Optimal Control, Lecture 5: Reinforcement Learning (RL)

Anders Hansson

Division of Automatic Control
Linkoping University

Contents

Optimal Control Problem

$$\begin{array}{ll} \text{minimize} & \phi(x_N) + \sum_{k=0}^{N-1} f_k(x_k, u_k) \\ \text{subject to} & x_{k+1} = F_k(x_k, u_k), \quad k \in \mathbf{Z}_{N-1} \end{array} \tag{1}$$

for a given initial value x_0 with variables $(u_0, x_1, \ldots, u_{N-1}, x_N)$.

Terminology

Optimal control	**Reinforcement learning**
System	Environment
Controller	Agent
Control	Action
Incremental cost	Stage reward
Cost function	Reward function

Dynamic Programming

Suppose there exist finite solution to the backward Dynamic Programming recursion

$$V_N(x) = \phi(x)$$

$$V_k(x) = \min_u \{f_k(x,u) + V_{k+1}(F_k(x,u))\} \quad (2)$$

$k = N-1, N-2, \ldots, 0$ Then there exists an optimal solution to (1) and

- (a) $J_k^*(x) = V_k(x)$ for all $k = 0, 1, \ldots, N,\ x \in X_n$
- (b) The optimal feedback control in each stage is the minimizing argument in (2)

The Q-function

Let

$$Q_k(x,u) = f_k(x,u) + V_{k+1}(F_k(x,u)), \quad k = 0, 1, \ldots, N-1$$

Then the dynamic programming recursion is

$$V_k(x) = \min_u Q_k(x,u)$$

where $V_N(x) = \phi(x)$.

Approximation of the V-function

Approximate V_k with regression

$$\tilde{V}_k(x, a_k) = a_k^T \varphi_k(x)$$

or an ANN and then approximate Q_k with

$$\tilde{Q}_k(x, u, a) = \begin{cases} f_k(x, u) + \tilde{V}_{k+1}(F_k(x, u), a), & k \in \mathbf{Z}_{N-2} \\ f_k(x, u) + \phi(F_k(x, u)), & k = N - 1. \end{cases}$$

Remark: No dependence on a for $k = N - 1$.

Fitted Value Iteration for the V-function

For $k = N-1, N-2, \dots, 0$:

1. Consider samples x_k^s, where $1 \leq s \leq r$ and let
$$\beta_k^s = \min_u \tilde{Q}_k(x_k^s, u, a_{k+1}), \tag{3}$$
where a_{k+1} is known from previous iterate.
2. Solv LS problem for next a_k:
$$\text{minimize } \frac{1}{2} \sum_{s=1}^{r} \left(\tilde{V}_k(x_k^s, a) - \beta_k^s \right)^2$$

When $\hat{V}_k$ linear regression model, the LS problem is a linear LS problem with closed form solution.

The approximate feedback function is given by
$$\mu_k(x) = \operatorname*{argmin}_u \tilde{Q}_k\left(x, u, a_{k+1}\right). \tag{4}$$

The choice of samples and regression heavily affects the obtained quality of approximation .

LQ Control

$$\begin{array}{ll} \text{minimize} & x_N^T S x_N + \sum_{k=0}^{N-1} x_k^T S x_k + u_k^T R u_k \\ \text{subject to} & x_{k+1} = A x_k + B u_k, \quad k \in \mathbf{Z}_{N-1} \end{array}$$

for given x_0, where $x_k \in \mathbf{R}^2$ and $u_k \in \mathbf{R}$.

Consider $\varphi(x) = (x_1^2, x_2^2, 2x_1x_2)$, let

$$\tilde{P} = \begin{bmatrix} a_1 & a_3 \\ a_3 & a_2 \end{bmatrix}$$

and let

$$\tilde{V}_k(x, a) = a^T \varphi(x) = x^T \tilde{P} x$$

Hence true value function $V_k(x) = x^T P_k x$ and approximate value function $\tilde{V}_k(x, a)$ agree if $\tilde{P} = P_k$.

