Optimal Control for Linear Quadratic Regulators
Recap: Dynamic Programming

\[ \pi^* = \{ \pi_0^*, \pi_1^*, \ldots, \pi_{H-1}^* \} \]

We use Dynamic Programming, and do DP backward in time; start at \( H - 1 \)

\[ Q_{H-1}^*(s, a) = r(s, a) \quad \pi_{H-1}^*(s) = \arg \max_a Q_{H-1}^*(s, a) \]

\[ V_{H-1}^*(s) = \max_a Q_{H-1}^*(s, a) = Q_{H-1}^*(s, \pi_{H-1}^*(s)) \]
Recap: Dynamic Programming

\[ \pi^* = \{\pi_0^*, \pi_1^*, \ldots, \pi_{H-1}^*\} \]

We use Dynamic Programming, and do DP backward in time; start at \( H - 1 \)

\[ Q_{H-1}^*(s, a) = r(s, a) \quad \pi_{H-1}^*(s) = \arg \max_a Q_{H-1}^*(s, a) \]

\[ V_{H-1}^*(s) = \max_a Q_{H-1}^*(s, a) = Q_{H-1}^*(s, \pi_{H-1}^*(s)) \]

Now assume that we have already computed \( V_{h+1}^*, h \leq H - 2 \) (i.e., we know how to perform optimally at \( h + 1 \))

\[ Q_h^*(s, a) = r(s, a) + \mathbb{E}_{s' \sim P(\cdot|s, a)} V_{h+1}^*(s') \]

\[ \pi_h^*(s) = \arg \max_a Q_h^*(s, a) \]
Recap: The Linear Quadratic Regulator (LQR)

\[
\min_{\pi_0, \ldots, \pi_{H-1}} \mathbb{E} \left[ x_H^T Q x_H + \sum_{h=0}^{H-1} (x_h^T Q x_h + u_h^T R u_h) \right]
\]

such that \( x_{h+1} = A x_h + B u_h + w_h, \quad u_h = \pi_h(x_h) \quad x_0 \sim \mu_0, \ w_h \sim N(0, \sigma^2 I), \)

Here, \( x_h \in \mathbb{R}^d, \ u_h \in \mathbb{R}^k, \)

the disturbance \( w_t \in \mathbb{R}^d \) is Gaussian noise

\( A \in \mathbb{R}^{d \times d} \) and \( B \in \mathbb{R}^{d \times k}, \)

\( Q \in \mathbb{R}^{d \times d} \) and \( R \in \mathbb{R}^{k \times k} \) are pd matrices
V/Q functions:

- Value function \( V_h^\pi : \mathbb{R}^d \to \mathbb{R} \) as
  \[
  V_h^\pi(x) = \mathbb{E} \left[ x_H^\top Q x_H + \sum_{t=h}^{H-1} (x_t^\top Q x_t + u_t^\top R u_t) \mid \pi, x_h = x \right],
  \]

- And \( Q_h^\pi : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R} \) as:
  \[
  Q_h^\pi(x, u) = \mathbb{E} \left[ x_H^\top Q x_H + \sum_{t=h}^{H-1} (x_t^\top Q x_t + u_t^\top R u_t) \mid \pi, x_h = x, u_h = u \right],
  \]
Optimal Value functions:

\[ V_h^*(x) = \min_{\pi_h, \pi_{h+1}, \ldots, \pi_{H-1}} \mathbb{E} \left[ x_H^T Q x_H + \sum_{i=h}^{H-1} x_i^T Q x_i + u_i^T R u_i \mid u_t = \pi_t(x_t), x_h = x \right] \]
Optimal Value functions:

\[ V_h^*(x) = \min_{\pi_h, \pi_{h+1}, \ldots, \pi_{H-1}} \mathbb{E} \left[ x_H^T Q x_H + \sum_{i=h}^{H-1} x_i^T Q x_i + u_i^T R u_i \mid u_t = \pi_t(x_t), x_h = x \right] \]

