KFCS Theory

Animesh Biswas, Kevin Palmowski

Last compiled July 24, 2014

1 Model

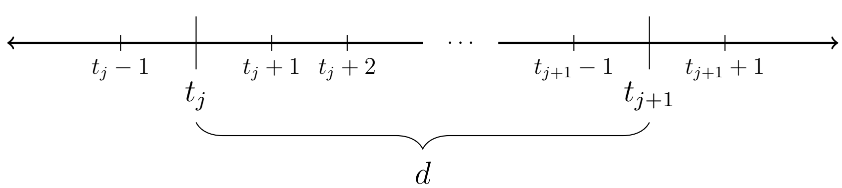
At each time $t \geq t_0$, we have

$$\begin{aligned} y_t &= Ax_t + w_t \\ x_{t+1} &= x_t + \nu_{t+1} \qquad \textbf{Can we use } \nu_t\textbf{?} \end{aligned}$$

Here, $\mathbb{E}[w_t] = \mathbf{0}$, $\mathrm{cov}(w_t) = \mathbb{E}[w_t w_t'] = \sigma^2_{\text{obs}} I_n$, iid and independent of x_t; $x_{t_0} = x_0 \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\text{sys},0} I_{N_0})$; and $\nu_t \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\text{sys}} I_{N_t})$ iid.

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

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



At the addition times $t_j = t_0 + jd$ for some t_0, 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

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 $\|y_{t_0} - Ax\|_2 < \xi$
$\hat{N}_{t_0} = \{j : |(\hat{x}_{t_0,\text{init}})_j| > \alpha\}$
$P_{t_0-1} = \sigma^2_{\text{sys},0} I_{\hat{N}_{t_0}}$
$Q_{t_0} = 0$
$\hat{x}_{t_0-1} = \mathbf{0}$

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

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

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

    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

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?

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**
      $\Delta_A = N_t \setminus N_{t-1}$
      $\left(\tilde{P}_{t-1}\right)_{\Delta_A,\Delta_A} = \sigma^2_{\textbf{sys}} I_{|\Delta_A|}$
    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

4 Candes RIP – C_1 Computation for α

We need to add this as a theorem or something – cite [1] Thm 1.3 and explicitly give the value of C_1 and the commentary below.

THEOREM / RESULT: [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 Kailath, Sayed, Hassibi – Linear Estimation

5.1 Appendix C, Section 3

Definitions and concepts from system theory.

$$x_{i+1} = F_i x_i + G_i u_i$$
$$y_i = H_i x_i + K_i u_i$$

Shorthand (F_i, G_i, H_i, K_i). **PROBLEM:** y is tied also to u here instead of some other variable...

Stable: F is stable if $|\lambda_i| < 1$ for all $\lambda_i \in \sigma_p(F)$, equivalently, $\rho(F) < 1$.

Controllable: $\{F, G\}$ is controllable if and only if the controllability matrix $\mathcal{C} = [G, FG, F^2G, \ldots, F^{n-1}G]$ is full rank n.

Unit-circle Controllable: $\{F, G\}$ is unit-circle controllable if $\text{rank}(\lambda I - FG) = n$ at all unit-circle eigenvalues λ of F.

P. 502 – unit-circle controllable $\equiv$ there exists K with $F - GK$ has no unit-circle eigenvalues. Is this even true? Nontrivial if so.

Stabilizable: $\{F, G\}$ is stabilizable if and only if $\text{rank}(\lambda I - FG) = n$ for all $\lambda \in \sigma_p(F)$ with $|\lambda| \geq 1$.

Observable: $\{F, H\}$ is observable if and only if $\{F^*, H^*\}$ is controllable, i.e. $\mathcal{C} = [H^*, F^*H^*, (F^*)^2H^*, \ldots, (F^*)^{n-1}H^*]$ is full rank n.

Detectable: $\{F, H\}$ is detectable if and only if $\{F^*, H^*\}$ is stabilizable, i.e. $\text{rank}(\lambda I - F^*H^*) = n$ for all $\lambda \in \sigma_p(F)$ with $|\lambda| \geq 1$.

