Spectrahedral Regression

Eliza O'Reilly

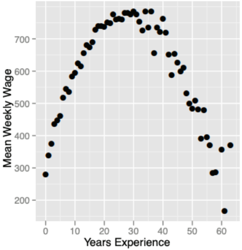
Caltech

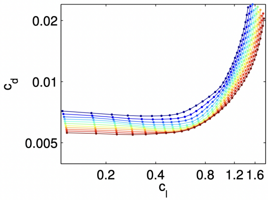
Joint work with Venkat Chandrasekaran

Convex Regression: Motivating Applications

Fitting a convex function to data is natural in many applications

- Economics[1]
  - Natural convex relationships in data
- Engineering Design[2]
  - Ex: Aircraft profile drag, Circuit design
  - Ultimate goal is to optimize function
  - Convexity useful for computational efficiency



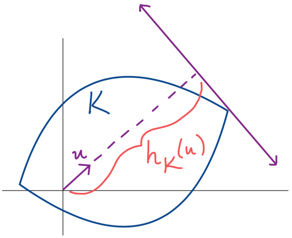


[1] [Afriat, 1967; Varian, 1982, 1984; Hannah and Dunson, 2013]
[2] [Hannah and Dunson, 2012; Hoburg and Abbeel, 2014]

Special Case: Support Function Estimation

- Goal is to reconstruct the convex hull of an object
- Measurements are support function evaluations
- Support function of convex set $K \subset \mathbb{R}^d$ is

$$h_K(u) := \max_{x \in K} \langle x, u \rangle, \quad u \in \mathbb{S}^{d-1}$$



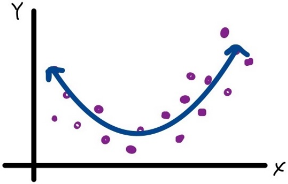
Applications: Radar, MRI, Computed Tomography[3]

[3] [Lele, Kulkani, and Willsky, 1992; Gregor and Rannou, 2002; Prince and Willsky, 1990]

Convex Regression

Goal: Estimate convex function $\hat{f}_n$ from $\{(x_i, y_i)\}_{i=1}^n$ in $\mathbb{R}^d \times \mathbb{R}$ such that

$$y_i \approx \hat{f}_n(x_i)$$

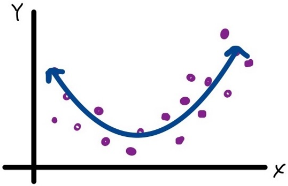


[4] [Prince and Willsky, 1990; Seijo and Sen, 2011]

Convex Regression

Goal: Estimate convex function $\hat{f}_n$ from $\{(x_i, y_i)\}_{i=1}^n$ in $\mathbb{R}^d \times \mathbb{R}$ such that

$$y_i \approx \hat{f}_n(x_i)$$



First consider the least squares estimator (LSE):

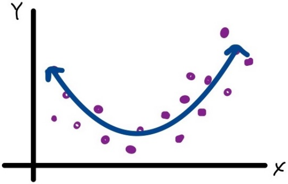
$$\hat{f}_n \in \operatorname{argmin}_{g:\mathbb{R}^d \to \mathbb{R} \text{ is convex}} \frac{1}{n} \sum_{i=1}^{n} (g(x_i) - y_i)^2$$

[4] [Prince and Willsky, 1990; Seijo and Sen, 2011]

Convex Regression

Goal: Estimate convex function $\hat{f}_n$ from $\{(x_i, y_i)\}_{i=1}^n$ in $\mathbb{R}^d \times \mathbb{R}$ such that

$$y_i \approx \hat{f}_n(x_i)$$



First consider the least squares estimator (LSE):

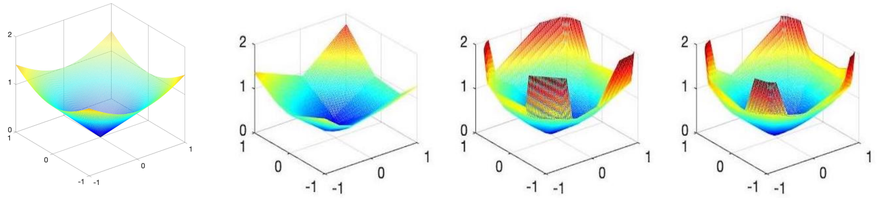
$$\hat{f}_n \in \operatorname{argmin}_{g:\mathbb{R}^d \to \mathbb{R} \text{ is convex}} \frac{1}{n} \sum_{i=1}^{n} (g(x_i) - y_i)^2$$

