Kalman Filter

Aleksandar Dogandžić

April 27, 2017

Contents

Reading: §13 in the textbook, [Hero 2015, §6.7.3], [Künsch 2001].

Preliminaries

It is easy to marginalize Gaussian random vectors: If

$$f(\boldsymbol{w} \mid \boldsymbol{x}) = \mathcal{N}(\boldsymbol{w} \mid A\boldsymbol{x}, \Sigma) \tag{1a}$$

conditional

$$f(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}, C) \tag{1b}$$

marginal

then the marginal probability density function (pdf) of $\boldsymbol{w}$ is

$$\begin{aligned} f(\boldsymbol{w}) &= \int f_{\boldsymbol{W}\mid\boldsymbol{X}}(\boldsymbol{w} \mid \boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ &= \mathcal{N}\big(\boldsymbol{w} \mid A\boldsymbol{\mu}, ACA^{\mathsf{T}} + \Sigma\big) \end{aligned} \tag{1c}$$

where "$^{\mathsf{T}}$" denotes a transpose. Of course, this also holds if we condition on a realization $\boldsymbol{y}$ of some random vector $\boldsymbol{Y}$:[1] If

[1] say the observed data in the Bayesian setting

$$f(\boldsymbol{w} \mid \boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}(\boldsymbol{w} \mid A\boldsymbol{x}, \Sigma)$$

conditional

$$f(\boldsymbol{x} \mid \boldsymbol{y}) = \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}, C)$$

marginal

then

$$f(\boldsymbol{w} \mid \boldsymbol{y}) = \mathcal{N}(\boldsymbol{w} \mid A\boldsymbol{\mu}, ACA^{\mathsf{T}} + \Sigma).$$

✻ MATRIX inversion lemma:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}. \tag{2a}$$

✻ A useful identity:

$$(A + BCD)^{-1}BC = A^{-1}B(C^{-1} + DA^{-1}B)^{-1} \tag{2b}$$

which follows from $BC(C^{-1} + DA^{-1}B) = (A + BCD)A^{-1}B$.

Measurement and State Equations

MEASUREMENT equation:

$$\boldsymbol{y}_k = \Phi\boldsymbol{\beta}_k + \underbrace{\boldsymbol{\nu}_k}_{\text{interference}} + \underbrace{\boldsymbol{\epsilon}_k}_{\text{noise}} \tag{3}$$

measurement equation

where k denotes the time index and the covariance matrices

$$V = \operatorname{cov}_{\boldsymbol{\nu}}(\boldsymbol{\nu}_k) \tag{4a}$$

$$R = \operatorname{cov}_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}_k) \tag{4b}$$

are assumed known. The matrix Φ is assumed known as well.

✻ STATE equation:

$$\boldsymbol{\beta}_k = H\boldsymbol{\beta}_{k-1} + J\boldsymbol{\eta}_k \tag{5}$$

state equation

where the covariance matrix

$$Q = \operatorname{cov}_{\boldsymbol{\eta}}(\boldsymbol{\eta}_k) \tag{6}$$

is assumed known. The matrices H and J are assumed known as well.

We assume that the random sequences $\boldsymbol{\nu}_k, \boldsymbol{\epsilon}_k$, and $\boldsymbol{\eta}_k$ are

- independent, identically distributed (i.i.d.) and zero-mean,
- Gaussian, and
- mutually independent.

The measurement and state equations (3) and (5) imply

$$f(\boldsymbol{y}_k \mid \boldsymbol{\beta}_k) = \mathcal{N}(\boldsymbol{y}_k \mid \Phi\boldsymbol{\beta}_k, V + R) \tag{7}$$

measurement equation

$$f(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}_{k-1}) = \mathcal{N}(\boldsymbol{\beta}_k \mid H\boldsymbol{\beta}_{k-1}, JQJ^{\mathsf{T}}) \tag{8}$$

state equation

where $k = 1, 2, \ldots$

✻ WE adopt the following prior pdf for the initial state:

$$f(\boldsymbol{\beta}_0) = \mathcal{N}\big(\boldsymbol{\beta}_0 \,\big|\, \boldsymbol{\beta}(0 \mid 0), P(0 \mid 0)\big). \tag{9}$$

Choosing $\widehat{\boldsymbol{\beta}}(0 \mid 0) = 0$ and a "large" prior covariance matrix $P(0 \mid 0)$ corresponds to a noninformative prior on $\boldsymbol{\beta}_0$.

