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):

$$\bar{\mathbf{x}} = \arg\min_{\|\mathbf{x}\|_0 \le s^*} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \qquad \mathbf{A} \in \mathbb{R}^{n\times d}, \mathbf{y} \in \mathbb{R}^n.$$

- 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
  - ρ_s^-: Restricted Convexity (RSC) constant
- **Key quantity: Restricted Condition Number** – $\tilde{\kappa}_s = \frac{\rho_1^+}{\rho_s^-}$
- Assumption: Generative model. $\text{supp}(\bar{\mathbf{x}}) = \mathbf{S}^*, |\mathbf{S}^*| = s^*$.

$$\mathbf{y} = \mathbf{A}\bar{\mathbf{x}} + \boldsymbol{\eta} = \mathbf{A}_{\mathbf{S}^*}\bar{\mathbf{x}}_{\mathbf{S}^*} + \boldsymbol{\eta} \qquad \text{where } \boldsymbol{\eta} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}_{n\times n}\right).$$

Goals of SLR

1. **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 $\mathbf{S}$ s.t. $\mathbf{S} \supseteq \mathbf{S}^*$.

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)
- Greedy and Pursuit methods,
  - Hard Thresholding Pursuit (HTP)
  - Partial Hard Thresholding (PHT)
  - **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: $j \leftarrow \arg\max_{i \notin \mathbf{S}_{k-1}} |\mathbf{A}_i^T \mathbf{r}_{k-1}|$,
- Incrementally grow support: $\mathbf{S}_k \leftarrow \mathbf{S}_{k-1} \cup \{j\}$,
- Optimize on current support: $\mathbf{x}_k \leftarrow \arg\min_{\text{supp}(\mathbf{x}) \subseteq \mathbf{S}_k} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$,
- Update residual: $\mathbf{r}_k \leftarrow \mathbf{y} - \mathbf{A}\mathbf{x}_k$,

Return $\mathbf{x}_s$.
The size of the support of OMP = number of iterations.

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.,

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

- 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.,

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

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 $\mathbf{S}^* \setminus \text{supp}(\hat{\mathbf{x}}_s) \neq \phi$, 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 $\left\|\mathbf{A}_{\mathbf{S}^*\setminus\mathbf{S}}^T \mathbf{A}_{\mathbf{S}} \left(\mathbf{A}_{\mathbf{S}}^T\mathbf{A}_{\mathbf{S}}\right)^{-1}\right\|_\infty \le \gamma$.

- Objective function can be lower bounded $\therefore$ 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.,

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

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)	$\lvert\bar{x}_{\min}\rvert$ **lower bound**
Yuan et al. [HTP]	$\kappa_{2s}^2 s^*$	$\frac{\sigma\sqrt{s}}{\sqrt{\rho_{2s}^-}}$
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)]	$s^* + \kappa_{2s}^2 \min\{s^*, r\}$	$\frac{\sigma\sqrt{\kappa_{2s}}\sqrt{\rho_1^+ s}}{\rho_{2s}^-}$
Jain et al. [IHT]	$\kappa_{2s+s^*}^2 s^*$	–
Zhang [OMP]	$\tilde{\kappa}_{s+s^*} s^* \log \kappa_{s+s^*}$	–
Our's [OMP]	$\tilde{\kappa}_{s+s^*} s^* \log \kappa_{s+s^*}$	$\gamma \cdot \frac{\sigma\sqrt{\rho_1^+}}{\rho_{s+s^*}^-}$

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

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

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

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

Lower bound instance construction

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

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

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

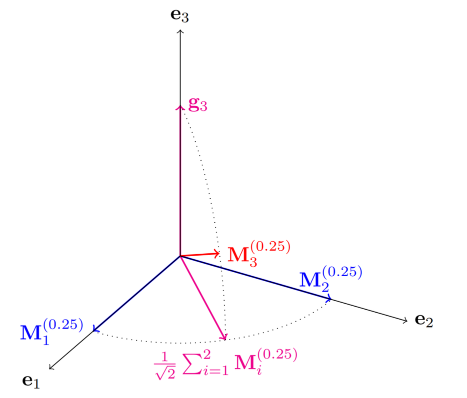

Figure 1:Visualizing in $d = 3$ with $s^* = 2$ and $\epsilon = 0.25$.

