

LS-CS-Residual (LS-CS): Compressive Sensing on Least Squares Residual

Namrata Vaswani

Abstract—We consider the problem of recursively and causally reconstructing time sequences of sparse signals (with unknown and time-varying sparsity patterns) from a limited number of noisy linear measurements. The sparsity pattern is assumed to change slowly with time. The key idea of our proposed solution, LS-CS-residual (LS-CS), is to replace compressed sensing (CS) on the observation by CS on the least squares (LS) residual computed using the previous estimate of the support. We bound CS-residual error and show that when the number of available measurements is small, the bound is much smaller than that on CS error if the sparsity pattern changes slowly enough. Most importantly, under fairly mild assumptions, we show “stability” of LS-CS over time for a signal model that allows support additions and removals, and that allows coefficients to gradually increase (decrease) until they reach a constant value (become zero). By “stability,” we mean that the number of misses and extras in the support estimate remain bounded by time-invariant values (in turn implying a time-invariant bound on LS-CS error). Numerical experiments, and a dynamic MRI example, backing our claims are shown.

Index Terms—Compressive sensing, least squares, recursive reconstruction, sparse reconstructions.

I. INTRODUCTION

CONSIDER the problem of recursively and causally reconstructing time sequences of spatially sparse signals (with unknown and time-varying sparsity patterns) from a limited number of linear incoherent measurements with additive noise. The signals are sparse in some transform domain referred to as the “sparsity basis” [3]. An important example of this problem occurs in dynamic magnetic resonance (MR) image reconstruction of deforming brain shapes or the beating heart in real-time applications such as MR image guided surgery or other interventional radiology procedures [4] or functional MRI. Human organ images are piecewise smooth [see Fig. 1], and so the wavelet transform is a valid sparsity basis [5]. MRI captures a limited number of Fourier coefficients of the image, which are incoherent with respect to the wavelet transform [5], [6]. Other examples include real-time single-pixel video imaging [7] or real-time sensor network based sensing of time-varying fields. Due to strong temporal dependencies in the signal sequence, it is usually valid to assume that its *sparsity pattern (support of*

Manuscript received August 14, 2009; accepted March 09, 2010. Date of publication April 12, 2010; date of current version July 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Pierre Vanderghyest. This research was partially supported by NSF Grants ECCS-0725849 and CCF-0917015. A part of this work appeared in the *Proceedings of the ICIP 2008* and the *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009.

The author is with the Electrical and Computer Engineering Department, Iowa State University, Ames, IA 50011 USA (e-mail: namrata@iastate.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2010.2048105

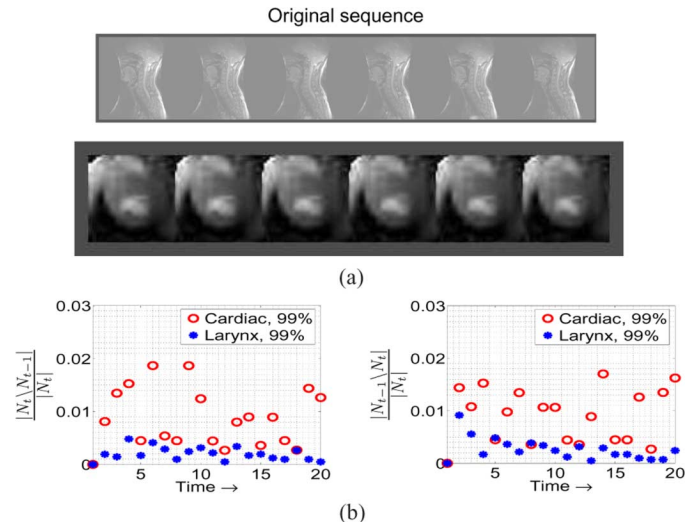


Fig. 1. In (a), we show two medical image sequences. In (b), N_t refers to the 99% energy support of the two-level Daubechies-4 2-D discrete wavelet transform (DWT) of these sequences. $|N_t|$ varied between 4121–4183 (≈ 0.07 m) for larynx and between 1108–1127 (≈ 0.06 m) for cardiac. We plot the number of additions (left) and the number of removals (right) as a fraction of $|N_t|$. Notice that all changes are less than 2% of the support size. (a) Top: larynx image sequence; bottom: cardiac sequence. (b) Slow support change plots. Left: additions; right: removals.

the sparsity transform vector) changes slowly over time. This is verified in Fig. 1. See also [8] and [9].

The solution to the static version of the above problem is provided by compressed sensing (CS) [3], [10]. CS for noisy observations, e.g., Dantzig selector [6], basis pursuit denoising (BPDN) [11], [12] or Lasso [13], [14] have been shown to have small error as long as incoherence assumptions hold. Most existing solutions for the dynamic problem, e.g., [7] and [15], are noncausal and batch solutions. Batch solutions process the entire time sequence in one go and thus have much higher reconstruction complexity. An alternative would be to apply CS at each time separately (simple CS), which is online and low-complexity, but since it does not use past observations, its reconstruction error is much larger when the number of available observations is small [see [5, Table I or Figs. 4(a)]].

The question is, For a time sequence of sparse signals, how can we obtain a recursive solution that improves the accuracy of simple CS by using past observations? By “recursive,” we mean a solution that uses only the previous signal estimate and the current observation vector at the current time. The key idea of our proposed solution, LS-CS-residual (LS-CS), is to replace CS on the observation by CS on the least squares (LS) residual computed using the previous support estimate. Its complexity is equal to that of simple CS which is $O(Nm^3)$ where m is the signal length and N is the time duration [16, Table 1]. Compare this to $O(N^3m^3)$ for a batch solution.

Other somewhat related work includes [17] and [18] (use the previous estimate to speed up the current optimization, but not to improve reconstruction error), and [19] (does not allow the support to change over time). Both [18] and [19] appeared after [1]. The work of [20] gives an approximate batch solution for dynamic MRI which is quite fast (but offline). Some other related work, but all for reconstructing a single sparse signal, includes [21] (uses a recursive algorithm) and [22] (related model, but offline algorithm). None of these works bound the reconstruction error or show its stability over time. We do both of these things for LS-CS in this paper.

In this work, we do ‘‘CS’’, whether in simple CS or in CS-residual, using the Dantzig selector (DS) [6]. This choice was motivated by the fact that its guarantees are stronger and its results are simpler to apply/modify (they depend only on signal support size) as compared to those for BPDN given in [12] (these depend on the actual support elements). In later work [8], [9], for practical experiments with larger sized images, we have also used BPDN since it runs faster. Between DS and Lasso (ℓ_2 constrained ℓ_1 minimization) [13], [14], either can be used. If Lasso is used, by starting with results of [14], results analogous to our Theorems 1 and 2 can be proved in the same way and their implications will also remain the same. With a little more work, this can also be done for BPDN.

Paper Organization. The LS-CS-residual (LS-CS) algorithm is developed in Section II. We bound its error and compare it with CS in Section III. Conditions for ‘‘stability’’ are obtained in Section IV. Numerical experiments are given in Section V and conclusions and future work in Section VI.

A. Notation

The set operations \cup , \cap , and \setminus have the usual meanings. T^c denotes the complement of T w.r.t. $[1, m] := [1, 2, \dots, m]$, i.e., $T^c := [1, m] \setminus T$. $|T|$ denotes the size (cardinality) of T .

For a vector, v , and a set, T , v_T denotes the $|T|$ length subvector containing the elements of v corresponding to the indexes in the set T . $\|v\|_k$ denotes the ℓ_k norm of a vector v . If just $\|v\|$ is used, it refers to $\|v\|_2$. Also, $v_{(i)}$ refers to the i th largest magnitude element of v (notation taken from [3]). Thus, for an m length vector, $|v_{(1)}| \geq |v_{(2)}| \geq \dots \geq |v_{(m)}|$. We use the notation $v(S)$ to denote the subvector of v containing the S smallest magnitude elements of v .

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 tall matrix, $M, M^\dagger := (M'M)^{-1}M'$.

For a fat matrix A , A_T denotes the submatrix obtained by extracting the columns of A corresponding to the indexes in T . The S -restricted isometry constant [3], δ_S , for an $n \times m$ matrix (with $n < m$), A , is the smallest real number satisfying

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

for all subsets $T \subset [1, m]$ of cardinality $|T| \leq S$ and all real vectors c of length $|T|$. The restricted orthogonality constant [3], $\theta_{S,S'}$, is the smallest real number satisfying

$$|c_1' A_{T_1}' A_{T_2} c_2| \leq \theta_{S,S'} \|c_1\| \|c_2\| \quad (2)$$

for all disjoint sets $T_1, T_2 \subset [1, m]$ with $|T_1| \leq S, |T_2| \leq S', S + S' \leq m$, and for all vectors c_1, c_2 of length $|T_1|, |T_2|$.

B. Problem Definition and Some More Notation

Let $(z_t)_{m \times 1}$ denote the spatial signal at time t and $(y_t)_{n \times 1}$, with $n < m$, denote its noise-corrupted measurements' vector at t , i.e., $y_t = Hz_t + w_t$ where w_t is measurement noise. The signal, z_t , is sparse in a given sparsity basis (e.g., wavelet) with orthonormal basis matrix, $\Phi_{m \times m}$, i.e., $x_t \triangleq \Phi' z_t$ is sparse. Denote its support by N_t . Thus, the observation model is

$$y_t = Ax_t + w_t, \quad A \triangleq H\Phi, \quad \mathbb{E}[w_t] = 0, \quad \mathbb{E}[w_t w_t'] = \sigma^2. \quad (3)$$

We assume that A has unit norm columns. Our goal is to recursively estimate x_t (or equivalently the signal, $z_t = \Phi x_t$) 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 state our assumptions after the following definition.

Definition 1 (Define S_ , S_{**}):* For $A := H\Phi$,

- 1) let S_* denote the largest S for which $\delta_S < 1/2$ (the choice $1/2$ is arbitrary, we can replace it by any $b < 1$);
- 2) let S_{**} denote the largest S for which $\delta_{2S} + \theta_{S,2S} < 1$.

Assumption 1: In the entire paper, we assume the following.

- 1) *Sparsity and Slow Support Change.* The support size, $|N_t| \approx |N_0| \ll m$ and the additions, $|N_t \setminus N_{t-1}| \leq S_a \ll |N_0|$ and the removals, $|N_{t-1} \setminus N_t| \leq S_a \ll |N_0|$.
- 2) *Incoherence.* The measurement matrix, A , satisfies $S_a < S_{**}$ and $|N_t| + k < S_*$ for some $k > 0$ (as we argue later $k \ll |N_t|$ suffices).

Sparsity and slow support change is verified in Fig. 1. Incoherence (approximate orthonormality of S -column submatrices of A) is known to hold with high probability (w.h.p.) when A is a random Gaussian, Rademacher, or partial Fourier matrix and n is large enough [3], [6].

1) *More Notation:* We use \hat{x}_t to denote the estimate of x_t given by our algorithm at time t and \hat{N}_t to denote its support estimate. To keep notation simple, we avoid using the subscript t wherever possible. We will use the following sets often.

Definition 2 (Define T, Δ, Δ_e): We use $T := \hat{N}_{t-1}$ to denote the support estimate from the previous time. This serves as an initial estimate of the current support. We use $\Delta := N_t \setminus T$ to denote the unknown part of the support at the current time. We use $\Delta_e := T \setminus N_t$ to denote the ‘‘erroneous’’ part of T . We attach the subscript, t , e.g., T_t, Δ_t , etc., where necessary.

Definition 3 (Define $\tilde{T}, \tilde{\Delta}, \tilde{\Delta}_e$): We use $\tilde{T} := \hat{N}_t$ to denote the final estimate of the current support. We use $\tilde{\Delta} := N_t \setminus \tilde{T}$ to denote the ‘‘misses’’ in the final estimate and $\tilde{\Delta}_e := \tilde{T} \setminus N_t$ to denote the ‘‘extras’’.

Notice that $N_t = (T \cup \Delta) \setminus \Delta_e = (\tilde{T} \cup \tilde{\Delta}) \setminus \tilde{\Delta}_e$. Some more notation— $\hat{x}_{CSres}, \tilde{\Delta}_{det}, \tilde{T}_{det}, \alpha, \alpha_{del}$ —is defined in Section II.

II. LEAST SQUARES CS-RESIDUAL (LS-CS)

We develop the LS-CS-residual (LS-CS) algorithm in Section II-A and summarize it in Algorithm 1. We discuss heuristics for threshold selection in Section II-B. A useful extension, Kalman filtered CS-residual, is introduced in Section II-C.

A. LS-CS-Residual Idea and Algorithm

Given observation, y , the Dantzig selector [6] solves

$$\min_{\zeta} \|\zeta\|_1 \text{ s.t. } \|A'(y - A\zeta)\|_\infty < \lambda. \quad (4)$$

Now consider the recursive reconstruction problem. If the support of x_t , N_t , were known at each t , we could simply compute its least squares (LS) estimate along N_t while setting all other values to zero. We refer to this estimate as the “*genie-aided*” LS estimate. When N_t is not known, one could do simple CS at each time, i.e., solve (4) with $y = y_t$, followed by thresholding the output to estimate its support, and then do the same thing using the support estimate \hat{N}_t instead of N_t . But in doing so, we are throwing away the information contained in past observations. If the available number of measurements, n , is small, this incurs large error [see Table I in Section V].