**Theorem:**

\( V_h^* \) is a quadratic function, i.e., \( V_h^*(x) = x^T P_h x + p_h \),

and optimal policy is linear:

\( \pi_h^*(x) = -K_h^* x, \)

and \( P_h \) & \( K_h^* \) can be computed exactly.
Optimal Value functions:

\[ V_h^*(x) = \min_{\pi_h, \pi_{h+1}, \ldots, \pi_{H-1}} \mathbb{E} \left[ x_H^T Q x_H + \sum_{i=h}^{H-1} x_i^T Q x_i + u_i^T R u_i \bigg| u_t = \pi_t(x_t), x_h = x \right] \]

**Theorem:**

\( V_h^* \) is a quadratic function, i.e., \( V_h^*(x) = x^T P_h x + p_h \),

and optimal policy is linear:

\( \pi_h^*(x) = -K_h^* x \),

and \( P_h \) & \( K_h^* \) can be computed exactly.

Today: we prove the above theorem and derive optimal policies.
Key Steps to Deriving Optimal Control

Again, we will do dynamic programming **backward in time**, i.e., from $H$ to 0.
Key Steps to Deriving Optimal Control

Again, we will do dynamic programming \textbf{backward in time}, i.e., from $H$ to 0

1. **Base case:**
   Show that $V_{H}^{*}(x)$ is quadratic for all $x \in \mathbb{R}^{d}$
Key Steps to Deriving Optimal Control

Again, we will do dynamic programming **backward in time**, i.e., from $H$ to 0

1. **Base case:**
   Show that $V^*_H(x)$ is quadratic for all $x \in \mathbb{R}^d$

2. **Inductive hypothesis:** Assume $V^*_{h+1}(x)$ is quadratic $\forall x$:
   - show that $Q^*_h(x, u)$ is quadratic in both $(x, u)$
   - Derive the optimal policy $\pi^*_h(x) = \arg\min_u Q^*_h(x, u)$, and show that it’s linear
Key Steps to Deriving Optimal Control

Again, we will do dynamic programming **backward in time**, i.e., from $H$ to 0

1. **Base case:**
   Show that $V^*_H(x)$ is quadratic for all $x \in \mathbb{R}^d$

2. **Inductive hypothesis:** Assume $V^*_{h+1}(x)$ is quadratic $\forall x$:
   - show that $Q^*_h(x, u)$ is quadratic in both $(x, u)$
   - Derive the optimal policy $\pi^*_h(x) = \arg\min_u Q^*_h(x, u)$, and show that it’s linear

3. **Conclusion:**
   show $V^*_h(x)$ is quadratic for all $x$;
Base case at $H$

Recall our cost functions:

$$\min_{\pi_0, \ldots, \pi_{H-1}} \mathbb{E}\left[ x_H^T Q x_H + \sum_{h=0}^{H-1} (x_h^T Q x_h + u_h^T R u_h) \right]$$
Base case at $H$

Recall our cost functions:

$$\min_{\pi_0, \ldots, \pi_{H-1}} \mathbb{E} \left[ x_H^TQx_H + \sum_{h=0}^{H-1} (x_h^TQx_h + u_h^TRu_h) \right]$$

So, at time step $H$, given $x$, the cost-to-go is $x^TQx$ regardless.

$$V^*_H(x) = x^TQx, \forall x \in \mathbb{R}^d$$
Base case at $H$

Recall our cost functions:

$$\min_{\pi_0, \ldots, \pi_{H-1}} \mathbb{E}\left[ x_H^\top Q x_H + \sum_{h=0}^{H-1} (x_h^\top Q x_h + u_h^\top R u_h) \right]$$

So, at time step $H$, given $x$, the cost-to-go is $x^\top Q x$ regardless.

