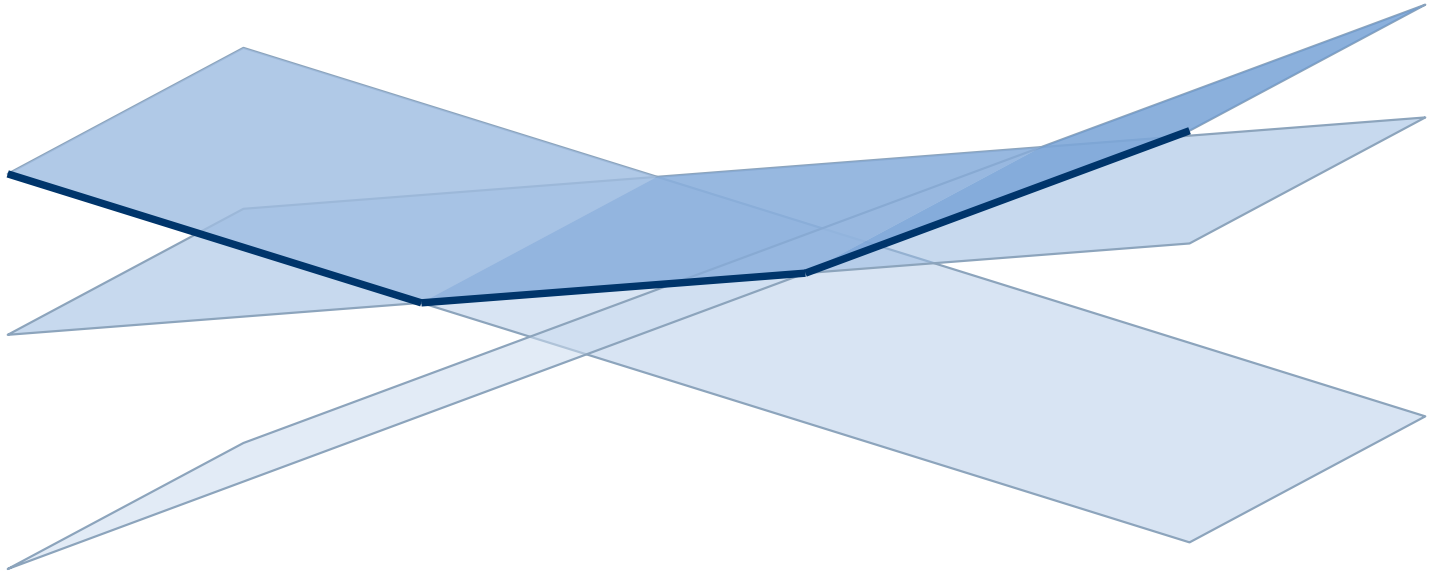


Linear Regression with Self-Selection Bias in Polynomial Time

Felix Zhou

Yale University

Table of Contents

Motivation

Problem Definition

Prior Works

Our Contributions

Proof Overview

Table of Contents

Motivation

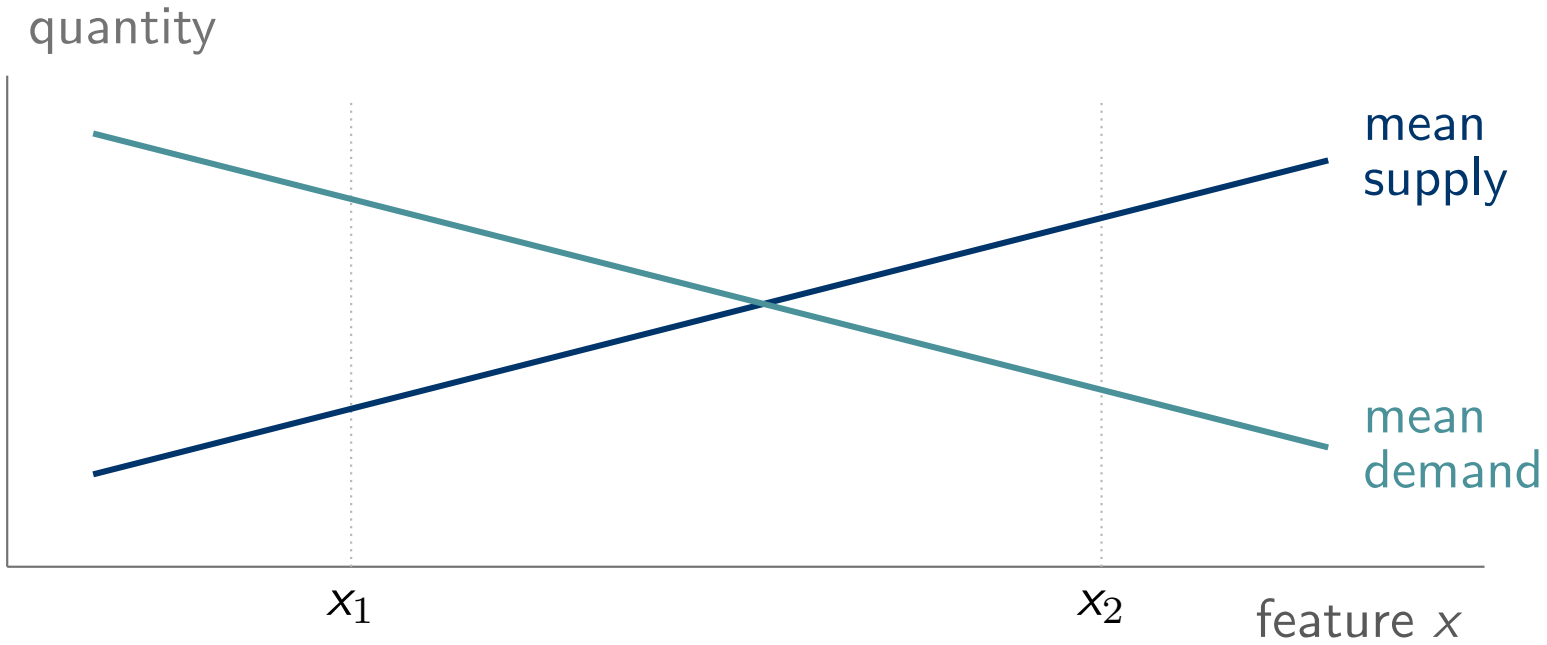
Problem Definition

Prior Works

Our Contributions

Proof Overview

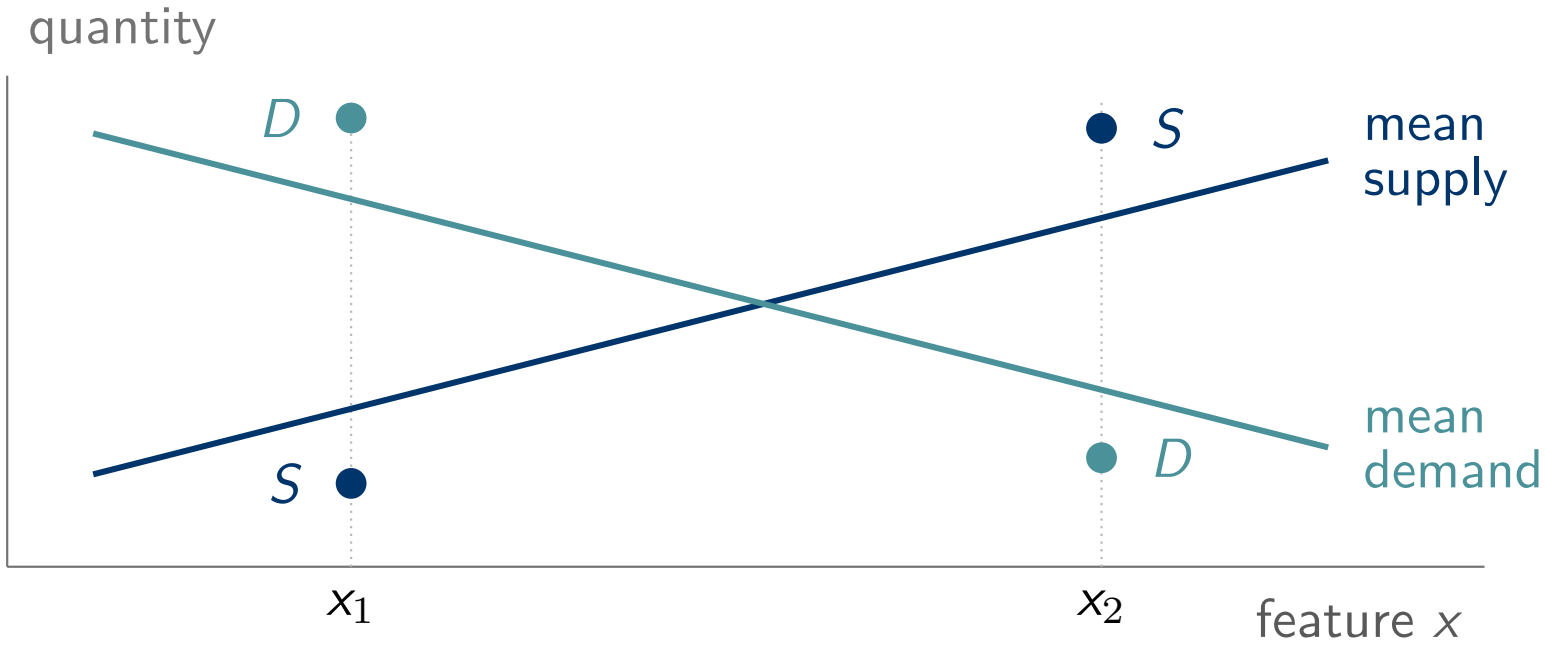
Motivation: Learning from Markets in Disequilibrium



- Mean supply S and demand D vary with features.

Fair and Jaffee [FJ72].

