Linearly-constrained Linear Quadratic Regulator from the viewpoint of kernel methods

Pierre-Cyril Aubin-Frankowski[*]

February 6, 2021

Abstract

The linear quadratic regulator problem is central in optimal control and was investigated since the very beginning of control theory. Nevertheless, when it includes affine state constraints, it remains very challenging from the classical "maximum principle" perspective. In this study we present how matrix-valued reproducing kernels allow for an alternative viewpoint. We show that the quadratic objective paired with the linear dynamics encode the relevant kernel, defining a Hilbert space of controlled trajectories. Drawing upon kernel formalism, we introduce a strengthened continuous-time convex optimization problem which can be tackled exactly with finite dimensional solvers, and which solution is interior to the constraints. When refining a time-discretization grid, this solution can be made arbitrarily close to the solution of the state-constrained Linear Quadratic Regulator. We illustrate the implementation of this method on a path-planning problem.

1 Introduction

In its simplest form, the problem of time-varying linear quadratic optimal control with finite horizon and affine inequality state constraints writes as

$$
\begin{aligned}
\min_{\mathbf{x}(\cdot),\mathbf{u}(\cdot)} \quad & g(\mathbf{x}(T)) + \int_0^T [\mathbf{x}(t)^\top \mathbf{Q}(t)\mathbf{x}(t) + \mathbf{u}(t)^\top \mathbf{R}(t)\mathbf{u}(t)]\mathrm{d}t \\
\text{s.t.} \quad & \mathbf{x}(0) = \mathbf{x}_0, \\
& \mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t), \text{ a.e. in } [0,T], \\
& \mathbf{c}_i(t)^\top \mathbf{x}(t) \le d_i(t), \, \forall\, t \in [0,T], \forall\, i \in \{1,\dots,P\},
\end{aligned}
\tag{$\mathcal{P}_0$}
$$

where the state $\mathbf{x}(t) \in \mathbb{R}^N$, the control $\mathbf{u}(t) \in \mathbb{R}^M$, $\mathbf{A}(t) \in \mathbb{R}^{N,N}$, $\mathbf{B}(t) \in \mathbb{R}^{N,M}$, $\mathbf{C}(t) = [\mathbf{c}_1(t)^\top; \dots; \mathbf{c}_P(t)^\top] \in \mathbb{R}^{P,N}$ ($\mathbf{c}_i(t) \in \mathbb{R}^N$), $\mathbf{d}(t) = (d_i(t))_i \in \mathbb{R}^P$, while $\mathbf{Q}(t) \in \mathbb{R}^{N,N}$ and $\mathbf{R}(t) \in \mathbb{R}^{M,M}$ are positive semidefinite matrices.

Below, for $q \in \{1, 2, \infty\}$, $L^q(0,T)$ denotes the L^q-space of functions over $[0,T]$ with integrable norms of the function values (resp. square-integrable, resp. bounded). We shall henceforth assume that, for all $t \in [0,T]$, $\mathbf{R}(t) \succcurlyeq r\mathrm{Id}_M$ with $r > 0$, as well as $\mathbf{A}(\cdot) \in L^1(0,T)$, $\mathbf{B}(\cdot) \in L^2(0,T)$,

---

[*]École des Ponts ParisTech and CAS, MINES ParisTech, PSL Research University, France. Email: pierre-cyril.aubin@mines-paristech.fr

$\mathbf{Q}(\cdot) \in L^1(0, T)$, and $\mathbf{R}(\cdot) \in L^2(0, T)$. To have a finite objective, it is natural to restrict our attention to measurable controls satisfying $\mathbf{R}(\cdot)^{1/2}\mathbf{u}(\cdot) \in L^2(0, T)$.

Without state constraints, under mild assumptions, the unconstrained Linear Quadratic Regulator (LQR) enjoys an explicit solution defined through the Hamiltonian matrix and the related Riccati equation (see e.g. Speyer and Jacobson (2010)). With state constraints, little can be said as Pontryagin's Maximum Principle involves not only an adjoint vector but also measures supported on the constraint boundary. A comprehensive review of this approach can be found in Hartl et al. (1995). One has to guess beforehand when the state-constraint is active (at the so-called junction times) in order to write the first-order necessary condition (Hermant, 2009). Secondly one has to impose assumptions to derive the magnitude of the discontinuities of the adjoint vector. This has proven to be intractable and made state-constrained continuous-time optimization a difficult problem. Let us provide an intuition for the appearance of discontinuities. If one follows an optimal trajectory of the LQR starting in the interior of state constraints, one may reach the boundary while the unconstrained Hamiltonian system of the Maximum Principle may incite to use a control leading to violate the constraint. One has then to apply a different control to remain in the constraint set, possibly generating a discontinuity in the adjoint vector.

Although LQR problems stand at the origin of control theory, research is still active in the field, not only for its numerous applications (see e.g. the examples of Burachik et al. (2014) and references within), but also for its theoretical aspects, even without constraints (Bourdin and Trélat, 2017) or just control constraints (Burachik et al., 2014). Many of these improvements are motivated by model predictive control, considered for instance in a time-invariant discrete-time state-constrained setting in Grüne and Guglielmi (2018) or continuous-time (van Keulen, 2020). In particular, Kojima and Morari (2004) proved that the solutions of a time-invariant LQR with discretized constraints converge to the solution of $(\mathcal{P}_0)$, putting emphasis on function spaces of controls. As a matter of fact, the aforementioned approaches focus on the control, used to parametrize the trajectories. In the present study, trajectories are instead at the core of the analysis.

When seeking a continuous-time numerical solution, one has to face an infinite number of pointwise constraints, and has either to relax the computationally intractable optimization problem or tighten it. *Relaxing* means either enforcing the constraint only at a finite number of points, without guarantees elsewhere (as with any time discretization method, e.g. Kojima and Morari, 2004), or through soft penalties (Gerdts and Hüpping, 2012), such as approximations of barrier functions (Dower et al., 2019). *Tightening* usually implies either choosing $\mathbf{u}(\cdot)$ in a convenient subspace of $L^2(0, T)$, such as the one of piecewise constant functions[1] or of splines with prescribed knots (Mercy et al., 2016), or through hard penalties, such as logarithmic barriers (Chaplais et al., 2011). Let us illustrate the difference between relaxing and tightening state constraints. Consider the problem of a traffic regulator whose aim is to enforce a speed limit over a highway. The drivers for their part want to go as far as possible in a given time. Deploying speed cameras ensures at best that the speed constraint is satisfied locally (relaxing). However if a smaller maximum speed is imposed at the camera level (tightening), then the cars cannot accelerate enough to break the speed limit before reaching the next camera. In a nutshell, the kernel methods framework we advocate allows to compute both a threshold and the resulting trajectories.

Kernel methods, being related to Green's functions, belong to a branch of functional analysis.

---

[1]This is known as *sampled-data* or *digital control* (Ackermann, 1985), the *sampled-data* terminology does not refer to machine learning techniques.

Their history was already sketched by Aronszajn (1950), to whom we owe the modern formulation of the theory. The regain of interest thanks to support vector machines (see e.g. Schölkopf and Smola, 2002) has reinstated kernel methods as the most principled technique in machine learning. There have been many attempts since then at bridging kernel methods and control theory. The two are already related in the works of Parzen (1970); Kailath (1971). More recently Steinke and Schölkopf (2008); Marco et al. (2017) have considered kernels for control systems, mainly to encode the input $\mathbf{u}(\cdot)$ or for system identification purposes (see e.g. the reviews of Pillonetto et al. (2014); Chiuso and Pillonetto (2019)). Kernels have also been applied to approximate the Koopman operator over observers of uncontrolled nonlinear systems, in connection with spectral analysis (Williams et al., 2015; Fujii and Kawahara, 2019) or for given controls (Sootla et al., 2018). The kernel Hilbert spaces have also been used to define suitable domains for operators (Rosenfeld et al., 2019; Giannakis et al., 2019). In most cases, the kernel is taken off-the-shelf, as with Gaussian kernels in connection with Bayesian inference (Singh et al., 2018; Bertalan et al., 2019).

On the contrary, departing from the prevalent perspective of using kernel methods as nonlinear embeddings, this article rekindles with a long standing tradition of engineering kernels for specific uses. This view has been mainly supported by the statistics community, especially in connection with splines and Sobolev spaces (Wahba, 1990; Heckman, 2012).[2] For $(\mathcal{P}_0)$, we show below that the quadratic objective paired with the linear dynamics encode the relevant kernel, which defines the Hilbert space of controlled trajectories. As kernel methods deal with a special class of Hilbert spaces, they are natural to consider for linear systems or for linearizations of nonlinear systems. Nonetheless the interactions run deeper. For instance we prove below that the controllability Gramian is directly related to matrix-valued kernels, and we recover the transversality condition merely through a representer theorem. This approach was further extended in Aubin-Frankowski (2020) to the connexion between the Linear-Quadratic matrix-valued kernel defined below and the dual Riccati equation. Since the Riccati equation is often used for the online purpose of finding the control, Aubin-Frankowski (2020) also discusses how the kernel formalism effectively allows for optimal synthesis, favoring an offline trajectory-focused viewpoint.

Our main result is to show that through the theory of kernel methods we can solve a novel strengthened version of $(\mathcal{P}_0)$, without changing the function spaces involved. This is achieved by introducing a finite number of second-order cone constraints, stronger than the infinite number of affine constraints. We furthermore show that the solution of the strengthened problem can be made arbitrarily close to the solution of $(\mathcal{P}_0)$, and that it enjoys a finite representation. One can thus exactly solve the continuous-time problem using only finite dimensional convex optimization solvers. This computable trajectory, which is interior to the affine state constraints, can also foster intuitions on the behavior of the optimal solution of $(\mathcal{P}_0)$. The considered tightening of $(\mathcal{P}_0)$ relies solely on the kernel formalism. It was first introduced in Aubin-Frankowski and Szabó (2020b) and thoroughly extended in Aubin-Frankowski and Szabó (2020a).

In Section 2 we present our strengthened problem and show how the LQR problem can be expressed as a regression problem over a vector-valued reproducing kernel Hilbert space. Section 3 details the consequences of this framework and identifies the corresponding Linear-Quadratic kernel. Our main result on approximation guarantees is stated in Section 4. Section 5 discusses the implementation and the numerical behavior of the strengthened constraints. The setting is

---

[2]Drawing inspiration from linear control theory, *control theoretic splines* were devised (Magnus Egerstedt, 2009; Fujioka and Kano, 2013), in particular for path-planning problems (Kano and Fujioka, 2018). Possibly unbeknownst to non-kernel users, kernel theory, sometimes known as *abstract splines*, is the natural generalization of splines (see e.g. Aubin-Frankowski et al., 2020).

also extended to intermediate or terminal equality constraints, as in path-planning problems. The annex pertains to conditions ensuring the existence of interior trajectories for ($\mathcal{P}_0$).

2 Theoretical preliminaries and problem formulation

In this section, we present the tools from the theory of kernel methods that we shall apply. We then introduce our strengthened problem with second-order cone constraints.

Notations: We use the shorthand $[\![1, P]\!] = \{1, \dots, P\}$. $\mathbb{R}^N_+$ is the subset of $\mathbb{R}^N$ of elements with nonnegative components. $\mathbb{B}_N$ denotes the closed unit ball of $\mathbb{R}^N$ for the Euclidean inner product, $\mathbf{1}_N$ the vector of all-ones. For a matrix $\mathbf{A} \in \mathbb{R}^{N,N}$, we write by $\|\mathbf{A}\|$ its operator norm. Id_N is the identity matrix of $\mathbb{R}^{N,N}$. We chose not to explicit the output space for the function spaces to avoid cumbersome notations, as it can be always deduced from the context. The space of functions with continuous derivatives up to order s is denoted by $\mathcal{C}^s(0, T)$. For a function $K(\cdot, \cdot)$ defined over a subset of $\mathbb{R} \times \mathbb{R}$, $\partial_1 K(\cdot, \cdot)$ denotes the partial derivative w.r.t. the first variable. For a Hilbert space $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_K)$, $\mathbb{B}_K$ is the closed unit ball of $\mathcal{H}_K$, $\| \cdot \|_K$ denoting the corresponding norm. Given a subspace $V \subset \mathcal{H}_K$, we denote by $V^\perp$ its orthogonal complement w.r.t. $\langle \cdot, \cdot \rangle_K$. A μ-strongly convex function $\mathcal{L} : \mathcal{H}_K \mapsto \mathbb{R}$ is a function satisfying, for all $f_1, f_2 \in \mathcal{H}_K$, $\alpha \in [0, 1]$, $\mathcal{L}(\alpha f_1 + (1 - \alpha) f_2) + \alpha(1 - \alpha) \frac{\mu}{2} \|f_1 - f_2\|^2_K \leq \alpha \mathcal{L}(f_1) + (1 - \alpha) \mathcal{L}(f_2)$.

Definition 1. Let $\mathcal{T}$ be a non-empty set. A Hilbert space $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_K)$ of $\mathbb{R}^N$-vector-valued functions defined on $\mathcal{T}$ is called a vector-valued reproducing kernel Hilbert space (vRKHS) if there exists a matrix-valued kernel $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}^{N,N}$ such that the *reproducing property* holds: for all $t \in \mathcal{T}$, $\mathbf{p} \in \mathbb{R}^N$, $K(\cdot, t)\mathbf{p} \in \mathcal{H}_K$ and for all $\mathbf{f} \in \mathcal{H}_K$, $\mathbf{p}^\top \mathbf{f}(t) = \langle \mathbf{f}, K(\cdot, t)\mathbf{p} \rangle_K$.

Many properties of real-valued RKHSs have been known since Aronszajn (1950), the general theory having been developed by Schwartz (1964). By Riesz's theorem, an equivalent definition of a vRKHS is that, for every $t \in \mathcal{T}$ and $\mathbf{p} \in \mathbb{R}^N$, the evaluation functional $\mathbf{f} \in \mathcal{H}_K \mapsto \mathbf{p}^\top \mathbf{f}(t) \in \mathbb{R}$ is continuous. There is also a one-to-one correspondence between the kernel K and the vRKHS $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_K)$ (see e.g. (Micheli and Glaunès, 2014, Theorem 2.6)), hence modifying $\mathcal{T}$ or changing the inner product changes the kernel. We shall use several classical properties: by symmetry of the scalar product, the matrix-valued kernel has a Hermitian symmetry, i.e. $K(s, t) = K(t, s)^\top$ for any $s, t \in \mathcal{T}$. Moreover, if the vRKHS can be written as $\mathcal{H}_K = \mathcal{H}_0 \oplus \mathcal{H}_1$, then $\mathcal{H}_0$ and $\mathcal{H}_1$ equipped with $\langle \cdot, \cdot \rangle_K$ are also vRKHSs, as closed subspaces of $\mathcal{H}_K$ for $\| \cdot \|_K$, and their kernels K_0 and K_1 satisfy $K = K_0 + K_1$.

