Sparse Channel Estimation for Massive MIMO with Compressed Sensing via Convex Optimization

Rittwik Sood
Advisor: Prof. Christopher Brinton

PUID: 0036403763

Department of E&CE
PURDUE UNIVERSITY

DATE SUBMITTED
[May 9, 2025]

Declaration

"I certify that the work for this project was done entirely in this class, even if it is related to my own research".

Abstract

This project implements a compressed sensing approach for sparse channel estimation in massive MIMO systems, addressing the challenge of reducing pilot overhead while maintaining estimation accuracy. In this massive MIMO channel estimation problem, I am solving a sparse signal recovery problem using convex optimization. In the end, we compare several methods to see which is more accurate and better in terms of Normalised Mean Square Error (NMSE) and support recovery rate.

Project Link (Code included)

github/rittwiksood/ECE647 Final Project

1 Introduction

The exponential growth in wireless data traffic has driven the development of advanced communication technologies capable of meeting these increasing demands. Among these innovations, Massive Multiple-Input Multiple-Output (MIMO) systems have emerged as a cornerstone technology for fifth-generation (5G) and beyond wireless networks due to their potential to significantly enhance spectral efficiency, link reliability, network capacity and data rates [1, 2]. Massive MIMO systems, characterized by base stations equipped with hundreds or thousands of antennas serving multiple users simultaneously, promise substantial performance gains through spatial multiplexing and beamforming [3]. A critical challenge in realizing the full potential of massive MIMO systems lies in accurate channel state information (CSI) estimation. In conventional MIMO systems, pilot-based channel estimation techniques are commonly employed, where orthogonal pilot sequences are transmitted to enable the receiver to estimate the channel coefficients [4].

A key challenge in massive MIMO is channel estimation, particularly due to the high-dimensional nature of the channel matrix. Traditional least-squares (LS) estimation methods require a large number of pilot symbols, leading to excessive overhead and reduced spectral efficiency. As the number of antennas scales up in massive MIMO deployments, traditional channel estimation approaches become increasingly impractical due to prohibitive pilot overhead and computational complexity [5, 6]. Fortunately, wireless channels often exhibit sparsity in certain transform domains, particularly in millimeter-wave (mmWave) and higher frequency bands [7]. This inherent sparsity presents an opportunity to leverage compressed sensing (CS) techniques for efficient channel estimation [8]. Compressed sensing, a signal processing paradigm that enables the reconstruction of sparse signals from fewer measurements than traditionally required by the Nyquist sampling theorem, offers a promising approach to address the channel estimation challenges in massive MIMO systems [9, 10]. The key advantage of compressed sensing is its ability to break the Nyquist sampling barrier, allowing accurate channel estimation even when the number of pilots (P) is much smaller than the number of transmit antennas N_t. This is particularly crucial for frequency-division duplexing (FDD) systems, where downlink channel estimation requires extensive feedback. The integration of compressed sensing with massive MIMO channel estimation has attracted significant research interest in recent years [11]. By exploiting channel sparsity, CS-based approaches can substantially reduce pilot overhead while maintaining estimation accuracy. Various CS algorithms, including orthogonal matching pursuit (OMP) [12], approximate message passing (AMP) [13], and basis pursuit denoising (BPDN) [14], have been adapted for sparse channel estimation in massive MIMO contexts. Additionally, structured sparsity models that incorporate prior knowledge about channel characteristics have been developed to further enhance estimation performance [15]. Furthermore, the integration of machine learning approaches with compressed sensing has opened new avenues for intelligent and adaptive channel estimation frameworks [16, 17].

Since wireless channels in millimeter-wave (mmWave) and massive MIMO systems exhibit inherent sparsity—where only a few dominant propagation paths exist, CS techniques enable accurate channel recovery with significantly fewer pilot symbols than conventional methods. This project explores sparse signal recovery algorithms for massive MIMO channel estimation, including:

- Basis Pursuit **(LASSO)** – Formulating channel estimation as an l_1-regularized convex optimization problem [14]
- **Group LASSO**– Exploiting structured sparsity in antenna correlations [18]
- **Orthogonal Matching Pursuit (OMP)** – A greedy algorithm for low-complexity sparse recovery

This project investigates advanced sparse channel estimation techniques for massive MIMO systems using compressed sensing, with a focus on developing practical solutions that balance estimation accuracy, computational efficiency, and pilot overhead requirements. This problem of sparse channel estimation in massive MIMO can be naturally formulated as a convex optimization problem by exploiting the inherent sparsity of wireless channels. In this project, we formulate this convex optimization problem and try to solve this by regulating parameters. The further paper is divided in following sections. Section 2 talks about the optimization problem and discuss its properties. Section 3 presents the algorithms used and their implementation. Section 4 presents the simulation results and analysis and, Section 5 provides Conclusion and discussion on the project's Future scope.

2 Formulation

The problem of sparse channel estimation in massive MIMO can be naturally formulated as a convex optimization problem by exploiting the inherent sparsity of wireless channels. Sparse channel estimation in massive MIMO systems is fundamentally an under determined linear inverse problem, where the number of pilot symbols ***(P)*** is much smaller than the number of transmit antennas (N_t). Traditional least-squares estimation fails in this regime because the system of equations is ill-posed. Compressed sensing provides a solution by exploiting channel sparsity, allowing accurate recovery even when $P << N_t$.

