Estimation and Control under Constraints through Kernel Methods

Pierre-Cyril Aubin

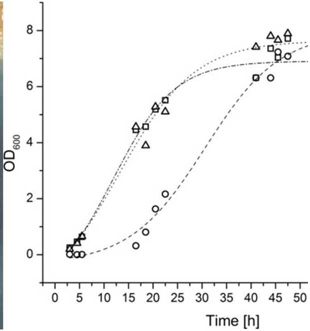
PhD Defence

Supervisor: Nicolas Petit

5th of July, 2021

What are shape constraints?

Nonparametric estimation



Shape constraints

- nonnegativity
  $f(x) \geq 0$
- directional monotonicity
  $\partial_i f(x) \geq 0$
- directional convexity
  $\partial^2_{i,i} f(x) \geq 0$

Side information/Requirements

$\hookrightarrow$ compensates small number of samples or excessive noise

Applied in many fields: Biology, Chemistry, Statistics, Economics,...
With many techniques: Isotonic regression, density estimation with splines,...

What are state constraints?

Optimal control

State constraints

- “avoid the wall“
  $x(t) \in [x_{low}, x_{high}]$
- “abide by the speed limit“
  $x'(t) \in [v_{low}, v_{high}]$
- “do not stress the pilot“
  $x"(t) \in [a_{low}, a_{high}]$

Physical constraints

$\hookrightarrow$ provides feasible trajectories in path-planning

Shape/state constraints are ubiquitous and handled through optimization:
in this thesis constraints are
affine pointwise inequality constraints over Hilbert spaces

Content of the thesis

Optimization in infinite dimensions with infinitely many constraints

- LQ optimal control is usually solved approximately through time discretization, whereas state constraints are theoretically difficult
- kernel methods only provide exact numerical solutions through representer theorems for finitely many constraints

Challenges to tackle

- handle infinitely many constraints in kernel methods with guarantees
- apply kernel methods to state-constrained LQ optimal control

Contributions of this thesis

- use finite coverings of compact sets in infinite dimensions to tighten infinitely many constraints by finitely many constraints of another type
- identify the LQ reproducing kernel corresponding to LQ optimal control

Table of Contents

1. Finding the RKHS of LQ optimal control
2. Tightening infinitely many constraints through finite coverings
3. Apply the kernel-based constraint tightening to LQ optimal control

This talk summarizes

- *Kernel Regression for Vehicle Trajectory Reconstruction under Speed and Inter-vehicular Distance Constraints*, Aubin, Petit and Szabó, **IFAC WC 2020**
- *Hard Shape-Constrained Kernel Machines*, Aubin and Szabó, **NeurIPS, 2020**
- *Linearly-constrained Linear Quadratic Regulator from the viewpoint of kernel methods*, Aubin, **SICON, 2021** (to appear)
- *Interpreting the dual Riccati equation through the LQ reproducing kernel*, Aubin, **CRM, 2021**

Time-varying state-constrained LQ optimal control

$$\min_{\mathbf{z}(\cdot),\mathbf{u}(\cdot)} \quad \chi_{\mathbf{z}_0}(\mathbf{z}(t_0)) + g(\mathbf{z}(T))$$

$$+\mathbf{z}(t_{ref})^\top \mathbf{J}_{ref}\mathbf{z}(t_{ref}) + \int_{t_0}^{T} \left[\mathbf{z}(t)^\top \mathbf{Q}(t)\mathbf{z}(t) + \mathbf{u}(t)^\top \mathbf{R}(t)\mathbf{u}(t)\right] \mathrm{d}t$$

$$\text{s.t.} \quad \mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z}(t) + \mathbf{B}(t)\mathbf{u}(t), \text{ a.e. in } [t_0, T],$$

$$\mathbf{c}_i(t)^\top \mathbf{z}(t) \leq d_i(t), \forall\, t \in \mathcal{T}_c, \forall\, i \in [\mathcal{I}] = [\![1, \mathcal{I}]\!],$$

- state $\mathbf{z}(t) \in \mathbb{R}^Q$, control $\mathbf{u}(t) \in \mathbb{R}^P$,
- reference time $t_{ref} \in [t_0, T]$, set of constraint times $\mathcal{T}_c \subset [t_0, T]$,
- $\mathbf{A}(\cdot) \in L^1(t_0, T)$, $\mathbf{B}(\cdot) \in L^2(t_0, T)$, $\mathbf{Q}(\cdot) \in L^1(t_0, T)$, $\mathbf{R}(\cdot) \in L^2(t_0, T)$,
- $\mathbf{Q}(t) \succcurlyeq 0$ and $\mathbf{R}(t) \succcurlyeq r\mathrm{Id}_M$ $(r > 0)$, $\mathbf{c}_i(\cdot), d_i(\cdot) \in C^0(t_0, T)$, $\mathbf{J}_{ref} \succ \mathbf{0}$,
- lower-semicontinuous terminal cost $g : \mathbb{R}^Q \to R \cup \{\infty\}$, indicator function $\chi_{\mathbf{z}_0}$,
- $\mathbf{z}(\cdot) : [t_0, T] \to \mathbb{R}^Q$ absolutely continuous, $\mathbf{R}(\cdot)^{1/2}\mathbf{u}(\cdot) \in L^2([t_0, T])$

Time-varying state-constrained LQ optimal control

$$\min_{\mathbf{z}(\cdot),\mathbf{u}(\cdot)} \quad \chi_{\mathbf{z}_0}(\mathbf{z}(t_0)) + g(\mathbf{z}(T)) \qquad \to L(\mathbf{z}(t_j)_{j\in[J]})$$

$$+\mathbf{z}(t_{ref})^\top \mathbf{J}_{ref}\mathbf{z}(t_{ref}) + \int_{t_0}^{T} \left[\mathbf{z}(t)^\top \mathbf{Q}(t)\mathbf{z}(t) + \mathbf{u}(t)^\top \mathbf{R}(t)\mathbf{u}(t)\right] \mathrm{d}t \to \|\mathbf{z}(\cdot)\|_{\mathcal{S}}^2$$

$$\text{s.t.} \quad \mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z}(t) + \mathbf{B}(t)\mathbf{u}(t), \text{ a.e. in } [t_0, T],$$

$$\mathbf{c}_i(t)^\top \mathbf{z}(t) \le d_i(t), \forall\, t \in \mathcal{T}_c, \forall\, i \in [\mathcal{I}] = [\![1, \mathcal{I}]\!],$$

- state $\mathbf{z}(t) \in \mathbb{R}^Q$, control $\mathbf{u}(t) \in \mathbb{R}^P$,
- reference time $t_{ref} \in [t_0, T]$, set of constraint times $\mathcal{T}_c \subset [t_0, T]$,
- $\mathbf{A}(\cdot) \in L^1(t_0, T)$, $\mathbf{B}(\cdot) \in L^2(t_0, T)$, $\mathbf{Q}(\cdot) \in L^1(t_0, T)$, $\mathbf{R}(\cdot) \in L^2(t_0, T)$,
- $\mathbf{Q}(t) \succcurlyeq 0$ and $\mathbf{R}(t) \succcurlyeq r\mathrm{Id}_M$ $(r > 0)$, $\mathbf{c}_i(\cdot), d_i(\cdot) \in C^0(t_0, T)$, $\mathbf{J}_{ref} \succ \mathbf{0}$,
- lower-semicontinuous terminal cost $g : \mathbb{R}^Q \to R \cup \{\infty\}$, indicator function $\chi_{\mathbf{z}_0}$, "loss function" $L : (\mathbb{R}^Q)^J \to \mathbb{R} \cup \{\infty\}$,
- $\mathbf{z}(\cdot) : [t_0, T] \to \mathbb{R}^Q$ absolutely continuous, $\mathbf{R}(\cdot)^{1/2}\mathbf{u}(\cdot) \in L^2([t_0, T])$

LQ optimal control as optimization over vector spaces

Optimization over Hilbert space $\mathcal{F}$

$$\begin{aligned} \min_{\mathbf{f}(\cdot)} \quad & \mathcal{L}(\mathbf{f}(\cdot)) \\ \text{s.t.} \quad & \\ & \mathbf{f} \in \mathcal{F}, \\ & l_t(\mathbf{f}(\cdot)) \leq 0 \ , \forall t \in \mathcal{T}_c \end{aligned}$$

$\mathcal{L}(\mathbf{f}(\cdot)) := L(\mathbf{f}(x_j)_{j\in[J]}) + R(\|\mathbf{f}\|_{\mathcal{F}})$
$l_t : \mathcal{F} \to \mathbb{R}$, e.g. $\mathcal{F} = H^1(\mathbb{R}^d, \mathbb{R}^Q)$

Linear Quadratic Optimal Control

$$\begin{aligned} \min_{\mathbf{z}(\cdot) \in W^{1,1}, \mathbf{u}(\cdot) \in L^2} \quad & \mathcal{L}(\mathbf{z}(\cdot), \mathbf{u}(\cdot)) \\ \text{s.t.} \quad & \\ & \mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z} + \mathbf{B}(t)\mathbf{u}, \text{a.e.}\, t \in [t_0, T], \\ & \mathbf{c}_i(t)^\top \mathbf{z}(t) \leq d_i(t), \ \forall t \in \mathcal{T}_c, \forall\, i \in [\mathcal{I}] \end{aligned}$$

$$\mathcal{L}(\mathbf{z}(\cdot), \mathbf{u}(\cdot)) := L(\mathbf{z}(t_j)_{j\in[J]}) + \|\mathbf{Q}^{1/2}\mathbf{z}(\cdot)\|_{L^2}^2 + \|\mathbf{R}^{1/2}\mathbf{u}(\cdot)\|_{L^2}^2$$

LQ optimal control as optimization over vector spaces

Optimization over Hilbert space $\mathcal{F}$

$$\min_{\mathbf{f}(\cdot)} \quad \mathcal{L}(\mathbf{f}(\cdot))$$

$$\text{s.t.}$$

$$\mathbf{f} \in \mathcal{F},$$

$$l_t(\mathbf{f}(\cdot)) \leq 0 \;, \forall t \in \mathcal{T}_c$$

$\mathcal{L}(\mathbf{f}(\cdot)) := L(\mathbf{f}(x_j)_{j\in[J]}) + R(\|\mathbf{f}\|_{\mathcal{F}})$
$l_t : \mathcal{F} \to \mathbb{R}$, e.g. $\mathcal{F} = H^1(\mathbb{R}^d, \mathbb{R}^Q)$

Linear Quadratic Optimal Control

$$\min_{\mathbf{z}(\cdot) \in W^{1,1}, \mathbf{u}(\cdot) \in L^2} \quad \mathcal{L}(\mathbf{z}(\cdot), \mathbf{u}(\cdot))$$

$$\text{s.t.}$$

$$\mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z} + \mathbf{B}(t)\mathbf{u}, \text{a.e.} t \in [t_0, T],$$

$$\mathbf{c}_i(t)^\top \mathbf{z}(t) \leq d_i(t), \forall t \in \mathcal{T}_c, \forall\, i \in [\mathcal{I}]$$

$$\mathcal{L}(\mathbf{z}(\cdot), \mathbf{u}(\cdot)) := L(\mathbf{z}(t_j)_{j\in[J]}) + \|\mathbf{Q}^{1/2}\mathbf{z}(\cdot)\|_{L^2}^2 + \|\mathbf{R}^{1/2}\mathbf{u}(\cdot)\|_{L^2}^2$$

Approximately solvable through finite elements

Approximately solvable through time discretization

LQ optimal control as optimization over vector spaces

Optimization over Hilbert space $\mathcal{F}$

$$\min_{\mathbf{f}(\cdot)} \quad \mathcal{L}(\mathbf{f}(\cdot))$$

$$\text{s.t.}$$

$$\mathbf{f} \in \mathcal{F},$$

$$l_t(\mathbf{f}(\cdot)) \leq 0 \ , \forall t \in \mathcal{T}_c$$

$\mathcal{L}(\mathbf{f}(\cdot)) := L(\mathbf{f}(x_j)_{j\in[J]}) + R(\|\mathbf{f}\|_{\mathcal{F}})$
$l_t : \mathcal{F} \to \mathbb{R}$, e.g. $\mathcal{F} = H^1(\mathbb{R}^d, \mathbb{R}^Q)$

Linear Quadratic Optimal Control

$$\min_{\mathbf{z}(\cdot)\in W^{1,1}, \mathbf{u}(\cdot)\in L^2} \quad \mathcal{L}(\mathbf{z}(\cdot), \mathbf{u}(\cdot))$$

$$\text{s.t.}$$

$$\mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z} + \mathbf{B}(t)\mathbf{u}, \text{ a.e. } t \in [t_0, T],$$

$$\mathbf{c}_i(t)^\top \mathbf{z}(t) \leq d_i(t), \ \forall t \in \mathcal{T}_c, \forall\, i \in [\mathcal{I}]$$

$$\mathcal{L}(\mathbf{z}(\cdot), \mathbf{u}(\cdot)) := L(\mathbf{z}(t_j)_{j\in[J]}) + \|\mathbf{Q}^{1/2}\mathbf{z}(\cdot)\|^2_{L^2} + \|\mathbf{R}^{1/2}\mathbf{u}(\cdot)\|^2_{L^2}$$

Exactly solvable if $\mathcal{F}$ RKHS, $R \nearrow$, $l_t \in \operatorname{span}(\{\delta_x, \delta'_x, \dots\}_{x\in\mathbb{R}^d})$, $\mathcal{T}_c$ finite

Approximately solvable through time discretization

LQ optimal control as optimization over vector spaces

Optimization over Hilbert space $\mathcal{F}$

$$\min_{\mathbf{f}(\cdot)} \quad \mathcal{L}(\mathbf{f}(\cdot))$$

$$\text{s.t.}$$

$$\mathbf{f} \in \mathcal{F},$$

$$l_t(\mathbf{f}(\cdot)) \leq 0 \ , \forall t \in \mathcal{T}_c$$

$\mathcal{L}(\mathbf{f}(\cdot)) := L(\mathbf{f}(x_j)_{j\in[J]}) + R(\|\mathbf{f}\|_{\mathcal{F}})$
$l_t : \mathcal{F} \to \mathbb{R}$, e.g. $\mathcal{F} = H^1(\mathbb{R}^d, \mathbb{R}^Q)$

Linear Quadratic Optimal Control

$$\min_{\mathbf{z}(\cdot) \in W^{1,1}, \mathbf{u}(\cdot) \in L^2} \quad \mathcal{L}(\mathbf{z}(\cdot), \mathbf{u}(\cdot))$$

$$\text{s.t.}$$

$$\mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z} + \mathbf{B}(t)\mathbf{u}, \text{a.e.} t \in [t_0, T],$$

$$\mathbf{c}_i(t)^\top \mathbf{z}(t) \leq d_i(t), \forall t \in \mathcal{T}_c, \forall i \in [\mathcal{I}]$$

$$\mathcal{L}(\mathbf{z}(\cdot), \mathbf{u}(\cdot)) := L(\mathbf{z}(t_j)_{j\in[J]}) + \|\mathbf{Q}^{1/2}\mathbf{z}(\cdot)\|^2_{L^2} + \|\mathbf{R}^{1/2}\mathbf{u}(\cdot)\|^2_{L^2}$$

Exactly solvable if $\mathcal{F}$ RKHS, $R \nearrow$,
$l_t \in \text{span}(\{\delta_x, \delta'_x, \dots\}_{x\in\mathbb{R}^d})$, $\mathcal{T}_c$ finite

Exactly solvable?

