Regularized Modified BPDN for Noisy Sparse Reconstruction with Partial Erroneous Support and Signal Value Knowledge

Wei Lu and Namrata Vaswani

Abstract

We study the problem of sparse reconstruction from noisy undersampled measurements when the following knowledge is available. (1) We are given partial, and partly erroneous, knowledge of the signal's support, denoted by T. (2) We are also given an erroneous estimate of the signal values on T, denoted by $(\hat{\mu})_T$. In practice, both of these may be available from prior knowledge. Alternatively, in recursive reconstruction applications, like real-time dynamic MRI, one can use the support estimate and the signal value estimate from the previous time instant as T and $(\hat{\mu})_T$. In this work, we introduce regularized modified-BPDN (reg-mod-BPDN) to solve this problem and obtain computable bounds on its reconstruction error. Reg-mod-BPDN tries to find the signal that is sparsest outside the set T, while being "close enough" to $(\hat{\mu})_T$ on T and while satisfying the data constraint. Corresponding results for modified-BPDN and BPDN follow as direct corollaries. A second key contribution is an approach to obtain computable error bounds that hold without any sufficient conditions. This makes it easy to compare the bounds for the various approaches. Empirical reconstruction error comparisons with many existing approaches are also provided.

Index Terms

compressive sensing, sparse reconstruction, modified-CS, partially known support

I. INTRODUCTION

The goal of this work is to solve the sparse recovery problem [2], [3], [4], [5], [6]. We try to reconstruct an m-length sparse vector, x, with support, N, from an $n < m$ length noisy measurement vector, y,

satisfying

$$y \triangleq Ax + w \tag{1}$$

when the following two things are available: (i) partial, and partly erroneous, knowledge of the signal's support, denoted by T; and (ii) an erroneous estimate of the signal values on T, denoted by $(\hat{\mu})_T$. In (1), w is the measurement noise and A is the measurement matrix. For simplicity, in this work, we just refer to x *as the signal* and to A *as the measurement matrix*. However, in general, x is the sparsity basis vector (which is either the signal itself or some linear transform of the signal) and $A = H\Phi$ where H is the measurement matrix and Φ is the sparsity basis matrix. If Φ is the identity matrix then x is the signal itself.

The true support of the signal, N, can be rewritten as

$$N = T \cup \Delta \setminus \Delta_e \tag{2}$$

where

$$\Delta \triangleq N \setminus T \text{ and } \Delta_e \triangleq T \setminus N \tag{3}$$

are the errors in the support estimate, T^c is the complement set of T and $\setminus$ is the set difference notation $(N \setminus T \triangleq N \cap T^c)$.

The signal estimate is assumed to be zero along T^c, i.e.

$$\hat{\mu} = \begin{bmatrix} (\hat{\mu})_T \\ \mathbf{0}_{T^c} \end{bmatrix} \tag{4}$$

and the signal itself can be rewritten as

$$(x)_{N \cup T} = (\hat{\mu})_{N \cup T} + e$$

$$(x)_{N^c} = 0 \tag{5}$$

where e denotes the error in the prior signal estimate. It is assumed that the error energy, $\|e\|_2^2$, is small compared to the signal energy, $\|x\|_2^2$.

In practical applications, T and $\hat{\mu}$ may be available from prior knowledge. Alternatively, in applications requiring recursive reconstruction of (approximately) sparse signal or image sequences, with slow time-varying sparsity patterns and slow changing signal values, one can use the support estimate and the signal value estimate from the previous time instant as the "prior knowledge". A key domain where this problem occurs is in fast (recursive) dynamic MRI reconstruction from highly undersampled measurements. In MRI, we typically assume that the images are wavelet sparse. We show slow support and signal value change for two medical image sequences in Fig. 1. From the figure, we can see that the maximum support changes for both sequences are less than 2% of the support size and almost all signal values' changes are less than 0.16% of the signal energy. Slow signal value change also implies that a signal value is small before it gets removed from the support. Other potential applications include single-pixel camera based real-time video imaging [7]; video compression; ReProCS (recursive projected CS) based video
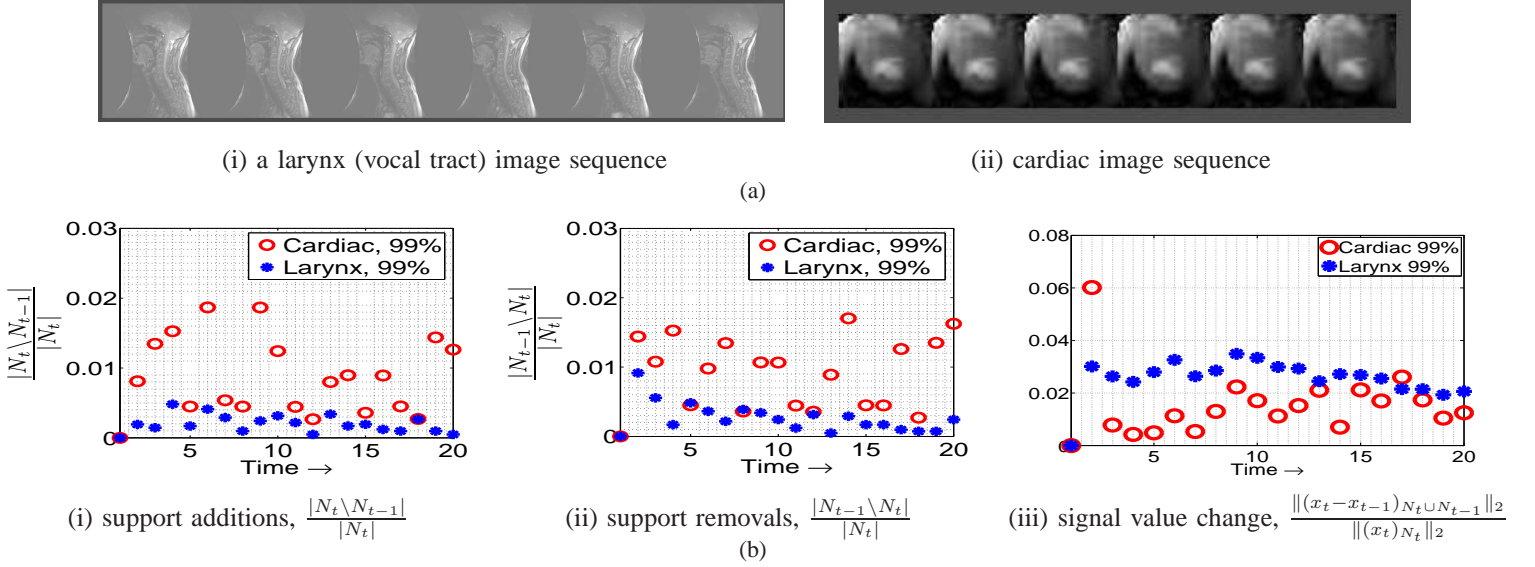
(i) a larynx (vocal tract) image sequence   (ii) cardiac image sequence

(a)

(i) support additions, $\frac{|N_t \setminus N_{t-1}|}{|N_t|}$ (ii) support removals, $\frac{|N_{t-1} \setminus N_t|}{|N_t|}$ (iii) signal value change, $\frac{\|(x_t - x_{t-1})_{N_t \cup N_{t-1}}\|_2}{\|(x_t)_{N_t}\|_2}$

(b)

Fig. 1. In (a), we show two medical image sequences (a cardiac and a larynx sequence). In (b), x_t is the two-level Daubechies-4 2D discrete wavelet transform (DWT) of the cardiac or the larynx image at time t and the set N_t is its 99% energy support (the smallest set containing 99% of the vector's energy). Its size, $|N_t|$ varied between 4121-4183 ($\approx 0.07m$) for larynx and between 1108-1127 ($\approx 0.06m$) for cardiac. *Notice that all support changes are less than 2% of the support size and almost all signal values changes are less than 4% of $\|(x_t)_{N_t}\|_2$.*

denoising or video layering (separating video in foreground and background layers) [8], [9]; and spectral domain optical coherence tomography [10] based dynamic imaging.

This work has the following contributions.

1) We introduce regularized modified-BPDN (reg-mod-BPDN) and obtain a computable bound on its reconstruction error using an approach motivated by [3]. Reg-mod-BPDN solves

$$\min_b \quad \gamma \|b_{T^c}\|_1 + \frac{1}{2}\|y - Ab\|_2^2 + \frac{1}{2}\lambda\|b_T - \hat{\mu}_T\|_2^2 \qquad (6)$$

i.e. it tries to find the signal that is sparsest outside the set T, while being "close enough" to $\hat{\mu}_T$ on T, and while satisfying the data constraint. Reg-mod-BPDN uses the fact that T is a good estimate of the true support, N, and that $\hat{\mu}_T$ is a good estimate of x_T. In particular, for $i \in \Delta_e$, this implies that $|\hat{\mu}_i|$ is close to zero (since $x_i = 0$ for $i \in \Delta_e$).

2) Our second key contribution is to show how to use the reconstruction error bound result to obtain another computable bound that holds without any sufficient conditions and is tighter. This allows easy bound comparisons of the various approaches. A similar result for mod-BPDN and BPDN follows as a direct corollary.

3) Reconstruction error comparisons with these and many other existing approaches are also shown.

A. Notations and Problem Definition

For any set T and vector b, b_T denotes a sub-vector containing the elements of b with indices in T. $\|b\|_k$ refers to the ℓ_k norm of the vector b. Also, $\|b\|_0$ counts the number of nonzero elements of b.

The notation T^c denotes the set complement of T, i.e., $T^c = \{i \in [1, ..., m], i \notin T\}$. $\emptyset$ is the empty set.

We use $'$ for transpose. For the matrix A, A_T denotes the sub-matrix containing the columns of A with indices in T. The matrix norm $\|A\|_p$, is defined as $\|A\|_p \triangleq \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$. I_T is an identity matrix on the set of rows and columns indexed by elements in T. $\mathbf{0}_{T,S}$ is a zero matrix on the set of rows and columns indexed by elements in T and S respectively.

The notation $\nabla L(b)$ denotes the gradient of the function $L(b)$ with respect to b.

When we say b is supported on $T \cup S$ we mean that the support of b (set of indices where b is nonzero) is a subset of $T \cup S$.

Our goal is to reconstruct a sparse vector, x, with support, N, from the noisy measurement vector, y satisfying (1). We assume partial knowledge of the support, denoted by T, and of the signal estimate on T, denoted by $(\hat{\mu})_T$. The support estimate may contain errors – misses, Δ, and extras, Δ_e, defined in (3). The signal estimate, $\hat{\mu}$, is assumed to be zero along T^c, i.e it satisfies (4) and the signal, x, satisfies (5).

B. Related Work

The sparse reconstruction problem, without using any support or signal value knowledge, has been studied for a long time [2], [3], [4], [5], [6]. It tries to find the sparsest signal among all signals that satisfy the data constraint, i.e. it solves $\min_b \|b\|_0$ $s.t.$ $y = A\beta$. This brute-force search has exponential complexity. One class of practical approaches to solve this is basis pursuit which replaces $\|b\|_0$ by $\|b\|_1$ [2]. The ℓ_1 norm is the closest norm to ℓ_0 that makes the problem convex. For noisy measurements, the data constraint becomes an inequality constraint. However, this assumes that the noise is bounded and the noise bound is available. In practical applications where this may not be available, one can use the Lagrangian version which solves

$$\min_b \ \gamma\|b\|_1 + \frac{1}{2}\|y - Ab\|_2^2 \tag{7}$$

This is called *basis pursuit denoising (BPDN) [2]*. Since this solves an unconstrained optimization problem, it is also faster. An error bound of BPDN was obtained in [3]. Error bounds for its constrained version were obtained in [11], [12].

The problem of sparse reconstruction with partial support knowledge was introduced in our work [13], [14]; and also in parallel in Khajehnejad et al [15] and in vonBorries et al [16]. In [13], [14], we proposed an approach called *modified-CS* which tries to find the signal that is sparsest outside the set T and satisfies the data constraint. We obtained exact reconstruction conditions for it by using the restricted isometry approach [17]. When measurements are noisy, for the same reasons as above, one can use the

Lagrangian version:

$$\min_b \quad \gamma \|b_{T^c}\|_1 + \frac{1}{2}\|y - Ab\|_2^2 \tag{8}$$

We call this *modified-BPDN (mod-BPDN)*. Its error was bounded in the conference version of this work [1], while the error of its constrained version was bounded in Jacques [18].

In [15], Khajehnejad et al assumed a probabilistic support prior and proposed a weighted ℓ_1 solution. They also obtained exact reconstruction thresholds for weighted ℓ_1 by using the overall approach of Donoho [19]. In Fig. 2, we show comparisons with the noisy Lagrangian version of *weighted ℓ_1* which solves:

$$\min_b \quad \gamma \|b_{T^c}\|_1 + \gamma'\|b_T\|_1 + \frac{1}{2}\|y - Ab\|_2^2 \tag{9}$$

Our earlier work on Least Squares CS-residual (LS-CS) and Kalman Filtered CS-residual (KF-CS) [20], [21] can also be interpreted as a possible solution for the current problem, although it was proposed in the context of recursive reconstruction of sparse signal sequences.

Reg-mod-BPDN may also be interpreted as a Bayesian CS or a model-based CS approach. Recent work in this area includes [22], [23], [24], [25], [26], [27], [28].

C. Some Related Approaches**

Before going further, we discuss below *a few approaches that are related to, but different from reg-mod-BPDN, and we argue when and why these will be worse than reg-mod-BPDN*. This section may be skipped on a quick reading. We show comparisons with all these in Fig. 2.

