

Support Recovery for Orthogonal Matching Pursuit: Upper and Lower bounds

Raghav Somani, Chirag Gupta, Prateek Jain and Praneeth Netrapalli

Microsoft Research, India and Carnegie Mellon University



Abstract

We study **Sparse Linear Regression (SLR)** under the Restricted Strong Convexity (RSC) assumption for Orthogonal Matching Pursuit (OMP).

Main Results:

- Improved bound for Support Recovery
- Improved Generalization Error bound
- Matching (up to log factors) lower bound for Generalization Error

Introduction

• Sparse Linear Regression (SLR):

$$\bar{\mathbf{x}} = \underset{\|\mathbf{x}\|_{0} < s^{*}}{\operatorname{arg \, min}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_{2}^{2} \qquad \mathbf{A} \in \mathbb{R}^{n \times d}, \mathbf{y} \in \mathbb{R}^{n}.$$

- Applications: Resource Constrained ML, High-dimensional Statistics, Image De-noising, etc.
- NP hard in general [1]. We study it under RSC:

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

- ρ_s^+ : Restricted Smoothness (RSS) constant
- Key quantity: Restricted Condition Number $-\tilde{\kappa}_s = \frac{\rho_1^+}{\rho^-}$
- Assumption: Generative model. $supp(\bar{\mathbf{x}}) = \mathbf{S}^*, |\mathbf{S}^*| = s^*.$
- $\mathbf{y} = \mathbf{A} \bar{\mathbf{x}} + oldsymbol{\eta} = \mathbf{A}_{\mathbf{S}^*} \bar{\mathbf{x}}_{\mathbf{S}^*} + oldsymbol{\eta} \qquad ext{where } oldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_{n imes n}\right).$

Goals of SLR

- **Bound Generalization Error**: $G(\mathbf{x}) := \frac{1}{n} \|\mathbf{A}(\bar{\mathbf{x}} \mathbf{x})\|_2^2$.
- **2 Support Recovery** Recover a small support set S s.t. $S \supseteq S^*$.

Existing SLR algorithms

- ℓ_1 minimization based (LASSO based) algorithms, for example: Dantzig selector.
- Non-convex penalty based,
- SCAD/MCP penalty based Iterative Hard Thresholding (IHT)

The size of the support of OMP = number of iterations.

- Greedy and Pursuit methods,
- Hard Thresholding Pursuit (HTP)
 Orthogonal Matching Pursuit (OMP)

Orthogonal Matching Pursuit

OMP – greedy algorithm. Estimates support of $\bar{\mathbf{x}}$ by adding one feature at a time. Start with empty set: $\mathbf{S}_0 = \phi$, $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{r}_0 = \mathbf{y}$. For iterations k = 1 to s• Select next index greedily: $j \leftarrow \underset{i \notin \mathbf{S}_{k-1}}{\arg \max} |\mathbf{A}_i^T \mathbf{r}_{k-1}|$,
• Incrementally grow support: $\mathbf{S}_k \leftarrow \mathbf{S}_{k-1} \cup \{j\}$,
• Optimize on current support: $\mathbf{x}_k \leftarrow \underset{\sup (\mathbf{x}) \subseteq \mathbf{S}_k}{\arg \min} ||\mathbf{A}\mathbf{x} - \mathbf{y}||_2^2$,
• Update residual: $\mathbf{r}_k \leftarrow \mathbf{y} - \mathbf{A}\mathbf{x}_k$,
Return \mathbf{x}_s .

Upper bound on Generalization Error

- (T. Zhang'10 [2]) If $\hat{\mathbf{x}}_s$ is OMP's output after $s \gtrsim s^* \tilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, then w.h.p., $\frac{1}{\pi} \|\mathbf{A}(\hat{\mathbf{x}}_s \bar{\mathbf{x}})\|_2^2 \lesssim \frac{1}{\pi} \sigma^2 s^* \tilde{\kappa}_{s+s^*}^2 \log \kappa_{s+s^*}$
- We adapt the analysis to improve the upper bound

Generalization error upper bound

After $s \gtrsim s^* \tilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations, w.h.p.,

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

Support Recovery upper bound

• If any support is unrecovered, there is a large additive decrease in the function value.

