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

Jinchun Zhan and Namrata Vaswani
Dep. of Electrical & Computer Eng., Iowa State University, Ames, Iowa, USA
Email: {jzhan,namrata}@iastate.edu

***Abstract*—*"THIS PAPER IS ELIGIBLE FOR THE STUDENT PAPER AWARD"*. In this work, we obtain performance guarantees for modified-CS and for its improved version, modified-CS-Add-LS-Del, for recursive reconstruction of sparse signal sequences from noisy measurements. Under mild assumptions, and for a realistic signal change model, 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, our results hold under weaker assumptions on the number of measurements than what simple compressive sensing (basis pursuit denoising) needs. Also, the result for modified-CS-add-LS-del holds under weaker assumptions on the signal magnitude increase rate than the result for modified-CS. Similar results were obtained in an earlier work, however the signal change model assumed there was very simple and not practically valid.**

I. Introduction

Starting with the seminal papers of Candes et al and Donoho [1], [2] there has been a large amount of recent work on sparse recovery/ compressed sensing (CS). Since 2008, the problem of recursively recovering a time sequence of sparse signals, with slowly changing sparsity patterns has also been extensively studied [3], [4], [5], [6], [7], [8], [9], [10]. In [7], the authors study a multiple measurement vector (MMV) version of the recursive recovery problem and obtains conditions under which the support of the sparse signals can be exactly tracked over time in the noise-free case.

A key assumption introduced in [3] and empirically verified in [4], is that for many natural signal/image sequences, the sparsity pattern (support in the sparsity basis) changes slowly over time. In [6], 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. Other related ideas for support recovery with prior knowledge about the support entries, that appeared in parallel, include [11], [12].

Error bounds for modified-CS for noisy measurements were obtained in [13] and [14]. However, when modified-CS is used for recursive reconstruction, the most important question is, under what conditions can we obtain time-invariant error bounds, i.e. show error stability over time? In [14], we first answered this question for modified-CS and for an improved version of modified-CS which we called "modified-CS with add-LS-del". However, the signal model assumed in [14] was highly simplified. For example, it assumed that the magnitude of a newly added coefficient to the support increased at the exact same rate at all times and for all new coefficients. A similar assumption was made for the magnitude to decrease before it got removed from the support. For typical sequences, neither of these assumptions holds in practice.

Contribution. In this work, we obtain conditions for error stability of modified-CS and modified-CS-Add-LS-Del for a realistic signal change model that allows different rates of magnitude increase and decrease at different times and for different coefficients. Unlike [14], it also allows different numbers of coefficients to get added or removed at different times. We verify that our model is indeed valid for MRI image sequences. For the above signal change model, under mild assumptions (enough number of measurements and large enough initial magnitude or large enough rate of magnitude increase) 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, we argue that our results hold under weaker assumptions on the number of measurements than what simple compressive sensing (basis pursuit denoising) needs. Also, the result for modified-CS-add-LS-del holds under weaker assumptions on the signal magnitude increase rate than the result for modified-CS.

A. Notation and Problem Definition

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

For a vector, v, and a set, $\mathcal{T}$, $v_\mathcal{T}$ denotes the $|\mathcal{T}|$ length sub-vector containing the elements of v corresponding to the indices in the set $\mathcal{T}$. $\|v\|_k$ denotes the ℓ_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, $M_\mathcal{T}$ denotes the sub-matrix obtained by extracting the columns of M corresponding to indices in $\mathcal{T}$.

At all times, $t > 0$, we assume the following observation model:

$$y_t = Ax_t + w_t, \ \|w_t\| \le \epsilon$$

where x_t is an m length sparse vector with support set $\mathcal{N}_t$, i.e. $\mathcal{N}_t := \{i : (x_t)_i \neq 0\}$; y_t is the $n < m$ length observation vector at time t; and w_t is the observation noise. Our algorithms need more measurements at the initial time, $t = 0$. We use n_0 to denote the number of measurements used at $t = 0$ and we use A_0 to denote the corresponding $n_0 \times m$ measurement matrix, i.e. at $t = 0$, we have

