Kernel representation of non-negative functions with applications in non-convex optimization and beyond

Pierre-Cyril Aubin-Frankowski and Alessandro Rudi

DI, Ecole normale supérieure, Université PSL, CNRS,
INRIA SIERRA, Paris, France

EURO-ESPOO 2022

6th of July, 2022

Examples of constraints in function optimization - 1

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“
  $x''(t) \in [a_{low}, a_{high}]$

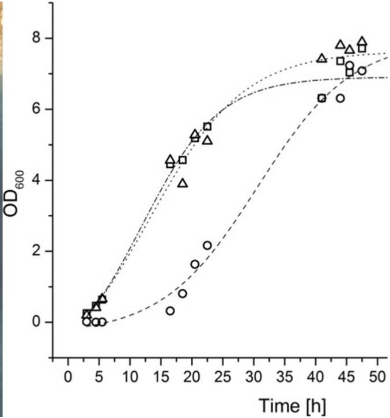
Physical constraints

$\hookrightarrow$ provides feasible trajectories in path-planning

This consists in an infinite number of pointwise constraints!

Examples of constraints in function optimization - 2

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,...

Examples of constraints in function optimization - 3

- **Global optimization of smooth (nonconvex) g:**

$$\max_{\substack{c\in\mathbb{R}\\ c\geq g(x),\,\forall x\in\mathcal{X}}} c \qquad (=\min_{x\in\mathcal{X}} g(x))$$

- **Density estimation with relative entropy:**

$$\min_{\substack{f\in\mathcal{C}(\mathcal{X},\mathbb{R}),\,\int_{\mathcal{X}} f(x)\mathrm{d}x=1\\ f(x)\geq 0,\,\forall x\in\mathcal{X}}} -\int_{\mathcal{X}} \log(f(x))\mathrm{d}\mu(x) \qquad (=\mathrm{KL}(\mu,\mu_f)+\mathrm{cst})$$

- **Optimal transport in its dual formulation:**

$$\max_{\substack{u,v\in C(\mathcal{X},\mathbb{R})\\ u(x)+v(y)\leq c(x,y),\,\forall x,y\in\mathcal{X}\times\mathcal{X}}} \int_{\mathcal{X}} u(x)\mathrm{d}\mu(x)+\int_{\mathcal{X}} v(y)\mathrm{d}\nu(y) \qquad (=\mathrm{OT}_c(\mu,\nu))$$

Other problems/extensions: Joint Quantile Regression (JQR), handling constrained derivatives, vector or SDP-valued functions,. . . methods presented in this talk used in [Aubin-Frankowski and Sz Marteau-Ferey et al., 2020a, Vacher et al., 2021, Rudi et al., 2020, Muzellec et al., 2021]

Dealing with an infinite number of constraints: an overview

$$\bar{f} \in \operatorname{argmin}_{f \in \mathcal{H}_k} \mathcal{L}(f) \text{ s.t. } "0 \leq f(x), \forall x \in \mathcal{K}", \mathcal{K} \subset \mathbb{R}^d \text{ non-finite (compact)}$$

Relaxing

- Discretize constraint at “virtual“ samples $\{\tilde{x}_m\}_{m \in [M]} \subset \mathcal{K}$,
  $\hookrightarrow$ no guarantees out-of-samples [Agrell, 2019, Takeuchi et al., 2006]
- Add constraint-inducing penalty, $R_{\text{cons}}(f) = -\lambda \int_{\mathcal{K}} \min(0, f(x)) \mathrm{d}x$
  $\hookrightarrow$ no guarantees, changes the problem objective [Brault et al., 2019]
- Replace inequality by equality to nonnegative function $\Phi(x)^\top A \Phi(x)$ then discretize
  $\hookrightarrow$ **generic**: bounded amount of violation, extra SDP variable A [Muzellec et al., 2021]

Tightening

- Replace inequality by equality to nonnegative function $\Phi(x)^\top A \Phi(x)$ and optimize over A
  $\hookrightarrow$ **non-generic**: only specific classes of functions [Marteau-Ferey et al., 2020b];
- Discretize but replace 0 by $\eta_m \|f\|$ [Aubin-Frankowski and Szabó, 2020a]
  $\hookrightarrow$ **generic**: no violation, second-order cone constraints, but extra tightening

Table of Contents

Our battle horse: the Reproducing kernel Hilbert space (RKHS)

A RKHS $(\mathcal{H}_k, \langle \cdot, \cdot \rangle_{\mathcal{H}_k})$ is a Hilbert space of real-valued 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. $k_x(\cdot) = k(x, \cdot) \in \mathcal{H}_k$ and $f(x) = \langle f(\cdot), k_x(\cdot) \rangle_{\mathcal{H}_k}$ for all $x \in \mathcal{X}$ and $f \in \mathcal{H}_k$ (reproducing property)

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

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

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

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

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

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

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

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

Kernel trick

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

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

A nice class of nonnegative functions: kernel Sum-of-Squares/PSD models

How to build a nonnegative function given a kernel $\Phi_k(x) = k(\cdot, x)$? Square it!

$$f : x \mapsto \langle \Phi_k(x), \Phi_k(x) \rangle_{\mathcal{H}_k} = k(x, x) \geq 0$$

More generally take a positive semidefinite operator $A \in S^+(\mathcal{H}_k)$,

$$f_A : x \mapsto \langle \Phi_k(x), A\Phi_k(x) \rangle_{\mathcal{H}_k} \geq 0$$

$$\text{(PSD model) } A = \sum_{i,j=1}^{N} a_{ij} \Phi_k(x_i) \otimes \Phi_k(x_j) \implies f_A(x) = \sum_{i,j=1}^{N} a_{ij} k(x, x_i) k(x, x_j)$$

$$\text{(kernel SoS) } [a_{ij}]_{i,j} = \sum_i \mathbf{u}_i \mathbf{u}_i^\top \text{ (SVD)} \implies f_A(x) = \sum_{i=1}^{N} (\sum_{j=1}^{N} u_{i,j} k(x, x_j))^2$$

Note that in general $f_A \notin \mathcal{H}_k$ but $f_A \in \mathcal{H}_k \odot \mathcal{H}_k$ (Hadamard product). If $\text{span}(\{k_x(\cdot)\}_{x \in \mathcal{X}})$ is dense in continuous functions, so are the $\{f_A\}_{A \in S^+(\mathcal{H}_k)}$ in nonnegative functions

What I am looking for: an approximation framework

Example: optimizing over vector fields $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^P$ constrained over compact set $\mathcal{K}$ to belong to a set $\mathbf{F}(x)$

Optimization over $\mathcal{F} \overset{e.g}{=} \mathcal{C}(\mathcal{X}, \mathbb{R})$ or $L(\mu)$

$$\min_{\mathbf{f} = [f_1; \ldots; f_P] \in \mathcal{F}^P} \int l(x, \mathbf{f}(x)) \mathrm{d}\mu(x)$$

$$\text{s.t.} \quad \mathbf{f}(x) \in \mathbf{F}(x), \forall x \in \mathcal{K}$$

What I am looking for: an approximation framework

Example: optimizing over vector fields $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^P$ constrained over compact set $\mathcal{K}$ to belong to a **convex** set $\mathbf{F}(x)$

Optimization over $\mathcal{F} \overset{e.g}{=} \mathcal{C}(\mathcal{X}, \mathbb{R})$ or $L(\mu)$

$$\min_{\mathbf{f} = [f_1; \ldots; f_P] \in \mathcal{F}^P} \int l(x, \mathbf{f}(x)) \mathrm{d}\mu(x)$$

$$\text{s.t.} \quad \mathbf{c}_i(x)^\top \mathbf{f}(x) + d_i(x) \geq 0, \ \forall\, i \in [I], \ \forall\, x \in \mathcal{K}$$

Infinite number of affine constraints!

