When Kernel Machines Meet Shape Constraints

Zoltán Szabó

Joint work with: Pierre-Cyril Aubin-Frankowski
@ MINES ParisTech → INRIA

Gatsby Unit, UCL
Jan 26, 2022

Shape constraints

Pattern

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

Shape constraints

Pattern

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

Examples:

1. non-negativity: $\boxed{0 \leq f(x)}$,

Shape constraints

Pattern

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

Examples:

1. non-negativity: $\boxed{0 \leq f(x)}$,
2. monotonicity ($\nearrow$): $\boxed{0 \leq f'(x)}$,

Shape constraints

Pattern

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

Examples:

1. non-negativity: $\boxed{0 \leq f(x)}$,
2. monotonicity ($\nearrow$): $\boxed{0 \leq f'(x)}$,
3. convexity: $\boxed{0 \leq f''(x)}$,

Shape constraints

Pattern

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

Examples:

1. non-negativity: $\boxed{0 \leq f(x)}$,
2. monotonicity ($\nearrow$): $\boxed{0 \leq f'(x)}$,
3. convexity: $\boxed{0 \leq f''(x)}$,
4. n-monotonicity: $\boxed{0 \leq f^{(n)}(x)}$,

Shape constraints

Pattern

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

Examples:

1. non-negativity: $\boxed{0 \le f(x)}$,
2. monotonicity ($\nearrow$): $\boxed{0 \le f'(x)}$,
3. convexity: $\boxed{0 \le f''(x)}$,
4. n-monotonicity: $\boxed{0 \le f^{(n)}(x)}$,
5. $(n-1)$-alternating monotonicity: for $n \ge 2$

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

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

Examples continued

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).

Examples continued

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

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

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

$\mathbf{u} \preccurlyeq \mathbf{v}$ 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).

Examples continued

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

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

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

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

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

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

Shape constraints are omnipresent [Johnson and Jiang, 2018, Guntuboyina and Sen, 2018, Chetverikov et al., 2018]

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

Shape constraints are omnipresent [Johnson and Jiang, 2018, Guntuboyina and Sen, 2018, Chetverikov et al., 2018]

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

Shape constraints are omnipresent [Johnson and Jiang, 2018, Guntuboyina and Sen, 2018, Chetverikov et al., 2018]

- 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].

Shape constraints are omnipresent [Johnson and Jiang, 2018, Guntuboyina and Sen, 2018, Chetverikov et al., 2018]

- 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].
- Statistics: quantile function $\nearrow$ w.r.t. the quantile level, pdfs are non-negative and often log-concave.

Shape constraints are omnipresent [Johnson and Jiang, 2018, Guntuboyina and Sen, 2018, Chetverikov et al., 2018]

- 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].
- Statistics: quantile function $\nearrow$ w.r.t. the quantile level, pdfs are non-negative and often log-concave.
- Finance:
  - European and American call option prices: convex & monotone in the underlying stock price and $\nearrow$ in volatility [Aït-Sahalia and Duarte, 2003].

Shape constraints are omnipresent [Johnson and Jiang, 2018, Guntuboyina and Sen, 2018, Chetverikov et al., 2018]

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

Shape constraints are omnipresent [Johnson and Jiang, 2018, Guntuboyina and Sen, 2018, Chetverikov et al., 2018]

- 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].
- Statistics: quantile function $\nearrow$ w.r.t. the quantile level, pdfs are non-negative and often log-concave.
- Finance:
  - European and American call option prices: convex & monotone in the underlying stock price and $\nearrow$ in volatility [Aït-Sahalia and Duarte, 2003].
- RL and stochastic optimization: value functions are often convex [Keshavarz et al., 2011, Shapiro et al., 2014].
- Supply chain models, game theory: supermodularity [Topkis, 1998, Simchi-Levi et al., 2014].

Focus

- Find $f \in \mathcal{H}$ such that $f(\mathbf{x}_n) \approx y_n$, $0 \leq Df(\mathbf{x})$ $\forall \mathbf{x} \in K$.

