

Error-Correction for Sparse Support Recovery Algorithms

Mohammad Mehrabi and Aslan Tchamkerten

Abstract

Consider the compressed sensing setup where the support \mathbf{s}^* of an m -sparse d -dimensional signal \mathbf{x} is to be recovered from n linear measurements with a given algorithm. Suppose that the measurements are such that the algorithm does not guarantee perfect support recovery and that true features may be missed. Can they efficiently be retrieved?

This paper addresses this question through a simple error-correction module referred to as LiRE. LiRE takes as input an estimate \mathbf{s}_{in} of the true support \mathbf{s}^* , and outputs a refined support estimate \mathbf{s}_{out} . In the noiseless measurement setup, sufficient conditions are established under which LiRE is guaranteed to recover the entire support, that is $\mathbf{s}_{\text{out}} \supseteq \mathbf{s}^*$. These conditions imply, for instance, that in the high-dimensional regime LiRE can correct a sublinear in m number of errors made by Orthogonal Matching Pursuit (OMP). The computational complexity of LiRE is $\mathcal{O}(mnd)$.

Experimental results with random Gaussian design matrices show that LiRE substantially reduces the number of measurements needed for perfect support recovery via Compressive Sampling Matching Pursuit, Basis Pursuit (BP), and OMP. Interestingly, adding LiRE to OMP yields a support recovery procedure that is more accurate and significantly faster than BP. This observation carries over in the noisy measurement setup where the combination of LiRE and OMP is faster and more accurate than LASSO. Finally, as a standalone support recovery algorithm with a random initialization, experiments show that LiRE's reconstruction performance lies between OMP and BP.

These results suggest that LiRE may be used generically, on top of any suboptimal baseline support recovery algorithm, to improve support recovery or to operate with a smaller number of measurements, at the cost of a relatively small computational overhead. Alternatively, LiRE may be used as a standalone support recovery algorithm that is competitive with respect to OMP.

Index Terms

Compressed sensing, error-correction, feature selection, high dimension, linear model, support recovery

I. INTRODUCTION

Consider the support recovery problem of an m -sparse signal $\mathbf{x}^* \in \mathbb{R}^d$ from $n < d$ linear measurements:

$$\mathbf{y} = \Phi \mathbf{x}^*, \quad (1)$$

where $\Phi \in \mathbb{R}^{n \times d}$ refers to the design matrix. Suppose an algorithm A is used to recover the support \mathbf{s}^* of \mathbf{x}^* in a regime where errors may happen; for instance, the Restricted Isometry Property (RIP) constant of Φ need not be small enough to guarantee perfect support recovery—see, *e.g.*, [1] for such a condition for OMP. Can potential errors be efficiently corrected?

We address this question through a simple low complexity error-correction module, referred to as LiRE—for List Regression Error-correction. LiRE takes as input an initial support estimate

$$\mathbf{s}_{\text{in}} \stackrel{\text{def}}{=} A(\mathbf{y}, \Phi)$$

provided by a baseline algorithm A , and produces a second support estimate

$$\mathbf{s}_{\text{out}} \stackrel{\text{def}}{=} \text{LiRE}(\mathbf{y}, \mathbf{s}_{\text{in}}, \Phi) \stackrel{\text{notation}}{=} \text{LiRE} \circ A(\mathbf{y}, \Phi)$$

of size m . Under certain RIP conditions that depend on the number of missed features $|\mathbf{s}^* \setminus \mathbf{s}_{\text{in}}|$, LiRE's estimate \mathbf{s}_{out} includes \mathbf{s}^* (see Theorem 1 and Corollaries 1,2 in Section II). In the high-dimensional regime, these conditions imply, for instance, that we can use OMP in a regime where perfect support recovery is not guaranteed, and yet LiRE will recover all missed features as long as their number grows sublinearly in the sparsity level.

In a second part of the paper, we further assess the performance of LiRE first as an error-correction module, then as a standalone support recovery algorithm with a random initialization. We present four sets of numerical experiments that address the following questions: Can LiRE improve the support recovery of (good) baseline algorithms? Can LiRE be combined with a baseline algorithm of similarly low complexity to achieve the performance of more complex reconstruction algorithms? Is LiRE robust to noise? And, how efficient is LiRE as a standalone support recovery algorithm? These questions are addressed by considering random Gaussian supports and design matrices.

M. Mehrabi is with the Data Sciences and Operations Department, Marshall School of Business, University of Southern California, Los Angeles, CA 90089-0809, US.

A. Tchamkerten is with the Department of Communications and Electronics, Telecom Paris, Institut Polytechnique de Paris, 19 Place Marguerite Perey, 91120 Palaiseau, FR.

- The first experiment considers OMP, BP ([2]), and Compressive Sampling Matching Pursuit (CoSaMP, [3]) as baseline algorithms. Results show that a few (five) iterations of LiRE increases and never decreases the average percentage of exact support recovery, for a non-trivial range of undersampling-sparsity operating regimes. In particular, LiRE reduces the number of measurements needed for perfect support recovery via CoSaMP, BP, and OMP by up to 15%, 25%, and 40%, respectively, depending on the the sparsity level.
- The second set of simulations compares LiRE_{OMP} against BP and shows that even though LiRE_{OMP} has a significantly lower complexity than BP ($\mathcal{O}(mnd)$ vs. $\mathcal{O}(n^2d^{1.5})$, see [4]), it achieves an average percentage of successful support recovery that is at least as large as BP, and sometimes larger.
- The third set of simulations evaluates the robustness of LiRE against noise in the Gaussian additive model

$$\mathbf{y} = \Phi \mathbf{x}^* + \mathbf{z}.$$

By repeating the second set of experiments but now with LiRE_{OMP} against the LASSO solution, we observe that LiRE_{OMP} is superior to LASSO as the noise level increases even though computing the LASSO has a computational cost that is at least quadratic in d [5].

- The fourth set of simulations evaluates the performance of LiRE as a standalone support recovery algorithm with a random initialization. Results show that in terms of percentage of exact support recovery LiRE lies between OMP and BP.

A. Related works

The problem of solving the under-determined system of equations given by (1) to recover the planted solution \mathbf{x}^* , a.k.a. compressed sensing, has a vast literature. By assuming some structural properties of \mathbf{x}^* , there exists a unique solution to (1) which may be found efficiently depending on Φ . The most common property is that \mathbf{x}^* is m -sparse. Finding the sparsest solution

$$\min_{\mathbf{x}: \Phi \mathbf{x} = \mathbf{y}} \|\mathbf{x}\|_0$$

is a non-convex NP-hard problem [6], but convex optimization can recover \mathbf{x}^* if the design matrix Φ satisfies certain conditions. For instance, in [7] it is shown that if Φ satisfies a certain RIP condition, then the sparsest solution corresponds to \mathbf{x}^* but also corresponds to the solution of the convex optimization problem $\min_{\mathbf{x}: \Phi \mathbf{x} = \mathbf{y}} \|\mathbf{x}\|_1$, a.k.a. Basis Pursuit. For alternative structural properties and related conditions see, e.g., [8], [9], [10].

Over the past fifteen years, a significant amount of work has gone into the design of ever more efficient reconstruction algorithms, and conditions on Φ under which reconstruction is possible, see, e.g., [2], [11], [3], [12], [13], [14], [15], [16], [17]. The performance of these algorithms is typically quantified in terms of computational complexity and the sparsity-undersampling tradeoff, that is the n vs. m curve (at fixed d) that characterizes the regimes where support recovery is attained, possibly within some prescribed distortion [18], [7], [19], [20], [21], [22], [23], [24], [25]. Recall that in high dimensions (say, $d = \omega(n^3)$), the computational bottleneck for computing \mathbf{x}^* lies in finding \mathbf{s}^* . Given \mathbf{s}^* , the signal \mathbf{x}^* is obtained by minimizing $\|\mathbf{y} - \Phi \mathbf{x}\|_2$ over all \mathbf{x} with support \mathbf{s}^* . In turn, this least square estimate is equal to Penrose's pseudo-inverse Φ^\dagger of Φ applied to \mathbf{y} , which can be performed with order $\mathcal{O}(m^2n)$ computations using direct methods (e.g., QR factorization) or cost $\mathcal{O}(mn)$ using approximation methods, e.g., Richardson's method (see, e.g., [3, Section 5.1]).

There are two main categories of signal reconstruction algorithms. In the first category, algorithms attempt to solve a convex relaxation of the original non-convex optimization problem, similarly to BP. In the second category, a support estimate is first constructed, typically in an iterative and greedy manner, then the signal is estimated. Prominent algorithms here include OMP and CoSaMP. It is generally accepted that algorithms based on convex relaxations achieve better sparsity-undersampling performance than greedy algorithms at the cost of an increase in computational complexity—see, for instance, [26] for complexity/performance comparison of BP and OMP. Beyond greedy algorithms and in the quest of ever faster reconstruction algorithms, a line of works takes advantage of distributed and parallel computation to further reduce computational time (see, e.g., [27], [28]).

Although many recovery algorithms have been shown to perform well in certain settings, the tradeoff between reconstruction performance and computational complexity remains elusive in general. The present work is a further exploration of this tradeoff by providing a means to increase reconstruction performance at the cost of a relatively small computational overhead.

Remark. *This paper is an extended version of the ISIT 2021 conference submission [29]. The main difference with the present paper is that the ISIT submission states Theorem 1 but without proof. The proof consists of a sequence of eight Lemmas and one proposition which shed light on the role of the list, the key component of LiRE. The ISIT version did not include Corollary 1 and Example 1, and included only part of the simulations presented here. In particular, simulations pertaining to LiRE's robustness to noise and pertaining to LiRE as a standalone support recovery algorithm are not present in the ISIT submission.*

B. Paper organization

We end this section with notational conventions. In Section II, we introduce LiRE and state sufficient conditions under which LiRE corrects all errors made by the baseline algorithm. In Section III, we provide experimental results. In Section IV, we prove the results of Section II, and in Section V we draw concluding remarks and outline open problems.

C. Notational conventions

The set of possible features is denoted as $\mathcal{F}_d \stackrel{\text{def}}{=} \{1, 2, \dots, d\}$. A *support vector* refers to a vector whose entries in \mathcal{F}_d are listed the *ascending order*. Given a support vector \mathbf{s} , $\mathbf{s}[i]$ denotes the i th entry of \mathbf{s} , $\mathbf{s}[-i]$ denotes the vector obtained by removing the i th entry of \mathbf{s} , and $\bar{\mathbf{s}}$ denotes a vector with entries in \mathcal{F}_d and not in \mathbf{s} . We use $j \in \mathbf{s}$ whenever $\mathbf{s}[i] = j$ for some $i \in \{1, 2, \dots, d\}$. The length and the support of a vector $\mathbf{x} \in \mathbb{R}^d$ are denoted as $|\mathbf{x}|$ and $\text{supp}(\mathbf{x})$, respectively. Vector \mathbf{x} is said to be m -sparse if $|\text{supp}(\mathbf{x})| \leq m$. Given support vectors \mathbf{r} and \mathbf{s} , we use $\mathbf{r} \setminus \mathbf{s}$ to denote the support vector whose entries belong to \mathbf{r} and not to \mathbf{s} . We write $\mathbf{r} \subset \mathbf{s}$ whenever $\mathbf{r} \setminus \mathbf{s}$ is the null vector. Further, we use $\mathbf{r} \cap \mathbf{s}$ to denote the support vector whose entries appear in both \mathbf{r} and \mathbf{s} , and use $\mathbf{r} \cup \mathbf{s}$ to denote the support vector whose entries appear in \mathbf{s} or \mathbf{r} . For example, given support vectors $\mathbf{r} = [1, 2, 3, 4]$ and $\mathbf{s} = [1, 4, 5]$ in \mathcal{F}_6 , we have $\mathbf{r}[-1] = [2, 3, 4]$, $\mathbf{s} \setminus \mathbf{r} = [5]$, $\mathbf{s} \cap \mathbf{r} = [1, 4]$, $\mathbf{r} \cup \mathbf{s} = [1, 2, 3, 4, 5]$, and that $\mathbf{r}[-1]$ and $\mathbf{s} \setminus \mathbf{r}$ are disjoint.

