CS 295: Optimal Control and Reinforcement Learning

Winter 2020

Lecture 4: Stochastic Optimal Control

Roy Fox

Department of Computer Science

Bren School of Information and Computer Sciences

University of California, Irvine

1 LQR with Gaussian process noise

In the *stochastic control* setting, the state transition is not deterministic, but stochastic. Since the space is continuous, the distribution of x_{t+1} given x_t and u_t should be continuous as well, with some density function $p(x_{t+1}|x_t, u_t)$. To continue looking at a family of models that is principled in physics and engineering, and that has nice computational properties, we consider the Gaussian distribution.

Suppose that

$$x_{t+1} = Ax_t + Bu_t + \omega_t,$$

where ω_t is Gaussian noise with mean 0 and covariance Σ_ω (again we omit any drift in the mean of the noise for the sake of simplicity). The variables ω_t for different times are taken to be independent and identically distributed, similar to the Markov property of MDPs. The noise corresponds to fluctuation in the state transition, i.e. uncertainty that results from the state being described at some intermediate level of accuracy, and not in full physical detail. In the continuous-time limit, this state transition equation becomes the Langevin equation, with Bu as an external field.

Now we are interested in minimizing the *expected* cost-to-go

$$\begin{aligned}\mathcal{J}_t(x_t, u) = \sum_{t'=t}^{T-1} \mathbb{E}[c(x_{t'}, u_{t'})|x_t, u] &= \sum_{t'=t}^{T-1} \mathbb{E}[\tfrac{1}{2}x_{t'}^\intercal Q x_{t'} + \tfrac{1}{2}u_{t'}^\intercal R u_{t'}|x_t, u] \\ &= \mathbb{E}[\tfrac{1}{2}x_t^\intercal Q x_t + \tfrac{1}{2}u_t^\intercal R u_t + \mathcal{J}_{t+1}(x_{t+1}, u)|x_t, u_t].\end{aligned}$$

The Bellman equation becomes

$$\begin{aligned}\mathcal{J}_t^*(x_t) &= \min_{u_t} \mathbb{E}_{x_{t+1}|x_t,u_t \sim \mathcal{N}(Ax_t+Bu_t,\Sigma_\omega)}[c(x_t, u_t) + \mathcal{J}_{t+1}^*(x_{t+1})] \\ &= \min_{u_t} \mathbb{E}_{\omega_t \sim \mathcal{N}(0,\Sigma_\omega)}[\tfrac{1}{2}x_t^\intercal Q x_t + \tfrac{1}{2}u_t^\intercal R u_t + \mathcal{J}_{t+1}^*(Ax_t + Bu_t + \omega_t)].\end{aligned}$$

Again we'll solve it for a T-step finite horizon, at the same time that we prove by induction that $\mathcal{J}_t^*$ is quadratic, this time with a constant term. Again $\mathcal{J}_T^* = 0$. Suppose that

$$\mathcal{J}_{t+1}^*(x_{t+1}) = \tfrac{1}{2}x_{t+1}^\intercal S_{t+1}x_{t+1} + \mathcal{J}_{t+1}^*(0),$$

with some positive semidefinite Hessian S_{t+1}.

An important fact about a random vector x with mean μ_x and covariance Σ_x is that for any matrix S

$$\mathbb{E}[x^\intercal Sx] = \mu_x^\intercal S\mu_x + \operatorname{tr}(S\,\Sigma_x).$$

So

$$\begin{aligned}\mathcal{J}_t^*(x_t) &= \min_{u_t} \mathbb{E}_{x_{t+1}|x_t,u_t\sim\mathcal{N}(Ax_t+Bu_t,\Sigma_\omega)}[\tfrac{1}{2}x_t^\intercal Qx_t + \tfrac{1}{2}u_t^\intercal Ru_t + \tfrac{1}{2}x_{t+1}^\intercal S_{t+1}x_{t+1} + \mathcal{J}_{t+1}^*(0)] \\ &= \min_{u_t}(\tfrac{1}{2}x_t^\intercal Qx_t + \tfrac{1}{2}u_t^\intercal Ru_t + \tfrac{1}{2}(Ax_t+Bu_t)^\intercal S_{t+1}(Ax_t+Bu_t)) + \tfrac{1}{2}\operatorname{tr}(S_{t+1}\,\Sigma_\omega) + \mathcal{J}_{t+1}^*(0).\end{aligned}$$

The minimization objective is the same as in the deterministic case, so the control policy is the same, and so is the Hessian S_t. The only difference is that now $\mathcal{J}_t^*$ has an additional constant term

$$\mathcal{J}_t^*(x_t) = \tfrac{1}{2}x_t^\intercal S_t x_t + \sum_{t'=t+1}^{T} \tfrac{1}{2}\operatorname{tr}(S_{t'}\,\Sigma_\omega).$$

This constant is a noise–cost term, representing the cost-to-go of the process noise. It cannot be controlled by the immediate control signal u_t, although its magnitude is determined by $S_{t'}$, which assumes future optimal control.

