Support Recovery for Orthogonal Matching Pursuit

Upper and Lower bounds

Raghav Somani, Chirag Gupta, Prateek Jain and Praneeth Netrapalli

September 24, 2018

Sparse Regression

$$\bar{\mathbf{x}} = \arg\min_{\|\mathbf{x}\|_0 \leq s^*} f(\mathbf{x}) \tag{1.1}$$

$\mathbf{x} \in \mathbb{R}^d$ and $s^* << d$.
ℓ_0 norm counts the number of non-zero elements.

Applications

- Resource constrained Machine Learning
- High dimensional Statistics
- Bioinformatics

Sparse Regression

$$\bar{\mathbf{x}} = \underset{\|\mathbf{x}\|_0 \leq s^*}{\arg\min} f(\mathbf{x}) \tag{1.1}$$

$\mathbf{x} \in \mathbb{R}^d$ and $s^* << d$.
ℓ_0 norm counts the number of non-zero elements.

Applications

- Resource constrained Machine Learning
- High dimensional Statistics
- Bioinformatics

Sparse Linear Regression (SLR)

- Sparse *Linear* Regression is a representative problem. Results typically extend easily to general case.
- With $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$, SLR's objective is to find

$$\bar{\mathbf{x}} = \arg\min_{\|\mathbf{x}\|_0 \leq s^*} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \tag{2.1}$$

  where $\mathbf{A} \in \mathbb{R}^{n \times d}, \mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} = \mathbb{R}^n$.
- Unconditionally, it is NP hard (reduction to 3 set cover problem).

Sparse Linear Regression (SLR)

- Sparse *Linear* Regression is a representative problem. Results typically extend easily to general case.
- With $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$, SLR's objective is to find

$$\bar{\mathbf{x}} = \underset{\|\mathbf{x}\|_0 \leq s^*}{\arg\min} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \tag{2.1}$$

  where $\mathbf{A} \in \mathbb{R}^{n \times d}, \mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} = \mathbb{R}^n$.
- Unconditionally, it is NP hard (reduction to 3 set cover problem).

Sparse Linear Regression (SLR)

- Sparse *Linear* Regression is a representative problem. Results typically extend easily to general case.
- With $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$, SLR's objective is to find

$$\bar{\mathbf{x}} = \arg\min_{\|\mathbf{x}\|_0 \leq s^*} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \tag{2.1}$$

  where $\mathbf{A} \in \mathbb{R}^{n \times d}, \mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} = \mathbb{R}^n$.
- Unconditionally, it is NP hard (reduction to 3 set cover problem).

Assumptions of interest

Despite being NP hard, SLR is tractable under certain assumptions.

- **Incoherence** -
  If $\Sigma = \mathbf{A}^T\mathbf{A}$, then $\max_{i \neq j} |\Sigma_{ij}| \leq M$
  If $M \leq \frac{1}{2s^*-1}$ and $\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} \implies \bar{\mathbf{x}}$ is unique sparsest solution, and OMP can recover $\bar{\mathbf{x}}$ in s^* steps.
- **Restricted Isometry Property (RIP)** -
  $\left\|\mathbf{A}_{\mathrm{S}}^T\mathbf{A}_{\mathrm{S}} - \mathbf{I}\right\|_2 \leq \delta_{|\mathrm{S}|} \qquad (\delta_s \leq M(s-1) \quad \forall\, s \geq 2)$
  $\implies (1-\delta_s)\|\mathbf{v}\|_2^2 \leq \|\mathbf{A}\mathbf{v}\|_2^2 \leq (1+\delta_s)\|\mathbf{v}\|_2^2 \quad \forall\, \mathbf{v}$ s.t. $\|\mathbf{v}\|_0 \leq s$.
- **Null space property** -
  $\forall\, \mathrm{S} \in [d]$ s.t. $|\mathrm{S}| \leq s$, if $\mathbf{v} \in \mathrm{Null}(\mathbf{A}) \setminus \{\mathbf{0}\}$, then $\|\mathbf{v}_{\mathrm{S}}\|_1 \leq \|\mathbf{v}_{\mathrm{S}^c}\|_1$
  $\implies \{\mathbf{v} \in \mathbb{R}^d \mid \mathbf{A}\mathbf{v} = \mathbf{0}\} \cap \{\mathbf{v} \in \mathbb{R}^d \mid \|\mathbf{v}_{\mathrm{S}^c}\|_1 \leq \|\mathbf{v}_{\mathrm{S}}\|_1\} = \mathbf{0}$
- **Restricted Strong Convexity (RSC)** -
  $\|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z}\|_2^2 \geq \rho_s^- \|\mathbf{x}-\mathbf{z}\|_2^2 \;\forall\, \mathbf{x}, \mathbf{z} \in \mathbb{R}^d$ s.t. $\|\mathbf{x}-\mathbf{z}\|_0 \leq s$

Incoherence $\implies$ RIP $\implies$ Null space property $\implies$ RSC
RSC is the weakest and the most popular assumption.

Assumptions of interest

Despite being NP hard, SLR is tractable under certain assumptions.

- **Incoherence** -
  If $\mathbf{\Sigma} = \mathbf{A}^T\mathbf{A}$, then $\max_{i \neq j} |\mathbf{\Sigma}_{ij}| \leq M$
  If $M \leq \frac{1}{2s^*-1}$ and $\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} \implies \bar{\mathbf{x}}$ is unique sparsest solution, and OMP can recover $\bar{\mathbf{x}}$ in s^* steps.
- **Restricted Isometry Property (RIP)** -
  $\left\|\mathbf{A}_\mathbf{S}^T\mathbf{A}_\mathbf{S} - \mathbf{I}\right\|_2 \leq \delta_{|\mathbf{S}|} \qquad (\delta_s \leq M(s-1) \quad \forall\, s \geq 2)$
  $\implies (1-\delta_s)\|\mathbf{v}\|_2^2 \leq \|\mathbf{A}\mathbf{v}\|_2^2 \leq (1+\delta_s)\|\mathbf{v}\|_2^2 \quad \forall\, \mathbf{v}$ s.t. $\|\mathbf{v}\|_0 \leq s$.
- **Null space property** -
  $\forall\, \mathbf{S} \in [d]$ s.t. $|\mathbf{S}| \leq s$, if $\mathbf{v} \in$ Null($\mathbf{A}$) $\setminus \{\mathbf{0}\}$, then $\|\mathbf{v}_\mathbf{S}\|_1 \leq \|\mathbf{v}_{\mathbf{S}^c}\|_1$
  $\implies \left\{\mathbf{v} \in \mathbb{R}^d \mid \mathbf{A}\mathbf{v} = \mathbf{0}\right\} \cap \left\{\mathbf{v} \in \mathbb{R}^d \mid \|\mathbf{v}_{\mathbf{S}^c}\|_1 \leq \|\mathbf{v}_\mathbf{S}\|_1\right\} = \mathbf{0}$
- **Restricted Strong Convexity (RSC)** -
  $\|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z}\|_2^2 \geq \rho_s^- \|\mathbf{x} - \mathbf{z}\|_2^2 \;\forall\, \mathbf{x}, \mathbf{z} \in \mathbb{R}^d$ s.t. $\|\mathbf{x} - \mathbf{z}\|_0 \leq s$

Incoherence $\implies$ RIP $\implies$ Null space property $\implies$ RSC
RSC is the weakest and the most popular assumption.

Assumptions of interest

Despite being NP hard, SLR is tractable under certain assumptions.

- **Incoherence** -
  If $\mathbf{\Sigma} = \mathbf{A}^T\mathbf{A}$, then $\max_{i \neq j} |\mathbf{\Sigma}_{ij}| \leq M$
  If $M \leq \frac{1}{2s^*-1}$ and $\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} \implies \bar{\mathbf{x}}$ is unique sparsest solution, and OMP can recover $\bar{\mathbf{x}}$ in s^* steps.
- **Restricted Isometry Property (RIP)** -
  $\left\|\mathbf{A}_{\mathbf{S}}^T\mathbf{A}_{\mathbf{S}} - \mathbf{I}\right\|_2 \leq \delta_{|\mathbf{S}|} \qquad (\delta_s \leq M(s-1) \quad \forall\, s \geq 2)$
  $\implies (1-\delta_s)\|\mathbf{v}\|_2^2 \leq \|\mathbf{A}\mathbf{v}\|_2^2 \leq (1+\delta_s)\|\mathbf{v}\|_2^2 \quad \forall\, \mathbf{v}$ s.t. $\|\mathbf{v}\|_0 \leq s$.
- **Null space property** -
  $\forall\, \mathbf{S} \in [d]$ s.t. $|\mathbf{S}| \leq s$, if $\mathbf{v} \in \mathrm{Null}(\mathbf{A}) \setminus \{\mathbf{0}\}$, then $\|\mathbf{v}_{\mathbf{S}}\|_1 \leq \|\mathbf{v}_{\mathbf{S}^c}\|_1$
  $\implies \left\{\mathbf{v} \in \mathbb{R}^d \mid \mathbf{A}\mathbf{v} = \mathbf{0}\right\} \cap \left\{\mathbf{v} \in \mathbb{R}^d \mid \|\mathbf{v}_{\mathbf{S}^c}\|_1 \leq \|\mathbf{v}_{\mathbf{S}}\|_1\right\} = \mathbf{0}$
- **Restricted Strong Convexity (RSC)** -
  $\|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z}\|_2^2 \geq \rho_s^- \|\mathbf{x} - \mathbf{z}\|_2^2 \ \forall\, \mathbf{x}, \mathbf{z} \in \mathbb{R}^d$ s.t. $\|\mathbf{x} - \mathbf{z}\|_0 \leq s$

Incoherence $\implies$ RIP $\implies$ Null space property $\implies$ RSC
RSC is the weakest and the most popular assumption.

Assumptions of interest

Despite being NP hard, SLR is tractable under certain assumptions.

