

# KF-CS: Compressive Sensing on Kalman Filtered Residual

## A. Notation and Problem Definition

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 sub-vector containing the elements of  $v$  corresponding to the indices in the set  $T$ .  $\|v\|_k$  denotes the  $\ell_k$  norm of a vector  $v$ . If just  $\|v\|$  is used, it refers to  $\|v\|_2$ . For a matrix  $M$ ,  $\|M\|_k$  denotes its induced  $k$ -norm, while just  $\|M\|$  refers to  $\|M\|_2$ .  $M'$  denotes the transpose of  $M$ . For a tall matrix,  $M$ ,  $M^\dagger := (M'M)^{-1}M'$ . For symmetric matrices,  $M_1 \leq M_2$  means that  $M_2 - M_1$  is positive semidefinite. For a fat matrix  $A$ ,  $A_T$  denotes the sub-matrix obtained by extracting the columns of  $A$  corresponding to the indices in  $T$ . The  $S$ -restricted isometry property (RIP) constant,  $\delta_S$ , and the  $S, S'$ -restricted orthogonality constant,  $\theta_{S,S'}$ , are as defined in equations 1.3 and 1.5 of [?] respectively.

For a square matrix,  $Q$ , we use  $(Q)_{T_1, T_2}$  to denote the sub-matrix of  $Q$  containing rows and columns corresponding to the entries in  $T_1$  and  $T_2$  respectively.  $I$  denotes an appropriate sized identity matrix. The  $m \times m$  matrix  $I_T$  is defined as

$$(I_T)_{T, T} = I, (I_T)_{T^c, [1, m]} = 0, (I_T)_{[1, m], T^c} = 0 \quad (1)$$

We use 0 to denote a vector or matrix of all zeros of appropriate size. The notation  $z \sim \mathcal{N}(\mu, \Sigma)$  means that  $z$  is Gaussian distributed with mean  $\mu$  and covariance  $\Sigma$ .

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 observation vector at  $t$ , i.e.  $y_t = Hz_t + w_t$ . 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. We denote its support by  $N_t$  and we use  $S_t := |N_t|$  to denote its size. Thus the observation model is

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

where  $\mathbb{E}[\cdot]$  denotes expectation. We assume that  $A$  has unit norm columns. The observation noise,  $w_t$ , is independent identically distributed (i.i.d.) over  $t$  and is independent of  $x_t$ . 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  $t$ .

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

- 1) let  $S_*$  denote the largest  $S$  for which  $\delta_S < 1/2$ ,
- 2) let  $S_{**}$  denote the largest  $S$  for which  $\delta_{2S} < \sqrt{2} - 1$ .

**Definition 2 (Define  $\hat{x}_t$ ,  $\hat{N}_t$ ):** We use  $\hat{x}_t$  to denote the final estimate of  $x_t$  at time  $t$  and  $\hat{N}_t$  to denote its support estimate.

**Definition 3 (Define  $T$ ,  $\Delta$ ,  $\Delta_e$ ):** We use  $T \equiv T_t := \hat{N}_{t-1}$  to denote the support estimate from the previous time. This

serves as an initial estimate of the current support. Now we will define the difference of support sets of two successive time instances in the following way. We use

$\Delta_{add} \equiv \Delta_{add, t} = N_t \setminus N_{t-1}$  to denote the set of actual addition at time  $t$ .

$\Delta_{del} \equiv \Delta_{del, t} = N_{t-1} \setminus N_t$  to denote the set of actual deletion at time  $t$ .

$\Delta \equiv \Delta_t := N_t \setminus T_t$  to denote the unknown part of the support at the current time.

$\Delta_e \equiv \Delta_{e, t} := T_t \setminus N_t$  to denote the “erroneous” part of  $T_t$ .

$\Delta_A \equiv \Delta_{A, t} = \hat{N}_t \setminus \hat{N}_{t-1}$  to denote the estimated addition.

$\Delta_D \equiv \Delta_{D, t} = \hat{N}_{t-1} \setminus \hat{N}_t$  to denote the estimated deletion.

To keep notation simple, we remove the subscript  $t$  in most places.

## I. KALMAN FILTERED MODCS RESIDUAL

### A. Signal Model

In this section we will describe a simple Random walk model. We call it as Signal Model 1. This model will help us to understand the basic KF-ModCS algorithm in the next section. The complete signal model is described in the below :

**Signal Model 1:**  $t = t_0$  is the initial time instant.

- 1) At  $t = t_0$ ,  $x_0$  is  $S_0$  sparse with support  $N_0$  and  $(x_0)_{N_0} \sim \mathcal{N}(0, \sigma_{sys, 0}^2 I)$ . The condition for signal mean to be zero is not a necessary condition for our algorithm.
- 2) At time  $t_{add, j} = t_0 + jd_1$ , for some  $d_1 > 0$ , there are  $S_a$  number of new additions to the support indices. Every new addition index also follows a Gaussian distribution with zero mean and variance  $\sigma_{sys, 0}^2$ . We pick these new indices randomly with uniform probability from the set  $[1, m] \setminus N_{t-1}$ , where  $N_{t-1}$  is the support of the signal at the previous time instant. Denote the set of indices of the co-efficient added at  $t_{add, j}$  by  $\Delta_{add, t}$ .
- 3) At  $t_{del, j} = t_{add, j} + d_2$ , for some  $d_2 > 0$  and for all  $j > 0$ , there are  $S_d$  number of new deletions from the support. We pick these deletion indices randomly with uniform probability from the set  $N_{t-1}$ . Denote the set of co-efficient detected at  $t_{del, j}$  by  $\Delta_{del, t}$ .
- 4) The currently non-zero indices of  $x_t$ , follows an independent Gaussian Random walk model with zero mean and variance  $\sigma_{sys}^2$ .
- 5) We limit the maximum size of support set to  $S_{max}$ . Similarly for minimum size, it is  $S_{min}$ .

The above model can be summarized as follows.

$$\begin{aligned}
|N_t \setminus N_{t-1}| &= \begin{cases} S_a & \text{if } t = t_{add,j} \\ 0 & \text{otherwise} \end{cases} \\
|N_{t-1} \setminus N_t| &= \begin{cases} S_d & \text{if } t = t_{del,j} \\ 0 & \text{otherwise} \end{cases} \\
x_0 &\sim \mathcal{N}(0, \Pi_0), \text{ where } \Pi_0 = \sigma_{sys,0}^2 I_{N_0} \\
\nu_t &\sim \mathcal{N}(0, \Pi_t), \text{ where } \Pi_t = \sigma_{sys}^2 I_{N_t} \\
(x_t)_{N_t} &= (x_{t-1})_{N_t} + (\nu_t)_{N_t} \\
(x_t)_{N_t^c} &= (\nu_t)_{N_t^c} = 0
\end{aligned} \tag{3}$$

*Discussion 1:* In some practical situation the support of the signal often changes very slowly. Keeping that in mind we can set the values of  $d_1$  and  $d_2$  large. Another important thing about this particular signal model is that  $d_1$  and  $d_2$  are constant. So between two successive additions, or between two successive deletions or between one addition and deletion the time differences are always constant. But in practical situation these may not happen. So we can make those differences random while we simulate our algorithm.

#### B. Giene-Aided Kalman Filter(Ga-KF)

In the Giene-Aided Kalman Filter we have the exact information about the support of the signal. so in presence of Gaussian Noise Ga-KF will give us the optimal solution. That is why Ga-KF is our benchmark to compare with. We will often compare our Kalman Filter Modified Compressed Sensing Result with the Ga-KF result. So before we go into the details of the Kf-ModCS algorithm lets have a review of the Ga-KF algorithm. First we will give an elaborate description and then we will present its algorithmic form.

In kalman filter decoding there are some parameters that need to be described first. One important parameter is  $P_t$  which is the covariance of the error signal  $(x_t - \hat{x}_t)$  updated at time  $t$ .  $Q_t$  is the covariance matrix of the signal  $\nu_t$  updated at time  $t$ . Suppose at time  $t$ , the support of the signal is  $N_t$  and there is no addition or deletion at that time. Then it is a regular kalman filter. We have some  $P_{t-1}$ ,  $\hat{x}_{t-1}$  estimate from the previous time instant. We can evaluate the  $Q_t$  depending upon the variance of the  $\nu_t$  and the support set size  $|N_t|$ . From these data we calculate  $P_{t|t-1}$ , which is  $P_{t|t-1} = P_{t-1} + Q_t$ . Once we have  $P_{t|t-1}$  and  $\hat{x}_{t-1}$  we can use the following equations to get  $P_t$ ,  $\hat{x}_t$ .

$$\begin{aligned}
K_t &= P_{t|t-1} A' (A P_{t|t-1} A' + \sigma_{obs}^2 I)^{-1} \\
P_t &= (I - K_t A) P_{t|t-1} \\
\hat{x}_t &= (I - K_t A) \hat{x}_{t-1} + K_t y_t
\end{aligned} \tag{4}$$

Equations. (4) are the celebrated kalman filter equations which are frequently used to find optimal solution. We will use these equations again and again when there is addition or deletion or no support change. These equations will remain same. Only change is in the pre-updates of  $P_{t|t-1}$  and  $\hat{x}_{t-1}$ .

So first think of the starting of decoding process, i.e. at time  $t = t_0$ . Initialize  $\hat{x}_{t_0-1}$  as all zero vector. At time  $t_0$  there are  $|N_0|$  number of indices added to the support with each index is a Gaussian random variable with mean 0 and variance  $\sigma_{sys,0}^2$ .

We can incorporate this information either in  $P_{t_0-1}$  matrix or  $Q_{t_0}$  matrix. In either case one of them would be all zero matrix. Lets update the  $P_{t_0-1}$  matrix with that information because  $P_t$  matrix is associated with the signal  $x_t$ . Then  $P_{t_0-1} = \sigma_{sys,0}^2 I_{N_0}$  and leave  $Q_{t_0}$  as all-zero matrix because  $\nu_{t_0}$  is all-zero. So  $P_{t_0|t_0-1} = P_{t_0-1} + Q_{t_0}$ . Now use the Equations. (4) to get  $P_{t_0}$  and  $\hat{x}_{t_0}$ .

Now consider the case when additions occur at time  $t_{add,j}$ . The new support set would be  $T_{new} = N_t$ . And  $T = \hat{N}_{t-1} = N_{t-1}$ ,  $\Delta_{add} = N_t \setminus N_{t-1}$ . Let  $P_{t-1}$  and  $\hat{x}_{t-1}$  are the covariance matrix and the input estimate from the previous time instant. At time  $t_{add,j}$  actually two things are happening. The first one is that some new indices with mean 0 and variance  $\sigma_{sys,0}^2$  are added. The second one is that each of the old indices would be added by a Gaussian random variable with mean 0 and variance  $\sigma_{sys}^2$ . The first phenomena is captured in the  $P_{t-1}$  matrix and the second one is captured in the  $Q_t$  matrix. The updates are :

$$\begin{aligned}
Q_t &= \sigma_{sys}^2 I_T \\
(\hat{x}_{t-1})_{\Delta_{add}} &= 0 \\
P_{t|t-1} &= P_{t-1} + Q_t
\end{aligned}$$

Now use Equations .(4) to get  $\hat{x}_t$  and  $P_t$ .

The last scenario is when a deletion occurs. Support set at time  $t = t_{del,j}$  is  $T_{new} = N_t$ . And  $T = \hat{N}_{t-1} = N_{t-1}$ ,  $\Delta_{del} = N_{t-1} \setminus N_t$ . Then the updates would be

$$\begin{aligned}
(P_{t-1})_{\Delta_{del}, [1,m]} &= 0 \\
(P_{t-1})_{[1,m], \Delta_{del}} &= 0 \\
Q_t &= \sigma_{sys}^2 I_{T_{new}} \\
P_{t|t-1} &= P_{t-1} + Q_t \\
(\hat{x}_{t-1})_{\Delta_{del}} &= 0
\end{aligned}$$

Finally use the Equations. (4) to get  $\hat{x}_t$  and  $P_t$ .

In the following we will try to present the above algorithm in a compressed and proper form :

#### GaKF-Algorithm 1:

- 1) Initialization at time  $t = t_0$ 
  - **Support**  $= N_0$ .
  - $P_{t-1} = \sigma_{sys,0}^2 I_{N_0}$ ,  $Q_t = 0$ ,  $\hat{x}_{t-1} = 0$ .
  - $P_{t|t-1} = P_{t-1} + Q_t$ .
  - **Use Equations. (4) to get  $\hat{x}_t$ ,  $P_t$ .**
- 2) Decoding at time  $t > t_0$ 

**for  $t = t_0 + 1$  : end**

$T = N_{t-1}$

  - a) **If  $t \neq t_{add,j}$  and  $t \neq t_{del,j}$** 
    - $Q_t = \sigma_{sys}^2 I_T$ .
    - $P_{t|t-1} = P_{t-1} + Q_t$ .
    - **Use Equations. (4) to get  $\hat{x}_t$ ,  $P_t$**
  - b) **If  $t = t_{add,j}$** 
    - $\Delta_{add} = N_t \setminus N_{t-1}$ ,
    - $Q_t = \sigma_{sys}^2 I_T$ ,  $(P_{t-1})_{\Delta_{add}, \Delta_{add}} = \sigma_{sys,0}^2$
    - $P_{t|t-1} = P_{t-1} + Q_t$ ,
    - **Use Equations. (4) to get  $\hat{x}_t$ ,  $P_t$**

- c) **If**  $t = t_{del,j}$
- $\Delta_{del} = N_{t-1} \setminus N_t, T_{new} = N_t, Q_t = \sigma_{sys}^2 I_{T_{new}}$ .
  - $(P_{t-1})_{\Delta_{del}, [1,m]} = 0, (P_{t-1})_{[1,m], \Delta_{del}} = 0$
  - $P_{t|t-1} = Q_t + P_{t-1}, (\hat{x}_{t-1})_{\Delta_{del}} = 0$ .
  - **Use Equations. (4) to get**  $\hat{x}_t, P_t$

### C. Kalman Filtered Modified Compressive Sensing

Here in this section we will discuss about the kalman Filtered Modified Compressive Sensing, abbreviated as Kf-ModCS, for the Signal Model 1. Estimate of addition set is denoted as  $\Delta_A = \hat{N}_t \setminus \hat{N}_{t-1}$  whereas the estimate of deletion set is denoted as  $\Delta_D = \hat{N}_{t-1} \setminus \hat{N}_t$ . And by using  $T$  we are actually denoting  $\hat{N}_{t-1}$  as mentioned in the notation. We will see that the difference between Ga-KF and Kf-ModCS is in the size of the support set of the signal. In Ga-KF we know the support but in Kf-ModCS we have to estimate the support with the help of compressed sensing. After we get the support we will apply Kalman Filter like we do in Ga-KF.

At time  $t = t_0$ , we have  $y_t$ , the measurement matrix  $A$ , the variance of signal co-efficient  $\sigma_{sys,0}^2$ , noise variance  $\sigma_{obs}^2$ . But we do not have any information about  $N_{t_0}$ . So we can't simply apply kalman filter on  $y_t$ . For that at least we have to estimate the support set. Here compressive sensing will help us. Compressive sensing gives a reconstructed signal using the output and measurement matrix. We use the normal BPDN to compute the CS reconstructed signal. The BPDN algorithm can be described as

$$\begin{aligned} & \min_x \|x\|_1 \\ & s.t \ \|y - Ax\|_2 < \xi \end{aligned} \quad (5)$$

$\xi$  depends on  $\sigma_{obs}^2$ . Let the CS-reconstructed signal at time  $t_0$  is denoted as  $\hat{x}_{t_0, CSres}$  and  $\hat{N}_{t_0}$  is the support estimate using an addition threshold  $\alpha_{t_0, add}$ . We will do lot of discussion of how to determine suitable addition threshold in the next section. But for now let us assume that we already have the addition threshold. Once we get the support size we will do exactly the same thing as we do for Ga-KF decoding. Create  $P_{t-1}, Q_t, \hat{x}_{t-1}$  in the same way as Ga-KF. Evaluate  $P_t$  and  $\hat{x}_t$  from there using Equations (4).