What I am looking for: an approximation framework

Example: optimizing over vector fields $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^P$ constrained over compact set $\mathcal{K}$ to belong to a **convex** set $\mathbf{F}(x)$

Optimization over $\mathcal{F} \overset{e.g}{=} \mathcal{C}(\mathcal{X}, \mathbb{R})$ or $L(\mu)$

$$\min_{\mathbf{f} = [f_1; \ldots; f_P] \in \mathcal{F}^P} \int l(x, \mathbf{f}(x)) \mathrm{d}\mu(x)$$

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

$$\mathbf{c}_i(x)^\top \mathbf{f}(x) + d_i(x) \geq 0, \ \forall\, i \in [I], \ \forall\, x \in \mathcal{K}$$

Empirical approx. through RKHS $\mathcal{H}_k$

$$\min_{\mathbf{f} \in \mathcal{H}_k^P} \sum_{n \in [N]} l(x_n, \mathbf{f}(x_n)) + \lambda \|\mathbf{f}\|^2_{\mathcal{H}_k^P}$$

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

$$\mathbf{c}_i(x_m)^\top \mathbf{f}(x_m) + d_i(x_m) \underset{=}{\overset{\geq}{=}} ?, \ \forall\, i \in [I], \ \forall\, m \in [M]$$

Infinite number of affine constraints!

What I am looking for: an approximation framework

Example: optimizing over vector fields $\mathbf{f} : \mathbb{R}^d \to \mathbb{R}^P$ constrained over compact set $\mathcal{K}$ to belong to a **convex** set $\mathbf{F}(x)$

Optimization over $\mathcal{F} \overset{e.g}{=} \mathcal{C}(\mathcal{X}, \mathbb{R})$ or $L(\mu)$

$$\min_{\mathbf{f} = [f_1; \dots; f_P] \in \mathcal{F}^P} \int l(x, \mathbf{f}(x)) \mathrm{d}\mu(x)$$

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

$$\mathbf{c}_i(x)^\top \mathbf{f}(x) + d_i(x) \geq 0, \forall\, i \in [I], \forall\, x \in \mathcal{K}$$

Infinite number of affine constraints!

Empirical approx. through RKHS $\mathcal{H}_k$

$$\min_{\mathbf{f} \in \mathcal{H}_k^P} \sum_{n \in [N]} l(x_n, \mathbf{f}(x_n)) + \lambda \|\mathbf{f}\|^2_{\mathcal{H}_k^P}$$

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

$$\mathbf{c}_i(x_m)^\top \mathbf{f}(x_m) + d_i(x_m) \overset{\geq}{\underset{=}{=}} ?, \forall\, i \in [I], \forall\, m \in [M]$$

Finite number of constraints for $\{x_m\}_m \subset \mathcal{K}$!

Statement of simpler problem

Given points $(x_n)_{n\in[N]} \in \mathcal{X}^N$, a *loss* $L : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$, a *regularizer* $R : \mathbb{R}_+ \to \mathbb{R}$, a RKHS $\mathcal{H}_k$ of smooth functions from $\mathcal{X}$ to $\mathbb{R}$ and a compact set $\mathcal{K} \subset \mathcal{X}$.

$$\begin{aligned} \bar{f}^0 \in \underset{f \in \mathcal{H}_k}{\arg\min}\ & \mathcal{L}(f) = L\left(\left(f(x_n)\right)_{n\in[N]}\right) + R\left(\|f\|_{\mathcal{H}_k}\right) \\ \text{s.t.}\quad & 0 \le f(x), \quad \forall x \in \mathcal{K}. \end{aligned}$$

Idea to overcome non-finiteness: Discretize constraint at "virtual" samples $\{\tilde{x}_m\}_{m\in[M]} \subset \mathcal{K}$, use the fill distance: $h_M = \sup_{x\in\mathcal{K}} d(x, \{\tilde{x}_m\}_{m\in[M]})$ to bound $|L(\bar{\mathbf{f}}^{approx}) - L(\bar{\mathbf{f}}^0)|$

Statement of simpler problem

Given points $(x_n)_{n\in[N]} \in \mathcal{X}^N$, a *loss* $L : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$, a *regularizer* $R : \mathbb{R}_+ \to \mathbb{R}$, a RKHS $\mathcal{H}_k$ of smooth functions from $\mathcal{X}$ to $\mathbb{R}$ and a compact set $\mathcal{K} \subset \mathcal{X}$.

$$\bar{f}^0 \in \underset{f \in \mathcal{H}_k}{\arg\min} \; \mathcal{L}(f) = L\left((f(x_n))_{n\in[N]}\right) + R\left(\|f\|_{\mathcal{H}_k}\right)$$
$$\text{s.t.} \qquad 0 \le f(x), \quad \forall x \in \mathcal{K}.$$

Idea to overcome non-finiteness: Discretize constraint at "virtual" samples $\{\tilde{x}_m\}_{m\in[M]} \subset \mathcal{K}$, use the fill distance: $h_M = \sup_{x\in\mathcal{K}} d(x, \{\tilde{x}_m\}_{m\in[M]})$ to bound $|L(\bar{\mathbf{f}}^{approx}) - L(\bar{\mathbf{f}}^0)|$

Second-order cone (SOC) tightening [Aubin-Frankowski and Szabó, 2020a]

$$\eta_M \|f\| \le f(\tilde{x}_m)$$

e.g. for $k(x,y) = \psi(x-y)$
$\eta_M := \sqrt{\psi(0) - \psi(h_M)} \propto h_M \ll 1$
Tighten constraint by at most $C\|f\| \cdot h_M$

Semi-positive definite (SDP) relaxation [Rudi et al., 2020]

$$\langle \Phi(\tilde{x}_m), A\Phi(\tilde{x}_m)\rangle_k = f(\tilde{x}_m)$$

with extra variable $A \in S^+(\mathcal{H}_k)$
Relax by at most $C(\|f\| + \mathrm{Tr}(A)) \cdot (h_M)^s$ for s-smooth Sobolev spaces

Deriving SOC constraints through continuity moduli

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

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

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

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

$$\text{“}0 \leq f(x),\ \forall x \in \mathcal{K}\text{“} \Leftarrow \text{“}\omega_m(f,\delta) \leq f(\tilde{x}_m),\ \forall m \in [M]\text{“}$$

Deriving SOC constraints through continuity moduli

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

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

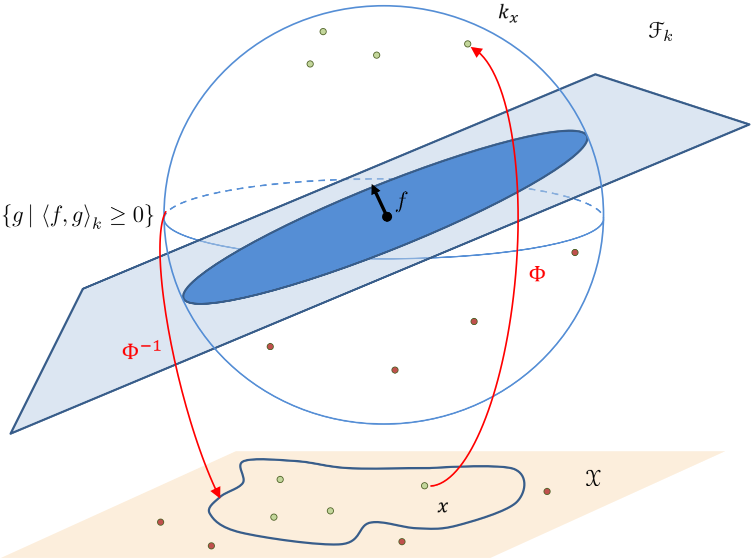
$$\omega_m(f,\delta) := \sup_{\{x \,|\, \|x-\tilde{x}_m\| \leq \delta\}} |f(x) - f(\tilde{x}_m)| \leq \eta_m(\delta)\|f(\cdot)\|_k$$