In the infinite-horizon setting, under some conditions, the Hessian will converge to a self-consistent solution of the Ricatti equation, which is then called an *algebraic Ricatti equation*

$$S = Q + A^\intercal(S - SB(R + B^\intercal SB)^{-1}B^\intercal S)A.$$

Then the relative weight of the term $\frac{1}{2}x^\intercal Sx$ in the cost-to-go tends to 0, and the average cost becomes $\frac{1}{2}\operatorname{tr}(S\,\Sigma_\omega)$. Note how the average infinite cost doesn't depend on the state in any finite time.

2 Linear–Quadratic Estimation (LQE)

In a continuous space problem, when observability is partial, the agent's *belief* of what the state may be is a continuous distribution over the world state given the agent's past observations. We again focus on the simplest interesting case, where the dynamics are linear and the distributions Gaussian, and for now we assume that the system is uncontrollable. This model is a special case of a Hidden Markov Model (HMM), but with continuous state and observation spaces, instead of discrete ones.

In addition to the Gaussian process noise, we now also have Gaussian observation noise. Let the observation in time t, denoted by $y_t \in \mathbb{R}^k$, be given by $y_t = Cx_t + \psi_t$, where $C \in \mathbb{R}^{k\times n}$, and the observation noise ψ_t is Gaussian with mean 0 and covariance matrix Σ_ψ. All the noises ω_t, ψ_t are independent random variables (they have no parents in the Bayesian network describing the process).

The whole stochastic process of states and observations can be expressed as a giant linear transformation of x_0, plus some high-dimensional Gaussian noise. The process is therefore a *Gaussian process*: all the variables are *jointly Gaussian*, while also keeping their Markovian independence properties.

An important property of jointly Gaussian variables, is that their conditional distributions are also Gaussian. If x and y are jointly Gaussian with mean $\begin{bmatrix}\mu_x\\ \mu_y\end{bmatrix}$ and covariance $\begin{bmatrix}\Sigma_x & \Sigma_{xy}\\ \Sigma_{yx} & \Sigma_y\end{bmatrix}$, where Σ_y has full rank, we have the following formula for the mean and the covariance of x when $y = y_0$ is given:

$$\begin{aligned}\mu_{x|y_0} &= \mathbb{E}[x|y_0] = \mu_x + \Sigma_{xy}\,\Sigma_y^{-1}(y_0 - \mu_y)\\ \Sigma_{x|y_0} &= \operatorname{cov}[x|y_0] = \Sigma_x - \Sigma_{xy}\,\Sigma_y^{-1}\,\Sigma_{yx}\,.\end{aligned}$$

One way to prove this is by direct computation from the density function of the Gaussian distribution. An easier way is to see that the Gaussian variable $z = x + \Sigma_{xy}\,\Sigma_y^{-1}(y_0 - y)$, representing "$x|y_0$", is independent of y:

$$\operatorname{cov}[z, y] = \operatorname{cov}[x + \Sigma_{xy}\,\Sigma_y^{-1}(y_0 - y), y] = \operatorname{cov}[x, y] - \Sigma_{xy}\,\Sigma_y^{-1}\operatorname{cov}[y] = 0.$$

This means that

$$\begin{aligned}\mu_{x|y_0} &= \mathbb{E}[x|y_0] = \mathbb{E}[z - \Sigma_{xy}\,\Sigma_y^{-1}(y_0 - y)|y = y_0] = \mathbb{E}[z|y_0]\\ &= \mathbb{E}[z] = \mathbb{E}[x] + \Sigma_{xy}\,\Sigma_y^{-1}(y_0 - \mathbb{E}[y]) = \mu_x + \Sigma_{xy}\,\Sigma_y^{-1}(y_0 - \mu_y),\end{aligned}$$

