1.7 LQR and the Kalman Filter

Linear Quadratic Regularization (LQR) is a special case of dynamic programming where we have a quadratic objective and a linear dynamic. [Note many smooth dynamics are linear over small time steps and smooth objectives are quadratic close to their minimum.] LQR has a solution with relatively simple form given by the Riccati equation. It can be generalized in a few different ways: random noise and incomplete state information. Even in these settings the optimal control remain essentially the same, however, we may need to replace the variable state, $\boldsymbol{x}$ with its mean $\boldsymbol{x}$, this is called certainty equivalent control. If noise is Gaussian, then estimating $\boldsymbol{x}$ is a relatively straight forward recursion which is given by the Kalman filter. We define and discuss each of these steps in subsequent sections below.

Linear Quadratic Regularization.

Def 90 (Linear Quadratic Regularization). *We consider the following optimization:*

$$\begin{aligned} V_0(\boldsymbol{x}_0) = \text{minimize} \quad & \sum_{t=0}^{T-1} \{\boldsymbol{x}_t^\top R \boldsymbol{x}_t + \boldsymbol{a}_t^\top Q \boldsymbol{a}_t\} + \boldsymbol{x}_T^\top R \boldsymbol{x}_T \\ \text{subject to} \quad & \boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + Ba_{t-1}, \quad t = 1, ..., T \\ \text{over} \quad & \boldsymbol{a}_0, ..., \boldsymbol{a}_{T-1} \end{aligned} \tag{LQR}$$

Here the actions $\boldsymbol{a}$ belong to $\mathbb{R}^m$ and the states $\boldsymbol{x}$ belong to $\mathbb{R}^n$. Here A and B are matrices and R and Q are positive semi-definite matrices.[4]

The objective above is quadratic and its constraints are linear. For this reason, this problem is called a Linear Quadratic Regulator problem and its solution is a Linear Quadratic Regular (LQR).t

Why LQR? This optimization is very common in control. This is because many dynamical systems are approximately linear [over small time steps] and many [smooth] objectives are approximately quadratic when close to their minima. So a wide variety of control problems are approximately LQR problems.

Riccati Equation. An important recursion that is needed to solve LQR problems is the Riccati Equation:

[4] Recall, a matrix M is positive semi-definite if $\boldsymbol{x}^\top M \boldsymbol{x} > 0$ for $x \neq 0$.

Def 91 (Riccati Equation and Gain Matrix). *The Riccati equation is the following matrix recursion*

$$\Lambda_t = R + A^\top \Lambda_{t+1} A - A^\top \Lambda_{t+1} B [Q - B^\top \Lambda_{t+1} B]^{-1} B^\top \Lambda_{t+1} A \qquad \text{(Riccati)}$$

for $t = 0, ..., T-1$ *and with* $\Lambda_T = R$. *The gain matrix is defined to be*

$$G_t = [Q + B^\top \Lambda_t B]^{-1} B^\top \Lambda_t A \qquad \text{(Gain)}$$

Essentially $a = -G_t x$ gives the optimal control at time t.

Solution for LQR. We let $V_\tau(\boldsymbol{x})$ be the optimal solution to (LQR), where the summation is started from time $t = \tau$ in state $\boldsymbol{x}_\tau = \boldsymbol{x}$ The following result gives the solution to an LQR problem.

Thrm 92. *The value function for* (LQR) *satisfies*

$$V_t(\boldsymbol{x}) = \boldsymbol{x}^\top \Lambda_t \boldsymbol{x}$$

where Λ_t *is the solution to the Riccati Equation, see* (Riccati). *Moreover, the optimal control action is given by*

$$\boldsymbol{a}_t^\star = -G_t \boldsymbol{x}_t$$

where G *is the Gain Matrix, see* (Gain).

Proof. The Bellman equation is

$$L_{t-1}(x) = \min_a \{\boldsymbol{x}^\top R \boldsymbol{x} + \boldsymbol{a}^\top Q \boldsymbol{a} + L_t(A\boldsymbol{x} + B\boldsymbol{a})\}$$

We now argue by induction that $L_t(x) = \boldsymbol{x}^\top \Lambda_t \boldsymbol{x}$ for all t. This is certainly true at time T where $L_T(x) = \boldsymbol{x}^\top R \boldsymbol{x}$.