Reproducing kernel Hilbert spaces (RKHS)

A RKHS $(\mathcal{F}_k, \langle \cdot, \cdot \rangle_{\mathcal{F}_k})$ is a Hilbert space of real-valued functions over a set $\mathcal{T}$ if one of the following equivalent conditions is satisfied [Aronszajn, 1950]

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

the topology of $(\mathcal{F}_k, \langle \cdot, \cdot \rangle_{\mathcal{F}_k})$ is stronger than pointwise convergence i.e. $\delta_t : f \in \mathcal{F}_k \mapsto f(t)$ is continuous for all $t \in \mathcal{T}$.

$$|f(t) - f_n(t)| = |\langle f - f_n, k_t \rangle_{\mathcal{F}_k}| \leq \|f - f_n\|_{\mathcal{F}_k} \|k_t\|_{\mathcal{F}_k} = \|f - f_n\|_{\mathcal{F}_k} \sqrt{k(t, t)}$$

Reproducing kernel Hilbert spaces (RKHS)

A RKHS $(\mathcal{F}_k, \langle \cdot, \cdot \rangle_{\mathcal{F}_k})$ is a Hilbert space of real-valued functions over a set $\mathcal{T}$ if one of the following equivalent conditions is satisfied [Aronszajn, 1950]

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

the topology of $(\mathcal{F}_k, \langle \cdot, \cdot \rangle_{\mathcal{F}_k})$ is stronger than pointwise convergence i.e. $\delta_t : f \in \mathcal{F}_k \mapsto f(t)$ is continuous for all $t \in \mathcal{T}$.

$$|f(t) - f_n(t)| = |\langle f - f_n, k_t \rangle_{\mathcal{F}_k}| \leq \|f - f_n\|_{\mathcal{F}_k} \|k_t\|_{\mathcal{F}_k} = \|f - f_n\|_{\mathcal{F}_k} \sqrt{k(t,t)}$$

k is s.t. $\exists\, \Phi_k : \mathcal{T} \to \mathcal{F}_k$ s.t. $k(t,s) = \langle \Phi_k(t), \Phi_k(s) \rangle_{\mathcal{F}_k}$, $\Phi_k(t) = k_t(\cdot)$

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

Two essential tools for computations

Representer Theorem (e.g. [Schölkopf et al., 2001a])

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

$$\bar{f} \in \underset{f \in \mathcal{F}_k}{\arg\min}\, L\left(\left(f(t_n)\right)_{n\in[N]}\right) + \Omega\left(\|f\|_k\right)$$

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 $\mathcal{F}_k$.

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

Kernel trick

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

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

Vector-valued reproducing kernel Hilbert space (vRKHS)

Definition (vRKHS)

Let $\mathcal{T}$ be a non-empty set. A Hilbert space $(\mathcal{F}_K, \langle \cdot, \cdot \rangle_K)$ of $\mathbb{R}^Q$-vector-valued functions defined on $\mathcal{T}$ is a vRKHS if there exists a matrix-valued kernel $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}^{Q \times Q}$ such that the reproducing property holds:

$$K(\cdot, t)\mathbf{p} \in \mathcal{F}_K, \quad \mathbf{p}^\top \mathbf{f}(t) = \langle \mathbf{f}, K(\cdot, t)\mathbf{p} \rangle_K, \quad \text{for } t \in \mathcal{T}, \mathbf{p} \in \mathbb{R}^Q, \mathbf{f} \in \mathcal{F}_K$$

Necessarily, K has a Hermitian symmetry: $K(s, t) = K(t, s)^\top$

There is a one-to-one correspondence between K and $(\mathcal{F}_K, \langle \cdot, \cdot \rangle_K)$ [Micheli and Glaunès, 2014], so changing $\mathcal{T}$ or $\langle \cdot, \cdot \rangle_K$ changes K.

For $\mathcal{T} \subset \mathbb{R}^d$, Sobolev spaces $\mathcal{H}^s(\mathcal{T}, \mathbb{R}^Q)$ satisfying $s > d/2$ are RKHSs. One can take $K(s, t) = k(s, t)\mathrm{Id}_Q$, with real-valued k such as

$$k_{\text{Gauss}}(t, s) = \exp\left(-\|t - s\|_{\mathbb{R}^d}^2 / (2\sigma^2)\right) \quad k_{\text{poly}}(t, s) = (1 + \langle t, s \rangle_{\mathbb{R}^d})^2.$$

Representer theorem in vRKHSs

Theorem (Representer theorem with constraints, P.-C. Aubin, 2021)

Let $(\mathcal{F}_K, \langle\cdot,\cdot\rangle_K)$ *be a vRKHS defined on a set* $\mathcal{T}$*. For a "loss"* $L : \mathbb{R}^{N_0} \to \mathbb{R} \cup \{+\infty\}$*, strictly increasing "regularizer"* $\Omega : \mathbb{R}_+ \to \mathbb{R}$*, and constraints* $d_i : \mathbb{R}^{N_i} \to \mathbb{R}$*, consider the optimization problem*

$$\begin{aligned} \bar{\mathbf{f}} \in \underset{\mathbf{f}\in\mathcal{F}_K}{\arg\min} \quad & L\left(\mathbf{c}_{0,1}^\top \mathbf{f}(t_{0,1}), \dots, \mathbf{c}_{0,N_0}^\top \mathbf{f}(t_{0,N_0})\right) + \Omega\left(\|\mathbf{f}\|_K\right) \\ \text{s.t.} \quad & \\ & \lambda_i \|\mathbf{f}\|_K \leq d_i(\mathbf{c}_{i,1}^\top \mathbf{f}(t_{i,1}), \dots, \mathbf{c}_{i,N_i}^\top \mathbf{f}(t_{i,N_i})), \ \forall\, i \in [\![1, P]\!]. \end{aligned}$$

Then there exists $\{\mathbf{p}_{i,m}\}_{m\in[\![1,N_i]\!]} \subset \mathbb{R}^Q$ *and* $\alpha_{i,m} \in \mathbb{R}$ *such that*

$$\bar{\mathbf{f}} = \sum_{i=0}^{P} \sum_{m=1}^{N_i} K(\cdot, t_{i,m})\mathbf{p}_{i,m} \text{ with } \mathbf{p}_{i,m} = \alpha_{i,m}\mathbf{c}_{i,m}.$$

Objective: Turn the state-constrained LQR into "KRR"

We have a vector space $\mathcal{S}$ of controlled trajectories $\mathbf{z}(\cdot) : [t_0, T] \to \mathbb{R}^Q$

$$\mathcal{S}_{[t_0,T]} := \{\mathbf{z}(\cdot) \,|\, \exists\, \mathbf{u}(\cdot) \in L^2(t_0, T) \text{ s.t. } \mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z}(t) + \mathbf{B}(t)\mathbf{u}(t) \text{ a.e. } \}$$

Given $\mathbf{z}(\cdot) \in \mathcal{S}_{[t_0,T]}$, for the pseudoinverse $\mathbf{B}(t)^\ominus$ of $\mathbf{B}(t)$, set

$$\mathbf{u}(t) := \mathbf{B}(t)^\ominus[\mathbf{z}'(t) - \mathbf{A}(t)\mathbf{z}(t)] \text{ a.e. in } [t_0, T].$$

$$\begin{aligned}\langle \mathbf{z}_1(\cdot), \mathbf{z}_2(\cdot)\rangle_{\mathcal{S}} &:= \mathbf{z}_1(t_{ref})^\top \mathbf{J}_{ref}\mathbf{z}_2(t_{ref}) \\ &\quad + \int_{t_0}^{T} \left[\mathbf{z}_1(t)^\top \mathbf{Q}(t)\mathbf{z}_2(t) + \mathbf{u}_1(t)^\top \mathbf{R}(t)\mathbf{u}_2(t)\right] \mathrm{d}t\end{aligned}$$

LQR for $\mathbf{Q} \equiv \mathbf{0}$, $\mathbf{R} \equiv \mathrm{Id}$

$$\min_{\substack{\mathbf{z}(\cdot)\in\mathcal{S} \\ \mathbf{u}(\cdot)\in L^2}} L(\mathbf{z}(t_j)_{j\in[J]}) + \|\mathbf{u}(\cdot)\|^2_{L^2(t_0,T)}$$

$$\mathbf{c}_i(t)^\top \mathbf{z}(t) \le d_i(t), \forall t \in \mathcal{T}_c, i \in [\mathcal{I}]$$

"KRR" (Kernel Ridge Regression)

$$\min_{\mathbf{z}(\cdot)\in\mathcal{S}} L(\mathbf{z}(t_j)_{j\in[J]}) + \|\mathbf{z}(\cdot)\|^2_{\mathcal{S}}$$

$$\mathbf{c}_i(t)^\top \mathbf{z}(t) \le d_i(t), \forall t \in \mathcal{T}_c, i \in [\mathcal{I}]$$

Is $(\mathcal{S}, \langle\cdot,\cdot\rangle_{\mathcal{S}})$ a RKHS?

Objective: Turn the state-constrained LQR into "KRR"

We have a vector space $\mathcal{S}$ of controlled trajectories $\mathbf{z}(\cdot) : [t_0, T] \to \mathbb{R}^Q$

$$\mathcal{S}_{[t_0,T]} := \{\mathbf{z}(\cdot) \,|\, \exists\, \mathbf{u}(\cdot) \in L^2(t_0, T) \text{ s.t. } \mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z}(t) + \mathbf{B}(t)\mathbf{u}(t) \text{ a.e. } \}$$

Given $\mathbf{z}(\cdot) \in \mathcal{S}_{[t_0,T]}$, for the pseudoinverse $\mathbf{B}(t)^{\ominus}$ of $\mathbf{B}(t)$, set

$$\begin{aligned}
\mathbf{u}(t) &:= \mathbf{B}(t)^{\ominus}[\mathbf{z}'(t) - \mathbf{A}(t)\mathbf{z}(t)] \;\text{ a.e. in } [t_0, T].\\
\langle \mathbf{z}_1(\cdot), \mathbf{z}_2(\cdot)\rangle_{\mathcal{S}} &:= \mathbf{z}_1(t_{ref})^\top \mathbf{J}_{ref}\, \mathbf{z}_2(t_{ref})\\
&\quad + \int_{t_0}^{T} \left[\mathbf{z}_1(t)^\top \mathbf{Q}(t)\mathbf{z}_2(t) + \mathbf{u}_1(t)^\top \mathbf{R}(t)\mathbf{u}_2(t)\right] \mathrm{d}t
\end{aligned}$$

Lemma (P.-C. Aubin, 2021)

$(\mathcal{S}_{[t_0,T]}, \langle\cdot,\cdot\rangle_{\mathcal{S}})$ *is a vRKHS over* $[t_0, T]$ *with uniformly continuous* $K(\cdot,\cdot;[t_0,T])$.

Splitting $\mathcal{S}_{[t_0,T]}$ into subspaces and identifying their kernels

It is hard to identify K, but take $\mathbf{Q} \equiv \mathbf{0}$, $\mathbf{R} \equiv \mathrm{Id}$, $t_{ref} = t_0$, $\mathbf{J}_{ref} = \mathrm{Id}$

$$\langle \mathbf{z}_1(\cdot), \mathbf{z}_2(\cdot) \rangle_{\mathcal{S}} := \mathbf{z}_1(t_0)^\top \mathbf{z}_2(t_0) + \int_{t_0}^{T} \mathbf{u}_1(t)^\top \mathbf{u}_2(t) \mathrm{d}t.$$

$$\mathcal{S}_0 := \{\mathbf{z}(\cdot) \,|\, \mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z}(t), \text{ a.e. in } [t_0, T]\} \qquad \|\mathbf{z}(\cdot)\|_{K_0}^2 = \|\mathbf{z}(t_0)\|^2$$

$$\mathcal{S}_u := \{\mathbf{z}(\cdot) \,|\, \mathbf{z}(\cdot) \in \mathcal{S} \text{ and } \mathbf{z}(t_0) = 0\} \qquad \|\mathbf{z}(\cdot)\|_{K_1}^2 = \|\mathbf{u}(\cdot)\|_{L^2(t_0,T)}^2.$$

As $\mathcal{S} = \mathcal{S}_0 \oplus \mathcal{S}_u$, $K = K_0 + K_1$.

