10 LQR with partial observations

We will now consider the LQR problem with partial observations. We have the state space $\mathcal{X} = \mathbb{R}^n$, the action space $\mathcal{U} = \mathbb{R}^m$, and the observation space $\mathcal{Y} = \mathbb{R}^p$; the linear state dynamics and the observation model are given by

$$X_{t+1} = AX_t + BU_t + W_t \tag{1a}$$
$$Y_t = CX_t + V_t \tag{1b}$$

where, as usual, the primitive random variables consisting of the initial state $X_0 \sim \mathcal{N}(0, \Sigma_0)$, the input disturbances $W_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma_W)$, and the output disturbances $V_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma_V)$ are assumed to be independent. Once again, we work with the quadratic cost

$$c(x, u) = x^\top Q x + u^\top R u, \tag{2}$$

with $Q = Q^\top \succeq 0$ and $R = R^\top \succ 0$. Because all the primitive random variables are Gaussian, we also refer to this model as the Linear Quadratic Gaussian model, or LQG for short.

Now admissible policies are of the form $g = (g_t)_{t \geq 0}$ with $g_t : \mathcal{Y}_0^t \times \mathcal{U}_0^{t-1} \to \mathcal{U}$, i.e., the action at time t is to be chosen as a function of the history of observations Y_0^t and previously taken actions U_0^{t-1}. Given a policy g, we will consider two types of cost criteria:

- finite-horizon expected cost

$$J_T(g) := \mathbf{E}^g \left[\sum_{t=0}^{T-1} (X_t^\top Q X_t + U_t^\top R U_t) + X_T^\top Q_T X_T \right], \tag{3}$$

  where we may also have the terminal cost $c_T(x) = x^\top Q_T x$ for some symmetric positive semidefinite matrix Q_T;

- long-term average cost

$$\bar{J}(g) := \limsup_{T \to \infty} \frac{1}{T} \mathbf{E}^g \left[\sum_{t=0}^{T-1} (X_t^\top Q X_t + U_t^\top R U_t) \right]. \tag{4}$$

We will prove that, in both cases, the optimal[1] policy has a very simple structure: $U_t = G_t \widehat{X}_t$ for some sequence of gain matrices $G_t \in \mathbb{R}^{m \times n}$ in the finite-horizon case and $U_t = G\widehat{X}_t$ for some matrix $G \in \mathbb{R}^{m \times n}$ in the long-term average-cost case. In the finite-horizon case, $\widehat{X}_t := \mathbf{E}[X_t | Y_0^t]$ is the minimum mean-square error (MMSE) estimate of the state X_t given the observation history Y_0^t, while in the average-cost case $\widehat{X}_t$ is the asymptotic MMSE estimate computed via the steady-state Kalman filter.

[1] With the usual caveats in the average-cost case.

10.1 Estimation of the state in the presence of control

Fix an arbitrary policy $g = (g_t)_{t \geq 0}$. We will show that the conditional means $\mu_t^g := \mathbf{E}^g[X_t | Y_0^t, U_0^{t-1}]$ and the conditional covariances $K_t^g := \mathbf{E}^g[(X_t - \mu_t^g)(X_t - \mu_t^g)^\top | Y_0^t, U_0^{t-1}]$ can be computed recursively. Moreover, the update rule for μ_t^g is independent of the policy g, and the covariances K_t^g are also independent of the data Y_0^t, U_0^{t-1}. We start with the following simple lemma:

Lemma 10.1 *The state X_t and the observation Y_t can be written as*

$$X_t = \tilde{X}_t^g + \bar{X}_t, \qquad Y_t = \tilde{Y}_t^g + \bar{Y}_t, \tag{5}$$

where $\tilde{X}_t^g$, $\bar{X}_t$, $\tilde{Y}_t^g$, $\bar{Y}_t$ evolve according to

$$\tilde{X}_{t+1}^g = A\tilde{X}_t^g + BU_t \tag{6a}$$
$$\tilde{Y}_t^g = C\tilde{X}_t^g \tag{6b}$$
$$\bar{X}_{t+1} = A\bar{X}_t + W_t \tag{6c}$$
$$\bar{Y}_t = C\bar{X}_t + V_t \tag{6d}$$

with the initial conditions $\tilde{X}_0^g = 0$, $\bar{X}_t \sim \mathcal{N}(0, \Sigma_0)$.