The first is what can be called *CS-residual or CS-diff* which computes

$$\hat{x} = \hat{\mu} + \hat{b}, \text{ where } \hat{b} \text{ solves}$$

$$\min_b \quad \gamma \|b\|_1 + \frac{1}{2}\|y - A\hat{\mu} - Ab\|_2^2 \tag{10}$$

This has the following limitation. It does not use the fact that when T is an accurate estimate of the true support, $(x)_{T^c}$ is much more sparse compared with the full $(x - \hat{\mu})$ (the support size of x_{T^c} is $|\Delta|$ while that of $(x - \hat{\mu})$ is $|T| + |\Delta|$ which is much larger). The exception is if the signal value prior is so strong that $(x - \hat{\mu})$ is zero (or very small) on all or a part of T.

CS-residual is also related to LS-CS and KF-CS. LS-CS solves (10) but with $\hat{\mu}_T$ being the LS estimate computed assuming that the signal is supported on T and with $(\hat{\mu})_{T^c} = 0$. For a static problem, KF-CS can be interpreted as computing the regularized LS estimate on T and using that as $\hat{\mu}_T$. LS-CS and KF-CS also have a limitation similar to CS-residual.

Another seemingly related approach is what can be called *CS-mod-residual*. It computes

$$\hat{x}_T = \hat{\mu}_T, \ \hat{x}_{T^c} = \hat{b}_c, \text{ where } \hat{b}_c \text{ solves}$$

$$\min_{b_c} \quad \frac{1}{2}\|y - A_T\hat{\mu}_T - A_{T^c}b_c\|_2^2 + \gamma\|b_c\|_1 \tag{11}$$

where b_c stands for $(b)_{T^c}$. This is solving a sparse recovery problem on T^c, i.e. it is implicitly assuming that x_T is either equal to $\hat{\mu}_T$ or very close to it. Thus, this also works only when the signal value prior is very strong.

Both CS-residual and CS-mod-residual can be interpreted as extensions of BPDN, and [3, Theorem 8] can be used to bound their error. In either case, the bound will contain terms proportional to $\|(x_T - \hat{\mu}_T)\|_2$ and as a result, it will be large whenever the prior is not strong enough[1]. This is also seen from our simulation experiments shown in Fig. 2 where we provide comparisons for the case of good signal value prior (0.1% error in initial signal estimate) and bad signal value prior (10% error in initial signal estimate). We vary support errors from 5% to 20% misses, while keeping the extras fixed at 10%.

Reg-mod-BPDN can also be confused with *modified-CS-residual* which computes[29]

$$\hat{x} = \hat{\mu} + \hat{b}, \text{ where } \hat{b} \text{ solves}$$
$$\min_b \quad \frac{1}{2}\|y - A\hat{\mu} - Ab\|_2^2 + \gamma\|b_{T^c}\|_1 \tag{12}$$

This is indeed related to reg-mod-BPDN and in fact this inspired it. We studied this empirically in [29]. However, one cannot get good error bounds for it in any easy fashion. Notice that the minimization is over the entire vector b, while the ℓ_1 cost is only on b_{T^c}.

One may also consider solving the following variant of reg-mod-BPDN (we call this *reg-mod-BPDN-var*):

$$\min_b \quad \gamma\|b_{T^c}\|_1 + \frac{1}{2}\|y - Ab\|_2^2 + \frac{1}{2}\lambda\|b - \hat{\mu}\|_2^2 \tag{13}$$

Since $\hat{\mu}$ is supported on T, the regularization term can be rewritten as $\lambda\|b - \hat{\mu}\|_2^2 = \lambda\|b_T - \hat{\mu}_T\|_2^2 + \lambda\|b_{T^c}\|_2^2$. Thus, in addition to the ℓ_1 norm cost on b_{T^c} imposed by the first term, this last term is also imposing an ℓ_2 norm cost on it. If λ is large enough, the ℓ_2 norm cost will encourage the energy of the solution to be spread out on T^c, thus causing it to be less sparse. Since the true x is very sparse on T^c ($|\Delta|$ is small compared to the support size also), we will end up with a larger recovery error[2]. [see Fig. 2(a)]. However, if we compare the two approaches for compressible signal sequences, e.g. the larynx sequence, it is difficult to say which will be better [see Fig. 4].

Finally, one may solve the following (*we can call it reg-BPDN*)

$$\min_b \quad \gamma\|b\|_1 + \frac{1}{2}\|y - Ab\|_2^2 + \frac{1}{2}\lambda\|b - \hat{\mu}\|_2^2 \tag{14}$$

This has two limitations. (1) Like CS-residual, this also does not use the fact that when T is an accurate estimate of the true support, $(x)_{T^c}$ is much more sparse compared with the full $(x - \hat{\mu})$. (2) Its last term is the same as that of reg-mod-BPDN-var which also causes the same problem as above.

---

[1]In either case, one can assume that $(x - \hat{\mu})$ is supported on Δ and the "noise" is $w + A_T(x_T - \hat{\mu}_T)$. Thus, CS-residual error can be bounded by $C(A, \Delta)(\|w\|_2 + \|A_T(x_T - \hat{\mu}_T)\|_2)$ while CS-mod-residual error can be bounded by $\|x_T - \hat{\mu}_T\|_2 + C(A_{T^c}, \Delta)(\|w\|_2 + \|A_T(x_T - \hat{\mu}_T)\|_2)$.

[2]In the limit if $\sqrt{\lambda/2}$ is much larger than γ, we may get a completely non-sparse solution.

D. Paper Organization

We introduce reg-mod-BPDN in Sec. II. We obtain computable bounds on its reconstruction error in Sec. III. The simultaneous comparison of upper bounds of multiple approaches becomes difficult because their results hold under different sufficient conditions. In Sec. IV, we address this issue by showing how to obtain a tighter error bound that also holds without any sufficient conditions and is still computable. In both sections, the bounds for mod-BPDN and BPDN follow as direct corollaries. In Sec V, the above result is used for easy numerical comparisons between the upper bounds of various approaches – reg-mod-BPDN, mod-BPDN, BPDN and LS-CS and for numerically evaluating tightness of the bounds with both Gaussian measurements and partial Fourier measurements. We also provide reconstruction error comparisons with CS-residual, LS-CS, KF-CS, CS-mod-residual, mod-CS-residual and reg-mod-BPDN-var, as well as with weighted ℓ_1, mod-BPDN and BPDN for (a) static sparse recovery from random-Gaussian measurements; and for (b) recovering a larynx image sequence from simulated MRI measurements. Conclusions are given in Sec. VI.

II. REGULARIZED MODIFIED-BPDN (REG-MOD-BPDN)

Consider the sparse recovery problem when partial support knowledge is available. As explained earlier, one can use mod-BPDN given in (8). When the support estimate is accurate, i.e. $|\Delta|$ and $|\Delta_e|$ are small, mod-BPDN provides accurate recovery with fewer measurements than what BPDN needs. However, it puts no cost on b_T except the cost imposed by the data term. Thus, when very few measurements are available or when the noise is large, b_T can become larger than required (in order to reduce the data term). A similar, though lesser, bias will occur with weighted ℓ_1 also when $\gamma' < \gamma$. To address this, when reliable prior signal value knowledge is available, we can instead solve

$$\min_b \quad L(b) \triangleq \gamma\|b_{T^c}\|_1 + \frac{1}{2}\|y - Ab\|_2^2 + \frac{1}{2}\lambda\|b_T - \hat{\mu}_T\|_2^2 \tag{15}$$

which we call *reg-mod-BPDN*. Its solution, denoted by $\hat{x}$, serves as the reconstruction of the unknown signal, x. Notice that the first term helps to find the solution that is sparsest outside T, the second term imposes the data constraint while the third term imposes closeness to $\hat{\mu}$ along T.

Mod-BPDN is the special case of (15) when $\lambda = 0$. BPDN is also a special case with $\lambda = 0$ and $T = \emptyset$ (so that $\Delta = N$).

A. Limitations and Assumptions

A limitation of adding the regularizing term, $\lambda\|b_T - \hat{\mu}_T\|_2^2$ is as follows. It encourages the solution to be close to $(\hat{\mu})_{\Delta_e}$ which is not zero. As a result, $(\hat{x})_{\Delta_e}$ will also not be zero (except if λ is very small) even though $(x)_{\Delta_e} = 0$. Thus, even in the noise-free case, reg-mod-BPDN will not achieve exact reconstruction. In both noise-free and noisy cases, if $(\hat{\mu})_{\Delta_e}$ is large, $(\hat{x})_{\Delta_e}$ being close to $(\hat{\mu})_{\Delta_e}$ can result in large error. Thus, we need the assumption that $(\hat{\mu})_{\Delta_e}$ is small.

For the reason above, when we estimate the support of $\hat{x}$, we need to use a nonzero threshold, i.e. compute

$$\hat{N} = \{i : |\hat{x}_i| > \rho\} \tag{16}$$

with a $\rho > 0$. We note that thresholding as above is done *only* for support estimation and not for improving the actual reconstruction. Support estimation is required in dynamic reg-mod-BPDN (described below) where we use the support estimate from the previous time instant as the support knowledge, T, for the current time.

In summary, to get a small error reconstruction, reg-mod-BPDN requires the following (this can also be seen from the result of Theorem 1):

1) T is a good estimate of the true signal's support, N, i.e. $|\Delta|$ and $|\Delta_e|$ are small compared to $|N|$; and

2) $\hat{\mu}_T$ is a good estimate of x_T. For $i \in \Delta_e$, this implies that $|\hat{\mu}_i|$ is close to zero (since $x_i = 0$ for $i \in \Delta_e$).

3) For accurate support estimation, we also need that most nonzero elements of x are larger than $\max_{i \in \Delta_e} |\hat{\mu}_i|$ (for exact support estimation, we need this to hold for all nonzero elements of x).

The smallest nonzero elements of x are usually on the set Δ. In this case, the third assumption is equivalent to requiring that most elements of x_Δ are larger than $\max_{i \in \Delta_e} |\hat{\mu}_i|$.

B. Dynamic Reg-Mod-BPDN for Recursive Recovery

An important application of reg-mod-BPDN is for recursively reconstructing a time sequence of sparse signals from undersampled measurements, e.g. for dynamic MRI. To do this, at time t we solve (15) with $T = \hat{N}_{t-1}$, $(\hat{\mu})_T = (\hat{x}_{t-1})_T$ and $(\hat{\mu})_{T^c} = \mathbf{0}$. Here $\hat{N}_{t-1}$ is the support estimate of the previous reconstruction, $\hat{x}_{t-1}$. At the initial time, $t = 0$, we can either initialize with BPDN, or with mod-BPDN using T from prior knowledge, e.g. for wavelet sparse images, T could be the set of indices of the approximation coefficients. We summarize the stepwise dynamic reg-mod-BPDN approach in Algorithm 1. Notice that at $t = 0$, one may need more measurements since the prior knowledge of T may not be very accurate. Hence, we use $y_0 = A_0 x_0 + w_0$ where A_0 is an $n_0 \times m$ measurement matrix with $n_0 > n$.

In Algorithm 1, we should reiterate that for support estimation, we need to use a threshold $\rho > 0$. The threshold should be large enough so that most elements of $\Delta_{e,t} := T \setminus N_t = \hat{N}_{t-1} \setminus N_t$ do not get detected into the support.

We briefly discuss here the stability of dynamic reg-mod-BPDN (reconstruction error and support estimation errors bounded by a time-invariant and small value at all times). Using an approach similar to that of [30], it should be possible to show the following. If (i) ρ is large enough (so that $\hat{N}_t$ does not falsely detect any element that got removed from N_t); (ii) the newly added elements to the current support, N_t, either get added at a large enough value to get detected immediately, or within a finite delay their magnitude becomes large enough to get detected; and (iii) the matrix A satisfies certain conditions (for a given support size and support change size); reg-mod-BPDN will be stable.

Algorithm 1 Dynamic Reg-mod-BPDN

At $t = 0$, compute $\hat{x}_0$ as the solution of $\min_b \quad \gamma\|(b)_{T^c}\|_1 + \frac{1}{2}\|y_0 - Ab\|_2^2$, where T is either empty or is available from prior knowledge. Compute $\hat{N}_0 = \{i \in [1, ..., m] : |(\hat{x}_0)_i| > \rho\}$. Set $T \leftarrow \hat{N}_0$ and $(\hat{\mu})_T \leftarrow (\hat{x}_0)_T$ For $t > 0$, do

1) *Reg-Mod-BPDN.* Let $T = \hat{N}_{t-1}$ and let $\hat{\mu}_T = (\hat{x}_{t-1})_T$. Compute $\hat{x}_t$ as the solution of (15).
2) *Estimate Support.* $\hat{N}_t = \{i \in [1, ..., m] : |(\hat{x}_t)|_i > \rho\}$.
3) Output the reconstruction $\hat{x}_t$.

Feedback $\hat{N}_t$ and $\hat{x}_t$; increment t, and go to step 1.

III. BOUNDING THE RECONSTRUCTION ERROR

In this section, we bound the reconstruction error of reg-mod-BPDN. Since mod-BPDN and BPDN are special cases, their results follow as direct corollaries. The result for BPDN is the same as [3, Theorem 8]. In Sec. III-A, we define the terms needed to state our result. In III-B we state our result and discuss its implications. In III-C, we give the proof outline.

