KFCS Theory

Animesh Biswas, Kevin Palmowski

Last compiled August 11, 2014

1 Model

Major change: time indexing redone to match NV original. t_0 is now the first addition and we assume there's an initial $(t_0 - 1)$ step.

At each time $t \geq (t_0 - 1)$ (**do we have a y_{t_0-1}?**), we have

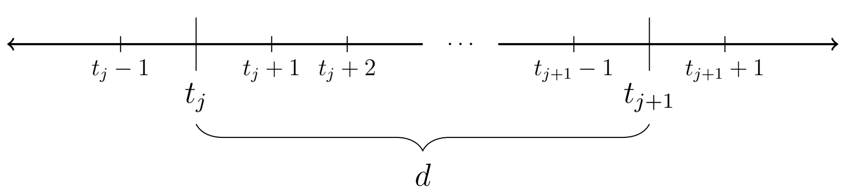
$$\begin{aligned} y_t &= Ax_t + w_t \\ x_{t+1} &= x_t + \nu_{t+1} \end{aligned}$$

Here, $\mathbb{E}[w_t] = \mathbf{0}$, $\text{cov}(w_t) = \mathbb{E}[w_t w_t'] = R = \sigma^2_{\text{obs}} I_{n\times n}$, iid and independent of x_t; $x_{t_0-1} \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\text{sys},0} I_{N_{t_0-1}})$; and $\nu_t \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\text{sys}} I_{N_t})$ iid. for $t \geq t_0$

$y_t, w_t \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times m}$, $x_t, \nu_t \in \mathbb{R}^m$.

Time indices are discrete. Make the distinction between sampling times (used) and continuous time (not used).

Update picture?



For $j \geq 0$, we have the addition times $\{t_j\}$. The initial time is $t = (t_0 - 1)$. At the addition times $t_j = t_0 + jd$, the support of x_t changes: $N_t = N_{t_j}$ for all $t \in [t_j : t_{j+1} - 1]$, and $N_{t_j} \subset N_{t_{j+1}}$.

2 Algorithm – KFCS with LS

This algorithm applies to the case where there are no support deletions.

Issues:

P_{t_0-1} **and** Q_t **– is this an identity of size** $|Nhat|$ **or is it a full-blown identity with nonzeros on diagonal entries corresponding to** $Nhat$

Is this algorithm transcribed correctly? There are 3 versions of it that I have (NV original, AB+NV typed draft, and AB handwritten) and all 3 are different.

Look for places to simplify – this is long and contains repeat steps, which is non-ideal

Needs to be redone for the new timescale

Input: σ_{sys}, σ_{obs}, $\sigma_{\text{sys},0}$, A, $\{t_j\}$, $\{N_t\}$, $\{y_t\}$

$\hat{x}_{t_0,\text{init}} = \arg\min_x \|x\|_1$ subject to $\|y_{t_0} - Ax\|_2 < \xi$
$\hat{N}_{t_0} = \{k : |(\hat{x}_{t_0,\text{init}})_k| > \alpha\}$
$P_{t_0-1} = \sigma^2_{\text{sys},0} I_{\hat{N}_{t_0}}$
$Q_{t_0} = 0$
$\hat{x}_{t_0-1} = \mathbf{0}$

$P_{t_0|t_0-1} = P_{t_0-1} + Q_{t_0}$
$K_{t_0} = P_{t_0|t_0-1} A' \left(A P_{t_0|t_0-1} A' + \sigma^2_{\text{obs}} I\right)^{-1}$
$J_{t_0} = I - K_{t_0} A$
$P_{t_0} = J_{t_0} P_{t_0|t_0-1}$
$\hat{x}_{t_0} = J_{t_0} \hat{x}_{t_0-1} + K_{t_0} y_{t_0}$

for $t > t_0$ **do**
  $Q_t = \sigma^2_{\text{sys}} I_{\hat{N}_{t-1}}$
  $P_{t|t-1} = P_{t-1} + Q_t$
  $K_t = P_{t|t-1} A' \left(A P_{t|t-1} A' + \sigma^2_{\text{obs}} I\right)^{-1}$
  $J_t = I - K_t A$
  $P_t = J_t P_{t|t-1}$
  $\hat{x}_{t,\text{init}} = J_t \hat{x}_{t-1} + K_t y_t$

  $y_{t,\text{res}} = y_t - A\hat{x}_{t,\text{init}}$
  $\hat{\beta}_t = \arg\min_\beta \|\beta\|_1$ subject to $\|y_{t,\text{res}} - A\beta\|_2 < \xi$
  $\hat{x}_{t,\text{CSres}} = \hat{x}_{t,\text{init}} + \hat{\beta}_t$

  $\Delta_A = \{k : |(\hat{x}_{t,\text{CSres}})_k| > \alpha\}$
  $\hat{N}_t = \hat{N}_{t-1} \cup \Delta_A$

  if $\Delta_A = \emptyset$ **then**
    $\hat{x}_t = \hat{x}_{t,\text{init}}$
  else
    $\hat{x}_t = \mathbf{0}$
    $(\hat{x}_t)_{\hat{N}_t} = (A_{[1:n],\hat{N}_t})^\dagger y_t$
    $P_t = 0_{m\times m}$
    $(P_t)_{\hat{N}_t,\hat{N}_t} = \left[(A_{[1:n],\hat{N}_t})'(A_{[1:n],\hat{N}_t})\right]^{-1} \sigma^2_{\text{obs}} I_{|\hat{N}_t|}$
  end
end

Algorithm 1: Kalman-Filtered Compressed Sensing (KFCS)

3 Algorithm – Genie-Aided Kalman Filtering (GAKF)

This algorithm applies to the case where there are no support deletions.

Issues:

Check blue piece below – do we want all-ones, identity of size $|DeltaA|$, or identity restricted to DeltaA and zero else?

Needs to be redone for the new timescale

Input: σ_{sys}, σ_{obs}, $\sigma_{\text{sys},0}$, A, $\{t_j\}$, $\{N_t\}$, $\{y_t\}$

for $t \geq t_0$ **do**
  if $t = t_0$ **then**
    $T = N_0$
    $\tilde{P}_{t-1} = \sigma^2_{\text{sys},0} I_T$
    $\tilde{x}_{t-1} = \mathbf{0}$
    $\tilde{Q}_t = 0$
  else
    $T = N_{t-1}$
    $\tilde{Q}_t = \sigma^2_{\text{sys}} I_T$
    if $t = t_j$ *for some* $j > 0$ **then**
      $\Delta_A = N_t \setminus N_{t-1}$
      $\left(\tilde{P}_{t-1}\right)_{\Delta_A, \Delta_A} = \sigma^2_{\mathbf{sys}} I_{|\Delta_A|}$
    end
  end
  $\tilde{P}_{t|t-1} = \tilde{P}_{t-1} + \tilde{Q}_t$
  $\tilde{K}_t = \tilde{P}_{t|t-1} A' \left(A \tilde{P}_{t|t-1} A' + \sigma^2_{\text{obs}} I \right)^{-1}$
  $\tilde{J}_t = I - \tilde{K}_t A$
  $\tilde{P}_t = \tilde{J}_t \tilde{P}_{t|t-1}$
  $\tilde{x}_t = \tilde{J}_t \tilde{x}_{t-1} + \tilde{K}_t y_t$
end

Algorithm 2: Genie-Aided Kalman Filter (GAKF)

4 Candes RIP – C_1 Computation for α

[1], Theorem 1.3: Suppose $y = Ax + \eta$, $|\operatorname{supp}(x)| = s$, $\delta_{2s} = \delta_{2s}(A) < \sqrt{2} - 1$, and $\|\eta\|_2 \leq \xi$. Then

$$\hat{x} = \arg\min_z \|z\|_1 \text{ subject to } \|y - Az\|_2 \leq \xi$$

satisfies

$$\|x - \hat{x}\|_2 \leq C_1(s)\xi,$$

where

$$C_1(s) = \frac{4\sqrt{1+\delta_{2s}}}{1-(1+\sqrt{2})\delta_{2s}} \ .$$

Claim / Note: It can be shown that C_1 is an increasing function of δ_{2s}, and δ_{2s} is an increasing function of s, so C_1 is an increasing function of s.