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

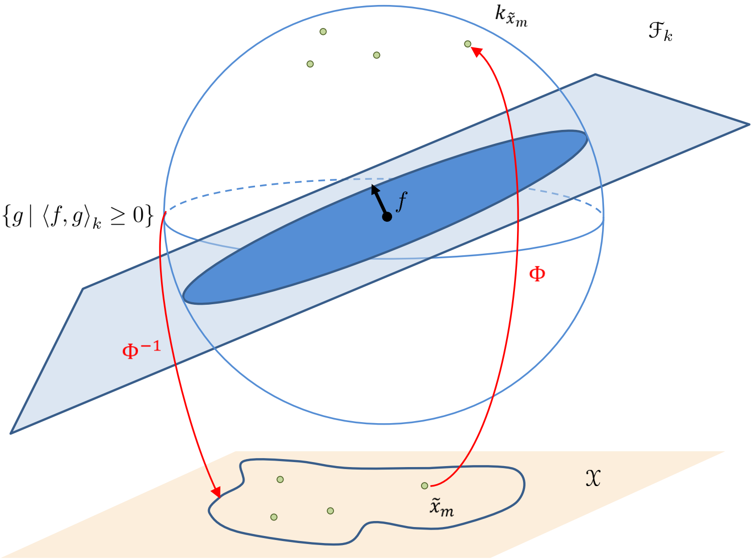
$$\text{“}0 \leq f(x),\ \forall x \in \mathcal{K}\text{“} \Leftarrow \text{“}\omega_m(f,\delta) \leq f(\tilde{x}_m),\ \forall m \in [M]\text{“}$$
$$\Leftarrow \text{“}\eta_m(\delta)\|f(\cdot)\| \leq f(\tilde{x}_m),\ \forall m \in [M]$$

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

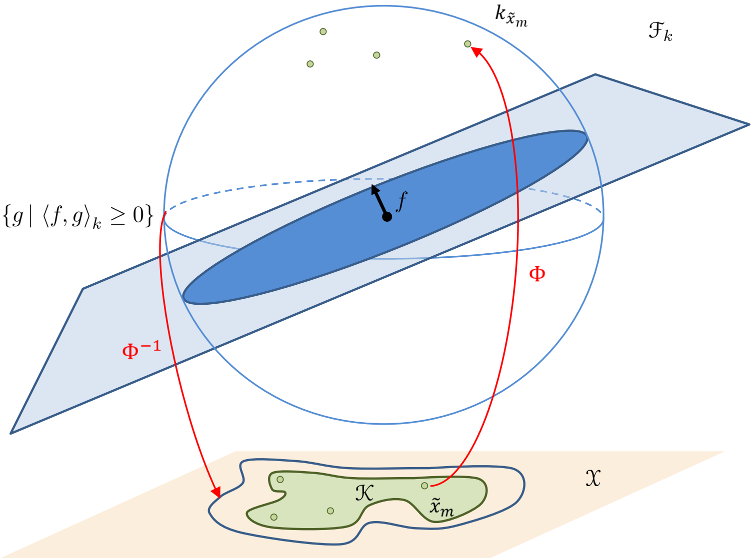
There is also a geometrical interpretation for this choice of η_m.



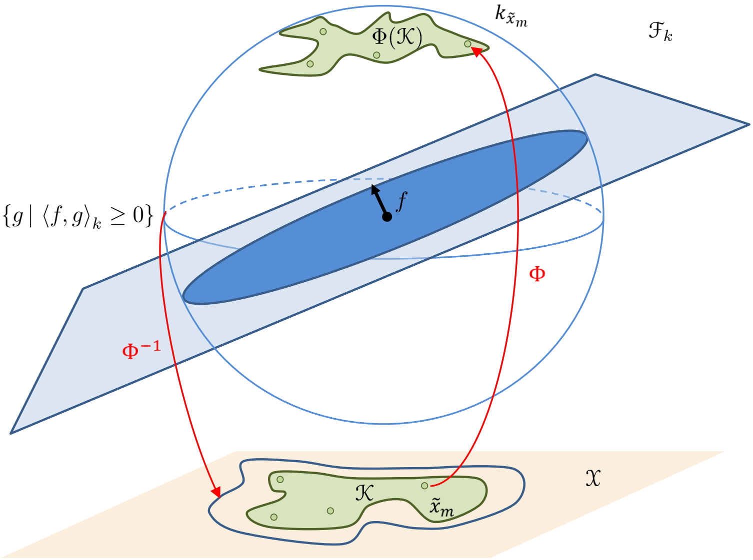
Support Vector Machine (SVM) is about separating red and green points by blue hyperplane.



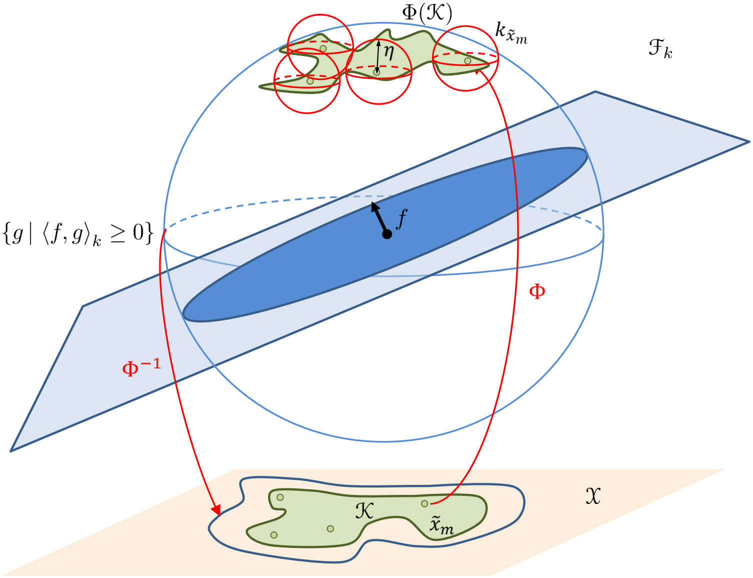
Using the nonlinear embedding $\Phi : x \mapsto D_x k(x, \cdot)$, the idea is the same. With only the green points, it is a one-class SVM [Schölkopf et al., 2001b]



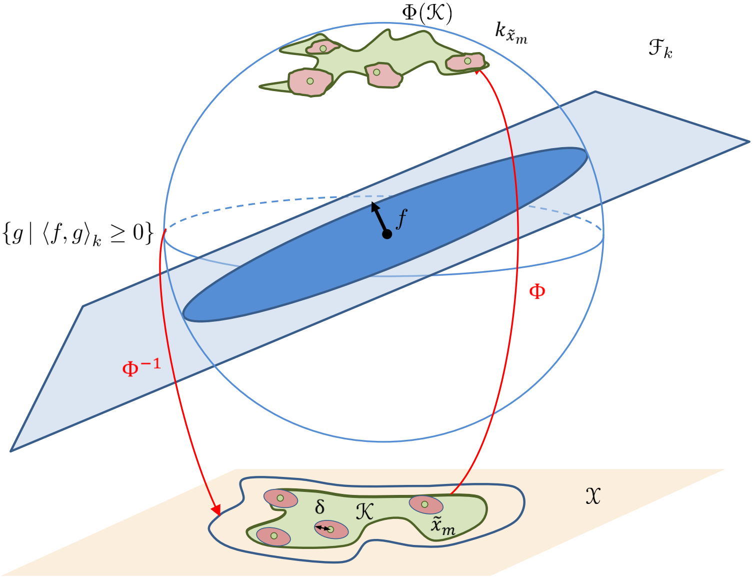
The green points are now samples of a compact set $\mathcal{K}$.