Firstly, We consider a massive MIMO system with:

$\rightarrow$ (N_t) transmit antennas
$\rightarrow$ (N_r) receive antennas
$\rightarrow$ S sparse multi path components
$\rightarrow$ P pilot symbols where $P << N_t$

Since the channel matrix H, where

$$H \in \mathbb{C}(N_t \times N_r)$$

has only a few dominant multipath components (i.e., it is sparse), we can leverage l_1-norm minimization techniques from compressed sensing to recover H efficiently from limited pilot observations. The received signal model is as follows:

$$y = \phi H + n \tag{1}$$

Where:

- $y \in \mathbb{C}(P \times N_r)$ is the received signal
- $\phi \in \mathbb{C}(P \times N_t)$ is the pilot matrix
- $H \in \mathbb{C}(N_t \times N_r)$
- $n \in \mathbb{C}(P \times N_r)$ is additive white Gaussian noise

Core Optimization Problem

We're solving a regularized least squares problem:

$$\text{minimize } ||y - \phi H||_2^2 + \lambda \cdot R(H) \tag{2}$$

where R(H) is a regularization term enforcing sparsity and $\lambda > 0$ is the regularization parameter. The below subsection now talks in detail for all the three methods, how the problem is convex and its properties.

2.1 LASSO (Basis Pursuit Denoising)

The Least Absolute Shrinkage and Selection Operator (LASSO) formulation for channel estimation is given by:

$$\min_{H} ||y - \phi H||_2^2 + \lambda ||H||_1 \tag{3}$$

Here,

- $\|y - \phi H\|_2^2$ is the least-squares data fidelity term, ensuring the estimated channel matches the received pilot signals.
- $\|H\|_1$ is the l_1-norm sparsity penalty, promoting solutions where most entries of H are zero.
- $\lambda > 0$ is a regularization parameter that balances sparsity and estimation accuracy

This problem is ***convex*** because:

(a) The squared l_2-norm is strictly convex as The Hessian $\nabla^2 f(H) = 2\Phi^H \Phi$ is positive semidefinite (PSD).

(b) The l_1-norm is convex (as it is a sum of absolute values).

(c) The sum of two convex functions remains convex.

2.1.1 Properties

1. **Convexity**: As discussed above, problem is convex.
2. **Non-smoothness**: The ℓ_1-norm is non-smooth at points where any component of H is zero, which makes the optimization problem non-differentiable at those points.
3. **Sparsity**: The l_1-norm regularization tends to produce sparse solutions, making it suitable for compressed sensing applications.

4. **Solution uniqueness**: The solution may not be unique if the measurement matrix Φ does not satisfy certain conditions, such as the restricted isometry property (RIP).

2.2 Group LASSO

In realistic massive MIMO systems, antennas exhibit correlated sparsity due to physical propagation characteristics. The Group LASSO extends LASSO [19] by enforcing group-wise sparsity:

$$\text{minimize } ||y - \phi H||_2^2 + \lambda \cdot \Sigma_{k=1}^{K} ||H_k||^2 \tag{4}$$

where H_k represents a subgroup of channel coefficients (e.g., from adjacent antennas).

This problem is ***convex*** because:

(a) The **group l_2 norm** is convex (since norms are convex functions).

(b) The sum of convex functions remains convex, so the problem is convex.

(c) Unlike standard LASSO, Group LASSO enforces group-wise sparsity (entire groups may be zero).

2.2.1 Properties

1. **Convexity**: Group LASSO is also a convex optimization problem
2. **Group sparsity**: It promotes sparsity at the group level, where either all elements in a group are zero or all are non-zero.
3. **Structured solution**: Group LASSO yields solutions that respect the inherent structure in the channel, which can improve estimation performance when the channel exhibits group-sparse characteristics.
4. **Complexity**: Generally more complex than standard LASSO but still solvable using various convex optimization techniques.
5. **Non-smooth but separable**: The regularization term is non-smooth but separable across groups, which can be exploited in optimization algorithms.

2.3 Orthogonal Matching Pursuit (OMP)

Unlike LASSO, which is formulated as a continuous optimization problem, OMP is a greedy algorithm that approximately solves:

$$\text{minimize } ||y - \phi H||_2^2 \quad \text{subject to } ||H||_0 \leq S \tag{5}$$

where $||H||_0$ counts the number of non zeros (sparsity constraint).

This problem is ***non-convex*** because:

(a) The l_0 norm makes the problem combinatorially hard (NP-complete).

(b) OMP approximates the solution via greedy selection of support (non-zero indices).

2.3.1 Properties

1. **Convexity**: Non Convex
2. **Guaranteed recovery**: Under certain conditions on Φ (e.g., restricted isometry property) and sparsity level, OMP can provably recover the true sparse signal.
3. **Structured solution**: Group LASSO yields solutions that respect the inherent structure in the channel, which can improve estimation performance when the channel exhibits group-sparse characteristics.
4. **Complexity**: Generally more complex than standard LASSO but still solvable using various convex optimization techniques.
5. **Non-smooth but separable**: The regularization term is non-smooth but separable across groups, which can be exploited in optimization algorithms.