For any support size s in this paper, we will have $s \leq S_{\max}$ and thus $C_1(s) \leq C_1(S_{\max})$.

5 Linear Systems Theory

5.1 Definitions

We present some basic definitions from linear systems theory. These can be found in [3], Appendix C. Throughout, let $F, G, H \in \mathbb{R}^{n\times n}$.

A matrix F is **stable** if $\rho(F) < 1$.

The pair $\{F, G\}$ is **controllable** if the matrix $[G, FG, \dots, F^{n-1}G]$ is full rank n. An equivalent characterization of controllability is that $\mathrm{rank}([\lambda I - F,\ G]) = n$ for all eigenvalues λ of F.

The pair $\{F, G\}$ is **stabilizable** if $\mathrm{rank}([\lambda I - F,\ G]) = n$ for all eigenvalues λ of F with $|\lambda| \geq 1$.

The pair $\{F, H\}$ is **detectable** if and only if $\{F', H'\}$ is stabilizable.

Consider the case where $F = I$. Then $\lambda = 1$ is the only eigenvalue of $F = F'$ and the matrix $[\lambda I - F,\ G] = [0, G]$ has rank n if and only if G has rank n. Therefore, if G is full rank, then $\{I, G\}$ is controllable and stabilizable. Additionally, since $\mathrm{rank}(H) = \mathrm{rank}(H')$, we can use the same argument to conclude that $\{I, H\}$ is detectable if H is full rank.

5.2 Theoretical Results

Here we present two important theoretical results from linear systems theory.

The general form of a **discrete-time algebraic Riccati equation (DARE)** is

$$P = FPF' + GQG' - (FPH' + GS)(R + HPH')^{-1}(FPH' + GS)', \tag{1}$$

where $P, F, G, H, Q, R, S \in \mathbb{R}^{n\times n}$.

[2], Theorem 7.5.1.b: Consider the DARE (1), where $\{F, H\}$ is detectable and

$$\begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \succ 0.$$

If, in addition, $\{F - GSR^{-1}H, GQ - GSR^{-1}S'\}$ is stabilizable, then the DARE always has a unique Hermitian and positive semi-definite stabilizing solution P such that $F - K_pH$ is stable, where $K_p = (FPH' + GS)(R + HPH')^{-1}$.

The general form of a **discrete-time algebraic Riccati recursion (DARR)** is

$$P_{i+1} = FP_iF' + GQG' - K_{p,i}R_{e,i}K_{p,i}', \quad i \geq 0 \tag{2}$$

where $K_{p,i} = (FP_iH' + GS)(R + HP_iH')^{-1}$, $R_{e,i} = R + HP_iH'$, and $\{P_k\}, F, G, H, Q, R, S \in \mathbb{R}^{n\times n}$.

[2], Lemma 8.7.3: Consider the Riccati recursion (2) with positive semi-definite initial condition $P_0 \succeq 0$. If $Q \succ 0$, $R \succ 0$, $\{F, H\}$ is detectable and $\{F - GSR^{-1}H,\, GQ - GSR^{-1}S'\}$ is stabilizable then P_i converges to the unique positive semi-definite matrix, P, that satisfies the discrete-time algebraic Riccati equation (1).

6 Proofs

Lemma 1. *Assume that* $\{x_t\}$ *and* $\{y_t\}$ *follow the signal model above,* $\{t\}$ *is a discrete set of sampling times, only additions to true support* ($N_t \subseteq N_{t+1}$ *for all* t)*, etc.*

Further assume that

i) *The true solution is exactly recovered at the initial time* $t = (t_0 - 1)$: $\hat{x}_{t_0-1} = x_{t_0-1}$, *so* $\hat{N}_{t_0-1} = N_{t_0-1}$; ***Can we relax this to just the true support is recovered?***

ii) *The maximum support size* S_{max} *satisfies* $S_{max} \leq S_{**} = \max\{s : \delta_{2s}(A) < \sqrt{2} - 1\}$;

iii) *The observation noise* w_t *is bounded in magnitude:* $\|w_t\|_2 < \xi$ *for all* t *and some* $\xi > 0$;

iv) *The addition thresholds* α_t *satisfy* $\alpha_t = \alpha = C\xi$ *for all* t*, where*

$$C = C(S_{max}) = \frac{4\sqrt{1 + \delta_{2S_{max}}}}{1 - \left(1 + \sqrt{2}\right)\delta_{2S_{max}}}$$

with $\delta_{2S_{max}} = \delta_{2S_{max}}(A)$; *and*

v) *The addition delay* d *satisfies* $d > \tau_{det}$*, where the detection delay* τ_{det} *is defined by*

$$\tau_{det} = \tau_{det}(\alpha, \varepsilon) = \left\lceil \left(\frac{2\alpha}{\sigma_{sys} \mathcal{Q}^{-1}\left(\frac{(1-\varepsilon)^{1/S_{add}}}{2}\right)} \right)^2 - 1 \right\rceil .$$

Here, $\mathcal{Q}^{-1}(x)$ *is the inverse of the Gaussian* $\mathcal{Q}$*-function,* $\mathcal{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,dt$.

Then

1) $\|x_t - \hat{x}_{t,CSres}\|_2 \leq \alpha$ *for all sampling times* $t \geq (t_0 - 1)$;

2) *There are no false support additions:* $\hat{N}_t \subseteq N_t$ *for all* $t \geq (t_0 - 1)$; *and*

3) *For any* $j \geq 0$, $\boldsymbol{Pr}(\mathbf{E_j} \,|\, \mathbf{F_j}) \geq 1 - \varepsilon$*, where* $\mathbf{E_j} = \{\hat{N}_t = N_t \text{ for all } t \in [t_j + \tau_{det} : t_{j+1} - 1]\}$, $\mathbf{F_j} = \{\hat{N}_{t_j-1} = N_{t_j-1}\}$*, and* $\varepsilon > 0$ *is arbitrary.*

Proof. **Need to find some way to get Candes Thm 1.3 in here and make the connection that $\hat{x}_{t,\mathbf{CSres}}$ in our notation is $x^\star$ in his. Also need to point out that the way we chose α, we have any $C_1\xi \leq C_1(S_{\mathbf{max}})\xi = \alpha$.**

To prove claims 1 and 2, we proceed by induction on the value of t.

Consider the base case, where $t = (t_0 - 1)$. Claim 1 follows from [1], Theorem 1.3 and assumptions (ii), (iii), and (iv) **(Not immediate – need to connect to Candes as above)**, and assumption (i) trivially proves claim 2.

Suppose now that claims 1 and 2 are both true for some time $(t-1)$. We show that the claims are true at time t.

