Stability of Modified-CS and LS-CS for Recursive Reconstruction of Sparse Signal Sequences

Namrata Vaswani
Dept. of Electrical and Computer Engineering
Iowa State University, Ames, IA 50011, USA
Email: namrata@iastate.edu

Abstract—In this work, we obtain sufficient conditions for the “stability” of our recently proposed algorithms, Least Squares Compressive Sensing residual (LS-CS) and modified-CS, for recursively reconstructing sparse signal sequences from noisy measurements. By “stability” we mean that the number of misses from the current support estimate and the number of extras in it remain bounded by a time-invariant value at all times. We show that, for a signal model with fixed signal power and support set size; support set changes allowed at every time; and gradual coefficient magnitude increase/decrease, “stability” holds under mild assumptions – bounded noise, high enough minimum nonzero coefficient magnitude increase rate, and large enough number of measurements at every time. A direct corollary is that the reconstruction error is also bounded by a time-invariant value at all times. If the support of the sparse signal sequence changes slowly over time, our results hold under weaker assumptions than what simple compressive sensing (CS) needs for the same error bound. Also, our support error bounds are small compared to the support size. Our discussion is backed up by Monte Carlo simulation based comparisons.

I. INTRODUCTION

The static sparse reconstruction problem has been studied for a while [2], [3], [4]. The recent papers on compressive sensing (CS) [5], [6], [7], [8], [9], [10] (and many other more recent works) provide the missing theoretical guarantees – conditions for exact recovery and error bounds when exact recovery is not possible. But for recovering a time sequence of sparse signals, with time-varying sparsity patterns, most existing approaches are batch methods, e.g. [11], [12]. Our recent work on Least Squares CS-residual (LS-CS) and Kalman filtered CS-residual (KF-CS) [13], [14], and later on modified-CS [15], [16], first studied the problem of recursively recovering a time sequence of sparse signals, with time-varying sparsity patterns, using much fewer measurements than what simple CS (CS done at each time separately) needs. By “recursive” reconstruction, we mean that we want to use only the current measurements’ vector and the previous reconstructed signal to reconstruct the current signal. The storage and computational complexity of these solutions is only as much as that of simple CS, but their reconstruction performance is significantly better. LS-CS and modified-CS only use the assumption that the sparsity pattern (support in the sparsity basis) changes slowly over time. As we show in Fig. 1 and in [16], this is a valid assumption for many medical image sequences. KF-CS also uses slow signal value change.

Denote the support estimate from the previous time by \( T \). Modified-CS tries to find a signal that is sparsest outside of \( T \) among all signals that satisfy the data constraint. It was first introduced in [15], [16], where we studied the noise-free case and obtained exact recovery conditions for it. LS-CS uses a different approach. It replaces CS on the observation by CS on the least squares (LS) residual computed by assuming that \( T \) is the correct support [13], [14]. In this work, we obtain the conditions required for “stability” of LS-CS, modified-CS and of an improved version of modified-CS which we call “modified-CS with add-LS-del” (improves the support estimation step of modified-CS). By “stability” we mean that the number of misses from the current support estimate and the number of extras in it remain bounded by a time-invariant value at all times. A direct corollary is that the reconstruction errors are also bounded by a time-invariant value at all times.

A. Related Work

LS-CS and modified-CS are causal and recursive approaches that only rely on the slow support change assumption. Another causal and recursive approach, that uses approximate belief propagation, has been proposed in very recent work [17]. This is a fully Bayesian approach that assumes prior probabilistic models on both slow support and slow signal value change. Some very interesting numerical experiments are shown.

“Recursive sparse reconstruction” also sometimes refers to homotopy methods, e.g. [18], [19], whose goal is to use the past reconstructions and homotopy to speed up the current optimization, but not to achieve accurate recovery from fewer measurements (than what simple CS needs). Algorithms that improve the reconstruction of a single signal recursively as more measurements come in, such as those in [20], [21], [19], are also sometimes referred to as “recursive sparse recovery” algorithms. Clearly, the goals in the above works are quite different from ours.

Also, causal but batch algorithms for recovering sparse signal sequences, with time-invariant support, from fewer measurements were proposed in [22].

Other related ideas in literature include the following. Two approaches related to modified-CS are [23] and weighted \( \ell_1 \) [24]. But both of these focus only on static sparse recovery with prior support knowledge. The work of [24] obtains exact...
recover thresholds for weighted \( \ell_1 \), similar to those in [7], for the case when a probabilistic prior on the signal support is available. Iterative support estimation approaches (using the recovered support from the first iteration for a second weighted \( \ell_1 \) step and doing this iteratively) have been studied in recent work [25], [26], [27]. This is done for iteratively improving the recovery of a single signal.

To the best of our knowledge, stability over time has not been studied in the above works for recursive sparse recovery, except in [28] (KF-CS and LS-CS) or [14] (LS-CS). Our result from [28] is under strong assumptions, e.g. it is for a random walk signal change model (which has unbounded signal power and hence is the easier but unrealistic case), and it requires strong assumptions on the measurement matrix. Our result for LS-CS stability from [14] holds under mild assumptions and is for a fairly realistic signal change model. The only limitation is that it assumes that support changes occur “every-so-often” (every \( d \) time units, there are \( S_n \) support additions and removals). But from testing the slow support change assumption for real data (medical image sequences), it has been observed that support changes usually occur at every time, e.g. see Fig. 1. This important case is the focus of the current work. Moreover, in [14], we only studied LS-CS. In this work we study both LS-CS and modified-CS and also modified-CS with add-LS-del.

II. NOTATION, PROBLEM DEFINITION AND OVERVIEW OF RESULTS

We define notation and give the problem formulation in Sec. II-A. We give a brief overview of our results in Sec. II-B.

A. Notation and Problem Definition

We let \([1, m] \):= \([1, 2, \ldots, m]\). We use \( T^c \) to denote the complement of a set \( T \) w.r.t. \([1, m] \), i.e. \( T^c := \{ i \in [1, m] : i \notin T \} \). We use \(|T|\) to denote the cardinality of \( T \). Also, \( \emptyset \) denotes the empty set. The set operations \( \cup, \cap, \setminus \) have their usual meanings (recall that \( A \setminus B := A \cap B^c \)).

For a vector, \( v \), and a set, \( T \), \( v_T \) denotes the \(|T|\) length sub-vector containing the elements of \( v \) corresponding to the indices in the set \( T \). \( \|v\|_k \) denotes the \( k \) norm of a vector \( v \). If just \( \|v\| \) is used, it refers to \( \|v\|_2 \). Similarly, for a matrix \( M \), \( \|M\|_k \) denotes its induced \( k \)-norm, while just \( \|M\| \) refers to \( \|M\|_2 \). \( M^t \) denotes the transpose of \( M \) and \( M^{-1} \) denotes the Moore-Penrose pseudo-inverse of \( M \) (when \( M \) is tall, \( M^t M^{-1} \)). Also, \( M_{\setminus y} \) denotes the sub-matrix obtained by extracting the columns of \( M \) corresponding to indices in \( T \).

At all times, \( t > 0 \), we assume the following observation model:

\[
y_t = Ax_t + w_t, \quad \|w_t\| \leq \epsilon
\]

where \( x_t \) is an \( n \) length sparse vector with support \( S_t \); \( y_t \) is the \( n \times m \) length observation vector at time \( t \); and \( w_t \) is the observation noise. As we explain later, our algorithms need more measurements at the initial time, \( t = 0 \). We use \( n_0 \) to denote the number of measurements used at \( t = 0 \) and we use \( A_0 \) to denote the corresponding \( n_0 \times m \) measurement matrix, i.e. at \( t = 0 \), we have

\[
y_0 = A_0 x_0 + y_0, \quad \|w_0\| \leq \epsilon
\]

The term “support”, as usual, refers to the set of indices of the nonzero elements of \( x_t \).

Our goal is to recursively estimate \( x_t \) using \( y_1, \ldots, y_t \). By recursively, we mean, use only \( y_t \) and the estimate from \( t - 1 \), \( \hat{x}_{t-1} \), to compute the estimate at \( t \).

The \( S \)-restricted isometry constant (RIC) \( \delta_2 \), for the matrix, \( A \), is the smallest real number satisfying

\[
(1 - \delta_2)\|c\|^2 \leq \|A c\|^2 \leq (1 + \delta_2)\|c\|^2
\]

for all sets \( T \subset [1, m] \) of cardinality \(|T| \leq S \) and all real vectors \( c \) of length \(|T| \). The restricted orthogonality constant (ROC) \( \theta_{S_1, S_2} \), is the smallest real number satisfying

\[
|c_1' A T_1, c_2' A T_2| \leq \theta_{S_1, S_2} ||c_1|| ||c_2||
\]

for all disjoint sets \( T_1, T_2 \subset [1, m] \) with \(|T_1| \leq S_1 \), \(|T_2| \leq S_2 \) and \( S_1 + S_2 \leq m \), and for all vectors \( c_1, c_2 \) of length \(|T_1| \), \(|T_2| \) respectively.

In this work, \( \delta_2, \theta_{S_1, S_2} \) always refer to the RIC, ROC for the measurement matrix \( A \) which is used at \( t > 0 \). If we refer to the RIC of any other matrix, e.g. \( A_0 \), we use \( \delta(T, A_0) \).

We use \( \alpha \) to denote the support estimation threshold used by modified-CS and we use \( \alpha_{\text{add}}, \alpha_{\text{del}} \) to denote the support addition and deletion thresholds used by modified-CS with add-LS-del and by LS-CS. We use \( \hat{N}_t \) to denote the support

\[
\text{Fig. 1. Slow support change in medical image sequences.}
\]

The two-level Daubechies-4 2D discrete wavelet transform (DWT) served as the sparsity basis. Since real image sequences are only approximately sparse, we use \( N_l \) to denote the 99%-energy support of the DWT of these sequences. The support size, \(|N_l|\), was 6-7% of the image size for both sequences. We plot the number of additions (left) and the number of removals (right) as a fraction of \(|N_l|\). Notice that all changes are less than 2% of the support size.
estimate at time $t$. To keep notation simple, we avoid using the subscript $t$ wherever possible.

Definition 1 ($T_i$, $\Delta_i$, $\Delta_{e,t}$): We use $T_i := \hat{N}_{t-1}$ to denote the support estimate from the previous time. This serves as the predicted support at time $t$. We use $\Delta_i := N_t \setminus T_i$ to denote the unknown part of $T_i$ and $\Delta_{e,t} := T_i \setminus N_t$ to denote the "erroneous" part of $T_i$. In many places in the manuscript, we remove the subscript $t$ to keep notation simple. With the above definition, clearly,

$$N_t = T_i \cup \Delta_i \setminus \Delta_{e,t}.$$  

Definition 2 ($\hat{T}_t$, $\hat{\Delta}_t$, $\hat{\Delta}_{e,t}$): We use $\hat{T}_t := \hat{N}_t$ to denote the final estimate of the current support; $\hat{\Delta}_t := \hat{N}_t \setminus \hat{T}_t$ to denote the "misses" in $\hat{N}_t$ and $\hat{\Delta}_{e,t} := \hat{T}_t \setminus \hat{N}_t$ to denote the "extraneous".

We sometimes refer to $\Delta_i, \Delta_e$ as the predicted support errors and to $\Delta_i, \Delta_e$ as the final (or estimated) support errors. The sets $T_{\text{add}}, \Delta_{\text{add}}, \Delta_{e, \text{add}}$ are defined in Definition 4 (Sec. V).

If two sets $B, C$ are disjoint, we just write $D \cup B \setminus C$ instead of writing $(D \cup B) \setminus C$, e.g. $N_t = T_i \cup \Delta_e \setminus \Delta_e$.

We refer to the left (right) hand side of an equation or inequality as LHS (RHS).

In this work, "modified-CS" refers to the solution of (9). Also, simple CS refers to the solution of (9) with $T = \emptyset$.

B. Overview of Results

When measurements are noisy, the reconstruction errors of modified-CS and of LS-CS can easily be bounded as a function of the support size, $|N_i|$, and of the predicted support error sizes, $|\Delta_i|$ and $|\Delta_{e,t}|$ [28], [14]. The bound is small at small time $t$ if $|\Delta_i|$ and $|\Delta_{e,t}|$ are small enough. But smallness of the predicted support errors, $\Delta_i, \Delta_{e,t}$, depends on the accuracy of the previous reconstruction, and thus, in general, it may happen that, over time, the error bound keeps increasing. Such a result is of limited use for a recursive reconstruction problem. There is thus a need to obtain conditions under which one can show "stability", i.e. ensure that a time-invariant bound holds on the sizes of these support errors. Combining this with the error bound result will imply that the reconstruction error is also bounded by a time-invariant value at all times.

In this work, we obtain results for the stability of three algorithms: (a) modified-CS; (b) "modified-CS with add-LS-del" and (c) LS-CS. "Modified-CS with add-LS-del" improves the support estimation step of modified-CS by using a three step approach first introduced in [13], [14] and in [30], [31] – support addition with a smaller threshold, followed by LS estimation on the new support, and finally support deletion using the LS estimate. Using add-LS-del significantly improves both the stability result we can prove (as argued in Sec. V-B) and the empirical reconstruction performance we get (see Sec. VII).

