Practice Course 2
Theory and Computation Methods of Approximate DP I

Mengdi Wang

July 3rd, 2012

Laboratory for Information and Decision Systems, M.I.T.

Theory of Approximate DP

Infinite-Horizon DP Problem

Minimize over policies

$$\pi = \{\mu_0, \mu_1, \ldots\}$$

the objective cost function

$$J_\pi(x_0) = \lim_{N\to\infty} \mathbf{E}_{w_k, k=0,1,\ldots} \left\{ \sum_{k=0}^{N-1} \alpha^k g\left(x_k, \mu_k(x_k), w_k\right) \right\}$$

How to Approximate DP

- Approximation: parameterize policies/cost vectors, aggregation, etc.
- Simulation: Use simulation-generated trajectories $\{x_k\}$ to calculate DP quantities, without knowing the system

Markovian Decision Process

Assume the system is an n-state (controlled) Markov chain

Change to Markov chain notation

- States $i = 1, \ldots, n$ (instead of x)
- Transition probabilities $p_{i_k i_{k+1}}(u_k)$ [instead of $x_{k+1} = f(x_k, u_k, w_k)$]
- Cost per stage $g(i, u, j)$ [instead of $g(x_k, u_k, w_k)$]
- Cost of a policy $\pi = \{\mu_0, \mu_1, \ldots\}$

$$J_\pi(i) = \lim_{N\to\infty} \mathbf{E}_{\substack{w_k \\ k=0,1,\ldots}} \left\{ \sum_{k=0}^{N-1} \alpha^k g\left(i_k, \mu_k(i_k), i_{k+1}\right) \mid i_0 = i \right\}$$

MDP Continued

The optimal cost vector satisfies the Bellman equation for all i

$$J^*(i) = \min_{u \in U} \sum_{j=1}^{n} p_{ij}(u)(g(i,u,j) + \alpha J^*(j)),$$

or in matrix form

$$J^* = \min_{\mu:\{1,\ldots,n\} \mapsto U} \{g_\mu + \alpha P_\mu J^*\}.$$

Shorthand notation for DP mappings

$$(TJ)(i) = \min_{u \in U(i)} \sum_{j=1}^{n} p_{ij}(u)\big(g(i,u,j) + \alpha J(j)\big), \quad i = 1, \ldots, n,$$

$$(T_\mu J)(i) = \sum_{j=1}^{n} p_{ij}\big(q(i)\big)\big(g\left(i, \mu(i), j\right) + \alpha J(j)\big), \quad i = 1, \ldots, n$$

Approximation Architecture

Approximation in Policy Space

Parameterize the set of policies μ using a vector r, and then optimize over r.

Approximation in Value Space

Approximate J^* and J_μ from a family of functions parameterized by r, e.g., a linear approximation

$$J \approx \Phi r, \qquad J(i) \approx \phi(i)' r.$$

Approximate DP Algorithms: A Roadmap

Approximate PI (*)

Implement the two steps of PI in an approximate sense:

- Policy Evaluation $J_{\mu_t} = T_{\mu_t} J_{\mu_t}$ by approximation/simulation
  - Direct Approach (*), e.g., simulation-based least squares
  - Indirect Approach, solve $J_{\mu_t} = \Pi T_{\mu_t} J_{\mu_t}$ by TD/LSTD/LSPE.
- Policy Improvement $T_{\mu_{t+1}} J_{\mu_t} = T J_{\mu_t}$ using the approximate cost vector/Q-factors.

Approximate J^* and Q^*

Solve $J^* = TJ^*$ or $Q^* = FQ^*$ directly by simulation, e.g.,

- Q- Learning, Bellman Error Minimization, LP approach

Call Options

A call option gives the buyer of the option the right to buy the underlying asset at a fixed price (strike price or K). The buyer pays a price for this right.

- At or before expiration,
  - If the value of the underlying asset (S)$>$ Strike Price(K) Buyer makes the difference: S - K
  - If the value of the underlying asset (S) $<$ Strike Price (K) Buyer does not exercise

Valuing American Call Options

Variables

- Strike Price: K
- Time till Expiration: T
- Price of underlying asset: S
- Volatility, Dividends, etc.

Valuing American options requires the solution of an optimal stopping problem:

$$\text{Option Price} = \mathbf{E}\left[S(t^*) - K \mid \text{Option eventually exercised}\right]$$

where

$$t^* = \text{optimal exercising time.}$$

If the option writers do not solve t^* correctly, the option buyers will have an arbitrage opportunity to exploit the option writers.