From the next time instance, as we don't know the support addition time or deletion time, we depends on the Mod-CS technique to find if there is any addition or deletion. Using  $T$ ,  $\sigma_{sys}^2$  first we will calculate  $Q_{tmp}$ , which is  $Q_{tmp} = \sigma_{sys}^2 I_T$ . Then  $(P_{tmp})_{t|t-1} = P_{t-1} + Q_{tmp}$ . Then compute  $\hat{x}_{t, init}$  using Equations. (4). The residual of  $y_t, \tilde{y}_{t, res}$  is defined as  $\tilde{y}_{t, res} = y_t - A\hat{x}_{t, init}$ . We will apply Mod-CS on  $\tilde{y}_{t, res}$  according to following equation :

$$\begin{aligned} & \min_x \|(x)_{T^c}\|_1 \\ & s.t \ \|y - Ax\|_2 < \xi \end{aligned} \quad (6)$$

which will give us  $\hat{\beta}_t$  as output.  $\hat{x}_{t, CSres} = \hat{x}_{t, init} + \hat{\beta}_t$  is our reconstructed signal at time  $t$ . Now we try to find if there is any addition in the support by applying the addition threshold  $\alpha_{t, add}$  on  $\hat{x}_{t, CSres}$ . We denote  $\Delta_A$  to be the set of new indices. Then the new support set for the signal would be  $T_{det} = T \cup \Delta_A$ . Once we have  $P_{t-1}, T, T_{det}, \hat{x}_{t-1}$  we

can easily apply the kalman Filter technique as we do in Ga-KF. If  $\Delta_A$  is empty we have to apply normal kalman filter or if  $\Delta_A$  is non-empty we have to apply kalman filter with addition. At the end of this operation we have  $\hat{x}_{inter}$  as our reconstructed signal. We have the deletion threshold  $\alpha_{t, del}$  to check if there is any index to be deleted from the set  $T_{det}$ . The method of finding the deletion threshold would be discussed in detail in the later section. But for the time being we believe that there is some appropriate threshold for deletion which determines which indices should be deleted and which should be kept in the support set. In this way we will get another support set  $T_{new} = T_{det} \setminus \Delta_D$ , where  $\Delta_D$  is the set of indices deleted at time  $t$ . If  $\Delta_D$  is an empty set then we don't have to do anything and  $\hat{x}_t = \hat{x}_{inter}, P_t$  would be same as we get during the previous step. But if  $\Delta_t$  is not an empty set then we have to be little bit careful during the updating of  $P_{t|t-1}$ . If we see the previous operation we notice that whether  $\Delta_A$  is empty or not,  $P_{t|t-1}$  matrix has been updated. So that will be our starting point and we will directly update the  $P_{t|t-1}$  and  $\hat{x}_{t-1}$  using  $\Delta_D$ . Then apply the usual kalman filter technique to get  $\hat{x}_t, P_t$ . The complete algorithm in compressed form is given below :

#### Kf-ModCS Algorithm 1:

- 1) Initialization at time  $t = t_0$ 
  - **Apply CS on  $y_{t_0}$  using equation (5) to get**  $\hat{x}_{t_0, CSres}$
  - **Use  $\alpha_{t_0, add}$  to get Support**  $= \hat{N}_{t_0}$ .
  - $P_{t-1} = \sigma_{sys,0}^2 I_{\hat{N}_{t_0}}, Q_t = 0, \hat{x}_{t-1} = 0$ .
  - $P_{t|t-1} = P_{t-1} + Q_t$ .
  - **Use Equations. (4) to get**  $\hat{x}, P_t$ .
- 2) Decoding at time  $t > t_0$ 

**for  $t = t_0 + 1$  : end**

  - $T = \hat{N}_{t-1}$
  - $Q_{tmp} = \sigma_{sys}^2 I_T, (P_{tmp})_{t|t-1} = P_{t-1} + Q_{tmp}$
  - **Use Equations. (4) to get**  $\hat{x}_{t, init}$ .
  - $\tilde{y}_{t, res} = y_t - A\hat{x}_{t, init}$ .
  - **Apply ModCS on  $\tilde{y}_{t, res}$  using  $T$  to get the output**  $\hat{\beta}_t$
  - $\hat{x}_{t, CSres} = \hat{x}_{t, init} + \hat{\beta}_t$
  - **apply  $\alpha_{t, add}$  on  $\hat{x}_{t, CSres}$  to measure  $\Delta_A$**
  - **If  $\Delta_A$  is empty**
    - $T_{det} = T$ .
    - $Q_t = \sigma_{sys}^2 I_T$ .
    - $P_{t|t-1} = P_{t-1} + Q_t$ .
    - **Use Equations. (4) to get**  $\hat{x}_{inter}, P_t$
  - **If  $\Delta_A$  is non-empty**
    - $T_{det} = T \cup \Delta_A$ .
    - $Q_t = \sigma_{sys}^2 I_T, (P_{t-1})_{\Delta_A, \Delta_A} = \sigma_{sys,0}^2$
    - $P_{t|t-1} = P_{t-1} + Q_t$ ,
    - **Use Equations. (4) to get**  $\hat{x}_{inter}, P_t$
  - **Use  $\alpha_{t, del}$  on  $\hat{x}_{inter}$  to find  $\Delta_D$  from the support set**
  - **If  $\Delta_D$  is empty**
    - $T_{new} = T_{det}$
    - $\hat{x}_t = \hat{x}_{inter}$

- **If  $\Delta_D$  is non-empty**
  - $T_{new} = T_{det} \setminus \Delta_D$ ,
  - $(P_{t|t-1})_{\Delta_D, [1, m]} = 0$ ,  $(P_{t|t-1})_{[1, m], \Delta_D} = 0$
  - $(\hat{x}_{t-1})_{\Delta_D} = 0$ .
  - **Use Equations. (4) to get  $\hat{x}_t$ ,  $P_t$**
- $\hat{N}_t = T_{new}$ .

## II. KF-MODCS ALGORITHM WITH MORE GENERAL SIGNAL MODEL

In the previous section we have described our Kf-ModCS algorithm with a simple random walk model where we know some information about the signal. In our decoding we use those information like the signal variances of  $x_t$  and  $\nu_t$ . Although we don't use any other given information like the sparsity size or the sparsity change rate, the signal model is such that some information are implicit. Consider the following signal model equation

$$x_{t+1} = (1 - \lambda_{cor})x_t + \beta_{cor}\nu_{t+1} \quad (7)$$

where  $\lambda_{cor}$  and  $\beta_{cor}$  are correlation factors which determine how much the signal is correlated with itself over time. In our signal model. 1 those factors are fixed at 0 and 1. And we use those information during decoding. But in some practical situation the scenario can be very bad when we don't have any knowledge about  $\lambda_{cor}$ ,  $\beta_{cor}$ ,  $\sigma_{sys,0}^2$ ,  $\sigma_{sys}^2$ . For those cases our algorithm should be robust enough to work efficiently. Before we move into the discussion about algorithm lets create a new signal model.

**Signal Model 2:** Initial Time is denoted as  $t_0$ .

- 1) At  $t = t_0$ ,  $x_0$  is  $S_0$  sparse with support  $N_0$  and  $(x_0)_{N_0} \sim \mathcal{N}(0, \sigma_{sys,0}^2 I)$ .
- 2) At each time  $t > t_0$ , we calculate  $N_t$  and  $N_t^c$ . We are associating a probability  $p_{10}$  to every indices of  $N_t^c$  so that some of the indices become non-zero from 0, and similarly associating a probability  $p_{01}$  to every indices of  $N_t$  so that some indices of  $N_t$  become 0.
- 3) Denote the time indices for addition as  $t_{add}$  and for deletion as  $t_{del}$ .
- 4) Every new indices getting added at time  $t$ , follows a Gaussian Distribution of mean 0 and variance  $\sigma_{sys,0}^2$ .
- 5) The currently non-zero indices of  $x_t$ , follows the below update model :

$$x_t = (1 - \lambda_{cor})x_{t-1} + \beta_{cor}\nu_t \quad (8)$$

where  $(\nu_t)_{N_t} \sim \mathcal{N}(0, \sigma_{sys}^2 I)$ .

### A. Ga-KF with Signal Model 2

Ga-KF with Signal Model 2 will not be very much different from Signal Model 1. There are only little changes in the equations of Kalman Filter. At time  $t = t_0$  the update of  $P_{t-1}$ ,  $Q_t$ ,  $\hat{x}_{t-1}$  would be exactly same as with Ga-KF with Signal Model 1. For time  $t > t_0$  there are some changes in the updates. Let at time  $t > t_0$ , the support set is  $N_t$  and we have  $P_{t-1}$ ,  $\hat{x}_{t-1}$  from previous time instant. If there is no addition

or deletion Ga-KF will follow the below equations :

$$\begin{aligned} \hat{x}_{filter} &= (1 - \lambda_{cor})\hat{x}_{t-1} \\ Q_t &= \sigma_{sys}^2 I_{N_t} \\ P_{t|t-1} &= (1 - \lambda_{cor})^2 P_{t-1} + \beta_{cor}^2 Q_t \end{aligned}$$

After getting  $P_{t|t-1}$  and  $\hat{x}_{filter}$  we can use the following equation to reconstruct  $\hat{x}_t$ .

$$\begin{aligned} K_t &= P_{t|t-1} A' (A P_{t|t-1} A' + \sigma_{obs}^2 I)^{-1} \\ P_t &= (I - K_t A) P_{t|t-1} \\ \hat{x}_t &= (I - K_t A) \hat{x}_{filter} + K_t y_t \end{aligned} \quad (9)$$

If there is addition of indices then the kalman filter equation would be :

$$\begin{aligned} \hat{x}_{filter} &= (1 - \lambda_{cor})\hat{x}_{t-1} \\ Q_t &= \sigma_{sys}^2 I_{N_t} \\ P_{t|t-1} &= (1 - \lambda_{cor})^2 P_{t-1} + \beta_{cor}^2 Q_t \\ (P_{t|t-1})_{\Delta_{add}, \Delta_{add}} &= \sigma_{sys,0}^2 \end{aligned}$$

Then use equ.(9) to get  $\hat{x}_t$ ,  $P_t$ .

Finally if there is deletion then we will start from  $P_{t|t-1}$  and  $\hat{x}_{filter}$  and modify them according to the deletion set  $\Delta_{del}$ .

$$\begin{aligned} (P_{t|t-1})_{\Delta_{del}, [1, m]} &= 0 \\ (P_{t|t-1})_{[1, m], \Delta_{del}} &= 0 \\ (\hat{x}_{filter})_{\Delta_{del}} &= 0 \end{aligned}$$

again use equ. (9) to get  $\hat{x}_t$ . The update equations for this signal model are essentially same as for signal model 1. We see one difference because of different correlation co-efficient. Another minor difference is that addition and deletion of indices can happen at the same time instant here in signal model 2. So the algorithm can be written in the following form :

### GaKF-Algorithm 2:

- **Initialization at time  $t = t_0$** 
  - **Support** =  $N_0$ .
  - $P_{t-1} = \sigma_{sys,0}^2 I_{N_0}$ ,  $Q_t = 0$ ,  $\hat{x}_{t-1} = 0$ .
  - $P_{t|t-1} = P_{t-1} + Q_t$ .
  - **Use Equations. (4) to get  $\hat{x}_t$ ,  $P_t$ .**
- **Decoding at time  $t > t_0$** 

**for  $t = t_0 + 1$  : end**

$T = N_{t-1}$

  - **If  $t \neq t_{add}$  and  $t \neq t_{del}$** 
    - \*  $Q_t = \sigma_{sys}^2 I_T$ .
    - \*  $P_{t|t-1} = (1 - \lambda_{cor})^2 P_{t-1} + \beta_{cor}^2 Q_t$ .
    - \*  $\hat{x}_{filter} = (1 - \lambda_{cor})\hat{x}_{t-1}$
    - \* **Use Equations. (9) to get  $\hat{x}_t$ ,  $P_t$**
  - **Else If  $t = t_{add}$** 
    - \*  $\hat{x}_{filter} = (1 - \lambda_{cor})\hat{x}_{t-1}$
    - \*  $Q_t = \sigma_{sys}^2 I_{N_t}$
    - \*  $P_{t|t-1} = (1 - \lambda_{cor})^2 P_{t-1} + \beta_{cor}^2 Q_t$
    - \*  $(P_{t|t-1})_{\Delta_{add}, \Delta_{add}} = \sigma_{sys,0}^2$
    - \* **Use Equations. (9) to get  $\hat{x}_t$ ,  $P_t$**

- If  $t = t_{del}$ 
  - \*  $Q_t = \sigma_{sys}^2 I_T$ ,  $P_{t|t-1} = (1 - \lambda_{cor})^2 P_{t-1} + \beta_{cor}^2 Q_t$ ,
  - $\hat{x}_{filter} = (1 - \lambda_{cor}) \hat{x}_{t-1}$
  - \*  $(P_{t|t-1})_{\Delta_{del}, [1, m]} = 0$
  - \*  $(P_{t|t-1})_{[1, m], \Delta_{del}} = 0$
  - \*  $(\hat{x}_{filter})_{\Delta_{del}} = 0$
  - \* **Use Equations. (9) to get  $\hat{x}_t, P_t$**

### B. Kf-ModCS With Known Information

Kf-ModCS algorithm with information follows the Ga-KF algorithm in every step with the exception is that the support of the signal is to be estimated in each time instant. Once we estimate the support set at time  $t$  we can atleast take a decision if there is any addition or deletion or nothing. After that the kalman filter algorithm would be same as Ga-KF.

#### Kf-ModCS Algorithm 2:

- Initialization at time  $t = t_0$ 
  - Apply CS on  $y_{t_0}$  using equation (5) to get  $\hat{x}_{t_0, CSres}$
  - Use  $\alpha_{t_0, add}$  to get Support =  $\hat{N}_{t_0}$ .
  - $P_{t-1} = \sigma_{sys, 0}^2 I_{\hat{N}_{t_0}}$ ,  $Q_t = 0$ ,  $\hat{x}_{t-1} = 0$ .
  - $P_{t|t-1} = P_{t-1} + Q_t$ .
  - **Use Equations. (4) to get  $\hat{x}_t, P_t$ .**
- Decoding at time  $t > t_0$ 

**for  $t = t_0 + 1$  : end**

  - $T = \hat{N}_{t-1}$
  - $Q_{tmp} = \sigma_{sys}^2 I_T$ ,  $(P_{tmp})_{t|t-1} = (1 - \lambda_{cor})^2 P_{t-1} + \beta_{cor}^2 Q_{tmp}$
  - $\hat{x}_{tmp, filter} = (1 - \lambda_{cor}) \hat{x}_{t-1}$
  - **Use Equations. (9) to get  $\hat{x}_{t, init}$ .**
  - $\tilde{y}_{t, res} = y_t - A \hat{x}_{t, init}$ .
  - **Apply ModCS on  $\tilde{y}_{t, res}$  using  $T$  to get the output  $\hat{\beta}_t$**
  - $\hat{x}_{t, CSres} = \hat{x}_{t, init} + \hat{\beta}_t$
  - **apply  $\alpha_{t, add}$  on  $\hat{x}_{t, CSres}$  to measure  $\Delta_A$**
  - **If  $\Delta_A$  is empty**
    - \*  $T_{det} = T$ .
    - \*  $Q_t = \sigma_{sys}^2 I_T$ .
    - \*  $P_{t|t-1} = (1 - \lambda_{cor})^2 P_{t-1} + \beta_{cor}^2 Q_t$ .
    - \*  $\hat{x}_{filter} = (1 - \lambda_{cor}) \hat{x}_{t-1}$
    - \* **Use Equations. (9) to get  $\hat{x}_{inter}, P_t$**
  - **If  $\Delta_A$  is non-empty**
    - \*  $T_{det} = T \cup \Delta_A$ .
    - \*  $\hat{x}_{filter} = (1 - \lambda_{cor}) \hat{x}_{t-1}$
    - \*  $Q_t = \sigma_{sys}^2 I_T$
    - \*  $P_{t|t-1} = (1 - \lambda_{cor})^2 P_{t-1} + \beta_{cor}^2 Q_t$
    - \*  $(P_{t|t-1})_{\Delta_A, \Delta_A} = \sigma_{sys, 0}^2$
    - \* **Use Equations. (9) to get  $\hat{x}_{inter}, P_t$**
  - **Use  $\alpha_{t, del}$  on  $\hat{x}_{inter}$  to find  $\Delta_D$  from the support set**
  - **If  $\Delta_D$  is empty**
    - \*  $T_{new} = T_{det}$
    - \*  $\hat{x}_t = \hat{x}_{inter}$
  - **If  $\Delta_D$  is non-empty**
    - \*  $T_{new} = T_{det} \setminus \Delta_D$ ,

- \*  $(P_{t|t-1})_{\Delta_D, [1, m]} = 0$
- \*  $(P_{t|t-1})_{[1, m], \Delta_D} = 0$
- \*  $(\hat{x}_{filter})_{\Delta_D} = 0$
- \* **Use Equations. (9) to get  $\hat{x}_t, P_t$**
- $\hat{N}_t = T_{new}$

In the above compressed form of the algorithm we use one variable  $\hat{x}_{inter}$  which is similar to variable of the same name defined during Kf-ModCS algorithm 1.

### C. Kf-ModCS without Information

When we don't have any information about the signal model then task is more challenging. We have to estimate the signal variance at each time instant to run the kalman filter. But for  $\lambda_{cor}$  and  $\beta_{cor}$  we will assume them to be 0 and 1 respectively. That means we are actually assuming the random walk model for Kf-ModCS decoding like we did in Signal Model 1.

#### 1) Estimation of Variance and Mean of $x_t$ and $\nu_t$ :

Estimates of Mean and Variance of  $x_t$  and  $\nu_t$  changes with time as we use more and more data as the time progress. Hence with the change in time the estimation would be more perfect. Following are the assumptions that we use during the estimation:

- Each of the new indices which are added to the support set of  $x_t$ , at any time instant, follows an i.i.d random distribution with some mean and variance.
- Each of the non-zero indices of  $\nu_t$  is zero mean random variable with some variance. So for  $\nu_t$  we don't have to estimate the mean.

We will rely mostly on the kalman filter output to estimate the mean and variance. But for some initial estimation we have to use CS or ModCS output. As we have no data we have to start with CS or ModCS. The estimate from CS data will be used in the same time instant, whereas the estimate from kalman data will be used in the next time instant.

- **Estimation of Mean And Variance for  $x_t$  :** At time  $t = t_0$ ,  $\hat{x}_{t_0, CSres}$  would be used to measure the mean and the variance of  $\hat{x}_{t_0}$ . Let  $\hat{N}_{t_0}$  is the support estimated at time  $t_0$ . Then

$$\mu_{t_0, x} = \frac{1}{|\hat{N}_{t_0}|} [\sum_{i \in \hat{N}_{t_0}} (\hat{x}_{t_0, CSres})_i]$$

$$\sigma_{t_0, x}^2 = \frac{1}{|\hat{N}_{t_0}|} [\sum_{i \in \hat{N}_{t_0}} (\hat{x}_{t_0, CSres})_i^2] - (\mu_{t_0, x})^2 \quad (10)$$

We will use these mean and variance in the kalman Filter update at time  $t = t_0$ . Now say  $\hat{x}_{t_0}$  is the kalman-Filter output at the end of time instant  $t_0$ . Then

$$\mu_{t_0+1, x} = \frac{1}{|\hat{N}_{t_0}|} [\sum_{i \in \hat{N}_{t_0}} (\hat{x}_{t_0})_i]$$

$$\sigma_{t_0+1, x}^2 = \frac{1}{|\hat{N}_{t_0}|} [\sum_{i \in \hat{N}_{t_0}} (\hat{x}_{t_0})_i^2] - (\mu_{t_0+1, x})^2 \quad (11)$$

Although the above mean and the variance are calculated at the end of time  $t_0$  with support estimate  $\hat{N}_{t_0}$  they will be used at  $t = t_0 + 1$  when there is any new addition.

