



**HAL**  
open science

# Infinite-Dimensional Sums-of-Squares for Optimal Control

Eloïse Berthier, Justin Carpentier, Alessandro Rudi, Francis Bach

► **To cite this version:**

Eloïse Berthier, Justin Carpentier, Alessandro Rudi, Francis Bach. Infinite-Dimensional Sums-of-Squares for Optimal Control. 61st IEEE Conference on Decision and Control, IEEE, Dec 2022, Cancun, Mexico. hal-03377120

**HAL Id: hal-03377120**

**<https://hal.science/hal-03377120>**

Submitted on 14 Oct 2021

**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.

# INFINITE-DIMENSIONAL SUMS-OF-SQUARES FOR OPTIMAL CONTROL

ELOÏSE BERTHIER, JUSTIN CARPENTIER, ALESSANDRO RUDI AND FRANCIS BACH

*Inria - Ecole Normale Supérieure  
PSL Research University, Paris, France*

ABSTRACT. We introduce an approximation method to solve an optimal control problem *via* the Lagrange dual of its weak formulation. It is based on a sum-of-squares representation of the Hamiltonian, and extends a previous method from polynomial optimization to the generic case of smooth problems. Such a representation is infinite-dimensional and relies on a particular space of functions – a reproducing kernel Hilbert space – chosen to fit the structure of the control problem. After subsampling, it leads to a practical method that amounts to solving a semi-definite program. We illustrate our approach by a numerical application on a simple low-dimensional control problem.

## 1. INTRODUCTION

The continuous-time optimal control problem (OCP) is a versatile framework modelling a wide variety of nonlinear systems and optimization criteria, with countless industrial applications, including aerospace [Tré12] or robotics [MLS94]. Developing efficient numerical methods for solving such a general problem is a daunting task, especially for high-dimensional systems. Among current methods, *indirect methods* exploit optimality criteria derived from Pontryagin’s maximum principle and give precise results but need to be initialized properly, while the less accurate *direct methods* reformulate the problem as a nonlinear program without specific initialization requirements [Tré05, Chapter 9].

In this paper, we focus on a direct method that computes the optimal value function of the problem as the maximal subsolution of the Hamilton-Jacobi-Bellman (HJB) equation. It is described in [HHHLT96, LHPT08] and is obtained by taking the dual of the *weak formulation* of the OCP, involving occupation measures [Vin93]. In [LHPT08], the numerical resolution of this formulation is based on polynomial optimization [Las15], and hence is restricted to polynomial dynamics and cost functions, with a possibly costly extension to smooth functions, involving a hierarchy of semi-definite programs (SDPs).

Our main contribution is to extend the numerical method of [LHPT08] to non-polynomial and smooth OCPs. To this end, we consider a space of smooth functions, called a reproducing kernel Hilbert space (RKHS) [Aro50], and use representations of non-negative functions in this space [MFBR20] with sums-of-squares. This is detailed in Sections 2 and 3. In Section 4, we notably prove that this representation is exact in two particular cases: the time-invariant linear quadratic regulator (LQR) and smooth control-affine systems. This results in a practical numerical method derived in Section 5, that requires to solve an SDP to approximate the optimal value function of the OCP. Finally, in Section 6 we illustrate the practical application of this method to a simple two-dimensional OCP.

## 2. BACKGROUND

First, we introduce the three building blocks that are then combined in Section 3 to design our approximation method.

---

*E-mail address:* `firstname.lastname@inria.fr`.

**2.1. Formulation of OCP with Maximal Subsolutions of HJB.** Let  $\mathcal{X}$  and  $\mathcal{U}$  be compact subsets of  $\mathbb{R}^d$  and  $\mathbb{R}^p$ , for integers  $d, p \geq 1$ , and assume that  $\mathcal{U}$  is convex. We define the dynamics  $f : [0, T] \times \mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R}^d$ , the running cost  $L : [0, T] \times \mathcal{X} \times \mathcal{U} \rightarrow \mathbb{R}$ , and the terminal cost  $M : \mathcal{X} \rightarrow \mathbb{R}$  occurring at the fixed terminal time  $T > 0$ . In addition, we assume that the state trajectories remain in the compact set  $\mathcal{X}$ , and that sampling points from  $\mathcal{X}$  is easy. We do not consider problems with explicit state-constraints which are left as future work.

Assume the existence of a smooth function  $V^* \in C^1(\mathcal{X}_T)$ , meaning that  $V^*$  is smoothly differentiable on  $\mathcal{X}_T := [0, T] \times \mathcal{X}$ , that is a solution of the HJB equation:  $\forall (t, x) \in \mathcal{X}_T$ ,

$$\begin{aligned} \frac{\partial V^*}{\partial t}(t, x) + \inf_{u \in \mathcal{U}} (L(t, x, u) + \nabla V^*(t, x)^\top f(t, x, u)) &= 0 \\ V^*(T, x) &= M(x), \end{aligned} \tag{HJB}$$

where  $\nabla V^*$  refers to the gradient of  $V^*$  w.r.t.  $x$ . For  $t \in [0, T]$ , let  $\mathcal{U}_t$  the set of admissible controls such that  $\forall s \in [t, T], (x(s), u(s)) \in \mathcal{X} \times \mathcal{U}$ . Then  $V^*$  is the value function [Lib11] of the following OCP:  $\forall (t_0, x_0) \in \mathcal{X}_T$ ,

$$\begin{aligned} V^*(t_0, x_0) &= \inf_{u \in \mathcal{U}_{t_0}} \int_{t_0}^T L(t, x(t), u(t)) dt + M(x(T)) \\ \forall t \in [t_0, T], \dot{x}(t) &= f(t, x(t), u(t)), \quad x(0) = x_0. \end{aligned} \tag{OCP}$$

Let  $\mu_0$  be a probability measure on  $\mathcal{X}$ . We are interested in the value of the stochastic initial point problem, where  $x_0$  is drawn according to  $\mu_0$ , that is,  $\mathbb{E}_{x_0 \sim \mu_0} [V^*(0, x_0)] = \int V^*(0, x_0) d\mu_0(x_0)$ .

In this paper, instead of directly looking for solutions of the HJB equation, we will focus on the following alternative problem (P), namely, finding a maximal subsolution of HJB:

$$\begin{aligned} \sup_{V \in C^1(\mathcal{X}_T)} \int V(0, x_0) d\mu_0(x_0) \\ \forall (t, x, u), \frac{\partial V}{\partial t}(t, x) + L(t, x, u) + \nabla V(t, x)^\top f(t, x, u) &\geq 0 \\ \forall x, V(T, x) &\leq M(x). \end{aligned} \tag{P}$$

This is the dual of the *weak formulation* of the OCP with occupation measures, which is a linear program over the space of measures [Vin93, Las09, KSEL17]. Moreover, subsolutions of HJB also play a key role in the theory of viscosity solutions [CL83] of partial differential equations. The first constraint in (P) is the positivity of a certain Hamiltonian associated to  $V$ , namely:

$$H(t, x, u) := \frac{\partial V}{\partial t}(t, x) + L(t, x, u) + \nabla V(t, x)^\top f(t, x, u).$$

If  $V = V^*$  is the optimal value function, then the optimal controller  $u = u^*(t, x)$  minimizes the positivity constraint and for all  $(t, x) \in \mathcal{X}_T$ ,  $H^*(t, x, u^*(t, x)) = 0$ .