3 Analysis and algorithm

In this section, I talk about the dual problem and whether a duality gap can be expected or not. Also I define the performance parameters upon which, we can judge the best method for optimizing our problem of Sparse channel estimation.

3.1 LASSO

3.1.1 Dual Problem Formulation

The Lagrangian is:

$$\min_H ||y - \phi H||_2^2 + \lambda ||H||_1 - u^T(H - z) \tag{6}$$

where z is an auxiliary variable. The dual problem is derived by minimizing the Lagrangian w.r.t. H. After simplification which becomes,

$$\max_{||\Phi^T \mathbf{u}||_\infty \leq \lambda} -\frac{1}{4}|u|_2^2 + u^T y \tag{7}$$

3.1.2 Duality Gap

For LASSO, there is *no duality gap* under typical conditions because:

1. The primal problem is convex
2. The objective function is continuous
3. Slater's condition is satisfied (and therefore, Strong Duality holds)

3.1.3 Algorithm

Since LASSO, is convex we can use Algorithm: *Proximal Gradient Descent (FISTA)* taken from [20].

Algorithm 1 LASSO Optimization (Taken from [20])

1: Initialize $\mathbf{H}_0$, step size $\eta = 1/L$ (where $L = 2\lambda_{\max}(\Phi^T\Phi)$).

2: Gradient step:

$$\mathbf{G}_k = H_k - \eta\Phi^T(\Phi H_k - \mathbf{y})$$

3: Proximal (shrinkage) step:

$$H_{k+1} = \text{sign}(G_k) \odot \max\left(|G_k| - \lambda\eta, 0\right)$$

4: Concluding step:

$$H_{k+1} = H_{k+1} + \frac{k}{k+3}(H_{k+1} - H_k)$$

5: Terminate step:

$$||H_{k+1} - H_k||_2 \leq \epsilon$$

3.2 Group LASSO

3.2.1 Dual Problem Formulation

Similar to LASSO, we can derive the dual problem by introducing auxiliary variables and forming the Lagrangian. The dual is derived via conjugate functions:

$$\max_{||\Phi^T\mathbf{u}||_{2,\star} \leq \lambda} -\frac{1}{4}|u|_2^2 + u^T y \tag{8}$$

, where $||\cdot||_{2,\star}$ is the dual group norm (block-wise l_2-norm).

3.2.2 Duality Gap

As with LASSO, Group LASSO is a convex optimization problem, and under standard conditions:

1. The objective function is convex and continuous
2. Slater's condition is satisfied

Therefore, strong duality holds, and there is *no duality gap* between the primal and dual problems. The optimal primal solution can be obtained from the dual solution.

3.2.3 Algorithm

We use BCD method [21] to optimize the objective function of the Group LASSO method. Algorithm 2 discusses the same.

3.3 OMP

3.3.1 Dual Problem Formulation

The dual problem formulation for the OMP optimization is challenging because the ℓ_0 "norm" makes the problem non-convex. Unlike convex problems, there's no standard duality theory that applies directly. Therefore, a dual does not exist in this problem.

Algorithm 2 Block Coordinate Descent (BCD)

1: Partition H into K groups of size G.
2: Initialize: $H = 0$.
3: Iterate over groups $k = 1, \ldots, K$:
- Compute residual:
$$\mathbf{r} = \mathbf{y} - \Phi_{-k} H_{-k}$$
- Update group k:
$$H_k = \left(1 - \frac{\lambda}{||\Phi_k^T \mathbf{r}||_2}\right)_+ \Phi_k^T \mathbf{r}$$
where $(\cdot)_+$ denotes the positive part, i.e., $\max(\cdot, 0)$.

4: Terminate when:
$$||H^{(t)} - H^{(t-1)}||_2 < \epsilon$$

Algorithm 3 Greedy Support Selection

1: **Initialize:**
$$\mathbf{r} = \mathbf{y}, \quad \mathcal{S} = \emptyset$$
2: **Iterate for** $s = 1, \ldots, S$:
- Find most correlated column:
$$i = \arg\max_j \left|\phi_j^T \mathbf{r}\right|$$
- Update support set:
$$\mathcal{S} = \mathcal{S} \cup \{i\}$$
- Solve least squares on support:
$$\mathbf{H}_{\mathcal{S}} = \left(\Phi_{\mathcal{S}}^T \Phi_{\mathcal{S}}\right)^{-1} \Phi_{\mathcal{S}}^T \mathbf{y}$$
- Update residual:
$$\mathbf{r} = \mathbf{y} - \Phi_{\mathcal{S}} \mathbf{H}_{\mathcal{S}}$$

3: **Terminate when** $|\mathcal{S}| = S$.

3.3.2 Duality Gap

For OMP, since it solves a non-convex problem (due to the ℓ_0 norm), there typically exists a duality gap between the primal and dual problems. This means that the optimal value of the dual problem provides only a lower bound on the optimal value of the primal problem.

3.3.3 Algorithm

Algorithm 3 provides implements OMP algorithm in this project. Being a greedy algorithm for a non-convex problem, OMP does not guarantee global optimality. However, it does provide local optimality in the sense that each iteration makes the locally optimal choice by selecting the column most correlated with the current residual. Under specific conditions on the measurement matrix Φ (such as satisfying the Restricted Isometry Property with appropriate constants), OMP can provably recover the true sparse signal or find a solution close to the global optimum.

