Lecture 10
Linear Quadratic Stochastic Control with Partial State Observation

- partially observed linear-quadratic stochastic control problem
- estimation-control separation principle
- solution via dynamic programming
Linear stochastic system

- linear dynamical system, over finite time horizon:

\[ x_{t+1} = Ax_t + Bu_t + w_t, \quad t = 0, \ldots, N - 1 \]

with state \( x_t \), input \( u_t \), and process noise \( w_t \)

- linear noise corrupted observations:

\[ y_t = Cx_t + v_t, \quad t = 0, \ldots, N \]

\( y_t \) is output, \( v_t \) is measurement noise

- \( x_0 \sim \mathcal{N}(0, X), w_t \sim \mathcal{N}(0, W), v_t \sim \mathcal{N}(0, V) \), all independent
Causal output feedback control policies

- causal feedback policies:
  - input must be function of past and present outputs
  - roughly speaking: current state $x_t$ is *not known*

- $u_t = \phi_t(Y_t), \quad t = 0, \ldots, N - 1$
  - $Y_t = (y_0, \ldots, y_t)$ is output history at time $t$
  - $\phi_t : \mathbb{R}^{p(t+1)} \to \mathbb{R}^m$ called the control policy at time $t$

- closed-loop system is

\[
x_{t+1} = Ax_t + B\phi_t(Y_t) + w_t, \quad y_t = Cx_t + v_t
\]

- $x_0, \ldots, x_N, y_0, \ldots, y_N, u_0, \ldots, u_{N-1}$ are all random
Stochastic control with partial observations

- objective:

\[ J = \mathbb{E} \left( \sum_{t=0}^{N-1} (x_t^T Q x_t + u_t^T R u_t) + x_N^T Q x_N \right) \]

with \( Q \geq 0, \ R > 0 \)

- partially observed linear quadratic stochastic control problem (a.k.a. LQG problem):

  choose output feedback policies \( \phi_0, \ldots, \phi_{N-1} \) to minimize \( J \)
Solution

• optimal policies are $\phi_t(Y_t) = K_t \mathbb{E}(x_t|Y_t)$
  
  – $K_t$ is optimal feedback gain matrix for associated LQR problem
  – $\mathbb{E}(x_t|Y_t)$ is the MMSE estimate of $x_t$ given measurements $Y_t$
    (can be computed using Kalman filter)

• called *separation principle*: optimal policy consists of
  
  – estimating state via MMSE (ignoring the control problem)
  – using estimated state as if it were the actual state, for purposes of control
LQR control gain computation

- define $P_N = Q$, and for $t = N, \ldots, 1$,

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

- set $K_t = -(R + B^T P_{t+1} B)^{-1} B^T P_{t+1} A$, $t = 0, \ldots, N - 1$

- $K_t$ does not depend on data $C, X, W, V$
Kalman filter current state estimate

- define
  - $\hat{x}_t = \mathbb{E}(x_t|Y_t)$ (current state estimate)
  - $\Sigma_t = \mathbb{E}(x_t - \hat{x}_t)(x_t - \hat{x}_t)^T$ (current state estimate covariance)
  - $\Sigma_{t+1|t} = A\Sigma_tA^T + W$ (next state estimate covariance)

- start with $\Sigma_{0|-1} = X$; for $t = 0, \ldots, N$,
  
  $$
  \Sigma_t = \Sigma_{t|t-1} - \Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + V)^{-1}C\Sigma_{t|t-1},
  $$
  
  $$
  \Sigma_{t+1|t} = A\Sigma_tA^T + W
  $$

- define $L_t = \Sigma_{t|t-1}C^T(C\Sigma_{t|t-1}C^T + V)^{-1}$, $t = 0, \ldots, N$
• set $\hat{x}_0 = L_0 y_0$; for $t = 0, \ldots, N - 1$, 

$$\hat{x}_{t+1} = A\hat{x}_t + B u_t + L_{t+1} e_{t+1}, \quad e_{t+1} = y_{t+1} - C(A\hat{x}_t + B u_t)$$

- $e_{t+1}$ is next output prediction error
- $e_{t+1} \sim \mathcal{N}(0, C\Sigma_{t+1|t} C^T + V)$, independent of $Y_t$

• Kalman filter gains $L_t$ do not depend on data $B, Q, R$
Solution via dynamic programming

