

ANALYZING LEAST SQUARES AND KALMAN FILTERED COMPRESSED SENSING

Namrata Vaswani

Dept. of ECE, Iowa State University, Ames, IA, namrata@iastate.edu

ABSTRACT

In recent work, we studied the problem of causally reconstructing time sequences of spatially sparse signals, with unknown and slow time-varying sparsity patterns, from a limited number of linear “incoherent” measurements. We proposed a solution called Kalman Filtered Compressed Sensing (KF-CS). The key idea is to run a reduced order KF only for the current signal’s estimated nonzero coefficients’ set, while performing CS on the Kalman filtering error to estimate new additions, if any, to the set. KF may be replaced by Least Squares (LS) estimation and we call the resulting algorithm LS-CS. In this work, (a) we bound the error in performing CS on the LS error and (b) we obtain the conditions under which the KF-CS (or LS-CS) estimate converges to that of a genie-aided KF (or LS), i.e. the KF (or LS) which knows the true nonzero sets.

Keywords: compressed sensing, kalman filter, least squares

1. INTRODUCTION

In recent work [1], we studied the problem of causally reconstructing time sequences of sparse signals, with unknown and slow time-varying sparsity patterns, from a limited number of noise-corrupted “incoherent” measurements. We proposed a solution called Kalman Filtered Compressed Sensing (KF-CS). With the exception of CS [2] and of [3], most other work [4, 5] treats the entire time sequence of signals/images as a single spatiotemporal signal and performs CS to reconstruct it. This is a non-causal solution and also has high computational cost. On the other hand, if the number of observations is small, performing CS [2] at each time (simple CS) incurs much larger error than KF-CS, see Fig. 1. Potential applications of KF-CS include making dynamic MRI real-time (causal and fast enough) [4, 6]; real-time video imaging using a single-pixel camera [5]; or real-time tracking of temperature, or other, time-varying fields using sensor networks that transmit random projections of the field [7].

In this work, in Sec. 2, we describe a simple modification of KF-CS [1] and introduce its non-Bayesian version, Least Squares (LS)-CS. Our *key contributions* are: (a) in Sec. 3, we bound the error in performing CS on the LS error in the observation and compare it with that for performing CS on the observation (simple CS), and (b) in Sec. 4, we obtain the conditions under which the KF-CS (or LS-CS) estimate converges to that of a genie-aided KF (or LS). Simulation comparisons are given in Sec. 5.

Problem Definition. The problem definition is the same as in [1]. Let $(z_t)_{m \times 1}$ denote the spatial signal of interest at time t and $(y_t)_{n \times 1}$, with $n < m$, denote its noise-corrupted observation vector at t . 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 a sparse vector (only $S_t \ll m$ elements of x_t are non-zero). The observations are “incoherent” w.r.t. the sparsity basis of the signal, i.e.

$$y_t = Ax_t + w_t, A \triangleq H\Phi, \mathbb{E}[w_t] = 0, \mathbb{E}[w_t w_t'] = \sigma_{obs}^2 I \quad (1)$$

This research was partially supported by NSF grant ECCS-0725849

and all S_t -column sub-matrices of A “approximately orthonormal”, i.e. $\delta_{S_t} < 1$ [2, eq. (1.3)]. w_t is independent of x_t and is i.i.d, $\forall t$.

Let N_t denote the support set of x_t , i.e. the set of its non-zero coordinates and let $T_t \triangleq \hat{N}_t$ denote its estimate. Also, let Δ_t denote the undetected nonzero set at time t , i.e. $\Delta_t \triangleq N_t \setminus T_{t-1}$ and let $\hat{\Delta}_t$ denote its estimate. Thus $T_t = T_{t-1} \cup \hat{\Delta}_t$. Let $S_t \triangleq |N_t|$ where $|\cdot|$ denotes the size of a set. Also, for any set T , let $(v)_T$ denote the $|T|$ length sub-vector containing the elements of v corresponding to the indices in the set T . For a matrix A , A_T denotes the sub-matrix obtained by extracting the columns of A corresponding to the indices in T . We use the notation $(Q)_{T_1, T_2}$ to denote the sub-matrix of Q containing rows and columns corresponding to the entries in T_1 and T_2 respectively. T^c denotes the complement of T w.r.t. $[1 : m]$, i.e. $T^c \triangleq [1 : m] \setminus T$. ϕ refers to the empty set. $'$ denotes transpose. The $m \times m$ matrix I_T is defined as follows: $(I_T)_{T, T} = I$ where I is a $|T|$ -identity matrix while $(I_T)_{T^c, [1:m]} = \mathbf{0}$, $(I_T)_{[1:m], T^c} = \mathbf{0}$.