Splitting $\mathcal{S}_{[t_0,T]}$ into subspaces and identifying their kernels

It is hard to identify K, but take $\mathbf{Q} \equiv \mathbf{0}$, $\mathbf{R} \equiv \mathrm{Id}$, $t_{ref} = t_0$, $\mathbf{J}_{ref} = \mathrm{Id}$

$$\langle \mathbf{z}_1(\cdot), \mathbf{z}_2(\cdot) \rangle_{\mathcal{S}} := \mathbf{z}_1(t_0)^\top \mathbf{z}_2(t_0) + \int_{t_0}^{T} \mathbf{u}_1(t)^\top \mathbf{u}_2(t) \mathrm{d}t.$$

$$\mathcal{S}_0 := \{\mathbf{z}(\cdot) \mid \mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z}(t), \text{ a.e. in } [t_0, T]\} \qquad \|\mathbf{z}(\cdot)\|_{K_0}^2 = \|\mathbf{z}(t_0)\|^2$$

$$\mathcal{S}_u := \{\mathbf{z}(\cdot) \mid \mathbf{z}(\cdot) \in \mathcal{S} \text{ and } \mathbf{z}(t_0) = 0\} \qquad \|\mathbf{z}(\cdot)\|_{K_1}^2 = \|\mathbf{u}(\cdot)\|_{L^2(t_0,T)}^2.$$

As $\mathcal{S} = \mathcal{S}_0 \oplus \mathcal{S}_u$, $K = K_0 + K_1$. Since $\dim(\mathcal{S}_0) = Q$, for $\mathbf{\Phi}_{\mathbf{A}}(t,s) \in \mathbb{R}^{Q \times Q}$ the state-transition matrix $s \to t$ of $\mathbf{z}'(\tau) = \mathbf{A}(\tau)\mathbf{z}(\tau)$

$$K_0(s,t) = \mathbf{\Phi}_{\mathbf{A}}(s,t_0)\mathbf{\Phi}_{\mathbf{A}}(t,t_0)^\top.$$

Splitting $\mathcal{S}_{[t_0,T]}$ into subspaces and identifying their kernels

It is hard to identify K, but take $\mathbf{Q} \equiv \mathbf{0}$, $\mathbf{R} \equiv \mathrm{Id}$, $t_{ref} = t_0$, $\mathbf{J}_{ref} = \mathrm{Id}$

$$\langle \mathbf{z}_1(\cdot), \mathbf{z}_2(\cdot)\rangle_{\mathcal{S}} := \mathbf{z}_1(t_0)^\top \mathbf{z}_2(t_0) + \int_{t_0}^{T} \mathbf{u}_1(t)^\top \mathbf{u}_2(t)\mathrm{d}t.$$

$$\mathcal{S}_0 := \{\mathbf{z}(\cdot) \,|\, \mathbf{z}'(t) = \mathbf{A}(t)\mathbf{z}(t), \text{ a.e. in } [t_0, T]\} \qquad \|\mathbf{z}(\cdot)\|_{K_0}^2 = \|\mathbf{z}(t_0)\|^2$$

$$\mathcal{S}_u := \{\mathbf{z}(\cdot) \,|\, \mathbf{z}(\cdot) \in \mathcal{S} \text{ and } \mathbf{z}(t_0) = 0\} \qquad \|\mathbf{z}(\cdot)\|_{K_1}^2 = \|\mathbf{u}(\cdot)\|_{L^2(t_0,T)}^2.$$

As $\mathcal{S} = \mathcal{S}_0 \oplus \mathcal{S}_u$, $K = K_0 + K_1$. Since $\dim(\mathcal{S}_0) = Q$, for $\mathbf{\Phi}_{\mathbf{A}}(t,s) \in \mathbb{R}^{Q\times Q}$ the state-transition matrix $s \to t$ of $\mathbf{z}'(\tau) = \mathbf{A}(\tau)\mathbf{z}(\tau)$

$$K_0(s,t) = \mathbf{\Phi}_{\mathbf{A}}(s,t_0)\mathbf{\Phi}_{\mathbf{A}}(t,t_0)^\top.$$

K_1 obtained using only the reproducing property and variation of constants

$$K_1(s,t) = \int_{t_0}^{\min(s,t)} \mathbf{\Phi}_{\mathbf{A}}(s,\tau)\mathbf{B}(\tau)\mathbf{B}(\tau)^\top \mathbf{\Phi}_{\mathbf{A}}(t,\tau)^\top \mathrm{d}\tau.$$

Examples: controllability Gramian/transversality condition

Steer a point from $(0, \mathbf{0})$ to $(T, \mathbf{z}_T)$, with e.g. $g(\mathbf{z}(T)) = \|\mathbf{z}_T - \mathbf{z}(T)\|_N^2$

Exact planning $(\mathbf{z}(T) = \mathbf{z}_T)$

$$\min_{\substack{\mathbf{z}(\cdot)\in\mathcal{S}\\ \mathbf{z}(0)=\mathbf{0}}} \quad \chi_{\mathbf{z}_T}(\mathbf{z}(T)) + \frac{1}{2}\|\mathbf{u}(\cdot)\|^2_{L^2(t_0,T)}$$

Relaxed planning $(\mathbf{g} \in \mathcal{C}^1$ convex$)$

$$\min_{\substack{\mathbf{z}(\cdot)\in\mathcal{S}\\ \mathbf{z}(0)=\mathbf{0}}} \quad g(\mathbf{z}(T)) + \frac{1}{2}\|\mathbf{u}(\cdot)\|^2_{L^2(t_0,T)}$$

$\mathbf{z}(0) = \mathbf{0} \Leftrightarrow \mathbf{z}(\cdot) \in \mathcal{S}_u$. By representer theorem: $\exists\, \mathbf{p}_T$, $\bar{\mathbf{z}}(\cdot) = K_1(\cdot, T)\mathbf{p}_T$

Controllability Gramian

$$K_1(T,T) = \int_0^T \mathbf{\Phi_A}(T,\tau)\mathbf{B}(\tau)\mathbf{B}(\tau)^\top \mathbf{\Phi_A}(T,\tau)^\top \mathrm{d}\tau$$

$\bar{\mathbf{z}}(T) = \mathbf{z}_T \Leftrightarrow \mathbf{z}_T \in \operatorname{Im}(K_1(T,T))$

Transversality Condition

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

Sufficient to take $\mathbf{p}_T = -\nabla g(\bar{\mathbf{z}}(T))$

Relation with the differential Riccati equation

Take $t_{ref} = T$, $\mathbf{J}_{ref} = \mathbf{J}_T \succ \mathbf{0}$. Let $J(t, T)$ be the solution of

$$\begin{aligned}
-\partial_1 \mathbf{J}(t, T) &= \mathbf{A}(t)^\top \mathbf{J}(t, T) + \mathbf{J}(t, T)\mathbf{A}(t) \\
&\quad - \mathbf{J}(t, T)\mathbf{B}(t)\mathbf{R}(t)^{-1}\mathbf{B}(t)^\top \mathbf{J}(t, T) + \mathbf{Q}(t), \\
\mathbf{J}(T, T) &= \mathbf{J}_T,
\end{aligned}$$

Theorem (P.-C. Aubin, 2021)

Let $K_{\text{diag}} : t_0 \in]-\infty, T] \mapsto K(t_0, t_0; [t_0, T])$. Then $K_{\text{diag}}(t_0) = \mathbf{J}(t_0, T)^{-1}$. More generally, $K(\cdot, t; [t_0, T])$ is given by a matrix Hamiltonian system for all $t \in [t_0, T]$

$$\begin{aligned}
\partial_1 K(s, t) &= \mathbf{A}(s) K(s, t) + \mathbf{B}(s)\mathbf{R}(s)^{-1}\mathbf{B}(s)^\top \begin{cases} \mathbf{\Pi}(s, t) + \mathbf{\Phi}_{\mathbf{A}}(t_0, s)^\top - \mathbf{\Phi}_{\mathbf{A}}(t, s)^\top, s \geq t, \\ \mathbf{\Pi}(s, t) + \mathbf{\Phi}_{\mathbf{A}}(t_0, s)^\top, s < t. \end{cases} \\
\partial_1 \mathbf{\Pi}(s, t) &= -\mathbf{A}(s)^\top \mathbf{\Pi}(s, t) + \mathbf{Q}(s) K(s, t), \\
\mathbf{\Pi}(t_0, t) &= -Id_N, \\
K(t, T) &= -\mathbf{J}_T^{-1}(\mathbf{\Pi}(T, t)^\top + \mathbf{\Phi}_{\mathbf{A}}(t, T) - \mathbf{\Phi}_{\mathbf{A}}(t_0, T)).
\end{aligned}$$

Relation with the differential Riccati equation

$$\bar{\mathbf{z}}(\cdot) := \underset{\mathbf{z}(\cdot)\in\mathcal{S}_{[t_0,T]}}{\arg\min} \quad \underbrace{\mathbf{z}(T)^\top \mathbf{J}_T\, \mathbf{z}(T) + \int_{t_0}^{T} [\mathbf{z}(t)^\top \mathbf{Q}(t)\mathbf{z}(t) + \mathbf{u}(t)^\top \mathbf{R}(t)\mathbf{u}(t)]\mathrm{d}t}_{\|\mathbf{z}(\cdot)\|_{\mathcal{S}}^2}$$

$$\text{s.t.}$$

$$\mathbf{z}(t_0) = \mathbf{z}_0,$$

Pontryagine's maximum principle (PMP)

$\mathbf{p}(t) = -\mathbf{J}(t,T)\bar{\mathbf{z}}(t)$ and

$\bar{\mathbf{u}}(t) = \mathbf{R}(t)^{-1}\mathbf{B}(t)^\top \mathbf{p}(t) = -\mathbf{R}(t)^{-1}\mathbf{B}(t)^\top \mathbf{J}(t,T)\bar{\mathbf{z}}(t) =: \mathbf{G}(t)\bar{\mathbf{z}}(t)$

$\hookrightarrow$ **online and differential** approach

Representer theorem from kernel methods

$\bar{\mathbf{z}}(t) = K(t, t_0; [t_0, T])\mathbf{p}_0$, with $\mathbf{p}_0 = K(t_0, t_0; [t_0, T])^{-1}\mathbf{z}_0 \in \mathbb{R}^Q$

$\hookrightarrow$ **offline and integral** approach ($\sim$ Green kernel in PDEs)

Numerical example: submarine in a cavern

Original control problem

$$\min_{z(\cdot)\in W^{2,2},u(\cdot)\in L^2} \quad \int_0^1 |u(t)|^2 \mathrm{d}t$$

$$\text{s.t.}$$

$$z(0) = 0, \quad \dot{z}(0) = 0,$$

$$\ddot{z}(t) = -\dot{z}(t) + u(t), \ \forall t \in [0,1],$$

$$z(t) \in [z_{\text{low}}(t), z_{\text{up}}(t)], \ \forall\, t \in [0,1].$$

Numerical example: submarine in a cavern

Original control problem

$$\min_{z(\cdot)\in W^{2,2},u(\cdot)\in L^2} \quad \int_0^1 |u(t)|^2 \mathrm{d}t$$

s.t.

$$z(0) = 0, \quad \dot{z}(0) = 0,$$

$$\ddot{z}(t) = -\dot{z}(t) + u(t), \forall t \in [0,1],$$

$$z(t) \in [z_{\mathrm{low}}(t), z_{\mathrm{up}}(t)], \forall\, t \in [0,1].$$

Rewriting in standard form

$$\min_{\mathbf{z}(\cdot)\in W^{1,2},u(\cdot)\in L^2} \quad \int_0^1 |u(t)|^2 \mathrm{d}t$$

s.t.

$$\mathbf{z}(0) = 0,$$

$$\mathbf{z}'(t) \overset{\text{a.e.}}{=} \mathbf{A}\mathbf{z}(t) + \mathbf{B}u(t),$$

$$z_1(t) \in [z_{\mathrm{low}}(t), z_{\mathrm{up}}(t)], \forall\, t \in [0,1]$$

$$\mathbf{z} = \begin{pmatrix} z \\ \dot{z} \end{pmatrix}, \mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Numerical example: submarine in a cavern

RKHS regression

$$\min_{\mathbf{z}(\cdot)\in\mathcal{S}_u} \quad \|\mathbf{z}(\cdot)\|_{K_1}^2$$
$$\text{s.t.}$$
$$z_1(t)\in[z_{\text{low}}(t),z_{\text{up}}(t)],\ \forall\, t\in[0,1]$$

Rewriting in standard form

$$\min_{\mathbf{z}(\cdot)\in W^{1,2},u(\cdot)\in L^2} \quad \int_0^1 |u(t)|^2 \mathrm{d}t$$
$$\text{s.t.}$$
$$\mathbf{z}(0)=0,$$
$$\mathbf{z}'(t)\overset{\text{a.e.}}{=}\mathbf{A}\mathbf{z}(t)+\mathbf{B}u(t),$$
$$z_1(t)\in[z_{\text{low}}(t),z_{\text{up}}(t)],\ \forall\, t\in[0,1]$$

$$\mathcal{S}_u := \{\mathbf{z}(\cdot)\,|\,\mathbf{z}(\cdot)\in\mathcal{S} \text{ and } \mathbf{z}(0)=0\} \quad \|\mathbf{z}(\cdot)\|_{K_1}^2 = \|\mathbf{u}(\cdot)\|_{L^2(0,1)}^2.$$

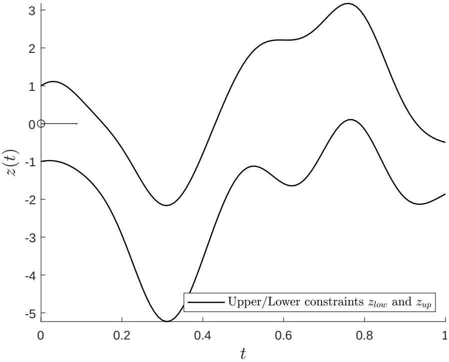
Numerical example: submarine in a cavern

RKHS regression

$$\min_{\mathbf{z}(\cdot) \in \mathcal{S}_u} \quad \|\mathbf{z}(\cdot)\|_{K_1}^2$$

s.t.

$$z_1(t) \in [z_{\text{low}}(t), z_{\text{up}}(t)], \ \forall\, t \in [0, 1]$$



$$\mathcal{S}_u := \{\mathbf{z}(\cdot) \mid \mathbf{z}(\cdot) \in \mathcal{S} \text{ and } \mathbf{z}(0) = 0\} \quad \|\mathbf{z}(\cdot)\|_{K_1}^2 = \|\mathbf{u}(\cdot)\|_{L^2(0,1)}^2.$$

$$K_1(s,t) = \int_0^{\min(s,t)} \mathbf{e}^{(s-\tau)\mathbf{A}} \mathbf{B}\mathbf{B}^\top \mathbf{e}^{(t-\tau)\mathbf{A}^\top} \mathrm{d}\tau$$

Numerical example: submarine in a cavern

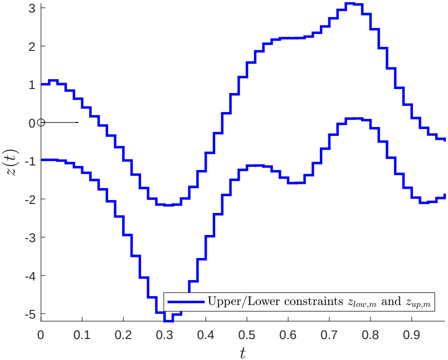
RKHS regression

$$\min_{\mathbf{z}(\cdot)\in\mathcal{S}_u} \quad \|\mathbf{z}(\cdot)\|_{K_1}^2$$

$$\text{s.t.}$$

$$z_1(t) \in [z_{\mathrm{low},m}, z_{\mathrm{up},m}],$$

$$\forall\, t \in [t_m - \delta_m, t_m + \delta_m],\ \forall m \in [M]$$



$$\mathcal{S}_u := \{\mathbf{z}(\cdot)\,|\,\mathbf{z}(\cdot)\in\mathcal{S} \text{ and } \mathbf{z}(0)=0\} \quad \|\mathbf{z}(\cdot)\|_{K_1}^2 = \|\mathbf{u}(\cdot)\|_{L^2(0,1)}^2.$$

$$K_1(s,t) = \int_0^{\min(s,t)} \mathbf{e}^{(s-\tau)\mathbf{A}}\mathbf{B}\mathbf{B}^\top \mathbf{e}^{(t-\tau)\mathbf{A}^\top}\mathrm{d}\tau$$