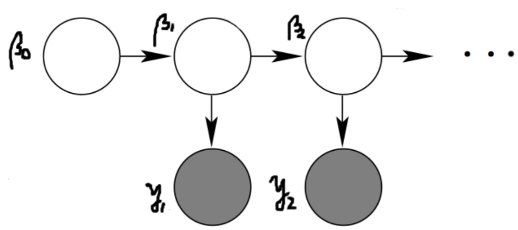

Figure 1: A directed acyclic graph (DAG) representation of a HMM.

HMM graphical model

OUR assumptions are depicted by the hidden-Markov model (HMM) graph in Fig. 1 implying, for example,

$$f(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{y}_1, \boldsymbol{y}_2) \propto f(\boldsymbol{\beta}_0) f(\boldsymbol{\beta}_1 \,|\, \boldsymbol{\beta}_0) f(\boldsymbol{\beta}_2 \,|\, \boldsymbol{\beta}_1) \cdot f(\boldsymbol{y}_1 \,|\, \boldsymbol{\beta}_1) f(\boldsymbol{y}_2 \,|\, \boldsymbol{\beta}_2). \tag{10}$$

Note the special conditional independence structure

$$\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_k, \boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_{k-1}\} \perp\!\!\!\perp \{\boldsymbol{Y}_{k+1}, \boldsymbol{Y}_{k+2}, \ldots, \boldsymbol{\beta}_{k+1}, \boldsymbol{\beta}_{k+2}, \ldots\} \,|\, \boldsymbol{\beta}_k \tag{11}$$

see Fig. 2.

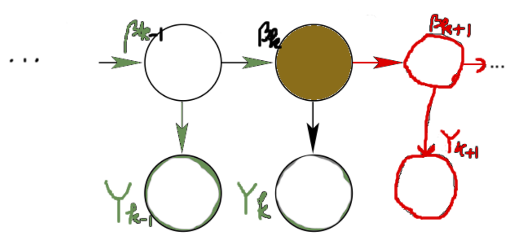

Figure 2: Conditional independence in (11).

Notation

WE introduce the following notation:

$$\boldsymbol{y}_{1:k} = (y_i)_{i=1}^{k}$$

and denote the conditional density of $\boldsymbol{\beta}_k$ given $\boldsymbol{y}_{1:\ell}$ by

$$f(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:\ell}).$$

- ☞ IF $k > \ell$, then $f(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:\ell})$ is a prediction density.
- ☞ IF $k = \ell$, then $f(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:k})$ is the filtering density.
- ☞ IF $k < \ell$, then $f(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:\ell})$ is a smoothing density.

Our Goals

GOAL: Estimate $\boldsymbol{\beta}_k$ *on-line (in real time)*.

☞ WE need to determine the *filtering pdf* $f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:k})$, which is Gaussian. Then, its mean is the minimum mean-square error (MMSE) (online) filtering estimate:

$$\widehat{\boldsymbol{\beta}}(k \mid k) = \mathrm{E}(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:k}).$$

☞ WE need the *one-step posterior-predictive pdf*

$$f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)})$$

also Gaussian. Its mean is the best one-step predictor:

$$\widehat{\boldsymbol{\beta}}(k \mid k-1) = \mathrm{E}(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)}).$$

The Gaussian smoothing density $f_{\boldsymbol{\beta}_k \mid \boldsymbol{Y}_{1:(k+s)}}(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k+s)})$ may also be of interest. Its mean is the best delayed (smoothing) estimate:

$$\widehat{\boldsymbol{\beta}}(k \mid k+s) = \mathrm{E}(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k+s)})$$

HW: Compute the smoothing pdfs.

for some positive index s.

How do we compute these pdfs and corresponding estimates? Here, we answer this question for filtering and one-step posterior-predictive densities under the linear observation and state-space Gaussian models (described above). This answer is known as the *Kalman filter*.