Large decrease in objective

After $s \gtrsim s^* \tilde{\kappa}_{s+s^*} \log \kappa_{s+s^*}$ iterations if $\mathbf{S}^* \setminus \operatorname{supp}(\mathbf{\hat{x}}_s) \neq \phi$, and $|\bar{x}_{\min}| \gtrsim \gamma \frac{\sigma \sqrt{\rho_1^+}}{\rho_{s+s^*}^-}$, then w.h.p.,

$$\|\mathbf{A}\hat{\mathbf{x}}_{s} - \mathbf{y}\|_{2}^{2} - \|\mathbf{A}\hat{\mathbf{x}}_{s+1} - \mathbf{y}\|_{2}^{2} \gtrsim \sigma^{2}$$
where $\|\mathbf{A}_{\mathbf{S}^{*}\backslash\mathbf{S}}^{T}\mathbf{A}_{\mathbf{S}}(\mathbf{A}_{\mathbf{S}}^{T}\mathbf{A}_{\mathbf{S}})^{-1}\|_{\infty} \leq \gamma$.

• Objective function can be lower bounded : extra iterations cannot be too large.

Support Recovery upper bound

If $\hat{\mathbf{x}}_s$ is OMP's output then under similar conditions, w.h.p.,

$$\mathbf{S}^* \subseteq \operatorname{supp}(\mathbf{\hat{x}}_s)$$
 and $\|\mathbf{\hat{x}}_s - \mathbf{\bar{x}}\|_{\infty} \lesssim \sigma_{\sqrt{\frac{\log s}{\rho_s^-}}}$.

Remarks and Comparison

- Our assumption on $|\bar{x}_{\min}|$ is better by at least $\sqrt{\tilde{\kappa}}$ than ones in recent works.
- The γ parameter is similar to the standard incoherence parameter. Existing results for OMP require the incoherence parameter to be < 1. Our analysis holds for arbitrary values of γ .
- The only known support recovery result is for LASSO under RSC, that provides similar result [3].

Related Work	Support expansion (s)	$ \bar{x}_{\min} $ lower bound
Yuan et al. [HTP]	$\kappa_{2s}^2 s^*$	$\frac{\sigma\sqrt{s}}{\sqrt{ ho_{2s}^{-}}}$
Shen et al. [HTP]	$\kappa_{2s}^2 s^*$	$\frac{\sigma\sqrt{\kappa_{2s}}\sqrt{\rho_1^+s}}{\rho_{s+s^*}^-}$
Shen et al. $[PHT(r)]$	$s^* + \kappa_{2s}^2 \min\left\{s^*, r\right\}$	$\frac{\sigma\sqrt{\kappa_{2s}}\sqrt{\rho_1^+s}}{\rho_{2s}^-}$
Jain et al. [IHT]	$\kappa_{2s+s^*}^2s^*$	
Zhang [OMP]	$\tilde{\kappa}_{s+s^*} s^* \log \kappa_{s+s^*}$	
Our's [OMP]	$\tilde{\kappa}_{s+s^*} s^* \log \kappa_{s+s^*}$	$\gamma \cdot rac{\sigma \sqrt{ ho_1^+}}{ ho_{s+s^*}^-}$

- X. Yuan, P. Li, and T. Zhang. Exact recovery of hard thresholding pursuit, NIPS'16.
- J. Shen and P. Li. On the iteration complexity of support recovery via hard thresholding pursuit, ICML'17.
- J. Shen and P. Li. Partial hard thresholding: Towards a principled analysis of support recovery, NIPS'17.
- P. Jain, A. Tewari, and P. Kar. On iterative hard thresholding methods for high-dimensional m-estimation, NIPS'14.
- T. Zhang. Sparse recovery with orthogonal matching pursuit under RIP. IEEE Transactions on Information Theory, 2011.