Now lets see the situation at time  $t_1 = t_0 + 1$ . If  $\Delta_{t_1,A}$  is the set of new indices added to the support set  $\hat{N}_{t_0}$  then the equations can be written as :

$$\begin{aligned}\mu_{t_0+2,x} &= \frac{\sum_{i \in \hat{N}_{t_0}} (\hat{x}_{t_0})_i + \sum_{i \in \Delta_{t_1,A}} (\hat{x}_{t_0+1})_i}{|\hat{N}_{t_0}| + |\Delta_{t_1,A}|} \\ \sigma_{t_0+2,x}^2 &= \frac{\sum_{i \in \hat{N}_{t_0}} (\hat{x}_{t_0})_i^2 + \sum_{i \in \Delta_{t_1,A}} (\hat{x}_{t_0+1})_i^2}{|\hat{N}_{t_0}| + |\Delta_{t_1,A}|} - \mu_{t_0+2,x}^2\end{aligned}\quad (12)$$

Then at the end of any time  $t = t_0 + k$ , if set of new indices is denoted by  $\Delta_{t_j,A}$  for  $j = 1, 2, \dots, k$  then the general equation of mean and variance estimate can be written as :

$$\begin{aligned}\mu_{t_0+k+1,x} &= \frac{\sum_{i \in \hat{N}_{t_0}} (\hat{x}_{t_0})_i + \sum_{j=1}^k \sum_{i \in \Delta_{t_j,A}} (\hat{x}_{t_0+j})_i}{|\hat{N}_{t_0}| + \sum_{j=1}^k |\Delta_{t_j,A}|} \\ \sigma_{t_0+k+1,x}^2 &= \frac{\sum_{i \in \hat{N}_{t_0}} (\hat{x}_{t_0})_i^2 + \sum_{j=1}^k \sum_{i \in \Delta_{t_j,A}} (\hat{x}_{t_0+j})_i^2}{|\hat{N}_{t_0}| + \sum_{j=1}^k |\Delta_{t_j,A}|} \\ &\quad - \mu_{t_0+k+1,x}^2\end{aligned}\quad (13)$$

- **Estimation of Variance of  $\nu_t$**  : For  $\nu_t$  we have to estimate only the variance, as we already assume that the mean is zero. At time  $t = t_0$ ,  $\hat{\nu}_t$  is all zero. So for  $t = t_0$  we don't have to do anything for  $\hat{\nu}_t$ .

But at time  $t = t_0 + 1$ , we need to rely on Mod-CS for variance estimation. If we want to do the Initial kalman filter estimation step before residual-ModCS we need to know the variance of  $\hat{\nu}_{t_0+1}$ . For that at the very beginning we need to apply Mod-CS on  $y_{t_0+1}$ , which implies that we have to apply ModCS twice at time  $t_0 + 1$ . That will reduce the speed of the operation. So we can skip the residual ModCS for  $t = t_0 + 1$  and simply apply ModCS on  $y_{t_0+1}$ , which gives us  $\hat{x}_{t_0+1,CSres}$ . Then

$$\begin{aligned}\hat{\nu}_{t_0+1} &= \hat{x}_{t_0+1,CSres} - \hat{x}_{t_0} \\ \sigma_{t_0+1,\nu}^2 &= \frac{1}{|\hat{N}_{t_0}|} \left[ \sum_{i \in \hat{N}_{t_0}} \{(\hat{x}_{t_0+1,CSres})_i - (\hat{x}_{t_0})_i\}^2 \right]\end{aligned}\quad (14)$$

We will use this variance for the time instant  $t = t_0 + 1$  for different kalman filter operations. At the end of the time instant  $t = t_0 + 1$ , we will get

$$\sigma_{t_0+2,\nu}^2 = \frac{1}{|\hat{N}_{t_0}|} \left[ \sum_{i \in \hat{N}_{t_0}} \{(\hat{x}_{t_0+1})_i - (\hat{x}_{t_0})_i\}^2 \right]\quad (15)$$

Notice that in the above we calculate over the old support set because the support set of  $\hat{\nu}_{t_0+1}$  at time  $t_0 + 1$  is  $\hat{N}_{t_0}$ . Once we get  $\sigma_{t_0+2,\nu}^2$  we can do the temporary kalman filter operation at time  $t_0 + 2$  using this variance. At the end of time interval  $t = t_0 + 2$ , the estimation

would be

$$\begin{aligned}\sigma_{t_0+3,\nu}^2 &= \frac{\left[ \sum_{i \in \hat{N}_{t_0}} \{(\hat{x}_{t_0+1})_i - (\hat{x}_{t_0})_i\}^2 \right]}{|\hat{N}_{t_0}| + |\hat{N}_{t_0+1}|} \\ &\quad + \frac{\left[ \sum_{i \in \hat{N}_{t_0+1}} \{(\hat{x}_{t_0+2})_i - (\hat{x}_{t_0+1})_i\}^2 \right]}{|\hat{N}_{t_0}| + |\hat{N}_{t_0+1}|}\end{aligned}\quad (16)$$

Then the general equation for estimation at the end of time interval  $t = t_0 + k + 1$ ,

$$\sigma_{t_0+k+2,\nu}^2 = \frac{\sum_{j=0}^k \left[ \sum_{i \in \hat{N}_{t_0+j}} \{(\hat{x}_{t_0+j+1})_i - (\hat{x}_{t_0+j})_i\}^2 \right]}{\sum_{j=0}^k |\hat{N}_{t_0+j}|}\quad (17)$$

**2) Threshold Seletion And simulation:** As we analyse our algorithms we see that there are actually two different types of thresholds calculation, one is addition threshold( $\alpha_{add}$ ) and another is deletion threshold( $\alpha_{del}$ ). To determine all these thresholds first we will review the threshold determination techniques in the paper **"Time Invariant Error Bounds for Modied-CS based Sparse Signal Sequence Recovery"** by **Jinchun Zhan and Namrata Vaswani**. So we start with defining  $x_{min,t}$  as  $x_{min,t} = \min_{j \in N_t} (|x_t|)$  and  $\hat{x}_{min,t} = \min_{j \in \hat{N}_{t-1}} (|\hat{x}_t|)$ . We know that  $\hat{N}_{t-1}$  is the support estimate at time  $t = t - 1$ .

- $\alpha_{add}$  :

**Addition Threshold Evaluation Method 1:** We detect the additional indices by the following method, Let at time  $t$ , the pre-estimated support size  $T = \hat{N}_{t-1}$  and Mod-CS reconstructed output is  $\hat{x}_{t,CSres}$ . Then the set of additional indices is defined as

$$\Delta_A = \{i \in [1, m] : |(\hat{x}_{t,CSres})_i| > \alpha_{add}\} \setminus T$$

In the above paper the authors have suggested that the  $\alpha_{add}$  in the above equation would be the smallest number such that, the minimum singular value of the sub-matrix  $A_{T \cup \Delta_A}$  is greater than a threshold value. We denote singular value by  $\vartheta$ , then the condition is

$$\vartheta_{min}(A_{T \cup \Delta_A}) \geq \text{fixed value}\quad (18)$$

That fixed value could be any value greater than 0, which can be set depending upon the nature of the simulation. The authors got the above condition by bounding the error between the actual signal and the ModCS-LS estimate. The above condition will preserve the rank of the sub-matrix  $A_{T \cup \Delta_A}$ . And we are denoting  $\alpha_{add}$  as the infimum of the absolute value of the new indices.

$$\begin{aligned}\alpha_{add} &= \inf_{i \in \Delta_A} |(\hat{x}_{t,CSres})_i| \text{ if } \Delta_A \text{ non-empty} \\ &= \infty \text{ if } \Delta_A \text{ empty}\end{aligned}\quad (19)$$

- $\alpha_{add}$

**Addition Threshold Evaluation Method 2:** Now we can slightly modify the above algorithm to make a more random  $\alpha_{add}$  selection. We bring back the

subscript  $t$  for  $\alpha_{add}$ . Here also we have to use the ModCS output for detecting the new indices and calculation of addition threshold. At time  $t = t_0$ ,  $T = \hat{N}_{t_0-1} = \Phi$ . Apply equ.(18) and equ.(19) to get  $\Delta_A$  and  $\alpha_{t,add}$ . Let  $A_j$  is the  $j$ -th column of  $A$ . At time  $t_0$ , we set fixed value to a value such that fixed value  $\leq \min_j [\vartheta_{min}(A_j)] = \min_j \|A_j\|$ . Then  $\hat{N}_{t_0} = \Delta_A$  can not be an empty set. From the next time instant, we will follow the below procedures :

- 1)  $T = \hat{N}_{t-1}$
- 2) set two new variables  $\vartheta_{high}, \vartheta_{low}$  such that  $\vartheta_{high} > \text{fixed value}$  and  $\vartheta_{low} < \text{fixed value}$ .
- 3) Set  $\alpha_{add,tmp} = \alpha_{t-1,add}$ .
- 4) Calculate  $\Delta_{A,tmp}$ .

$$\Delta_{A,tmp} = \{i \in [1, m] : |(\hat{x}_{t,CSres})_i| > \alpha_{add,tmp}\} \setminus T$$

- 5)  $T_{det,tmp} = T \cup \Delta_{A,tmp}$
- 6) if  $\vartheta_{low} \leq \vartheta_{min}(A_{T_{det,tmp}}) \leq \vartheta_{high}$ 
  - $\Delta_A = \Delta_{A,tmp}$
  - $T_{det} = T_{det,tmp}$
  - $\alpha_{t,add} = \alpha_{add,tmp}$
  - end
- 7) if  $\vartheta_{low} > \vartheta_{min}(A_{T_{det,tmp}})$ 
  - Get a new  $\Delta_A$  and  $\alpha_{t,add}$  using equ.(18) and equ.(19) with fixed value =  $\vartheta_{low}$ .
  - end,
- 8) if  $\vartheta_{min}(A_{T_{det}}) > \vartheta_{high}$ 
  - Get a new  $\Delta_A$  and  $\alpha_{t,add}$  using equ.(18) and equ.(19) with fixed value =  $\vartheta_{high} - \delta$ , where  $\delta$  is small positive value.
  - end,

we can see that this way of estimating the addition threshold is faster.

- **Motivation Behind the Method.2:** In the first method we have only one parameter that is controlling the set of extra indices  $\Delta_A$ . That parameter is the minimum singular value of the sub matrix, formed by the columns of  $A$ , indexed by the total support set. Now assume at time  $t = t_0$  we have set the fixed value to some small value  $\vartheta_{low}$  so that large number of indices can be added to the support set  $\hat{N}_{t_0-1} = \Phi$ . But after time  $t > t_0$ , we set the fixed value to a larger value  $\vartheta_{high}$  so that lesser number of indices would be added in the support. But if a situation occurs when number of addition in the true signal support is large then we don't have any reverse mechanism to go back to small fixed value. In the other word, we don't have any control over setting the fixed value parameter. Here we think our second threshold calculation method is better than previous one. Because in this algorithm three parameters are actually controlling the  $\Delta_A$ . Two parameters are  $\vartheta_{low}, \vartheta_{high}$  and the third one is  $\alpha_{add}$  itself. So two parameters are singular values and the third one is magnitude parameter. We get the initial  $\alpha_{add}$  using fixed value at time  $t = t_0$ .

From the next time instant we are using that  $\alpha_{add}$  for calculating  $\Delta_A$ . If the new  $T_{det}$  created by  $\alpha_{add}$  satisfies the singular value condition of step.6 then we are keeping  $\alpha_{add}$  same for the next time instant, which is fine because all our conditions are maintained. If the minimum singular value  $\vartheta_{min}(A_{T_{det}})$  is less than  $\vartheta_{low}$  our algorithm starts to search for a higher threshold and  $\Delta_A$  so that the lower bound on the minimum singular value is satisfied. Again when  $\vartheta_{min} > \vartheta_{high}$  our threshold is too high. Then this method will automatically adjust the  $\alpha_{add}$  to a lower value by executing the step.8. In that way both the magnitude parameter and the singular value parameter are influencing each other to get the best estimate of  $\Delta_A$ . Another advantage of the later algorithm is the speed. It is little bit faster than the previous method.

#### • $\alpha_{del}$ :

**Deletion Threshold Evaluation Method 1:** Deletion threshold detection is based on the fact that we don't want to miss any indices of the signal  $x_t$ . So if we set  $\alpha_{del} < x_{min,t}$  as our deletion threshold then no true index will be deleted because the threshold is less than the minimum value of  $x_t$ . But in that case we have to know the signal values correctly, which is not possible. The best we can do is to estimate the minimum value  $\hat{x}_{t-1}$ , which is denoted as  $\hat{x}_{min,t}$  and then set the threshold as  $\alpha_{del} < \hat{x}_{min,t}$ . Now the question is how small  $\alpha_{del}$  would be than  $\hat{x}_{min,t}$ . In our KF-ModCS decoding we do the ModCS first, then we detect if any addition occurs or not. Depending on that we apply the kalamn filter based on the new support size. We denote the output of the kalman filter as  $\hat{x}_{inter}$  and the new support set as  $T_{det}$  in the earlier section. We want to apply  $\alpha_{del}$  on  $\hat{x}_{inter}$  to check deletion of indices. If

$$\alpha_{del} = \text{Inf} [\hat{x}_{min,t} - \|(x_t - \hat{x}_{inter})_{T_{det}}\|_\infty] \quad (20)$$

then we can avoid misses in the support. In the above mentioned paper the authors have introduced similar lower bound,

$$\alpha_{del} = \text{Inf} [\hat{x}_{min,t} - \|(x_t - \hat{x}_{T_{det}}^{LS})_{T_{det}}\|_\infty] \quad (21)$$

where  $\hat{x}_{T_{det}}^{LS}$  is the ADD-LS estimate at time  $t$ . They are applying deletion threshold on the Add-LS estimate. Our bound equ.(20) is inspired from that bound. But there is a problem while evaluating the norm  $\|(x_t - \hat{x}_{inter})_{T_{det}}\|_\infty$ . Let us denote the  $e_t$  as the error  $x_t - \hat{x}_{inter}$ . Then

$$\begin{aligned} e_t &= x_t - \hat{x}_{inter} \\ &= x_t - [(I - K_t A) \hat{x}_{t-1} + K_t y_t] \\ &= x_t - (I - k_t A) \hat{x}_{t-1} - k_t (A x_t + w_t) \\ &= (I - K_t A) x_t - (I - K_t A) \hat{x}_{t-1} - K_t w_t \end{aligned}$$

The problem with the last line of the equation is that we can't simply write  $x_t = x_{t-1} + \nu_t$  because there may be addition or deletion of indices, which makes the recursion formula very difficult to track. That is why we reject that bound. Hence we will rely on a simple assumption. Let at time  $t$ , after we get our new support  $T_{det}$  we evaluate  $\hat{x}_{T_{det}}^{LS} = A_{T_{det}}^\dagger y_t$ , where  $A_{T_{det}}^\dagger$  is the pseduo-inverse of

$A_{T_{det}}$ . Now we assume that  $\|x_t - \hat{x}_{inter}\| \approx \|x_t - \hat{x}_{T_{det}}^{LS}\|$  which will make the calculation a lot easier. We are applying this heuristic because both  $\hat{x}_{inter}$  and  $\hat{x}_{T_{det}}^{LS}$  are evaluated based on the support size  $T_{det}$  and with submatrix  $A_{T_{det}}$ . Now if we follow the procedure of evaluation of the bound in the above paper [ ] we find that

$$\begin{aligned} \|(x_t - \hat{x}_{T_{det}}^{LS})_{T_{det}}\|_{\infty} &\approx 0.3\hat{x}_{min,t} \\ &+ \|A^{\dagger}(y_t - A\hat{x}_{t,CSres})\|_{\infty} \end{aligned} \quad (22)$$

which gives us

$$\alpha_{del} = \text{Inf}[0.7\hat{x}_{t,min} - \|A^{\dagger}(y_t - A\hat{x}_{t,CSres})\|_{\infty}] \quad (23)$$

- $\alpha_{del}$

**Deletion Threshold Evaluation Method 2:** If we go back to equ.(20)  $\|x_t - \hat{x}_{inter}\|$  can get replaced with similar norm relation.  $\hat{x}_{t,CSres}$  is the ModCS output at time  $t$ . If we assume that we detect the extra indices correctly then we can write  $\|x_t - \hat{x}_{inter}\| \approx \|x_t - \hat{x}_{t,CSres}\|$  from the stability analysis. From the paper [Time Invariant Error Bounds for Modified-CS based Sparse Signal Sequence Recovery],  $\|x_t - \hat{x}_{t,CSres}\|_{\infty} \leq C_3\hat{x}_{min,t}$ , for some  $C_3 < 1$ . Then,

$$\alpha_{del} = (1 - C_3)\hat{x}_{min,t} \quad (24)$$

3) **Algorithm Revisited:** As we described how to estimate the mean and variance or the addition and deletion threshold we can introduce our algorithm in this section. First we define a new notation for a  $m$ -length vector  $V$  which is all-zero but 1 in some indices. For a support set  $T$ , the vector is denoted as  $V_T$ . That means :

$$\begin{aligned} V_T &= 1 \\ V_{T^c} &= 0 \end{aligned}$$

Let at time  $t = t_0$ , we apply CS on the measurement  $y_{t_0}$  and get the output  $\hat{x}_{t_0,CSres}$ . Apply  $\alpha_{t_0,add}$  to estimate the support  $\hat{N}_{t_0}$ . Get the mean and variance of  $(\hat{x}_{t_0,CSres})_{\hat{N}_{t_0}}$  which are denoted as  $\mu_{t_0,x}$  and  $\sigma_{t_0,x}^2$  respectively from the equ.(10). Now update the  $P_{t_0-1}$  matrix as  $P_{t_0-1} = \sigma_{t_0,x}^2 I_{\hat{N}_{t_0}}$ . Leave the  $Q_{t_0}$  matrix as all-zero one. But for  $\hat{x}_{t_0-1}$  now it would be  $(\hat{x}_{t_0-1})_{\hat{N}_{t_0}} = \mu_{t_0,x}$ . So the equations will be

$$\begin{aligned} P_{t-1} &= \sigma_{t_0,x}^2 I_{\hat{N}_{t_0}} \\ Q_t &= 0 \\ \hat{x}_{t-1} &= \mu_{t_0,x} V_{\hat{N}_{t_0}} \\ \hat{x}_{filter} &= \hat{x}_{t-1} \\ P_{t|t-1} &= P_{t-1} + Q_t. \end{aligned} \quad (25)$$

After that we will use equation (9) to get  $\hat{x}_{t_0}$  and  $P_{t_0}$ . Use equ.(11) to get  $\mu_{t_0+1,x}$  and  $\sigma_{t_0+1,x}^2$ . At time  $t > t_0$ , if  $t = t_0 + 1$ , apply Mod-CS on  $y_t$  to get a output  $\hat{x}_{t,CSres}$ .  $\hat{v}_t = \hat{x}_{t,CSres} - \hat{x}_{t-1}$ . There is no initial kalman filter operation at  $t = t_0 + 1$ , and hence no residual

mod-cs. Use equ.(14) to get the variance of  $\hat{v}_{t_0+1}$  and use that variance for updating  $Q_t$  matrix.