and

$$\begin{aligned}\Sigma_{x|y_0} &= \operatorname{cov}[x|y_0] = \operatorname{cov}[z - \Sigma_{xy}\,\Sigma_y^{-1}(y_0 - y)|y = y_0] = \operatorname{cov}[z|y_0]\\ &= \operatorname{cov}[z] = \operatorname{cov}[x] - \operatorname{cov}[x, y]\,\Sigma_y^{-1}\,\Sigma_{yx} - \Sigma_{xy}\,\Sigma_y^{-1}\operatorname{cov}[y, x] + \Sigma_{xy}\,\Sigma_y^{-1}\operatorname{cov}[y]\,\Sigma_y^{-1}\,\Sigma_{yx}\\ &= \Sigma_x - \Sigma_{xy}\,\Sigma_y^{-1}\,\Sigma_{yx},\end{aligned}$$

and similarly higher normalized moments are 0, as required.

The algorithm that performs *belief propagation*, that is, updates the belief with each new observation, is called the *Kalman filter* in this case (in continuous time: the *Kalman-Bucy filter*), and the belief it keeps is called a *Linear–Quadratic estimator* (LQE).

The posterior distribution $b_t(x_t|y_{\leqslant t})$, and similarly the predictive distribution $b'_t(x_t|y_{<t})$, are Gaussian. This means that they can be represented by their means $\hat{x}_t$ and $\hat{x}'_t$, and their covariances Σ_t and Σ'_t, respectively. We can compute these sequentially with Bayesian inference, but we need to see that the computation is practical.

The prediction step is

$$\begin{aligned}\hat{x}'_{t+1} &= \mathbb{E}[x_{t+1}|y_{\leqslant t}] = \mathbb{E}[Ax_t + \omega_t|y_{\leqslant t}] = A\hat{x}_t\\ \Sigma'_{t+1} &= \operatorname{cov}[x_{t+1}|y_{\leqslant t}] = \operatorname{cov}[Ax_t + \omega_t|y_{\leqslant t}] = A\,\Sigma_t\,A^\intercal + \Sigma_\omega\,.\end{aligned}$$

The update step computes the mean

$$\begin{aligned}\hat{x}_t &= \mathbb{E}[x_t|y_{\leqslant t}] = \mathbb{E}[(x_t|y_{<t})|(y_t|y_{<t})]\\ &= \mathbb{E}[x_t|y_{<t}] + \operatorname{cov}[x_t, y_t|y_{<t}]\operatorname{cov}[y_t|y_{<t}]^{-1}(y_t - \mathbb{E}[y_t|y_{<t}])\\ &= \mathbb{E}[x_t|y_{<t}] + \operatorname{cov}[x_t|y_{<t}]C^\intercal(C\operatorname{cov}[x_t|y_{<t}]C^\intercal + \Sigma_\psi)^{-1}(y_t - C\,\mathbb{E}[x_t|y_{<t}])\\ &= \hat{x}'_t + \Sigma'_t\,C^\intercal(C\,\Sigma'_t\,C^\intercal + \Sigma_\psi)^{-1}(y_t - C\hat{x}'_t).\end{aligned}$$

Recall our assumption that Σ_ψ is full-rank, otherwise the observation can be represented in a lower dimension. The term $e_t = y_t - C\hat{x}'_t$ is the *prediction error* (also called *innovation*), i.e. the error in predicting y_t based on previous observations. To update the covariance we compute

$$\begin{aligned}\Sigma_t &= \operatorname{cov}[x_t|y_{\leqslant t}] = \operatorname{cov}[(x_t|y_{<t})|(y_t|y_{<t})]\\ &= \operatorname{cov}[x_t|y_{<t}] - \operatorname{cov}[x_t, y_t|y_{<t}]\operatorname{cov}[y_t|y_{<t}]^{-1}\operatorname{cov}[y_t, x_t|y_{<t}]\\ &= \operatorname{cov}[x_t|y_{<t}] - \operatorname{cov}[x_t|y_{<t}]C^\intercal(C\operatorname{cov}[x_t|y_{<t}]C^\intercal + \Sigma_\psi)^{-1}C\operatorname{cov}[x_t|y_{<t}]\\ &= \Sigma'_t - \Sigma'_t\,C^\intercal(C\,\Sigma'_t\,C^\intercal + \Sigma_\psi)^{-1}C\,\Sigma'_t\,.\end{aligned}$$