Our goal is to find an approximate solution  $V$  of (P). Under some additional assumptions, (P) is equivalent to the OCP. Such regularity and convexity assumptions were first studied by [Vin93] and are detailed in [LHPT08]. In this particular case, the value of problem (P) coincides with the one of the stochastic initial point problem:

$$\sup P = \mathbb{E}_{x_0 \sim \mu_0} [V^*(0, x_0)].$$

**2.2. Parameterization of the Value Function.** A first difficulty in problem (P) is searching  $V$  in the infinite-dimensional set  $C^1(\mathcal{X}_T)$ . One option is to search  $V$  in a finitely-parameterized set  $\mathcal{F}_\Theta$ . A common practice, notably in approximate dynamic programming and reinforcement learning [SB18, Chapter 9], is to use a linear approximation of  $V$ , with a feature vector  $\psi(t, x) \in \mathbb{R}^m$  and a parameter  $\theta$  in a convex subset  $\Theta$  of  $\mathbb{R}^m$ , for  $m \geq 1$ .

Since we assumed that  $V^*$  is a solution of the HJB equation, we can restrict that search space in (P) to functions  $V$  such that  $V(T, \cdot) = M(\cdot)$ . Then we assume that the parameterization is such that  $\forall \theta, V_\theta(T, \cdot) = M(\cdot)$ , so that we can remove the explicit constraint in (P). Hence our parameterized set is:

$$\mathcal{F}_\Theta := \{(t, x) \mapsto V_\theta(t, x) = \theta^\top \psi(t, x) + M(x) \mid \theta \in \Theta\},$$

with  $\psi$  such that  $\psi(T, \cdot) = 0$ . To simplify the evaluations of  $\nabla V_\theta$  and  $\frac{\partial V_\theta}{\partial t}$ , it is convenient (but not necessary) to use a separable feature vector  $\psi(t, x) = \kappa(t)\varphi(x)$ , with  $\kappa(T) = 0$ .

**2.3. Representing Non-Negative Functions as Sum-of-Squares.** Problem (P) is constrained by a dense set of inequalities indexed by  $(t, x, u) \in [0, T] \times \mathcal{X} \times \mathcal{U}$ , which cannot be directly handled by numerical algorithms. Hence we look for a finite – possibly approximate – representation of the non-negative function  $(t, x, u) \mapsto H(t, x, u) \geq 0$ .

If  $f, L, M$  are polynomials, one way is to use sum-of-squares (SoS) polynomials [Las15], *i.e.*, to represent  $H(t, x, u)$  as the sum of the squares of polynomials of a given degree. This is a sufficient but not necessary condition for being a non-negative polynomial. This technique has been applied to problem (P) in [LHPT08], although it is presented in its dual version using the method of moments [Hen13]. In any case, this representation is not exact in general and gives a lower approximation of (P). To numerically solve the problem, one needs to build a hierarchy of SDPs obtained by this SoS representation with polynomials of increasing degree  $r$ . Under generic conditions, this hierarchy converges to the value of (P), but the speed of convergence is not explicitly controlled [PHL17]. This is a potentially critical issue, since the size of the SDP at rank  $r$  is defined by the number of monomials of degree less than  $r$  in the dimension of  $(t, x, u)$ , which is  $\binom{d+p+1+r}{r}$ , a quantity growing exponentially with  $r$ .

In this paper, we opt for another option inspired by recently-introduced machine learning techniques [RMFB20]: representing a non-negative function as a SoS in a reproducing kernel Hilbert space (RKHS). Hereafter, we briefly define an RKHS and mention its main properties, and refer to [PR16] for a thorough description. Consider a set  $E$ , a function  $k : E \times E \rightarrow \mathbb{R}$  is a positive definite kernel if  $\forall n \in \mathbb{N}, y_1, \dots, y_n \in E$ , the matrix  $K := (k(y_i, y_j))_{i,j=1}^n$  is positive semi-definite. Associated to a positive definite kernel  $k$ , there exists a unique RKHS  $\mathcal{H}$ , a Hilbert space of functions  $E \rightarrow \mathbb{R}$ , with an inner product  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ , such that the following properties hold (the second one is the so-called “reproducing property”):

- $\forall y \in E, k_y := k(y, \cdot) \in \mathcal{H}$ ;
- $\forall g \in \mathcal{H}, \forall y \in E, \langle g, k_y \rangle_{\mathcal{H}} = g(y)$ .

In addition, there exists a feature map  $\Phi : E \rightarrow \mathcal{H}$ , possibly infinite dimensional, defined by  $\Phi(y) = k_y$ , which maps a point in  $E$  to a function in  $\mathcal{H}$ , and in particular we have  $k(y, y') = \langle \Phi(y), \Phi(y') \rangle_{\mathcal{H}}$ . Conversely, any feature map defines an RKHS associated to the former kernel.

Here we mention two classical kernels from the RKHS literature, which we use later to build our function representations. Assume that  $E$  is a subset of  $\mathbb{R}^\ell$ , for  $\ell \geq 1$ . The polynomial kernel of degree  $r$  is defined on  $E \times E$  by  $k(y, y') = (1 + y^\top y')^r$ , and the corresponding embedding  $\Phi(y)$  is the vector of  $\binom{\ell+r}{r}$  multivariate monomials of degree less than  $r$ . In this case  $\mathcal{H}$  is finite-dimensional. The exponential kernel is defined by  $k(y, y') = \exp(-\|y - y'\|_2 / \sigma)$ , with  $\sigma > 0$ . If  $E$  is bounded and has locally Lipschitz boundary, the corresponding RKHS is the Sobolev space of functions whose weak-derivatives up to order  $s = \ell/2 + 1/2$  are square-integrable [BTA11].

Functions that are SoS in an RKHS  $\mathcal{H}$  can be represented using an infinite-dimensional positive semi-definite operator [RMFB20, Corollary 1]. Indeed, assume that a function  $h \in \mathcal{H}$  is written as a sum-of-squares of functions  $h_j \in \mathcal{H}$ :

$$\forall y \in E, \quad h(y) = \sum_{j=1}^q h_j(y)^2.$$

For any  $v, w \in \mathcal{H}$ , we have:

$$\langle w, (v \otimes v)w \rangle = \langle w, vv^*w \rangle = \text{Tr}(w^*vv^*w) = (\langle v, w \rangle)^2,$$

where  $\otimes$  denotes the outer product. Because of the reproducing property:  $\forall j \in \{1, \dots, q\}, y \in E$ ,

$$h_j(y)^2 = (\langle \Phi(y), h_j \rangle_{\mathcal{H}})^2 = \langle \Phi(y), (h_j \otimes h_j)\Phi(y) \rangle.$$

And then:

$$h(y) = \langle \Phi(y), \mathcal{A}\Phi(y) \rangle_{\mathcal{H}},$$

with  $\mathcal{A} := \sum_{j=1}^q h_j \otimes h_j \in \mathbb{S}_+(\mathcal{H})$  and  $\mathcal{A}$  has rank at most  $q$ , where  $\mathbb{S}_+(\mathcal{H})$  is the set of bounded self-adjoint positive semi-definite operators on  $\mathcal{H}$ .

