# Kalman Filter

Reading:

- Ch. 13 in Kay-I.
- Ch. 13 in Moon & Stirling.
- For a general exposition on state space and hidden Markov models, see
  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 the covariance matrices

$$
\begin{aligned}
V &= \operatorname{cov}(\boldsymbol{\nu}_k) && (2)\\
R &= \operatorname{cov}(\boldsymbol{\epsilon}_k) && (3)
\end{aligned}
$$

are assumed known. The matrix $\Phi$ 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

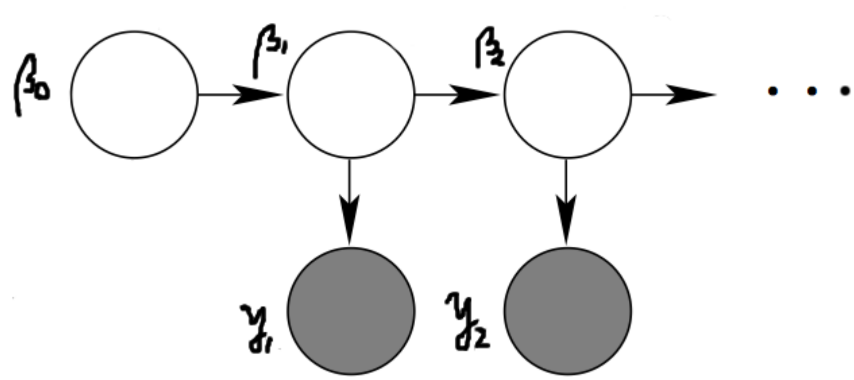
- independent, identically distributed (i.i.d.) and zero-mean,
- Gaussian, and
- mutually independent.

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

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

Choosing $\widehat{\boldsymbol{\beta}}(0|0) = 0$ (or some other value that is not too large in magnitude/norm) and a "large" prior covariance matrix $P(0|0)$ corresponds to a noninformative prior for $\boldsymbol{\beta}_0$.

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



implying, for example,

$$p(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, y_1, y_2) \propto \pi(\boldsymbol{\beta}_0)\, p(\boldsymbol{\beta}_1 \,|\, \boldsymbol{\beta}_0)\, p(\boldsymbol{\beta}_2 \,|\, \boldsymbol{\beta}_1) \cdot p(y_1 \,|\, \boldsymbol{\beta}_1) \cdot p(y_2 \,|\, \boldsymbol{\beta}_2).$$

The above model provides us with

$$p_{\boldsymbol{y}_k \,|\, \boldsymbol{\beta}_k}(\boldsymbol{y}_k \,|\, \boldsymbol{\beta}_k) = \mathcal{N}(\Phi \boldsymbol{\beta}_k, V + R), \quad k = 1, 2, \ldots \text{ (obs. eqn.)}$$

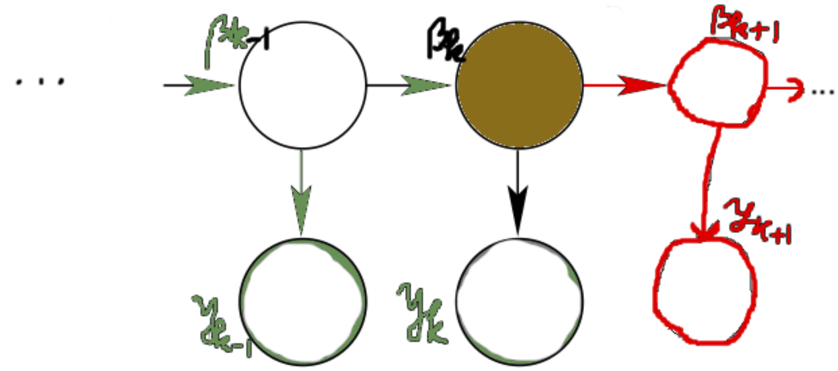
and

$$p_{\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}) = \mathcal{N}(H \boldsymbol{\beta}_{k-1}, J Q J^T),\ k = 1, 2, \ldots \text{ (state eqn.)}.$$

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:



**A Useful Fact.** It is really easy to marginalize Gaussian random vectors: if

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

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

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

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

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

then

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

Let us introduce the following notation:

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

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

$$p_{\beta_k \,|\, \boldsymbol{y}_{1:l}}(\beta \,|\, \boldsymbol{y}_{1:l}).$$

If $k > l$ then $p_{\beta_k \,|\, \boldsymbol{y}_{1:l}}(\beta \,|\, \boldsymbol{y}_{1:l}) \equiv$ prediction density.
If $k = l$ then $p_{\beta_k \,|\, \boldsymbol{y}_{1:k}}(\beta \,|\, \boldsymbol{y}_{1:k}) \equiv$ filtering density.
If $k < l$ then $p_{\beta_k \,|\, \boldsymbol{y}_{1:l}}(\beta \,|\, \boldsymbol{y}_{1:l}) \equiv$ smoothing density.