First, we verify claim 1 at time t. Referring to Algorithm 1, we have

$$\begin{aligned} \beta_t &= x_t - \hat{x}_{t,\text{init}} \\ \hat{\beta}_t &= \arg\min_{\beta} \|\beta\|_1 \text{ subject to } \|y_{t,\text{res}} - A\beta\|_2 < \xi \\ \hat{x}_{t,\text{CSres}} &= \hat{x}_{t,\text{init}} + \hat{\beta}_t, \end{aligned}$$

where $\text{supp}(\hat{x}_{t,\text{init}}) = \hat{N}_{t-1}$.

By the induction hypothesis, $\hat{N}_{t-1} \subseteq N_{t-1}$, and by our model assumptions we have $N_{t-1} \subseteq N_t$. Therefore, $\text{supp}(\beta_t) \subseteq N_t \cup N_{t-1} = N_t$, so $|\text{supp}(\beta_t)| \leq |N_t| \leq S_{\text{max}}$. With this, we can apply [1], Theorem 1.3 to see that $\|\beta_t - \hat{\beta}_t\|_2 \leq \alpha$ **(AGAIN, need to make this connection)**. By the definitions of β_t and $\hat{x}_{t,\text{CSres}}$, we see that $\|\beta_t - \hat{\beta}_t\|_2 = \|x_t - \hat{x}_{t,\text{CSres}}\|_2$, so claim 1 follows.

Next, we verify claim 2 at time t. Suppose that $(x_t)_i = 0$ for some index i, so that $i \notin \text{supp}(x_t) = N_t$. Since $N_{t-1} \subseteq N_t$, we must also have $i \notin N_{t-1}$; by the induction hypothesis, this implies that $i \notin \hat{N}_{t-1}$.

Applying the result of claim 1,

$$|(\hat{x}_{t,\text{CSres}})_i| = |(x_t - \hat{x}_{t,\text{CSres}})_i| \leq \|x_t - \hat{x}_{t,\text{CSres}}\|_2 \leq \alpha.$$

Referring to Algorithm 1, $\hat{N}_t = \hat{N}_{t-1} \cup \{k : |(\hat{x}_{t,\text{CSres}})_k| > \alpha\}$. Since $i \notin \hat{N}_{t-1}$ and $|(\hat{x}_{t,\text{CSres}})_i| \leq \alpha$, it follows that $i \notin \hat{N}_t$. Thus if $i \notin N_t$, then $i \notin \hat{N}_t$; equivalently, if $i \in \hat{N}_t$, then $i \in N_t$. Therefore, $\hat{N}_t \subseteq N_t$, which proves claim 2 and completes our induction proof.

Now, we prove claim 3. Let $\Delta_t = N_t \setminus \hat{N}_{t-1}$ denote the set of indices of the true support at time t which have not been detected before time t. Fix $j \geq 0$ and suppose that $\mathbf{F_j}$ holds, that is, $\hat{N}_{t_j - 1} = N_{t_j - 1}$.

Since $\mathbf{F_j}$ holds, $\Delta_t \subseteq \Delta_{\text{add},t_j}$ for all $t \in [t_j : t_{j+1} - 1]$.

Let $i \in \Delta_t$ for some $t \in [t_j : t_{j+1} - 1]$ and suppose that $|(x_t)_i| > 2\alpha$. Applying the result from claim 1,

$$0 \leq |(x_t - \hat{x}_{t,\text{CSres}})_i| \leq \|x_t - \hat{x}_{t,\text{CSres}}\|_2 \leq \alpha < 2\alpha < |(x_t)_i|,$$

so that

$$\begin{aligned}
|(\hat{x}_{t,\text{CSres}})_i| &= \left|(x_t)_i - [(x_t)_i - (\hat{x}_{t,\text{CSres}})_i]\,\right| \\
&\geq \left||(x_t)_i| - |(x_t - \hat{x}_{t,\text{CSres}})_i|\right| \\
&= |(x_t)_i| - |(x_t - \hat{x}_{t,\text{CSres}})_i| \\
&> 2\alpha - \alpha \\
&= \alpha.
\end{aligned}$$

We see that if $|(x_t)_i| > 2\alpha$, then $|(\hat{x}_{t,\text{CSres}})_i| > \alpha$, so $i \in \hat{N}_t = \hat{N}_{t-1} \cup \{k : |(\hat{x}_{t,\text{CSres}})_k| > \alpha\}$.

If $|(x_t)_i| > 2\alpha$ for all $i \in \Delta_{\text{add},t_j}$, then $\Delta_t \subseteq \Delta_{\text{add},t_j} \subseteq \hat{N}_t$; in words, we will detect all "missing" indices at time t, so $\hat{N}_t = N_t$.

From the above discussion, we see that the event $\{|(x_t)_i| > 2\alpha \text{ for all } i \in \Delta_{\text{add},t_j}\}$ is contained within the event $\{|(x_t)_i| > 2\alpha \text{ for all } i \in \Delta_t \mid \mathbf{F_j}\}$, which in turn is contained within the event $\{\hat{N}_t = N_t \mid \mathbf{F_j}\}$.

All of the above is still kind of weak in places. It all makes sense in words and is true, but the math / set theory is kind of wonky.

Our model asserts that the entries $(x_t)_i$ for $i \in \Delta_{\text{add},t_j}$ are independent and identically distributed $\mathcal{N}\left(0, (t - t_j + 1)\sigma_{\text{sys}}^2\right)$ random variables. With this in mind, we see that

$$\begin{aligned}
\mathbf{Pr}\left(\hat{N}_t = N_t \,\middle|\, \mathbf{F_j}\right) &\geq \mathbf{Pr}\left(|(x_t)_i| > 2\alpha \text{ for all } i \in \Delta_t \,|\, \mathbf{F_j}\right) \\
&\geq \mathbf{Pr}\left(|(x_t)_i| > 2\alpha \text{ for all } i \in \Delta_{\text{add},t_j}\right) \\
&= \left[\mathbf{Pr}\left(|(x_t)_k| > 2\alpha\right)\right]^{S_{\text{add}}}, \quad k \in \Delta_{\text{add},t_j} \text{ arbitrary} \\
&= \left[2\mathcal{Q}\left(\frac{2\alpha}{\sigma_{\text{sys}}\sqrt{t - t_j + 1}}\right)\right]^{S_{\text{add}}}.
\end{aligned}$$

We examine the particular case where $t = t_j + \tau_{\text{det}}$. In this case,

$$\begin{aligned}
\mathbf{Pr}\left(\hat{N}_{t_j+\tau_{\text{det}}} = N_{t_j+\tau_{\text{det}}} \,\middle|\, \mathbf{F_j}\right) &\geq \left[2\mathcal{Q}\left(\frac{2\alpha}{\sigma_{\text{sys}}\sqrt{(t_j+\tau_{\text{det}}+1)-t_j}}\right)\right]^{S_{\text{add}}} \\
&= \left[2\mathcal{Q}\left(\frac{2\alpha}{\sigma_{\text{sys}}\sqrt{\tau_{\text{det}}+1}}\right)\right]^{S_{\text{add}}} \\
&\geq 1-\varepsilon,
\end{aligned}$$

where the final inequality is easily verified and follows from the ceiling in the definition of τ_{det} and the fact that $\mathcal{Q}$ is a decreasing function.

