THE REPRODUCING KERNEL HILBERT SPACES UNDERLYING LINEAR SDE ESTIMATION, KALMAN FILTERING AND THEIR RELATION TO OPTIMAL CONTROL

PIERRE-CYRIL AUBIN-FRANKOWSKI AND ALAIN BENSOUSSAN

In honor of Prof. Wendell Fleming,

ABSTRACT. It is often said that control and estimation problems are in duality. Recently, in [1], we found new reproducing kernels in Linear-Quadratic optimal control by focusing on the Hilbert space of controlled trajectories, allowing for a convenient handling of state constraints and meeting points. We now extend this viewpoint to estimation problems where it is known that kernels are the covariances of stochastic processes. Here, the Markovian Gaussian processes stem from the linear stochastic differential equations describing the continuous-time dynamics and observations. Taking extensive care to require minimal invertibility requirements on the operators, we give novel explicit formulas for these covariances. We also determine their reproducing kernel Hilbert spaces, stressing the symmetries between a space of forward-time trajectories and a space of backward-time information vectors. The two spaces play an analogue role for filtering to Sobolev spaces in variational analysis, and allow to recover the Kalman estimate through a direct variational argument. For comparison, we then recover the Kalman filter and smoother formulas through more classical arguments based on the innovation process. Extension to discrete-time observations or infinite-dimensional state, tough technical, would be straightforward.

1. INTRODUCTION

Context. In preceding papers [1, 2, 3], connections have been made between the theory of reproducing kernel Hilbert spaces and control theory. Framing control problems as optimizing over Banach vector spaces of functions with a dynamic constraint was already discussed in [21, p255]. However, in the linear-quadratic case, more specific Hilbertian structures actually emerge, the quadratic cost defining a norm over the vector space of linearly controlled trajectories. The latter has then an explicit reproducing kernel which is completely determined by the cost and dynamics. Conversely, this kernel summarizes all the information about the system, and was shown in some cases to coincide with the inverse of the solution of the backward-time differential Riccati equation [2] or with the controllability Gramian [1]. Note that this approach does not rely on the value function, which does not live in a vector space but rather in a max-plus one, and has thus ties with

tropical kernels [4].

Another very related field is the problem of estimation. Its "duality" with control theory has been studied extensively since the seminal paper by [16], but the word dual is unfortunately often abusively used, and does not always correspond to rigorous dual spaces in a mathematical sense or to dual min-max/max-min problems. We refer to [14, Chapter 15] for a formalization of this duality in discrete-time and to [19, Chapter 3] for valuable insights in continuous-time. Unlike in control theory, kernels have been known in estimation problems since the origin of the field. Indeed reproducing kernels are the covariances of stochastic processes in this context [24, 6], and, in the Gaussian case, summarize all the information. In this Bayesian field of Gaussian process regression, kernels are ubiquitous, forming the deterministic counterpart of the stochastic processes [25, 17]. This regression framework was in particular applied to stochastic differential equations [see 28, for an introduction and recent review] in close relation with Kalman filtering. Nevertheless, to the best of our knowledge, the reproducing kernel Hilbert spaces (RKHSs) associated with the continuous-time Gaussian processes stemming from linear stochastic differential equations (SDEs) have not been written explicitly.

Main results. To achieve the identification of the RKHSs, we extensively use the more recent theory of operator-valued kernels [see 9, and references therein]. We prove that the linear SDEs define two vector-valued RKHSs over the time interval: one, forward, for the reconstructible trajectories, and one, backward, for the information vectors. The values of the trajectory kernel coincide with the covariance of the minimal estimation error, and thus with the Kalman filter and Rauch-Tung-Striebel Smoother. For the information kernel, the connection is with the Gramian of observability. We provide new closed-form expressions for the covariance kernels of these Markovian Gaussian processes. By drawing upon the connection between covariance and RKHSs, we also generalize the kernel formulas obtained previously for optimal control and define two dual deterministic optimization problems associated with the smoothing task. Since we work in an estimation context, extensive care was taken to provide minimal invertibility requirements on the matrices involved. To simplify a little, we limit ourselves to finite-dimensional systems, but generalization to infinite-dimensional systems (i.e. nonstationnary spatiotemporal Gaussian processes, [29, 20]) can be done in the spirit of [3].

Related work. For discrete-time differential equations, the benefits and limitations of kernel regression, defined over the time axis, were discussed in [31]. This contrasts with the wider use of kernels, defined over observation and state spaces, as off-the-shelf tools for estimation as in [18]. Our setting is closer to the former, with all our kernels defined over time. Characterization of Markovian Gaussian processes by the form of their kernel was recalled in [23, Chapter 3.2] and [6, Example 2, p58], and further studied in [12]. The formulas for the covariance kernel of a linear SDE over the state variable but without an observation process were given in [28, Section 6.4, p.89]. Up to our knowledge, while every linear SDE leads to a Markovian Gaussian process, the converse is not known in the general case.

Concerning duality, [19] summarizes the traditional input-output viewpoint, with the duality appearing through the adjoint of the map sending the initial condition to the output. [19] then delves into a promising duality framework for the nonlinear setting. We adopt a different viewpoint here, kernel-based, focusing on covariances and Hilbert spaces of trajectories.

The paper is structured as follows. The filtering and smoothing settings and their linear estimators are presented in Section 2 and related to a general perturbed two-point boundary value problem. Section 3 gives the formulas of the Green kernels for this problem, based on impulse-response. In Section 4, these kernels are shown to be the reproducing kernels of two functional spaces, of controlled trajectories and of information vectors respectively, and two dual optimization problems defined. In Section 5, for completeness, we rederive various known formulas for the Kalman filter and RTS smoothing. The results are finally summarized in Section 6 and compared to those previously obtained by the authors in optimal control. We provide in the Appendix implementable formulas to compute the kernels based on the Hamiltonian matrix.

2. Kalman filtering and RTS smoothing in continuous time

2.1. **Statement of the problem.** We follow the presentation of [5, Chapter 7]. We consider on a probability space $(\Omega, \mathcal{A}, P)$ a filtration $\mathcal{F}^t$ and two independent standard Wiener spaces $w(t)$, $b(t)$ with values in $\mathbb{R}^p$ and $\mathbb{R}^m$, and respective covariances Q and R in the sense that the following correlation formulas hold

$$(2.1) \qquad \mathbb{E}[w(s)w(t)^*] = \int_{t_0}^{\min(s,t)} Q(\tau)d\tau, \ \mathbb{E}[b(s)b(t)^*] = \int_{t_0}^{\min(s,t)} R(\tau)d\tau$$

where $w(t)^* \in \mathbb{R}^{p,*}$ denotes the transpose and in which $Q(\cdot) \in L^1([t_0, T], \mathcal{L}(\mathbb{R}^{d,*}; \mathbb{R}^d))$, $R(\cdot) \in L^2([t_0, T], \mathcal{L}(\mathbb{R}^{m,*}; \mathbb{R}^m))$ and, for some $r > 0$, $R(\cdot) \succcurlyeq r \operatorname{Id}$ a.e. The processes are assumed to be adapted to the filtration $\mathcal{F}^t$. We observe a dynamic system with partial information. The dynamic system is characterized by its state $x(t) \in \mathbb{R}^n$, which evolves according to the model

$$(2.2) \quad dx(t) = (F(t)x(t) + f(t))dt + G(t)dw(t), \quad x(t_0) = x_0 + \xi, \quad \xi \sim \mathcal{N}(0, \Pi_0)$$

in which $F(\cdot) \in L^1([t_0, T], \mathcal{L}(\mathbb{R}^n; \mathbb{R}^n))$, $G(\cdot) \in L^2([t_0, T], \mathcal{L}(\mathbb{R}^d; \mathbb{R}^n))$ and $f(\cdot) \in L^1([t_0, T], \mathbb{R}^n)$ are fixed deterministic functions, which are known, and $x_0 \in \mathbb{R}^n$, is a deterministic known vector. The random variable ξ has values in $\mathbb{R}^n$, and is assumed Gaussian with zero mean and covariance matrix Π_0 such that ξ, $w(\cdot)$, $b(\cdot)$ are mutually independent. Note that we require R to be invertible but do not require it for Q or Π_0. The state $x(t)$ is not observed, we have instead a continuous-time observation process $y(t)$ with values in $\mathbb{R}^m$, related to $x(t)$ by the following relation[1]

$$(2.3) \qquad\qquad dy(t) = (H(t)x(t) + h(t))dt + db(t), \quad y(t_0) = y_0$$

---

[1]A discrete-time observation process $y(t_i) = H(t_i)x(t_i) + \varepsilon_i$ could easily be considered, without changing formalism, with the same operator H and independent Gaussian noises $\varepsilon_i \sim \mathcal{N}(0, R_i)$.

in which $H(\cdot) \in L^2([t_0, T], \mathcal{L}(\mathbb{R}^n; \mathbb{R}^m))$, $h(\cdot) \in L^1([t_0, T], \mathbb{R}^m)$. At any final time T, the available information is the trajectory $(y(t))_{t \in [t_0, T]}$ and we want to estimate the value of $x(s)$ for $s \in [t_0, T]$. When $s < T$, the best estimate is called the RTS smoother, and when $s = T$, it is the Kalman filter.

2.2. **Best estimate.** From (2.2) and (2.3), we deduce that the processes $x(t)$ and $y(t)$ are Gaussian, with mean $\bar{x}(t)$ and $\bar{y}(t)$ given by the equations

$$(2.4) \qquad \frac{d\bar{x}}{dt}(t) = F(t)\bar{x}(t) + f(t), \qquad\qquad \bar{x}(t_0) = x_0,$$

$$(2.5) \qquad \frac{d\bar{y}}{dt}(t) = H(t)\bar{x}(t) + h(t), \qquad\qquad \bar{y}(t_0) = y_0.$$

We introduce the processes $\widetilde{x}(t)$ and $\widetilde{y}(t)$ with zero mean, $\widetilde{x}(t) := x(t) - \bar{x}(t)$, $\widetilde{y}(t) := y(t) - \bar{y}(t)$. They are solutions of the SDEs

$$(2.6) \qquad d\widetilde{x}(t) = F(t)\widetilde{x}(t)dt + G(t)dw(t), \qquad\qquad \widetilde{x}(t_0) = \xi,$$

$$(2.7) \qquad d\widetilde{y}(t) = H(t)\widetilde{x}(t)dt + db(t), \qquad\qquad \widetilde{y}(t_0) = 0.$$

The formal problem is to estimate $x(s)$ with the σ-algebra $\mathcal{Y}^T = \sigma(y(\tau), 0 \leq \tau \leq T)$. For the minimum mean square estimator, the solution is well-known [23], it is the conditional expectation:

$$(2.8) \qquad\qquad \hat{x}(s|T) = E[x(s)|\mathcal{Y}^T].$$

The random variable $\hat{x}(s|T)$, for $s < T$, is the RTS smoother estimate, while $\hat{x}(T|T)$ is the Kalman filter estimate. However the expression (2.8) is not operational and we need more implementable formulas. The most important element we want to exploit is that, thanks to the fact that the processes $x(\cdot)$ and $y(\cdot)$ are Gaussian, the conditional expectation coincides with the best unbiased linear estimate, a.k.a. the minimum variance linear estimator. A linear unbiased estimate is characterized by an operator $S(t) \in \mathcal{L}(\mathbb{R}^m; \mathbb{R}^n)$. We estimate $x(s)$ by the minimum variance linear estimator $x_S(s|T)$ defined by the formula, related to Wiener filtering,

$$(2.9) \qquad\qquad x_S(s|T) := \bar{x}(s) + \int_{t_0}^T S_s(t|T)d\widetilde{y}(t)$$

which is obviously unbiased. The estimation error $\epsilon_S(s|T)$ thus satisfies

$$(2.10) \qquad \epsilon_S(s|T) := x(s) - x_S(s|T) = \widetilde{x}(s) - \int_{t_0}^T S_s(t|T)d\widetilde{y}(t).$$

The objective is to find $\hat{S}_s(\cdot|T)$ minimizing the covariance matrix of the error

$$(2.11) \qquad\qquad \hat{S}_s(\cdot|T) \in \underset{S(\cdot|T)}{\mathrm{argmin}}\, \Gamma_S(s|T) = \mathbb{E}[\epsilon_S(s|T)(\epsilon_S(s|T))^*].$$

This minimization must be interpreted in the sense of the operator norm of positive matrices. The fact that this problem has a solution is a fundamental result of Kalman smoothing and filtering theory. In this work we obtain a new expression for the optimal operator $\hat{S}_s(\cdot|T)$ by finding a closed-form formula for the following proper covariance function,

$$(2.12) \qquad\qquad K(s, t|T) = \mathbb{E}[\epsilon_{\hat{S}_s}(s|T)(\epsilon_{\hat{S}_t}(t|T))^*] \in \mathcal{L}(\mathbb{R}^{n,*}, \mathbb{R}^n)$$

We show below in Corollary 3.2 that $\hat{S}_s(t|T) = K(s,t|T)H^*(t)R^{-1}(t)$. From a Gaussian process perspective, K can be interpreted as the posterior covariance given $\widetilde{y}$ with the prior (2.1) over $x(\cdot)$. The fact that covariances K are also reproducing kernels is well-known since [24]. [6, Chapter 2] summarized the deep relations between positive semidefinite kernels and stochastic processes, albeit only in the real-valued case. [17] also reviewed the connections between kernel methods and Gaussian processes. We meet here a similar relation to the classical "duality" between estimation and control in the Linear-Quadratic-Gaussian setting, which we will come back to in Sections 2.3 and 4.4. The kernels have a "dual" nature, both deterministic, as Green functions of differential equations, and stochastic, as covariances of second order processes.

2.3. **A detour through estimators.** In this section only, we consider infinite-dimensional operators to give the high-level idea before introducing the various quantities studied in the next sections. We will proceed formally, following the introduction of [5, Chapter 4.5] on finite-dimensional best estimators and the short presentation of [6, Chapter 2.4] of Hilbert spaces generated by a process with finite second order moments. For such a process $X = (X_t)_{t \in [t_0,T]}$, we write $\mathcal{C}_X$ its covariance operator. For simplicity, we consider only zero-mean processes X, Y and only quantities having finite second order moments. Setting $\mathscr{T} = [t_0, T]$, the (Bayesian) minimum mean square estimator (MMSE) is defined as

$$\text{(MMSE)} \qquad \min_{\hat{X} \in L^2(\Omega \times \mathscr{T}, \mathbb{R}^n), \, \Phi \text{ meas.}, \, \hat{X} = \Phi(Y)} \mathbb{E}((X - \hat{X})^\top (X - \hat{X})).$$

where Φ is a measurable function from $L^2(\Omega \times \mathscr{T}, \mathbb{R}^m)$ to $L^2(\Omega \times \mathscr{T}, \mathbb{R}^n)$, whence $\hat{X}$ belongs to the space $\bar{\mathcal{N}}(Y)$ of nonlinear functionals of Y. The linear MMSE, which coincides with the minimum variance linear estimator (MVLE), instead restricts the search space to the space $\bar{\mathcal{L}}(Y)$ of linear functionals of the process Y

$$\text{(MVLE)} \qquad \min_{\hat{X} \in L^2(\Omega \times \mathscr{T}, \mathbb{R}^n), \, \mathcal{S} \in \mathcal{L}(L^2(\Omega \times \mathscr{T}, \mathbb{R}^m), L^2(\Omega \times \mathscr{T}, \mathbb{R}^n)), \, \hat{X} = \mathcal{S}Y} \mathbb{E}((X - \hat{X})^\top (X - \hat{X})).$$

Gaussian process regression for real-valued x, i.e. $n = 1$, further restricts the search space, by introducing the canonical congruence ψ_Y between the process Y and its RKHS $\mathcal{H}_Y$, i.e. the linear isomorphism satisfying $\psi_Y(v^\top Y_t)(s) = \mathbb{E}[Y_s Y_t^\top]v$ for all $v \in \mathbb{R}^m$,

$$\text{(GP-reg)} \qquad \min_{\hat{X} \in L^2(\Omega \times \mathscr{T}, \mathbb{R}^n), \, g \in \mathcal{H}_Y, \, \hat{X} = \psi_Y^{-1}(g)} \mathbb{E}((X - \hat{X})^\top (X - \hat{X})).$$

By convex duality, introducing a process Λ which acts as a Lagrange multiplier, the dual problem of (MVLE) can be written as

$$\max_{\substack{\Lambda \in L^2(\Omega \times \mathscr{T}, \mathbb{R}^{n,*})}} \min_{\substack{\hat{X} \in L^2(\Omega \times \mathscr{T}, \mathbb{R}^n), \\ \mathcal{S} \in \mathcal{L}(L^2(\Omega \times \mathscr{T}, \mathbb{R}^m), L^2(\Omega \times \mathscr{T}, \mathbb{R}^n))}} \mathbb{E}((X - \hat{X})^\top (X - \hat{X})) + 2\left\langle \Lambda, \hat{X} - \mathcal{S}Y \right\rangle_{L^2(\Omega \times \mathscr{T}, \mathbb{R}^n)}.$$

