EE 126 — Probability and Random Processes
April 19, 2018 — Kalman Filter

1 Lecture

Kalman Filter

Theorem.

$$L[X \mid Y, Z] = L[X \mid Y] + L[X \mid Z]$$

if Y & Z are uncorrelated (i.e. $cov(Y, Z) = 0$, $Y \perp Z$, $\mathbb{E}[YZ] = 0$). Then the error is orthogonal to both Y and Z.

Theorem. If Y and Z are not orthogonal, we should use Gram-Schmidt to make them so.

$$\begin{aligned} L[X \mid Y, Z] &= L[X \mid Y, Z^{\perp}] \\ &= L[X \mid Y] + L[X \mid Z^{\perp}] \end{aligned}$$

where $Z^{\perp} = Z - Z^{\parallel} = Z - L[Z \mid Y] =$ the "innovative" component of Z.

Why wouldn't the vectors be orthogonal? Perhaps we observe Y, then Z.

In words, we orthogonalize the basis $\{Y, Z\}$ before projecting. The equivalence is clear because $\text{span}\{Y, Z\} = \text{span}\{Y, Z^{\perp}\}$, i.e. $\{Y, Z^{\perp}\}$ is a basis for the same space $\mathcal{L}(Y, Z)$.

As setup for the **Kalman filter**, we have a sequence of noisy observations Y_i's $Y_1, ..., Y_n$ and want to do linear estimation of the causes $X_1, ..., X_n$. We will denote $\{Y_1, ..., Y_n\}$ as Y^n. The Kalman filter will perform iterative linear estimation of $L[X_n \mid Y^n]$ in an online fashion (i.e. we can't wait for all of the Y's before spitting out X's).

Note: the Kalman filter "filters" out noise from the Y_i's to produce X_i's.

Our iterative estimates of $L[X_n \mid Y^n]$ adhere to the following structure:

1) $L[X_1 \mid Y^1]$

2) $L[X_2 \mid Y^2] = L[X_2 \mid Y_1] + L[X_2 \mid Y_2^{\perp}]$
   - $Y_2^{\perp} = Y_2 - L[Y_2 \mid Y_1]$

3) $L[X_3 \mid Y^2, Y_3] = L[X_3 \mid Y^2] + L[X_3 \mid Y_3^{\perp}] = L[X_3 \mid Y_1] + L[X_3 \mid Y_2^{\perp}] + L[X_3 \mid Y_3^{\perp}]$
   - $Y_3^{\perp} = Y_3 - L[Y_3 \mid Y^2]$

⋮

n) $L[X_n \mid Y^n] = L[X_n \mid Y^{(n-1)}] + L[X_n \mid Y_n^{\perp}]$

In words, we first estimate $L[X_1 \mid Y_1]$. Then we estimate $L[X_2 \mid Y_1, Y_2] = L[X_2 \mid Y_1] + L[X_2 \mid Y_2^{\perp}]$. And so on...

State-Space Equations

Here we will study the scalar versions of the state-space equations. For the vector version, read the book.

$$X_n = aX_{n-1} + V_n$$
$$Y_n = cX_n + W_n$$

where a is the linear estimation coefficient and V_n, W_n are i.i.d. zero mean noise.

The goal is to estimate X_n given Y_n. We will refer to $L[X_n \mid Y^m]$ as $\hat{X}_{n|m}$, i.e. the estimate of X_n at time m. Then $L[X_n \mid Y^{n-1}] = \hat{X}_{n|n-1}$. Furthermore, we define

$$\sigma^2_{n|m} \coloneqq \mathbb{E}[(X_n - \hat{X}_{n|m})^2] \quad \text{(error variance)}$$
$$\Delta_{n|m} \coloneqq X_n - \hat{X}_{n|m} \quad \text{(error)}$$

Note that if the noise is Gaussian, then the LLSE $L[X_n \mid Y^n]$ (Kalman filter) is equal to the MMSE and is optimal.

Kalman Equations (Scalar Case)

(1) $$\underbrace{\hat{X}_{n|n}}_{\text{target of estimation}} = \underbrace{\hat{X}_{n|n-1}}_{\text{precomputed (known) estimate}} + K_n \underbrace{(Y_n - c\hat{X}_{n|n-1})}_{\tilde{Y}_n\text{: innovation in } Y_n}$$

We estimate X_n based on what we've already seen (before n), and then we update the preliminary estimate after observing n. So the first component is the "predict" part and the second component is the "update" part. (Also, $\tilde{Y}_n \perp Y^{(n-1)}$.)

(2) $K_n = \frac{c\sigma^2_{n|n-1}}{c^2\sigma^2_{n|n-1} + \sigma^2_W}$ (Kalman gain)

σ^2_W is the variance of W_n.

(3) $\sigma^2_{n|n-1} = a^2\sigma^2_{n-1|n-1} + \sigma^2_V$

(4) $\sigma^2_{n|n} = \sigma^2_{n|n-1}(1 - K_n c)$

(2), (3), and (4) can be precomputed because they only depend on constants and known values.

On round n, we **input** $\hat{X}_{n-1|n-1}$, $\sigma^2_{n-1|n-1}$, and Y_n and **output** $\hat{X}_{n|n}$ and $\sigma^2_{n|n}$.

