Linear Quadratic Stochastic Control

- linear-quadratic stochastic control problem
- solution via dynamic programming

Linear stochastic system

- linear dynamical system, over finite time horizon:

$$x_{t+1} = Ax_t + Bu_t + w_t, \qquad t = 0, \ldots, N-1$$

- w_t is the process noise or disturbance at time t
- w_t are IID with $\mathbf{E}\, w_t = 0$, $\mathbf{E}\, w_t w_t^T = W$
- x_0 is independent of w_t, with $\mathbf{E}\, x_0 = 0$, $\mathbf{E}\, x_0 x_0^T = X$

Control policies

- state-feedback control: $u_t = \phi_t(x_t), \quad t = 0, \ldots, N-1$
- $\phi_t : \mathbb{R}^n \to \mathbb{R}^m$ called the control **policy** at time t
- roughly speaking: we choose input *after* knowing the current state, but *before* knowing the disturbance
- closed-loop system is
$$x_{t+1} = Ax_t + B\phi_t(x_t) + w_t, \qquad t = 0, \ldots, N-1$$
- $x_0, \ldots, x_N, \ u_0, \ldots, u_{N-1}$ are random

Stochastic control problem

- objective:

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

with Q, $Q_f \geq 0$, $R > 0$

- J depends (in complex way) on control policies $\phi_0, \ldots, \phi_{N-1}$
- linear-quadratic **stochastic control problem**: choose control policies $\phi_0, \ldots, \phi_{N-1}$ to minimize J

  ('linear' refers to the state dynamics; 'quadratic' to the objective)
- an infinite dimensional problem: variables are *functions* $\phi_0, \ldots, \phi_{N-1}$

Solution via dynamic programming

- let $V_t(z)$ be optimal value of objective, from t on, starting at $x_t = z$

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

  subject to $x_{\tau+1} = Ax_\tau + Bu_\tau + w_\tau$, $u_\tau = \phi_\tau(x_\tau)$

- we have
  - $V_N(z) = z^T Q_f z$
  - $J^\star = \mathbf{E}\, V_0(x_0)$ (expectation over x_0)

- V_t can be found by backward recursion: for $t = N-1, \ldots, 0$

$$V_t(z) = z^\mathsf{T} Q z + \min_v \left\{ v^\mathsf{T} R v + \mathbf{E}\, V_{t+1}(Az + Bv + w_t) \right\}$$

- expectation is over w_t

- we do not know where we will land, when we take $u_t = v$

- optimal policies have form

$$\phi_t^\star(x_t) = \underset{v}{\operatorname{argmin}} \left\{ v^\mathsf{T} R v + \mathbf{E}\, V_{t+1}(Ax_t + Bv + w_t) \right\}$$

Explicit form

- let's show (via recursion) value functions are quadratic, with form

$$V_t(x_t) = x_t^T P_t x_t + q_t, \quad t = 0, \ldots, N,$$

  with $P_t \geq 0$

- $P_N = Q_N$, $q_N = 0$
- now assume that $V_{t+1}(z) = z^T P_{t+1} z + q_{t+1}$
- Bellman recursion is

$$\begin{aligned} V_t(z) &= z^T Q z + \min_v \{v^T R v + \mathbf{E}((Az + Bv + w_t)^T P_{t+1}(Az + Bv + w_t) + q_{t+1})\} \\ &= z^T Q z + \mathbf{trace}(W P_{t+1}) + q_{t+1} + \min_v \{v^T R v + (Az + Bv)^T P_{t+1}(Az + Bv)\} \end{aligned}$$

  - we use $\mathbf{E}(w_t^T P_{t+1} w_t) = \mathbf{trace}(W P_{t+1})$
  - same recursion as deterministic LQR, with added constant

- optimal policy is linear state feedback: $\phi_t^\star(x_t) = K_t x_t$,

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

  (same form as in deterministic LQR)

- plugging in optimal w gives $V_t(z) = z^\mathsf{T} P_t z + q_t$, with

$$\begin{aligned} P_t &= A^\mathsf{T} P_{t+1} A - A^\mathsf{T} P_{t+1} B (B^\mathsf{T} P_{t+1} B + R)^{-1} B^\mathsf{T} P_{t+1} A + Q \\ q_t &= q_{t+1} + \mathbf{trace}(W P_{t+1}) \end{aligned}$$

  - first recursion same as for deterministic LQR
  - second term is just a running sum

- we conclude that
  - P_t, K_t are *same* as in deterministic LQR
  - strangely, optimal policy is same as LQR, and independent of X, W

