter theory. *Rep. RE-9, Experimental Astronomy Laboratory, Massachusetts Institute of Technology* (1965).

- [45] M.A. Rami and X.Y. Zhou, Linear matrix inequalities, Riccati equations, and indefinite stochastic linear quadratic controls. *IEEE Trans. Autom. Control* **45** (2000) 1131–1143.
- [46] R.T. Rockafellar and R.J.B. Wets, Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time. *SIAM J. Control Optim.* **28** (1990) 810–822.
- [47] A. Shapiro and A. Nemirovski, On complexity of stochastic programming problems, in Continuous optimization. Springer (2005) pp. 111–146.
- [48] I.A. Shvartsman and R.B. Vinter, Regularity properties of optimal controls for problems with time-varying state and control constraints. *Nonlinear Anal.: Theory, Methods Appl.* **65** (2006) 448–474.
- [49] S. Tang, General linear quadratic optimal stochastic control problems with random coefficients: linear stochastic Hamilton systems and backward stochastic Riccati equations. *SIAM J. Control Optim.* **42** (2003) 53–75.
- [50] E. Trélat, Some properties of the value function and its level sets for affine control systems with quadratic cost. *J. Dyn. Control Syst.* **6** (2000) 511–541.
- [51] E. Trélat, Optimal control and applications to aerospace: some results and challenges. *J. Optim. Theory Appl.* **154** (2012) 713–758.
- [52] Y. Wang, D. Yang, J. Yong and Z. Yu, Exact controllability of linear stochastic differential equations and related problems. *Am. Inst. Math. Sci.* **7** (2017) 305–345.
- [53] D.D. Yao, S. Zhang and X.Y. Zhou, Stochastic linear-quadratic control via semidefinite programming. *SIAM J. Control Optim.* **40** (2001) 801–823.
- [54] J. Yong and X.Y. Zhou, Vol. 43 of Stochastic controls: Hamiltonian systems and HJB equations. Springer Science & Business Media (1999).

## Subscribe to Open (S2O)

A fair and sustainable open access model



This journal is currently published in open access under a Subscribe-to-Open model (S2O). S2O is a transformative model that aims to move subscription journals to open access. Open access is the free, immediate, online availability of research articles combined with the rights to use these articles fully in the digital environment. We are thankful to our subscribers and sponsors for making it possible to publish this journal in open access, free of charge for authors.

### **Please help to maintain this journal in open access!**

Check that your library subscribes to the journal, or make a personal donation to the S2O programme, by contacting [subscribers@edpsciences.org](mailto:subscribers@edpsciences.org)

More information, including a list of sponsors and a financial transparency report, available at: <https://www.edpsciences.org/en/math-s2o-programme>