$$V_H^*(x) = x^\top Q x, \ \forall x \in \mathbb{R}^d$$

Denote $P_H := Q, p_H = 0$,

we write $V_H^*(x) = x^\top P_H x + p_H$

(Goal: derive recursive formulation of $P_h$, $p_h$)
Induction Step:

\[
\begin{align*}
\min_{\pi_0, \ldots, \pi_{H-1}} & \quad E \left[ x_H^T Q x_H + \sum_{h=0}^{H-1} (x_h^T Q x_h + u_h^T R u_h) \right] \\
\text{such that} & \quad x_{h+1} = A x_h + B u_h + w_h, \quad u_h = \pi_h(x_h), \quad x_0 \sim \mu_0, \quad w_h \sim N(0, \sigma^2 I),
\end{align*}
\]

Assume \( V_{h+1}^*(x) = x^T P_{h+1} x + p_{h+1} \), for all \( x \), where \( P_{h+1} \in \mathbb{R}^{d \times d}, p_{h+1} \in \mathbb{R} \)
Induction Step:

\[
\min_{\pi_0, \ldots, \pi_{H-1}} E \left[ x_H^T Q x_H + \sum_{h=0}^{H-1} (x_h^T Q x_h + u_h^T R u_h) \right] \\
\text{such that} \quad x_{h+1} = A x_h + B u_h + w_h, \quad u_h = \pi_h(x_h) \quad x_0 \sim \mu_0, \ w_h \sim N(0, \sigma^2 I),
\]

Assume \( V_{h+1}^*(x) = x^T P_{h+1} x + p_{h+1}, \) for all \( x, \) where \( P_{h+1} \in \mathbb{R}^{d \times d}, p_{h+1} \in \mathbb{R} \)

\[
Q_h^*(x, u) = c(x, u) + \mathbb{E}_{x' \sim P(x,u)} V_{h+1}^*(x')
\]
Induction Step:

\[
\min_{\pi_0, \ldots, \pi_{H-1}} \mathbb{E} \left[ x_H^T Q x_H + \sum_{h=0}^{H-1} (x_h^T Q x_h + u_h^T R u_h) \right]
\]

such that \( x_{h+1} = A x_h + B u_h + w_h, \ u_h = \pi_h(x_h) \), \( x_0 \sim \mu_0, \ w_h \sim N(0, \sigma^2 I) \).

Assume \( V_{h+1}^*(x) = x^T P_{h+1} x + p_{h+1} \), for all \( x \), where \( P_{h+1} \in \mathbb{R}^{d \times d} \), \( p_{h+1} \in \mathbb{R} \).

\[
Q_h^*(x, u) = c(x, u) + \mathbb{E}_{x' \sim P(x, u)} V_{h+1}^*(x')
= x^T Q x + u^T R u + \mathbb{E}_{x' \sim P(x, u)} V_{h+1}^*(x')
\]
Induction Step:

\[
\min_{\pi_0, \ldots, \pi_{H-1}} \mathbb{E} \left[ x_H^T Q x_H + \sum_{h=0}^{H-1} (x_h^T Q x_h + u_h^T R u_h) \right]
\]

such that

\[ x_{h+1} = A x_h + B u_h + w_h, \quad u_h = \pi_h(x_h), \quad x_0 \sim \mu_0, \quad w_h \sim N(0, \sigma^2 I), \]

Assume \( V_{h+1}^*(x) = x^T P_{h+1} x + p_{h+1} \), for all \( x \), where \( P_{h+1} \in \mathbb{R}^{d \times d} \), \( p_{h+1} \in \mathbb{R} \)

\[
Q_h^*(x, u) = c(x, u) + \mathbb{E}_{x' \sim P(x, u)} V_{h+1}^*(x')
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{x' \sim P(x, u)} V_{h+1}^*(x')
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{w \sim N(0, \sigma^2 I)} \left[ V_{h+1}^* (Ax + Bu + w) \right]
\]
Induction Step:

\[
\min_{\pi_0, \ldots, \pi_{H-1}} E \left[ x_H^T Q x_H + \sum_{h=0}^{H-1} (x_h^T Q x_h + u_h^T R u_h) \right]
\]

such that \( x_{h+1} = A x_h + B u_h + w_h, \quad u_h = \pi_h(x_h) \quad x_0 \sim \mu_0, \quad w_h \sim N(0, \sigma^2 I) \),