To use the information contained in past observations, along with the knowledge that support changes slowly, we propose the following idea. Assume for a moment that the support has not changed from $t - 1$ to t . Use $T := \hat{N}_{t-1}$ to compute an initial LS estimate and compute the LS residual, i.e., compute

$$\begin{aligned} (\hat{x}_{t,\text{init}})_T &= A_T^\dagger y_t, & (\hat{x}_{t,\text{init}})_{T^c} &= 0, \\ \tilde{y}_{t,\text{res}} &= y_t - A \hat{x}_{t,\text{init}}. \end{aligned}$$

Notice that the LS residual, $\tilde{y}_{t,\text{res}}$, can be rewritten as

$$\tilde{y}_{t,\text{res}} = A\beta_t + w_t, \quad \beta_t := x_t - \hat{x}_{t,\text{init}} \quad (5)$$

where β_t is a $|T \cup \Delta|$ -sparse vector with $(\beta_t)_{(T \cup \Delta)^c} = 0$,

$$(\beta_t)_T = (x_t)_T - A_T^\dagger y_t = -A_T^\dagger (w_t + A_\Delta (x_t)_\Delta) \quad (6)$$

$$(\beta_t)_\Delta = (x_t)_\Delta - 0 = (x_t)_\Delta. \quad (7)$$

In the above, (6) follows because $A_T^\dagger A_T = I$ and $y_t = Ax_t + w_t = A_N(x_t)_N + w_t = A_{N \cap T}(x_t)_{N \cap T} + A_{N \cap T^c}(x_t)_{N \cap T^c} + w_t = A_T(x_t)_T + A_\Delta(x_t)_\Delta + w_t$. The last equality holds because $N \cap T \subseteq T$ and $\Delta = N \cap T^c$. Here $N \equiv N_t$.

Notice that $\Delta \subseteq (N_t \setminus N_{t-1}) \cup (N_{t-1} \setminus T)$ and $\Delta_e \subseteq (N_{t-1} \setminus N_t) \cup (T \setminus N_{t-1})$. From Section I-B, $|N_t \setminus N_{t-1}| \leq S_a$ and $|N_{t-1} \setminus N_t| \leq S_a$. If S_a is small enough and T is an accurate enough estimate of N_{t-1} (so that $|\Delta|$ and $|\Delta_e|$ are small enough) and A is incoherent enough (so that $\|A_T' A_\Delta\|$ is small enough), β_t will be small (compressible) along T . In other words, β_t will be only $|\Delta|$ -approximately-sparse. In this case, doing CS on $\tilde{y}_{t,\text{res}}$ should incur much less error than doing CS on y_t (simple CS), which needs to reconstruct a $|N_t|$ -sparse signal, x_t . This is the key idea of our approach.

Thus, we propose to do CS on the LS residual (CS-residual), i.e., solve (4) with $y = \tilde{y}_{t,\text{res}}$ and denote its output by $\hat{\beta}_t$. Now,

$$\hat{x}_{t,\text{CSres}} := \hat{\beta}_t + \hat{x}_{t,\text{init}} \quad (8)$$

can serve as one possible estimate of x_t . But, as explained in [6], since $\hat{\beta}_t$ is obtained after ℓ_1 norm minimization, it will be biased towards zero. Thus, $\hat{x}_{t,\text{CSres}}$ will also be biased (its elements along Δ will be biased towards zero, while those along T may be biased away from zero). We can use the Gauss–Dantzig selector trick of [6] to reduce the bias. To do that, we first detect the new additions as follows:

$$\tilde{T}_{\text{det}} = T \cup \{i \in [1, m] : |(\hat{x}_{t,\text{CSres}})_i| > \alpha\} \quad (9)$$

and then we use \tilde{T}_{det} to compute an LS estimate

$$(\hat{x}_{t,\text{det}})_{\tilde{T}_{\text{det}}} = A_{\tilde{T}_{\text{det}}}^\dagger y_t, \quad (\hat{x}_{t,\text{det}})_{(\tilde{T}_{\text{det}})^c} = 0. \quad (10)$$

If $\tilde{T}_{\text{det}} = N_t$, $\hat{x}_{t,\text{det}}$ will be unbiased. In fact, it will be the best linear unbiased estimate, in terms of minimizing the mean squared error (MSE). But even if \tilde{T}_{det} is roughly accurate, the bias and MSE will be significantly reduced.

If the addition threshold, α , is not large enough, occasionally there will be some false detections (coefficients whose true value is zero but they wrongly get detected due to error in the CS-residual step). Also, there may have been actual removals from the true support. This necessitates a “deletion” step to delete these elements from the support estimate. Deletion can be done by thresholding $\hat{x}_{t,\text{det}}$, i.e., we compute

$$\hat{N}_t = \tilde{T}_{\text{det}} \setminus \{i \in \tilde{T}_{\text{det}} : |(\hat{x}_{t,\text{det}})_i| \leq \alpha_{\text{del}}\}. \quad (11)$$

The above is better than deleting using $\hat{x}_{t,\text{CSres}}$ which, as explained above, usually has a larger MSE than that of $\hat{x}_{t,\text{det}}$.

A final LS estimate can be computed using $\tilde{T} := \hat{N}_t$ as

$$(\hat{x}_t)_{\tilde{T}} = A_{\tilde{T}}^\dagger y_t, \quad (\hat{x}_t)_{\tilde{T}^c} = 0. \quad (12)$$

We summarize the complete algorithm in Algorithm 1.

Algorithm 1: LS-CS-Residual (LS-CS) Algorithm

Initialization ($t = 0$): At the initial time, $t = 0$, we run simple CS with a large enough number of measurements, $n_0 > n$ (usually much larger), i.e., we solve (4) with $y = y_0$ and $A = A_0 = H_0 \Phi$ (H_0 , and hence A_0 , will be an $n_0 \times m$ matrix). This is followed by support estimation and then LS estimation as in the Gauss–Dantzig selector [6]. We denote the final output by \hat{x}_0 and its estimated support by \hat{N}_0 . For $t > 0$ do,

1) *Initial LS*. 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 (5).

2) *CS-residual*. Do CS (Dantzig selector) on the LS residual, i.e., solve (4) with $y = \tilde{y}_{t,\text{res}}$ and denote its output by $\hat{\beta}_t$. Compute $\hat{x}_{t,\text{CSres}}$ using (8).

3) *Detection and LS*. Use (9) to detect additions to the support to get \tilde{T}_{det} . Compute the LS estimate, $\hat{x}_{t,\text{det}}$, using \tilde{T}_{det} , as given in (10).

4) *Deletion and LS*. Use (11) to detect deletions from the support to get $\tilde{T} := \hat{N}_t$. Compute the LS estimate, \hat{x}_t , using \tilde{T} , as given in (12).

5) Output \hat{x}_t and $\hat{z}_t = \Phi \hat{x}_t$. Feedback \hat{N}_t .

Increment t and go to step 1.

In most places in the paper, we use “*addition*” (“*removal*”) to refer to additions (removals) from the actual support, while using “*detection*” (“*deletion*”) to refer to additions (removals) from the support estimate. Occasionally, this is not followed.

We define the following sets which will be used in Section IV.

Definition 4 (Define \tilde{T}_{det} , $\hat{\Delta}_{\text{det}}$, $\hat{\Delta}_{e,\text{det}}$): The set \tilde{T}_{det} is defined in (9), $\hat{\Delta}_{\text{det}} := N_t \setminus \tilde{T}_{\text{det}}$ and $\hat{\Delta}_{e,\text{det}} := \tilde{T}_{\text{det}} \setminus N_t$.

B. Selecting the Thresholds α and α_{del}

A thumb rule from literature is to set α at the noise level [6]. But, for our problem, the more elements of Δ that one can detect without making $A_{\tilde{T}_{\text{det}}}$ badly conditioned, the better it is in

terms of reducing the MSE of $\hat{x}_{t,\text{det}}$. Thus, a better option than setting α explicitly, may be to keep adding the largest magnitude elements of $\hat{x}_{t,\text{det}}$ to \hat{T}_{det} until $A_{\hat{T}_{\text{det}}}$ just exceeds a condition number threshold (for the given noise level).

If α is appropriately set (either explicitly or implicitly) and the noise is small enough, the MSE of $\hat{x}_{t,\text{det}}$ will be significantly lower than that of $\hat{x}_{t,\text{CSres}}$. So, one can set α_{del} to a larger value than α , and still not have too many false deletions, while ensuring that most false detections get deleted. A good heuristic is to set α_{del} to a fraction of the minimum nonzero coefficient value. But if the noise is large, the MSE of $\hat{x}_{t,\text{det}}$ may not be much lower and also one would have used a larger value of α . In this case it is better to just set $\alpha_{\text{del}} = \alpha$.

Another heuristic, which ensures robustness to occasional large noise, is to limit the maximum number of detections at a given time to a little more than S_a (if S_a is known).

C. Kalman Filtered CS-Residual (KF-CS): Regularized LS-CS

Now, LS-CS does not use $(\hat{x}_{t-1})_T$ to improve the current estimate. But, often, in practice, coefficient values also change slowly. To use this fact, we can replace the initial LS estimate by a regularized LS estimate. If training data is available to learn a linear prior model for signal coefficients' change, this can be done by using a Kalman filter (KF). We develop and study the KF-CS algorithm in [1], [23]. As we demonstrate in [23], KF-CS significantly improves upon LS-CS when n (condition number of A_T is larger) or noise is larger.

III. BOUNDING CS-RESIDUAL ERROR

We first bound the CS-residual reconstruction error and compare it with the bound on CS error. In Section III-C, we give a tighter bound on the CS-residual error, but which holds under stronger assumptions. All bounds depend on $|T|$, $|\Delta|$.

To simplify notation, in this section, we remove the subscript t . Consider reconstructing x with support, N , from $y := Ax + w$. The support can be written as $N = (T \cup \Delta) \setminus \Delta_e$ where T is the “known” part of the support (equal to support estimate from the previous time), $\Delta_e := T \setminus N$ and $\Delta := N \setminus T$.

A. Bounding CS-Residual Error

If n is large enough so that $|T| + |\Delta| = |N| + |\Delta_e| \leq S_{**}$, then we can use the bounds given in [6, Theorems 1.1 or 1.2] to bound the CS-residual error. But recall that CS-residual is primarily designed for situations where n is smaller. It applies CS to the observation residual $\tilde{y} = A\beta + w$ where $\beta := x - \hat{x}_{\text{init}}$ is a $(|T| + |\Delta|)$ -sparse signal, that is compressible (small) along T . To bound its error, we first prove Lemma 1 which modifies [6, Theorem 1.3] to apply it to “sparse-compressible signals”, i.e., sparse signals that are partly (or fully) compressible. Next, we bound $\|\beta_T\|$ (the

“compressible” part of β). Finally we use this lemma along with the bound on $\|\beta_T\|$ to obtain the CS-residual error bound.

Lemma 1 (CS Error Bound—Sparse-Compressible Signal): Assume that $\|w\|_\infty \leq (\lambda/\|A\|_1)$ (bounded noise). Let ζ is an S_{nz} -sparse vector with support T_{nz} , and we measure $y := A\zeta + w$. Its estimate, $\hat{\zeta}$, obtained by solving (4), obeys the following. For all sets $T_{\text{rest}} \subseteq T_{nz}$ of size $|T_{\text{rest}}| = S_{nz} - S$ and for all $1 \leq S \leq \min(S_{**}, S_{nz})$,

$$\begin{aligned} \|\zeta - \hat{\zeta}\|^2 &\leq C_2(S)S\lambda^2 + C_3(S)\frac{\|(\zeta)_{T_{\text{rest}}}\|_1^2}{S} \\ &\leq C_2(S)S\lambda^2 + C_3(S)\frac{(S_{nz} - S)}{S}\|(\zeta)_{T_{\text{rest}}}\|^2 \end{aligned}$$

$$\text{where } C_2(S) \triangleq \frac{48}{(1 - \delta_{2S} - \theta_{S,2S})^2}$$

$$\text{and } C_3(S) \triangleq 8 + \frac{24\theta_{S,2S}^2}{(1 - \delta_{2S} - \theta_{S,2S})^2}. \quad (13)$$

The proof is given in Appendix A. Recall that S_{**} is defined in Definition 1.