Assuming by induction that $L_t(x) = \boldsymbol{x}^\top \Lambda_t \boldsymbol{x}$, we have that

$$\begin{aligned} L_{t-1}(x) &= \min_a \{\boldsymbol{x}^\top R \boldsymbol{x} + \boldsymbol{a}^\top Q \boldsymbol{a} + (A\boldsymbol{x} + B\boldsymbol{a})^\top \Lambda_t (A\boldsymbol{x} + B\boldsymbol{a})\} \\ &= \min_a \{\boldsymbol{x}^\top R \boldsymbol{x} + \boldsymbol{a}^\top Q \boldsymbol{a} + \boldsymbol{x}^\top A^\top \Lambda_t A \boldsymbol{x} + 2\boldsymbol{a}^\top B^\top \Lambda_t A \boldsymbol{x} + \boldsymbol{a}^\top B^\top \Lambda_t B \boldsymbol{a}\} \end{aligned}$$

Differentiating the above objective with respect to $\boldsymbol{a}$ and setting equal to zero, minimizes the above objective and gives the condition

$$0 = 2Q\boldsymbol{a} + 2B^\top \Lambda_t A \boldsymbol{x} + 2B^\top \Lambda_t B a$$

This implies that the optimal action is

$$\boldsymbol{a}^\star = -[Q + B^\top \Lambda_t B]^{-1} B^\top \Lambda_t A \boldsymbol{x}\,.$$

In other words we see that $\boldsymbol{a}_t^\star = -G_t\boldsymbol{x}_t$, as require above. However, we still need to verify that $V_{t-1}(x) = \boldsymbol{x}^\top\Lambda_{t-1}\boldsymbol{x}$ to complete the induction step. Substituting our expression for $\boldsymbol{a}^\star$ into the above minimization gives.

$$\begin{aligned} L_{t-1}(\boldsymbol{x}) &= \boldsymbol{x}^\top R\boldsymbol{x} + \boldsymbol{x}^\top A^\top \Lambda_t A\boldsymbol{x} + \boldsymbol{a}^{\star\top} Q\boldsymbol{a}^\star + 2\boldsymbol{a}^{\star\top} B^\top \Lambda_t A\boldsymbol{x} + \boldsymbol{a}^{\star\top} B^\top \Lambda_t B\boldsymbol{a}^\star \\ &= \boldsymbol{x}^\top R\boldsymbol{x} + \boldsymbol{x}^\top A^\top \Lambda_t A\boldsymbol{x} - \boldsymbol{x}^\top A^\top \Lambda_t B[Q + B^\top \Lambda_t B]^{-1} B^\top \Lambda_t A\boldsymbol{x} \\ &= \boldsymbol{x}^\top \Lambda_{t-1}\boldsymbol{x} \end{aligned}$$

where the last inequality follows by our definition of Λ_{t-1} from the Riccati equation. □

LQR with Noise.

We consider a small variation on the LQR problem. In particular we assume that x_t is randomly perturbed. We consider following optimization:

$$\begin{aligned} L_0(\boldsymbol{x}_0) = &\text{minimize} \quad \sum_{t=0}^{T-1} \{\boldsymbol{x}_t^\top R\boldsymbol{x}_t + \boldsymbol{a}_t^\top Q\boldsymbol{a}_t\} + \boldsymbol{x}_T^\top R\boldsymbol{x}_T \qquad \text{(Noisy LQR)} \\ &\text{subject to} \quad \boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + Ba_{t-1} + \epsilon_{t-1}, \quad t = 1, ..., T \\ &\text{over} \quad \boldsymbol{a}_0, ..., \boldsymbol{a}_{T-1} \end{aligned}$$

The only change with respect to (LQR) is that we add a random variable ϵ_{t-1}. Here we assume that ϵ_t is independent over time and has mean zero and covariance matrix N. That is

$$\mathbb{E}[\boldsymbol{\epsilon}_t] = \boldsymbol{0} \quad \text{and} \quad \mathbb{E}[\boldsymbol{\epsilon}_t^\top \boldsymbol{\epsilon}_t] = N.$$

The next result shows that optimal control remains the same when we add noise, only the value function changes a little bit.

Thrm 93. *The optimal control for the noisy LQR problem is identical to the LQR problem [without noise] in Theorem 92. The value function now has the form*

$$L_t(x) = \boldsymbol{x}^\top \Lambda_t \boldsymbol{x} + \gamma_t$$

where $\gamma_{t-1} = tr(\Lambda_t N) + \gamma_t$ *and* $\gamma_T = 0$.

Proof. The Bellman equation is

$$L_{t-1}(x) = \min_a \{\boldsymbol{x}^\top R\boldsymbol{x} + \boldsymbol{a}^\top Q\boldsymbol{a} + \mathbb{E}[L_t(A\boldsymbol{x} + B\boldsymbol{a} + \boldsymbol{\epsilon})]\} .$$

We now argue by induction that $L_t(x) = \boldsymbol{x}^\top \Lambda_t \boldsymbol{x} + \gamma_t$ for all t. This is certainly true at time T where $L_T(x) = \boldsymbol{x}^\top R\boldsymbol{x}$.