At time  $t > t_0 + 1$ , we use the residual mod-cs technique to improve our result. For that we have to run initial kalman filter on the support size  $T = \hat{N}_{t-1}$ . The kalman filter output is used to evaluate  $\tilde{y}_{t,res}$  and then we apply residual-ModCS on  $\tilde{y}_{t,res}$ . The output of this operation is  $\hat{x}_{t,CSres}$  same as for other Kf-ModCS algorithm. On  $\hat{x}_{t,CSres}$  we apply the  $\alpha_{t,add}$  to find new addition.

Now we will show how  $P_t$ ,  $Q_t$ ,  $\hat{x}_t$  are being updated. Let  $\Delta_A$  is the set of all indices added at time  $t$ . If  $\Delta_A$  is an empty set then we will directly apply the normal kalman filter without addition or deletion with proper  $P_{t|t-1}$ ,  $\hat{x}_{t-1}$ . The updates are described in the below:

$$\begin{aligned} Q_t &= \sigma_{t,\nu}^2 I_T \\ P_{t|t-1} &= P_{t-1} + Q_t \\ \hat{x}_{filter} &= \hat{x}_{t-1} \end{aligned} \quad (26)$$

Then use Equations. (9) to get  $\hat{x}_t$ . The output vector  $\hat{x}_t$  of the operation is denoted as  $\hat{x}_{inter}$ . On the other hand if  $\Delta_A$  is non-empty the updates of  $P_{t|t-1}$  and  $Q_t$  are

$$\begin{aligned} Q_t &= \sigma_{t,\nu}^2 I_T \\ P_{t|t-1} &= P_{t-1} + Q_t \\ (P_{t|t-1})_{\Delta_A, \Delta_A} &= \sigma_{t,x}^2 \\ \hat{x}_{filter} &= \hat{x}_{t-1} \\ (\hat{x}_{filter})_{\Delta_A} &= \mu_{t,x} \end{aligned} \quad (27)$$

Apply Equations (9) with these  $P_{t|t-1}$ ,  $\hat{x}_{filter}$  to get the output vector  $\hat{x}_{inter}$ .

For the deletion case again we will do the same as we do for other Kf-ModCS algorithm. We apply the deletion threshold  $\alpha_{t,del}$  on  $\hat{x}_{inter}$  to find  $\Delta_D$ . Then start with  $P_{t|t-1}$  and  $\hat{x}_{filter}$  and properly update them with  $\Delta_D$ . The update equations are

$$\begin{aligned} (P_{t|t-1})_{\Delta_D, [1,m]} &= 0 \\ (P_{t|t-1})_{[1,m], \Delta_D} &= 0 \\ (\hat{x}_{filter})_{\Delta_D} &= 0 \end{aligned} \quad (28)$$

Apply Equations. (9) to get the  $\hat{x}_t$ . So in compressed form the algorithm would be :

**Kf-ModCS Algorithm 3:**

- Initialization at time  $t = t_0$ 
  - Apply CS on  $y_{t_0}$  using equation (5) to get  $\hat{x}_{t_0,CSres}$
  - Use  $\alpha_{add}$ —Method 1 or Method 2 to get  $\alpha_{t_0,add}$ .
  - Use  $\alpha_{t_0,add}$  to get Support  $= \hat{N}_{t_0}$ .
  - Calculate  $\mu_{t_0,x}$  and  $\sigma_{t_0,x}^2$  using equ.(10)
  - Use Equation. (25) to get  $\hat{x}_{filter}$ ,  $P_{t|t-1}$
  - Use Equations. (9) to get  $\hat{x}_{t_0}$ ,  $P_t$ .
  - Calculate  $\mu_{t_0+1,x}$  and  $\sigma_{t_0+1,x}^2$  using equ.(13)
- Decoding at time  $t > t_0$ 
  - for  $t = t_0 + 1$  : end
    - 1)  $T = \hat{N}_{t-1}$
    - 2) If  $t = t_0 + 1$ 
      - Apply ModCS on  $y_t$  using  $T$  to get  $\hat{x}_{t,CSres}$



- $\hat{\nu}_t = \hat{x}_{t,CSres} - \hat{x}_{t-1}$
- Use equ.(14) to get  $\sigma_{t,\nu}^2$
- 3) else
  - $Q_{tmp} = \sigma_{t,\nu}^2 I_T$ ,  $\hat{x}_{tmp} = \hat{x}_{t-1}$
  - $(P_{tmp})_{t|t-1} = P_{t-1} + Q_{tmp}$
  - Use Equations. (9) to get  $\hat{x}_{t,init}$ .
  - $\tilde{y}_{t,res} = y_t - A\hat{x}_{t,init}$ .
  - Apply ModCS on  $\tilde{y}_{t,res}$  using  $T$  to get  $\hat{\beta}_t$
  - $\hat{x}_{t,CSres} = \hat{x}_{t,init} + \hat{\beta}_t$
- 4) Use  $\alpha_{add}$ —Method 1 or Method 2 to get  $\alpha_{t,add}$ .
- 5) apply  $\alpha_{t,add}$  on  $\hat{x}_{t,CSres}$  to measure  $\Delta_A$
- 6) If  $\Delta_A$  is empty
  - $T_{det} = T$ .
  - Use Equation. (26) to get  $\hat{x}_{filter}$ ,  $P_{t|t-1}$
  - Use Equations. (9) to get  $\hat{x}_{inter}$ ,  $P_t$
- 7) If  $\Delta_A$  is non-empty
  - $T_{det} = T \cup \Delta_A$ .
  - Use Equation. (27) to get  $\hat{x}_{filter}$ ,  $P_{t|t-1}$
  - Use Equations. (9) to get  $\hat{x}_{inter}$ ,  $P_t$
- 8) Use  $\alpha_{del}$ —Method 1 or Method 2 to get  $\alpha_{t,del}$ .
- 9) Apply  $\alpha_{t,del}$  on  $\hat{x}_{inter}$  to find  $\Delta_D$  from the support set  $T_{det}$
- 10) If  $\Delta_D$  is empty
  - $T_{new} = T_{det}$
  - $\hat{x}_t = \hat{x}_{inter}$
- 11) If  $\Delta_D$  is non-empty
  - $T_{new} = T_{det} \setminus \Delta_D$ ,
  - Use Equation. (28) to get  $\hat{x}_{filter}$ ,  $P_{t|t-1}$
  - Use Equations. (9) to get  $\hat{x}_t$ ,  $P_t$
- 12)  $\hat{N}_t = T_{new}$
- 13) Use equ.(13) and equ.(17) to get  $\mu_{t+1,x}$ ,  $\sigma_{t+1,x}^2$ ,  $\sigma_{t+1,\nu}^2$  for the next time instant

### III. KF-CS ERROR STABILITY

#### A. Kalman Filter Compressive Sensing

We will discuss the stability result for KF-CS algorithm. If we can prove the stability result for KF-CS that will also be valid for KF-ModCS algorithm. So first we will present the simplest KF-CS algorithm with Signal Model 1. The algorithm will be exactly similar as Kf-ModCS 1 with the exception of CS instead of ModCS.

##### Kf-CS Algorithm 1:

- 1) Initialization at time  $t = t_0$ 
  - Apply CS on  $y_{t_0}$  using equation (5) to get  $\hat{x}_{t_0,CSres}$
  - Use  $\alpha_{t_0,add}$  to get Support =  $\hat{N}_{t_0}$ .
  - $P_{t-1} = \sigma_{sys,0}^2 I_{\hat{N}_{t_0}}$ ,  $Q_t = 0$ ,  $\hat{x}_{t-1} = 0$ .
  - $P_{t|t-1} = P_{t-1} + Q_t$ .
  - Use Equations. (4) to get  $\hat{x}$ ,  $P_t$ .
- 2) Decoding at time  $t > t_0$ 

for  $t = t_0 + 1$  : end

  - $T = \hat{N}_{t-1}$
  - $Q_{tmp} = \sigma_{sys}^2 I_T$ ,  $(P_{tmp})_{t|t-1} = P_{t-1} + Q_{tmp}$
  - Use Equations. (4) to get  $\hat{x}_{t,init}$ .
  - $\tilde{y}_{t,res} = y_t - A\hat{x}_{t,init}$ .

- Apply CS on  $\tilde{y}_{t,res}$  using equ.(5) to get the output  $\hat{\beta}_t$
- $\hat{x}_{t,CSres} = \hat{x}_{t,init} + \hat{\beta}_t$
- apply  $\alpha_{t,add}$  on  $\hat{x}_{t,CSres}$  to measure  $\Delta_A$
- If  $\Delta_A$  is empty
  - $T_{det} = T$ .
  - $Q_t = \sigma_{sys}^2 I_T$ .
  - $P_{t|t-1} = P_{t-1} + Q_t$ .
  - Use Equations. (4) to get  $\hat{x}_{inter}$ ,  $P_t$
- If  $\Delta_A$  is non-empty
  - $T_{det} = T \cup \Delta_A$ .
  - $Q_t = \sigma_{sys}^2 I_T$ ,  $(P_{t-1})_{\Delta_A, \Delta_A} = \sigma_{sys,0}^2$
  - $P_{t|t-1} = P_{t-1} + Q_t$ ,
  - Use Equations. (4) to get  $\hat{x}_{inter}$ ,  $P_t$
- Use  $\alpha_{t,del}$  on  $\hat{x}_{inter}$  to find  $\Delta_D$  from the support set
- If  $\Delta_D$  is empty
  - $T_{new} = T_{det}$
  - $\hat{x}_t = \hat{x}_{inter}$
- If  $\Delta_D$  is non-empty
  - $T_{new} = T_{det} \setminus \Delta_D$ ,
  - $(P_{t|t-1})_{\Delta_D, [1,m]} = 0$ ,  $(P_{t|t-1})_{[1,m], \Delta_D} = 0$
  - $(\hat{x}_{t-1})_{\Delta_D} = 0$ .
  - Use Equations. (4) to get  $\hat{x}_t$ ,  $P_t$
- $\hat{N}_t = T_{new}$ .

#### B. Kalman Filter Compressive Sensing with Add-Del LS method

Analyzing the KF-CS algorithm of the previous section, which includes the deletion step, is difficult using the approach that we outline below. Thus, in this section, we will first introduce a new algorithm with LS-initialization, which we will call KF-CS-LS. In this method after the support estimation using CS we can use the LS estimation process to get the new  $\hat{x}_t$ . The complete algorithm is :

##### Kf-CS Algorithm 2:

- 1) Initialization at time  $t = t_0$ 
  - Apply CS on  $y_{t_0}$  using equation (5) to get  $\hat{x}_{t_0,CSres}$
  - Use  $\alpha_{t_0,add}$  to get Support =  $\hat{N}_{t_0}$ .
  - $P_{t-1} = \sigma_{sys,0}^2 I_{\hat{N}_{t_0}}$ ,  $Q_t = 0$ ,  $\hat{x}_{t-1} = 0$ .
  - $P_{t|t-1} = P_{t-1} + Q_t$ .
  - Use Equations. (4) to get  $\hat{x}$ ,  $P_t$ .
- 2) Decoding at time  $t > t_0$ 

for  $t = t_0 + 1$  : end

  - $T = \hat{N}_{t-1}$
  - $Q_{tmp} = \sigma_{sys}^2 I_T$ ,  $(P_{tmp})_{t|t-1} = P_{t-1} + Q_{tmp}$
  - Use Equations. (4) to get  $\hat{x}_{t,init}$ .
  - $\tilde{y}_{t,res} = y_t - A\hat{x}_{t,init}$ .
  - Apply CS on  $\tilde{y}_{t,res}$  using equ.(5) to get the output  $\hat{\beta}_t$
  - $\hat{x}_{t,CSres} = \hat{x}_{t,init} + \hat{\beta}_t$
  - apply  $\alpha_{t,add}$  on  $\hat{x}_{t,CSres}$  to measure  $\Delta_A$
  - If  $\Delta_A$  is empty
    - $T_{det} = T$ .

- $Q_t = \sigma_{sys}^2 I_T$ .
- $P_{t|t-1} = P_{t-1} + Q_t$ .
- **Use Equations. (4) to get  $\hat{x}_{inter}$ ,  $P_t$**
- **If  $\Delta_A$  is non-empty**
  - $T_{det} = T \cup \Delta_A$ .
  - $(\hat{x}_{inter})_{T_{det}} = A_{T_{det}}^\dagger y_t$
  - $(\hat{x}_{inter})_{T_{det}^c} = 0$
  - $(P_t)_{T_{det}, T_{det}} = (A_{T_{det}}' A_{T_{det}})^{-1} \sigma_{obs}^2$
  - $(P_t)_{T_{det}^c, [1, m]} = 0$ ,  $(P_t)_{[1, m], T_{det}^c} = 0$
- **Use  $\alpha_{t, del}$  on  $\hat{x}_{inter}$  to find  $\Delta_D$  from the support set**
- **If  $\Delta_D$  is empty**
  - $T_{new} = T_{det}$
  - $\hat{x}_t = \hat{x}_{inter}$
- **If  $\Delta_D$  is non-empty**
  - $T_{new} = T_{det} \setminus \Delta_D$ ,
  - $(\hat{x}_t)_{T_{new}} = A_{T_{new}}^\dagger y_t$
  - $(\hat{x}_t)_{T_{new}^c} = 0$
  - $(P_t)_{T_{new}, T_{new}} = (A_{T_{new}}' A_{T_{new}})^{-1} \sigma_{obs}^2$
  - $(P_t)_{T_{new}^c, [1, m]} = 0$ ,  $(P_t)_{[1, m], T_{new}^c} = 0$
- $\hat{N}_t = T_{new}$ .

*Remark 1:* Notice that the LS step re-initializes the KF whenever the estimated support changes. This ensures less dependence of the current error on the past, and makes the stability analysis easier.

*Remark 2:* For ease of notation, in (??), we write the KF equations for the entire  $x_t$ . But the algorithm actually runs a reduced order KF for only  $(x_t)_T$  at time  $t$ , i.e. we actually have  $(\hat{x}_t)_{T^c} = 0$ ,  $(K_t)_{T^c, [1, n]} = 0$ ,  $(P_{t|t-1})_{[1, m], T^c} = 0$ ,  $(P_{t-1})_{[1, m], T^c} = 0$ ,  $(P_{t|t-1})_{T^c, [1, m]} = 0$ , and  $(P_{t-1})_{T^c, [1, m]} = 0$ . For computational speedup, the reduced order KF should be explicitly implemented.

*Remark 3:* The KF in KF-CS does not always run with correct model parameters. Thus, even when  $\sigma_{sys}^2/\sigma_{obs}^2$  is small, it is not clear if KF-CS will always outperform LS-CS [?]. This will hold at times when the support is accurately estimated and the KF has stabilized

Again while analysing the stability of the KF-CS-LS method we assume that there is no deletion to make the analysis simple. So here we will study the KF-CS without the deletion step, i.e. we set  $\alpha_{del} = 0$ . KF-CS without deletion assumes that there are few and bounded number of removals and false detects. For simplicity, in this work, we just assume  $S_r = 0$  in Signal Model 1 and we will select  $\alpha$  so that there are zero false detects.  $S_r = 0$  along with the assumption that the maximum sparsity size is  $S_{max}$  implies that there are only a finite number of addition times,  $K$ , i.e. for all  $t \geq t_{[add, K-1]}$ ,  $N_t = N_{t_{[add, K-1]}}$ . We summarize this in the following signal model.

**Signal Model 3:** Assume Signal Model 1 with  $S_r = 0$ . This implies that there are only a finite number of addition times,  $t_{add, j}$ ,  $j = 0, 1, \dots, (K-1)$  and  $K = \lceil \frac{S_{max} - S_0}{S_a} \rceil$ . Let  $t_K := \infty$ .

From now on in the stability discussion, we will drop the subscript  $\{add\}$  and only use integer in the subscript. In this section, we find sufficient conditions under which, with

high probability (w.h.p.), KF-CS for Signal Model 3 and observation model given by (2) gets to within a small error of the genie-KF for the same system, within a finite delay of the new addition time. Since the genie-KF error is itself stable w.h.p., as long as  $\delta_{S_{max}} < 1$ , this also means that the KF-CS reconstruction error is stable w.h.p.

Our approach involves two steps. Consider  $t \in [t_j, t_{j+1})$ . First, we find the conditions under which w.h.p. all elements of the current support,  $N_t = N_{t_j}$  get detected before the next addition time,  $t_{j+1}$ . Denote the detection delay by  $\tau_{det}$ . If this happens, then during  $[t_j + \tau_{det}, t_{j+1})$ , both KF-CS and genie-KF run the same fixed dimensional and fixed parameter KF, but with different initial conditions. Next, we show that if this interval is large enough, then, w.h.p, KF-CS will stabilize to within a small error of the genie-KF within a finite delay after  $t_j + \tau_{det}$ . Combining these two results gives our stability result.

