KF-CS: Compressive Sensing on Kalman Filtered Residual

I. LEMMA AND THEOREM

A. Lemma 1

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 α is high enough, there are no false detections. If the delay between addition times also satisfies $d > \tau_{\text{det}}(\epsilon, S_a)$, where τ_{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 $t_j + \tau_{\text{det}}(\epsilon, S_a) < t_{j+1}$.

Lemma 1: Assume that x_t follows Signal Model **??**. 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, $d > \tau_{\text{det}}(\epsilon, S_a)$,

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

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

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.

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 $|\Delta_{e,t}| = 0$. 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] $\|x_t - \hat{x}_{t,\text{CSres}}\|^2 \leq B_*$. 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 $t - 1$, $|\Delta_{e,t-1}| = 0$. Thus β_t is $|N_t \cup \Delta_{e,t-1}| = |N_t|$ 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 $x_t - \hat{x}_{t,\text{CSres}} = \beta_t - \hat{\beta}_t$ 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,

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

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

C. Lemma 2

Lemma 2: Assume that x_t follows Signal Model **??**; $\delta_{S_{\max}} < 1$; and $\alpha_{del} = 0$. Define the event $D_f := \{\hat{N}_t = N_t = N_*, \ \forall \ t \in [t_*, \ t_{**}]\}$. For a given $\epsilon, \epsilon_{\text{err}}$, there exists a $\tau_{KF}(\epsilon, \epsilon_{\text{err}}, N_*)$ s.t. for all $t \in [t_* + \tau_{KF}, \ t_{**}]$, $Pr(\|\text{diff}_t\|^2 \leq \epsilon_{\text{err}} \mid D_f) > 1 - \epsilon$. 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 III-D. After review, this can be significantly shortened. The proof involves two parts. First, we use the results from [**?**] and [**?**] to show that (a) $P^{\ddagger}_{t|t-1}, P^{\ddagger}_t, K^{\ddagger}_t$ and $J_t := I - K^{\ddagger}_t A_{N_*}$, where $P^{\ddagger}_{t|t-1} = (P_{t|t-1})_{N_*,N_*}, P^{\ddagger}_t = (P_t)_{N_*,N_*}, K^{\ddagger}_t = (K_t)_{N_*,[1:m]}$, 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 $\|.\|_\rho$, s.t. $\|J_*\|_\rho < 1$. Second, we use (a) and (b) to show that the difference in the KF-CS and genie-KF estimates, diff_t, converges to zero in mean square, and hence also in probability (by Markov's inequality).

D. 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 $\hat{x}_t = \hat{x}_{t,\text{init}}$. 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

$$\begin{aligned} e_t &= J_t e_{t-1} + J_t \nu_t - K_t w_t \\ \tilde{e}_t &= \tilde{J}_t \tilde{e}_{t-1} + \tilde{J}_t \nu_t - \tilde{K}_t w_t \\ \text{diff}_t &= J_t \text{diff}_{t-1} + (J_t - \tilde{J}_t)(\tilde{e}_{t-1} + \nu_t) + (\tilde{K}_t - K_t) w_t \end{aligned} \tag{3}$$

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

$$\begin{aligned} x_t &= x_{t-1} + \nu_t \\ y_t &= A x_t + w_t \end{aligned} \tag{4}$$

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 **??** . The state-noise in appendix **??** is denoted as ν 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 $F \equiv I, \ Q \equiv (\sigma^2_{sys} I_{N_*})_{N_*,N_*}, \ A \equiv A_{N_*}, \ R \equiv \sigma^2_{obs} I$, *but with different initial conditions*. KF-CS uses $\hat{x}_{t_*}$, $P_{t_*+1|t_*} \neq \mathbb{E}[e_{t_*+1} e'_{t_*+1} | y_1 \ldots y_{t_*}]$ while GA-KF uses the correct initial conditions, $\hat{x}_{t_*,GAKF}$, $\tilde{P}_{t_*+1|t_*} = \mathbb{E}[\tilde{e}_{t_*+1} \tilde{e}'_{t_*+1} | y_1, \ldots y_{t_*}]$ Since $|N_*| \leq S_{\max}$ and $\delta_{S_{\max}} < 1$, $A \equiv A_{N_*}$ is full rank.We can rewrite the equ. 32 in the following form