Recall that $\beta = x - \hat{x}_{\text{init}}$ can be rewritten as

$$\begin{aligned} \beta_T &= -(A_T' A_T)^{-1} A_T' (A_\Delta x_\Delta + w) \\ \beta_\Delta &= x_\Delta, \quad (\beta)_{(T \cup \Delta)^c} = 0. \end{aligned} \quad (14)$$

As long as $\delta_{|T|} < 1$, $\|\beta_T\|^2$ can be bounded as follows.

$$\begin{aligned} \|\beta_T\|^2 &\leq 2 \left[\left\| (A_T' A_T)^{-1} A_T' A_\Delta x_\Delta \right\|^2 \right. \\ &\quad \left. + \left\| (A_T' A_T)^{-1} A_T' w \right\|^2 \right] \\ &\leq 2 \frac{\theta_{|T|,|\Delta|}^2}{(1 - \delta_{|T|})^2} \|x_\Delta\|^2 + \frac{2}{(1 - \delta_{|T|})} \|w\|^2. \end{aligned} \quad (15)$$

The above follows by using (a) $\left\| (A_T' A_T)^{-1} A_T' \right\|^2 = \left\| (A_T' A_T)^{-1} \right\| \leq 1/(1 - \delta_{|T|})$ (if $\delta_{|T|} < 1$) and (b) $\left\| A_T' A_\Delta \right\| \leq \theta_{|T|,|\Delta|}$. (a) follows because $\left\| (A_T' A_T)^{-1} \right\|$ is the inverse of the minimum eigenvalue of $A_T' A_T$. (b) follows from (2) by setting $c_1 = A_{T_1}' A_{T_2} c_2$, $T_1 = T$, and $T_2 = \Delta$.

From (15), it is clear that if the noise is small and if $|\Delta|$, $|\Delta_e|$ are small enough so that θ is small, $\|\beta_T\|$ is small. Using (15) along with Lemma 1 and the definition of S_* , we can prove the following result.

Theorem 1 (CS-Residual Error Bound): Assume that $|T| := |N| + |\Delta_e| - |\Delta| \leq S_*$ and $\|w\|_\infty \leq (\lambda/\|A\|_1)$. Then $\|x - x_{\text{CSres}}\|_2^2$ satisfies (16), shown at the bottom of the page, where $\theta := \theta_{|T|,|\Delta|}$ and $C_2(S)$, $C_3(S)$ are defined in (13).

The proof is given in Appendix A. Recall: $x_\Delta(K)$ is the vector containing the K smallest magnitude elements of x_Δ .

A simple corollary of the above result follows by applying it for a particular value of S , $S = |\Delta|$ when $|\Delta| > 0$. This will

$$\begin{aligned} \|x - \hat{x}_{\text{CSres}}\|^2 &\leq \min_{1 \leq S \leq \min(S_{**}, |T| + |\Delta|)} F_{\text{CSres}}(S), \text{ where} \\ F_{\text{CSres}}(S) &:= \left[C_2(S)S\lambda^2 + C_3(S)\frac{|T| + |\Delta| - S}{S} B(S) \right], \\ B(S) &= \begin{cases} 8\theta^2 \|x_\Delta\|^2 + 4\|w\|^2 & \text{if } S \geq |\Delta| \\ 8\theta^2 \|x_\Delta\|^2 + 4\|w\|^2 + \|x_\Delta(|\Delta| - S)\|^2 & \text{if } S < |\Delta| \end{cases} \end{aligned} \quad (16)$$

usually result in the smallest bound in situations where S_{**} is not much larger than $|\Delta|$. When $|\Delta| = 0$, $\beta_T = A_T^\dagger w$ which is anyway quite small. It is not immediately clear which value of S is the best. We retain the \min in this case. We also bound $\|w\|$ by its maximum value, $\sqrt{n}\lambda/\|A\|_1$.

Corollary 1 (Simpler CS-Residual Error Bound): Assume that $\|w\|_\infty \leq \lambda/\|A\|_1$, $|\Delta| \leq S_{**}$ and $|T| \leq S_*$.

1) If $|\Delta| > 0$,

$$\begin{aligned} \|x - \hat{x}_{\text{CSres}}\|^2 &\leq C' + C''\theta_{|T|,|\Delta|}^2\|x_\Delta\|^2, \text{ where} \\ C' &\equiv C'(|T|, |\Delta|) \\ &:= C_2(|\Delta|)|\Delta|\lambda^2 + 4C_3(|\Delta|)\frac{|T|}{|\Delta|}\frac{n\lambda^2}{\|A\|_1^2} \\ C'' &\equiv C''(|T|, |\Delta|) := 8C_3(|\Delta|)|T|. \end{aligned} \quad (17)$$

2) If $|\Delta| = 0$, $\|x - \hat{x}_{\text{CSres}}\|^2 \leq B_0$ where

$$B_0 := \min_{1 \leq S \leq S_{**}} \left[C_2(S)S\lambda^2 + C_3(S)\frac{|T| - S}{S}\frac{4n\lambda^2}{\|A\|_1^2} \right] \quad (18)$$

where $C_2(S)$, $C_3(S)$ are defined in (13).

This corollary is used in the LS-CS stability result.

Remark 1: It is easy to modify the above results for *Gaussian noise*, $w \sim \mathcal{N}(0, \sigma^2 I)$, in a fashion analogous to the results of [6]. Like [6], we will also get “large probability” results. We do not do this here because any large probability result will make the study of stability over time difficult.

Remark 2: In the bounds of Theorem 1 or Corollary 1, there is a term that is proportional to $(|T| + |\Delta| - S)$ or to $|T|$ respectively. This comes from Lemma 1 when we bound the ℓ_1 norm term, $\|(\zeta)_{T_{\text{rest}}}\|_1^2$, by $(S_{nz} - S)$ times $\|(\zeta)_{T_{\text{rest}}}\|_2^2$. A similar term is also there in the CS bound given below in (19) (bound for CS error which holds under the same weak assumptions as those used by our result¹).

B. Comparing CS-Residual and CS Error Bounds

We now compare the CS-residual bound with that of CS.

Remark 3: By showing that the upper bound on CS-residual error is much smaller, we only show that the performance guarantees for CS-residual are better than those for CS. To actually compare their errors, we use simulations.

To compare the CS-residual bound with CS, first note that the CS error bounds for sparse signals given in [6, Theorems 1.1 and 1.2] apply only when $\delta_{2S} + \theta_{S,2S} < 1$ holds for $S = |N|$, i.e., when $|N| \leq S_{**}$. When n is small and this does not hold, these results are not applicable. On the other hand, Lemma 1 does not assume anything about S_{**} . Let \hat{x}_{CS} denote the simple CS output. Using Lemma 1, it is easy to see that

$$\begin{aligned} \|x - \hat{x}_{\text{CS}}\|^2 &\leq \min_{1 \leq S \leq \min(S_{**}, |N|)} F_{\text{CS}}(S), \text{ where} \\ F_{\text{CS}}(S) &:= \left[C_2(S)S\lambda^2 + C_3(S)\frac{|N| - S}{S}B_{\text{CS}}(S) \right], \\ B_{\text{CS}}(S) &:= \|x_N(|N| - S)\|^2. \end{aligned} \quad (19)$$

¹A term containing the ℓ_1 norm of the “compressible” part appears in all bounds for CS for compressible signals, e.g., [6, Theorem 1.3] or [14, Theorem 1], and hence also appears when we bound CS error for sparse signals with not enough measurements.

Compare (16) with (19) under the following assumptions.

- 1) The magnitude of the largest element of x_Δ is smaller than or equal to that of the smallest element of $x_{N \setminus \Delta}$, i.e., $|(x_\Delta)_{(1)}| \leq |(x_{N \setminus \Delta})_{(N \setminus \Delta)}|$. This is reasonable since Δ contains the recently added elements which will typically be smaller while $N \setminus \Delta$ contains the previously added elements which should have a larger value.
- 2) $|\Delta|$, $|\Delta_e|$ are small enough (i.e., S_a is small enough and $T \approx N_{t-1}$) and the noise is small enough so that
 - a) $|\Delta| \leq 0.1|N|$ and $|\Delta_e| \leq 0.1|N|$ (this ensures that $|T| \leq 1.1|N|$);
 - b) $\theta_{|T|,|\Delta|}^2 < 1/8$, and $\|w\|^2 \leq (\|x_{N \setminus \Delta}(|\Delta|)\|^2 (1 - 8\theta^2)/4)$.
- 3) n is small so that $S_{**} = 0.2|N|$, but is just large enough so that $S_* \geq 1.1|N|$. $S_* \geq 1.1|N|$ along with assumption 2a ensures that $\delta_{|T|} < 1/2$.

The above assumptions ensure that $\|\beta_T\|^2 \leq 8\theta^2\|x_\Delta\|^2 + 4\|w\|^2 \leq 8\theta^2\|x_{N \setminus \Delta}(|\Delta|)\|^2 + 4\|w\|^2 \leq \|x_{N \setminus \Delta}(|\Delta|)\|^2$, i.e., β_T is “compressible enough.”

Under the above assumptions, we show that $F_{\text{CSres}}(S)$ in (16) is significantly smaller than $F_{\text{CS}}(S)$ in (19) for each value of S and hence the same will hold for the upper bounds.² For any S , the first term in $F_{\text{CSres}}(S)$ and $F_{\text{CS}}(S)$ is the same. In the second term, the main difference is in $B(S)$ versus $B_{\text{CS}}(S)$. The constants are almost the same, their ratio is $(|N| + |\Delta_e| - S)/|N| - S = 1 + (|\Delta_e|/|N| - S) \leq 9/8$ (follows since $S \leq S_{**} = 0.2|N|$ and $|\Delta_e| \leq 0.1|N|$). Thus, if we can show that $B(S)$ is much smaller than $B_{\text{CS}}(S)$, we will be done. First consider $|\Delta| \leq S \leq 0.2|N|$. In this case,

$$\begin{aligned} B(S) &= 8\theta^2\|x_\Delta\|^2 + 4\|w\|^2 \\ &\leq \|x_{N \setminus \Delta}(|\Delta|)\|^2 \leq \|x_{N \setminus \Delta}(0.1|N|)\|^2 \\ B_{\text{CS}}(S) &\geq B_{\text{CS}}(0.2|N|) = \|x_N(0.8|N|)\|^2 \\ &= \|x_\Delta\|^2 + \|x_{N \setminus \Delta}(0.8|N| - |\Delta|)\|^2 \\ &> 0 + \|x_{N \setminus \Delta}(0.7|N|)\|^2 \\ &\geq 7\|x_{N \setminus \Delta}(0.1|N|)\|^2 \geq 7B(S). \end{aligned} \quad (20)$$

Now consider $1 \leq S < |\Delta|$

$$\begin{aligned} B(S) &= 8\theta^2\|x_\Delta\|^2 + 4\|w\|^2 + \|x_\Delta(|\Delta| - S)\|^2 \\ &\leq \|x_{N \setminus \Delta}(|\Delta|)\|^2 + \|x_\Delta\|^2 \\ &\leq 2\|x_{N \setminus \Delta}(|\Delta|)\|^2 \leq 2\|x_{N \setminus \Delta}(0.1|N|)\|^2 \\ B_{\text{CS}}(S) &> B_{\text{CS}}(|\Delta|) \\ &\geq B_{\text{CS}}(0.1|N|) = \|x_N(0.9|N|)\|^2 \\ &\geq \|x_{N \setminus \Delta}(0.8|N|)\|^2 \\ &\geq 8\|x_{N \setminus \Delta}(0.1|N|)\|^2 \geq 4B(S). \end{aligned} \quad (21)$$

Thus, $(B(S)/B_{\text{CS}}(S)) \leq 1/4$ in all cases. Denote the common first term in F_{CSres} and F_{CS} by T1. Denote the second terms by T2_{CSres} and T2_{CS}. Thus, $(T2_{\text{CSres}}/T2_{\text{CS}}) = (B(S)/B_{\text{CS}}(S))(|N| + |\Delta_e| - S)/|N| - S \leq (1/4)(9/8) =$

²Notice that $(F_{\text{CSres}}(S)/F_{\text{CS}}(S)) \leq a$ for all S implies that $(\min_S F_{\text{CSres}}(S)/F_{\text{CS}}(S)) \leq a$ for all S . Since this holds for all S , it also holds for the \max taken over S , i.e., $(\min_S F_{\text{CSres}}(S)/\min_S F_{\text{CS}}(S)) = \max_S (\min_S F_{\text{CSres}}(S)/F_{\text{CS}}(S)) \leq a$.

$9/32$. Thus, $F_{CSres}(S) \leq F_{CS}(S)$ for all $S \leq S_{**}$. By footnote 2, the same holds for the bounds.

Furthermore, if the noise is small enough, and for $S \leq S_{**}$, the second term is the dominant term in $F_{CS}(S)$, i.e., $T1 \ll T2_{CS}$. Then $(F_{CSres}(S)/F_{CS}(S)) = (T1/T1 + T2_{CS}) + (T2_{CSres}/T1 + T2_{CS}) \approx (T2_{CSres}/T2_{CS}) \leq 9/32$, i.e., $F_{CSres}(S)/F_{CS}(S)$ is also roughly less than $9/32$ for all S . From footnote 2, this means that *the CS-residual bound is also roughly (9/32) times the CS bound (is significantly smaller)*.

If $|\Delta| = 0$, assumption 2b does not hold and so the above comparison does not hold. But clearly, under high enough SNR, B_0 in Corollary 1 is much smaller than (19).