The proof is a straightforward application of linearity. The main message here is that the evolution of $\tilde{X}_t^g$ and $\tilde{Y}_t^g$ is deterministic and depends on g, while the evolution of $\bar{X}_t$ and $\bar{Y}_t$ is stochastic and does not depend on g. Since the process $\{(\bar{X}_t, \bar{Y}_t)\}_{t=0}^\infty$ is of the form that was discussed in the lecture on the Kalman filter. Thus, for each t, the conditional means

$$\bar{\mu}_{t|t} := \mathbf{E}[\bar{X}_t | \bar{Y}_0^t] \qquad \text{and} \qquad \bar{\mu}_{t+1|t} := \mathbf{E}[\bar{X}_{t+1} | Y_0^t] \tag{7}$$

and the conditional covariances

$$K_{t|t} := \mathrm{Cov}(\bar{X}_t | \bar{Y}_0^t) \qquad \text{and} \qquad K_{t+1|t} := \mathrm{Cov}(\bar{X}_{t+1} | \bar{Y}_0^t) \tag{8}$$

can be computed via the Kalman filter:

$$\bar{\mu}_{t+1|t+1} = A\bar{\mu}_{t|t} + L_{t+1}(\bar{Y}_{t+1} - CA\bar{\mu}_{t|t}) \tag{9}$$
$$\bar{\mu}_{t+1|t} = A\bar{\mu}_{t|t-1} + AL_t(\bar{Y}_t - C\bar{\mu}_{t|t-1}) \tag{10}$$
$$K_{t+1|t} = AK_{t|t-1}A^\top - AL_tCK_{t|t-1}A^\top + \Sigma_W \tag{11}$$
$$K_{t+1|t+1} = (I - L_{t+1}C)K_{t+1|t} \tag{12}$$

where

$$L_t := K_{t|t-1}C^\top(\Sigma_V + CK_{t|t-1}C^\top)^{-1}, \tag{13}$$

with the initial conditions $\bar{\mu}_{0|0} = L_0\bar{Y}_0$, $\bar{\mu}_{0|-1} = 0$ and $K_{0|-1} = \Sigma_0$. Next, we make the following key observation:

Lemma 10.2 *For each t, the σ-algebras generated by (Y_0^t, U_0^{t-1}) and by $\bar{Y}_0^t$ are the same — that is, $\bar{Y}_0^t$ contains the same information as the observation-action history (Y_0^t, U_0^{t-1}).*

Proof: By construction, $\tilde{X}_t^g$ and $\tilde{Y}_t^g$ are deterministic functions of U_0^{t-1}. Thus, $\bar{Y}_t = Y_t - \tilde{Y}_t^g$ is also a function of (Y_0^t, U_0^{t-1}). This implies that $\bar{Y}_0^t$ is a function of (Y_0^t, U_0^{t-1}), and so $\sigma(\bar{Y}_0^t) \subseteq \sigma(Y_0^t, U_0^{t-1})$.

We will prove the reverse inclusion $\sigma(Y_0^t, U_0^{t-1}) \subseteq \sigma(\bar{Y}_0^t)$ by induction. For $t = 0$, $Y_0 = \tilde{Y}_0^g + \bar{Y}_0$ is a function of $\bar{Y}_0$, since $\tilde{Y}_0^g = C\tilde{X}_0^g = 0$. Now suppose that $Z_t^g := (Y_0^t, U_0^{t-1})$ is a function of $\bar{Y}_0^t$. Since