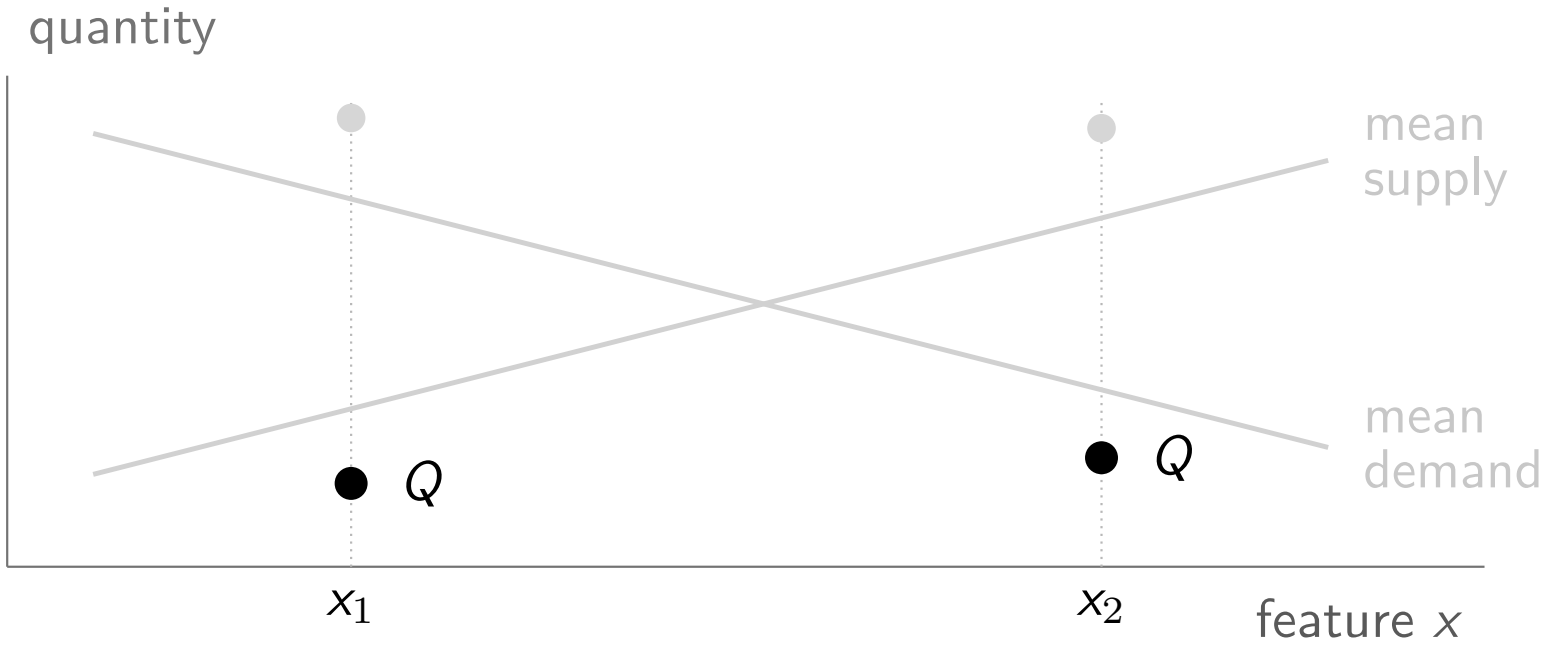
Motivation: Learning from Markets in Disequilibrium



- Mean supply S and demand D vary with features.
- Transactions still occur due to random shocks.

Fair and Jaffee [FJ72].

Motivation: Learning from Markets in Disequilibrium



- Mean supply S and demand D vary with features.
- Transactions still occur due to random shocks.
- Only "self-selected" quantity $Q = \min\{S, D\}$ is observed.

Fair and Jaffee [FJ72].

A Simplified Fair–Jaffee Model

- **Observed transaction** $(\boldsymbol{X}, Q)$, where $Q = \min\{S, D\}$.

Fair and Jaffee [FJ72].

A Simplified Fair–Jaffee Model

- **Observed transaction** $(\boldsymbol{X}, Q)$, where $Q = \min\{S, D\}$.
- **Supply** $S = \boldsymbol{w}_S^\top \boldsymbol{X} + \xi_S$.

Fair and Jaffee [FJ72].

A Simplified Fair–Jaffee Model

- **Observed transaction** $(\boldsymbol{X}, Q)$, where $Q = \min\{S, D\}$.
- **Supply** $S = \boldsymbol{w}_S^\top \boldsymbol{X} + \xi_S$.
- **Demand** $D = \boldsymbol{w}_D^\top \boldsymbol{X} + \xi_D$.

Fair and Jaffee [FJ72].

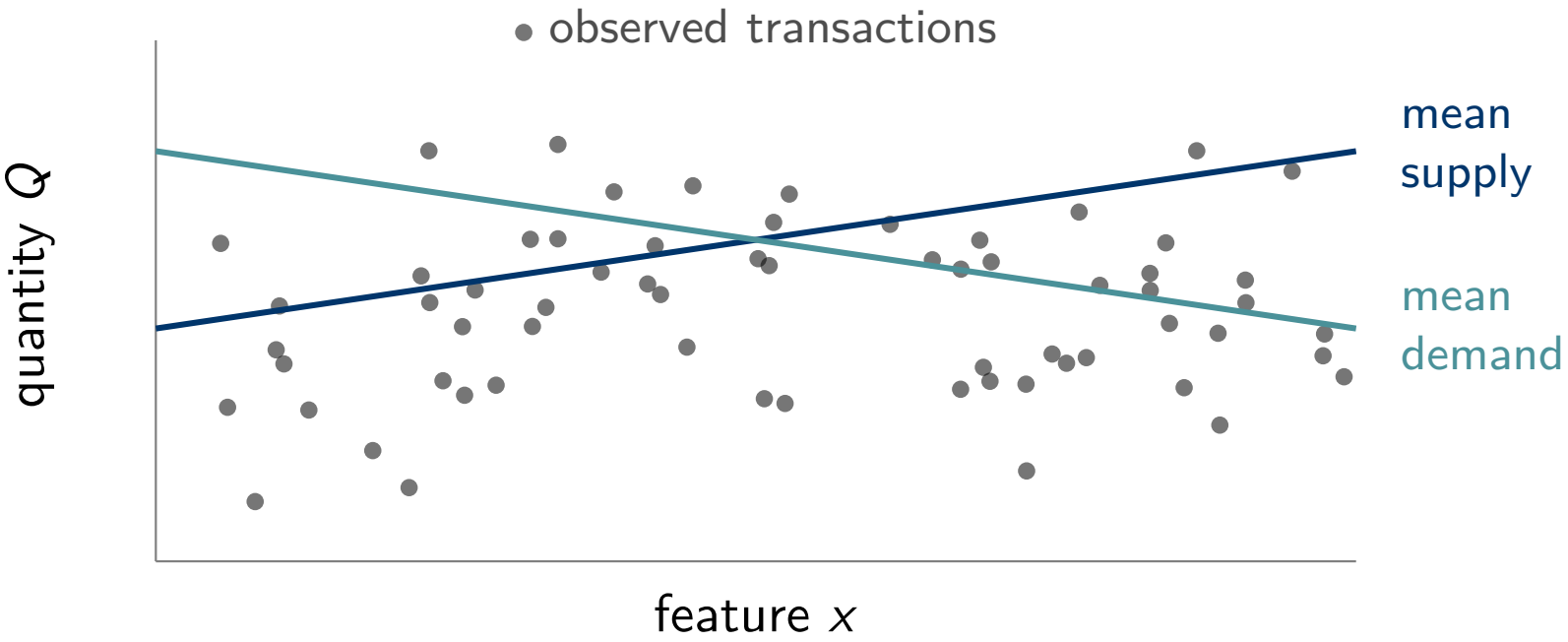
A Simplified Fair–Jaffee Model

- **Observed transaction** $(\boldsymbol{X}, Q)$, where $Q = \min\{S, D\}$.
- **Supply** $S = \boldsymbol{w}_S^\top \boldsymbol{X} + \xi_S$.
- **Demand** $D = \boldsymbol{w}_D^\top \boldsymbol{X} + \xi_D$.
- Wish to learn $\boldsymbol{w}_S, \boldsymbol{w}_D$ given observations $(\boldsymbol{X}, Q)$

Fair and Jaffee [FJ72].

Challenges

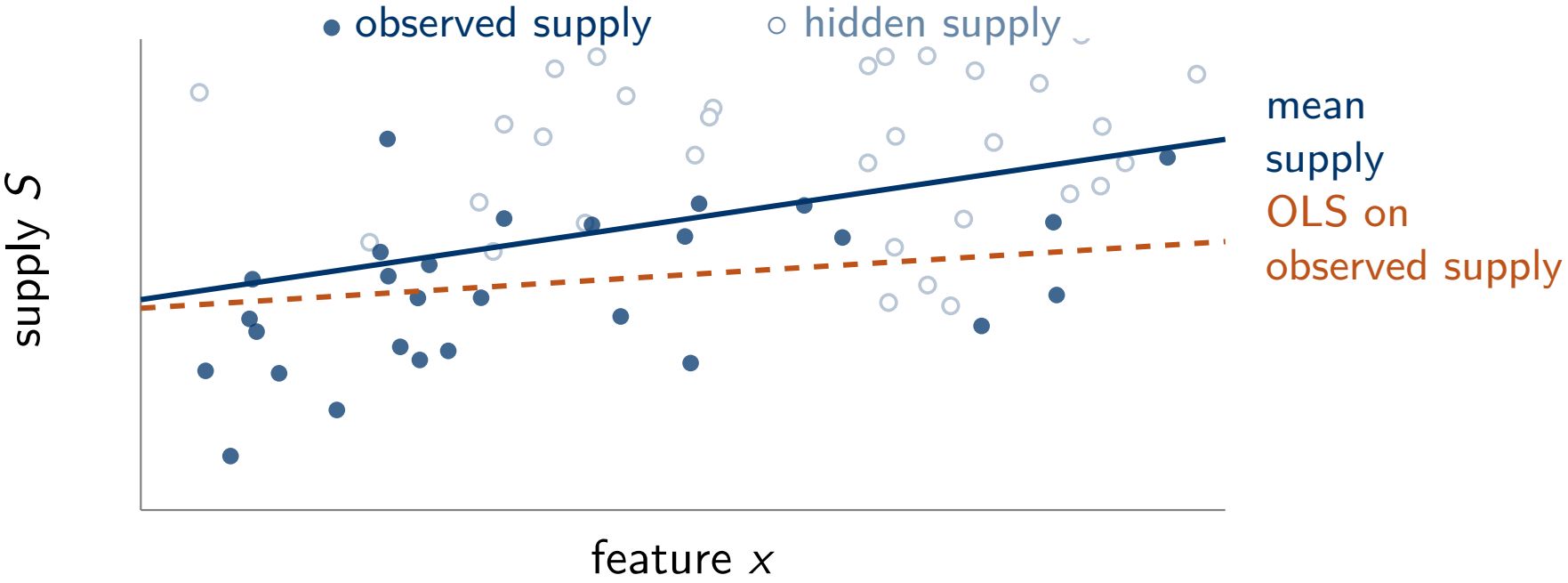
Goal Learn supply and demand from transaction records.



