Shape-Constrained Estimation in Reproducing Kernel Hilbert Spaces

Zoltán Szabó

Joint work with:

- Pierre-Cyril Aubin-Frankowski @ MINES ParisTech,
- Nicolas Petit @ MINES ParisTech

ML Journal Club,
Palaiseau
March 5, 2020

Outline

- Shape constraints.
- Kernels and RKHS-s.
- Task & results.

Shape constraints

Basic examples

1. Non-negativity:

$$0 \leq f(x) \quad (\forall x).$$

Basic examples

1. Non-negativity:

$$0 \leq f(x) \quad (\forall x).$$

2. Monotonicity ($\nearrow$):

$$x \leq y \Rightarrow f(x) \leq f(y).$$

Equivalently, $0 \leq f'(x)$ for all x.

Basic examples

1. Non-negativity:

$$0 \leq f(x) \quad (\forall x).$$

2. Monotonicity ($\nearrow$):

$$x \leq y \Rightarrow f(x) \leq f(y).$$

   Equivalently, $0 \leq f'(x)$ for all x.

3. Convexity:

$$0 \leq f''(x) \quad (\forall x).$$

Other examples

4. n-monotonicity: $0 \leq f^{(n)}(x)$ $(\forall x)$.

Other examples

4. n-monotonicity: $0 \leq f^{(n)}(x)$ $(\forall x)$.
5. $(n-1)$-alternating monotonicity: for $n \geq 2$

$$(-1)^j f^{(j)} : \geq 0, \nearrow \text{ and convex } \forall j \in [\![0, n-2]\!].$$

Other examples

4. n-monotonicity: $0 \leq f^{(n)}(x)$ $(\forall x)$.
5. $(n-1)$-alternating monotonicity: for $n \geq 2$

$$(-1)^j f^{(j)} : \geq 0, \nearrow \text{ and convex } \forall j \in [\![0, n-2]\!].$$

Example: generator of a d-variate Archimedean copula is $(d-2)$-alternating monotone.

Other examples from economics, game theory

6. Monotonicity w.r.t. partial ordering ($\mathbf{u} \preccurlyeq \mathbf{v} \Rightarrow f(\mathbf{u}) \leq f(\mathbf{v})$):

$\mathbf{u} \preccurlyeq \mathbf{v}$ iff

- $u_i \leq v_i$ ($\forall i$; product ordering),
- $\sum_{j \in [i]} u_j \leq \sum_{j \in [i]} v_j$ ($\forall i$; unordered weak majorization).

Other examples from economics, game theory

6. Monotonicity w.r.t. partial ordering ($\mathbf{u} \preccurlyeq \mathbf{v} \Rightarrow f(\mathbf{u}) \leq f(\mathbf{v})$):

$$0 \leq \partial^{\mathbf{e}_j} f(\mathbf{x}), \quad (\forall j \in [d], \forall \mathbf{x}),$$

$$0 \leq \partial^{\mathbf{e}_d} f(\mathbf{x}) \leq \ldots \leq \partial^{\mathbf{e}_1} f(\mathbf{x}) \quad (\forall \mathbf{x}).$$

$\mathbf{u} \preccurlyeq \mathbf{v}$ iff

- $u_i \leq v_i$ ($\forall i$; product ordering),
- $\sum_{j \in [i]} u_j \leq \sum_{j \in [i]} v_j$ ($\forall i$; unordered weak majorization).

Other examples from economics, game theory

6. Monotonicity w.r.t. partial ordering ($\mathbf{u} \preccurlyeq \mathbf{v} \Rightarrow f(\mathbf{u}) \leq f(\mathbf{v})$):

$$0 \leq \partial^{\mathbf{e}_j} f(\mathbf{x}), \quad (\forall j \in [d], \forall \mathbf{x}),$$

$$0 \leq \partial^{\mathbf{e}_d} f(\mathbf{x}) \leq \ldots \leq \partial^{\mathbf{e}_1} f(\mathbf{x}) \quad (\forall \mathbf{x}).$$

$\mathbf{u} \preccurlyeq \mathbf{v}$ iff

- $u_i \leq v_i$ ($\forall i$; product ordering),
- $\sum_{j \in [i]} u_j \leq \sum_{j \in [i]} v_j$ ($\forall i$; unordered weak majorization).

7. Supermodularity:

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

i.e. $f(\mathbf{u} \vee \mathbf{v}) + f(\mathbf{u} \wedge \mathbf{v}) \geq f(\mathbf{u}) + f(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$.

Other examples from economics, game theory

6. Monotonicity w.r.t. partial ordering ($\mathbf{u} \preccurlyeq \mathbf{v} \Rightarrow f(\mathbf{u}) \leq f(\mathbf{v})$):

$$0 \leq \partial^{\mathbf{e}_j} f(\mathbf{x}), \quad (\forall j \in [d], \forall \mathbf{x}),$$

$$0 \leq \partial^{\mathbf{e}_d} f(\mathbf{x}) \leq \ldots \leq \partial^{\mathbf{e}_1} f(\mathbf{x}) \quad (\forall \mathbf{x}).$$

$\mathbf{u} \preccurlyeq \mathbf{v}$ iff

- $u_i \leq v_i$ ($\forall i$; product ordering),
- $\sum_{j \in [i]} u_j \leq \sum_{j \in [i]} v_j$ ($\forall i$; unordered weak majorization).

7. Supermodularity:

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

i.e. $f(\mathbf{u} \vee \mathbf{v}) + f(\mathbf{u} \wedge \mathbf{v}) \geq f(\mathbf{u}) + f(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$.

Pattern

$$0 \leq Df(\mathbf{x}) \quad \forall \mathbf{x}.$$

Our high-level goal & challenge

- Given: $\{(\mathbf{x}_n, y_n)\}_{n\in[N]} \subset \mathbb{R}^d \times \mathbb{R}$ samples.
- Goal: find $f \in \mathcal{H}$ such that

$$f(\mathbf{x}_n) \approx y_n,$$
$$0 \leq Df(\mathbf{x}) \quad \forall \mathbf{x} \in K.$$

Our high-level goal & challenge

- Given: $\{(\mathbf{x}_n, y_n)\}_{n\in[N]} \subset \mathbb{R}^d \times \mathbb{R}$ samples.
- Goal: find $f \in \mathcal{H}$ such that

$$f(\mathbf{x}_n) \approx y_n,$$
$$0 \leq Df(\mathbf{x}) \quad \forall \mathbf{x} \in K.$$

Typical approaches:

- soft constraints: finite many points.
- constraint-specific parametrization: exponential/quadratic.
- restricted function classes: polynomials, or polynomial splines.
- asymptotic guarantees.

Our high-level goal & challenge

- Given: $\{(\mathbf{x}_n, y_n)\}_{n\in[N]} \subset \mathbb{R}^d \times \mathbb{R}$ samples.
- Goal: find $f \in \mathcal{H}$ such that

$$f(\mathbf{x}_n) \approx y_n,$$
$$0 \leq Df(\mathbf{x}) \quad \forall \mathbf{x} \in K.$$

Typical approaches:

- soft constraints: finite many points.
- constraint-specific parametrization: exponential/quadratic.
- restricted function classes: polynomials, or polynomial splines.
- asymptotic guarantees.

In our work

$\mathcal{H}$: RKHS; rich but tractable. Hard constraint & performance guarantees.

Kernel, RKHS

Motivation

Extension of $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}$ leads to kernels.

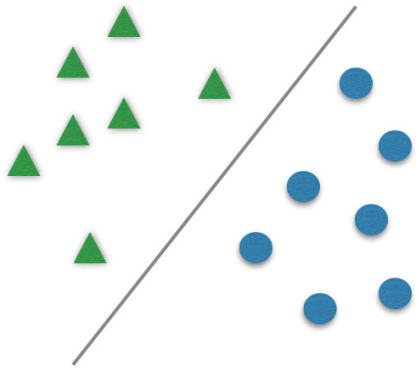
Motivation

Extension of $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}$ leads to kernels. Why?

Motivation

Extension of $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}$ leads to kernels. **Why?**

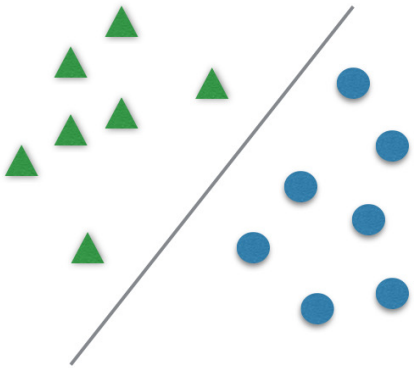
- Classification (SVM):

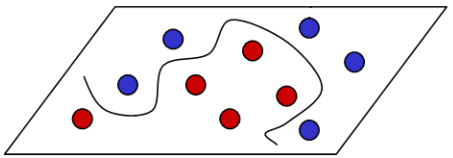


Motivation

Extension of $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}$ leads to kernels. **Why?**

- Classification (SVM):



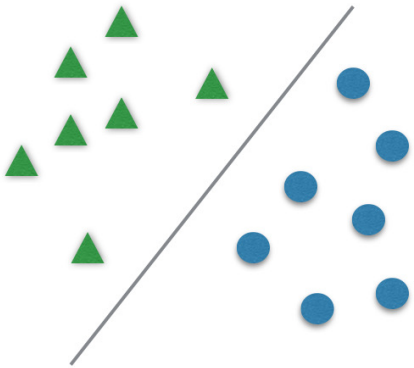


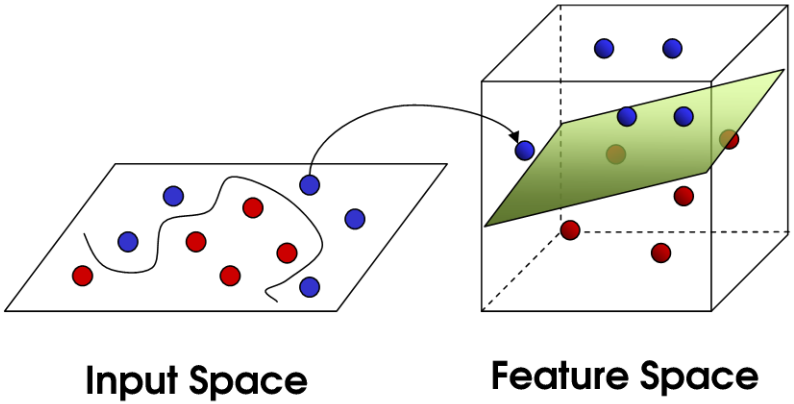
Input Space

Motivation

Extension of $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}$ leads to kernels. **Why?**

- Classification (SVM):

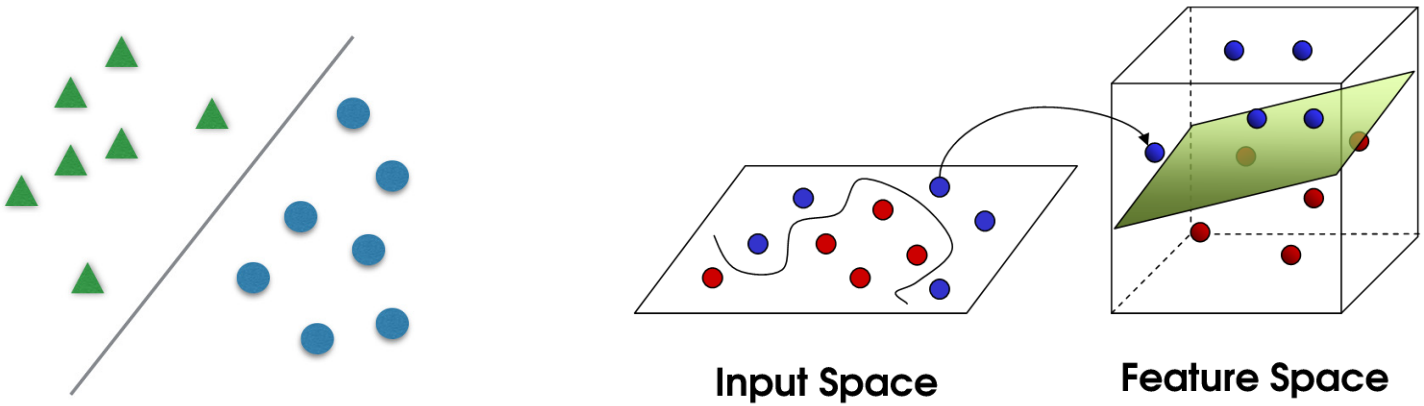




Motivation

Extension of $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}$ leads to kernels. **Why?**