If $\hat{N}_t = N_t$ for $t = t_j + \tau_{\text{det}}$, then the model assumptions of no support deletions and no support additions until time t_{j+1}, in addition to the result of claim 2, imply that $\hat{N}_t = N_t$ for all $t \in [t_j + \tau_{\text{det}} : t_{j+1} - 1]$, which is exactly the event $\mathbf{E_j}$. Therefore, $\mathbf{Pr}\left(\mathbf{E_j} \,|\, \mathbf{F_j}\right) = \mathbf{Pr}\left(\hat{N}_{t_j+\tau_{\text{det}}} = N_{t_j+\tau_{\text{det}}} \,\middle|\, \mathbf{F_j}\right) \geq 1-\varepsilon$, which completes the proof. $\square$

Lemma 2. *Assume that $\{x_t\}$ and $\{y_t\}$ follow the signal model above, $\{t\}$ is a discrete set of sampling times, only additions to true support ($N_t \subseteq N_{t+1}$ for all t), etc.*

$\delta_{S_{max}}(A) < 1$, $\alpha_{del} = 0$.

Define the event $\mathbf{D} = \{\hat{N}_t = N_t = N_* \text{ for all } t \in [t_* : t_{**}]\}$, where N_* is some fixed index set.

At each time t, let $\hat{x}_t = \hat{x}_{t,KFCS}$ be the KFCS estimate of x_t (Algorithm 1) and let $\tilde{x}_t = \tilde{x}_{t,GAKF}$ be the GAKF estimate of x_t (Algorithm 2).

Then given any $\varepsilon > 0$ there exists some $t_{ms} \geq t_*$ such that for all $t \in [t_{ms} : t_{**}]$, we have $\mathbb{E}\left[\|\tilde{x}_t - \hat{x}_t\|_2^2 \,|\, \mathbf{D}\right] < \varepsilon$, i.e., $\hat{x}_t$ converges to $\tilde{x}_t$ in mean square.

Proof. Throughout, we assume that the event $\mathbf{D}$ occurs and $t \in [t_* : t_{**}]$.

Where possible, we consider variables and parameters only along the support set N_*, but to simplify notation will omit the subscript N_*. Thus, $\nu_t = (\nu_t)_{N_*}$, $A = A_{[1:n],N_*}$, $Q = Q_{N_*,N_*}$, $\hat{x}_t = (\hat{x}_t)_{N_*}$, $J_t = (J_t)_{N_*,N_*}$, $K_t = (K_t)_{N_*,[1:n]}$, $P_{t|t-1} = (P_{t|t-1})_{N_*,N_*}$, $P_t = (P_t)_{N_*,N_*}$, and analogously for $\tilde{x}_t$, $\tilde{J}_t$, $\tilde{K}_t$, $\tilde{P}_{t|t-1}$, and $\tilde{P}_t$.

Note, however, that y_t and w_t may be be supported on $[1 : n]$ and are thus not truncated when they appear; similarly, R is not truncated.

For $t > t_*$, both KFCS and GAKF run the same fixed-dimensional and fixed-parameter Kalman filter for $(x_t)_{N_*}$, but with different initial conditions. **Elaborate...**

$\downarrow$———— **moved to enhance the flow of the proof**

Suppose that $t \in [t_* : t_{**}]$. We see that

$$
\begin{aligned}
P_{t+1|t} &= P_t + Q \\
&= (I - K_t A)P_{t|t-1} + Q \\
&= P_{t|t-1} + Q - P_{t|t-1}A'(AP_{t|t-1}A' + R)^{-1}AP_{t|t-1},
\end{aligned}
$$

which is a discrete algebraic Riccati recursion (2) with $F = I$, $G = I$, $Q = \sigma^2_{\text{sys}} I_{|N_*|\times|N_*|} \succ 0$, $R = \sigma^2_{\text{obs}} I_{n\times n} \succ 0$, and $S = 0$. **Verify Q, R – goes back to the algorithm issues.** Note that Q is constant on $[t_* : t_{**}]$ since we assume that $\mathbf{D}$ occurs.

Since $|N_*| \leq S_{\max}$ and $\delta_{S_{\max}} < 1$, $A = (A_{[1:n],N_*})$ is full rank. Therefore, using the results from Section 5.1, $\{I, A\}$ is detectable. Further, since $Q = \sigma^2_{\text{sys}} I$ is full rank, $\{I, Q\}$ is stabilizable.

Referring to the algorithm **(which one?)**, we see that $P_0 = P_{t_0-1} = \sigma^2_{\text{sys},0} I \succ 0$. **is this even true? Need to get the algorithms and model set up correctly. I think we want the initial step to be $P_{t_0|t_0-1} = \sigma^2_{\text{sys},0} I + Q_{t_0} \succeq 0$, but the two algorithms disagree on what Q_{t_0} is.**

Therefore, by [2], Lemma 8.7.3, the DARR converges to a positive semi-definite matrix $P_\star$ which satisfies the corresponding DARE. This implies that $K_t \to K_\star = P_\star A'(AP_\star A' + R)^{-1}$

and $J_t \to J_\star = (I - K_\star A)$. Further, by [2], Theorem 7.5.1.b, $\rho(J_\star) = \rho(I - K_\star A) < 1$.

Since GAKF and KFCS run the same Kalman filter, these results also apply to the GAKF iterates, i.e. $\tilde{P}_{t|t-1} \to P_\star$, $\tilde{K}_t \to K_\star$, and $\tilde{J}_t \to J_\star$.

Define $\rho = \rho(J_\star)$ and let $\varepsilon_0 = (1-\rho)/2$. A standard result from linear algebra states that there exists a matrix norm $\|\cdot\|_\rho$ such that $\|J_\star\|_\rho \leq \rho + \varepsilon_0 = (1+\rho)/2 < 1$. Further, by the equivalence of matrix norms on a finite-dimensional space, there exists some constant $c_{2,\rho}$ such that $\|M\|_2 \leq c_{2,\rho}\|M\|_\rho$ for any matrix M.

Since $\tilde{J}_t \to J_\star$, there exists some $t_c \geq t_0$ such that for all $t \geq t_c$, $\|\tilde{J}_t\|_2 < \|J_\star\|_2 + 1$. Therefore, for any $t \geq t_0$, we have $\|\tilde{J}_t\|_2 \leq \max\{\|\tilde{J}_{t_0}\|_2, \|\tilde{J}_{t_0+1}\|_2, \ldots, \|\tilde{J}_{t_c-1}\|_2, \|J_\star\|_2 + 1\}$, i.e. there exists some value $\tilde{B}_J > 0$ such that $\|\tilde{J}_t\|_2 < \tilde{B}_J$ for all t. Since $\|\tilde{J}_t\|_2 < \infty$ for all t and $\|J_\star\|_2 < \infty$, we must also have $\tilde{B}_J < \infty$.

By similar arguments, since J_t converges to $J_\star$ and $P_{t|t-1}$ and $\tilde{P}_{t|t-1}$ converge to $P_\star$, there exist some $0 < B_J, B_P, \tilde{B}_P < \infty$ such that $\|J_t\|_2 < B_J$, $\|P_{t|t-1}\|_2 < B_P$, and $\|\tilde{P}_{t|t-1}\|_2 < \tilde{B}_P$ for all t.

Let $\varepsilon > 0$ be arbitrary.

