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/

MLSS Tübingen, July 2020

Shape constraints, ex: monotonic kernel ridge regression

Shape constraints = priors on the form of the solution of the problem

$\hookrightarrow$ compensates lack of samples or excessive noise

$\hookrightarrow$ incorporates physical constraints

ex: monotone, convex functions or non-crossing quantiles are priors

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

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

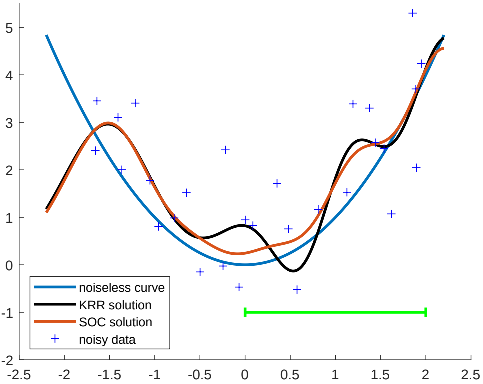
$$0 \leq Df(x), \quad \forall x \in K.$$

D is a differential operator (e.g. $Df = f'$), K a compact set (e.g. $[0, T]$).

For $K = [0, T]$, we have an infinite number of constraints!

Discretize constraint at $\{\tilde{x}_m\}_{m \leq M} \subset K$? No guarantees out-of-samples!

Goal: devise a technique for constraints to be satisfied



1D KRR with monotonic constraint over $[0, 2]$:

Unconstrained KRR

vs

Second-Order Cone (SOC) constrained

Let us add a buffer to the discretization (interior solution)

$$\text{“}0 \leq Df(x),\ \forall x \in K\text{“} \Leftarrow \text{“}\eta_{K,m}\|f(\cdot)\| \leq Df(\tilde{x}_m),\ \forall m \in [\![1, M]\!]\text{“}$$

How to choose $\eta_{K,m}$?

Reproducing kernel Hilbert spaces (RKHS) in one slide

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

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

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

$$k \text{ is s.t. } \mathbf{G} = [k(x_i, x_j)]_{i,j=1}^n \succcurlyeq 0 \text{ and } \mathcal{F}_k := \overline{\operatorname{span}(\{k_x(\cdot)\}_{x \in X})}$$

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

- There is a one-to-one correspondence between kernels k and RKHSs $(\mathcal{F}_k, \langle \cdot, \cdot \rangle_{\mathcal{F}_k})$. Changing X or $\langle \cdot, \cdot \rangle_{\mathcal{F}_k}$ changes the kernel k.
- if X is an open set, $k \in \mathcal{C}^{m,m}(X, X)$, D a differential operator of order at most m, then kernel trick for derivatives holds

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

Back to Second-Order Cone Constraints

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

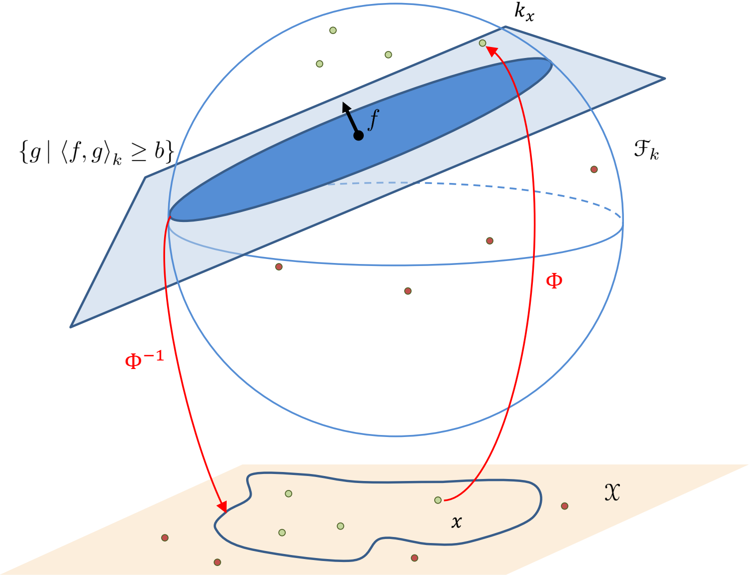
$$Df(x) = Df(\tilde{x}_m) + \langle f(\cdot), D_x k(x,\cdot) - D_x k(\tilde{x}_m,\cdot)\rangle_k$$
$$Df(x) \geq Df(\tilde{x}_m) - \|f(\cdot)\|_k \|D_x k(x,\cdot) - D_x k(\tilde{x}_m,\cdot)\|_k$$
$$Df(x) \geq Df(\tilde{x}_m) - \|f(\cdot)\|_k \underbrace{\sup_{\{x \,|\, \|x-\tilde{x}_m\|\leq\delta\}} \|D_x k(x,\cdot) - D_x k(\tilde{x}_m,\cdot)\|_k}_{\eta_{K,m}(\delta)}$$

For smooth kernels, $\delta \to 0$ gives $\eta_{K,m}(\delta) \to 0$.

