KALMAN FILTERED COMPRESSED SENSING

Namrata Vaswani

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

ABSTRACT

We consider the problem of reconstructing time sequences of spatially *sparse* signals (with unknown and time-varying sparsity patterns) from a *limited number* of linear *"incoherent"* measurements, *in real-time*. The signals are sparse in some transform domain referred to as the sparsity basis. For a single spatial signal, the solution is provided by Compressed Sensing (CS). The question that we address is, for a sequence of sparse signals, can we do better than CS, if (a) the sparsity pattern of the signal's transform coefficients' vector changes slowly over time, and (b) a simple prior model on the temporal dynamics of its current non-zero elements is available. The overall idea of our solution is to use CS to estimate the support set of the initial signal's transform vector. At future times, run a reduced order Kalman filter with the currently estimated support and *estimate new additions to the support set by applying CS to the Kalman innovations or filtering error (whenever it is "large")*.

Index Terms/Keywords: compressed sensing, Kalman filtering, compressive sampling, sequential MMSE estimation

1. INTRODUCTION

We consider the problem of reconstructing time sequences of spatially *sparse* signals (with unknown and time-varying sparsity patterns) from a *limited number* of linear *"incoherent"* measurements, *in real-time*. The signals are sparse in some transform domain referred to as the "sparsity basis" [1]. A common example of such a problem is dynamic MRI or CT to image deforming human organs or to image brain neural activation patterns (in response to stimuli) using fMRI. The ability to perform real-time MRI capture and reconstruction can make interventional MR practical [2]. Human organ images are usually piecewise smooth and thus the wavelet transform is a valid sparsity basis [1, 3]. Due to strong temporal dependencies, the sparsity pattern usually changes *slowly* over time. MRI captures a small (sub-Nyquist) number of Fourier transform coefficients of the image, which are known to be "incoherent" with respect to the wavelet transform [1, 3]. Other example problems include sequentially estimating optical flow of a single deforming object (sparse in Fourier domain) from a set of randomly spaced optical flow measurements (e.g. those at high intensity variation points [4]), or real-time video reconstruction using the single-pixel camera [5].

The solution to the static version of the above problem is provided by Compressed Sensing (CS) [1, 6, 7]. The noise-free observations case [1] is exact, with high probability (w.h.p.), while the noisy case [7] has a small error w.h.p.. But existing solutions for the dynamic problem [5, 8] treat the entire time sequence as a single spatiotemporal signal and perform CS to reconstruct it. This is a batch solution (need to wait to get the entire observation sequence) and has very high complexity. An alternative would be to apply CS at each time separately, which is online and low-complexity, but will require many more measurements to achieve low error. The question

that we address is: *can we do better than performing CS at each time separately, if (a) the sparsity pattern (support set) of the transform coefficients' vector changes slowly, i.e. every time, none or only a few elements of the support change, and (b) a prior model on the temporal dynamics of its current non-zero elements is available.*

Our solution is motivated by reformulating the above problem as causal minimum mean squared error (MMSE) estimation with a slow time-varying set of dominant basis directions (or equivalently the support of the transform vector). If the support is known, the MMSE solution is given by the Kalman filter (KF) [9] for this support. But what happens if the support is unknown and time-varying? The initial support can be estimated using CS [7]. If at a given time, there is an addition to the support set, but we run the KF for the old model, there will be a model mismatch and the innovation (and filtering) error will increase. Whenever it does, *the change in support can be estimated by running CS on the innovation or the filtering error*, followed by thresholding. A Kalman update step is run using the new support set. If some coefficients become and remain nearly zero (or nearly constant), they can be removed from the support set.

If, for a moment, we assume that CS [7] gives the correct estimate of the support at all times, then the above approach will give the MMSE estimate of the signal at all times. The reason it is very likely that CS [7] gives the correct estimate is because we use it to fit a very sparse "model change" signal to the filtering error. Also note that a full Kalman filter [9], that does not use the fact that the signal is sparse, is meaningless here, because the number of observations available is smaller than the signal dimension, and thus many elements of the signal transform will be unobservable. Unless all unobservable modes are stable, the error will blow up. Other recent work that also attempts to use prior knowledge with CS, but to reconstruct only a single signal is [10, 11, 12].