A. Definitions

We begin by defining the function that we want to minimize as

$$L(b) \triangleq L_1(b) + \gamma\|b_{T^c}\|_1 \tag{17}$$

where

$$L_1(b) \triangleq \frac{1}{2}\|y - Ab\|_2^2 + \frac{1}{2}\lambda\|b_T - \hat{\mu}_T\|_2^2 \tag{18}$$

contains the two ℓ_2 norm terms (data fidelity term and the regularization term). If we constrain b to be supported on $T \cup S$ for some $S \subset T^c$, then the minimizer of $L_1(b)$ will be the regularized least squares (LS) estimator obtained when we put a weight λ on $\|b_T - \hat{\mu}_T\|_2^2$ and a weight zero on $\|b_S - \hat{\mu}_S\|_2^2$.

Let S be a given subset of Δ. Next, we define three matrices which will be frequently used in our results. Let

$$Q_{T,\lambda}(S) \triangleq A_{T \cup S}{}' A_{T \cup S} + \lambda \begin{bmatrix} I_T & \mathbf{0}_{T,S} \\ \mathbf{0}_{S,T} & \mathbf{0}_{S,S} \end{bmatrix} \tag{19}$$

$$M_{T,\lambda} \triangleq I - A_T(A_T{}'A_T + \lambda I_T)^{-1}A_T{}' \tag{20}$$

$$P_{T,\lambda}(S) \triangleq (A_S{}'M_{T,\lambda}A_S)^{-1} \tag{21}$$

where I_T is a $|T| \times |T|$ identity matrix and $\mathbf{0}_{T,S}$, $\mathbf{0}_{S,T}$, $\mathbf{0}_{S,S}$ are all zeros matrices with sizes $|T| \times |S|$, $|S| \times |T|$ and $|S| \times |S|$.

Assumption 1: We assume that $Q_{T,\lambda}(\Delta)$ is invertible. This implies that, for any $S \subseteq \Delta$, the functions $L(b)$ and $L_1(b)$ are strictly convex over the set of all vectors supported on $T \cup S$.

Proposition 1: When $\lambda > 0$, $Q_{T,\lambda}(S)$ is invertible if A_S has full rank. When $\lambda = 0$ (mod-BPDN), this will hold if $A_{T \cup S}$ has full rank.

The proof is easy and is given in Appendix A.

Let $S \subseteq \Delta$. Consider minimizing $L(b)$ over b supported on $T \cup S$. When $b_{(T \cup S)^c} = 0$ and Assumption 1 holds, $L(b_{T \cup S})$ is strictly convex and thus has a unique minimizer. The same holds for $L_1(b_{T \cup S})$. Define their respective unique minimizers as

$$d_{T,\lambda}(S) \triangleq \arg\min_b \ L(b) \quad \text{subject to} \quad b_{(T \cup S)^c} = \mathbf{0} \tag{22}$$

$$c_{T,\lambda}(S) \triangleq \arg\min_b \ L_1(b) \quad \text{subject to} \quad b_{(T \cup S)^c} = \mathbf{0} \tag{23}$$

As explained earlier, $c_{T,\lambda}(S)$ is the regularized LS estimate of x when assuming that x is supported on $T \cup S$ and with the weights mentioned earlier. It is easy to see that

$$[c_{T,\lambda}(S)]_{T \cup S} = Q_{T,\lambda}(S)^{-1} \left(A_{T \cup S}'y + \begin{bmatrix} \lambda\hat{\mu}_T \\ \mathbf{0}_S \end{bmatrix} \right)$$

$$[c_{T,\lambda}(S)]_{(T \cup S)^c} = \mathbf{0} \tag{24}$$

In a fashion similar to [3], define

$$ERC_{T,\lambda}(S) \triangleq 1 - \max_{\omega \notin T \cup S} \|P_{T,\lambda}(S)A_S'M_{T,\lambda}A_\omega\|_1 \tag{25}$$

This is different from the ERC of [3] but simplifies to it when $T = \emptyset$, $S = N$ and $\lambda = 0$. In [3], the ERC, which in our notation is $ERC_{\emptyset,0}(N)$, being strictly positive, along with γ approaching zero, ensured exact recovery of BPDN in the noise-free case. Hence, in [3], ERC was an acronym for *Exact Recovery Coefficient*. In this work, the same holds for mod-BPDN. If $ERC_{T,0}(\Delta) > 0$, the solution of mod-BPDN approaches the true x as γ approaches zero. We explain this further in Remark 2 below. However, no similar claim can be made for reg-mod-BPDN. On the other hand, for the reconstruction error bounds, ERC serves the exact same purpose for reg-mod-BPDN as it does for BPDN in [3]: $ERC_{T,\lambda}(\Delta) > 0$ and γ greater than a certain lower bound ensures that the reg-mod-BPDN (or mod-BPDN) error can be bounded by modifying the approach of [3].

B. Reconstruction error bound

The reconstruction error can be bounded as follows.

Theorem 1: If $Q_{T,\lambda}(\Delta)$ is invertible, $ERC_{T,\lambda}(\Delta) > 0$ and

$$\gamma \geq \gamma_{T,\lambda}^*(\Delta) \triangleq \frac{\|A_{(T \cup \Delta)^c}'(y - Ac_{T,\lambda}(\Delta))\|_\infty}{ERC_{T,\lambda}(\Delta)} \tag{26}$$

then,

1) $L(b)$ has a unique minimizer, $\hat{x}$.
2) The minimizer, $\hat{x}$, is equal to $d_{T,\lambda}(\Delta)$, and thus is supported on $T \cup \Delta$.

3) Its error can be bounded as

$$\|x - \hat{x}\|_2 \leq \gamma\sqrt{|\Delta|}f_1(\Delta) + \lambda f_2(\Delta)\|x_T - \hat{\mu}_T\|_2$$
$$+ f_3(\Delta)\|w\|_2$$

where

$$f_1(\Delta) \triangleq$$
$$\sqrt{\|(A_T{}'A_T + \lambda I_T)^{-1}A_T{}'A_\Delta P_{T,\lambda}(\Delta)\|_2^2 + \|P_{T,\lambda}(\Delta)\|_2^2},$$
$$f_2(\Delta) \triangleq \|Q_{T,\lambda}(\Delta)^{-1}\|_2,$$
$$f_3(\Delta) \triangleq \|Q_{T,\lambda}(\Delta)^{-1}A_{T\cup\Delta}{}'\|_2, \tag{27}$$

$P_{T,\lambda}(\Delta)$ is defined in (21) and $Q_{T,\lambda}(\Delta)$ in (19).

Corollary 1 (corollaries for mod-BPDN and BPDN): The result for mod-BPDN follows by setting $\lambda = 0$ in Theorem 1. The result for BPDN follows by setting $\lambda = 0$, $T = \emptyset$ (and so $\Delta = N$). This result is the same as [3, Theorem 8].

Remark 1 (smallest γ): Notice that the error bound above is an increasing function of γ. Thus $\gamma = \gamma_{T,\lambda}^*(\Delta)$ gives the smallest bound.

In words, Theorem 1 says that, if $Q_{T,\lambda}(\Delta)$ is invertible, $ERC_{T,\lambda}(\Delta)$ is positive, and γ is large enough (larger than γ^*), then $L(b)$ has a unique minimizer, $\hat{x}$, and $\hat{x}$ is supported on $T \cup \Delta = N \cup \Delta_e$. This means that the only wrong elements that can possibly be part of the support of $\hat{x}$ are elements of Δ_e. Moreover, the error between $\hat{x}$ and the true x is bounded by a value that is small as long as the noise, $\|w\|_2$, is small, the prior term, $\|x_T - \hat{\mu}_T\|_2$, is small and $\gamma_{T,\lambda}^*(\Delta)$ is small. By rewriting $y - Ac_{T,\lambda}(\Delta) = A(x - c_{T,\lambda}(\Delta)) + w$ and using Lemma 2 (given in the Appendix) one can upper bound γ^* by terms that are increasing functions of $\|w\|_2$ and $\|x_T - \hat{\mu}_T\|_2$. Thus, as long as these are small, the bound is small.

As shown in Proposition 1, $Q_{T,\lambda}(\Delta)$ is invertible if $\lambda > 0$ and A_Δ is full rank or if $A_{T\cup\Delta}$ is full rank.

Next, we use the idea of [3, Corollary 10] to show that $ERC_{T,0}(\Delta)$ is an *Exact Recovery Coefficient* for mod-BPDN.

Remark 2 (ERC and exact recovery of mod-BPDN): For mod-BPDN, $c_{T,0}(\Delta)$ is the LS estimate when x is supported on $T \cup \Delta$. Using (24), (1), and the fact that x is supported on $N \subseteq T \cup \Delta$, it is easy to see that in the noise-free ($w = 0$) case, $c_{T,0}(\Delta) = x_{T\cup\Delta}$. Hence the numerator of $\gamma_{T,0}^*(\Delta)$ will be zero. Thus, using Theorem 1, if $ERC_{T,0}(\Delta) > 0$, the mod-BPDN error satisfies $\|x - \hat{x}\|_2 \leq \gamma\sqrt{|\Delta|}f_1(\Delta)$. Thus the mod-BPDN solution, $\hat{x}$, will approach the true x as γ approaches zero. Moreover, as long as $\gamma < \frac{\min_{i\in N}|x_i|}{\sqrt{|\Delta|}f_1(\Delta)}$, at least the support of $\hat{x}$ will equal the true support, N [3].

We show a numerical comparison of the results of reg-mod-BPDN, mod-BPDN and BPDN in Table I (simulation details given in Sec. V). Notice that BPDN needs 90% of the measurements for its sufficient

---

[3]If we bounded the ℓ_∞ norm of the error as done in [3] we would get a looser upper bound on the allowed γ's for this.

conditions to start holding (ERC to become positive) whereas mod-BPDN only needs 19%. Moreover, even with 90% of the measurements, the ERC of BPDN is just positive and very small. As a result, its error bound is large (27% normalized mean squared error (NMSE)). Similarly, notice that mod-BPDN needs $n \geq 19\%m$ for its sufficient conditions to start holding ($A_{T \cup \Delta}$ to become full rank which is needed for $Q_{T,0}(\Delta)$ to be invertible). For reg-mod-BPDN which only needs A_Δ to be full rank, $n = 13\%m$ suffices.

Remark 3: A sufficient conditions comparison only provides a comparison of when a given result can be applied to provide a bound on the reconstruction error. In other words, it tells us under what conditions we can guarantee that the reconstruction error of a given approach will be small (below a bound). Of course this does not mean that we cannot get small error even when the sufficient condition does not hold, e.g., in simulations, BPDN provides a good reconstruction using much less than 90% of the measurements. However, when $n < 90\%m$ we cannot bound its reconstruction error using Theorem 1 above.

C. Proof Outline

To prove Theorem 1, we use the following approach motivated by that of [3].

1) We first bound $\|d_{T,\lambda}(\Delta) - c_{T,\lambda}(\Delta)\|_2$ by simplifying the necessary and sufficient condition for it to be the minimizer of $L(b)$ when b is supported on $T \cup \Delta$. This is done in Lemma 1 in Appendix B.

2) We bound $\|c_{T,\lambda}(\Delta) - x\|_2$ using the expression for $c_{T,\lambda}(\Delta)$ in (24) and substituting $y = A_{T \cup \Delta} x_{T \cup \Delta} + w$ in it (recall that x is zero outside $T \cup \Delta$). This is done in Lemma 2 in Appendix B.

3) We can bound $\|d_{T,\lambda}(\Delta) - x\|_2$ using the above two bounds and the triangle inequality.

4) We use an approach similar to [3, Lemma 6] to find the sufficient conditions under which $d_{T,\lambda}(\Delta)$ is also the unconstrained unique minimizer of $L(b)$, i.e. $\hat{x} = d_{T,\lambda}(\Delta)$. This is done in Lemma 3 in Appendix B.

The last step (Lemma 3) helps prove the first two parts of Theorem 1. Combining the above four steps, we get the third part (error bound). We give the lemmas in Appendix B. They are proved in Appendix D1, D2 and D3.

Two key differences in the above approach with respect to the result of [3] are

- $c_{T,\lambda}(\Delta)$ is the regularized LS estimate instead of the LS estimate in [3]. This helps obtain a better and simpler error bound of reg-mod-BPDN than when using the LS estimate. Of course, when $\lambda = 0$ (mod-BPDN or BPDN), $c_{T,0}(\Delta)$ is just the LS estimate again.

- For reg-mod-BPDN (and also for mod-BPDN), the subgradient set of the ℓ_1 term is $\partial \|b_{T^c}\|_1|_{b=d_{T,\lambda}(\Delta)}$ and so any ϕ in this set is zero on T, and only has $\|\phi_\Delta\|_\infty \leq 1$. Since $|\Delta| \ll |N|$, this helps to get a tighter bound on $\|c_{T,\lambda}(\Delta) - d_{T,\lambda}(\Delta)\|_2$ in step 1 above as compared to that for BPDN [3] (see proof of Lemma 1 for details).

IV. Tighter Bounds without Sufficient Conditions

The problem with the error bounds for reg-mod-BPDN, mod-BPDN, BPDN or LS-CS [31] is that they all hold under different sufficient conditions. This makes it difficult to compare them. Moreover, the bound is particularly loose when n is such that the sufficient conditions just get satisfied. This is because the ERC is just positive but very small (resulting in a very large γ^* and hence a very large bound). To address this issue, in this section, we obtain a bound that holds without any sufficient conditions and that is also tighter, while still being computable. The key idea that we use is as follows:

- we modify Theorem 1 to hold for "sparse-compressible" signals [31], i.e. for sparse signals, x, in which some nonzero coefficients out of the set Δ are small ("compressible") compared to the rest; and then
- we minimize the resulting bound over all allowed split-ups of x into non-compressible and compressible parts.

