

Macro Topics: Introduction to Matlab

Fall semester 2016

Lecture notes (November 11)

Olga Bychkova

Topics Covered Today

Discrete time dynamic programming: Infinite horizon

- ▶ Preliminaries
- ▶ “Shorthand” notation
- ▶ Properties of a functional operator in shorthand
- ▶ Key results in bounded case: Convergence
- ▶ Key results in bounded case: Bellman equation
- ▶ Key results in bounded case: Optimality of policy

This lecture is based on Dimitri Bertsekas' book, Ch. 1 of V. 2.

DP Algorithm in Infinite Time: Preliminaries I

Consider the following problem:

$$\min J(x_0) = \lim_{N \rightarrow \infty} \mathbb{E}_{\{w_k\}_{k=0}^{N-1}} \left[\sum_{k=0}^{N-1} \alpha^k g(x_k, u_k, w_k) \right],$$

$$\text{s.t. } x_{k+1} = f(x_k, u_k, w_k), \quad u_k \in U(x_k).$$

- ▶ This is a discounted problem, where cost at period k is discounted as α^k , $0 < \alpha < 1$.
- ▶ In infinite time, we are usually interested in stationary problems. In particular, this means that $f_k \equiv f$, $g_k \equiv g$, $\forall k$.
- ▶ The theory is easiest if we assume that the cost g is bounded for any x , u , and w :

$$g(x, u, w) \leq M < \infty.$$

DP Algorithm in Infinite Time: Preliminaries II

- ▶ Suppose policy $\pi = \{\mu_0, \mu_1, \dots\}$ is admissible (meaning that $\mu_k(x_k) \in U(x_k) \forall k$).
- ▶ Split the cost function associated with the policy π , $J_\pi(x)$, into two components. One is the cost accumulated up to period N and another, cost-to-go after N , denoted as $\alpha^N J(x_N)$:

$$J_\pi(x) = \mathbb{E}_{\{w_k\}_{k=0}^{N-1}} \left[\alpha^N J(x_N) + \sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k) \right].$$

- ▶ We will treat the infinite horizon problem as a finite horizon problem with a (so far unknown) terminal cost function $\alpha^N J(x_N)$.
- ▶ The DP algorithm gives

$$J_k(x) = \min_{u \in U(x)} \mathbb{E}_w \left[\alpha^k g(x, u, w) + J_{k+1}(f(x, u, w)) \right],$$

$$k = N - 1 : -1 : 0, \quad J_N(x) = \alpha^N J(x_N).$$

DP Algorithm in Infinite Time: Preliminaries III

- ▶ Now consider value functions

$$V_k(x) = \frac{J_{N-k}(x)}{\alpha^{N-k}}.$$

We get

$$V_N(x) = \frac{J_0(x)}{\alpha^0} = J_0(x) :$$

the optimal cost for the N -stage problem.

- ▶ Introduce a new variable $\bar{k} = N - k$, note that

$$J_{\bar{k}}(x) = \alpha^{\bar{k}} V_k(x),$$

and re-write the DP step as

$$\alpha^{\bar{k}} V_k(x) = \min_{u \in U(x)} \mathbb{E}_w \left[\alpha^{\bar{k}} g(x, u, w) + \alpha^{\bar{k}+1} V_{k-1}(f(x, u, w)) \right],$$

which after division by $\alpha^{\bar{k}}$ becomes

$$V_k(x) = \min_{u \in U(x)} \mathbb{E}_w [g(x, u, w) + \alpha V_{k-1}(f(x, u, w))], \quad V_0(x) = J(x).$$

DP Algorithm in Infinite Time: Preliminaries IV

- ▶ Above, $J(x)$ is the (unknown) function that represents cost-to-go after N , $\alpha^N J(x_N)$.
- ▶ Notice that the solution to the DP problem is much easier than the original backward algorithm: starting from some terminal cost function (cost-to-go), we could solve the $N + 1$ -stage DP problem simply by doing one more iteration of the above expression.
- ▶ This great simplification is possible because of stationarity assumptions.