Derivation

WE DERIVE the Kalman filter by induction, starting with $k = 1$:

$$\begin{aligned} f(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)})\big|_{k=1} &= f(\boldsymbol{\beta}_0 \mid \underbrace{\boldsymbol{y}_{1:0}}_{\text{nothing}}) \\ &= f(\boldsymbol{\beta}_0) \\ &= \mathcal{N}\Big(\boldsymbol{\beta}_0 \,\big|\, \widehat{\boldsymbol{\beta}}(0 \mid 0), P(0 \mid 0)\Big). \end{aligned} \tag{12}$$

✻ AT time index $k-1$, our knowledge about $\boldsymbol{\beta}_{k-1}$ is given by the filtering pdf

$$f(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)}) = \mathcal{N}\Big(\boldsymbol{\beta}_{k-1} \,\Big|\, \widehat{\boldsymbol{\beta}}(k-1 \mid k-1), P(k-1 \mid k-1)\Big)$$

induction hypothesis

where

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(k-1 \mid k-1) &\triangleq \mathrm{E}\big(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)}\big) \\ P(k-1 \mid k-1) &\triangleq \mathrm{cov}\big(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)}\big). \end{aligned} \tag{13}$$

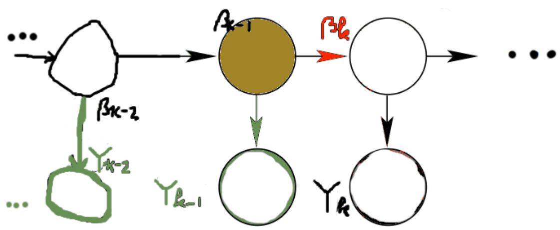

Figure 3: HMM graph implying (14).

One-step posterior-predictive pdf $f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)})$ from the filtering pdf $f(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)})$

SUPPOSE that we are at time $k-1$ and wish to predict $\boldsymbol{\beta}_k$. Assume that the filtering pdf $f(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)})$ is known.

✻ GOAL: Compute an update from the filtering pdf $f(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)})$ to the posterior-predictive pdf $f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)})$.

✻ KEY insight: By the HMM graph in Fig. 3, we have

$$\boldsymbol{\beta}_k \perp\!\!\!\perp \boldsymbol{Y}_{1:(k-1)} \mid \boldsymbol{\beta}_{k-1} \tag{14a}$$

or, equivalently,

$$f(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}_{k-1}, \boldsymbol{y}_{1:(k-1)}) = f(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}_{k-1}). \tag{14b}$$

Now,

$$\begin{aligned} f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)}) &= \int f_{\boldsymbol{B}_k, \boldsymbol{B}_{k-1} \mid \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k, \boldsymbol{\beta} \mid \boldsymbol{y}_{1:(k-1)})\, \mathrm{d}\boldsymbol{\beta} \\ &= \int \underbrace{f_{\boldsymbol{B}_k \mid \boldsymbol{B}_{k-1}, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}, \boldsymbol{y}_{1:(k-1)})}_{f_{\boldsymbol{B}_k \mid \boldsymbol{B}_{k-1}}(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}),\ \text{see (14)}} \\ &\qquad \cdot f_{\boldsymbol{B}_{k-1} \mid \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{B} \mid \boldsymbol{y}_{1:(k-1)})\, \mathrm{d}\boldsymbol{\beta} \\ &= \int f_{\boldsymbol{B}_k \mid \boldsymbol{B}_{k-1}}(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}) f_{\boldsymbol{B}_{k-1} \mid \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta} \mid \boldsymbol{y}_{1:(k-1)})\, \mathrm{d}\boldsymbol{\beta} \end{aligned} \tag{15}$$

Both $f(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}_{k-1}) = f(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}_{k-1}, \boldsymbol{y}_{1:(k-1)})$ and $f(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)})$ are Gaussian:

$$\begin{aligned} f(\boldsymbol{\beta}_k \mid \boldsymbol{\beta}_{k-1}, \boldsymbol{y}_{1:(k-1)}) &= \mathcal{N}\big(\boldsymbol{\beta}_k \mid H \boldsymbol{\beta}_{k-1}, JQJ^{\mathsf{T}}\big) \\ f(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)}) &= \mathcal{N}\big(\boldsymbol{\beta}_{k-1} \mid \widehat{\boldsymbol{\beta}}(k-1 \mid k-1), P(k-1 \mid k-1)\big) \end{aligned}$$

conditional

marginal

and we evaluate the integral (15) using (1):

$$f\big(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)}\big) = \mathcal{N}\big(\boldsymbol{\beta}_k \mid H \widehat{\boldsymbol{\beta}}(k-1 \mid k-1), HP(k-1 \mid k-1)\, H^{\mathsf{T}} + JQJ^{\mathsf{T}}\big).$$

