About TA sections:

TA: Ekaterina Zubova (ez268@cornell.edu)

Section time and location: 8:40am - 9:55am Rockefeller Hall 132

Office hours: Tuesday 4:30-5:30 pm in Uris Hall 451; other times available by appointment (just send me an email).

In these notes:

1 Kalman Filter: State-Space Representation and Notation

For state-space systems of the form:

$$X_t = A_t X_{t-1} + C_t u_t$$
$$Z_t = D_t X_t + v_t$$

where:

- X_t is the unobserved (latent) state vector at time t
- $u_t \sim \mathcal{N}(0, I)$ is process noise (we assume the standard Gaussian)
- Z_t is the observed measurement at time t
- $v_t \sim \mathcal{N}(0, \Sigma_v)$ is observation noise

The Kalman filter recursively computes estimates of the state X_t conditional on the history of observations $Z_t, Z_{t-1}, ..., Z_0$, and an initial prior $X_{0|0}$ with covariance $P_{0|0}$.

The **filtering equation** for the optimal estimate $X_{t|t}$ is:

$$X_{t|t} = A_t X_{t-1|t-1} + K_t \left(Z_t - D_t A_t X_{t-1|t-1} \right),$$

where

- $X_{t|t} \equiv \mathbb{E}[X_t \mid Z^t]$ is the estimate of the state using all available data up to time t
- K_t is the **Kalman gain**, which balances how much weight we give to the prediction versus the observation
- The term $Z_t - D_t A_t X_{t-1|t-1}$ is the **innovation**, i.e., what we observed minus what we expected.

Intuition: The Kalman gain K_t is chosen to minimize the posterior variance, i.e., to make our estimate $X_{t|t}$ as accurate as possible in terms of its mean squared error.

2 Kalman Filter: Scalar Case

Consider the scalar process

$$x_t = \rho x_{t-1} + u_t, \quad z_t = x_t + v_t, \quad \begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}\right), \quad x_{0|0} = \bar{x}_0, \quad \mathbb{E}[(\bar{x}_0 - x_0)^2] = p_{0|0}.$$

With $x_{0|0}$ given, we can compute a prior for x_1.

Using the transition equation:

$$x_{1|0} = \mathbb{E}[x_1 \mid x_0] = \rho x_{0|0}.$$

the variance of this estimate is:

$$\mathbb{E}[(x_1 - x_{1|0})^2] = \mathbb{E}[(\rho(x_0 - \bar{x}_0) + u_1)^2] = \rho^2 \mathbb{E}[(x_0 - \bar{x}_0)^2] + \mathbb{E}[u_1^2] + 2\underbrace{\rho \mathbb{E}[(x_0 - \bar{x}_0)u_1]}_{=0} = \rho^2 p_{0|0} + \sigma_u^2$$

So we define prior variance as:

$$p_{1|0} = \rho^2 p_{0|0} + \sigma_u^2$$

Intuition:

- $\rho^2 p_0$ reflects propagated uncertainty from x_0,
- σ_u^2 is innovation uncertainty in x_1.

Combining the prior and the signal, we write the updated estimate as a convex combination:

$$x_{1|1} = (1 - k_1) x_{1|0} + k_1 z_1$$

Recall that our goal is to find a Kalman gain k_t to minimize the posterior variance. Define the posterior error:

$$x_1 - x_{1|1} = x_1 - \big((1 - k_1)x_{1|0} + k_1 z_1\big) = (1 - k_1)(x_1 - x_{1|0}) - k_1 v_1$$

Then the posterior variance is:

$$\mathbb{E}[(x_1 - x_{1|1})^2] = (1 - k_1)^2 \mathbb{E}[(x_1 - x_{1|0})^2] + k_1^2 \mathbb{E}[v_1^2] = (1 - k_1)^2 p_{1|0} + k_1^2 \sigma_v^2$$

To find the optimal gain, minimize this expression:

$$\min_{k_1} \left[(1-k_1)^2 p_{1|0} + k_1^2 \sigma_v^2\right]$$

Take the derivative w.r.t. k_1 and set it to zero:

$$-2(1-k_1)p_{1|0} + 2k_1\sigma_v^2 = 0 \implies k_1(p_{1|0}+\sigma_v^2) = p_{1|0} \implies \boxed{k_1 = \frac{p_{1|0}}{p_{1|0}+\sigma_v^2}}$$

Plug this into the expression for the posterior variance:

$$p_{1|1} = \mathbb{E}[(x_1 - x_{1|1})^2] = \left(1 - \frac{p_{1|0}}{p_{1|0}+\sigma_v^2}\right)^2 p_{1|0} + \left(\frac{p_{1|0}}{p_{1|0}+\sigma_v^2}\right)^2 \sigma_v^2 = p_{1|0}\left(1 - \frac{p_{1|0}}{p_{1|0}+\sigma_v^2}\right)$$

