Stability of LS-CS-residual and modified-CS for sparse signal sequence reconstruction

Namrata Vaswani
ECE Dept., Iowa State University, Ames, IA 50011, USA, Email: namrata@iastate.edu

***Abstract*— In this work, we show the "stability" of two of our recently proposed algorithms, LS-CS-residual (LS-CS) and the noisy version of modified-CS, designed for recursive reconstruction of sparse signal sequences from noisy measurements. By "stability" we mean that the number of misses from the current support estimate and the number of extras in it remain bounded by a time-invariant value at all times. The concept is meaningful only if the bound is small compared to the current signal support size. A direct corollary is that the reconstruction errors are also bounded by a time-invariant value.**

I. Introduction

We study the "stability" of two of our recently proposed algorithms, LS-CS-residual (LS-CS) [1] and the noisy version of modified-CS [2], [3], designed for recursive reconstruction of sparse signal sequences from noisy measurements. The key assumption used by both algorithms is that the support changes slowly over time. This was verified in [2]. LS-CS replaces compressive sensing (CS) on the observation (simple CS) by CS on the least squares (LS) residual computed using the previous support estimate, denoted by T. Modified-CS uses a different approach. It finds a signal that satisfies the data constraint and is sparsest outside of T.

In [1] and [3], we bounded the reconstruction errors of LS-CS and of modified-CS, respectively, in terms of the sizes of the unknown part of the support and of the erroneous part of T. The sizes of these sets, and consequently the error bound, depend on the accuracy of the previous reconstruction, i.e. they are not time-invariant. To get a time-invariant error bound, we need a time-invariant bound on these sets' sizes, or equivalently on the size of the misses and extras in the final support estimate at any given time, i.e. we need "stability".

In [1], we proved LS-CS stability for a signal model with bounded signal power and support size. But our model allowed the delay between support addition times to be more than one (it needed to be large enough to allow all previously added elements to get detected). But, in practice, in most applications, this will not hold. The delay will be one, i.e. the support will change at every time, e.g. see Fig. 1 of [2]. *This important case is the focus of the current work.*

We show the stability of both modified-CS and LS-CS under mild assumptions (bounded noise, high enough SNR and enough measurements at every time) for a signal model that allows *equal number of support additions/removals at every time* and gradual coefficient increase/decrease. Our approach is similar in both cases. The proof for modified-CS is simpler and the result holds under fewer assumptions. Hence, in Sec. III, we give that result first. We then discuss the key modifications needed to prove LS-CS stability, followed by discussing and comparing the two results with each other and with CS. Simulation results demonstrating stability are given in Sec. IV and conclusions in Sec. V.

Related algorithms include [4] (assumes time-invariant support), [5], [6] (use past reconstructions to speed up current optimization but not to improve error), KF-CS [7] and regularized modified-CS [8]. Except KF-CS, none of these study stability over time. The KF-CS result is under very strong assumptions, e.g. it is for a random walk model with only support additions (no removals).

A. Notation and Problem Definition

We use T^c to denote the complement of T w.r.t. $[1,m] := [1,2,\dots m]$, i.e. $T^c := [1,m] \setminus T$. $|T|$ denotes the cardinality of T. 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 ℓ_k norm of a vector v. *If just* $\|v\|$ *is used, it refers to* $\|v\|_2$. For a matrix M, $\|M\|_k$ denotes its induced k-norm, while just $\|M\|$ refers to $\|M\|_2$. M' denotes the transpose of M. For a fat matrix A, A_T denotes the sub-matrix obtained by extracting the columns of A corresponding to the indices in T. The S-restricted isometry constant [9], $\delta(S)$, for an $n \times m$ matrix (with $n < m$), A, and the S, S' restricted orthogonality constant [9], $\theta(S, S')$, are as defined in [9, eq 1.3] and [9, eq 1.5] respectively.

We assume the following observation model:

$$y_t = Ax_t + w_t, \ \|w_t\| \le \epsilon \tag{1}$$

where x_t is an m length sparse vector with support N_t, y_t is the $n < m$ length observation vector at time t and w_t is observation noise. The noise is assumed to be bounded: $\|w_t\| \le \epsilon$. 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.

We use $\hat{x}_t$ to denote the final estimate of x_t at time t and $\hat{N}_t$ to denote its support estimate. To keep notation simple, we avoid using the subscript t wherever possible.

Definition 1 (T, Δ, Δ_e): We use $T := \hat{N}_{t-1}$ to denote the support estimate from the previous time. We use $\Delta := N_t \setminus T$ to denote the unknown part of the support at the current time and $\Delta_e := T \setminus N_t$ to denote the "erroneous" part of T. We attach the subscript t to the set, e.g. T_t or Δ_t, where necessary.

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

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

II. Dynamic Modified-CS and LS-CS

Here, we briefly review dynamic modified-CS and LS-CS.