Let $\tilde{\Delta} \subseteq \Delta$ be such that the conditions of Theorem 1 hold for it. Then the first step involves modifying Theorem 1 to bound the error for reconstructing x when we treat $x_{\Delta \setminus \tilde{\Delta}}$ as the "compressible" part. The main difference here is in bounding $\|c_{T,\lambda}(\tilde{\Delta}) - x\|_2$ which now has a larger bound because of $x_{\Delta \setminus \tilde{\Delta}}$. We do this in Lemma 4 in the Appendix C. Notice from the proofs of Lemma 1 and Lemma 3 in Appendix D1 and D3 that nothing in their result changes if we replace Δ by a $\tilde{\Delta} \subseteq \Delta$. Combining Lemma 4 with Lemmas 1 and 3 applied for $\tilde{\Delta}$ instead of Δ leads to the following corollary.

Corollary 2: Consider a $\tilde{\Delta} \subseteq \Delta$. If $Q_{T,\lambda}(\tilde{\Delta})$ is invertible, $ERC_{T,\lambda}(\tilde{\Delta}) > 0$, and $\gamma = \gamma^*_{T,\lambda}(\tilde{\Delta})$, then

$$\|x - \hat{x}\|_2 \leq f(T, \lambda, \Delta, \tilde{\Delta}, \gamma^*_{T,\lambda}(\tilde{\Delta})) \tag{28}$$

where

$$f(T, \lambda, \Delta, \tilde{\Delta}, \gamma) \triangleq \gamma \sqrt{|\tilde{\Delta}|} f_1(\tilde{\Delta}) + \lambda f_2(\tilde{\Delta}) \|x_T - \hat{\mu}_T\|_2$$
$$+ f_3(\tilde{\Delta}) \|w\|_2 + f_4(\tilde{\Delta}) \|x_{\Delta \setminus \tilde{\Delta}}\|_2, \tag{29}$$

$$f_4(\tilde{\Delta}) \triangleq \sqrt{\|Q_{T,\lambda}(\tilde{\Delta})^{-1} A_{T \cup \tilde{\Delta}}{}' A_{\Delta \setminus \tilde{\Delta}}\|_2^2 + 1}, \tag{30}$$

$f_1(\cdot), f_2(\cdot), f_3(\cdot)$ are defined in (27) and $\gamma^*_{T,\lambda}(\tilde{\Delta})$ in (26).

Proof: The proof is given in Appendix C1.

In order to get a bound that depends only on $\|x_T - \hat{\mu}_T\|_2$, $\|x_{\Delta \setminus \tilde{\Delta}}\|_2$, the noise, w, and the sets T, Δ, Δ_e, we can further bound $\gamma^*_{T,\lambda}(\tilde{\Delta})$ by rewriting $y - Ac_{T,\lambda}(\tilde{\Delta}) = A(x - c_{T,\lambda}(\tilde{\Delta})) + w$ and then bounding $\|x - (c_{T,\lambda}(\tilde{\Delta}))\|_2$ using Lemma 4. Doing this gives the following corollary.

Corollary 3: If $Q_{T,\lambda}(\tilde{\Delta})$ is invertible, $ERC_{T,\lambda}(\tilde{\Delta}) > 0$, and $\gamma = \gamma^*_{T,\lambda}(\tilde{\Delta})$, then

$$\|x - \hat{x}\|_2 \leq g(\tilde{\Delta}) \tag{31}$$

where

$$g(\tilde{\Delta}) \triangleq g_1 \|x_T - \hat{\mu}_T\|_2 + g_2 \|w\|_2 + g_3 \|x_{\Delta \setminus \tilde{\Delta}}\|_2 + g_4 \tag{32}$$

$$g_1 \triangleq \lambda f_2(\tilde{\Delta}) \left(\frac{\sqrt{|\tilde{\Delta}|} f_1(\tilde{\Delta}) \mathrm{maxcor}(\tilde{\Delta})}{ERC_{T,\lambda}(\tilde{\Delta})} + 1 \right),$$

$$g_2 \triangleq \frac{\sqrt{|\tilde{\Delta}|} f_1(\tilde{\Delta}) f_3(\tilde{\Delta}) \mathrm{maxcor}(\tilde{\Delta})}{ERC_{T,\lambda}(\tilde{\Delta})} + f_3(\tilde{\Delta}),$$

$$g_3 \triangleq \frac{\sqrt{|\tilde{\Delta}|} f_1(\tilde{\Delta}) f_4(\tilde{\Delta}) \mathrm{maxcor}(\tilde{\Delta})}{ERC_{T,\lambda}(\tilde{\Delta})} + f_4(\tilde{\Delta}),$$

$$g_4 \triangleq \frac{\sqrt{|\tilde{\Delta}|} \|A_{(T \cup \tilde{\Delta})^c}{}' w\|_\infty f_1(\tilde{\Delta})}{ERC_{T,\lambda}(\tilde{\Delta})},$$

$$\mathrm{maxcor}(\tilde{\Delta}) \triangleq \max_{i \notin (T \cup \tilde{\Delta})^c} \|A_i{}' A_{T \cup \Delta}\|_2,$$

$f_1(\cdot), f_2(\cdot), f_3(\cdot)$ and $f_4(\cdot)$ are defined in (27) and (30), and $\gamma_{T,\lambda}^*(\tilde{\Delta})$ in (26).

Proof: The proof is given in Appendix C2.

Using the above corollary and minimizing over all allowed $\tilde{\Delta}$'s, we get the following result.

Theorem 2: Let

$$\tilde{\Delta}^* \triangleq \operatorname*{arg\,min}_{\tilde{\Delta} \in \mathcal{G}} g(\tilde{\Delta}) \tag{33}$$

where

$$\mathcal{G} \triangleq \{\tilde{\Delta} : \tilde{\Delta} \subseteq \Delta, ERC_{T,\lambda}(\tilde{\Delta}) > 0, Q_{T,\lambda}(\tilde{\Delta}) \text{ is invertible}\} \tag{34}$$

If $\gamma = \gamma_{T,\lambda}^*(\tilde{\Delta}^*)$, then

1) $L(b)$ has a unique minimizer, $\hat{x}$, supported on $T \cup \tilde{\Delta}^*$.

2) The error bound is

$$\|x - \hat{x}\|_2 \leq g(\tilde{\Delta}^*) \tag{35}$$

($\gamma_{T,\lambda}^*(\tilde{\Delta})$ is defined in (26)).

Proof: This result follows by minimizing over all allowed $\tilde{\Delta}$'s from Corollary 3.

Compare Theorem 2 with Theorem 1. Theorem 1 holds only when the complete set Δ belongs to $\mathcal{G}$, whereas Theorem 2 holds always (we only need to set γ appropriately). Moreover, even when Δ does belong to $\mathcal{G}$, Theorem 1 gives the error bound by choosing $\tilde{\Delta}^* = \Delta$. However, Theorem 2 minimizes over all allowed $\tilde{\Delta}$'s, thus giving a tighter bound, especially for the case when the sufficient conditions of Theorem 1 just get satisfied and $ERC_{T,\lambda}(\Delta)$ is positive but very small. A similar comparison also holds for the mod-BPDN and BPDN results.

The problem with Theorem 2 is that its bound is not computable (the computational cost is exponential in $|\Delta|$). Notice that $g(\tilde{\Delta}^*) := \min_{\tilde{\Delta} \in \mathcal{G}} g(\tilde{\Delta})$ can be rewritten as

$$g(\tilde{\Delta}^*) \triangleq \min_{\tilde{\Delta} \in \mathcal{G}} g(\tilde{\Delta}) = \min_{0 \le k \le |\Delta|} \min_{\mathcal{G}_k} g(\tilde{\Delta}) \quad \text{where}$$

$$\mathcal{G}_k \triangleq \mathcal{G} \cap \{\tilde{\Delta} \subseteq \Delta : |\tilde{\Delta}| = k\} \tag{36}$$

Let $d := |\Delta|$. The minimization over $\mathcal{G}_k$ is expensive since it requires searching over all $\binom{d}{k}$ size k subsets of Δ to first find which ones belong to $\mathcal{G}_k$ and then find the minimum over all $\tilde{\Delta} \subseteq \mathcal{G}_k$. The total computation cost to do the former for all sets $\mathcal{G}_0, \mathcal{G}_1, \ldots \mathcal{G}_d$ is $O(\sum_{k=0}^{d} \binom{d}{k}) = O(2^d)$, i.e. it is *exponential in* d. This makes the bound computation intractable for large problems.

A. Obtaining a Computable Bound

In most cases of practical interest, the term that has the maximum variability over different sets in $\mathcal{G}_k$ is $\|x_{\Delta \setminus \tilde{\Delta}}\|_2$. The multipliers g_1, g_2, g_3 and g_4 vary very slightly for different sets in a given $\mathcal{G}_k$. Using this fact, we can obtain the following upper bound on $\min_{\mathcal{G}_k} g(\tilde{\Delta})$ which is only slightly looser and also holds without sufficient conditions, but is computable in polynomial time.

Define $\tilde{\Delta}^{**}(k)$ and B_k as follows

$$\tilde{\Delta}^{**}(k) \triangleq \arg \min_{\{\tilde{\Delta} \subseteq \Delta, |\tilde{\Delta}| = k\}} \|x_{\Delta \setminus \tilde{\Delta}}\|_2$$

$$B_k \triangleq \begin{cases} g(\tilde{\Delta}^{**}(k)) & \text{if } \tilde{\Delta}^{**}(k) \in \mathcal{G}_k \\ \infty & \text{otherwise} \end{cases} \tag{37}$$

Then, clearly

$$\min_{\mathcal{G}_k} g(\tilde{\Delta}) \le B_k \tag{38}$$

since $\min_{\mathcal{G}_k} g(\tilde{\Delta}) \le g(\tilde{\Delta})$ for any $\tilde{\Delta} \in \mathcal{G}_k$ and it is also less than infinity. For any k, the set $\tilde{\Delta}^{**}(k)$ can be obtained by sorting the elements of x_Δ in decreasing order of magnitude and letting $\tilde{\Delta}^{**}(k)$ contain the indices of the k largest elements. Doing this takes $O(d \log d)$ time since sorting takes $O(d \log d)$ time. Computation of B_k requires matrix multiplications and inversions which are $O(k^3)$. Thus, the total cost of doing this is at most $O(d^4)$ which is still polynomial in d.

Therefore, we get the following bound that is *computable in polynomial time and that still holds without sufficient conditions and is much tighter than Theorem 1.*

Theorem 3: Let

$$k_{\min} \triangleq \arg \min_{0 \le k \le |\Delta|} B_k \quad \text{and}$$

$$\tilde{\Delta}^{**} \triangleq \tilde{\Delta}^{**}(k_{\min}) \tag{39}$$

where B_k and $\tilde{\Delta}^{**}(k)$ are defined in (37). If $\gamma = \gamma^*_{T,\lambda}(\tilde{\Delta}^{**})$,

1) $L(b)$ has a unique minimizer, $\hat{x}$, supported on $T \cup \tilde{\Delta}^{**}$.

2) The error bound is

$$\|x - \hat{x}\|_2 \leq g(\tilde{\Delta}^{**}) \tag{40}$$

($\gamma^*_{T,\lambda}(\tilde{\Delta})$ is defined in (26)).

Corollary 4 (corollaries for mod-BPDN and BPDN): The result for mod-BPDN follows by setting $\lambda = 0$ in Theorem 3. The result for BPDN follows by setting $\lambda = 0$, $T = \emptyset$ (and so $\Delta = N$) in Theorem 3.

When n and $s \triangleq |N|$ are large enough, the above bound is either only slightly larger, or often actually equal, to that of Theorem 2 (e.g. in Fig. 5(a), $m = 256$, $n = 0.13m = 33$, $s = 0.1m = 26$). The reason for the equality is that the minimizing value of k is the one that is small enough to ensure that g_1, g_2, g_3, g_4 are small. When k is small, g_1, g_2, g_3, g_4, ERC and $Q(\tilde{\Delta})$ have very similar values for all sets $\tilde{\Delta}$ of the same size k. In (32), the only term with significant variability for different sets $\tilde{\Delta}$ of the same size k is $\|x_{\Delta \setminus \tilde{\Delta}}\|_2$. Thus, (a) $\arg\min_{\mathcal{G}_k} g(\tilde{\Delta}) = \arg\min_{\mathcal{G}_k} \|x_{\Delta \setminus \tilde{\Delta}}\|_2$ and (b) $\mathcal{G}_k$ is equal to $\{\tilde{\Delta} \subseteq \Delta, |\tilde{\Delta}| = k\}$. Thus, (38) holds with equality and so the bounds from Theorems 3 and 2 are equal. As n and $s \triangleq |N|$ approach infinity, *it is possible to use a law of large numbers (LLN) argument to prove that both bounds will be equal with high probability (w.h.p.).* The key idea will be the same as above: show that as n, s go to infinity, w.h.p., g_1, g_2, g_3, g_4, Q and ERC are equal for all sets $\tilde{\Delta}$ of any given size k. We will develop this result in future work.

V. NUMERICAL EXPERIMENTS

In this section, we show both upper bound comparisons and actual reconstruction error comparisons. The upper bound comparison only tells us that the performance guarantees of reg-mod-BPDN are better than those for the other methods. To actually demonstrate that reg-mod-BPDN outperforms the others, we need to compare the actual reconstruction errors. This section is organized as follows. After giving the simulation model in Sec V-A, we show the reconstruction error comparisons for recovering simulated sparse signals from random Gaussian measurements in Sec V-B. In Sec V-C, we show comparisons for recursive dynamic MRI reconstruction of a larynx image sequence. In this comparison, we also show the usefulness of the Theorem 3 in helping us select a good value of γ. In the last three subsections, we show numerical comparisons of the results of the various theorems. The upper bound comparisons of Theorem 3 and the comparison of the corresponding reconstruction errors suggests that the bounds for reg-mod-BPDN and BPDN are tight under the scenarios evaluated. Hence, they can be used as a proxy to decide which algorithm to use when. We show this for both random Gaussian and partial Fourier measurements.