Minimizing over $\mathcal{S}$ imposes $\Lambda \in \bar{\mathcal{L}}(Y)^\perp$, the orthogonal space of $\bar{\mathcal{L}}(Y)$ in $L^2(\Omega \times \mathscr{T}, \mathbb{R}^n)$. Minimizing over $\hat{X}$ gives $\hat{X}^\top = X^\top - \hat{\Lambda}$ so

$$\text{(MVLE-dual)} \qquad \min_{\Lambda \in \bar{\mathcal{L}}(Y)^\perp} \mathbb{E}((X^\top - \Lambda)^\top (X^\top - \Lambda)).$$

This covector or adjoint process does not exist in the Gaussian process framework because the duality is not taken in L^2 but in $\mathcal{H}_Y$, seen as self-adjoint. For such orthogonal spaces to appear, one has to work in a larger space than $\mathcal{H}_Y$, such as L^2. Similarly, in optimal control, working in the space of absolutely continuous trajectories leads to defining a covector function. Working in the RKHS of linearly controlled trajectories as in [1] bypasses the need of such covector.[2]

Coming back to (MVLE), the first order optimality condition writes as the orthogonality criterion $\mathbb{E}((X - \hat{X})Y^\top) = 0$ which gives that $\hat{\mathcal{S}} = \mathcal{C}_{XY}\mathcal{C}_Y^{-1}$ if $\mathcal{C}_Y$ is invertible. If $Y = \mathcal{H}X + Z$, with invertible $\mathcal{C}_Z = \mathcal{R}$, then $\mathcal{C}_Y$ has an inverse and, setting $\mathcal{C}_\epsilon = \mathcal{C}_X - \mathcal{C}_{\hat{X},\hat{\mathcal{S}}}$, further assuming $\mathcal{C}_X$ to be invertible and using the Woodbury identity, we obtain that $\hat{\mathcal{S}} = \mathcal{C}_\epsilon \mathcal{H}^\top \mathcal{R}^{-1}$. We note also that for any deterministic $\bar{\lambda}(\cdot) \in L^2([t_0, T], \mathbb{R}^{n,*})$, $\hat{\mathcal{S}}$ also minimizes

$$\min_{\hat{X} \in L^2(\Omega \times \mathcal{T}, \mathbb{R}^n), \mathcal{S} \in \mathcal{L}(L^2(\Omega \times \mathcal{T}, \mathbb{R}^m), L^2(\Omega \times \mathcal{T}, \mathbb{R}^n)), \hat{X} = \mathcal{S}Y} \mathbb{E}(\|\bar{\lambda}(\cdot)^\top (X - \hat{X})\|_{L^2}^2).$$

Defining the deterministic $v(\cdot) = \mathcal{S}^\top \bar{\lambda}(\cdot)$, $\hat{\mathcal{S}}$ minimizes a "control" problem over v, with a specific rewriting for $Y = \mathcal{H}X + Z$,

$$\text{(MVLE-det)} \quad \min_{\substack{\mathcal{S} \in \mathcal{L}(L^2(\mathcal{T}, \mathbb{R}^m), L^2(\mathcal{T}, \mathbb{R}^n)), \\ v(\cdot) = \mathcal{S}^\top \bar{\lambda}(\cdot)}} \underbrace{\frac{\langle \bar{\lambda}(\cdot), \mathcal{C}_X \bar{\lambda}(\cdot) \rangle_{L^2} + \langle v(\cdot), \mathcal{C}_Y v(\cdot) \rangle_{L^2} - 2\langle \bar{\lambda}(\cdot), \mathcal{C}_{XY} v(\cdot) \rangle_{L^2}}{\langle \bar{\lambda}(\cdot) - \mathcal{H}^\top v(\cdot), \mathcal{C}_X(\bar{\lambda}(\cdot) - \mathcal{H}^\top v(\cdot)) \rangle_{L^2} + \langle v(\cdot), \mathcal{R}v(\cdot) \rangle_{L^2}}}.$$

Another approach would have been to consider a Bayesian posterior estimate when considering jointly Gaussian processes X, Y. However, as recalled in [17, p.9 and 16], Bayes' rule is more involved in infinite dimensions and the likelihood may be degenerate as, for $\mathcal{C}_X$ non-invertible, X does not have a density w.r.t. the Lebesgue measure. However, for invertible joint covariance $\mathcal{C}_{X,Y}$, we have that $\mathcal{C}_Y$ and $\mathcal{C}_\epsilon = \mathcal{C}_X - \mathcal{C}_{XY}\mathcal{C}_Y^{-1}\mathcal{C}_{YX}$ are invertible. Nevertheless, for continuous-time observations, the absence of a Lebesgue measure requires extensive care [10], leading to a less common viewpoint on reproducing kernels in relation with Gaussian measures and Cameron-Martin spaces [22, Definition 2.3.4]. We bypass these difficulties by expressing the maximum log-likelihood estimator for a realization $y(\cdot)$ of Y as its resulting least

---

[2]Note also that the theory of Gaussian processes (GPs), despite being a subfield of Bayesian statistics, does not require to manipulate a proper likelihood function to solve a regression problem. Uncannily we can draw another parallel with linear-quadratic (LQ) optimal control, where introducing the value function is not necessary to derive the solution. In a nutshell, the GP (resp. kernel) approach to estimation (resp. control) in the Gaussian (resp. LQ) case does not involve a larger space of functions to work in, but restricts the analysis to the space generated by the process (resp. RKHS). In our context, we will see that one can formally move from stochastic Bayesian GPs to deterministic frequentist kernels by replacing the Brownian noise $dw(t)$ by a control $u(t)dt$. This formal change is related to moving from Itô to Stratonovitch calculus and allows to derive exact properties of the stochastic system such as set invariance based on the deterministic counterpart [11], introducing a Stratonovitch drift in case of nonlinear diffusion terms.

squares problem[3]

$$\text{(LSE)} \quad \min_{\hat{x}(\cdot)\in L^2(\mathscr{T},\mathbb{R}^n)} \left\langle (\hat{x}(\cdot), y(\cdot)), \mathcal{C}_{X,Y}^{-1}(\hat{x}(\cdot), y(\cdot)) \right\rangle_{L^2(\Omega\times\mathscr{T},\mathbb{R}^n\times\mathbb{R}^m)}$$

$$= \underbrace{\left\langle \hat{x}(\cdot) - \mathcal{C}_{XY}\mathcal{C}_Y^{-1}y(\cdot), \mathcal{C}_\epsilon^{-1}(\hat{x}(\cdot) - \mathcal{C}_{XY}\mathcal{C}_Y^{-1}y(\cdot)) \right\rangle_{L^2} + \left\langle y(\cdot), \mathcal{C}_Y^{-1}y(\cdot) \right\rangle_{L^2}}_{\left\langle \hat{x}(\cdot), \mathcal{C}_X^{-1}\hat{x}(\cdot) \right\rangle_{L^2} + \left\langle y(\cdot) - \mathcal{H}\hat{x}(\cdot), \mathcal{R}^{-1}(y(\cdot) - \mathcal{H}\hat{x}(\cdot)) \right\rangle_{L^2}}.$$

Remark 2.1 (Stochastic and deterministic dual or equivalent problems)**.** [14, Chapter 15, Table 15.1, p568] describes the four problems we obtained as either "dual" or "equivalent". Here we intend duality strictly in the Fenchel context, obtained by permutation of max and min. The problems (MVLE)-(MVLE-dual) are indeed stochastic dual problems. That (LSE)-(MVLE-det) are deterministic dual problems as claimed in [5, p59] is not so straightforward (this is justified through dual bases in [14]). We will return to this in Section 4.4 and prove that when expressed on RKHSs, there is indeed a Fenchel duality. Similarly, that (MVLE)-(LSE) and (MVLE-dual)-(MVLE-det) are "equivalent" stochastic and deterministic problems, as defined in [14, Section 15.3], will get clearer from the fact that both problems share the same reproducing kernel (see Section 4). Note that (LSE)-(MVLE-det) do not have the same assumptions concerning the invertibility of $\mathcal{C}_{X,Y}$ and of $\mathcal{C}_Y$, a limitation which we will not meet when using kernels.

The problem with the approaches developed so far is that manipulating abstract covariances and continuous linear maps over L^2 can be cumbersome and is not operable for non-discrete time. Instead, Laurent Schwartz's kernel theorem allows us to consider all these operators as kernel integral operators. Thus, in the following, we will see formulas very similar to the ones derived in this section, but written explicitly on the time axis and with recursive versions in the spirit of Kalman filtering. Formally one can consider in (MVLE-det) the test function $\bar{\lambda}(\cdot) = \delta_s(\cdot)\bar{\lambda}$ for some $\bar{\lambda} \in \mathbb{R}^{n,*}$ to recover (2.11), but we will see that choosing $\bar{\lambda}(\cdot)$ based on the adjoint process $\hat{\Lambda}$ will allow us to avoid having to manipulate $\mathcal{C}_X$ in (MVLE-det).

2.4. **A related optimal control problem.** We return to the formulas of Section 2.2 and start by expressing $\left\langle \Gamma_S(s|T)\bar{\lambda}, \bar{\lambda} \right\rangle$ more explicitly,

$$\left\langle \bar{\lambda}, \epsilon_{\hat{S}_s}(s|T) \right\rangle = \left\langle \bar{\lambda}, \widetilde{x}(s) \right\rangle - \int_{t_0}^T \left\langle S_s^*(t|T)\bar{\lambda}, d\widetilde{y}(t) \right\rangle$$

$$= \left\langle \bar{\lambda}, \widetilde{x}(s) \right\rangle - \int_{t_0}^T \left\langle S_s^*(t|T)\bar{\lambda}, H(t)\widetilde{x}(t)dt + db(t) \right\rangle.$$

---

[3]In continuous time, $\tilde{y}(t) \in L^2(\mathscr{T},\mathbb{R}^m)$ "is reminiscent of the observation process, in fact rather the derivative of the observation process (which, as we know, does not exist)" [5, p180]. It is as if we claim to observe the derivative, but $\tilde{y}(t)$ will always appear within integrals.

We seek an expression where $\widetilde{x}$ does not appear. We thus introduce the following adjoint equation over an adjoint variable λ_s, a.k.a. the information vector,

$$(2.13) \qquad\qquad -\frac{d\lambda_s}{dt} = F^*(t)\lambda_s(t) - H^*(t)S_s^*(t|T)\bar{\lambda},$$

$$(2.14) \qquad \lambda_s(T) = \left| \begin{array}{ll} 0 & \text{if } s < T \\ \bar{\lambda} & \text{if } s = T \end{array} \right. , \quad \lambda_s(s) - \lambda_s(s^+) = \bar{\lambda}, \text{if } s < T,$$

where $\lambda_s(s^+) = \lim_{h \to s^+} \lambda_s(h)$ with the convention $\lambda_s(T^+) = 0$. Then a simple calculation shows that

$$\left\langle \bar{\lambda}, \epsilon_{\hat{S}_s}(s|T) \right\rangle = \left\langle \bar{\lambda}, \widetilde{x}(s) \right\rangle + \int_{t_0}^{T} \left\langle -\frac{d\lambda}{dt} - F^*(t)\lambda_s(t), \widetilde{x}(t)dt \right\rangle$$

$$- \int_{t_0}^{T} \left\langle S_s^*(t|T)\bar{\lambda}, db(t) \right\rangle$$

$$= \left\langle \lambda_s(s) - \lambda_s(s^+), \widetilde{x}(s) \right\rangle + \int_{t_0}^{T} \left\langle -\frac{d\lambda}{dt}, \widetilde{x}(t)dt \right\rangle$$

$$+ \int_{t_0}^{T} \left\langle \lambda_s(t), -d\widetilde{x}(t) + G(t)dw(t) \right\rangle - \int_{t_0}^{T} \left\langle S_s^*(t|T)\bar{\lambda}, db(t) \right\rangle$$

$$= \underbrace{\left\langle \lambda_s(s) - \lambda_s(s^+), \widetilde{x}(s) \right\rangle - [\langle \lambda_s(\tau), \widetilde{x}(\tau) \rangle]_{t_0}^{s} - [\langle \lambda_s(\tau), \widetilde{x}(\tau) \rangle]_{s^+}^{T}}_{=\langle \lambda_s(t_0), \xi \rangle}$$

$$+ \int_{t_0}^{T} \langle G^*(t)\lambda_s(t), dw(t) \rangle - \int_{t_0}^{T} \left\langle S_s^*(t|T)\bar{\lambda}, db(t) \right\rangle$$

and therefore

$$(2.15) \quad \left\langle \Gamma_S(s|T)\bar{\lambda}, \bar{\lambda} \right\rangle = \langle \Pi_0 \lambda_s(t_0), \lambda_s(t_0) \rangle + \int_{t_0}^{T} \langle G(t)Q(t)G^*(t)\lambda_s(t), \lambda_s(t) \rangle \, dt$$

$$+ \int_{t_0}^{T} \left\langle R(t)S_s^*(t|T)\bar{\lambda}, S_s^*(t|T)\bar{\lambda} \right\rangle dt$$

More generally, beyond linear feedbacks $(S_s^*(\cdot|T)\bar{\lambda})$, for a general control input $v(\cdot)$ with $R(\cdot)^{\frac{1}{2}}v(\cdot) \in L^2(t_0, T; \mathbb{R}^m)$, it is natural to consider the following control problem extending (2.14) and corresponding to (MVLE-det)

$$(2.16) \qquad\qquad -\frac{d\lambda_s}{dt} = F^*(t)\lambda_s(t) + H^*(t)v(t),$$

$$\lambda_s(T) = \left| \begin{array}{ll} 0 & \text{if } s < T \\ \bar{\lambda} & \text{if } s = T \end{array} \right. , \quad \lambda_s(s) - \lambda_s(s^+) = \bar{\lambda}, \text{if } s < T;$$

$$(2.17) \qquad J(v(\cdot)) = \langle \Pi_0 \lambda_s(t_0), \lambda_s(t_0) \rangle + \int_{t_0}^{T} \langle G(t)Q(t)G^*(t)\lambda_s(t), \lambda_s(t) \rangle \, dt$$

$$+ \int_{t_0}^{T} \langle R(t)v(t), v(t) \rangle \, dt.$$

The classical way to solve such a problem is through Pontryagin's Maximum Principle [see e.g. 5, Chapter 10], i.e. by using Lagrange-Fenchel duality, which effectively

leads to a Hamiltonian system, taking the form of a two-point boundary value problem:

$$(2.18) \quad \frac{d\hat{\gamma}_s}{dt} = F(t)\hat{\gamma}_s(t) - G(t)Q(t)G^*(t)\hat{\lambda}_s(t)$$

$$-\frac{d\hat{\lambda}_s}{dt} = F^*(t)\hat{\lambda}_s(t) + H^*(t)R^{-1}(t)H(t)\hat{\gamma}_s(t)$$

$$\hat{\gamma}_s(t_0) = -\Pi_0\hat{\lambda}_s(t_0), \ \lambda_s(T) = \left| \begin{array}{ll} 0 & \text{if } s < T \\ \bar{\lambda} & \text{if } s = T \end{array} \right. , \ \lambda_s(s) - \lambda_s(s^+) = \bar{\lambda}, \text{if } s < T.$$

and where the optimal control of problem (2.4)-(2.17) is given by

$$(2.19) \qquad\qquad \hat{v}_s(t) = R^{-1}(t)H(t)\hat{\gamma}_s(t).$$

Since the pair $(\hat{\lambda}_s(t), \hat{\gamma}_s(t))$ depends linearly on $\bar{\lambda}$, we obtain immediately that there exists a single $\widehat{S}_s(\cdot|T)$ which minimizes (2.15) for any $\bar{\lambda}$, namely

$$(2.20) \qquad\qquad \widehat{S}_s^*(t|T)\bar{\lambda} = -R^{-1}(t)H(t)\hat{\gamma}_s(t).$$

This suggests that one has to solve (2.18) to get an expression of $\widehat{S}^*(s,t|T)$.