A. Dynamic modified-CS algorithm

Modified-CS was first proposed in [10] for exact reconstruction from noiseless measurements when part of the support is known. It solved $\min_\beta \|\beta_{T^c}\|_1 \ s.t. \ y = A\beta$. For noisy measurements, we can relax the data constraint in many possible ways. In this work, we stick to the following, because its error bounds have the simplest form and this makes the corresponding stability result less messy.

$$\min_\beta \|\beta_{T^c}\|_1 \ s.t. \ \|y - A\beta\|^2 \le \epsilon^2 \tag{2}$$

Denote its output by $\hat{x}_{modcs}$. In this work, whenever we refer to CS, we refer to the following

$$\min_\beta \|\beta\|_1 \ s.t. \ \|y - A\beta\|^2 \le \epsilon^2 \tag{3}$$

We summarize below the dynamic modified-CS algorithm [2] for time sequences. At $t = 0$, we use large enough number of measurements, $n_0 > n$, and do CS.
Set $\hat{N}_{-1} = \phi$ (empty set). For $t \ge 0$ do,

1) If $t = 0$, set $A := A_{n_0}$, else set $A := A_n$.
2) *Modified-CS.* Solve (2) with $T = \hat{N}_{t-1}$ and $y = y_t$. Denote its output by $\hat{x}_{t,modcs}$.
3) *Detections / LS.* Compute $\tilde{T}_{\text{det}}$ and LS estimate using it:

$$\tilde{T}_{\text{det}} = T \cup \{i \in T^c : |(\hat{x}_{t,modcs})_i| > \alpha_{add}\}$$
$$(\hat{x}_{t,\text{det}})_{\tilde{T}_{\text{det}}} = A_{\tilde{T}_{\text{det}}}{}^\dagger y_t, \ \ (\hat{x}_{t,\text{det}})_{\tilde{T}_{\text{det}}^c} = 0 \tag{4}$$

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

$$\tilde{T} = \tilde{T}_{\text{det}} \setminus \{i \in \tilde{T}_{\text{det}} : |(\hat{x}_{t,\text{det}})_i| \le \alpha_{del}\}$$
$$(\hat{x}_t)_{\tilde{T}} = A_{\tilde{T}}{}^\dagger y_t, \ \ (\hat{x}_t)_{\tilde{T}^c} = 0 \tag{5}$$

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

Notice that one could also replace the addition and deletion steps above by a single step that computes $\hat{N}_t = \{i : |(\hat{x}_{t,modcs})_i| > \alpha_*\}$. This would be sufficient when the noise is small and $|\Delta_e|$ is small. But when either of these does not hold, the bias in $\hat{x}_{t,modcs}$ creates the following problem. Along T^c, the solution will be biased towards zero, while along T it may be biased away from zero (since there is no constraint on $(\beta)_T$). The set T contains Δ_e which needs to be deleted. Since the estimates along Δ_e may be biased away from zero, one will need a higher threshold to delete them. But that would make detecting additions more difficult especially since the estimates along $\Delta \subseteq T^c$ are biased towards zero.

By adapting the approach of [11], the error of the modified-CS step can be bounded as a function of $|T| = |N| + |\Delta_e| - |\Delta|$ and $|\Delta|$ [12]. We state a modified version of [12]'s result.

Theorem 1 (modified-CS error bound [12]): If $\|w\| \le \epsilon$ and $\delta(\max(3|\Delta|, |N| + |\Delta| + |\Delta_e|)) < \sqrt{2} - 1$, then

$$\|x_t - \hat{x}_{t,modcs}\| \le C_1(\max(3|\Delta|, |N| + |\Delta| + |\Delta_e|))\epsilon, \ \text{where}$$
$$C_1(S) \triangleq \frac{4\sqrt{1 + \delta(S)}}{1 - (\sqrt{2} + 1)\delta(S)} \tag{6}$$

B. LS-CS (dynamic CS-residual) algorithm

LS-CS uses partial knowledge of support in a different way than modified-CS. The LS-CS algorithm [1] is the same as the dynamic modified-CS algorithm but with step 2 replaced by

- *CS-residual step.*
  - Use $T := \hat{N}_{t-1}$ to compute the initial LS estimate, $\hat{x}_{t,\text{init}}$, and the LS residual, $\tilde{y}_{t,\text{res}}$, using

$$(\hat{x}_{t,\text{init}})_T = A_T{}^\dagger y_t, \ \ (\hat{x}_{t,\text{init}})_{T^c} = 0$$
$$\tilde{y}_{t,\text{res}} = y_t - A\hat{x}_{t,\text{init}} \tag{7}$$

  - Do CS on the LS residual, i.e. solve (3) with $y = \tilde{y}_{t,\text{res}}$ and denote its output by $\hat{\beta}_t$. Compute

$$\hat{x}_{t,\text{CSres}} := \hat{\beta}_t + \hat{x}_{t,\text{init}}. \tag{8}$$

The CS-residual step error can be bounded as follows. The proof follows in exactly the same way as that given in [1] where CS is done using Dantzig selector instead of (3). We use (3) here to keep the comparison with modified-CS easier.