Infinite Horizon DP Algorithm: “Shorthand” Notation I

- ▶ Introduce mapping T (a functional operator), operating on function to produce a function:

$$(TJ)(x) = \min_{u \in U(x)} \mathbb{E}_w[g(x, u, w) + \alpha J(f(x, u, w))], \quad \forall x.$$

- ▶ TJ is the optimal cost function for the one-stage problem with stage cost g and terminal cost αJ .
- ▶ For any stationary policy $\pi = \{\mu, \mu, \dots\}$,

$$(T_\mu J)(x) = \mathbb{E}_w[g(x, \mu(x), w) + \alpha J(f(x, \mu(x), w))], \quad \forall x.$$

- ▶ We define k th iteration of mapping T as

$$(T^k J)(x) \equiv (T(T^{k-1} J))(x).$$

Infinite Horizon DP Algorithm: “Shorthand” Notation II

- ▶ $(T^k J)(x)$ is the optimal cost for k -stage, α -discounted problem with initial state x , period cost g , and terminal cost $\alpha^k J(x)$.
- ▶ With thus defined T and T_μ , we can write cost functions associated with policy $\pi = \{\mu_0, \mu_1, \dots\}$ and stationary policy $\pi = \{\mu, \mu, \dots\}$ as

$$J_\pi(x) = \lim_{k \rightarrow \infty} (T_{\mu_0} T_{\mu_1} \dots T_{\mu_k} J_0)(x),$$

$$J_\mu(x) = \lim_{k \rightarrow \infty} (T_\mu^k J_0)(x).$$

Infinite Horizon DP Algorithm: Properties of T in Shorthand I

Monotonicity: For $\forall J$ and J' such that $J(x) \leq J'(x) \forall x$, and for $\forall \mu$,

$$(TJ)(x) \leq (TJ')(x), \quad \forall x,$$

$$(T_\mu J)(x) \leq (T_\mu J')(x), \quad \forall x.$$

Operator T preserves inequality relations between functions.

Proof: Both TJ and $T_\mu J$ are one-stage optimal costs for the DP problem with the same stage cost g . If the terminal cost increases, total optimal cost **cannot** decrease.

Infinite Horizon DP Algorithm: Properties of T in Shorthand II

Additivity: For $\forall J$ and \forall scalar r and for $\forall \mu$, the following holds:

$$(T(J + re))(x) = (TJ)(x) + \alpha r, \quad \forall x,$$

$$(T_\mu(J + re))(x) = (T_\mu J)(x) + \alpha r, \quad \forall x.$$

Here $e(x) \equiv 1$ is the unit function.

Operator T translates vertical shift of r into a vertical shift of αr .

Infinite Horizon DP Algorithm: Properties of T in Shorthand III

Proof: Use the definition:

$$(TJ)(x) = \min_{u \in U(x)} \mathbb{E}_w[g + \alpha J],$$

$$\begin{aligned}(T(J+re))(x) &= \min_{u \in U(x)} \mathbb{E}_w[g + \alpha(J+re)] = \min_{u \in U(x)} \mathbb{E}_w[g + \alpha J + \alpha r] = \\ &= \min_{u \in U(x)} \mathbb{E}_w[g + \alpha J] + \alpha r = (TJ)(x) + \alpha r.\end{aligned}$$

Proof for $T_\mu J$ is identical.

Infinite Horizon DP Algorithm: Properties of T in Shorthand IV

Contraction: For any bounded J and J' and $\forall \mu$,

$$\max_x |(TJ)(x) - (TJ')(x)| \leq \alpha \max_x |J(x) - J'(x)|,$$

$$\max_x |(T_\mu J)(x) - (T_\mu J')(x)| \leq \alpha \max_x |J(x) - J'(x)|.$$

Under operator T the maximum distance between images of functions is not more than α times the maximum distance between the original functions.