A. Simulation Model

The notation $z = \pm a$ means that we generate each element of the vector z independently and each is either $+a$ or $-a$ with probability 1/2. The notation $\nu \sim \mathcal{N}(0, \Sigma)$ means that ν is generated from a

Gaussian distribution with mean 0 and covariance matrix Σ. We use $\lfloor a \rfloor$ to denote the largest integer less than or equal to a. Independent and identically distributed is abbreviated as iid. Also, N-RMSE refers to the normalized root mean squared error.

We use the recursive reconstruction application [20], [14] to motivate the simulation model. In this case, assuming that slow support and slow signal value change hold [see Fig. 1], we can use the reconstructed value of the signal at the previous time as $\hat{\mu}$ and its support as T. To simulate the effect of slow signal value change, we let $x_N = \mu_N + \nu$ where ν is a small iid Gaussian deviation and we let $\hat{\mu}_{T\cap N} = \mu_{T\cap N}$ (and so $x_{T\cap N} = \hat{\mu}_{T\cap N} + \nu_{T\cap N}$).

The extras set, $\Delta_e = T \setminus N$, contains elements that got removed from the support at the current time or at a few previous times (but so far did not get removed from the support estimate). In most practical applications, only small valued elements at the previous time get removed from the support and hence the magnitude of $\hat{\mu}$ on Δ_e will be small. We use β_s to denote this small magnitude, i.e. we simulate $(\hat{\mu})_{\Delta_e} = \pm\beta_s$.

The misses' set at time t, Δ, definitely includes the elements that just got added to the support at t or the ones that previously got added but did not get detected into the support estimate so far. The new elements typically get added at a small value and their value slowly increases to a large one. Thus, elements in Δ will either have small magnitude (corresponding to the current newly added ones), or will have larger magnitude but still smaller than that of elements already in $N \cap T$. To simulate this, we do the following. (a) We simulate the elements on $N \cap T$ to have large magnitude, β_l, i.e. we let $(\mu)_{N\cap T} = \pm\beta_l$. (b) We split the set Δ into two disjoint parts, Δ_1 and $\Delta_2 = \Delta \setminus \Delta_1$. The set Δ_1 contains the small (e.g. newly added) elements, i.e. $(\mu)_{\Delta_1} = \pm\beta_s$. The set Δ_2 contains the larger elements, though still with magnitudes smaller than those in $N \cap T$, i.e. $(\mu)_{\Delta_2} = \pm\beta_m$, where $\beta_l \geq \beta_m \geq \beta_s$.

In summary, we use the following simulation model.

$$(x)_N = (\mu)_N + \nu, \quad \nu \sim \mathcal{N}(0, \sigma_p^2 I)$$

$$(x)_{N^c} = 0 \tag{41}$$

$$\text{where} \quad (\mu)_{N\cap T} = \pm\beta_l$$

$$(\mu)_{\Delta_1} = \pm\beta_s, \quad (\mu)_{\Delta_2} = \pm\beta_m$$

$$(\mu)_{N^c} = 0 \tag{42}$$

and

$$(\hat{\mu})_{T\cap N} = (\mu)_{T\cap N} = \pm\beta_l$$

$$(\hat{\mu})_{\Delta_e} = \pm\beta_s$$

$$(\hat{\mu})_{T^c} = 0 \tag{43}$$

We generate the support of x, N, of size $|N|$, uniformly at random from $[1, ..., m]$. We generate Δ with size $|\Delta|$ and Δ_e with size $|\Delta_e|$ uniformly at random from N and from N^c respectively. The set
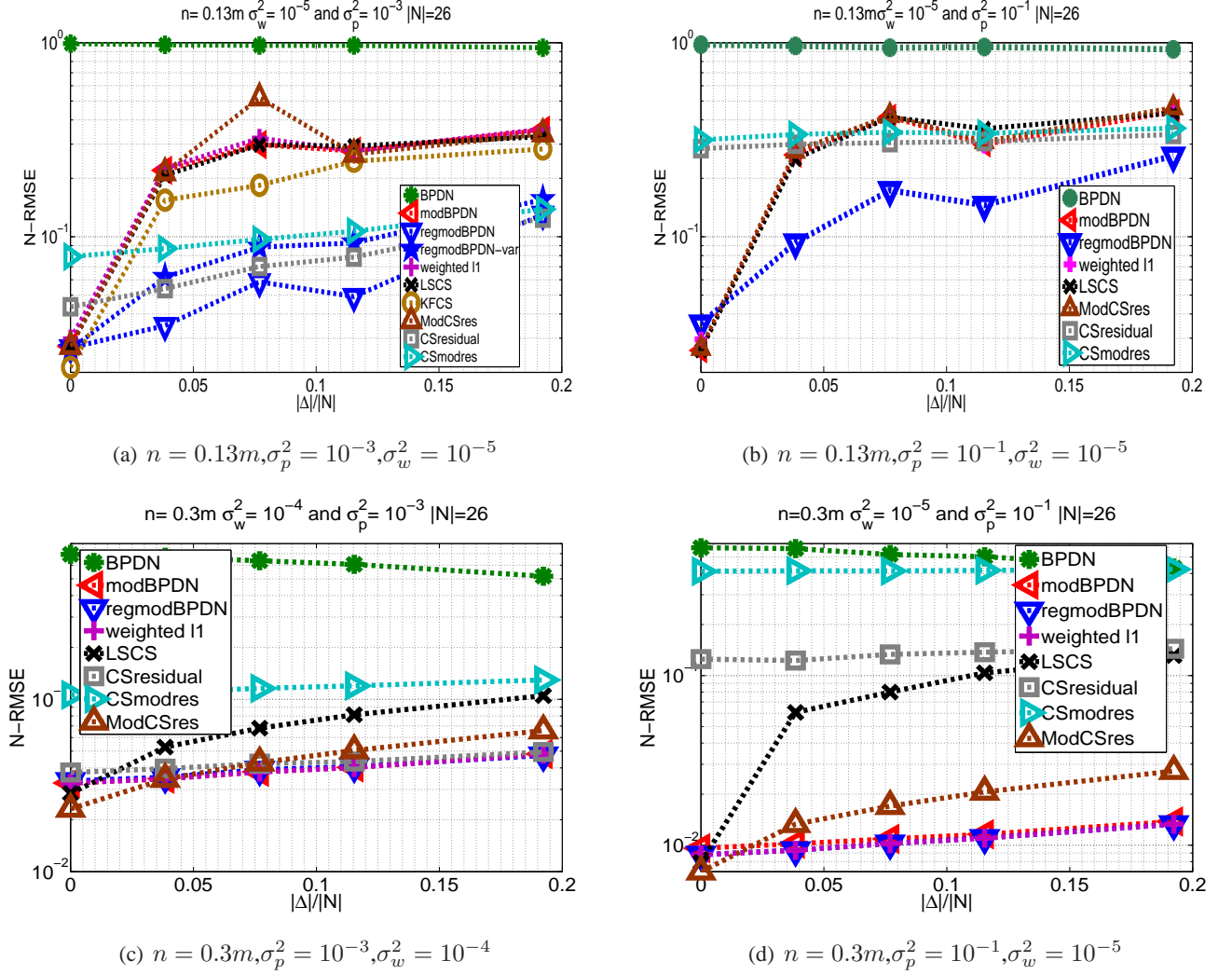
Fig. 2. The N-RMSE for reg-mod-BPDN, mod-BPDN, BPDN, LS-CS, KF-CS, weighted ℓ_1, CS-residual, CS-mod-residual and modified-CS-residual are plotted. For $n = 0.13m$, reg-mod-BPDN has smaller errors than those of mod-BPDN and the gap is larger when the signal estimate is good. For $n = 0.3m$, the errors of reg-mod-BPDN, mod-BPDN and weighted ℓ_1 are close and all small.

Δ_1 of size $|\Delta_1| = \lfloor |\Delta|/2 \rfloor$ is generated uniformly at random from Δ. The set $\Delta_2 = \Delta \setminus \Delta_1$. We let $T = N \cup \Delta_e \setminus \Delta$. We generate μ and then x using (42) and (41). We generate $\hat{\mu}$ using (43).

In some simulations, we simulated the more difficult case where $\beta_m = \beta_s$. In this case, all elements on Δ were identically generated and hence we did not need Δ_1.

B. Reconstruction Error Comparisons

In Fig. 2, we compare the Monte Carlo average of the reconstruction error of reg-mod-BPDN with that of mod-BPDN, BPDN, weighted ℓ_1 [15] given in (9), CS-residual given in (10), CS-mod-residual
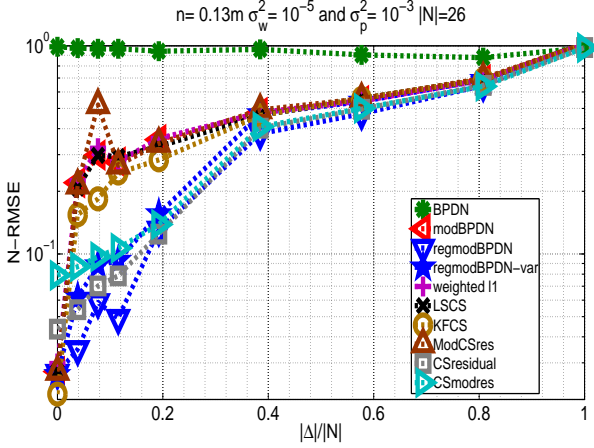
Fig. 3. Plot of Fig 2(a) extended all the way to $|\Delta|/|N| = 1$ (which is the same as $\Delta = N$). Notice that if $|\Delta_e| = 0$, then the point $|\Delta|/|N| = 1$ of reg-mod-BPDN (or of mod-BPDN) is the same as BPDN. But in our plot, $|\Delta_e| = 3$ and hence the two points are different, even though the errors are quite similar.

given in (11) and modified-CS-residual[29] given in (12). Simulation was done according to the model specified above. We used random Gaussian measurements in this simulation, i.e. we generated A as an $n \times m$ matrix with iid zero mean Gaussian entries and normalized each column to unit ℓ_2 norm.

We experimented with two choices of n, $n = 0.13m$ (where reg-mod-BPDN outperforms mod-BPDN) and $n = 0.3m$ (where both are similar) and two values of σ_p^2, $\sigma_p^2 = 0.001$ (good prior) and $\sigma_p^2 = 0.1$ (bad prior). For the cases of Fig 2(a) ($n = 0.13m$, $\sigma_p^2 = 0.001$) and Fig 2(b) ($n = 0.13m$, $\sigma_p^2 = 0.1$), we used signal length $m = 256$, support size $|N| = 0.1m = 26$ and support extras size, $|\Delta_e| = 0.1|N| = 3$. The misses' size, $|\Delta|$, was varied between 0 and $0.2|N|$ (these numbers were motivated by the medical imaging application, we used larger numbers than what are shown in Fig. 1). We used $\beta_l = 1$, $\beta_m = 0.4$ and $\beta_s = 0.2$. The noise variance was $\sigma_w^2 = 10^{-5}$. For the last two figures, Fig 2(c) ($n = 0.3m$, $\sigma_p^2 = 0.001$) and Fig 2(d) ($n = 0.3m$, $\sigma_p^2 = 0.1$), for which n was larger, we used $\beta_m = \beta_s = 0.25$ which is a more difficult case for reg-mod-BPDN. For Fig. 2(c), we also used a larger noise variance $\sigma_w^2 = 10^{-4}$. All other parameters were the same.

In Fig. 3, we show a plot of reg-mod-BPDN and BPDN from Fig 2(a) extended all the way to $|\Delta|/|N| = 1$ (which is the same as $\Delta = N$). Notice that if $|\Delta_e| = 0$, then the point $|\Delta|/|N| = 1$ of reg-mod-BPDN (or of mod-BPDN) is the same as BPDN. But in this plot, $|\Delta_e| = 3$ and hence the two points are different, even though the errors are quite similar.

For applications where some training data is available, γ and λ for reg-mod-BPDN can be chosen by interpreting the reg-mod-BPDN solution as the maximum a posteriori (MAP) estimate under a certain prior signal model (assume x_T is Gaussian with mean $\hat{\mu}_T$ and variance σ_p^2 and x_{T^c} is independent of x_T and is iid Laplacian with parameter b). This idea is explained in detail in [14]. However, there is no easy way to do this for the other methods. Alternatively, choosing γ and λ according to Theorem 3 gives
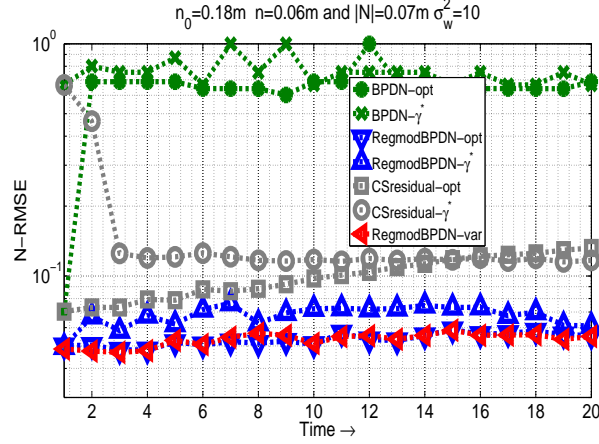
Fig. 4. Reconstructing a 32×32 block of the actual (compressible) larynx sequence from partial Fourier measurements. Measurements $n = 0.18m$ for $t = 0$ and $n = 0.06m$ for $t > 0$. Reg-mod-BPDN has the smallest reconstruction error among all methods.

another good start point. We can do this for mod-BPDN and BPDN, but we cannot do this for the other methods (we show examples using this approach later). For a fair error comparison, for each algorithm, we selected γ from a set of values $[0.00001\ 0.00005\ 0.0001\ 0.0005\ 0.001\ 0.005\ 0.01\ 0.1]$. We tried all these values for a small number of simulations (10 simulations) and then picked the best one (one with the smallest N-RMSE) for each algorithm. For weighted ℓ_1 reconstruction, we also pick the best γ' in (9) from the same set in the same way[4]. For reg-mod-BPDN, λ should be larger when the signal estimate is good and should be decreased when the signal estimate is not so good. We can use $\lambda = \alpha\sigma_w^2/\sigma_p^2$ to adaptively determine its value for different choices of σ_w^2 and σ_p^2. In our simulations, we used $\alpha = 0.2$ for Fig. 2 (a), (b) and (d) and $\alpha = 0.05$ for Fig. 2(c).

