Shape Restricted Splines via Constrained Optimization: Computation and Statistical Analysis

Jinglai Shen

Department of Mathematics and Statistics
University of Maryland Baltimore County (UMBC), Baltimore, MD

6th Int'l Conference on Complementarity Problems
Berlin, Germany, August 5, 2014

joint with Teresa Lebair (UMBC) and Xiao Wang (Purdue)

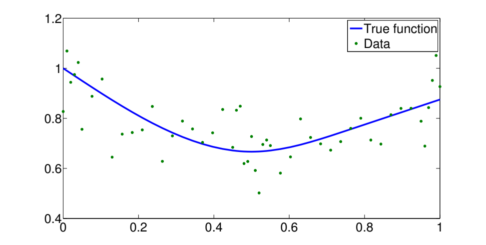
Acknowledgements: J.-S. Pang (USC) and A. Draganescu (UMBC)

Outline

Shape Constrained Curve-fitting/Estimation

Motivation

1. Various static or dynamic models of biologic, engineering and economic systems contain shape constrained functions
2. Example: convex shape constraint



Applications

- Biology: dose response, drug combination, and genetic networks
- Engineering: path planning, lifetime estimation in reliability engr.
- Statistics: isotonic regression, log-concave density estimation

Focused Topics

Topic I: Computation of shape constrained smoothing splines

1. Formulated as a constrained optimal control or constrained optimization problem with nonsmooth features
2. Efficient numerical schemes

Topic II: Statistical analysis of shape constrained estimators

1. Convergence of an estimator to the true function: consistency and convergence rate
2. Optimal rate estimation and minimax optimal estimation

T. Robertson, F.T. Wright, and R.L. Dykstra. *Order Restricted Statistical Inference*. John Wiley & Sons Ltd., 1988.

Smoothing Splines

Smoothing spline model: unconstrained case

1. Classical smoothing splines (Wahba): $\min_{f \in \mathcal{S}} J(f)$, where $f : [0,1] \to \mathbb{R}$, $(t_i, y_i)_{i=1}^n$ are samples, and

$$J(f) := \frac{1}{n} \sum_{i=1}^{n} \Big(f(t_i) - y_i \Big)^2 + \lambda \int_0^1 \Big(f^{(m)}(t) \Big)^2 dt$$

2. Control theoretical splines (Egerstedt and Martin)

$$\min \frac{1}{n} \sum_{i=1}^{n} \Big(f(t_i) - y_i \Big)^2 + \lambda \int_0^1 u^2(t) dt$$

where

$$\dot{x}(t) = Ax(t) + bu(t), \quad f(t) = c^T x(t), \quad A \in \mathbb{R}^{\ell \times \ell}, \quad b, c \in \mathbb{R}^{\ell}.$$

Example: when $m = 2$, $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $c = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and $u(t) = f''(t)$.

Shape Constrained Smoothing Splines

Example: convex smoothing spline

- $\min J(f) := \frac{1}{n}\sum_{i=1}^{n}\big(f(t_i)-y_i\big)^2+\lambda\int_0^1\big(f^{(2)}(t)\big)^2dt,\quad f^{(2)}\geq 0\ a.e.\ [0,1]$
- equivalently, $\min J(f) := \frac{1}{n}\sum_{i=1}^{n}\big(f(t_i)-y_i\big)^2+\lambda\int_0^1 u^2(t)dt$ subject to

$$\dot{x}(t)=Ax(t)+bu(t),\quad f(t)=c^Tx(t),\quad u(t)\in\Omega:=\mathbb{R}_+\ a.e.\ [0,1]$$

Formulation of shape constrained smoothing spline

Given a (constrained) linear control system $\Sigma(A,B,C,\Omega)$ on $\mathbb{R}^\ell$:

$$\dot{x}=Ax+Bu,\qquad u\in\mathcal{W}:=\{u\in L_2([0,1];\mathbb{R}^m)\,|\,u(t)\in\Omega\ a.e.\},$$

where $A\in\mathbb{R}^{\ell\times\ell}$, $B\in\mathbb{R}^{\ell\times m}$, $C\in\mathbb{R}^{p\times\ell}$, $\Omega\subseteq\mathbb{R}^m$ is closed and convex. Given $\{(t_i,y_i)\}_{i=1}^n$ and weights $w_i>0$ with $\sum_{i=1}^n w_i=1$, define the cost functional

$$J(u,x_0)\ :=\ \sum_{i=1}^{n}w_i\big\|y_i-Cx(t_i;u,x_0)\big\|_2^2+\lambda\int_0^1\|u(t)\|_2^2dt$$

