State-constrained Linear-Quadratic Optimal Control as a Kernel Regression with Hard Shape Constraints

Pierre-Cyril Aubin-Frankowski

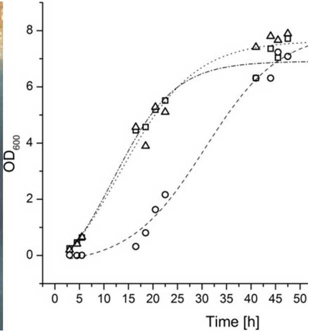
PhD student (École des Ponts ParisTech),
CAS - Centre Automatique et Systèmes, MINES ParisTech

https://pcaubin.github.io/

October 2020

What are shape/state constraints?

Estimation



Side information

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

Control

Physical constraints

$\hookrightarrow$ provides feasible trajectories in path-planning

Ubiquitous and both handled as a constrained optimization problem

Table of Contents

Based on

- *Kernel Regression for Vehicle Trajectory Reconstruction under Speed and Inter-vehicular Distance Constraints*, PCAF, Nicolas Petit and Zoltán Szabó, *IFAC World Congress 2020*
- *Hard Shape-Constrained Kernel Machines*, PCAF and Zoltán Szabó, NeurIPS 2020, `https://arxiv.org/abs/2005.12636`
- *Linearly-constrained Linear Quadratic Regulator from the viewpoint of kernel methods*, PCAF, June 2020, Accepted with minor revision in SIAM Journal on Control and Optimization

Problem statement

Given samples $(x_n, y_n)_{n\in[N]} \in (\mathcal{X}\times\mathbb{R})^N$, a *loss* $L : (\mathcal{X}\times\mathbb{R}\times\mathbb{R})^N \to \mathbb{R}\cup\{\infty\}$, a *regularizer* $\Omega : \mathbb{R}_+ \to \mathbb{R}$. For $x \in \mathcal{X} \subset \mathbb{R}^d$, $f \in \mathcal{C}^s(\mathcal{X},\mathbb{R})$, consider

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

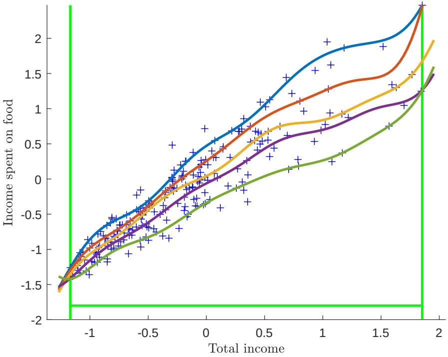
where $\mathcal{F}$ is a Hilbert space of smooth functions from $\mathcal{X}$ to $\mathbb{R}$, D_i is a differential operator ($D_i = \sum_j \gamma_j \partial^{r_j}$), $b_i \in \mathbb{R}$ is a lower bound, $\mathcal{K}_i$ is compact.

For non-finite $\mathcal{K}_i$, we have an infinite number of constraints!

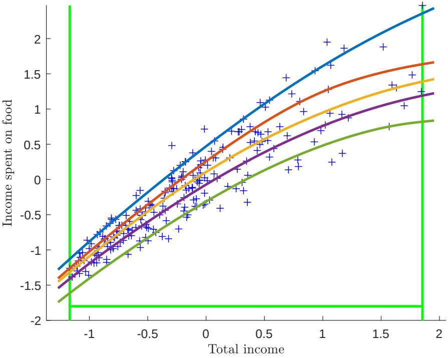
How can we make this optimization problem computationally tractable?

In practice: nonparametric estimation under constraints

In statistics: nonnegative densities, non-crossing quantiles



non-crossing+increasing



non-crossing+increasing+concave

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

Glimpse of content of the talk

From dealing with a real-valued problem $f : x \in \mathcal{X} \subset \mathbb{R}^d \to y \in \mathbb{R}$

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

ex: least-squares with monotonicity constraint

to a path-planning vector-valued problem $f : t \in [0, T] \to y \in \mathbb{R}^Q$

Take $\mathcal{F}$ to be a Hilbert space of trajectories (e.g. Sobolev space)

$$\begin{aligned}\min_{f(\cdot) \in \mathcal{F}} \quad & g(f(T)) + \|f\|_{\mathcal{F}}^2\\ \text{s.t.} \quad & f(0) = y_0,\\ & c_i(t)^\top f(t) \le d_i(t), \quad \forall\, t \in [0, T], \forall i \in [\mathcal{I}].\end{aligned}$$

ex: $g(f(T)) = \|y_T - f(T)\|_{\mathbb{R}^Q}^2$

Dealing with an infinite number of constraints: an overview

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

Relaxing

- Discretize constraint at “virtual“ samples $\{\tilde{x}_{m,i}\}_{m\leq M} \subset \mathcal{K}_i$,
  $\hookrightarrow$ no guarantees out-of-samples [Agrell, 2019, Takeuchi et al., 2006]
- Add constraint-inducing penalty, $\Omega_{\text{cons}}(f) = -\lambda \int_{\mathcal{K}_i} \min(0, D_i f(x) - b_i)\mathrm{d}x$
  $\hookrightarrow$ 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, $\Phi(x)^\top A\Phi(x)$, Sum-Of-Squares [Hall, 2018]
- Use only spaces $\mathcal{F}$ s.t. constraints have a “simple“ writing, e.g. splines
  $\hookrightarrow$ highly restricted functions classes [Papp and Alizadeh, 2014]
- **Our solution:** discretize $\mathcal{K}_i$ but replace b_i using RKHS geometry

Reproducing kernel Hilbert spaces (RKHS) at a glance (1)

A RKHS $(\mathcal{F}_k, \langle\cdot,\cdot\rangle_{\mathcal{F}_k})$ is a Hilbert space of real-valued[1] functions over a set $\mathcal{X}$ if one of the following equivalent conditions is satisfied [Aronszajn, 1950]

$\exists\, k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ s.t. $k_x(\cdot) = k(x,\cdot) \in \mathcal{F}_k$ and $f(x) = \langle f(\cdot), k_x(\cdot)\rangle_{\mathcal{F}_k}$

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

$$|f(x) - f_n(x)| = |\langle f - f_n, k_x\rangle_k| \leq \|f - f_n\|_k \|k_x\|_k = \|f - f_n\|_k \sqrt{k(x,x)}$$

k is s.t. $\exists\, \Phi_k : \mathcal{X} \to \mathcal{F}_k$ s.t. $k(x,y) \;=\; \langle \Phi_k(x), \Phi_k(y)\rangle_{\mathcal{F}_k}$, $\Phi_k(x) = k_x(\cdot)$

k is s.t. $\mathbf{G} = [k(x_i,x_j)]_{i,j=1}^n \succcurlyeq 0$ and $\mathcal{F}_k := \overline{\mathrm{span}(\{k_x(\cdot)\}_{x\in\mathcal{X}})}$, i.e. the completion for the pre-scalar product $\langle k_x(\cdot), k_y(\cdot)\rangle_{k,0} = k(x,y)$

[1] There is a natural extension to vector-valued RKHSs (more on this later).