5.2 Chapter 8, section 3

Solutions of a DARE

Section 8.1 – Time-Invariant State-Space Model (p.266)

$$x_{i+1} = Fx_i + Gu_i$$
$$y_i = Hx_i + v_i$$
$$Q = \text{cov}(u), \qquad R = \text{cov}(v), \qquad S = \text{cov}(u, v)$$

Discrete Algebraic Riccati Equation (DARE):

$$P = FPF^* + GQG^* - K_pR_eK_p^*$$

$$R_e = R + HPH^*$$
$$K_p = (FPH^* + GS)R_e^{-1}$$

Define $F^s = F - GSR^{-1}H$ and $Q^s = Q - SR^{-1}S^*$.

Theorem 8.3.1: Assume that F is stable (otherwise, $\{F, H\}$ is detectable), $\{F^s, GQ^{s/2}\}$ controllable on the unit circle,

$$\begin{bmatrix} Q & S \\ S^* & R \end{bmatrix} \succeq 0, \qquad R \succ 0.$$

Under these conditions, the DARE has a unique solution P such that $F - K_pH$ is stable. Moreover, this so-called stabilizing solution is positive semi-definite and results in a positive definite $R_e = R + HPH^*$.

5.3 Chapter 14

Lemma 14.2.1: Consider the zero-initial-condition Riccati recursion and assume $\{F, H\}$ detectable and $\{F, GQ^{1/2}\}$ unit-circle controllable so that the unique stabilizing solution P exists. Then P_i^0 converges to P if and only if $\{F, GQ^{1/2}\}$ is stabilizable.

I don't think we're in the zero-initial-condition case; rather, $P^0 = \sigma I$.

Exercise 14.4? p.546

5.4 Appendix E

(Confirm the statement – p.783)

Theorem E.5.1: Consider the DARE (above). The following are equivalent: (i) $\{F, H\}$ is detectable and $\{F^s, GQ^{s/2}\}$ is controllable on the unit circle; (ii) the DARE has a stabilizing solution P, i.e., one for which the matrix $F - K_pH$ is stable. Moreover, any such stabilizing solution is unique and positive semi-definite.

6 Callier / Desoer – Linear System Theory

Model – p. 57 – discrete-time system representation

$$x_{k+1} = A_k x_k + B_k u_k$$

$$y_k = C_k x_k + D_k u_k$$

Problem: we have w_k instead of $D_k u_k$. We can always find an appropriate D_k unless $u_k = 0$, which happens with probability 0, presumably, so it should be okay?

Time-invariant: (A, B, C, D) constant in time.

(Thms: pp. 293–294)

Theorem 8.d.62.ii Consider a discrete-time time-invariant system representation $R_d = [A, B, C, D]$. If A and B are real, there exists $F \in \mathbb{R}^{n_i \times n}$ such that

$$\sigma_p(A + BF) \subset \mathcal{D}(0, 1) = \{z \in \mathbb{C} : |z| < 1\}$$

if and only if the pair (A, B) is stabilizable.

Theorem 8.d.65.ii Consider a discrete-time time-invariant system representation $R_d = [A, B, C, D]$. If C and A are real, there exists $L \in \mathbb{R}^{n \times n_o}$ such that

$$\sigma_p(A + LC) \subset \mathcal{D}(0, 1) = \{z \in \mathbb{C} : |z| < 1\}$$

if and only if the pair (C, A) is detectable.

Problem: These theorems are for a time-invariant system. We have time-invariance on everything except D_k due to our w_k issue. This may be a non-issue, but I have no idea.

Translate to (F, G, H, K) notation:

Detectable: $\{F, H\}$ is detectable if and only if $\rho(F + LH) < 1$ for some L.

Stabilizable: $\{F, G\}$ is stabilizable if and only if $\rho(F + GL) < 1$ for some L.

Translate to our problem, $F = I$, $H = A$, $G = I$, $Q = \sigma_{\text{sys}}^2 I$:

Detectable: $\{I, A\}$ is detectable if and only if $\rho(I + LA) < 1$ for some L.

Stabilizable: $\{I, \sigma_{\text{sys}} I\}$ is stabilizable if and only if $\rho(I + \sigma_{\text{sys}} L) < 1$ for some L.

7 Hassibi – PhD Thesis

Results on Riccati equations.