Infinite Horizon DP Algorithm: Properties of T in Shorthand V

Proof: Again, using the definition,

$$(TJ)(x) = \min_{u \in U(x)} \mathbb{E}_w[g + \alpha J],$$

$$(TJ')(x) = \min_{u \in U(x)} \mathbb{E}_w[g + \alpha J'],$$

$$(TJ)(x) - (TJ')(x) \geq -\alpha \max_x |J(x) - J'(x)|,$$

$$(TJ)(x) - (TJ')(x) \leq \alpha \max_x |J(x) - J'(x)|,$$

$$\max_x |(TJ)(x) - (TJ')(x)| \leq \alpha \max_x |J(x) - J'(x)|.$$

For $T_\mu J$ the proof is identical.

Infinite Horizon DP Algorithm: Properties of T in Shorthand VI

Contraction Mapping Theorem: The fact that both T and T_μ are contractions implies (by the contraction mapping theorem) that:

- (1) T has a unique fixed point J^* : $\forall x, (TJ^*)(x) = J^*(x)$;
- (2) T_μ has a unique fixed point J_μ : $\forall x, (T_\mu J_\mu)(x) = J_\mu(x)$.

Convergence Rate: For $\forall k$,

$$\max_x |(T^k J)(x) - J^*(x)| \leq \alpha^k \max_x |J(x) - J^*(x)|.$$

Infinite Horizon DP. Key Results in Bounded Case.

Convergence I

Convergence: For \forall bounded J and $\forall x$,

$$J^*(x) = \lim_{N \rightarrow \infty} (T^N J)(x).$$

This forms the basis for value function iteration method of solving infinite horizon DP problems: start with arbitrary J , iterate until convergence.

Infinite Horizon DP. Key Results in Bounded Case.

Convergence II

Another method (also relying on the convergence result) is policy function iterations:

- ▶ Policy evaluation: given a candidate optimal policy μ_k , evaluate associated cost J_{μ_k} by solving

$$J_{\mu_k} = T_{\mu_k} J_{\mu_k}.$$

- ▶ Policy improvement: find a new candidate optimal policy μ_{k+1} by applying T to J_{μ_k} (minimizing expected value of $g + \alpha J_{\mu_k}$):

$$T_{\mu_{k+1}} J_{\mu_k} = T J_{\mu_k}.$$

- ▶ Repeat until convergence.

Infinite Horizon DP. Key Results in Bounded Case.

Convergence III

Proof (value iteration only): For \forall policy π ,

$$\begin{aligned} J_{\pi}(x_0) &= \lim_{N \rightarrow \infty} \mathbb{E}_{\{w_k\}_{k=0}^{N-1}} \left[\sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k) \right] = \\ &= \mathbb{E}_{\{w_k\}_{k=0}^{K-1}} \left[\sum_{k=0}^{K-1} \alpha^k g(x_k, \mu_k(x_k), w_k) \right] + \\ &+ \lim_{N \rightarrow \infty} \mathbb{E}_{\{w_k\}_{k=K}^{N-1}} \left[\sum_{k=K}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k) \right]. \end{aligned}$$

Absolute value of the last term is less than $M \sum_{k=K}^{\infty} \alpha^k = \frac{\alpha^K M}{1 - \alpha}$.

Infinite Horizon DP. Key Results in Bounded Case.

Convergence IV

Subtract the lim term from the RHS, add $\alpha^K J(x)$ and bound:

$$\begin{aligned} J_\pi(x_0) - \frac{\alpha^K M}{1 - \alpha} - \alpha^K \max |J(x)| &\leq \\ &\leq \mathbb{E}_{\{w_k\}_{k=0}^{K-1}} \left[\alpha^K J(x) + \sum_{k=0}^{K-1} \alpha^k g(x_k, \mu_k(x_k), w_k) \right], \\ J_\pi(x_0) + \frac{\alpha^K M}{1 - \alpha} + \alpha^K \max |J(x)| &\geq \\ &\geq \mathbb{E}_{\{w_k\}_{k=0}^{K-1}} \left[\alpha^K J(x) + \sum_{k=0}^{K-1} \alpha^k g(x_k, \mu_k(x_k), w_k) \right]. \end{aligned}$$

Infinite Horizon DP. Key Results in Bounded Case.