Hidden index Which side constrained each transaction?

Challenges

Goal Learn supply and demand from transaction records.



Selection bias Supply is observed only when $S \leq D$.
Even given the labels, OLS underestimates the supply slope here.

Table of Contents

Motivation

Problem Definition

Prior Works

Our Contributions

Proof Overview

Problem Definition

DEFINITION (MAXIMUM SELF-SELECTION; [CDIZ23,STOC])

1. Recover unknown regressors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k \in \mathbb{R}^n$ to error $\varepsilon > 0$.

Maximum self-selection (Def. 1.1). Boundedness and uncoveredness (Assumption 1.2).

Problem Definition

DEFINITION (MAXIMUM SELF-SELECTION; [CDIZ23,STOC])

1. Recover unknown regressors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k \in \mathbb{R}^n$ to error $\varepsilon > 0$.
2. Observe samples $(\boldsymbol{X}, Y_{\max})$ where

$$\boldsymbol{X} \sim N(0, I_n), \qquad Y_{\max} = \max(Y_1, \ldots, Y_k),$$
$$Y_i = \boldsymbol{w}_i^\top \boldsymbol{X} + \xi_i, \qquad \xi_i \overset{\text{iid}}{\sim} N(0,1).$$

Maximum self-selection (Def. 1.1). Boundedness and uncoveredness (Assumption 1.2).

Problem Definition

DEFINITION (MAXIMUM SELF-SELECTION; [CDIZ23,STOC])

1. Recover unknown regressors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k \in \mathbb{R}^n$ to error $\varepsilon > 0$.
2. Observe samples $(\boldsymbol{X}, Y_{\max})$ where

$$\boldsymbol{X} \sim N(0, I_n), \qquad Y_{\max} = \max(Y_1, \ldots, Y_k),$$
$$Y_i = \boldsymbol{w}_i^\top \boldsymbol{X} + \xi_i, \qquad \xi_i \overset{\text{iid}}{\sim} N(0,1).$$

3. Argmax index not observed!

Maximum self-selection (Def. 1.1). Boundedness and uncoveredness (Assumption 1.2).

Problem Definition

DEFINITION (MAXIMUM SELF-SELECTION; [CDIZ23,STOC])

1. Recover unknown regressors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k \in \mathbb{R}^n$ to error $\varepsilon > 0$.
2. Observe samples $(\boldsymbol{X}, Y_{\max})$ where

$$\boldsymbol{X} \sim N(0, I_n), \qquad Y_{\max} = \max(Y_1, \ldots, Y_k),$$
$$Y_i = \boldsymbol{w}_i^\top \boldsymbol{X} + \xi_i, \qquad \xi_i \overset{\text{iid}}{\sim} N(0,1).$$

3. Argmax index not observed!

ASSUMPTION (GEOMETRIC CONSTANTS)

$$\|\boldsymbol{w}_i\|_2 \le B, \qquad \frac{\langle \boldsymbol{w}_i, \boldsymbol{w}_j \rangle}{\|\boldsymbol{w}_i\|_2 \, \|\boldsymbol{w}_j\|_2} \le 1 - \Delta \quad (i \neq j).$$

Maximum self-selection (Def. 1.1). Boundedness and uncoveredness (Assumption 1.2).

Table of Contents

Motivation

Problem Definition

Prior Works

Our Contributions

Proof Overview

History

- **Roy (1951)** Occupational choice and earnings.

Roy, Oxford Economic Papers. Fair–Jaffee and Heckman, Econometrica.

History

- **Roy (1951)** Occupational choice and earnings.
- **Fair–Jaffee (1972)** Markets in disequilibrium.

Roy, Oxford Economic Papers. Fair–Jaffee and Heckman, Econometrica.

History

- **Roy (1951)** Occupational choice and earnings.
- **Fair–Jaffee (1972)** Markets in disequilibrium.
- **Heckman (1979)** Algorithm with asymptotic guarantees.

Roy, Oxford Economic Papers. Fair–Jaffee and Heckman, Econometrica.

History

- **Roy (1951)** Occupational choice and earnings.
- **Fair–Jaffee (1972)** Markets in disequilibrium.
- **Heckman (1979)** Algorithm with asymptotic guarantees.

Can we develop computationally efficient algorithms
with finite-sample guarantees?

Roy, Oxford Economic Papers. Fair–Jaffee and Heckman, Econometrica.

Prior Works

Fixed geometric constants.

Work	Samples	Running time
CDIZ23 STOC	$\mathrm{poly}(n)\exp(\mathrm{poly}(k))$	$\mathrm{poly}(n)\exp(\mathrm{poly}(k))$

CDIZ23 gives accuracy $\varepsilon = \mathrm{poly}(1/k)$.

Prior Works

Fixed geometric constants.

Work	Samples	Running time
CDIZ23 STOC	$\mathrm{poly}(n)\exp(\mathrm{poly}(k))$	$\mathrm{poly}(n)\exp(\mathrm{poly}(k))$
GM24	$\mathrm{poly}(n, k, 1/\varepsilon)$	$\mathrm{poly}(n, k, 1/\varepsilon) + (1/\varepsilon)^{\widetilde{O}(k)}$

CDIZ23 gives accuracy $\varepsilon = \mathrm{poly}(1/k)$.

Prior Works

Fixed geometric constants.

Work	Samples	Running time
CDIZ23 STOC	$\mathrm{poly}(n)\exp(\mathrm{poly}(k))$	$\mathrm{poly}(n)\exp(\mathrm{poly}(k))$
GM24	$\mathrm{poly}(n, k, 1/\varepsilon)$	$\mathrm{poly}(n, k, 1/\varepsilon) + (1/\varepsilon)^{\widetilde{O}(k)}$
GM24 + KMZ26, COLT	$\mathrm{poly}(n, k, 1/\varepsilon)$	$\mathrm{poly}(n, k, 1/\varepsilon) + 2^{\widetilde{O}(k)}$

CDIZ23 gives accuracy $\varepsilon = \mathrm{poly}(1/k)$.

Prior Works

Fixed geometric constants.

Work	Samples	Running time
CDIZ23 STOC	$\mathrm{poly}(n)\exp(\mathrm{poly}(k))$	$\mathrm{poly}(n)\exp(\mathrm{poly}(k))$
GM24	$\mathrm{poly}(n, k, 1/\varepsilon)$	$\mathrm{poly}(n, k, 1/\varepsilon) + (1/\varepsilon)^{\widetilde{O}(k)}$
GM24 + KMZ26, COLT	$\mathrm{poly}(n, k, 1/\varepsilon)$	$\mathrm{poly}(n, k, 1/\varepsilon) + 2^{\widetilde{O}(k)}$
This work	$\mathrm{poly}(n, k, 1/\varepsilon)$	$\mathrm{poly}(n, k, 1/\varepsilon)$

CDIZ23 gives accuracy $\varepsilon = \mathrm{poly}(1/k)$.

Table of Contents

Motivation

Problem Definition

Prior Works

Our Contributions

Proof Overview

Our Contributions

THEOREM 1.3. POLYNOMIAL-TIME SELF-SELECTION

For fixed geometric constants, with high probability,

- Recover all k regressors.

Our Contributions

THEOREM 1.3. POLYNOMIAL-TIME SELF-SELECTION

For fixed geometric constants, with high probability,

- Recover all k regressors.
- Accuracy[1]

$$\|\widehat{\boldsymbol{w}}_i - \boldsymbol{w}_i\|_2 \leq \varepsilon \qquad \text{for every } i.$$

[1] The estimates are indexed up to a permutation.

Our Contributions

THEOREM 1.3. POLYNOMIAL-TIME SELF-SELECTION

For fixed geometric constants, with high probability,

- Recover all k regressors.
- Accuracy[1]

$$\|\widehat{\boldsymbol{w}}_i - \boldsymbol{w}_i\|_2 \leq \varepsilon \qquad \text{for every } i.$$

- Sample complexity and running time

$$\text{poly}(n, k, 1/\varepsilon).$$

[1] The estimates are indexed up to a permutation.