The nonzero coefficients’ set N_t changes slowly over time. For the currently nonzero coefficients of x_t , $(x_t)_{N_t}$, we assume a spatially i.i.d. Gaussian random walk model, while the rest of the coefficients remain constant, i.e.

$$x_0 = \mathbf{0}, x_t = x_{t-1} + \nu_t, \nu_t \sim \mathcal{N}(0, Q_t), Q_t = \sigma_{sys}^2 I_{N_t} \quad (2)$$

where ν_t is temporally i.i.d.. The current nonzero set, N_t , is unknown $\forall t$. Our goal is to recursively get the best estimates of N_t and x_t (or equivalently of the signal, $z_t = \Phi x_t$) using y_1, \dots, y_t .

2. KALMAN FILTERED CS AND LEAST SQUARES CS

We describe a simple modification of KF-CS [1] and introduce Least Squares CS. Let $\hat{x}_{t|t-1}$, \hat{x}_t , K_t and $P_{t|t-1}$, P_t denote the predicted and updated state estimates at time t , the Kalman gain and the prediction and updated error covariances given by the KF in KF-CS (since KF-CS does not always use the correct value of Q_t , $P_{t|t-1}$ or P_t are not equal to the actual covariances of $x_t - \hat{x}_{t|t-1}$ or $x_t - \hat{x}_t$).

2.1. Modified Kalman Filtered Compressed Sensing (KF-CS)

KF-CS can be summarized as running a KF for the system in (1), (2) but with Q_t replaced by $\hat{Q}_t = \sigma_{sys}^2 I_{T_t}$. The new additions, if any, are estimated by performing CS on the Kalman filtering error, $\tilde{y}_{t,f}$.

At time t , we first run a “temporary” Kalman prediction and update step using $\hat{Q}_t = \sigma_{sys}^2 I_{T_{t-1}}$, i.e. we compute

$$\begin{aligned} K_{t,tmp} &= (P_{t-1} + \hat{Q}_t)A'(A(P_{t-1} + \hat{Q}_t)A' + \sigma_{obs}^2 I)^{-1} \\ \hat{x}_{t,tmp} &= (I - K_{t,tmp}A)\hat{x}_{t-1} + K_{t,tmp} y_t \end{aligned} \quad (3)$$

Let $T \triangleq T_{t-1}$. The filtering error is

$$\tilde{y}_{t,f} \triangleq y_t - A\hat{x}_{t,tmp} = A_{\Delta_t}(x_t)_{\Delta_t} + A_T(x_t - \hat{x}_t)_T + w_t \quad (4)$$

As explained in [1], if the filtering error norm is large, there is a need to estimate Δ_t . One can rewrite $\tilde{y}_{t,f}$ as $\tilde{y}_{t,f} = A\beta_t + w_t$, where $\beta_t \triangleq [(x_t - \hat{x}_t)_T, (x_t)_{\Delta_t}, 0_{(T \cup \Delta_t)^c}]$ is a “sparse-compressible”

Algorithm 1 Kalman Filtered Compressive Sensing (KF-CS)

Initialization: Set $\hat{x}_0 = 0$, $P_0 = 0$, $T_0 = \phi$ (if unknown) or equal to the known support. For $t > 0$, do,

1. **Temporary Kalman prediction and update.** Implement (3) using $\hat{Q}_t = \sigma_{sys}^2 I_{T_{t-1}}$.
 2. **Compute Additions using CS.** Compute the KF error, $\tilde{y}_{t,f} \triangleq y_t - A\hat{x}_{t,tmp}$. Check if $FEN \triangleq \tilde{y}_{t,f}' \Sigma_{f,t}^{-1} \tilde{y}_{t,f} > \alpha_{fe}$. If it is,
 - (a) Do CS on $\tilde{y}_{t,f}$ followed by thresholding, i.e. compute $\hat{\Delta}_t$ using (5), (6). The new estimated support is $T_t = T_{t-1} \cup \hat{\Delta}_t$.
 3. **Kalman prediction and update.** Implement (7) using $\hat{Q}_t = \sigma_{sys}^2 I_{T_t}$.
 - (a) (KF-CS with final LS): If $T_t \neq T_{t-1}$, implement (7) using $\hat{Q}_t = \infty I_{T_t}$, i.e. set $\hat{x}_t = (A_{T_t}' A_{T_t})^{-1} A_{T_t}' y_t$ and $(P_t)_{T_t, T_t} = (A_{T_t}' A_{T_t})^{-1} \sigma_{obs}^2$, $(P_t)_{T_t^c, :} = 0$, $(P_t)_{:, T_t^c} = 0$.
 4. **Compute Deletions.** If $T_t = T_{t-1} \dots = T_{t-k}$ (nonzero set has not changed for long enough, i.e. w.h.p. KF stabilized),
 - (a) Check for “zero” coefficients, i.e. compute $\hat{\Delta}_Z = \{i \in T_t : \sum_{\tau=t-k'+1}^t (\hat{x}_{\tau,i})^2 / k' < \alpha_z\}$ with $k' < k$. Set $T_t \leftarrow T_t \setminus \hat{\Delta}_Z$. Set $(\hat{x}_t)_{\hat{\Delta}_Z} = 0$. Set $(P_t)_{\hat{\Delta}_Z, [1:m]} = 0$ and $(P_t)_{[1:m], \hat{\Delta}_Z} = 0$.
 5. **Output T_t , \hat{x}_t and the signal estimate, $\hat{z}_t = \Phi \hat{x}_t$.** Increment t and go to the first step.
-