# What Are Our Goals?

**Our goal may be** to estimate $\boldsymbol{\beta}_k$ *on-line (in real time)*.

**Best (MMSE) on-line (filtering) estimate:**

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

$\Longrightarrow$ we need to determine the filtering density $p_{\beta_k \,|\, \boldsymbol{y}_{1:k}}(\beta_k \,|\, \boldsymbol{y}_{1:k})$ (which provides us much more information than just the mean — it gives us all we wish to know about $\boldsymbol{\beta}_k$ given $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_{k-1}$).

**Best one-step predictor:**

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

$\Longrightarrow$ we need to determine the one-step posterior-predictive pdf $p_{\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)})$.

**Best delayed (smoothing) estimate:**

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

for some positive index $s \Longrightarrow$ we need to obtain the smoothing density $p_{\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k+s)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k+s)})$.

How do we compute these estimates and corresponding pdfs? 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*.

Computing the smoothing pdf will be in your HW # 7, where you will derive the Raunch-Tung-Striebel Kalman-smoothing recursion.

# Kalman Filter: Derivation

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

$$p_{\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{y}_{1:(k-1)}}(\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{y}_{1:(k-1)})\Big|_{k=1} = p_{\boldsymbol{\beta}_0|\boldsymbol{y}_{1:0}}(\boldsymbol{\beta}_0 \,|\, \underbrace{\boldsymbol{y}_{1:0}}_{\text{nothing}})$$

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

We now assume that, at time $k-1$, our knowledge about $\boldsymbol{\beta}_{k-1}$ is given by the filtering pdf

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

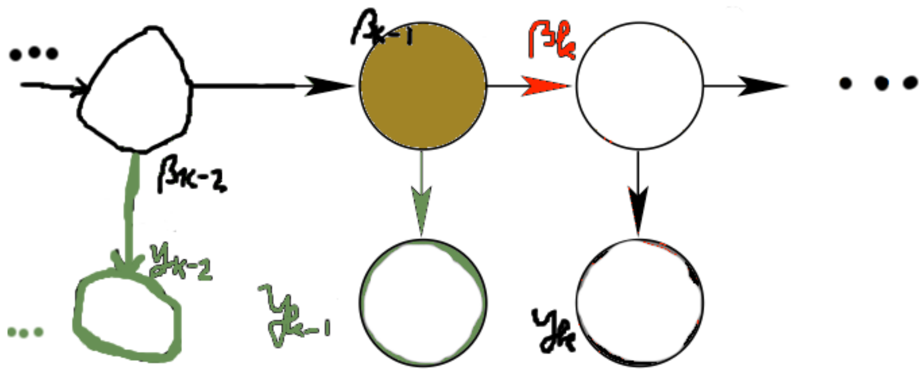
where

$$\begin{aligned} \widehat{\boldsymbol{\beta}}(k-1 \,|\, k-1) \quad &\stackrel{\triangle}{=} \quad \mathrm{E}\,[\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{y}_{1:(k-1)}] \\ P(k-1 \,|\, k-1) \quad &\stackrel{\triangle}{=} \quad \mathrm{cov}(\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{y}_{1:(k-1)}). \end{aligned} \tag{7}$$

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

$$
\begin{aligned}
p_{\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)}) \\
&= \int p_{\boldsymbol{\beta}_k, \boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{y}_{1:(k-1)}}(\boldsymbol{\beta}_k, \boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:(k-1)}) \, d\boldsymbol{\beta} \\
&= \int \underbrace{p_{\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}, \boldsymbol{y}_{1:(k-1)}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}, \boldsymbol{y}_{1:(k-1)})}_{p_{\boldsymbol{\beta}_k|\boldsymbol{\beta}_{k-1}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta})} \, p_{\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{y}_{1:(k-1)}}(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:(k-1)}) \, d\boldsymbol{\beta} \\
&= \int p_{\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}}(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}) \, p_{\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{y}_{1:(k-1)}}(\boldsymbol{\beta} \,|\, \boldsymbol{y}_{1:(k-1)}) \, d\boldsymbol{\beta} \qquad (8)
\end{aligned}
$$

see the HMM graph below where we observe:




implying

$$
\boldsymbol{\beta}_k \perp\!\!\!\perp \boldsymbol{y}_{1:(k-1)} \,|\, \boldsymbol{\beta}_{k-1} \iff p(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}, \boldsymbol{y}_{1:(k-1)}) = p(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}).
$$

