Practice Course 1
Theory and Computation Methods of Exact DP

Mengdi Wang

June 27th, 2012

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

Theory of Exact DP: A Review

Descrete-time System

$$x_{k+1} = f_k(x_k, u_k, w_k), \qquad k = 0, 1, \ldots, N-1$$

- x_k: State; summarizes past information that is relevant for future optimization
- u_k: Control; decision to be selected at time k from a given set U_k
- w_k: Random disturbance or noise

Objective - control the system to minimize overall cost

$$E\left\{ g_N(x_N) + \sum_{k=0}^{N-1} \alpha^k g_k(x_k, u_k, w_k) \right\},$$

where $N < \infty$ or $N = \infty$, and $\alpha \in (0, 1]$.

Finite-Horizon DP Problem $(N < \infty, \alpha = 1)$

The Problem

- Policies: collection of functions about what to do at each time and state

$$\pi = \{\mu_0, \dots, \mu_{N-1}\}, \qquad \mu_k : X_k \mapsto U_k.$$

- Cost function: the expected cost of specified policy if the system starts with given state

$$J_\pi(x_0) = E\left\{ g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k) \right\}$$

- Objective: find $\pi^* = \mathrm{argmin}_\pi J_\pi(x_0)$ for all x_0.

Optimality Condition = DP Algorithm

Principle of optimality

the tail part of an optimal policy is optimal for the tail subproblem

DP algorithm

Start with $J_N(x_N) = g_N(x_N)$, and go backwards for $k = N-1, \ldots, 0$ using

$$J_k(x_k) = \min_{u_k \in U_k} \mathbf{E}_{w_k} \left\{ g_k(x_k, u_k, w_k) + J_{k+1}\left(f_k(x_k, u_k, w_k)\right) \right\}.$$

Proof by induction that the principle of optimality is always satisfied.

Infinite-Horizon DP Problem $(N = \infty, \alpha \le 1)$

The Problem

Minimize

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

- Want to find $\pi = \{\mu_0, \mu_1, \ldots\}$ that is optimal for the entire future.
- Need to make sure $J_{\pi^*}(x_0) < \infty$:
    - Discounted problem: $\alpha < 1$.
    - Stochastic Shortest Path problem (SSP): termination in finite time.

Example: Discounted Markovian Decision Problem

States: $\{1, \ldots, n\}$.
Given state i_t, control u, the next state evolves according to transition probabilities

$$\mathbf{P}(i_{t+1} = j \mid i_t = i, u) = p_{ij}(u), \qquad t = 1, 2, \ldots.$$

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

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

The T Mappings

The operator T : space of $J \to$ space of J:

$$(TJ)(x) = \min_{u \in U(x)} \mathbf{E}_w \left\{ g(x,u,w) + \alpha J\left(f(x,u,w)\right) \right\}, \ \forall\, x$$

Key Properties for Developing the Theory

- Monotonicity

$$J_1 \leq J_2, \text{ then } TJ_1 \leq TJ_2$$

- Contraction

$$\|TJ_1 - TJ_2\| \leq \alpha \|J_1 - J_2\|, \qquad \|T_\mu J_1 - T_\mu J_2\| \leq \alpha \|J_1 - J_2\|$$

Notations

- TJ: The one-step-backward cost by improving the cost J
- J_μ: The total cost if using μ throughout

Theory and Algorithm of Infinite-Horizon DP

The DP problem is essentially the Bellman equation

$$TJ^* = J^*$$

Map of Key Results

- Optimal cost vector implies optimal policy, and vice versa:

$$J^* = TJ^* = T_{\mu^*}J^* = J_{\mu^*}.$$

- Value Iteration $J_{k+1} = TJ_k$ is convergent ($J_k \to J^*$).
- Polity Iteration $T_{\mu_{k+1}}J_{\mu_k} = TJ_{\mu_k}$ is convergent ($J_{\mu_k} \to J^*$, $\mu_k \to \mu^*$).
- Error bounds, Asynchronous implementation, etc.

