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

\[ \rightarrow \text{ provides feasible trajectories in path-planning} \]

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

\[ \rightarrow \] 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) $g$:**
  \[
  \max_{c \in \mathbb{R}} c \quad (= \min_{x \in X} g(x)) \quad \text{subject to} \quad c \geq g(x), \forall x \in X
  \]

- **Density estimation with relative entropy:**
  \[
  \min_{f \in C(X, \mathbb{R}), \int_X f(x) dx = 1} \quad -\int_X \log(f(x)) d\mu(x) \quad (= \text{KL} (\mu, \mu_f) + \text{cst})
  \]
  \[
  \text{subject to} \quad f(x) \geq 0, \forall x \in X
  \]

- **Optimal transport in its dual formulation:**
  \[
  \max_{u, v \in C(X, \mathbb{R})} \quad \int_X u(x) d\mu(x) + \int_X v(y) d\nu(y) \quad (= \text{OT}_c (\mu, \nu))
  \]
  \[
  \text{subject to} \quad u(x) + v(y) \leq c(x, y), \forall x, y \in X \times X
  \]

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 Szabó, 2020b, 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

\[ \tilde{f} \in \arg\min_{f \in \mathcal{H}} \mathcal{L}(f) \text{ s.t. } "0 \leq f(x), \forall x \in \mathcal{K}" , \mathcal{K} \subset \mathbb{R}^d \text{ non-finite (compact)} \]

### Relaxing
- Discretize constraint at “virtual“ samples \( \{\tilde{x}_m\}_{m \in [M]} \subset \mathcal{K} \),
  - no guarantees out-of-samples [Agrell, 2019, Takeuchi et al., 2006]
- Add constraint-inducing penalty, \( R_{\text{cons}}(f) = -\lambda \int_{\mathcal{K}} \min(0, f(x)) \, dx \)
  - 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
  - **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 \( A \)
  - **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]
  - **generic**: no violation, second-order cone constraints, but extra tightening
1. Introduction to constrained problems
2. Kernel methods for problem approximation
3. Deriving bounds on the optimization error
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]

\[ \exists k : \mathcal{X} \times \mathcal{X} \to \mathbb{R} \text{ s.t. } k_x(\cdot) = k(x, \cdot) \in \mathcal{H}_k \text{ and } f(x) = \langle f(\cdot), k_x(\cdot) \rangle_{\mathcal{H}_k} \text{ for all } x \in \mathcal{X} \text{ and } f \in \mathcal{H}_k \text{ (reproducing property)} \]

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

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

\(k\) is s.t. \(G = [k(x_i, x_j)]_{i,j=1}^n \succeq 0\) 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 \( L : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{\infty\} \), strictly increasing \( \Omega : \mathbb{R}_+ \rightarrow \mathbb{R} \), and

\[
\tilde{f} \in \arg\min_{f \in \mathcal{H}_k} L \left( (f(x_n))_{n \in [N]} \right) + \Omega \left( \|f\|_k \right)
\]

Then \( \exists (a_n)_{n \in [N]} \in \mathbb{R}^N \) s.t.

\[
\tilde{f}(\cdot) = \sum_{n \in [N]} a_n k(\cdot, x_n)
\]

\( \hookleftarrow \) Optimal solutions lie in a finite dimensional subspace of \( \mathcal{H}_k \).

**Finite number of evaluations \( \implies \) finite number of coefficients**

Kernel trick

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

\( \hookleftarrow \) 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 $A \in S^+(\mathcal{H}_k)$,

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

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

(kernel SoS) $[a_{ij}]_{i,j} = \sum_i u_i u_i^\top$ (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 \( f : \mathbb{R}^d \rightarrow \mathbb{R}^p \) constrained over compact set \( K \) to belong to a set \( F(x) \)

Optimization over \( F = \mathcal{C}(\mathcal{X}, \mathbb{R}) \) or \( L(\mu) \)

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

s.t.
\[
f(x) \in F(x), \ \forall x \in K
\]
What I am looking for: an approximation framework

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

Optimization over $F \equiv \mathcal{C}(\mathcal{X}, \mathbb{R})$ or $L(\mu)$

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

s.t.

$$c_i(x)^T f(x) + d_i(x) \geq 0, \forall i \in [l], \forall x \in \mathcal{K}$$

Infinite number of affine constraints!
What I am looking for: an approximation framework

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

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

s.t.