Throughout the paper Φ refers to an $n \times d$ real matrix and we use $\Phi_{\mathbf{s}}$ to denote the matrix Φ restricted to the set of columns indexed by the entries of \mathbf{s} . The transpose of matrix Φ is denoted as Φ^T and $\Phi_{\mathbf{s}}^T$ denotes $(\Phi_{\mathbf{s}})^T$.

Given $\Psi \in \mathbb{R}^{n \times m}$, the (least square) residual of $\mathbf{y} \in \mathbb{R}^n$ is defined as

$$\mathbf{y}^{\perp \Psi} \stackrel{\text{def}}{=} \mathbf{y} - \Psi \cdot \arg \min_{\mathbf{z} \in \mathbb{R}^m} \|\mathbf{y} - \Psi \mathbf{z}\|_2.$$

Recall that if $\Psi^T \Psi$ is invertible then

$$\mathbf{y}^{\perp \Psi} = (I - P\{\Psi\})\mathbf{y}$$

where

$$P\{\Psi\} \stackrel{\text{def}}{=} \Psi(\Psi^T \Psi)^{-1} \Psi^T$$

is the projection operator (see, e.g., [30]).

With a slight abuse of notation, the residual with respect to a support vector \mathbf{s} is defined as

$$\mathbf{y}^{\perp \mathbf{s}} \stackrel{\text{def}}{=} \mathbf{y}^{\perp \Phi_{\mathbf{s}}}.$$

The support vector consisting of the ℓ most correlated features with respect to $\mathbf{y}^{\perp \mathbf{s}}$ is defined as

$$\mathcal{L}(\ell, \mathbf{s}) \stackrel{\text{def}}{=} \arg \max_{\mathbf{q}, |\mathbf{q}|=\ell} \|\Phi_{\mathbf{q}}^T \mathbf{y}^{\perp \mathbf{s}}\|_1,$$

where the maximum is intended to be over support vectors in \mathcal{F}_d . The least square estimate with respect to a support vector \mathbf{s} is defined as

$$\mathcal{E}(\mathbf{s}) \stackrel{\text{def}}{=} \arg \min_{\mathbf{x} \in \mathbb{R}^d, \text{supp}(\mathbf{x}) \subseteq \mathbf{s}} \|\mathbf{y} - \Phi \mathbf{x}\|_2.$$

II. LIST REGRESSION ERROR-CORRECTION (LiRE)

Consider the noiseless linear model

$$\mathbf{y} = \Phi \mathbf{x}^* \tag{2}$$

with $\mathbf{x}^* \in \mathbb{R}^d$, $\mathbf{y} \in \mathbb{R}^n$, $n \leq d$, and where the design matrix $\Phi \in \mathbb{R}^{n \times d}$ is assumed to have unit l_2 columns without loss of generality. Given Φ , \mathbf{y} , and knowing that

$$|\mathbf{s}^*| \stackrel{\text{def}}{=} |\text{supp}(\mathbf{x}^*)| \leq m \leq n,$$

we want to find \mathbf{s} such that $\mathbf{s}^* \subseteq \mathbf{s}$ and $|\mathbf{s}| = m$.

Given an initial estimate \mathbf{s}_{in} of \mathbf{s}^* that potentially misses true features, LiRE attempts to produce a second estimate \mathbf{s}_{out} that contains all true features, that is $\mathbf{s}_{\text{out}} \supseteq \mathbf{s}^*$. At the heart of LiRE is a leave-one-out procedure for error-correction purpose which checks, for each feature of the initial support estimate, whether it can be replaced by a better one. The pseudo-code of LiRE is given below and is followed by comments in light of existing support recovery algorithms:

Algorithm 1 LiRE (List Regression Error-Correction)

Input:

- $n \times d$ real design matrix Φ
- n dimensional data vector \mathbf{y}
- upper bound m on support size of \mathbf{x}^*
- list size $1 \leq \ell \leq m$ (internal parameter)
- initial estimate \mathbf{s}_{in} of \mathbf{s}^* with $|\mathbf{s}_{\text{in}}| \leq m$

Initialization:

If $|\mathbf{s}_{\text{in}}| < m$, add any $m - |\mathbf{s}_{\text{in}}|$ features from \mathcal{F}_d to \mathbf{s}_{in} such that $|\mathbf{s}_{\text{in}}| = m$. Set $\mathbf{s}_{\text{out}} = \mathbf{s}_{\text{in}}$.

Procedure:

for $i = 1$ **to** m **do**

- If $\mathbf{y}^{\perp \mathbf{s}_{\text{out}}[-i]} = 0$, exit the **for** loop. Else, find the ℓ most correlated features with respect to $\mathbf{y}^{\perp \mathbf{s}_{\text{out}}[-i]}$:

$$\mathbf{l} \stackrel{\text{def}}{=} \mathcal{L}(\ell, \mathbf{s}_{\text{out}}[-i]).$$

- Compute the least square estimate with respect to features $\mathbf{s}_{\text{out}}[-i] \cup \mathbf{l}$:

$$\hat{\mathbf{x}} \stackrel{\text{def}}{=} \mathcal{E}(\mathbf{s}_{\text{out}}[-i] \cup \mathbf{l})$$

- Pick $j \in \mathbf{l}$ such that $|\hat{\mathbf{x}}_j| = \|\hat{\mathbf{x}}_{\mathbf{l}}\|_{\infty}$.
- Replace $\mathbf{s}_{\text{out}}[i]$ by j .

end for

Output:

- Support vector \mathbf{s}_{out}
-

Feature i is first removed from the current support estimate and the residual is computed. If the residual is non-zero, the ℓ features that are most correlated with the residual are added to the support. This results in an expanded support of size $m - 1 + \ell$, from which a signal estimate is computed. Finally, feature i is replaced with the most relevant feature of this estimate, restricted to the list elements—in particular, feature i could replace itself if it belongs to the list.

The support expansion through the list is reminiscent of the signal proxy formation in several greedy algorithms, including OMP, gOMP [31], and CoSaMP. This step, however, serves here the purpose of error-correction as it allows to test whether a particular feature should be replaced or not. Intuitively, a wrong feature is less likely to be corrected if the list size is small. But a correct feature is also more likely to get replaced by a wrong feature if the list is large. Accordingly, the theoretical guarantees for successful error-correction provided below (Section II-B) tie the number of errors and the list size in an attempt to strike a balance between these two types of error.

A. Computational complexity of LiRE

For each of its m rounds, LiRE involves:

- Two least square problems, cost $\mathcal{O}(m^2n)$ using direct methods or cost $\mathcal{O}(mn)$ using approximation methods (see Section I-A).
- d inner products of n dimensional vectors, cost $\mathcal{O}(nd)$.
- Two sortings of d numbers, cost $\mathcal{O}(d \log d)$ (e.g., by Merge Sort).

Hence, LiRE's computational cost is $\mathcal{O}(mnd)$, with the restriction $m \leq \sqrt{d}$ if we use direct methods for the least square problems. Notice that the second and third computations can be performed efficiently through parallelization.

B. Sufficient conditions for error-correction

Theorem 1 below provides a sufficient condition under which one pass of LiRE recovers all missed true features of the initial support estimate.

Theorem 1. Fix integers $1 \leq \ell \leq m \leq n \leq d$, and consider the model (2) for a given design matrix $\Phi \in \mathbb{R}^{n \times d}$. Let \mathbf{s}_{in} , with $|\mathbf{s}_{\text{in}}| \leq m$, denote an estimate of the true support \mathbf{s}^* , let $e \stackrel{\text{def}}{=} |\mathbf{s}^* \setminus \mathbf{s}_{\text{in}}|$ denote the number of missed true features, and let

$$t \stackrel{\text{def}}{=} \max\{m + e, \ell + e + 1\},$$

$$\eta_t \stackrel{\text{def}}{=} \frac{\sqrt{2}\delta_t(1 - \delta_t^2)}{(1 - \delta_t - \delta_t^2)(1 - 2\delta_t)},$$

where δ_t denotes the order- t RIP constant of Φ .¹

Suppose that e , ℓ , and Φ satisfy the following inequalities:

$$\ell \leq \max\{e, 1\} \quad (3)$$

$$\sqrt{e+1} \leq \frac{\sqrt{2}(1-\delta_t-\delta_t^2)(1-2\delta_t)}{\delta_t(1+\delta_t)(1+2\delta_t-\delta_t^2)} \quad (4)$$

$$\sqrt{\ell} > \frac{(1-\delta_t^2+\delta_t)\eta_t\sqrt{e+1}-1+\delta_t}{1-\delta_t-\delta_t\eta_t\sqrt{e+1}}\sqrt{e+1} \quad (5)$$

$$\delta_{\ell+m-1} < 0.5. \quad (6)$$

Then, given \mathbf{s}_{in} , LiRE with list size ℓ outputs \mathbf{s}_{out} such that $\mathbf{s}_{out} \supset \mathbf{s}^*$ and $|\mathbf{s}_{out}| = m$.

A more explicit sufficient condition for error correction is obtained by choosing $\ell = \max\{e, 1\}$, which implies $t \leq m+e+2$ since $e \leq m$, and $\ell \leq e+1$. Condition (4) in Theorem 1 with strict inequality implies Condition (5). Furthermore, Condition (6) is implied by the condition $\delta_{m+e} < 0.5$ since $\ell \leq e+1$. It then follows:

Corollary 1. *LiRE with list size $\ell = \max\{e, 1\}$ corrects exactly e errors (i.e., $|\mathbf{s}_{out}| = m$, $\mathbf{s}_{out} \supset \mathbf{s}^*$, $|\mathbf{s}^* \setminus \mathbf{s}_{in}| = e$) if Φ satisfies the following RIP conditions:*

$$\delta_{m+e} < 0.5,$$

$$e+1 < \left(\frac{\sqrt{2}(1-\delta_{m+e+2}-\delta_{m+e+2}^2)(1-2\delta_{m+e+2})}{\delta_{m+e+2}(1+2\delta_{m+e+2}-\delta_{m+e+2}^2)(1+\delta_{m+e+2})} \right)^2. \quad (7)$$

Theorem 1 and Corollary 1 assume that the number of errors to be corrected is known to be exactly e . If we want LiRE to correct any number of errors up to some number $\bar{e} \geq 1$, then by Condition (3) we should pick $\ell = 1$ which yields the following result:

Corollary 2. *LiRE with list size $\ell = 1$ corrects up to \bar{e} errors if Φ satisfies the following RIP conditions:*

$$\delta_m < 0.5, \quad (8)$$

$$\bar{e}+1 \leq \left(\frac{(1-\delta_{m+\bar{e}+1}-\delta_{m+\bar{e}+1}^2)(1-2\delta_{m+\bar{e}+1})}{\delta_{m+\bar{e}+1}\sqrt{2}(1+\delta_{m+\bar{e}+1}-\delta_{m+\bar{e}+1}^2)(1+\delta_{m+\bar{e}+1})} \right)^2. \quad (9)$$

Proof of Corollary 2. Inequality (9) implies (4) and renders inequality (5) vacuous as its right-hand side becomes non-positive. Finally, note that (6) is satisfied if $\ell = 1$ and (8) holds. \square

For small values of δ 's the upperbound (9) is about four times larger than the upperbound (7), but is sufficient to show that LiRE can actually correct errors beyond the regime of exact support recovery of certain baseline algorithms. The following example shows that LiRE corrects a sublinear in m number of errors in a regime where OMP may produce errors:

Example 1 (OMP, sublinear number of errors). *Corollary 2 implies that LiRE corrects up to \bar{e} errors if*

$$\delta_{m+1+\bar{e}} \leq \frac{(1+o(1))}{\sqrt{\bar{e}}} \quad \text{as } \bar{e}, m \rightarrow \infty \text{ with } \bar{e} \leq m. \quad (10)$$

On the other hand, the RIP necessary (and sufficient) condition for OMP to recover the support is (see [32], [33])²

$$\delta_{m+1} < \frac{1}{\sqrt{m+1}}. \quad (11)$$

Now, using the property $\delta_{ab} \leq b \cdot \delta_{2a}$ for positive integers a and b [3, Corollary 3.4] with $a = (m+1)/2$ and $b = 2(1+\bar{e}/(m+1))$, we deduce that in the regime $\bar{e}, m \rightarrow \infty$ with $\bar{e} = o(m)$, Condition (11) implies

$$\delta_{m+1+\bar{e}} \leq \frac{2(1+o(1))}{\sqrt{m+1}},$$

¹Given an integer $t \geq 1$, the order- t Restricted Isometry Property (RIP) constant δ_t of matrix Φ is defined as the smallest δ such that the inequality

$$(1-\delta)\|\mathbf{x}\|_2^2 \leq \|\Phi\mathbf{x}\|_2^2 \leq (1+\delta)\|\mathbf{x}\|_2^2$$

holds for all t -sparse vectors \mathbf{x} .