Assuming by induction that $L_t(x) = \boldsymbol{x}^\top \Lambda_t \boldsymbol{x} + \gamma_t$, we have that

$$\begin{aligned} L_{t-1}(x) &= \min_a \{\boldsymbol{x}^\top R\boldsymbol{x} + \boldsymbol{a}^\top Q\boldsymbol{a} + \mathbb{E}[(A\boldsymbol{x} + B\boldsymbol{a} + \boldsymbol{\epsilon})^\top \Lambda_t (A\boldsymbol{x} + B\boldsymbol{a} + \boldsymbol{\epsilon})] + \gamma_t\} \\ &= \min_a \{\boldsymbol{x}^\top R\boldsymbol{x} + \boldsymbol{a}^\top Q\boldsymbol{a} + \boldsymbol{x}^\top A^\top \Lambda_t A\boldsymbol{x} + 2\boldsymbol{a}^\top B^\top \Lambda_t A\boldsymbol{x} + \boldsymbol{a}^\top B^\top \Lambda_t B\boldsymbol{a}\} \\ &\quad + \mathbb{E}[\boldsymbol{\epsilon}^\top \Lambda_t \boldsymbol{\epsilon}] + \gamma_t \end{aligned}$$

First, observe the minimization above is identical to the LQR minimization in Theorem 92, and so equals $\boldsymbol{x}\Lambda_{t-1}\boldsymbol{x}$ by the proof given in Theorem 92. Second observe, a quick calculation shows that

$$\mathbb{E}[\boldsymbol{\epsilon}^\top \Lambda_t \boldsymbol{\epsilon}] = \sum_{ij} \mathbb{E}[\epsilon_i \epsilon_j \Lambda_{t,ij}] = tr(N\Lambda).$$

Thus $\gamma_t + \mathbb{E}[\boldsymbol{\epsilon}^\top \Lambda_t \boldsymbol{\epsilon}] = \gamma_{t-1}$ as defined above. These two observations give that

$$L_{t-1}(x) = \boldsymbol{x}^\top \Lambda_{t-1}\boldsymbol{x} + \gamma_{t-1}$$

as required. □

Linear Quadratic Gaussian.

We consider a Linear Quadratic Regularization problem but were there is both noise and imperfect state observation. In particular, we do not directly observe the state $\boldsymbol{x}$ but instead some measurement $\boldsymbol{y}$ which we must use to control $\boldsymbol{x}$. Further both $\boldsymbol{x}$ and $\boldsymbol{y}$ are subject to noise.

$$\begin{aligned} L_0(\boldsymbol{x}_0) = &\text{minimize} && \sum_{t=0}^{T-1} \{\boldsymbol{x}_t^\top R\boldsymbol{x}_t + \boldsymbol{a}_t^\top Q\boldsymbol{a}_t\} + \boldsymbol{x}_T^\top R\boldsymbol{x}_T \qquad \text{(LQG)} \\ &\text{subject to} && \boldsymbol{x}_t = A\boldsymbol{x}_{t-1} + B\boldsymbol{a}_{t-1} + \boldsymbol{\epsilon}_{t-1}, \\ & && \boldsymbol{y}_t = C\boldsymbol{x}_{t-1} + \boldsymbol{\delta}_{t-1} \quad t = 1, \dots, T \\ &\text{over} && \boldsymbol{a}_0, \dots, \boldsymbol{a}_{T-1} \end{aligned}$$

In addition to terms in definition of LQR and LQR with Noise, we introduce a matrix C and a noise term δ_t which is independent over time, has mean zero and covariance M, that is

$$\mathbb{E}[\boldsymbol{\delta}] = 0, \qquad \text{and} \qquad \mathbb{E}[\boldsymbol{\delta}\boldsymbol{\delta}^\top] = M.$$

Later [when considering the Kalman Filter], we will need the random variables for δ and ϵ to be Gaussian [hence the name LQG] but we do not require this yet.

Here the state $\boldsymbol{x}$ is not directly observed. So we must base decisions from data of past decision and measurements, that is

$$F_t = (\boldsymbol{y}_t, ..., \boldsymbol{y}_1, \boldsymbol{a}_{t-1}, ..., \boldsymbol{a}_0)\,.$$

Result on LQG. The key result on LQG is that if we can estimate the mean state given F_t, i.e. $\boldsymbol{x}_t := \mathbb{E}[\boldsymbol{x}|F_t]$, then the optimal control is that same as from the LQR problem i.e. $a^\star = -G\bar{x}_t$.