Focus

- Find $f \in \mathcal{H}$ such that $f(\mathbf{x}_n) \approx y_n$, $0 \leq Df(\mathbf{x})$ $\forall \mathbf{x} \in K$.
- Various exciting approaches with asymptotic guarantees [Han and Wellner, 2016, Chen and Samworth, 2016, Freyberger and Reeves, 2018, Lim, 2020, Deng and Zhang, 2020, Kur et al., 2020], but
  1. they are often 'soft': restriction at finite many points,

Focus

- Find $f \in \mathcal{H}$ such that $f(\mathbf{x}_n) \approx y_n$, $0 \leq Df(\mathbf{x})$ $\forall \mathbf{x} \in K$.
- Various exciting approaches with asymptotic guarantees [Han and Wellner, 2016, Chen and Samworth, 2016, Freyberger and Reeves, 2018, Lim, 2020, Deng and Zhang, 2020, Kur et al., 2020], but
  1. they are often 'soft': restriction at finite many points,
  2. use simplistic function classes: polynomials, polynomial splines,

Focus

- Find $f \in \mathcal{H}$ such that $f(\mathbf{x}_n) \approx y_n$, $0 \leq Df(\mathbf{x})$ $\forall \mathbf{x} \in K$.
- Various exciting approaches with asymptotic guarantees [Han and Wellner, 2016, Chen and Samworth, 2016, Freyberger and Reeves, 2018, Lim, 2020, Deng and Zhang, 2020, Kur et al., 2020], but
  1. they are often 'soft': restriction at finite many points,
  2. use simplistic function classes: polynomials, polynomial splines,
  3. apply hard-wired parameterizations: exponential, quadratic, or

Focus

- Find $f \in \mathcal{H}$ such that $f(\mathbf{x}_n) \approx y_n$, $0 \leq Df(\mathbf{x})$ $\forall \mathbf{x} \in K$.
- Various exciting approaches with asymptotic guarantees [Han and Wellner, 2016, Chen and Samworth, 2016, Freyberger and Reeves, 2018, Lim, 2020, Deng and Zhang, 2020, Kur et al., 2020], but
  1. they are often 'soft': restriction at finite many points,
  2. use simplistic function classes: polynomials, polynomial splines,
  3. apply hard-wired parameterizations: exponential, quadratic, or
  4. only work for (a few) fixed Ds.

Focus

- Find $f \in \mathcal{H}$ such that $f(\mathbf{x}_n) \approx y_n$, $0 \le Df(\mathbf{x})$ $\forall \mathbf{x} \in K$.
- Various exciting approaches with asymptotic guarantees [Han and Wellner, 2016, Chen and Samworth, 2016, Freyberger and Reeves, 2018, Lim, 2020, Deng and Zhang, 2020, Kur et al., 2020], <u>but</u>
  1. they are often 'soft': restriction at finite many points,
  2. use simplistic function classes: polynomials, polynomial splines,
  3. apply hard-wired parameterizations: exponential, quadratic, or
  4. only work for (a few) fixed Ds.

Today: optimization framework

rich $\mathcal{H}$, hard ($\forall \mathbf{x} \in K$) shape constraints, modularity in D.

Focus

- Find $f \in \mathcal{H}$ such that $f(\mathbf{x}_n) \approx y_n$, $0 \leq Df(\mathbf{x})$ $\forall \mathbf{x} \in K$.
- Various exciting approaches with asymptotic guarantees [Han and Wellner, 2016, Chen and Samworth, 2016, Freyberger and Reeves, 2018, Lim, 2020, Deng and Zhang, 2020, Kur et al., 2020], but
  1. they are often 'soft': restriction at finite many points,
  2. use simplistic function classes: polynomials, polynomial splines,
  3. apply hard-wired parameterizations: exponential, quadratic, or
  4. only work for (a few) fixed Ds.

Today: optimization framework

rich $\mathcal{H}$, hard ($\forall \mathbf{x} \in K$) shape constraints, modularity in D.

Towards flexible $\mathcal{H}$-s ...

