MultiPass Lasso Algorithms for Sparse Signal Recovery

Yuzhe Jin
Department of Electrical and Computer Eng.
University of California, San Digeo
9500 Gilman Drive, La Jolla, CA 92093-0407, USA
Email: yujin@ucsd.edu

Bhaskar D. Rao
Department of Electrical and Computer Eng.
University of California, San Digeo
9500 Gilman Drive, La Jolla, CA 92093-0407, USA
Email: brao@ece.ucsd.edu

***Abstract*—We develop the MultiPass Lasso (MPL) algorithm for sparse signal recovery. MPL applies the Lasso algorithm in a novel, sequential manner and has the following important attributes. First, MPL improves the estimation of the support of the sparse signal by combining high quality estimates of its partial supports which are sequentially recovered via the Lasso algorithm in each iteration/pass. Second, the algorithm is capable of exploiting the dynamic range in the nonzero magnitudes. Preliminary theoretic analysis shows the potential performance improvement enabled by MPL over Lasso. In addition, we propose the Reweighted MultiPass Lasso algorithm which substitutes Lasso with MPL in each iteration of Reweighted ℓ_1 Minimization. Experimental results favorably support the advantages of the proposed algorithms in both reconstruction accuracy and computational efficiency, thereby supporting the potential of the MultiPass framework for algorithmic development.**

***Index Terms*—Sparse signal recovery, MultiPass Lasso, Reweighted MultiPass Lasso, group detector, multiuser detection**

I. Introduction

Consider the estimation of a sparse signal through its linear measurements in the presence of noise, namely based on the model

$$\mathbf{Y} = A\mathbf{x} + \mathbf{N}$$

where $\mathbf{x} \in \mathbb{R}^m$ is the signal of interest, $A \in \mathbb{R}^{n\times m}$ is the measurement matrix, $\mathbf{N} \in \mathbb{R}^n$ is the measurement noise, and $\mathbf{Y} \in \mathbb{R}^n$ is the noisy measurement. We assume $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I)$. Denote by k the number of nonzero entries in $\mathbf{x}$, i.e., $k = |\text{supp}(\mathbf{x})|$, where $\text{supp}(\mathbf{x})$ denotes the support set of $\mathbf{x}$. Then, $\mathbf{x}$ is said to be sparse when $k \ll m$. Given the measurement $\mathbf{Y}$ and the measurement matrix A, the goal is to reconstruct the sparse signal $\mathbf{x}$. This problem of sparse signal recovery has recently received much attention and has many applications such as compressed sensing, biomagnetic inverse problems, image processing, channel estimation, and wireless communication [1], [2], [3]. Among existing algorithms for reconstructing sparse solutions, one can broadly classify them into two categories according to their underlying principles. The first class of algorithms employ a greedy search approach and the locations of the nonzero entries in $\mathbf{x}$ are sequentially determined via a number of iterations. At each iteration, the algorithm finds the columns of A that best correlate with the current residual signal and then removes their contributions from the current residual signal to form the new residual signal for the next iteration. The algorithm terminates employing a stopping criterion such as the number of iterations or the strength of the residual signal, among others. Algorithms related to this principle include Matching Pursuit [4], Orthogonal Matching Pursuit [5], Stagewise Orthogonal Matching Pursuit [6], Subspace Pursuit [7], etc. The second class of algorithms involve solving an optimization problem with a carefully chosen cost function to which the minimizers are considered reasonable estimates of the sparse signals of interest. As opposed to the class of sequential selection algorithms above, an algorithm of the second class jointly estimates all the nonzero entries. Basis Pursuit (BP) [8], FOCUSS [9], Lasso [10], and Reweighted ℓ_1 Minimization [11] are examples of this latter type of joint recovery algorithms. Especially, with important relevance to this work, the Lasso algorithm solves for

$$\mathbf{X}_{\text{Lasso}} = \arg\min_{\tilde{\mathbf{x}}} \frac{1}{2n}\|\mathbf{Y} - A\tilde{\mathbf{x}}\|_2^2 + \lambda\|\tilde{\mathbf{x}}\|_1 \qquad (1)$$

where $\lambda \geq 0$ is the regularization parameter. Note that (1) is convex in nature and can be solved by many existing convex optimization routines [12], [13], [14].

In this work, we explore the opportunity of merging these different design principles together for performance improvement. The main contributions of the paper are summarized as follows. First, we propose the MultiPass Lasso algorithm in which the Lasso algorithm is applied in a novel, sequential manner. Specifically, at each iteration, Lasso is performed based on the current residual signal with the goal of estimating a subset of the support of the true sparse signal. Then, the current residual signal is projected onto the orthogonal complement of the subspace spanned by columns of A corresponding to all previously selected indices to form the residual signal for the next iteration. Second, we derive the Reweighted MultiPass Lasso algorithm, which essentially replaces the Lasso routine with MultiPass Lasso in Reweighted ℓ_1 Minimization [11]. Third, motivated by the techniques presented in [15], we conduct preliminary theoretic analysis to demonstrate the potential performance improvement of the MultiPass Lasso algorithm. Fourth, the experimental study indicates that Multipass Lasso and Reweighted MultiPass Lasso possess advantages over Lasso and Reweighted ℓ_1 Minimization, respectively, in both reconstruction accuracy and computational efficiency. Overall, these observations support the potential of the MultiPass framework and further suggest its usage in conjunction with other joint recovery methods for algorithmic innovations.

The rest of the paper is organized as follows. We motivate

the MultiPass Lasso algorithm in Section II-A, and develop the MultiPass Lasso algorithm in Section II-B, with observations in Section II-C and theoretic analysis in Section II-D. The Reweighted MultiPass Lasso algorithm is proposed in Section II-E. Experimental study is presented in Section III. Section IV concludes the paper.

Notations. Let $\mathbb{R}^m$ denote the m-dimensional real Euclidean space. Let $\mathbb{N}$ denote the set of positive integers. Let $[m]$ denote the set $\{1, 2, ..., m\}$. Let $|\mathcal{S}|$ denote the cardinality of set $\mathcal{S}$. For a vector $\mathbf{x}$, $\|\mathbf{x}\|_p$ denotes the ℓ_p norm of $\mathbf{x}$, i.e., $\|\mathbf{x}\|_p \triangleq (\sum_i |x_i|^p)^{1/p}$. Especially, $\|\mathbf{x}\|_\infty = \max_i |x_i|$. Let $\text{supp}(\mathbf{x}) \triangleq \{i : x_i \neq 0\}$. For a vector $\mathbf{x}$, $\mathbf{x}_\mathcal{S}$ denotes the subvector made of the elements of $\mathbf{x}$ indexed by the set $\mathcal{S}$. For a matrix A, $A_\mathcal{S}$ denotes the submatrix formed by columns of A indexed by the set $\mathcal{S}$, and $A^\dagger$ denotes the Moore-Penrose pseudoinverse of A. Let $\text{diag}(x_i)$ denote a diagonal matrix with elements of x_i as its ith diagonal element.

II. The MultiPass Lasso Algorithms

A. Motivation of the MultiPass Lasso Algorithm

The motivation of the MultiPass Lasso algorithm is many fold.

First, we are inspired by a theoretical advantage of Lasso which was recently unveiled in [15]. Basically, with other model parameters fixed, a larger regularization parameter λ increases the probability of the support of the reconstructed signal being a subset of the true support. Based on this fact, the MultiPass Lasso algorithm suitably reflects the design idea that at each iteration a partial support can be correctly reconstructed with high probability so that the union of these sequentially recovered partial supports has a good chance of being the true support.

