Abstract and Semicontractive Dynamic Programming

Dimitri P. Bertsekas

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Lecture 1 of 5

July 2016

Classical Total Cost Stochastic Optimal Control (SOC)

System: $x_{k+1} = f(x_k, u_k, w_k)$

- x_k: State at time k, from some space X
- u_k: Control at time k, from some space U
- w_k: Random "disturbance" at time k, from a countable space W, with $p(w_k \mid x_k, u_k)$ given

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

- Each μ_k maps states x_k to controls $u_k = \mu_k(x_k) \in U(x_k)$ (a constraint set)
- Cost of π starting at x_0, with discount factor $\alpha \in (0, 1]$:

$$J_\pi(x_0) = \limsup_{N\to\infty} E\left\{\sum_{k=0}^{N} \alpha^k g(x_k, \mu_k(x_k), w_k)\right\}$$

- Optimal cost starting at x_0: $J^*(x_0) = \inf_\pi J_\pi(x_0)$
- Optimal policy π^*: Satisfies $J_{\pi^*}(x) = J^*(x)$ for all $x \in X$
- Stationary policies, those of the form $\{\mu, \mu, \ldots\}$, play a special role (typically, there are stationary optimal policies that are optimal)

Bellman's Equation

- The cost of a stationary policy μ starting from state x, denoted $J_\mu(x)$, typically satisfies

$$J_\mu(x) = E\big\{g(x,\mu(x),w) + \alpha J_\mu\big(f(x,\mu(x),w)\big)\big\}, \qquad \forall\, x \in X$$

  This is called Bellman's equation for policy μ

- The optimal cost starting from state x, denoted $J^*(x)$, typically satisfies

$$J^*(x) = \inf_{u \in U(x)} E\big\{g(x,u,w) + \alpha J^*\big(f(x,u,w)\big)\big\}, \qquad \forall\, x \in X$$

  This is called Bellman's equation

- Both types of Bellman's equation are functional equations in J_μ or J^*
- They can be viewed abstractly as having the form

$$J_\mu = T_\mu J_\mu \qquad \text{or} \qquad J^* = TJ^*$$

- In a given DP problem it is significant when Bellman's equation has a unique solution. This is true if all T_μ are contraction mappings (with a common modulus). If this is true, the problem is called contractive and otherwise noncontractive
- Contractive problems are much nicer!

Two Main Classes of Total Cost SOC Problems

Contractive Problems:

- $\alpha < 1$ and bounded g
- Date to 50s (Bellman, Shapley)
- Nicest results; key fact is the contraction property of the mapping in Bellman's equation

Noncontractive Problems - Stochastic Shortest Path (SSP):

- Date to 60s (Eaton-Zadeh, Derman, Pallu de la Barriere)
- Also known as first passage or transient programming
- Aim is to reach a special termination state at min expected cost
- Under favorable assumptions, the results are almost as strong as for the discounted case (when the noncontractive policies cannot be optimal)
- In general, very complex behavior is possible

Some Additional Noncontractive Problems:

- Discounted problems with unbounded g
- Undiscounted problems with positive and negative cost ($g \leq 0$ or $g \geq 0$)

Semicontractive Problems

Intermediate Problem Types: Between Contractive and Noncontractive

- Problems where some policies are “well-behaved" and some are not
- “Well-behaved" has a problem-dependent meaning. The most common example of “well-behaved" policy is one that is contractive
- Pathological behaviors are due to policies that are not “well-behaved"

Our Approach

- Select a class of well-behaved policies (we call them regular and define them in a precise way later)
- Define a restricted optimization problem over the regular policies only
- Show that the restricted problem has nice theoretical and algorithmic properties
- Relate the restricted problem to the original
- Under reasonable conditions, obtain strong theoretical and algorithmic results

References

Research Monograph

D. P. Bertsekas, Abstract Dynamic Programming, Athena Scientific, 2013; updates on-line.

Subsequent Papers

- D. P. Bertsekas, “Regular Policies in Abstract Dynamic Programming," Lab. for Information and Decision Systems Report LIDS-P-3173, MIT, May 2015.
- D. P. Bertsekas, “Affine Monotonic and Risk-Sensitive Models in Dynamic Programming", Lab. for Information and Decision Systems Report LIDS-3204, MIT, June 2016.
- D. P. Bertsekas, ‘Robust Shortest Path Planning and Semicontractive Dynamic Programming", Naval Research Logistics J., to appear.
- D. P. Bertsekas, “Value and Policy Iteration in Optimal Control and Adaptive Dynamic Programming," IEEE Transactions on Neural Networks and Learning Systems, to appear.
- D. P. Bertsekas and H. Yu, “Stochastic Shortest Path Problems Under Weak Conditions," Lab. for Information and Decision Systems Report LIDS-P-2909, MIT, January 2016.

