LINEAR QUADRATIC ESTIMATOR

5.1: Setting up the optimal state estimator

- We now start to put the pieces together
  - Least-squares estimation and dynamic systems ➠ Observer.
- Discrete case first, then convert to continuous case.
- The filtering aspect will become apparent as we progress.

MODEL: Same as before

$$x_{k+1} = A_d x_k + w_k$$

$$y_k = C_d x_k + v_k.$$

- Periodic measurements of system behavior available, but corrupted by noise v_k ➠ sensor noise.
- Process noise w_k and sensor noise v_k are white. (If not, need to use shaping stuff from before and augment system vector x).

ASSUME:

$$\mathbb{E}[w_k] = 0; \quad \mathbb{E}[w_k w_j^T] = \Xi_w \Delta(k-j)$$

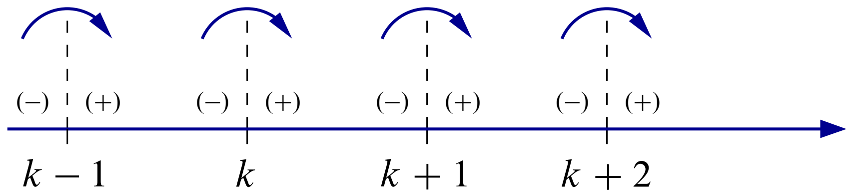
$$\mathbb{E}[v_k] = 0; \quad \mathbb{E}[v_k v_j^T] = \Xi_v \Delta(k-j)$$

$$\mathbb{E}[w_k v_j^T] = 0 \quad \forall\, k, j$$

$$\mathbb{E}[x_0] = \hat{x}_0.$$

GOAL: Use these periodic measurements of the system output to develop an optimal estimate of the state $x_k \to \hat{x}_k$, and develop a measure of confidence in this estimate.

TERMINOLOGY:



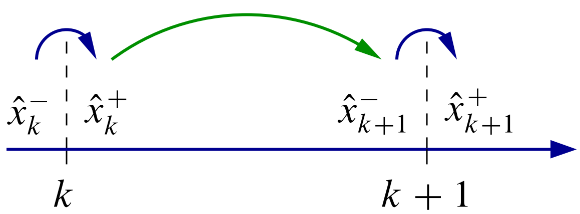
- State estimate is "$\hat{x}_k$". Two estimates possible and used:
  - One that uses all *PRIOR* information

$$\cdots y_{k-2}, y_{k-1} \Rightarrow \hat{x}_k^-.$$

  - One that uses all *POSSIBLE* information

$$\cdots y_{k-2}, y_{k-1}, y_k \Rightarrow \hat{x}_k^+.$$

- Let us analyze one segment



STEP I: *Measurement update:* Use y_k to go from $\hat{x}_k^-$ to $\hat{x}_k^+$.

STEP II: *Time update:* (propagation) Use the (assumed known) plant model to propagate statistics for $\hat{x}_k^+$ to $\hat{x}_{k+1}^-$.

- At each step, we have an estimation error:

$$\underbrace{\tilde{x}_k^+}_{\text{error}} = \underbrace{x_k}_{\text{truth}} - \underbrace{\hat{x}_k^+}_{\text{estimate}},$$

with associated error covariance matrix (assuming/enforcing $\mathbb{E}[\tilde{x}_k^+] = 0$).

$$\Xi_{x,k}^+ = \mathbb{E}\left[\tilde{x}_k^+(\tilde{x}_k^+)^T\right] = \mathbb{E}\left[(x_k - \hat{x}_k^+)(x_k - \hat{x}_k^+)^T\right].$$

- Want $\Xi^+_{x,k}$ to be "small".

STEP I: Assume we have the initial values $(\hat{x}_k^-, \Xi^-_{x,k})$.

- Now, use new measurement to find optimal estimates of $(\hat{x}_k^+, \Xi^+_{x,k})$ ➠ Key part of recursive solution.
- Enforce the constraint that the solution to the update problem $\hat{x}_k^- \to \hat{x}_k^+$ is a linear equation of the form

$$\hat{x}_k^+ = L'_k \hat{x}_k^- + L_k y_k.$$

- Updated estimate is a linear combination of previous estimate and most recent measurement.
- L_k, L'_k time varying matrices; different; yet to be determined.
- We are looking for the best *LINEAR* filter.
- If w_k, v_k are also Gaussian, it will also turn out this is the *BEST* filter possible. No nonlinear filter can do better!

5.2: Deriving the optimal state estimator

STEP I: Solution

- Assume initial estimate unbiased ➟ $\mathbb{E}[\tilde{x}_k^-] = 0$.
- Constrain solution so that $\mathbb{E}[\tilde{x}_k^+] = 0$ also.

$$\begin{aligned}
\mathbb{E}[x_k - \hat{x}_k^+] &= \mathbb{E}[x_k - L_k'\hat{x}_k^- - L_k y_k] \\
&= \mathbb{E}[x_k - L_k'\hat{x}_k^- - L_k y_k] + \overbrace{\mathbb{E}[L_k' x_k] - \mathbb{E}[L_k' x_k]}^{0} \\
\mathbb{E}[\tilde{x}_k^+] &= \mathbb{E}[(I - L_k' - L_k C_d)x_k + \underbrace{L_k' x_k - L_k'\hat{x}_k^-}_{L_k'\tilde{x}_k^-}] \\
&= \mathbb{E}[(I - L_k' - L_k C_d)x_k].
\end{aligned}$$

- To ensure that this estimate is unbiased for all x_k, must have $L_k' + L_k C_d = I$. Therefore

$$\begin{aligned}
\hat{x}_k^+ &= L_k'\hat{x}_k^- + L_k y_k \\
&= (I - L_k C_d)\hat{x}_k^- + L_k y_k \\
&= \hat{x}_k^- + L_k \underbrace{\left[y_k - C_d\hat{x}_k^-\right]}_{\text{Innovation}}
\end{aligned}$$

  where the innovation is the error in our estimate, or "what is new" in this measurement.

- Exactly the same form as we have seen before in the updates we have looked at! Need a good way to choose "blending factor" L_k.

Estimate covariance

- Since the estimation error has zero mean, the covariance matrix is $\Xi_{x,k} = \mathbb{E}[\tilde{x}_k\tilde{x}_k^T]$.

- Need to look at $\Xi^{-}_{\tilde{x},k}$ and $\Xi^{+}_{\tilde{x},k}$ to see how covariance is changed by a measurement.
- Then, pick L_k for "optimal blending".

Optimization

- We will choose L_k to minimize the sum of the squares of the estimation error vector:

$$J_k = \mathbb{E}[\left(\tilde{x}_k^+\right)^T \tilde{x}_k^+].$$

- Equivalent to minimizing the "length" squared of the error vector, which seems like an appealing objective.
- Note that we can rewrite the cost in terms of the error covariance matrix
$$J_k = \text{trace}[\Xi^{+}_{\tilde{x},k}]$$
(recall $x^T x = \text{trace}[x x^T]$).

Covariance update

- First, we must write $\Xi^{+}_{\tilde{x},k}$ in terms of $\Xi^{-}_{\tilde{x},k}$ and the measurement noise.
- Know $\hat{x}_k^+ = (I - L_k C_d)\hat{x}_k^- + L_k y_k$. Therefore, estimation error dynamics given by

$$\begin{aligned}
\tilde{x}_k^+ &= x_k - \hat{x}_k^+ \\
&= x_k - (I - L_k C_d)\hat{x}_k^- - L_k(C_d x_k + v_k) \\
&= (I - L_k C_d)\tilde{x}_k^- - L_k v_k.
\end{aligned}$$

- Now, form

$$\begin{aligned}\Xi_{x,k}^{+} &= \mathbb{E}[\tilde{x}_k^{+}\left(\tilde{x}_k^{+}\right)^T] \\ &= \mathbb{E}[\{(I - L_kC_d)\tilde{x}_k^{-} - L_kv_k\}\{(I - L_kC_d)\tilde{x}_k^{-} - L_kv_k\}^T].\end{aligned}$$

- As before, three terms,

1. $\mathbb{E}[(I - L_kC_d)\tilde{x}_k^{-}\left(\tilde{x}_k^{-}\right)^T(I - L_kC_d)^T] = (I - L_kC_d)\Xi_{x,k}^{-}(I - L_kC_d)^T$.
2. $\mathbb{E}[L_kv_kv_k^TL_k^T] = L_k\Xi_vL_k^T$.
3. $\mathbb{E}[\tilde{x}_k^{-}v_k^T] = 0$ since prior estimate error is uncorrelated with the measurement noise at this time step.