2.5. **A general two-point boundary value problem.** At this stage, rather than coping with the difficulty of dealing with jumps as in (2.18), we consider another two-point boundary value problem similar to (2.18) but which crucially does not depend on s. We also had not so far considered the possibility of having a Gaussian prior with covariance $\Sigma_T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^{n,*})$ on the terminal information $\lambda_s(T)$, so we introduce it here for greater generality. This term will act a special weight on the terminal point $x(T)$ in relation with a terminal cost in optimal control. Consider the pair $(\hat{\mu}(t), \hat{\nu}(t))$ solution of the coupled system

$$(2.21) \qquad \frac{d\hat{\mu}}{dt} = F(t)\hat{\mu}(t) - G(t)Q(t)G^*(t)\hat{\nu}(t) + l_\mu(t)$$

$$-\frac{d\hat{\nu}}{dt} = F^*(t)\hat{\nu}(t) + H^*(t)R^{-1}(t)H(t)\hat{\mu}(t) - l_\nu(t)$$

$$\hat{\mu}(t_0) = -\Pi_0\hat{\nu}(t_0), \ \hat{\nu}(T) = \Sigma_T\hat{\mu}(T).$$

where $\hat{\mu}$ (resp. $\hat{\nu}$) plays the role of γ_s (resp. λ_s) and $l_\mu(\cdot) \in L^2(t_0, T; \mathbb{R}^n)$ and $l_\nu(\cdot) \in L^2(t_0, T; \mathbb{R}^{n,*})$ are two test functions acting as perturbations of the differential equations. Notice that, for $l_\mu(\cdot) \equiv 0$ and $l_{\nu,s}(\tau) = -\bar{\lambda}\delta_s(\tau)$, the system (2.18) with a jump condition corresponds precisely to the two-point boundary system (2.21). The important result is the following

Proposition 2.2. *For $\Sigma_T = 0$, $l_\mu(\cdot) \equiv 0$ and $l_\nu(t) = H^*(t)R^{-1}(t)g(t)$ with $g(\cdot) \in L^2(t_0, T; \mathbb{R}^m)$, we have the formula*

$$(2.22) \qquad\qquad \hat{\mu}(s) = \int_{t_0}^T \hat{S}_s(t|T)g(t)dt.$$

Proof. From (2.20) we have

$$\left\langle \bar{\lambda}, \int_{t_0}^T \widehat{S}_s(t|T)g(t)dt \right\rangle = -\int_{t_0}^T \left\langle \hat{\gamma}_s(t), H^*(t)R^{-1}(t)g(t) \right\rangle dt.$$

We use the second equation of (2.21) and proceed with integration by parts between the systems (2.18) and (2.21). Note that $\hat{\mu}$, $\hat{\nu}$, $\hat{\gamma}_s$ are all continuous functions, but that $\hat{\lambda}_s$ is not, so that we have to split some integrals to take care of the singularity at s as we did before to derive (2.15).

$$
\left\langle \bar{\lambda}, \int_{t_0}^T \widehat{S}_s(t|T)g(t)dt \right\rangle
$$

$$
= \int_{t_0}^T \left\langle \hat{\gamma}_s(t), -\frac{d\hat{\nu}}{dt} - F^*(t)\hat{\nu}(t) - H^*(t)R^{-1}(t)H(t)\hat{\mu}(t) \right\rangle dt
$$

$$
= \int_{t_0}^T \left\langle \hat{\gamma}_s(t), -\frac{d\hat{\nu}}{dt} \right\rangle dt - \int_{t_0}^T \left\langle \frac{d\hat{\gamma}_s}{dt} + G(t)Q(t)G^*(t)\hat{\lambda}_s(t), \hat{\nu}(t) \right\rangle dt
$$

$$
+ \int_{t_0}^T \left\langle \frac{d\hat{\lambda}_s}{dt} + F^*(t)\hat{\lambda}_s(t), \hat{\mu}(t) \right\rangle dt
$$

$$
= -[\langle \hat{\gamma}_s(t), \hat{\nu}(t) \rangle]_{t_0}^T + \int_{t_0}^T \left\langle \hat{\lambda}_s(t), \frac{d\hat{\mu}}{dt} - F(t)\hat{\mu}(t) \right\rangle dt
$$

$$
+ \int_{t_0}^T \left\langle \frac{d\hat{\lambda}_s}{dt} + F^*(t)\hat{\lambda}_s(t), \hat{\mu}(t) \right\rangle dt
$$

$$
= -\left\langle \hat{\gamma}_s(T), \underbrace{\hat{\nu}(T)}_{=\Sigma_T\hat{\mu}(T)=0} \right\rangle + \left\langle \underbrace{\hat{\gamma}_s(t_0)}_{=-\Pi_0\hat{\lambda}_s(t_0)}, \hat{\nu}(t_0) \right\rangle + \left\langle \underbrace{\hat{\lambda}_s(T)}_{=0}, \hat{\mu}(T) \right\rangle \mathbb{1}_{s<T}
$$

$$
- \left\langle \hat{\lambda}_s(s^+) - \hat{\lambda}_s(s), \hat{\mu}(s) \right\rangle - \left\langle \hat{\lambda}_s(t_0), \underbrace{\hat{\mu}(t_0)}_{=-\Pi_0\hat{\nu}(t_0)} \right\rangle
$$

$$
= \left\langle \hat{\lambda}_s(s) - \hat{\lambda}_s(s^+), \hat{\mu}(s) \right\rangle = \left\langle \bar{\lambda}, \hat{\mu}(s) \right\rangle.
$$

$\square$

To derive another formula for $\widehat{S}_s(t|T)$, we are now going to find another matrix satisfying (2.22) for all $g(\cdot)$ with $\hat{\mu}(\cdot)$ defined as per (2.21).

3. Solution through Riccati equations and kernels

3.1. **Deriving the kernels.** When $l_\mu(\cdot) \equiv 0$ and $l_\nu(\cdot) \equiv 0$, the classical approach to solve (2.21) is by variation of constants, introducing two matrices $\Sigma(t)$ and $\Pi(t)$ satisfying $\hat{\mu}(t) = -\Pi(t)\hat{\nu}(t)$ and $\hat{\nu}(t) = \Sigma(t)\hat{\mu}(t)$. It is then straightforward to show that they must satisfy two (dual) differential Riccati equations

(3.1)

$$
-\frac{d}{dt}\Sigma = \Sigma(t)F(t) + F^*(t)\Sigma(t) - \Sigma(t)G(t)Q(t)G^*(t)\Sigma(t) + H^*(t)R^{-1}(t)H(t),
$$
$$
\Sigma(T) = \Sigma_T;
$$

(3.2) $\quad \dfrac{d}{dt}\Pi = F(t)\Pi(t) + \Pi(t)F^*(t) - \Pi(t)H^*(t)R^{-1}(t)H(t)\Pi(t) + G(t)Q(t)G^*(t),$
$$
\Pi(t_0) = \Pi_0.
$$

The information filter matrix $\Sigma(t)$ satisfies a backward equation depending on T whereas the estimation filter matrix $\Pi(t)$ satisfies a forward one and depends on t_0. The solutions for these equations exist at all times on $[t_0, T]$. This is a classical result whose proof we give in Lemma 6.1 in Appendix A.1.[4]

When $l_\mu(\cdot) \not\equiv 0$ or $l_\nu(\cdot) \not\equiv 0$, we are going instead to follow a Green kernel approach, related to Wiener filtering, and look for kernel integral operators satisfying respectively:[5]

$$(3.3) \qquad \hat{\mu}(s) = \int_{t_0}^T K(s, t|T) l_\nu(t) dt \quad \text{for } l_\mu(\cdot) \equiv 0,$$

$$(3.4) \qquad \hat{\nu}(s) = \int_{t_0}^T \Lambda(s, t|T) l_\mu(t) dt \quad \text{for } l_\nu(\cdot) \equiv 0.$$

These formulas may recall the ones appearing in the innovations approach [14, Section 16.4.2]; however we do not consider cross-covariances between trajectories and observations. We will show (see Proposition 4.6 below) that K satisfying (3.3) is precisely the covariance of the optimal error (2.12). On the other hand Λ as in (3.4) is related to observability problems. To solve (2.21) in general, we can compute the distorsion w.r.t. the solutions with null perturbations, introducing two variables:

$$(3.5) \qquad r(t) = \hat{\mu}(t) + \Pi(t)\hat{\nu}(t), \quad \eta(t) = \hat{\nu}(t) - \Sigma(t)\hat{\mu}(t).$$

From (2.21), (3.1) and (3.2), we deduce that $r(\cdot)$ and $\eta(\cdot)$ satisfy the following differential equations[6]

$$(3.6) \quad \frac{d}{dt}r = (F(t) - \Pi(t)H^*(t)R^{-1}(t)H(t))r(t) + \Pi(t)l_\nu(t) + l_\mu(t), \qquad r(t_0) = 0;$$

$$(3.7)$$
$$-\frac{d}{dt}\eta = (F^*(t) - \Sigma(t)G(t)Q(t)G^*(t))\eta(t) + \Sigma(t)l_\mu(t) - l_\nu(t), \qquad \eta(T) = 0.$$

This suggests to introduce semigroups associated with the matrix function $F(t) - G(t)Q(t)G^*(t)\Sigma(t)$ denoted $\Phi_{F,\Sigma}(s, t)$ (resp. $F(s) - \Pi(s)H^*(s)R^{-1}(s)H(s)$, denoted

---

[4]If Π_0 is invertible, then so is $\Pi(t)$. Indeed the Riccati equation preserves positive definiteness [see e.g. 14, Example 16.3.4, p629]. Then $\Pi(t)^{-1}$ also satisfies (3.1), and, if we choose $\Sigma_T = \Pi(T)^{-1}$, then we can identify the matrices $\Sigma(t)$ and $\Pi(t)^{-1}$.

[5]An alternative scheme would be to consider exponentials of the Hamiltonian matrix as in [30, Chapter 5]. We use and recall this approach in the Appendix A.2 to give a numerical method to evaluate the kernel $K(s, t|T)$.

[6]We give below the computation for (3.6), eq.(3.7) is obtained similarly

$dr/dt = F\hat{\mu} - GQG^*\hat{\nu} + l_\mu + (F\Pi + \Pi F^* - \Pi H^*R^{-1}H\Pi + GQG^*)\hat{\nu} - \Pi(F^*\hat{\nu} + H^*R^{-1}H\hat{\mu} - l_\nu)$.

$\Phi_{F,\Pi}(s,t))$

$$\frac{d}{d\tau}\Phi_{F,\Sigma}(\tau,t) = (F(\tau) - G(\tau)Q(\tau)G^*(\tau)\Sigma(\tau))\Phi_{F,\Sigma}(\tau,t), \qquad \Phi_{F,\Sigma}(t,t) = \mathrm{Id};$$

$$-\frac{d}{dt}\Phi^*_{F,\Sigma}(\tau,t) = (F^*(t) - \Sigma(t)G(t)Q(t)G^*(t))\Phi^*_{F,\Sigma}(\tau,t), \qquad \Phi_{F,\Sigma}(\tau,\tau) = \mathrm{Id};$$

$$\frac{d}{d\tau}\Phi_{F,\Pi}(\tau,t) = (F(\tau) - \Pi(\tau)H^*(\tau)R^{-1}(\tau)H(\tau))\Phi_{F,\Pi}(\tau,t), \quad \Phi_{F,\Pi}(t,t) = \mathrm{Id};$$

$$-\frac{d}{dt}\Phi^*_{F,\Pi}(\tau,t) = (F^*(t) - H^*(t)R^{-1}(t)H(t)\Pi(t))\Phi^*_{F,\Pi}(\tau,t), \qquad \Phi^*_{F,\Pi}(\tau,\tau) = \mathrm{Id}.$$

Note that if $Q(\cdot) \equiv 0$ and $\Pi_0 = 0$, then $\Pi(\cdot) \equiv 0$ and $\Phi_{F,\Pi}(t,s) = \Phi_F(t,s)$, the semi-group associated with the operator $F(t)$, i.e. $\partial_t\Phi_F(t,s) = F(t)\Phi_F(t,s)$ and $\Phi_F(t,t) = \mathrm{Id}$. Similarly, if $H(\cdot) \equiv 0$ and $\Sigma_T = 0$, then $\Phi_{F,\Sigma}(t,s) = \Phi_F(t,s)$. In the general case, we have

Theorem 3.1. *The kernels K and Λ satisfying (3.3) and (3.4) are given by*

$$(3.8) \quad K(s,t|T) = \Phi_{F,\Sigma}(s,t_0)\Pi_0^{\frac{1}{2}}(\mathrm{Id} + \Pi_0^{\frac{1}{2}}\Sigma(t_0)\Pi_0^{\frac{1}{2}})^{-1}\Pi_0^{\frac{1}{2}}\Phi^*_{F,\Sigma}(t,t_0)$$

$$+ \int_{t_0}^{\min(s,t)} \Phi_{F,\Sigma}(s,\tau)G(\tau)Q(\tau)G^*(\tau)\Phi^*_{F,\Sigma}(t,\tau)d\tau;$$

$$(3.9) \quad \Lambda(s,t|T) = \Phi^*_{F,\Pi}(T,s)\Sigma_T^{\frac{1}{2}}(\mathrm{Id} + \Sigma_T^{\frac{1}{2}}\Pi(T)\Sigma_T^{\frac{1}{2}})^{-1}\Sigma_T^{\frac{1}{2}}\Phi_{F,\Pi}(T,t)$$

$$+ \int_{\max(s,t)}^{T} \Phi^*_{F,\Pi}(\tau,s)H^*(\tau)R^{-1}(\tau)H(\tau)\Phi_{F,\Pi}(\tau,t)d\tau.$$

Remark. The matrix inverses we consider always exist since Π_0, Σ_T, $\Sigma(t_0)$ and $\Pi(T)$ are all positive semidefinite. Note that if Π_0 is invertible, then the first term in K boils down to $\Phi_{F,\Sigma}(s,t_0)(\Pi_0^{-1} + \Sigma(t_0))^{-1}\Phi^*_{F,\Sigma}(t,t_0)$. If $\Sigma_T = 0$, then the first term in Λ vanishes. If furthermore $H(\cdot) \equiv 0$, the observations are then pure Brownian noise independent from the state, and thus dispensable. We hence recover the formula of [28, Section 6.4, p.89] for the (prior) covariance of a linear SDE. In [14, Example 3.3], the case $G(\cdot) \equiv 0$ was further discussed (see also Section 4.2) as it makes the second term in K vanish. Both kernels satisfy a Hermitian symmetry, i.e. $K(s,t|T) = K(t,s|T)^*$. In Appendix A.2, we give a numerical method to evaluate the kernel $K(s,t|T)$ through the Hamiltonian semigroup. For time-invariant systems, this boils down to matrix exponentials.

Proof. Let us first identify K. By the variation of constants formula, (3.7) yields

$$(3.10) \qquad \hat{\nu}(t) - \Sigma(t)\hat{\mu}(t) = \eta(t) = \int_t^T \Phi^*_{F,\Sigma}(\tau,t)(\Sigma(\tau)l_\mu(\tau) - l_\nu(\tau))d\tau.$$

Consequently (2.21) becomes

$$(3.11) \quad \frac{d\hat{\mu}}{dt} = F(t)\hat{\mu}(t) - G(t)Q(t)G^*(t)\left(\Sigma(t)\hat{\mu}(t)\right.$$

$$\left. + \int_t^T \Phi^*_{F,\Sigma}(\tau,t)(\Sigma(\tau)l_\mu(\tau) - l_\nu(\tau))d\tau\right) + l_\mu(t).$$

Now consider the case where $l_\mu(\cdot) \equiv 0$. Introducing the square root $\Pi_0^{\frac{1}{2}}$, the variation of constants formula gives

(3.12)

$$
\hat{\mu}(s) = \Phi_{F,\Sigma}(s,t_0)\hat{\mu}(t_0) + \int_{t_0}^{s} \Phi_{F,\Sigma}(s,t)G(t)Q(t)G^*(t) \int_{t}^{T} \Phi_{F,\Sigma}^*(\tau,t)l_\nu(\tau)d\tau dt
$$

$$
= -\Phi_{F,\Sigma}(s,t_0)\Pi_0^{\frac{1}{2}}\Pi_0^{\frac{1}{2}}\hat{\nu}(t_0)
$$

$$
+ \int_{t_0}^{T} \int_{t_0}^{\min(s,t)} \Phi_{F,\Sigma}(s,\tau)G(\tau)Q(\tau)G^*(\tau)\Phi_{F,\Sigma}^*(t,\tau)l_\nu(t)d\tau dt.
$$

To compute $\hat{\nu}(t_0)$, we use (3.10)

(3.13) $\quad (\mathrm{Id} + \Pi_0^{\frac{1}{2}}\Sigma(t_0)\Pi_0^{\frac{1}{2}})\Pi_0^{\frac{1}{2}}\hat{\nu}(t_0) = \Pi_0^{\frac{1}{2}}(\hat{\nu}(t_0) + \Sigma(t_0)\Pi_0\hat{\nu}(t_0)) = \Pi_0^{\frac{1}{2}}\eta(t_0)$

$$
= -\Pi_0^{\frac{1}{2}} \int_0^T \Phi_{F,\Sigma}^*(\tau,t_0)l_\nu(\tau)d\tau.
$$

Inserting (3.13) in (3.12), we then identify K through (3.3) which yields (3.8). Now for Λ, the procedure is the same, (3.6) and (2.21) yield

(3.14) $\quad \hat{\mu}(t) \;+\; \Pi(t)\hat{\nu}(t) \;\;=\;\; r(t) \;\;=\;\; \int_{t_0}^{t} \Phi_{F,\Pi}(t,\tau)(\Pi(\tau)l_\nu(\tau) \;+\; l_\mu(\tau))d\tau,$

(3.15) $\quad -\dfrac{d\hat{\nu}}{dt} = F^*(t)\hat{\nu}(t) + H^*(t)R^{-1}(t)H(t)\left(-\Pi(t)\hat{\nu}(t)\right.$

$$
\left. + \int_{t_0}^{t} \Phi_{F,\Pi}(t,\tau)(\Pi(\tau)l_\nu(\tau) + l_\mu(\tau))d\tau \right) - l_\nu(t).
$$

Now consider the case where $l_\nu(\cdot) \equiv 0$,

(3.16) $\qquad \hat{\nu}(s) = \Phi_{F,\Pi}^*(T,s)\hat{\nu}(T)$

$$
+ \int_{t}^{T} \Phi_{F,\Pi}^*(t,s)H^*(t)R^{-1}(t)H(t) \int_{t_0}^{t} \Phi_{F,\Pi}(t,\tau)l_\mu(\tau)d\tau dt
$$

$$
= \Phi_{F,\Pi}^*(T,s)\Sigma_T^{\frac{1}{2}}\Sigma_T^{\frac{1}{2}}\hat{\mu}(T)
$$

$$
+ \int_{t_0}^{T} \int_{\max(s,t)}^{T} \Phi_{F,\Pi}^*(\tau,s)H^*(\tau)R^{-1}(\tau)H(\tau)\Phi_{F,\Pi}(\tau,t)d\tau l_\mu(t)dt.
$$

To compute $\hat{\mu}(T)$, we use (3.14)

(3.17) $\quad (\mathrm{Id} + \Sigma_T^{\frac{1}{2}}\Pi(T)\Sigma_T^{\frac{1}{2}})\Sigma_T^{\frac{1}{2}}\hat{\mu}(T) = \Sigma_T^{\frac{1}{2}}(\hat{\mu}(T) + \Pi(T)\Pi_0\hat{\mu}(T)) = \Sigma_T^{\frac{1}{2}}r(T)$

$$
= -\Sigma_T^{\frac{1}{2}} \int_0^T \Phi_{F,\Pi}(\tau,t_0)l_\nu(\tau)d\tau.
$$

Inserting (3.17) in (3.16), we then identify Λ through (3.4) which yields (3.9).