$$Z_{t+1}^g = (Y_0^{t+1}, U_0^t) = (Z_t^g, Y_{t+1}, U_t) \tag{14}$$

where $U_t = g_t(Y_0^t, U_0^{t-1})$, so U_t is a function of $\bar{Y}_0^t$ by the induction hypothesis. Next, since $\tilde{X}_{t+1}^g$ is a function of U_0^t, it follows that $\tilde{X}_{t+1}^g$ and $\tilde{Y}_{t+1}^g = C\tilde{X}_{t+1}^g$ are functions of $\bar{Y}_0^t$. Finally, $Y_{t+1} = \tilde{Y}_{t+1}^g + \bar{Y}_{t+1}$, so Y_{t+1}^g is a function of $\bar{Y}_0^{t+1}$. Consequently, Z_{t+1}^g is a function of $\bar{Y}_0^{t+1}$, and the result follows by induction. ∎

Lemma 10.2 is of tremendous consequence: since $\tilde{X}_t^g$ is a function of U_0^{t-1}, it follows that the conditional probability laws of $X_t = \tilde{X}_t^g + \bar{X}_t$ given (Y_0^t, U_0^{t-1}) and of X_t given $\bar{Y}_0^t$ coincide, i.e.,

$$\mathbf{P}[X_t \in \cdot | Y^t, U_0^{t-1}] = \mathbf{P}[\tilde{X}_t^g + \bar{X}_t \in \cdot | \bar{Y}_0^t]. \tag{15}$$

By the same token,

$$\mathbf{P}[X_{t+1} \in \cdot | Y_0^t, U_0^{t-1}] = \mathbf{P}[\tilde{X}_{t+1}^g + \bar{X}_{t+1} \in \cdot | \bar{Y}_0^t]. \tag{16}$$

Thus, in particular, the conditional distribution of X_t given (Y_0^t, U_0^{t-1}) is Gaussian with mean

$$\widehat{X}_{t|t} := \mathbf{E}[X_t | Y^t, U_0^{t-1}] = \bar{\mu}_{t|t} + \tilde{X}_t^g \tag{17}$$

and covariance $\mathrm{Cov}(X_t | Y_0^t, U_0^{t-1}) = K_{t|t}$; likewise, the conditional distribution of X_{t+1} given (Y_0^t, U_0^{t-1}) is Gaussian with mean $\widehat{X}_{t+1|t} = \bar{\mu}_{t+1|t} + \tilde{X}_{t+1}^g$ and covariance $K_{t+1|t}$. These observations lead to the following key result:

Theorem 10.1 *For any policy g, the state estimates $\widehat{X}_{t|t}$ and $\widehat{X}_{t|t-1}$ evolve according to the recursion*

$$\widehat{X}_{t+1|t+1} = A\widehat{X}_{t|t} + BU_t + L_{t+1}(Y_{t+1} - C(A\widehat{X}_{t|t} + BU_t)) \tag{18}$$

$$\widehat{X}_{t+1|t} = A\widehat{X}_{t|t-1} + BU_t + L_{t+1}(Y_t - C\widehat{X}_{t|t-1}) \tag{19}$$

with initial conditions $\widehat{X}_{0|0} = L_0 Y_0$ and $\widehat{X}_{0|-1} = 0$.

Proof: From (17) and (9), we have

$$\widehat{X}_{t+1|t+1} = \bar{\mu}_{t+1|t+1} + \tilde{X}_{t+1}^g \tag{20}$$

$$= A\bar{\mu}_{t|t} + L_{t+1}(\bar{Y}_{t+1} - CA\bar{\mu}_{t|t}) + A\tilde{X}_t^g + BU_t \tag{21}$$

$$= A(\bar{\mu}_{t|t} + \tilde{X}_t^g) + BU_t + L_{t+1}(\bar{Y}_{t+1} - CA\bar{\mu}_{t|t}) \tag{22}$$

$$= A\widehat{X}_t + BU_t + L_{t+1}(Y_{t+1} - C(A\widehat{X}_t + BU_t)). \tag{23}$$