Assume \( V_{h+1}^*(x) = x^T P_{h+1} x + p_{h+1} \), for all \( x \), where \( P_{h+1} \in \mathbb{R}^{d \times d}, p_{h+1} \in \mathbb{R} \)

\[
Q^*_h(x, u) = c(x, u) + \mathbb{E}_{x' \sim P(x, u)} V^*_{h+1}(x')
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{x' \sim P(x, u)} V^*_{h+1}(x')
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{w \sim \mathcal{N}(0, \sigma^2 I)} [V^*_{h+1} (A x + B u + w)]
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{w \sim \mathcal{N}(0, \sigma^2 I)} [(A x + B u + w)^T P_{h+1} (A x + B u + w) + p_{h+1}]
\]
Induction Step:

\[
\min_{\pi_0, \ldots, \pi_{H-1}} E \left[ x_H^T Q x_H + \sum_{h=0}^{H-1} (x_h^T Q x_h + u_h^T R u_h) \right]
\]

such that \( x_{h+1} = A x_h + B u_h + w_h, \ u_h = \pi_h(x_h), \ x_0 \sim \mu_0, \ w_h \sim N(0, \sigma^2 I), \)

Assume \( V_{h+1}^*(x) = x^T P_{h+1} x + p_{h+1}, \) for all \( x, \) where \( P_{h+1} \in \mathbb{R}^{d \times d}, \ p_{h+1} \in \mathbb{R} \)

\[
Q_h^*(x, u) = c(x, u) + \mathbb{E}_{x' \sim P(x, u)} V_{h+1}^*(x')
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{x' \sim P(x, u)} V_{h+1}^*(x')
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{w \sim \mathcal{N}(0, \sigma^2 I)} \left[ V_{h+1}^* (Ax + Bu + w) \right]
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{w \sim \mathcal{N}(0, \sigma^2 I)} \left[ (Ax + Bu + w)^T P_{h+1} (Ax + Bu + w) + p_{h+1} \right]
\]

\[
= x^T (Q + A^T P_{h+1} A) x + u^T (R + B^T P_{h+1} B) u + 2x^T A^T P_{h+1} Bu + \mathbb{E}_{w \sim \mathcal{N}(0, \sigma^2 I)} \left[ w^T P_{h+1} w \right] + p_{h+1}
\]
Induction Step:

\[
\min_{\pi_0, \ldots, \pi_{H-1}} E \left[ x_H^T Q x_H + \sum_{h=0}^{H-1} (x_h^T Q x_h + u_h^T R u_h) \right]
\]

such that \( x_{h+1} = A x_h + B u_h + w_h \), \( u_h = \pi_h(x_h) \) \( x_0 \sim \mu_0, w_h \sim N(0, \sigma^2 I) \),

Assume \( V^*_h(x) = x^T P_{h+1} x + p_{h+1} \), for all \( x \), where \( P_{h+1} \in \mathbb{R}^{d \times d}, p_{h+1} \in \mathbb{R} \)

\[
Q^*_h(x, u) = c(x, u) + \mathbb{E}_{x' \sim P(x, u)} V^*_h(x')
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{x' \sim P(x, u)} V^*_h(x')
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{w \sim \mathcal{N}(0, \sigma^2 I)} [V^*_h(A x + B u + w)]
\]

\[
= x^T Q x + u^T R u + \mathbb{E}_{w \sim \mathcal{N}(0, \sigma^2 I)} [(A x + B u + w)^T P_{h+1} (A x + B u + w) + p_{h+1}]
\]

