Kalman Filter

Reading:

- Ch. 13 in Kay-I.
- Ch. 13 in Moon & Stirling.
- A general exposition on state space and hidden Markov models:

  H.R. Künsch, "State space and hidden Markov models," in *Complex Stochastic Systems*, O.E. Barndorff-Nielsen, D.R. Cox, and C. Klüpelberg, Eds., London UK: Chapman & Hall, 2001, ch. 3, pp. 109–173.

Kalman Filter: Model

Measurement equation:

$$\boldsymbol{y}_k = \Phi\, \boldsymbol{\beta}_k + \underbrace{\boldsymbol{\nu}_k}_{\text{interference}} + \underbrace{\boldsymbol{\epsilon}_k}_{\text{noise}} \tag{1}$$

where k denotes the time index and the covariance matrices

$$V = \operatorname{cov}(\boldsymbol{\nu}_k) \tag{2}$$

$$R = \operatorname{cov}(\boldsymbol{\epsilon}_k) \tag{3}$$

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

State equation:

$$\boldsymbol{\beta}_k = H\, \boldsymbol{\beta}_{k-1} + J\, \boldsymbol{\eta}_k \tag{4}$$

where the covariance matrix

$$Q = \operatorname{cov}(\boldsymbol{\eta}_k) \tag{5}$$

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

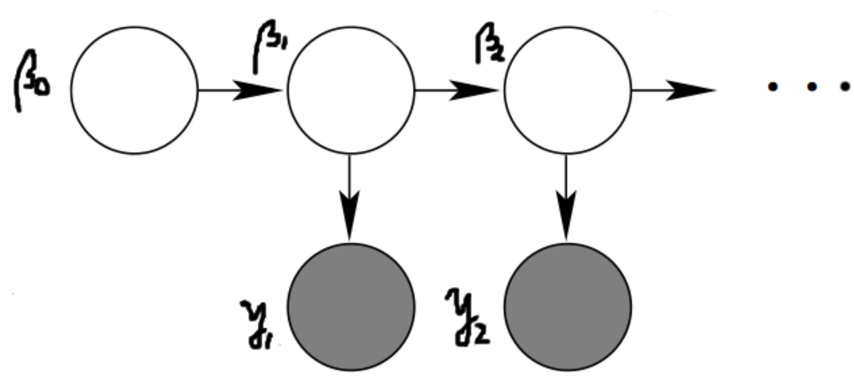
- i.i.d. and zero-mean,
- Gaussian, and
- mutually independent.

We also adopt the following prior pdf for the initial state:

$$f_{\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0) = \mathcal{N}\Big(\boldsymbol{\beta}_0 \,|\, \widehat{\boldsymbol{\beta}}(0\,|\,0), P(0\,|\,0)\Big).$$

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

These assumptions are depicted by the following hidden-Markov-model (HMM) graph:



implying, for example,

$$f_{\beta_0,\beta_1,\beta_2,y_1,y_2}(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, y_1, y_2)$$
$$\propto f_{\beta_0}(\boldsymbol{\beta}_0)\, f_{\beta_1 \,|\, \beta_0}(\boldsymbol{\beta}_1 \,|\, \boldsymbol{\beta}_0)\, f(\boldsymbol{\beta}_2 \,|\, \boldsymbol{\beta}_1)\, f_{y_1 \,|\, \beta_1}(y_1 \,|\, \boldsymbol{\beta}_1)\, f_{y_2 \,|\, \beta_2}(y_2 \,|\, \boldsymbol{\beta}_2).$$

The above model provides

$$f_{\boldsymbol{Y}_k \,|\, \boldsymbol{\beta}_k}(\boldsymbol{y}_k \,|\, \boldsymbol{\beta}_k) = \mathcal{N}(\boldsymbol{y}_k \,|\, \Phi\, \boldsymbol{\beta}_k, V + R) \quad \text{(obs. eqn.)}$$