Kernel

- Def-1 (feature space): $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ kernel if

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

- Examples ($\gamma > 0$, $c \geq 0$, $p \in \mathbb{Z}^+$):

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

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

Kernel, RKHS

- Def-1 (feature space): $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ kernel if

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

- Def-2 (reproducing kernel):

$$k(\cdot, x) := [x' \mapsto k(x', x)] \in \mathcal{H}, \qquad f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}}.$$

Constructively, $\mathcal{H}_k = \overline{\left\{\sum_{i=1}^n \alpha_i k(\cdot, x_i) \,:\, \alpha_i \in \mathbb{R}, x_i \in \mathcal{X}, n \in \mathbb{N}^*\right\}}$.

- Examples ($\gamma > 0$, $c \geq 0$, $p \in \mathbb{Z}^+$):

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

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

Kernel, RKHS

- Def-1 (feature space): $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ kernel if

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

- Def-2 (reproducing kernel):

$$k(\cdot, x) := \left[x' \mapsto k(x', x)\right] \in \mathcal{H}, \qquad f(x) = \langle f, k(\cdot, x) \rangle_{\mathcal{H}}.$$

Constructively, $\mathcal{H}_k = \overline{\left\{\sum_{i=1}^{n} \alpha_i k(\cdot, x_i) \,:\, \alpha_i \in \mathbb{R}, x_i \in \mathcal{X}, n \in \mathbb{N}^*\right\}}$.

- Equivalent definitions, $k \overset{1:1}{\leftrightarrow} \mathcal{H}_k$.
- Included: Fourier analysis, polynomials, splines, ...
- Examples ($\gamma > 0$, $c \geq 0$, $p \in \mathbb{Z}^+$):

$$k_p(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + c)^p, \qquad k_G(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x}-\mathbf{y}\|_2^2},$$
$$k_L(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x}-\mathbf{y}\|_1}, \qquad k_e(\mathbf{x}, \mathbf{y}) = e^{\gamma \langle \mathbf{x}, \mathbf{y} \rangle}.$$

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], probability distributions [Berlinet and Thomas-Agnan, 2004, Hein and Bousquet, 2005, Smola et al., 2007, Sriperumbudur et al., 2010],
- 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., 2020, Borgwardt et al., 2020].

Task-1: joint quantile regression (JQR)

- Given: $(\tau_q)_{q\in[Q]} \subset (0,1)$ 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-1: joint quantile regression (JQR)

- Given: $(\tau_q)_{q\in[Q]} \subset (0,1)$ levels $\nearrow$, $\{(\mathbf{x}_n, y_n)\}_{n\in[N]}$ samples.
- Estimate jointly the τ_q-quantiles of $\mathbb{P}(Y|X=\mathbf{x})$ [Sangnier et al., 2016].
- Objective:

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

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

Task-1: joint quantile regression (JQR)

- Given: $(\tau_q)_{q\in[Q]} \subset (0,1)$ levels $\nearrow$, $\{(\mathbf{x}_n, y_n)\}_{n\in[N]}$ samples.
- Estimate jointly the τ_q-quantiles of $\mathbb{P}(Y|X=\mathbf{x})$ [Sangnier et al., 2016].
- Objective:

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

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

- 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-1: joint quantile regression (JQR)

- Given: $(\tau_q)_{q\in[Q]} \subset (0,1)$ levels $\nearrow$, $\{(\mathbf{x}_n, y_n)\}_{n\in[N]}$ samples.
- Estimate jointly the τ_q-quantiles of $\mathbb{P}(Y|X=\mathbf{x})$ [Sangnier et al., 2016].
- Objective:

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

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

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

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

Constraints

function values (f_q) with interaction $(f_{q+1} - f_q)$, bias terms (b_q) with interaction $(b_q - b_{q+1})$.

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

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

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

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

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

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

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

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

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

Constraints

function values (f_q) and derivatives (f_q') with interaction ($f_q - f_{q+1}$), bias terms (b_q) with interaction ($b_{q+1} - b_q$).