All our results are obtained under a bounded observation noise assumption and for a signal model with

1) support changes ($S_a$ additions and $S_a$ removals) occurring at every time, $t$,

2) magnitude of the newly added coefficients increasing gradually, and similarly for decrease before removal,

3) support size, $|N_t| = S_0$ at all times and the signal power $\|x_t\|^2$, also constant at all times.

Our results have the following form. For a given number and type of measurements (i.e. for a given measurement matrix, $A$), and for a given noise bound, $\epsilon$, if,

1) the support estimation threshold(s) is/are appropriately set,

2) the support size, $S_0$, and the newly added (or removed) support size, $S_a$, are small enough,

3) the newly added coefficients’ increase rate (existing large coefficients’ decrease rate), $r$, is large enough, and

4) the initial number of measurements, $n_0$, is large enough for accurate initial reconstruction using simple CS, then, the support error sizes are bounded by time-invariant values: we show that $|\Delta_i| \leq 2S_a, |\Delta_{e,t}| = 0$ and $|\Delta_i| \leq 2S_a, |\Delta_{e,t}| \leq S_a$. A direct corollary is that the reconstruction error is also bounded by a time-invariant value at all times.

Remark 1: The reason we need to assume bounded noise is as follows. When the noise is unbounded, e.g. Gaussian, all error bounds for CS and, similarly, all error bounds for LS-CS or modified-CS hold with "large probability", e.g. see [22], [14]. For stability, we need the error bound for LS-CS or modified-CS to hold at all times, $0 \leq t < \infty$ (this, in turn, is used to ensure that the support gets estimated with bounded error at all times). Clearly, this will be a zero probability event. As an aside, most existing works which use the RIC based approach of Candes et al to bound the error of noisy sparse recovery, or of noisy sparse recovery with partial support knowledge, also assume bounded noise, e.g. [9], [10], [29], [30].

Remark 2: We should mention that constant or bounded signal power is both the more practical case (since, in practice, signal power never keeps increasing unboundedly) and is also the more difficult case. This is because the accuracy of the reconstruction at time $t + 1$ relies heavily on the correct detection of the small elements at time $t$. Correct detection will become easier for larger signal power (or, to be precise, for larger power of the smallest nonzero coefficients).

For our signal model, slow support change translates to $S_a \ll S_0$. Under this assumption, clearly, $2S_a \ll S_0$, and so our support error bounds are small compared to the support size, $S_0$, making our stability results meaningful. We can argue that our results hold under weaker assumptions (allow larger values of $S_0$), for a given measurement matrix $A$, than the corresponding simple CS (CS done at each time separately) result. Since simple CS is not a recursive approach, the CS error bound from [10] (or other works) also serves as a stability result for it. Also, we can argue that modified-CS with add-LS-del needs the weakest conditions on the number of measurements, $n$, and on the rate of coefficient magnitude increase/decrease, $r$. Modified-CS needs similar conditions on $n$, but needs a larger $r$. LS-CS needs the strongest conditions on both $n$ and $r$. Since we can only compare sufficient conditions or upper bounds, we back up all our discussion with simulation experiments to compare actual reconstruction
III. SIGNAL MODEL FOR STUDYING STABILITY

The modified-CS or LS-CS algorithms do not assume any signal model. But for showing stability, we need certain assumptions on the signal change over time.

Signal Model 1: Assume the following.

1) (addition) At each \( t > 0 \), \( S_a \) new coefficients get added to the support at magnitude \( r \). Denote this set by \( A_t \).

2) (increase) At each \( t > 0 \), the magnitude of \( S_a \) coefficients out of all those which had magnitude \((j-1)r\) at \( t-1 \) increases to \( jr \). This occurs for all \( 2 \leq j \leq d \). Thus the maximum magnitude reached by any coefficient is \( M := dr \).

3) (decrease) At each \( t > 0 \), the magnitude of \( S_a \) coefficients out of all those which had magnitude \((j+1)r\) at \( t-1 \) decreases to \( jr \). This occurs for all \( 1 \leq j \leq (d-2) \).

4) (removal) At each \( t > 0 \), \( S_a \) coefficients out of all those which had magnitude \( r \) at \( t-1 \) get removed from the support (magnitude becomes zero). Denote this set by \( R_t \).

5) (initial time) At \( t = 0 \), the support size is \( S_0 \) and it contains \( 2S_a \) elements each with magnitude \( r, 2r, \ldots, (d-1)r \), and \((S_0 - (2d-2)S_a)\) elements with magnitude \( M \).

We show an example of the above signal model in Fig. 2.

The above model has the following realistic features – (a) equal number, \( S_a \), of additions and removals to/from the support occur at every time, \( t \); (b) a newly added coefficient gets added at a small magnitude; (c) magnitude of any nonzero element either remains constant, or increases gradually at rate \( r \), but not beyond a maximum magnitude \( M := dr \), or decreases gradually at rate \( r \); and (d) at all times, the signals have the same support set size, \( |N_t| = S_0 \) and the same signal power, \( \|x_t\|^2 = (S_0 - (2d-2)S_a)M^2 + 2S_a \sum_{j=1}^{d-1} j^2r^2 \).

In practice, the number of additions/removals to the support is never exactly equal, but varies in a small range over time. A similar thing holds for the coefficient increase/decrease rate, \( r \), or for the stable magnitude, \( M \). But for notational simplicity, we ignore these variations. Also, in practice, different nonzero elements may have different magnitude increase rates, \( r_i \), and different stable magnitudes, \( M_i \). It will be possible to extend our results to this latter case fairly easily, and we expect that the result will require a lower bound on \( \min_i r_i \).

Signal Model II does not specify a particular generative model. Two examples of signal models that satisfy the above assumptions are given in Appendix A. Briefly, in the first model, at each \( t \), \( S_a \) new elements, randomly selected from \( N_{t-1} \), get added to the support at initial magnitude, \( r \), and equally likely sign. Their magnitude keeps increasing gradually, at rate \( r \), for \( d \) time units after which it becomes constant at \( M := dr \). The sign does not change. Also, at each time, \( t \), \( S_a \) randomly selected elements out of the “stable” elements’ set (set of elements which have magnitude \( M \) at \( t-1 \)), begin to decrease at rate \( r \) and this continues until their magnitude becomes zero, i.e., they get removed from the support. A second possible generative model randomly selects \( S_a \) out of the \( 2S_a \) current elements with magnitude \( jr \) and increases them, and decreases the other \( S_a \) elements.

To understand the implications of the assumptions in Signal Model I we define the following sets.

Definition 3: Define the following.

1) For all \( 0 \leq j \leq d-1 \), let

\[ D_t(j) := \{ i : |x_{t,i}| = jr, \ |x_{t-1,i}| = (j+1)r \} \]

denote the set of elements that decrease from \((j+1)r\) to \( jr \) at time, \( t \).

2) For all \( 1 \leq j \leq d \), let

\[ I_t(j) := \{ i : |x_{t,i}| = jr, \ |x_{t-1,i}| = (j-1)r \} \]

denote the set of elements that increase from \((j-1)r\) to \( jr \) at time, \( t \).

3) For all \( 1 \leq j \leq d-1 \), let

\[ S_t(j) := \{ i : 0 < |x_{t,i}| < jr \} \]

denote the set of small but nonzero elements, with smallness threshold \( jr \).

4) Clearly,

a) The newly added set,

\[ A_t = I_t(1) \]

b) The newly removed set,

\[ R_t = D_t(0) \]

c) \[ |I_t(j)| = S_a, \ |D_t(j)| = S_a, \ |S_t(j)| = 2(j-1)S_a \]

Consider a \( 1 < j < d \). From Signal Model I it is clear that at any time, \( t \), \( S_a \) elements enter the small elements’ set, \( S(j) \), from the bottom (set \( A_t \)) and \( S_a \) enter from the top (set \( D_t(j-1) \)). Similarly \( S_a \) elements leave \( S(j) \) from the bottom (set \( R_t \)) and \( S_a \) from the top (set \( I_t(j) \)). Thus,

\[ S_t(j) = S_{t-1}(j) \cup (A_t \cup D_t(j-1)) \setminus (R_t \cup I_t(j)) \]

\( ^2 \)To model the variations over time compactly, a probabilistic signal change model will be a better one. But that will make our analysis a lot more tricky since the reconstruction error bounds, which form the starting point for our stability results, do not assume any randomness \([10,14,29]\). In particular, they do not treat the sparse signal as a random variable.
To look at an example, see Fig. 3. Consider $j = 3$. Notice that $S_{t-1}(3) = \{2, 91, 12, 74\}$ and $S_t(3) = \{79, 12, 2, 66\}$. Also, $A_t = \{79\}$, $R_t = \{91\}$, $I_t(3) = \{74\}$ and $D_t(2) = \{66\}$. Clearly $\{2, 91, 12, 74\} \cup \{\{79\} \cup \{66\}\} \cup \{91\} \cup \{74\} = \{79, 12, 2, 66\}$, i.e. (5) holds.

Since $A_t$, $R_t$, $D_t(j - 1)$, $I_t(j)$ are mutually disjoint, $R_t \subseteq S_{t-1}(j)$ and $I_t(j) \subseteq S_{t-1}(j)$, thus, (5) implies that

$$S_{t-1}(j) \cup A_t \backslash R_t = S_t(j) \cup I_t(j) \backslash D_t(j - 1) \quad (6)$$

Also, clearly, from Signal Model 1

$$N_t = N_{t-1} \cup A_t \backslash R_t \quad (7)$$

We will use these in the proof of the results of Sec. VII

IV. STABILITY OF MODIFIED-CS

Modified-CS was first proposed in [13], [16] as a solution to the problem of sparse reconstruction with partial, and possibly erroneous, knowledge of the support. Denote this “known” support by $T$. Modified-CS tries to find a signal that is sparsest outside of the set $T$ among all signals satisfying the data constraint. In the noisy case, it solves $\min_x \|y - Ax\| \text{ s.t. } \|y - A\hat{x}\|_1 \leq \epsilon$. For recursively reconstructing a time sequence of sparse signals, we use the support estimate from the previous time, $N_{t-1}$, as the set $T$. The support is estimated by thresholding the output of modified-CS. At the initial time, $t = 0$, we let $T$ be the empty set, $\emptyset$, i.e. we do simple CS. Alternatively, as explained in [16], we can use prior knowledge of the initial signal’s support as the set $T$ at $t = 0$, e.g. for wavelet sparse images with no (or a small) black background, the set of indices of the approximation coefficients can form the set $T$. This prior knowledge is usually not as accurate. Thus, in either case, at $t = 0$ we need more measurements, i.e. $n_0 > n$.

In this work, for simplicity, we assume that simple CS is done at $t = 0$. We summarize the algorithm in Algorithm 1

Algorithm 1 Modified-CS

For $t \geq 0$, do

1) Simple CS. If $t = 0$, set $T = \emptyset$ and compute $\hat{x}_{t,t -,} \text{modcs}$ as the solution of

$$\min_{\beta} \|\beta\|_1 \text{ s.t. } \|y_0 - A\beta\|_1 \leq \epsilon \quad (8)$$

2) Modified-CS. If $t > 0$, set $T = \hat{N}_{t-1}$ and compute $\hat{x}_{t,t -,} \text{modcs}$ as the solution of

$$\min_{\beta} \|\beta\| \text{ s.t. } \|y_t - T\beta\|_1 \leq \epsilon \quad (9)$$

3) Estimate the Support. Compute $\hat{T}$ as

$$\hat{T} = \{i \in [1, m] : |(\hat{x}_{t,t -,} \text{modcs})_i| > \alpha\} \quad (10)$$

4) Set $\hat{N}_t = \hat{T}$. Output $\hat{x}_{t,t -,} \text{modcs}$, Feedback $\hat{N}_t$.

By adapting the approach of [10], the error of modified-CS can be bounded as a function of $T = |N| + |\Delta_\epsilon| - |\Delta|$ and $|\Delta|$. This was done in [29]. We state a modified version here.

Lemma 1 (modified-CS error bound): Let $x$ be a sparse vector with support $N$ and let $y := Ax + w$ with $\|w\| \leq \epsilon$. Also, let $\Delta := N \setminus T$ and $\Delta_\epsilon := T \setminus N$. Assume that $|N| = S_N$, $|\Delta_\epsilon| \leq S_{\Delta_\epsilon}$ and $|\Delta| \leq S_\Delta$.

Consider Algorithm 1

Also, let $\Delta := N \setminus T$ and $\Delta_\epsilon := T \setminus N$. Let $\hat{x}_{\text{modcs}}$ denote the solution of $\hat{x}_{t,t -,} \text{modcs}$.

$$\|x - \hat{x}_{\text{modcs}}\| \leq C_1(|N| + |\Delta| + |\Delta_\epsilon|)\epsilon, \quad \text{where}$$

$$C_1(S) = \frac{4\sqrt{1 + \delta_S}}{1 - (\sqrt{2} + 1)\delta_S} \quad (11)$$

For the sake of completeness, and for ease of review, we provide a proof in the last appendix, Appendix [F].