2. THE MODEL AND PROBLEM FORMULATION

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 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 $S_t << m$ elements of x_t are non-zero). Here $'$ denotes transpose. The observations are "incoherent" w.r.t. the sparsity basis of the signal, i.e. $y_t = H z_t + w_t = H\Phi x_t + w_t$, where $H_{n \times m}$ is such that the correlation between the columns of $A \triangleq H\Phi$ is small enough to ensure that, for any $S \leq S_t$, any S-column sub-matrix of A is "approximately orthonormal" (its nonzero singular values are between $\sqrt{1-\delta}$ to $\sqrt{1+\delta}$ for $\delta < 1$) [7]. w_t is i.i.d. Gaussian measurement noise. Thus the measurement model is:

$$y_t = A x_t + w_t, \ A \triangleq H\Phi, \ w_t \sim \mathcal{N}(0, \sigma_{obs}^2 I) \qquad (1)$$

We refer to x_t as the state at t. Our goal is to get the "best" causal estimate of x_t (or equivalently of the signal, $z_t = \Phi x_t$) at each t.

Let T_t denote the the support set of x_t, i.e. the set of its non-zero coordinates and let $S_t = size(T_t)$. In other words, $T_t =$

$[i_1, i_2, \ldots i_{S_t}]$ where i_k are the non-zero coordinates of x_t. For any set T, let $(v)_T$ denote the $size(T)$ length sub-vector containing the elements of v corresponding to the indices in the set T. For another set, γ, we also use the notation T_γ which treats T as a vector and selects the elements of T corresponding to the indices in the set γ. 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. The set operations $\cup$, $\cap$, and $\setminus$ have the usual meanings (note $T_1 \setminus T_2$ denotes elements of T_1 not in T_2). We use $'$ to denote transpose. T^c denotes the complement of T w.r.t. $[1:m]$, i.e. $T^c \triangleq [1:m] \setminus T$. Also $||v||_p$ is the l_p norm of the vector v, i.e. $||v||_p \triangleq (\sum_i |v_i|^p)^{1/p}$.

Assumption 1. We assume slow changes in sparsity patterns, i.e. the maximum size of the change in the support set at any time is smaller (usually much smaller) than S_t at any t, i.e. $S_{diff,max} \triangleq \max_t[size(T_t \setminus T_{t-1}) + size(T_{t-1} \setminus T_t)] < \min_t S_t$.

Assumption 2. We also assume that A satisfies $\delta_{2S_{max}} + \delta_{3S_{max}} < 1$ where δ_S is the RIP constant defined in equation 1.3 of [7] and $S_{max} \triangleq \max_t S_t$. It should be possible to apply the proposed algorithm even under a slightly weaker assumption that only requires $\delta_{2S_{max}} < 1$ (required to ensure any S_{max} or less column sub-matrix of A is full rank and hence the state is observable) and $\delta_{2S_{diff,max}} + \delta_{3S_{diffmax}} < 1$. This is part of ongoing work.

System Model for x_t. For the currently non-zero coefficients of x_t, we assume a spatially i.i.d. Gaussian random walk model, with noise variance σ_{sys}^2. At the first time instant at which $(x_t)_i$ becomes non-zero, it is assumed to be generated from a zero mean Gaussian with variance σ_{init}^2. Thus, we have the model: $x_0 = 0$,

$$(x_t)_i = (x_{t-1})_i + (\nu_t)_i, \; (\nu_t)_i \sim \mathcal{N}(0, \sigma_{sys}^2), \text{ if } i \in T_t, \; i \in T_{t-1}$$
$$(x_t)_i = (x_{t-1})_i + (\nu_t)_i, \; (\nu_t)_i \sim \mathcal{N}(0, \sigma_{init}^2) \text{ if } i \in T_t, \; i \notin T_{t-1}$$
$$(x_t)_i = (x_{t-1})_i \text{ if } i \notin T_t \tag{2}$$