Let us define our candidate for a vRKHS, the space $\mathcal{S}$ of trajectories satisfying the dynamical system of ($\mathcal{P}_0$):

$$\mathcal{S} := \{ \mathbf{x}(\cdot) \,|\, \exists\, \mathbf{u}(\cdot) \text{ s.t. } \mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \text{ a.e. and } \int_0^T \mathbf{u}(t)^\top \mathbf{R}(t)\mathbf{u}(t)\mathrm{d}t < \infty \}. \quad (1)$$

There is not necessarily a unique choice of $\mathbf{u}(\cdot)$ for a given $\mathbf{x}(\cdot) \in \mathcal{S}$ (for instance if $\mathbf{B}(t)$ is not injective for some t). Therefore, with each $\mathbf{x}(\cdot) \in \mathcal{S}$, we associate the control $\mathbf{u}(\cdot)$ having minimal norm based on the pseudoinverse of $\mathbf{B}(t)^\ominus$ of $\mathbf{B}(t)$ for the $\mathbb{R}^M$-norm $\| \cdot \|_{\mathbf{R}(t)} := \|\mathbf{R}(t)^{1/2} \cdot \|$:

$$\mathbf{u}(t) = \mathbf{B}(t)^\ominus [\mathbf{x}'(t) - \mathbf{A}(t)\mathbf{x}(t)] \text{ a.e. in } [0, T]. \quad (2)$$

The vector space $\mathcal{S}$ has then a natural scalar-product. As a matter of fact, the expression

$$\langle \mathbf{x}_1(\cdot), \mathbf{x}_2(\cdot) \rangle_K := \mathbf{x}_1(0)^\top \mathbf{x}_2(0) + \int_0^T [\mathbf{x}_1(t)^\top \mathbf{Q}(t)\mathbf{x}_2(t) + \mathbf{u}_1(t)^\top \mathbf{R}(t)\mathbf{u}_2(t)]\mathrm{d}t \quad (3)$$

4

is bilinear and symmetric over $\mathcal{S} \times \mathcal{S}$.[3] It is positive definite over $\mathcal{S}$ as $\|\mathbf{x}(\cdot)\|_K^2 = \mathbf{0}$ implies that $\mathbf{u}(\cdot) \overset{a.e}{\equiv} 0$ and $\mathbf{x}(0) = \mathbf{0}$, hence that $\mathbf{x}(\cdot) \equiv \mathbf{0}$. Combining (2) and (3), we can express $\| \cdot \|_K$ as a Sobolev-like norm split into two semi-norms $\| \cdot \|_{K_0}$ and $\| \cdot \|_{K_1}$

$$\|\mathbf{x}(\cdot)\|_K^2 = \underbrace{\|\mathbf{x}(0)\|^2}_{\|\mathbf{x}(\cdot)\|_{K_0}^2} + \underbrace{\int_0^T [\|\mathbf{x}(t)\|_{\mathbf{Q}(t)}^2 + \|\mathbf{B}(t)^{\ominus}(\mathbf{x}'(t) - \mathbf{A}(t)\mathbf{x}(t))\|_{\mathbf{R}(t)}^2]\mathrm{d}t}_{\|\mathbf{x}(\cdot)\|_{K_1}^2} . \tag{4}$$

By Cauchy-Lipschitz's theorem, $\|\mathbf{x}(\cdot)\|_{K_0} = \|\mathbf{x}(0)\|$ defines a norm over the finite-dimensional subspace $\mathcal{S}_0$ of trajectories with null quadratic cost (hence null control):

$$\mathcal{S}_0 := \{\mathbf{x}(\cdot) \mid \int_0^T \mathbf{x}(t)^\top \mathbf{Q}(t)\mathbf{x}(t)\mathrm{d}t = 0 \text{ and } \mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t), \text{ a.e. in } [0,T]\}. \tag{5}$$

Let us define its (infinite-dimensional) orthogonal complement $\mathcal{S}_u := (\mathcal{S}_0)^\perp$ in $\mathcal{S}$ w.r.t. $\| \cdot \|_K$. From now on we equip $\mathcal{S}$ (resp. $\mathcal{S}_0$, $\mathcal{S}_u$) with $\| \cdot \|_K$ (resp. $\| \cdot \|_{K_0}$, $\| \cdot \|_{K_1}$). These will be shown to all be vRKHSs. Suppose we identified the matrix-valued kernels spawning them (a procedure to be found in Section 3). In a recent approach in kernel methods (Aubin-Frankowski and Szabó, 2020b), developed for regression problems with constraints over derivatives, we suggested replacing "$\mathbf{C}(t)\mathbf{x}(t) \le \mathbf{d}(t) \, (\forall \, t \in [0,T])$" by the following strengthened second-order cone (SOC)[4] constraints:

$$\eta_i(\delta_m, t_m)\|\mathbf{x}(\cdot)\|_K + \mathbf{c}_i(t_m)^\top \mathbf{x}(t_m) \le d_i(\delta_m, t_m), \, \forall \, m \in [\![1, N_P]\!], \forall \, i \in [\![1, P]\!] \tag{6}$$

where the $(t_m)_{m \in [\![1, N_P]\!]} \in [0,T]^{N_P}$ are N_P time points associated to radii $\delta_m > 0$ satisfying $[0,T] \subset \cup_{m \in [\![1, N_P]\!]}[t_m - \delta_m, t_m + \delta_m]$. The constants $\eta_i(\delta_m, t_m)$ and $d_i(\delta_m, t_m)$ are then defined as:

$$\eta_i(\delta_m, t_m) := \sup_{t \in [t_m - \delta_m, t_m + \delta_m] \cap [0,T]} \|K(\cdot, t_m)\mathbf{c}_i(t_m) - K(\cdot, t)\mathbf{c}_i(t)\|_K,$$

$$d_i(\delta_m, t_m) := \inf_{t \in [t_m - \delta_m, t_m + \delta_m] \cap [0,T]} d_i(t).$$

This tightening of the constraints stems from interpreting $\eta_i(\delta_m, t_m)\|\mathbf{x}(\cdot)\|_K$ as an upper bound of the modulus of continuity of the unknown $\mathbf{C}(\cdot)\mathbf{x}(\cdot)$ defined as follows

$$\omega_i^{(\mathbf{Cx})}(\delta_m, t_m) := \sup_{t \in [t_m - \delta_m, t_m + \delta_m] \cap [0,T]} \underbrace{|\mathbf{c}_i(t)^\top \mathbf{x}(t) - \mathbf{c}_i(t_m)^\top \mathbf{x}(t_m)|}_{|\langle x(\cdot), K(\cdot,t)c(t) - K(\cdot,t_m)c(t_m)\rangle_K|} \le \eta_i(\delta_m, t_m)\|\mathbf{x}(\cdot)\|_K. \tag{7}$$

This inequality is obtained applying successively the reproducing property and the Cauchy–Schwarz inequality. Since the intractable modulus of continuity controls the variations of $\mathbf{C}(t)\mathbf{x}(t)$, the SOC upper bound provides a tractable tightening. This interpretation through moduli of continuity was extensively developed in Section 3.2 of Aubin-Frankowski and Szabó (2020a).

---

[3]The description by $\mathcal{S}$ of the optimization variables effectively pushes controls in the background while bringing forth trajectories as the main object of study. This describes $(\mathcal{P}_0)$ more as a regression problem over $\mathcal{S}$ than as an optimal control problem over controls.

[4]The "second-order cone" terminology is classical in optimization following the similarity between (6) and the definition of the Lorentz cone $\{(\mathbf{z}, r) \in \mathbb{R}^{N+1} \mid \|\mathbf{z}\| \le r\}$.

With the above notations, our strengthened time-varying linear quadratic optimal control problem with finite horizon and finite number of SOC constraints is[5]

$$\min_{\substack{\mathbf{x}(\cdot)\in\mathcal{S},\\ \mathbf{x}(0)=\mathbf{x}_0}} \quad g(\mathbf{x}(T)) + \|\mathbf{x}(\cdot)\|_K^2$$

$$(\mathcal{P}_{\delta,\text{fin}})$$

$$\text{s.t.} \quad \eta_i(\delta_m,t_m)\|\mathbf{x}(\cdot)\|_K + \mathbf{c}_i(t_m)^\top\mathbf{x}(t_m) \leq d_i(\delta_m,t_m), \ \forall m \in [\![1,N_P]\!], \forall i \in [\![1,P]\!].$$

The introduction of $(\mathcal{P}_{\delta,\text{fin}})$ as an approximation of $(\mathcal{P}_0)$ entirely stems from the vRKHS formalism and does not result from optimal control considerations. It relies on an inner approximation of a convex set in an infinite-dimensional Hilbert space (see Aubin-Frankowski and Szabó, 2020a, Section 3.1). From a machine learning perspective, the initial condition and the terminal cost act as a "loss" function, whereas the quadratic cost is turned into a norm over $\mathcal{S}$ and can thus be interpreted as a "regularizer". Departing also from optimal control, the tightening is obtained by incorporating the quadratic part of the objective (4) in the state constraints to form (6). As discussed in Aubin-Frankowski and Szabó (2020b), introducing (6) leads to a finite number of evaluations of the variable $\mathbf{x}(\cdot)$ in $(\mathcal{P}_{\delta,\text{fin}})$ which allows for a representer theorem (Theorem 1 in Section 3.1).

Our goal is to show that $(\mathcal{P}_{\delta,\text{fin}})$ is indeed a tightening of $(\mathcal{P}_0)$, enjoying a representer theorem providing a finite-dimensional representation of the solution of problem $(\mathcal{P}_{\delta,\text{fin}})$. We also bound the distance between the trajectories solutions of $(\mathcal{P}_0)$ and $(\mathcal{P}_{\delta,\text{fin}})$, and prove that it can be made as small as desired by refining the time-discretization grid.

3 Revisiting LQ control through kernel methods

This section presents a step-by-step approach to identify the matrix-valued kernel K of the Hilbert space $\mathcal{S}$ of solutions of a linear control system, equipped with the scalar product (3). This is done independently from the state constraints which effect is only to select a closed convex subset of the space of trajectories. In Section 3.1, we consider the case $\mathbf{Q} \equiv \mathbf{0}$ which enjoys explicit formulas. We also express a representer theorem (Theorem 1 suited for problems of the form $(\mathcal{P}_{\delta,\text{fin}})$. This allows us to revisit, through the kernel framework, classical notions, such as the solution of the unconstrained LQR problem, or the definition of the Gramian of controllability. In Section 3.2, we consider the case $\mathbf{Q} \not\equiv \mathbf{0}$ and relate our solution to an adjoint equation over matrices. Furthermore, the identification of kernels developed in Section 3 is by no means restricted to finite T, hence the kernel formalism can also tackle infinite-horizon problems.[6]

Let us denote by $\boldsymbol{\Phi}_\mathbf{A}(t,s) \in \mathbb{R}^{N,N}$ the state-transition matrix of $\mathbf{x}'(\tau) = \mathbf{A}(\tau)\mathbf{x}(\tau)$, defined from s to t. The key property used throughout this section is the variation of constants formula, a.k.a. Duhamel's principle, stating that for any absolutely continuous $\mathbf{x}(\cdot)$ such that $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)$ a.e. we have

$$\mathbf{x}(t) = \boldsymbol{\Phi}_\mathbf{A}(t,0)\mathbf{x}(0) + \int_0^t \boldsymbol{\Phi}_\mathbf{A}(t,\tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\mathrm{d}\tau. \tag{8}$$

Lemma 1. $(\mathcal{S}, \langle\cdot,\cdot\rangle_K)$ *is a vRKHS.*

---

[5]Since $\|\mathbf{x}(\cdot)\|_{K_0}^2 = \|\mathbf{x}(0)\|^2$ and $\mathbf{x}(0)$ is fixed, we replaced the integral $\|\mathbf{x}(\cdot)\|_{K_1}^2$ by $\|\mathbf{x}(\cdot)\|_K^2$ in the objective. Recently, in Aubin-Frankowski (2020), another choice of inner product was introduced, by incorporating quadratic terminal costs g into $\|\mathbf{x}(\cdot)\|_K^2$ for another choice of inner product than (3).

[6]We need to assume T to be finite for the results of Section 4 to hold, as we use a procedure based on compact coverings to deal with the state constraints.

Proof: We have to show that: *i*) $(\mathcal{S}, \langle \cdot, \cdot \rangle_K)$ is a Hilbert space, *ii*) for every $t \in [0, T]$ and $\mathbf{p} \in \mathbb{R}^N$, the evaluation functional $\mathbf{x}(\cdot) \in \mathcal{S} \mapsto \mathbf{p}^\top \mathbf{x}(t) \in \mathbb{R}$ is continuous.

i) From (3) and the discussion of Section 2, it is obvious that $\langle \cdot, \cdot \rangle_K$ is a scalar product. We just have to show that $\mathcal{S}$ is complete. Let $(\mathbf{x}_n(\cdot))_n$ be a Cauchy sequence in $\mathcal{S}$, with associated controls $(\mathbf{u}_n(\cdot))_n$. Then $(\|\mathbf{x}_n(\cdot)\|_K)_n$ is a Cauchy sequence in $\mathbb{R}$ and thus converges, so $(\|\mathbf{x}_n(0)\|)_n$ and $(\|\mathbf{R}(\cdot)^{1/2}\mathbf{u}_n(\cdot)\|_{L^2(0,T)})_n$ are bounded. Since $\mathbf{R}(t) \succcurlyeq r\mathrm{Id}_M$ with $r > 0$, $(\|\mathbf{u}_n(\cdot)\|_{L^2(0,T)})_n$ is thus bounded too, and we can take a subsequence $(\mathbf{u}_{n_i}(\cdot))_i$ weakly converging to some $\mathbf{u}(\cdot)$. Let $s, t \in [0, T]$,

$$\mathbf{x}_n(t) - \mathbf{x}_n(s) \overset{(8)}{=} (\boldsymbol{\Phi}_{\mathbf{A}}(t,s) - \mathrm{Id}_N)\mathbf{x}_n(s) + \int_s^t \boldsymbol{\Phi}_{\mathbf{A}}(t,\tau)\mathbf{B}(\tau)\mathbf{u}_n(\tau)\mathrm{d}\tau. \tag{9}$$

Taking $s = 0$, as $(\|\mathbf{x}_n(0)\|)_n$ is bounded, $\mathbf{A}(\cdot) \in L^1(0,T)$ and $\mathbf{B}(\cdot) \in L^2(0,T)$, we have that $\boldsymbol{\Phi}_{\mathbf{A}}(\cdot, \cdot)$ is continuous and $\{\mathbf{x}_n(\cdot)\}_n$ is uniformly bounded in $\mathcal{C}(0,T)$. Thus (9) implies that the sequence $(\mathbf{x}_{n_i}(\cdot))_i$ is equicontinuous. By Ascoli's theorem, we can take a $(\mathbf{x}_{n_{i_j}}(\cdot))_j$ uniformly converging to some $\mathbf{x}(\cdot)$ satisfying (8) for $\mathbf{u}(\cdot)$, thus $\mathbf{x}(\cdot) \in \mathcal{S}$.

ii) Let $t \in [0, T]$, $\mathbf{p} \in \mathbb{R}^N$, and $\mathbf{x}(\cdot) \in \mathcal{S}$. By (8) and Cauchy-Schwarz inequality,

$$\mathbf{p}^\top \mathbf{x}(t) = \mathbf{p}^\top \boldsymbol{\Phi}_{\mathbf{A}}(t,0)\mathbf{x}(0) + \int_0^t \mathbf{p}^\top \boldsymbol{\Phi}_{\mathbf{A}}(t,\tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\mathrm{d}\tau$$

$$|\mathbf{p}^\top \mathbf{x}(t)| \leq \sup_{\tau \in [0,T]} \|\mathbf{p}^\top \boldsymbol{\Phi}_{\mathbf{A}}(t,\tau)\| \cdot \left(\|\mathbf{x}(0)\| + \left\| \int_0^t \mathbf{B}(\tau)\mathbf{u}(\tau)\mathrm{d}\tau \right\| \right)$$

$$\leq \sup_{\tau \in [0,T]} \sqrt{2}\|\mathbf{p}^\top \boldsymbol{\Phi}_{\mathbf{A}}(t,\tau)\| \cdot \left(\|\mathbf{x}(0)\|^2 + \|\mathbf{B}(\cdot)\|^2_{L^2(0,T)}\|\mathbf{u}(\cdot)\|^2_{L^2(0,T)} \right)^{\frac{1}{2}}$$

$$\leq \sup_{\tau \in [0,T]} \sqrt{2}\|\mathbf{p}^\top \boldsymbol{\Phi}_{\mathbf{A}}(t,\tau)\| \cdot \left(1 + \frac{\|\mathbf{B}(\cdot)\|_{L^2(0,T)}}{\sqrt{r}} \right) \left(\|\mathbf{x}(0)\|^2 + \|\mathbf{R}(\cdot)^{\frac{1}{2}}\mathbf{u}(\cdot)\|^2_{L^2(0,T)} \right)^{\frac{1}{2}}.$$

Hence the linear map $\mathbf{x}(\cdot) \in \mathcal{S} \mapsto \mathbf{p}^\top \mathbf{x}(t) \in \mathbb{R}$ is continuous.

∎

By Definition 1, we know that a matrix-valued reproducing kernel $K(\cdot, \cdot)$ exists. We now repeatedly use (8) to identify it.

3.1 Case $\mathbf{Q} \equiv 0$

The space $\mathcal{S}$ of controlled trajectories is defined as in (1) equipped with the quadratic norm

$$\|\mathbf{x}(\cdot)\|^2_K = \|\mathbf{x}(0)\|^2 + \int_0^T \mathbf{u}(t)^\top \mathbf{R}(t)\mathbf{u}(t)\mathrm{d}t = \|\mathbf{x}(\cdot)\|^2_{K_0} + \|\mathbf{x}(\cdot)\|^2_{K_1}, \tag{10}$$

where $\mathbf{u}(\cdot)$ is defined as in (2). We can further specify its subspaces

$$\mathcal{S}_0 = \{\mathbf{x}(\cdot) \,|\, \mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t), \text{ a.e. in } [0,T]\} \qquad \mathcal{S}_u = \{\mathbf{x}(\cdot) \,|\, \mathbf{x}(\cdot) \in \mathcal{S} \text{ and } \mathbf{x}(0) = 0\}.$$

By uniqueness of the reproducing kernel, we only have to exhibit a candidate $K(s,t)$ satisfying $\mathbf{p}^\top \mathbf{x}(t) = \langle \mathbf{x}(\cdot), K(\cdot,t)\mathbf{p} \rangle_K$ ($\forall t \in [0,T], \mathbf{x} \in \mathcal{S}, \mathbf{p} \in \mathbb{R}^N$) and $K(\cdot,t)\mathbf{p} \in \mathcal{S}$ ($\forall t \in [0,T], \mathbf{p} \in \mathbb{R}^N$). The space $(\mathcal{S}_0, \|\cdot\|_{K_0})$ being finite dimensional, we can right away identify its kernel[7]

$$K_0(s,t) = \boldsymbol{\Phi}_{\mathbf{A}}(s,0)\boldsymbol{\Phi}_{\mathbf{A}}(t,0)^\top. \tag{11}$$

---

[7]This is a classical result for finite dimensional vRKHSs. Fix any family $\{v_j\}_{j \in [\![1,N]\!]}$ spanning $\mathcal{S}_0$, let $\mathbf{V}(s) := [v_j(s)]_{j \in [\![1,N]\!]} \in \mathbb{R}^{N,N}$ and $\mathbf{G}_v := (\langle v_i, v_j \rangle_{K_0})_{i,j \in [\![1,N]\!]}$. The matrix $\mathbf{G}_v$ is invertible as $\|\cdot\|_{K_0}$ is a norm over $\mathcal{S}_0$ and thus $K_0(s,t) = \mathbf{V}(s)^\top \mathbf{G}_v^{-1}\mathbf{V}(t)$. Here we have $\mathbf{V}(s) = \boldsymbol{\Phi}_{\mathbf{A}}(s,0)^\top$ and $\mathbf{G}_v = \mathrm{Id}_N$.