In [MFBR20], this SoS representation in an RKHS is used to model non-negative functions, *e.g.*, for signal processing or statistics applications. In some cases, *e.g.*, in Sobolev spaces as we will see hereafter, the representation is exact in the sense that all non-negative functions can be written as a SoS in  $\mathcal{H}$ , whereas the polynomial SoS representation is tight only for a restricted class of

polynomials [Las09]. Besides, SoS polynomials are a particular case of what has just been described, if  $k$  is the polynomial kernel. In the rest of the paper, we will extend the method of [LHPT08] from polynomials to any RKHS, at the expense of possibly infinite-dimensional representations.

### 3. DENSE SET OF INEQUALITY CONSTRAINTS

In this section, we start by providing a basic relaxation of problem (P) and a motivation for preferring a SoS representation of the non-negativity constraints in (P). Then we present the resulting problem and its main features.

**3.1. Relaxed formulation by subsampling.** A straightforward relaxation of (P) is obtained by finitely subsampling the non-negativity constraints. Let us sample values of  $(t^{(i)}, x^{(i)}, u^{(i)})_{i \in I}$  in  $[0, T] \times \mathcal{X} \times \mathcal{U}$ , with  $I$  a finite set of cardinality  $n \geq 1$ . For simplicity, assume that  $\mu_0$  is the mean of  $n_0$  Diracs at points  $\{x_0^{(i)}\}_{i \in \{1, \dots, n_0\}}$ . Besides, let us use a linear parameterization of  $V$  as described in Section 2.2, with  $\Theta = \mathbb{R}^m$ . We then obtain a linear program, with a possibly unbounded solution in the overparameterized setting ( $m \gg n$ ). To circumvent this effect, we add a quadratic regularizer on  $\theta$  with parameter  $\lambda_\theta \geq 0$ , and obtain the following problem:

$$\begin{aligned} & \sup_{\theta \in \mathbb{R}^m} \frac{1}{n_0} \sum_{k=1}^{n_0} \theta^\top \psi(0, x_0^{(k)}) + M(x_0^{(k)}) - \lambda_\theta \|\theta\|_2^2 \\ & \forall i \in I, \quad \theta^\top \frac{\partial \psi}{\partial t}(t^{(i)}, x^{(i)}) + L(t^{(i)}, x^{(i)}, u^{(i)}) \\ & + \left( \theta^\top \nabla \psi(t^{(i)}, x^{(i)}) + \nabla M(x^{(i)}) \right)^\top f(t^{(i)}, x^{(i)}, u^{(i)}) \geq 0. \end{aligned} \quad (\text{LP})$$

Although this is not exactly a linear program if  $\lambda_\theta > 0$ , it can still be solved easily by standard solvers, and we will refer to it as the LP problem. A similar finite-dimensional LP formulation has been proposed in [GQ09] for discounted infinite-horizon control problems. It is part of a long series of LP formulations for optimal control (see, e.g., [GPS17] and references therein), for dynamic programming and more recently for reinforcement learning [LMMN21].

This problem will be used as a baseline in Section 6, to be compared with the SoS formulation below. It is a relaxation that gives an upper-bound on (P), but it is not straightforward to relate the number of samples  $n$  and how they are spread to the quality of the approximation. Yet in the example below, this can be evaluated explicitly.

*Example 1:* Let  $g : \mathbb{R}^p \rightarrow \mathbb{R}$  be a smooth function with a unique minimizer  $u^*$ . With  $T = 1$ ,  $L : (t, x, u) \mapsto g(u)$ ,  $f = 0$  and  $M = 0$ , then  $V^*(t, x) = (1 - t)g(u^*)$  and solving the OCP is essentially equivalent to finding the global minimizer of  $g$ . If  $V$  is parameterized by  $V_\theta(t, x) = \theta(1 - t)$ , the LP formulation with  $\lambda_\theta = 0$  writes:

$$\sup \theta \text{ s.t. } \forall i \in I, -\theta + g(u_i) \geq 0,$$

which is readily solved by  $\theta = \min\{g(u_i)\}_{i \in I}$ . In general, this method requires  $O(\varepsilon^{-p})$  samples to approximate  $g(u^*)$  with precision  $\varepsilon$  [Nov06], and yet when  $g$  is smooth, this is not an optimal way to perform zero-th order optimization. Indeed, [RMFB20] use a SoS representation of  $g - \theta$  to solve this exact problem, and alleviate the curse of dimensionality when  $g$  is sufficiently smooth: the number of samples reduces to  $O(\varepsilon^{-p/s})$  for  $g \in C^s(\mathbb{R}^p)$ . In the rest of this paper, we propose to use the same approach and to generalize it to any OCP. We expect similar benefits when  $H$  is smooth.

**3.2. Strengthened formulation by SoS representation.** Consider an RKHS  $\mathcal{H}$  of real-valued functions on  $[0, T] \times \mathcal{X} \times \mathcal{U}$ , with positive definite kernel  $k$ , and  $\Phi : E \rightarrow \mathcal{H}$  the corresponding embedding. We use a SoS representation in  $\mathcal{H}$ , or “kernel SoS”, for the constraint  $H \geq 0$  in (P):

$$\begin{aligned} & \sup_{\substack{V \in C^1(\mathcal{X}_T), \\ \mathcal{A} \in \mathbb{S}_+(\mathcal{H})}} \int V(0, x_0) d\mu_0(x_0) \\ & \forall (t, x, u), \quad H(t, x, u) = \langle \Phi(t, x, u), \mathcal{A}\Phi(t, x, u) \rangle. \end{aligned} \quad (\text{KSOS})$$

This is a strengthening of the constraint in (P), since being SoS is stronger than being non-negative. So in general, (KSOS) is a lower-approximation of (P). However, in certain cases, (KSOS)

can be equivalent to (P), as we will prove in Section 4. A sufficient condition is the existence of  $\mathcal{A} \in \mathbb{S}_+(\mathcal{H})$  such that, at the optimal  $V^*$ :

$$\forall(t, x, u), \quad H^*(t, x, u) = \langle \Phi(t, x, u), \mathcal{A}\Phi(t, x, u) \rangle.$$

#### 4. TIGHT SUM-OF-SQUARES REPRESENTATIONS

We study the tightness of problem (KSOS) in two particular cases: the time-invariant LQR and smooth value functions.

