Semicontractive Dynamic Programming

Dimitri P. Bertsekas

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Lecture 3 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.
Outline of this Lecture

1. Abstract Dynamic Programming
2. Results Overview
3. Semicontractive Models
4. Semicontractive Analysis
Main Objective

- **Unification** of the core theory and algorithms of total cost sequential decision problems
- Simultaneous treatment of a variety of problems: stochastic optimal control, Markovian decision problems (MDP), sequential games, sequential minimax, multiplicative cost, risk-sensitive, etc

Methodology

- Define a problem by its "**mathematical signature**": the mapping defining the optimality/Bellman equation
- Structure of this mapping (monotonicity, contraction, "semicontractive" properties, etc) determines the analytical and algorithmic theory of the problem
- **Fixed point theory**: An important connection
Abstract DP Mappings

- State and control spaces: \( X, U \)
- Control constraint: \( u \in U(x) \)
- Stationary policies: \( \mu : X \mapsto U \), with \( \mu(x) \in U(x) \) for all \( x \)

Monotone Mappings

- Abstract monotone mapping \( H : X \times U \times E(X) \mapsto \mathbb{R} \)

\[
J \leq J' \implies H(x, u, J) \leq H(x, u, J'), \quad \forall x, u
\]

where \( E(X) \) is the set of functions \( J : X \mapsto [-\infty, \infty] \)

- Mappings \( T_\mu \) and \( T \)

\[
(T_\mu J)(x) = H(x, \mu(x), J), \quad \forall x \in X, J \in R(X)
\]

\[
(TJ)(x) = \inf_{\mu} (T_\mu J)(x) = \inf_{u \in U(x)} H(x, u, J), \quad \forall x \in X, J \in R(X)
\]

Stochastic Optimal Control Mapping: A Special Case

\[
H(x, u, J) = E \{ g(x, u, w) + \alpha J(f(x, u, w)) \}
\]
Abstract Problem Formulation

**Abstract DP Problem**

- Given an initial function \( \bar{J} \in \mathbb{R}(X) \) and policy \( \mu \), define

\[
J_\mu(x) = \limsup_{N \to \infty} (T^N_\mu \bar{J})(x), \quad x \in X
\]

- Find \( J^*(x) = \inf_\mu J_\mu(x) \) and an optimal \( \mu \) attaining the infimum

**Notes**

- Theory revolves around fixed point properties of mappings \( T_\mu \) and \( T \):

\[
J_\mu = T_\mu J_\mu, \quad J^* = TJ^*
\]

These are generalized forms of **Bellman's equation**

- Algorithms are special cases of fixed point algorithms

- We restrict attention (initially) to issues involving only stationary policies
Examples With a Dynamic System $x_{k+1} = f(x_k, \mu(x_k), w_k)$

### Stochastic Optimal Control

\[
\bar{J}(x) \equiv 0, \quad (T_\mu J)(x) = E_w \{ g(x, \mu(x), w) + \alpha J(f(x, \mu(x), w)) \}
\]

\[
J_\mu(x_0) = \limsup_{N \to \infty} E_{w_0, w_1, \ldots} \left\{ \sum_{k=0}^{N} \alpha^k g(x_k, \mu(x_k), w_k) \right\}
\]

### Minimax - Sequential Games

\[
\bar{J}(x) \equiv 0, \quad (T_\mu J)(x) = \sup_{w \in W(x)} \{ g(x, u, w) + \alpha J(f(x, u, w)) \}
\]

\[
J_\mu(x_0) = \limsup_{N \to \infty} \sup_{w_0, w_1, \ldots} \sum_{k=0}^{N} \alpha^k g(x_k, \mu(x_k), w_k)
\]

### Multiplicative Cost Problems

\[
\bar{J}(x) \equiv 1, \quad (T_\mu J)(x) = E_w \{ g(x, \mu(x), w) J(f(x, \mu(x), w)) \}
\]

\[
J_\mu(x_0) = \limsup_{N \to \infty} E_{w_0, w_1, \ldots} \left\{ \prod_{k=0}^{N} g(x_k, \mu(x_k), w_k) \right\}
\]
Examples With a Markov Chain: Transition Probs. $p_{i_k,i_{k+1}}(u_k)$