Task-3: safety-critical control

- Trajectory of an underwater vehicle:

$$t \in \mathcal{T} := [0, 1] \mapsto [x(t); z(t)] \in \mathbb{R}^2.$$

Task-3: safety-critical control

- Trajectory of an underwater vehicle:

$$t \in \mathcal{T} := [0, 1] \mapsto [x(t); z(t)] \in \mathbb{R}^2.$$

- Simplifying assumption: $x(0) = 0, \dot{x}(t) = 1 \, \forall t \in \mathcal{T} \Rightarrow x(t) = t$.

Task-3: safety-critical control

- Trajectory of an underwater vehicle:

$$t \in \mathcal{T} := [0, 1] \mapsto [x(t); z(t)] \in \mathbb{R}^2.$$

- Simplifying assumption: $x(0) = 0, \dot{x}(t) = 1 \, \forall t \in \mathcal{T} \Rightarrow x(t) = t.$
- Requirement: stay between the floor and the ceiling of the cavern

$$z(t) \in [z_{\text{low}}(t), z_{\text{up}}(t)] \;\; \forall t \in \mathcal{T}.$$

Task-3: safety-critical control

- Trajectory of an underwater vehicle:

$$t \in \mathcal{T} := [0, 1] \mapsto [x(t); z(t)] \in \mathbb{R}^2.$$

- Simplifying assumption: $x(0) = 0, \dot{x}(t) = 1 \, \forall t \in \mathcal{T} \Rightarrow x(t) = t$.
- Requirement: stay between the floor and the ceiling of the cavern

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

- Initial condition: $z(0) = 0$ and $\dot{z}(0) = 0$.

Task-3: safety-critical control

- Trajectory of an underwater vehicle:

$$t \in \mathcal{T} := [0,1] \mapsto [x(t); z(t)] \in \mathbb{R}^2.$$

- Simplifying assumption: $x(0) = 0, \dot{x}(t) = 1 \, \forall t \in \mathcal{T} \Rightarrow x(t) = t$.
- Requirement: stay between the floor and the ceiling of the cavern

$$z(t) \in [z_{\text{low}}(t), z_{\text{up}}(t)] \;\; \forall t \in \mathcal{T}.$$

- Initial condition: $z(0) = 0$ and $\dot{z}(0) = 0$.
- Control task (LQ = linear dynamics & quadratic cost):

$$\begin{aligned}
&\min_{u \in L^2(\mathcal{T},\mathbb{R})} \quad \int_{\mathcal{T}} |u(t)|^2 \mathrm{d}t \\
&\text{s.t.} \\
&z(0) = 0, \quad \dot{z}(0) = 0, \\
&\ddot{z}(t) = -\dot{z}(t) + u(t), \, \forall t \in \mathcal{T}, \\
&z_{\text{low}}(t) \leq z(t) \leq z_{\text{up}}(t), \, \forall \, t \in \mathcal{T}.
\end{aligned}$$

Task-3: safety-critical control – continued

- With full state $\mathbf{f}(t) := [z(t); \dot{z}(t)] \in \mathbb{R}^2$

$$\dot{\mathbf{f}}(t) = \mathbf{A}\mathbf{f}(t) + \mathbf{B}u(t),\ \mathbf{f}(0) = \mathbf{0},\ \mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \in \mathbb{R}^{2\times 2},\ \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \mathbb{R}^2$$

Task-3: safety-critical control – continued

- With full state $\mathbf{f}(t) := [z(t); \dot{z}(t)] \in \mathbb{R}^2$

$$\dot{\mathbf{f}}(t) = \mathbf{A}\mathbf{f}(t) + \mathbf{B}u(t),\ \mathbf{f}(0) = \mathbf{0},\ \mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \in \mathbb{R}^{2\times 2},\ \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \mathbb{R}^2$$