- Solution to the LSE is the maximum of n affine functions[4]
- Can be computed using convex quadratic programming

[4] [Prince and Willsky, 1990; Seijo and Sen, 2011]

Drawbacks of LSE

- For n input-output pairs LSE is maximum of n affine functions
- Complexity increases with amount of data
- LSE is minimax suboptimal for Lipschitz convex regression ($d \geq 5$) and support function estimation ($d \geq 6$)[5]

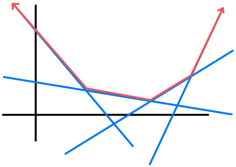


LSE Reconstruction of the function $y = \|x\|_2$ from $n = 20, 50, 200$ noisy measurements

[5] [Guntaboyina, 2012; Kur, Gao, Guntuboyina, and Sen, 2020; Kur, Rakhlin, and Guntuboyina, 2020]

Polyhedral Regression

- f is **m-polyhedral** if $f(x) = \max_{i=1,\ldots,m}\{\langle a_i, x\rangle + b_i\}$
- Constrain LSE over m-polyhedral functions



$$\hat{f}_n \in \operatorname{argmin}_{g:\mathbb{R}^d \to \mathbb{R} \text{ is } m\text{-polyhedral}} \frac{1}{n}\sum_{i=1}^{n}(g(x_i) - y_i)^2$$

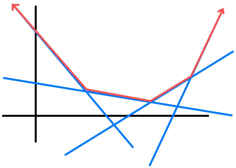
[6] [Guntaboyina, 2012; Han and Wellner, 2016]

[7] [Magnani and Boyd, 2009; Hannah and Dunson, 2013; Balazs et al, 2015; Ghosh et al., 2020]

Polyhedral Regression

- f is **m-polyhedral** if $f(x) = \max_{i=1,\dots,m}\{\langle a_i, x\rangle + b_i\}$
- Constrain LSE over m-polyhedral functions

$$\hat{f}_n \in \operatorname{argmin}_{g:\mathbb{R}^d \to \mathbb{R} \text{ is } m\text{-polyhedral}} \frac{1}{n}\sum_{i=1}^{n}(g(x_i) - y_i)^2$$

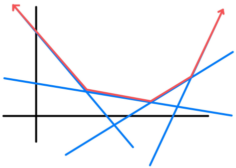


- Obtains minimax rates[6]

[6] [Guntaboyina, 2012; Han and Wellner, 2016]

[7] [Magnani and Boyd, 2009; Hannah and Dunson, 2013; Balazs et al, 2015; Ghosh et al., 2020]

Polyhedral Regression



- f is **m-polyhedral** if $f(x) = \max_{i=1,\ldots,m}\{\langle a_i, x\rangle + b_i\}$
- Constrain LSE over m-polyhedral functions

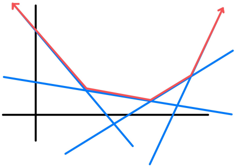
$$\hat{f}_n \in \operatorname{argmin}_{g:\mathbb{R}^d \to \mathbb{R} \text{ is } m\text{-polyhedral}} \frac{1}{n}\sum_{i=1}^{n}(g(x_i) - y_i)^2$$

- Obtains minimax rates[6]
- Tractable methods for computing estimator [7]

[6] [Guntaboyina, 2012; Han and Wellner, 2016]
[7] [Magnani and Boyd, 2009; Hannah and Dunson, 2013; Balazs et al, 2015; Ghosh et al., 2020]

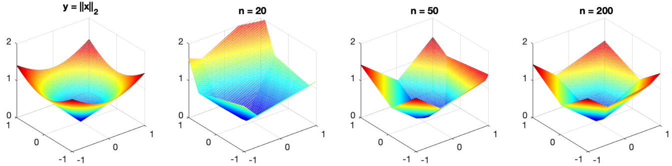
Polyhedral Regression

- f is **m-polyhedral** if $f(x) = \max_{i=1,\ldots,m}\{\langle a_i, x\rangle + b_i\}$
- Constrain LSE over m-polyhedral functions



$$\hat{f}_n \in \operatorname{argmin}_{g:\mathbb{R}^d \to \mathbb{R} \text{ is } m\text{-polyhedral}} \frac{1}{n}\sum_{i=1}^{n}(g(x_i) - y_i)^2$$

- Obtains minimax rates[6]
- Tractable methods for computing estimator [7]



- Drawback: polyhedral approximations of non-polyhedral functions and sets

[6] [Guntaboyina, 2012; Han and Wellner, 2016]