A shape constrained smoothing spline $\widehat{f}$ is determined by an optimal solution of $\inf J(u,x_0)$ subject to $\Sigma(A,B,C,\Omega)$, i.e., $\widehat{f}(t)=Cx(t;u^*,x_0^*)$.

Optimality Conditions

Existence and uniqueness of optimal solution

Suppose

$$\mathbf{H.1}: \ \text{rank} \begin{pmatrix} Ce^{At_1} \\ Ce^{At_2} \\ \vdots \\ Ce^{At_n} \end{pmatrix} = \ell.$$

Then there exists a unique optimal solution $(u^*, x_0^*) \in \mathcal{W} \times \mathbb{R}^\ell$ for any (t_i, y_i), (w_i), and $\lambda > 0$.

Optimality conditions in term of VI

$$\begin{aligned} u^*(t) &= \Pi_\Omega\Big(-\lambda^{-1}\sum_{i=1}^n w_i P_i^T(t)\big(\widehat{f}(t_i) - y_i\big)\Big), \quad \text{and} \\ 0 &= \sum_{i=1}^n w_i\big(Ce^{A_i t_i}\big)^T\big(\widehat{f}(t_i) - y_i\big), \end{aligned}$$

where $\widehat{f}(t_i) = Cx(t_i; u^*(t_i), x_0^*)$, and $P_i(t) := Ce^{A(t_i - t)}B \cdot \mathbf{I}_{[0,t_i]}$.

More on Optimality Conditions

Facts

1. On each $[t_k, t_{k+1})$, $u^*(t)$ depends on $\widehat{f}(t_i)$ with $t_i < t_k$ only.
2. The optimal initial condition x_0^* completely determines u^* and $\widehat{f}$ on $[0,1]$ (may write $\widehat{f}$ as $\widehat{f}(t, x_0^*)$)
3. Given (t_i, y_i) and (w_i) and λ, define $H_{y,n} : \mathbb{R}^\ell \to \mathbb{R}^\ell$

$$H_{y,n}(z) := \sum_{i=1}^{n} w_i \big(Ce^{A_i t_i}\big)^T \Big(\widehat{f}(t_i, z) - y_i\Big)$$

   Then the equation $H_{y,n}(z) = 0$ has a unique solution (under **H.1**), which is the optimal initial condition x_0^*.

Nonsmoothness of $\widehat{f}(t, \cdot)$ and $H_{y,n}$

1. If Π_Ω is directionally differentiable on $\mathbb{R}^m$, then $\widehat{f}(t, z)$ is B-differentiable in z for any fixed $t \in [0,1]$;
2. If Π_Ω is semismooth on $\mathbb{R}^m$, then $\widehat{f}(t, z)$ is semismooth in z for any fixed $t \in [0,1]$.

Boundedness of Level Sets

Level set of $H_{y,n}$

Given $z_* \in \mathbb{R}^\ell$, define $S_{z_*} := \big\{ z \in \mathbb{R}^\ell \,|\, \|H_{y,n}(z)\| \le \|H_{y,n}(z_*)\| \big\}$

Proposition (Boundedness of level sets)

Let $\Omega \subseteq \mathbb{R}^m$ be closed and convex. For any given $(t_i, y_i), (w_i), \lambda > 0$ and z_* such that **H.1** holds, the level set S_{z_*} is bounded.