Q-Learning

Optimal Q-factors - the function $Q(i,u)$ that satisfies the following Bellman equation

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

or in short

$$Q^* = FQ^*$$

Q-Learning: apply DP to Q-values instead of J values

- Value Iteration: $Q_{k+1} = FQ_k$
- Policy Iteration: $F_{\mu_{k+1}} Q_{\mu_k} = FQ_{\mu_k}$
- VI and PI are convergent
- Model-free.

Options

An option provides the holder with the right to buy or sell a specified quantity of an underlying asset at a fixed price (called a strike price or an exercise price) at or before the expiration date of the option.

Since it is a right and not an obligation, the holder can choose not to exercise the right and allow the option to expire.
There are two types of options - call options (right to buy) and put options (right to sell).

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 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
- More generally,
  - the value of a call increases as the value of the underlying asset increases
  - the value of a call decreases as the value of the underlying asset decreases

Put Options

A put option gives the buyer of the option the right to sell the underlying asset at a fixed price. The buyer pays a price for this right.

- At 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
- More generally,
  - the value of a call increases as the value of the underlying asset decreases
  - the value of a call decreases as the value of the underlying asset increases

European Options vs. American Options

An American option can be exercised at any time prior to its expiration, while a European option can be exercised only at expiration.

- The possibility of early exercise makes American options more valuable than otherwise similar European options.

Early exercise is preferred in many cases, e.g.,

- when the underlying asset pays large dividends.
- when an investor holds both the underlying asset and deep in-the-money puts on that asset, at a time when interest rates are high.

Valuing European Call Options

Variables

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

Valuing European options involves solving a stochastic calculus equation, e.g, the Black-Scholes model.
In the simplest case, the option is priced as a conditional expectation relating to an exponentiated normal distribution:

$$\text{Option Price} = \mathbf{E}\left[S_T - K \mid S_T \geq K\right],$$

where

$$\log \frac{S_T}{S(0)} \sim \mathcal{N}(0, \sigma\sqrt{T}).$$

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:

$$\text{Option Price} = S(t^*) - K,$$

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.

DP Formulation

- 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: $g_t = 0$

Bellman Equation

When $t = T$, $J_T(S_T) = \max\{S_T - K, 0\}$, and when $t < T$

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

where the optimal cost vector $J_t(S)$ is the option price at the tth day when the current stock price is S.

A Simple Binomial Model

We focus on American call options.

- Strike price: K
- Duration: T days
- Stock price of tth day: S_t
- Growth rate: $u \in (1, \infty)$
- Diminish rate: $d \in (0, 1)$
- Probability of growth: $p \in [0, 1]$

Binomial Model of Stock Price

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

As the discretization of time becomes finer, the binomial model approaches the Brownian motion model.

DP Formulation for Binomial Model

- Given S_0, T, K, u, r, p.
- State: S_t, finite number of possible values
- Cost vector: $J_t(S)$, the value of option at the tth day when the current stock price is S.

Bellman equation for binomial option

$$J_t(S_t) = \max\left\{S_t - K,\ \ pJ_{t+1}(uS_t) + (1-p)J_{t+1}(dS_t)\right\},$$