**Finite-State Markov and Semi-Markov Decision Processes**

\[
\begin{align*}
\bar{J}(x) & \equiv 0, \\
(T_\mu J)(i) & = \sum_{i=1}^{n} p_{ij}(\mu(i)) (g(i, \mu(i), j) + \alpha_{ij}(\mu(i)) J(j)) \\
J_\mu(i_0) & = \lim_{N \to \infty} \sup E \left\{ \sum_{k=0}^{N} (\alpha_{i_0}(\mu(i_0)) \cdots a_{i_k i_{k+1}}(\mu(i_k))) g(i_k, \mu(i_k), i_{k+1}) \right\}
\end{align*}
\]

where $\alpha_{ij}(u)$ are state and control-dependent discount factors

**Risk-Sensitive Shortest Path: Exponential Cost with Termination State $t$**

\[
\begin{align*}
J_\mu(x_0) & = \lim_{N \to \infty} \sup \ E \left\{ e^{g(i_0, \mu(i_0), i_1)} + \cdots + e^{g(i_N, \mu(i_N), i_{N+1})} \right\} \\
\bar{J}(x) & \equiv 1, \\
(T_\mu J)(i) & = p_{it}(\mu(i)) e^{g(i, \mu(i), t)} + \sum_{i=1}^{n} p_{ij}(\mu(i)) e^{g(i, \mu(i), j)} J(j)
\end{align*}
\]
Models Classified According to Properties of $T_\mu$

**Contractive (C)**

All $T_\mu$ are contractions within the set of bounded functions $B(X)$, w.r.t. a common (weighted) sup-norm and contraction modulus (e.g., discounted problems).

**Monotone Increasing (I) and Monotone Decreasing (D)**

$\bar{J} \leq T_\mu \bar{J}$ (e.g., negative DP problems)

$\bar{J} \geq T_\mu \bar{J}$ (e.g., positive DP problems)

**Semicontractive (SC)**

$T_\mu$ has “contraction-like” properties for some $\mu$ - to be discussed (e.g., SSP problems)

**Semicontractive Nonnegative (SC⁺)**

Semicontractive, and in addition $\bar{J} \geq 0$ and

$$J \geq 0 \quad \Rightarrow \quad H(x, u, J) \geq 0, \quad \forall \ x, u$$

(e.g., affine monotonic, exponential/risk-sensitive problems)
Bellman’s Equation:

\[ J_\mu = T_\mu J_\mu \quad \text{and} \quad J^* = TJ^* \]

hold often (but not always) under our assumptions.

Bellman’s Equation: Cases (C), (I), and (D)

\[ J_\mu = T_\mu J_\mu \quad \text{and} \quad J^* = TJ^* \]

always hold.

Bellman’s Equation: Case (SC)

\[ J_\mu = T_\mu J_\mu \]

holds only for \( \mu \): “regular”

\( \hat{J} \), the “restricted optimal” cost function, solves Bellman’s Eq. under our assumptions. We may have \( J^* \neq \hat{J} \).
### Case (C)

$T$ is a contraction within $B(X)$ and $J^*$ is its unique fixed point.

### Cases (I), (D)

$T$ has multiple fixed points (some partial results hold).

### Case (SC)

$\hat{J}$ is the unique fixed point of $T$ within a subset of $J \in R(X)$ with “regular” behavior.

### Case (SC+)

$J^*$ is the unique positive (or nonnegative) fixed point of $T$. 
Cases (C), (I), and (SC - under one set of assumptions)

\[ \mu^* \text{ is optimal if and only if } T_{\mu^*} J^* = TJ^* \]

Case (SC - under another set of assumptions)

A “regular” \( \mu^* \) is optimal if and only if \( T_{\mu^*} J^* = TJ^* \)

Case (D)

\[ \mu^* \text{ is optimal if and only if } T_{\mu^*} J_{\mu^*} = TJ_{\mu^*} \]
Convergence of Value Iteration: $J_{k+1} = TJ_k$

**Case (C)**

$T^k J \to J^*$ for all $J \in B(X)$

**Case (D)**

$T^k \bar{J} \to J^*$

**Case (I)**

$T^k \bar{J} \to J^*$ under additional “compactness” conditions

**Case (SC)**

$T^k J \to \hat{J}$ and possibly $T^k J \to J^*$ for all $J \in R(X)$ within a set of “regular” behavior