Monte Carlo Comparison. We used Monte Carlo to compare CS-residual reconstruction error with that of CS for a static problem. This is discussed in Section V-A and Table I.

C. Tighter CS-Residual Bound Under Stronger Assumptions **

To address an anonymous reviewer's comment, we give below a tighter error bound for CS-residual (does not contain a term proportional to $|T|$). But this also holds under a stronger assumption. This section can be skipped in a quick reading.

Using (14), $\|(\beta)_T\|_1 \leq \|A_T^\dagger A_\Delta\|_1 \|x_\Delta\|_1 + \|A_T^\dagger w\|_1$. Notice that if the noise is small and $|\Delta|$ is small, $\|(\beta)_T\|_1$ will be small. In particular, if $\|(\beta)_T\|_1 \leq b\|x_\Delta\|_1$, by applying the first inequality of Lemma 1 with $\zeta = \beta$, $S_{nz} = |T| + |\Delta|$, $S = |\Delta|$ and $T_{rest} = T$; using $\|x_\Delta\|_1 \leq \sqrt{|\Delta|}\|x_\Delta\|_2$; and combining the resulting bound with that given in Corollary 1, we get the following.

Corollary 2: Assume the following:

- 1) $\|w\|_\infty \leq (\lambda/\|A\|_1)$, $|\Delta| \leq S_{**}$, $|T| \leq S_*$;
- 2) $\|A_T^\dagger A_\Delta\|_1 < c$ and $\|x_\Delta\|_1 > \left(\|A_T^\dagger w\|_1/b - \|A_T^\dagger A_\Delta\|_1\right)$ for a $b > c$

then,

$$\|x - \hat{x}_{CSres}\|^2 \leq C_2(|\Delta|)|\Delta|\lambda^2 + C_3(|\Delta|) \times \min\left(b^2\|x_\Delta\|^2, 8|T|\theta^2\|x_\Delta\|^2 + 4|T|\frac{n\lambda^2}{\|A\|_1^2}\right). \quad (22)$$

If $|\Delta| = 0$, condition 2 cannot hold. In this case, $\|x - \hat{x}_{CSres}\|^2 \leq B_0$ with B_0 defined in Corollary 1. ■

Notice that *the first term in the min does not contain $|T|$* . The above bound is tighter when this first term is smaller, i.e., b is small enough (happens if $|\Delta|$ small but $\|x_\Delta\|_\infty$ large).

IV. LS-CS STABILITY

So far we bounded CS-residual error as a function of $|T|$, $|\Delta|$. The bound is small as long as $|\Delta_e|$ and $|\Delta|$ are small. A bound on LS-CS error as a function of $|\tilde{T}|$, $|\tilde{\Delta}|$ is also easy to obtain. The next questions are:

- 1) Under what conditions on the measurement model and the signal model, will the number of extras, $|\tilde{\Delta}_e|$, and the number of misses, $|\tilde{\Delta}|$, and hence also $|\Delta_e|$, $|\Delta|$, be bounded by a time-invariant value, i.e., be ‘‘stable’’? This will imply a time-invariant bound on LS-CS error.
- 2) If additions/removals occur every-so-often, under what conditions can we claim that $|\tilde{\Delta}_e|$, $|\tilde{\Delta}|$ will become zero within a finite delay of an addition time? This will mean that the LS-CS estimate becomes equal to the genie-aided LS estimate (LS estimate computed using N_t).

The answers to both questions are, of course, interrelated.

We first describe our signal model for studying stability in Section IV-A. The three key lemmas leading to our stability result are given in Section IV-B and the result itself is derived in Section IV-C. We discuss its implications in Section IV-D. We argue that if Assumption 1 holds, a) the bounds on the misses and extras are small compared to the support size (‘‘stability’’ is meaningful) and b) stability holds under weaker assumptions on the measurement matrix than those required for simple CS.

A. Signal Model

For stability, we need a signal model. We assume the following deterministic model that a) assumes a nonzero delay between new coefficient addition and removal times, b) allows new coefficients' magnitudes to gradually increase from zero for sometime and finally reach a constant value, and c) allows coefficients to gradually decrease and become zero (get removed from support). At $t = 0$, we assume that x_0 is $(S_0 - S_a)$ sparse with all ‘‘large’’ coefficients with values $\pm M$.

Signal Model 1: The model is as follows.

- 1) *Initialization.* At $t = 0$, x_0 is $(S_0 - S_a)$ sparse. All its nonzero coefficients have values $\pm M$.
- 2) *Addition.* At $t = t_j = 1 + (j - 1)d$, for all $j \geq 1$, S_a new coefficients get added. Denote the set of indexes of coefficients added at $t = t_j$ by $\mathcal{A} = \mathcal{A}(j)$. A new coefficient, $i \in \mathcal{A}$, gets added at an initial magnitude a_i (its sign can be ± 1) and then its magnitude increases at a rate a_i until it either reaches M or for d time units. Thus, the maximum magnitude of the i th coefficient is $\min(M, da_i)$ for $i \notin N_0$, and is M for $i \in N_0$.
- 3) *Removal.* S_a coefficients get removed at $t = t_{j+1} - 1 = jd$ for all $j \geq 1$. Denote the set of indexes of coefficients which get removed at $t_{j+1} - 1$ by $\mathcal{R} = \mathcal{R}(j)$. During $[t_{j+1} - r, t_{j+1} - 1]$, the elements of \mathcal{R} start to decrease and become zero at $t = t_{j+1} - 1$. For coefficient, i , the rate of decrease is $\min(M, da_i)/r$ per unit time.
- 4) The sets $\mathcal{A}(j)$ and $\mathcal{R}(j)$ are disjoint, i.e., the coefficients that just got added do not get removed.

Thus at any $t \in [t_j, t_{j+1} - r - 1]$, the support can be split as \mathcal{A} (*increasing coefficients*) and $N \setminus \mathcal{A}$ (*constant coefficients*), where $N = N_t = N_{t_j}$. At any $t \in [t_{j+1} - r, t_{j+1} - 2]$, it can be split as \mathcal{A} (*increasing*), \mathcal{R} (*decreasing*), $N \setminus (\mathcal{A} \cup \mathcal{R})$ (*constant*). At $t = t_{j+1} - 1$, $N = N_t = N_{t_j} \setminus \mathcal{R}$ (*all constant*).

Notice that in the above model the signal support size remains roughly constant. It is S_0 or $(S_0 - S_a)$ at all times. Also, the maximum signal power is bounded by $S_0 M^2$.

B. Three Key Lemmas

Proving stability, i.e., showing that the number of misses, $|\tilde{\Delta}|$, and extras, $|\tilde{\Delta}_e|$, remain bounded, requires finding sufficient conditions for the following three things to hold at the current time: a) one, or a certain number of, large undetected coefficients definitely get detected; b) large enough detected coefficients definitely do not get falsely deleted, and c) every-so-often the extras (false detects or true removals) definitely do get deleted. a) and b) are used to ensure that $|\tilde{\Delta}|$ remains bounded while (c) is used to ensure that $|\tilde{\Delta}_e|$, and hence $|\tilde{T}| \leq |N| + |\tilde{\Delta}_e|$, remains bounded. These three things are done in the following three lemmas.

Lemma 2 (Detection Condition): Assume that $|T| \leq S_T$, $|\Delta| \leq S_\Delta$, and $\|w\|_\infty \leq \lambda/\|A\|_1$. The current largest magnitude undetected element, $(x_\Delta)_{(1)}$, will definitely get detected at the current time if $S_T \leq S_*$, $S_\Delta \leq S_{**}$,

$$\frac{2\theta_{S_T, S_\Delta}^2 S_\Delta C'''(S_T, S_\Delta) < 1, \text{ and}}{\max_{|\Delta| \leq S_\Delta} \frac{2\alpha^2 + 2C'(S_T, |\Delta|)}{1 - 2\theta_{S_T, |\Delta|}^2 |\Delta| C'''(S_T, |\Delta|)}} < (x_\Delta)_{(1)}. \quad (23)$$

Lemma 3 (No False Deletion Condition): Assume that $\|w\|_\infty \leq \lambda/\|A\|_1$, $|\tilde{T}_{\text{det}}| \leq S_T$ and $|\tilde{\Delta}_{\text{det}}| \leq S_\Delta$. For a given b_1 , let $T_l := \{i \in \tilde{T}_{\text{det}} : x_i \geq b_1\}$. All $i \in T_l$ will not get (falsely) deleted at the current time if $S_T \leq S_*$, and

$$b_1^2 > 2\alpha_{\text{del}}^2 + \frac{8n\lambda^2}{\|A\|_1^2} + 16\theta_{S_T, S_\Delta}^2 |\tilde{\Delta}_{\text{det}}| \|x_{\tilde{\Delta}_{\text{det}}}\|_\infty^2. \quad (24)$$

Lemma 4 (Deletion Condition): Assume that $\|w\|_\infty \leq \lambda/\|A\|_1$, $|\tilde{T}_{\text{del}}| \leq S_T$ and $|\tilde{\Delta}_{\text{del}}| \leq S_\Delta$. All elements of $\tilde{\Delta}_{e, \text{del}}$ will get deleted if $S_T \leq S_*$ and $\alpha_{\text{del}}^2 \geq (4n\lambda^2/\|A\|_1^2) + 8\theta_{S_T, S_\Delta}^2 |\tilde{\Delta}_{\text{del}}| \|x_{\tilde{\Delta}_{\text{del}}}\|_\infty^2$.

These lemmas follow easily from Corollary 1 and a few simple facts. They are proved in Appendix B.

C. The Main Result

By running simple CS at $t = 0$ with an appropriate number of measurements, $n_0 > n$ (usually much larger), we assume that we detect all nonzero coefficients and there are no false detects, i.e., $\hat{N}_0 = N_0$. This assumption is made for simplicity. For stability, we need to ensure that within a finite delay of a new addition time, all new additions definitely get detected (call this delay the ‘‘worst case detection delay’’) and that this delay plus the coefficient decrease time, r , is smaller than the delay between two addition times, d . This needs to be done while ensuring that there are no false deletions of either the constant or the definitely detected increasing coefficients. Also, the deletion threshold needs to be high enough to definitely delete the extras every-so-often (ensure $|\tilde{T}|$ bounded).

To obtain our result, the above is done by splitting $[t_j, t_{j+1} - 1]$ into the four subintervals shown in Fig. 2 and using the lemmas from the previous subsection to find sufficient conditions so that the following hold for some $d_0 < d$:

- 1) At all $t \in [t_j, t_j + d_0 - 1]$, there is no false deletion of the constant coefficients (during this time the increasing coefficients may be too small and we do not care if they get detected or not). This prevents the number of misses from increasing.
- 2) At every $t = t_j + d_0 + i - 1$, for $i = 1, \dots, S_a$, (a) the i th largest increasing coefficient is definitely detected, and (b) all constant coefficients and the first i largest increasing coefficients are not falsely deleted. This ensures that by $t = t_j + d_0 + S_a - 1$, the number of misses becomes zero, i.e., the ‘‘worst case detection delay’’ is $d_0 + S_a - 1$.
- 3) At $t = t_j + d_0 + S_a - 1$, all false detects get deleted. This is needed to keep $|T|$ bounded.
- 4) At all $t \in [t_j + d_0 + S_a, t_{j+1} - r - 1]$, (a) the current falsely detected set is immediately deleted and (b) none of the constant or increasing coefficients get falsely deleted.

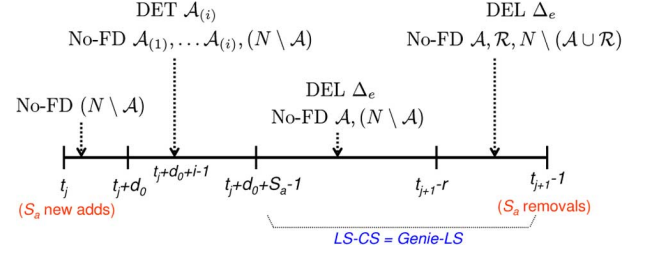


Fig. 2. Our approach to show stability (prove Theorem 2). We split $[t_j, t_{j+1} - 1]$ into the four subintervals shown above and ensure no false deletion (No-FD)/detection (DET)/deletion (DEL) of the coefficients listed in each. The notation $\mathcal{A}_{(i)}$ refers to the i th largest increasing coefficient. Recall: in $[t_j, t_{j+1} - r - 1]$, \mathcal{A} is the increasing coefficients’ set and $N \setminus \mathcal{A}$ is the constant coefficients’ set.

- 5) At all $t \in [t_{j+1} - r, t_{j+1} - 2]$, (a) the current falsely detected set is deleted and (b) none of the decreasing, constant or increasing coefficients are falsely deleted.
- 6) At $t_{j+1} - 1$, all falsely detected and removed coefficients are deleted and there is no false deletion.

Doing the above leads to the following result.