• let $V_t(Y_t)$ be optimal value of LQG problem, from $t$ on, conditioned on the output history $Y_t$:

$$V_t(Y_t) = \min_{\phi_t, \ldots, \phi_{N-1}} \mathbb{E} \left( \sum_{\tau=t}^{N-1} (x_\tau^T Q x_\tau + u_\tau^T R u_\tau) + x_N^T Q x_N \mid Y_t \right)$$

• we’ll show that $V_t$ is a quadratic function plus a constant, in fact,

$$V_t(Y_t) = \hat{x}_t^T P_t \hat{x}_t + q_t, \quad t = 0, \ldots, N,$$

where $P_t$ is the LQR cost-to-go matrix ($\hat{x}_t$ is a linear function of $Y_t$)
• we have

\[ V_N(Y_N) = \mathbb{E}(x_N^T Q x_N | Y_N) = \hat{x}_N^T Q \hat{x}_N + \text{Tr}(Q \Sigma_N) \]

(using \( x_N | Y_N \sim \mathcal{N}(\hat{x}_N, \Sigma_N) \)) so \( P_N = Q, \ q_N = \text{Tr}(Q \Sigma_N) \)

• dynamic programming (DP) equation is

\[ V_t(Y_t) = \min_{u_t} \mathbb{E} \left( x_t^T Q x_t + u_t^T R u_t + V_{t+1}(Y_{t+1}) | Y_t \right) \]

(and argmin, which is a function of \( Y_t \), is optimal input)

• with \( V_{t+1}(Y_{t+1}) = \hat{x}_{t+1}^T P_{t+1} \hat{x}_{t+1} + q_{t+1} \), DP equation becomes

\[
V_t(Y_t) = \min_{u_t} \mathbb{E} \left( x_t^T Q x_t + u_t^T R u_t + \hat{x}_{t+1}^T P_{t+1} \hat{x}_{t+1} + q_{t+1} | Y_t \right) \\
= \mathbb{E}(x_t^T Q x_t | Y_t) + q_{t+1} + \min_{u_t} \left( u_t^T R u_t + \mathbb{E}(\hat{x}_{t+1}^T P_{t+1} \hat{x}_{t+1} | Y_t) \right)
\]
using $x_t | Y_t \sim \mathcal{N}(\hat{x}_t, \Sigma_t)$, the first term is

$$E(x_t^T Q x_t | Y_t) = \hat{x}_t^T Q \hat{x}_t + \text{Tr}(Q \Sigma_t)$$

using

$$\hat{x}_{t+1} = A\hat{x}_t + Bu_t + L_{t+1} e_{t+1},$$

with $e_{t+1} \sim \mathcal{N}(0, C\Sigma_{t+1|t} C^T + V)$, independent of $Y_t$, we get

$$E(\hat{x}_{t+1}^T P_{t+1} \hat{x}_{t+1} | Y_t) = \hat{x}_t^T A^T P_{t+1} A \hat{x}_t + u_t^T B^T P_{t+1} Bu_t + 2\hat{x}_t^T A^T P_{t+1} Bu_t$$

$$+ \text{Tr} \left( (L_{t+1}^T P_{t+1} L_{t+1})(C\Sigma_{t+1|t} C^T + V) \right)$$

using $L_{t+1} = \Sigma_{t+1|t} C^T (C\Sigma_{t+1|t} C^T + V)^{-1}$, last term becomes

$$\text{Tr}(P_{t+1} \Sigma_{t+1|t} C^T (C\Sigma_{t+1|t} C^T + V)^{-1} C\Sigma_{t+1|t}) = \text{Tr} \ P_{t+1} (\Sigma_{t+1|t} - \Sigma_{t+1})$$
• combining all terms we get

\[
V_t(Y_t) = \hat{x}_t^T (Q + A^T P_{t+1} A) \hat{x}_t + q_{t+1} + \text{Tr}(Q \Sigma_t) \\
+ \text{Tr} P_{t+1} (\Sigma_{t+1|t} - \Sigma_{t+1}) \\
+ \min_{u_t} (u_t^T (R + B^T P_{t+1} B) u_t + 2 \hat{x}_t A^T P_{t+1} B u_t)
\]

• minimization same as in deterministic LQR problem
• thus optimal policy is \( \phi_t^*(Y_t) = K_t \hat{x}_t \), with

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

• plugging in optimal \( u_t \) we get \( V_t(Y_t) = \hat{x}_t^T P_t \hat{x}_t + q_t \), where

\[
P_t = A^T P_{t+1} A + Q - A^T P_{t+1} B (R + B^T P_{t+1} B)^{-1} B^T P_{t+1} A \\
q_t = q_{t+1} + \text{Tr}(Q \Sigma_t) + \text{Tr} P_{t+1} (\Sigma_{t+1|t} - \Sigma_{t+1})
\]