The convergence results above and standard analysis techniques can be used to show that there exists some $t_\varepsilon > t_*$ such that for all $t \geq t_\varepsilon$, all of the following conditions hold:

- $\|K_t - \tilde{K}_t\|_2 < \varepsilon$;
- $\|J_t - \tilde{J}_t\|_2 < \varepsilon$; and
- $\|J_t\|_\rho \leq \|J_\star\|_\rho + (1-\rho)/4$.

$\star$ **Why do we care if t_ε does not depend on y?**

Problem: NV proof says $\hat{x}_{t_*}$ is independent of y_{t_*}, but by definition it's not. AB draft: $t_* - 1$. So I agree that we're independent of $y_1 \ldots y_{t_*-1}$, but we are dependent on $y_{t_*} \ldots y_t$ because $\hat{x}_t = J_t\hat{x}_{t-1} + K_t y_t$ for $t > t_*$. **All of this independence stuff needs to be very carefully worked and verified; also, why do we care? I think $\hat{x}$ is useless here, it does not affect the choice of t_ε.**

Attempted fix: Examining the algorithms, we see that K_t, $\tilde{K}_t$, J_t, $\tilde{J}_t$, $P_{t|t-1}$ and $\tilde{P}_{t|t-1}$ do not depend on $\{y_k\}$, hence, neither do $K_\star$, $J_\star$, and $P_\star$. It follows that t_ε also does not depend on $\{y_k\}$.

↑——— /moved

Let $\hat{e}_t = x_t - \hat{x}_t$ and $\tilde{e}_t = x_t - \tilde{x}_t$. Define $\text{diff}_t = \hat{e}_t - \tilde{e}_t$ and notice that $\text{diff}_t = \tilde{x}_t - \hat{x}_t$.

Let $t > t_\varepsilon > t_*$. By Algorithm 1 and the model, we see that

$$\begin{aligned}
\hat{e}_t &= x_t - \hat{x}_t \\
&= (x_{t-1} + \nu_t) - (J_t\hat{x}_{t-1} + K_t y_t) \\
&= x_{t-1} + \nu_t - J_t\hat{x}_{t-1} - K_t(Ax_t + w_t) \\
&= x_{t-1} + \nu_t - J_t\hat{x}_{t-1} - K_t A(x_{t-1} + \nu_t) - K_t w_t \\
&= (I - K_t A)x_{t-1} - J_t\hat{x}_{t-1} + (I - K_t A)\nu_t - K_t w_t \\
&= J_t(x_{t-1} - \hat{x}_{t-1}) + J_t\nu_t - K_t w_t \\
&= J_t\hat{e}_{t-1} + J_t\nu_t - K_t w_t.
\end{aligned}$$

Similarly, using Algorithm 2 and the model, we can verify that

$$\tilde{e}_t = \tilde{J}_t\tilde{e}_{t-1} + \tilde{J}_t\nu_t - \tilde{K}_t w_t.$$

Combining these results yields

$$\mathrm{diff}_t = J_t\mathrm{diff}_{t-1} + (J_t - \tilde{J}_t)(\tilde{e}_{t-1} + \nu_t) + (\tilde{K}_t - K_t)w_t.$$

Let

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

so that $\mathrm{diff}_t = J_t\mathrm{diff}_{t-1} + u_t$. Recursively applying this identity, we see that

$$\begin{aligned}
\mathrm{diff}_t &= J_t\mathrm{diff}_{t-1} + u_t \\
&= J_t\left(J_{t-1}\mathrm{diff}_{t-2} + u_{t-1}\right) + u_t \\
&= J_t J_{t-1}\mathrm{diff}_{t-2} + J_t u_{t-1} + u_t \\
&= J_t J_{t-1}\left(J_{t-2}\mathrm{diff}_{t-3} + u_{t-2}\right) + J_t u_{t-1} + u_t \\
&= J_t J_{t-1} J_{t-2}\mathrm{diff}_{t-3} + J_t J_{t-1} u_{t-2} + J_t u_{t-1} + u_t \\
&\vdots \\
&= J_t J_{t-1} \cdots J_{t_\varepsilon+1}\mathrm{diff}_{t_\varepsilon} + J_t J_{t-1} \cdots J_{t_\varepsilon+2} u_{t_\varepsilon+1} + \ldots + J_t u_{t-1} + u_t.
\end{aligned}$$

If we define

$$M_k^t = \begin{cases} J_t J_{t-1} \cdots J_{k+1} J_k & k \le t \\ I & k > t \end{cases}$$

then we can more compactly write

$$\mathrm{diff}_t = M_{t_\varepsilon+1}^t \mathrm{diff}_{t_\varepsilon} + \sum_{k=t_\varepsilon+1}^{t} M_{k+1}^t u_k.$$

Therefore, applying the triangle and Cauchy-Schwarz inequalities for expectation and noting that the matrices $\{M_k^t\}$ are deterministic,

$$
\begin{aligned}
\mathbb{E}\left[\|\mathrm{diff}_t\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} &= \mathbb{E}\left[\left\|M_{t_\varepsilon+1}^t \mathrm{diff}_{t_\varepsilon} + \sum_{k=t_\varepsilon+1}^{t} M_{k+1}^t u_k\right\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} \\
&\leq \mathbb{E}\left[\left\|M_{t_\varepsilon+1}^t \mathrm{diff}_{t_\varepsilon}\right\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} + \mathbb{E}\left[\|u_t\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} + \sum_{k=t_\varepsilon+1}^{t-1} \mathbb{E}\left[\left\|M_{k+1}^t u_k\right\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} \\
&\leq \mathbb{E}\left[\left\|M_{t_\varepsilon+1}^t\right\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} \mathbb{E}\left[\|\mathrm{diff}_{t_\varepsilon}\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} + \\
&\qquad \mathbb{E}\left[\|u_t\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} + \sum_{k=t_\varepsilon+1}^{t-1} \mathbb{E}\left[\left\|M_{k+1}^t\right\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} \mathbb{E}\left[\|u_k\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} \\
\mathbb{E}\left[\|\mathrm{diff}_t\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} &\leq \|M_{t_\varepsilon+1}^t\|_2 \, \mathbb{E}\left[\|\mathrm{diff}_{t_\varepsilon}\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} + \\
&\qquad \left(1 + \sum_{\ell=t_\varepsilon+2}^{t} \|M_\ell^t\|_2\right) \max_{\tau \in [t_\varepsilon+1:t]} \left\{\mathbb{E}\left[\|u_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}}\right\}. \qquad (3)
\end{aligned}
$$

Recall that, for $k \geq t_\varepsilon$, we have

$$
\|J_k\|_\rho \leq \|J_\star\|_\rho + (1-\rho)/4 \leq (1+\rho)/2 + (1-\rho)/4 = (3+\rho)/4 < 1.
$$

Let $a = (3+\rho)/4$. Then for $t_\varepsilon \leq k \leq t$,

$$
\begin{aligned}
\|M_k^t\|_2 &\leq c_{2,\rho}\|M_k^t\|_\rho \\
&= \|J_t J_{t-1} \cdots J_k\|_\rho \\
&\leq \|J_t\|_\rho \|J_{t-1}\|_\rho \cdots \|J_k\|_\rho \\
\|M_k^t\|_2 &\leq c_{2,\rho} a^{t-k+1}. \qquad (4)
\end{aligned}
$$

With this, we see that

$$
\begin{aligned}
\left(1 + \sum_{\ell=t_\varepsilon+2}^{t} \|M_\ell^t\|_2\right) &\leq \left(1 + \sum_{\ell=t_\varepsilon+2}^{t} c_{2,\rho} a^{t-\ell+1}\right) \\
&\leq \max\{1, c_{2,\rho}\} \cdot \sum_{\ell=0}^{\infty} a^\ell \\
\left(1 + \sum_{\ell=t_\varepsilon+2}^{t} \|M_\ell^t\|_2\right) &\leq \max\{1, c_{2,\rho}\} \cdot \frac{1}{1-a}\ . \qquad (5)
\end{aligned}
$$

Let $\tau \in [t_\varepsilon+1 : t]$ be arbitrary. Since $\tau > t_\varepsilon$, we have $\|\tilde{K}_\tau - K_\tau\|_2 < \varepsilon$ and $\|\tilde{J}_\tau - J_\tau\|_2 < \varepsilon$.