- **Incoherence** -

  If $\mathbf{\Sigma} = \mathbf{A}^T\mathbf{A}$, then $\max_{i \neq j} |\mathbf{\Sigma}_{ij}| \leq M$

  If $M \leq \frac{1}{2s^*-1}$ and $\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} \implies \bar{\mathbf{x}}$ is unique sparsest solution, and OMP can recover $\bar{\mathbf{x}}$ in s^* steps.

- **Restricted Isometry Property (RIP)** -

  $\left\|\mathbf{A}_{\mathbf{S}}^T\mathbf{A}_{\mathbf{S}} - \mathbf{I}\right\|_2 \leq \delta_{|\mathbf{S}|} \qquad (\delta_s \leq M(s-1) \quad \forall\ s \geq 2)$

  $\implies (1-\delta_s)\|\mathbf{v}\|_2^2 \leq \|\mathbf{A}\mathbf{v}\|_2^2 \leq (1+\delta_s)\|\mathbf{v}\|_2^2 \quad \forall\ \mathbf{v}$ s.t. $\|\mathbf{v}\|_0 \leq s$.

- **Null space property** -

  $\forall\ \mathbf{S} \in [d]$ s.t. $|\mathbf{S}| \leq s$, if $\mathbf{v} \in \mathrm{Null}(\mathbf{A}) \setminus \{\mathbf{0}\}$, then $\|\mathbf{v}_{\mathbf{S}}\|_1 \leq \|\mathbf{v}_{\mathbf{S}^c}\|_1$

  $\implies \left\{\mathbf{v} \in \mathbb{R}^d \mid \mathbf{A}\mathbf{v} = \mathbf{0}\right\} \cap \left\{\mathbf{v} \in \mathbb{R}^d \mid \|\mathbf{v}_{\mathbf{S}^c}\|_1 \leq \|\mathbf{v}_{\mathbf{S}}\|_1\right\} = \mathbf{0}$

- **Restricted Strong Convexity (RSC)** -

  $\|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z}\|_2^2 \geq \rho_s^- \|\mathbf{x} - \mathbf{z}\|_2^2\ \forall\ \mathbf{x}, \mathbf{z} \in \mathbb{R}^d$ s.t. $\|\mathbf{x} - \mathbf{z}\|_0 \leq s$

Incoherence $\implies$ RIP $\implies$ Null space property $\implies$ RSC
RSC is the weakest and the most popular assumption.

Assumptions of interest

Despite being NP hard, SLR is tractable under certain assumptions.

- **Incoherence** -
  If $\mathbf{\Sigma} = \mathbf{A}^T\mathbf{A}$, then $\max_{i \neq j} |\mathbf{\Sigma}_{ij}| \leq M$
  If $M \leq \frac{1}{2s^*-1}$ and $\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} \implies \bar{\mathbf{x}}$ is unique sparsest solution, and OMP can recover $\bar{\mathbf{x}}$ in s^* steps.
- **Restricted Isometry Property (RIP)** -
  $\left\|\mathbf{A}_{\mathbf{S}}^T\mathbf{A}_{\mathbf{S}} - \mathbf{I}\right\|_2 \leq \delta_{|\mathbf{S}|} \qquad (\delta_s \leq M(s-1) \quad \forall\, s \geq 2)$
  $\implies (1-\delta_s)\|\mathbf{v}\|_2^2 \leq \|\mathbf{A}\mathbf{v}\|_2^2 \leq (1+\delta_s)\|\mathbf{v}\|_2^2 \quad \forall\, \mathbf{v}$ s.t. $\|\mathbf{v}\|_0 \leq s$.
- **Null space property** -
  $\forall\, \mathbf{S} \in [d]$ s.t. $|\mathbf{S}| \leq s$, if $\mathbf{v} \in \text{Null}(\mathbf{A}) \setminus \{\mathbf{0}\}$, then $\|\mathbf{v}_{\mathbf{S}}\|_1 \leq \|\mathbf{v}_{\mathbf{S}^c}\|_1$
  $\implies \left\{\mathbf{v} \in \mathbb{R}^d \mid \mathbf{A}\mathbf{v} = \mathbf{0}\right\} \cap \left\{\mathbf{v} \in \mathbb{R}^d \mid \|\mathbf{v}_{\mathbf{S}^c}\|_1 \leq \|\mathbf{v}_{\mathbf{S}}\|_1\right\} = \mathbf{0}$
- **Restricted Strong Convexity (RSC)** -
  $\|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z}\|_2^2 \geq \rho_s^- \|\mathbf{x} - \mathbf{z}\|_2^2 \;\forall\, \mathbf{x}, \mathbf{z} \in \mathbb{R}^d$ s.t. $\|\mathbf{x} - \mathbf{z}\|_0 \leq s$

Incoherence $\implies$ RIP $\implies$ Null space property $\implies$ RSC
RSC is the weakest and the most popular assumption.

Assumptions of interest

Despite being NP hard, SLR is tractable under certain assumptions.

- **Incoherence** -
  If $\mathbf{\Sigma} = \mathbf{A}^T\mathbf{A}$, then $\max_{i \neq j} |\mathbf{\Sigma}_{ij}| \leq M$
  If $M \leq \frac{1}{2s^*-1}$ and $\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} \implies \bar{\mathbf{x}}$ is unique sparsest solution, and OMP can recover $\bar{\mathbf{x}}$ in s^* steps.
- **Restricted Isometry Property (RIP)** -
  $\left\|\mathbf{A}_\mathbf{S}^T\mathbf{A}_\mathbf{S} - \mathbf{I}\right\|_2 \leq \delta_{|\mathbf{S}|} \qquad (\delta_s \leq M(s-1) \quad \forall\, s \geq 2)$
  $\implies (1-\delta_s)\|\mathbf{v}\|_2^2 \leq \|\mathbf{A}\mathbf{v}\|_2^2 \leq (1+\delta_s)\|\mathbf{v}\|_2^2 \quad \forall\, \mathbf{v}$ s.t. $\|\mathbf{v}\|_0 \leq s$.
- **Null space property** -
  $\forall\, \mathbf{S} \in [d]$ s.t. $|\mathbf{S}| \leq s$, if $\mathbf{v} \in \mathrm{Null}(\mathbf{A}) \setminus \{\mathbf{0}\}$, then $\|\mathbf{v}_\mathbf{S}\|_1 \leq \|\mathbf{v}_{\mathbf{S}^c}\|_1$
  $\implies \left\{\mathbf{v} \in \mathbb{R}^d \mid \mathbf{A}\mathbf{v} = \mathbf{0}\right\} \cap \left\{\mathbf{v} \in \mathbb{R}^d \mid \|\mathbf{v}_{\mathbf{S}^c}\|_1 \leq \|\mathbf{v}_\mathbf{S}\|_1\right\} = \mathbf{0}$
- **Restricted Strong Convexity (RSC)** -
  $\|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z}\|_2^2 \geq \rho_s^- \|\mathbf{x}-\mathbf{z}\|_2^2 \;\forall\, \mathbf{x}, \mathbf{z} \in \mathbb{R}^d$ s.t. $\|\mathbf{x}-\mathbf{z}\|_0 \leq s$

Incoherence $\implies$ RIP $\implies$ Null space property $\implies$ RSC
RSC is the weakest and the most popular assumption.

Goals of SLR

SLR can be modelled as

$$\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} + \boldsymbol{\eta} \tag{2.2}$$

where $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n \times n})$, $\text{supp}(\bar{\mathbf{x}}) = \mathbf{S}^*$ and $|\mathbf{S}^*| = s^*$.

$$\implies \mathbf{y} = \mathbf{A}_{\mathbf{S}^*}\bar{\mathbf{x}}_{\mathbf{S}^*} + \boldsymbol{\eta} \tag{2.3}$$

Model with deterministic conditions on η can also be analyzed.

Goals of SLR

1. **Bounding Generalization error** - Upper bound $G(\mathbf{x}) := \frac{1}{n}\|\mathbf{A}(\mathbf{x} - \bar{\mathbf{x}})\|_2^2$ where the rows of $\mathbf{A}$ are i.i.d.
2. **Support Recovery** - Recover the true features of $\mathbf{A}$, i.e., find a $\mathbf{S} \supseteq \mathbf{S}^*$.

Goals of SLR

SLR can be modelled as

$$\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} + \boldsymbol{\eta} \tag{2.2}$$

where $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n\times n})$, supp$(\bar{\mathbf{x}}) = \mathbf{S}^*$ and $|\mathbf{S}^*| = s^*$.

$$\implies \mathbf{y} = \mathbf{A}_{\mathbf{S}^*}\bar{\mathbf{x}}_{\mathbf{S}^*} + \boldsymbol{\eta} \tag{2.3}$$

Model with deterministic conditions on $\boldsymbol{\eta}$ can also be analyzed.

Goals of SLR

1. **Bounding Generalization error** - Upper bound $G(\mathbf{x}) := \frac{1}{n}\|\mathbf{A}(\mathbf{x} - \bar{\mathbf{x}})\|_2^2$ where the rows of $\mathbf{A}$ are i.i.d.
2. **Support Recovery** - Recover the true features of $\mathbf{A}$, i.e., find a $\mathbf{S} \supseteq \mathbf{S}^*$.

Goals of SLR

SLR can be modelled as

$$\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} + \boldsymbol{\eta} \tag{2.2}$$

where $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n \times n})$, supp$(\bar{\mathbf{x}}) = \mathbf{S}^*$ and $|\mathbf{S}^*| = s^*$.

$$\implies \mathbf{y} = \mathbf{A}_{\mathbf{S}^*}\bar{\mathbf{x}}_{\mathbf{S}^*} + \boldsymbol{\eta} \tag{2.3}$$

Model with deterministic conditions on η can also be analyzed.

