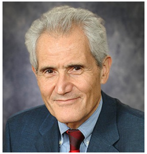
The RKHSs underlying linear SDE Estimation, Kalman filtering and their relation to optimal control

Pierre-Cyril Aubin
joint work with Alain Bensoussan (ex-INRIA, UT Dallas & CU Hong Kong)

Postdoc at INRIA Paris - SIERRA

Retraite SIERRA/WILLOW, October 2022
Paper accepted at *Pure and Applied Functional Analysis* (available on arXiv)



#1 Where's ~~Waldo/Charlie~~ the kernel? For optimal control

Time-varying state-constrained LQ optimal control

$$
\begin{aligned}
\min_{x(\cdot),u(\cdot)} \quad & g(x(T)) \\
& + \left\langle \Pi_0^{-1} x(t_0), x(t_0) \right\rangle + \langle \Sigma_T x(T), x(T) \rangle \\
& \quad + \int_{t_0}^{T} \|u(\tau)\|^2 d\tau + \int_{t_0}^{T} \left\langle H^* R^{-1} H x(\tau), x(\tau) \right\rangle d\tau \\
\text{s.t.} \quad & \frac{d}{d\tau} x = F x(\tau) + G Q^{\frac{1}{2}} u(\tau), \text{ a.e. in } [t_0, T]
\end{aligned}
$$

What's the kernel?

#1 Where's ~~Waldo/Charlie~~ the kernel? For optimal control

Time-varying state-constrained LQ optimal control

$$\min_{x(\cdot),u(\cdot)} \quad g(x(T)) \qquad \rightarrow L(x(t_j)_{j\in[J]})$$

$$+ \left\langle \Pi_0^{-1} x(t_0), x(t_0) \right\rangle + \langle \Sigma_T x(T), x(T) \rangle$$

$$+ \int_{t_0}^{T} \|u(\tau)\|^2 d\tau + \int_{t_0}^{T} \left\langle H^* R^{-1} H x(\tau), x(\tau) \right\rangle d\tau \qquad \rightarrow \|x(\cdot)\|^2_{\mathcal{S}^x_{[t_0,T]}}$$

$$\text{s.t.} \quad \frac{d}{d\tau} x = F x(\tau) + G Q^{\frac{1}{2}} u(\tau), \text{ a.e. in } [t_0, T] \qquad \rightarrow x(\cdot) \in \mathcal{S}^x_{[t_0,T]}$$

What's the kernel? It is the one describing the Sobolev-like space of trajectories

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

#2 Where's ~~Waldo/Charlie~~ the kernel? For Kalman estimation

Continuous-time estimation problem (smoothing/filtering) over GPs with linear SDE

$$dx(t) = Fx(t)dt + Gdw(t), \qquad x(t_0) = \xi, \tag{2}$$

$$dy(t) = Hx(t)dt + db(t), \qquad y(t_0) = 0. \tag{3}$$

Problem: Estimate $x(s)$ with the σ-algebra $\mathcal{Y}^T = \sigma(y(\tau), 0 \leq \tau \leq T)$ by (linear) minimum mean square estimator, a.k.a. the minimum variance linear estimator

$$\hat{x}(s|T) = E[x(s)|\mathcal{Y}^T] = x_S(s|T) := \bar{x}(s) + \int_{t_0}^{T} S_s(t|T)dy(t). \tag{4}$$

$$\epsilon_S(s|T) := x(s) - x_S(s|T) = x(s) - \int_{t_0}^{T} S_s(t|T)dy(t). \tag{5}$$

$$\hat{S}_s(\cdot|T) \in \operatorname*{argmin}_{S(\cdot|T)} \Gamma_S(s|T) = \mathbb{E}[\epsilon_S(s|T)(\epsilon_S(s|T))^*]. \tag{6}$$

The kernel is the covariance of $\epsilon_{\hat{S}_s}(\cdot|T)$ and we have $\hat{S}_s(t|T) = K(s,t|T)H^*R^{-1}$,

$$K(s,t|T) = \mathbb{E}[\epsilon_{\hat{S}_s}(s|T)(\epsilon_{\hat{S}_t}(t|T))^*] \in \mathcal{L}(\mathbb{R}^{n,*}, \mathbb{R}^n) \tag{7}$$