²If (11) is reversed, then there exists design matrices for which OMP is not guaranteed to always recover \mathbf{s}^* .

which is more stringent than (10). In summary, if the design matrix satisfies (10), OMP alone is not guaranteed to recover the support and LiRE will retrieve the missed features as long as their number is known and sublinear in the sparsity.

In general, to quantify the benefits due to LiRE through Corollary 2 we need to identify a meaningful regime, namely conditions on the design matrix under which the baseline algorithm potentially makes errors, and an upper bound on the number of errors. (Note that from Corollary 2 it is unclear how LiRE performs when the actual number of errors is above \bar{e} or different than e for Theorem 1 and Corollary 1.)

Unfortunately, such conditions are hardly available. And even when they are, such as for OMP, non-trivial upper bounds on the number of errors remain elusive.³ Hence, to provide a practical assessment of the performance of LiRE without bounds on the number of errors we resorted to numerical simulations which are presented in the next section.

III. NUMERICAL EXPERIMENTS

The simulations described next evaluate the performance of LiRE both as an error correction module and as a standalone support recovery algorithm, with random Gaussian support and design matrices. Following the guidelines of reproducible research, the code for the simulations described below is available at:

<https://github.com/mehrab4/LiRE>.

List size

Theorem 1 suggests that the list size should be tied to the number of errors. Since this number is typically unknown, for all the numerical experiments we chose ℓ empirically to be equal to $0.5m$ if $1.5m \leq n$, and $n - m$ otherwise.

A. Improvement over OMP, CoSaMP, and BP

The first set of experiments shows how LiRE improves the support recovery of OMP, CoSaMP, and BP for different normalized sparsity levels m/d and number of samples n/d . Experiments were carried for $d \in \{256, 512, 1024\}$ for different values of m and n with step size of about $0.015d$ for m and $0.03d$ for n . Given m, n, d , a design matrix Φ was first randomly generated with i.i.d. normal $\mathcal{N}(0, \frac{1}{n})$ entries. The signal \mathbf{x} was randomly generated with a uniformly chosen support of size m and with independently generated i.i.d. $\mathcal{N}(0, 1)$ entries. The percentages of perfect support recovery over 50 experiments were computed with and without LiRE—each experiment is run with and without LiRE. For each experiment, LiRE was run 5 consecutive times. The number of iterations of CoSaMP was chosen to be $d/4$. In Fig. 1, levels of red indicate improvement—percentage of exact support recovery with LiRE minus percentage of exact support recovery without LiRE—from 0% (white squares) to 100% (dark red squares).

Discussion: The lower white region in each plot of Fig. 1 corresponds to no recovery (0%) with and without LiRE, whereas the upper white region corresponds to maximal (100%) recovery with and without LiRE. Between these two regions LiRE improves recovery, up to 100% for the dark red regions. We observe that LiRE never degrades performance, and numerical values at $d = 1024$ reveal that LiRE reduces the number of samples needed to reach perfect (100%) reconstruction by 30 – 40% for OMP, 10 – 15% for CoSaMP, and 15 – 25% for BP. Notice that the dark red regions correspond to regimes where the baseline algorithm always produces errors, and LiRE correct them all.

The choice of running LiRE five times was based on empirical trials. This guaranteed a non-negative improvement of the percentage of exact support recovery across the entire range of sparsity levels (from 0.08 to 0.23). By contrast, we observed that fewer iterations could result in a slight decrease in performance for certain sparsity levels. This is possibly due to the fact that the list size was not optimized.

B. LiRE \circ OMP versus BP

The second set of experiments is perhaps the most interesting as it shows how LiRE may boost the performance of a low complexity reconstruction algorithm to achieve the performance of significantly more complex ones. We ran the same experiments as in the previous section but now compared BP against LiRE \circ OMP. Levels of red in Fig. 2 correspond to the percentage of exact support recovery of LiRE \circ OMP minus the percentage of exact support recovery of BP, with LiRE run 1, 3, and 5 consecutive times.

Fig. 3 shows the percentage of exact support recovery as a function of the number of measurements, for LiRE \circ OMP and BP, with LiRE run only once, $d = 512$, and three sparsity levels (35, 70, and 100).

Discussion: Referring to Fig. 2, we observe that LiRE \circ OMP performs at least as well BP—in certain regimes strictly better—despite the significantly lower computational cost: $\mathcal{O}(mnd)$ for LiRE \circ OMP vs. $\mathcal{O}(n^2d^{1.5})$ for BP [4]. We also note that the region of improvement (non-white region) does not substantially change with the number of iterations. Finally, the numerical data for Fig. 3 shows that LiRE \circ OMP reduces by 5 – 10% the number of measurements needed by BP to achieve perfect support recovery.

³Note that here we are interested in conditions on the design matrix under which a suboptimal algorithm produces at most a certain number of errors, in the worst-case over \mathbf{x}^* . In fact, several works investigate the fundamental limitations of non-zero error support recovery, or recovery with distortion, in probabilistic setups (see, e.g., [24], [25]).

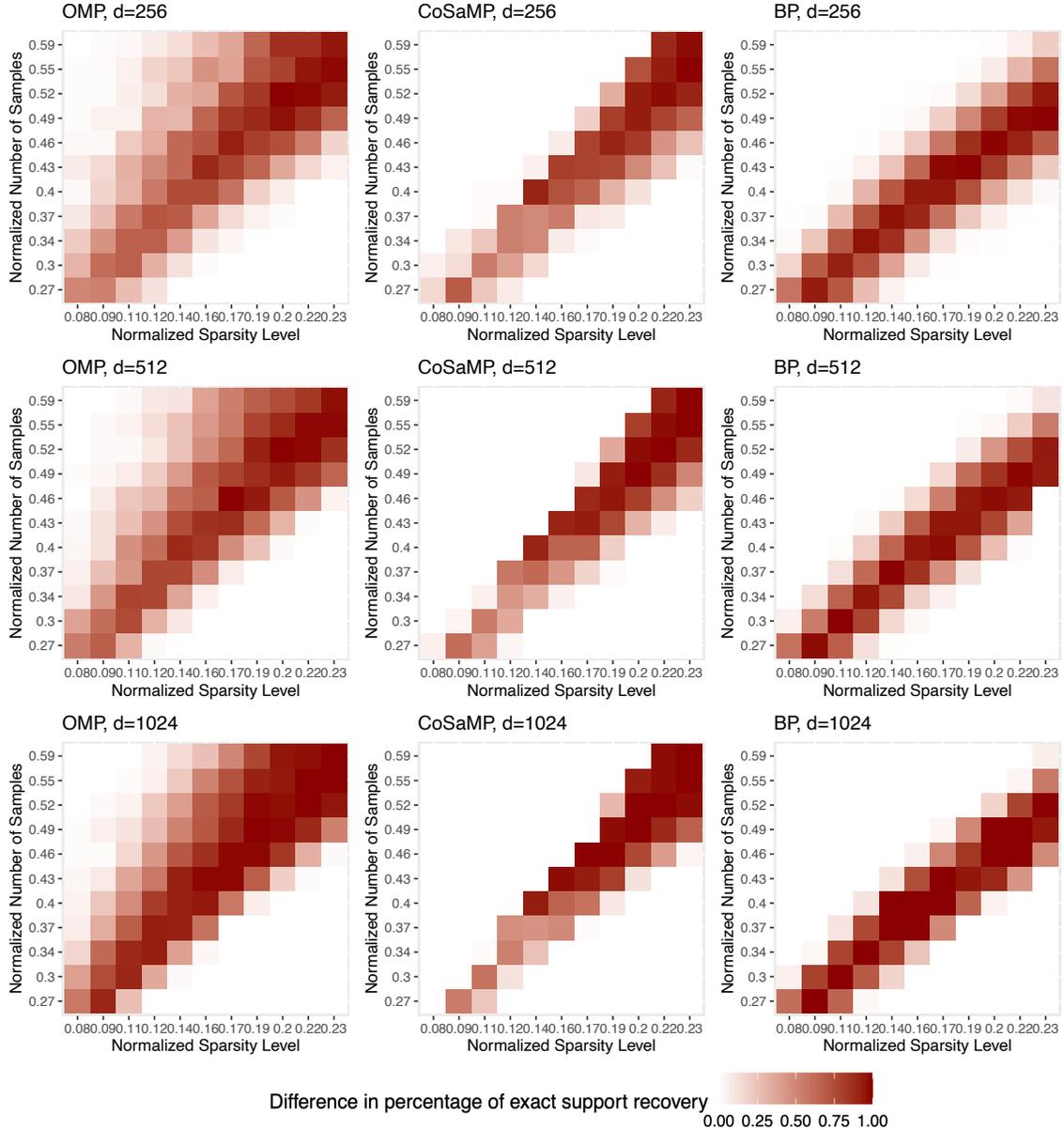


Fig. 1: Improvement due to LiRE. Levels of red indicate the difference in the percentage of exact support recovery with and without LiRE, for OMP, CoSaMP, and BP—differences are always nonnegative. White cells indicate no improvement and dark red cells indicate regimes where LiRE allows to entirely recover the underlying support whereas the baseline algorithm alone makes errors.

C. Robustness to noise: $\text{LiRE} \circ \text{OMP}$ vs. LASSO

In the third set of experiments, see Fig.4, we considered the noisy measurement model

$$\mathbf{y} = \Phi \mathbf{x}^* + \mathbf{z},$$

and compared $\text{LiRE} \circ \text{OMP}$ (with LiRE run only once) against LASSO, similarly as for the previous set of experiments (difference of percentage of exact recovery computed over 50 experiments for any given pair of normalized sparsity level and number of samples). The noise vector \mathbf{z} was i.i.d. Gaussian across components, with zero mean and variance $\sigma^2 \in \{0.0005, 0.001, 0.002\}$. The variance values were chosen empirically; high enough to put LASSO outside its comfort zone of perfect reconstruction, but not too high to allow LiRE to correct errors. The ℓ_1 -regularization parameter of LASSO was chosen by performing a ten-fold cross-validation.

