Semicontractive Dynamic Programming

Dimitri P. Bertsekas

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Lecture 2 of 5

July 2016

Contents of the Lecture Series

- Semicontractive Examples.
- Semicontractive Analysis for Stochastic Optimal Control.
- Extensions to Abstract DP Models.
- Applications to Stochastic Shortest Path and Other Problems.
- Algorithms.

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: $\pi = \{\mu_0, \mu_1, \dots\}$

- 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 $\alpha \in (0, 1]$:

$$J_\pi(x_0) = \limsup_{k \to \infty} E\left\{\sum_{m=0}^{k} \alpha^m g(x_m, \mu_m(x_m), w_m)\right\}$$

- 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 $\{\mu, \mu, \dots\}$, play a special role

Bellman's Equations In Shorthand

- The cost of a stationary policy μ starting from state x, denoted $J_\mu(x)$, and the optimal cost starting from state x, denoted $J^*(x)$, typically satisfy Bellman's equations

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

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

Denote for all $x \in X$ and μ,

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

$$(T_\mu J)(x) = H\big(x,\mu(x),J\big), \qquad (TJ)(x) = \inf_{u \in U(x)} H(x,u,J)$$

A key property is that T_μ and T are monotone

- The Bellman equations can be viewed abstractly as the fixed point equations $J_\mu = T_\mu J_\mu$ and $J^* = TJ^*$
- We are considering semicontractive problems where some T_μ are "contraction-like" and others are not

Results that we Aim for in Total Cost SOC

- Characterization of the set of solutions of Bellman's equations

$$J = T_\mu J, \qquad J = TJ$$

  Are J_μ and J^* solutions?

- If $\mu^*(x)$ attains the min for all x,

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

  (i.e., $T_{\mu^*} J^* = TJ^*$ in shorthand), then μ^* is optimal

- The value iteration (VI) method converges: $\{T^k J\} \to J^*$ for appropriate initial J

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

$$J_{\mu^k} = T_{\mu^k} J_{\mu^k}, \qquad \text{(policy evaluation)}$$

  and

$$T_{\mu^{k+1}} J_{\mu^k} = TJ_{\mu^k}, \qquad \text{(policy improvement)}$$

We Gave Four Pathological Examples: What is the Root of the Anomalies?

A (partial) answer

The presence of policies that are not well-behaved in terms of VI (e.g., involve zero length cycles in shortest path problems, or are unstable in linear quadratic problems)

We call these policies “irregular" and we investigate

- What problems can they cause?
- Under what assumptions are they “harmless"?

Our Approach for Semicontractive Problems

- Select a class of well-behaved/regular policies
- 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

Research Monograph

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

Outline of this Lecture

1. *S*-Regular Policies
2. *S*-Regular Restricted Optimization
3. Weak and Strong PI Properties
4. Stochastic Shortest Path Example
5. Deterministic Optimal Control Example

Shorthand Notation for Cost Functions

- $\Re$ the set of real numbers $(-\infty, \infty)$
- $R(X)$ the set of real-valued functions $J : X \mapsto \Re$
- $E(X)$ the set of extended real-valued functions $J : X \mapsto [-\infty, \infty]$
- $\bar{J}$ is the identically 0-function, $\bar{J}(x) \equiv 0$
- N-stage cost of policy μ starting from x with terminal cost function J:

$$(T_\mu^N J)(x) = E\left\{\alpha^N J(x_N) + \sum_{k=0}^{N-1} \alpha^k g\big(x_k, \mu(x_k), w_k\big)\right\}$$

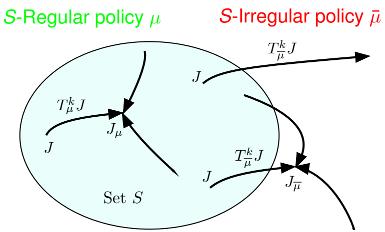
- k-stage cost of policy μ starting from x with 0 terminal cost:

$$(T_\mu^k \bar{J})(x) = E\left\{\sum_{m=0}^{k-1} \alpha^m g\big(x_m, \mu(x_m), w_m\big)\right\}$$

- Infinite horizon cost of policy μ starting from x:

$$J_\mu(x) = \limsup_{k \to \infty} (T_\mu^k \bar{J})(x)$$

S-Regular Policies



Definition of S-Regular Policy

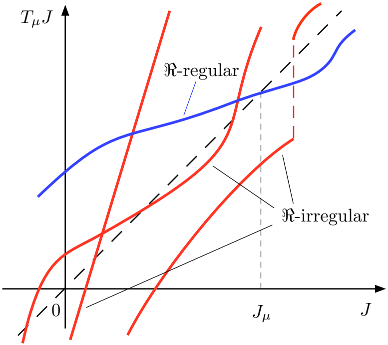
Given a set of functions $S \subset E(X)$, we say that a stationary policy μ is S-regular if:

- $J_\mu \in S$ and $J_\mu = T_\mu J_\mu$
- $T_\mu^k J \to J_\mu$ for all $J \in S$

A policy that is not S-regular is called S-irregular.

