Semicontractive Dynamic Programming

Dimitri P. Bertsekas

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Lecture 4 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

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 $H : X \times U \times E(X) \mapsto \Re$

$$J \leq J' \qquad \Longrightarrow \qquad H(x,u,J) \leq H(x,u,J'), \qquad \forall\, x,u$$

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

- Mappings T_μ and T

$$(T_\mu J)(x) = H\big(x, \mu(x), J\big), \qquad \forall\, x \in X,\ J \in E(X)$$

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

Stochastic Optimal Control Mapping: A Special Case

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

Abstract Problem Formulation

Abstract DP Problem

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

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

- Find $J^*(x) = \inf_\mu J_\mu(x)$ and an optimal μ 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 μ, usually $T_\mu J_\mu = TJ_\mu$.
- Algorithms, such as value iteration (VI) and policy iteration (PI), and their convergence issues.

Semicontractive Models:

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

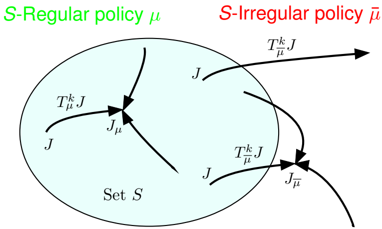
Our Approach for Semicontractive Problems

- 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

Key idea: We have a set of functions $S \subset E(X)$, which we view as the "domain of regularity"



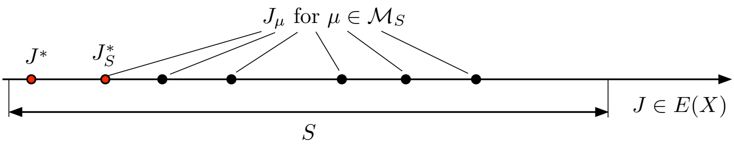
Definition of S-Regular Policy

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.

S-Regular Restricted Problem



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

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

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

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

$$J^* \leq J^*_S$$

A Principal Assumption that Guarantees “Good Behavior"

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 μ, there exists $x \in X$ such that
$$\limsup_{k \to \infty} \,(T_\mu^k J)(x) = \infty$$
- 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 $\big\{u \in U(x) \mid H(x, u, J) \leq \lambda\big\}$ is compact for every $J \in S$, $x \in X$, and $\lambda \in \Re$
- For each sequence $\{J_m\} \subset S$ with $J_m \uparrow J$ for some $J \in S$,
$$\lim_{m \to \infty} H(x, u, J_m) = H\,(x, u, J)\,, \qquad \forall\, x \in X,\ u \in U(x)$$
- For each function $J \in S$, there exists a function $J' \in S$ such that $J' \leq J$ and $J' \leq TJ'$

It is worth checking which parts of these assumptions are violated in the counterexamples of Lecture 1.

Main Result

Proposition: Under the preceding assumption

- (Bellman Eq.) $J^* = TJ^*$. Moreover, J^* is the unique fixed point of T within S
- (VI Convergence) We have $T^k J \to J^*$ for all $J \in S$
- (Optimality Condition) μ is optimal if and only if $T_\mu J^* = TJ^*$, and there exists an optimal S-regular μ
- (PI Convergence) If in addition for each $\{J_m\} \subset E(X)$ with $J_m \downarrow J$ for some $J \in E(X)$,

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

  then every sequence $\{\mu^k\}$ generated by the PI algorithm starting from an S-regular policy μ^0 satisfies $J_{\mu^k} \downarrow J^*$
- (Optimization-Based Solution of Bellman's Eq.) For any $J \in S$, if $J \le TJ$ we have $J \le J^*$, and if $J \ge TJ$ we have $J \ge J^*$

Note: Nearly as strong results as for contractive problems.