Again we see that the Bayesian belief can be computed sequentially, using only the previous estimator and the current observation. It involves simple linear algebra, with the most computation-intensive operator being matrix inversion. The inference step consists of updating the mean vector linearly

$$\hat{x}_t = A\hat{x}_{t-1} + K_t e_t = (I - K_t C)A\hat{x}_{t-1} + K_t y_t,$$

where K_t is the *Kalman gain*

$$K_t = \Sigma'_t\,C^\intercal(C\,\Sigma'_t\,C^\intercal + \Sigma_\psi)^{-1},$$

and updating the predictive covariance matrix with a Ricatti equation

$$\Sigma'_{t+1} = A(\Sigma'_t - \Sigma'_t\,C^\intercal(C\,\Sigma'_t\,C^\intercal + \Sigma_\psi)^{-1}C\,\Sigma'_t)A^\intercal + \Sigma_\omega\,.$$

Note that no actual observations are needed to compute the estimator covariance. The prior mean is 0 and the prior covariance is

$$\Sigma_{x_{t+1}} = A\,\Sigma_{x_t}\,A^\intercal + \Sigma_\omega\,.$$

Obviously the estimator is improved and the covariance reduced by considering the observations. But we don't need the observations in order to compute *by how much* the estimator is improved. In an infinite horizon, the covariance is the solution to the algebraic Ricatti equation

$$\Sigma' = A(\Sigma' - \Sigma'\,C^\intercal(C\,\Sigma'\,C^\intercal + \Sigma_\psi)^{-1}C\,\Sigma')A^\intercal + \Sigma_\omega,$$

which can be hard-coded into the agent, eliminating the need to invert matrices in real time during execution.

The mean $\hat{x}_t$ is an unbiased estimator for x_t, so

$$\hat{x}_t = x_t + \epsilon_t,$$

where ϵ_t is some Gaussian estimation noise with mean 0 and covariance Σ_{ϵ_t}. The Bayesian estimator is a sufficient statistic of the observable history for this hidden state, which implies that it minimizes the variance $x_t^\intercal \Sigma_t x_t$ along any component x_t, and so it minimizes the estimator covariance $\Sigma_{\hat{x}_t}$, i.e. the estimation noise Σ_{ϵ_t}.

In fact, there's a deep connection between inference and control, and in particular between sufficient inference and optimal control. Let's take another look at the recursions for the estimator covariance and the cost Hessian

$$\begin{aligned}\Sigma'_{t+1} &= A(\Sigma'_t - \Sigma'_t C^\intercal (C \Sigma'_t C^\intercal + \Sigma_\psi)^{-1} C \Sigma'_t) A^\intercal + \Sigma_\xi \\ S_t &= Q + A^\intercal (S_{t+1} - S_{t+1} B (R + B^\intercal S_{t+1} B)^{-1} B^\intercal S_{t+1}) A.\end{aligned}$$

Obviously they have a common structure, so that one problem can be mapped to the other by mapping their components. This is a classic demonstration of the *duality* between inference and control, summed up in the following table.

LQR	LQE
backward	forward
S_{T-t}	Σ'_t
A	$A^\intercal$
B	$C^\intercal$
Q	Σ_ξ
R	Σ_ψ

A different view of the duality can be obtained through a variant of the Kalman filter called the *information filter*, where instead of the estimator covariance Σ'_t we compute its inverse, the estimator *precision*. The resulting duality between S and $(\Sigma')^{-1}$ is actually the better one to consider.

3 The full Linear–Quadratic–Gaussian (LQG) setting

Putting together inference and control, we get a Partially Observable Markov Decision Process (POMDP). The family of continuous-space POMDPs with linear dynamics, quadratic cost rate and Gaussian noises is called LQG. LQG is the only widely-applicable class of POMDPs for which we have a complete analytic solution.

The dynamics now include the control signal

$$x_{t+1} = Ax_t + Bu_t + \omega_t \qquad \omega_t \sim \mathcal{N}(0, \Sigma_\omega).$$

This is like stochastic control, but observability is now partial

$$y_t = Cx_t + \psi_t \qquad \psi_t \sim \mathcal{N}(0, \Sigma_\psi).$$

So we need both an inference policy, which we already suspect will optimally be linear

$$\hat{x}_t = G_t \hat{x}_{t-1} + K_t y_t,$$

and a control policy, which we also suspect will optimally be linear, but can now only depend on the belief, not directly on the hidden state

$$u_t = L_t \hat{x}_t.$$

This set of 4 equations completely determines the stochastic dynamics of the entire system, and our objective is to minimize a total cost with the same rate

$$c(x_t, u_t) = \tfrac{1}{2} x_t^\intercal Q x_t + \tfrac{1}{2} u_t^\intercal R u_t.$$

The sufficient Bayesian inference is just like before, except we have an additional control term Bu_t to take into account. When the control signal is given, either nominally or as a function of the estimator, it adds a constant drift to the process, affecting the mean but not the posterior covariance. We now have the prediction

$$\hat{x}'_{t+1} = A\hat{x}_t + Bu_t$$

and the update

$$\hat{x}_t = A\hat{x}_{t-1} + Bu_{t-1} + K_t e_t = (I - K_t C)(A\hat{x}_{t-1} + Bu_{t-1}) + K_t y_t,$$

with the same K_t and Σ'_t as before.