Section 7.3, footnote 1: Def: $\{F, H\}$ is detectable if there exists a matrix K such that $F - KH$ is stable.

Linking this with the definition of detectable from KSHLinear:

$$\exists K with F - KH stable \equiv detectable \equiv \{F^*, H^*\} stabilizable$$

Lemma 8.7.3 Consider the Riccati recursion with positive semi-definite initial condition

$$P_{i+1} = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^* \ , \qquad P_0 \succeq 0.$$

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

$$P = FPF^* + GQG^* - (FPH^* + GS)(R + HPH^*)^{-1}(FPH^* + GS)^*.$$

Thereom 8.7.1 Consider the Riccati recursion

$$P_{i+1} = FP_iF^* + GQG^* - K_{p,i}R_{e,i}K_{p,i}^* \ , \qquad P_0$$

where $R \succ 0$, $Q - SR^{-1}S^* \succ 0$, $\{F, H\}$ is detectable and $\{F - GSR^{-1}H, GQ - GSR^{-1}S^*\}$ is stabilizable. Suppose, moreover, that the initial condition P_0 is such that

$$I + (P^a)^{*/2}P_0(P^a)^{1/2} \succ 0,$$

where P^a is the unique positive semi-definite solution to the dual Riccati recursion. Then P_i converges to the unique positive semi-definite matrix, P, that satisfies the discrete-time algebraic Riccati equation

$$P = FPF^* + GQG^* - (FPH^* + GS)(R + HPH^*)^{-1}(FPH^* + GS)^*.$$

8 Proofs

Lemma 1. *Assume that* $\{x_t\}$ *and* $\{y_t\}$ *follow the signal model above,* $\{t_0, t_0+1, t_0+2, \ldots\}$ *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_0*:* $\hat{x}_{t_0} = x_{t_0}$*, so* $\hat{N}_{t_0} = N_{t_0} = N_0$*;* ***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:* $\|w_t\| < \xi$ *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*

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

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*;*

2) there are no false support additions: $\hat{N}_t \subseteq N_t$ *for all sampling times* t*; and*

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

We may want to split claim 3 into its own piece because its proof relies on the other 2 parts, which are proved separately with induction.

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 $t = t_0$. Claim 1 follows from Theorem 1.3 in [1] 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. Let **TYPO HERE? $y_t \to y_{t,\mathbf{res}}$? Can probably remove after getting algorithm typed up.**

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

where $\hat{x}_{t,\text{init}}$ and $y_{t,\text{res}}$ are defined in the KFCS with LS algorithm and $\hat{x}_{t,\text{init}}$ satisfies $\text{supp}(\hat{x}_{t,\text{init}}) = \hat{N}_{t-1}$.

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 $|\text{supp}(\beta_t)| \leq |N_t| \leq S_{\max}$. With this, we can apply Theorem 1.3 in [1] 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 the algorithm, $\hat{N}_t = \hat{N}_{t-1} \cup \{j : |(\hat{x}_{t,\text{CSres}})_j| > \alpha\}$. 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. Suppose that $\mathbf{F_j}$ holds, that is, $\hat{N}_{t_j-1} = N_{t_j-1}$.

Since $\mathbf{F_j}$ holds, $\Delta_t \subseteq \Delta_{\mathrm{add},t_j}$ 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,

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

so that

$$\begin{aligned}
|(\hat{x}_{t,\mathrm{CSres}})_i| &= |(x_t)_i - [(x_t)_i - (\hat{x}_{t,\mathrm{CSres}})_i]| \\
&\ge \big||(x_t)_i| - |(x_t - \hat{x}_{t,\mathrm{CSres}})_i|\big| \\
&= |(x_t)_i| - |(x_t - \hat{x}_{t,\mathrm{CSres}})_i| \\
&> 2\alpha - \alpha \\
&= \alpha.
\end{aligned}$$