Second, from a practical perspective, it has been observed that real signals usually do not have nonzero entries with similar magnitudes. According to the theoretic analysis for Basis Pursuit and Lasso [16], [17], they are capable of dealing well with signals with similar magnitude but they do not exploit the variation in the dynamic range of the nonzero magnitudes. Sequential selection methods perform notably better when such variation in nonzero entries exists [18]. The proposed MultiPass Lasso aims to make the best of both worlds. We model the nonzero entries as clusters with each cluster comprising of a group of nonzero entries with comparable magnitudes. The clusters are identified in a sequential manner and the nonzero entries within a group are detected jointly. Note that, at each internal iteration, existing sequential selection methods usually apply some correlation based detector [4], [5], [6], which has performance inferior to more sophisticated methods for simultaneous detection of multiple nonzero entries when the columns of A are correlated [19]. Therefore, MultiPass Lasso enables the potential of using Lasso for each internal iteration with the goal of improving the reconstruction of a partial support.

Moreover, recently a connection between sparse signal recovery and reliable communication over the Multiple Access Channel (MAC) has been studied [20]. From an information theoretic perspective, a nonzero entry can be viewed as a user and the measurement matrix as the codebook, thus bridging sparse signal recovery and multiple user detection in wireless communication. A sequential selection method for sparse signal recovery can be viewed as a Successive Interference Cancellation (SIC) scheme for multiple user detection, and methods such as Lasso as joint detection schemes [21]. When one examines the capacity region of a MAC, similar magnitudes of the nonzero entries suggest joint detection and disparate magnitudes suggest sequential detection. The development of group detectors [22] in multiple user communication motivates us to extend this design principle to the problem of sparse signal recovery.

B. The MultiPass Lasso Algorithm (MPL)

We propose the MultiPass Lasso algorithm as follows.

Step 1: Let $\mathbf{Y}^{(0)} = \mathbf{Y}$, $\mathcal{S}^{(0)} = \emptyset$, $A^{(0)} = A$, and $l = 1$.

Step 2: At the lth iteration. Choose a regularization parameter $\lambda^{(l)}$ within the range $[0, \frac{1}{n}\|A^{(l-1)\intercal}\mathbf{Y}^{(l-1)}\|_\infty)$. Solve for

$$\mathbf{X}^{(l)} = \arg\min_{\tilde{\mathbf{x}}} \ \frac{1}{2n}\|\mathbf{Y}^{(l-1)} - A^{(l-1)}\tilde{\mathbf{x}}\|_2^2 + \lambda^{(l)}\|\tilde{\mathbf{x}}\|_1.$$

Let

$$\mathcal{S}^{(l)} = \mathcal{S}^{(l-1)} \cup \text{supp}(\mathbf{X}^{(l)}), \ \mathbf{Y}^{(l)} = P^\perp_{\mathcal{S}^{(l)}}\mathbf{Y}^{(l-1)}$$
$$A^{(l)} = P^\perp_{\mathcal{S}^{(l)}}A^{(l-1)}$$

where $P^\perp_\mathcal{S} \triangleq I - A_\mathcal{S}(A_\mathcal{S}^\intercal A_\mathcal{S})^{-1}A_\mathcal{S}^\intercal$.

Step 3: If the predefined termination condition is satisfied, output the final estimate of the sparse signal as $\mathbf{X}_{\text{MPL}} = A^\dagger_{\mathcal{S}^{(l)}}\mathbf{Y}$. Otherwise, set $l \to l + 1$ and go to Step 2.

C. Observations on MPL

1) Selection of Regularization Parameter $\lambda^{(l)}$: At the lth iteration, MPL solves for a Lasso solution based on the current residual signal $\mathbf{Y}^{(l-1)}$ and the current measurement matrix $A^{(l-1)}$. According to the discussion in [23], generally, $\mathbf{X}^{(l)} \neq \mathbf{0}$ if and only if $\lambda^{(l)} \in [0, \frac{1}{n}\|A^{(l-1)\intercal}\mathbf{Y}^{(l-1)}\|_\infty)$. Meanwhile, based on the analysis in [24], as $\lambda^{(l)} \to 0_+$, in general more nonzero entries will be included in $\mathbf{X}^{(l)}$. One feasible way in implementation is to use the form $\lambda^{(l)} = \frac{\gamma}{n}\|A^{(l)\intercal}\mathbf{Y}^{(l)}\|_\infty$ with some fixed $\gamma \in (0, 1)$ for all iterations. Our experiences indicate that $\gamma \in (0.4, 0.9)$ typically gives good results. With larger γ, the solution in each iteration tends to be sparser, and generally more iterations will be carried out.

2) Termination Condition: We propose two possible criteria for algorithm termination, which are inspired by matching pursuit algorithms [4], [5], [25]. In the first criterion, we choose some $\delta > 0$ such that the algorithm stops if $\frac{1}{n}\|\mathbf{Y}^{(l)}\|_2^2 \leq \delta$. When the measurement noise is absent, i.e., $\mathbf{N} = \mathbf{0}$, δ can be chosen as a small positive quantity related to machine precision. When the measurement is noisy, δ should be chosen based on the noise variance σ^2. As the second termination criterion, a threshold $s_{\max} \in \mathbb{N}$ is set, and the algorithm terminates if $|\mathcal{S}^{(l)}| \geq s_{\max}$.

3) Update of Measurement Matrix $A^{(l)}$: The update for $A^{(l)}$ in Step 2 is also employed in Order Recursive Matching Pursuit [25] for removing the contribution from previously selected columns. Note that the columns of $A^{(l)}$ indexed by $\mathcal{S}^{(l)}$ are actually zero vectors due to the orthogonal projection

$P_{\mathcal{S}^{(l)}}^{\perp}$. For purpose of implementation, one can use the sub-matrix of $A^{(l)}$ by removing all the zero columns and properly re-indexing the remaining columns.

D. Preliminary Analysis of MPL

Recent theoretical work on Lasso [15] characterizes the probability lower bound for the event that the support of the recovered signal being a subset of the support of the true sparse signal. We can leverage the analytical techniques therein with necessary modifications to accommodate the dependencies among iterations to demonstrate the performance improvement enabled by MPL.

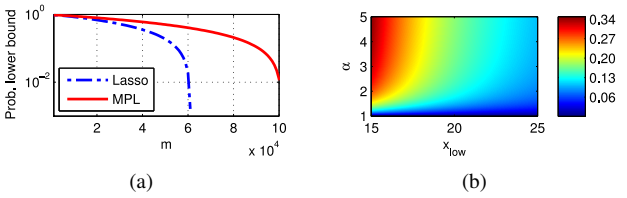
For ease of exposition, let us consider a simple but representative scenario, where the nonzero entries of a sparse signal can only take one of two possible values. Formally, let $\mathcal{S}_h, \mathcal{S}_l \subset [m]$ satisfying $\mathcal{S}_h \cap \mathcal{S}_l = \emptyset, |\mathcal{S}_h| = k_h, |\mathcal{S}_l| = k_l$. For $0 < x_{\text{low}} \leq x_{\text{high}}$, the signal $\mathbf{x}$ is given as $x_i = x_{\text{high}}$ if $i \in \mathcal{S}_h$; $x_i = x_{\text{low}}$ if $i \in \mathcal{S}_l$; and $x_i = 0$ otherwise. Suppose the elements of the measurement matrix A is independently generated according to $\mathcal{N}(0,1)$. Let

$$\lambda_1 \triangleq \frac{x_{\text{high}}}{8\sqrt{\frac{k}{n}} + \sqrt{\frac{128k\log(m-k)}{n}} + \sqrt{18} + 1}$$

$$\lambda_2 \triangleq \frac{x_{\text{low}}}{\frac{n}{n-k}\left(8\left(\frac{k_l}{n-k_h}\right)^{1/4} + \sqrt{128} + 1\right) + \sqrt{18}}.$$

Fix some $\delta > \sigma^2$. Let $\rho > 0$ be an arbitrarily small constant. For tractable analysis, we only employ the first termination criterion discussed in Section II-C-2).