Define

$$\widehat{\boldsymbol{\beta}}(k \mid k-1) \triangleq H \widehat{\boldsymbol{\beta}}(k-1 \mid k-1) \tag{16a}$$

$$P(k \mid k-1) \triangleq HP(k-1 \mid k-1) H^{\mathsf{T}} + JQJ^{\mathsf{T}} \tag{16b}$$

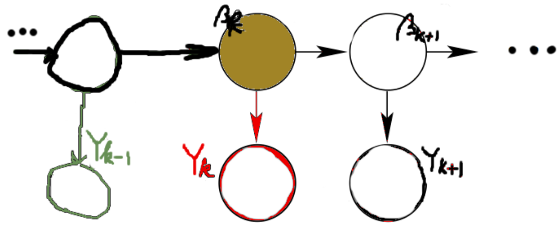

Figure 4: HMM graph implying (17b).

which leads to compact notation for the one-step posterior-predictive pdf of the hidden process $\boldsymbol{\beta}_k$:

$$f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)}) = \mathcal{N}\Big(\boldsymbol{\beta}_k \,\Big|\, \widehat{\boldsymbol{\beta}}(k \mid k-1), P(k \mid k-1)\Big).$$

Filtering pdf $f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:k})$ from the one-step posterior-predictive pdf $f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)})$

SUPPOSE now that time k has arrived and that we have collected a new observation $\boldsymbol{y}_k$. Here, the one-step posterior-predictive pdf $f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)})$ is known.

☞ GOAL: Compute an update from the one-step posterior-predictive pdf $f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)})$ to the filtering pdf $f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:k})$.

✻ KEY insight: By the HMM graph in Fig. 4, we have

$$\boldsymbol{Y}_k \perp\!\!\!\perp \boldsymbol{Y}_{1:(k-1)} \mid \boldsymbol{\beta}_k \tag{17a}$$

or, equivalently,

$$f(\boldsymbol{y}_k \mid \boldsymbol{\beta}_k, \boldsymbol{y}_{1:(k-1)}) = f(\boldsymbol{y}_k \mid \boldsymbol{\beta}_k). \tag{17b}$$

Now,

$$\begin{aligned}
f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:k}) &= f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_k, \boldsymbol{y}_{1:(k-1)}) \\
&\propto f(\boldsymbol{\beta}_k, \boldsymbol{y}_k \mid \boldsymbol{y}_{1:(k-1)}) \\
&\propto \underbrace{f(\boldsymbol{y}_k \mid \boldsymbol{\beta}_k, \boldsymbol{y}_{1:(k-1)})}_{f(\boldsymbol{y}_k \mid \boldsymbol{\beta}_k)} \; f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)}) && \text{see (17b)} \\
&\propto \underbrace{f(\boldsymbol{y}_k \mid \boldsymbol{\beta}_k)}_{\mathcal{N}(\boldsymbol{y}_k \mid \Phi\boldsymbol{\beta}_k, V+R)} \quad \underbrace{f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)})}_{\mathcal{N}(\boldsymbol{\beta}_k \mid \widehat{\boldsymbol{\beta}}(k \mid k-1), P(k \mid k-1))} \\
&\propto \exp\big[-0.5(\boldsymbol{y}_k - \Phi\boldsymbol{\beta}_k)^{\mathsf{T}}(V+R)^{-1}(\boldsymbol{y}_k - \Phi\boldsymbol{\beta}_k)\big] \\
&\qquad \cdot \exp\Big\{-0.5\big[\boldsymbol{\beta}_k - \widehat{\boldsymbol{\beta}}(k \mid k-1)\big]^{\mathsf{T}} P^{-1}(k \mid k-1)\big[\boldsymbol{\beta}_k - \widehat{\boldsymbol{\beta}}(k \mid k-1)\big]\Big\}
\end{aligned}$$