Contents of the Lecture Series

1. Semicontractive Examples.
2. Semicontractive Analysis for Stochastic Optimal Control.
3. Extensions to Abstract DP Models.
4. Applications to Stochastic Shortest Path and Other Problems.
5. Algorithms.

Results that we Aim for in Total Cost DP Problems

Denote

$$H(x,u,J) = E\big\{g(x,u,w) + \alpha J\big(f(x,u,w)\big)\big\}$$

- J^* satisfies Bellman's equation

$$J^*(x) = \inf_{u \in U(x)} H(x,u,J^*), \qquad \forall\, x \in X$$

  and if $\mu^*(x)$ attains the min for all x, μ^* is optimal

- The value iteration (VI) method converges: $J_k \to J^*$, where

$$J_{k+1}(x) = \inf_{u \in U(x)} H(x,u,J_k)$$

- The policy iteration (PI) method converges: $J_{\mu^k} \to J^*$, where $\{\mu^k\}$ is generated by

$$J_{\mu^k}(x) = H\big(x,\mu^k(x),J_{\mu^k}\big), \quad \forall\, x \in X, \qquad \text{(policy evaluation)}$$

$$\mu^{k+1}(x) \in \arg\min_{u \in U(x)} H(x,u,J_{\mu^k}), \quad \forall\, x \in X. \qquad \text{(policy improvement)}$$

- In all examples, we introduce a set of “well-behaved" or “regular" policies (in shortest path problems, regular policies will be the ones that reach the termination state in finite time).
- Let

$$J^*(x)\text{: Optimal cost (over all policies) starting from } x$$

$$\hat{J}(x)\text{: Optimal cost over the regular policies only, starting from } x$$

The Four Examples

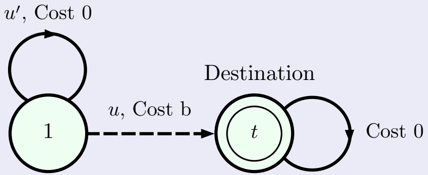
- A finite-state, finite-control deterministic shortest path problem. Here Bellman's equation may have multiple solutions (including J^* and $\hat{J}$), and VI and PI may not converge to J^* or to $\hat{J}$
- A finite-state, finite-control stochastic shortest path problem. Here J^* does not satisfy Bellman's equation, while $\hat{J}$ does
- A finite-state, infinite-control stochastic shortest path problem. Here there is no optimal policy, and VI and PI exhibit some peculiarities
- A linear-quadratic optimal control problem. Here Bellman's equation has two solutions, J^* and $\hat{J}$, and VI and PI typically converge to $\hat{J}$

A Deterministic Shortest Path Problem

Stationary policy costs

$J_\mu(1) = b,\ J_{\mu'}(1) = 0$

Optimal cost $J^*(1) = \min\{b, 0\}$



Bellman's equation: $J(1) = \min\{b,\ J(1)\}$. Set of solutions: All $J(1) \leq b$

μ is well-behaved/regular, but μ' is not; here $\hat{J} = b$, $J^* = \min\{b, 0\}$

Value iteration (VI) starting from any $J_0(1)$: $J_{k+1}(1) = \min\{b, J_k(1)\}$

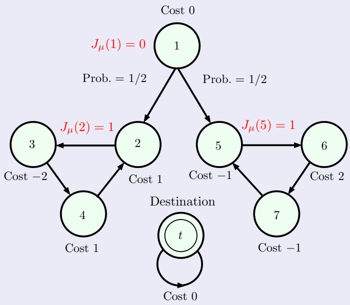
- If $b < 0$: $J_k(1) \to J^*(1)$ starting with $J_0(1) \geq b$ (works depending on J_0)
- If $b > 0$: $J_k(1) \to J^*(1)$ only if $J_0(1) = 0$; starting from $J_0(1) \geq b$, $J_k(1) \to \hat{J}(1)$
- VI for the regular policy μ: $J_{\mu,\,k}(1) = b$ (works)
- VI for the irregular policyμ': $J_{\mu',\,k+1}(1) = J_{\mu',\,k}(1)$ (fails)