For the first iteration of MPL, we choose the regularization parameter as $\lambda^{(1)} = \min(\lambda_1, \frac{1}{n}\|A^{\mathsf{T}}\mathbf{Y}\|_\infty) - \rho$. If the algorithm proceeds to the second iteration, we choose $\lambda^{(2)} = \min(\lambda_2, \frac{1}{n}\|A^{(1)\mathsf{T}}\mathbf{Y}^{(1)}\|_\infty) - \rho$. For any possible further iteration l, we pick an arbitrary $\lambda^{(l)} \in (0, \frac{1}{n}\|A^{(l-1)\mathsf{T}}\mathbf{Y}^{(l-1)}\|_\infty)$. Then, we can actually compute a lower bound for $\mathsf{P}(\text{supp}(\mathbf{X}_{\text{MPL}}) = \text{supp}(\mathbf{x}))$[1]. Due to the complex nature of this lower bound, we visualize it as well as the probability lower bound for support recovery by Lasso [15] in different ways in Fig.1 to compare the performance guarantees offered by different algorithms.



(a) (b)

Fig. 1. Comparisons of performance lower bounds for Lasso and MPL. Parameters: .
(a) $n = 800, \sigma^2 = 1, \delta = 3, x_{\text{high}} = 20, k_h = 10, x_{\text{low}} = 10, k_l = 5$.
(b) $n = 800, m = 10^5, x_{\text{high}} = \alpha x_{\text{low}}, \sigma^2 = 1, \delta = 3, k_h = 10, k_l = 5$.

First, we examine the impact of m by the simulation in Fig.1(a). Note that, for large m (i.e., more possible locations to monitor), MPL provides better performance guarantee than Lasso while holding other parameters fixed. Next, we study the impact of the dynamic range of nonzero magnitudes on the reconstruction performance in Fig.1(b). Note that each point (x_{low}, α) corresponds to the nonzero signal value pair $(x_{\text{low}}, x_{\text{high}})$ where $x_{\text{high}} = \alpha x_{\text{low}}$. The color of a point indicates the difference between the probability lower bound for MPL and that of Lasso. As we can see, given the noise level, when x_{low} is relatively small, a nontrivial distance between x_{high} and x_{low} (i.e., a large dynamic range) enables MPL to enjoy better performance guarantee.

[1]The derivation of this lower bound is presented in Appendix B.

E. The Reweighted MultiPass Lasso Algorithm (RMPL)

The MPL algorithm enables the opportunity of substituting the Lasso routine in existing algorithms for obtaining their alternative MPL versions. Inspired by Reweighted ℓ_1 Minimization [11], we develop the Reweighted MultiPass Lasso algorithm as follows.

Step 1: Set $q = 1$, $w_i^{(1)} = 1$ for $i \in [m]$. Choose $\epsilon > 0$, $q_{\max} \in \mathbb{N}$.

Step 2: At the qth iteration. Run MultiPass Lasso based on the modified measurement matrix $A \cdot \text{diag}((w_i^{(q)})^{-1})$ and the measurement $\mathbf{Y}$, and obtain the output as $\mathbf{Z}^{(q)}$. For $i \in [m]$, compute

$$x_i^{(q)} = (w_i^{(q)})^{-1} z_i^{(q)}, \quad w_i^{(q+1)} = (|x_i^{(q)}| + \epsilon)^{-1}.$$

Step 3: If $\mathbf{X}^{(q)} = \mathbf{X}^{(q-1)}$ or $q = q_{\max}$, the algorithm terminates. Otherwise, set $q \to q+1$ and go to Step 2.

III. Experimental Study

We perform experiments to empirically study the performance of the proposed MultiPass Lasso algorithms. The common experimental setup is as follows. We set $n = 100$ and $m = 256$. The measurement matrix A has elements i.i.d. according to $\mathcal{N}(0,1)$. The number of nonzero entries k increases from 9 to 57 with a step size 6, and the nonzero entries are independently drawn from $\mathcal{N}(0,1)$. Note that the Gaussian distribution for the nonzero entries is commonly employed in the literature and it leads to variety in magnitude. Both termination criteria in Section II-C-2) are employed. We claim a success in sparse signal recovery if $\|\hat{\mathbf{X}} - \mathbf{x}\|_2 / \|\mathbf{x}\|_2 \leq \tau$, where $\hat{\mathbf{X}}$ denotes the estimated sparse signal by some algorithm, and τ is a pre-defined constant. Each performance curve in this section is averaged over 200 random trials.

A. MPL: Selection of Regularization Parameter $\lambda^{(l)}$

First, we study the selection of the regularization parameter $\lambda^{(l)}$ in each internal iteration of MPL. As earlier mentioned, the form $\lambda^{(l)} = \frac{\gamma}{n}\|A^{(l)\mathsf{T}}\mathbf{Y}^{(l)}\|_\infty$ with a fixed $\gamma \in (0,1)$ for all iterations is employed. We set $\sigma = 0$, $\delta = 10^{-14}$, $s_{\max} = 0.7n$, $\tau = 10^{-3}$, and a zero entry is determined using the threshold 10^{-7}. The MPL algorithm is implemented using FPC [12]. Fig.2 illustrates the behavior of MPL with different choices of γ.

Note that as γ increases, $\lambda^{(l)}$ increases, and a partial support of the sparse signal can be detected with higher probability [15]. In Fig.2(a), for the cases with more nonzero entries, i.e., larger k, using larger λ gives slight performance improvement. Meanwhile, as $\lambda^{(l)}$ increases, in general fewer nonzero entries

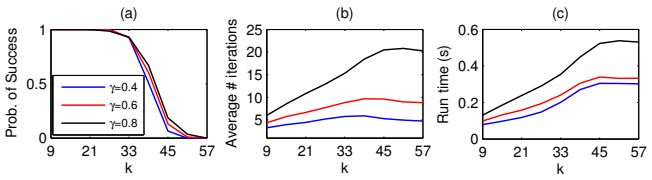
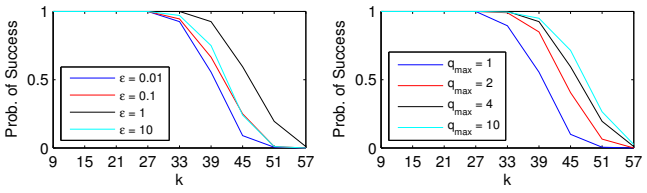


Fig. 2. The role of γ in $\lambda^{(l)}$ for MPL.

appear in the solution vector $\mathbf{X}^{(l)}$ [24]. Hence, the average number of iterations needed for each experiment increases, leading to higher computational cost. Fig.2(b), (c) agree with this analysis.

B. RMPL: Selection of ϵ and q_{max}

We explore the selection of parameters for RMPL. First, we set $q_{\max} = 4$, and focus on the effect of ϵ. For the component MPL, we choose $\gamma = 0.45$, $s_{\max} = 0.95n$. Other parameters are the same as in Section III-A. Fig.3(a) shows the result.



(a) The role of ϵ (b) The role of $q_{\max}$

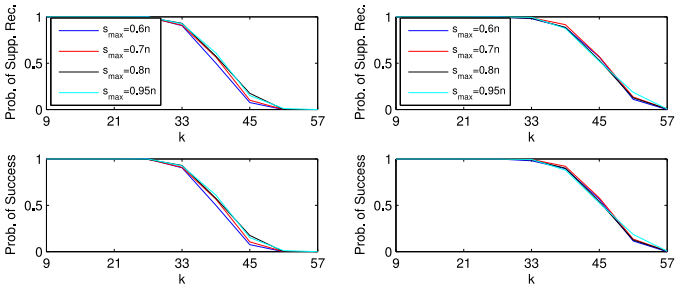
Fig. 3. Parameter selection for RMPL.

