Lecture 1

Linear quadratic regulator: Discrete-time finite horizon

- LQR cost function
- multi-objective interpretation
- LQR via least-squares
- dynamic programming solution
- steady-state LQR control
- extensions: time-varying systems, tracking problems
LQR problem: background

discrete-time system \( x(t + 1) = Ax(t) + Bu(t), x(0) = x_0 \)

problem: choose \( u(0), u(1), \ldots \) so that

- \( x(0), x(1), \ldots \) is ‘small’, \( i.e. \), we get good regulation or control
- \( u(0), u(1), \ldots \) is ‘small’, \( i.e. \), using small input effort or actuator authority
- we’ll define ‘small’ soon
- these are usually competing objectives, \( e.g. \), a large \( u \) can drive \( x \) to zero fast

linear quadratic regulator (LQR) theory addresses this question
LQR cost function

we define quadratic cost function

\[ J(U) = \sum_{\tau=0}^{N-1} (x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau)) + x(N)^T Q_f x(N) \]

where \( U = (u(0), \ldots, u(N - 1)) \) and

\[ Q = Q^T \geq 0, \quad Q_f = Q_f^T \geq 0, \quad R = R^T > 0 \]

are given state cost, final state cost, and input cost matrices
• $N$ is called time horizon (we’ll consider $N = \infty$ later)

• first term measures state deviation

• second term measures input size or actuator authority

• last term measures final state deviation

• $Q$, $R$ set relative weights of state deviation and input usage

• $R > 0$ means any (nonzero) input adds to cost $J$

**LQR problem:** find $u_{lqr}(0), \ldots, u_{lqr}(N - 1)$ that minimizes $J(U)$
Comparison to least-norm input

c.f. least-norm input that steers $x$ to $x(N) = 0$:

- no cost attached to $x(0), \ldots, x(N-1)$
- $x(N)$ must be exactly zero

we can approximate the least-norm input by taking

$$R = I, \quad Q = 0, \quad Q_f \text{ large}, \ e.g., \ Q_f = 10^8 I$$
Multi-objective interpretation

common form for $Q$ and $R$:

$$R = \rho I, \quad Q = Q_f = C^T C$$

where $C \in \mathbb{R}^{p \times n}$ and $\rho \in \mathbb{R}$, $\rho > 0$

cost is then

$$J(U) = \sum_{\tau=0}^{N} \|y(\tau)\|^2 + \rho \sum_{\tau=0}^{N-1} \|u(\tau)\|^2$$

where $y = Cx$

here $\sqrt{\rho}$ gives relative weighting of output norm and input norm
Input and output objectives

fix $x(0) = x_0$ and horizon $N$; for any input $U = (u(0), \ldots, u(N - 1))$ define

- **input cost** $J_{\text{in}}(U) = \sum_{\tau=0}^{N-1} \|u(\tau)\|^2$
- **output cost** $J_{\text{out}}(U) = \sum_{\tau=0}^{N} \|y(\tau)\|^2$

these are (competing) objectives; we want both small

LQR quadratic cost is $J_{\text{out}} + \rho J_{\text{in}}$
plot \((J_{\text{in}}, J_{\text{out}})\) for all possible \(U\):

- shaded area shows \((J_{\text{in}}, J_{\text{out}})\) achieved by some \(U\)
- clear area shows \((J_{\text{in}}, J_{\text{out}})\) not achieved by any \(U\)
three sample inputs $U_1$, $U_2$, and $U_3$ are shown

- $U_3$ is worse than $U_2$ on both counts ($J_{\text{in}}$ and $J_{\text{out}}$)

- $U_1$ is better than $U_2$ in $J_{\text{in}}$, but worse in $J_{\text{out}}$

interpretation of LQR quadratic cost:

$$J = J_{\text{out}} + \rho J_{\text{in}} = \text{constant}$$

corresponds to a line with slope $-\rho$ on $(J_{\text{in}}, J_{\text{out}})$ plot
LQR optimal input is at boundary of shaded region, just touching line of smallest possible $J$

$u_2$ is LQR optimal for $\rho$ shown

by varying $\rho$ from 0 to $+\infty$, can sweep out optimal tradeoff curve
LQR via least-squares

LQR can be formulated (and solved) as a least-squares problem

\[ X = (x(0), \ldots, x(N)) \] is a linear function of \( x(0) \) and \( U = (u(0), \ldots, u(N - 1)) \):

\[
\begin{bmatrix}
    x(0) \\
    \vdots \\
    x(N)
\end{bmatrix} =
\begin{bmatrix}
    0 & \cdots & 0 \\
    B & 0 & \cdots \\
    AB & B & 0 & \cdots \\
    \vdots & \vdots & \ddots & \vdots \\
    A^{N-1}B & A^{N-2}B & \cdots & B
\end{bmatrix}
\begin{bmatrix}
    u(0) \\
    \vdots \\
    u(N - 1)
\end{bmatrix} +
\begin{bmatrix}
    I \\
    A \\
    \vdots \\
    A^N
\end{bmatrix}
\begin{bmatrix}
    x(0)
\end{bmatrix}
\]

express as \( X = GU + Hx(0) \), where \( G \in \mathbb{R}^{Nn \times Nm} \), \( H \in \mathbb{R}^{Nn \times n} \)
express LQR cost as

\[
J(U) = \left\| \text{diag}(Q^{1/2}, \ldots, Q^{1/2}, Q_f^{1/2})(GU + Hx(0)) \right\|^2 \\
+ \left\| \text{diag}(R^{1/2}, \ldots, R^{1/2})U \right\|^2
\]

this is just a (big) least-squares problem

this solution method requires forming and solving a least-squares problem with size \(N(n + m) \times Nm\)

using a naive method (e.g., QR factorization), cost is \(O(N^3nm^2)\)
Dynamic programming solution

• gives an efficient, recursive method to solve LQR least-squares problem; cost is $O(Nn^3)$

• (but in fact, a less naive approach to solve the LQR least-squares problem will have the same complexity)

• useful and important idea on its own

• same ideas can be used for many other problems
for $t = 0, \ldots, N$ define the value function $V_t : \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$V_t(z) = \min_{u(t), \ldots, u(N-1)} \sum_{\tau = t}^{N-1} (x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau)) + x(N)^T Q_f x(N)$$

subject to $x(t) = z$, $x(\tau + 1) = Ax(\tau) + Bu(\tau)$, $\tau = t, \ldots, T$

- $V_t(z)$ gives the minimum LQR cost-to-go, starting from state $z$ at time $t$
- $V_0(x_0)$ is min LQR cost (from state $x_0$ at time 0)
we will find that

- $V_t$ is quadratic, i.e., $V_t(z) = z^T P_t z$, where $P_t = P_t^T \geq 0$
- $P_t$ can be found recursively, working backward from $t = N$
- the LQR optimal $u$ is easily expressed in terms of $P_t$

cost-to-go with no time left is just final state cost:

$$V_N(z) = z^T Q_f z$$

thus we have $P_N = Q_f$
Dynamic programming principle

• now suppose we know $V_{t+1}(z)$

• what is the optimal choice for $u(t)$?

• choice of $u(t)$ affects
  – current cost incurred (through $u(t)^T R u(t)$)
  – where we land, $x(t + 1)$ (hence, the min-cost-to-go from $x(t + 1)$)

• dynamic programming (DP) principle:

$$ V_t(z) = \min_w \left( z^T Q z + w^T R w + V_{t+1}(A z + B w) \right) $$

– $z^T Q z + w^T R w$ is cost incurred at time $t$ if $u(t) = w$
– $V_{t+1}(A z + B w)$ is min cost-to-go from where you land at $t + 1$
follows from fact that we can minimize in any order:

$$\min_{w_1, \ldots, w_k} f(w_1, \ldots, w_k) = \min_{w_1} \left( \min_{w_2, \ldots, w_k} f(w_1, \ldots, w_k) \right)$$

a fct of $w_1$

in words:
min cost-to-go from where you are \(=\) min over
(current cost incurred + min cost-to-go from where you land)
Example: path optimization

- edges show possible flights; each has some cost
- want to find min cost route or path from SF to NY
dynamic programming (DP):

• $V(i)$ is min cost from airport $i$ to NY, over all possible paths

• to find min cost from city $i$ to NY: minimize sum of flight cost plus min cost to NY from where you land, over all flights out of city $i$ (gives optimal flight out of city $i$ on way to NY)

• if we can find $V(i)$ for each $i$, we can find min cost path from any city to NY

• DP principle: $V(i) = \min_j(c_{ji} + V(j))$, where $c_{ji}$ is cost of flight from $i$ to $j$, and minimum is over all possible flights out of $i$
HJ equation for LQR

\[ V_t(z) = z^T Q z + \min_w \left( w^T R w + V_{t+1}(A z + B w) \right) \]

- called DP, Bellman, or Hamilton-Jacobi equation
- gives \( V_t \) recursively, in terms of \( V_{t+1} \)
- any minimizing \( w \) gives optimal \( u(t) \):

\[ u_{lq}(t) = \arg\min_w \left( w^T R w + V_{t+1}(A z + B w) \right) \]
let’s assume that $V_{t+1}(z) = z^T P_{t+1} z$, with $P_{t+1} = P_{t+1}^T \geq 0$

we’ll show that $V_t$ has the same form

by DP,

$$V_t(z) = z^T Q z + \min_w \left( w^T R w + (A z + B w)^T P_{t+1} (A z + B w) \right)$$

can solve by setting derivative w.r.t. $w$ to zero:

$$2w^T R + 2(A z + B w)^T P_{t+1} B = 0$$

hence optimal input is

$$w^* = -(R + B^T P_{t+1} B)^{-1} B^T P_{t+1} A z$$
• and so (after some ugly algebra)

\[ V_t(z) = z^T Q z + w^*^T R w^* + (Az + B w^*)^T P_{t+1} (Az + B w^*) \]
\[ = z^T (Q + A^T P_{t+1} A - A^T P_{t+1} B (R + B^T P_{t+1} B)^{-1} B^T P_{t+1} A) z \]
\[ = z^T P_t z \]

where

\[ P_t = Q + A^T P_{t+1} A - A^T P_{t+1} B (R + B^T P_{t+1} B)^{-1} B^T P_{t+1} A \]

• easy to show \( P_t = P_t^T \geq 0 \)
Summary of LQR solution via DP

1. set $P_N := Q_f$

2. for $t = N, \ldots, 1$,

   $$P_{t-1} := Q + A^T P_t A - A^T P_t B (R + B^T P_t B)^{-1} B^T P_t A$$

3. for $t = 0, \ldots, N - 1$, define $K_t := -(R + B^T P_{t+1} B)^{-1} B^T P_{t+1} A$

4. for $t = 0, \ldots, N - 1$, optimal $u$ is given by $u_{\text{lqr}}(t) = K_t x(t)$

- optimal $u$ is a linear function of the state (called linear state feedback)
- recursion for min cost-to-go runs backward in time
LQR example

2-state, single-input, single-output system

\[ x(t + 1) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t), \quad y(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) \]

with initial state \( x(0) = (1, 0) \), horizon \( N = 20 \), and weight matrices

\[ Q = Q_f = C^T C, \quad R = \rho I \]
optimal trade-off curve of $J_{\text{in}}$ vs. $J_{\text{out}}$: 

![Graph showing the trade-off curve with circles indicating LQR solutions for $\rho = 0.3$ and $\rho = 10$.]

circles show LQR solutions with $\rho = 0.3$, $\rho = 10$
$u$ & $y$ for $\rho = 0.3$, $\rho = 10$:
optimal input has form \( u(t) = K(t)x(t) \), where \( K(t) \in \mathbb{R}^{1 \times 2} \)

state feedback gains vs. \( t \) for various values of \( Q_f \) (note convergence):

![Graphs showing state feedback gains vs. t for various values of Q_f](image-url)
Steady-state regulator

usually $P_t$ rapidly converges as $t$ decreases below $N$

limit or steady-state value $P_{ss}$ satisfies

$$P_{ss} = Q + A^T P_{ss} A - A^T P_{ss} B (R + B^T P_{ss} B)^{-1} B^T P_{ss} A$$

which is called the (DT) algebraic Riccati equation (ARE)

- $P_{ss}$ can be found by iterating the Riccati recursion, or by direct methods

- for $t$ not close to horizon $N$, LQR optimal input is approximately a linear, constant state feedback

$$u(t) = K_{ss} x(t), \quad K_{ss} = -(R + B^T P_{ss} B)^{-1} B^T P_{ss} A$$

(very widely used in practice; more on this later)
Time-varying systems

LQR is readily extended to handle time-varying systems

\[ x(t + 1) = A(t)x(t) + B(t)u(t) \]

and time-varying cost matrices

\[
J = \sum_{\tau=0}^{N-1} \left( x(\tau)^T Q(\tau)x(\tau) + u(\tau)^T R(\tau)u(\tau) \right) + x(N)^T Q_f x(N)
\]

(so \( Q_f \) is really just \( Q(N) \))

DP solution is readily extended, but (of course) there need not be a steady-state solution
Tracking problems

we consider LQR cost with state and input offsets:

\[
J = \sum_{\tau=0}^{N-1} (x(\tau) - \bar{x}(\tau))^T Q (x(\tau) - \bar{x}(\tau)) \\
+ \sum_{\tau=0}^{N-1} (u(\tau) - \bar{u}(\tau))^T R (u(\tau) - \bar{u}(\tau))
\]

(we drop the final state term for simplicity)

here, \( \bar{x}(\tau) \) and \( \bar{u}(\tau) \) are given desired state and input trajectories

DP solution is readily extended, even to time-varying tracking problems
Gauss-Newton LQR

**nonlinear** dynamical system: \( x(t + 1) = f(x(t), u(t)), x(0) = x_0 \)

Objective is

\[
J(U) = \sum_{\tau=0}^{N-1} (x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau)) + x(N)^T Q_f x(N)
\]

where \( Q = Q^T \geq 0, \ Q_f = Q_f^T \geq 0, \ R = R^T > 0 \)

Start with a guess for \( U \), and alternate between:

- linearize around current trajectory
- solve associated LQR (tracking) problem

Sometimes converges, sometimes to the globally optimal \( U \)
some more detail:

- let $u$ denote current iterate or guess
- simulate system to find $x$, using $x(t + 1) = f(x(t), u(t))$
- linearize around this trajectory: $\delta x(t + 1) = A(t)\delta x(t) + B(t)\delta u(t)$

$$A(t) = D_x f(x(t), u(t)) \quad B(t) = D_u f(x(t), u(t))$$

- solve time-varying LQR tracking problem with cost

$$J = \sum_{\tau=0}^{N-1} (x(\tau) + \delta x(\tau))^T Q (x(\tau) + \delta x(\tau))$$

$$+ \sum_{\tau=0}^{N-1} (u(\tau) + \delta u(\tau))^T R (u(\tau) + \delta u(\tau))$$

- for next iteration, set $u(t) := u(t) + \delta u(t)$
Lecture 2
LQR via Lagrange multipliers

- useful matrix identities
- linearly constrained optimization
- LQR via constrained optimization
Some useful matrix identities

let’s start with a simple one:

$$Z(I + Z)^{-1} = I - (I + Z)^{-1}$$

(provided $I + Z$ is invertible)

to verify this identity, we start with

$$I = (I + Z)(I + Z)^{-1} = (I + Z)^{-1} + Z(I + Z)^{-1}$$

re-arrange terms to get identity
an identity that’s a bit more complicated:

\[
(I + XY)^{-1} = I - X(I + YX)^{-1}Y
\]

(if either inverse exists, then the other does; in fact
\[
\det(I + XY) = \det(I + YX)
\]

to verify:

\[
(I - X(I + YX)^{-1}Y) (I + XY) = I + XY - X(I + YX)^{-1}Y(I + XY)
\]

\[
= I + XY - X(I + YX)^{-1}(I + YX)Y
\]

\[
= I + XY - XY = I
\]
another identity:

\[ Y(I + XY)^{-1} = (I + YX)^{-1}Y \]

to verify this one, start with \( Y(I + XY) = (I + YX)Y \)

then multiply on left by \((I + YX)^{-1}\), on right by \((I + XY)^{-1}\)

• note dimensions of inverses not necessarily the same

• mnemonic: lefthand \( Y \) moves into inverse, pushes righthand \( Y \) out . . .
and one more:

\[(I + XZ^{-1}Y)^{-1} = I - X(Z + YX)^{-1}Y\]

let's check:

\[(I + X(Z^{-1}Y))^{-1} = I - X(I + Z^{-1}YX)^{-1}Z^{-1}Y\]
\[= I - X(Z(I + Z^{-1}YX))^{-1}Y\]
\[= I - X(Z + YX)^{-1}Y\]
Example: rank one update

- Suppose we’ve already calculated or know $A^{-1}$, where $A \in \mathbb{R}^{n \times n}$

- We need to calculate $(A + bc^T)^{-1}$, where $b, c \in \mathbb{R}^n$
  ($A + bc^T$ is called a rank one update of $A$)

We’ll use another identity, called matrix inversion lemma:

$$(A + bc^T)^{-1} = A^{-1} - \frac{1}{1 + c^T A^{-1} b} (A^{-1} b)(c^T A^{-1})$$

Note that RHS is easy to calculate since we know $A^{-1}$
more general form of matrix inversion lemma:

\[(A + BC)^{-1} = A^{-1} - A^{-1}B (I + CA^{-1}B)^{-1} CA^{-1}\]

let’s verify it:

\[
\begin{align*}
(A + BC)^{-1} &= (A(I + A^{-1}BC))^{-1} \\
&= (I + (A^{-1}B)C)^{-1}A^{-1} \\
&= (I - (A^{-1}B)(I + C(A^{-1}B))^{-1}C) A^{-1} \\
&= A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}
\end{align*}
\]
Another formula for the Riccati recursion

\[ P_{t-1} = Q + A^T P_t A - A^T P_t B (R + B^T P_t B)^{-1} B^T P_t A \]
\[ = Q + A^T P_t (I - B (R + B^T P_t B)^{-1} B^T P_t) A \]
\[ = Q + A^T P_t (I - B ((I + B^T P_t B R^{-1}) R)^{-1} B^T P_t) A \]
\[ = Q + A^T P_t (I - B R^{-1} (I + B^T P_t B R^{-1})^{-1} B^T P_t) A \]
\[ = Q + A^T P_t (I + B R^{-1} B^T P_t)^{-1} A \]
\[ = Q + A^T (I + P_t B R^{-1} B^T)^{-1} P_t A \]

or, in pretty, symmetric form:

\[ P_{t-1} = Q + A^T P_t^{1/2} \left( I + P_t^{1/2} B R^{-1} B^T P_t^{1/2} \right)^{-1} P_t^{1/2} A \]
Linearly constrained optimization

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad Fx = g
\end{align*}
\]

- \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is smooth \textit{objective function}
- \( F \in \mathbb{R}^{m \times n} \) is fat

form \textit{Lagrangian} \( L(x, \lambda) = f(x) + \lambda^T(g - Fx) \) (\( \lambda \) is \textit{Lagrange multiplier})

if \( x \) is optimal, then

\[
\nabla_x L = \nabla f(x) - F^T \lambda = 0, \quad \nabla_\lambda L = g - Fx = 0
\]

\( \text{i.e.,} \ \nabla f(x) = F^T \lambda \) for some \( \lambda \in \mathbb{R}^m \)

(generalizes optimality condition \( \nabla f(x) = 0 \) for unconstrained minimization problem)
\[ f(x) = \text{constant} \]

\[ \nabla f(x) = F^T \lambda \text{ for some } \lambda \iff \nabla f(x) \in \mathcal{R}(F^T) \iff \nabla f(x) \perp \mathcal{N}(F) \]
Feasible descent direction

suppose $x$ is current, feasible point (i.e., $Fx = g$)

consider a small step in direction $v$, to $x + hv$ ($h$ small, positive)

when is $x + hv$ better than $x$?

need $x + hv$ feasible: $F(x + hv) = g + hFv = g$, so $Fv = 0$

$v \in N(F)$ is called a feasible direction

we need $x + hv$ to have smaller objective than $x$:

$$f(x + hv) \approx f(x) + h \nabla f(x)^T v < f(x)$$

so we need $\nabla f(x)^T v < 0$ (called a descent direction)

(if $\nabla f(x)^T v > 0$, $-v$ is a descent direction, so we need only $\nabla f(x)^T v \neq 0$)

$x$ is not optimal if there exists a feasible descent direction
if $x$ is optimal, every feasible direction satisfies $\nabla f(x)^T v = 0$

\[
Fv = 0 \Rightarrow \nabla f(x)^T v = 0 \iff \mathcal{N}(F) \subseteq \mathcal{N}(\nabla f(x)^T) \\
\iff \mathcal{R}(F^T) \supseteq \mathcal{R}(\nabla f(x)) \\
\iff \nabla f(x) \in \mathcal{R}(F^T) \\
\iff \nabla f(x) = F^T \lambda \quad \text{for some } \lambda \in \mathbb{R}^m \\
\iff \nabla f(x) \perp \mathcal{N}(F)
\]
LQR as constrained minimization problem

minimize \[ J = \frac{1}{2} \sum_{t=0}^{N-1} (x(t)^T Q x(t) + u(t)^T R u(t)) + \frac{1}{2} x(N)^T Q f x(N) \]
subject to \[ x(t + 1) = A x(t) + B u(t), \quad t = 0, \ldots, N - 1 \]

• variables are \( u(0), \ldots, u(N - 1) \) and \( x(1), \ldots, x(N) \)
  \((x(0) = x_0 \text{ is given})\)

• objective is (convex) quadratic
  (factor \(1/2\) in objective is for convenience)

introduce Lagrange multipliers \( \lambda(1), \ldots, \lambda(N) \in \mathbb{R}^n \) and form Lagrangian

\[ L = J + \sum_{t=0}^{N-1} \lambda(t + 1)^T (A x(t) + B u(t) - x(t + 1)) \]
Optimality conditions

we have \( x(t + 1) = Ax(t) + Bu(t) \) for \( t = 0, \ldots, N - 1, \ x(0) = x_0 \)

for \( t = 0, \ldots, N - 1, \ \nabla_{u(t)} L = R u(t) + B^T \lambda(t + 1) = 0 \)

hence, \( u(t) = -R^{-1} B^T \lambda(t + 1) \)

for \( t = 1, \ldots, N - 1, \ \nabla_{x(t)} L = Q x(t) + A^T \lambda(t + 1) - \lambda(t) = 0 \)

hence, \( \lambda(t) = A^T \lambda(t + 1) + Q x(t) \)

\( \nabla_{x(N)} L = Q_f x(N) - \lambda(N) = 0 \), so \( \lambda(N) = Q_f x(N) \)

these are a set of linear equations in the variables

\[ u(0), \ldots, u(N - 1), \ x(1), \ldots, x(N), \ \lambda(1), \ldots, \lambda(N) \]
Co-state equations

Optimality conditions break into two parts:

\[ x(t + 1) = Ax(t) + Bu(t), \quad x(0) = x_0 \]

This recursion for state \( x \) runs forward in time, with initial condition

\[ \lambda(t) = A^T \lambda(t + 1) + Qx(t), \quad \lambda(N) = Qfx(N) \]

This recursion for \( \lambda \) runs backward in time, with final condition

- \( \lambda \) is called co-state
- Recursion for \( \lambda \) sometimes called adjoint system
Solution via Riccati recursion

we will see that $\lambda(t) = P_t x(t)$, where $P_t$ is the min-cost-to-go matrix
defined by the Riccati recursion

thus, Riccati recursion gives clever way to solve this set of linear equations

it holds for $t = N$, since $P_N = Q_f$ and $\lambda(N) = Q_f x(N)$

now suppose it holds for $t + 1$, i.e., $\lambda(t + 1) = P_{t+1} x(t + 1)$

let’s show it holds for $t$, i.e., $\lambda(t) = P_t x(t)$

using $x(t + 1) = A x(t) + B u(t)$ and $u(t) = -R^{-1} B^T \lambda(t + 1)$,

$$\lambda(t + 1) = P_{t+1} (A x(t) + B u(t)) = P_{t+1} (A x(t) - B R^{-1} B^T \lambda(t + 1))$$

so

$$\lambda(t + 1) = (I + P_{t+1} B R^{-1} B^T)^{-1} P_{t+1} A x(t)$$
using $\lambda(t) = \lambda(t) = A^T \lambda(t + 1) + Qx(t)$, we get

$$\lambda(t) = A^T(I + P_{t+1}BR^{-1}B^T)^{-1}P_{t+1}Ax(t) + Qx(t) = P_t x(t)$$

since by the Riccati recursion

$$P_t = Q + A^T(I + P_{t+1}BR^{-1}B^T)^{-1}P_{t+1}A$$

this proves $\lambda(t) = P_t x(t)$
let’s check that our two formulas for $u(t)$ are consistent:

$$u(t) = -R^{-1} B^T \lambda(t + 1)$$

$$= -R^{-1} B^T (I + P_{t+1} BR^{-1} B^T)^{-1} P_{t+1} Ax(t)$$

$$= -R^{-1} (I + B^T P_{t+1} BR^{-1})^{-1} B^T P_{t+1} Ax(t)$$

$$= -((I + B^T P_{t+1} BR^{-1}) R)^{-1} B^T P_{t+1} Ax(t)$$

$$= -(R + B^T P_{t+1} B)^{-1} B^T P_{t+1} Ax(t)$$

which is what we had before
Lecture 3
Infinite horizon linear quadratic regulator

• infinite horizon LQR problem

• dynamic programming solution

• receding horizon LQR control

• closed-loop system
Infinite horizon LQR problem

discrete-time system \( x(t + 1) = Ax(t) + Bu(t), \; x(0) = x_0 \)

problem: choose \( u(0), u(1), \ldots \) to minimize

\[
J = \sum_{\tau=0}^{\infty} (x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau))
\]

with given constant state and input weight matrices

\[
Q = Q^T \geq 0, \quad R = R^T > 0
\]

\dots an infinite dimensional problem
problem: it’s possible that $J = \infty$ for all input sequences $u(0), \ldots$

$$x(t + 1) = 2x(t) + 0u(t), \quad x(0) = 1$$

let’s assume $(A, B)$ is controllable

then for any $x_0$ there’s an input sequence

$$u(0), \ldots, u(n - 1), 0, 0, \ldots$$

that steers $x$ to zero at $t = n$, and keeps it there

for this $u$, $J < \infty$

and therefore, $\min_u J < \infty$ for any $x_0$
Dynamic programming solution

define **value function** $V : \mathbb{R}^n \rightarrow \mathbb{R}$

$$V(z) = \min_{u(0),...} \sum_{\tau=0}^{\infty} (x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau))$$

subject to $x(0) = z$, $x(\tau + 1) = Ax(\tau) + Bu(\tau)$

- $V(z)$ is the minimum LQR cost-to-go, starting from state $z$
- doesn’t depend on time-to-go, which is always $\infty$; infinite horizon problem is *shift invariant*
**Hamilton-Jacobi equation**

**fact:** $V$ is quadratic, i.e., $V(z) = z^T P z$, where $P = P^T \geq 0$
(can be argued directly from first principles)

**HJ equation:**

$$V(z) = \min_w (z^T Q z + w^T R w + V(Az + Bw))$$

or

$$z^T P z = \min_w (z^T Q z + w^T R w + (Az + Bw)^T P (Az + Bw))$$

minimizing $w$ is $w^* = -(R + B^T P B)^{-1} B^T P A z$

so HJ equation is

$$z^T P z = z^T Q z + w^{*T} R w^* + (Az + Bw^*)^T P (Az + Bw^*)$$
$$= z^T (Q + A^T P A - A^T P B (R + B^T P B)^{-1} B^T P A) z$$

Infinite horizon linear quadratic regulator
this must hold for all $z$, so we conclude that $P$ satisfies the ARE

$$P = Q + A^T P A - A^T P B (R + B^T P B)^{-1} B^T P A$$

and the optimal input is constant state feedback $u(t) = Kx(t)$,

$$K = -(R + B^T P B)^{-1} B^T P A$$

compared to finite-horizon LQR problem,

• value function and optimal state feedback gains are time-invariant
• we don’t have a recursion to compute $P$; we only have the ARE
fact: the ARE has only one positive semidefinite solution $P$

i.e., ARE plus $P = P^T \geq 0$ uniquely characterizes value function

consequence: the Riccati recursion

$$P_{k+1} = Q + A^T P_k A - A^T P_k B (R + B^T P_k B)^{-1} B^T P_k A, \quad P_1 = Q$$

converges to the unique PSD solution of the ARE

(when $(A, B)$ controllable)

(later we’ll see direct methods to solve ARE)

thus, infinite-horizon LQR optimal control is same as steady-state finite horizon optimal control
Receding-horizon LQR control

consider cost function

\[
J_t(u(t), \ldots, u(t + T - 1)) = \sum_{\tau = t}^{\tau = t + T} (x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau))
\]

• \(T\) is called horizon
• same as infinite horizon LQR cost, truncated after \(T\) steps into future

if \((u(t)^*, \ldots, u(t + T - 1)^*)\) minimizes \(J_t\), \(u(t)^*\) is called \((T\)-step ahead) optimal receding horizon control

in words:

• at time \(t\), find input sequence that minimizes \(T\)-step-ahead LQR cost, starting at current time
• then use only the first input
example: 1-step ahead receding horizon control

find $u(t), u(t + 1)$ that minimize

$$J_t = x(t)^T Q x(t) + x(t + 1)^T Q x(t + 1) + u(t)^T R u(t) + u(t + 1)^T R u(t + 1)$$

first term doesn’t matter; optimal choice for $u(t + 1)$ is 0; optimal $u(t)$ minimizes

$$x(t+1)^T Q x(t+1) + u(t)^T R u(t) = (Ax(t) + Bu(t))^T Q (Ax(t) + Bu(t)) + u(t)^T R u(t)$$

thus, 1-step ahead receding horizon optimal input is

$$u(t) = -(R + B^T Q B)^{-1} B^T Q Ax(t)$$

... a constant state feedback
in general, optimal $T$-step ahead LQR control is

$$u(t) = K_T x(t), \quad K_T = -(R + B^T P_T B)^{-1} B^T P_T A$$

where

$$P_1 = Q, \quad P_{i+1} = Q + A^T P_i A - A^T P_i B (R + B^T P_i B)^{-1} B^T P_i A$$

i.e.: same as the optimal finite horizon LQR control, $T - 1$ steps before the horizon $N$

- a constant state feedback
- state feedback gain converges to infinite horizon optimal as horizon becomes long (assuming controllability)
Closed-loop system

suppose $K$ is LQR-optimal state feedback gain

$$x(t + 1) = Ax(t) + Bu(t) = (A + BK)x(t)$$

is called *closed-loop system*

$(x(t + 1) = Ax(t)$ is called *open-loop system*)

is closed-loop system stable? consider

$$x(t + 1) = 2x(t) + u(t), \quad Q = 0, \quad R = 1$$

optimal control is $u(t) = 0x(t)$, *i.e.*, closed-loop system is unstable

**fact:** if $(Q, A)$ observable and $(A, B)$ controllable, then closed-loop system is stable
Lecture 4
Continuous time linear quadratic regulator

- continuous-time LQR problem
- dynamic programming solution
- Hamiltonian system and two point boundary value problem
- infinite horizon LQR
- direct solution of ARE via Hamiltonian
Continuous-time LQR problem

continuous-time system $\dot{x} = Ax + Bu, \ x(0) = x_0$

problem: choose $u : \mathbb{R}_+ \rightarrow \mathbb{R}^m$ to minimize

$$J = \int_0^T x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau) \ d\tau + x(T)^T Q_f x(T)$$

- $T$ is time horizon
- $Q = Q^T \geq 0$, $Q_f = Q_f^T \geq 0$, $R = R^T > 0$ are state cost, final state cost, and input cost matrices

... an infinite-dimensional problem: (trajectory $u : \mathbb{R}_+ \rightarrow \mathbb{R}^m$ is variable)
we’ll solve LQR problem using dynamic programming

for $0 \leq t \leq T$ we define the value function $V_t : \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$V_t(z) = \min_u \int_t^T x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau) \, d\tau + x(T)^T Q_f x(T)$$

subject to $x(t) = z$, $\dot{x} = Ax + Bu$

• minimum is taken over all possible signals $u : [t, T] \rightarrow \mathbb{R}^m$
• $V_t(z)$ gives the minimum LQR cost-to-go, starting from state $z$ at time $t$
• $V_T(z) = z^T Q_f z$
**fact:** $V_t$ is quadratic, *i.e.*, $V_t(z) = z^T P_t z$, where $P_t = P_t^T \geq 0$

similar to discrete-time case:

- $P_t$ can be found from a *differential equation* running backward in time from $t = T$
- the LQR optimal $u$ is easily expressed in terms of $P_t$
we start with $x(t) = z$

let’s take $u(t) = w \in \mathbb{R}^m$, a constant, over the time interval $[t, t + h]$, where $h > 0$ is small

cost incurred over $[t, t + h]$ is

$$
\int_{t}^{t+h} x(\tau)^T Q x(\tau) + w^T R w \, d\tau \approx h(z^T Q z + w^T R w)
$$

and we end up at $x(t + h) \approx z + h(Az + Bw)$
min-cost-to-go from where we land is approximately

\[
V_{t+h}(z + h(Az + Bw)) \\
= (z + h(Az + Bw))^T P_{t+h}(z + h(Az + Bw)) \\
\approx (z + h(Az + Bw))^T (P_t + h\hat{P}_t)(z + h(Az + Bw)) \\
\approx z^T P_t z + h\left((Az + Bw)^T P_t z + z^T P_t (Az + Bw) + z^T \hat{P}_t z\right)
\]

(dropping \(h^2\) and higher terms)

cost incurred plus min-cost-to-go is approximately

\[
z^T P_t z + h\left(z^T Q z + w^T R w + (Az + Bw)^T P_t z + z^T P_t (Az + Bw) + z^T \hat{P}_t z\right)
\]

minimize over \(w\) to get (approximately) optimal \(w\):

\[
2hw^T R + 2hz^T P_t B = 0
\]
so

\[ w^* = -R^{-1} B^T P_t z \]

thus optimal \( u \) is time-varying linear state feedback:

\[ u_{\text{lqr}}(t) = K_t x(t), \quad K_t = -R^{-1} B^T P_t \]
HJ equation

now let’s substitute \( w^* \) into HJ equation:

\[
\begin{align*}
  z^T P_t z &\approx z^T P_t z + \\
  &+ h \left( z^T Q z + w^*^T R w^* + (A z + B w^*)^T P_t z + z^T P_t (A z + B w^*) + z^T \dot{P}_t z \right)
\end{align*}
\]

yields, after simplification,

\[
-\dot{P}_t = A^T P_t + P_t A - P_t B R^{-1} B^T P_t + Q
\]

which is the \textit{Riccati differential equation} for the LQR problem.

we can solve it (numerically) using the \textit{final condition} \( P_T = Q_f \)
Summary of cts-time LQR solution via DP

1. solve Riccati differential equation

\[
-\dot{P}_t = A^T P_t + P_t A - P_t B R^{-1} B^T P_t + Q, \quad P_T = Q_f
\]

(backward in time)

2. optimal \( u \) is \( u_{\text{lqr}}(t) = K_t x(t), \quad K_t := -R^{-1} B^T P_t \)

DP method readily extends to time-varying \( A, B, Q, R, \) and tracking problem
Steady-state regulator

usually $P_t$ rapidly converges as $t$ decreases below $T$

limit $P_{ss}$ satisfies (cts-time) algebraic Riccati equation (ARE)

$$A^T P + PA - PB R^{-1} B^T P + Q = 0$$

a quadratic matrix equation

• $P_{ss}$ can be found by (numerically) integrating the Riccati differential equation, or by direct methods

• for $t$ not close to horizon $T$, LQR optimal input is approximately a linear, constant state feedback

$$u(t) = K_{ss} x(t), \quad K_{ss} = -R^{-1} B^T P_{ss}$$
Derivation via discretization

let’s discretize using small step size $h > 0$, with $Nh = T$

$$x((k + 1)h) \approx x(kh) + h\dot{x}(kh) = \left(I + hA\right)x(kh) + hBu(kh)$$

$$J \approx \frac{h}{2} \sum_{k=0}^{N-1} \left(x(kh)^TQx(kh) + u(kh)^TRu(kh)\right) + \frac{1}{2}x(Nh)^TQ_fx(Nh)$$

this yields a discrete-time LQR problem, with parameters

$$\tilde{A} = I + hA, \quad \tilde{B} = hB, \quad \tilde{Q} = hQ, \quad \tilde{R} = hR, \quad \tilde{Q}_f = Q_f$$
solution to discrete-time LQR problem is \( u(kh) = \tilde{K}_k x(kh) \),

\[
\tilde{K}_k = -(\tilde{R} + \tilde{B}^T \tilde{P}_{k+1} \tilde{B})^{-1} \tilde{B}^T \tilde{P}_{k+1} \tilde{A}
\]

\[
\tilde{P}_{k-1} = \tilde{Q} + \tilde{A}^T \tilde{P}_k \tilde{A} - \tilde{A}^T \tilde{P}_k \tilde{B} (\tilde{R} + \tilde{B}^T \tilde{P}_k \tilde{B})^{-1} \tilde{B}^T \tilde{P}_k \tilde{A}
\]

substituting and keeping only \( h^0 \) and \( h^1 \) terms yields

\[
\tilde{P}_{k-1} = h Q + \tilde{P}_k + h A^T \tilde{P}_k + h \tilde{P}_k A - h \tilde{P}_k B R^{-1} B^T \tilde{P}_k
\]

which is the same as

\[
-\frac{1}{h}(\tilde{P}_k - \tilde{P}_{k-1}) = Q + A^T \tilde{P}_k + \tilde{P}_k A - \tilde{P}_k B R^{-1} B^T \tilde{P}_k
\]

letting \( h \to 0 \) we see that \( \tilde{P}_k \to P_{kh} \), where

\[
-\dot{P} = Q + A^T P + PA - PBR^{-1} B^T P
\]
similarly, we have

\[
\tilde{K}_k = -(R + \tilde{B}^T \tilde{P}_{k+1} \tilde{B})^{-1} \tilde{B}^T \tilde{P}_{k+1} \tilde{\Lambda} \\
= -(hR + h^2 B^T \tilde{P}_{k+1} B)^{-1} h B^T \tilde{P}_{k+1} (I + hA) \\
\rightarrow -R^{-1} B^T P_{kh}
\]

as \( h \rightarrow 0 \)
Derivation using Lagrange multipliers

pose as constrained problem:

\[
\begin{align*}
\text{minimize} \quad & J = \frac{1}{2} \int_0^T x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau) \, d\tau + \frac{1}{2} x(T)^T Q_f x(T) \\
\text{subject to} \quad & \dot{x}(t) = A x(t) + B u(t), \quad t \in [0, T]
\end{align*}
\]

• optimization variable is function \( u : [0, T] \rightarrow \mathbb{R}^m \)
• infinite number of equality constraints, one for each \( t \in [0, T] \)

introduce Lagrange multiplier function \( \lambda : [0, T] \rightarrow \mathbb{R}^n \) and form

\[
L = J + \int_0^T \lambda(\tau)^T (A x(\tau) + B u(\tau) - \dot{x}(\tau)) \, d\tau
\]
Optimality conditions

(note: you need distribution theory to really make sense of the derivatives here . . . )

from $\nabla_{u(t)} L = Ru(t) + B^T \lambda(t) = 0$ we get $u(t) = -R^{-1}B^T \lambda(t)$

to find $\nabla_{x(t)} L$, we use

$$\int_0^T \lambda(\tau)^T \dot{x}(\tau) \, d\tau = \lambda(T)^T x(T) - \lambda(0)^T x(0) - \int_0^T \dot{\lambda}(\tau)^T x(\tau) \, d\tau$$

from $\nabla_{x(t)} L = Qx(t) + A^T \lambda(t) + \dot{\lambda}(t) = 0$ we get

$$\dot{\lambda}(t) = -A^T \lambda(t) - Qx(t)$$

from $\nabla_{x(T)} L = Q_f x(T) - \lambda(T) = 0$, we get $\lambda(T) = Q_f x(T)$
Co-state equations

optimality conditions are

\[ \dot{x} = Ax + Bu, \quad x(0) = x_0, \quad \dot{\lambda} = -A^T \lambda - Qx, \quad \lambda(T) = Q_f x(T) \]

using \( u(t) = -R^{-1} B^T \lambda(t) \), can write as

\[ \frac{d}{dt} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} \]

• 2n \times 2n matrix above is called Hamiltonian for problem

• with conditions \( x(0) = x_0, \lambda(T) = Q_f x(T) \), called two-point boundary value problem
as in discrete-time case, we can show that \( \lambda(t) = P_tx(t) \), where

\[
-\dot{P}_t = A^T P_t + P_tA - P_tBR^{-1}B^TP_t + Q, \quad P_T = Q_f
\]

in other words, value function \( P_t \) gives simple relation between \( x \) and \( \lambda \) to show this, we show that \( \lambda = Px \) satisfies co-state equation

\[
\dot{\lambda} = -A^T\lambda - Qx
\]

\[
\dot{\lambda} = \frac{d}{dt}(Px) = \dot{P}x + P\dot{x} = -(Q + A^T P + PA - PBR^{-1}B^TP)x + P(Ax - BR^{-1}B^T\lambda) = -Qx - A^TPx + PBR^{-1}B^TPx - PBR^{-1}B^TPx = -Qx - A^T\lambda
\]
**Solving Riccati differential equation via Hamiltonian**

the (quadratic) Riccati differential equation

$$-\dot{P} = A^T P + PA - PBR^{-1}B^TP + Q$$

and the (linear) Hamiltonian differential equation

$$\frac{d}{dt} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} \begin{bmatrix} x(t) \\ \lambda(t) \end{bmatrix}$$

are closely related

$$\lambda(t) = P_t x(t)$$ suggests that $P$ should have the form $P_t = \lambda(t)x(t)^{-1}$

(but this doesn’t make sense unless $x$ and $\lambda$ are scalars)
consider the Hamiltonian matrix (linear) differential equation

$$\frac{d}{dt} \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{bmatrix} \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix}$$

where $X(t), Y(t) \in \mathbb{R}^{n \times n}$

then, $Z(t) = Y(t)X(t)^{-1}$ satisfies Riccati differential equation

$$-\dot{Z} = A^T Z + ZA - ZBR^{-1}B^T Z + Q$$

hence we can solve Riccati DE by solving (linear) matrix Hamiltonian DE, with final conditions $X(T) = I$, $Y(T) = Q_f$, and forming $P(t) = Y(t)X(t)^{-1}$
\[
\dot{Z} = \frac{d}{dt} Y X^{-1}
\]
\[
= \dot{Y} X^{-1} - Y X^{-1} \dot{X} X^{-1}
\]
\[
= (-QX - A^T Y)X^{-1} - Y X^{-1} (AX - BR^{-1} B^T Y) X^{-1}
\]
\[
= -Q - A^T Z - ZA + ZBR^{-1} B^T Z
\]

where we use two identities:

- \( \frac{d}{dt} (F(t)G(t)) = \dot{F}(t)G(t) + F(t)\dot{G}(t) \)
- \( \frac{d}{dt} (F(t)^{-1}) = -F(t)^{-1} \dot{F}(t)F(t)^{-1} \)
Infinite horizon LQR

we now consider the infinite horizon cost function

\[ J = \int_0^{\infty} x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau) \, d\tau \]

we define the value function as

\[ V(z) = \min_u \int_0^{\infty} x(\tau)^T Q x(\tau) + u(\tau)^T R u(\tau) \, d\tau \]

subject to \( x(0) = z, \dot{x} = Ax + Bu \)

we assume that \((A, B)\) is controllable, so \(V\) is finite for all \(z\)

can show that \(V\) is quadratic: \( V(z) = z^T P z \), where \( P = P^T \geq 0 \)
optimal $u$ is $u(t) = Kx(t)$, where $K = -R^{-1}B^TP$ (i.e., a constant linear state feedback)

HJ equation is ARE

$$Q + A^TP + PA - PBR^{-1}B^TP = 0$$

which together with $P \geq 0$ characterizes $P$

can solve as limiting value of Riccati DE, or via direct method
Closed-loop system

with $K$ LQR optimal state feedback gain, closed-loop system is

$$\dot{x} = Ax + Bu = (A + BK)x$$

**fact:** closed-loop system is stable when $(Q, A)$ observable and $(A, B)$ controllable

we denote eigenvalues of $A + BK$, called *closed-loop eigenvalues*, as $\lambda_1, \ldots, \lambda_n$

with assumptions above, $\Re \lambda_i < 0$
Solving ARE via Hamiltonian

\[
\begin{bmatrix}
A & -BR^{-1}B^T \\
-Q & -A^T
\end{bmatrix}
\begin{bmatrix}
I \\
P
\end{bmatrix}
= \begin{bmatrix}
A - BR^{-1}B^TP \\
-Q - A^TP
\end{bmatrix}
= \begin{bmatrix}
A + BK \\
-Q - A^TP
\end{bmatrix}
\]

and so

\[
\begin{bmatrix}
I & 0 \\
-P & I
\end{bmatrix}
\begin{bmatrix}
A & -BR^{-1}B^T \\
-Q & -A^T
\end{bmatrix}
\begin{bmatrix}
I & 0 \\
P & I
\end{bmatrix}
= \begin{bmatrix}
A + BK & -BR^{-1}B^T \\
0 & -(A + BK)^T
\end{bmatrix}
\]

where 0 in lower left corner comes from ARE

note that

\[
\begin{bmatrix}
I & 0 \\
P & I
\end{bmatrix}^{-1} = \begin{bmatrix}
I & 0 \\
-P & I
\end{bmatrix}
\]
we see that:

- eigenvalues of Hamiltonian $H$ are $\lambda_1, \ldots, \lambda_n$ and $-\lambda_1, \ldots, -\lambda_n$

- hence, closed-loop eigenvalues are the eigenvalues of $H$ with negative real part
let’s assume $A + BK$ is diagonalizable, i.e.,

$$T^{-1}(A + BK)T = \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$$

then we have $T^T(-A - BK)^T T^{-T} = -\Lambda$, so

$$\begin{bmatrix} T^{-1} & 0 \\ 0 & T^T \end{bmatrix} \begin{bmatrix} A + BK & -BR^{-1}B^T \\ 0 & -(A + BK)^T \end{bmatrix} \begin{bmatrix} T & 0 \\ 0 & T^{-T} \end{bmatrix}$$

$$= \begin{bmatrix} \Lambda & -T^{-1}BR^{-1}BT TT^{-T} \\ 0 & -\Lambda \end{bmatrix}$$
putting it together we get

\[
\begin{bmatrix}
T^{-1} & 0 \\
0 & TT
\end{bmatrix}
\begin{bmatrix}
I & 0 \\
-P & I
\end{bmatrix}
H
\begin{bmatrix}
I & 0 \\
P & I
\end{bmatrix}
\begin{bmatrix}
T & 0 \\
0 & T^{-T}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
T^{-1} & 0 \\
-TTP & TT
\end{bmatrix}
H
\begin{bmatrix}
T & 0 \\
PT & T^{-T}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\Lambda & -T^{-1}BR^{-1}B^TT^{-T} \\
0 & -\Lambda
\end{bmatrix}
\]

and so

\[
H\begin{bmatrix}
T \\
PT
\end{bmatrix} = \begin{bmatrix}
T \\
PT
\end{bmatrix}\Lambda
\]

thus, the \( n \) columns of \( \begin{bmatrix} T \\ PT \end{bmatrix} \) are the eigenvectors of \( H \) associated with the stable eigenvalues \( \lambda_1, \ldots, \lambda_n \)
Solving ARE via Hamiltonian

- find eigenvalues of $H$, and let $\lambda_1, \ldots, \lambda_n$ denote the $n$ stable ones (there are exactly $n$ stable and $n$ unstable ones)

- find associated eigenvectors $v_1, \ldots, v_n$, and partition as

$$
\begin{bmatrix}
v_1 & \cdots & v_n
\end{bmatrix} = 
\begin{bmatrix}
X \\
Y
\end{bmatrix} \in \mathbb{R}^{2n \times n}
$$

- $P = YX^{-1}$ is unique PSD solution of the ARE

(this is very close to the method used in practice, which does not require $A + BK$ to be diagonalizable)
Lecture 5
Observability and state estimation

- state estimation
- discrete-time observability
- observability – controllability duality
- observers for noiseless case
- continuous-time observability
- least-squares observers
- statistical interpretation
- example
State estimation set up

we consider the discrete-time system

\[ x(t + 1) = Ax(t) + Bu(t) + w(t), \quad y(t) = Cx(t) + Du(t) + v(t) \]

- \( w \) is state disturbance or noise
- \( v \) is sensor noise or error
- \( A, B, C, \) and \( D \) are known
- \( u \) and \( y \) are observed over time interval \([0, t - 1]\)
- \( w \) and \( v \) are not known, but can be described statistically or assumed small
State estimation problem

state estimation problem: estimate \( x(s) \) from

\[
u(0), \ldots, u(t - 1), y(0), \ldots, y(t - 1)\]

- \( s = 0 \): estimate initial state
- \( s = t - 1 \): estimate current state
- \( s = t \): estimate (i.e., predict) next state

an algorithm or system that yields an estimate \( \hat{x}(s) \) is called an observer or state estimator

\( \hat{x}(s) \) is denoted \( \hat{x}(s|t - 1) \) to show what information estimate is based on (read, “\( \hat{x}(s) \) given \( t - 1 \)”)
Noiseless case

let’s look at finding \( x(0) \), with no state or measurement noise:

\[
x(t + 1) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t)
\]

with \( x(t) \in \mathbb{R}^n, u(t) \in \mathbb{R}^m, y(t) \in \mathbb{R}^p \)

then we have

\[
\begin{bmatrix}
  y(0) \\
  \vdots \\
  y(t - 1)
\end{bmatrix} = \mathcal{O}_t x(0) + \mathcal{T}_t \begin{bmatrix}
  u(0) \\
  \vdots \\
  u(t - 1)
\end{bmatrix}
\]
where

\[
O_t = \begin{bmatrix}
  C & \cdot & \cdot \\
  CA & \cdot & \cdot \\
  \vdots & \cdot & \cdot \\
  CA^{t-1} & \cdot & \cdot \\
\end{bmatrix}, \quad T_t = \begin{bmatrix}
  D & 0 & \cdots \\
  CB & D & 0 & \cdots \\
  \vdots & \cdot & \cdot & \cdot \\
  CA^{t-2}B & CA^{t-3}B & \cdots & CB & D \\
\end{bmatrix}
\]

- \(O_t\) maps initials state into resulting output over \([0, t-1]\)
- \(T_t\) maps input to output over \([0, t-1]\)

hence we have

\[
O_t x(0) = \begin{bmatrix}
  y(0) \\
  \vdots \\
  y(t-1)
\end{bmatrix} - T_t \begin{bmatrix}
  u(0) \\
  \vdots \\
  u(t-1)
\end{bmatrix}
\]

RHS is known, \(x(0)\) is to be determined
hence:

• can uniquely determine $x(0)$ if and only if $\mathcal{N}(\mathcal{O}_t) = \{0\}$

• $\mathcal{N}(\mathcal{O}_t)$ gives ambiguity in determining $x(0)$

• if $x(0) \in \mathcal{N}(\mathcal{O}_t)$ and $u = 0$, output is zero over interval $[0, t - 1]$

• input $u$ does not affect ability to determine $x(0)$; its effect can be subtracted out
Observability matrix

by C-H theorem, each $A^k$ is linear combination of $A^0, \ldots, A^{n-1}$

hence for $t \geq n$, $\mathcal{N}(\mathcal{O}_t) = \mathcal{N}(\mathcal{O})$ where

$$\mathcal{O} = \mathcal{O}_n = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}$$

is called the observability matrix

if $x(0)$ can be deduced from $u$ and $y$ over $[0, t - 1]$ for any $t$, then $x(0)$ can be deduced from $u$ and $y$ over $[0, n - 1]$

$\mathcal{N}(\mathcal{O})$ is called unobservable subspace; describes ambiguity in determining state from input and output

system is called observable if $\mathcal{N}(\mathcal{O}) = \{0\}$, i.e., $\text{Rank}(\mathcal{O}) = n$
Observability – controllability duality

let \((\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D})\) be dual of system \((A, B, C, D)\), i.e.,

\[
\tilde{A} = A^T, \quad \tilde{B} = C^T, \quad \tilde{C} = B^T, \quad \tilde{D} = D^T
\]

controllability matrix of dual system is

\[
\tilde{C} = \begin{bmatrix} \tilde{B} & \tilde{A}\tilde{B} & \cdots & \tilde{A}^{n-1}\tilde{B} \\ \end{bmatrix}
= \begin{bmatrix} C^T & A^T C^T & \cdots & (A^T)^{n-1} C^T \end{bmatrix}
= \mathcal{O}^T,
\]

transpose of observability matrix

similarly we have \(\tilde{\mathcal{O}} = C^T\)
thus, system is observable (controllable) if and only if dual system is controllable (observable)

in fact,

\[ N(\mathcal{O}) = \text{range}(\mathcal{O}^T) \perp = \text{range}(\tilde{\mathcal{C}}) \perp \]

i.e., unobservable subspace is orthogonal complement of controllable subspace of dual
Observers for noiseless case

suppose $\text{Rank}(O_t) = n$ (i.e., system is observable) and let $F$ be any left inverse of $O_t$, i.e., $FO_t = I$

then we have the observer

$$x(0) = F \begin{bmatrix} y(0) \\ \vdots \\ y(t-1) \end{bmatrix} - T_t \begin{bmatrix} u(0) \\ \vdots \\ u(t-1) \end{bmatrix}$$

which deduces $x(0)$ (exactly) from $u$, $y$ over $[0, t - 1]$

in fact we have

$$x(\tau - t + 1) = F \begin{bmatrix} y(\tau - t + 1) \\ \vdots \\ y(\tau) \end{bmatrix} - T_t \begin{bmatrix} u(\tau - t + 1) \\ \vdots \\ u(\tau) \end{bmatrix}$$
i.e., our observer estimates what state was \( t - 1 \) epochs ago, given past \( t - 1 \) inputs & outputs

observer is (multi-input, multi-output) finite impulse response (FIR) filter, with inputs \( u \) and \( y \), and output \( \hat{x} \)
Invariance of unobservable set

**fact:** the unobservable subspace $\mathcal{N}(\mathcal{O})$ is invariant, i.e., if $z \in \mathcal{N}(\mathcal{O})$, then $Az \in \mathcal{N}(\mathcal{O})$

**proof:** suppose $z \in \mathcal{N}(\mathcal{O})$, i.e., $CA^k z = 0$ for $k = 0, \ldots, n - 1$

evidently $CA^k(Az) = 0$ for $k = 0, \ldots, n - 2$;

$$CA^{n-1}(Az) = CA^n z = -\sum_{i=0}^{n-1} \alpha_i CA^i z = 0$$

(by C-H) where

$$\det(sI - A) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_0$$
Continuous-time observability

continuous-time system with no sensor or state noise:

\[ \dot{x} = Ax + Bu, \quad y = Cx + Du \]

can we deduce state \( x \) from \( u \) and \( y \)?

let’s look at derivatives of \( y \):

\[
\begin{align*}
y &= Cx + Du \\
\dot{y} &= C\dot{x} + D\dot{u} = CAx + CBu + D\dot{u} \\
\ddot{y} &= CA^2x + CABu + CB\ddot{u} + D\ddot{u}
\end{align*}
\]

and so on
hence we have
\[
\begin{bmatrix}
  y \\
  \dot{y} \\
  \vdots \\
  y^{(n-1)}
\end{bmatrix}
= \mathcal{O}x + \mathcal{T}
\begin{bmatrix}
  u \\
  \dot{u} \\
  \vdots \\
  u^{(n-1)}
\end{bmatrix}
\]

where \( \mathcal{O} \) is the observability matrix and and

\[
\mathcal{T} = \begin{bmatrix}
  D & 0 & \cdots \\
  CB & D & 0 & \cdots \\
  \vdots & \vdots & \ddots & \ddots \\
  CA^{n-2}B & CA^{n-3}B & \cdots & CB & D
\end{bmatrix}
\]

(same matrices we encountered in discrete-time case!)
rewrite as

\[ \mathcal{O} x = \begin{bmatrix}
  y \\
  \dot{y} \\
  \vdots \\
  y^{(n-1)}
\end{bmatrix} - T \begin{bmatrix}
  u \\
  \dot{u} \\
  \vdots \\
  u^{(n-1)}
\end{bmatrix} \]

RHS is known; \( x \) is to be determined

hence if \( \mathcal{N}(\mathcal{O}) = \{0\} \) we can deduce \( x(t) \) from derivatives of \( u(t), y(t) \) up to order \( n - 1 \)

in this case we say system is observable

can construct an observer using any left inverse \( F \) of \( \mathcal{O} \):

\[ x = F \left( \begin{bmatrix}
  y \\
  \dot{y} \\
  \vdots \\
  y^{(n-1)}
\end{bmatrix} - T \begin{bmatrix}
  u \\
  \dot{u} \\
  \vdots \\
  u^{(n-1)}
\end{bmatrix} \right) \]
• reconstructs $x(t)$ (exactly and instantaneously) from

$$u(t), \ldots, u^{(n-1)}(t), y(t), \ldots, y^{(n-1)}(t)$$

• derivative-based state reconstruction is dual of state transfer using impulsive inputs
A converse

suppose \( z \in \mathcal{N}(\mathcal{O}) \) (the unobservable subspace), and \( u \) is any input, with \( x, y \) the corresponding state and output, \( i.e., \)

\[
\dot{x} = Ax + Bu, \quad y = Cx + Du
\]

then state trajectory \( \tilde{x} = x + e^{At}z \) satisfies

\[
\dot{\tilde{x}} = A\tilde{x} + Bu, \quad y = C\tilde{x} + Du
\]