Sketch of the proof

Suppose not. Then there exists (z_k) in S_{z_*} with $\|z_k\| \to \infty$ and $z_k/\|z_k\| \to v_* \neq 0$. It can be shown

$$\lim_{k\to\infty} \frac{H_{y,n}(z_k)}{\|z_k\|} = \widetilde{H}_{\widetilde{y},n}(v_*)\big|_{\widetilde{y}=0},$$

where $\widetilde{H}_{\widetilde{y},n}(z) = \sum_{i=1}^n w_i (Ce^{A_i t_i})^T \big(\widetilde{f}(t_i, z) - \widetilde{y}_i\big)$, $\widetilde{f}$ is obtained from the linear control system $\Sigma(A, B, C, \Omega^\infty)$, and $\widetilde{y}_i = 0, \forall i$. Since $\widetilde{H}_{0,n}(z) = 0$ has a unique solution $z = 0$, $\widetilde{H}_{0,n}(v_*) \neq 0$ and $\|H_{y,n}(z_k)\| \to \infty$, contradiction.

Solving $H_{y,n}(z) = 0$ for Polyhedral Ω (I)

Notation

- Define $F(z) := B \circ \Pi_\Omega \circ B^T$
- For each $k = 1, 2, \ldots, n-1$, let

$$v_k(z) := \frac{1}{\lambda} \sum_{i=1}^{k} w_i \big(Ce^{A_i t}\big)^T \Big(\widehat{f}(t_i, z) - y_i\Big), \quad q(t, v) := e^{-A^T t} v$$

Then $Bu^*(t, z) = F(q(t, v_k(z)))$ for all $t \in [t_k, t_{k+1})$.

Non-degenerate case

1. $F : \mathbb{R}^\ell \to \mathbb{R}^\ell$ is continuous and piecewise affine, and admits a polyhedral subdivision Ξ.
2. For any v and k, $q(t, v)$ has finitely many switchings on Ξ in $[t_k, t_{k+1}]$.
3. $q(t, v)$ is called non-degenerate on $[t_k, t_{k+1}]$ if it is in the interior of a polyhedron of Ξ between any consecutive switching times; otherwise, $q(t, v)$ is called degenerate.

Solving $H_{y,n}(z) = 0$ for Polyhedral Ω (II)

More assumptions and notation

- Let $\rho_1 > 0$ and $\rho_2 > 0$ be such that $\|Ce^{A(t-s)}\|_\infty \leq \rho_1, \forall\, t, s \in [0,1]$ and $\max_i \|E_i\|_\infty \leq \rho_2$, where each matrix E_i corresponds to an affine piece of F.
- Assumption **H.2**: there exist $\rho_t > 0$ and $\mu \geq \nu > 0$ such that for all n,

$$\max_{0 \leq i \leq n-1} |t_{i+1} - t_i| \leq \frac{\rho_t}{n}, \qquad \frac{\nu}{n} \leq w_i \leq \frac{\mu}{n}, \quad \forall\, i.$$

Theorem (Non-degenerate case)

Let Ω be a polyhedron in $\mathbb{R}^m$. Assume that **H.1** – **H.2** hold and $\lambda \geq \mu^2 \rho_1^2 \rho_2 \rho_t / (4\nu)$. Suppose that $q(t, v_k(z))$ is non-degenerate on $[t_k, t_{k+1}]$ for each $k = 1, 2 \ldots, n-1$. Then there exists a unique direction vector $d \in \mathbb{R}^\ell$ such that

$$H_{y,n}(z) + H'_{y,n}(z; d) = 0.$$

Solving $H_{y,n}(z) = 0$ for Polyhedral Ω (III)

Proposition (Degenerate case)

Assume additionally that (C, A) is an observable pair. If $q(t, v_k(z))$ is degenerate on $[t_k, t_{k+1}]$ for some $k \in \{1, \ldots, n-1\}$, then for any $\varepsilon > 0$, there exists $d \in \mathbb{R}^\ell$ with $0 < \|d\| \leq \varepsilon$ such that $q(t, v_k(z+d))$ is non-degenerate on $[t_k, t_{k+1}]$ for each $k = 1, \ldots, n-1$.

Modified Nonsmooth Newton's Method w. Line Search

- Apply the modified nonsmooth Newton's method with line search based on (Pang, 1990) to solve $H_{y,n}(z) = 0$
- Numerical convergence is proved under suitable conditions

J.-S. Pang. Newton's method for B-differentiable equations. *Mathematics of Operations Research*, Vol. 15, pp. 311–341, 1990.