signal with a “large” or “non-compressible” nonzero part, $(x_t)_{\Delta_t}$, and a “small” or “compressible” nonzero part, $(x_t - \hat{x}_t)_T$. The Dantzig selector (DS) [2] followed by thresholding can be applied to detect the “non-compressible” nonzero part as follows:

$$\hat{\beta}_t = \arg \min_{\beta} \|\beta\|_1, s.t. \|A'(\tilde{y}_{t,f} - A\beta)\|_{\infty} \leq \lambda_m \sigma_{obs} \quad (5)$$

$$\hat{\Delta}_t = \{i \in T_{t-1}^c : \hat{\beta}_{t,i}^2 > \alpha_a\} \quad (6)$$

where $\lambda_m \triangleq \sqrt{2 \log m}$ and α_a is the addition threshold. Thus, the estimated support set at time t is $T_t = T \cup \hat{\Delta}_t = T_{t-1} \cup \hat{\Delta}_t$.

Next we run the Kalman prediction/update using $\hat{Q}_t = \sigma_{sys}^2 I_{T_t}$:

$$\begin{aligned} P_{t|t-1} &= P_{t-1} + \hat{Q}_t, \quad K_t = P_{t|t-1} A' (A P_{t|t-1} A' + \sigma_{obs}^2 I)^{-1} \\ P_t &= (I - K_t A) P_{t|t-1} \\ \hat{x}_t &= (I - K_t A) \hat{x}_{t-1} + K_t y_t \end{aligned} \quad (7)$$

with initialization $P_0 = \mathbf{0}_{[1:m], [1:m]}$, $\hat{x}_0 = \mathbf{0}_{[1:m]}$.

Remark 1 For easy notation, in (3),(7) we write the KF equations for the entire x_t . But actually we are running a reduced order KF for only the coefficients in T ($T \equiv T_{t-1}$ for (3) and $T \equiv T_t$ for (7)).

2.1.1. Deleting Zero Coefficients

If the addition threshold, α_a , is not large enough, occasionally there will be some false additions (coefficients whose true value is zero but they wrongly get added due to error in the CS step). Also, there may be coefficients that actually become and remain zero. All such coefficients need to be detected and removed from T_t to prevent unnecessary increase in $|T_t|$. Increased $|T_t|$ implies smaller minimum eigenvalue of $A_{T_t}' A_{T_t}$ and thus increased estimation error. The increase is especially large if $A_{T_t}' A_{T_t}$ is close to becoming singular.

One possible way to detect if a coefficient, i , is zero is to check if the magnitude of its estimates in the last few time instants is small, e.g. one can check if $\sum_{\tau=t-k'+1}^t (\hat{x}_{\tau,i})^2 / k' < \alpha_z$. This scheme would be fairly accurate (small enough false alarm and miss probabilities), if the estimation error, $e_{\tau,i} = x_{\tau,i} - \hat{x}_{\tau,i}$ is small enough, for all $\tau \in [t - k' + 1, t]$. If we check for zeroing only when T_t has not changed for long enough (w.h.p. this implies that all past additions have been detected, i.e. $T_t = N_t$, and the KF for T_t has stabilized), the variance of $e_{\tau,i}$ would be approximately equal to $(P_t)_{i,i} < \sigma_{obs}^2 / \lambda_{min}(A_T' A_T)$, i.e. it would be small enough.

When a coefficient, i , is detected as being zero, we remove it from T_t , we set $\hat{x}_{t,i} = 0$ and we set $(P_t)_{i, [1:m]} = 0$, $(P_t)_{[1:m], i} = 0$. We summarize the entire KF-CS algorithm in Algorithm 1.

2.2. Least Squares CS: Non-Bayesian KF-CS

In applications where training data is not be available to learn the prior model parameters required by KF-CS, one can use a non-Bayesian version of KF-CS i.e. replace the KF in KF-CS by Least Squares (LS) estimation. The LS step is also faster than the KF step.

