Optimal Control Theory and Linear Quadratic Regulators

Sham Kakade and Wen Sun

CS 6789: Foundations of Reinforcement Learning
Today

• Recap:
  • LQR model + Ricatti equations

• Today: LQRs
  • Infinite horizon model + SDP formulations
  • (convex) SLS parameterization
Recap
Robotics and Controls

Dexterous Robotic Hand Manipulation
OpenAI, 2019
Optimal Control

- $f_t$ maps a state $x_t \in R^d$, a control (the action) $u_t \in R^k$, and a disturbance $w_t$, to the next state $x_{t+1} \in R^d$, starting from an initial state $x_0$.

\[ x_{t+1} = f_t(x_t, u_t, w_t) \]

- The objective is to find the control policy $\pi$ which minimizes the long term cost, where $H$ is the time horizon (which can be finite or infinite).

\[ \text{minimize } E_\pi \left[ \sum_{t=0}^{H-1} c_t(x_t, u_t) \right] \]

such that

\[ x_{t+1} = f_t(x_t, u_t, w_t) \]

where $H$ is the time horizon (which can be finite or infinite).

- often solved by considering the linearized control (sub-)problem where the dynamics are approximated by

\[ x_{t+1} = A_t x_t + B_t u_t + w_t, \]

with the matrices $A_t$ and $B_t$ are derivatives of the dynamics $f$ (around some trajectory) and where the costs are approximated by a quadratic function in $x_t$ and $u_t$. 

The Linear Quadratic Regulator (LQR) (finite horizon case)

• Let’s suppose this local approximation to a non-linear model is globally valid. (clearly false but this is an effective approach once when we ‘close’).

• The finite horizon LQR problem is given by

\[
\text{minimize } E \left[ x_H^T Q x_H + \sum_{t=0}^{H-1} (x_t^T Q x_t + u_t^T R u_t) \right]
\]

such that \( x_{t+1} = A_t x_t + B_t u_t + w_t, \) \( x_0 \sim D, \) \( w_t \sim N(0, \sigma^2 I), \)

where initial state \( x_0 \sim D \) is randomly distributed according \( D; \)
the disturbance \( w_t \in \mathbb{R}^d \) is multi-variate normal, with covariance \( \sigma^2 I; \)
\( A_t \in \mathbb{R}^{d \times d} \) and \( B_t \in \mathbb{R}^{d \times k} \) are referred to as system (or transition) matrices;
\( Q \in \mathbb{R}^{d \times d} \) and \( R \in \mathbb{R}^{k \times k} \) are psd matrices that parameterize the quadratic costs.

• Note that this model is a finite horizon MDP, where the \( S = R^d \) and \( A = R^k. \)
Same defs (but for costs)

• define the value function $V_\pi^h : \mathbb{R}^d \to \mathbb{R}$ as

$$V_\pi^h(x) = E\left[x_H^\top Q x_H + \sum_{t=h}^{H-1} (x_t^\top Q x_t + u_t^\top R u_t) \mid \pi, x_h = x\right],$$

• and the state-action value $Q_\pi^h : \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}$ as:

$$Q_\pi^h(x, u) = E\left[x_H^\top Q x_H + \sum_{t=h}^{H-1} (x_t^\top Q x_t + u_t^\top R u_t) \mid \pi, x_h = x, u_h = u\right],$$
Value Iteration and the Ricatti Equations

Theorem: (for the finite horizon case, with time homogenous $A_t = A, B_t = B$)

The optimal policy is a linear controller specified by:

$$\pi^*(x_t) = -K_t^*x_t$$

where $K_t^*$ is given by:

$$K_t^* = (B^TP_{t+1}B + R)^{-1}B^TP_{t+1}A$$

where $P_t$ can be computed iteratively, in a backwards manner, using the following algebraic Ricatti equations, where for $t \in [H]$,

$$P_t = A^TP_{t+1}A + Q - A^TP_{t+1}B(B^TP_{t+1}B + R)^{-1}B^TP_{t+1}A$$

$$= A^TP_{t+1}A + Q - (K_{t+1}^*)^T(B^TP_{t+1}B + R)K_{t+1}^*$$

and where $P_H = Q$.

The above equation is simply the value iteration algorithm.

Furthermore, for $t \in [H]$, we have that:

$$V_t^*(x) = x^TP_tx + \sigma^2 \sum_{h=t+1}^{H} \text{Trace}(P_h)$$
Proof: optimal control at $h = H - 1$

- Bellman equations $\Rightarrow$ there is an optimal policy which is deterministic and only a function of $x_t$ and $t$.

