Time Invariant Error Bounds for Modified-CS based Sparse Signal Sequence Recovery

Jinchun Zhan and 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 performance guarantees for modified-CS and for its improved version, modified-CS-Add-LS-Del, for recursive reconstruction of a time sequence of sparse signals from a reduced set of noisy measurements available at each time. Under mild assumptions, we show that the support recovery error of both algorithms is bounded by a time-invariant and small value at all times. The same is also true for the reconstruction error. Under a slow support change assumption, (i) the support recovery error bound is small compared to the support size; and (ii) our results hold under weaker assumptions on the number of measurements than what simple compressive sensing (CS) / basis pursuit denoising needs. We do the above for two types of signal change models. The first is a simple model that may often not be realistic. However it is used to illustrate the key ideas and it allows for easy comparison of the various results. The second one is a more complicated but realistic signal change model and includes the first model as a special case.

I. INTRODUCTION

The static sparse reconstruction problem has been studied for a while [2], [3], [4]. The papers on compressive sensing (CS) from 2005 [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. In more recent works, the problem of recursively recovering a time sequence of sparse signals, with slowly changing sparsity patterns has also been studied [11], [12], [13], [14], [15], [16], [17], [18]. By “recursive” reconstruction, we mean that we want to use only the current measurements’ vector and the previous reconstructed signal to recover the current signal. This problem occurs in many applications such as real-time dynamic magnetic resonance imaging (MRI); single-pixel camera based real-time video imaging; recursively separating the region of the brain that is activated in response to a stimulus from brain functional MRI (fMRI) sequences [19] and recursively extracting sparse foregrounds (e.g. moving objects) from slow-changing (low-dimensional) backgrounds in video sequences [20]. For other potential applications, see [21], [22].

An important assumption introduced and empirically verified in [11], [12] is that for many natural signal/image sequences, the sparsity pattern (support set of its projection into the sparsity basis) changes slowly over time. In [13], the authors exploited this fact to reformulate the above problem as one of sparse recovery with partially known support and introduced a solution approach called modified-CS. Given the partial support knowledge \( \mathcal{T} \), modified-CS tries to find a signal that is sparsest outside of \( \mathcal{T} \) among all signals that satisfy the data constraint. Exact recovery conditions were obtained for modified-CS and it was argued that these are weaker than those for simple CS (basis pursuit) under the slow support change assumption. Related ideas for support recovery with prior knowledge about the support entries, that appeared in parallel, include [23], [24]. All of [13], [23] and [24] studied the noise-free measurements’ case. Later work includes [25], [26].

Error bounds for modified-CS for noisy measurements were obtained in [27], [28], [1]. When modified-CS is used for recursive reconstruction, these bounds tell us that the reconstruction error bound at the current time is proportional to the support recovery error (misses and extras in the support estimate) from the previous time. Unless we impose extra conditions, this support error can keep increasing over time, in which case the bound is not useful. Thus, for recursive reconstruction, the important question is, under what conditions can we obtain time-invariant bounds on the support error (which will, in turn, imply time-invariant bounds on the reconstruction error)? In other words, when can we ensure “stability” over time? Notice that, even if we did nothing, i.e. we set \( \hat{x}_2 = 0 \), the support error will be bounded by the support size. If the support size is bounded, then this is a naive stability result too, but is not useful. Here, we look for results in which the support error bound is small compared to the support size.

This work was supported by NSF grants ECCS-0725849 and CCF-0917015. A part of this work was presented at Allerton 2010 [1] and another part will be presented at ISIT 2013. The results in this paper are a significant generalization of both conference papers and include all the proofs (which are missing in the ISIT paper due to lack of space).
We explain the differences of our results w.r.t. the LS-CS result in detail later in Sec IV-E.

A. Contributions

In this work, we introduce modified-CS-add-LS-del which is a modified-CS based algorithm for recursive recovery with an improved support estimation step and we explain how to set its parameters in practice. The main contribution of this work is to obtain conditions for stability of modified-CS and modified-CS-add-LS-del for recursive recovery of a time sequence of sparse signals. We compare the two results with each other and with the result for simple CS. Here and in the rest of the paper, simple CS refers to the solution of (4) (this is often referred to as basis pursuit denoising). Under mild assumptions, and for two types of signal change models, we show that the support recovery error, and reconstruction error, of both algorithms is bounded under slow support change, this bound is small compared to the support size, making our result meaningful. Similar arguments can be made for the reconstruction error also. The assumptions we need are: mild restricted isometry conditions on the measurement matrix; for any new element that is added to the support, either its initial magnitude is large enough or for the first few time instants, its magnitude increases at a large enough rate; and a similar assumption for magnitude decrease and removal from the support; appropriately set algorithm parameters; and a special start condition.

Under the slow support change assumption, we can also argue that our results hold under weaker restricted isometry assumptions than what simple CS needs. Also, the result for modified-CS-add-LS-del holds under weaker assumptions on the initial nonzero magnitude or the magnitude increase rate than that for modified-CS.

We do the above for two signal change models. The first is deliberately chosen to be a simple model that may often not be realistic. However, it helps to illustrate the key ideas of our results, and it allows for easy comparison of the results. The second model is a realistic, but more complicated, signal change model that allows different initial magnitudes; different rates of magnitude increase and of magnitude decrease at different times and for different elements; it also allows different numbers of support additions and removals at various times. We use MRI image sequences to demonstrate that this model is indeed realistic. However, it helps to illustrate the key ideas of our results, and it allows for easy comparison of the results. The second model is the focus of the current work.

B. Notation and Problem Definition

We let $[1, m] := [1, 2, \ldots, m]$. We let $\emptyset$ denote an empty set. 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 $A \setminus B := A \cap B^c$). If two sets $B, C$ are disjoint, we just write $D \cup B \setminus C$ instead of writing $(D \cup B) \setminus C$.

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 $\ell_k$ 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 $\|M\|_2$. $M'$ denotes the transpose of $M$ and $M^\dagger$ denotes the...
Moore-Penrose pseudo-inverse of $M$ (when $M$ is full column rank, $M^\dagger := (M'M)^{-1}M'$). Also, $M_T$ denotes the sub-matrix obtained by extracting the columns of $M$ corresponding to indices in $T$.

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

We assume the following observation model:

$$y_t = A_t x_t + w_t, \quad \|w_t\| \leq \epsilon$$

(1)

where $x_t$ is an $m$ length sparse vector with support set $\mathcal{N}_t$, i.e. $\mathcal{N}_t := \{i : (x_t)_i \neq 0\}$; $A_t$ is a $n_t \times m$ measurement matrix; $y_t$ is the $n_t$ length observation vector at time $t$ (with $n_t < m$); and $w_t$ is the observation noise. For $t > 0$, we fix $n_t = n$.

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 $t$.

Remark 1 (Why bounded noise): All results for bounding CS (CS solved using $\ell_1$ minimization) error in noise, and hence all results for bounding modified-CS error in noise, either assume a deterministic noise bound and then bound $\|\hat{x} - x\|$, e.g., [33], [30], [27], [31], [32]; or assume unbounded, e.g. Gaussian noise and then bound $\|\hat{x} - x\|$ with “large” probability, e.g. [33], [54] Sec IV, [12] Section III-A, [32]. The latter approach is not useful for recovering a time sequence of sparse signals because the error bound will hold for all times $0 \leq t < \infty$ with probability zero.

To get a meaningful error stability result with unbounded, e.g. Gaussian noise, one needs a way to compute or bound the expected value of the error. Moreover, even if one could do this for a given time, it would not tell us anything about the support recovery error (for the given noise sequence realization) and hence would not be useful for analyzing modified-CS.

As a sidenote, we should point out that, in most applications, the noise is typically bounded (because of finite sensing power available). One often chooses to model the noise as Gaussian because it simplifies performance analysis in many situations.

C. More Notation

For any matrix, $A$, the $S$-restricted isometry constant (RIC) $[8]$, $\delta_S(A)$ is the smallest real number satisfying

$$(1 - \delta_S)\|c\|^2 \leq \|A_T c\|^2 \leq (1 + \delta_S)\|c\|^2$$

(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) $[8]$, $\theta_{S_1,S_2}(A)$, is the smallest real number satisfying

$$|c_1'A_T^tA_Tc_2| \leq \theta_{S_1,S_2}\|c_1\|\|c_2\|$$

(3)

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, we need the same condition on the RIC and ROC of all measurement matrices $A_t$ for $t > 0$. Thus, in the rest of this paper, we let

$$\delta_S := \max_{t>0} \delta_S(A_t), \quad \theta_{S_1,S_2} := \max_{t>0} \theta_{S_1,S_2}(A_t).$$

If we need the RIC of ROC of any other matrix, then we specify it explicitly.

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-add-LS-del. In either case, we use $\hat{\mathcal{N}}_t$ to denote the support estimate at time $t$.

Definition 1 ($\hat{T}_t$, $\Delta_0$, $\Delta_{e,t}$): We use $\hat{T}_t := \hat{\mathcal{N}}_{t-1}$ to denote the support estimate from the previous time. This serves as the predicted support at time $t$. We use $\Delta_t := \mathcal{N}_t \setminus \hat{T}_t$ to denote the unknown part of $\mathcal{N}_t$ and $\Delta_{e,t} := \hat{T}_t \setminus \hat{\mathcal{N}}_t$ to denote the “erroneous” part of $\hat{\mathcal{N}}_t$.

With the above definition, clearly, $\hat{\mathcal{N}}_t = \hat{T}_t \cup \Delta_t \setminus \Delta_{e,t}$.

Definition 2 ($\tilde{T}_t$, $\Delta_t$, $\Delta_{e,t}$): We use $\tilde{T}_t := \hat{\mathcal{N}}_t$ to denote the final estimate of the current support; $\tilde{\Delta}_t := \mathcal{N}_t \setminus \tilde{T}_t$ to denote the “misses” in $\mathcal{N}_t$ and $\tilde{\Delta}_{e,t} := \tilde{T}_t \setminus \hat{\mathcal{N}}_t$ to denote the “extras”.

The sets $T_{\text{add}}, \Delta_{\text{add}}, \Delta_{e,\text{add}}$ are defined in Definition [3] which is given in the next section.
D. Other Related Work

“Recursive sparse reconstruction” also sometimes refers to homotopy methods, e.g. [35], 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). The goals in the above works are quite different from ours.

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 [36], [37], [38], [39]. This is done for iteratively improving the recovery of a single signal.

E. Paper Organization

The rest of the paper is organized as follows. The algorithms – modified-CS and modified-CS-add-LS-del – are introduced in Sec II. In Sec III, we give a very simple signal change model and derive stability results for modified-CS and modified-CS-add-LS-del under this set of model assumptions. In Sec IV we do the same for a realistic signal change model. The results are discussed in Sec III-D and IV-D respectively. In Sec V, we demonstrate that the signal model assumptions of Sec IV are indeed valid for medical imaging data. In Sec VI, we explain how to set the algorithm parameters automatically for both modified-CS and modified-CS-add-LS-del. In this section, we also give simulation experiments that back up some of our discussions from earlier sections. Conclusions and future work are given in Sec VII.

II. MODIFIED-CS AND MODIFIED-CS-ADD-LS-DEL FOR RECURSIVE RECONSTRUCTION

A. Modified-CS

Modified-CS was first proposed in [13] 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

\[
\min_{\beta} \| (\beta)_{T^c} \|_1 \text{ s.t. } \| y_0 - A_0 \beta \| \leq \epsilon
\]  

(4)

For recursively reconstructing a time sequence of sparse signals, we use the support estimate from the previous time, \(\hat{N}_{t-1}\), as the set \(T\). The simplest way to estimate the support is by thresholding the output of modified-CS. We summarize the complete algorithm in Algorithm 1.

At the initial time, \(t = 0\), we let \(T\) be the empty set, \(\emptyset\), i.e. we do simple CS. Alternatively, as explained in [13], 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.

Algorithm 1 Modified-CS

For \(t \geq 0\), do

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

\[
\min_{\beta} \| (\beta) \|_1 \text{ s.t. } \| y_0 - A_0 \beta \| \leq \epsilon
\]  

(4)

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

\[
\min_{\beta} \| (\beta)_{T^c} \|_1 \text{ s.t. } \| y_t - A_t \beta \| \leq \epsilon
\]  

(5)

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

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

(6)

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

B. Limitation: Biased solution

Modified-CS uses single step thresholding for estimating the support \(\hat{N}_t\). The threshold, \(\alpha\), needs to be large enough to ensure that all (or most) removed elements are correctly deleted and there are no (or very few) false detections. But this means that the new additions to the support set will either have to be added at a large value, or their magnitude will need to increase to a large value quickly enough to ensure correct detection within a small delay. This issue is further exaggerated by the fact that \(\hat{x}_{t,\text{modcs}}\) is a biased estimate of \(x_t\). Along \(T^c_t\), the values of \(\hat{x}_{t,\text{modcs}}\) will be biased towards zero (because we minimize
This also means that one can use a larger deletion threshold, which will be less biased than the modified-CS output. As a result, deletion will be more accurate when done using this estimate. A not critical. The addition step threshold, followed by LS estimation on the new support estimate, will have smaller error and will be less biased than the modified-CS output. As a result, deletion will be more accurate when done using this estimate. This also means that one can use a larger deletion threshold, which will ensure quicker deletion of extras.

Related ideas were introduced in our older work [12], [11] for KF-CS and LS-CS, and in [40], [30] for a greedy algorithm for static sparse reconstruction.

We explain how to automatically set the parameters for both modified-CS-add-LS-del and modified-CS in Sec VI-A.

Algorithm 2 Modified-CS-Add-LS-Del

For $t \geq 0$, do

1) Simple CS. If $t = 0$, set $\mathcal{T}_t = \emptyset$ and compute $\hat{x}_{t,\text{modcs}}$ as the solution of (4).

2) Modified-CS. If $t > 0$, set $\mathcal{T}_t = \hat{\mathcal{N}}_{t-1}$ and compute $\hat{x}_{t,\text{modcs}}$ as the solution of (5).