Reproducing kernel Hilbert spaces (RKHS) at a glance (2)

- There is a one-to-one correspondence between kernels k and RKHSs $(\mathcal{F}_k, \langle \cdot, \cdot \rangle_{\mathcal{F}_k})$. Changing $\mathcal{X}$ or $\langle \cdot, \cdot \rangle_{\mathcal{F}_k}$ changes the kernel k.[2]
- for $\mathcal{X} \subset \mathbb{R}^d$, Sobolev spaces $\mathcal{H}^s(\mathcal{X})$ satisfying $s > d/2$ are RKHSs. For $\mathcal{X} = \mathbb{R}^d$ their (Matérn) kernels are well known. Classical kernels include

$$k_{\text{Gauss}}(x, y) = \exp\left(-\|x - y\|^2_{\mathbb{R}^d}/(2\sigma^2)\right) \quad k_{\text{lin}}(x, y) = \langle x, y \rangle_{\mathbb{R}^d}$$

- if $\mathcal{X} \subset \mathbb{R}^d$ is contained in the closure of its interior (e.g. $[0, +\infty[$, for $d = 1$), $k \in \mathcal{C}^{s,s}(\mathcal{X} \times \mathcal{X}, \mathbb{R})$, $D = \sum_j \gamma_j \partial^{\mathbf{r}_j}$ a differential operator of order at most s, then $\mathcal{F}_k \subset \mathcal{C}^s(\mathcal{X}, \mathbb{R})$ and reproducing formula for derivatives:

$$D_x k(x, \cdot) \in \mathcal{F}_k \quad ; \quad Df(x) = \langle f(\cdot), D_x k(x, \cdot) \rangle_{\mathcal{F}_k}$$

[2] It is hard to identify $\mathcal{F}_k$ given k, or k given $\mathcal{F}_k$ (more on this later).

Two essential tools for computations

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

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

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

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

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

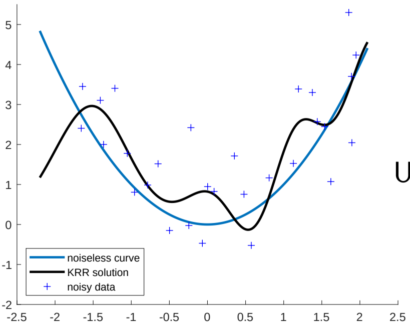
Finite number of evaluations $\implies$ finite number of coefficients

Kernel trick

$$\langle \sum_{n\in[N]} a_n k(x_n, \cdot), \sum_{m\in[M]} a'_m k(x'_m, \cdot)\rangle_k = \sum_{n\in[N]} \sum_{m\in[N']} a_n a'_m k(x_n, x'_m)$$

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

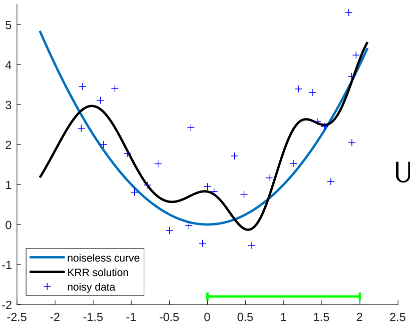
Example: 1D monotonic kernel ridge regression (KRR)



$$\min_{f \in \mathcal{F}_k} \frac{1}{N} \sum_{n=1}^{N} |y_n - f(x_n)|^2 + \lambda \|f\|_{\mathcal{F}_k}^2$$

Applying the representer theorem

Unconstrained KRR $\bar{f} = \sum_{n=1}^{N} \alpha_n k_{x_n}$,
$\boldsymbol{\alpha} = (\boldsymbol{G} + N\lambda \cdot \mathrm{Id})^{-1} \boldsymbol{y}$



$$\min_{f \in \mathcal{F}_k} \frac{1}{N} \sum_{n=1}^{N} |y_n - f(x_n)|^2 + \lambda \|f\|_{\mathcal{F}_k}^2$$

s.t. $0 \le f'(x), \forall x \in [0, 2]$

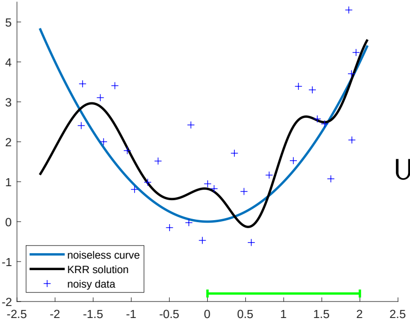
Unconstrained KRR $\bar{f} = \sum_{n=1}^{N} \alpha_n k_{x_n}$,

$$\boldsymbol{\alpha} = (\boldsymbol{G} + N\lambda \cdot \mathrm{Id})^{-1} \boldsymbol{y}$$

here is not monotonic on $[0, 2]$!

Infinite number of evaluations $\Rightarrow$ no representer theorem!
How to modify the problem to ensure constraint satisfaction?

Example: 1D monotonic kernel ridge regression (KRR)



$$\min_{f\in\mathcal{F}_k} \frac{1}{N}\sum_{n=1}^{N} |y_n - f(x_n)|^2 + \lambda \|f\|^2_{\mathcal{F}_k}$$

$$\text{s.t. } \boxed{\eta_m \|f\|_k \leq f'(\tilde{x}_m),\ \forall m \in [M]}$$

Unconstrained KRR $\bar{f} = \sum_{n=1}^{N} \alpha_n k_{x_n}$,
$\boldsymbol{\alpha} = (\boldsymbol{G} + N\lambda \cdot \mathrm{Id})^{-1}\boldsymbol{y}$

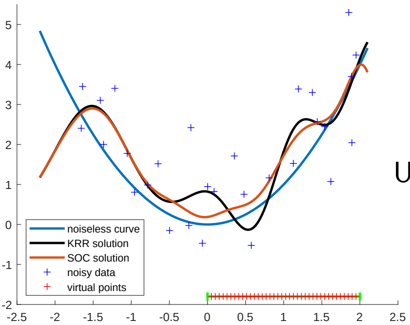
vs

Second-Order Cone (SOC) constrained KRR

Second-Order Cone constraints: $\{f \mid \|Af + b\|_k \leq c^\top f + d\}$
SOC comes from adding a buffer, $\eta_m > 0$, to a discretization, $\{\tilde{x}_m\}_{m\in[M]}$

LP$\subset$ QP $\subset$ SOCP $\subset$ SDP

Example: 1D monotonic kernel ridge regression (KRR)



$$\min_{f \in \mathcal{F}_k} \frac{1}{N} \sum_{n=1}^{N} |y_n - f(x_n)|^2 + \lambda \|f\|_{\mathcal{F}_k}^2$$

$$\text{s.t. } \boxed{\eta_m \|f\|_k \leq f'(\tilde{x}_m),\ \forall m \in [M]}$$

Unconstrained KRR $\bar{f} = \sum_{n=1}^{N} \alpha_n k_{x_n}$,

$$\boldsymbol{\alpha} = (\boldsymbol{G} + N\lambda \cdot \mathrm{Id})^{-1} \boldsymbol{y}$$

vs

Second-Order Cone (SOC) constrained KRR

Second-Order Cone constraints: $\{f \mid \|Af + b\|_k \leq c^\top f + d\}$

SOC comes from adding a buffer, $\eta_m > 0$, to a discretization, $\{\tilde{x}_m\}_{m \in [M]}$

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

This choice is related to continuity moduli.

