Hard Shape-Constrained Kernel Regression

Pierre-Cyril Aubin-Frankowski[1], Zoltán Szabó[2]

[1]Ecole des Ponts ParisTech and CAS, MINES ParisTech, PSL, France [2]CMAP, CNRS, École Polytechnique, Institut Polytechnique de Paris, France

Quick Summary

- Shape constraints: priors on the form (e.g. nonnegativity) to
  - compensate lack of samples or excessive noise
  - incorporate physical constraints in an optimization problem
- Guarantee constraint satisfaction in kernel regression in a "hard" way?
- Define a strengthened problem through SOC constraints

Shape-constrained kernel regression

$$\bar{f} \in \underset{f \in \mathcal{F}_k}{\arg\min} \; L\left((\mathbf{x}_n, y_n, f(\mathbf{x}_n))_{n\in[N]}\right) + \Omega\left(\|f|_k\right)$$
$$\text{s.t.} \quad 0 \le Df(\mathbf{x}), \quad \forall \mathbf{x} \in K.$$

- D is a differential operator of order s, $K \subset \mathbb{R}^d$ a compact set (e.g. $[0, T]$)
- $\mathcal{F}_k$ is an RKHS, i.e. Hilbert space of real-valued functions, e.g. $W^{2,2}(\mathbb{R}^d)$

RKHSs are defined by a positive definite kernel $k(\cdot,\cdot)$ with, for $k \in \mathcal{C}^{s,s}$, the reproducing property: $Df(\mathbf{x}) = \langle f(\cdot), D_x k(\mathbf{x},\cdot)\rangle_{\mathcal{F}_k}$

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

Examples

- Kernel ridge regression with monotonicity constraint:

$$\mathcal{L}(f) := \frac{1}{N}\sum_{n\in[N]} |y_n - f(x_n)|^2 + \lambda_f\|f\|_k^2, \text{ s.t. } f'(x) \ge 0, \forall x \in [x_l, x_u]$$

- Joint quantile regression with non-crossing constraints, over $\{f_q + b_q\}_{q\in[Q]}$

$$\mathcal{L}(\mathbf{f}, \mathbf{b}) = \frac{1}{N}\sum_{q\in[Q]}\sum_{n\in[N]} l_{\tau_q}(y_n - [f_q(\mathbf{x}_n) + b_q]) + \lambda_{\mathbf{b}}\|\mathbf{b}\|_2^2 + \lambda_f \sum_{q\in[Q]}\|f_q\|_k^2$$
$$\text{s.t. } f_{q+1}(\mathbf{x}) + b_{q+1} \ge f_q(\mathbf{x}) + b_q, \forall q \in [Q-1], \forall \mathbf{x} \in [x_l, x_u]^d.$$

Idea

Soft way: discretize the shape constraint at $\{\tilde{\mathbf{x}}_m\}_{m\le M} \subset K$
$\hookrightarrow$ **No guarantees out-of-samples!**

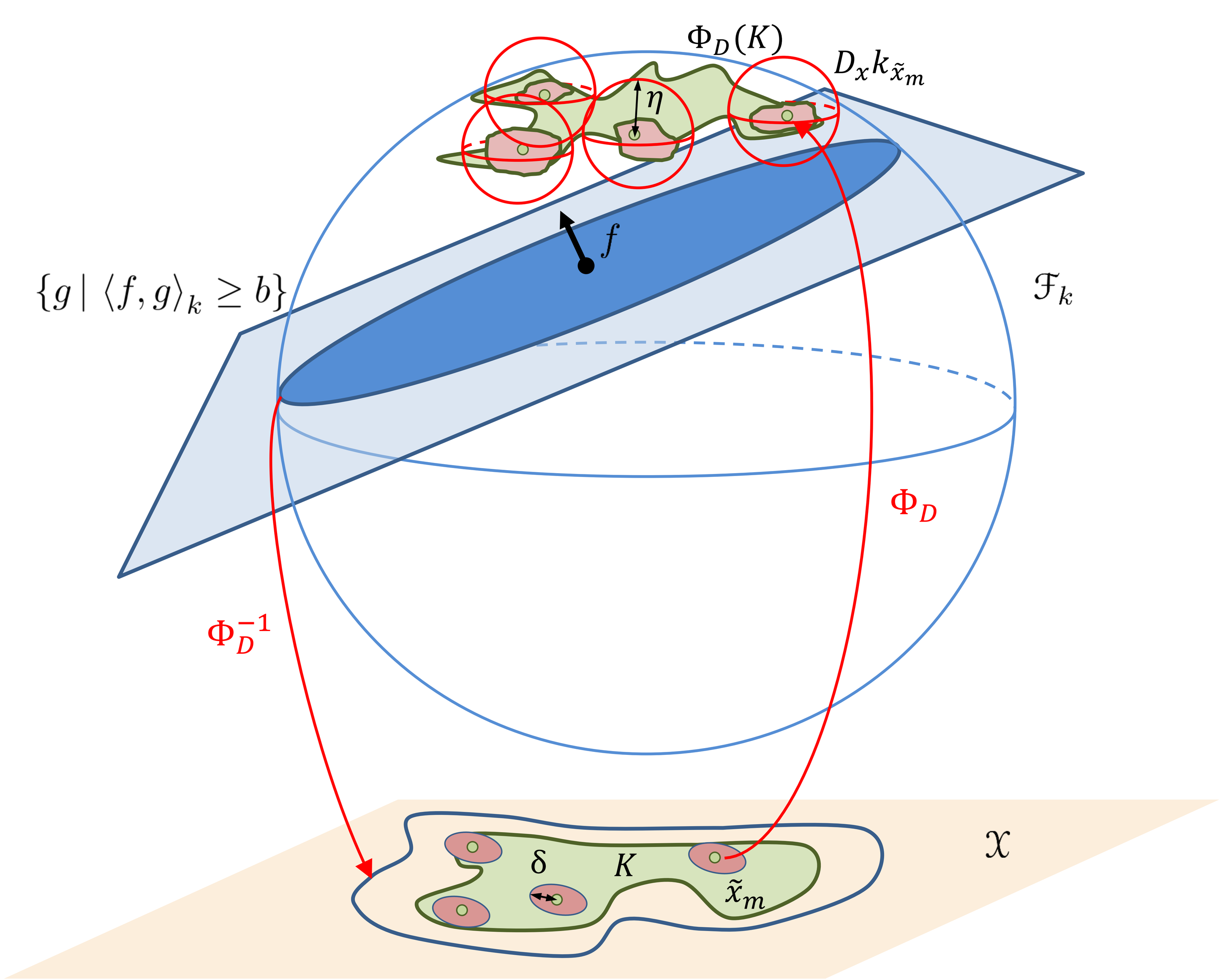
Hard way: take instead $\delta > 0$ and $\mathbf{x}$ s.t. $\|\mathbf{x} - \tilde{\mathbf{x}}_m\| \le \delta$