Thrm 94. *For an LQG problem the optimal control at time t is*

$$a^\star_t = -G_t\bar{\boldsymbol{x}}_t$$

where $\bar{\boldsymbol{x}}_t = \mathbb{E}[\boldsymbol{x}_t|F_t]$ *and* G_t *is* (Gain)*. Further, the optimal value function is*

$$L_t(F_t) = \mathbb{E}[\boldsymbol{x}_t^\top \Lambda_t \boldsymbol{x}_t|F_t] + I_t + \gamma_t$$

where

$$I_t = \sum_{\tau=t}^{T-1} \mathbb{E}\Big[\Delta_\tau^\top (R + A^\top \Lambda_{\tau+1} A - \Lambda_\tau)\Delta_\tau|F_t\Big] \qquad \textit{and} \qquad \Delta_t = \boldsymbol{x}_t - \bar{\boldsymbol{x}}_t\,.$$

Before proving this result, we take a moment to discuss.

Certainty Equivalence. The last result is interesting because even though there is noise and we do not observe the system state. We still apply the same control as in the case where we have full information for a deterministic system. When we treat the mean as if it was the "true" state, we call this *certainty equivalence*.

In general applying a certainty equivalent estimate is not optimal, bit it is for LQG systems. So why is certainty equivalence optimal here. In particular, if we look at the new term I_t in the value function, it looks like we need to estimate future values of Δ_τ which in principle should depend on the future actions and states that we visit. This would likely make for a complex dependence on the current action taken. However, it turns out, because the problem is linear, that Δ_t does not depend on the actions and states taken. So in the Bellman equation I_t is effectively a constant as far as the action taken is concerned. This simplifies the problem considerably and means we are still within the scope of our original LQR solution.

The following lemma shows that Δ_t does not depend on actions taken and states visited.

Lemma 5. *Δ_τ is a constant with respect to $\boldsymbol{a}_0, \dots, \boldsymbol{a}_\tau$.*

Proof. We recursively consider the update equation for $\boldsymbol{x}_\tau$. Note that

$$\begin{aligned}\boldsymbol{x}_\tau &= A\boldsymbol{x}_{\tau-1} + B\boldsymbol{a}_{\tau-1} + \boldsymbol{\epsilon}_{\tau-1} \\ &= A[A\boldsymbol{x}_{\tau-2} + B\boldsymbol{a}_{\tau-2} + \boldsymbol{\epsilon}_{\tau-2}] + B\boldsymbol{a}_{\tau-1} + \boldsymbol{\epsilon}_{\tau-1} \\ &= A^2\boldsymbol{x}_{\tau-2} + AB\boldsymbol{a}_{\tau-2} + B\boldsymbol{a}_{\tau-1} + A\boldsymbol{\epsilon}_{\tau-2} + \boldsymbol{\epsilon}_{\tau-1} \\ &\vdots \\ &= A^\tau \boldsymbol{x}_0 + \sum_{t=0}^{\tau-1} A^{\tau-1-t}B\boldsymbol{a}_t + \sum_{t=0}^{\tau-1} A^{\tau-1-t}\boldsymbol{\epsilon}_t\,.\end{aligned}$$

Consequently notice,

$$\bar{\boldsymbol{x}}_\tau = \mathbb{E}[\boldsymbol{x}_\tau | F_\tau] = A^\tau \boldsymbol{x}_0 + \sum_{t=0}^{\tau-1} A^{\tau-1-t}B\boldsymbol{a}_t + \mathbb{E}\Big[\sum_{t=0}^{\tau-1} A^{\tau-1-t}\boldsymbol{\epsilon}_t \Big| F_\tau\Big]$$

So

$$\boldsymbol{x}_\tau - \bar{\boldsymbol{x}}_\tau = \sum_{t=0}^{\tau-1} A^{\tau-1-t}\boldsymbol{\epsilon}_t - \mathbb{E}\Big[\sum_{t=0}^{\tau-1} A^{\tau-1-t}\boldsymbol{\epsilon}_t \Big| F_\tau\Big].$$

It seems like we are done, we have removed all dependence on the actions taken. But remember $F_t = (\boldsymbol{y}_t, \dots, \boldsymbol{y}_1, \boldsymbol{a}_{t-1}, \dots, \boldsymbol{a}_0)$ which we condition on above. In principle, we could modify the set of actions that we take to infer information about $\sum_{t=0}^{\tau-1} A^{\tau-1-t}\boldsymbol{\epsilon}_t$ and thus there would be dependence on the actions taken in the conditional expectation above. However, this turns out not to be the case.