Deriving SOC constraints through continuity moduli

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

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

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

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. $\|x - \tilde{x}_m\| \leq \delta$

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

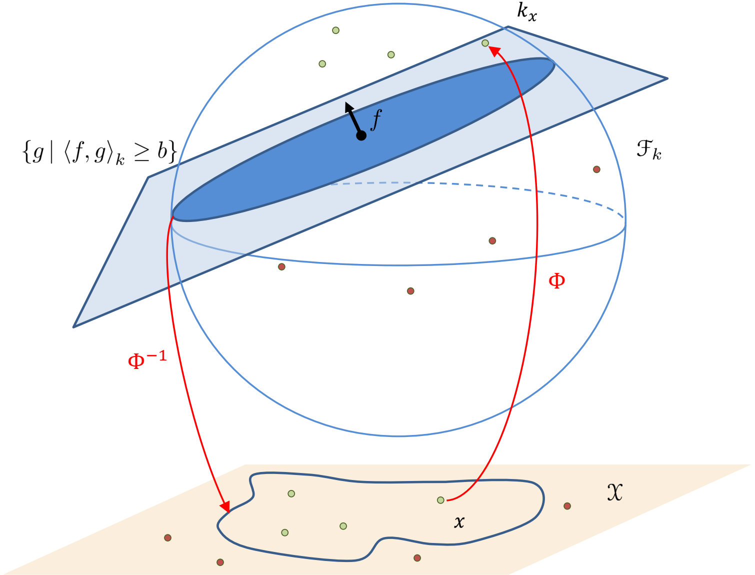
$$\omega_m(Df,\delta) := \sup_{\{x \mid \|x-\tilde{x}_m\|\leq\delta\}} |Df(x) - Df(\tilde{x}_m)| \leq \eta_m(\delta)\|f(\cdot)\|_k$$

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

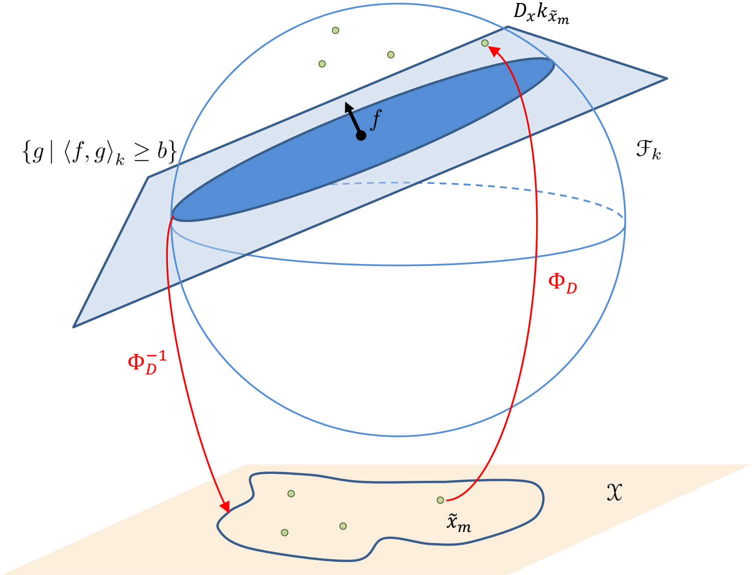
$$\text{“}b \leq Df(x),\ \forall x \in \mathcal{K}\text{“} \Leftarrow \text{“}b + \omega_m(Df,\delta) \leq Df(\tilde{x}_m),\ \forall m \in [M]\text{“}$$
$$\Leftarrow \text{“}b + \eta_m\|f(\cdot)\| \leq Df(\tilde{x}_m),\ \forall m \in [M]$$

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

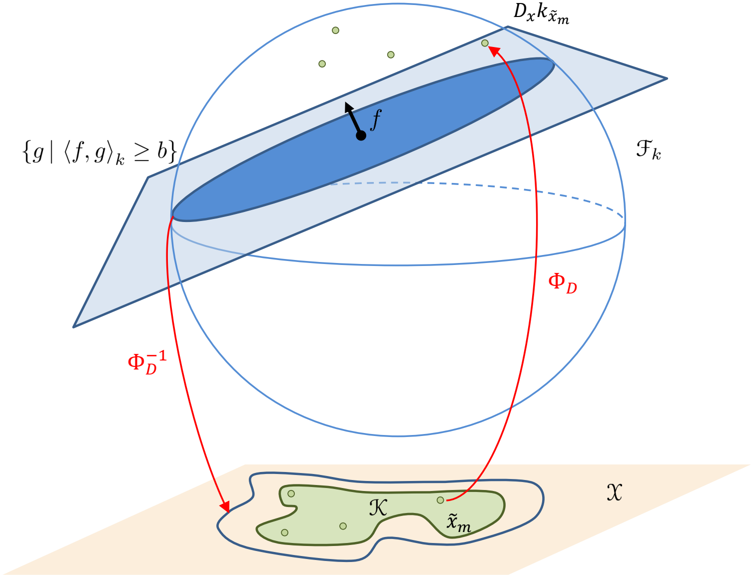
There is also a geometrical interpretation for this choice of η_m.



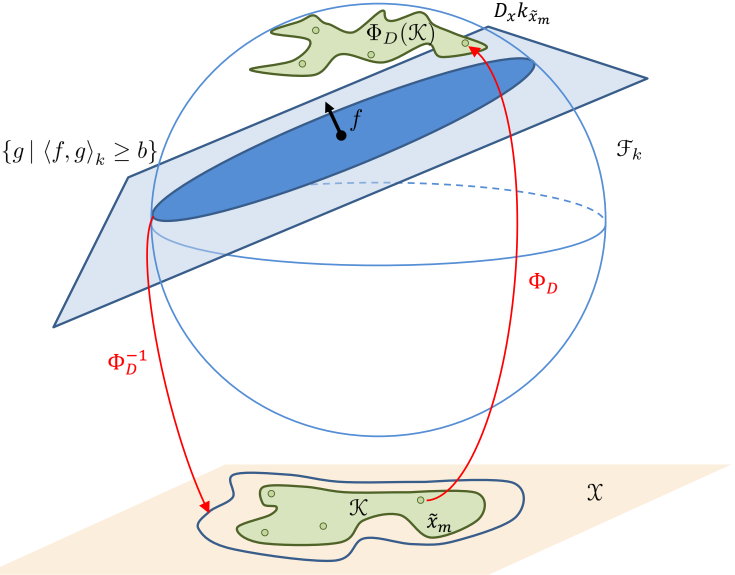
Support Vector Machine (SVM) is about separating red and green points by blue hyperplane.