LQ Control ctd.

Approximate Q-function:

$$\begin{aligned}\tilde{Q}_k(x,u,a) &= x^T S x + u^T R u + (Ax+Bu)^T \tilde{P}(Ax+Bu) \\ &= \begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} S + A^T \tilde{P} A & A^T \tilde{P} B \\ B^T \tilde{P} A & R + B^T \tilde{P} B \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}.\end{aligned} \tag{5}$$

For $k = N-1$ down to $k = 0$ we solve the linear LS problem in (3) to obtain

$$\begin{aligned}\beta_k^s = (x_k^s)^T \Big\{ S + A^T \tilde{P}_{k+1} A - A^T \tilde{P}_{k+1} B \left(R + B^T \tilde{P}_{k+1} B \right)^{-1} \\ \times B^T \tilde{P}_{k+1} A \Big\} x_k^s\end{aligned}$$

assuming $R + B^T \tilde{P}_{k+1} B$ positive definite. Here $\tilde{P}_N = S$.

LQ Control ctd.

We then obtain a_k as solution to the linear LS problem

$$\text{minimize } \frac{1}{2}\sum_{s=1}^{r}\left(\varphi^T(x_k^s)a - \beta_k^s\right)^2$$

with variable a. This defines $\tilde{P}_k$. The solution a_k satisfies the normal equations

$$\Phi_k^T\Phi_k a_k = \Phi_k^T\beta_k,$$

where

$$\Phi_k = \begin{bmatrix}\varphi^T(x_k^1)\\ \vdots \\ \varphi^T(x_k^r)\end{bmatrix}; \qquad \beta_k = \begin{bmatrix}\beta_k^1\\ \vdots \\ \beta_k^r\end{bmatrix}.$$

It is here crucial to choose x_k^s such that $\Phi_k^T\Phi_k$ is invertible. We realize that we need $r \geq 3$ for this hold.

From (4) and (5) we obtain

$$u_k = -\left(R_k + B_k^T\tilde{P}_{k+1}B_k\right)^{-1}B_k^T\tilde{P}_{k+1}A_kx_k$$

Finite horizon value iteration for the Q-function

Remember

$$V_k(x) = \min_u Q_k(x, u)$$

and hence

$$V_{k+1}(F_k(x, \bar{u})) = \min_u Q_{k+1}(F_k(x, \bar{u}), u)$$

Add $f_k(x, \bar{u})$ to both sides to obtain

$$Q_k(x, \bar{u}) = f_k(x, \bar{u}) + \min_u Q_{k+1}(F_k(x, \bar{u}), u), \quad k = N-2, N-3, \dots, 0$$

where $Q_{N-1}(x, u) = f_{N-1}(x, u) + \phi(F_{N-1}(x, u))$.

Observations

- Iteration for Q-function equivalent to dynamic programming recursion.
- Do not need to know F_k.
- Sufficient to be able to evaluate $F_k(x, \bar{u})$ using xperiments or digital twin.
- Q-function more complicated than value function V since V only function of x but Q also function of u.

Fitted value iteration for the Q-function

Approximate Q_k as

$$\tilde{Q}_k(x, u, a_k) = a_k^T \varphi(x, u)$$

for $k \in \mathbf{Z}_{N-1}$ or with an ANN.

Consider samples (x_k^s, u_k^s) and define

$$\beta_{N-1}^s = \phi(F_{N-1}(x_{N-1}^s, u_{N-1}^s))$$

and

$$\beta_k^s = \min_u \tilde{Q}_{k+1}(F_k(x_k^s, u_k^s), u, a_{k+1}), \tag{6}$$

where a_{k+1} is a known value from previous iterate.

- We do not need an analytical expression for F_k in order to define β_k^s.
- Depending on how the feature vectors are chosen the minimization above could become very tractable.

Fitted value iteration for the Q-function ctd.