- optimal cost is

$$
\begin{aligned}
J^\star &= \mathbf{E}\, V_0(x_0) \\
&= \textbf{trace}(X P_0) + q_0 \\
&= \textbf{trace}(X P_0) + \sum_{t=1}^{N} \textbf{trace}(W P_t)
\end{aligned}
$$

- interpretation:
  - $x_0^T P_0 x_0$ is optimal cost of deterministic LQR, with $w_0 = \cdots = w_{N-1} = 0$
  - $\textbf{trace}(X P_0)$ is average optimal LQR cost, with $w_0 = \cdots = w_{N-1} = 0$
  - $\textbf{trace}(W P_t)$ is average optimal LQR cost, for $\mathbf{E}\, x_t = 0$, $\mathbf{E}\, x_t x_t^T = W$, $w_t = \cdots = w_{N-1} = 0$

Infinite horizon

- choose policies to minimize average stage cost

$$J = \lim_{N\to\infty} \frac{1}{N} \mathbf{E} \sum_{t=0}^{N-1} \left(x_t^T Q x_t + u_t^T R u_t\right)$$

- optimal average stage cost is

$$J^\star = \mathbf{trace}(W P_{\text{ss}})$$

  where P_{ss} satisfies the ARE

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

  - optimal average stage cost doesn't depend on X

- (an) optimal policy is constant linear state feedback

$$u_t = K_{\text{ss}} x_t$$

  where

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

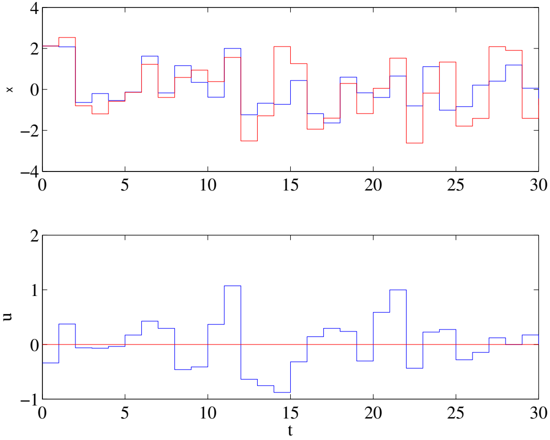
  - K_{ss} is steady-state LQR feedback gain
  - doesn't depend on X, W

Example

- system with $n = 5$ states, $m = 2$ inputs, horizon $N = 30$
- A, B chosen randomly; A scaled so $\mathbf{max}_i |\lambda_i(A)| = 1$
- $Q = I$, $Q_f = 10I$, $R = I$
- $x_0 \sim \mathcal{N}(0, X)$, $X = 10I$
- $w_t \sim \mathcal{N}(0, W)$, $W = 0.5I$

Sample trajectories

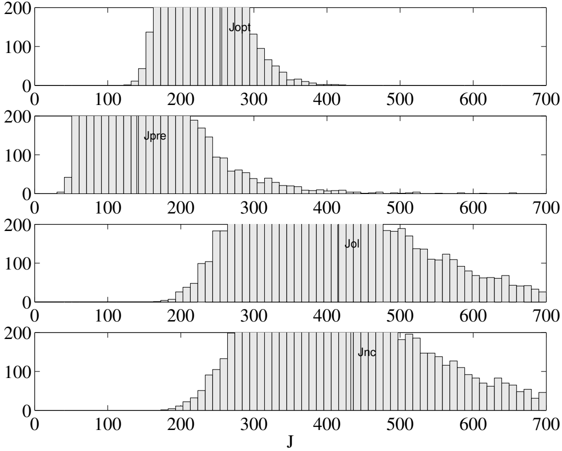
sample trace of $(x_t)_1$ and $(u_t)_1$



blue: optimal stochastic control, red: no control $(u_0 = \cdots = u_{N-1} = 0)$

Cost histogram

cost histogram for 1000 simulations



Comparisons

we compared optimal stochastic control ($J^\star = 224.2$) with

- 'prescient' control
  - decide input sequence with full knowledge of future disturbances
  - $u_0, \dots, u_{N-1}$ computed assuming *all* w_t are known
  - $J^{\rm pre} = 137.6$
- 'open-loop' control
  - $u_0, \dots, u_{N-1}$ depend only on x_0
  - $u_0, \dots, u_{N-1}$ computed assuming $w_0 = \cdots = w_{N-1} = 0$
  - $J^{\rm ol} = 423.7$
- no control
  - $u_0 = \cdots = u_{N-1} = 0$
  - $J^{\rm nc} = 442.0$