$\square$

By Proposition 2.2, we have shown (2.22) which matches with (3.3) and should hold for all test functions $g(\cdot) \in L^2(t_0, T; \mathbb{R}^m)$, so we deduce immediately that:

Corollary 3.2. *If $\Sigma_T = 0$, then, for all $s, t \in [t_0, T]$, we have the formula*

$$(3.18) \qquad \hat{S}_s(t|T) = K(s, t|T)H^*(t)R^{-1}(t).$$

3.2. **Control interpretation as impulse response.** A natural quantity to consider in (3.3) is the impulse-response for $l_\nu(\tau) = z\delta_t(\tau)$ and some $z \in \mathbb{R}^{n,*}$, which gives $x_{zt}(s) = K(s, t|T)z$. Now (2.21) then writes as

$$(3.19) \qquad \frac{d}{d\tau}x_{zt}(\tau) = F(\tau)x_{zt}(\tau) - G(\tau)Q(\tau)G^*(\tau)\nu_{zt}(\tau)$$

$$-\frac{d}{d\tau}\nu_{zt}(\tau) = F^*(\tau)\nu_{zt}(\tau) + H^*(\tau)R^{-1}(\tau)H(\tau)x_{zt}(\tau) - z\delta_t(\tau)$$

$$x_{zt}(t_0) = -\Pi_0\nu_{zt}(t_0), \ \nu_{zt}(T) = \Sigma_T x_{zt}(T),$$

where the discontinuity occurs in the second line rather than the first. This, we are going to reverse through a change of variables. Considering the function $\chi_{zt}(\tau)$ solution of

$$(3.20) \qquad -\frac{d}{d\tau}\chi_{zt}(\tau) = F^*(\tau)\chi_{zt}(\tau), \tau < t, \quad \chi_{zt}(t) = z$$

we can write formally that $\chi_{zt}(\tau)\mathbb{1}_{\tau<t}$ is solution of

$$(3.21) \qquad -\frac{d}{d\tau}(\chi_{zt}(\tau)\mathbb{1}_{\tau<t}) = F^*(\tau)(\chi_{zt}(\tau)\mathbb{1}_{\tau<t}) + z\delta_t(\tau), \quad \chi_{zt}(T)\mathbb{1}_{T<t} = 0$$

With $l_\nu(\tau) = z\delta_t(\tau)$, we define $q_{zt}(\tau) = \nu_{zt}(\tau) + \chi_{zt}(\tau)\mathbb{1}_{\tau<t}$, then from (3.19) the pair $x_{zt}(\tau), q_{zt}(\tau)$ is solution of the system

(3.22)

$$\frac{d}{d\tau}x_{zt}(\tau) = F(\tau)x_{zt}(\tau) - G(\tau)Q(\tau)G^*(\tau)q_{zt}(\tau) + G(\tau)Q(\tau)G^*(\tau)\chi_{zt}(\tau)\mathbb{1}_{\tau<t}$$

$$-\frac{d}{d\tau}q_{zt}(\tau) = F^*(\tau)q_{zt}(\tau) + H^*(\tau)R^{-1}(\tau)H(\tau)x_{zt}(\tau)$$

$$x_{zt}(t_0) = -\Pi_0 q_{zt}(t_0) + \Pi_0\chi_{zt}(t_0), \ q_{zt}(T) = \Sigma_T x_{zt}(T).$$

Similarly to (2.17), we interpret the system (3.22) as the necessary and sufficient optimality condition of the following control problem

$$(3.23) \quad \frac{d}{d\tau}\zeta(\tau) = F(\tau)\zeta(\tau) + G(\tau)Q^{\frac{1}{2}}(\tau)u(\tau) + G(\tau)Q(\tau)G^*(\tau)\chi_{zt}(\tau)\mathbb{1}_{\tau<t},$$

$$\zeta(t_0) = \Pi_0^{\frac{1}{2}}\xi + \Pi_0\chi_{zt}(t_0)$$

in which $u(\cdot)$ and ξ are controls, and the objective to minimize is

$$(3.24) \quad J_x(\xi, u(\cdot)) = \|\xi\|^2 + \langle\Sigma_T\zeta(T), \zeta(T)\rangle + \int_{t_0}^{T}\|u(\tau)\|^2 d\tau$$

$$+ \int_{t_0}^{T}\left\langle H^*(\tau)R^{-1}(\tau)H(\tau)\zeta(\tau), \zeta(\tau)\right\rangle d\tau$$

If we call $x_{zt}(\tau)$ the optimal state and define $q_{zt}(\tau)$ as in (3.22) we obtain by standard methods that the optimal controls $\hat{\xi}$ and $\hat{v}(\cdot)$ are given by

$$(3.25) \qquad \hat{\xi} = -\Pi_0^{\frac{1}{2}}q_{zt}(t_0), \ \hat{v}(\tau) = -Q^{\frac{1}{2}}(\tau)G^*(\tau)q_{zt}(\tau)$$

and the pair $x_{zt}(\tau), q_{zt}(\tau)$ solves (3.22).

4. Ientifying the RKHS

We can draw a parallel between the above reasoning of Section 3.2 and going from a partial differential equation (PDE) to its calculus of variations formulation. Similarly, to solve a linear PDE one can look for its Green kernel, a.k.a. its fundamental solution. This is precisely what we have done in (2.5) for (2.21). Now we will take advantage of (3.24) to identify the spaces of functions we are optimizing over. This is the same thing as identifying the reproducing kernel Hilbert spaces of the kernels of Theorem 3.1. In general it is a hard task given a PDE to find its Green kernel and the vector space of functions where the solutions live [27, Chapter 1]. But this task will be made much simpler owing to the fact that we are not looking at any Hilbert space, but at an RKHS:

Definition 4.1. Let $\mathscr{T}$ be a non-empty set. A Hilbert space $(\mathcal{H}_K(\mathscr{T}), \langle \cdot, \cdot \rangle_K)$ of $\mathbb{R}^n$-vector-valued functions defined on $\mathscr{T}$ is called a vRKHS if there exists a matrix-valued kernel $K_{\mathscr{T}} : \mathscr{T} \times \mathscr{T} \to \mathcal{L}(\mathbb{R}^{n,*}, \mathbb{R}^n)$ such that the *reproducing property* holds: for all $t \in \mathscr{T}$, $\mathbf{p} \in \mathbb{R}^{n,*}$, we have $K_{\mathscr{T}}(\cdot, t)\mathbf{p} \in \mathcal{H}_K(\mathscr{T})$ and for all $\mathbf{f} \in \mathcal{H}_K(\mathscr{T})$, $\langle \mathbf{p}, \mathbf{f}(t) \rangle = \langle \mathbf{f}, K_{\mathscr{T}}(\cdot, t)\mathbf{p} \rangle_K$.

Remark: It is well-known that by Riesz's theorem, an equivalent definition of a vRKHS is that, for every $t \in \mathscr{T}$ and $\mathbf{p} \in \mathbb{R}^n$, the evaluation functional $\mathbf{f} \in \mathcal{H}_K(\mathscr{T}) \mapsto \langle \mathbf{p}, \mathbf{f}(t) \rangle \in \mathbb{R}$ is continuous. There is also a one-to-one correspondence between the kernel $K_{\mathscr{T}}$ and the vRKHS $(\mathcal{H}_K(\mathscr{T}), \langle \cdot, \cdot \rangle_K)$ [see e.g. 8, Proposition 2.3]. Moreover, by symmetry of the scalar product, the matrix-valued kernel has a Hermitian symmetry, i.e. $K_{\mathscr{T}}(s, t) = K_{\mathscr{T}}(t, s)^*$ for any $s, t \in \mathscr{T}$. We refer to [9] and references therein for more on this topic of operator-valued kernels.

4.1. K**: A primal RKHS of trajectories.** We want to show that with the kernel $K(s, t | T)$ defined in (3.8),

$$(4.1) \quad K(s, t | T) = \Phi_{F,\Sigma}(s, t_0)\Pi_0^{\frac{1}{2}}(\mathrm{Id} + \Pi_0^{\frac{1}{2}}\Sigma(t_0)\Pi_0^{\frac{1}{2}})^{-1}\Pi_0^{\frac{1}{2}}\Phi_{F,\Sigma}^*(t, t_0)$$
$$+ \int_{t_0}^{\min(s,t)} \Phi_{F,\Sigma}(s, \tau)G(\tau)Q(\tau)G^*(\tau)\Phi_{F,\Sigma}^*(t, \tau)d\tau$$

we can associate a Hilbert space, for which it is the reproducing kernel. As intuited from (3.23), it is a space of functions in $H^1(t_0, T; \mathbb{R}^n)$ defined as follows

$$(4.2) \quad \mathcal{S}_{[t_0,T]}^x = \{x(\cdot) \in H^1(t_0, T; \mathbb{R}^n) | \exists\, u(\cdot) \in L^2(t_0, T; \mathbb{R}^p), \xi \in \mathbb{R}^n \text{ s.t.}$$
$$\frac{d}{d\tau}x = F(\tau)x(\tau) + G(\tau)Q^{\frac{1}{2}}(\tau)u(\tau), x(t_0) = \Pi_0^{\frac{1}{2}}\xi\}.$$

Recall that we only required $R(\cdot)$ to be invertible, improving over [3]. For a given trajectory $x(\cdot)$, in case there are several pairs $(u(\cdot), \xi)$ which satisfy the relations (4.2) we call the representative of $x(\cdot)$ the pair with minimal norm $\int_{t_0}^T \|u(t)\|^2 dt + \|\xi\|^2$. It is uniquely defined. We can identify this representative as the unique pair

$(u(\cdot), \xi)$ satisfying (4.2) for a given $x(\cdot)$ and

$$(4.3) \quad \int_{t_0}^{T} \langle u(s), \widetilde{u}(s) \rangle \, ds = 0, \left\langle \xi, \widetilde{\xi} \right\rangle = 0,$$

$$\forall \widetilde{u}(\cdot), \widetilde{\xi} \text{ such that } G(s) Q^{\frac{1}{2}}(s) \widetilde{u}(\cdot) = 0, \text{for a.e. } s \in (t_0, T), \Pi_0^{\frac{1}{2}} \widetilde{\xi} = 0$$

The vector space $\mathcal{S}^x_{[t_0, T]}$ is a subspace of the Sobolev space $H^1(t_0, T; \mathbb{R}^n)$. We equip $\mathcal{S}^x_{[t_0, T]}$ with a scalar product derived from (3.24)

$$(4.4) \quad \left\langle x^1(\cdot), x^2(\cdot) \right\rangle_{\mathcal{S}^x_{[t_0, T]}} = \left\langle \xi^1, \xi^2 \right\rangle + \left\langle \Sigma_T x^1(T), x^2(T) \right\rangle + \int_{t_0}^{T} \left\langle u^1(s), u^2(s) \right\rangle ds$$

$$+ \int_{t_0}^{T} \left\langle H^*(s) R^{-1}(s) H(s) x^1(s), x^2(s) \right\rangle ds$$

where $(\xi^1, u^1(\cdot))$; $(\xi^2, u^2(\cdot))$ are the representatives of $x^1(\cdot), x^2(\cdot)$ respectively.

Theorem 4.2. *The vector space* $(\mathcal{S}^x_{[t_0, T]}, \langle \cdot, \cdot \rangle_{\mathcal{S}^x_{[t_0, T]}})$ *is a reproducing kernel Hilbert space, and* $K(s, t | T)$ *defined by (3.8) is its corresponding reproducing kernel.*

Proof. The vector space $(\mathcal{S}^x_{[t_0, T]}, \langle \cdot, \cdot \rangle_{\mathcal{S}^x_{[t_0, T]}})$ is clearly a pre-complete Hilbert space by bilinearity of its inner product. It is furthermore complete since every Cauchy sequence converges within. Indeed, consider a Cauchy sequence $(x^k(\cdot))_{k \in \mathbb{N}}$ in $\mathcal{S}^x_{[t_0, T]}$, namely one for which there exists $(u^k(\cdot), \xi^k)$, which is the representative of $x^k(\cdot)$, satisfying

$$\frac{d}{ds} x^k = F(s) x^k(s) + G(s) Q^{\frac{1}{2}}(s) u^k(s), \quad x^k(t_0) = \Pi_0^{\frac{1}{2}} \xi^k$$

and

$$\|\xi^k - \xi^l\|^2 + \|\Sigma_T^{\frac{1}{2}}(x^k(T) - x^l(T))\|^2 + \int_{t_0}^{T} \|u^k(s) - u^l(s)\|^2 ds$$

$$+ \int_{t_0}^{T} \left\langle H^*(s) R^{-1}(s) H(s)(x^k(s) - x^l(s)), (x^k(s) - x^l(s)) \right\rangle ds \xrightarrow[k,l \to +\infty]{} 0.$$

Consequently $\xi^k, u^k(\cdot)$ are Cauchy sequences in $\mathbb{R}^n, L^2(t_0, T; \mathbb{R}^m)$ respectively, so $\xi^k \to \xi, u^k(\cdot) \to u(\cdot)$ and necessarily $x^k(\cdot) \to x(\cdot)$ in $H^1(t_0, T; \mathbb{R}^n)$ with

$$\frac{d}{ds} x = F(s) x(s) + G(s) Q^{\frac{1}{2}}(s) u(s), x(t_0) = \Pi_0^{\frac{1}{2}} \xi.$$

Since $(\xi^k, u^k(\cdot))$ are the representatives of $x^k(\cdot)$ they satisfy $\int_{t_0}^{T} \left\langle u^k(s), \widetilde{u}(s) \right\rangle ds = 0$, $\left\langle \xi^k, \widetilde{\xi} \right\rangle = 0$ for any pair $(\widetilde{u}(\cdot), \widetilde{\xi})$ such that $G(s) Q^{\frac{1}{2}}(s) \widetilde{u}(s) = 0$, a.e. $s \in (t_0, T), \Pi_0^{\frac{1}{2}} \widetilde{\xi} = 0$. But then $(u(\cdot), \xi)$ satisfies also $\int_{t_0}^{T} \langle u(s), \widetilde{u}(s) \rangle ds = 0, \left\langle \xi, \widetilde{\xi} \right\rangle = 0$, which implies that $(u(\cdot), \xi)$ is the representative of $x(\cdot)$. Hence $x^n(\cdot) \to x(\cdot)$ in the sense of the norm of $\mathcal{S}^x_{[t_0, T]}$ and $\mathcal{S}^x_{[t_0, T]}$ is a Hilbert space. We now want to show that $\mathcal{S}^x_{[t_0, T]}$ is a RKHS and that its corresponding kernel is $K(s, t | T)$. We must check two facts as

per Definition 4.1

$$(4.5) \qquad\qquad K(\cdot, t|T)z \in \mathcal{S}^x_{[t_0, T]}$$

$$(4.6) \qquad \langle x(\cdot), K(\cdot, t|T)z \rangle_{\mathcal{S}^x_{[t_0, T]}} = \langle x(t), z \rangle, \forall x(\cdot) \in \mathcal{S}^x_{[t_0, T]}, t \in [t_0, T], z \in \mathbb{R}^{n,*}$$

By definition, $K(\cdot, t|T)z = x_{zt}(\cdot)$. Now, from formulas (3.22) $x_{zt}(\cdot)$ belongs to $\mathcal{S}^x_{[t_0, T]}$, with

$$(4.7) \quad v_{zt}(\tau) = -Q^{\frac{1}{2}}(\tau)G^*(\tau)q_{zt}(\tau) + Q^{\frac{1}{2}}(\tau)G^*(\tau)\chi_{zt}(\tau)\mathbb{1}_{\tau < t},$$

$$\xi_{zt} = -\Pi_0^{\frac{1}{2}}q_{zt}(t_0) + \Pi_0^{\frac{1}{2}}\chi_{zt}(t_0)$$