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

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

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

s.t.

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

Infinite number of affine constraints!
What I am looking for: an approximation framework

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

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

$$\begin{align*}
\min_{f = [f_1; \ldots; f_P] \in \mathcal{F}^P} & \int l(x, f(x))d\mu(x) \\
\text{s.t.} & \quad c_i(x)^T f(x) + d_i(x) \geq 0, \forall i \in [I], \forall x \in \mathcal{K}
\end{align*}$$

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

$$\begin{align*}
\min_{f \in \mathcal{H}_k^P} & \sum_{n \in [N]} l(x_n, f(x_n)) + \lambda \|f\|_2^2 \\
\text{s.t.} & \quad c_i(x_m)^T f(x_m) + d_i(x_m) \geq ? , \forall i \in [I], \forall m \in [M]
\end{align*}$$

Infinite number of affine constraints!

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 $L : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{\infty\}$, a regularizer $R : \mathbb{R}_+ \rightarrow \mathbb{R}$, a RKHS $\mathcal{H}_k$ of smooth functions from $\mathcal{X}$ to $\mathbb{R}$ and a compact set $\mathcal{K} \subset \mathcal{X}$.

$$\overline{f}^0 \in \underset{f \in \mathcal{H}_k}{\operatorname{arg min}} \; \mathcal{L}(f) = L\left((f(x_n))_{n \in [N]}\right) + R\left(\|f\|_{\mathcal{H}_k}\right)$$

s.t. $0 \leq f(x), \; \forall x \in \mathcal{K}$.

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 $|L(\overline{f}^\text{approx}) - L(\overline{f}^0)|$
Statement of simpler problem

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

\[
\bar{f}^0 \in \arg\min_{f \in \mathcal{H}_k} \mathcal{L}(f) = L\left((f(x_n))_{n \in [N]}\right) + R\left(\|f\|_{\mathcal{H}_k}\right)
\]

s.t. \(0 \leq f(x), \quad \forall x \in \mathcal{K}\).

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 \(|L(\bar{f}^{\text{approx}}) - L(\bar{f}^0)|\)

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

\[
\eta_M \|f\| \leq f(\tilde{x}_m)
\]

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 \(C(\|f\| + \text{Tr}(A)) \cdot (h_M)^s\) for \(s\)-smooth Sobolev spaces
Deriving SOC constraints through continuity moduli

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

$$|f(x) - f(\tilde{x}_m)| = |\langle f(\cdot), k(x, \cdot) - k(\tilde{x}_m, \cdot) \rangle_k|$$

$$\leq \|f(\cdot)\|_k \sup_{\{x \mid \|x - \tilde{x}_m\| \leq \delta\}} \|k(x, \cdot) - k(\tilde{x}_m, \cdot)\|_k$$

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

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

"$0 \leq f(x), \forall x \in \mathcal{K}$" $\iff$ "$\omega_m(f, \delta) \leq f(\tilde{x}_m), \forall m \in [M]$"
Deriving SOC constraints through continuity moduli

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

$$|f(x) - f(\tilde{x}_m)| = |\langle f(\cdot), k(x, \cdot) - k(\tilde{x}_m, \cdot) \rangle_k|$$

$$\leq \|f(\cdot)\|_k \sup_{\{x \mid \|x - \tilde{x}_m\| \leq \delta\}} \|k(x, \cdot) - k(\tilde{x}_m, \cdot)\|_k$$

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

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

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

$\iff$ "$\eta_m(\delta) \|f(\cdot)\| \leq f(\tilde{x}_m)$, $\forall m \in [M]$"

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

There is also a geometrical interpretation for this choice of $\eta_m$. 