3.4 Performance Metrics

I have evaluated the above three methods on below performance metrics. These metrics allows us to evaluate the above three methods on the basis of different parameters and gives a clear indication which method can be used in what test case. It also highlights trade offs between these three different methods

3.4.1 Normalized Mean Squared Error (NMSE)

- NMSE is a fundamental metric that quantifies the average squared difference between the estimated channel and the true channel, normalized by the energy of the true channel
- Measures estimation accuracy of the channel matrix:

$$\textbf{NMSE} \; = \frac{||H_{est} - H_{true}|\,|_F^2}{||H_{true}|\,|_F^2} \tag{9}$$

- In the context of massive MIMO channel estimation, NMSE serves as a comprehensive measure of estimation accuracy that directly impacts the achievable system performance.
- A lower NMSE generally translates to higher spectral efficiency, better beamforming gain, and improved interference mitigation
- Lower NMSE indicates better reconstruction of channel coefficients.
- Methods that are able to accomplish global optimization has a typically lower NMSE.

3.4.2 Support Recovery Error

- While NMSE provides a comprehensive measure of estimation accuracy, it does not explicitly evaluate the algorithm's ability to identify the correct support, which is crucial in sparse channel estimation
- Support Recovery Error addresses this limitation by focusing on the algorithm's capability to correctly identify the positions of non-zero elements in the channel vector.
- Mathematically, it can be defined as:

$$\textbf{Support Error} = \frac{\text{Number of misclassified non-zeros}}{\text{Total entries}} \tag{10}$$

- Lower support error highlight better identification of active paths.

4 Simulation results

This section provides results and plots of the optimization problems discussed above. Table 1 provides the summarized comparison of all the methods. Lasso and Group Lasso are convex functions whereas OMP is a non-convex function. Also Table 1 sumarises the dual problems if existing and give an indication of Duality gap in three of the algorithms.

4.1 Summary Comparison of Dual Problems and Solvers

Method	**Convexity**	**Dual Problem**	**Duality Gap**	**Algorithm**	**Optimality Guarantee**
LASSO	Convex	$\max_{\|\|\Phi^T \mathbf{u}\|\|_\infty \leq \lambda} -\frac{1}{4}\|u\|_2^2 + u^T y$	No gap (strong duality holds)	Proximal Gradient Descent (FISTA)	Global optimality
Group LASSO	Convex	$\max_{\|\|\Phi^T \mathbf{u}\|\|_{2,\star} \leq \lambda} -\frac{1}{4}\|u\|_2^2 + u^T y$	No gap (strong duality holds)	Block Coordinate Descent	Global optimality
OMP	Non-convex	No standard dual (related to basis pursuit dual)	Typically has duality gap	Greedy iterative algorithm	Local optimality

Table 1: Comparison of methods: convexity, dual problems, and solution guarantees

4.1.1 Choice of Regularization Parameter λ

- Too large λ: Over-sparsifies indicating high estimation error.
- Too small λ: Under-sparsifies indicating poor pilot compression.

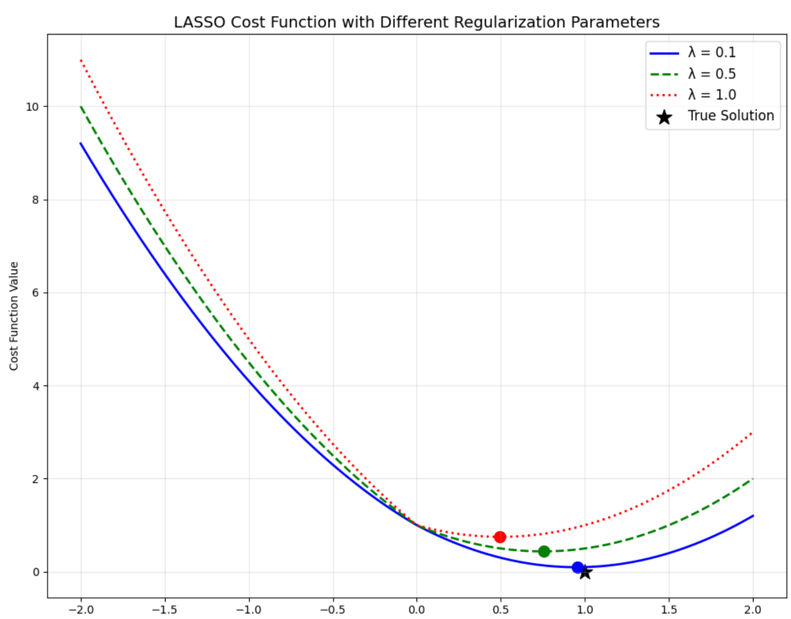

Figure 1: Choice of Regularisation Parameter influencing convergence

Figure 1 depicts how, for a large lambda, it over sparsifies and for a correct $\lambda = 0.1$, it is near the true solution.