$$y_0 = A_0 x_0 + w_0, \ \|w_0\| \le \epsilon$$

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

The S-restricted isometry constant (RIC) [15], δ_S, for the matrix, A, is the smallest real number satisfying

$$(1 - \delta_S)\|c\|^2 \le \|A_{\mathcal{T}} c\|^2 \le (1 + \delta_S)\|c\|^2 \quad (1)$$

for all sets $\mathcal{T} \subset [1, m]$ of cardinality $|\mathcal{T}| \le S$ and all real vectors c of length $|\mathcal{T}|$. The restricted orthogonality constant (ROC) [15], θ_{S_1,S_2}, is the smallest real number satisfying

$$|{c_1}' {A_{\mathcal{T}_1}}' A_{\mathcal{T}_2} c_2| \le \theta_{S_1,S_2} \|c_1\| \ \|c_2\| \quad (2)$$

for all disjoint sets $\mathcal{T}_1, \mathcal{T}_2 \subset [1, m]$ with $|\mathcal{T}_1| \le S_1$, $|\mathcal{T}_2| \le S_2$ and $S_1 + S_2 \le m$, and for all vectors c_1, c_2 of length $|\mathcal{T}_1|$, $|\mathcal{T}_2|$ respectively.

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

We use α to denote the support estimation threshold used by modified-CS and we use $\alpha_{\text{add}}, \alpha_{\text{del}}$ to denote the support addition and deletion thresholds used by modified-CS with add-LS-del.We use $\hat{\mathcal{N}}_t$ to denote the support estimate at time t. To keep notation simple, we avoid using the subscript t wherever possible.

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

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

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

The sets $\mathcal{T}_{\text{add}}, \Delta_{\text{add}}, \Delta_{e,\text{add}}$ are defined in Definition 3 which is given in the next section.

If two sets $\mathcal{B}$, $\mathcal{C}$ are disjoint, we just write $\mathcal{D} \cup \mathcal{B} \setminus \mathcal{C}$ instead of writing $(\mathcal{D} \cup \mathcal{B}) \setminus \mathcal{C}$, e.g. $\mathcal{N}_t = \mathcal{T}_t \cup \Delta_t \setminus \Delta_{e,t}$.

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

Remark 1: The reason we need the bounded noise assumption is as follows. When the noise is unbounded, e.g. Gaussian, all error bounds for CS and, similarly, all error bounds for modified-CS hold with "large probability" [4], [16], [17], [18]. To show stability, we need the error bound for modified-CS to hold at all times, $0 \le t \le \infty$ (this, in turn, is used to ensure that the support gets estimated with bounded error at all times). Clearly this is a zero probability event.

II. Modified-CS and Modified-CS-Add-LS-Del

Modified-CS was introduced in [6] as a solution to the problem of sparse reconstruction with partial and possibly erroneous knowledge of the support. It tries to find a signal that is sparest outside of the known support among all signals satisfying the data constraint. For a time sequence of sparse signals, we use the support estimate from the previous time as known support. This was studied in [14]. We summarize the algorithm in Algorithm 1. In Algorithm 1, we use thresholding to compute the current support estimate. However, as explained in [14], the modified-CS estimate is biased towards zero along $\mathcal{T}^c$ and may be biased away from zero along $\mathcal{T}$ and this causes single step thresholding to be less accurate. To address this issue, in [14], we introduced a three step add-LS-delete procedure for support estimation. Similar ideas were also used earlier in [3] and [19], [20] in related contexts. We summarize the resulting algorithm called "modified-CS-add-LS-del" in Algorithm 2. In add-LS-del, one uses a small addition threshold α_{add}; followed by LS estimation on the new support; and finally a larger threshold α_{del} applied to the LS estimate to delete elements. α_{add} needs to be just large enough to ensure that $A_{\mathcal{T}_{\text{add}}}$ is well conditioned.