\( i.e., \) input/output signals \( u, y \) consistent with both state trajectories \( x, \tilde{x} \) hence if system is unobservable, no signal processing of any kind applied to \( u \) and \( y \) can deduce \( x \)

unobservable subspace \( \mathcal{N}(\mathcal{O}) \) gives fundamental ambiguity in deducing \( x \) from \( u, y \)
Least-squares observers

discrete-time system, with sensor noise:

\[ x(t + 1) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t) + v(t) \]

we assume \( \text{Rank}(O_t) = n \) (hence, system is observable)

\textit{least-squares} observer uses pseudo-inverse:

\[
\hat{x}(0) = O_t^\dagger \begin{bmatrix} y(0) \\ \vdots \\ y(t - 1) \end{bmatrix} - T_t \begin{bmatrix} u(0) \\ \vdots \\ u(t - 1) \end{bmatrix}
\]

where \( O_t^\dagger = (O_t^T O_t)^{-1} O_t^T \)
since $O_t^\dagger O_t = I$, we have

$$\hat{x}_{1s}(0) = x(0) + O_t^\dagger \begin{bmatrix} v(0) \\ \vdots \\ v(t - 1) \end{bmatrix}$$

in particular, $\hat{x}_{1s}(0) = x(0)$ if sensor noise is zero (i.e., observer recovers exact state in noiseless case)
**interpretation:** $\hat{x}_{ls}(0)$ minimizes discrepancy between

- output $\hat{y}$ that *would be* observed, with input $u$ and initial state $x(0)$ (and no sensor noise), and

- output $y$ that *was* observed,

measured as

$$\sum_{\tau=0}^{t-1} \|\hat{y}(\tau) - y(\tau)\|^2$$

can express least-squares initial state estimate as

$$\hat{x}_{ls}(0) = \left( \sum_{\tau=0}^{t-1} (A^T)^\tau C^T C A^\tau \right)^{-1} \sum_{\tau=0}^{t-1} (A^T)^\tau C^T \tilde{y}(\tau)$$

where $\tilde{y}$ is observed output with portion due to input subtracted:

$$\tilde{y} = y - h \ast u$$

where $h$ is impulse response
Statistical interpretation of least-squares observer

suppose sensor noise is IID $\mathcal{N}(0, \sigma I)$

- called *white noise*
- each sensor has noise variance $\sigma$

then $\hat{x}_{ls}(0)$ is MMSE estimate of $x(0)$ when $x(0)$ is deterministic (or has ‘infinite’ prior variance)

estimation error $z = \hat{x}_{ls}(0) - x(0)$ can be expressed as

$$z = O_t^\dagger \begin{bmatrix} v(0) \\ \vdots \\ v(t - 1) \end{bmatrix}$$

hence $z \sim \mathcal{N}(0, \sigma O^\dagger O^\dagger T)$
\[ \sigma O^\dagger O^\dagger T = \sigma \left( \sum_{\tau=0}^{t-1} (A^T)^\tau C^T C A^\tau \right)^{-1} \]

we’ll assume \( \sigma = 1 \) to simplify

matrix \( \left( \sum_{\tau=0}^{t-1} (A^T)^\tau C^T C A^\tau \right)^{-1} \) gives measure of ‘how observable’ the state is, over \( [0, t - 1] \)
Infinite horizon error covariance

the matrix

\[ P = \lim_{t \to \infty} \left( \sum_{\tau=0}^{t-1} (A^T)^\tau C^T CA^\tau \right)^{-1} \]

always exists, and gives the limiting error covariance in estimating \( x(0) \) from \( u, y \) over longer and longer periods:

\[ \lim_{t \to \infty} \mathbf{E}(\hat{x}_{1s}(0|t-1) - x(0))(\hat{x}_{1s}(0|t-1) - x(0))^T = P \]

- if \( A \) is stable, \( P > 0 \)
  - \( i.e. \), can’t estimate initial state perfectly even with infinite number of measurements \( u(t), y(t), t = 0, \ldots \) (since memory of \( x(0) \) fades . . . )

- if \( A \) is not stable, then \( P \) can have nonzero nullspace
  - \( i.e. \), initial state estimation error gets arbitrarily small (at least in some directions) as more and more of signals \( u \) and \( y \) are observed
Observability Gramian

suppose system

\[ x(t + 1) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t) \]

is observable and stable

\[ \sum_{\tau=0}^{t-1} (A^T)\tau C^T C A^\tau \] converges as \( t \to \infty \) since \( A^\tau \) decays geometrically

the matrix \( W_o = \sum_{\tau=0}^{\infty} (A^T)\tau C^T C A^\tau \) is called the observability Gramian

\( W_o \) satisfies the matrix equation

\[ W_o - A^T W_o A = C^T C \]

which is called the observability Lyapunov equation (and can be solved exactly and efficiently)
Current state estimation

we have concentrated on estimating $x(0)$ from

$$u(0), \ldots, u(t - 1), y(0), \ldots, y(t - 1)$$

now we look at estimating $x(t - 1)$ from this data

we assume

$$x(t + 1) = Ax(t) + Bu(t), \quad y(t) = Cx(t) + Du(t) + v(t)$$

• no state noise
• $v$ is white, i.e., IID $\mathcal{N}(0, \sigma I)$

using

$$x(t - 1) = A^{t-1}x(0) + \sum_{\tau=0}^{t-2} A^{t-2-\tau} Bu(\tau)$$
we get current state least-squares estimator:

\[
\hat{x}(t - 1|t - 1) = A^{t-1}\hat{x}_{ls}(0|t - 1) + \sum_{\tau=0}^{t-2} A^{t-2-\tau} Bu(\tau)
\]

righthand term \((i.e., \text{effect of input on current state})\) is known estimation error \(z = \hat{x}(t - 1|t - 1) - x(t - 1)\) can be expressed as

\[
z = A^{t-1}O_t^\dagger \begin{bmatrix} v(0) \\ \vdots \\ v(t - 1) \end{bmatrix}
\]

hence \(z \sim \mathcal{N}(0, \sigma A^{t-1}O^\dagger O^{\dagger T}(A^T)^{t-1})\)

\(i.e., \text{covariance of least-squares current state estimation error is} \)

\[
\sigma A^{t-1}O^\dagger O^{\dagger T}(A^T)^{t-1} = \sigma A^{t-1} \left( \sum_{\tau=0}^{t-1} (A^T)^{\tau} C^T C A^\tau \right)^{-1} (A^T)^{t-1}
\]
this matrix measures ‘how observable’ current state is, from past $t$ inputs & outputs

- decreases (in matrix sense) as $t$ increases
- hence has limit as $t \to \infty$ (gives limiting error covariance of estimating current state given all past inputs & outputs)
Example

- particle in $\mathbb{R}^2$ moves with uniform velocity
- (linear, noisy) range measurements from directions $-15^\circ, 0^\circ, 20^\circ, 30^\circ$, once per second
- range noises IID $\mathcal{N}(0,1)$
- no assumptions about initial position & velocity

**problem:** estimate initial position & velocity from range measurements
express as linear system

\[
x(t + 1) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x(t), \quad y(t) = \begin{bmatrix} k_1^T \\ \vdots \\ k_4^T \end{bmatrix} x(t) + v(t)
\]

• \((x_1(t), x_2(t))\) is position of particle

• \((x_3(t), x_4(t))\) is velocity of particle

• \(v(t) \sim \mathcal{N}(0, I)\)

• \(k_i\) is unit vector from sensor \(i\) to origin

true initial position & velocities: \(x(0) = (1 - 3 - 0.04 0.03)\)
range measurements (& noiseless versions):

![Graph showing range measurements from sensors 1 to 4 over time.](image-url)
• estimate based on \((y(0), \ldots, y(t))\) is \(\hat{x}(0|t)\)

• actual RMS position error is

\[
\sqrt{(\hat{x}_1(0|t) - x_1(0))^2 + (\hat{x}_2(0|t) - x_2(0))^2}
\]

(similarly for actual RMS velocity error)

• position error std. deviation is

\[
\sqrt{\mathbf{E}((\hat{x}_1(0|t) - x_1(0))^2 + (\hat{x}_2(0|t) - x_2(0))^2)}
\]

(similarly for velocity)
Example ctd: state prediction

predict particle position 10 seconds in future:

\[ \hat{x}(t + 10|t) = A^{t+10} \hat{x}_1(0|t) \]

\[ x(t + 10) = A^{t+10} x(0) \]

plot shows estimates (dashed), and actual value (solid) of position of particle 10 steps ahead, for \(10 \leq t \leq 110\)
\[ \hat{x}_1(t|t-10), \hat{x}_2(t|t-10) \]
Continuous-time least-squares state estimation

assume $\dot{x} = Ax + Bu$, $y = Cx + Du + v$ is observable

least-squares observer is

$$\hat{x}_{ls}(0) = \left( \int_0^t e^{AT}\tau C^T C e^{AT} \, d\tau \right)^{-1} \int_0^t e^{AT\bar{t}} C^T \tilde{y}(\bar{t}) \, d\bar{t}$$

where $\tilde{y} = y - h * u$ is observed output minus part due to input

then $\hat{x}_{ls}(0) = x(0)$ if $v = 0$

$\hat{x}_{ls}(0)$ is limiting MMSE estimate when $v(t) \sim \mathcal{N}(0, \sigma I)$ and $\mathbf{E} v(t)v(s)^T = 0$ unless $t - s$ is very small

(called white noise — a tricky concept)
Lecture 5
Invariant subspaces

• invariant subspaces

• a matrix criterion

• Sylvester equation

• the PBH controllability and observability conditions

• invariant subspaces, quadratic matrix equations, and the ARE
Invariant subspaces

suppose \( A \in \mathbb{R}^{n \times n} \) and \( \mathcal{V} \subseteq \mathbb{R}^n \) is a subspace

we say that \( \mathcal{V} \) is \( A \)-invariant if \( A\mathcal{V} \subseteq \mathcal{V} \), i.e., \( v \in \mathcal{V} \implies Av \in \mathcal{V} \)

examples:

- \( \{0\} \) and \( \mathbb{R}^n \) are always \( A \)-invariant
- \( \text{span}\{v_1, \ldots, v_m\} \) is \( A \)-invariant, where \( v_i \) are (right) eigenvectors of \( A \)
- if \( A \) is block upper triangular,

\[
A = \begin{bmatrix}
A_{11} & A_{12} \\
0 & A_{22}
\end{bmatrix},
\]

with \( A_{11} \in \mathbb{R}^{r \times r} \), then \( \mathcal{V} = \left\{ \begin{bmatrix} \tilde{z} \\ 0 \end{bmatrix} \mid \tilde{z} \in \mathbb{R}^r \right\} \) is \( A \)-invariant
Examples from linear systems

- if $B \in \mathbb{R}^{n \times m}$, then the controllable subspace

$$\mathcal{R}(C) = \mathcal{R} ([B \ AB \ \cdots \ A^{n-1}B])$$

is $A$-invariant

- if $C \in \mathbb{R}^{p \times n}$, then the unobservable subspace

$$\mathcal{N}(O) = \mathcal{N} \left( \begin{bmatrix} C \\
\vdots \\
CA^{n-1} \end{bmatrix} \right)$$

is $A$-invariant
Dynamical interpretation

consider system $\dot{x} = Ax$

$\mathcal{V}$ is $A$-invariant if and only if

$x(0) \in \mathcal{V} \implies x(t) \in \mathcal{V}$ for all $t \geq 0$

(same statement holds for discrete-time system)
A matrix criterion for $A$-invariance

suppose $\mathcal{V}$ is $A$-invariant

let columns of $M \in \mathbb{R}^{n \times k}$ span $\mathcal{V}$, i.e.,

$$\mathcal{V} = \mathcal{R}(M) = \mathcal{R}([t_1 \cdots t_k])$$

since $A t_1 \in \mathcal{V}$, we can express it as

$$A t_1 = x_{11} t_1 + \cdots + x_{k1} t_k$$

we can do the same for $A t_2, \ldots, A t_k$, which gives

$$A[t_1 \cdots t_k] = [t_1 \cdots t_k] \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{k1} & \cdots & x_{kk} \end{bmatrix}$$

or, simply, $A M = M X$
in other words: if $\mathcal{R}(M)$ is $A$-invariant, then there is a matrix $X$ such that $AM = MX$

converse is also true: if there is an $X$ such that $AM = MX$, then $\mathcal{R}(M)$ is $A$-invariant

now assume $M$ is rank $k$, i.e., $\{t_1, \ldots, t_k\}$ is a basis for $V$

then every eigenvalue of $X$ is an eigenvalue of $A$, and the associated eigenvector is in $V = \mathcal{R}(M)$

if $Xu = \lambda u, u \neq 0$, then $Mu \neq 0$ and $A(Mu) = MXu = \lambda Mu$

so the eigenvalues of $X$ are a subset of the eigenvalues of $A$

more generally: if $AM = MX$ (no assumption on rank of $M$), then $A$ and $X$ share at least $\text{Rank}(M)$ eigenvalues
Sylvester equation

the **Sylvester equation** is \( AX + XB = C \), where \( A, B, C, X \in \mathbb{R}^{n \times n} \)

when does this have a solution \( X \) for every \( C \)?

express as \( S(X) = C \), where \( S \) is the linear function \( S(X) = AX + XB \)
(\( S \) maps \( \mathbb{R}^{n \times n} \) into \( \mathbb{R}^{n \times n} \) and is called the **Sylvester operator**)

so the question is: when is \( S \) nonsingular?

\( S \) is singular if and only if there exists a nonzero \( X \) with \( S(X) = 0 \)

this means \( AX + XB = 0 \), so \( AX = X(-B) \), which means \( A \) and \(-B\) share at least one eigenvalue (since \( X \neq 0 \))

so we have: if \( S \) is singular, then \( A \) and \(-B\) have a common eigenvalue
let’s show the converse: if $A$ and $-B$ share an eigenvalue, $S$ is singular

suppose

$$Av = \lambda v, \quad w^T B = -\lambda w^T, \quad v, w \neq 0$$

then with $X = vw^T$ we have $X \neq 0$ and

$$S(X) = AX + XB = Avw^T + vw^T B = (\lambda v)w^T + v(-\lambda w^T) = 0$$

which shows $S$ is singular

so, Sylvestor operator is singular if and only if $A$ and $-B$ have a common eigenvalue

or: Sylvestor operator is nonsingular if and only if $A$ and $-B$ have no common eigenvalues
Uniqueness of stabilizing ARE solution

suppose $P$ is any solution of ARE

\[ A^T P + PA + Q - PBR^{-1}B^TP = 0 \]

and define $K = -R^{-1}B^TP$

we say $P$ is a stabilizing solution of ARE if

\[ A + BK = A - BR^{-1}B^TP \]

is stable, i.e., its eigenvalues have negative real part

**fact:** there is at most one stabilizing solution of the ARE (which therefore is the one that gives the value function)
to show this, suppose $P_1$ and $P_2$ are both stabilizing solutions

subtract AREs to get

$$A^T(P_1 - P_2) + (P_1 - P_2)A - P_1BR^{-1}B^TP_1 + P_2BR^{-1}B^TP_2 = 0$$

rewrite as Sylvester equation

$$(A + BK_2)^T(P_1 - P_2) + (P_1 - P_2)(A + BK_1) = 0$$

since $A + BK_2$ and $A + BK_1$ are both stable, $A + BK_2$ and $-(A + BK_1)$ cannot share any eigenvalues, so we conclude $P_1 - P_2 = 0$
Change of coordinates

suppose \( \mathcal{V} = \mathcal{R}(M) \) is \( A \)-invariant, where \( M \in \mathbb{R}^{n \times k} \) is rank \( k \)

find \( \tilde{M} \in \mathbb{R}^{n \times (n-k)} \) so that \( [M \ \tilde{M}] \) is nonsingular

\[
A[M \ \tilde{M}] = [AM \ A\tilde{M}] = [M \ \tilde{M}] \begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix}
\]

where

\[
\begin{bmatrix} Y \\ Z \end{bmatrix} = [M \ \tilde{M}]^{-1}A\tilde{M}
\]

with \( T = [M \ \tilde{M}] \), we have

\[
T^{-1}AT = \begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix}
\]
in other words: if $\mathcal{V}$ is $A$-invariant we can change coordinates so that

- $A$ becomes block upper triangular in the new coordinates
- $\mathcal{V}$ corresponds to $\left\{ \begin{bmatrix} \tilde{z} \\ 0 \end{bmatrix} \mid \tilde{z} \in \mathbb{R}^k \right\}$ in the new coordinates
Revealing the controllable subspace

consider \( \dot{x} = Ax + Bu \) (or \( x(t+1) = Ax(t) + Bu(t) \)) and assume it is not controllable, so \( \mathcal{V} = \mathcal{R}(C) \neq \mathbb{R}^n \)

let columns of \( M \in \mathbb{R}^k \) be basis for controllable subspace (e.g., choose \( k \) independent columns from \( C \))

let \( \tilde{M} \in \mathbb{R}^{n \times (n-k)} \) be such that \( T = [M \ \tilde{M}] \) is nonsingular

then

\[
T^{-1}AT = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} \\ 0 & \tilde{A}_{22} \end{bmatrix}, \quad T^{-1}B = \begin{bmatrix} \tilde{B}_1 \\ 0 \end{bmatrix} \\
\tilde{C} = T^{-1}C = \begin{bmatrix} \tilde{B}_1 & \cdots & \tilde{A}_{11}^{-1} \tilde{B}_1 \\ 0 & \cdots & 0 \end{bmatrix}
\]

in the new coordinates the controllable subspace is \( \{(z, 0) \mid z \in \mathbb{R}^k\} \); \((\tilde{A}_{11}, \tilde{B}_1)\) is controllable
we have changed coordinates to reveal the controllable subspace:

\[
\begin{align*}
\tilde{B}_1 & \\
1/s & \quad \tilde{x}_1 \\
\tilde{A}_{11} & \\
\tilde{A}_{12} & \quad \tilde{x}_2 \\
1/s & \\
\tilde{A}_{22}
\end{align*}
\]

roughly speaking, \( \tilde{x}_1 \) is the controllable part of the state
Revealing the unobservable subspace

Similarly, if \((C, A)\) is not observable, we can change coordinates to obtain

\[
T^{-1}AT = \begin{bmatrix}
\tilde{A}_{11} & 0 \\
\tilde{A}_{21} & \tilde{A}_{22}
\end{bmatrix}, \quad CT = \begin{bmatrix}
\tilde{C}_1 & 0
\end{bmatrix}
\]

and \((\tilde{C}_1, \tilde{A}_{11})\) is observable.
Popov-Belevitch-Hautus controllability test

PBH controllability criterion: \((A, B)\) is controllable if and only if

\[
\text{Rank} \left[ sI - A \ B \right] = n \quad \text{for all} \quad s \in \mathbb{C}
\]

equivalent to:

\((A, B)\) is uncontrollable if and only if there is a \(w \neq 0\) with

\[
\begin{align*}
    w^T A &= \lambda w^T, \\
    w^T B &= 0
\end{align*}
\]

\(i.e.,\) a left eigenvector is orthogonal to columns of \(B\)
to show it, first assume that \( w \neq 0, \ w^T A = \lambda w^T, \ w^T B = 0 \)
then for \( k = 1, \ldots, n - 1, \ w^T A^k B = \lambda^k w^T B = 0, \) so
\[
\begin{align*}
w^T[B \ AB \cdots A^{n-1}B] &= w^T C = 0
\end{align*}
\]
which shows \((A, B)\) not controllable

conversely, suppose \((A, B)\) not controllable
change coordinates as on p.5–15, let \( z \) be any left eigenvector of \( \tilde{A}_{22}, \) and
define \( \tilde{w} = (0, z) \)
then \( \tilde{w}^T \tilde{A} = \lambda \tilde{w}^T, \ \tilde{w}^T \tilde{B} = 0 \)

it follows that \( w^T A = \lambda w^T, \ w^T B = 0, \) where \( w = T^{-T} \tilde{w} \)
PBH observability test

PBH observability criterion: \((C, A)\) is observable if and only if

\[
\text{Rank} \begin{bmatrix} sI - A & C \end{bmatrix} = n \text{ for all } s \in \mathbb{C}
\]

is equivalent to:

\((C, A)\) is unobservable if and only if there is a \(v \neq 0\) with

\[
Av = \lambda v, \quad Cv = 0
\]

i.e., a (right) eigenvector is in the nullspace of \(C\)
Observability and controllability of modes

the PBH tests allow us to identify unobservable and uncontrollable modes

the mode associated with right and left eigenvectors \( v, w \) is

- uncontrollable if \( w^T B = 0 \)
- unobservable if \( Cv = 0 \)

(classification can be done with repeated eigenvalues, Jordan blocks, but gets tricky)
Controllability and linear state feedback

we consider system $\dot{x} = Ax + Bu$ (or $x(t + 1) = Ax(t) + Bu(t)$)

we refer to $u = Kx + w$ as a linear state feedback (with auxiliary input $w$), with associated closed-loop system $\dot{x} = (A + BK)x + Bw$
suppose $w^T A = \lambda w^T$, $w \neq 0$, $w^T B = 0$, i.e., $w$ corresponds to uncontrollable mode of open loop system

then $w^T (A + BK) = w^T A + w^T BK = \lambda w^T$, i.e., $w$ is also a left eigenvector of closed-loop system, associated with eigenvalue $\lambda$

i.e., eigenvalues (and indeed, left eigenvectors) associated with uncontrollable modes cannot be changed by linear state feedback

conversely, if $w$ is left eigenvector associated with uncontrollable closed-loop mode, then $w$ is left eigenvector associated with uncontrollable open-loop mode

in other words: state feedback preserves uncontrollable eigenvalues and the associated left eigenvectors
Invariant subspaces and quadratic matrix equations

Suppose \( \mathcal{V} = \mathcal{R}(M) \) is \( A \)-invariant, where \( M \in \mathbb{R}^{n \times k} \) is rank \( k \), so 
\[ AM = MX \text{ for some } X \in \mathbb{R}^{k \times k} \]

conformably partition as

\[
\begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}
\begin{bmatrix}
M_1 \\
M_2
\end{bmatrix}
= 
\begin{bmatrix}
M_1 \\
M_2
\end{bmatrix}X
\]

\[ A_{11}M_1 + A_{12}M_2 = M_1X, \quad A_{21}M_1 + A_{22}M_2 = M_2X \]

Eliminate \( X \) from first equation (assuming \( M_1 \) is nonsingular):

\[ X = M_1^{-1}A_{11}M_1 + M_1^{-1}A_{12}M_2 \]

Substituting this into second equation yields

\[ A_{21}M_1 + A_{22}M_2 = M_2M_1^{-1}A_{11}M_1 + M_2M_1^{-1}A_{12}M_2 \]
multiply on right by $M_1^{-1}$:

$$A_{21} + A_{22}M_2M_1^{-1} = M_2M_1^{-1}A_{11} + M_2M_1^{-1}A_{12}M_2M_1^{-1}$$

with $P = M_2M_1^{-1}$, we have

$$-A_{22}P + PA_{11} - A_{21} + PA_{12}P = 0,$$

a general quadratic matrix equation

if we take $A$ to be Hamiltonian associated with a cts-time LQR problem, we recover the method of solving ARE via stable eigenvectors of Hamiltonian...
Lecture 6
Estimation

- Gaussian random vectors
- minimum mean-square estimation (MMSE)
- MMSE with linear measurements
- relation to least-squares, pseudo-inverse
Gaussian random vectors

random vector $x \in \mathbb{R}^n$ is *Gaussian* if it has density

$$p_x(v) = (2\pi)^{-n/2}(\det \Sigma)^{-1/2} \exp \left( -\frac{1}{2}(v - \bar{x})^T \Sigma^{-1} (v - \bar{x}) \right),$$

for some $\Sigma = \Sigma^T > 0$, $\bar{x} \in \mathbb{R}^n$

- denoted $x \sim \mathcal{N}(\bar{x}, \Sigma)$
- $\bar{x} \in \mathbb{R}^n$ is the *mean* or *expected* value of $x$, *i.e.*,

$$\bar{x} = \mathbb{E} x = \int v p_x(v) dv$$

- $\Sigma = \Sigma^T > 0$ is the *covariance* matrix of $x$, *i.e.*,

$$\Sigma = \mathbb{E}(x - \bar{x})(x - \bar{x})^T$$
\[
\begin{align*}
  &= \mathbf{E} xx^T - \bar{x}x^T \\
  &= \int (v - \bar{x})(v - \bar{x})^T p_x(v) dv
\end{align*}
\]

density for \( x \sim \mathcal{N}(0, 1) \):
• mean and variance of scalar random variable $x_i$ are

\[ E x_i = \bar{x}_i, \quad E(x_i - \bar{x}_i)^2 = \Sigma_{ii} \]

hence standard deviation of $x_i$ is $\sqrt{\Sigma_{ii}}$

• covariance between $x_i$ and $x_j$ is $E(x_i - \bar{x}_i)(x_j - \bar{x}_j) = \Sigma_{ij}$

• correlation coefficient between $x_i$ and $x_j$ is $\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$

• mean (norm) square deviation of $x$ from $\bar{x}$ is

\[ E \|x - \bar{x}\|^2 = E \text{Tr}(x - \bar{x})(x - \bar{x})^T = \text{Tr} \Sigma = \sum_{i=1}^{n} \Sigma_{ii} \]

(using $\text{Tr} AB = \text{Tr} BA$)

**example:** $x \sim \mathcal{N}(0, I)$ means $x_i$ are independent identically distributed (IID) $\mathcal{N}(0, 1)$ random variables
Confidence ellipsoids

- \( p_x(v) \) is constant for \( (v - \bar{x})^T \Sigma^{-1} (v - \bar{x}) = \alpha \), i.e., on the surface of ellipsoid
  \[
  \mathcal{E}_\alpha = \{ v \mid (v - \bar{x})^T \Sigma^{-1} (v - \bar{x}) \leq \alpha \}
  \]
  - thus \( \bar{x} \) and \( \Sigma \) determine shape of density

- \( \eta \)-confidence set for random variable \( z \) is smallest volume set \( S \) with
  \[
  \Pr(z \in S) \geq \eta
  \]
  - in general case confidence set has form \( \{ v \ p_z(v) \geq \beta \} \)

- \( \mathcal{E}_\alpha \) are the \( \eta \)-confidence sets for Gaussian, called confidence ellipsoids
  - \( \alpha \) determines confidence level \( \eta \)
Confidence levels

the nonnegative random variable $(x - \bar{x})^T \Sigma^{-1} (x - \bar{x})$ has a $\chi^2_n$ distribution, so $\text{Prob}(x \in \mathcal{E}_\alpha) = F_{\chi^2_n}(\alpha)$ where $F_{\chi^2_n}$ is the CDF

some good approximations:

- $\mathcal{E}_n$ gives about 50% probability
- $\mathcal{E}_{n+2\sqrt{n}}$ gives about 90% probability
geometrically:

- mean $\bar{x}$ gives center of ellipsoid

- semiaxes are $\sqrt{\alpha \lambda_i} u_i$, where $u_i$ are (orthonormal) eigenvectors of $\Sigma$ with eigenvalues $\lambda_i$
example: $x \sim \mathcal{N}(\bar{x}, \Sigma)$ with $\bar{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$

- $x_1$ has mean 2, std. dev. $\sqrt{2}$
- $x_2$ has mean 1, std. dev. 1
- correlation coefficient between $x_1$ and $x_2$ is $\rho = 1/\sqrt{2}$
- $E \| x - \bar{x} \|^2 = 3$

90% confidence ellipsoid corresponds to $\alpha = 4.6$:

(there, 91 out of 100 fall in $\mathcal{E}_{4.6}$)
Affine transformation

suppose \( x \sim \mathcal{N}(\bar{x}, \Sigma_x) \)

consider affine transformation of \( x \):

\[ z = Ax + b, \]

where \( A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \)

then \( z \) is Gaussian, with mean

\[ \mathbf{E} z = \mathbf{E}(Ax + b) = A \mathbf{E} x + b = A\bar{x} + b \]

and covariance

\[ \Sigma_z = \mathbf{E}(z - \bar{z})(z - \bar{z})^T = \mathbf{E} A(x - \bar{x})(x - \bar{x})^T A^T = A\Sigma_x A^T \]
examples:

- if $w \sim \mathcal{N}(0, I)$ then $x = \Sigma^{1/2} w + \bar{x}$ is $\mathcal{N}(\bar{x}, \Sigma)$
  useful for simulating vectors with given mean and covariance

- conversely, if $x \sim \mathcal{N}(\bar{x}, \Sigma)$ then $z = \Sigma^{-1/2}(x - \bar{x})$ is $\mathcal{N}(0, I)$
  (normalizes & decorrelates; called whitening or normalizing)
suppose \( x \sim \mathcal{N}(\bar{x}, \Sigma) \) and \( c \in \mathbb{R}^n \)

scalar \( c^T x \) has mean \( c^T \bar{x} \) and variance \( c^T \Sigma c \)

thus (unit length) direction of minimum variability for \( x \) is \( u \), where

\[
\Sigma u = \lambda_{\text{min}} u, \quad \|u\| = 1
\]

standard deviation of \( u^T x \) is \( \sqrt{\lambda_{\text{min}}} \)