Theorem 2 (CS-residual error bound [1]): If $\|w\| \le \epsilon$, $\delta_{2|\Delta|} < \sqrt{2} - 1$ and $\delta_{|T|} < 1/2$,

$$\|x_t - \hat{x}_{t,\text{CSres}}\| \le C'(|T|, |\Delta|)\epsilon + \theta_{|T|,|\Delta|} C''(|T|, |\Delta|)\|x_\Delta\|$$
$$C'(|T|, |\Delta|) \triangleq C_1(2|\Delta|) + \sqrt{2} C_2(2|\Delta|)\sqrt{\frac{|T|}{|\Delta|}}$$
$$C''(|T|, |\Delta|) \triangleq 2C_2(2|\Delta|)\sqrt{|T|}, \ \text{where}$$
$$C_1(S) \text{ is defined in (6)}, \ C_2(S) \triangleq 2\frac{1 + (\sqrt{2} - 1)\delta(S)}{1 - (\sqrt{2} + 1)\delta(S)} \tag{9}$$

III. Stability of Dynamic modified-CS and LS-CS

So far we bounded the modified-CS and CS-residual error as a function of $|\Delta|$ and $|\Delta_e|$. Similarly the final LS step error in either case can be bounded as a function of $|\tilde{\Delta}|$ and $|\tilde{\Delta}_e|$. In this section, we find the conditions under which we can obtain a time-invariant bound on the sizes of these sets, i.e. ensure "stability". This ensures a time-invariant bound on the reconstruction errors.

A. Signal model for studying stability

We assume a simple signal model that (a) allows equal and nonzero number of additions/removals from the support at every time, (b) allows a new coefficient magnitude to gradually increase from zero, at a rate M/d, for a duration, d, and finally reach a constant value, M, (c) allows coefficients to gradually decrease and become zero (get removed from support) at the same rate, and (d) has constant signal power and support size.

Signal Model 1: Assume the following.

1) At $t = 0$, support size is S_0 and it contains $2S_a$ elements each with magnitude $M/d, 2M/d, \dots (d-1)M/d$, and $(S_0 - (2d-2)S_a)$ elements with magnitude M.
2) At each $t > 0$, S_a coefficients get added to the support at magnitude M/d. Denote this set by $\mathcal{A}_t$.
3) At each $t > 0$, S_a coefficients which had magnitude M/d at $t-1$ get removed from the support (magnitude becomes zero). Denote this set by $\mathcal{R}_t$.
4) At each $t > 0$, the magnitude of S_a coefficients which had magnitude $(j-1)M/d$ at $t-1$ increases to jM/d. This occurs for all $2 \le j \le d$.
5) At each $t > 0$, the magnitude of S_a coefficients which had magnitude $(j+1)M/d$ at $t-1$ decreases to jM/d. This occurs for all $1 \le j \le (d-1)$.

In the above model, the size and composition of the support at any t is the same as that at $t = 0$. Also, at each t, there are S_a new additions and S_a removals and the signal power is $(S_0 - (2d-2)S_a)M^2 + S_a \sum_{j=1}^{d-1} j^2 M^2/d^2$.

To understand the model better, define the following sets.

Definition 3: Define the following

1) $\mathcal{D}_t(j) := \{i : |x_{t,i}| = jM/d,\ |x_{t-1,i}| = (j+1)M/d\}$.
2) $\mathcal{I}_t(j) := \{i : |x_{t,i}| = jM/d,\ |x_{t-1,i}| = (j-1)M/d\}$.
3) Small elements' set, $\mathcal{S}_t(j) := \{i : 0 < |x_{t,i}| < jM/d\}$.

With these definitions, clearly, the newly added set, $\mathcal{A}_t := \mathcal{I}_t(1)$, and the newly removed set, $\mathcal{R}_t := \mathcal{D}_t(0)$.

Consider a $d_0 \le d$. From the signal model, it is clear that at any time, t, S_a elements enter the small coefficients' set, $\mathcal{S}_t(d_0)$, from the bottom (set $\mathcal{A}_t$) and S_a enter from the top (set $\mathcal{D}_t(d_0-1)$). Similarly S_a elements leave $\mathcal{S}_t(d_0)$ from the bottom (set $\mathcal{R}_t$) and S_a from the top (set $\mathcal{I}_t(d_0)$). Thus,

$$\mathcal{S}_t(d_0) = \mathcal{S}_{t-1}(d_0) \cup (\mathcal{A}_t \cup \mathcal{D}_t(d_0-1)) \setminus (\mathcal{R}_t \cup \mathcal{I}_t(d_0)) \quad (10)$$

We will use this in our stability result.

Notice that the above model does not specify a particular generative model, e.g. at time t, out of the $2S_a$ elements with magnitude jM/d, for any $1 < j < d$, one can arbitrarily pick any S_a elements to increase and the other S_a to decrease. Also, it does not specify the signs of the nonzero elements. One simple generative model, which we use for our simulations, is as follows. At each t, select $\mathcal{A}_t \subseteq N_{t-1}^c$ and $\mathcal{D}_t(d-1) \subseteq N_t \cap \{i : |x_{t,i}| = M\}$ of size S_a, uniformly at random. Then let the same set of elements increase (decrease) until they become constant at M (become constant at zero). Set the sign to ± 1 with equal probability when the element gets added and retain the same sign at all future times.

B. Stability result for dynamic modified-CS

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 $\tilde{T}_{\text{det}}$ to definitely get deleted. These can be obtained using Theorem 1 and the following facts.