**4.1. Case 1: Infinite-Horizon Time-Invariant LQR.** First we look at a very simple OCP, where every quantity can be computed almost in closed form, and with infinite-horizon so that there is no dependence in  $t$ . Let  $f(x, u) = A_0x + B_0u$ , for  $A_0 \in \mathbb{R}^{d \times d}$ ,  $B_0 \in \mathbb{R}^{d \times p}$ , with  $(A_0, B_0)$  controllable,  $L(x, u) = x^\top Q_0x + u^\top R_0u$ ,  $Q_0 \in \mathbb{S}_+(\mathbb{R}^{d \times d})$ ,  $R_0 \in \mathbb{S}_+(\mathbb{R}^{p \times p})$ ,  $R_0 \succ 0$ . The optimal value function is  $V^*(x) = x^\top S_0x$ , where  $S_0$  is the unique positive semi-definite solution of the algebraic Riccati equation:

$$0 = -Q_0 - A_0^\top S_0 - S_0 A_0 + S_0 B_0 R_0^{-1} B_0^\top S_0.$$

The optimal controller is  $u^*(x) = -R_0^{-1} B_0^\top S_0 x =: -K_0 x$ .

$$\begin{aligned} H^*(x, u) &= x^\top Q_0 x + u^\top R_0 u + 2x^\top S_0 (A_0 x + B_0 u) \\ &= u^\top R_0 u + x^\top S_0 B_0 u + u^\top B_0^\top S_0 x + x^\top S_0 B_0 K_0 x. \end{aligned}$$

This is a SoS of degree-one polynomials in  $(x, u)$ :

$$\begin{aligned} H^*(x, u) &= (u + K_0 x)^\top R_0 (u + K_0 x) \\ &= \begin{pmatrix} x^\top & u^\top \end{pmatrix} \begin{pmatrix} K_0^\top \\ I_p \end{pmatrix} R_0 \begin{pmatrix} K_0 & I_p \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} \\ &= \sum_{j=1}^p [q_j(x, u)]^2, \end{aligned}$$

with  $q_j(x, u) := [R_0^{1/2}]_j \cdot \begin{pmatrix} K_0 & I_p \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix}$ .

Hence, an infinite-horizon, time-invariant LQR with unknown parameters can be equivalently expressed by:

$$\begin{aligned} \sup_{V \in C^1(\mathcal{X}), N \succeq 0} \int V(x_0) d\mu_0(x_0) \\ \forall(x, u), \quad L(x, u) + \nabla V(x)^\top f(x, u) = \begin{pmatrix} x \\ u \end{pmatrix}^\top N \begin{pmatrix} x \\ u \end{pmatrix}. \end{aligned}$$

In the next section, we prove that similar SoS constructions exist for sufficiently smooth OCPs, possibly with an infinite-dimensional embedding (*v.s.* a  $(d + p)$ -dimensional one here).

**4.2. Sum-of-Squares Decomposition with Smooth Functions.** We show that, for smooth and control-affine OCPs,  $H^*$  is a SoS of smooth functions. Let  $\Omega_2 := \text{Int}(\mathcal{U})$ ,  $\Omega_1 := \text{Int}\{(t, x) \in \mathcal{X}_T \mid \arg\min_{u \in \mathcal{U}} H^*(t, x, u) \subset \Omega_2\}$ , and  $\Omega := \Omega_1 \times \Omega_2$ .

**Theorem 1.** *Let  $s \in \mathbb{N}$ ,  $s \geq 1$ . Assume that:*

- *$f$  is control-affine:  $\forall(t, x, u) \in [0, T] \times \mathcal{X} \times \mathcal{U}$ ,*

$$f(t, x, u) = g(t, x) + B(t, x)u.$$

- *For all  $(t, x) \in \Omega_1$ ,  $u \mapsto L(t, x, u)$  is twice differentiable on  $\Omega_2$  and strongly convex:  $\nabla_u^2 L(t, x, u) \succeq \rho I$  for some  $\rho > 0$ , and  $(t, x, u) \mapsto \nabla_u^2 L(t, x, u) \in C^s(\Omega)$ .*
- *$(t, x, u) \mapsto \nabla_u L(t, x, u) + B(t, x)^\top \nabla_x V^*(t, x) \in C^s(\Omega)$ .*

*Then there exist  $p$  functions  $(w_j)_{1 \leq j \leq p} \in C^s(\Omega)$  such that:*

$$\forall(t, x, u) \in \Omega, \quad H^*(t, x, u) = \sum_{j=1}^p w_j(t, x, u)^2.$$

The proof is in the appendix. This result motivates the use of exponential kernels, inducing a Sobolev space RKHS, to represent the non-negativity constraints in (P) for smooth OCPs. When  $s > d/2 + 1$ , by applying a technique similar to the one used under Corollary 2 of [RMFB20], it is possible to obtain a SoS representation in terms of the exponential kernel. Then (KSOS) is equivalent to (P).

**4.3. Stochastic Smoothing of the Optimal Value Function.** However in general,  $V^*$  is not necessarily smooth (e.g. minimal time problems), nor is  $u^*$  (e.g. *bang-bang* controllers that are not even continuous). Here, we provide a generic technique to give some regularity to  $V^*$ . For  $\eta > 0$ , we can define a perturbed version of the control system [FR12], where the state is a random variable  $X_t$ :

$$dX_t = f(t, X_t, u_t)dt + (2\eta)^{1/2}dB_t,$$

where  $B_t$  is a standard Brownian motion independent of  $X_t$ . We define the optimal value function, with  $X_0 = x_0$ :

$$V^\eta(t_0, x_0) = \inf_{u \in \mathcal{U}} \mathbb{E} \left[ \int_{t_0}^T L(t, X_t, u_t)dt + M(X_T) \right].$$

$V^\eta$  is the unique  $C^{1,2}(\mathcal{X}_T)$  ( $C^1$  in  $t$ ,  $C^2$  in  $x$ ) solution [FR12] of the following regularized HJB equation:  $\forall (t, x) \in \mathcal{X}_T$ ,

$$\begin{aligned} \inf_{u \in \mathcal{U}} \{L(t, x, u) + \nabla V^\eta(t, x)^\top f(t, x, u)\} \\ + \frac{\partial V^\eta}{\partial t}(t, x) + \eta \Delta V^\eta(t, x) = 0, \end{aligned}$$

with  $V^\eta(T, x) = M(x)$ .  $\Delta V^\eta$  refers to the Laplacian with respect to  $x$ . Contrary to HJB, the solutions are at least  $C^{1,2}(\mathcal{X}_T)$  because this is a quasilinear parabolic partial differential equation [Lie96]. The regularization  $\eta \Delta V$  is a vanishing viscosity term, and the optimal controller is still in  $\operatorname{argmin}_u L + \nabla V^\eta{}^\top f$ . Generically,  $V^\eta$  converges to  $V^*$  as  $\eta \rightarrow 0$ , following a reasoning similar to the theory of viscosity solutions [CL83].

## 5. SDP FORMULATION AND ITS NUMERICAL RESOLUTION

**5.1. Finite-Dimensional Formulation via Subsampling.** Similarly to the (LP) formulation that relaxes (P), we will now derive a relaxation of problem (KSOS), which is another relaxation of problem (P) if the SoS representation of  $H^*$  is tight. Going through (KSOS) as an intermediate step will help to exploit the structure of (P). Using a parameterization of  $V$  in  $\mathcal{F}_\Theta$  with  $\Theta = \mathbb{R}^m$ , and a set of sampled points  $(t^{(i)}, x^{(i)}, u^{(i)})_{i \in I}$  in  $[0, T] \times \mathcal{X} \times \mathcal{U}$ , with  $|I| = n$ , we obtain:

$$\begin{aligned} \sup_{\mathcal{A} \in \mathbb{S}_+(\mathcal{H}), \theta \in \Theta} \quad & c^\top \theta - \lambda_\theta \|\theta\|_2^2 - \lambda \operatorname{Tr}(\mathcal{A}) + C \\ \text{such that} \quad & \forall i \in \{1, \dots, n\}, \\ & b_i + a_i^\top \theta = \langle \Phi(t^{(i)}, x^{(i)}, u^{(i)}), \mathcal{A} \Phi(t^{(i)}, x^{(i)}, u^{(i)}) \rangle, \end{aligned}$$

with  $c := \sum_{i=1}^n \mu_0^{(i)} \psi(0, x^{(i)})$ ,  $C := \sum_i \mu_0^{(i)} M(x^{(i)})$ ,  $b_i := L(t^{(i)}, x^{(i)}, u^{(i)}) + \nabla M(x^{(i)})^\top f(t^{(i)}, x^{(i)}, u^{(i)}) + \eta \Delta M(x^{(i)})$ , and  $a_i := J_\psi(t^{(i)}, x^{(i)}) f(t^{(i)}, x^{(i)}, u^{(i)}) + \frac{\partial \psi}{\partial t}(t^{(i)}, x^{(i)}) + \eta \Delta \psi(t^{(i)}, x^{(i)})$ , where  $J_\psi$  denotes the Jacobian matrix of  $\psi$  with respect to  $x$  only. Note that we integrate the stochastic smoothing process in this formulation, with parameter  $\eta$  that can be eventually set to 0.