Table of Contents

Motivation

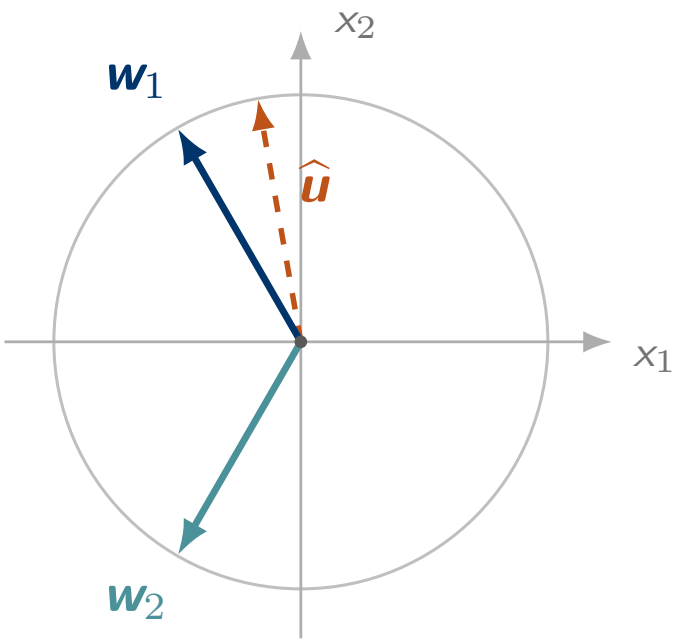
Problem Definition

Prior Works

Our Contributions

Proof Overview
- Structural Lemma
- Algorithmic Implications
- Technical Challenges

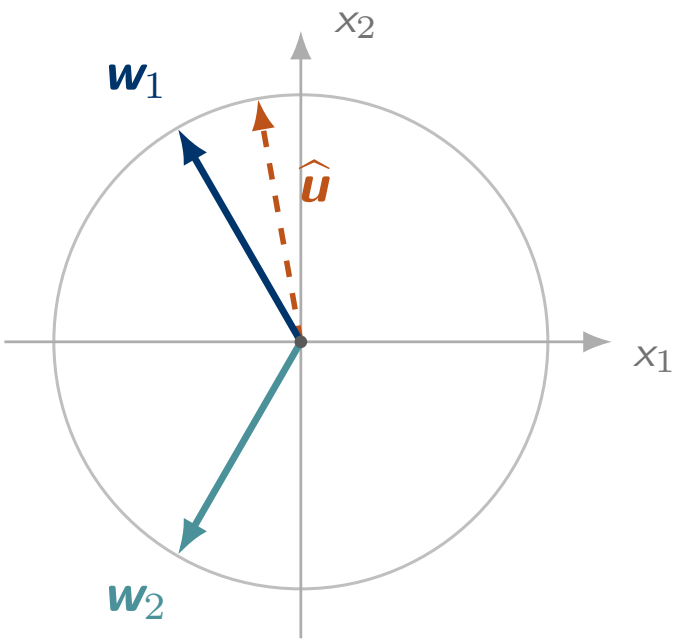
Prior Works: Moment-Testing Paradigm



- Start from a candidate $\widehat{u} \approx w_i / \|w_i\|_2$.

Paper, §2.1, p. 4. Prior moment tests [CDIZ23, GM24].

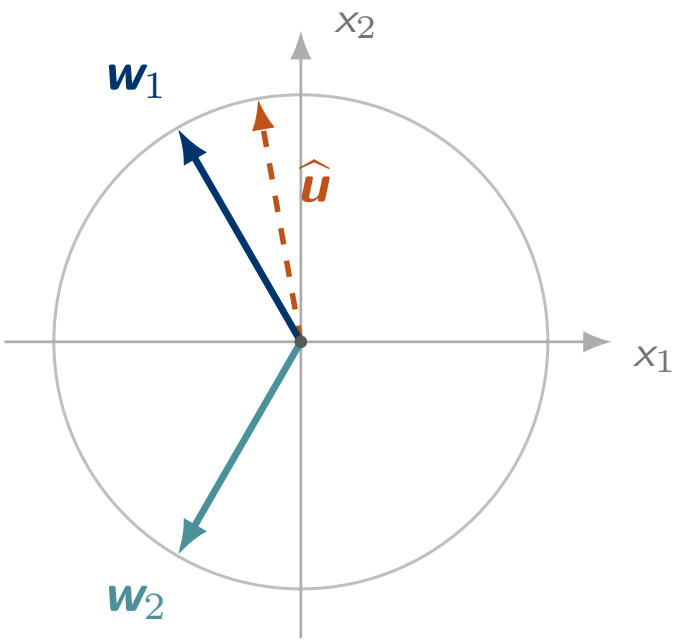
Prior Works: Moment-Testing Paradigm



- Start from a candidate $\widehat{\boldsymbol{u}} \approx w_i / \|w_i\|_2$.
- Condition on a large $\widehat{\boldsymbol{u}}^\top \boldsymbol{X}$.

Paper, §2.1, p. 4. Prior moment tests [CDIZ23, GM24].

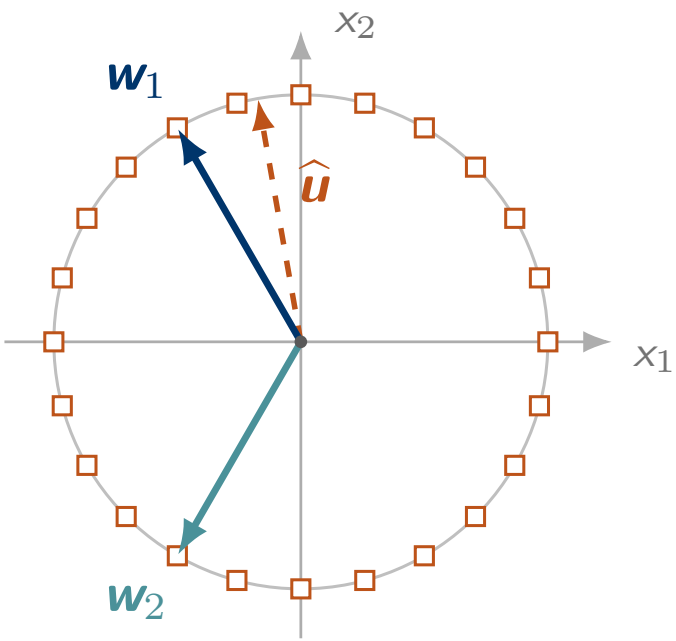
Prior Works: Moment-Testing Paradigm



- Start from a candidate $\widehat{\boldsymbol{u}} \approx w_i / \|w_i\|_2$.
- Condition on a large $\widehat{\boldsymbol{u}}^\top \boldsymbol{X}$.
- Expect $Y_i \approx Y_{\max}$.

Paper, §2.1, p. 4. Prior moment tests [CDIZ23, GM24].

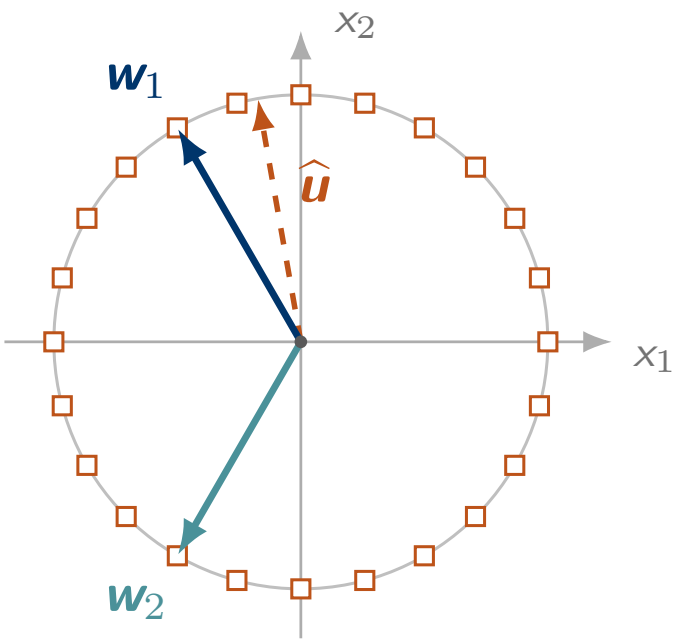
Prior Works: Moment-Testing Paradigm



- Start from a candidate $\widehat{\boldsymbol{u}} \approx w_i / \|w_i\|_2$.
- Condition on a large $\widehat{\boldsymbol{u}}^\top \boldsymbol{X}$.
- Expect $Y_i \approx Y_{\max}$.
- Test conditional moments of $Y_{\max}$.

Paper, §2.1, p. 4. Prior moment tests [CDIZ23, GM24].

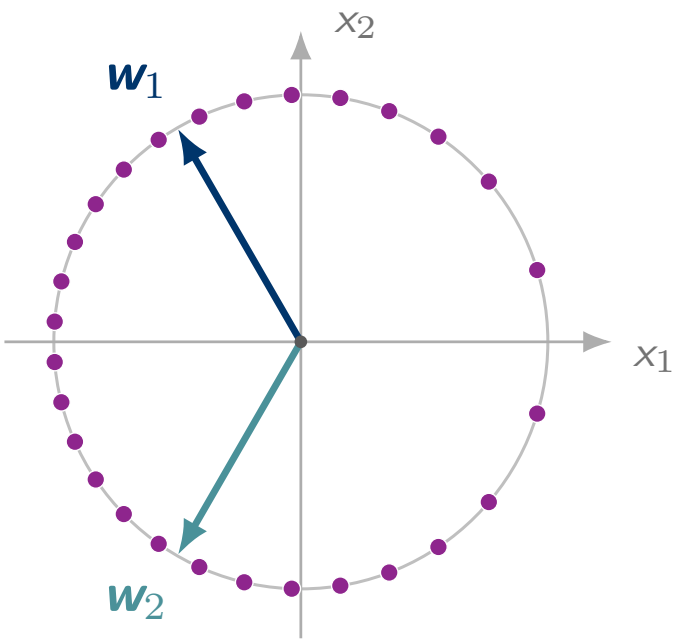
Prior Works: Moment-Testing Paradigm



- Start from a candidate $\widehat{\boldsymbol{u}} \approx w_i / \|w_i\|_2$.
- Condition on a large $\widehat{\boldsymbol{u}}^\top \boldsymbol{X}$.
- Expect $Y_i \approx Y_{\max}$.
- Test conditional moments of $Y_{\max}$.
- Cost of searching the span exponential in k.

Paper, §2.1, p. 4. Prior moment tests [CDIZ23, GM24].

Main Idea: Converse Holds!



Filter for $Y_{\max} \in [T, T+1]$.
Expect large $w_i^\top \boldsymbol{X}$ for some w_i.