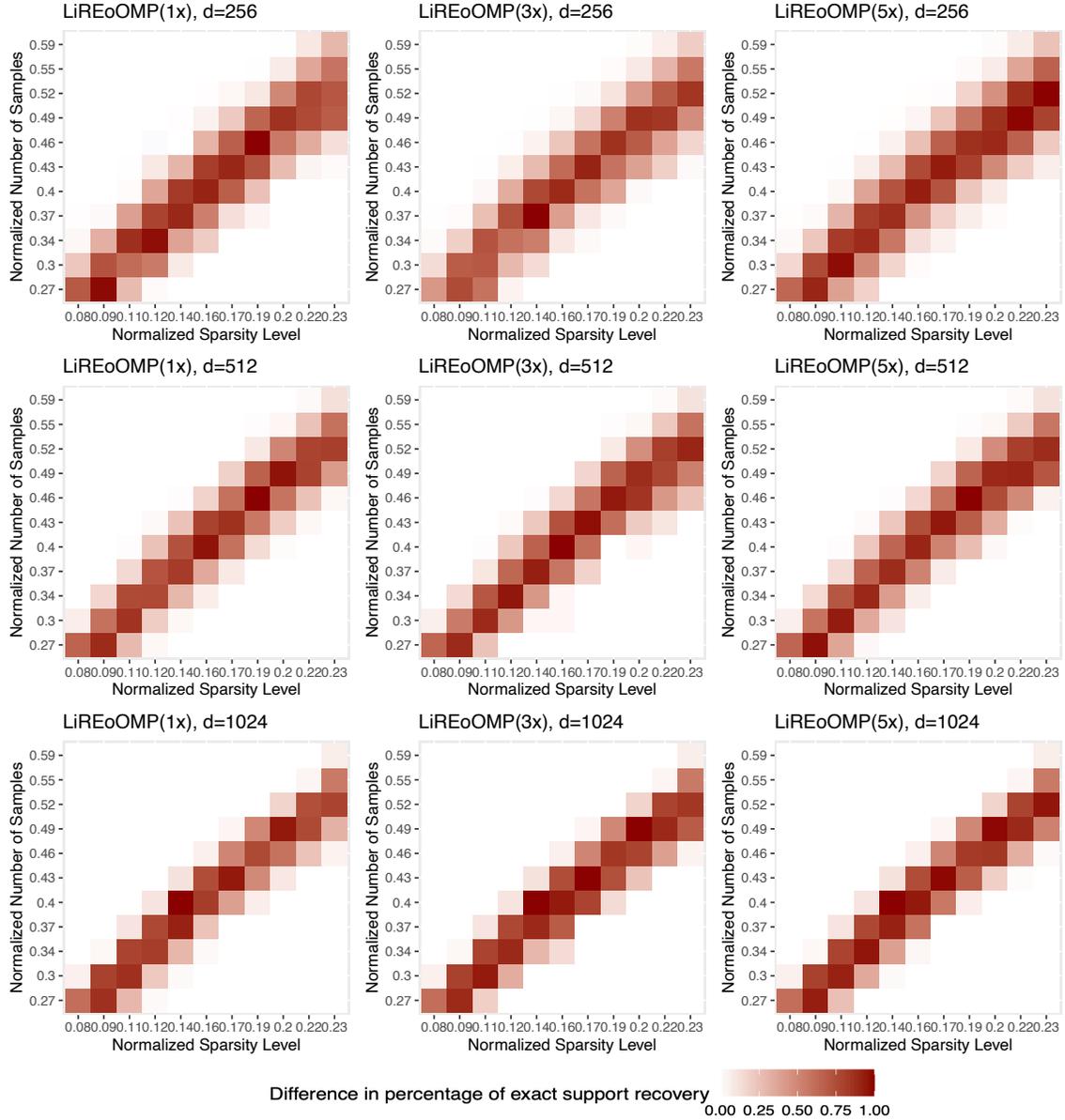


Fig. 2: Same experiment as illustrated in Fig. 1, but for LiRE \circ OMP vs. BP, with 1, 3 and 5 iterations of LiRE.

Discussion

In Fig. 4, we observe that LiRE \circ OMP always improves support recovery over LASSO (except for the white region where the support is never fully recovered, with and without LiRE) and is also significantly faster since LASSO solver's complexity is quadratic or cubic in d , depending on the sparsity level [5]. Outside the white region the improvement is significant, mostly by more than 50%.

D. LiRE as a standalone support recovery algorithm?

In the fourth set of experiments (Fig. 5), we compared LiRE (one pass) as a standalone support recovery algorithm against BP and OMP, and run similar experiments as described in Section III-B. LiRE was initialized with a randomly and uniformly selected \mathbf{s}_{in} . Fig. 5a gives the percentage of exact support recovery with BP minus the percentage of exact support recovery with LiRE. Red favors BP and blue favors LiRE, and similarly for LiRE against OMP in Fig. 5b where red favors LiRE. Recall that the computational complexities of LiRE and OMP are similar, of order $\mathcal{O}(mnd)$, whereas it is order $\mathcal{O}(n^2d^{1.5})$ for BP.

Discussion

Fig. 5a shows that BP is superior to LiRE at higher sparsity levels. At lower sparsity levels performances appear to be close. Fig. 5b shows that LiRE achieves better performance than OMP, and that the difference gets more pronounced as the sparsity

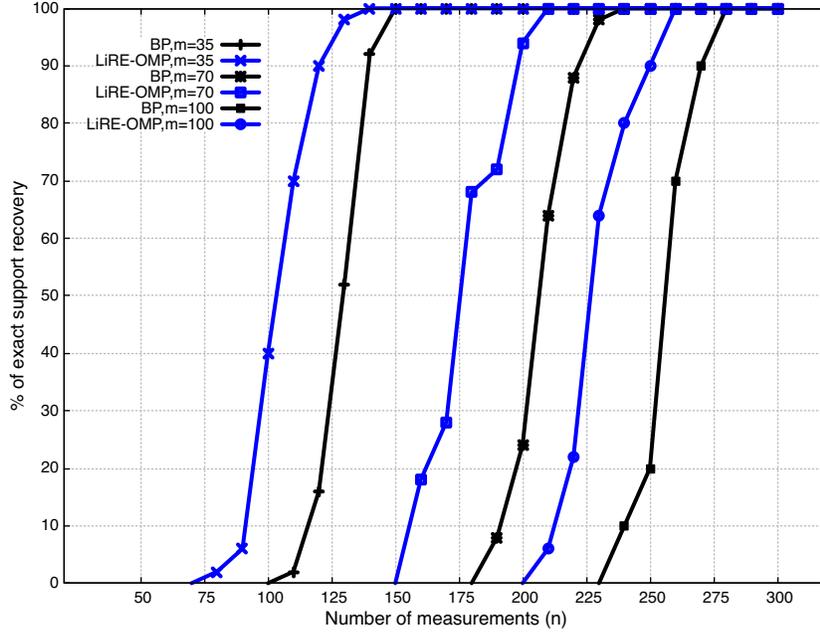


Fig. 3: LiRE \circ OMP vs. BP, for $d = 512$.

level increases.

Note here that with a random initialization \mathbf{s}_{in} may contain no correct feature, that is $\bar{e} = m$. Applying Corollary 2 with the trivial upper bound $\bar{e} = m$ yields the sufficient condition

$$\delta_{2m+1} \leq \frac{1}{\sqrt{2m}}(1 + o(1)) \quad m \rightarrow \infty \quad (12)$$

for LiRE to recover an arbitrary number of errors. This condition is significantly more stringent than the sufficient condition for OMP to recover the underlying support: $\delta_{m+1} < \frac{1}{\sqrt{m+1}}$. In light of Fig. 5b, LiRE's sufficient condition (12) to recover an arbitrary number of errors appears to be loose.

IV. ANALYSIS

Lemma 1 (Lemmas 1 and 2 in [12]). *Let \mathbf{l}, \mathbf{s} be disjoint support vectors and suppose $\delta_{|\mathbf{l}+\mathbf{s}|} < 1$. Then, for all $\mathbf{a} \in \mathbb{R}^{|\mathbf{l}|}$ and $\mathbf{b} \in \mathbb{R}^{|\mathbf{s}|}$ we have*

$$\begin{aligned} |\mathbf{a}^\top \Phi_{\mathbf{l}}^\top \Phi_{\mathbf{s}} \mathbf{b}| &\leq \delta_{|\mathbf{l}+\mathbf{s}|} \|\mathbf{a}\|_2 \|\mathbf{b}\|_2, \\ \|\Phi_{\mathbf{l}}^\top \Phi_{\mathbf{s}} \mathbf{b}\|_2 &\leq \delta_{|\mathbf{l}+\mathbf{s}|} \|\mathbf{b}\|_2. \end{aligned}$$

Lemma 2. *Let \mathbf{s}, \mathbf{q} be support vectors that are disjoint from support vector \mathbf{l} . If $\max\{\delta_{|\mathbf{s}+\mathbf{l}|}, \delta_{|\mathbf{q}+\mathbf{l}|}\} < 1$, then for any $\mathbf{a} \in \mathbb{R}^{|\mathbf{s}|}$ we have*

$$\begin{aligned} \|P\{\Phi_{\mathbf{l}}\} \Phi_{\mathbf{s}} \mathbf{a}\|_2 &\leq \frac{\delta_{|\mathbf{s}+\mathbf{l}|}}{\sqrt{1-\delta_{|\mathbf{l}|}}} \|\mathbf{a}\|_2, \\ \|\Phi_{\mathbf{q}}^\top P\{\Phi_{\mathbf{l}}\} \Phi_{\mathbf{s}} \mathbf{a}\|_2 &\leq \frac{\delta_{|\mathbf{s}+\mathbf{l}|} \delta_{|\mathbf{q}+\mathbf{l}|}}{1-\delta_{|\mathbf{l}|}} \|\mathbf{a}\|_2. \end{aligned}$$

Proof. We know that

$$\|P\{\Phi_{\mathbf{l}}\} \Phi_{\mathbf{s}} \mathbf{a}\|_2^2 = \mathbf{u}^\top (\Phi_{\mathbf{l}}^\top \Phi_{\mathbf{l}})^{-1} \mathbf{u},$$

where $\mathbf{u} \stackrel{\text{def}}{=} \Phi_{\mathbf{l}}^\top \Phi_{\mathbf{s}} \mathbf{a}$. From the definition of the restricted isometry property we get

$$\mathbf{u}^\top (\Phi_{\mathbf{l}}^\top \Phi_{\mathbf{l}})^{-1} \mathbf{u} \leq \frac{\|\mathbf{u}\|_2^2}{1-\delta_{|\mathbf{l}|}}.$$