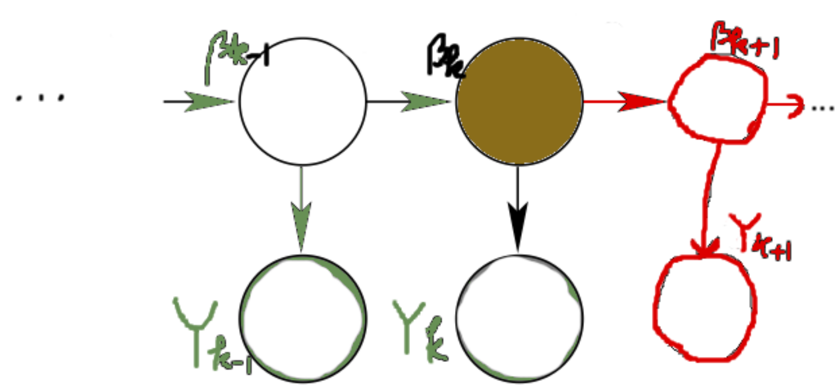
and

$$f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}) = \mathcal{N}(\boldsymbol{\beta}_k \,|\, H\boldsymbol{\beta}_{k-1}, J\, Q\, J^T) \quad \text{(state eqn.)}$$

where $k = 1, 2, \ldots$ Note the special conditional-independence structure

$$\{Y_1, \ldots, Y_k, \beta_0, \ldots, \beta_{k-1}\} \perp\!\!\!\perp \{Y_{k+1}, Y_{k+2}, \ldots, \beta_{k+1}, \beta_{k+2}, \ldots\} \mid \beta_k$$

depicted by the following graph:



Useful Facts

It is really easy to marginalize Gaussian random vectors: if

$$\begin{aligned} f_{\boldsymbol{W}\,|\,\boldsymbol{X}}(\boldsymbol{w}\,|\,\boldsymbol{x}) &= \mathcal{N}(\boldsymbol{w}\,|\,A\,\boldsymbol{x}, \Sigma) \quad \text{(conditional)} \\ f_{\boldsymbol{X}}(\boldsymbol{x}) &= \mathcal{N}(\boldsymbol{x}\,|\,\boldsymbol{\mu}, C) \quad \text{(marginal)} \end{aligned}$$

then the marginal pdf of $\boldsymbol{w}$ is

$$f_{\boldsymbol{W}}(\boldsymbol{w}) = \int f_{\boldsymbol{W}\,|\,\boldsymbol{X}}(\boldsymbol{w}\,|\,\boldsymbol{x})\, f_{\boldsymbol{X}}(\boldsymbol{x})\, d\boldsymbol{x} = \mathcal{N}(\boldsymbol{w}\,|\,A\,\boldsymbol{\mu}, A\,C\,A^T + \Sigma) \tag{6}$$

where "T" denotes a transpose. Of course, this also holds if we condition on a realization $\boldsymbol{y}$ of some random vector $\boldsymbol{Y}$ (say the observed data in the Bayesian setting): if

$$\begin{aligned} f_{\boldsymbol{W}\,|\,\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{w}\,|\,\boldsymbol{x},\boldsymbol{y}) &= \mathcal{N}(\boldsymbol{w}\,|\,A\,\boldsymbol{x}, \Sigma) \quad \text{(conditional)} \\ f_{\boldsymbol{X}\,|\,\boldsymbol{Y}}(\boldsymbol{x}\,|\,\boldsymbol{y}) &= \mathcal{N}(\boldsymbol{x}\,|\,\boldsymbol{\mu}, C) \quad \text{(marginal)} \end{aligned}$$

then

$$f_{\boldsymbol{W}\,|\,\boldsymbol{Y}}(\boldsymbol{w}\,|\,\boldsymbol{y}) = \mathcal{N}(\boldsymbol{w}\,|\,A \underbrace{\boldsymbol{\mu}}_{\substack{\text{marginal}\\ \text{mean}\\ \text{of } \boldsymbol{x}}}, A \underbrace{C}_{\substack{\text{marginal}\\ \text{covariance}\\ \text{of } \boldsymbol{x}}} A^T + \underbrace{\Sigma}_{\substack{\text{conditional}\\ \text{covariance}\\ \text{of } \boldsymbol{w}}}).$$

We introduce the following notation:

$$\boldsymbol{y}_{1:k} = [y_1, y_2, \dots, y_k]^T$$

and denote the conditional density of β_k given $\boldsymbol{y}_{1:l}$ by

$$f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_{1:l}}(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:l}).$$

If $k > l$, then $f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_{1:l}}(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:l})$ is a prediction density.
If $k = l$, then $f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_{1:k}}(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:k})$ is the filtering density.
If $k < l$, then $f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_{1:l}}(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:l})$ is a smoothing density.