Consider

$$\begin{aligned}
\mathbb{E}\left[\|u_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} &= \mathbb{E}\left[\left\|(J_\tau - \tilde{J}_\tau)(\tilde{e}_{\tau-1} + \nu_\tau) + (\tilde{K}_\tau - K_\tau)w_\tau\right\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} \\
&\leq \|J_\tau - \tilde{J}_\tau\|_2 \, \mathbb{E}\left[\|\tilde{e}_{\tau-1} + \nu_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} + \|\tilde{K}_\tau - K_\tau\|_2 \, \mathbb{E}\left[\|w_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} \\
&< \varepsilon \cdot \mathbb{E}\left[\|\tilde{e}_{\tau-1} + \nu_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} + \varepsilon \cdot \mathbb{E}\left[\|w_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} \\
\mathbb{E}\left[\|u_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} &\leq \varepsilon \left(\mathbb{E}\left[\|\tilde{e}_{\tau-1}\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} + \mathbb{E}\left[\|\nu_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} + \mathbb{E}\left[\|w_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}}\right), \qquad (6)
\end{aligned}$$

where we have used the triangle and Cauchy-Schwarz inequalities for expectation.

By the properties of the Kalman filter, for any k, we have

$$\begin{aligned}
\tilde{P}_k &= \mathbb{E}\left[\left(\tilde{x}_k - \mathbb{E}\left[\tilde{x}_k \,|\, y_1, y_2, \ldots, y_k\right]\right)\left(\tilde{x}_k - \mathbb{E}\left[\tilde{x}_k \,|\, y_1, y_2, \ldots, y_k\right]\right)' \,\middle|\, y_1, y_2, \ldots, y_k\right] \\
&= \mathbb{E}\left[\left(\tilde{x}_k - x_k\right)\left(\tilde{x}_k - x_k\right)' \,\middle|\, y_1, y_2, \ldots, y_k\right] \\
&= \mathbb{E}\left[\tilde{e}_k \tilde{e}_k' \,|\, y_1, y_2, \ldots, y_k\right] \\
&= \mathbb{E}\left[\tilde{e}_k \tilde{e}_k'\right],
\end{aligned}$$

where the independence on the last line follows because $\tilde{P}_k$ has no dependence on any of the $\{y_i\}$, a well-known property of the Kalman filter (and consequence of the algorithm). Therefore,

$$\begin{aligned}
\mathbb{E}\left[\|\tilde{e}_k\|_2^2 \,\middle|\, \mathbf{D}\right] &= \operatorname{tr}\left(\mathbb{E}\left[\tilde{e}_k' \tilde{e}_k \,|\, \mathbf{D}\right]\right) \\
&= \mathbb{E}\left[\operatorname{tr}(\tilde{e}_k' \tilde{e}_k) \,|\, \mathbf{D}\right] \\
&= \mathbb{E}\left[\operatorname{tr}(\tilde{e}_k \tilde{e}_k') \,|\, \mathbf{D}\right] \\
&= \operatorname{tr}\left(\mathbb{E}\left[\tilde{e}_k \tilde{e}_k' \,|\, \mathbf{D}\right]\right) \\
&= \operatorname{tr}\left(\mathbb{E}\left[\tilde{e}_k \tilde{e}_k'\right]\right) \\
&= \operatorname{tr}\left(\tilde{P}_k\right).
\end{aligned}$$

Here, we used the fact that the occurrence of $\mathbf{D}$ is independent of the value of $\tilde{P}_k = \mathbb{E}\left[e_k e_k'\right]$. **Make sure this is legitimate.**

We see that

$$\|\tilde{P}_k\|_2 = \|\tilde{J}_k \tilde{P}_{k|k-1}\|_2 \leq \|\tilde{J}_k\|_2 \|\tilde{P}_{k|k-1}\|_2 < \tilde{B}_J \tilde{B}_P < \infty,$$

where we recall that $\|\tilde{P}_{k|k-1}\|_2 < \tilde{B}_P$ and $\|\tilde{J}_k\|_2 < \tilde{B}_J$. Since $\tilde{P}_k$ is Hermitian, $\|\tilde{P}_k\|_2 = \lambda_{\max}(\tilde{P}_k)$. Therefore,

$$\operatorname{tr}(\tilde{P}_k) = \sum_i \lambda_i(\tilde{P}_k) \leq |N_*| \lambda_{\max}(\tilde{P}_k) = |N_*| \|\tilde{P}_k\|_2 < |N_*| \tilde{B}_J \tilde{B}_P < \infty,$$

so there exists some $0 < \tilde{B} < \infty$ such that $\operatorname{tr}(\tilde{P}_k) < \tilde{B}$ for all k.

Therefore,

$$\mathbb{E}\left[\|\tilde{e}_{\tau-1}\|_2^2 \,\middle|\, \mathbf{D}\right] = \operatorname{tr}\left(\tilde{P}_{\tau-1}\right) < \tilde{B}.$$

Since $\mathbf{D}$ occurs, ν_τ is supported on N_*, so the covariance of $\nu_\tau = (\nu_\tau)_{N_*}$ is $\mathbb{E}[\nu_\tau \nu_\tau'] = \mathbb{E}\left[\nu_\tau \nu_\tau' \,|\, \mathbf{D}\right] = \sigma_{\text{sys}}^2 I_{|N_*|\times|N_*|}$. **VERIFY this claim.** Therefore,

$$\begin{aligned}
\mathbb{E}\left[\|\nu_\tau\|_2^2 \,\middle|\, \mathbf{D}\right] &= \operatorname{tr}\left(\mathbb{E}\left[\nu_\tau' \nu_\tau \,|\, \mathbf{D}\right]\right) \\
&= \mathbb{E}\left[\operatorname{tr}(\nu_\tau' \nu_\tau) \,|\, \mathbf{D}\right] \\
&= \mathbb{E}\left[\operatorname{tr}(\nu_\tau \nu_\tau') \,|\, \mathbf{D}\right] \\
&= \operatorname{tr}\left(\mathbb{E}\left[\nu_\tau \nu_\tau' \,|\, \mathbf{D}\right]\right) \\
&= \operatorname{tr}\left(\sigma_{\text{sys}}^2 I_{|N_*|\times|N_*|}\right) \\
&= |N_*|\sigma_{\text{sys}}^2.
\end{aligned}$$

A similar computation proves that $\mathbb{E}\left[\|w_\tau\|_2^2 \,\middle|\, \mathbf{D}\right] = n\sigma_{\text{obs}}^2$.