We claim that $(v_{zt}(\cdot), \xi_{zt})$ is the representative of $x_{zt}(\cdot)$. Indeed the pair $(v_{zt}(\cdot), \xi_{zt})$ satisfies (4.3) by using a transposition. It remains to check (4.6). Consider $x(\cdot) \in \mathcal{S}^x_{[t_0, T]}$, such that

$$\frac{d}{d\tau}x = F(\tau)x(\tau) + G(\tau)Q^{\frac{1}{2}}(\tau)u(\tau), x(t_0) = \Pi_0^{\frac{1}{2}}\xi$$

assuming that $(u(\cdot), \xi)$ is the representative. We have

$$\langle x(\cdot), x_{zt}(\cdot) \rangle_{\mathcal{S}^x_{[t_0, T]}} = \left\langle \xi, (-\Pi_0^{\frac{1}{2}}q_{zt}(t_0) + \Pi_0^{\frac{1}{2}}\chi_{zt}(t_0)) \right\rangle + \langle \Sigma_T x(T), x_{zt}(T) \rangle$$

$$+ \int_{t_0}^T \left\langle u(\tau), -Q^{\frac{1}{2}}(\tau)G^*(\tau)q_{zt}(\tau) + Q^{\frac{1}{2}}(\tau)G^*(\tau)\chi_{zt}(\tau)\mathbb{1}_{\tau < t} \right\rangle d\tau$$

$$+ \int_{t_0}^T \left\langle H^*(\tau)R^{-1}(\tau)H(\tau)x(\tau), x_{zt}(\tau) \right\rangle d\tau$$

Using relations (3.22) and the equation for $x(\cdot)$ with integration by parts, it is easy to conclude that

$$\langle x(\cdot), x_{zt}(\cdot) \rangle_{\mathcal{S}^x_{[t_0, T]}} = \langle x(t_0), \chi_{zt}(t_0) - q_{zt}(t_0) \rangle + \langle x(T), q_{zt}(T) \rangle$$

$$+ \int_{t_0}^T \left\langle \frac{d}{d\tau}x(\tau) - F(\tau)x(\tau), \chi_{zt}(\tau)\mathbb{1}_{\tau < t} - q_{zt}(\tau) \right\rangle d\tau$$

$$+ \int_{t_0}^T \left\langle x(\tau), -\frac{d}{d\tau}q_{zt}(\tau) - F^*(\tau)q_{zt}(\tau) \right\rangle d\tau$$

$$= [\langle x(\tau), q_{zt}(\tau) \rangle]_{\tau=T}^{t_0} + \langle x(T), q_{zt}(T) \rangle + \int_{t_0}^t \left\langle \frac{d}{d\tau}x(\tau), \chi_{zt}(\tau) \right\rangle d\tau$$

$$+ \int_{t_0}^t \left\langle x(\tau), \frac{d}{d\tau}\chi_{zt}(\tau) \right\rangle]d\tau$$

$$= \langle x(t), z \rangle$$

which proves (4.6) and concludes the proof of the theorem.

$\square$

Remark 4.3 (No measurements and Gramian of controllability)**.** In the simpler case where $\Sigma_T = 0$, $H(\cdot) \equiv 0$, we have that $\Sigma(\cdot) \equiv 0$. For $\Pi_0 = 0$ and $Q(\cdot) \equiv \mathrm{Id}$,

(3.24) becomes $J_x(\xi, u(\cdot)) = \int_{t_0}^T \|u(\tau)\|^2 d\tau$ and we obtain

$$K(T, T|T) = \int_{t_0}^T \Phi_F(T, \tau) G(\tau) Q(\tau) G^*(\tau) \Phi_F^*(T, \tau) d\tau$$

which is precisely the Gramian of controlability. Its straightforward relation with kernels for optimal control was further discussed in [1, p8].

4.2. Λ: A dual RKHS of information vectors.

We have seen with formula (2.20) that the optimal operator $\hat{S}_s(t|T)$ is directly related to an optimal control problem (2.4),(2.17). This problem concerns a backward evolution. Similarly, formula (2.22) has been related to a two-point boundary value problem (2.21). This one is easily interpreted as the necessary and sufficient conditions of optimality of another control problem, namely

$$(4.8) \qquad -\frac{d}{dt}\lambda(t) = F^*(t)\lambda(t) + H^*(t)v(t), \quad \lambda(T) = \Sigma_T^{\frac{1}{2}} z$$

and the payoff to minimize is a slight variation of (2.17),

$$(4.9) \quad J_\lambda(z, v(\cdot)) = \langle \Pi_0 \lambda(t_0), \lambda(t_0) \rangle + \|z\|^2 + \int_{t_0}^T \langle G(t)Q(t)G^*(t)\lambda(t), \lambda(t) \rangle \, dt$$
$$+ \int_{t_0}^T \langle R(t)v(t), v(t) \rangle \, dt.$$

Recall that we did not require Σ_T to be invertible but that $R(t)$ is invertible. These problems lead to the kernel

$$(4.10) \quad \Lambda(s, t|T) = \Phi_{F,\Pi}^*(T, s) \Sigma_T^{\frac{1}{2}} (\mathrm{Id} + \Sigma_T^{\frac{1}{2}} \Pi(T) \Sigma_T^{\frac{1}{2}})^{-1} \Sigma_T^{\frac{1}{2}} \Phi_{F,\Pi}(T, t)$$
$$+ \int_{\max(s,t)}^T \Phi_{F,\Pi}^*(\tau, s) H^*(\tau) R^{-1}(\tau) H(\tau) \Phi_{F,\Pi}(\tau, t) d\tau$$

and to a space of information vectors $\mathcal{S}_{[t_0,T]}^\lambda$

$$(4.11) \quad \mathcal{S}_{[t_0,T]}^\lambda = \{\lambda(\cdot) \in H^1(t_0, T; \mathbb{R}^n) | \, v(\cdot) \in L^2(t_0, T; \mathbb{R}^m), z \in \mathbb{R}^n \ \text{s.t.}$$
$$-\frac{d}{dt}\lambda(t) = F^*(t)\lambda(t) + H^*(t)v(t), \ \lambda(T) = \Sigma_T^{\frac{1}{2}} z\}.$$

Again if there are several $(z, v(\cdot))$ for the same trajectory $\lambda(\cdot)$, we define the representative of $\lambda(\cdot)$ as the unique $(z, v(\cdot))$ satisfying (4.11) which minimizes $\int_{t_0}^T \langle R(t)v(t), v(t) \rangle \, dt + \|z\|^2$ and, similarly to (4.3), we have

$$(4.12) \quad \int_{t_0}^T \langle R(s)v(s), \widetilde{v}(s) \rangle \, ds = 0, \langle z, \widetilde{z} \rangle = 0,$$

$$\forall \widetilde{v}(\cdot), \widetilde{z} \text{ such that } H^*(s)\widetilde{v}(\cdot) = 0, \text{for a.e. } s \in (t_0, T), \Sigma_T^{\frac{1}{2}} \widetilde{z} = 0.$$

We then equip $\mathcal{S}_{[t_0,T]}^\lambda$ with the inner product derived from the quadratic form (4.9).

Theorem 4.4. *The Hilbert space $(\mathcal{S}_{[t_0,T]}^\lambda, \langle \cdot, \cdot \rangle_{\mathcal{S}_{[t_0,T]}^\lambda})$ is a reproducing kernel Hilbert space and the corresponding kernel is $\Lambda(s, t|T)$.*

Proof. The proof is exactly the same as that of Theorem 4.2, noticing that the RKHS $\mathcal{S}^x_{[t_0,T]}$ corresponds to $\mathcal{S}^\lambda_{[t_0,T]}$ when considering the permutations $t \leftrightarrow T - (t - t_0)$, $F(t) \leftrightarrow F^*(t)$, $\Pi_0 \leftrightarrow \Sigma_T$, $G(t)Q^{\frac{1}{2}}(t) \leftrightarrow H^*(t)R^{-\frac{1}{2}}(t)$. $\square$

Remark 4.5 (No plant noise and Gramian of observability)**.** In the simpler case where $\Pi_0 = 0$, $G(\cdot) \equiv 0$, we have that $\Pi(\cdot) \equiv 0$ and $\mathcal{S}^x_{[t_0,T]}$ is finite-dimensional. For $\Sigma_T = 0$ and $R(\cdot) \equiv \mathrm{Id}$, (4.9) becomes $J_\lambda(z, v(\cdot)) = \int_{t_0}^T \|v(t)\|^2 dt$ and we obtain

$$\Lambda(t_0, t_0|T) = \int_{t_0}^T \Phi_F^*(\tau, t_0) H^*(\tau) H(\tau) \Phi_F(\tau, t_0) d\tau$$

which is precisely the Gramian of observability. This setting is discussed in [14, Example 16.3.3]. The covector $\lambda(\cdot)$, through its interpretation as a Lagrange multiplier as in Section 2.3, is related to a notion of sensitivity of the solution $\hat{x}(\cdot)$ to the "constraints" $(y_t)_{t \in [t_0,T]}$, and thus here to the ability to recover x_0 from the observations.

By Loéve's theorem [6, Theorem 27, p57], we know that $\Lambda(s, t)$ is the covariance of a stochastic process. It is clearly not forward Markovian. It is possible to introduce a backward Markovian Gaussian process which would allow to recover (MVLE-dual), with technicalities covered in [14, Appendix Chap. 16]. We leave this identification to future developments.

4.3. K **as error covariance, formulas relating** K **and** Λ**.** We can now prove the claimed result (2.12) stating that the kernel K corresponds to the covariance of the error. We can also relate through a simple formula the two kernels K and Λ.

Proposition 4.6. *Let* $\Sigma_T = 0$ *and recall that* $\epsilon_{\hat{S}_s}(s|T) = x(s) - \hat{x}(s|T)$, *then*

$$(4.13) \qquad\qquad K(s, t|T) = \mathbb{E}[\epsilon_{\hat{S}_s}(s|T)(\epsilon_{\hat{S}_t}(t|T))^*]$$

Moreover, we have, for all $s, t \in [t_0, T]$,

$$(4.14) \quad K(s, t|T) = \Pi(s)\Phi_{F,\Pi}^*(t, s)1\!\!1_{s \leq t} + \Phi_{F,\Pi}(s, t)\Pi(t)1\!\!1_{s > t} - \Pi(s)\Lambda(s, t|T)\Pi(t).$$

In particular $K(T, T|T) = \Pi(T)$ *and* $K(t, t|T) = \Pi(t) - \Pi(t)\Lambda(t, t|T)\Pi(t)$.

Equation (4.14) can be seen as an extension to $s \neq t$ of Bryson-Frazier formulas [14, Theorem 16.5.1]. For $s = t$, it gives that $\mathbb{E}[\epsilon_{\hat{S}_s}(s|T)(\epsilon_{\hat{S}_s}(s|T))^*] \preccurlyeq \Pi(s) = \mathbb{E}[\epsilon_{\hat{S}_s}(s|s)(\epsilon_{\hat{S}_s}(s|s))^*]$. In other words, observations on the interval $[s, T]$ allow to lower the variance of the error at time s. For $t = T$ and $\Sigma_T = 0$, $\Lambda(s, T|T) = 0$ and we recover the result of [14, Lemma 16.5.1].

Proof. Since for $l_\mu(\cdot) \equiv 0$ and $l_{\nu,s}(\tau) = -\bar{\lambda}\delta_s(\tau)$, the system (2.18) with a jump condition corresponds to the two-point boundary system (2.21). With the same computations that led to (2.15), using (2.20) then integration by parts, we obtain

that

$$\left\langle \mathbb{E}[\epsilon_{\hat{S}_s}(s|T)(\epsilon_{\hat{S}_s}(t|T))^*]\bar{\lambda}, \bar{\lambda} \right\rangle = \langle \Pi_0 \lambda_s(t_0), \lambda_t(t_0) \rangle$$

$$+ \int_{t_0}^{T} \langle G(\tau)Q(\tau)G^*(\tau)\lambda_s(\tau), \lambda_t(\tau) \rangle \, d\tau + \int_{t_0}^{T} \left\langle H^*(\tau)R^{-1}(\tau)H(\tau)\hat{\gamma}_s(\tau)\bar{\lambda}, \hat{\gamma}_t(\tau) \right\rangle d\tau$$

$$\overset{(2.18)}{=} -\langle \lambda_s(t_0), \hat{\gamma}_t(t_0) \rangle + \int_{t_0}^{T} \left\langle \lambda_s(\tau), F(\tau)\hat{\gamma}_s(\tau) - \frac{d\hat{\gamma}_s}{du} \right\rangle d\tau$$

$$- \int_{t_0}^{T} \left\langle \frac{d\hat{\lambda}_s}{du} + F^*(\tau)\hat{\lambda}_s(\tau), \hat{\gamma}_t(\tau) \right\rangle d\tau$$

$$= -\langle \lambda_s(t_0), \hat{\gamma}_t(t_0) \rangle - [\langle \lambda_s(\tau), \hat{\gamma}_t(\tau) \rangle]_{t_0}^{s} - [\langle \lambda_s(\tau), \hat{\gamma}_t(\tau) \rangle]_{s+}^{T}$$

$$\overset{(2.18)}{=} -\left\langle \bar{\lambda}, \hat{\gamma}_t(s) \right\rangle \overset{(3.3)}{=} \left\langle \bar{\lambda}, K(s,t|T)\bar{\lambda} \right\rangle.$$

This proves (4.13). For (4.14), define $r_t(\tau) = \hat{\gamma}_t(\tau) + \Pi(\tau)\lambda_t(\tau)$ as in (3.5). Then similarly to (3.14)-(3.15)

$$(4.15) \quad r_t(\tau) \quad = \quad -\int_{t_0}^{\tau} \Phi_{F,\Pi}(\tau,\sigma)(\Pi(\sigma)\bar{\lambda}\delta_t(\sigma))d\sigma \quad = \quad -\Phi_{F,\Pi}(\tau,t)\Pi(t)\mathbb{1}_{\tau>t}\bar{\lambda}.$$

$$(4.16) \quad -\frac{d}{d\tau}\lambda_t = (F^*(\tau) - H^*(\tau)R^{-1}(\tau)H(\tau))\lambda_t(\tau) + \bar{\lambda}\delta_t(\tau)$$
$$- H^*(\tau)R^{-1}(\tau)H(\tau)\Phi_{F,\Pi}(\tau,t)\Pi(t)\mathbb{1}_{\tau>t}\bar{\lambda}.$$

Consequently, by the variation of constants formula,

$$(4.17) \quad \lambda_t(s) = \Phi_{F,\Pi}^*(t,s)\bar{\lambda}\mathbb{1}_{s\le t}$$
$$- \int_{\max(s,t)}^{T} \Phi_{F,\Pi}^*(\tau,s)H^*(\tau)R^{-1}(\tau)H(\tau)\Phi_{F,\Pi}(\tau,t)\Pi(t)\bar{\lambda}d\tau.$$

Since $\left\langle \bar{\lambda}, K(s,t|T)\bar{\lambda} \right\rangle = -\left\langle \bar{\lambda}, \hat{\gamma}_t(s) \right\rangle = \left\langle \bar{\lambda}, \Pi(s)\lambda_t(s) - r_t(s) \right\rangle$, (4.15)-(4.17) yield (4.14).

$\square$

Remark 4.7 (Relation with Fisher information). From (4.13), we know that the kernel K is the optimal error covariance for linear estimators. The Fisher information matrix of estimating $(x_t)_{t\in[t_0,T]}$ given $(y_t)_{t\in[t_0,T]}$ is well-defined if the unknown $(x_t)_{t\in[t_0,T]}$ is deterministic. In that case, since $(y_t)_{t\in[t_0,T]}$ is a Gaussian process and the measurements are linear, the integral operator with kernel K thus saturates the Cramér-Rao inequality and coincides with the inverse of the Fisher information matrix. For random $(x_t)_{t\in[t_0,T]}$, defining properly the Bayesian posterior analogue of the Cramér-Rao lower bound [34] in this infinite-dimensional context would require an undue level of generality, as hinted at in Section 2.3. However, for a deterministic $(x_t)_{t\in[t_0,T]}$, we have $G(\cdot) \equiv 0$ and $\Pi_0 = 0$ as considered previously. The only parameter of $(x_t)_{t\in[t_0,T]}$ to identify is x_0, so the Fisher information matrix writes as $\int_{t_0}^{T} \frac{\partial \bar{y}(t)}{\partial x_0} \frac{\partial \bar{y}(t)}{\partial x_0}^* dt$ which through simple calculations is shown to be equal to the Gramian of observability (already noticed in [16, Remark (h)]). The fact

that the Fisher information defines a kernel was unveiled by [13]. Since the posterior Cramér-Rao lower bound of our estimation problem does not seem to have been written yet, much remains to be done in this direction of research to draw connections with information theory and with the score vector. Applications would already exist in sensor placement and D-optimal design.