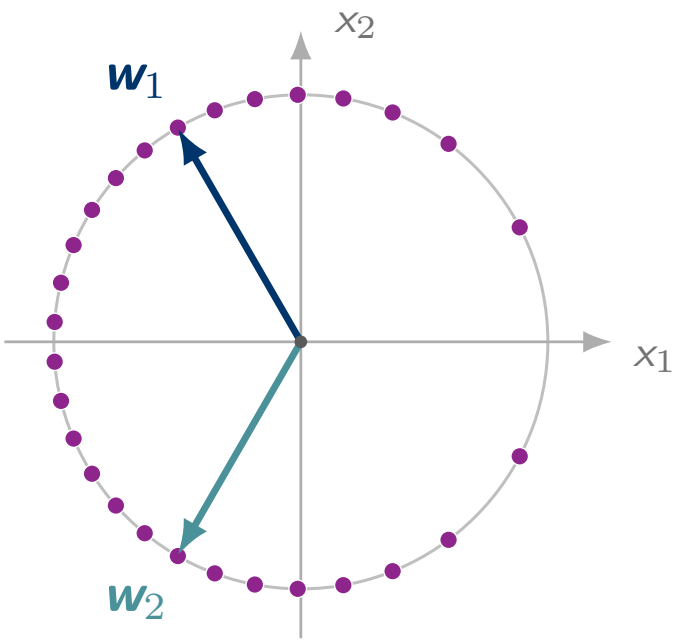
$$T = 1$$

Roughly uniform angles.

Selected covariate directions.

Paper, §2.2.1 and Figure 1, p. 5. Two-dimensional Gaussian example.

Main Idea: Converse Holds!



Filter for $Y_{\max} \in [T, T+1]$.
Expect large $w_i^\top \boldsymbol{X}$ for some w_i.

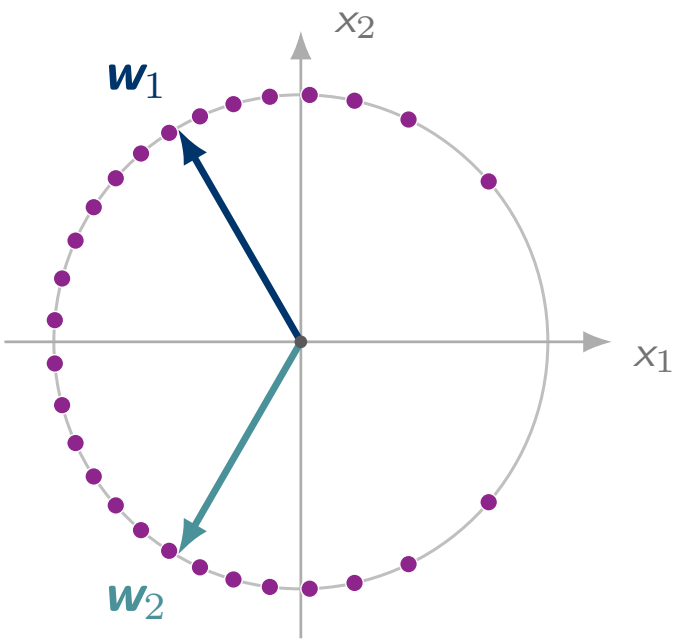
$$T = 2$$

Mass shifts toward regressors.

Selected covariate directions.

Paper, §2.2.1 and Figure 1, p. 5. Two-dimensional Gaussian example.

Main Idea: Converse Holds!



Filter for $Y_{\max} \in [T, T+1]$.
Expect large $w_i^\top \boldsymbol{X}$ for some w_i.

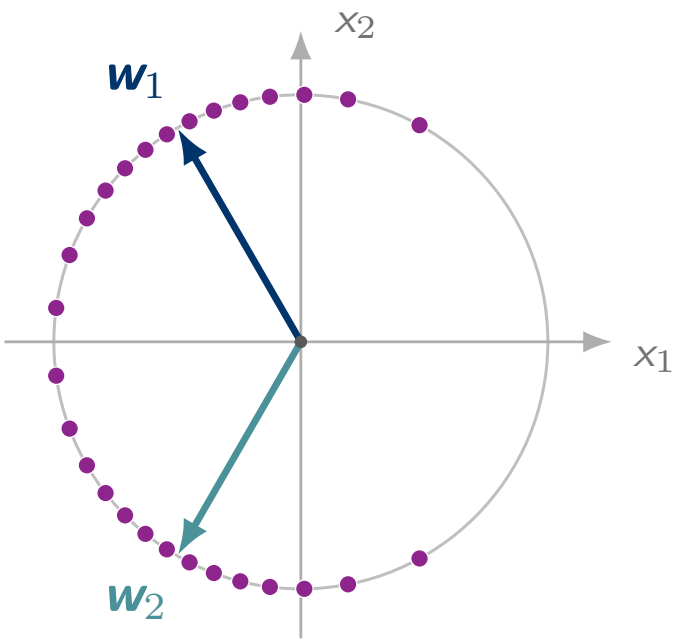
$$T = 3$$

Mass shifts toward regressors.

Selected covariate directions.

Paper, §2.2.1 and Figure 1, p. 5. Two-dimensional Gaussian example.

Main Idea: Converse Holds!



Filter for $Y_{\max} \in [T, T+1]$.
Expect large $w_i^\top \boldsymbol{X}$ for some w_i.

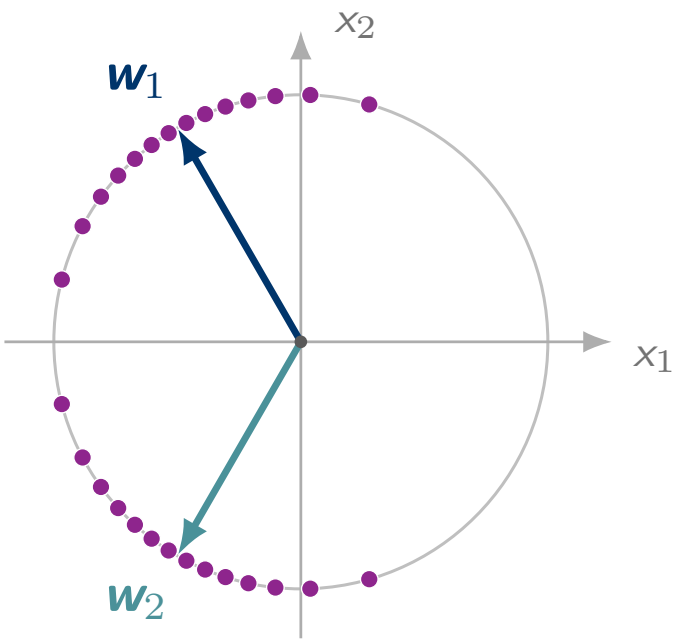
$$T = 4$$

Mass shifts toward regressors.

Selected covariate directions.

Paper, §2.2.1 and Figure 1, p. 5. Two-dimensional Gaussian example.

Main Idea: Converse Holds!



Filter for $Y_{\max} \in [T, T+1]$.
Expect large $w_i^\top \boldsymbol{X}$ for some w_i.

$$T = 5$$

Mass concentrates near regressors.

Selected covariate directions.

Paper, §2.2.1 and Figure 1, p. 5. Two-dimensional Gaussian example.

Background: Hermite Moments

Draw $\boldsymbol{X}' \sim N(\mathbf{0}, I_n)$ independently and set

$$\widetilde{\boldsymbol{X}} = \boldsymbol{X} + \mathrm{i}\boldsymbol{X}'.$$

We study Hermite moments:

$$\mathbb{E}_{\boldsymbol{X}}[\mathsf{H}_d(\boldsymbol{X})] = \frac{1}{\sqrt{d!}}\,\mathbb{E}_{\boldsymbol{X},\boldsymbol{X}'}\left[\widetilde{\boldsymbol{X}}^{\otimes d}\right].$$

Paper, §2.2.1–2.2.2.

Background: Hermite Moments

Draw $\boldsymbol{X}' \sim N(\mathbf{0}, I_n)$ independently and set

$$\widetilde{\boldsymbol{X}} = \boldsymbol{X} + \mathrm{i}\boldsymbol{X}'.$$

We study Hermite moments:

$$\mathbb{E}_{\boldsymbol{X}}[\mathrm{H}_d(\boldsymbol{X})] = \frac{1}{\sqrt{d!}}\,\mathbb{E}_{\boldsymbol{X},\boldsymbol{X}'}\left[\widetilde{\boldsymbol{X}}^{\otimes d}\right].$$

- Different Hermite degrees are orthogonal under a Gaussian.

Paper, §2.2.1–2.2.2.

Background: Hermite Moments

Draw $\boldsymbol{X}' \sim N(\mathbf{0}, I_n)$ independently and set

$$\widetilde{\boldsymbol{X}} = \boldsymbol{X} + \mathrm{i}\boldsymbol{X}'.$$

We study Hermite moments:

$$\mathbb{E}_{\boldsymbol{X}}[\mathsf{H}_d(\boldsymbol{X})] = \frac{1}{\sqrt{d!}}\,\mathbb{E}_{\boldsymbol{X},\boldsymbol{X}'}\left[\widetilde{\boldsymbol{X}}^{\otimes d}\right].$$

- Different Hermite degrees are orthogonal under a Gaussian.
- Positive-degree moments have Gaussian mean zero.

Paper, §2.2.1–2.2.2.

Background: Hermite Moments

Draw $\boldsymbol{X}' \sim N(\mathbf{0}, I_n)$ independently and set

$$\widetilde{\boldsymbol{X}} = \boldsymbol{X} + \mathrm{i}\boldsymbol{X}'.$$

We study Hermite moments:

$$\mathbb{E}_{\boldsymbol{X}, Y_{\max}}[\mathrm{H}_d(\boldsymbol{X})\mathbf{1}\{Y_{\max} \in [T, T+1]\}] = \frac{1}{\sqrt{d!}}\,\mathbb{E}_{\boldsymbol{X}, \boldsymbol{X}', Y_{\max}}\left[\widetilde{\boldsymbol{X}}^{\otimes d}\mathbf{1}\{Y_{\max} \in [T, T+1]\}\right]$$

- Different Hermite degrees are orthogonal under a Gaussian.
- Positive-degree moments have Gaussian mean zero.