- Classification (SVM):



- Representation of distributions:

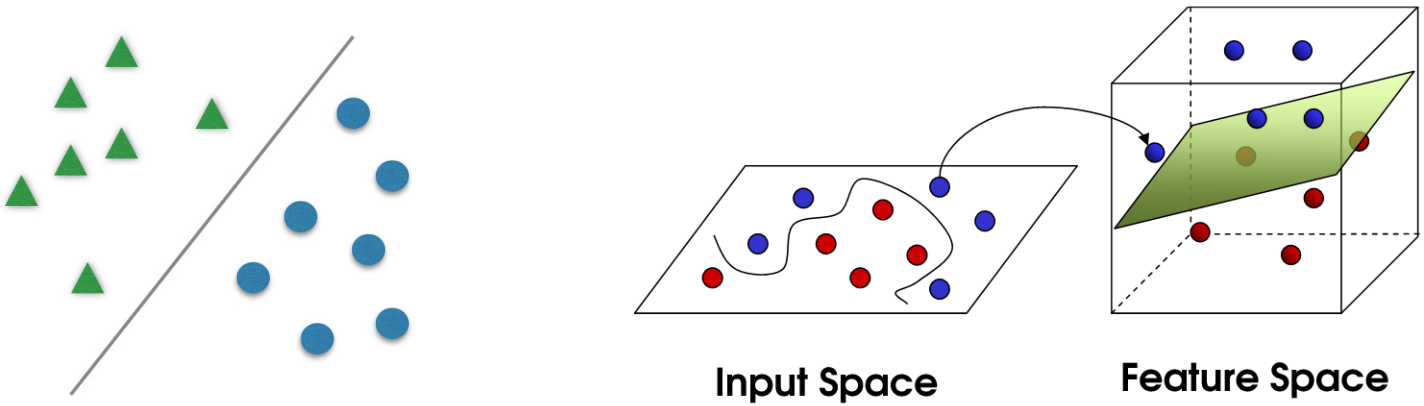
$$\mathbb{P} \mapsto \mathbb{E}_{\mathbf{x}\sim\mathbb{P}}\varphi(\mathbf{x}),$$

divergence measures (MMD), independence measures (HSIC, KCCA, KGV), hypothesis testing.

Motivation

Extension of $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}$ leads to kernels. **Why?**

- Classification (SVM):



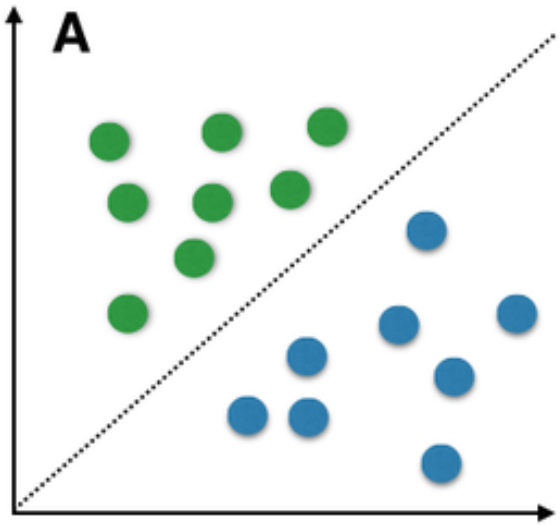
- Representation of distributions:

$$\mathbb{P} \mapsto \mathbb{E}_{\mathbf{x}\sim\mathbb{P}}\varphi(\mathbf{x}),$$

  divergence measures (MMD), independence measures (HSIC, KCCA, KGV), hypothesis testing.
- Gaussian processes (covariance function), Fourier analysis, ...

Classification motivation: linear separability

Idealized situation

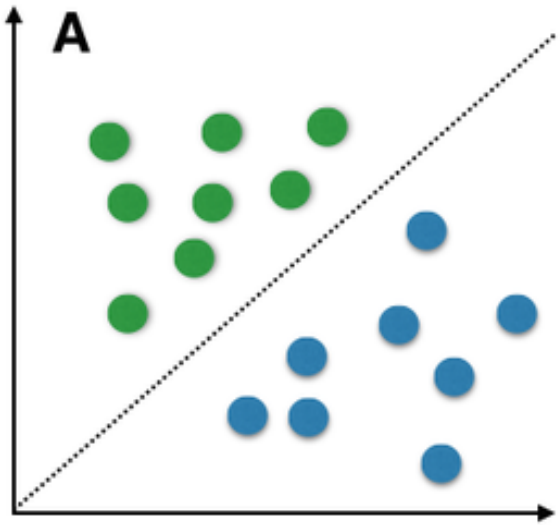


Decision surface:

$$\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle = 0\}$$

Classification motivation: linear separability

Idealized situation



Decision surface:

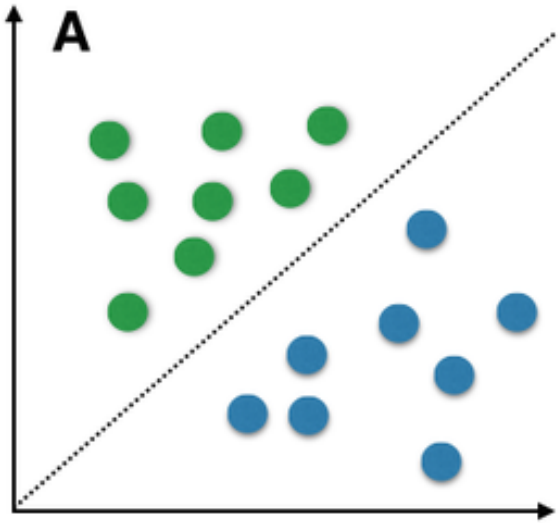
$$\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle = 0\} \Rightarrow$$

classes:

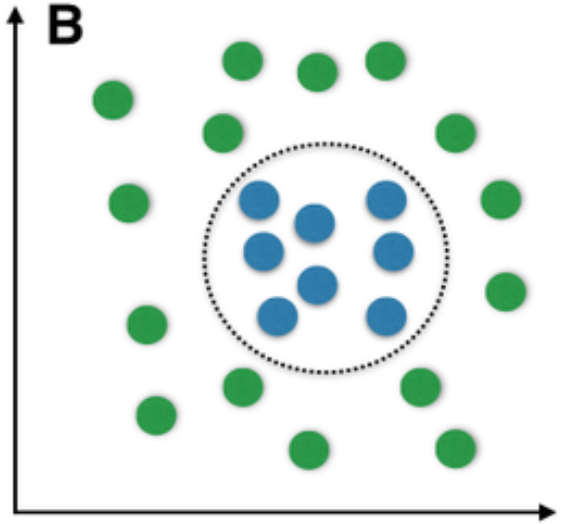
$$\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle \geq 0\} \qquad \{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle < 0\}$$

Classification motivation: non-linear separability

Idealized situation



Real world



Decision surface (left):

$$\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle = 0\} \Rightarrow$$

classes:

$$\{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle \geq 0\} \qquad \{\mathbf{x} : \langle \mathbf{w}, \mathbf{x} \rangle < 0\}.$$

Non-linear separability – continued

On the ellipse

$$\left\{ \mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} = 1 \right\}$$

Non-linear separability – continued

On the ellipse, outside

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} = 1\right\},$$

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} > 1\right\}$$

Non-linear separability – continued

On the ellipse, outside, inside:

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} = 1\right\},$$

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} > 1\right\},$$

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} < 1\right\}.$$

Non-linear separability – continued

On the ellipse, outside, inside:

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} = 1\right\},$$
$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} > 1\right\},$$
$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} < 1\right\}.$$

With polynomial feature: $\varphi(\mathbf{x}) = \left(x_1^2, x_1, 1, x_2^2, x_2\right)$:

- Decision surface: $\{\mathbf{x} : \langle \mathbf{w}, \varphi(\mathbf{x}) \rangle = 0\}$.

Non-linear separability – continued

On the ellipse, outside, inside:

$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} = 1\right\},$$
$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} > 1\right\},$$
$$\left\{\mathbf{x} : \frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{b^2} < 1\right\}.$$

With polynomial feature: $\varphi(\mathbf{x}) = \left(x_1^2, x_1, 1, x_2^2, x_2\right)$:

- Decision surface: $\{\mathbf{x} : \langle \mathbf{w}, \varphi(\mathbf{x})\rangle = 0\}$.
- Classes: $\{\mathbf{x} : \langle \mathbf{w}, \varphi(\mathbf{x})\rangle \geq 0\}$, $\{\mathbf{x} : \langle \mathbf{w}, \varphi(\mathbf{x})\rangle < 0\}$.

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle$ =?

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle = \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle$$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle = \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle$$

$$= x_1^2(x_1')^2 + \underbrace{\sqrt{2}\sqrt{2}}_{2} x_1x_2(x_1')(x_2') + x_2^2(x_2')^2$$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle = \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle$$

$$= x_1^2(x_1')^2 + \underbrace{\sqrt{2}\sqrt{2}}_{2} x_1x_2(x_1')(x_2') + x_2^2(x_2')^2$$

$$= \left(x_1x_1' + x_2x_2'\right)^2$$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle = \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle$$

$$= x_1^2(x_1')^2 + \underbrace{\sqrt{2}\sqrt{2}}_{2} x_1x_2(x_1')(x_2') + x_2^2(x_2')^2$$

$$= \left(x_1x_1' + x_2x_2'\right)^2$$

$$= \left\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} \right\rangle^2 = \langle \mathbf{x}, \mathbf{x}' \rangle^2 =: k(\mathbf{x}, \mathbf{x}').$$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle = \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle$$

$$= x_1^2(x_1')^2 + \underbrace{\sqrt{2}\sqrt{2}}_{2} x_1x_2(x_1')(x_2') + x_2^2(x_2')^2$$

$$= \left(x_1x_1' + x_2x_2'\right)^2$$

$$= \left\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} \right\rangle^2 = \langle \mathbf{x}, \mathbf{x}' \rangle^2 =: k(\mathbf{x}, \mathbf{x}').$$

$\langle \mathbf{x}, \mathbf{x}' \rangle^d = \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle$: $\varphi(\mathbf{x}) =$ d-order polynomial. $\Rightarrow$

Quadratic & polynomial features

Still in $\mathbb{R}^2$:

$$\varphi(\mathbf{x}) = \left(x_1^2, \sqrt{2}x_1x_2, x_2^2\right),$$

$$\langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle = \left\langle \begin{bmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{bmatrix}, \begin{bmatrix} (x_1')^2 \\ \sqrt{2}(x_1')(x_2') \\ (x_2')^2 \end{bmatrix} \right\rangle$$

$$= x_1^2(x_1')^2 + \underbrace{\sqrt{2}\sqrt{2}}_{2} x_1x_2(x_1')(x_2') + x_2^2(x_2')^2$$

$$= \left(x_1x_1' + x_2x_2'\right)^2$$

$$= \left\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} \right\rangle^2 = \langle \mathbf{x}, \mathbf{x}' \rangle^2 =: k(\mathbf{x}, \mathbf{x}').$$

$\langle \mathbf{x}, \mathbf{x}' \rangle^d = \langle \varphi(\mathbf{x}), \varphi(\mathbf{x}') \rangle$: $\varphi(\mathbf{x}) = d$-order polynomial. $\Rightarrow$ Explicit computation would be heavy!

Kernel, RKHS

- Def-1 (feature space):

$$k(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}} .$$

Kernel, RKHS

- Def-1 (feature space):

$$k(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}} .$$

- Def-2 (reproducing kernel):

$$k(\cdot, x) \in \mathcal{H}, \qquad f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}}.$$

Constructively, $\mathcal{H}_k = \overline{\{\sum_{i=1}^{n} \alpha_i k(\cdot, x_i)\}}$.

Kernel, RKHS

- Def-1 (feature space):

$$k(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}} .$$

- Def-2 (reproducing kernel):

$$k(\cdot, x) \in \mathcal{H}, \qquad f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}}.$$

  Constructively, $\mathcal{H}_k = \overline{\{\sum_{i=1}^{n} \alpha_i k(\cdot, x_i)\}}$.