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

If $|(x_t)_i| > 2\alpha$ for all $i \in \Delta_{\mathrm{add},t_j}$, then $\Delta_t \subseteq \Delta_{\mathrm{add},t_j} \subseteq \hat{N}_t$; 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 $\{|(x_t)_i| > 2\alpha \text{ for all } i \in \Delta_{\mathrm{add},t_j}\}$ 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 $(x_t)_i$ of x_t are independent and identically distributed $\mathcal{N}\left(0, (t-t_j)\sigma_{\mathrm{sys}}^2\right)$ random variables. With this in mind, we see that

$$\begin{aligned}
\mathbf{Pr}\left(\hat{N}_t = N_t \,\middle|\, \mathbf{F_j}\right) &\ge \mathbf{Pr}\left(|(x_t)_i| > 2\alpha \text{ for all } i \in \Delta_t \,|\, \mathbf{F_j}\right) \\
&\ge \mathbf{Pr}\left(|(x_t)_i| > 2\alpha \text{ for all } i \in \Delta_{\mathrm{add},t_j}\right) \\
&= \left[\mathbf{Pr}\left(|(x_t)_1| > 2\alpha\right)\right]^{S_{\mathrm{add}}} \\
&= \left[2\mathcal{Q}\left(\frac{2\alpha}{\sigma_{\mathrm{sys}}\sqrt{t-t_j}}\right)\right]^{S_{\mathrm{add}}}.
\end{aligned}$$

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

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

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, $\{t_0, t_0+1, t_0+2, \ldots\}$ 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 $\mathbf{D} = \{\hat{N}_t = N_t = N_*$ 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 and let $\tilde{x}_t = \hat{x}_{t,GAKF}$ be the GAKF estimate of x_t.

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., $\tilde{x}_t$ converges to $\hat{x}_t$ in mean square.

Proof. Throughout, we assume that the event 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, $\nu_t = (\nu_t)_{N_*}$, $A = A_{[1:n],N_*}$, $\hat{x}_t = (\hat{x}_t)_{N_*}$, $J_t = (J_t)_{N_*,N_*}$, $K_t = (K_t)_{N_*,[1:n]}$, $P_{t|t-1} = (P_{t|t-1})_{N_*,N_*}$, $P_t = (P_t)_{N_*,N_*}$, 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.

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 + possibly combine with first paragraph about DARE stuff below**

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

Notice that for $t \in [t_* : t_{**}]$ our model is a discrete-time time-invariant linear system with $(F, G, H, K) = (I, I, A, I)$. We can apply elements of linear system theory **discussed in SECTION** to generate results relating to $P_{t|t-1}$. **As noted above: techncially K is not time-invariant because of w_t not being Ku_t, so this may break down.**

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 (DARR) with $F = I$, $G = I$, $Q = \sigma_{\text{sys}}^2 I_{N_*}$, $R = \sigma_{\text{obs}}^2 I_{n\times n}$, and $S = 0$. **Verify Q, R.** Note that Q and R are fixed in time since we assume that D occurs.

Since $|N_*| \leq S_{\max}$ and $\delta_{S_{\max}} < 1$, $A = (A_{[1:n],N_*})$ is full rank. **Need to link this to detectability.**

Since $\sigma_{\text{sys}} > 0$, $Q^{1/2} = \sigma_{\text{sys}} I$ is full rank, where $Q^{1/2}$ denotes the matrix square root of Q. We see that the matrix $L = -\frac{1}{\sigma_{\text{sys}}+1} I$ satisfies $\rho(I + \sigma_{\text{sys}} I L) = \rho\left[\left(\frac{1}{\sigma_{\text{sys}}+1}\right) I\right] < 1$, so

$\{I, \sigma_{\text{sys}} I\}$ is stabilizable. **Previously had thought we need $\sigma_{\text{sys}} \neq 1$ (see the rank definition - rank of a 0 matrix is never n), but I think it cleaned itself up...**

Point out $P_{0|-1} \succeq 0$.

Therefore, by **HASSIBI PHD RESULT**, the DARR converges to a positive semi-definite matrix $P_\star$. This implies that $K_t \to K_\star = P_\star A'(AP_\star A' + \mathbf{R})^{-1}$ and $J_t \to J_\star = (I - K_\star A)$. Further, by **LINEAR THM E.5.1**, $\rho(J_\star) < 1$.

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 $\tilde{J}_t \to J_\star$.

Let $\rho = \rho(J_\star)$ 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 $c_{\rho,2}$ such that $\|M\|_\rho \leq c_{\rho,2}\|M\|_2$ for any matrix M.