3. ANALYZING CS ON LS ERROR (LSE)

Let $T \triangleq T_{t-1}$ and $\Delta \triangleq \Delta_t = N_t \setminus T_{t-1}$. The true nonzero sets at any time, N_t , are assumed to be non-random. But $T = T_{t-1}$ is a random variable since its value depends on y_{t-1} and T_{t-2} (or equivalently on $y_{1:t-1}$). We use $\mathbb{E}[\cdot]$ to denote expectation w.r.t. all random quantities $(y_{1:t}, x_{1:t})$ at time t while using $\mathbb{E}[\cdot | y_{1:t-1}]$ to denote the expected value conditioned on $y_{1:t-1}$. Conditioned on $y_{1:t-1}$, the set T , and hence also the set $\Delta = N_t \setminus T$, is known.

The key difference between simple CS and LS-CS is that simple CS applies (5) on $y_t = Ax_t + w_t$ to estimate the $|N_t|$ -sparse signal, x_t , while LS-CS applies (5) on the LS error (LSE), $\tilde{y}_{t,f} := y_t - A\hat{x}_{t,tmp} = A\beta_t + w_t$ to estimate $\beta_t := x_t - \hat{x}_{t,tmp}$, where $\hat{x}_{t,tmp} = (A_T' A_T)^{-1} A_T' y_t$. $\beta_t = [(x_t - \hat{x}_{t,tmp})_T, (x_t)_{\Delta}, 0_{T \cup \Delta^c}] = (A_T' A_T)^{-1} A_T' (A_{\Delta}(x_t)_{\Delta} + w_t), (x_t)_{\Delta}, 0_{T \cup \Delta^c}]$ is what we call a “sparse-compressible” signal: it is $|T \cup \Delta|$ -sparse but, if the sparsity pattern changes slowly enough, it is compressible along T . We use this idea to bound the error in CS on LSE and to show that if the sparsity pattern changes slowly enough, the CS-LSE error bound is much smaller than that of simple CS.

We use the following definition of compressibility of the random process $\beta_t = \beta_t(x_t, y_{1:t})$.

Definition 1 We say that β_t is **compressible** if the maximum over T of the average of $(\beta_t)_i^2$, conditioned on past observations, is smaller than the minimum average squared value of any currently nonzero component of x_t , i.e. if $\max_{i \in T} \mathbb{E}[(\beta_t)_i^2 | y_{1:t-1}] < \min_{i \in N_t} \mathbb{E}[(x_t)_i^2]$. This is a valid definition since $\min_{i \in N_t} \mathbb{E}[(x_t)_i^2] \leq \min_{i \in \Delta} \mathbb{E}[(x_t)_i^2] = \min_{i \in \Delta} \mathbb{E}[(\beta_t)_i^2]$ for all choices $\Delta = \Delta(y_{1:t-1})$.

Assumption 1 (model, algorithm) Assume that

1. y_t , x_t follow (1), (2); w_t , v_t are independent of each other and over time; and w_t has bounded support (e.g. truncated Gaussian) with cutoffs at $\pm \frac{\lambda_m \sigma_{obs}}{\max_i \|A_i\|_1}$ in all dimensions.

2. $N_{t-1} \subseteq N_t$ for all t and $S_t := |N_t| \leq S_{max}$.
3. The number of false additions is bounded, i.e. $|T_t \setminus N_t| \leq S_{fa}$ for all t . This implies that $|T_t| \leq S_t + S_{fa} \leq S_{max} + S_{fa}$.
4. $\delta_{S_{max}+S_{fa}} < 1$. $\delta_S = \delta_S(A)$ is defined in [2, eq. (1.3)].

Bounded measurement noise (Assumption 1.1) is usually valid. Assumption 1.3 is observed in all our simulations, as long as the addition threshold α_a is large enough. Assumption 1.4 quantifies the required amount of incoherency of the measurement matrix w.r.t. the sparsity basis. Consider Assumption 1.2. While this assumption is not strictly true, it is observed (for medical image sequences) that it is approximately true: the set $N_t \setminus N_{t-1}$, and the total set $N_t \setminus N_0$, are both small. Also, if we relax the definition of support to denote any set containing all nonzero elements of x_t , then this is true.

Under the above assumptions, we can prove the following [8]:

Theorem 1 Assume that Assumption 1 holds. Let $t_a = t_a(t)$ denote the last addition time before or at t .