- Def-3 (Gram matrix): $\mathbf{G} = [k(x_i, x_j)]_{i,j=1}^{n} \succeq 0$.

Kernel, RKHS

- Def-1 (feature space):

$$k(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}} .$$

- Def-2 (reproducing kernel):

$$k(\cdot, x) \in \mathcal{H}, \qquad f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}}.$$

  Constructively, $\mathcal{H}_k = \overline{\{\sum_{i=1}^{n} \alpha_i k(\cdot, x_i)\}}$.
- Def-3 (Gram matrix): $\mathbf{G} = [k(x_i, x_j)]_{i,j=1}^{n} \succeq 0$.
- Def-4 (evaluation): $\delta_x(f) = f(x)$ is continuous for all x.

Kernel, RKHS

- Def-1 (feature space):

$$k(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}} .$$

- Def-2 (reproducing kernel):

$$k(\cdot, x) \in \mathcal{H}, \qquad f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}}.$$

  Constructively, $\mathcal{H}_k = \overline{\{\sum_{i=1}^{n} \alpha_i k(\cdot, x_i)\}}$.
- Def-3 (Gram matrix): $\mathbf{G} = [k(x_i, x_j)]_{i,j=1}^{n} \succeq 0$.
- Def-4 (evaluation): $\delta_x(f) = f(x)$ is continuous for all x.

- All these definitions are equivalent, $k \overset{1:1}{\leftrightarrow} \mathcal{H}_k$.
- $f(x) = \langle \mathbf{w}, \varphi(x) \rangle_{\mathbb{R}^M} = \sum_{m=1}^{M} w_m \varphi_m(x)$: Fourier analysis, splines, ...

Kernel examples on $\mathbb{R}^d$ ($\gamma, \sigma, \nu > 0$, $c \geq 0$, $p \in \mathbb{Z}^+$)

$$k_p(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + c)^p$$

Kernel examples on $\mathbb{R}^d$ ($\gamma, \sigma, \nu > 0$, $c \geq 0$, $p \in \mathbb{Z}^+$)

$$k_p(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + c)^p, \qquad k_G(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x}-\mathbf{y}\|_2^2},$$

$$k_e(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x}-\mathbf{y}\|_2}, \qquad k_L(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x}-\mathbf{y}\|_1},$$

$$k_C(\mathbf{x}, \mathbf{y}) = \frac{1}{1 + \gamma \|\mathbf{x} - \mathbf{y}\|_2^2}, \qquad k_{\tilde{e}}(\mathbf{x}, \mathbf{y}) = e^{\gamma \langle \mathbf{x}, \mathbf{y} \rangle}.$$

Kernel examples on $\mathbb{R}^d$ ($\gamma, \sigma, v > 0$, $c \geq 0$, $p \in \mathbb{Z}^+$)

$$k_p(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + c)^p, \qquad k_G(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x}-\mathbf{y}\|_2^2},$$

$$k_e(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x}-\mathbf{y}\|_2}, \qquad k_L(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x}-\mathbf{y}\|_1},$$

$$k_C(\mathbf{x}, \mathbf{y}) = \frac{1}{1 + \gamma \|\mathbf{x} - \mathbf{y}\|_2^2}, \qquad k_{\tilde{e}}(\mathbf{x}, \mathbf{y}) = e^{\gamma \langle \mathbf{x}, \mathbf{y} \rangle}.$$

Or the flexible Matérn family:

$$k_M(\mathbf{x}, \mathbf{y}) = \frac{2^{1-v}}{\Gamma(v)} \left(\frac{\sqrt{2v}\, \|\mathbf{x} - \mathbf{y}\|_2}{\sigma} \right)^v K_v \left(\frac{\sqrt{2v}\, \|\mathbf{x} - \mathbf{y}\|_2}{\sigma} \right),$$

where

- K_v: modified Bessel function of the second kind of order v,
- Specific cases: For $v = \frac{1}{2}$ one gets $k(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|_2}{\sigma}}$. Gaussian kernel: $v \to \infty$.

Kernels on other domains ($\mathcal{X}$)

- Strings [Watkins, 1999, Lodhi et al., 2002, Leslie et al., 2002, Kuang et al., 2004, Leslie and Kuang, 2004, Saigo et al., 2004, Cuturi and Vert, 2005],
- time series [Rüping, 2001, Cuturi et al., 2007, Cuturi, 2011, Király and Oberhauser, 2019],
- trees [Collins and Duffy, 2001, Kashima and Koyanagi, 2002],
- groups and specifically rankings [Cuturi et al., 2005, Jiao and Vert, 2016],
- sets [Haussler, 1999, Gärtner et al., 2002],
- various generative models [Jaakkola and Haussler, 1999, Tsuda et al., 2002, Seeger, 2002, Jebara et al., 2004],
- fuzzy domains [Guevara et al., 2017], or
- graphs [Kondor and Lafferty, 2002, Gärtner et al., 2003, Kashima et al., 2003, Borgwardt and Kriegel, 2005, Shervashidze et al., 2009, Vishwanathan et al., 2010, Kondor and Pan, 2016, Bai et al., 2018].

Properties of k control that of $\mathcal{H}_k$ and its usefulness

- k: bounded $[\sup_{x,y \in \mathcal{X}} k(x,y) \leq C] \Rightarrow \forall f \in \mathcal{H}_k$ is bounded

Properties of k control that of $\mathcal{H}_k$ and its usefulness

- k: bounded $[\sup_{x,y\in\mathcal{X}} k(x,y) \leq C] \Rightarrow \forall f \in \mathcal{H}_k$ is bounded:

$$|f(x)| = \left|\langle f, k(\cdot,x)\rangle_{\mathcal{H}_k}\right| \overset{\text{CBS}}{\leq} \|f\|_{\mathcal{H}_k} \underbrace{\|k(\cdot,x)\|_{\mathcal{H}_k}}_{\sqrt{k(x,x)}}.$$

Properties of k control that of $\mathcal{H}_k$ and its usefulness

- k: bounded $[\sup_{x,y\in\mathcal{X}} k(x,y) \leq C] \Rightarrow \forall f \in \mathcal{H}_k$ is bounded:

$$|f(x)| = \left|\langle f, k(\cdot,x)\rangle_{\mathcal{H}_k}\right| \overset{\text{CBS}}{\leq} \|f\|_{\mathcal{H}_k} \underbrace{\|k(\cdot,x)\|_{\mathcal{H}_k}}_{\sqrt{k(x,x)}}.$$

- k: continuous $\Rightarrow \mathcal{H}_k$: separable $\left[\ell^2(\mathbb{N})\right]$.

Properties of k control that of $\mathcal{H}_k$ and its usefulness

- k: bounded $[\sup_{x,y\in\mathcal{X}} k(x,y) \leq C] \Rightarrow \forall f \in \mathcal{H}_k$ is bounded:

$$|f(x)| = \left|\langle f, k(\cdot,x)\rangle_{\mathcal{H}_k}\right| \overset{\text{CBS}}{\leq} \|f\|_{\mathcal{H}_k} \underbrace{\|k(\cdot,x)\|_{\mathcal{H}_k}}_{\sqrt{k(x,x)}}.$$

- k: continuous $\Rightarrow \mathcal{H}_k$: separable $\left[\ell^2(\mathbb{N})\right]$.
- k: bounded and continuous $\Rightarrow \forall f \in \mathcal{H}_k$ is bounded & continuous.

Properties of k control that of $\mathcal{H}_k$ and its usefulness

- k: bounded $[\sup_{x,y\in\mathcal{X}} k(x,y) \leq C] \Rightarrow \forall f \in \mathcal{H}_k$ is bounded:

$$|f(x)| = \left|\langle f, k(\cdot,x)\rangle_{\mathcal{H}_k}\right| \overset{\text{CBS}}{\leq} \|f\|_{\mathcal{H}_k} \underbrace{\|k(\cdot,x)\|_{\mathcal{H}_k}}_{\sqrt{k(x,x)}}.$$

- k: continuous $\Rightarrow \mathcal{H}_k$: separable $\left[\ell^2(\mathbb{N})\right]$.
- k: bounded and continuous $\Rightarrow \forall f \in \mathcal{H}_k$ is bounded & continuous.
- $k \in \mathcal{C}^m \Rightarrow \forall f \in \mathcal{H}_k$ is m-times continuously differentiable.

Properties of k control that of $\mathcal{H}_k$ and its usefulness

- k: bounded $[\sup_{x,y\in\mathcal{X}} k(x,y) \leq C] \Rightarrow \forall f \in \mathcal{H}_k$ is bounded:

$$|f(x)| = \left|\langle f, k(\cdot,x)\rangle_{\mathcal{H}_k}\right| \overset{\text{CBS}}{\leq} \|f\|_{\mathcal{H}_k} \underbrace{\|k(\cdot,x)\|_{\mathcal{H}_k}}_{\sqrt{k(x,x)}}.$$

- k: continuous $\Rightarrow \mathcal{H}_k$: separable $\left[\ell^2(\mathbb{N})\right]$.
- k: bounded and continuous $\Rightarrow \forall f \in \mathcal{H}_k$ is bounded & continuous.
- $k \in \mathcal{C}^m \Rightarrow \forall f \in \mathcal{H}_k$ is m-times continuously differentiable.
- k: analytic $\Rightarrow \forall f \in \mathcal{H}_k$ is analytic.

Properties of k control that of $\mathcal{H}_k$ and its usefulness

- k: bounded $[\sup_{x,y\in\mathcal{X}} k(x,y) \leq C] \Rightarrow \forall f \in \mathcal{H}_k$ is bounded:

$$|f(x)| = \left|\langle f, k(\cdot,x)\rangle_{\mathcal{H}_k}\right| \overset{\text{CBS}}{\leq} \|f\|_{\mathcal{H}_k} \underbrace{\|k(\cdot,x)\|_{\mathcal{H}_k}}_{\sqrt{k(x,x)}}.$$

- k: continuous $\Rightarrow \mathcal{H}_k$: separable $\left[\ell^2(\mathbb{N})\right]$.
- k: bounded and continuous $\Rightarrow \forall f \in \mathcal{H}_k$ is bounded & continuous.
- $k \in \mathcal{C}^m \Rightarrow \forall f \in \mathcal{H}_k$ is m-times continuously differentiable.
- k: analytic $\Rightarrow \forall f \in \mathcal{H}_k$ is analytic.
- k: universal $\Leftrightarrow \mathcal{H}_k$ is dense in $\mathcal{C}_b(\mathcal{X})$.

Properties of k control that of $\mathcal{H}_k$ and its usefulness

- k: bounded $[\sup_{x,y\in\mathcal{X}} k(x,y) \leq C] \Rightarrow \forall f \in \mathcal{H}_k$ is bounded:

$$|f(x)| = \left|\langle f, k(\cdot,x)\rangle_{\mathcal{H}_k}\right| \overset{\text{CBS}}{\leq} \|f\|_{\mathcal{H}_k} \underbrace{\|k(\cdot,x)\|_{\mathcal{H}_k}}_{\sqrt{k(x,x)}}.$$

- k: continuous $\Rightarrow \mathcal{H}_k$: separable $\left[\ell^2(\mathbb{N})\right]$.
- k: bounded and continuous $\Rightarrow \forall f \in \mathcal{H}_k$ is bounded & continuous.
- $k \in \mathcal{C}^m \Rightarrow \forall f \in \mathcal{H}_k$ is m-times continuously differentiable.
- k: analytic $\Rightarrow \forall f \in \mathcal{H}_k$ is analytic.
- k: universal $\Leftrightarrow \mathcal{H}_k$ is dense in $\mathcal{C}_b(\mathcal{X})$.
- k: characteristic $\Leftrightarrow \mathbb{P} \mapsto \int_{\mathcal{X}} \varphi(x)\mathrm{d}\mathbb{P}(x) \in \mathcal{H}_k$ is injective.

Kernel puzzle

Let

$$\mathcal{X} = \{0, 1\},$$

$$k(x, x') = \left\{ \begin{array}{ll} 1, & \text{if } x \neq x' \\ -1, & \text{if } x = x' \end{array} \right\}.$$

Kernel puzzle

Let

$$\mathcal{X} = \{0, 1\},$$
$$k(x, x') = \left\{ \begin{array}{ll} 1, & \text{if } x \neq x' \\ -1, & \text{if } x = x' \end{array} \right\}.$$

Puzzle

Is k a kernel?