Expanding the quadratic forms in the exponent and grouping the linear and quadratic terms yields

$$\begin{aligned} f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:k}) &\propto \exp\Big\{-0.5\boldsymbol{\beta}_k^\top \underbrace{\left[\Phi^\top (V+R)^{-1}\Phi + P^{-1}(k \mid k-1)\right]}_{P^{-1}(k \mid k)} \boldsymbol{\beta}_k \\ &\qquad + \boldsymbol{\beta}_k^\top\left[\Phi^\top (V+R)^{-1}\boldsymbol{y}_k + P^{-1}(k \mid k-1)\,\widehat{\boldsymbol{\beta}}(k \mid k-1)\right]\Big\} \\ &= \mathcal{N}\Big(\boldsymbol{\beta}_k \,\Big|\, P(k \mid k)\left[\Phi^\top (V+R)^{-1}\boldsymbol{y}_k + P^{-1}(k \mid k-1)\widehat{\boldsymbol{\beta}}(k \mid k-1)\right], P(k \mid k)\Big) \end{aligned}$$

where[2]

[2] based on the definition (13)

$$P(k \mid k) = \left[\Phi^\top (V+R)^{-1}\Phi + P^{-1}(k \mid k-1)\right]^{-1} \tag{18a}$$

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(k \mid k) &= P(k \mid k)\left[\Phi^\top (V+R)^{-1}\boldsymbol{y}_k + P^{-1}(k \mid k-1)\widehat{\boldsymbol{\beta}}(k \mid k-1)\right] \\ &= P(k \mid k)\,\Phi^\top (V+R)^{-1}\boldsymbol{y}_k \\ &\quad + P(k \mid k)P^{-1}(k \mid k-1)\widehat{\boldsymbol{\beta}}(k \mid k-1). \end{aligned} \tag{18b}$$

Recall the *matrix inversion lemma*:

$$(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1}$$

and apply it as follows:

$$\underbrace{\Big[\overbrace{P^{-1}(k \mid k-1)}^{A} + \overbrace{\Phi^\top}^{B} \overbrace{(V+R)^{-1}}^{C} \overbrace{\Phi}^{D}\Big]^{-1}}_{P(k \mid k)} = P(k \mid k-1) - \underbrace{P(k \mid k-1)\Phi^\top[V+R+\Phi P(k \mid k-1)\Phi^\top]^{-1}}_{\triangleq \mathcal{K}(k)}\,\Phi P(k \mid k-1)$$

yielding

$$P(k \mid k) = P(k \mid k-1) - \mathcal{K}(k)\Phi P(k \mid k-1) \tag{19a}$$

where

$$\mathcal{K}(k) \triangleq P(k \mid k-1)\Phi^\top\left[V+R+\Phi P(k \mid k-1)\Phi^\top\right]^{-1} \tag{19b}$$

is known as the *Kalman gain*. Apply the identity

$$(A+BCD)^{-1}BC = A^{-1}B(C^{-1}+DA^{-1}B)^{-1}$$

as follows:

$$\begin{aligned} \overbrace{\Big[\underbrace{P^{-1}(k \mid k-1)}_{A} + \underbrace{\Phi^\top}_{B} \underbrace{(V+R)^{-1}}_{C} \underbrace{\Phi}_{D}\Big]^{-1}}^{P(k \mid k)} \underbrace{\Phi^\top}_{B} \underbrace{(V+R)^{-1}}_{C} &= P(k \mid k-1)\Phi^\top\left[V+R+\Phi P(k \mid k-1)\Phi^\top\right]^{-1} \\ &= \mathcal{K}(k). \end{aligned} \tag{20}$$

Now, use the identities (19a) and (20) to simplify $\widehat{\boldsymbol{\beta}}(k \mid k)$ in (18b):