4.4. **Dual deterministic problems on RKHSs.** So far, we identified the RKHSs associated with the two Green functions K and Λ. However the primary modern use of reproducing kernels is as convenient spaces for optimization problems such as the GP regression (GP-reg) highlighted earlier on. This is exemplified in the following formalization of (LSE). Since we do not require any invertibility except for $R(t)$, we will use pseudo-inverses w.r.t. the Euclidean norm for the other operators, such as $\Pi_0^\ominus$, but this does not change the formalism. For $\widetilde{x}(\cdot) \in \mathcal{S}_{[t_0,T]}^x$, after making the covariances explicit, (LSE) writes as the traditional least-squares problem:

$$(4.18) \quad L_x(\widetilde{x}(\cdot)) := \int_{t_0}^T \|\widetilde{y}(t) - H(t)\widetilde{x}(t)\|_{R(t)^{-1}}^2 dt$$

$$+ \|G(t)^\ominus \left(\frac{d}{dt}\widetilde{x} - F(t)\widetilde{x}(t)\right)\|_{Q(t)^\ominus}^2 dt + \left\langle \Pi_0^\ominus \widetilde{x}(t_0), \widetilde{x}(t_0) \right\rangle + \left\langle \Sigma_T \widetilde{x}(T), \widetilde{x}(T) \right\rangle$$

$$= \int_{t_0}^T \|\widetilde{y}(t) - H(t)\widetilde{x}(t)\|_{R(t)^{-1}}^2 dt + \|\widetilde{x}(\cdot)\|_{\mathcal{S}_{[t_0,T]}^x}^2 - \int_{t_0}^T \|H(t)\widetilde{x}(t)\|_{R(t)^{-1}}^2 dt,$$

where we refer to footnote 3 for the interpretation of $\widetilde{y}(t)$. The space $\mathcal{S}_{[t_0,T]}^x$ is actually the set of trajectories such that L_x is finite. The term in Σ_T can be seen as imposing a Gaussian prior on the terminal point.

Proposition 4.8. *For a realization $\widetilde{y}(\cdot)$, the zero-mean part of the solution to the smoothing problem is given by*

$$(4.19) \qquad \int_{t_0}^T K(\cdot, t|T) H^*(t) R^{-1}(t) \widetilde{y}(t) dt = \operatorname*{argmin}_{\widetilde{x}(\cdot) \in \mathcal{S}_{[t_0,T]}^x} L_x(\widetilde{x}(\cdot))$$

with $L_x(\widetilde{x}(\cdot)) = \|R(t)^{-1/2}\widetilde{y}(\cdot)\|_{L^2}^2 + \|\widetilde{x}(\cdot)\|_{\mathcal{S}_{[t_0,T]}^x}^2 - 2\left\langle H^*(\cdot)R^{-1}(\cdot)\widetilde{y}(\cdot), \widetilde{x}(\cdot)\right\rangle_{L^2([t_0,T])}.$

Proof. By Corollary 3.2, we know that the l.h.s. is the solution of the smoothing problem. We just have to prove the equality. Set $\mathscr{T} = [t_0, T]$. By our assumptions on F, G, Q, H, we obtain that $K(\cdot, \cdot|T) \in L^\infty(\mathscr{T} \times \mathscr{T}, \mathcal{L}(\mathbb{R}^{n,*}, \mathbb{R}^n))$. Consequently the following kernel integral operator $\mathscr{K} : L^2(\mathscr{T}, \mathbb{R}^{n,*}) \to L^2(\mathscr{T}, \mathbb{R}^n)$ is self-adjoint and bounded [8, Proposition 4.2],

$$(\mathscr{K}f)(s) := \int_{t_0}^T K(s, t|T) f(t) dt \text{ for } f \in L^2(\mathscr{T}, \mathbb{R}^{n,*}).$$

Furthermore, by the reproducing property (4.6), we have that

$$\forall f \in L^2(\mathscr{T}, \mathbb{R}^{n,*}), \, g \in \mathcal{S}_{[t_0,T]}^x, \, \langle g, f \rangle_{L^2(\mathscr{T})} = \int_{t_0}^T \langle f(t), g(t) \rangle \, dt$$

$$= \left\langle g, \int_{t_0}^T K(\cdot, t|T) f(t) dt \right\rangle_{\mathcal{S}_{[t_0,T]}^x} = \langle g, \mathscr{K}f \rangle_{\mathcal{S}_{[t_0,T]}^x}.$$

Consequently (4.19) boils down to minimizing the strongly convex function $\|\widetilde{x}(\cdot)\|^2_{\mathcal{S}^x_{[t_0,T]}} - 2\left\langle \mathscr{K}[H^*R^{-1}\widetilde{y}(\cdot)], \widetilde{x}\right\rangle_{\mathcal{S}^x_{[t_0,T]}}$, which gives $\hat{x}(\cdot) = \mathscr{K}[H^*R^{-1}\widetilde{y}(\cdot)]$ as expected. This concludes the proof. □

Remark 4.9 (Analogy with Sobolev spaces in calculus of variations)**.** The above problem (4.19) may seem unfamiliar to kernel practitioners. Indeed it would be a kernel ridge regression if not for the last term in (4.18), as it is not a quadratic data fitting term with a quadratic regularizer, but a linear data fitting term. Nevertheless it precisely matches the variational formulation $(\int\|\nabla u(t)\|^2 dt + \langle f, u\rangle_{L^2})$ of a PDE like Poisson's equation ($\Delta u = f$) with null boundary condition. The emphasis we have put on the Sobolev-like space $\mathcal{S}^x_{[t_0,T]}$ over which we optimize is no different than the focus on the Sobolev space H^1_0 to study Poisson's equation. [26] treated Linear-Quadratic optimal control as a form of quadratic programming. The Hilbert spaces of trajectories and information vectors we defined are complementary to his discussion.[7] Considering a dual convex problem to an optimal control one is admittedly less frequent in the control community, even though it may have some computational advantages [see 7, and references therein]. Here we derive the dual problem to (4.19), which corresponds formally to (MVLE-det).

Proposition 4.10. *Decompose $R^{-1}(t)\widetilde{y}(t)$ into the sum of two vectors of $\mathbb{R}^{m,*}$:* $\mathrm{proj}^{\|\cdot\|_{R(t)}}_{\mathrm{Im}\,H(t)}(\widetilde{y}(t)) \in R(t)^{-1}\mathrm{Im}\,H(t)$ *and* $\mathrm{proj}^{\|\cdot\|_{R(t)}}_{\mathrm{Ker}\,H^*(t)}(\widetilde{y}(t)) \in \mathrm{Ker}\,H^*(t)$. *The convex dual problem to (4.19) is, for the adjoint control $v(\cdot)$ defined as in (4.11),*

$$(4.20)\quad \min_{\lambda(\cdot)\in\mathcal{S}^\lambda_{[t_0,T]}} L_\lambda(\lambda(\cdot)) = \|\lambda(\cdot)\|^2_{\mathcal{S}^\lambda_{[t_0,T]}} - 2\int_{t_0}^T \left\langle R(t)\,\mathrm{proj}^{\|\cdot\|_{R(t)}}_{\mathrm{Im}\,H(t)}(\widetilde{y}(t)), v(t)\right\rangle dt$$

$$- \|R(\cdot)^{1/2}\,\mathrm{proj}^{\|\cdot\|_{R(\cdot)}}_{\mathrm{Ker}\,H^*(\cdot)}(\widetilde{y}(\cdot))\|^2_{L^2}$$

$$= \int_{t_0}^T \|\mathrm{proj}^{\|\cdot\|_{R(t)}}_{\mathrm{Im}\,H(t)}(\widetilde{y}(t)) - v(t)\|^2_{R(t)}dt + \int_{t_0}^T \langle G(t)Q(t)G^*(t)\lambda(t), \lambda(t)\rangle\, dt$$

$$+ \langle \Pi_0\lambda(t_0), \lambda(t_0)\rangle + \left\langle \Sigma^\ominus_T\lambda(T), \lambda(T)\right\rangle - \|R(\cdot)^{-1/2}\widetilde{y}(\cdot)\|^2_{L^2}.$$

Proof. For any given $\lambda(\cdot) \in C^1(\mathscr{T}, \mathbb{R}^{n,*})$ and $\lambda_0 \in \mathbb{R}^{n,*}$, we introduce the Lagrangian L over $x(\cdot) \in C^1(\mathscr{T}, \mathbb{R}^n)$, $u(\cdot) \in L^2(\mathscr{T}, \mathbb{R}^d)$, $\xi \in \mathbb{R}^n$,

$$L_{tot}(\xi, u(\cdot), x(\cdot), \lambda(\cdot), \lambda_0) = \|\xi\|^2 + \langle \Sigma_T x(T), x(T)\rangle + \int_{t_0}^T \|u(\tau)\|^2 d\tau$$

$$+ \int_{t_0}^T \left\langle H^*(\tau)R^{-1}(\tau)(H(\tau)x(\tau) - 2\widetilde{y}(\tau)), x(\tau)\right\rangle d\tau$$

$$+ 2\int_{t_0}^T \left\langle \lambda(t), \frac{d}{d\tau}x - F(\tau)x(\tau) - G(\tau)Q^{\frac{1}{2}}(\tau)u(\tau)\right\rangle d\tau$$

$$+ 2\left\langle x(t_0) - \Pi_0^{\frac{1}{2}}\xi, \lambda_0\right\rangle + \|R(\cdot)^{-1/2}\widetilde{y}(\cdot)\|^2_{L^2}.$$

---

[7][26] takes the control in L^1 and considers additional constraints, making it closer to optimization problems over Banach spaces $W^{1,\infty}$ or $W^{1,1}$ which we do not cover in our Hilbertian setting.

Integrating by parts, and minimizing over $(\xi, x(\cdot), u(\cdot), x(T), x(t_0))$, and we obtain that

$$
\begin{aligned}
&\partial_\xi \quad \xi = \Pi_0^{\frac{1}{2}} \lambda_0, \qquad \partial_{u(\cdot)} \quad u = Q^{\frac{1}{2}} G^* \lambda, \\
&\partial_{x(T)} \quad \lambda(T) = -\Sigma_T x(T), \qquad \partial_{x(t_0)} \quad \lambda_0 = \lambda(t_0), \\
&\partial_{x(\cdot)} \quad 0 = \frac{d}{dt} \lambda(t) + F^*(t)\lambda(t) - H^*(t)R^{-1}(t)(H(t)x(t) - \widetilde{y}(t)),
\end{aligned}
$$

whence $\lambda(\cdot) \in \mathcal{S}_{[t_0, T]}^\lambda$. Decompose $R^{-1}(t)\widetilde{y}(t)$ into $v_{H,y}(t) + v_{H*,y}(t)$ with $R(t)v_{H,y}(t) \in \operatorname{Im} H(t)$ and $v_{H*,y}(t) \in \operatorname{Ker} H^*(t)$. Denoting by $\hat{x}$ the optimum of $x(\cdot)$ and setting $v(t) = -R^{-1}(t)H(t)\hat{x}(t) + v_{H,y}(t)$, $z = -\Sigma_T \hat{x}(T)$, we obtain that $v(t)$ is the representative of $\lambda(\cdot) \in \mathcal{S}_{[t_0, T]}^\lambda$, since $H^*(t)R^{-1}(t)\widetilde{y}(t) = H^*(t)v_{H,y}(t)$ and for all $\widetilde{v}$ satisfying (4.11) and $H^*(t)\widetilde{v}(t) = 0$, we have obviously $\langle R(t)v(t), \widetilde{v}(t) \rangle = 0$, so (4.12) holds, and for $z = -\Sigma_T \hat{x}(T)$ as well, showing $(v_{H,y}(\cdot, z))$ is the representative of $\lambda(\cdot)$. Consequently

$$
\begin{aligned}
L(\hat{x}(\cdot), \lambda(\cdot)) = &-\langle \Pi_0 \lambda(t_0), \lambda(t_0) \rangle - \|z\|^2 - \int_{t_0}^T \langle G(t)Q(t)G^*(t)\lambda(t), \lambda(t) \rangle \, dt \\
&- \int_{t_0}^T \left\langle H^*(\tau)R^{-1}(\tau)(H(\tau)\hat{x}(\tau), \hat{x}(\tau) \right\rangle d\tau + \|R(\cdot)^{-1/2}\widetilde{y}(\cdot)\|_{L^2}^2,
\end{aligned}
$$

since $\int_{t_0}^T \langle H^*(\tau)R^{-1}(\tau)(H(\tau)\hat{x}(\tau), \hat{x}(\tau) \rangle \, d\tau = \langle R(v - v_{H,y}), v - v_{H,y} \rangle_{L^2}$,

$$
\begin{aligned}
L(\hat{x}(\cdot), \lambda(\cdot)) = &-\|\lambda(\cdot)\|_{\mathcal{S}_{[t_0,T]}^\lambda}^2 + 2\int_{t_0}^T \langle R(t)v_{H,y}(t), v(t) \rangle \, dt \\
&- \|R(\cdot)^{1/2}\widetilde{v}_{H,y}(\cdot)\|_{L^2}^2 + \|R(\cdot)^{-1/2}\widetilde{y}(\cdot)\|_{L^2}^2.
\end{aligned}
$$

Since, by orthogonality $\|R(\cdot)^{-1/2}\widetilde{y}(\cdot)\|_{L^2}^2 = \|R(\cdot)^{1/2}\widetilde{v}_{H,y}(\cdot)\|_{L^2}^2 + \|R(\cdot)^{1/2}\widetilde{v}_{H*,y}(\cdot)\|_{L^2}^2$, this concludes the proof.

$\square$

The two problems we obtained (4.19)-(4.20) are the proper formalization of (LSE)-(MVLE-det) when the function spaces are properly defined. Not using the RKHSs would be similar to not defining the Sobolev spaces when doing variational analysis of quadratic energies. Furthermore, unlike (LSE) and (MVLE-det), which stem from stochastic or Bayesian viewpoints as presented in Section 2.3, (4.19) and (4.20) can be easily modified to account for further side information, such as the sign of the trajectory to be reconstructed, which is a form of state constraint, or other objective functions. We can then use the full machinery of kernel representer theorems to derive a numerical solution, as in [1].

5. COMPLEMENTS ON THE FILTER AND SMOOTHER

In this section, for completeness, we derive the expression of the Kalman filter in continuous time and the expression of the smoother based on the innovation process, both being well-known. We take $\Sigma_T = 0$ and $l_\nu(t) = H^*(t)R^{-1}(t)g(t)$ with $g(\cdot) \in L^2(t_0, T; \mathbb{R}^m)$, and later we posit $g(t)dt = d\widetilde{y}(t)$. Going back to the expression

(2.9) of the best estimate, and to the expression (3.18) of the operator $\widehat{S}_s(t|T)$), we can write the formula

$$(5.1) \qquad \hat{x}(s|T) = \bar{x}(s) + \int_{t_0}^{T} K(s,t|T)H^*(t)R^{-1}(t)d\tilde{y}(t),$$

which gives for $s = T$ the Kalman filter formula

$$(5.2) \qquad \hat{x}(T|T) = \bar{x}(T) + \int_{t_0}^{T} K(T|T,t)H^*(t)R^{-1}(t)d\tilde{y}(t).$$

We want to show that the smoother at s can be recursively expressed in terms of the sequence of filters between s and T. As we deal with Linear-Quadratic-Gaussian estimation problems, we naturally find ourselves in a very thoroughly explored field and some expressions below are well-known, and we point out to where they could be found in textbooks [14, 5]. General kernel formulas are in any case new, and so is the connection of Kalman filtering and smoothing with the explicit RKHSs discussed in Section 4.