$$\begin{aligned} x_t &= x_{t-1} + Q^{1/2} \eta_t \\ y_t &= A_{N_*} x_t + w_t \end{aligned} \tag{5}$$

where η_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 **??**. 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

$$\begin{aligned} & P_{t+1|t} \\ &= P_t + Q \\ &= (I - K_t A_{N_*}) P_{t|t-1} + Q \\ &= P_{t|t-1} + Q - K_t A_{N_*} P_{t|t-1} \\ &= P_{t|t-1} + Q - P_{t|t-1} A'_{N_*} (A_{N_*} P_{t|t-1} A'_{N_*} + R)^{-1} A_{N_*} P_{t|t-1} \end{aligned} \tag{6}$$

So that is how we got our Riccati equation. Now the observability matrix is $[A_{N_*} \quad A_{N_*}I \quad A_{N_*}I^2 \quad . \quad . \quad . \quad A_{N_*}I^{n_*-1}]'$, 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. Similarily our controllability matrix $[Q^{1/2} \quad IQ^{1/2} \quad ...I^{n_*-1}Q^{1/2}]$ is also of full rank as $Q^{1/2}$ matrix is full rank. So $(I, Q^{1/2})$ is controllable. Thus, according to Lemma **??**,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

$$P_{t+1|t} = P_{t|t-1} + Q - P_{t-1|t}A'_{N_*}[A_{N_*}P_{t|t-1}A'_{N_*} + R]^{-1} \\ A_{N_*}P_{t|t-1} \tag{7}$$

Consequently K_t and J_t will also converge to $K_* \triangleq P_*{A_{N_*}}'(A_{N_*}P_*{A_{N_*}}' + \sigma^2_{obs}I)^{-1}$ 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-deifinte solution and the matrix $I - K_*A_{N_*}$ is stable using Theorem **??**. That means as J_* is stable, i.e. its spectral radius $\rho = \rho(J_*) < 1$. Let $\epsilon_0 = (1-\rho)/2$. By Lemma **??**, there exists a matrix norm, denoted $\|.\|_\rho$, s.t. $\|J_*\|_\rho \le \rho + \epsilon_0 = (1+\rho)/2 < 1$.

Consider any $\epsilon_1 < (1-\rho)/4$. Depending upon the value of ϵ_1 we assume that there exists a t_{ϵ_1} s.t. for all $t \ge t_{\epsilon_1}$, $\|K_t - \tilde{K}_t\| < \epsilon_1$, $\|J_t - \tilde{J}_t\| < \epsilon_1$ 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 τ_1, which depends on ϵ_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. $\hat{x}_{t_*} = {A_{N_*}}^\dagger y_{t_*}$, $P_{t_*} = (A'_{N_*}A_{N_*})^{-1}\sigma^2_{obs}$, $K_{t_*} = (A'_{N_*}A_{N_*})^{-1}A'_{N_*}$ and $J_{t_*} = 0$ None of these depend on $y_1 \ldots y_{t_*-1}$ and hence the future values of $\hat{x}_t$ or of P_t, J_t, K_t etc also do not. Hence t_{ϵ_1} also does not.

Since $\tilde{P}_{t|t-1} \to P_*$, $\tilde{P}_{t|t-1}$ is bounded. Since $\tilde{P}_t = (I - K_tA_{N_*})\tilde{P}_{t|t-1} \le \tilde{P}_{t|t-1}$, $\tilde{P}_t$ 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 $E[\tilde{e}_t\tilde{e}'_t|y_1 \ldots y_t] = E[\tilde{e}_t\tilde{e}'_t|y_1 \ldots y_t, D_f]$. Since

$$\mathbb{E}[\tilde{e}_t\tilde{e}'_t|y_1 \ldots y_t] = \tilde{P}_t = \mathbb{E}[\tilde{e}_t\tilde{e}'_t] \tag{8}$$