Kernel puzzle

Let

$$\mathcal{X} = \{0, 1\},$$
$$k(x, x') = \left\{ \begin{array}{ll} 1, & \text{if } x \neq x' \\ -1, & \text{if } x = x' \end{array} \right\}.$$

Puzzle

Is k a kernel?

No!

$k(x, x) = \langle \varphi(x), \varphi(x) \rangle_{\mathcal{H}}$

Kernel puzzle

Let

$$\mathcal{X} = \{0, 1\},$$
$$k(x, x') = \left\{ \begin{array}{ll} 1, & \text{if } x \neq x' \\ -1, & \text{if } x = x' \end{array} \right\}.$$

Puzzle

Is k a kernel?

No!

$$k(x, x) = \langle \varphi(x), \varphi(x) \rangle_{\mathcal{H}} = \|\varphi(x)\|^2_{\mathcal{H}} \geq 0 \quad \text{(Gram with } n = 1\text{)},$$

Kernel puzzle

Let

$$\mathcal{X} = \{0, 1\},$$
$$k(x, x') = \left\{ \begin{array}{ll} 1, & \text{if } x \neq x' \\ -1, & \text{if } x = x' \end{array} \right\}.$$

Puzzle

Is k a kernel?

No!

$k(x, x) = \langle \varphi(x), \varphi(x) \rangle_{\mathcal{H}} = \|\varphi(x)\|^2_{\mathcal{H}} \geq 0$ (Gram with $n = 1$),

$k(0, 0) = k(1, 1) = -1$ (in our case).

Kernel puzzle

Let

$$\mathcal{X} = \{0, 1\},$$
$$k(x, x') = \left\{ \begin{array}{ll} 1, & \text{if } x \neq x' \\ -1, & \text{if } x = x' \end{array} \right\}.$$

Puzzle

Is k a kernel?

No!

$k(x, x) = \langle \varphi(x), \varphi(x) \rangle_{\mathcal{H}} = \|\varphi(x)\|^2_{\mathcal{H}} \geq 0$ (Gram with $n = 1$),

$k(0, 0) = k(1, 1) = -1$ (in our case).

Easy-to-check conditions for a $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ function to be kernel ?

Kernel factory

Let k, k_m kernels, $\alpha, \alpha_m \in \mathbb{R}^{\geq 0}$. Then, the followings are also kernels:

1. Non-negative shift: $k + \alpha$.
2. Cone: $\sum_{m=1}^{M} \alpha_m k_m$.
3. Limit: $k(x, x') := \lim_{m \to \infty} k_m(x, x')$.
4. Pre-post multiplication: $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, $f : \mathcal{X} \to \mathbb{R}$

$$\tilde{k}(x, y) = f(x) k(x, y) f(y).$$

5. Product:

$$\left(\otimes_{m=1}^{M} k_m\right)\left((x_1, \ldots, x_M), \left(x'_1, \ldots, x'_M\right)\right) = \prod_{m=1}^{M} k_m\left(x_m, x'_m\right).$$

Task & Results

Task-1: convoy localization, one vehicle ($Q = 1$)

- Given: noisy time-location samples $\{(t_n, x_n)\}_{n\in[N]} \subset \underbrace{[0, T]}_{=:\mathcal{T}} \times \mathbb{R}$.
- Goal: learn the (t, x) relation.
- Constraint: lower bound on speed ($v_{\min}$).

Task-1: convoy localization, one vehicle ($Q = 1$)

- Given: noisy time-location samples $\{(t_n, x_n)\}_{n\in[N]} \subset \underbrace{[0, T]}_{=:\mathcal{T}} \times \mathbb{R}$.
- Goal: learn the (t, x) relation.
- Constraint: lower bound on speed ($v_{\min}$).
- Objective:

$$\min_{b \in \mathbb{R},\, f \in \mathcal{H}_k} \left[\frac{1}{N} \sum_{n\in[N]} |x_n - (b + f(t_n))|^2 + \lambda \|f\|^2_{\mathcal{H}_k} \right]$$

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

$$v_{\min} \leq f'(t), \quad \forall t \in \mathcal{T}.$$

Task-2: convoy localization, multiple vehicles ($Q \geq 1$)

- Data: $\left\{(t_{q,n}, x_{q,n})_{n\in[N_q]}\right\} \subseteq \mathcal{T} \times \mathbb{R}$, $q \in [Q]$.
- Constraints: speed ($v_{\min}$), inter-vehicular distance ($d_{\min}$).
- Objective:

$$\min_{\substack{f_1,\ldots,f_Q \in \mathcal{H}_k,\\ b_1,\ldots,b_Q \in \mathbb{R}}} \frac{1}{Q}\sum_{q=1}^{Q}\left[\left(\frac{1}{N_q}\sum_{n=1}^{N_q}|x_{q,n} - (b_q + f_q(t_{q,n}))\,|^2\right) + \lambda\|f_q\|^2_{\mathcal{H}_k}\right]$$

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

$$d_{\min} + b_{q+1} + f_{q+1}(t) \leq b_q + f_q(t), \forall q \in [Q-1],\ t \in \mathcal{T},$$

$$v_{\min} \leq f_q'(t), \qquad \forall q \in [Q],\ t \in \mathcal{T}.$$

Task-2: joint quantile regression

- Given: $(\tau_q)_{q\in[Q]}$ levels $\nearrow$, $\{(\mathbf{x}_n, y_n)\}_{n\in[N]}$ samples.
- Estimate jointly the τ_q-quantiles of $\mathbb{P}(Y|X=\mathbf{x})$.

Task-2: joint quantile regression

- Given: $(\tau_q)_{q\in[Q]}$ levels $\nearrow$, $\{(\mathbf{x}_n, y_n)\}_{n\in[N]}$ samples.
- Estimate jointly the τ_q-quantiles of $\mathbb{P}(Y|X=\mathbf{x})$: f_q [Sangnier et al., 2016].
- Objective:

$$\mathcal{L}(\mathbf{f}, \mathbf{b}) = \frac{1}{N} \sum_{q\in[Q]} \sum_{n\in[N]} l_{\tau_q}(y_n - [f_q(\mathbf{x}_n) + b_q]) + \lambda_{\mathbf{b}} \|\mathbf{b}\|_2^2 + \lambda_f \sum_{q\in[Q]} \|f_q\|_k^2,$$

$$l_\tau(e) = \max(\tau e, (\tau - 1)e).$$

Task-2: joint quantile regression

- Given: $(\tau_q)_{q\in[Q]}$ levels $\nearrow$, $\{(\mathbf{x}_n, y_n)\}_{n\in[N]}$ samples.
- Estimate jointly the τ_q-quantiles of $\mathbb{P}(Y|X=\mathbf{x})$: f_q [Sangnier et al., 2016].
- Objective:

$$\mathcal{L}(\mathbf{f}, \mathbf{b}) = \frac{1}{N} \sum_{q\in[Q]} \sum_{n\in[N]} l_{\tau_q}(y_n - [f_q(\mathbf{x}_n) + b_q]) + \lambda_{\mathbf{b}} \|\mathbf{b}\|_2^2 + \lambda_f \sum_{q\in[Q]} \|f_q\|_k^2,$$

$$l_\tau(e) = \max(\tau e, (\tau - 1)e).$$

- Constraint (non-crossing): $K :=$ smallest rectangle containing $\{\mathbf{x}_n\}_{n\in[N]}$,

$$f_q(\mathbf{x}) + b_q \leq f_{q+1}(\mathbf{x}) + b_{q+1},\ \forall q \in [Q-1],\ \forall \mathbf{x} \in K.$$

Task: general

$$(\bar{\mathbf{f}}, \bar{\mathbf{b}}) = \underset{\substack{\mathbf{f}=(f_q)_{q\in[Q]} \in (\mathcal{H}_k)^Q, \\ \mathbf{b}=(b_q)_{q\in[Q]} \in \mathcal{B}, \\ (\mathbf{f},\mathbf{b}) \in C}}{\arg\min} \mathcal{L}(\mathbf{f}, \boldsymbol{b}),$$

Task: general

$$(\bar{\mathbf{f}}, \bar{\mathbf{b}}) = \underset{\substack{\mathbf{f}=(f_q)_{q\in[Q]} \in (\mathcal{H}_k)^Q, \\ \mathbf{b}=(b_q)_{q\in[Q]} \in \mathcal{B}, \\ (\mathbf{f},\mathbf{b}) \in C}}{\arg\min} \mathcal{L}(\mathbf{f}, \boldsymbol{b}),$$

$$\mathcal{L}(\mathbf{f}, \boldsymbol{b}) = L\left(\mathbf{b}, \{\mathbf{x}_n, y_n, (f_q(\mathbf{x}_n))_{q\in[Q]}\}_{n\in[N]}\right) + \Omega\left((\|f_q\|_{\mathcal{H}_k})_{q\in[Q]}\right),$$

Task: general

$$(\bar{\mathbf{f}}, \bar{\mathbf{b}}) = \underset{\substack{\mathbf{f}=(f_q)_{q\in[Q]} \in (\mathcal{H}_k)^Q, \\ \mathbf{b}=(b_q)_{q\in[Q]} \in \mathcal{B}, \\ (\mathbf{f},\mathbf{b}) \in C}}{\arg\min} \mathcal{L}(\mathbf{f}, \boldsymbol{b}),$$

$$\mathcal{L}(\mathbf{f}, \boldsymbol{b}) = L\left(\mathbf{b}, \{\mathbf{x}_n, y_n, (f_q(\mathbf{x}_n))_{q\in[Q]}\}_{n\in[N]}\right) + \Omega\left((\|f_q\|_{\mathcal{H}_k})_{q\in[Q]}\right),$$

$$C = \{(\mathbf{f}, \mathbf{b}) \,|\, (\mathbf{b}_0 - \mathbf{U}\mathbf{b})_i \leq D_i(\mathbf{W}\mathbf{f} - \mathbf{f}_0)_i(\mathbf{x}), \quad \forall \mathbf{x} \in K_i, \forall i \in [I]\},$$

Task: general

$$(\bar{\mathbf{f}}, \bar{\mathbf{b}}) = \underset{\substack{\mathbf{f}=(f_q)_{q\in[Q]} \in (\mathcal{H}_k)^Q, \\ \mathbf{b}=(b_q)_{q\in[Q]} \in \mathcal{B}, \\ (\mathbf{f},\mathbf{b}) \in C}}{\arg\min} \mathcal{L}(\mathbf{f}, \boldsymbol{b}),$$

$$\mathcal{L}(\mathbf{f}, \boldsymbol{b}) = L\left(\mathbf{b}, \{\mathbf{x}_n, y_n, (f_q(\mathbf{x}_n))_{q\in[Q]}\}_{n\in[N]}\right) + \Omega\left((\|f_q\|_{\mathcal{H}_k})_{q\in[Q]}\right),$$

$$C = \{(\mathbf{f}, \mathbf{b}) \,|\, (\mathbf{b}_0 - \mathbf{U}\mathbf{b})_i \leq D_i(\mathbf{W}\mathbf{f} - \mathbf{f}_0)_i(\mathbf{x}), \quad \forall \mathbf{x} \in K_i, \forall i \in [I]\},$$

$$(\mathbf{W}\mathbf{f})_i = \sum_{q\in[Q]} W_{i,q} f_q,$$

Task: general

$$(\bar{\mathbf{f}}, \bar{\mathbf{b}}) = \underset{\substack{\mathbf{f}=(f_q)_{q\in[Q]} \in (\mathcal{H}_k)^Q, \\ \mathbf{b}=(b_q)_{q\in[Q]} \in \mathcal{B}, \\ (\mathbf{f},\mathbf{b}) \in C}}{\arg\min} \mathcal{L}(\mathbf{f}, \boldsymbol{b}),$$

$$\mathcal{L}(\mathbf{f}, \boldsymbol{b}) = L\left(\mathbf{b}, \{\mathbf{x}_n, y_n, (f_q(\mathbf{x}_n))_{q\in[Q]}\}_{n\in[N]}\right) + \Omega\left((\|f_q\|_{\mathcal{H}_k})_{q\in[Q]}\right),$$

$$C = \{(\mathbf{f}, \mathbf{b}) \,|\, (\mathbf{b}_0 - \mathbf{U}\mathbf{b})_i \leq D_i(\mathbf{W}\mathbf{f} - \mathbf{f}_0)_i(\mathbf{x}), \quad \forall \mathbf{x} \in K_i, \forall i \in [I]\},$$

$$(\mathbf{W}\mathbf{f})_i = \sum_{q\in[Q]} W_{i,q} f_q,$$

$$D_i = \sum_{j\in[n_{i,j}]} \gamma_{i,j} \partial^{\mathbf{r}_{i,j}}, \; |\mathbf{r}_{i,j}| \leq s, \; \gamma_{i,j} \in \mathbb{R}, \; \partial^{\mathbf{r}} f(\mathbf{x}) = \frac{\partial^{\sum_{j=1}^d r_j} f(\mathbf{x})}{\partial_{x_1}^{r_1} \cdots \partial_{x_d}^{r_d}}.$$

Blanket assumptions

1. Domain $\mathcal{X} \subseteq \mathbb{R}^d$: open. Kernel $k \in \mathcal{C}^s(\mathcal{X} \times \mathcal{X})$.
2. $K_i \subset \mathcal{X}$: compact, $\forall i$.
3. Bias domain $\mathcal{B} \subseteq \mathbb{R}^Q$: convex.
4. Loss L restricted to $\mathcal{B}$: strictly convex in $\mathbf{b}$.
5. Regularizer Ω: strictly increasing in each of its argument.