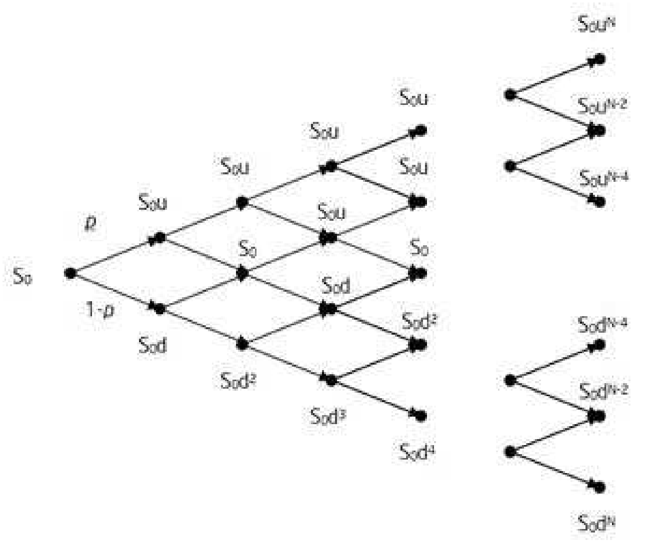
$$J_T(S_T) = \max\{S_T - K, 0\}.$$

Use Exact DP to Evaluate Options

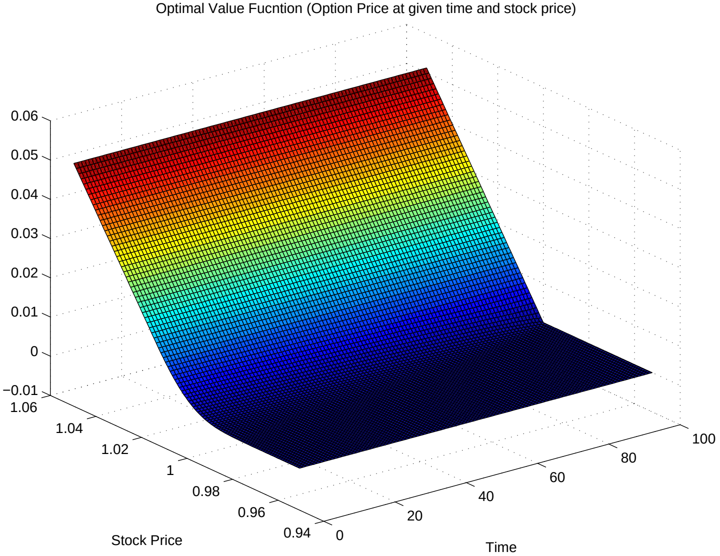
Exercise 1

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

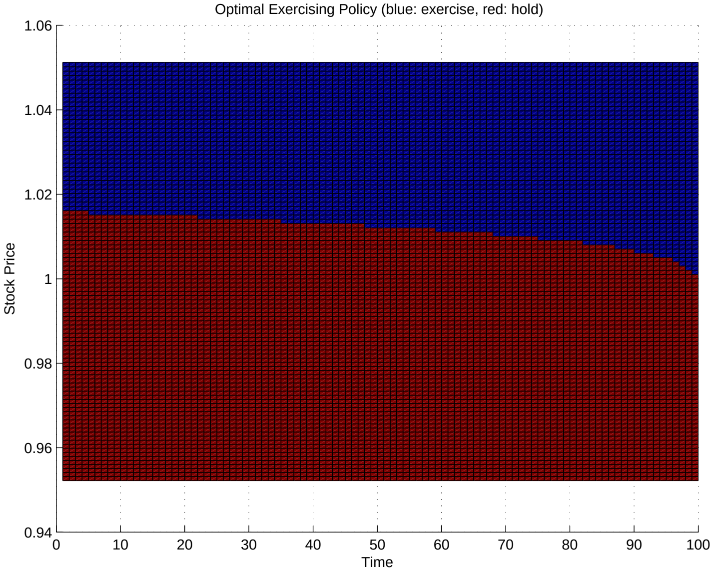
Algorithm Structure: Binomial Tree



Computation Results - Option Prices



Computation Results - Exercising strategy



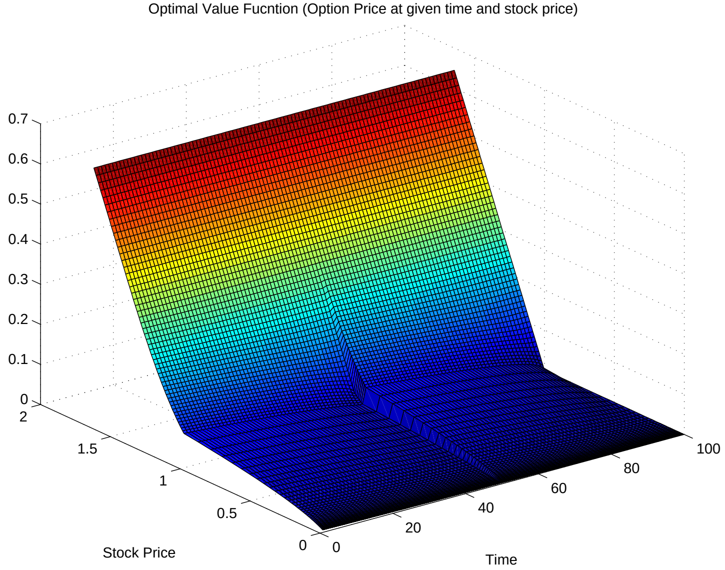
Exercise 2

Option with Dividend

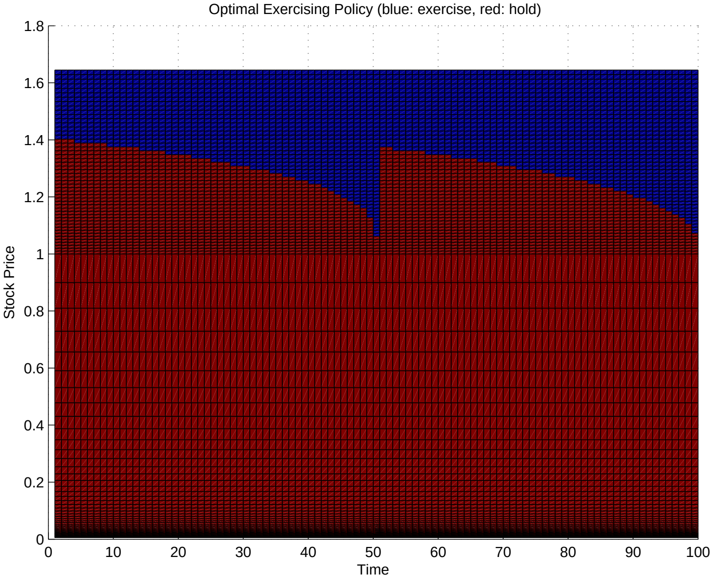
Assume that at time $t = T/2$, the stock will yield a dividend to its shareholders. As a result, the stock price will decrease by $1 - d^{10}$ at time t.

Use this information to modify the program and price the option.

Option prices when there is dividend