\[
= x^T (Q + A^T P_{h+1} A) x + u^T (R + B^T P_{h+1} B) u + 2 x^T A^T P_{h+1} B u + \mathbb{E}_{w \sim \mathcal{N}(0, \sigma^2 I)} [w^T P_{h+1} w] + p_{h+1}
\]

\[
= x^T (Q + A^T P_{h+1} A) x + u^T (R + B^T P_{h+1} B) u + 2 x^T A^T P_{h+1} B u + \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1}
\]
Induction Step (continue)

\[ Q_h^*(x, u) = c(x, u) + \mathbb{E}_{x' \sim P_{(x,u)}}[V_{h+1}^*(x')] \]

\[ = x^\top (Q + A^\top P_{h+1}A) x + u^\top (R + B^\top P_{h+1}B) u + 2x^\top A^\top P_{h+1}Bu + \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]
Induction Step (continue)

\[ Q_h^*(x, u) = c(x, u) + \mathbb{E}_{x' \sim P(x, u)} \left[ V_{h+1}^*(x') \right] \]

\[ = x^\top \left( Q + A^\top P_{h+1} A \right) x + u^\top \left( R + B^\top P_{h+1} B \right) u + 2x^\top A^\top P_{h+1} Bu + \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]

\[ \pi_h^*(x) = \arg \min_u Q_h^*(x, u) \]
Induction Step (continue)

\[ Q_h^*(x, u) = c(x, u) + \mathbb{E}_{x' \sim P(x, u)} \left[ V_{h+1}^*(x') \right] \]

\[ = x^\top \left( Q + A^\top P_{h+1} A \right) x + u^\top \left( R + B^\top P_{h+1} B \right) u + 2x^\top A^\top P_{h+1} Bu + \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]

\[ \pi_h^*(x) = \arg \min_u Q_h^*(x, u) \]

Set \( \nabla_u Q_h^*(x, u) = 0 \), and solve for \( u \):
Induction Step (continue)

\[ Q_h^*(x, u) = c(x, u) + \mathbb{E}_{x' \sim p(x, u)} [V_{h+1}^*(x')] \]
\[ = x^\top (Q + A^\top P_{h+1} A) x + u^\top (R + B^\top P_{h+1} B) u + 2x^\top A^\top P_{h+1} Bu + \text{tr} (\sigma^2 P_{h+1}) + p_{h+1} \]

\[ \pi_h^*(x) = \arg\min_u Q_h^*(x, u) \]

Set \( \nabla_u Q_h^*(x, u) = 0 \), and solve for \( u \):

\[ \pi_h^*(x) = - (R + B^\top P_{h+1} B)^{-1} B^\top P_{h+1} A x \]

\[ := K_h^* \]
Induction Step (continue)

\[ Q_h^*(x, u) = c(x, u) + \mathbb{E}_{x' \sim P(x,u)} \left[ V_{h+1}^*(x') \right] \]

\[ = x^\top \left( Q + A^\top P_{h+1} A \right) x + u^\top \left( R + B^\top P_{h+1} B \right) u + 2x^\top A^\top P_{h+1} Bu + \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]

\[ \pi_h^*(x) = \arg \min_u Q_h^*(x, u) \]

Set \( \nabla_u Q_h^*(x, u) = 0 \), and solve for \( u \):

\[ \pi_h^*(x) = - \left( R + B^\top P_{h+1} B \right)^{-1} B^\top P_{h+1} A x \]

\[ := K_h^* \]

\[ := - K_h^* x \]
Concluding the Induction step:

\[ Q_h^*(x, u) = x^\top (Q + A^\top P_{h+1}A) x + u^\top (R + B^\top P_{h+1}B) u + 2x^\top A^\top P_{h+1}Bu + \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]

\[ \pi_h^*(x) = - (R + B^\top P_{h+1}B)^{-1} B^\top P_{h+1}A x \]

\[ := K_h^* \]
Concluding the Induction step:

\[ Q_h^*(x, u) = x^\top (Q + A^\top P_{h+1}A) x + u^\top (R + B^\top P_{h+1}B) u + 2x^\top A^\top P_{h+1}Bu + \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]

\[ \pi_h^*(x) = - (R + B^\top P_{h+1}B)^{-1} B^\top P_{h+1}A x \]

\[ \underbrace{:=K_h^*}_{\text{:=}} \]

\[ V_h^*(x) = Q_h^*(x, \pi_h^*(x)) \]
Concluding the Induction step:

\[ Q_h^*(x, u) = x^\top (Q + A^\top P_{h+1}A) x + u^\top (R + B^\top P_{h+1}B) u + 2x^\top A^\top P_{h+1}Bu + \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]

\[ \pi_h^*(x) = -(R + B^\top P_{h+1}B)^{-1}B^\top P_{h+1}A x \]

\[ := K_h^* \]

\[ V_h^*(x) = Q_h^*(x, \pi_h^*(x)) \]

We can express \( V_h^*(x) \) as \( V_h^*(x) = x^\top P_h x + p_h \), where:
Concluding the Induction step:

\[ Q^*_h(x, u) = x^T \left( Q + A^T P_{h+1} A \right) x + u^T \left( R + B^T P_{h+1} B \right) u + 2x^T A^T P_{h+1} Bu + \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]

\[ \pi^*_h(x) = - \left( R + B^T P_{h+1} B \right)^{-1} B^T P_{h+1} A x \]

\[ := K^*_h \]

\[ V^*_h(x) = Q^*_h(x, \pi^*_h(x)) \]

We can express \( V^*_h(x) \) as \( V^*_h(x) = x^T P_h x + p_h \), where:

\[ P_h = Q + A^T P_{h+1} A - A^T P_{h+1} B (R + B^T P_{h+1} B)^{-1} B^T P_{h+1} A, \]

\[ p_h = \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]
Concluding the Induction step:

\[ Q_h^*(x, u) = x^T \left( Q + A^T P_{h+1} A \right) x + u^T \left( R + B^T P_{h+1} B \right) u + 2x^T A^T P_{h+1} Bu + \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]

\[ \pi_h^*(x) = - (R + B^T P_{h+1} B)^{-1} B^T P_{h+1} A x \]

\[ := \kappa_h^* \]

\[ V_h^*(x) = Q_h^*(x, \pi_h^*(x)) \]

We can express \( V_h^*(x) \) as \( V_h^*(x) = x^T P_h x + p_h \), where:

\[ P_h = Q + A^T P_{h+1} A - A^T P_{h+1} B(R + B^T P_{h+1} B)^{-1} B^T P_{h+1} A, \quad \text{Ricatti Equation} \]

\[ p_h = \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1} \]
Summary:
Summary:

\( V_H^*(x) = x^\top Q x, \) define \( P_H = Q, p_H = 0, \)
Summary:

\[ V_H^*(x) = x^\top Q x, \text{ define } P_H = Q, p_H = 0, \]

We have shown that \( V_h^*(x) = x^\top P_h x + p_h \), where:

\[
P_h = Q + A^\top P_{h+1} A - A^\top P_{h+1} B (R + B^\top P_{h+1} B)^{-1} B^\top P_{h+1} A,
\]

\[
p_h = \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1}
\]
Summary:

\[ V^*_H(x) = x^\top Q x, \ \text{define} \ P_H = Q, p_H = 0, \]

We have shown that \( V^*_h(x) = x^\top P_H x + p_h \), where:

\[
P_h = Q + A^\top P_{h+1} A - A^\top P_{h+1} B (R + B^\top P_{h+1} B)^{-1} B^\top P_{h+1} A,
\]

\[
p_h = \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1}
\]