- Therefore, for any L_k,

$$\Xi_{x,k}^{+} = (I - L_kC_d)\Xi_{x,k}^{-}(I - L_kC_d)^T + L_k\Xi_vL_k^T.$$

- These two equations tell us how to update $\tilde{x}_k$ and $\Xi_{x,k}$, given any L_k.
- Now, optimize to find best L_k.
- The "trace" operator makes life quite easy since we have

$$\frac{\mathrm{d}\,\mathrm{trace}(AB)}{\mathrm{d}A} = B^T, \qquad AB \text{ square}$$

$$\frac{\mathrm{d}\,\mathrm{trace}(ACA^T)}{\mathrm{d}A} = 2AC, \qquad C \text{ symmetric.}$$

- Form $\dfrac{\mathrm{d}J_k}{\mathrm{d}L_k} = 0$ for optimal gain. Recall

$$\begin{aligned}J_k &= \mathbb{E}[\tilde{x}_k^{T+}\tilde{x}_k^{+}] = \mathrm{trace}(\Xi_{x,k}^{+}) \\ &= \mathrm{trace}\left((I - L_kC_d)\Xi_{x,k}^{-}(I - L_kC_d)^T\right) + \mathrm{trace}(L_k\Xi_vL_k^T)\end{aligned}$$

- Then,

$$\frac{\mathrm{d}J_k}{\mathrm{d}L_k} = -2\Xi_{x,k}^{-}C_d^T + 2L_k(C_d\Xi_{x,k}^{-}C_d^T + \Xi_v) = 0$$

which gives

$$L_k^* = \Xi_{\tilde{x},k}^- C_d^T \left[C_d \Xi_{\tilde{x},k}^- C_d^T + \Xi_v \right]^{-1}$$

which is the <u>Kalman gain matrix</u>—minimizing gain.

- Can show that with this gain

$$\begin{aligned}\Xi_{\tilde{x},k}^+ &= (I - L_k^* C_d)\Xi_{\tilde{x},k}^- \\ &= \begin{bmatrix}\text{The optimized value of the updated} \\ \text{error covariance matrix.}\end{bmatrix}\end{aligned}$$

- Thus, in terms of the mean-square estimation error,

$$L_k = \Xi_{\tilde{x},k}^- C_d^T \left[C_d \Xi_{\tilde{x},k}^- C_d^T + \Xi_v \right]^{-1}$$

  is the optimal blending ratio ➠ Very similar form to the gain matrices studied earlier.

(End of STEP I, measurement update).

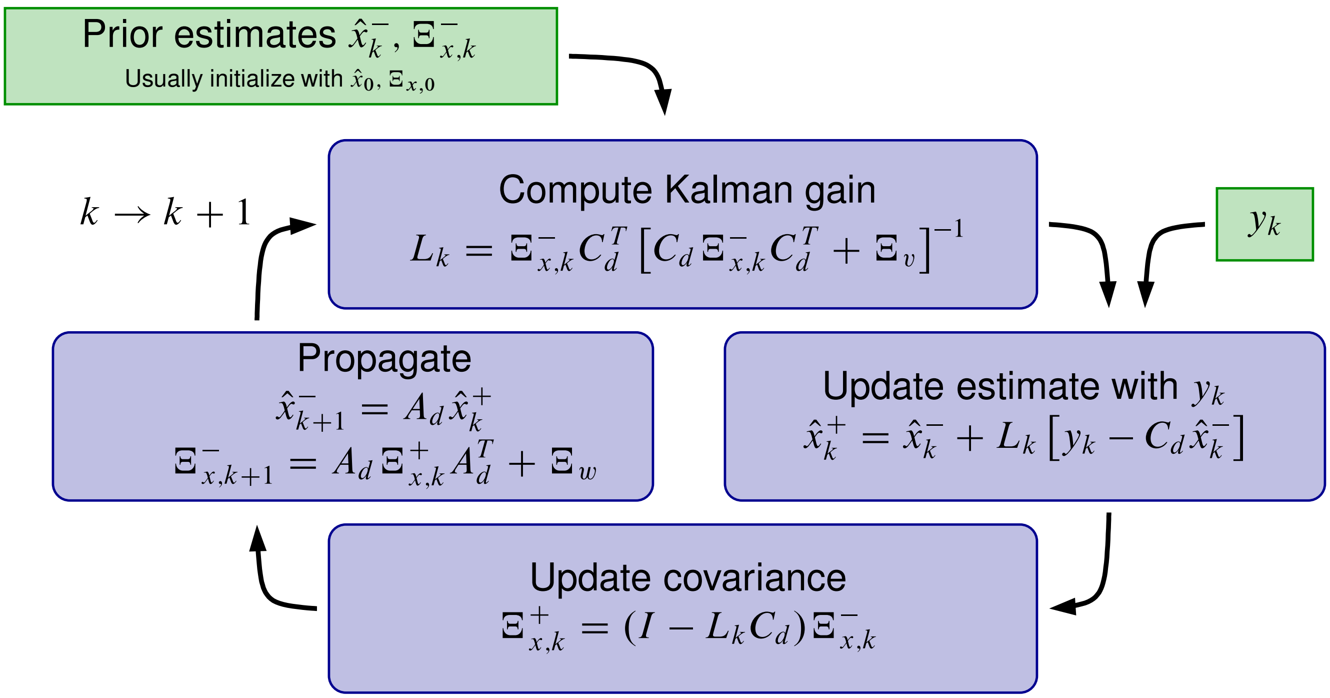
STEP II: Time update (propagation).

- Have developed update estimates at time "k^+".
- Need to propagate to develop predictions of the best estimate at time "$(k+1)^-$".
- Have already studied this case. Using known model and process noise covariance,

$$\begin{aligned}\hat{x}_{k+1}^- &= A_d \hat{x}_k^+ \qquad (+\text{ deterministic input}) \\ \Xi_{\tilde{x},k+1}^- &= A_d \Xi_{\tilde{x},k}^+ A_d^T + \Xi_w.\end{aligned}$$

5.3: Kalman filter loop

- We can combine the parts that make up this one segment $(\cdot)_k^- \Rightarrow (\cdot)_{k+1}^-$ to form a recursion:



- "Simple" to perform on a digital computer.

EXAMPLE: $\dot{x}(t) = w(t)$, $\Delta t = 1$, $S_w = 1$.

- Independent noisy measurements $y(t) = x(t) + v(t)$.
- Standard deviation of measurement noise $\sigma_v = 1/2$.
- Convert continuous to discrete

$$x_k = A_d x_{k-1} + w_{k-1}, \qquad \mathbb{E}[w_k w_l] = \Xi_w \Delta(k-l).$$

  - Find A_d from zero-order-hold equivalent: $A_d = 1$.
  - Find Ξ_w using $\Xi_w = \displaystyle\int_0^1 e^0 S_w e^0 \, dt = 1$.
  - Measurement equation, $y_k = C_d x_k + v_k$.

$$\mathbb{E}[v_k v_k] = \Xi_v = \sigma_v^2 = 1/4, \qquad C_d = 1.$$

  - Assume that there is no initial uncertainty. $\hat{x}_0^- = 0$ and $\Xi_{x,0}^- = 0$.