The Three Applications of this Lecture

Common Characteristics

- They all involve a finite number of states ($X = \{1, \ldots, n\}$), and a finite number of controls at each state (so the number of policies is finite).
- The set $R(X)$ of real-valued functions on X is identified with $\Re^n$.
- In all cases S is a subset of $\Re^n$.
- Usually there is a termination state.
- Because of this structure, the complicated assumption given earlier simplifies, and is nonrestrictive and intuitive.
- The results are almost as strong as for discounted problems.

The Applications in Summary

- Stochastic Shortest Path (SSP) Problems: Transition probs. $p_{ij}(u)$,

$$\bar{J}(i) \equiv 0, \qquad (T_\mu J)(i) = \sum_{j=1}^{n} p_{ij}\big(\mu(i)\big)\big(g(i, \mu(i), j) + J(j)\big)$$

- Affine Monotonic (AM) Problems:

$$\bar{J} \geq 0, \qquad T_\mu J = b_\mu + A_\mu J,$$

where $b_\mu \geq 0$, $A_\mu \geq 0$. A special case is SSP with exponential cost.

- Minimax Shortest Path (MSP) Problems: Disturbance has a nonprobabilistic set-membership description, $w \in W(i)$,

$$\bar{J}(i) \equiv 0, \qquad (T_\mu J)(i) = \max_{w \in W(i)} \big\{ g(i, \mu(i), w) + \alpha J(f(i, \mu(i), w)) \big\}$$

$$J_\mu(i_0) = \limsup_{N \to \infty} \max_{w_0, w_1, \ldots} \sum_{k=0}^{N} g\big(i_k, \mu(i_k), w_k\big)$$

We Specialize our Analysis to the Finite Spaces Context

S consists of real-valued functions

- For SSP and MSP, we use $S = \Re^n$.
- For AM, we use $S = \Re^n_+$, the nonnegative orthant.

Thanks to the finite spaces structure, and the choices of S, the complicated multipart assumption simplifies to the following:

- There exists at least one S-regular policy.
- Infinite cost condition: For all $J \in S$ and S-irregular μ, there exists i such that

$$\limsup_{k \to \infty} (T_\mu^k J)(i) = \infty$$

All other parts of the assumption are automatically satisfied.

A Common Approach for All Three Applications

Define the set S

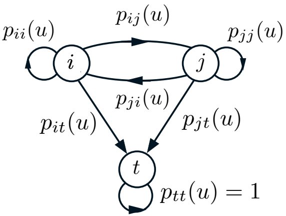
For SSP and MSP, we use $S = \Re^n$. For AM, we use $S = \Re^n_+$, the nonnegative orthant.

Characterize the S-Regular Policies

- For SSP, the S-regular μ are the proper policies (those that terminate with prob. 1).
- For AM, the S-regular μ are those for which T_μ is a contraction, i.e., all eigenvalues of A_μ are strictly within the unit circle.
- For MSP, the S-regular μ turn out to be those that guarantee termination regardless of the adversarial actions $w_0, w_1, \ldots$, but also some others.

Assume that there exists an S-regular policy and that each S-irregular policy has infinite cost.

Apply the theorem: J^* solves uniquely Bellman's Eq., VI, PI, and optimization approach work, etc.



A graph of n nodes plus the destination t

- At each node i we choose one of m probability distributions $p_{ij}(u)$, $u = 1, \ldots, m$, over the successor nodes j.
- Transition cost $g(i, u, j)$.
- Minimize total expected cost up to termination.

$$\bar{J}(i) \equiv 0, \qquad (T_\mu J)(i) = \sum_{i=1}^{n} p_{ij}\big(\mu(i)\big)\big(g(i, \mu(i), j) + J(j)\big)$$

Analysis for the SSP Problem

Proper Policies

- A policy μ is proper if it terminates from every initial state with probability 1.
- Equivalent definition: Starting at any node i, there exists a sequence of positive probability transitions under μ that starts at i and ends at t.
- Then $J_\mu(i)$ is the expected cost starting from i up to termination.