Goals of SLR

1. **Bounding Generalization error** - Upper bound $G(\mathbf{x}) := \frac{1}{n}\|\mathbf{A}(\mathbf{x} - \bar{\mathbf{x}})\|_2^2$ where the rows of $\mathbf{A}$ are i.i.d.
2. **Support Recovery** - Recover the true features of $\mathbf{A}$, i.e., find a $\mathrm{S} \supseteq \mathrm{S}^*$.

Goals of SLR

SLR can be modelled as

$$\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} + \boldsymbol{\eta} \tag{2.2}$$

where $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n\times n})$, $\text{supp}(\bar{\mathbf{x}}) = \mathbf{S}^*$ and $|\mathbf{S}^*| = s^*$.

$$\implies \mathbf{y} = \mathbf{A}_{\mathbf{S}^*}\bar{\mathbf{x}}_{\mathbf{S}^*} + \boldsymbol{\eta} \tag{2.3}$$

Model with deterministic conditions on η can also be analyzed.

Goals of SLR

1. **Bounding Generalization error** - Upper bound $G(\mathbf{x}) := \frac{1}{n}\|\mathbf{A}(\mathbf{x} - \bar{\mathbf{x}})\|_2^2$ where the rows of $\mathbf{A}$ are i.i.d.
2. **Support Recovery** - Recover the true features of $\mathbf{A}$, i.e., find a $\mathbf{S} \supseteq \mathbf{S}^*$.

Algorithms to solve SLR

The literature mainly studies 3 classes of algorithms

Existing SLR algorithms

- ℓ_1 minimization based (LASSO based). E.g. - Dantzig selector
- Non-convex penalty based. E.g. - IHT, SCAD penalty, Log-sum penalty
- Greedy methods. E.g. - Orthogonal Matching Pursuit (OMP)

We study SLR under RSC assumption for OMP algorithm.

Algorithms to solve SLR

The literature mainly studies 3 classes of algorithms

Existing SLR algorithms

- ℓ_1 minimization based (LASSO based). E.g. - Dantzig selector
- Non-convex penalty based. E.g. - IHT, SCAD penalty, Log-sum penalty
- Greedy methods. E.g. - Orthogonal Matching Pursuit (OMP)

We study SLR under RSC assumption for OMP algorithm.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.

$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.

At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\ |\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1}\rangle| \quad |\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1}\rangle| \quad \cdots \quad |\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1}\rangle| \quad \cdots \quad |\langle \mathbf{A}_{i_p}, \mathbf{r}_{k-1}\rangle|\ \right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.
$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.
At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\, |\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1} \rangle| \;\; |\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1} \rangle| \;\cdots\; |\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1} \rangle| \;\cdots\; |\langle \mathbf{A}_{i_{n-k+1}}, \mathbf{r}_{k-1} \rangle| \,\right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.
$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.
At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\left|\left\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1} \right\rangle\right| \ \left|\left\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1} \right\rangle\right| \ \ldots \ \left|\left\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1} \right\rangle\right| \ \ldots \ \left|\left\langle \mathbf{A}_{i_{d-k+1}}, \mathbf{r}_{k-1} \right\rangle\right| \right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.
$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.
At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\, |\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1} \rangle| \;\; |\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1} \rangle| \;\; \ldots \;\; \left|\left\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1} \right\rangle\right| \;\; \ldots \;\; \left|\left\langle \mathbf{A}_{i_{d-k+1}}, \mathbf{r}_{k-1} \right\rangle\right| \,\right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.
$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.
At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\left|\left\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1} \right\rangle\right| \;\; \left|\left\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1} \right\rangle\right| \;\; \ldots \;\; \left|\left\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1} \right\rangle\right| \;\; \ldots \;\; \left|\left\langle \mathbf{A}_{i_{d-k+1}}, \mathbf{r}_{k-1} \right\rangle\right| \right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.
$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.
At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\left|\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1} \rangle\right| \ \left|\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1} \rangle\right| \ \dots \left|\left\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1} \right\rangle\right| \dots \ \left|\left\langle \mathbf{A}_{i_{d-k+1}}, \mathbf{r}_{k-1} \right\rangle\right| \right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.
$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.
At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\left|\left\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1} \right\rangle\right| \; \left|\left\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1} \right\rangle\right| \; \ldots \; \left|\left\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1} \right\rangle\right| \; \ldots \; \left|\left\langle \mathbf{A}_{i_{d-k+1}}, \mathbf{r}_{k-1} \right\rangle\right| \right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.
$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.
At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\left|\left\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1} \right\rangle\right| \ \left|\left\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1} \right\rangle\right| \ \ldots \ \left|\left\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1} \right\rangle\right| \ \ldots \ \left|\left\langle \mathbf{A}_{i_{d-k+1}}, \mathbf{r}_{k-1} \right\rangle\right| \right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.
$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.
At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\left|\left\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1} \right\rangle\right| \; \left|\left\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1} \right\rangle\right| \; \ldots \; \left|\left\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1} \right\rangle\right| \; \ldots \; \left|\left\langle \mathbf{A}_{i_{d-k+1}}, \mathbf{r}_{k-1} \right\rangle\right| \right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.
$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.
At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\left|\left\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1} \right\rangle\right| \ \left|\left\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1} \right\rangle\right| \ \ldots \ \left|\left\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1} \right\rangle\right| \ \ldots \ \left|\left\langle \mathbf{A}_{i_{d-k+1}}, \mathbf{r}_{k-1} \right\rangle\right| \right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Set initial support set $\mathbf{S}_0 = \phi$ & $\mathbf{x}_0 = \mathbf{0}$.
$\therefore$ residual $\mathbf{r}_0 = \mathbf{y} - \mathbf{A}\mathbf{x}_0 = \mathbf{y}$.
At k^{th} iteration ($k \geq 1$)

- From the left-over columns of $\mathbf{A}$ (in $\mathbf{A}_{\mathbf{S}^c_{k-1}}$), find the column with maximum absolute inner product with $\mathbf{r}_{k-1}$.

$$\left[\left|\left\langle \mathbf{A}_{i_1}, \mathbf{r}_{k-1} \right\rangle\right| \ \left|\left\langle \mathbf{A}_{i_2}, \mathbf{r}_{k-1} \right\rangle\right| \ \ldots \ \left|\left\langle \mathbf{A}_{i_j}, \mathbf{r}_{k-1} \right\rangle\right| \ \ldots \ \left|\left\langle \mathbf{A}_{i_{d-k+1}}, \mathbf{r}_{k-1} \right\rangle\right| \right]$$

- Include i_j into the set: $\mathbf{S}_k = \mathbf{S}_{k-1} \cup \{i_j\}$.
- Fully optimize on $\mathbf{S}_k$: $\mathbf{x}_k = \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ (Simple least squares).
- Update residual: $\mathbf{r}_k = \mathbf{y} - \mathbf{A}\mathbf{x}_k$.

Orthogonal Matching Pursuit for SLR

Result: OMP sparse estimate $\hat{\mathbf{x}}_s^{\text{OMP}} = \mathbf{x}_s$

$\mathbf{S}_0 = \phi$, $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{r}_0 = \mathbf{y}$

for $k = 1, 2, \ldots, s$ **do**

$j \leftarrow \arg\max_{i \notin \mathbf{S}_{k-1}} |\mathbf{A}_i^T \mathbf{r}_{k-1}|$ (Greedy selection)

$\mathbf{S}_k \leftarrow \mathbf{S}_{k-1} \cup \{j\}$

$\mathbf{x}_k \leftarrow \arg\min_{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$

$\mathbf{r}_k \leftarrow \mathbf{y} - \mathbf{A}\mathbf{x}_k$

end

Algorithm 1: Orthogonal Matching Pursuit (OMP) for SLR

Note that $\mathbf{A}_i^T \mathbf{r}_{k-1} \propto [\nabla f(\mathbf{x}_{k-1})]_i$ for $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$

Orthogonal Matching Pursuit for SLR

Result: OMP sparse estimate $\hat{\mathbf{x}}_s^{\text{OMP}} = \mathbf{x}_s$

$\mathbf{S}_0 = \phi$, $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{r}_0 = \mathbf{y}$

for $k = 1, 2, \ldots, s$ **do**

$\quad j \leftarrow \underset{i \notin \mathbf{S}_{k-1}}{\arg\max} |\mathbf{A}_i^T \mathbf{r}_{k-1}|$ (Greedy selection)

$\quad \mathbf{S}_k \leftarrow \mathbf{S}_{k-1} \cup \{j\}$

$\quad \mathbf{x}_k \leftarrow \underset{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k}{\arg\min} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$

$\quad \mathbf{r}_k \leftarrow \mathbf{y} - \mathbf{A}\mathbf{x}_k$

end

Algorithm 1: Orthogonal Matching Pursuit (OMP) for SLR

Note that $\mathbf{A}_i^T \mathbf{r}_{k-1} \propto [\nabla f(\mathbf{x}_{k-1})]_i$ for $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$

Orthogonal Matching Pursuit for general $f(\mathbf{x})$

Result: OMP sparse estimate $\hat{\mathbf{x}}_s^{\text{OMP}} = \mathbf{x}_s$
$\mathbf{S}_0 = \phi$, $\mathbf{x}_0 = \mathbf{0}$
for $k = 1, 2, \ldots, s$ **do**

$$j := \arg\max_{i \notin \mathbf{S}_{k-1}} |[\nabla f(\mathbf{x}_{k-1})]_i|$$

$$\mathbf{S}_k := \mathbf{S}_{k-1} \cup \{j\}$$

$$\mathbf{x}_k := \arg\min_{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k} f(\mathbf{x})$$

end

Algorithm 2: OMP for a general function $f(\mathbf{x})$

Key quantities

- Restricted Smoothness (ρ^+) & Restricted Strong Convexity (ρ^-)

$$\rho_s^- \|\mathbf{x} - \mathbf{z}\|_2^2 \leq \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z}\|_2^2 \leq \rho_s^+ \|\mathbf{x} - \mathbf{z}\|_2^2 \tag{4.1}$$

$\forall\, \mathbf{x}, \mathbf{z} \in \mathbb{R}^d$ s.t. $\|\mathbf{x} - \mathbf{z}\|_0 \leq s$.