We now propagate forward:

$$p_{2|1} = \rho^2 p_{1|1} + \sigma_u^2$$

Apply the Kalman update at $t = 2$:

$$k_2 = \frac{p_{2|1}}{p_{2|1}+\sigma_v^2}$$

$$p_{2|2} = p_{2|1}\left(1 - \frac{p_{2|1}}{p_{2|1}+\sigma_v^2}\right)$$

Let $p_{t|t-1}$ be the prior variance at time t, and $p_{t|t}$ the posterior variance. Then, by an induction type argument, we have:

$$p_{t|t-1} = \rho^2 p_{t-1|t-1} + \sigma_u^2$$

$$k_t = \frac{p_{t|t-1}}{p_{t|t-1}+\sigma_v^2}$$

$$p_{t|t} = p_{t|t-1}\left(1 - \frac{p_{t|t-1}}{p_{t|t-1}+\sigma_v^2}\right)$$

The final form of the filtered estimate is:

$$x_{t|t} = (1-k_t)x_{t|t-1} + k_t z_t = x_{t|t-1} + k_t(z_t - x_{t|t-1}) = \rho x_{t-1|t-1} + k_t(z_t - \rho x_{t-1|t-1})$$

- If the observation noise variance is zero, i.e., $\sigma_v^2 = 0$, we observe a perfect signal $z_t = x_t$. Then:

  $$k_t = \frac{p_{t|t-1}}{p_{t|t-1}+0} = 1$$

  Interpretation: The Kalman filter fully trusts the signal, since it is noise-free. The update becomes:

  $$x_{t|t} = 0 \times x_{t|t-1} + 1 \times z_t = z_t.$$

- If the observation noise variance is infinite, i.e., $\sigma_v^2 = \infty$, then the signal z_t is uninformative. Then:

$$k_t = \frac{p_{t|t-1}}{p_{t|t-1} + \infty} = 0$$

Interpretation: The Kalman filter completely ignores the signal as it is "useless" and relies only on its prior prediction:

$$x_{t|t} = 1 \times x_{t|t-1} + 0 \times z_t = x_{t|t-1}$$

- If the process noise is infinitely large, i.e., $\sigma_u^2 = \infty$, then:

$$p_{t|t-1} = \rho^2 p_{t-1|t-1} + \sigma_u^2 \to \infty \Rightarrow k_t = \frac{\infty}{\infty + \sigma_v^2} \approx 1$$

Interpretation: The state evolves almost randomly, making prediction unreliable. The Kalman filter fully relies on the signal z_t regardless of its noise level. The update becomes:

$$x_{t|t} = z_t$$

3 Kalman Filter: Multivariate Case

We consider the multivariate linear Gaussian system:

$$X_t = AX_{t-1} + Cu_t, \quad u_t \sim \mathcal{N}(0, I)$$
$$Z_t = DX_t + v_t, \quad v_t \sim \mathcal{N}(0, \Sigma_v)$$

with (by analogy with the scalar case above, but now with careful attention to dimensions):

- $X_t \in \mathbb{R}^n$ is a latent state,
- $Z_t \in \mathbb{R}^l$ is an observed measurement,
- $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{n\times m}$, $D \in \mathbb{R}^{l\times n}$, $\Sigma_v \in \mathbb{R}^{l\times l}$.

Again, by analogy with scalar case, but now in matrix form:

- prior covariance: $P_{t|t-1} = \mathbb{E}[(X_t - X_{t|t-1})(X_t - X_{t|t-1})'] = \mathbb{E}\left[(X_t - \mathbb{E}(X_t \mid Z^{t-1}))(X_t - \mathbb{E}(X_t \mid Z^{t-1}))'\right]$,
- posterior covariance: $P_{t|t} = \mathbb{E}[(X_t - X_{t|t})(X_t - X_{t|t})'] = \mathbb{E}\left[(X_t - \mathbb{E}(X_t \mid Z^{t}))(X_t - \mathbb{E}(X_t \mid Z^{t}))'\right]$

The Kalman update rule is:

$$X_{t|t} = X_{t|t-1} + K_t(Z_t - DX_{t|t-1}).$$

Again, we want to choose the Kalman gain K_t to minimize the posterior error covariance $P_{t|t}$.

We compute the innovation as:

$$\tilde{Z}_t \equiv Z_t - DX_{t|t-1}$$

The innovation covariance is:

$$S_t = DP_{t|t-1}D' + \Sigma_v$$

The cross-covariance between the state and the innovation is:

$$\text{Cov}(X_t, \tilde{Z}_t) = P_{t|t-1}D'$$

Let $\tilde{X}_t \equiv X_t - X_{t|t-1}$, the prediction error. Then

$$X_t - X_{t|t} = \tilde{X}_t - K_t\tilde{Z}_t$$

Then the posterior error covariance is:

$$P_{t|t} = \mathbb{E}[(X_t - X_{t|t})(X_t - X_{t|t})'] = \mathbb{E}[(\tilde{X}_t - K_t\tilde{Z}_t)(\tilde{X}_t - K_t\tilde{Z}_t)']$$