Numerical example: submarine in a cavern

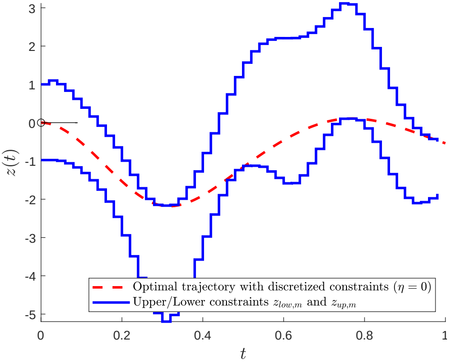
RKHS regression

$$\min_{\mathbf{z}(\cdot)\in\mathcal{S}_u} \quad \|\mathbf{z}(\cdot)\|_{K_1}^2$$

$$\text{s.t.}$$

$$z_1(t_m) \in [z_{\text{low},m}, z_{\text{up},m}],$$

$$\cancel{\forall t \in [t_m-\delta_m, t_m+\delta_m]}, \forall m \in [M]$$



$$\bar{\mathbf{z}}(\cdot) = \sum_{m=1}^{M} K_1(\cdot, t_m)\mathbf{p}_m = \sum_{m=1}^{M} \alpha_m K_1(\cdot, t_m)\mathbf{e}_1$$

$$K_1(s,t) = \int_0^{\min(s,t)} \mathbf{e}^{(s-\tau)\mathbf{A}}\mathbf{B}\mathbf{B}^\top \mathbf{e}^{(t-\tau)\mathbf{A}^\top} \mathrm{d}\tau$$

Numerical example: submarine in a cavern

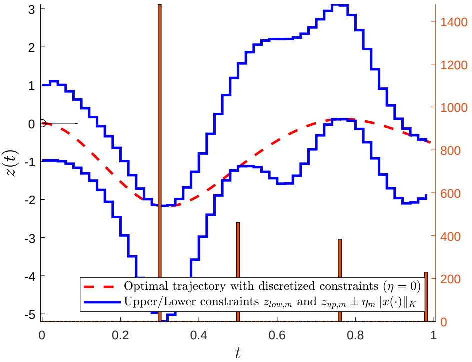
Quadratic programming

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^M} \sum_{n,m=1}^{M} \alpha_n \alpha_m \mathbf{e}_1^\top K_1(t_n, t_m)\mathbf{e}_1$$

s.t.

$$\sum_{n=1}^{M} \alpha_n K_1(t_m, t_n)\mathbf{e}_1 \in [z_{\mathrm{low},m}, z_{\mathrm{up},m}],$$

$$\cancel{\forall t \in [t_m - \delta_m, t_m + \delta_m]}, \forall m \in [M]$$



$$\bar{\mathbf{z}}(\cdot) = \sum_{m=1}^{M} K_1(\cdot, t_m)\mathbf{p}_m = \sum_{m=1}^{M} \alpha_m K_1(\cdot, t_m)\mathbf{e}_1$$

$$K_1(s,t) = \int_0^{\min(s,t)} \mathbf{e}^{(s-\tau)\mathbf{A}} \mathbf{B}\mathbf{B}^\top \mathbf{e}^{(t-\tau)\mathbf{A}^\top} \mathrm{d}\tau$$

Numerical example: submarine in a cavern

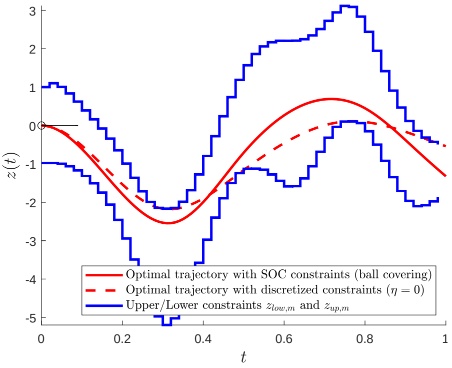
RKHS regression

$$\min_{\mathbf{z}(\cdot)\in\mathcal{S}_u} \quad \|\mathbf{z}(\cdot)\|_{K_1}^2$$

s.t.

$$z_1(t_m) \in [z_{\text{low},m}, z_{\text{up},m}] \pm \eta_m \|\mathbf{z}(\cdot)\|_{K_1},$$

$$\cancel{\forall t \in [t_m - \delta_m, t_m + \delta_m]}, \forall m \in [M]$$



$$\bar{\mathbf{z}}(\cdot) = \sum_{m=1}^{M} K_1(\cdot, t_m)\mathbf{p}_m = \sum_{m=1}^{M} \alpha_m K_1(\cdot, t_m)\mathbf{e}_1$$

$$K_1(s,t) = \int_0^{\min(s,t)} \mathbf{e}^{(s-\tau)\mathbf{A}}\mathbf{B}\mathbf{B}^\top \mathbf{e}^{(t-\tau)\mathbf{A}^\top} \mathrm{d}\tau$$

Table of Contents

1. Finding the RKHS of LQ optimal control
2. Tightening infinitely many constraints through finite coverings
3. Apply the kernel-based constraint tightening to LQ optimal control

Problem statement in machine learning terms

For simplicity, we consider a real-valued kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, for $\mathcal{X} \subset \mathbb{R}^d$.

Given points $(x_n)_{n\in[N]} \in \mathcal{X}^N$, a *loss* $L : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$, a *regularizer* $R : \mathbb{R}_+ \to \mathbb{R}$. Consider

$$\begin{aligned} \bar{f} \in \underset{f \in \mathcal{F}_k}{\arg\min}\ & \mathcal{L}(f) = L\left((f(x_n))_{n\in[N]}\right) + R\left(\|f\|_{\mathcal{F}_k}\right) \\ \text{s.t.}\ & \\ & b_i \leq D_i f(x), \quad \forall\, x \in \mathcal{K}_i, \forall i \in [\mathcal{I}] = [\![1, \mathcal{I}]\!], \end{aligned}$$

where $\mathcal{F}_k$ is a RKHS of smooth functions from $\mathcal{X}$ to $\mathbb{R}$, D_i is a differential operator ($D_i = \sum_j \gamma_j \partial^{r_j}$), $b_i \in \mathbb{R}$ is a lower bound, $\mathcal{K}_i$ is compact.

For non-finite $\mathcal{K}_i$, we have an infinite number of constraints!
$\hookrightarrow$ No representer theorem to work in finite dimensions!

How can we make this optimization problem computationally tractable?

Dealing with an infinite number of constraints: an overview

$$\bar{f} \in \underset{f \in \mathcal{F}_k}{\arg\min}\, \mathcal{L}(f) \text{ s.t. } "b_i \leq D_i f(x),\ \forall\, x \in \mathcal{K}_i,\ \forall i \in [\mathcal{I}]",\ \mathcal{K}_i \text{ non-finite}$$

Relaxing

- Discretize constraint at “virtual“ samples $\{\tilde{x}_{i,m}\}_{m \leq M} \subset \mathcal{K}_i$,
  $\hookrightarrow$ no guarantees out-of-samples [Agrell, 2019, Takeuchi et al., 2006]
- Add constraint-inducing penalty, $R_{\text{cons}}(f) = -\lambda \int_{\mathcal{K}_i} \min(0, D_i f(x) - b_i)\mathrm{d}x$
  $\hookrightarrow$ no guarantees, changes the problem objective [Brault et al., 2019]

Tightening

- Replace $\mathcal{F}$ by algebraic subclass of functions satisfying the constraints
  $\hookrightarrow$ hard to stack constraints, $\Phi(x)^\top A \Phi(x)$ [Marteau-Ferey et al., 2020]
- **Our solution:** discretize $\mathcal{K}_i$ but replace b_i using kernel theory

Deriving SOC constraints through continuity moduli

Take $\delta \geq 0$ and x s.t. $\|x - \tilde{x}_m\| \leq \delta$

$$\begin{aligned} |Df(x) - Df(\tilde{x}_m)| &= |\langle f(\cdot), D_x k(x,\cdot) - D_x k(\tilde{x}_m,\cdot)\rangle_k| \\ &\leq \|f(\cdot)\|_k \underbrace{\sup_{\{x \,|\, \|x-\tilde{x}_m\|\leq\delta\}} \|D_x k(x,\cdot) - D_x k(\tilde{x}_m,\cdot)\|_k}_{\eta_m(\delta)} \end{aligned}$$

$$\omega_m(Df,\delta) := \sup_{\{x \,|\, \|x-\tilde{x}_m\|\leq\delta\}} |Df(x) - Df(\tilde{x}_m)| \leq \eta_m(\delta)\|f(\cdot)\|_k$$

For a covering $\mathcal{K} = \bigcup_{m\in[M]} \mathbb{B}_{\mathcal{X}}(\tilde{x}_m, \delta_m)$

$$\text{“}b \leq Df(x),\ \forall x \in \mathcal{K}\text{“} \Leftrightarrow \text{“}b + \omega_m(Df,\delta) \leq Df(\tilde{x}_m),\ \forall m \in [M]\text{“}$$

Deriving SOC constraints through continuity moduli

Take $\delta \geq 0$ and x s.t. $\|x - \tilde{x}_m\| \leq \delta$

$$\begin{aligned}|Df(x) - Df(\tilde{x}_m)| &= |\langle f(\cdot), D_x k(x,\cdot) - D_x k(\tilde{x}_m,\cdot)\rangle_k| \\ &\leq \|f(\cdot)\|_k \underbrace{\sup_{\{x \,|\, \|x-\tilde{x}_m\|\leq\delta\}} \|D_x k(x,\cdot) - D_x k(\tilde{x}_m,\cdot)\|_k}_{\eta_m(\delta)}\end{aligned}$$

$$\omega_m(Df,\delta) := \sup_{\{x \,|\, \|x-\tilde{x}_m\|\leq\delta\}} |Df(x) - Df(\tilde{x}_m)| \leq \eta_m(\delta)\|f(\cdot)\|_k$$

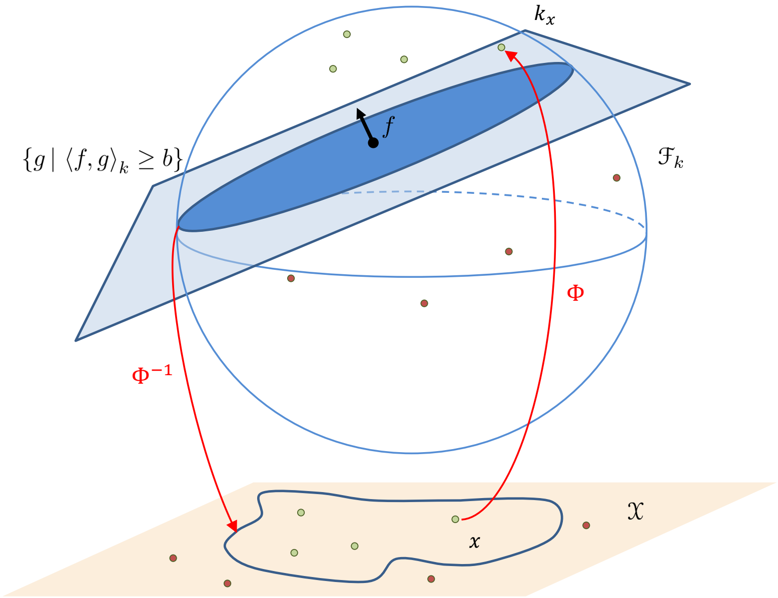
For a covering $\mathcal{K} \subset \bigcup_{m\in[M]} \mathbb{B}_{\mathcal{X}}(\tilde{x}_m, \delta_m)$

$$\begin{aligned}\text{“}b \leq Df(x),\ \forall x \in \mathcal{K}\text{“} &\Leftarrow \text{“}b + \omega_m(Df,\delta) \leq Df(\tilde{x}_m),\ \forall m \in [M]\text{“} \\ &\Leftarrow \text{“}b + \eta_m(\delta)\|f(\cdot)\| \leq Df(\tilde{x}_m),\ \forall m \in [M]\end{aligned}$$

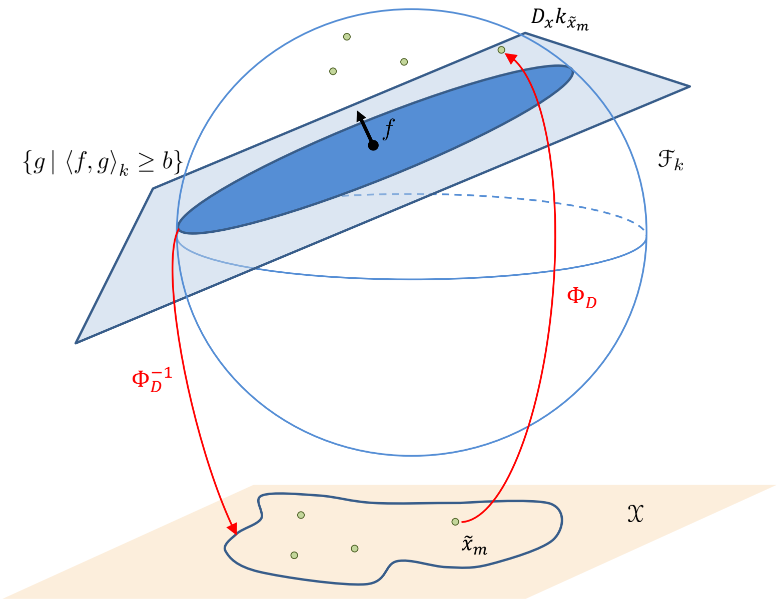
Second-Order Cone (SOC) constraints: $\{f \,|\, \|Af + b\|_K \leq c^\top f + d\}$

Since the kernel is smooth, $\delta \to 0$ gives $\eta_m(\delta) \to 0$.

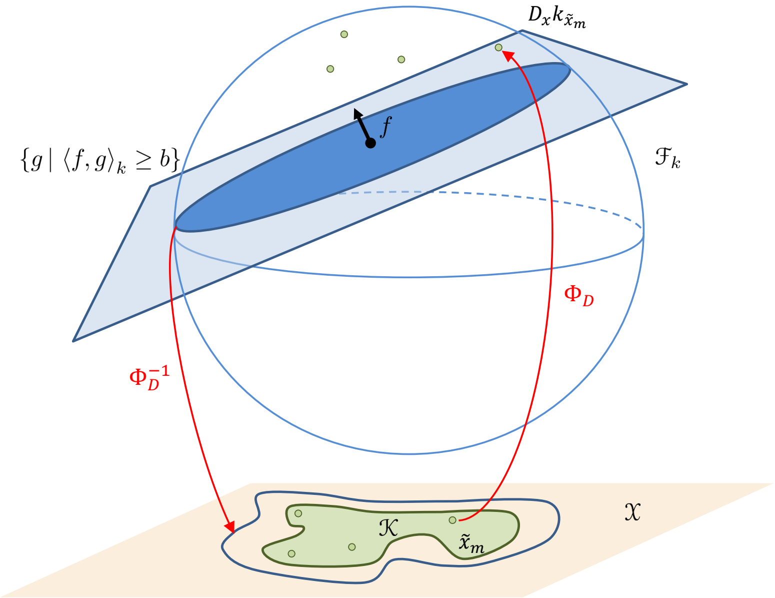
There is also a geometrical interpretation for this choice of η_m.