By similar reasoning,

$$\widehat{X}_{t+1|t} = \bar{\mu}_{t+1|t} + \tilde{X}^g_{t+1} \tag{24}$$
$$= A\bar{\mu}_{t|t-1} + AL_t(\bar{Y}_t - C\bar{\mu}_{t|t-1}) + A\tilde{X}^g_t + BU_t \tag{25}$$
$$= A(\bar{\mu}_{t|t-1} + \tilde{X}^g_t) + BU_t + AL_t(\bar{Y}_t - C\bar{\mu}_{t|t-1}) \tag{26}$$
$$= A\widehat{X}_{t|t-1} + BU_t + AL_t(Y_t - C\widehat{X}_{t|t-1}). \tag{27}$$

The initial conditions are given by $\widehat{X}_{0|0} = \mathbf{E}[X_0|Y_0] = L_0Y_0$ and $\widehat{X}_{0|-1} = \mathbf{E}[X_0] = 0$. ■

The update rule for $\widehat{X}_t := \widehat{X}_{t|t}$ has clear intuitive appeal: at each time t, we use the current estimate $\widehat{X}_t$ and the known action U_t to *predict* the next oobservation Y_{t+1} by $\widehat{Y}_{t+1} = C(A\widehat{X}_t + BU_t)$; once the new observation Y_{t+1} is received, we multiply the error $E_{t+1} := Y_{t+1} - \widehat{Y}_{t+1}$ by the gain matrix L_{t+1} and add it to the prediction $\widehat{X}_{t+1|t} = A\widehat{X}_t + BU_t$.

Now we make one additional observation: the error E_{t+1} can be further expressed as

$$E_{t+1} = Y_{t+1} - C(A\widehat{X}_t + BU_t) \tag{28}$$
$$= CX_{t+1} + V_{t+1} - C(A\widehat{X}_t + BU_t) \tag{29}$$
$$= C(AX_t + BU_t + W_t - A\widehat{X}_t - BU_t) + V_{t+1} \tag{30}$$
$$= C(A(X_t - \widehat{X}_t) + W_t) + V_{t+1}, \tag{31}$$

and, since all the primitive random variables are Gaussian and $\widehat{X}_t$ is a linear function of the primitive random variables, we see that E_{t+1} is also Gaussian with zero mean and covariance matrix $C(AK_{t|t}A^\top + \Sigma_W)C^\top + \Sigma_V = CK_{t+1|t}C^\top + \Sigma_V$. Moreover, it can be shown that E_{t+1} is *independent* of $\widehat{X}_t$ and U_t. The main takeaway here is that, for any policy g, the estimated state $\widehat{X}_t$ evolves according to the linear Gaussian model

$$\widehat{X}_{t+1} = A\widehat{X}_t + BU_t + \bar{W}_{t+1}, \tag{32}$$

where $\bar{W}_{t+1} := L_{t+1}E_{t+1}$ is a sequence of independent Gaussian random vectors with

$$\mathbf{E}[\bar{W}_{t+1}] = 0 \qquad \text{and} \qquad \mathrm{Cov}(\bar{W}_{t+1}) = L_{t+1}(CK_{t+1|t}C^\top + \Sigma_V)L^\top_{t+1}. \tag{33}$$

Note, moreover, that the information state $\widehat{X}_t$ is *fully observed*. Another important observation is that the covariance matrices $K_{t|t}$ and $K_{t|t-1}$ do not depend on the data (Y_0^t, U_0^{t-1}) or on the policy g, and therefore can be precomputed offline.

10.2 Finite horizon

Now we are ready to tackle the partially observed LQG problem in the finite-horizon setting with the stage costs $c_t(x,u) = x^\top Qx + u^\top Ru$, $0 \le t \le T-1$, and the terminal cost $c_T(x) = x^\top Qx$.[2] We are interested in minimizing the horitzon-T expected cost

$$J_T(g) := \mathbf{E}^g\left[\sum_{t=0}^{T-1}(X_t^\top QX_t + U_t^\top RU_t) + X_T^\top QX_T\right] \tag{34}$$

[2]We have set $Q_T = Q$ for simplicity.

over all admissible policies g. In particular, we seek a policy g^* that achives $J_T^* := \min_g J_T(g)$.