(similarly for maximum variability)
Degenerate Gaussian vectors

• it is convenient to allow $\Sigma$ to be singular (but still $\Sigma = \Sigma^T \geq 0$)
  
  – in this case density formula obviously does not hold
  – meaning: in some directions $x$ is not random at all
  – random variable $x$ is called a degenerate Gaussian

• write $\Sigma$ as

$$
\Sigma = \left[ \begin{array}{cc}
Q_+ & Q_0 \\
\end{array} \right] \left[ \begin{array}{cc}
\Sigma_+ & 0 \\
0 & 0 \\
\end{array} \right] \left[ \begin{array}{cc}
Q_+ & Q_0 \\
\end{array} \right]^T
$$

where $Q = [Q_+ \ Q_0]$ is orthogonal, $\Sigma_+ > 0$

  – columns of $Q_0$ are orthonormal basis for $\mathcal{N}(\Sigma)$
  – columns of $Q_+$ are orthonormal basis for $\text{range}(\Sigma)$
• then

\[ Q^T x = \begin{bmatrix} z \\ w \end{bmatrix}, \quad x = Q_+ z + Q_0 w \]

− \( z \sim \mathcal{N}(Q_+^T \bar{x}, \Sigma_+) \) is (nondegenerate) Gaussian (hence, density formula holds)

− \( w = Q_0^T \bar{x} \in \mathbb{R}^n \) is not random, called \textit{deterministic component} of \( x \)
Linear measurements

linear measurements with noise:

\[ y = Ax + v \]

- \( x \in \mathbb{R}^n \) is what we want to measure or estimate
- \( y \in \mathbb{R}^m \) is measurement
- \( A \in \mathbb{R}^{m \times n} \) characterizes sensors or measurements
- \( v \) is sensor noise
common assumptions:

- \( x \sim \mathcal{N}(\bar{x}, \Sigma_x) \)
- \( v \sim \mathcal{N}(\bar{v}, \Sigma_v) \)
- \( x \) and \( v \) are independent

- \( \mathcal{N}(\bar{x}, \Sigma_x) \) is the prior distribution of \( x \) (describes initial uncertainty about \( x \))
- \( \bar{v} \) is noise bias or offset (and is usually 0)
- \( \Sigma_v \) is noise covariance
thus

\[
\begin{bmatrix}
  x \\
  v
\end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix}
  \bar{x} \\
  \bar{v}
\end{bmatrix}, \begin{bmatrix}
  \Sigma_x & 0 \\
  0 & \Sigma_v
\end{bmatrix} \right)
\]

using

\[
\begin{bmatrix}
  x \\
  y
\end{bmatrix} = \begin{bmatrix}
  I & 0 \\
  A & I
\end{bmatrix} \begin{bmatrix}
  x \\
  v
\end{bmatrix}
\]

we can write

\[
\mathbf{E} \begin{bmatrix}
  x \\
  y
\end{bmatrix} = \begin{bmatrix}
  \bar{x} \\
  A\bar{x} + \bar{v}
\end{bmatrix}
\]

and

\[
\mathbf{E} \begin{bmatrix}
  x - \bar{x} \\
  y - \bar{y}
\end{bmatrix}^T \begin{bmatrix}
  x - \bar{x} \\
  y - \bar{y}
\end{bmatrix} = \begin{bmatrix}
  I & 0 \\
  A & I
\end{bmatrix} \begin{bmatrix}
  \Sigma_x & 0 \\
  0 & \Sigma_v
\end{bmatrix} \begin{bmatrix}
  I & 0 \\
  A & I
\end{bmatrix}^T
\]

\[
= \begin{bmatrix}
  \Sigma_x & \Sigma_x A^T \\
  A\Sigma_x & A\Sigma_x A^T + \Sigma_v
\end{bmatrix}
\]
covariance of measurement $y$ is $A\Sigma_x A^T + \Sigma_v$

- $A\Sigma_x A^T$ is ‘signal covariance’
- $\Sigma_v$ is ‘noise covariance’
Minimum mean-square estimation

suppose \( x \in \mathbb{R}^n \) and \( y \in \mathbb{R}^m \) are random vectors (not necessarily Gaussian)

we seek to estimate \( x \) given \( y \)

thus we seek a function \( \phi : \mathbb{R}^m \rightarrow \mathbb{R}^n \) such that \( \hat{x} = \phi(y) \) is near \( x \)

one common measure of nearness: mean-square error,

\[
\mathbb{E} \| \phi(y) - x \|^2
\]

minimum mean-square estimator (MMSE) \( \phi_{\text{mmse}} \) minimizes this quantity

general solution: \( \phi_{\text{mmse}}(y) = \mathbb{E}(x|y) \), i.e., the conditional expectation of \( x \) given \( y \)
MMSE for Gaussian vectors

now suppose $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are jointly Gaussian:

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_y \end{bmatrix} \right)$$

(after a lot of algebra) the conditional density is

$$p_{x|y}(v|y) = (2\pi)^{-n/2}(\det \Lambda)^{-1/2} \exp\left( -\frac{1}{2}(v - w)^T \Lambda^{-1}(v - w) \right),$$

where

$$\Lambda = \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{xy}^T, \quad w = \bar{x} + \Sigma_{xy} \Sigma_y^{-1} (y - \bar{y})$$

hence MMSE estimator (i.e., conditional expectation) is

$$\hat{x} = \phi_{\text{mmse}}(y) = \mathbb{E}(x|y) = \bar{x} + \Sigma_{xy} \Sigma_y^{-1} (y - \bar{y})$$
$\phi_{\text{mmse}}$ is an affine function

MMSE estimation error, $\hat{x} - x$, is a Gaussian random vector

$$\hat{x} - x \sim \mathcal{N}(0, \Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{xy}^T)$$

note that

$$\Sigma_x - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{xy}^T \leq \Sigma_x$$

i.e., covariance of estimation error is always less than prior covariance of $x$
Best linear unbiased estimator

estimator

\[ \hat{x} = \phi_{\text{blu}}(y) = \bar{x} + \Sigma_{xy}\Sigma_y^{-1}(y - \bar{y}) \]

makes sense when \( x, y \) aren’t jointly Gaussian

this estimator

- is *unbiased*, i.e., \( \mathbb{E}\hat{x} = \mathbb{E}\bar{x} \)
- often works well
- is widely used
- has minimum mean square error among all *affine* estimators

sometimes called *best linear unbiased* estimator
MMSE with linear measurements

consider specific case

\[ y = Ax + v, \quad x \sim \mathcal{N}(\bar{x}, \Sigma_x), \quad v \sim \mathcal{N}(\bar{v}, \Sigma_v), \]

\( x, v \) independent

MMSE of \( x \) given \( y \) is affine function

\[ \hat{x} = \bar{x} + B(y - \bar{y}) \]

where \( B = \Sigma_x A^T(A\Sigma_x A^T + \Sigma_v)^{-1} \), \( \bar{y} = A\bar{x} + \bar{v} \)

interpretation:

- \( \bar{x} \) is our best prior guess of \( x \) (before measurement)

- \( y - \bar{y} \) is the discrepancy between what we actually measure (\( y \)) and the expected value of what we measure (\( \bar{y} \))
• estimator modifies prior guess by $B$ times this discrepancy

• estimator blends prior information with measurement

• $B$ gives gain from observed discrepancy to estimate

• $B$ is small if noise term $\Sigma_v$ in ‘denominator’ is large
MMSE error with linear measurements

MMSE estimation error, $\tilde{x} = \hat{x} - x$, is Gaussian with zero mean and covariance

$$\Sigma_{\text{est}} = \Sigma_x - \Sigma_x A^T (A \Sigma_x A^T + \Sigma_v)^{-1} A \Sigma_x$$

- $\Sigma_{\text{est}} \leq \Sigma_x$, i.e., measurement always decreases uncertainty about $x$
- difference $\Sigma_x - \Sigma_{\text{est}}$ (or some other comparison) gives value of measurement $y$ in estimating $x$
  - $(\Sigma_{\text{est}} ii / \Sigma_x ii)^{1/2}$ gives fractional decrease in uncertainty of $x_i$ due to measurement
  - $(\text{Tr} \Sigma_{\text{est}} / \text{Tr} \Sigma)^{1/2}$ gives fractional decrease in uncertainty in $x$, measured by mean-square error
Estimation error covariance

- error covariance $\Sigma_{\text{est}}$ can be determined \textit{before} measurement $y$ is made!

- to evaluate $\Sigma_{\text{est}}$, only need to know
  - $A$ (which characterizes sensors)
  - prior covariance of $x$ (\textit{i.e.,} $\Sigma_x$)
  - noise covariance (\textit{i.e.,} $\Sigma_v$)

- you \textit{do not} need to know the measurement $y$ (or the means $\bar{x}$, $\bar{v}$)

- useful for \textit{experiment design} or \textit{sensor selection}
Information matrix formulas

we can write estimator gain matrix as

\[ B = \Sigma_x A^T (A \Sigma_x A^T + \Sigma_v)^{-1} \]
\[ = (A^T \Sigma_v^{-1} A + \Sigma_x^{-1})^{-1} A^T \Sigma_v^{-1} \]

- \( n \times n \) inverse instead of \( m \times m \)

- \( \Sigma_x^{-1}, \Sigma_v^{-1} \) sometimes called information matrices

corresponding formula for estimator error covariance:

\[ \Sigma_{est} = \Sigma_x - \Sigma_x A^T (A \Sigma_x A^T + \Sigma_v)^{-1} A \Sigma_x \]
\[ = (A^T \Sigma_v^{-1} A + \Sigma_x^{-1})^{-1} \]
can interpret $\Sigma_{\text{est}}^{-1} = \Sigma_{x}^{-1} + A^T \Sigma_{v}^{-1} A$ as:

- posterior information matrix ($\Sigma_{\text{est}}^{-1}$)
  - = prior information matrix ($\Sigma_{x}^{-1}$)
  + information added by measurement ($A^T \Sigma_{v}^{-1} A$)
proof: multiply

\[ \Sigma_x A^T (A \Sigma_x A^T + \Sigma_v)^{-1} \overset{?}{=} (A^T \Sigma_v^{-1} A + \Sigma_x^{-1})^{-1} A^T \Sigma_v^{-1} \]

on left by \((A^T \Sigma_v^{-1} A + \Sigma_x^{-1})\) and on right by \((A \Sigma_x A^T + \Sigma_v)\) to get

\[ (A^T \Sigma_v^{-1} A + \Sigma_x^{-1}) \Sigma_x A^T \overset{?}{=} A^T \Sigma_v^{-1} (A \Sigma_x A^T + \Sigma_v) \]

which is true
Relation to regularized least-squares

suppose $\bar{x} = 0$, $\bar{v} = 0$, $\Sigma_x = \alpha^2 I$, $\Sigma_v = \beta^2 I$

estimator is $\hat{x} = By$ where

$$B = (A^T \Sigma_v^{-1} A + \Sigma_x^{-1})^{-1} A^T \Sigma_v^{-1}$$
$$= (A^T A + (\beta/\alpha)^2 I)^{-1} A^T$$

. . . which corresponds to regularized least-squares

MMSE estimate $\hat{x}$ minimizes

$$\|Az - y\|^2 + (\beta/\alpha)^2 \|z\|^2$$

over $z$
Example

navigation using range measurements to distant beacons

\[ y = Ax + v \]

- \( x \in \mathbb{R}^2 \) is location
- \( y_i \) is range measurement to \( i \)th beacon
- \( v_i \) is range measurement error, IID \( \mathcal{N}(0, 1) \)
- \( i \)th row of \( A \) is unit vector in direction of \( i \)th beacon

prior distribution:

\[
x \sim \mathcal{N}(\bar{x}, \Sigma_x), \quad \bar{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \Sigma_x = \begin{bmatrix} 2^2 & 0 \\ 0 & 0.5^2 \end{bmatrix}
\]

\( x_1 \) has std. dev. 2; \( x_2 \) has std. dev. 0.5
90% confidence ellipsoid for prior distribution
\[ \{ x \mid (x - \bar{x})^T \Sigma_x^{-1} (x - \bar{x}) \leq 4.6 \} : \]
**Case 1:** one measurement, with beacon at angle $30^\circ$

fewer measurements than variables, so combining prior information with measurement is critical

resulting estimation error covariance:

$$\Sigma_{\text{est}} = \begin{bmatrix} 1.046 & -0.107 \\ -0.107 & 0.246 \end{bmatrix}$$
90% confidence ellipsoid for estimate $\hat{x}$: (and 90% confidence ellipsoid for $x$)

**interpretation**: measurement

- yields essentially no reduction in uncertainty in $x_2$
- reduces uncertainty in $x_1$ by a factor about two
Case 2: 4 measurements, with beacon angles $80^\circ, 85^\circ, 90^\circ, 95^\circ$

resulting estimation error covariance:

$$\Sigma_{est} = \begin{bmatrix} 3.429 & -0.074 \\ -0.074 & 0.127 \end{bmatrix}$$
90% confidence ellipsoid for estimate $\hat{x}$: (and 90% confidence ellipsoid for $x$)

**interpretation:** measurement yields

- little reduction in uncertainty in $x_1$
- small reduction in uncertainty in $x_2$
Lecture 7
The Kalman filter

- Linear system driven by stochastic process
- Statistical steady-state
- Linear Gauss-Markov model
- Kalman filter
- Steady-state Kalman filter
Linear system driven by stochastic process

we consider linear dynamical system $x(t + 1) = Ax(t) + Bu(t)$, with $x(0)$ and $u(0)$, $u(1), \ldots$ random variables

we’ll use notation

$$\bar{x}(t) = \mathbf{E} x(t), \quad \Sigma_x(t) = \mathbf{E} (x(t) - \bar{x}(t))(x(t) - \bar{x}(t))^T$$

and similarly for $\bar{u}(t), \Sigma_u(t)$

taking expectation of $x(t + 1) = Ax(t) + Bu(t)$ we have

$$\bar{x}(t + 1) = A\bar{x}(t) + B\bar{u}(t)$$

i.e., the means propagate by the same linear dynamical system
now let’s consider the covariance

\[ x(t + 1) - \bar{x}(t + 1) = A(x(t) - \bar{x}(t)) + B(u(t) - \bar{u}(t)) \]

and so

\[ \Sigma_x(t + 1) = \mathbf{E}(A(x(t) - \bar{x}(t)) + B(u(t) - \bar{u}(t))) \cdot (A(x(t) - \bar{x}(t)) + B(u(t) - \bar{u}(t)))^T \]

\[ = A\Sigma_x(t)A^T + B\Sigma_u(t)B^T + A\Sigma_{ux}(t)B^T + B\Sigma_{ux}(t)A^T \]

where

\[ \Sigma_{xu}(t) = \Sigma_{ux}(t)^T = \mathbf{E}(x(t) - \bar{x}(t))(u(t) - \bar{u}(t))^T \]

thus, the covariance \( \Sigma_x(t) \) satisfies another, Lyapunov-like linear dynamical system, driven by \( \Sigma_{xu} \) and \( \Sigma_u \)
consider special case $\Sigma_{xu}(t) = 0$, i.e., $x$ and $u$ are uncorrelated, so we have Lyapunov iteration

$$\Sigma_x(t + 1) = A\Sigma_x(t)A^T + B\Sigma_u(t)B^T,$$

which is stable if and only if $A$ is stable.

If $A$ is stable and $\Sigma_u(t)$ is constant, $\Sigma_x(t)$ converges to $\Sigma_x$, called the *steady-state covariance*, which satisfies Lyapunov equation

$$\Sigma_x = A\Sigma_xA^T + B\Sigma_uB^T$$

thus, we can calculate the steady-state covariance of $x$ exactly, by solving a Lyapunov equation

(useful for starting simulations in statistical steady-state)
Example

we consider \( x(t + 1) = Ax(t) + w(t) \), with

\[
A = \begin{bmatrix}
0.6 & -0.8 \\
0.7 & 0.6
\end{bmatrix},
\]

where \( w(t) \) are IID \( \mathcal{N}(0, I) \)
eigenvalues of \( A \) are \( 0.6 \pm 0.75j \), with magnitude 0.96, so \( A \) is stable
we solve Lyapunov equation to find steady-state covariance

\[
\Sigma_x = \begin{bmatrix}
13.35 & -0.03 \\
-0.03 & 11.75
\end{bmatrix}
\]

covariance of \( x(t) \) converges to \( \Sigma_x \) no matter its initial value
two initial state distributions: $\Sigma_x(0) = 0$, $\Sigma_x(0) = 10^2 I$

plot shows $\Sigma_{11}(t)$ for the two cases
$x_1(t)$ for one realization from each case:
Linear Gauss-Markov model

we consider linear dynamical system

\[ x(t + 1) = Ax(t) + w(t), \quad y(t) = Cx(t) + v(t) \]

- \( x(t) \in \mathbb{R}^n \) is the state; \( y(t) \in \mathbb{R}^p \) is the observed output
- \( w(t) \in \mathbb{R}^n \) is called process noise or state noise
- \( v(t) \in \mathbb{R}^p \) is called measurement noise
**Statistical assumptions**

- $x(0), w(0), w(1), \ldots$, and $v(0), v(1), \ldots$ are jointly Gaussian and independent
- $w(t)$ are IID with $\mathbf{E} w(t) = 0$, $\mathbf{E} w(t)w(t)^T = W$
- $v(t)$ are IID with $\mathbf{E} v(t) = 0$, $\mathbf{E} v(t)v(t)^T = V$
- $\mathbf{E} x(0) = \bar{x}_0$, $\mathbf{E} (x(0) - \bar{x}_0)(x(0) - \bar{x}_0)^T = \Sigma_0$

(it’s not hard to extend to case where $w(t), v(t)$ are not zero mean)

we’ll denote $X(t) = (x(0), \ldots, x(t))$, etc.

since $X(t)$ and $Y(t)$ are linear functions of $x(0)$, $W(t)$, and $V(t)$, we conclude they are all jointly Gaussian (i.e., the process $x$, $w$, $v$, $y$ is Gaussian)
Statistical properties

- sensor noise $v$ independent of $x$
- $w(t)$ is independent of $x(0), \ldots, x(t)$ and $y(0), \ldots, y(t)$
- **Markov property**: the process $x$ is Markov, i.e.,

  $$x(t)|x(0), \ldots, x(t-1) = x(t)|x(t-1)$$

  roughly speaking: if you know $x(t-1)$, then knowledge of $x(t-2), \ldots, x(0)$ doesn't give any more information about $x(t)$
Mean and covariance of Gauss-Markov process

mean satisfies $\bar{x}(t + 1) = A\bar{x}(t)$, $\bar{x}(0) = \bar{x}_0$, so $\bar{x}(t) = A^t\bar{x}_0$

covariance satisfies

$$\Sigma_x(t + 1) = A\Sigma_x(t)A^T + W$$

if $A$ is stable, $\Sigma_x(t)$ converges to steady-state covariance $\Sigma_x$, which satisfies Lyapunov equation

$$\Sigma_x = A\Sigma_x A^T + W$$
Conditioning on observed output

we use the notation

\[ \hat{x}(t|s) = \mathbb{E}(x(t)|y(0), \ldots, y(s)), \]
\[ \Sigma_{t|s} = \mathbb{E}(x(t) - \hat{x}(t|s))(x(t) - \hat{x}(t|s))^T \]

• the random variable \( x(t)|y(0), \ldots, y(s) \) is Gaussian, with mean \( \hat{x}(t|s) \) and covariance \( \Sigma_{t|s} \)

• \( \hat{x}(t|s) \) is the minimum mean-square error estimate of \( x(t) \), based on \( y(0), \ldots, y(s) \)

• \( \Sigma_{t|s} \) is the covariance of the error of the estimate \( \hat{x}(t|s) \)
State estimation

we focus on two state estimation problems:

• finding $\hat{x}(t|t)$, i.e., estimating the current state, based on the current and past observed outputs
• finding $\hat{x}(t + 1|t)$, i.e., predicting the next state, based on the current and past observed outputs

since $x(t)$, $Y(t)$ are jointly Gaussian, we can use the standard formula to find $\hat{x}(t|t)$ (and similarly for $\hat{x}(t + 1|t)$)

$$\hat{x}(t|t) = \bar{x}(t) + \Sigma_{x(t)Y(t)} \Sigma_{Y(t)}^{-1} (Y(t) - \bar{Y}(t))$$

the inverse in the formula, $\Sigma_{Y(t)}^{-1}$, is size $pt \times pt$, which grows with $t$

the Kalman filter is a clever method for computing $\hat{x}(t|t)$ and $\hat{x}(t + 1|t)$ recursively
Measurement update

let’s find $\hat{x}(t|t)$ and $\Sigma_{t|t}$ in terms of $\hat{x}(t|t-1)$ and $\Sigma_{t|t-1}$

start with $y(t) = Cx(t) + v(t)$, and condition on $Y(t-1)$:

$$y(t)|Y(t-1) = Cx(t)|Y(t-1) + v(t)|Y(t-1) = Cx(t)|Y(t-1) + v(t)$$

since $v(t)$ and $Y(t-1)$ are independent

so $x(t)|Y(t-1)$ and $y(t)|Y(t-1)$ are jointly Gaussian with mean and covariance

$$\begin{bmatrix} \hat{x}(t|t-1) \\ C\hat{x}(t|t-1) \end{bmatrix}, \quad \begin{bmatrix} \Sigma_{t|t-1} & \Sigma_{t|t-1}C^T \\ C\Sigma_{t|t-1} & C\Sigma_{t|t-1}C^T + V \end{bmatrix}$$
now use standard formula to get mean and covariance of

\[(x(t)|Y(t - 1))(y(t)|Y(t - 1)),\]

which is exactly the same as \(x(t)|Y(t)\):

\[
\hat{x}(t|t) = \hat{x}(t|t - 1) + \Sigma_{t|t-1} C^T (C \Sigma_{t|t-1} C^T + V)^{-1} (y(t) - C \hat{x}(t|t - 1))
\]

\[
\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} C^T (C \Sigma_{t|t-1} C^T + V)^{-1} C \Sigma_{t|t-1}
\]

this gives us \(\hat{x}(t|t)\) and \(\Sigma_{t|t}\) in terms of \(\hat{x}(t|t - 1)\) and \(\Sigma_{t|t-1}\)

this is called the measurement update since it gives our updated estimate of \(x(t)\) based on the measurement \(y(t)\) becoming available
Time update

now let's increment time, using \( x(t + 1) = Ax(t) + w(t) \)

condition on \( Y(t) \) to get

\[
x(t + 1)|Y(t) = Ax(t)|Y(t) + w(t)|Y(t)
\]

\[
= Ax(t)|Y(t) + w(t)
\]

since \( w(t) \) is independent of \( Y(t) \)

therefore we have \( \hat{x}(t + 1|t) = A\hat{x}(t|t) \) and

\[
\Sigma_{t+1|t} = \mathbf{E}(\hat{x}(t + 1|t) - x(t + 1))(\hat{x}(t + 1|t) - x(t + 1))^T
\]

\[
= \mathbf{E}(A\hat{x}(t|t) - Ax(t) - w(t))(A\hat{x}(t|t) - Ax(t) - w(t))^T
\]

\[
= A\Sigma_{t|t}A^T + W
\]
Kalman filter

measurement and time updates together give a recursive solution
start with prior mean and covariance, \( \hat{x}(0| - 1) = \bar{x}_0, \Sigma(0| - 1) = \Sigma_0 \)
apply the measurement update

\[
\hat{x}(t|t) = \hat{x}(t|t-1) + \Sigma_{t|t-1} C^T (C \Sigma_{t|t-1} C^T + V)^{-1} (y(t) - C \hat{x}(t|t-1)) \\
\Sigma_t = \Sigma_{t|t-1} - \Sigma_{t|t-1} C^T (C \Sigma_{t|t-1} C^T + V)^{-1} C \Sigma_{t|t-1}
\]

to get \( \hat{x}(0|0) \) and \( \Sigma_{0|0} \); then apply time update

\[
\hat{x}(t + 1|t) = A \hat{x}(t|t), \quad \Sigma_{t+1|t} = A \Sigma_{t|t} A^T + W
\]

to get \( \hat{x}(1|0) \) and \( \Sigma_{1|0} \)

now, repeat measurement and time updates . . .
Riccati recursion

to lighten notation, we’ll use $\hat{x}(t) = \hat{x}(t|t-1)$ and $\hat{\Sigma}_t = \Sigma_{t|t-1}$

we can express measurement and time updates for $\hat{\Sigma}$ as

$$\hat{\Sigma}_{t+1} = A\hat{\Sigma}_t A^T + W - A\hat{\Sigma}_t C^T (C\hat{\Sigma}_t C^T + V)^{-1} C\hat{\Sigma}_t A^T$$

which is a Riccati recursion, with initial condition $\hat{\Sigma}_0 = \Sigma_0$

• $\hat{\Sigma}_t$ can be computed \textit{before any observations are made}

• thus, we can calculate the estimation error covariance \textit{before} we get any observed data
Comparison with LQR

in LQR,

- Riccati recursion for $P(t)$ (which determines the minimum cost to go from a point at time $t$) runs \textit{backward} in time
- we can compute cost-to-go before knowing $x(t)$

in Kalman filter,

- Riccati recursion for $\hat{\Sigma}_t$ (which is the state prediction error covariance at time $t$) runs \textit{forward} in time
- we can compute $\hat{\Sigma}_t$ before we actually get any observations
we can express KF as

\[
\hat{x}(t + 1) = A\hat{x}(t) + A\hat{\Sigma}_t C^T (C\hat{\Sigma}_t C^T + V)^{-1} (y(t) - C\hat{x}(t))
\]

\[
= A\hat{x}(t) + L_t(y(t) - \hat{y}(t))
\]

where \( L_t = A\hat{\Sigma}_t C^T (C\hat{\Sigma}_t C^T + V)^{-1} \) is the observer gain, and \( \hat{y}(t) \) is \( \hat{y}(t|t - 1) \)

- \( \hat{y}(t) \) is our output prediction, i.e., our estimate of \( y(t) \) based on \( y(0), \ldots, y(t - 1) \)
- \( e(t) = y(t) - \hat{y}(t) \) is our output prediction error
- \( A\hat{x}(t) \) is our prediction of \( x(t + 1) \) based on \( y(0), \ldots, y(t - 1) \)
- our estimate of \( x(t + 1) \) is the prediction based on \( y(0), \ldots, y(t - 1) \), plus a linear function of the output prediction error

Observer form
The Kalman filter block diagram
Steady-state Kalman filter

as in LQR, Riccati recursion for $\hat{\Sigma}_t$ converges to steady-state value $\hat{\Sigma}$, provided $(C, A)$ is observable and $(A, W)$ is controllable

$\hat{\Sigma}$ gives steady-state error covariance for estimating $x(t + 1)$ given $y(0), \ldots, y(t)$

note that state prediction error covariance converges, even if system is unstable

$\hat{\Sigma}$ satisfies ARE

$$\hat{\Sigma} = A\hat{\Sigma}A^T + W - A\hat{\Sigma}C^T (C\hat{\Sigma}C^T + V)^{-1} C\hat{\Sigma}A^T$$

(which can be solved directly)
steady-state filter is a time-invariant observer:

\[
\hat{x}(t+1) = A\hat{x}(t) + L(y(t) - \hat{y}(t)), \quad \hat{y}(t) = C\hat{x}(t)
\]

where \( L = A\hat{\Sigma}C^{T}(C\hat{\Sigma}C^{T} + V)^{-1} \)

define state estimation error \( \tilde{x}(t) = x(t) - \hat{x}(t) \), so

\[
y(t) - \hat{y}(t) = Cx(t) + v(t) - C\hat{x}(t) = C\tilde{x}(t) + v(t)
\]

and

\[
\tilde{x}(t+1) = x(t+1) - \hat{x}(t+1)
\]

\[
= Ax(t) + w(t) - A\hat{x}(t) - L(C\tilde{x}(t) + v(t))
\]

\[
= (A - LC')\tilde{x}(t) + w(t) - Lv(t)
\]
thus, the estimation error propagates according to a linear system, with closed-loop dynamics $A - LC$, driven by the process $w(t) - LCv(t)$, which is IID zero mean and covariance $W + LVLT$

provided $A, W$ is controllable and $C, A$ is observable, $A - LC$ is stable
Example

system is

\[ x(t + 1) = Ax(t) + w(t), \quad y(t) = Cx(t) + v(t) \]

with \( x(t) \in \mathbb{R}^6, y(t) \in \mathbb{R} \)

we’ll take \( E x(0) = 0, E x(0)x(0)^T = \Sigma_0 = 5^2 I; W = (1.5)^2 I, V = 1 \)

eigenvalues of \( A \):

\[ 0.9973 \pm 0.0730j, \quad 0.9995 \pm 0.0324j, \quad 0.9941 \pm 0.1081j \]