Along the way, we also have shown that \( \pi^*_h(x) = -K^*_h x \) where:

\[
\pi^*_h(x) = - \left( R + B^\top P_{h+1} B \right)^{-1} B^\top P_{h+1} A x
\]

\[ :=K^*_h \]
Summary:

$$V^*_H(x) = x^\top Q x, \text{ define } P_H = Q, p_H = 0,$$

We have shown that $$V^*_h(x) = x^\top P_h x + p_h$$, where:

$$P_h = Q + A^\top P_{h+1} A - A^\top P_{h+1} B (R + B^\top P_{h+1} B)^{-1} B^\top P_{h+1} A,$$

$$p_h = \text{tr} (\sigma^2 P_{h+1}) + p_{h+1}$$

Along the way, we also have shown that $$\pi^*_h(x) = - K^*_h x$$ where:

$$\pi^*_h(x) = - (R + B^\top P_{h+1} B)^{-1} B^\top P_{h+1} A x$$

$$:=K^*_h$$

Optimal control has nothing to do with initial distribution, and the noise!
Some Basic Extension:

Time dependent costs and transitions:

\[
\min_{\pi_0, \ldots, \pi_{H-1}} \mathbb{E}\left[ x_H^T Q_H x_H + \sum_{h=0}^{H-1} (x_h^T Q_h x_h + u_h^T R_h u_h) \right]
\]

such that

\[
x_{h+1} = A_h x_h + B_h u_h + w_h, \quad u_h = \pi_h(x_h) \quad x_0 \sim \mu_0, \quad w_h \sim N(0, \sigma^2 I),
\]
Some Basic Extension:

Time dependent costs and transitions:

\[
\begin{aligned}
\min_{\pi_0, \ldots, \pi_{H-1}} \quad & E \left[ x_H^T Q_H x_H + \sum_{h=0}^{H-1} (x_h^T Q_h x_h + u_h^T R_h u_h) \right] \\
\text{such that} \quad & x_{h+1} = A_h x_h + B_h u_h + w_h, \quad u_h = \pi_h(x_h) \quad x_0 \sim \mu_0, \quad w_h \sim N(0, \sigma^2 I), \\
\end{aligned}
\]

Same derivation, we will have the following Ricatti Equation:

\[
\begin{aligned}
P_h &= Q_h + A_h^T P_{h+1} A_h - A_h^T P_{h+1} B_h (R_h + B_h^T P_{h+1} B_h)^{-1} B_h^T P_{h+1} A_h, \\
p_h &= \text{tr} \left( \sigma^2 P_{h+1} \right) + p_{h+1}
\end{aligned}
\]
Some Basic Extension:

More generally...

$$\min_{\pi_0, \ldots, \pi_{H-1}} \mathbb{E} \left[ x_H^T Q_H x_H + x_H^T q_H + c_H + \sum_{h=0}^{H-1} (x_h^T Q_h x_h + u_h^T R_h u_h + u_h^T M_h x_h + x_h^T q_h + u_h^T r_h + c_h) \right]$$

such that $$x_{h+1} = A_h x_h + B_h u_h + v_h + w_h, \ u_h = \pi_h (x_h), \ x_0 \sim \mu_0, \ w_h \sim N(0, \sigma^2 I),$$
Some Basic Extension:

More generally…

\[
\min_{\pi_0, \ldots, \pi_{H-1}} \mathbb{E} \left[ x_H^T Q_H x_H + x_H^T q_H + c_H + \sum_{h=0}^{H-1} (x_h^T Q_h x_h + u_h^T R_h u_h + u_h^T M_h x_h + x_h^T q_h + u_h^T r_h + c_h) \right]
\]

such that

\[
x_{h+1} = A_h x_h + B_h u_h + v_h + w_h, \quad u_h = \pi_h(x_h) \quad x_0 \sim \mu_0, \quad w_h \sim N(0, \sigma^2 I),
\]

Same DP idea and similar derivation
(HW1 question)
Some Basic Extension:

Tracking a pre-defined trajectory:

$$\min_{\pi_0, \ldots, \pi_{H-1}} E \left[ (x_H - x_H^*)^\top Q_H (x_H - x_H^*) + \sum_{h=0}^{H-1} (x_h - x_h^*)^\top Q_h (x_h - x_h^*) + (u_h - u_h^*)^\top R_h (u_h - u_h^*) \right]$$

such that

$$x_{h+1} = A_h x_h + B_h u_h + w_h, \quad u_h = \pi_h(x_h) \quad x_0 \sim \mu_0, \quad w_h \sim N(0, \sigma^2 I),$$
Some Basic Extension:

Tracking a pre-defined trajectory:

\[
\min_{\pi_0, \ldots, \pi_{H-1}} E \left[ (x_H - x_H^*)^\top Q_H (x_H - x_H^*) + \sum_{h=0}^{H-1} (x_h - x_h^*)^\top Q_h (x_h - x_h^*) + (u_h - u_h^*)^\top R_h (u_h - u_h^*) \right]
\]

such that

\[
x_{h+1} = A_h x_h + B_h u_h + w_h, \quad u_h = \pi_h (x_h) \quad x_0 \sim \mu_0, \quad w_h \sim N(0, \sigma^2 I),
\]

We can simply complete the square
and we reduce back to the setting in the previous slide!
Known transition versus black-box access

So far, we studied Policy Evaluation, Policy Iteration, Value Iteration, and DP-based approach, we have assumed that transition is known, i.e., $P(s'|s, a), \forall s, a, s'$ is known, or $A, B, \mathcal{N}(0, \sigma^2 I)$ are known.
So far, we studied Policy Evaluation, Policy Iteration, Value Iteration, and DP-based approach, we have assumed that transition is **known**, i.e., $P(s'|s,a), \forall s,a,s'$ is known, or $A, B, \mathcal{N}(0,\sigma^2I)$ are known

Starting from this Thursday, we start considering **unknown** transition:
Known transition versus black-box access

So far, we studied Policy Evaluation, Policy Iteration, Value Iteration, and DP-based approach, we have assumed that transition is known, i.e., \( P(s'|s,a) \), \( \forall s, a, s' \) is known, or \( A, B, \mathcal{N}(0,\sigma^2 I) \) are known.

Starting from this Thursday, we start considering unknown transition:

We start w/ black-box access to \( P \), or \( f(x, u, w) \):
Known transition versus black-box access

So far, we studied Policy Evaluation, Policy Iteration, Value Iteration, and DP-based approach, we have assumed that transition is **known**, i.e., $P(s' | s, a)$, $\forall s, a, s'$ is known, or $A, B, \mathcal{N}(0, \sigma^2 I)$ are known.

Starting from this Thursday, we start considering **unknown** transition:

We start w/ black-box access to $P$, or $f(x, u, w)$:

We can **reset the system to any** $(s, a)$, and observe $s' \sim P( \cdot | s, a)$,
Known transition versus black-box access

So far, we studied Policy Evaluation, Policy Iteration, Value Iteration, and DP-based approach, we have assumed that transition is known, i.e., $P(s' | s, a), \forall s, a, s'$ is known, or $A, B, \mathcal{N}(0, \sigma^2 I)$ are known.

Starting from this Thursday, we start considering unknown transition:

We start w/ black-box access to $P$, or $f(x, u, w)$:

We can reset the system to any $(s, a)$, and observe $s' \sim P( \cdot | s, a)$,

Or we can reset to any $(x, u)$, and observe $x' = f(x, u, w)$

($w$ being some unknown noisy disturbance)
Summary today:

1. We use DP to derive the optimal control for LQR (Ricatti equation)!

2. Never try to remember the exact form!
   Only need to understand the way we derive it (again DP!)
Next Lecture:

Control for Nonlinear system w/ black-box access to $f(x, u)$
(In general, very hard, we will study approximate algorithm
and only aim for locally optimal solutions)