- The S-regular μ are the ones for which J_μ is the unique and stable equilibrium point of T_μ within S
- The S-regular μ are well-behaved" with respect to VI (at least within S)

S-Regular Policies: Illustration for $S = \Re$



Definition of S-Regularity

Given a set of functions $S \subset E(X)$, a stationary policy μ is S-regular if:

- $J_\mu \in S$ and $J_\mu = T_\mu J_\mu$
- $T_\mu^k J \to J_\mu$ for all $J \in S$

Examples of S-Regular Policies in SOC

Definition of S-Regularity

Given a set of functions $S \subset E(X)$, a stationary policy μ is S-regular if:

- $J_\mu \in S$ and $J_\mu = T_\mu J_\mu$
- $T_\mu^k J \to J_\mu$ for all $J \in S$

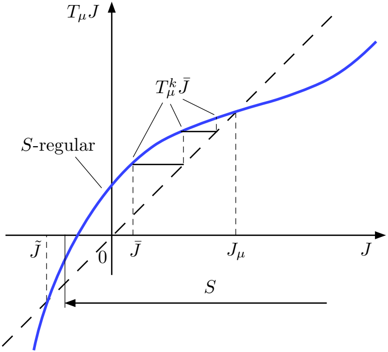
Examples

- All μ such that $J_\mu \in S$ and T_μ is a contraction mapping over S are S-regular: i.e., $T_\mu : S \mapsto S$ and

  $$\| T_\mu J - T_\mu J' \| \leq \beta \| J - J' \|, \qquad \forall\, J, J' \in S$$

  where $\beta \in (0, 1)$
- n-state shortest path problems: If $S = \Re^n$, the S-regular policies are precisely the terminating policies
- In linear quadratic problems: If S is the set of positive semidefinite quadratic functions, the linear stabilizing controllers are S-regular

S-Regular Policies: Dependence on the Choice of S

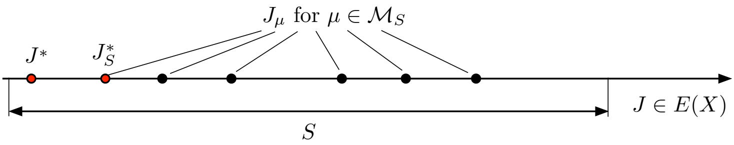


Definition of S-Regularity

Given a set of functions $S \subset E(X)$, a stationary policy μ is S-regular if:

- $J_\mu \in S$ and $J_\mu = T_\mu J_\mu$
- $T_\mu^k J \to J_\mu$ for all $J \in S$

S-Regular Restricted Problem



Given a set $S \subset E(X)$

- Let $\mathcal{M}_S$ be the set of all S-regular policies
- Consider the restricted optimization problem: Minimize J_μ over the S-regular μ
- Let J_S^* be the optimal cost function over S-regular policies only:

$$J_S^*(x) = \inf_{\mu \in \mathcal{M}_S} J_\mu(x), \qquad x \in X$$

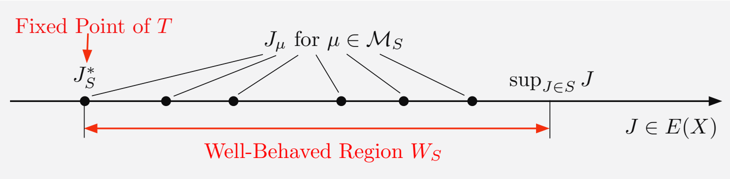
- Since S-regular policies is a subset of the set of all policies,

$$J^* \leq J_S^*$$

Given a set $S \subset E(X)$ consider

$$J_S^*(x) = \inf_{\mu \in \mathcal{M}_S} J_\mu(x), \qquad x \in X$$

where $\mathcal{M}_S$ is the set of all S-regular policies

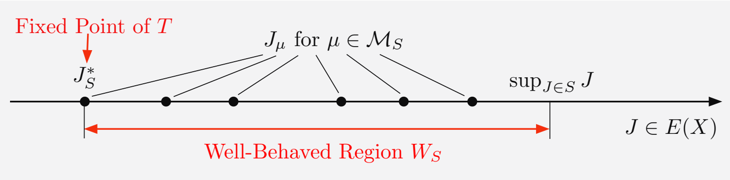


Proposition

Assume that J_S^* is a fixed point of T. Then:

- (Uniqueness of fixed point) J_S^* is the only fixed point of T within the set $W_S = \{J \in E(X) \mid J_S^* \leq J \leq \tilde{J} \text{ for some } \tilde{J} \in S\}$
- (VI convergence) $T^k J \to J_S^*$ for every $J \in W_S$
- (Optimality condition) If μ^* is S-regular, $J_S^* \in S$, and $T_{\mu^*} J_S^* = T J_S^*$, then μ^* is $\mathcal{M}_S$-optimal. Conversely, if μ^* is $\mathcal{M}_S$-optimal, then $T_{\mu^*} J_S^* = T J_S^*$.

Proof Argument



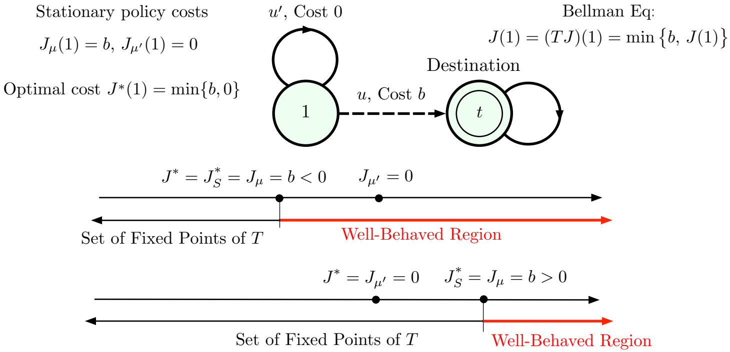
- Let $J \in W_S$, so that $J_S^* \leq J \leq \tilde{J}$ for some $\tilde{J} \in S$. We have for all k and S-regular μ,

$$J_S^* = T^k J_S^* \leq T^k J \leq T^k \tilde{J} \leq T_\mu^k \tilde{J} \qquad \Longrightarrow \qquad T^k J \to J_S^*$$

- If J' is another fixed point of T that belongs to W_S, start VI at J' to get that $J' = \lim_{k\to\infty} T^k J' = J_S^*$.
- If μ^* satisfies $T_{\mu^*} J_S^* = T J_S^*$, then $T_{\mu^*} J_S^* = J_S^*$, implying that $J_S^* = J_{\mu^*}$. Thus μ^* is $\mathcal{M}_S$-optimal
- Conversely, if μ^* is M_S-optimal, we have $J_{\mu^*} = J_S^*$, so

$$T J_S^* = J_S^* = J_{\mu^*} = T_{\mu^*} J_{\mu^*} = T_{\mu^*} J_S^*$$

Deterministic Shortest Path Example; $S = \Re$



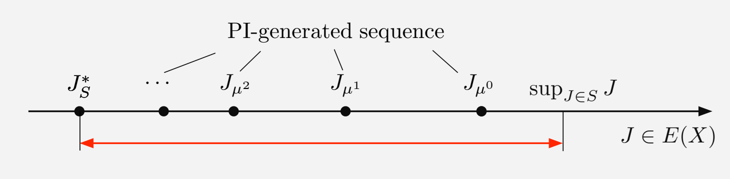
- J^*_S is the only fixed point of T in the well-behaved region
- $T^k J \to J^*_S$ for every J in the well-behaved region $J \in E(X)$

How do we Choose S so that J_S^* is a Fixed Point of T?