We are able to do the second step because, whenever  $\hat{N}_t \neq \hat{N}_{t-1}$ , the final LS step re-initializes the KF with  $P_t$ ,  $\hat{x}_t$  given by (??). This ensures that the KF-CS estimate,  $\hat{x}_t$ , and the Kalman gain,  $K_t$ , at  $t+1$  and future times depend on the past observations only through  $T := \hat{N}_t$ . Thus, conditioned on the event  $\{\hat{N}_t = N_t, \forall t \in [t_j + \tau_{det}, t_{j+1})\}$ , there will be no dependence of either  $\hat{x}_t$  or of  $K_t$  on observations before  $t_j + \tau_{det}$ .

### C. The Stability Result

We begin by stating Lemma 1 which shows two things. First, if accurate initialization is assumed, the noise is bounded,  $S_{max} \leq S_{**}$ ,  $\alpha_{del} = 0$  and  $\alpha$  is high enough, there are no false detections. If the delay between addition times also satisfies  $d > \tau_{det}(\epsilon, S_a)$ , where  $\tau_{det}$  is what we call the “high probability detection delay”, then the following holds. If before  $t_j$ , the support was perfectly estimated, then w.p.  $\geq 1 - \epsilon$ , all the additions which occurred at  $t_j$  will get detected by  $t_j + \tau_{det}(\epsilon, S_a) < t_{j+1}$ .

*Lemma 1:* Assume that  $x_t$  follows Signal Model 3. If

- 1) (*initialization* ( $t = 0$ )) all elements of  $x_0$  get correctly detected and there are no false detects, i.e.  $\hat{N}_0 = N_0$ ,
- 2) (*measurements*)  $S_{max} \leq S_{**}$  and  $\|w\|_2 \leq \xi$ ,
- 3) (*algorithm*) we set  $\alpha_{del} = 0$  and  $\alpha^2 = B_* := (C_1 \xi)^2$ , where  $C_1$  is defined in [ ][Restricted Isometry Property]
- 4) (*signal model*) delay between addition times,  $d > \tau_{det}(\epsilon, S_a)$ ,

$$\text{where } \tau_{det}(\epsilon, S) := \left\lceil \frac{4B_*}{\sigma_{sys}^2 [\mathcal{Q}^{-1}(\frac{(1-\epsilon)^{1/S}}{2})]^2} \right\rceil - 1, \quad (29)$$

$\lceil \cdot \rceil$  denotes the greatest integer function and  $\mathcal{Q}(z) := \int_z^\infty (1/\sqrt{2\pi}) e^{-x^2/2} dx$  is the Gaussian Q-function,

then

- 1) at each  $t$ ,  $\hat{N}_t \subseteq N_t \subseteq N_{t+1}$  and so  $|\Delta_{e, t+1}| = 0$
- 2) at each  $t$ ,  $\|x_t - \hat{x}_{t, CSres}\|^2 \leq B_*$
- 3)  $Pr(E_j | F_j) \geq 1 - \epsilon$  where  $F_j := \{\hat{N}_t = N_t \text{ for } t = t_j - 1\}$  and  $E_j := \{\hat{N}_t = N_t, \forall t \in [t_j + \tau_{det}(\epsilon, S), t_{j+1} - 1]\}$ .

The proof is given in Appendix B. The initialization assumption is made only for simplicity. It can be easily satisfied by using  $n_0 > n$  to be large enough. Next we give another lemma, lemma. 2 which states that if the true support set does not change after a certain time,  $t_{nc}$ , and if it gets correctly detected by a certain time,  $t_* \geq t_{nc}$ , then KF-CS converges to the genie-KF in mean-square and hence also in probability.

*Lemma 2:* Assume that  $x_t$  follows Signal Model 3;  $\delta_{S_{\max}} < 1$ ; and  $\alpha_{del} = 0$ . Define the event  $D_f := \{\hat{N}_t = N_t = N_*, \forall t \in [t_*, t_{**}]\}$ . For a given  $\epsilon, \epsilon_{err}$ , there exists a  $\tau_{KF}(\epsilon, \epsilon_{err}, N_*)$  s.t. for all  $t \in [t_* + \tau_{KF}, t_{**}]$ ,  $Pr(\|\text{diff}_t\|^2 \leq \epsilon_{err} \mid D_f) > 1 - \epsilon$ . Clearly if  $t_{**} < t_* + \tau_{KF}$ , this is an empty interval.

The proof is similar to what we think should be a standard result for a KF with wrong initial conditions (here, KF-CS with  $t = t_*$  as the initial time) to converge to a KF with correct initial conditions (here, genie-KF) in mean square. A similar (actually stronger) result is proved for the continuous time KF in [?]. We could not find an appropriate citation for the discrete time KF and hence we just give our proof in Appendix C. After review, this can be significantly shortened. The proof involves two parts. First, we use the results from [?] and [?] to show that (a)  $P_{t|t-1}^\dagger, P_t^\dagger, K_t^\dagger$  and  $J_t := I - K_t^\dagger A_{N_*}$ , where  $P_{t|t-1}^\dagger = (P_{t|t-1})_{N_*, N_*}, P_t^\dagger = (P_t)_{N_*, N_*}, K_t^\dagger = (K_t)_{N_*, [1:m]}$ , converge to steady state values which are the same as those for the corresponding genie-KF; and (b) the steady state value of  $J_t$ , denoted  $J_*$ , has spectral radius less than 1 and because of this, there exists a matrix norm, denoted  $\|\cdot\|_\rho$ , s.t.  $\|J_*\|_\rho < 1$ . Second, we use (a) and (b) to show that the difference in the KF-CS and genie-KF estimates,  $\text{diff}_t$ , converges to zero in mean square, and hence also in probability (by Markov's inequality).

The stability result then follows by applying Lemma 2 for each addition time,  $t_j$ .

*Theorem 1 (KF-CS Stability):* Assume that  $x_t$  follows Signal Model 3. Let  $\text{diff}_t := \hat{x}_t - \hat{x}_{t, GAKF}$  where  $\hat{x}_{t, GAKF}$  is the genie-aided KF estimate and  $\hat{x}_t$  is the KF-CS estimate. For a given  $\epsilon, \epsilon_{err}$ , if the conditions of Lemma 1 hold, and if the delay between addition times,  $d > \tau_{det}(\epsilon, S_a) + \tau_{KF}(\epsilon, \epsilon_{err}, N_{t_j})$ , where  $\tau_{det}(\cdot, \cdot)$  is defined in (29) in Lemma 1 and  $\tau_{KF}(\cdot, \cdot, \cdot)$  in Lemma 2, then

- 1)  $Pr(\|\text{diff}_t\|^2 \leq \epsilon_{err}) > (1 - \epsilon)$ , for all  $t \in [t_j + \tau_{det}(\epsilon, S_a) + \tau_{KF}(\epsilon, \epsilon_{err}, N_{t_j}), t_{j+1} - 1]$ , for all  $j = 0, \dots, (K - 1)$ , and for some  $\epsilon > 0$ .
- 2)  $Pr(|\Delta| \leq S_a \text{ and } |\Delta_e| = 0, \forall t) \geq (1 - \epsilon)^K$  for some  $\epsilon > 0$ .
- 3)  $Pr(|\Delta| = 0 \text{ and } |\Delta_e| = 0, \forall t \in [t_j + \tau_{det}(\epsilon, S_a), t_{j+1} - 1], \forall j = 0, \dots, K - 1) \geq (1 - \epsilon)^K$  for some  $\epsilon > 0$ .

The proof is given in Appendix D. A direct corollary is that after  $t_{K-1}$  KF-CS will converge to the genie-KF in probability. This is because for  $t \geq t_{K-1}$ ,  $N_t$  remains constant ( $t_K = \infty$ ).

#### D. Discussion

Consider a  $t \in [t_j, t_{j+1})$ . Notice that  $\tau_{KF}$  depends on the current support,  $N_t = N_{t_j}$  while  $\tau_{det}$  depends only on the number of additions at  $t_j, S_a$ . Theorem 1 says that if  $n$  is large enough so that  $S_{\max} \leq S_{**}$ ;  $\alpha_{del} = 0$  (ensures no deletions);

$\alpha = \sqrt{B_*}$  (ensures no false detects); and if the time needed for the current KF to stabilize,  $\tau_{KF}(\epsilon, \epsilon_{err}, N_{t_j})$ , plus the high probability detection delay,  $\tau_{det}(\epsilon, S_a)$ , is smaller than  $d$ , then w.p.  $\geq (1 - \epsilon)$ , KF-CS will stabilize to within a small error,  $\epsilon_{err}$ , of the genie-KF before the next addition time,  $t_{j+1}$ . If the current  $\tau_{KF}$  is too large, this cannot be claimed. But as long as  $\tau_{det}(\epsilon, S_a) < d$ , the unknown support size,  $|\Delta|$  remains bounded by  $S_a$ , w.p.  $\geq (1 - \epsilon)^K$ .

We give our result for the case of zero removals and zero false detects, but the same idea will extend even if  $|\Delta_e|$  is just bounded.

## IV. SIMULATION

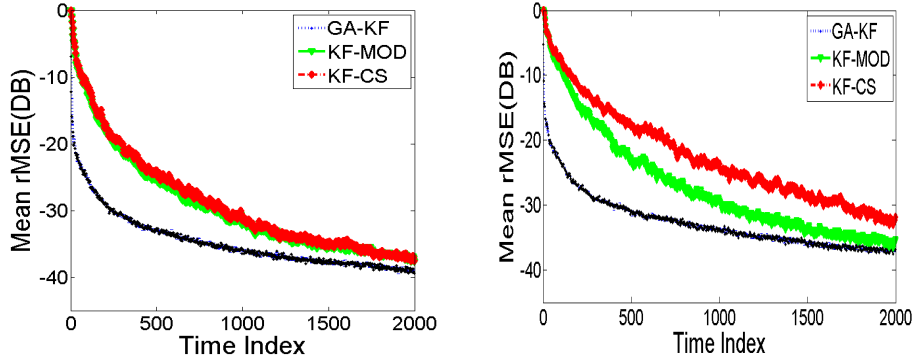
In this section we will discuss the simulation results to prove the stability result and then to demonstrate our algorithm with addition and deletion for simulated data and for real images. The primary performance parameter that we used in all of our experiments, is denoted as the normalized MSE, which is defined as

$$\text{NMSE} = \frac{\|\hat{x}_t - x_t\|^2}{\|x_t\|^2}$$

Here we are getting NMSE for each value of time instant  $t$ .

#### A. Stability Result Simulation

The first set of simulation deals with the stability discussion for the KF-CS algorithm. We will use the Signal Model 3 to generate the input signal. In our simulation the following parameters we have used.  $S_0 = 10$  denotes the initial support size. As there is no deletion, we will consider only addition of size  $S_a = 2$  at a regular interval of  $D = 20$ . Maximum support size is  $S_{\max} = 20$ , which signifies if the support size reaches  $S_{\max}$  there will be no more addition. Every new index follows a Gaussian distribution with mean  $\eta_0 = 0$  and variance  $\sigma_{sys,0}^2 = 1$ . On the hand the currently non-zero indices follows an independent Gaussian random walk model with mean 0 and variance  $\sigma_{sys}^2 = 1$ . We create the entries of our measurement matrix  $A$  from a Gaussian distribution of mean 0 and variance  $\frac{1}{n}$ , where  $n$  is the number of measurement. Then we normalize each of the column of  $A$ . We set the length of signal  $m = 256$ . In the stability analysis we have to set only one threshold, i.e. addition threshold. The addition threshold will be set to a high value, at least same as  $\sqrt{B_*}$  [ $B_*$  is defined in Lemma 1] so that there is no extra in the support estimate. We create the measurement noise from a uniform distribution of mean 0 and variance  $\sigma_{obs}^2$ . During the simulation we actually set a SNR value and from there we calculate the  $\sigma_{obs}^2$ . The SNR value is set to be 30DB. We notice that in fig.1(a) both KF-CS and KF-ModCS converges to GA-KF when the measurement number is  $n = 70$ . On the other hand we notice, in fig.1(b), that KF-ModCS performs better than KF-CS when the measurement number reduces to  $n = 45$ .



(a) Signal Model 3 with  $S_0 = 10$ ,  $S_a = 2$ ,  $d = 20$ ,  $S_{\max} = 20$ ,  $n = 70$  (b) Signal Model 3 with  $S_0 = 10$ ,  $S_a = 2$ ,  $d = 20$ ,  $S_{\max} = 20$ ,  $n = 45$

Fig. 1. Comparison of Mean relative MSE or NMSE of GA-KF, KF-CS, KF-ModCS for Signal Model 3.

### B. Practical Case

In this section we simulate our algorithm for the Signal Model 2. Unless otherwise noted, the following parameter settings are used for all the experiments. The value of the initial support size is set to be  $S_0 = 20$ . The transition probabilities are  $p_{01} = 0.1$  and  $p_{10} = 0.0086$ , which maintain the number of deletion and addition to roughly equal to  $0.1 * 20 = 2$  and  $0.0086 * 236 = 2.02 \approx 2$  respectively at each time. Input signal is a  $m = 256$  length signal which is  $S_0$  sparse at time  $t = t_0$ . The new indices of  $x_t$  follow a Gaussian Distribution of mean 0 and variance  $\sigma_{sys,0}^2 = 1$  at time  $t = t_0$ . The non-zero indices at time  $t > t_0$  will follow  $(x_t) = (1 - \lambda_{cor})x_{t-1} + \beta_{cor}w_t$ .  $w_t$  is also a Gaussian random variable with mean 0 and variance  $\sigma_{sys}^2$ . We set  $\lambda_{cor} = 0.05$ ,  $\beta_{cor} = 0.05$  and  $\sigma_{sys}^2 = 39$ . We will vary the number of measurement number,  $SNR$  value and sometimes  $\lambda_{cor}$  to find the performance of our Algorithm. And also we take the average over 100 realizations.

Another important objective is setting the `fixed value`. Whether we use the addition threshold method 1 or 2 we have to fix a value for that parameter to start the decoding. Once we get that value we can set other parameters. In the first simulation we set the  $SNR$  value to be  $25\text{DB}$  and the measurement number  $n = 70$ . We fix the value of `fixed value` to be 0.4,  $\vartheta_{low} = 0.3$ ,  $\vartheta_{high} = 0.8$  and  $C = 0.8$ . Then we run our decoding algorithms using addition threshold method 1 and 2 and for the following case when we don't have any knowledge about the signal model. Another simulation is also run in similar conditions except the measurement number has been changed. If  $n = 45$  we set the value of `fixed value` to be 0.3. Depending on the value of `fixed value` we change the values of other parameter. Now the next question arises about how we are changing the value of `fixed value` as we change the measurement number. Unfortunately we have not got any explicit relationship between that parameter and measurement number. But we observe that there is some relationship between the two, i.e more precisely between `fixed value` and  $S_0/n$ . For a fixed  $m = 256$ , we have tried to give the following heuristics to determine `fixed value`. As we don't know how much is  $S_0$  first we estimate the  $\hat{N}_0$  from  $\hat{x}_{CSres,0}$  using 99.9% energy. Then we calculate

$|\hat{N}_0|/n$  and use the following chart :

- If  $|\hat{N}_0|/n > 0.27$  then `fixed value`= 0.3
- If  $0.27 \geq |\hat{N}_0|/n > 0.25$  then `fixed value`= 0.4
- If  $0.25 \geq |\hat{N}_0|/n > 0.22$  then `fixed value`= 0.5
- If  $0.22 \geq |\hat{N}_0|/n > 0.2$  then `fixed value`= 0.6
- If  $|\hat{N}_0|/n \leq 0.2$  then `fixed value`= 0.7

[Note : I am still trying to find the best possible range] After getting these values we can set the values for the other parameters in the following way.

$$\begin{aligned}\vartheta_{low} &= \text{fixed value} - 0.15 \\ \vartheta_{high} &= \vartheta_{low} + 0.5 \\ C &= 1.2 - \text{fixed value}\end{aligned}$$

We see in both figures 2(a) and 2(b) addition threshold method 2 gives better result than the addition threshold method 1. So from now onwards we will only use addition threshold method 2 when we will talk about the *KF-ModCS* algorithm. Another important thing is that in this algorithm we don't have any information about the signal model except the measurement noise variance. So now we will see how algorithm work in low  $SNR$  environment. For that purpose we will decrease the value of  $p_{01}$  to 0.05 and  $p_{10}$  to 0.0043 and also decrease the value of  $S_0$  to 15. Measurement number is here  $n = 60$  which is less than 25% of the full measurement. As we don't have the whole idea of how the parameter `fixed value` should be tuned in accordance with  $SNR$  we rely on only  $n$  and  $S_0$  to set its value. Here we will show the results for two different values of  $SNR$ .

### C. Real Image

We tested the performance of our algorithm on a real Image also. The experiment consists of 20 frame of a Cardiac Image. Each of the image is of size  $64 \times 64$ . We use a mask and the 2-D fourier matrix as our measurement matrix. To elaborate more on that let  $X$  be the 2-D image. And the 2-D fourier transform of that image is defined as  $Y = F_r X F_r$ , where  $F_r$  is the fourier matrix. Now the 2-D image is compressible in the 2-D, 2-level *Daubechies - 4* wavelet domain. Let  $W$  is the wavelet matrix, then  $X_W = W X W'$

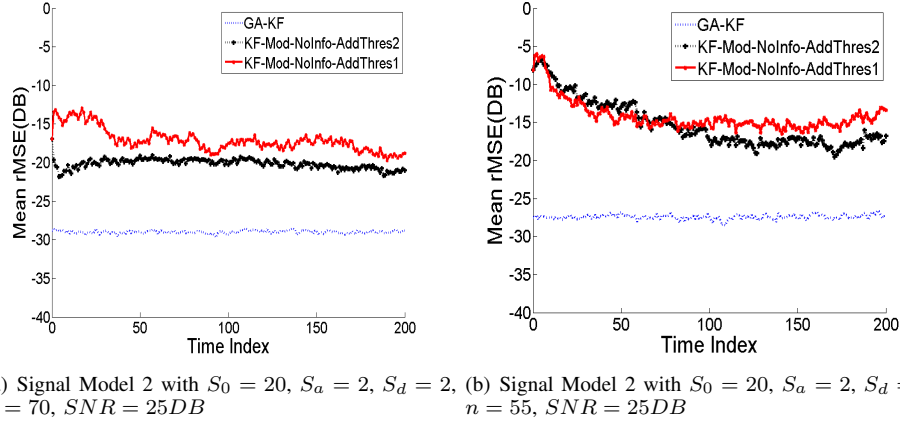


Fig. 2. Comparison of Mean relative MSE or NMSE at different measurement for Signal Model 2.