The regularization parameter  $\lambda > 0$  controls the trace of the infinite-dimensional operator  $\mathcal{A}$ , and allows for subsampling to provably recover the non-sampled program when  $n$  tends to infinity, and  $\lambda$  goes to zero at the proper rate (see [RMFB20] for the precise dependence). In the limit  $\lambda \rightarrow 0$ , we recover the LP formulation where we assume nothing about the SoS representation of  $H^*$  in  $\mathcal{H}$ .

Both the operator  $\mathcal{A}$  and the  $\Phi(t^{(i)}, x^{(i)}, u^{(i)})$  can be infinite dimensional, depending on the RKHS  $\mathcal{H}$ . Yet, following [RMFB20], we can reformulate the problem equivalently in finite dimension. Using the representer theorem in [MFBR20], one can prove that  $\mathcal{A}$  can be sought in the form: for  $D \in \mathbb{R}^{n \times n}$ ,  $D \succeq 0$ ,

$$\mathcal{A} = \sum_{i,j=1}^n D_{ij} \Phi(t^{(i)}, x^{(i)}, u^{(i)}) \otimes \Phi(t^{(j)}, x^{(j)}, u^{(j)}).$$

Simple computations detailed in [RMFB20] show that:

$$\begin{cases} \forall i, \langle \Phi(t^{(i)}, x^{(i)}, u^{(i)}), \mathcal{A}\Phi(t^{(i)}, x^{(i)}, u^{(i)}) \rangle = [KDK]_{ii}, \\ \text{Tr}(\mathcal{A}) = \text{Tr}(DK), \end{cases}$$

where  $K$  is the kernel matrix with entry  $(i, j)$  equal to  $k((t^{(i)}, x^{(i)}, u^{(i)}), (t^{(j)}, x^{(j)}, u^{(j)}))$ . Assume that  $K \succ 0$ . We denote by  $K = R^\top R$  the Cholesky decomposition of  $K$ , with  $R$  an invertible upper-triangular matrix.

Let  $B := RDR^\top$  and for  $1 \leq i \leq n$ ,  $\Phi_i := R_{\cdot i}$ . Then:

$$\begin{cases} \text{Tr}(B) = \text{Tr}(DK) = \text{Tr}(\mathcal{A}), \\ [KDK]_{ii} = [R^\top BR]_{ii} = \Phi_i^\top B \Phi_i. \end{cases}$$

The problem can now be reformulated as a finite-dimensional SDP over the positive semi-definite matrix  $B \in \mathbb{R}^{n \times n}$ :

$$\begin{aligned} \sup_{B \succcurlyeq 0, \theta \in \mathbb{R}^m} \quad & c^\top \theta - \lambda_\theta \|\theta\|_2^2 - \lambda \text{Tr}(B) + C \quad (\text{SDP}) \\ \text{such that} \quad & \forall i \in \{1, \dots, n\}, b_i + a_i^\top \theta = (\Phi_i)^\top B \Phi_i. \end{aligned}$$

An important question is to estimate the number of subsampled inequalities sufficient to ensure that (SDP)  $\Leftrightarrow$  (KSOS). If nothing is assumed on the structure of  $H$ , as in the LP method, this number is infinite. In contrast, the kernel SoS representation can reduce it or make it finite. If  $H$  is a polynomial of degree  $2r$ ,  $k$  is the polynomial kernel of degree  $r$ , then  $n \geq 2r$  distinct sampled points are enough to interpolate  $H$ , and (SDP)  $\Leftrightarrow$  (KSOS). Another example is global optimization of smooth functions (see *Example 1*) with the exponential kernel. We refer to [RMFB20] for the analysis of the convergence rates, with a lower dependence in the dimension for the kernel SoS when compared to direct inequality subsampling (corresponding to the LP approach).

**5.2. Interior Point Method with the Damped Newton Method.** Problem (SDP) can be readily solved by any off-the-shelf SDP solver. However, for large  $n$ , this quickly becomes too computationally demanding. Here, we propose a numerical scheme based on [RMFB20] that scales better with the number of subsamples  $n$ . First, we introduce a slack variable  $\delta \in \mathbb{R}^n$  allowing the constraints to be slightly violated (*e.g.* because  $\mathcal{F}_\Theta$  is not a perfect model), controlled by a large parameter  $\gamma > 0$ . Second, we introduce a log-barrier term controlled by a small  $\varepsilon > 0$ , useful to form the dual of the SDP. We obtain the following problem:

$$\begin{aligned} \sup_{\substack{B \succcurlyeq 0, \\ \theta, \delta}} \quad & c^\top \theta - \lambda \text{Tr}(B) - \lambda_\theta \|\theta\|_2^2 - \gamma \|\delta\|_2^2 + \varepsilon \log \det B + C \\ \text{such that} \quad & \forall i \in \{1, \dots, n\}, b_i + a_i^\top \theta = (\Phi_i)^\top B \Phi_i + \delta_i. \end{aligned}$$

The Lagrange dual of this problem reads:

$$\begin{aligned} \inf_{\alpha \in \mathbb{R}^n} \quad & \sum_{i=1}^n \alpha_i b_i + \frac{1}{4\lambda_\theta} \sum_{j=1}^m \left( c_j + \sum_{i=1}^n \alpha_i a_{ij} \right)^2 \\ & - \varepsilon \log \det U(\alpha) + \frac{1}{4\gamma} \|\alpha\|_2^2 + \varepsilon n \log(\varepsilon/e) + C, \end{aligned}$$

where  $U(\alpha) := \lambda I_n + \Phi^\top \text{Diag}(\alpha) \Phi$ , and  $\Phi := R^\top$  is the matrix with rows  $(\Phi_i)_{1 \leq i \leq n}$ . Let us call the objective  $F(\alpha)$ .