Corollary 1 (modified-CS error bound): Let $x$ be a sparse vector with support $N$ and let $y := Ax + w$ with $\|w\| \leq \epsilon$. Also, let $\Delta := N \setminus T$ and $\Delta_\epsilon := T \setminus N$. Let $\hat{x}_{\text{modcs}}$ denote the solution of $\hat{x}_{t,t -,} \text{modcs}$.

If $\delta |N| + |\Delta| + |\Delta_\epsilon| \text{ is just smaller than } \sqrt{2} - 1$, the error bound will be very large because the denominator of $C_1(S)$ will be very large. To keep the bound small, we need to assume that $\delta |N| + |\Delta| + |\Delta_\epsilon| < b(\sqrt{2} - 1)$ with $b < 1$. For simplicity, let $b = 1/2$. Then we get the following corollary, which we will use in our stability results.

Corollary 1 (modified-CS error bound): Let $x$ be a sparse vector with support $N$ and let $y := Ax + w$ with $\|w\| \leq \epsilon$. Also, let $\Delta := N \setminus T$ and $\Delta_\epsilon := T \setminus N$. Let $\hat{x}_{\text{modcs}}$ denote the solution of $\hat{x}_{t,t -,} \text{modcs}$.

If $\delta |N| + |\Delta| + |\Delta_\epsilon| < (\sqrt{2} - 1)/2$ and $|\Delta| \leq |N|/3$,

$$\|x - \hat{x}_{\text{modcs}}\| \leq C_1(|N| + |\Delta| + |\Delta_\epsilon|)\epsilon \leq 8.79\epsilon \quad (12)$$

Proof: Notice that $C_1(S)$ is an increasing function of $\delta_S$. The above corollary follows by using $\delta |N| + |\Delta| + |\Delta_\epsilon| < (\sqrt{2} - 1)/2$ to bound $C_1(S)$ by $C_1((\sqrt{2} - 1)/2) = 8.79$.

We can state a similar version of the result for CS [10].

Corollary 2 (CS error bound): Let $x$ be a sparse vector with support $N$ and let $y := Ax + w$ with $\|w\| \leq \epsilon$. Let $\hat{x}_{\text{cs}}$ denote the solution of $\hat{x}_{t,t -,} \text{cs}$ with $T = \emptyset$. If

$$\delta |N| < (\sqrt{2} - 1)/2,$$

then

$$\|x - \hat{x}_{\text{cs}}\| \leq C_1(2|N|)\epsilon \leq 8.79\epsilon \quad (13)$$

A. Stability result for modified-CS

The first step to show stability is to find sufficient conditions for a certain set of large coefficients to definitely get detected, and for the elements of $\Delta_\epsilon$ to definitely get deleted. This can be obtained using Corollary 1 and the following simple facts which we state as a proposition.

Proposition 1 (simple facts): Consider Algorithm 1

1) An $i \in N$ will definitely get detected in step 1 if $|x_i| > \alpha + \|x - \hat{x}_{\text{modcs}}\|$. This follows since $\|x - \hat{x}_{\text{modcs}}\| \geq \|x - \hat{x}_{\text{modcs}}\|_{\infty} \geq \|x - \hat{x}_{\text{modcs}}\|_{1}$.

2) Similarly, all $i \in \Delta_\epsilon$ (the zero elements of $T$) will definitely get deleted in step 2 if $\alpha \geq \|x - \hat{x}_{\text{modcs}}\|_{\infty}$.

Combining the above facts with Corollary 1 we get the following lemma.

Lemma 2: Let $x$ be a sparse vector with support $N$ and let $y := Ax + w$ with $\|w\| \leq \epsilon$. Also, let $\Delta := N \setminus T$ and $\Delta_\epsilon := T \setminus N$. Assume that $|N| = S_N$, $|\Delta_\epsilon| \leq S_{\Delta_\epsilon}$ and $|\Delta| \leq S_\Delta$. Consider Algorithm 1

\[ \]
1) Let $L := \{i \in N : |x_i| \geq b_1\}$. All elements of $L$ will get detected in step 3 if
   a) $\delta_{S_0 + S_\Delta + S_\Delta} < (\sqrt{2} - 1)/2$ and $S_\Delta \leq S_N/3$, and
   b) $b_1 > \alpha + 8.79\epsilon$.
2) In step 3 there will be no false additions, and all the true removals from the support (the set $\Delta_\epsilon$) get deleted at the current time, if
   a) $\delta_{S_0 + S_\Delta + S_\Delta} < (\sqrt{2} - 1)/2$ and $S_\Delta \leq S_N/3$, and
   b) $\alpha \geq 8.79\epsilon$.

In the above lemma and proposition, for ease of notation, we have removed the subscript $t$ from $x_i$, $N_t$, $T_t$ and $\Delta_t$.

We use the above lemma to obtain the stability result as follows. Let us fix a bound on the maximum allowed magnitude of a missed coefficient. Suppose we want to ensure that only coefficients with magnitude less than $2r$ are part of the final set of misses, $\Delta_t$, at any time, $t$, and that the final set of extras, $\Delta_{x,t}$ is an empty set. In other words, we want to find conditions to ensure that $\Delta_t \subseteq S_t(2)$ and $|\Delta_{x,t}| = 0$.

Using Signal Model 1, $|S_t(2)| = 2S_0$ and thus $\Delta_t \subseteq S_t(2)$ will imply that $|\Delta_t| \leq 2S_0$. This leads to the following result.

The result can be easily generalized to ensure that, for some $d_0 \leq d$, $\Delta_t \subseteq S_t(d_0)$, and thus $|\Delta_t| \leq (2d_0 - 2)S_0$, holds at all times $t$. We show how to do this for the result of the next section in Appendix B, an analogous thing can be done for Theorem 1 as well.

**Theorem 1 (Stability of modified-CS):** Assume Signal Model 1 on $x_t$. Also assume that $y_t$ satisfies (1) with $\|w_t\| \leq \epsilon$. Consider Algorithm 1. If the following hold
1) **(support estimation threshold)** set $\alpha = 8.79\epsilon$
2) **(support size, support change rate)** $S_0$, $S_\alpha$ satisfy $\delta_{S_0 + 3S_\alpha} < (\sqrt{2} - 1)/2$ and $S_\alpha \leq S_0/6$.
3) **(new element increase rate)** $r \geq G$, where
   $$G \triangleq \frac{\alpha + 8.79\epsilon}{2} = 8.79\epsilon \quad (14)$$
4) **(initial time)** at $t = 0$, $n_0$ is large enough to ensure that $\Delta_0 \subseteq S_0(2)$, $|\Delta_0| \leq 2S_\alpha$, $|\Delta_{x,0}| = 0$ and $|T_0| \leq S_0$.
   a) at all $t \geq 0$, $|T_t| \leq S_0$, $|\Delta_{x,t}| = 0$, $\Delta_t \subseteq S_t(2)$ and so $|\Delta_t| \leq 2S_\alpha$.
   b) at all $t \geq 0$, $|T_t| \leq S_0$, $|\Delta_{x,t}| \leq S_\alpha$, and $|\Delta_t| \leq 2S_\alpha$.
   c) at all $t \geq 0$, $\|x_t - \hat{x}_{t,modcs}\| \leq 8.79\epsilon$

**Proof:** The complete proof is given in Appendix B. It follows using induction. We use the induction assumption; the fact that $T_t = T_{t-1} = N_{t-1}$, and the fact that $N_t = N_{t-1} \cup A_{\ell} \setminus R_t$ to bound $|T_t|$, $|\Delta_t|$ and $|\Delta_{x,t}|$. Next, we use these bounds and Lemma 2 to bound $|\Delta_t|$ and $|\Delta_{x,t}|$. Finally $|T_t| \leq |N_t| + |\Delta_{x,t}|$ helps to bound $|T_t|$.

**B. Discussion**

**Remark 3:** We note that condition 3 is not restrictive. It is easy to see that this will hold if the number of measurements at $t = 0$, $n_0$, is large enough to ensure that the measurement matrix at $t = 0$, $A_0$, satisfies $\delta_{2S_0}(A_0) < (\sqrt{2} - 1)/2$ and conditions 1 and 2 hold.

Notice that all the support errors are bounded by $2S_\alpha$ or less. Under slow support change, $S_\alpha \ll S_0$ and so $2S_\alpha$ is also small compared to the support size, $S_0$, making the above result a meaningful stability result.

Let us compare the results for modified-CS and simple CS. Since simple CS is not a recursive approach (each time instant is handled separately), Corollary 2 is also a stability result for it. From Corollary 2 simple CS needs $\delta_{2S_0} < (\sqrt{2} - 1)/2$ to ensure that its error is bounded by $8.79\epsilon$ for all $t$. On the other hand, for $t > 0$, our result from Theorem 1 only needs $S_\alpha \leq S_0/6$ and $\delta_{S_0 + 3S_\alpha} < (\sqrt{2} - 1)/2$ to get the same error bound. Under $S_\alpha \ll S_0$ (slow support change), $S_\alpha \leq S_0/6$ easily holds and $\delta_{S_0 + 3S_\alpha} < (\sqrt{2} - 1)/2$ is clearly weaker than the simple CS condition. Thus, at $t > 0$, for a given measurement matrix $A$, modified-CS error is guaranteed to remain below $8.79\epsilon$ for larger support sizes, $S_0$, than for simple CS. Said another way, for a given $S_0$, modified-CS needs fewer measurements (only enough to satisfy $\delta_{S_0 + 3S_\alpha} < (\sqrt{2} - 1)/2$), than simple CS (which needs enough to satisfy $\delta_{2S_0} < (\sqrt{2} - 1)/2$).

At $t = 0$, the modified-CS algorithm of Algorithm 1 needs the same number of measurements as simple CS. If reliable prior support knowledge were available at $t = 0$, one would need fewer measurements even at $t = 0$.

The above discussion only compares sufficient conditions. We back it up with actual simulation comparisons in Fig. 3(a) and 3(b) where we compare the average reconstruction error when $n$ is just large enough to ensure small (less than 0.5%) error for modified-CS. With this $n$, the CS error is between 20-30%. Here “error” refers to normalized mean squared error (NMSE). The simulation details are given in Sec. VII.

**C. Limitations**

Before going further, let us discuss the limitations of the above result and of modified-CS itself. First, in Proposition 1 and hence everywhere after that, we bound the $\ell_2$ norm of the error by the $\ell_2$ norm. This is often a loose bound and results in a loose lower bound on the required threshold $\alpha$ and consequently a larger than required lower bound on the minimum required rate of coefficient increase/decrease, $r$.

Second, modified-CS uses single step thresholding for estimating the support $N_t$. The threshold, $\alpha$, needs to be large enough to ensure correct deletion of all the removed elements and no false detection of zero elements (condition 1). But this means that the magnitude increase rate, $r$, needs to be even larger to ensure correct detection, and no false deletion, of all but the smallest $2S_\alpha$ nonzero elements (condition 2).

There is another related issue which is not seen in the theoretical analysis because we only bound the $\ell_2$ norm of the error, but is actually more important since it affects the reconstruction itself, not just the sufficient conditions for its stability. This has to do with the fact that $\hat{x}_{t,modcs}$ is a biased estimate of $x_t$. A similar issue for noisy CS, and a possible solution (Gauss-Dantzig selector), was first discussed in [32]. In our context, along $T$, the values of $\hat{x}_{t,modcs}$ will be biased towards zero (because we minimize $\|\hat{x}_{t,modcs}\|_1$), while, along $T$, they may be biased away from zero (since there is no constraint on $\|\hat{x}_{t,modcs}\|_1$). The bias will be larger when the noise is larger. This will create the following problem. The set $T$
contains the set $\Delta_c$, which needs to be deleted. Since the estimates along $\Delta_c$ may be biased away from zero, one will need a higher threshold to delete them. But that would make detection more difficult, especially since the estimates along $\Delta \subseteq T_c$ will be biased towards zero. In the next section, we discuss a partial solution to this and the previous issue.

V. MODIFIED-CS WITH ADD-LS-DEL AND ITS STABILITY

The last two issues mentioned above in Sec. IV-C can be partly addressed by replacing the single support estimation step by a three step Add-LS-Del procedure summarized in Algorithm 2. This idea was first introduced in our older work \cite{14,13} for recursive sparse reconstruction and simultaneously also in \cite{30,31} for greedy sparse algorithms for static sparse reconstruction. It involves a support addition step (that uses a smaller threshold), as in \cite{15}, followed by LS estimation on the new support estimate, $T_{\text{add}}$, as in \cite{16}, and then a deletion step that thresholds the LS estimate, as in \cite{7}. This can be followed by a second LS estimation using the final support estimate, as in \cite{15}, although this last step is not critical. The addition step threshold, $\alpha_{\text{add}}$, needs to be just large enough to ensure that the matrix used for LS estimation, $A_{\text{add}}$, is well-conditioned. If $\alpha_{\text{add}}$ is chosen properly and if $n$ is large enough, the LS estimate on $T_{\text{add}}$ will have smaller error than the modified-CS output. As a result, deletion will be more accurate when done using this estimate. This also means that one can also use a larger deletion threshold, $\alpha_{\text{del}}$, which will ensure quicker deletion of extras. We summarize the algorithm in Algorithm 2. Notice the reduction in error of modified-CS with add-LS-del as compared to modified-CS in Fig. 4.