Convergence V

Now take minimum over all π (J_π becomes J^* , and the min \mathbb{E} term, by definition, is $T^K J$):

$$\begin{aligned} J^*(x_0) - \frac{\alpha^K M}{1 - \alpha} - \alpha^K \max |J(x)| &\leq (T^K J)(x_0) \leq \\ &\leq J^*(x_0) + \frac{\alpha^K M}{1 - \alpha} + \alpha^K \max |J(x)|. \end{aligned}$$

And finally, let $K \rightarrow \infty$:

$$J^*(x_0) = \lim_{K \rightarrow \infty} (T^K J)(x_0).$$

Infinite Horizon DP. Key Results in Bounded Case.

Bellman Equation I

Bellman Equation: J^* satisfies

$$J^*(x) = \min_{u \in U(x)} \mathbb{E}_w[g(x, u, w) + \alpha J^*(f(x, u, w))], \quad \forall x,$$

$$J^* = TJ^*.$$

Moreover, J^* is a unique solution in a class of bounded functions.

Infinite Horizon DP. Key Results in Bounded Case.

Bellman Equation II

Proof:

- ▶ Using the previous result for $J_0(x) \equiv 0$,

$$J^*(x) - \frac{\alpha^K M}{1 - \alpha} \leq (T^K J_0)(x) \leq J^*(x) + \frac{\alpha^K M}{1 - \alpha}, \quad \forall x.$$

- ▶ Apply T to all sides, use monotonicity property to get

$$(TJ^*)(x) - \frac{\alpha^{K+1} M}{1 - \alpha} \leq (T^{K+1} J_0)(x) \leq (TJ^*)(x) + \frac{\alpha^{K+1} M}{1 - \alpha}, \quad \forall x.$$

- ▶ Take limit as $K \rightarrow \infty$, note that because of convergence the middle part becomes J^* :

$$(TJ^*)(x) \leq J^*(x) \leq (TJ^*)(x),$$

$$J^*(x) = (TJ^*)(x), \quad \forall x.$$

Infinite Horizon DP. Key Results in Bounded Case.

Bellman Equation III

- ▶ To prove uniqueness, take any bounded J that satisfies $J = TJ$, then

$$J = \lim_{N \rightarrow \infty} T^N J = J^*$$

by the convergence property. Therefore, J^* is unique.

For a stationary policy μ ($\pi = \{\mu, \mu, \dots\}$), the proof with J_μ is identical.

Infinite Horizon DP. Key Results in Bounded Case.

Optimality of Policy I

Optimality of Policy: Stationary policy μ is optimal if and only if $\mu(x)$ attains minimum in the Bellman equation, or, in shorthand notation,

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

Infinite Horizon DP. Key Results in Bounded Case.

Optimality of Policy II

Proof: (\Leftarrow)

Suppose that $TJ^* = T_\mu J^*$. Then we have

$$TJ^* = J^*(BE) = T_\mu J^*.$$

But J_μ is the unique solution of the BE $J_\mu = TJ_\mu$, therefore, $J^* = T_\mu J^*$ implies $J^* = J_\mu$, or μ is indeed optimal policy.

Infinite Horizon DP. Key Results in Bounded Case.

Optimality of Policy III

Proof: (\Rightarrow)

Suppose now that μ is optimal, meaning $J^* = J_\mu = T_\mu J_\mu$. Again, J_μ is unique, therefore,

$$J^* = TJ^*(BE) = T_\mu J^*,$$

$$TJ^* = T_\mu J^*,$$

or μ attains minimum in the Bellman equation.

We are guaranteed that optimal stationary policy exists if we can attain a minimum in the Bellman equation. In particular, it is true if the set U is finite.