Exercising strategy when there is dividend



Infinite-Time Options Pricing

Assume that:

- The option never expires.
- There exists a discount factor $\alpha \in (0, 1)$ (e.g., discount of future, inflation rate, growth rate of alternative investment, etc).

The infinite-time model:

- extends to evaluation of a broad class of securities, including stocks, mortgages, etc.
- is a good approximation to finite-time options when T is large.

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

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

Value Iteration

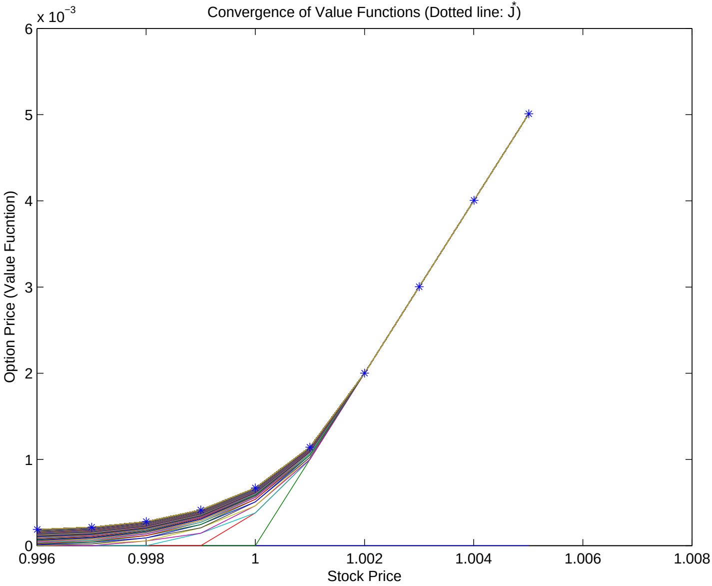
Exercise 3

Use VI to evaluate an infinite-time American call options. The program should be a function of $S_0, \alpha, p, U, D, u, d, K$.