Algorithm 2 Modified-CS with Add-LS-Del

For $t \geq 0$, do

1) **Simple CS.** If $t = 0$, set $T = \emptyset$ and compute $\hat{x}_{t,\text{modcs}}$ as the solution of \cite{5}.

2) **Modified-CS.** If $t > 0$, set $T = \hat{N}_{t-1}$ and compute $\hat{x}_{t,\text{modcs}}$ as the solution of \cite{9}.

3) **Additions / LS.** Compute $T_{\text{add}}$ and the LS estimate using it:

\[ T_{\text{add}} = T \cup \{i \in T^c : |(\hat{x}_{t,\text{modcs}})_i| > \alpha_{\text{add}}\} \quad (15) \]

\[ (\hat{x}_{t,\text{add}})_{T_{\text{add}}} = A_{T_{\text{add}}}^{-1} y_t, \quad (\hat{x}_{t,\text{add}})_{T_{\text{add}}} = 0 \quad (16) \]

4) **Deletions / LS.** Compute $\hat{T}$ and LS estimate using it:

\[ \hat{T} = T_{\text{add}} \setminus \{i \in T_{\text{add}} : |(\hat{x}_{t,\text{add}})_i| \leq \alpha_{\text{del}}\} \quad (17) \]

\[ (\hat{x}_t)_{\hat{T}} = A_{\hat{T}}^{-1} y_t, \quad (\hat{x}_t)_{\hat{T}} = 0 \quad (18) \]

5) Set $\hat{N}_t = \hat{T}$. Feedback $\hat{N}_t$. Output either $\hat{x}_t$ or $\hat{x}_{t,\text{modcs}}$.

Definition 4 (Define $T_{\text{add},t}, \Delta_{\text{add},t}, \Delta_{c,\text{add},t}$): The set $T_{\text{add},t}$ is the support estimate obtained after the support addition step. It is defined in \cite{15} in Algorithm 2. The set $\Delta_{\text{add},t} := \hat{N}_t \setminus T_{\text{add},t}$ denotes the set of missing elements from $T_{\text{add},t}$ and the set $\Delta_{c,\text{add},t} := T_{\text{add},t} \setminus \hat{N}_t$ denotes the set of extras in it. We remove the subscript $t$ where not needed.

A. Stability result for Modified-CS with Add-LS-Del

The first step to show stability is to find sufficient conditions for (a) a certain set of large coefficients to definitely get detected, and (b) to definitely not get falsely deleted, and (c) for the zero coefficients in $T_{\text{add}}$ to definitely get deleted. These can be obtained using Corollary 1 and the following simple facts which we state as a proposition, in order to easily refer to them later. In the proposition and the three lemmas below, we remove the subscript $t$ for ease of notation.

**Proposition 2 (Simple facts):** Consider Algorithm 2.

1) An $i \in \Delta$ will definitely get detected in step 3 if $|x_i| > \alpha_{\text{add}} + \|x - \hat{x}_{\text{modcs}}\|_\infty$. This follows since $\|x - \hat{x}_{\text{modcs}}\|_\infty \geq \|x - \hat{x}_{\text{modcs}}\|_1$.

2) Similarly, an $i \in T_{\text{add}}$ will definitely not get falsely deleted in step 4 if $|x_i| > \alpha_{\text{del}} + \|x - \hat{x}_{\text{add}}\|_{\text{LS}}$.

3) All $i \in \Delta_{c,\text{add}}$ (the zero elements of $T_{\text{add}}$) will definitely get deleted if $\alpha_{\text{del}} \geq \|x - \hat{x}_{\text{add}}\|_{\text{LS}}$.

4) Consider LS estimation on the known part of support $T$. Assume that $x \equiv \hat{x}_{\text{LS}}T$ and detect element $i$ if $|\|w\|_T| < \theta_{\text{LS}} |(\hat{x}_{\text{LS}})_i|$. These bounds are derived in \cite{14} (equation (15)\).

Combining the above facts with Corollary 1 we can easily get the following three lemmas.

**Lemma 3 (Detection condition):** Let $x$ be a sparse vector with support $N$ and let $y := Ax + w$ with $\|w\| \leq \epsilon$. Also, let $\Delta := N \setminus T$ and $\Delta_{c} := T \setminus N$.

Assume that $|N| = S_N$, $|\Delta| \leq S_{\Delta_c}$, $|\Delta| \leq S_{\Delta}$.

Consider Algorithm 2 For a given $b_1$, let

\[ L := \{i \in \Delta : |x_i| \geq b_1\}. \]

All elements of $L$ will get detected in step 3 if

1) $\delta_{S_N + S_{\Delta_c} + S_{\Delta}} < (\sqrt{2} - 1)/2$ and $S_{\Delta} \leq S_{N}/3$, and
2) $b_1 > \alpha_{\text{add}} + 8.79\epsilon$.

**Proof:** This lemma follows from fact 1 of Proposition 2 and Corollary 1.

**Lemma 4 (No false deletion condition):** Let $x$ be a sparse vector with support $N$ and let $y := Ax + w$ with $\|w\| \leq \epsilon$. Also, let $T_{\text{add}}, \Delta_{\text{add}}, \Delta_{c,\text{add}}$ be as defined in Definition 4.

Assume that $|T_{\text{add}}| \leq S_T$ and $|\Delta_{\text{add}}| \leq S_{\Delta}$.

Consider Algorithm 2 For a given $b_1$, let

\[ L := \{i \in T_{\text{add}} : |x_i| \geq b_1\}. \]

No element of $L$ will get (falsely) deleted in step 4 if

1) $\delta_{S_T} < 1/2$ and
2) $b_1 > \alpha_{\text{del}} + 2\sqrt{2} + 2\theta_{\text{LS}}S_{\Delta} \|x_{\Delta_{c}}\|_\infty$.

**Proof:** This lemma follows directly from fact 2 and fact 4 (applied with $T \equiv T_{\text{add}}$ and $\Delta \equiv \Delta_{\text{add}}$) of Proposition 2.

**Lemma 5 (Deletion condition):** Let $x$ be a sparse vector with support $N$ and let $y := Ax + w$ with $\|w\| \leq \epsilon$. Also, let $T_{\text{add}}, \Delta_{\text{add}}, \Delta_{c,\text{add}}$ be as defined in Definition 4.

Assume that $|T_{\text{add}}| \leq S_T$ and $|\Delta_{\text{add}}| \leq S_{\Delta}$.

Instead of $\delta_{|T|} < 1/2$, one can pick any $b < 1$ and the constants in the bound will change appropriately.
Consider Algorithm $2$. All elements of $\Delta_{r, \text{add}}$ will get deleted in step $4$ if

1. $\delta_{S_T} < 1/2$ and
2. $\alpha_{\text{add}} \geq 2\sqrt{2\epsilon} + 2\sqrt{S_0} \|x_0\|/x_{\Delta_{e, r}}$.

**Proof:** This lemma follows directly from fact $3$ and fact $4$ (applied with $T = T_{\text{add}}$ and $\Delta = \Delta_{\text{add}}$) of Proposition $2$.

Using the above lemmas and the signal model, we can obtain sufficient conditions to ensure that, for some $d_0 \leq d$, at each time $t$, $\Delta_t \subseteq S_t(d_0)$ (so that $|\Delta_t| < (2d_0 - 2)S_0$) and $|\Delta_{e, r}| = 0$, i.e. only elements smaller than $d_0$ may be missed and there are no extras. For notational simplicity, we state the special case below which uses $d_0 = 2$. The general case is given in Appendix $3$ in Corollary $4$. In fact, this is the generalized version of Corollary $3$ which relaxes some assumptions of the result below.

**Theorem 2** *(Stability of modified-CS with add-LS-del)*: Assume Signal Model $1$ on $x_t$. Also assume that $y_t$ satisfies $1$ with $\|w_t\| \leq \epsilon$. Consider Algorithm $2$ If

1) *(addition and deletion thresholds)*

   a) $\alpha_{\text{add}}$ is large enough so that there are at most $S_0$ false additions per unit time,
2) *(support size, support change rate)* $S_0$, $\Delta_0$ satisfy

   a) $\delta_{S_0 + 3S_a} < (\sqrt{2} - 1)/2$ and $S_0 \leq S_0/6, \delta$, and
3) *(new element increase rate)* $\Delta > \max(G_1, G_2)$, where

   $G_1 \triangleq \frac{\alpha_{\text{add}} + 8.79\epsilon}{2} \sqrt{2\epsilon}$

   $G_2 \triangleq \frac{2}{1 - 2\sqrt{S_a} \theta_{S_0 + 2S_a} S_a} \sqrt{2\epsilon}$

   \[ \text{(19)} \]

4) *(initial time)* at $t = 0$, $n_0$ is large enough to ensure that $\Delta_0 \subseteq S_0(2)$, $|\Delta_0| \geq 2S_a$, $|\Delta_{e, 0}| = 0$, and $|T_0| \leq S_0$.

   then,

1) at all $t \geq 0$, $|\tilde{T}_t| \leq S_0$, $|\tilde{\Delta}_t| = 0$, $\Delta_t \subseteq S_t(2)$ and so $|\Delta_t| \leq 2S_0$,
2) at all $t > 0$, $|\tilde{T}_t| \leq S_0$, $|\tilde{\Delta}_t| \leq S_0$, and $|\Delta_t| \leq 2S_2$,
3) at all $t > 0$, $|\tilde{T}_{\Delta_{\text{add}}, t}| \leq S_0 + 2S_a$, $|\tilde{\Delta}_{\text{add}, t}| \leq 2S_a$, and
4) at all $t > 0$, $|x_t - \tilde{x}_t| \leq \sqrt{2\epsilon} + (2\theta_{S_0 + 2S_a} + 1)\sqrt{2S_0} \epsilon$
5) at all $t > 0$, $|x_t - \tilde{x}_t, \text{mod } T_t| \leq 8.79\epsilon$

**Proof:** The complete proof is given in Appendix $4$. This proof also follows by induction, but is more complicated than that of Theorem $1$. The induction step consists of three parts.

- **First**, we use the induction assumption; the fact that $T_t = T_{t-1} = N_{t-1}$; and the fact that $N_t = N_{t-1} \setminus A_t \setminus R_t$ to bound $|T_t|$, $|\Delta_{e, t}|$, $|\Delta_t|$. This part of the proof is the same as that of Theorem $1$. The next two parts are quite different.

- **We use the bounds from the first part**; equation $6$; Lemma $3$ the limit on the number of false detections from condition $1(a)$ and $|T_{\text{add}}| \leq |N| + |\Delta_{e, \text{add}}|$ to bound $|\Delta_{\text{add}, t}|, |\Delta_{e, \text{add}, t}|, |T_{\text{add}, t}|$.

- **Finally**, we use the bounds from the second part; Lemmas $4$ and $5$ and $|\tilde{T}| \leq |N| + |\Delta_e|$ to bound $|\Delta_t|, |\Delta_{e, t}|, |\tilde{T}_t|$.

**B. Discussion**

Notice that condition $2(b)$ may become difficult to satisfy as soon as $S_a$ increases, which will happen when the problem dimension, $m$, increases, and consequently $S_0$ increases, even though $S_a$ and $S_0$ remain small fractions of $m$, e.g. typically $S_0 \approx 10\% m$ and $S_a \approx 2\% - 10\% S_0 \approx 0.2\% - 1\% m$. The reason we get this condition is because in facts $2$ and $3$ of Proposition $2$ and hence also in Lemmas $4$ and $5$ and in the final result, we bound the $\ell_2$ norm of the LS step error, $(x - \hat{x}_{\text{add}})T_{\text{add}}$, by its $\ell_2$ norm. This is clearly a loose bound. It holds with equality only when the entire LS step error is concentrated in one dimension.

In practice, as observed in our simulations, the LS step error is actually quite spread out, since the LS step tends to reduce the bias in the estimate, at least as long as the number of misses in $T_{\text{add}}$ is small and $A_{T_{\text{add}}}$ is well conditioned (which are required conditions for stability anyway and are enforced by conditions $5$ and $6$ of Theorem $2$). Thus, it is not unreasonable to assume that $\|(x - \hat{x}_{\text{add}})T_{\text{add}}\|_2 \leq C\|(x - \hat{x}_{\text{add}})T_{\text{add}}\|$ for some $C < 1$. From simulations, it is observed that $C = \frac{\zeta_m}{\sqrt{S_0}}$ works. Here $\zeta_m$ is slightly more than one and increases very slowly with $m$, e.g. for $m = 200$, $\zeta_m = 1.11$, for $m = 1000$, $\zeta_m = 1.23$ and for $m = 2000$, $\zeta_m = 1.38$. The above numbers were obtained when we simulated according to the generative model for Signal Model $1$ given in Appendix $A$. We used $S_0 = 0.1m, S_a = 0.01m$, and $r = 1$; the matrix $A$ was random Gaussian, with $n = \frac{0.3861S_0 \log_2 m}{\epsilon}$; the noise, $w_t$, was independent identically distributed (i.i.d.) uniform $(-e, e)$ in various dimensions and over time and we used $e = 0.1266$; and we set $\alpha_{\text{add}} = e/2$ and $\alpha_{\text{del}} = r/2$. Similar conclusions were obtained for $r = 3/4$ and $2/3$.

