Optimal Control, Lecture 3: Value Iteration and Policy Iteration

Anders Hansson

Division of Automatic Control
Linkoping University

Contents

- Bellman Equation
- Value Iteration (VI)
- Policy Iteration (PI)
- Approximate PI
- Linear Quadratic Control

Infinite Time Horizon Oprimization

$$
\begin{array}{ll}
J^*(x_0) = \text{min} & \sum_{k=0}^{\infty} \gamma^k f(x_k, u_k) \\
\text{subject to} & x_{k+1} = F(x_k, u_k) \\
& x_0 \text{ given} \\
& u_k \in U(x_k)
\end{array}
\tag{1}
$$

with $0 < \gamma \leq 1$ discount factor.

Bellman Equation

Assume $0 \in U(0)$, $f(0,0) = 0$, $F(0,0) = 0$ and that f is strictly positive definite. If there exists a strictly positive definite V such that the Bellman equation

$$V(x) = \min_{u \in U(x)} \{f(x,u) + \gamma V(F(x,u))\}$$

holds, then

- (a) $J^*(x) = V(x)$
- (b) The minimizing argument in the Bellman equation is an optimal feedback for (3) that results in a globally convergent closed loop system if γ is sufficiently close to one.

Value Iterations (VI)

Change iteration index in the dynamic programming recursion so that we iterate forward instead:

$$V_{k+1}(x) = \min_{u \in U(x)} \{ f_k(x,u) + \gamma V_k \left(F_k(x,u) \right) \} \tag{2}$$

with initial value $V_0(x) = 0$.

If one has a clever guess of an approximate solution to the Bellman equation, this can be used as initial value instead.

VI for LQ Control

$$
\begin{array}{ll}
\text{min} & \sum_{k=0}^{\infty} \gamma^k \left(x_k^T S x_k + u_k^T R u_k \right) \\
\text{subject to} & x_{k+1} = A x_k + B u_k \\
 & x_0 \text{ given}
\end{array}
\tag{3}
$$

Let $V_k(x) = x^T P_k x$ with $P_0 = 0$. Similarly as in LQ Example in previous lecture

$$
P_{k+1} = S + \gamma A^T P_k A - \gamma^2 A^T P_k B \left(R + B^T P_k B \right)^{-1} B^T P_k A
$$

Proof of Convergence

Based on Bellman operator:

$$T(V)(x) = \min_{u \in U(x)} \left\{ f(x,u) + \gamma V(F(x,u)) \right\}. \tag{4}$$

Details on white board.

Policy Iterations (PI)

Bellman policy operator:

$$T_\mu(V)(x) = f(x, \mu(x)) + \gamma V\left(F(x, \mu\left(x\right))\right) \tag{5}$$

for a given function μ.

Iterate starting with initial μ_0:

1. Solve (policy evaluation step)

$$V_k(x) = T_{\mu_k}(V_k)(x), \tag{6}$$

2. Solve (policy improvement step)

$$\mu_{k+1}(x) = \underset{u \in U(x)}{\operatorname{argmin}} \left\{ f(x, u) + \gamma V_k\left(F(x, u)\right)\right\}. \tag{7}$$

Proof of convergence on white board.

LQ Control

Guess that $V_k(x) = x^T P_k x$ and that $\mu_k(x) = -L_k x$.

Policy evaluation step:

$$x^T P_k x = x^T S x + x^T L_k^T R L_k x + \gamma x^T \left(A - BL_k\right)^T P_k \left(A - BL_k\right) x$$

for given L_k. Solution from Lyapunov equation

$$P_k - \gamma \left(A - BL_k\right)^T P_k \left(A - BL_k\right) = S + L_k^T R L_k,$$