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

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

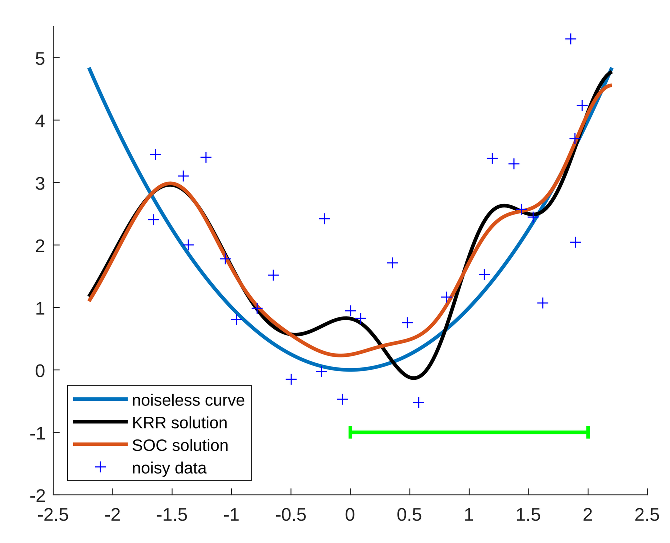
Acks: ZSz benefited from the support of the Europlace Institute of Finance and that of the Chair Stress Test RISK Management and Financial Steering, led by the French École Polytechnique and its Foundation and sponsored by BNP Paribas.