Define a_k as solution to

$$\text{minimize} \quad \frac{1}{2}\sum_{s=1}^{r}\left(\tilde{Q}_k(x_k^s, u_k^s, a) - f_k(x_k^s, u_k^s) - \beta_k^s\right)^2 \tag{7}$$

with variable a for $k \in \mathbf{Z}_{N-1}$.

The iterations start with $k = N-1$ and goes down to $k = 0$, where we alternate between solving (7) and (6).

The approximate optimal control is

$$u_k^\star = \mu_k(x) = \underset{u}{\operatorname{argmin}}\, \tilde{Q}_k\left(x, u, a_k\right). \tag{8}$$

Remark: We notice that using the Q-function instead of using the value function comes at the price of also having to sample the control signal space.

LQ Control

$$\begin{array}{ll}\text{minimize} & x_N^T S x_N + \sum_{k=0}^{N-1} x_k^T S x_k + u_k^T R u_k \\ \text{subject to} & x_{k+1} = A x_k + B u_k, \quad k \in \mathbf{Z}_{N-1}\end{array}$$

for given x_0, where $x_k \in \mathbf{R}^2$ and $u_k \in \mathbf{R}$.

Let $\varphi(x,u) = (x_1^2, x_2^2, u^2, 2x_1x_2, 2x_1u, 2x_2u)$ and

$$\tilde{Q}_k(x,u,a) = a^T \varphi(x,u),$$

With

$$\begin{bmatrix} \tilde{P} & \tilde{r} \\ \tilde{r}^T & \tilde{q} \end{bmatrix} = \begin{bmatrix} a_1 & a_4 & a_5 \\ a_4 & a_2 & a_6 \\ a_5 & a_6 & a_3 \end{bmatrix}$$

we may write

$$\tilde{Q}_k(x,u,a) = \begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} \tilde{P} & \tilde{r} \\ \tilde{r}^T & \tilde{q} \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}. \qquad (9)$$

LQ Control ctd.

For $k = N - 1$ we define

$$\beta_{N-1}^s = \left(x_+^s\right)^T S x_+^s$$

where $x_+^s = Ax_{N-1}^s + Bu_{N-1}^s$. For $k = N - 2$ down to $k = 0$ we solve the linear LS problem in (6) to obtain

$$\beta_k^s = \left(x_+^s\right)^T \left(\tilde{P}_{k+1} - \tilde{r}_{k+1}\tilde{q}_{k+1}^{-1}\tilde{r}_{k+1}^T\right) x_+^s,$$

where $x_+^s = Ax_k^s + Bu_k^s$.

LQ Control ctd.

We then obtain a_k for $k \in \mathbf{Z}_{N-1}$ as the solution to:

$$\text{minimize} \quad \frac{1}{2}\sum_{s=1}^{r}\left(\varphi^T(x_k^s, u_k^s)a - (x_k^s)^T S x_k^s - (u_k^s)^T R u_k^s - \beta_k^s\right)^2.$$

The solution a_k satisfies the normal equations

$$\Phi_k^T \Phi_k a_k = \Phi_k^T \gamma_k,$$

where

$$\Phi_k = \begin{bmatrix} \varphi^T(x_k^1, u_k^1) \\ \vdots \\ \varphi^T(x_k^r, u_k^r) \end{bmatrix}, \qquad \gamma_k = \begin{bmatrix} (x_k^1)^T S x_k^1 + (u_k^1)^T R u_k^1 + \beta_k^1 \\ \vdots \\ (x_k^r)^T S x_k^r + (u_k^r)^T R u_k^r + \beta_k^r \end{bmatrix}.$$

Crucial to choose (x_k^s, u_k^s) such that $\Phi_k^T \Phi_k$ is invertible. We realize that we need $r \geq 6$ for this to hold.

Optimal feedback function is by (8) and (9) given by

$$\mu_k(x) = -\tilde{q}_k^{-1} \tilde{r}_k^T x,$$