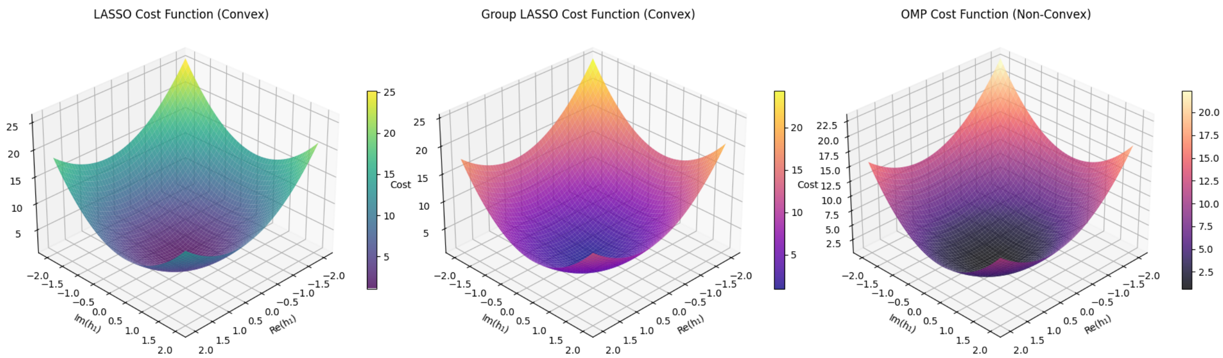

Figure 2: Convexity of the methods used in this project

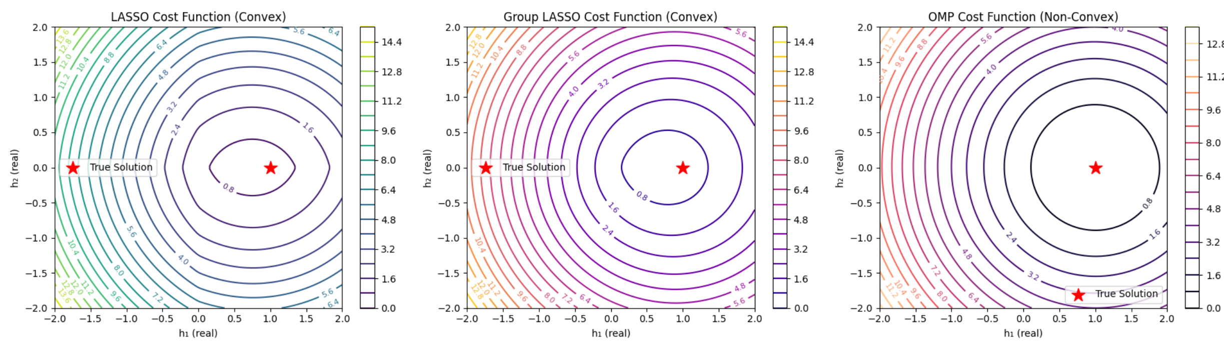

Figure 3: Cost functions of the methods on a 2-D plane

4.1.2 Computational Complexity

Method	Complexity (per iteration)	Convergence Guarantee
LASSO	$\mathcal{O}(N_t^3)$	Global optimum
Group LASSO	$\mathcal{O}(KN_t^2)$ (block-wise)	Global optimum
OMP	$\mathcal{O}(SPN_t)$	Local (greedy) solution

Table 2: Comparison of computational complexity and convergence guarantees

Table 2 provides the time complexity of each of the method discussed in this project.

4.2 Convexity of the methods

Figure 2 shows the cost functions of each of the method. We can conclude that LASSO and Group LASSO are convex in nature whereas OMP is non-convex. The figure shows cost as a function of real and imaginary parts of channel coefficients. Figure 3 whereas, plot the same functions on a 2-D visualization for a better understanding.

