Linear Systems, Estimation, and Control

Linear quadratic regulator and linear quadratic Gaussian control

Aditya Mahajan
McGill University

Lecture Notes for ECSE 506: Stochastic Control and Decision Theory
March 24, 2016

Centralized Linear Quadratic Control

The MDP setup — State feedback

The POMDP setup — Output feedback

The MDP setup – State feedback

Linear Quadratic Regulation (LQR)

Notation

State : $X_t \in \mathbb{R}^n$

Action: $U_t \in \mathbb{R}^m$

Dynamics

$X_{t+1} = \mathbf{A}_t X_t + \mathbf{B}_t U_t$, where $\mathbf{A}_t \in \mathbb{R}^{n \times n}$, $\mathbf{B}_t \in \mathbb{R}^{n \times m}$.

Cost

Per step cost : $c_t(x_t, u_t) = \|x_t\|^2_{\mathbf{Q}_t} + \|u_t\|^2_{\mathbf{R}_t}$

Terminal reward: $c_T(x_T) = \|x_T\|^2_{\mathbf{Q}_T}$

where $\|X\|_{\mathbf{Q}} = X^\top \mathbf{Q} X$ and for all t, $\mathbf{Q}_t = \mathbf{Q}_t^\top \succeq 0$ and $\mathbf{R}_t = \mathbf{R}_t^\top \succ 0$.

Control objective

Choose $U_t = g_t(X_{1:t}, U_{1:t-1})$ so as to minimize

$$J(\mathbf{g}) = \sum_{t=1}^{T-1} c_t(X_t, U_t) + c_T(x_T)$$

- Regulation problem: keep the state of the system close to the origin.
- Tracking problem: To keep the state of the system close to a reference trajectory $\{x^\circ_t\}$, use the cost

$$c_t(x_t, u_t) = \|x_t - x^\circ_t\|_{\mathbf{Q}_t} + \|u_t\|_{\mathbf{R}_t}, \qquad c_T(x_T) = \|x_T - x^\circ_T\|_{\mathbf{Q}_T}.$$

Deterministic LQR is a MDP

	MDP Dynamic Model	Deterministic LQR
System Dynamics	$X_{t+1} = f_t(X_t, U_t, W_t)$	$X_{t+1} = \mathbf{A}_t X_t + \mathbf{B}_t U_t$
Information Structure	$U_t = g_t(X_{1:t}, U_{1:t-1})$	$U_t = g_t(X_{1:t}, U_{1:t-1})$
Objective Function	$\mathbb{E}\left[\sum_{t=1}^{T-1} c_t(X_t, U_t) + c_T(X_T)\right]$	$\sum_{t=1}^{T-1} c_t(X_t, U_t) + c_T(X_T)$

Structure of Controller

Using Markov strategies does not entail any loss of optimality

$$U_t = g_t(X_t)$$

Dynamic program

$$V_T(x_T) = c_T(x_T);$$

$$V_t(x_t) = \max_{u_t \in \mathcal{U}_t(x_t)} \Big\{ c_t(x_t, u_t) + \mathbb{E}[V_{t+1}(f_t(x_t, u_t, W_t))] \Big\},$$

$$t = T-1, \dots, 1.$$

Structure of optimal deterministic LQR

Theorem The value function at time t is

$$V_t(X_t) = \|X_t\|^2_{\mathbf{S}_t}$$

and the optimal control action is

$$U_t = -\mathbf{H}_t X_t$$

where the gain matrices $\mathbf{H}_t$ are determined recursively as follows:

$$\mathbf{H}_T = 0$$

$$\mathbf{H}_t = [\mathbf{R}_t + \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{B}_t]^{-1} \boldsymbol{\Lambda}_t$$

where

$$\boldsymbol{\Lambda}_t = \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{A}_t$$

and $\mathbf{S}_t$ are determined by the backward Riccati difference equations:

$$\mathbf{S}_T = \mathbf{Q}_T$$

$$\mathbf{S}_t = \mathbf{A}_t^\top \mathbf{S}_{t+1} \mathbf{A}_t + \mathbf{Q}_t - \boldsymbol{\Lambda}_t^\top [\mathbf{R}_t + \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{B}_t]^{-1} \boldsymbol{\Lambda}_t$$

Riccati equations are named after Count Jacopo Francesco Riccati (1670--1754) who studied the differential equations of the form

$$\dot{x} = ax^2 + bt + ct^2$$

and its variations. In modern control, such equations arise in the calculus of variations and optimal filtering. The discrete-time version of these equations are also named after Riccati.