We can expand it as

$$\begin{aligned} P_{t|t} &= \mathbb{E}\left[\tilde{X}_t\tilde{X}_t' - \tilde{X}_t\tilde{Z}_t'K_t' - K_t\tilde{Z}_t\tilde{X}_t' + K_t\tilde{Z}_t\tilde{Z}_t'K_t'\right] \\ &= P_{t|t-1} - K_t\underbrace{\mathbb{E}[\tilde{Z}_t\tilde{X}_t']}_{=DP_{t|t-1}} - \underbrace{\mathbb{E}[\tilde{X}_t\tilde{Z}_t']}_{=P_{t|t-1}D'}K_t' + K_t\underbrace{\mathbb{E}[\tilde{Z}_t\tilde{Z}_t']}_{=DP_{t|t-1}D'+\Sigma_v}K_t' \end{aligned}$$

So we have:

$$P_{t|t} = P_{t|t-1} - K_tDP_{t|t-1} - P_{t|t-1}D'K_t' + K_t\underbrace{(DP_{t|t-1}D' + \Sigma_v)}_{\equiv S_t}K_t'$$

Take derivative of $P_{t|t}$ with respect to K_t:

$$\frac{\partial P_{t|t}}{\partial K_t} = -2P_{t|t-1}D' + 2K_tS_t$$

Setting this to zero gives:

$$-2P_{t|t-1}D' + 2K_tS_t = 0 \quad \Rightarrow \quad K_tS_t = P_{t|t-1}D'$$

Solving for K_t yields the optimal Kalman gain:

$$\boxed{K_t = P_{t|t-1}D'(DP_{t|t-1}D' + \Sigma_v)^{-1}}$$

Intuition:

- If $\Sigma_v \to 0$: high trust in signals $\Rightarrow K_t \to D^{-1}$ if invertible.
- If $\Sigma_v \to \infty$: no trust in signals $\Rightarrow K_t \to 0$.

Note that as we define the posterior covariance:

$$P_{t|t} = \mathbb{E}\left[(X_t - \mathbb{E}(X_t \mid Z^t))(X_t - \mathbb{E}(X_t \mid Z^t))'\right],$$

we have $P_{t|t} = 0$, if and only if the state X_t is known exactly given observations up to time t, i.e., $X_t = \mathbb{E}(X_t \mid Z^t)$ almost surely. This happens when:

- The observation noise is zero: $\Sigma_v = 0$
- The observation matrix D is invertible, so that $X_t = D^{-1}Z_t$.

We can also write covariances in terms of parameter matrices:

- Prior covariance:

$$P_{t|t-1} = AP_{t-1|t-1}A' + CC'$$

- Posterior covariance:

$$P_{t|t} = P_{t|t-1} - K_t S_t K_t', \quad \text{where } S_t = DP_{t|t-1}D' + \Sigma_v$$

This allows us to think about their bounds:

Covariance	**Lower Bound**	**Upper Bound**	**Conditions**
$P_{t\|t}$	0	$P_{t\|t-1}$	LB: $\Sigma_v = 0$, D invertible UB: $\Sigma_v \to \infty$ or low rank D
$P_{t\|t-1}$	CC'	No general UB	LB: $P_{t-1\|t-1} = 0$ (see above)

Intuition:

- $P_{t|t-1}$: grows with model uncertainty, i.e., large process noise CC' or unstable dynamics (i.e., if eigenvalues of $A > 1$). However, no general finite upper bound exists (simple example is instability).

- $P_{t|t}$: grows with uninformative observations, i.e., large Σ_v, low-rank D.

4 Optional: Exam Practice

Try to answer the following questions:

- For the scalar process

$$x_t = \rho x_{t-1} + u_t$$
$$z_t = x_t + v_t$$
$$\begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{bmatrix}\right)$$

$$x_{0|0} = \bar{x}_0, \quad \mathbb{E}\left(\bar{x}_0 - x_0\right)^2 = p_{0|0}$$

1. Find the Kalman gain k_t such that $x_{t|t}$ is given by

$$x_{t|t} = \rho x_{t-1|t-1} + k_t \left[z_t - \rho x_{t-1|t-1}\right]$$

is the expected value of x_t conditional on $\bar{x}_0$ and the history of z_t.

2. What is k_t if $\sigma_v^2 = 0$? Interpret.

3. What is k_t if $\sigma_v^2 = \infty$? Interpret.

4. What is k_t if $\sigma_u^2 = \infty$? Interpret.

- Consider the state space system of the form

$$X_t = AX_{t-1} + Cu_t, \quad u_t \sim \mathcal{N}(0, I)$$
$$Z_t = DX_t + v_t$$

and define

$$P_{t|t-s} \equiv \mathbb{E}\left[X_t - \mathbb{E}(X_t \mid Z^{t-s})\right]\left[X_t - \mathbb{E}(X_t \mid Z^{t-s})\right]'$$

What restrictions on A, C, D and Σ_v would imply that:

1. $P_{t|t} = 0$?

2. What are the upper and lower bounds of $P_{t|t}$ and $P_{t|t-1}$? For what values of A, C, D and Σ_v would these bounds be attained?