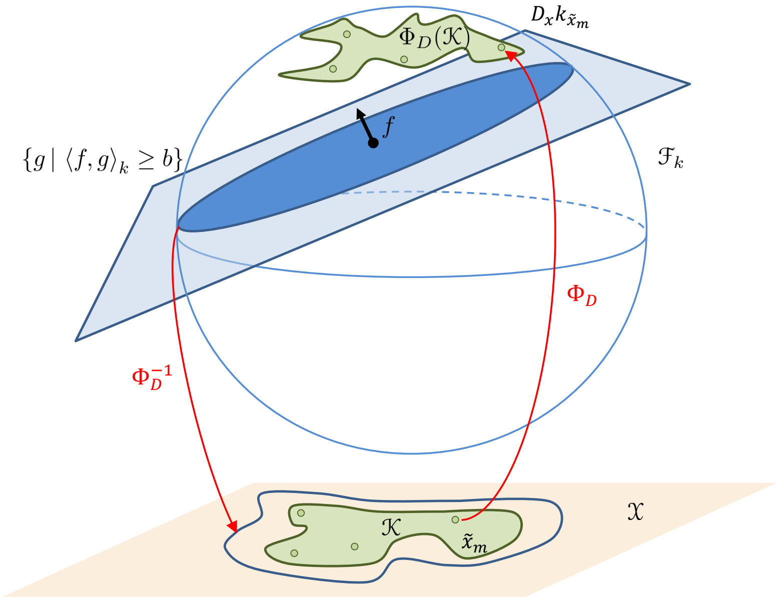
Support Vector Machine (SVM) is about separating red and green points by blue hyperplane.



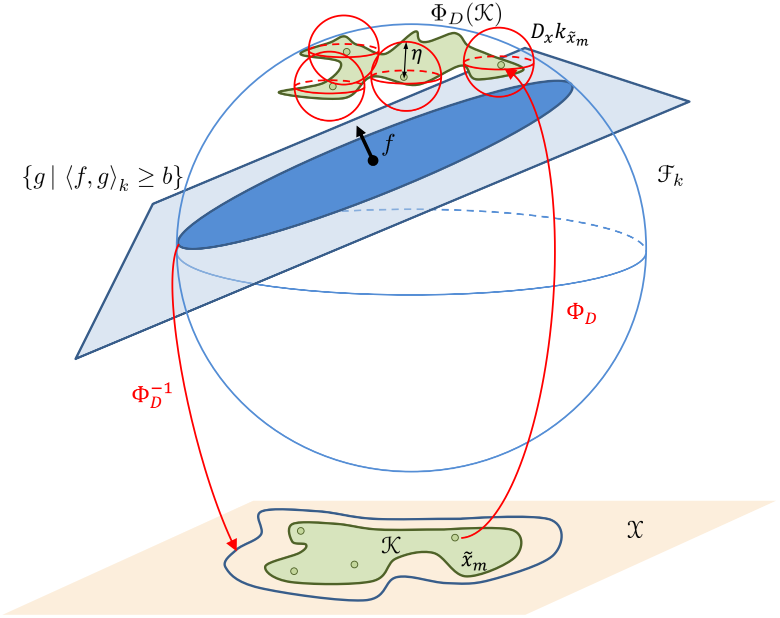
Using the nonlinear embedding $\Phi_D : x \mapsto D_x k(x, \cdot)$, the idea is the same. With only the green points, it is a one-class SVM [Schölkopf et al., 2001b]



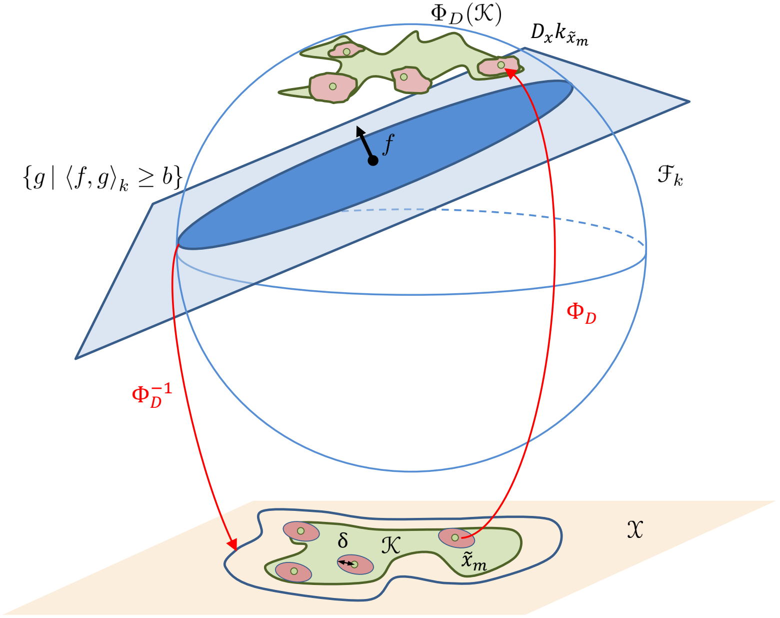
The green points are now samples of a compact set $\mathcal{K}$.



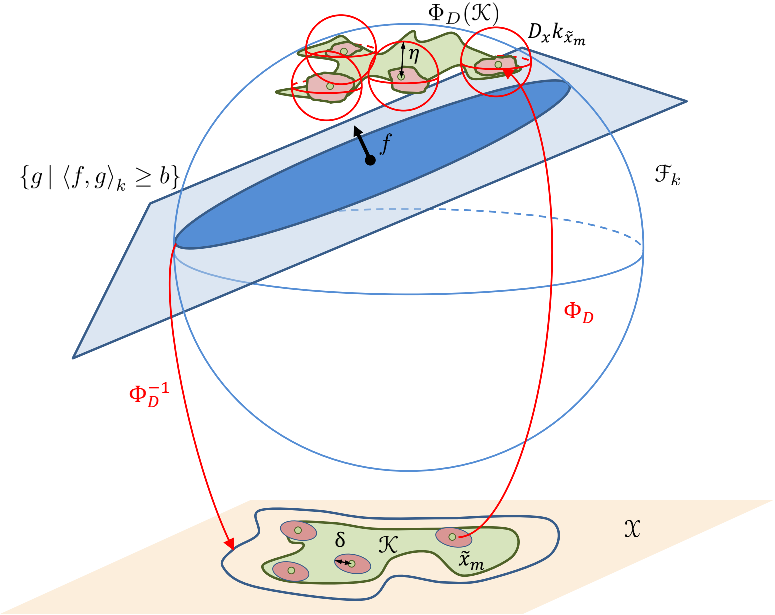
The image $\Phi_D(\mathcal{K})$ is not convex...



The image $\Phi_D(\mathcal{K})$ is not convex, can we cover it by balls of radius η?



First cover $\mathcal{K} \subset \bigcup\{\tilde{x}_m + \delta\mathbb{B}\}$, and then look at the images $\Phi_D(\{\tilde{x}_m + \delta\mathbb{B}\})$



Cover the $\Phi_D(\{\tilde{x}_m + \delta\mathbb{B}\})$ with tiny balls! This is how SOC was defined.

Main contribution in Aubin and Szabó, NeurIPS, 2020

$$(f_\eta, b_\eta) \in \underset{f \in \mathcal{F}_k, \mathbf{b} \in \mathcal{B}}{\arg\min} \quad \mathcal{L}_{\mathbf{b}}(f) = L\left(\left(f(x_n)\right)_{n\in[N]}\right) + R\left(\|f\|_k\right) + \mu\|\mathbf{b}\|^2$$

$$\text{s.t.} \qquad b_i + \eta_{i,m}\|f(\cdot)\|_k \leq D_i f(\tilde{x}_{i,m}), \quad \forall\, m \in [M_i], \forall i \in [\mathcal{I}].$$

where $\mathcal{B}$ is a closed convex constraint set. If $R(\cdot)$ is strictly increasing, then

Theorem (Theoretical guarantees, P.-C. Aubin and Z. Szabó, 2020)

i) *The finite number of SOC constraints is* ***tighter*** *than the infinite number of affine constraints.*

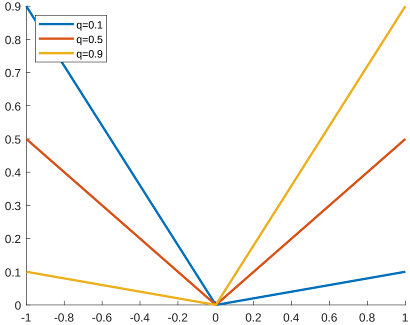
ii) ***Representer theorem*** *(optimal solutions have a finite expression)* $f_\eta = \sum_{i\in[\mathcal{I}], m\in[M_i]} \tilde{a}_{i,m} D_{i,x} k\left(\tilde{x}_{i,m}, \cdot\right) + \sum_{n\in[N]} a_n k(x_n, \cdot)$

iii) *If $\mathcal{L}_{\mathbf{b}}$ is μ-strongly convex, we have* ***bounds****: computable/theoretical*

$$\|f_\eta - \bar{f}\|_k \leq \min\left(\sqrt{\frac{2(\mathcal{L}_{\mathbf{b}_\eta}(f_\eta) - \mathcal{L}_{\mathbf{b}_{\eta=0}}(f_{\eta=0}))}{\mu}}, \sqrt{\frac{L_{\bar{f}}\|\boldsymbol{\eta}\|_\infty}{\mu}}\right)$$

(Assuming $\mathcal{B} = \mathbb{R}^{\mathcal{I}}$ for the a priori bound, $\bar{f}$ the argmin of $\mathcal{L}_{\mathbf{b}}$ with original constraints.)

Joint Quantile Regression (JQR)



$f_\tau(x)$ conditional quantile over (X, Y): $P(Y \leq f_\tau(x) | X = x) = \tau \in]0, 1[$.

Estimation through convex optimization over "pinball loss" $l_\tau(\cdot)$ (i.e. tilted absolute value [Koenker, 2005]).

Known fact: quantile functions can cross when estimated independently.

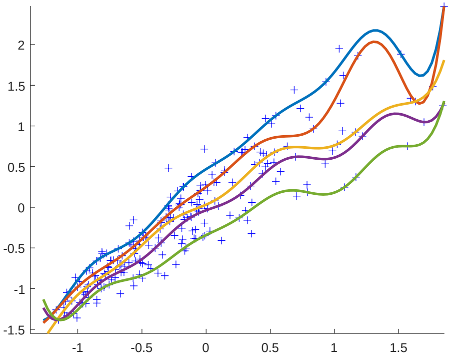
Joint quantile regression with non-crossing constraints

$$\min_{(f_q)_{q\in[Q]}\in\mathcal{F}_k^Q} \mathcal{L}(f_1, \ldots, f_Q) = \frac{1}{N} \sum_{q\in[Q]} \sum_{n\in[N]} l_{\tau_q}(y_n - f_q(\mathbf{x}_n)) + \lambda_f \sum_{q\in[Q]} \|f_q\|_k^2$$

$$\text{s.t. } f_{q+1}(\mathbf{x}) \geq f_q(\mathbf{x}), \forall q \in [Q-1], \forall \mathbf{x} \in [\min_{n\in[N], i\in[d]}\{x_{n,i}\}, \max_{n\in[N], i\in[d]}\{x_{n,i}\}]^d.$$

Pairing non-crossing quantiles with other shape constraints

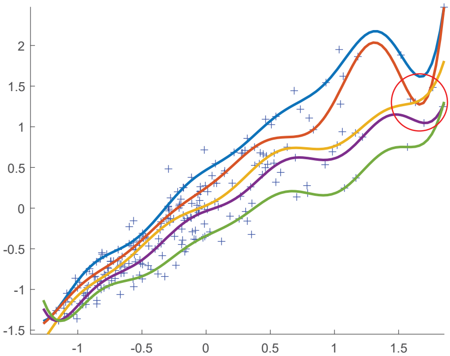
Engel's law (1857): "As income rises, the proportion of income spent on food falls, but absolute expenditure on food rises."



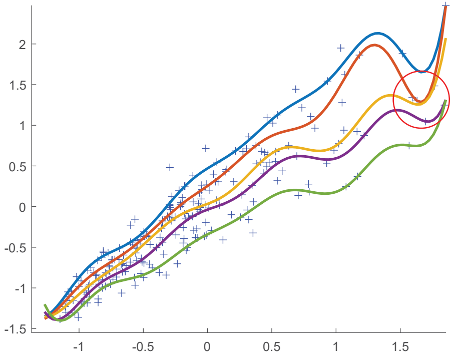
without constraints

Pairing non-crossing quantiles with other shape constraints

Engel's law (1857): "As income rises, the proportion of income spent on food falls, but absolute expenditure on food rises."



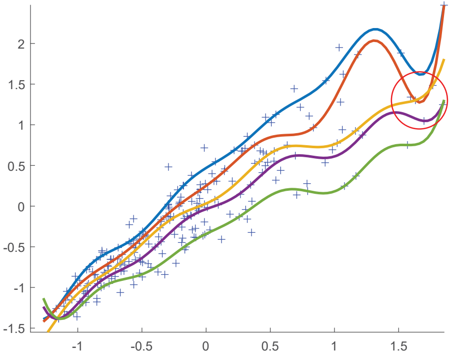
without constraints



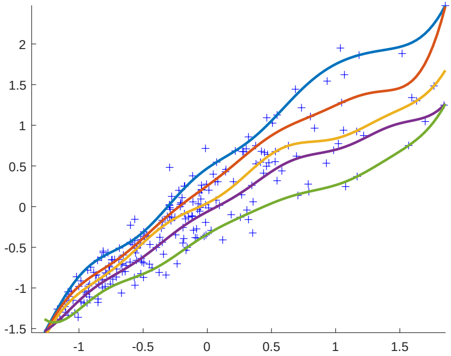
non-crossing constraints

Pairing non-crossing quantiles with other shape constraints

Engel's law (1857): "As income rises, the proportion of income spent on food falls, but absolute expenditure on food rises."