The main message here is that g^* has the special form

$$U_t = G_t \widehat{X}_t, \tag{35}$$

where $\widehat{X}_t$ is the information state that evolves according to (32), while G_t is the optimal LQR gain matrix at time t for the *deterministic* control problem

$$\min_{u_0,\ldots,u_{T-1}\in\mathbb{R}^m} \left\{ \sum_{t=0}^{T-1} (x_t^\top Q x_t + u_t^\top R u_t) + x_T^\top Q_T x_T \right\} \tag{36}$$

$$\text{subject to } x_{s+1} = Ax_s + Bu_s, \qquad 0 \le s \le T-1. \tag{37}$$

As we already saw in the previous lectures, the computation of G_t involves solving the T-step Riccati recursion starting at Q.

Let $\widehat{X}_t$ be the information state at time t. Given a policy g, for any t we have

$$\begin{aligned}
&\mathbf{E}^g[c_t(X_t, U_t)] \\
&= \mathbf{E}^g[X_t^\top Q X_t + U_t^\top R U_t] &(38)\\
&= \mathbf{E}^g[(\widehat{X}_t + X_t - \widehat{X}_t)^\top Q (\widehat{X}_t + X_t - \widehat{X}_t) + U_t^\top R U_t] &(39)\\
&= \mathbf{E}^g[\widehat{X}_t^\top Q \widehat{X}_t + U_t^\top R U_t] + 2 \cdot \mathbf{E}^g[(X_t - \widehat{X}_t)^\top Q \widehat{X}_t] + \mathbf{E}^g[(X_t - \widehat{X}_t)^\top Q (X_t - \widehat{X}_t)]. &(40)
\end{aligned}$$

Note that the first term is equal to $\mathbf{E}^g[c_t(\widehat{X}_t, U_t)]$, the second term is zero (why?), and the third term is equal to

$$\mathbf{E}^g[(X_t - \widehat{X}_t)^\top Q (X_t - \widehat{X}_t)] = \mathrm{tr}[Q \cdot \mathrm{Cov}(X_t - \widehat{X}_t)] = \mathrm{tr}[Q K_t], \tag{41}$$

where the error covariance matrix K_t does not depend on the data (Y_0^t, U_0^{t-1}) or on the policy g. The same applies to the terminal cost:

$$\mathbf{E}^g[c_T(X_T)] = \mathbf{E}^g[c_T(\widehat{X}_T)] + \mathrm{tr}[Q K_T]. \tag{42}$$

Armed with these observations, we see that the control problem in the partially observed LQG model reduces to controlling the *information state* $\widehat{X}_t$, which evolves according to the fully observed linear dynamics (32). The only difference is that the disturbance process $\{\bar{W}_t\}_{t\ge1}$ consists of independent zero-mean Gaussian random vectors whose covariance matrices depend on t. However, a moment of reflection tells us that the (time-varying) optimal policy in the fully observed LQR model depends only on the system matrices A and B, as well as on the cost matrices Q and R.

Thus, everything reduces to analyzing the fully observed LQG problem

$$\widehat{X}_{t+1} = A\widehat{X}_t + BU_t + \bar{W}_{t+1}, \tag{43}$$

with the initial conditon $\widehat{X}_0 = L_0 Y_0$, where the error process $\{\bar{W}_t\}_{t\ge1}$ consists of independent zero-mean Gaussian random vectors $\bar{W}_t$ with $\Sigma_t := \mathrm{Cov}(\bar{W}_t) = L_t(CK_{t|t-1}C^\top + \Sigma_V)L_t^\top$. Recalling our results on the finite-horizon LQR problem, we can immediately write down the optimal policy:

Theorem 10.2 *The optimal policies and the minimum expected cost are given by*

$$g_t^*(Y^t, U_0^{t-1}) = G_t \widehat{X}_t \tag{44}$$

$$J_T^* = \mathbf{E}[\widehat{X}_0^\top P_0 \widehat{X}_0] + \sum_{t=1}^{T} \mathrm{tr}(P_t \Sigma_t) + \sum_{t=0}^{T} \mathrm{tr}(Q K_t), \tag{45}$$

where the matrices G_t, P_t can be computed recursively as follows: $P_T = Q$ and, for $t = T-1, T-2, \ldots, 0$,

$$G_t = -(R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1} A \tag{46}$$

$$P_t = Q + A^\top (P_{t+1} - P_{t+1} B (R + B^\top P_{t+1} B)^{-1} B^\top P_{t+1}) A \tag{47}$$

and the information state $\widehat{X}_t$ and the error covariance matrices K_t are computed using the Kalman filter recursion.