[7] [Magnani and Boyd, 2009; Hannah and Dunson, 2013; Balazs et al, 2015; Ghosh et al., 2020]

Spectrahedral Regression [O. & Chandrasekaran, '22]

- $\mathbb{S}^m$: $m \times m$ real symmetric matrices
- A function f is **m-spectrahedral** if for some $A_0, \ldots, A_d \in \mathbb{S}^m$,

$$f(x) = \lambda_{\mathbf{max}} \left(\sum_{i=1}^{d} x_i A_i + A_0 \right)$$

- Constrain LSE over m-spectrahedral functions

$$\hat{f}_n \in \mathrm{argmin}_{g:\mathbb{R}^d \to \mathbb{R} \text{ is } m\text{-spectrahedral}} \frac{1}{n} \sum_{i=1}^{n} (g(x_i) - y_i)^2$$

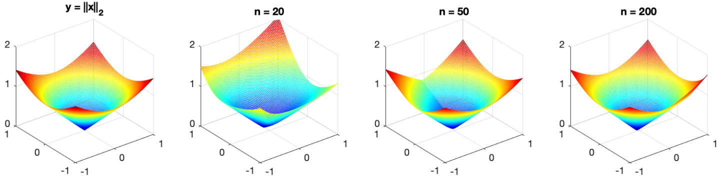
Spectrahedral Regression [O. & Chandrasekaran, '22]

- $\mathbb{S}^m$: $m \times m$ real symmetric matrices
- A function f is **m-spectrahedral** if for some $A_0, \ldots, A_d \in \mathbb{S}^m$,

$$f(x) = \lambda_{\mathbf{max}} \left(\sum_{i=1}^{d} x_i A_i + A_0 \right)$$

- Constrain LSE over m-spectrahedral functions

$$\hat{f}_n \in \operatorname{argmin}_{g:\mathbb{R}^d \to \mathbb{R} \text{ is } m\text{-spectrahedral}} \frac{1}{n} \sum_{i=1}^{n} (g(x_i) - y_i)^2$$



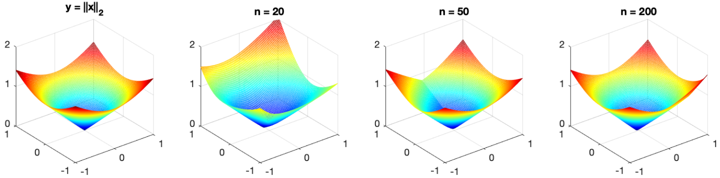
Spectrahedral Regression [O. & Chandrasekaran, '22]

- $\mathbb{S}^m$: $m \times m$ real symmetric matrices
- A function f is **m-spectrahedral** if for some $A_0, \ldots, A_d \in \mathbb{S}^m$,

$$f(x) = \lambda_{\max}\left(\sum_{i=1}^{d} x_i A_i + A_0\right)$$

- Constrain LSE over m-spectrahedral functions

$$\hat{f}_n \in \operatorname{argmin}_{g:\mathbb{R}^d \to \mathbb{R} \text{ is } m\text{-spectrahedral}} \frac{1}{n}\sum_{i=1}^{n}(g(x_i) - y_i)^2$$

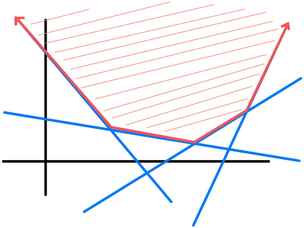


Linear → Semidefinite programming

Can optimize *m*-polyhedral function

$$f(x) = \max_{i=1,\ldots,m} \{\langle a_i, x\rangle + b_i\}$$

using linear programming



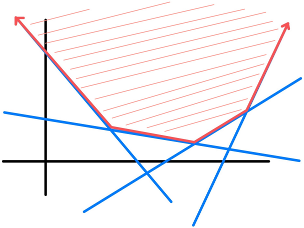
- $\{(x, y) \in \mathbb{R}^{d+1} : f(x) \leq y\}$ is a **polyhedron**

  $\{(x, y) \in \mathbb{R}^{d+1} : 0 \leq y - \langle a_i, x\rangle - b_i,\ i = 1, \ldots, m\}$

Linear → Semidefinite programming

Can optimize m-polyhedral function

$$f(x) = \max_{i=1,\ldots,m} \{\langle a_i, x\rangle + b_i\}$$



using linear programming

- $\{(x,y) \in \mathbb{R}^{d+1} : f(x) \le y\}$ is a **polyhedron**

  $$\{(x,y) \in \mathbb{R}^{d+1} : 0 \le y - \langle a_i, x\rangle - b_i,\ i = 1,\ldots,m\}$$