From Fig.3(a), we see that RMPL with $\epsilon = 1$ achieves the best performance. Next, let us fix $\epsilon = 1$, and study the impact of $q_{\max}$ on the performance of the algorithm. This is illustrated in Fig.3(b). Clearly, allowing more reweighted iterations helps improve the performance at the expense of computational cost.

C. MPL and RMPL: Selection of s_{max}

Next, we study the role of the parameter $s_{\max}$ for MPL and RMPL. We choose $s_{\max} = 0.6n, 0.7n, 0.8n, 0.95n$, respectively. The experiment setup is the same as in Sections III-A and III-B, except we fix $\gamma = 0.6$ for MPL, and $\epsilon = 1$ and $q_{\max} = 4$ for RMPL. The performance metrics are the success rate and the support recovery rate (using 10^{-7} to determine zero entry for all final solutions). Fig.4 summarizes the results.

From Fig.4(a), we see that the increase of $s_{\max}$ can slightly improve the performance of MPL for the cases with more nonzero entries. This is reasonable in the following sense. A large $s_{\max}$ may allow more indices to be selected, and it is more likely that $\text{supp}(\mathbf{x}) \subseteq \mathcal{S}^{(l)}$ (assuming l iterations in total). Then, in the noiseless setting, the final least squares estimation in Step 3 of MPL may set the coefficients very close to zero on the indices outside the true support. These indices, though selected in $\mathcal{S}^{(l)}$, will be judged as corresponding to zero entries via thresholding. Next, based on Fig.4(b), we can see that the performance of RMPL with different $s_{\max}$ are very similar, due to the fact that the multiple runs of MPL in RMPL has made the selection of $s_{\max}$ less important. In summary, this set



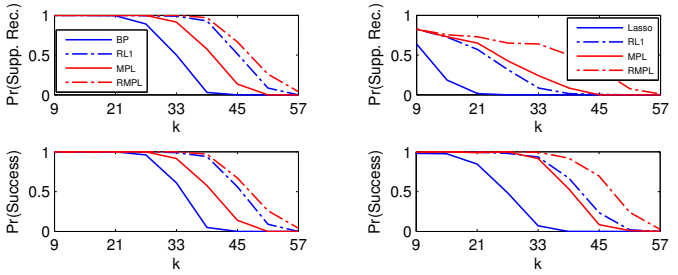
(a) MPL (b) RMPL

Fig. 4. The role of $s_{\max}$ for MPL and RMPL. In each case, the upper plot shows support recovery rate, whereas the lower plot shows success rate.

of experiments indicates that the performances of MPL and RMPL, in terms of support recovery and estimation accuracy, are relatively insensitive to the selection of $s_{\max}$.

D. Comparison with BP, Lasso, and Reweighted ℓ_1 Minimization

First, we consider the noiseless scenario, i.e., $\sigma = 0$. We compare the performance between BP and MPL, and between Reweighted ℓ_1 Minimization (RL1) and RMPL. In this case, BP and RL1 are implemented via ℓ_1-MAGIC [26]. For MPL, $\gamma = 0.6$. For both RL1 and RMPL, $\epsilon = 1$ and $q_{\max} = 10$. Other parameters are chosen the same way as in Sections III-A and III-B. Fig.5(a) summarizes the results. It can be seen that MPL and RMPL outperform BP and RL1, respectively, in terms of both rate of success and rate of support recovery (using 10^{-5} to determine zero entry for all final solutions).



(a) Noiseless case (b) Noisy case

Fig. 5. Sparse signal recovery in different settings. In each case, the upper plot shows support recovery rate, whereas the lower plot shows success rate.

Next, we consider the noisy setting with $\sigma = 0.01$. We compare among Lasso, RL1, MPL, and RMPL. All algorithms are implemented using the same Lasso routine [12]. For Lasso, we choose $\lambda = \sigma\sqrt{(2\log m)/n}$ as suggested in [8], [15]. This choice of λ is also employed by RL1. For MPL and RMPL, $\delta = 2.25\sigma^2$, and we use 10^{-4} as the threshold for determining zero entries. Meanwhile, $\tau = 10^{-2}$. Other parameters remain the same as in the noiseless case above. Fig.5(b) shows the result. We can see that MPL outperforms Lasso, and RMPL gives better performance than RL1 in both criteria (using 10^{-3} for determining zero entry in all final solutions). This

set of experiments supports the design goal of MPL from which better support recovery is expected. The effectiveness of the proposed MPL and RMPL in various settings are also illustrated.

E. Computational Efficiency of MPL and RMPL

We examine the computational efficiency of MPL and RMPL with comparison to Lasso and RL1, respectively, in the noisy setting. The parameters are chosen the same as in the noisy case in Section III-D, except that for RL1 and RMPL, we set $q_{\max} = 4$.

Three different Lasso solvers are employed, namely the Fixed-Point Continuation method (FPC) [12], the Truncated Newton Interior-Point method with Preconditioned Conjugate Gradients (L1LS) [13], and the Basic Gradient Projection method (GPSR) [14]. With each of the three Lasso solvers, we implement Lasso, RL1, MPL, and RMPL with the same initialization, termination condition, and precision control to ensure fair comparison. The computer in use runs MATLAB R2010a in Windows XP environment. Fig.6 summarizes the results.

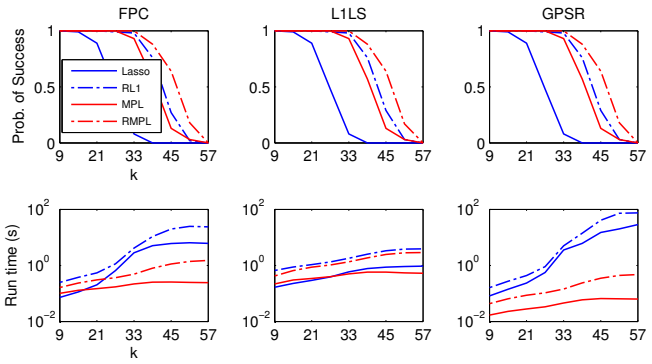


Fig. 6. Comparison of algorithm efficiency. Each column presents the performance and the corresponding run time of algorithms implemented via the same Lasso solver specified in the title. The upper row indicates the empirical probability of success, and the lower row illustrates the average time consumed per experiment.

According to Fig.6, the empirical probabilities of success for each algorithm using different Lasso solvers are almost identical. Most implementations of MPL perform faster than Lasso for a large range of k. Especially, MPL with GPSR implementation is faster than Lasso for the whole range of k tested here, and it is faster than other implementations of Lasso as well. Meanwhile, with each optimization solver, RMPL achieves better performance and lower computational cost than RL1 built upon the same solver. The GPSR version of RMPL achieves the lowest computational cost among the tested cases. Overall, MPL and RMPL achieve both better performance and lower computational cost than their Lasso counterparts, respectively, suggesting their potential as effective and efficient algorithmic choices for sparse signal recovery.

IV. Conclusion

We proposed the MultiPass Lasso algorithm and the Reweighted MultiPass algorithm for sparse signal recovery. The results are promising and compare favorably with their one-pass, Lasso counterparts. This MultiPass framework is quite general and can accommodate a plethora of options in the various stages. This can be a fertile ground for future algorithmic work and analysis.