To see this, first, let $\boldsymbol{y}_t^0$ be the sequence of observations made when actions are chosen to be zero, that is, since

$$\boldsymbol{y}_t = C\boldsymbol{x}_t = CA^\tau \boldsymbol{x}_0 + \sum_{t=0}^{\tau-1} CA^{\tau-1-t}B\boldsymbol{a}_t + \sum_{t=0}^{\tau-1} CA^{\tau-1-t}\boldsymbol{\epsilon}_t$$

then $\boldsymbol{y}_t^0$ is given by

$$\boldsymbol{y}_t^0 = CA^\tau \boldsymbol{x}_0 + \sum_{t=0}^{\tau-1} CA^{\tau-1-t}\boldsymbol{\epsilon}_t + \boldsymbol{\delta}_t$$

Since we know which actions we take we can always construct $\boldsymbol{y}_t^0$ from $\boldsymbol{y}_t$ and vice versa. I.e. conditioning on $F_t = (\boldsymbol{y}_t, \dots, \boldsymbol{y}_1, \boldsymbol{a}_{t-1}, \dots, \boldsymbol{a}_0)$ is the same as conditioning on $F_t^0 = (\boldsymbol{y}_t^0, \dots, \boldsymbol{y}_1^0, \boldsymbol{a}_{t-1}, \dots, \boldsymbol{a}_0)$. Further, suppose we let π be the policy that we use to select the actions. In particular, suppose that $a_t = \pi(F_t)$ where here π is some deterministic function. However given the discussion above, equally we could view actions as a function of F_t^0, that is $a_t = \pi^0(F_t^0)$ where π^0 is again a deterministic. Thus all information required to choose each action is determined by the function π^0 and the vectors $(y_t^0, ..., y_1^0)$. In other words, conditioning on F_t^0 is the same as conditioning on $(y_t^0, ..., y_1^0)$ and the deterministic function π^0. However, π^0 is deterministic and thus independent of the random variable $\sum_{t=0}^{\tau-1} A^{\tau-1-t}\epsilon_t$, thus it plays no role in determining its conditional expectation. In summary we have found that

$$\begin{aligned}\mathbb{E}\Big[\sum_{t=0}^{\tau-1} A^{\tau-1-t}\epsilon_t \Big| F_\tau\Big] &= \mathbb{E}\Big[\sum_{t=0}^{\tau-1} A^{\tau-1-t}\epsilon_t \Big| F_\tau^0\Big] \\ &= \mathbb{E}\Big[\sum_{t=0}^{\tau-1} A^{\tau-1-t}\epsilon_t \Big| y_t^0, ..., y_1^0, \pi^0\Big] = \mathbb{E}\Big[\sum_{t=0}^{\tau-1} A^{\tau-1-t}\epsilon_t \Big| y_t^0, ..., y_1^0\Big]\end{aligned}$$

The right hand expression does not depend on the action taken and so the same is true of

$$x_\tau - \bar{x}_\tau = \sum_{t=0}^{\tau-1} A^{\tau-1-t}\epsilon_t - \mathbb{E}\Big[\sum_{t=0}^{\tau-1} A^{\tau-1-t}\epsilon_t \Big| F_\tau\Big].$$

□

A further slightly more minor observation is the following Lemma.

Lemma 6.

$$\mathbb{E}[x_t^\top M x_t | F_t] = \bar{x}_t^\top M \bar{x}_t - \mathbb{E}[\Delta_t^\top M \Delta_t | F_t]$$

Proof.

$$\begin{aligned}\mathbb{E}[x_t^\top M x_t | F_t] &= \mathbb{E}[(\bar{x}_t + \Delta)^\top M (\bar{x}_t + \Delta) | F_t] \\ &= \bar{x}_t M \bar{x}_t + 2\bar{x}_t M \mathbb{E}[\Delta_t | F_t] + \mathbb{E}[x_t^\top M x_t | F_t] \\ &= \bar{x}_t^\top M \bar{x}_t - \mathbb{E}[\Delta_t^\top M \Delta_t | F_t]\end{aligned}$$

□

Proof of Theorem 94 We can now use the above lemma to prove Theorem 94.

Proof of Theorem 94. The result of the Theorem is certainly true at time T, where $L(F_T) = \boldsymbol{x}_T^\top R \boldsymbol{x}_T$. Let's work back inductively assuming the form $L_{t+1}(F_{t+1})$ holds to prove the result.

$$
\begin{aligned}
L_t(F_t) &= \min_{a_t} \mathbb{E}[\boldsymbol{x}_t^\top R \boldsymbol{x}_t + \boldsymbol{a}_t^\top Q \boldsymbol{a}_t + L_{t+1}(F_{t+1}) | F_t] \\
&= \min_{a_t} \mathbb{E}\Big[\underbrace{\boldsymbol{x}_t^\top R \boldsymbol{x}_t}_{(a)} + \boldsymbol{a}_t^\top Q \boldsymbol{a}_t + \underbrace{\mathbb{E}\Big[\boldsymbol{x}_{t+1}^\top \Lambda_{t+1} \boldsymbol{x}_{t+1} | F_{t+1}\Big]}_{(b)} + \underbrace{I_{t+1} + \gamma_{t+1}}_{(c)} \Big| F_t \Big]
\end{aligned}
$$

Let's deal with the three terms (a), (b) and (c) above.