Geometrical intuition

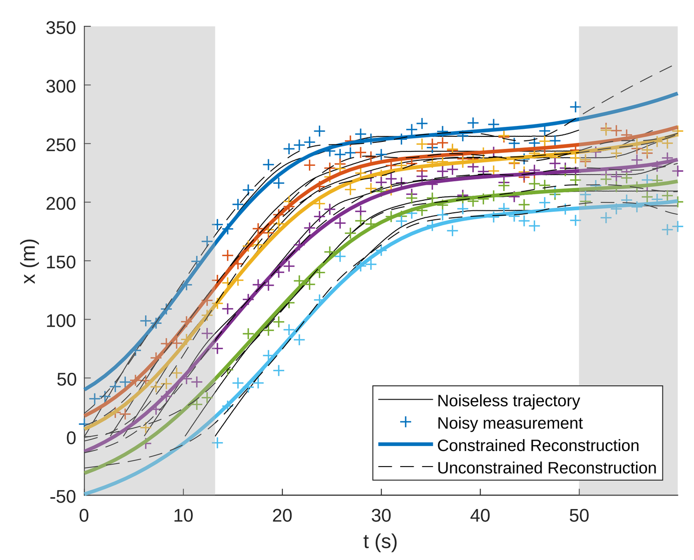


Numerical Illustrations

Kernel Ridge Regression with monotone constraint

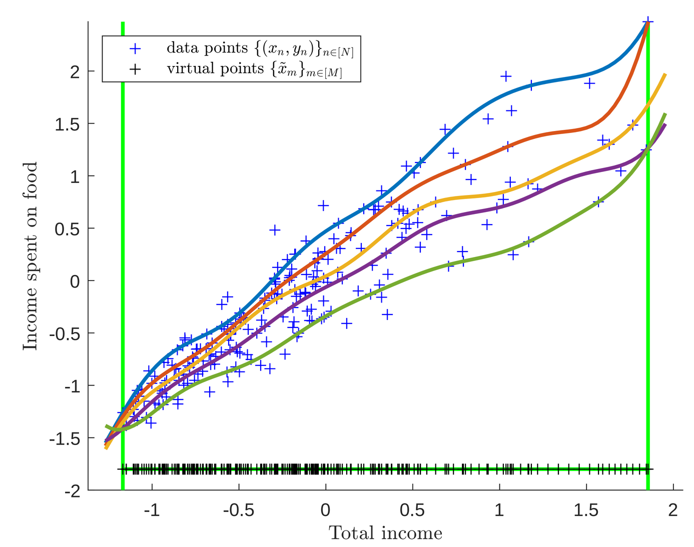


(a) Toy data with Gaussian noise [1]

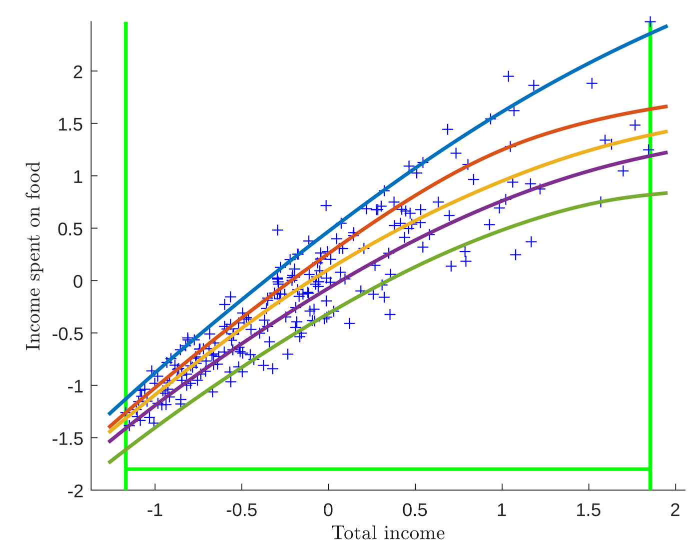


(b) Car data with traffic jam [2]

Joint Quantile Regression with non-crossing *and* increasing constraints



(c) Engel's law over household income [1]



(d) Adding a concavity constraint [1]

Goal

- Add a buffer to the discretization (interior solution)
  "$0 \le Df(\mathbf{x}), \forall \mathbf{x} \in K$" $\Leftarrow$ "$\eta_{K,m}\|f(\cdot)\|_k \le Df(\tilde{\mathbf{x}}_m), \forall m \in [\![1, M]\!]$"
  $\hookrightarrow$ This generates a SOC (second-order cone) constraint.
- Discuss geometrically the choice of $\eta_{K,m}$ and $\{\tilde{\mathbf{x}}_m\}_{m\le M}$
- Apply the method to various shape constraints
  - Trajectory reconstruction under speed and inter-vehicular distance constraints
  - Engel's law in economics (non-crossing/monotone/concave quantile functions)

Theoretical guarantees

Denote by v_{disc} the optimal value for the discretization ($\eta = 0$) and by v_η that of the SOC version

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

ii) Finite number of evaluations $\Rightarrow$ **representer theorem** (optimal solutions have a finite expression)

iii) If $\mathcal{L}$ is μ-strongly convex, we have a **computable bound**

$$\|f_\eta - \bar{f}\|_k \le \sqrt{\frac{2(v_\eta - v_{\text{disc}})}{\mu_f}}$$

Discussion

(i) This holds for given samples (optimization rather than statistical properties)
(ii) The representer theorem provides an equivalent finite-dimensional problem depending on the number N of samples $\mathbf{x}_n$ and M of virtual points $\tilde{\mathbf{x}}_m$
(iii) The smaller η = the smaller δ = the larger M = the costlier
(iv) The virtual points can be chosen among the samples (*recycling*)

Extensions

- SDP constraints (e.g. convexity for $d \ge 2$): $\mathbf{0} \preccurlyeq \mathbf{Hess}(f)(\mathbf{x})$
- Vector-valued functions $\mathbf{f} : \mathcal{X} \to \mathbb{R}^P$
- Other applications: finance, control theory..

References

[1] Pierre-Cyril Aubin-Frankowski and Zoltán Szabó. Hard shape-constrained kernel machines, 2020. https://arxiv.org/abs/2005.12636.

[2] Pierre-Cyril Aubin-Frankowski, Nicolas Petit, and Zoltán Szabó. Kernel regression for vehicle trajectory reconstruction under speed and inter-vehicular distance constraints. In *IFAC World Congress*, volume (to appear), 2020.