Our strenghtened SOC-constrained formulation

$$(\mathbf{f}_{\boldsymbol{\eta}}, \mathbf{b}_{\boldsymbol{\eta}}) = \underset{\mathbf{f} \in (\mathcal{H}_k)^Q,\, \boldsymbol{b} \in \mathcal{B}}{\arg\min} \quad \mathcal{L}(\mathbf{f}, \boldsymbol{b}) \tag{$\mathcal{P}_{\boldsymbol{\eta}}$}$$

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

$$\begin{aligned} &(\mathbf{b}_0 - \mathbf{U}\mathbf{b})_i + \eta_i \|(\mathbf{W}\mathbf{f} - \mathbf{f}_0)_i\|_{\mathcal{H}_k} \\ &\leq \min_{m \in [M_i]} D_i(\mathbf{W}\mathbf{f} - \mathbf{f}_0)_i \left(\tilde{\mathbf{x}}_{i,m}\right), \ \forall i \in [I], \end{aligned} \tag{$\mathcal{C}_{\boldsymbol{\eta}}$}$$

where

- $\{\tilde{\mathbf{x}}_{i,m}\}_{m \in [M_i]}$: a δ_i-net of K_i in $\|\cdot\|_{\mathcal{X}}$,
- $\eta_i = \sup_{m \in [M_i], \mathbf{u} \in \mathbb{B}_{\|\cdot\|_{\mathcal{X}}}(\mathbf{0},1)} \|D_{i,\mathbf{x}} k(\tilde{\mathbf{x}}_{i,m}, \cdot) - D_{i,\mathbf{x}} k(\tilde{\mathbf{x}}_{i,m} + \delta_i \mathbf{u}, \cdot)\|_{\mathcal{H}_k}$.

Theorem

- Minimal values: v_{disc} = value of $(\mathcal{P}_{\boldsymbol{\eta}})$ with '$\boldsymbol{\eta} = \mathbf{0}$', $\bar{v} = \mathcal{L}\left(\bar{\mathbf{f}}, \bar{\boldsymbol{b}}\right)$, $v_{\boldsymbol{\eta}} = \mathcal{L}\left(\mathbf{f}_{\boldsymbol{\eta}}, \mathbf{b}_{\boldsymbol{\eta}}\right)$.
- Let $\mathbf{f}_{\boldsymbol{\eta}} = \left(f_{\boldsymbol{\eta},q}\right)_{q\in[Q]}$.

Theorem

- Minimal values: v_{disc} = value of $(\mathcal{P}_{\boldsymbol{\eta}})$ with '$\boldsymbol{\eta} = \mathbf{0}$', $\bar{v} = \mathcal{L}\left(\bar{\mathbf{f}}, \bar{\boldsymbol{b}}\right)$, $v_{\boldsymbol{\eta}} = \mathcal{L}\left(\mathbf{f}_{\boldsymbol{\eta}}, \mathbf{b}_{\boldsymbol{\eta}}\right)$.
- Let $\mathbf{f}_{\boldsymbol{\eta}} = \left(f_{\boldsymbol{\eta},q}\right)_{q\in[Q]}$.

Then,

- (i) Tightening: any $(\mathbf{f}, \boldsymbol{b})$ satisfying $(\mathcal{C}_{\boldsymbol{\eta}})$ also satisfies $(\mathcal{C})$, hence

$$v_{\text{disc}} \leq \bar{v} \leq v_{\boldsymbol{\eta}}.$$

Theorem

- Minimal values: v_{disc} = value of $(\mathcal{P}_{\boldsymbol{\eta}})$ with '$\boldsymbol{\eta} = \mathbf{0}$', $\bar{v} = \mathcal{L}\left(\bar{\mathbf{f}}, \bar{\boldsymbol{b}}\right)$, $v_{\boldsymbol{\eta}} = \mathcal{L}\left(\mathbf{f}_{\boldsymbol{\eta}}, \mathbf{b}_{\boldsymbol{\eta}}\right)$.
- Let $\mathbf{f}_{\boldsymbol{\eta}} = \left(f_{\boldsymbol{\eta},q}\right)_{q\in[Q]}$.

Then,

- (i) Tightening: any $(\mathbf{f}, \boldsymbol{b})$ satisfying $(\mathcal{C}_{\boldsymbol{\eta}})$ also satisfies $(\mathcal{C})$, hence

$$v_{\text{disc}} \leq \bar{v} \leq v_{\boldsymbol{\eta}}.$$

- (ii) Representer theorem: For $\forall q \in [Q]$, $\exists \tilde{a}_{i,0,q}, \tilde{a}_{i,m,q}, a_{n,q} \in \mathbb{R}$ s.t.

$$f_{\boldsymbol{\eta},q} = \sum_{i\in[I]} \left[\tilde{a}_{i,0,q} f_{0,i} + \sum_{m\in[M_i]} \tilde{a}_{i,m,q} D_{i,\mathbf{x}} k\left(\tilde{\mathbf{x}}_{i,m}, \cdot\right) \right] + \sum_{n\in[N]} a_{n,q} k(\mathbf{x}_n, \cdot).$$

Theorem – continued

- (iii) Performance guarantee: if $\mathcal{L}$ is $(\mu_{f_q}, \mu_{\boldsymbol{b}})$-strongly convex w.r.t. $(f_q, \mathbf{b})$ for any $q \in [Q]$, then

$$\|f_{\boldsymbol{\eta},q} - \bar{f}_q\|_{\mathcal{H}_k} \leq \sqrt{\frac{2(v_{\boldsymbol{\eta}} - v_{\text{disc}})}{\mu_{f_q}}}, \quad \|\boldsymbol{b}_{\boldsymbol{\eta}} - \bar{\boldsymbol{b}}\|_2 \leq \sqrt{\frac{2(v_{\boldsymbol{\eta}} - v_{\text{disc}})}{\mu_{\boldsymbol{b}}}}.$$

Theorem – continued

- (iii) Performance guarantee: if $\mathcal{L}$ is $(\mu_{f_q}, \mu_{\boldsymbol{b}})$-strongly convex w.r.t. $(f_q, \mathbf{b})$ for any $q \in [Q]$, then

$$\|f_{\boldsymbol{\eta},q} - \bar{f}_q\|_{\mathcal{H}_k} \leq \sqrt{\frac{2(v_{\boldsymbol{\eta}} - v_{\text{disc}})}{\mu_{f_q}}}, \quad \|\boldsymbol{b}_{\boldsymbol{\eta}} - \bar{\boldsymbol{b}}\|_2 \leq \sqrt{\frac{2(v_{\boldsymbol{\eta}} - v_{\text{disc}})}{\mu_{\boldsymbol{b}}}}.$$

If in addition $\mathbf{U}$ is surjective, $\mathcal{B} = \mathbb{R}^Q$, and $\mathcal{L}(\bar{\mathbf{f}}, \cdot)$ is L_b-Lipschitz continuous on $\mathbb{B}_{\|\cdot\|_2}\left(\bar{\mathbf{b}}, c_f \|\boldsymbol{\eta}\|_\infty\right)$ where $c_f = \sqrt{d} \left\| \left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T \right\| \max_{i\in[I]} \left\|(\mathbf{W}\bar{\mathbf{f}} - \mathbf{f}_0)_i\right\|_{\mathcal{H}_k}$, then

$$\|f_{\boldsymbol{\eta},q} - \bar{f}_q\|_{\mathcal{H}_k} \leq \sqrt{\frac{2L_b c_f \|\boldsymbol{\eta}\|_\infty}{\mu_{f_q}}}, \|\boldsymbol{b}_{\boldsymbol{\eta}} - \bar{\boldsymbol{b}}\|_2 \leq \sqrt{\frac{2L_b c_f \|\boldsymbol{\eta}\|_\infty}{\mu_{\boldsymbol{b}}}}.$$

Theorem – continued

- (iii) Performance guarantee: if $\mathcal{L}$ is $(\mu_{f_q}, \mu_{\boldsymbol{b}})$-strongly convex w.r.t. $(f_q, \mathbf{b})$ for any $q \in [Q]$, then

$$\|f_{\boldsymbol{\eta},q} - \bar{f}_q\|_{\mathcal{H}_k} \leq \sqrt{\frac{2(v_{\boldsymbol{\eta}} - v_{\text{disc}})}{\mu_{f_q}}}, \quad \|\boldsymbol{b}_{\boldsymbol{\eta}} - \bar{\boldsymbol{b}}\|_2 \leq \sqrt{\frac{2(v_{\boldsymbol{\eta}} - v_{\text{disc}})}{\mu_{\boldsymbol{b}}}}.$$

If in addition $\mathbf{U}$ is surjective, $\mathcal{B} = \mathbb{R}^Q$, and $\mathcal{L}(\bar{\mathbf{f}}, \cdot)$ is L_b-Lipschitz continuous on $\mathbb{B}_{\|\cdot\|_2}\left(\bar{\mathbf{b}}, c_f\|\boldsymbol{\eta}\|_\infty\right)$ where $c_f = \sqrt{d}\left\|\left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\right\| \max_{i\in[I]}\left\|(\mathbf{W}\bar{\mathbf{f}} - \mathbf{f}_0)_i\right\|_{\mathcal{H}_k}$, then

$$\|f_{\boldsymbol{\eta},q} - \bar{f}_q\|_{\mathcal{H}_k} \leq \sqrt{\frac{2L_b c_f\|\boldsymbol{\eta}\|_\infty}{\mu_{f_q}}}, \|\boldsymbol{b}_{\boldsymbol{\eta}} - \bar{\boldsymbol{b}}\|_2 \leq \sqrt{\frac{2L_b c_f\|\boldsymbol{\eta}\|_\infty}{\mu_{\boldsymbol{b}}}}.$$

1st bound: computable. 2nd: Larger $M_i \Rightarrow$ smaller $\delta_i \Rightarrow$ smaller η_i $\Rightarrow$ tighter bound.