Acknowledgment

This research was supported by NSF Grant CCF-0830612.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
[2] E. J. Candes, "Compressive sampling," *Proceedings of the Int. Congress of Mathematicians*, pp. 1433–1452, 2006.
[3] S. F. Cotter and B. D. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *IEEE Trans. on Communications*, vol. 50, pp. 374–377, 2002.
[4] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Sig. Proc.*, vol. 41, no. 12, pp. 3397–3415, 1993.
[5] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," *27th Annual Asilomar Conf. Sig. Sys. Comp.*, 1993.
[6] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *preprint*, 2006.
[7] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: Closing the gap between performance and complexity," *preprint*, 2008.
[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIREV*, vol. 43, no. 1, 2001.
[9] I. Gorodnitsky and B. Rao, "Sparse signal reconstruction from limited data using focuss: A re-weighted norm minimization algorithm," *IEEE Trans. Sig. Proc.*, vol. 45, no. 3, 1997.
[10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.
[11] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J Fourier Anal Appl.*, vol. 14, pp. 877–905, 2008.
[12] E. T. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing," *TR07-07, available at http://www.caam.rice.edu/ optimization/L1/fpc/.*
[13] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l1-regularized least squares," *IEEE JSTSP*, vol. 1, pp. 606–617, 2007.
[14] M. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE JSTSP*, vol. 1, pp. 586–597, 2007.
[15] M. Wainwright, "Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso)," *IEEE Trans. IT*, vol. 55, no. 5, 2009.
[16] D. M. Malioutov, M. Çetin, and A. S. Willsky, "Optimal sparse representations in general overcomplete bases," *ICASSP*, 2004.
[17] E. Candes, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. Paris*, 2008.
[18] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On-off random access channels: A compressed sensing framework," *submitted to IEEE Trans. Information Theory*.
[19] S. Verdu, *Multiuser Detection*. Cambridge university Press, 1998.
[20] Y. Jin and B. D. Rao, "Insights into the stable recovery of sparse solutions in overcomplete representations using network information theory," *ICASSP*, 2008.
[21] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
[22] F. Hasegawa, J. Luo, K. R. Pattipati, P. Willett, and D. Pham, "Speed and accuracy comparison of techniques for multiuser detection in synchronous cdma," *IEEEE Trans. Comm.*, vol. 52, pp. 540–545, 2004.
[23] M. R. Osborne, B. Presnell, and B. A. Turlach, "On the lasso and its dual," *Journal of Computational and Graphical Statistics*, vol. 9, pp. 319–337, 1999.
[24] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 1976.
[25] S. Cotter, J. Adler, B. D. Rao, and K. Kreutz-Delgado, "Forward sequential algorithms for best basis selection," *IEE Proceedings on Vision, Image and Signal Processing*, vol. 146, no. 5, pp. 235–244, 1999.
[26] E. J. Candes and J. Romberg, "ℓ1-magic : Recovery of sparse signals via convex programming," *Technical Report*, 2005.
[27] Y. Jin, Y.-H. Kim, and B. D. Rao, "Support recovery of sparse signals," *arXiv:1003.0888v1[cs.IT]*, 2010.

APPENDIX A
PROBABILITY LOWER BOUND FOR LASSO

This appendix presents the probability lower bound for Lasso on support recovery, which is employed to produce the figures in Section II-D. This lower bound for Lasso is based on Theorem 3 in [15], with slight modification to accommodate flexible settings, to offer a probabilistic characterization of the support recovery performance. Define the auxiliary function for notation simplicity, for $\eta \in (0, \frac{1}{2})$, $\lambda > 0$,

$$g(\eta, \lambda) \triangleq \left(1 + \max(\eta, 8\sqrt{k/n})\right)\left(k/n + \sigma^2(n-k)/\lambda^2 n^2\right).$$

Theorem 1 (Modified based on Theorem 3 of [15])**.** *Suppose the elements of the measurement matrix A is independently generated according to $\mathcal{N}(0,1)$. Choose $\lambda > 0$. Let*

$$\begin{aligned}\mathcal{E}_1 &\triangleq \{supp(\mathbf{X}_{Lasso}) \subseteq supp(\mathbf{x})\}\\ \mathcal{E}_2 &\triangleq \{\|\mathbf{X}_{Lasso} - \mathbf{x}\|_\infty \\ &\qquad \le \lambda(8\sqrt{k/n} + \sqrt{128k\log(m-k)/n} + \sqrt{18} + 1)\}.\end{aligned}$$

Then, for $\eta \in (0, \frac{1}{2})$,

$$\begin{aligned}\mathsf{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \ge\ & 1 - 2(m-k)\exp\left(-\frac{1}{2g(\eta,\lambda)}\right)\\ & - 2\exp\left(-\frac{3(n-k)\eta^2}{16}\right) - 4\exp\left(-\frac{k}{2}\right)\\ & - 2k\exp\left(-\log(m-k)\right) - 2k\exp\left(-\frac{n\lambda^2}{\sigma^2}\right)\\ & - 2\exp\left(-\frac{n}{2}\right). \qquad (2)\end{aligned}$$

Proof Sketch: The proof follows Section V of [15] with mainly the following modification. We choose $t = \sqrt{18}\lambda$ (as opposed to $t = 20\sqrt{\sigma^2 \log k/n}$ in the reference) to reach a bound similar to (42) of [15]. Thus, (2) can be reached via straightforward derivation parallel to [15] with minor clarifications on some terms. ∎

Remark 1. This theorem provides a lower bound, for a given choice of λ, on the probability that the reconstructed support being contained in the true support and the elementwise maximum distance of the reconstructed signal to the true signal being upper bounded. For the scenario defined in Section II-D, successful support recovery is implied by $\mathcal{E}_1 \cap \mathcal{E}_2$ if we require

$$\lambda(8\sqrt{k/n} + \sqrt{128k\log(m-k)/n} + \sqrt{18} + 1) < x_{\text{low}}.$$

This relation naturally gives the selection of λ for Lasso to maximize this probability lower bound for support recovery.

Remark 2. Note that Theorem 3 in [15] mainly focuses on the relations among model parameters for successful signal reconstruction in the asymptotic sense. It states that $\mathsf{P}(\mathcal{E}_1 \cap \mathcal{E}_2) \ge 1 - c_1\exp(-c_2\min(k, \log(m-k)))$ for some constants $c_1, c_2 > 0$ with certain family of λ and implicitly under the scaling $n = \Omega(k\log(m-k))$. In contrast, our modification (2) emphasizes the roles of n, m, k in finite settings. Meanwhile, it can afford a more flexible upper bound for $\|\mathbf{X}_{\text{Lasso}} - \mathbf{x}\|_\infty$, in contrast to the original bound (33) of [15] which may not be smaller than $20\sqrt{\sigma^2 \log k/n}$. Further, we explicitly work out all the constants to facilitate further comparisons.

The internal free parameter is chosen as $\eta = 0.49$ for plotting the figures in Section II-D.

APPENDIX B
DERIVATION OF PROBABILITY LOWER BOUND FOR MPL

In this appendix, we derive the probability lower bound for successful support recovery using MultiPass Lasso in the scenario defined in Section II-D. This lower bound is then employed to produce the figures in that section.

Define $\mathcal{S} \triangleq \text{supp}(\mathbf{x}) = \mathcal{S}_h \cup \mathcal{S}_l$, and the events

$$\begin{aligned}\mathcal{G} \triangleq\ & \left\{\text{supp}(\mathbf{X}^{(1)}) \subseteq \mathcal{S}\right\} \cap \left\{\mathcal{S}_h \subseteq \text{supp}(\mathbf{X}^{(1)})\right\}\\ & \cap \left\{\lambda_1 < \frac{1}{n}\|A^\intercal \mathbf{Y}\|_\infty\right\} \cap \left\{\frac{1}{n}\|\mathbf{N}\|_2^2 \ \le \delta\right\}\\ \mathcal{T} \triangleq\ & \left\{\text{supp}(\mathbf{X}^{(2)}) \subseteq (\mathcal{S} \backslash \text{supp}(\mathbf{X}^{(1)}))\right\}\\ & \cap \left\{(\mathcal{S}_l \backslash \text{supp}(\mathbf{X}^{(1)})) \subseteq \text{supp}(\mathbf{X}^{(2)})\right\}\\ & \cap \left\{\lambda_2 < \frac{1}{n}\|A^{(1)\intercal}\mathbf{Y}^{(1)}\|_\infty\right\} \cap \left\{\frac{1}{n}\|\mathbf{Y}^{(1)}\|_2^2 > \delta\right\}.\end{aligned}$$