- The controlled trajectories $\mathbf{f}$ belong to a $\mathbb{R}^2$-valued RKHS with kernel

$$k(s,t) := \int_0^{\min(s,t)} e^{(s-\tau)\mathbf{A}}\mathbf{B}\mathbf{B}^\top e^{(t-\tau)\mathbf{A}^\top}\,\mathrm{d}\tau, \quad s,t \in \mathcal{T},$$

and the task is

$$\begin{aligned} \min_{\mathbf{f}=[f_1;f_2]\in\mathcal{H}_k} \quad & \|\mathbf{f}\|^2_{\mathcal{H}_k} \\ \text{s.t.} \quad & \\ z_{\text{low}}(t) \le f_1(t) \le z_{\text{up}}(t),\ \forall\, t \in \mathcal{T}. & \end{aligned}$$

Task-3: safety-critical control – finished

- Assume for simplicity: z_{low} and z_{up} are piece-wise constant.
- Task:

$$
\begin{aligned}
&\min_{\mathbf{f}=[f_1;f_2]\in\mathcal{H}_k} \quad \|\mathbf{f}\|^2_{\mathcal{H}_k}\\
&\quad\text{s.t.}\\
&z_{\text{low},m} \le f_1(t) \le z_{\text{up},m},\ \forall\, t \in \mathcal{T}_m,\ \forall m \in [M].
\end{aligned}
$$

Task-3: safety-critical control – finished

- Assume for simplicity: z_{low} and z_{up} are piece-wise constant.
- Task:

$$\begin{aligned} &\min_{\mathbf{f}=[f_1;f_2]\in\mathcal{H}_k} \quad \|\mathbf{f}\|^2_{\mathcal{H}_k} \\ &\text{s.t.} \\ &z_{\text{low},m} \le f_1(t) \le z_{\text{up},m},\ \forall\, t \in \mathcal{T}_m,\ \forall m \in [M]. \end{aligned}$$

Constraints

linear transformation of functions (f_1), with matrix-valued kernel.

Our task

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

Our task

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

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

Our task

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

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

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

Our task

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

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

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

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

Our task

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

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

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

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

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

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. $\mathbf{f}_{0,i} \in \mathcal{H}_k$ for $\forall i$.
4. Bias domain $\mathcal{B} \subseteq \mathbb{R}^Q$: convex.
5. Loss L restricted to $\mathcal{B}$: strictly convex in $\mathbf{b}$.
6. Regularizer Ω: strictly increasing in each of its argument.

Our strenghtened SOC-constrained formulation

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

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

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

where

- $\{\tilde{\mathbf{x}}_{i,m}\}_{m \in [M_i]}$: a δ_i-net of K_i in $\|\cdot\|_{\mathcal{X}}$,
- $\eta_i = \sup_{m \in [M_i], \mathbf{u} \in \mathbb{B}_{\|\cdot\|_{\mathcal{X}}}(\mathbf{0},1)} \|D_{i,\mathbf{x}} k(\tilde{\mathbf{x}}_{i,m}, \cdot) - D_{i,\mathbf{x}} k(\tilde{\mathbf{x}}_{i,m} + \delta_i \mathbf{u}, \cdot)\|_{\mathcal{H}_k}$,
- $D_{i,\mathbf{x}} k(\mathbf{x}_0, \cdot) := \mathbf{y} \mapsto D_i(\mathbf{x} \mapsto k(\mathbf{x}, \mathbf{y}))(\mathbf{x}_0)$.

Tightening idea

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

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

Tightening idea

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

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

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

Tightening idea

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

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

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

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

Tightening idea

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

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

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

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

Theorem

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

Then,

- (i) Tightening: any $(\mathbf{f}, \mathbf{b})$ 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 '$\boldsymbol{\eta} = \mathbf{0}$', $\bar{v} = \mathcal{L}\left(\bar{\mathbf{f}}, \bar{\mathbf{b}}\right)$, $v_{\boldsymbol{\eta}} = \mathcal{L}(\mathbf{f}_{\boldsymbol{\eta}}, \mathbf{b}_{\boldsymbol{\eta}})$.
- Let $\mathbf{f}_{\boldsymbol{\eta}} = (f_{\boldsymbol{\eta},q})_{q\in[Q]}$.