Firstly, for (a) we have by Lemma 6 that

$$
\mathbb{E}[\boldsymbol{x}_t^\top R \boldsymbol{x}_t | F_t] = \bar{\boldsymbol{x}}_t^\top R \bar{\boldsymbol{x}}_t + \mathbb{E}[\Delta_t^\top R \Delta_t | F_t]
$$

Second, for term (b):

$$
\begin{aligned}
&\mathbb{E}\Big[\boldsymbol{x}_{t+1}^\top \Lambda_{t+1} \boldsymbol{x}_{t+1} \Big| F_{t+1}\Big] \\
&= \mathbb{E}\Big[(A\boldsymbol{x}_t + B\boldsymbol{a}_t + \boldsymbol{\epsilon}_t)^\top \Lambda_{t+1} (A\boldsymbol{x}_t + B\boldsymbol{a}_t + \boldsymbol{\epsilon}_t) \Big| F_{t+1}\Big] \\
&= \mathbb{E}[\boldsymbol{x}_t^\top A^\top \Lambda_{t+1} A \boldsymbol{x}_t | F_t] + 2\bar{\boldsymbol{x}}_t A^\top \Lambda_{t+1} B a_t + \boldsymbol{a} B^\top \Lambda_{t+1} B \boldsymbol{a} + tr(N\Lambda_{t+1}) \\
&= \bar{\boldsymbol{x}}_t^\top A^\top \Lambda_{t+1} A \bar{\boldsymbol{x}}_t + \mathbb{E}[\Delta_t^\top A^\top \Lambda_{t+1} A \Delta_t | F_t] \\
&\quad + 2\bar{\boldsymbol{x}}_t A^\top \Lambda_{t+1} B a_t + \boldsymbol{a} B^\top \Lambda_{t+1} B \boldsymbol{a} + tr(N\Lambda_{t+1})
\end{aligned}
$$

Third, for term (c),

$$
\begin{aligned}
\mathbb{E}[I_{t+1} + \gamma_{t+1} | F_t] &= \sum_{\tau=t+1}^{T-1} \mathbb{E}\left[\mathbb{E}\Big[\Delta_\tau^\top (R + A^\top \Lambda_{\tau+1} A - \Lambda_\tau) \Delta_\tau \Big| F_{t+1}\Big] \Big| F_t \right] + \gamma_{t+1} \\
&= \sum_{\tau=t+1}^{T-1} \mathbb{E}\left[\Delta_\tau^\top (R + A^\top \Lambda_{\tau+1} A - \Lambda_\tau) \Delta_\tau \Big| F_t \right] + \gamma_{t+1}\,.
\end{aligned}
$$

Applying the last three terms to $L_t(F_t)$, above, we get that

$$
\begin{aligned}
&L_t(F_t) \\
&= \min_{a} \{ \bar{\boldsymbol{x}}_t^\top R \bar{\boldsymbol{x}}_t + \boldsymbol{a}^\top Q \boldsymbol{a} + \bar{\boldsymbol{x}}_t^\top A^\top \Lambda_{t+1} A \bar{\boldsymbol{x}}_t + 2\boldsymbol{a}^\top B^\top \Lambda_{t+1} A \bar{\boldsymbol{x}}_t + \boldsymbol{a}^\top B^\top \Lambda_{t+1} B \boldsymbol{a} \} \\
&\quad + \mathbb{E}[\Delta_t^\top [R + A^\top \Lambda_{t+1} A] \Delta_t | F_t] + \sum_{\tau=t+1}^{T-1} \mathbb{E}\left[\Delta_\tau^\top (R + A^\top \Lambda_{\tau+1} A - \Lambda_\tau) \Delta_\tau \Big| F_t \right] \\
&\quad + tr(N\Lambda_{t+1}) + \gamma_{t+1}
\end{aligned}
$$

Critically, we have applied Lemma 5, to take terms involving Δ_t from the minimization. An important consequence is that the minimization above is the same as for deterministic LQR problems. So the optimal control is $a_t^\star = -G\bar{\boldsymbol{x}}_t$, by the same calculation done in Theorem 92. So it is equal to $\bar{\boldsymbol{x}}_t^\top \Lambda_t \bar{\boldsymbol{x}}_t$. So applying this and Lemma , i.e. that $\bar{\boldsymbol{x}}_t \Lambda_t \bar{\boldsymbol{x}}_t = \mathbb{E}[\boldsymbol{x}_t^\top \Lambda_t \boldsymbol{x}_t | F_t] - \mathbb{E}[\Delta_t^\top \Lambda_t \Delta_t | F_t]$. This gives that