Remark 10.1 It is also possible to derive this result from first principles using dynamic programming. You will do this in the homework.

We can express the optimal cost (45) in a different, more suggestive, way. First, using the fact that $L_t = K_{t|t-1} C^\top (\Sigma_V + C K_{t|t-1} C^\top)^{-1}$, we can write

$$\mathrm{tr}(P_t \Sigma_t) = \mathrm{tr}(P_t L_t (\Sigma_V + C K_{t|t-1} C^\top) L_t^\top) \tag{48}$$

$$= \mathrm{tr}(P_t K_{t|t-1} C^\top L_t^\top). \tag{49}$$

Next, since $K_t = K_{t|t-1} - L_t C K_{t|t-1}$, it follows that the matrix $L_t C K_{t|t-1}$ is symmetric and equal to $K_{t|t-1} - K_t$. Thus, for each t,

$$\mathrm{tr}(P_t \Sigma_t) = \mathrm{tr}(P_t (K_{t|t-1} - K_t)), \tag{50}$$

where $K_t - K_{t|t-1}$ quantifies the reduction of uncertainty about X_t when the information Y_0^{t-1} is augmented by the new observation Y_t. A similar argument leads to the expression

$$\mathbf{E}[\widehat{X}_0^\top P_0 \widehat{X}_0] = \mathrm{tr}(P_0 (\Sigma_0 - K_0)) = \mathrm{tr}(P_0 (K_{0|-1} - K_0)). \tag{51}$$

Thus, the optimal cost (45) can be written as

$$J_T^* = \sum_{t=0}^{T} \mathrm{tr}(P_t (K_{t|t-1} - K_t)) + \sum_{t=0}^{T} \mathrm{tr}(Q K_t). \tag{52}$$

Finally, since $K_{t|t-1} = AK_{t-1}A^\top + \Sigma_W$, we can rewrite J_T^* as

$$J_T^* = \sum_{t=0}^{T} \mathrm{tr}(QK_t) + \sum_{t=1}^{T} \mathrm{tr}(P_t(AK_{t-1}A^\top + \Sigma_W - K_t)) + \mathrm{tr}(P_0(\Sigma_0 - K_0)) \tag{53}$$

$$= \underbrace{\mathrm{tr}(P_0\Sigma_0) + \sum_{t=1}^{T} \mathrm{tr}(P_t\Sigma_W)}_{:=J^*_{\mathsf{LQR}}}$$

$$+ \underbrace{\mathrm{tr}((Q-P_0)K_0) + \sum_{t=1}^{T} \Big[\mathrm{tr}((Q-P_t)K_t) + \mathrm{tr}(P_tAK_{t-1}A^\top)\Big]}_{:=J_{\mathsf{est}}}, \tag{54}$$

where J^*_{LQR} is the optimal cost of the fully observed LQR problem with the same parameters, while J_{est} is the additional penalty for not knowing the true state. Note that, in the fully observed case $C = I$ and $\Sigma_V = 0$, we have $K_t = 0$ for all t, and therefore $J_T^* = J^*_{\mathsf{LQR}}$, as expected.