Paper, §2.2.1–2.2.2.

Structural Lemma

Lemma (Restricted moment structure)

Let $\boldsymbol{u}_i = \boldsymbol{w}_i / \|\boldsymbol{w}_i\|_2$. For weights $\lambda_i > 0$,

$$\mathbb{E}[\mathsf{H}_d(\boldsymbol{X})\mathbf{1}\{Y_{\max} \in [T, T+1]\}] = \sum_{i=1}^{k} \lambda_i \boldsymbol{u}_i^{\otimes d} + \mathsf{R}_d.$$

Paper, §2.2.1 and Theorem 5.1.

Structural Lemma

Lemma (Restricted moment structure)

Let $\boldsymbol{u}_i = \boldsymbol{w}_i / \|\boldsymbol{w}_i\|_2$. For weights $\lambda_i > 0$,

$$\mathbb{E}[\mathsf{H}_d(\boldsymbol{X})\mathbf{1}\{Y_{\max} \in [T, T+1]\}] = \sum_{i=1}^{k} \lambda_i \boldsymbol{u}_i^{\otimes d} + \mathsf{R}_d.$$

1. For $T = \Theta(\sqrt{\log n})$,
$$\|\mathsf{R}_d\|_F \leq \mathrm{poly}(1/n).$$

Paper, §2.2.1 and Theorem 5.1.

Structural Lemma

Lemma (Restricted moment structure)

Let $\boldsymbol{u}_i = \boldsymbol{w}_i / \|\boldsymbol{w}_i\|_2$. For weights $\lambda_i > 0$,

$$\mathbb{E}[\mathsf{H}_d(\boldsymbol{X})\mathbf{1}\{Y_{\max} \in [T, T+1]\}] = \sum_{i=1}^{k} \lambda_i \boldsymbol{u}_i^{\otimes d} + \mathsf{R}_d.$$

1. For $T = \Theta(\sqrt{\log n})$,
$$\|\mathsf{R}_d\|_F \leq \mathrm{poly}(1/n).$$
2. For $d = \Theta(\log n)$ and $i \neq j$,
$$\left|\langle \boldsymbol{u}_i^{\otimes d}, \boldsymbol{u}_j^{\otimes d} \rangle\right| \leq \mathrm{poly}(1/n).$$

Paper, §2.2.1 and Theorem 5.1.

One response

$$\mathbb{E}[\mathsf{H}_d(\boldsymbol{X})\mathbf{1}\{Y_i \in [T, T+1]\}]$$
$$= \frac{1}{\sqrt{d!}}\mathbb{E}[\widetilde{\boldsymbol{X}}^{\otimes d}\mathbf{1}\{Y_i \in [T, T+1]\}]$$

Paper, §2.2.1 and Corollary 5.3.

One response

$$
\begin{aligned}
&\mathbb{E}[\mathsf{H}_d(\boldsymbol{X})\mathbf{1}\{Y_i \in [T, T+1]\}] \\
&\quad = \frac{1}{\sqrt{d!}}\mathbb{E}[\widetilde{\boldsymbol{X}}^{\otimes d}\mathbf{1}\{Y_i \in [T, T+1]\}] \\
&\quad = \underbrace{\frac{1}{\sqrt{d!}}\mathbb{E}[\langle \boldsymbol{u}_i, \widetilde{\boldsymbol{X}}\rangle^d \mathbf{1}\{Y_i \in [T, T+1]\}]}_{\lambda_i}\, \boldsymbol{u}_i^{\otimes d}.
\end{aligned}
$$

Paper, §2.2.1 and Corollary 5.3.

k responses

$$\mathbf{1}\{Y_{\max} \in [T, T+1]\} = \sum_i \mathbf{1}\{Y_i \in [T, T+1]\} - \sum_i \mathbf{1}\{Y_i \in [T, T+1]\}\mathbf{1}\{Y_i \neq Y_{\max}\}.$$

Paper, §2.2.1 and Corollary 5.3.

k responses

$$\mathbf{1}\{Y_{\max} \in [T, T+1]\} = \sum_i \mathbf{1}\{Y_i \in [T, T+1]\} - \sum_i \mathbf{1}\{Y_i \in [T, T+1]\}\mathbf{1}\{Y_i \neq Y_{\max}\}.$$

$$\mathsf{R}_d = -\sum_i \mathbb{E}[\mathsf{H}_d(\boldsymbol{X})\mathbf{1}\{Y_i \in [T, T+1]\}\mathbf{1}\{Y_i \neq Y_{\max}\}]\,.$$

Paper, §2.2.1 and Corollary 5.3.

1. Remainder is Small

Fix $Y_i = t \in [T, T+1]$. Using the angular separation condition,

$$\Pr(Y_{\max} > Y_i \mid Y_i = t) \leq \sum_{j \neq i} \Pr(Y_j > t \mid Y_i = t)$$

Paper, §2.2.1 and §5.1.2.

1. Remainder is Small

Fix $Y_i = t \in [T, T+1]$. Using the angular separation condition,

$$\Pr(Y_{\max} > Y_i \mid Y_i = t) \leq \sum_{j \neq i} \Pr(Y_j > t \mid Y_i = t)$$

$$\leq k e^{-\Omega(T^2)}$$

Paper, §2.2.1 and §5.1.2.

1. Remainder is Small

Fix $Y_i = t \in [T, T+1]$. Using the angular separation condition,

$$\Pr(Y_{\max} > Y_i \mid Y_i = t) \leq \sum_{j \neq i} \Pr(Y_j > t \mid Y_i = t)$$

$$\leq k e^{-\Omega(T^2)} \leq \mathrm{poly}(1/n).$$

Paper, §2.2.1 and §5.1.2.

2. Nearly Orthogonal Directions

$$\left|\langle \boldsymbol{u}_i^{\otimes d}, \boldsymbol{u}_j^{\otimes d} \rangle\right| = |\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle|^d$$

Paper, §2.2.1–2.2.2.

2. Nearly Orthogonal Directions

$$\left|\langle \boldsymbol{u}_i^{\otimes d}, \boldsymbol{u}_j^{\otimes d} \rangle\right| = |\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle|^d$$
$$\leq (1 - \Delta)^d$$

Paper, §2.2.1–2.2.2.

2. Nearly Orthogonal Directions

$$\begin{aligned}\left|\langle \boldsymbol{u}_i^{\otimes d}, \boldsymbol{u}_j^{\otimes d}\rangle\right| &= |\langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle|^d \\ &\leq (1-\Delta)^d \\ &\leq e^{-\Delta d} \leq \mathrm{poly}(1/n) \qquad (d = \Theta(\log n)).\end{aligned}$$

Paper, §2.2.1–2.2.2.

2. Nearly Orthogonal Directions

$$\begin{aligned}\left|\langle \boldsymbol{u}_i^{\otimes d}, \boldsymbol{u}_j^{\otimes d}\rangle\right| &= |\langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle|^d \\ &\leq (1-\Delta)^d \\ &\leq e^{-\Delta d} \leq \mathrm{poly}(1/n) \qquad (d = \Theta(\log n)).\end{aligned}$$

Lifted vectors have n^d entries.

Paper, §2.2.1–2.2.2.

Table of Contents

Motivation

Problem Definition

Prior Works

Our Contributions

Proof Overview
- Structural Lemma
- Algorithmic Implications
- Technical Challenges

Algorithm

1. **Implicit projections**

$$\widetilde{\boldsymbol{X}}^{\otimes d} \longmapsto C_d \widetilde{\boldsymbol{X}}^{\otimes d}.$$

C_d projects into a small space.

Algorithm

1. **Implicit projections**

$$\widetilde{\boldsymbol{X}}^{\otimes d} \longmapsto C_d \widetilde{\boldsymbol{X}}^{\otimes d}.$$

   C_d projects into a small space.

2. **Spectral recovery**

$$\left(\widetilde{\boldsymbol{X}}, C_d \widetilde{\boldsymbol{X}}^{\otimes d}, Y_{\max}\right) \overset{\text{from samples}}{\longmapsto} \widehat{\boldsymbol{u}}_i \approx \boldsymbol{u}_i.$$

Algorithm

1. **Implicit projections**

$$\widetilde{\boldsymbol{X}}^{\otimes d} \longmapsto C_d \widetilde{\boldsymbol{X}}^{\otimes d}.$$

   C_d projects into a small space.