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(k \mid k) &= \overbrace{P(k \mid k)\Phi^\top (V+R)^{-1}}^{\mathcal{K}(k),\text{ see (20)}}\, \boldsymbol{y}_k + \overbrace{P(k \mid k)P^{-1}(k \mid k-1)}^{I-\mathcal{K}(k)\Phi,\text{ see (19a)}}\,\widehat{\boldsymbol{\beta}}(k \mid k-1) \\ &= \mathcal{K}(k)\,\boldsymbol{y}_k + [I - \mathcal{K}(k)\Phi]\widehat{\boldsymbol{\beta}}(k \mid k-1) \\ &= \widehat{\boldsymbol{\beta}}(k \mid k-1) + \mathcal{K}(k)\left[\boldsymbol{y}_k - \Phi\widehat{\boldsymbol{\beta}}(k \mid k-1)\right]. \end{aligned}$$

Summary

We now summarize the Kalman-filtering scheme:

$$\widehat{\boldsymbol{\beta}}(k \mid k-1) = H\widehat{\boldsymbol{\beta}}(k-1 \mid k-1) \tag{21a}$$
$$P(k \mid k-1) = HP(k-1 \mid k-1)H^{\mathsf{T}} + JQJ^{\mathsf{T}} \tag{21b}$$

prediction

and complete the recursion as follows:

$$\widehat{\boldsymbol{\beta}}(k \mid k) = \widehat{\boldsymbol{\beta}}(k \mid k-1) + \mathcal{K}(k)\big[\underbrace{\boldsymbol{y}_k - \Phi\widehat{\boldsymbol{\beta}}(k \mid k-1)}_{\text{prediction error}}\big] \tag{21c}$$

filtering

$$P(k \mid k) = P(k \mid k-1) - \mathcal{K}(k)\Phi P(k \mid k-1) \tag{21d}$$

where

$$\mathcal{K}(k) = P(k \mid k-1)\Phi^{\mathsf{T}}\big[V + R + \Phi P(k \mid k-1)\Phi^{\mathsf{T}}\big]^{-1}. \tag{22}$$

Both the one-step posterior-predictive and filtering pdfs are multivariate Gaussian, implying that they are completely described by their mean vectors and covariance matrices:

$$f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:(k-1)}) = \mathcal{N}\big(\boldsymbol{\beta}_k \mid \widehat{\boldsymbol{\beta}}(k \mid k-1), P(k \mid k-1)\big)$$

(one-step posterior-predictive pdf)

$$f(\boldsymbol{\beta}_k \mid \boldsymbol{y}_{1:k}) = \mathcal{N}\big(\boldsymbol{\beta}_k \mid \widehat{\boldsymbol{\beta}}(k \mid k), P(k \mid k)\big).$$

(filtering pdf)

An alternative expression for $\widehat{\boldsymbol{\beta}}(k \mid k)$

Note that

$$\mathcal{K}(k) = P(k \mid k)\Phi^{\mathsf{T}}(V + R)^{-1} \tag{23}$$

see (20). Now, the expression for the posterior mean $\widehat{\boldsymbol{\beta}}(k \mid k)$ can be written as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(k \mid k) &= \widehat{\boldsymbol{\beta}}(k \mid k-1) + \mathcal{K}(k)[\boldsymbol{y}_k - \Phi\widehat{\boldsymbol{\beta}}(k \mid k-1)] \\ &= H\widehat{\boldsymbol{\beta}}(k-1 \mid k-1) \\ &\quad + P(k \mid k)\Phi^{\mathsf{T}}(V + R)^{-1}\big[\boldsymbol{y}_k - \Phi H\widehat{\boldsymbol{\beta}}(k-1 \mid k-1)\big]. \end{aligned} \tag{24}$$

RLS Algorithm

To establish a relationship between the Kalman recursion and recursive least-squares (RLS) algorithm, choose

$$H = I, \qquad J = 0. \tag{25}$$

Then, the state equation (5) reduces to the statement that the "state" is constant:

$$\boldsymbol{\beta}_k = H\boldsymbol{\beta}_{k-1} + J\boldsymbol{\eta}_k = \boldsymbol{\beta}_{k-1} \triangleq \boldsymbol{\beta}$$

and (21b) simplifies to

$$P(k \mid k-1) = P(k-1 \mid k-1). \tag{26}$$

☞ SIMPLIFIED notation. Considering that we do not have meaningful prediction steps any more, we can simplify the notation and define

$$P(k) = P(k \mid k) \tag{27a}$$
$$\widehat{\boldsymbol{\beta}}(\boldsymbol{y}_{1:k}) = \widehat{\boldsymbol{\beta}}(k \mid k). \tag{27b}$$