Numerical Results: Example I

Consider $y_i - f(t_i) \sim \mathcal{N}(0, \sigma^2)$

Example 1: Convex constraint w. unevenly spaced design pts

$$f(t) = \begin{cases} \frac{4}{3}t^3 - t + 1 & \text{if } t \in [0, \frac{1}{2}) \\ -\frac{8}{3}t^3 + 6t^2 - 4t + \frac{3}{2} & \text{if } t \in [\frac{1}{2}, \frac{3}{4}) \\ \frac{1}{2}t + \frac{3}{8} & \text{if } t \in [\frac{3}{4}, 1] \end{cases}$$

$$u(t) = f''(t) = \begin{cases} 8t & \text{if } t \in [0, \frac{1}{2}) \\ 12 - 16t & \text{if } t \in [\frac{1}{2}, \frac{3}{4}) \in \Omega := [0, \infty), \\ 0 & \text{if } t \in [\frac{3}{4}, 1] \end{cases}$$

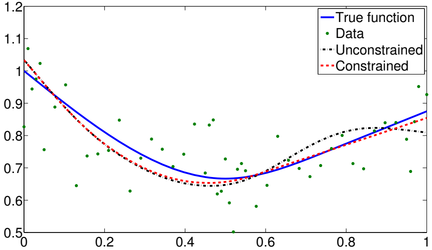
$$z^0 = (2, 3)^T, \qquad \sigma = 0.1, \qquad \frac{\sigma}{|f_{\max} - f_{\min}|} = 30\%, \qquad \lambda = 10^{-4},$$

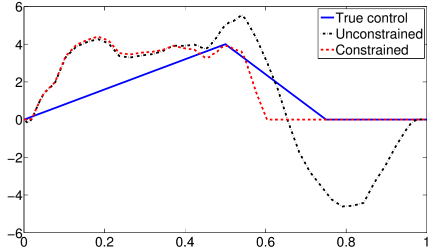
Design points (t_i):

$$\Big\{0, \frac{1}{2n}, \ldots, \frac{1}{20}, \frac{1}{20} + \frac{4}{3n}, \ldots, \frac{9}{20}, \frac{9}{20} + \frac{1}{2n}, \ldots, \frac{11}{20}, \frac{11}{20} + \frac{4}{3n}, \ldots, \frac{19}{20}, \frac{19}{20} + \frac{1}{2n}, \ldots, 1$$

$$x_0 = (1, -1)^T, \qquad A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 \end{bmatrix}^T, \qquad C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

Numerical Results: Example I with $n = 50$





Numerical Results: Example II

Example 2: General dynamics and constraint with unevenly spaced design points $u(t) \in \Omega := [8, \infty)$

$$f(t) = \begin{cases} 11.60967t(e^{-t} + e^{-2t}) - 27.21935e^{-t} + 25.21945e^{-2t} + 2 & \text{if } t \in [0, \frac{1}{4}) \\ -6.23368e^{-t} + 3.25670e^{-2t} + 3 & \text{if } t \in [\frac{1}{4}, \frac{1}{2}) \\ -11.60967t(e^{-t} + e^{-2t}) + 18.22245e^{-t} - 21.69226e^{-2t} + 3 & \text{if } t \in [\frac{1}{2}, \frac{3}{4}) \\ -3.34450e^{-t} + 1.30615e^{-2t} + 2 & \text{if } t \in [\frac{3}{4}, 1] \end{cases}$$

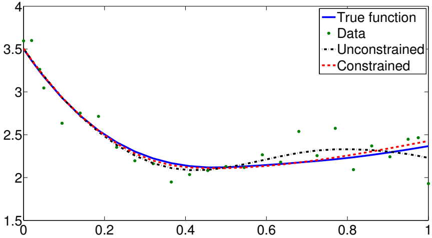
$$u(t) = f''(t) + 3f'(t) + 2f(t) = \begin{cases} 23.21935(e^{-t} - e^{-2t}) + 8 & \text{if } t \in [0, \frac{1}{4}) \\ 12 & \text{if } t \in [\frac{1}{4}, \frac{1}{2}) \\ -38.28223e^{-t} + 63.11673e^{-2t} + 6 & \text{if } t \in [\frac{1}{2}, \frac{3}{4}) \\ 8 & \text{if } t \in [\frac{3}{4}, 1] \end{cases}$$