- Restricted condition number ($\widetilde{\kappa}_s$)

$$\widetilde{\kappa}_s = \frac{\rho_1^+}{\rho_s^-} \tag{4.2}$$

- We also define

$$\kappa_s = \frac{\rho_s^+}{\rho_s^-} \geq \widetilde{\kappa}_s \tag{4.3}$$

Key quantities

- Restricted Smoothness (ρ^+) & Restricted Strong Convexity (ρ^-)

$$\rho_s^- \|\mathbf{x} - \mathbf{z}\|_2^2 \leq \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z}\|_2^2 \leq \rho_s^+ \|\mathbf{x} - \mathbf{z}\|_2^2 \tag{4.1}$$

$\forall\, \mathbf{x}, \mathbf{z} \in \mathbb{R}^d$ s.t. $\|\mathbf{x} - \mathbf{z}\|_0 \leq s$.

- Restricted condition number ($\widetilde{\kappa}_s$)

$$\widetilde{\kappa}_s = \frac{\rho_1^+}{\rho_s^-} \tag{4.2}$$

- We also define

$$\kappa_s = \frac{\rho_s^+}{\rho_s^-} \geq \widetilde{\kappa}_s \tag{4.3}$$

Key quantities

- Restricted Smoothness (ρ^+) & Restricted Strong Convexity (ρ^-)

$$\rho_s^- \|\mathbf{x} - \mathbf{z}\|_2^2 \leq \|\mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{z}\|_2^2 \leq \rho_s^+ \|\mathbf{x} - \mathbf{z}\|_2^2 \tag{4.1}$$

$\forall\, \mathbf{x}, \mathbf{z} \in \mathbb{R}^d$ s.t. $\|\mathbf{x} - \mathbf{z}\|_0 \leq s$.

- Restricted condition number ($\widetilde{\kappa}_s$)

$$\widetilde{\kappa}_s = \frac{\rho_1^+}{\rho_s^-} \tag{4.2}$$

- We also define

$$\kappa_s = \frac{\rho_s^+}{\rho_s^-} \geq \widetilde{\kappa}_s \tag{4.3}$$

Lower bounds for Fast rates

- If $\hat{\mathbf{x}}_{\ell_0}$ is the best ℓ_0 estimate in the set of s^*-sparse vectors then one can show

$$\sup_{\|\bar{\mathbf{x}}\|_0 \le s^*} \frac{1}{n}\mathbb{E}\left[\|\mathbf{A}(\hat{\mathbf{x}}_{\ell_0} - \bar{\mathbf{x}})\|_2^2\right] \lesssim \frac{\sigma^2 s^*}{n} \tag{4.4}$$

- Non-tractable since computing $\hat{\mathbf{x}}_{\ell_0}$ involves searching all $\binom{d}{s^*}$ subsets.
- (Y. Zhang, Wainwright & Jordan'15) $\exists\ \mathbf{A} \in \mathbb{R}^{n\times d}$ s.t. any poly-time algorithm satisfies

$$\sup_{\|\bar{\mathbf{x}}\|_0 \le s^*} \frac{1}{n}\mathbb{E}\left[\|\mathbf{A}(\hat{\mathbf{x}}_{\text{poly}} - \bar{\mathbf{x}})\|_2^2\right] \gtrsim \frac{\sigma^2 {s^*}^{1-\delta}\tilde{\kappa}_{s^*}}{n} \qquad \forall\ \delta > 0 \tag{4.5}$$

Consequence - Any estimator $\hat{\mathbf{x}}$ achieving fast rate must either **not be poly-time** or must return $\hat{\mathbf{x}}_{\text{poly}}$ that is **not** s^*-sparse.

$\lesssim$ and $\gtrsim$ are inequalities up to constant & poly-$\log d$ factors

Lower bounds for Fast rates

- If $\hat{\mathbf{x}}_{\ell_0}$ is the best ℓ_0 estimate in the set of s^*-sparse vectors then one can show

$$\sup_{\|\bar{\mathbf{x}}\|_0 \leq s^*} \frac{1}{n}\mathbb{E}\left[\|\mathbf{A}(\hat{\mathbf{x}}_{\ell_0} - \bar{\mathbf{x}})\|_2^2\right] \lesssim \frac{\sigma^2 s^*}{n} \tag{4.4}$$

- Non-tractable since computing $\hat{\mathbf{x}}_{\ell_0}$ involves searching all $\binom{d}{s^*}$ subsets.
- (Y. Zhang, Wainwright & Jordan'15) $\exists\ \mathbf{A} \in \mathbb{R}^{n\times d}$ s.t. any poly-time algorithm satisfies

$$\sup_{\|\bar{\mathbf{x}}\|_0 \leq s^*} \frac{1}{n}\mathbb{E}\left[\|\mathbf{A}(\hat{\mathbf{x}}_{\text{poly}} - \bar{\mathbf{x}})\|_2^2\right] \gtrsim \frac{\sigma^2 {s^*}^{1-\delta}\widetilde{\kappa}_{s^*}}{n} \qquad \forall\ \delta > 0 \tag{4.5}$$

Consequence - Any estimator $\hat{\mathbf{x}}$ achieving fast rate must either **not be poly-time** or must return $\hat{\mathbf{x}}_{\text{poly}}$ that is **not** s^*-sparse.

$\lesssim$ and $\gtrsim$ are inequalities up to constant & poly-$\log d$ factors

Lower bounds for Fast rates

- If $\hat{\mathbf{x}}_{\ell_0}$ is the best ℓ_0 estimate in the set of s^*-sparse vectors then one can show

$$\sup_{\|\bar{\mathbf{x}}\|_0 \leq s^*} \frac{1}{n} \mathbb{E}\left[\|\mathbf{A}(\hat{\mathbf{x}}_{\ell_0} - \bar{\mathbf{x}})\|_2^2\right] \lesssim \frac{\sigma^2 s^*}{n} \tag{4.4}$$

- Non-tractable since computing $\hat{\mathbf{x}}_{\ell_0}$ involves searching all $\binom{d}{s^*}$ subsets.
- (Y. Zhang, Wainwright & Jordan'15) $\exists\ \mathbf{A} \in \mathbb{R}^{n \times d}$ s.t. any poly-time algorithm satisfies

$$\sup_{\|\bar{\mathbf{x}}\|_0 \leq s^*} \frac{1}{n} \mathbb{E}\left[\left\|\mathbf{A}(\hat{\mathbf{x}}_{\text{poly}} - \bar{\mathbf{x}})\right\|_2^2\right] \gtrsim \frac{\sigma^2 {s^*}^{1-\delta} \tilde{\kappa}_{s^*}}{n} \qquad \forall\ \delta > 0 \tag{4.5}$$

Consequence - Any estimator $\hat{x}$ achieving fast rate must either **not be poly-time** or must return $\hat{x}_{\text{poly}}$ that is **not** s^*-sparse.

$\lesssim$ and $\gtrsim$ are inequalities up to constant & poly-$\log d$ factors

Lower bounds for Fast rates

- If $\hat{\mathbf{x}}_{\ell_0}$ is the best ℓ_0 estimate in the set of s^*-sparse vectors then one can show

$$\sup_{\|\bar{\mathbf{x}}\|_0 \leq s^*} \frac{1}{n} \mathbb{E}\left[\left\|\mathbf{A}(\hat{\mathbf{x}}_{\ell_0} - \bar{\mathbf{x}})\right\|_2^2\right] \lesssim \frac{\sigma^2 s^*}{n} \tag{4.4}$$

- Non-tractable since computing $\hat{\mathbf{x}}_{\ell_0}$ involves searching all $\binom{d}{s^*}$ subsets.
- (Y. Zhang, Wainwright & Jordan'15) $\exists\ \mathbf{A} \in \mathbb{R}^{n \times d}$ s.t. any poly-time algorithm satisfies

$$\sup_{\|\bar{\mathbf{x}}\|_0 \leq s^*} \frac{1}{n} \mathbb{E}\left[\left\|\mathbf{A}(\hat{\mathbf{x}}_{\text{poly}} - \bar{\mathbf{x}})\right\|_2^2\right] \gtrsim \frac{\sigma^2 {s^*}^{1-\delta} \widetilde{\kappa}_{s^*}}{n} \qquad \forall\ \delta > 0 \tag{4.5}$$

Consequence - Any estimator $\hat{\mathbf{x}}$ achieving fast rate must either **not be poly-time** or must return $\hat{\mathbf{x}}_{\text{poly}}$ that is **not** s^*-sparse.

$\lesssim$ and $\gtrsim$ are inequalities up to constant & poly-$\log d$ factors

Upper bounds on Generalization error

- Tightest known upper bounds for poly-time algorithms like IHT, OMP and Lasso were at least $\widetilde{\kappa}$ times worse than known lower bounds (Jain'14, T. Zhang'10, Y. Zhang'17).
- (T. Zhang'10) If $\hat{\mathrm{x}}_s$ be the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, then with high probability

$$\frac{1}{n}\|\mathbf{A}(\mathbf{x}_s - \bar{\mathrm{x}})\|_2^2 \lesssim \frac{1}{n}\sigma^2 s^* \widetilde{\kappa}^2_{s+s^*} \log \kappa_{s+s^*}. \tag{4.6}$$

- With slight modification to T. Zhang's analysis we get

Generalization error for OMP

If $\hat{\mathrm{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, then with high probability

$$\frac{1}{n}\|\mathbf{A}(\mathbf{x}_s - \bar{\mathrm{x}})\|_2^2 \lesssim \frac{1}{n}\sigma^2 s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}. \tag{4.7}$$

Matches the fast rate lower bound up to log factors.