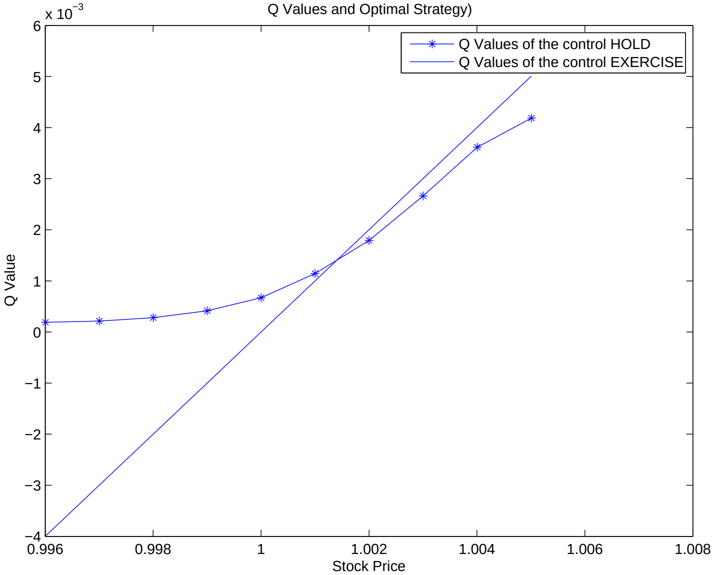
Value Iteration Algorithm

- Start with an arbitrary J_0.
- Repeat the following until convergence:
  - Given J_k.
  - Apply value iteration $J_{k+1} = TJ_k$.
- Obtain the optimal stopping policy from J^*.

Convergence of Value Functions J_t



Optimal Q Values and Exercising Policies



Policy Iteration for Option Pricing

Algorithm

- Starts with any μ_0.
- Policy evaluation:
  - Evaluate J_{μ_t} by VI: applying $T_{\mu_k} J \mapsto J$ till convergence, and yielding J_{μ_k}.
  - Evaluate the Q-values by $Q_{\mu_t}(i_t) = \alpha \mathbf{E}\left[J_{\mu_t}(i_{t+1})\right]$.
- Policy improvement:

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

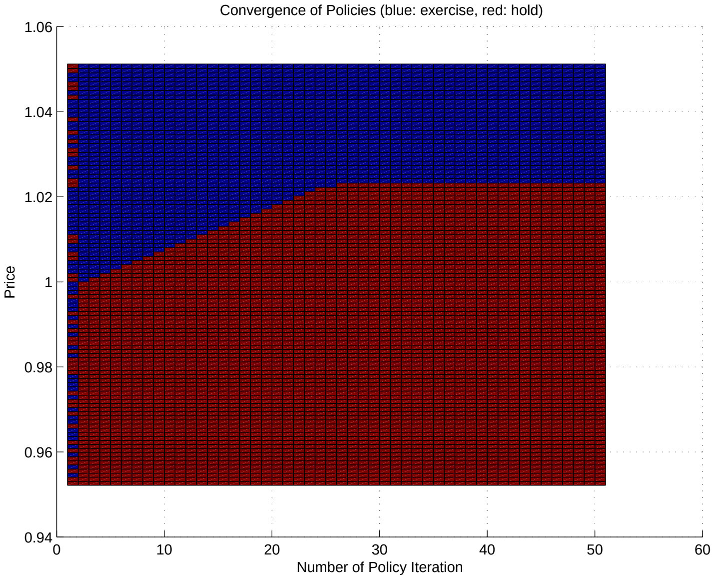
Exercise 4

Modify the code of Exercise 3, to use PI to evaluate an infinite-time American call options. The program should be a function of $S_0, \alpha, p, U, D, u, d, K$.

Suggestions:

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

Convergence of Policies



The end

Thank You Very Much!
Any Question is Welcome :-)