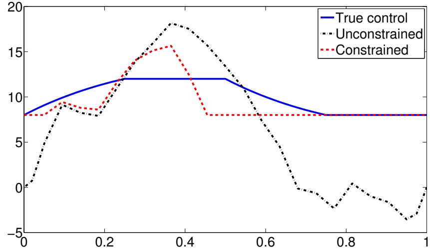
$$z^0 = (0, 1/2)^T, \qquad \sigma = 0.2, \qquad \frac{\sigma}{|f_{\max} - f_{\min}|} = 14.5\%, \qquad \lambda = 10^{-4},$$

$$\text{Design points } (t_i) = \left\{0, \frac{1}{2n}, \frac{2}{2n}, \ldots, \frac{1}{20}, \frac{1}{20} + \frac{9}{8n}, \ldots, \frac{19}{20}, \frac{19}{20} + \frac{1}{2n}, \ldots, 1\right\},$$

$$x_0 = (7/2, -7)^T, \qquad A = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 \end{bmatrix}^T, \qquad C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$

Numerical Results: Example II with $n = 25$





Numerical Performance

Constrained vs. unconstrained smoothing splines

Shape constrained smoothing splines outperform their unconstrained counterparts

		$\|f - \widehat{f}\|_{L_2}$		$\|f - \widehat{f}\|_{L_\infty}$		$\|x(0) - \widehat{x}_0\|_2$	
		const.	unconst.	const.	unconst.	const.	unconst.
I	$n = 25$	0.00696	0.00723	0.06809	0.07216	0.25985	0.30825
	$n = 50$	0.00351	0.00362	0.04971	0.05218	0.19141	0.22549
	$n = 100$	0.00177	0.00180	0.03487	0.03588	0.14021	0.15958
II	$n = 25$	0.01302	0.01492	0.12639	0.15609	0.76778	1.45583
	$n = 50$	0.00704	0.00791	0.09998	0.12474	0.70899	1.41832
	$n = 100$	0.00387	0.00436	0.08048	0.10519	0.75410	1.54277

Numerical convergence of modified Newton's method

- Depends heavily on examples but appears to be superlinear
- Typically ranges between 10 and 30 iterations
- Iterations for convergence increase slightly with sample size n

Shape Constrained Regression

Regression model

$$y_i = f(t_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$

where $f : [0,1] \to \mathbb{R}$ is the underlying true function subject to the constraint $f \in \mathcal{C}$, t_i are design points, y_i are samples, and ε_i are i.i.d. random variables with $\varepsilon_i \sim N(0, \sigma^2)$.

Constraints

1. Shape constraint: $f \in \mathcal{S}$, where for some $m \in \mathbb{N}$,

   $$\mathcal{S} := \left\{ f : [0,1] \to \mathbb{R} \,|\, (f^{(m-1)}(t_1) - f^{(m-1)}(t_2)) \cdot (t_1 - t_2) \geq 0, \forall\, t_1, t_2 \in [0,1] \right\}.$$

2. Smoothness constraint: f is in the Hölder class $H(r, L)$ with $r \in (m-1, m]$, $L > 0$, i.e., the family of $\ell := (m-1)$ times continuously differentiable functions whose ℓ-th derivative is uniformly Hölder continuous with exponent $\gamma := r - \ell \in (0,1]$, i.e.,

   $$|f^{(\ell)}(t_1) - f^{(\ell)}(t_2)| \leq L \cdot |t_1 - t_2|^{\gamma}, \quad \forall\, t_1, t_2 \in [0,1].$$

Minimax Optimal Estimation

Key issues on a given function class $\mathcal{C}$

- What is the "best rate" of convergence of estimators uniformly on $\mathcal{C}$?
- How can one construct an estimator that achieves the "best rate" of convergence on $\mathcal{C}$? (minimax upper bound)
- Is the "best rate" of convergence strict on $\mathcal{C}$ for any permissible estimator? (minimax lower bound)

Optimal rate of convergence on $H(r, L)$ in the sup-norm

$$\inf_{\widehat{f}} \sup_{f \in H(r,L)} \mathbb{E}\left(\|\widehat{f} - f\|_\infty\right) \asymp L^{\frac{1}{2r+1}} \sigma^{\frac{2r}{2r+1}} \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}},$$

where $\widehat{f}$: estimate of a true function f, and $a \asymp b$: a/b is bounded by two positive constants from below and above for all n sufficiently large.