Upper bounds on Generalization error

- Tightest known upper bounds for poly-time algorithms like IHT, OMP and Lasso were at least $\widetilde{\kappa}$ times worse than known lower bounds (Jain'14, T. Zhang'10, Y. Zhang'17).
- (T. Zhang'10) If $\hat{\mathbf{x}}_s$ be the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, then with high probability

$$\frac{1}{n}\|\mathbf{A}(\mathbf{x}_s - \bar{\mathbf{x}})\|_2^2 \lesssim \frac{1}{n}\sigma^2 s^* \widetilde{\kappa}^2_{s+s^*} \log \kappa_{s+s^*}. \tag{4.6}$$

- With slight modification to T. Zhang's analysis we get

Generalization error for OMP

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, then with high probability

$$\frac{1}{n}\|\mathbf{A}(\mathbf{x}_s - \bar{\mathbf{x}})\|_2^2 \lesssim \frac{1}{n}\sigma^2 s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}. \tag{4.7}$$

Matches the fast rate lower bound up to log factors.

Upper bounds on Generalization error

- Tightest known upper bounds for poly-time algorithms like IHT, OMP and Lasso were at least $\widetilde{\kappa}$ times worse than known lower bounds (Jain'14, T. Zhang'10, Y. Zhang'17).
- (T. Zhang'10) If $\hat{\mathbf{x}}_s$ be the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, then with high probability

$$\frac{1}{n}\|\mathbf{A}(\mathbf{x}_s - \bar{\mathbf{x}})\|_2^2 \lesssim \frac{1}{n}\sigma^2 s^* \widetilde{\kappa}^2_{s+s^*} \log \kappa_{s+s^*}. \quad (4.6)$$

- With slight modification to T. Zhang's analysis we get

Generalization error for OMP

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, then with high probability

$$\frac{1}{n}\|\mathbf{A}(\mathbf{x}_s - \bar{\mathbf{x}})\|_2^2 \lesssim \frac{1}{n}\sigma^2 s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}. \quad (4.7)$$

Matches the fast rate lower bound up to log factors.

Upper bounds on Generalization error

- Tightest known upper bounds for poly-time algorithms like IHT, OMP and Lasso were at least $\widetilde{\kappa}$ times worse than known lower bounds (Jain'14, T. Zhang'10, Y. Zhang'17).
- (T. Zhang'10) If $\hat{\mathbf{x}}_s$ be the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, then with high probability

$$\frac{1}{n} \|\mathbf{A}(\mathbf{x}_s - \bar{\mathbf{x}})\|_2^2 \lesssim \frac{1}{n} \sigma^2 s^* \widetilde{\kappa}^2_{s+s^*} \log \kappa_{s+s^*}. \tag{4.6}$$

- With slight modification to T. Zhang's analysis we get

Generalization error for OMP

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, then with high probability

$$\frac{1}{n} \|\mathbf{A}(\mathbf{x}_s - \bar{\mathbf{x}})\|_2^2 \lesssim \frac{1}{n} \sigma^2 s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}. \tag{4.7}$$

Matches the fast rate lower bound up to log factors.

Support Recovery upper bound

- Support recovery results are known for SCAD/MCP penalty based methods under bounded incoherence (Loh'14).
- For greedy algorithms like HTP and PHT, known support recovery results require poor dependence of $\widetilde{\kappa}$ on $|\bar{x}_{\min}|$ (Shen'17).
- If S is the support set of the s^{th} OMP iterate $\hat{\mathbf{x}}_s$, and if $\mathbf{S}^* \setminus \mathbf{S} \neq \phi$, then there is a **large additive decrease** in objective if $|\bar{x}_{\min}|$ is larger than the appropriate noise level.

Large decrease in objective

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations s.t. $\mathbf{S}^* \setminus \mathbf{S} \neq \phi$ and $|\bar{x}_{\min}| \gtrsim \frac{\sigma\gamma\sqrt{\rho_1^+}}{\rho_{s+s^*}^-}$ then with high probability

$$\|\mathbf{A}\hat{\mathbf{x}}_s - \mathbf{y}\|_2^2 - \|\mathbf{A}\hat{\mathbf{x}}_{s+1} - \mathbf{y}\|_2^2 \gtrsim \sigma^2 \tag{4.8}$$

where $\left\|\mathbf{A}_{\mathbf{S}^*\setminus\mathbf{S}}^T \mathbf{A}_\mathbf{S} \left(\mathbf{A}_\mathbf{S}^T \mathbf{A}_\mathbf{S}\right)^{-1}\right\|_\infty \leq \gamma$ and $\mathbf{S} = \text{supp}(\hat{\mathbf{x}}_s)$.

γ is similar to standard incoherence condition.

Support Recovery upper bound

- Support recovery results are known for SCAD/MCP penalty based methods under bounded incoherence (Loh'14).
- For greedy algorithms like HTP and PHT, known support recovery results require poor dependence of $\widetilde{\kappa}$ on $|\bar{x}_{\min}|$ (Shen'17).
- If S is the support set of the s^{th} OMP iterate $\hat{\mathbf{x}}_s$, and if $\mathrm{S}^* \setminus \mathrm{S} \neq \phi$, then there is a **large additive decrease** in objective if $|\bar{x}_{\min}|$ is larger than the appropriate noise level.

Large decrease in objective

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations s.t. $\mathrm{S}^* \setminus \mathrm{S} \neq \phi$ and $|\bar{x}_{\min}| \gtrsim \frac{\sigma\gamma\sqrt{\rho_1^+}}{\rho_{s+s^*}^-}$ then with high probability

$$\|\mathbf{A}\hat{\mathbf{x}}_s - \mathbf{y}\|_2^2 - \|\mathbf{A}\hat{\mathbf{x}}_{s+1} - \mathbf{y}\|_2^2 \gtrsim \sigma^2 \tag{4.8}$$

where $\left\|\mathbf{A}_{\mathrm{S}^*\setminus\mathrm{S}}^T \mathbf{A}_\mathrm{S} \left(\mathbf{A}_\mathrm{S}^T \mathbf{A}_\mathrm{S}\right)^{-1}\right\|_\infty \leq \gamma$ and $\mathrm{S} = \mathrm{supp}(\hat{\mathbf{x}}_s)$.

γ is similar to standard incoherence condition.

Support Recovery upper bound

- Support recovery results are known for SCAD/MCP penalty based methods under bounded incoherence (Loh'14).
- For greedy algorithms like HTP and PHT, known support recovery results require poor dependence of $\widetilde{\kappa}$ on $|\bar{x}_{\min}|$ (Shen'17).
- If S is the support set of the s^{th} OMP iterate $\hat{\mathbf{x}}_s$, and if $\mathrm{S}^* \setminus \mathrm{S} \neq \phi$, then there is a **large additive decrease** in objective if $|\bar{x}_{\min}|$ is larger than the appropriate noise level.

Large decrease in objective

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations s.t. $\mathrm{S}^* \setminus \mathrm{S} \neq \phi$ and $|\bar{x}_{\min}| \gtrsim \frac{\sigma\gamma\sqrt{\rho_1^+}}{\rho_{s+s^*}^-}$ then with high probability

$$\|\mathbf{A}\hat{\mathbf{x}}_s - \mathbf{y}\|_2^2 - \|\mathbf{A}\hat{\mathbf{x}}_{s+1} - \mathbf{y}\|_2^2 \gtrsim \sigma^2 \tag{4.8}$$

where $\left\|\mathbf{A}_{\mathrm{S}^*\setminus\mathrm{S}}^T \mathbf{A}_\mathrm{S} \left(\mathbf{A}_\mathrm{S}^T \mathbf{A}_\mathrm{S}\right)^{-1}\right\|_\infty \leq \gamma$ and $\mathrm{S} = \mathrm{supp}(\hat{\mathbf{x}}_s)$.

γ is similar to standard incoherence condition.

Support Recovery upper bound

- Support recovery results are known for SCAD/MCP penalty based methods under bounded incoherence (Loh'14).
- For greedy algorithms like HTP and PHT, known support recovery results require poor dependence of $\widetilde{\kappa}$ on $|\bar{x}_{\min}|$ (Shen'17).
- If $\mathbf{S}$ is the support set of the s^{th} OMP iterate $\hat{\mathbf{x}}_s$, and if $\mathbf{S}^* \setminus \mathbf{S} \neq \phi$, then there is a **large additive decrease** in objective if $|\bar{x}_{\min}|$ is larger than the appropriate noise level.

Large decrease in objective

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations s.t. $\mathbf{S}^* \setminus \mathbf{S} \neq \phi$ and $|\bar{x}_{\min}| \gtrsim \frac{\sigma\gamma\sqrt{\rho_1^+}}{\rho_{s+s^*}^-}$ then with high probability

$$\|\mathbf{A}\hat{\mathbf{x}}_s - \mathbf{y}\|_2^2 - \|\mathbf{A}\hat{\mathbf{x}}_{s+1} - \mathbf{y}\|_2^2 \gtrsim \sigma^2 \tag{4.8}$$

where $\left\|\mathbf{A}_{\mathbf{S}^*\setminus\mathbf{S}}^T \mathbf{A}_{\mathbf{S}} \left(\mathbf{A}_{\mathbf{S}}^T \mathbf{A}_{\mathbf{S}}\right)^{-1}\right\|_\infty \leq \gamma$ and $\mathbf{S} = \text{supp}(\hat{\mathbf{x}}_s)$.

γ is similar to standard incoherence condition.

Support Recovery upper bound

- Support recovery results are known for SCAD/MCP penalty based methods under bounded incoherence (Loh'14).
- For greedy algorithms like HTP and PHT, known support recovery results require poor dependence of $\widetilde{\kappa}$ on $|\bar{x}_{\min}|$ (Shen'17).
- If $\mathbf{S}$ is the support set of the s^{th} OMP iterate $\hat{\mathbf{x}}_s$, and if $\mathbf{S}^* \setminus \mathbf{S} \neq \phi$, then there is a **large additive decrease** in objective if $|\bar{x}_{\min}|$ is larger than the appropriate noise level.