1) An $i \in \Delta$ will get detected if $|x_i| > \alpha_{add} + \|x - \hat{x}_{modcs}\|$. This follows since $\|x - \hat{x}_{modcs}\| \ge |(x - \hat{x}_{\text{det}})_i|$.
2) Similarly, an $i \in \tilde{T}_{\text{det}}$ will not get falsely deleted if $|x_i| > \alpha_{del} + \|(x - \hat{x}_{\text{det}})_{\tilde{T}_{\text{det}}}\|$.
3) All $i \in \tilde{\Delta}_{e,\text{det}}$ (the zero elements of $\tilde{T}_{\text{det}}$) will get deleted if $\alpha_{del} \ge \|(x - \hat{x}_{\text{det}})_{\tilde{T}_{\text{det}}}\|$.
4) If $\|w\| \le \epsilon$ and if $\delta_{|\tilde{T}_{\text{det}}|} < 1/2$ (or pick any constant less than one and the error bound will change appropriately), then $\|(x - \hat{x}_{\text{det}})_{\tilde{T}_{\text{det}}}\| \le \sqrt{2}\epsilon + 2\theta_{|\tilde{T}_{\text{det}}|, |\tilde{\Delta}_{\text{det}}|} \|x_{\tilde{\Delta}_{\text{det}}}\|$.

Combining the above facts with Theorem 1, we can get the following three lemmas.

Lemma 1 (Detection condition): Assume that $\|w\| \le \epsilon$, $|N| \le S_N$, $|\Delta_e| \le S_{\Delta_e}$, $|\Delta| \le S_\Delta$. Let $\Delta_1 := \{i \in \Delta : |x_i| \ge b_1\}$. All elements of Δ_1 will get detected at the current time if $\delta(\max(3S_\Delta, S_N + S_{\Delta_e} + S_\Delta)) < \sqrt{2} - 1$ and

$$b_1 > \alpha_{add} + C_1(\max(3S_\Delta, S_N + S_{\Delta_e} + S_\Delta))\epsilon \quad (11)$$

where $C_1(S)$ is defined in Theorem 1.

Proof: The proof follows from fact 1 and Theorem 1 and the fact that $C_1(.)$ is a non-decreasing function of $|N|, |\Delta|, |\Delta_e|$.

Lemma 2 (No false deletion condition): Assume that $\|w\| \le \epsilon$, $|\tilde{T}_{\text{det}}| \le S_T$ and $|\tilde{\Delta}_{\text{det}}| \le S_\Delta$. For a given b_1, let $T_1 := \{i \in \tilde{T}_{\text{det}} : |x_i| \ge b_1\}$. All $i \in T_1$ will not get (falsely) deleted at the current time if $\delta(S_T) < 1/2$ and

$$b_1 > \alpha_{del} + \sqrt{2}\epsilon + 2\theta_{S_T, S_\Delta} \|x_{\tilde{\Delta}_{\text{det}}}\|. \quad (12)$$

Proof: The lemma follows directly from facts 2, 4.

Lemma 3 (Deletion condition): Assume that $\|w\| \le \epsilon$, $|\tilde{T}_{\text{det}}| \le S_T$ and $|\tilde{\Delta}_{\text{det}}| \le S_\Delta$. All elements of $\tilde{\Delta}_{e,\text{det}}$ will get deleted if $\delta(S_T) < 1/2$ and $\alpha_{del} \ge \sqrt{2}\epsilon + 2\theta_{S_T, S_\Delta} \|x_{\tilde{\Delta}_{\text{det}}}\|$.

Proof: The lemma follows directly from facts 3, 4.

Using the above lemmas and the signal model, we obtain sufficient conditions to ensure that, for some $d_0 \le d$, at each time t, $\tilde{\Delta} \subseteq \mathcal{S}_t(d_0)$ (so that $|\tilde{\Delta}| \le (2d_0 - 2)S_a$) and $|\tilde{\Delta}_e| = 0$, i.e. only elements smaller than $d_0 M/d$ may be missed and there are no extras. This leads to the following result.

Theorem 3 (Stability of dynamic modified-CS): Assume Signal Model 1 and bounded noise, i.e. $\|w\| \le \epsilon$. If the following hold for some $1 \le d_0 \le d$,

1) *(addition and deletion thresholds)*
   a) α_{add} is large enough so that there are at most f false additions per unit time,
   b) $\alpha_{del} = \sqrt{2}\epsilon + 2\theta(S_0 + S_a + f, k_2(d_0))e(d_0)$,
2) *(no. of measurements, n)* n is large enough so that $\delta(\max(3k_1(d_0), S_0 + S_a + k_1(d_0))) < \sqrt{2} - 1$ and also $\delta(S_0 + S_a + f) < 1/2$,
3) *(SNR and n)* $(d_0 M/d) \ge \max(G_1, G_2)$, where

$$G_1 \triangleq \alpha_{add} + C_1(\max(3k_1(d_0), S_0 + S_a + k_1(d_0)))\epsilon$$
$$G_2 \triangleq \alpha_{del} + \sqrt{2}\epsilon + 2\theta(S_0 + S_a + f, k_2(d_0))e(d_0) \quad (13)$$

