Last time we explored the robustness properties of LQR, and we found out that for example if 

\[ \| \Delta^{-1} - I \|_\infty < 1 \]

the closed loop system is stable.

Notice that this is the weak worst-case analysis, that is no matter what \( \Delta \) is, as long as \( \| \Delta^{-1} - I \|_\infty < 1 \), stability is guaranteed.
In general, there are two popular measures for uncertainty quantification:

- **Worst-case**
- **Probabilistic**

**Worst-case:** there is a bound on the uncertainty set, and the parameters can take any value in this set.

**Probabilistic:** the uncertain parameters are represented as random variables/processes, and one reason is to reason about probabilistic properties of the corresponding systems.

Today we explore this second approach in the context of LTI systems.
Our model is as follows

\[ u \rightarrow \begin{align*}
\dot{x} &= Ax + Bu + Gu \\
y &= Cx + v
\end{align*} \rightarrow y
\]

- \( w(t) \): process noise
- \( v(t) \): measurement noise

realization of a random process

- ensemble average
  - if time average
    - is equal to ensemble average

from which we assume the random process is ergodic.

\[ \text{white: uncorrelated across time} \]

\[ \text{Gaussian: the random variable at each instance has a Gaussian distribution} \]

\[ \text{expressed in terms of the density} \]

\[ \text{mean} \]

\[ \text{Covariance} \]
With respect to the Gaussian density function we can define the notion of \( \mathbb{E} \) expected value of the random variable:

\[
\mathbb{E}\{X\} = \int \mathbf{y} \, p_x(\mathbf{y}) \, d\mathbf{y}
\]

\[
p_x(\mathbf{y}) = \frac{1}{(2\pi)^{n/2} \det \Sigma} e^{-\frac{1}{2} (\mathbf{y} - \bar{x})^T \Sigma^{-1} (\mathbf{y} - \bar{x})}
\]

where \( \Sigma \) is the covariance matrix, positive definite, \( \bar{x} \) is the mean, \( \bar{x} = \mathbb{E}\{x\} \) (first moment),

\[
(2nd \, moment) \Rightarrow \Sigma = \mathbb{E}\{(x - \bar{x})(x - \bar{x})^T\}
\]

An important property of the "expectation" is that it is a linear operator:

\[
\mathbb{E}\{x + y\} = \mathbb{E}\{x\} + \mathbb{E}\{y\}
\]
For our setting we will assume that

\[ \mathbb{E}\{w(t)^2\} = 0 \quad \mathbb{E}\{w(t)w(t)^T\} = W(t) \delta(t-t') \]

\[ \mathbb{E}\{v(t)^2\} = 0 \quad \mathbb{E}\{v(t)v(t)^T\} = V(t) \delta(t-t') \]

for all \( t, t' \).

\[ \mathbb{E}\{v(t)w(t)^T\} \text{ independent} \]

\[ \text{stationary } \implies W \]

\[ \text{un correlated across time } \implies V \]

we will also assume that \( x_0 \) the initial condition is also Gaussian, independent of \( v \) \& \( w \).

**example:** suppose \( w \sim N(\bar{w}, W) \)

normal \( \bar{w} \) mean \( W \) covariance

Gaussian

\[ z = Aw + b \] \( \mathbb{E}\{z\} = A \mathbb{E}\{w\} + b = A \bar{w} + b \)

\[ \text{cov } z = \mathbb{E}\{(z - \bar{z})(z - \bar{z})^T\} = \mathbb{E}\{A(w - \bar{w})(w - \bar{w})A^T\} = \bar{A}WA^T \]
Note that this allows us to consider
\[ \dot{x} = Ax + Bu + \bar{w} \quad (*) \]
instead of
\[ \dot{x} = Ax + Bu + Gw \quad (**) \]
since if \( w \) in (**) is \( N(\bar{w}, \bar{W}) \) then
the equivalent \( \bar{w} \) in (*) is \( N(G\bar{w}, GWG^T) \).

Let us now consider the set up of optimal estimation
Parallel to the setup of the observer, we form the filter of the form

\[ \dot{x} = A \hat{x} + Bu + K_f \overline{y} (y - C \hat{x}) \]

where \( \overline{y} \) is filter gain.

Let \( e(t) = x(t) - \hat{x}(t) \), then

\[ e = x - \hat{x} = (A - K_f C) \nu + \nu \]

where \( \nu = w - K_f v \)

Note that

\[ E\{\nu\} = E\{w\} - K_f E\{v\} = 0 \]

\[ \text{Cov}(\nu) = W - K_f V K_f^T \]
We consider the problem
\[ \min \lim_{t \to \infty} \mathbb{E} \left\{ e(t)^{\top} Q e(t) \right\} \]

Recall that we can write
\[ e^{\top} Q e = \text{trace} \left( e e^{\top} Q \right) \]

If \( Q = I \), then the optimization aims to minimize \( e(t)^{\top} e(t) \) as \( t \to \infty \), i.e.,
the steady state variance of the estimation error.
Since this problem looks similar to LQR, let us see if we can get some insights into how we can streamline the solution.