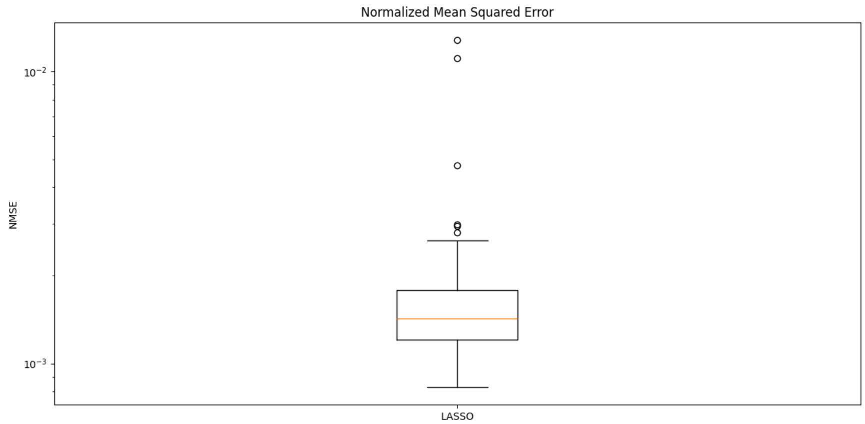

Figure 4: NMSE range for LASSO cost function

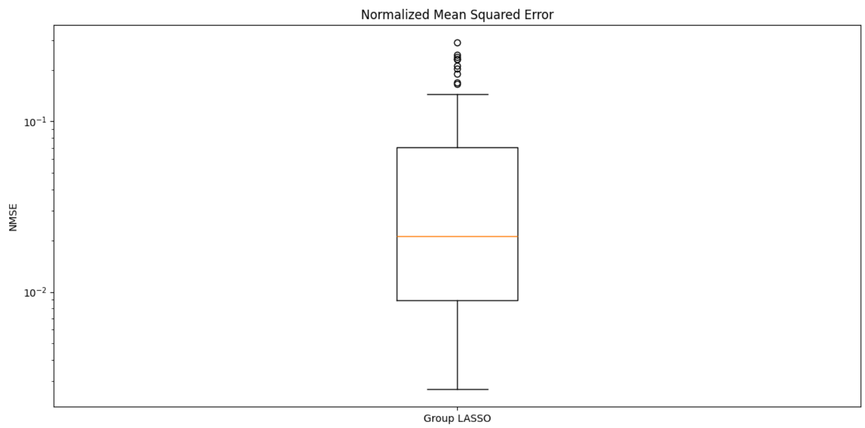

Figure 5: NMSE range for Group LASSO cost function

4.3 NMSE comparisons

Figure 4,5,6 represents the NMSE values for Lasso, Group Lasso and OMP respectively. On Y-axis (log scale): NMSE values are plotted on a logarithmic scale, which helps visualize a wide range of errors more clearly. The central box represents the interquartile range (IQR), i.e., the middle 50% of NMSE values. The Orange line is the median NMSE, a measure of its central tendency. The circles, which are the data points outside the whiskers, represents cases where the NMSE was significantly worse than the norm.

Below table talks about the Average NMSE and the Average Support errors for all the algorithms.

```
Average NMSE - LASSO: 0.0018
Average NMSE - Group LASSO: 0.0546
Average NMSE - OMP: 0.0383
Average Support Error - LASSO: 0.2450
Average Support Error - Group LASSO: 0.5769
Average Support Error - OMP: 0.1101
```

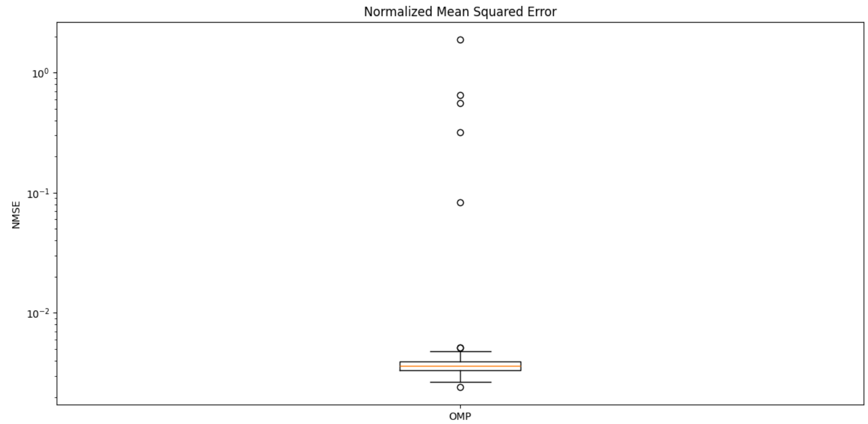

Figure 6: NMSE range for OMP cost function

For LASSO, the NMSE performance is generally good when the channel exhibits element-wise sparsity without specific group structures. However, LASSO may struggle to accurately estimate the magnitudes of the non-zero coefficients due to the shrinkage effect, especially when the regularization parameter is large. Group LASSO, on the other hand, is expected to achieve superior NMSE performance when the channel exhibits inherent group sparsity, such as in millimeter-wave MIMO systems where channel paths tend to concentrate in specific angular or spatial clusters. By leveraging the prior knowledge of group structure, Group LASSO can better preserve the energy distribution within the active groups, potentially leading to more accurate amplitude estimation. OMP follows a fundamentally different approach by greedily building the support set through iterative selection. This approach often results in excellent NMSE when the channel is exactly sparse and the measurement matrix satisfies certain conditions (such as the Restricted Isometry Property). However, OMP's performance can degrade rapidly when the true channel is only approximately sparse or when the noise level is high, as the greedy selections may miss important components or include spurious ones.

4.4 Support Recovery Error (SRE)

In massive MIMO systems, accurate support recovery is particularly important for several reasons. First, it directly impacts the sparsity level of the estimated channel, which affects the efficiency of subsequent channel-based operations such as precoding and detection. LASSO's performance in terms of SRE is influenced by the choice of the regularization parameter. LASSO may struggle with closely spaced non-zero elements or groups, as the L1 norm tends to select one representative from correlated variables. Group LASSO inherently performs better in support recovery when the true channel exhibits the assumed group structure. By enforcing group-wise sparsity, it can more effectively identify the active groups, especially when the energy is distributed across multiple elements within a group. This property makes Group LASSO particularly suitable for massive MIMO systems operating at millimeter-wave frequencies, where the channel paths often cluster in specific angular sectors. Figure 7 shows us the NMSE and SRE plots for all the algorithms. OMP, with its direct focus on building