We fixed the chosen γ, γ' and λ and did Monte Carlo averaging over 100 simulations. We conclude the following. (1) When the signal estimate is not good (Fig. 2(b),(d)) or when n is small (Fig. 2(a),(b)), CS-residual and CS-mod-residual have significantly larger error than reg-mod-BPDN. (2) In case of Fig. 2(d) ($n = 0.3m$), they also have larger error than mod-BPDN. (3) In all four cases, weighed ℓ_1 and mod-BPDN have similar performance. This is also similar to that of reg-mod-BPDN in case of $n = 0.3m$, but is much worse in case of $n = 0.13m$. (4) We also show a comparison with regmodBPDN-var in Fig. 2(a). Notice that it has larger errors than reg-mod-BPDN for reasons explained in Sec. I-C.

---

[4]To give an example, our finally selected numbers for Fig. 2(d) were $\gamma = 0.01, 0.001, 0.001, 0.001, 0.001, 0.001, 0.01, 0.01$ for BPDN, mod-BPDN, reg-mod-BPDN, weighted ℓ_1, LS-CS, CS-residual, CS-mod-residual, mod-CS-residual respectively and $\gamma' = 0.0001$

C. Dynamic MRI application using γ from Theorem 3

In Fig. 4, we show comparisons for simulated dynamic MR imaging of an actual larynx image sequence (Fig. 1 (a)(i)). The larynx image is not exactly sparse but is only compressible in the wavelet domain. We used a two-level Daubechies-4 2D discrete wavelet transform (DWT). The 99%-energy support size of its wavelet transform vector, $|N_t| \approx 0.07m$. Also, $|\Delta_t| \approx 0.001m$ and $|\Delta_{e,t}| \approx 0.002m$. We used a 32×32 block of this sequence and at each time and simulated undersampled MRI, i.e. we selected n 2D discrete Fourier transform (DFT) coefficients using the variable density sampling scheme of [32], and added iid Gaussian noise with zero mean and variance $\sigma_w^2 = 10$ to each of them. Using a small 32×32 block allows easy implementation using CVX (for full sized image sequences, one needs specialized code). We used $n_0 = 0.18m$ at $t = 0$ and $n = 0.06m$ at $t > 0$.

We implemented dynamic reg-mod-BPDN as described in Algorithm 1. In this problem, the matrix $A = F_u \cdot W^{-1}$ where F_u contains the selected rows of the 2D-DFT matrix and W is the inverse 2D-DWT matrix (for a two-level Daubechies-4 wavelet). Reg-mod-BPDN was compared with similarly implemented reg-mod-BPDN-var and CS-residual algorithms (CS-residual only solved simple BPDN at $t = 0$). We also compared with simple BPDN (BPDN done for each frame separately). For reg-mod-BPDN and reg-mod-BPDN-var, the support estimation threshold, ρ, was chosen as suggested in [14]: we used $\rho = 20$ which is slightly larger than the smallest magnitude element in the 99%-energy support which is 15. At $t = 0$, we used T_0 to be the set of indices of the wavelet approximation coefficients. To choose γ and λ we tried two different things. (a) We used λ and γ from the set $[0.00001\ 0.00005\ 0.0001\ 0.0005\ 0.001\ 0.005\ 0.01\ 0.1]$ to do the reconstruction for a short training sequence (5 frames), and used the average error to pick the best λ and γ. We call the resulting reconstruction error plot reg-mod-BPDN-opt. (b) We computed the average of the γ^* obtained from Theorem 3 for the 5-frame training sequence and used this as γ for the test sequence. We selected λ from the above set by choosing the one that minimizes the average of the bound of Theorem 3 for the 5 frames. We call the resulting error plot reg-mod-BPDN-γ^*. The same two things were also done for BPDN and CS-residual as well. For reg-mod-BPDN-var, we only did (a).

From Fig. 4, we can conclude the following. (1) Reg-mod-BPDN significantly outperforms the other methods when using so few measurements. (2) Reg-mod-BPDN-var and reg-mod-BPDN have similar performance in this case. (3) The reconstruction performance of reg-mod-BPDN using γ^* from Theorem 3 is close to that of reg-mod-BPDN using the best γ chosen from a large set. This indicates that Theorem 3 provides a good way to select γ in practice.

D. Comparing the result of Theorem 1

In Table I, we compare the result of Theorem 1 for reg-mod-BPDN, mod-BPDN and BPDN. We used $m = 256$, $|N| = 26 = 0.1m$, $|\Delta| = 0.04|N| = |\Delta_e|$, $\sigma_p^2 = 10^{-3}$, $\beta_l = 1$ and $\beta_m = \beta_s = 0.25$. Also, $\sigma_w^2 = 10^{-5}$ and we varied n. For each experiment with a given n, we did the following. We did 100 Monte Carlo simulations. Each time, we evaluated the sufficient conditions for the bound of reg-mod-BPDN to hold. We say the bound *holds* if all the sufficient conditions hold for at least 98 realizations.

If this did not happen, we record *not hold* in Table I. If this did happen, then we recorded $\sqrt{\frac{\mathbb{E}[\text{bound}^2]}{\mathbb{E}[\|x\|_2^2]}}$ where $\mathbb{E}[\cdot]$ denotes the Monte Carlo average computed over those realizations for which the sufficient conditions do hold. Here, "bound" refers to the right hand side of (27) computed with $\gamma = \gamma_{T,\lambda}^*(\Delta)$ given in (26). An analogous procedure was followed for both mod-BPDN and BPDN.

The comparisons are summarized in Table I. For reg-mod-BPDN, we selected λ from the set $[0.00001 \ 0.00005 \ 0.0001 \ 0.0$ by picking the one that gave the smallest bound. Clearly the reg-mod-BPDN result holds with the smallest n, while the BPDN result needs a very large n ($n \geq 90\%$). Also even with $n = 90\%$, the BPDN error bound is very large.

n	Reg-mod-BPDN	Mod-BPDN	BPDN
$0.13m$	0.885	not hold	not hold
$0.19m$	0.161	0.303	not hold
$0.5m$	0.0199	0.0199	not hold
$0.9m$	0.014	0.014	0.27

TABLE I

SUFFICIENT CONDITIONS AND NORMALIZED BOUNDS COMPARISON OF REG-MOD-BPDN, MOD-BPDN AND BPDN. SIGNAL LENGTH $m = 256$, SUPPORT SIZE $|N| = 0.1m$, $|\Delta| = 4\%|N|$, $\Delta_e = 4\%|N|$, $\sigma_w^2 = 10^{-5}$ AND $\sigma_p^2 = 10^{-3}$. "NOT HOLD" MEANS THE ONE OR ALL OF THE SUFFICIENT CONDITIONS DOES NOT HOLD.

E. Comparing Theorems 1, 2, 3

In Fig. 5 (a), we compare the results from Theorems 1, 2 and 3 for one simulation. We plot $\frac{\text{bound}}{\|x\|_2}$ for $|\Delta|/|N|$ ranging from 0 to 0.2. Also, we used $m = 256$, $|N| = 26$, $|\Delta_e| = 0.1|N|$, $\sigma_p^2 = 10^{-3}$, $\beta_l = 1$ and $\beta_m = \beta_s = 0.25$. Also, $n = 0.13m$ and $\sigma_w^2 = 10^{-5}$. We used $\gamma = \gamma^*$ given in the respective theorems, and we set $\lambda = 10\sigma_w^2/\sigma_p^2$. We notice the following. (1) The bound of Theorem 1 is much larger than that of Theorem 2 or 3, even for $|\Delta| = 0.04|N|$. (2) For larger values of $|\Delta|$, the sufficient conditions of Theorem 1 do not hold and hence it does not provide a bound at all. (3) For reasons explained in Sec. IV, in this case, the bound of Theorem 3 is equal to that of Theorem 2. Recall that the computational complexity of the bound from Theorem 2 is exponential in $|\Delta|$. However if $|\Delta|$ is small, e.g. in our simulations $|\Delta| \leq 5$, this is still doable.

F. Upper bound comparisons using Theorem 3

In Fig. 5(b), we do two things. (1) We compare the reconstruction error bounds from Theorem 3 for reg-mod-BPDN, mod-BPDN and BPDN and compare them with the bounds for LS-CS error given in [31, Corollary 1]. All bounds hold without any sufficient conditions which is what makes this comparison possible. (2) We also use the γ^* given by Theorem 3 to obtain the reconstructions and compute the Monte Carlo averaged N-RMSE. Comparing this with the Monte Carlo averaged upper bound on the
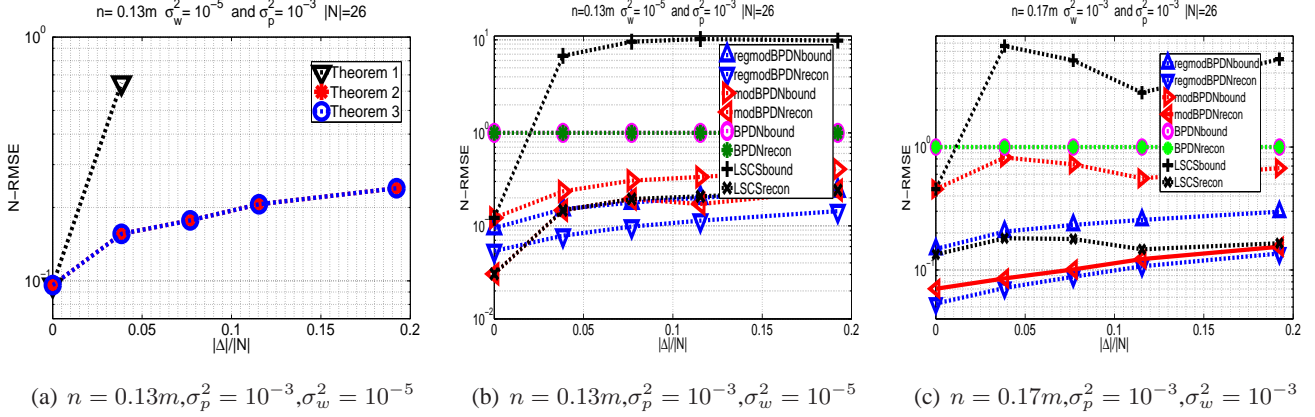
Fig. 5. In (a), we compare the three bounds from Theorem 1, 2 and 3 for one realization of x. In (b) and (c), we compare the normalized average bounds from Theorem 3 and reconstruction errors with random Gaussian and partial Fourier measurements respectively.

N-RMSE, $\sqrt{\frac{\mathbb{E}[\text{bound}^2]}{\mathbb{E}[\|x\|_2^2]}}$, allows us to evaluate the tightness of a bound. Here $\mathbb{E}[\cdot]$ denotes the mean computed over 100 Monte Carlo simulations and "bound" refers to the right hand side of (40). We used $m = 256$, $|N| = 26$, $|\Delta_e| = 0.1|N|$, $\sigma_p^2 = 10^{-3}$, $\beta_l = 1$, $\beta_m = \beta_s = 0.25$, and $|\Delta|$ was varied from 0 to $0.2|N|$. Also, $n = 0.13m$ and $\sigma_w^2 = 10^{-5}$.

From the figure, we can observe the following. (1) Reg-mod-BPDN has much smaller bounds than those of mod-BPDN, BPDN and LS-CS. The differences between reg-mod-BPDN and mod-BPDN bounds is minor when $|\Delta|$ is small but increases as $|\Delta|$ increases. (2) The conclusions from the reconstruction error comparisons are similar to those seen from the bound comparisons, indicating that the bound can serve as a useful proxy to decide which algorithm to use when (notice bound computation is much faster than computing the reconstruction error). (3) Also, reg-mod-BPDN and mod-BPDN bounds are quite tight as compared to the LS-CS bound. BPDN bound and error are both 100%. 100% error is seen because the reconstruction is the all zeros' vector.

In Fig. 5(c), we did a similar set of experiments for the case where A corresponds to a simulated MRI experiment, i.e. $A = F_u \cdot W^{-1}$ where F_u contains randomly selected rows of the 2D-DFT matrix and W is the inverse 2D-DWT matrix (for a two-level Daubechies-4 wavelet). We used $n = 0.17m$ and $\sigma_w^2 = 10^{-3}$. All other parameters were the same as in Fig. 5(b). Our conclusions are also the same.