Matrix inversion lemma:

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}. \quad (7)$$

A useful identity:

$$(A + B\,C\,D)^{-1}\,B\,C = A^{-1}\,B\,(C^{-1} + D\,A^{-1}\,B)^{-1} \quad (8)$$

which follows from $B\,C\,(C^{-1} + D\,A^{-1}\,B) = (A + B\,C\,D)\,A^{-1}\,B$.

What Are Our Goals?

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

We need to determine the filtering density $f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:k}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:k})$, which is Gaussian. Then, its mean is the **MMSE (online) filtering estimate:**

$$\widehat{\boldsymbol{\beta}}(k \,|\, k) = \mathrm{E}_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_k}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_k).$$

We need the one-step posterior-predictive pdf

$$f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)})$$

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

$$\widehat{\boldsymbol{\beta}}(k \,|\, k-1) = \mathrm{E}_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_{k-1}}[\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{k-1}].$$

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

$$\widehat{\boldsymbol{\beta}}(k \,|\, k+s) = \mathrm{E}_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots \boldsymbol{Y}_{k+s}}[\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{k+s}]$$

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*.

HW: Compute the smoothing pdfs.

Kalman Filter: Derivation

We derive the Kalman filter by induction, starting with $k = 1$:

$$f_{\boldsymbol{\beta}_{k-1} \mid \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)})\big|_{k=1} = f_{\boldsymbol{\beta}_0 \mid \boldsymbol{y}_{1:0}}(\boldsymbol{\beta}_0 \mid \underbrace{\boldsymbol{y}_{1:0}}_{\text{nothing}})$$

$$= \quad f_{\boldsymbol{\beta}_0}(\boldsymbol{\beta}_0) = \mathcal{N}\big(\boldsymbol{\beta}_0 \,\big|\, \widehat{\boldsymbol{\beta}}(0 \mid 0), P(0 \mid 0)\big).$$

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

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

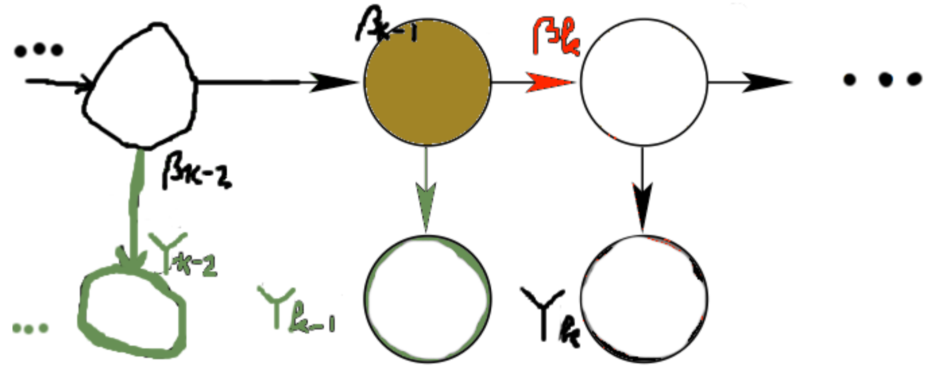
where

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(k-1 \mid k-1) \quad &\stackrel{\triangle}{=} \quad \mathrm{E}_{\boldsymbol{\beta}_{k-1} \mid \boldsymbol{Y}_{1:(k-1)}}[\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)}] \\ P(k-1 \mid k-1) \quad &\stackrel{\triangle}{=} \quad \mathrm{cov}_{\boldsymbol{\beta}_{k-1} \mid \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)}). \end{aligned} \tag{9}$$