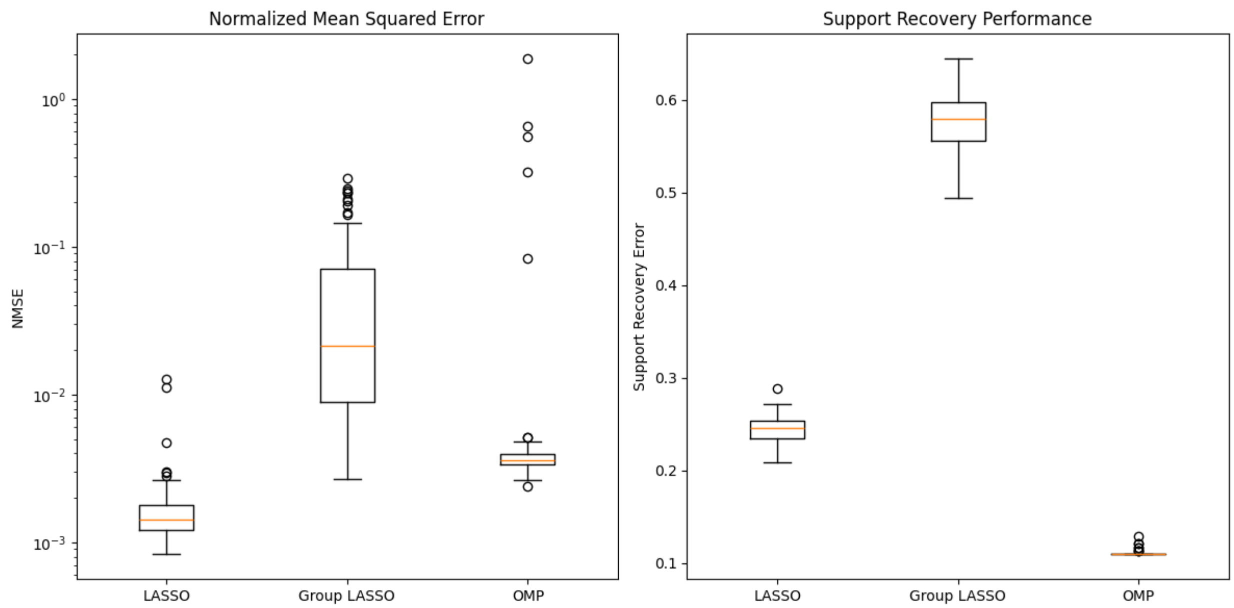

Figure 7: NMSE and SRP plots for LASSO, Group Lasso and OMP

the support set iteratively, can achieve excellent support recovery performance when the stopping criterion is well-tuned to the true sparsity level. Its greedy nature ensures that each newly selected element provides the maximum immediate reduction in the residual error. However, OMP may suffer from error propagation, where early incorrect selections can mislead subsequent iterations, potentially resulting in support sets that significantly deviate from the true ones.

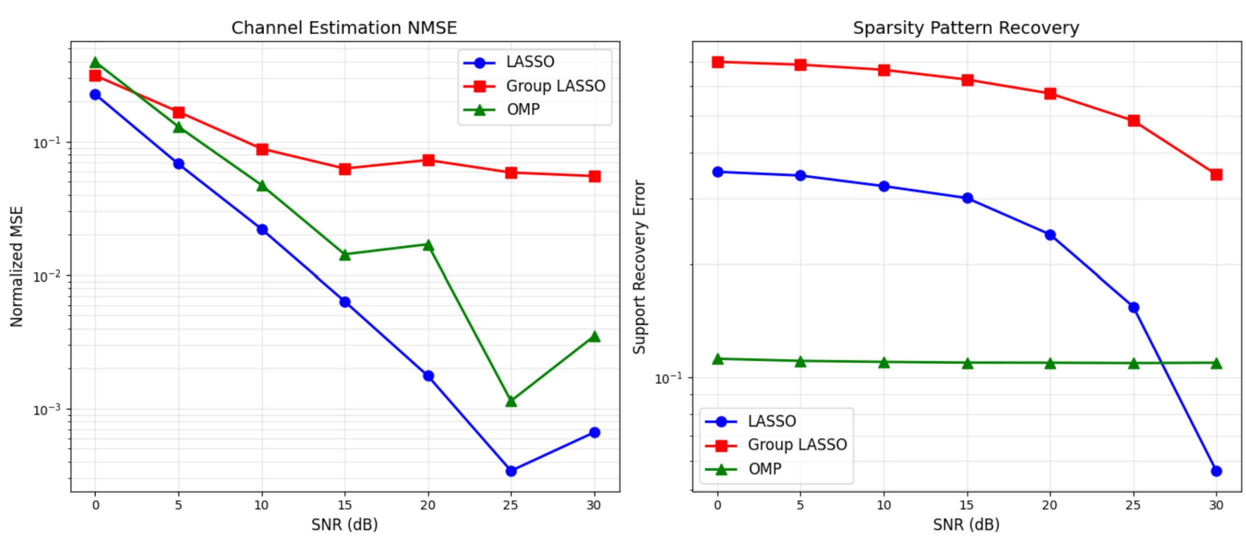

Figure 8: NMSE and SRE Comparative analysis

Figure 8 represents NMSE and SRE analysis across different SNRs (from 0 dB - 30 dB)