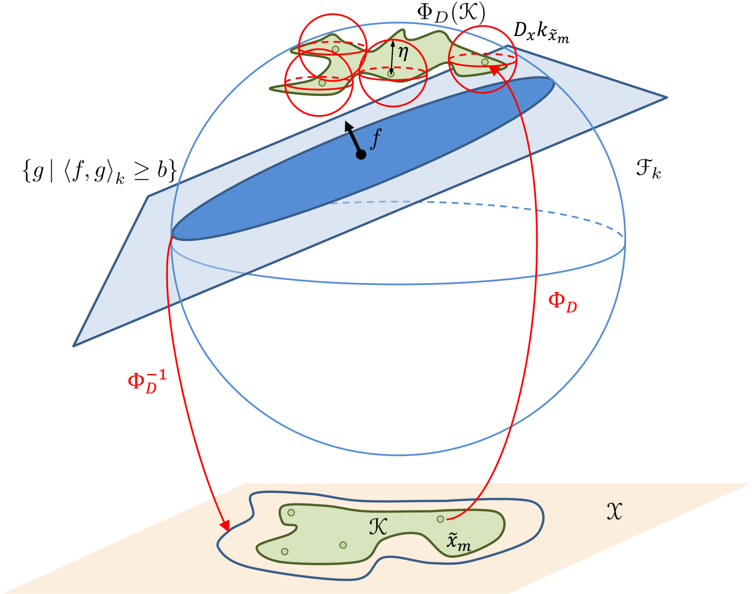
Using the nonlinear embedding $\Phi_D : x \mapsto D_x k(x, \cdot)$, the idea is the same. Consider only the green points, it looks like one-class SVM.



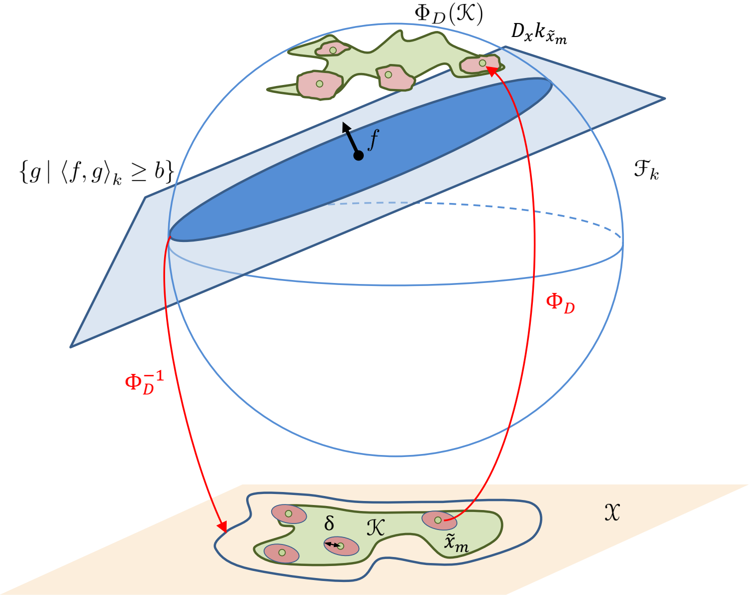
The green points are now samples of a compact set $\mathcal{K}$.



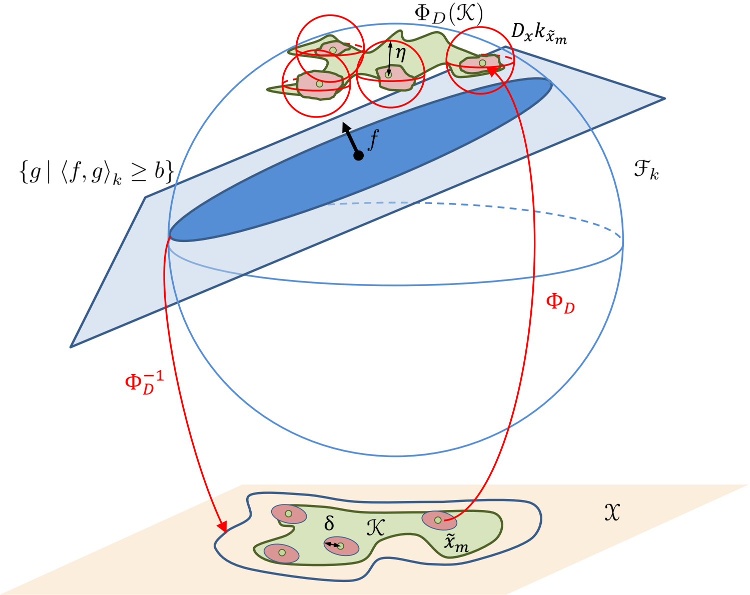
The image $\Phi_D(\mathcal{K})$ looks ugly...



The image $\Phi_D(\mathcal{K})$ looks ugly, can we cover it by balls? How to choose 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 theorem

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

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

where $\mathcal{B}$ is a closed convex constraint set over $(b_i)_{i\in[\mathcal{I}]}$. If $\Omega(\cdot)$ is strictly increasing, then

Theoretical guarantees [Aubin-Frankowski and 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 $\mathcal{L}$ is μ-strongly convex, we have **bounds**: computable/theoretical[a]

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

[a] Assuming $\mathcal{B} = \mathbb{R}^{\mathcal{I}}$ for the *a priori* bound to hold.

Discussion

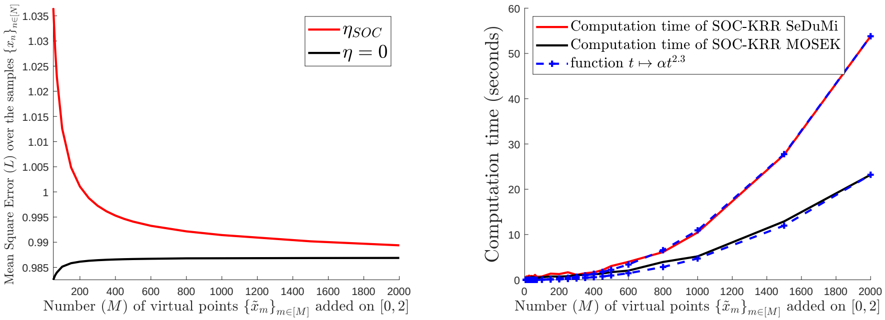
(i) This theorem holds for given samples $(x_n, y_n)_{n\in[N]}$ (optimization rather than statistical properties - no asymptotics)