As $\mathcal{S} = \mathcal{S}_0 \oplus \mathcal{S}_u$, from the properties of sums of kernels, we derive that we should look for K of the form $K_0 + K_1$ for which the reproducing property, with $\langle \cdot, \cdot \rangle_K$ defined in (4), writes as follows, for all $t \in [0,T]$, $\mathbf{p} \in \mathbb{R}^N$, $\mathbf{x}(\cdot) \in \mathcal{S}$,

$$\mathbf{p}^\top \mathbf{x}(t) = (K(0,t)\mathbf{p})^\top \mathbf{x}(0) + \int_0^T \left[(\mathbf{B}(t)^\ominus (\partial_1(K(s,t)\mathbf{p}) - \mathbf{A}(t)K(s,t)\mathbf{p}))^\top \mathbf{R}(s)\mathbf{u}(s) \right] \mathrm{d}s. \quad (12)$$

Setting $\partial_1 K(s,t) : \mathbf{p} \mapsto \partial_1(K(s,t)\mathbf{p})$, let us define formally $\mathbf{U}_t(s) := \mathbf{B}(s)^\ominus(\partial_1 K_1(s,t) - \mathbf{A}(s)K_1(s,t))$. By the Hermitian symmetry of K and the fact that $K_0(\cdot,t)\mathbf{p}$ belongs to $\mathcal{S}_0$ and $K_1(0,t)\mathbf{p} = \mathbf{0}$, (12) holds if and only if for all $t \in [0,T]$ and $\mathbf{x}(\cdot) \in \mathcal{S}$

$$\mathbf{x}(t) = K_0(t,0)\mathbf{x}(0) + \int_0^T \mathbf{U}_t(s)^\top \mathbf{R}(s)\mathbf{u}(s)\mathrm{d}s.$$

This expression can be identified with (8) when defining $\mathbf{U}_t(s)$ as follows

$$\mathbf{U}_t(s) := \begin{cases} \mathbf{R}(s)^{-1}\mathbf{B}(s)^\top \mathbf{\Phi_A}(t,s)^\top & \forall s \leq t, \\ \mathbf{0} & \forall s > t. \end{cases} \quad (13)$$

Consequently, to ensure that, for all $t \in [0,T]$, $\mathbf{p} \in \mathbb{R}^N$, $K_1(\cdot,t)\mathbf{p} \in \mathcal{S}_u$, K_1 has to satisfy the following differential equation for any given $t \in [0,T]$:

$$\partial_1 K_1(s,t) = \mathbf{A}(s)K_1(s,t) + \mathbf{B}(s)\mathbf{U}_t(s) \text{ a.e. in } [0,T] \text{ with } K_1(0,t) = \mathbf{0}. \quad (14)$$

Since $\mathbf{A}(\cdot) \in L^1(0,T)$, $\mathbf{B}(\cdot) \in L^2(0,T)$, and $\mathbf{R}(t) \succcurlyeq r\mathrm{Id}_M$ with $r > 0$, we have that $\mathbf{U}_t(\cdot) \in L^2(0,T)$. By applying the variation of constants formula to (14), with $\mathbf{U}_t(\cdot)$ defined in (13), we get an explicit expression for K_1, satisfying a Hermitian symmetry when permuting s and t,

$$K_1(s,t) = \int_0^{\min(s,t)} \mathbf{\Phi_A}(s,\tau)\mathbf{B}(\tau)\mathbf{R}(\tau)^{-1}\mathbf{B}(\tau)^\top \mathbf{\Phi_A}(t,\tau)^\top \mathrm{d}\tau. \quad (15)$$

Remark (Gramian): Formula (15) for $K_1(T,T)$ corresponds to the Gramian of controllability. The link is straightforward as the controllability problem of steering a point from $(0,\mathbf{0})$ to $(T,\mathbf{x}_T)$ simply writes as, with $\mathbf{u}(\cdot)$ defined as in (2) and $\mathbf{R}(\cdot) \equiv \mathrm{Id}_M$,

$$\begin{aligned} \min_{\mathbf{x}(\cdot) \in \mathcal{S}} \quad & \int_0^T \|\mathbf{u}(t)\|^2 \mathrm{d}t \\ \text{s.t.} \quad & \mathbf{x}(0) = \mathbf{0}, \\ & \mathbf{x}(T) = \mathbf{x}_T, \end{aligned} \quad (16)$$

which set of solutions can actually be made explicit.[8] As a matter of fact, in kernel methods, it is classical to look for a "representer theorem", i.e. a necessary condition to ensure that the solutions of an optimization problem live in a finite dimensional subspace of $\mathcal{S}$ and consequently enjoy a finite representation. Such theorems are usually stated without constraints and for real-valued kernels (e.g. Schölkopf et al., 2001). Here we formulate a representer theorem for conic constraints and matrix-valued kernels, as it will prove instrumental to derive a finite formulation for the SOC-strengthening of $(\mathcal{P}_0)$.

---

[8]When choosing the terminal cost $g(\cdot)$ to be equal to the indicator function of $\{\mathbf{x}_T\}$, (16) does correspond to $(\mathcal{P}_0)$ in the absence of state constraints.

Theorem 1 (Representer theorem)**.** *Let* $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_K)$ *be a vRKHS defined on a set* $\mathcal{T}$*. Let* $P \in \mathbb{N}$ *and, for* $i \in [\![0, P]\!]$ *and given* $N_i \in \mathbb{N}$*,* $\{t_{i,j}\}_{j \in [\![1,N_i]\!]} \subset \mathcal{T}$*. Consider the following optimization problem with "loss" function* $L : \mathbb{R}^{N_0} \to \mathbb{R} \cup \{+\infty\}$*, strictly increasing "regularizer" function* $\Omega : \mathbb{R}_+ \to \mathbb{R}$*, and constraints* $d_i : \mathbb{R}^{N_i} \to \mathbb{R}$*,* $\lambda_i \geq 0$ *and* $\{\mathbf{c}_{i,m}\}_{m \in [\![1,N_i]\!]} \subset \mathbb{R}^N$*,*

$$\bar{\mathbf{f}} \in \arg\min_{\mathbf{f} \in \mathcal{H}_K} \quad L\left(\mathbf{c}_{0,1}^\top \mathbf{f}(t_{0,1}), \ldots, \mathbf{c}_{0,N_0}^\top \mathbf{f}(t_{0,N_0})\right) + \Omega\left(\|\mathbf{f}\|_K\right)$$

$$\text{s.t.} \quad \lambda_i \|\mathbf{f}\|_K \leq d_i(\mathbf{c}_{i,1}^\top \mathbf{f}(t_{i,1}), \ldots, \mathbf{c}_{i,N_i}^\top \mathbf{f}(t_{i,N_i})), \, \forall\, i \in [\![1, P]\!].$$

Then, for any minimizer $\bar{\mathbf{f}}$*, there exists* $\{\mathbf{p}_{i,m}\}_{m \in [\![1,N_i]\!]} \subset \mathbb{R}^N$ *such that* $\bar{\mathbf{f}} = \sum_{i=0}^{P} \sum_{m=1}^{N_i} K(\cdot, t_{i,m}) \mathbf{p}_{i,m}$ *with* $\mathbf{p}_{i,m} = \alpha_{i,m} \mathbf{c}_{i,m}$ *for some* $\alpha_{i,m} \in \mathbb{R}$*.*

Proof: Let $\bar{\mathbf{f}}$ be an optimal solution and let $V := \operatorname{span}\left(\{K(\cdot, t_{i,m})\mathbf{c}_{i,m}\}_{m \leq N_i,\, i \leq P}\right)$. Take $\mathbf{v} \in V$ and $\mathbf{w} \in V^\perp$ such that $\bar{\mathbf{f}} = \mathbf{v} + \mathbf{w}$. As $\mathbf{c}_{i,m}^\top \mathbf{w}(t_{i,m}) = \langle \mathbf{w}(\cdot), K(\cdot, t_{i,m})\mathbf{c}_{i,m} \rangle_K = 0$, the terms appearing in L and d_i are the same for $\bar{\mathbf{f}}$ and $\mathbf{v}$. Moreover, $\|\mathbf{v}\|_K \leq \|\bar{\mathbf{f}}\|_K$, hence $\mathbf{v}$ belongs to the constraint set since $\bar{\mathbf{f}}$ does. Furthermore $\Omega\left(\|\mathbf{v}\|_K\right) \leq \Omega\left(\|\bar{\mathbf{f}}\|_K\right)$, so, by optimality of $\bar{\mathbf{f}}$, $\mathbf{w} = 0$ which concludes the proof.

∎

In other words, Theorem 1 states that to each time t where the variable $\mathbf{f}$ is evaluated corresponds a multiplier $\mathbf{p}_t \in \mathbb{R}^N$ in the expression of the optimal solutions. Hence, if the number of such evaluations is finite, then the representation of $\bar{\mathbf{f}}$ is finite.[9] Besides, a representer theorem, like Pontryagin's Maximum Principle, is only a necessary condition on the form of the solutions. Theorem 1 guarantees the existence or uniqueness of an optimal solution only when coupled with other assumptions (e.g. that L and Ω are convex, and L is lower semi-continuous). This further highlights the analogy with the Maximum Principle in the quadratic case. Theorem 1 is instrumental to obtain a finite-dimensional equivalent of $(\mathcal{P}_{\delta,\text{fin}})$.

Theorem 1 applied to (16) implies that any candidate optimal solution $\bar{\mathbf{x}}(\cdot)$ can be written as $\bar{\mathbf{x}}(s) = K(s, 0)\mathbf{p}_0 + K(s, T)\mathbf{p}_T$, with $\mathbf{p}_0, \mathbf{p}_T \in \mathbb{R}^N$. As $\bar{\mathbf{x}}(0) = \mathbf{0}$, $\operatorname{proj}_{\mathcal{S}_0}(\bar{\mathbf{x}}(\cdot)) = \mathbf{0}$, and as $K_1(\cdot, 0) \equiv \mathbf{0}$, $\bar{\mathbf{x}}(s) = K_1(s, T)\mathbf{p}_T$. So $\bar{\mathbf{x}}(T) = \mathbf{x}_T$ is satisfied if and only if $\mathbf{x}_T \in \operatorname{Im}(K_1(T, T))$ where the operator $K_1(T, T)$ is defined by (14), setting $\mathbf{R} \equiv \text{Id}$. Hence (16) has a solution for any $\mathbf{x}_T$ (i.e. the system is controllable) if and only if the Gramian of controllability $K_1(T, T)$ is invertible.

Remark (LQR without state constraints): We derive also from the kernel framework the transversality condition, as well as the classical solution of the LQR problem without state constraints, defined as follows, with $\mathbf{u}(\cdot)$ again defined as in (2),

$$\min_{\mathbf{x}(\cdot) \in \mathcal{S}} \quad g(\mathbf{x}(T)) + \frac{1}{2} \int_0^T \mathbf{u}(t)^\top \mathbf{R}(t)\mathbf{u}(t)\mathrm{d}t \qquad (\mathcal{P}_{\text{uncons}})$$

$$\text{s.t.} \quad \mathbf{x}(0) = \mathbf{0}.$$

Similarly, through the representer theorem, we deduce that $\bar{\mathbf{x}}(\cdot) = K_1(\cdot, T)\mathbf{p}_T$. Hence, by the reproducing property,

$$\int_0^T \bar{\mathbf{u}}^\top(t)\mathbf{R}(t)\bar{\mathbf{u}}(t)\mathrm{d}t = \|K_1(\cdot, T)\mathbf{p}_T\|_K^2 = \mathbf{p}_T^\top K_1(T, T)\mathbf{p}_T.$$

---

[9]The property of having a finite number of evaluations is precisely what distinguishes the unconstrained controllability problem (16) or the SOC-constrained problem $(\mathcal{P}_{\delta,\text{fin}})$ from the original state-constrained problem $(\mathcal{P}_0)$ which has an infinite number of affine constraints.

Assume that $g(\cdot) \in \mathcal{C}^1(\mathbb{R}^N, \mathbb{R})$ and that it is convex. Applying the first-order optimality condition, we conclude that

$$\mathbf{0} = \nabla \left(\mathbf{p} \mapsto g(K_1(T,T)\mathbf{p}) + \frac{1}{2}\mathbf{p}^\top K_1(T,T)\mathbf{p} \right)(\mathbf{p}_T) = K_1(T,T)(\nabla g(K_1(T,T)\mathbf{p}_T) + \mathbf{p}_T). \quad (17)$$

So it is sufficient to take $\mathbf{p}_T = -\nabla g(K_1(T,T)\mathbf{p}_T) = -\nabla g(\bar{\mathbf{x}}(T))$, i.e. to have the transversality condition satisfied. However this formula says more than that, as it covers the problem of degeneracies of the "controllability Gramian" $K_1(T,T)$ and gives an explicit equation (17) to be satisfied by $\mathbf{p}_T$. Notice that we do not consider any adjoint equation, only adjoint vectors that are not explicitly propagated. In our framework, the Hamiltonian is implicit.

3.2 Case $\mathbf{Q} \not\equiv \mathbf{0}$

For the case with $\mathbf{Q} \not\equiv \mathbf{0}$, we have a more intricate formula. The reproducing property for K, in which the term $\mathbf{U}_t(\cdot)$ will be explicitly specified below[10], writes as follows, for all $t \in [0,T]$, $\mathbf{p} \in \mathbb{R}^N$, $\mathbf{x}(\cdot) \in \mathcal{S}$,

$$\mathbf{p}^\top \mathbf{x}(t) = (K(0,t)\mathbf{p})^\top \mathbf{x}(0) + \int_0^T (K(s,t)\mathbf{p})^\top \mathbf{Q}(s)\mathbf{x}(s)\mathrm{d}s + \int_0^T (\mathbf{U}_t(s)\mathbf{p}))^\top \mathbf{R}(s)\mathbf{u}(s)\mathrm{d}s. \quad (18)$$

By the Hermitian symmetry of K and the variation of constants formula (8), we can rewrite (18) as, for all $t \in [0,T]$, $\mathbf{x}(\cdot) \in \mathcal{S}$,

$$\mathbf{x}(t) = K(t,0)\mathbf{x}(0) + \int_0^T K(t,s)\mathbf{Q}(s)\left(\mathbf{\Phi_A}(s,0)\mathbf{x}(0) + \int_0^s \mathbf{\Phi_A}(s,\tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\mathrm{d}\tau\right)\mathrm{d}s$$
$$+ \int_0^T \mathbf{U}_t(s)^\top \mathbf{R}(s)\mathbf{u}(s)\mathrm{d}s.$$

After some regrouping of terms, a change of integration bounds and identification with (8), we get the integral equations:

$$\mathbf{\Phi_A}(t,0) = K(t,0) + \int_0^T K(t,s)\mathbf{Q}(s)\mathbf{\Phi_A}(s,0)\mathrm{d}s,$$

$$\forall s \leq t, \; \mathbf{\Phi_A}(t,s)\mathbf{B}(s) = \mathbf{U}_t(s)^\top \mathbf{R}(s) + \int_s^T K(t,\tau)\mathbf{Q}(\tau)\mathbf{\Phi_A}(\tau,s)\mathbf{B}(s)\mathrm{d}\tau$$

$$\forall s > t, \; \mathbf{0} = \mathbf{U}_t(s)^\top \mathbf{R}(s) + \int_s^T K(t,\tau)\mathbf{Q}(\tau)\mathbf{\Phi_A}(\tau,s)\mathbf{B}(s)\mathrm{d}\tau,$$

which can be summarized as

$$K(t,0) = \mathbf{\Phi_A}(t,0) - \tilde{K}(t,0) \text{ with } \tilde{K}(t,s) := \int_s^T K(t,\tau)\mathbf{Q}(\tau)\mathbf{\Phi_A}(\tau,s)\mathrm{d}\tau,$$

$$\mathbf{U}_t(s)^\top \mathbf{R}(s) = \begin{cases} (\mathbf{\Phi_A}(t,s) - \tilde{K}(t,s))\mathbf{B}(s) & \forall s \leq t, \\ -\tilde{K}(t,s)\mathbf{B}(s) & \forall s > t. \end{cases} \quad (19)$$