12/28
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 $\eta$?
First cover $\mathcal{K} \subset \bigcup \{\bar{x}_m + \delta B\}$, and then look at the images $\Phi(\{\bar{x}_m + \delta B\})$. 

$$\{g \mid \langle f, g \rangle_k \geq 0\}$$
Cover the $\Phi(\{\tilde{x}_m + \delta B\})$ 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

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

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

Leverages smooth interpolation and relaxing, thus “faster“ convergence

In both cases, SOC or SDP constraints instead of affine $\Rightarrow$ extra computational price
Nested constraint sets

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

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

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

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

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

\[
\mathcal{V}_{\epsilon} := \{ f \in \mathcal{H}_k \mid f(x) \geq \epsilon, \forall x \in \mathcal{K} \}.
\]
Nested constraint sets

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

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

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 \( \beta \)-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

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

\[ \mathcal{V}_{SDP} := \{ f \in \mathcal{H}_k \mid \exists A \in \mathcal{S}^+(\mathcal{H}_k), \]

\[ f(\tilde{x}_m) = \langle \Phi(\tilde{x}_m), A\Phi(\tilde{x}_m) \rangle_k, \forall m \in [M] \}, \]

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

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

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

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

\[ \mathcal{L}(\bar{f}^-) \leq \mathcal{L}(\bar{f}_{SDP}^R) \leq \mathcal{L}(\bar{f}^0) \leq \mathcal{L}(\bar{f}_{SOC}) \leq \mathcal{L}(\bar{f}^-) \]

**Idea:** find a \( g_\epsilon \in \mathcal{H}_k \) such that \( \|g_\epsilon\|_k \leq \omega(\epsilon) \) where \( \omega : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \uparrow \), and such that \( \bar{f}^- + g_\epsilon \in \mathcal{V}_0 \), thus under some \( \beta \)-Lipschitz assumption on \( \mathcal{L} \),

\[
\mathcal{L}(\bar{f}^-) \leq \mathcal{L}(\bar{f}_{SDP}^R) \\
\leq \mathcal{L}(\bar{f}^0) \\
\leq \mathcal{L}(\bar{f}^- + g_\epsilon) \\
\leq \mathcal{L}(\bar{f}^-) + \beta \omega(\epsilon).
\]

**SOC:** \( \epsilon \geq C_{SOC} \cdot R \cdot h_M \)

**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

\[
\begin{align*}
\min_{z(\cdot) \in W^{2,2}, u(\cdot) \in L^2} & \quad \int_0^1 |u(t)|^2 dt \\
\text{s.t.} & \quad z(0) = 0, \quad \dot{z}(0) = 0, \\
& \quad \ddot{z}(t) = -\dot{z}(t) + u(t), \quad \forall t \in [0, 1], \\
& \quad z(t) \in [z_{\text{low}}(t), z_{\text{up}}(t)], \quad \forall t \in [0, 1].
\end{align*}
\]
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.

[Graphs showing production functions for different constraints]
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.
“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
  - 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/
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/](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

$$\min_{(f_q)_{q \in [Q]}} \mathcal{L}(f_1, \ldots, f_Q) = \frac{1}{N} \sum_{q \in [Q]} \sum_{n \in [N]} l_\tau(y_n - f_q(x_n)) + \lambda f \sum_{q \in [Q]} \|f_q\|^2_k$$
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.”
Engel’s law (1857): “As income rises, the proportion of income spent on food falls, but absolute expenditure on food rises.”
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 + concave
Time-varying state-constrained LQ optimal control

$$\min_{x(\cdot), u(\cdot)} \chi x_0(x(t_0)) + g(x(T))$$

$$+ x(t_{\text{ref}})^\top J_{\text{ref}} x(t_{\text{ref}}) + \int_{t_0}^T \left[ x(t)^\top Q(t) x(t) + u(t)^\top R(t) u(t) \right] dt$$

s.t.  