(ii) The representer theorem provides an equivalent finite-dimensional problem of size $N+M$ with SOC constraints $\sim \mathcal{O}((N+M)^3)$

(iii) Better bound $\equiv$ smaller $\eta \equiv$ smaller $\delta \equiv$ larger $M \equiv$ costly in time

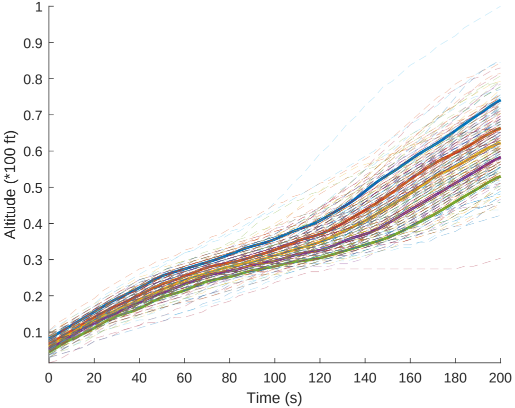
(iv) The virtual points can be chosen among the samples (*recycling*)

KRR example



Joint quantile regression (JQR): airplane data

Airplane trajectories at takeoff have increasing altitude.



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

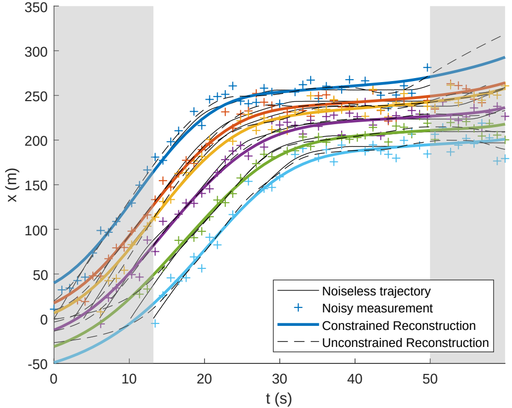
Increasing quantiles should be non-crossing

Data provided by ENAC (flights Paris→Toulouse) [Nicol, 2013]

Two shape constraints jointly handled with $15k$ samples.
Works with higher dimensions too!

Kernel ridge regression (KRR): trajectory reconstruction

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



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

Forward trajectories also maintain security distance

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

(In *Kernel Regression for Vehicle Trajectory Reconstruction under Speed and Inter-vehicular Distance Constraints*, PCAF and Nicolas Petit and Zoltán Szabó *IFAC World Congress 2020*)

Teaser slide

This approach works as well for

- Other compact coverings than balls, iterative covering
- SDP constraints (e.g. convexity for $d \geq 2$): $0 \preccurlyeq \mathbf{Hess}(f)(x)$
- Vector-valued functions $f : \mathcal{X} \to \mathbb{R}^Q$
- Other applications: finance, control theory,...

Control: Take $\mathcal{F}_k$ to be a Hilbert space of trajectories $[0, T] \to \mathbb{R}^Q$

$$\begin{aligned} \min_{x(\cdot) \in \mathcal{F}_k} \quad & g(x(T)) + \|x(\cdot)\|_k^2 \\ \text{s.t.} \quad & x(0) = x_0, \\ & c_i(t)^\top x(t) \leq d_i(t), \quad \forall\, t \in [0, T], \forall i \in [\mathcal{I}]. \end{aligned}$$

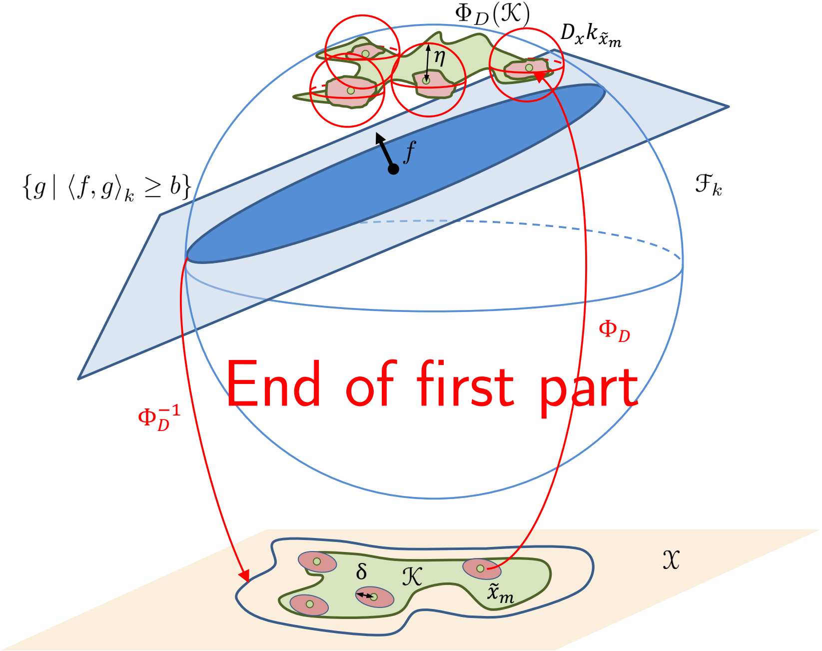
Partial conclusion/Take-home message

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 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!
$\mathcal{X} \subset \mathbb{R}^d$, $d \gg 1$

But the control problem is only defined over $\mathcal{X} = [0, T]$ ($d = 1$)!



End of first part

Linearly-constrained Linear Quadratic Regulator (LQR)

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

$$\begin{aligned} \min_{x(\cdot),u(\cdot)} \quad & g(x(T)) + \int_0^T [x(t)^\top Q(t)x(t) + u(t)^\top R(t)u(t)]\mathrm{d}t \\ \text{s.t.} \quad & x(0) = x_0, \\ & x'(t) = A(t)x(t) + B(t)u(t), \text{ a.e. in } [0,T], \\ & c_i(t)^\top x(t) \le d_i(t), \forall\, t \in [0,T], \forall\, i \in [\![1,P]\!], \end{aligned}$$

with state $x(t) \in \mathbb{R}^N$, control $u(t) \in \mathbb{R}^M$, $A(\cdot) \in L^1(0,T)$, $B(\cdot) \in L^2(0,T)$, $Q(t) \succcurlyeq 0$ and $R(t) \succcurlyeq r\mathrm{Id}_M$

Linearly-constrained Linear Quadratic Regulator (LQR)

Problem of time-varying linear quadratic optimal control with finite horizon and affine inequality state constraints with $Q \equiv 0$ and $R \equiv \mathrm{Id}_M$

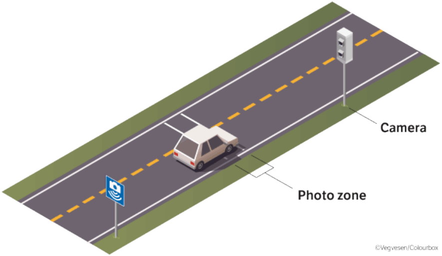
$$
\begin{aligned}
\min_{x(\cdot),u(\cdot)} \quad & g(x(T)) + \int_0^T \|u(t)\|_{\mathbb{R}^M}^2 \,\mathrm{d}t \\
\text{s.t.} \quad & x(0) = x_0, \\
& x'(t) = A(t)x(t) + B(t)u(t), \text{ a.e. in } [0, T], \\
& c_i(t)^\top x(t) \leq d_i(t), \,\forall\, t \in [0, T], \forall\, i \in \{1, \dots, P\},
\end{aligned}
$$

with state $x(t) \in \mathbb{R}^N$, control $u(t) \in \mathbb{R}^M$, $A(\cdot) \in L^1(0, T)$, $B(\cdot) \in L^2(0, T)$, $x(\cdot) : [0, T] \to \mathbb{R}^N$ absolutely continuous and $u(\cdot) \in L^2(0, T)$.