3) Additions / LS. Compute $\mathcal{T}_{\text{add},t}$ and the LS estimate using it:

$$\mathcal{A}_t = \{ i : |(\hat{x}_{t,\text{modcs}})_i | > \alpha_{\text{add}} \}$$

$$\mathcal{T}_{\text{add},t} = \mathcal{T}_t \cup \mathcal{A}_t$$

(7)

$$(\hat{x}_{t,\text{add}})_{\mathcal{T}_{\text{add},t}^t} = A_{\mathcal{T}_{\text{add},t}^t}^t y_t, \quad (\hat{x}_{t,\text{add}})_{\mathcal{T}_{\text{add},t}^t} = 0$$

(8)

4) Deletions / LS. Compute $\tilde{\mathcal{T}}_t$ and LS estimate using it:

$$\tilde{\mathcal{R}}_t = \{ i \in \mathcal{T}_{\text{add},t} : |(\hat{x}_{t,\text{add}})_i | \leq \alpha_{\text{del}} \}$$

$$\tilde{\mathcal{T}}_t = \mathcal{T}_{\text{add},t} \setminus \tilde{\mathcal{R}}_t$$

$$(\hat{x}_t)_{\tilde{\mathcal{T}}_t} = A_{\tilde{\mathcal{T}}_t^t}^t y_t, \quad (\hat{x}_t)_{\tilde{\mathcal{T}}_t^c} = 0$$

(9)

(10)

5) Set $\hat{\mathcal{N}}_t = \tilde{\mathcal{T}}_t$. Feedback $\hat{\mathcal{N}}_t$. Output $\hat{x}_t$.

**Definition 3 (Define $\mathcal{T}_{\text{add},t}, \Delta_{\text{add},t}, \Delta_{c,\text{add},t}$):** The set $\mathcal{T}_{\text{add},t}$ is the support estimate obtained after the support addition step. It is defined in (7) in Algorithm 2. The set $\Delta_{\text{add},t} := \mathcal{N}_t \setminus \mathcal{T}_{\text{add},t}$ denotes the set of missing elements from $\mathcal{N}_t$ and the set $\Delta_{c,\text{add},t} := \mathcal{T}_{\text{add},t} \setminus \mathcal{N}_t$ denotes the set of extras in it. We remove the subscript $t$ where not needed.

D. Modified-CS error bound at time $t$

By adapting the approach of (10), the error of modified-CS can be bounded as a function of $|\mathcal{T}| = |\mathcal{N}| + |\Delta_c| - |\Delta|$ and $|\Delta|$. This was done in (41). We state a modified version here.

**Lemma 1 (modified-CS error bound):** Assume that $y_t$ satisfies (1) and the support of $x_t$ is $\mathcal{N}_t$. Consider step 2 of Algorithm 1 or 2. If $\delta_{|\mathcal{T}|+3|\Delta|} = \delta_{|\mathcal{N}_t|+|\Delta_c|+2|\Delta|} < (\sqrt{2} - 1)/2$, then

$$\|x_t - \hat{x}_{t,\text{modcs}}\| \leq C_1(|\mathcal{T}| + 3|\Delta|) \epsilon \leq 7.50 \epsilon, \quad C_1(S) \leq \frac{4 \sqrt{1 + \delta_S}}{1 - 2 \delta_S}.$$

For completeness, we provide a proof in Appendix A.

Notice that the bound by $C_1(|\mathcal{T}| + 3|\Delta|) \epsilon$ will hold as long as $\delta_{|\mathcal{T}|+3|\Delta|} < 1/2$. By enforcing that $\delta_{|\mathcal{T}|+3|\Delta|} \leq 1/2c$ for a $c < 1$, we ensure that $C_1(.)$ is bounded by a fixed constant. To state the above lemma we pick $c = \sqrt{2} - 1$ and this gives $C_1(.) = 7.50$. We can state a similar result for CS [10].
Lemma 2 (CS error bound [10]): Let $x$ be a sparse vector with support $\mathcal{N}$ and let $y := Ax + w$ with $\|w\| \leq \epsilon$. Let $\hat{x}_{cs}$ denote the solution of (5) with $T = \emptyset$. If $\delta_{2|\mathcal{N}|} < (\sqrt{2} - 1)/2$, then

$$\|x - \hat{x}_{cs}\| \leq C_1(2|\mathcal{N}|)\epsilon \leq 8.57\epsilon$$

E. LS step error bound at time $t$

We can claim the following about the LS step error in step 3 of Algorithm 2.

**Lemma 3:** Assume that $y_t$ satisfies (1) and the support of $x_t$ is $\mathcal{N}_t$. Consider step 3 of Algorithm 2

1. $(x_t - \hat{x}_{t,\text{add}})T_{\text{add},t} = (A^{T\text{add},t}A^{T\text{add},t})^{-1}[A^{T\text{add},t}w_t + A^{T\text{add},t}A^{T\text{add},t}x_t - \hat{x}_{t,\text{add}}], (x_t - \hat{x}_{t,\text{add}}) = (x_t)_{\Delta_{\text{add},t}}$, and $(x_t - \hat{x}_{t,\text{add}})_i = 0$, if $i \notin T_{\text{add},t} \cup \Delta_{\text{add},t}$.

2. $\| (x_t - \hat{x}_{t,\text{add}})T_{\text{add},t} \| \leq \frac{1}{\sqrt{1 - \delta_{|\mathcal{T}|}}} \epsilon + \theta_{|\mathcal{T}|,|\Delta_{\text{add},t}|} \| (x_t)_{\Delta_{\text{add},t}} \|.$

3. $\| (x_t - \hat{x}_{t,\text{add}}) \| \leq \frac{1}{\sqrt{1 - \delta_{|\mathcal{T}|}}} \epsilon + (1 + \theta_{|\mathcal{T}|,|\Delta_{\text{add},t}|}) \| (x_t)_{\Delta_{\text{add},t}} \|.$

Proof: The first claim follows directly from the expression for $\hat{x}_{t,\text{add}}$. The second claim uses the first claim and the facts that $\|A^{-1}\|_2 \leq 1/\sqrt{1 - \delta_{|\mathcal{T}|}}, \|(A^TA)^{-1}\| \leq 1/(1 - \delta_{|\mathcal{T}|})$ and $\|A_{T,\Delta}\|_2 \leq \theta_{|\mathcal{T}|,|\Delta|}$ [12].

III. STABILITY RESULTS: SIMPLE SIGNAL MODEL

The modified-CS algorithms do not assume any signal model. However for showing stability, we need certain assumptions on the signal change over time. In this section, we assume a very simple signal change model which allows us to illustrate the key ideas and allows for easy comparison of the results.

A. Simple Signal Change Model

This simple model uses a single parameter, $r$, for the newly added elements’ magnitude and the magnitude increase and decrease rate of all elements at all times. It also fixes the number of support additions and removals to be $S_a$.

**Signal Model 1:** Assume the following.

1. **(addition and increase)** At each $t > 0$, $S_a$ new coefficients get added to the support at magnitude $r$. Denote this set by $\mathcal{A}_t$. 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 $M := dr$.

2. **(decrease and removal)** 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 $1 \leq j \leq (d - 2)$. 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 $\mathcal{R}_t$.

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

To understand the implications of the assumptions in Signal Model [11] we define the following sets. For $0 \leq j \leq d - 1$, let

$$\mathcal{D}_t(j) := \{ i : |x_{t,i}| = jr, \ |x_{t-1,i}| = (j + 1)r \}$$

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

$$\mathcal{I}_t(j) := \{ i : |x_{t,i}| = jr, \ |x_{t-1,i}| = (j - 1)r \}$$

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

$$\mathcal{S}_t(j) := \{ i : 0 < |x_{t,i}| < jr \}$$

denote the set of small but nonzero elements, with smallness threshold $jr$. Clearly, the newly added set, $\mathcal{A}_t = \mathcal{I}_t(1)$ and the newly removed set, $\mathcal{R}_t = \mathcal{D}_t(0)$. Also, $|\mathcal{I}_t(j)| = S_a$, $|\mathcal{D}_t(j)| = S_a$, $|\mathcal{S}_t(j)| = 2(j - 1)S_a$.

Consider a $1 < j \leq d$. From Signal Model [11] it is clear that at any time, $t$, $S_a$ elements enter the small elements’ set, $\mathcal{S}_t(j)$, from the bottom (set $\mathcal{A}_t$) and $S_a$ enter from the top (set $\mathcal{D}_t(j - 1)$). Similarly $S_a$ elements leave $\mathcal{S}_t(j)$ from the bottom (set $\mathcal{R}_t$) and $S_a$ from the top (set $\mathcal{I}_t(j)$). Thus,

$$\mathcal{S}_t(j) = \mathcal{S}_{t-1}(j) \cup (\mathcal{A}_t \cup \mathcal{D}_t(j - 1)) \setminus (\mathcal{R}_t \cup \mathcal{I}_t(j))$$

(11)
Since $A_t, R_t, D_t(j - 1), J_t(j)$ are mutually disjoint, $R_t \subseteq S_{t-1}(j)$ and $J_t(j) \subseteq S_{t-1}(j)$, thus, \[(1)\] implies that
\[S_{t-1}(j) \cup A_t \setminus R_t = S_t(j) \cup J_t(j) \setminus D_t(j - 1)\] (12)
Also, clearly, from Signal Model \[1\]
\[N_t = N_{t-1} \cup A_t \setminus R_t\] (13)

B. 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 $\Delta$ to definitely get deleted. These are obtained in Lemma \[3\] by using Lemma \[4\] and the following simple facts. Next, we use Lemma \[3\] to ensure that all new additions to the support get detected within a finite delay, and all removals from the support get deleted immediately.

**Proposition 1** (simple facts): Consider Algorithm \[4\]
1. An $i \in N_t$ will definitely get detected in step \[2\] if $|(x_t)_i| > \alpha + \|x_t - \hat{x}_{t,\text{modcs}}\|_\infty$.
2. Similarly, all $i \in \Delta_{e,t}$ (the zero elements of $T_t$) will definitely get deleted in step \[2\] if $\alpha \geq \|x_t - \hat{x}_{t,\text{modcs}}\|_\infty$.

In general, for any vector $z$, $\|z\|_\infty \leq \|z\|$ with equality holding only if $z$ is one-sparse (exactly one element of $z$ is nonzero). If the energy of $z$ is more spread out, $\|z\|_\infty$ will be smaller than $\|z\|$. Typically the error $x_t - \hat{x}_{t,\text{modcs}}$ will not be one-sparse, but will be more spread out. The assumption below states this.

**Assumption 1**: Consider Algorithm \[4\] Assume that the Modified-CS reconstruction error is spread out enough so that
\[
\|x_t - \hat{x}_{t,\text{modcs}}\|_\infty \leq \frac{\zeta_M}{\sqrt{S_a}} \|x_t - \hat{x}_{t,\text{modcs}}\|
\]
for some $\zeta_M \leq \sqrt{S_a}$.

Combining the above proposition and assumption with Lemma \[4\] we get the following lemma.

**Lemma 4**: Consider Algorithm \[4\] Assume Assumption \[4\] Assume that $|N_t| = S_{N_t}, |\Delta_{e,t}| \leq S_{\Delta_{e,t}}$ and $|\Delta_t| \leq S_{\Delta_t}$.
1. All elements of the set \{\(i \in N_t : |(x_t)_i| \geq b_1\)\} will get detected in step \[2\] if
   \[
   \delta_{S_{N_t} + S_{\Delta_{e,t}}} \leq 0.207, \quad \text{and} \quad b_1 > \alpha + \frac{\zeta_M}{\sqrt{S_a}} 7.50 \epsilon.
   \]
2. In step \[3\] there will be no false additions, and all the true removals from the support (the set $\Delta_{e,t}$) will get deleted at the current time, if
   \[
   \delta_{S_{N_t} + S_{\Delta_{e,t}}} \leq 0.207, \quad \text{and} \quad \alpha \geq \frac{\zeta_M}{\sqrt{S_a}} 7.50 \epsilon.
   \]

We use the above lemma to obtain sufficient conditions to ensure the following: for some $d_0 \leq d$, at all times, $t$, (i) only coefficients with magnitude less than $d_0 \epsilon$ are part of the final set of misses, $\Delta_t$, and (ii) the final set of extras, $\Delta_{e,t}$, is an empty set. In other words, we find conditions to ensure that $\Delta_t \subseteq S_t(d_0)$ and $|\Delta_{e,t}| = 0$. Using Signal Model \[4\] $|S_t(d_0)| = 2(d_0 - 1)S_a$ and thus $\Delta_t \subseteq S_t(d_0)$ will imply that $|\Delta_t| \leq 2(d_0 - 1)S_a$.

**Theorem 1** (Stability of modified-CS): Consider Algorithm \[4\] Assume Signal Model \[4\] on $x_t$. Also assume that $y_t$ satisfies \[4\]. Assume that Assumption \[4\] holds. If, for some $d_0 \leq d$, the following hold
1. (support estimation threshold) set $\alpha = \frac{\zeta_M}{\sqrt{S_a}} 7.50 \epsilon$
2. (number of measurements) $\delta_{S_t + (2k_1 + 1)S_a} \leq 0.207$,
3. (new element increase rate) \(r \geq G\), where
\[
G \triangleq \frac{\alpha + \frac{\zeta_M}{\sqrt{S_a}} 7.50 \epsilon}{d_0} \epsilon
\]
4. (initial time) at $t = 0$, $n_0$ is large enough to ensure that $\hat{\Delta}_0 \subseteq S_0(d_0)$, $|\hat{\Delta}_0| \leq 2(d_0 - 1)S_a$, $|\hat{\Delta}_{e,0}| = 0$ and $|\hat{T}_0| \leq S$ where
\[
k_1 \triangleq \max(1, 2d_0 - 2)
\]
then,
1. at all $t \geq 0$, $|\hat{T}_t| \leq S$, $|\hat{\Delta}_{e,t}| = 0$, $\hat{\Delta}_t \subseteq S_t(d_0)$ and so $|\hat{\Delta}_t| \leq 2(d_0 - 1)S_a$,
2. at all $t > 0$, $|\hat{T}_t| \leq S$, $|\hat{\Delta}_{e,t}| \leq S_a$, and $|\Delta_t| \leq k_1 S_a$,
3. at all $t > 0$, $\|x_t - \hat{x}_{t,\text{modcs}}\| \leq 7.50 \epsilon$.