Both $p(\boldsymbol{\beta}_k|\boldsymbol{\beta}_{k-1})$ and $p(\boldsymbol{\beta}_{k-1}|\boldsymbol{y}_{1:(k-1)})$ are Gaussian and we can evaluate the integral (8) using (6) and obtain

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

$$\begin{aligned}
\overbrace{p(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}, \boldsymbol{y}_{1:(k-1)})}^{\text{conditional}} &= p(\boldsymbol{\beta}_k \,|\, \boldsymbol{\beta}_{k-1}) = \mathcal{N}(H\boldsymbol{\beta}_{k-1}, JQJ^T) \\
\underbrace{p(\boldsymbol{\beta}_{k-1} \,|\, \boldsymbol{y}_{1:(k-1)})}_{\text{marginal}} &= \mathcal{N}\Big(\widehat{\boldsymbol{\beta}}(k-1 \,|\, k-1), P(k-1|k-1)\Big)
\end{aligned}$$

which implies

$$\begin{aligned}
p(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)}) &= \mathcal{N}\Big(H\widehat{\boldsymbol{\beta}}(k-1|k-1), \\
&\qquad HP(k-1|k-1)H^T + JQJ^T\Big).
\end{aligned}$$

Define

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

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

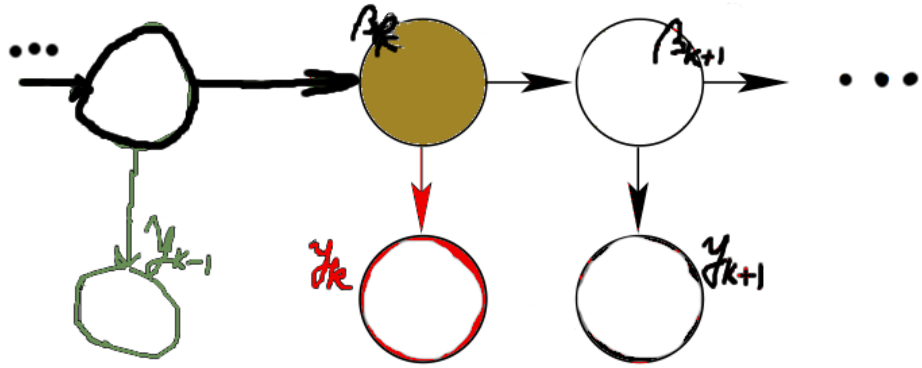
$$p(\boldsymbol{\beta}_k \,|\, \boldsymbol{y}_{1:(k-1)}) = \mathcal{N}\Big(\widehat{\boldsymbol{\beta}}(k \,|\, k-1), P(k \,|\, k-1)\Big).$$

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 $p_{\boldsymbol{\beta}_k|\boldsymbol{y}_{1:k}}(\boldsymbol{\beta}_k|\boldsymbol{y}_{1:k})$:

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

$$
\begin{aligned}
&\propto \exp[-\tfrac{1}{2}\,(\boldsymbol{y}_k - \Phi\boldsymbol{\beta}_k)^T (V+R)^{-1} (\boldsymbol{y}_k - \Phi\boldsymbol{\beta}_k)] \\
&\quad \cdot \exp\{-\tfrac{1}{2}\,[\boldsymbol{\beta}_k - \widehat{\boldsymbol{\beta}}(k \,|\, k-1)]^T P(k \,|\, k-1)^{-1} [\boldsymbol{\beta}_k - \widehat{\boldsymbol{\beta}}(k \,|\, k-1)]\}
\end{aligned}
$$

see the HMM graph below where we observe:




implying

$$
\boldsymbol{y}_k \perp\!\!\!\perp \boldsymbol{y}_{1:(k-1)} \,|\, \boldsymbol{\beta}_k \quad \Longleftrightarrow \quad p(\boldsymbol{y}_k \,|\, \boldsymbol{\beta}_k, \boldsymbol{y}_{1:(k-1)}) = p(\boldsymbol{y}_k \,|\, \boldsymbol{\beta}_k).
$$