(which have magnitude one)

goal: predict \( y(t + 1) \) based on \( y(0), \ldots, y(t) \)
first let’s find variance of $y(t)$ versus $t$, using Lyapunov recursion

$$E y(t)^2 = C \Sigma_x(t) C^T + V, \quad \Sigma_x(t+1) = A \Sigma_x(t) A^T + W, \quad \Sigma_x(0) = \Sigma_0$$
now, let’s plot the prediction error variance versus $t$,

$$E e(t)^2 = E(\hat{y}(t) - y(t))^2 = C\hat{\Sigma}_t C^T + V,$$

where $\hat{\Sigma}_t$ satisfies Riccati recursion

$$\hat{\Sigma}_{t+1} = A\hat{\Sigma}_t A^T + W - A\hat{\Sigma}_t C^T (C\hat{\Sigma}_t C^T + V)^{-1} C\hat{\Sigma}_t A^T, \quad \hat{\Sigma}_0 = \Sigma_0$$

prediction error variance converges to steady-state value 18.7
now let’s try the Kalman filter on a realization $y(t)$

top plot shows $y(t)$; bottom plot shows $e(t)$ (on different vertical scale)
Lecture 8
The Extended Kalman filter

- Nonlinear filtering
- Extended Kalman filter
- Linearization and random variables
Nonlinear filtering

• nonlinear Markov model:

\[ x(t + 1) = f(x(t), w(t)), \quad y(t) = g(x(t), v(t)) \]

- \( f \) is (possibly nonlinear) dynamics function
- \( g \) is (possibly nonlinear) measurement or output function
- \( w(0), w(1), \ldots, v(0), v(1), \ldots \) are independent
- even if \( w, v \) Gaussian, \( x \) and \( y \) need not be

• nonlinear filtering problem: find, e.g.,

\[ \hat{x}(t|t-1) = \mathbf{E}(x(t)|y(0), \ldots, y(t-1)), \quad \hat{x}(t|t) = \mathbf{E}(x(t)|y(0), \ldots, y(t)) \]

• general nonlinear filtering solution involves a PDE, and is not practical
Extended Kalman filter

- extended Kalman filter (EKF) is *heuristic* for nonlinear filtering problem
- often works well (when tuned properly), but sometimes not
- widely used in practice
- based on
  - linearizing dynamics and output functions at current estimate
  - propagating an approximation of the conditional expectation and covariance
Linearization and random variables

• consider $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$

• suppose $\mathbb{E} x = \bar{x}$, $\mathbb{E}(x - \bar{x})(x - \bar{x})^T = \Sigma_x$, and $y = \phi(x)$

• if $\Sigma_x$ is small, $\phi$ is not too nonlinear,

$$y \approx \tilde{y} = \phi(\bar{x}) + D\phi(\bar{x})(x - \bar{x})$$

$$\tilde{y} \sim \mathcal{N}(\phi(\bar{x}), D\phi(\bar{x})\Sigma_x D\phi(\bar{x})^T)$$

• gives approximation for mean and covariance of nonlinear function of random variable:

$$\bar{y} \approx \phi(\bar{x}), \quad \Sigma_y \approx D\phi(\bar{x})\Sigma_x D\phi(\bar{x})^T$$

• if $\Sigma_x$ is not small compared to ‘curvature’ of $\phi$, these estimates are poor
• a good estimate can be found by Monte Carlo simulation:

\[
\bar{y} \approx \bar{y}^{mc} = \frac{1}{N} \sum_{i=1}^{N} \phi(x^{(i)})
\]

\[
\Sigma_y \approx \frac{1}{N} \sum_{i=1}^{N} \left( \phi(x^{(i)}) - \bar{y}^{mc} \right) \left( \phi(x^{(i)}) - \bar{y}^{mc} \right)^T
\]

where \( x^{(1)}, \ldots, x^{(N)} \) are samples from the distribution of \( x \), and \( N \) is large

• another method: use Monte Carlo formulas, with a small number of nonrandom samples chosen as ‘typical’, e.g., the 90% confidence ellipsoid semi-axis endpoints

\[
x^{(i)} = \bar{x} \pm \beta v_i, \quad \Sigma_x = V \Lambda V^T
\]
Example

\( x \sim \mathcal{N}(0, 1), \ y = \exp(x) \)

(for this case we can compute mean and variance of \( y \) exactly)

<table>
<thead>
<tr>
<th></th>
<th>( \bar{y} )</th>
<th>( \sigma_y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>exact values</td>
<td>( e^{1/2} = 1.649 )</td>
<td>( \sqrt{e^2 - e} = 2.161 )</td>
</tr>
<tr>
<td>linearization</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Monte Carlo (( N = 10 ))</td>
<td>1.385</td>
<td>1.068</td>
</tr>
<tr>
<td>Monte Carlo (( N = 100 ))</td>
<td>1.430</td>
<td>1.776</td>
</tr>
<tr>
<td>Sigma points (( x = \bar{x}, \ \bar{x} \pm 1.5\sigma_x ))</td>
<td>1.902</td>
<td>2.268</td>
</tr>
</tbody>
</table>
Extended Kalman filter

• **initialization:** \( \hat{x}(0| - 1) = \bar{x}_0, \Sigma(0| - 1) = \Sigma_0 \)

• **measurement update**
  - linearize output function at \( x = \hat{x}(t|t - 1) \):

\[
C = \frac{\partial g}{\partial x}(\hat{x}(t|t - 1), 0) \\
V = \frac{\partial g}{\partial v}(\hat{x}(t|t - 1), 0) \Sigma_v \frac{\partial g}{\partial v}(\hat{x}(t|t - 1), 0)^T
\]

  - measurement update based on linearization

\[
\hat{x}(t|t) = \hat{x}(t|t - 1) + \Sigma_{t|t-1} C^T (C \Sigma_{t|t-1} C^T + V)^{-1} \ldots \\
\ldots (y(t) - g(\hat{x}(t|t - 1), 0)) \\
\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} C^T (C \Sigma_{t|t-1} C^T + V)^{-1} C \Sigma_{t|t-1}
\]
- **time update**

  - linearize dynamics function at $x = \hat{x}(t|t)$:

    $$A = \frac{\partial f}{\partial x}(\hat{x}(t|t), 0)$$

    $$W = \frac{\partial f}{\partial w}(\hat{x}(t|t), 0) \Sigma_w \frac{\partial f}{\partial w}(\hat{x}(t|t), 0)^T$$

  - time update based on linearization

    $$\hat{x}(t + 1|t) = f(\hat{x}(t|t), 0), \quad \Sigma_{t+1|t} = A\Sigma_{t|t}A^T + W$$

  - replacing linearization with Monte Carlo yields *particle filter*

  - replacing linearization with sigma-point estimates yields *unscented Kalman filter (UKF)*
**Example**

- $p(t), u(t) \in \mathbb{R}^2$ are position and velocity of vehicle, with $(p(0), u(0)) \sim \mathcal{N}(0, I)$

- Vehicle dynamics:

$$p(t+1) = p(t) + 0.1u(t), \quad u(t+1) = \begin{bmatrix} 0.85 & 0.15 \\ -0.1 & 0.85 \end{bmatrix} u(t) + w(t)$$

$w(t)$ are IID $\mathcal{N}(0, I)$

- Measurements: noisy measurements of distance to 9 points $p_i \in \mathbb{R}^2$

$$y_i(t) = \|p(t) - p_i\| + v_i(t), \quad i = 1, \ldots, 9,$$

$v_i(t)$ are IID $\mathcal{N}(0, 0.3^2)$
EKF results

- EKF initialized with $\hat{x}(0| - 1) = 0$, $\Sigma(0| - 1) = I$, where $x = (p, u)$
- $p_i$ shown as stars; $p(t)$ as dotted curve; $\hat{p}(t|t)$ as solid curve
Current position estimation error

\[ \| \hat{p}(t|t) - p(t) \| \text{ versus } t \]
Current position estimation predicted error

\[(\Sigma(t|t)_{11} + \Sigma(t|t)_{22})^{1/2} \text{ versus } t\]
Lecture 9
Invariant sets, conservation, and dissipation

• invariant sets
• conserved quantities
• dissipated quantities
• derivative along trajectory
• discrete-time case
Invariant sets

we consider autonomous, time-invariant nonlinear system $\dot{x} = f(x)$

a set $C \subseteq \mathbb{R}^n$ is invariant (w.r.t. system, or $f$) if for every trajectory $x$,

$$x(t) \in C \implies x(\tau) \in C \text{ for all } \tau \geq t$$

- if trajectory enters $C$, or starts in $C$, it stays in $C$
- trajectories can cross into boundary of $C$, but never out of $C$
Examples of invariant sets

general examples:

• \{x_0\}, where \(f(x_0) = 0\) (i.e., \(x_0\) is an equilibrium point)

• any trajectory or union of trajectories, e.g.,
\[\{x(t) \mid x(0) \in D, \ t \geq 0, \ \dot{x} = f(x)\}\]

more specific examples:

• \(\dot{x} = Ax, \ C = \text{span}\{v_1, \ldots, v_k\}\), where \(Av_i = \lambda_i v_i\)

• \(\dot{x} = Ax, \ C = \{z \mid 0 \leq w^T z \leq a\}\), where \(w^T A = \lambda w^T, \ \lambda \leq 0\)
Invariance of nonnegative orthant

when is nonnegative orthant $\mathbb{R}^n_+$ invariant for $\dot{x} = Ax$? (i.e., when do nonnegative trajectories always stay nonnegative?)

**answer:** if and only if $A_{ij} \geq 0$ for $i \neq j$

first assume $A_{ij} \geq 0$ for $i \neq j$, and $x(0) \in \mathbb{R}^n_+$; we’ll show that $x(t) \in \mathbb{R}^n_+$ for $t \geq 0$

$$x(t) = e^{tA}x(0) = \lim_{k \to \infty} (I + (t/k)A)^k x(0)$$

for $k$ large enough the matrix $I + (t/k)A$ has all nonnegative entries, so $(I + (t/k)A)^k x(0)$ has all nonnegative entries

hence the limit above, which is $x(t)$, has nonnegative entries
now let’s assume that $A_{ij} < 0$ for some $i \neq j$; we’ll find trajectory with $x(0) \in \mathbb{R}^n_+$ but $x(t) \not\in \mathbb{R}^n_+$ for some $t > 0$

let’s take $x(0) = e_j$, so for small $h > 0$, we have $x(h) \approx e_j + hAe_j$

in particular, $x(h)_i \approx hA_{ij} < 0$ for small positive $h$, i.e., $x(h) \not\in \mathbb{R}^n_+$

this shows that if $A_{ij} < 0$ for some $i \neq j$, $\mathbb{R}^n_+$ isn’t invariant
Conserved quantities

scalar valued function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ is called integral of the motion, a conserved quantity, or invariant for $\dot{x} = f(x)$ if for every trajectory $x$, $\phi(x(t))$ is constant

classical examples:

- total energy of a lossless mechanical system
- total angular momentum about an axis of an isolated system
- total fluid in a closed system

level set or level surface of $\phi$, $\{ z \in \mathbb{R}^n \mid \phi(z) = a \}$, are invariant sets

e.g., trajectories of lossless mechanical system stay in surfaces of constant energy
Example: nonlinear lossless mechanical system

\[ m\ddot{q} = -F = -\phi(q), \text{ where } m > 0 \text{ is mass, } q(t) \text{ is displacement, } F \text{ is restoring force, } \phi \text{ is nonlinear spring characteristic with } \phi(0) = 0 \]

with \( x = (q, \dot{q}) \), we have

\[
\dot{x} = \begin{bmatrix} \dot{q} \\ \ddot{q} \end{bmatrix} = \begin{bmatrix} x_2 \\ -(1/m)\phi(x_1) \end{bmatrix}
\]
potential energy stored in spring is

\[ \psi(q) = \int_0^q \phi(u) \, du \]

total energy is kinetic plus potential: \( E(x) = (m/2)\dot{q}^2 + \psi(q) \)

\( E \) is a conserved quantity: if \( x \) is a trajectory, then

\[
\frac{d}{dt} E(x(t)) = \frac{d}{dt} \left( \frac{m}{2} \dot{q}^2 + \psi(q) \right) \\
= m\ddot{q} + \phi(q)\dot{q} \\
= m\ddot{q} - (1/m)\phi(q) + \phi(q)\dot{q} \\
= 0
\]

\( i.e., E(x(t)) \) is constant
Derivative of function along trajectory

we have function $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ and $\dot{x} = f(x)$

if $x$ is trajectory of system, then

$$\frac{d}{dt} \phi(x(t)) = D\phi(x(t)) \frac{dx}{dt} = \nabla \phi(x(t))^T f(x)$$

we define $\dot{\phi} : \mathbb{R}^n \rightarrow \mathbb{R}$ as

$$\dot{\phi}(z) = \nabla \phi(z)^T f(z)$$

interpretation: $\dot{\phi}(z)$ gives $\frac{d}{dt} \phi(x(t))$, if $x(t) = z$

e.g., if $\dot{\phi}(z) > 0$, then $\phi(x(t))$ is increasing when $x(t)$ passes through $z$
if \( \phi \) is conserved, then \( \phi(x(t)) \) is constant along any trajectory, so

\[
\dot{\phi}(z) = \nabla \phi(z)^T f(x) = 0
\]

for all \( z \)

this means the vector field \( f(z) \) is everywhere orthogonal to \( \nabla \phi \), which is normal to the level surface
we say that $\phi : \mathbb{R}^n \to \mathbb{R}$ is a *dissipated quantity* for system $\dot{x} = f(x)$ if for all trajectories, $\phi(x(t))$ is (weakly) decreasing, i.e., $\phi(x(\tau)) \leq \phi(x(t))$ for all $\tau \geq t$

classical examples:

- total energy of a mechanical system with damping
- total fluid in a system that leaks

condition: $\dot{\phi}(z) \leq 0$ for all $z$, i.e., $\nabla \phi(z)^T f(z) \leq 0$

$-\dot{\phi}$ is sometimes called the *dissipation function*

if $\phi$ is dissipated quantity, *sublevel sets* $\{z \mid \phi(z) \leq a\}$ are invariant
Geometric interpretation

$\phi = \text{const.}$

- Vector field points \textit{into} sublevel sets
- $\nabla \phi(z)^T f(z) \leq 0$, \textit{i.e.}, $\nabla \phi$ and $f$ always make an obtuse angle
- Trajectories can only “slip down” to lower values of $\phi$
Example

linear mechanical system with damping: \( M\ddot{q} + D\dot{q} + Kq = 0 \)

- \( q(t) \in \mathbb{R}^n \) is displacement or configuration
- \( M = M^T > 0 \) is mass or inertia matrix
- \( K = K^T > 0 \) is stiffness matrix
- \( D = D^T \geq 0 \) is damping or loss matrix

we’ll use state \( x = (q, \dot{q}) \), so

\[
\dot{x} = \begin{bmatrix}
\dot{q} \\
\ddot{q}
\end{bmatrix} = \begin{bmatrix}
0 & I \\
-M^{-1}K & -M^{-1}D
\end{bmatrix} x
\]
consider total (potential plus kinetic) energy

$$E = \frac{1}{2} q^T K q + \frac{1}{2} \dot{q}^T M \dot{q} = \frac{1}{2} x^T \begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix} x$$

we have

$$\dot{E}(z) = \nabla E(z)^T f(z)$$

$$= z^T \begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix} \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}D \end{bmatrix} z$$

$$= z^T \begin{bmatrix} 0 & K \\ -K & -D \end{bmatrix} z$$

$$= -\dot{q}^T D \dot{q} \leq 0$$

makes sense: $$\frac{d}{dt} (\text{total stored energy}) = - (\text{power dissipated})$$
Trajectory limit with dissipated quantity

suppose \( \phi : \mathbb{R}^n \to \mathbb{R} \) is dissipated quantity for \( \dot{x} = f(x) \)

- \( \phi(x(t)) \to \phi^* \) as \( t \to \infty \), where \( \phi^* \in \mathbb{R} \cup \{-\infty\} \)

- if trajectory \( x \) is bounded and \( \dot{\phi} \) is continuous, \( x(t) \) converges to the zero-dissipation set:

\[
x(t) \to D_0 = \{ z \mid \dot{\phi}(z) = 0 \}
\]

i.e., \( \text{dist} (x(t), D_0) \to 0 \), as \( t \to \infty \) (more on this later)
Linear functions and linear dynamical systems

we consider linear system $\dot{x} = Ax$

when is a linear function $\phi(z) = c^T z$ conserved or dissipated?

$$\dot{\phi} = \nabla \phi(z)^T f(z) = c^T A z$$

$$\dot{\phi}(z) \leq 0 \text{ for all } z \iff \dot{\phi}(z) = 0 \text{ for all } z \iff A^T c = 0$$

i.e., $\phi$ is dissipated if only if it is conserved, if and only if if $A^T c = 0$

($c$ is left eigenvector of $A$ with eigenvalue 0)
Quadratic functions and linear dynamical systems

we consider linear system $\dot{x} = Ax$

when is a quadratic form $\phi(z) = z^T P z$ conserved or dissipated?

$$\dot{\phi}(z) = \nabla \phi(z)^T f(z) = 2 z^T P A z = z^T (A^T P + P A) z$$

i.e., $\dot{\phi}$ is also a quadratic form

- $\phi$ is conserved if and only if $A^T P + P A = 0$
  (which means $A$ and $-A$ share at least $\text{Rank}(P)$ eigenvalues)
- $\phi$ is dissipated if and only if $A^T P + P A \leq 0$
A criterion for invariance

suppose $\phi : \mathbb{R}^n \to \mathbb{R}$ satisfies $\phi(z) = 0 \implies \dot{\phi}(z) < 0$

then the set $C = \{z \mid \phi(z) \leq 0\}$ is invariant

idea: all trajectories on boundary of $C$ cut into $C$, so none can leave

to show this, suppose trajectory $x$ satisfies $x(t) \in C$, $x(s) \notin C$, $t \leq s$

consider (differentiable) function $g : \mathbb{R} \to \mathbb{R}$ given by $g(\tau) = \phi(x(\tau))$

g satisfies $g(t) \leq 0$, $g(s) > 0$

any such function must have at least one point $T \in [t, s]$ where $g(T) = 0$, $g'(T) \geq 0$ (for example, we can take $T = \min\{\tau \geq t \mid g(\tau) = 0\}$)

this means $\phi(x(T)) = 0$ and $\dot{\phi}(x(T)) \geq 0$, a contradiction
Discrete-time systems

we consider nonlinear time-invariant discrete-time system or recursion
\[ x(t + 1) = f(x(t)) \]

we say \( C \subseteq \mathbb{R}^n \) is invariant (with respect to the system) if for every trajectory \( x \),
\[ x(t) \in C \implies x(\tau) \in C \text{ for all } \tau \geq t \]
i.e., trajectories can enter, but cannot leave set \( C \)
equivalent to: \( z \in C \implies f(z) \in C \)

**example:** when is nonnegative orthant \( \mathbb{R}_+^n \) invariant for \( x(t + 1) = Ax(t) \)?
answer: \( \Leftrightarrow A_{ij} \geq 0 \text{ for } i, j = 1, \ldots, n \)
Conserved and dissipated quantities

\( \phi : \mathbb{R}^n \rightarrow \mathbb{R} \) is conserved under \( x(t + 1) = f(x(t)) \) if \( \phi(x(t)) \) is constant, i.e., \( \phi(f(z)) = \phi(z) \) for all \( z \)

\( \phi \) is a dissipated quantity if \( \phi(x(t)) \) is (weakly) decreasing, i.e., \( \phi(f(z)) \leq \phi(z) \) for all \( z \)

we define \( \Delta \phi : \mathbb{R}^n \rightarrow \mathbb{R} \) by \( \Delta \phi(z) = \phi(f(z)) - \phi(z) \)

\( \Delta \phi(z) \) gives change in \( \phi \), over one step, starting at \( z \)

\( \phi \) is conserved if and only if \( \Delta \phi(z) = 0 \) for all \( z \)

\( \phi \) is dissipated if and only if \( \Delta \phi(z) \leq 0 \) for all \( z \)
Quadratic functions and linear dynamical systems

we consider linear system \( x(t + 1) = Ax(t) \)

when is a quadratic form \( \phi(z) = z^T P z \) conserved or dissipated?

\[
\Delta \phi(z) = (Az)^T P(Az) - z^T P z = z^T (A^T P A - P) z
\]

i.e., \( \Delta \phi \) is also a quadratic form

• \( \phi \) is conserved if and only if \( A^T P A - P = 0 \)
  (which means \( A \) and \( A^{-1} \) share at least \( \text{Rank}(P) \) eigenvalues, if \( A \) invertible)

• \( \phi \) is dissipated if and only if \( A^T P A - P \leq 0 \)
Lecture 10
Basic Lyapunov theory

• stability
• positive definite functions
• global Lyapunov stability theorems
• Lasalle’s theorem
• converse Lyapunov theorems
• finding Lyapunov functions
Some stability definitions

we consider nonlinear time-invariant system $\dot{x} = f(x)$, where $f : \mathbb{R}^n \to \mathbb{R}^n$

a point $x_e \in \mathbb{R}^n$ is an equilibrium point of the system if $f(x_e) = 0$

$x_e$ is an equilibrium point $\iff x(t) = x_e$ is a trajectory

suppose $x_e$ is an equilibrium point

- system is globally asymptotically stable (G.A.S.) if for every trajectory $x(t)$, we have $x(t) \to x_e$ as $t \to \infty$
  (implies $x_e$ is the unique equilibrium point)

- system is locally asymptotically stable (L.A.S.) near or at $x_e$ if there is an $R > 0$ s.t. $\|x(0) - x_e\| \leq R \implies x(t) \to x_e$ as $t \to \infty$
• often we change coordinates so that \( x_e = 0 \) (i.e., we use \( \tilde{x} = x - x_e \))

• a linear system \( \dot{x} = Ax \) is G.A.S. (with \( x_e = 0 \)) \( \Leftrightarrow \Re \lambda_i(A) < 0, \ i = 1, \ldots, n \)

• a linear system \( \dot{x} = Ax \) is L.A.S. (near \( x_e = 0 \)) \( \Leftrightarrow \Re \lambda_i(A) < 0, \ i = 1, \ldots, n \)

  (so for linear systems, L.A.S. \( \Leftrightarrow \) G.A.S.)

• there are many other variants on stability (e.g., stability, uniform stability, exponential stability, . . . )

• when \( f \) is nonlinear, establishing any kind of stability is usually very difficult
Energy and dissipation functions

consider nonlinear system \( \dot{x} = f(x) \), and function \( V : \mathbb{R}^n \rightarrow \mathbb{R} \)

we define \( \dot{V} : \mathbb{R}^n \rightarrow \mathbb{R} \) as \( \dot{V}(z) = \nabla V(z)^T f(z) \)

\( \dot{V}(z) \) gives \( \frac{d}{dt} V(x(t)) \) when \( z = x(t) \), \( \dot{x} = f(x) \)

we can think of \( V \) as generalized energy function, and \( -\dot{V} \) as the associated generalized dissipation function
Positive definite functions

A function $V : \mathbb{R}^n \rightarrow \mathbb{R}$ is positive definite (PD) if

- $V(z) \geq 0$ for all $z$
- $V(z) = 0$ if and only if $z = 0$
- all sublevel sets of $V$ are bounded

The last condition is equivalent to $V(z) \to \infty$ as $z \to \infty$

Example: $V(z) = z^T P z$, with $P = P^T$, is PD if and only if $P > 0$
Lyapunov theory

Lyapunov theory is used to make conclusions about trajectories of a system \( \dot{x} = f(x) \) (e.g., G.A.S.) \textit{without finding the trajectories} (i.e., solving the differential equation)

A typical Lyapunov theorem has the form:

- \textbf{if} there exists a function \( V : \mathbb{R}^n \rightarrow \mathbb{R} \) that satisfies some conditions on \( V \) and \( \dot{V} \)
- \textbf{then}, trajectories of system satisfy some property

If such a function \( V \) exists we call it a \textit{Lyapunov function} (that proves the property holds for the trajectories)

Lyapunov function \( V \) can be thought of as \textit{generalized energy function} for system
A Lyapunov boundedness theorem

suppose there is a function $V$ that satisfies

- all sublevel sets of $V$ are bounded
- $\dot{V}(z) \leq 0$ for all $z$

then, all trajectories are bounded, i.e., for each trajectory $x$ there is an $R$ such that $\|x(t)\| \leq R$ for all $t \geq 0$

in this case, $V$ is called a Lyapunov function (for the system) that proves the trajectories are bounded
to prove it, we note that for any trajectory \( x \)

\[
V(x(t)) = V(x(0)) + \int_0^t \dot{V}(x(\tau)) \, d\tau \leq V(x(0))
\]

so the whole trajectory lies in \( \{ z \mid V(z) \leq V(x(0)) \} \), which is bounded

also shows: every sublevel set \( \{ z \mid V(z) \leq a \} \) is invariant
A Lyapunov global asymptotic stability theorem

Suppose there is a function $V$ such that

- $V$ is positive definite
- $\dot{V}(z) < 0$ for all $z \neq 0$, $\dot{V}(0) = 0$

then, every trajectory of $\dot{x} = f(x)$ converges to zero as $t \to \infty$ (i.e., the system is globally asymptotically stable)

interpretation:

- $V$ is positive definite generalized energy function
- energy is always dissipated, except at 0
Proof

Suppose trajectory $x(t)$ does not converge to zero.

$V(x(t))$ is decreasing and nonnegative, so it converges to, say, $\epsilon$ as $t \to \infty$.

Since $x(t)$ doesn’t converge to 0, we must have $\epsilon > 0$, so for all $t$, $\epsilon \leq V(x(t)) \leq V(x(0))$.

$C = \{z \mid \epsilon \leq V(z) \leq V(x(0))\}$ is closed and bounded, hence compact. So $\dot{V}$ (assumed continuous) attains its supremum on $C$, i.e., $\sup_{z \in C} \dot{V} = -a < 0$. Since $\dot{V}(x(t)) \leq -a$ for all $t$, we have

$$V(x(T)) = V(x(0)) + \int_0^T \dot{V}(z) \, dz \leq V(x(0)) - aT$$

which for $T > V(x(0))/a$ implies $V(x(0)) < 0$, a contradiction.

So every trajectory $x(t)$ converges to 0, i.e., $\dot{x} = f(x)$ is G.A.S.
A Lyapunov exponential stability theorem

suppose there is a function $V$ and constant $\alpha > 0$ such that

- $V$ is positive definite
- $\dot{V}(z) \leq -\alpha V(z)$ for all $z$

then, there is an $M$ such that every trajectory of $\dot{x} = f(x)$ satisfies

$$\|x(t)\| \leq M e^{-\alpha t/2} \|x(0)\|$$

(this is called global exponential stability (G.E.S.))

idea: $\dot{V} \leq -\alpha V$ gives guaranteed minimum dissipation rate, proportional to energy
Example

consider system

\[ \dot{x}_1 = -x_1 + g(x_2), \quad \dot{x}_2 = -x_2 + h(x_1) \]

where \(|g(u)| \leq |u|/2, \ |h(u)| \leq |u|/2\)

two first order systems with nonlinear cross-coupling
let’s use Lyapunov theorem to show it’s globally asymptotically stable

we use $V = \frac{1}{2}(x_1^2 + x_2^2)$

required properties of $V$ are clear ($V \geq 0$, etc.)

let’s bound $\dot{V}$:

$$
\dot{V} = x_1 \dot{x}_1 + x_2 \dot{x}_2
= -x_1^2 - x_2^2 + x_1 g(x_2) + x_2 h(x_1)
\leq -x_1^2 - x_2^2 + |x_1 x_2|
\leq -(1/2)(x_1^2 + x_2^2)
= -V
$$

where we use $|x_1 x_2| \leq (1/2)(x_1^2 + x_2^2)$ (derived from $(|x_1| - |x_2|)^2 \geq 0$)

we conclude system is G.A.S. (in fact, G.E.S.)