Infinite-Horizon DP Formulation

Assume that:

- Dynamics of underlying asset $S_{t+1} = f(S_t, w_t)$
- State: S_t, price of the underlying asset
- Control: $u_t \in \{\text{Exercise}, \text{Hold}\}$
- Transition cost: $g_t(\text{HOLD}) = 0$, $g_t(\text{Exercise}) = S_t - K$.
- The option never expires.
- There exists a discount factor $\alpha \in (0, 1)$

Bellman Equation

Let $J_t(S)$ be the option price at the tth day when the current stock price is S

$$J(S_t) = \max\{S_t - K, \alpha \mathbf{E}\left[J(S_{t+1})\right]\}.$$

Binomial Model

For simplicity, consider a model with a finite number of states:

$$S_{t+1} = \begin{cases} \min\{U, uS_t\} & \text{with probability} \quad \text{p} \\ \max\{D, dS_t\} & \text{with probability} \quad \text{1-p} \end{cases}$$

The Bellman equation is $J = TJ$ where

$$TJ(S) = \max\Big\{S - K, \\ \alpha\left[pJ(\min\{U, uS_t\}) + (1-p)J(\max\{D, dS_t\})\right]\Big\}.$$

Features

We will approximate the option prices J^*, J_μ using two set of features, each consisting of 3 features/basis functions.

Simple Polynomial

$$L_0(S) = 1, \quad L_1(S) = S, \quad , L_2(S) = S^2.$$

Laguerre Polynomial

$$L_0(S) = \exp(-S), \quad L_1(S) = \exp(-S)(1 - S),$$
$$L_2(S) = \exp(-S)(1 - S + S^2/2).$$

The basis matrix Φ is an $n \times 3$ matrix.

Approximate Policy Evaluation

Exercise 1.A (Direct Approach)

Use the direct least squares approach

$$\min_r \frac{1}{2} \sum_{k=0}^{N-1} \left(\phi(i_k)' r - \sum_{t=k}^{N-1} \alpha^{t-k} g\left(i_t, \mu(i_t), i_{t+1}\right) \right)^2$$

to evaluate the profits of a specified exercising strategy.

- Construct a simulator that generates trajectories of $\{i_k\}$.
- Plot the approximate cost vector as a function of the stock price.

Exercise 1 Continued

Formula of the Solution $J_\mu \approx \Phi r_\mu$

$$
\begin{aligned}
r_\mu &= \left(\sum_{k=0}^{N-1} \phi(i_k)\phi(i_k)'\right)^{-1} \left(\sum_{k=0}^{N-1} \phi(i_k) \sum_{t=k}^{N-1} \alpha^{t-k} g\left(i_t, \mu(i_t), i_{t+1}\right)\right) \\
&= A^{-1} b
\end{aligned}
$$

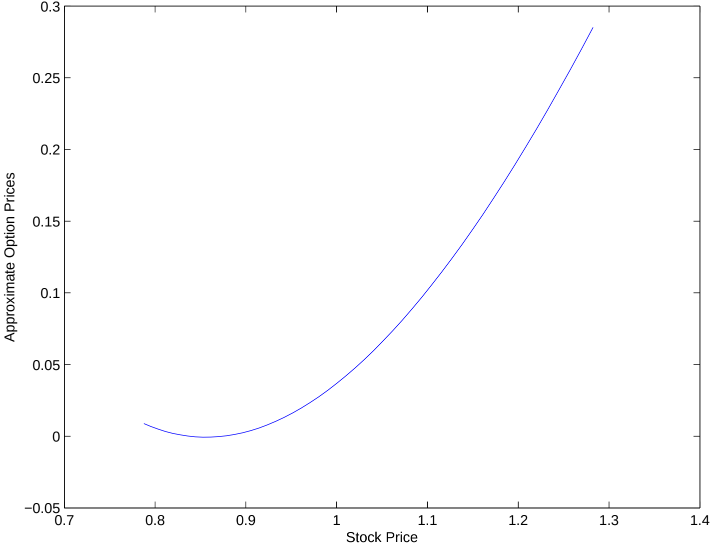
where

$$
A = \sum_{k=0}^{N-1} \phi(i_k)\phi(i_k)',
$$

$$
b = \sum_{k=0}^{N-1} \alpha^{t_\mu(k)-k} (S(t_\mu(k)) - K)\phi(i_k)
$$

where $t_\mu(k)$ is the first time of triggering the 'Exercise' control using policy μ after time k.

Computation Results - Option Prices



Approximate Policy Evaluation

Exercise 1.B (Optional)

Suppose that holding the option always incurs a cost $g(i,j) = -10^{-4}$. Modify the program of Exercise 1.A to price the American call option.

Exercise 1.C (Optional)

Use an indirect approach to price an American call option.

- Choose any one of the three algorithms:

$$TD/LSTD/LSPE$$

Use Approximate PI to Evaluate Options

Exercise 2

Use approximate PI to price an American call option.
The program should be a function of S_0, T, p, u, d, K.