Completing the squares lemma

Lemma Let

- $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$
- $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times m}$
- $\mathbf{R} \in \mathbb{R}^{m \times m}$ and $\mathbf{Q} \in \mathbb{R}^{n \times n}$ such that $\mathbf{R} = \mathbf{R}^\top \succ 0$ and $\mathbf{Q} = \mathbf{Q}^\top \succeq 0$.

Then

$$\|u\|_{\mathbf{R}}^2 + \|\mathbf{A}x + \mathbf{B}u\|_{\mathbf{Q}}^2 = \|u + \mathbf{H}x\|_{\mathbf{K}}^2 + \|x\|_{\mathbf{L}}^2.$$

where

$$\mathbf{K} = \mathbf{R} + \mathbf{B}^\top \mathbf{Q}\mathbf{B}, \qquad \mathbf{K} = \mathbf{K}^\top \succ 0$$

$$\mathbf{H} = \mathbf{K}^{-1}\Lambda, \qquad \text{where } \Lambda = \mathbf{B}^\top \mathbf{Q}\mathbf{A}$$

$$\mathbf{L} = \mathbf{A}^\top \mathbf{Q}\mathbf{A} - \Lambda^\top \mathbf{K}\Lambda$$

Proof The result follows by completing the squares in two different ways.

$$\text{LHS} = u^\top \mathbf{R}u + u^\top \mathbf{B}^\top \mathbf{Q}\mathbf{B}u + x^\top \mathbf{A}^\top \mathbf{Q}\mathbf{A}x + x^\top \mathbf{A}^\top \mathbf{Q}\mathbf{B}u + u^\top \mathbf{B}^\top \mathbf{Q}\mathbf{A}x$$

$$\text{RHS} = u^\top \mathbf{K}u + x^\top \mathbf{H}^\top \mathbf{K}u + u^\top \mathbf{K}\mathbf{H}x + x^\top \mathbf{H}^\top \mathbf{K}\mathbf{H}x + x^\top \mathbf{L}x$$

Compare the coefficients of $u^\top \cdots u$, $u^\top \cdots x$, and $x^\top \cdots x$,

Proof of the structure of optimal determistic LQR

Proof Proceed by backward induction.

- Basis: $V_T(x) = c_T(x) = \|x\|^2_{\mathbf{Q}_T}$.
- Induction hypothesis: $V_{t+1}(x) = \|x\|^2_{\mathbf{S}_{t+1}}$.
- Induction step:

$$
\begin{aligned}
V_t(x) &= \min_u \big[\|x\|^2_{\mathbf{Q}_t} + \|u\|^2_{\mathbf{R}_t} + V_{t+1}(\mathbf{A}_t x + \mathbf{B}_t u)\big] \\
&= \min_u \big[\|x\|^2_{\mathbf{Q}_t} + \underbrace{\|u\|^2_{\mathbf{R}_t} + \|\mathbf{A}_t x + \mathbf{B}_t u\|^2_{\mathbf{S}_{t+1}}}_{\text{completion of squares}}\big] \\
&= \min_u \big[\|x\|^2_{\mathbf{Q}_t} + \overbrace{\|u + \mathbf{H}_t x\|^2_{\mathbf{K}_t} + \|x\|^2_{\mathbf{L}_t}}\big]
\end{aligned}
$$

where

$$\mathbf{H}_t = [\mathbf{R}_t + \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{B}_t]^{-1} \mathbf{\Lambda}_t, \qquad \text{where } \mathbf{\Lambda}_t = \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{A}_t.$$

$$\mathbf{L}_t = \mathbf{A}^\top \mathbf{S}_{t+1} \mathbf{A}_t - \mathbf{\Lambda}_t^\top [\mathbf{R}_t + \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{B}_t]^{-1} \mathbf{\Lambda}_t.$$

Thus, the optimal control action is $u = -\mathbf{H}_t x$ and the optimal cost is

$$V_t(x) = \|x\|^2_{\mathbf{Q}_t} + \|x\|^2_{\mathbf{L}_t} = \|x\|^2_{\mathbf{S}_t}, \qquad \text{where } \mathbf{S}_t = \mathbf{Q}_t + \mathbf{L}_t.$$

Note that the update equation of $\mathbf{S}_t$ is same as that in the Theorem.