Next, Lemma 1 gives us

$$\|\mathbf{u}\|_2 \leq \delta_{|\mathbf{l}+\mathbf{s}|} \|\mathbf{a}\|_2.$$

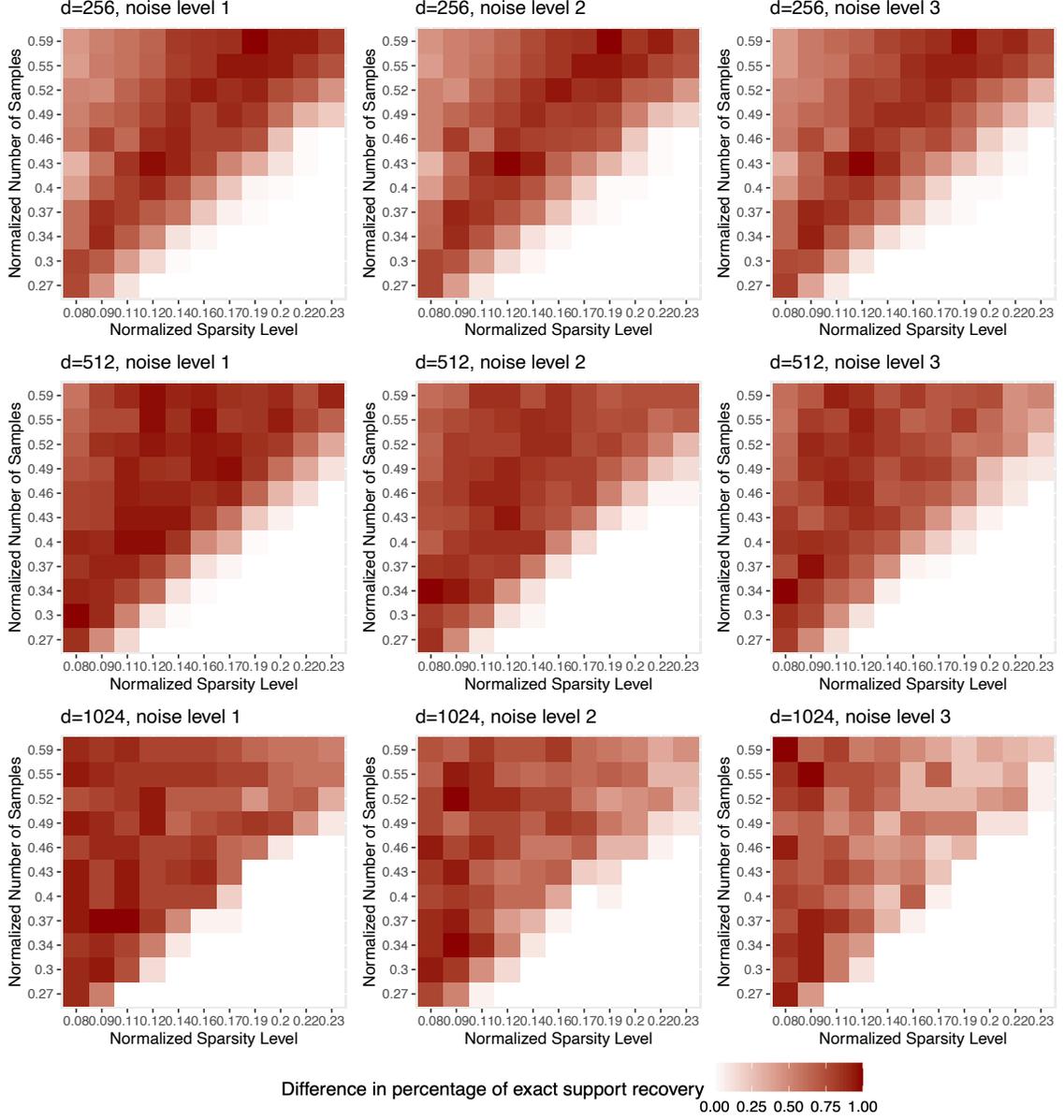


Fig. 4: Same experiment as illustrated in Fig. 1, but for LiReOMP vs. LASSO at different noise levels.

This implies

$$\|P\{\Phi_l\}\Phi_s a\|_2 \leq \frac{\delta_{|l|+|s|}}{\sqrt{1-\delta_{|l|}}}\|a\|_2.$$

For the proof of the second inequality in Lemma 2 first assume $\|\mathbf{b}\|_2 \leq 1$. Then, the Cauchy inequality yields

$$\mathbf{b}^\top \Phi_q^\top P\{\Phi_l\}\Phi_s \mathbf{a} \leq \|P\{\Phi_l\}\Phi_q \mathbf{b}\|_2 \|P\{\Phi_l\}\Phi_s \mathbf{a}\|_2.$$

Next, by invoking the first inequality in Lemma 2 we obtain

$$\max_{\|\mathbf{b}\|_2 \leq 1} \mathbf{b}^\top \Phi_q^\top P\{\Phi_l\}\Phi_s \mathbf{a} \leq \frac{\delta_{|s|+|l|}\delta_{|q|+|l|}}{1-\delta_{|l|}}\|a\|_2,$$

which establishes the desired result. \square

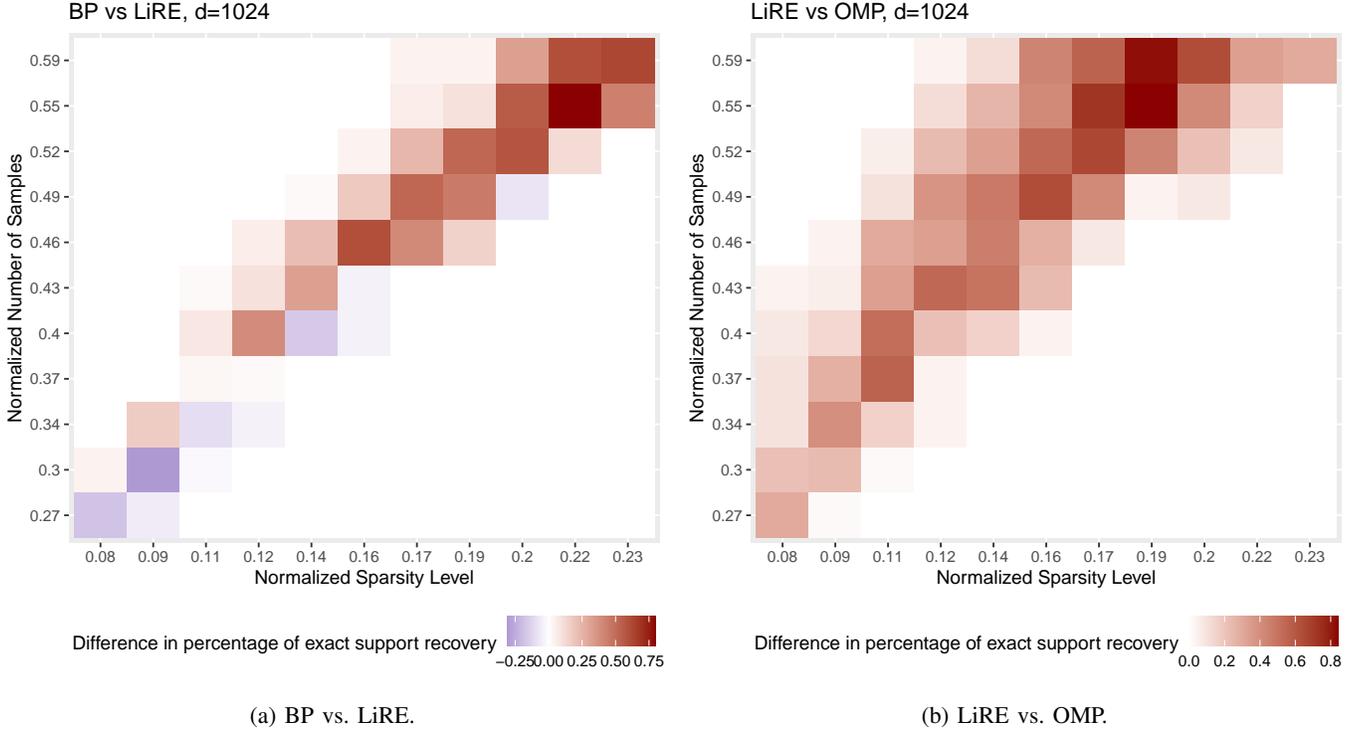


Fig. 5: LiRE as a standalone support recovery algorithm. Red squares favor BP in Fig. 5a and favor LiRE in Fig. 5b.

Lemma 3. Let $\mathbf{s}_1, \mathbf{s}_2$ be two disjoint support vectors with $|\mathbf{s}_1| = k_1$ and $|\mathbf{s}_2| = k_2$. If $\delta_{k_1+k_2} \leq 0.5$, then $\Phi_{\mathbf{s}_1}^\top (I - P\{\Phi_{\mathbf{s}_2}\}) \Phi_{\mathbf{s}_1}$ is invertible and

$$(\Phi_{\mathbf{s}_1}^\top (I - P\{\Phi_{\mathbf{s}_2}\}) \Phi_{\mathbf{s}_1})^{-1} \succeq \frac{1}{1 + \delta_{k_1}} I_{k_1}.$$

Proof. Given $\mathbf{z} \in \mathbb{R}^{k_1}$ we have

$$\mathbf{z}^\top (\Phi_{\mathbf{s}_1}^\top (I - P\{\Phi_{\mathbf{s}_2}\}) \Phi_{\mathbf{s}_1}) \mathbf{z} = \|\Phi_{\mathbf{s}_1} \mathbf{z}\|_2^2 - \|P\{\Phi_{\mathbf{s}_2}\} \Phi_{\mathbf{s}_1} \mathbf{z}\|_2^2,$$

which is no larger than $\|\Phi_{\mathbf{s}_1} \mathbf{z}\|_2^2$. Next, recall that the RIP property implies

$$\|\Phi_{\mathbf{s}_1} \mathbf{z}\|_2^2 \leq (1 + \delta_{k_1}) \|\mathbf{z}\|_2^2. \quad (13)$$

Therefore,

$$(\Phi_{\mathbf{s}_1}^\top (I - P\{\Phi_{\mathbf{s}_2}\}) \Phi_{\mathbf{s}_1}) \preceq (1 + \delta_{k_1}) I_{k_1}.$$

Moreover, Lemma 2 gives us

$$\|P\{\Phi_{\mathbf{s}_2}\} \Phi_{\mathbf{s}_1} \mathbf{z}\|_2^2 \leq \frac{\delta_k^2 \|\mathbf{z}\|_2^2}{1 - \delta_{k_2}}. \quad (14)$$

By combining (13) and (14) we get

$$\begin{aligned} (1 - \delta_{k_1} - \frac{\delta_k^2}{1 - \delta_{k_2}}) I_{k_1} &\preceq (\Phi_{\mathbf{s}_1}^\top (I - P\{\Phi_{\mathbf{s}_2}\}) \Phi_{\mathbf{s}_1}) \\ &\preceq (1 + \delta_{k_1}) I_{k_1}. \end{aligned}$$

Therefore $(\Phi_{\mathbf{s}_1}^\top (I - P\{\Phi_{\mathbf{s}_2}\}) \Phi_{\mathbf{s}_1})$ is invertible and we have the desired property. \square

Lemma 4. Let $\mathbf{s}_1, \mathbf{s}_2$ be two disjoint support vectors, let $\mathbf{s} = \mathbf{s}_1 \cup \mathbf{s}_2$, and suppose $\delta_{|\mathbf{s}_1|+|\mathbf{s}_2|} \leq 0.5$. Then,

$$\mathbf{x}_{\mathbf{s}_1} = (\Phi_{\mathbf{s}_1}^\top (I - P\{\Phi_{\mathbf{s}_2}\}) \Phi_{\mathbf{s}_1})^{-1} \Phi_{\mathbf{s}_1}^\top \mathbf{y}^{\perp \mathbf{s}_2}.$$

Proof. Since $\delta_{|\mathbf{s}_1|+|\mathbf{s}_2|} < 1$ we have that $\Phi_{\mathbf{s}}$ is invertible and $\mathbf{x}_{\mathbf{s}} = (\Phi_{\mathbf{s}}^\top \Phi_{\mathbf{s}})^{-1} \Phi_{\mathbf{s}}^\top \mathbf{y}$. Also, since $\mathbf{s} = \mathbf{s}_1 \cup \mathbf{s}_2$ we have

$$\begin{bmatrix} \mathbf{x}_{\mathbf{s}_1} \\ \mathbf{x}_{\mathbf{s}_2} \end{bmatrix} = \begin{bmatrix} \Phi_{\mathbf{s}_1}^\top \Phi_{\mathbf{s}_1} & \Phi_{\mathbf{s}_1}^\top \Phi_{\mathbf{s}_2} \\ \Phi_{\mathbf{s}_2}^\top \Phi_{\mathbf{s}_1} & \Phi_{\mathbf{s}_2}^\top \Phi_{\mathbf{s}_2} \end{bmatrix}^{-1} \begin{bmatrix} \Phi_{\mathbf{s}_1}^\top \mathbf{y} \\ \Phi_{\mathbf{s}_2}^\top \mathbf{y} \end{bmatrix}.$$

Let Ψ denote the Schur complement of block $\Phi_{s_2}^\top \Phi_{s_2}$ of the matrix $\Phi_s^\top \Phi_s$. It then follows that

$$\Psi = \Phi_{s_1}^\top \Phi_{s_1} - \Phi_{s_1}^\top \Phi_{s_2} (\Phi_{s_2}^\top \Phi_{s_2})^{-1} \Phi_{s_2}^\top \Phi_{s_1},$$

i.e., $\Psi = \Phi_{s_1}^\top (I - P\{\Phi_{s_2}\}) \Phi_{s_1}$. Since $\delta_{|s_1|+|s_2|} \leq 0.5$, we have that Ψ is invertible by Lemma 3. On the other hand, from the properties of invertible Schur complement we know that

$$\mathbf{x}_{s_1} = \Psi^{-1} \Phi_{s_1}^\top \mathbf{y} - \Psi^{-1} \Phi_{s_1}^\top \Phi_{s_2} (\Phi_{s_2}^\top \Phi_{s_2})^{-1} \Phi_{s_2}^\top \mathbf{y},$$

which can be rewritten as $\mathbf{x}_{s_1} = \Psi^{-1} \Phi_{s_1}^\top \mathbf{y}^{\perp s_2}$, thereby completing the proof. \square

Lemma 5. Fix a length k support vector \mathbf{l} , let $\ell \stackrel{\text{def}}{=} \max\{m+1, k\}$, and let \mathbf{x} be the regression of \mathbf{y} on \mathbf{l} , i.e., $\mathbf{x} = \mathcal{E}(\mathbf{l})$. If $\delta_\ell \leq 0.5$, then for any $1 \leq j \leq k$ we have

$$|\mathbf{x}_{\mathbf{l}[j]}| \leq \frac{\delta_\ell}{1 - \delta_\ell - \delta_\ell^2} \|\mathbf{x}_{s^* / \mathbf{l}}\|_2$$

whenever $\mathbf{l}[j] \notin s^*$.