Algorithm 1 Modified-CS

For $t \ge 0$, do

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

$$\min_\beta \|(\beta)\|_1 \text{ s.t. } \|y_0 - A_0 \beta\| \le \epsilon \quad (3)$$

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

$$\min_\beta \|(\beta)_{\mathcal{T}_t^c}\|_1 \text{ s.t. } \|y_t - A\beta\| \le \epsilon \quad (4)$$

3) *Estimate the Support.* Compute $\tilde{\mathcal{T}}_t$ as

$$\tilde{\mathcal{T}}_t = \{i \in [1, m] : |(\hat{x}_{t,modcs})_i| > \alpha\} \quad (5)$$

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

Definition 3 (Define $\mathcal{T}_{add,t}, \Delta_{add,t}, \Delta_{e,add,t}$): The set $\mathcal{T}_{\text{add},t}$ is the support estimate obtained after the support addition step 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 $\mathcal{T}_{\text{add},t}$ and the set $\Delta_{e,\text{add},t} := \mathcal{T}_{\text{add},t} \setminus \mathcal{N}_t$ denotes the set of extras in it.

The following lemma bounds the modified-CS error at t.

Lemma 1 (modified-CS error bound): Let x be a sparse vector with support $\mathcal{N}$ and let $y := Ax + w$ with $\|w\| \le \epsilon$. Also, let $\Delta := \mathcal{N} \setminus \mathcal{T}$ and $\Delta_e := \mathcal{T} \setminus \mathcal{N}$. Let $\hat{x}_{modcs}$ denote the solution of (4). If $\delta_{|\mathcal{T}|+3|\Delta|} < (\sqrt{2}-1)/2$, then $\|x - \hat{x}_{modcs}\| \le C_1(|\mathcal{T}| + 3|\Delta|)\epsilon \le 8.79\epsilon$, where $C_1(S) \triangleq \frac{4\sqrt{1+\delta_S}}{1-(\sqrt{2}+1)\delta_S}$.

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 $\hat{x}_{t,modcs}$ as the solution of (3).
2) *Modified-CS.* If $t > 0$, set $\mathcal{T}_t = \hat{\mathcal{N}}_{t-1}$ and compute $\hat{x}_{t,modcs}$ as the solution of (4).
3) *Additions / LS.* Compute $\mathcal{T}_{\text{add},t}$ and the LS estimate using it:

$$\hat{\mathcal{A}}_t = \{i \in [1, m] : |(\hat{x}_{t,modcs})_i| > \alpha_{\text{add}}\}$$

$$\mathcal{T}_{\text{add},t} = \mathcal{T}_t \cup \hat{\mathcal{A}}_t \quad (6)$$

$$(\hat{x}_{t,\text{add}})_{\mathcal{T}_{\text{add},t}} = {\mathcal{A}_{\mathcal{T}_{\text{add},t}}}^{\dagger} y_t, \quad (\hat{x}_{t,\text{add}})_{\mathcal{T}^c_{\text{add},t}} = 0 \quad (7)$$

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

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

$$\tilde{\mathcal{T}}_t = \mathcal{T}_{\text{add},t} \setminus \hat{\mathcal{R}}_t \quad (8)$$

$$(\hat{x}_t)_{\tilde{\mathcal{T}}_t} = {\mathcal{A}_{\tilde{\mathcal{T}}_t}}^{\dagger} y_t, \quad (\hat{x}_t)_{\tilde{\mathcal{T}}_t^c} = 0 \quad (9)$$

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

Proof: The proof follows using approach of [17]. It is given in the Appendix of [21].

The following lemma bounds the simple CS error.

Lemma 2 (CS error bound [17]): 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 (4) with $\mathcal{T} = \emptyset$. If $\delta_{2|\mathcal{N}|} < (\sqrt{2}-1)/2$, then $\|x - \hat{x}_{cs}\| \leq C_1(2|\mathcal{N}|)\epsilon \leq 8.79\epsilon$.

III. Signal Change Model

The algorithms described above do not assume any signal change model. But to obtain error bounds over time, we need a model for signal change. 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. A new element j gets added at time t_j at an initial magnitude $a_{j,t}$ and its magnitude increases for the next $d_{j,t} \geq d_{\min}$ time units. Notice that $d_{j,t}$ can be ∞ too, i.e. there is no maximum limit on how large a coefficient can become. For element j, the magnitude increase at time t is $r_{j,t}$ with $r_{\min} \leq r_{j,t} \leq r_{\max}$. Also, at each time t at most S_a elements out of the "large elements" set (the set of elements with magnitude at least $a_{\min} + d_{\min} r_{\min}$) leave the set and begin to decrease. These elements keep decreasing and get removed from the support in at most b time units. As demonstrated in Section V, the above assumptions are practically valid for MRI sequences. We specify our model precisely below.