Linear Quadratic regulator example

Generalized LQR: Cross-term in cost

Minimizing output error

Suppose that instead of minimizing the norm of the state X_t, we are interested in minimizing the norm of the output $Y_t = \mathbf{C}_t X_t + \mathbf{D}_t U_t$. In such a case, the per-step cost function will be of the form

$$c_t(X_t, U_t) = \|X_t\|^2_{\mathbf{Q}_t} + \|U_t\|^2_{\mathbf{R}_t} + 2X_t^\top \mathbf{N}_t U_t$$

Assume that the terminal cost function does not change, and

$$\begin{bmatrix} \mathbf{Q}_t & \mathbf{N}_t \\ \mathbf{N}_t^\top & \mathbf{R}_t \end{bmatrix} \succeq 0, \qquad \text{or equivalently} \qquad \mathbf{Q} - \mathbf{N}\mathbf{R}^{-1}\mathbf{N}^\top \succeq 0.$$

Key Lemma

$$\|x\|^2_{\mathbf{Q}} + \|u\|^2_{\mathbf{R}} + 2x^\top \mathbf{N} u = \|x\|^2_{\tilde{\mathbf{Q}}} + \|u + \mathbf{R}^{-1}\mathbf{N}^\top x\|^2_{\mathbf{R}}.$$

where $\tilde{\mathbf{Q}} = \mathbf{Q} - \mathbf{N}\mathbf{R}^{-1}\mathbf{N}^\top$.

Change of variables

Let $\tilde{U}_t = U_t + \mathbf{R}_t^{-1}\mathbf{N}_t^\top X_t$. Then

$$X_{t+1} = \tilde{\mathbf{A}}_t X_t + \mathbf{B}_t \tilde{U}_t, \qquad \text{where } \tilde{\mathbf{A}}_t = \mathbf{A}_t - \mathbf{B}_t \mathbf{R}_t^{-1}\mathbf{N}_t^\top$$

- Thus, the system is in the same form as the standard LQR.

Generalized LQR: Cross-term in cost

Theorem The value function at time t is

$$V_t(X_t) = \|X_t\|_{S_t}$$

and the optimal control action is

$$U_t = -\mathbf{H}_t X_t$$

where the gain matrices $\mathbf{H}_t$ are computed recursively as follows:

$$\mathbf{H}_T = 0$$

$$\mathbf{H}_t = \left[\mathbf{R}_t + \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{B}_t\right]^{-1} \boldsymbol{\Lambda}_t$$

where

$$\boldsymbol{\Lambda}_t = \mathbf{N}_t^\top + \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{A}_t$$

and S_t are determined by the modified backward Riccati equations:

$$\mathbf{S}_T = \mathbf{Q}_T$$

$$\mathbf{S}_t = \mathbf{A}_t^\top \mathbf{S}_{t+1} \mathbf{A}_t + \mathbf{Q}_t - \boldsymbol{\Lambda}_t^\top \left[\mathbf{R}_t + \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{B}_t\right]^{-1} \boldsymbol{\Lambda}_t$$

- Note that the only change from the standard LQR equations is in the definition of $\boldsymbol{\Lambda}_t$.

Generalized LQR: Proof for cross-term in cost

Proof Consider the system with the change of variables. The structure of the optimal controller and the form of the value function are given as before. Recall that $U_t = \tilde{U}_t - R_t^{-1} N_t^\top X_t$. Hence,

$$H_t = \big[R_t + B_t^\top S_{t+1} B_t\big]^{-1} B_t^\top S_{t+1} \tilde{A}_t + R_t^{-1} N_t = \big[R_t + B_t^\top S_{t+1} B_t\big]^{-1} \Lambda$$

where

$$\begin{aligned} \Lambda_t &= B_t^\top S_{t+1} \tilde{A}_t + \big[R_t + B_t^\top S_{t+1} B_t\big] R_t^{-1} N_t \\ &= B_t^\top S_{t+1} \big[A_t - B_t R_t^{-1} N_t^\top\big] + \big[R_t + B_t^\top S_{t+1} B_t\big] R_t^{-1} N_t \\ &= B_t^\top S_{t+1} A_t + N_t^\top \end{aligned}$$

Furthermore, since the terminal cost is the same as before, the initial condition of the backward Riccati equation does not change. The Riccati update is given by

$$S_t = \tilde{A}_t^\top S_{t+1} \tilde{A}_t + \tilde{Q}_t - \big[B_t^\top S_{t+1} \tilde{A}_t\big]^\top \big[R_t + B_t^\top S_{t+1} B_t\big]^{-1} \big[B_t^\top S_{t+1} \tilde{A}_t\big]$$

Substituting the value of $\tilde{A}_t$ and $\tilde{Q}_t$ and some (messy) algebraic manipulation gives the result (see next page).

Generalized LQR: Proof for cross-term in cost (cont.)

Proof (cont.) Ignore the subscripts for ease of notation.