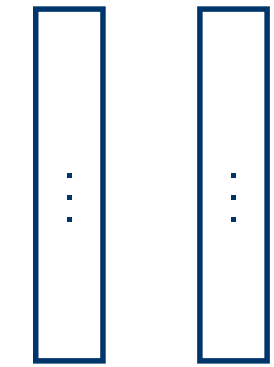
2. **Spectral recovery**

$$\left(\widetilde{\boldsymbol{X}}, C_d \widetilde{\boldsymbol{X}}^{\otimes d}, Y_{\max}\right) \overset{\text{from samples}}{\longmapsto} \widehat{\boldsymbol{u}}_i \approx \boldsymbol{u}_i.$$

3. **Restricted least squares**

$$\widehat{\boldsymbol{u}}_i \longmapsto \widehat{\boldsymbol{w}}_i.$$

Step 1. Implicit Projections



lifted vectors

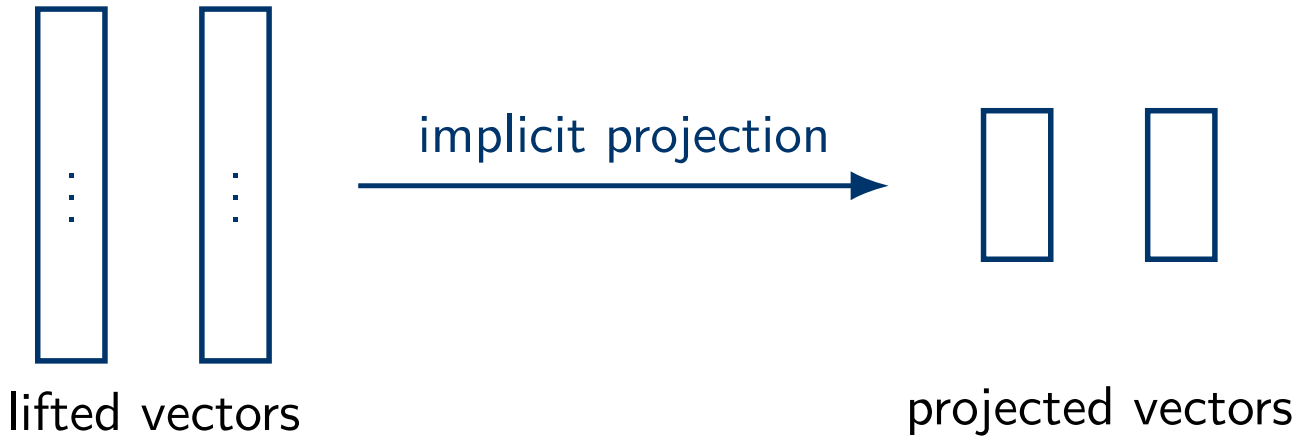
- Given $(\widetilde{\boldsymbol{X}}, Y_{\max})$, compute $C_d \widetilde{\boldsymbol{X}}^{\otimes d}$ in polynomial time.

Step 1. Implicit Projections



- Given $(\widetilde{\boldsymbol{X}}, Y_{\max})$, compute $C_d \widetilde{\boldsymbol{X}}^{\otimes d}$ in polynomial time.
- Implicit projections approximately preserve inner products.
  [LL24, SICOMP] [DK25, FOCS]

Step 1. Implicit Projections



- Given $(\widetilde{\boldsymbol{X}}, Y_{\max})$, compute $C_d \widetilde{\boldsymbol{X}}^{\otimes d}$ in polynomial time.
- Implicit projections approximately preserve inner products. [LL24, SICOMP] [DK25, FOCS]
  - New analysis showing subroutine is robust to bias.

Step 2. Spectral Recovery

Even $\displaystyle\sum_i \lambda_i C_{2d} \boldsymbol{u}_i^{\otimes 2d}$

Odd $\displaystyle\sum_i \lambda_i C_{2d} \boldsymbol{u}_i^{\otimes 2d} \otimes \boldsymbol{u}_i$

Project the two degree-d blocks of each moment.

Paper, §2.2.2–2.2.3.

Step 2. Spectral Recovery

Even $\displaystyle\sum_i \lambda_i C_{2d} \boldsymbol{u}_i^{\otimes 2d}$

Odd $\displaystyle\sum_i \lambda_i \left[C_{2d} \boldsymbol{u}_i^{\otimes 2d} \right] \underbrace{\langle \boldsymbol{u}_i, \boldsymbol{b} \rangle}_{\text{contract by } \boldsymbol{b}}$

Contract the last factor along the probe $\boldsymbol{b}$.

Paper, §2.2.2–2.2.3.

Step 2. Spectral Recovery

$$E = \sum_i \lambda_i (C_d \boldsymbol{u}_i^{\otimes d})(C_d \boldsymbol{u}_i^{\otimes d})^\top$$

$$O(\boldsymbol{b}) = \sum_i \lambda_i \underbrace{\langle \boldsymbol{u}_i, \boldsymbol{b} \rangle}_{\text{contracted factor}} (C_d \boldsymbol{u}_i^{\otimes d})(C_d \boldsymbol{u}_i^{\otimes d})^\top$$

Flatten the two projected blocks into a matrix.

Paper, §2.2.2–2.2.3.

Step 2. Rescaling

Treat the projected lifted directions $C\boldsymbol{u}_i^{\otimes d}$ as orthonormal.

$$E^{-1/2}(C\boldsymbol{u}_i^{\otimes d}) = \lambda_i^{-1/2}(C\boldsymbol{u}_i^{\otimes d}).$$

Step 2. Rescaling

Treat the projected lifted directions $C\boldsymbol{u}_i^{\otimes d}$ as orthonormal.

$$E^{-1/2}(C\boldsymbol{u}_i^{\otimes d}) = \lambda_i^{-1/2}(C\boldsymbol{u}_i^{\otimes d}).$$

$$E^{-1/2}O(\boldsymbol{b})E^{-1/2} = \sum_i \underbrace{\langle \boldsymbol{u}_i, \boldsymbol{b}\rangle}_{\text{eigenvalue}} (C\boldsymbol{u}_i^{\otimes d})(C\boldsymbol{u}_i^{\otimes d})^\top.$$

- Given $E, O(\boldsymbol{b})$, recover $\widehat{\boldsymbol{u}_1}, \ldots, \widehat{\boldsymbol{u}_n}$.
  [AHK12, COLT] [HK13, ITCS]

Step 2. Rescaling

Treat the projected lifted directions $C\boldsymbol{u}_i^{\otimes d}$ as orthonormal.

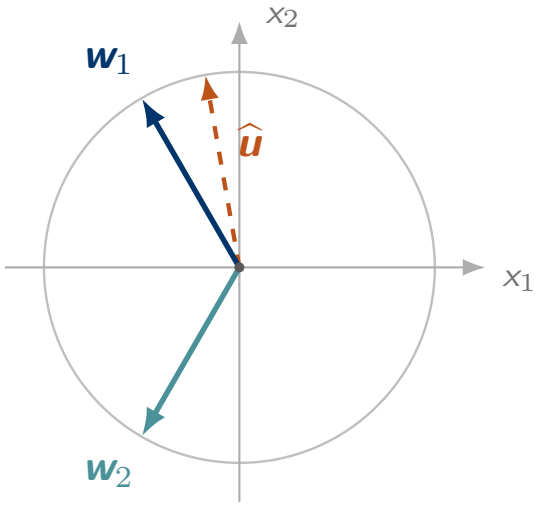
$$E^{-1/2}(C\boldsymbol{u}_i^{\otimes d}) = \lambda_i^{-1/2}(C\boldsymbol{u}_i^{\otimes d}).$$

$$E^{-1/2}O(\boldsymbol{b})E^{-1/2} = \sum_i \underbrace{\langle \boldsymbol{u}_i, \boldsymbol{b}\rangle}_{\text{eigenvalue}} (C\boldsymbol{u}_i^{\otimes d})(C\boldsymbol{u}_i^{\otimes d})^\top.$$

- Given $E, O(\boldsymbol{b})$, recover $\widehat{\boldsymbol{u}_1}, \ldots, \widehat{\boldsymbol{u}_n}$.
  [AHK12, COLT] [HK13, ITCS]
  - New analysis to handle small eigenvalues and bias.

Step 3. Restricted Least Squares

Recover $\boldsymbol{w}_i$ given good estimate $\widehat{\boldsymbol{u}} \approx \boldsymbol{w}_i / \|\boldsymbol{w}_i\|_2$,

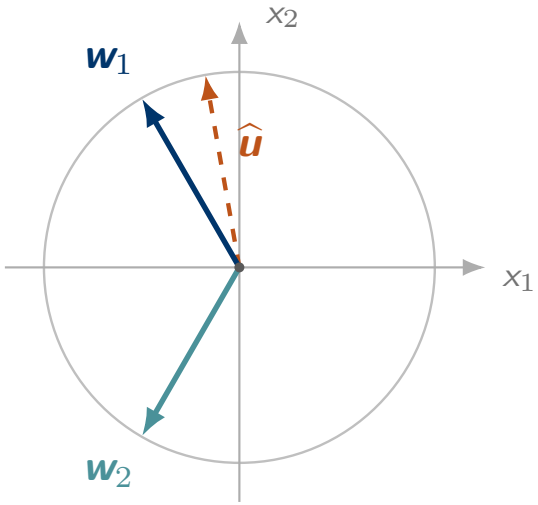


$$\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top]\boldsymbol{w}_i = \mathbb{E}[\boldsymbol{X}Y_i].$$

Conditioning idea [GM24]. Restricted least squares: paper, §2.2.4.

Step 3. Restricted Least Squares

Recover $\boldsymbol{w}_i$ given good estimate $\widehat{\boldsymbol{u}} \approx \boldsymbol{w}_i / \|\boldsymbol{w}_i\|_2$,



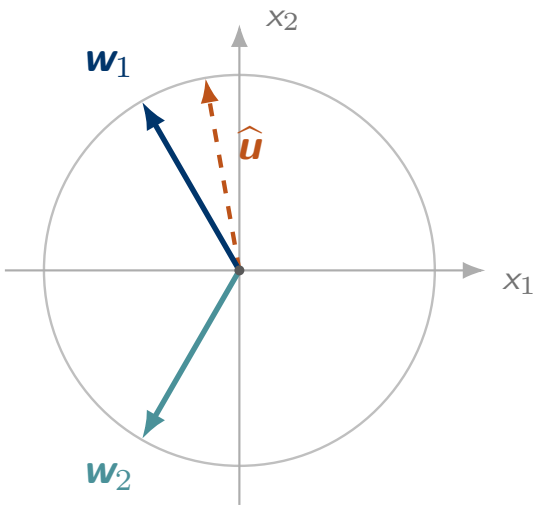
$$\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top \mid \mathcal{E}]\boldsymbol{w}_i = \mathbb{E}[\boldsymbol{X}Y_i \mid \mathcal{E}].$$

$$\mathcal{E} = \{\Gamma \le \langle \widehat{\boldsymbol{u}}, \boldsymbol{X} \rangle \le \Gamma + 1\}.$$

Conditioning idea [GM24]. Restricted least squares: paper, §2.2.4.