Although not as explicit as formula (15) from the case $\mathbf{Q} \equiv \mathbf{0}$, this integral expression for $\mathbf{Q} \not\equiv \mathbf{0}$ will still prove valuable to investigate the regularity of $K(\cdot, \cdot)$ (Lemma 2 below). To provide further

---

[10]Recall that $\mathbf{U}_t(s) := \mathbf{B}(s)^{\ominus}[\partial_1 K(s,t) - \mathbf{A}(s)K(s,t)]$.

insight on this expression, again for fixed t, let us introduce formally an adjoint equation for a variable $\mathbf{\Pi}(s,t) \in \mathbb{R}^{N,N}$,

$$\partial_1 \mathbf{\Pi}(s,t) = -\mathbf{A}(s)^\top \mathbf{\Pi}(s,t) + \mathbf{Q}(s)K(s,t) \qquad \mathbf{\Pi}(T,t) = \mathrm{Id}_N. \tag{20}$$

Again, applying the variation of constants formula to $\mathbf{\Pi}(s,t)$, taking the transpose and owing to the symmetries of $K(\cdot,\cdot)$ and $\mathbf{\Phi}(\cdot,\cdot)$, we derive that

$$\mathbf{\Pi}(s,t) = \mathbf{\Phi}_{(-\mathbf{A}^\top)}(s,T)\mathbf{\Pi}(T,t) + \int_T^s \mathbf{\Phi}_{(-\mathbf{A}^\top)}(s,\tau)\mathbf{Q}(\tau)K(\tau,t)\mathrm{d}\tau$$

$$\mathbf{\Pi}(s,t)^\top = \mathbf{\Phi}_\mathbf{A}(T,s) - \int_s^T K(t,\tau)\mathbf{Q}(\tau)\mathbf{\Phi}_\mathbf{A}(\tau,s)\mathrm{d}\tau = \mathbf{\Phi}_\mathbf{A}(T,s) - \tilde{K}(t,s).$$

Since $\partial_1 K(s,t) = \mathbf{A}(s)K(s,t) + \mathbf{B}(s)\mathbf{U}_t(s)$, by (19), for any given time t,

$$\begin{aligned}
\partial_1 K(s,t) &= \mathbf{A}(s)K(s,t) + \mathbf{B}(s)\mathbf{R}(s)^{-1}\mathbf{B}(s)^\top \begin{cases} \mathbf{\Pi}(s,t) - \mathbf{\Phi}_\mathbf{A}(T,s)^\top + \mathbf{\Phi}_\mathbf{A}(t,s)^\top \ \forall s \leq t, \\ \mathbf{\Pi}(s,t) - \mathbf{\Phi}_\mathbf{A}(T,s)^\top \quad \forall s > t. \end{cases} \\
K(0,t) &= \mathbf{\Pi}(0,t) + \mathbf{\Phi}_\mathbf{A}(t,0)^\top - \mathbf{\Phi}_\mathbf{A}(T,0)^\top.
\end{aligned} \tag{21}$$

The difference of behavior between the two cases $\mathbf{Q} \equiv \mathbf{0}$ and $\mathbf{Q} \not\equiv \mathbf{0}$ is classical in optimal control. While the control equation runs forward in time, the adjoint equation runs backward. For $\mathbf{Q} \equiv \mathbf{0}$, the adjoint equation can be solved independently from K, which is why $\mathbf{\Pi}_t$ was not introduced. For $\mathbf{Q} \not\equiv \mathbf{0}$, we have two coupled differential equations (20)-(21) over $K(\cdot,t)$ and $\mathbf{\Pi}_t(\cdot)$. This system does not enjoy an explicit expression, however its solutions can still be computed as a two-point boundary value problem. For quadratic terminal costs, see Aubin-Frankowski (2020), where the connections between computing the kernel and solving the Hamiltonian system or its Riccati equation counterpart are stressed.

4 Theoretical approximation guarantees

In this section, we show that the SOC-constrained problem $(\mathcal{P}_{\delta,\mathrm{fin}})$ is a tightening of the original problem $(\mathcal{P}_0)$. We also provide bounds on the $\|\cdot\|_K$-distance between the optimal trajectory of $(\mathcal{P}_0)$ and that of $(\mathcal{P}_{\delta,\mathrm{fin}})$. This shows that the SOC-tightening is consistent in a numerical analysis sense, as, for bounded kernels K, convergence in $\|\cdot\|_K$ is stronger than uniform convergence of the states, and also implies convergence of the L^2-norms of the controls. We prove that the kernels $K(\cdot,\cdot)$ identified in Section 3 are indeed C^0-continuous.

We shall manipulate various forms of state constraints. We thus write our generic problem $(\mathcal{P}_*)$, with objective $\mathcal{L}$ and constraints defined through the constraint set $\mathcal{V}_*$, as follows

$$\begin{aligned}
\bar{\mathbf{x}}^*(\cdot) \in \arg\min_{\mathbf{x}(\cdot) \in \mathcal{V}_*} \quad &\mathcal{L}(\mathbf{x}(\cdot)) := g(\mathbf{x}(T)) + \|\mathbf{x}(\cdot)\|_K^2 \\
\mathrm{s.t.} \quad &\mathbf{x}(0) = \mathbf{x}_0.
\end{aligned} \tag{$\mathcal{P}_*$}$$

Existence and uniqueness of the solution for each $(\mathcal{P}_*)$ will be discussed below. Recall that $\mathcal{S}$ is defined in (1). Let $(t_m)_{m \in [\![1,N_0]\!]} \in [0,T]^{N_0}$ be N_0 time points associated to radii $\delta_m > 0$ chosen so that they form a covering $[0,T] \subset \cup_{m \in [\![1,N_0]\!]}[t_m - \delta_m, t_m + \delta_m]$. The vectors $\mathbf{d}_m(\delta_m,t_m) := (d_i(\delta_m,t_m))_{i \in [\![1,P]\!]} \in \mathbb{R}^P$, $\overrightarrow{\boldsymbol{\eta}}(\delta,t) := (\eta_i(\delta,t))_{i \in [\![1,P]\!]} \in \mathbb{R}^P$ and $\overrightarrow{\boldsymbol{\omega}}(\delta,t) := (\omega_i(\delta,t))_{i \in [\![1,P]\!]} \in \mathbb{R}^P$ are

11

defined component-wise:[11]

$$\eta_i(\delta, t) := \sup_{s \in [t-\delta, t+\delta] \cap [0,T]} \|K(\cdot, t)\mathbf{c}_i(t) - K(\cdot, s)\mathbf{c}_i(s)\|_K, \tag{22}$$

$$\omega_i(\delta, t) := \sup_{s \in [t-\delta, t+\delta] \cap [0,T]} |d_i(t) - d_i(s)|, \tag{23}$$

$$d_i(\delta_m, t_m) := \inf_{s \in [t_m-\delta_m, t_m+\delta_m] \cap [0,T]} d_i(s). \tag{24}$$

For $\vec{\epsilon} \in \mathbb{R}_+^P$, we shall consider the following constraints

$$\mathcal{V}_0 := \{\mathbf{x}(\cdot) \in \mathcal{S} \mid \mathbf{C}(t)\mathbf{x}(t) \le \mathbf{d}(t), \, \forall\, t \in [0,T]\},$$
$$\mathcal{V}_{\delta,\mathrm{fin}} := \{\mathbf{x}(\cdot) \in \mathcal{S} \mid \vec{\boldsymbol{\eta}}(\delta_m, t_m)\|\mathbf{x}(\cdot)\|_K + \mathbf{C}(t_m)\mathbf{x}(t_m) \le \mathbf{d}(\delta_m, t_m), \, \forall\, m \in [\![1, N_0]\!]\},$$
$$\mathcal{V}_{\delta,\mathrm{inf}} := \{\mathbf{x}(\cdot) \in \mathcal{S} \mid \vec{\boldsymbol{\eta}}(\delta, t)\|\mathbf{x}(\cdot)\|_K + \vec{\boldsymbol{\omega}}(\delta, t) + \mathbf{C}(t)\mathbf{x}(t) \le \mathbf{d}(t), \, \forall\, t \in [0,T]\},$$
$$\mathcal{V}_\epsilon := \{\mathbf{x}(\cdot) \in \mathcal{S} \mid \vec{\epsilon} + \mathbf{C}(t)\mathbf{x}(t) \le \mathbf{d}(t), \, \forall\, t \in [0,T]\}.$$

To these closed constraint sets correspond the problems $(\mathcal{P}_0)$, $(\mathcal{P}_{\delta,\mathrm{fin}})$ $(\mathcal{P}_{\delta,\mathrm{inf}})$, and $(\mathcal{P}_\epsilon)$. In particular, $\bar{\mathbf{x}}^0(\cdot)$ denotes the optimal solution of $(\mathcal{P}_0)$. When $\mathbf{C}(\cdot)$ and $\mathbf{d}(\cdot)$ are $\mathcal{C}^0$-continuous, we prove right away that $\eta_i(\cdot, t)$ and $\omega_i(\cdot, t)$ converge uniformly in t to 0 as $\delta \to 0^+$, so that the SOC inequalities defining the set $\mathcal{V}_{\delta,\mathrm{inf}}$ converge to the original affine constraints.

Lemma 2 (Uniform continuity of K). *If $\mathbf{A}(\cdot) \in L^1(0,T)$ and $\mathbf{B}(\cdot) \in L^2(0,T)$, then $K(\cdot, \cdot)$ is uniformly continuous. Assume furthermore that $\mathbf{C}(\cdot)$ and $\mathbf{d}(\cdot)$ are $\mathcal{C}^0$-continuous. Then for all $i \in [\![1, P]\!]$, the increasing functions $\eta_i(\cdot, t)$ and $\omega_i(\cdot, t)$ converge to 0 uniformly w.r.t. t as $\delta \to 0^+$.*

Proof: Since $\mathbf{A}(\cdot) \in L^1(0,T)$, $\mathbf{\Phi_A}(\cdot, \cdot)$ is uniformly continuous. Hence, for $\mathbf{Q} \equiv \mathbf{0}$, through the explicit formulas (11) and (15), we deduce that $K(\cdot, \cdot)$ is uniformly continuous. Consequently, for any $i \in [\![1, P]\!]$, and $s, t \in [0,T]$, recalling that $\mathbb{B}_N$ denotes the closed unit ball of $\mathbb{R}^N$,

$$\|K(\cdot, t)\mathbf{c}_i(t) - K(\cdot, s)\mathbf{c}_i(s)\|_K \le \|(K(\cdot, t) - K(\cdot, s))\mathbf{c}_i(t)\|_K + \|K(\cdot, s)(\mathbf{c}_i(t) - \mathbf{c}_i(s))\|_K$$
$$\le \|\mathbf{c}_i(t)\| \sup_{\mathbf{p} \in \mathbb{B}_N} \|(K(\cdot, t) - K(\cdot, s))\mathbf{p}\|_K + \|\mathbf{c}_i(t) - \mathbf{c}_i(s)\| \sup_{\mathbf{p} \in \mathbb{B}_N} \|\mathbf{p}^\top K(s,s)\mathbf{p}\|^{1/2},$$

which proves the statement for $\eta_i(\cdot, t)$, whereas the result for $\omega_i(\cdot, t)$ stems directly from the uniform continuity of $\mathbf{d}(\cdot)$. Obviously, the components $\eta_i(\cdot, t)$ and $\omega_i(\cdot, t)$ are increasing for any given t.

For $\mathbf{Q} \not\equiv \mathbf{0}$, we do not have explicit formulas such as (11) and (15). Nonetheless, the $\|\cdot\|_K$-norm (4) for $\mathbf{Q} \not\equiv \mathbf{0}$ is stronger than the $\|\cdot\|_K$-norm for $\mathbf{Q} \equiv \mathbf{0}$. Since, for $\mathbf{Q} \equiv \mathbf{0}$, K is uniformly continuous, owing to (Schwartz, 1964, Proposition 24), the topology induced by K over $\mathcal{S}$ is stronger than the topology of uniform convergence over $[0,T]$. Hence the topology induced by K for $\mathbf{Q} \not\equiv \mathbf{0}$ is also stronger. Therefore, using again the result of Schwartz (1964), for $\mathbf{Q} \not\equiv \mathbf{0}$, $K(\cdot, \cdot)$ is continuous w.r.t. each variable and locally bounded.[12] Hence $K(\cdot, \cdot)$ is bounded on the compact set $[0,T] \times [0,T]$. Let us prove the continuity of $t \mapsto K(t,t)$. Since K is bounded, by (19), $\tilde{K}$ is bounded, so

---

[11]The computation of η_i can be performed using that, by the reproducing property, $\|K(\cdot, t)\mathbf{c}_i(t) - K(\cdot, s)\mathbf{c}_i(s)\|_K^2 = \mathbf{c}_i(t)^\top K(t,t)\mathbf{c}_i(t) + \mathbf{c}_i(s)^\top K(s,s)\mathbf{c}_i(s) - 2\mathbf{c}_i(t)^\top K(t,s)\mathbf{c}_i(s)$. We chose to overload the notation of d_i to define the constants $d_i(\delta_m, t_m)$ in order to draw the parallel with the other perturbations of the constraints, $\eta_i(\delta, t)$ and $\omega_i(\delta, t)$.

[12]This allows to derive the continuity of $\eta_i(\cdot, t)$ but does not provide a uniform bound w.r.t to t.

$\mathbf{U}_t(\cdot) \in L^2(0, T)$. Let $\mathbf{p} \in \mathbb{B}_N$ and $t \in [0, T]$, then, by definition of K, $K(\cdot, t)\mathbf{p} \in \mathcal{S}$ is associated to the control $\mathbf{U}_t(\cdot)\mathbf{p}$. Let $\delta > 0$, by the variation of constants formula (8), we have

$$\|(K(t+\delta, t) - K(t, t))\mathbf{p}\| \leq \|\mathbf{\Phi_A}(t+\delta, 0) - \mathbf{\Phi_A}(t, 0)\| \cdot \|K(0, t)\mathbf{p}\| + \int_t^{t+\delta} \|\mathbf{\Phi_A}(t, s)\mathbf{B}(s)\mathbf{U}_t(s)\mathbf{p}\| \mathrm{d}s.$$

Let $\lambda > 0$. With a similar computation when permuting t and $t + \delta$, taking the supremum over $\mathbf{p} \in \mathbb{B}_N$, one can find $\Delta > 0$ such that for any $\delta \in [0, \Delta]$,

$$\max(\|K(t+\delta, t) - K(t, t)\|, \|K(t, t+\delta) - K(t+\delta, t+\delta)\|) \leq \lambda/2.$$

Hence, owing to the Hermitian symmetry of K, for any $\delta \in [0, \Delta]$,

$$\|K(t+\delta, t+\delta) - K(t, t)\| \leq \|K(t+\delta, t) - K(t, t)\| + \|K(t, t+\delta) - K(t+\delta, t+\delta)\| \leq \lambda.$$

This shows that K is indeed continuous on the diagonal. As underlined by Laurent Schwartz, showing the continuity of $t \mapsto K(t, t)$ is enough to conclude. We reproduce briefly his argument (see Schwartz, 1964, p194): for any $\mathbf{p} \in \mathbb{R}^N$, whenever t converges to t_0, $K(\cdot, t)\mathbf{p}$ weakly converges in $\mathcal{S}$ to $K(\cdot, t_0)\mathbf{p}$, however, by continuity on the diagonal, the norm $\mathbf{p}^\top K(t, t)\mathbf{p}$ converges to $\mathbf{p}^\top K(t_0, t_0)\mathbf{p}$, so $K(\cdot, t)\mathbf{p}$ strongly converges in $\mathcal{S}$ and by extension in $\mathcal{C}^0(0, T)$, exactly showing that $K(\cdot, \cdot)$ is continuous, hence uniformly continuous.

∎

Generically, under the minimal assumptions of Lemma 2, one can apply the SOC-scheme to obtain $(\mathcal{P}_{\delta, \mathrm{fin}})$, which is equivalent to a finite dimensional problem owing to the representer theorem (Theorem 1). By Lemma 2, the scheme is coherent since for δ decreasing to zero, the coefficients of the SOC constraints converge uniformly in t to those of the original problem $(\mathcal{P}_0)$. However, ensuring that the solution $\bar{\mathbf{x}}^{\delta, \mathrm{fin}}(\cdot)$ converges to $\bar{\mathbf{x}}^0(\cdot)$ requires a more thorough analysis.