According to the total probability rule,

$$\begin{aligned}&\mathsf{P}(\text{supp}(\mathbf{X}_{\text{MPL}}) = \mathcal{S})\\ &\ge \mathsf{P}\left(\text{supp}(\mathbf{X}_{\text{MPL}}) = \mathcal{S}\middle|\mathcal{G}\right)\mathsf{P}\left(\mathcal{G}\right)\\ &\ge \Big(\mathsf{P}(\text{supp}(\mathbf{X}_{\text{MPL}}) = \mathcal{S}, \text{supp}(\mathbf{X}^{(1)}) \subset \mathcal{S}|\mathcal{G})\\ &\quad + \mathsf{P}(\text{supp}(\mathbf{X}_{\text{MPL}}) = \mathcal{S}, \text{supp}(\mathbf{X}^{(1)}) = \mathcal{S}|\mathcal{G})\Big)\mathsf{P}(\mathcal{G})\\ &= \Big(\mathsf{P}(\text{supp}(\mathbf{X}_{\text{MPL}}) = \mathcal{S}, \text{supp}(\mathbf{X}^{(1)}) \subset \mathcal{S}|\mathcal{G})\\ &\quad + \mathsf{P}(\text{supp}(\mathbf{X}^{(1)}) = \mathcal{S}|\mathcal{G})\Big)\mathsf{P}(\mathcal{G}) \qquad (3)\\ &\ge \Big(\mathsf{P}(\text{supp}(\mathbf{X}_{\text{MPL}}) = \mathcal{S}, \text{supp}(\mathbf{X}^{(1)}) \subset \mathcal{S}, \mathcal{T}|\mathcal{G})\\ &\quad + \mathsf{P}(\text{supp}(\mathbf{X}^{(1)}) = \mathcal{S}|\mathcal{G})\Big)\mathsf{P}(\mathcal{G})\\ &\ge \Big(\mathsf{P}\left(\text{supp}(\mathbf{X}_{\text{MPL}}) = \mathcal{S}|\mathcal{T}, \text{supp}(\mathbf{X}^{(1)}) \subset \mathcal{S}, \mathcal{G}\right)\\ &\quad \cdot \mathsf{P}(\mathcal{T}, \text{supp}(\mathbf{X}^{(1)}) \subset \mathcal{S}\big|\mathcal{G}) + \mathsf{P}(\text{supp}(\mathbf{X}^{(1)}) = \mathcal{S}|\mathcal{G})\Big)\mathsf{P}(\mathcal{G})\\ &= \Big(\mathsf{P}(\mathcal{T}, \text{supp}(\mathbf{X}^{(1)}) \subset \mathcal{S}\big|\mathcal{G}) + \mathsf{P}(\text{supp}(\mathbf{X}^{(1)}) = \mathcal{S}|\mathcal{G})\Big)\mathsf{P}(\mathcal{G})\\ &\qquad (4)\end{aligned}$$

where (3) follows from

$$\mathsf{P}(\text{supp}(\mathbf{X}_{\text{MPL}}) = \mathcal{S}|\text{supp}(\mathbf{X}^{(1)}) = \mathcal{S}, \mathcal{G}) = 1$$

and (4) follows from

$$\mathsf{P}(\text{supp}(\mathbf{X}_{\text{MPL}}) = \mathcal{S}|\mathcal{T}, \text{supp}(\mathbf{X}^{(1)}) \subset \mathcal{S}, \mathcal{G}) = 1.$$

Let $\mathbf{1}$ denote a vector of all ones with a proper length. Define the following auxiliary random variable and events, for any

fixed set $\mathcal{V} \subset \mathcal{S}$ satisfying $\mathcal{S}_h \subseteq \mathcal{V}$,

$$\begin{aligned}
\mathbf{Q}_{\mathcal{V}} &\triangleq P_{\mathcal{V}}^{\perp}(x_{\text{low}} A_{\mathcal{S}\setminus\mathcal{V}}\mathbf{1} + \mathbf{N}) \\
\mathcal{F}_{\mathcal{V}} &\triangleq \left\{ \lambda_2 < \frac{1}{n}\|(P_{\mathcal{V}}^{\perp}A)^{\mathsf{T}}\mathbf{Q}_{\mathcal{V}}\|_{\infty} \right\} \\
\mathcal{H}_{\mathcal{V}} &\triangleq \left\{ \frac{1}{n}\|\mathbf{Q}_{\mathcal{V}}\|_2^2 > \delta \right\} \\
\mathcal{J}_{\mathcal{V}} &\triangleq \left\{ \text{supp}(\mathbf{X}^{(1)}) = \mathcal{V} \right\}.
\end{aligned}$$

Thus, using the fact that $\mathsf{P}(\mathcal{A}\cap\mathcal{B}) \geq \mathsf{P}(\mathcal{A}) - \mathsf{P}(\mathcal{B}^c)$, one can have

(please see the long equations (5) on the next page,)

where $\text{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}}))$ indicates that the reconstruction is based on $\mathbf{Q}_{\mathcal{V}}$, which is a possible form of $\mathbf{Y}^{(1)}$. Due to the facts that $\sum_{\mathcal{V}:\mathcal{V}\subset\mathcal{S},\mathcal{S}_h\subseteq\mathcal{V}} \mathsf{P}(\mathcal{J}_{\mathcal{V}}|\mathcal{G}) + \mathsf{P}(\text{supp}(\mathbf{X}^{(1)}) = \mathcal{S}|\mathcal{G}) = 1$, and that $\mathsf{P}((\mathcal{A}\cap\mathcal{B})^c) \leq \mathsf{P}(\mathcal{A}^c) + \mathsf{P}(\mathcal{B}^c)$, we have

(please see the long equations (6) on the next page.)

It remains to work out proper bounds for the terms in (6), respectively. First, we consider $\mathsf{P}(\mathcal{G})$, which concerns the behavior in the first iteration of the MultiPass Lasso algorithm. Note that

$$\begin{aligned}
\mathsf{P}(\mathcal{G}) \geq 1 &- \mathsf{P}\left(\|\mathbf{N}\|_2^2/n \ > \delta\right) - \mathsf{P}\left(\lambda_1 \geq \|A^{\mathsf{T}}\mathbf{Y}\|_{\infty}/n\right) \\
&- \mathsf{P}((\text{supp}(\mathbf{X}^{(1)}) \subseteq \mathcal{S}, \mathcal{S}_h \subseteq \text{supp}(\mathbf{X}^{(1)}))^c). \qquad (7)
\end{aligned}$$

First, using the Chernoff bound for χ^2 random variables as in Lemma 1 of [27], we have, for $\delta > \sigma^2$,

$$\mathsf{P}\left(\|\mathbf{N}\|_2^2/n > \delta\right) \leq \exp\left(-n\cdot h(\sigma^2,\delta)\right). \qquad (8)$$

where $h(\alpha,\beta) \triangleq \frac{\beta}{2\alpha} - \frac{1}{2} + \frac{1}{2}\log\frac{\alpha}{\beta}$.