*without knowing the trajectories*
Lasalle’s theorem

Lasalle’s theorem (1960) allows us to conclude G.A.S. of a system with only $\dot{V} \leq 0$, along with an observability type condition

we consider $\dot{x} = f(x)$

suppose there is a function $V : \mathbb{R}^n \to \mathbb{R}$ such that

- $V$ is positive definite
- $\dot{V}(z) \leq 0$
- the only solution of $\dot{w} = f(w), \dot{V}(w) = 0$ is $w(t) = 0$ for all $t$

then, the system $\dot{x} = f(x)$ is G.A.S.
• last condition means no nonzero trajectory can hide in the “zero dissipation” set

• unlike most other Lyapunov theorems, which extend to time-varying systems, Lasalle’s theorem *requires* time-invariance
A Lyapunov instability theorem

suppose there is a function $V : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

- $\dot{V}(z) \leq 0$ for all $z$ (or just whenever $V(z) \leq 0$)
- there is $w$ such that $V(w) < V(0)$

then, the trajectory of $\dot{x} = f(x)$ with $x(0) = w$ does not converge to zero (and therefore, the system is not G.A.S.)

to show it, we note that $V(x(t)) \leq V(x(0)) = V(w) < V(0)$ for all $t \geq 0$

but if $x(t) \rightarrow 0$, then $V(x(t)) \rightarrow V(0)$; so we cannot have $x(t) \rightarrow 0$
A Lyapunov divergence theorem

suppose there is a function \( V : \mathbb{R}^n \rightarrow \mathbb{R} \) such that

- \( \dot{V}(z) < 0 \) whenever \( V(z) < 0 \)
- there is \( w \) such that \( V(w) < 0 \)

then, the trajectory of \( \dot{x} = f(x) \) with \( x(0) = w \) is unbounded, i.e.,

\[
\sup_{t \geq 0} \|x(t)\| = \infty
\]

(this is not quite the same as \( \lim_{t \to \infty} \|x(t)\| = \infty \))
Proof of Lyapunov divergence theorem

let $\dot{x} = f(x)$, $x(0) = w$. let’s first show that $V(x(t)) \leq V(w)$ for all $t \geq 0$.

if not, let $T$ denote the smallest positive time for which $V(x(T)) = V(w)$. then over $[0, T]$, we have $V(x(t)) \leq V(w) < 0$, so $\dot{V}(x(t)) < 0$, and so

$$\int_0^T \dot{V}(x(t)) \, dt < 0$$

the lefthand side is also equal to

$$\int_0^T \dot{V}(x(t)) \, dt = V(x(T)) - V(x(0)) = 0$$

so we have a contradiction.

it follows that $V(x(t)) \leq V(x(0))$ for all $t$, and therefore $\dot{V}(x(t)) < 0$ for all $t$.

now suppose that $\|x(t)\| \leq R$, i.e., the trajectory is bounded.

$\{z \mid V(z) \leq V(x(0)), \|z\| \leq R\}$ is compact, so there is a $\beta > 0$ such that $\dot{V}(z) \leq -\beta$ whenever $V(z) \leq V(x(0))$ and $\|z\| \leq R$. 

Basic Lyapunov theory
we conclude $V(x(t)) \leq V(x(0)) - \beta t$ for all $t \geq 0$, so $V(x(t)) \to -\infty$, a contradiction.
Converse Lyapunov theorems

a typical *converse Lyapunov theorem* has the form

- **if** the trajectories of system satisfy some property
- **then** there exists a Lyapunov function that proves it

a sharper converse Lyapunov theorem is more specific about the form of the Lyapunov function

*example:* if the linear system $\dot{x} = Ax$ is G.A.S., then there is a quadratic Lyapunov function that proves it (we’ll prove this later)
A converse Lyapunov G.E.S. theorem

suppose there is $\beta > 0$ and $M$ such that each trajectory of $\dot{x} = f(x)$ satisfies

$$\|x(t)\| \leq Me^{-\beta t}\|x(0)\| \text{ for all } t \geq 0$$

(called global exponential stability, and is stronger than G.A.S.)

then, there is a Lyapunov function that proves the system is exponentially stable, i.e., there is a function $V : \mathbb{R}^n \to \mathbb{R}$ and constant $\alpha > 0$ s.t.

- $V$ is positive definite
- $\dot{V}(z) \leq -\alpha V(z)$ for all $z$
Proof of converse G.A.S. Lyapunov theorem

suppose the hypotheses hold, and define

\[ V(z) = \int_0^\infty \|x(t)\|^2 \, dt \]

where \( x(0) = z, \dot{x} = f(x) \)

since \( \|x(t)\| \leq Me^{-\beta t}\|z\| \), we have

\[ V(z) = \int_0^\infty \|x(t)\|^2 \, dt \leq \int_0^\infty M^2 e^{-2\beta t}\|z\|^2 \, dt = \frac{M^2}{2\beta} \|z\|^2 \]

(which shows integral is finite)
let's find \( \dot{V}(z) = \left. \frac{d}{dt} \right|_{t=0} V(x(t)) \), where \( x(t) \) is trajectory with \( x(0) = z \)

\[
\dot{V}(z) = \lim_{t \to 0} \left( \frac{1}{t} \left( V(x(t)) - V(x(0)) \right) \right)
= \lim_{t \to 0} \left( \frac{1}{t} \left( \int_t^\infty \|x(\tau)\|^2 d\tau - \int_0^\infty \|x(\tau)\|^2 d\tau \right) \right)
= \lim_{t \to 0} \left( -\frac{1}{t} \int_0^t \|x(\tau)\|^2 d\tau \right)
= -\|z\|^2
\]

now let's verify properties of \( V \)

\( V(z) \geq 0 \) and \( V(z) = 0 \iff z = 0 \) are clear

finally, we have \( \dot{V}(z) = -z^Tz \leq -\alpha V(z) \), with \( \alpha = 2\beta/M^2 \)
Finding Lyapunov functions

- there are many different types of Lyapunov theorems
- the key in all cases is to find a Lyapunov function and verify that it has the required properties
- there are several approaches to finding Lyapunov functions and verifying the properties

one common approach:

- decide form of Lyapunov function (e.g., quadratic), parametrized by some parameters (called a Lyapunov function candidate)
- try to find values of parameters so that the required hypotheses hold
Other sources of Lyapunov functions

- value function of a related optimal control problem
- linear-quadratic Lyapunov theory (next lecture)
- computational methods
- converse Lyapunov theorems
- graphical methods (really!)

(as you might guess, these are all somewhat related)
Lecture 11
Linear quadratic Lyapunov theory

• the Lyapunov equation
• Lyapunov stability conditions
• the Lyapunov operator and integral
• evaluating quadratic integrals
• analysis of ARE
• discrete-time results
• linearization theorem
The Lyapunov equation

the Lyapunov equation is

$$A^T P + PA + Q = 0$$

where $A, P, Q \in \mathbb{R}^{n \times n}$, and $P, Q$ are symmetric

interpretation: for linear system $\dot{x} = Ax$, if $V(z) = z^T P z$, then

$$\dot{V}(z) = (Az)^T P z + z^T P(Az) = - z^T Q z$$

i.e., if $z^T P z$ is the (generalized) energy, then $z^T Q z$ is the associated (generalized) dissipation

linear-quadratic Lyapunov theory: linear dynamics, quadratic Lyapunov function
we consider system $\dot{x} = Ax$, with $\lambda_1, \ldots, \lambda_n$ the eigenvalues of $A$

if $P > 0$, then

- the sublevel sets are ellipsoids (and bounded)
- $V(z) = z^T P z = 0 \iff z = 0$

**boundedness condition:** if $P > 0$, $Q \geq 0$ then

- all trajectories of $\dot{x} = Ax$ are bounded
  (this means $\Re \lambda_i \leq 0$, and if $\Re \lambda_i = 0$, then $\lambda_i$ corresponds to a Jordan block of size one)
- the ellipsoids $\{z \mid z^T P z \leq a\}$ are invariant
Stability condition

if $P > 0$, $Q > 0$ then the system $\dot{x} = Ax$ is (globally asymptotically) stable, i.e., $\Re \lambda_i < 0$

to see this, note that

$$\dot{V}(z) = -z^T Q z \leq -\lambda_{\min}(Q) z^T z \leq -\frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)} z^T P z = -\alpha V(z)$$

where $\alpha = \frac{\lambda_{\min}(Q)}{\lambda_{\max}(P)} > 0$
An extension based on observability

(Lasalle’s theorem for linear dynamics, quadratic function)

if $P > 0$, $Q \geq 0$, and $(Q, A)$ observable, then the system $\dot{x} = Ax$ is (globally asymptotically) stable

to see this, we first note that all eigenvalues satisfy $\Re \lambda_i \leq 0$

now suppose that $v \neq 0$, $Av = \lambda v$, $\Re \lambda = 0$

then $A\bar{v} = \bar{\lambda} \bar{v} = -\lambda \bar{v}$, so

$$\left\| Q^{1/2}v \right\|^2 = v^*Qv = -v^* \left( A^T P + PA \right) v = \lambda v^* P v - \lambda v^* P \bar{v} = 0$$

which implies $Q^{1/2}v = 0$, so $Qv = 0$, contradicting observability (by PBH)

interpretation: observability condition means no trajectory can stay in the “zero dissipation” set $\{z \mid z^T Q z = 0\}$
An instability condition

if $Q \geq 0$ and $P \not\geq 0$, then $A$ is not stable

to see this, note that $\dot{V} \leq 0$, so $V(x(t)) \leq V(x(0))$

since $P \not\geq 0$, there is a $w$ with $V(w) < 0$; trajectory starting at $w$ does not converge to zero

in this case, the sublevel sets $\{ z \mid V(z) \leq 0 \}$ (which are unbounded) are invariant
The Lyapunov operator

the Lyapunov operator is given by

\[ \mathcal{L}(P) = A^T P + PA \]

special case of Sylvester operator

\( \mathcal{L} \) is nonsingular if and only if \( A \) and \( -A \) share no common eigenvalues, i.e., \( A \) does not have pair of eigenvalues which are negatives of each other

• if \( A \) is stable, Lyapunov operator is nonsingular

• if \( A \) has imaginary (nonzero, \( j\omega \)-axis) eigenvalue, then Lyapunov operator is singular

thus if \( A \) is stable, for any \( Q \) there is exactly one solution \( P \) of Lyapunov equation

\[ A^T P + PA + Q = 0 \]
Solving the Lyapunov equation

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

we are given \( A \) and \( Q \) and want to find \( P \)

if Lyapunov equation is solved as a set of \( n^2 \) equations in \( n^2 \) variables, cost is \( O(n^6) \) operations

fast methods, that exploit the special structure of the linear equations, can solve Lyapunov equation with cost \( O(n^3) \)

based on first reducing \( A \) to Schur or upper Hessenberg form
The Lyapunov integral

if $A$ is stable there is an explicit formula for solution of Lyapunov equation:

$$P = \int_0^\infty e^{tA^T} Q e^{tA} \, dt$$

to see this, we note that

$$A^T P + PA = \int_0^\infty \left( A^T e^{tA^T} Q e^{tA} + e^{tA^T} Q e^{tA} A \right) \, dt$$

$$= \int_0^\infty \left( \frac{d}{dt} e^{tA^T} Q e^{tA} \right) \, dt$$

$$= e^{tA^T} Q e^{tA} \bigg|_0^\infty$$

$$= -Q$$
Interpretation as cost-to-go

if $A$ is stable, and $P$ is (unique) solution of $A^T P + PA + Q = 0$, then

$$V(z) = z^T P z$$

$$= z^T \left( \int_0^\infty e^{tA^T} Q e^{tA} \, dt \right) z$$

$$= \int_0^\infty x(t)^T Q x(t) \, dt$$

where $\dot{x} = Ax$, $x(0) = z$

thus $V(z)$ is cost-to-go from point $z$ (with no input) and integral quadratic cost function with matrix $Q$
if $A$ is stable and $Q > 0$, then for each $t$, $e^{tA^T}Qe^{tA} > 0$, so

$$P = \int_{0}^{\infty} e^{tA^T}Qe^{tA} dt > 0$$

meaning: if $A$ is stable,

- we can choose any positive definite quadratic form $z^TQz$ as the dissipation, i.e., $-\dot{V} = z^TQz$
- then solve a set of linear equations to find the (unique) quadratic form $V(z) = z^TPz$
- $V$ will be positive definite, so it is a Lyapunov function that proves $A$ is stable

in particular: a linear system is stable if and only if there is a quadratic Lyapunov function that proves it
**generalization:** if $A$ stable, $Q \geq 0$, and $(Q, A)$ observable, then $P > 0$

to see this, the Lyapunov integral shows $P \geq 0$

if $Pz = 0$, then

$$0 = z^T P z = z^T \left( \int_0^\infty e^{tA}^T Q e^{tA} \, dt \right) z = \int_0^\infty \|Q^{1/2} e^{tA} z\|^2 \, dt$$

so we conclude $Q^{1/2} e^{tA} z = 0$ for all $t \geq 0$

this implies that $Qz = 0$, $QAz = 0$, $QA^{n-1}z = 0$, contradicting $(Q, A)$ observable
Monotonicity of Lyapunov operator inverse

suppose \( A^T P_i + P_i A + Q_i = 0, \ i = 1, \ 2 \)

if \( Q_1 \geq Q_2 \), then for all \( t \), \( e^{tA^T} Q_1 e^{tA} \geq e^{tA^T} Q_2 e^{tA} \)

if \( A \) is stable, we have

\[
P_1 = \int_0^\infty e^{tA^T} Q_1 e^{tA} \ dt \geq \int_0^\infty e^{tA^T} Q_2 e^{tA} \ dt = P_2
\]

in other words: if \( A \) is stable then

\[
Q_1 \geq Q_2 \implies \mathcal{L}^{-1}(Q_1) \geq \mathcal{L}^{-1}(Q_2)
\]

\( i.e., \) inverse Lyapunov operator is monotonic, or preserves matrix inequality, when \( A \) is stable

(question: is \( \mathcal{L} \) monotonic?)
Evaluating quadratic integrals

suppose $\dot{x} = Ax$ is stable, and define

$$J = \int_0^\infty x(t)^T Q x(t) \, dt$$

to find $J$, we solve Lyapunov equation $A^T P + PA + Q = 0$ for $P$

then, $J = x(0)^T P x(0)$

in other words: we can evaluate quadratic integral exactly, by solving a set of linear equations, without even computing a matrix exponential
Controllability and observability Grammians

for $A$ stable, the controllability Grammian of $(A, B)$ is defined as

$$W_c = \int_0^\infty e^{tA} B B^T e^{tA^T} \, dt$$

and the observability Grammian of $(C, A)$ is

$$W_o = \int_0^\infty e^{tA^T} C^T C e^{tA} \, dt$$

these Grammians can be computed by solving the Lyapunov equations

$$AW_c + W_c A^T + B B^T = 0, \quad A^T W_o + W_o A + C^T C = 0$$

we always have $W_c \geq 0$, $W_o \geq 0$;
$W_c > 0$ if and only if $(A, B)$ is controllable, and
$W_o > 0$ if and only if $(C, A)$ is observable
Evaluating a state feedback gain

consider

\[ \dot{x} = Ax + Bu, \quad y = Cx, \quad u = Kx, \quad x(0) = x_0 \]

with closed-loop system \( \dot{x} = (A + BK)x \) stable

to evaluate the quadratic integral performance measures

\[ J_u = \int_0^\infty u(t)^T u(t) \, dt, \quad J_y = \int_0^\infty y(t)^T y(t) \, dt \]

we solve Lyapunov equations

\[
\begin{align*}
(A + BK)^T P_u + P_u (A + BK) + K^T K &= 0 \\
(A + BK)^T P_y + P_y (A + BK) + C^T C &= 0
\end{align*}
\]

then we have \( J_u = x_0^T P_u x_0, \quad J_y = x_0^T P_y x_0 \)
write ARE (with $Q \geq 0$, $R > 0$)
\[ A^T P + PA + Q - PBR^{-1}B^T P = 0 \]
as
\[ (A + BK)^T P + P(A + BK) + (Q + K^T RK) = 0 \]
with $K = -R^{-1}B^T P$
we conclude: if $A + BK$ stable, then $P \geq 0$ (since $Q + K^T RK \geq 0$)
\textit{i.e.}, any stabilizing solution of ARE is PSD
if also $(Q, A)$ is observable, then we conclude $P > 0$
to see this, we need to show that $(Q + K^T RK, A + BK)$ is observable
if not, there is $v \neq 0$ s.t.
\[ (A + BK)v = \lambda v, \quad (Q + K^T RK)v = 0 \]
which implies

\[ v^*(Q + K^T RK)v = v^*Qv + v^*K^T RKv = \|Q^{1/2}v\|^2 + \|R^{1/2}Kv\|^2 = 0 \]

so \( Qv = 0, \ Kv = 0 \)

\[ (A + BK)v = Av = \lambda v, \quad Qv = 0 \]

which contradicts \((Q, A)\) observable

the same argument shows that if \( P > 0 \) and \((Q, A)\) is observable, then \( A + BK \) is stable
Monotonic norm convergence

suppose that $A + A^T < 0$, i.e., (symmetric part of) $A$ is negative definite can express as $A^T P + PA + Q = 0$, with $P = I, Q > 0$

meaning: $x^T P x = \|x\|^2$ decreases along every nonzero trajectory, i.e.,

- $\|x(t)\|$ is always decreasing monotonically to 0
- $x(t)$ is always moving towards origin

this implies $A$ is stable, but the converse is false: for a stable system, we need not have $A + A^T < 0$

(for a stable system with $A + A^T \not< 0$, $\|x(t)\|$ converges to zero, but not monotonically)
for a stable system we can always change coordinates so we have monotonic norm convergence

let $P$ denote the solution of $A^T P + PA + I = 0$

take $T = P^{-1/2}$

in new coordinates $A$ becomes $\tilde{A} = T^{-1} AT$,

\[
\tilde{A} + \tilde{A}^T = P^{1/2} A P^{-1/2} + P^{-1/2} A^T P^{1/2} \\
= P^{-1/2} (PA + A^T P) P^{-1/2} \\
= -P^{-1} < 0
\]

in new coordinates, convergence is obvious because $\|x(t)\|$ is always decreasing
Discrete-time results

all linear quadratic Lyapunov results have discrete-time counterparts

the *discrete-time* Lyapunov equation is

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

*meaning*: if \( x(t + 1) = Ax(t) \) and \( V(z) = z^T P z \), then \( \Delta V(z) = -z^T Q z \)

- if \( P > 0 \) and \( Q > 0 \), then \( A \) is (discrete-time) stable (*i.e.*, \( |\lambda_i| < 1 \))
- if \( P > 0 \) and \( Q \geq 0 \), then all trajectories are bounded (*i.e.*, \( |\lambda_i| \leq 1 \); \( |\lambda_i| = 1 \) only for \( 1 \times 1 \) Jordan blocks)
- if \( P > 0 \), \( Q \geq 0 \), and \((Q, A)\) observable, then \( A \) is stable
- if \( P \neq 0 \) and \( Q \geq 0 \), then \( A \) is not stable
Discrete-time Lyapunov operator

the discrete-time Lyapunov operator is given by $\mathcal{L}(P) = A^T P A - P$

$\mathcal{L}$ is nonsingular if and only if, for all $i, j$, $\lambda_i \lambda_j \neq 1$
(roughly speaking, if and only if $A$ and $A^{-1}$ share no eigenvalues)

if $A$ is stable, then $\mathcal{L}$ is nonsingular; in fact

$$P = \sum_{t=0}^{\infty} (A^T)^t Q A^t$$

is the unique solution of Lyapunov equation $A^T P A - P + Q = 0$

the discrete-time Lyapunov equation can be solved quickly (i.e., $O(n^3)$)
and can be used to evaluate infinite sums of quadratic functions, etc.
Converse theorems

suppose \( x(t + 1) = Ax(t) \) is stable, \( A^T P A - P + Q = 0 \)

- if \( Q > 0 \) then \( P > 0 \)
- if \( Q \geq 0 \) and \((Q, A)\) observable, then \( P > 0 \)

in particular, a discrete-time linear system is stable if and only if there is a quadratic Lyapunov function that proves it
Monotonic norm convergence

suppose $A^T P A - P + Q = 0$, with $P = I$ and $Q > 0$

this means $A^T A < I$, i.e., $\|A\| < 1$

meaning: $\|x(t)\|$ decreases on every nonzero trajectory; indeed,
$\|x(t + 1)\| \leq \|A\| \|x(t)\| < \|x(t)\|$

when $\|A\| < 1$,

- stability is obvious, since $\|x(t)\| \leq \|A\|^t \|x(0)\|$
- system is called contractive since norm is reduced at each step

the converse is false: system can be stable without $\|A\| < 1$
now suppose $A$ is stable, and let $P$ satisfy $A^T PA - P + I = 0$

take $T = P^{-1/2}$

in new coordinates $A$ becomes $\tilde{A} = T^{-1} AT$, so

\[
\tilde{A}^T \tilde{A} = P^{-1/2} A^T P A P^{-1/2}
\]
\[
= P^{-1/2} (P - I) P^{-1/2}
\]
\[
= I - P^{-1} < I
\]

i.e., $\|\tilde{A}\| < 1$

so for a stable system, we can change coordinates so the system is contractive
Lyapunov’s linearization theorem

we consider nonlinear time-invariant system \( \dot{x} = f(x) \), where \( f : \mathbb{R}^n \rightarrow \mathbb{R}^n \)

suppose \( x_e \) is an equilibrium point, \( i.e., f(x_e) = 0 \), and let
\( A = Df(x_e) \in \mathbb{R}^{n \times n} \)

the linearized system, near \( x_e \), is \( \dot{\delta x} = A\delta x \)

linearization theorem:

- if the linearized system is stable, \( i.e., \Re \lambda_i(A) < 0 \) for \( i = 1, \ldots, n \), then
  the nonlinear system is locally asymptotically stable
- if for some \( i \), \( \Re \lambda_i(A) > 0 \), then the nonlinear system is not locally asymptotically stable
stability of the linearized system determines the local stability of the nonlinear system, except when all eigenvalues are in the closed left halfplane, and at least one is on the imaginary axis.

Examples like $\dot{x} = x^3$ (which is not LAS) and $\dot{x} = -x^3$ (which is LAS) show the theorem cannot, in general, be tightened.

**Examples:**

<table>
<thead>
<tr>
<th>Eigenvalues of $Df(x_e)$</th>
<th>Conclusion about $\dot{x} = f(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-3, -0.1 \pm 4j, -0.2 \pm j$</td>
<td>LAS near $x_e$</td>
</tr>
<tr>
<td>$-3, -0.1 \pm 4j, 0.2 \pm j$</td>
<td>not LAS near $x_e$</td>
</tr>
<tr>
<td>$-3, -0.1 \pm 4j, \pm j$</td>
<td>no conclusion</td>
</tr>
</tbody>
</table>
Proof of linearization theorem

let’s assume $x_e = 0$, and express the nonlinear differential equation as

$$\dot{x} = Ax + g(x)$$

where $\|g(x)\| \leq K\|x\|^2$

suppose that $A$ is stable, and let $P$ be unique solution of Lyapunov equation

$$A^TP + PA + I = 0$$

the Lyapunov function $V(z) = z^TPz$ proves stability of the linearized system; we’ll use it to prove local asymptotic stability of the nonlinear system
\[ \dot{V}(z) = 2z^T P (Az + g(z)) \]
\[ = z^T (A^T P + PA) z + 2z^T P g(z) \]
\[ = -z^T z + 2z^T P g(z) \]
\[ \leq -\|z\|^2 + 2\|z\|\|P\|\|g(z)\| \]
\[ \leq -\|z\|^2 + 2K\|P\|\|z\|^3 \]
\[ = -\|z\|^2 (1 - 2K\|P\|\|z\|) \]

so for \( \|z\| \leq 1/(4K\|P\|) \),

\[ \dot{V}(z) \leq -\frac{1}{2}\|z\|^2 \leq -\frac{1}{2\lambda_{\text{max}}(P)} z^T P z = -\frac{1}{2\|P\|} z^T P z \]
finally, using

$$\|z\|^2 \leq \frac{1}{\lambda_{\text{min}}(P)} z^T P z$$

we have

$$V(z) \leq \frac{\lambda_{\text{min}}(P)}{16K^2\|P\|^2} \implies \|z\| \leq \frac{1}{4K\|P\|} \implies \dot{V}(z) \leq -\frac{1}{2\|P\|} V(z)$$

and we’re done

comments:

• proof actually constructs an ellipsoid inside basin of attraction of $x_e = 0$, and a bound on exponential rate of convergence

• choice of $Q = I$ was arbitrary; can get better estimates using other $Q$s, better bounds on $g$, tighter bounding arguments . . .
Integral quadratic performance

consider $\dot{x} = f(x), \ x(0) = x_0$

we are interested in the integral quadratic performance measure

$$J(x_0) = \int_0^\infty x(t)^T Q x(t) \ dt$$

for any fixed $x_0$ we can find this (approximately) by simulation and numerical integration

(we’ll assume the integral exists; we do not require $Q \geq 0$)
suppose there is a function $V : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

- $V(z) \geq 0$ for all $z$
- $\dot{V}(z) \leq -z^T Q z$ for all $z$

then we have $J(x_0) \leq V(x_0)$, i.e., the Lyapunov function $V$ serves as an upper bound on the integral quadratic cost

(since $Q$ need not be PSD, we might not have $\dot{V} \leq 0$; so we cannot conclude that trajectories are bounded)
to show this, we note that

$$V(x(T)) - V(x(0)) = \int_0^T \dot{V}(x(t)) \, dt \leq -\int_0^T x(t)^T Q x(t) \, dt$$

and so

$$\int_0^T x(t)^T Q x(t) \, dt \leq V(x(0)) - V(x(T)) \leq V(x(0))$$

since this holds for arbitrary $T$, we conclude

$$\int_0^\infty x(t)^T Q x(t) \, dt \leq V(x(0))$$
Integral quadratic performance for linear systems

for a stable linear system, with $Q \geq 0$, the Lyapunov bound is sharp, i.e., there exists a $V$ such that

- $V(z) \geq 0$ for all $z$
- $\dot{V}(z) \leq -z^T Q z$ for all $z$

and for which $V(x_0) = J(x_0)$ for all $x_0$

(take $V(z) = z^T P z$, where $P$ is solution of $A^T P + PA + Q = 0$)
Lecture 12
Lyapunov theory with inputs and outputs

• systems with inputs and outputs
• reachability bounding
• bounds on RMS gain
• bounded-real lemma
• feedback synthesis via control-Lyapunov functions
Systems with inputs

we now consider systems with inputs, \(\dot{x} = f(x, u)\), where \(x(t) \in \mathbb{R}^n\), \(u(t) \in \mathbb{R}^m\)

if \(x, u\) is state-input trajectory and \(V : \mathbb{R}^n \to \mathbb{R}\), then

\[
\frac{d}{dt} V(x(t)) = \nabla V(x(t))^T \dot{x}(t) = \nabla V(x(t))^T f(x(t), u(t))
\]

so we define \(\dot{V} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}\) as

\[
\dot{V}(z, w) = \nabla V(z)^T f(z, w)
\]

(\(i.e., \dot{V}\) depends on the state and input)
Reachable set with admissible inputs

consider \( \dot{x} = f(x, u) \), \( x(0) = 0 \), and \( u(t) \in \mathcal{U} \) for all \( t \)

\( \mathcal{U} \subseteq \mathbb{R}^m \) is called the set of admissible inputs

we define the reachable set as

\[
\mathcal{R} = \left\{ x(T) \mid \dot{x} = f(x, u), \ x(0) = 0, \ u(t) \in \mathcal{U}, \ T > 0 \right\}
\]

i.e., the set of points that can be hit by a trajectory with some admissible input

applications:

- if \( u \) is a control input that we can manipulate, \( \mathcal{R} \) shows the places we can hit (so big \( \mathcal{R} \) is good)
- if \( u \) is a disturbance, noise, or antagonistic signal (beyond our control), \( \mathcal{R} \) shows the worst-case effect on \( x \) (so big \( \mathcal{R} \) is bad)
Lyapunov bound on reachable set

Lyapunov arguments can be used to bound reachable sets of nonlinear or time-varying systems.