Proof. Write \mathbf{l} as $\mathbf{l}[j] \cup \mathbf{l}[-j]$. Since $\delta_k \leq 0.5$ we can apply Lemma 4 and get

$$|\mathbf{x}_{\mathbf{l}[j]}| = \frac{|\Phi_{\mathbf{l}[j]}^\top \mathbf{y}^{\perp \mathbf{l}[-j]}|}{\|\Phi_{\mathbf{l}[j]}^{\perp \mathbf{l}[-j]}\|_2^2}. \quad (15)$$

We proceed by upper-bounding the numerator and lower-bounding the denominator of (15). For the numerator we have

$$\begin{aligned} |\Phi_{\mathbf{l}[j]}^\top \mathbf{y}^{\perp \mathbf{l}[-j]}| &= |\Phi_{\mathbf{l}[j]}^\top (I - P\{\Phi_{\mathbf{l}[-j]}\}) \Phi_{s^*} \mathbf{x}^*| \\ &= |\Phi_{\mathbf{l}[j]}^\top (I - P\{\Phi_{\mathbf{l}[-j]}\}) \Phi_{s^* / \mathbf{l}[-j]} \mathbf{x}_{s^* / \mathbf{l}[-j]}^*| \\ &\leq |\Phi_{\mathbf{l}[j]}^\top \Phi_{s^* / \mathbf{l}[-j]} \mathbf{x}_{s^* / \mathbf{l}[-j]}^*| + |\Phi_{\mathbf{l}[j]}^\top P\{\Phi_{\mathbf{l}[-j]}\} \Phi_{s^* / \mathbf{l}[-j]} \mathbf{x}_{s^* / \mathbf{l}[-j]}^*| \\ &\stackrel{(a)}{\leq} \delta_{m+1} \|\mathbf{x}_{s^* / \mathbf{l}[-j]}^*\|_2 + \|P\{\Phi_{\mathbf{l}[-j]}\} \Phi_{\mathbf{l}[j]}\|_2 \|P\{\Phi_{\mathbf{l}[-j]}\} \Phi_{s^* / \mathbf{l}[-j]} \mathbf{x}_{s^* / \mathbf{l}[-j]}^*\|_2 \\ &\stackrel{(b)}{\leq} \delta_{m+1} \|\mathbf{x}_{s^* / \mathbf{l}[-j]}^*\|_2 + \frac{\delta_k^2}{1 - \delta_{k-1}} \|\mathbf{x}_{s^* / \mathbf{l}[-j]}^*\|_2, \end{aligned}$$

where (a) follows from Lemma 1 and where (b) holds by Lemma 2 and the Cauchy inequality. Therefore, from the monotonicity of the RIP constant we get

$$|\Phi_{\mathbf{l}[j]}^\top \mathbf{y}^{\perp \mathbf{l}[-j]}| \leq \frac{\delta_\ell}{1 - \delta_\ell} \|\mathbf{x}_{s^* / \mathbf{l}[-j]}^*\|_2. \quad (16)$$

For the denominator of (15) we have

$$\|\Phi_{\mathbf{l}[j]}^{\perp \mathbf{l}[-j]}\|_2^2 = 1 - \|P\{\Phi_{\mathbf{l}[-j]}\} \Phi_{\mathbf{l}[j]}\|_2^2,$$

hence by Lemma 2 we get $\|\Phi_{\mathbf{l}[j]}^{\perp \mathbf{l}[-j]}\|_2^2 \geq 1 - \frac{\delta_k^2}{1 - \delta_{k-1}}$. This inequality together with inequalities (16) and (15) complete the proof. \square

Lemma 6. Let \mathbf{s} be a support vector of size k , and $e = |\mathbf{s}^* \setminus \mathbf{s}|$. Under setting (1), suppose $k + \ell > m$, $\ell \leq e$, and suppose the design matrix Φ satisfies $\delta_{k+\ell} \leq 0.5$. Let $\mathbf{l} \stackrel{\text{def}}{=} \mathcal{L}(\ell, \mathbf{s})$, $\mathbf{x} \stackrel{\text{def}}{=} \mathcal{E}(\mathbf{p})$, and $t \stackrel{\text{def}}{=} \max\{k+e, \ell+e\}$. Consider $j \in \mathbf{l}$ such that $|\hat{\mathbf{x}}_j| = \|\hat{\mathbf{x}}\|_\infty$. If $\mathbf{s}^* \not\subset \mathbf{p}$ and $j \notin \mathbf{s}^*$, then

$$\|\mathbf{x}_{\mathbf{l} \cap \mathbf{s}^*}\|_2 \leq \left(\frac{\delta_t (1 - \delta_t^2) \sqrt{2e}}{(1 - \delta_t - \delta_t^2)(1 - 2\delta_t)} - 1 \right) \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{p}}\|_2.$$

Proof. From Lemma 5 we have

$$|\hat{\mathbf{x}}_j| \leq \frac{\delta_{k+\ell}}{1 - \delta_{k+\ell} - \delta_{k+\ell}^2} \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{p}}\|_2. \quad (17)$$

On the other hand, we have

$$\begin{aligned}
\|\hat{\mathbf{x}}_l\|_\infty &= \|(\Phi_l^\top(I - P\{\Phi_s\})\Phi_l)^{-1}\Phi_l^\top\mathbf{y}^{\perp s}\|_\infty \\
&\geq \frac{1}{\sqrt{\ell}}\|(\Phi_l^\top(I - P\{\Phi_s\})\Phi_l)^{-1}\Phi_l^\top\mathbf{y}^{\perp s}\|_2 \\
&\stackrel{(a)}{\geq} \frac{1}{\sqrt{\ell}(1 + \delta_\ell)}\|\Phi_l^\top\mathbf{y}^{\perp s}\|_2 \\
&\stackrel{(b)}{\geq} \frac{1}{\sqrt{e}(1 + \delta_\ell)}\|\Phi_{s^*\setminus s}^\top\mathbf{y}^{\perp s}\|_2 \\
&\stackrel{(c)}{\geq} \frac{1}{\sqrt{e}(1 + \delta_\ell)}(\|\Phi_{s^*\setminus s}^\top\Phi_{s^*\setminus s}\mathbf{x}_{s^*\setminus s}^*\|_2 - \\
&\|\Phi_{s^*\setminus s}^\top P\{\Phi_s\}\Phi_{s^*\setminus s}\mathbf{x}_{s^*\setminus s}^*\|_2) \\
&\stackrel{(d)}{\geq} \frac{1}{\sqrt{e}(1 + \delta_\ell)}\left((1 - \delta_e)\|\mathbf{x}_{s^*\setminus s}^*\|_2 - \frac{\delta_{k+e}^2}{1 - \delta_k}\|\mathbf{x}_{s^*\setminus s}^*\|_2\right),
\end{aligned}$$

where (a) follows from Lemma 3; where (b) is a direct result from the definition of l along with the fact that $s^*\setminus p$ is non-empty and $|l| \leq e$; where (c) follows from definition of $\mathbf{y}^{\perp s}$ and the triangle inequality; and where (d) follows from the RIP definition together with Lemma 2. Next, by combining the above inequality with (17) and using the monotonicity of the RIP constant, we deduce that if $j \notin s^*$, then

$$\frac{\delta_t}{1 - \delta_t - \delta_t^2}\|\mathbf{x}_{s^*\setminus p}^*\|_2 \geq \frac{(1 - 2\delta_t)}{\sqrt{e}(1 - \delta_t^2)}\|\mathbf{x}_{s^*\setminus s}^*\|_2.$$

This implies

$$\|\mathbf{x}_{s^*\setminus s}^*\|_2 \leq \frac{\delta_t(1 - \delta_t^2)\sqrt{e}}{(1 - \delta_t - \delta_t^2)(1 - 2\delta_t)}\|\mathbf{x}_{s^*\setminus p}^*\|_2. \quad (18)$$

Now, since $s^*\setminus s = (l \cap s^*) \cup (s^*\setminus p)$, we have $\sqrt{2}\|\mathbf{x}_{s^*\setminus s}^*\|_2 \geq \|\mathbf{x}_{l \cap s^*}^*\|_2 + \|\mathbf{x}_{s^*\setminus p}^*\|_2$. Using this inequality in (18) completes the proof. \square

Lemma 7. Let s be a support vector of size k . Let $l \stackrel{\text{def}}{=} \mathcal{L}(\ell, s)$, $p \stackrel{\text{def}}{=} s \cup l$, $e \stackrel{\text{def}}{=} |s^*\setminus s|$, and $t \stackrel{\text{def}}{=} \max\{k + e, e + \ell\}$. If $s^* \not\subset p$, then

$$\left(\frac{1 + \delta_t - \delta_t^2}{1 - \delta_t}\right)\|\mathbf{x}_{l \cap s^*}^*\|_2 + \left(\frac{2\delta_t - \delta_t^2}{1 - \delta_k}\right)\|\mathbf{x}_{s^*\setminus p}^*\|_2 \geq \left(\frac{1 - 2\delta_t}{1 - \delta_t}\|\mathbf{x}_{s^*\setminus p}^*\|_2 - \frac{\delta_t}{1 - \delta_t}\|\mathbf{x}_{s^* \cap l}^*\|_2\right) \cdot \sqrt{\frac{\ell}{|s^*\setminus p|}}$$

Proof of Lemma 7. From the definition of $\mathcal{L}(\ell, s)$ and the fact that $s^*\setminus p$ is non-empty, we have $\frac{\|\Phi_l^\top\mathbf{y}^{\perp s}\|_2}{\sqrt{\ell}} \geq \|\Phi_{s^*\setminus p}^\top\mathbf{y}^{\perp s}\|_\infty$.