5.1. **Rederiving the Kalman filter.** Recall that we defined $r(t) = \hat{\mu}(t) + \Pi(t)\hat{\nu}(t)$ in (3.5). By (3.14),

$$(5.3) \quad \frac{dr}{dt} = (F(t) - \Pi(t)H^*(t)R^{-1}(t)H(t))r(t) + \Pi(t)H^*(t)R^{-1}(t)g(t),$$

$$r(t_0) = 0.$$

The importance of $r(t)$ is that, like $\Pi(t)$, it satisfies a forward differential equation and thus it does not depend on T. On the other hand, since $\hat{\nu}(T) = 0$, $\hat{\mu}(T) = r(T)$. But then from (3.14) we can write

$$(5.4) \quad r(T) \quad = \quad \int_{t_0}^{T} K(T|T,t)H^*(t)R^{-1}(t)g(t)dt \quad = \quad \int_{t_0}^{T} \widehat{S}(T|T,t)g(t)dt.$$

If in (5.4) we substitute $g(t)dt$ by $d\tilde{y}(t)$ we obtain the stochastic differential equation (SDE) as given in [5, Chapter 7]

$$(5.5) \quad dr = (F(t) - \Pi(t)H^*(t)R^{-1}(t)H(t))r(t)dt + \Pi(t)H^*(t)R^{-1}(t)d\tilde{y}(t),$$

$$r(t_0) = 0.$$

and

$$(5.6) \qquad r(T) = \int_{t_0}^{T} \widehat{S}(T|T,t)d\tilde{y}(t).$$

From (5.2) it follows that the Kalman filter over $[0,t]$, i.e. only filtering without smoothing, denoted to simplify notation $\hat{x}(t) = \hat{x}(t|t) = \bar{x}(t) + r(t)$, thus satisfies the SDE

$$d\hat{x}(t) = (F(t)\hat{x}(t) + f(t))dt + \Pi(t)H^*(t)R^{-1}(t)(d\tilde{y}(t) - H(t)r(t)dt).$$

Finally, we obtain the classical Kalman filter equation

$$(5.7) \quad d\hat{x}(t) = (F(t)\hat{x}(t) + f(t))dt + \Pi(t)H^*(t)R^{-1}(t)(dy(t) - (H(t)\hat{x}(t) + h(t))),$$

$$\hat{x}(t_0) = x_0.$$

5.2. **Expression of the smoother in terms of innovation.** From (3.15), we know that

$$(5.8) \quad -\frac{d}{d\tau}\widehat{\nu}(\tau) = (F^*(\tau) - H^*(\tau)R^{-1}(\tau)H(\tau)\Pi(\tau))\widehat{\nu}(\tau)$$
$$- H^*(\tau)R^{-1}(\tau)(g(\tau) - H(\tau)r(\tau)),$$
$$\widehat{\nu}(T) = 0$$

and thus, by the variation of constants formula,

$$(5.9) \qquad \widehat{\nu}(s) = -\int_s^T \Phi^*_{F,\Pi}(t,s)H^*(t)R^{-1}(t)(g(t) - H(t)r(t))dt.$$

Therefore, we have

$$(5.10) \qquad \widehat{\mu}(s) = r(s) + \Pi(s)\int_s^T \Phi^*_{F,\Pi}(t,s)H^*(t)R^{-1}(t)(g(t) - H(t)r(t))dt.$$

If we replace in the equation $g(t)dt$ by $d\widetilde{y}(t)$ we must interpret $r(t)$ as the solution of the SDE (5.5) . We then have

$$(5.11) \qquad \widehat{x}(s|T) = \overline{x}(s) + \widehat{\mu}(s)$$

$$\widehat{x}(s) = \overline{x}(s) + r(s).$$

From (5.10), we deduce that

$$(5.12) \ \widehat{x}(s|T) = \widehat{x}(s) + \Pi(s)\int_s^T \Phi^*_{F,\Pi}(t,s)H^*(t)R^{-1}(t)(dy(t) - (H(t)\widehat{x}(t) + h(t))dt).$$

Define the innovation process $e(\cdot)$ as follows

$$(5.13) \qquad e(t) = y(t) - \int_{t_0}^t (H(\tau)\widehat{x}(\tau) + h(\tau))d\tau.$$

It is well-known that the innovation process is a $\mathcal{Y}^t$ Wiener process with covariance matrix $R(t)$ [see e.g. 5, Lemma 7.1]. In a nutshell we obtained that

Proposition 5.1. *The Kalman smoother can be written as follows*

$$(5.14) \qquad \widehat{x}(s|T) = \widehat{x}(s) + \Pi(s)\int_s^T \Phi^*_{F,\Pi}(t,s)H^*(t)R^{-1}(t)de(t)$$

This result is well-known and can be found in [14, Lemma 16.5.1]. Consequently, the error of the Kalman smoother is given by

$$(5.15) \quad \widehat{\epsilon}(s|T) = x(s) - \widehat{x}(s) - \Pi(s)\int_s^T \Phi^*_{F,\Pi}(t,s)H^*(t)R^{-1}(t)de(t)$$
$$= \widehat{\epsilon}(s|s) - \Pi(s)\int_s^T \Phi^*_{F,\Pi}(t,s)H^*(t)R^{-1}(t)de(t)$$

It is obvious that, for $s < t$, the random variables $\widehat{\epsilon}(s|s)$ and $e(t)$ are independent. From (5.15), one can then compute $K(s,s|T) = \mathbb{E}[\widehat{\epsilon}(s|T)\widehat{\epsilon}^*(s|T)]$ and obtain (4.14) for the special case of $s = t$, as in [14, Theorem 16.5.1].

6. Summary of results

In this section we summarize the findings and key expressions of the article. We also highlight the improvements made over previous formulas for kernels and controlled linear systems. Our problem was to find the operator $\hat{S}_s(t|T)$ minimizing the variance of the estimation error,

$$\text{(6.1)} \qquad \epsilon_S(s|T) = x(s) - x_S(s|T) = \widetilde{x}(s) - \int_{t_0}^{T} S_s(t|T)d\widetilde{y}(t).$$

$$\text{(6.2)} \qquad \hat{S}_s(\cdot|T) \in \underset{S_s(\cdot|T)}{\operatorname{argmin}} \, \Gamma_S(s|T) = \mathbb{E}[\epsilon_S(s|T)(\epsilon_S(s|T))^*].$$

The latter is related to the posterior covariance through the expression $\hat{S}_s(t|T) = K(s,t|T)H^*(t)R^{-1}(t)$ (Corollary 3.2) with

$$\text{(6.3)} \qquad K(s,t|T) = \mathbb{E}[\epsilon_{\hat{S}}(s|T)(\epsilon_{\hat{S}}(t|T))^*] \in \mathcal{L}(\mathbb{R}^{n,*}, \mathbb{R}^n).$$

The explicit formula for K can be obtained by finding the Green functions of a two-point boundary value problem:

$$\text{(6.4)} \qquad \frac{d\hat{\mu}}{dt} = F(t)\hat{\mu}(t) - G(t)Q(t)G^*(t)\hat{\nu}(t) + l_\mu(t)$$

$$-\frac{d\hat{\nu}}{dt} = F^*(t)\hat{\nu}(t) + H^*(t)R^{-1}(t)H(t)\hat{\mu}(t) - l_\nu(t)$$

$$\hat{\mu}(t_0) = -\Pi_0\hat{\nu}(t_0), \ \hat{\nu}(T) = \Sigma_T\hat{\mu}(T).$$

The canonical resolution proceeds by introducing two Riccati differential equations

$$\text{(6.5)}$$
$$-\frac{d}{dt}\Sigma = \Sigma(t)F(t) + F^*(t)\Sigma(t) - \Sigma(t)G(t)Q(t)G^*(t)\Sigma(t) + H^*(t)R^{-1}(t)H(t),$$
$$\Sigma(T) = \Sigma_T;$$

$$\text{(6.6)} \quad \frac{d}{dt}\Pi = F(t)\Pi(t) + \Pi(t)F^*(t) - \Pi(t)H^*(t)R^{-1}(t)H(t)\Pi(t) + G(t)Q(t)G^*(t),$$
$$\Pi(t_0) = \Pi_0.$$

which we complement by looking for matrix-valued kernels K and Λ satisfying

$$\text{(6.7)} \qquad \hat{\mu}(s) = \int_{t_0}^{T} K(s,t|T)l_\nu(t)dt \quad \text{for } l_\mu(\cdot) \equiv 0,$$

$$\text{(6.8)} \qquad \hat{\nu}(s) = \int_{t_0}^{T} \Lambda(s,t|T)l_\mu(t)dt \quad \text{for } l_\nu(\cdot) \equiv 0.$$

where K corresponds to the covariance (6.3) as shown in Proposition 4.6. Introducing a semigroup associated with the matrix function $F(t) - G(t)Q(t)G^*(t)\Sigma(t)$

denoted $\Phi_{F,\Sigma}(s,t)$ (resp. $F(s) - \Pi(s)H^*(s)R^{-1}(s)H(s)$, denoted $\Phi_{F,\Pi}(s,t)$), we obtained two symmetric formulas (Theorem 3.1):

$$(6.9) \quad K(s,t|T) = \Phi_{F,\Sigma}(s,t_0)\Pi_0^{\frac{1}{2}}(\mathrm{Id} + \Pi_0^{\frac{1}{2}}\Sigma(t_0)\Pi_0^{\frac{1}{2}})^{-1}\Pi_0^{\frac{1}{2}}\Phi_{F,\Sigma}^*(t,t_0)$$
$$+ \int_{t_0}^{\min(s,t)} \Phi_{F,\Sigma}(s,\tau)G(\tau)Q(\tau)G^*(\tau)\Phi_{F,\Sigma}^*(t,\tau)d\tau$$

$$(6.10) \quad \Lambda(s,t|T) = \Phi_{F,\Pi}^*(T,s)\Sigma_T^{\frac{1}{2}}(\mathrm{Id} + \Sigma_T^{\frac{1}{2}}\Pi(T)\Sigma_T^{\frac{1}{2}})^{-1}\Sigma_T^{\frac{1}{2}}\Phi_{F,\Pi}(T,t)$$
$$+ \int_{\max(s,t)}^{T} \Phi_{F,\Pi}^*(\tau,s)H^*(\tau)R^{-1}(\tau)H(\tau)\Phi_{F,\Pi}(\tau,t)d\tau$$

For which we proved (Theorem 4.2), that K was the reproducing kernel of a space of controlled trajectories equipped with a quadratic norm

$$(6.11) \quad \mathcal{S}_{[t_0,T]}^x = \{x(\cdot) \in H^1(t_0,T;\mathbb{R}^n) | \exists\, u(\cdot) \in L^2(t_0,T;\mathbb{R}^p), \xi \in \mathbb{R}^n \text{ s.t.}$$
$$\frac{d}{d\tau}x = F(\tau)x(\tau) + G(\tau)Q^{\frac{1}{2}}(\tau)u(\tau), x(t_0) = \Pi_0^{\frac{1}{2}}\xi\}$$

$$(6.12) \quad \|x(\cdot)\|_{\mathcal{S}_{[t_0,T]}^x}^2 = J_x(\xi,u(\cdot)) = \|\xi\|^2 + \langle \Sigma_T x(T), x(T) \rangle + \int_{t_0}^{T} \|u(\tau)\|^2 d\tau$$
$$+ \int_{t_0}^{T} \left\langle H^*(\tau)R^{-1}(\tau)H(\tau)x(\tau), x(\tau) \right\rangle d\tau.$$

Similarly, Λ is associated with a Hilbert space of information vectors (Theorem 4.4)

$$(6.13) \quad \mathcal{S}_{[t_0,T]}^\lambda = \{\lambda(\cdot) \in H^1(t_0,T;\mathbb{R}^n) | v(\cdot) \in L^2(t_0,T;\mathbb{R}^m), z \in \mathbb{R}^n \text{ s.t.}$$
$$-\frac{d}{dt}\lambda(t) = F^*(t)\lambda(t) + H^*(t)v(t), \ \lambda(T) = \Sigma_T^{\frac{1}{2}}z\}$$

$$(6.14) \quad \|\lambda(\cdot)\|_{\mathcal{S}_{[t_0,T]}^\lambda}^2 = J_\lambda(z,v(\cdot)) = \langle \Pi_0\lambda(t_0), \lambda(t_0) \rangle + \|z\|^2$$
$$+ \int_{t_0}^{T} \langle G(t)Q(t)G^*(t)\lambda(t), \lambda(t) \rangle dt + \int_{t_0}^{T} \langle R(t)v(t), v(t) \rangle dt$$

The two kernels are related by the following formula (Proposition 4.6)

$$(6.15) \quad K(s,t|T) = \Pi(s)\Phi_{F,\Pi}^*(t,s)\mathbb{1}_{s \leq t} + \Phi_{F,\Pi}(s,t)\Pi(t)\mathbb{1}_{s > t} - \Pi(s)\Lambda(s,t|T)\Pi(t).$$

For any realization $\widetilde{y}(\cdot)$ of $Y - \bar{y}$, the kernels allow to define two dual optimization problems (see Propositions 4.8 and 4.10 for the definitions) formalizing (LSE)-(MVLE-det)

$$(6.16) \quad \min_{\tilde{x}(\cdot) \in \mathcal{S}^x_{[t_0,T]}} \|R(t)^{-1/2}\widetilde{y}(\cdot)\|^2_{L^2} + \|\widetilde{x}(\cdot)\|^2_{\mathcal{S}^x_{[t_0,T]}} - 2\left\langle H^*(\cdot)R^{-1}(\cdot)\widetilde{y}(\cdot), \widetilde{x}(\cdot)\right\rangle_{L^2([t_0,T])}$$

$$= \int_{t_0}^T \|\tilde{y}(t) - H(t)\tilde{x}(t)\|^2_{R(t)^{-1}} dt + \|\tilde{x}(\cdot)\|^2_{\mathcal{S}^x_{[t_0,T]}} - \int_{t_0}^T \|H(t)\tilde{x}(t)\|^2_{R(t)^{-1}} dt$$

$$(6.17) \quad \min_{\lambda(\cdot) \in \mathcal{S}^\lambda_{[t_0,T]}} \|\lambda(\cdot)\|^2_{\mathcal{S}^\lambda_{[t_0,T]}} - 2\int_{t_0}^T \left\langle R(t)\,\mathrm{proj}^{\|\cdot\|_{R(t)}}_{\mathrm{Im}\,H(t)}(\widetilde{y}(t)), v(t)\right\rangle dt$$

$$- \|R(\cdot)^{1/2}\,\mathrm{proj}^{\|\cdot\|_{R(\cdot)}}_{\mathrm{Ker}\,H^*(\cdot)}(\widetilde{y}(\cdot))\|^2_{L^2}$$

$$= \int_{t_0}^T \|\mathrm{proj}^{\|\cdot\|_{R(t)}}_{\mathrm{Im}\,H(t)}(\widetilde{y}(t)) - v(t)\|^2_{R(t)} dt + \int_{t_0}^T \langle G(t)Q(t)G^*(t)\lambda(t), \lambda(t)\rangle dt$$

$$+ \langle \Pi_0\lambda(t_0), \lambda(t_0)\rangle + \left\langle \Sigma^\ominus_T \lambda(T), \lambda(T)\right\rangle - \|R(\cdot)^{-1/2}\tilde{y}(\cdot)\|^2_{L^2}.$$

This relates Kalman filtering to optimization problems over RKHSs, which act in filtering problems as the Sobolev spaces in calculus of variations. Inspired by Table 15.1 in [14], which written for discrete-time estimation problems, we summarize in Table 1 the relations between the deterministic optimal control problems written over RKHSs (6.16)-(6.17) and the original stochastic smoothing problems. The lower line, consisting of problems (iii) and (iv) over dual variables, is arguably less studied both in estimation and control, and we leave to future work the identification of the backward Markovian Gaussian process underlying Λ.

TABLE 1. Summary of the four optimization problems considered in the article. One moves vertically by permuting min-max into max-min, the problems being (Fenchel) dual. One moves horizontally by formally setting $dw(t) = u(t)dt$ and considering the same kernel for (stochastic) covariance of optimal error and (deterministic) trajectories, the problems being (kernel) "equivalent".

	Stochastic problems	Deterministic problems
primal variables	(i) Given Gaussian processes $(X_t, Y_t)_{t\in[0,T]}$ Solve linear MMSE i.e. (MVLE) $\min_{\hat{X}\in\mathcal{L}(Y)} \mathbb{E}((X-\hat{X})^\top(X-\hat{X}))$ Optimum: $\hat{X} = \mathbb{E}[X\|Y] = \hat{S}Y$	(ii) Given RKHS $\mathcal{S}^x_{[t_0,T]}$ with kernel K, observations $\tilde{y}(\cdot)$ Solve optimal control problem over trajectories $\tilde{x}(\cdot) \in \mathcal{S}^x_{[t_0,T]}$ $\|\tilde{x}(\cdot)\|^2_{\mathcal{S}^x_{[t_0,T]}} - 2\left\langle H^*(\cdot)R^{-1}(\cdot)\tilde{y}(\cdot), \tilde{x}(\cdot)\right\rangle_{L^2([t_0,T])}$ Optimum: $\hat{x}(\cdot) = \int_{t_0}^T K(\cdot, t\|T)H^*(t)R^{-1}(t)\tilde{y}(t)dt$
dual variables	(iii) Given Gaussian processes $(X_t^\top, Y_T)_{t\in[0,T]}$ Solve over $(\Lambda_t)_{t\in[0,T]}$ (MVLE-dual) $\min_{\Lambda\in\mathcal{L}(Y)^\perp} \mathbb{E}((X^\top-\Lambda)^\top(X^\top-\Lambda))$ Optimum: $\hat{X}^\top = X^\top - \hat{\Lambda}$	(iv) Given RKHS $\mathcal{S}^\lambda_{[t_0,T]}$ with kernel Λ, observations $\tilde{y}(\cdot)$ Solve optimal control problem over adjoint $\lambda(\cdot) \in \mathcal{S}^\lambda_{[t_0,T]}$ $\|\lambda(\cdot)\|^2_{\mathcal{S}^\lambda_{[t_0,T]}} - 2\int_{t_0}^T \left\langle R(t)\,\mathrm{proj}^{\|\cdot\|_{R(t)}}_{\mathrm{Im}\,H(t)}(\tilde{y}(t)), v(t)\right\rangle dt$ Optimum: $\hat{v}(t) = -R^{-1}(t)H(t)\hat{x}(t) + \mathrm{proj}^{\|\cdot\|_{R(t)}}_{\mathrm{Im}\,H(t)}(\tilde{y}(t))$

Comparison with previous results on kernels and control: The trajectory space seen as an RKHS was first presented in previous articles on linear-quadratic optimal control:

- [1] first introduced the idea of considering the vector space $(\mathcal{S}^x_{[t_0,T]}, \langle \cdot, \cdot \rangle_{\mathcal{S}^x_{[t_0,T]}})$ as a reproducing kernel Hilbert space in the case $\Sigma_T = 0$. A two-point boundary system to obtain K was given but not solved explicitly. The closed form formula were only given for $\Pi_0 = \mathrm{Id}$ and $H \equiv 0$. The kernel was used to guarantee the satisfaction of state constraints;

- [2] observed that, for $\Sigma_T \neq 0$ and $\Pi_0 \to \infty$, the map $t_0 \mapsto K(t_0, t_0|[t_0, T])$ satisfied a forward Riccati equation and was the inverse of the usual backward Riccati matrix considered in linear-quadratic optimal control. It was underlined that kernels shift the focus on trajectories rather than their parametrization by controls;

- [3] gave a closed form formula for K, when $\Sigma_T = 0$, in the general case of an infinite-dimensional state with values in a Hilbert space to tackle linear PDE control. We emphasized there that this allowed for representer theorems and closed-form solutions when considering linear-quadratic optimal control problems.