**Case (SC+)**

$T^k J \to J^*$ for all $J > 0$ (or $J \geq 0$ under some conditions)
Policy Iteration: $T_{\mu^{k+1}}J_{\mu^k} = TJ_{\mu^k}$ (A Complicated Story)

Classical Form of Exact PI

- (C): Convergence starting with any $\mu$
- (SC): Convergence starting with a “regular” $\mu$ (not if “irregular” $\mu$ arise)
- (I), (D): Convergence fails

Optimistic/Modified PI (Combination of VI and PI)

- (C): Convergence starting with any $\mu$
- (SC): Convergence starting with any $\mu$ after a substantial modification in the policy evaluation step: Solving an “optimal stopping” problem instead of a linear equation
- (D): Convergence starting with initial condition $\bar{J}$
- (I): Convergence may fail (special conditions required)

Asynchronous Optimistic/Modified PI (Combination of VI and PI)

- (C): Fails in the standard form. Works after a substantial modification
- (SC): Works after a substantial modification
- (D), (I): Convergence may fail (special conditions required)
Approximate $J_\mu$ and $J^*$ within a subspace spanned by basis functions

- Aim for approximate versions of value iteration, and policy iteration
- Very large and complex problems has been addressed
- Simulation-based algorithms are common
- No mathematical model is necessary (a computer simulator of the controlled system is sufficient)
- Abstract DP applies when cost approximation is based on the aggregation method (then the aggregate DP model has the required monotonicity property)

Case (C)

- A wide variety of additional results thanks to the underlying contraction property
- Approximate value iteration and Q-learning
- Approximate policy iteration, pure and optimistic/modified

Cases (I), (D), (SC)

Hardly any results available. Some results for stochastic shortest path problems
Semicontractive Abstract Problem Formulation

- Abstract monotone mapping $H : X \times U \times E(X) \mapsto \mathbb{R}$

\[ J \leq J' \implies H(x, u, J) \leq H(x, u, J'), \quad \forall x, u \]

where $E(X)$ is the set of functions $J : X \mapsto [-\infty, \infty]$

- Mappings $T_{\mu}$ and $T$

\[ (T_{\mu}J)(x) = H(x, \mu(x), J), \quad \forall x \in X, J \in R(X) \]
\[ (TJ)(x) = \inf_{\mu} (T_{\mu}J)(x) = \inf_{u \in U(x)} H(x, u, J), \quad \forall x \in X, J \in R(X) \]

Abstract DP Problem

- Given an initial function $\bar{J} \in R(X)$ and policy $\mu$, define

\[ J_\mu(x) = \limsup_{N \to \infty} (T_{\mu}^N \bar{J})(x), \quad x \in X \]

- Find $J^*(x) = \inf_{\mu} J_\mu(x)$ and an optimal $\mu$ attaining the infimum
Key idea: We have a set of functions $S \subseteq E(X)$, which we view as the “domain of regularity”

Definition of S-Regular Policy

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

1. $J_\mu \in S$ and $J_\mu = T_\mu J_\mu$
2. $T^k_\mu J \to J_\mu$ for all $J \in S$

A policy that is not S-regular is called S-irregular.
Given a set $S \subset E(X)$

- Consider the \textbf{restricted optimization problem}: Minimize $J_\mu$ over $\mu$ in the set $\mathcal{M}_S$ of all $S$-regular policies.
- Let $J^*_S$ be the optimal cost function over $S$-regular policies only:
  \[
  J^*_S(x) = \inf_{\mu \in \mathcal{M}_S} J_\mu(x), \quad x \in X
  \]
- Since the set of $S$-regular policies is a subset of the set of all policies, $J^* \leq J^*_S$. 

- Consider the $J^*$ for $J^*$.
Given a set $S \subset E(X)$ consider

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

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) $T^k J \rightarrow J_S^*$ for every $J \in W_S$
- (Optimality condition) If $\mu^*$ is $S$-regular, $J_S^* \in S$, and $T_{\mu^*} J_S^* = T J_S^*$, then $\mu^*$ is $\mathcal{M}_S$-optimal. Conversely, if $\mu^*$ is $\mathcal{M}_S$-optimal, then $T_{\mu^*} J_S^* = T J_S^*$. 

Well-Behaved Region $W_S$
How do we Show that $J^*_S$ is a Fixed Point of $T$?