Why are state constraints difficult to study?

- **Theoretical obstacle**: Pontryagine's Maximum Principle involves not only an adjoint vector $p(t)$ 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.

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

We have a vector space $\mathcal{S}$ of trajectories $x(\cdot) : [0, T] \to \mathbb{R}^N$

$$\mathcal{S} := \{x(\cdot) \,|\, \exists\, u(\cdot) \in L^2(0, T) \text{ s.t. } x'(t) = A(t)x(t) + B(t)u(t) \text{ a.e. } \}$$

The space of trajectories $\mathcal{S}$ depends on T, $A(\cdot)$, $B(\cdot)$.

LQR (Linear Quadratic Regulator)

$$\min_{\substack{x(\cdot)\in\mathcal{S} \\ u(\cdot)\in L^2(0,T)}} \quad g(x(T)) + \|u(\cdot)\|^2_{L^2(0,T)}$$

$x(0) = x_0$
$c_i(t)^\top x(t) \leq d_i(t), t \in [0, T], i \leq P$

“KRR“ (Kernel Ridge Regression)

$$\min_{x(\cdot)\in\mathcal{S}} \; g(x(T)) + \lambda \|x(\cdot)\|^2_{\mathcal{S}}$$

$x(0) = x_0$
$c_i(t)^\top x(t) \leq d_i(t), t \in [0, T], i \leq P$

Is $\mathcal{S}$ a RKHS? For which inner product?

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 $\mathbb{R}^N$-vector-valued functions defined on $\mathcal{T}$ is a vRKHS if there exists a matrix-valued kernel $K : \mathcal{T} \times \mathcal{T} \to \mathbb{R}^{N \times N}$ such that the reproducing property holds:

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

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

There is also 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.

Representer theorem in vRKHSs

Theorem (Representer theorem with SOC constraints)

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*

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

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

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

Application to linear control systems with quadratic cost

$$\mathcal{S} := \{x(\cdot) \in W^{1,1} \,|\, \exists\, u(\cdot) \in L^2(0,T) \text{ s.t. } x'(t) = A(t)x(t) + B(t)u(t) \text{ a.e. }\}$$

Given $x(\cdot) \in \mathcal{S}$, for the pseudoinverse $B(t)^{\ominus}$ of $B(t)$, set

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

$$\langle x_1(\cdot), x_2(\cdot)\rangle_K := x_1(0)^\top x_2(0) + \int_0^T u_1(t)^\top u_2(t)\mathrm{d}t$$

Lemma

$(\mathcal{S}, \langle\cdot,\cdot\rangle_K)$ is a vRKHS with uniformly continuous $K(\cdot,\cdot)$.

$\|\cdot\|_K$ is a Sobolev-like norm split into two semi-norms

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

Splitting $\mathcal{S}$ into subspaces and identifying their kernels

$\mathcal{S}_0 := \{x(\cdot) \,|\, x'(t) = A(t)x(t), \text{ a.e. in } [0,T]\}$ $\qquad \|x(\cdot)\|_{K_0}^2 = \|x(0)\|^2$

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

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

Splitting $\mathcal{S}$ into subspaces and identifying their kernels

$\mathcal{S}_0 := \{x(\cdot) \,|\, x'(t) = A(t)x(t), \text{ a.e. in } [0, T]\}$ $\quad \|x(\cdot)\|^2_{K_0} = \|x(0)\|^2$

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

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

$$K_0(s, t) = \Phi_A(s, 0)\Phi_A(t, 0)^\top$$

Only using the reproducing property and that for $x(\cdot) \in \mathcal{S}$,

$$x(t) = \Phi_A(t, 0)x(0) + \int_0^t \Phi_A(t, \tau)B(\tau)u(\tau)\mathrm{d}\tau. \tag{1}$$

For fixed t, define control matrix $U_t(s) := \begin{cases} B(s)^\top \Phi_A(t, s)^\top & \forall s \leq t, \\ 0 & \forall s > t. \end{cases}$

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

$$K_1(s, t) = \int_0^{\min(s,t)} \Phi_A(s, \tau)B(\tau)B(\tau)^\top \Phi_A(t, \tau)^\top \mathrm{d}\tau.$$

Examples: controllability Gramian/transversality condition

Steer a point from $(0,0)$ to (T, x_T), with e.g. $g(x(T)) = \|x_T - x(T)\|_N^2$

Exact planning ($x(T) = x_T$)

$$\min_{\substack{x(\cdot)\in\mathcal{S}\\ x(0)=0}} \chi_{x_T}(x(T)) + \frac{1}{2}\|u(\cdot)\|^2_{L^2(0,T)}$$

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

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

As, $x(0) = 0$, applying the representer theorem: $\exists\, p_T$, $\bar{x}(\cdot) = K_1(\cdot, T)p_T$

Controllability Gramian

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

$\bar{x}(T) = x_T \Leftrightarrow x_T \in \operatorname{Im}(K_1(T,T))$

Transversality Condition

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

Take $p_T = -\nabla g(\bar{x}(T))$

From affine state constraints to SOC constraints

Take (t_m, δ_m) such that $[0, T] \subset \cup_{m \in [\![1, N_P]\!]} [t_m - \delta_m, t_m + \delta_m]$, take

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

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

We have strengthened SOC constraints that enable a representer theorem

$$\eta_i(\delta_m, t_m)\|x(\cdot)\|_K + c_i(t_m)^\top x(t_m) \leq d_i(\delta_m, t_m), \forall\, m \in [\![1, N_P]\!], \forall\, i \in [\![1, P]\!]$$

$$\Downarrow$$

$$c_i(t)^\top x(t) \leq d_i(t), \forall\, t \in [0, T], \forall\, i \in [\![1, P]\!]$$

Lemma (Uniform continuity of tightened constraints)

As $K(\cdot, \cdot)$ is UC, if $c_i(\cdot)$ and $d_i(\cdot)$ are $\mathcal{C}^0$-continuous, when $\delta \to 0^+$, $\eta_i(\cdot, t)$ converges to 0 and $d_i(\cdot, t)$ converges to $d_i(t)$, uniformly w.r.t. t.