Time step 0:

- $L_0 = \Xi^-_{x,0} C_d^T (C_d \Xi^-_{x,0} C_d^T + \Xi_v)^{-1} = 0.$
- $\hat{x}_0^+ = \hat{x}_0^- + L_0(y_0 - C_d \hat{x}_0^-) = \hat{x}_0^- = 0.$
- $\Xi^+_{x,0} = (I - L_0 C_d)\Xi^-_{x,0} = \Xi^-_{x,0} = 0.$
- $\hat{x}_1^- = A_d \hat{x}_0^+ = 0$ and $\Xi^-_{x,1} = A_d \Xi^+_{x,0} A_d^T + \Xi_w = 0 + 1 = 1.$

Time step 1:

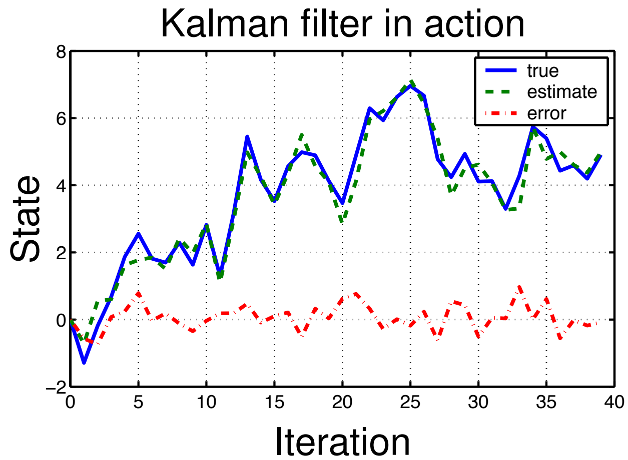
- $L_1 = \Xi^-_{x,1} C_d^T (C_d \Xi^-_{x,1} C_d^T + \Xi_v)^{-1} = 1 \cdot 1(1 \cdot 1 \cdot 1 + 1/4)^{-1} = 4/5.$
- $\hat{x}_1^+ = \hat{x}_1^- + L_1(y_1 - C_d \hat{x}_1^-) = 0 + (4/5)y_1 = (4/5)y_1.$
- $\Xi^+_{x,1} = (I - L_1 C_d)\Xi^-_{x,1} = 1/5.$
- $\hat{x}_2^- = A_d \hat{x}_1^+ = (4/5)y_1$ and $\Xi^-_{x,2} = A_d \Xi^+_{x,1} A_d^T + \Xi_w = (1/5) + 1 = 6/5.$

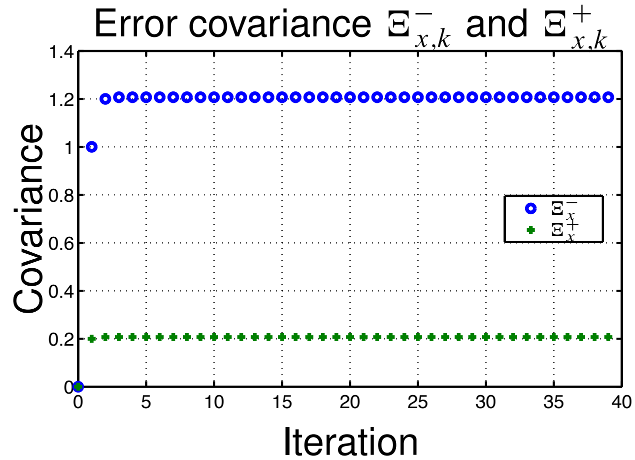
Time step 2:

- $L_2 = \Xi^-_{x,2} C_d^T (C_d \Xi^-_{x,2} C_d^T + \Xi_v)^{-1} = (6/5)(6/5 + 1/4)^{-1} = 24/29.$
- $\hat{x}_2^+ = \hat{x}_2^- + L_2(y_2 - C_d \hat{x}_2^-) = (4/29)y_1 + (24/29)y_2.$
- $\Xi^+_{x,2} = 6/29$ and so forth, hopefully to steady state.

$$\begin{array}{ccccccccccc}
\Xi^-_{x,0} = 0 & & \Xi^-_{x,1} = 1 & & \Xi^-_{x,2} = 1.2 & & \Xi^-_{x,3} = 1.207 & \\
\downarrow & \nearrow & \downarrow & \nearrow & \downarrow & \nearrow & \downarrow & \nearrow \\
\Xi^+_{x,0} = 0 & & \Xi^+_{x,1} = 0.2 & & \Xi^+_{x,2} = 0.207 & & \Xi^+_{x,3} = 0.2071 &
\end{array}$$

- Covariance increases during propagation and is then reduced by each measurement.
- Covariance still oscillates at steady state between $\Xi^-_{x,\text{ss}}$ and $\Xi^+_{x,\text{ss}}$.






```
% Kalman Filter Example, modified (slightly) from Claire Tomlin's code
a=1; c=1; % System dynamics
Xi_v=1/4; Xi_w=1;% White noise covariance
randn('seed',sum(40*clock)); % Init rand # generator
w = sqrt(Xi_w)*randn(40,1); % Generate disturbance
v = sqrt(Xi_v)*randn(40,1);% Generate noise

x(1) = 0;% init state for simulation
xhatminus(1) = 0;% init prior state info
Xi_x_minus(1) = 0;% init prior error covariance

for k = 1:40,
  % Compute Kalman gain
  L(k) = Xi_x_minus(k)*c'*inv(c*Xi_x_minus(k)*c'+Xi_v);

  x(k+1) = a*x(k)+w(k); % compute noisy state dynamic
  y(k) = c*x(k)+v(k); % simulate noisy measurements

  % Measurement update with y(k)
  xhatplus(k) = xhatminus(k)+L(k)*(y(k)-c*xhatminus(k));

  % Update covariance
  Xi_x_plus(k) = (1-L(k)*c)*Xi_x_minus(k);

  % Propagate
  xhatminus(k+1) = a*xhatplus(k);
  Xi_x_minus(k+1) = a*Xi_x_plus(k)*a'+Xi_w;
end;

figure(1); clf; k=0:39;
plot(k,x(1:40),'-',k,xhatplus(1:40),'--',0:39,x(1:40)-xhatplus,'-.');
legend('true','estimate','error'); title('Kalman filter in action');
```



```
xlabel('Iteration'); ylabel('State'); grid;

figure(2); clf;
plot(k,Xi_x_minus(1:40),'o',k,Xi_x_plus(1:40),'+');
grid; legend('Xi_x^{-}','Xi_x^{+}',0);
title('Error covariance before (o) and after ^{+} meas. update');
xlabel('Iteration'); ylabel('Covariance');
```

Discrete Kalman filter versus RLS

KALMAN: First the covariance matrix:

$$\begin{aligned}\Xi_{x,k}^{+} &= (I - L_k C_d)\Xi_{x,k}^{-} \\ &= \Xi_{x,k}^{-} - \Xi_{x,k}^{-} C_d^T \left[C_d \Xi_{x,k}^{-} C_d^T + \Xi_v\right]^{-1} C_d \Xi_{x,k}^{-}.\end{aligned}$$

- Now, use matrix inversion lemma (proof in book appendix)

$$\left[\bar{A}^{-1} + \bar{B}^T \bar{C}^{-1} \bar{B}\right]^{-1} = \bar{A} - \bar{A}\bar{B}^T \left[\bar{B}\bar{A}\bar{B}^T + \bar{C}\right]^{-1} \bar{B}\bar{A}$$

  where we substitute

$$\bar{A} = \Xi_{x,k}^{-} \qquad \bar{B} = C_d \qquad \bar{C} = \Xi_v.$$

- Then

$$\left[\Xi_{x,k}^{+}\right]^{-1} = \left[\Xi_{x,k}^{-}\right]^{-1} + C_d^T \Xi_v^{-1} C_d.$$

- For the gain, use the following form of the matrix inversion lemma:

$$\left[\bar{A}^{-1} + \bar{B}^T \bar{C}^{-1} \bar{B}\right]^{-1} \bar{B}^T \bar{C}^{-1} = \bar{A}\bar{B}^T \left[\bar{B}\bar{A}\bar{B}^T + \bar{C}\right]^{-1}$$

  with the same substitutions (proof in notes appendix).

- So:

$$\begin{aligned}L_k &= \Xi_{x,k}^{-} C_d^T \left[C_d \Xi_{x,k}^{-} C_d^T + \Xi_v\right]^{-1} \\ &= \left[\left(\Xi_{x,k}^{-}\right)^{-1} + C_d^T \Xi_v^{-1} C_d\right]^{-1} C_d^T \Xi_v^{-1} \\ &= \Xi_{x,k}^{+} C_d^T \Xi_v^{-1}.\end{aligned}$$

RLS: Covariance update was

$$Q_{k+1}^{-1} = Q_k^{-1} + H_{k+1}^T R_{k+1}^{-1} H_{k+1}.$$

- The gain was

$$\text{gain} = Q_{k+1} H_{k+1}^T R_{k+1}^{-1}.$$

- Therefore, measurement update step of Kalman filter is the same as the recursive form of the maximum-likelihood estimate (when v in $y = Hx + v$ is Gaussian, $\hat{x}_{ML}$ is linear).
  - Therefore, if w_k is Gaussian, a linear filter is the best filter possible! No nonlinear filter can do better!
- Time update is additional step necessary when estimating a moving (dynamic) state vector.