Suppose that we are at time $k-1$ and wish to predict $\boldsymbol{\beta}_k$. We assume that the filtering pdf $f_{\boldsymbol{\beta}_{k-1} \mid \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_{k-1} \mid \boldsymbol{y}_{1:(k-1)})$ is known. Our prediction task requires the computation of the one-step posterior-predictive pdf $f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k | \boldsymbol{y}_{1:(k-1)})$:

$$
\begin{aligned}
f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_{1:(k-1)}}&(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)}) \\
&= \int f_{\boldsymbol{\beta}_k, \boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k, \boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:(k-1)})\, d\boldsymbol{\beta} \\
= \int & \underbrace{f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}, \boldsymbol{y}_{1:(k-1)})}_{f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta})} \, f_{\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:(k-1)})\, d\boldsymbol{\beta} \\
= \int & f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}) \, f_{\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:(k-1)})\, d\boldsymbol{\beta} \qquad (10)
\end{aligned}
$$

see the HMM graph below, where we observe:



implying

$$\boldsymbol{\beta}_k \perp\!\!\!\perp \boldsymbol{Y}_{1:(k-1)} \,|\, \boldsymbol{\beta}_{k-1}$$

or, equivalently,

$$f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}, \boldsymbol{y}_{1:(k-1)}) = f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}).$$

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

$$\overbrace{f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}, \boldsymbol{y}_{1:(k-1)})}^{\text{conditional}} = \mathcal{N}(\boldsymbol{\beta}_k \,|\, H\, \boldsymbol{\beta}_{k-1}, J\, Q\, J^T)$$

$$\underbrace{f_{\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{y}_{1:(k-1)})}_{\text{marginal}}$$

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

and we evaluate the integral (10) using (6):

$$f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)}) = \mathcal{N}\big(\boldsymbol{\beta}_k \,|\, H\widehat{\boldsymbol{\beta}}(k-1 \,|\, k-1), \\ H\, P(k-1 \,|\, k-1)\, H^T + J\, Q\, J^T\big).$$

Define

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(k \,|\, k-1) &\stackrel{\triangle}{=} H\, \widehat{\boldsymbol{\beta}}(k-1|k-1) \\ P(k \,|\, k-1) &\stackrel{\triangle}{=} H\, P(k-1 \,|\, k-1)\, H^T + J\, Q\, J^T \end{aligned}$$

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

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

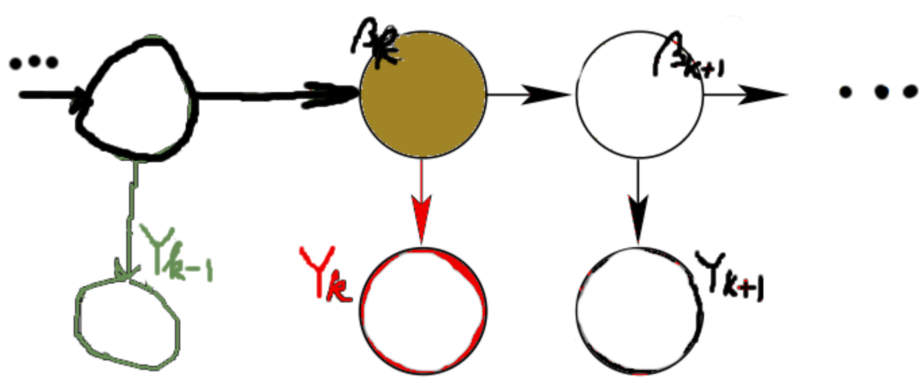
Suppose now that time k has arrived and that we have collected a new observation $\boldsymbol{y}_k$. Here, the one-step posterior predictive pdf $p_{\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)})$ is known. We wish to update our knowledge and incorporate $\boldsymbol{y}_k$ by computing the filtering density $f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:k}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:k})$:

$$
\begin{aligned}
f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_{1:k}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:k}) &= f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_k, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_k, \boldsymbol{y}_{1:(k-1)}) \\
&\propto f_{\boldsymbol{\beta}_k, \boldsymbol{Y}_k \,|\, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k, \boldsymbol{y}_k \,|\, \boldsymbol{y}_{1:(k-1)}) \\
&\propto \underbrace{f_{\boldsymbol{Y}_k \,|\, \boldsymbol{\beta}_k, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{y}_k \,|\, \boldsymbol{\beta}_k, \boldsymbol{y}_{1:(k-1)})}_{f_{\boldsymbol{Y}_k \,|\, \boldsymbol{\beta}_k}(\boldsymbol{y}_k \,|\, \boldsymbol{\beta}_k)} \, f_{\boldsymbol{\beta}_k | \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)}) \\
&\propto \underbrace{f_{\boldsymbol{Y}_k \,|\, \boldsymbol{\beta}_k}(\boldsymbol{y}_k \,|\, \boldsymbol{\beta}_k)}_{\mathcal{N}(\boldsymbol{y}_k \,|\, \Phi\, \boldsymbol{\beta}_k, V+R)} \cdot \underbrace{f_{\boldsymbol{\beta}_k \,|\, \boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)})}_{\mathcal{N}(\boldsymbol{\beta}_k \,|\, \widehat{\boldsymbol{\beta}}(k \,|\, k-1), P(k \,|\, k-1))}
\end{aligned}
$$

$$
\propto \exp[-\tfrac{1}{2}\, (\boldsymbol{y}_k - \Phi\, \boldsymbol{\beta}_k)^T \, (V + R)^{-1} \, (\boldsymbol{y}_k - \Phi\, \boldsymbol{\beta}_k)]
$$

$$
\cdot \exp\{-\tfrac{1}{2}\, [\boldsymbol{\beta}_k - \widehat{\boldsymbol{\beta}}(k \,|\, k-1)]^T \, P(k \,|\, k-1)^{-1} \, [\boldsymbol{\beta}_k - \widehat{\boldsymbol{\beta}}(k \,|\, k-1)]\}
$$

see the HMM graph below, where we observe:



implying

$$\mathbf{Y}_k \perp\!\!\!\perp \mathbf{Y}_{1:(k-1)} \,|\, \boldsymbol{\beta}_k$$

or, equivalently,

$$f_{\mathbf{Y}_k \,|\, \boldsymbol{\beta}_k, \mathbf{Y}_{1:(k-1)}}(\boldsymbol{y}_k \,|\, \boldsymbol{\beta}_k, \boldsymbol{y}_{1:(k-1)}) = f_{\mathbf{Y}_k \,|\, \boldsymbol{\beta}_k}(\boldsymbol{y}_k \,|\, \boldsymbol{\beta}_k).$$

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

$$\begin{aligned} & f_{\boldsymbol{\beta}_k \,|\, \mathbf{Y}_{1:k}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:k}) \\ & \qquad \propto \exp\Big\{ -\tfrac{1}{2}\boldsymbol{\beta}_k^T \underbrace{[\Phi^T (V+R)^{-1} \Phi + P^{-1}(k \,|\, k-1)]}_{P^{-1}(k \,|\, k)} \boldsymbol{\beta}_k \\ & \qquad + \boldsymbol{\beta}_k^T [\Phi^T (V+R)^{-1} \boldsymbol{y}_k + P^{-1}(k \,|\, k-1)\, \widehat{\boldsymbol{\beta}}(k \,|\, k-1)] \Big\} \\ & = \mathcal{N}\Big(\boldsymbol{\beta}_k \,|\, P(k \,|\, k)\, [\Phi^T (V+R)^{-1} \boldsymbol{y}_k + P^{-1}(k \,|\, k-1)\, \widehat{\boldsymbol{\beta}}(k \,|\, k-1)], \\ & \qquad P(k \,|\, k)\Big) \end{aligned}$$

implying [see also (9)]

$$P(k \,|\, k) = [\Phi^T (V+R)^{-1} \Phi + P^{-1}(k \,|\, k-1)]^{-1}$$

and

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(k \,|\, k) &= P(k \,|\, k)\, [\Phi^T (V+R)^{-1} \boldsymbol{y}_k + P^{-1}(k \,|\, k-1)\, \widehat{\boldsymbol{\beta}}(k \,|\, k-1)] \\ &= P(k \,|\, k)\, \Phi^T (V+R)^{-1} \boldsymbol{y}_k + P(k \,|\, k)\, P^{-1}(k \,|\, k-1)\, \widehat{\boldsymbol{\beta}}(k \,|\, k-1). \end{aligned}$$

