Semicontractive Dynamic Programming

Dimitri P. Bertsekas

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Lecture 5 of 5

September 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. Review of Abstract DP
2. Review of Semicontractive Analysis
Abstract DP Problem Formulation

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

Monotone Mappings

- Abstract monotone mapping \(H : X \times U \times E(X) \mapsto \mathbb{R}\)
  \[J \leq J' \implies H(x, u, J) \leq H(x, u, J'), \quad \forall x, u\]
- Mappings \(T_\mu\) and \(T\)
  \[(T_\mu J)(x) = H(x, \mu(x), J), \quad \forall x \in X, J \in E(X)\]
  \[(TJ)(x) = \inf_\mu (T_\mu J)(x) = \inf_{u \in U(x)} H(x, u, J), \quad \forall x \in X, J \in E(X)\]

Stochastic Optimal Control Mapping: A Special Case

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

We saw several other problems and mappings, e.g., exponential cost, minimax, etc.
Abstract DP Problem

- Given an initial function $\bar{J} \in E(X)$ 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

Results of Interest

- Bellman’s equation
  \[ J^* = TJ^* \]
  and its set of solutions. Usually $J^*$ is a solution.
- Conditions for optimality of a stationary policy $\mu$, usually $T_\mu J_\mu = TJ_\mu$.
- Algorithms and their convergence issues.

Semicontractive Models:

Some policies are “well-behaved” (have a regularity property), and others are not.
**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:

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

A policy that is not S-regular is called S-irregular.
Value Iteration (VI)
- Given an initial function $J_0$, generate $T^k J_0$, $k = 0, 1, \ldots$
- We hope and expect that $T^k J_0 \to J^*$ for all $J_0$, or for $J_0$ in some convenient subset of functions.
- There is a similar VI algorithm that aims to compute $J_\mu$ in the limit. It generates $T^k_\mu J_0$, $k = 0, 1, \ldots$
- Note the connection with $S$-regularity: essentially, $\mu$ is $S$-regular if VI is “well-behaved starting within $S$,” i.e., $T^k_\mu J_0 \to J_\mu$, for all $J_0 \in S$.

Policy Iteration (PI)
- $\{\mu^k\}$ is generated by a two-step iteration:
  - $J_{\mu^k} = T_{\mu^k} J_{\mu^k}$, \hspace{1cm} (policy evaluation)
  - and $T_{\mu^{k+1}} J_{\mu^k} = T J_{\mu^k}$, \hspace{1cm} (policy improvement)
- We aim to prove that $J_{\mu^k} \to J^*$, and perhaps $\mu^k \to \mu^*$, an optimal policy.
Given a set $S \subset E(X)$

- Consider the 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:

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

- $J^* \leq J^*_S$ with strict inequality possible.
- When $J^* \neq J^*_S$, we have seen that $J^*_S$ may be more “well-behaved” than $J^*$.
- Most of our analysis has focused on cases where $J^* = J^*_S$. 

S-Regular Restricted Problem
Our Main Assumption

Assume that $S$ consists of real-valued functions and:

- There exists at least one $S$-regular policy and $J_S^* = \inf_{\mu \in \mathcal{M}_S} J_{\mu}$ belongs to $S$.
- For every $J \in S$ and $S$-irregular policy $\mu$, there exists $x \in X$ such that

$$\limsup_{k \to \infty} (T^k_{\mu} J)(x) = \infty$$

- $S$ 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 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 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

- **(Bellman Eq.)** $J^* = TJ^*$. Moreover, $J^*$ is the unique fixed point of $T$ within $S$
- **(VI Convergence)** We have $T^kJ \to J^*$ for all $J \in S$
- **(Optimality Condition)** $\mu$ is optimal if and only if $T\mu J^* = TJ^*$, and there exists an optimal $S$-regular $\mu$
- **(PI Convergence)** 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 $J_{\mu^k} \downarrow J^*$
- **(Optimization-Based Solution of Bellman’s Eq.)** For any $J \in S$, if $J \leq TJ$ we have $J \leq J^*$, and if $J \geq TJ$ we have $J \geq J^*$ (this allows finding $J^*$ by linear programming for many types of problems with finite spaces)
Value Iteration Properties

- Under our main assumption, $T^k J \to J^*$ for all $J \in S$.
- Under weaker assumptions (centering on PI properties of $S$, cf. Lectures 2 and 3), $T^k J \to J_S^*$ for all $J$ such that $J_S^* \leq J \leq \tilde{J}$ for some $\tilde{J} \in S$.

Policy Iteration Properties (Assuming we Start with an $S$-Regular Policy)

- Under our main assumption, $J_{\mu^k} \to J^*$.
- Under weaker assumptions (a strong PI property of $S$, cf. Lectures 2 and 3), $J_{\mu^k} \to J_S^*$.
- Note a weakness: An initial $S$-regular policy is needed.