5.4: Continuous-time filters

GOAL: Develop an estimator $\hat{x}(t)$ that is a linear function of measurements $y(\tau)$, $0 \le \tau \le t$, and minimizes the function

$$\mathbb{E}\left[(x(t)-\hat{x}(t))^T(x(t)-\hat{x}(t))\right].$$

ASSUMED: Dynamics of linear system are

$$\dot{x}(t) = Ax(t) + B_w w(t)$$

$$y(t) = C_y x(t) + v(t)$$

and $\mathbb{E}[w(t)] = \mathbb{E}[v(t)] = 0$; $\mathbb{E}[w(t_1)w(t_2)^T] = S_w\delta(t_1-t_2)$, $\mathbb{E}[v(t_1)v(t_2)^T] = S_v\delta(t_1-t_2)$, $\mathbb{E}[w(t_1)v(t_2)] = 0$, $S_w > 0$, $S_v > 0$, $\hat{x}(0)$ and $\Xi_x(0)$ given.

APPROACH: As before, analyze discrete case as $\Delta t \to 0$.

- To proceed, we must relate the discrete and continuous sensor noises. (We already saw how to relate process noises).
- One approach: Use discrete noise (Δt small)

$$v_k = \frac{1}{\Delta t}\int_0^{\Delta t} v(t)\,\mathrm{d}t \qquad \text{(average value)}.$$

- Then,

$$\begin{aligned}\Xi_v = \mathbb{E}[v_k v_k^T] &= \frac{1}{\Delta t^2}\int_0^{\Delta t}\int_0^{\Delta t}\mathbb{E}[v(\gamma)v(\tau)^T]\,\mathrm{d}\gamma\,\mathrm{d}\tau \\ &= \frac{1}{\Delta t^2}\cdot\Delta t\cdot S_v = \frac{S_v}{\Delta t}.\end{aligned}$$

- As $\Delta t \to 0$, the discrete noise tends to infinite pulse covariance with weight S_v, similar to $S_v\delta(t)$.

Conversion

- Recall
  - $\Xi^-_{x,k+1} = A_d \, \Xi^+_{x,k} A_d^T + \Xi_w.$
  - $\Xi^+_{x,k} = (I - L_k C_d) \Xi^-_{x,k}.$
- With malice aforethought, consider

$$\begin{aligned}\frac{L_k}{\Delta t} &= \frac{\Xi^-_{x,k} C_d^T}{\Delta t}\left[C_d \, \Xi^-_{x,k} C_d^T + \Xi_v\right]^{-1}\\ &= \Xi^-_{x,k} C_d^T\left[C_d \, \Xi^-_{x,k} C_d^T \Delta t + \Xi_v \Delta t\right]^{-1}.\end{aligned}$$

- As $\Delta t \to 0$, if $\Xi^-_{x,k}$ finite,

$$\left.\begin{array}{l} C_d \, \Xi^-_{x,k} C_d^T \Delta t \to 0 \\ \Xi_v \Delta t \to S_v \\ \Xi^-_{x,k} \to \Xi_x(t) \\ C_d \to C_y \end{array}\right\} \begin{array}{l}\text{as } \Delta t \to 0 \\ \dfrac{L_k}{\Delta t} \to L(t) = \Xi_x(t) C_y^T S_v^{-1}.\end{array}$$

- Already looked at propagation case

$$\begin{aligned}\Xi^-_{x,k+1} &= A_d \, \Xi^+_{x,k} A_d^T + \Xi_w\\ &\approx (I + A\Delta t)\Xi^+_{x,k}(I + A\Delta t)^T + B_w S_w B_w^T \Delta t\\ &= \Xi^+_{x,k} + \Delta t\left[A\,\Xi^+_{x,k} + \Xi^+_{x,k} A^T + B_w S_w B_w^T\right] + o(\Delta t^2),\end{aligned}$$

but $\Xi^+_{x,k} = \Xi^-_{x,k} - L_k C_d \, \Xi^-_{x,k}$ ...combine...

$$\begin{aligned}\Xi^-_{x,k+1} = \Xi^-_{x,k} - L_k C_d \, \Xi^-_{x,k}&\\ +\Delta t\Big[A\,\Xi^-_{x,k} - A L_k C_d \, \Xi^-_{x,k} + \Xi^-_{x,k} A^T&\\ -L_k C_d \, \Xi^-_{x,k} A^T + B_w S_w B_w^T\Big]&\\ +o(\Delta t^2).&\end{aligned}$$

- Rearrange...

$$\frac{\Xi^-_{x,k+1} - \Xi^-_{x,k}}{\Delta t} = \frac{-L_k}{\Delta t} C_d \Xi^-_{x,k} + A\Xi^-_{x,k} + \Xi^-_{x,k} A^T + B_w S_w B_w^T$$
$$-AL_k C_d \Xi^-_{x,k} - L_k C_d \Xi^-_{x,k} A^T + o(\Delta t).$$

- As $\Delta t \to 0$, LHS $\to \dot{\Xi}_x(t)$, $C_d \to C_y$, $\dfrac{L_k}{\Delta t} \to L(t)$, $\Xi^-_{x,k} \to \Xi_x(t)$ and $AL_k C_d \Xi^-_{x,k} \to 0$.

Resulting expression

$$\dot{\Xi}_x(t) = \overbrace{\underbrace{A\Xi_x(t) + \Xi_x(t)A^T + B_w S_w B_w^T}_{\text{Lyapunov for propagation}} - \underbrace{\Xi_x(t) C_y^T S_v^{-1} C_y \Xi_x(t)}_{\geq 0}}^{\text{Riccati for measurement update}}.$$

- Matrix *RICCATI* equation for error covariance.
- Equation includes the following impact on $\dot{\Xi}_x(t)$:
  - $A\Xi_x(t) + \Xi_x(t)A^T$ ➠ Homogeneous part.
  - $B_w S_w B_w^T$ ➠ Increase due to process noise.
  - $\Xi_x(t) C_y^T S_v^{-1} C_y \Xi_x(t)$ ➠ Decrease due to measurements.

Culture

Count Jacopo Francesco Riccati (1676–1754) was an Italian savant who wrote on mathematics, physics, and philosophy. He was chiefly responsible for introducing the ideas of Newton to Italy. At one point he was offered the presidency of the St. Petersburg Academy of Sciences but understandably he preferred the leisure and comfort of his aristocratic life in Italy to administrative responsibilities in Russia. Though widely known in scientific circles of his time, he now survives only through the differential equation bearing his name. Even this was an accident of history, for Riccati merely discussed special cases of this equation without offering any solutions, and most of these special cases were successfully treated by various members of the Bernoulli family. The details of this complex story can be found in G. N. Watson "A treatise on the Theory of Bessel Functions." 2d ed., pp. 1–3, Cambridge, London, 1944. The special Riccati equation $\dot{y} + by^2 = cx^m$ is known to be solvable in finite terms if and only if the exponent m is -2 or of the form $-4k/(2k+1)$ for some integer k.

(Taken from "Differential Equations with Applications and Historical Notes" by George F. Simmons).

State-Estimate Update

- Recall

1. $\hat{x}_k^- = A_d \hat{x}_{k-1}^+$.
2. $\hat{x}_k^+ = \hat{x}_k^- + L_k(y_k - C_d \hat{x}_k^-)$.

- Substitute (1) into (2)

$$\hat{x}_k^+ = A_d \hat{x}_{k-1}^+ + L_k(y_k - C_d A_d \hat{x}_{k-1}^+).$$

- Substitute:

$$\hat{x}_k^+ = (I + A\Delta t)\hat{x}_{k-1}^+ + L_k\Big(y_k - C_d(I + A\Delta t)\hat{x}_{k-1}\Big).$$

- So,

$$\frac{\hat{x}_k^+ - \hat{x}_{k-1}^+}{\Delta t} = A\hat{x}_{k-1}^+ + \frac{L_k}{\Delta t}\Big(y_k - C_d \hat{x}_{k-1}^+ - C_d A \Delta t \hat{x}_{k-1}^+\Big).$$

- Then,

$$\dot{\hat{x}}(t) = A\hat{x}(t) + L(t)\left[y(t) - C_y \hat{x}(t)\right]$$

$$L(t) = \Xi_x(t) C_y^T S_v^{-1}$$

$$\dot{\Xi}_x(t) = A\Xi_x(t) + \Xi_x(t)A^T + B_w S_w B_w^T - \Xi_x(t) C_y^T S_v^{-1} C_y \Xi_x(t).$$

- This is called a "Kalman–Bucy" filter, or a continuous-time Kalman filter/estimator.