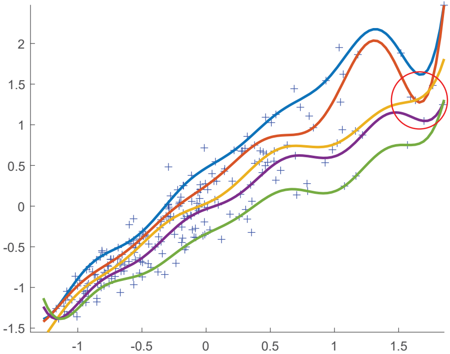
without constraints



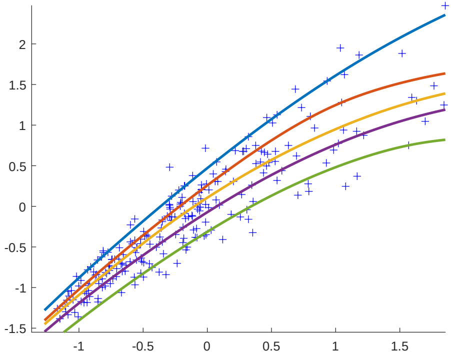
non-crossing+increasing

Pairing non-crossing quantiles with other shape constraints

Engel's law (1857): "As income rises, the proportion of income spent on food falls, but absolute expenditure on food rises."



without constraints

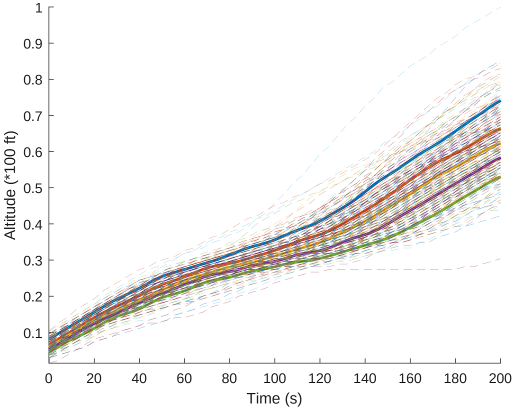


non-crossing+increasing+concave

Qualitative priors have a great effect on the shape of solutions!

Joint quantile regression (JQR): airplane data

Airplane trajectories at takeoff have increasing altitude



JQR with monotonic constraint over $[x_{\min}, x_{\max}]$:

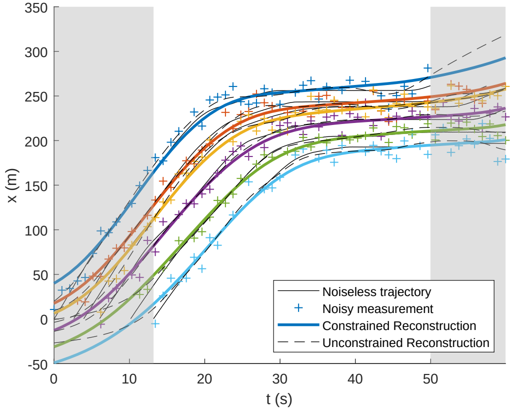
[P.-C. Aubin and Z. Szabó, 2020]

Increasing quantiles should be non-crossing

Data provided by ENAC (flights Paris$\to$Toulouse) [Nicol, 2013]

Kernel ridge regression (KRR): trajectory reconstruction

Very noisy GPS data: six non-overtaking cars in a traffic jam



KRR with monotonic constraint over $[t_{\min}, t_{\max}]$:

[P.-C. Aubin, N. Petit and Z. Szabó, 2020]

Forward trajectories also maintain security distance

Data from IFSTTAR (MOCoPo Project) [Buisson et al., 2016]

Partial conclusion

We have seen how to tighten an **infinite number of affine constraints over a compact set** into **finitely many SOC constraints** in RKHSs
$\hookrightarrow$ we thus have a representer theorem!

- tightening intractable constraints is the only way to have guarantees
- but tightening is “harder“ to perform (here computationally)

Covering schemes suffer from the curse of dimensionality! $\mathcal{X} \subset \mathbb{R}^d$, $d \gg 1$

Partial conclusion

We have seen how to tighten an **infinite number of affine constraints over a compact set** into **finitely many SOC constraints** in RKHSs
$\hookrightarrow$ we thus have a representer theorem!

- tightening intractable constraints is the only way to have guarantees
- but tightening is "harder" to perform (here computationally)

Covering schemes suffer from the curse of dimensionality! $\mathcal{X} \subset \mathbb{R}^d$, $d \gg 1$

However the control problem is only defined over $\mathcal{X} = [t_0, T]$ ($d = 1$)!

Table of Contents

1. Finding the RKHS of LQ optimal control
2. Tightening infinitely many constraints through finite coverings
3. Apply the kernel-based constraint tightening to LQ optimal control

SOC tightening of state-constrained LQ optimal control

Problem of time-varying linear quadratic optimal control with finite horizon and affine inequality state constraints

$$\min_{\mathbf{z}(\cdot)\in\mathcal{S}_{[t_0,T]}} \quad \chi_{\mathbf{z}_0}(\mathbf{z}(t_0)) + g(\mathbf{z}(T)) + \|\mathbf{z}(\cdot)\|_K^2$$

$$\text{s.t.}$$

$$\mathbf{c}_i(t)^\top \mathbf{z}(t) \le d_i(t),\ \forall\, t\in[t_0,T], \forall\, i\in[\mathcal{I}],$$

SOC tightening of state-constrained LQ optimal control

Problem of time-varying linear quadratic optimal control with finite horizon and affine inequality state constraints with SOC tightening

$$\min_{\mathbf{z}(\cdot) \in \mathcal{S}_{[t_0,T]}} \quad \chi_{\mathbf{z}_0}(\mathbf{z}(t_0)) + g(\mathbf{z}(T)) + \|\mathbf{z}(\cdot)\|_K^2$$

$$\text{s.t.}$$

$$\eta_i(\delta_m, t_m)\|\mathbf{z}(\cdot)\|_K + \mathbf{c}_i(t_{i,m})^\top \mathbf{z}(t_{i,m}) \leq d_{i,m},\ \forall\, m \in [M_i], \forall\, i \in [\mathcal{I}],$$

with $[t_0, T] \subset \bigcup_{m\in[M]}[t_m - \delta_m, t_m + \delta_m]$, and two values defined at each t_m

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

$$d_{i,m} := \inf_{t\,\in\,[t_m-\delta_m, t_m+\delta_m]\cap[t_0,T]} d_i(t).$$

Main theoretical result in P.-C. Aubin, SICON, 2021

(H-gen) $\mathbf{A}(\cdot), \mathbf{Q}(\cdot) \in L^1$ and $\mathbf{B}(\cdot), \mathbf{R}(\cdot) \in L^2$, $\mathbf{c}_i(\cdot)$ and $d_i(\cdot) \in \mathcal{C}^0$.

(H-sol) $\mathbf{c}_i(t_0)^\top \mathbf{z}_0 < d_i(t_0)$ and there exists a trajectory $\mathbf{z}^\epsilon(\cdot) \in \mathcal{S}$ satisfying strictly the affine constraints, as well as the initial condition.[1]

(H-obj) $g(\cdot)$ is convex and continuous.

Theorem (∃/Approximation by SOC constraints, P.-C. Aubin, 2021)

Both the original problem and its strengthening have unique optimal solutions. For any $\rho > 0$, there exists $\overline{\delta} > 0$ such that for all $(\delta_m)_{m\in[\![1,N_0]\!]}$, with $[t_0, T] \subset \cup_{m\in[\![1,N_0]\!]}[t_m - \delta_m, t_m + \delta_m]$ satisfying $\overline{\delta} \geq \max_{m\in[\![1,N_0]\!]} \delta_m$,

$$\frac{1}{\gamma_K} \sup_{t\in[t_0,T]} \|\bar{\mathbf{z}}_\eta(t) - \bar{\mathbf{z}}(t)\| \leq \|\bar{\mathbf{z}}_\eta(\cdot) - \bar{\mathbf{z}}(\cdot)\|_K \leq \rho$$

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

[1](H-sol) is implied for instance by an inward-pointing condition at the boundary.

Main practical result in P.-C. Aubin, SICON, 2021

Problem of time-varying linear quadratic optimal control with finite horizon and affine inequality state constraints with SOC tightening

$$\min_{\mathbf{z}(\cdot) \in \mathcal{S}_{[t_0,T]}} \quad \chi_{\mathbf{z}_0}(\mathbf{z}(t_0)) + g(\mathbf{z}(T)) + \|\mathbf{z}(\cdot)\|_K^2$$

$$\text{s.t.}$$

$$\eta_i(\delta_m, t_m)\|\mathbf{z}(\cdot)\|_K + \mathbf{c}_i(t_{i,m})^\top \mathbf{z}(t_{i,m}) \le d_{i,m},\ \forall\, m \in [M_i], \forall\, i \in [\mathcal{I}].$$

By the representer theorem, the optimal solution has the form

$$\bar{\mathbf{z}}(\cdot) = \sum_{j=0}^{P} \sum_{m=1}^{N_j} K(\cdot, t_{j,m})\bar{\mathbf{p}}_{j,m},$$

where $t_{0,1} = t_0$ and $t_{0,2} = T$, and the coefficients $(\bar{\mathbf{p}}_{j,m})_{j,m}$ solve a finite dimensional second-order cone problem.

Main practical result in P.-C. Aubin, SICON, 2021

More precisely, setting $t_{0,1} = t_0$ and $t_{0,2} = T$, the coefficients of the optimal solution $\bar{\mathbf{z}}(\cdot) = \sum_{j=0}^{P} \sum_{m=1}^{N_j} K(\cdot, t_{j,m})\bar{\mathbf{p}}_{j,m}$ solve

$$
\begin{aligned}
\min_{\substack{z\in\mathbb{R}_+,\\ \mathbf{p}_{j,m}\in\mathbb{R}^N,\\ \alpha_{j,m}\in\mathbb{R}}} \quad & \chi_{\mathbf{z}_0}\left(\sum_{j=0}^{P}\sum_{m=1}^{N_j} K(t_0, t_{j,m})\bar{\mathbf{p}}_{j,m}\right) + g\left(\sum_{j=0}^{P}\sum_{m=1}^{N_j} K(T, t_{j,m})\bar{\mathbf{p}}_{j,m}\right) + \gamma^2 \\
\text{s.t.} \quad & \gamma^2 = \sum_{i=0}^{P}\sum_{n=1}^{N_i}\sum_{j=0}^{P}\sum_{m=1}^{N_j} \mathbf{p}_{i,n}^\top K(t_{i,n}, t_{j,m})\mathbf{p}_{j,m}, \\
& \mathbf{p}_{j,m} = \alpha_{j,m}\mathbf{c}_j(t_m), \quad \forall\, m \in [\![1, N_j]\!], \forall\, j \in [\![1, P]\!], \\
& \begin{matrix} \eta_i(\delta_{i,m}, t_{i,m})\gamma + \sum_{j=0}^{P}\sum_{m=1}^{N_j} \mathbf{c}_i(t_{i,m})^\top K(t_{i,m}, t_{j,m})\mathbf{p}_{j,m} \\ \le d_i(\delta_{i,m}, t_{i,m}), \end{matrix} \quad \begin{matrix} \forall\, m \in [\![1, N_i]\!], \\ \forall\, i \in [\![1, P]\!], \end{matrix} .
\end{aligned}
$$

which can be written equivalently as a finite dimensional second-order cone problem (SOCP).

Future work: Pushing RKHSs beyond/Revisiting LQR

For RKHSs

- Control constraints do not correspond to continuous evaluations
  $\hookrightarrow$ limits of RKHS pointwise theory (e.g. $x' = u \in L^2([0, T], [-1, 1])$ a.e.)
- Successive linearizations of nonlinear system lead to changing kernels
  $\hookrightarrow$ a single kernel may not be sufficient (e.g. $x' = f_{[x_n(\cdot)]}x + f_{[u_n(\cdot)]}u$ a.e.)
- Non-quadratic costs for linear systems do not lead to Hilbert spaces
  $\hookrightarrow$ one may need Banach kernels (e.g. $\|\mathbf{u}(\cdot)\|^2_{L^2(0,T)} \to \|\mathbf{u}(\cdot)\|_{L^1(0,T)}$)

Future work: Pushing RKHSs beyond/Revisiting LQR

For RKHSs

- Control constraints do not correspond to continuous evaluations
  $\hookrightarrow$ limits of RKHS pointwise theory (e.g. $x' = u \in L^2([0,T],[-1,1])$ a.e.)

- Successive linearizations of nonlinear system lead to changing kernels
  $\hookrightarrow$ a single kernel may not be sufficient (e.g. $x' = f_{[x_n(\cdot)]}x + f_{[u_n(\cdot)]}u$ a.e.)

- Non-quadratic costs for linear systems do not lead to Hilbert spaces
  $\hookrightarrow$ one may need Banach kernels (e.g. $\|\mathbf{u}(\cdot)\|^2_{L^2(0,T)} \to \|\mathbf{u}(\cdot)\|_{L^1(0,T)}$)

For control theory

- To each evaluation at time t corresponds a covector $p_t \in \mathbb{R}^Q$
  $\hookrightarrow$ Representer theorem well adapted for state constraints, but unsuitable for control constraints. Reverts the difficulty w.r.t. PMP approach.

- The Gramian of controllability generates trajectories
  $\hookrightarrow$ This allows for close-form solutions in continuous-time for state constraints.

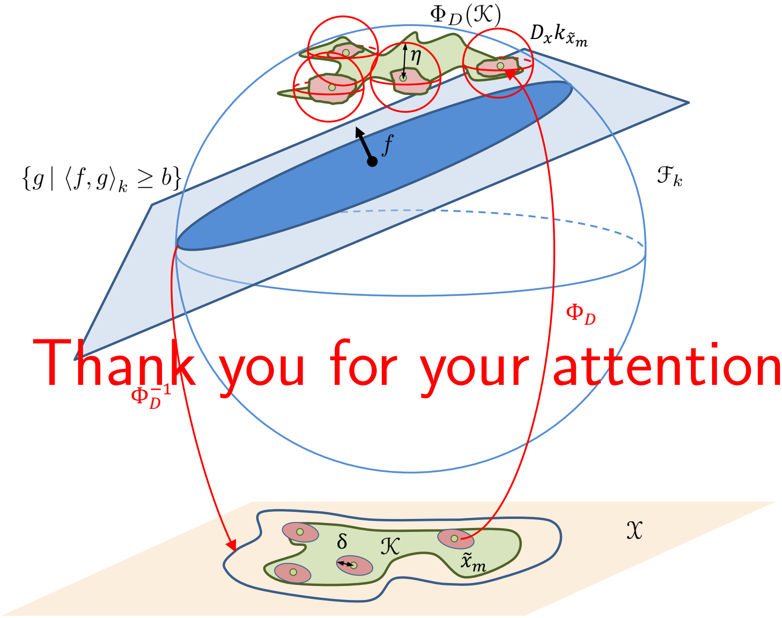
Final remarks

"Finite coverings in RKHSs can be used to turn an **infinite number of pointwise affine constraints** over a compact set into **finitely many SOC constraints**."

"State-constrained LQ Optimal Control is a shape-constrained kernel regression."

"In general, positive definite kernels are much too linear to tackle nonlinear control problems $\rightarrow$ **Linearize**! "

Not covered in this talk:
Chapter 6 (regularity of minimal time) and Chapter 7 (set approximation).



Thank you for your attention!

References I

Agrell, C. (2019).
Gaussian processes with linear operator inequality constraints.
Journal of Machine Learning Research, 20:1–36.