With (6), these results show that

$$\mathbb{E}\left[\|u_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} < \varepsilon\left(\sqrt{\tilde{B}} + \sqrt{|N_*|\sigma_{\text{sys}}^2} + \sqrt{n\sigma_{\text{obs}}^2}\right).$$

Since $\tau \in [t_\varepsilon + 1 : t]$ was arbitrary, we conclude that

$$\max_{\tau\in[t_\varepsilon+1:t]}\left\{\mathbb{E}\left[\|u_\tau\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}}\right\} < \varepsilon\left(\sqrt{\tilde{B}} + \sqrt{|N_*|\sigma_{\text{sys}}^2} + \sqrt{n\sigma_{\text{obs}}^2}\right). \tag{7}$$

We have seen that $\mathbb{E}\left[\|\tilde{e}_k\|_2^2 \,\middle|\, \mathbf{D}\right] < \operatorname{tr}(\tilde{P}_k) < \tilde{B}$ for some $\tilde{B}$ and all k; by similar work, we can conclude that there exists some B such that $\mathbb{E}\left[\|e_k\|_2^2 \,\middle|\, \mathbf{D}\right] < \operatorname{tr}(P_k) < B$ for all k. Therefore, by the triangle inequality for expectation,

$$\begin{aligned}
\mathbb{E}\left[\|\text{diff}_{t_\varepsilon}\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} &= \mathbb{E}\left[\|e_{t_\varepsilon} - \tilde{e}_{t_\varepsilon}\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} \\
&\le \mathbb{E}\left[\|e_{t_\varepsilon}\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} + \mathbb{E}\left[\|\tilde{e}_{t_\varepsilon}\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} \\
\mathbb{E}\left[\|\text{diff}_{t_\varepsilon}\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} &< B + \tilde{B}.
\end{aligned} \tag{8}$$

Combining (3) with (4), (5), (7), and (8), we see that

$$\mathbb{E}\left[\|\text{diff}_t\|_2^2 \,\middle|\, \mathbf{D}\right]^{\frac{1}{2}} < c_{2,\rho} a^{t-t_\varepsilon}(B + \tilde{B}) + C\varepsilon,$$

where $C = \max\{1, c_{2,\rho}\} \cdot \frac{1}{1-a} \cdot \left(\sqrt{\tilde{B}} + \sqrt{|N_*|\sigma_{\text{sys}}^2} + \sqrt{n\sigma_{\text{obs}}^2}\right)$.

If

$$t_{\text{ms}} = t_\varepsilon + \left\lceil \log_a \left(\frac{C\varepsilon}{c_{2,\rho}(B+\tilde{B})} \right) \right\rceil,$$

then we see that for all $t \geq t_{\text{ms}}$,

$$\mathbb{E}\left[\|\tilde{x}_t - \hat{x}_t\|_2^2 \,\middle|\, \mathbf{D} \right]^{\frac{1}{2}} = \mathbb{E}\left[\|\text{diff}_t\|_2^2 \,\middle|\, \mathbf{D} \right]^{\frac{1}{2}} < 2C\varepsilon,$$

and since C is constant and ε is arbitrary we have obtained our desired result. $\square$

Corollary 1. *Assume that the conditions of Lemma 2 hold.*

Then given any ε *and* ε_{err} *there exists some* $\tau_{KF} = \tau_{KF}(\varepsilon, \varepsilon_{err}, N_*)$ *such that for all* $t \in [t_* + \tau_{KF} : t_{**}]$, *we have* $\boldsymbol{Pr}\left(\|\tilde{x}_t - \hat{x}_t\|_2^2 \leq \varepsilon_{err} \,\middle|\, \mathbf{D}\right) > 1 - \varepsilon$. *Note that if* $t_* + \tau_{KF} > t_{**}$, *then this interval is empty and the result is vacuously true.*

Proof. Let $\varepsilon > 0$ and $\varepsilon_{\mathrm{err}} > 0$ be given and let $\tilde{\varepsilon} \leq \varepsilon \cdot \varepsilon_{\mathrm{err}}$. By Lemma 2, there exists some $t_{\mathrm{ms}} = t_{\mathrm{ms}}(\tilde{\varepsilon}, N_*)$ such that for all $t \geq t_{\mathrm{ms}}$,

$$\mathbb{E}\left[\|\tilde{x}_t - \hat{x}_t\|_2^2 \,\middle|\, \mathbf{D}\right] < \tilde{\varepsilon} \leq \varepsilon \cdot \varepsilon_{\mathrm{err}}.$$

Let $t \geq t_{\mathrm{ms}}$. By Markov's inequality,

$$\mathbf{Pr}\left(\|\tilde{x}_t - \hat{x}_t\|_2^2 > \varepsilon_{\mathrm{err}} \,\middle|\, \mathbf{D}\right) \leq \frac{\mathbb{E}\left[\|\tilde{x}_t - \hat{x}_t\|_2^2 \mid \mathbf{D}\right]}{\varepsilon_{\mathrm{err}}} < \frac{\tilde{\varepsilon}}{\varepsilon_{\mathrm{err}}} \leq \varepsilon.$$

Define $\tau_{\mathrm{KF}} = t_{\mathrm{ms}} - t_*$. Since t_{ms} is a function of $\tilde{\varepsilon}$, which is itself a function of ε and $\varepsilon_{\mathrm{err}}$, we have $\tau_{\mathrm{KF}} = \tau_{\mathrm{KF}}(\varepsilon, \varepsilon_{\mathrm{err}}, N_*)$, and for all $t \geq t_{\mathrm{ms}} = t_* + \tau_{\mathrm{KF}}$,

$$\mathbf{Pr}\left(\|\tilde{x}_t - \hat{x}_t\|_2^2 < \varepsilon_{\mathrm{err}} \,\middle|\, \mathbf{D}\right) > 1 - \varepsilon,$$

which is our desired result. □

Theorem 1. *Assume that the conditions of Lemma 1 and Lemma 2 hold.* ***Recall def of E_j and F_j?***

Let $\varepsilon > 0$, $\varepsilon_{err} > 0$ be given.

Let $\tau_{det} = \tau_{det}(\alpha, \varepsilon)$ be as in Lemma 1.

Choose $d > \tau_{det} + \max_j\{\tau_{KF}(\varepsilon, \varepsilon_{err}, N_{t_j})\}$.

1) *Given any $j \in [0 : K-1]$, $\boldsymbol{Pr}(\|\tilde{x}_t - \hat{x}_t\|_2^2 \leq \varepsilon_{err}) > (1-\varepsilon)^{j+2}$ for all $t \in [t_j + \tau_{det} + \tau_{KF}(\varepsilon, \varepsilon_{err}, N_{t_j}) : t_{j+1} - 1]$.*

2) *$\boldsymbol{Pr}(|\Delta_t| \leq S_{add}$ and $|\Delta_e| = 0$ for all $t \geq t_0) \geq (1-\varepsilon)^K$*

3) *$\boldsymbol{Pr}($For all $j \in [0 : K-1]$, $|\Delta_t| = 0$ and $|\Delta_e| = 0$ for all $t \in [t_j + \tau_{det} : t_{j+1} - 1])$ $\geq (1-\varepsilon)^K$*

Proof. We first show by induction that $\mathbf{Pr}(\mathbf{E_j}) \geq (1-\varepsilon)^{j+1}$ for all $j \geq 0$.