Signal Model 1: Assume the following.

1) At the initial time, $t = 0$, the support set, $\mathcal{N}_0$, contains S_0 nonzero elements, i.e. $|\mathcal{N}_0| = S_0$.
2) At time t, $S_{a,t}$ elements are added to the support. A new element j gets added to the support at initial magnitude $a_{j,t}$ [1] and its magnitude increases at rate $r_{j,t}$ for the next $d_{j,t}$ [2] time units.
3) We define the "large set" as $\mathcal{L}_t := \{j : |(x_t)_j| \geq a_{\min} + d_{\min} r_{\min}\}$. Elements in $\mathcal{L}_{t-1}$ either remain in $\mathcal{L}_t$ (while increasing or decreasing or remaining constant) or decrease enough to leave $\mathcal{L}_t$. We assume that at time t, $S_{d,t}$ elements out of $\mathcal{L}_{t-1}$ decrease enough to leave $\mathcal{L}_{t-1}$, i.e. $|\mathcal{L}_{t-1} \setminus \mathcal{L}_t| = S_{d,t}$. 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.
4) We assume that $0 \leq S_{a,t} \leq S_a$, $0 \leq S_{d,t} \leq S_a$, $0 \leq S_{r,t} \leq S_a$, $r_{\min} \leq r_{j,t} \leq r_{\max}$, $a_{\min} \leq a_{j,t} \leq a_{\max}$ and $d_{j,t} \geq d_{\min}$.
5) The support size at any time t, $S_t := |\mathcal{N}_t| \leq S$.
   - As we explain below, $S_t \leq S$ holds if $S_0 \leq S$ and $\sum_{\tau=1}^{t} S_{a,\tau} \leq \sum_{\tau=1}^{t-b} S_{d,\tau}$.
   - More simply, $S_t \leq S$ also holds if $S_0 \leq S$; for $1 \leq t \leq b$, $S_{a,t} = S_{r,t} = 0$, $S_{d,t} = S_a$ and for $t > b$, $S_{a,t} = S_{r,t} = S_{d,t} = S_a$

Let $\mathcal{A}_t := \mathcal{N}_t \setminus \mathcal{N}_{t-1}$ denote the newly added set and let $\mathcal{I}_t := \{j : |(x_t)_j| > |(x_{t-1})_j|\}$ denote the set of increasing elements. Condition 2 implies that (i) $|\mathcal{A}_t| = S_{a,t}$; (ii) if $j \in \mathcal{A}_{t-t_0}$ (i.e. if x_j is added at $t - t_0$) for a $t_0 \leq d_{\min}$, then $|(x_t)_j| = a_{j,t-t_0} + \sum_{\tau=t-t_0+1}^{t} r_{j,\tau}$; and (iii) $\mathcal{A}_t \subseteq \mathcal{I}_t \cap \mathcal{I}_{t+1} \cdots \cap \mathcal{I}_{t+d_{\min}}$.

Let $\mathcal{R}_t := \mathcal{N}_{t-1} \setminus \mathcal{N}_t$ denote the newly removed set and let $\mathcal{D}_t := \mathcal{L}_t^c \cap |\{i : 0 < |(x_t)_i| < |(x_{t-1})_i|\}|$ denote the set of decreasing elements. Condition 3 implies that (i) $|\mathcal{R}_t| = S_{r,t}$; (ii) $\mathcal{D}_t \subseteq \mathcal{D}_{t+1} \cup \mathcal{R}_{t+1}$; (iii) $|\mathcal{D}_t| \leq \sum_{\tau=t-b+1}^{t} S_{d,\tau} \leq bS_a$ and (iv) $\sum_{\tau=1}^{t} S_{r,\tau} \geq \sum_{\tau=1}^{t-b} S_{d,\tau}$.