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

Aubin-Frankowski, P.-C. (2020).
Lipschitz regularity of the minimum time function of differential inclusions with state constraints.
Systems & Control Letters, 139:104677.

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

References II

Aubin-Frankowski, P.-C. (2021b).
Linearly-constrained Linear Quadratic Regulator from the viewpoint of kernel methods.
SIAM Journal on Control and Optimization.
(to appear, https://arxiv.org/abs/2011.02196).

Aubin-Frankowski, P.-C., Petit, N., and Szabó, Z. (2020).
Kernel Regression for Vehicle Trajectory Reconstruction under Speed and Inter-vehicular Distance Constraints.
In *IFAC World Congress 2020*.
(https://hal-mines-paristech.archives-ouvertes.fr/hal-03021643).

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

References III

Aubin-Frankowski, P.-C. and Szabó, Z. (2020b).
Hard Shape-Constrained Kernel Machines.
In *Advances in Neural Information Processing Systems (NeurIPS)*.
(http://arxiv.org/abs/2005.12636).

Aubin-Frankowski, P.-C. and Vert, J.-P. (2020).
Gene regulation inference from single-cell RNA-seq data with linear differential equations and velocity inference.
Bioinformatics.

Berlinet, A. and Thomas-Agnan, C. (2004).
Reproducing Kernel Hilbert Spaces in Probability and Statistics.
Kluwer.

Brault, R., Lambert, A., Szabó, Z., Sangnier, M., and d'Alché Buc, F. (2019).
Infinite-task learning with RKHSs.
In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1294–1302.

References IV

Buisson, C., Villegas, D., and Rivoirard, L. (2016).
Using polar coordinates to filter trajectories data without adding extra physical constraints.
In *Transportation Research Board 95th Annual Meeting*.

Heckman, N. (2012).
The theory and application of penalized methods or reproducing kernel Hilbert spaces made easy.
Statistics Surveys, 6(0):113–141.

Koenker, R. (2005).
Quantile Regression.
Cambridge University Press.

Korba, A., Aubin-Frankowski, P.-C., Majewski, S., and Ablin, P. (2021).
Kernel Stein discrepancy descent.
International Conference on Machine Learning (ICML).
(to appear).

References V

Marteau-Ferey, U., Bach, F., and Rudi, A. (2020).
Non-parametric models for non-negative functions.
In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 12816–12826.

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

Nicol, F. (2013).
Functional principal component analysis of aircraft trajectories.
In *International Conference on Interdisciplinary Science for Innovative Air Traffic Management (ISIATM)*.

Rothenstein, D., Baier, J., Schreiber, T. D., Barucha, V., and Bill, J. (2012).
Influence of zinc on the calcium carbonate biomineralization of halomonas halophila.
Aquatic Biosystems, 8(1):31.

References VI

Saitoh, S. and Sawano, Y. (2016).
Theory of Reproducing Kernels and Applications.
Springer Singapore.

Sangnier, M., Fercoq, O., and d'Alche Buc, F. (2016).
Joint quantile regression in vector-valued RKHSs.
pages 3693–3701.

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

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. (2001b).
Estimating the support of a high-dimensional distribution.
Neural Computation, 13(7):1443–1471.

References VII

Takeuchi, I., Le, Q., Sears, T., and Smola, A. (2006).
Nonparametric quantile estimation.
Journal of Machine Learning Research, 7:1231–1264.

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

Annex: Extra list of shape constraints

- **Monotonicity w.r.t. partial ordering**: $\mathbf{u} \preccurlyeq \mathbf{v} \Rightarrow f(\mathbf{u}) \leq f(\mathbf{v})$ for $\mathbf{u} \preccurlyeq \mathbf{v}$ iff $\sum_{j\in[i]} u_j \leq \sum_{j\in[i]} v_j$ for all $i \in [d]$ (unordered weak majorization)

$$\partial^{\mathbf{e}_1} f(\mathbf{x}) \geq \ldots \geq \partial^{\mathbf{e}_d} f(\mathbf{x}) \geq 0 \quad (\forall \mathbf{x});$$

  $\mathbf{u} \preccurlyeq \mathbf{v} \Rightarrow f(\mathbf{u}) \leq f(\mathbf{v})$ for $\mathbf{u} \preccurlyeq \mathbf{v}$ iff $u_i \leq v_i$ $(\forall i \in [d])$ (product ordering),

$$\partial^{\mathbf{e}_j} f(\mathbf{x}) \geq 0, \quad (\forall j \in [d], \forall \mathbf{x}).$$

- **Supermodularity**: $f(\mathbf{u} \vee \mathbf{v}) + f(\mathbf{u} \wedge \mathbf{v}) \geq f(\mathbf{u}) + f(\mathbf{v})$, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, where $\mathbf{u} \vee \mathbf{v} := (\max(u_j, v_j))_{j\in[d]}$ and $\mathbf{u} \wedge \mathbf{v} := (\min(u_j, v_j))_{j\in[d]}$. For $f \in C^2$

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \geq 0 \quad (\forall i \neq j \in [d], \forall \mathbf{x}).$$

Annex: JQR performance over UCI datasets

- PDCD = Primal-Dual Coordinate Descent [Sangnier et al., 2016], JQR with parallel/heteroscedatic quantile penalization (see also ITL [Brault et al., 2019] for noncrossing inducer)
- mean $\pm$ std of 100$\times$value of the pinball loss (smaller is better)

Dataset	d	N	PDCD	SOC
engel	1	235	48 $\pm$ 8	53 $\pm$ 9
GAGurine	1	314	61 $\pm$ 7	65 $\pm$ 6
geyser	1	299	105 $\pm$ 7	108 $\pm$ 3
mcycle	1	133	66 $\pm$ 9	62 $\pm$ 5
ftcollinssnow	1	93	154 $\pm$ 16	148 $\pm$ 13
CobarOre	2	38	159 $\pm$ 24	151 $\pm$ 17
topo	2	52	69 $\pm$ 18	62 $\pm$ 14
caution	2	100	88 $\pm$ 17	98 $\pm$ 22
ufc	3	372	81 $\pm$ 4	87 $\pm$ 6

Annex: Estimation of production functions

Classical assumptions

$$\min_{f \in \mathcal{F}_K} \frac{1}{N} \sum_{n \in [N]} [y_n - f(\mathbf{x}_n)]^2 + \lambda \|f\|_K^2$$

s.t.

$$\begin{aligned}
0 &\leq f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{K}, \\
0 &\leq \partial^{\mathbf{e}_i} f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{K}, \forall i \in [d], \\
0 &= f(\mathbf{0}), \\
\mathbf{0}_{d \times d} &\preccurlyeq - \left[\left(\partial^{\mathbf{e}_i + \mathbf{e}_j} f \right) (\mathbf{x}) \right]_{i,j}, \forall \mathbf{x} \in \mathcal{K},
\end{aligned}$$

$y \to -\log(y)$

$$\min_{g \in \mathcal{F}_K} \frac{1}{N} \sum_{n \in [N]} [y_n - g(\mathbf{x}_n)]^2$$

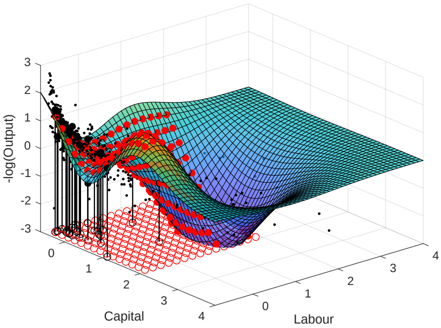
s.t.

$$\|g\|_K \leq \tilde{\lambda},$$

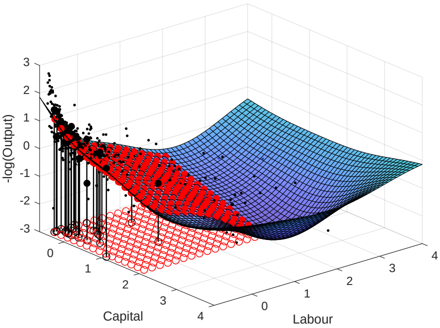
$$\begin{aligned}
0 &\leq -\partial^{\mathbf{e}_1} g(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{K}, \\
0 &\leq -\partial^{\mathbf{e}_2} g(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{K}, \\
\mathbf{0}_{2 \times 2} &\preccurlyeq \left[\left(\partial^{\mathbf{e}_i + \mathbf{e}_j} g \right) (\mathbf{x}) \right]_{i,j \in [2]}, \forall \mathbf{x} \in \mathcal{K}.
\end{aligned}$$

Annex: Estimation of production functions

Only 25 points selected out of 543, checking generalization properties for various constraints (used as side information)



(a) NoCons

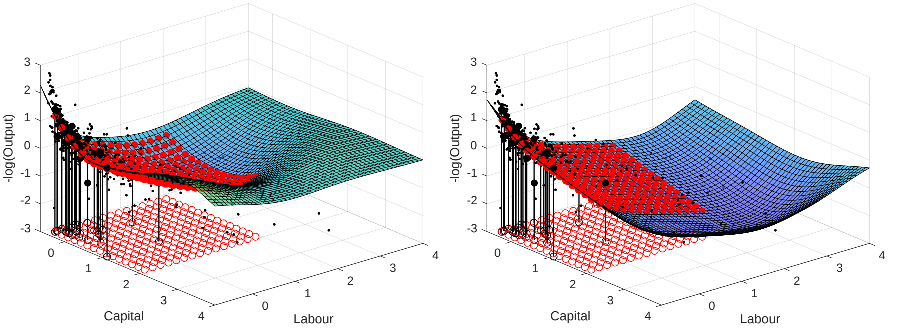


(b) SOC Monot.

Belgian labour data, 1996, `https://vincentarelbundock.github.io/Rdatasets/doc/Ecdat/Labour.html`

Annex: Estimation of production functions

Only 25 points selected out of 543, checking generalization properties for various constraints (used as side information)



(c) SOC Conv.

(d) SOC Conv.+Monot.

Belgian labour data, 1996, `https://vincentarelbundock.github.io/Rdatasets/doc/Ecdat/Labour.html`

Annex: Estimation of production functions

Only 25 points selected out of 543, checking generalization properties for various constraints (used as side information)

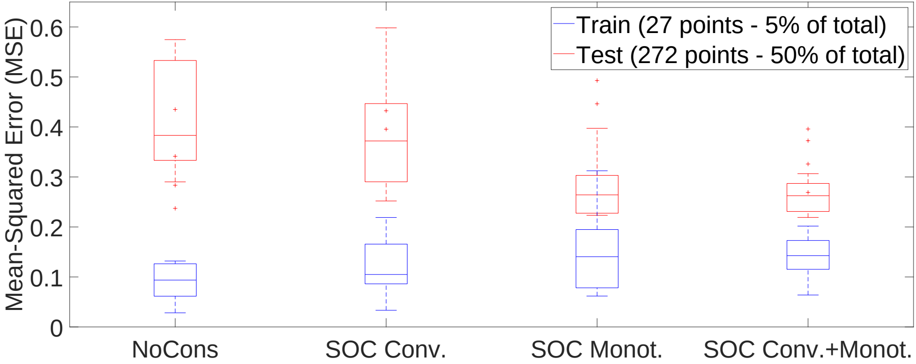


Figure: MSE as a function of incorporating shape constraints with the proposed SOC technique. NoCons: no constraint. SOC Monot.: two monotonicity constraints. SOC Conv.: one convexity constraint. SOC Conv.+Monot.: one convexity and two monotonicity constraints.

Annex: Green kernels and RKHSs

Let D be a differential operator, D^* its formal adjoint. Define the Green function $G_{D^*D,x}(y) : \Omega \to \mathbb{R}$ s.t. $D^*D\, G_{D^*D,x}(y) = \delta_z(y)$ then, if the integrals over the boundaries in Green's formula are null, for any $f \in \mathcal{F}_k$

$$f(x) = \int_\Omega f(y) D^* D G_{D^*D,x}(y) dy = \int_\Omega Df(y) D G_{D^*D,x}(y) =: \langle f, G_{D^*D,x} \rangle_{\mathcal{F}_k},$$

so $k(x, y) = G_{D^*D,x}(y)$ [Saitoh and Sawano, 2016, p61]. For vector-valued contexts, e.g. $\mathcal{F}_K = W^{s,2}(\mathbb{R}^d, \mathbb{R}^d)$ and $D^*D = (1 - \sigma^2 \Delta)^s$ component-wise, see [Micheli and Glaunès, 2014, p9].

Alternatively, in 1D, $D\, G_{D,x}(y) = \delta_z(y)$, the kernel associated to the inner product $\int_\Omega Df(y) Dg(y) dy$ for the space of f "null at the border" writes as

$$k(x, y) = \int_\Omega G_{D,x}(z) G_{D,y}(z) dz$$

see [Berlinet and Thomas-Agnan, 2004, p286] and [Heckman, 2012].

Annex: Alternative finite coverings

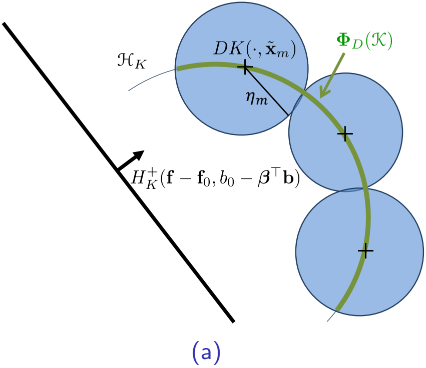
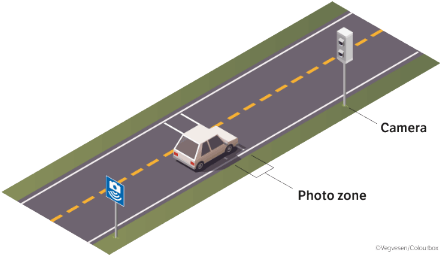


Figure: Two examples of coverings in $\mathcal{F}_K$ of $\mathbf{\Phi}_D(\mathcal{K})$ by a set $\bar{\Omega} = \cup_{m \in [M]} \bar{\Omega}_m$ contained in the halfspace $H_K^+(\mathbf{f} - \mathbf{f}_0, b_0 - \boldsymbol{\beta}^\top \mathbf{b})$. (a): covering through balls $\Omega_m = \mathring{\mathbb{B}}_K\left(DK(\cdot, \tilde{x}_m), \eta_m\right)$. (b): covering through a ball intersected with halfspaces.

Annex: Why are state constraints difficult to study?

- **Theoretical obstacle**: Pontryagine's maximum principle involves not only an adjoint vector $\mathbf{p}(t)$ but also measures/BV functions $\psi(t)$ supported at times where the constraints are saturated. You cannot just backpropagate the Hamiltonian system from the transversality condition.

- **Numerical obstacle**: Time discretization of constraints may fail e.g.



Speed cameras in traffic control

In between two cameras, drivers always break the speed limit.