Policy iteration (PI) starting from μ

If $b < 0$: Oscillates between μ and μ'. If $b > 0$: Converges to suboptimal μ

A Stochastic Shortest Path Problem (from Bertsekas and Yu, 2015)

A single policy μ. The only uncertainty is at the first stage starting at state 1.

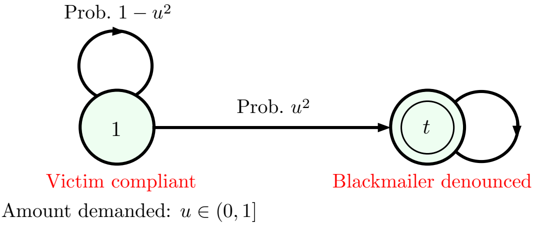


The Bellman Eq. is violated at 1: $J_\mu(1) \neq \frac{1}{2}J_\mu(2) + \frac{1}{2}J_\mu(5)$

A peculiar phenomenon

Consider the deterministic optimal control problem where at state 1 we may choose either to go to 2 or to 5 at zero cost

- Then $J^*(x) = 1$ for all x, including $J^*(1) = 1$
- Belman's equation $J^*(1) = \min\{J^*(2), J^*(5)\}$ is satisfied
- Randomization lowers the optimal cost and invalidates Bellman's equation



- Every policy μ terminates with probability 1, and $J_\mu(1) = -\frac{1}{\mu(1)}$
- We have $J^*(1) = -\infty$ and there exists no optimal policy
- Bellman's equation is

$$J(1) = \min_{0<u\leq 1} \left\{ -u + (1 - u^2)J(1) \right\}$$

  It is satisfied by $J^* = -\infty$ (also by $J = \infty$)
- VI converges to J^* starting from any scalar J
- In PI we have $J_{\mu^k} \to J^*$, but $\mu^k(1) \to 0$ (which is not an admissible policy)
- A variation of the problem: Replacing the probability u^2 by u. Then $J^*(1) = -1$ is a solution of Bellman's Eq., but all $J \leq -1$ are also solutions, and still there is no optimal policy

A Linear Quadratic Problem

System: $x_{k+1} = \gamma x_k + u_k$, Cost per stage: $g(x, u) = u^2$

- Here $J^*(x) \equiv 0$ and the optimal policy is $\mu^*(x) \equiv 0$
- Bellman's equation is

$$J(x) = \min_{u \in \Re} \left\{ u^2 + J(\gamma x + u) \right\}, \qquad x \in \Re,$$

  and is satisfied by J^*. Are there any other solutions?

Let $\gamma > 1$, so the system is unstable

- The optimal policy yields an unstable closed-loop system
- Bellman's equation has a second solution: $\hat{J}(x) = (\gamma^2 - 1)x^2$
- $\hat{J}$ is the optimal cost function over the class of policies that stabilize the system (these are the "well-behaved" or "regular" policies)
- Both VI and PI typically converge to $\hat{J}$ (not J^*!)

A Summary from the Examples

- Bellman's equation may have multiple solutions
- Often but not always, J^* is a solution
- A restricted problem, involving "well-behaved" policies, is meaningful and plays an important role
- The appropriate set of "well-behaved" policies is problem-dependent (e.g., terminating in shortest path problems, or stabilizing in the linear quadratic case)
- The optimal cost function over all policies, J^*, may differ from $\hat{J}$, the optimal cost function over the "well-behaved" policies
- $\hat{J}$ is the likely limit of the VI and the PI algorithms, starting from an appropriate set of initial conditions

In the next lecture, we will aim to:

- Explain this behavior through analysis
- Formalize the notion of "well-behaved" policy through a notion of regularity
- Introduce the kind of assumptions under which anomalous behavior can be avoided or mitigated
- Provide results of the type that are available for contractive problems