**Proof**: The proof is given in Appendix \[4\] It follows using induction.

**Remark 2**: The condition \[4\] is not restrictive. It is easy to see that this will hold if $n_0$ is large enough to ensure that $\delta_{2S}(A_0) \leq 0.207$. 
C. 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 Lemma 1 and simple facts similar to Proposition 1.

As explained before, we can assume that the modified-CS reconstruction error is not one-sparse but is more spread out. The same assumption should also be valid for the LS step error. We state these next.

**Assumption 2:** Consider Algorithm 2. Assume that the Modified-CS reconstruction error is spread out enough so that,

$$\|x_t - \hat{x}_{t, \text{modcs}}\|_\infty \leq \frac{\zeta_M}{\sqrt{S_n}} \|x_t - \hat{x}_{t, \text{modcs}}\|_\infty$$

at all times, $t$, for some $\zeta_M \leq \sqrt{S_a}$. Similarly, assume that the LS step error along $T_{\text{add}, t}$ is spread out enough so that

$$\|(x_t - \hat{x}_{\text{add}, t})_{T_{\text{add}, t}}\|_\infty \leq \frac{\zeta_L}{\sqrt{S_n}} \|(x_t - \hat{x}_{\text{add}, t})_{T_{\text{add}, t}}\|_\infty$$

at all times, $t$, for some $\zeta_L \leq \sqrt{S_a}$.

Combining the above assumption with Lemmas 1 and 5 we get the following lemmas.

**Lemma 5 (Detection condition):** Consider Algorithm 2. Assume Assumption 2. Assume that $|N_1| = S_{N_1}, |\Delta_{e, t}| \leq S_{\Delta_{e, t}}, |\Delta_t| \leq S_{\Delta_t}$. Pick a $b_1 > 0$. All elements of the set $\{i \in \Delta : |(x_t)_i| \geq b_1\}$ will get detected in step 3 if

- $\delta_{S_{N_1} + S_{\Delta_{e, t}} + 2S_{\Delta_t}} \leq 0.207$, and $b_1 > \alpha_{\text{add}} + \sqrt{S_a} \theta_{\text{add}}$.

**Lemma 6 (Deletion and No false-deletion condition):** Consider Algorithm 2. Assume Assumption 2. Assume that $|T_{\text{add}, t}| \leq S_{T_{\text{add}, t}}$ and $|\Delta_{\text{add}, t}| \leq S_{\Delta_{\text{add}, t}}$.

1) Pick a $b_1 > 0$. No element of the set $\{i \in T_{\text{add}, t} : |(x_t)_i| \geq b_1\}$ will get (falsely) deleted in step 3 if

- $\delta_{S_{T_{\text{add}, t}} < 1/2}$ and $b_1 > \alpha_{\text{del}} + \sqrt{S_a} \theta_{\text{del}} (\sqrt{2} \epsilon + 2\theta_{S_{T_{\text{add}, t}}, S_{\Delta_{\text{add}, t}}} \|(x_t)_{T_{\text{add}, t}}\|_\infty)$.

2) All elements of $\Delta_{e, \text{add}}$ will get deleted in step 3 if

- $\delta_{S_{T_{\text{add}, t}} < 1/2}$ and $\alpha_{\text{del}} \geq \sqrt{S_a} \theta_{\text{del}} (\sqrt{2} \epsilon + 2\theta_{S_{T_{\text{add}, t}}, S_{\Delta_{\text{add}, t}}} \|(x_t)_{T_{\text{add}, t}}\|_\infty)$.

Using the above lemmas, we can obtain sufficient conditions to ensure that, for some $d_0 \leq d$, at each time $t$, $\hat{\Delta}_t \subseteq S_t(d_0)$ (so that $|\hat{\Delta}_t| \leq (2d_0 - 2)S_a$) and $|\hat{\Delta}_{e, t}| = 0$.

**Theorem 2 (Stability of modified-CS with add-LS-del):** Consider Algorithm 2. Assume Signal Model 1 on $x_t$. Also assume that $y_t$ satisfies 1. Assume that Assumption 2 holds. If, for some $1 \leq d_0 \leq d$, the following hold

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

2) (number of measurements)
   a) $\delta_{S + S_a (1 + 2k_1)} \leq 0.207$,
   b) $\delta_{S + S_a + f} < 1/2$,
   c) $\theta_{S + S_a + f, k_2 S_a} < \frac{1}{2} \frac{d_0}{\sqrt{k_3 S_a}}$

3) (new element increase rate) $r \geq \max(G_1, G_2)$, where
   $$G_1 \triangleq \frac{\alpha_{\text{add}} + \sqrt{S_a} \theta_{\text{add}}}{d_0} \frac{7.50 \epsilon}{\sqrt{2}}$$
   $$G_2 \triangleq \frac{2 \sqrt{S_a} (d_0 - 2k_2 S_a)}{\sqrt{S_a} (d_0 - 4k_3 \theta_{S + S_a + f, k_2 S_a} \zeta_{LR})}$$

4) (initial time) $n_0$ is large enough to ensure that $\hat{\Delta}_0 \subseteq S_0(d_0), |\hat{\Delta}_0| \leq (2d_0 - 2)S_a, |\hat{\Delta}_{e, 0}| = 0, |\hat{T}_0| \leq S$,

where

$$k_1 \triangleq \max(1, 2d_0 - 2)$$
$$k_2 \triangleq \max(0, 2d_0 - 3)$$
$$k_3 \triangleq \sqrt{\sum_{j=1}^{d_0-1} j^2 + \sum_{j=1}^{d_0-2} j^2}$$

then, at all $t \geq 0$. 
D. Discussion

Notice that, with Signal Model 1 at all times, $t$, the signals have the same support set size, $|\mathcal{N}_t| = S$ and the same signal power, $\|x_t\|^2 = (S - (2d_0 - 2)S_a)M^2 + 2S_a \sum_{j=1}^{d-1} j^2 r^2$.

The support error bound in both results above is proportional to $S_a$. Thus, under slow support change, i.e. $S_a \ll S$, this bound is small compared to the support size $S$. Also, the reconstruction error is bounded below a constant times $\epsilon$. If signal power is large enough compared to the noise (high enough signal-to-noise ratio), this bound is also small compared to the signal power.

To make the comparisons simpler, let us fix $d_0 = 2$ and let $f = S_a$ in Theorem 2. Consider the conditions on the number of measurements. Modified-CS needs $\delta_{S+5S_a} \leq 0.207$. Modified-CS-add-LS-del needs $\delta_{S+5S_a} \leq 0.207$; $\delta_{S+2S_a} < 0.5$ (this is implied by the first condition) and $\theta_{S+2S_a,S_a} \leq \frac{1}{4}\epsilon$. Since $\theta_{S+2S_a,S_a} \leq \delta_{S+3S_a}$, the third condition is also implied by the first as long as $\zeta_j \leq 1.2$. In simulation tests (described in Sec IV-D) we observed that this was usually true. Then, both modified-CS and modified-CS-add-LS-del need the same condition on the number of measurements: $\delta_{S+5S_a} \leq 0.207$. Consider simple CS. Since simple CS is not a recursive approach (each time instant is handled separately), Lemma 2 is also a stability result for it. From Lemma 2 simple CS needs $\delta_{2S} \leq 0.207$ to get the same error bound. Under the slow support change assumption, $S_a \ll S$. In this case, clearly simple CS requires a stronger condition than either of the modified-CS algorithms.

Let us compare the requirement on $r$. In Theorem 2 for modified-cs-add-lds-del, since $\theta_{S+S_a+f,k_2S_a} \leq \frac{1}{4}\frac{d_0}{\sqrt{2} \sqrt{5} S_a}$, so $\delta_2 \leq \frac{4\sqrt{2}\alpha_{add} \epsilon}{\sqrt{S_a} d_0} \frac{\sqrt{5} r}{d_0} \leq \frac{7.50\epsilon}{d_0} \frac{\sqrt{5} r}{d_0} < G_1$ and thus $G_1$ is what decides the minimum allowed value of $r$. Thus, it needs $r \geq G_1 = \frac{d_0}{\frac{4\sqrt{2}\alpha_{add} \epsilon}{\sqrt{S_a} d_0}}$. On the other hand, modified-CS needs $r \geq G = \frac{d_0}{\frac{4\sqrt{2}\alpha_{add} \epsilon}{\sqrt{S_a} d_0}} \frac{7.50\epsilon}{d_0}$. If $\alpha_{add}$ is close to zero, this means that the minimum magnitude increase rate, $r$, required by Theorem 2 is almost half of that required by Theorem 1. In our simulation experiments, $\alpha_{add}$ was typically quite small: it was usually close to a small constant times $\epsilon/\sqrt{n}$ (see Sec VI).

Remark 3: In the discussion above, we fixed $d_0$. Now let us see the reason for allowing $d_0$ to be anything below $d$. If the rate of magnitude increase, $r$, is smaller, $r \geq G_1$ or $r \geq G$ 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 the RIC and ROC are needed (in other words, how many measurements, $n_t$, needed). Smaller $r$ means a larger $d_0$ is needed which, in turn, means that stronger conditions on the RIC and ROC (larger $n_t$) are needed. Thus, for a given $n_t = n$, 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 $d_0$).

IV. Stability Results: Realistic Signal Model

We introduce the signal change model in the next subsection and then give the results in the following two subsections. Discussion is given in the following two subsections.

A. Realistic Signal Change Model

Briefly, our model assumes the following. At any time the signal vector $x_t$ is a sparse vector with support set $\mathcal{N}_t$ of size $S$ or less. At most $S_a$ elements get added to the support at each time $t$ and at most $S_a$ elements get removed from it. At time $t = t_j$, a new element $j$ gets added at an initial magnitude $a_j$, and its magnitude increases for the next $d_j \geq d_{\text{min}}$ time units. Its magnitude increase at time $\tau$ (for any $t_j < \tau \leq t_j + d_j$ is $r_{j,\tau}$). Also, at each time $t$, at most $S_a$ elements out of the “large elements” set (defined in the signal model) leave the set and begin to decrease. These elements keep decreasing and get removed from the support in at most $b$ time units. In the model as stated above, we are implicitly allowing an element $j$ to get added to the support at most once. In general, $j$ can get added, then removed and then added again. To allow for this, we let $t_j$ be the set of time instants at which $j$ gets added; we replace $a_j$ by $a_{j,t}$ and we replace $d_j$ by $d_{j,t}$ (both of which are nonzero only for $t \leq t_j$).

As demonstrated in Section VI the above assumptions are practically valid for MRI sequences. We specify the model precisely below.
Signal Model 2: Assume the following.

1) At the initial time, \( t = 0 \), the support set, \( N_0 \), contains \( S_0 \) nonzero elements, i.e. \( |N_0| = S_0 \).
2) At time \( t \), \( S_{a,t} \) elements are added to the support set. Denote this set by \( A_t \). At time \( t \), a new element \( j \) gets added to the support at an initial magnitude \( a_{j,t} \) and its magnitude increases for at least the next \( d_{min} > 0 \) time instants. At time \( \tau \) (for \( t < \tau \leq t + d_{min} \)), the magnitude of element \( j \) increases by \( r_{j,\tau} \geq 0 \).
   - \( a_{j,t} \) is nonzero only if element \( j \) got added at time \( t \), for all other times, we set it to zero.
3) We define the “large set” as
   \[ L_t := \{ j \not\in \bigcup_{\tau=t-d_{min}+1}^t A_\tau : |(x_t)_j| \geq \ell \} \]
   for a given constant \( \ell \). Elements in \( L_{t-1} \) either remain in \( L_t \) (while increasing or decreasing or remaining constant) or decrease enough to leave \( L_t \).
4) At time \( t \), \( S_{d,t} \) elements out of \( L_{t-1} \) decrease enough to leave \( L_{t-1} \). Denote this set \( B_t \). All these elements continue to keep decreasing and become zero (removed from support) within at most \( b \) time units. Also, at time \( t \), \( S_{r,t} \) elements out of these decreasing elements are removed from the support. Denote this set by \( R_t \).
5) At all times \( t \), \( 0 \leq S_{a,t} \leq S_a \), \( 0 \leq S_{d,t} \leq \min \{ S_a, |L_{t-1}| \} \), \( 0 \leq S_{r,t} \leq S_a \) and the support size, \( S_t := |N_t| \leq S \) for constants \( S \) and \( S_a \) such that \( S + S_a \leq m \).

The above is not a generative model. It is only a set of assumptions on signal change. One possible generative model that satisfies these assumptions is given in Appendix C.

Remark 4: It is easy to see that Signal Model 2 is a special case of Signal Model 1 with \( a_{j,t} = r_{j,t} = r \), \( d_{min} = d \), \( b = d \), \( S_0 = S \), \( S_{a,t} = S_{d,t} = S_{r,t} = S_a \), \( \ell = dr \).

From the above model, the newly added elements’ set \( A_t := N_t \setminus N_{t-1} \); the newly removed elements’ set \( R_t := N_{t-1} \setminus N_t \); the set of elements that begin to start decreasing at \( t \), \( B_t := L_{t-1} \setminus L_t \). Define the following sets: the set of increasing (actually non-decreasing) elements at \( t \),
   \[ I_t := \{ j \in N_t : |(x_t)_j| \geq |(x_{t-1})_j| \} \]
and the set of small and decreasing elements,
   \[ SD_t := L_t^c \cap \{ i \in N_t : 0 < |(x_t)_i| < |(x_{t-1})_i| \} \]

Notice that \( I_t \) also includes \( j \) if its magnitude does not change from \( t - 1 \) to \( t \).

