KFCS Theory

Animesh Biswas, Kevin Palmowski

Last compiled July 18, 2014

1 Model

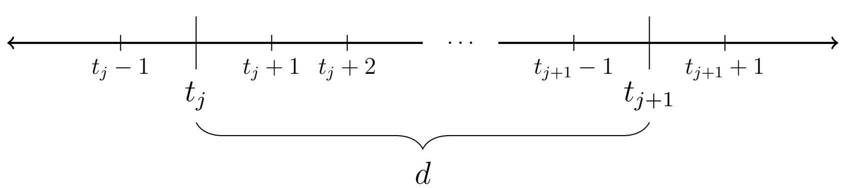
At each time $t \geq t_0$, we have

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

Here, $\mathbb{E}[w_t] = \mathbf{0}$, $\text{cov}(w_t) = \mathbb{E}[w_t w_t'] = \sigma^2_{\text{obs}} I_n$, 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 the diagonals for** $Nhat$**?**

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

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

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 $C_1\xi \le C_1(S_{\mathbf{max}})\xi = \alpha$.**

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 $|\text{supp}(\beta_t)| \le |N_t| \le S_{\max}$. With this, we can apply Theorem 1.3 in [1] to see that $\|\beta_t - \hat{\beta}_t\|_2 \le \alpha$ **(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,

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

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 $|(\hat{x}_{t,\text{CSres}})_i| \le \alpha$, 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 F_j holds, that is, $\hat{N}_{t_j-1} = N_{t_j-1}$.

Since 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

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

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

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 $\{|(x_t)_i| > 2\alpha \text{ for all } i \in \Delta_t \mid F_j\}$, which in turn is contained within the event $\{\hat{N}_t = N_t \mid F_j\}$.

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 $\mathcal{N}\left(0, (t-t_j)\sigma_{\text{sys}}^2\right)$ random variables. With this in mind, we see that

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

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

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

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 E_j. Therefore, $\mathbf{Pr}(E_j \mid F_j) = \mathbf{Pr}(\hat{N}_{t_j+\tau_{\text{det}}} = N_{t_j+\tau_{\text{det}}} \mid F_j) \geq 1 - \varepsilon$, 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.*

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

Define the event $D = \{\hat{N}_t = N_t = N_* \text{ for all } t \in [t_* : t_**]\}$, 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 and let $\tilde{x}_t = \hat{x}_{t,GAKF}$ be the GAKF estimate of x_t.

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 $\boldsymbol{Pr}\left(\|\tilde{x}_t - \hat{x}_t\|_2^2 \leq \varepsilon_{err} \,\middle|\, D\right) > 1 - \varepsilon$. Note that if $t_* + \tau_{KF} > t_{**}$, then this interval is empty and the result is vacuously true.

Proof. Throughout, we assume that the event D occurs and $t \in [t_* : t_{**}]$, so that all vectors are supported on N_*.

For simplicity of notation, we consider all variables and parameters only along the support set N_*. Thus, $\nu_t = (\nu_t)_{N_*}$, $A = A_{[1:n],N_*}$, $\hat{x}_t = (\hat{x}_t)_{N_*}$, $K_t = (K_t)_{N_*,[1:n]}$, $P_{t|t-1} = (P_{t|t-1})_{N_*,N_*}$, $J_t = (J_t)_{N_*,N_*}$, and analogously for $\tilde{x}_t$, $\tilde{K}_t$, $\tilde{P}_{t|t-1}$, and $\tilde{J}_t$.

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

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

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

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

[The next group of paragraphs in the writeup is heavy on exposition and can definitely be shortened. The main point we need to make is that the result from Hassibi's book ensures that the matrices $P_{t|t-1}$, etc. converge – I don't think we really need to give a crash course on control theory here, since this is the only place the results turn up.]

□

References

[1] Emmanuel 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.