Proposition 1 (Nested sequence). *Let $\delta_{max} := \max_{m \in [\![1, N_0]\!]} \delta_m$. For any $\delta \geq \delta_{max}$, if, for a given $y_0 \geq 0$, $\overrightarrow{\epsilon} \geq \sup_{t \in [0, T]} [\overrightarrow{\eta}(\delta, t)y_0 + \overrightarrow{\omega}(\delta, t)]$, then we have a nested sequence*

$$(\mathcal{V}_\epsilon \cap y_0 \mathbb{B}_K) \subset \mathcal{V}_{\delta, inf} \subset \mathcal{V}_{\delta, fin} \subset \mathcal{V}_0. \tag{25}$$

Proof: The inclusion $\mathcal{V}_\epsilon \cap y_0 \mathbb{B}_K \subset \mathcal{V}_{\delta, \mathrm{inf}}$ stems from the definition of the sets. Since $d_i(\delta_m, t_m) \geq d_i(t_m) - \omega_i(\delta_m, t_m)$, $\mathcal{V}_{\delta, \mathrm{inf}} \subset \mathcal{V}_{\delta, \mathrm{fin}}$. Recall that $[0, T] \subset \cup_{m \in [\![1, N_0]\!]} [t_m - \delta_m, t_m + \delta_m]$. Let $t \in [0, T]$ and $\mathbf{x}(\cdot) \in \mathcal{V}_{\delta, \mathrm{fin}}$. Take $m \in [\![1, N_0]\!]$ such that $t \in [t_m - \delta_m, t_m + \delta_m]$. For any $i \in [\![1, P]\!]$, applying the reproducing property and Cauchy-Schwarz inequality,

$$\mathbf{c}_i(t)^\top \mathbf{x}(t) = \mathbf{c}_i(t_m)^\top \mathbf{x}(t_m) + \langle \mathbf{x}(\cdot), K(\cdot, t)\mathbf{c}_i(t) - K(\cdot, t_m)\mathbf{c}_i(t_m) \rangle_K$$
$$\mathbf{c}_i(t)^\top \mathbf{x}(t) \leq \mathbf{c}_i(t_m)^\top \mathbf{x}(t_m) + \eta_i(\delta_m, t_m)\|\mathbf{x}(\cdot)\|_K \leq d_i(\delta_m, t_m) \leq d_i(t),$$

$d_i(\delta_m, t_m)$ being by definition the infimum of the i-th component $d_i(t)$ of $\mathbf{d}(t)$ on $[t_m - \delta_m, t_m + \delta_m]$. So $\mathbf{x}(\cdot) \in \mathcal{V}_0$, hence $\mathcal{V}_{\delta, \mathrm{fin}} \subset \mathcal{V}_0$.

∎

Proposition 1 states that, on the one hand, enforcing a finite number of SOC constraints with $\overrightarrow{\eta}$ as in (22) is more restrictive than enforcing an infinite number of affine constraints. On the other hand, SOC constraints are less restrictive than shrinking the affine constraints by some $\overrightarrow{\epsilon} > 0$. The nested property (25) is instrumental in our analysis. As a matter of fact, we shall focus

on ϵ-perturbations of affine constraints rather than on SOC constraints to construct a trajectory $\mathbf{x}^\epsilon(\cdot) \in \mathcal{V}_\epsilon$ close to $\bar{\mathbf{x}}^0(\cdot)$. We shall then resort to strong convexity arguments to derive bounds on $\|\bar{\mathbf{x}}^{\delta,\mathrm{fin}}(\cdot) - \bar{\mathbf{x}}^0(\cdot)\|_K$.

We now list the hypotheses used to prove our main result.

(H-gen) $\mathbf{A}(\cdot) \in L^1(0,T)$ and $\mathbf{B}(\cdot) \in L^2(0,T)$, $\mathbf{C}(\cdot)$ and $\mathbf{d}(\cdot)$ are $\mathcal{C}^0$-continuous.

(H-sol) $\mathbf{C}(0)\mathbf{x}_0 < \mathbf{d}(0)$ and there exists $\overrightarrow{\epsilon} > 0$ such that $\mathcal{V}_\epsilon \cap \{\mathbf{x}(\cdot)\,|\,\mathbf{x}(0) = \mathbf{x}_0\} \neq \emptyset$, i.e. there exists a trajectory $\mathbf{x}^\epsilon(\cdot) \in \mathcal{S}$ satisfying strictly the affine constraints, as well as the initial condition, with $\mathbf{x}_0$ interior to the state constraints.

(H-L) There exists $\mu > 0$ such that the objective function $\mathcal{L} : \mathbf{x}(\cdot) \in \mathcal{S} \mapsto g(\mathbf{x}(T)) + \|\mathbf{x}(\cdot)\|_K^2$ is μ-strongly convex. The terminal cost $g(\cdot)$ is continuous over $\mathbb{R}^N$, $\mathbf{Q}(\cdot) \in L^1(0,T)$, and $\mathbf{R}(\cdot) \in L^2(0,T)$. There exists $r > 0$ such that $\mathbf{R}(t) \succcurlyeq r\mathrm{Id}_M$ for all $t \in [0,T]$.

Discussion of the Assumptions: Assumption (H-gen) ensures the $\mathcal{C}^0$-continuity of the kernel K and of the functions η_i and ω_i (Lemma 2). Assumption (H-L) concerns the objective function, whereas (H-sol) ensures that the set of trajectories satisfying the state constraints is non-empty if the latter are shrunk:

- The existence requirement in (H-sol) can be derived from assumptions on the existence of interior viable trajectories. We provide in the Annex an example of such assumptions (Lemma 4) based on inward pointing conditions on the boundary and on regularity assumptions on the constraints and the dynamics. The assumption $\mathbf{C}(0)\mathbf{x}_0 < \mathbf{d}_0$ ensures that $\mathbf{x}_0$ is a suitable initial condition for the $\overrightarrow{\epsilon}$-tightening $\mathcal{V}_\epsilon$.[13]

- The strong convexity requirement in (H-L) is obviously satisfied whenever $g(\cdot)$ is convex.[14] It is required in order to bound the distance on solutions since the problems $(\mathcal{P}_*)$ share the same objective but different constraint sets. The terminal cost $g(\cdot)$ is supposed $\mathcal{C}^0$-continuous, but could be taken merely locally continuous in a neighborhood[15] of $\bar{\mathbf{x}}^0(T)$ and lower bounded over any compact subset of $\mathbb{R}^N$. By Lemma 3, $(\mathcal{P}_0)$ has a unique optimal solution $\bar{\mathbf{x}}^0(\cdot)$.

Lemma 3 (Existence and uniqueness of solutions)**.** *Under Assumptions (H-gen), (H-sol), and (H-L), $\bar{\mathbf{x}}^0(\cdot)$ exists and is unique. The same result holds true for $(\mathcal{P}_{\delta,\mathrm{fin}})$, $(\mathcal{P}_{\delta,\mathrm{inf}})$, and $(\mathcal{P}_\epsilon)$ for any $\delta \in [0,\delta_0]$, where $\delta_0 > 0$ satisfies that $\overrightarrow{\epsilon} \geq \sup_{t \in [0,T]}[\overrightarrow{\boldsymbol{\eta}}(\delta_0,t)\|\mathbf{x}^\epsilon(\cdot)\|_K + \overrightarrow{\boldsymbol{\omega}}(\delta_0,t)]$ for $\overrightarrow{\epsilon}$ as in (H-sol).*

Proof: The existence result is a consequence of Tonelli's direct method, usually stated for lower bounded and lower semi-continuous $g(\cdot)$. We detail the proof since our Assumptions are both slightly different and stronger. Since by Assumption (H-sol), $\mathcal{V}_0 \neq \emptyset$, let $(\mathbf{x}_n(\cdot), \mathbf{u}_n(\cdot))$ be a minimizing sequence of $(\mathcal{P}_0)$ converging to the optimal value $\bar{m}$. As $\mathcal{L}(\cdot)$ is μ-strongly convex, $\frac{\mathcal{L}(\mathbf{x}(\cdot))}{\|\mathbf{x}(\cdot)\|_K} \to +\infty$ as $\|\mathbf{x}(\cdot)\|_K \to +\infty$. Hence $\bar{m}$ is finite, and $(\mathbf{x}_n(\cdot))_n$ is a subset of a ball $M_0\mathbb{B}_K \subset \mathcal{S}$, for some $M_0 > 0$. Since $\mathbf{x}(\cdot) \in \mathcal{S} \mapsto \mathbf{x}(T)$ is continuous, $\{\mathbf{x}(T)\,|\,\mathbf{x}(\cdot) \in M_0\mathbb{B}_K\}$ is also bounded. By continuity of $g(\cdot)$, let $m_g := -\inf_{\mathbf{x}(\cdot) \in M_0\mathbb{B}_K} g(\mathbf{x}(T)) < +\infty$. Consequently, for n large enough,

$$r\|\mathbf{u}_n(\cdot)\|_{L^2(0,T)}^2 \leq \|\mathbf{x}_n(\cdot)\|_K^2 \leq \bar{m} + 1 - g(\mathbf{x}_n(T)) \leq \bar{m} + 1 + m_g,$$

---

[13]Since the SOC tightening lies in-between the $\overrightarrow{\epsilon}$-tightening and the original constraints (Proposition 1), it cannot be guaranteed that an initial condition on the border of the constraints would be suitable for the SOC tightening.

[14]More generally, g could be μ_0-semiconvex (i.e. $g(\cdot) + \frac{\mu_0}{2}\|\cdot\|^2$ is convex with $2 > \mu_0 \sup_{\mathbf{p} \in \mathbb{B}_N} \|\mathbf{p}^\top K(T,T)\mathbf{p}\|^{1/2}$.

[15]This neighborhood is considered with respect to the relative topology of the terminal constraint set $\{\mathbf{x} \in \mathbb{R}^N\,|\,\mathbf{C}(T)\mathbf{x} \leq \mathbf{d}(T)\}$.

so $(\mathbf{u}_n(\cdot))_n$ is bounded in L^2, and we can take a subsequence $(\mathbf{u}_{n_i}(\cdot))_i$ weakly converging to some $\mathbf{u}(\cdot)$. Let $s,t \in [0,T]$. By the variation of constants formula (8), we had derived (9). Since $\mathbf{A}(\cdot) \in L^1(0,T)$ and $\mathbf{B}(\cdot) \in L^2(0,T)$, and $\mathbf{x}_n(s)$ is uniformly bounded in n and s by the reproducing property and continuity of K (Lemma 2), $(\mathbf{x}_{n_i}(\cdot))_i$ is equicontinuous. By Ascoli's theorem, we thus have a subsequence $(\mathbf{x}_{n_{i_j}}(\cdot))_j$ uniformly converging to some $\mathbf{x}(\cdot)$ satisfying (8) for $\mathbf{u}(\cdot)$, thus $\mathbf{x}(\cdot) \in \mathcal{S}$. By continuity of $g(\cdot)$, $\mathcal{L}(\mathbf{x}(\cdot)) = \bar{m}$. Since $\mathcal{L}(\cdot)$ is strongly convex, the optimal trajectory is unique and belongs to the closed set $\mathcal{V}_0$. To conclude, replace $\mathcal{V}_0$ with $\mathcal{V}_{\delta,\mathrm{fin}}$ (resp. $\mathcal{V}_{\delta,\mathrm{inf}}$, and $\mathcal{V}_\epsilon$), the inequality satisfied by δ_0 shows that $\mathbf{x}^\epsilon(\cdot) \in \mathcal{V}_{\delta,\mathrm{fin}}$, consequently the constraint sets are non-empty. The same arguments as above yield the result.

∎

Theorem 2 (Main result - Approximation by SOC constraints). *Under Assumptions (H-gen), (H-sol), and (H-L), for any $\lambda > 0$, there exists $\bar{\delta} > 0$ such that for all $N_0 > 0$ and $(\delta_m)_{m \in [\![1,N_0]\!]}$, with $[0,T] \subset \cup_{m \in [\![1,N_0]\!]}[t_m - \delta_m, t_m + \delta_m]$ satisfying $\bar{\delta} \geq \max_{m \in [\![1,N_0]\!]} \delta_m$, we have*

$$\frac{1}{\gamma_K} \cdot \sup_{t \in [0,T]} \|\bar{\mathbf{x}}^{\delta,fin}(t) - \bar{\mathbf{x}}^0(t)\| \leq \|\bar{\mathbf{x}}^{\delta,fin}(\cdot) - \bar{\mathbf{x}}^0(\cdot)\|_K \leq \lambda \tag{26}$$

with $\gamma_K := \sup_{t \in [0,T],\, \mathbf{p} \in \mathbb{B}_N} \sqrt{\mathbf{p}^\top K(t,t)\mathbf{p}}$.

Proof: Let $\lambda > 0$. Consider any $\tilde{\lambda} > 0$ such that

$$2\tilde{\lambda} + \tilde{\lambda}(\tilde{\lambda} + 2\|\bar{\mathbf{x}}^0(\cdot)\|_{L^\infty(0,T)})\|\mathbf{Q}(\cdot)\|_{L^1(0,T)} \leq \lambda. \tag{27}$$

By Assumption (H-sol), pick $\mathbf{x}^\epsilon(\cdot) \in \mathcal{V}_\epsilon$ such that $\mathbf{x}^\epsilon(0) = \bar{\mathbf{x}}^0(0)$. Denote by $\mathbf{u}^\epsilon(\cdot)$ the associated control. Take $\alpha > 0$ small enough such that $\mathbf{x}^{\alpha\epsilon}(\cdot) := \alpha\mathbf{x}^\epsilon(\cdot) + (1-\alpha)\bar{\mathbf{x}}^0(\cdot) \in \mathcal{S}$ and $\mathbf{u}^{\alpha\epsilon}(\cdot) := \alpha\mathbf{u}^\epsilon(\cdot) + (1-\alpha)\bar{\mathbf{u}}^0(\cdot)$ satisfy

$$\|\bar{\mathbf{x}}^0(\cdot) - \mathbf{x}^{\alpha\epsilon}(\cdot)\|_{L^\infty(0,T)} = \alpha\|\bar{\mathbf{x}}^0(\cdot) - \mathbf{x}^\epsilon(\cdot)\|_{L^\infty(0,T)} \leq \tilde{\lambda}$$
$$|\|\mathbf{R}(\cdot)^{1/2}\bar{\mathbf{u}}(\cdot)\|^2_{L^2(0,T)} - \|\mathbf{R}(\cdot)^{1/2}\mathbf{u}^{\alpha\epsilon}(\cdot)\|^2_{L^2(0,T)}| \leq \tilde{\lambda}$$

and, by continuity of $g(\cdot)$, $|g(\bar{\mathbf{x}}^0(T)) - g(\mathbf{x}^{\alpha\epsilon}(T))| \leq \tilde{\lambda}$. Consequently $\mathbf{x}^{\alpha\epsilon}(0) = \bar{\mathbf{x}}^0(0)$ and for all $t \in [0,T]$, $\mathbf{C}(t)\mathbf{x}^{\alpha\epsilon}(t) \leq \alpha(\mathbf{d}(t) - \overrightarrow{\boldsymbol{\epsilon}}) + (1-\alpha)\mathbf{d}(t) = \mathbf{d}(t) - \alpha\overrightarrow{\boldsymbol{\epsilon}}$, so $\mathbf{x}^{\alpha\epsilon}(\cdot) \in \mathcal{V}_{\alpha\epsilon}$. Hence

$$\mathcal{L}(\mathbf{x}^{\alpha\epsilon}(\cdot)) - \mathcal{L}(\bar{\mathbf{x}}^0(\cdot)) \leq |g(\bar{\mathbf{x}}^0(T)) - g(\mathbf{x}^{\alpha\epsilon}(T))| + \left|\|\bar{\mathbf{x}}^0(\cdot)\|^2_K - \|\mathbf{x}^{\alpha\epsilon}(\cdot)\|^2_K\right|$$

$$\leq \tilde{\lambda} + \int_0^T \left|(\bar{\mathbf{x}}^0(t) - \mathbf{x}^{\alpha\epsilon}(t))^\top \mathbf{Q}(t)(\bar{\mathbf{x}}^0(t) + \mathbf{x}^{\alpha\epsilon}(t))\right|\mathrm{d}t + \left|\|\mathbf{R}(\cdot)^{1/2}\bar{\mathbf{u}}(\cdot)\|^2_{L^2(0,T)} - \|\mathbf{R}(\cdot)^{1/2}\mathbf{u}^{\alpha\epsilon}(\cdot)\|^2_{L^2(0,T)}\right|$$

$$\leq 2\tilde{\lambda} + \tilde{\lambda}(\tilde{\lambda} + 2\|\bar{\mathbf{x}}^0(\cdot)\|_{L^\infty(0,T)})\|\mathbf{Q}(\cdot)\|_{L^1(0,T)} \overset{(27)}{\leq} \lambda.$$