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:

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

yielding

$$P(k\,|\,k) = P(k\,|\,k-1) - K(k)\,\Phi P(k\,|\,k-1) \qquad (11)$$

where

$$K(k) \;\triangleq\; P(k\,|\,k-1)\,\Phi^T\,[V + R + \Phi P(k\,|\,k-1)\,\Phi^T]^{-1}$$

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

$$(A + B\,C\,D)^{-1}\,B\,C = A^{-1}\,B\,(C^{-1} + D\,A^{-1}\,B)^{-1}$$

as follows:

$$\overbrace{[\underbrace{P^{-1}(k\,|\,k-1)}_{A} + \underbrace{\Phi^T}_{B}\underbrace{(V+R)^{-1}}_{C}\underbrace{\Phi}_{D}]^{-1}}^{P(k\,|\,k)}\underbrace{\Phi^T}_{B}\underbrace{(V+R)^{-1}}_{C}$$

$$= P(k\,|\,k-1)\,\Phi^T[V + R + \Phi\,P(k\,|\,k-1)\,\Phi^T]^{-1} = K(k). \quad (12)$$

Now, we utilize the identities (11) and (12) to simplify $\widehat{\boldsymbol{\beta}}(k\,|\,k)$ in (11):

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(k\,|\,k) &= \overbrace{P(k\,|\,k)\;\Phi^T\,(V+R)^{-1}}^{K(k),\ \text{see (12)}}\boldsymbol{y}_k \\
&\quad + \overbrace{P(k\,|\,k)\,P^{-1}(k\,|\,k-1)}^{I-K(k)\Phi\ \text{see (11)}}\,\widehat{\boldsymbol{\beta}}(k\,|\,k-1) \\
&= K(k)\,\boldsymbol{y}_k + [I - K(k)\,\Phi]\,\widehat{\boldsymbol{\beta}}(k\,|\,k-1) \\
&= \widehat{\boldsymbol{\beta}}(k\,|\,k-1) + K(k)\,[\boldsymbol{y}_k - \Phi\,\widehat{\boldsymbol{\beta}}(k\,|\,k-1)].
\end{aligned}$$

We now summarize the Kalman-filtering scheme:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(k\,|\,k-1) &= H\widehat{\boldsymbol{\beta}}(k-1|k-1) \\
P(k\,|\,k-1) &= HP(k-1\,|\,k-1)H^T + JQJ^T
\end{aligned}$$

and complete the recursion as follows:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(k\,|\,k) &= \widehat{\boldsymbol{\beta}}(k\,|\,k-1) + K(k)\,\underbrace{[\boldsymbol{y}_k - \Phi\,\widehat{\boldsymbol{\beta}}(k\,|\,k-1)]}_{\text{prediction error}} \\
P(k\,|\,k) &= P(k\,|\,k-1) - K(k)\,\Phi P(k\,|\,k-1)
\end{aligned}$$

where

$$K(k) = P(k\,|\,k)\,\Phi^T\,(V+R)^{-1}.$$

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:

$$\begin{aligned}
& f_{\boldsymbol{\beta}_k\,|\,\boldsymbol{Y}_{1:(k-1)}}(\boldsymbol{\beta}_k\,|\,\boldsymbol{y}_{1:(k-1)}) \\
&\quad = \mathcal{N}\Big(\boldsymbol{\beta}_k\,|\,\widehat{\boldsymbol{\beta}}(k\,|\,k-1), P(k\,|\,k-1)\Big) \quad \text{(one-step post. pred. pdf)}
\end{aligned}$$

$$f_{\boldsymbol{\beta}_k\,|\,\boldsymbol{Y}_{1:k}}(\boldsymbol{\beta}_k\,|\,\boldsymbol{y}_{1:k}) = \mathcal{N}\Big(\boldsymbol{\beta}_k\,|\,\widehat{\boldsymbol{\beta}}(k\,|\,k), P(k\,|\,k)\Big) \quad \text{(filtering pdf)}.$$

Comment

Since Kalman filter is linear (affine, more precisely), it is also the LMMSE estimator.