Large decrease in objective

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations s.t. $\mathbf{S}^* \setminus \mathbf{S} \neq \phi$ and $|\bar{x}_{\min}| \gtrsim \frac{\sigma\gamma\sqrt{\rho_1^+}}{\rho_{s+s^*}^-}$ then with high probability

$$\|\mathbf{A}\hat{\mathbf{x}}_s - \mathbf{y}\|_2^2 - \|\mathbf{A}\hat{\mathbf{x}}_{s+1} - \mathbf{y}\|_2^2 \gtrsim \sigma^2 \tag{4.8}$$

where $\left\|\mathbf{A}_{\mathbf{S}^*\setminus\mathbf{S}}^T \mathbf{A}_{\mathbf{S}} \left(\mathbf{A}_{\mathbf{S}}^T \mathbf{A}_{\mathbf{S}}\right)^{-1}\right\|_\infty \leq \gamma$ and $\mathbf{S} = \text{supp}(\hat{\mathbf{x}}_s)$.

γ is similar to standard incoherence condition.

Support Recovery upper bound

Since $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \geq 0 \; \forall \; \mathbf{x} \in \mathbb{R}^d$, the number of extra iterations cannot be too large.

Support recovery and infinity norm bound

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, s.t. $|\bar{x}_{\min}| \gtrsim \frac{\sigma\gamma\sqrt{\rho_1^+}}{\rho_{s+s^*}^-}$, then with high probability

1. $\mathrm{S}^* \subseteq \text{supp}(\hat{\mathbf{x}}_s)$
2. $\|\hat{\mathbf{x}}_s - \bar{\mathbf{x}}\|_\infty \lesssim \sigma\sqrt{\frac{\log s}{\rho_s^-}}$

where $\left\|\mathbf{A}_{\mathrm{S}^*\setminus\mathrm{S}}^T \mathbf{A}_{\mathrm{S}} \left(\mathbf{A}_{\mathrm{S}}^T \mathbf{A}_{\mathrm{S}}\right)^{-1}\right\|_\infty \leq \gamma$ and $\mathrm{S} = \text{supp}(\hat{\mathbf{x}}_s)$.

- Condition on $|\bar{x}_{\min}|$ scales as $\frac{1}{\sqrt{n}}$ since both $\rho_{s+s^*}^-$ and ρ_1^+ have a factor of n. Also it is better by at-least $\sqrt{\bar{\kappa}}$ than that in other recent works.
- γ is allowed to be very large.

Support Recovery upper bound

Since $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \geq 0 \; \forall \; \mathbf{x} \in \mathbb{R}^d$, the number of extra iterations cannot be too large.

Support recovery and infinity norm bound

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, s.t. $|\bar{x}_{\min}| \gtrsim \frac{\sigma\gamma\sqrt{\rho_1^+}}{\rho^-_{s+s^*}}$, then with high probability

1. $\mathbf{S}^* \subseteq \text{supp}(\hat{\mathbf{x}}_s)$
2. $\|\hat{\mathbf{x}}_s - \bar{\mathbf{x}}\|_\infty \lesssim \sigma \sqrt{\frac{\log s}{\rho_s^-}}$

where $\left\| \mathbf{A}^T_{\mathbf{S}^* \setminus \mathbf{S}} \mathbf{A}_{\mathbf{S}} \left(\mathbf{A}^T_{\mathbf{S}} \mathbf{A}_{\mathbf{S}} \right)^{-1} \right\|_\infty \leq \gamma$ and $\mathbf{S} = \text{supp}(\hat{\mathbf{x}}_s)$.

- Condition on $|\bar{x}_{\min}|$ scales as $\frac{1}{\sqrt{n}}$ since both $\rho^-_{s+s^*}$ and ρ_1^+ have a factor of n. Also it is better by at-least $\sqrt{\widetilde{\kappa}}$ than that in other recent works.
- γ is allowed to be very large.

Support Recovery upper bound

Since $\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 \geq 0 \ \forall\ \mathbf{x} \in \mathbb{R}^d$, the number of extra iterations cannot be too large.

Support recovery and infinity norm bound

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, s.t. $|\bar{x}_{\min}| \gtrsim \frac{\sigma\gamma\sqrt{\rho_1^+}}{\rho_{s+s^*}^-}$, then with high probability

1. $\mathbf{S}^* \subseteq \text{supp}(\hat{\mathbf{x}}_s)$
2. $\|\hat{\mathbf{x}}_s - \bar{\mathbf{x}}\|_\infty \lesssim \sigma\sqrt{\frac{\log s}{\rho_s^-}}$

where $\left\|\mathbf{A}_{\mathbf{S}^*\setminus\mathbf{S}}^T \mathbf{A}_\mathbf{S} \left(\mathbf{A}_\mathbf{S}^T \mathbf{A}_\mathbf{S}\right)^{-1}\right\|_\infty \leq \gamma$ and $\mathbf{S} = \text{supp}(\hat{\mathbf{x}}_s)$.

- Condition on $|\bar{x}_{\min}|$ scales as $\frac{1}{\sqrt{n}}$ since both $\rho_{s+s^*}^-$ and ρ_1^+ have a factor of n. Also it is better by at-least $\sqrt{\widetilde{\kappa}}$ than that in other recent works.
- γ is allowed to be very large.

Support Recovery upper bound

Since $\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2 \geq 0 \; \forall \; \mathbf{x} \in \mathbb{R}^d$, the number of extra iterations cannot be too large.

Support recovery and infinity norm bound

If $\hat{\mathbf{x}}_s$ is the output of OMP after $s \gtrsim s^* \widetilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, s.t. $|\bar{x}_{\min}| \gtrsim \frac{\sigma\gamma\sqrt{\rho_1^+}}{\rho_{s+s^*}^-}$, then with high probability

1. $\mathbf{S}^* \subseteq \text{supp}(\hat{\mathbf{x}}_s)$
2. $\|\hat{\mathbf{x}}_s - \bar{\mathbf{x}}\|_\infty \lesssim \sigma\sqrt{\frac{\log s}{\rho_s^-}}$

where $\left\|\mathbf{A}_{\mathbf{S}^*\setminus\mathbf{S}}^T \mathbf{A}_{\mathbf{S}} \left(\mathbf{A}_{\mathbf{S}}^T \mathbf{A}_{\mathbf{S}}\right)^{-1}\right\|_\infty \leq \gamma$ and $\mathbf{S} = \text{supp}(\hat{\mathbf{x}}_s)$.

- Condition on $|\bar{x}_{\min}|$ scales as $\frac{1}{\sqrt{n}}$ since both $\rho_{s+s^*}^-$ and ρ_1^+ have a factor of n. Also it is better by at-least $\sqrt{\widetilde{\kappa}}$ than that in other recent works.
- γ is allowed to be very large.

Lower bound instance construction

- (Y. Zhang'15)'s lower bounds were for algorithms that output s^*-sparse solutions which does not apply for OMP when it is run for more than s^* iterations.
- We provide matching lower bounds for Support recovery as well as generalization error for OMP.
- Fool OMP into picking incorrect indexes. Large support size $\implies$ large generalization error.
- Construct an evenly distributed $\bar{\mathrm{x}}$

$$\bar{x}_i = \begin{cases} \frac{1}{\sqrt{s^*}} & \text{if } 1 \leq i \leq s^* \\ 0 & \text{if } i > s^* \end{cases} \qquad \implies \text{supp}(\bar{\mathrm{x}}) = \{1, 2, \ldots, s^*\}$$

- Construct $\mathrm{M}^{(\epsilon)} \in \mathbb{R}^{n \times d}$ parameterized by ϵ
  - $\mathrm{M}^{(\epsilon)}_{1:s^*}$ are random s^* orthogonal column vectors s.t. $\left\|\mathrm{M}^{(\epsilon)}_i\right\|_2^2 = n \; \forall \; i \in [s^*]$.
  - $\mathrm{M}^{(\epsilon)}_i = \sqrt{1-\epsilon}\left[\frac{1}{\sqrt{s^*}} \sum_{j=1}^{s^*} \mathrm{M}^{(\epsilon)}_j\right] + \sqrt{\epsilon}\mathrm{g}_i \; \forall \; i \notin [s^*]$ where g_i's are orthogonal to each other and $\mathrm{M}^{(\epsilon)}_{1:s^*}$ with $\|\mathrm{g}_i\|_2^2 = n$.

Lower bound instance construction

- (Y. Zhang'15)'s lower bounds were for algorithms that output s^*-sparse solutions which does not apply for OMP when it is run for more than s^* iterations.
- We provide matching lower bounds for Support recovery as well as generalization error for OMP.
- Fool OMP into picking incorrect indexes. Large support size $\implies$ large generalization error.
- Construct an evenly distributed $\bar{\mathrm{x}}$

$$\bar{x}_i = \begin{cases} \frac{1}{\sqrt{s^*}} & \text{if } 1 \le i \le s^* \\ 0 & \text{if } i > s^* \end{cases} \qquad \implies \text{supp}(\bar{\mathrm{x}}) = \{1, 2, \ldots, s^*\}$$

- Construct $\mathrm{M}^{(\epsilon)} \in \mathbb{R}^{n \times d}$ parameterized by ϵ
  - $\mathrm{M}^{(\epsilon)}_{1:s^*}$ are random s^* orthogonal column vectors s.t. $\left\|\mathrm{M}^{(\epsilon)}_i\right\|_2^2 = n \; \forall \, i \in [s^*]$.
  - $\mathrm{M}^{(\epsilon)}_i = \sqrt{1-\epsilon}\left[\frac{1}{\sqrt{s^*}} \sum_{j=1}^{s^*} \mathrm{M}^{(\epsilon)}_j\right] + \sqrt{\epsilon}\mathrm{g}_i \; \forall \, i \notin [s^*]$ where g_i's are orthogonal to each other and $\mathrm{M}^{(\epsilon)}_{1:s^*}$ with $\|\mathrm{g}_i\|_2^2 = n$.

