Nonparametric Convex Regression\textsuperscript{1}

Bodhisattva Sen  
Department of Statistics  
Columbia University, New York  
bodhi@stat.columbia.edu

University of North Carolina  
Chapel Hill  
4 April, 2014

\textsuperscript{1}Supported by NSF grant DMS-1150435 (CAREER)
Outline

1. Nonparametric convex regression: Introduction
2. Testing against a linear model
3. Theoretical results: Risk bounds
1 Nonparametric convex regression: Introduction

2 Testing against a linear model

3 Theoretical results: Risk bounds
What is convex regression?

Regression model of the form

$$Y_i = \phi_0(X_i) + \epsilon_i, \quad \text{for } i = 1, \ldots, n,$$

where

- $\phi_0 : \mathbb{R}^d \rightarrow \mathbb{R}, \ d \geq 1$, is an unknown convex function
- $X_i$ is an $X$-valued vector ($X \subseteq \mathbb{R}^d$ is a closed convex set)
- $\epsilon_i$ is a random variable with $\mathbb{E}(\epsilon_i|X_i) = 0$ and $\text{Var}(\epsilon_i) < \infty$

Goal: Estimate $\phi_0$ nonparametrically, assuming the convexity
What is a convex function?

- $f : \mathbb{R}^d \to \mathbb{R}$ is convex iff for all $X_1, X_2 \in \mathbb{R}^d$ and $t \in [0, 1]$,

  $$f(tX_1 + (1 - t)X_2) \leq t f(X_1) + (1 - t) f(X_2).$$
Arises frequently in *economics*, operations research, financial engineering, e.g., production functions are known to be *concave & ↑*

Approximating a function (objective function, constraint) for a convex optimization problem

Convexity is probably the simplest *shape* constraint that has a natural generalization to multi-dimension.

Two further applications of the method

- $\phi_0$ is *affine* (i.e. $\phi_0(X) = \alpha + \beta^T X$) iff $\phi_0$ both convex and concave
  - Test the goodness-of-fit of a linear model (Sen and Meyer (2013))

- Any *smooth* function on a compact domain can be represented as the difference of two convex functions
  - Estimate any smooth function (Work in progress)
Convex regression when $d = 1$

- **Model**
  \[ Y_i = \phi_0(x_i) + \epsilon_i, \quad \text{for } i = 1, 2, \ldots, n \]

- $\phi_0 : \mathbb{R} \to \mathbb{R}$ is an unknown **convex** (or concave) function
- $x_1 < x_2 < \cdots < x_n$ are known constants; $\epsilon_1, \ldots, \epsilon_n$ mean 0 errors.

- Want to estimate $\phi_0$ **nonparametrically**, given $\phi_0$ is **convex/concave**
The estimation procedure: least squares

- Model

\[ Y_i = \phi_0(X_i) + \epsilon_i \quad \text{for } i = 1, \ldots, n; \quad X_i \in \mathbb{R}^d; \quad d \geq 1 \]

- We estimate \( \phi_0 \) via the method of least squares

\[ \hat{\phi} = \arg\min_{\psi: \mathbb{R}^d \to \mathbb{R} \text{ convex}} \sum_{i=1}^{n} \{ Y_i - \psi(X_i) \}^2. \]

- \( \hat{\theta} = (\hat{\phi}(x_1), \ldots, \hat{\phi}(x_n)) \) is unique (Why? next slide ...)

- Long history: Hildreth (1954) ...

- Fully automated; does not require the choice of tuning parameter(s)
Sequence model: \( Y_i = \phi_0(X_i) + \epsilon_i, \; i = 1, 2, \ldots, n, \) is equivalent to
\[
Y = \theta_0 + \epsilon
\]

\( Y = (Y_1, \ldots, Y_n)^T, \; \theta_0 = (\phi_0(X_1), \ldots \phi_0(X_n))^T, \; \epsilon = (\epsilon_1, \ldots, \epsilon_n)^T \)

\[
\min_{\psi \text{ convex}} \sum_{i=1}^n \{Y_i - \psi(X_i)\}^2 \equiv \min_{\theta \in \mathcal{K}} \sum_{i=1}^n (Y_i - \theta_i)^2
\]

\( \mathcal{K} = \{\theta \in \mathbb{R}^n : \exists \psi \text{ convex s.t.} \; \psi(X_i) = \theta_i, \; \forall i = 1, \ldots, n\} \)