4.5 Interplay Between NMSE and SRE in Massive MIMO

The relationship between NMSE and SRE is complex and not always monotonic. In some cases, an algorithm might achieve excellent support recovery but still exhibit high NMSE due to poor amplitude estimation of the non-zero coefficients. Conversely, an algorithm might achieve reasonable NMSE by correctly estimating the dominant components while missing smaller ones, resulting in poor SRE. In the context of massive MIMO, the relative importance of NMSE versus SRE depends on the specific application scenario. The performance of LASSO, Group LASSO, and OMP on these metrics also varies with the characteristics of the massive MIMO channel. In scenarios with well-separated paths and high sparsity, OMP may outperform both LASSO and Group LASSO in terms of both NMSE and SRE. In contrast, when the channel exhibits more complex structures or is only approximately sparse, the convex optimization-based approaches of LASSO and Group LASSO might provide more robust estimation. The analysis explains why both metrics are necessary for a complete assessment of sparse channel estimation algorithms in massive MIMO systems, with NMSE focusing on overall estimation accuracy and SRE emphasizing the correct identification of the channel support structure.

5 Conclusion

- **LASSO** provides a convex, **globally optimal** solution for sparse MIMO channel estimation.
- **Group LASSO** extends this to structured sparsity, improving performance in correlated antenna systems.
- **OMP** is a non-convex but **fast alternative** for real-time applications
- Theoretical guarantees (RIP) ensure reliable recovery when pilots are well-designed.

This optimization framework enables low-pilot-overhead channel estimation, crucial for 5G/6G, mmWave, and IoT applications. Future work may integrate deep learning for adaptive regularization or distributed optimization for cell-free massive MIMO.

6 Code

Project Report: Report Link
Code Link: Code

References

[1] T. L. Marzetta, ”Noncooperative cellular wireless with unlimited numbers of base station antennas,” IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3590-3600, Nov. 2010.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, ”Massive MIMO for next generation wireless systems,” IEEE Commun. Mag., vol. 52, no. 2, pp. 186-195, Feb. 2014.

[3] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, ”An overview of massive MIMO: Benefits and challenges,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 742-758, Oct. 2014.

[4] J. Hoydis, S. ten Brink, and M. Debbah, ”Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?,” IEEE J. Sel. Areas Commun., vol. 31, no. 2, pp. 160-171, Feb. 2013.

[5] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, ”A coordinated approach to channel estimation in large-scale multiple-antenna systems,” IEEE J. Sel. Areas Commun., vol. 31, no. 2, pp. 264-273, Feb. 2013.

[6] X. Gao, L. Dai, S. Han, C.-L. I, and R. W. Heath Jr., ”Energy-efficient hybrid analog and digital precoding for mmWave MIMO systems with large antenna arrays,” IEEE J. Sel. Areas Commun., vol. 34, no. 4, pp. 998-1009, Apr. 2016.

[7] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath Jr., ”Channel estimation and hybrid precoding for millimeter wave cellular systems,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 831-846, Oct. 2014.

[8] D. L. Donoho, ”Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

[9] E. J. Candès and M. B. Wakin, ”An introduction to compressive sampling,” IEEE Signal Process. Mag., vol. 25, no. 2, pp. 21-30, Mar. 2008.

[10] Z. Gao, L. Dai, W. Dai, B. Shim, and Z. Wang, ”Structured compressive sensing-based spatio-temporal joint channel estimation for FDD massive MIMO,” IEEE Trans. Commun., vol. 64, no. 2, pp. 601-617, Feb. 2016.

[11] X. Rao and V. K. N. Lau, ”Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems,” IEEE Trans. Signal Process., vol. 62, no. 12, pp. 3261-3271, Jun. 2014.

[12] J. A. Tropp and A. C. Gilbert, ”Signal recovery from random measurements via orthogonal matching pursuit,” IEEE Trans. Inf. Theory, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[13] D. L. Donoho, A. Maleki, and A. Montanari, ”Message-passing algorithms for compressed sensing,” Proc. Nat. Acad. Sci. USA, vol. 106, no. 45, pp. 18914-18919, Nov. 2009.

[14] S. S. Chen, D. L. Donoho, and M. A. Saunders, ”Atomic decomposition by basis pursuit,” SIAM Rev., vol. 43, no. 1, pp. 129-159, Feb. 2001.

[15] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, ”Compressed channel sensing:

A new approach to estimating sparse multipath channels," Proc. IEEE, vol. 98, no. 6, pp. 1058-1076, Jun. 2010.

[16] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," IEEE Wireless Commun. Lett., vol. 7, no. 5, pp. 852-855, Oct. 2018.

[17] Y. Yang, F. Gao, G. Y. Li, and M. Jian, "Deep learning-based downlink channel prediction for FDD massive MIMO system," IEEE Commun. Lett., vol. 23, no. 11, pp. 1994-1998, Nov. 2019.

[18] Meier L, Van De Geer S, Bühlmann P. The group lasso for logistic regression. Journal of the Royal Statistical Society Series B: Statistical Methodology. 2008 Feb;70(1):53-71.

[19] Simon N, Friedman J, Hastie T, Tibshirani R. A sparse-group lasso. Journal of computational and graphical statistics. 2013 Apr 1;22(2):231-45.

[20] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring. In2009 IEEE International Conference on Acoustics, Speech and Signal Processing 2009 Apr 19 (pp. 693-696). IEEE.

[21] Tseng P. Convergence of a block coordinate descent method for nondifferentiable minimization. Journal of optimization theory and applications. 2001 Jun;109:475-94.