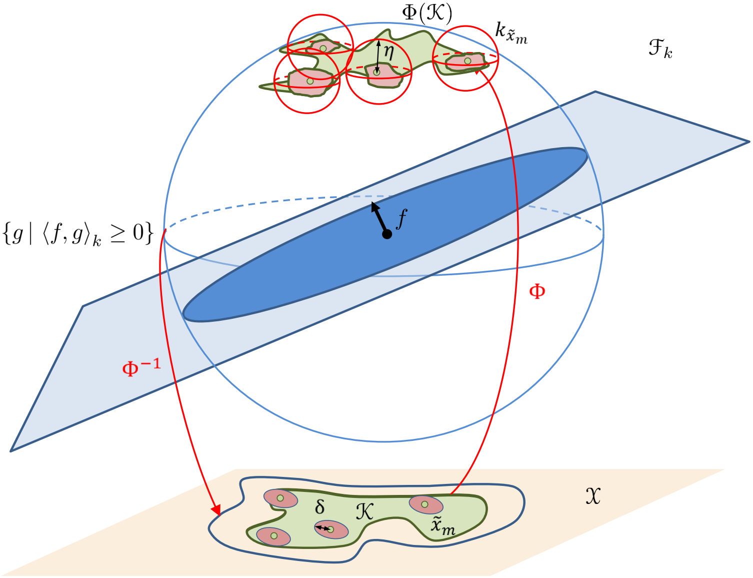
The image $\Phi(\mathcal{K})$ is not convex...



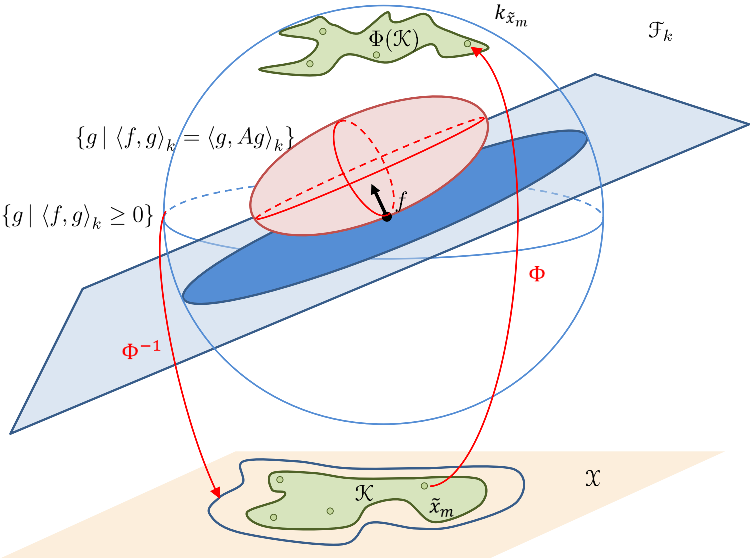
The image $\Phi(\mathcal{K})$ is not convex, can we cover it by balls of radius η?



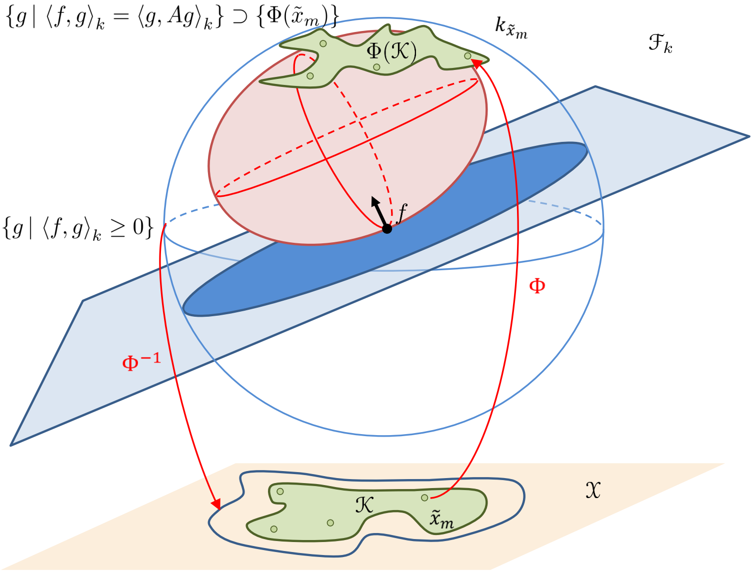
First cover $\mathcal{K} \subset \bigcup\{\tilde{x}_m + \delta\mathbb{B}\}$, and then look at the images $\Phi(\{\tilde{x}_m + \delta\mathbb{B}\})$



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



For SDP relaxation (a.k.a. kernel Sum-Of-Squares), it is rather like inflating an ellipsis. . .



For SDP relaxation (a.k.a. kernel Sum-Of-Squares), it is rather like inflating an ellipsis until it reaches all the points to interpolate

Second-order-cone (SOC) tightening
Ball covering in the RKHS

Semi-positive definite (SDP) relaxation
Kernel Sum-Of-Squares (kSOS)

Protecting the points from all sides, thus "slower" convergence

Leverages smooth interpolation and relaxing, thus "faster" convergence

In both cases, SOC or SDP constraints instead of affine $\Longrightarrow$ extra computational price

Nested constraint sets

Fill distance: $h_M = \sup_{x \in \mathcal{K}} d(x, \{\tilde{x}_m\}_{m \in [M]})$

$$\begin{aligned}
\mathcal{V}_{-\epsilon} &:= \{f \in \mathcal{H}_k \mid f(x) \geq -\epsilon, \forall\, x \in \mathcal{K}\} \\
\mathcal{V}_{SDP} &:= \{f \in \mathcal{H}_k \mid \exists A \in S^+(\mathcal{H}_k),\ f(\tilde{x}_m) = \langle \Phi(\tilde{x}_m), A\Phi(\tilde{x}_m)\rangle_k, \forall\, m \in [M]\}, \\
\mathcal{V}_0 &:= \{f \in \mathcal{H}_k \mid f(x) \geq 0, \forall\, x \in \mathcal{K}\}, \\
\mathcal{V}_{SOC} &:= \{f \in \mathcal{H}_k \mid f(\tilde{x}_m) \geq \eta_M \|f\|_K, \forall\, m \in [M]\}, \\
\mathcal{V}_{\epsilon} &:= \{f \in \mathcal{H}_k \mid f(x) \geq \epsilon, \forall\, x \in \mathcal{K}\}.
\end{aligned}$$

Nested constraint sets

Fill distance: $h_M = \sup_{x \in \mathcal{K}} d(x, \{\tilde{x}_m\}_{m \in [M]})$

$$\begin{aligned}
\mathcal{V}_{-\epsilon} &:= \{f \in \mathcal{H}_k \,|\, f(x) \geq -\epsilon, \forall\, x \in \mathcal{K}\} \\
\mathcal{V}_{SDP} &:= \{f \in \mathcal{H}_k \,|\exists A \in S^+(\mathcal{H}_k),\, f(\tilde{x}_m) = \langle \Phi(\tilde{x}_m), A\Phi(\tilde{x}_m)\rangle_k, \forall\, m \in [M]\}, \\
\mathcal{V}_0 &:= \{f \in \mathcal{H}_k \,|\, f(x) \geq 0, \forall\, x \in \mathcal{K}\}, \\
\mathcal{V}_{SOC} &:= \{f \in \mathcal{H}_k \,|\, f(\tilde{x}_m) \geq \eta_M \|f\|_K, \forall\, m \in [M]\}, \\
\mathcal{V}_{\epsilon} &:= \{f \in \mathcal{H}_k \,|\, f(x) \geq \epsilon, \forall\, x \in \mathcal{K}\}.
\end{aligned}$$