Annex: IPC gives strictly feasible trajectories

(H-sol) $\mathbf{C}(0)\mathbf{z}_0 < \mathbf{d}(0)$ and there exists a trajectory $\mathbf{z}^\epsilon(\cdot) \in \mathcal{S}$ satisfying strictly the affine constraints, as well as the initial condition.

(H1) $\mathbf{A}(\cdot), \mathbf{B}(\cdot) \in \mathcal{C}^0$, $\mathbf{c}_i(\cdot), d_i(\cdot) \in \mathcal{C}^1$ and $\mathbf{C}(0)\mathbf{z}_0 < \mathbf{d}(0)$.

(H2) There exists $M_u > 0$ s.t. , for all $t \in [t_0, T]$ and $\mathbf{z} \in \mathbb{R}^Q$ satisfying $\mathbf{C}(t)\mathbf{z} \leq \mathbf{d}(t)$, and $\|\mathbf{z}\| \leq (1 + \|\mathbf{z}_0\|)e^{T\|\mathbf{A}(\cdot)\|_{L^\infty(t_0,T)} + TM_u\|\mathbf{B}(\cdot)\|_{L^\infty(t_0,T)}}$, there exists $\mathbf{u}_{t,x} \in M_u\mathbb{B}_M$ such that

$$\forall\, i \in \{j \,|\, \mathbf{c}_j(t)^\top \mathbf{z} = d_j(t)\},\ \mathbf{c}_i'(t)^\top \mathbf{z} - d_i'(t) + \mathbf{c}_i(t)^\top (\mathbf{A}(t)\mathbf{z} + \mathbf{B}(t)\mathbf{u}_{t,x}) < 0.$$

This is an inward-pointing condition (IPC) at the boundary.

Lemma (Existence of interior trajectories)

If (H1) and (H2) hold, then (H-sol) holds.

Annex: control proof main idea, nested property

$$\eta_i(\delta, t) := \sup \|K(\cdot, t)\mathbf{c}_i(t) - K(\cdot, s)\mathbf{c}_i(s)\|_K, \quad \omega_i(\delta, t) := \sup |d_i(t) - d_i(s)|,$$
$$d_i(\delta_m, t_m) := \inf d_i(s), \qquad \text{over } s \in [t_m - \delta_m, t_m + \delta_m] \cap [t_0, T]$$

For $\overrightarrow{\boldsymbol{\epsilon}} \in \mathbb{R}_+^P$, the constraints we shall consider are defined as follows

$$\mathcal{V}_0 := \{\mathbf{z}(\cdot) \in \mathcal{S} \,|\, \mathbf{C}(t)\mathbf{z}(t) \leq \mathbf{d}(t), \,\forall\, t \in [t_0, T]\},$$
$$\mathcal{V}_{\delta,\text{fin}} := \{\mathbf{z}(\cdot) \in \mathcal{S} \,|\, \overrightarrow{\boldsymbol{\eta}}(\delta_m, t_m)\|\mathbf{z}(\cdot)\|_K + \mathbf{C}(t_m)\mathbf{z}(t_m) \leq \mathbf{d}(\delta_m, t_m), \,\forall\, m \in [\![1, M_0]\!]\},$$
$$\mathcal{V}_{\delta,\text{inf}} := \{\mathbf{z}(\cdot) \in \mathcal{S} \,|\, \overrightarrow{\boldsymbol{\eta}}(\delta, t)\|\mathbf{z}(\cdot)\|_K + \overrightarrow{\boldsymbol{\omega}}(\delta, t) + \mathbf{C}(t)\mathbf{z}(t) \leq \mathbf{d}(t), \,\forall\, t \in [t_0, T]\},$$
$$\mathcal{V}_{\overrightarrow{\boldsymbol{\epsilon}}} := \{\mathbf{z}(\cdot) \in \mathcal{S} \,|\, \overrightarrow{\boldsymbol{\epsilon}} + \mathbf{C}(t)\mathbf{z}(t) \leq \mathbf{d}(t), \,\forall\, t \in [t_0, T]\}.$$

Proposition (Nested sequence)

Let $\delta_{max} := \max_{m \in [\![1, M_0]\!]} \delta_m$. For any $\delta \geq \delta_{max}$, if, for a given $y_0 \geq 0$, $\epsilon_i \geq \sup_{t \in [t_0, T]}[\eta_i(\delta, t)y_0 + \omega_i(\delta, t)]$, then we have a nested sequence

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

Only the simpler $\mathcal{V}_{\overrightarrow{\boldsymbol{\epsilon}}}$ constraints matter!

Annex: Van Loan's trick for time-invariant Gramians

Use matrix exponentials as in [Van Loan, 1978]

$$\exp\left(\begin{pmatrix} \mathbf{A} & \mathbf{Q}_c \\ 0 & -\mathbf{A}^\top \end{pmatrix} \Delta\right) = \begin{pmatrix} \mathbf{F}_2(\Delta) & \mathbf{G}_2(\Delta) \\ 0 & \mathbf{F}_3(\Delta) \end{pmatrix}$$

$$\hat{\mathbf{F}}_2(t) = e^{\mathbf{A}t}$$

$$\hat{\mathbf{F}}_3(t) = e^{-\mathbf{A}^\top t}$$

$$\hat{\mathbf{G}}_2(t) = \int_0^t e^{(t-\tau)\mathbf{A}} \mathbf{Q}_c e^{-\tau \mathbf{A}^\top} \mathrm{d}\tau$$

$$K_1(s,t) = \int_0^{\min(s,t)} e^{(s-\tau)\mathbf{A}} \mathbf{B}\mathbf{B}^\top e^{(t-\tau)\mathbf{A}^\top} \mathrm{d}\tau$$

Set $\mathbf{Q}_C = \mathbf{B}\mathbf{R}^{-1}\mathbf{B}^\top$.

For $s \leq t$, $K_1(s,t) = \hat{\mathbf{G}}_2(s)\hat{\mathbf{F}}_2(t)^\top$

For $t \leq s$, $K_1(s,t) = \hat{\mathbf{F}}_2(s)\hat{\mathbf{G}}_2(t)^\top$

Annex: Example of constrained pendulum - definition

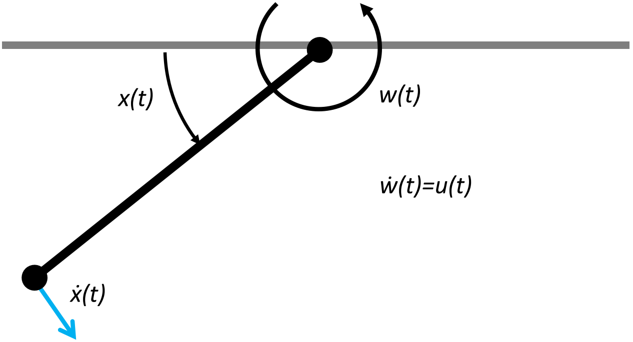
Constrained pendulum when controlling the third derivative of the angle

$$\min_{x(\cdot),w(\cdot),u(\cdot)} -\dot{x}(T) + \lambda \|u(\cdot)\|^2_{L^2(0,T)} \qquad \lambda \ll 1$$

$$x(0) = 0.5, \quad \dot{x}(0) = 0, \quad w(0) = 0, \quad x(T/3) = 0.5, \quad x(T) = 0$$

$$\ddot{x}(t) = -10\,x(t) + w(t), \quad \dot{w}(t) = u(t), \text{ a.e. in } [0, T]$$

$$\dot{x}(t) \in [-3, +\infty[, \quad w(t) \in [-10, 10], \forall\, t \in [0, T]$$



Annex: Example of constrained pendulum - definition

Constrained pendulum when controlling the third derivative of the angle

$$\boxed{\min_{x(\cdot),w(\cdot),u(\cdot)} -\dot{x}(T) + \lambda\|u(\cdot)\|^2_{L^2(0,T)}} \qquad \lambda \ll 1$$

$$\boxed{x(0) = 0.5, \quad \dot{x}(0) = 0, \quad w(0) = 0, \quad x(T/3) = 0.5, \quad x(T) = 0}$$

$$\boxed{\ddot{x}(t) = -10\,x(t) + w(t), \quad \dot{w}(t) = u(t), \text{ a.e. in } [0, T]}$$

$$\boxed{\dot{x}(t) \in [-3, +\infty[, \quad w(t) \in [-10, 10], \forall\, t \in [0, T]}$$

Converting affine state constraints to SOC constraints, applying rep. thm

$$\begin{aligned} \eta_{\dot{x}}\|\mathbf{z}(\cdot)\|_K - \dot{x}(t_m) &\leq 3, \\ \eta_w\|\mathbf{z}(\cdot)\|_K + w(t_m) &\leq 10, \\ \eta_w\|\mathbf{z}(\cdot)\|_K - w(t_m) &\leq 10 \end{aligned}$$

$$\begin{aligned} \bar{\mathbf{z}}(\cdot) = {} & K(\cdot, 0)\mathbf{p}_0 + K(\cdot, T/3)\mathbf{p}_{T/3} \\ & + K(\cdot, T)\mathbf{p}_T + \sum_{m=1}^{M} K(\cdot, t_m)\mathbf{p}_m \end{aligned}$$

Most of computational cost is related to the “controllability Gramians“ $K_1(s,t) = \int_0^{\min(s,t)} \mathbf{e}^{(s-\tau)\mathbf{A}}\mathbf{B}\mathbf{B}^\top \mathbf{e}^{(t-\tau)\mathbf{A}^\top} \mathrm{d}\tau$ which we have to approximate.

Annex: Example of constrained pendulum - illustration

Optimal solutions of the constrained pendulum “path-planning“ problem.
Red circles: equality constraints. Grayed areas: constraints over $[0, T]$.

Angle $x(\cdot)$ Velocity $\dot{x}(\cdot)$ Couple $w(\cdot)$

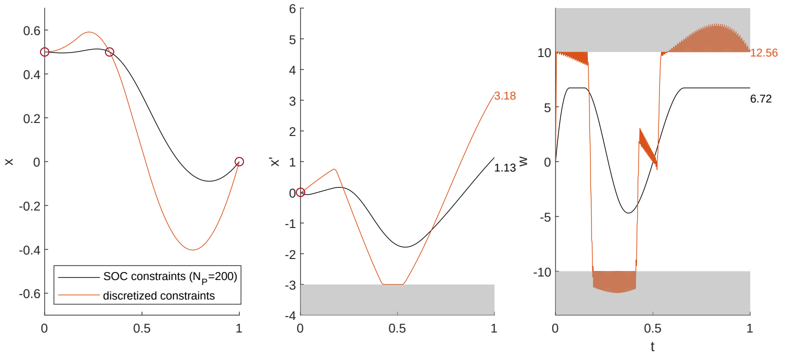


Figure: Comparison of SOC constraints (guaranteed η_w) vs discretized constraints ($\eta_w = 0$) for $N_P = 200$.

Annex: Example of constrained pendulum - illustration

Optimal solutions of the constrained pendulum “path-planning“ problem. Red circles: equality constraints. Grayed areas: constraints over $[0, T]$.

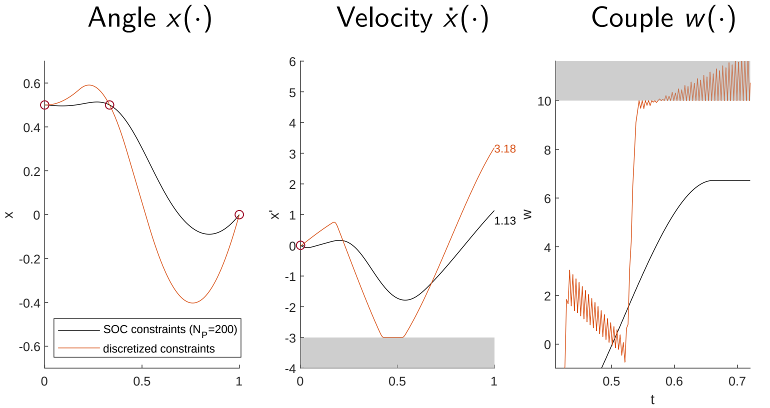


Figure: Comparison of SOC constraints (guaranteed η_w) vs discretized constraints ($\eta_w = 0$) for $N_P = 200$ - Chattering phenomenon like for traffic cameras!.

Annex: Example of constrained pendulum - illustration

Optimal solutions of the constrained pendulum “path-planning“ problem.
Red circles: equality constraints. Grayed areas: constraints over $[0, T]$.

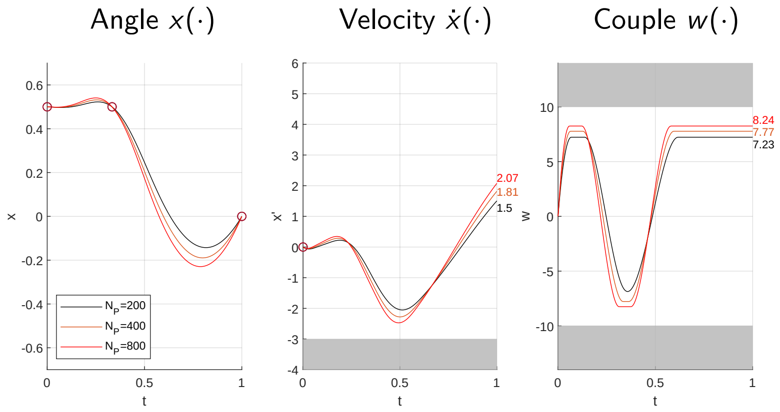


Figure: Comparison of SOC constraints for varying N_P and guaranteed η_w.

Annex: Example of constrained pendulum - illustration

Optimal solutions of the constrained pendulum “path-planning“ problem.
Red circles: equality constraints. Grayed areas: constraints over $[0, T]$.

Angle $x(\cdot)$ Velocity $\dot{x}(\cdot)$ Couple $w(\cdot)$

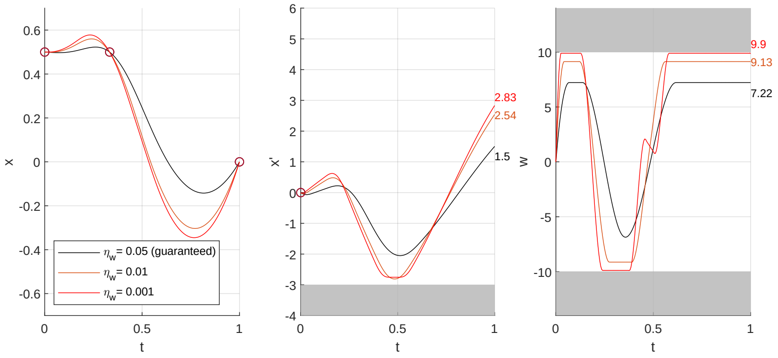


Figure: Comparison of SOC constraints for varying η_w and $N_P = 200$.