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

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

implying that [see also (7)]

$$
\begin{aligned}
P(k \,|\, k) &= [\Phi^T (V+R)^{-1} \Phi + P(k \,|\, k-1)^{-1}]^{-1} \\
\widehat{\boldsymbol{\beta}}(k \,|\, k) &= \overbrace{[\Phi^T (V+R)^{-1} \Phi + P(k \,|\, k-1)^{-1}]^{-1}}^{P(k \,|\, k)} \\
&\quad \cdot [\Phi^T (V+R)^{-1} \boldsymbol{y}_k + P(k \,|\, k-1)^{-1}\, \widehat{\boldsymbol{\beta}}(k \,|\, k-1)] \\
&= P(k \,|\, k)\, \Phi^T (V+R)^{-1} \boldsymbol{y}_k \\
&\quad + P(k \,|\, k)\, P(k \,|\, k-1)^{-1}\, \widehat{\boldsymbol{\beta}}(k \,|\, k-1). \qquad (9)
\end{aligned}
$$

Recall the *matrix inversion lemma*:

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

and apply it as follows:

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

$$= P(k\,|\,k-1)$$

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

yielding

$$P(k\,|\,k) = P(k\,|\,k-1) - K(k)\,\Phi P(k\,|\,k-1) \qquad (10)$$

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*. Let us (re)derive another useful identity (which we mentioned earlier in handout # 4):

$$\begin{aligned}(A + BCD)^{-1}BC &= A^{-1}BC - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}BC \\ &= A^{-1}B(C^{-1} + DA^{-1}B)^{-1}(C^{-1} + DA^{-1}B)C \\ &\quad -A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}BC \\ &= A^{-1}B(C^{-1} + DA^{-1}B)^{-1}\end{aligned}$$

and apply it as follows:

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

$$= P(k\,|\,k-1)\,\Phi^T [V + R + \Phi\, P(k\,|\,k-1)\,\Phi^T]^{-1} = K(k). \quad (11)$$

Now, we utilize the identities (10) and (11) to simplify $\widehat{\boldsymbol{\beta}}(k\,|\,k)$ in (9):

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(k\,|\,k) &= \overbrace{P(k\,|\,k)\,\Phi^T (V+R)^{-1}}^{K(k),\ \text{see (11)}} \boldsymbol{y}_k \\
&\quad + \overbrace{P(k\,|\,k)\,P(k\,|\,k-1)^{-1}}^{I-K(k)\,\Phi\ \text{see (10)}}\, \widehat{\boldsymbol{\beta}}(k\,|\,k-1) \\
&= K(k)\,\boldsymbol{y}_k + [I - K(k)\,\Phi]\,\widehat{\boldsymbol{\beta}}(k\,|\,k-1) \\
&= \widehat{\boldsymbol{\beta}}(k\,|\,k-1) + K(k)\,[\boldsymbol{y}_k - \Phi\,\widehat{\boldsymbol{\beta}}(k\,|\,k-1)].
\end{aligned}$$

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

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

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:

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

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

# Comment

Updating mean vectors and covariance matrices according to the Kalman-filtering equations makes sense (even) without imposing the Gaussian assumptions on the measurement-model and prior pdfs. Kalman filter belongs to the category of "best linear" estimators (predictors); we met this notion earlier when we introduced BLUE in handout # 3.

# Relationship with the LMS and RLS Algorithms

Note that

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

see (11). Now, the expression for the posterior mean $\widehat{\boldsymbol{\beta}}(k \,|\, k)$ can be written as

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

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:

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

Furthermore, replace the matrix $\Phi$ 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, (12) 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 $P(k\,|\,k)$:

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

which reduces to

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

If we define

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

then

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

To summarize, here is our RLS iteration:

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

where

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

Let us now compare the above recursion with the *least-mean-square (LMS)* algorithm:

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

where $\mu$ replaces $P(k-1 \mid k-1)/(V + R + \boldsymbol{x}_k^T\, \boldsymbol{h}_k)$ in the RLS iteration. Thus, the LMS algorithm can be viewed as an approximation to the Kalman filter.

# Kalman Filter: Example

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

$$y[n] = \sum_{k=0}^{p-1} \underbrace{h_n[k]}_{\text{time-varying channel}} \quad \underbrace{v[n-k]}_{\text{transmitted signal}} + \underbrace{w[n]}_{\text{noise}}. \qquad (13)$$

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

$$\underbrace{\boldsymbol{h}[n]}_{x[n]} = \boldsymbol{A}\boldsymbol{h}[n-1] + \boldsymbol{w}[n]$$

where

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

$\boldsymbol{A}$ 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 (13) in the matrix form:

$$y[n] = \Big[ v[n] \; v[n-1] \; \dots \; v[n-p+1] \Big] \cdot \boldsymbol{h}[n] + w[n].$$