- Due to that $x_H = Ax + Bu + w_{H-1}$, we have:
  \[
  Q_{H-1}(x,u) = E[(Ax + Bu + w_{H-1})^\top Q(Ax + Bu + w_{H-1})] + x^\top Qx + u^\top Ru
  \]
  \[= (Ax + Bu)^\top Q(Ax + Bu) + \sigma^2 \text{Trace}(Q) + x^\top Qx + u^\top Ru \]

- This is a quadratic function of $u$. Solving for the optimal control at $x$, gives:
  \[
  \pi^{\star}_{H-1}(x) = - (B^\top QB + R)^{-1}B^\top QAx = - K^{\star}_{H-1}x,
  \]
where the last step uses that $P_H := Q$. 

\[
\]
Proof: optimal value at $h = H - 1$

- (shorthand $K_{H-1}^* = K$). using the optimal control at:
  
  $V_{H-1}^*(x) = Q_{H-1}(x, -K_{H-1}^*x) = x^T(A - BK)^TQ(A - BK)x + x^TQx + x^TK^TRKx - \sigma^2\text{Trace}(Q)$

- Continuing
  
  $V_{H-1}^*(x) - \sigma^2\text{Trace}(Q) = x^T((A - BK)^TQ(A - BK) + Q + K^TRK)x$
  
  $= x^T(AQA + Q - 2K^TB^TQA + K^T(B^TQB + R)K)x$
  
  $= x^T(AQA + Q - 2K^T(B^TQB + R)K + K^T(B^TQB + R)K)x$
  
  $= x^T(AQA + Q - K^T(B^TQB + R)K)x$
  
  $= x^TP_{H-1}x$.

where the fourth step uses our expression for $K = K_{H-1}^*$. 
Proof: wrapping up...

• This implies that:
\[
Q^*_H(x, u) = E[V^*_H(Ax + Bu + w_{H-2})] + x^\top Q x + u^\top R u
\]

\[
= (Ax + Bu)^\top P_{H-1}(Ax + Bu) + \sigma^2 \left( \text{Trace}(P_{H-1}) + \text{Trace}(Q) \right) + x^\top Q x + u^\top R u.
\]

• The remainder of the proof follows from a recursive argument, which can be verified along identical lines to the \( t = H - 1 \) case.
Today
The Linear Quadratic Regulator (LQR)  
(infinite horizon case)

- The infinite horizon LQR problem is given by

\[
\begin{align*}
\text{minimize} & \quad \lim_{H \to \infty} \frac{1}{H} E \left[ \sum_{t=0}^{H} (x_t^\top Q x_t + u_t^\top R u_t) \right] \\
\text{such that} & \quad x_{t+1} = A x_t + B u_t + w_t, \quad x_0 \sim D, \quad w_t \sim N(0, \sigma^2 I).
\end{align*}
\]

where \(A\) and \(B\) are time homogenous.

- Studied often in theory, but less relevant in practice (?)  
  (largely due to that time homogenous, globally linear models are rarely good approximations)
- Discounted case never studied.  
  (discounting doesn’t necessarily make costs finite)
- Note that we can have ‘unbounded’ average cost.
Theorem:
Suppose that the optimal average cost is finite.
Let $P$ be a solution to the following algebraic Riccati equation:
\[
P = A^T PA + Q - A^T PB(B^T PB + R)^{-1} B^T PA.
\]
(Note that $P$ is a positive definite matrix).
Theorem:
Suppose that the optimal average cost is finite.
Let $P$ be a solution to the following algebraic Riccati equation:
\[ P = A^T PA + Q - A^T PB (B^T PB + R)^{-1} B^T PA. \]
(Note that $P$ is a positive definite matrix).
We have that the optimal policy is:
\[ \pi^*(x) = - K^* x \]
where the optimal control gain is:
\[ K^* = - (B^T PB + R)^{-1} B^T PA \]
We have that $P$ is unique and that the optimal average cost is $\sigma^2 \text{Trace}(P)$. 
Semidefinite Programs to find $P$
The Primal SDP:
(for the infinite horizon LQR)