Then,

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

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

- (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 $(\mu_{f_q}, \mu_{\mathbf{b}})$-strongly convex w.r.t. $(f_q, \mathbf{b})$ for any $q \in [Q]$, then

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

Theorem – continued

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

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

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

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

Theorem – continued

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

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

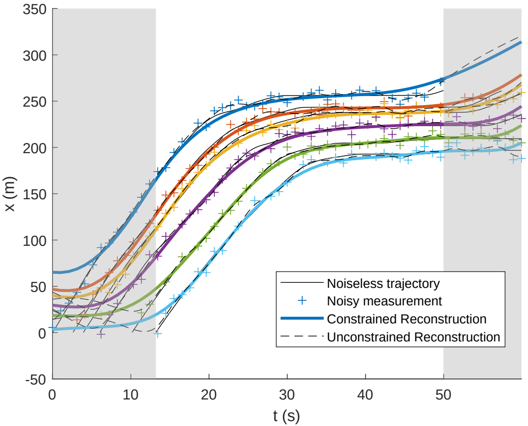
If in addition $\mathbf{U}$ is surjective, $\mathcal{B} = \mathbb{R}^Q$, and $\mathcal{L}(\bar{\mathbf{f}}, \cdot)$ is L_b−Lipschitz continuous on $\mathbb{B}_{\|\cdot\|_2}\left(\bar{\mathbf{b}}, c_f\|\boldsymbol{\eta}\|_\infty\right)$ where $c_f = \sqrt{d}\left\|\left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T\right\|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

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

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

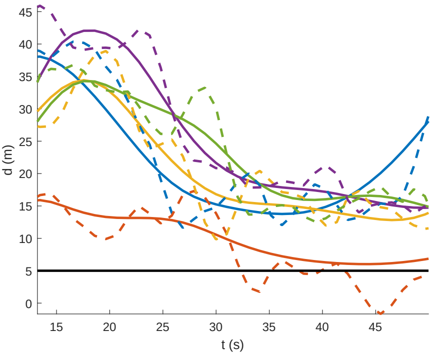
Demo (task-1): convoy localization with traffic jam

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

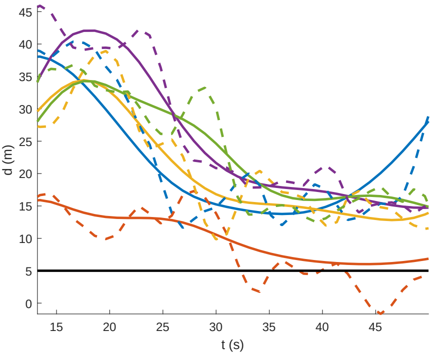


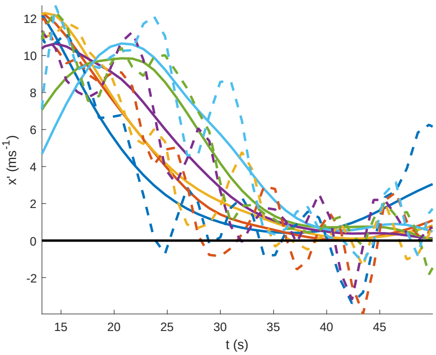
Demo (task-1): continued

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



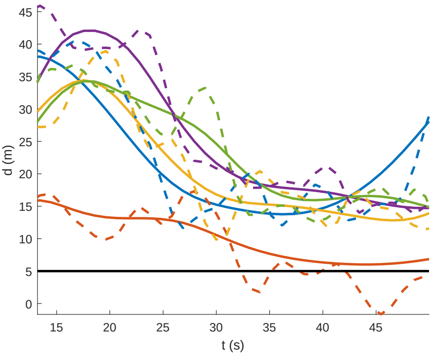
Pairwise distances: $t \mapsto f_q(t) - f_{q+1}(t)$ Speed: $t \mapsto f_q'(t)$