1. Let $\mathbf{K} = \mathbf{R} + \mathbf{B}^\top S \mathbf{B}$. Thus, $\mathbf{B}^\top S \mathbf{B} = \mathbf{K} - \mathbf{R}$.
2. $\tilde{\mathbf{A}} S \tilde{\mathbf{A}} = \mathbf{A} S \mathbf{A} + \mathbf{N} \mathbf{R}^{-1} (\mathbf{B}^\top S \mathbf{B}) \mathbf{R}^{-1} \mathbf{N}^\top - 2 \mathbf{A}^\top S \mathbf{B} \mathbf{R}^{-1} \mathbf{N}^\top$.
3. $\tilde{\mathbf{Q}} = \mathbf{Q} - \mathbf{N} \mathbf{R}^{-1} \mathbf{N}^\top$.
4. $\mathbf{B}^\top S \tilde{\mathbf{A}} = \mathbf{B}^\top S \mathbf{A} - (\mathbf{B}^\top S \mathbf{B}) \mathbf{R}^{-1} \mathbf{N}^\top = \boldsymbol{\Lambda} - \mathbf{K} \mathbf{R}^{-1} \mathbf{N}^\top$.
5. $(\mathbf{B}^\top S \tilde{\mathbf{A}})^\top \mathbf{K}^{-1} (\mathbf{B}^\top S \tilde{\mathbf{A}}) = \boldsymbol{\Lambda}^\top \mathbf{K}^{-1} \boldsymbol{\Lambda} + \mathbf{N} \mathbf{R}^{-1} \mathbf{K} \mathbf{R}^{-1} \mathbf{N}^\top - 2 \boldsymbol{\Lambda}^\top \mathbf{R}^{-1} \mathbf{N}^\top$.
6. (2) + (3) - (5) = Result - $2[\mathbf{N} + \mathbf{A}^\top S \mathbf{B} - \boldsymbol{\Lambda}^\top] \mathbf{R}^{-1} \mathbf{N}^\top$, where the last term is zero by definition of $\boldsymbol{\Lambda}$.

LQR Tracking problem

Tracking setup Suppose that we want to ensure that the output signal $Y_t = \mathbf{C}_t X_t$ is close to a reference trajectory $\{y^\circ_t\}_{t=1}^T$. Then, the cost functions are

$$c_t(X_t, U_t) = \|\mathbf{C}_t X_t - y^\circ_t\|^2_{\mathbf{Q}_t} + \|U_t\|^2_{\mathbf{R}_t}, \qquad c_T(X_T) = \|\mathbf{C}_T X_T - y^\circ_T\|^2_{\mathbf{Q}_T}.$$

Theorem The value function at time t is

$$V_t(X_t) = \|X_t\|^2_{S_t} + \alpha_t$$

and the optimal control action is

$$U_t = -H_t X_t + H^\circ_t r_{t+1}$$

Recursive computations

- $\{\mathbf{S}_t\}_{t=1}^T$ and $\{\mathbf{H}_t\}_{t=1}^T$ follow the same recursion as before;
- The gain matrices H°_t are given by $H^\circ_t = \big[\mathbf{R}_t + \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{B}_t\big]^{-1} \mathbf{B}_t^\top$
- The correction terms r_t are given by

$$r_T = \mathbf{C}_T^\top Q_T y^\circ_T, \qquad r_t = \big[\mathbf{A}_t - \mathbf{B}_t \mathbf{H}_t\big]^\top r_{t+1} + \mathbf{C}_t^\top \mathbf{Q}_t y^\circ_t$$

- The tracking error α_t is given by

$$\alpha_T = \|y^\circ_T\|^2_{\mathbf{Q}_T}, \qquad \alpha_t = \|y^\circ_t\|^2_{\mathbf{Q}_t} - 2 r_{t+1}^\top \mathbf{B}_t \big[\mathbf{R}_t + \mathbf{B}_t^\top \mathbf{S}_{t+1} \mathbf{B}_t\big]^{-1} \mathbf{B}_t^\top r_{t+1} + \alpha_{t+1}$$

Stochastic Linear Quadratic Regulator (LQR) setup

Stochastic dynamics

$$X_{t+1} = \mathbf{A}_t X_t + \mathbf{B}_t U_t + W_t, \qquad \text{where } \mathbb{E}[W_T \top W] = \Sigma_W$$

The above model is similar to the deterministic LQR setup with the exception that the state evolution is stochastic. The structure of the optimal controller is as given below.

Theorem

The value function at time t is

$$V_t(X_t) = \|X_t\|^2_{\mathbf{S}_t} + \alpha_t$$

and the optimal control action is

$$U_t = -\mathbf{H}_t X_t$$

where $\mathbf{S}_t$ and $\mathbf{H}_t$ follow the same recursion as before and

$$\alpha_T = 0$$

$$\alpha_t = \alpha_{t+1} + \mathrm{Tr}[\mathbf{\Sigma}_W \mathbf{S}_{t+1}] = \sum_{\tau=t+1}^{T} \mathrm{Tr}[\mathbf{\Sigma}_W \mathbf{S}_\tau]$$

Thus, the optimal controller is the same as in the deterministic case. The only effect of the noise is to increase the value function. (This phenomenon is unique to LQG systems).