Suggestions:

- Start with a randomly generated policy $\mu_0 : \{1, \ldots, n\} \mapsto \{HOLD, EXERCISE\}$.
- Use approximate policy evaluation (Exercise 1) to evaluate J_{μ_t} and Q_{μ_t} for a given policy μ_t.
- Plot the trajectories of μ_t.

Policy Iteration for Option Pricing

Algorithm (starts with any μ_0)

- Policy evaluation:
  - Evaluate $J_{\mu_t} \approx \Phi r_\mu$ by approximate policy evaluation: use the program of Exercise 1 to compute r_μ
  - Evaluate the Q-values. For example, for $i_t \in [2, n-1]$,

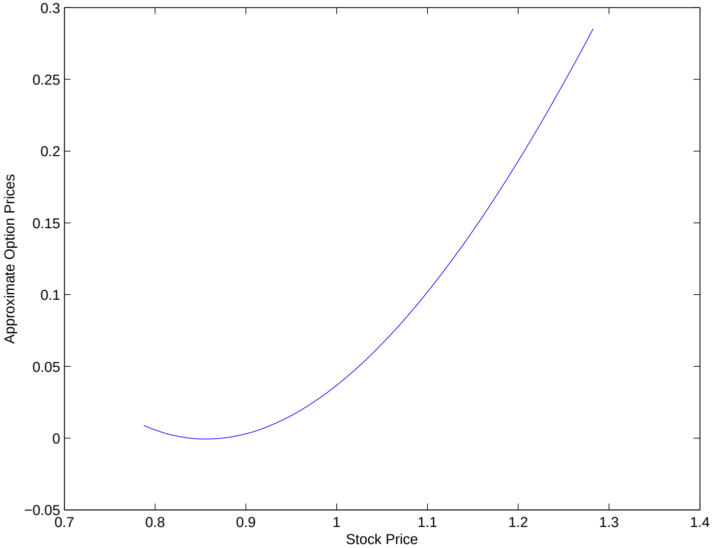
$$\begin{aligned} Q_{\mu_t}(i_t) &= \alpha \mathbf{E}\left[J_{\mu_t}(i_{t+1})\right] \approx \alpha \mathbf{E}\left[\tilde{J}_{\mu_t}(i_{t+1})\right] \\ &= \alpha\left(p\tilde{J}_{\mu_t}(i_t+1) + (1-p)\tilde{J}_{\mu_t}(i_t-1)\right). \end{aligned}$$

  Note $\tilde{J}_\mu(i) = \phi(i)' r_\mu$.

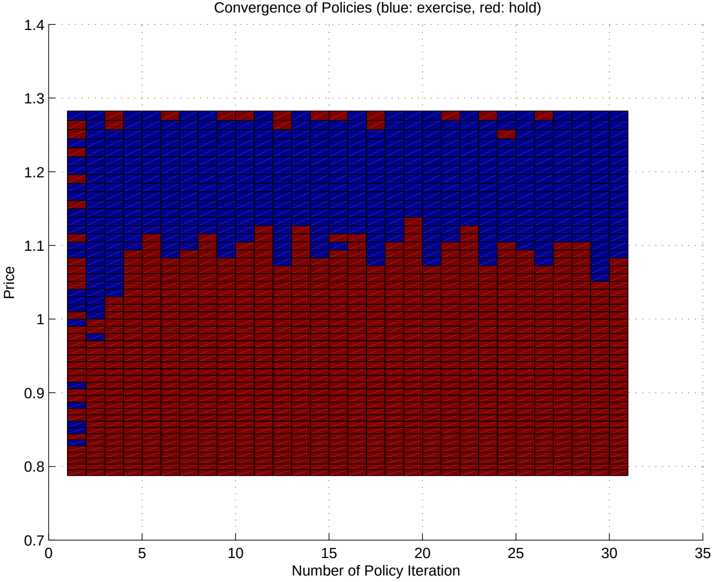
- Policy improvement:

$$\mu_{t+1}(i) = \begin{cases} \text{HOLD} & \text{if } S(i) - K \leq Q_{\mu_t}(i), \\ \text{EXERCISE} & \text{Otherwise.} \end{cases}$$

Computation Results - Option Prices



Convergence of Exercising Policies



Online Approximate PI for Q Factors

Exercise 3

Modify the program of Exercise 2, so that the policy improvement step uses approximate evaluation of Q-factors (instead of exact Q values calculated using known p).

- For each state i, calculate

$$Q(i) = \mathbf{E}\left[\alpha \tilde{J}(i_{k+1}) \mid i_k = i\right]$$

by averaging the samples obtained from the trajectory

$$Q(i) \approx \frac{\sum_{k=0}^{k=N} \mathbf{1}(i_k = i)\alpha \tilde{J}(i_{k+1})}{\sum_{k=0}^{k=N} \mathbf{1}(i_k = i)}$$

- Note $J(i_{k+1}) = \phi(i_{k+1})' r$.

The end

Thank You Very Much!

Any Question is Welcome :-)