Proposition (Informal nestedness)

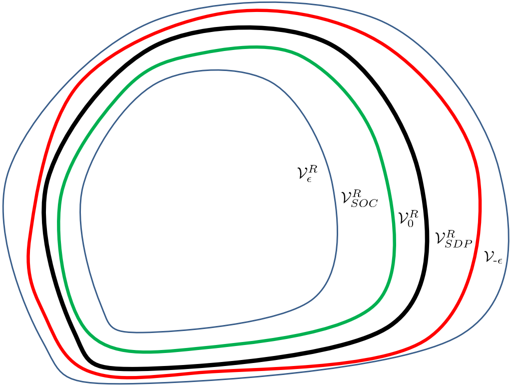
Under some assumptions on the kernel (e.g. Sobolev), there exists explicit constants C_{SOC} *and* C_{SDP}*, such that for* $h_M = \sup_{x \in \mathcal{K}} d(x, \{\tilde{x}_m\}_{m \in [M]})$ *and any* $R \geq 0$

$$\epsilon \geq C_{SOC} \cdot R \cdot h_M \implies (\mathcal{V}_\epsilon \cap R\mathbb{B}_k) \subset \mathcal{V}_{SOC} \subset \mathcal{V}_0$$

$$\epsilon \geq C_{SDP} \cdot R \cdot (h_M)^s \implies (R\mathbb{B}_k \cap \mathcal{V}_0) \subset (R\mathbb{B}_k \cap \mathcal{V}_{SDP}) \subset \mathcal{V}_{-\epsilon}$$

If $\mathcal{L}$ *is* β*-Lipschitz, then* $|\mathcal{L}(\bar{f}^0) - \mathcal{L}(\bar{f}^{SOC})| \leq \beta C_{SOC} \cdot R \cdot h_M$*. If* $\bar{f}^0$ *has a quadratic expression, then* $|\mathcal{L}(\bar{f}^0) - \mathcal{L}(\bar{f}^{SDP})| \leq \beta C_{SDP} \cdot R \cdot (h_M)^s$

Nested constraint sets - decreasing optima sequence



$$\mathcal{V}_{-\epsilon} := \{f \in \mathcal{H}_k \,|\, f(x) \geq -\epsilon, \forall\, x \in \mathcal{K}\}$$

$$\mathcal{V}_{SDP} := \{f \in \mathcal{H}_k \,|\exists A \in S^+(\mathcal{H}_k), \; f(\tilde{x}_m) = \langle \Phi(\tilde{x}_m), A\Phi(\tilde{x}_m)\rangle_k, \forall\, m \in [M]\},$$

$$\mathcal{V}_0 := \{f \in \mathcal{H}_k \,|\, f(x) \geq 0, \forall\, x \in \mathcal{K}\},$$

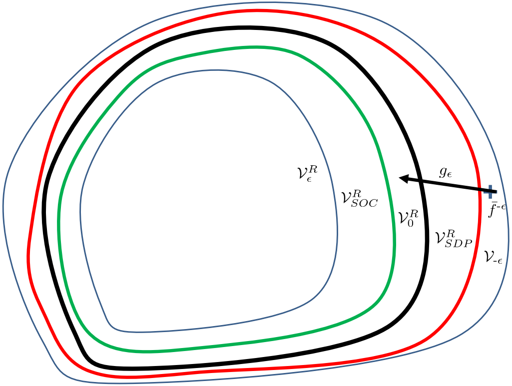
$$\mathcal{V}_{SOC} := \{f \in \mathcal{H}_k \,|\, f(\tilde{x}_m) \geq \eta_M \|f\|_K, \forall\, m \in [M]\},$$

$$\mathcal{V}_\epsilon := \{f \in \mathcal{H}_k \,|\, f(x) \geq \epsilon, \forall\, x \in \mathcal{K}\}.$$

For $R \geq \|\bar{f}^0\|_k$, we have

$$\mathcal{L}(\bar{f}^{-\epsilon}) \leq \mathcal{L}(\bar{f}^{SDP}_R) \leq \mathcal{L}(\bar{f}^0) \leq \mathcal{L}(\bar{f}^{SOC}) \leq \mathcal{L}(\bar{f}^\epsilon)$$

Nested constraint sets - decreasing optima sequence



$$\mathcal{L}(\bar{f}^{-\epsilon}) \le \mathcal{L}(\bar{f}_R^{SDP}) \le \mathcal{L}(\bar{f}^{0}) \le \mathcal{L}(\bar{f}^{SOC}) \le \mathcal{L}(\bar{f}^{\epsilon})$$

Idea: find a $g_\epsilon \in \mathcal{H}_k$ such that $\|g_\epsilon\|_k \le \omega(\epsilon)$ where $\omega : \mathbb{R}_+ \to \mathbb{R}_+ \nearrow$, and such that $\bar{f}^{-\epsilon} + g_\epsilon \in \mathcal{V}_0$, thus under some β-Lipschitz assumption on $\mathcal{L}$,

$$\begin{aligned}\mathcal{L}(\bar{f}^{-\epsilon}) &\le \mathcal{L}(\bar{f}_R^{SDP}) \\ &\le \mathcal{L}(\bar{f}^{0}) \\ &\le \mathcal{L}(\bar{f}^{-\epsilon} + g_\epsilon) \\ &\le \mathcal{L}(\bar{f}^{-\epsilon}) + \beta\omega(\epsilon).\end{aligned}$$

SOC: $\epsilon \ge C_{SOC} \cdot R \cdot h_M$

SDP/kSoS: $\epsilon \approx C_{SDP} \cdot R \cdot (h_M)^s$

Example 1: solving LQ control with state constraints through KRR

Original control problem

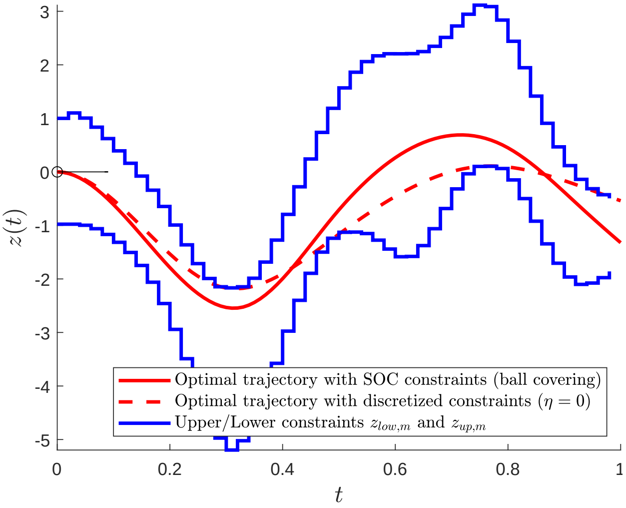
$$\min_{z(\cdot)\in W^{2,2},u(\cdot)\in L^2} \quad \int_0^1 |u(t)|^2 \mathrm{d}t$$

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

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

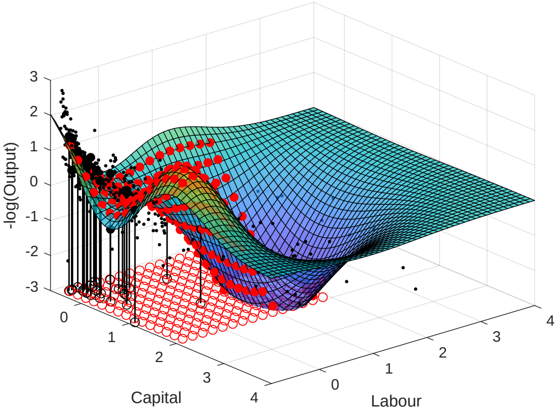
$$\ddot{z}(t) = -\dot{z}(t) + u(t), \forall t \in [0,1],$$