\( \hat{\theta} \) is the projection of \( Y \) onto \( \mathcal{K} \), i.e.,
\[
\hat{\theta} := \Pi(Y|\mathcal{K}) = \arg\min_{\theta \in \mathcal{K}} \|Y - \theta\|^2
\]
• $\mathcal{K}$ is a closed convex cone in $\mathbb{R}^n$ ($\mathcal{T}$ is a convex cone iff $\alpha_1 \theta_1 + \alpha_2 \theta_2 \in \mathcal{T}, \forall \alpha_1, \alpha_2 > 0 \ & \theta_1, \theta_2 \in \mathcal{T}$)

• The projection theorem gives the existence and uniqueness of $\hat{\theta}$.

Computation when $d = 1$

• $\mathcal{K}$ has a straightforward characterization when $d = 1$, i.e.,

$$\mathcal{K} = \left\{ \theta \in \mathbb{R}^n : \frac{\theta_2 - \theta_1}{x_2 - x_1} \leq \frac{\theta_3 - \theta_2}{x_3 - x_2} \leq \cdots \leq \frac{\theta_n - \theta_{n-1}}{x_n - x_{n-1}} \right\}$$

• Quadratic program with $(n-2)$ linear constraints

• $\hat{\phi}$ ends up being a piece-wise linear function constructed by linearly interpolating the pairs $(x_1, \hat{\theta}_1), \ldots, (x_n, \hat{\theta}_n)$

What happens when $d > 1$?
• If \( f \) is \textit{differentiable}, then \( f \) is convex iff

\[
f(x) + \nabla f(x)^\top (y - x) \leq f(y), \quad \text{for all } x, y \in \mathbb{R}^d
\]

where \( \nabla f(x) \) is the \textit{gradient} of \( f \) at \( x \).

• The first-order Taylor approximation is a \textit{global under-estimator} of \( f \).

• In fact, for \textit{every} convex function \( f \), there exists \( \xi_x \in \mathbb{R}^d \) such that

\[
f(x) + \xi_x^\top (y - x) \leq f(y) \quad \text{for all } x, y \in \mathbb{R}^d.
\]

• \( \xi_x \) is called a \textit{sub-gradient} at \( x \).
Sub-differentials and sub-gradients

- The sub-differential of \( f \) at \( x \in \mathbb{R}^d \) is given by
\[
\partial f(x) := \{ \xi \in \mathbb{R}^d : f(x) + \xi^T(y - x) \leq f(y) \ \forall y \in \mathbb{R}^d \}.
\]

- The elements of the sub-differential are called sub-gradients.

- Example: If \( f(x) = \|x\| \), \( \partial f(0) = \{ \xi \in \mathbb{R}^d : \|\xi\| \leq 1 \} \).

- \( f \) is differentiable at \( x \) iff \( \partial f(x) = \{ \nabla f(x) \} \).
Computation of $\hat{\theta}$

The vector $\hat{\theta}$ can now be computed by solving the following quadratic program with linear constraints:

$$
\begin{align*}
\min & \quad \|Y - \theta\|^2 \\
\text{subject to} & \quad \theta_i + \xi_i^T (X_j - X_i) \leq \theta_j \quad \forall \ i, j = 1, \ldots, n \\
& \quad \xi_1, \ldots, \xi_n \in \mathbb{R}^d, \theta \in \mathbb{R}^n
\end{align*}
$$

This problem has a unique solution in $\theta$ but not in the $\xi_j$’s (i.e., for any two solutions $(\xi_1, \ldots, \xi_n, \theta)$ and $(\tau_1, \ldots, \tau_n, \zeta)$ we have $\theta = \zeta$).

Quadratic program with $n(n-1)$ linear constraints and $n(d+1)$ variables.

Can use MOSEK, cvx or any off-the-shelf solver for $n \sim 200$

Scalable algorithm? (work in progress; can work with $n \sim 5000$)
Estimator at a point $X$: $\hat{\phi}(X) = \max_{i=1,...,n}\{\hat{\theta}_i + (X - X_i)^\top \hat{\xi}_i\}$

Figure: The scatter plot and LSE: (a) $\phi_0(X) = \|X\|^2$ (left panel); (b) $\phi_0(X) = -X_1 + X_2$ (right panel). For both cases, $n = 256$ and $\epsilon \sim \mathcal{N}(0, 1)$.