Next, we choose an arbitrary $i \in \mathcal{S}_h$. Then, for any η_1, η_2 such that $(1+\sqrt{18})^{-1} < \eta_1 < 1, \eta_2 > 1$,

$$\begin{aligned}
&\mathsf{P}\left(\lambda_1 \geq \|A^{\mathsf{T}}\mathbf{Y}\|_{\infty}/n\right) \\
&\leq \mathsf{P}\left(\lambda_1 \geq \|A^{\mathsf{T}}\mathbf{Y}\|_{\infty}/n \middle| \eta_1 < \|\mathbf{A}_i\|_2^2/n < \eta_2\right) \\
&\quad + \mathsf{P}\left(\|\mathbf{A}_i\|_2^2/n \leq \eta_1\right) + \mathsf{P}\left(\|\mathbf{A}_i\|_2^2/n \geq \eta_2\right).
\end{aligned}$$

where $\mathbf{A}_i$ denotes the ith column of A. It can be readily seen that

$$\begin{aligned}
\mathsf{P}(\|\mathbf{A}_i\|_2^2/n \leq \eta_1) &\leq \exp\left(-n\cdot h(1,\eta_1)\right) \\
\mathsf{P}(\|\mathbf{A}_i\|_2^2/n \geq \eta_2) &\leq \exp\left(-n\cdot h(1,\eta_2)\right).
\end{aligned}$$

Due to the fact that

$$\begin{aligned}
&\frac{1}{n}\|A^{\mathsf{T}}\mathbf{Y}\|_{\infty} \\
&\geq \left| \frac{x_{\text{high}}}{n}\|\mathbf{A}_i\|_2^2 + \frac{1}{n}\mathbf{A}_i^{\mathsf{T}}(x_{\text{high}} A_{\mathcal{S}_h\setminus\{i\}}\mathbf{1} + x_{\text{low}} A_{\mathcal{S}_l}\mathbf{1} + \mathbf{N}) \right|
\end{aligned}$$

and the independence among $\mathbf{A}_i$, $A_{\mathcal{S}\setminus\{i\}}$ and $\mathbf{N}$, we employ the tail bound [15, Appendix-A] to obtain

$$\begin{aligned}
&\mathsf{P}\left(\lambda_1 \geq \frac{1}{n}\|A^{\mathsf{T}}\mathbf{Y}\|_{\infty} \middle| \mathbf{A}_i, \eta_1 < \|\mathbf{A}_i\|_2^2/n < \eta_2\right) \\
&\leq \exp\left(-\frac{n(\lambda_1 - x_{\text{high}}\eta_1)^2}{2(x_{\text{high}}^2(k_h-1) + x_{\text{low}}^2 k_l + \sigma^2)\eta_2}\right).
\end{aligned}$$

Putting pieces together, one can have

$$\begin{aligned}
&\mathsf{P}\left(\lambda_1 \geq \frac{1}{n}\|A^{\mathsf{T}}\mathbf{Y}\|_{\infty}\right) \\
&\leq \exp\left(-\frac{n(\lambda_1 - x_{\text{high}}\eta_1)^2}{2(x_{\text{high}}^2(k_h-1) + x_{\text{low}}^2 k_l + \sigma^2)\eta_2}\right) \\
&\quad + \exp\left(-n\cdot h(1,\eta_1)\right) + \exp\left(-n\cdot h(1,\eta_2)\right). \qquad (9)
\end{aligned}$$

Next, note that, for any μ satisfying $0 < \mu < x_{\text{high}}$, the event $\{\text{supp}(\mathbf{X}^{(1)}) \subseteq \mathcal{S}\} \cap \{\|\mathbf{X}^{(1)} - \mathbf{x}\|_{\infty} \leq \mu\}$ implies the event $\{\text{supp}(\mathbf{X}^{(1)}) \subseteq \mathcal{S}\} \cap \{\mathcal{S}_h \subseteq \text{supp}(\mathbf{X}^{(1)})\}$. This observation leads to

$$\begin{aligned}
&\mathsf{P}((\text{supp}(\mathbf{X}^{(1)}) \subseteq \mathcal{S}, \mathcal{S}_h \subseteq \text{supp}(\mathbf{X}^{(1)}))^c) \\
&\leq \mathsf{P}((\text{supp}(\mathbf{X}^{(1)}) \subseteq \mathcal{S}, \|\mathbf{X}^{(1)} - \mathbf{x}\|_{\infty} < x_{\text{high}})^c) \\
&\leq \mathsf{P}((\text{supp}(\mathbf{X}^{(1)}) \nsubseteq \mathcal{S}) + \mathsf{P}(\|\mathbf{X}^{(1)} - \mathbf{x}\|_{\infty} \geq x_{\text{high}}). \qquad (10)
\end{aligned}$$

We recognize that the probability bound above belongs to the scenario considered in Theorem 1 with the choice $\lambda = \lambda_1 - \rho$. Therefore, $\mathsf{P}(\text{supp}(\mathbf{X}^{(1)}) \subseteq \mathcal{S}, \mathcal{S}_h \subseteq \text{supp}(\mathbf{X}^{(1)}))$ can be lower bounded by the right hand side of (2) with the choice $\lambda = \lambda_1 - \rho$. Thus far, we obtained a lower bound for $\mathsf{P}(\mathcal{G})$.

Next, we consider the remaining terms in the brackets of (6). From an analytical perspective, these terms seem similar to the terms in (7), except that now the effect of the orthogonal projection $P_{\mathcal{V}}^{\perp}$ after the first iteration must be considered accordingly. Here, we demonstrate a useful technique to address this issue by working out one remaining term in (6) as an example. Let us focus on $\mathsf{P}(\mathcal{H}_{\mathcal{V}}^c)$. Let $|\mathcal{V}| = v$. First, we condition the analysis on $A_{\mathcal{V}}$, which implies the conditioning on $P_{\mathcal{V}}^{\perp}$. It is easy to verify that $P_{\mathcal{V}}^{\perp}P_{\mathcal{V}}^{\perp} = P_{\mathcal{V}}^{\perp}$ and $P_{\mathcal{V}}^{\perp\mathsf{T}} = P_{\mathcal{V}}^{\perp}$. One can decompose $P_{\mathcal{V}}^{\perp} = U^{\mathsf{T}}U$ where $U \in \mathbb{R}^{(n-v)\times n}$ and $UU^{\mathsf{T}} = I$. With these at hand,

$$\begin{aligned}
\frac{1}{n}\|\mathbf{Q}_{\mathcal{V}}\|_2^2 &= \frac{1}{n}\|P_{\mathcal{V}}^{\perp}(x_{\text{low}} A_{\mathcal{S}\setminus\mathcal{V}}\mathbf{1} + \mathbf{N})\|_2^2 \\
&= \frac{1}{n}\|U(x_{\text{low}} A_{\mathcal{S}\setminus\mathcal{V}}\mathbf{1} + \mathbf{N})\|_2^2 \qquad (11)
\end{aligned}$$

where $U(x_{\text{low}} A_{\mathcal{S}\setminus\mathcal{V}}\mathbf{1} + \mathbf{N}) \in \mathbb{R}^{n-v}$ and $U(x_{\text{low}} A_{\mathcal{S}\setminus\mathcal{V}}\mathbf{1} + \mathbf{N}) \sim \mathcal{N}(\mathbf{0}, ((k-v)x_{\text{low}}^2 + \sigma^2)I)$. Hence,

$$\begin{aligned}
&\mathsf{P}(\mathcal{H}_{\mathcal{V}}^c | A_{\mathcal{V}}) \\
&= \mathsf{P}\left(\frac{1}{n}\|\mathbf{Q}_{\mathcal{V}}\|_2^2 \leq \delta \middle| A_{\mathcal{V}}\right) \\
&= \mathsf{P}\left(\frac{1}{n-v}\|U(x_{\text{low}} A_{\mathcal{S}\setminus\mathcal{V}}\mathbf{1} + \mathbf{N})\|_2^2 \leq \frac{n}{n-v}\delta \middle| A_{\mathcal{V}}\right) \\
&\leq \exp\left(-(n-v)\cdot h((k-v)x_{\text{low}}^2 + \sigma^2, n\delta/(n-v))\right). \qquad (12)
\end{aligned}$$

Note that (12) is independent of $A_{\mathcal{V}}$ (or, $P_{\mathcal{V}}^{\perp}$). Therefore, $\mathsf{P}(\mathcal{H}_{\mathcal{V}}^c)$ can be also upper bounded by (12).