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

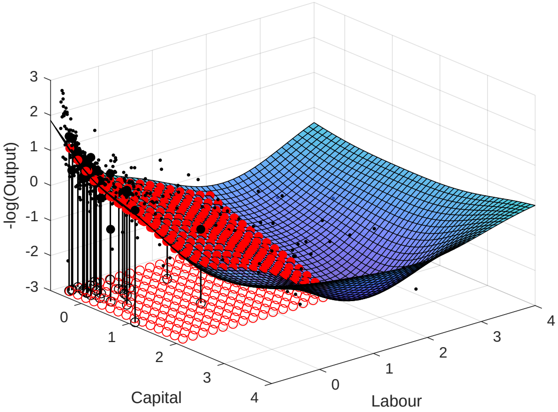


Example 2: Estimation of monotone/convex production functions

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



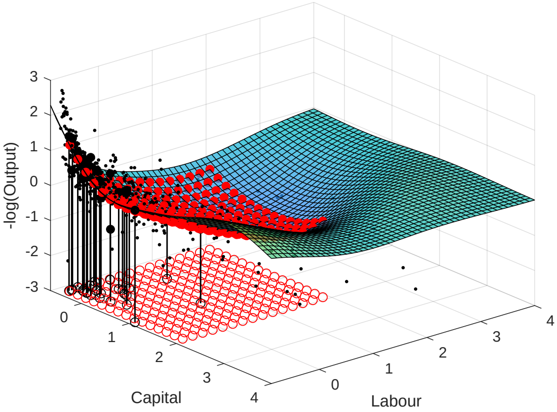
(a) NoCons



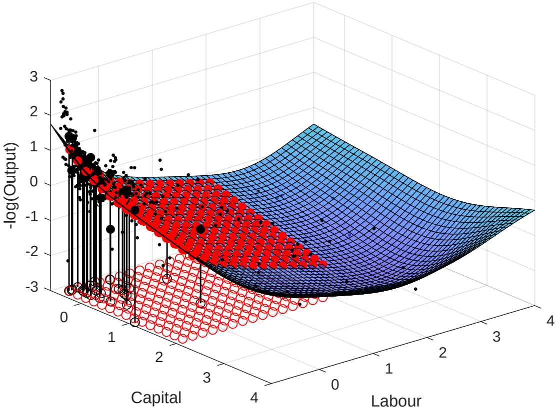
(b) SOC Monot.

Example 2: Estimation of monotone/convex 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.

Example 2: Estimation of monotone/convex production functions

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

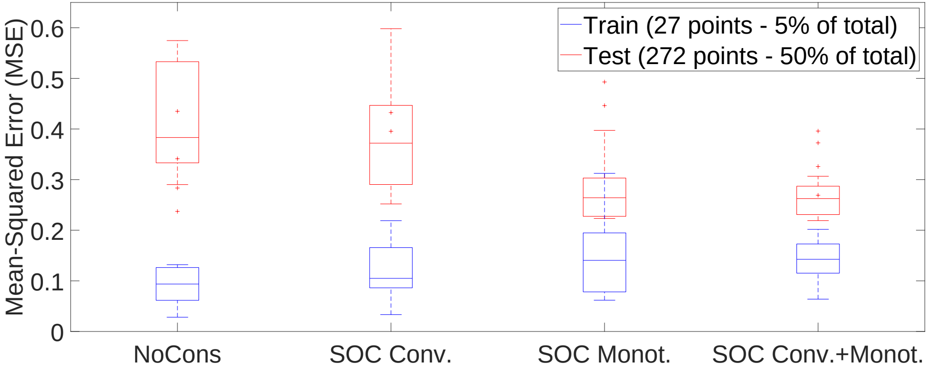


Figure: MSE as a function of incorporating shape constraints with the proposed SOC technique. NoCons: no constraint. SOC Monot.: two monotonicity constraints. SOC Conv.: one convexity constraint. SOC Conv.+Monot.: one convexity and two monotonicity constraints.

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 inequality/SDP equality constraints**.“

“**Bounding the constraint perturbation** made by discretizing allows to easily **assess rates of convergence**.“

To go beyond

- Handle state constraint in LQ control through the LQ kernel

  $\hookrightarrow$ PCAF, *Linearly-constrained Linear Quadratic Regulator from the viewpoint of kernel methods*, SIAM Journal on Control and Optimization, 2021

- Tackle SDP and derivative constraints with SOC constraints

  $\hookrightarrow$ PCAF and Zoltán Szabó, *Handling Hard Affine Shape Constraints in RKHSs*, under review, 2021

- Use kernels for learning vector fields and nonlinear systems

  $\hookrightarrow$ Coming in soon!

More to be found on `https://pcaubin.github.io/`

To go beyond

- Handle state constraint in LQ control through the LQ kernel

  ↪ PCAF, *Linearly-constrained Linear Quadratic Regulator from the viewpoint of kernel methods*, SIAM Journal on Control and Optimization, 2021

- Tackle SDP and derivative constraints with SOC constraints

  ↪ PCAF and Zoltán Szabó, *Handling Hard Affine Shape Constraints in RKHSs*, under review, 2021

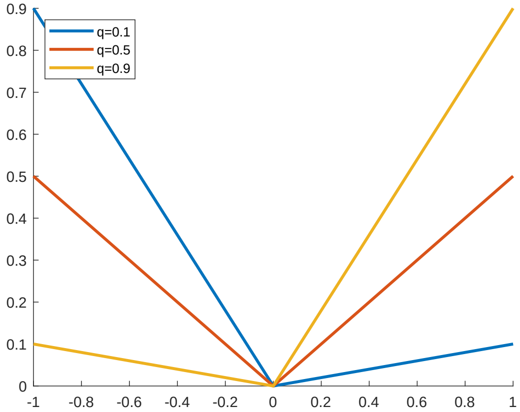
- Use kernels for learning vector fields and nonlinear systems

  ↪ Coming in soon!

More to be found on `https://pcaubin.github.io/`

Thank you for your attention!

Example 3: 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]).

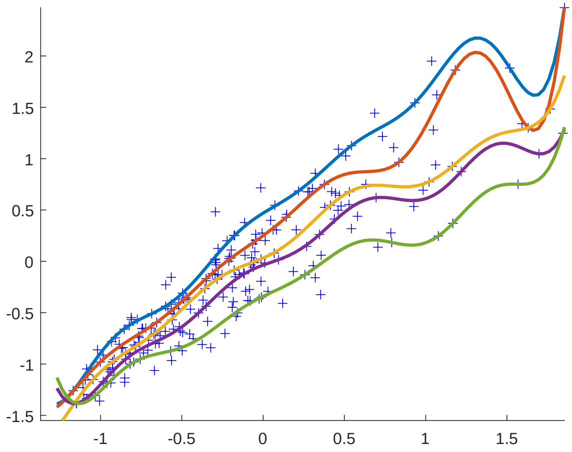
Known fact: quantile functions can cross when estimated independently.