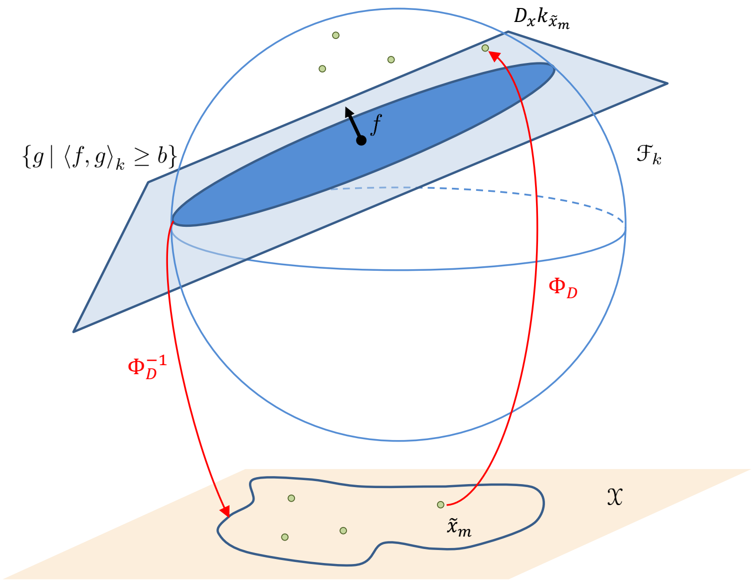
Shift-invariant kernel ($k(x,y) = k_0(x-y)$) gives

$$\eta = \sup_{u \in \mathbb{B}_{\|\cdot\|_{\mathcal{X}}}(0,1)} \sqrt{|2D_x D_y k_0(0) - 2D_x D_y k_0(\delta u)|}$$

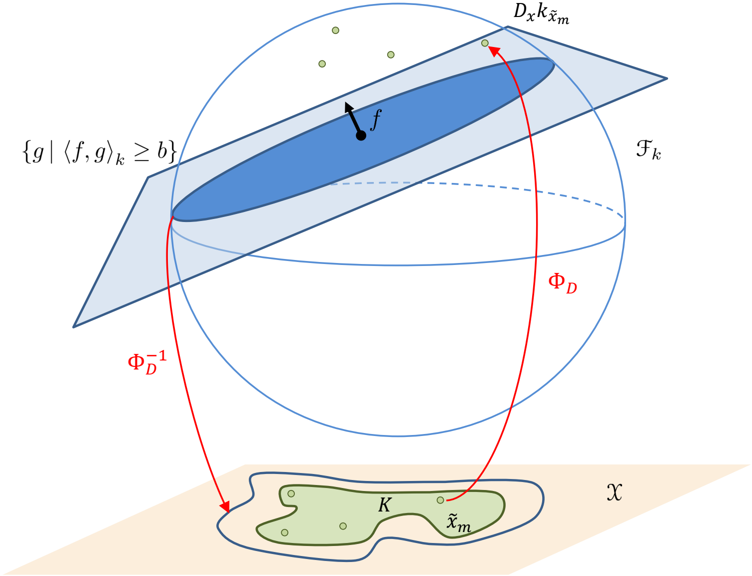
Other buffers were possible (e.g. constant), why choose "$\eta_{K,m}\|f(\cdot)\|$"?