LMS and RLS Algorithms

Note that

$$K(k) = P(k \mid k)\, \Phi^T (V + R)^{-1}$$

see (12). Now, the expression for the posterior mean $\widehat{\boldsymbol{\beta}}(k \mid k)$ can be written as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}(k \mid k) &= \widehat{\boldsymbol{\beta}}(k \mid k-1) + K(k)\,[\boldsymbol{y}_k - \Phi\, \widehat{\boldsymbol{\beta}}(k \mid k-1)] \\ &= H\, \widehat{\boldsymbol{\beta}}(k-1 \mid k-1) \\ &\quad + P(k \mid k)\, \Phi^T\, (V + R)^{-1}\, [\boldsymbol{y}_k - \Phi\, H\, \widehat{\boldsymbol{\beta}}(k-1 \mid k-1)]. \qquad (13)\end{aligned}$$

To establish a relationship between the Kalman recursion and RLS and LMS algorithms, choose $H = I$ and $J = 0$, in which case the state equation (4) reduces to the statement that the "state" is constant:

$$\boldsymbol{\beta}_k = H\, \boldsymbol{\beta}_{k-1} + J\, \boldsymbol{\eta}_k = \boldsymbol{\beta}_{k-1} \stackrel{\triangle}{=} \boldsymbol{\beta}.$$

Replace the matrix Φ by the time-varying vector $\boldsymbol{x}_k^T$:[1]

$$\Phi = \boldsymbol{x}_k^T.$$

Then, the measurement equation (1) simplifies to

$$y_k = \boldsymbol{x}_k^T\, \boldsymbol{\beta} + \nu_k + \epsilon_k.$$

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

Under the above assumptions, (13) simplifies to

$$\widehat{\boldsymbol{\beta}}(k \,|\, k) = \widehat{\boldsymbol{\beta}}(k-1 \,|\, k-1) + \frac{P(k \,|\, k)\, \boldsymbol{x}_k}{V + R} \cdot [y_k - \boldsymbol{x}_k^T \, \widehat{\boldsymbol{\beta}}(k-1 \,|\, k-1)]$$

which (almost) corresponds to the basic form of the *recursive least-squares (RLS) algorithm*. We need an update equation for $P(k \,|\, k)$:

$$\begin{aligned} P(k \,|\, k) \quad &= \quad \underbrace{P(k \,|\, k-1)}_{H P(k-1 \,|\, k-1) H^T + J Q J^T} \\ &\quad - \underbrace{K(k)}_{P(k \,|\, k-1)\, \Phi^T \, [V + R + \Phi P(k \,|\, k-1)\, \Phi^T]^{-1}} \Phi \, P(k \,|\, k-1) \end{aligned}$$

which reduces to

$$\begin{aligned} P(k \,|\, k) &= P(k-1 \,|\, k-1) - \frac{P(k \,|\, k-1)\, \boldsymbol{x}_k \boldsymbol{x}_k^T \, P(k \,|\, k-1)}{V + R + \boldsymbol{x}_k^T P(k \,|\, k-1) \boldsymbol{x}_k} \\ &\overset{P(k \,|\, k-1) = P(k-1 \,|\, k-1)}{=} \\ &\quad P(k-1 \,|\, k-1) - \frac{P(k-1 \,|\, k-1)\, \boldsymbol{x}_k \boldsymbol{x}_k^T \, P(k-1 \,|\, k-1)}{V + R + \boldsymbol{x}_k^T P(k-1 \,|\, k-1)\, \boldsymbol{x}_k}. \end{aligned}$$