Joint quantile regression with non-crossing constraints

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

Pairing non-crossing quantiles with other shape constraints

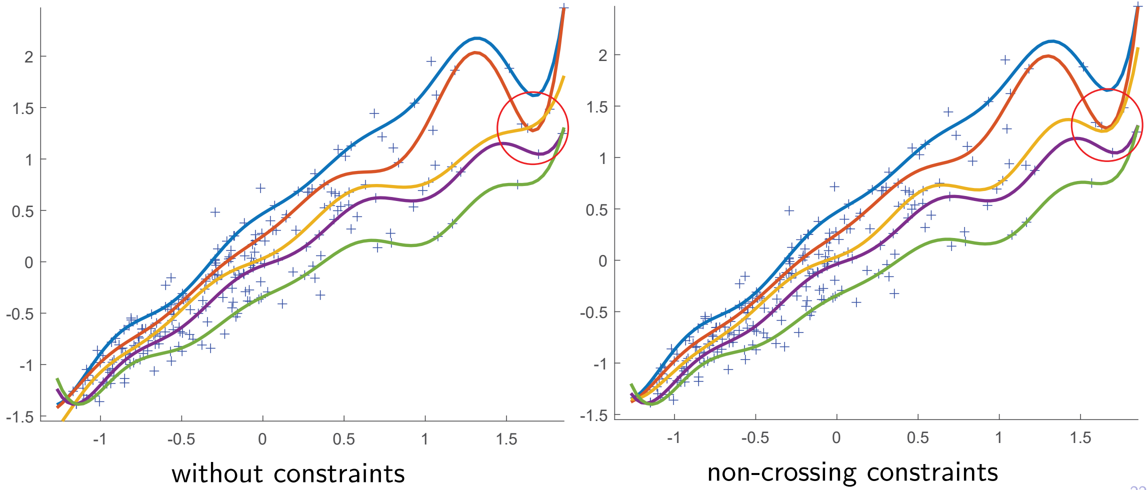
Engel's law (1857): "As income rises, the proportion of income spent on food falls, but absolute expenditure on food rises."



without constraints

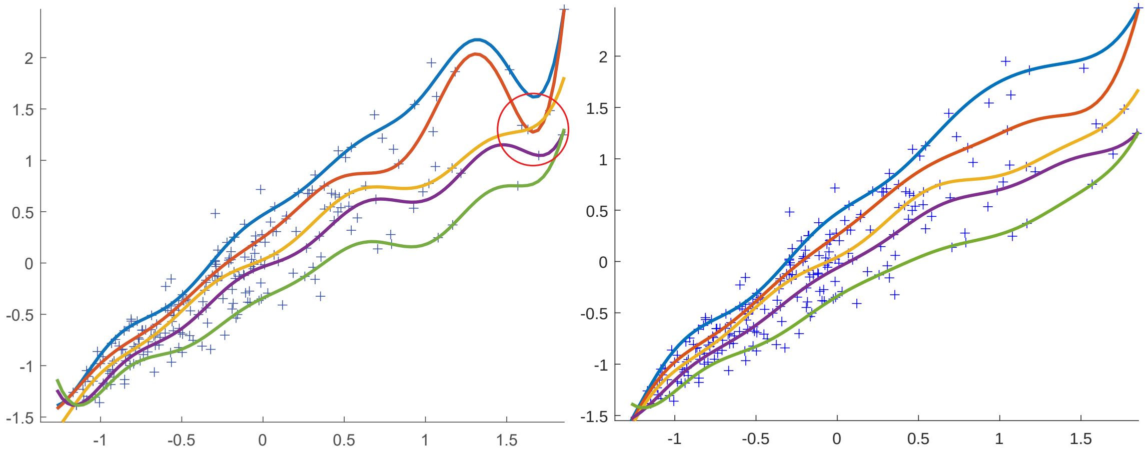
Pairing non-crossing quantiles with other shape constraints

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



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."

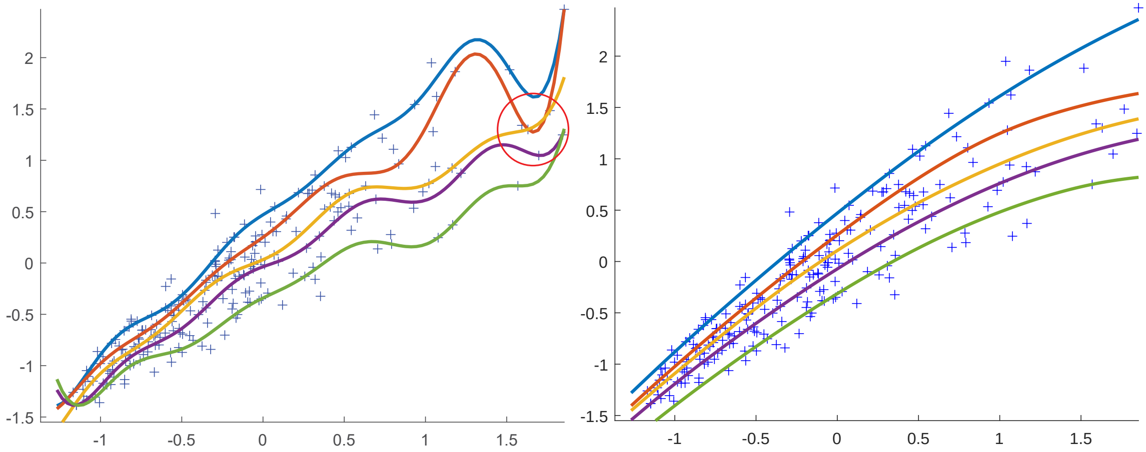


without constraints

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

Time-varying state-constrained LQ optimal control

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

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

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

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

- state $\mathbf{x}(t) \in \mathbb{R}^Q$, control $\mathbf{u}(t) \in \mathbb{R}^P$,
- reference time $t_{ref} \in [t_0, T]$, set of constraint times $\mathcal{T}_c \subset [t_0, T]$,
- $\mathbf{x}(\cdot) : [t_0, T] \to \mathbb{R}^Q$ absolutely continuous, $\mathbf{R}(\cdot)^{1/2}\mathbf{u}(\cdot) \in L^2([t_0, T])$

Time-varying state-constrained LQ optimal control

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

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

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

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

- state $\mathbf{x}(t) \in \mathbb{R}^Q$, control $\mathbf{u}(t) \in \mathbb{R}^P$,
- reference time $t_{ref} \in [t_0, T]$, set of constraint times $\mathcal{T}_c \subset [t_0, T]$,
- $\mathbf{x}(\cdot) : [t_0, T] \to \mathbb{R}^Q$ absolutely continuous, $\mathbf{R}(\cdot)^{1/2}\mathbf{u}(\cdot) \in L^2([t_0, T])$

$$\mathcal{S} := \{\mathbf{x} : [t_0, T] \to \mathbb{R}^Q \,|\, \exists\, \mathbf{R}(\cdot)^{1/2}\mathbf{u}(\cdot) \in L^2(t_0, T) \text{ s.t. } \mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \text{ a.e. } \}$$

Given $\mathbf{x}(\cdot) \in \mathcal{S}$, for the pseudoinverse $\mathbf{B}(t)^\ominus$ for $\|\cdot\|_{\mathbf{R}}$, set $\mathbf{u}(t) \overset{a.e.}{=} \mathbf{B}(t)^\ominus[\mathbf{x}'(t) - \mathbf{A}(t)\mathbf{x}(t)]$.
$(\mathcal{S}, \langle\cdot,\cdot\rangle_{\mathcal{S}})$ is a (vector-valued) RKHS with an explicit kernel [Aubin-Frankowski, 2021]!