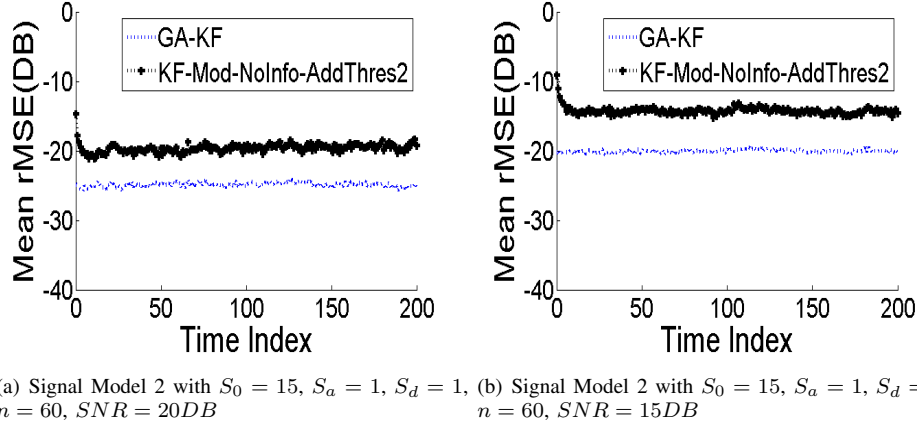


Fig. 3. Comparison of Mean relative MSE or NMSE at different SNR for Signal Model 2.

or  $X = W'X_WW$ , because  $W$  is unitary matrix. Then we get  $F_r X F_R = F_r W' X_W W F_r$ . Now we want to treat the 2-D image  $X$  as a row-ordered vector  $X_v$  and also its Fourier Transform  $Y$  as a row-ordered vector  $Y_v$ . Which implies that  $Y_v = F_r \otimes F_r X_v$  where  $\otimes$  represents the kronecker product. Again as  $W$  is real matrix,  $X_v = W' \otimes W' X_{Wv}$  where  $X_{Wv}$  is the row-ordered vector of the matrix  $X_W$ . So  $Y_v$  can be written as  $Y_v = (F_r \otimes F_r)(W' \otimes W')X_{Wv}$ . The preceding equation gives us the full sampled measurement of a sparse vector  $X_{Wv}$ . To make it a partial sampled measurement we can use a logical mask, which is randomly generated. The logical mask will indicates which of the indices of  $Y_v$  should be sampled. By adding random Gaussian noise the measured vector would be of the following form :

$$Y_{v,s} = A_m X_{Wv} + \nu$$

where  $A_m$  is the collection of rows of  $(F_r \otimes F_r)(W' \otimes W')$  indexed by the logical mask.

Nest task is to set the value of fixed value. We can't use the same range of values of fixed value as we did for the Gaussian matrix. Because the measurement matrix here is  $(F \otimes F)(W' \otimes W')$  we have to find again the appropriate value of fixed value using 99.9% energy technique. We find for 30% measurement the value should be 0.05. Using that we set the values for  $\vartheta_{low}$ ,  $\vartheta_{high}$  and  $C$  as 0.03, 0.2 and 1 respectively. All the simulations with cardiac images use the above-mentioned values of the parameter. Another important

observation we find is that as the sparsity to length ratio for the cardiac image( in Wavelet domain) is very close to 10%, taking 20% measurement with Fourier matrix will not give us good recovery. So I used 30% measurement and also I used same measurement for all frames. Here also we take the average over 100 realizations.

#### D. Comparison With Other Contemporary Works

KF-ModCS algorithm works fine when the input signal varies in the following way :

$$(x_t)_i = (1 - \lambda_{cor})(x_{t-1})_i + \beta_{cor}(\nu_t)_i \text{ where } i \in N_t$$

For this kind of situation our algorithm does not need any prior information about the signal. But If the signal varies following a generalized matrix model like

$$(x_t)_{N_t} = F_{mat}(x_{t-1})_{N_{t-1}} + \nu_t$$

Then we need to know  $F_{mat}$  or at least some estimate of it to perform decoding. In our studies we actually ignore the later model and focuss our discussion only on the first model.

When we go through other algorithms which tackle similar kind of problem we notice that most of those algorithms do need some estimation about the signal variance or about the value of  $\lambda_{cor}$  or  $\beta_{cor}$ . Sometimes those algorithms need some exact estimation about the initial sparsity size or about the sparsity change rate. If the initialization is not perfect sometimes those algorithm fail to decode the signal. Here in the

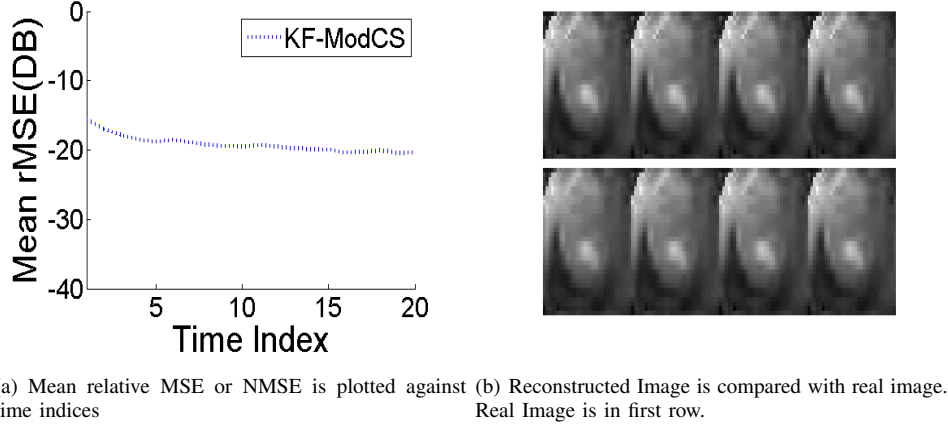


Fig. 4. Cardiac Image Recovery with 30% measurement with 20DB SNR

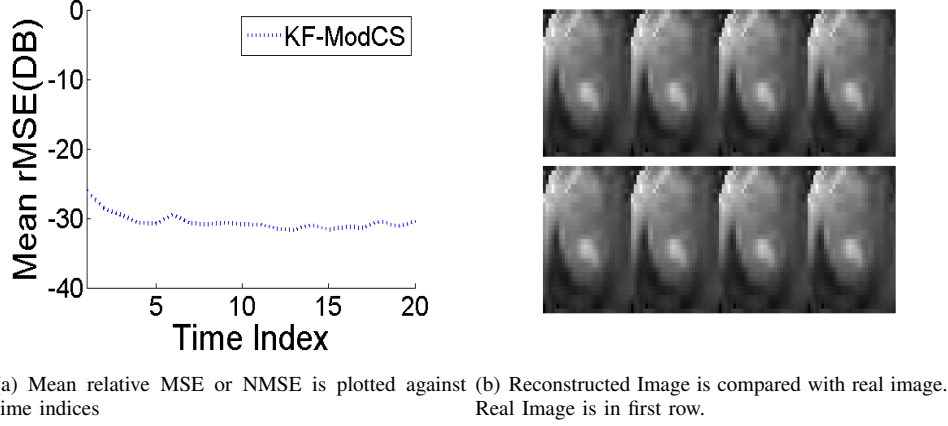


Fig. 5. Cardiac Image Recovery with 30% measurement with 35DB SNR

following discussion we will discuss about some contemporary algorithms and will compare our result with them. All of the comparisons will be done with their signal models and sometimes with our signal models.

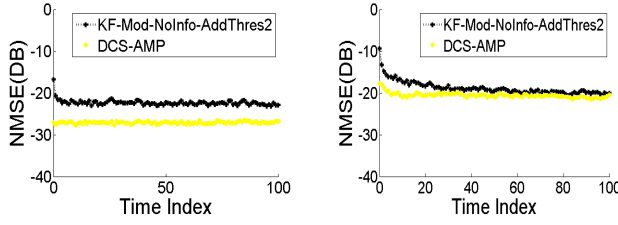
1) *Ohio State Group*: Justin Zeinel and others have introduced an algorithm called *DCS – AMP* or Dynamics Compressed Sensing with Approximate Message Passing Algorithm to reconstruct  $x_t$ . It is a powerful algorithm which works very efficiently when it has some knowledge about  $\lambda_{cor}$ ,  $\sigma_{sys,0}$ ,  $\sigma_{sys}$ . In their paper [ ] they have formulated a signal model which is very close to Signal Model 2. The only difference is that in our signal model the value of new indices are sampled from a Gaussian Distribution of mean 0 and variance  $\sigma_{sys,0}^2$  but in their signal model that is not the case. They proposed two different random processes. The first random process  $r_t$  follows the following temporal change

$$r_t = (1 - \lambda_{cor})r_{t-1} + \lambda_{cor}\nu_t$$

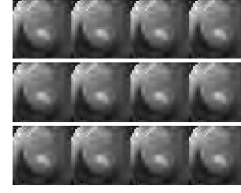
where each entry of  $\nu_t$  follows an i.i.d Gaussian Random distribution with mean 0 and variance  $\sigma_{sys}^2$  and  $r_0$  is a Gaussian Vector of i.i.d entries with mean  $\eta$  and variance  $\sigma_{sys,0}^2$ . On the other hand, another random process  $\theta_t$  is a vector with 0 and 1. The zero entry of  $\theta_t$  can change to 1 with probability  $p_{10}$  and the non-zero entry can change to 0 with probability  $p_{01}$ .  $\theta_0$  has  $S_0$  number of non-zero entries depending upon a probability value  $q$  where  $S_0$  is roughly equal to  $qm$ . They establish some relationships among the

parameters such as  $\sigma_{sys}^2 = \frac{(2-\lambda_{cor})\sigma_{sys,0}^2}{\lambda_{cor}}$  and  $p_{10} = \frac{qp_{01}}{(1-q)}$ . Here we will show two different comparisons. In the first comparison their algorithm is properly initialized with the exact values of different signal parameters  $\lambda_{cor}$ ,  $\sigma_{sys,0}^2$ ,  $\sigma_{sys}^2$ ,  $p_{01}$ ,  $p_{10}$ ,  $S_0$ . We use their signal model to generate the input data. Then we run DCS-AMP and KF-ModCS on the output and try to reconstruct the input signal. We observe that *DCS – AMP* performs really better than *KF – ModCS*. In the second comparison our Signal Model 2 is used to generate data. As we mentioned that our signal model is not very different from their. But we incorporate some extra features in our signal model. It can create input signal and also noise which are sampled from a uniform distribution. We add one extra parameter  $\beta_{cor}$  and  $\beta_{cor}$  is not necessarily equal to  $\alpha_{cor}$ . Similarly  $\sigma_{sys}$  can be totally independent of  $\sigma_{sys,0}$  which can be true for the pair  $p_{01}$  and  $p_{10}$ . So we generate a series of input signals which are uniform random vector and outputs using signal Model 2, where we keep the value of *SNR*, measurement same as the first scenario. We notice that for the later scenario DCS-AMP performs nearly as good as KF-ModCS. There is not much difference in the performances. We deliberately choose the uniform signal to show that our algorithm is totally independent of the nature of the input. Our claim gets stronger when we compare the Cardiac Data Recovery both with *DCS – AMP* and KF-ModCS. When we use simulated data we set the parameter as  $q = 0.04$ ,





(a) Signal follows the change model mentioned in paper [1] (b) Signal is generated using Signal Model 2



(c) Image comparison. From top- wise first row True Image, second Image row KF-ModCS, third Row DCS-AMP

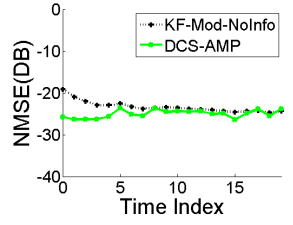


Fig. 6. Comparison with *DCS-AMP* algorithm

$p_{01} = 0.05$ ,  $\sigma_{sys,0} = 1$ ,  $\lambda_{cor} = 0.05$ ,  $SNR = 20DB$  for their signal model and for our signal model these parameters are same with some additional parameters are set to  $p_{10} = 0.003$ ,  $\beta_{cor} = 0.5$ ,  $\sigma_{sys} = 1$  and also the signal and noise are uniform random variables. On the other hand when we compared for real image we use cardiac data with 30% measurement and in 25DB SNR environment. In this simulation the DCS-AMP algorithm has been initialized with the nearly exactly values of parameters such as  $\sigma_{sys}^2$ ,  $\sigma_{sys,0}^2$ ,  $q$ ,  $p_{01}$ ,  $p_{10}$  using nearly all true support indices in the same way as the authors had done for Larynx Image. On the other hand in KF-ModCS algorithm any signal parameter needs not to be initialized. The algorithm automatically set the parameter fixed value using 99.9% energy. We observe that in fig.6(a) the *DCS-AMP* algorithm beats the *KF-ModCS* algorithm with NMSE difference of nearly 8 DB. But when we change the signal and noise to uniform random variable both the algorithm are very much comparable. The same conclusion can be drawn for the recovery of cardiac image. But the important difference is that *DCS-AMP* needs some information about the Cardiac Image but again the *KF-ModCS* algorithm needs no information about the Image.

2) *Georgia Tech Group*: In their paper [1] Salman Asif and others have tackled a similar problem as ours. Again the signal model they have used is

$$\begin{aligned} x_{t+1} &= Fx_t + \nu_t \\ y_{t+1} &= Ax_{t+1} + w_t \end{aligned}$$

where they have set  $F$  a permutation matrix. But for any permutation matrix our algorithm *KF-ModCS* will not work properly until and unless it has knowledge about  $F$ . So we use a special case when  $F$  is a diagonal matrix, not necessary an identity matrix. They have introduces three algorithms in the paper [1] and two algorithms in the paper [2] for similar problem. We run all the algorithms and our algorithms for same input data which follows their signal model. For the simulations we have set  $n = 256$ ,  $m = 70$ ,  $s_{num} = 20$ ,  $p = 2$ ,  $dyn_{var} = 0.01$ ,  $obs_{var} = 0.001$ . The average SNR is 22DB. And we take the average over 100 realizations. We observe in Fig 7 *Kf-ModCS* performs pretty well in that environment, better than any algorithm except *RWL1-DF*.

Then I tried to recover the cardiac image using these algorithms after trying to initialize some parameters in the way mentioned in the paper and their codes. But those algorithms

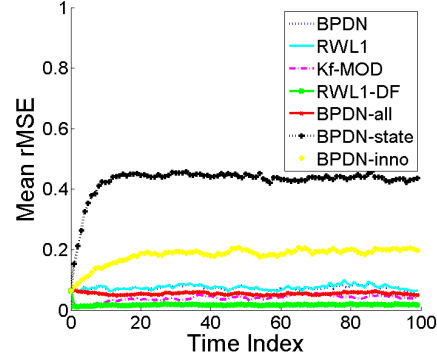


Fig. 7. Comparison with *Homotopy* algorithms for simulated data shows the following problems "Degenerate indices in previous iteration!" and gives erroneous outputs.

3) *Avishy's Group*: In their paper [3] Avishy and Others have introduced their algorithms, *CS-EmbeddedKalmanFiltering*, for the similar problem. Their algorithm also do some kalman-filtering steps embedded with CS. They have introduced two variations of that algorithm -one is for 1-norm and the other is for  $p$ -pseudo norm. Then they implemented their algorithm for static and dynamic cases. Here we will compare for both scenario with their input data for the following parameter,  $n = 256$ ,  $m = 72$ ,  $SNR = 25$ ,  $s_{num} = 10$ . In the static case they have created the input from a uniform distribution of the range  $[-10, 10]$ , and in the dynamic case the non-zero indices of the support set follow a random walk model with a distribution of mean 0 and variance 1. In their model they don't do any support addition or deletion. From the fig. 8(a) and fig. 8(b) we notice that *KF-ModCS* performs much better than *CSKF* algorithms.

Note : Avishy sent me three subroutines for their algorithms. Using those subroutines and algorithms presented in their paper I tried to simulate the static case and dynamic case scenario. But the results I got are bit different than in their paper. I asked Avishy for all the codes but he has not sent me those.

## APPENDIX

### A. Some Definitions and Useful Theorems

In this section we will discuss some useful definitions and theorems . But before we give those defintions let us

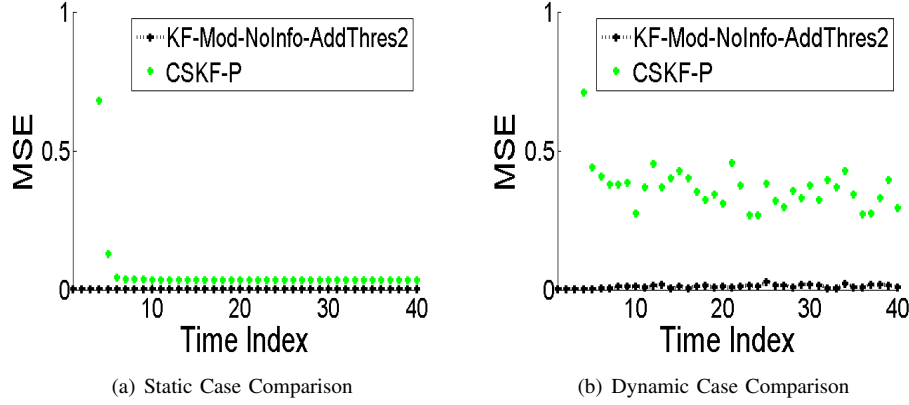


Fig. 8. Comparison with *CSKF* algorithm for Simulated Data

define a discrete-time linear time-invariant system. Typically a discrete-time linear time-invariant system is defined in the following way, which is basically a state-space model. If  $x_i, y_i, u_i$  denote the state, output and deterministic control input respectively at time  $i$  then

$$\begin{aligned} x_{i+1} &= Fx_i + Gu_i \\ y_i &= Ax_i + Bu_i, i \geq i_0 \end{aligned} \quad (30)$$

where,  $x_{i_0}$  = initial condition at time  $i_0$ ,  $F, G, A, B$  have dimensions  $n \times n, n \times q, p \times n, p \times q$ , respectively. Likewise  $u_i$  is  $q \times 1$  and  $y_i$  is  $p \times 1$ . The  $n$ -dimensional vector  $x$  is called the state of the system.