Can optimize m-spectrahedral function $f(x) = \lambda_{\max}\left(\sum_{i=1}^{d} x_i A_i + A_0\right)$ using semidefinite programming

- $\{(x,y) \in \mathbb{R}^{d+1} : f(x) \le y\}$ is a **spectrahedron**

$$\{(x,y) \in \mathbb{R}^{d+1} : 0 \preccurlyeq yI - \sum_{i=1}^{d} x_i A_i - A_0\}$$

Spectrahedral Regression: Average Weekly Wages

Data set: 1988 Current Population Survey: 25,361 records of weekly wages with (i) Experience (ii) Education

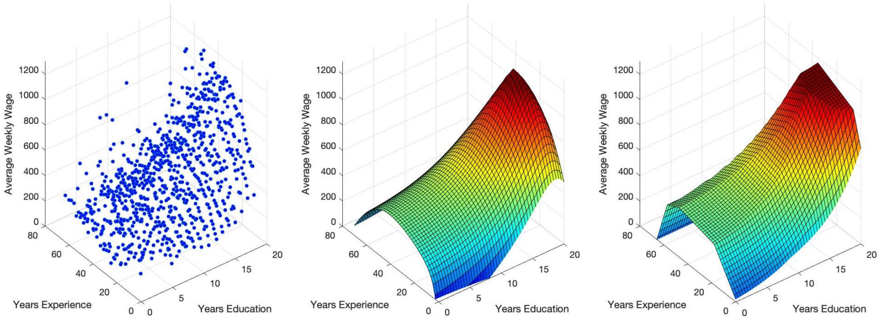
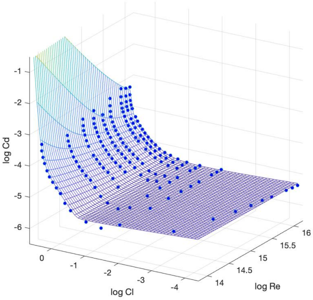


Figure: Spectrahedral ($m = 3$) and Polyhedral ($m = 6$) estimators of average weekly wages versus years experience and education

Spectrahedral Regression: Aircraft Design

Data set: XFOIL simulated data of airplane wing profile drag coefficient as a function of the Reynolds number and lift coefficient



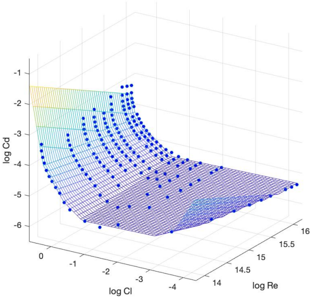


Figure: Spectrahedral ($m = 3$) and Polyhedral ($m = 6$) estimators

Support Function Estimation [Soh & Chandrasekaran, 21]

- Polyhedral regression $\rightarrow$ Constrain LSE over polytopes with m vertices:

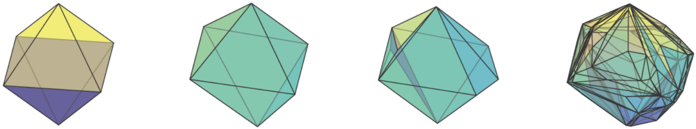


Figure: $m = 6, m = 12$ polytope, and LSE reconstructions of the unit ℓ_1-ball from 200 noisy support function measurements

Support Function Estimation [Soh & Chandrasekaran, 21]

- Polyhedral regression $\rightarrow$ Constrain LSE over polytopes with m vertices:

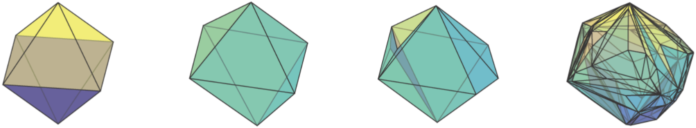


Figure: $m = 6$, $m = 12$ polytope, and LSE reconstructions of the unit ℓ_1-ball from 200 noisy support function measurements