Tightening idea

Let $s = 0$, $l = 1$. Recall constraint $(\mathcal{C})$:

$$\left\{(\mathbf{f}, \mathbf{b}) \mid \underbrace{(b_0 - \mathbf{Ub})}_{\beta} \leq \underbrace{\underbrace{(\mathbf{Wf} - f_0)}_{\phi}(\mathbf{x})}_{\langle \phi, k(\mathbf{x}, \cdot) \rangle_{\mathcal{H}_k}}, \quad \forall \mathbf{x} \in K \right\}$$

Tightening idea

Let $s = 0$, $l = 1$. Recall constraint $(\mathcal{C})$:

$$\left\{(\mathbf{f}, \mathbf{b}) \mid \underbrace{(b_0 - \mathbf{Ub})}_{\beta} \leq \underbrace{\underbrace{(\mathbf{Wf} - f_0)}_{\phi}(\mathbf{x})}_{\langle \phi, k(\mathbf{x}, \cdot) \rangle_{\mathcal{H}_k}}, \quad \forall \mathbf{x} \in K \right\}, \text{i.e.}$$

$$\Phi(K) := \{k(\mathbf{x}, \cdot) : \mathbf{x} \in K\} \subseteq H^+_{\phi,\beta} := \{g \in \mathcal{H}_k \mid \beta \leq \langle \phi, g \rangle_{\mathcal{H}_k}\}$$

Tightening idea

Let $s = 0$, $l = 1$. Recall constraint $(\mathcal{C})$:

$$\left\{(\mathbf{f}, \mathbf{b}) \mid \underbrace{(b_0 - \mathbf{Ub})}_{\beta} \leq \underbrace{\underbrace{(\mathbf{Wf} - f_0)}_{\phi}(\mathbf{x})}_{\langle \phi, k(\mathbf{x}, \cdot) \rangle_{\mathcal{H}_k}}, \quad \forall \mathbf{x} \in K \right\}, \text{i.e.}$$

$$\Phi(K) := \{k(\mathbf{x}, \cdot) : \mathbf{x} \in K\} \subseteq H^+_{\phi,\beta} := \{g \in \mathcal{H}_k \mid \beta \leq \langle \phi, g \rangle_{\mathcal{H}_k}\}$$

- $(\mathcal{C}_\eta)$ means: covering of $\Phi(K)$ by balls with η-radius centered at the $k(\tilde{\mathbf{x}}_m, \cdot)$ is in the halfspace $H^+_{\phi,\beta}$; hence it is tightening.

Tightening idea

Let $s = 0$, $l = 1$. Recall constraint $(\mathcal{C})$:

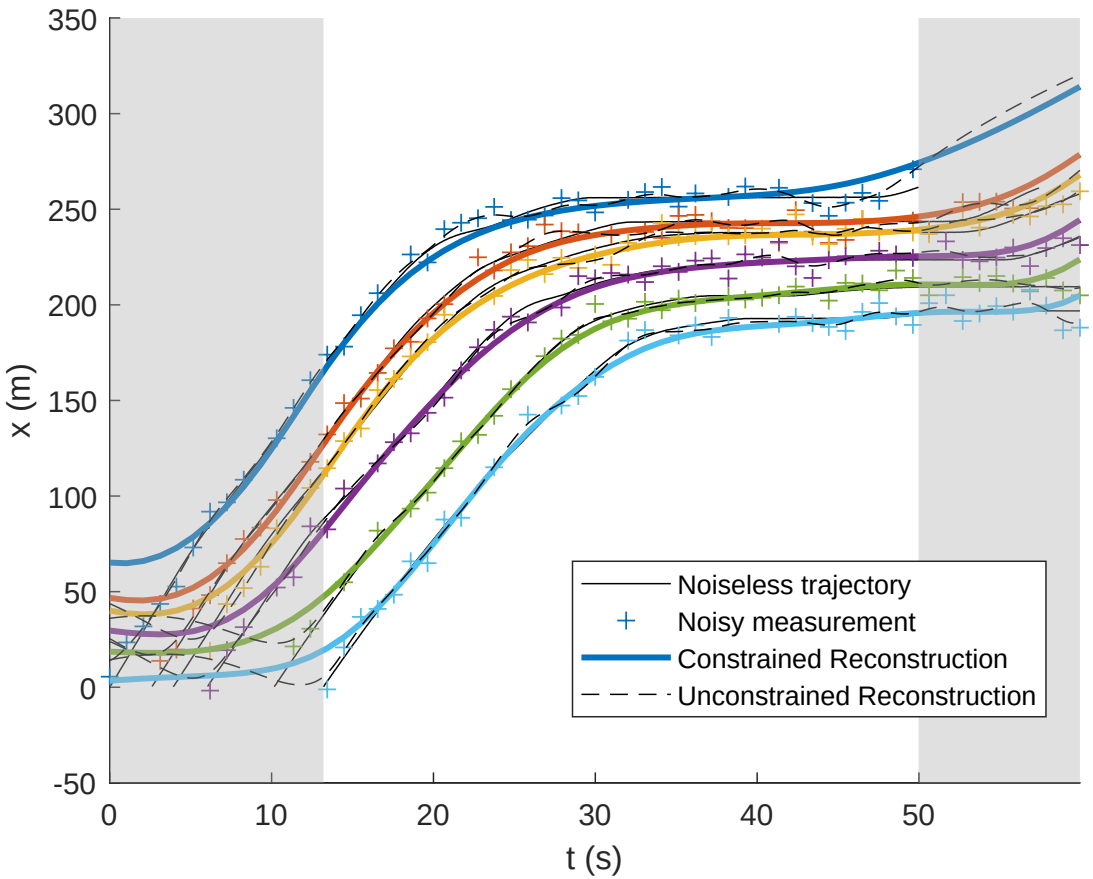
$$\{(\mathbf{f}, \mathbf{b}) \mid \underbrace{(b_0 - \mathbf{Ub})}_{\beta} \leq \underbrace{\underbrace{(\mathbf{Wf} - f_0)}_{\phi}(\mathbf{x})}_{\langle \phi, k(\mathbf{x}, \cdot) \rangle_{\mathcal{H}_k}}, \quad \forall \mathbf{x} \in K\}, \text{ i.e.}$$

$$\Phi(K) := \{k(\mathbf{x}, \cdot) : \mathbf{x} \in K\} \subseteq H^+_{\phi,\beta} := \{g \in \mathcal{H}_k \mid \beta \leq \langle \phi, g \rangle_{\mathcal{H}_k}\}$$

- $(\mathcal{C}_\eta)$ means: covering of $\Phi(K)$ by balls with η-radius centered at the $k(\tilde{\mathbf{x}}_m, \cdot)$ is in the halfspace $H^+_{\phi,\beta}$; hence it is tightening.
- η is obtained as the minimal radius.

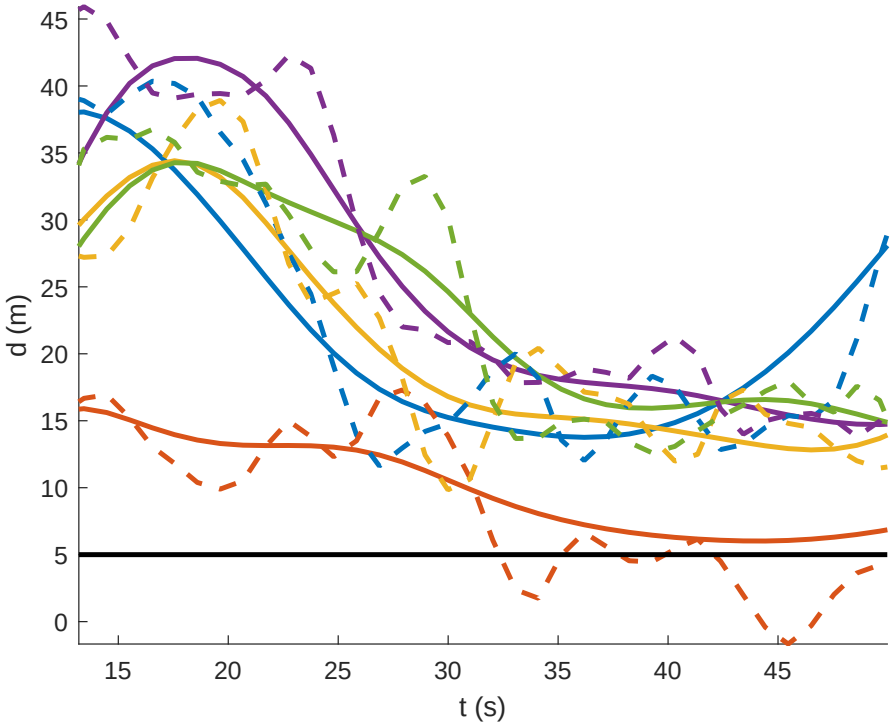
Demo: task-1 = convoy localization with traffic jam

Setting: $q = 6,\ d_{\min} = 5m,\ v_{\min} = 0.$



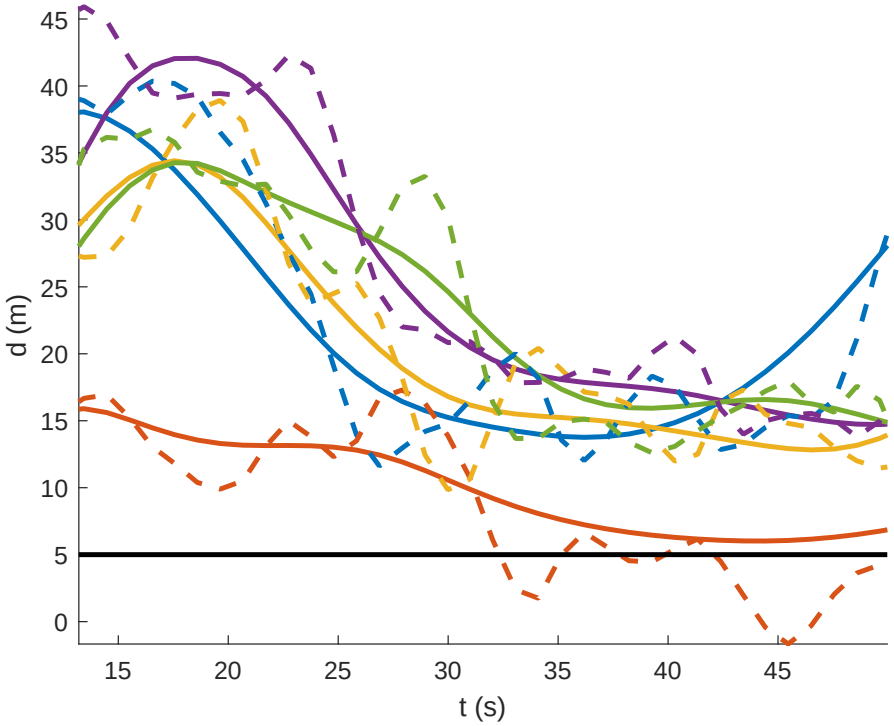
Convoy localization: continued

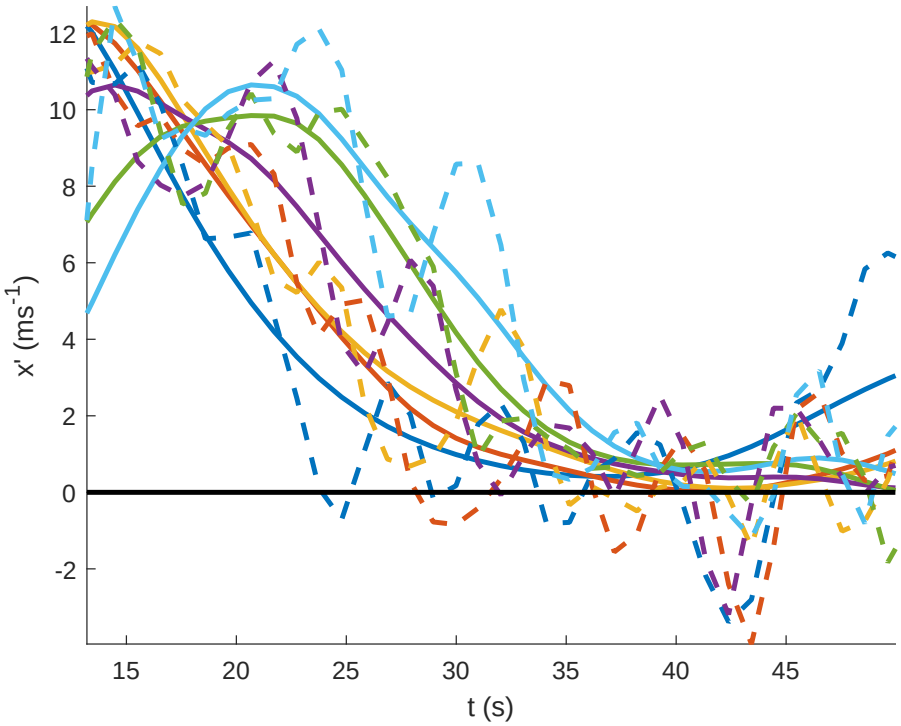
Pairwise distances: $t \mapsto f_q(t) - f_{q+1}(t)$