Condition 2 of the above model implies that (i) \( |A_t| = S_{a,t} \); (ii) if \( j \in A_{t-t_0} \) (i.e. if \( j \) is added at \( t - t_0 \)) for a \( t_0 \leq d_{min} \), then \( |(x_t)_j| = a_{j,t-t_0} + \sum_{\tau=t-t_0}^{t} r_{j,\tau} \); and (iii) \( A_t \subseteq I_t \cap I_{t+1} \cdots \cap I_{t+d_{min}} \) (all newly added elements increase for at least \( d_{min} \) time instants).

Condition 3 implies that \( L_{t-1} \subseteq L_t \cup SD_t \). It also implies that \( (\cup_{\tau=t-d_{min}+1}^t A_\tau) \cap L_t = \emptyset \). This, along with condition 2 means that \( \cup_{\tau=t-d_{min}+1}^t A_\tau \subseteq I_t \).

Condition 4 implies that \( |B_t| = S_{d,t} \); \( L_{t-1} \setminus B_t \subseteq L_t \); \( SD_t = SD_{t-1} \cup B_t \setminus R_t \); \( \sum_{\tau=1}^t S_{r,\tau} \geq \sum_{\tau=1}^{t-b} S_{d,\tau} \); \( |SD_t| \leq \sum_{\tau=t-b+1}^{t} S_{d,\tau} \); and \( |R_t| = S_{r,\tau} \).

Condition 5, along with the above, implies that \( |SD_t| \leq b S_a \).

Finally, it is easy to see that \( N_t = I_t \cup L_t \cup SD_t \). The sets \( I_t \), \( L_t \) are not disjoint, but both are disjoint with \( SD_t \).

The above model tells us the following. Consider an element \( j \) that got added at time \( t_i \), i.e. \( j \in A_{t_i} \). At \( \tau = t_i + 1, \ldots t_i + d_{min} - 1 \), \( j \in L_{\tau_i} \) and \( j \not\in L_{\tau} \). At \( \tau = t_i + d_{min} \), \( j \in L_{\tau_i} \); if \( |(x_{\tau_i})_j| \geq \ell \) then \( j \in L_{\tau_i} \) as well. For \( \tau > t_i + d_{min} \), what happens depends on \( \tau - 1 \). If \( j \in L_{\tau_i} \), then either \( j \in L_{\tau} \) or it decreases enough to enter the small and decreasing set, i.e. \( j \in B_{\tau_i} \cup SD_{\tau_i} \). If \( j \in SD_{\tau_i} \), then either it keeps decreasing or gets removed, i.e. either \( j \in SD_{\tau} \) or \( j \in R_{\tau} \subseteq N_{\tau}^c \). If \( j \in L_{\tau_i} \cap I_{\tau_i} \), then, if \( |(x_{\tau_i})_j| \geq \ell \) then \( j \in L_{\tau_i} \cap I_{\tau_i} \), else \( j \in L_{\tau_i}^c \cap I_{\tau_i} \).

We now discuss sufficient conditions for condition 5 of the signal model to hold.

Remark 5: Since \( S_t = S_{t-1} + S_{a,t} - S_{r,t} = S_0 + \sum_{\tau=1}^t S_{a,\tau} - \sum_{\tau=1}^t S_{r,\tau} \), thus, \( S_t \leq S \) holds if \( S_0 \leq S \) and \( \sum_{\tau=1}^t S_{a,\tau} \leq \sum_{\tau=1}^{t-b} S_{d,\tau} \).

Notice that an element \( j \) could get added, then removed and added again later. Let
   \[ t_j := \{ t : a_{j,t} \neq 0 \} \]
denote the set of time instants at which \( j \) gets added. Clearly, \( t_j = \emptyset \) if \( j \) never got added. Let
   \[ a_{\min} := \min_{j \in t_j, t_j \neq \emptyset} \min_{t \in t_j, t_j \neq 0} a_{j,t} \]
denote the minimum of $a_{j,t}$ over all elements $j$ that got added at $t > 0$. We are excluding coefficients that never got added and those that got added at $t = 0$. Let

$$r_{\min}(d) := \min_{j: t_j \neq 0} \min_{t \in [0, t_j]} \min_{\tau \neq t} r_{j,\tau}$$

denote the minimum, over all elements $j$ that got added at $t > 0$, of the minimum of $r_{j,\tau}$ over the first $d$ time instants after $j$ got added.

Define

$$\ell := a_{\min} + d_{\min} r_{\min}(d_{\min}).$$

With $\ell$ defined this way, clearly, $N_t = (\cup_{\tau = t - d_{\min} + 1}^{t} A_{\tau}) \cup \mathcal{L}_t \cup SD_t$ where the three sets are mutually disjoint.

Also, with $\ell$ as above, it is clear that for $t > d_{\min}$, $\mathcal{L}_t = \mathcal{L}_{t-1} \cup \mathcal{A}_{t-d_{\min}-1} \setminus \mathcal{B}_t$, and for $t \leq d_{\min}$, $\mathcal{L}_t = \mathcal{L}_{t-1} \setminus \mathcal{B}_t$. Here, by definition, $\mathcal{L}_{t-1}$ and $\mathcal{A}_{t-d_{\min}-1}$ are disjoint and $\mathcal{B}_t \subseteq \mathcal{L}_{t-1}$. Thus,

$$|\mathcal{L}_t| = |\mathcal{L}_0| + \sum_{\tau = 1}^{t-d_{\min}} S_{a,\tau} - \sum_{\tau = 1}^{t} S_{d,\tau}.$$

Also notice that $|\mathcal{L}_0| \leq S_0$. Using these facts and Remark 5, we can conclude the following.

**Remark 6:** Let $\ell := a_{\min} + d_{\min} r_{\min}(d_{\min})$. Then, condition 5 of Signal Model 3 holds if

1. $0 \leq S_{a,t} \leq S_a$ and $0 \leq S_{d,t} \leq S_d$,
2. $(d_{\min} + b + 1) S_a \leq |\mathcal{L}_0| \leq S_0 \leq S$, and
3. $\sum_{\tau = 1}^{\ell} S_{a,\tau} \leq \sum_{\tau = 1}^{t-d_{\min}} S_{d,\tau} \leq |\mathcal{L}_0| + \sum_{\tau = 1}^{t} S_{d,\tau}.$

The leftmost lower bound of the second condition ensures that the upper bound of the third condition is not smaller than the lower bound. The upper bound of the third condition ensures that $S_{d,t} \leq |\mathcal{L}_{t-1}|$ always (it is actually written to ensure $S_{d,t} \leq |\mathcal{L}_{t-1}|$). $S_0 \leq S$ and the lower bound of the third condition ensures that $S_t \leq S$ (as explained in Remark 5).

A simpler sufficient condition is as follows.

**Remark 7:** Let $\ell := a_{\min} + d_{\min} r_{\min}(d_{\min})$. Then, condition 5 of Signal Model 3 holds if $(d_{\min} + b + 1) S_a \leq |\mathcal{L}_0| \leq S_0 \leq S; S_{d,t} = S_d$ for all $t$; and for $1 \leq t \leq b$, $S_{a,t} = 0$, and for $t > b$, $S_{a,t} = S_a$.

In the above model, we only assume that all coefficients will get removed in at most $b$ time units. However, it can happen that some coefficients get removed earlier than that and hence it is fair to include this in the signal model. We do this below.

**Signal Model 3:** Assume Signal Model 3 with the following extra assumption.

- Out of the $S_{d,t}$ elements that started decreasing at time $t$, at least $\frac{1}{2} S_{d,t}$ of them get removed by $t + \tau$ for $\tau < b$.

All implications of the above model are the same as those of Signal Model 3 except that now, $|SD_t| \leq S_{d,t} + \frac{b+1}{b} S_{d,t-1} + \ldots + \frac{1}{b} S_{d,t-b+1} \leq \frac{b+1}{b} S_a$; while for Signal Model 3 $|SD_t| \leq b S_a$.

**B. Modified-CS Stability Result**

For the above signal model, we can claim the following.

**Theorem 3:** Consider Algorithm 11 Assume Signal Model 3 on $x_t$. Also assume that $y_t$ satisfies (11). Assume that Assumption 1 holds. If there exists a $d_0 \leq d_{\min}$ such that the following hold:

1. algorithm parameters
   a) $\alpha = \frac{\zeta_M}{\sqrt{S_a}} 7.50 \epsilon,$
   b) number of measurements
      a) $\delta_{S + (b+1) S_a} \leq 0.207,$
      b) $t + d_0 \epsilon + \sum_{\tau = t+1}^{T} r_{j,\tau}$

3. initial magnitude and magnitude increase rate:

   $$\min\{\ell, \min_{j: t_j \neq 0} \min_{t \in [0, t_j]} r_{j,\tau}\}$$

   $$\geq \alpha + \frac{\zeta_M}{\sqrt{S_a}} 7.50 \epsilon,$$

4. at $t = 0$, $\nu_0$ is large enough to ensure that $|\Delta_t| \leq \frac{b+1}{b} S_a + \nu_0 S_a; |\Delta_{e,t}| = 0, |\tilde{T}| \leq S$, then, for all $t$,

   1) $|\Delta_t| \leq \frac{b+1}{b} S_a + \nu_0 S_a; |\Delta_{e,t}| = 0, |\tilde{T}| \leq S,$
2) \(|\Delta_t| \leq \frac{(b+1)}{2}S_a + d_0S_a + S_a, \ |T_t| \leq S, \ |\Delta_{e,t}| \leq S_a,\)
3) and \(\|x_t - \hat{x}_t\| \leq 7.50\epsilon\)

Proof: See Appendix D.

Corollary 1: Under Signal Model 2, the result of Theorem 3 changes in the following way: replace \(\frac{(b+1)}{2}S_a\) by \(bS_a\) everywhere in the result.

Remark 8: Condition 4 of the above result is not restrictive. It is easy to see that it will hold if \(\delta_{2\sigma}(A_0) \leq 0.207\) and if \([L_0] \geq [S_0 - (\frac{(b+1)}{2}S_a + d_0S_a)].\)

Remark 9: A simpler sufficient condition for condition 3 is: \(\min(\ell, a_{\min} + d_0r_{\min}(d_0)) > \alpha + \frac{\zeta M}{\sqrt{S_a}}7.50\epsilon.\)

C. Modified-CS-Add-LS-Del Stability Result

Finally we study Modified-CS-Add-LS-Del.

Theorem 4: Consider Algorithm 2 Assume Signal Model 2 on \(x_t\). Also assume that \(y_t\) satisfies (1). Assume that Assumption 2 holds. If there exists a \(d_0 \leq d_{\text{min}}\) such that the following hold:

1) algorithm parameters
   a) \(\alpha_{\text{add}}\) is large enough so that there are at most \(f\) false adds at time \(t\), i.e. \(|\hat{A}_t \setminus N_t| \leq f\)
   b) \(\alpha_{\text{det}} = 1.12 \frac{\zeta L}{\sqrt{S_a}} + 0.261\zeta L h, \) where \(h^2 = (\frac{(b+1)}{2} + d_0)(\alpha_{\text{add}} + \frac{\zeta M}{\sqrt{S_a}}7.50\epsilon)^2\)

2) number of measurements
   a) \(\delta_{S+f(S+d_0S_a)} \leq 0.207\)
   b) \(\delta_{S+f(S+d_0S_a)} \leq 0.207\)
   c) \(\theta_{S+f(S+d_0S_a)} \leq 0.207\)

3) initial magnitude and magnitude increase rate:
\[
\min\{\ell, \min_{j: \tau_j \neq 0} \min_{t \in t_j} (a_{j,t} + \sum_{\tau = t+1}^{t+d_0} r_{j,\tau})\} > \min\{\alpha_{\text{add}} + \frac{\zeta M}{\sqrt{S_a}}7.50\epsilon, 2\alpha_{\text{det}}\}\] (19)

4) at \(t = 0, n_0\) is large enough to ensure that \(|\hat{A}_t| \leq \frac{(b+1)}{2}S_a + d_0S_a, |\hat{A}_{e,t}| = 0,\) then

1) \(\hat{A}_t \subseteq SP_t \cup A_t \cup A_{t-1} \ldots A_{t-d_0+1}\)
2) \(|\hat{A}_t| \leq \frac{(b+1)}{2}S_a + d_0S_a, |\hat{A}_{e,t}| = 0, |\hat{T}_t| \leq S\)
3) \(|\Delta_t| \leq \frac{(b+1)}{2}S_a + d_0S_a + S_a, |T_t| \leq S\)
4) \(\|x_t - \hat{x}_{t,\text{modes}}\| \leq 7.50\epsilon,\)
5) \(\|x_t - \hat{x}_t\| \leq 1.12\epsilon + 1.261\sqrt{(\frac{(b+1)}{2} + d_0)(\alpha_{\text{det}} + 7.50\epsilon)}S_a.\)

Proof: See Appendix E.

Remark 10: Claims similar to Corollary 1 and Remarks 8 and 9 hold for the above result also.

D. Discussion

Remark 11: Consider the noise-free case, i.e. the case when \(\epsilon = 0\). In this case, our results say the following: if RIP of order \(S + kS_a\) holds with \(\delta_{S+kS_a} < 0.207\), and if the support thresholds are set to zero, then both algorithms will exactly recover all sparse signal sequences with support size at most \(S\), and number of support additions and removals per unit time at most \(S_a\). It is easy to show that RIP (or actually left-RIP) of order \(S + S_a\) at all times \(t > 0\) is also necessary for the above. We give a proof for this in Appendix E. Thus the sufficient condition that our results need is only slightly stronger and clearly, cannot be improved much further since RIP(\(S + S_a\)) is necessary. Thus, for example, RIP of order \(S + k'\sqrt{S_a}\) or \(S + k'S_a\) will not work. This remark is inspired by a concern of an anonymous reviewer.