With using the extra assumption $\|(x - \hat{x}_{\text{add}})T_{\text{add}}\|_2 \leq \frac{\zeta_m}{\sqrt{S_0}}\|(x - \hat{x}_{\text{add}})T_{\text{add}}\|$ in facts $2$ and $3$ of Proposition $2$, Lemmas $4$ and $5$ get replaced by the following two lemmas. With using these new lemmas, condition $2(b)$ of Theorem $2$ will get replaced by $\theta_{S_0 + 2S_a} S_a < 1/m$ which is an easily satisfiable condition. Moreover, this also makes the lower bound on the required value of $r$ (rate of coefficient increase/decrease) smaller.

**Lemma 6** *(No false deletion condition – weaker)*: Let $x$ be a sparse vector with support $N$ and let $y := Ax + w$, with $\|w\| \leq \epsilon$. Also, let $T_{\text{add}}, \Delta_{\text{add}}, \Delta_{e, \text{add}}$ be as defined in Definition $1$.

Assume that $|T_{\text{add}}| \leq S_T$ and $|\Delta_{\text{add}}| \leq S_{\Delta}$.

Consider Algorithm $2$. Assume that the LS step error is spread out enough to ensure that

$\|(x - \hat{x}_{\text{add}})T_{\text{add}}\|_2 \leq \frac{\zeta_m}{\sqrt{S_0}}\|(x - \hat{x}_{\text{add}})T_{\text{add}}\|_2$.

For a given $b_1$, let

$L := \{i \in T_{\text{add}} : |x_i| \geq b_1\}$.

No elements of $L$ will get (falsely) deleted in step $4$ if the matrix $A$ was chosen once and fixed. We sampled over the distributions of $w_t$ and $x_t$.\footnote{We computed $\zeta_m$ by computing the maximum of $\|(x - \hat{x}_{\text{add}})T_{\text{add}}\|_2$ over time and over $500$ independent simulations for $m = 200$ (and over $50$ for $m = 1000, 2000$). The matrix $A$ was chosen once and fixed.}
1) \( \delta_{S_{T}} < 1/2, \) and
2) \( b_1 > \alpha_{del} + \frac{\zeta_m}{\sqrt{s_a}} (\sqrt{2} + 2 \theta_{S_T} \Delta_{S} \|x_{S_{T}}\|_2) \).

**Lemma 7 (Deletion condition weaker):** Let \( x \) be a sparse vector with support \( N \) and let \( y := Ax + w, \) with \( \|w\| \leq \epsilon. \)
Also, let \( T_{add}, \Delta_{S_{add}}, \Delta_{add} \) be as defined in Definition 4.
Assume that \( |T_{add}| \leq S_{T} \) and \( |\Delta_{add}| \leq S_{S} \).

Consider Algorithm 2. Assume that the LS step error is spread out enough to ensure that
\[
\| (x - \hat{x}_{add})_{S_{add}} \| \leq \frac{\zeta_m}{\sqrt{s_a}} \| (x - x_{add})_{S_{add}} \|.
\]
All elements of \( \Delta_{S_{add}} \) will get deleted in step 4 if
1) \( \delta_{S_{T}} < 1/2 \) and
2) \( \alpha_{add} \geq \frac{\sqrt{2} \zeta_m}{\sqrt{s_a}} (\sqrt{2} + 2 \theta_{S_T} \Delta_{S} \|x_{S_{T}}\|_2) \).

By using Lemmas 6 and 7 instead of Lemmas 4 and 5 respectively, and doing everything else exactly as in the proof of Theorem 2 we get the following corollary.

**Corollary 3 (Stability of modified-CS with add-LS-del - 2):** Assume Signal Model 1 on \( x_t \). Also assume that \( y_t \) satisfies 1 with \( \|w_k\| \leq \epsilon. \) Let
\[ e_t := (x_t - \hat{x}_{add,t})_{T_{add}}, \]
denote the LS step error. Assume that the LS step error is spread out enough so that
\[ \| e_t \|_{\infty} \leq \frac{\zeta_m}{\sqrt{s_a}} \| e_t \| \]  
(20)
at all times, \( t. \) Consider Algorithm 2. If
1) (addition and deletion thresholds)
   a) \( \alpha_{add} \) is large enough so that there are at most \( S_a \) false additions per unit time,
   b) \( \alpha_{add} = \sqrt{2} \zeta_m \epsilon + 2 \theta_{S_0 + 2S_a} \zeta_m r \).
2) (support size, support change rate) \( S_0, S_a \) satisfy
   a) \( \delta_{S_0 + 3S_a} < (\sqrt{2} - 1)/2 \) and \( S_a \leq S_0/6, \) and
   b) \( \theta_{S_0 + 2S_a} \leq \frac{1}{\sqrt{2}}. \)
3) (new element increase rate) \( r \geq \max (G_1, G_2) \), where
   \[ G_1 \triangleq \frac{\alpha_{add} + 8.79 \epsilon}{2} \]
   \[ G_2 \triangleq \frac{\sqrt{2} \zeta_m \epsilon}{\sqrt{S_a} (1 - 2 \theta_{S_0 + 2S_a} S_a \zeta_m)} \]  
(21)
4) (initial time) at \( t = 0, n_0 \) is large enough to ensure that
   \( \Delta \subseteq S_0(2), |\Delta| \leq 2S_a, |\Delta_\sigma| = 0, |T| \leq S_0, \)
then all conclusions of Theorem 2 hold.

A generalization of the above corollary, that allows the support error to stabilize at \((2d_0 - 2)S_a, \) for some \( d_0 \leq d, \) is given in Appendix D in Corollary 4.

Recall that \( \zeta_m/\sqrt{s_a} \) is smaller than one. For example, in our simulations when \( m = 2000, \zeta_m = 1.38, \) while \( \sqrt{s_a} = \sqrt{20} = 4.47. \) Also, \( \zeta_m \) increases very slowly with \( m \) (slower than \( O(\log m) \)) where as \( \sqrt{s_a} \) typically increases as \( \sqrt{m}. \) Thus, conditions 11, 22, and 4 are significantly weaker compared to those in Theorem 3 while others are the same.

In particular, now condition 22 is easy to satisfy.

Let us compare this result with that for modified-CS given in Theorem 1. Consider the lower bound on \( r \) required by both results. In the above result, since \( \theta_{S_0 + 2S_a} S_a < 1/(4\zeta_m), \) so \( G_2 < \frac{\sqrt{2} \zeta_m \epsilon}{\sqrt{S_a}} \in \left( \frac{2.9}{\sqrt{S_a}} \right) \) is smaller than \( G_1 \) and thus \( G_1 \) is what decides the minimum allowed value of \( r. \) Because of add-LS-del, the addition threshold, \( \alpha_{add} \) can now be much smaller, as long as the number of false adds is small. If \( \alpha_{add} \) is close to zero, the value of \( G_1 \) is almost half of that of \( G \) in Theorem 1.

Theorem 2 allows for the minimum coefficient increase rate, \( r, \) required by the above result is almost half of that required by Theorem 1.

On the other hand, the above result also requires condition 23 on \( \theta \) which Theorem 1 does not, but this condition is typically weaker than condition 23 as \( \theta_{S_0 + 2S_a} S_a \) is smaller than \( \theta_{S_0 + 3S_a} \) where as the right hand sides do not differ by much.

The above is also demonstrated in Figs. 3(b) and 3(c) for \( r = 1, \) both are stable, but for \( r = 2/3, \) modified-CS is unstable while modified-CS with add-LS-del is still stable.

Finally, let us compare our result with the simple CS result given in Corollary 2. Corollary 2 needs \( \delta_{S_0} < (\sqrt{2} - 1)/2 = 0.207 \) to achieve the same error bound as our result. On the other hand, if the LS step error is spread out enough, we only need \( \delta_{S_0 + 3S_a} < (\sqrt{2} - 1)/2 = 0.207 \) and \( \theta_{S_0 + 2S_a} S_a < 1/(4\zeta_m). \) When \( S_a \ll S_0 \) (slow support change), the first condition is clearly weaker than what CS needs. The second condition is also weaker since \( \theta_{S_0 + 2S_a} S_a \) is significantly smaller than \( \delta_{S_0} \) where as the right hand sides 0.207 and 0.25/\( \zeta_m \) are roughly equal. A quantitative comparison can be done by using the upper bounds \( \theta_{u,k} \leq \delta_{u,k} + \delta \) and \( \delta_{u,k} \leq \alpha_{add} \) \[3(1)\]. If \( S_a = 0.02S_0, \) then \( \delta_{S_0} = \delta_{100S_0} \leq 100\delta_{2S_0} \) and \( \theta_{S_0 + 2S_a} S_a \leq \delta_{S_0 + 3S_a} < 53\delta_{S_0}. \) Thus, the CS condition is stronger as long as \( \zeta_m < (100/53)(0.25/0.207) = 2.28. \) If \( S_a = 0.1S_0, \) then the CS condition is stronger if \( \zeta_m < 1.9. \)

**Remark 4:** In the discussion so far we have used the special case stability results where we find conditions to ensure that the misses remain below \( 2S_a. \) Let us look at the general form of the result – Corollary 4 in Appendix D – where we provide conditions to ensure that, for some \( d_0 \leq d, \) the misses are below \((2d_0 - 2)S_a. \) In Corollary 4 using an argument similar to the one above, \( \tilde{G}_2 < \tilde{G}_1 \) holds for any \( d_0. \) Also, notice that, if the rate of coefficient increase, \( r, \) is smaller, \( r \geq \tilde{G}_1 \) will hold for a larger value of \( d_0. \) This means that the support error bound, \((2d_0 - 2)S_a, \) will be larger. This, in turn, decides what conditions on \( \delta \) and \( \theta \) are needed (in other words, how many measurements, \( n, \) are needed). Smaller \( r \) means a larger \( d_0 \) is needed which, in turn, means that stronger conditions on \( \delta, \theta \) (larger \( n \)) are needed. Thus, for a given \( n, \) as \( r \) is reduced, the algorithm will stabilize to larger and larger support error levels (larger \( d_0 \)) and finally become unstable (because the given \( n \) does not satisfy the conditions on \( \delta, \theta \) for the larger \( d_0 \)).

The above is demonstrated empirically in Fig. 3. The last three rows of this figure used \( n = 59. \) When \( r = 1, \) modified-CS with add-LS-del is stable at zero support errors. When \( r \) is reduced to 2/3, it is stable at mean support errors less than 0.3%. When \( r \) is reduced to 2/5 it becomes unstable.
VI. Stability of LS-CS

In [14], [28], we introduced Least Squares CS-residual (LS-CS) as one of the first solutions to the problem of recursively reconstructing sparse signal sequences with slow time-varying sparsity patterns. We summarize the complete LS-CS algorithm in Algorithm 3. LS-CS uses partial knowledge of support, \( T \), in a different way than modified-CS. It first computes an initial LS estimate on the set \( T \), as in [22], and then computes the observation residual, as in (23). Noisy CS is done on this observation residual, as in (24), and the solution is added back to the initial LS estimate, as in (25). The add-LS-del approach described earlier is used for support estimation.

Algorithm 3 Least Squares CS-residual (LS-CS)

For \( t \geq 0 \), do
1) Simple CS. Do as in Algorithm 2
2) CS-residual.
   a) Use \( T := \hat{T}_{t-1} \) to compute the initial LS estimate, \( \hat{x}_{t,\text{init}} \), and the LS residual, \( \hat{y}_{t,\text{res}} \), as follows.
      \[
      (\hat{x}_{t,\text{init}})T = A_T^T \hat{y}_t, \quad (\hat{x}_{t,\text{init}})_{T^c} = 0
      \]
      \[
      \hat{y}_{t,\text{res}} = y_t - A \hat{x}_{t,\text{init}}
      \]
   b) Do noisy CS on the LS residual, i.e. solve
      \[
      \min_\beta \| \beta \|_1 \ s.t. \ \| \hat{y}_{t,\text{res}} - A \beta \| \leq \epsilon
      \]
      and denote its output by \( \hat{\beta}_t \). Compute
      \[
      \hat{x}_{t,\text{CSres}} := \hat{\beta}_t + \hat{x}_{t,\text{init}}.
      \]
3) Additions / LS. Compute \( T_{\text{add}} \) and the LS estimate on it as in Algorithm 2. Use \( \hat{x}_{t,\text{CSres}} \) instead of \( \hat{x}_{t,\text{modcs}} \) for estimating \( T_{\text{add}} \).
4) Deletions / LS. Compute \( \hat{T} \) and the LS estimate on it as in Algorithm 2.
5) Set \( \hat{N}_t = \hat{T} \). Output \( \hat{x}_t \). Feedback \( \hat{N}_t \).

The CS-residual step error, \( x_t - \hat{x}_{t,\text{CSres}} \), where \( \hat{x}_{t,\text{CSres}} \) is defined in (25), can be bounded as follows. The proof is easy and follows in the same way as that for [14] Corollary 1 where noisy CS is done using Dantzig selector instead of (23). We use (24) here to keep the comparison with modified-CS easier.