S-Regularity

- A policy is S-regular, where $S = \Re^n$, if and only if it is proper.
- We just verify the regularity definition ($T_\mu^k J \rightarrow J_\mu$ for all $J \in S$): We have that $T_\mu^k J$ does not depend on J for k large if and only if μ terminates.
- Assume there exists a proper policy.

Assume each Improper Policy has Infinite Cost Starting at Some Initial State

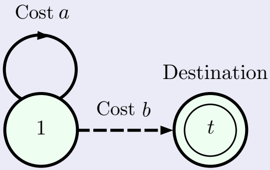
Check that "cycling has positive cost"; true if every transition has positive cost.

Apply the theorem

J^* solves uniquely Bellman's Eq., VI and PI converge to J^*, LP approach works, etc.

Back to the Pathological Examples

Deterministic shortest path problem



One proper policy (from 1 go to t), and one improper policy (self-cycle)

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

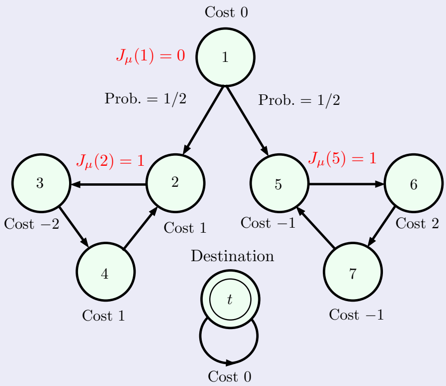
- Unique solution, $J^*(1) = b$ if $a > 0$ (assumptions satisfied)
- All $J(1) \leq b$ if $a = 0$ (assumptions violated)
- No real-valued solution if $a < 0$ (assumptions violated; consider changing S)

The assumption $a > 0$ corresponds to the classical conditions:

- There exists a path to the destination starting from every node.
- All cycles have positive length.

The SSP Problem where J^* does not Satisfy Bellman's Equation

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



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

Here the infinite cost condition is violated.

Affine Monotonic Problems

T_μ maps $J \in \Re^n_+$ into $T_\mu J \in \Re^n_+$ and is affine:

$$T_\mu J = b_\mu + A_\mu J,$$

where $b_\mu \geq 0$, $A_\mu \geq 0$. Also assume $\bar{J} \in \Re^n_+$ (but may have $\bar{J} \neq 0$)

Some special cases

- An SSP problem with nonnegative cost per transition. Corresponds to $\bar{J} = 0$ and

$$b_\mu(i) = g\big(i, \mu(i), j\big), \qquad A_\mu(i,j) = p_{ij}\big(\mu(i)\big)$$

- An SSP problem with exponential cost for the length of a path, so

$$J_\mu(i) = E\big\{\,\exp(\text{Length of path starting at } i \text{ up to reaching destination } t)\big\}$$

  Corresponds to the affine monotonic problem defined by

$$\bar{J}(i) \equiv 1, \qquad (T_\mu J)(i) = p_{it}\big(\mu(i)\big)e^{g\left(i,\mu(i),t\right)} + \sum_{i=1}^{n} p_{ij}\big(\mu(i)\big)e^{g\left(i,\mu(i),j\right)} J(j)$$

- Multiplicative cost function (contains the exponential cost SSP as a special case)

Characterization of S-Regular Policies; Here $S = \Re^n_+$

Cost Function of a Policy μ

By repeatedly applying the equation $T_\mu J = b_\mu + A_\mu J$, we have

$$T_\mu^N J = A_\mu^N J + \sum_{k=0}^{N-1} A_\mu^k b_\mu, \qquad \forall\, J \in E^+(X),\ N = 1, 2, \ldots,$$

$$J_\mu = \limsup_{N\to\infty} T_\mu^N \bar{J} = \limsup_{N\to\infty} A_\mu^N \bar{J} + \sum_{k=0}^{\infty} A_\mu^k b_\mu$$

Contractive policies: Those for which $\limsup_{N\to\infty} A_\mu^N J = 0$ for all $J \in \Re^n$ (equivalently A_μ has eigenvalues strictly within the unit circle).