To summarize, the kernels considered appear in linear-quadratic optimal control because of Hilbertian vector spaces of trajectories, while, for estimation problems, they appear through covariances of Gaussian processes. It is this "dual", deterministic and stochastic, nature of kernels which underlies the "duality" between optimal control and estimation in the Linear-Quadratic case.

Conclusion. We improved on our previous results by considering the dual RKHS $(\mathcal{S}^\lambda_{[t_0,T]}, \langle \cdot, \cdot \rangle_{\mathcal{S}^\lambda_{[t_0,T]}})$ of covectors/information vectors, by relaxing as much as possible the invertibility requirements on the matrices Q, Π_0, Σ_T, and most of all by considering an estimation, rather than optimal control, problem and consequently defining the kernel K as the covariance of the estimation error. This way, we derived novel formulas for the covariances of the Markovian Gaussian processes induced by linear SDEs. The kernels presented should also in principle allow for incorporating constraints or considering various sampling times of observations, and are computable through the matrix exponential formulas given in Appendix A.2. We considered here a continuous-time observation process, the extension to discrete-time measurements is straightforward by replacing integrals with finite sums when dealing with the observation operator H. Similarly generalization to an infinite-dimensional state space can be done in the spirit of [3]. We could not discuss here the question of duality as a formal change of variables as in [16, 32], as well as the relation between value function and likelihood. Indeed, the two Riccati equations outlined correspond respectively to the backward evolution of the Hessian of the quadratic value function through the Hamilton-Jacobi-Bellman equation, and to the forward evolution of the Hessian of the Gaussian posterior density through the Fokker-Planck equation. This paves the way to extending to nonlinear filtering the kernel viewpoint. We leave these interesting directions to future work.

APPENDIX

A.1 Existence of the solution of the Riccati equation.

Lemma 6.1. *The Riccati equation (3.1) has a solution $\Sigma(t)$ which is symmetric and positive semi-definite on the interval $[t_0, T]$.*

Proof. The differential equation (3.1) has a unique local solution on an interval $[t_1, T]$ for t_1 sufficiently close to T. The solution is symmetric, since the transpose satisfies the equation. We can then consider the differential equation

$$(6.18) \qquad \frac{d\gamma}{dt} = (F(t) - G(t)Q(t)G^*(t)\Sigma(t))\gamma(t), \quad \gamma(t_1) = \gamma_0.$$

We can then compute

$$\frac{d\langle \gamma(t), \Sigma(t)\gamma(t)\rangle}{dt} = 2\left\langle \gamma(t), \Sigma(t)\frac{d\gamma(t)}{dt}\right\rangle + \left\langle \gamma(t), \frac{d\Sigma(t)}{dt}\gamma(t)\right\rangle$$

$$= -\left\langle \gamma(t), \left(\Sigma(t)G(t)Q(t)G^*(t)\Sigma(t) + H^*(t)R^{-1}(t)H(t)\right)\gamma(t)\right\rangle \leq 0$$

and since $\Sigma(T) \succcurlyeq 0$, by integration between t_1 and T, we obtain $\langle \gamma_0, \Sigma(t_1)\gamma_0\rangle \geq 0$. Since γ_0 is arbitrary, we obtain $\Sigma(t_1) \geq 0$. We could have started in (6.18) at any point on the interval (t_1, T). Therefore the local solution satisfies $\Sigma(t) \geq 0$. Consider next the semi group $\Phi_F(t, t_1)$, on the interval $t \in (t_1, T)$ defined by the differential equation

$$(6.19) \qquad \frac{d}{dt}\Phi_F(t, t_1) = F(t)\Phi_F(t, t_1), \quad \Phi_F(t_1, t_1) = \mathrm{Id}.$$

If we consider

$$\frac{d\langle \Phi_F(t, t_1)\gamma_0, \Sigma(t)\Phi_F(t, t_1)\gamma_0\rangle}{dt}$$

$$= \left\langle \Phi_F(t, t_1)\gamma_0, \left(F^*(t)\Sigma(t) + \Sigma(t)F(t) + \frac{d}{dt}\Sigma(t)\right)\Phi_F(t, t_1)\gamma_0\right\rangle,$$

from the Riccati equation, we obtain

$$\frac{d\langle \Phi_F(t, t_1)\gamma_0, \Sigma(t)\Phi_F(t, t_1)\gamma_0\rangle}{dt} \geq -\left\langle \Phi_F(t, t_1)\gamma_0, H^*(t)R^{-1}(t)H(t)\Phi_F(t, t_1)\gamma_0\right\rangle$$

Integrating between t_1 and T, we get

$$(6.20) \quad \langle \gamma_0, \Sigma(t)\gamma_0\rangle \leq \int_{t_1}^{T}\left\langle \Phi_F(t, t_1)\gamma_0, H^*(t)R^{-1}(t)H(t)\Phi_F(t, t_1)\gamma_0\right\rangle \leq C\|\gamma_0\|^2$$

for some constant $C > 0$. It follows that the solution $\Sigma(t)$ can be extended beyond t_1 and finally up to t_0. This completes the proof. $\qquad\square$

A.2 Computation of kernels of linear SDEs through exponentials of the Hamiltonian matrix. The following method to compute Gramians can be traced at least back to [33] in the time-invariant case and was mentioned in [15, Section 8] as a periodically rediscovered way to solve differential Riccati equations. It made its way in the linear SDE literature [see e.g. 28, Section 6.3, p.84] where $H \equiv 0$ and $\Sigma \equiv 0$. However a more systematic presentation through the Hamiltonian matrix can be found in [30] for linear-quadratic control. In particular this allows to consider

the important case of $GQG^* \not\equiv 0$ and $H \not\equiv 0$. We focus below on computing K but similar operations can be performed to obtain Λ by changing variables, as was done in the proof of Theorem 4.4. Consider the equations

$$\frac{d\hat{\mu}}{dt} = F(t)\hat{\mu}(t) - G(t)Q(t)G^*(t)\hat{\nu}(t), \qquad\qquad \hat{\mu}(t_0) = -\Pi_0\hat{\nu}(t_0)$$

$$\frac{d\hat{\nu}}{dt} = -F^*(t)\hat{\nu}(t) - H^*(t)R^{-1}(t)H(t)\hat{\mu}(t), \qquad \hat{\nu}(T) = \Sigma_T\hat{\mu}(T)$$

and write them in matrix form introducing the Hamiltonian matrix $\mathbb{H}(t)$

$$(6.21) \qquad \frac{d}{dt}\begin{pmatrix}\hat{\mu}(t)\\\hat{\nu}(t)\end{pmatrix} = \underbrace{\begin{pmatrix}F(t) & -G(t)Q(t)G^*(t)\\-H^*(t)R^{-1}(t)H(t) & -F^*(t)\end{pmatrix}}_{\mathbb{H}(t)}\begin{pmatrix}\hat{\mu}(t)\\\hat{\nu}(t)\end{pmatrix}.$$

Denote by $\Phi_{\mathbb{H}}(T,t)$ the transition matrix $\frac{d}{dt}\Phi_{\mathbb{H}}(T,t) = -\mathbb{H}(t)\Phi_{\mathbb{H}}(T,t)$, $\Phi_{\mathbb{H}}(T,T) = $ Id and set as in [30, eq.(5.72)]

$$\Phi_{\mathbb{H}}(T,t) := \begin{pmatrix}\Phi_{11}(T,t) & \Phi_{12}(T,t)\\\Phi_{21}(T,t) & \Phi_{22}(T,t)\end{pmatrix}$$

$$\bar{\Phi}_{\mathbb{H}}(T,t) := \begin{pmatrix}\bar{\Phi}_{11}(T,t) & \bar{\Phi}_{12}(T,t)\\\bar{\Phi}_{21}(T,t) & \bar{\Phi}_{22}(T,t)\end{pmatrix} = \begin{pmatrix}\text{Id} & 0\\-\Sigma_T & \text{Id}\end{pmatrix}\cdot\Phi_{\mathbb{H}}(T,t).$$

Assume that $\bar{\Phi}_{22}(T,t)$ is invertible, then through some calculations one can show that [see 30, Chapter 5, eq.(5.77,5.160, 5.163)]

$$\Sigma(t) = -\bar{\Phi}_{22}^{-1}(T,t)\bar{\Phi}_{21}(T,t),$$

$$\Phi_{F,\Sigma}(t,T) = \bar{\Phi}_{22}^*(T,t),$$

$$\int_T^s \Phi_{F,\Sigma}(T,\tau)G(\tau)Q(\tau)G^*(\tau)\Phi_{F,\Sigma}^*(T,\tau)d\tau = \bar{\Phi}_{12}(T,s)\bar{\Phi}_{22}^{-1}(T,s).$$

where $\Sigma(t)$ is the solution of the backward Riccati equation (6.5) and $\Phi_{F,\Sigma}(s,t)$ the semigroup associated with $F(t) - G(t)Q(t)G^*(t)\Sigma(t)$. Define two auxiliary kernels which we will compute independently.

$$(6.22)\quad K(s,t|T) = \underbrace{\Phi_{F,\Sigma}(s,t_0)\Pi_0^{\frac{1}{2}}(\text{Id} + \Pi_0^{\frac{1}{2}}\Sigma(t_0)\Pi_0^{\frac{1}{2}})^{-1}\Pi_0^{\frac{1}{2}}\Phi_{F,\Sigma}^*(t,t_0)}_{=:K_0(s,t)}$$

$$+ \underbrace{\int_{t_0}^{\min(s,t)}\Phi_{F,\Sigma}(s,\tau)G(\tau)Q(\tau)G^*(\tau)\Phi_{F,\Sigma}^*(t,\tau)d\tau}_{=:K_1(s,t)}$$

By the above expressions, we have for $s \leq t$

$$K_0(s,t) = \bar{\Phi}_{22}^*(T,s)\bar{\Phi}_{22}^{-1,*}(T,t_0)\Pi_0^{\frac{1}{2}}(\text{Id} + \Pi_0^{\frac{1}{2}}\bar{\Phi}_{22}^{-1}(T,t_0)\bar{\Phi}_{21}(T,t_0)\Pi_0^{\frac{1}{2}})^{-1}$$

$$\times \Pi_0^{\frac{1}{2}}\bar{\Phi}_{22}^{-1}(T,t_0)\bar{\Phi}_{22}(T,t)$$

$$K_1(s,t) = \bar{\Phi}_{22}^*(T,s)[\bar{\Phi}_{12}(T,s)\bar{\Phi}_{22}^{-1}(T,s) - \bar{\Phi}_{12}(T,t_0)\bar{\Phi}_{22}^{-1}(T,t_0)]\bar{\Phi}_{22}(T,t)$$

Formulas for $s > t$ can be obtained by Hermitian symmetry, $K(s,t|T) = K(t,s|T)^*$. These formulas simplify drastically when $H(\cdot) \equiv 0$ and $\Sigma_T = 0$, so $\Sigma(\cdot) \equiv 0$,

$\bar{\Phi}_{22}^*(T,t) = \Phi_F(t,T)$ (alternative computations could be done in this case with $\Phi_{\mathbb{H}}(t,t_0)$) which inverse is easily computable. Otherwise every inversion should be computed numerically. In the time-invariant case, we have $\Phi_{\mathbb{H}}(T,t) = e^{(T-t)\mathbb{H}}$ and all quantities can be computed through matrix exponentials.

References

[1] P.-C. Aubin-Frankowski, *Linearly constrained linear quadratic regulator from the viewpoint of kernel methods*, SIAM Journal on Control and Optimization **59** (2021), 2693–2716.

[2] P.-C. Aubin-Frankowski, *Interpreting the dual Riccati equation through the LQ reproducing kernel*, Comptes Rendus. Mathématique **359** (2021), 199–204.

[3] P.-C. Aubin-Frankowski and A. Bensoussan, *Operator-valued kernels and control of infinite dimensional dynamic systems*, in: Control and Decision Conference (CDC), 2022.

[4] P.-C. Aubin-Frankowski and S. Gaubert, *Tropical reproducing kernels and optimization*, Integral Equations and Operator Theory, 2024.

[5] A. Bensoussan, *Estimation and Control of Dynamical Systems*, Springer International Publishing, 2018.

[6] A. Berlinet and C. Thomas-Agnan, *Reproducing Kernel Hilbert Spaces in Probability and Statistics*, Kluwer, 2004.

[7] R. S. Burachik, C. Y. Kaya and S. N. Majeed, *A duality approach for solving control-constrained linear-quadratic optimal control problems*, SIAM Journal on Control and Optimization **52** (2014), 1423–1456.

[8] C. Carmeli, E. De Vito and A. Toigo, *Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem*, Analysis and Applications **4** (2006), 377–408.

[9] C. Carmeli, E. De Vito, A. Toigo and V. Umanitá, *Vector valued reproducing kernel Hilbert spaces and universality*, Analysis and Applications **8** (2010), 19–61.

[10] G. Da Prato, *An Introduction to Infinite-Dimensional Analysis*, Springer Berlin Heidelberg, 2006.

[11] G. Da Prato and H. Frankowska, *Invariance of stochastic control systems with deterministic arguments*, Journal of Differential Equations **200** (2004), 18–52.

[12] N. Eisenbaum and H. Kaspi, *A characterization of the infinitely divisible squared Gaussian processes*, The Annals of Probability **34** (2006), 728–742.

[13] T. Jaakkola and D. Haussler, *Exploiting generative models in discriminative classifiers*, in: Advances in Neural Information Processing Systems 11, MIT Press, 1998, pp. 487–493.

[14] T. Kailath, A. H. Sayed and B. Hassibi, *Linear Estimation*, Prentice Hall information and system sciences series. Pearson, 2000.

[15] R. Kalman, *Contribution to the theory of optimal control*, Bol. Soc. Mat. Mexicana **5** (1960), 102–119.

[16] R. E. Kálmán and R. S. Bucy, *New results in linear filtering and prediction theory*, Journal of Basic Engineering **83** (1961), 95–108.

[17] M. Kanagawa, P. Hennig, D. Sejdinovic and B. K. Sriperumbudur, *Gaussian processes and kernel methods: A review on connections and equivalences*, 2018. URL https://arxiv.org/abs/1807.02582.

[18] M. Kanagawa, Y. Nishiyama, A. Gretton and K. Fukumizu, *Filtering with state-observation examples via kernel Monte Carlo filter*, Neural Comput. **28** (2016), 382–444.

[19] J. W. Kim, *Duality for nonlinear filtering*, PhD thesis, University of Illinois Urbana-Champaign, 2022.

[20] F. Lindgren, D. Bolin and H. Rue, *The SPDE approach for Gaussian and non-Gaussian fields: 10 years and still running*, Spatial Statistics **50** (2022): 100599.

[21] D. Luenberger, *Optimization by Vector Space Methods*, Wiley, New York, 1968.

[22] A. Lunardi, M. Miranda and D. Pallara, Infinite dimensional analysis (lecture notes), 2016, URL https://www.mathematik.tu-darmstadt.de/media/analysis/lehrmaterial_anapde/hallerd/Lectures.pdf.

[23] J. Neveu, *Processus Aléatoires Gaussiens*, Seminaire Math. Sup., Les presses de I'Universite de Montreal, 1968.

[24] E. Parzen, *An approach to time series analysis*, The Annals of Mathematical Statistics **32** (1961), 951–989.

[25] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning series. MIT Press, 2005.

[26] R. T. Rockafellar, *Linear-quadratic programming and optimal control*, SIAM Journal on Control and Optimization **25** (1987), 781–814.

[27] S. Saitoh and Y. Sawano, *Theory of Reproducing Kernels and Applications*, Springer Singapore, 2016.

[28] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*, Cambridge University Press, 2019.

[29] S. Sarkka, A. Solin and J. Hartikainen, *Spatiotemporal learning via infinite-dimensional Bayesian filtering and smoothing: A look at Gaussian process regression through Kalman filtering*, IEEE Signal Processing Magazine **30** (2013), 51–61.

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

[31] F. Steinke and B. Scholkopf, *Kernels, regularization and differential equations*, Pattern Recognition **41** (2008), 3271–3286.

[32] E. Todorov, *General duality between optimal control and estimation*, in: IEEE Conference on Decision and Control (CDC), 2008, pp. 4286–4292.

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

[34] H. L. Van Trees, *Detection, Estimation, and Modulation Theory. 1: Detection, Estimation, and Linear Modulation Theory*, Wiley, New York, 1968.

P.-C. Aubin-Frankowski
INRIA-Département d'Informatique de l'École Normale Supérieure, PSL, Research University, Paris, France
    E-mail address: `pierre-cyril.aubin@inria.fr`

A. Bensoussan
International Center for Decision and Risk Analysis, Jindal School of Management, University of Texas at Dallas, also with the School of Data Science, City University Hong Kong. Research supported by the National Science Foundation under grants NSF-DMS-1905449 and 2204795, and grant from the SAR Hong Kong RGC GRF 14301321.
    E-mail address: `axb046100@utdallas.edu`