thus

$$\mathbb{E}[\|\tilde{e}_t\|^2|D_f] = \text{tr}(\tilde{P}_t) < B. \tag{9}$$

Using (31), we get for all $t \ge t_{\epsilon_1}$

$$\begin{aligned} &\text{diff}_t \\ &= J_t\text{diff}_{t-1} + (J_t - \tilde{J}_t)(\tilde{e}_{t-1} + \nu_t) + \\ &\quad (\tilde{K}_t - K_t)w_t \\ &= J_t[J_{t-1}\text{diff}_{t-2} + (J_{t-1} - \tilde{J}_{t-1})(\tilde{e}_{t-2} + \\ &\quad \nu_{t-1}) + (\tilde{K}_{t-1} - K_{t-1})w_{t-1}] + (J_t - \tilde{J}_t)(\tilde{e}_{t-1} + \nu_t) + \\ &\quad (\tilde{K}_t - K_t)w_t \\ &= J_tJ_{t-1}\text{diff}_{t-2} + Iu_t + J_tu_{t-1} \\ &= J_tJ_{t-1}...J_{t-(t_{\epsilon_1}+1)}\text{diff}_{t-t_{\epsilon_1}} + Iu_t + J_tu_{t-1} + \\ &\quad J_tJ_{t-1}u_{t-2} + ..(\prod_{k=t_{\epsilon_1}+1}^{t} J_k)u_{t_{\epsilon_1}} \end{aligned} \tag{10}$$

where $u_t = (J_t - \tilde{J}_t)(\tilde{e}_{t-1} + \nu_t) + (K_t - \tilde{K}_t)w_t$ Thus, using (31) and using Cauchy-Schwartz for all $t \geq t_{\epsilon_1}$, we get

$$\begin{aligned}
&\mathbb{E}[\|\text{diff}_t\|^2|D_f]^{1/2} \\
&\leq \|M_{t,t_{\epsilon_1}}\| \ \mathbb{E}[\|\text{diff}_{t_{\epsilon_1}}\|^2|D_f]^{1/2} \\
&+\|L_{t,t_{\epsilon_1}}\| \sup_{t_{\epsilon_1} \leq \tau \leq t} \mathbb{E}[\|u_\tau\|^2|D_f]^{1/2}, \\
&\text{where} \\
&M_{t,t_{\epsilon_1}} \triangleq \prod_{k=t_{\epsilon_1}+1}^{t} J_k, \\
&L_{t,t_{\epsilon_1}} \triangleq I + J_t + J_t J_{t-1} + .. \prod_{k=t_{\epsilon_1}+1}^{t} J_k
\end{aligned} \tag{11}$$

Since neither t_{ϵ_1}, nor the matrices J_t or K_t for $t > t_*$, depend on $y_1, \ldots y_{t_*}$, we do not need to condition the expectation on $y_1, \ldots y_{t_*}$.

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

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

$$\mathbb{E}[\|\text{diff}_t\|^2|D_f]^{1/2} \leq c_{\rho,2} a^{t-t_{\epsilon_1}} \mathbb{E}[\|\text{diff}_{t_{\epsilon_1}}\|^2|D_f]^{1/2} + C\epsilon_1 \tag{12}$$

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 $\mathbb{E}[\|\text{diff}_{t_\epsilon}\|^2|D_f]^{1/2}$ 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 $t \geq t_{\tilde{\epsilon}/2C} + \frac{\log(\mathbb{E}[\|\text{diff}_{t_{\tilde{\epsilon}/2C}}\|^2|D_f]^{1/2}) + \log(2c_{\rho,2}) - \log\tilde{\epsilon}}{\log(1/a)}$,

$$\mathbb{E}[\|\text{diff}_t\|^2|D_f]^{1/2} \leq \tilde{\epsilon} \tag{13}$$

Name this delay $t - t_\epsilon$ as τ_2 , which depends on ϵ_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 $t \in [t_* + \tau_1 + \tau_2, t_{**}]$

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

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 τ_1 and τ_2 depends on ϵ_1, N_*. Hence they will depend on ϵ,ϵ_{err} and N_*. So for a given ϵ and a given ϵ_{err} there exists a $\tau_{KF}(\epsilon, \epsilon_{\text{err}}, N_*) > \tau_1 + \tau_2$ s.t. for all $t \geq t_* + \tau_{KF}(\epsilon, \epsilon_{\text{err}}, N_*)$, $Pr(\|\text{diff}_t\|^2 < \epsilon_{\text{err}} \mid D_f) \geq (1-\epsilon)$.