Let $\delta_0 > 0$ such that $\alpha\overrightarrow{\boldsymbol{\epsilon}} \geq \sup_{t \in [0,T]}[\overrightarrow{\boldsymbol{\eta}}(\delta_0, t)\|\mathbf{x}^{\alpha\epsilon}(\cdot)\|_K + \overrightarrow{\boldsymbol{\omega}}(\delta_0, t)]$. Then $\mathbf{x}^{\alpha\epsilon}(\cdot) \in \mathcal{V}_{\delta_0,\mathrm{inf}} \subset \mathcal{V}_{\delta_0,\mathrm{fin}}$, the sets thus being non-empty. Notice that, for any $\delta \in [0, \delta_0]$, as $\bar{\mathbf{x}}^*(\cdot)$ is optimal for $(\mathcal{P}_*)$, from the nested property (25), we derive that

$$\mathcal{L}(\bar{\mathbf{x}}^{\delta,\mathrm{fin}}(\cdot)) \leq \mathcal{L}(\bar{\mathbf{x}}^{\delta,\mathrm{inf}}(\cdot)) \leq \mathcal{L}(\bar{\mathbf{x}}^{\delta_0,\mathrm{inf}}(\cdot)).$$

As $\mathcal{L}(\cdot)$ is μ-convex, $\mathcal{L}^{-1}(]-\infty, \mathcal{L}(\bar{\mathbf{x}}^{\delta_0,\mathrm{inf}}(\cdot))])$ is a bounded set, contained in a ball $M_0\mathbb{B}_K$ for some $M_0 > 0$, and containing all the $\{\bar{\mathbf{x}}^{\delta,\mathrm{inf}}(\cdot)\}_{\delta \in [0,\delta_0]}$. Since $\mathcal{S}$ is a vRKHS, $\mathbf{x}(\cdot) \in \mathcal{S} \mapsto \mathbf{x}(T)$ is

15

continuous. So $\{\mathbf{x}(T) \,|\, \mathbf{x}(\cdot) \in M_0 \mathbb{B}_K\}$ is also bounded. Hence, for any $\delta \in [0, \delta_0]$,

$$g(\bar{\mathbf{x}}^{\delta,\inf}(T)) + \|\bar{\mathbf{x}}^{\delta,\inf}(\cdot)\|_K^2 \leq \mathcal{L}(\bar{\mathbf{x}}^{\delta_0,\inf}(\cdot)) \leq |g(\bar{\mathbf{x}}^{\delta_0,\inf}(T))| + \|\bar{\mathbf{x}}^{\delta_0,\inf}(\cdot)\|_K^2$$

$$\|\bar{\mathbf{x}}^{\delta,\inf}(\cdot)\|_K \leq \|\bar{\mathbf{x}}^{\delta_0,\inf}(\cdot)\|_K + \sqrt{|g(\bar{\mathbf{x}}^{\delta_0,\inf}(T))|} + |\inf_{\mathbf{x}(\cdot)\in M_0\mathbb{B}_K} g(\mathbf{x}(T))|^{\frac{1}{2}} =: y_0.$$

As $\|\bar{\mathbf{x}}^{\delta_0,\inf}(\cdot)\|_K \leq y_0$, $\mathbf{x}^{\alpha\epsilon}(\cdot)$ and $\bar{\mathbf{x}}^{\delta_0,\inf}(\cdot)$ are both admissible for the following problem,

$$\min_{\substack{\mathbf{x}(\cdot)\in\mathcal{V}_{\delta_0,\inf} \\ \|\mathbf{x}(\cdot)\|_K \leq y_0 + \|\bar{\mathbf{x}}^{\alpha\epsilon}(\cdot)\|_K}} \mathcal{L}(\mathbf{x}(\cdot))$$

with $\bar{\mathbf{x}}^{\delta_0,\inf}(\cdot)$ being optimal by definition, hence we have $\mathcal{L}(\bar{\mathbf{x}}^{\delta_0,\inf}(\cdot)) \leq \mathcal{L}(\mathbf{x}^{\alpha\epsilon}(\cdot))$. To conclude, let $\bar{\delta} \in]0, \delta_0]$ such that $\alpha\overrightarrow{\boldsymbol{\epsilon}} \geq \sup_{t\in[0,T]}[\overrightarrow{\boldsymbol{\eta}}(\bar{\delta},t)y_0 + \overrightarrow{\boldsymbol{\omega}}(\bar{\delta},t)]$. Then, for any $\delta \in]0, \bar{\delta}]$, by strong convexity of $\mathcal{L}(\cdot)$, $\bar{\mathbf{x}}^0(\cdot)$ being optimal for $(\mathcal{P}_0)$,

$$\frac{\mu}{2}\|\bar{\mathbf{x}}^{\delta,\mathrm{fin}}(\cdot) - \bar{\mathbf{x}}^0(\cdot)\|_K^2 \leq \mathcal{L}(\bar{\mathbf{x}}^{\delta,\mathrm{fin}}(\cdot)) - \mathcal{L}(\bar{\mathbf{x}}^0(\cdot)) \leq \mathcal{L}(\mathbf{x}^{\alpha\epsilon}(\cdot)) - \mathcal{L}(\bar{\mathbf{x}}^0) \leq \lambda.$$

Replacing λ by $\sqrt{2\lambda/\mu}$, we deduced that $\|\bar{\mathbf{x}}^{\delta,\mathrm{fin}}(\cdot) - \bar{\mathbf{x}}^0(\cdot)\|_K \leq \lambda$. By Cauchy-Schwarz inequality, for any $t \in [0,T]$, $\|\bar{\mathbf{x}}^{\delta,\mathrm{fin}}(t) - \bar{\mathbf{x}}^0(t)\| \leq \|\bar{\mathbf{x}}^{\delta,\mathrm{fin}}(\cdot) - \bar{\mathbf{x}}^0(\cdot)\|_K \sup_{\mathbf{p}\in\mathbb{B}_N} \sqrt{\mathbf{p}^\top K(t,t)\mathbf{p}}$. By definition of γ_K, taking the supremum over $[0,T]$, we derive the remaining inequality.

$\blacksquare$

Theorem 2 states that, when the discretization steps $(\delta_{\mathrm{m}})_m$ go to zero, then the solution $\bar{\mathbf{x}}^{\delta,\mathrm{fin}}(\cdot)$ of the SOC-approximation $(\mathcal{P}_{\delta,\mathrm{fin}})$ can be made arbitrarily close to the solution $\bar{\mathbf{x}}^0(\cdot)$ of the original problem $(\mathcal{P}_0)$, uniqueness being ensured by Assumption (H-L). Concerning the stability of solutions under shrinking perturbation of the state constraints, the result of Theorem 2 actually also holds when replacing $\bar{\mathbf{x}}^{\delta,\mathrm{fin}}(\cdot)$ by $\bar{\mathbf{x}}^\epsilon(\cdot)$, showing that, when $\overrightarrow{\boldsymbol{\epsilon}}$ goes to zero, $\bar{\mathbf{x}}^\epsilon(\cdot)$ converges to $\bar{\mathbf{x}}^0(\cdot)$.

5 Finite-dimensional implementation and numerical example

In this section, we express the finite-dimensional equivalent of problem $(\mathcal{P}_{\delta,\mathrm{fin}})$ owing to the representer theorem (Theorem 1) and discuss its implementation on a numerical example. In general, the SOC transformation requires only the minimal hypotheses of Lemma 2 to be conceptually grounded. Theorem 2 essentially states theoretical guarantees of convergence for small discretization steps. For numerical applications, the problem can thus be extended to incorporate costs or equality constraints at any finite number of intermediate times, as in path planning problems. This situation was already met when discussing the controllability Gramian (16). We consequently enrich $(\mathcal{P}_{\delta,\mathrm{fin}})$ to optimization problems considered in Theorem 1, of the following form, with $\|\mathbf{x}(\cdot)\|_K^2 = \|\mathbf{x}(0)\|^2 + \|\mathbf{x}(\cdot)\|_{K_1}^2$,

$$\min_{\mathbf{x}(\cdot)\in\mathcal{S}} \; L\left(\mathbf{c}_{0,1}^\top \mathbf{x}(t_{0,1}), \ldots, \mathbf{c}_{0,N_0}^\top \mathbf{x}(t_{0,N_0})\right) + \|\mathbf{x}(\cdot)\|_K^2$$

$$\text{s.t. } \eta_i(\delta_{i,m}, t_{i,m})\|\mathbf{x}(\cdot)\|_K + \mathbf{c}_i(t_{i,m})^\top \mathbf{x}(t_{i,m}) \leq d_i(\delta_{i,m}, t_{i,m}), \; \forall m \in [\![1, N_i]\!], \forall i \in [\![1, P]\!],$$
$$(\mathcal{P}_{\mathrm{SOC}})$$

for $\{\mathbf{c}_{0,m}\}_{m\in[\![1,N_0]\!]} \subset \mathbb{R}^N$. The differences between $(\mathcal{P}_{\delta,\mathrm{fin}})$ and $(\mathcal{P}_{\mathrm{SOC}})$ are that $g(\cdot)$, which depended only on the terminal point, is now replaced with a loss $L : \mathbb{R}^{N_0} \to \mathbb{R} \cup \{+\infty\}$, defined on a finite number $(t_{0,m})_{m\in[\![1,N_0]\!]}$ of intermediate points (taken with repetition), and that different

discretization grids $(\delta_{i,m}, t_{i,m})_{i,m\in[\![1,N_i]\!]}$ are used for each constraint $i \in [\![1, P]\!]$. As no structural assumptions are imposed on L, it may incorporate indicator functions to account for the initial condition or for rendezvous points.[16] In Section 4, the SOC constraints were introduced to turn an infinite number of affine constraints over $[0, T]$ into a finite number of SOC constraints. The logic is therefore to separate the discrete pointwise requirements (which go to L) from the constraints that should hold on $[0, T]$ (which are approximated by SOC constraints). Since the constraints on $[0, T]$ may apply to different components of the state, we may consider different grid steps for each i.

In general, adding the SOC terms leads to more conservative solutions w.r.t. the one with affine constraints. The SOC constraints formulation only requires estimating $\eta_i(\delta_{i,m}, t_{i,m})$ and $-d_i(\delta_{i,m}, t_{i,m})$ defined in (22) and (24). Both quantities should in principle be overestimated for the guarantees to hold. However, since the tightening results from a worst-case Cauchy-Schwarz scenario as discussed in (7), in practice, underestimation does not affect the numerical results. Moreover the definition (22) of η_i is not the only possible formulation.[17] Its choice results from geometrical considerations on coverings[18] of compact sets in infinite-dimensional Hilbert spaces (see Section 3 of Aubin-Frankowski and Szabó (2020a) for more details). Even for other values of η_i than (22), considering SOC terms in the constraints proves to be beneficial in terms of local satisfaction of the constraints on a neighborhood of $t_{i,m}$. Besides, the discretization grids considered here are 'static' in the sense that they are fixed before solving $(\mathcal{P}_{\mathrm{SOC}})$. Extensions to 'dynamic' grids, refined depending on the optimization steps, can be found in Section 5 of Aubin-Frankowski and Szabó (2020a).

By Theorem 1, for $\mathbf{x}(\cdot) = \sum_{j=0}^{P}\sum_{m=1}^{N_j} K(\cdot, t_{j,m})\mathbf{p}_{j,m}$ and $z = \|\mathbf{x}(\cdot)\|_K$,[19] $(\mathcal{P}_{\mathrm{SOC}})$ is equivalent to

$$
\min_{\substack{z\in\mathbb{R}_+,\\ j\in[\![0,P]\!],\, m\in[\![1,N_j]\!],\\ \mathbf{p}_{j,m}\in\mathbb{R}^N,\, \alpha_{j,m}\in\mathbb{R}}} \quad L\left(\left(\left(\sum_{j=0}^{P}\sum_{m=1}^{N_j} K(t_{0,n}, t_{j,m})\mathbf{p}_{j,m}\right)_{n\in[\![1,N_0]\!]}\right)\right) + z^2
$$

$$
\text{s.t.} \quad z^2 = \sum_{i=0}^{P}\sum_{n=1}^{N_i}\sum_{j=0}^{P}\sum_{m=1}^{N_j} \mathbf{p}_{i,n}^\top K(t_{i,n}, t_{j,m})\mathbf{p}_{j,m},
$$

$$
\mathbf{p}_{j,m} = \alpha_{j,m}\mathbf{c}_j(t_m), \quad \forall\, m \in [\![1, N_j]\!], \forall\, j \in [\![1, P]\!],
$$

$$
\eta_i(\delta_{i,m}, t_{i,m})z + \sum_{j=0}^{P}\sum_{m=1}^{N_j} \mathbf{c}_i(t_{i,m})^\top K(t_{i,m}, t_{j,m})\mathbf{p}_{j,m} \quad \forall\, m \in [\![1, N_i]\!], \forall\, i \in [\![1, P]\!].
$$
$$
\leq d_i(\delta_{i,m}, t_{i,m}),
$$

For quadratic L with indicator functions, after incorporating the quadratic objective into the constraints through a change of variables, $(\mathcal{P}_{\mathrm{SOC}})$ writes as a second-order cone program (SOCP). Hence this problem can be straightforwardly implemented in convex solvers or modeling frameworks such as CVXGEN (Mattingley and Boyd, 2012). SOCP is slightly more expensive computationally than

---

[16]To write $(\mathcal{P}_{\delta,\mathrm{fin}})$ as $(\mathcal{P}_{\mathrm{SOC}})$, take $N_0 = 2N$, $t_{0,1} = \cdots = t_{0,N} = 0$, $t_{0,N+1} = \cdots = t_{0,N_0} = T$. Denoting by $\chi_{\mathbf{x}_0}(\cdot)$ the indicator function of $\mathbf{x}_0$, set $L\left(\mathbf{c}_{0,1}^\top\mathbf{x}(t_{0,1}), \ldots, \mathbf{c}_{0,N_0}^\top\mathbf{x}(t_{0,N_0})\right) := \chi_{\mathbf{x}_0}(\mathbf{x}(t_{0,1})) + g(\mathbf{x}(t_{0,N_0}))$.

[17]Using that $\mathbf{x}(\cdot) = \mathbf{z}_0(\cdot) + \mathbf{z}_1(\cdot) \in \mathcal{S}_0 \oplus \mathcal{S}_u$ one could consider two η-terms instead of one to derive a less conservative tightening. In the same spirit $\mathbf{d}(t)$ could be projected onto $\mathcal{S}$, and the projection incorporated in the scalar products of the left hand-side. For $\mathbf{Q}(\cdot) \not\equiv \mathbf{0}$, $\|\mathbf{x}(0)\|^2$ could be replaced in (4) by an $\mathcal{S}_0$-norm $\|\mathbf{S}_0\mathbf{x}(0)\|^2$ with a surjective $\mathbf{S}_0 \in \mathbb{R}^{N_0,N}$ where $N_0 = \dim(\mathcal{S}_0)$. This would not change the formulation, but lead to a "tighter" norm $\|\mathbf{x}(\cdot)\|_K$.

[18]Definition (22) corresponds to a covering made of balls in $\mathcal{S}$.

[19]The reproducing property, $\langle K(\cdot, t_1)\mathbf{p}_1, K(\cdot, t_2)\mathbf{p}_2\rangle_K = \mathbf{p}_2^\top K(t_2, t_1)\mathbf{p}_1$, applied to $\mathbf{x}(\cdot)$ allows to explicit z.

the quadratic programs (QP) classically derived for LQR (Kojima and Morari, 2004). Both QP and SOCP have polynomial computation times, however the exponent is problem-dependent, so achieving a theoretical comparison is challenging. From a numerical viewpoint, it seems to be also very much framework-dependent. For instance, when using CVXGEN, the baseline cost of calling CVX blurs the difference in computation times between QP and SOCP. As the more time points, the more coefficients, it is beneficial to define the grids as subsets of a 'master grid'. Furthermore when computing η_i or $K(s,t)$, one has to approximate the supremum in (22) or the integral in (15), for instance through sampling. For time-invariant dynamics and $\mathbf{Q}(\cdot) \equiv \mathbf{0}$, one can use the method proposed in Van Loan (1978) to quickly compute $K_1(s,t)$ in (15). If sought for, the analysis of the approximation error would correspond to that of the stability of the solution when perturbing the constants and matrices appearing in the SOCP. This leads to a large body of articles (e.g. Bonnans and Ramírez (2005) and quoting articles) which exceeds the purposes of this study. If one overestimates the value of η and adds a term $\lambda_{\mathrm{cond}} \sum_{i=0}^{P} \sum_{n=1}^{N_i} \|\mathbf{p}_{i,n}\|^2$ to z^2, then the state constraints are further tightened. This is done in practice to improve the conditioning of the matrices, and discussed along other numerical implementation details of a SOC-constrained kernel regression in Aubin-Frankowski and Szabó (2020b). For a sufficiently small error in K, constraints satisfaction could be guaranteed, even when facing numerical errors, through this further tightening involving λ_{cond}.