The above model can be compactly written as: $x_0 = 0$,

$$x_t = x_{t-1} + \nu_t, \; \nu_t \sim \mathcal{N}(0, Q_t),$$
$$(Q_t)_{T_t \cap T_{t-1}, T_t \cap T_{t-1}} = \sigma_{sys}^2 I$$
$$(Q_t)_{T_t \setminus T_{t-1}, T_t \setminus T_{t-1}} = \sigma_{init}^2 I$$
$$(Q_t)_{T_t^c, T_t^c} = 0 \tag{3}$$

where *the set T_t is unknown* $\forall t$. If T_t were known at each t, i.e. the system model was completely defined, the MMSE estimate of x_t from $y_1, y_2, \ldots y_t$ would be given by a reduced order KF defined for $(x_t)_{T_t}$. But, as explained in Sec. 1, in most practical problems, T_t is in fact unknown and time-varying. Often, it may be possible to get a rough prior estimate of T_1 by thresholding the eigenvalues of the covariance of x_1 (possible to do if multiple realizations of x_1 are available to estimate its covariance). But without multiple i.i.d. realizations of the entire $\{x_t\}$, which are impossible to obtain in most cases, it is not possible to get a-priori estimates of T_t for all t. But note that, it is possible to estimate $\sigma_{sys}^2, \sigma_{init}^2$ for the model of (3) using just one "training" realization of $\{x_t\}$ (which is usually easy to get) by setting the near-zero elements to zero in each x_t and using the rest to obtain an ML estimate.

Assuming known values of $\sigma_{sys}^2, \sigma_{init}^2$, our goal here is to get the best estimates of T_t and x_t at each t using $y_1, \ldots y_t$. Specifically,

1. At each time, t, get the best estimate of the support set, T_t, i.e. get an estimate $\hat{T}_t$ with smallest possible $[size(\hat{T}_t \setminus T_t) + size(\hat{T}_t \setminus T_t)]$ using $y_1, y_2 \ldots y_t$.

2. Assuming the estimates of $T_1, \ldots T_t$ are perfect (have zero error), get the MMSE estimate of x_t using $y_1, y_2 \ldots y_t$.

3. KALMAN FILTERED COMPRESSED SENSING (KF-CS)

We explain Kalman Filtered Compressed Sensing (KF-CS) below. We misuse notation to also denote the estimated nonzero set by T_t.

Running the KF. Assume, for now, that the support set at $t = 1$, T_1, is known. Consider the situation where the first change in the support occurs at a $t = t_a$, i.e. for $t < t_a$, $T_t = T_1$, and that the change is an addition to the support. This means that for $t < t_a$, we need to just run a regular KF, which assumes the following reduced order measurement and system models: $y_t = A_T(x_t)_T + w_t$, $(x_t)_T = (x_{t-1})_T + (\nu_t)_T$, with $T = T_1$. The KF prediction and update steps for this model are [9]: $\hat{x}_0 = 0$, $P_0 = 0$,

$$\hat{x}_{t|t-1} = \hat{x}_{t-1}$$
$$(P_{t|t-1})_{T,T} = (P_{t-1})_{T,T} + \sigma_{sys}^2 I \tag{4}$$

$$K_{t,T} \triangleq (P_{t|t-1})_{T,T} A_T' \Sigma_{ie,t}^{-1}, \; \Sigma_{ie,t} \triangleq A_T(P_{t|t-1})_{T,T} A_T' + \sigma_{obs}^2 I$$
$$(\hat{x}_t)_T = (\hat{x}_{t|t-1})_T + K_{t,T}[y_t - A\hat{x}_{t|t-1}]$$
$$(\hat{x}_t)_{T^c} = (\hat{x}_{t|t-1})_{T^c} = (\hat{x}_{t-1})_{T^c}$$
$$(P_t)_{T,T} = [I - K_{t,T} A_T](P_{t|t-1})_{T,T} \tag{5}$$

Detecting If Addition to Support Set Occurred. The Kalman innovation error is $\tilde{y}_t \triangleq y_t - A\hat{x}_{t|t-1}$. For $t < t_a$, $\tilde{y}_t = [A(x_t - \hat{x}_{t|t-1}) + w_t] \sim \mathcal{N}(0, \Sigma_{ie,t})$ [9]. At $t = t_a$, a new set, Δ, gets added to the support of x_t, i.e. $y_t = A_T(x_t)_T + A_\Delta(x_t)_\Delta + w_t$, where the set Δ is unknown. Since the old model is used for the KF prediction, at $t = t_a$, $\tilde{y}_t$ will have non-zero mean, $A_\Delta(x_t)_\Delta$, i.e.