- Spectrahedral regression $\rightarrow$ Constrain LSE over m-spectratopes
- m-spectratopes are linear images of an m-dimensional spectroplex:

$$\{X \in \mathbb{S}^m : X \succeq 0, \langle X, I \rangle = 1\}$$

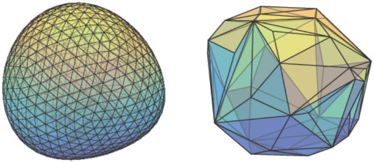
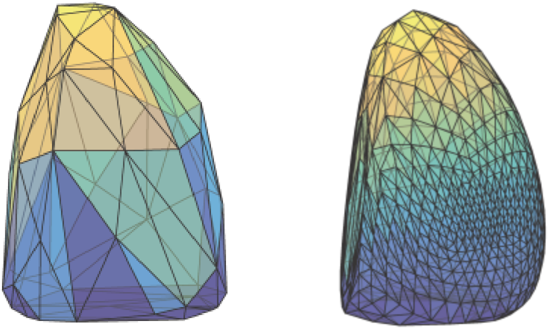


Figure: $m = 3$ spectratope and LSE reconstructions of ℓ_2-ball from 50 noisy support function measurements

Support Function Estimation: Lung Reconstruction



(a) $n = 50$, LSE (b) $n = 50, m = 3$

Figure: LSE and m-spectrahedral ($m = 3$) lung reconstruction[8]

[8] [Soh and Chandrasekaran, 2021]

Block Spectrahedral Regression

- f is **(m, k)-spectrahedral** if $f(x) = \lambda_{\max}\left(\sum_{i=1}^{d} x_i A_i + A_0\right)$ where $A_0, \ldots, A_d \in \mathbb{S}_k^m$ are block-diagonal with blocks of size k

Block Spectrahedral Regression

- f is **(m, k)-spectrahedral** if $f(x) = \lambda_{\max}\left(\sum_{i=1}^{d} x_i A_i + A_0\right)$ where $A_0, \ldots, A_d \in \mathbb{S}_k^m$ are block-diagonal with blocks of size k

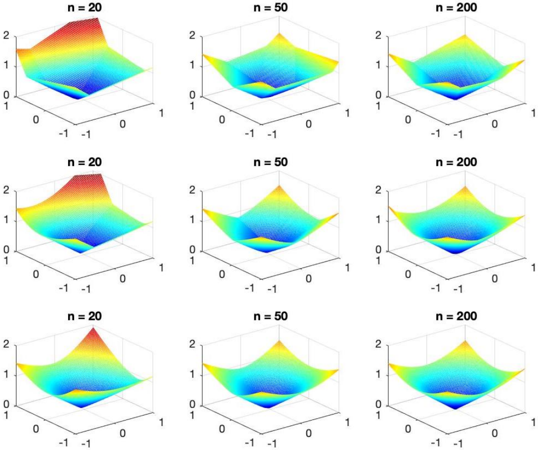


Figure: Polyhedral $(m, k) = (6, 1)$, block spectrahedral $(m, k) = (4, 2)$, and spectrahedral $m = 3$ reconstructions of $y = \|x\|_2$.

Block Spectrahedral Regression

- f is **(m, k)-spectrahedral** if $f(x) = \lambda_{\max}\left(\sum_{i=1}^{d} x_i A_i + A_0\right)$ where $A_0, \ldots, A_d \in \mathbb{S}_k^m$ are block-diagonal with blocks of size k

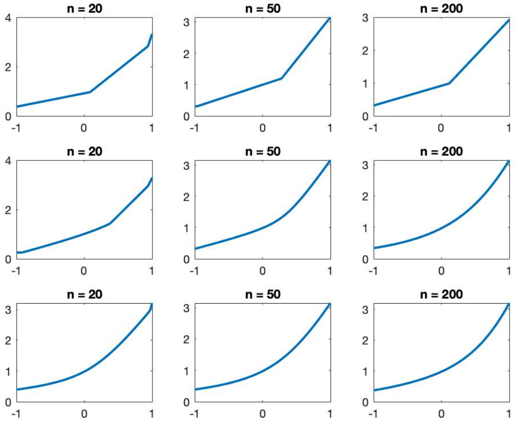


Figure: Polyhedral $(m, k) = (6, 1)$, block spectrahedral $(m, k) = (4, 2)$, and spectrahedral $m = 3$ reconstructions of $y = \exp(\langle x, b \rangle)$.