- $\hat{\phi}$ is a polyhedral convex function.

- Belgium labour firms example (MATLAB figures)
The LSE is *consistent* in *fixed* and *stochastic* design settings.

**Consistency theorem (Seijo and Sen (2011))**

Under appropriate conditions we have:

(i) \( \mathbb{P}(\sup_{X \in \mathcal{X}} \{|\hat{\phi}(X) - \phi_0(X)|\} \to 0 \text{ for any compact set } X \subset \mathcal{X}^\circ) = 1. \)

(ii) Denoting by \( \mathcal{B} \) the unit ball (w.r.t. the Euclidian norm) we have

\[
\mathbb{P}(\partial \hat{\phi}(X) \subset \partial \phi_0(X) + \epsilon \mathcal{B} \text{ almost always}) = 1 \quad \forall \epsilon > 0, \forall X \in \mathcal{X}^\circ.
\]

(iii) If \( \phi_0 \) is differentiable on \( \mathcal{X}^\circ \), then

\[
\mathbb{P} \left( \sup_{\xi \in \partial \hat{\phi}(X)} \|\xi - \nabla \phi_0(X)\| \to 0 \text{ for any compact set } X \subset \mathcal{X}^\circ \right) = 1.
\]
1. Nonparametric convex regression: Introduction

2. Testing against a linear model

3. Theoretical results: Risk bounds
Testing against a linear model

- Let \( Y := (Y_1, Y_2, \ldots, Y_n) \in \mathbb{R}^n \) and

\[ Y \sim \mathcal{N}(\theta_0, \sigma^2 I_n), \quad \sigma^2 > 0 \]

- Problem: Test \( H_0 : \theta_0 \in S \) where

\[ S := \{ \theta \in \mathbb{R}^n : \theta = X\beta, \beta \in \mathbb{R}^k, \text{ for some } k \geq 1 \} \]

and \( X \) is the *design* matrix.

- Important problem with a long history

- Most methods fit the *parametric* and a *nonparametric* model to estimate the regression function

- *Compare* the *two fits* to come up with a test statistic
Drawbacks of most methods

- Involves the *choice* of delicate tuning parameters (e.g., bandwidths)

- Calibration (i.e., the level) of the test is usually done by resampling the test statistics (bootstrap), and thus only *approximate*.

We develop a *likelihood ratio test* for $H_0$ using ideas from shape restricted regression.

- Fully *automated*, no tuning parameters required

- Test has *exact* level

- *Power* of the test goes to 1, under (almost) *all* alternatives
Testing against linear regression

\[ Y_i = \phi_0(X_i) + \epsilon_i, \quad i = 1, 2, \ldots, n. \]

- \( \phi_0 : \mathbb{R}^d \to \mathbb{R}, \ d \geq 1, \) is an unknown smooth function

- Design points \( X_1, X_2, \ldots, X_n \) in \( \mathbb{R}^d; \epsilon_1, \ldots, \epsilon_n \) are i.i.d. \( \mathcal{N}(0, \sigma^2) \)

- Test \( H_0 : \phi_0 \) is affine, i.e., \( \phi_0(X) = a + b^\top X, \) for some unknown \( a \in \mathbb{R}, b \in \mathbb{R}^d \)

Thus, \( Y \sim \mathcal{N}(\theta_0, \sigma^2 I_n) \) where \( \theta_0 := (\phi_0(X_1), \ldots, \phi_0(X_n))^\top \in \mathbb{R}^n \)

- Test \( H_0 : \theta_0 \in S \) where \( S := \{ \theta \in \mathbb{R}^n : \theta = X\beta, \beta \in \mathbb{R}^{d+1} \} \), and \( X = [e|x] \) where \( x = (X_1, \ldots, X_n)^\top \in \mathbb{R}^{n \times d}, \ e = (1, \ldots, 1)^\top \in \mathbb{R}^n. \)
A function is *affine* iff it is *both* convex and concave.