A PI-Based Approach

- The approach applies when $S$ is “well-behaved” with respect to PI: roughly, starting from an $S$-regular policy $\mu^0$, PI generates $S$-regular policies.
- The significance of $S$-regularity is that $\{J_{\mu^k}\}$ is monotonically nonincreasing,
  
  $$J_{\mu^k} = T_{\mu^k} J_{\mu^k} \geq T J_{\mu^k} = T_{\mu^{k+1}} J_{\mu^k} \geq J_{\mu^{k+1}}$$

  so it has a limit $J_\infty$.
- It is natural to expect that $J_\infty$ will be equal to $J^*_S$ and will be a fixed point of $T$.

We introduce weak and strong PI properties and obtain corresponding weaker and stronger results for $J^*_S$ to be a fixed point of $T$. 
We say that $S$ has the **weak PI property** if there exists a sequence of $S$-regular policies $\{\mu^k\}$ generated by PI.

**Assume:**
- The weak PI property
- A “continuity from above” property for $H$: For each sequence $\{J_m\} \subset E(X)$ with $J_m \downarrow J$ for some $J \in E(X)$, we have
  \[
  H(x, u, J) = \lim_{m \to \infty} H(x, u, J_m), \quad \forall x \in X, u \in U(x)
  \]

Then $J_S^*$ is the only fixed point of $T$ within $W_S$, and VI converges to $J_S^*$ starting from within $W_S$. 
The Strong PI Property

We say that $S$ has the **strong PI property** if the weak PI property holds, and PI generates exclusively $S$-regular policies, when started with an $S$-regular policy.

**Verifying the Strong PI Property for $S \subset R(X)$**

$S$ 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 $\lambda \in \mathbb{R}$.
- For every $J \in S$ and $S$-irregular policy $\mu$, 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

Assume the conditions of the preceding slide hold (so that the strong PI property also holds), and also that $J^*_S \in S$. Then:

- $J^*_S$ is the unique fixed point of $T$ within $S$
- We have $T^k J \to J^*_S$ for every $J$ in the well-behaved region $W_S$
- Every policy $\mu$ that satisfies $T\mu J^*_S = TJ^*_S$ is $\mathcal{M}_S$-optimal and there exists at least one such policy

Note the stronger conclusions:

- $J^*_S$ is the unique fixed point of $T$ within $S$ (not just from within $W_S$)
- An optimality condition and existence of an $\mathcal{M}_S$-optimal policy
The conditions for verifying the strong PI property hold:

- \( S \subseteq R(X) \)
- 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 \( \lambda \in \mathbb{R} \)
- For every \( J \in S \) and \( S \)-irregular policy \( \mu \), there exists a state \( x \in X \) such that

\[
\limsup_{k \to \infty} (T_{\mu}^k J)(x) = \infty
\]

and also:

- \( S \) contains \( \bar{J} \), and has the property that if \( J_1, J_2 \) are two functions in \( S \), then \( S \) contains all functions \( J \) with \( J_1 \leq J \leq J_2 \)
- The function \( J_S^* = \inf_{\mu \in \mathcal{M}_S} J_{\mu} \) belongs to \( S \)
- For each sequence \( \{J_m\} \subset S \) with \( J_m \uparrow J \) for some \( J \in S \),

\[
\lim_{m \to \infty} H(x, u, J_m) = H(x, u, J), \quad \forall x \in X, \ u \in U(x)
\]

- For each function \( J \in S \), there exists a function \( J' \in S \) such that \( J' \leq J \) and \( J' \leq TJ' \)
Proposition: Under the preceding assumption

- $J^*$ is the unique fixed point of $T$ within the set $S$
- We have $T^k J \rightarrow J^*$ for all $J \in S$
- $\mu$ is optimal if and only if $T\mu J^* = TJ^*$, and there exists an optimal $S$-regular $\mu$
- For any $J \in S$, if $J \leq TJ$ we have $J \leq J^*$, and if $J \geq TJ$ we have $J \geq J^*$
- If in addition for each $\{J_m\} \subset E(X)$ with $J_m \downarrow J$ for some $J \in E(X)$,
  \[ H(x, u, J) = \lim_{m \to \infty} H(x, u, J_m), \quad \forall x \in X, u \in U(x) \]
  then every sequence $\{\mu^k\}$ generated by the PI algorithm starting from an $S$-regular policy $\mu^0$ satisfies $J_{\mu^k} \downarrow J^*$