Optimal control: constrained pendulum - definition

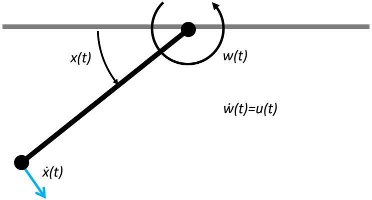
Constrained pendulum when controlling the third derivative of the angle

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

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

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

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



Optimal control: constrained pendulum - definition

Constrained pendulum when controlling the third derivative of the angle

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

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

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

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

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

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

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

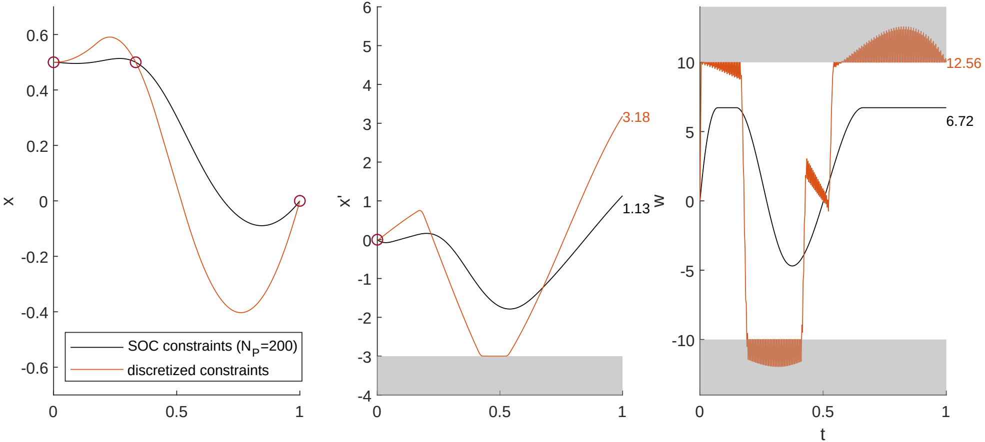
Most of computational cost is related to the “controllability Gramians“
$K_1(s, t) = \int_0^{\min(s,t)} \mathbf{e}^{(s-\tau)\mathbf{A}}\mathbf{B}\mathbf{B}^\top\mathbf{e}^{(t-\tau)\mathbf{A}^\top}\mathrm{d}\tau$ which we have to approximate.

Optimal control: 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 $w(\cdot)$

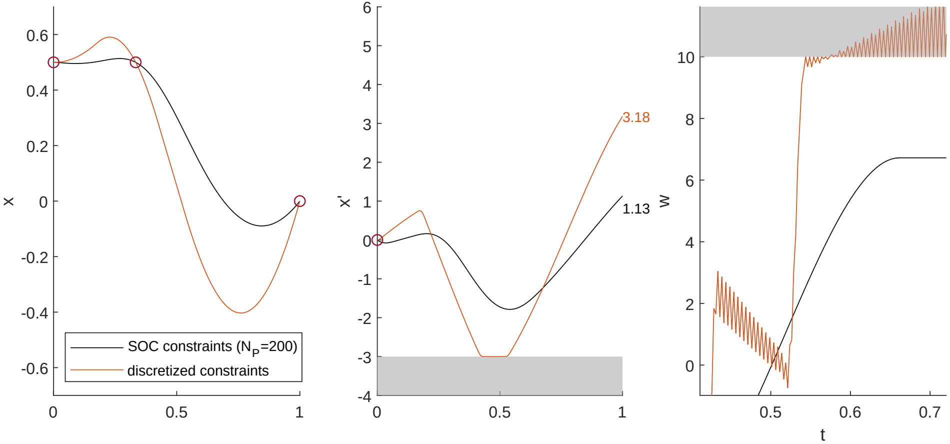


Optimal control: constrained pendulum - illustration

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 $w(\cdot)$

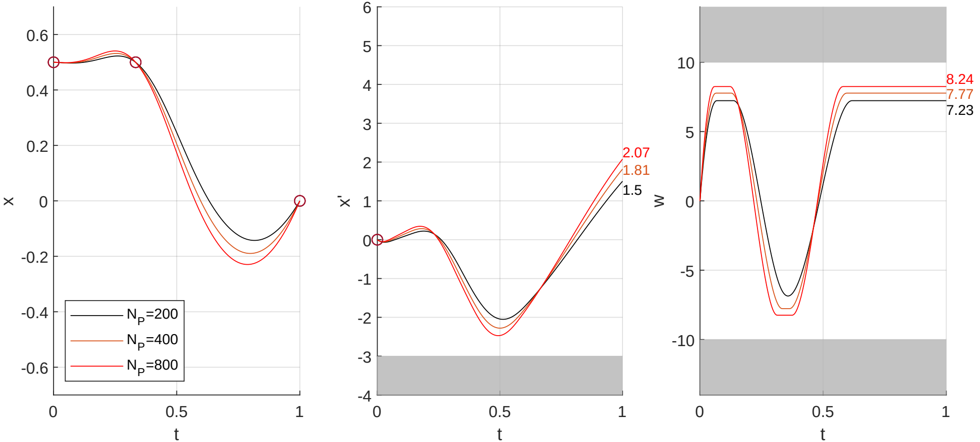


Optimal control: constrained pendulum - illustration

Optimal solutions of the constrained pendulum “path-planning“ problem.

Red circles: equality constraints. Grayed areas: constraints over $[0, T]$.

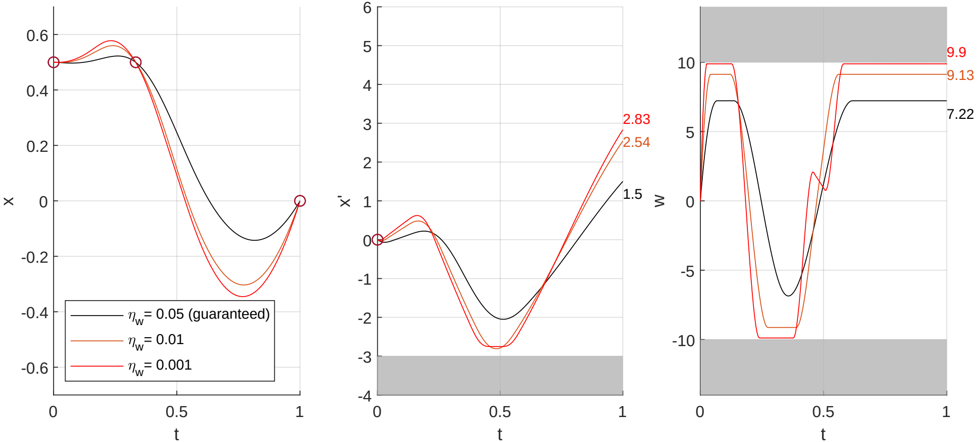
Angle $x(\cdot)$ Velocity $\dot{x}(\cdot)$ Couple $w(\cdot)$



Optimal control: 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 $w(\cdot)$



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. (2021).
Linearly constrained linear quadratic regulator from the viewpoint of kernel methods.
SIAM Journal on Control and Optimization, 59(4):2693–2716.

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 II

Aubin-Frankowski, P.-C. and Szabó, Z. (2020b).
Hard shape-constrained kernel machines.
In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 384–395.
(http://arxiv.org/abs/2005.12636).

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.

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

Marteau-Ferey, U., Bach, F., and Rudi, A. (2020a).
Non-parametric models for non-negative functions.
In *Advances in Neural Information Processing Systems*, volume 33, pages 12816–12826. Curran Associates, Inc.

References III

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

Muzellec, B., Bach, F., and Rudi, A. (2021).
Learning psd-valued functions using kernel sums-of-squares.

Rudi, A., Marteau-Ferey, U., and Bach, F. (2020).
Finding global minima via kernel approximations.

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 IV

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

Vacher, A., Muzellec, B., Rudi, A., Bach, F. R., and Vialard, F. (2021).
A dimension-free computational upper-bound for smooth optimal transport estimation.
In *Conference on Learning Theory, COLT 2021*, volume 134 of *Proceedings of Machine Learning Research*, pages 4143–4173. PMLR.