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

Finally, notice that $\mathcal{N}_t = \mathcal{I}_t \cup \mathcal{D}_t \cup \mathcal{L}_t$.

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 2: Assume Signal Model 1 with the following extra assumptions.

- Out of the $S_{d,t}$ elements that started decreasing at time t, at least $\frac{\tau}{b} S_{d,t}$ of them get removed by $t + \tau$ for $\tau < b$.
  - Thus, at time t, the total number of decreasing elements, $|\mathcal{D}_t| \leq S_{d,t} + \frac{b-1}{b} S_{d,t-1} + \ldots \frac{1}{b} S_{d,t-b+1} \leq S_a(b+1)/2$.

IV. Time Invariant Error Bounds

A. Modified-CS result

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

Theorem 1: Consider Algorithm 1. Assume that the noise is bounded, i.e. $\|w\| \leq \epsilon$ and that x_t satisfies Signal Model 2.

[1] $a_{j,t}$ is nonzero only when x_j begin to get added at time t.

[2] $d_{j,t}$ is nonzero only when x_j begin to get added at time t.

Also, assume that the modified-CS error is spread out enough so that

$$\|x_t - \hat{x}_t\|_\infty \leq \frac{\zeta_M}{\sqrt{S_a}} \|x_t - \hat{x}_t\| \tag{10}$$

for some $\zeta_M \leq \sqrt{S_a}$.

If there exists a $d_0 \leq d_{\min}$ such that the following hold:

1) algorithm parameters
   - $\alpha = \frac{\zeta_M}{\sqrt{S_a}} 8.79\epsilon$,
2) number of measurements
   - $\delta_{S+3(\frac{(b+1)}{2}+d_0+1)S_a} \leq (\sqrt{2}-1)/2$,
3) initial magnitude and magnitude increase rate
   - the following holds $a_{\min} + d_0 r_{\min} > \alpha + \frac{\zeta_M}{\sqrt{S_a}} 8.79\epsilon = \frac{\zeta_M}{\sqrt{S_a}} 17.58\epsilon$,
4) at $t = 0, n_0$ is large enough to ensure that $|\tilde{\Delta}_t| \leq \frac{b+1}{2} S_a + d_0 S_a, |\tilde{\Delta}_{e,t}| = 0$,

then, for all t,

1) $|\tilde{\Delta}_t| \leq \frac{(b+1)}{2} S_a + d_0 S_a$, $|\tilde{\Delta}_{e,t}| = 0$, $|\tilde{\mathcal{T}}_t| \leq S$,
2) $|\Delta_t| \leq \frac{(b+1)}{2} S_a + d_0 S_a + S_a$, $|\mathcal{T}_t| \leq S$, $|\Delta_{e,t}| \leq S_a$,
3) and $\|x_t - \hat{x}_t\| \leq 8.79\epsilon$

Proof: See [21].

Theorem 1 claims that if x_t satisfies Signal Model 2, if enough number of measurement is available and if each nonzero coefficient has either a large enough initial magnitude or a large enough rate of magnitude increase, then the number of misses and extras from current support estimate are bounded by a time-invariant value. Also, the reconstruction error is bounded by a time-invariant value. Notice that the above result bounds the extras and misses by a constant times S_a. Under the slow support change assumption, $S_a \ll S_t$. Thus, in this case, the support error sizes are much smaller than the support size, making the above a meaningful result.

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

Remark 2: In general, for any vector z, $\|z\|_\infty \leq \|z\|_2$ 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 $\|z\|_2$. There is no reason for the error $x_t - \hat{x}_t$ to be one-sparse. The assumption, $\|x_t - \hat{x}_t\|_\infty \leq \frac{\zeta_M}{\sqrt{S_a}} \|x_t - \hat{x}_t\|$ for some $\zeta_M \leq \sqrt{S_a}$, just quantifies this. Notice that if $\zeta_M = \sqrt{S_a}$, then the inequality always holds.

Remark 3: Notice that in our signal model, an element j can be added more than once.

Remark 4: Notice that condition 4 of Theorem 1 is not restrictive. It is easy to see that it will hold if n_0 is large enough to ensure that $\delta_{2S} \leq 0.207$.