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 $t \geq t_\varepsilon$, the following conditions hold:

- $\|K_t - \tilde{K}_t\|_2 < \varepsilon$;
- $\|J_t - \tilde{J}_t\|_2 < \varepsilon$;
- $\|J_t\|_\rho \leq \|J_\star\|_\rho + (1-\rho)/4$;
- $\|\tilde{J}_t\|_2 < \|J_\star\|_2 + 1$; and $\leftarrow$ **added this for bounding $\mathrm{tr}(\tilde{P}_t)$**
- $\|\tilde{P}_{t|t-1}\|_2 < \|P_\star\|_2 + 1$. $\leftarrow$ **added this for bounding $\mathrm{tr}(\tilde{P}_t)$**

Point out that t_ε is independent of 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.**

$\uparrow$———— **/moved**

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

Let $t > t_\varepsilon > t_*$. By the KFCS with LS algorithm and the model, we see that

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

Similarly, using the GAKF algorithm 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

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

Let $u_t = (J_t - \tilde{J}_t)(\tilde{e}_{t-1} + \nu_t) + (\tilde{K}_t - K_t) w_t$, so that $\text{diff}_t = J_t \text{diff}_{t-1} + u_t$. Recursively applying this identity, we see that

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

If we define $M_k^t = J_t J_{t-1} \cdots J_{k+1} J_k$ for $k \le t$, then we can more compactly write

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

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

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

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

$$\begin{aligned}
\|M_k^t\|_\rho &= \|J_t J_{t-1} \cdots J_k\|_\rho \\
&\leq \|J_t\|_\rho \|J_{t-1}\|_\rho \cdots \|J_k\|_\rho \\
&\leq a^{t-k+1},
\end{aligned}$$

so $\|M_k^t\|_2 \leq c_{\rho,2} a^{t-k+1}$. With this, we see that

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

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

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

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

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

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

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.**

Suppose that $k \geq t_\varepsilon$. Then

$$\|\tilde{P}_k\|_2 = \|\tilde{J}_k \tilde{P}_{k|k-1}\|_2 \leq \|\tilde{J}_k\|_2 \|\tilde{P}_{k|k-1}\|_2 < (\|J_\star\|_2 + 1)(\|P_\star\|_2 + 1) < \infty.$$

Since $\tilde{P}_k$ is Hermitian, $\|\tilde{P}_k\|_2 = \lambda_{\max}(\tilde{P}_k)$. Therefore,

$$\operatorname{tr}(\tilde{P}_k) = \sum_i \lambda_i(\tilde{P}_k) \leq |N_*| \lambda_{\max}(\tilde{P}_k) = |N_*| \|\tilde{P}_k\|_2 < |N_*|(\|J_\star\|_2 + 1)(\|P_\star\|_2 + 1) < \infty,$$

so there exists some $B > 0$ such that $\operatorname{tr}(\tilde{P}_k) < B$ for all $k \geq t_\varepsilon$. **PROBLEM: B is no longer technically independent of t_ε, in fact, choice of t_ε was done so that B can be expressed. This affects C down the line and may affect the conclusion.**

Since $(\tau - 1) \geq t_\varepsilon$, we can combine these results to see that

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

Recall that $\nu_\tau = (\nu_\tau)_{N_*}$ and $\operatorname{cov}(\nu_\tau) = \mathbb{E}\left[\nu_\tau \nu_\tau'\right]$ is independent of D. **Elaborate why?** Notice that

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

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

We now see that

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

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

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

Combining these results with **reference prev equations....**, we see that

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

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

Problem: why is $\mathbb{E}\left[\|\text{diff}_{t_\varepsilon}\|_2^2 \,|\, \mathbf{D}\right]^{\frac{1}{2}}$ finite? NV ignores it, and AB says "because they're finite," which needs more explanation. Moving on, assume it's finite.

If

$$t_{\text{ms}} = t_\varepsilon + \left\lceil \log_a\left(\frac{C\varepsilon}{c_{\rho,2}\mathbb{E}\left[\|\text{diff}_{t_\varepsilon}\|_2^2 \,|\, \mathbf{D}\right]^{\frac{1}{2}}}\right)\right\rceil,$$

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 (B technically not constant)** and ε is arbitrary we have obtained our desired result. □

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,

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

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. □

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.