Theorem 2 (LS-CS Stability): Under Signal Model 1, if there exists a $d_0 < d$, so that the following conditions hold:

- 1) (*initialization*) all elements of x_0 get correctly detected and there are no false additions, i.e., $\hat{N}_0 = N_0$;
- 2) (*algorithm—thresholds*) we set $\alpha_{\text{del}} = 2\sqrt{n}\lambda/\|A\|_1$ and we set α large enough so that there are at most f false detections per unit time;
- 3) (*measurement model*)
 - a) $\|w\|_\infty \leq \lambda/\|A\|_1$, $S_a \leq S_{**}$, $S_0 + f(d_0 + S_a) \leq S_*$; and
 - b) $\theta_{S_T, S_\Delta}^2 S_\Delta C'''(S_T, S_\Delta) < 1$ with $S_T = S_0 + f(d_0 + S_a)$ and $S_\Delta = S_a$;
- 4) (*signal model—additions & no false deletions of increasing coefficients*) the following hold for all $i = 1, \dots, S_a$ and for all $\mathcal{A} = \mathcal{A}(j)$ for all j :
 - a) with $S_T = S_0 + f(d_0 + i - 1)$ and $S_\Delta = S_a - i + 1$,

$$\min(M, (d_0 + i)(a_{\mathcal{A}})_{(i)})^2 > \max_{|\Delta| \leq S_\Delta} \frac{2\alpha^2 + 2C'(S_T, |\Delta|)}{1 - 2(\theta_{S_T, |\Delta|})^2 |\Delta| C'''(S_T, |\Delta|)};$$

$$\text{b) with } S_T = S_0 + f(d_0 + i), S_\Delta = S_a - i \text{ and } (a_{\mathcal{A}})_{(S_a+1)} \equiv 0,$$

$$\min(M, (d_0 + i)(a_{\mathcal{A}})_{(i)})^2 > 2\alpha_{\text{del}}^2 + \left(\frac{8n\lambda^2}{\|A\|_1^2} \right) + 16\theta_{S_T, S_\Delta}^2 (S_a - i) \min(M, (d_0 + i)(a_{\mathcal{A}})_{(i+1)})^2;$$

- 5) (*signal model—no false deletions of constant coefficients*) with $S_T = S_0 + f(d_0 + S_a)$, $S_\Delta = S_a$,

$$\min(M, d \min_i a_i)^2 > 2\alpha_{\text{del}}^2 + \left(\frac{8n\lambda^2}{\|A\|_1^2} \right) + 16\theta_{S_T, S_\Delta}^2 S_a \min(M, (d_0 + S_a) \max_i a_i)^2;$$

- 6) (*signal model—no false deletions of decreasing coeff’s*)

$$\min(M, d \min_i a_i)^2 > r^2 \left(2\alpha_{\text{del}}^2 + \left(\frac{4n\lambda^2}{\|A\|_1^2} \right) \right);$$

7) (signal model—delay b/w addition times large enough)

$$d \geq d_0 + S_a + r$$

where $C'(\cdot, \cdot), C''(\cdot, \cdot)$ are defined in (17), then,

- 1) at all t , $|\tilde{\Delta}| \leq S_a$, $|\tilde{\Delta}_e| \leq f(S_a + d_0)$ and $|\tilde{T}| \leq S_0 + f(S_a + d_0)$ and the same bounds also hold for $|\Delta|$, $|\Delta_e|$, $|T|$ respectively; and
- 2) for all $t \in [t_j + d_0 + S_a - 1, t_{j+1} - 1]$, $|\tilde{\Delta}| = 0 = |\tilde{\Delta}_e|$, and thus $\tilde{N}_t = N_t$ (LS-CS estimate = genie-LS estimate).

The proof is given in Appendix C. Note that in $\min_i a_i$, the min is taken over $i \in [1, m]$ and same for $\max_i a_i$. We now give a simple corollary of Theorem 2 (proved in Appendix D).

Corollary 3: If the conditions given in Theorem 2 hold,

- 1) at all t , the LS-CS error satisfies

$$\begin{aligned} \|(x_t - \hat{x}_t)_{\tilde{\Delta}}\|^2 &\leq S_a \min(M, (d_0 + S_a) \max_i a_i)^2 \\ \|(x_t - \hat{x}_t)_{\tilde{T}}\|^2 &\leq 8\theta^2 S_a \min(M, (d_0 + S_a) \max_i a_i)^2 \\ &\quad + \left(\frac{4n\lambda^2}{\|A\|_1^2} \right) \end{aligned}$$

with θ computed at $S_T = S_0 + f(d_0 + S_a)$, $S_\Delta = S_a$;

- 2) at all t , the CS-residual error, $\|x_t - \hat{x}_{t, \text{CSres}}\|^2 \leq$

$$\max(B_0, C' + \theta^2 C'' S_a \min(M, (d_0 + S_a) \max_i a_i)^2)$$

with θ, C', C'' computed at $S_T = S_0 + f(d_0 + S_a)$, $S_\Delta = S_a$, and B_0 defined in (18).

Remark 4: Note that the initialization assumption is not restrictive. Denote the bound given by [6, Theorem 1.1] for $S = S_0 - S_a$ by $B1$. It is easy to see that this assumption will hold if the addition threshold at $t = 0$ is $\alpha_{\text{init}} = \sqrt{B1}$ (ensures no false detects) and if $M > \alpha_{\text{init}} + \sqrt{B1} = 2\sqrt{B1}$ (ensures all true adds detected). If the noise is small enough, by choosing n_0 large enough, we can make $B1$ small enough.

Even if this cannot be done, our result will only change slightly. The misses can be combined with the new additions at $t = 1$. Extras will at most increase the bound on $|T|$ by f .

D. Discussion and Extensions

Notice that Signal Model 1 results in *bounded SNR and roughly constant signal support size* at all times. Theorem 2 and Corollary 3 show that under Signal Model 1 and under the initialization assumption (made only for simplicity), if

- 1) the noise is bounded and n is large enough so that condition 3 holds;
- 2) the addition/deletion thresholds are appropriately set (condition 2);
- 3) for a given noise bound and n ,
 - a) the smallest constant coefficient magnitude is large enough (condition 5);
 - b) the rates of coefficient magnitude increase and decrease are large enough (conditions 4, 6);
 - c) and the delay between addition times, d , is larger than the “worst case detection delay” plus coefficient decrease time, $d_0 + S_a + r$ (condition 7);

then,

- 1) the number of misses, $|\tilde{\Delta}| \leq S_a$, and the number of extras, $|\tilde{\Delta}_e| \leq f(S_a + d_0)$ and the same bounds hold for $|\Delta|$, $|\Delta_e|$

(here $d_0 \leq d$ is the smallest integer for which conditions of Theorem 2 hold);

- 2) within a finite delay, $d_0 + S_a - 1$, all new additions get detected and not falsely deleted ($|\tilde{\Delta}| = 0$), and the extras get deleted ($|\tilde{\Delta}_e| = 0$), i.e., the LS-CS estimate becomes equal to the genie-LS estimate;
- 3) the LS-CS error and the CS-residual error are bounded by the time-invariant values given in Corollary 3.

From Assumption 1 (given in Section I-B), $S_a \ll S_0$. When n is large enough (as required above), it is easy to set α so that f is small, e.g., in our simulations the average f was often less than 1 while $S_0 = 20$. With a fast enough signal increase (as required above), d_0 will also be small. Thus, we can claim that $|\tilde{\Delta}|$ and $|\tilde{\Delta}_e|$ will be bounded by a small value compared to the signal support size, S_0 , i.e., “stability” is meaningful.

Under the above assumptions, compare our requirements on n (condition 3 of Theorem 2) to those of the CS error bound [6], which needs $S_0 \leq S_{**}$. The comparison is easier to do if we slightly modify the definition of S_{**} to be the largest S for which $\delta_{2S} < 1/2$ and $\delta_{2S} + \theta_{S,2S} < 1$ (this will imply that $2S_{**} \leq S_*$). Clearly $S_a \leq S_{**}$ is much weaker than $S_0 \leq S_{**}$. If S_a, d_0 and f are small enough (argued above), both $S_0 + f(d_0 + S_a) \leq S_*$ and condition 3b will also be much weaker than $2S_0 \leq 2S_{**} \leq S_*$.

Notice that our signal model assumes that support changes occur every d time instants. This may be slightly restrictive. But it is necessary in order to answer our second question (do the support errors ever become zero?). If we do not care about answering this, we can assume a signal model with $d = 1$ and modify our arguments to still ensure stability. But the support errors may never become zero. We do this in [24].

Also, note that if r is large (slow rate of decrease), condition 6 becomes difficult to satisfy. If we remove this, we may not be able to prevent false deletion of the decreasing coefficients when they become too small (go below $\alpha_{\text{del}} + (2\sqrt{n}\lambda/\|A\|_1)$). But since they are small, this will increase the CS-residual error at the next time instant only slightly. With small changes to our arguments, it should be possible to still prove stability.

V. NUMERICAL EXPERIMENTS

In Section V-A, we study a static problem and compare CS-residual error with that of CS. In Section V-B, we verify LS-CS stability. In Section V-C, we simulate lower SNRs and faster additions. In all these simulations, A was random-Gaussian. We averaged over 100 simulations (noise and signal supports for all times randomly generated) for all the time-series simulations and over 50 for the static one. In Section V-D, we show a dynamic MRI reconstruction example. Our MATLAB code is posted at <http://www.ece.ias-tate.edu/~namrata/research/SequentialCS.html>. All our code used CVX, www.stanford.edu/~boyd/cvx/.

A. Comparing CS-Residual With CS

We simulated a single time instant reconstruction problem (reconstruct x from $y := Ax + w$) with $m = 200$, $|N| = 20$, and with $|\Delta| = 0.1|N| = 2 = |\Delta_e|$. The noise w was zero mean i.i.d Gaussian. The nonzero signal values, x_N , were i.i.d. ± 1 with equal probability. The sets $N, \Delta \subseteq N$ and $\Delta_e \subseteq N^c$ were uniformly randomly generated each time. We used four

TABLE I

COMPARING NORMALIZED MSE OF CS-RESIDUAL (WITH $\lambda = 4\sigma$) WITH THAT OF CS (DANTZIG SELECTOR (DS)) WITH THREE DIFFERENT λ 'S. WE USED $m = 200$, $|N| = 20$, $|\Delta| = |\Delta_e| = 2$. COMPARISON SHOWN FOR THREE CHOICES OF $n = 45, 59, 100$ IN THE THREE

($n = 45$)	$n=45$ $\sigma=0.04$	$n=45$ $\sigma=0.09$	$n=45$ $\sigma=0.18$	$n=45$ $\sigma=0.44$	($n = 59$)	$n=59$ $\sigma=0.04$	$n=59$ $\sigma=0.09$	$n=59$ $\sigma=0.18$	$n=59$ $\sigma=0.44$	($n = 100$)	$n=100$ $\sigma=0.04$	$n=100$ $\sigma=0.09$
DS, $\lambda=12\sigma$	0.8235	0.8952	0.9794	1.0000	DS, $\lambda=12\sigma$	0.7572	0.8402	0.9937	1.0000	DS, $\lambda=12\sigma$	0.5856	0.8547
DS, $\lambda=4\sigma$	0.7994	0.8320	0.8642	0.9603	DS, $\lambda=4\sigma$	0.6545	0.6759	0.7991	0.9607	DS, $\lambda=4\sigma$	0.2622	0.4975
DS, $\lambda=0.4\sigma$	0.8071	0.8476	0.8762	1.0917	DS, $\lambda=0.4\sigma$	0.5375	0.5479	0.7086	1.0525	DS, $\lambda=0.4\sigma$	0.0209	0.0929
CS-residual	0.1397	0.1685	0.2270	0.5443	CS-residual	0.0866	0.1069	0.1800	0.4102	CS-residual	0.0402	0.0687

different noise standard deviations ($\sigma = 0.0439 * [1, 2, 4, 10]$) and three different choices of n (45, 59, 100). In Table I, we compare the normalized MSE (NMSE) of CS-residual output with that of CS. CS (Dantzig selector) was run with different choices of λ while for CS-residual we fixed $\lambda = 4\sigma$. Except when $n = 100$, in all other cases CS-residual outperforms CS significantly. For $n = 100$ (large n), if $\sigma = 0.04$, CS (with smallest λ) is better, and if $\sigma = 0.09$, both are similar.

A few other observations. Equation (1) When n is small, the best CS error occurs when we run it with the smallest λ . Smaller λ reduces the size of the feasible set and thus the ℓ_1 norm of the minimizer, \hat{x} , is larger, i.e., more of its elements are nonzero (if λ is too large, $\hat{x} = 0$ will be feasible and will be the solution). Equation (2) We also compared the CS-residual error with the error of the final LS-CS output (not shown). Only when CS-residual error was small, the support estimation was accurate and in this situation the final LS-CS error was much smaller.

B. Verifying LS-CS Stability

In Fig. 3, we verify LS-CS stability. We simulated Signal Model 1 with $m = 200$, $S_0 = 20$, $S_a = 2$ and with $d = 8$, $r = 2$, $M = 3$. Half the a_i 's were 0.5, the other half were 0.25. We used $n = 59$ and the noise was *uniform*($-c, c$) with $c = 0.0528$. The LS-CS algorithm used $\lambda = c\|A\|_1 = 0.35$, $\alpha = c$ and a larger $\alpha_{del} = 2.28c$. We assumed that the initialization condition holds, i.e., we started LS-CS with $\hat{N}_0 = N_0$.