1. If $|\Delta|$ is small enough to ensure that $(t - t_a + 1)\sigma_{sys}^2 \geq \frac{\theta_{|T|,|\Delta|}^2}{(1-\delta_{|T|})^2} \lambda_{max}(\mathbb{E}[(x_t)_\Delta (x_t)_{\Delta'}' | y_{1:t-1}]) + \frac{\sigma_{obs}^2}{1-\delta_{|T|}}$, then $\beta_t := x_t - \hat{x}_{t,tmp}$ is compressible. $\theta_{S,S'}$ is defined in [2, eq. (1.5)].
2. The following bound on the CS-LSE error holds

$$\mathbb{E}[\|x_t - \hat{x}_{t,CSLSE}\|_2^2 | y_{1:t-1}] \leq \min_{1 \leq S \leq S_\infty} B_{CSLSE}(S)$$

$$B_{CSLSE}(S) := C_2(S) S \sigma_{obs}^2 + C_3(S) \frac{(|T| + |\Delta| - S)}{S} L0$$

$$L0 \triangleq \begin{cases} \frac{\theta_{|T|,|\Delta|}^2}{(1-\delta_{|T|})^2} \mathbb{E}[\|(x_t)_\Delta\|^2 | y_{1:t-1}] + \frac{\sigma_{obs}^2}{1-\delta_{|T|}} & \text{if } S \geq |\Delta| \\ \left(\frac{\theta_{|T|,|\Delta|}^2}{(1-\delta_{|T|})^2} + 1 \right) \mathbb{E}[\|(x_t)_\Delta\|^2 | y_{1:t-1}] + \frac{\sigma_{obs}^2}{1-\delta_{|T|}} & \text{if } S < |\Delta| \end{cases}$$

where $\hat{x}_{t,CSLSE}$ is the output of (5) with $\tilde{y}_{t,f} = y_t - A \hat{x}_{t,tmp}$ and $\hat{x}_{t,tmp} = (A_T' A_T)^{-1} A_T' y_t$.

Notice that $\mathbb{E}[(x_t)_\Delta (x_t)_{\Delta'}' | y_{1:t-1}]$, and its trace, $\mathbb{E}[\|(x_t)_\Delta\|^2 | y_{1:t-1}]$, can be computed by running a genie-aided KF.

In [8], we also derive a bound on the unconditional CS-LSE error, i.e. the error averaged over all values of the past observations, under slightly stronger assumptions. We also discuss why the bound on the CS-LSE error is much smaller than that on simple CS error.

4. CONVERGENCE TO GENIE-AIDED KF (OR LS)

Consider the genie-aided KF, i.e. the KF which knows the true nonzero set, N_t , at each t . It is the linear MMSE estimator of x_t from y_1, \dots, y_t if the nonzero sets, N_t 's, are known. It would be the MMSE estimator (i.e. it would be the best estimator among all possible estimators) if the observation noise were Gaussian instead of truncated Gaussian. The genie-aided KF can be summarized as running (7) with $\hat{Q}_t = \sigma_{sys}^2 I_{N_t}$. In this section, we obtain conditions under which the KF-CS estimate converges to the genie-aided KF estimate in probability. As a corollary, we also get conditions for LS-CS to converge to genie-aided LS.

We begin by giving Lemma 1 states that if the true nonzero set does not change after a certain time, and if eventually it is correctly detected, then KF-CS converges to GA-KF. This is followed by Lemmas 2 and 3 which prove that, if the addition threshold is high enough, the probability of false addition is zero and the probability of correct set detection approaches one with t . Combining these lemmas gives the final result.

Lemma 1 [8] Assume that there exists a t_0 s.t. $\forall t \geq t_0$, $T_t = N_t = N_*$ and assume that $\delta_{|N_*|} < 1$. Consider KF-CS without the deletion step, i.e. with $\alpha_z = 0$, and with the step 3a (KF-CS with final LS) replacing step 3 if $T_t \neq T_{t-1}$. The difference in the KF-CS and GA-KF estimates, $d_t \triangleq |\hat{x}_{t,GA-KF} - \hat{x}_t|$, converges to zero in mean square and hence also in probability.

Assume that Assumption 1 holds. The bounded support assumption on w_t ensures that $|A_i' w_t| \leq \|w_t\|_\infty \|A_i\|_1 \leq \lambda_m \sigma_{obs}$, $\forall i$. With this, the theorems of [2] can be directly modified to hold with probability one. This helps prove the following.

Lemma 2 Assume that (i) Assumption 1 holds and that $\delta_{2S_{max}} + \delta_{3S_{max}} < 1$ (stronger incoherency requirement than earlier); and (ii) in Algorithm 1, we set $\alpha_a = B_1 \triangleq C_1^2 \lambda_m^2 S_{max} \sigma_{obs}^2$ (C_1 is defined in [2, Thm. 1.1]). Then, at each t , the following hold:

$$\|x_t - \hat{x}_{t,tmp} - \hat{\beta}_t\|^2 \leq B_1 \triangleq C_1^2 \lambda_m^2 S_{max} \sigma_{obs}^2 \quad (8)$$

$$\hat{\Delta}_t \subseteq N_t, \text{ and so } T_t \subseteq N_t, \text{ and so } T_t \cup \Delta_{t+1} = N_{t+1} \quad (9)$$