Lemma 8 (CS-residual error bound [14]): Let \( x \) be a sparse vector with support \( N \) and let \( y := Ax + w \) with \( \| w \| \leq \epsilon \). Also, let \( \Delta := N \setminus T \) and \( \Delta_a := T \setminus N \). Consider step 2 of Algorithm 3. If
\[
\delta_{2|\Delta|} < (\sqrt{2} - 1)/2
\]
\[
\delta_{|T|} < 1/2,
\]
then
\[
\| x - \hat{x}_{\text{CSres}} \| \leq C' \| y \|_2 + \theta_{|T|,|\Delta|} C'' \| x_{\Delta} \|,
\]
where
\[
C' \equiv C'(|T|, |\Delta|) \leq C_1 \frac{1}{1 - \sqrt{2}} \delta_S^2
\]
\[
C'' \equiv C''(|T|, |\Delta|) \leq C_2 \frac{1}{1 - \sqrt{2}} \delta_S^2
\]
(26)

A. Stability result for LS-CS

Our overall approach is similar to the one discussed in the previous section for modified-CS with add-LS-del. The key difference is in the detection condition lemma, which we give below. Its proof is given in Appendix B. This lemma is different from Lemma 3 because, unlike modified-CS, the CS-residual error bound at time \( t \) also depends on the magnitudes of the elements in the initial missed set \( \Delta \).

Lemma 9 (Detection condition for LS-CS): Let \( x \) be a sparse vector with support \( N \) and let \( y := Ax + w \) with \( \| w \| \leq \epsilon \). Also, let \( \Delta := N \setminus T \) and \( \Delta_a := T \setminus N \). Assume that \( |T| \leq S_T \) and \( |\Delta| \leq S_\Delta \). Assume that \( \| x_{\Delta} \|_\infty \leq b \).

Consider step 3 of Algorithm 3. For a \( \gamma \leq 1 \), let
\[
L_1 := \{ i \in \Delta : \gamma b \leq |x_i| \leq b \}
\]
and let
\[
L_2 := \Delta \setminus L_1 = \{ i \in \Delta : |x_i| < \gamma b \}.
\]
Assume that \( |L_1| \leq S_{L_1} \) and \( \| x_{L_2} \| \leq \kappa b \). All \( i \in L_1 \) will definitely get detected at the current time if
\[
\delta_{2S_\Delta} < (\sqrt{2} - 1)/2,
\]
\[
\delta_{S_T} < 1/2,
\]
\[
\max_{|\Delta| \leq S_\Delta} \theta_{|T|,|\Delta|} C''(S_T, |\Delta|) \leq \frac{2}{\sqrt{2}} \delta_S^2 \quad \text{and}
\]
\[
\max_{|\Delta| \leq S_\Delta} \frac{\alpha_{\text{add}} + C'(S_T, |\Delta|) \epsilon}{\gamma - \theta_{S_T,|\Delta|} C''(S_T, |\Delta|) (\sqrt{S_{L_1} + \kappa})} < b
\]
where \( C'(\cdot, \cdot), C''(\cdot, \cdot) \) are defined in Lemma 8.

Proof: The proof is given in Appendix B.

The stability result then follows in the same fashion as Theorem 2. The only difference is that instead of Lemma 3 we apply Lemma 2 with \( S_T = S_0, S_\Delta = 2S_a, b = 2r, \gamma = 1 \), \( S_{L_1} = S_0 \) and \( \kappa = \frac{2}{\sqrt{2}} \delta_S^2 \).

Theorem 3 (Stability of LS-CS): Assume Signal Model 1 on \( x_t \). Also assume that \( y_t \) satisfies (1) with \( \| w_t \| \leq \epsilon \). Consider Algorithm 3.

1) (addition and deletion thresholds)
   a) \( \alpha_{\text{add}} \) is large enough so that there are at most \( S_a \) false additions per unit time,
   b) \( \alpha_{\text{add}} = \sqrt{2} \epsilon + 2\sqrt{S_a} \theta_{S_0 + 2S_a, S_a, r} \)

2) (support size, support change rate) \( S_0, S_a \) satisfy
   a) \( \delta_{S_a} < (\sqrt{2} - 1)/2 \)
   b) \( \delta_{S_0 + 2S_a} < 1/2 \)
   c) \( \max_{|\Delta| \leq 2S_a} \theta_{S_0,|\Delta|} C''(S_0, |\Delta|) \leq \frac{1}{\sqrt{2} S_a} \)
get relaxed to something like

\[ \max \{ \tilde{G}_1, \tilde{G}_2 \} \]

where

\[ \tilde{G}_1 \triangleq \max_{|\Delta| \leq 2S_a} \left[ \frac{\alpha_{\text{add}} + C'(S_0, |\Delta|)\epsilon}{2 - 3\theta_{S_0,|\Delta|}\sqrt{S_aC''(S_0, |\Delta|)}} \right] \]

\[ \tilde{G}_2 \triangleq \frac{\sqrt{2\epsilon}}{1 - 2\sqrt{S_a}\theta_{S_0+2S_a, S_a}} \]

(27)

4) \textit{(initialization)} \ (\text{same condition as in Theorem 2})

then, all conclusions of Theorem 2 hold for LS-CS, except the last one. This is replaced by \( \|x_t - \tilde{x}_t, \text{CSres} \| \leq \max_{|\Delta| \leq 2S_a} C'(S_0, |\Delta|) + (\theta_{S_0,|\Delta|}C''(S_0, |\Delta|) + 1)\sqrt{2S_aT} \).

B. Discussion

Notice that conditions 2c and 2d are the difficult conditions to satisfy as the problem size, \( m \), increases and consequently \( S_0 \) and \( S_a \) increase. We get condition 2d because we bound the \( \ell_2 \) norm of the addition step error by its \( \ell_2 \) norm. This can be relaxed to \( \theta_{S_0+2S_a, S_a} < 1/(4\epsilon_m) \) in the same fashion as in the previous section.

Consider condition 2a. We get this condition because (i) we upper bound the \( \ell_2 \) norm of the CS-residual step error, \( x_t - \tilde{x}_t, \text{CSres} \), by its \( \ell_2 \) norm in Lemma 2 and (ii) in the proof of Lemma 2 we upper bound the \( \ell_1 \) norm of the initial LS step error, \( x_t - \tilde{x}_t, \text{init} \), by \( \sqrt{T} \) times its \( \ell_2 \) norm (this results in the expression for \( C'' \) given in Lemma 8). If we can relax (i), we can try to weaken the required condition, but it will still be stronger than what modified-CS with add-LS-del or modified-CS need. For example, if we can assume a bound similar to (20) for the CS-residual step error, and if additionally, we assume that, in the range \( |\Delta| \leq 2S_a \), \( \theta_{S_0,|\Delta|}C''(S_0, |\Delta|) \) is largest for \( |\Delta| = 2S_a \), condition 2c will get relaxed to something like \( \theta_{S_0,2S_a}C_2(2S_a) \leq \frac{1}{\alpha_m}\sqrt{2S_a} \). This is still stronger than condition 2a of Corollary 3 primarily because of \( \sqrt{S_a/S_0} \).

The above is also observed in our simulations. In Fig. 3 LS-CS needs a larger \( n \) (\( n = 65 \)) for stability where as for modified-CS with add-LS-del or modified-CS, \( n = 59 \) suffices. We show the results for \( r = 1 \) or lower, but even when we increased \( r \) to \( r = 2 \) or \( r = 3 \), LS-CS was still unstable with \( n = 59 \). The simulation details are given in Sec. VII.

VII. Simulation Results

We compared modified-CS (mod-CS), as given in Algorithm 1 modified-CS with Add-LS-Del (mod-CS-add-LS-del), as given in Algorithm 2, and simple CS for a few different choices of \( n \) and \( r \). The results are shown in Fig. 4 where we show four rows of plots. In each row, we plot the normalized mean squared error (NMSE), \( \frac{\mathbb{E}[N]\|\hat{x}_t - x_t\|^2}{\|x_t\|^2} \), the normalized mean extras, \( \frac{\mathbb{E}[N\{\|\hat{x}_t - x_t\| - |x_t|\}]}{\mathbb{E}[N]} \), and the normalized mean misses, \( \frac{\mathbb{E}[N\{\|\hat{x}_t - x_t\| - |x_t|\}]}{\mathbb{E}[N]} \), in the left, middle, and right columns respectively. Here \( \mathbb{E}[\cdot] \) denotes the empirical mean over the 500 realizations.

In all rows, we used the generative model for Signal Model 1 from Appendix A with \( m = 200 \), \( S_0 = 20 \), \( S_a = 2 \). The measurement noise, \( (w_t)_t \sim \text{i.i.d. uniform}(-c, c) \) with \( c = 0.1266 \), i.e. it was i.i.d. uniform in all dimensions and over time. Each element of the measurement matrix, \( A \), was i.i.d. zero mean random Gaussian. Fig. 3(a) used \( n = 65 \), \( r = 1 \) and \( d = 3 \), while the other three rows used \( n = 59 \). Fig. 3(b) used \( n = 59 \), \( r = 2/3 \), \( d = 3 \); and Fig. 3(d) used \( n = 59 \), \( r = 2/5 \), \( d = 5 \). Our simulations selected \( A \) once and kept it fixed, but Monte Carlo averaged over \( \omega_t \) and \( x_t \).

We set the addition threshold, \( \alpha_{\text{add}} \), to be at the noise level \( -\) we set it to \( c/2 \). Assuming that the LS step after support addition gives a fairly accurate estimate of the nonzero values, one can set the deletion threshold, \( \alpha_{\text{del}} \), to a larger value of \( \alpha_{\text{del}} = r/2 \) and still ensure that there are no (or very few) false deletions. Larger deletion threshold ensures that all (or most) of the false additions and removals get deleted. Modified-CS used a single threshold, \( \alpha \), somewhere in between \( \alpha_{\text{add}} \) and \( \alpha_{\text{del}} \). We set \( \alpha = ((c/2) + (r/2))/2 \) (we picked this after trying a few different options for \( \alpha \)). Also, we did not do anything at \( t = 0 \). We just started our simulation at \( t = 1 \) with the assumption that \( \bar{A}_0 = 2 \), \( \bar{A}_{r,0} = 0 \) and hence \( \bar{T}_0 = S_0 - 2 \) (i.e. the initial time condition of all our theorems holds).

Notice, from the plots, that LS-CS needs at least \( n = 65 \) for stability (compare Fig. 3(a) with Fig. 3(b)) where as mod-CS-add-LS-del and mod-CS are stable even with \( n = 59 \). We also tried using \( n = 59 \) and larger values of \( r \), but even with \( r = 3 \) LS-CS was still unstable in a few cases.

Secondly, even with \( n = 65 \), simple CS NMSE is about 20% where as mod-CS-add-LS-del and mod-CS are stable at 0.1% and mod-CS is stable at 0.3%. We do not show support recovery errors for simple CS since they were very large. With \( n = 59 \), simple CS NMSE goes up to 30%. We also show the NMSE plot for simple Gauss-CS (CS followed by a final LS step on the estimated support, done in a fashion similar to Gauss-Dantzig selector [52]). Since the CS error itself is so large, this debiasing step does not help.

When \( n = 65 \) and \( r = 1 \), mod-CS is stable, but has larger error than both LS-CS and mod-CS-add-LS-del. When \( n = 59 \) and \( r = 1 \), LS-CS becomes unstable. But, mod-CS and mod-CS-add-LS-del are still stable, with mod-CS being stable at a larger error (both larger support error and MSE) than mod-CS-add-LS-del. When \( r \) is reduced to 2/3, mod-CS also becomes unstable. But mod-CS-add-LS-del is still stable, though at higher error values than when \( r = 1 \). When \( r \) is further reduced to 2/5, even mod-CS-add-LS-del becomes unstable.

Mod-cs-add-LS-del uses a better support estimation method and thus its extras and misses are both much smaller than those of mod-CS. As a result, (a) it can remain stable for smaller values of \( r \) than mod-CS; and (b) when both are stable, its reconstruction error is smaller than that of mod-CS.

VIII. Conclusions

Under mild assumptions, we showed the “stability” of modified-CS and its improved version, modified-CS with add-LS-del, and of LS-CS for recursive sparse signal sequence reconstruction. By “stability” we mean that the number of misses from the current support estimate and the number of extras in it remain bounded by a time-invariant value at all
times. Under slow support change, the results are meaningful, i.e. the bound is small compared to the support size. A direct corollary is that the reconstruction errors are also bounded by time-invariant and small values.

We can argue that our results ensure stability under weaker assumptions that those required by simple CS. We are able to compare the implications of the results for the three recursive algorithms and argue that modified-CS with add-time-invariant and small values. Under slow support change, the results are meaningful, on conditions on both $n$ and $r$, with equal probability when the element is how to show stability for a stochastic model of signal change, $\delta$-approximate modified-CS \cite{16} or of \cite{17}. That also use slow signal value change, e.g. KF-CS \cite{13}, \cite{14}.

A. Generative Models for Signal Model

To help understand Signal Model better, we provide here two possible generative models that satisfy its assumptions. In both cases, at $t = 0$, the support size is $S_t$ and it contains $2S_a$ elements each with magnitude $r$, $2r \ldots (d − 1)r$, and $(S_0 − (2d − 2)S_a)$ elements with magnitude $M$.