Since  $F/\varepsilon$  is self-concordant [BV04] like in [RMFB20], we propose to use damped Newton iterations [Nem04] on  $F/\varepsilon$ :

$$\alpha \leftarrow \alpha + \frac{1}{1 + \lambda(\alpha)} \Delta \alpha,$$

where  $\Delta(\alpha) := -[F''(\alpha)]^{-1} F'(\alpha)$  is the Newton direction and  $\lambda(\alpha) := \sqrt{\Delta \alpha^\top F''(\alpha) \Delta \alpha} / \varepsilon$  is the Newton decrement. The gradient and Hessian of  $F$  are computed by:

$$\begin{aligned} \frac{\partial F}{\partial \alpha_i} = b_i + \frac{1}{2\lambda_\theta} \sum_{j=1}^m a_{ij} \left( c_j + \sum_{k=1}^n a_{kj} \alpha_k \right) + \frac{1}{2\gamma} \alpha_i \\ - \varepsilon \Phi_i^\top U(\alpha)^{-1} \Phi_i. \end{aligned}$$



$$\frac{\partial^2 F}{\partial \alpha_i \partial \alpha_j} = \frac{1}{2\lambda_\theta} \sum_{k=1}^n a_{ik} a_{jk} + \varepsilon [\Phi_i^\top U(\alpha)^{-1} \Phi_j]^2 + \frac{\mathbf{1}_{i=j}}{2\gamma}.$$

At optimum, the value function is recovered by

$$\theta^* = \frac{1}{2\lambda_\theta} \left( \sum_{i=1}^n \alpha_i^* a_i + c \right),$$

and the dual variable  $\alpha^*$  plays a role similar to an occupation measure [Vin93], although it is not necessarily non-negative. To improve numerical stability in the experiments hereafter, we used an homotopy heuristics that progressively decreases the parameters  $\lambda_\theta$  and  $\varepsilon$ . Moreover, parallel implementations are possible because no singular value decomposition is needed, only matrix operations and system inversions.

## 6. NUMERICAL EXAMPLE

In this section, we apply the kernel SoS method along with the basic LP method, on a two-dimensional control problem, namely the double integrator with finite horizon.

Setting. The problem is an LQR, as in Section 4.1, but with finite-horizon  $T = 1$ ,  $d = 2$ ,  $p = 1$ ,  $M(x) = \|x\|_2^2$ ,

$$A_0 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, B_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, Q_0 = I_2, R_0 = 0.1.$$

The optimal value function and controller are  $V^*(t, x) = x^\top S(t)x$ ,  $u^*(t, x) = -R_0^{-1} B_0^\top S(t)x =: -K(t)x$ , where  $S(\cdot)$  is the positive semi-definite solution of  $S(T) = I_2$  and:

$$\dot{S}(t) = -Q_0 - A_0^\top S(t) - S(t)A_0 + S(t)B_0R_0^{-1}B_0^\top S(t).$$

Parameterization of  $V$ . Let  $V_\theta(t, x) = \theta^\top \psi(t, x)$ , where each entry of  $\psi$  is a product of basis functions on  $\mathcal{X}$  and  $[0, T]$ . Let  $\varphi(x) := (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)^\top$ , because we know  $V^*$  is quadratic in  $x$ . For  $\kappa$  on  $[0, T]$ , we only know that it is a smooth function, so we use an approximate basis of the Sobolev space of functions with squared integrable derivatives: a sequence of sines and cosines with decreasing periods beginning with  $2T$  to avoid constraining  $V(0, \cdot) = V(T, \cdot)$ , and  $\kappa(T) = 0$ , ensures that  $V(T, \cdot) = M(\cdot)$ :

$$\kappa(t) := \left( \frac{1}{\omega} \sin \left( \frac{\omega\pi t - T}{2} \right) \right)_{1 \leq \omega \leq m_t}^\top.$$

Finally,  $\psi_{i+6j}(t, x) := \varphi_i(x)\kappa_j(t)$ , and  $\theta \in \mathbb{R}^m$ ,  $m = 6m_t$ . We choose  $m_t = 10$ , for which the performance of the policy of the projection of  $V^*$  on  $\mathcal{F}_\Theta$  is almost perfect.

Evaluation. We give two criteria to evaluate the quality of an approximation  $V$ . First, the distance to  $V^*$ :  $\|\bar{V} - \bar{V}^*\|^2$ , where  $\bar{V}$  is the vector of its values on a regular grid on  $[0, T] \times [-1, 1]^2$  with  $10 \times 10 \times 10$  points. Second, the cost of the policy on a  $10 \times 10$  regular grid of initial points.

Sampling. The set of samples  $(t^{(i)}, x^{(i)}, u^{(i)})_{i \in I}$  is built as follows. The  $x^{(i)}$  are  $n_x$  points in  $[-1, 1]^2$  generated by the Sobol sequence [Sob67], the  $(u^{(i)})_{1 \leq i \leq n_u}$  are on a uniform grid on  $[-10, 10]$  and the  $(t^{(i)})_{1 \leq i \leq n_t}$  on  $[0, T]$ . The sample set is the Cartesian product of the three previous ones, and has  $n = n_t n_x n_u$  elements. We also use the same samples as initial points  $(t_0^{(i)}, x_0^{(i)})$  in the objective function of problem (P). Note that we have replaced it with  $\sum_{i=1}^n V(t^{(i)}, x^{(i)})/n$ , as we found it more efficient in our experiments to optimize over  $V$  at intermediate time steps rather than at  $t_0$  only. Indeed, we ultimately evaluate our approximation by the accuracy of  $V$  on the whole  $\mathcal{X}_T$  and not only on  $\{t_0\} \times \mathcal{X}$ . In a discrete states and actions setting, this effect is analyzed in [DFVR03], where  $\mu_0$  is denoted as ‘‘state-relevance weights’’.

Methods. We compare three methods: the LP, the guided SoS and the kernel SoS. The LP method is detailed in Section 3.1, and as for the kernel SoS method, we add a slackness parameter on the constraints, with a penalization controlled by  $\gamma > 0$  ( $\gamma \rightarrow \infty$  recovers the original LP).

The guided SoS method is the same as the (SDP) problem, except that the embeddings  $\Phi_i$  of the samples are replaced by vectors  $\Psi_i$  of fixed dimension, which are computed explicitly, without a kernel. Motivated by the fact that:

$$H^*(t, x, u) = (u + K(t)x)^\top R(u + K(t)x),$$

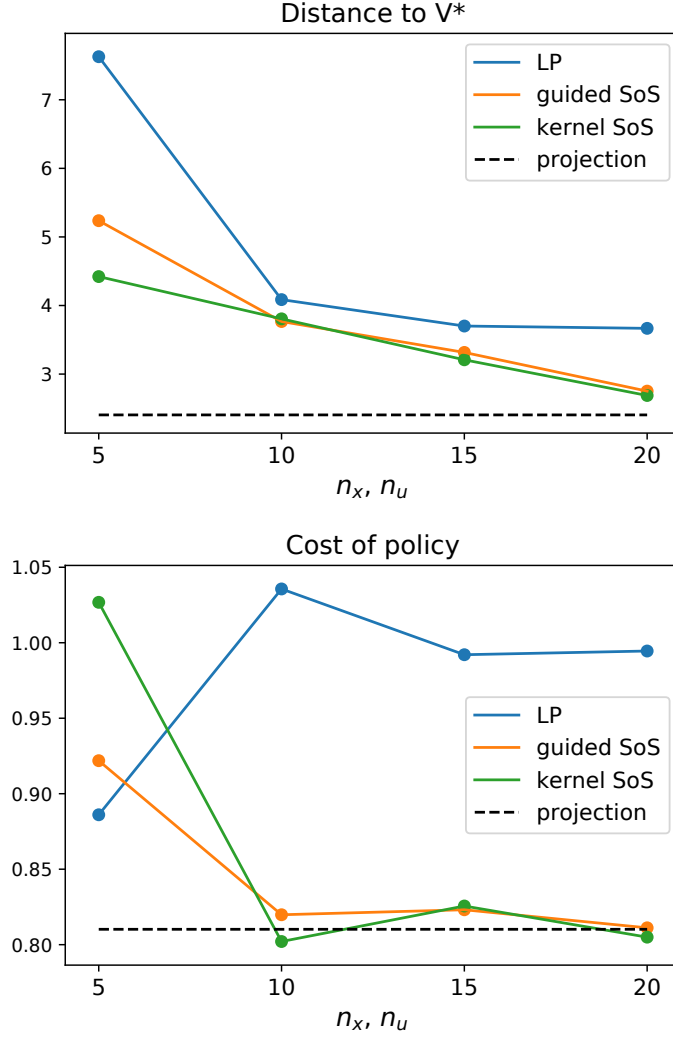


FIGURE 1. Comparison of the performances of the value function and the policy of the three methods, as a function of the number of samples  $n_x = n_u$ .

we choose the embedding vectors as follows:

$$\Psi_i := (u^{(i)}/10, x^{(i)}, (1/\omega \sin(\omega\pi/2(t/T - 1))x)_{1 \leq \omega \leq q_t})^\top,$$

where the last  $q_t$  scalar terms (without the vector  $x$ ) approximately model  $K(\cdot)$  as a smooth function of  $t$ . For computational efficiency, we choose  $q_t = 5$  and we checked that this basis can approximate the entries of  $K(t)$  well. Then we solve an SDP of size  $(p + q_t d) \times (p + q_t d) = 11 \times 11$  instead of  $n \times n$  for the kernel version.

The kernel SoS method is as described in the previous sections, with the following kernel:

$$k((t, x, u), (t', x', u')) = \langle u, u' \rangle / 100 + \langle x, x' \rangle \times \exp(-|t - t'|).$$

This kernel is also designed to match the shape of  $H^*$ , with a smooth term in  $t$  modelled by the exponential kernel. The matrix  $K$  can be singular, so we replace it by  $K + 10^{-8}I_n$ .

Results. We compare the performance of the three methods to a baseline: the projection of  $V^*$  on  $\mathcal{F}_\Theta$ , which is a proxy for the best performance to expect with a fixed  $\mathcal{F}_\Theta$ . We keep the best set of hyper-parameters after a grid search on  $(\lambda_\theta, \gamma)$  for the LP method, and on  $(\lambda_\theta, \gamma, \lambda)$  for the two others, with  $\varepsilon = 10^{-4}$ . We keep  $\eta = 0$  and  $n_t = 20$  in all the experiments, and a varying number  $n_x = n_u$  of sample points. For example, with  $n_x = n_u = 20$ , the dual variable  $\alpha$  of the largest problem here has dimension  $n = 8000$ , and solving the numerical problem written in Python takes a few minutes on a standard laptop.

The results are presented in Figure 1. The guided and kernel SoS methods perform similar, and better than the LP: they better exploit a fixed number of samples than the LP. Note that the kernel SoS tends to the LP when  $\lambda$  tends to 0, hence using a positive  $\lambda$  improves the results. We believe that the design of a kernel adapted to prior knowledge on the problem is crucial to benefit from this effect. Finally, the kernel SoS has the same performance as the guided SoS, but it is computationally more expensive as soon as  $n > 11$ . Yet the kernel version extends way beyond such fixed finite-dimensional embeddings, to infinite-dimensional embeddings represented by any positive definite kernel, including the exponential kernel, the polynomial kernel and many others.

## CONCLUSION

The kernel SoS approximation method generalizes the polynomial SoS method for OCPs. Like the simple LP method, it is *black-box* in the sense that it is based only on function evaluations of the dynamics and loss, without requiring any gradients. Moreover, it enables to exploit prior knowledge on the structure of an OCP, by choosing an appropriate kernel. The problem reduces to an SDP, whose size can be computationally limiting, but parallel implementations are possible. There are several sources of approximation in this method: the parameterization  $\mathcal{F}_\Theta$  of  $V$  might not be exact, the SoS representation of  $H^*$  is not exact in general (although we have proved it is in a few particular cases), and we subsample a finite number of constraints (further work is needed to evaluate the effect of this step). In particular, it seems essential to assess in which cases the subsampling step is tight with a finite number of samples, or approximately so, in such a way that the overall process gives a certified lower-bound on the OCP, similarly to [LHPT08]. For all these reasons, the method will probably not reach high precision solutions, but can be used to initialize direct shooting methods, and returns an approximate solution even with very few samples. Furthermore, we believe it is possible to extend the method to also account for state constraints, similarly to [LHPT08]. One could also parameterize the value function directly in an RKHS. Another interesting extension is to apply the method to Markov decision processes, where we could deal with states or actions that are more complex objects (graphs, trajectories, DNA sequences...) with appropriate kernels.

## ACKNOWLEDGEMENTS

This work was supported by the Direction Générale de l'Armement, and by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). We also acknowledge support from the European Research Council (grants SEQUOIA 724063 and REAL 947908). We thank Didier Henrion for interesting discussions related to this work.

## REFERENCES

- [Aro50] Nachman Aronszajn, *Theory of reproducing kernels*, Transactions of the American Mathematical Society **68** (1950), no. 3, 337–404.
- [BTA11] Alain Berlinet and Christine Thomas-Agnan, *Reproducing kernel hilbert spaces in probability and statistics*, Springer, 2011.
- [BV04] Stephen P. Boyd and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
- [CL83] Michael G. Crandall and Pierre-Louis Lions, *Viscosity solutions of Hamilton-Jacobi equations*, Trans. Am. Math. Soc. **277** (1983), no. 1, 1–42.
- [DFVR03] Daniela Pucci De Farias and Benjamin Van Roy, *The linear programming approach to approximate dynamic programming*, Operations Research **51** (2003), no. 6, 850–865.
- [FR12] Wendell H. Fleming and Raymond W. Rishel, *Deterministic and stochastic optimal control*, Springer Science and Business Media, 2012.
- [GPS17] Vladimir Gaitsgory, Alex Parkinson, and Ilya Shvartsman, *Linear programming formulations of deterministic infinite horizon optimal control problems in discrete time*, Discrete & Continuous Dynamical Systems **22** (2017), no. 10, 3821–3838.
- [GQ09] Vladimir Gaitsgory and Marc Quincampoix, *Linear programming approach to deterministic infinite horizon optimal control problems with discounting*, SIAM Journal on Control and Optimization **48** (2009), 2480–2512.
- [Hen13] Didier Henrion, *Optimization on linear matrix inequalities for polynomial systems control*, arXiv preprint arXiv:1309.3112 (2013).
- [HHHLT96] Daniel Hernández-Hernández, Onésimo Hernández-Lerma, and Michael Taksar, *The linear programming approach to deterministic optimal control problems*, Applicationes Mathematicae **24** (1996), no. 1, 17–33.