Remark 12: Notice that Signal Models 2 or 3 allow for both slow and fast signal magnitude increase or decrease. Slow magnitude increase/decrease would happen, for example, in an imaging problem when one object slowly morphs into another with gradual intensity changes. Or, in case of brain regions becoming “active” in response to stimuli, the activity level gradually increases from zero to a certain maximum value within a few milliseconds (10-12 frames of fMRI data), and similarly the “activity” level decays to zero within a few milliseconds. In both of the above examples, a new coefficient will get added to the support at time \(t\) at a small magnitude \(a_{j,t}\) and increase by \(r_{j,t}\) per unit time for sometime after that. Similarly for the
decay to zero of the brain’s activity level. On the other hand, the signal model also allows support changes resulting from motion of objects, e.g., translation. In this case, the signal magnitude changes will typically not be slow. As the object moves, a set of new pixels enter the support and another set leave. The entering pixels may have large enough pixel intensity and their intensity may never change. For our model this means that the pixel enters at a large enough initial magnitude $a_{j,t}$ but its magnitude never changes i.e. $r_{j,\tau} = 0$ for all $\tau$. If all pixels exit the support without their magnitude first decreasing, then $b = 1$.

The only thing that the above results (Theorem 3 and 4) require is that (i) for any element $j$ that is added, either $a_{j,t}$ is large enough or $r_{j,\tau}$ is large enough for the initial few $(d_0)$ time instants so that condition 3 holds; and (ii) a decaying coefficient decays to zero within a short delay, $b$. (i) ensures that every newly added support element gets detected either immediately or within a finite delay; while (ii) ensures removal within finite delay of a decreasing element. For the moving object case, this translates to requiring that $a_{j,t}$ be large enough. For the first two examples above, this translates to requiring that $r_{j,\tau}$ be large enough for the first few time instants after $j$ gets added and that $b$ be small enough.

Recall that $\delta_S := \max_{\tau > 0} \delta_S(A_t)$. Other than the above assumption, the results also need that the support estimation thresholds are set appropriately; enough number of measurements, $n_t$, are available at all times $t > 0$ so that condition 2 holds (this number depends on the support size, $S$, the support change size, $S_a$ and on $b$); and condition 4 holds.

For the above results, the support errors are bounded by a constant times $S_t$. Thus, under slow support change, the bound is small compared to the support size, $S_t$, making the above a meaningful result. The reconstruction error is bounded by a constant times $\epsilon$. Under high enough SNR, this bound is small compared to the signal power. In fact, for Signal Models 2 or 3 the signal power is not bounded. To compare the results, let us fix some of the parameters. Suppose that $b = 3$, $f = S_a$, $S_0 = S$, $S_{a,t} = S_{r,t} = S_{d,t} = S_a$. Let $d_0 = 2$. The modified-CS result says the following. If

1. $\delta_{S_{15}S_a} \leq 0.207$, and
2. LHS of condition 3 > $\frac{c_d}{\sqrt{S_a}} 15\epsilon$,
then $|\hat{\Delta}| \leq 4S_a$ and $|\hat{\Delta}_{x,t}| = 0$ and $\|x_t - \hat{x}_{t,modcs}\| \leq 7.50\epsilon$. The Modified-CS-add-LS-del result says the following. If

1. $\delta_{S_{15}S_a} \leq 0.207$(the other two conditions are implied by this), and
2. LHS of condition 3 $> \max(\alpha_{add} + \frac{c_d}{\sqrt{S_a}} 7.50\epsilon, 2.24 \frac{c_d}{\sqrt{S_a}} \epsilon + 0.522\zeta_S h)$, where $h^2 = 4(\alpha_{add} + \frac{c_d}{\sqrt{S_a}} 7.50\epsilon)^2$.
then $|\hat{\Delta}| \leq 4S_a$ and $|\hat{\Delta}_{x,t}| = 0$ and $\|x_t - \hat{x}_{t,modcs}\| \leq 7.50\epsilon$.

The CS result from Lemma 2 says the following. If

1. $\delta_{S_2} \leq 0.207$
then $\|x_t - \hat{x}_{t,cs}\| \leq 8.57\epsilon$.

Thus, both modified-CS and modified-CS-add-LS-del need the same restricted isometry condition (condition on the number of measurements). Under the slow support change assumption, $S_a \ll S_t \leq S$. In this case, both the modified-CS algorithms hold under a weaker restricted isometry condition (potentially fewer number of measurements required) than what simple CS needs for the same reconstruction error bound.

Next we compare the lower bounds on the LHS of condition 3 needed by modified-CS and by modified-CS-add-LS-del. This requires knowing $\zeta_M$ and $\zeta_L$. To get an idea of the values of $\zeta_M$ and $\zeta_L$, we did simulations based on Signal Model 2 with $S = 0.1m, S_{a,t} = S_{d,t} = S_{r,t} = S_a = 0.01m, b = d_{\text{min}} = 3, r_{j,t} = 1, a_{j,t} = 1$ (we generated it using the generative model given in Appendix A of [1]). The measurement matrices $A_t$ were zero mean random Gaussian $n_t \times m$ matrices with columns normalized to unit norm. For $t = 0$, $n_0 = 160$; for $t > 0$, $n_t = n = 57$. The measurement noise, $(w_t)_j \sim \text{i.i.d. uniform}(-c_t, c_t)$ for $1 \leq j \leq m$. For $t = 0$, $c_t = 0.01266$; for $t > 0$, $c_t = 0.1266$. We used the same measurement Gaussian matrix $A$ for $t > 0$. We generated 500 realizations respectively with different choices of $m$, and used both algorithms for reconstruction. When $m = 200$, we got, $\zeta_M = 0.9328\sqrt{S_a}, \zeta_L = 0.8734\sqrt{S_a}$; when $m = 1000$, $\zeta_M = 0.8295\sqrt{S_a}, \zeta_L = 0.8628\sqrt{S_a}$; when $m = 2000$, $\zeta_M = 0.8497\sqrt{S_a}, \zeta_L = 0.8628\sqrt{S_a}$.

For our comparison, we pick the largest values we got from the above experiment: let $\zeta_M = 0.9328\sqrt{S_a}$ and $\zeta_L = 0.8734\sqrt{S_a}$. With these values, modified-CS needs LHS of condition 3 $> 13.99\epsilon$ and modified-CS-Add-LS-Del needs LHS of condition 3 $> \max(\alpha_{add} + 7.00\epsilon, 10.978\epsilon + 3.246\alpha_{add}) = 10.978\epsilon + 3.246\alpha_{add}$. With $\alpha_{add}$ small enough, clearly modified-CS-add-LS-del requires a weaker assumption. As explained earlier and also in [1], $\alpha_{add}$ is a small threshold that is typically proportional to the noise bound $c_i, i.e., \epsilon/\sqrt{m}$. Thus the mod-CS-Add-LS-Del condition is weaker.

The comparison between modified-CS and modified-CS-add-LS-del above is not as clear-cut as that in the simple model case (Signal Model 1). The reason is that the simple model tells us exactly how many support additions and removals occur at each time; and it also tells us the exact number of elements with a certain magnitude. As a result, it is possible to get a better bound on $\|x_{\Delta_{a,d}}\|^2$: this is needed to bound the LS step error. The LS error decides the value of $\alpha_{del}$ and $\alpha_{del}$. In turn, decides the lower bound on the LHS of condition 3. The current models Signal Model 2 or 3 are much more flexible, but
between the time that a coefficient began to decrease and when it was removed was to be just all times; and, for decreasing coefficients, it assumed a very specific and fixed rate of magnitude decrease. None of these is a very practical assumption. Our realistic signal change models (Signal Model 2 or 3) do not fix any of these things.

Each time instant (except in the simplest case where all new elements are added at very large magnitudes). Thus, one cannot change at every time which is more realistic, but is also more difficult to analyze. With this, one always has some misses at deletion threshold.

Every vector and computed its 99.9% energy support set. All elements not in this set were set to zero. This gave us an exactly sparse wavelet transform (DWT) domain. A two level db4 wavelet was used there. We computed this 2D DWT, re-arranged it as a most of the changes were visible. As shown in earlier work [13], this sequence is approximately sparse in the 2D discrete wavelet transform (DWT) domain. A two level db4 wavelet was used there. We computed this 2D DWT, re-arranged it as a vector and computed its 99.9% energy support set. All elements not in this set were set to zero. This gave us an exactly sparse sequence \( x_t \). Its dimension \( m = 36^2 = 1296 \). For this sequence, we observed the following. The support size \( N_t \) satisfied \( |N_t| \leq S = 113 \) for all \( t \). The number of additions from \( t - 1 \) to \( t \) satisfied \( |N_t \setminus N_{t-1}| \leq 21 \) and the number of removals, \( |N_{t-1} \setminus N_t| \leq 26 \). Thus, \( S_a = 26 \). Also, the initial nonzero value, \( a_{j,t} \), ranged from 13 to 37, the rate of magnitude increase, \( r_{j,t} \), ranged from 1 to 37, and the duration for which the increase occurred, \( d_{j,t} \), ranged from 0 to 4. Also, the maximum delay between the time that a coefficient began to decrease and when it was removed was \( b = 7 \).

Next we consider a 64x64 functional MRI sequence. fMRI is a technique that is used to investigate brain function. The 10 frame sequence and extracted out a 36x36 region of this sequence selected as the region that includes the part where most of the changes were visible. As shown in earlier work [13], this sequence is approximately sparse in the 2D discrete wavelet transform (DWT) domain. A two level db4 wavelet was used there. We computed this 2D DWT, re-arranged it as a vector and computed its 99.9% energy support set. All elements not in this set were set to zero. This gave us an exactly sparse sequence \( x_t \). Its dimension \( m = 36^2 = 1296 \). For this sequence, we observed the following. The support size \( N_t \) satisfied \( |N_t| \leq S = 113 \) for all \( t \). The number of additions from \( t - 1 \) to \( t \) satisfied \( |N_t \setminus N_{t-1}| \leq 21 \) and the number of removals, \( |N_{t-1} \setminus N_t| \leq 26 \). Thus, \( S_a = 26 \). Also, the initial nonzero value, \( a_{j,t} \), ranged from 13 to 37, the rate of magnitude increase, \( r_{j,t} \), ranged from 1 to 37, and the duration for which the increase occurred, \( d_{j,t} \), ranged from 0 to 4. Also, the maximum delay between the time that a coefficient began to decrease and when it was removed was \( b = 7 \).

Next we consider a 64x64 functional MRI sequence. fMRI is a technique that is used to investigate brain function. The sequence we study here is for the brain responding to a certain type of stimulus (light being turned on and off). This sequence consisted of a rest state brain sequence to which activation was added based on the models suggested in [42]. The goal is to be able to accurately extract out the activation region from this sequence. As is done in [19], one can use the undersampled ReProCS algorithm to extract out the sparse activation regions from the low rank background brain image sequence, as long as an initial background brain training sequence is available. In our example, the activation started at frame 71. For the purpose of ReProCS, the active region “image” (the image that is zero everywhere except in the active region), is the sparse signal of interest. For a 23 pixel region that is known to correspond to the part of the brain that is affected by the above stimulus, the activation was added follows [42]. The 23 pixel region was split into 2 sub-regions so that the activation intensity was
smallest at the boundary of the region and slowly increased as one moved inwards. We show the 2 regions in Fig. 2(b) $R_1$ is the innermost region, $R_2$ is the outermost. The activation in these regions satisfied the following model. For $j \in R_1$, $(x_i)_j = b(t)M_a$. For $j \in R_2$, $(x_i)_j = 0.2b(t)^2M_a$. Here $M_a = 1783$ is the maximum magnitude in the active region and $b(t)$ is the blood oxygenation level dependent (BOLD) signal taken from $[42]$. It is plotted in Fig. 2(a) This image sequence was of size 64x64, i.e. its dimension $m = 64^2 = 4096$. We computed its 99.9% energy support and set all elements not in this set to zero. This gave us our sparse sequence $x_t$. The support size of $x_t$, $N_t$, satisfied $|N_t| \leq S = 23$ for all $t$. The number of additions from $t - 1$ to $t$ satisfied $|N_t \setminus N_{t-1}| \leq S_a = 13$ and the number of removals, $|N_{t-1} \setminus N_t| \leq S_a = 13$. Also, the initial nonzero value, $a_{j,t}$, ranged from 57 to 97, the rate of magnitude increase, $r_{j,t}$, ranged from 1 to 637, and the duration for which the increase occurred, $d_{j,t}$, ranged from 6 to 7. Also, the maximum delay between the time that a coefficient began to decrease and when it was removed was $b = 7$.

![Graph showing BOLD signal and its square](image)

Fig. 2: (a) plot of the BOLD signal and of its square. (b) active, transient and inactive brain regions

VI. Setting Algorithm Parameters and Simulation Results

A. Setting algorithm parameters automatically

Algorithm 1 has one parameter $\alpha$. Algorithm 2 has two parameters $\alpha_{add}$, $\alpha_{del}$. We explain here how to set these thresholds automatically. It is often fair to assume that the noise bound on $\epsilon$ is known, e.g. it can be estimated using a short initial noise-only training sequence. We assume this here. In cases where it is not known or can change with time, one can approximate it by $\|y_{t-1} - A_{t-1}x_{t-1}\|_2$ (assuming accurate recovery at $t - 1$).

Define the minimum nonzero value at time $t$, $x_{min,t} = \min_{j \in N_t} |(x_t)_j|$. This can be estimated as $\hat{x}_{min,t} = \min_{j \in N_t} |(\hat{x}_{t-1})|$

When setting the thresholds automatically, they will change with time. We set $\alpha_{add,t}$ using the following heuristic. By Lemma 3 we have $(x_t - \hat{x}_{t,add})T_{add,t} = (A_{T_{add,t}}A_{T_{add,t}}')^{-1}A_{T_{add,t}}A_{T_{add,t}}'(x_t)A_{T_{add,t}}$. To ensure that this is bounded, we need $\|A_{T_{add,t}}\|\|A_{T_{add,t}}'\|_2$ to be bounded. Since $\|A_{T_{add,t}}\|_2 = \frac{1}{\sigma_{min}(A_{T_{add,t}})}$ and $\|A_{T_{add,t}}'\|_2 = \frac{1}{\sigma_{min}(A_{T_{add,t}})}$, we pick $\alpha_{add,t}$ as smallest number such that $\sigma_{min}(A_{T_{add,t}}) \geq 0.4$.