4) *(initialization)* at $t = 0$, n_0 is large enough to ensure that $\tilde{\Delta} \subseteq \mathcal{S}_0(d_0)$, $|\tilde{\Delta}| \le (2d_0 - 2)S_a$, $|\tilde{\Delta}_e| = 0$, $|\tilde{T}| \le S_0$,

where

$$k_1(d_0) \triangleq \max(1, 2d_0 - 2)S_a$$
$$k_2(d_0) \triangleq \max(0, 2d_0 - 3)S_a$$
$$e(d_0) \triangleq \sqrt{2S_a \sum_{j=1}^{d_0-1} j^2 M^2/d^2} \quad (14)$$

then,

1) at all $t \geq 0$, $|\tilde{T}| \leq S_0$, $|\tilde{\Delta}_e| = 0$, and $\tilde{\Delta} \subseteq \mathcal{S}_t(d_0)$ and so $|\tilde{\Delta}| \leq (2d_0 - 2)S_a$,
2) at all $t > 0$, $|T| \leq S_0$, $|\Delta_e| \leq S_a$, and $|\Delta| \leq k_1(d_0)$,
3) at all $t > 0$, $|\tilde{T}_{\text{det}}| \leq S_0 + S_a + f$, $|\tilde{\Delta}_{e,\text{det}}| \leq S_a + f$, and $|\tilde{\Delta}_{\text{det}}| \leq k_2(d_0)$

The proof follows by induction. We use the induction assumption; $T_t = \tilde{T}_{t-1}$; and the signal model to bound $|\Delta|, |\Delta_e|, |T|$. Then we use Lemma 1; the limit on number of false detections; and $|\tilde{T}_{\text{det}}| \leq |N| + |\tilde{\Delta}_{e,\text{det}}|$ to bound $|\tilde{\Delta}_{\text{det}}|, |\tilde{\Delta}_{e,\text{det}}|, |\tilde{T}_{\text{det}}|$. Finally, we use Lemmas 2 and 3 to bound $|\tilde{\Delta}|, |\tilde{\Delta}_e|, |\tilde{T}|$. The complete proof is given in the Appendix.

Corollary 1: Under assumptions of Theorem 3, at all $t \geq 0$,

1) $\|x_t - \hat{x}_t\| \leq \sqrt{2}\epsilon + (2\theta(S_0, (2d_0 - 2)S_a) + 1)e(d_0)$
2) $\|x_t - \hat{x}_{t,modcs}\| \leq C_1(\max(3k_1(d_0), S_0 + S_a + k_1(d_0)))\epsilon$ (recall: $\hat{x}_{t,modcs}$ is output of step 2) of Sec. II-A).

Remark 1: Note that condition 4 is not restrictive. It is easy to see that it will hold if n_0 is large enough to ensure that $\delta(2S_0) \leq \sqrt{2} - 1$; $\alpha_{add,0}$ is large enough s.t. there are at most f false detects; $\alpha_{del,0} = \sqrt{2}\epsilon + 2\theta(S_0 + f, k_1(d_0))e(d_0)$; and $(d_0M/d) > \max(\alpha_{add,0} + C_1(2S_0)\epsilon, 2\alpha_{del,0})$.

C. Stability result for LS-CS (dynamic CS-residual)

The overall approach is similar to the one discussed above for modified-CS. The key difference is in the detection condition lemma, which we give below.

Lemma 4 (Detection condition for LS-CS): Assume that $\|w\| \leq \epsilon$, $|T| \leq S_T$ and $|\Delta| \leq S_\Delta$. Let $b := \|x_\Delta\|_\infty$. For a $\gamma \leq 1$, let $\Delta_1 := \{i \in \Delta : \gamma b \leq |x_i| \leq b\}$ and let $\Delta_2 := \Delta \setminus \Delta_1$. Assume that $|\Delta_1| \leq S_{\Delta_1}$ and $\|x_{\Delta_2}\| \leq \kappa$. All $i \in \Delta_1$ will definitely get detected at the current time if $\delta(2S_\Delta) < \sqrt{2} - 1$, $\delta(S_T) < 1/2$,

$$\theta(S_T, S_\Delta)\sqrt{S_{\Delta_1}}C''(S_T, S_\Delta) < \gamma \quad \text{and}$$

$$\max_{|\Delta| \leq S_\Delta} \frac{\alpha_{add} + C'(S_T, |\Delta|)\epsilon + \theta(S_T, |\Delta|)C''(S_T, |\Delta|)\kappa}{\gamma - \theta(S_T, |\Delta|)\sqrt{S_{\Delta_1}}C''(S_T, |\Delta|)} < b$$

where $C'(.,.)$, $C''(.,.)$ are defined in Theorem 2.

Proof: From Theorem 2, if $\|w\| \leq \epsilon$, $\delta(2|\Delta|) < \sqrt{2} - 1$, $\delta(|T|) < 1/2$, then $\|x - \hat{x}_{CSres}\| \leq C'(|T|, |\Delta|)\epsilon + \theta(|T|, |\Delta|)C''(|T|, |\Delta|)\|x_\Delta\|$. Using $\|x_\Delta\| \leq \sqrt{|\Delta_1|}b + \|x_{\Delta_2}\|$, fact 1 of Sec. III-B and the fact that for all $i \in \Delta_1$, $|x_i| \geq \gamma b$, we can conclude that all $i \in \Delta_1$ will get detected if $\delta(2|\Delta|) < \sqrt{2} - 1$, $\delta(|T|) < 1/2$ and $\alpha_{add} + C'\epsilon + \theta C''\|x_{\Delta_2}\| + \theta C''\sqrt{|\Delta_1|}b < \gamma b$. The third inequality holds if $\theta\sqrt{|\Delta_1|}C'' < \gamma$ and $\frac{\alpha_{add} + C'\epsilon + \theta C''\|x_{\Delta_2}\|}{\gamma - \theta\sqrt{|\Delta_1|}C''} < b$.