Reproducing kernel Hilbert spaces (RKHS)

A RKHS $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_{\mathcal{H}_K})$ is a Hilbert space of real-valued functions over a set $\mathfrak{T}$ if one of the following equivalent conditions is satisfied (Aronszajn's theorem)

$\exists\, k : \mathfrak{T} \times \mathfrak{T} \to \mathbb{R}$ s.t. $k_t(\cdot) = k(\cdot, t) \in \mathcal{H}_K$ and $F = \langle f(\cdot), k_t(\cdot) \rangle_{\mathcal{H}_K}$ for all $t \in \mathfrak{T}$ and $f \in \mathcal{H}_K$ (reproducing property)

k is s.t. $\mathbf{G} = [k\,(t_i, t_j)]_{i,j=1}^{n} \succcurlyeq 0$ and $\mathcal{H}_K := \overline{\text{span}(\{k_t(\cdot)\}_{t \in \mathfrak{T}})}$, i.e. the completion for the pre-scalar product $\langle k_t(\cdot), k_s(\cdot) \rangle_{k,0} = k(t, s)$

Loeve's theorem: a kernel is p.s.d. if and only if it is the (proper) covariance of second-order stochastic process

Two essential tools for computations

Representer Theorem

Let $L : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$, strictly increasing $\Omega : \mathbb{R}_+ \to \mathbb{R}$, and

$$\bar{f} \in \underset{f \in \mathcal{H}_K}{\operatorname{argmin}}\, L\left((f(t_n))_{n\in[N]}\right) + \Omega\left(\|f\|_k\right)$$

Then $\exists\, (a_n)_{n\in[N]} \in \mathbb{R}^N$ s.t. $\bar{f}(\cdot) = \sum_{n\in[N]} a_n k(\cdot, t_n)$

$\hookrightarrow$ Optimal solutions lie in a finite dimensional subspace of $\mathcal{H}_K$.

Finite number of evaluations $\implies$ finite number of coefficients

Kernel trick

$$\langle \sum_{n\in[N]} a_n k(\cdot, t_n), \sum_{m\in[M]} b_m k(\cdot, s_m) \rangle_{\mathcal{H}_K} = \sum_{n\in[N]} \sum_{m\in[M]} a_n b_m k(t_n, s_m)$$

$\hookrightarrow$ On this finite dimensional subspace, no need to know $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_{\mathcal{H}_K})$.

In a nutshell

- finding an RKHS somewhere allows for simpler computations
- in LQ optimal control, RKHSs come from vector spaces of trajectories
- in linear estimation, kernels come from covariances of optimal errors

A detour through estimators - stochastic versions

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

$$\min_{\hat{X}\in L^2(\Omega\times\mathcal{T}),\,\mathcal{S}\in\mathcal{L}(L^2,L^2),\,\hat{X}=\mathcal{S}Y} \mathbb{E}((X-\hat{X})^\top(X-\hat{X})). \tag{MVLE}$$

Gaussian process regression for real-valued x introduces the canonical congruence ψ_Y between Y and its RKHS $\mathcal{H}_Y$, i.e. $\psi_Y(v^\top Y_t)(s)=\mathbb{E}[Y_sY_t^\top]v$ for all $v\in\mathbb{R}^m$,

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

By convex duality, introducing a process Λ which acts as a Lagrange multiplier,

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

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

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

A detour through estimators - the deterministic viewpoint

Note that for any deterministic $\bar{\lambda}(\cdot) \in L^2([t_0, T], \mathbb{R}^{n,*})$, $\hat{\mathcal{S}}$ also minimizes

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

For $Y = \mathcal{H}X + Z$, defining the deterministic $v(\cdot) = \mathcal{S}^\top \bar{\lambda}(\cdot)$, $\hat{\mathcal{S}}$ minimizes a "control" problem