Motivating question

For a given $m \in \mathbb{N}$, what are the minimax upper and lower bounds over $\mathcal{S}_H(r, L) := H(r, L) \cap \mathcal{S}$ as $n \to \infty$ (when the sup-norm is used)?

Constrained B-spline Estimator (I)

Constrained B-spline estimator

$$\widehat{f}(t) = \sum_{k=1}^{K_n+m-1} \widehat{b}_k B_k(t)$$

where $t_i = i/n$, B_k are B-splines of $(m-1)$th degree with knots $\kappa_i = i/K_n$, and the optimal spline coefficient $\widehat{b} = \{\widehat{b}_k, k = 1, \ldots, K_n + m - 1\}$ is

$$\widehat{b} = \arg \min_{D_m b \geq 0} \sum_{i=1}^{n} \left[y_i - \sum_{k=1}^{K_n+m-1} b_k B_k(t_i) \right]^2$$

Here $D_m \in \mathbb{R}^{(K_n-1)\times(K_n+m-1)}$ corresponds to the m-th difference operator.

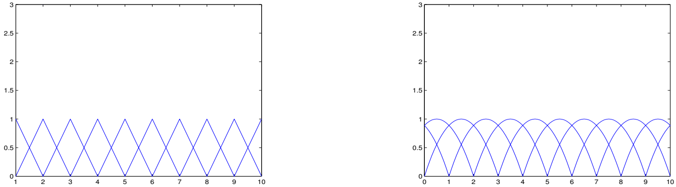


Figure: Left: B-splines of degree 1; Right: B-splines of degree 2

Constrained B-spline Estimator (II)

Quadratic program for optimal spline coefficients

$$\widehat{b} = \arg \min_{D_m b \geq 0} \frac{1}{2} b^T \Lambda_{K_n} b - b^T \bar{y},$$

where

$$\Lambda_{K_n} = \frac{1}{\beta_n} X^T X, \quad \bar{y} = \frac{1}{\beta_n} X^T y, \quad y = (y_1, \ldots, y_n)^T.$$

Here $\beta_n := \sum_{i=1}^n B_k^2(t_i)$ for any $k = m, \ldots, K_n$, and $X = \left[B_k(t_j)\right]_{j,k}$.

Key questions for statistical asymptotic analysis

Since the number of knots K_n depends on n and $K_n \to \infty$ as $n \to \infty$, it is desired to know how to choose K_n for favorable asymptotic properties:

- uniform convergence on $[0, 1]$, including consistency on the boundary (and in the interior)
- optimal convergence rate

Piecewise Linear Formulation of $\widehat{b}$

Properties of $\widehat{b}$ for fixed K_n

1. Optimality condition:

$$\Lambda_{K_n}\widehat{b} - \bar{y} - D_m^T\lambda = 0, \qquad 0 \le \lambda \perp D_m\widehat{b} \ge 0.$$

2. $\widehat{b} : \mathbb{R}^{K_n+m-1} \to \mathbb{R}^{K_n+m-1}$ is a continuous, piecewise linear function of $\bar{y}$ with 2^{K_n-1} linear selection functions (may write $\widehat{b}$ as $\widehat{b}^{(K_n)}$)
3. $\widehat{b}$ is Lipschitz in $\bar{y}$, and the Lipschitz constant may depend on K_n and a norm (e.g., the ℓ_∞-norm).