Certainty Equivalence principle (Simon, 1948)

Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and X be a $\mathbb{R}^n$ valued random variable. If f is quadratic in its arguments, then

$$u^* = \arg\min_{u \in \mathbb{R}^m} \mathbb{E}[f(X, u)] = \arg\min_{u \in \mathbb{R}^m} f(\mathbb{E}[X], u)$$

Note: f is quadratic means $f(x, u) = \mathbf{A}x + \mathbf{B}u + \|x\|^2_{\mathbf{Q}} + \|u\|^2_{\mathbf{R}} + x^\top \mathbf{G}u + \alpha$ where all matrices are of appropriate dimensions.

Proof of the structure of optimal stochastic LQR

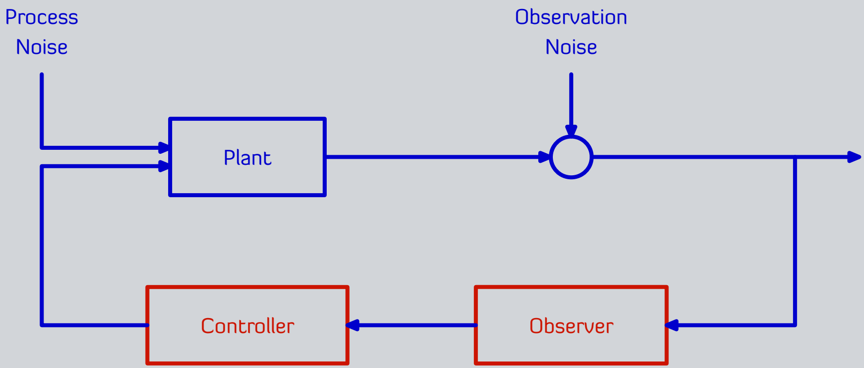
The proof is similar to that of deterministic LQR and follows from the following observation.

Lemma For any particular values $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$, matrices $\mathbf{A}, \mathbf{Q} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times m}$, and a random variable W such that $\mathbb{E}[W^\top W] = \Sigma_W$

$$\mathbb{E}[\|\mathbf{A}x + \mathbf{B}u + W\|^2_{\mathbf{Q}}] = \|\mathbf{A}x + \mathbf{B}u\|^2_{\mathbf{Q}} + \operatorname{Tr}(\mathbf{Q}\Sigma_W)$$

Proof

The POMDP setup — Output feedback



Problem formulation

Notation
State : $X_t \in \mathbb{R}^n$ Action: $U_t \in \mathbb{R}^m$
Observation: $Y_t \in \mathbb{R}^p$

Dynamics $X_{t+1} = \mathbf{A}_t X_t + \mathbf{B}_t U_t + W_t$, where $\mathbf{A}_t \in \mathbb{R}^{n \times n}$, $\mathbf{B}_t \in \mathbb{R}^{n \times m}$, $W_t \in \mathbb{R}^n$.

Observations $Y_t = \mathbf{C}_t X_t + N_t$, where $\mathbf{C}_t \in \mathbb{R}^{p \times n}$, $N_t \in \mathbb{R}^p$.

Random variables Primitive R.V.s $\{X_1, N_{1:T}, W_{1:T}\}$ are independent and Gaussian with

$$X_1 \sim \mathcal{N}(0, \Sigma_X), \quad N_t \sim \mathcal{N}(0, \Sigma_N), \quad W_t \sim \mathcal{N}(0, \Sigma_W).$$

Cost
Per step cost : $c_t(x_t, u_t) = \|x_t\|^2_{\mathbf{Q}_t} + \|u_t\|^2_{\mathbf{R}_t}$
Terminal reward: $c_T(x_T) = \|x_T\|^2_{\mathbf{Q}_T}$
where $\|X\|_Q = X^\top \mathbf{Q} X$ and for all t, $\mathbf{Q}_t = \mathbf{Q}_t^\top \succeq 0$ and $\mathbf{R}_t = \mathbf{R}_t^\top \succ 0$.

Control objective Choose $U_t = g_t(Y_{1:t}, U_{1:t-1})$ so as to minimize

$$J(g) = \mathbb{E}^g \Big[\sum_{t=1}^{T-1} c_t(X_t, U_t) + c_T(x_T) \Big]$$

Sufficient statistic for control

Theorem (Sufficient statistic) The state-estimate $\hat{X}_t = \mathbb{E}[X_t \mid Y_{1:t}, U_{1:t-1}]$ is a sufficient statistic for control, i.e., there is no loss of optimality in restricting attention to control laws of the form: $U_t = g_t(\hat{X}_t)$.