**Controllability** : Roughly the concept of controllability denotes the ability to move a system around in its entire configuration space using only certain admissible inputs. So the controllable conditions are like when the input signal will be able to influence or control the evolution of each individual entry of the state vector and hence the name controllable for the pair  $\{F, G\}$  [?][Linear System Book, P.762]. The so-called controllability matrix  $C$  is defined as  $C = [G \ FG \ F^2G \dots F^{n-1}G]$ . If  $C$  matrix is of full rank then the system is stated as controllable.

**Observability** : Observability is basically the dual of the controllability. It is a measure for how well the internal states of a system can be inferred with the knowledge of its external outputs. Observability matrix is defined as  $\mathcal{O} = [A' \ F'A' \dots (F')^{n-1}A']'$ . The system is stated as observable when the  $\mathcal{O}$  matrix is of full rank [?][Linear System P.764]

**Stabilizability** : The eigenvalues of the matrix  $F$  are sometimes called the modes of the realization  $(F, G, A, B)$ . Let  $\lambda$  be an eigenvalue of the matrix  $F$ . In the discrete-time case, a stable mode is one that satisfies  $|\lambda| < 1$ . For the rest of the eigenvalues of  $F$  the modes are called unstable modes. Now for any pair  $\{F, G\}$ , we classify the modes of  $F$  as controllable or uncontrollable according to the following criterion.

Modes at which  $\text{rank}([\lambda I - FG]) < n$ , are said to be uncontrollable. Otherwise, they are said to be controllable. Now a pair  $\{F, G\}$  will be said to be stabilizable if all its unstable modes are controllable. That is, if all the modes at which the above rank condition is satisfied are stable modes. [?][Linear System, P.763]

**Detectability** :  $F, A$  will be detectable if, and only if, all the unstable modes will be observable at the output of the realization. [?][Linear System, p.764].

Now we will extend our noise-less model to a noisy-model and will define the controllability, observability criterion again for this case. Typically the noisy Kalman-filter model is defined in the following way

$$\begin{aligned} x_{i+1} &= Fx_i + Gu_i \\ y_i &= Ax_i + w_i, \end{aligned} \quad (31)$$

where  $u_i$  is called the state-noise and  $w_i$  is called the measurement noise.

And the variance of  $u$  is  $Q$ , variance of  $w$  is  $R$ .

We also define the covariance between  $u$  and  $w$  is  $S$ . Typically in the kalman filter stability discussion we often assume that  $S = 0$ .

Now again the controllability and observability criterions for the above kalman filter are as follows.

**Controllability** : Here the controllability matrix is again defined as  $C = [G \ FG \ F^2G \dots F^{n-1}G]$ . If  $C$  matrix is of full rank then the system is stated as controllable. [?][Kalman Filter and Extended Kalman Filter, Maria Isabel Ribero, P 24]

**Observability** : Observability matrix is defined as  $\mathcal{O} = [A' \ F'A' \dots (F')^{n-1}A']'$ . The system is stated as observable when the  $\mathcal{O}$  matrix is of full rank. [?][Kalman Filter and Extended Kalman Filter, Maria Isabel Ribero, P 24].

Now we will derive the algebraic Riccati equation for the kalman filter and will discuss the solution of that Riccati equation. We will derive the Riccati equation when  $S = 0$ . We observe that the update equations for the kalman filter are

**Prediction Steps** : [?][?][Kalman Filter and Extended Kalman Filter, Maria Isabel Ribero, P 21]

$$\begin{aligned} \hat{x}_{i+1|i} &= F\hat{x}_{i|i} \\ P_{i+1|i} &= FP_{i|i}F' + GQG' \end{aligned}$$

And the Filter Steps: [?][Kalman Filter and Extended Kalman Filter, Maria Isabel Ribero, P 21]

$$\begin{aligned} \hat{x}_{i|i} &= \hat{x}_{i|i-1} + K_i[y_i - A\hat{x}_{i|i-1}] \\ K_i &= P_{i|i-1}A'[AP_{i|i-1}A' + R]^{-1} \\ P_{i|i} &= [I - K_iA]P_{i|i-1} \end{aligned}$$



If we combine the above two steps then we get the following equation

$$P_{i+1|i} = FP_{i|i-1}F' + GQG' - FK_i[AP_{i|i-1}A' + R]K_i'F' \quad (32)$$

which is basically the discrete Riccati recursive equation.

Now we will see that if the system satisfies some conditions then it will converge to a positive definite solution from any initial condition. Sometimes we will use the word "DARE" to denote Discrete Algebraic Riccati Equation.

**Lemma 3:** (Existence to the solutions to the DARE) [, Kalman Filter and Extended Kalman Filter , Maria Isabel Ribero , P 24]: Consider the system dynamics mentioned above and make the following assumptions that

- The matrix  $Q > 0$
- The matrix  $R > 0$ .
- The pair  $\{F, G\}$  is controllable.
- The pair  $\{F, A\}$  is observable.

Under above conditions

- The prediction matrix  $P_{i+1|i}$  converges to a constant matrix  $\tilde{P}$  which is positive semi-definite.
- $\tilde{P}$  is the unique positive semi-definite solution of the discrete algebraic Riccati equation mentioned above.
- $\tilde{P}$  is independent of the initial condition given the initial covariance matrix is positive semi-definite.

**Theorem 2:** (Algebraic Riccati Equation) [[Linear Estimation P.783] : Consider the discrete-time Algebraic Riccati Equation

$$P = FPF' + GQG' - (FPA')(R + APA')^{-1}(FPA')'$$

Then the following two statements are equivalent.

- $\{F, A\}$  is detectable and  $\{F, GQ^{1/2}\}$  is controllable.
- The DARE has a stabilizing solution  $P$ , i.e., one for which the matrix  $F - K_p A$  is stable, where  $K_p = (FPA')(R + APA')^{-1}$

When all the eigenvalues of  $F - K_p A$  lie inside the closed unit disc then we denote  $F - K_p A$  as stable.

**Spectral Radius:** The spectral radius of a matrix  $A \in M_n$ , where  $M_n$  is the space of all  $n \times n$  matrix, is defined as the non-negative real number  $\rho(A) = \max|\lambda| : \lambda \in \sigma(A)$ , where  $\sigma(A)$  is the collection of all eigenvalues of  $A$ . This is just the radius of the smallest disc centered at the origin in the complex plane that includes all the eigen-values of  $A$ . [?][Horn and Jhonson, P.35]

**Lemma 4:** ([, Horn and Jhonson P.297, Lemma 5.6.10]) : Let  $A$  is a square matrix of size  $n \times n$  and  $\epsilon > 0$  be given. There is a matrix norm  $\|\cdot\|$  such that  $\rho(A) \leq \|A\| \leq \rho(A) + \epsilon$ .

## B. Proof of Lemma 1

With  $\|w\|_2 < \xi$ , from [Theorem 1.2, Restricted Isometry Property :Candes], if a signal is  $S$ -sparse and if  $S \leq S_{**}$ , then, the error after running the BPDN selector is bounded by  $B_*$ .

We will prove the first two claims of lemma 1 by induction method. Consider the base case, when  $t = 0$ . The first assumption says that at  $t = 0$ , all elements of  $x_0$  get correctly

detected and there is no false detect. So  $\hat{N}_0 = N_0$ . As in signal model there is no support deletion, only addition process occurs,  $N_0 \subseteq N_1$ , so  $\hat{N}_0 \subseteq N_1$  and  $|\Delta_{e,t+1}| = 0$ . From the claim 2,  $S_{max} \leq S_{**}$ ,  $\|w\|_2 \leq \xi$  and from claim 3,  $\alpha_{del} = 0$ ,  $\alpha^2 = B_*$ . So from [Theorem 1.2, Restricted Isometry Property, Candes]  $\|x_t - \hat{x}_{t,CSres}\|^2 \leq B_*$ . So the first two claims are proved for  $t = 0$ .

Now suppose the first two claims are proved for  $t = t - 1$ . Using the first claim for  $t - 1$ ,  $|\Delta_{e,t}| = 0$ . Thus  $\beta_t$  is  $|N_t \cup \Delta_{e,t}| = |N_t|$  sparse. Since  $|N_t| \leq S_{max}$  and condition 2 holds, we can apply theorem [Theorem 1.2, Restricted Isometry Property :Candes] to get  $\|\beta_t - \hat{\beta}_t\|^2 \leq B_*$ . But  $x_t - \hat{x}_{t,CSres} = \beta_t - \hat{\beta}_t$  and so the second claim follows for  $t$ . By setting  $\alpha = \sqrt{B_*}$  (condition 3), we ensure that for any index  $i$  with  $(x_t)_i = 0$ ,  $(\hat{x}_{t,CSres})_i^2 = ((x_t)_i - (\hat{x}_{t,CSres})_i)^2 \leq \|x_t - \hat{x}_{t,CSres}\|^2 \leq B_* = \alpha^2$  (no false detects). Using this and  $S_r = 0$ , the first claim follows for  $t$ . For the third claim, it is easy to see that for any  $i \in \Delta$ , if, at  $t$ ,  $(\hat{x}_t)_i^2 > \alpha$  then  $i$  will definitely get detected. Now  $(x_t)_i^2 = ((x_t)_i - (\hat{x}_t)_i)^2 + (\hat{x}_t)_i^2 + 2((x_t)_i - (\hat{x}_t)_i)(\hat{x}_t)_i$ . So if  $(x_t)_i^2 > 2\alpha^2 + 2B_* = 4B_*$ , then  $i$  will get detected at  $t$ . Consider a  $t \in [t_j, t_{j+1} - 1]$ . Since  $F_j$  holds, so at  $t = t_j$ ,  $\Delta = \mathcal{A}(j)$ . Also, since  $\alpha_{del} = 0$ , there cannot be false deletions and thus for any  $t \in [t_j, t_{j+1} - 1]$ ,  $|\Delta| \leq S_a$ . Consider the worst case: no coefficient has got detected until  $t$ , i.e.  $\Delta_t = \mathcal{A}(j)$  and so  $|\Delta_t| = S_a$ . All  $i \in \mathcal{A}(j)$  will definitely get detected at  $t$  if  $(x_t)_i^2 > 4B_*$  for all  $i \in \mathcal{A}(j)$ . From our model, the different coefficients are independent, and for any  $i \in \mathcal{A}(j)$ ,  $(x_t)_i \sim \mathcal{N}(0, (t - t_j)\sigma_{sys}^2)$ . Thus,

$$\begin{aligned} &Pr((x_t)_i^2 > 4B_*, \forall i \in \mathcal{A}(j) \mid F_j) \\ &= \left(2Q\left(\sqrt{\frac{4B_*}{(t - t_j)\sigma_{sys}^2}}\right)\right)^{S_a} \end{aligned} \quad (33)$$

Using the first claim,  $Pr(\hat{N}_t = N_t \mid F_j)$  is equal to this. Thus for  $t = t_j + \tau_{det}(\epsilon, S_a)$ ,  $Pr(\hat{N}_t = N_t \mid F_j) \geq 1 - \epsilon$ . Since there are no false detects; no deletions and no new additions until  $t_{j+1}$ ,  $\hat{N}_t = N_t$  for  $t = t_j + \tau_{det}$  implies that  $E_j$  occurs. This proves the third claim.

## C. Proof of Lemma 2

Let  $\hat{x}_{t,GAKF}$  denote the genie-aided KF (GA-KF) estimate at  $t$ .

Assume that the event  $D_f$  occurs. Then, for  $t \in [t_*, t_{**}]$ ,  $\hat{N}_t = N_t = N_*$ , i.e.  $\Delta_t := N_t \setminus \hat{N}_{t-1} = N_* \setminus N_* = \phi$  (empty set) and so  $\hat{x}_t = \hat{x}_{t,init}$ . Let  $e_t \triangleq x_t - \hat{x}_t$  and  $\tilde{e}_t \triangleq x_t - \hat{x}_{t,GAKF}$ .

For simplicity of notation we assume in this proof that all variables and parameters are only along  $N_*$ , i.e. we let  $\hat{x}_t \equiv (\hat{x}_t)_{N_*}$ ,  $e_t \equiv (e_t)_{N_*}$ ,  $\nu_t \equiv (\nu_t)_{N_*}$ ,  $P_{t|t-1} \equiv (P_{t|t-1})_{N_*, N_*}$ ,  $K_t \equiv (K_t)_{N_*, [1:n]}$ . Let  $J_t \triangleq I - K_t A_{N_*}$ . Similarly for  $\hat{x}_{t,GAKF}$ ,  $\tilde{e}_t$ ,  $\tilde{P}_{t|t-1}$ ,  $\tilde{K}_t$ ,  $\tilde{J}_t$ . Here  $\tilde{P}_{t|t-1}$ ,  $\tilde{K}_t$ ,  $\tilde{J}_t$  are the corresponding matrices for GA-KF.

From (??), for  $t \in [t_*, t_{**}]$ ,  $e_t$ ,  $\tilde{e}_t$  and  $\text{diff}_t = e_t - \tilde{e}_t$  satisfy

$$\begin{aligned} e_t &= J_t e_{t-1} + J_t \nu_t - K_t w_t \\ \tilde{e}_t &= \tilde{J}_t \tilde{e}_{t-1} + \tilde{J}_t \nu_t - \tilde{K}_t w_t \\ \text{diff}_t &= J_t \text{diff}_{t-1} + (J_t - \tilde{J}_t)(\tilde{e}_{t-1} + \nu_t) + (\tilde{K}_t - K_t)w_t \end{aligned} \quad (34)$$

Now we will model our system with the notation used to introduce our problem. The noisy state space model for the Kalman Filter is

$$\begin{aligned} x_t &= x_{t-1} + \nu_t \\ y_t &= Ax_t + w_t \end{aligned} \quad (35)$$

where  $F \equiv I, G \equiv I, E[\nu_t \nu_t^*] = Q, E[w_t w_t^*] = R$ . Here  $F, G$  are used in appendix A. The state-noise in appendix A is denoted as  $\nu$  to be consistent with our problem formulation. For  $t > t_*$  both KF-CS and GA-KF run the same fixed dimensional and fixed parameter KF for  $(x_t)_{N_*}$  with parameters  $F \equiv I, Q \equiv (\sigma_{sys}^2 I_{N_*})_{N_*, N_*}, A \equiv A_{N_*}, R \equiv \sigma_{obs}^2 I$ , but with different initial conditions. KF-CS uses  $\hat{x}_{t_*}, P_{t_*+1|t_*} \neq \mathbb{E}[e_{t_*+1} e'_{t_*+1} | y_1 \dots y_{t_*}]$  while GA-KF uses the correct initial conditions,  $\hat{x}_{t_*, GAKF}, \tilde{P}_{t_*+1|t_*} = \mathbb{E}[\tilde{e}_{t_*+1} \tilde{e}'_{t_*+1} | y_1, \dots, y_{t_*}]$ . Since  $|N_*| \leq S_{\max}$  and  $\delta_{S_{\max}} < 1, A \equiv A_{N_*}$  is full rank. We can rewrite the equ. 35 in the following form

$$\begin{aligned} x_t &= x_{t-1} + Q^{1/2} \eta_t \\ y_t &= A_{N_*} x_t + w_t \end{aligned} \quad (36)$$

where  $\eta_t$  is the admissible Gaussian input of unit variance, and  $w_t$  is the noise of variance  $R$ . Here we define  $G = Q^{1/2}$ .  $G$  is again from appendix A. Before we discuss about the solutions of the Discrete Algebraic Riccati equation we will show how the Riccati equation comes into the picture for our particular Kalman Filter. We observed that

$$\begin{aligned} P_{t+1|t} &= P_t + Q \\ &= (I - K_t A_{N_*}) P_{t|t-1} + Q \\ &= P_{t|t-1} + Q - K_t A_{N_*} P_{t|t-1} \\ &= P_{t|t-1} + Q - P_{t|t-1} A'_{N_*} (A_{N_*} P_{t|t-1} A'_{N_*} + R)^{-1} A_{N_*} P_{t|t-1} \end{aligned} \quad (37)$$

So that is how we got our Riccati equation. Now the observability matrix is  $[A_{N_*} \ A_{N_*} I \ A_{N_*} I^2 \ \dots \ A_{N_*} I^{n_*-1}]'$ , where  $n_*$  is the dimension of  $N_*$ . As  $A_{N_*}$  is a matrix of full rank, so our observability matrix must have column rank and so row rank  $n_*$ . Thus  $(I, A_{N_*})$  is observable. Similarly our controllability matrix  $[Q^{1/2} \ I Q^{1/2} \ \dots I^{n_*-1} Q^{1/2}]$  is also of full rank as  $Q^{1/2}$  matrix is full rank. So  $(I, Q^{1/2})$  is controllable. Thus, according to Lemma 3, starting from any initial condition,  $P_{t+1|t}$  will converge to a positive semi-definite,  $P_*$ , which is the unique solution of the discrete algebraic Riccati equation

$$P_{t+1|t} = P_{t|t-1} + Q - P_{t-1|t} A'_{N_*} [A_{N_*} P_{t|t-1} A'_{N_*} + R]^{-1} A_{N_*} P_{t|t-1} \quad (38)$$

Consequently  $K_t$  and  $J_t$  will also converge to  $K_* \triangleq P_* A_{N_*}' (A_{N_*} P_* A_{N_*}' + \sigma_{obs}^2 I)^{-1}$  and  $J_* \triangleq I - K_* A_{N_*}$  respectively. For  $t > t_*$ , the GA-KF also runs the same KF. Thus,  $\tilde{P}_{t|t-1}, \tilde{K}_t, \tilde{J}_t$  will also converge to  $P_*, K_*, J_*$  respectively.