Proof: We prove this result by induction. At any t , when solving (5), $\hat{x}_{t,tmp,i} = 0$, $\forall i \in T_{t-1}$. The sparse vector to be estimated is $\beta_t \triangleq [(x_t - \hat{x}_{t,tmp})_{T_{t-1}}, (x_t)_{\Delta_t}, 0_{N_c^c}]$. First consider the base case, $t = 1$. At $t = 1$, $T_{t-1} = T_0 = \phi$ (empty) and so $\hat{x}_{1,tmp,i} = 0$, $\forall i$. Thus $\beta_1 = x_1$ with nonzero set $\Delta_1 = N_1$. Since $|N_1| \leq S_{max}$ and since the observation noise, w_t , satisfies $|A_i' w_t| \leq \lambda_m \sigma_{obs}$, we can apply Theorem 1.1 of [2] to get (8) to always hold at $t = 1$.

Also, for any $i \in N_1^c$, $x_{1,i} = 0$ and so $\hat{\beta}_{1,i}^2 = (x_{1,i} - \hat{\beta}_{1,i})^2 \leq \|x_1 - \hat{x}_{1,tmp} - \hat{\beta}_1\|^2 \leq B_1$ (from (8)). But $\alpha_a = B_1$. Thus, from (6), $\hat{\Delta}_1 \subseteq N_1$. Thus $T_1 \triangleq T_0 \cup \hat{\Delta}_1 \subseteq N_1$. But $N_1 \subseteq N_2$. Thus, $T_1 \subseteq N_2$. Since $\Delta_1 \triangleq N_1 \setminus T_0$, this implies that $T_1 \cup \Delta_2 = N_2$. Thus (9) also holds for $t = 1$. This proves the base case.

For the inductive step, assume that (9) and (8) hold for $t - 1$. Thus, $T_{t-1} \cup \Delta_t = N_t$, which is the nonzero set for β_t . But $|N_t| \leq S_{max}$. Thus Theorem 1.1 of [2] can be applied to get (8) to hold for t . Also, for any $i \in N_t^c$, $x_{t,i} = 0$ and so $\hat{\beta}_{t,i}^2 = (x_{t,i} - \hat{\beta}_{t,i})^2 \leq \|x_t - \hat{x}_{t,tmp} - \hat{\beta}_t\|^2 \leq B_1 = \alpha_a$. Thus from (6), $\hat{\Delta}_t \subseteq N_t$. Thus $T_t \triangleq T_{t-1} \cup \hat{\Delta}_t \subseteq N_t \subseteq N_{t+1}$. Since $\Delta_{t+1} \triangleq N_{t+1} \setminus T_t$, this means that $T_t \cup \Delta_{t+1} = N_{t+1}$. Thus (9) holds for t . This proves the induction step and thus the result holds.

Lemma 3 Assume that (i) Assumption 1 holds and that $\delta_{2S_{max}} + \delta_{3S_{max}} < 1$; (ii) in Algorithm 1, we set $\alpha_a = B_1 \triangleq C_1^2 \lambda_m^2 S_{max} \sigma_{obs}^2$; and (iii) all additions occur before a finite time, $t_{a,max}$, i.e. $N_t = N_{t_{a,max}}$, $\forall t \geq t_{a,max}$. Let $N_* \triangleq N_{t_{a,max}}$. Then, $\lim_{t \rightarrow \infty} \Pr(T_{t+\tau} = N_*, \forall \tau \geq 0) = 1$

Proof: Since (i) and (ii) hold, Lemma 2 holds. For any $i \in \Delta_t$, $\hat{x}_{t,tmp,i} = 0$. Thus, (8) implies that $(x_{t,i} - \hat{\beta}_{t,i})^2 \leq B_1$ and so $|\hat{\beta}_{t,i}| \geq |x_{t,i}| - \sqrt{B_1}$. Thus, if $|x_{t,i}| > \sqrt{B_1} + \sqrt{\alpha_a} = 2\sqrt{B_1}$, then $\hat{\beta}_{t,i}^2 > \alpha_a$, i.e. $i \in \hat{\Delta}_t$. In other words, $\Pr(\{i \in \hat{\Delta}_t | x_{t,i}^2 > 4B_1\}) = 1$. The same argument applies even if we consider all $i \in \Delta_t$. Thus, $\Pr(\{\Delta_t \subseteq \hat{\Delta}_t\} | \{x_{t,i}^2 > 4B_1 \forall i \in \Delta_t\}) = 1$.