Kalman filtering This sufficient statistic is updated using the Kalman filtering equations

$$\hat{X}_{t+1} = \mathbf{A}_t \hat{X}_t + \mathbf{B}_t U_t + \mathbf{K}_{t+1} \big[Y_{t+1} - \mathbf{C}_{t+1} \hat{X}_t \big]$$

where $\mathbf{K}_t$ is the Kalman gain given by

$$\mathbf{K}_{t+1} = \mathbf{L}_t \big[\mathbf{\Sigma}_N + \mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^\top \big]^{-1} \qquad \text{with} \quad \mathbf{L}_t = \mathbf{A}_t \mathbf{P}_t \mathbf{C}_t^\top .$$

The initial mean is given by

$$\hat{X}_1 = \dots$$

and the covariance matrices $\mathbf{P}_t$ are precomputable and are given by forward Riccati difference equation

$$\mathbf{P}_{t+1} = \mathbf{A}_t \mathbf{P}_t \mathbf{A}_t^\top + \mathbf{\Sigma}_W - \mathbf{L}_t \big[\mathbf{\Sigma}_N + \mathbf{C}_t \mathbf{P}_t \mathbf{C}_t^\top \big]^{-1} \mathbf{L}_t^\top$$

with

$$\mathbf{P}_1 = \dots$$

Structure of optimal controller

Theorem The value function at time t is

$$V_t(\hat{X}_t) = \|\hat{X}_t\|^2_{\mathbf{S}_t} + \alpha_t$$

and the optimal control action is

$$U_t = -\mathbf{H}_t \hat{X}_t$$

where $\mathbf{S}_t$ and $\mathbf{H}_t$ follow the same recursion as before and

$$\alpha_T = \mathrm{Tr}[\mathbf{P}_T \mathbf{Q}_T]$$

$$\begin{aligned} \alpha_t &= \alpha_{t+1} + \mathrm{Tr}[\mathbf{P}_t \mathbf{Q}_t + (\mathbf{\Sigma}_W + \mathbf{A}_t \mathbf{P}_t \mathbf{A}_t^\top - \mathbf{P}_{t+1})\mathbf{S}_{t+1}] \\ &= \mathrm{Tr}[\mathbf{P}_T \mathbf{Q}_T] + \sum_{\tau=t}^{T-1} \mathrm{Tr}[\mathbf{P}_t \mathbf{Q}_t + (\mathbf{\Sigma}_W + \mathbf{A}_t \mathbf{P}_t \mathbf{A}_t^\top - \mathbf{P}_{t+1})\mathbf{S}_{t+1}] \end{aligned}$$

Further Reading

1. The Kalman filter (and its generalization to non-linear systems called extended Kalman filter) was used by NASA in the Ranger, Mariner, and Apollo missions, including the lunar module of Apollo 11. For a history of Kalman filtering in NASA see:

   Leonard A. McGee and Stanley F. Schmidt, "Discovery of the Kalman Filter as a practical tool for aerospace and industry," NASA Technical Memorandum TM-86847, Nov 1985.

   http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19860003843_1986003843.pdf

2. A slightly more cumbersome form of "Kalman filter" was derived in:

   Peter Swerling, "A proposed stagewise differential correlation procedure for satellite tracking and prediction," RAND Corporation report P-1292, Jan 1958; also published in Journal of Astronomical Science, vol 6, 1959

Decentralized Linear Quadratic Control

Static teams

Witsenhausen Counterexample

Appendices on background material

Static teams

Problem formulation

The system consists of n agents.

Primitive random variables $(X, Y^1, \dots, Y^n)$ jointly Gaussian with

$$\mathbb{E}[X] = \bar{x}, \quad \mathbb{E}[Y^i] = \bar{y}^i, \quad cov(X, Y^i) = \Xi_i, \quad cov(Y^i, Y^j) = \Sigma_{ij}.$$

Notation
Observation of agent i: Y^i
Action of agent i : U^i

Cost Let $U = \text{vec}(U^1, \dots, U^n)$.

$$c(X, U) = U^\top \mathbf{R} U + U^\top \mathbf{P} X = \sum_{i=1}^{n} \sum_{j=1}^{n} (U^i)^\top \mathbf{R}_{ij} U^j + \sum_{i=1}^{n} (U^i)^\top \mathbf{P}_i X$$

where $\mathbf{R} = [\mathbf{R}_{ij}]$ is symmetric and positive definite and $\mathbf{P} = [\mathbf{P}_1^\top, \dots, \mathbf{P}_n^\top]^\top$.