Lower bound instance construction

- (Y. Zhang'15)'s lower bounds were for algorithms that output s^*-sparse solutions which does not apply for OMP when it is run for more than s^* iterations.
- We provide matching lower bounds for Support recovery as well as generalization error for OMP.
- Fool OMP into picking incorrect indexes. Large support size $\implies$ large generalization error.
- Construct an evenly distributed $\bar{\mathrm{x}}$

$$\bar{x}_i = \begin{cases} \frac{1}{\sqrt{s^*}} & \text{if } 1 \leq i \leq s^* \\ 0 & \text{if } i > s^* \end{cases} \qquad \implies \text{supp}(\bar{\mathrm{x}}) = \{1, 2, \ldots, s^*\}$$

- Construct $\mathrm{M}^{(\epsilon)} \in \mathbb{R}^{n \times d}$ parameterized by ϵ
  - $\mathrm{M}^{(\epsilon)}_{1:s^*}$ are random s^* orthogonal column vectors s.t. $\left\|\mathrm{M}^{(\epsilon)}_i\right\|_2^2 = n \ \forall\, i \in [s^*]$.
  - $\mathrm{M}^{(\epsilon)}_i = \sqrt{1-\epsilon}\left[\frac{1}{\sqrt{s^*}} \sum_{j=1}^{s^*} \mathrm{M}^{(\epsilon)}_j\right] + \sqrt{\epsilon}\mathrm{g}_i \ \forall\, i \notin [s^*]$ where g_i's are orthogonal to each other and $\mathrm{M}^{(\epsilon)}_{1:s^*}$ with $\|\mathrm{g}_i\|_2^2 = n$.

Lower bound instance construction

- (Y. Zhang'15)'s lower bounds were for algorithms that output s^*-sparse solutions which does not apply for OMP when it is run for more than s^* iterations.
- We provide matching lower bounds for Support recovery as well as generalization error for OMP.
- Fool OMP into picking incorrect indexes. Large support size $\implies$ large generalization error.
- Construct an evenly distributed $\bar{\mathbf{x}}$

$$\bar{x}_i = \begin{cases} \frac{1}{\sqrt{s^*}} & \text{if } 1 \leq i \leq s^* \\ 0 & \text{if } i > s^* \end{cases} \qquad \implies \text{supp}(\bar{\mathbf{x}}) = \{1, 2, \ldots, s^*\}$$

- Construct $\mathbf{M}^{(\epsilon)} \in \mathbb{R}^{n \times d}$ parameterized by ϵ
  - $\mathbf{M}^{(\epsilon)}_{1:s^*}$ are random s^* orthogonal column vectors s.t. $\left\|\mathbf{M}^{(\epsilon)}_i\right\|_2^2 = n \; \forall \, i \in [s^*]$.
  - $\mathbf{M}^{(\epsilon)}_i = \sqrt{1-\epsilon}\left[\frac{1}{\sqrt{s^*}}\sum_{j=1}^{s^*} \mathbf{M}^{(\epsilon)}_j\right] + \sqrt{\epsilon}\mathbf{g}_i \; \forall \, i \notin [s^*]$ where $\mathbf{g}_i$'s are orthogonal to each other and $\mathbf{M}^{(\epsilon)}_{1:s^*}$ with $\|\mathbf{g}_i\|_2^2 = n$.

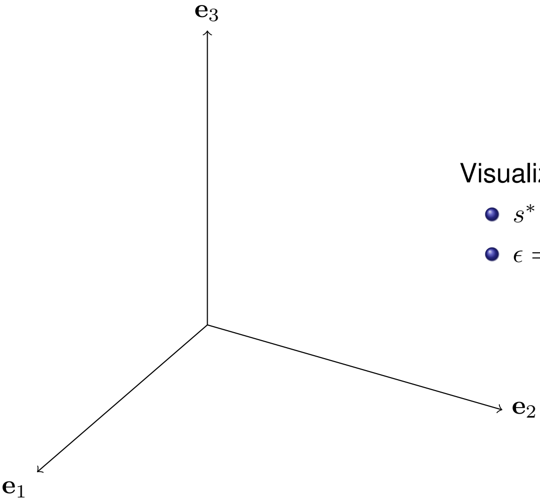
Lower bound instance construction

- (Y. Zhang'15)'s lower bounds were for algorithms that output s^*-sparse solutions which does not apply for OMP when it is run for more than s^* iterations.
- We provide matching lower bounds for Support recovery as well as generalization error for OMP.
- Fool OMP into picking incorrect indexes. Large support size $\implies$ large generalization error.
- Construct an evenly distributed $\bar{\mathbf{x}}$

$$\bar{x}_i = \begin{cases} \frac{1}{\sqrt{s^*}} & \text{if } 1 \leq i \leq s^* \\ 0 & \text{if } i > s^* \end{cases} \qquad \implies \text{supp}(\bar{\mathbf{x}}) = \{1, 2, \ldots, s^*\}$$

- Construct $\mathbf{M}^{(\epsilon)} \in \mathbb{R}^{n \times d}$ parameterized by ϵ
  - $\mathbf{M}^{(\epsilon)}_{1:s^*}$ are random s^* orthogonal column vectors s.t. $\left\|\mathbf{M}^{(\epsilon)}_i\right\|_2^2 = n \; \forall \, i \in [s^*]$.
  - $\mathbf{M}^{(\epsilon)}_i = \sqrt{1-\epsilon}\left[\frac{1}{\sqrt{s^*}} \sum_{j=1}^{s^*} \mathbf{M}^{(\epsilon)}_j\right] + \sqrt{\epsilon}\mathbf{g}_i \; \forall \, i \notin [s^*]$ where $\mathbf{g}_i$'s are orthogonal to each other and $\mathbf{M}^{(\epsilon)}_{1:s^*}$ with $\|\mathbf{g}_i\|_2^2 = n$.