Then, since $\|\Phi_{s^*\setminus p}^\top\mathbf{y}^{\perp s}\|_\infty \geq \frac{\|\Phi_{s^*\setminus p}^\top\mathbf{y}^{\perp s}\|_2}{\sqrt{|s^*\setminus p|}}$, we get

$$\|\Phi_l^\top\mathbf{y}^{\perp s}\|_2 \geq \sqrt{\frac{\ell}{|s^*\setminus p|}}\|\Phi_{s^*\setminus p}^\top\mathbf{y}^{\perp s}\|_2. \quad (19)$$

Now, on the hand we have

$$\begin{aligned}
\|\Phi_{s^*\setminus p}^\top\mathbf{y}^{\perp s}\|_2 &= \|\Phi_{s^*\setminus p}^\top(I - P\{\Phi_s\})\Phi_{s^*\setminus s}\mathbf{x}_{s^*\setminus s}^*\|_2 \\
&\geq \|\Phi_{s^*\setminus p}^\top\Phi_{s^*\setminus s}\mathbf{x}_{s^*\setminus s}^*\|_2 - \|\Phi_{s^*\setminus p}^\top P\{\Phi_s\}\Phi_{s^*\setminus s}\mathbf{x}_{s^*\setminus s}^*\|_2 \\
&\stackrel{(a)}{\geq} \|\Phi_{s^*\setminus p}^\top\Phi_{s^*\setminus s}\mathbf{x}_{s^*\setminus s}^*\|_2 - \frac{\delta_{k+e}^2}{1 - \delta_k}\|\mathbf{x}_{s^*\setminus s}^*\|_2 \\
&\stackrel{(b)}{\geq} \|\Phi_{s^*\setminus p}^\top\Phi_{s^*\setminus p}\mathbf{x}_{s^*\setminus p}^*\|_2 - \|\Phi_{s^*\setminus p}^\top\Phi_{s^* \cap l}\mathbf{x}_{s^* \cap l}^*\|_2 - \frac{\delta_{k+e}^2}{1 - \delta_k}\|\mathbf{x}_{s^*\setminus s}^*\|_2 \\
&\stackrel{(c)}{\geq} \|\Phi_{s^*\setminus p}^\top\Phi_{s^*\setminus p}\mathbf{x}_{s^*\setminus p}^*\|_2 - \delta_e\|\mathbf{x}_{s^* \cap l}^*\|_2 - \frac{\delta_{k+e}^2}{1 - \delta_k}\|\mathbf{x}_{s^*\setminus s}^*\|_2 \\
&\stackrel{(d)}{\geq} (1 - \delta_{|s^*\setminus p|})\|\mathbf{x}_{s^*\setminus p}^*\|_2 - \delta_e\|\mathbf{x}_{s^* \cap l}^*\|_2 - \frac{\delta_{k+e}^2}{1 - \delta_k}\|\mathbf{x}_{s^*\setminus s}^*\|_2 \\
&\stackrel{(e)}{\geq} (1 - \delta_e)\|\mathbf{x}_{s^*\setminus p}^*\|_2 - \delta_e\|\mathbf{x}_{s^* \cap l}^*\|_2 - \frac{\delta_{k+e}^2}{1 - \delta_k}\|\mathbf{x}_{s^*\setminus s}^*\|_2 \\
&\stackrel{(b)}{\geq} (1 - \delta_e - \frac{\delta_{k+e}^2}{1 - \delta_k})\|\mathbf{x}_{s^*\setminus p}^*\|_2 - (\delta_e + \frac{\delta_{k+e}^2}{1 - \delta_k})\|\mathbf{x}_{s^* \cap l}^*\|_2,
\end{aligned} \quad (20)$$

where (a) holds by Lemma 2; where (b) follows from the triangle inequality and the identity $\mathbf{s}^* \setminus \mathbf{s} = (\mathbf{s}^* \cap \mathbf{l}) \cup (\mathbf{s}^* \setminus \mathbf{p})$; where (c) is a direct result of Lemma 1; where (d) follows from the definition of RIP constant; and where (e) holds by the monotonicity of the RIP constant and the fact that $|\mathbf{s}^* \setminus \mathbf{p}| \leq e$.

On the other hand, we have

$$\begin{aligned}
\|\Phi_{\mathbf{l}}^{\top} \mathbf{y}^{\perp \mathbf{s}}\|_2 &= \|\Phi_{\mathbf{l}}^{\top} (I - P\{\Phi_{\mathbf{s}}\}) \Phi_{\mathbf{s}^*} \mathbf{x}_{\mathbf{s}^*}^*\|_2 \\
&= \|\Phi_{\mathbf{l}}^{\top} (I - P\{\Phi_{\mathbf{s}}\}) \Phi_{\mathbf{s}^* \setminus \mathbf{s}} \mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 \\
&\leq \|\Phi_{\mathbf{l}}^{\top} \Phi_{\mathbf{s}^* \setminus \mathbf{s}} \mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 + \|\Phi_{\mathbf{l}}^{\top} P\{\Phi_{\mathbf{s}}\} \Phi_{\mathbf{s}^* \setminus \mathbf{s}} \mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 \\
&\stackrel{(a)}{\leq} \|\Phi_{\mathbf{l}}^{\top} \Phi_{\mathbf{s}^* \setminus \mathbf{s}} \mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 + \frac{\delta_{e+k} \delta_{k+\ell}}{1 - \delta_k} \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 \\
&\stackrel{(b)}{\leq} \|\Phi_{\mathbf{l} \setminus \mathbf{s}^*}^{\top} \Phi_{\mathbf{s}^* \setminus \mathbf{s}} \mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 + \|\Phi_{\mathbf{l} \cap \mathbf{s}^*}^{\top} \Phi_{\mathbf{s}^* \setminus \mathbf{s}} \mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 + \frac{\delta_{e+k} \delta_{k+\ell}}{1 - \delta_k} \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 \\
&\stackrel{(c)}{\leq} (\delta_{\ell+e} + \frac{\delta_{e+k} \delta_{k+\ell}}{1 - \delta_k}) \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 + \|\Phi_{\mathbf{l} \cap \mathbf{s}^*}^{\top} \Phi_{\mathbf{s}^* \setminus \mathbf{s}} \mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 \\
&\stackrel{(d)}{\leq} (\delta_{\ell+e} + \frac{\delta_{e+k} \delta_{k+\ell}}{1 - \delta_k}) \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 + \|\Phi_{\mathbf{l} \cap \mathbf{s}^*}^{\top} \Phi_{\mathbf{l} \cap \mathbf{s}^*} \mathbf{x}_{\mathbf{l} \cap \mathbf{s}^*}^*\|_2 + \|\Phi_{\mathbf{l} \cap \mathbf{s}^*}^{\top} \Phi_{\mathbf{s}^* \setminus \mathbf{p}} \mathbf{x}_{\mathbf{s}^* \setminus \mathbf{p}}^*\|_2 \\
&\stackrel{(e)}{\leq} (\delta_{\ell+e} + \frac{\delta_{e+k} \delta_{k+\ell}}{1 - \delta_k}) \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 + (1 + \delta_e) \|\mathbf{x}_{\mathbf{l} \cap \mathbf{s}^*}^*\|_2 + \|\Phi_{\mathbf{l} \cap \mathbf{s}^*}^{\top} \Phi_{\mathbf{s}^* \setminus \mathbf{p}} \mathbf{x}_{\mathbf{s}^* \setminus \mathbf{p}}^*\|_2 \\
&\stackrel{(f)}{\leq} (\delta_{\ell+e} + \frac{\delta_{e+k} \delta_{k+\ell}}{1 - \delta_k}) \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{s}}^*\|_2 + (1 + \delta_e) \|\mathbf{x}_{\mathbf{l} \cap \mathbf{s}^*}^*\|_2 + \delta_e \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{p}}^*\|_2 \\
&\leq (1 + \delta_e + \delta_{\ell+e} + \frac{\delta_{e+k} \delta_{k+\ell}}{1 - \delta_k}) \|\mathbf{x}_{\mathbf{l} \cap \mathbf{s}^*}^*\|_2 + (\delta_e + \delta_{\ell+e} + \frac{\delta_{e+k} \delta_{k+\ell}}{1 - \delta_k}) \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{p}}^*\|_2, \tag{21}
\end{aligned}$$

where (a) follows from Lemma 2; where (b) follows from the triangle inequality applied to identity $\mathbf{l} = (\mathbf{l} \setminus \mathbf{s}^*) \cup (\mathbf{l} \cap \mathbf{s}^*)$; where (c) follows from Lemma 1 and from the fact that $|\mathbf{l} \setminus \mathbf{s}^*| \leq \ell$; where (d) follows from the triangle inequality applied to identity $\mathbf{s}^* \setminus \mathbf{s} = (\mathbf{s}^* \setminus \mathbf{p}) \cup (\mathbf{s}^* \cap \mathbf{l})$; where (e) follows from the definition of RIP and the fact that $|\mathbf{s}^* \cap \mathbf{l}| \leq e$; and where (f) follows from Lemma 1 and the fact that $e = |\mathbf{s}^* \cap \mathbf{l}| + |\mathbf{s}^* \setminus \mathbf{p}|$.

Finally, combine (20), (21), and (19) along with $t = \max\{k + e, \ell + e\}$ and the monotonicity of the RIP constant to get the desired result. \square

Proposition 1. Consider the setting in (1). Let \mathbf{s} be a support vector of size k and let $e \stackrel{\text{def}}{=} |\mathbf{s}^* \setminus \mathbf{s}|$. For an integer ℓ that satisfies $\ell \leq e$ and $k + \ell > m$ let $\mathbf{l} \stackrel{\text{def}}{=} \mathcal{L}(\ell, \mathbf{s})$, $\mathbf{p} \stackrel{\text{def}}{=} \mathbf{l} \cup \mathbf{s}$, $\hat{\mathbf{x}} \stackrel{\text{def}}{=} \mathcal{E}(\mathbf{p})$, and $t \stackrel{\text{def}}{=} \max\{k + e, \ell + e\}$. Suppose $j \in \mathbf{l}$ is such that $|\hat{\mathbf{x}}_j| = \|\hat{\mathbf{x}}_{\mathbf{l}}\|_{\infty}$. Furthermore, assume $\delta_{k+\ell} \leq 0.5$ and $\mathbf{s}^* \not\subset \mathbf{p}$. If the two inequalities

$$\sqrt{e} \leq \frac{\sqrt{2}(1 - \delta_t - \delta_t^2)(1 - 2\delta_t)}{(1 - \delta_t^2 + 2\delta_t)\delta_t(1 + \delta_t)},$$

and

$$\sqrt{\ell} > \frac{(1 - \delta_t^2 + \delta_t)(\eta\sqrt{e} - 1) + 2\delta_t - \delta_t^2}{1 - 2\delta_t - \delta_t(\eta\sqrt{e} - 1)} \sqrt{e}$$

hold with $\eta \stackrel{\text{def}}{=} \frac{(1 - \delta_t^2)\delta_t\sqrt{2}}{(1 - \delta_t - \delta_t^2)(1 - 2\delta_t)}$, then $j \in \mathbf{s}^*$.

Proof. Assume $j \notin \mathbf{s}^*$. Then, from Lemma 6 we get

$$\|\mathbf{x}_{\mathbf{l} \cap \mathbf{s}^*}^*\|_2 \leq (\eta\sqrt{e} - 1) \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{p}}^*\|_2. \tag{22}$$

From (22) and Lemma 7 we get

$$\|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{p}}^*\|_2 \left(\frac{1 + \delta_t - \delta_t^2}{1 - \delta_t} (\eta\sqrt{e} - 1) + \frac{2\delta_t - \delta_t^2}{1 - \delta} \right) \geq \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{p}}^*\|_2 \left(\frac{1 - 2\delta_t}{1 - \delta_t} - \frac{\delta_t}{1 - \delta_t} (\eta\sqrt{e} - 1) \right) \cdot \sqrt{\frac{\ell}{|\mathbf{s}^* \setminus \mathbf{p}|}}.$$

Next, since $1 - 2\delta_t \geq \delta_t(\eta\sqrt{e} - 1)$ and $|\mathbf{s}^* \setminus \mathbf{p}| \leq e$ we get

$$\left(\frac{1 + \delta_t - \delta_t^2}{1 - \delta_t} (\eta\sqrt{e} - 1) + \frac{2\delta_t - \delta_t^2}{1 - \delta} \right) \geq \left(\frac{1 - 2\delta_t}{1 - \delta_t} - \frac{\delta_t}{1 - \delta_t} (\eta\sqrt{e} - 1) \right) \sqrt{\frac{\ell}{e}},$$

which contradicts the assumed lower bound on ℓ . Therefore the initial assumption cannot hold and j must belong to \mathbf{s}^* . \square

Proof of Theorem 1. We start with the first step of LiRE ($i = 1$). Let $\mathbf{s} \stackrel{\text{def}}{=} \mathbf{s}_{\text{out}}[-1]$, $\mathbf{l} \stackrel{\text{def}}{=} \mathcal{L}(\ell, \mathbf{s})$, $\mathbf{p} \stackrel{\text{def}}{=} \mathbf{l} \cup \mathbf{s}$, and $\hat{\mathbf{x}} \stackrel{\text{def}}{=} \mathcal{E}(\mathbf{p})$. Pick $j \in \mathbf{l}$ such that $|\hat{\mathbf{x}}_j| = \|\hat{\mathbf{x}}_{\mathbf{l}}\|_{\infty}$. Note that since $|\mathbf{s}^* \setminus \mathbf{s}| \leq |\mathbf{s}^* \setminus \mathbf{s}_{\text{out}}| + 1$, we have $|\mathbf{s}^* \setminus \mathbf{s}| \leq e + 1$. We have the following three possible cases:

Case 1: $\mathbf{s}^* \not\subset \mathbf{p}$. Here the assumptions of Proposition 1 are satisfied and therefore $j \in \mathbf{s}^*$. This implies $\mathbf{s}_{\text{out}}^*[1] \in \mathbf{s}^*$, hence the first component of the estimated support is corrected.