Policy improvement step:

$$\mu_{k+1}(x) = \underset{u}{\operatorname{argmin}} \left\{ x^T S x + u^T R u + \gamma \left(Ax - Bu\right)^T P_k \left(Ax - Bu\right) \right\}.$$

with solution $\mu_{k+1}(x) = -L_{k+1} x$, where

$$L_{k+1} = \gamma \left(R + \gamma B^T P_k B\right)^{-1} B^T P_k A.$$

Approximate Evaluation of V_k

It holds that (6) implies

$$\begin{aligned} V_k(x_0) &= f\left(x_0, \mu_k(x_0)\right) + \gamma V_k\left(F\left(x_0, \mu_k(x_0)\right)\right) \\ &= f\left(x_0, \mu_k(x_0)\right) + \gamma V_k\left(x_1\right) \\ &= f\left(x_0, \mu_k(x_0)\right) + \gamma f\left(x_1, \mu_k(x_1)\right) + \gamma^2 V_k\left(x_2\right) \\ &\vdots \\ &= \sum_{i=0}^{N-1} \gamma^i f\left(x_i, \mu_k(x_i)\right) + \gamma^N V_k\left(x_N\right), \end{aligned} \tag{8}$$

where $x_{i+1} = F(x_i, \mu_k(x_i))$.

In case N is large and μ_k is stabilizing we have that x_N is close to zero and that also $V_k(x_N)$ is close to zero.

Approximate evaluation of $V_k(x_0)$ obtained by simulating the dynamical system and add up the incremental costs.

Approximation of V

Let

$$\beta_k^s = \sum_{i=0}^{N-1} \gamma^i f\left(x_i, \mu_k(x_i)\right),$$

where $x_{i+1} = F(x_i, \mu_k(x_i))$, $x_0 = x^s$, $1 \leq s \leq r$.

Let $\tilde{V}(x, a)$ be a linear regression or an Aritifical Neural Network (ANN) with parameter a that should approximate V_k in (6).

Find the approximation of V_k by solving

$$\text{minimize } \frac{1}{2} \sum_{s=1}^{r} \left(\tilde{V}(x^s, a) - \beta_k^s\right)^2$$

with variable a. The solution is denoted a_k. Then perform exact policy improvement step

$$\mu_{k+1}(x) = \underset{u \in U(x)}{\operatorname{argmin}} \left\{ f(x, u) + \gamma \tilde{V}\left(F(x, u), a_k\right) \right\}. \qquad (9)$$

Approximate LQ Control

Assume one input and two states and let $\varphi(x) = (x_1^2, x_2^2, 2x_1x_2)$.[1] Let

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

With

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

we have

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

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

[1] Notice that the indices refer to components of the vector and not to time.

Approximate LQ Control

With an abuse of notation $a_k \in \mathbf{R}^3$, which defines $\tilde{P}_k$, is obtained as solution to Least Squares (LS) problem

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

with variable a. The solution a_k satisfies normal equations

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

where

$$\Phi_k = \begin{bmatrix} \varphi^T(x^1) \\ \vdots \\ \varphi^T(x^r) \end{bmatrix}, \qquad \beta_k = \begin{bmatrix} \beta_k^1 \\ \vdots \\ \beta_k^r \end{bmatrix},$$

and where

$$\beta_k^s = \sum_{i=0}^{N-1} \gamma^i \left(x_i^T S x_i + \mu_k(x_i)^T R \mu_k(x_i) \right),$$

and where $x_{i+1} = A x_i + B \mu_k(x_i)$ with initial values x^s.

Approximate LQ Control

It is crucial to choose x^s such that $\Phi_k^T \Phi_k$ is invertible, which holds if $r \geq 3$.

We define

$$\begin{aligned}\tilde{Q}_k(x,u,a) &= f(x,u) + \gamma \tilde{V}(Ax+Bu,a) \\ &= x^T S x + u^T R u + \gamma (Ax+Bu)^T \tilde{P}(Ax+Bu) \\ &= \begin{bmatrix} x \\ u \end{bmatrix}^T \begin{bmatrix} S + \gamma A^T \tilde{P} A & \gamma A^T \tilde{P} B \\ \gamma B^T \tilde{P} A & R + \gamma B^T \tilde{P} B \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}.\end{aligned} \tag{10}$$

She solution to (9) is given by

$$\mu_{k+1}(x) = \operatorname*{argmin}_u \tilde{Q}_k(x,u,a_k) = -\gamma (R + \gamma B^T \tilde{P}_k B)^{-1} B^T \tilde{P}_k A x$$

assuming $R + \gamma B^T \tilde{P}_k B$ positive definite. Hence

$$\mu_{k+1}(x) = -L_{k+1} x,$$

where $L_{k+1} = \gamma (R + \gamma B^T \tilde{P}_k B)^{-1} B^T \tilde{P}_k A$.