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 $\mathbf{y} = \mathbf{M}^{(\epsilon)}\bar{\mathbf{x}}$ for $s \le d - s^*$ iterations, then

$$\tilde{\kappa}_s\left(\mathbf{M}^{(\epsilon)}\right) \lesssim \frac{s}{s^*},\ \gamma\left(\mathbf{M}^{(\epsilon)}\right) = \mathcal{O}(1) \qquad \text{and} \qquad \mathbf{S}^* \cap \text{supp}(\hat{\mathbf{x}}_s) = \phi$$

Lower bound: Noise case

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

$$\frac{1}{n}\|\mathbf{A}(\hat{\mathbf{x}}_s - \bar{\mathbf{x}})\|_2^2 \gtrsim \frac{1}{n}\sigma^2 s^*\tilde{\kappa}_{s+s^*} \qquad \text{and} \qquad \mathbf{S}^* \cap \text{supp}(\hat{\mathbf{x}}_s) = \phi.$$

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

Simulations

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

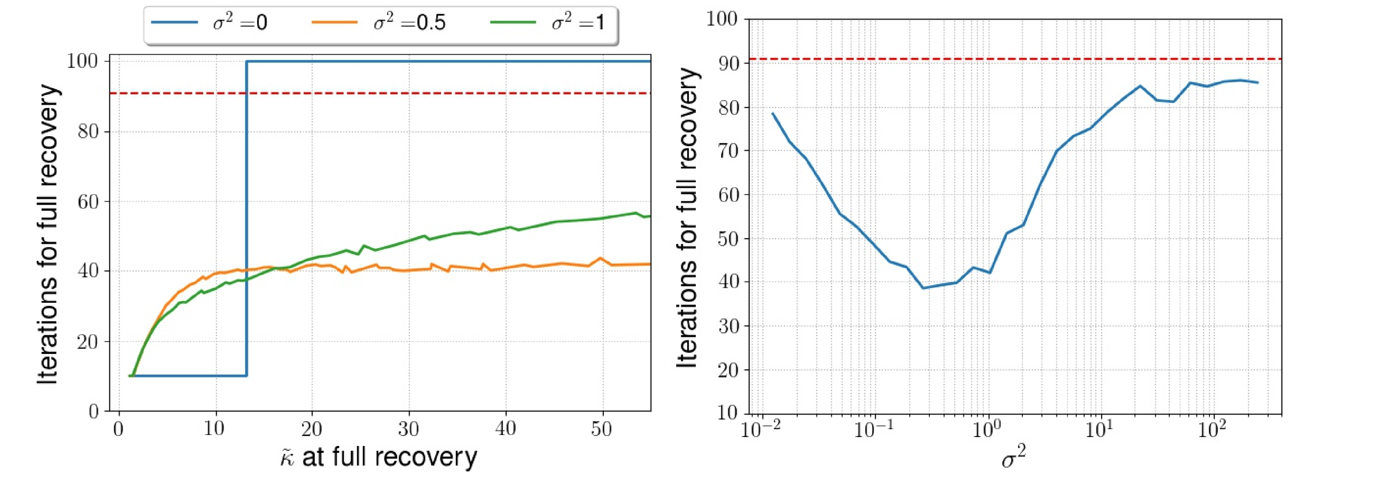

Figure 2:Number of iterations required for recovering the full support of $\bar{\mathbf{x}}$ w.r.t $\tilde{\kappa}_{s+s^*}$ of the design matrix and the sub-Gaussian parameter of the noise term (σ^2).

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 & polylog d factors.