Results for Gaussian ensemble

- When rows of **A** are sampled from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ s.t. $\Sigma_{ii} \leq 1 \ \forall \ i \in [d]$.
- Generalization error: If $\hat{\mathbf{x}}_s$ is OMP's output
- after $s = \Theta\left(\kappa\left(\Sigma\right)\log\kappa\left(\Sigma\right)s^*\right)$ iterations,
 $n = \Omega\left(\frac{s\log d}{\sigma(\Sigma)}\right)$

$$\frac{1}{n} \|\mathbf{A} \left(\mathbf{\hat{x}}_s - \mathbf{\bar{x}}\right)\|_2^2 \lesssim \frac{1}{n} \sigma^2 \kappa \left(\mathbf{\Sigma}\right) \log \kappa \left(\mathbf{\Sigma}\right) s^* \quad \text{w.h.p.}$$

- Support Recovery: For $\Sigma = I$,
- $s = \Omega(s^*)$,
 $|\bar{x}_{\min}| = \Omega\left(\sigma\sqrt{\frac{\log d}{n}}\right)$, and
 $n = \Omega\left((s^*)^2\log d\right)$

$$\mathbf{S}^* \subseteq \operatorname{supp}(\mathbf{\hat{x}}_s)$$
, and $\|\mathbf{\hat{x}}_s - \mathbf{\bar{x}}\|_{\infty} \lesssim \sigma \sqrt{\frac{\log s}{n}}$ w.h.p.

Lower bound instance construction

- Fool OMP into picking incorrect indexes. Large support size \implies large generalization error.
- ullet Construct evenly distributed $ar{\mathbf{x}}$

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

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

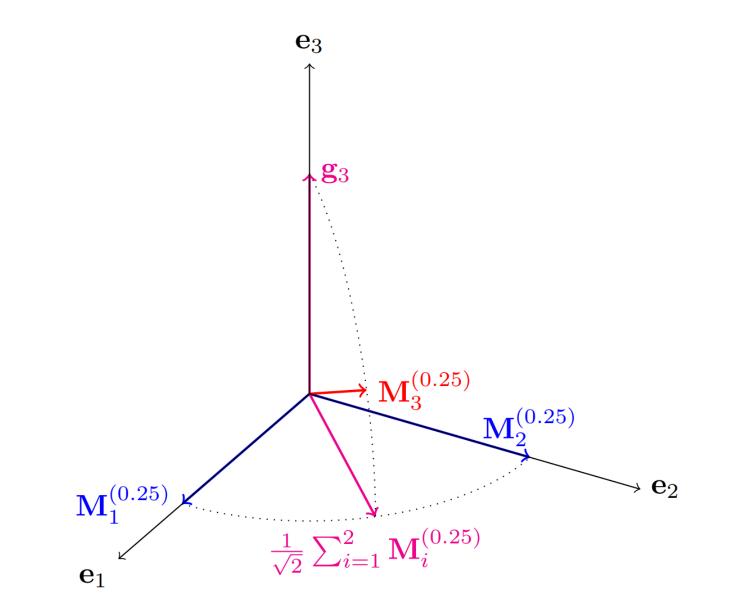


Figure 1:Visualizing in d=3 with $s^*=2$ and $\epsilon=0.25$.

Lower bound results

Lower bound: Noiseless case

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

$$\widetilde{\kappa}_s\left(\mathbf{M}^{(\epsilon)}\right) \lesssim \frac{s}{s^*}, \ \gamma\left(\mathbf{M}^{(\epsilon)}\right) = \mathcal{O}\left(1\right) \quad \text{and} \quad \mathbf{S}^* \cap s$$

Lower bound: Noise case

For $s^* \leq s \leq d^{1-\alpha}$, $\alpha \in (0,1)$, $\exists \epsilon > 0$ s.t. when OMP is executed on the SLR problem with $\mathbf{y} = \mathbf{M}^{(\epsilon)} + \boldsymbol{\eta}$ where $\boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_{n \times n}\right)$, then w.h.p., $\tilde{\kappa}_s\left(\mathbf{M}^{(\epsilon)}\right) \lesssim \frac{s}{s^*}$, $\gamma\left(\mathbf{M}^{(\epsilon)}\right) = \mathcal{O}\left(1\right)$,

$$\|\mathbf{A}\left(\mathbf{\hat{x}}_{s}-\mathbf{\bar{x}}\right)\|_{2}^{2} \gtrsim \frac{1}{n}\sigma^{2}s^{*}\widetilde{\kappa}_{s+s^{*}} \quad \text{and} \quad \mathbf{S}^{*} \cap \operatorname{supp}(\mathbf{\hat{x}}_{s}) = \phi.$$

These lower bounds show that $s \gtrsim \tilde{\kappa}_s s^*$ iterations are indeed necessary for support recovery.