Since we only know that $|T| \leq S_T$, $|\Delta| \leq S_\Delta$, $|\Delta_1| \leq S_{\Delta_1}$ and $\|x_{\Delta_2}\| \leq \kappa$, we need the above inequalities to hold for all values of $|T|, |\Delta|, |\Delta_1|, \|x_{\Delta_2}\|$ satisfying these upper bounds. This leads to the conclusion of the lemma. The left hand sides (LHS) of all the required inequalities, except the last one, are non-decreasing functions of $|\Delta|, |T|, |\Delta_1|$ and thus we just use their upper bounds. The LHS of the last one is non-decreasing in $|T|, |\Delta_1|, \|x_{\Delta_2}\|$, but is not monotonic in $|\Delta|$ (since $C'(|T|, |\Delta|)$ is not monotonic in $|\Delta|$). Hence we explicitly maximize over $|\Delta| \leq S_\Delta$. ■

The stability result then follows in the same fashion as Theorem 3. The only difference is that instead of Lemma 1, we use Lemma 4 applied with $S_T = S_0, S_\Delta = k_1(d_0)$, $b = d_0M/d$, $\gamma = 1$, $S_{\Delta_1} = S_a$ and $\kappa = e(d_0)$.

Theorem 4 (Stability of LS-CS): Assume Signal Model 1 and $\|w\| \leq \epsilon$. If the following hold for some $1 \leq d_0 \leq d$,

1) *(add/del thresholds)* (same condition as in Theorem 3)
2) *(no. of measurements, n)* n is large enough so that
   a) $\delta(2k_1(d_0)) < \sqrt{2} - 1$ and $\delta(S_0 + S_a + f) < 1/2$
   b) $\theta(S_0, k_1(d_0))\sqrt{S_a}C''(S_0, k_1(d_0)) < 1$
3) *(SNR and n)* $(d_0M/d) \geq \max(\tilde{G}_1, \tilde{G}_2)$, where

$$\tilde{G}_1 \triangleq \max_{|\Delta| \leq k_1(d_0)} \left[\frac{\alpha_{add} + C'(S_0, |\Delta|)\epsilon + \theta(S_0, |\Delta|)C''(S_0, |\Delta|)e(d_0)}{\gamma - \theta(S_0, |\Delta|)\sqrt{S_a}C''(S_0, |\Delta|)}\right]$$

$$\tilde{G}_2 \triangleq \alpha_{del} + \sqrt{2}\epsilon + 2\theta(S_0 + S_a + f, k_2(d_0))e(d_0)$$

4) *(initialization)* (same condition as in Theorem 3)

then, all conclusions of Theorem 3 and Corollary 1 hold for LS-CS, except the second claim of Corollary 1, which is replaced by $\|x_t - \hat{x}_{t,\text{CSres}}\| \leq \max_{|\Delta| \leq k_1(d_0)}[C'(S_0, |\Delta|)\epsilon + \theta(S_0, |\Delta|)C''(S_0, |\Delta|)[e(d_0) + S_a(d_0M/d)^2]]$.

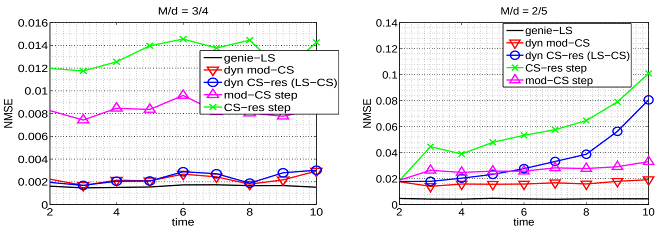
D. Implications of the Results and Comparisons

Both the stability results provide sufficient conditions under which the number of misses is less than $(2d_0 - 2)S_a$ and the number of extras is zero. From our discussion in the introduction, stability is meaningful only if $(2d_0 - 2)S_a \ll S_0$. From [2], we know that support changes slowly, i.e. $S_a \ll S_0$. Also, usually it is valid to assume that either $2dS_a \ll S_0$ (only a small number of coefficients are increasing or decreasing) or M is large enough to ensure that $d_0 \ll d$. Either of these ensures $(2d_0 - 2)S_a \ll S_0$.

For modified-CS, under Signal Model 1, "stability" is ensured under the following fairly mild assumptions: (a) the addition/deletion thresholds are appropriately set (conditions 1a and 1b); (b) the noise is bounded and the number of measurements, n, is large enough for condition 2 to hold; and (c) for a given support change rate, S_a, and magnitude change rate, M/d, the worst-case noise power, ϵ^2 is small enough and n is large enough to ensure that condition 3 holds.

The main difference for LS-CS stability is that the detection condition is much stronger (compare Lemma 4 with Lemma 1). This is because CS-residual error depends on $\|x_\Delta\|$, while there is no such dependence for modified-CS. Thus LS-CS needs an extra condition on $\theta(S, S')$ given in condition 2b of Theorem 4, which may not hold if n is too small. Also, when n is just large enough for condition 2b to hold, $\tilde{G}_1$ (defined in condition 3) may be very large and thus LS-CS will need higher signal increase rate, M/d, for its condition 3 to hold.