E. Theorem 1

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 **??**. 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 $\epsilon, \epsilon_{\text{err}}$, if the conditions of Lemma 1 hold, and if the delay between addition times, $d > \tau_{\text{det}}(\epsilon, S_a) + \tau_{KF}(\epsilon, \epsilon_{\text{err}}, N_{t_j})$, where $\tau_{\text{det}}(.,.)$ is defined in (29) in Lemma 1 and $\tau_{KF}(.,.,.)$ in Lemma 2, then

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

The proof is given in Appendix III-F. 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 ($t_K = \infty$).

F. 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, $Pr(E_0) = \left(2\mathcal{Q}\left(\sqrt{\frac{4B_*}{(\tau_{det})\sigma_{sys}^2}}\right)\right)^{S_a} > 1-\epsilon$ for some $\epsilon > 0$. Let denote $Pr(E_0) = 1-\epsilon+\delta = 1-\epsilon_2$ for some $\delta > 0$. Now $Pr(E_1) = Pr(E_1 \cap E_0) + Pr(E_1 \cap E_0^c) = Pr(E_1|E_0)Pr(E_0) + Pr(E_1 \cap E_0^c)$. 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 $Pr(E_1|E_0^c) = \left(2\mathcal{Q}\left(\sqrt{\frac{4B_*}{(d+\tau_{det})\sigma_{sys}^2}}\right)\right)^{S_a} \times \left(2\mathcal{Q}\left(\sqrt{\frac{4B_*}{(\tau_{det})\sigma_{sys}^2}}\right)\right)^{S_a}$. Now as $d + \tau_{det} > \tau_{det}$ so $\left(2\mathcal{Q}\left(\sqrt{\frac{4B_*}{(d+\tau_{det})\sigma_{sys}^2}}\right)\right)^{S_a} > \left(2\mathcal{Q}\left(\sqrt{\frac{4B_*}{(\tau_{det})\sigma_{sys}^2}}\right)\right)^{S_a}$. 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 ϵ_2 is arbitrarily small so we can neglect ϵ_2^2 and ϵ_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. $Pr(E_j \cap E_{j-1}^c) = Pr(E_j \cap E_{j-1}^c \cap E_{j-2}) + Pr(E_j \cap E_{j-1}^c \cap E_{j-2}^c) = Pr(E_j \cap E_{j-1}^c|E_{j-2})Pr(E_{j-2}) + Pr(E_j \cap E_{j-1}^c \cap E_{j-2}^c) = Pr(E_j|E_{j-1}^c, E_{j-2})Pr(E_{j-1}^c|E_{j-2})Pr(E_{j-2}) + Pr(E_j \cap E_{j-1}^c \cap E_{j-2}^c)$. 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 similarily 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 $Pr(E_j) > (1-\epsilon_3)(1-\epsilon_2) + \epsilon_3(1-\epsilon_2)^2(1-\epsilon_4) +$ some positive term, where for some $\epsilon_4 > 0$, $Pr(E_{j-2}) = 1-\epsilon_4$. As ϵ_2, ϵ_3 and ϵ_4 are arbitrarily small , so neglecting the higher order of ϵ_2,ϵ_3 and ϵ_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 τ_{det} depends on ϵ . Lemma 2 gives us $Pr(\|\text{diff}_t\|^2 \leq \epsilon_{\text{err}}|D_f) > 1-\epsilon'$ 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 $t_* = t_j + \tau_{\text{det}}(\epsilon, S_a)$ 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 τ_{KF} depends on ϵ' and ϵ_{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 τ_{det} and τ_{KF} both depend on ϵ'' . So the first claim is proved.

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

[1]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