In all 100 simulations, the number of misses and extras became exactly zero within $d_0 + S_a - 1 = 4$ time units of the addition time, i.e., the LS-CS estimate became equal to that of the genie-LS. Thus, according to the simulations, d_0 was 3.

The NMSE of LS-CS was stable below 0.4%. Compare this to that of CS or Gauss-CS with small λ , which was 30%–40%.

C. Lower SNR and Faster Additions

Next we ran two sets of simulations—slow-adds and fast-adds—with much lower SNRs. Slow-adds used $d = 8$, while fast-adds used $d = 3$. In all simulations, $m = 200$, $S_0 = 20$, $S_a = 2$ and the noise was *uniform*($-c, c$). Also, we used a smaller λ , $\lambda = c\|A\|_1/2$ since it encourages more additions.

We define two quantities: minimum average signal to noise ratio (min-SNR) and maximum average signal to noise ratio (max-SNR). Min (max) SNR is the ratio of minimum (maximum) average signal magnitude to the noise standard deviation. For *unif*($-c, c$) noise, the standard deviation is $c/\sqrt{3}$. Min-SNR, which occurs right after a new addition, decides how quickly new additions start getting detected (decides d_0). Max-SNR decides whether $|\Delta|$ becomes zero before the next addition. Both also depend on n of course.

For the previous subsection (Fig. 3), $c/\sqrt{3} = 0.03$. Minimum average signal magnitude was $(0.5 + 0.25)/2 = 0.375$ while

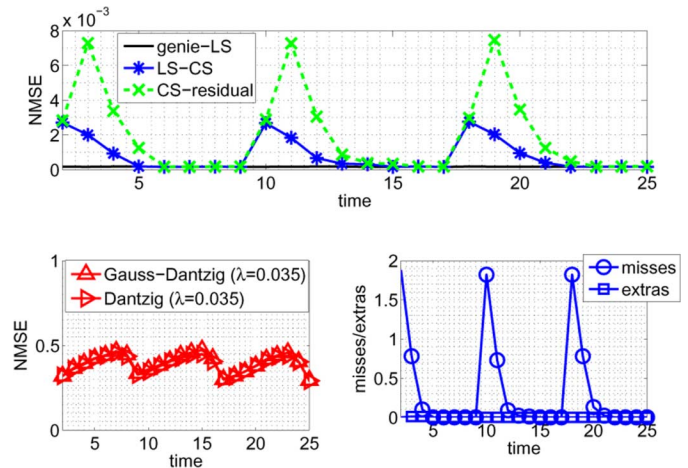


Fig. 3. Verifying LS-CS stability.

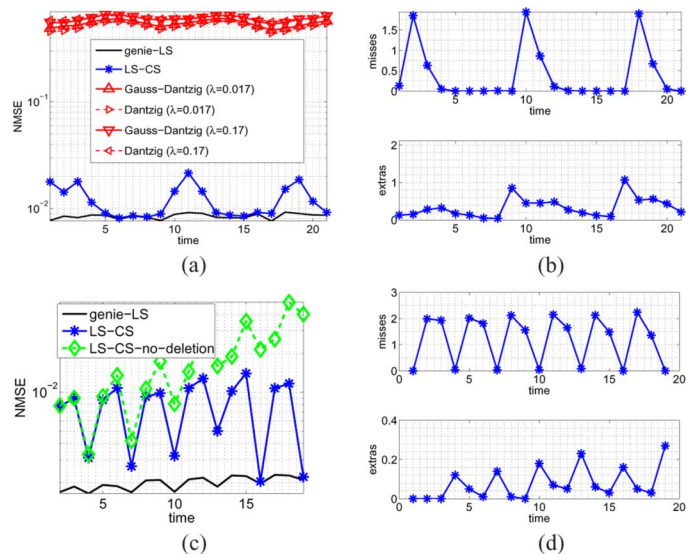


Fig. 4. Lower SNRs and Faster additions. LS-CS-no-deletion refers to LS-CS without deletion step. y axis is log scale in (a), (c). (a) Low SNR, Slow adds, $n = 59$. (b) Low SNR, Slow adds, $n = 59$. (c) Low SNR, Fast adds, $n = 59$. (d) Low SNR, Fast adds, $n = 59$.

maximum was $(M + d \min_i a_i)/2 = (3 + 8 * 0.25)/2 = 2.5$. Thus, *min-SNR* was 12.3 while *max-SNR* was 82.

In slow-adds [Fig. 4(a) and (b)], we use $n = 59$, $c = 0.1266$ and Signal Model 1 with $a_i = 0.2$, $M = 1$, $d = 8$ and $r = 3$. Thus, *min-SNR* was $0.2 * \sqrt{3}/0.1266 = 2.73$ while *max-SNR* was $1 * \sqrt{3}/0.1266 = 13.7$ (both are much smaller than 12.3 and 82, respectively). LS-CS used $\lambda = 0.176$, $\alpha = c/2 = 0.06 = \alpha_{del}$. We restricted the maximum number of detects at a time to $S_a + 1 = 3$. We also evaluated our assumption that CS

at $t = 0$ done with large enough n_0 finds the support without any error. With $n_0 = 150$, this was true 90% of the times, while in other cases there were one to two errors. Notice the following from Fig. 4(a) and (b): 1) Most additions get detected within two time units and there are occasionally a few extras, and 2) as long as $A_T' A_T$ remains well-conditioned, a few extras do not increase the error visibly above that of the genie-LS. Notice from the plots that even when $LS-CS \approx$ genie-LS, the average extras, $|\tilde{\Delta}_e|$, are not zero. (3) LS-CS error (NMSE) is stable at 2.5% while the CS errors are much larger at 40%–60%.

In fast-adds [Fig. 4(c), (d)], we use $n = 59$, $c = 0.0528$ and a slightly modified Signal Model 1 with $a_i = 0.2$, $M = 1$, $d = 3$ and $r = 2$. Thus, $\min SNR$ was $0.2 * \sqrt{3}/0.0528 = 6.6$ while $\max SNR$ was $0.6 * \sqrt{3}/0.0528 = 19.7$. Both are smaller than the stability simulation, but larger than the slow-adds simulation. This was needed because in this case the delay between addition times was only 3, and so quick detection was needed to ensure error stability. LS-CS error (NMSE) is still stable at 1%. LS-CS used $\lambda = 0.176$, $\alpha = c = 0.05 = \alpha_{del}$ and maximum allowed detects at a time of $S_a = 2$.

D. Dynamic MRI Reconstruction Example

To address a reviewer comment, in Fig. 5, we show the applicability of LS-CS to accurately reconstruct a sparsified cardiac image sequence from only 35% (simulated) MRI measurements. Clearly, LS-CS error is stable at 2% while that of CS is much larger. Detailed comparisons for actual (not sparsified) image sequences, using practical MR data acquisition schemes, and with using BPDN are given in [9].

For Fig. 5, the sparsity basis was the two-level Daubechies-4 2-D DWT. Images were 32×32 ($m = 1024$) and were sparsified by retaining the largest magnitude DWT coefficients that make up 99.5% of the total image energy and computing the inverse DWT. The support size of the sparsified DWT vector varied between 106–110, and the number of additions to (or removals from) the support from any $t - 1$ to t varied between 1–3. Denote the 1-D DWT matrix by W and the DFT matrix by F . Then $\Phi = W \otimes W$ and the measurement matrix, $H = M_{rs}(F \otimes F)/32$ where M_{rs} is an $n \times m$ random row selection matrix and \otimes denotes the Kronecker product. We used $n = 0.35 m$ and $n_0 = 0.8 m$. Noise was zero mean i.i.d. Gaussian with variance $\sigma^2 = 0.125$. Both LS-CS and CS used $\lambda = 1.5\sigma$. We also tried running CS with smaller values of λ : $\lambda = 0.15\sigma$ and $\lambda = 0.3\sigma$, but these resulted in (4) being infeasible.

VI. CONCLUSIONS AND FUTURE WORK

We formulated the problem of recursive reconstruction of sparse signal sequences from noisy observations as one of noisy CS with partly known support (the support estimate from the previous time serves as the “known” part). Our proposed solution, LS CS-residual (LS-CS), replaces CS on the raw observation by CS on the LS residual, computed using the known part of the support. We obtained bounds on CS-residual error. When the number of available measurements, n , is small, we showed that our bound is much smaller than the CS error bound if $|\Delta|$, $|\Delta_e|$ are small enough. We used this bound to prove the stability of LS-CS over time. By “stability” we mean that the support estimation errors, $|\tilde{\Delta}|$, $|\tilde{\Delta}_e|$, remain bounded by time-invariant

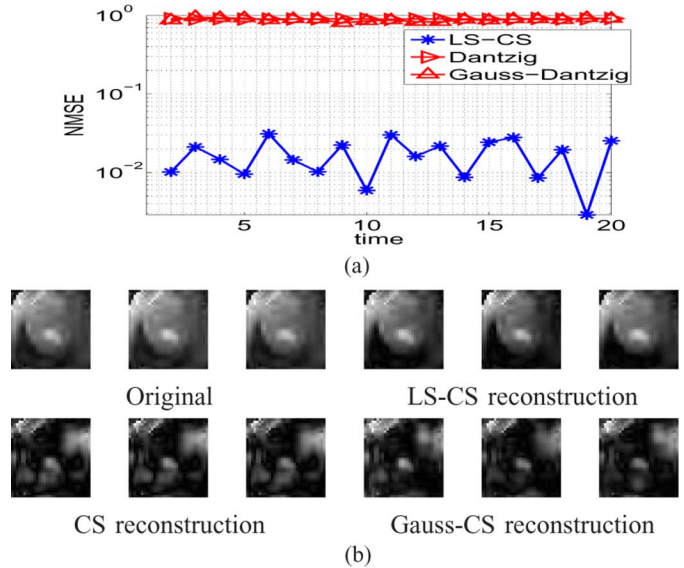


Fig. 5. Dynamic MRI. Reconstructing a sparsified cardiac sequence. (a) NMSE comparison (y axis is log scale). (b) Frames 2, 11, 20: original and reconstructed.

values. Extensive numerical experiments backing our claims are shown.

An open question is how to prove stability of LS-CS for a stochastic signal model that uses a random walk model with drift given by the current model for coefficient increase/decrease while using a (statistically) stationary model for “constant” coefficients, and that assumes a prior on support change, e.g., one can modify the model of [25]. An even more difficult open question is how to extend this analysis to show stability of KF-CS [1], [23] under similar assumptions (this is more difficult since KF-CS error also depends on the previous reconstructed signal, not just its support estimate).

In this work, we did not study exact reconstruction using much fewer noise-free measurements. We do this in [26] and [27].

APPENDIX

A. CS-Residual Bound: Proof of Lemma 1 and Theorem 1

Proof of Lemma 1: The proof is a modification of the proof of Theorem 1.3 given in [6]. Let $\delta \equiv \delta_{2S}$, $\theta \equiv \theta_{S,2S}$. Let $\hat{\zeta} = \zeta + h$. Let $T_0 \subseteq T_{nz}$ be a size S subset with $S \leq \min(S_{**}, S_{nz})$ and let $T_{rest} = T_{nz} \setminus T_0$. $\|w\|_\infty \leq (\lambda/\|A\|_1)$ implies that $|A_i' w| \leq \|A_i\|_1 \|w\|_\infty \leq \lambda$. Thus, [6, eq. (3.1)] holds with probability (w.p.) 1 and so ζ is feasible. Thus,

$$\|(h)_{T_0^c}\|_1 \leq \|(h)_{T_0}\|_1 + 2\|(\zeta)_{T_0^c}\|_1 \quad (25)$$

$$\|A' A h\|_\infty \leq 2\lambda. \quad (26)$$

The second equation is [6, eq. (3.3)]. The first follows by simplifying $\|\hat{\zeta}\|_1 \leq \|\zeta\|_1$ [6].