If we define

$$\boldsymbol{h}_k \overset{\triangle}{=} P(k-1 \,|\, k-1)\, \boldsymbol{x}_k$$

then

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

To summarize, here is our RLS iteration:

$$\widehat{\boldsymbol{\beta}}(k\,|\,k) = \widehat{\boldsymbol{\beta}}(k-1\,|\,k-1) + \frac{\boldsymbol{h}_k}{V+R+\boldsymbol{x}_k^T\,\boldsymbol{h}_k}\cdot[y_k - \boldsymbol{x}_k^T\,\widehat{\boldsymbol{\beta}}(k-1\,|\,k-1)]$$

where

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

Compare the above recursion with the *least-mean-square (LMS)* algorithm:

$$\widehat{\boldsymbol{\beta}}(k\,|\,k) = \widehat{\boldsymbol{\beta}}(k-1\,|\,k-1) + \mu\,\boldsymbol{x}_k\,[y_k - \boldsymbol{x}_k^T\,\widehat{\boldsymbol{\beta}}(k-1\,|\,k-1)]$$

where μ replaces $P(k-1\,|\,k-1)/(V+R+\boldsymbol{x}_k^T\,\boldsymbol{h}_k)$ in the RLS iteration. Thus, the LMS algorithm can be viewed as an approximation to the Kalman filter, for a simple constant-state dynamic model.

Kalman Filter: Example

Example 13.3 in Kay-I. Time-varying channel estimation:

$$y[n] = \sum_{k=0}^{p-1} \underbrace{h_n[k]}_{\substack{\text{time-varying}\\ \text{channel}}} \underbrace{v[n-k]}_{\substack{\text{transmitted}\\ \text{signal}}} + \underbrace{w[n]}_{\text{noise}}. \qquad (14)$$

If the channel coefficients are not changing too fast, we can model their variation by the following state equation for $\boldsymbol{h}[n]$:

$$\boldsymbol{h}[n] = A\,\boldsymbol{h}[n-1] + \boldsymbol{w}[n]$$

where

$$\boldsymbol{h}[n] = \begin{bmatrix} h_n[0] \\ h_n[1] \\ \vdots \\ h_n[p-1] \end{bmatrix}$$

A is assumed to be a known $p \times p$ matrix, and $\boldsymbol{w}[n]$ is a noise vector with covariance matrix $\sigma^2 \boldsymbol{I}$, where $\boldsymbol{I}$ denotes the identity matrix of appropriate dimensions. The measurement equation follows by rewriting (14) in the matrix form:

$$y[n] = \big[v[n]\ v[n-1]\ \cdots\ v[n-p+1]\big]\,\boldsymbol{h}[n] + w[n].$$

Kalman Filter: Example

Consider the following scalar linear system in state-space form:

$$\begin{aligned} y_k &= \beta_k + \epsilon_k \\ \beta_k &= h\,\beta_{k-1} + J\,\eta_k \quad k = 1, 2, \ldots \end{aligned}$$

where the random sequences ϵ_k, and η_k are

- i.i.d. and zero-mean with variances σ_ϵ^2 and 1 (respectively),
- Gaussian, and
- mutually independent

and the prior pdf for the initial state is

$$f_{\beta_0}(\beta_0) = \mathcal{N}(\beta_0 \,|\, 0, 0).$$

(a) Write the Kalman-filtering equations for this system.

(b) Interpret the prior pdf $f_{\beta_0}(\beta_0) = \mathcal{N}(\beta_0 \,|\, 0, 0)$, i.e. how much do we know about the initial state β_0?

(c) What happens when

$$\frac{J^2}{\sigma_\epsilon^2} \ll 1$$

and when

$$\frac{J^2}{\sigma_\epsilon^2} \gg 1?$$

Interpret the behavior of the Kalman gain in these two scenarios.

(d) Consider the steady-state scenario where $k \nearrow +\infty$ and obtain the corresponding Kalman-filtering equations. Find the Kalman gain $K(k)$ as well as $P(k \,|\, k)$ and $P(k \,|\, k-1)$ under this scenario.