Remark 5: In the signal model, we assume that only elements out of the large set, $\mathcal{L}_{t-1}$, can begin to decrease and get removed. This is again an assumption made for simplicity. As we can see from the proof, the result of the above theorem will hold, even if this was not true, i.e. even if increasing elements were allowed to begin to decrease. However the assumption that once an element begins decreasing it gets removed within b time units is essential.

B. Modified-CS-Add-LS-Del result

Theorem 2: Consider Algorithm 2. Assume that the noise is bounded, i.e. $\|w\| \leq \epsilon$ and that x_t satisfies Signal Model 2. Also, assume that

- the modified-CS error is spread out enough so that

$$\|x_t - \hat{x}_t\|_\infty \leq \frac{\zeta_M}{\sqrt{S_a}} \|x_t - \hat{x}_t\| \tag{11}$$

  for some $\zeta_M \leq \sqrt{S_a}$, and
- the LS step error is spread out enough so that

$$\|(x_t - \hat{x}_{t,\text{add}})_{\mathcal{T}_{\text{add},t}}\|_\infty \leq \frac{\zeta_L}{\sqrt{S_a}} \|(x_t - \hat{x}_{t,\text{add}})_{\mathcal{T}_{\text{add},t}}\| \tag{12}$$

  for some $\zeta_L \leq \sqrt{S_a}$.

If there exists a $d_0 \leq d_{\min}$ such that the following hold:

1) algorithm parameters
   - α_{add} is large enough so that there are at most f false adds at time t, i.e. $|\hat{\mathcal{A}}_t \setminus \mathcal{N}_t| \leq f$
   - $\alpha_{\text{del}} = 1.12 \frac{\zeta_L}{\sqrt{S_a}} \epsilon + 0.261 \zeta_L h$, where $h^2 = \frac{(b+1)}{2} (\alpha_{\text{add}} + \frac{\zeta_M}{\sqrt{S_a}} 8.79\epsilon)^2 + (d_0 a_{\max}^2 + a_{\max} r_{\max} d_0 (d_0 - 1) + r_{\max}^2 \frac{d_0(d_0-1)(2d_0-1)}{6})$.
2) number of measurements
   - $\delta_{S+3(\frac{(b+1)}{2} S_a + d_0 S_a + S_a)} \leq 0.207$
   - $\delta_{S+S_a+f} \leq 0.207$
   - $\theta_{S+S_a+f, \frac{(b+1)}{2} S_a + d_0 S_a} \leq 0.207$
3) initial magnitude and magnitude increase rate:

$$a_{\min} + d_0 r_{\min} > \max\{\alpha_{\text{add}} + \frac{\zeta_M}{\sqrt{S_a}} 8.79\epsilon, 2\alpha_{\text{del}}\} \tag{13}$$

4) at $t = 0, n_0$ is large enough to ensure that $|\tilde{\Delta}_t| \leq \frac{b+1}{2} S_a + d_0 S_a, |\tilde{\Delta}_{e,t}| = 0$,

then

1) $\tilde{\Delta}_t \subseteq \mathcal{D}_t \cup \mathcal{A}_t \cup \mathcal{A}_{t-1} \dots \mathcal{A}_{t-d_0+1}$
2) $|\tilde{\Delta}_t| \leq \frac{(b+1)}{2} S_a + d_0 S_a$, $|\tilde{\Delta}_{e,t}| = 0$, $|\tilde{\mathcal{T}}_t| \leq S$
3) $|\Delta_t| \leq \frac{(b+1)}{2} S_a + d_0 S_a + S_a$, $|\mathcal{T}_t| \leq S$
4) and $\|x_t - \hat{x}_t\| \leq 8.79\epsilon$

Proof: See [21].

C. Discussion

First let us compare Modified-CS with Modified-CS-Add-LS-Del. Consider some simplifications to the signal model to reduce the number of parameters. Suppose that $S_{a,t} = S_{r,t} = 0$, $S_{d,t} = S_a$ for $1 \leq t \leq b$, and $S_{a,t} = S_{r,t} = S_{d,t} = S_a$ for $t > b$. Also, let $a_{\min} = r_{\min}, a_{\max} = r_{\max}$, $b = 3, d_0 = 2$ and suppose that $f = S_a$.