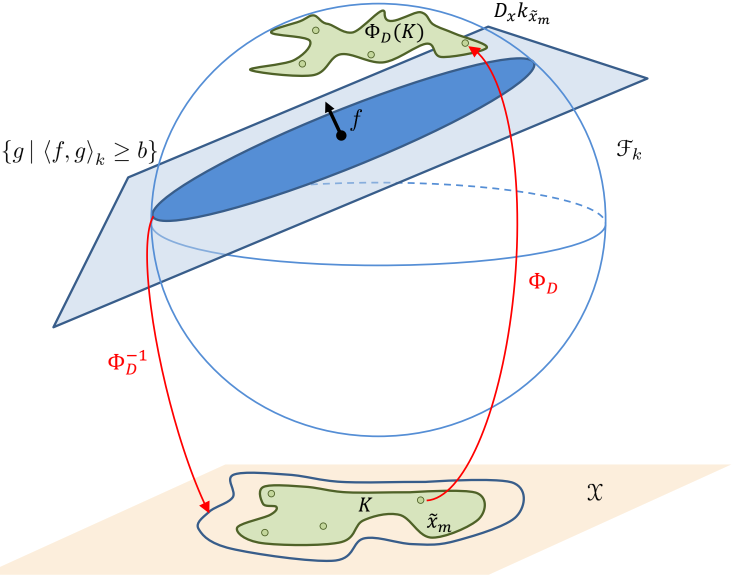
SVM is about separating red and green points by blue hyperplane.



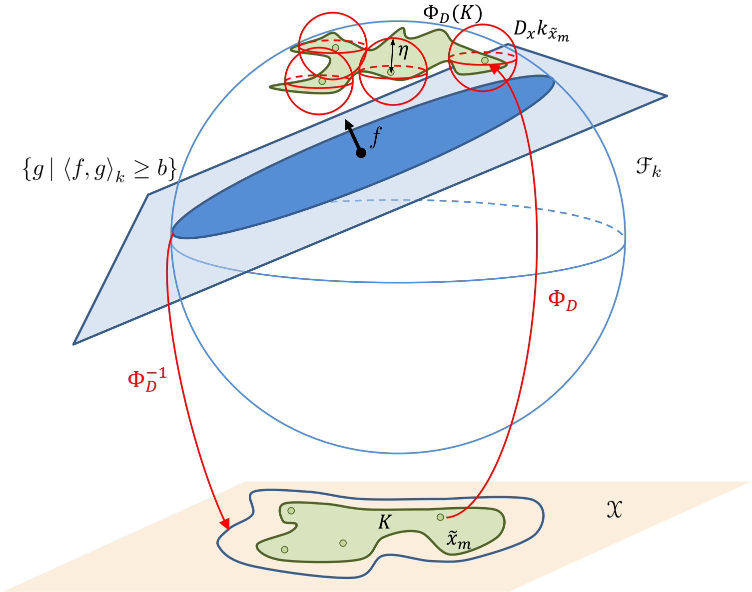
Using the nonlinear embedding $\Phi_D : x \mapsto D_x k(x, \cdot)$, the idea is the same.
Consider only the green points, it looks like one-class SVM.



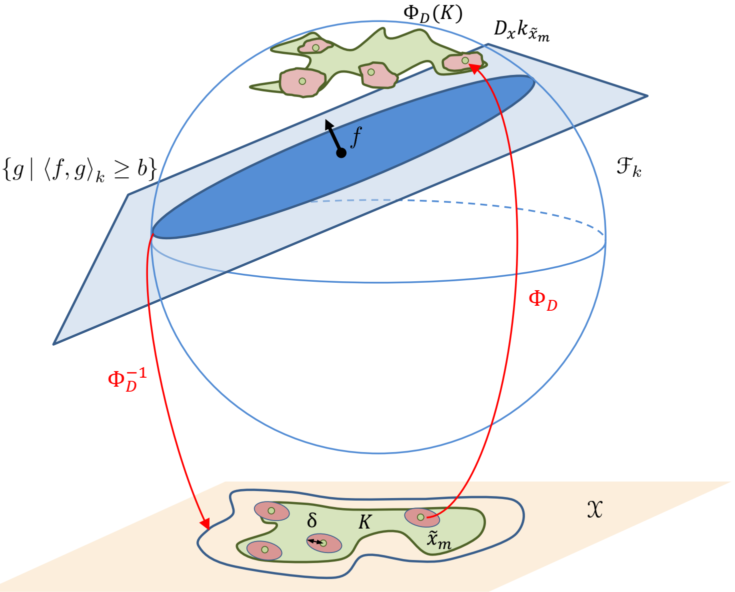
The green points are now samples of a compact set K.