The cost-to-go $\mathcal{J}$ can be expressed in terms of the state of the system, which consists of both the world state x_t and the memory state, the estimator $\hat{x}_t$. In principle, the memory state is a belief state, so it also includes the posterior covariance Σ'_t. However, we saw that Σ'_t doesn't depend on any actual states or observations, so it's not a random variable, and any dependence on it is just in the form of the cost-to-go function $\mathcal{J}$, which is now

$$\mathcal{J}_t(x_t, \hat{x}_t, u) = \sum_{\tau=t}^{T-1} \mathbb{E}[c(x_\tau, u_\tau) | x_t, \hat{x}_t] = \mathbb{E}[c(x_t, u_t) + \mathcal{J}_{t+1}(x_{t+1}, \hat{x}_{t+1}, u) | x_t, \hat{x}_t].$$

There's a delicate point here which is often glossed over. It's important that we took the full system state in this recursive equation. Given $(x_t, \hat{x}_t)$, the past and the future of the process are completely independent, so the recursive term $\mathbb{E}[\mathcal{J}_{t+1}(x_{t+1}, \hat{x}_{t+1}, u) | x_t, \hat{x}_t]$ properly describes the dynamics of the process. If we tried, for example, to write down a recursive equation for the expected cost given only the world state, we would get

$$\mathbb{E}[\mathcal{J}_t(x_t, \hat{x}_t, u) | x_t] = \mathbb{E}[c(x_t, u_t) + \mathcal{J}_{t+1}(x_{t+1}, \hat{x}_{t+1}, u) | x_t].$$

We can't directly compute the last term from the recursive term $\mathbb{E}[\mathcal{J}_{t+1}(x_{t+1}, \hat{x}_{t+1}, u) | x_{t+1}]$, because x_t and $\hat{x}_{t+1}$ are not independent given x_{t+1}.

On the other hand, taking the full state in the Bellman equation for $\mathcal{J}^*$ is incorrect as well. The minimum in the Bellman equation is taken given the specific state for which we optimize, but u_t can't be optimized for $(x_t, \hat{x}_t)$, because it can only depend on $\hat{x}_t$.

The trick is that the cost recursion *can* be written in terms of $\hat{x}_t$ alone, when the inference is sufficient. We have

$$\begin{aligned}\bar{\mathcal{J}}_t(\hat{x}_t, u) &= \mathbb{E}[\mathcal{J}_t(x_t, \hat{x}_t, u)|\hat{x}_t] = \mathbb{E}[c(x_t, u_t) + \mathcal{J}_{t+1}(x_{t+1}, \hat{x}_{t+1}, u)|\hat{x}_t] \\ &= \mathbb{E}[c(x_t, u_t) + \bar{\mathcal{J}}_{t+1}(\hat{x}_{t+1})|\hat{x}_t],\end{aligned}$$

where the last step follows from the fact that $\hat{x}_{t+1}$ is a sufficient statistic of the observable history for x_{t+1}, and so x_{t+1} only depends on the previous estimator $\hat{x}_t$ through the updated estimator $\hat{x}_{t+1}$.

The Bellman equation is now

$$\mathcal{J}_t^*(\hat{x}_t) = \min_{u_t} \mathbb{E}_{\substack{x_t|\hat{x}_t \sim \mathcal{N}(\hat{x}_t, \Sigma_t) \\ e_{t+1}|\hat{x}_t \sim \mathcal{N}(0, C\,\Sigma'_{t+1}\,C^\intercal + \Sigma_\psi)}} [c(x_t, u_t) + \mathcal{J}_{t+1}^*(\hat{x}_{t+1})].$$

As we did in LQR, we'll assume and prove by induction that

$$\mathcal{J}_t^*(\hat{x}_t) = \tfrac{1}{2}\hat{x}_t^\intercal S_t \hat{x}_t + \mathcal{J}_t^*(0).$$