But from (9) and (6), $\hat{\Delta}_t \subseteq \Delta_t$. Thus, if $x_{t,i}^2 > 4B_1$, $\forall i \in \Delta_t$, $\hat{\Delta}_t = \Delta_t$ and so $T_t \triangleq T_{t-1} \cup \hat{\Delta}_t = N_t$. Thus, $\Pr(T_t = N_t | \{x_{t,i}^2 > 4B_1 \forall i \in \Delta_t\}) = 1$. Now, $\forall t \geq t_{a,max}$, $N_t = N_*$. Thus for $t > t_{a,max}$, $T_t = N_*$ implies that $\Delta_{t+1} = \phi$. This implies that $\hat{\Delta}_t = \phi$ and so $T_{t+1} = T_t = N_*$. Thus, $T_t = N_*$ implies that $T_{t+k} = N_*$, $\forall k \geq 0$. Thus, for all $t > t_{a,max}$,

$$\Pr(T_{t+\tau} = N_* \forall \tau \geq 0 | \{x_{t,i}^2 > 4B_1 \forall i \in \Delta_t\}) = 1 \quad (10)$$

Now, $x_{t,i}^2 \sim \mathcal{N}(0, (t - t_i)\sigma_{sys}^2)$ where t_i is the time at which element i got added. Note that $t_i \leq t_{a,max}$. Thus,

$$Pr(x_{t,i}^2 > 4B_1) \geq 2Q\left(\sqrt{\frac{4B_1}{(t - t_{a,max})\sigma_{sys}^2}}\right) \quad (11)$$

where Q is the Gaussian Q-function. Combining (10), (11) and using the fact that the different $x_{t,i}$'s are independent,

$$Pr(T_{t+\tau} = N_* \forall \tau \geq 0) \geq \left(2Q\left(\sqrt{\frac{4B_1}{(t - t_{a,max})\sigma_{sys}^2}}\right)\right)^{S_{max}} \quad (12)$$

Thus for any $\epsilon > 0$, $Pr(T_{t+\tau} = N_*, \forall \tau \geq 0) \geq 1 - \epsilon$ if $t \geq t_{a,max} + \tau_\epsilon$, $\tau_\epsilon \triangleq \left\lceil \frac{4B_1}{\sigma_{sys}^2 [Q^{-1}((1-\epsilon)^{1/S_{max}})]^2} \right\rceil$, where $\lceil \cdot \rceil$ is the greatest integer function. Thus the claim follows. ■

Combining Lemma 3 with Lemma 1 we get the final result.

Theorem 2 Assume that (i) Assumption 1 holds and that $\delta_{2S_{max}} + \delta_{3S_{max}} < 1$; (ii) in Algorithm 1, we set $\alpha_a = B_1 \triangleq C_1^2 \lambda_m^2 S_{max} \sigma_{obs}^2$; and (iii) all additions occur before a finite time, $t_{a,max}$, i.e. $N_t = N_{t_{a,max}}, \forall t \geq t_{a,max}$. Consider KF-CS without the deletion step, i.e. with $\alpha_z = 0$, and with the step 3a (KF-CS with final LS) replacing step 3 if $T_t \neq T_{t-1}$. Then, $d_t \triangleq \hat{x}_{t,GA-KF} - \hat{x}_t$ converges to zero in probability, i.e. the KF-CS estimate converges to the Genie-Aided KF estimate in probability, as $t \rightarrow \infty$. Also, the LS-CS estimate converges to the Genie-Aided LS estimate, in probability, as $t \rightarrow \infty$ (this follows directly from Lemma 3).

The assumption $\delta_{2S_{max}} + \delta_{3S_{max}} < 1$ is just stronger incoherency requirement on A than Assumption 1. The assumption of all additions occurring before a finite time is a valid one for problems where the system is initially in its transient state (nonstationary), but later stabilizes to a stationary state. Alternatively, the above theorem can be applied to claim that the KF-CS (or LS-CS) estimate stabilizes to within a small error the GA-KF (or GA-LS) estimate, if additions occur slowly enough, i.e. if the delay between two addition times is long enough to allow it to stabilize [8].

5. SIMULATION RESULTS

Lemma 2 says that if the addition threshold was set high enough ($\alpha_a = B_1$ where B_1 is the CS error upper bound), then there would be no false additions. But if we set the addition threshold very high, then for the initial time instants, the KF-CS estimation error would be large and it will do worse than simple CS. Thus, in practice, we set α_a lower, but we implement the false addition detection and removal scheme described in Sec. 2.1.1. We evaluated its performance using the following set of simulations. We simulated a time sequence of sparse $m=256$ length signals, $z_t = x_t$ which follow (2) with $\sigma_{sys}^2 = 1$ and nonzero sets, $N_{t-1} \subseteq N_t, \forall t$ satisfying $N_t = N_1, \forall t < 10$, $N_t = N_{10}, \forall 10 \leq t < 20$, $N_t = N_{20}, \forall 20 \leq t < 30$, $N_t = N_{30}, \forall 30 \leq t < 100$ and $|N_1| = 8, |N_{10}| = 12, |N_{20}| = 16, |N_{30}| = 20$. Thus $S_{max} = 20$. The set N_1 and all the additions were generated uniformly at random from the remaining elements out of $[1 : m]$. The measurement matrix, $A = H$ was simulated as in [2] by generating $n \times m$ i.i.d. Gaussian entries (with $n = 72$) and normalizing each column of the resulting matrix. The observation noise variance was $\sigma_{obs}^2 = ((1/3)\sqrt{16/n})^2$ (this is taken from [2]) and we simulated Gaussian noise (not truncated).