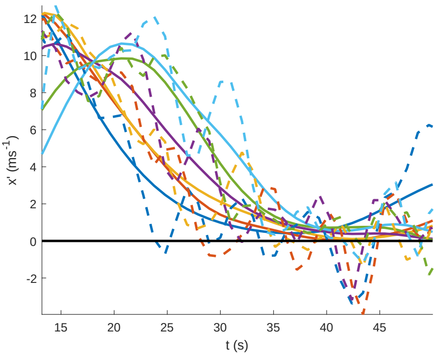




Demo (task-1): continued

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



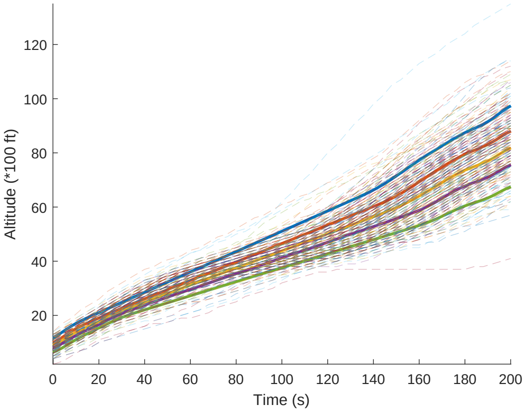


Shape constraints: especially relevant in noisy situations.

Demo (task-2): joint quantile regression

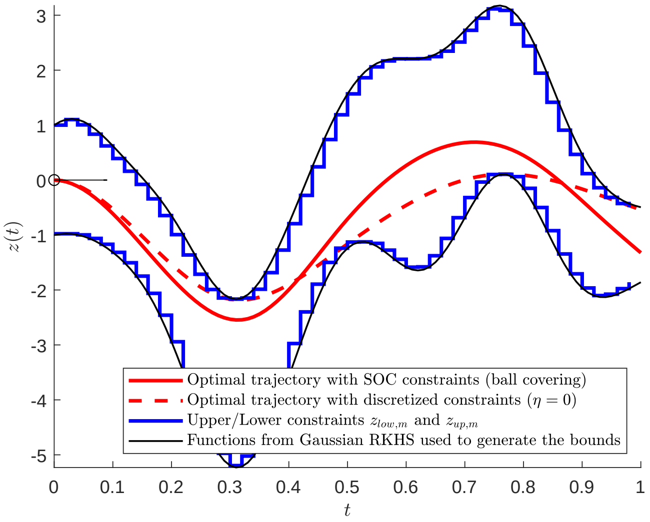
Analysis of aircraft trajectories, ENAC [Nicol, 2013]

- 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).



Demo (task-3): control of underwater vehicle

Vs discretization-based approach (which might crash):



Summary

- Focus: hard affine shape constraints on derivatives & RKHS.
- Proposed framework: SOC-based tightening.
- Applications:
  - convoy localization,
  - joint quantile regression: aircraft trajectories,
  - safety-critical control.

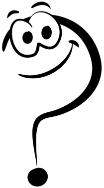
References & acknowledgements

- Transportation systems [Aubin-Frankowski et al., 2020].
- Control aspect [Aubin-Frankowski, 2020].
- Method:
  - $dim(y) = 1$: [Aubin-Frankowski and Szabó, 2020]. Code @ GitHub.
  - $dim(y) \geq 1$ and SDP constraints (say joint convexity, production functions): [Aubin-Frankowski and Szabó, 2021].

Acknowledgements: 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.

References & acknowledgements

- Transportation systems [Aubin-Frankowski et al., 2020].
- Control aspect [Aubin-Frankowski, 2020].
- Method:
  - $dim(y) = 1$: [Aubin-Frankowski and Szabó, 2020]. Code @ GitHub.
  - $dim(y) \geq 1$ and SDP constraints (say joint convexity, production functions): [Aubin-Frankowski and Szabó, 2021].



Acknowledgements: 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.