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

Expressing the maximum log-likelihood estimator for a realization $y(\cdot)$ of Y as

$$\min_{\hat{x}(\cdot) \in L^2(\mathcal{T}, \mathbb{R}^n)} \underbrace{\left\langle \hat{x}(\cdot) - \mathcal{C}_{XY} \mathcal{C}_Y^{-1} y(\cdot), \mathcal{C}_\epsilon^{-1} (\hat{x}(\cdot) - \mathcal{C}_{XY} \mathcal{C}_Y^{-1} y(\cdot)) \right\rangle_{L^2} + \left\langle y(\cdot), \mathcal{C}_Y^{-1} y(\cdot) \right\rangle_{L^2}}_{\left\langle \hat{x}(\cdot), \mathcal{C}_X^{-1} \hat{x}(\cdot) \right\rangle_{L^2} + \langle y(\cdot) - \mathcal{H}\hat{x}(\cdot), \mathcal{R}^{-1} (y(\cdot) - \mathcal{H}\hat{x}(\cdot)) \rangle_{L^2}}. \quad \text{(LSE)}$$

Justifying the (dual) optimal control

We start by expressing $\left\langle \Gamma_S(s|T)\bar{\lambda}, \bar{\lambda} \right\rangle$ more explicitly,

$$\left\langle \bar{\lambda}, \epsilon_{\hat{S}_s}(s|T) \right\rangle = \left\langle \bar{\lambda}, x(s) \right\rangle - \int_{t_0}^{T} \left\langle S_s^*(t|T)\bar{\lambda}, dy(t) \right\rangle = \left\langle \bar{\lambda}, x(s) \right\rangle - \int_{t_0}^{T} \left\langle S_s^*(t|T)\bar{\lambda}, Hx(t)dt + db(t) \right\rangle.$$

We seek an expression where x does not appear. We introduce the adjoint equation over λ_s

$$-\frac{d\lambda_s}{dt} = F^*\lambda_s(t) - H^* S_s^*(t|T)\bar{\lambda}, \quad \lambda_s(T) = \left| \begin{array}{ll} 0 & \text{if } s < T \\ \bar{\lambda} & \text{if } s = T \end{array} \right. , \quad \lambda_s(s) - \lambda_s(s^+) = \bar{\lambda}, \text{if } s < T, \tag{8}$$

Then a simple calculation shows that $\left\langle \bar{\lambda}, \epsilon_{\hat{S}_s}(s|T) \right\rangle =$ ... integration by parts ...

$$\left\langle \Gamma_S(s|T)\bar{\lambda}, \bar{\lambda} \right\rangle = \langle \Pi_0 \lambda_s(t_0), \lambda_s(t_0) \rangle + \int_{t_0}^{T} \langle GQG^*\lambda_s(t), \lambda_s(t) \rangle \, dt + \int_{t_0}^{T} \left\langle RS_s^*(t|T)\bar{\lambda}, S_s^*(t|T)\bar{\lambda} \right\rangle dt \tag{9}$$

A two-point boundary problem

More generally, beyond linear feedbacks $(S_s^*(\cdot|T)\bar{\lambda})$, for a general control input $v(\cdot)$ minimize

$$-\frac{d\lambda_s}{dt} = F^*\lambda_s(t) + H^*v(t),\ \lambda_s(T) = \left| \begin{array}{ll} 0 & \text{if } s < T \\ \bar{\lambda} & \text{if } s = T \end{array} \right., \quad \lambda_s(s) - \lambda_s(s^+) = \bar{\lambda}, \text{if } s < T;$$

$$J(v(\cdot)) = \langle \Pi_0 \lambda_s(t_0), \lambda_s(t_0)\rangle + \int_{t_0}^{T} \langle GQG^*\lambda_s(t), \lambda_s(t)\rangle\, dt + \int_{t_0}^{T} \langle Rv(t), v(t)\rangle\, dt.$$

We get $\widehat{S}_s^*(t|T)\bar{\lambda} = \hat{v}_s(t) = R^{-1}H\hat{\gamma}_s(t)$ and the Hamiltonian system:

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

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

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

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

Kernels and Riccati equations

$$\frac{d\hat{\mu}}{dt} = F\hat{\mu}(t) - GQG^*\hat{\nu}(t) + l_\mu(t); \qquad -\frac{d\hat{\nu}}{dt} = F^*\hat{\nu}(t) + H^*R^{-1}H\hat{\mu}(t) - l_\nu(t)$$
$$\hat{\mu}(t_0) = -\Pi_0\hat{\nu}(t_0), \qquad \hat{\nu}(T) = \Sigma_T\hat{\mu}(T).$$

Set $\hat{\mu}(t) = -\Pi(t)\hat{\nu}(t)$ and $\hat{\nu}(t) = \Sigma(t)\hat{\mu}(t)$? Such as:

$$-\frac{d}{dt}\Sigma = \Sigma(t)F + F^*\Sigma(t) - \Sigma(t)GQG^*\Sigma(t) + H^*R^{-1}H, \qquad \Sigma(T) = \Sigma_T;$$
$$\frac{d}{dt}\Pi = F\Pi(t) + \Pi(t)F^* - \Pi(t)H^*R^{-1}H\Pi(t) + GQG^*, \qquad \Pi(t_0) = \Pi_0.$$

Look for kernels K and Λ such that

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

A primal space of trajectories

Introduce semigroup of $F - GQG^*\Sigma(t)$ denoted $\Phi_{F,\Sigma}(s,t)$

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

$$\mathcal{S}^x_{[t_0,T]} = \{x(\cdot) \in H^1 \mid \exists\, u(\cdot) \in L^2 \text{ s.t. } \frac{d}{d\tau}x = Fx(\tau) + GQ^{\frac{1}{2}}u(\tau)\}$$

with square-norm

$$\|x(\cdot)\|^2_{\mathcal{S}^x_{[t_0,T]}} = \left\langle \Pi_0^{-1}x(t_0), x(t_0)\right\rangle + \langle \Sigma_T x(T), x(T)\rangle + \int_{t_0}^T \|u(\tau)\|^2 d\tau + \int_{t_0}^T \left\langle H^* R^{-1} H x(\tau), x(\tau)\right\rangle d\tau$$

has kernel

$$K(s,t|T) = \Phi_{F,\Sigma}(s,t_0)(\Pi_0^{-1} + \Sigma(t_0))^{-1}\Phi^*_{F,\Sigma}(t,t_0) + \int_{t_0}^{\min(s,t)} \Phi_{F,\Sigma}(s,\tau)GQG^*\Phi^*_{F,\Sigma}(t,\tau)d\tau$$

A dual space of information vectors

Introduce semigroup of $F - \Pi(s)H^* R^{-1} H$, denoted $\Phi_{F,\Pi}(s,t)$)

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

$$\mathcal{S}^{\lambda}_{[t_0,T]} = \{\lambda(\cdot) \in H^1 \mid v(\cdot) \in L^2 \ \text{ s.t. } -\frac{d}{dt}\lambda(t) = F^*\lambda(t) + H^* v(t)\}$$

with square-norm

$$\|\lambda(\cdot)\|^2_{\mathcal{S}^{\lambda}_{[t_0,T]}} = \langle \Pi_0 \lambda(t_0), \lambda(t_0)\rangle + \left\langle \Sigma_T^{-1}\lambda(T), \lambda(T)\right\rangle + \int_{t_0}^{T} \langle GQG^*\lambda(t), \lambda(t)\rangle\, dt + \int_{t_0}^{T} \langle Rv(t), v(t)\rangle\, dt$$

has kernel

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

Dual deterministic problems

$$L_x(x(\cdot)) := \int_{t_0}^{T} \|y(t) - Hx(t)\|^2_{R^{-1}} dt + \| G^{\ominus} \left(\frac{d}{dt} x - Fx(t) \right) \|^2_{Q^{\ominus}} dt$$

$$+ \left\langle \Pi_0^{\ominus} x(t_0), x(t_0) \right\rangle + \langle \Sigma_T x(T), x(T) \rangle$$

$$\int_{t_0}^{T} K(\cdot, t|T) H^* R^{-1} y(t) dt$$

$$= \underset{x(\cdot) \in \mathcal{S}^x_{[t_0,T]}}{\operatorname{argmin}} L_x(x(\cdot)) = \|R^{-1/2} y(\cdot)\|^2_{L^2} + \|x(\cdot)\|^2_{\mathcal{S}^x_{[t_0,T]}} - 2 \left\langle H^*(\cdot) R^{-1}(\cdot) y(\cdot), x(\cdot) \right\rangle_{L^2([t_0,T])}$$