- [KSEL17] Angeliki Kamoutsi, Tobias Sutter, Peyman Mohajerin Esfahani, and John Lygeros, *On infinite linear programming and the moment approach to deterministic infinite horizon discounted optimal control problems*, IEEE Control Systems Letters **1** (2017), no. 1, 134–139.
- [Las09] Jean-Bernard Lasserre, *Moments, positive polynomials and their applications*, vol. 1, World Scientific, 2009.
- [Las15] ———, *An introduction to polynomial and semi-algebraic optimization*, vol. 52, Cambridge University Press, 2015.
- [LHPT08] Jean-Bernard Lasserre, Didier Henrion, Christophe Prieur, and Emmanuel Trélat, *Nonlinear optimal control via occupation measures and LMI-relaxations*, SIAM Journal on Control and Optimization **47** (2008), no. 4, 1643–1666.
- [Lib11] Daniel Liberzon, *Calculus of variations and optimal control theory: A concise introduction*, Princeton University Press, 2011.
- [Lie96] Gary M. Lieberman, *Second order parabolic differential equations*, World scientific, 1996.
- [LMMN21] Fan Lu, Prashant G. Mehta, Sean P. Meyn, and Gergely Neu, *Convex Q-learning*, 2021 American Control Conference (ACC), 2021.
- [MFBR20] Ulysse Marteau-Ferey, Francis Bach, and Alessandro Rudi, *Non-parametric models for non-negative functions*, Advances in Neural Information Processing Systems, 2020.
- [MLS94] Richard M. Murray, Zexiang Li, and S. Shankar Sastry, *A mathematical introduction to robotic manipulation*, CRC press, 1994.
- [Nem04] Arkadi Nemirovski, *Interior point polynomial time methods in convex programming*, Lecture notes (2004).
- [Nov06] Erich Novak, *Deterministic and stochastic error bounds in numerical analysis*, Springer, 2006.
- [PHL17] Edouard Pauwels, Didier Henrion, and Jean-Bernard Lasserre, *Positivity certificates in optimal control*, Geometric and Numerical Foundations of Movements, Springer, 2017.
- [PR16] Vern I. Paulsen and Mrinal Raghupathi, *An introduction to the theory of reproducing kernel hilbert spaces*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2016.
- [RMFB20] Alessandro Rudi, Ulysse Marteau-Ferey, and Francis Bach, *Finding global minima via kernel approximations*, Tech. Report 2012.11978, arXiv, 2020.
- [SB18] Richard S. Sutton and Andrew G. Barto, *Reinforcement learning: An introduction*, MIT press, 2018.
- [Sch81] Laurent Schwartz, *Cours d’analyse*, vol. 1, Hermann, 1981.
- [Sob67] I. M. Sobol’, *On the distribution of points in a cube and the approximate evaluation of integrals*, USSR Comp. Math. Math. Phys. (1967).
- [Tré05] Emmanuel Trélat, *Contrôle optimal: Théorie & applications*, Vuibert Paris, 2005.
- [Tré12] ———, *Optimal control and applications to aerospace: some results and challenges*, Journal of Optimization Theory and Applications **154** (2012), no. 3, 713–758.
- [Vin93] Richard Vinter, *Convex duality and nonlinear optimal control*, SIAM Journal on Control and Optimization **31** (1993), no. 2, 518–538.

## APPENDIX A. PROOF OF THEOREM 1

*Proof.* Consider the Hamiltonian with  $f$  control-affine:

$$H^*(t, x, u) = \nabla V^*(t, x)^\top g(t, x) + \frac{\partial V^*}{\partial t}(t, x) + L(t, x, u) + \nabla V^*(t, x)^\top B(t, x)u.$$

Pontryagin’s maximum principle states that:

$$\forall(t, x) \in \Omega_1, \quad \inf_{u \in \mathcal{U}} H^*(t, x, u) = 0.$$

Since  $L$  is strongly convex in  $u$ , the minimizer is unique and we call it  $u^*(t, x)$ . By definitions of  $\Omega_1$  and  $\Omega_2$ , we have a mapping  $u^* : \Omega_1 \rightarrow \Omega_2$  and it is characterized by:

$$\begin{aligned} \nabla_u H^*(t, x, u) &= \nabla_u L(t, x, u^*(t, x)) + B(t, x)^\top \nabla_x V^*(t, x) \\ &= 0. \end{aligned}$$

Since  $(t, x, u) \mapsto \nabla_u^2 L(t, x, u)$  is continuous on  $\Omega$  and invertible, and  $(t, x, u) \mapsto \nabla_u H^*(t, x, u) \in C^s(\Omega)$ , then the implicit function theorem ensures that  $u^* \in C^s(\Omega_1)$  (see [Sch81], Chapter 8, Theorems 25 & 31).

For  $(t, x, u) \in \Omega$ , we use Taylor’s formula around  $u^*(t, x)$ :

$$\begin{aligned} H^*(t, x, u) &= H^*(t, x, u^*(t, x)) \\ &\quad + \nabla_u H^*(t, x, u^*(t, x))^\top (u - u^*(t, x)) \\ &\quad + (u - u^*(t, x))^\top R(t, x, u)(u - u^*(t, x)), \end{aligned}$$

$$R(t, x, u) := \int_0^1 (1 - \tau) \nabla_u^2 H^*(t, x, (1 - \tau)u^*(t, x) + \tau u) d\tau.$$

Since  $H^*(t, x, u^*(t, x)) = 0$ ,  $\nabla_u H^*(t, x, u^*(t, x)) = 0$  (by definition), and  $\nabla_u^2 H^*(t, x, \cdot) = \nabla_u^2 L(t, x, \cdot)$ , we have:

$$H^*(t, x, u) = (u - u^*(t, x))^\top R(t, x, u) (u - u^*(t, x)), \text{ and}$$

$$R(t, x, u) = \int_0^1 (1 - \tau) \nabla_u^2 L(t, x, (1 - \tau)u^*(t, x) + \tau u) d\tau \succcurlyeq \frac{\rho}{2} I.$$

For  $(t, x, u) \in \Omega$ ,  $R(t, x, u)$  has a positive-definite square root  $\sqrt{R(t, x, u)}$ . Also,  $\forall \tau \in [0, 1]$ ,  $(1 - \tau)u^*(t, x) + \tau u \in \Omega_2$  because  $\text{Int}(\mathcal{U})$  is convex like  $\mathcal{U}$ .

Since  $\forall i, j, \frac{\partial^2 L}{\partial u_i \partial u_j} \in C^s(\Omega)$ ,  $u^* \in C^s(\Omega_1)$ , and  $\sqrt{\cdot}$  is  $C^\infty$  on  $\{M \mid M^\top = M, M \succcurlyeq \frac{\rho}{2} I\}$ , then  $r_{i,j} : (t, x, u) \mapsto e_i^\top \sqrt{R(t, x, u)} e_j \in C^s(\Omega)$ , and we have the decomposition:

$$H^*(t, x, u) = \sum_{i=1}^p w_i(t, x, u)^2, \quad \text{with}$$

$$\begin{aligned} w_i(t, x, u) &:= \sqrt{R(t, x, u)}_{i \cdot} (u - u^*(t, x)) \\ &= \sum_{j=1}^p r_{i,j}(t, x, u) (e_j^\top (u - u^*(t, x))), \end{aligned}$$

and each  $w_i \in C^s(\Omega)$ . □