EXAMPLE: Estimate the value of a constant given continuous measurements of x corrupted by Gaussian white noise.

$$\dot{x} = 0 \qquad \text{(trivial dynamics of constant)}$$

$$y = x + v,$$

where $v \sim \mathcal{N}(0, S_v)$. Then $A = B_w = S_w = 0$.

- The Riccati equation becomes: $\dot{\Xi}_x(t) = -\Xi_x^2(t)/S_v$.
- Integrate:

$$\int_{\Xi_{x,0}}^{\Xi_x(t)} \frac{\mathrm{d}\Xi_x(\tau)}{\Xi_x^2(\tau)} = -\frac{1}{S_v}\int_0^t \mathrm{d}\tau$$

$$-\Xi_x^{-1}(\tau)\Big|_{\Xi_{x,0}}^{\Xi_x(t)} = -t/S_v$$

$$\frac{1}{\Xi_{x,0}} - \frac{1}{\Xi_x(t)} = -t/S_v$$

$$\frac{1}{\Xi_{x,0}} + \frac{t}{S_v} = \frac{1}{\Xi_x(t)}$$

$$\Xi_x(t) = \frac{\Xi_{x,0}S_v}{S_v + \Xi_{x,0}t};$$

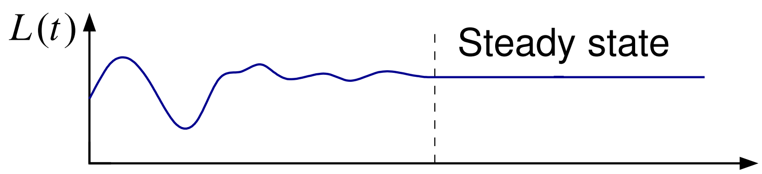
$$L(t) = \Xi_x(t)/S_v = \frac{\Xi_{x,0}}{S_v + \Xi_{x,0}t}.$$

- As $t \to \infty$, $L(t) \to 0$, so that $\hat{x}(t) \to$ constant.

$$\dot{\hat{x}}(t) = \underbrace{\frac{\Xi_{x,0}}{S_v + \Xi_{x,0}t}}_{L(t)}[y(t) - \hat{x}(t)] \to 0.$$

5.5: Steady-state Kalman filters

- Several examples have begun to demonstrate that the filter gains can have a transient and then settle to constant values.



- Since the estimator will typically be running for a long time, most of the operations will occur with the filter in the steady-state region.
- Therefore, recursive/differential equations do not need to be solved!

Steady-state solution

- The Kalman gain depends on state covariance, which for a discrete-time system has two steady-state solutions: $\Xi^-_{x,ss}$ and $\Xi^+_{x,ss}$. Then,

$$L_{ss} = \Xi^+_{x,ss} C_d^T \Xi_v^{-1}.$$

- Consider the filter loop as $k \to \infty$

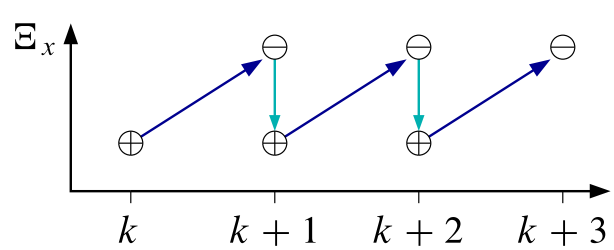
$$\Xi^-_{x,ss} = A_d \Xi^+_{x,ss} A_d^T + \Xi_w$$

$$\Xi^+_{x,ss} = \Xi^-_{x,ss} - \Xi^-_{x,ss} C_d^T \left[C_d \Xi^-_{x,ss} C_d^T + \Xi_v \right]^{-1} C_d \Xi^-_{x,ss}.$$

- Combine these two to get

$$\Xi^-_{x,ss} = \Xi_w + A_d \Xi^-_{x,ss} A_d^T - A_d \Xi^-_{x,ss} C_d^T \left[C_d \Xi^-_{x,ss} C_d^T + \Xi_v \right]^{-1} C_d \Xi^-_{x,ss} A_d^T.$$

- One matrix equation, one matrix unknown ➠ `dlqe.m`



EXAMPLE: $x_{k+1} = x_k + w_k$. $y_k = x_k + v_k$. Steady-state covariance?

- Let $\mathbb{E}[w_k] = \mathbb{E}[v_k] = 0$, $\mathbb{E}[w_k w_l] = 1\Delta(k-l)$, $\mathbb{E}[v_k v_l] = 2\Delta(k-l)$
- $\mathbb{E}[x_0] = \hat{x}_0$ and $\mathbb{E}[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)] = \Xi_{x,0}$.
- Notice $A_d = 1$, $C_d = 1$, $\Xi_w = 1$ and $\Xi_v = 2$.

$$\left(\Xi^+_{x,k}\right)^{-1} = \left(\Xi^-_{x,k}\right)^{-1} + C_d^T \Xi_v^{-1} C_d$$

$$\Xi^+_{x,k} = \frac{1}{\frac{1}{\Xi^-_{x,k}} + \frac{1}{2}} = \frac{2\Xi^-_{x,k}}{2 + \Xi^-_{x,k}}.$$

$$\Xi^-_{x,k+1} = A_d \Xi^+_{x,k} A_d^T + \Xi_w = \Xi^+_{x,k} + 1 = \frac{2\Xi^-_{x,k}}{2 + \Xi^-_{x,k}} + 1$$

$$\Xi^-_{x,\text{ss}} = \frac{2\Xi^-_{x,\text{ss}}}{2 + \Xi^-_{x,\text{ss}}} + 1$$

which gives us $\Xi^-_{x,\text{ss}} = 2$ and $\Xi^+_{x,\text{ss}} = 1$.

$$\begin{array}{ccccccccccc}
\Xi^-_{x,0} = \infty & & \Xi^-_{x,1} = 3 & & \Xi^-_{x,2} = 11/5 & & \Xi^-_{x,3} = 43/21 & \\
\downarrow & \nearrow & \downarrow & \nearrow & \downarrow & \nearrow & \downarrow & \nearrow \\
\Xi^+_{x,0} = 2 & & \Xi^+_{x,1} = 6/5 & & \Xi^+_{x,2} = 22/21 & & \Xi^+_{x,3} = 86/85 &
\end{array}$$

Steady-state in continuous time

- The same kind of result often holds for LTI continuous-time systems.
- The steady-state Kalman gain is associated with $\dot{\Xi}_x(t) \to 0$, giving the algebraic Riccati equation (ARE)

$$A\Xi_x + \Xi_x A^T + B_w S_w B_w^T - \Xi_x C_y^T S_v^{-1} C_y \Xi_x = 0$$

and

$$L = \Xi_x C_y^T S_v^{-1}.$$

- Same as time-varying case, but with constant Ξ_x and L. `care.m`
- Tradeoff between sensor and process noise now explicit: $L = \Xi_x C_y^T S_v^{-1}$.
  - If uncertain about estimate (Ξ_x large), then innovation $(y - C_y\hat{x})$ weighted heavily (big L).
  - If S_v small (measurements accurate) then new measurements highly weighted (big L).
- We can think of $\Xi_x C_y^T S_v^{-1}$ as analogous to SNR.

NOTE: Continuous case requires $S_v > 0$ (S_v^{-1} explicit) but discrete does not (Ξ_v^{-1} can be avoided).

Stability of steady-state Kalman filter

- Steady-state gain is much easier to compute/use than time-varying solution.
- Small performance reduction if we use steady-state value for *ALL* time since transient region is quite short (relatively). But, what can we say about stability?

CTS SYSTEM:

$$\dot{x}(t) = Ax(t) + B_w w(t)$$

$$y(t) = C_y x(t) + v(t)$$

where $w \sim \mathcal{N}(0, S_w)$ and $v \sim \mathcal{N}(0, S_v)$; w, v white and independent.

ASSUME:

1. $S_v > 0$, $S_w > 0$.