The complexity for Theorem 3 is polynomial in $|\Delta|$ whereas that of the LS-CS bound [31, Corollary 1] is exponential in $|\Delta|$. To also show comparison with the LS-CS bound, we had to choose a small value of $m = 256$ so that the maximum value of $|\Delta| = 0.2|N| = 5$ was small enough. In terms of MATLAB time, computation of the Theorem 3 bound for reg-mod-BPDN took 0.2 seconds while computing the LS-CS bound took 1.2 seconds. For all methods except LS-CS, we were able to do the same thing fairly quickly even for $m = 4096$, or even larger. It took only 8 seconds to compute the bound of Theorem 3 when $m = 4096$, $n = 0.13m$, $|N| = 410 = 0.1m$ and $|\Delta| = |\Delta_e| = 0.1|N| = 41$.

VI. Conclusions and Future Work

In this work we studied the problem of sparse reconstruction from noisy undersampled measurements when partial and partly erroneous, knowledge of the signal's support and an erroneous estimate of the signal values on the "partly known support" is also available. Denote the support knowledge by T and the signal value estimate on T by $\hat{\mu}_T$. We proposed and studied a solution called regularized modified-BPDN which tries to find the signal that is sparsest outside T, while being "close enough" to $\hat{\mu}_T$ on T, and while satisfying the data constraint. We showed how to obtain computable error bounds that hold without any sufficient conditions. This made it easy to compare bounds for the various approaches (corresponding results for modified-BPDN and BPDN follow as direct corollaries). Empirical error comparisons with these and many other existing approaches are also provided.

In ongoing work, we are evaluating the utility of reg-mod-BPDN for recursive functional MR imaging to detect brain activation patterns in response to stimuli [33]. On the other end, we are also working on obtaining conditions under which it will remain "stable" (its error will be bounded by a time-invariant and small value) for a recursive recovery problem. In [30], this has been done for the constrained version of reg-mod-BPDN. That result uses the restricted isometry constants (RIC) and the restricted orthogonality constants (ROC) [17], [11] in its sufficient conditions and bounds. However, this means that the conditions and bounds are not computable. Also, since the stability holds under a different set of sufficient conditions and has a different error bound than that for mod-CS [34] or LS-CS [20] or CS [11], comparison of the various results is difficult. An open question is how to extend the results of the current work (which are computable) to show the stability of unconstrained reg-mod-BPDN.

APPENDIX

A. Proof of Proposition 1

When $\lambda = 0$, $Q_{T,0}(S) = A_{T\cup S}{}' A_{T\cup S}$. Thus, $Q_{T,\lambda}(S)$ is invertible iff $A_{T\cup S}$ is full rank. When $\lambda > 0$, $Q_{T,\lambda}(S)$ is as defined in (19). Apply block matrix inversion lemma

$$
\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} =
$$
$$
\begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}
$$

with $\mathbf{A} = A_T{}'A_T + \lambda I_T$, $\mathbf{B} = A_T{}'A_S$, $\mathbf{C} = A_S{}'A_T$ and $\mathbf{D} = A_S{}'A_S$, clearly $Q_{T,\lambda}(S)$ is invertible iff $A_S{}'A_S$ and $A_T{}'RA_T + \lambda I_T$ are invertible where $R := [I - A_S(A_S{}'A_S)^{-1}A_S']$. When A_S is full rank, (i) $A_S{}'A_S$ is full rank; and (ii) R is a projection matrix. Thus $R = R'R$ and so $A_T{}'RA_T = (RA_T)'(RA_T)$ is positive semi-definite. As a result, $A_T{}'RA_T + \lambda I_T$ is positive definite and thus invertible. Hence, when A_S is invertible, $Q_{T,\lambda}(S)$ is also invertible.

B. Proof of Theorem 1

In this subsection, we give the three lemmas for the proof of Theorem 1. *To keep notation simple we remove the subscripts $_{T,\lambda}$ from $Q(\Delta)$, M, $P(\Delta)$, $d(\Delta)$, $c(\Delta)$, $ERC(\Delta)$ in this and other Appendices.*

Lemma 1: Suppose that $Q(\Delta)$ is invertible, then

$$\|d(\Delta) - c(\Delta)\|_2 \leq \gamma\sqrt{|\Delta|} \cdot f_1(\Delta) \tag{44}$$

Lemma 1 can be obtained by setting $\nabla L(b) = 0$ and then using block matrix inversion on $Q(\Delta)$. The proof of Lemma 1 is in Appendix D1. Next, $\|c(\Delta) - x\|_2$ can be bounded using the following lemma.

Lemma 2: Suppose that $Q(\Delta)$ is invertible. Then

$$\|c(\Delta) - x\|_2 \leq \lambda f_2(\Delta)\|x_T - \hat{\mu}_T\|_2 + f_3(\Delta)\|w\|_2 \tag{45}$$

The proof of Lemma 2 is in Appendix D2.

Lemma 3: If $Q(\Delta)$ is invertible, $ERC(\Delta) > 0$, and $\gamma \geq \gamma^*(\Delta)$, then $L(b)$ has a unique minimizer which is equal to $d(\Delta)$.

Lemma 3 can be obtained in a fashion similar to [3], [1]. Its proof is given in Appendix D3.

Combining Lemmas 1, 2 and 3, and using the fact $\|d(\Delta) - x\|_2 \leq \|d(\Delta) - c(\Delta)\|_2 + \|c(\Delta) - x\|_2$, we get Theorem 1.

C. Proof of Theorem 2

The following lemma is needed for the proof of the corollaries leading to Theorem 2.

Lemma 4: Suppose that $Q(\tilde{\Delta})$ is invertible. Then

$$\|c(\tilde{\Delta}) - x\|_2 \leq$$
$$\lambda f_2(\tilde{\Delta})\|x_T - \hat{\mu}_T\|_2 + f_3(\tilde{\Delta})\|w\|_2 + f_4(\tilde{\Delta})\|x_{\Delta\setminus\tilde{\Delta}}\|_2 \tag{46}$$

Since $c(\tilde{\Delta})$ is only supported on $T \cup \tilde{\Delta}$ and $y = A_{T\cup\tilde{\Delta}}x_{T\cup\tilde{\Delta}} + A_{\Delta\setminus\tilde{\Delta}}x_{\Delta\setminus\tilde{\Delta}} + w$, the last term of (46) can be obtained by separating $x_{\Delta\setminus\tilde{\Delta}}$ out. The proof of Lemma 4 is given in Appendix D4.

Using Lemma 4, we can obtain Corollary 1 and then Corollary 2. Then minimize over all allowed $\tilde{\Delta}$'s in Corollary 2, we get Theorem 2. The proof of Corollary 1 and 2 are given as follows.

1) Proof of Corollary 1: Notice from the proof of Lemma 1 and Lemma 3 that nothing in the result changes if we replace Δ by a $\tilde{\Delta} \subseteq \Delta$. By Lemma 1 for $\tilde{\Delta}$, we are able to bound $\|d(\tilde{\Delta}) - c(\tilde{\Delta})\|_2$. Hence, we get the first term of (29). Next, invoke Lemma 4 to bound $\|c(\tilde{\Delta}) - x\|_2$ and we can obtain the rest three terms of (29). Lemma 3 for $\tilde{\Delta}$ gives the sufficient conditions under which $d(\tilde{\Delta})$ is the unique unconstrained minimizer of $L(b)$.

2) Proof of Corollary 2: Corollary 2 is obtained by bounding $\gamma^*(\tilde{\Delta})$. $\gamma^*(\tilde{\Delta}) = \|A_{(T\cup\tilde{\Delta})^c}'(y - Ac(\tilde{\Delta}))\|_\infty / ERC(\tilde{\Delta})$ can be bounded by rewriting $y - Ac(\tilde{\Delta}) = A_{T\cup\Delta}(x_{T\cup\Delta} - (c(\tilde{\Delta}))_{T\cup\Delta}) + w$ and then bounding $\|x_{T\cup\Delta} - (c(\tilde{\Delta}))_{T\cup\Delta}\|_2 = \|x - c(\tilde{\Delta})\|_2$ using Lemma 4. Doing this, we get

$$\|A_{(T\cup\tilde{\Delta})^c}'(y - Ac(\tilde{\Delta}))\|_\infty$$

$$\leq \max_{i\notin T\cup\tilde{\Delta}} |A_i' A_{T\cup\Delta}(x_{T\cup\Delta} - (c(\tilde{\Delta}))_{T\cup\Delta})| + |A_i'w|$$

$$\leq \max_{i\notin T\cup\tilde{\Delta}} \|A_i' A_{T\cup\Delta}\|_2 \|x_{T\cup\Delta} - (c(\tilde{\Delta}))_{T\cup\Delta}\|_2 + |A_i'w|$$

$$\leq \text{maxcor}(\tilde{\Delta})\lambda f_2(\tilde{\Delta})\|x_T - \mu_T\|_2 + \text{maxcor}(\tilde{\Delta})f_3(\tilde{\Delta})\|w\|_2$$

$$+\text{maxcor}(\tilde{\Delta})f_4(\tilde{\Delta})\|x_{\Delta\setminus\tilde{\Delta}}\|_2 + \|A_{(T\cup\tilde{\Delta})^c}'w\|_\infty$$

Using the above inequality to bound $\gamma^*(\tilde{\Delta})$ and replacing γ in $f(T, \lambda, \Delta, \tilde{\Delta}, \gamma)$, given in (29), by this bound, we can get (31).

D. Proof of Lemmas 1, 2, 3, 4

1) Proof of Lemma 1: We use the approach of [3, Lemma 3]. We can minimize the function $L(b)$ over all vectors supported on set $T \cup \Delta$ by minimizing:

$$F(b) = \frac{1}{2}\|y - A_{T\cup\Delta}b_{T\cup\Delta}\|_2^2 + \frac{1}{2}\lambda\|b_T - \hat{\mu}_T\|_2^2 + \gamma\|b_\Delta\|_1 \tag{47}$$

Since $Q(\Delta)$ is invertible, $F(b)$ is strictly convex as a function of $b_{T\cup\Delta}$. Then at the unique minimizer, $d(\Delta)$, $0 \in \nabla F(b)|_{b=d(\Delta)}$. Let $\partial\|b_{T^c}\|_1|_{b=d(\Delta)}$ denote the subgradient set of $\|b_{T^c}\|_1$ at $b = d(\Delta)$. Then clearly any ϕ in this set satisfies

$$\phi_T = 0 \tag{48}$$

$$\|\phi_{T^c}\|_\infty \leq 1 \tag{49}$$

Now, $0 \in \nabla F(b)|_{b=d(\Delta)}$ implies that

$$(A_{T\cup\Delta}'A_{T\cup\Delta})[d(\Delta)]_{T\cup\Delta} - A_{T\cup\Delta}'y$$

$$+\lambda \begin{bmatrix} [d(\Delta)]_T - \hat{\mu}_T \\ \mathbf{0}_\Delta \end{bmatrix} + \gamma\phi_{T\cup\Delta} = 0 \tag{50}$$

Simplifying the above equation, we get

$$[d(\Delta)]_{T\cup\Delta} = Q(\Delta)^{-1}(A_{T\cup\Delta}'y + \lambda \begin{bmatrix} \hat{\mu}_T \\ \mathbf{0}_\Delta \end{bmatrix} - \gamma\phi_{T\cup\Delta}) \tag{51}$$

Therefore, using (48) and (24), we have

$$[c(\Delta)]_{T\cup\Delta} - [d(\Delta)]_{T\cup\Delta} = Q(\Delta)^{-1} \begin{bmatrix} \mathbf{0}_T \\ \gamma\phi_\Delta \end{bmatrix} \tag{52}$$

Since

$$Q(\Delta) = \begin{bmatrix} A_T{'}A_T + \lambda I_T & A_T{'}A_\Delta \\ A_\Delta{'}A_T & A_\Delta{'}A_\Delta \end{bmatrix}, \tag{53}$$

using the block matrix inversion lemma

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} =$$

$$\begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix}$$

with $\mathbf{A} = A_T{'}A_T + \lambda I_T$, $\mathbf{B} = A_T{'}A_\Delta$, $\mathbf{C} = A_\Delta{'}A_T$ and $\mathbf{D} = A_\Delta{'}A_\Delta$ and using $\phi_T = 0$, we obtain

$$[c(\Delta)]_{T\cup\Delta} - [d(\Delta)]_{T\cup\Delta} =$$

$$\begin{bmatrix} -\gamma(A_T{'}A_T + \lambda I_{|T|})^{-1}A_T A_\Delta (A_\Delta{'}MA_\Delta)^{-1}\phi_\Delta \\ \gamma(A_\Delta{'}MA_\Delta)^{-1}\phi_\Delta \end{bmatrix}$$

Since $\|\phi_\Delta\|_\infty \leq 1$, the bound of (44) follows.

2) Proof of Lemma 2: Recall $c(\Delta)$ is given in (24). Since both x and $c(\Delta)$ are zero outside $T \cup \Delta$, then $\|c(\Delta) - x\|_2 = \|[c(\Delta)]_{T\cup\Delta} - x_{T\cup\Delta}\|_2$. With $y = Ax + w$ and $Ax = A_{T\cup\Delta}x_{T\cup\Delta}$, we have

$$A_{T\cup\Delta}{'}y = A_{T\cup\Delta}{'}(A_{T\cup\Delta}x_{T\cup\Delta} + w) \tag{54}$$

Notice $A'_{T\cup\Delta}A_{T\cup\Delta} = Q(\Delta) - \lambda \begin{bmatrix} I_T & \mathbf{0}_{T,S} \\ \mathbf{0}_{S,T} & \mathbf{0}_{S,S} \end{bmatrix}$. Using (54), we obtain the following equation

$$A_{T\cup\Delta}{'}y = Q(\Delta)x_{T\cup\Delta} - \lambda \begin{bmatrix} x_T \\ \mathbf{0}_\Delta \end{bmatrix} + A_{T\cup\Delta}{'}w \tag{55}$$

Then, using (24) we can obtain

$$[c(\Delta)]_{T\cup\Delta} - x_{T\cup\Delta} = \lambda Q(\Delta)^{-1} \begin{bmatrix} \hat{\mu}_T - x_T \\ \mathbf{0}_\Delta \end{bmatrix} + Q(\Delta)^{-1}A_{T\cup\Delta}{'}w$$

Finally, this gives (45).