Simulations

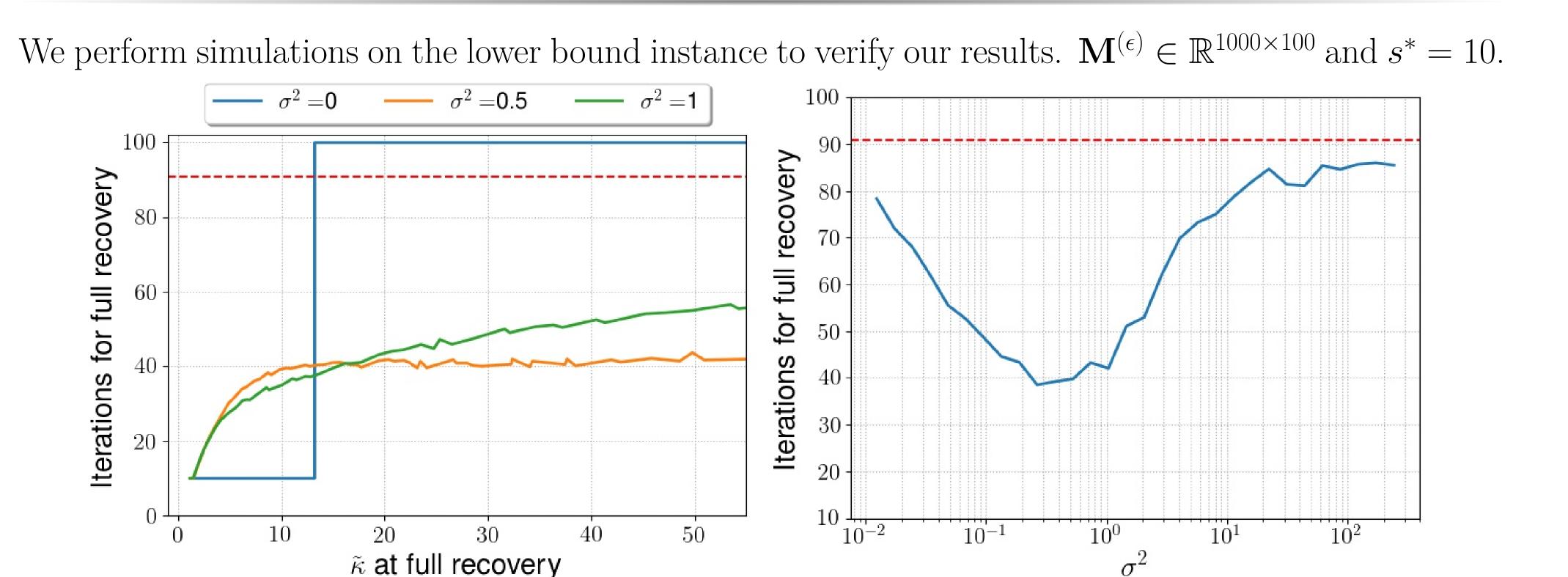


Figure 2:Number of iterations required for recovering the full support of $\bar{\mathbf{x}}$ w.r.t $\tilde{\kappa}_{s+s^*}$ of the design matrix and the sub-Gaussian parameter of the noise term (σ^2) .

When noise is very high, the selection step of OMP tends to become a uniform random selector. This is represented by the red dashed line.

Conclusion

- We obtain support recovery & generalization guarantees for OMP under RSC for SLR.
- We provide lower bounds for OMP showing that our results are tight up to logarithmic factors.
- Our results match the best known results for **SLR** that use non-convex penalty based methods.

References

1] Balas Kausik Natarajan.

Sparse approximate solutions to linear systems. SIAM journal on computing, 24(2):227–234, 1995.

[2] Tong Zhang.

Sparse recovery with orthogonal matching pursuit under rip.

IEEE Transactions on Information Theory, 57(9):6215–6221, 2011.

[3] Po-Ling Loh, Martin J Wainwright, et al.

Support recovery without incoherence: A case for nonconvex regularization. *The Annals of Statistics*, 45(6):2455–2482, 2017.

 \lesssim and \gtrsim are inequalities up to constants & $\operatorname{polylog} d$ factors.