Questions

1. How do you compute the spectrahedral estimator?

2. What is the expressive power of spectrahedral functions?

Questions

1. **How do you compute the spectrahedral estimator?**

2. What is the expressive power of spectrahedral functions?

Alternating Minimization

- For $\mathcal{A} = (A_0, \ldots, A_d) \in (\mathbb{S}_k^m)^{d+1}$, define $\mathcal{A}[\xi] = \sum_{j=0}^{d} \xi_j A_j$
- Let $\xi^{(i)} = (x^{(i)}, 1) \in \mathbb{R}^{d+1}$. Want to compute:

$$\hat{\mathcal{A}} \in \operatorname{argmin}_{\mathcal{A} \in (\mathbb{S}_k^m)^{d+1}} \frac{1}{n} \sum_{i=1}^{n} \left[y^{(i)} - \lambda_{\max} \left(\mathcal{A}[\xi^{(i)}] \right) \right]^2 ,$$

Alternating Minimization

- For $\mathcal{A} = (A_0, \ldots, A_d) \in (\mathbb{S}_k^m)^{d+1}$, define $\mathcal{A}[\xi] = \sum_{j=0}^d \xi_j A_j$
- Let $\xi^{(i)} = (x^{(i)}, 1) \in \mathbb{R}^{d+1}$. Want to compute:

$$\hat{\mathcal{A}} \in \text{argmin}_{\mathcal{A} \in (\mathbb{S}_k^m)^{d+1}} \frac{1}{n} \sum_{i=1}^n \left[y^{(i)} - \lambda_{\max} \left(\mathcal{A}[\xi^{(i)}] \right) \right]^2,$$

Input: Data collection $\{(x^{(i)}, y^{(i)})\}_{i=1}^n$; initialization $\mathcal{A} \in (\mathbb{S}_k^m)^{d+1}$
Algorithm: Repeat until convergence

- **Step 1**: Update optimal eigenvectors $u^{(i)} \leftarrow \lambda_{\max}(\mathcal{A}[\xi^{(i)}])$
- **Step 2**: Update $\mathcal{A}$ by solving

$$\mathrm{argmin}_{\mathcal{A} \in (\mathbb{S}_k^m)^{d+1}} \frac{1}{n} \sum_{i=1}^n \left(y^{(i)} - \langle u^{(i)}, \mathcal{A}[\xi^{(i)}] \rangle \right)^2,$$

i.e. $\mathcal{A}^+ \leftarrow (\Xi_{\mathcal{A}}^T \Xi_{\mathcal{A}})^{-1} \Xi_{\mathcal{A}}^T y$, where $\Xi_{\mathcal{A}}^T = (\xi^{(1)} \otimes u^{(1)} | \cdots | \xi^{(n)} \otimes u^{(n)})$

Output: Final iterate $\mathcal{A}$

Local Convergence Guarantee for AM

Assumption: $\{(x_i, y_i)\}_{i=1}^n$ i.i.d. samples from $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ such that

$$X \sim \mathcal{N}(0, \mathcal{I}), \quad Y = \lambda_{\max}(\mathcal{A}_*[\xi]) + \varepsilon, \quad \xi = (X, 1), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Local Convergence Guarantee for AM

Assumption: $\{(x_i, y_i)\}_{i=1}^n$ i.i.d. samples from $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ such that

$$X \sim \mathcal{N}(0, \mathcal{I}), \quad Y = \lambda_{\max}(\mathcal{A}_*[\xi]) + \varepsilon, \quad \xi = (X, 1), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Suppose that the true parameter $\mathcal{A}_* \in (\mathbb{S}_k^m)^{d+1}$ satisfies:

$$\inf_{u \in \mathbb{S}^d} \lambda_1(\mathcal{A}_*[u]) - \lambda_2(\mathcal{A}_*[u]) := \kappa > 0$$

Local Convergence Guarantee for AM

Assumption: $\{(x_i, y_i)\}_{i=1}^n$ i.i.d. samples from $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ such that

$$X \sim \mathcal{N}(0, \mathcal{I}), \quad Y = \lambda_{\max}(\mathcal{A}_*[\xi]) + \varepsilon, \quad \xi = (X, 1), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Suppose that the true parameter $\mathcal{A}_* \in (\mathbb{S}_k^m)^{d+1}$ satisfies:

$$\inf_{u \in \mathbb{S}^d} \lambda_1(\mathcal{A}_*[u]) - \lambda_2(\mathcal{A}_*[u]) := \kappa > 0$$

Theorem (O. and Chandrasekaran)

If the initial parameter $\mathcal{A}^{(0)}$ satisfies

$$\|\mathcal{A}^{(0)} - \mathcal{A}_*\|_F^2 \leq \frac{c_1 \kappa^2}{(d+1)m},$$

and n is large enough, then the error at all iterations $t \geq 1$ simultaneously satisfies

$$\|\mathcal{A}^{(t)} - \mathcal{A}_*\|_F^2 \leq \left(\frac{3}{4}\right)^t \|\mathcal{A}^{(0)} - \mathcal{A}_*\|_F^2 + \frac{c_2 m^3 (d+1) \sigma^2 \log(n)^2}{n}$$

with high probability, where c_1 and c_2 are absolute constants.