One approach: Choose S so that $J_S^* = J^*$, and prove that J^* is a fixed point of T. J^* is a fixed point of T for several broad classes of problems, e.g., all deterministic problems, all problems where the cost per stage g is uniformly nonnegative, or uniformly nonpositive, etc

We will Follow Another Approach Based on Policy Iteration (PI)

- The approach applies when S is “well-behaved" with respect to PI: roughly, starting from an S-regular policy μ^0, PI generates S-regular policies
- The significance of S-regularity is that $\{J_{\mu^k}\}$ is monotonically nonincreasing, and its limit is a fixed point of T



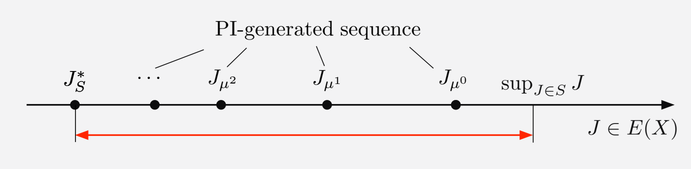
The Weak PI property

Key Fact

If $\{\mu^k\}$ is generated by PI and consists of S-regular policies then

$$J_S^* \leq J_{\mu^{k+1}} \leq J_{\mu^k}, \qquad \forall\, k$$

Proof: $J_{\mu^k} = T_{\mu^k} J_{\mu^k} \geq T J_{\mu^k} = T_{\mu^{k+1}} J_{\mu^k} \geq \lim_{m\to\infty} T^m_{\mu^{k+1}} J_{\mu^k} = J_{\mu^{k+1}}$

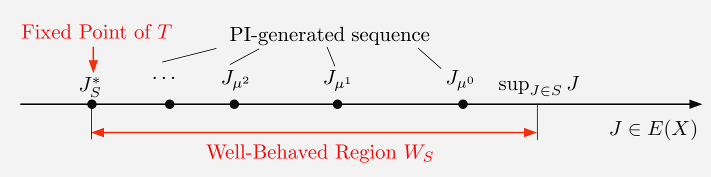


We distinguish between two versions of PI-related assumptions, strong and weak, which lead to corresponding strong and weak results

Definition

We say that S has the weak PI property if there exists a sequence of S-regular policies $\{\mu^k\}$ that can be generated by the PI algorithm

Weak PI Property Theorem



Let S have the weak PI property and $\{\mu^k\}$ be a generated sequence of S-regular policies. Then:

- $J_{\mu^k} \downarrow J_S^*$ and J_S^* is a fixed point of T
- The well-behaved theorem applies, so J_S^* is the only fixed point of T within the well-behaved region, and VI converges to J_S^* starting from within that region

Proof Argument

- S-regularity guarantees that J_{μ^k} is monotonically nonincreasing so $J_{\mu^k} \downarrow J_\infty$
- We have

$$J_{\mu^{k+1}} \leq T_{\mu^{k+1}} J_{\mu^k} = T J_{\mu^k} \leq T_{\mu^k} J_{\mu^k} = J_{\mu^k}$$

- By a limit and continuity argument, J_∞ is a fixed point of T and $J_\infty = J_S^*$

The Strong PI property

Definition

We say that S has the strong PI property if it has the weak PI property and starting from an S-regular policy, PI generates only S-regular policies (i.e., the set of S-regular policies is nonempty and is closed under PI)

Verifying the Strong PI Property for $S = R(X)$

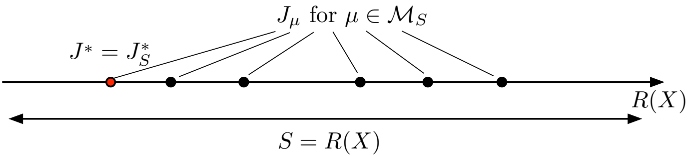
S has the strong PI property if:

- There exists at least one S-regular policy
- The set
$$\{u \in U(x) \mid H(x, u, J) \leq \lambda\}$$
is compact for every $J \in S$, $x \in X$, and $\lambda \in \Re$.
- For every $J \in S$ and S-irregular policy μ, there exists a state $x \in X$ such that
$$\limsup_{k \to \infty} (T_\mu^k J)(x) = \infty$$
(so S-irregular policies cannot be optimal)

Strong PI Property Theorem for $S = R(X)$

Assume the conditions of the preceding slide hold (so that the strong PI property holds), and also $J_S^* \in R(X)$ and is bounded below. Then:

- $J_S^* = J^*$, and J^* is the unique fixed point of T within $R(X)$
- We have $T^k J \to J^*$ for every $J \in R(X)$
- A policy μ^* is optimal if and only if $T_{\mu^*} J^* = T J^*$, and there exists at least one optimal policy
- Starting from an S-regular policy μ^0, PI generates a sequence $\{\mu^k\}$ of S-regular policies such that $J_{\mu^k} \downarrow J^*$



Note the stronger conclusions:

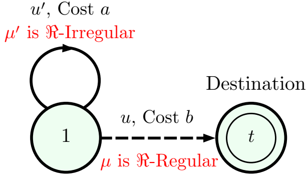
- $J_S^* = J^*$ and is the unique fixed point of T within $R(X)$ (not just within W_S)
- VI converges starting anywhere within $R(X)$
- PI converges assuming we know an initial S-regular policy

Revisit: Deterministic Shortest Path Example, $S = \Re$

Bellman Eq: $J(1) = \min\{b,\ a + J(1)\}$

PI Improvement Eq:

$\mu^1 \in \arg\min\{b,\ a + J_{\mu^0}\} = \arg\min\{b,\ a + b\}$



Case $a = 0$, Weak PI Property Holds

- $J_S^* = b$, $J^* = \min\{b, 0\}$
- Set of fixed points of $T = (-\infty, b]$
- Well-behaved region is $W_S = \{J \mid J_S^* \leq J\} = [b, \infty)$
- J_S^* is the unique fixed point of T within W_S (but if $b > 0$, we have $J^* < J_S^*$)
- VI converges to J^* starting anywhere within W_S
- PI convergence is problematic

Case $a > 0$, Strong PI Property Holds

- $J_S^* = J^*$ is the unique fixed point of T within S
- VI converges to J^* for all initial conditions. PI also converges ...

Stochastic Shortest Path (SSP) Problems

Problem Formulation

- Finite state space $X = \{1, \ldots, n\}$ plus a termination state t
- Transition probabilities $p_{xy}(u)$
- $U(x)$ is finite for all $x \in X$
- Cost per stage $g(x, u)$ and no discounting ($\alpha = 1$)

Proper policies

- μ is proper if the termination state t is reached w.p.1 under μ (is improper otherwise)
- Let $S = R(X) = \Re^n$. Then μ is S-regular if and only if it is proper

Contraction properties

- The mapping T_μ of a policy μ is a weighted sup-norm contraction iff μ proper
- If all stationary policies are proper, then T is a sup-norm contraction, and the problem behaves like a discounted problem
- SSP is a prime example of a semicontractive model (when some policies are proper/contractions/regular while others are not)

Stochastic Shortest Path Problems Results

Case where improper policies have infinite cost (strong PI property holds)

If there exists a proper policy and for every improper μ, $J_\mu(x) = \infty$ for some x, then:

- J^* is the unique fixed point of T in $\Re^n$
- VI converges to J^* starting from every $J \in \Re^n$
- PI converges to an optimal proper policy, if started with a proper policy

Case where improper policies have finite cost (due to zero length “cycles")

Let $\hat{J}$ be the optimal cost function over proper stationary policies only, and assume that $\hat{J}$ and J^* are real-valued. Then, by the weak PI property theorem:

- $\hat{J}$ is the unique fixed point of T in the set $\{J \in \Re^n \mid J \geq \hat{J}\}$
- VI converges to $\hat{J}$ starting from any $J \geq \hat{J}$
- PI need not converge to an optimal policy even if started with a proper policy
- A related line of analysis also applies: Use a “perturbed" version of the problem (add a $\delta_k > 0$ to g, with $\delta_k \downarrow 0$)
- A “perturbed" version of PI (add a $\delta_k > 0$ to g, with $\delta_k \downarrow 0$) converges to an optimal policy within the class of proper policies, if started with a proper policy
- An improper policy may be (overall) optimal, while J^* need not be a fixed point of T

Application to Nonnegative Cost Deterministic Optimal Control

Classic problem of regulation to a terminal set

- System: $x_{k+1} = f(x_k, u_k)$. Cost per stage: $g(x_k, u_k) \geq 0$
- Cost-free and absorbing terminal set of states X_0 that we aim to reach or approach asymptotically at minimum cost
- Let $S = \{J \in E^+(X) \mid J(x) = 0, \ \forall \, x \in X_0\}$
- The terminating policies (reach X_0 in a finite number of steps from all x with $J^*(x) < \infty$) are S-regular

Assumptions (implying that the strong PI property holds)

- $J^*(x) > 0$ for all $x \notin X_0$ (implies that S-irregular policies have infinite cost)
- Controllability: For all x with $J^*(x) < \infty$ and $\epsilon > 0$, there exists a terminating policy μ that reaches X_0 starting from x with cost $J_\mu(x) \leq J^*(x) + \epsilon$
- A compactness condition

Results

- J^* is the unique solution of Bellman's equation within S
- VI converges to J^* starting from any $J \in S$
- PI converges to J^* starting from a terminating policy