Let $I = \{ \theta \in \mathbb{R}^n : \exists \psi \text{ convex s.t. } \psi(X_i) = \theta_i, \forall i = 1, \ldots, n \}$

Let $D = -I = \{ \theta \in \mathbb{R}^n : \exists \psi \text{ concave s.t. } \psi(X_i) = \theta_i \}$

$I, D$ are *cones* that contain $S$

If $H_0$ is true, both $\Pi(Y|I)$ and $\Pi(Y|D)$ must be *close* to $\Pi(Y|S)$. 
**Figure:** Constant (dotted blue), convex (dashed red) and concave (solid black) fits to a scatterplot generated from $Y = \phi_0(x) + \epsilon$ with independent standard normal errors.
Method

- Model: \( \mathbf{Y} \sim \mathcal{N}(\theta_0, \sigma^2 I_n) \)

- Consider testing \( H_0 : \theta_0 \in S \) versus \( H_1 : \theta_0 \in I \cup D \setminus S \)

- \( S \) is a \textit{linear subspace}.

- \( I \subset \mathbb{R}^n \) is a closed \textit{convex cone} containing \( S \); \( D = -I \).

- \( I \cap D = S \)

- The log-likelihood function (up to a constant) is
  \[
  \ell(\theta, \sigma^2) = -\frac{1}{2\sigma^2} \| \mathbf{Y} - \theta \|^2 - \frac{n}{2} \log \sigma^2.
  \]

- The \textit{likelihood ratio} statistic is
  \[
  \Lambda = 2 \left\{ \max_{\theta \in I \cup D, \sigma^2 > 0} \ell(\theta, \sigma^2) - \max_{\theta \in S, \sigma^2 > 0} \ell(\theta, \sigma^2) \right\}
An equivalent test is to reject $H_0$ if
\[ T(Y) := \frac{\max\{\|\hat{\theta}_S - \hat{\theta}_I\|^2, \|\hat{\theta}_S - \hat{\theta}_D\|^2\}}{\|Y - \hat{\theta}_S\|^2} \quad \text{is large} \]

\[ \hat{\theta}_I = \Pi(Y|I), \quad \hat{\theta}_D = \Pi(Y|D) \quad \text{and} \quad \hat{\theta}_S = \Pi(Y|S). \]

Questions

- How do we find the critical value of the test?
- Power against all alternatives?
How to find the critical value of the test?

**Result:** The distribution of $T$ is *invariant* to translations in $S$.

**Lemma (Sen and Meyer (2013))**

For any $s \in S$ and $z \in \mathbb{R}^n$,\

$$T(z + s) = T(z).$$

As a consequence, for any $\theta_0 \in S$,\

$$T(Y) \overset{d}{=} T(Z),$$

where $Y \sim \mathcal{N}(\theta_0, \sigma^2 I_n)$ and $Z \sim \mathcal{N}(0, I_n)$.

**Proof**

For $z \in \mathbb{R}^n$ and $s \in S$,\

$$\Pi(z + s|\mathcal{I}) = \Pi(z|\mathcal{I}) + s, \quad \text{and} \quad \Pi(z + s|S) = \Pi(z|S) + s.$$\

Note that $Y = \sigma Z + \theta_0$, where $Z \sim \mathcal{N}(0, I_n)$. 
The test statistics \( T \) is *pivotal*, under \( H_0 \)

Suppose \( T(Z) \sim F \). Then, we *reject* \( H_0 \) if

\[
T(Y) > F^{-1}(1 - \alpha),
\]

where \( \alpha \in (0, 1) \) is the desired level of the test.

The distribution \( F \) can be approximated, up to any desired precision, by *Monte Carlo* simulations.

The test procedure has *exact* level \( \alpha \).

The test is also *unbiased*.
Theorem (Sen and Meyer (2013))

Under mild conditions, the *power* of the test converges to 1.

Example: Testing against simple linear regression

\[ Y_i = \phi_0(x_i) + \epsilon_i; \quad 0 \leq x_1 < \cdots < x_n \leq 1; \quad \epsilon_1, \ldots, \epsilon_n \text{ i.i.d. } \mathcal{N}(0, \sigma^2) \]

\[ \phi_0 : [0, 1] \rightarrow \mathbb{R} \text{ is an unknown } \textit{smooth} \text{ function} \]

Test \( H_0 : \phi_0 \) is an \textit{affine} function

Recall \( T(Y) = \frac{\max\{\frac{1}{n}\|\hat{\theta}_S - \hat{\theta}_I\|^2, \frac{1}{n}\|\hat{\theta}_S - \hat{\theta}_D\|^2\}}{\frac{1}{n}\|Y - \hat{\theta}_S\|^2} \)

Under \( H_0 \), \( T = O_p(\log n^{5/4}/n) \).