Key fact is that μ is $R^+(X)$-regular if and only if T_μ is contractive. Justification:

$$J_\mu = \limsup_{N\to\infty} T_\mu^N J = \limsup_{N\to\infty} \sum_{k=0}^{N-1} A_\mu^k b_\mu, \qquad \forall\, \mu\text{: contractive},\ J \in \Re^n_+$$

Hence, if μ is contractive it is also $R^+(X)$-regular. The reverse can also be shown to be true.

Assumptions and Results for AM Problems

Assume that:

- There exists at least one contractive policy
- Each noncontractive policy has infinite cost for some initial state.

Then the standard results hold:

- Bellman's Eq. has J^* as its unique solution
- VI and PI converge to J^*
- Standard optimality conditions hold
- Solution by linear programming is possible

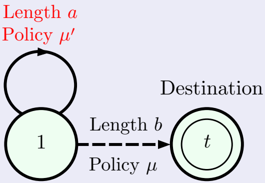
Some notes for the exponential cost SSP

- Every proper policy is contractive but the reverse is not true (consider a deterministic problem and a policy with a negative length cycle)
- In exponential cost SSP policies that include cycles with “negative cost" do not cause difficulties (but “zero cost cycles" may cause a problem)

Back to the Deterministic Shortest Path Problem

$J_{\mu'}(1) = \lim_{N\to\infty} \exp(aN)$

$(T_{\mu'}J)(1) = \exp(a)J(1)$



$J_{\mu}(1) = \exp(b)$

$(T_{\mu}J)(1) = \exp(b)$

Bellman's equation: $J(1) = \min\{\exp(b), \exp(a)J(1)\}$

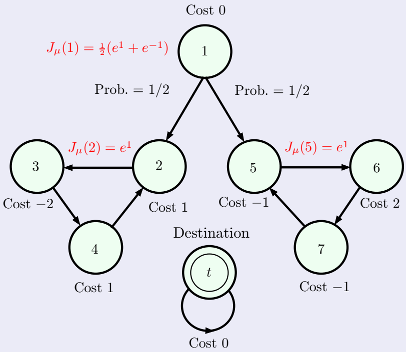
- If $a > 0$ (assumptions satisfied), $J^*(1) = \exp(b)$ solves uniquely Bellman's Eq., μ is optimal
- If $a < 0$ (assumptions satisfied, both policies are contractive, even though μ' is improper!), $J^*(1) = 0$ solves uniquely Bellman's Eq., μ' is optimal
- If $a = 0$ (assumptions violated), all $J(1)$ in the interval $0 \leq J(1) \leq \exp(b)$ solve Bellman's Eq., $J^*(1) = \min\{\exp(b), 1\}$

The assumption $a > 0$ corresponds to the classical conditions:

- There exists a path to the destination starting from every node.
- All cycles have positive length.

An Exponential Cost SSP Problem where J^* does not Satisfy Bellman's Equation

This is the exponential cost version of the earlier SSP counterexample, which involved zero length cycles.



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

Here the policy is noncontractive and hence $\Re^n_+$-irregular, while the infinite cost condition is violated.

Problem Formulation

- A graph with set of nodes $X = \{1, \ldots, n\}$ plus a destination t, and a set of directed arcs (i, j), where $i, j \in X \cup \{t\}$.
- At each node i we may choose a control u from a finite set $U(i)$.
- The destination t is absorbing and cost-free.
- At node i, a successor node j is selected by an antagonistic opponent from a given set $Y(i, u) \subset X \cup \{t\}$ and a cost $g(i, u, j)$ is incurred.
- Mappings:

$$H(i, u, J) = \max_{j \in Y(i,u)} \left[g(i, u, j) + \tilde{J}(j) \right], \qquad \forall\ x,\ u,\ J \in \Re^n,$$

where $\tilde{J}(j) = J(j)$ if $j \in X$ and $\tilde{J}(j) = 0$ if $j = t$. We have