When n is large enough (for stability), it is easy to set α_{add} so that there are few false detections, f, compared to S_0, e.g. in our simulations, the average value of f was less than 1 when $S_0 = 20$. Now, $f \ll S_0$ along with $(2d_0 - 2)S_a \ll S_0$ (discussed earlier) implies that the requirements of Theorem 3



(a) $M = 3,\ d = 4$ (b) $M = 2,\ d = 5$

Fig. 1. Normalized MSE (NMSE) plots for modified-CS and LS-CS. In all cases, NMSE for CS was between 0.22-0.30 (not plotted).

on n (condition 2) are *significantly weaker* than those for the corresponding CS result [11] which requires $\delta(2S_0) \le \sqrt{2}-1$. The same is true when comparing the LS-CS requirements with those for CS, as long as M/d is large enough to ensure that d_0 is small enough for its conditions 2b and 3 to hold.

IV. Simulation Results

We compared dynamic modified-CS, LS-CS and CS for a few choices of M/d. In all cases, we used $m = 200$, $S_0 = 20$, $S_a = 2$, $n = 59$ and $uniform(-c, c)$ noise with $c = 0.1266$. We averaged over 50 simulations. We show the normalized MSE plot for $M = 3, d = 4$ in Fig. 1(a). The signal power for $M/d = 3/4$ is "large enough" compared to n and ϵ, to ensure stability of both LS-CS and modified-CS. From simulations, $|\Delta| \le 3$ for both modified-CS and LS-CS. When M/d was reduced to $2/3$ (not shown), again both were stable, although LS-CS errors were a bit higher than modified-CS. When signal power is decreased to $M/d = 2/5$ (Fig. 1(b)), it is still large enough to ensure stability of modified-CS, although for a larger value of d_0 (results in larger NMSE). But it is too small for LS-CS, thus resulting in its instability.

The NMSE for simple CS was much larger - between 22%-30% in all cases - since $n = 59$ is too small for CS.

V. Conclusions

We showed the "stability" of LS-CS and dynamic modified-CS for signal sequence reconstruction, under mild assumptions. By "stability" we mean that the number of misses from the current support estimate and the number of extras in it remain bounded by a time-invariant value at all times.

Appendix: Proof of Theorem 3

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. $|\tilde{\Delta}_{e,t-1}| = 0$, $|\tilde{T}_{t-1}| \le S_0$, and $\tilde{\Delta}_{t-1} \subseteq \mathcal{S}_{t-1}(d_0)$ so that $|\tilde{\Delta}_{t-1}| \le (2d_0 - 2)S_a$. Using this assumption we prove that the claim holds at t. We use (10) and the following facts often: (a) $\mathcal{R}_t \subseteq N_{t-1}$ and $\mathcal{A}_t \subseteq N_{t-1}^c$, (b) $N_t = N_{t-1} \cup \mathcal{A}_t \setminus \mathcal{R}_t$, and (c) if two sets B, C are disjoint, then, $(D \cap B^c) \cup C = D \cup C \setminus B$ for any set D.

Since $T_t = \tilde{T}_{t-1}$, so $|T_t| \le S_0$. Since $\Delta_{e,t} = \hat{N}_{t-1} \setminus N_t = \hat{N}_{t-1} \cap [(N_{t-1}^c \cap \mathcal{A}_t^c) \cup \mathcal{R}_t] \subseteq \tilde{\Delta}_{e,t-1} \cup \mathcal{R}_t = \mathcal{R}_t$. Thus $|\Delta_{e,t}| \le |\mathcal{R}_t| = S_a$.

Next we bound $|\Delta_t|$. Note that $\Delta_t = N_t \setminus \hat{N}_{t-1} = (N_{t-1} \cap \hat{N}_{t-1}^c \cap \mathcal{R}_t^c) \cup (\mathcal{A}_t \cap \mathcal{R}_t^c \cap \hat{N}_{t-1}^c) = (\tilde{\Delta}_{t-1} \cap \mathcal{R}_t^c) \cup (\mathcal{A}_t \cap \hat{N}_{t-1}^c) \subseteq (\mathcal{S}_{t-1}(d_0) \cap \mathcal{R}_t^c) \cup \mathcal{A}_t = \mathcal{S}_{t-1}(d_0) \cup \mathcal{A}_t \setminus \mathcal{R}_t$. First consider the case when the conditions of the theorem hold for a $d_0 > 1$. Since $\mathcal{R}_t$ is a subset of $\mathcal{S}_{t-1}(d_0)$ and $\mathcal{A}_t$ is disjoint with $\mathcal{S}_{t-1}(d_0)$, thus $|\Delta_t| \le |\mathcal{S}_{t-1}(d_0)| + |\mathcal{A}_t| - |\mathcal{R}_t| = (2d_0 - 2)S_a + S_a - S_a$. If $d_0 = 1$, $\mathcal{S}_{t-1}(d_0)$ is empty and so $\Delta_t = \mathcal{A}_t$ and thus $|\Delta_t| = S_a$. Thus in all cases $|\Delta_t| \le k_1(d_0)$.