Main theorem

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

(H-sol) $c_i(0)x_0 < d_i(0)$ and there exists a trajectory $x^\epsilon(\cdot) \in \mathcal{S}$ satisfying strictly the affine constraints, as well as the initial condition.[3]

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

Theorem (Existence and Approximation by SOC constraints)

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* $[0,T] \subset \cup_{m\in[\![1,N_0]\!]}[t_m - \delta_m, t_m + \delta_m]$ *satisfying* $\overline{\delta} \geq \max_{m\in[\![1,N_0]\!]} \delta_m$*,*

$$\frac{1}{\gamma_K} \cdot \sup_{t\in[0,T]} \|\bar{x}_\eta(t) - \bar{x}(t)\| \leq \|\bar{x}_\eta(\cdot) - \bar{x}(\cdot)\|_K \leq \rho.$$

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

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

Numerical example: constrained pendulum - definition

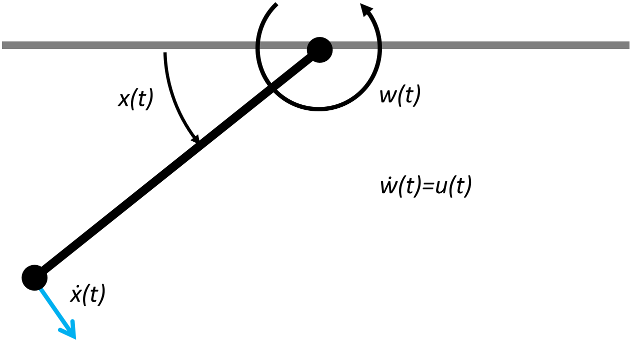
Constrained pendulum when controlling the third derivative of the angle

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

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

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

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



Numerical example: constrained pendulum - definition

Constrained pendulum when controlling the third derivative of the angle

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

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

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

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

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

$$\eta_{\dot{x}}\|x(\cdot)\|_K - \dot{x}(t_m) \leq \text{-}3,$$
$$\eta_w\|x(\cdot)\|_K + w(t_m) \leq 10,$$
$$\eta_w\|x(\cdot)\|_K - w(t_m) \leq 10$$

$$\bar{x}(\cdot) = K(\cdot, 0)p_0 + K(\cdot, T/3)p_{T/3} + K(\cdot, T)p_T + \sum_{m=1}^{M} K(\cdot, t_m)p_m$$

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

Numerical example: constrained pendulum - illustration

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

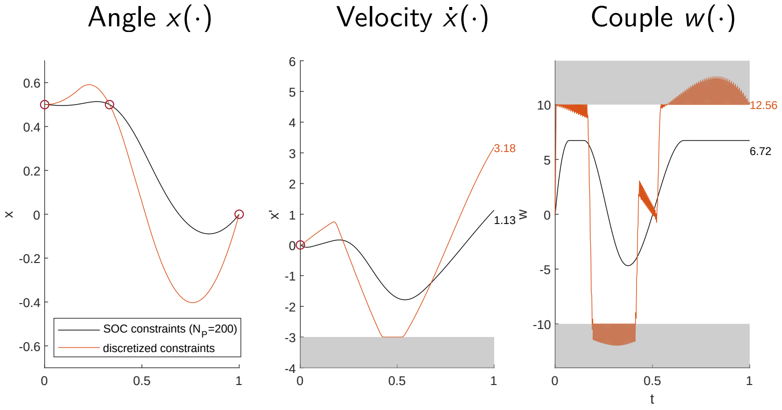


Figure: Comparison of SOC constraints (guaranteed η_w) vs discretized constraints ($\eta_w = 0$) for $N_P = 200$.

Numerical example: constrained pendulum - illustration

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

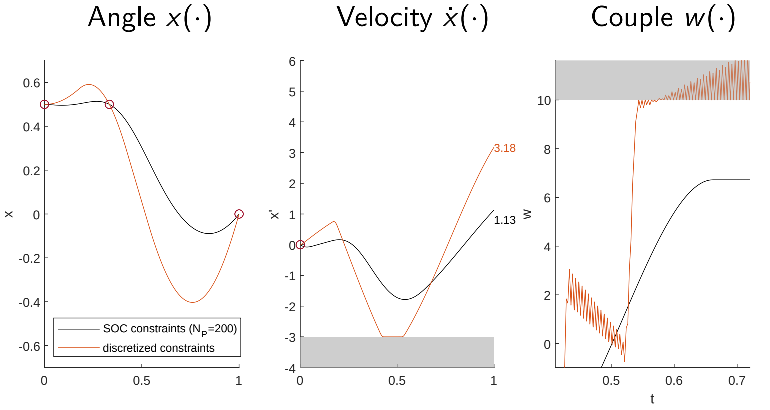


Figure: Comparison of SOC constraints (guaranteed η_w) vs discretized constraints ($\eta_w = 0$) for $N_P = 200$ - Chattering phenomenon!.

Numerical example: 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 $w(\cdot)$

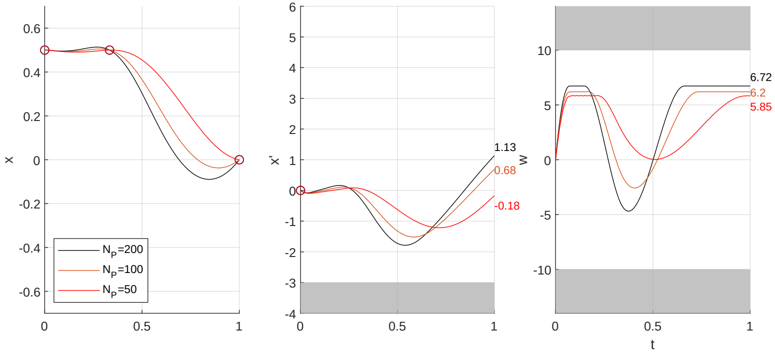


Figure: Comparison of SOC constraints for varying N_P and guaranteed η_w.

Numerical example: 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 $w(\cdot)$

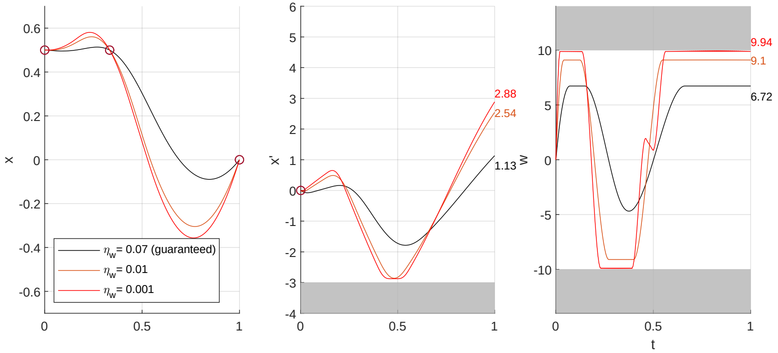


Figure: Comparison of SOC constraints for varying η_w and $N_P = 200$.