We implemented KF-CS with $\lambda_m = \sqrt{2 \log_2 m} = 4$, $\alpha_a = 9\sigma_{obs}^2$, $\alpha_{fe} = 2n$, $\alpha_z = \sigma_{obs}^2$, $k = 5$, $k' = 3$. Since the observation noise was not truncated, occasionally the addition step resulted in a very large number of false additions, which made $A_{T_t} A_{T_t}^T$ singular

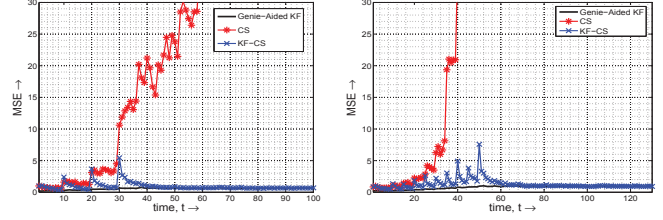


Fig. 1. MSE plots comparison. Fig. 1(a): Large KF-CS error occurs at and after the new addition times, $t = 10, 20, 30$. But once the addition is detected, the error gradually reduces to that of GA-KF (or slightly higher). The error of simple CS (labeled as CS) is much larger (max value 45). Fig. 1(b): Simple CS error beyond $t = 50$ when $S_t \geq 26$ than is much larger in the Fig. 1(a) (max value 425).

(or almost singular) resulting in large errors at all future t . To prevent this, we set a maximum value for the number of allowed additions: we allowed at most $(1.25n / \log_2 m)$ largest magnitude coefficient estimates larger than α_a to be added. Also, typically an addition took 2-3 time instants to get detected. Thus we set $\sigma_{init}^2 = 3\sigma_{sys}^2$ (σ_{init}^2 is used instead of σ_{sys}^2 the first time a new coefficient gets added). We simulated the above system 100 times and compared the MSE of KF-CS with that of GA-KF and of simple CS (followed by thresholding and least squares estimation as in Gauss-Dantzig [2]).

In a second set of simulations, shown in Fig. 1(b), we started with $S_1 = 8$ and for $10 \leq t \leq 50$, we added 2 new elements every 5 time units. Thus $S_{max} = 26 = S_t, \forall t \geq 50$. Note $26 > n/3 = 24$, i.e. $\delta_{3S_{max}}$ cannot be smaller than 1.

6. DISCUSSION AND ONGOING WORK

In this work, we introduced Least Squares CS and analyzed why CS on the LS error in the observation will have lower error than CS on the raw observations (simple CS), when sparsity patterns change slowly enough. We also showed that if all additions occur before a finite time, if the addition threshold is set high enough, if UUP holds for S_{max} , and if the noise has bounded support, KF-CS (or LS-CS), converge to the genie-aided KF (or LS) in probability. In ongoing work, we are working on relaxing the first three assumptions used in the above result. We are also working on developing KF-CS for real-time dynamic MR imaging [6].

7. REFERENCES

- [1] N. Vaswani, "Kalman filtered compressed sensing," in *IEEE Intl. Conf. Image Proc. (ICIP)*, 2008.
- [2] E. Candes and T. Tao, "The dantzig selector: statistical estimation when p is much larger than n," *Annals of Statistics*, 2006.
- [3] C. Rozell, D. Johnson, R. Baraniuk, and B. Olshausen, "Locally competitive algorithms for sparse approximation," in *IEEE Intl. Conf. Image Proc. (ICIP)*, 2007.
- [4] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic mri," *Magnetic Resonance in Medicine*, vol. 59(2), pp. 365–373, January 2008.
- [5] M. Wakin, J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk, "An architecture for compressive imaging," in *IEEE Intl. Conf. Image Proc. (ICIP)*, 2006.
- [6] C. Qiu, W. Lu, and N. Vaswani, "Real-time dynamic mr image reconstruction using kalman filtered compressed sensing," in *IEEE Intl. Conf. Acoustics, Speech, Sig. Proc. (ICASSP)*, 2009.
- [7] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Info. Th.*, Sept. 2006.
- [8] N. Vaswani, "Analyzing least squares and kalman filtered compressed sensing," in *long version, available at http://www.ece.iastate.edu/~namrata/AnalyzeKFCS.pdf*, 2008.