Set $R^{-1} y(t) = \operatorname{proj}^{\|\cdot\|_R}_{\operatorname{Im} H}(y(t)) \in R^{-1} \operatorname{Im} H \; + \; \operatorname{proj}^{\|\cdot\|_R}_{\operatorname{Ker} H^*(t)}(y(t)) \in \operatorname{Ker} H^*(t)$.

$$\min_{\lambda(\cdot) \in \mathcal{S}^{\lambda}_{[t_0,T]}} \|\lambda(\cdot)\|^2_{\mathcal{S}^{\lambda}_{[t_0,T]}} - 2 \int_{t_0}^{T} \left\langle R \operatorname{proj}^{\|\cdot\|_R}_{\operatorname{Im} H}(y(t)), v(t) \right\rangle dt - \|R^{1/2} \operatorname{proj}^{\|\cdot\|_R}_{\operatorname{Ker} H^*(\cdot)}(y(\cdot))\|^2_{L^2}$$

$$= \int_{t_0}^{T} \|\operatorname{proj}^{\|\cdot\|_R}_{\operatorname{Im} H}(y(t)) - v(t)\|^2_R dt + \int_{t_0}^{T} \langle GQG^* \lambda(t), \lambda(t) \rangle \, dt + \langle \Pi_0 \lambda(t_0), \lambda(t_0) \rangle$$

$$+ \left\langle \Sigma_T^{\ominus} \lambda(T), \lambda(T) \right\rangle - \|R^{-1/2} y(\cdot)\|^2_{L^2}.$$

Summary

	Stochastic problems	Deterministic problems
primal variables	(i) Given Gaussian processes $(X_t)_{t\in[0,T]}$, $(Y_t)_{t\in[0,T]}$ 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 $y(\cdot)$ Solve primal optimal control problem over trajectories $\min_{x(\cdot)\in\mathcal{S}^x_{[t_0,T]}} \\|x(\cdot)\\|^2_{\mathcal{S}^x_{[t_0,T]}} - 2\left\langle H^*R^{-1}y(\cdot), x(\cdot)\right\rangle_{L^2([t_0,T])}$ Optimum: $\hat{x}(\cdot)=\int_{t_0}^T K(\cdot,t\|T)H^*R^{-1}y(t)dt$
dual variables	(iii) Given Gaussian processes $(X_t^\top)_{t\in[0,T]}$, $(Y_t)_{t\in[0,T]}$ 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 $y(\cdot)$ Solve dual optimal control problem over adjoint/information $\min_{\lambda(\cdot)\in\mathcal{S}^\lambda_{[t_0,T]}} \\|\lambda(\cdot)\\|^2_{\mathcal{S}^\lambda_{[t_0,T]}} - 2\int_{t_0}^T \left\langle R\,\mathrm{proj}^{\\|\cdot\\|_R}_{\mathrm{Im}\,H}(y(t)), v(t)\right\rangle dt$ Optimum: $\hat{v}(t)=-R^{-1}H\hat{x}(t)+\mathrm{proj}^{\\|\cdot\\|_R}_{\mathrm{Im}\,H}(y(t))$

Table: Vertical: permute min-max into max-min. Horizontal: set $dw(t)=u(t)dt$.

Kernels of LQ optimal control come from Hilbertian vector spaces of trajectories. For estimation problems, they are covariances of GPs. The “Dual”, deterministic and stochastic, nature of kernels leads to “duality” between optimal control and estimation in the LQ case.

Conclusion

In a nutshell

- finding an RKHS somewhere allows for simpler computations
- in LQ optimal control, RKHSs come from vector spaces of trajectories
- in linear estimation, kernels come from covariances of optimal errors

Objective:

- re-read known optimal control/estimation problems through kernel lens
- use nonlinear embeddings on the state, apply it to stochastic optimal control, and optimization over measures