Local Convergence Guarantee for AM (Version 2)

Assumption: $\{(x_i, y_i)\}_{i=1}^n$ i.i.d. samples from $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ such that

$$\|X\|_\infty \leq \eta, \quad Y = \lambda_{\mathsf{max}}(\mathcal{A}_*[\xi]) + \varepsilon, \quad \xi = (X, 1), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Local Convergence Guarantee for AM (Version 2)

Assumption: $\{(x_i, y_i)\}_{i=1}^n$ i.i.d. samples from $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ such that

$$\|X\|_\infty \leq \eta, \quad Y = \lambda_{\max}(\mathcal{A}_*[\xi]) + \varepsilon, \quad \xi = (X, 1), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{A}_*[\xi] = \begin{bmatrix} \mathcal{A}_*^{(1)}[\xi] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{A}_*^{(m/k)}[\xi] \end{bmatrix}$$

Local Convergence Guarantee for AM (Version 2)

Assumption: $\{(x_i, y_i)\}_{i=1}^n$ i.i.d. samples from $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ such that

$$\|X\|_\infty \leq \eta, \quad Y = \lambda_{\max}(\mathcal{A}_*[\xi]) + \varepsilon, \quad \xi = (X, 1), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{A}_*[\xi] = \begin{bmatrix} \mathcal{A}_*^{(1)}[\xi] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{A}_*^{(m/k)}[\xi] \end{bmatrix}$$

- There exists $\kappa > 0$ and $\delta \in (0, 1)$ such that for all $j, \ell \in \{1, \dots, m/k\}$,

$$\inf_{j \neq \ell} \mathbb{E}\left[\left|\lambda_1(\mathcal{A}_*^{(j)}[\xi]) - \lambda_1(\mathcal{A}_*^{(\ell)}[\xi])\right|\right] \geq \frac{m\kappa}{k\delta}.$$

and if $k \geq 2$, for all $j \in \{1, \dots, m/k\}$

$$\inf_{j=1,\dots,m/k} \inf_{u \in \mathbb{S}^d} \lambda_1(\mathcal{A}_*^{(j)}[u]) - \lambda_2(\mathcal{A}_*^{(j)}[u]) := \kappa > 0.$$

Local Convergence Guarantee for AM (Version 2)

Assumption: $\{(x_i, y_i)\}_{i=1}^n$ i.i.d. samples from $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ such that

$$\|X\|_\infty \leq \eta, \quad Y = \lambda_{\max}(\mathcal{A}_*[\xi]) + \varepsilon, \quad \xi = (X, 1), \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\mathcal{A}_*[\xi] = \begin{bmatrix} \mathcal{A}_*^{(1)}[\xi] & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathcal{A}_*^{(m/k)}[\xi] \end{bmatrix}$$

- There exists $\kappa > 0$ and $\delta \in (0, 1)$ such that for all $j, \ell \in \{1, \dots, m/k\}$,

  $$\inf_{j \neq \ell} \mathbb{E}\left[|\lambda_1(\mathcal{A}_*^{(j)}[\xi]) - \lambda_1(\mathcal{A}_*^{(\ell)}[\xi])|\right] \geq \frac{m\kappa}{k\delta}.$$

  and if $k \geq 2$, for all $j \in \{1, \dots, m/k\}$

  $$\inf_{j=1,\dots,m/k} \inf_{u \in \mathbb{S}^d} \lambda_1(\mathcal{A}_*^{(j)}[u]) - \lambda_2(\mathcal{A}_*^{(j)}[u]) := \kappa > 0.$$

- Assume that there is a constant $c > 0$ such that for all $\mathcal{A} \neq \mathcal{B} \in (\mathbb{S}^k)^{d+1}$,

  $$\mathbb{P}\left(|\lambda_1(\mathcal{A}[\xi]) - \lambda_1(\mathcal{B}[\xi])| \leq \rho \mathbb{E}\left[|\lambda_1(\mathcal{A}[\xi]) - \lambda_1(\mathcal{B}[\xi])|\right]\right) \leq c\rho, \quad \text{for all } \rho > 0.$$

Local Convergence Guarantee for AM (Version 2)

Theorem (O. and Chandrasekaran)

If the initial parameter choice $\mathcal{A}_0$ satisfies

$$\|\mathcal{A}^{(0)} - \mathcal{A}_*\|_F^2 \leq \frac{c_1 \kappa^2 k^3}{(d+1)^3 m^5},$$

and n is large enough, then the error at all iterations $t \geq 1$ simultaneously satisfies

$$\|\mathcal{A}^{(t)} - \mathcal{A}_*\|_F^2 \leq \left(\frac{3}{4}\right)^t \|\mathcal{A}^{(0)} - \mathcal{A}_*\|_F^2 + \frac{c_2 m^3 (d+1) \sigma^2 \log(n)^2}{n},$$

with probability $\to 1 - 3\delta$ as $n \to \infty$.