Pushing RKHSs beyond/Revisiting classical 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
  $\hookrightarrow$ you may need Banach kernels (e.g. $\|u(\cdot)\|^2_{L^2(0,T)} \to \|u(\cdot)\|_{L^1(0,T)}$)

For control theory

- To each evaluation at time t corresponds a covector $p_t \in \mathbb{R}^N$
  $\hookrightarrow$ 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

General conclusion

Shape constraints in RKHSs

We have seen how to tighten in RKHSs an infinite number of pointwise affine constraints over a compact set into finitely many SOC constraints.

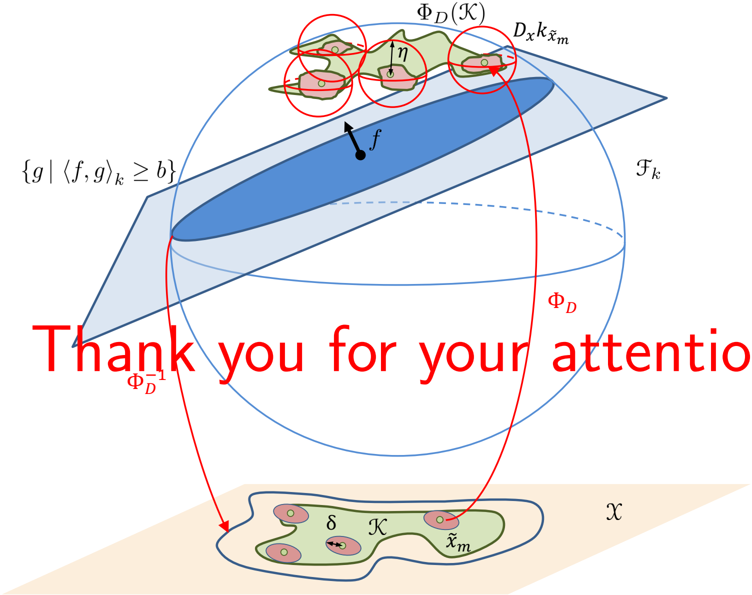
- tightening intractable constraints is the only way to have guarantees
- compact coverings in infinite dimensional spaces provide a solution

Linear Quadratic Regulator as a kernel regression

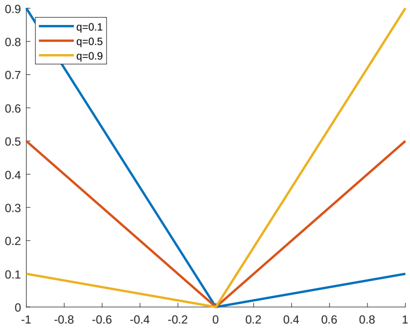
We have seen that state-constrained LQR is a non-trivial 1D example of shape constraints that

- allows to revisit classical notions from the kernel viewpoint
- allows to deal with the difficult problem of state constraints

Thank you for your attention!



Appendix: 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, over $(f_q)_{q\in[Q]}$:

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

$$\text{s.t. } f_{q+1}(x) \geq f_q(x), \forall q \in [Q-1], \forall x \in [\min x_n, \max x_n]^d.$$

Appendix: 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 ± std of 100×value of the pinball loss (smaller is better)

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

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. $D^*D\, G_{D^*D,x}(y) = \delta_x(y)$ then, if the integrals over the boundaries in Green's formula are null, for any $f \in \mathcal{F}_k$

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

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, $D\, G_{D,x}(y) = \delta_x(y)$, 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: IPC gives strictly feasible trajectories

(H-sol) $C(0)x_0 < d(0)$ and there exists a trajectory $x^\epsilon(\cdot) \in \mathcal{S}$ satisfying strictly the affine constraints, as well as the initial condition.

(H1) $A(\cdot)$ and $B(\cdot)$ are $\mathcal{C}^0$. $C(\cdot)$ and $d(\cdot)$ are $\mathcal{C}^1$ and $C(0)x_0 < d(0)$.

(H2) There exists $M_u > 0$ s.t. , for all $t \in [0, T]$ and $x \in \mathbb{R}^N$ satisfying $C(t)x \leq d(t)$, and $\|x\| \leq (1 + \|x_0\|)e^{T\|A(\cdot)\|_{L^\infty(0,T)} + TM_u\|B(\cdot)\|_{L^\infty(0,T)}}$, there exists $u_{t,x} \in M_u\mathbb{B}_M$ such that

$$\forall\, i \in \{j \,|\, c_j(t)^\top x = d_j(t)\},\; c_i'(t)^\top x - d_i'(t) + c_i(t)^\top(A(t)x + B(t)u_{t,x}) < 0.$$

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

$$\eta_i(\delta, t) := \sup \|K(\cdot, t)c_i(t) - K(\cdot, s)c_i(s)\|_K, \quad \omega_i(\delta, t) := \sup |d_i(t) - d_i(s)|,$$
$$d_i(\delta_m, t_m) := \inf d_i(s), \qquad \text{over } s \in [t_m - \delta_m, t_m + \delta_m] \cap [0, T]$$

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

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

Proposition (Nested sequence)

Let $\delta_{max} := \max_{m\in[\![1,N_0]\!]} \delta_m$*. For any* $\delta \ge \delta_{max}$*, if, for a given* $y_0 \ge 0$*,* $\epsilon_i \ge \sup_{t\in[0,T]}[\eta_i(\delta, t)y_0 + \omega_i(\delta, t)]$*, then we have a nested sequence*

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

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

Annex: List of shape constraints

- **Monotonicity w.r.t. partial ordering**:

$$\partial^{e_1} f(x) \geq \ldots \geq \partial^{e_d} f(x) \geq 0 \quad (\forall x).$$

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

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

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

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. and Szabó, Z. (2020).
Hard shape-constrained kernel machines.
https://arxiv.org/abs/2005.12636.

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

Brault, R., Lambert, A., Szabo, Z., Sangnier, M., and d'Alche Buc, F. (2019).
Infinite task learning in RKHSs.
volume 89 of *Proceedings of Machine Learning Research*, pages 1294–1302. PMLR.

References II

Buisson, C., Villegas, D., and Rivoirard, L. (2016).
Using polar coordinates to filter trajectories data without adding extra physical constraints.

Hall, G. (2018).
Optimization over nonnegative and convex polynomials with and without semidefinite programming.
PhD Thesis, Princeton University.

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.
Econometric Society Monographs. Cambridge University Press.

References III

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)*.

Papp, D. and Alizadeh, F. (2014).
Shape-constrained estimation using nonnegative splines.
Journal of Computational and Graphical Statistics, 23(1):211–231.

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

References IV

Sangnier, M., Fercoq, O., and d'Alché Buc, F. (2016).
Joint quantile regression in vector-valued RKHSs.
Advances in Neural Information Processing Systems (NIPS), pages 3693–3701.

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

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