2. All plant dynamics constant
3. $[A, C_y]$ detectable (unstable modes observable)
4. $[A, B_w]$ stabilizable (unstable modes controllable)

- Time-varying Kalman gains for this time-invariant system satisfy

$$\dot{\Xi}_x = A\Xi_x + \Xi_x A^T + B_w S_w B_w^T - \Xi_x C_y^T S_v^{-1} C_y \Xi_x$$
$$L = \Xi_x C_y^T S_v^{-1},$$

$\Xi_{x,0}$ given.

- With the above assumptions, $\Xi_x(t) \to \Xi_{x,\mathrm{ss}}$ regardless of $\Xi_{x,0}$ as $t \to \infty$, where $\Xi_{x,\mathrm{ss}}$ solves

$$0 = A\Xi_{x,\mathrm{ss}} + \Xi_{x,\mathrm{ss}} A^T + B_w S_w B_w^T - \Xi_{x,\mathrm{ss}} C_y^T S_v^{-1} C_y \Xi_{x,\mathrm{ss}}.$$

- Stabilizable/detectable implies $\Xi_{x,\mathrm{ss}} \ge 0$ (unique)
- $\Xi_{x,\mathrm{ss}} > 0$ iff $[A, B_w]$ controllable.
- If $\Xi_{x,\mathrm{ss}}$ exists, the steady-state optimal estimator

$$\dot{\hat{x}} = A\hat{x} + L_{\mathrm{ss}}(y(t) - C_y \hat{x}(t))$$
$$L_{\mathrm{ss}} = \Xi_{x,\mathrm{ss}} C_y^T S_v^{-1}$$

is asymptotically stable if (1)–(4) hold.

Big conclusion

1. Time-varying Kalman filter gains provide *PERFORMANCE OPTIMAL* filter design, but,
2. Under *MILD* conditions, we can show that the closed-loop estimation error dynamics are stable when we use the steady-state gains *for all time* ➠ *BIG* computational savings.

5.6: Frequency-domain interpretation

- We are interested in finding the estimation-error dynamics. We start with the plant dynamics:

$$\dot{x} = Ax + B_u u + B_w w$$

$$y = C_y x + v$$

where $w \sim \mathcal{N}(0, S_w)$, $B_w S_w B_w^T > 0$, $v \sim \mathcal{N}(0, S_v)$, $S_v > 0$; w, v white and independent, $[A, C_y]$ observable.

- Strong form of assumptions.
- Steady-state Kalman filter:

$$\dot{\hat{x}} = A\hat{x} + B_u u + L(y - C_y \hat{x}).$$

- This gives estimation error dynamics: $\tilde{x} = x - \hat{x}$

$$\dot{\tilde{x}} = \dot{x} - \dot{\hat{x}} = Ax + B_w w - \big[(A - LC_y)\hat{x} + L(C_y x + v)\big]$$

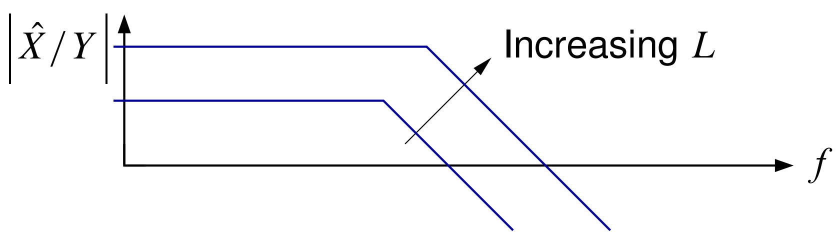
$$= (A - LC_y)\tilde{x} + B_w w - Lv.$$

- Filter stability governed by eigenvalues of $(A - LC_y)$.
- Equation makes explicit the tradeoff between:
  - Speed of estimator error decay (L big, eig$(A - LC_y)$ far in LHP)
  - Susceptibility of estimator error to being corrupted by sensor noise. (L small so Lv small).
- Kalman filter selects the *optimal balance* between these two goals.
- Have $\dot{\hat{x}} = A\hat{x} + L(y - C_y \hat{x})$ … (steady-state filter)

- Laplace transform both sides... (assume scalar state for now)

$$\frac{\hat{X}(s)}{Y(s)} = \frac{L}{s - A + LC_y} = \frac{\frac{L}{LC_y - A}}{\frac{s}{LC_y - A} + 1}.$$

- Pole at $s = -(LC_y - A)$ and dc-gain of $L/(LC_y - A)$.
- This is the transfer function of the filter applied to the measurements to form the estimate $\hat{x}$ (low-pass).
- Increasing L pushes the filter transfer function up and out.



- Eventually, estimate would be too corrupted by the noise in the measurements.
- Note that balancing the sensor-noise impact done with respect to the process noise $(B_w w)$.
- It turns out that the ratio S_w/S_v plays a key role in the selection of L.

EXAMPLE: System driven by noise

$$\dot{x} = w$$

$$y = x + v$$

where $\mathbb{E}[w] = \mathbb{E}[v] = 0$ and w and v independent; $\mathbb{E}[w(t)w(t+\tau)] = S_w\delta(\tau)$ and $\mathbb{E}[v(t)v(t+\tau)] = S_v\delta(\tau)$.

- $\mathbb{E}[x(0)] = 0$, $\mathbb{E}[x(0)^2] = \Xi_x(0)$.

- The optimal filter is

$$\dot{\hat{x}} = \frac{\Xi_x(t)}{S_v}(y - \hat{x})$$

$$\Xi_x(t) = \sqrt{S_w S_v}\frac{1 + be^{-2\alpha t}}{1 - be^{-2\alpha t}},$$

where $\alpha = \sqrt{S_w/S_v}$ and $b = (\Xi_x(0) - \sqrt{S_w S_v})/(\Xi_x(0) + \sqrt{S_w S_v})$.

- Convergence speed of $\Xi_x(t)$ determined by S_w/S_v.
- Steady-state:

$$\Xi_x(t) \to \Xi_{x,\text{ss}} = \sqrt{S_w S_v}$$

$$L_{\text{ss}} = \frac{\Xi_{x,\text{ss}}}{S_v} = \sqrt{S_w/S_v}.$$

Therefore, steady-state filter:

$$\dot{\hat{x}} = -\sqrt{\frac{S_w}{S_v}}\hat{x} + \sqrt{\frac{S_w}{S_v}}y,$$

and the closed-loop pole location determined by S_w/S_v.

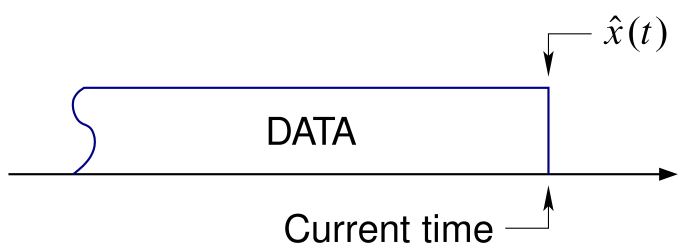
- If S_w/S_v small, sensors are relatively noisy.
- If S_w/S_v large, sensors relatively clean.

More "culture"

- There are three main types of Kalman filter:

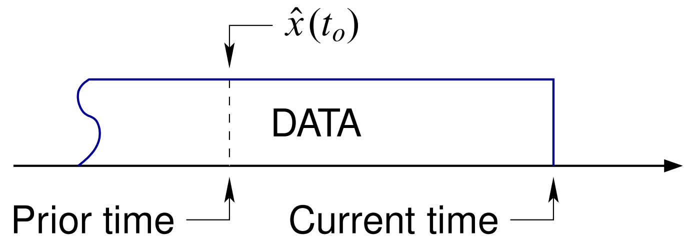
1. We have concentrated on the *"filtering problem"*.

   - Use data up to and including the current time to provide an estimate for the *current* time.



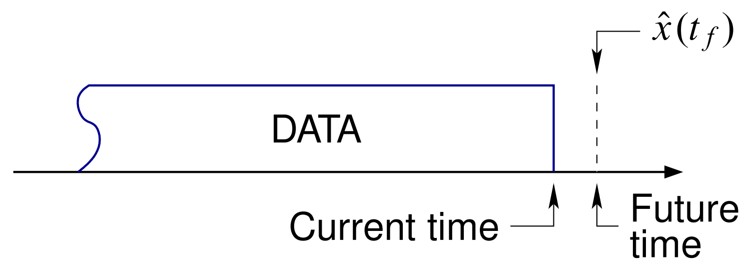
2. The *"smoothing problem"* (offline, post-analysis)

   - Use data up to and including the current time to provide an estimate for a *past* time.