$$\tilde{y}_t = A_\Delta(x_t)_\Delta + \tilde{w}_t = A_{T^c}(x_t)_{T^c} + \tilde{w}_t, \; where$$
$$\tilde{w}_t \triangleq [A_T(x_t - \hat{x}_{t|t-1})_T + w_t] \sim \mathcal{N}(0, \Sigma_{ie,t}) \tag{6}$$

where $\Delta \subseteq T^c$ is the undetected nonzero set at the current time. Thus, the problem of detecting if a new set has been added or not gets transformed into the problem of detecting if the Gaussian distributed $\tilde{y}_t$ has non-zero or zero mean. Note that $A_\Delta(x_t)_\Delta = A_{T^c}(x_t)_{T^c}$ and thus the generalized Likelihood Ratio Test (G-LRT) for this problem simplifies to detecting if the weighted innovation error norm, $IEN \triangleq \tilde{y}_t' \Sigma_{ie,t}^{-1} \tilde{y}_t \gtrless$ threshold. Alternatively, one can apply G-LRT to the filtering error, $\tilde{y}_{t,f} \triangleq y_t - A\hat{x}_t$. $\tilde{y}_{t,f}$ can be written:

$$\tilde{y}_{t,f} = A_\Delta(x_t)_\Delta + A_T(x_t - \hat{x}_t)_T + w_t$$
$$= [I - A_T K_{t,T}]A_\Delta(x_t)_\Delta + \tilde{w}_{t,f}, \; \tilde{w}_{t,f} \triangleq [I - A_T K]\tilde{w}_t$$
$$\tilde{w}_{t,f} \sim \mathcal{N}(0, \Sigma_{fe,t}), \; \Sigma_{fe,t} \triangleq [I - A_T K_{t,T}]\Sigma_{ie,t}[I - A_T K_{t,T}]'\tag{7}$$

The filtering error covariance, $\Sigma_{fe,t} < \Sigma_{ie,t}$. Thus, on average, in $\tilde{y}_{t,f}$, the noise, $\tilde{w}_{t,f}$, is smaller than that in $\tilde{y}_t$ (since the change, $(x_t - x_{t-1})_T$, has been estimated and subtracted out), but the new component, $A_\Delta(x_t)_\Delta$, is also partially suppressed. The suppression is small because $A_T K_{t,T} A_\Delta(x_t)_\Delta = A_T(P_{t|t-1}^{-1} \sigma_{obs}^2 + A_T' A_T)^{-1} A_T' A_\Delta(x_t)_\Delta$ (follows by rewriting $K_{t,T}$ using the matrix inversion lemma) and $A_T' A_\Delta(x_t)_\Delta$ is small (because of restricted orthogonality [7, eq. 1.5]). Assuming the suppression is small enough, using $\tilde{y}_{t,f}$ will result in lower misses for a given false alarm rate. Thus we use G-LRT on $\tilde{y}_{t,f}$.

Estimating the New Additions (using CS). If the filtering error norm, $FEN \triangleq \tilde{y}_{t,f}' \Sigma_{fe,t}^{-1} \tilde{y}_{t,f}$, is "high", there is a need to estimate the new additions' set, Δ. This can be done by applying the Dantzig selector (DS) [7] to $\tilde{y}_{t,f}$ followed by thresholding the output of the DS (as is also done in the Gauss-Dantzig selector), i.e. we compute

$$\hat{\beta}_t = \arg \min_\beta ||\beta||_1, \text{s.t.} ||A_{T^c}'(\tilde{y}_{t,f} - A_{T^c}\beta)||_\infty \leq \lambda_m \sigma_{obs}$$

$$\hat{\Delta} = (T^c)_{nz}, \; where \; nz \triangleq \{i : \hat{\beta}_{t,i}^2 > \alpha_a\}, \tag{8}$$

894

Algorithm 1 Kalman Filtered Compressive Sensing (KF-CS)

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

1. Set $T \leftarrow T_{t-1}$.

2. **KF prediction and update.** Run (4) and (5) using the current T. Compute the filtering error, $\tilde{y}_{t,f} \triangleq y_t - A\hat{x}_t$.

3. **Addition (using CS).** Compute $FEN \triangleq \tilde{y}_{t,f}' \Sigma_{fe,t}^{-1} \tilde{y}_{t,f}$, and check if it is greater than its threshold. If it is,

   (a) **Run CS on the filtering error followed by thresholding**, i.e. compute $\hat{\Delta}$ using (8) [or use (9)].

   (b) The new estimated support is $T_{new} = T \cup \hat{\Delta}$.

   (c) Set $T \leftarrow T_{new}$. Set $(P_{t|t-1})_{\hat{\Delta},\hat{\Delta}} = \sigma_{init}^2 I$. **Run the KF update** given in (5) for the current T.