We define  $J_* = I - K_* A_{N_*}$ . As the system is controllable and observable we see that the Algebraic Riccati equation has a positive semi-definite solution and the matrix  $I - K_* A_{N_*}$  is stable using Theorem 2. That means as  $J_*$  is stable, i.e. its spectral radius  $\rho = \rho(J_*) < 1$ . Let  $\epsilon_0 = (1 - \rho)/2$ . By Lemma 4, there exists a matrix norm, denoted  $\|\cdot\|_\rho$ , s.t.  $\|J_*\|_\rho \leq \rho + \epsilon_0 = (1 + \rho)/2 < 1$ .

Consider any  $\epsilon_1 < (1 - \rho)/4$ . Depending upon the value of  $\epsilon_1$  we assume that there exists a  $t_{\epsilon_1}$  s.t. for all  $t \geq t_{\epsilon_1}$ ,  $\|K_t - \tilde{K}_t\| < \epsilon_1, \|J_t - \tilde{J}_t\| < \epsilon_1$  and  $\|J_t\|_\rho < \|J_*\|_\rho + \epsilon_1 < (1 + \rho)/2 + (1 - \rho)/4 = (3 + \rho)/4 < 1$ . Let name this delay  $t_{\epsilon_1} - t_*$  as  $\tau_1$ , which depends on  $\epsilon_1, N_*$ . So we can say that for any  $t \in [t_* + \tau_1, t_{**}]$  all the above inequalities hold. Now, the last set of undetected elements of  $N_*$  are detected at  $t_*$ . Thus at  $t_*$ , KF-CS computes a final LS estimate, i.e.  $\hat{x}_{t_*} = A_{N_*}^\dagger y_{t_*}, P_{t_*} = (A'_{N_*} A_{N_*})^{-1} \sigma_{obs}^2, K_{t_*} = (A'_{N_*} A_{N_*})^{-1} A'_{N_*}$  and  $J_{t_*} = 0$ . None of these depend on  $y_1 \dots y_{t_*-1}$  and hence the future values of  $\hat{x}_t$  or of  $P_t, J_t, K_t$  etc also do not. Hence  $t_{\epsilon_1}$  also does not.

Since  $\tilde{P}_{t|t-1} \rightarrow P_*, \tilde{P}_{t|t-1}$  is bounded. Since  $\tilde{P}_t = (I - K_t A_{N_*}) \tilde{P}_{t|t-1} \leq \tilde{P}_{t|t-1}$ ,  $\tilde{P}_t$  is also bounded, i.e. there exists a  $B < \infty$  s.t.  $\text{tr}(\tilde{P}_t) < B, \forall t \in [t_*, t_{**}]$ .

Now as the event  $D_f$  occurs in the interval  $t \in [t_*, t_{**}]$ , the error  $E[\tilde{e}_t \tilde{e}'_t | y_1 \dots y_t] = E[\tilde{e}_t \tilde{e}'_t | y_1 \dots y_t, D_f]$ . Since

$$\mathbb{E}[\tilde{e}_t \tilde{e}'_t | y_1 \dots y_t] = \tilde{P}_t = \mathbb{E}[\tilde{e}_t \tilde{e}'_t] \quad (39)$$

thus

$$\mathbb{E}[\|\tilde{e}_t\|^2 | D_f] = \text{tr}(\tilde{P}_t) < B. \quad (40)$$

Using (34), we get for all  $t \geq t_{\epsilon_1}$

$$\begin{aligned} \text{diff}_t &= J_t \text{diff}_{t-1} + (J_t - \tilde{J}_t)(\tilde{e}_{t-1} + \nu_t) + (\tilde{K}_t - K_t) w_t \\ &= J_t [J_{t-1} \text{diff}_{t-2} + (J_{t-1} - \tilde{J}_{t-1})(\tilde{e}_{t-2} + \nu_{t-1}) + (\tilde{K}_{t-1} - K_{t-1}) w_{t-1}] + (J_t - \tilde{J}_t)(\tilde{e}_{t-1} + \nu_t) + (\tilde{K}_t - K_t) w_t \\ &= J_t J_{t-1} \text{diff}_{t-2} + I u_t + J_t u_{t-1} \\ &= J_t J_{t-1} \dots J_{t-(t_{\epsilon_1}+1)} \text{diff}_{t-t_{\epsilon_1}} + I u_t + J_t u_{t-1} + \dots \\ &\quad + J_t J_{t-1} u_{t-2} + \dots \left( \prod_{k=t_{\epsilon_1}+1}^t J_k \right) u_{t_{\epsilon_1}} \end{aligned} \quad (41)$$

where  $u_t = (J_t - \tilde{J}_t)(\tilde{e}_{t-1} + \nu_t) + (K_t - \tilde{K}_t) w_t$ . Thus, using

(34) and using Cauchy-Schwartz for all  $t \geq t_{\epsilon_1}$ , we get

$$\begin{aligned}
& \mathbb{E}[\|\text{diff}_t\|^2 | D_f]^{1/2} \\
& \leq \|M_{t,t_{\epsilon_1}}\| \mathbb{E}[\|\text{diff}_{t_{\epsilon_1}}\|^2 | D_f]^{1/2} \\
& \quad + \|L_{t,t_{\epsilon_1}}\| \sup_{t_{\epsilon_1} \leq \tau \leq t} \mathbb{E}[\|u_\tau\|^2 | D_f]^{1/2}, \\
& \text{where} \\
& M_{t,t_{\epsilon_1}} \triangleq \prod_{k=t_{\epsilon_1}+1}^t J_k, \\
& L_{t,t_{\epsilon_1}} \triangleq I + J_t + J_t J_{t-1} + \dots \prod_{k=t_{\epsilon_1}+1}^t J_k
\end{aligned} \tag{42}$$

Since neither  $t_{\epsilon_1}$ , nor the matrices  $J_t$  or  $K_t$  for  $t > t_*$ , depend on  $y_1, \dots, y_{t_*}$ , we do not need to condition the expectation on  $y_1, \dots, y_{t_*}$ .

Notice that

- 1)  $\sup_{t_{\epsilon_1} \leq \tau \leq t} \mathbb{E}[\|u_\tau\|^2 | D_f]^{1/2} \leq \epsilon_1(\sqrt{B} + \sqrt{|N_*| \sigma_{sys}^2} + \sqrt{n \sigma_{obs}^2})$ .
- 2)  $\|M_{t,t_{\epsilon_1}}\|_\rho \leq \prod_{\tau=t_{\epsilon_1}+1}^t \|J_\tau\|_\rho < a^{t-t_{\epsilon_1}}$  with  $a \triangleq (3 + \rho)/4 < 1$ . Thus  $\|M_{t,t_{\epsilon_1}}\| \leq c_{\rho,2} a^{t-t_{\epsilon_1}}$  where  $c_{\rho,2}$  is the smallest real number satisfying  $\|M\| \leq c_{\rho,2} \|M\|_\rho$ , for all size  $|N_*|$  square matrices  $M$  (holds because of equivalence of norms).
- 3)  $\|L_{t,t_{\epsilon_1}}\|_\rho \leq 1 + a + \dots a^{t-t_{\epsilon_1}} < \frac{1}{(1-a)}$ .

Thus  $\|L_{t,t_{\epsilon_1}}\| \leq \frac{c_{\rho,2}}{(1-a)}$ .

Combining the above facts, for all  $t \geq t_{\epsilon_1}$ ,

$$\mathbb{E}[\|\text{diff}_t\|^2 | D_f]^{1/2} \leq c_{\rho,2} a^{t-t_{\epsilon_1}} \mathbb{E}[\|\text{diff}_{t_{\epsilon_1}}\|^2 | D_f]^{1/2} + C \epsilon_1 \tag{43}$$

where  $a := (3 + \rho)/4$ ,  $C := \frac{c_{\rho,2}}{1-a}(\sqrt{B} + \sqrt{|N_*| \sigma_{sys}^2} + \sqrt{n \sigma_{obs}^2})$  and  $\mathbb{E}[\|\text{diff}_{t_{\epsilon_1}}\|^2 | D_f]^{1/2}$  is bounded as it is finite. Notice that  $a < 1$ . Consider an  $\tilde{\epsilon} = 2C\epsilon_1$ . It is easy to see that for all  $t \geq t_{\tilde{\epsilon}/2C} + \frac{\log(\mathbb{E}[\|\text{diff}_{t_{\tilde{\epsilon}/2C}\|}^2 | D_f]^{1/2}) + \log(2c_{\rho,2}) - \log \tilde{\epsilon}}{\log(1/a)}$ ,

$$\mathbb{E}[\|\text{diff}_t\|^2 | D_f]^{1/2} \leq \tilde{\epsilon} \tag{44}$$

Name this delay  $t - t_{\epsilon}$  as  $\tau_2$ , which depends on  $\epsilon_1, N_*$ . So we see that for any  $t \in [t_* + \tau_1 + \tau_2, t_{**}]$  the mean-square error is less than  $\tilde{\epsilon}$ .

From Markov's inequality, we have for any  $t \in [t_* + \tau_1 + \tau_2, t_{**}]$

$$\begin{aligned}
P(\|\text{diff}_t\| > \epsilon_{err} | D_f) & \leq \frac{\mathbb{E}[\|\text{diff}_t\|^2 | D_f]^{1/2}}{\epsilon_{err}} \\
& \leq \frac{\tilde{\epsilon}}{\epsilon_{err}}
\end{aligned}$$

So we can say that  $P(\|\text{diff}_t\| > \epsilon_{err} | D_f) \leq \epsilon$  where  $\epsilon = \frac{\tilde{\epsilon}}{\epsilon_{err}}$ . Now we see that both  $\tau_1$  and  $\tau_2$  depends on  $\epsilon_1, N_*$ . Hence they will depend on  $\epsilon, \epsilon_{err}$  and  $N_*$ . So for a given  $\epsilon$  and a given  $\epsilon_{err}$  there exists a  $\tau_{KF}(\epsilon, \epsilon_{err}, N_*) > \tau_1 + \tau_2$  s.t. for all  $t \geq t_* + \tau_{KF}(\epsilon, \epsilon_{err}, N_*)$ ,  $P(\|\text{diff}_t\|^2 < \epsilon_{err} | D_f) \geq (1 - \epsilon)$ .

#### D. Proof of Theorem 1

The events  $E_j$  and  $F_j$  are defined in Lemma 1. At the first addition time,  $t_0 = 1$ , using the initialization condition,  $\hat{N}_{t_0-1} = N_{t_0-1}$ , i.e.  $F_0$  holds. Thus, by Lemma 1,  $Pr(E_0) = \left(2Q\left(\sqrt{\frac{4B_*}{(\tau_{det})\sigma_{sys}^2}}\right)\right)^{S_a} > 1 - \epsilon$  for some  $\epsilon > 0$ . Let denote  $Pr(E_0) = 1 - \epsilon + \delta = 1 - \epsilon_2$  for some  $\delta > 0$ . Now  $Pr(E_1) = Pr(E_1 \cap E_0) + Pr(E_1 \cap E_0^c) = Pr(E_1|E_0)Pr(E_0) + Pr(E_1|E_0^c)Pr(E_0^c)$ . As we get from the Lemma 1,  $Pr(E_1|E_0) = Pr(E_1|F_1) = Pr(E_0|F_0) = Pr(E_0) = 1 - \epsilon_2$ . Now to calculate  $Pr(E_1 \cap E_0^c)$  we have to think that at time  $t \in [t_0 + \tau_{det}, t_1 - 1]$  not every new indices are detected, but all those indices are detected within the next detection time, i.e during the time interval  $t \in [t_0 + \tau_{det}, t_2 - 1]$  all those indices will be detected. And also the new addition indices will be detected within the detection time  $t \in [t_1 + \tau_{det}, t_1 - 1]$ .

As every index detection is an independent process so we can conclude that  $Pr(E_1|E_0^c) = \left(2Q\left(\sqrt{\frac{4B_*}{(d+\tau_{det})\sigma_{sys}^2}}\right)\right)^{S_a} \times \left(2Q\left(\sqrt{\frac{4B_*}{(\tau_{det})\sigma_{sys}^2}}\right)\right)^{S_a}$ . Now as  $d + \tau_{det} > \tau_{det}$  so  $\left(2Q\left(\sqrt{\frac{4B_*}{(d+\tau_{det})\sigma_{sys}^2}}\right)\right)^{S_a} > \left(2Q\left(\sqrt{\frac{4B_*}{(\tau_{det})\sigma_{sys}^2}}\right)\right)^{S_a}$ . Then we have  $Pr(E_1 \cap E_0^c) = Pr(E_1|E_0^c)Pr(E_0^c) > \epsilon_2(1 - \epsilon_2)^2$ . Hence  $Pr(E_1) > (1 - \epsilon_2)^2 + \epsilon_2(1 - \epsilon_2)^2 = 1 - \epsilon_2 - \epsilon_2^2 + \epsilon_2^3$ . As  $\epsilon_2$  is arbitrarily small so we can neglect  $\epsilon_2^2$  and  $\epsilon_2^3$ . That means  $Pr(E_1) > 1 - \epsilon_2 > 1 - \epsilon$ . Now to prove the same for any time  $t = t_j$  we will use the induction method. Let for  $j - 1$ ,  $Pr(E_{j-1}) > 1 - \epsilon$ . So again for some  $\delta_1 > 0$ ,  $Pr(E_{j-1}) = 1 - \epsilon + \delta_1 = 1 - \epsilon_3$  and  $Pr(E_j|E_{j-1}) = 1 - \epsilon_2$  from Lemma 1.  $Pr(E_j \cap E_{j-1}^c) = Pr(E_j \cap E_{j-1}^c \cap E_{j-2}) + Pr(E_j \cap E_{j-1}^c \cap E_{j-2}^c) = Pr(E_j \cap E_{j-1}^c | E_{j-2})Pr(E_{j-2}) + Pr(E_j \cap E_{j-1}^c | E_{j-2}^c)Pr(E_{j-2}^c) = Pr(E_j|E_{j-1}^c, E_{j-2})Pr(E_{j-1}^c | E_{j-2})Pr(E_{j-2}) + Pr(E_j \cap E_{j-1}^c \cap E_{j-2}^c)$ . If we notice the first term we see that  $Pr(E_j|E_{j-1}^c, E_{j-2}) = Pr(E_1|E_0^c)$  and  $Pr(E_{j-1}^c | E_{j-2}) = \epsilon_3$ . We can similarly try to split the second term  $Pr(E_j \cap E_{j-1}^c \cap E_{j-2}^c)$  conditioned on the event  $E_{j-3}$  and so on. Then we have  $Pr(E_j) > (1 - \epsilon_3)(1 - \epsilon_2) + \epsilon_3(1 - \epsilon_2)^2(1 - \epsilon_4) + \dots$  some positive term, where for some  $\epsilon_4 > 0$ ,  $Pr(E_{j-2}) = 1 - \epsilon_4$ . As  $\epsilon_2, \epsilon_3$  and  $\epsilon_4$  are arbitrarily small, so neglecting the higher order of  $\epsilon_2, \epsilon_3$  and  $\epsilon_4$  the above inequality get the following simplified form:  $Pr(E_j) > 1 - \epsilon_2 > 1 - \epsilon$ .

So we observe that  $Pr(E_j) > 1 - \epsilon$  for some  $\epsilon > 0$ . The detection delay  $\tau_{det}$  depends on  $\epsilon$ . Lemma 2 gives us  $Pr(\|\text{diff}_t\|^2 \leq \epsilon_{err} | D_f) > 1 - \epsilon'$  for some  $\epsilon' > 0$  and  $D_f$  is the event which is denoted as  $D_f := \{\hat{N}_t = N_t = N_*, \forall t \in [t_*, t_{**}]\}$ . Assume that  $E_j$  occurs and apply Lemma 2 with  $t_* = t_j + \tau_{det}(\epsilon, S_a)$  and  $t_{**} = t_{j+1} - 1$ . From Lemma 2 we get  $Pr(\|\text{diff}_t\|^2 \leq \epsilon_{err} | E_j) \geq (1 - \epsilon')$ . The kalman Filter delay  $\tau_{KF}$  depends on  $\epsilon'$  and  $\epsilon_{err}$ . So combining these two results we get  $Pr(\|\text{diff}_t\|^2 \leq \epsilon_{err}) \geq Pr(\|\text{diff}_t\|^2 \leq \epsilon_{err}, E_j) \geq (1 - \epsilon)(1 - \epsilon')$ . Again neglecting the term  $\epsilon\epsilon'$ ,  $Pr(\|\text{diff}_t\|^2 \leq \epsilon_{err}) \geq 1 - \epsilon - \epsilon'$ . Define  $\epsilon'' = \epsilon + \epsilon'$ . Then  $Pr(\|\text{diff}_t\|^2 \leq \epsilon_{err}) \geq 1 - \epsilon''$  and also we notice that  $\tau_{det}$  and  $\tau_{KF}$  both depend on  $\epsilon''$ . So the first claim is proved.

Clearly <sup>1</sup>  $Pr(E_j|E_0, E_1, \dots, E_{j-1}) = Pr(E_j|E_{j-1}) = Pr(E_j|F_j)$ . By Lemma 1,  $Pr(E_j|F_j) \geq 1 - \epsilon$ . Combining this with  $Pr(E_0) \geq 1 - \epsilon$ , we get  $Pr(E_j \cap E_{j-1} \cap \dots \cap E_0) = Pr(E_0)Pr(E_1|F_1) \dots Pr(E_j|F_j) \geq (1 - \epsilon)^{j+1}$ . The second and the third claim follow directly from the before-mentioned arguments.

<sup>1</sup>since  $E_j = \{(x_{t_j + \tau_{\text{det}}})_i^2 > 4B_*, \forall i \in \Delta_{t_j + \tau_{\text{det}}}\}$  and the sequence of  $x_t$ 's is a Markov process