The modified-CS result says the following. If

1) $\delta_{S+15S_a} \leq 0.207$,
2) $3r_{\min} > \frac{\zeta_M}{\sqrt{S_a}} 17.58\epsilon$,

then $|\tilde{\Delta}_t| \leq 4S_a$ and $|\tilde{\Delta}_{e,t}| = 0$.

The Modified-CS-Add-LS-Del result says the following. If

1) $\delta_{S+15S_a} \leq 0.207$,
2) $3r_{\min} > \alpha_{\text{add}} + \frac{\zeta_M}{\sqrt{S_a}} 8.79\epsilon$ and $3r_{\min} > 2.24 \frac{\zeta_L}{\sqrt{S_a}} \epsilon + 2.52\theta \zeta_L h$, where $h^2 = 2(\alpha_{\text{add}} + \frac{\zeta_M}{\sqrt{S_a}} 8.79\epsilon)^2 + 5r_{\max}^2$.

then $|\tilde{\Delta}_t| \le 4S_a$ and $|\tilde{\Delta}_{e,t}| = 0$.

To get an idea of the values of ζ_M and ζ_L, we did simulations based on Signal Model 1 with $m = 200$, $S_0 = 20, S_{a,t} = 2, S_{d,t} = 2, S_{r,t} = 2, b = 3, r_{j,t} = 1, a_{j,t} = 1$; a random Gaussian measurement matrix $A_{n\times m}$ with $m = 200, n = 72$; and measurement noise, w_t that was i.i.d. uniformly distributed between $\pm c$ with $c = 0.1266$. Thus $\epsilon = \sqrt{n}c$. We generated 500 realizations with the above parameters, and used both algorithms for reconstruction. We got $\zeta_M = 0.93\sqrt{S_a}, \zeta_L = 0.87\sqrt{S_a}$. Similar or smaller values were obtained if m was increased and S_0, S_a increased linearly.

Notice that both modified-CS and modified-CS-add-LS-del need the same assumptions on the number of measurements. With the above values for ζ_M and ζ_L, modified-CS needs $r_{\min} > 5.45\epsilon$. Modified-CS-Add-LS-Del needs $r_{\min} > \max\{0.33\alpha_{\text{add}} + 2.72\epsilon, 2.40\epsilon + 0.21\alpha_{\text{add}} + 0.34r_{\max}\}$. If $r_{\max} = 1.5r_{\min}$, this inequality gives $r_{\min} > 4.90\epsilon + 0.44\alpha_{\text{add}}$. Thus, with α_{add} small enough, clearly modified-CS-add-LS-del requires a weaker assumption on $r_{\min}$. As explained earlier and also in [14], α_{add} is a small threshold that is typically taken to be proportional to the noise per signal element, e.g. in our simulations, we took $\alpha_{\text{add}} = c/2 + r/16$. Using $c = \epsilon/\sqrt{n}$, this means that $\alpha_{\text{add}} \approx \epsilon/\sqrt{n}$. With this value for α_{add}, clearly the mod-cs-add-LS-del condition is weaker.

To compare with the CS result given in Lemma 2, notice that CS needs $\delta_{2S} < 0.207$, whereas both the modified-CS algorithms only need $\delta_{S+15S_a} < 0.207$. Under the slow support change assumption, $S_a \ll S$ and in this case, the modified-CS algorithms hold under a weaker restricted isometry condition (potentially fewer number of measurements required).