   Performance can be improved by iterating the above four steps until $size(\hat{\Delta}) = 0$ or FEN less than its threshold.

4. **Deletion.** Compute the set $\hat{\Delta}_D = \{i \in T : \sum_{\tau=t-k+1}^{t} (\hat{x}_\tau)_i^2 < k\alpha_d\}$. The new estimated support set is $T_{new} = T \setminus \hat{\Delta}_D$.

   (a) Set $T \leftarrow T_{new}$. Set $(\hat{x}_t)_{\Delta_D} = 0$, $(P_{t|t-1})_{\hat{\Delta}_D,[1:m]} = 0$, $(P_{t|t-1})_{[1:m],\hat{\Delta}_D} = 0$. **Run the KF update** given in (5).

5. Assign $T_t \leftarrow T$. **Output T_t, $\hat{x}_t$ and the signal estimate, $\hat{z}_t = \Phi\hat{x}_t$.** Increment t and go to step 1.

$\lambda_m \triangleq \sqrt{2\log m}$ and α_a is the zeroing threshold for addition. Thus, the new estimated support set is $T_{new} = T \cup \hat{\Delta}$. We initialize the prediction covariance along $\hat{\Delta}$ as $(P_{t|t-1})_{\hat{\Delta},\hat{\Delta}} = \sigma_{init}^2 I$. Since it typically takes a few time instants before a new addition gets detected, it is useful to set σ_{init}^2 to a higher value compared to σ_{sys}^2.

Note that the above ignores the fact that the "noise" in $\tilde{y}_{t,f}$, $\tilde{w}_{t,f}$, is colored and that the "signal" to be estimated is partially suppressed (explained earlier). Since the suppression is small, the algorithm still works in practice, but the error bound results for the DS cannot be applied. Alternatively, as we explain in ongoing work [13], one can rewrite $\tilde{y}_{t,f} = A\beta_t + w_t$ where $\beta_t \triangleq [(x_t - \hat{x}_t)_T, (x_t)_{T^c}]$ is a "sparse-compressible" signal with a "large" nonzero part, $(x_t)_\Delta$, a "small" or "compressible" nonzero part, $(x_t - \hat{x}_t)_T$ and the zero part, $(x_t)_{(T\cup\Delta)^c}$. Then, DS can be applied to estimate the "large" nonzero part as follows (this will correctly detect elements whose value is above the filtering error level):

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

$$\hat{\Delta} = \{i \in T^c : \hat{\beta}_{t,i}^2 > \alpha_a\} \tag{9}$$

As we discuss in [13], the above can be analyzed by adapting Theorem 1.2 and Theorem 1.3 of [7]. If the sparsity pattern changes slowly enough and the filtering error is small enough (slow time varying system), it should be possible to show that performing *CS on the filtering error, $\tilde{y}_{t,f}$, to only detect new additions* is *more accurate than performing regular CS at each t on y_t to detect the entire vector x_t* (without using knowledge of the previous support set).

KF Update. We run the KF update given in (5) with $T = T_{new}$. This can be interpreted as a Bayesian version of Gauss-Dantzig [7].

Iterating CS and KF-update. Often, it may happen that not all the elements of the true Δ get estimated in one run of the CS step. To address this, CS and KF update can be iterated until FEN goes below a threshold or until $\hat{\Delta}$ is empty. But there is also a risk of adding too many wrong coefficients.

Deleting Near-Zero Coefficients. Over time, some coefficients may become and remain zero. Alternatively, some coefficients may wrongly get added in the addition step, due to CS error. In both cases, the coefficients need to be removed from the support set T_t. One possible way to do this would be to check if $(\hat{x}_t)_i^2 < \alpha_d$ or to average its value over the last few time instants. When a coefficient, i, is removed, we need to modify T_t, set $(\hat{x}_t)_i = 0$ and set

$(P_{t|t-1})_{i,[1:m]} = 0$ and $(P_{t|t-1})_{[1:m],i} = 0$. As we explain in [13], to prevent too many deletion errors, deletion should be done only when the KF has stabilized (T_t has not changed for long enough).