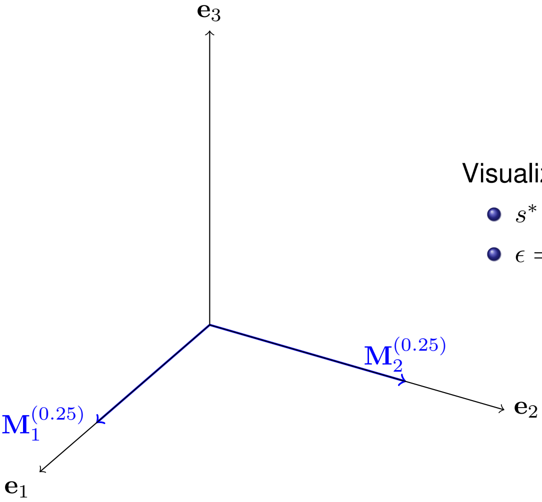
Lower bound instance construction



Visualizing in $d = 3$ with

- $s^* = 2$
- $\epsilon = 0.25$

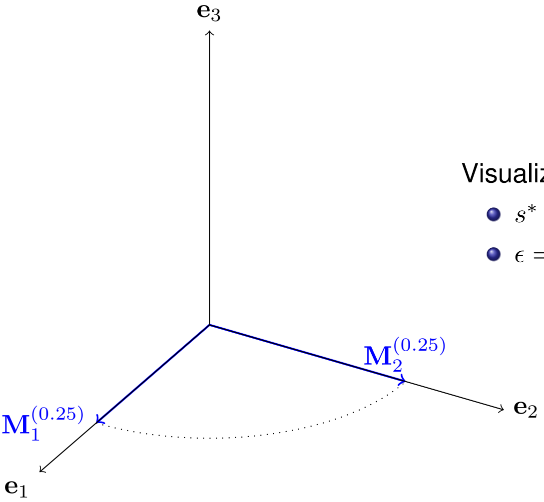
Lower bound instance construction



Visualizing in $d = 3$ with

- $s^* = 2$
- $\epsilon = 0.25$

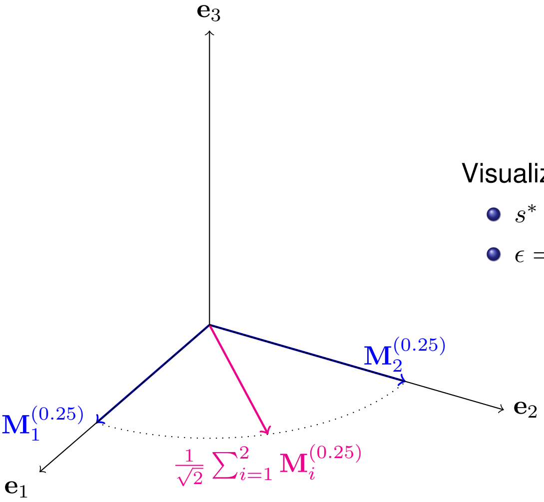
Lower bound instance construction



Visualizing in $d = 3$ with

- $s^* = 2$
- $\epsilon = 0.25$

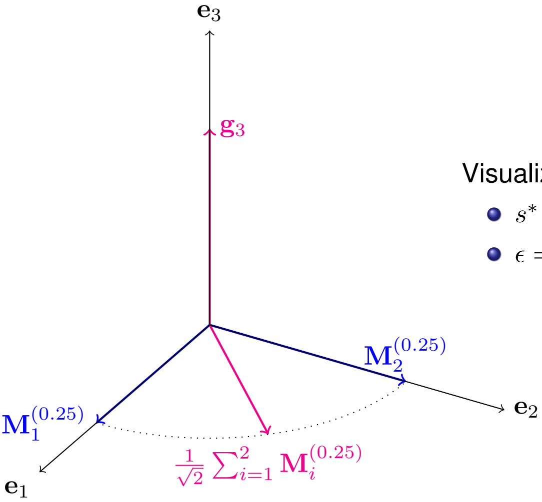
Lower bound instance construction



Visualizing in $d = 3$ with

- $s^* = 2$
- $\epsilon = 0.25$

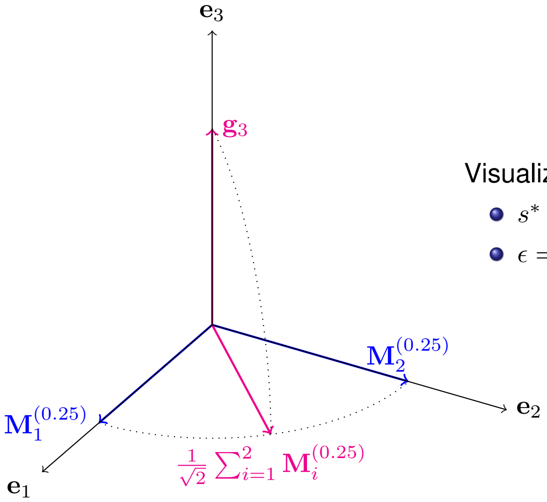
Lower bound instance construction



Visualizing in $d = 3$ with

- $s^* = 2$
- $\epsilon = 0.25$

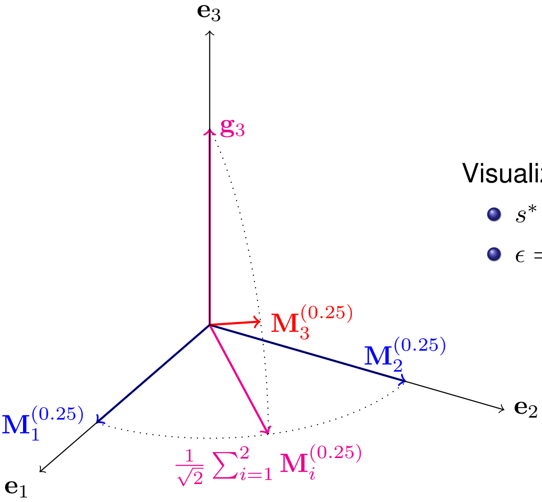
Lower bound instance construction



Visualizing in $d = 3$ with

- $s^* = 2$
- $\epsilon = 0.25$

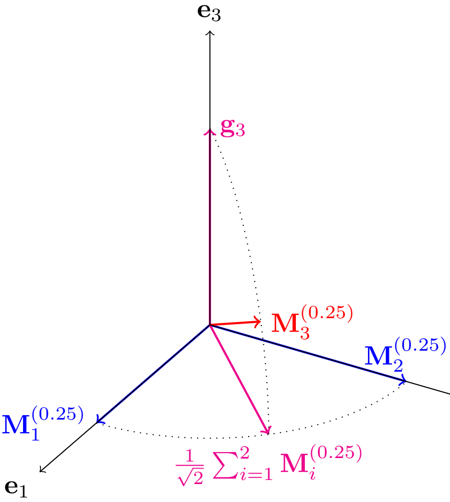
Lower bound instance construction



Visualizing in $d = 3$ with

- $s^* = 2$
- $\epsilon = 0.25$

Lower bound instance construction



Visualizing in $d = 3$ with

- $s^* = 2$
- $\epsilon = 0.25$

Smaller $\epsilon \implies$ more correlation.
$\therefore \epsilon$ is a proxy for $\widetilde{\kappa}_s$.

Lower bounds

Noiseless case

For $s^* \leq d \leq n$, $\exists\ \epsilon > 0$ s.t. when OMP is executed on SLR problem with $y = \mathbf{M}^{(\epsilon)}\bar{\mathbf{x}}$ for $s \leq d - s^*$ iterations

- $\widetilde{\kappa}_s(\mathbf{M}^{(\epsilon)}) \lesssim \frac{s}{s^*}$ and $\gamma \leq \sqrt{\frac{3}{2}}$
- $\mathbf{S}^* \cap \text{supp}(\hat{\mathbf{x}}_s) = \phi$

Noisy case

For $s^* \leq s \leq d^{1-\alpha}$ where $\alpha \in (0, 1)$, $\exists\ \epsilon > 0$ s.t. when OMP is executed on SLR problem with $y = \mathbf{M}^{(\epsilon)}\bar{\mathbf{x}} + \boldsymbol{\eta}$ where $\boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}_{n\times n}\right)$, then

- $\widetilde{\kappa}_s(\mathbf{M}^{(\epsilon)}) \lesssim \frac{s}{s^*}$ and $\gamma \leq \frac{1}{2}$
- with high probability $\frac{1}{n}\left\|\mathbf{A}\hat{\mathbf{x}}_s - \mathbf{A}\bar{\mathbf{x}}\right\|_2^2 \gtrsim \frac{1}{n}\sigma^2\widetilde{\kappa}_{s+s^*}s^*$
- with high probability $\mathbf{S}^* \cap \text{supp}(\hat{\mathbf{x}}_s) = \phi$

$\implies s \gtrsim \widetilde{\kappa}_s s^*$ iterations are indeed necessary.
Addition of noise can only help.

Lower bounds

Noiseless case

For $s^* \leq d \leq n$, $\exists\, \epsilon > 0$ s.t. when OMP is executed on SLR problem with $y = \mathbf{M}^{(\epsilon)}\bar{\mathbf{x}}$ for $s \leq d - s^*$ iterations

- $\widetilde{\kappa}_s(\mathbf{M}^{(\epsilon)}) \lesssim \frac{s}{s^*}$ and $\gamma \leq \sqrt{\frac{3}{2}}$
- $\mathbf{S}^* \cap \mathrm{supp}(\hat{\mathbf{x}}_s) = \phi$

Noisy case

For $s^* \leq s \leq d^{1-\alpha}$ where $\alpha \in (0,1)$, $\exists\, \epsilon > 0$ s.t. when OMP is executed on SLR problem with $y = \mathbf{M}^{(\epsilon)}\bar{\mathbf{x}} + \boldsymbol{\eta}$ where $\boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_{n\times n}\right)$, then

- $\widetilde{\kappa}_s(\mathbf{M}^{(\epsilon)}) \lesssim \frac{s}{s^*}$ and $\gamma \leq \frac{1}{2}$
- with high probability $\frac{1}{n}\|\mathbf{A}\hat{\mathbf{x}}_s - \mathbf{A}\bar{\mathbf{x}}\|_2^2 \gtrsim \frac{1}{n}\sigma^2\widetilde{\kappa}_{s+s^*}s^*$
- with high probability $\mathbf{S}^* \cap \mathrm{supp}(\hat{\mathbf{x}}_s) = \phi$

$\implies s \gtrsim \widetilde{\kappa}_s s^*$ iterations are indeed necessary.
Addition of noise can only help.

Lower bounds

Noiseless case

For $s^* \leq d \leq n$, $\exists\, \epsilon > 0$ s.t. when OMP is executed on SLR problem with $y = \mathbf{M}^{(\epsilon)}\bar{\mathbf{x}}$ for $s \leq d - s^*$ iterations

- $\widetilde{\kappa}_s(\mathbf{M}^{(\epsilon)}) \lesssim \frac{s}{s^*}$ and $\gamma \leq \sqrt{\frac{3}{2}}$
- $\mathbf{S}^* \cap \mathrm{supp}(\hat{\mathbf{x}}_s) = \phi$

Noisy case

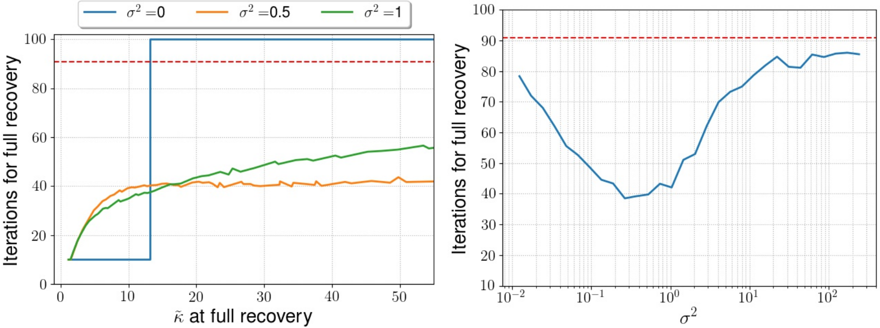
For $s^* \leq s \leq d^{1-\alpha}$ where $\alpha \in (0,1)$, $\exists\, \epsilon > 0$ s.t. when OMP is executed on SLR problem with $y = \mathbf{M}^{(\epsilon)}\bar{\mathbf{x}} + \boldsymbol{\eta}$ where $\boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}_{n\times n}\right)$, then

- $\widetilde{\kappa}_s(\mathbf{M}^{(\epsilon)}) \lesssim \frac{s}{s^*}$ and $\gamma \leq \frac{1}{2}$
- with high probability $\frac{1}{n}\|\mathbf{A}\hat{\mathbf{x}}_s - \mathbf{A}\bar{\mathbf{x}}\|_2^2 \gtrsim \frac{1}{n}\sigma^2\widetilde{\kappa}_{s+s^*}s^*$
- with high probability $\mathbf{S}^* \cap \mathrm{supp}(\hat{\mathbf{x}}_s) = \phi$

$\implies s \gtrsim \widetilde{\kappa}_s s^*$ iterations are indeed necessary.
Addition of noise can only help.

Simulations

We perform simulations on the lower bound instance class.
$\mathbf{M}^{(\epsilon)} \in \mathbb{R}^{1000\times 100}$ and $s^* = 10$.



(a) Varying condition number

(b) Varying noise variance

Figure: Number of iterations required for recovering the full support of $\bar{\mathbf{x}}$ with respect to the restricted condition number ($\widetilde{\kappa}_{s+s^*}$) of design matrix and the variance of noise (σ^2).