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

- Let $\bar{J}$ be the zero function, so

$$J_\mu(i_0) = \limsup_{N \to \infty} \max_{w_0, w_1, \ldots} \sum_{k=0}^{N} g\big(i_k, \mu(i_k), w_k\big)$$

Cost Function and Other Properties of a Policy μ

- A possible path under μ starting at node $i_0 \in X$ is an arc sequence $p = \{(i_0, i_1), (i_1, i_2), \dots\}$, such that $i_{k+1} \in Y(i_k, \mu(i_k))$ for all $k \geq 0$. The set of all possible paths under μ starting at i_0 is denoted by $P(i_0, \mu)$.
- The length of a path $p \in P(i_0, \mu)$ is $\limsup_{N \to \infty} \sum_{k=0}^{N} g(i_k, \mu(i_k), i_{k+1})$.
- Similar definitions for the length of a portion of a path p, consisting of a finite number of consecutive arcs.
- For any μ and i, $(T_\mu^k \bar{J})(i)$ is the length of the longest path under μ that starts at i and consists of k arcs, and can be computed with a k-stage DP algorithm.
- Of special interest are cycles, i.e., paths of the form $\{(i_i, i_{i+1}), \dots, (i_{i+m}, i_i)\}$, and paths that terminate, i.e., have the form $p = \{(i_0, i_1), \dots, (i_m, t), (t, t), \dots\}$.

Proper Policies

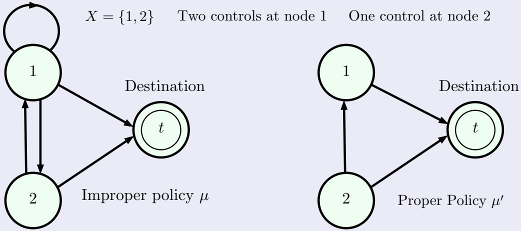
A policy μ is proper if for all i, all the paths in $P(i, \mu)$ contain no cycle and terminate.

S-Regular Policies ($S = \Re^n$)

It is easy to see that all proper policies are $\Re^n$-regular. The reverse is not true.

Characterization of S-Regular Policies ($S = \Re^n$)

The Characteristic Graph of a Policy μ: $\mathcal{A}_\mu = \cup_{i \in X}\big\{(i,j) \mid j \in Y\big(i,\mu(i)\big)\big\}$



- We say that $\mathcal{A}_\mu$ is destination-connected if for each $i \in X$ there exists a terminating path in $P(i,\mu)$.

Characterization of $\Re^n$-Regular Policies

- μ is $\Re^n$-regular if and only if $\mathcal{A}_\mu$ is destination-connected and all its cycles have negative length. (Note that a proper policy is $\Re^n$-regular.)
- μ is $\Re^n$-irregular if and only if it is improper, and either is destination-disconnected or $\mathcal{A}_\mu$ has a cycle with length ≥ 0. (Note that there exist improper policies that are $\Re^n$-regular.)

Assumptions and Results for MSP

Assume that:

- There exists at least one proper policy (implies that there exists an $\Re^n$-regular policy).
- For every improper policy μ, all cycles in the characteristic graph $\mathcal{A}_\mu$ have positive length (implies that every $\Re^n$-irregular policy has infinite cost for some initial state).

Then the standard results hold:

- Bellman's Eq. has J^* as its unique solution.
- VI, PI, converge to J^*.
- Standard optimality conditions hold, etc.

Some notes

- The positive cycle condition can be relaxed to nonnegativity, using a perturbation approach (add a $\delta > 0$ to each $g(i, u, j)$ and take $\delta \downarrow 0$; see the next lecture).
- There is a finitely terminating Dijkstra-like algorithm for MSP problems with nonnegative arc lengths (this is a consequence of the shortest path character of the problem, not its semicontractive character).