To highlight the behavior of the SOC-transformation ($\mathcal{P}_{\mathrm{SOC}}$) of problem ($\mathcal{P}_0$), we consider the problem ($\mathcal{P}_{\mathrm{pend}}$) of a 'linear' pendulum with angle $x(t)$ where we control the derivative $u(t)$ of a forcing term $w(t)$, with state constraints both on $w(t)$ and on $\dot{x}(t)$, the full state being $\mathbf{x} := [x, \dot{x}, w] \in \mathbb{R}^3$,

$$\min_{x(\cdot), u(\cdot)} \quad -\lambda_T \, \dot{x}(T) + \lambda_u \|u(\cdot)\|_{L^2(0,T)}^2 + \|\mathbf{x}(0)\|^2$$

$$\begin{aligned} \text{s.t.} \quad & x(0) = 0.5, \quad \dot{x}(0) = 0, \quad w(0) = 0, \\ & x(T/3) = 0.5, \quad x(T) = 0, \\ & \ddot{x}(t) = -10\,x(t) + w(t), \quad \dot{w}(t) = u(t), \text{ a.e. in } [0,T], \\ & \dot{x}(t) \in [-3, +\infty[, \quad w(t) \in [-10, 10], \, \forall\, t \in [0,T]. \end{aligned} \qquad (\mathcal{P}_{\mathrm{pend}})$$

The objective of ($\mathcal{P}_{\mathrm{pend}}$) can be interpreted as trying to maximize the terminal velocity $\dot{x}(T)$ with an L^2-cost over the control. This example underlines how ($\mathcal{P}_{\mathrm{SOC}}$) effectively allows for more possibilities than ($\mathcal{P}_{\delta,\mathrm{fin}}$). We have both an initial, an intermediate and a final condition on $x(\cdot)$. Since the full state $\mathbf{x} := [x, \dot{x}, w]$ is not fixed at the intermediate time $T/3$, the problem cannot be split into independent problems over two time intervals. We take $\mathbf{R}(\cdot) \equiv \lambda_u \in \mathbb{R}$ in order to have ($\mathcal{P}_{\mathrm{pend}}$) written as ($\mathcal{P}_{\mathrm{SOC}}$). We identify the kernels K_0 and K_1 as in (11) and (15) and use a uniform grid of N_P points for the two affine constraints over $\dot{x}(t)$ and $w(t)$ and turn them into SOC constraints

$$\eta_{\dot{x}}\|\mathbf{x}(\cdot)\|_K - \dot{x}(t_m) \le -3, \quad \eta_w\|\mathbf{x}(\cdot)\|_K + w(t_m) \le 10, \quad \eta_w\|\mathbf{x}(\cdot)\|_K - w(t_m) \le 10, \, \forall\, m \in [\![1, N_P]\!]$$

where $\eta_{\dot{x}}$ and η_w are defined as in (22) for $\mathbf{C} = [0\text{-}1\,0\,;\,0\,0\,1\,;\,0\,0\text{-}1]$. For the experiment of Fig. 1, we take $T = 1$, $\lambda_T = 10^6$ and $\lambda_u = 10^4$. All computations took less than 30 seconds.

We first compare SOC constraints with discretized constraints ($\eta_w = 0$) for a moderate value of $N_P = 200$. Whereas the SOC-constrained optimal trajectory is fairly conservative w.r.t. to the bounds over $w(t)$, the optimal trajectory for discretized constraints suffers from chattering and does not satisfy the w-constraints. This was already hinted at in the speed limit example of the introduction: the control sticks to the w-constraints at the points $(t_m)_m$ but violates them in between. Hence the optimal value of $\dot{x}(T) = 3.18$ is attained only by repeated violation of the
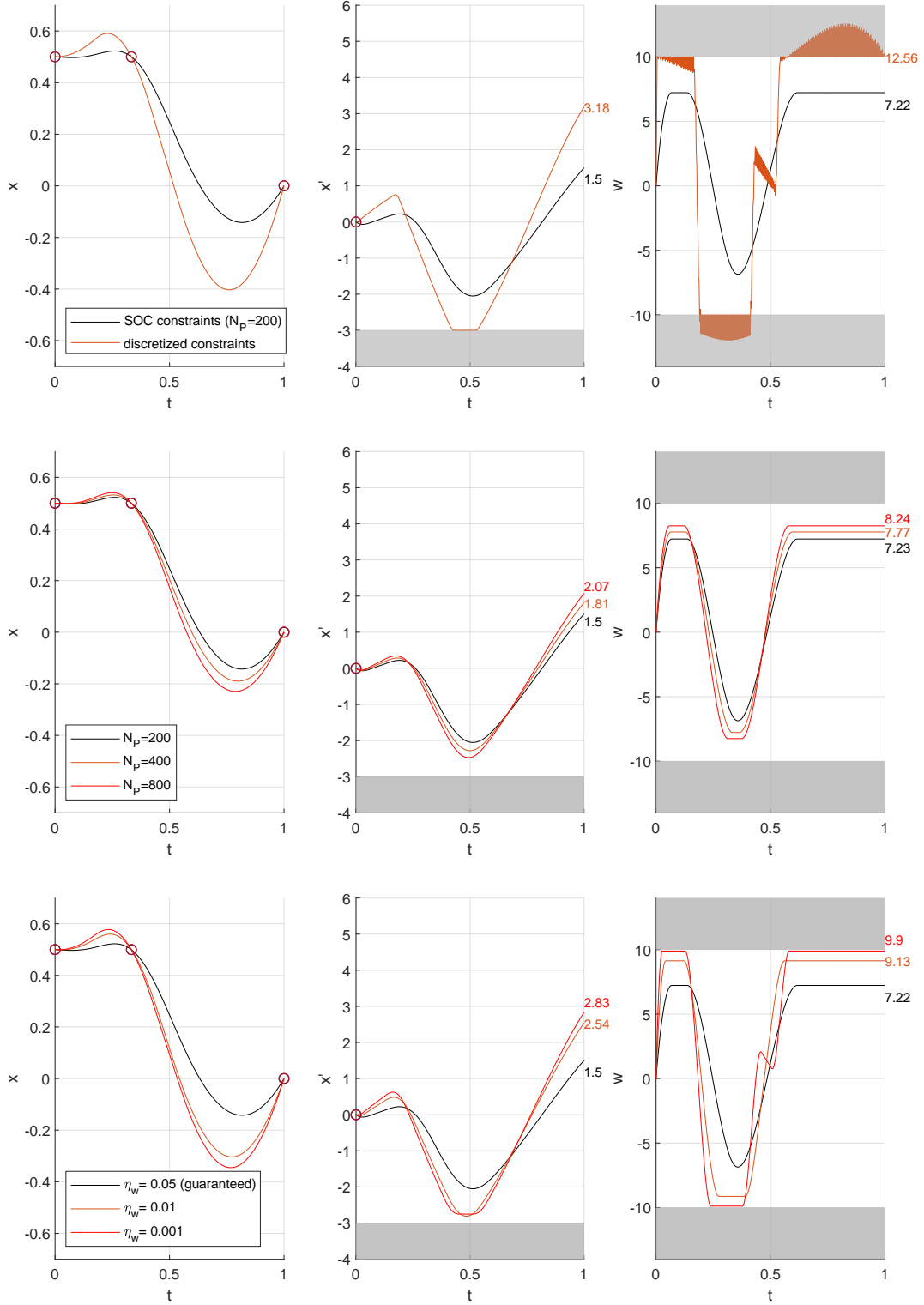
18

Figure 1: Optimal solutions of $(\mathcal{P}_{\text{pend}})$ for varying N_P and η_w. The red circles indicate the equality-constrained points, the grayed areas the constraints over $[0, T]$. We report the values of $\dot{x}(T)$ and of the maximum of $w(\cdot)$. **Top:** Comparison of SOC constraints (guaranteed η_w) versus discretized constraints ($\eta_w = 0$) for $N_P = 200$. **Center:** Comparison of SOC constraints for varying N_P and guaranteed η_w. **Bottom:** Comparison of SOC constraints for varying η_w and $N_P = 200$.

19

w-constraints. We then present the consequences of changing the number of grid points N_P and of lowering η_w w.r.t. to the value of its definition. The parameter $\eta_{\dot{x}}$ is kept fixed to its nominal value as it has little influence on the optimal solutions. We first investigate the effect of changing N_P while keeping the guaranteed value of η_w, defined as in (22). We see that the threshold applied to the w-constraints decreases, albeit slowly. Secondly, for $N_P = 200$, we divide η_w first by a factor 5, then by a factor 10. The w-constraints threshold drastically diminishes and, for $\eta_w = 0.001$, the inflexion at $t = 0.5$ appears, as a result of the constrained arc over $\dot{x}(\cdot)$. In the limit case, we would recover the trajectory with discretized constraints ($\eta_w = 0$).

We conclude from this example that incorporating SOC terms in the constraints proves to be beneficial, even for $\overrightarrow{\eta}$ chosen smaller than its nominal value. Nevertheless increasing N_P leads, as stated in Theorem 2, to convergence to the optimal trajectory $\bar{\mathbf{x}}^0(\cdot)$ with affine state constraints through trajectories $\bar{\mathbf{x}}^{\delta,\mathrm{fin}}(\cdot)$ that are always both feasible and interior. Moreover the shapes of the SOC-optimal trajectories provide intuition on the times and properties of constrained arcs.

Annex: Existence of interior trajectories

We provide here conditions ensuring the existence of interior trajectories for $(\mathcal{P}_0)$. For any $\overrightarrow{\epsilon} \in \mathbb{R}_+^P$, let $\mathcal{A}_\epsilon := \{(t, \mathbf{x}) \mid t \in [0, T], \overrightarrow{\epsilon} + \mathbf{C}(t)\mathbf{x} \leq \mathbf{d}(t)\}$ and $\mathcal{A}_{\epsilon,t} := \{\mathbf{x} \mid (t, \mathbf{x}) \in \mathcal{A}_\epsilon\}$. Below, for $(t, \mathbf{x}) \in \mathcal{A}_\epsilon$, $T_{\mathcal{A}_\epsilon}(t, \mathbf{x})$ denotes the contingent cone to the set $\mathcal{A}_\epsilon$ at point $(t, \mathbf{x})$ (see e.g. Haddad (1981)).

(H1) $\mathbf{A}(\cdot)$ and $\mathbf{B}(\cdot)$ are $\mathcal{C}^0$-continuous. $\mathbf{C}(\cdot)$ and $\mathbf{d}(\cdot)$ are $\mathcal{C}^1$-continuous and $\mathbf{C}(0)\mathbf{x}_0 < \mathbf{d}(0)$.

(H2) There exists $M_u > 0$ such that, for all $t \in [0, T]$ and $\mathbf{x} \in R\mathbb{B}_N$ satisfying $\mathbf{C}(t)\mathbf{x} \leq \mathbf{d}(t)$, with $R := (1 + \|\mathbf{x}_0\|)e^{T\|\mathbf{A}(\cdot)\|_{L^\infty(0,T)} + TM_u\|\mathbf{B}(\cdot)\|_{L^\infty(0,T)}}$, there exists $\mathbf{u}_{t,x} \in M_u\mathbb{B}_M$ such that

$$\forall\, i \in I_{t,x} := \{i \mid \mathbf{c}_i(t)^\top \mathbf{x} = d_i(t)\}, \ \mathbf{C}_i'(t)^\top \mathbf{x} - d_i'(t) + \mathbf{c}_i(t)^\top(\mathbf{A}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u}_{t,x}) < 0. \qquad (28)$$

The inward-pointing condition (H2) is a geometrical assumption on the boundary of the constraints. In particular, (H2) implies that the constraint set is non degenerate, i.e. $\mathcal{A}_{0,t}$ is the closure of its interior at all times t.

Lemma 4 (Existence of interior trajectories). *Under Assumptions (H1) and (H2), the following properties are satisfied*

i) there exists $\overrightarrow{\epsilon}_0 > 0$, $M_v > 0$, $\xi > 0$, and $\eta > 0$ such that for all $\overrightarrow{\epsilon} \leq \overrightarrow{\epsilon}_0$ and all $(t, \mathbf{x}) \in (\partial \mathcal{A}_\epsilon + (0, \eta\mathbb{B}_N)) \cap \mathcal{A}_\epsilon \cap ([0, T] \times (R - 1)\mathbb{B}_N)$, there exists $\mathbf{u}_{t,x} \in M_u\mathbb{B}_M$ such that $\mathbf{v} = \mathbf{A}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u}_{t,x} \in M_v\mathbb{B}_N$ and

$$\mathbf{y} + \delta(\mathbf{v} + \xi\mathbb{B}_N) \subset \mathcal{A}_{\epsilon,t+\delta} \qquad (29)$$

for all $\delta \in [0, \xi]$ and all $\mathbf{y} \in \mathbf{x} + \xi\mathbb{B}_N$ such that $\mathbf{y} \in \mathcal{A}_{\epsilon,t}$. Hence $\mathbf{v} \in T_{\mathcal{A}_\epsilon}(t, \mathbf{x})$, and

ii) there exists a trajectory in $\mathcal{S}$ satisfying strictly the affine constraints, as well as the initial condition.

The proof below is composed of two parts. The first part translates an inward-pointing condition written in normal form (28) to its tangent form (29). This allows in the second part to invoke a viability argument to derive the existence of a strictly feasible trajectory. This approach may seem remindful of a result by Soner (1986). However, unlike Soner (1986), we do not need to explicitly construct a feasible trajectory approximating a reference one, nor derive an estimate

on their distance. Moreover the proof leverages the linear structure of the dynamics and of the constraints, allowing for time-varying constraints which Soner (1986) did not consider.

Proof: *i)* Define $\mathbf{h}(t, \mathbf{x}) := \mathbf{C}(t)\mathbf{x} - \mathbf{d}(t)$ and $\boldsymbol{\xi}(t, \mathbf{x}, \mathbf{u}) := \mathbf{C}'(t)\mathbf{x} - \mathbf{d}'(t) + \mathbf{C}(t)(\mathbf{A}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u})$.

Step 1: We claim that there exists $\vec{e} > 0$, $\rho > 0$, and $M_v > 0$ such that for all $t \in [0, T]$ and $\mathbf{x} \in R\mathbb{B}_N$ satisfying $\mathbf{h}(t, \mathbf{x}) \leq 0$ with some active constraints (i.e. $I_{t,x} \neq \emptyset$), there exists $\mathbf{u}_{t,x} \in \mathbb{R}^M$ such that $\mathbf{v} = \mathbf{A}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u} \in M_v\mathbb{B}_N$ and

$$\forall i \in I^e_{t,x} := \{i \mid h_i(t, \mathbf{x}) > -e_i\}, \ \xi_i(t, \mathbf{x}, \mathbf{u}_{t,x}) \leq -\rho. \tag{30}$$

Let $\mathbf{u}_{t,x}$ as in (28), and set

$$\rho_{t,x} := -\max_{i \in I_{t,x}} \xi_i(t, \mathbf{x}, \mathbf{u}_{t,x}) > 0 \qquad e_{t,x} := -\max_{i \notin I_{t,x}} h_i(t, \mathbf{x}, \mathbf{u}_{t,x}) > 0.$$

Since $\xi(\cdot, \cdot, \mathbf{u}_{t,x})$ and $\mathbf{h}(\cdot, \cdot)$ are continuous, we can find $\Delta_{t,x} > 0$ such that

$$\sup_{\substack{\delta \in [t - \Delta_{t,x}, t + \Delta_{t,x}] \cap [0,T], \\ \mathbf{w} \in \mathbb{B}_N}} \|\mathbf{h}(t, \mathbf{x}) - \mathbf{h}(t + \delta, \mathbf{x} + \Delta_{t,x}\mathbf{w})\|_\infty \leq \frac{e_{t,x}}{2}$$

$$\sup_{\substack{\delta \in [t - \Delta_{t,x}, t + \Delta_{t,x}] \cap [0,T], \\ \mathbf{w} \in \mathbb{B}_N, i \in I_{t,x}}} |\xi_i(t, \mathbf{x}, \mathbf{u}_{t,x}) - \xi_i(t + \delta, \mathbf{x} + \Delta_{t,x}\mathbf{w}, \mathbf{u}_{t,x})| \leq \frac{\rho_{t,x}}{2}.$$

This implies that the index set of active constraints does not increase in size for points in the open set $\Omega_{t,x} := (t, \mathbf{x}) + \Delta_{t,x}] - 1, 1[\times \mathring{\mathbb{B}}_N$, denoting by $\mathring{\mathbb{B}}_N$ the open unit ball of $\mathbb{R}^N$. Moreover, for any point in $\Omega_{t,x}$, $\mathbf{u}_{t,x}$ satisfies our claim for the constants $\rho_{t,x}/2$ and $(e_{t,x}/2)\,\mathbf{1}_N$. Since we are considering a compact set, we can select a finite number of $(t_j, \mathbf{x}_j)_{j \in [\![1,J]\!]}$ such that

$$([0, T] \times R\mathbb{B}_N) \cap \mathcal{A}_0 \subset \bigcup_{j \in [\![1,J]\!]} \Omega_{t_j, \mathbf{x}_j}.$$

To conclude, simply take[20]

$$\rho := \min_{j \in [\![1,J]\!]} \frac{\rho_{t_j, \mathbf{x}_j}}{2}, \qquad \mathbf{e} := \left(\min_{j \in [\![1,J]\!]} \frac{e_{t_j, \mathbf{x}_j}}{2}\right)\mathbf{1}_N, \qquad M_v := R\|\mathbf{A}(\cdot)\|_{L^\infty(0,T)} + M_u\|\mathbf{B}(\cdot)\|_{L^\infty(0,T)}.$$

Step 2: Thanks to the uniform constants of the normal form (30), we derive the constants of the tangent form (29). Let us choose $\xi, \eta \in]0, 1]$ such that for all $(t, \mathbf{x}) \in \partial\mathcal{A}_0 \cap ([0, T] \times R\mathbb{B}_N)$, setting $\mathbf{v}_{t,x} = \mathbf{A}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u}_{t,x} \in M_v\mathbb{B}_N$, for

$$\mathbf{e}^{t,x,\delta}_{\alpha,\beta,\gamma} := (\mathbf{C}(t+\delta) - \mathbf{C}(t))(\mathbf{x} + \eta\boldsymbol{\gamma} + \xi\boldsymbol{\alpha}) + \delta\mathbf{C}(t+\delta)(\mathbf{v}_{t,x} + \eta\mathbf{A}(t)\boldsymbol{\gamma} + \xi\boldsymbol{\beta}) + \mathbf{d}(t) - \mathbf{d}(t+\delta),$$

we have

$$\frac{\mathbf{e}}{2} > \sup_{\substack{\delta \in [t-\xi, t+\xi] \cap [0,T] \\ \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{B}_N}} \mathbf{e}^{t,x,\delta}_{\alpha,\beta,\gamma} \quad , \quad \frac{-\delta\rho}{2} > \sup_{\substack{\delta \in [t-\xi, t+\xi] \cap [0,T] \\ \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{B}_N, i \in I^e_{t,x}}} (\mathbf{e}^{t,x,\delta}_{\alpha,\beta,\gamma})_i. \tag{31}$$

---

[20]If $\mathcal{A}_0$ is bounded, take $M_u := 1 + \max_{j \in [\![1,J]\!]} \|\mathbf{u}_{t_j, \mathbf{x}_j}\|$.