• The primal optimization problem is given as:

\[
\begin{align*}
\text{maximize} & \quad \sigma^2 \text{Trace}(P) \\
\text{subject to} & \quad \begin{bmatrix} A^T P A + Q - I & A^T P B \\ B^T P A & B^T P B + R \end{bmatrix} \succeq 0, \quad P \succeq 0
\end{align*}
\]

where the optimization variable is \( P \).
The Primal SDP:
(for the infinite horizon LQR)

• The primal optimization problem is given as:
  \[
  \begin{align*}
  \text{maximize } & \quad \sigma^2 \text{Trace}(P) \\
  \text{subject to } & \quad \begin{bmatrix} A^T P A + Q - I & A^T P B \\ B^T P A & B^T P B + R \end{bmatrix} \succeq 0, \quad P \succeq 0
  \end{align*}
  \]
  where the optimization variable is $P$.

• This SDP has a unique solution, $P^*$, which implies:
  • $P^*$ satisfies the Ricatti equations.
  • The optimal average cost of the infinite horizon LQR is $\sigma^2 \text{Trace}(P^*)$
  • The optimal policy use the gain matrix: $K^* = -(B^T P B + R)^{-1}B^T P A$
The Primal SDP:
(for the infinite horizon LQR)

• The primal optimization problem is given as:

\[
\begin{aligned}
\text{maximize} & \quad \sigma^2 \text{Trace}(P) \\
\text{subject to} & \quad \begin{bmatrix}
A^T P A + Q - I & A^T P B \\
B^T P A & B^T P B + R
\end{bmatrix} \succeq 0, \quad P \succeq 0
\end{aligned}
\]

where the optimization variable is \(P\).

• This SDP has a unique solution, \(P^*\), which implies:
  • \(P^*\) satisfies the Ricatti equations.
  • The optimal average cost of the infinite horizon LQR is \(\sigma^2 \text{Trace}(P^*)\)
  • The optimal policy use the gain matrix: \(K^* = - (B^T P B + R)^{-1} B^T P A\)

• Proof idea: Following from the Ricatti equation, we have the relaxation that for all matrices \(K\), the matrix \(P\) must satisfy:

\[
P \succeq A^T P A + Q - A^T P B (B^T P B + R)^{-1} B^T P A
\]
The Dual SDP:

- The dual optimization problem is:

  \[
  \text{minimize } \text{Trace} \left( \Sigma \cdot \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \right)
  \]

  subject to \( \Sigma_{xx} = (A \ B)\Sigma(A \ B)^\top + \sigma^2 I, \quad \Sigma \succeq 0 \)

  where the optimization variable is \( \Sigma \), a \((d + k) \times (d + k)\) matrix, with the block structure:

  \[
  \Sigma = \begin{bmatrix} 
  \Sigma_{xx} & \Sigma_{xu} \\
  \Sigma_{ux} & \Sigma_{uu} 
  \end{bmatrix}
  \]
The Dual SDP:

- The dual optimization problem is:
  
  \[
  \text{minimize} \quad \text{Trace} \left( \Sigma \cdot \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \right) \\
  \text{subject to} \quad \Sigma_{xx} = (A \ B) \Sigma (A \ B)^\top + \sigma^2 I, \quad \Sigma \succeq 0
  \]

  where the optimization variable is \( \Sigma \), a \((d + k) \times (d + k)\) matrix, with the block structure:

  \[
  \Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xu} \\ \Sigma_{ux} & \Sigma_{uu} \end{bmatrix}
  \]

- The interpretation of \( \Sigma \) is that it is the covariance matrix of the stationary distribution. This analogous to state-action visitation distributions (the dual variables in the MDP LP).
The Dual SDP:

- The dual optimization problem is:

$$\text{minimize} \quad \text{Trace} \left( \Sigma \cdot \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix} \right)$$

subject to

$$\Sigma_{xx} = (A \ B) \Sigma (A \ B)^\top + \sigma^2 I, \quad \Sigma \succeq 0$$

where the optimization variable is $\Sigma$, a $(d + k) \times (d + k)$ matrix, with the block structure:

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xu} \\ \Sigma_{ux} & \Sigma_{uu} \end{bmatrix}$$

- The interpretation of $\Sigma$ is that it is the covariance matrix of the stationary distribution. This analogous to state-action visitation distributions (the dual variables in the MDP LP).

- This SDP has a unique solution, say $\Sigma^\star$. The optimal gain matrix is then given by:

$$K^\star = - \Sigma^\star_{ux} (\Sigma^\star_{xx})^{-1}$$