V. Model Verification

We verified that two different types of MRI image sequences – a larynx (vocal tract) MRI sequence and a brain functional MRI sequence – do indeed satisfy Signal Model 1. Both are discussed in [21]. Here we describe model verification for the larynx sequence. We used a 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 [6], 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 $\mathcal{N}_t$ satisfied $|\mathcal{N}_t| \le S = 113$ for all t. The number of additions from $t-1$ to t satisfied $|\mathcal{N}_t \setminus \mathcal{N}_{t-1}| \le 21$ and the number of removals, $|\mathcal{N}_{t-1} \setminus \mathcal{N}_t| \le 26$. Thus, $S_a = 26$. Also, the initial nonzero value, $a_{j,t}$, ranged from $a_{\min} = 13$ to $a_{\max} = 37$, the rate of magnitude increase, $r_{j,t}$, ranged from $r_{\min} = 1$ to $r_{\max} = 37$, and the duration for which the increase occurred, $d_{j,t}$, ranged from $d_{\min} = 1$ to $d_{\max} = 4$. Also, the maximum delay between the time that a coefficient began to decrease and when it was removed was $b = 7$.

VI. Conclusions

Under mild assumptions and for a realistic signal model, we showed that both the support recovery errors of both modified-cs and modified-cs-add-ls-del are bounded by a time-invariant and small value at all times. We also argued that our results hold under weaker assumptions on n than simple CS. Also, typically, the modified-cs-add-ls-del holds under weaker assumptions than the mod-cs result. Monte Carlo simulations backing our conclusions are shown in [21].

References

[1] 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.

[2] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Th.*, vol. 52(4), pp. 1289–1306, April 2006.

[3] N. Vaswani, "Kalman filtered compressed sensing," in *ICIP*, 2008.

[4] N. Vaswani, "Least Squares CS-residual (LS-CS): Compressive Sensing on Least Squares residual," *IEEE Trans. Sig. Proc.*, August 2010.

[5] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using kalman filtering with embedded pseudo-measurement norms and quasi-norms," *IEEE Trans. Sig. Proc.*, pp. 2405–2409, April 2010.

[6] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Processing*, September 2010 (shorter version in ISIT 2009).

[7] J. M. Kim, O. K. Lee, and J. C. Ye, "Exact Dynamic Support Tracking with Multiple Measurement Vectors using Compressive MUSIC," in *ArXiv preprint arXiv:1110.0378 [cs.IT]*, 2011.

[8] J. Ziniel, L. C. Potter, and P. Schniter, "Tracking and smoothing of time-varying sparse signals via approximate belief propagation," in *Asilomar Conf. on Sig. Sys. Comp.*, 2010.

[9] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning," *IEEE J. Sel. Topics Sig. Proc., Special Issue on Adaptive Sparse Representation of Data and Applications in Signal and Image Processing*, vol. 5, no. 5, pp. 912–926, Sept 2011.

[10] A. Charles, M. S. Asif, J. Romberg, and C. Rozell, "Sparsity penalties in dynamical system estimation," 2011.

[11] A. Khajehnejad, W. Xu, A. Avestimehr, and B. Hassibi, "Analyzing weighted l1 minimization for sparse recovery with nonuniform sparse models," *IEEE Trans. Sig. Proc.*, vol. 59, pp. 1985–2001, 2011.

[12] C. J. Miosso, R. von Borries, M. Argaez, L. Velazquez, C. Quintero, and C. Potes, "Compressed sensing reconstruction with prior information by iteratively reweighted least-squares," *IEEE Trans. Sig. Proc.*, vol. 57, pp. 2424–2431, January 2009.

[13] L. Jacques, "A short note on compressed sensing with partially known signal support," *Signal Processing*, December 2010.

[14] N. Vaswani, "Stability (over time) of Modified-CS for Recursive Causal Sparse Reconstruction," in *Allerton*, 2010.

[15] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Info. Th.*, vol. 51(12), pp. 4203 – 4215, Dec. 2005.

[16] E. Candes and T. Tao, "The dantzig selector: statistical estimation when p is much larger than n," *Annals of Statistics*, vol. 35 (6), 2007.

[17] 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.

[18] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals," *IEEE Trans. Info. Th.*, pp. 1030–1051, March 2006.

[19] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: Closing the gap between performance and complexity," *Preprint, 2008*.

[20] D. Needell and J.A. Tropp., "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harmonic Anal.*, vol. 26, pp. 301–321, 2008.

[21] J. Zhan and N. Vaswani, "Time invariant error bounds for modified-cs based sparse signal sequence recovery (long version)," *http://home.engineering.iastate.edu/ namrata/stability_modcs_ald.pdf.*