Replace the matrix Φ by the time-varying vector $\boldsymbol{\phi}_k^\top$:[3]

[3] The time-varying extension of the Kalman recursion is trivial.

$$\Phi = \boldsymbol{\phi}_k^\top. \tag{28}$$

Then, the measurement equation (3) simplifies to

$$y_k = \boldsymbol{\phi}_k^\top \boldsymbol{\beta} + v_k + \epsilon_k.$$

Under the above assumptions, (24) and (22) simplify to

basic form of the *RLS algorithm*

see (25) and (26)

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{y}_{1:k}) = \widehat{\boldsymbol{\beta}}(\boldsymbol{y}_{1:(k-1)}) + \frac{P(k)\boldsymbol{\phi}_k}{V+R}\big[y_k - \boldsymbol{\phi}_k^\top \widehat{\boldsymbol{\beta}}(\boldsymbol{y}_{1:(k-1)})\big] \tag{29}$$
$$\mathcal{K}(k) = \frac{P(k-1)\boldsymbol{\phi}_k}{V+R+\boldsymbol{\phi}_k^T P(k-1)\boldsymbol{\phi}_k} \tag{30}$$

and (21d) becomes

see (26) and (30)

$$\begin{aligned} P(k) &= P(k-1) - \mathcal{K}(k)\boldsymbol{\phi}_k^T P(k-1) \\ &= P(k-1) - \frac{P(k-1)\boldsymbol{\phi}_k\boldsymbol{\phi}_k^\top P(k-1)}{V+R+\boldsymbol{\phi}_k^\top P(k-1)\boldsymbol{\phi}_k}. \end{aligned}$$

If we define

$$\boldsymbol{h}_k \triangleq P(k-1)\boldsymbol{\phi}_k$$

then

$$P(k)\boldsymbol{\phi}_k = \boldsymbol{h}_k - \boldsymbol{h}_k \frac{\boldsymbol{\phi}_k^\top P(k-1)\boldsymbol{\phi}_k}{V+R+\boldsymbol{\phi}_k^\top P(k-1)\boldsymbol{\phi}_k} = \frac{V+R}{V+R+\boldsymbol{\phi}_k^\top \boldsymbol{h}_k}\boldsymbol{h}_k.$$

✻ We now summarize our RLS iteration:

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{y}_{1:k}) = \widehat{\boldsymbol{\beta}}(\boldsymbol{y}_{1:(k-1)}) + \frac{\boldsymbol{h}_k}{V + R + \boldsymbol{\phi}_k^\top \boldsymbol{h}_k}\big[y_k - \boldsymbol{\phi}_k^\top \widehat{\boldsymbol{\beta}}(\boldsymbol{y}_{1:(k-1)})\big]$$

where

$$\begin{aligned} \boldsymbol{h}_k &= P(k-1)\boldsymbol{\phi}_k \\ P(k) &= P(k-1) - \frac{\boldsymbol{h}_k \boldsymbol{h}_k^\top}{V + R + \boldsymbol{\phi}_k^\top \boldsymbol{h}_k}. \end{aligned}$$

We aim to solve the linear system

$$\boldsymbol{y}_{1:k} = \begin{bmatrix} \boldsymbol{\phi}_1^\top \\ \vdots \\ \boldsymbol{\phi}_k^\top \end{bmatrix} \boldsymbol{\beta} + \boldsymbol{\nu}_{1:k} + \boldsymbol{\epsilon}_{1:k}$$

recursively, with regularization: recall the prior pdf in (9).

References

Hero, Alfred O. (2015). *Statistical Methods for Signal Processing*. Lecture notes. Univ. Michigan, Ann Arbor, MI (cit. on p. 1).

Künsch, Hans R. (2001). "State space and hidden Markov models". In: *Complex stochastic systems*. Ed. by O. E. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg. Vol. 87. Monogr. Statist. Appl. Probab. Chapman & Hall/CRC, Boca Raton, FL, pp. 109–173 (cit. on p. 1).