Incorporating the techniques above for dealing with $P_{\mathcal{V}}^{\perp}$ into similar steps we performed for lower bounding $\mathsf{P}(\mathcal{G})$, one should be able to work out the following results via

$$
\begin{aligned}
&\mathsf{P}(\mathcal{T}, \mathrm{supp}(\mathbf{X}^{(1)}) \subset \mathcal{S} \big| \mathcal{G}) \\
&= \mathsf{P}(\mathcal{T}, \mathrm{supp}(\mathbf{X}^{(1)}) \subset \mathcal{S}, \mathcal{G}) / \mathsf{P}(\mathcal{G}) \\
&= \sum_{\mathcal{V}: \mathcal{V} \subset \mathcal{S}, \mathcal{S}_h \subseteq \mathcal{V}} \frac{\mathsf{P}(\mathcal{T}, \mathcal{J}_{\mathcal{V}}, \mathcal{G})}{\mathsf{P}(\mathcal{G})} \\
&= \sum_{\mathcal{V}: \mathcal{V} \subset \mathcal{S}, \mathcal{S}_h \subseteq \mathcal{V}} \frac{\mathsf{P}(\mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})) \subseteq \mathcal{S} \backslash \mathcal{V}, \mathcal{S}_l \backslash \mathcal{V} \subseteq \mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})), \mathcal{F}_{\mathcal{V}}, \mathcal{H}_{\mathcal{V}}, \mathcal{J}_{\mathcal{V}}, \mathcal{G})}{\mathsf{P}(\mathcal{G})} \\
&\geq \sum_{\mathcal{V}: \mathcal{V} \subset \mathcal{S}, \mathcal{S}_h \subseteq \mathcal{V}} \frac{\mathsf{P}(\mathcal{J}_{\mathcal{V}}, \mathcal{G})) - \mathsf{P}((\mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})) \subseteq \mathcal{S} \backslash \mathcal{V}, \mathcal{S}_l \backslash \mathcal{V} \subseteq \mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})), \mathcal{F}_{\mathcal{V}}, \mathcal{H}_{\mathcal{V}})^c)}{\mathsf{P}(\mathcal{G})} \\
&= \sum_{\mathcal{V}: \mathcal{V} \subset \mathcal{S}, \mathcal{S}_h \subseteq \mathcal{V}} \mathsf{P}(\mathcal{J}_{\mathcal{V}} | \mathcal{G}) - \frac{\mathsf{P}((\mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})) \subseteq \mathcal{S} \backslash \mathcal{V}, \mathcal{S}_l \backslash \mathcal{V} \subseteq \mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})), \mathcal{F}_{\mathcal{V}}, \mathcal{H}_{\mathcal{V}})^c)}{\mathsf{P}(\mathcal{G})}
\end{aligned} \tag{5}
$$

$$
\begin{aligned}
&\mathsf{P}(\mathrm{supp}(\mathbf{X}_{\mathrm{MPL}}) = \mathcal{S}) \\
&\geq \left(1 - \sum_{\mathcal{V}: \mathcal{V} \subset \mathcal{S}, \mathcal{S}_h \subseteq \mathcal{V}} \frac{\mathsf{P}\left((\mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})) \subseteq \mathcal{S} \backslash \mathcal{V}, \mathcal{S}_l \backslash \mathcal{V} \subseteq \mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})), \mathcal{F}_{\mathcal{V}}, \mathcal{H}_{\mathcal{V}})^c\right)}{\mathsf{P}(\mathcal{G})}\right) \mathsf{P}(\mathcal{G}) \\
&\geq \mathsf{P}(\mathcal{G}) - \sum_{\mathcal{V}: \mathcal{V} \subset \mathcal{S}, \mathcal{S}_h \subseteq \mathcal{V}} \left[\mathsf{P}\left((\mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})) \subseteq \mathcal{S} \backslash \mathcal{V}, \mathcal{S}_l \backslash \mathcal{V} \subseteq \mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})))^c\right) + \mathsf{P}(\mathcal{F}_{\mathcal{V}}^c) + \mathsf{P}(\mathcal{H}_{\mathcal{V}}^c)\right]
\end{aligned} \tag{6}
$$

straightforward derivations,

$$
\begin{aligned}
\mathsf{P}(\mathcal{F}_{\mathcal{V}}^c) \leq \exp&\left(-\frac{n(\lambda_2 - x_{\mathrm{low}}\eta_1)^2}{2(x_{\mathrm{low}}^2(k-v-1)+\sigma^2)\eta_2}\right) \\
&+ \exp\Big(-(n-v)\cdot h\Big(1, n\eta_1/(n-v)\Big)\Big) \\
&+ \exp\big(-(n-v)\cdot h(1, n\eta_2/(n-v))\big).
\end{aligned} \tag{13}
$$

Next, slightly different approach should be taken to analyze $\mathsf{P}((\mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})) \subseteq \mathcal{S}\backslash\mathcal{V}, \mathcal{S}_l\backslash\mathcal{V} \subseteq \mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})))^c)$. The reason is that the number of remaining nonzero entries to recover ranges from 1 to k_l, whereas the probability lower bound (2) scales as $1 - \exp(-\min(k, \log(m-k)))$. We need to modify the derivation in order to provide meaningful results for the worst case $v = k-1$ in the summation in (6). Therefore, we choose $t = (k/n)^{1/4}$ when applying Lemma 9 of [15] (as opposed to $t = \sqrt{k/n}$). Further, in Appendix G of [15], we instead work with the probability $\mathsf{P}(|||D|||_2 \geq 8(k/n)^{1/4}) \leq 2\exp(-\sqrt{nk}/2)$, later condition on $\{|||D|||_2 < 8(k/n)^{1/4}\}$, and continue the proof therein with the choice $t = \sqrt{128}$ to reach a bound similar to the last inequality of that section. It can be verified that $(\lambda_2 - \rho)(\frac{n}{n-v}(8(\frac{k-v}{n-v})^{1/4} + \sqrt{128} + 1) + \sqrt{18}) < x_{\mathrm{low}}$ for $k_h \leq v < k$. With these modifications, one should be able to obtain that, for $\eta \in (0, \frac{1}{2})$,

$$
\begin{aligned}
&\mathsf{P}\left((\mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})) \subseteq \mathcal{S}\backslash\mathcal{V}, \mathcal{S}_l\backslash\mathcal{V} \subseteq \mathrm{supp}(\mathbf{X}^{(2)}(\mathbf{Q}_{\mathcal{V}})))^c\right) \\
&\leq 2(m-k)\exp\Big(-\frac{1}{2g_1(v,\eta,\lambda_2-\rho)}\Big) \\
&\quad + 2\exp\Big(-\frac{3(n-k)\eta^2}{16}\Big) + 4\exp\Big(-\frac{\sqrt{(n-v)(k-v)}}{2}\Big) \\
&\quad + 2(k-v)\exp\Big(-\sqrt{\frac{n-v}{k-v}}\Big) \\
&\quad + 2(k-v)\exp\Big(-\frac{(n-v)(\lambda_2-\rho)^2}{\sigma^2}\Big) \\
&\quad + 2\exp\Big(-\frac{n-v}{2}\Big)
\end{aligned} \tag{14}
$$

where $g_1(v,\eta,\lambda) \triangleq (1+\max(\eta, 8(\frac{k-v}{n-v})^{1/4}))(\frac{k-v}{n-v} + \frac{\sigma^2(n-k)}{\lambda^2 n^2})$.

Finally, $\mathsf{P}(\mathrm{supp}(\mathbf{X}_{\mathrm{MPL}}) = \mathcal{S})$ can be lower bounded by substituting (7) [with (8), (9), and (10) by replacing $\lambda = \lambda_1 - \rho$ in (2)], (12), (13), (14) into (6). For figures shown in Section II-D, we choose the internal free parameters as $\eta_1 = 0.8, \eta_2 = 1.7, \eta = 0.49$.