The first inequality can be derived from the $\mathcal{C}^1$-smoothness of $\mathbf{C}(\cdot)$ and $\mathbf{d}(\cdot)$, and from the fact that both $\mathbf{x}$, $\mathbf{v}_{t,x}$, $\mathbf{A}(\cdot)$ and $\mathbf{C}(\cdot)$ are uniformly bounded. The second inequality stems from (30), as $\lim_{\delta \to 0^+} \frac{\mathbf{e}_{0,0,0}^{t,x,\delta}}{\delta} = \xi(t, \mathbf{x}, \mathbf{u}_{t,x})$. The two inequalities of (31) state that, for $\mathbf{y} := \mathbf{x} + \eta\boldsymbol{\gamma} + \xi\boldsymbol{\alpha}$, if $\mathbf{y} \in \mathcal{A}_{0,t}$ then

$$\mathbf{y} + \delta(\mathbf{v} + \xi\mathbb{B}_N) \subset \mathcal{A}_{0,t+\delta} \text{ with } \mathbf{v} := \mathbf{v}_{t,x} + \eta\mathbf{A}(t)\boldsymbol{\gamma} = \mathbf{A}(t)(\mathbf{x} + \eta\boldsymbol{\gamma}) + \mathbf{B}(t)\mathbf{u}_{t,x}.$$

Step 3: There just remains to extend the result of Step 2 to the perturbed constraint sets $\mathcal{A}_\epsilon$. For any $t \in [0, T]$, let $\mathbf{w}_t$ be an eigenvector of the largest eigenvalue μ_t of $\mathbf{C}(t)\mathbf{C}(t)^T$ satisfying $\|\mathbf{w}_t\|_\infty = \max_{i \in [\![1, P]\!]} w_{t,i} = 1$. Let $\bar{\mu} := \min_{t \in [0, T]} \mu_t > 0$. Take any $\overrightarrow{\boldsymbol{\epsilon}}_0 > 0$ such that

$$\sup_{t \in [0, T]} \frac{2\|\overrightarrow{\boldsymbol{\epsilon}}_0\|_\infty \|\mathbf{C}(t)^\top \mathbf{w}_t\|}{\bar{\mu}} \le \frac{\eta}{2}.$$

Let $\overrightarrow{\boldsymbol{\epsilon}} \le \overrightarrow{\boldsymbol{\epsilon}}_0$, take $(t, \mathbf{x}) \in \partial\mathcal{A}_\epsilon \cap ([0, T] \times (R-1)\mathbb{B}_N)$. Then

$$\exists i \in [\![1, P]\!], \ \mathbf{c}_i(t)^\top \left(\mathbf{x} + \frac{2\|\overrightarrow{\boldsymbol{\epsilon}}_0\|_\infty \mathbf{C}(t)^\top \mathbf{w}_t}{\bar{\mu}} \right) = d_i(t) - \overrightarrow{\boldsymbol{\epsilon}}_i + \frac{2\mu_t}{\bar{\mu}}\|\overrightarrow{\boldsymbol{\epsilon}}_0\|_\infty > 0.$$

Since $\eta \le 1$, one can therefore find $\check{\mathbf{x}} \in R\mathbb{B}_N \cap \partial\mathcal{A}_{0,t}$ such that $\|\mathbf{x} - \check{\mathbf{x}}\| = d_{\mathcal{A}_{0,t}}(\mathbf{x}) \le \eta/2$, to which the conclusions of Step 2 apply. So the set considered in (29) at $\mathbf{x}$, taking as constants $\xi/2$ and $\eta/2$, is a subset of the one defined by $\check{\mathbf{x}}$, ξ and η. Since only differences of $\mathbf{d}(\cdot)$ appeared in Step 2 and that $\mathbf{x} \in \partial\mathcal{A}_{\epsilon,t}$, adding $\overrightarrow{\boldsymbol{\epsilon}}$ has no effect on the computations, so it follows that (29) is satisfied. For $\mathbf{y} = \mathbf{x}$, by definition of $T_{\mathcal{A}_\epsilon}(t, \mathbf{x})$, we have that $(1, \mathbf{v}) \in T_{\mathcal{A}_\epsilon}(t, \mathbf{x})$.

ii) Let $\mathbf{F}(t, \mathbf{x}) := \{(1, \mathbf{A}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u}) \mid \mathbf{u} \in M_u\mathbb{B}_M\}$. Since $\mathbf{A}(\cdot)$ and $\mathbf{B}(\cdot)$ are $\mathcal{C}^0$-continuous, the bounded set-valued map $\mathbf{F}(\cdot, \cdot)$ is upper semicontinuous. Let $\Omega := \mathcal{A}_\epsilon \cap ([0, T] \times (R-1)\mathring{\mathbb{B}}_N)$. The open ball $\mathring{\mathbb{B}}_N$ being open in $\mathbb{R}^N$, Ω is locally compact in $[0, T] \times \mathbb{R}^N$. In *i)* we have shown that the local viability condition is satisfied for $\mathcal{A}_\epsilon$. The intersection with an open set does not add boundary points, so, for $(t, \mathbf{x}) \in \Omega$, $\mathbf{F}(t, \mathbf{x}) \cap T_\Omega(t, \mathbf{x}) \ne \emptyset$. We may therefore apply (Haddad, 1981, Theorem 1) which provides a trajectory $\mathbf{x}^\epsilon(\cdot)$ satisfying $\mathbf{x}^\epsilon(0) = \mathbf{x}_0$, $\mathbf{x}^{\epsilon,'}(t) \in \mathbf{F}(t, \mathbf{x}^\epsilon(t))$, and $\mathbf{h}(t, \mathbf{x}^\epsilon(t)) + \overrightarrow{\boldsymbol{\epsilon}} \le 0$. Let $[0, t_1[$ be the maximal interval of existence of $\mathbf{x}^\epsilon(\cdot)$. Since $\mathbf{x}^{\epsilon,'}(\cdot)$ is measurable, by (Vinter, 1990, Theorem 2.3.13), we can find some measurable $\mathbf{u}^\epsilon(\cdot)$ with values bounded by M_u, s.t. $\mathbf{x}^{\epsilon,'}(\cdot) = \mathbf{A}(t)\mathbf{x} + \mathbf{B}(t)\mathbf{u}^\epsilon(t)$ a.e. in $[0, T]$. The dynamics being sublinear as $\mathbf{A}(\cdot)$, $\mathbf{B}(\cdot)$, and $\mathbf{u}^\epsilon(\cdot)$ are bounded, $\mathbf{x}^\epsilon(\cdot)$ can be continuously extended to $(t_1, \mathbf{x}^\epsilon(t_1)) \in \Omega$. So $t_1 \ge T$, otherwise the viability condition would allow to extend $\mathbf{x}^\epsilon(\cdot)$ beyond t_1. Since $\mathbf{u}^\epsilon(\cdot)$ is measurable and bounded, $\mathbf{u}^\epsilon(\cdot) \in L^2(0, T)$. Hence $\mathbf{x}^\epsilon(\cdot) \in \mathcal{S}$ satisfies the required properties.

■

References

Ackermann, J. (1985). *Sampled-Data Control Systems.* Springer Berlin Heidelberg.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404.

Aubin-Frankowski, P.-C. (2020). Interpreting the dual Riccati equation through the LQ reproducing kernel. *Comptes Rendus. Mathématique.* (`https://arxiv.org/abs/2012.12940`).

Aubin-Frankowski, P.-C., Petit, N., and Szabó, Z. (2020). Kernel regression for vehicle trajectory reconstruction under speed and inter-vehicular distance constraints. In *IFAC World Congress*, volume (to appear).

Aubin-Frankowski, P.-C. and Szabó, Z. (2020a). Handling hard affine SDP shape constraints in RKHSs. Technical report. (`https://arxiv.org/abs/2101.01519`).

Aubin-Frankowski, P.-C. and Szabó, Z. (2020b). Hard shape-constrained kernel machines. In *Advances in Neural Information Processing Systems (NIPS)*.

Bertalan, T., Dietrich, F., Mezić, I., and Kevrekidis, I. G. (2019). On learning Hamiltonian systems from data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(12):121107–1 – 121107–9.

Bonnans, J. F. and Ramírez, H. (2005). Perturbation analysis of second-order cone programming problems. *Mathematical Programming*, 104(2-3):205–227.

Bourdin, L. and Trélat, E. (2017). Linear–quadratic optimal sampled-data control problems: Convergence result and Riccati theory. *Automatica*, 79:273–281.

Burachik, R. S., Kaya, C. Y., and Majeed, S. N. (2014). A Duality Approach for Solving Control-Constrained Linear-Quadratic Optimal Control Problems. *SIAM Journal on Control and Optimization*, 52(3):1423–1456.

Chaplais, F., Malisani, P., and Petit, N. (2011). Design of penalty functions for optimal control of linear dynamical systems under state and input constraints. In *Conference on Decision and Control (CDC)*, pages 6697–6704.

Chiuso, A. and Pillonetto, G. (2019). System Identification: A Machine Learning Perspective. *Annual Review of Control, Robotics, and Autonomous Systems*, 2(1):281–304.

Dower, P. M., McEneaney, W. M., and Cantoni, M. (2019). Game representations for state constrained continuous time linear regulator problems. *arXiv:1904.05552*. http://arxiv.org/abs/1904.05552.

Fujii, K. and Kawahara, Y. (2019). Dynamic mode decomposition in vector-valued reproducing kernel Hilbert spaces for extracting dynamical structure among observables. *Neural Networks*, 117:94–103.

Fujioka, H. and Kano, H. (2013). Control theoretic B-spline smoothing with constraints on derivatives. pages 2115–2120.

Gerdts, M. and Hüpping, B. (2012). Virtual control regularization of state constrained linear quadratic optimal control problems. *Computational Optimization and Applications*, 51(2):867–882.

Giannakis, D., Das, S., and Slawinska, J. (2019). Reproducing kernel Hilbert space compactification of unitary evolution groups. http://arxiv.org/abs/1808.01515.

Grüne, L. and Guglielmi, R. (2018). Turnpike properties and strict dissipativity for discrete time linear quadratic optimal control problems. *SIAM Journal on Control and Optimization*, 56(2):1282–1302.

Haddad, G. (1981). Monotone trajectories of differential inclusions and functional differential inclusions with memory. *Israel Journal of Mathematics*, 39(1-2):83–100.

Hartl, R. F., Sethi, S. P., and Vickson, R. G. (1995). A Survey of the Maximum Principles for Optimal Control Problems with State Constraints. *SIAM Review*, 37(2):181–218.

Heckman, N. (2012). The theory and application of penalized methods or Reproducing Kernel Hilbert Spaces made easy. *Statistics Surveys*, 6(0):113–141.

Hermant, A. (2009). Stability analysis of optimal control problems with a second-order state constraint. *SIAM Journal on Optimization*, 20(1):104–129.

Kailath, T. (1971). RKHS approach to detection and estimation problems–I: Deterministic signals in Gaussian noise. *IEEE Transactions on Information Theory*, 17(5):530–549.

Kano, H. and Fujioka, H. (2018). B-spline trajectory planning with curvature constraint. In *2018 Annual American Control Conference (ACC)*, pages 1963–1968.

Kojima, A. and Morari, M. (2004). LQ control for constrained continuous-time systems. *Automatica*, 40(7):1143–1155.

Magnus Egerstedt, C. M. (2009). *Control Theoretic Splines: Optimal Control, Statistics, and Path Planning (Princeton Series in Applied Mathematics)*. Princeton University Press.

Marco, A., Hennig, P., Schaal, S., and Trimpe, S. (2017). On the design of LQR kernels for efficient controller learning. In *Conference on Decision and Control (CDC)*, pages 5193–5200. IEEE.

Mattingley, J. and Boyd, S. (2012). CVXGEN: a code generator for embedded convex optimization. *Optimization and Engineering*, 13(1):1–27.

Mercy, T., Van Loock, W., and Pipeleers, G. (2016). Real-time motion planning in the presence of moving obstacles. In *European Control Conference (ECC)*, pages 1586–1591.

Micheli, M. and Glaunès, J. A. (2014). Matrix-valued kernels for shape deformation analysis. *Geometry, Imaging and Computing*, 1(1):57–139.

Parzen, E. (1970). Statistical inference on time series by RKHS methods. In *Proceedings 12th Biennial Seminar, Canadian Mathematical Congress*.

Pillonetto, G., Dinuzzo, F., Chen, T., Nicolao, G. D., and Ljung, L. (2014). Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3):657 – 682.

Rosenfeld, J. A., Kamalapurkar, R., Gruss, L. F., and Johnson, T. T. (2019). Dynamic Mode Decomposition for Continuous Time Systems with the Liouville Operator. http://arxiv.org/abs/1910.03977.

Schölkopf, B., Herbrich, R., and Smola, A. J. (2001). A generalized representer theorem. In *Computational Learning Theory (CoLT)*, pages 416–426. Springer Berlin Heidelberg.

Schölkopf, B. and Smola, A. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press.

Schwartz, L. (1964). Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *Journal d'Analyse Mathématique*, 13:115–256.

Singh, S., Sindhwani, V., Slotine, J.-J., and Pavone, M. (2018). Learning stabilizable dynamical systems via control contraction metrics. In *Workshop on Algorithmic Foundations of Robotics (WAFR).* https://arxiv.org/abs/1808.00113.

Soner, H. M. (1986). Optimal control with state-space constraint i. *SIAM Journal on Control and Optimization*, 24(3):552–561.

Sootla, A., Mauroy, A., and Ernst, D. (2018). Optimal control formulation of pulse-based control using Koopman operator. *Automatica*, 91:217–224.

Speyer, J. L. and Jacobson, D. H. (2010). *Primer on Optimal Control Theory.* Society for Industrial and Applied Mathematics, USA.

Steinke, F. and Schölkopf, B. (2008). Kernels, regularization and differential equations. *Pattern Recognition*, 41(11):3271–3286.

van Keulen, T. (2020). Solution for the continuous-time infinite-horizon linear quadratic regulator subject to scalar state constraints. *IEEE Control Systems Letters*, 4(1):133–138.

Van Loan, C. (1978). Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control*, 23(3):395–404.

Vinter, R. (1990). *Optimal Control.* Birkhäuser, Basel.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM, CBMS-NSF Regional Conference Series in Applied Mathematics.

Williams, M. O., , Rowley, C. W., Kevrekidis, I. G., and and (2015). A kernel-based method for data-driven Koopman spectral analysis. *Journal of Computational Dynamics*, 2(2):247–265.