Suppose there is a $V : \mathbb{R}^n \rightarrow \mathbb{R}$ and $a > 0$ such that

$$\dot{V}(z, w) \leq -a \text{ whenever } V(z) = b \text{ and } w \in \mathcal{U}$$

and define $C = \{z \mid V(z) \leq b\}$

Then, if $\dot{x} = f(x, u)$, $x(0) \in C$, and $u(t) \in \mathcal{U}$ for $0 \leq t \leq T$, we have $x(T) \in C$

i.e., every trajectory that starts in $C = \{z \mid V(z) \leq b\}$ stays there, for any admissible $u$

In particular, if $0 \in C$, we conclude $\mathcal{R} \subseteq C$
idea: on the boundary of $C$, every trajectory cuts into $C$, for all admissable values of $u$

proof: suppose $\dot{x} = f(x, u)$, $x(0) \in C$, and $u(t) \in \mathcal{U}$ for $0 \leq t \leq T$, and $V$ satisfies hypotheses

suppose that $x(T) \notin C$

consider scalar function $g(t) = V(x(t))$

$g(0) \leq b$ and $g(T) > b$, so there is a $t_0 \in [0, T]$ with $g(t_0) = b$, $g'(t_0) \geq 0$

but

$$g'(t_0) = \frac{d}{dt}V(x(t)) = \dot{V}(x(t), u(t)) \leq -a < 0$$

by the hypothesis, so we have a contradiction
Reachable set with integral quadratic bounds

we consider \( \dot{x} = f(x, u), \ x(0) = 0 \), with an integral constraint on the input:

\[
\int_0^\infty u(t)^T u(t) \, dt \leq a
\]

the reachable set with this integral quadratic bound is

\[
\mathcal{R}_a = \left\{ x(T) \mid \dot{x} = f(x, u), \ x(0) = x_0, \ \int_0^T u(t)^T u(t) \, dt \leq a \right\}
\]

i.e., the set of points that can be hit using at most \( a \) energy
Example

consider stable linear system $\dot{x} = Ax + Bu$

minimum energy (i.e., integral of $u^T u$) to hit point $z$ is $z^T W_c^{-1} z$, where $W_c$ is controllability Grammian

reachable set with integral quadratic bound is (open) ellipsoid

$$\mathcal{R}_a = \{ z | z^T W_c^{-1} z < a \}$$
Lyapunov bound on reachable set with integral constraint

suppose there is a $V : \mathbb{R}^n \to \mathbb{R}$ such that

- $V(z) \geq 0$ for all $z$, $V(0) = 0$
- $\dot{V}(z, w) \leq w^T w$ for all $z, w$

then $\mathcal{R}_a \subseteq \{z \mid V(z) \leq a\}$

proof:

$$V(x(T)) - V(x(0)) = \int_0^T \dot{V}(x(t), u(t)) \, dt \leq \int_0^T u(t)^T u(t) \, dt \leq a$$

so, using $V(x(0)) = V(0) = 0$, $V(x(T)) \leq a$
interpretation:

- $V$ is (generalized) internally stored energy in system
- $u(t)^T u(t)$ is power supplied to system by input
- $\dot{V} \leq u^T u$ means stored energy increases by no more than power input
- $V(0) = 0$ means system starts in zero energy state
- conclusion is: if energy $\leq a$ applied, can only get to states with stored energy $\leq a$
Stable linear system

consider stable linear system \( \dot{x} = Ax + Bu \)

we’ll show Lyapunov bound is tight in this case, with \( V(z) = z^TW_c^{-1}z \)
multiply \( AW_c + W_cA^T + BB^T = 0 \) on left & right by \( W_c^{-1} \) to get

\[ W_c^{-1}A + A^TW_c^{-1} + W_c^{-1}BB^TW_c^{-1} = 0 \]

now we can find and bound \( \dot{V} \):

\[
\dot{V}(z, w) = 2z^TW_c^{-1}(Az + Bw)
= z^T(W_c^{-1}A + A^TW_c^{-1})z + 2z^TW_c^{-1}Bw
= -z^TW_c^{-1}BB^TW_c^{-1}z + 2z^TW_c^{-1}Bw
= -\|B^TW_c^{-1}z - w\|^2 + w^Tw
\leq w^Tw
\]
for $V(z) = z^T W_c^{-1} z$, Lyapunov bound is

$$\mathcal{R}_a \subseteq \{z \mid z^T W_c^{-1} z \leq a\}$$

righthand set is closure of lefthand set, so bound is tight

roughly speaking, for a stable linear system, a point is reachable with an integral quadratic bound if and only if there is a quadratic Lyapunov function that proves it
(except for points right on the boundary)
RMS gain

recall that the RMS value of a signal is given by

\[
\text{rms}(z) = \left( \lim_{T \to \infty} \frac{1}{T} \int_0^T \|z(t)\|^2 \, dt \right)^{1/2}
\]

assuming the limit exists

now consider a system with input signal \( u \) and output signal \( y \)

we define its RMS gain as the maximum of \( \text{rms}(y)/\text{rms}(u) \), over all \( u \) with nonzero RMS value
Lyapunov method for bounding RMS gain

now consider the nonlinear system

$$\dot{x} = f(x, u), \quad x(0) = 0, \quad y = g(x, u)$$

with $x(t) \in \mathbb{R}^n$, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$

we can use Lyapunov methods to bound its RMS gain

suppose $\gamma \geq 0$, and there is a $V : \mathbb{R}^n \rightarrow \mathbb{R}$ such that

- $V(z) \geq 0$ for all $z$, $V(0) = 0$
- $\dot{V}(z, w) \leq \gamma^2 w^T w - y^T y$ for all $z, w$
  (i.e., $\dot{V}(z, w) \leq \gamma^2 w^T w - g(z, w)^T g(z, w)$ for all $z, w$)

then, the RMS gain of the system is no more than $\gamma$
proof:

\[ V(x(T)) - V(x(0)) = \int_0^T \dot{V}(x(t), u(t)) \, dt \]

\[ \leq \int_0^T (\gamma^2 u(t)^T u(t) - y(t)^T y(t)) \, dt \]

using \( V(x(0)) = V(0) = 0 \), \( V(x(T)) \geq 0 \), we have

\[ \int_0^T y(t)^T y(t) \, dt \leq \gamma^2 \int_0^T u(t)^T u(t) \, dt \]

dividing by \( T \) and taking the limit \( T \to \infty \) yields \( \text{rms}(y)^2 \leq \gamma^2 \text{rms}(u)^2 \)
Bounded-real lemma

let’s use a quadratic Lyapunov function $V(z) = z^T P z$ to bound the RMS gain of the stable linear system $\dot{x} = Ax + Bu$, $x(0) = 0$, $y = Cx$

the conditions on $V$ give $P \geq 0$

the condition $\dot{V}(z, w) \leq \gamma^2 w^T w - g(z, w)^T g(z, w)$ becomes

$$\dot{V}(z, w) = 2 z^T P (Az + Bw) \leq \gamma^2 w^T w - (Cz)^T Cz$$

for all $z$, $w$

let’s write that as a quadratic form in $(z, w)$:

$$\begin{bmatrix} z \\ w \end{bmatrix}^T \begin{bmatrix} A^T P + PA + C^T C & PB \\ B^T P & -\gamma^2 I \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix} \leq 0$$
so we conclude: if there is a $P \geq 0$ such that

$$
\begin{bmatrix}
A^T P + PA + C^T C & PB \\
B^T P & -\gamma^2 I
\end{bmatrix} \leq 0
$$

then the RMS gain of the linear system is no more than $\gamma$

it turns out that for linear systems this condition is not only sufficient, but also necessary

(this result is called the \textit{bounded-real lemma})

by taking Schur complement, we can express the block $2 \times 2$ matrix inequality as

$$A^T P + PA + C^T C + \gamma^{-2} PBB^T P \leq 0$$

(which is a Riccati-like quadratic matrix \textit{inequality} . . . )
Nonlinear optimal control

we consider $\dot{x} = f(x, u), \ u(t) \in \mathcal{U} \subseteq \mathbb{R}^m$

here we consider $u$ to be an input we can manipulate to achieve some desired response, such as minimizing, or at least making small,

$$J = \int_0^\infty x(t)^T Q x(t) \, dt$$

where $Q \geq 0$

(many other choices for criterion will work)
we can solve via dynamic programming: let \( V : \mathbb{R}^n \rightarrow \mathbb{R} \) denote value function, \( i.e., \)

\[
V(z) = \min \{ J \mid \dot{x} = f(x, u), \quad x(0) = z, \quad u(t) \in \mathcal{U} \}
\]

then the optimal \( u \) is given by

\[
u^*(t) = \arg \min_{w \in \mathcal{U}} \dot{V}(x(t), w)
\]

and with the optimal \( u \) we have

\[
\dot{V}(x(t), u^*) = -x(t)^T Q x(t)
\]

but, it can be very difficult to find \( V \), and therefore \( u^* \)

Lyapunov theory with inputs and outputs 12–18
Feedback design via control-Lyapunov functions

suppose there is a function $V : \mathbb{R}^n \to \mathbb{R}$ such that

- $V(z) \geq 0$ for all $z$
- for all $z$, $\min_{w \in U} \dot{V}(z, w) \leq -z^T Q z$

then, the state feedback control law $u(t) = g(x(t))$, with

$$g(z) = \arg\min_{w \in U} \dot{V}(z, w)$$

results in $J \leq V(x(0))$

in this case $V$ is called a control-Lyapunov function for the problem
• if $V$ is the value function, this method recovers the optimal control law

• we’ve used Lyapunov methods to generate a suboptimal control law, but one with a guaranteed bound on the cost function

• the control law is a greedy one, that simply chooses $u(t)$ to decrease $V$ as quickly as possible (subject to $u(t) \in \mathcal{U}$)

• the inequality $\min_{w \in \mathcal{U}} \dot{V}(z,w) \leq -z^TQz$ is the inequality form of $\min_{w \in \mathcal{U}} \dot{V}(z,w) = -z^TQz$, which holds for the optimal input, and $V$ the value function

control-Lyapunov methods offer a good way to generate suboptimal control laws, with performance guarantees, when the optimal control is too hard to find
Lecture 13
Linear matrix inequalities and the S-procedure

• Linear matrix inequalities
• Semidefinite programming
• S-procedure for quadratic forms and quadratic functions
Linear matrix inequalities

suppose $F_0, \ldots, F_n$ are symmetric $m \times m$ matrices

an inequality of the form

$$F(x) = F_0 + x_1 F_1 + \cdots + x_n F_n \geq 0$$

is called a linear matrix inequality (LMI) in the variable $x \in \mathbb{R}^n$

here, $F : \mathbb{R}^n \rightarrow \mathbb{R}^{m \times m}$ is an affine function of the variable $x$
LMIs:

• can represent a wide variety of inequalities
• arise in many problems in control, signal processing, communications, statistics, . . .

most important for us: **LMIs can be solved very efficiently** by newly developed methods (EE364)

“solved” means: we can find $x$ that satisfies the LMI, or determine that no solution exists
Example

\[ F(x) = \begin{bmatrix} x_1 + x_2 & x_2 + 1 \\ x_2 + 1 & x_3 \end{bmatrix} \geq 0 \]

\[ F_0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad F_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad F_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad F_3 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \]

LMI \( F(x) \geq 0 \) equivalent to

\[ x_1 + x_2 \geq 0, \quad x_3 \geq 0 \]
\[ (x_1 + x_2)x_3 - (x_2 + 1)^2 = x_1x_3 + x_2x_3 - x_2^2 - 2x_2 - 1 \geq 0 \]

\[ \ldots \text{a set of } \textit{nonlinear} \text{ inequalities in } x \]
Certifying infeasibility of an LMI

• if $A, B$ are symmetric PSD, then $\text{Tr}(AB) \geq 0$:

$$\text{Tr}(AB) = \text{Tr} \left( A^{1/2} B^{1/2} B^{1/2} A^{1/2} \right) = \left\| A^{1/2} B^{1/2} \right\|_F^2$$

• suppose $Z = Z^T$ satisfies

$$Z \geq 0, \quad \text{Tr}(F_0Z) < 0, \quad \text{Tr}(F_iZ) = 0, \quad i = 1, \ldots, n$$

• then if $F(x) = F_0 + x_1 F_1 + \cdots + x_n F_n \geq 0$,

$$0 \leq \text{Tr}(ZF(x)) = \text{Tr}(ZF_0) < 0$$

a contradiction

• $Z$ is *certificate* that proves LMI $F(x) \geq 0$ is infeasible
Example: Lyapunov inequality

suppose $A \in \mathbb{R}^{n \times n}$

the Lyapunov inequality $A^T P + PA + Q \leq 0$ is an LMI in variable $P$

meaning: $P$ satisfies the Lyapunov LMI if and only if the quadratic form $V(z) = z^T P z$ satisfies $\dot{V}(z) \leq z^T Q z$, for system $\dot{x} = Ax$

the dimension of the variable $P$ is $n(n + 1)/2$ (since $P = P^T$)

here, $F(P) = -A^T P - PA - Q$ is affine in $P$

(we don’t need special LMI methods to solve the Lyapunov inequality; we can solve it analytically by solving the Lyapunov equation $A^T P + PA + Q = 0$)
Extensions

multiple LMIs: we can consider multiple LMIs as one, large LMI, by forming block diagonal matrices:

\[ F^{(1)}(x) \geq 0, \ldots, F^{(k)}(x) \geq 0 \iff \text{diag} \left( F^{(1)}(x), \ldots, F^{(k)}(x) \right) \geq 0 \]

example: we can express a set of linear inequalities as an LMI with diagonal matrices:

\[ a_1^T x \leq b_1, \ldots, a_k^T x \leq b_k \iff \text{diag}(b_1 - a_1^T x, \ldots, b_k - a_k^T x) \geq 0 \]

linear equality constraints: \( a^T x = b \) is the same as the pair of linear inequalities \( a^T x \leq b, a^T x \geq b \)
Example: bounded-real LMI

suppose \( A \in \mathbb{R}^{n \times n} \), \( B \in \mathbb{R}^{n \times m} \), \( C \in \mathbb{R}^{p \times n} \), and \( \gamma > 0 \)

the *bounded-real LMI* is

\[
\begin{bmatrix}
A^T P + PA + C^T C & PB \\
B^T P & -\gamma^2 I
\end{bmatrix} \leq 0, \quad P \geq 0
\]

with variable \( P \)

meaning: if \( P \) satisfies this LMI, then the quadratic Lyapunov function \( V(z) = z^T P z \) proves the RMS gain of the system \( \dot{x} = Ax + Bu \), \( y = Cx \) is no more than \( \gamma \)

(in fact we can solve this LMI by solving an ARE-like equation, so we don’t need special LMI methods . . . )
Strict inequalities in LMIs

sometimes we encounter strict matrix inequalities

\[ F(x) \geq 0, \quad F_{\text{strict}}(x) > 0 \]

where \( F, F_{\text{strict}} \) are affine functions of \( x \)

• practical approach: replace \( F_{\text{strict}}(x) > 0 \) with \( F_{\text{strict}}(x) \geq \epsilon I \), where \( \epsilon \) is small and positive

• if \( F \) and \( F_{\text{strict}} \) are homogenous (i.e., linear functions of \( x \)) we can replace with

\[ F(x) \geq 0, \quad F_{\text{strict}}(x) \geq I \]

example: we can replace \( A^T P + PA \leq 0, \ P > 0 \) (with variable \( P \)) with
\[ A^T P + PA \leq 0, \ P \geq I \]
Quadratic Lyapunov function for time-varying LDS

we consider time-varying linear system \( \dot{x}(t) = A(t)x(t) \) with

\[
A(t) \in \{ A_1, \ldots, A_K \}
\]

• we want to establish some property, such as all trajectories are bounded

• this is hard to do in general (cf. time-invariant LDS)

• let’s use quadratic Lyapunov function \( V(z) = z^TPz \); we need \( P > 0 \), and \( \dot{V}(z) \leq 0 \) for all \( z \), and all possible values of \( A(t) \)

• gives

\[
P > 0, \quad A_i^TP + PA_i \leq 0, \quad i = 1, \ldots, K
\]

• by homogeneity, can write as LMIs

\[
P \succeq I, \quad A_i^TP + PA_i \leq 0, \quad i = 1, \ldots, K
\]
• in this case \( V \) is called *simultaneous Lyapunov function* for the systems 
\[
\dot{x} = A_i x, \quad i = 1, \ldots, K
\]

• there is no analytical method (e.g., using AREs) to solve such an LMI, but it is easily done numerically

• if such a \( P \) exists, it proves boundedness of trajectories of 
\[
\dot{x}(t) = A(t)x(t), \text{ with }
\]
\[
A(t) = \theta_1(t)A_1 + \cdots + \theta_K(t)A_K
\]
where \( \theta_i(t) \geq 0, \quad \theta_1(t) + \cdots + \theta_K(t) = 1 \)

• in fact, it works for the *nonlinear* system \( \dot{x} = f(x) \) provided for each 
\( z \in \mathbb{R}^n \),
\[
Df(z) = \theta_1(z)A_1 + \cdots + \theta_K(z)A_K
\]
for some \( \theta_i(z) \geq 0, \quad \theta_1(z) + \cdots + \theta_K(z) = 1 \)
Semidefinite programming

A semidefinite program (SDP) is an optimization problem with linear objective and LMI and linear equality constraints:

\[
\begin{align*}
\text{minimize} & \quad c^T x \\
\text{subject to} & \quad F_0 + x_1 F_1 + \cdots + x_n F_n \succeq 0 \\
& \quad Ax = b
\end{align*}
\]

Most important property for us:

We can solve SDPs globally and efficiently

Meaning: we either find a globally optimal solution, or determine that there is no \( x \) that satisfies the LMI & equality constraints.
Example: Let $A \in \mathbb{R}^{n \times n}$ be stable, $Q = Q^T \geq 0$

then the LMI $A^T P + PA + Q \leq 0$, $P \geq 0$ in $P$ means the quadratic Lyapunov function $V(z) = z^T P z$ proves the bound

$$\int_0^\infty x(t)^T Q x(t) \, dt \leq x(0)^T P x(0)$$

now suppose that $x(0)$ is fixed, and we seek the best possible such bound

this can be found by solving the SDP

$$\text{minimize} \quad x(0)^T P x(0)$$

$$\text{subject to} \quad A^T P + PA + Q \leq 0, \quad P \geq 0$$

with variable $P$ (note that the objective is linear in $P$)

(in fact we can solve this SDP analytically, by solving the Lyapunov equation)
S-procedure for two quadratic forms

let \( F_0 = F_0^T, F_1 = F_1^T \in \mathbb{R}^{n \times n} \)

when is it true that, for all \( z \), \( z^T F_1 z \geq 0 \) \( \Rightarrow z^T F_0 z \geq 0 \)?

in other words, when does nonnegativity of one quadratic form imply nonnegativity of another?

simple condition: there exists \( \tau \in \mathbb{R}, \tau \geq 0 \), with \( F_0 \geq \tau F_1 \)

then for sure we have \( z^T F_1 z \geq 0 \) \( \Rightarrow z^T F_0 z \geq 0 \)

(since if \( z^T F_1 z \geq 0 \), we then have \( z^T F_0 z \geq \tau z^T F_1 z \geq 0 \))

**fact:** the converse holds, provided there exists a point \( u \) with \( u^T F_1 u > 0 \)

this result is called the *lossless* S-procedure, and is *not* easy to prove

(condition that there exists a point \( u \) with \( u^T F_1 u > 0 \) is called a *constraint qualification*)
S-procedure with strict inequalities

when is it true that, for all \( z \), \( z^T F_1 z \geq 0 \), \( z \neq 0 \) \( \Rightarrow \) \( z^T F_0 z > 0 \)?

in other words, when does nonnegativity of one quadratic form imply positivity of another for nonzero \( z \)?

simple condition: suppose there is a \( \tau \in \mathbb{R}, \tau \geq 0 \), with \( F_0 > \tau F_1 \)

fact: the converse holds, provided there exists a point \( u \) with \( u^T F_1 u > 0 \)

again, this is not easy to prove
Example

let’s use quadratic Lyapunov function $V(z) = z^T P z$ to prove stability of

$$\dot{x} = Ax + g(x), \quad \|g(x)\| \leq \gamma \|x\|$$

we need $P > 0$ and $\dot{V}(x) \leq -\alpha V(x)$ for all $x$ ($\alpha > 0$ is given)

$$\dot{V}(x) + \alpha V(x) = 2x^T P (Ax + g(x)) + \alpha x^T P x$$
$$= x^T (A^T P + PA + \alpha P) x + 2x^T P z$$
$$= \begin{bmatrix} x \\ z \end{bmatrix}^T \begin{bmatrix} A^T P + PA + \alpha P & P \\ P & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix}$$

where $z = g(x)$
$z$ satisfies $z^T z \leq \gamma^2 x^T x$

so we need $P > 0$ and

$$-\begin{bmatrix} x \\ z \end{bmatrix}^T \begin{bmatrix} A^T P + PA + \alpha P & P \\ P & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} \geq 0$$

whenever

$$\begin{bmatrix} x \\ z \end{bmatrix}^T \begin{bmatrix} \gamma^2 I & 0 \\ 0 & -I \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} \geq 0$$

by S-procedure, this happens if and only if

$$-\begin{bmatrix} A^T P + PA + \alpha P & P \\ P & 0 \end{bmatrix} \geq \tau \begin{bmatrix} \gamma^2 I & 0 \\ 0 & -I \end{bmatrix}$$

for some $\tau \geq 0$

(constraint qualification holds here)
thus, necessary and sufficient conditions for the existence of quadratic Lyapunov function can be expressed as LMI

\[ P > 0, \quad \begin{bmatrix} A^T P + PA + \alpha P + \tau \gamma^2 I & P \\ P & -\tau I \end{bmatrix} \leq 0 \]

in variables \( P, \tau \) (note condition \( \tau \geq 0 \) is automatic from 2, 2 block)

by homogeneity, we can write this as

\[ P \geq I, \quad \begin{bmatrix} A^T P + PA + \alpha P + \tau \gamma^2 I & P \\ P & -\tau I \end{bmatrix} \leq 0 \]

• solving this LMI to find \( P \) is a powerful method
• it beats, for example, solving the Lyapunov equation
  \[ A^T P + PA + I = 0 \]
  and hoping the resulting \( P \) works
S-procedure for multiple quadratic forms

let $F_0 = F_0^T, \ldots, F_k = F_k^T \in \mathbb{R}^{n \times n}$

when is it true that

$$
\text{for all } z, \quad z^T F_1 z \geq 0, \ldots, z^T F_k z \geq 0 \Rightarrow z^T F_0 z \geq 0
$$

(1)

in other words, when does nonnegativity of a set of quadratic forms imply nonnegativity of another?

simple sufficient condition: suppose there are $\tau_1, \ldots, \tau_k \geq 0$, with

$$
F_0 \geq \tau_1 F_1 + \cdots + \tau_k F_k
$$

then for sure the property (1) above holds

(in this case this is only a sufficient condition; it is not necessary)
using the matrix inequality condition

\[ F_0 \geq \tau_1 F_1 + \cdots + \tau_k F_k, \quad \tau_1, \ldots, \tau_k \geq 0 \]

as a sufficient condition for

for all \( z \), \( z^T F_1 z \geq 0, \ldots, z^T F_k z \geq 0 \) \( \Rightarrow z^T F_0 z \geq 0 \)

is called the (lossy) S-procedure

the matrix inequality condition is an LMI in \( \tau_1, \ldots, \tau_k \), therefore easily solved

the constants \( \tau_i \) are called multipliers
Lecture 14
Analysis of systems with sector nonlinearities

- Sector nonlinearities
- Lur’e system
- Analysis via quadratic Lyapunov functions
- Extension to multiple nonlinearities
Sector nonlinearities

A function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is said to be in sector $[l, u]$ if for all $q \in \mathbb{R}$, $p = \phi(q)$ lies between $lq$ and $uq$.

This can be expressed as quadratic inequality

$$(p - uq)(p - lq) \leq 0 \text{ for all } q, \ p = \phi(q)$$
examples:

- sector $[-1, 1]$ means $|\phi(q)| \leq |q|$
- sector $[0, \infty]$ means $\phi(q)$ and $q$ always have same sign (graph in first & third quadrants)

some equivalent statements:

- $\phi$ is in sector $[l, u]$ iff for all $q$,

$$|\phi(q) - \frac{u + l}{2}q| \leq \frac{u - l}{2}|q|$$

- $\phi$ is in sector $[l, u]$ iff for each $q$ there is $\theta(q) \in [l, u]$ with $\phi(q) = \theta(q)q$
Nonlinear feedback representation

linear dynamical system with nonlinear feedback

\[
\dot{x} = Ax + Bp \\
q = Cx
\]

closed-loop system: \( \dot{x} = Ax + B\phi(Cx) \)

- a common representation that separates linear and nonlinear parts
- often \( p, q \) are scalar signals
Lur’e system

A (single nonlinearity) *Lur’e system* has the form

\[
\dot{x} = Ax + Bp, \quad q = Cx, \quad p = \phi(q)
\]

where \( \phi : \mathbb{R} \to \mathbb{R} \) is in sector \([l, u]\)

Here \( A, B, C, l, \) and \( u \) are given; \( \phi \) is otherwise not specified

- A common method for describing nonlinearity and/or uncertainty
- Goal is to prove stability, or derive a bound, using only the sector information about \( \phi \)
- If we succeed, the result is strong, since it applies to a large family of nonlinear systems
Stability analysis via quadratic Lyapunov functions

let’s try to establish global asymptotic stability of Lur’e system, using quadratic Lyapunov function $V(z) = z^TPz$

we’ll require $P > 0$ and $\dot{V}(z) \leq -\alpha V(z)$, where $\alpha > 0$ is given

second condition is:

$$\dot{V}(z) + \alpha V(z) = 2z^TP(Az + B\phi(Cz)) + \alpha z^TPz \leq 0$$

for all $z$ and all sector $[l, u]$ functions $\phi$

same as:

$$2z^TP(Az + Bp) + \alpha z^TPz \leq 0$$

for all $z$, and all $p$ satisfying $(p - uq)(p - lq) \leq 0$, where $q = Cz$
we can express this last condition as a quadratic inequality in $(z, p)$:

$$
\begin{bmatrix}
  z \\
p
\end{bmatrix}^T
\begin{bmatrix}
  \sigma C^T C & -\nu C^T \\
-\nu C & 1
\end{bmatrix}
\begin{bmatrix}
  z \\
p
\end{bmatrix} \leq 0
$$

where $\sigma = lu$, $\nu = (l + u)/2$

so $\dot{V} + \alpha V \leq 0$ is equivalent to:

$$
\begin{bmatrix}
  z \\
p
\end{bmatrix}^T
\begin{bmatrix}
  A^T P + PA + \alpha P & PB \\
B^T P & 0
\end{bmatrix}
\begin{bmatrix}
  z \\
p
\end{bmatrix} \leq 0
$$

whenever

$$
\begin{bmatrix}
  z \\
p
\end{bmatrix}^T
\begin{bmatrix}
  \sigma C^T C & -\nu C^T \\
-\nu C & 1
\end{bmatrix}
\begin{bmatrix}
  z \\
p
\end{bmatrix} \leq 0
$$
by (lossless) S-procedure this is equivalent to: there is a \( \tau \geq 0 \) with

\[
\begin{bmatrix}
A^TP + PA + \alpha P & PB \\
B^TP & 0
\end{bmatrix} \leq \tau
\begin{bmatrix}
\sigma C^T C & -\nu C^T \\
-\nu C & 1
\end{bmatrix}
\]

or

\[
\begin{bmatrix}
A^TP + PA + \alpha P - \tau \sigma C^T C & PB + \tau \nu C^T \\
B^TP + \tau \nu C & -\tau
\end{bmatrix} \leq 0
\]

an LMI in \( P \) and \( \tau \) (2, 2 block automatically gives \( \tau \geq 0 \))

by homogeneity, we can replace condition \( P > 0 \) with \( P \geq I \)

our final LMI is

\[
\begin{bmatrix}
A^TP + PA + \alpha P - \tau \sigma C^T C & PB + \tau \nu C^T \\
B^TP + \tau \nu C & -\tau
\end{bmatrix} \leq 0, \quad P \geq I
\]

with variables \( P \) and \( \tau \)
• hence, can efficiently determine if there exists a quadratic Lyapunov function that proves stability of Lur’e system

• this LMI can also be solved via an ARE-like equation, or by a graphical method that has been known since the 1960s

• this method is more sophisticated and powerful than the 1895 approach:
  – replace nonlinearity with \( \phi(q) = \nu q \)
  – choose \( Q > 0 \) (e.g., \( Q = I \)) and solve Lyapunov equation

\[
(A + \nu BC)^T P + P(A + \nu BC) + Q = 0
\]

for \( P \)

– hope \( P \) works for nonlinear system
Multiple nonlinearities

we consider system

\[
\dot{x} = Ax + Bp, \quad q = Cx, \quad p_i = \phi_i(q_i), \quad i = 1, \ldots, m
\]

where \( \phi_i : \mathbb{R} \rightarrow \mathbb{R} \) is sector \([l_i, u_i]\)

we seek \( V(z) = z^T P z \), with \( P > 0 \), so that \( \dot{V} + \alpha V \leq 0 \)

last condition equivalent to:

\[
\begin{bmatrix} z \\ p \end{bmatrix}^T \begin{bmatrix} A^T P + PA + \alpha P & PB \\ B^T P & 0 \end{bmatrix} \begin{bmatrix} z \\ p \end{bmatrix} \leq 0
\]

whenever

\[
(p_i - u_i q_i)(p_i - l_i q_i) \leq 0, \quad i = 1, \ldots, m
\]
we can express this last condition as

\[
\begin{bmatrix}
z \\
p
\end{bmatrix}^T
\begin{bmatrix}
\sigma c_i^T c_i & -\nu_i c_i^T e_i^T \\
-\nu_i e_i c_i & e_i e_i^T
\end{bmatrix}
\begin{bmatrix}
z \\
p
\end{bmatrix} \leq 0, \quad i = 1, \ldots, m
\]

where \( c_i \) is the \( i \)th row of \( C \), \( e_i \) is the \( i \)th unit vector, \( \sigma_i = l_i u_i \), and \( \nu_i = (l_i + u_i)/2 \)

now we use (lossy) S-procedure to get a sufficient condition: there exists \( \tau_1, \ldots, \tau_m \geq 0 \) such that

\[
\begin{bmatrix}
A^T P + PA + \alpha P - \sum_{i=1}^{m} \tau_i \sigma_i c_i^T c_i & PB + \sum_{i=1}^{m} \tau_i \nu_i c_i^T \\
B^T P + \sum_{i=1}^{m} \tau_i \nu_i c_i & -\sum_{i=1}^{m} \tau_i e_i e_i^T
\end{bmatrix} \leq 0
\]
we can write this as:

\[
\begin{bmatrix}
A^T P + PA + \alpha P - C^T DF C & PB + C^T DG \\
B^T P + DGC & -D
\end{bmatrix} \leq 0
\]

where

\[D = \text{diag}(\tau_1, \ldots, \tau_m), \quad F = \text{diag}(\sigma_1, \ldots, \sigma_m), \quad G = \text{diag}(\nu_1, \ldots, \nu_m)\]