Otherwise, \( T \rightarrow_p c, \ c > 0. \)

Thus, \( \mathbb{P}_{\phi_0}(T > F^{-1}(1 - \alpha)) \rightarrow 1, \text{ if } \phi_0 \notin H_0 \)
1 Nonparametric convex regression: Introduction

2 Testing against a linear model

3 Theoretical results: Risk bounds
Convex regression in 1-dimension

- \( \phi_0 : [0, 1] \rightarrow \mathbb{R} \) is an unknown convex function.

- Observe
  \[
  Y_i = \phi_0(x_i) + \epsilon_i, \quad \text{for } i = 1, 2, \ldots, n.
  \]

- \( x_1 < x_2 < \cdots < x_n \) are fixed uniform grid points in \([0, 1]\).

- \( \epsilon_1, \ldots, \epsilon_n \) are i.i.d. \( \mathcal{N}(0, \sigma^2) \), \( \sigma^2 > 0 \).

- Want to study the LSE \( \hat{\phi}_{ls} \) of \( \phi_0 \), i.e.,
  \[
  \hat{\phi}_{ls} = \arg\min_{\psi \text{ is convex}} \sum_{i=1}^{n} \{Y_i - \psi(x_i)\}^2.
  \]

- \( \hat{\theta} = (\hat{\phi}_{ls}(x_1), \ldots, \hat{\phi}_{ls}(x_n)) \) is unique.
Theoretical results: Convex LSE when $d = 1$

**Known results**

- Asymptotic Behavior on Interior Compacta:
  - Consistency (Hanson and Pledger (1976))
  - Rate of convergence (Dümbgen et al. (2004))

- Asymptotic Behavior at a Point:
  - Rate of convergence (Mammen (1991))
  - Limiting distribution (Groeneboom et al. (2001))

- Not much is known about the *global* behavior of the LSE.

- Natural global loss function:

$$\ell^2(\psi, \phi) = \frac{1}{n} \sum_{i=1}^{n} \{\psi(x_i) - \phi(x_i)\}^2.$$
Worst case risk upper bound

- Let $\theta_0 = (\phi_0(x_1), \ldots, \phi_0(x_n))$, $\hat{\theta} = (\hat{\phi}_{ls}(x_1), \ldots, \hat{\phi}_{ls}(x_n))$

- For every $\phi_0$, the risk

$$
\mathbb{E}_{\phi_0} \ell^2(\hat{\phi}_{ls}, \phi_0) = \frac{1}{n} \mathbb{E}_{\theta_0} \|\hat{\theta} - \theta_0\|^2,
$$

is bounded from above by $n^{-4/5}$ up to logarithmic factors in $n$.

**Risk upper bound (Guntuboyina and Sen (2013b))**

We have

$$
\mathbb{E}_{\phi_0} \ell^2(\hat{\phi}_{ls}, \phi_0) \lesssim n^{-4/5} \log n
$$

- *No smoothness* assumptions on $\phi_0$ necessary. Only convexity needed.

- *Finite sample* result, holds for every $n$. 
Question: Can $\mathbb{E}_{\phi_0} \ell_2(\hat{\phi}_{ls}, \phi_0)$ be much smaller than $n^{-4/5}$?

(1) **NO**: If $\phi_0''(x)$ exists and is bounded from above and below by positive constants on a subinterval of $(0, 1)$.

Then, in a very strong sense, no estimator (not just the LSE) can have risk smaller than $n^{-4/5}$.

The local minimax risk at $\phi_0$:

$$R_n(\phi_0) := \inf_{\hat{\phi}} \sup_{\phi \in N(\phi_0)} \mathbb{E}_{\phi} \ell_2(\phi, \hat{\phi})$$

where $N(\phi_0) := \{\phi \text{ convex} : \|\phi - \phi_0\|_\infty \leq n^{-2/5}\}$.

Local minimax lower bound (Guntuboyina and Sen (2013b))

$$R_n(\phi_0) \gtrsim n^{-4/5}$$
Question: Can \( \mathbb{E}_{\phi_0} \ell^2(\hat{\phi}_{ls}, \phi_0) \) be much smaller than \( n^{-4/5} \)?

(2) **YES**: \( \phi_0 \) is a piecewise affine convex function.