Questions

1. How do you compute the spectrahedral estimator?

2. **What is the expressive power of spectrahedral functions?**

Expressiveness of Spectrahedral Functions

How well do (m, k)-spectrahedral functions approximate Lipschitz convex functions?

- A result of Dudley (1974) implies that for **polyhedral** functions,

$$\sup_{\substack{f:\Omega\to\mathbb{R}\text{ convex}\\ \text{and } L\text{-Lipschitz}}} \inf_{g \text{ is } m\text{-polyhedral}} \|g - f\|_\infty = O(m^{-\frac{2}{d}})$$

Expressiveness of Spectrahedral Functions

How well do (m, k)-spectrahedral functions approximate Lipschitz convex functions?

- A result of Dudley (1974) implies that for **polyhedral** functions,

$$\sup_{\substack{f:\Omega\to\mathbb{R}\text{ convex}\\ \text{and } L\text{-Lipschitz}}} \inf_{g \text{ is } m\text{-polyhedral}} \|g - f\|_\infty = O(m^{-\frac{2}{d}})$$

Theorem (O. and Chandrasekaran)

Suppose $k_m = O(m^t)$ *for* $t \in [0, 1]$. *For all* $\varepsilon > 0$,

$$O\left(m^{-\frac{2(1+t)}{d} - \varepsilon}\right) \leq \sup_{\substack{f:\Omega\to\mathbb{R}\text{ convex}\\ \text{and } L\text{-Lipschitz}}} \inf_{g \text{ is } (m, k_m)\text{-spectrahedral}} \|g - f\|_\infty \leq O\left(m^{-\frac{2}{d}}\right)$$

- For constant k, approximation rate is same as for m-polyhedral functions

Proof Idea: Use Statistical Risk Bound

- Define

$$\hat{f}_{m,k}^{(n)} \in \operatorname{argmin}_{g \text{ is } (m,k)\text{-spectrahedral}} \frac{1}{n} \sum_{i=1}^{n} (g(x_i) - y_i)^2,$$

where $\{(x_i, y_i)\}_{i=1}^n$ are i.i.d. samples of a random pair $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ s.t.

$$Y = f(X) + \varepsilon$$

Theorem (O. and Chandrasekaran)

$$\mathbb{E}[(\hat{f}_{m,k}^{(n)}(X) - f(X))^2] \leq O\left(\inf_{g \text{ is } (m,k)\text{-spectrahedral}} \|g - f\|_\infty^2 + \frac{km \log(n)}{n} \right)$$

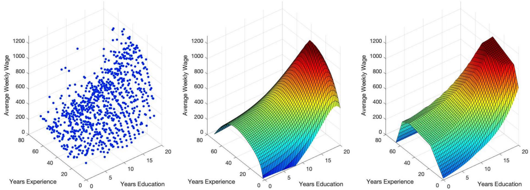
- Minimax rate for class of convex and L-Lipschitz functions[9] implies

$$\sup_{\substack{f:\Omega \to \mathbb{R} \text{ convex} \\ \text{and } L\text{-Lipschitz}}} \mathbb{E}[(\hat{f}_{m,k}^{(n)}(X) - f(X))^2] \geq O(n^{-\frac{4}{d+4}})$$

[9] [Balázs, György, and Szepesvári, 2015]

Summary

- **Spectrahedral regression** is a new approach for fitting convex functions to data that generalizes polyhedral regression
- Returns convex estimators that exhibit **both smooth and singular** features
- Expressiveness of spectrahedral functions has implications for how well semidefinite relaxations approximate general convex optimization
- **Empirical evidence**: m-spectrahedral regression performs comparably to $m(m+1)/2$-polyhedral regression



Future Work

- Guidance for parameter selection and tuning
- Computational Guarantees: initialization, extend other approaches for polyhedral regression
- Approximation power of (m, k)-spectrahedral functions
- Other shape-constrained regression, density estimation applications

Thank you!

Questions?