Optimization Approach

- Under our main assumption $J^*$ maximizes over $J$ the sum $\sum_{i \in X} J(i)$ subject to $J \leq TJ$.
- Under weaker assumptions, $J_S^*$ maximizes over $J$ the sum $\sum_{i \in X} J(i)$ subject to $J \leq TJ$. 
A Mixture of VI and PI

Start with some $J_0 \in E(X)$ such that $J_0 \geq TJ_0$, and generate a sequence $\{J_k, \mu^k\}$ according to

$$T_{\mu^k}J_k = TJ_k, \quad J_{k+1} = T_{\mu^k}^{m_k}J_k, \quad k = 0, 1, \ldots,$$

where $m_k$ is a positive integer for each $k$.

Convergence under the Main Assumption

- We have $J_k \downarrow J^*$.
- The sequence $\{\mu^k\}$ generated by the algorithm consists of $S$-regular policies.

Notes

- Generally tends to converge faster than both VI and PI.
- Still requires a $J_0$ such that $J_0 \geq TJ_0$.
- There are interesting asynchronous variations for which this is not a requirement. Moreover this algorithm can deal with irregular policies as well.
A Motivating Example

Stationary policy costs

\[ J_\mu(1) = b, \ J_\mu'(1) = 0 \]

\[ J_\nu(1) = b, \ J^*(1) = 0 \]

Bellman Eq:

\[ J(1) = (TJ)(1) = \min \{b, \ J(1)\} \]

Suppose that we add \( \delta > 0 \) to the two costs 0 and \( b \):

- The cost of the improper policy \( \mu' \) becomes \( \infty \).
- The cost of the proper policy \( \mu \) increases by \( \delta \).
- By letting \( \delta \downarrow 0 \), we obtain \( J_\nu^*(1) = b \).

This Motivates a Perturbation Approach

- For each policy \( \mu \) and \( \delta \geq 0 \), we consider the mappings

\[ (T_{\mu,\delta}J)(x) = H(x, \mu(x), J) + \delta, \quad x \in X, \quad T_{\delta}J = \inf_{\mu \in \mathcal{M}} T_{\mu,\delta}J. \]

- Solve the \( \delta \)-perturbed problem with a sequence \( \delta_k \downarrow 0 \).
We define the cost functions of policies $\mu \in \mathcal{M}$, and optimal cost function $J^*_\delta$ of the $\delta$-perturbed problem by

$$J_{\mu, \delta}(x) = \limsup_{k \to \infty} T_{\mu, \delta}^k \bar{J}, \quad J^*_\delta = \inf_{\mu \in \mathcal{M}} J_{\mu, \delta}. $$

\[\]

Proposition:

Given a set $S \subset E(X)$, assume that:

- For every $\delta > 0$, we have $J^*_\delta = T_\delta J^*_\delta$, and there exists an $S$-regular policy $\mu^*_\delta$ that is optimal for the $\delta$-perturbed problem, i.e., $J_{\mu^*_\delta, \delta} = J^*_\delta$.

- For every $S$-regular policy $\mu$, we have

$$J_{\mu, \delta} \leq J_{\mu} + w_{\mu}(\delta), \quad \forall \delta > 0,$$

where $w_{\mu}$ is a function such that $\lim_{\delta \downarrow 0} w_{\mu}(\delta) = 0$.

- $H$ has the property that for every sequence $\{J_m\} \subset S$ with $J_m \downarrow J$, we have

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

Then $\lim_{\delta \downarrow 0} J^*_\delta = J^*_S$, and $J^*_S$ is a fixed point of $T$ (which brings to bear a main result from Lectures 2 and 3).
Value Iteration

- $J_{k+1} = T_{\delta_k} J_k$, with $\delta_k \downarrow 0$.
- There is an asynchronous version of the algorithm

Policy Iteration for SSP Assuming that $J^*(i) > -\infty$ for all $i$

Let $\delta_k \downarrow 0$, and let $\mu^0$ be a proper policy. Given a proper policy $\mu^k$, and we generate $\mu^{k+1}$ according to

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

Then:

- We have $J_{\mu^k} \to J^*_S$.
- $\mu^k$ is an optimal policy for sufficiently large $k$ (this depends on the finiteness of the state and the control spaces).
Some Final Remarks

On Abstract DP
- Abstraction leads to an economical analysis and promotes a deeper understanding.
- Focuses on the fundamental issues.

Semicontractive Models: An Interesting Special Class of Abstract DP Models
- Include important classes of practical problems.
- Involves unusual/pathological behavior.
- Aims to discover simple assumptions that preclude the pathological behavior, and allow the use of reliable algorithms.