Control Objective Choose $U^i = g^i(Y^i)$ so as to minimize

$$J(\mathbf{g}) = \mathbb{E}^{\mathbf{g}}[c(X, U)]$$

Witsenhausen Counterexample

Problem Formulation

Primitive variables

- (X_1, N_2) zero mean independent (scalar) Gaussian.
- $var(X_1) = \sigma^2$ and $var(N_2) = 1$.

State equations

$$X_2 = X_1 + U_1$$

$$X_3 = X_2 - U_2$$

Observations

$$Y_1 = X_1$$

$$Y_2 = X_2 + N_2$$

Cost $k^2 U_1^2 + X_3^2$

Objective Choose $U_i = g_i(Y_i)$ to minimize the expected cost.

Note The dynamics are linear, cost is quadratic, and noise is Gaussian. But the information structure is non-classical. Controller 2 does not know the observation of controller 1.

Reformulation using a change of variables

Reformulation Let $X = X_1$, $N = N_2$, and

$$f(x) = x + g_1(x)$$

$$g(y) = g_2(y)$$

Then,

$$J(f, g) = \mathbb{E}\left[k^2(X - f(x))^2 + \big(f(X) - g\big(f(X) + N\big)\big)^2\right]$$

The optimization problem is

$$J^* = \inf_{(f,g)} J(f, g)$$

Theorem An optimal solution exists.

Remark The existence of an optimal solution was proved by Witsenhausen. For the purpose of what we want to show, the existence result is not that important.

Structure of optimal strategies

Lemma

1. $0 \leqslant J^* \leqslant \min(1, k^2\sigma^2)$.
2. For any (f, g), there exists $(\tilde{f}, \tilde{g})$ such that $\mathbb{E}[\tilde{f}(X)] = 0$, $\mathbb{E}[(X - \tilde{f}(X))^2] \leqslant \sigma^2$ and $J(\tilde{f}, \tilde{g}) \leqslant J(f, g)$.

Proof

1. Since the cost is positive, $J^* \geqslant 0$. Now consider the following two strategies:
   - $f(x) = 0$, $g(y) = 0$. $J(f, g) = \mathbb{E}[k^2X^2] = k^2\sigma^2$.
   - $f(x) = x$, $g(y) = y$. $J(f, g) = \mathbb{E}[(X - (X + N))^2] = 1$.

   Hence, $J^* \leqslant \min(1, k^2\sigma^2)$.

2. If $\mathbb{E}[(X - f(X))^2] \geqslant \sigma^2$, then $J(f, g) \geqslant \sigma^2$. Set $\tilde{f}(x) = 0$ and $g(y) = 0$. Then $\tilde{f}$ satifies the required conditions and $J(\tilde{f}, \tilde{g}) = \sigma^2 \leqslant J(f, g)$.

   If $\mathbb{E}[(X-f(X))^2] \leqslant \sigma^2$, then $\mathbb{E}[f(X)^2] \leqslant 4\sigma^2$ (why?). Therefore, $\mathbb{E}[f(X)]$ exists. Let $m = \mathbb{E}[f(X)]$.
   - Define $\tilde{f}(x) = f(x) - m$ and $\tilde{g}(y) = g(y + m) - m$.
   - $\mathbb{E}[\tilde{f}(X)] = 0$ and $\mathbb{E}[(X - \tilde{f}(X))^2] = \mathbb{E}[(X - f(X))^2] - m^2 \leqslant \sigma^2$.
   - $\tilde{f}(X) - \tilde{g}(\tilde{f}(X) + N) = f(X) - g(f(X) + N)$.

   Consequently $J(\tilde{f}, \tilde{g}) \leqslant J(f, g)$.

Performance of an affine strategy

Affine strategy From the previous lemma, we can restrict attention to strategies such that $\mathbb{E}[f(X)] = 0$. Thus, when considering affine strategies for the first stage, we only need to consider $f(x) = \lambda x$.

The best response to $f(x) = \lambda x$ is $g(y) = \mu y$ where

$$\mu = \frac{\sigma^2\lambda^2}{1 + \sigma^2\lambda^2}$$

The corresponding performance is

$$J(\lambda) = k^2\sigma^2(1 - \lambda)^2 + \frac{\sigma^2\lambda^2}{1 + \sigma^2\lambda^2}$$