1) Generative Model 1: This assumes that when a new element gets added to the support, its magnitude keeps increasing at rate $r$ until it reaches $M := dr$. An analogous model is assumed for decrease until removal from support. The sign is selected as $+1$ or $−1$ with equal probability when the element gets added to the support, but remains the same after that.

Mathematically this can be described as follows. Let $(x_t)_i = (m_t)_i(s_t)_i$ where $(m_t)_i$ denotes the magnitude and $(s_t)_i$ denotes the sign of $(x_t)_i$ at time $t$.

At any $t > 0$, do the following.

1) Update

$$I_t(j) = I_{t−1}(j − 1), \text{ for all } 2 \leq j \leq d, \text{ and}$$

$$D_t(j) = D_{t−1}(j + 1), \text{ for all } 0 \leq j \leq d − 2 \quad (28)$$

where $I_t(j)$ and $D_t(j)$ are defined in Definition

Recall that the removed set, $R_t = D_t(0)$.

2) Generate

a) the new addition set, $A_t = I_t(1)$, of size $S_a$ uniformly at random from $N_{t−1}$, and

b) the new decreasing set, $D_t(d − 1)$, of size $S_a$ uniformly at random from $\{i \in N_{t−1} : (x_{t−1})_i = M\}$.

3) Update the coefficients’ magnitudes as follows.

$$m_t = \begin{cases} (m_{t−1})_i + r, & i \in U_{j=1}^d I_t(j) \\ (m_{t−1})_i − r, & i \in U_{j=1}^d D_t(j) \\ (m_{t−1})_i, & i \notin C_t \end{cases} \quad (29)$$

where $C_t := N_{t} \cap \{U_{j=1}^d I_t(j) \cup U_{j=1}^d D_t(j)\}$.

4) Update the signs as follows.

$$s_t = \begin{cases} (s_{t−1})_i, & i \in N_t \setminus A_t \\ i \in A_t \\ 0, & i \in N^c_t \end{cases} \quad (30)$$

where $iid(±1)$ refers to generating the sign as $+1$ or $−1$ with equal probability and doing this independently for each element $i$.

5) Set $(x_t)_i = (m_t)_i(s_t)_i$ for all $i$.

Our simulations used the above model.

2) Generative Model 2: A second reasonable generative model selects any $S_a$ out of the $2S_a$ elements with current magnitude $jr$ and increase them, and decreases the other $S_a$ elements. In other words, it replaces the first step above by the following, while keeping the rest of the steps the same.

1) Generate

a) $I_t(j)$ of size $S_a$ uniformly at random from $\{i \in N_{t−1} : (x_{t−1})_i = (j−1)r\}$ for all $2 \leq j \leq d$.

b) $D_t(j)$ of size $S_a$ uniformly at random from $\{i \in N_{t−1} : (x_{t−1})_i = (j+1)r\}$ for all $0 \leq j \leq d − 2$.

B. Appendix: Proof of Theorem

We prove the first claim by induction. Using condition 2 of the theorem, the claim holds for $t = 0$. This proves the base case. For the induction step, assume that the claim holds at $t − 1$, i.e. $|\Delta_{c,e,t−1}| = 0$, $|T_{e,t−1}| \leq S_0$, and $\Delta_{c,e,t−1} \subseteq S_{t−1}(2)$ so that $|\Delta_{c,e,t−1}| \leq 2S_a$. Using this we prove that the claim holds at $t$. In the proof, we use the following facts often: (a) $R_t \subseteq N_{t−1}$ and $A_t \subseteq N^c_{t−1}$, (b) $N_t = N_{t−1} \cup A_t \setminus R_t$, and (c) if two sets $B, C$ are disjoint, then $D \cup C \setminus B := (D \cup C) \setminus B = (D \cap B^c) \cup C$ for any set $D$.

We first bound $|T_{e,t}|, |\Delta_{c,e,t}|, |\Delta_t|$. Since $T_{e,t} = \tilde{T}_{e,t−1} = \tilde{N}_{t−1}$, so $|T_{e,t}| \leq S_0$. Also, $\Delta_{c,e,t−1} = \tilde{N}_{t−1} \setminus N_{t−1} \supseteq (N_{t−1} \cap A_t^c) \cup R_t \subseteq \Delta_{c,e,t−1} \cup R_t = R_t$. The last equality follows since $|\Delta_{c,e,t−1}| = 0$. Thus $|\Delta_{c,e,t}| \leq |R_t| = S_a$.

Consider $|\Delta_t|$. Notice that $\Delta_t = N_t \setminus \tilde{N}_{t−1} = (N_{t−1} \cap \tilde{N}_{t−1} \cap R_t^c) \cup (A_t \cap \tilde{N}_{t−1}^c) = (\Delta_{c,e,t−1} \cap R_t^c) \cup (A_t \cap \tilde{N}_{t−1}^c) \subseteq (S_{t−1} \cap R_t^c) \cup (A_t \cap \tilde{N}_{t−1}^c) \subseteq S_{t−1}(2) \cap R_t$. Here we used $\Delta_{c,e,t−1} \subseteq S_{t−1}(2)$. Since $R_t \subseteq S_{t−1}(2)$ and $A_t$ is disjoint with $S_{t−1}(2)$, thus $|\Delta_t| \leq |S_{t−1}(2)| + |A_t| − |R_t| = 2S_a + S_0 − S_a$.

Next we bound $|\Delta_{c,t}|, |\Delta_{c,e,t}|, |T_{e,t}|$. Consider the support estimation step. Apply the first claim of Lemma \cite{2} with $S_N = S_0$, $S_{\Delta e} = S_a$, $S_{\Delta} = 2S_a$, and $b_1 = 2r$. Since conditions 2 and 3 of the theorem hold, all elements of $N_t$ with magnitude equal to or greater than $2r$ will get detected. Thus, $\Delta_{c,e,t} \subseteq S_{t−1}(2)$. Apply the second claim of the lemma. Since conditions 2 and 4 hold, all zero elements will get deleted and there will be no false detections, i.e. $|\Delta_{c,e,t}| = 0$. Finally, $|T_{e,t}| \leq |N_t| + |\Delta_{c,e,t}| \leq S_0 + 0$.

The second claim for time $t$ follows using the first claim for time $t−1$ and the arguments from the para above. The third claim follows using the second claim and Corollary 4.
C. Appendix: Proof of Theorem 2

We prove the first claim of the theorem by induction. Using condition 4 of the theorem, the claim holds for \( t = 0 \). This proves the base case. For the induction step, assume that the claim holds at \( t - 1 \), i.e., \( |\Delta_{t-1}| = 0 \), \( |T_{t-1}| \leq S_0 \), and \( \Delta_{t-1} \subseteq S_{t-1}(2) \) so that \( |\Delta_{t-1}| \leq 2S_0 \). Using this, we prove that the claim holds at \( t \). We will make the following facts often:

(a) \( R_t \subseteq N_t-1 \), (b) \( A_t \subseteq N_{t-1}^c \), (c) \( N_t = N_{t-1} \cup A_t \setminus R_t \), and (d) if two sets \( B, C \) are disjoint, then, \( D \cup C \cup B := (D \cup B) \setminus C \) for any set \( D \).

The bounding of \( |T_t|, |\Delta_t|, |\Delta_{t,t}| \) is exactly as in the proof of Theorem 1. Since \( T_t = T_{t-1} \), so \( |T_t| \leq S_0 \). Also, \( \Delta_{t,t} = N_{t-1} \setminus N_t = N_{t-1} \setminus \left[ (N_{t-1} \cap A_t) \cup R_t \right] \subseteq \Delta_{t-1,t} \cup R_t = R_t \). Thus \( |\Delta_{t,t}| \leq |R_t| = S_0 \). Finally, \( \Delta_t = N_t \setminus N_{t-1} = (\Delta_{t-1} \cap R_t^c) \cup (A_t \cap N_{t-1}^c) \subseteq \left( S_{t-1}(2) \cap R_t^c \right) \cup A_t \). Thus,

\[
\Delta_t \subseteq S_{t-1}(2) \cup A_t \setminus R_t \tag{31}
\]

Since \( R_t \subseteq S_{t-1}(2) \) and \( A_t \) is disjoint with \( S_{t-1}(2) \), thus \( |\Delta_t| \leq |S_{t-1}(2)| + |A_t| = |R_t| = 2S_0 + S_0 - S_0 \).

Consider the deletion step. We prove the first claim of the theorem by induction. Using addition and deletion thresholds, \( \Delta_{t-1} \subseteq S_{t-1}(2) \cup A_t \setminus R_t \).

The first case is \((\Delta_{t-1} \subseteq S_{t-1}(2) \cup A_t \setminus R_t) \). The second case follows by using conditions 4 of Proposition 2 applied directly from the first claim and fact 4 of Proposition 2 applied.

D. Appendix: Generalized version of Corollary 2

Corollary 4 (Stability of modified-CS with add-LS-del – 3): Assume Signal Model 1 and \( \|u_t\| \leq \epsilon \). Let \( e_t := (x_t - \hat{x}_{add,t}) \text{sat}, \). Assume that the LS step error is spread out enough so that

\[
\|e_t\| \leq \frac{\zeta}{\sqrt{\alpha}}
\]

at all \( t \). Consider Algorithm 2. If, for some \( 1 \leq d_0 \leq d \),

1. (addition and deletion thresholds)
   a) \( \alpha_{add} \) is large enough so that there are at most \( f \) false additions per unit time,
   b) \( \alpha_{del} = \sqrt{\frac{2}{S_0}} \zeta_m \epsilon + 2k_3 \theta_{S_0 + S_0 + f,k_2 \zeta_m \epsilon} \),

2. (support size, support change rate) \( S_0, S_A \) satisfy
   a) \( \delta_{S_0 + S_1} < (\sqrt{2} - 1)/2 \) and \( S_0 \leq \frac{S_0}{\delta_{S_0}} \),
   b) \( \delta_{S_0 + S_0 + f} < 1/2 \),
   c) \( \delta_{S_0 + f,k_2 S_A} < \frac{d_0}{4k_3 \zeta_m} \),

3. (new element increase bound) \( r \geq \max(\tilde{G}_1, \tilde{G}_2) \), where

\[
\tilde{G}_1 = \frac{\alpha_{add} + 8.79 \epsilon}{d_0} \]
\[
\tilde{G}_2 = \frac{2\sqrt{\zeta_m \epsilon}}{\sqrt{\alpha}} (d_0 - 4k_3 \theta_{S_0 + S_0 + f,k_2 \zeta_m}) \tag{33}
\]

4. (initial time) \( n_0 \) is large enough to ensure \( \tilde{A}_0 \leq S_0(d_0), \ |\Delta_{e,0}| \leq (2d_0 - 2)S_0, \ |\Delta_{e,0}| = 0, \ |\tilde{T}_0| \leq S_0, \)

where

\[
k_1 = \max(1, 2d_0 - 2) \]
\[
k_2 = \max(0, 2d_0 - 3) \]
\[
k_3 = \sqrt{\frac{d_0 - 1}{4k_3 \theta_{S_0 + S_0 + f,k_2 \zeta_m}}} \tag{34}
\]

then,

1. at all \( t \geq 0, \ |\tilde{T}_t| \leq S_0, \ |\Delta_{e,t}| = 0, \) and \( \tilde{A}_t \subseteq S_t(d_0) \) and so \( |\Delta_t| \leq (2d_0 - 2)S_0 \),
2. at all \( t > 0, \ |\tilde{T}_t| \leq S_0, \ |\Delta_{e,t}| \leq S_0, \) and \( |\tilde{A}_t| \leq k_1 S_A \),
3. at all \( t > 0, \ |T_{add,t}| \leq S_0 + S_0 + f, \ |\Delta_{e,add,t}| \leq S_0 + f, \) and \( |\Delta_{add,t}| \leq k_2 S_A \)
4. at all \( t > 0, \ |x_t - \hat{x}_t| \leq \sqrt{2 \epsilon} + k_3 \sqrt{\zeta_m \epsilon} (2\theta_{S_0}(2d_0 - 2)S_0 + 1) \)
5. at all \( t > 0, \ |x_t - \hat{x}_{t,modeLS}| \leq C_1 (S_0 + S_A + k_1 S_A) \epsilon \leq 8.79 \epsilon \).