Explanation / Derivation

WLOG, assume $c = 1$ (otherwise we can scale $Y_n = cX_n + W_n$).

(1)
$$\hat{X}_{n|n} = L[X_n \mid Y^n] = L[X_n \mid Y^{n-1}] + L[X_n \mid \tilde{Y}_n]$$

where $\tilde{Y}_n = Y_n - L[Y_n \mid Y^{n-1}]$.

Observe:

$$\begin{aligned} L[Y_n \mid Y^{n-1}] &= L[cX_n + W_n \mid Y^{n-1}] \\ &= cL[X_n \mid Y^{n-1}] + L[W_n \mid Y^{n-1}] \\ &= c\hat{X}_{n|n-1} + 0 \quad (W_n \text{ is independent of everything else}) \end{aligned}$$

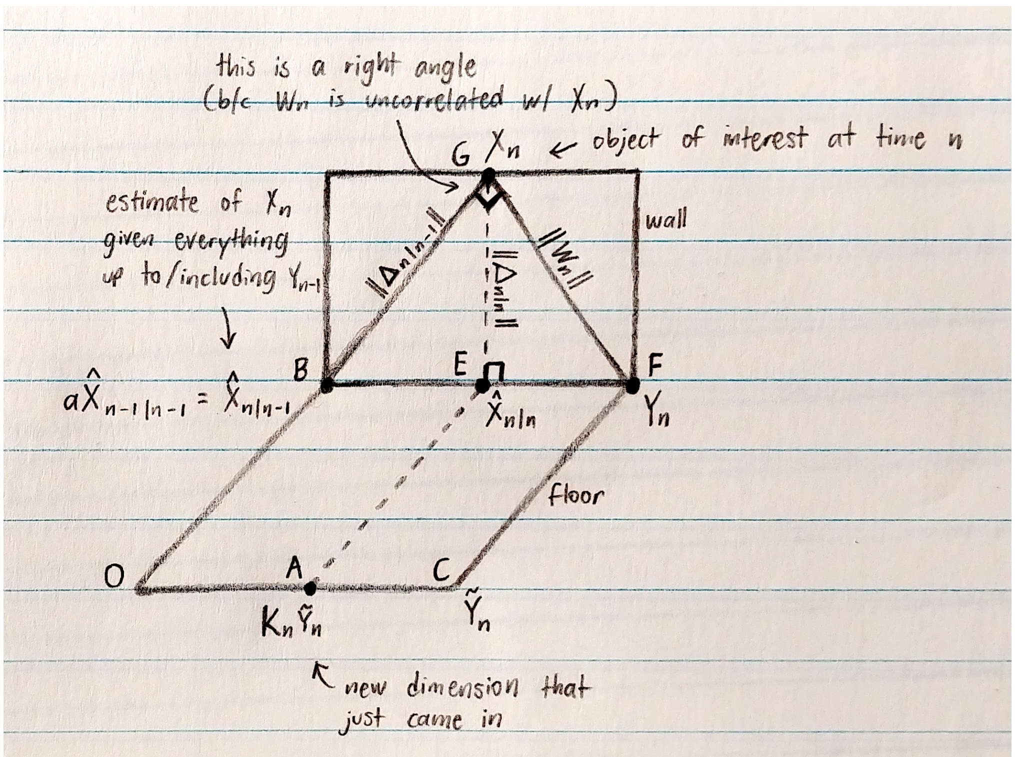
And then

$$\begin{aligned} \hat{X}_{n|n-1} &= L[X_n \mid Y^{n-1}] \\ &= L[aX_{n-1} + V_n \mid Y^{n-1}] \\ &= aL[X_{n-1} \mid Y^{n-1}] + L[V_n \mid Y^{n-1}] \\ &= aL[X_{n-1} \mid Y^{n-1}] \\ &= a\hat{X}_{n-1|n-1} \quad \to L[Y_n \mid Y^{n-1}] = ca\hat{X}_{n-1|n-1} \end{aligned}$$

(2) Note that K_n is the Kalman gain (it's just a gain).

$$\begin{aligned}
K_n\tilde{Y}_n &= L[X_n \mid \tilde{Y}_n] \\
&= \text{Proj}_{\tilde{Y}_n} X_n \ \ (= \text{“}b\tilde{Y}_n\text{”}) \\
&= \frac{\mathbb{E}[X_n\tilde{Y}_n]}{\mathbb{E}[\tilde{Y}_n^2]} \\
&= \ldots \\
&= \frac{\sigma_{n|n-1}}{\sigma_{n|n-1}^2 + \sigma_W^2}
\end{aligned}$$

Geometry



The Kalman gain can be found in $BE = \|K_n\tilde{Y}_n\| = K_n \cdot BF$. By similar triangles,

$$\frac{BE}{BG} = \frac{BG}{BF} \implies BE = \frac{BG^2}{BF}$$

$$BE = K_n \cdot BF \implies K_n = \frac{BG^2}{BF^2} = \frac{\sigma_{n|n-1}^2}{\sigma_{n|n-1}^2 + \sigma_W^2}$$

where BG^2 is the square of the error at the previous time.