Convoy localization: continued

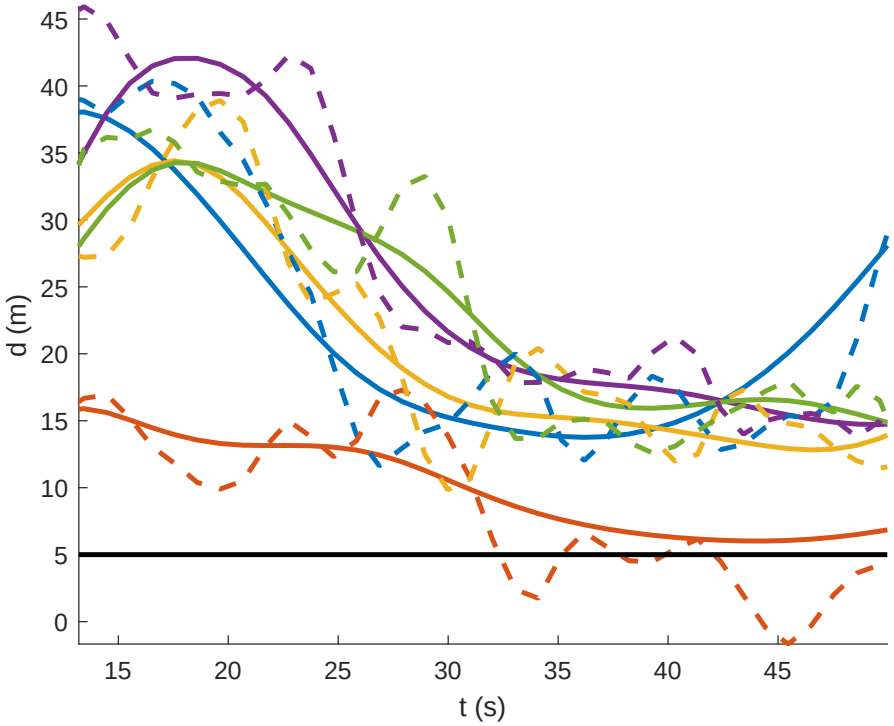
Pairwise distances: $t \mapsto f_q(t) - f_{q+1}(t)$ Speed: $t \mapsto f_q'(t)$

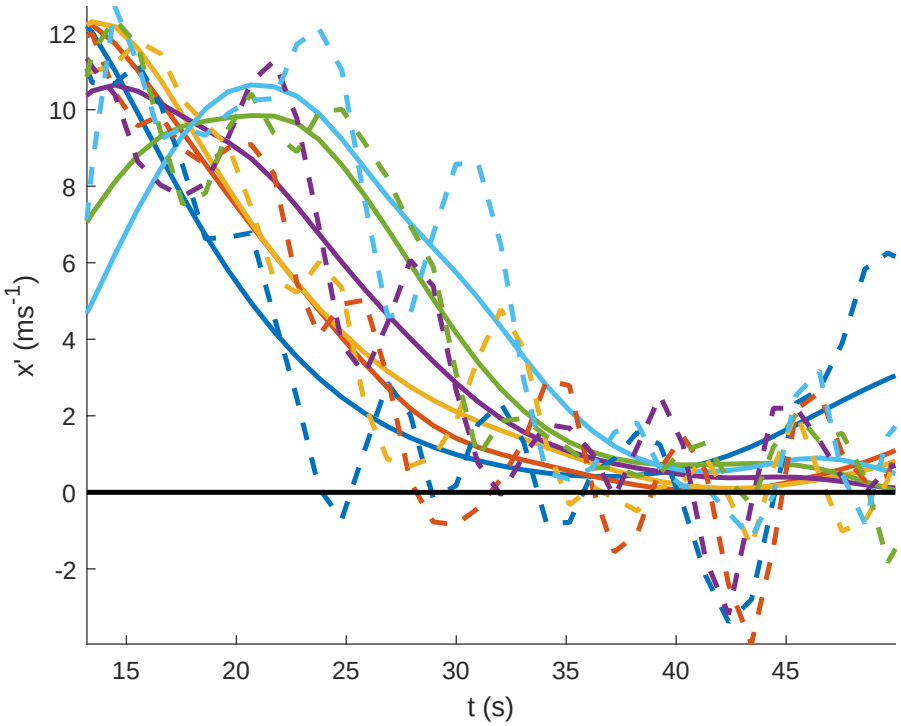




Convoy localization: continued

Pairwise distances: $t \mapsto f_q(t) - f_{q+1}(t)$ Speed: $t \mapsto f_q'(t)$



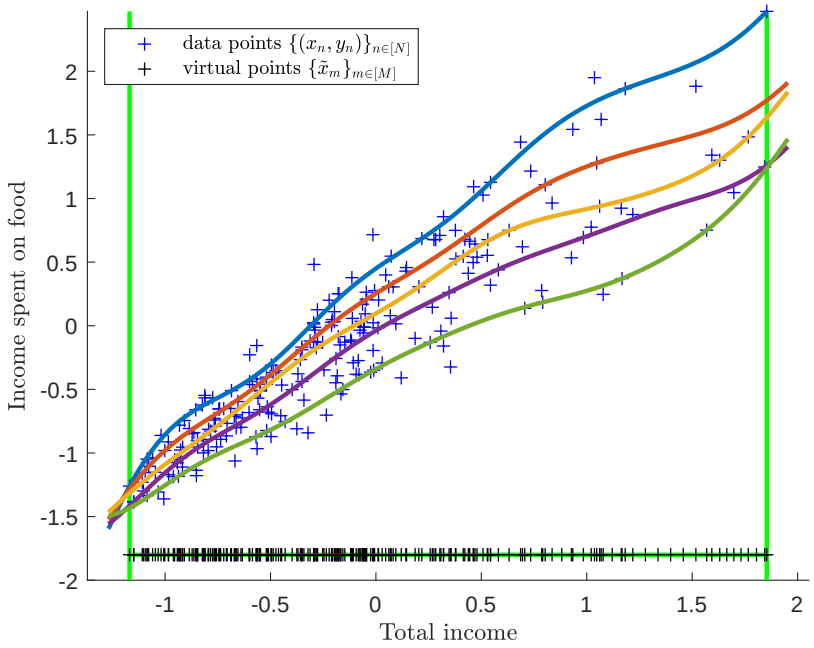


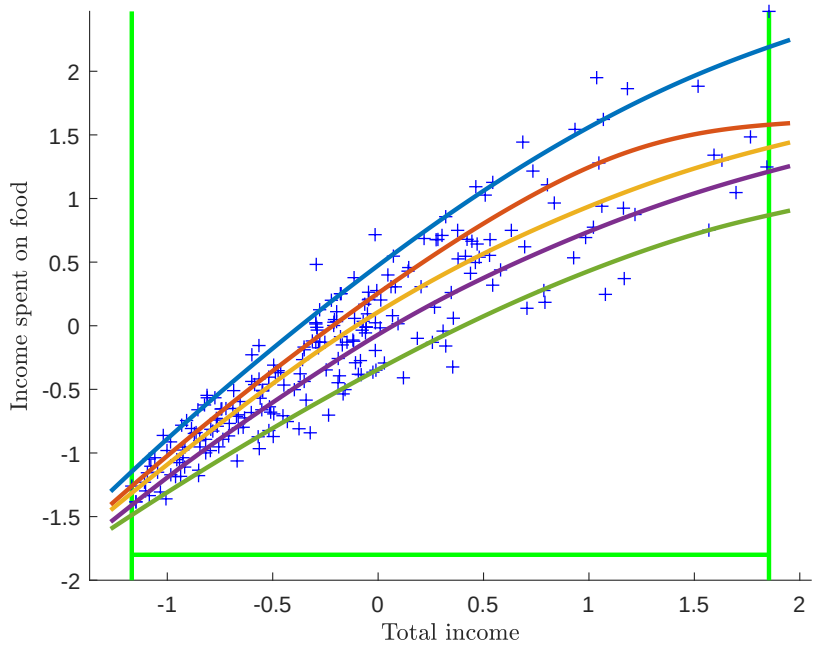
Shape constraints: especially relevant in noisy situations.

Demo: task-2 = joint quantile regression

Economics:

- x: annual household income, y: food expenditure. $d = 1$, $N = 235$.
- Engel's law $\Rightarrow \nearrow$, concave.
- Demo: $\tau_q \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.
- Left: non-crossing, $\nearrow$. Right: non-crossing, $\nearrow$, concave.

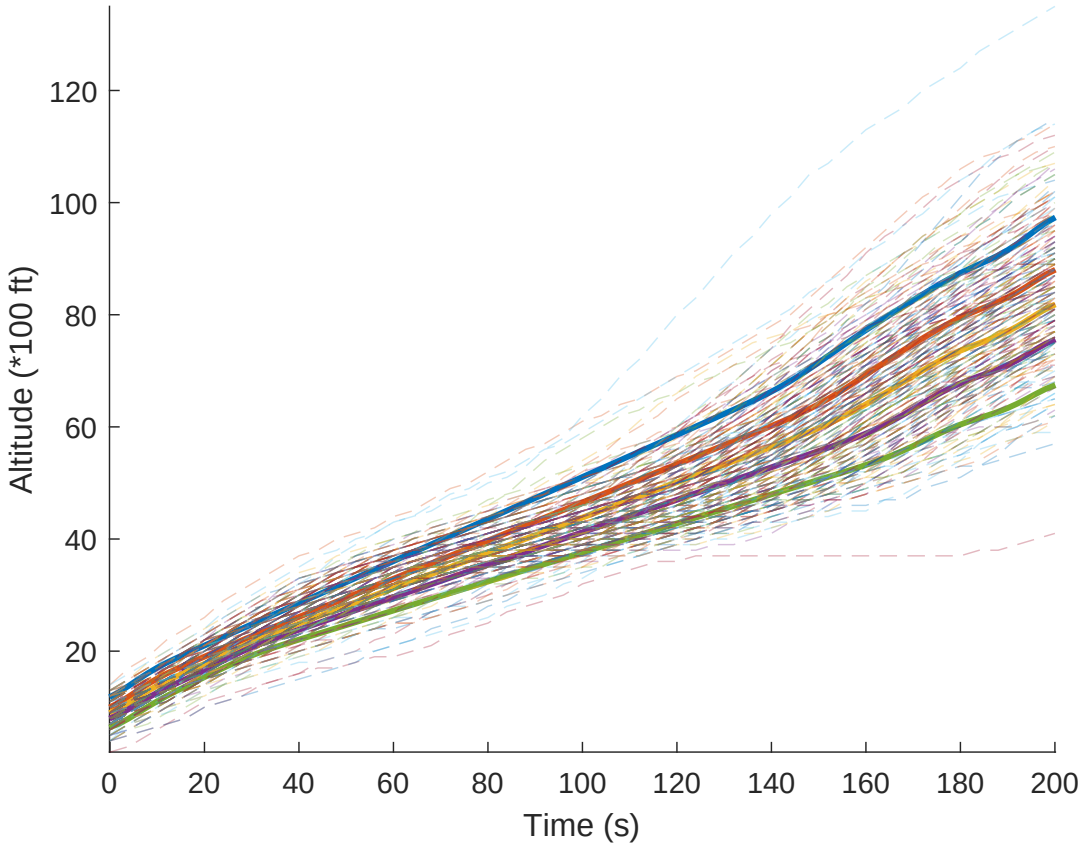




Demo: task-2 = joint quantile regression

Analysis of aircraft trajectories, ENAC:

- y: radar-measured altitude of aircrafts flying between two cities (Paris & Toulouse); x: time. $d = 1$, $N = 15657$.
- Demo: $\tau_q \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.
- Constraint: non-crossing, $\nearrow$ (takeoff).



Summary

- Focus: shape constraints in RKHSs.
- Contribution: tightened SOC-constrained reformulation.
- Application:
  - convoy localization,
  - joint quantile regression: economics, aircraft trajectories.

Summary

- Focus: shape constraints in RKHSs.
- Contribution: tightened SOC-constrained reformulation.
- Application:
  - convoy localization,
  - joint quantile regression: economics, aircraft trajectories.
- Dissemination:
  - [Aubin-Frankowski and Szabó, 2020] (ICML-2020, submitted),
  - [Aubin-Frankowski et al., 2020] (IFAC WC-2020, accepted).

Summary

- Focus: shape constraints in RKHSs.
- Contribution: tightened SOC-constrained reformulation.
- Application:
  - convoy localization,
  - joint quantile regression: economics, aircraft trajectories.
- Dissemination:
  - [Aubin-Frankowski and Szabó, 2020] (ICML-2020, submitted),
  - [Aubin-Frankowski et al., 2020] (IFAC WC-2020, accepted).