**Adaptation**: Suppose \( \phi_0 \) has \( k \) affine pieces, then

\[
\mathbb{E}_{\phi_0} \ell^2(\hat{\phi}_{ls}, \phi_0) \lesssim \frac{k}{n} (\log n)^{5/4}.
\]

- **Parametric rate** up to logarithmic factors.
- **Surprising** as the LSE minimizes the LS criterion over *all* convex fns.

\[
\phi_0(x) = x^2
\]

\[
\phi_0(x) = |x|
\]
Key proof ideas - Risk upper bounds


- Risk behavior of $\hat{\phi}_{ls}$ is determined by

$$
\mathbb{E} \left[ \sup_{\phi \text{ convex}} \sum_{i=1}^{n} \epsilon_i \{\phi(x_i) - \phi_0(x_i)\} \right]
$$

- *Local balls* of convex functions, defined as

$$
S(\phi_0, r) := \{\phi \text{ convex} : \ell^2(\phi, \phi_0) \leq r^2\}
$$

- Need good upper bounds on *expected maxima* of Gaussian processes.

- Using Dudley’s theorem, we can bound the above if we can compute the *metric entropy* of these local balls.
Covering numbers and metric entropy

Definitions

For a metric space \((\mathcal{X}, \rho)\) and a subset \(T \subseteq \mathcal{X}\) we say that \(\hat{T} \subseteq \mathcal{X}\) is an \(\epsilon\)-cover of \(T\) if \(\forall t \in T, \exists \hat{t} \in \hat{T}\) such that \(\rho(t, \hat{t}) \leq \epsilon\).

There obviously exist many different covers of \(T\), but we concern ourselves with the fewest elements. We call the size of such a cover the covering number.

Formally, the \(\epsilon\)-covering number of \(T\) is \(N(\epsilon, T, \rho) = \min\{|\hat{T}|: \hat{T}\) is an \(\epsilon\)-cover\}\).

A proper cover is one where \(\hat{T} \subseteq T\), and a proper covering number is defined in terms of the size of the minimum proper cover.

It can be shown that covering numbers and proper covering numbers are related by the following inequality.

\[ N(\epsilon, T) \leq N_{\text{proper}}(\epsilon, T) \leq N(\epsilon^2, T) \]

The metric entropy, defined as \(H(\epsilon, T) = \log N(\epsilon, T, \rho)\), is a natural representation of how many bits you need to send in order to identify an element of a set up to precision \(\epsilon\).

Example

For the space \((\mathbb{X}, L_2(P_n))\), the distance is defined as

- Recall: For a subset \(\mathcal{F}\) of a metric space \((\mathcal{X}, \rho)\) the \(\epsilon\)-covering number \(N(\epsilon, \mathcal{F}, \rho)\) is defined as the smallest number of balls of radius \(\epsilon\) whose union contains \(\mathcal{F}\).
- Covering numbers capture the size of the underlying set \(\mathcal{F}\).
- Metric entropy: logarithm of the covering numbers \(\log N(\epsilon, \mathcal{F}, \rho)\)
- We need good bounds on the metric entropy \(\log N(\epsilon, S(\phi_0, r), \ell)\)
Bronšteǐn (1976) showed that the metric entropy of uniformly Lipschitz and uniformly bounded convex functions on $[0, 1]^d$ under the supremum metric grows as $\epsilon^{-d/2}$.

But functions in these local balls are neither uniformly bounded nor uniformly Lipschitz.

We modify Bronshtein’s result through a series of peeling arguments to obtain the metric entropy of these balls (Dryanov (2009) and Guntuboyina and Sen (2013a)).
Ongoing research and future work

- Determine the *global* and *local* (point-wise) rates of convergence and the asymptotic distribution of the LSE in convex regression.

- Determine the *optimal* rates of convergence in these problems.

- How do we estimate and compute other multi-dimensional shape restricted functions, e.g., *multi-dimensional* monotone regression, quasi convex functions, etc.?

- Risk bounds for a general *polyhedral cone* (connection to non-negative least squares)?

- Investigate *additive models* with unknown monotone/convex functions.

- Study *single index models* with unknown monotone/convex links.
Collaborators

Emilio Seijo  
former Ph.D. student  
Start-up

Mary Meyer  
Professor  
Colorado State Uni.

Aditya Guntuboyina  
Assistant Professor  
UC Berkeley

and others ...
References


Thank you very much!

Questions?