Let us start by recalling how certain cost functions can be computed using linear algebra:

1. If \( A \) is Hurwitz, \( x = Ax \), what is \( J = \int_0^\infty x^T Q x \, dt \)?

\[ J = x_0^T P x_0 \quad \text{where} \quad A^T P + PA + Q = 0. \]

So if we solve \( \min \, \text{trace} \, P (x_0 x_0^T) \)

\[ A^T P + PA + Q = 0 \]

\& \( A \) is Hurwitz we get \( J \).
What about \( J = \min_K \int_0^\infty x^TQx + u^TRu \, dt \)
\[
\dot{x} = Ax + Bu \quad , \quad u = -Kx \\
x_0 \text{ given}
\]

we can write this as:
\[
\min_K \int_0^\infty x^T(Q + KRK)x \, dt
\]
\[
\dot{x} = (A - BK)x
\]

therefore we can write \( J \) as
\[
\min_K \quad \text{trace } P x_0 x_0^T
\]
\[
K, P \quad (A - BK)^T P + P (A - BK) = -Q - KRK
\]

we know the solution is
\[
K = -RBP
\]
\[
A^TP + PA + Q - BP \, R^{-1}BP = 0.
\]
So what is the connection?

If we look at

\[
\min_e \{ \text{trace } ee^TQ \} \\
\hat{e} = (A - KfC)e + \eta.
\]

we want to connect this solution of LQR with

If we need the following fact:

\[
J = \lim_{t \to \infty} \{ x^TQx \} \\
\hat{e} = Ax + \eta \\
\mathbb{E}\{w\} = 0 \\
\mathbb{E}\{-ww^T\} = W
\]

\[
J: \text{trace } PW \\
A^TP + PA + Q = 0.
\]
In this case
\[ P = \int_0^\infty e^{At}Qe^{At} \, dt \]

so
\[ \text{trace } PW = \text{trace} \left( \int_0^\infty e^{At}Qe^{At} \, dt \right) W \]
\[ = \text{trace} \int_0^\infty e^{At}We^{At} \, dtQ \]
\[ \approx \Sigma \text{ where } A\Sigma + \Sigma^T A = -W \]

so if we have
\[ \lim_{t \to \infty} E[xQx^T] = A\Sigma + \Sigma^T A = -W, \text{ steady state covariance} \]
so if we want to \[ \sum_{k_f} \text{min } \text{trace } \{ \mathbf{e} \mathbf{e}^T \mathbf{Q} \} \]

\[ \mathbf{e} = (\mathbf{A} - k_f \mathbf{C}) \mathbf{e} + \mathbf{y} \]

\[ \mathbf{W} + k_f \mathbf{V}_f^T \]

we can solve:

\[ \begin{align*}
\min_{\Sigma, k_f} & \quad \text{trace } \Sigma \mathbf{Q} \\
\text{subject to } & \quad (\mathbf{A} - k_f \mathbf{C}) \Sigma + \Sigma (\mathbf{A} - k_f \mathbf{C})^T \mathbf{W} + k_f \mathbf{V}_f^T = \mathbf{0}
\end{align*} \]

compare this with

\[ \begin{align*}
\min_{\mathbf{P}, \mathbf{K}} & \quad \text{trace } \mathbf{P} \mathbf{x}_0 \mathbf{x}_0^T \\
\text{subject to } & \quad (\mathbf{A} - \mathbf{B} \mathbf{K})^T \mathbf{P} + \mathbf{P} (\mathbf{A} - \mathbf{B} \mathbf{K}) + \mathbf{Q} + \mathbf{K} \mathbf{R} \mathbf{K}^T = \mathbf{0}
\end{align*} \]

with solution \[ \mathbf{K} = \mathbf{R}^{-1} \mathbf{B} \mathbf{P} \]
If we compare

\[(A - K_f C) \Sigma + \Sigma (A - K_f C)^T + K_f \Sigma K_f^T + W = 0\]

\[(A - B K)^T P + P (A - B K) + K^T R K + Q = 0\]

we realize that

\[A \rightarrow A^T\]
\[K \rightarrow K_f^T\]
\[Q \rightarrow W\]
\[V \rightarrow R\]
\[B^T \rightarrow \zeta\]

So, optimal \(K_f\) is

\[K_f = \Sigma C V^{-1}\]
\[ \Sigma \] is the solution of the filter Riccati equation:

\[ A\Sigma + \Sigma A^T + W - \Sigma C^T V^{-1} C\Sigma = 0. \]

\[ K_f \] is called the Kalman (filter) gain.

\[ \hat{x} = A\hat{x} + Bu + K_f (y - C\hat{x}) \]

is called the Kalman filter.
there are two important properties of the Kalman filter that we use soon:

\[ \mathbb{E}\{e(t) \hat{x}(t)^T\} = 0 \]

\&

\[ \mathbb{E}\{y - C\hat{x}\} = 0 \]

\[ \mathbb{E}\{(y - C\hat{x})(y - C\hat{x})^T\} = V \]

space of all (causal) random processes consistent w/ linear models.
We will see that these two properties can be used to show the separation principle.

\( \dot{x} = Ax + Bu + w \)
\( y = Cx + v \)

Kalman control gain

Kalman filter gain.

\( \hat{x} = Ax + Bu + K_f (\hat{y} - Cy) \)

\( u \)

\( \hat{x} \)

\( K \) and \( K_f \) can be designed separately.