3) Proof of Lemma 3: The proof is similar to that in [3] and [1]. Recall that $d(\Delta)$ minimizes the function $L(b)$ over all b supported on $T \cup \Delta$. We need to show that if $\gamma \geq \gamma^*(\Delta)$, then $d(\Delta)$ is the unique global minimizer of $L(b)$.

The idea is to prove under the given condition, any small perturbation h on $d(\Delta)$ will increase function $L(d(\Delta))$,i.e. $L(d(\Delta) + h) - L(d(\Delta)) > 0, \forall \|h\|_\infty \leq \epsilon$ for ϵ small enough. Then since $L(b)$ is a convex function, $d(\Delta)$ will be the unique global minimizer[3].

Similar to [1], we first split the perturbation into two parts $h = u + v$ where u is supported on $T \cup \Delta$ and v is supported on $(T \cup \Delta)^c$. Clearly $\|u\|_\infty \leq \|h\|_\infty \leq \epsilon$. We consider the case $v \neq 0$ since the case

$v = 0$ is already covered in Lemma 1. Then

$$L(d(\Delta) + h) = \frac{1}{2}\|y - A(d(\Delta) + u) - Av\|_2^2 +$$

$$\frac{1}{2}\lambda\|[d(\Delta)]_T + u_T + v_T - \hat{\mu}_T\|_2^2 + \gamma\|(d(\Delta) + u)_{T^c} + v_{T^c}\|_1$$

Then, we can obtain

$$L(d(\Delta) + h) - L(d(\Delta)) = L(d(\Delta) + u) - L(d(\Delta))$$

$$+\frac{1}{2}\|Av\|_2^2 - \langle y - Ad(\Delta), Av\rangle + \langle Au, Av\rangle + \gamma\|v_{T^c}\|_1$$

Since $d(\Delta)$ minimizes $L(b)$ over all vectors supported on $T \cup \Delta$, $L(d(\Delta)+u)-L(d(\Delta)) \geq 0$. Then since $L(d(\Delta) + u) - L(d(\Delta)) \geq 0$ and $\|Av\|_2^2 \geq 0$, we need to prove that the rest are positive,i.e.,$\gamma\|v_{T^c}\|_1 - \langle y - Ad(\Delta), Av\rangle + \langle Au, Av\rangle \geq 0$. Instead, we can prove this by proving a stronger condition $\gamma\|v_{T^c}\|_1 - |\langle y - Ad(\Delta), Av\rangle| - |\langle Au, Av\rangle| \geq 0$. Since $\langle y - Ad(\Delta), Av\rangle = v'A'(y - Ad(\Delta))$ and v is supported on $(T \cup \Delta)^c$,

$$|\langle y - Ad(\Delta), Av\rangle| = |v_{(T\cup\Delta)^c}'A_{(T\cup\Delta)^c}'(y - Ad(\Delta))|$$

$$\leq \|v\|_1\|A_{(T\cup\Delta)^c}'(y - Ad(\Delta))\|_\infty$$

Thus,

$$|\langle y - Ad(\Delta), Av\rangle| \leq \max_{\omega\notin T\cup\Delta} |\langle y - Ad(\Delta), A_\omega\rangle\|v\|_1$$

Meanwhile,

$$|\langle Au, Av\rangle| \leq \|A'Au\|_\infty\|v\|_1 \leq \epsilon\|A'A\|_\infty\|v\|_1 \tag{56}$$

And $\|v\|_1 = \|v_{T^c}\|_1$ since v is supported on $(T \cup \Delta)^c \subseteq T^c$. Then what we need to prove is

$$\left[\gamma - \max_{\omega\notin T\cup\Delta} |\langle y - Ad(\Delta), A_\omega\rangle| - \epsilon\|A'A\|_\infty\right]\|v\|_1 > 0 \tag{57}$$

Since we can select $\epsilon > 0$ as small as possible, then we just need to show

$$\gamma - \max_{\omega\notin T\cup\Delta} |\langle y - Ad(\Delta), A_\omega\rangle| > 0 \tag{58}$$

Since $y - Ad(\Delta) = (y - Ac(\Delta)) + A(c(\Delta) - d(\Delta))$, and by Lemma 1 we know $A(c(\Delta) - d(\Delta)) = \gamma MA_\Delta(A_\Delta'MA_\Delta)^{-1}\phi_\Delta$ and since $\|\phi_\Delta\|_\infty \leq 1$, we conclude that $d(\Delta)$ is the unique global minimizer if

$$\|A_{(T\cup\Delta)^c}'(y - Ac(\Delta))\|_\infty < \gamma\left[1 - \max_{\omega\notin T\cup\Delta} \|P(\Delta)A_\Delta'MA_\omega\|_1\right] \tag{59}$$

Next, we will show that $d(\Delta)$ is also the unique global minimizer under the following condition

$$\|A_{(T\cup\Delta)^c}'(y - Ac_{T,\lambda}(\Delta))\|_\infty = \gamma\left[1 - \max_{\omega\notin T\cup\Delta} \|P(\Delta)A_\Delta'MA_\omega\|_1\right] \tag{60}$$

Since the perturbation $h \neq 0$, then $u \neq 0$ or $v \neq 0$. Therefore, we will discuss the following three cases.

1) $u \neq 0$. In this case, we know $L(d(\Delta) + u) - L(d(\Delta)) > 0$ since $d(\Delta)$ is the unique minimizer over all vectors supported on $T \cup \Delta$. Therefore, $L(d(\Delta) + h) - L(d(\Delta)) > 0$ if (60) holds.

2) $u = 0$, $v \neq 0$ and v is not in the null space of A, i.e., $Av \neq 0$. In this case, we know $\|Av\|_2^2 > 0$. Hence, $L(d(\Delta) + h) - L(d(\Delta)) > 0$ when (60) holds.

3) $u = 0$, $v \neq 0$ and $Av = 0$. In this case, $L(d(\Delta) + h) - L(d(\Delta)) = \gamma\|v_{T^c}\|_1$. Thus, $L(d(\Delta) + h) - L(d(\Delta)) > 0$ if $\gamma > 0$. Clearly, $L(d(\Delta) + h) - L(d(\Delta)) > 0$ when (60) holds.

Finally, combining (59) and (60), we can conclude that $d(\Delta)$ is the unique global minimizer if the following condition holds

$$\|A_{(T \cup \Delta)^c}{}'(y - Ac(\Delta))\|_\infty \leq \gamma \mathrm{ERC}(\Delta) \tag{61}$$

4) Proof of Lemma 4: Consider a $\tilde{\Delta} \subseteq \Delta$ such that $A_{\tilde{\Delta}}$ has full rank. Since $A_{T \cup \tilde{\Delta}}{}'y = A_{T \cup \tilde{\Delta}}{}'(A_{T \cup \tilde{\Delta}}x_{T \cup \tilde{\Delta}} + w + A_{\Delta \setminus \tilde{\Delta}}x_{\Delta \setminus \tilde{\Delta}})$, expanding these terms we have

$$A_{T \cup \tilde{\Delta}}{}'y = Q(\Delta)x_{T \cup \tilde{\Delta}} - \lambda \begin{bmatrix} x_T \\ \mathbf{0}_{\tilde{\Delta}} \end{bmatrix} + A_{T \cup \tilde{\Delta}}{}'w + A_{T \cup \tilde{\Delta}}{}'A_{\Delta \setminus \tilde{\Delta}}x_{\Delta \setminus \tilde{\Delta}} \tag{62}$$

Then, using this in the expression for $c(\tilde{\Delta})$ from (24), we get

$$\begin{aligned}
[c(\tilde{\Delta})]_{T \cup \Delta} - x_{T \cup \Delta} &= \begin{bmatrix} \lambda Q(\tilde{\Delta})^{-1} \begin{bmatrix} \hat{\mu}_T - x_T \\ \mathbf{0}_{\tilde{\Delta}} \end{bmatrix} \\ \mathbf{0}_{\Delta \setminus \tilde{\Delta}} \end{bmatrix} \\
&+ \begin{bmatrix} Q(\tilde{\Delta})^{-1}A_{T \cup \tilde{\Delta}}{}'w \\ \mathbf{0}_{\Delta \setminus \tilde{\Delta}} \end{bmatrix} + \begin{bmatrix} Q(\tilde{\Delta})^{-1}A_{T \cup \tilde{\Delta}}{}'A_{\Delta \setminus \tilde{\Delta}}x_{\Delta \setminus \tilde{\Delta}} \\ -x_{\Delta \setminus \tilde{\Delta}} \end{bmatrix}
\end{aligned} \tag{63}$$

Therefore, we get (46).

References

[1] W. Lu and N. Vaswani, "Modified basis pursuit denoising (modified-bpdn) for noisy compressive sensing with partially known support," in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, 2010.

[2] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal of Scientific Computation*, vol. 20, pp. 33 – 61, 1998.

[3] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. on Information Theory*, vol. 52(3), pp. 1030 – 1051, March 2006.

[4] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Th.*, vol. 52(2), pp. 489 – 509, February 2006.

[5] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52(4), pp. 1289 – 1306, April 2006.

[6] "Rice compressive sensing resources [online]," in *Available: http://www.dsp.rice.edu/cs*.

[7] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, "Compressive imaging for video representation and coding," in *Proc. Picture Coding Symposium (PCS), Beijing, China*, April 2006.

[8] C. Qiu and N. Vaswani, "Reprocs: A missing link between recursive robust pca and recursive sparse recovery in large but correlated noise," *arXiv: 1106.3286*, 2011.

[9] ——, "Support-predicted modified-cs for principal components' pursuit," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2011.

[10] X. Liu and J. U. Kang, "Compressive sd-oct: the application of compressed sensing in spectral domain optical coherence tomography," *Optics Express*, vol. 18, pp. 22 010 – 22 019, 2010.

[11] E. Candes, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de lAcademie des Sciences, Paris, Serie I*, 2008.

[12] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59(8), pp. 1207–1223, August 2006.

[13] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2009.

[14] ——, "Modified-cs: Modifying compressive sensing for problems with partially known support," *IEEE Trans. on Sig. Proc.*, vol. 58(9), pp. 4595 – 4607, September 2010.

[15] A. Khajehnejad, W. Xu, A. Avestimehr, and B. Hassibi, "Weighted l1 minimization for sparse recovery with prior information," in *IEEE Intl. Symp. Info. Theory(ISIT)*, 2009.

[16] R. von Borries, C. J. Miosso, and C. Potes, "Compressive sensing reconstruction with prior information by iteratively reweighted least-squares," *IEEE Transactions on Signal Processing*, vol. 57(6), pp. 2424 – 2431, June 2009.

[17] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. on Infomation Theory*, vol. 51(12), pp. 4203 – 4215, 2005.

[18] L. Jacques, "A short note on compressed sensing with partially known signal support," *Signal Processing*, vol. 90(12), pp. 3308–3312, December 2010.

[19] D. Donoho and J. Tanner, "High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension," vol. 102(27), pp. 617–652, 2006.

[20] N. Vaswani, "Ls-cs-residual (ls-cs): Compressive sensing on the least squares residual," *IEEE Trans. Signal Process.*, vol. 58(8), pp. 4108 – 4120, August 2010.

[21] ——, "Kalman filtered compressed sensing," in *IEEE Intl. Conf. Image Proc. (ICIP)*, 2008.

[22] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. on Information Theory*, vol. 56, pp. 1982–2001, April 2010.

[23] Y. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. on Information Theory*, vol. 55(11), pp. 5302 – 5316, November 2009.

[24] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56(6), pp. 2346–2356, June 2008.

[25] P. Schniter, L. Potter, and J. Ziniel, "Fast bayesian matching pursuit: Model uncertainty and parameter estimation for sparse linear models," in *Information Theory and Applications (ITA)*, 2008.

[26] S. Som, L. C. Potter, and P. Schniter, "Compressive imaging using approximate message passing and a markov-tree prior," in *Asilomar Conf. on Sig. Sys. Comp.*, 2010.

[27] C. La and M. Do, "Signal reconstruction using sparse tree representations," in *SPIE Wavelets XI, San Diego, California*, September 2005.

[28] M. Duarte, M. Wakin, and R. Baraniuk, "Fast reconstruction of piecewise smooth signals from random projections," in *SPARS Workshop*, November 2005.

[29] W. Lu and N. Vaswani, "Modified compressive sensing for real-time dynamic mr imaging," in *IEEE Intl. Conf. Image Proc. (ICIP)*, 2009.

[30] F. Raisali, "Stability (over time) of regularized modified-cs (noisy) for recursive causal sparse reconstruction," in *45th Annual Conference on Information Sciences and Systems (CISS)*, 2011.

[31] C. Qiu and N. Vaswani, "Ls-cs-residual (ls-cs): Compressive sensing on the least squares residual," in *http://home.engineering.iastate.edu/~chenlu/KFLS_v3*.

[32] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine*, vol. 58(6), pp. 1182–1195, December 2007.

[33] W. Lu, T. Li, I. C. Atkinson, and N. Vaswani, "Modified-cs-residual for recursive reconstruction of highly undersampled functional mri sequences," in *IEEE Intl. Conf. Image Proc. (ICIP)*, 2011.

[34] N. Vaswani, "Stability (over time) of Modified-CS for Recursive Causal Sparse Reconstruction," in *Allerton Conf. Communication, Control, and Computing*, 2010.