If one could set $\alpha_{del}$ equal to the lower bound on $x_{min,t} - \|x_t - \hat{x}_{t,add}\|_{T_{add,t}}$, there will be zero misses. Using this idea, we let $\alpha_{del,t}$ be an estimate of the lower bound of this quantity. Notice that

$$
\|x_t - \hat{x}_{t,add}\|_{T_{add,t}} \leq \|A_{T_{add,t}}A_{T_{add,t}}x_t,\Delta_{add} + A_{T_{add,t}}\Delta_{add}\|_2
\leq \|A_{T_{add,t}}A_{T_{add,t}}\|_2\|x_t,\Delta_{add}\|_2 + \|A_{T_{add,t}}\\|_2\|\Delta_{add}\|_2
\approx \|A_{T_{add,t}}A_{T_{add,t}}\|_2\|x_t,\Delta_{add}\|_2 + \|A_{T_{add,t}}\\|_2\|\Delta_{add}\|_2
$$

where $C_1$, $C_2$ are some constant larger than 1. Here we use the fact that for any matrix $B$, $\|B\|_{\infty} \leq C_1\|B\|$ for some constant $C_1$ and that only small elements are missed and hence we can approximate $\|x_t,\Delta_{add}\|_\infty$ by $C_2$ times $\hat{x}_{min,t}$ where $C_2$ is a small constant larger than 1. We cannot compute $\theta_{T_{add,t},|\Delta_{add}|}$, but it is fair to assume that it is small (significantly smaller than one). If we assume that

$$
C_1C_2\|A_{T_{add,t}}A_{T_{add,t}}\|_2\|x_t,\Delta_{add}\|_2 \leq 0.3
$$
then the above bound simplifies to $0.3\hat{x}_{\min,t} + \|A^T_{\text{modcs}} \hat{\omega}_t\|_\infty$. We can approximate $\hat{\omega}_t$ by $y_t - A\hat{x}_{\text{modcs}}$. Thus, we set $\alpha_{\text{del},t} = 0.7\hat{x}_{\min,t} - \|A^T_{\text{modcs}} (y_t - A\hat{x}_{\text{modcs}})\|_\infty$.

For Algorithm 1, we set $\alpha_t$ as follows. If $\|x_t - \hat{x}_{\text{modcs}}\|_\infty \leq C\hat{x}_{\min,t}$ for some $C < 1$, then setting $\alpha_t = (1 - C)\hat{x}_{\min,t}$ will ensure that there are no misses. If this bound holds for most entries $i$, then most entries will be correctly recovered, i.e., there will be few misses. If we ensure $\sigma_{\min}(A_{\hat{T}_t}) \geq 0.4$ then the number of extras will be bounded. To try to ensure that both the above hold, we let $\alpha_t$ to be the smallest value such that $\min_{j \in \hat{T}_t} |(\hat{x}_{\text{modcs}})_j| \geq (1 - C)\hat{x}_{\min,t} = 0.5\hat{x}_{\min,t}$ (we pick $C = 0.5$), and $\sigma_{\min}(A_{\hat{T}_t}) \geq 0.4$.

To get a more robust estimate of the minimum nonzero value of $x_t$, we use a short-time average of $\{\hat{x}_{\min,t}, t - t_0 \leq \tau \leq t\}$ as the estimate of $x_{\min,t}$. In our experiments, $t_0 = 10$.

**B. Simulation Results**

In the discussion so far, we only compared sufficient conditions required by different algorithms. The general conclusion obtained by comparing the sufficient conditions was that modified-CS-add-LS-del is the best algorithm followed by modified-CS and then simple-CS. In this section, we use simulations to demonstrate the same thing. We compared simple CS (solves (4) at each time instant), modified-CS(mod-CS) as given in Algorithm 1 and modified-CS-add-LS-del (mod-CS-Add-LS-Del) as given in Algorithm 2. The parameters for the algorithms were set as explained in Sec VI-A above.

The data was generated as follows. We used Signal Model 2 generated as explained in Appendix C with $m = 200$, $S = 20$, $d_{\min} = 3$, $a_{\min} = r_{\min}(d_{\min}) = r$, $S_0 = 2$, $b = 3$, $\ell = a_{\min} + d_{\min}/r_{\min}(d_{\min}) = 4r$ and $r$ was varied. The measurement matrices $A_t$ were zero mean random Gaussian $n_t \times m$ matrices with columns normalized to unit norm. We used $n_0 = 160$ and $n_t = n = 57$ for $t > 0$. The measurement noise, $(w_t)_j \sim \text{i.i.d. uniform}(-c_1, c_1)$ for $1 \leq j \leq m$. For $t = 0$, $c_1 = 0.01266$; for $t \geq 1$, $c_1 = c = 0.1266$. In the first set of experiments shown in Fig. 3 we used the same measurement matrix $A_t = A$ for all $t \geq 1$. In the second experiment shown in Fig. 4 $A_t$ was time varying.

The normalized mean squared error (NMSE), $\frac{\|x_t - \hat{x}_t\|^2}{\|x_t\|^2}$, the normalized mean extras, $\frac{E[|X_t \setminus \hat{X}_t|]}{E[|X_t|]}$, and the normalized mean misses, $\frac{E[|\hat{X}_t \setminus X_t|]}{E[|X_t|]}$, are used to compare the reconstruction performance. Here $E[.]$ denotes the empirical mean over the 500 realizations. Consider the results of Fig 3. Clearly, both mod-CS and mod-CS-Add-LS-Del significantly outperform simple CS. This is because for $t > 0$, the number of measurements, $n_t = 57$ is too small for a 200 length 20 sparse signal. When $a_{\min} = r_{\min}(d_{\min}) = r$ is large enough, both mod-CS and mod-CS-Add-LS-Del are stable at 5% error or less. When $r$ is reduced, mod-CS becomes unstable. Of course when $r$ is reduced even further to $r = 0.2$, both become unstable (not shown). If Fig 4 we show results for the case when $A_t$ changes with time and all other parameters are the same as Fig 3(a). Clearly in this case, the performance of both mod-CS and mod-CS-add-LS-del has improved significantly.

In Fig. 5 we plot the average value of $\alpha_{\text{add},t}$ for the simulations corresponding to Fig 4. As can be seen, this threshold is close to $4c = 4c/\sqrt{r}$ at all times.

For solving the minimization problems given in (4) and (5), we used the YALL1 software, which is provided in http://yall1blogs.rice.edu/. Both the modified-CS algorithms and simple CS took roughly the same amount of time. For the results of Fig 4 when running the code in MATLAB on the same server, simple CS needed 0.0466 seconds per frame; mod-CS needed 0.0432 seconds per frame and mod-CS-Add-LS-Del needed 0.0517 seconds. These numbers are computed by averaging over all 500 realizations and over the 200 time instants per realization.

![Fig. 3: Error Comparison with Fixed Measurement Matrix](image-url)
In this work we obtained performance guarantees for recursive noisy modified-CS which has been shown in earlier work to be a practically useful algorithm [13], [43], [28]. We show that, under a realistic practically valid signal model and mild assumptions – a lower bound on either the initial nonzero magnitude or the magnitude increase rate; some RIP conditions (which imply conditions on the required number of measurements); appropriately set algorithm parameters; and a special start condition – the support and signal recovery error of modified-CS and its improvement can be bounded by time-invariant and small values.

The special start condition is a possible limitation of our analysis. This can be removed in various ways. If some prior knowledge about signal support is available, that can be used at $t = 0$ as suggested and demonstrated in [13]. Or, one can solve a batch problem (multiple measurement vector (MMV) problem) for the first set of $k$ frames. If we let $\mathcal{N} = \bigcup_{t=1}^{d_0} \mathcal{N}_t$, then we have an MMV problem with row support $\mathcal{N}$ that can be solved using mixed norm minimization [44], simultaneous-OMP [45], [46], compressive MUSIC [47], iterative MUSIC [48], block sparsity approaches [49] or M-SBL (Sparse Bayesian Learning) [50]. In this case one could adopt guarantees for the chosen batch method for the initialization.

In this work, we used a deterministic set of assumptions on signal change. Notice however that one can assume any probabilistic model that ensures that $a_{j,t} \geq a_{\min}$ and $r_{j,\tau}$ is anything larger than $r_{\min}(d_0)$ for the first $d_0$ frames after a new addition; and at later times, $r_{j,\tau}$ can be anything between zero and infinity. Similarly, any probabilistic model for coefficient decrease that ensures removal within at most $b$ frames after decrease begins will suffice. We can fix $d_0$ to be any integer between zero and $d_{\min}$ and our result will then hold for that particular value of $d_0$.

Other ongoing and future work includes designing and analyzing better support prediction techniques rather than just using the previous support estimate as the prediction for the current support. Some initial ideas are presented in [51].

**APPENDIX**

A. **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 $h := \hat{x}_{\text{modcs}} - x$. We adapt the approach of [10] to bound the reconstruction error, $\|h\| := \|\hat{x}_{\text{modcs}} - x\|$. A similar result was obtained in [27].

Let $\Delta_1$ denote the set of indices of $h$ with the $|\Delta|$ largest values outside of $\mathcal{T} \cup \Delta$, let $\Delta_2$ denote the indices of the next $|\Delta|
Combining this with (20), and using the definition of $h_{\Delta, t}$, we get

$$\|h_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}\| \leq \sum_{j \geq 2} \|h_{\Delta_j}\| \leq \frac{1}{\sqrt{|\Delta|}} \|h_{(\mathcal{T} \cup \Delta)^c}\|_1$$

(20)

Since $\hat{x}_{\text{modcs}} = x + h$ is the minimizer of (5) and since both $x$ and $\hat{x}_{\text{modcs}}$ are feasible; and since $x$ is supported on $\mathcal{N} \subseteq \mathcal{T} \cup \Delta$,

$$\|x\|_1 = \|x_{\mathcal{T}}\|_1 \geq \||x + h\|_{\mathcal{T}^c}\|_1$$

$$\geq \|x\|_1 - \|h\|_1 + \|h_{(\mathcal{T} \cup \Delta)^c}\|_1$$

(21)

Thus,

$$\|h_{(\mathcal{T} \cup \Delta)^c}\|_1 \leq \|h\|_1$$

(22)

Combining this with (20), and using $\frac{\|h\|}{\sqrt{|\Delta|}} \leq \|h\|$, we get

$$\|h_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}\| \leq \sum_{j \geq 2} \|h_{\Delta_j}\| \leq \|h\|$$

(23)

Next, since both $x$ and $\hat{x}_{\text{modcs}}$ are feasible,

$$\|Ah\| = \|A(x - \hat{x}_{\text{modcs}})\|$$

$$\leq \|y - Ax\| + \|y - A\hat{x}_{\text{modcs}}\| \leq 2\epsilon$$

(24)

In this proof, let

$$\delta \triangleq \delta_{|\mathcal{T}| + 3|\Delta|}$$

(25)

Now, we upper bound $\|h_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}\|$. By $\delta_{|\mathcal{T}| + 2|\Delta|} \leq \delta$, we have

$$(1 - \delta)\|h_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}\| \leq \|Ah_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}\|$$

(26)

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

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

Using (24) and the definition of $\delta_S$ given in (3) and $\delta_{|\mathcal{T}| + 2|\Delta|} \leq \delta$,

$$|\langle Ah_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}, Ah \rangle| \leq 2\epsilon \sqrt{1 + \delta} \|h_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}\|$$

(27)

Using the definition of $\theta_{S_1, S_2}$ given in (3); equation (23); and the fact that $\|h_{\mathcal{T}}\| + \|h_{\Delta \cup \Delta_1}\| \leq \sqrt{2}\|h_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}\|$, we get the following. Using $\theta_{|\mathcal{T}|, |\Delta|} \leq \delta_{|\mathcal{T}| + |\Delta|} \leq \delta_{|\mathcal{T}| + 3|\Delta|}$, $\|Ah\| \leq \|h\|_1 + \|h_{\Delta \cup \Delta_1}\| \leq \delta_{|\mathcal{T}| + 3|\Delta|} \leq \delta_{|\mathcal{T}| + 3|\Delta|}$, we get

$$\sum_{j \geq 2} \langle Ah_{\Delta_j}, Ah \rangle$$

$$\leq \sum_{j \geq 2} \|h_{\Delta_j}\|$$

$$\leq \delta \|h_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}\| \|h\|$$

(28)

Combining the last six equations above, using $\|h\| \leq \|h_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}\|$, we can simplify the above to get

$$\|h\| \leq 2\|h_{(\mathcal{T} \cup \Delta \cup \Delta_1)^c}\| \leq \frac{4\sqrt{1 + \delta}}{1 - 2\delta} \epsilon$$

(29)

Clearly, all of the above discussion holds only if the RHS is positive which is true only if $2\delta_{|\mathcal{T}| + 3|\Delta|} < 1$. Thus, we can get Lemma [1]
B. Proof of Theorem 7

We prove the first claim 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., $|\Delta_{t-1}| = 0$, $|T_{t-1}| \leq S$, and $\Delta_{t-1} \subseteq S_{t-1}(d_0)$ so that $|\Delta_{t-1}| \leq 2(d_0-1)S_a$. 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 $A_t \subseteq N_t \cap \bigcap_{i \in A_t} R_i$, (b) $N_t = N_{t-1} \cup A_t \setminus R_t$, and (c) if two sets $B, C$ are disjoint, then, $D \cup C \setminus B := (D \cup C) \setminus B = (D \cap B') \cup C$ for any set $D$.

We first bound $|T_t|, |\Delta_{t,t}|, |\Delta_t|$. Since $T_t = \tilde{T}_{t-1} = \tilde{N}_{t-1}$, so $|T_t| \leq S$. Thus, $\Delta_{t,t} = \tilde{N}_{t-1} \cap (N_{t-1} \cap \bigcap_{i \in A_t} R_i) \subseteq \tilde{N}_{t-1} \cap (N_{t-1} \cap \bigcap_{i \in A_t} R_i) \subseteq \tilde{N}_{t-1} \cap R_t = \tilde{N}_{t-1} \cap R_t$. The last equality follows since $|\Delta_{t-1}| = 0$. Thus $|\Delta_{t,t}| \leq |R_t| = S_a$.