Then

$$\begin{aligned}\mathcal{J}_t^*(\hat{x}_t) &= \min_{u_t} \mathbb{E}_{\substack{x_t|\hat{x}_t \sim \mathcal{N}(\hat{x}_t, \Sigma_t) \\ e_{t+1}|\hat{x}_t \sim \mathcal{N}(0, C\,\Sigma'_{t+1}\,C^\intercal + \Sigma_\psi)}} [\tfrac{1}{2}x_t^\intercal Q x_t + \tfrac{1}{2}u_t^\intercal R u_t + \mathcal{J}_{t+1}^*(A\hat{x}_t + Bu_t + K_{t+1}e_{t+1})] \\ &= \min_{u_t}(\tfrac{1}{2}\hat{x}_t^\intercal Q \hat{x}_t + \tfrac{1}{2}u_t^\intercal R u_t + \tfrac{1}{2}(A\hat{x}_t + Bu_t)^\intercal S_{t+1}(A\hat{x}_t + Bu_t)) \\ &\qquad + \tfrac{1}{2}\operatorname{tr}(Q\,\Sigma_t) + \tfrac{1}{2}\operatorname{tr}(S_{t+1}K_{t+1}(C\,\Sigma'_{t+1}\,C^\intercal + \Sigma_\psi)K_{t+1}^\intercal) + \mathcal{J}_{t+1}^*(0).\end{aligned}$$

Interestingly, we have the same optimal control as in LQR, this time as a function of the estimator $\hat{x}_t$ instead of the actual (now hidden) state x_t

$$\begin{aligned}u_t &= L_t\hat{x}_t \\ L_t &= -(R + B^\intercal S_{t+1}B)^{-1}B^\intercal S_{t+1}A.\end{aligned}$$

The cost Hessian S_t is also the same, obtained recursively by the same Ricatti equation as in LQR

$$S_t = Q + A^\intercal(S_{t+1} - S_{t+1}B(R + B^\intercal S_{t+1}B)^{-1}B^\intercal S_{t+1})A.$$

The only difference is that we have an additional constant uncontrollable noise–cost term

$$\mathcal{J^*}_t(0) = \tfrac{1}{2}\sum_{\tau=t}^{T}(\operatorname{tr}(Q\,\Sigma_\tau) + \operatorname{tr}(S_{\tau+1}K_{\tau+1}(C\,\Sigma'_{\tau+1}\,C^\intercal + \Sigma_\psi)K_{\tau+1}^\intercal)).$$

Substituting the optimal control in the equations for sufficient inference, we get the linear inference step

$$\begin{aligned}G_t &= (I - K_tC)(A + BL_t) \\ K_t &= \Sigma'_t\,C^\intercal(C\,\Sigma'_t\,C^\intercal + \Sigma_\psi)^{-1},\end{aligned}$$

with the same Σ'_t as in LQE and the same Kalman gain. We can rewrite the constant noise–cost term as

$$\begin{aligned}\mathcal{J}_t^*(0) &= \tfrac{1}{2}\sum_{\tau=t}^{T}(\operatorname{tr}(Q\,\Sigma_\tau) + \operatorname{tr}(S_{\tau+1}\,\Sigma'_{\tau+1}\,C^\intercal(C\,\Sigma'_{\tau+1}\,C^\intercal + \Sigma_\psi)^{-1}C\,\Sigma'_{\tau+1})) \\ &= \tfrac{1}{2}\sum_{\tau=t}^{T}(\operatorname{tr}(Q\,\Sigma_\tau) + \operatorname{tr}(S_{\tau+1}(\Sigma'_{\tau+1} - \Sigma_{\tau+1}))).\end{aligned}$$

In fully-observable stochastic control, the constant cost term represented the cost-to-go of the process noise. Here it breaks additively into two dual parts: the immediate cost of the uncertainty in x_t accumulated from the past, and the future cost-to-go of the uncertainty that the immediate observation noise adds to $\hat{x}_t$.

In infinite horizon, the cost rate converges to

$$\tfrac{1}{2}\operatorname{tr}(Q\,\Sigma) + \tfrac{1}{2}\operatorname{tr}(S(\Sigma' - \Sigma))),$$

where S and Σ' each solves its own algebraic Ricatti equation, and Σ is computed from Σ'.

The fact that in LQG we have the same cost Hessian and feedback gain as in LQR, and the same estimator covariance and Kalman gain as in LQE, means that the control and estimation parts of LQG can be optimized separately. This *separation principle* is essentially unique to LQG, and is largely responsible for it being a useful model in a variety of applications.