3. The *"prediction problem"*

   - Use data up to and including the current time to provide an estimate for a *future* time.



- All three filters have application to control theory, but #2 and #3 are used more often by signal-processing people.

5.7: Symmetric root locus

- Can extend this analysis to more complex cases and determine where all the closed-loop poles end up.
- Map pole-loci of $A - L_{\text{ss}}C_y$ for valid L_{ss}.
- Model of plant

$$\dot{x} = Ax + B_u u + B_w w, \qquad x \in \mathbb{R}^n$$

$$y = C_y x + v.$$

- For now, (u, w, y, v) scalars. v, w white. $w \sim \mathcal{N}(0, S_w)$, $v \sim \mathcal{N}(0, S_v)$. $[A, B_w]$ stabilizable; $[A, C_y]$ detectable.
- Define scalar transfer function

$$G_{yw}(s) = C_y(sI - A)^{-1}B_w = \frac{N(s)}{D(s)} \ldots \text{(process noise to clean sensor).}$$

- Steady-state optimal observer poles are the LHP zeros of

$$(-1)^n D(s)D(-s)\left[1 + \frac{S_w}{S_v}G_{yw}(s)G_{yw}(-s)\right] = 0$$

or

$$D(s)D(-s) \pm \frac{S_w}{S_v}N(s)N(-s) = 0$$

where $\pm$ is chosen so no poles on imaginary axis.

- Symmetric root locus in the ratio S_w/S_v.
- Changing S_w/S_v uniquely determines the location of the estimator poles.

1. As $S_w/S_v \to 0$ the n poles approach (where $p_i = \lambda_i(A)$)

$$\hat{p}_i = \begin{cases} p_i, & \text{if } \mathbb{R}(p_i) \le 0 \ldots \text{stable;} \\ -p_i, & \text{if } \mathbb{R}(p_i) > 0 \ldots \text{unstable.} \end{cases}$$

- As sensors become relatively noisy, filter gains very low. Essentially open-loop, but must be stable.

2. As $S_w/S_v \to \infty$, —low sensor noise, high bandwidth—assume $N(s)$ has m roots z_i

   (a) m of the n poles approach

$$\hat{p}_i = \begin{cases} z_i, & \text{if } \mathbb{R}(z_i) \le 0; \\ -z_i, & \text{if } \mathbb{R}(z_i) > 0, \end{cases}$$

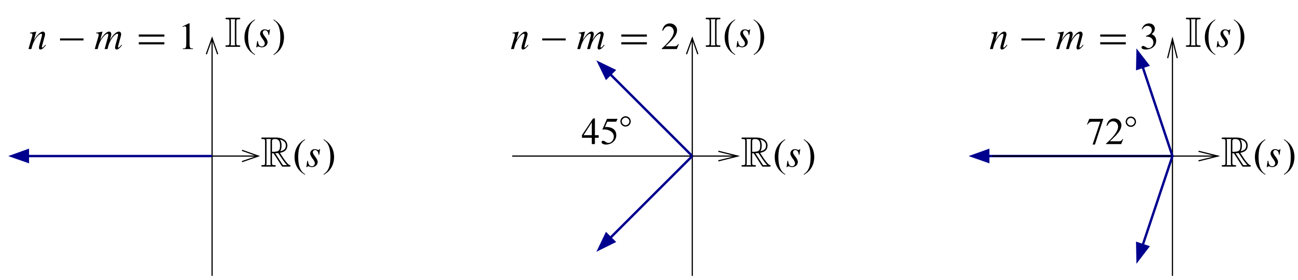
   the finite zeros of $N(s)/D(s)$.

   (b) Remaining $n-m$ poles asymptotically approach straight lines that intersect at origin with angles

$$\pm l\frac{\pi}{n-m} \qquad l = 0, 1, \ldots (n-m-1)/2 \qquad n-m \text{ odd}$$
$$\pm(l+1/2)\frac{\pi}{n-m} \qquad l = 0, 1, \ldots (n-m)/2 - 1 \qquad n-m \text{ even},$$

   a *BUTTERWORTH* configuration.

Butterworth examples



- Remember, these are only asymptotic trajectories for the poles. We can use standard root-locus arguments to find out where the rest of the poles end up at intermediate values of S_w/S_v.
- Note the key role played by the ratio S_w/S_v.

EXAMPLE: Undamped harmonic oscillator.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega_o^2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w$$

and $w \sim \mathcal{N}(0,1)$, $v \sim \mathcal{N}(0, S_v)$.

$$y = x_2 + v.$$

- Location of optimal poles?

$$G_{yw}(s) = \begin{bmatrix} 0 & 1 \end{bmatrix} \left\{ \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -\omega_o^2 & 0 \end{bmatrix} \right\}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

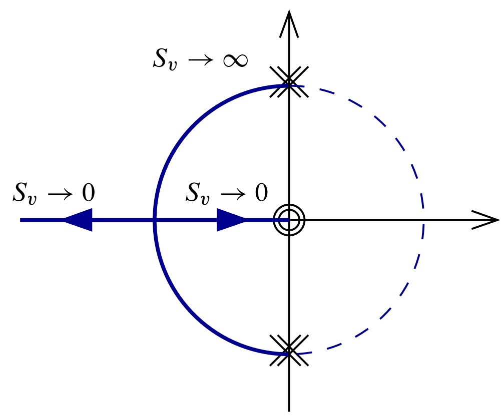
$$= \frac{s}{s^2 + \omega_o^2} = \frac{N(s)}{D(s)}.$$

- Need to find stable solutions of

$$(s^2 + \omega_o^2)^2 - \frac{1}{S_v}(s^2) = 0$$

or

$$1 - \frac{1}{S_v} \frac{s^2}{(s^2 + \omega_o^2)^2} = 0.$$



- As $S_v \to 0$ (low sensor noise), one observer pole tends to the origin (shielded from process noise by system zero at origin). The other pole $\to -\infty$, which essentially results in differentiation of the output to determine the state.

- As $S_v \to \infty$ (high sensor noise), poles revert to the open-loop poles (open-loop observer).

5.8: Relationship between LQE and LQR

- For LQR, we solve the Riccati equation

$$-\dot{P}(t) = P(t)A + A^T P(t) + Q - P(t)B_u R^{-1} B_u^T P(t).$$

  $P(t)$ is the linear function relating the costate to the state ($\lambda(t) = P(t)x(t)$), Q is the state-weighting penalty and R is the control-weighting penalty.

- For LQE, we found that we need to solve the Riccati equation

$$\dot{\Xi}_x(t) = \Xi_x(t)A^T + A\Xi_x(t) + B_w S_w B_w^T - \Xi_x(t)C_y^T S_v^{-1} C_y \Xi_x(t).$$

  $\Xi_x(t)$ is the error covariance matrix, S_w is the process noise spectral density and S_v is the measurement noise spectral density.

- The LQR and LQE Riccati equations are *duals*. First note that LQR is solved backward in time. If it were to be solved forward in time we would have

$$\dot{P}(t) = P(t)A + A^T P(t) + Q - P(t)B_u R^{-1} B_u^T P(t).$$

- Then, replace $A \Leftrightarrow A^T$, $B_u \Leftrightarrow C_y^T$, $K \Leftrightarrow L$, $Q \Leftrightarrow B_w S_w B_w^T$, $R \Leftrightarrow S_v$ and $P \Leftrightarrow \Xi_x$ to notice the duality.

- Duality is explored in further detail in the text to show that LQE is $\mathscr{H}_2$ optimal estimation when cast in a certain framework just as LQR is $\mathscr{H}_2$ optimal control in a similar specific framework.

Solving for the gain matrix $L(t)$

- Here, we use duality to formulate an estimator Hamiltonian matrix, which may be used to solve the estimator Riccati equation.