$$
\begin{aligned}
&L_t(F_t)\\
&= \mathbb{E}[\boldsymbol{x}_t^\top \Lambda_t \boldsymbol{x}_t | F_t]\\
&\quad + \mathbb{E}[\Delta_t^\top [R + A^\top \Lambda_{t+1} A - \Lambda_t] \Delta_t | F_t] + \sum_{\tau=t+1}^{T-1} \mathbb{E}\left[\Delta_\tau^\top (R + A^\top \Lambda_{\tau+1} A - \Lambda_\tau) \Delta_\tau \middle| F_t\right]\\
&\quad + tr(N\Lambda_{t+1}) + \gamma_{t+1}\\
&= \mathbb{E}[\boldsymbol{x}_t^\top \Lambda_t \boldsymbol{x}_t | F_t] + I_t + \gamma_t\,,
\end{aligned}
$$

where we apply the definitions of I_t and γ_t. This gives the required expression of $L_t(F_t)$. □

Kalman Filter

Kalman filtering (and filtering in general) considers the following setting: we have a sequence of states $\boldsymbol{x}_t$, which evolves under random perturbations over time. Unfortunately we cannot observe $\boldsymbol{x}_t$, we can only observe some noisy function of $\boldsymbol{x}_t$, namely, $\boldsymbol{y}_t$. Our task is to find the best estimate of $\boldsymbol{x}_t$ given our observations of $\boldsymbol{y}_t$.

Consider the equations

$$
\begin{aligned}
\boldsymbol{x}_{t+1} &= A\boldsymbol{x}_t + B\boldsymbol{a}_t + \boldsymbol{\epsilon}_t\\
\boldsymbol{y}_{t+1} &= C\boldsymbol{x}_{t+1} + \boldsymbol{\delta}_{t+1}\,.
\end{aligned}
$$

where $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \Sigma^\epsilon)$, $\boldsymbol{\delta}_{t+1} \sim \mathcal{N}(0, \Sigma^\delta)$ and ϵ_t and ν_t are independent. (We let Σ^ϵ be the sub-matrix of the covariance matrix corresponding to ϵ and so forth...)

The Kalman filter has two update stages: a prediction update and a measurement update. These are

$$
\begin{aligned}
\bar{\boldsymbol{x}}_{t+1|t} &= A\bar{\boldsymbol{x}}_{t+1|t} + Ba_t && \text{(Predict-1)}\\
P_{t+1|t} &= AP_{t|t}A^\top + \Sigma_t^\epsilon && \text{(Predict-2)}
\end{aligned}
$$

and

$$
\begin{aligned}
\bar{\boldsymbol{x}}_{t+1} &= \bar{\boldsymbol{x}}_{t+1|t} + K_t(\boldsymbol{y}_{t+1} - C\bar{\boldsymbol{x}}_{t+1|t}) && \text{(Measure-1)}\\
P_{t+1} &= P_{t+1|t} - K_t C P_{t+1|t} && \text{(Measure-2)}
\end{aligned}
$$

where

$$K_t = P_{t+1|t}C^\top(CP_{t+1|t}C^\top + \Sigma_t^\delta).$$

The matrix K_t is often referred to as the Kalman Gain. Assuming the initial state $\boldsymbol{x}_0$ is known and deterministic $P_{0|0} = 0$ in the above.

We will use the following proposition, which is a standard result on normally distributed random vectors, variances and covariances,

Prop 95. *Let $\boldsymbol{u}$ be normally distributed vector with mean $\bar{\boldsymbol{u}}$ and covariance $\Sigma_{\boldsymbol{u}}$, i.e.*

$$\boldsymbol{u} \sim \mathcal{N}(\bar{\boldsymbol{u}}, \Sigma_{\boldsymbol{u}}).$$

i) For any matrix A and (constant) vector $\boldsymbol{c}$, we have that

$$A\boldsymbol{u} + \boldsymbol{c} \sim \mathcal{N}(A\bar{\boldsymbol{u}} + \boldsymbol{c}, A\Sigma_{\boldsymbol{u}}A^\top).$$

ii) If we take $\boldsymbol{u} = (\boldsymbol{v}, \boldsymbol{w})$ then $\boldsymbol{w}$ conditional on $\boldsymbol{v}$ gives

$$(\boldsymbol{w}|\boldsymbol{v}) \sim \mathcal{N}(\bar{\boldsymbol{w}} + \Sigma_{wv}\Sigma_{vv}^{-1}(\boldsymbol{v} - \bar{\boldsymbol{v}}), \Sigma_{ww} - \Sigma_{wv}\Sigma_{vv}^{-1}\Sigma_{vw})$$

iii) $Var(A\boldsymbol{u}) = A\Sigma_u A^\top$, $Cov(A\boldsymbol{u}, B\boldsymbol{u}) = A\Sigma_u B^\top$.

We can justify the Kalman filtering steps by proving that the conditional distribution of x_{t+1} is given by the Prediction and measurement steps. Specifically we have the following.