Step 3. Restricted Least Squares

Recover $\boldsymbol{w}_i$ given good estimate $\widehat{\boldsymbol{u}} \approx \boldsymbol{w}_i / \|\boldsymbol{w}_i\|_2$,



$$\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\top \mid \mathcal{E}]\boldsymbol{w}_i \approx \mathbb{E}[\boldsymbol{X} Y_{\max} \mid \mathcal{E}].$$

$$\mathcal{E} = \{\Gamma \le \langle \widehat{\boldsymbol{u}}, \boldsymbol{X} \rangle \le \Gamma + 1\}.$$

On this interval, $Y_i = Y_{\max}$ with high probability.
Fit regressor using least squares.

Conditioning idea [GM24]. Restricted least squares: paper, §2.2.4.

Table of Contents

Motivation

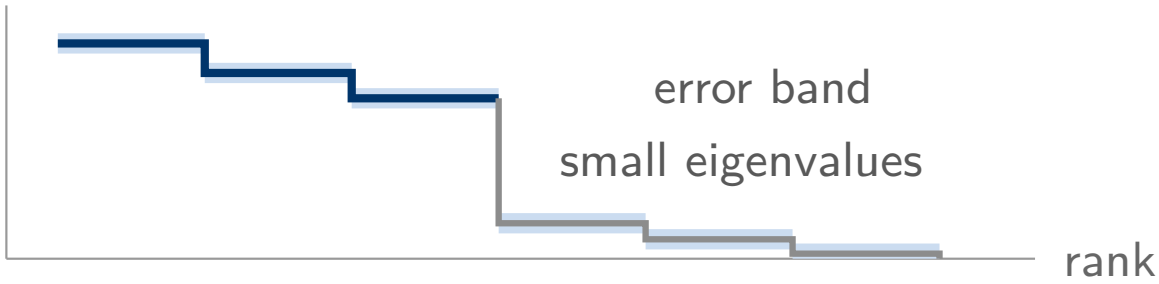
Problem Definition

Prior Works

Our Contributions

Proof Overview
- Structural Lemma
- Algorithmic Implications
- Technical Challenges

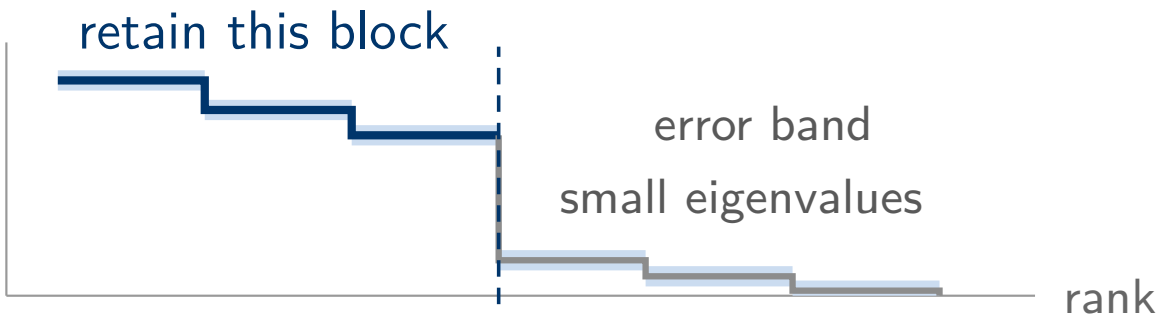
Technical Challenges



- Irreducible remainder can significantly distort smaller eigenspaces.

Paper, §2.2.3 and §2.2.5.

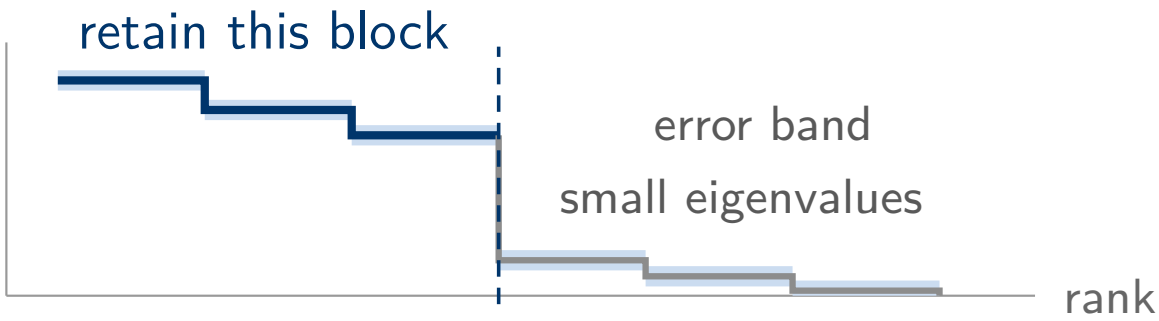
Technical Challenges



- Irreducible remainder can significantly distort smaller eigenspaces.
- Invert only the leading block.
  Recover a subset of the regressors.

Paper, §2.2.3 and §2.2.5.

Technical Challenges



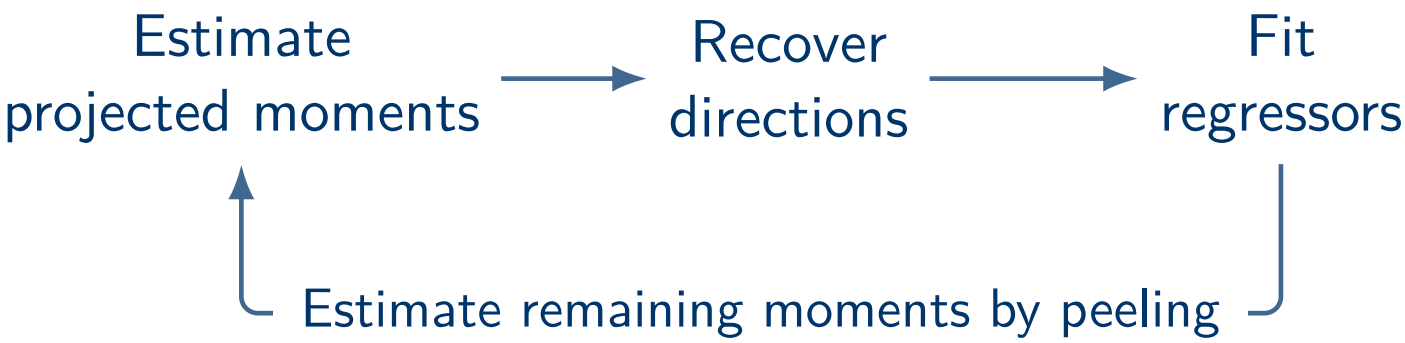
- Irreducible remainder can significantly distort smaller eigenspaces.
- Invert only the leading block.
  Recover a subset of the regressors.
- Peel and repeat. Use the fitted regressors to reweight observations and estimate the remaining restricted moments.

Paper, §2.2.3 and §2.2.5.

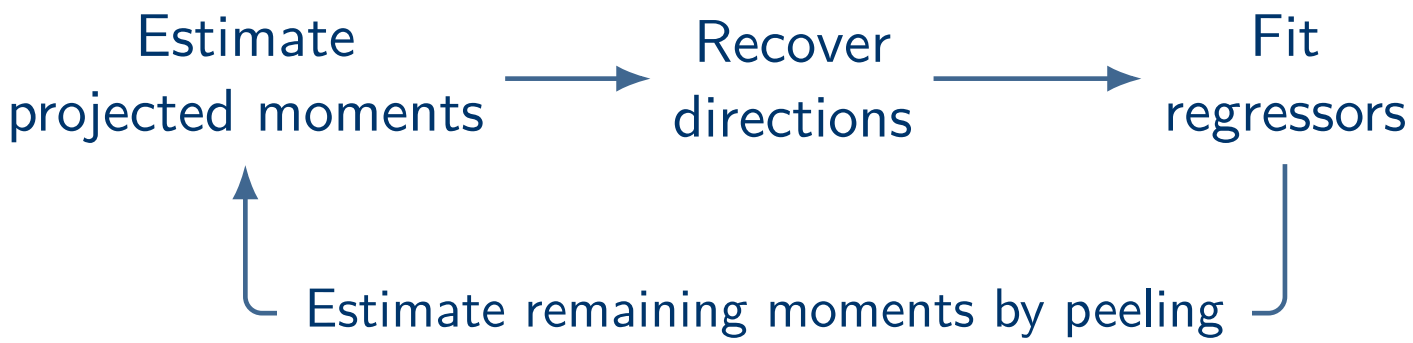
Overall Algorithm



Overall Algorithm



Overall Algorithm



At least one new regressor per round. At most k rounds.

Summary

- **Structural lemma.** Restricted Hermite moments uncover the regressor directions.

Summary

- **Structural lemma.** Restricted Hermite moments uncover the regressor directions.
- **Tensor decomposition.** Implicit projections and spectral algorithm recovers regressor directions.

Summary

- **Structural lemma.** Restricted Hermite moments uncover the regressor directions.
- **Tensor decomposition.** Implicit projections and spectral algorithm recovers regressor directions.
- **Restricted least squares.** Full regressors from rough directions.
  - Constant warm-start radius [cf. KMZ26, COLT].

Summary

- **Structural lemma.** Restricted Hermite moments uncover the regressor directions.
- **Tensor decomposition.** Implicit projections and spectral algorithm recovers regressor directions.
- **Restricted least squares.** Full regressors from rough directions.
  - Constant warm-start radius [cf. KMZ26, COLT].

First polynomial-time algorithm
for maximum self-selection with unknown index.

Future Directions

These techniques may also apply to

- Mixtures of linear regressions [CLS20, STOC] [DK25, FOCS]

Future Directions

These techniques may also apply to

- Mixtures of linear regressions [CLS20, STOC] [DK25, FOCS]
- Max-affine regression [GPGR22, IEEE Trans. Inf. Theory]

Future Directions

These techniques may also apply to

- Mixtures of linear regressions [CLS20, STOC] [DK25, FOCS]
- Max-affine regression [GPGR22, IEEE Trans. Inf. Theory]
- Gaussian multi-index models [DIKZ25, FOCS]