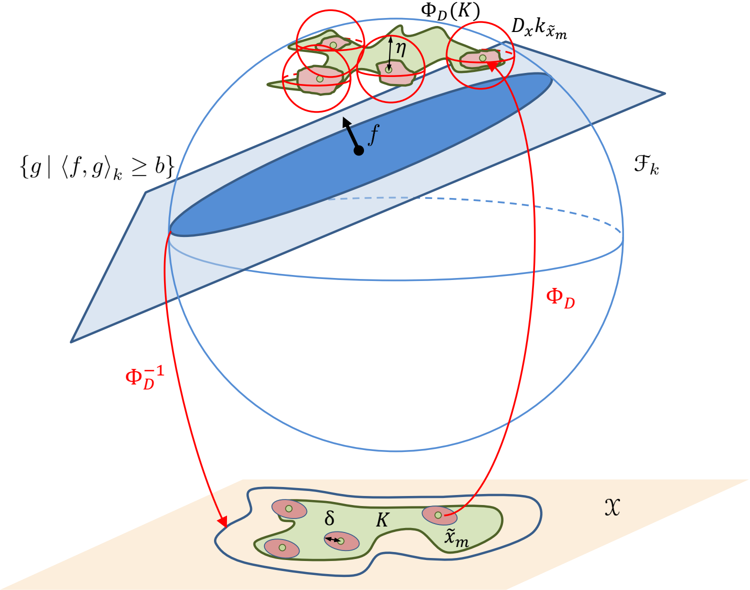
The image $\Phi_D(K)$ looks ugly...



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



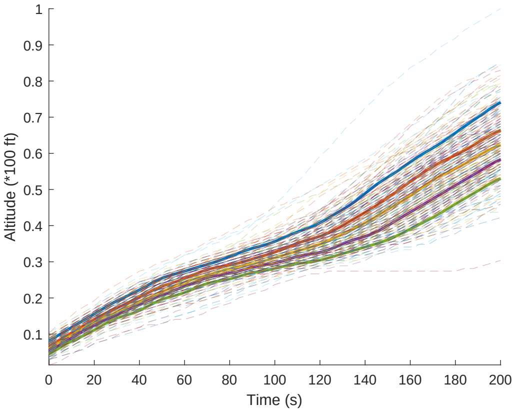
First cover $K \subset \bigcup\{\tilde{x}_m + \delta\mathbb{B}\}$, and then look at the images $\Phi_D(\{\tilde{x}_m + \delta\mathbb{B}\})$



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

Joint quantile regression (JQR): airplane data

Airplane trajectories at takeoff have increasing altitude.



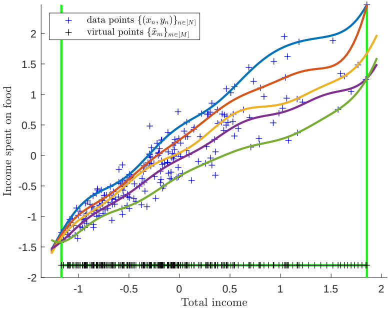
JQR with monotonic constraint over $[x_{\min}, x_{\max}]$:

Quantiles should be non-crossing

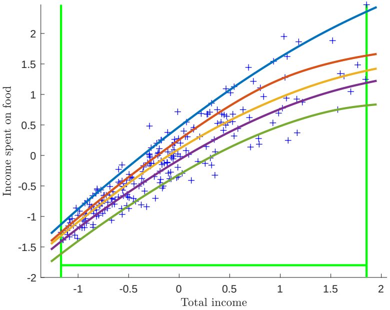
Two shape constraints jointly handled with $15k$ samples.
Works with higher dimensions too!

Joint quantile regression (JQR): Engel's law

As income rises, the proportion of income spent on food falls, but absolute expenditure on food rises.



Increasing+non-crossing



Increasing+non-crossing+concave

Priors have a great effect on the shape of solutions!

Open problems:

- other types of compact coverings? Convex hull of union of sets?
- modify coverings while optimizing? e.g. greedily adding new samples

Deeply grateful to:

Nicolas Petit (Mines ParisTech), my PhD advisor

Jean-Philippe Vert (Mines ParisTech-Google), who got me to love kernels

Zoltán Szabó (Ecole polytechnique), my co-author for quantile regression

See *Hard Shape-Constrained Kernel Machines*, PCAF and Zoltán Szabó
`https://arxiv.org/abs/2005.12636`