Deleting Constant Coefficients. If a coefficient, i, becomes constant (this may happen in certain applications), one can keep improving the estimate of its constant value by changing the prediction step for it to $(P_{t|t-1})_{i,i} = (P_{t-1})_{i,i}$. Either one can keep doing this forever (the error in its estimate will go to zero with t) or one can assume that the estimation error has become negligibly small after a finite time and then remove the coefficient index from T_t. It is not clear what is the correct thing to do in this case.

Initialization. Initially, the support set, T_1 may be roughly known (estimated by thresholding the eigenvalues of the covariance of x_1, which is computable if its multiple realizations are available) or unknown. We initialize KF-CS by setting $\hat{x}_0 = 0$, $P_0 = 0$ and T_0 = roughly known support or T_0 = empty (if support is completely unknown). In the latter case, automatically at $t = 1$, the IEN (or FEN) will be large, and thus CS will run to estimate T_1.

The entire KF-CS algorithm is summarized in Algorithm 1.

KF-CS Error Analysis. In ongoing work [13], we are working on finding sufficient conditions under which KF-CS error will converge to that of the genie-aided KF (KF with known nonzero set at each t). This can be used to show KF-CS error stability. The key idea is to analyze the effect of missed and false additions (or false and missed deletions). The extra error due to a missed element, $(x_t)_i$, cannot be larger than a constant times the CS error at the current time (which itself is upper bounded by a small value w.h.p. [7]) plus α_a (due to thresholding). Also, eventually, when the magnitude of $(x_t)_i$ becomes large enough (exceeds CS error plus threshold), it will get detected by CS at that time w.h.p.. Thus, w.h.p., the detection delay will be finite.

We can prevent too many extra coordinates from getting wrongly estimated by having a rough idea of the maximum sparsity of x_t and using thresholding to only select that many, or a few more, highest magnitude non-zero elements. The deletion scheme is currently being improved. Note that if some true element gets missed by CS (or gets wrongly deleted) because its value was too small, it will, w.h.p., get detected by CS at a future time. Also, as long as $rank(A_T) > size(T)$ for the currently estimated T (which may contain some extra coordinates), the estimation error will increase
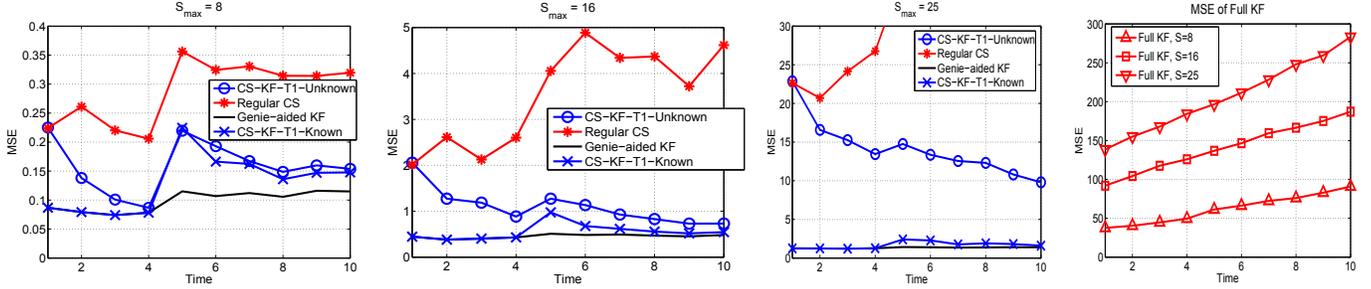
895

Fig. 1. MSE plots of KF-CS (labeled CS-KF by mistake) with initial nonzero set, T_1, unknown and known cases, compared against regular CS in the first 3 figures and against the Full 256-dim KF in the last figure (its MSE is so large that we cannot plot it in the same scale as the others). The benchmark (MMSE estimate with known T_1, T_5) is the genie-aided KF. The simulated signal's energy at t is $\mathbb{E}[||x_t||_2^2] = S_1\sigma_{init}^2 + (\sum_{\tau=2}^{t} S_\tau)\sigma_{sys}^2$.

beyond MMSE, but will not blow up.