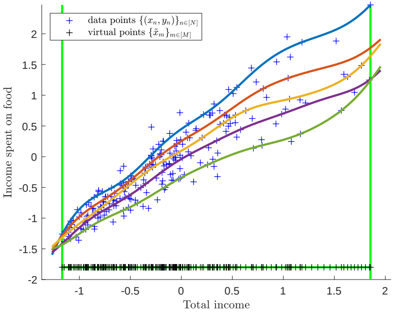
INSTITUT Louis Bachelier

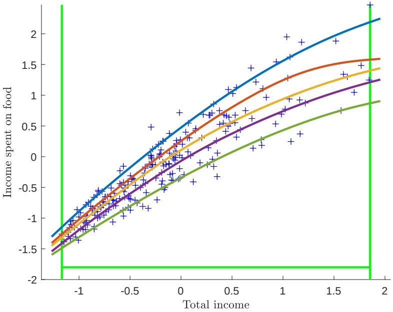
Stress Test RISK Management and Financial Steering

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.





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. (2020).
Linearly-constrained linear quadratic regulator from the viewpoint of kernel methods.
Technical report.
(`https://arxiv.org/abs/2011.02196`).

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.

Aubin-Frankowski, P.-C. and Szabó, Z. (2020).
Hard shape-constrained kernel machines.
In *Advances in Neural Information Processing Systems (NeurIPS)*.

Aubin-Frankowski, P.-C. and Szabó, Z. (2021).
Handling hard affine SDP shape constraints in RKHSs.
Technical report.
(http://arxiv.org/abs/2101.01519).

Bai, L., Cui, L., Rossi, L., Xu, L., Bai, X., and Hancock, E. (2020).
Local-global nested graph kernels using nested complexity traces.
Pattern Recognition Letters, 134:87–95.

Berlinet, A. and Thomas-Agnan, C. (2004).
Reproducing Kernel Hilbert Spaces in Probability and Statistics.
Kluwer.

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., Ghisu, E., Llinares-López, F., O'Bray, L., and Riec, B. (2020).
Graph kernels: State-of-the-art and future challenges.
Foundations and Trends in Machine Learning, 13(5-6):531–712.

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.

Chetverikov, D., Santos, A., and Shaikh, A. M. (2018).
The econometrics of shape restrictions.
Annual Review of Economics, 10(1):31–63.

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.

Deng, H. and Zhang, C.-H. (2020).
Isotonic regression in multi-dimensional spaces and graphs.
Annals of Statistics, 48(6):3672–3698.

Freyberger, J. and Reeves, B. (2018).
Inference under shape restrictions.
Technical report, University of Wisconsin-Madison.
(https://www.ssc.wisc.edu/~jfreyberger/Shape_Inference_Freyberger_Reeves.pdf).

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.

Guntuboyina, A. and Sen, B. (2018).
Nonparametric shape-restricted regression.
Statistical Science, 33(4):568–594.

Han, Q. and Wellner, J. A. (2016).
Multivariate convex regression: global risk bounds and adaptation.
Technical report.
(`https://arxiv.org/abs/1601.06844`).

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`).

Hein, M. and Bousquet, O. (2005).
Hilbertian metrics and positive definite kernels on probability measures.
In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 136–143.

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.

Johnson, A. L. and Jiang, D. R. (2018).
Shape constraints in economics and operations research.
Statistical Science, 33(4):527–546.

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.

Kur, G., Dagan, Y., and Rakhlin, A. (2020).

Optimality of maximum likelihood for log-concave density estimation and bounded convex regression.
Technical report.
(https://arxiv.org/abs/1903.05315).

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.

Lim, E. (2020).

The limiting behavior of isotonic and convex regression estimators when the model is misspecified.
Electronic Journal of Statistics, 14:2053–2097.

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.

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

Nicol, F. (2013).
Functional principal component analysis of aircraft trajectories.

In *International Conference on Interdisciplinary Science for Innovative Air Traffic Management (ISIATM)*.

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.

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

Smola, A., Gretton, A., Song, L., and Schölkopf, B. (2007).
A Hilbert space embedding for distributions.
In *Algorithmic Learning Theory (ALT)*, pages 13–31.

Sriperumbudur, B., Gretton, A., Fukumizu, K., Schölkopf, B., and Lanckriet, G. (2010).
Hilbert space embeddings and metrics on probability measures.

Journal of Machine Learning Research, 11:1517–1561.

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.