Recall that S_{**} is the largest value of S for which $\delta + \theta < 1$. Thus, we can apply [6, Lemma 3.1] for any $S \leq S_{**}$. Let T_1 contain the indexes of the S largest magnitude elements of $h := \hat{\zeta} - \zeta$ outside of T_0 . Let $T_{01} := T_0 \cup T_1$. Thus, $|T_{01}| = 2S$ and $\|h_{T_0}\|_k \leq \|h_{T_{01}}\|_k$ for any ℓ_k norm. Apply [6, Lemma 3.1] and use (25) and (26) to upper bound its first inequality. Then use $\|h_{T_0}\|_1/\sqrt{S} \leq \|h_{T_0}\| \leq \|h_{T_{01}}\|$ to simplify the resulting

inequality, and then use $(a+b)^2 \leq 2a^2 + 2b^2$ to square it. Finally, use $\|(\zeta)_{T_0^c}\|_1 = \|(\zeta)_{T_{\text{rest}}}\|_1$ to get

$$\|h_{T_0}\|^2 \leq \frac{16S\lambda^2}{(1-\delta-\theta)^2} + \frac{8\theta^2\|\zeta_{T_{\text{rest}}}\|_1^2}{(1-\delta-\theta)^2S}. \quad (27)$$

Using $(a+b)^2 \leq 2a^2 + 2b^2$ to simplify the square of (25); using the resulting bound in the second inequality of [6, Lemma 3.1]; and then finally using (27), we get

$$\|h\|^2 \leq \frac{48S\lambda^2}{(1-\delta-\theta)^2} + \left(8 + 24\frac{\theta^2}{(1-\delta-\theta)^2}\right) \frac{\|\zeta_{T_{\text{rest}}}\|_1^2}{S}. \quad (28)$$

Since $T_{\text{rest}} = T_{nz} \setminus T_0$ and $|T_0| = S$, thus $|T_{\text{rest}}| = S_{nz} - S$. Thus, $\|\zeta_{T_{\text{rest}}}\|_1^2 \leq (S_{nz} - S)\|\zeta_{T_{\text{rest}}}\|^2$. This gives our result which holds for any set $T_0 \subseteq T_{nz}$ of size $S \leq \min(S_{**}, S_{nz})$. ■

Proof of Theorem 1: The result follows by applying Lemma 1 with $\zeta = \beta$, $S_{nz} = |T| + |\Delta|$ and picking the set T_{rest} of size $|T| + |\Delta| - S$ as follows. For $S \geq |\Delta|$, pick $T_{\text{rest}} \subseteq T$ of size $|T| + |\Delta| - S$ and bound $\|\beta_{T_{\text{rest}}}\|$ by $\|\beta_T\|$. Use (15) to bound $\|\beta_T\|$, and use $\delta_{|T|} < 1/2$ to simplify the final expression. For $S < |\Delta|$, pick the set T_{rest} as the set T union with $|\Delta| - S$ smallest elements of x_Δ . Finally use $\hat{x}_{\text{CSres}} = \hat{\beta} + \hat{x}_{\text{init}}$ and $\beta = x - \hat{x}_{\text{init}}$ to get $\beta - \hat{\beta} = x - \hat{x}_{\text{CSres}}$. Lastly, from the definitions, $|T| = |N| + |\Delta_e| - |\Delta|$.

B. LS-CS Stability: Proofs of the Key Lemmas for Theorem 2

The proofs of the three lemmas essentially follow from Corollary 1 and the following simple facts.

- 1) An $i \in \Delta$ (an undetected element) will definitely get detected at current time if $x_i^2 > 2\alpha^2 + 2\|x - \hat{x}_{\text{CSres}}\|^2$.
- 2) An $i \in (\tilde{T}_{\text{det}} \setminus \tilde{\Delta}_{e,\text{det}})$ (a nonzero element of the current detected set) will definitely not get falsely deleted at the current time if $x_i^2 > 2\alpha_{\text{del}}^2 + 2\|(x - \hat{x}_{\text{det}})\tilde{T}_{\text{det}}\|^2$.
- 3) All $i \in \tilde{\Delta}_{e,\text{det}}$ (a zero element of the current detected set) will get deleted if $\alpha_{\text{del}}^2 \geq \|(x - \hat{x}_{\text{det}})\tilde{T}_{\text{det}}\|^2$.
- 4) If $\|w\|_\infty \leq \lambda/\|A\|_1$ and $|\tilde{T}_{\text{det}}| \leq S_*$, then $\|(x - \hat{x}_{\text{det}})\tilde{T}_{\text{det}}\|^2 \leq (4n\lambda^2/\|A\|_1^2) + 8\theta|\tilde{T}_{\text{det}}, \tilde{\Delta}_{\text{det}}|^2\|x_{\tilde{\Delta}_{\text{det}}}\|^2 \leq (4n\lambda^2/\|A\|_1^2) + 8\theta|\tilde{T}_{\text{det}}, \tilde{\Delta}_{\text{det}}|^2\|\tilde{\Delta}_{\text{det}}\| \|x_{\tilde{\Delta}_{\text{det}}}\|_\infty^2$.
- 5) The bound in fact 4 is non-decreasing in $|\tilde{T}_{\text{det}}|$ and $|\tilde{\Delta}_{\text{det}}|$.

Proof of Lemma 2: From Corollary 1 and the fact that $\|x_\Delta\|^2 \leq |\Delta|(x_\Delta)_{(1)}$, if $\|w\|_\infty \leq \lambda/\|A\|_1$, $|T| \leq S_*$ and $|\Delta| \leq S_{**}$, then $\|x - \hat{x}_{\text{CSres}}\|^2 \leq C' + C''\theta^2|\Delta|(x_\Delta)_{(1)}$ with C', C'', θ computed at $|T|, |\Delta|$. C', C'' are defined in (17).

Using fact 1 from above, the largest undetected element, $(x_\Delta)_{(1)}$, will definitely get detected at the current time if $(x_\Delta)_{(1)}^2 > 2\alpha^2 + 2C' + 2C''\theta^2|\Delta|(x_\Delta)_{(1)}$. Clearly this holds if $2\theta^2|\Delta|C'' < 1$ and $(2\alpha^2 + 2C'/1 - 2\theta^2|\Delta|C'') < (x_\Delta)_{(1)}$. If it is only known that $|T| \leq S_T$ and $|\Delta| \leq S_\Delta$ then our conclusion will hold if the maximum of the left-hand sides (LHS) over $|T| \leq S_T$ and $|\Delta| \leq S_\Delta$ is less than the right side. This gives the lemma. The LHS of the first inequality is non-decreasing in $|T|, |\Delta|$ and hence is maximized for S_T, S_Δ . The LHS of the second one is non-decreasing in $|T|$ but is not monotonic in $|\Delta|$.

³An $i \in \Delta$ will get detected if $|(\hat{x}_{\text{CSres}})_i| > \alpha$. Since $|(\hat{x}_{\text{CSres}})_i| \geq |x_i| - |x_i - (\hat{x}_{\text{CSres}})_i| \geq |x_i| - \|x - \hat{x}_{\text{CSres}}\|$, this holds if $|x_i| > \alpha + \|x - \hat{x}_{\text{CSres}}\|$. This, in turn, holds if $x_i^2 > 2\alpha^2 + 2\|x - \hat{x}_{\text{CSres}}\|^2$.

Proof of Lemma 3: It follows from facts 2, 4, and 5.

Proof of Lemma 4: It follows from facts 3, 4, and 5.

C. LS-CS Stability: Proof of Theorem 2

Let $t_0 = 0$ (call it the zeroth addition time). The first addition time, $t_1 = 1$. We prove Theorem 2 by induction. At $t = t_0 = 0$, all the $S_0 - S_a$ coefficients are correctly detected (according to the initialization condition), and thus $|\hat{\Delta}| = |\hat{\Delta}_e| = 0$ and $|\tilde{T}| = |N| = S_0 - S_a$. Thus, for the initial interval $t \in [t_0, t_1 - 1]$, our result holds. This proves the base case. Now for the induction step, assume that

Assumption 2 (Induction Step Assumption): The result holds for all $t \in [t_{j-1}, t_j - 1]$. Thus, at $t = t_j - 1$, $|\hat{\Delta}| = |\hat{\Delta}_e| = 0$ and $|\tilde{T}| = |N| = S_0 - S_a$.

Then prove that the result holds for $t \in [t_j, t_{j+1} - 1]$. The following facts will be frequently used in the proof.

- 1) Recall that $t_{j+1} = t_j + d$. Also, coefficient decrease of the elements of \mathcal{R} begins at $t_{j+1} - r = t_j + d - r$ and the coefficients get removed at $t_{j+1} - 1$. Since $d \geq d_0 + S_a + r$ (condition 7 of the theorem), thus, coefficient decrease does not begin until $t_j + d_0 + S_a$ or later.
- 2) At all $t \in [t_j, t_{j+1} - 2]$, $|N| = S_0$, while at $t = t_{j+1} - 1$, $|N| = S_0 - S_a$. Also, there are S_a additions at $t = t_j$ and none in the rest of the interval $[t_j, t_{j+1} - 1]$. There are S_a removals at $t = t_{j+1} - 1$, and none in the rest of the interval before that.
- 3) $\Delta_t \subseteq \tilde{\Delta}_{t-1} \cup (N_t \setminus N_{t-1})$ and $\Delta_{e,t} \subseteq \tilde{\Delta}_{e,t-1} \cup (N_{t-1} \setminus N_t)$. If there are no new additions, $\Delta_t = \tilde{\Delta}_{t-1}$. Similarly, if there are no new removals, $\Delta_{e,t} = \tilde{\Delta}_{e,t-1}$.

The induction step proof follows by combining the results of the following six claims. In each claim, we bound $|\hat{\Delta}|, |\hat{\Delta}_e|, |\tilde{T}|$ in one of the subintervals shown in Fig. 2. Using the last two facts above, the bounds for $|\Delta|, |\Delta_e|, |T|$ follow directly.

Claim 1: At all $t = t_j + i$, for all $i = 0, 1, \dots, d_0 - 1$, $|\tilde{\Delta}| \leq S_a, |\hat{\Delta}_e| \leq (i+1)f, |\tilde{T}| \leq S_0 + (i+1)f$.

Proof: We prove this by induction. Consider the base case, $t = t_j$. At this time there are S_a new additions and $|N| = S_0$. Using Assumption 2 (induction step assumption), $|\Delta| = S_a, |\Delta_e| = 0$. In the detection step, $|\hat{\Delta}_{\text{det}}| \leq |\Delta| = S_a$ and so $\|x_{\tilde{\Delta}_{\text{det}}}\|_\infty \leq \min(M, (a_{\mathcal{A}})_{(1)})$. There are at most f false detects (condition 2), so that $|\tilde{\Delta}_{e,\text{det}}| \leq 0 + f$. Thus, $|\tilde{T}_{\text{det}}| = |N| + |\tilde{\Delta}_{e,\text{det}}| - |\hat{\Delta}_{\text{det}}| \leq S_0 + f$.

The smallest constant coefficient has magnitude $\min(M, d \min_i a_i)$. Apply Lemma 3 with $S_T = S_0 + f, S_\Delta = S_a, b_1 = \min(M, d \min_i a_i)$. It is applicable since conditions 3a and 5 hold. Thus, none of the constant coefficients will get falsely deleted and so $|\hat{\Delta}| \leq S_a$. Also, clearly $|\hat{\Delta}_e| \leq |\hat{\Delta}_{e,\text{det}}| \leq f$. Thus, $|\tilde{T}| \leq S_0 + f$.

For the induction step, assume that the result holds for $t_j + i - 1$. Thus, at $t = t_j + i, |\Delta_{e,t}| = |\hat{\Delta}_{e,t-1}| \leq if$ and $|\Delta_t| = |\hat{\Delta}_{t-1}| \leq S_a$. Using condition 2, after the detection step, $|\hat{\Delta}_{e,\text{det}}| \leq (i+1)f$. Thus, $|\tilde{T}_{\text{det}}| \leq S_0 + (i+1)f$. Also, $|\hat{\Delta}_{\text{det}}| \leq S_a$ and so $\|x_{\tilde{\Delta}_{\text{det}}}\|_\infty \leq \min(M, (i+1)(a_{\mathcal{A}})_{(1)})$. Applying Lemma 3 with $S_T = S_0 + (i+1)f, S_\Delta = S_a, b_1 = \min(M, d \min_i a_i)$ (applicable since conditions 3a and 5 hold), none of the constant coefficients will get falsely deleted. Thus, $|\hat{\Delta}| \leq S_a$. Also, clearly $|\hat{\Delta}_e| \leq |\hat{\Delta}_{e,\text{det}}| \leq (i+1)f$. Thus, $|\tilde{T}| \leq S_0 + (i+1)f$.

Claim 2: At $t = t_j + d_0 + i - 1$, for all $i = 1, \dots, S_a$, $|\tilde{\Delta}| \leq S_a - i$, $|\tilde{\Delta}_e| \leq (d_0 + i)f$, $|\tilde{T}| \leq S_0 + (d_0 + i)f$, and the first i largest increasing coefficients are definitely detected.

Proof: We prove this by induction. Consider the base case $t = t_j + d_0$. Using the previous claim, $|\Delta| \leq S_a$, $|\Delta_e| \leq d_0f$, $|T| \leq S_0 + d_0f$. At this time, either the largest element of \mathcal{A} , which has magnitude $\min(M, (d_0 + 1)(a_{\mathcal{A}})_{(1)})$, has already been detected so that the number of undetected elements already satisfies $|\Delta| \leq S_a - 1$ or it has not been detected. If it has been detected, then $|\tilde{\Delta}_{\text{det}}| \leq |\Delta| \leq S_a - 1$. If it has not been detected, then $(x_{\Delta})_{(1)} = \min(M, (d_0 + 1)(a_{\mathcal{A}})_{(1)})$. Apply Lemma 2 with $S_{\Delta} = S_a$, $S_T = S_0 + d_0f$. It is applicable since conditions 3a and 3b hold and condition 4a holds for $i = 1$. Thus, the largest element will definitely get detected. Thus, in all cases, $|\tilde{\Delta}_{\text{det}}| \leq S_a - 1$ and so $\|x_{\tilde{\Delta}_{\text{det}}}\|_{\infty} \leq \min(M, (d_0 + 1)(a_{\mathcal{A}})_{(2)})$. Using condition 2, $|\tilde{\Delta}_{e,\text{det}}| \leq (d_0 + 1)f$ and so $|\tilde{T}_{\text{det}}| \leq S_0 + (d_0 + 1)f$.