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

\[
c_i(t)^\top x(t) \leq d_i(t), \forall t \in T_c, \forall i \in [I] = [1, I],
\]

- state \(x(t) \in \mathbb{R}^Q\), control \(u(t) \in \mathbb{R}^P\),
- reference time \(t_{\text{ref}} \in [t_0, T]\), set of constraint times \(T_c \subset [t_0, T]\),
- \(x(\cdot) : [t_0, T] \to \mathbb{R}^Q\) absolutely continuous, \(R(\cdot)^{1/2} u(\cdot) \in L^2([t_0, T])\)
Time-varying state-constrained LQ optimal control

\[
\begin{align*}
\min_{x(\cdot), u(\cdot)} & \quad x_0(x(t_0)) + g(x(T)) + x(t_{\text{ref}})^T J_{\text{ref}} x(t_{\text{ref}}) + \int_{t_0}^T \left[ x(t)^T Q(t) x(t) + u(t)^T R(t) u(t) \right] dt \\
\text{s.t.} & \quad x'(t) = A(t) x(t) + B(t) u(t), \quad \text{a.e. in } [t_0, T], \\
& \quad c_i(t)^T x(t) \leq d_i(t), \quad \forall t \in T, \forall i \in [I] = [1, I], \\
\end{align*}
\]

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

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

Given \( x(\cdot) \in S \), for the pseudoinverse \( B(t)^\ominus \) for \( \| \cdot \|_R \), set \( u(t) \overset{a.e.}{=} B(t)^\ominus [x'(t) - A(t) x(t)] \).

\((S, \langle \cdot, \cdot \rangle_S)\) 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

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

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

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

\[
\dot{x}(t) \in [-3, +\infty[, \quad w(t) \in [-10, 10], \forall t \in [0, T]
\]
Optimal control: constrained pendulum - definition

Constrained pendulum when controlling the third derivative of the angle

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

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

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

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

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

\[
\eta_{\dot{x}}\|x(\cdot)\|_K - \dot{x}(t_m) \leq 3,
\]

\[
\eta_w\|x(\cdot)\|_K + w(t_m) \leq 10,
\]

\[
\eta_w\|x(\cdot)\|_K - w(t_m) \leq 10
\]

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

Most of computational cost is related to the “controllability Gramians“

\[K_1(s, t) = \int_0^{\min(s,t)} e^{(s-\tau)A}BB^T e^{(t-\tau)A^T} d\tau\]

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 \(w(\cdot)\)

Figure: Comparison of SOC constraints (\(\eta_w\)) vs discretized constraints (\(\eta_w=0\)) for \(N_P=200\).
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 \(w(\cdot)\)

Figure: Comparison of SOC constraints (\(N_P=200\)) vs discretized constraints (\(\eta_w=0\)) for \(N_P=200\) - Chattering phenomenon like for traffic cameras!
Optimal control: constrained pendulum - illustration


Angle $x(\cdot)$

Velocity $\dot{x}(\cdot)$

Couple $w(\cdot)$

Figure: Comparison of SOC constraints for varying $N_P$ and guaranteed $\eta_w$. 
Optimal control: constrained pendulum - illustration

Optimal solutions of the constrained pendulum “path-planning“ problem. Red circles: \textcolor{red}{equality constraints}. Grayed areas: \textcolor{gray}{constraints over \([0, T]\).}

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

\begin{align*}
\eta_w & = 0.05 \text{ (guaranteed)} \\
\eta_w & = 0.01 \\
\eta_w & = 0.001
\end{align*}
References I

Gaussian processes with linear operator inequality constraints.

Theory of reproducing kernels.

Linearly constrained linear quadratic regulator from the viewpoint of kernel methods.

Handling hard affine SDP shape constraints in RKHSs.
Technical report.
Hard shape-constrained kernel machines.

Infinite-task learning with RKHSs.
In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1294–1302.

*Quantile Regression.*
Cambridge University Press.

Non-parametric models for non-negative functions.
Non-parametric models for non-negative functions.

Learning psd-valued functions using kernel sums-of-squares.

Finding global minima via kernel approximations.

A generalized representer theorem.

Estimating the support of a high-dimensional distribution.
References IV