Consider the base case, where $j = 0$. By assumption, $\hat{N}_{t_0-1} = N_{t_0-1}$, so $\mathbf{F_0}$ occurs. We have

$$\mathbf{Pr}(\mathbf{E_0}) = \mathbf{Pr}\left(\mathbf{E_0} \,|\, \mathbf{F_0}\right) \geq 1 - \varepsilon$$

by Lemma 1, which proves the base case.

Now assume that the claim is true for $j = (k-1)$ for some $k \geq 1$, that is, $\mathbf{Pr}(\mathbf{E_{k-1}}) \geq (1-\varepsilon)^k$. Consider

$$\begin{aligned}
\mathbf{Pr}(\mathbf{E_k}) &= \mathbf{Pr}(\mathbf{E_k} \cap \mathbf{E_{k-1}}) + \mathbf{Pr}(\mathbf{E_k} \cap (\mathbf{E_{k-1}})^c) \\
&\geq \mathbf{Pr}(\mathbf{E_k} \cap \mathbf{E_{k-1}}) \\
&= \mathbf{Pr}\left(\mathbf{E_k} \,|\, \mathbf{E_{k-1}}\right) \mathbf{Pr}(\mathbf{E_{k-1}}) \\
&= \mathbf{Pr}\left(\mathbf{E_k} \,|\, \mathbf{F_k}\right) \mathbf{Pr}(\mathbf{E_{k-1}}) \qquad \textbf{WHY is this true?} \\
&\geq (1-\varepsilon)(1-\varepsilon)^k \\
&= (1-\varepsilon)^{k+1},
\end{aligned}$$

where we applied Lemma 1 to conclude that $\mathbf{Pr}\left(\mathbf{E_k} \,|\, \mathbf{F_k}\right) \geq 1 - \varepsilon$. Therefore, by the principle of mathematical induction, we conclude that

$$\mathbf{Pr}(\mathbf{E_j}) \geq (1-\varepsilon)^{j+1}$$

for all $j \geq 0$.

Fix $j \in [0 : K-1]$.

Choosing $t_* = t_j + \tau_{\text{det}}$ and $t_{**} = t_{j+1} - 1$, the event $\mathbf{D} = \{\hat{N}_t = N_t = N_*$ for all $t \in [t_* : t_{**}]\}$ is identically the event $\mathbf{E_j} = \{\hat{N}_t = N_t$ for all $t \in [t_j + \tau_{\text{det}} : t_{j+1} - 1]\}$. Corollary 1 thus yields

$$\mathbf{Pr}\left(\|\tilde{x}_t - \hat{x}_t\|_2^2 \leq \varepsilon_{\text{err}} \,\middle|\, \mathbf{E_j}\right) > 1 - \varepsilon$$

for all $t \in [t_j + \tau_{\text{det}} + \tau_{\text{KF}}(\varepsilon, \varepsilon_{\text{err}}, N_{t_j}) : t_{j+1} - 1]$.

Note that since $d > \tau_{\text{det}} + \tau_{\text{KF}}(\varepsilon, \varepsilon_{\text{err}}, N_{t_j})$ for all j, the interval $[t_j + \tau_{\text{det}} + \tau_{\text{KF}}(\varepsilon, \varepsilon_{\text{err}}, N_{t_j}) : t_{j+1} - 1]$ is nonempty.

For any $t \in [t_j + \tau_{\text{det}} + \tau_{\text{KF}} : t_{j+1} - 1]$, we see that

$$\begin{aligned}
\mathbf{Pr}\left(\|\tilde{x}_t - \hat{x}_t\|_2^2 \leq \varepsilon_{\text{err}}\right) &= \mathbf{Pr}\left(\{\|\tilde{x}_t - \hat{x}_t\|_2^2 \leq \varepsilon_{\text{err}}\} \cap \mathbf{E_j}\right) + \mathbf{Pr}\left(\{\|\tilde{x}_t - \hat{x}_t\|_2^2 \leq \varepsilon_{\text{err}}\} \cap (\mathbf{E_j})^c\right) \\
&\geq \mathbf{Pr}\left(\{\|\tilde{x}_t - \hat{x}_t\|_2^2 \leq \varepsilon_{\text{err}}\} \cap \mathbf{E_j}\right) \\
&= \mathbf{Pr}\left(\|\tilde{x}_t - \hat{x}_t\|_2^2 \leq \varepsilon_{\text{err}} \,\middle|\, \mathbf{E_j}\right) \mathbf{Pr}(\mathbf{E_j}) \\
&> (1-\varepsilon)(1-\varepsilon)^{j+1} \\
&= (1-\varepsilon)^{j+2},
\end{aligned}$$

which verifies **the first claim.**

I think that the third claim's probability equals the one below. Either way, we need this.

$$\begin{aligned}
\mathbf{Pr}(\mathbf{E_0} \cap \mathbf{E_1} \cap \ldots \cap \mathbf{E_{K-1}}) &= \mathbf{Pr}(\mathbf{E_0})\mathbf{Pr}\left(\mathbf{E_1} \,|\, \mathbf{E_0}\right)\mathbf{Pr}\left(\mathbf{E_2} \,|\, \mathbf{E_0} \cap \mathbf{E_1}\right) \cdots \mathbf{Pr}\left(\mathbf{E_{K-1}} \,\middle|\, \bigcap_{j=0}^{K-1} \mathbf{E_j}\right) \\
&= \mathbf{Pr}(\mathbf{E_0})\mathbf{Pr}\left(\mathbf{E_1} \,|\, \mathbf{E_0}\right)\mathbf{Pr}\left(\mathbf{E_2} \,|\, \mathbf{E_1}\right) \cdots \mathbf{Pr}\left(\mathbf{E_{K-1}} \,|\, \mathbf{E_{K-2}}\right) \\
&= \mathbf{Pr}(\mathbf{E_0})\mathbf{Pr}\left(\mathbf{E_1} \,|\, \mathbf{F_1}\right)\mathbf{Pr}\left(\mathbf{E_2} \,|\, \mathbf{F_2}\right) \cdots \mathbf{Pr}\left(\mathbf{E_{K-1}} \,|\, \mathbf{F_{K-1}}\right) \\
&\geq (1-\varepsilon)^K
\end{aligned}$$

Stuff to verify:

$\mathbf{Pr}\left(E_j \,|\, E_{j-1}\right) = \mathbf{Pr}\left(E_j \,|\, F_j\right)$: if E_{j-1} happens, then F_j definitely happens, but not seeing why these are equal yet.

Markov property used on $\{E_j\}$: justification

Second claim: the event seems to be a superset of the event $E_0 \cap \ldots \cap E_{K-1}$, so obviously the probability is bigger than $(1-\varepsilon)^K$. □

References

[1] E. J. Candès, *The restricted isometry property and its implications for compressed sensing*, Comptes Rendus Mathematique, Volume 346, Issues 9–10, May 2008, Pages 589-592, ISSN 1631-073X, http://dx.doi.org/10.1016/j.crma.2008.03.014. **Verify this citation, the journal name may be wrong**

[2] B. Hassibi, *Indefinite Metric Spaces in Estimation, Control and Adaptive Filtering*, Ph.D. Dissertation, Stanford University, August, 1996. Available `http://www.ee2.caltech.edu/Faculty/babak/pubs/thesis.html`.

[3] T. Kailath, A.H. Sayed and B. Hassibi, *Linear Estimation*, Prentice-Hall, 2000. ISBN 0130224642.

[4] F. M. Callier and C. A. Desoer, *Linear System Theory*, Springer-Verlag, 1991. **UNUSED**