Consider $|\Delta_t|$. Notice that $\Delta_t = N_t \setminus \tilde{N}_{t-1} = (N_{t-1} \cap \tilde{N}_{t-1} \cap \bigcap_{i \in A_t} R_i) \subseteq (N_{t-1} \cap \bigcap_{i \in A_t} R_i) \subseteq (S_{t-1}(d_0) \cap \bigcup_{i \in A_t} R_i) = S_{t-1}(d_0) \cup A_t \cap R_t \subseteq S_{t-1}(d_0) \cup A_t \cap R_t$. Here we used $\Delta_{t-1} \subseteq S_{t-1}(d_0)$.

We next bound $|\tilde{T_t}|, |\Delta_{t,t}|, |\tilde{T_t}|$. Consider the support estimation step. Apply the first claim of Lemma 4 with $S_N = S$, $S_{\Delta t} = S_a$, $S_{\Delta s} = S_a$, and $b_1 = d_0r$. Since conditions 2 and 1 of the theorem hold, all elements of $N_t$ with magnitude equal or greater than $d_0r$ will get detected. Thus, $|\tilde{T_t}| \leq |T_t| + |\Delta_{t,t}| \leq S + S_a$.

The second claim for time $t$ follows using the first claim for time $t - 1$ and the arguments from the paras above. The third claim follows using the second claim and Lemma 1.

C. 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., $|\Delta_{t-1}| = 0$, $|T_{t-1}| \leq S$, and $\Delta_{t-1} \subseteq S_{t-1}(d_0)$ so that $|\Delta_{t-1}| \leq 2(d_0-1)S_a$. Using this, we prove that the claim holds at $t$. In the proof, we use the following facts often: (a) $R_t \subseteq N_{t-1}$, (b) $A_t \subseteq N_t \cap \bigcap_{i \in A_t} R_i$, (c) $N_t = N_{t-1} \cup A_t \setminus R_t$, and (d) if two sets $B, C$ are disjoint, then, $D \cup C \setminus B := (D \cup C) \setminus B = (D \cap B') \cup C$ for any set $D$.

The bounding of $|T_t|, |\Delta_t|, |\Delta_{t,t}|$ is exactly as in the proof of Theorem 1. Since $T_t = \tilde{T}_{t-1}$, so $|T_t| \leq S$. Thus, $\Delta_{t,t} = \tilde{N}_{t-1} \setminus N_t = \tilde{N}_{t-1} \setminus (N_{t-1} \cap \bigcap_{i \in A_t} R_i) \subseteq \tilde{N}_{t-1} \setminus \bigcap_{i \in A_t} R_i = R_t$. Thus $|\Delta_{t,t}| \leq |R_t| = S_a$.

Finally, $\Delta_t = N_t \setminus \tilde{N}_{t-1} = (\tilde{N}_{t-1} \cap \bigcap_{i \in A_t} R_i) \subseteq (S_{t-1}(d_0) \cap \bigcap_{i \in A_t} R_i) = S_{t-1}(d_0) \cup A_t \cap R_t$.

$$\Delta_t \subseteq S_{t-1}(d_0) \cup A_t \cap R_t \tag{30}$$

When $d_0 \geq 2$, $R_t \subseteq S_{t-1}(d_0)$ and $A_t$ is disjoint with $S_{t-1}(d_0)$, so $|\Delta_t| \leq |S_{t-1}(d_0)| + |A_t| - |R_t| = 2(d_0-1)S_a + S_a - S_a$. When $d_0 = 1$, $S_{t-1}(d_0) = \emptyset$, and $A_t$ is disjoint with $R_t$, so $|\Delta_t| \leq |A_t| - |R_t| = |A_t| = S_a$. Thus, $|\Delta_t| \leq k_1S_a$.

Consider the detection step. There are at most $f$ false detects (from condition 11 and thus $|\Delta_{t,t}| \leq |\Delta_{t,t}| - f \leq S_a + f$.

Thus $|T_{t+1}| \leq |T_t| + |\Delta_{add,t}| \leq S + S_a + f$.

Next, consider $|\Delta_{add,t}|$. Notice that

$$\Delta_t \subseteq S_{t-1}(d_0) \cup A_t \setminus R_t \subseteq S_t(d_0) \cup \mathcal{I}_t(d_0) \setminus D_t(d_0-1). \tag{31}$$

The first $\subseteq$ is from (30), the second one follows by using (12) for $j = d_0$. Now, apply Lemma 5 with $S_{N_t} = S$, $S_{\Delta_t} = S_a$, $S_{\Delta s} = S_a$, $S_{\Delta s} = k_1S_a$, and $b_1 = d_0r$. Using (31), $|\Delta_t| \leq |\Delta_{add,t}| \leq \Delta_t \setminus \mathcal{I}_t(d_0)$. Since conditions 2 and 1 hold, by Lemma 5 all elements of $\{i \in \Delta_t : |x_i| \geq b_1\}$ will definitely get detected at time $t$. Thus $|\Delta_{add,t}| \subseteq \Delta_t \setminus \mathcal{I}_t(d_0)$. From (31), $|\Delta_t| \leq |\mathcal{I}_t(d_0) \setminus D_t(d_0-1)| = 2(d_0-1)S_a - S_a$, when $d_0 = 1$, $D_t(d_0-1) = S_t(d_0) = \emptyset$, then $|\Delta_{add,t}| = 0$. Thus $|\Delta_{add,t}| \leq k_2S_a$.

Consider the deletion step. Apply Lemma 6 with $S_{\Delta s,t} = S_a$, $S_{\Delta s,t} = k_2S_a$. Since condition 26 holds, $\delta_{S + S_a + f} < 1/2$ holds. Since $\Delta_{add,t} \subseteq \Delta_t \setminus \mathcal{I}_t(d_0) \setminus D_t(d_0-1)$, $\Delta_{add,t}$ contains only $2S_a$ elements of magnitude $\{r, 2r, \ldots, (d_0 - 2)r\}$ and $S_a$ elements of magnitude $(d_0 - 1)r$. Thus, $||x_t||_{\Delta_{add,t}} \leq k_3\sqrt{S_a}$. Using these facts and condition 11 by Lemma 6 all elements of $\Delta_{add,t}$ will get deleted. Thus $|\Delta_{add,t}| = 0$. Thus $|T_t| \leq |N_t| + |\Delta_{t,t}| \leq S$.

To bound $|\Delta_t|$, apply Lemma 6 with $S_{\Delta s,t} = S + S_a + f$, $S_{\Delta s,t} = k_2S_a$, $b_1 = d_0r$. By Lemma 6 to ensure that all elements of $\{i \in \mathcal{T}_{add,t} : |x_i| \geq b_1\}$ do not get falsely deleted, we need $\delta_{S_0 + S_a + f} < 1/2$ and $d_0r > \alpha_{del} + \frac{\sqrt{2}\epsilon}{\sqrt{S_a}}$. From condition 11 $\alpha_{del} = \frac{\sqrt{2}\epsilon}{\sqrt{S_a}} + k_3S_a \eta_{\epsilon}$. Thus, we need $\delta_{S_0 + S_a + f} < 1/2$ and
Recall that $d_0 r > 2(\sqrt{\frac{d_0}{S_0}} \zeta \ell + 2k_3 \theta S_d + S_a + f, k_2 S_a \zeta r), \delta_{S_0 + S_a + f} < 1/2$ holds since condition $22$ holds. The second one holds since condition $22$ and $r \geq G_2$ of condition $3$ hold. Thus, we can ensure that all elements of $\{ i \in T_{\text{add}, t} : |(x_i)_t| \geq b_1 \}$, i.e. all elements of $T_{\text{add}, t}$ with magnitude greater than or equal to $b_1 = d_0 r$ do not get falsely deleted. But nothing can be said about the elements smaller than $d_0 r$ (in the worst case all of them may get falsely deleted). Thus, $\Delta_t \subseteq \mathcal{S}(d_0)$ and so $|\Delta_t| \leq 2(d_0 - 1)S_a$.

This finishes the proof of the first claim. To prove the second and third claims for any $t > 0$: use the first claim for $t - 1$ and the arguments from the paragraphs above to show that the second and third claim hold for $t$. The fourth claim follows using the previous claims and Lemma $1$. The fifth claim follows using previous claims, Lemma $3$ and a bound on $\| (x_t)_{\Delta_t} \|_2$.

It is easy to see that $\| (x_t)_{\Delta_t} \|_2 \leq k_3 \sqrt{S_0 r}$.

D. Proof of Theorem $2$

Recall from the signal model that $|\mathcal{N}_t| \leq S$ for all $t$, and that $|SD_t| \leq \frac{(b+1)}{2} S_d$. Also $\mathcal{N}_t = \cup_{t = t - d_{\text{min}} + 1}^t A_t \cup \cup_{t} \cup SD_t$, noting that the first two sets might not be disjoint.

The proof follows using induction. The base case is easy. Assume that the result holds at $t - 1$. At $t$, at most $S_a$ new elements get added to the support, thus $|\Delta_t| \leq |\Delta_{t-1}| + S_a \leq \frac{(b+1)}{2} S_d + d_0 S_a + S_a$. Also, since $T_t = \bar{T}_{t-1}$, thus $|T_t| \leq S$. And $\Delta_{t-1}(t) = \Delta_{t-1}(t) \cup R_t$, indicating $|\Delta_{t-1}(t)| \leq |\Delta_{t-1}(t)| + |R_t| \leq S'$. The second condition of the theorem ensures that $\delta_{|T_t| + 1/2} |\Delta_t| \leq \sqrt{2 - 1}/2$. Thus using Lemma $1$, $|x_t - \hat{x}_t| \leq S_0 r$.

Consider the support detection step. Consider an $i \notin \mathcal{N}_t$, i.e. $(x_i)_t = 0$. Since $\alpha = \frac{\sqrt{M}}{\sqrt{S_a}} 7.50\epsilon \geq \frac{\sqrt{r}}{\sqrt{S_a}} |x_t - \hat{x}_t| \geq \|x_t - \hat{x}_t\|_\infty \geq \|x_t - \hat{x}_t\|_\infty$, thus $i$ will never get detected into the support estimate. Thus, $|\Delta_{t, t}| = 0$. Thus $|T_t| = |\mathcal{N}_t| + |\Delta_{t, t}| \leq S$.

The third condition ensures that any newly added element exceeds $\alpha + \frac{\sqrt{M}}{\sqrt{S_a}} 7.50\epsilon$ within $d_0$ time units and any element of $\mathcal{L}_t$ exceeds $\alpha + \frac{\sqrt{M}}{\sqrt{S_a}} 7.50\epsilon$ as $\ell > \alpha + \frac{\sqrt{M}}{\sqrt{S_a}} 7.50\epsilon$. Consider any such element $j$. This means that $|\hat{x}_t|_j \geq |(x_t)_j| - |(x_t - \hat{x}_t)| \geq |(x_t)_j| - \frac{\sqrt{M}}{\sqrt{S_a}} |x_t - \hat{x}_t| \geq |(x_t)_j| - \frac{\sqrt{M}}{\sqrt{S_a}} 7.50\epsilon \geq \alpha$. Thus such an element will definitely get detected into the support. This means that the only nonzero elements that are missed are those that got added in the last $d_0$ frames or those that are currently decreasing. The maximum number of elements that got added in the last $d_0$ time units is $d_0 S_a$. The maximum number of decreasing elements at $t$ is less than or equal to $\frac{b+1}{2} S_d$. Thus, $|\Delta_t| \leq \frac{b+1}{2} S_d + d_0 S_a$. This finishes the proof of the induction step and hence of the theorem.

E. Proof of Theorem $2$

**Proposition 2 (simple facts):** Consider Algorithm $2$

1) An $i \in \mathcal{N}_t$ will definitely get detected if $|(x_t)_i| > \alpha_{\text{del}} + \frac{\sqrt{M}}{\sqrt{S_a}} \|x_t - \hat{x}_t, \text{modcs}\|_\infty$.

2) An $i \notin \mathcal{N}_t$ will definitely not be deleted if $|(x_t)_i| > \alpha_{\text{del}} + \frac{\sqrt{M}}{\sqrt{S_a}} \|x_t - \hat{x}_t, \text{add}\|_\infty$.

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

Recall from the signal model that $\mathcal{N}_t = \cup_{t = t - d_{\text{min}} + 1}^t A_t \cup \cup_{t} \cup SD_t$, noting that the first two sets might not be disjoint. By the induction assumption, $|\bar{T}_{t-1}| \leq S$. Since $\bar{T}_t = \bar{T}_{t-1} = \hat{N}_{t-1}$, thus,

$$|\bar{T}_{t-1}| \leq S \tag{32}$$

Also, by the induction assumption,

$$\Delta_{t-1} \subseteq SD_{t-1} \cup A_{t-1} \cdots A_{t-d_0} \tag{33}$$

Recall that $\mathcal{N}_t = N_{t-1} \cup A_t \setminus R_t$. Also, $SD_t \subseteq SD_t \cup R_t$. Thus, $SD_{t-1} \cap R_t \subseteq SD_t$. Thus,

$$\Delta_t = N_t \cap \hat{N}_{t-1} = (N_t \cap N_{t-1} \cap \hat{N}_{t-1}) \cup (A_t \cap \hat{N}_{t-1}) \subseteq (\Delta_{t-1} \cap R_t) \cup A_t \subseteq SD_t \cup A_{t-1} \cdots A_{t-d_0} \cup A_t \tag{34}$$

Thus,

$$|\Delta_t| \leq \frac{b+1}{2} S_a + d_0 S_a + S_a \tag{35}$$

Using the above bounds on $|\bar{T}_t|$ and $|\Delta_t|$ and the RIP condition of the theorem, we can apply Lemma $1$ to show that

$$\|x_t - \hat{x}_t, \text{modcs}\|_\infty \leq 7.50\epsilon \tag{36}$$
Thus, using the Proposition \[2\] and condition 3, all elements of \(A_{t-d_0}\) are definitely detected in the add step at \(t\), i.e.
\[
A_{t-d_0} \subseteq \hat{A}_t
\] (37)
Also since \(\ell\) satisfies condition 3, all elements of \(L_{\ell}\) will be detected in the add step at \(t\).