Case 2: $\mathbf{s}^* \subset \mathbf{p}$ and $\mathbf{s}^* \subset \mathbf{s}_{\text{out}}[-1]$. Since $\mathbf{s}^* \subset \mathbf{s}_{\text{out}}[-1]$, we have $\mathbf{y}^{\perp \mathbf{s}_{\text{out}}[-1]} = 0$ and LiRE will exit the **for** loop and output \mathbf{s}_{out} . Moreover, we get $\mathbf{s}^* \subset \mathbf{s}_{\text{out}}$ which completes the proof.

Case 3: $\mathbf{s}^* \subset \mathbf{p}$ and $\mathbf{s}^* \not\subset \mathbf{s}_{\text{out}}[-1]$. The next lemma characterizes any true feature in the support estimate.

Lemma 8. *If $\mathbf{s}^* \subset \mathbf{p}$ and $\mathbf{s}^* \not\subset \mathbf{s}_{\text{out}}[-1]$, then for every i in \mathbf{l} we have $i \in \mathbf{s}^*$ if and only if $|\hat{\mathbf{x}}_i| > 0$.*

We have $\mathbf{s}^* \subset \mathbf{s}_{\text{out}}[-1] \cup \mathbf{l}$ and since $\mathbf{s}^* \not\subset \mathbf{s}_{\text{out}}[-1]$ we deduce that there must exist some feature $i \in \mathbf{l} \cap \mathbf{s}^*$. By Lemma 8, we get $|\hat{\mathbf{x}}_i| > 0$. Moreover, the assumption that $|\hat{\mathbf{x}}_j| = \|\hat{\mathbf{x}}_{\mathbf{l}}\|_{\infty}$ implies $|\hat{\mathbf{x}}_j| \geq |\hat{\mathbf{x}}_i|$, hence $|\hat{\mathbf{x}}_j| > 0$. From Lemma 8 we then conclude that $j \in \mathbf{s}^*$. This means that LiRE will add a true feature in this case as well.

The above argument holds for the first iteration of the **for** loop of LiRE. However, as we are not adding errors to \mathbf{s}_{out} , the same argument can be used for the other iterations as well. As there are at most m missed true features outside \mathbf{s}_{in} , after one round all of them will be included and therefore $\mathbf{s}^* \subset \mathbf{s}_{\text{out}}$.

Proof of Lemma 8. Pick $i \in \mathbf{l}$. If $i \notin \mathbf{s}^*$, next since $\delta_{m+\ell} < 0.5$, we can apply Lemma 5 and get $|\hat{\mathbf{x}}_i| \leq \frac{\delta_{m+\ell}}{1 - \delta_{m+\ell} - \delta_{m+\ell}^2} \|\mathbf{x}_{\mathbf{s}^* \setminus \mathbf{p}}^*\|_2$, then this implies $|\hat{\mathbf{x}}_i| = 0$. It remains to show the converse, namely that if $i \in \mathbf{s}^* \cap \mathbf{l}$ then $|\hat{\mathbf{x}}_i| > 0$. Suppose that there exists $i \in \mathbf{s}^* \cap \mathbf{l}$ such that $|\hat{\mathbf{x}}_i| = 0$. From Lemma 3 we get

$$|\hat{\mathbf{x}}_i| = \frac{|\Phi_i^T \mathbf{y}^{\perp \mathbf{p}/\{i\}}|}{\|\Phi_i^{\perp \mathbf{p}/\{i\}}\|_2^2}, \quad (23)$$

which yields $|\Phi_i^T \mathbf{y}^{\perp \mathbf{p}/\{i\}}| = 0$. Next, by expanding (23) we get

$$|\Phi_i^T (I - P\{\Phi_{\mathbf{p}/\{i\}}\}) \Phi_{\mathbf{s}^*} \mathbf{x}_{\mathbf{s}^*}^*| = 0.$$

Then, since $\mathbf{s}^* \subset \mathbf{p}$ and $i \in \mathbf{s}^*$, we deduce that

$$|\Phi_i^T (I - P\{\Phi_{\mathbf{p}/\{i\}}\}) \Phi_i \mathbf{x}_i^*| = 0. \quad (24)$$

On the other hand, we have

$$\begin{aligned} |\Phi_i^T (I - P\{\Phi_{\mathbf{p}/\{i\}}\}) \Phi_i \mathbf{x}_i^*| &\geq |\Phi_i^T \Phi_i \mathbf{x}_i^*| - |\Phi_i^T P\{\Phi_{\mathbf{p}/\{i\}}\} \Phi_i \mathbf{x}_i^*| \\ &\stackrel{(a)}{\geq} |\mathbf{x}_i^*| - \frac{\delta_{m+\ell}^2}{1 - \delta_{m+\ell}} |\mathbf{x}_i^*|, \end{aligned}$$

where for (a) we used Lemma 2. This implies $|\Phi_i^T (I - P\{\Phi_{\mathbf{p}/\{i\}}\}) \Phi_i \mathbf{x}_i^*| > 0$ and contradicts (24). Hence, if $i \in \mathbf{l} \cap \mathbf{s}^*$, then $|\hat{\mathbf{x}}_i| > 0$, which completes the proof. \square

\square

V. CONCLUDING REMARKS

We proposed LiRE, a low complexity error-correction module for sparse recovery algorithms, and provided sufficient conditions under which LiRE corrects all errors made by the baseline algorithm. Simulations show that LiRE may boost the performance of low complexity greedy algorithms to attain the performance of significantly more complex ones, as we saw in the comparison of LiRE \circ OMP vs. BP. Alternatively, LiRE may be used as a fast standalone support recovery algorithm that is competitive against OMP. Interesting venues for future research are in order:

- Theorem 1 provides a sufficient condition for error-correction that appears conservative in light of the numerical experiments. For instance, Theorem 1 implies that if we want to correct up to \bar{e} errors (as opposed to correcting exactly e errors), then because of Condition (3) the list size should be equal to one. This, in turn, yields Corollary 2 which appears to be loose when the number of potential errors is large, linear in m —see Section III-D. We also observe that the list size used for the numerical experiments appears to be a robust choice across sparsity levels. Improving Theorem 1 by potentially relaxing Condition (3) is a natural direction for future investigation.
- A very interesting problem is to quantify the successive refinement of the support estimate obtained by running LiRE multiple times.
- LiRE keeps the list size constant throughout its iterations. However, as iterations proceed there are fewer and fewer errors to be corrected and since the list size impacts the likelihood of changing a feature, it might be possible to improve performance by considering an adaptive list size.
- Theoretical guarantees in the noisy measurement setup would yield interesting and non-trivial extensions of the results presented in this paper.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Venkat Chandar for his insightful comments.

REFERENCES

- [1] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, "A sharp condition for exact support recovery with orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1370–1382, 2016.
- [2] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus mathématique*, vol. 346, no. 9-10, pp. 589–592, 2008.
- [3] D. Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and computational harmonic analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [4] M. Fornasier, "Numerical methods for sparse recovery," *Theoretical foundations and numerical methods for sparse recovery*, vol. 14, pp. 93–200, 2010.
- [5] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani *et al.*, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [6] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM journal on computing*, vol. 24, no. 2, pp. 227–234, 1995.
- [7] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [8] S. N. Negahban, P. Ravikumar, M. J. Wainwright, B. Yu *et al.*, "A unified framework for high-dimensional analysis of m -estimators with decomposable regularizers," *Statistical science*, vol. 27, no. 4, pp. 538–557, 2012.
- [9] A. Agarwal, S. Negahban, and M. J. Wainwright, "Fast global convergence rates of gradient methods for high-dimensional statistical recovery," in *Advances in Neural Information Processing Systems*, 2010, pp. 37–45.
- [10] A. Bora, A. Jalal, E. Price, and A. G. Dimakis, "Compressed sensing using generative models," in *International Conference on Machine Learning*, 2017, pp. 537–546.
- [11] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [12] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [13] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and computational harmonic analysis*, vol. 27, no. 3, pp. 265–274, 2009.
- [14] S. Foucart, "Hard thresholding pursuit: an algorithm for compressive sensing," *SIAM Journal on Numerical Analysis*, vol. 49, no. 6, pp. 2543–2563, 2011.
- [15] S. Kwon, J. Wang, and B. Shim, "Multipath matching pursuit," *IEEE Trans. Information Theory*, vol. 60, no. 5, pp. 2986–3001, 2014.
- [16] R. Khanna and A. Kyrillidis, "lht dies hard: Provable accelerated iterative hard thresholding," *arXiv preprint arXiv:1712.09379*, 2017.
- [17] J. Shen and S. Mousavi, "Least sparsity of p -norm based optimization problems with $p > 1$," *SIAM Journal on Optimization*, vol. 28, no. 3, pp. 2721–2751, 2018.
- [18] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on information theory*, vol. 52, no. 1, pp. 6–18, 2005.
- [19] S. Negahban and M. J. Wainwright, "Restricted strong convexity and weighted matrix completion: Optimal bounds with noise," *Journal of Machine Learning Research*, vol. 13, no. May, pp. 1665–1697, 2012.
- [20] R. Somani, C. Gupta, P. Jain, and P. Netrapalli, "Support recovery for orthogonal matching pursuit: upper and lower bounds," in *Advances in Neural Information Processing Systems*, 2018, pp. 10814–10824.
- [21] P. Jain, A. Tewari, and I. S. Dhillon, "Partial hard thresholding," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 3029–3038, 2017.
- [22] Y.-B. Zhao and Z.-Q. Luo, "Analysis of optimal thresholding algorithms for compressed sensing," *arXiv preprint arXiv:1912.10258*, 2019.
- [23] S. Foucart and S. Subramanian, "Iterative hard thresholding for low-rank recovery from rank-one projections," *Linear Algebra and its Applications*, vol. 572, pp. 117–134, 2019.
- [24] S. Aeron, V. Saligrama, and M. Zhao, "Information theoretic bounds for compressed sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5111–5130, 2010.
- [25] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3065–3092, 2012.
- [26] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.
- [27] B. Mirzasoleiman, A. Karbasi, R. Sarkar, and A. Krause, "Distributed submodular maximization: Identifying representative elements in massive data," *Advances in Neural Information Processing Systems*, vol. 26, pp. 2049–2057, 2013.
- [28] R. Khanna, E. Elenberg, A. Dimakis, S. Negahban, and J. Ghosh, "Scalable greedy feature selection via weak submodularity," in *Artificial Intelligence and Statistics*, 2017, pp. 1560–1568.
- [29] M. Mohammad and A. Tchamkerten, "Error-correction for sparse support recovery algorithms," *submitted to ISIT 2021*.
- [30] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [31] J. Wang, S. Kwon, and B. Shim, "Generalized orthogonal matching pursuit," *IEEE Transactions on signal processing*, vol. 60, no. 12, pp. 6202–6216, 2012.
- [32] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo, "A sharp condition for exact support recovery with orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1370–1382, 2017.
- [33] J. Wang and B. Shim, "On the recovery limit of sparse signals using orthogonal matching pursuit," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4973–4976, 2012.