Formulation of linear pieces of $\widehat{b}$

1. For each $\bar{y}$, define the index set

$$\alpha := \{i \,|\, (D_m\widehat{b}(\bar{y}))_i = 0\} \subseteq \{1, \ldots, K_n - 1\}$$

2. For each α, a row linearly independent matrix F_α exists such that

$$\widehat{b}(\bar{y}) = F_\alpha^T (F_\alpha \Lambda_{K_n} F_\alpha^T)^{-1} F_\alpha \bar{y}$$

Uniform Lipschitz Property of $\widehat{b}$

Theorem (Uniform Lipschitz property)

The family of piecewise linear functions $\{\widehat{b}^{(K_n)} \mid K_n \in \mathbb{N}\}$ is uniformly Lipschitz in the ℓ_∞-norm, i.e., there exists a constant $L_m > 0$ s.t.

$$\sup_{K_n \in \mathbb{N}} \sup_{u \neq v \in \mathbb{R}^{K_n+m-1}} \frac{\left\|\widehat{b}^{(K_n)}(u) - \widehat{b}^{(K_n)}(v)\right\|_\infty}{\|u-v\|_\infty} \leq L_m$$

Sufficient condition for uniform Lipschitz property

In light of the piecewise linear formulation of $\widehat{b}^{(K_n)}$, it suffices to show

$$\sup_{K_n,\alpha} \|F_\alpha^T (F_\alpha \Lambda_{K_n} F_\alpha^T)^{-1} F_\alpha\|_\infty < \infty$$

T. Lebair and J. Shen. Uniform Lipschitz property of constrained B-splines subject to general shape constraints. 2014.

Proof of Uniform Lipschitz Property

Sketch of the proof

1. Cornerstone result

Theorem (de Boor's Conjecture (Shadrin, 2001))

Let $\mathcal{T} = (t_k)_{k=0}^n$ be a knot sequence on $[a, b]$, let $N_{m,k}^{\mathcal{T},E} := (\widetilde{N}_k)_{k=1}^{n+m-1}$ be B-splines of degree $(m-1)$ defined by $\mathcal{T}$ and some extension E. Let $\widetilde{M}_k := \|\widetilde{N}_k\|_{L_1}^{-1} \cdot \widetilde{N}_k$ for each k, and G be the Grammian matrix given by $G_{ij} = \langle \widetilde{M}_i, \widetilde{N}_j \rangle$. Then $\|G^{-1}\|_\infty$ is bounded independent of a, b, n, and $\mathcal{T}$.

2. Main idea: for any K_n and α, relate $F_\alpha^T (F_\alpha \Lambda_{K_n} F_\alpha^T)^{-1} F_\alpha$ to a suitable Grammian defined by some B-splines with certain knot sequence satisfying the shape constraint, and apply the above theorem to obtain a uniform bound on $\|F_\alpha^T (F_\alpha \Lambda_{K_n} F_\alpha^T)^{-1} F_\alpha\|_\infty$.

A.Y. Shadrin. The L_∞-norm of the L_2-spline projector is bounded independently of the knot sequence: A proof of de Boor's conjecture. *Acta Mathematica*, Vol. 187(1), pp. 59–137, 2001.

Implications of Uniform Lipschitz Property (I)

Uniform convergence and optimal estimation on $\mathcal{S}_H(r, L)$

1. Asymptotic performance in the sup-norm:

$$\mathbb{E}(\|\widehat{f} - f\|_\infty) \quad = \quad O\left(L K_n^{-r} + \sigma \sqrt{\frac{K_n \log n}{n}} \right)$$

2. Optimal rate of convergence in the sup-norm (minimax upper bound): Let $K_n = \left\lceil \left(\frac{L}{\sigma}\right)^{\frac{2}{2r+1}} \left(\frac{n}{\log n}\right)^{\frac{1}{2r+1}} \right\rceil$, then $\exists$ a constant $C > 0$ s.t.

$$\sup_{f \in \mathcal{S}_H(r,L)} \mathbb{E}\left(\|\widehat{f} - f\|_\infty\right) \quad \leq \quad C \cdot L^{\frac{1}{2r+1}} \sigma^{\frac{2r}{2r+1}} \left(\frac{\log n}{n}\right)^{\frac{r}{2r+1}}, \ \forall\, n$$

3. $\widehat{f}$ is consistent on the boundary of $[0, 1]$ as $K_n, n \to \infty$

X. Wang and J. Shen. Uniform convergence and rate adaptive estimation of convex functions via constrained optimization. *SIAM Journal on Control and Optimization*, Vol. 51(4), pp. 2753–2787, 2013.