Using \[37\],
\[
\Delta_{\text{add}, t} = \Delta_t \setminus \hat{A}_t = SD_t \cup A_t \cup A_{t-1} \cdots \cup A_{t-d_0} \setminus \hat{A}_t \\
\subseteq SD_t \cup A_t \cup A_{t-1} \cdots \cup A_{t-d_0+1}
\] (38)
Thus,
\[
|\Delta_{\text{add}, t}| \leq \frac{(b+1)}{2} S_a + d_0 S_a
\] (39)
Also, \(T_{\text{add}, t} \subseteq N_{\ell} \cup \Delta_{\text{e, add}, t}\) and
\[
\Delta_{\text{e, add}, t} = \Delta_{\text{e}, t} \cup (\hat{A}_t \setminus N_t) \subseteq \hat{\Delta}_{\text{e}, t-1} \cup R_t \cup (\hat{A}_t \setminus N_t)
\] (40)
Thus, \(|\Delta_{\text{e, add}, t}| \leq S_a + f\) and so
\[
|T_{\text{add}, t}| \leq S + |\Delta_{\text{e, add}, t}| \leq S + S_a + f
\] (41)
By Lemma \[3\] and condition 2c of the Theorem, we have
\[
\|(x_t - \hat{x}_{t, \text{add}})\| \leq 1.12\epsilon + (1 + 1.261\theta |T_{\text{add}, t}|)|\Delta_{\text{add}, t}|\| (x_t)_{\text{add}, t}\|
\leq 1.12\epsilon + 1.261 \| (x_t)_{\text{add}, t}\|
\] (42)
Recall that, by Proposition \[2\] any element of \(x_{\Delta_{\text{add}, t}}\) will have magnitude smaller than \(\alpha_{\text{add}} + \frac{\zeta M}{\sqrt{S_a}} 7.50\epsilon\). By \[39\], we have
\[
\|x_{\Delta_{\text{add}, t}}\| \leq \sqrt{|\Delta_{\text{add}, t}|(\alpha_{\text{add}} + \frac{\zeta M}{\sqrt{S_a}} 7.50\epsilon)}
\leq \sqrt{(\frac{(b+1)}{2} S_a + d_0 S_a)(\alpha_{\text{add}} + \frac{\zeta M}{\sqrt{S_a}} 7.50\epsilon)}
\] (43)
Let \(h = \sqrt{(\frac{(b+1)}{2} + d_0)(\alpha_{\text{add}} + \frac{\zeta M}{\sqrt{S_a}} 7.50\epsilon)}\). Combining this with the bound on \(|T_{\text{add}, t}|\) and \(|\Delta_{\text{add}, t}|\) we can bound the LS step error by a time-invariant quantity,
\[
\|(x_t - \hat{x}_{t, \text{add}})_{\text{T}_{\text{add}, t}}\| \leq 1.12\epsilon + 1.261 h \sqrt{S_a}
\] (44)
Using Assumption \[2\] we have,
\[
\|(x_t - \hat{x}_{t, \text{add}})_{\text{T}_{\text{add}, t}}\|_{\infty} \leq 1.12 \frac{\zeta L}{\sqrt{S_a}} + 0.261 \zeta_L h
\] (45)
Using the fact that \(\alpha_{\text{det}}\) is equal to the RHS of the above equation and proposition fact 3, if \((x_t)_j = 0\), then \(j \in \hat{R}_t\). Thus,
\[
N_{\ell}^c \subseteq \hat{R}_t
\] (46)
Next, using \[19, 45\], fact 2 of Proposition \[2\] and the value of \(\alpha_{\text{det}}\), we can conclude the following: if \(j \in L_{\ell}, j\) will not get falsely deleted; the same is true if \(j \in A_{\tau}, \tau \leq t - d_0\). Thus,
\[
\hat{R}_t \subseteq N_{\ell}^c \cup SD_t \cup A_t \cup A_{t-1} \cdots \cup A_{t-d_0+1}
\] (47)
Recall that \(\hat{N}_t = \hat{N}_{t-1} \cup \hat{A}_t \setminus \hat{R}_t\). Thus
\[
\Delta_t = N_t \setminus \hat{N}_t = (N_t \cap \hat{N}_{t-1} \cap \hat{A}_t) \cup (N_t \setminus \hat{R}_t)
\subseteq (\Delta_t \cap \hat{A}_t^c) \cup (SD_t \cup A_t \cup A_{t-1} \cdots A_{t-d_0+1})
\] (48)
Since \(A_{t-d_0} \subset \hat{A}_t\), using \[34\], we get
\[
\Delta_t \cap \hat{A}_t^c \subseteq SD_t \cup A_t \cup A_{t-1} \cdots \cup A_{t-d_0+1}
\] (49)
Thus, using \[48\],
\[
\hat{\Delta}_t \subseteq SD_t \cup A_t \cup A_{t-1} \cdots \cup A_{t-d_0+1}
\] (50)
Thus,
\[
|\tilde{\Delta}_t| \leq \frac{(b+1)}{2} S_a + d_0 S_a
\]  
(51)

Now consider $\tilde{\Delta}_{e,t}$.
\[
\tilde{\Delta}_{e,t} = \tilde{\mathcal{N}}_t \setminus \mathcal{N}_t
= (\mathcal{N}_{t-1} \cap \tilde{\mathcal{R}}_t \cap \mathcal{N}_t^c) \cup (\tilde{\mathcal{A}}_t \cap \tilde{\mathcal{R}}_t \cap \mathcal{N}_t^c)
\]

As $\mathcal{N}_t^c \subseteq \tilde{\mathcal{R}}_t$, we have $\tilde{\mathcal{R}}_t \subseteq \mathcal{N}_t$. Thus,
\[
\tilde{\Delta}_{e,t} = \emptyset
\]  
(52)

Thus,
\[
|\tilde{\Delta}_{e,t}| = 0
\]  
(53)

Since $|\mathcal{N}_t| \leq S$ and since $|\tilde{\mathcal{T}}_t| \leq |\mathcal{N}_t| + |\tilde{\Delta}_{e,t}|$, thus
\[
|\tilde{\mathcal{T}}_t| \leq S
\]  
(54)

By condition 2,
\[
\theta_{|\tilde{\mathcal{T}}_t|, |\tilde{\Delta}_t|} \leq \theta_{S, S + d_0 S_a + d_0 S_a}
\leq \delta_{S + 3 S + d_0 S_a + S_a} \leq 0.207
\]  
(55)

and
\[
\delta_{|\tilde{\mathcal{T}}_t|} \leq \delta_S \leq \delta_{S + d_0 S_a + f} \leq 0.207
\]

Using the same way as getting $\| (x_t - \hat{x}_{t, \text{add}}) \|$, we have
\[
\| (x_t - \hat{x}_t) \| \leq 1.12 \epsilon + 1.261 \| x_{\tilde{\Delta}_t} \|
\]

Also, using Proposition 2 any element of $x_{\tilde{\Delta}_t}$ will have magnitude smaller than $\alpha_{\text{del}} + 1.12 \frac{\epsilon}{\sqrt{S_a}}$. By (51), we have
\[
\| x_{\tilde{\Delta}_t} \| \leq \sqrt{\frac{(b+1)}{2} S_a + d_0 S_a} (\alpha_{\text{del}} + 1.12 \frac{\epsilon}{\sqrt{S_a}})
\]

Thus, the final claim is proved.

\(F. \text{ Proof of Remark 11: Necessary conditions}\)

Consider the noise-free case, i.e. $\epsilon = 0$ and Algorithm 1. We claim that left-RIP$(S + S_a)$ at all times $t > 0$ is necessary to ensure exact recovery of all sparse signal sequences with support size at most $S$, and number of support additions and removals at most $S_a$. We prove this here. Assume exact recovery at $t - 1$. Assume also that the support size at $t - 1$ is $S$, there are $S_a$ new additions and $S_a$ new removals at time $t$. Thus support size at time $t$ is also $S$.

Suppose that left-RIP$(S + S_a)$ does not hold. This means there is a set, $R$, of size $S + S_a$ for which $\text{rank}((A_t)_{R}) < S + S_a$. Pick a $z$ so that $z_R \in \text{null}((A_t)_{R})$ (i.e. $(A_t)_{R} z_R = 0$) and $z_R = 0$. Partition $R$ into three sets $R = D \cup D_1 \cup D_2$ s.t. all are disjoint; $|D| = S - S_a$, $|D_1| = S_a = |D_2|$ and $\| z_{D_2} \|_1 \leq \| z_{D_1} \|_1$. Create two sparse vectors $x^1$ and $x^2$ supported on $D \cup D_1$ and $D \cup D_2$ respectively as follows. Let $(x^1)_D = z_D/2$, $(x^1)_{D_1} = z_{D_1}$, $(x^1)_{(D \cup D_1)}^c = 0$. Let $(x^2)_D = -z_D/2$, $(x^2)_{D_2} = z_{D_2}$, $(x^2)_{(D \cup D_2)}^c = 0$. Then both $x^1$ and $x^2$ have support size $S$.

Suppose that the signal at time $t$ is $x^1$, i.e. $x_t = x^1$ so that $y_t = A_t x^1$, and suppose that the support (equal to support estimate) from $t - 1$ is $T = D \cup \Delta_t$, where $\Delta_t$ is a subset of $(D \cup D_1 \cup D_2)^c$ of size $S_a$. Consider the solution of modified-CS with $\epsilon = 0$. In this case, both $x^1$ and $x^2$ are feasible since $A_t(x^1 - x^2) = (A_t)_{D} z_D/2 + (A_t)_{D_1} z_{D_1} - (A_t)_{D} (-z_D/2) - (A_t)_{D_2} (-z_{D_2}) = (A_t)_{R} z_R$. But, $\| (x^1)_{D^c} \| = \| (x^2)_{D} \|_1 = \| z_{D_1} \|_1 \geq \| z_{D_2} \|_1 = \| (x^2)_{D^c} \|$. Thus, clearly $x^1$ will not be the unique solution to modified-CS with $\epsilon = 0$. This proves that left-RIP$(S + S_a)$ is necessary.
G. Generative model for Signal Model

This model requires that when a new element \( j \) gets added to the support, its magnitude keeps increasing at rate \( r_{j,t} \) until it reaches large set, and that an element \( i \) of the large set starts to decrease at rate \( r_{i,t} \) until it reaches 0. 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. We can choose values for \( a_{\min}, d_{\min}, r_{\min} (d_{\min}), S_a, m, b \) during simulation.

Mathematically, it can be described as follows. Let \((x_t) = (M_t)j(s_t)j\) where \((M_t)j\) denotes the magnitude and \((s_t)j\) denotes the sign of \((x_t)\) at time \( t \). \( x_t \) is a \( m \times 1 \) vector; \( S_0 = [\mu_1 S] \), here \( \mu_1 \) is a random number between 0.9 and 1.

For \( 1 \leq t \leq b \), let \( S_{a,t} = 0 \), \( S_{r,t} = 0 \), \( S_{d,t} = S_a \); For any \( t > b \), do the following.

1) Generate
   a) the new addition set, \( A_t \), of size \( S_{a,t} = [\mu_2 (\Sigma_{r=t}^{t-1} S_{d,r} - \Sigma_{r=t}^{t-1} S_{a,r})] \) (here \( \mu_2 \) is a random number between 0.9 and 1) uniformly at random from \( N_t^{-c} \),
   b) the new decreasing set, \( B_t \), of size \( S_{d,t} = [\mu_3 S_a] \) (here \( \mu_3 \) is a random number between 0.5 and 1) uniformly at random from \( L_{t-1} \), and
   c) the new deleted set, \( R_t \), of size \( S_{r,t} = [\mu_4 |SD_{t-1}] \) (here \( \mu_4 \) is a random number between 0.1 and 0.3), as the smallest \( S_{r,t} \) elements of \( SD_{t-1} \).

2) Update the coefficients’ magnitudes as follows.
   \[
   \begin{align*}
   (M_t)i &= \begin{cases}
   (M_{t-1})i + r_{i,t}, & i \in A_{t-d_{\min}} \cup L_{t-1} \setminus B_t, r_{i,t} = \mu_5; \\
   (M_{t-1})i + r_{i,t}, & i \in A_{t-d_{\min}} \cup L_{t-1} \setminus B_t, r_{i,t} = \mu_6 r_{\min}(d_{\min}); \\
   (M_{t-1})i - r_{i,t}, & i \in SD_{t-1} \setminus R_t, r_{i,t} = \mu_7 f_i; \\
   (M_{t-1})i - r_{i,t}, & i \in B_t, r_{i,t} = \mu_8 (M_{t-1} - \ell); \\
   0, & i \in N_t^c.
   \end{cases}
   \end{align*}
   \]
   where \( \mu_6, \mu_7 \) and \( \mu_8 \) are random numbers between 1 and 1.44; \( \mu_5 \) is a random number larger than \( -((M_{t-1})_i - \ell) \).

3) Update the signs as follows.
   \[
   (s_t)i = \begin{cases}
   (s_{t-1})i, & i \in N_t^c \setminus A_t \\
   iid(\pm 1), & i \in A_t \\
   0, & i \in N_t^c
   \end{cases}
   \]
   where \( iid(\pm 1) \) refers to generating the sign as +1 or -1 with equal probability and doing this independently for each element \( i \).

4) Set \((x_t)i = (M_t)i(s_t)i\) for all \( i \).

5) Update
   \[
   \begin{align*}
   L_t &= A_{t-d_{\min}} \cup L_{t-1} \setminus B_t, \\
   SD_t &= SD_{t-1} \cup B_t \setminus R_t.
   \end{align*}
   \]

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 focuss: 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 App. 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] N. Vaswani, “Kalman filtered compressed sensing,” in IEEE Intl. Conf. Image Proc. (ICIP), 2008.

[12] 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.