• recursion for \( P_t \) is exactly the same as for deterministic LQR
Optimal objective

• optimal LQG cost is

\[
J^* = \mathbb{E} V_0(y_0) = q_0 + \mathbb{E} \hat{x}_0^T P_0 \hat{x}_0 = q_0 + \text{Tr} P_0 (X - \Sigma_0)
\]

using \( \hat{x}_0 \sim \mathcal{N}(0, X - \Sigma_0) \)

• using \( q_N = \text{Tr} Q \Sigma_N \) and

\[
q_t = q_{t+1} + \text{Tr}(Q \Sigma_t) + \text{Tr} P_{t+1}(\Sigma_{t+1}|t - \Sigma_{t+1})
\]

we get

\[
J^* = \sum_{t=0}^{N} \text{Tr}(Q \Sigma_t) + \sum_{t=0}^{N} \text{Tr} P_t(\Sigma_{t|t-1} - \Sigma_t)
\]

using \( \Sigma_{0|0} = X \)
we can write this as

$$J^* = \sum_{t=0}^{N} \text{Tr}(Q\Sigma_t) + \sum_{t=1}^{N} \text{Tr} P_t(A\Sigma_{t-1}A^T + W - \Sigma_t) + \text{Tr}(P_0(X - \Sigma_0))$$

which simplifies to

$$J^* = J_{lqr} + J_{est}$$

where

$$J_{lqr} = \text{Tr}(P_0X) + \sum_{t=1}^{N} \text{Tr}(P_tW),$$

$$J_{est} = \text{Tr}((Q - P_0)\Sigma_0) + \sum_{t=1}^{N} \text{Tr}((Q - P_t)\Sigma_t) + \text{Tr}(P_tA\Sigma_{t-1}A^T)$$

– $J_{lqr}$ is the stochastic LQR cost, i.e., the optimal objective if you knew the state
– $J_{est}$ is the cost of not knowing (i.e., estimating) the state
• when state measurements are exact \((C = I, V = 0)\), we have \(\Sigma_t = 0\), so we get

\[
J^* = J_{lqr} = \text{Tr}(P_0X) + \sum_{t=1}^{N} \text{Tr}(P_tW)
\]
Infinite horizon LQG

• choose policies to minimize infinite horizon average stage cost

\[ J = \lim_{N \to \infty} \frac{1}{N} \mathbb{E} \sum_{t=0}^{N-1} (x_t^T Q x_t + u_t^T R u_t) \]

• optimal average stage cost is

\[ J^* = \text{Tr}(Q \Sigma) + \text{Tr}(P(\tilde{\Sigma} - \Sigma)) \]

where \( P \) and \( \tilde{\Sigma} \) are PSD solutions of AREs

\[
\begin{align*}
P &= Q + A^T PA - A^T PB (R + B^T PB)^{-1} B^T PA, \\
\tilde{\Sigma} &= A \tilde{\Sigma} A^T + W - A \tilde{\Sigma} C^T (C \tilde{\Sigma} C^T + V)^{-1} C \tilde{\Sigma} A^T
\end{align*}
\]

and \( \Sigma = \tilde{\Sigma} - \tilde{\Sigma} C^T (C \tilde{\Sigma} C^T + V)^{-1} C \tilde{\Sigma} \)
• optimal average stage cost doesn’t depend on $X$

• (an) optimal policy is

$$u_t = K\hat{x}_t, \quad \hat{x}_{t+1} = A\hat{x}_t + Bu_t + L(y_{t+1} - C(A\hat{x}_t + Bu_t))$$

where

$$K = -(R + B^T PB)^{-1}B^TPA, \quad L = \tilde{\Sigma}C^T(C\tilde{\Sigma}C^T + V)^{-1}$$

• $K$ is steady-state LQR feedback gain

• $L$ is steady-state Kalman filter gain
Example

• system with $n = 5$ states, $m = 2$ inputs, $p = 3$ outputs; infinite horizon

• $A, B, C$ chosen randomly; $A$ scaled so $\max_i |\lambda_i(A)| = 1$

• $Q = I$, $R = I$, $X = I$, $W = 0.5I$, $V = 0.5I$

• we compare LQG with the case where state is known (stochastic LQR)
Sample trajectories

sample trace of \((x_t)_1\) and \((u_t)_1\) in steady state

blue: LQG, red: stochastic LQR
Cost histogram

histogram of stage costs for 5000 steps in steady state