Applying Lemma 3 with $S_T = S_0 + (d_0 + 1)f$, $S_{\Delta} = S_a - 1$, $b_1 = \min(M, (d_0 + 1)(a_{\mathcal{A}})_{(1)})$ (applicable since condition 3a holds and 4b holds for $i = 1$), the largest increasing coefficient will not get falsely deleted. Further, applying Lemma 3 with $b_1 = \min(M, d \min_i a_i)$ (applicable since conditions 3a and 5 hold), none of the constant coefficients will get falsely deleted. Thus, $|\tilde{\Delta}| \leq S_a - 1$. Also $|\tilde{\Delta}_e| \leq |\tilde{\Delta}_{e,\text{det}}| \leq (d_0 + 1)f$ and so $|\tilde{T}| \leq S_0 + (d_0 + 1)f$.

For the induction step, assume that the result holds for $t_j + d_0 + i - 2$. Thus, at $t = t_j + d_0 + i - 1$, $|\Delta| \leq S_a - i + 1$, $|\Delta_e| \leq (d_0 + i - 1)f$, $|T| \leq S_0 + (d_0 + i - 1)f$ and the first $i - 1$ largest elements have already definitely been detected. Either the i th largest element has also been already detected, in which case $|\Delta| \leq S_a - i$ or it has not been detected. If it has, then $|\tilde{\Delta}_{\text{det}}| \leq |\Delta| \leq S_a - i$. If it has not been detected, then $(x_{\Delta})_{(1)} = \min(M, (d_0 + i)(a_{\mathcal{A}})_{(i)})$. As before, use conditions 3a, 3b, and 4a and apply Lemma 2 to claim that the i th largest element will definitely get detected. Thus, in all cases, $|\tilde{\Delta}_{\text{det}}| \leq S_a - i$ and so $\|x_{\tilde{\Delta}_{\text{det}}}\|_{\infty} \leq \min(M, (d_0 + i)(a_{\mathcal{A}})_{(i+1)})$. Using condition 2, $|\tilde{\Delta}_{e,\text{det}}| \leq (d_0 + i)f$ and so $|\tilde{T}_{\text{det}}| \leq S_0 + (d_0 + i)f$. Also as before, apply Lemma 3 first with $b_1 = \min(M, (d_0 + i)(a_{\mathcal{A}})_{(i)})$ and then with $b_1 = \min(M, d \min_i a_i)$ (applicable since conditions 3a, 4b, and 5 hold) to claim that all constant coefficients and all the i largest increasing coefficients will not get falsely deleted. Thus, $|\tilde{\Delta}| \leq S_a - i$. Also, $|\tilde{\Delta}_e| \leq (d_0 + i)f$ and $|\tilde{T}| \leq S_0 + (d_0 + i)f$.

Claim 3: At $t = t_j + d_0 + S_a - 1$, $|\tilde{\Delta}_e| = 0$.

Proof: In the previous proof we have shown that at $t = t_j + d_0 + S_a - 1$, i.e., for $i = S_a$, $|\tilde{\Delta}_{\text{det}}| = 0$ and $|\tilde{T}_{\text{det}}| \leq S_0 + (d_0 + S_a)f$. Apply Lemma 4 with $S_{\Delta} = 0$, $S_T = S_0 + (d_0 + S_a)f$ (applicable since conditions 3a and 2 hold). Thus, all false detects will get deleted, i.e., $|\tilde{\Delta}_e| = 0$.

Claim 4: At all $t \in [t_j + d_0 + S_a - 1, t_{j+1} - r - 1]$, $|\tilde{\Delta}| = 0$, $|\tilde{\Delta}_e| = 0$. Thus, $\tilde{T} = N_t$ and $|\tilde{T}| = |N_t| = S_0$.

Proof: Using the previous two claims, the result holds for $t = t_j + d_0 + S_a - 1$ (base case). For the induction step, assume that it holds for $t_j + d_0 + S_a + i - 1$. Thus, at $t = t_j + d_0 + S_a + i$, $|\Delta| = 0$, $|\Delta_e| = 0$ and $|T| = S_0$. Since $|\tilde{\Delta}_{\text{det}}| \leq |\Delta|$, $|\tilde{\Delta}_{\text{det}}| = 0$ and thus $\|x_{\tilde{\Delta}_{\text{det}}}\|_{\infty} = 0$. Using condition 2, $|\tilde{\Delta}_{e,\text{det}}| \leq 0 + f$ and thus $|\tilde{T}_{\text{det}}| \leq S_0 + f$. Use conditions 3a, 4b (for $i = S_a$), and 5 to first apply Lemma 3 with $S_T = S_0 + f$, $S_{\Delta} = 0$,

$b_1 = \min(M, (d_0 + S_a + i + 1)(a_{\mathcal{A}})_{(S_a)})$ (smallest increasing coefficient) and then with $b_1 = \min(M, d \min_i a_i)$ (smallest constant coefficient) to show that there are no false deletions of either constant or increasing coefficients. Thus, $|\tilde{\Delta}| = 0$. Use conditions 3a and 2 and apply Lemma 4 with $S_{\Delta} = 0$, to show that $|\tilde{\Delta}_e| = 0$.

Claim 5: At $t \in [t_{j+1} - r, t_{j+1} - 1]$, $|\tilde{\Delta}| = 0$, $|\tilde{\Delta}_e| = 0$. Thus, $\tilde{T} = N_t$ and $|\tilde{T}| = |N_t| = S_0$.

Proof: The proof again follows by induction and arguments similar to those of the previous claim. The only difference is the following. At any $t = t_{j+1} - r + i - 1$, one applies Lemma 3 three times: the first two times for increasing and constant coefficients (as before) and then a third time with $S_T = S_0 + f$, $S_{\Delta} = 0$, $b_1 = ((i - 1)/r) \min(M, d(a_{\mathcal{R}})_{(S_a)})$ (for the current smallest decreasing coefficient). This last one is applicable since conditions 3a and 6 hold.

Claim 6: At $t = t_{j+1} - 1$, $|\tilde{\Delta}| = 0$, $|\tilde{\Delta}_e| = 0$. Thus, $\tilde{T} = N_t$ and $|\tilde{T}| = |N_t| = S_0 - S_a$.

The only difference at this time is that the decreasing coefficients get removed. As a result, $|N_t| = S_0 - S_a$, $|\Delta_e| = S_a$ and $|\Delta_{e,\text{det}}| = S_a + f$. But $|\Delta_{\text{det}}| = |\Delta| = 0$. As before, using conditions 3a and 2 and applying Lemma 4 with $S_{\Delta} = 0$, all extras will still get removed and so still $|\tilde{\Delta}_e| = 0$. Everything else is the same as before.

D. LS-CS Stability: Proof of Corollary 3

We have shown that $|\tilde{T}| \leq S_0 + f(d_0 + S_a)$ and $|\tilde{\Delta}| \leq S_a$. We can bound $\|x_{\tilde{\Delta}}\|$ as follows. In the first subinterval, $|\tilde{\Delta}| \leq S_a$ and the maximum value of any element of $\tilde{\Delta}_t$ at any t in this interval is $\min(M, d_0(a_{\mathcal{A}(j)})_{(1)})$ so that $\|x_{\tilde{\Delta}}\|^2 \leq S_a \min(M, d_0(a_{\mathcal{A}(j)})_{(1)})^2$. In the second subinterval, at $t = t_j + d_0 + i - 1$, $|\tilde{\Delta}| \leq S_a - i$ and $\|x_{\tilde{\Delta}}\|_{\infty} \leq \min(M, (d_0 + i)(a_{\mathcal{A}(j)})_{(i+1)})$. In the last two subintervals, $|\tilde{\Delta}| = 0$. Thus,

$$\begin{aligned} \|x_{\tilde{\Delta}}\|^2 &\leq \max_j \max_{i=0, \dots, S_a} (S_a - i) \min(M, (d_0 + i)(a_{\mathcal{A}(j)})_{(i+1)})^2 \\ &\leq S_a \min(M, (d_0 + S_a) \max_i a_i)^2. \end{aligned} \quad (29)$$

This gives the LS-CS error bound. In a similar fashion, we can argue that $\|x_{\Delta}\|^2 \leq S_a \min(M, (d_0 + S_a) \max_i a_i)^2$. Using this in Corollary 1 gives the CS-residual error bound.

REFERENCES

- [1] N. Vaswani, "Kalman filtered compressed sensing," in *Proc. ICIP*, Oct. 2008, pp. 893–896.
- [2] N. Vaswani, "Analyzing least squares and Kalman filtered compressed sensing," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Apr. 2009, pp. 3013–3016.
- [3] E. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [4] A. J. Martin, O. M. Weber, D. Saloner, R. Higashida, M. Wilson, M. Saeed, and C. B. Higgins, "Application of MR technology to endovascular interventions in an XMR suite," *Medica Mundi*, vol. 46, pp. 28–34, Dec. 2002.
- [5] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, Dec. 2007.
- [6] E. Candes and T. Tao, "The dantzig selector: Statistical estimation when p is much larger than n ," *Ann. Stat.*, vol. 35, no. 6, pp. 2313–2351, 2007.
- [7] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, "An architecture for compressive imaging," in *Proc. ICIP*, Oct. 2006, pp. 1273–1276.

- [8] C. Qiu, W. Lu, and N. Vaswani, "Real-time dynamic MRI reconstruction using Kalman filtered CS," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Apr. 2009, pp. 393–396.
- [9] C. Qiu and N. Vaswani, "Compressive sensing on the least squares and Kalman filtering residual for real-time dynamic MRI and video reconstruction," *IEEE Trans. Image Process.*, 2009, submitted for publication.
- [10] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [11] S. S. Chen, Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33–61, 1998.
- [12] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
- [13] E. Candes, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.
- [14] 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.
- [15] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med.*, vol. 59, no. 2, pp. 365–373, Jan. 2008.
- [16] S. Sarvotham, D. Baron, and R. Baraniuk, "Compressed sensing reconstruction via belief propagation," *Electr. Comput. Eng. Dept., Rice Univ., Houston, TX, Tech. Rep. ECE-06-01*, 2006, .
- [17] C. Rozell, D. Johnson, R. Baraniuk, and B. Olshausen, "Locally competitive algorithms for sparse approximation," in *Proc. ICIP*, Oct. 2007, pp. 169–172.
- [18] M. S. Asif and J. Romberg, "Dynamic updating for sparse time varying signals," in *Proc. CISS*, Mar. 2009, pp. 3–8.
- [19] D. Angelosante and G. B. Giannakis, "RLS-weighted Lasso for adaptive estimation of sparse signals," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Apr. 2009, pp. 3245–3248.
- [20] H. Jung, K. H. Sung, K. S. Nayak, E. Y. Kim, and J. C. Ye, " k - t FO-CUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magn. Reson. Med.*, vol. 61, no. 1, pp. 103–116, Jan. 2009.
- [21] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed sensing image reconstruction via recursive spatially adaptive filtering," in *Proc. ICIP*, Oct. 2007, pp. 549–552.
- [22] D. Baron, M. Wakin, M. Duarte, S. Sarvotham, and R. Baraniuk, "Distributed compressed sensing," in *Proc. Allerton Conf.*, Sep. 2005, pp. 814–825.
- [23] N. Vaswani, "KF-CS: Compressive sensing on Kalman filtering residual," *ArXiv Preprint arXiv:0912.1628*, 2009.
- [24] N. Vaswani, "Stability (over time) of modified-CS and LS-CS for recursive casual sparse reconstruction," *ArXiv Preprint*, 2010.
- [25] P. Schniter, L. Potter, and J. Ziniel, "Fast Bayesian matching pursuit: Model uncertainty and parameter estimation for sparse linear models," *Proc. Information Theory Applications (ITA)*, pp. 326–333, Feb. 2008.
- [26] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," in *Proc. IEEE Int. Symp. Info. Theory (ISIT)*, Jul. 2009, pp. 488–492.
- [27] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Signal Process.*, 2010, to be published.



Namrata Vaswani received the B.Tech. degree from the Indian Institute of Technology (IIT), Delhi, in 1999 and the Ph.D. degree from the University of Maryland, College Park, in 2004, both in electrical engineering.

She spent one year as a Postdoctoral Researcher and Research Scientist at Georgia Tech. Since fall 2005, she has been an Assistant Professor in the Electrical and Computer Engineering Department at Iowa State University, Ames. Her research interests

are in estimation problems in statistical and sequential signal processing and biomedical imaging. Her current research focus is on recursive sparse reconstruction, sequential compressive sensing, and large dimensional tracking problems.

Dr. Vaswani has been serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING since 2009.