- this is an LMI in variables $P$ and $D$
- $2, 2$ block automatically gives us $\tau_i \geq 0$
- by homogeneity, we can add $P \geq I$ to ensure $P > I$
- solving these LMIs allows us to (sometimes) find quadratic Lyapunov functions for Lur’e system with multiple nonlinearities (which was impossible until recently)
we consider system

\[
\begin{align*}
\dot{x}_2 &= \phi_1(x_1), \\
\dot{x}_3 &= \phi_2(x_2), \\
\dot{x}_1 &= \phi_3(-2(x_1 + x_2 + x_3))
\end{align*}
\]

where \(\phi_1, \phi_2, \phi_3\) are sector \([1 - \delta, 1 + \delta]\)

- \(\delta\) gives the percentage nonlinearity
- for \(\delta = 0\), we have (stable) linear system \(\dot{x} = \begin{bmatrix} -2 & -2 & -2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x\)
let's put system in Lur'e form:

\[
\dot{x} = Ax + Bp, \quad q = Cx, \quad p_i = \phi_i(q_i)
\]

where

\[
A = 0, \quad B = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & -2 & -2 \end{bmatrix}
\]

the sector limits are \( l_i = 1 - \delta, u_i = 1 + \delta \)

define \( \sigma = l_i u_i = 1 - \delta^2 \), and note that \( (l_i + u_i)/2 = 1 \)
we take \( x(0) = (1, 0, 0) \), and seek to bound \( J = \int_0^\infty \|x(t)\|^2 \, dt \)

(for \( \delta = 0 \) we can calculate \( J \) exactly by solving a Lyapunov equation)

we’ll use quadratic Lyapunov function \( V(z) = z^T P z \), with \( P \geq 0 \)

Lyapunov conditions for bounding \( J \): if \( \dot{V}(z) \leq -z^T z \) whenever the sector conditions are satisfied, then \( J \leq x(0)^T P x(0) = P_{11} \)

use S-procedure as above to get sufficient condition:

\[
\begin{bmatrix}
A^T P + PA + I - \sigma C^T D C & PB + C^T D \\
B^T P + DC & -D
\end{bmatrix} \leq 0
\]

which is an LMI in variables \( P \) and \( D = \text{diag}(\tau_1, \tau_2, \tau_3) \)

note that LMI gives \( \tau_i \geq 0 \) automatically
to get best bound on $J$ for given $\delta$, we solve SDP

$$\begin{align*}
\text{minimize} & \quad P_{11} \\
\text{subject to} & \quad \begin{bmatrix}
A^T P + PA + I - \sigma C^T D C & PB + C^T D \\
B^T P + DC & -D
\end{bmatrix} \leq 0 \\
P & \geq 0
\end{align*}$$

with variables $P$ and $D$ (which is diagonal)

optimal value gives best bound on $J$ that can be obtained from a quadratic Lyapunov function, using S-procedure
• top plot shows bound on $J$; bottom points show results for constant linear $\phi_i$'s chosen at random in interval $1 \pm \delta$

• bound is exact for $\delta = 0$; for $\delta \geq 0.15$, LMI is infeasible
Lecture 15
Perron-Frobenius Theory

- Positive and nonnegative matrices and vectors
- Perron-Frobenius theorems
- Markov chains
- Economic growth
- Population dynamics
- Max-min and min-max characterization
- Power control
- Linear Lyapunov functions
- Metzler matrices
Positive and nonnegative vectors and matrices

we say a matrix or vector is

- **positive** (or **elementwise positive**) if all its entries are positive
- **nonnegative** (or **elementwise nonnegative**) if all its entries are nonnegative

we use the notation $x > y$ ($x \geq y$) to mean $x - y$ is elementwise positive (nonnegative)

**warning:** if $A$ and $B$ are square and symmetric, $A \geq B$ can mean:

- $A - B$ is PSD (*i.e.*, $z^T A z \geq z^T B z$ for all $z$), or
- $A - B$ elementwise positive (*i.e.*, $A_{ij} \geq B_{ij}$ for all $i$, $j$)

in this lecture, $>$ and $\geq$ mean elementwise
Application areas

nonnegative matrices arise in many fields, *e.g.*, 

- economics
- population models
- graph theory
- Markov chains
- power control in communications
- Lyapunov analysis of large scale systems
Basic facts

if $A \geq 0$ and $z \geq 0$, then we have $Az \geq 0$

conversely: if for all $z \geq 0$, we have $Az \geq 0$, then we can conclude $A \geq 0$

in other words, matrix multiplication preserves nonnegativity if and only if the matrix is nonnegative

if $A > 0$ and $z \geq 0$, $z \neq 0$, then $Az > 0$

conversely, if whenever $z \geq 0$, $z \neq 0$, we have $Az > 0$, then we can conclude $A > 0$

if $x \geq 0$ and $x \neq 0$, we refer to $d = (1/1^T x)x$ as its distribution or normalized form

$d_i = x_i / (\sum_j x_j)$ gives the fraction of the total of $x$, given by $x_i$
Regular nonnegative matrices

suppose $A \in \mathbb{R}^{n \times n}$, with $A \geq 0$

$A$ is called regular if for some $k \geq 1$, $A^k > 0$

meaning: form directed graph on nodes 1, \ldots, $n$, with an arc from $j$ to $i$ whenever $A_{ij} > 0$

then $(A^k)_{ij} > 0$ if and only if there is a path of length $k$ from $j$ to $i$

$A$ is regular if for some $k$ there is a path of length $k$ from every node to every other node
examples:

• any positive matrix is regular

\[
\begin{bmatrix}
1 & 1 \\
0 & 1
\end{bmatrix}
\]
and
\[
\begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix}
\]
are not regular

\[
\begin{bmatrix}
1 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0
\end{bmatrix}
\]
is regular
Perron-Frobenius theorem for regular matrices

suppose $A \in \mathbb{R}^{n \times n}$ is nonnegative and regular, i.e., $A^k > 0$ for some $k$

then

- there is an eigenvalue $\lambda_{\text{pf}}$ of $A$ that is real and positive, with positive left and right eigenvectors
- for any other eigenvalue $\lambda$, we have $|\lambda| < \lambda_{\text{pf}}$
- the eigenvalue $\lambda_{\text{pf}}$ is simple, i.e., has multiplicity one, and corresponds to a $1 \times 1$ Jordan block

the eigenvalue $\lambda_{\text{pf}}$ is called the *Perron-Frobenius* (PF) eigenvalue of $A$

the associated positive (left and right) eigenvectors are called the (left and right) PF eigenvectors (and are unique, up to positive scaling)
**Perron-Frobenius theorem for nonnegative matrices**

suppose $A \in \mathbb{R}^{n \times n}$ and $A \geq 0$

then

- there is an eigenvalue $\lambda_{pf}$ of $A$ that is real and nonnegative, with associated nonnegative left and right eigenvectors
- for any other eigenvalue $\lambda$ of $A$, we have $|\lambda| \leq \lambda_{pf}$

$\lambda_{pf}$ is called the *Perron-Frobenius* (PF) eigenvalue of $A$

the associated nonnegative (left and right) eigenvectors are called (left and right) PF eigenvectors

in this case, they need not be unique, or positive
Markov chains

we consider stochastic process $X(0), X(1), \ldots$ with values in $\{1, \ldots, n\}$

$\text{Prob}(X(t + 1) = i | X(t) = j) = P_{ij}$

$P$ is called the transition matrix; clearly $P_{ij} \geq 0$

let $p(t) \in \mathbb{R}^n$ be the distribution of $X(t)$, i.e., $p_i(t) = \text{Prob}(X(t) = i)$

then we have $p(t + 1) = Pp(t)$

*note*: standard notation uses transpose of $P$, and row vectors for probability distributions

$P$ is a stochastic matrix, i.e., $P \geq 0$ and $\mathbf{1}^T P = \mathbf{1}^T$

so $\mathbf{1}$ is a left eigenvector with eigenvalue 1, which is in fact the PF eigenvalue of $P$
Equilibrium distribution

let \( \pi \) denote a PF (right) eigenvector of \( P \), with \( \pi \geq 0 \) and \( 1^T \pi = 1 \)

since \( P \pi = \pi \), \( \pi \) corresponds to an invariant distribution or equilibrium distribution of the Markov chain

now suppose \( P \) is regular, which means for some \( k \), \( P^k > 0 \)

since \( (P^k)_{ij} \) is \( \text{Prob}(X(t + k) = i|X(t) = j) \), this means there is positive probability of transitioning from any state to any other in \( k \) steps

since \( P \) is regular, there is a unique invariant distribution \( \pi \), which satisfies \( \pi > 0 \)

the eigenvalue 1 is simple and dominant, so we have \( p(t) \to \pi \), no matter what the initial distribution \( p(0) \)

in other words: the distribution of a regular Markov chain always converges to the unique invariant distribution
Rate of convergence to equilibrium distribution

rate of convergence to equilibrium distribution depends on second largest eigenvalue magnitude, i.e.,

\[ \mu = \max\{|\lambda_2|, \ldots, |\lambda_n|\} \]

where \(\lambda_i\) are the eigenvalues of \(P\), and \(\lambda_1 = \lambda_{pf} = 1\)

(\(\mu\) is sometimes called the SLEM of the Markov chain)

the mixing time of the Markov chain is given by

\[ T = \frac{1}{\log(1/\mu)} \]

(roughly, number of steps over which deviation from equilibrium distribution decreases by factor \(e\))
Dynamic interpretation

consider $x(t + 1) = Ax(t)$, with $A \geq 0$ and regular

then by PF theorem, $\lambda_{pf}$ is the unique dominant eigenvalue

let $v, w > 0$ be the right and left PF eigenvectors of $A$, with $1^Tv = 1, w^Tv = 1$

then as $t \to \infty$, $(\lambda_{pf}^{-1}A)^t \to vw^T$

for any $x(0) \geq 0$, $x(0) \neq 0$, we have

$$\frac{1}{1^Tx(t)}x(t) \to v$$

as $t \to \infty$, i.e., the distribution of $x(t)$ converges to $v$

we also have $x_i(t + 1)/x_i(t) \to \lambda_{pf}$, i.e., the one-period growth factor in each component always always converges to $\lambda_{pf}$
we consider an economy, with activity level $x_i \geq 0$ in sector $i$, $i = 1, \ldots, n$

given activity level $x$ in period $t$, in period $t + 1$ we have $x(t + 1) = Ax(t)$, with $A \geq 0$

$A_{ij} \geq 0$ means activity in sector $j$ does not decrease activity in sector $i$, i.e., the activities are mutually noninhibitory

we’ll assume that $A$ is regular, with PF eigenvalue $\lambda_{pf}$, and left and right PF eigenvectors $w, v$, with $1^Tv = 1$, $w^Tv = 1$

PF theorem tells us:

• $x_i(t + 1)/x_i(t)$, the growth factor in sector $i$ over the period from $t$ to $t + 1$, each converge to $\lambda_{pf}$ as $t \to \infty$

• the distribution of economic activity (i.e., $x$ normalized) converges to $v$
• asymptotically the economy exhibits (almost) balanced growth, by the factor \( \lambda_{pf} \), in each sector

these hold independent of the original economic activity, provided it is nonnegative and nonzero

what does left PF eigenvector \( w \) mean?

for large \( t \) we have

\[
x(t) \sim \lambda^t_{pf} w^T x(0)v
\]

where \( \sim \) means we have dropped terms small compared to dominant term

so asymptotic economic activity is scaled by \( w^T x(0) \)

in particular, \( w_i \) gives the relative value of activity \( i \) in terms of long term economic activity
Population model

\( x_i(t) \) denotes number of individuals in group \( i \) at period \( t \)

groups could be by age, location, health, marital status, etc.

population dynamics is given by \( x(t + 1) = Ax(t) \), with \( A \geq 0 \)

\( A_{ij} \) gives the fraction of members of group \( j \) that move to group \( i \), or the number of members in group \( i \) created by members of group \( j \) (e.g., in births)

\( A_{ij} \geq 0 \) means the more we have in group \( j \) in a period, the more we have in group \( i \) in the next period

- if \( \sum_i A_{ij} = 1 \), population is preserved in transitions out of group \( j \)
- we can have \( \sum_i A_{ij} > 1 \), if there are births (say) from members of group \( j \)
- we can have \( \sum_i A_{ij} < 1 \), if there are deaths or attrition in group \( j \)
now suppose $A$ is regular

- PF eigenvector $v$ gives asymptotic population distribution
- PF eigenvalue $\lambda_{pf}$ gives asymptotic growth rate (if $> 1$) or decay rate (if $< 1$)
- $w^T x(0)$ scales asymptotic population, so $w_i$ gives relative value of initial group $i$ to long term population
Path count in directed graph

we have directed graph on $n$ nodes, with adjacency matrix $A \in \mathbb{R}^{n \times n}$

$$A_{ij} = \begin{cases} 
1 & \text{there is an edge from node } j \text{ to node } i \\
0 & \text{otherwise}
\end{cases}$$

$(A^k)_{ij}$ is number of paths from $j$ to $i$ of length $k$

now suppose $A$ is regular

then for large $k$,

$$A^k \sim \lambda_{\text{pf}}^k vw^T = \lambda_{\text{pf}}^k (1^T w) v (w / 1^T w)^T$$

($\sim$ means: keep only dominant term)

$v, w$ are right, left PF eigenvectors, normalized as $1^T v = 1, w^T v = 1$
total number of paths of length $k$: $1^T A^k 1 \approx \lambda_{pf}^k (1^T w)$

for $k$ large, we have (approximately)

- $\lambda_{pf}$ is factor of increase in number of paths when length increases by one
- $v_i$: fraction of length $k$ paths that end at $i$
- $w_j / 1^T w$: fraction of length $k$ paths that start at $j$
- $v_i w_j / 1^T w$: fraction of length $k$ paths that start at $j$, end at $i$

- $v_i$ measures importance/connectedness of node $i$ as a **sink**
- $w_j / 1^T w$ measures importance/connectedness of node $j$ as a **source**
(Part of) proof of PF theorem for positive matrices

suppose $A > 0$, and consider the optimization problem

\[
\text{maximize } \delta \\
\text{subject to } Ax \geq \delta x \text{ for some } x \geq 0, \quad x \neq 0
\]

note that we can assume $1^T x = 1$

*interpretation:* with $y_i = (Ax)_i$, we can interpret $y_i/x_i$ as the ‘growth factor’ for component $i$

problem above is to find the input distribution that maximizes the minimum growth factor

let $\lambda_0$ be the optimal value of this problem, and let $v$ be an optimal point, i.e., $v \geq 0$, $v \neq 0$, and $Av \geq \lambda_0 v$
we will show that $\lambda_0$ is the PF eigenvalue of $A$, and $v$ is a PF eigenvector

first let’s show $Av = \lambda_0 v$, i.e., $v$ is an eigenvector associated with $\lambda_0$

if not, suppose that $(Av)_k > \lambda_0 v_k$

now let’s look at $\tilde{v} = v + \epsilon e_k$

we’ll show that for small $\epsilon > 0$, we have $A\tilde{v} > \lambda_0 \tilde{v}$, which means that $A\tilde{v} \geq \delta \tilde{v}$ for some $\delta > \lambda_0$, a contradiction

for $i \neq k$ we have

$$(A\tilde{v})_i = (Av)_i + A_{ik}\epsilon > (Av)_i \geq \lambda_0 v_i = \lambda_0 \tilde{v}_i$$

so for any $\epsilon > 0$ we have $(A\tilde{v})_i > \lambda_0 \tilde{v}_i$

$$(A\tilde{v})_k - \lambda_0 \tilde{v}_k = (Av)_k + A_{kk}\epsilon - \lambda_0 v_k - \lambda_0 \epsilon$$

$$= (Av)_k - \lambda_0 v_k - \epsilon(\lambda_0 - A_{kk})$$
since \((Av)_k - \lambda_0 v_k > 0\), we conclude that for small \(\epsilon > 0\),
\((A\tilde{v})_k - \lambda_0 \tilde{v}_k > 0\)

to show that \(v > 0\), suppose that \(v_k = 0\)

from \(Av = \lambda_0 v\), we conclude \((Av)_k = 0\), which contradicts \(Av > 0\)
(which follows from \(A > 0\), \(v \geq 0\), \(v \neq 0\))

now suppose \(\lambda \neq \lambda_0\) is another eigenvalue of \(A\), \(i.e., Az = \lambda z\), where \(z \neq 0\)

let \(|z|\) denote the vector with \(|z|_i = |z_i|\)

since \(A \geq 0\) we have \(A|z| \geq |Az| = |\lambda||z|\)

from the definition of \(\lambda_0\) we conclude \(|\lambda| \leq \lambda_0\)

(to show strict inequality is harder)
Max-min ratio characterization

proof shows that PF eigenvalue is optimal value of optimization problem

\[
\begin{align*}
\text{maximize} & \quad \min_i \frac{(Ax)_i}{x_i} \\
\text{subject to} & \quad x > 0
\end{align*}
\]

and that PF eigenvector \( v \) is optimal point:

- PF eigenvector \( v \) maximizes the minimum growth factor over components
- with optimal \( v \), growth factors in all components are equal (to \( \lambda_{pf} \))

in other words: by maximizing minimum growth factor, we actually achieve balanced growth
Min-max ratio characterization

A related problem is

\[
\text{minimize } \max_i \frac{(Ax)_i}{x_i} \\
\text{subject to } x > 0
\]

Here we seek to minimize the maximum growth factor in the coordinates.

The solution is surprising: the optimal value is \( \lambda_{pf} \) and the optimal \( x \) is the PF eigenvector \( v \).

• If \( A \) is nonnegative and regular, and \( x > 0 \), the \( n \) growth factors \( (Ax)_i/x_i \) ‘straddle’ \( \lambda_{pf} \): at least one is \( \geq \lambda_{pf} \), and at least one is \( \leq \lambda_{pf} \).

• When we take \( x \) to be the PF eigenvector \( v \), all the growth factors are equal, and solve both max-min and min-max problems.
Power control

we consider \( n \) transmitters with powers \( P_1, \ldots, P_n > 0 \), transmitting to \( n \) receivers

path gain from transmitter \( j \) to receiver \( i \) is \( G_{ij} > 0 \)

signal power at receiver \( i \) is \( S_i = G_{ii} P_i \)

interference power at receiver \( i \) is \( I_i = \sum_{k \neq i} G_{ik} P_k \)

signal to interference ratio (SIR) is

\[
\frac{S_i}{I_i} = \frac{G_{ii} P_i}{\sum_{k \neq i} G_{ik} P_k}
\]

how do we set transmitter powers to maximize the minimum SIR?
we can just as well minimize the maximum interference to signal ratio, \( i.e. \), solve the problem

\[
\begin{align*}
\text{minimize} & \quad \max_i \frac{(\tilde{G}P)_i}{P_i} \\
\text{subject to} & \quad P > 0
\end{align*}
\]

where

\[
\tilde{G}_{ij} = \begin{cases} 
G_{ij}/G_{ii} & i \neq j \\
0 & i = j
\end{cases}
\]

since \( \tilde{G}^2 > 0 \), \( \tilde{G} \) is regular, so solution is given by PF eigenvector of \( \tilde{G} \)

PF eigenvalue \( \lambda_{pf} \) of \( \tilde{G} \) is the optimal interference to signal ratio, \( i.e. \), maximum possible minimum SIR is \( 1/\lambda_{pf} \)

with optimal power allocation, all SIRs are equal

note: \( \tilde{G} \) is the matrix of ratios of interference to signal path gains
Nonnegativity of resolvent

suppose $A$ is nonnegative, with PF eigenvalue $\lambda_{pf}$, and $\lambda \in \mathbb{R}$
then $(\lambda I - A)^{-1}$ exists and is nonnegative, if and only if $\lambda > \lambda_{pf}$
for any square matrix $A$ the power series expansion

$$(\lambda I - A)^{-1} = \frac{1}{\lambda} I + \frac{1}{\lambda^2} A + \frac{1}{\lambda^3} A^2 + \cdots$$

converges provided $|\lambda|$ is larger than all eigenvalues of $A$
if $\lambda > \lambda_{pf}$, this shows that $(\lambda I - A)^{-1}$ is nonnegative

to show converse, suppose $(\lambda I - A)^{-1}$ exists and is nonnegative, and let $v \neq 0, v \geq 0$ be a PF eigenvector of $A$
then we have

$$(\lambda I - A)^{-1} v = \frac{1}{\lambda - \lambda_{pf}} v \geq 0$$
and it follows that $\lambda > \lambda_{pf}$
Equilibrium points

consider \( x(t + 1) = Ax(t) + b \), where \( A \) and \( b \) are nonnegative

equilibrium point is given by \( x_{eq} = (I - A)^{-1}b \)

by resolvent result, if \( A \) is stable, then \( (I - A)^{-1} \) is nonnegative, so
equilibrium point \( x_{eq} \) is nonnegative for any nonnegative \( b \)

moreover, equilibrium point is monotonic function of \( b \): for \( \tilde{b} \geq b \), we have
\( \tilde{x}_{eq} \geq x_{eq} \)

conversely, if system has a nonnegative equilibrium point, for every
nonnegative choice of \( b \), then we can conclude \( A \) is stable
Iterative power allocation algorithm

we consider again the power control problem

suppose $\gamma$ is the desired or target SIR

de simple iterative algorithm: at each step $t$,

1. first choose $\tilde{P}_i$ so that

\[
\frac{G_{ii}\tilde{P}_i}{\sum_{k \neq i} G_{ik}P_k(t)} = \gamma
\]

$\tilde{P}_i$ is the transmit power that would make the SIR of receiver $i$ equal to $\gamma$, assuming none of the other powers change

2. set $P_i(t + 1) = \tilde{P}_i + \sigma_i$, where $\sigma_i > 0$ is a parameter i.e., add a little extra power to each transmitter)
each receiver only needs to know its current SIR to adjust its power: if current SIR is $\alpha$ dB below (above) $\gamma$, then increase (decrease) transmitter power by $\alpha$ dB, then add the extra power $\sigma$

i.e., this is a *distributed algorithm*

*question:* does it work? (we assume that $P(0) > 0$)

*answer:* yes, if and only if $\gamma$ is less than the maximum achievable SIR, i.e.,

$$\gamma < 1/\lambda_{pf}(\tilde{G})$$

to see this, algorithm can be expressed as follows:

- in the first step, we have $\tilde{P} = \gamma \tilde{G} P(t)$
- in the second step we have $P(t+1) = \tilde{P} + \sigma$

and so we have

$$P(t+1) = \gamma \tilde{G} P(t) + \sigma$$

a linear system with constant input
PF eigenvalue of $\gamma \tilde{G}$ is $\gamma \lambda_{pf}$, so linear system is stable if and only if $\gamma \lambda_{pf} < 1$

power converges to equilibrium value

$$P_{eq} = (I - \gamma \tilde{G})^{-1} \sigma$$

(which is positive, by resolvent result)

now let's show this equilibrium power allocation achieves SIR at least $\gamma$ for each receiver

we need to verify $\gamma \tilde{G} P_{eq} \leq P_{eq}$, i.e.,

$$\gamma \tilde{G} (I - \gamma \tilde{G})^{-1} \sigma \leq (I - \gamma \tilde{G})^{-1} \sigma$$

or, equivalently,

$$(I - \gamma \tilde{G})^{-1} \sigma - \gamma \tilde{G} (I - \gamma \tilde{G})^{-1} \sigma \geq 0$$

which holds, since the lefthand side is just $\sigma$
Linear Lyapunov functions

suppose $A \geq 0$

then $\mathbb{R}_+^n$ is invariant under system $x(t + 1) = Ax(t)$

suppose $c > 0$, and consider the linear Lyapunov function $V(z) = c^T z$

if $V(Az) \leq \delta V(z)$ for some $\delta < 1$ and all $z \geq 0$, then $V$ proves (nonnegative) trajectories converge to zero

**fact:** a nonnegative regular system is stable if and only if there is a linear Lyapunov function that proves it

to show the ‘only if’ part, suppose $A$ is stable, i.e., $\lambda_{pf} < 1$

take $c = w$, the (positive) left PF eigenvector of $A$

then we have $V(Az) = w^T Az = \lambda_{pf} w^T z$, i.e., $V$ proves all nonnegative trajectories converge to zero
Weighted $\ell_1$-norm Lyapunov function

to make the analysis apply to all trajectories, we can consider the weighted sum absolute value (or weighted $\ell_1$-norm) Lyapunov function

$$V(z) = \sum_{i=1}^{n} w_i |z_i| = w^T |z|$$

then we have

$$V(Az) = \sum_{i=1}^{n} w_i |(Az)_i| \leq \sum_{i=1}^{n} w_i (A|z|)_i = w^T A|z| = \lambda_{pf} w^T |z|$$

which shows that $V$ decreases at least by the factor $\lambda_{pf}$

conclusion: a nonnegative regular system is stable if and only if there is a weighted sum absolute value Lyapunov function that proves it
SVD analysis

suppose \( A \in \mathbb{R}^{m \times n}, A \geq 0 \)

then \( A^T A \geq 0 \) and \( AA^T \geq 0 \) are nonnegative

hence, there are nonnegative left & right singular vectors \( v_1, w_1 \) associated with \( \sigma_1 \)

in particular, there is an optimal rank-1 approximation of \( A \) that is nonnegative

if \( A^T A, AA^T \) are regular, then we conclude

- \( \sigma_1 > \sigma_2 \), i.e., maximum singular value is isolated
- associated singular vectors are positive: \( v_1 > 0, w_1 > 0 \)
Continuous time results

we have already seen that $\mathbb{R}^n_+$ is invariant under $\dot{x} = Ax$ if and only if $A_{ij} \geq 0$ for $i \neq j$

such matrices are called *Metzler matrices*

for a Metzler matrix, we have

- there is an eigenvalue $\lambda_{\text{metzler}}$ of $A$ that is real, with associated nonnegative left and right eigenvectors
- for any other eigenvalue $\lambda$ of $A$, we have $\Re \lambda \leq \lambda_{\text{metzler}}$
  
  *i.e.*, the eigenvalue $\lambda_{\text{metzler}}$ is dominant for system $\dot{x} = Ax$
- if $\lambda > \lambda_{\text{metzler}}$, then $(\lambda I - A)^{-1} \geq 0$
the analog of the stronger Perron-Frobenius results:

if \((\tau I + A)^k > 0\), for some \(\tau\) and some \(k\), then

- the left and right eigenvectors associated with eigenvalue \(\lambda_{\text{metzler}}\) of \(A\) are positive
- for any other eigenvalue \(\lambda\) of \(A\), we have \(\Re \lambda < \lambda_{\text{metzler}}\)

\(\text{i.e.},\) the eigenvalue \(\lambda_{\text{metzler}}\) is strictly dominant for system \(\dot{x} = Ax\)
Derivation from Perron-Frobenius Theory

suppose $A$ is Metzler, and choose $\tau$ s.t. $\tau I + A \geq 0$

(e.g., $\tau = 1 - \min_i A_{ii}$)

by PF theory, $\tau I + A$ has PF eigenvalue $\lambda_{pf}$, with associated right and left eigenvectors $v \geq 0$, $w \geq 0$

from $(\tau I + A)v = \lambda_{pf}v$ we get $Av = (\lambda_{pf} - \tau)v = \lambda_0v$, and similarly for $w$

we’ll show that $\Re \lambda \leq \lambda_0$ for any eigenvalue $\lambda$ of $A$

suppose $\lambda$ is an eigenvalue of $A$

suppose $\tau + \lambda$ is an eigenvalue of $\tau I + A$

by PF theory, we have $|\tau + \lambda| \leq \lambda_{pf} = \tau + \lambda_0$

this means $\lambda$ lies inside a circle, centered at $-\tau$, that passes through $\lambda_0$

which implies $\Re \lambda \leq \lambda_0$
Linear Lyapunov function

suppose \( \dot{x} = Ax \) is stable, and \( A \) is Metzler, with \( (\tau I + A)^k > 0 \) for some \( \tau \) and some \( k \)

we can show that all nonnegative trajectories converge to zero using a linear Lyapunov function

let \( w > 0 \) be left eigenvector associated with dominant eigenvalue \( \lambda_{\text{metzler}} \)

then with \( V(z) = w^T z \) we have

\[
\dot{V}(z) = w^T Az = \lambda_{\text{metzler}} w^T z = \lambda_{\text{metzler}} V(z)
\]

since \( \lambda_{\text{metzler}} < 0 \), this proves \( w^T z \rightarrow 0 \)