Supplement

- Joint quantile regression on 9 UCI benchmarks.
- Shape constraints: applications areas.
- Partial ordering.

Joint quantile regression: 9 UCI testbeds

- PDCD = Primal-Dual Coordinate Descent [Sangnier et al., 2016].
- 4-5th columns: mean $\pm$ std of 100$\times$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

Shape constraints in applications

- Finance:
  - European and American call option prices: convex & monotone in the underlying stock price and ↗ in volatility [Aït-Sahalia and Duarte, 2003].

Shape constraints in applications

- Finance:
  - European and American call option prices: convex & monotone in the underlying stock price and ↗ in volatility [Aït-Sahalia and Duarte, 2003].
- Statistics: quantile function ↗ w.r.t. the quantile level.

Shape constraints in applications

- Finance:
  - European and American call option prices: convex & monotone in the underlying stock price and ↗ in volatility [Aït-Sahalia and Duarte, 2003].
- Statistics: quantile function ↗ w.r.t. the quantile level.
- RL and stochastic optimization: value functions are often convex [Keshavarz et al., 2011, Shapiro et al., 2014].

Shape constraints in applications

- Finance:
  - European and American call option prices: convex & monotone in the underlying stock price and ↗ in volatility [Aït-Sahalia and Duarte, 2003].
- Statistics: quantile function ↗ w.r.t. the quantile level.
- RL and stochastic optimization: value functions are often convex [Keshavarz et al., 2011, Shapiro et al., 2014].
- Biology (monotone regression): identify genome interactions [Luss et al., 2012], dose-response studies [Hu et al., 2005].

Shape constraints in applications+

- Economics:
  - utility functions are  and concave [Matzkin, 1991].

Shape constraints in applications+

- Economics:
  - utility functions are ↗ and concave [Matzkin, 1991].
  - demand functions of normal goods are downward sloping [Lewbel, 2010, Blundell et al., 2012],

Shape constraints in applications+

- Economics:
  - utility functions are ↗ and concave [Matzkin, 1991].
  - demand functions of normal goods are downward sloping [Lewbel, 2010, Blundell et al., 2012],
  - production functions are concave [Varian, 1984] or S-shaped [Yagi et al., 2020].

Shape constraints in applications+

- Economics:
  - utility functions are ↗ and concave [Matzkin, 1991].
  - demand functions of normal goods are downward sloping [Lewbel, 2010, Blundell et al., 2012],
  - production functions are concave [Varian, 1984] or S-shaped [Yagi et al., 2020].
  - panel multinomial choice problems [Shi et al., 2018]: cyclic monotonicity,

Shape constraints in applications+

- Economics:
  - utility functions are $\nearrow$ and concave [Matzkin, 1991].
  - demand functions of normal goods are downward sloping [Lewbel, 2010, Blundell et al., 2012],
  - production functions are concave [Varian, 1984] or S-shaped [Yagi et al., 2020].
  - panel multinomial choice problems [Shi et al., 2018]: cyclic monotonicity,
  - single index model: most link functions are monotone [Li and Racine, 2007, Chen and Samworth, 2016, Balabdaoui et al., 2019].

Shape constraints in applications+

- Economics:
  - utility functions are $\nearrow$ and concave [Matzkin, 1991].
  - demand functions of normal goods are downward sloping [Lewbel, 2010, Blundell et al., 2012],
  - production functions are concave [Varian, 1984] or S-shaped [Yagi et al., 2020].
  - panel multinomial choice problems [Shi et al., 2018]: cyclic monotonicity,
  - single index model: most link functions are monotone [Li and Racine, 2007, Chen and Samworth, 2016, Balabdaoui et al., 2019].
- Supply chain models, stochastic multi-period inventory problems, pricing models and game theory: supermodularity [Topkis, 1998, Simchi-Levi et al., 2014].

Partial ordering

$(\mathcal{X}, \leq)$ is a partial ordering if

1. reflexity: $a \leq a$ for $\forall a \in \mathcal{X}$,
2. antisymmetry: $a \leq b$ and $b \leq a$ imply $a = b$, and
3. transitivity: if $a \leq b$ and $b \leq c$ imply $a \leq c$

hold.


Aït-Sahalia, Y. and Duarte, J. (2003).
Nonparametric option pricing under shape restrictions.
Journal of Econometrics, 116(1-2):9–47.


Aubin-Frankowski, P.-C., Petit, N., and Szabó, Z. (2020).
Kernel regression for vehicle trajectory reconstruction under speed and inter-vehicular distance constraints.
In *IFAC World Congress (IFAC WC)*, Berlin, Germany. (accepted).


Aubin-Frankowski, P.-C. and Szabó, Z. (2020).
Hard shape-constrained kernel machines.
In *International Conference on Machine Learning (ICML)*. (submitted).


Bai, L., Cui, L., Rossi, L., Xu, L., Bai, X., and Hancock, E. (2018).
Local-global nested graph kernels using nested complexity traces.
Pattern Recognition Letters.

(in press; available at https://doi.org/10.1016/j.patrec.2018.06.016).


Balabdaoui, F., Durot, C., and Jankowski, H. (2019).
Least squares estimation in the monotone single index model.
Bernoulli, 25(4B):3276–3310.


Blundell, R., Horowitz, J. L., and Parey, M. (2012).
Measuring the price responsiveness of gasoline demand: economic shape restrictions and nonparametric demand estimation.
Quantitative Economics, 3:29–51.


Borgwardt, K. M. and Kriegel, H.-P. (2005).
Shortest-path kernels on graphs.
In *International Conference on Data Mining (ICDM)*, pages 74–81.


Chen, Y. and Samworth, R. J. (2016).
Generalized additive and index models with shape constraints.

Journal of the Royal Statistical Society – Statistical Methodology, Series B, 78(4):729–754.


Collins, M. and Duffy, N. (2001).
Convolution kernels for natural language.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 625–632.


Cuturi, M. (2011).
Fast global alignment kernels.
In *International Conference on Machine Learning (ICML)*, pages 929–936.


Cuturi, M., Fukumizu, K., and Vert, J.-P. (2005).
Semigroup kernels on measures.
Journal of Machine Learning Research, 6:1169–1198.


Cuturi, M. and Vert, J.-P. (2005).
The context-tree kernel for strings.
Neural Networks, 18(8):1111–1123.


Cuturi, M., Vert, J.-P., Birkenes, O., and Matsui, T. (2007).
A kernel for time series based on global alignments.
In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 413–416.


Gärtner, T., Flach, P., Kowalczyk, A., and Smola, A. (2002).
Multi-instance kernels.
In *International Conference on Machine Learning (ICML)*, pages 179–186.


Gärtner, T., Flach, P., and Wrobel, S. (2003).
On graph kernels: Hardness results and efficient alternatives.
Learning Theory and Kernel Machines, pages 129–143.


Guevara, J., Hirata, R., and Canu, S. (2017).
Cross product kernels for fuzzy set similarity.
In *International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6.


Haussler, D. (1999).

Convolution kernels on discrete structures.
Technical report, University of California at Santa Cruz.
(http://cbse.soe.ucsc.edu/sites/default/files/convolutions.pdf).


Hu, J., Kapoor, M., Zhang, W., Hamilton, S. R., and Coombes, K. R. (2005).
Analysis of dose-response effects on gene expression data with comparison of two microarray platforms.
Bioinformatics, 21(17):3524–3529.


Jaakkola, T. S. and Haussler, D. (1999).
Exploiting generative models in discriminative classifiers.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 487–493.


Jebara, T., Kondor, R., and Howard, A. (2004).
Probability product kernels.
Journal of Machine Learning Research, 5:819–844.


Jiao, Y. and Vert, J.-P. (2016).

The Kendall and Mallows kernels for permutations.
In *International Conference on Machine Learning (ICML)*, volume 37, pages 2982–2990.


Kashima, H. and Koyanagi, T. (2002).
Kernels for semi-structured data.
In *International Conference on Machine Learning (ICML)*, pages 291–298.


Kashima, H., Tsuda, K., and Inokuchi, A. (2003).
Marginalized kernels between labeled graphs.
In *International Conference on Machine Learning (ICML)*, pages 321–328.


Keshavarz, A., Wang, Y., and Boyd, S. (2011).
Imputing a convex objective function.
In *IEEE Multi-Conference on Systems and Control*, pages 613–619.


Király, F. J. and Oberhauser, H. (2019).
Kernels for sequentially ordered data.

Journal of Machine Learning Research, 20:1–45.


Kondor, R. and Pan, H. (2016).
The multiscale Laplacian graph kernel.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 2982–2990.


Kondor, R. I. and Lafferty, J. (2002).
Diffusion kernels on graphs and other discrete input.
In *International Conference on Machine Learning (ICML)*, pages 315–322.


Kuang, R., Ie, E., Wang, K., Wang, K., Siddiqi, M., Freund, Y., and Leslie, C. (2004).
Profile-based string kernels for remote homology detection and motif extraction.
Journal of Bioinformatics and Computational Biology, 13(4):527–550.


Leslie, C., Eskin, E., and Noble, W. S. (2002).

The spectrum kernel: A string kernel for SVM protein classification.
Biocomputing, pages 564–575.


Leslie, C. and Kuang, R. (2004).
Fast string kernels using inexact matching for protein sequences.
Journal of Machine Learning Research, 5:1435–1455.


Lewbel, A. (2010).
Shape-invariant demand functions.
The Review of Economics and Statistics, 92(3):549–556.


Li, Q. and Racine, J. S. (2007).
Nonparametric Econometrics.
Princeton University Press.


Lodhi, H., Saunders, C., Shawe-Taylor, J., Cristianini, N., and Watkins, C. (2002).
Text classification using string kernels.
Journal of Machine Learning Research, 2:419–444.


Luss, R., Rossett, S., and Shahar, M. (2012).
Efficient regularized isotonic regression with application to gene-gene interaction search.
Annals of Applied Statistics, 6(1):253–283.


Matzkin, R. L. (1991).
Semiparametric estimation of monotone and concave utility functions for polychotomous choice models.
Econometrica, 59(5):1315–1327.


Rüping, S. (2001).
SVM kernels for time series analysis.
Technical report, University of Dortmund.
(`http://www.stefan-rueping.de/publications/rueping-2001-a.pdf`).


Saigo, H., Vert, J.-P., Ueda, N., and Akutsu, T. (2004).
Protein homology detection using string alignment kernels.
Bioinformatics, 20(11):1682–1689.


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.


Seeger, M. (2002).
Covariance kernels from Bayesian generative models.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 905–912.


Shapiro, A., Dentcheva, D., and Ruszczynski, A. (2014).
Lectures on Stochastic Programming: Modeling and Theory.
SIAM - Society for Industrial and Applied Mathematics.


Shervashidze, N., Vishwanathan, S. V. N., Petri, T., Mehlhorn, K., and Borgwardt, K. M. (2009).
Efficient graphlet kernels for large graph comparison.
In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 488–495.


Shi, X., Shum, M., and Song, W. (2018).
Estimating semi-parametric panel multinomial choice models using cyclic monotonicity.
Econometrica, 86(2):737–761.


Simchi-Levi, D., Chen, X., and Bramel, J. (2014).
The Logic of Logistics: Theory, Algorithms, and Applications for Logistics Management.
Springer.


Topkis, D. M. (1998).
Supermodularity and complementarity.
Princeton University Press.


Tsuda, K., Kin, T., and Asai, K. (2002).
Marginalized kernels for biological sequences.
Bioinformatics, 18:268–275.


Varian, H. R. (1984).
The nonparametric approach to production analysis.

Econometrica, 52(3):579–597.


Vishwanathan, S. N., Schraudolph, N., Kondor, R., and Borgwardt, K. (2010).
Graph kernels.
Journal of Machine Learning Research, 11:1201–1242.


Watkins, C. (1999).
Dynamic alignment kernels.
In *Advances in Neural Information Processing Systems (NIPS)*, pages 39–50.


Yagi, D., Chen, Y., Johnson, A. L., and Kuosmanen, T. (2020).
Shape-constrained kernel-weighted least squares: Estimating production functions for Chilean manufacturing industries.
Journal of Business & Economic Statistics, 38(1):43–54.