Proof Let $\tilde{X} = \lambda X \sim \mathcal{N}(0, \sigma^2\lambda^2)$. Then, $Y = \tilde{X} + N$, and we want to choose $g(y)$ to minimize $\mathbb{E}[(\tilde{X} - g(Y))^2]$. Since all random variables are Gaussian, the best estimator is $g(Y) = \mathbb{E}\,\tilde{X} + \Sigma_{\tilde{X}Y}\Sigma_{YY}^{-1}Y$ and the corresponding performance is $\Sigma_{\tilde{X}\tilde{X}} - \Sigma_{\tilde{X}Y}\Sigma_{YY}^{-1}\Sigma_{\tilde{X}Y}^{\top}$. The result follows from observing that

$$\Sigma_{\tilde{X}\tilde{X}} = \sigma^2\lambda^2, \quad \Sigma_{YY} = 1 + \sigma^2\lambda^2, \quad \text{and} \quad \Sigma_{\tilde{X}Y} = \sigma^2\lambda^2.$$

Best affine strategy

Best affine strategy Let $\lambda = t/\sigma$. Then, the best choice of t satisfies

$$\frac{t}{(1+t^2)^2} = k^2(\sigma - t)$$

Proof This follows by taking the derivative of $J(\lambda)$ wrt λ and setting it to zero.

A family of solutions Let $k^2 < \frac{1}{4}$ and $k\sigma = 1$. Then,

$$\lambda = \mu = \frac{1}{2}\left(1 \pm \sqrt{1-4k^2}\right) \quad \text{and} \quad J_a^* = 1 - k^2.$$

Proof It is easy to verify that this satisfies the solution above.

Non-linear strategies can outperform the best lienar strategy

A non-linear strategy

Note: This is different from the strategy presented in Witsenhausen's paper.

Consider

$$f(x) = \sigma \operatorname{sgn}(x) \quad \text{and} \quad g(y) = \lambda \operatorname{sgn}(y).$$

Let $\tilde{X} = f(X)$. Then,

$$\begin{aligned} J(f, g) &= k^2\, \mathbb{E}[(X - \tilde{X})^2] + \mathbb{E}[(\tilde{X} - (\tilde{X} + N))^2] \\ &= 2k^2\sigma^2 \left(1 - \mathbb{E}\left[\left|\frac{x}{\sigma}\right|\right]\right) + 4\sigma^2\, \mathbb{P}(N > \sigma) \\ &= 2k^2\sigma^2 \left(1 - \sqrt{\frac{2}{\pi}}\right) + 4\sigma^2 \operatorname{erfc}(\sigma) \end{aligned}$$

Comparison

Let $k\sigma = 1$ and consider $k \to 0$. Then

$$J_a^* = 1 \quad \text{but} \quad J(f, g) = 2\left(1 - \sqrt{\frac{2}{\pi}}\right) \approx 0.404$$

Thus, the non-linear strategy given above outperforms the best affine strategy!

Appendices on background material

Appendix: Positive definite matrices

Positive definite

A $n \times n$ symmetric matrix $\mathbf{M}$ is

- positive definite (written as $\mathbf{M} \succ 0$) if

$$\text{for any } x \neq 0, x \in \mathbb{R}^n, \quad x^\top \mathbf{M} x > 0.$$

- positive semi-definite (written as $\mathbf{M} \succeq 0$) if

$$\text{for any } x \neq 0, x \in \mathbb{R}^n, \quad x^\top \mathbf{M} x \geqslant 0.$$

Eigenvalues characterization

A symmetric matrix is positive definite (resp., positive semi-definite) if and only if all of its eigenvalues are positive (resp., non-negative).

Examples

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 2x_2^2 \implies \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \succ 0.$$

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_2^2 \implies \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} \succeq 0.$$

Appendix: Linear estimation of Gaussian signals

Conditional expectation of Gaussian vectors

Let (X, Y) be jointly Gaussian with mean $\mu = (\mu_X, \mu_Y)$ and covariance $\Sigma = \begin{bmatrix} \mathbf{\Sigma}_{XX} & \mathbf{\Sigma}_{XY} \\ \mathbf{\Sigma}_{XY}^\top & \mathbf{\Sigma}_{YY} \end{bmatrix}$. Then,

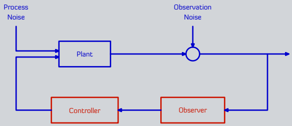
$$\mathbb{E}[X|Y] = \mu_X + \mathbf{\Sigma}_{XY}\mathbf{\Sigma}_{YY}^{-1}(Y - \mu_Y)$$

is a Gaussian random variable with mean μ_X and covariance $\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{XY}^\top$.

The mean square error $(X - \mathbb{E}[X|Y])^2$ is $\mathrm{Tr}[\mathbf{\Sigma}_{XX} - \mathbf{\Sigma}_{XY}\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{XY}^\top]$.

The MDP setup — State feedback

The POMDP setup — Output feedback



Static teams

Witsenhausen Counterexample

Appendices on background material