Consider the detection step. There are at most f false detects (from condition 1a) and thus $|\tilde{\Delta}_{e,\text{det},t}| \le |\Delta_{e,t}| + f \le S_a + f$. Thus $|\tilde{T}_{\text{det},t}| \le |N_t| + |\tilde{\Delta}_{e,\text{det},t}| \le S_0 + S_a + f$.

Next we bound $|\Delta_{\text{det},t}|$. Using the above discussion and (10), $\Delta_t \subseteq \mathcal{S}_{t-1}(d_0) \cup \mathcal{A}_t \setminus \mathcal{R}_t = \mathcal{S}_t(d_0) \cup \mathcal{I}_t(d_0) \setminus \mathcal{D}_t(d_0-1)$. Apply Lemma 1 with $S_N = S_0$, $S_{\Delta_e} = S_a$, $S_\Delta = k_1(d_0)$, and $b_1 = d_0 M/d$ (so that $\Delta_1 \subseteq \mathcal{I}_t(d_0)$). Since conditions 2 and 3 of the theorem hold, all the undetected elements of $\mathcal{I}_t(d_0)$ will definitely get detected at time t. Thus $\tilde{\Delta}_{\text{det},t} \subseteq \mathcal{S}_t(d_0) \setminus \mathcal{D}_t(d_0 - 1)$. If $d_0 > 1$, $|\tilde{\Delta}_{\text{det},t}| \le |\mathcal{S}_t(d_0)| - |\mathcal{D}_t(d_0-1)| = (2d_0-3)S_a$. This holds since $\mathcal{D}_t(d_0 - 1) \subseteq \mathcal{S}_t(d_0)$. If $d_0 = 1$, $\mathcal{S}_t(d_0)$ is empty and so $|\tilde{\Delta}_{\text{det},t}| = 0$. Thus in all cases $|\tilde{\Delta}_{\text{det},t}| \le k_2(d_0)$.

Consider the deletion step. Apply Lemma 3 with $S_T = S_0 + S_a + f$, $S_\Delta = k_2(d_0)$. Use $\tilde{\Delta}_{\text{det},t} \subseteq \mathcal{S}_t(d_0) \setminus \mathcal{D}_t(d_0-1) \subseteq \mathcal{S}_t(d_0)$ to bound $\|x_{\tilde{\Delta}_{\text{det}}}\|$ by $e(d_0)$. Since condition 1b holds, all elements of $\tilde{\Delta}_{e,\text{det},t}$ will get deleted. Thus $|\tilde{\Delta}_{e,t}| = 0$. Thus $|\tilde{T}_t| \le |N_t| + |\tilde{\Delta}_{e,t}| \le S_0$.

Finally, we bound $|\tilde{\Delta}_t|$. Apply Lemma 2 with $S_T = S_0 + S_a + f$, $S_\Delta = k_2(d_0)$, $b_1 = d_0 M/d$. Use $\tilde{\Delta}_{\text{det},t} \subseteq \mathcal{S}_t(d_0) \setminus \mathcal{D}_t(d_0 - 1) \subseteq \mathcal{S}_t(d_0)$ to bound $\|x_{\tilde{\Delta}_{\text{det}}}\|$ by $e(d_0)$. Since conditions 2 and 3 hold, all elements of $\tilde{T}_{\text{det}}$ with magnitude greater than or equal to $b_1 = d_0 M/d$ will definitely not get falsely deleted. But nothing can be said about the elements smaller than $d_0 M/d$. In the worst case $\tilde{\Delta}_t$ may contain all of these elements, i.e. it may be equal to $\mathcal{S}_t(d_0)$. Thus, $\tilde{\Delta}_t \subseteq \mathcal{S}_t(d_0)$ and so $|\tilde{\Delta}_t| \le (2d_0 - 2)S_a$.

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 paragraphs 2-5 above.

References

[1] N. Vaswani, "Ls-cs: Compressive sensing on least squares residual," *Accepted (AQ) to IEEE Trans. Sig. Proc., arXiv:0911.5524*, 2010.

[2] W. Lu and N. Vaswani, "Modified compressive sensing for real-time dynamic mr imaging," in *ICIP*, 2009.

[3] W. Lu and N. Vaswani, "Modified bpdn for noisy compressive sensing with partially known support," in *ICASSP*, 2010.

[4] D. Angelosante and G.B. Giannakis, "Rls-weighted lasso for adaptive estimation of sparse signals," in *ICASSP*, 2009.

[5] C. Rozell, D. Johnson, R. Baraniuk, and B. Olshausen, "Locally competitive algorithms for sparse approximation," in *ICIP*, 2007.

[6] M. S. Asif and J. Romberg, "Dynamic updating for sparse time varying signals," in *CISS*, 2009.

[7] N. Vaswani, "Analyzing least squares and kalman filtered compressed sensing," in *ICASSP*, 2009.

[8] W. Lu and N. Vaswani, "Regularized modified bpdn for compressive sensing with partially known support," in *ISIT*, 2010, submitted.

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

[10] N. Vaswani and W. Lu, "Modified-cs: Modifying compressive sensing for problems with partially known support," in *ISIT*, 2009.

[11] 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.

[12] L. Jacques, "A short note on compressed sensing with partially known signal support," *ArXiv preprint 0908.0660*, 2009.