Implications of Uniform Lipschitz Property (II)

Let $\overline{f}$ be the estimator based on noise free data, i.e.,

$$\overline{f}(t) = \sum_{k=1}^{K_n+m-1} \overline{b}_k B_k(t), \quad \text{where } \overline{b} := \arg \min_{D_m b \geq 0} \frac{1}{2} b^T \Lambda_{K_n} b - b^T \mathbb{E}(\bar{y})$$

Pointwise uniform bound

1. There exist positive constants C_1 and C_2 such that for any $t_0 \in (0, 1)$,

$$\mathbb{E}\big(|\widehat{f}(t_0) - \overline{f}(t_0)|^2\big) \leq C_1 \cdot \sigma^2 \frac{K_n}{n}$$
$$\mathbb{E}\big(|\widehat{f}(t_0) - \overline{f}(t_0)|^4\big) \leq C_2 \cdot \sigma^4 \left(\frac{K_n}{n}\right)^2$$

2. For any $t_0 \in (0, 1)$ and any $m - 1 \leq r' \leq r$,

$$\sup_{f \in \mathcal{S}_H(r,L)} \mathbb{E}\big(|\widehat{f}(t_0) - f(t_0)|^2\big) = O\left(C_1 \cdot \sigma^2 \frac{K_n}{n} + C_1' \frac{L^2}{K_n^{2r'}}\right)$$

Implications of Uniform Lipschitz Property (III)

Adaptive constrained estimation on $\mathcal{S}_H(r, L)$

1. Assume that the Hölder order $r \in [m-1, m]$ is unknown
2. Develop a constrained spline based adaptive estimator that achieves the optimal sup-norm risk:

$$\sup_{r\in[m-1,m]} \sup_{f\in\mathcal{S}_H(r,L)} \mathbb{E}\Big(\|\widehat{f}_{(\hat{r})} - f\|_\infty\Big) \leq \pi_2 \, L^{\frac{1}{2r+1}} \sigma^{\frac{2r}{2r+1}} \Big(\frac{\log n}{n}\Big)^{\frac{r}{2r+1}}.$$

3. Develop an adaptive estimator that achieves the optimal pointwise risk:

$$\sup_{r\in[m-1,m]} \sup_{f\in\mathcal{S}_H(r,L)} \mathbb{E}\Big(|\widetilde{f}(x_0) - f(x_0)|^2\Big) \leq \pi_3 L^{\frac{2}{(2r+1)}} \sigma^{\frac{4r}{(2r+1)}} n^{-\frac{2r}{(2r+1)}}.$$

Minimax Lower Bound

Background

Based on information theoretical results on probability measure distance.

Construction for lower bound

Construct a family of shape constrained functions $f_{j,n}, j = 0, 1, \ldots, M_n$ s.t.

(C1) each $f_{j,n} \in \mathcal{C}_H(r, L)$, $j = 0, 1, \ldots, M_n$;

(C2) once $j \neq k$, $\|f_{j,n} - f_{k,n}\|_\infty \geq 2s_n > 0$, where $s_n \asymp (\log n/n)^{r/(2r+1)}$;

(C3) there exists a fixed constant $c_0 \in (0, 1/8)$ s.t. for all large n,

$$\frac{1}{M_n}\sum_{j=1}^{M_n} K(P_j, P_0) \leq c_0 \log(M_n),$$

where P_j: distribution of $(Y_{j,1}, \ldots, Y_{j,n})$, $Y_{j,i} = f_{j,n}(X_i) + \xi_i$, $i = 1, \ldots, n$ with $X_i = i/n$ and ξ_i: iid r.v., and $K(P, Q)$: Kullback divergence between two probability measures P and Q.

T. Lebair, J. Shen, and X. Wang. Minimax optimal estimation of convex functions in the sup-norm. 2013.

Conclusions

Summary

1. Computation of general shape constrained smoothing splines via a nonsmooth Newton's method
2. Statistical analysis of constrained B-spline estimation: uniform Lipschitz property

Future research

1. Numerical issues: constrained smoothing splines subject to additional constraints
2. Statistical issues: minimax analysis under general constraints
3. Multivariable shape constrained estimation and computation

Thank you!