10.3 Long-term average cost

The LQG problem under the average-cost criterion admits similar treatment, but now we replace both the LQR optimal controller and the Kalman filter with their asymptotic (steady-state) variants. Specifically, we consider the information state update

$$\widehat{X}_{t+1} = A\widehat{X}_t + BU_t + L(Y_{t+1} - C(A\widehat{X}_t + BU_t)), \tag{55}$$

where the steady-state Kalman gain matrix L is given by

$$L = KC^\top(\Sigma_V + CKC^\top)^{-1} \tag{56}$$

and K is the solution of the DARE

$$K = A\left(K - KC^\top(\Sigma_V + CKC^\top)^{-1}CK\right)A^\top + \Sigma_W. \tag{57}$$

Clearly, $\widehat{X}_t$ computed according to (55) is *not* equal to the true information state $\mathbf{E}[X_t|Y_0^t, U_0^{t-1}]$, but it is asymptotically accurate (as $t \to \infty$) provided the DARE has a unique symmetric positive definite solution — that is, $\lim_{t\to\infty} \mathrm{Cov}(X_t - \widehat{X}_t) = (I - LC)K$.

The control law is of the estimated state feedback form $U_t = G\widehat{X}_t$, where the LQR gain matrix G is computed given by

$$G = -(R + BPB^\top)^{-1}B^\top PA \tag{58}$$

and P is the solution of the DARE

$$P = A^\top\left(P - PB(R + B^\top PB)^{-1}B^\top P\right)A + Q. \tag{59}$$

We can summarize the main result on the average-cost LQR problem in the following theorem:

Theorem 10.3 *Assume the following:*

1. *The pair (A, B) is controllable and there exists a matrix $\tilde{C}$, such that $Q = \tilde{C}^\top \tilde{C}$ and the pair $(A, \tilde{C})$ is observable.*
2. *The pair (A, C) is observable and there exists a matrix Γ, such that $\Sigma_W = \Gamma\Gamma^\top$ and the pair (A, Γ) is controllable.*

Then the estimated state feedback policy $U_t = G\widehat{X}_t$ is optimal among all admissible policies g satisfying $\mathbf{E}^g[\|X_T\|^2] = o(T)$, and achieves the cost

$$\bar{J} = \mathrm{tr}(Q\tilde{K}) + \mathrm{tr}(PK), \tag{60}$$

where $\tilde{K} = (I - LC)K$.

We do not give the proof, but some elements of it will be explored in the homework. Once again, if we set $C = I$ and $\Sigma_V = 0$ (the fully observed case), we get $K = \Sigma_W$, $L = I$, and $\tilde{K} = 0$, so we get $\bar{J} = \mathrm{tr}(P\Sigma_W)$, which is exactly what we had derived earlier by solving the ACOE.

10.4 The separation principle and the dual effect

In both cases, the optimal policy had the following structure: first, an estimate $\widehat{X}_t$ of the state X_t was computed according to an appropriate recursion (the exact Kalman filter in the finite-horizon case and the steady-state Kalman filter in the infinite-horizon average-cost case), and then the control gain was applied. In both cases, the control gain was exactly the same as one would have in the fully observed case. This is a manifestation of the *separation principle* between estimation and control, and in this particular case it arises from the fact that the covariance matrix of the information state does not depend on the history of observations and actions — in fact, it does not even depend on the policy! Now, since the covariance matrix K_t quantifies the uncertainty the controller has about the state X_t, the fact that K_t is unaffected by the control actions means that the only effect the control action has is on the information state $\widehat{X}_t$, not on the amount of uncertainty the controller has about the true state.

This is quite a fortuitous state of affairs; in general, though, when faced with a partially observed control problem, the control action may affect the belief state (i.e., the posterior probability law of the true state given the currently available information) in an arbitrary way. When this is the case, one speaks of the *dual effect* of control: the control action U_t applied at each time t affects both the true state X_t and the controller's uncertainty about X_t. Intuitively, the more uncertainty the controller has about the true state, the harder the controller's job becomes. In the partially observed LQR case, the belief state is Gaussian (this is even true for an arbitrary policy g). A Gaussian probability law is uniquely specified by the mean vector and the covariance matrix, and the controller's uncertainty about X_t is captured by the latter. The absence of the dual effect in the partially observed LQG model is explained by the fact that this covariance matrix is not affected by the actions of the controller.