- Recall that the LQR Riccati equation could be solved by dividing $P(t)$ into two parts. We do the same thing for $\Xi_x(t)$

$$\Xi_x(t) = T_2(t)T_1^{-1}(t).$$

- Substitute this result into the LQE Riccati equation

$$\frac{\mathrm{d}T_2(t)T_1^{-1}(t)}{\mathrm{d}t} = T_2(t)T_1^{-1}(t)A^T + AT_2(t)T_1^{-1}(t) + B_wS_wB_w^T$$
$$-T_2(t)T_1^{-1}(t)C_y^TS_v^{-1}C_yT_2(t)T_1^{-1}(t).$$

- Note that (the first from the product rule; the second from $\mathrm{d}[T_1(t)T_1^{-1}(t)]/\mathrm{d}t = \mathrm{d}I/\mathrm{d}t = 0$.)

$$\frac{\mathrm{d}T_2(t)T_1^{-1}(t)}{\mathrm{d}t} = \frac{\mathrm{d}T_2(t)}{\mathrm{d}t}T_1^{-1}(t) + T_2(t)\frac{\mathrm{d}[T_1^{-1}(t)]}{\mathrm{d}t},$$

and $$\frac{\mathrm{d}[T_1^{-1}(t)]}{\mathrm{d}t} = -T_1^{-1}(t)\dot{T}_1(t)T_1^{-1}(t).$$

- Then

$$\dot{T}_2(t)T_1^{-1}(t) - T_2(t)T_1^{-1}(t)\dot{T}_1(t)T_1^{-1}(t)$$
$$= T_2(t)T_1^{-1}(t)A^T + AT_2(t)T_1^{-1}(t) + B_wS_wB_w^T$$
$$- T_2(t)T_1^{-1}(t)C_y^TS_vC_yT_2(t)T_1^{-1}(t).$$

- Right-multiply by $T_1(t)$ and regroup

$$\dot{T}_2(t) - T_2(t)T_1^{-1}(t)[\dot{T}_1(t)] = B_wS_wB_w^TT_1(t) + AT_2(t)$$
$$-T_2(t)T_1^{-1}(t)\left[-A^TT_1(t) + C_y^TS_v^{-1}C_yT_2(t)\right].$$

- Comparing the two sides of the equation, a solution is obtained when $T_1(t)$ and $T_2(t)$ are solutions to the linear differential equations:

$$\dot{T}_1(t) = -A^TT_1(t) + C_y^TS_v^{-1}C_yT_2(t)$$
$$\dot{T}_2(t) = B_wS_wB_w^TT_1(t) + AT_2(t).$$

- Combining yields the Hamiltonian system for the Kalman filter

$$\begin{bmatrix} \dot{T}_1(t) \\ \dot{T}_2(t) \end{bmatrix} = \begin{bmatrix} -A^T & C_y^T S_v^{-1} C_y \\ B_w S_w B_w^T & A \end{bmatrix} \begin{bmatrix} T_1(t) \\ T_2(t) \end{bmatrix} = \mathscr{Y} \begin{bmatrix} T_1(t) \\ T_2(t) \end{bmatrix}$$

- The matrix $\mathscr{Y}$ is termed the Hamiltonian of the Kalman filter.
- The initial conditions of the Hamiltonian system must satisfy

$$\Xi_x(0) = T_2(0) T_1^{-1}(0),$$

so $T_1(0)$ and $T_2(0)$ are not uniquely specified. One choice is

$$\begin{bmatrix} T_1(0) \\ T_2(0) \end{bmatrix} = \begin{bmatrix} I \\ \Xi_x(0) \end{bmatrix}.$$

- The Hamiltonian system is then solved as

$$\begin{bmatrix} T_1(t) \\ T_2(t) \end{bmatrix} = e^{\mathscr{Y}t} \begin{bmatrix} I \\ \Xi_x(0) \end{bmatrix} = \begin{bmatrix} \Phi_{11}(t) & \Phi_{12}(t) \\ \Phi_{21}(t) & \Phi_{22}(t) \end{bmatrix} \begin{bmatrix} I \\ \Xi_x(0) \end{bmatrix}$$

or

$$\Xi_x(t) = \{\Phi_{21}(t) + \Phi_{22}(t)\Xi_x(0)\}\{\Phi_{11}(t) + \Phi_{12}(t)\Xi_x(0)\}^{-1}$$

and the Kalman gain is

$$L(t) = \{\Phi_{21}(t) + \Phi_{22}(t)\Xi_x(0)\}\{\Phi_{11}(t) + \Phi_{12}(t)\Xi_x(0)\}^{-1} C_y^T S_v^{-1}.$$

The steady-state Kalman gain

- We can also use the Kalman Hamiltonian matrix to find the steady-state gain matrix L_{ss} much the same way we used the LQR Hamiltonian matrix to find the steady-state matrix K_{ss}.
- The eigenequation of the Kalman Hamiltonian is

$$\begin{bmatrix} -A^T & C_y^T S_v^{-1} C_y \\ B_w S_w B_w^T & A \end{bmatrix} \Psi = \Psi \Lambda.$$

- Where Λ is a diagonal matrix containing the eigenvalues of the Hamiltonian (the first n_x of which are in the left-half plane) and Ψ is the matrix of eigenvectors. We can write Ψ as

$$\Psi = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}.$$

- Using properties of the matrix exponential,

$$\begin{aligned} \begin{bmatrix} T_1(t) \\ T_2(t) \end{bmatrix} &= e^{\mathscr{H}t} \begin{bmatrix} I \\ \Xi_x(0) \end{bmatrix} = e^{\Psi\Lambda\Psi^{-1}t} \begin{bmatrix} I \\ \Xi_x(0) \end{bmatrix} = \Psi e^{\Lambda t}\Psi^{-1} \begin{bmatrix} I \\ \Xi_x(0) \end{bmatrix} \\ &= \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix} \begin{bmatrix} e^{-\lambda t} & 0 \\ 0 & e^{\lambda t} \end{bmatrix} \Psi^{-1} \begin{bmatrix} I \\ \Xi_x(0) \end{bmatrix} \\ &= \begin{bmatrix} \Psi_{11}e^{-\lambda t} & \Psi_{12}e^{\lambda t} \\ \Psi_{21}e^{-\lambda t} & \Psi_{22}e^{\lambda t} \end{bmatrix} \begin{bmatrix} \text{(matrix top)} \\ \text{(matrix bottom)} \end{bmatrix}. \end{aligned}$$

- As $t \to \infty$, $e^{-\lambda t} \to 0$ and

$$\begin{bmatrix} T_1(t) \\ T_2(t) \end{bmatrix} \to \begin{bmatrix} \Psi_{12} \\ \Psi_{22} \end{bmatrix} e^{\lambda t}\text{(matrix bottom)}.$$

- The steady-state Riccati solution is then

$$\begin{aligned} \Xi_{x,\text{ss}} &= T_2(\infty)T_1^{-1}(\infty) \\ &= \Psi_{22}(\Psi_{12})^{-1}. \end{aligned}$$

  and the steady-state gain matrix is

$$L_{\text{ss}} = \Psi_{22}(\Psi_{12})^{-1}C_y^T S_v^{-1}.$$

- Note that the formula in the book is incorrect.

Appendix: Proof of matrix inversion lemma, form 2

- Want to show

$$\left[\bar{A}^{-1}+\bar{B}^T\bar{C}^{-1}\bar{B}\right]^{-1}\bar{B}^T\bar{C}^{-1}=\bar{A}\bar{B}^T\left[\bar{B}\bar{A}\bar{B}^T+\bar{C}\right]^{-1}.$$

- Multiply on left by $\left[\bar{A}^{-1}+\bar{B}^T\bar{C}^{-1}\bar{B}\right]$

$$\begin{aligned}\bar{B}^T\bar{C}^{-1}&=\left[\bar{A}^{-1}+\bar{B}^T\bar{C}^{-1}\bar{B}\right]\bar{A}\bar{B}^T\left[\bar{B}\bar{A}\bar{B}^T+\bar{C}\right]^{-1}\\&=\bar{B}^T\left[\bar{B}\bar{A}\bar{B}^T+\bar{C}\right]^{-1}+\bar{B}^T\bar{C}^{-1}\bar{B}\bar{A}\bar{B}^T\left[\bar{B}\bar{A}\bar{B}^T+\bar{C}\right]^{-1}\\&=\bar{B}^T\left\{I+\bar{C}^{-1}\bar{B}\bar{A}\bar{B}^T\right\}\left[\bar{B}\bar{A}\bar{B}^T+\bar{C}\right]^{-1}\\&=\bar{B}^T\bar{C}^{-1}\left\{\bar{C}+\bar{B}\bar{A}\bar{B}^T\right\}\left[\bar{B}\bar{A}\bar{B}^T+\bar{C}\right]^{-1}\\&=\bar{B}^T\bar{C}^{-1}.\end{aligned}$$