Thrm 96.

$$[\boldsymbol{x}_{t+1}|\boldsymbol{y}_{[0:t]}, \boldsymbol{a}_{[0:t]}] \sim \mathcal{N}(\bar{\boldsymbol{x}}_{t+1|t}, P_{t+1|t})$$
$$[\boldsymbol{x}_{t+1}|\boldsymbol{y}_{[0:t+1]}, \boldsymbol{a}_{[0:t]}] \sim \mathcal{N}(\bar{\boldsymbol{x}}_{t+1}, P_{t+1})$$

where $\boldsymbol{y}_{[0:t]} := (\boldsymbol{y}_0, ..., \boldsymbol{y}_t)$ and $\boldsymbol{a}_{[0:t]} := (\boldsymbol{a}_0, ..., \boldsymbol{a}_t)$. Thus

$$\mathbb{E}[\bar{x}_{t+1}|F_{t+1}] = \bar{\boldsymbol{x}}_{t+1}$$

where $\bar{\boldsymbol{x}}_{t+1}$ is given by (Measure-1).

Proof. We show the result by induction supposing that

$$[\boldsymbol{x}_t|\boldsymbol{y}_{[0:t]}, \boldsymbol{a}_{[0:t-1]}] \sim \mathcal{N}(\bar{\boldsymbol{x}}_t, P_t).$$

Since $\boldsymbol{x}_{t+1}$ is a linear function of $\boldsymbol{x}_t$, we have that

$$[\boldsymbol{x}_{t+1}|\boldsymbol{y}_{[0:t]}, \boldsymbol{a}_{[0:t]}] \sim \mathcal{N}(\bar{\boldsymbol{x}}_{t+1|t}, P_{t+1|t})\,.$$

where, by Prop 95ii), we have that

$$\bar{\boldsymbol{x}}_{t+1|t} = A\bar{\boldsymbol{x}}_t + Ba_t\,, \qquad P_{t+1|t} = AP_tA^\top + \Sigma^\epsilon\,.$$

Given $\boldsymbol{y}_{t+1} = C\boldsymbol{x}_{t+1} + \boldsymbol{\delta}_t$, we have by Prop 95iii) that $Var(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{[0,t]}, \boldsymbol{a}_{[0,t]}) = CP_{t+1|t}C^\top$ and $Cov(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1}|\boldsymbol{y}_{[0,t]}, \boldsymbol{a}_{[0,t]}) = P_{t+1|t}C^\top$. Thus

$$[(\boldsymbol{x}_{t+1}, \boldsymbol{y}_{t+1})|\boldsymbol{y}_{[0:t]}, \boldsymbol{a}_{[0:t]}] \sim \mathcal{N}\left([\bar{\boldsymbol{x}}_{t+1|t}, C\bar{\boldsymbol{x}}_{t+1|t}], \begin{bmatrix} P_{t+1|t} & P_{t+1}C^\top \\ CP_{t+1|t} & CP_{t+1}C^\top + \Sigma_t^\delta \end{bmatrix}\right).$$

Thus applying Prop 95ii), we get that

$$\begin{aligned}[x_{t+1}|y_{[0:t+1]}, a_{[0:t]}] &= [[x_{t+1}|y_{[0:t]}, a_{[0:t]}]|y_{t+1}] \\ &\sim \mathcal{N}\Big(\bar{\boldsymbol{x}}_{t+1|t} + P_{t+1|t}C^\top[CP_{t+1}C^\top + \Sigma_t^\delta]^{-1}(y_{t+1} - C\bar{\boldsymbol{x}}_{t+1|t})\,, \\ &\qquad P_{t+1|t} - P_{t+1|t}C^\top[CP_{t+1|t}C^\top + \Sigma_t^\delta]^{-1}CP_{t+1|t}\Big).\end{aligned}$$

That is, as required, $[\boldsymbol{x}_{t+1}|\boldsymbol{y}_{[0:t+1]}, \boldsymbol{a}_{[0:t]}] \sim \mathcal{N}(\bar{\boldsymbol{x}}_{t+1}, P_{t+1})$ for

$$\begin{aligned}\bar{\boldsymbol{x}}_{t+1} &= \bar{\boldsymbol{x}}_{t+1\,|\,t} + K_t(\boldsymbol{y}_{t+1} - C\bar{\boldsymbol{x}}_{t+1\,|\,t}) \\ P_{t+1} &= P_{t+1\,|\,t} - K_tCP_{t+1\,|\,t}\end{aligned}$$

where $K_t = P_{t+1\,|\,t}C^\top(CP_{t+1|t}C^\top + \Sigma_t^\delta)$. □

References

Bucy and Kalman developed the Kalman filter [10]. It is used extensively in control theory, for a recent text see Grenwal and Andrews [17]. For a machine learning and Bayesian perspective see Murphy [27].