Proof: The proof follows using exactly the same steps as in the proof of Theorem 2 but of course with Lemmas 4 and 5 replaced by Lemmas 6 and 7 respectively. The only difference is that, instead of ensuring \( |\Delta_{e,t}| = 0 \) and \( \tilde{A}_t \subseteq S_t(2) \), we try to ensure \( |\Delta_{e,t}| = 0 \) and \( \tilde{A}_t \subseteq S_t(d_0) \) for some \( d_0 \leq d \). For \( 1 < d_0 \leq d, \) notice that \( |S_t(d_0)| = (2d_0 - 2)S_0 \). Also, since, now, \( \Delta_{add,t} \subseteq S_t(d_0) \setminus D_t(d_0 - 1), \) so \( |\Delta_{add,t}| \leq (2d_0 - 3)S_0 \) and \( |\Delta_{add,t}| \leq k_3 S_A \). The case of \( d_0 = 1 \) is handled separately. In this case, \( S_t(d_0) \) is empty, but \( \tilde{A}_t \) is not empty, but is equal to \( A_t \). Also, \( \Delta_{add,t} \) and \( \tilde{A}_t \) are empty.
E. Proof of Lemma \[9\]

From Lemma \[8\] if \(|w| \leq \varepsilon, \delta_{2|\Delta|} < (\sqrt{2} - 1)/2\) and \(\delta_{T|T|} < 1/2\), then \(\|x - \hat{x}_{\text{CSres}}\| \leq C((|T|, |\Delta|)\varepsilon + \theta_{|T|,|\Delta|} C''((|T|, |\Delta|))\|x\|_1\). Using the fact that \(\|x\|_1 \leq \sqrt{|L_1| b + |x_{\Delta}|}\); fact \[1\] of Proposition \[2\] and the fact that for all \(i \in L_1, |x_i| \geq \gamma b\), we can conclude that all \(i \in L_1\) will get detected if

1. \(\delta_{2|\Delta|} < (\sqrt{2} - 1)/2\),
2. \(\delta_{T|T|} < 1/2\), and
3. \(\alpha_{\text{add}} + C'\varepsilon + \theta C''(\sqrt{|L_1| b + |x_{\Delta}|}) < \gamma b\). Using \(\|x_{\Delta}\| \leq \kappa b \text{ and } |L_1| \leq S_{L_1}\), this inequality holds if
   a. \(\theta C'' \leq \frac{\kappa}{2(\sqrt{|S_{L_1}|} + \kappa)}\) and
   b. \(\frac{\alpha_{\text{add}} + C'\varepsilon}{\theta C''(\sqrt{|S_{L_1}|} + \kappa)} < \gamma b\).

Since we only know that \(|T| \leq S_T, |\Delta| \leq S_{\Delta}\), we need the above inequalities to hold for all values of \(|T|, |\Delta|\) satisfying these upper bounds. This leads to the conclusion of the lemma. Notice that the LHS’s for the two inequalities are non-decreasing functions of \(|\Delta|, |T|\) and thus the lemma just uses their upper bounds. The LHS’s of the last two are non-decreasing in \(|T|\), but are not monotonic in \(|\Delta|\) (since \(C'(|T|, |\Delta|)\) and \(C''(|T|, |\Delta|)\) are not monotonic in \(|\Delta|\)). Hence we explicitly maximize over \(|\Delta| \leq S_{\Delta}\).

F. Proof of Lemma \[7\]

We provide the proof here for the sake of completeness and for ease of review. This will be removed later. Let \(b := \hat{x}_{\text{modes}} - x\). We adapt the approach of \[10\] to bound the reconstruction error, \(\|\hat{h}\| := \|\hat{x} - x\|\). A similar result was obtained in \[29\].

Let \(\Delta_1\) denote the set of indices of \(h\) with the \(|\Delta|\) largest values outside of \(T \cup \Delta\), let \(\Delta_2\) denote the indices of the next \(|\Delta|\) largest values and so on. Then using the same approach as that of \[10\],

\[
\|h_{(T \cup \Delta \cup \Delta_1)}\| \leq \sum_{j \geq 2} \|h_{\Delta_j}\| \leq \frac{1}{\sqrt{|\Delta|}} \|h_{(T \cup \Delta)}\|_1 \quad (35)
\]

Since \(\hat{x}_{\text{modes}} = x + h\) is the minimizer of \[2\] and since both \(x\) and \(\hat{x}_{\text{modes}}\) are feasible; and since \(x\) is supported on \(N \subseteq T \cup \Delta\),

\[
\|x_{\Delta}\|_1 = \|x_T\|_1 \geq \|(x + h)_T\|_1 \\
\geq \|x_T\|_1 - \|h_{\Delta}\|_1 + \|h_{(T \cup \Delta)}\|_1 \quad (36)
\]

Thus,

\[
\|h_{(T \cup \Delta)}\|_1 \leq \|h_{\Delta}\|_1 \quad (37)
\]

Combining this with \[35\] and using \(\frac{\|h_{\Delta}\|_1}{\sqrt{|\Delta|}} \leq \|h_{\Delta}\|\), we get

\[
\|h_{(T \cup \Delta \cup \Delta_1)}\| \leq \sum_{j \geq 2} \|h_{\Delta_j}\| \leq \|h_{\Delta}\| \quad (38)
\]

Next, since both \(x\) and \(\hat{x}_{\text{modes}}\) are feasible,

\[
\|Ah\| = \|A(x - \hat{x}_{\text{modes}})\| \\
\leq \|y - Ax\| + \|y - A\hat{x}_{\text{modes}}\| \leq 2\varepsilon \quad (39)
\]

In this proof, let

\[
\delta \triangleq \delta_{|T|+2|\Delta|} \text{ and } \theta \triangleq \theta_{|T|,|\Delta|} \quad (40)
\]

Now, we upper bound \(\|h_{(T \cup \Delta \cup \Delta_1)}\|\). To do that, notice that

\[
(1 - \delta)\|h_{(T \cup \Delta \cup \Delta_1)}\| \leq \|Ah_{(T \cup \Delta \cup \Delta_1)}\|^2 \leq \|Ah_{(T \cup \Delta \cup \Delta_1)}\|^2 \quad (41)
\]

To bound the RHS of the above, notice that \(Ah_{(T \cup \Delta \cup \Delta_1)} = Ah - \sum_{j \geq 2} Ah_{\Delta_j}\) and so

\[
\|Ah_{(T \cup \Delta \cup \Delta_1)}\|^2 = \langle Ah_{(T \cup \Delta \cup \Delta_1)}, Ah \rangle - \sum_{j \geq 2} \langle Ah_{\Delta_j}, Ah \rangle
\]

Using \[39\] and the definition of \(\delta_8\) given in \[3\],

\[
\langle Ah_{(T \cup \Delta \cup \Delta_1)}, Ah \rangle \leq 2\delta_1 + \|h_{(T \cup \Delta \cup \Delta_1)}\| \\
\text{Using the definition of } \theta_8 \text{ given in \[4\]; equation \[38\]; and the fact that } \|h_T\| + \|h_{\Delta}\| \leq \sqrt{2} \|h_{(T \cup \Delta \cup \Delta_1)}\|, \text{ we get the following. If } 2|\Delta| \leq |T|,
\]

\[
\|\sum_{j \geq 2} \langle Ah_{\Delta_j}, Ah \rangle \|
\leq (\theta_8 \|h_T\| + \theta_2|\Delta|, \|h_{\Delta}\|) \sum_{j \geq 2} \|h_{\Delta_j}\| \\
\leq \sqrt{2}\theta_8 \|h_{(T \cup \Delta \cup \Delta_1)}\| \|h_{\Delta}\| \quad (43)
\]

Combining the last four equations above, if \(2|\Delta| \leq |T|\),

\[
(1 - \delta)\|h_{(T \cup \Delta \cup \Delta_1)}\| \leq 2\delta_1 + \|h_{(T \cup \Delta \cup \Delta_1)}\| \\
\text{Using } \|h_{\Delta}\| \leq \|h_{(T \cup \Delta \cup \Delta_1)}\|, \text{ we can simplify the above to get}
\]

\[
\|h_{(T \cup \Delta \cup \Delta_1)}\| \leq \frac{2\delta_1 + \|h_{(T \cup \Delta \cup \Delta_1)}\|}{1 - \delta - \sqrt{2}\theta_8} \quad (45)
\]

Finally, using \[38\] and \(\|h_{\Delta}\| \leq \|h_{(T \cup \Delta \cup \Delta_1)}\|\) and the above,

\[
\|h\| \leq 2\|h_{(T \cup \Delta \cup \Delta_1)}\| \leq \frac{4\sqrt{1 + \delta}}{1 - \delta - \sqrt{2}\theta_8} \quad (46)
\]

Clearly, all of the above discussion holds only if the RHS is positive which is true only if \(\delta + \sqrt{2}\theta_8 < 1\). Also, \[42\] and hence everything after that needs \(2|\Delta| \leq |T|\). Since \(|T| = |N| + |\Delta| - |\Delta|\), this will hold if \(3|\Delta| \leq |N|\). Thus, we get the following result.

**Corollary 5:** If \(|\Delta| \leq |N|/3\) and if \(\delta_{|T|+2|\Delta|} + \sqrt{2}\theta_{|T|,|\Delta|} < 1\), then

\[
\|h\| \leq \frac{4\sqrt{1 + \delta}}{1 - \delta - \sqrt{2}\theta_{|T|,|\Delta|}} \quad (47)
\]

Using \(\theta_{|T|,|\Delta|} \leq \delta_{|T|+|\Delta|} \leq \delta_{|T|+2|\Delta|}\) in both the required sufficient condition and in the bound; and by substituting \(|T| = |N| + |\Delta| - |\Delta|\); and by using \(\sqrt{2} - 1 = 2\delta_8\), we get the notionally simpler result of Lemma \[7\].

**REFERENCES**

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

[2] S.G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” IEEE Trans. Sig. Proc., vol. 41(12), pp. 3397 – 3415, Dec 1993.

[3] S. Chen, D. Donoho, and M. Saunders, “Atomic decomposition by basis pursuit,” SIAM Journal of Scientific Computing, vol. 20, pp. 33–61, 1998.

[4] I.F. Gorodnitsky and B.D. Rao, “Sparse signal reconstruction from limited data using focus: A re-weighted norm minimization algorithm,” IEEE Trans. Sig. Proc., vol. 45, pp. 600–616, March 1997.
[5] D. Donoho, “Compressed sensing,” *IEEE Trans. Info. Th.*, vol. 52(4), pp. 1289–1306, April 2006.
[6] 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.
[7] D. Donoho, “For most large underdetermined systems of linear equations, the minimal ell-1 norm solution is also the sparsest solution,” *Comm. Pure and Appl. Math.*, vol. 59(6), pp. 797–829, June 2006.
[8] E. Candes and T. Tao, “Decoding by linear programming,” *IEEE Trans. Info. Th.*, vol. 51(12), pp. 4203 – 4215, Dec. 2005.
[9] 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.
[10] E. Candes, “The restricted isometry property and its implications for compressed sensing,” *Compte Rendus de l’Academie des Sciences, Paris, Serie I*, pp. 589–592, 2008.
[11] U. Gamper, P. Boesiger, and S. Kozerke, “Compressed sensing in dynamic mri,” *Magnetic Resonance in Medicine*, vol. 59(2), pp. 365–373, January 2008.
[12] 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.
[13] N. Vaswani, “Kalman filtered compressed sensing,” in *IEEE Intl. Conf. Image Proc. (ICIP)*, 2008.
[14] N. Vaswani, “LS-CS-residual (LS-CS): Compressive Sensing on Least Squares residual,” *IEEE Trans. Sig. Proc.*, vol. 58(8), pp. 4108–4120, August 2010.
[15] N. Vaswani and W. Lu, “Modified-cs: Modifying compressive sensing for problems with partially known support,” in *IEEE Intl. Symp. Info. Th. (ISIT)*, 2009.
[16] N. Vaswani and W. Lu, “Modified-cs: Modifying compressive sensing for problems with partially known support,” *IEEE Trans. Sig. Proc.*, vol. 58(9), pp. 4595–4607, September 2010.
[17] J. Ziniel, L. C. Potter, and P. Schniter, “Tracking and smoothing of time-varying sparse signals via approximate belief propagation,” in *Asilomar Conf. on Sig. Sys. Comp.*, 2010.
[18] M. S. Asif and J. Romberg, “Dynamic updating for sparse time varying signals,” in *Conf. Info. Sciences and Systems*, 2009.
[19] M. S. Asif and J. Romberg, “Dynamic Updating for ℓ1 minimization,” *IEEE J. Selected Topics in Signal Processing*, vol. 4(2), April 2010.
[20] E.J. Candes, M.B Wakin, and S.P. Boyd, “Enhancing sparsity by reweighted ℓ(1) minimization,” *Journal of Fourier Analysis and Applications*, vol. 14 (5-6), pp. 877–905, 2008.
[21] D. Angelosante, G.B. Giannakis, and E. Grossi, “Compressed sensing of time-varying signals,” in *Dig. Sig. Proc. Workshop*, 2009.
[22] C. J. Miosso, R. von Borries, M. Argez, L. Valazquez, C. Quintero, and C. Potes, “Compressive sensing reconstruction with prior information by iteratively reweighted least-squares,” *IEEE Trans. Sig. Proc.*, vol. 57 (6), pp. 2424–2431, June 2009.
[23] A. Khajehnejad, W. Xu, A. Avestimehr, and B. Hassibi, “Weighted l1 minimization for sparse recovery with prior information,” in *IEEE Intl. Symp. Info. Th. (ISIT)*, 2009.
[24] E. Candes and T. Tao, “The dantzig selector: statistical estimation when p is much larger than n,” *Annals of Statistics*, vol. 35 (6), pp. 2313–2351, 2007.
Fig. 3. Normalized MSE (NMSE), normalized number of extras and normalized number of misses over time for modified-CS (mod-CS), modified-CS with add-LS-del (mod-CS-add-LS-del), LS-CS and simple CS. In all cases, NMSE for simple CS was more than 20% (plotted only in (a) and (b)). We cannot use a logarithmic y-axis for plotting support errors since in some cases the errors are exactly zero.