4. SIMULATION RESULTS

We simulated a time sequence of sparse m=256 length signals, x_t, with maximum sparsity S_{max}. Three sets of simulations were run with S_{max}= 8, 16 and 25. The A matrix was simulated as in [7] by generating $n \times m$ i.i.d. Gaussian entries (with $n = 72$) and normalizing each column of the resulting matrix. Such a matrix has been shown to satisfy the UUP at a level $C \log m$ [7]. The observation noise variance, $\sigma_{obs}^2 = ((1/3)\sqrt{S_{max}/n})^2$ (this is taken from [7]). The prior model on x_t was (3) with $\sigma_{init}^2 = 9$ and $\sigma_{sys}^2 = 1$. T_1 (support set of x_1) was obtained by generating $S_{max} - 2$ unique indices uniformly randomly from $[1 : m]$. We simulated an increase in the support at $t = 5$, i.e. $T_t = T_1$, $\forall t < 5$, while at $t = 5$, we added two more elements to the support set. Thus, $T_t = T_5$, $\forall t \geq 5$ had size S_{max}. Only addition to the support was simulated.

We used the proposed KF-CS algorithm (Algorithm 1), without the deletion step, to compute the causal estimate $\hat{x}_t$ of x_t at each t. The resulting mean squared error (MSE) at each t, $\mathbb{E}_{x,y}[||x_t - \hat{x}_t||_2^2]$, was computed by averaging over 100 Monte Carlo simulations of the above model. The same matrix, A, was used in all the simulations, but we averaged over the joint pdf of x, y, i.e. we generated T_1, T_5, $(\nu_t)_{T_t}$, $w_t, t = 1, \ldots 10$ randomly in each simulation. Our simulation results are shown in Fig. 1 *(KF-CS is labeled as CS-KF in plots by mistake)*. Our benchmark was the genie-aided KF, i.e. an S_{max}-order KF with known T_1 and T_5, which generates the MMSE estimate of x_t. We simulated two types of KF-CS methods, one with known T_1, but unknown T_5 and the other with unknown T_1 and T_5. Both performed almost equally well for $S_{max} = 8$, but as S_{max} was increased much beyond the UUP level of A, the performance of the unknown T_1 case degraded more (the CS assumption did not hold). We also show comparison with regular CS at each t, which does not use the fact that T_t changes slowly (and does not assume known T_1 either). This had much higher MSE than KF-CS. The MSE become worse for larger S_{max}. We also implemented the full KF for the 256-dim state vector. This used (3) with $Q_t = \sigma_{sys}^2 I_{256 \times 256}$, i.e. it assumed no knowledge of the sparsity. Since we had only a 72-length observation vector, the full system is not observable. Since all non-zero modes are unstable, its error blows up.

5. CONCLUSIONS AND FUTURE DIRECTIONS

To the best of our knowledge, this is the first work on extending the CS idea to *causally* estimate a time sequence of spatially sparse signals. We do this by using CS to estimate the signal support at the

initial time instant, followed by running a KF for the reduced order model, until the innovation or filtering error increases. When it does, we estimate the "change in support" by running CS on the filtering error. This has much lower error since the "change" is much sparser than the actual signal. Open questions to be addressed in future are (a) the analysis of the stability of KF-CS, (b) comparison of KF-CS error with that of regular CS, (c) studying how and when to delete coefficients, (d) KF-CS for compressible signal sequences.

6. REFERENCES

[1] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Info. Th.*, vol. 52(2), pp. 489–509, February 2006.

[2] 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, December 2002.

[3] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine*, vol. 58(6), pp. 1182–1195, December 2007.

[4] J. Shi and C. Tomasi, "Good features to track," in *IEEE Conf. on Comp. Vis. Pat. Rec. (CVPR)*, 1994, pp. 593–600.

[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.*, 2006.

[6] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52(4), pp. 1289–1306, April 2006.

[7] E. Candes and T. Tao, "The dantzig selector: statistical estimation when p is much larger than n," *Annals of Statistics*, 2006.

[8] U. Gamper, P. Boesiger, and S. Kozerke, "Compressed sensing in dynamic mri," *Magnetic Resonance in Medicine*, vol. 59(2), pp. 365–373, January 2008.

[9] T. Kailath, A.H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, 2000.

[10] J. Garcia-Frias and I. Esnaola, "Exploiting prior knowledge in the recovery of signals from noisy random projections," in *Data Compression Conference*, 2007.

[11] K. Egiazarian, A. Foi, and V. Katkovnik, "Compressed sensing image reconstruction via recursive spatially adaptive filtering," in *IEEE Intl. Conf. Image Proc.*, 2007.

[12] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Sig. Proc.*, to appear.

[13] N. Vaswani, "Analyzing kalman filtered compressed sensing," in *submitted to Allerton*, 2008.