Abstract and Semicontractive DP: Stable Optimal Control

Dimitri P. Bertsekas

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

University of Connecticut

October 2017

Based on the Research Monograph

Abstract Dynamic Programming, 2nd Edition, Athena Scientific, 2017 (on-line)

Dynamic Programming

A UNIVERSAL METHODOLOGY FOR SEQUENTIAL DECISION MAKING

Applies to a very broad range of problems

- Deterministic <—> Stochastic
- Combinatorial optimization <—> Optimal control w/ infinite state and control spaces

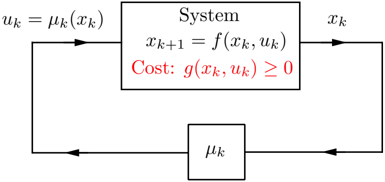
Approximate DP (Neurodynamic Programming, Reinforcement Learning)

- Allows the use of approximations
- Applies to very challenging/large scale problems
- Has proved itself in many fields, including some spectacular high profile successes

Standard Theory

- Analysis: Bellman's equation, conditions for optimality
- Algorithms: Value iteration, policy iteration, and approximate versions
- Abstract DP aims to unify the theory through mathematical abstraction
- Semicontractive DP an important special case - focus of new research

Outline



"Destination" t
(cost-free and absorbing)

An optimal control/regulation problem
or
An arbitrary space shortest path problem

- System: $x_{k+1} = f(x_k, u_k)$, $k = 0, 1$, where $x_k \in X$, $u_k \in U(x_k) \subset U$
- Policies: $\pi = \{\mu_0, \mu_1, \ldots\}$, $\mu_k(x) \in U(x)$, $\forall\, x$
- Cost $g(x, u) \geq 0$. Absorbing destination: $f(t, u) = t$, $g(t, u) = 0$, $\forall\, u \in U(t)$
- Minimize over policies $\pi = \{\mu_0, \mu_1, \ldots\}$

$$J_\pi(x_0) = \sum_{k=0}^{\infty} g\big(x_k, \mu_k(x_k)\big)$$

  where $\{x_k\}$ is the generated sequence using π and starting from x_0
- $J^*(x) = \inf_\pi J_\pi(x)$ is the optimal cost function

Classical example: Linear quadratic regulator problem; $t = 0$

$$x_{k+1} = Ax_k + Bu_k, \qquad g(x, u) = x'Qx + u'Ru$$

Optimality vs Stability - A Loose Connection

- Loose definition: A stable policy is one that drives $x_k \to t$, either asymptotically or in a finite number of steps
- Loose connection with optimization: The trajectories $\{x_k\}$ generated by an optimal policy satisfy $J^*(x_k) \downarrow 0$ (J^* acts like a Lyapunov function)
- Optimality does not imply stability (Kalman, 1960)

Classical DP for nonnegative cost problems (Blackwell, Strauch, 1960s)

- J^* solves Bellman's Eq.

$$J^*(x) = \inf_{u \in U(x)} \{g(x,u) + J^*\big(f(x,u)\big)\}, \quad x \in X, \qquad J^*(t) = 0,$$

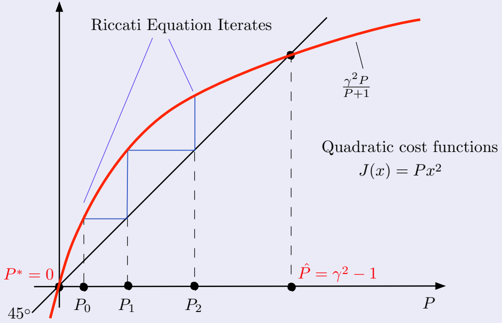
  and is the "smallest" (≥ 0) solution (but not unique)
- If $\mu^*(x)$ attains the min in Bellman's Eq., μ^* is optimal
- The value iteration (VI) algorithm

$$J_{k+1}(x) = \inf_{u \in U(x)} \{g(x,u) + J_k\big(f(x,u)\big)\}, \qquad x \in X,$$

  is erratic (converges to J^* under some conditions if started from $0 \leq J_0 \leq J^*$)
- The policy iteration (PI) algorithm is erratic

System: $x_{k+1} = \gamma x_k + u_k$ (unstable case, $\gamma > 1$). Cost: $g(x, u) = u^2$

- $J^*(x) \equiv 0$, optimal policy: $\mu^*(x) \equiv 0$ (which is not stable)
- Bellman Eq. $\rightarrow$ Riccati Eq. $P = \gamma^2 P/(P + 1)$ - $J^*(x) = P^* x^2$, $P^* = 0$ is a solution



- A second solution $\hat{P} = \gamma^2 - 1$: $\hat{J}(x) = \hat{P} x^2$
- $\hat{J}$ is the optimal cost over the stable policies
- VI and PI typically converge to $\hat{J}$ (not J^*!)
- Stabilization idea: Use $g(x, u) = u^2 + \delta x^2$. Then $J^*_\delta(x) = P^*_\delta x^2$ with $\lim_{\delta \downarrow 0} P^*_\delta = \hat{P}$

Summary of Analysis I: p-Stable Policies

Idea: Add a “small" perturbation to the cost function to promote stability

- Add to g a δ-multiple of a “forcing" function p with $p(x) > 0$ for $x \neq t$, $p(t) = 0$
- The resulting “perturbed" cost function of π is

$$J_{\pi,\delta}(x_0) = J_\pi(x_0) + \delta \sum_{k=0}^{\infty} p(x_k), \qquad \delta > 0$$

- Definition: A policy π is called p-stable if

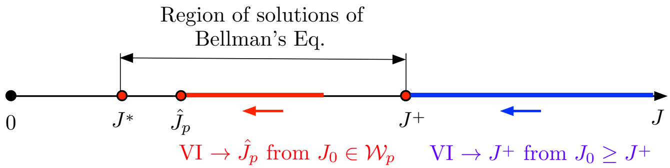
$$J_{\pi,\delta}(x_0) < \infty, \quad \forall\, x_0 \text{ with } J^*(x_0) < \infty \qquad \text{(this is independent of } \delta\text{)}$$

- The role of p:
  - Ensures that p-stable policies drive x_k to t (p-stable implies $p(x_k) \to 0$)
  - Differentiates stable policies by “speed of stability" (e.g., $p(x) = \|x\|$ vs $p(x) = \|x\|^2$)

The case $p(x) \equiv 1$ for $x \neq t$ is special

- Then the p-stable policies are the terminating policies (reach t in a finite number of steps for all x_0 with $J^*(x_0) < \infty$)
- The terminating policies are the “most stable" (they are p-stable for all p)

Summary of Analysis II: Restricted Optimality



J^*, $\hat{J}_p$, and J^+ are solutions of Bellman's Eq. with $J^* \leq \hat{J}_p \leq J^+$

- $\hat{J}_p(x)$: optimal cost J_π over the p-stable π, starting at x
- $J^+(x)$: optimal cost J_π over the terminating π, starting at x

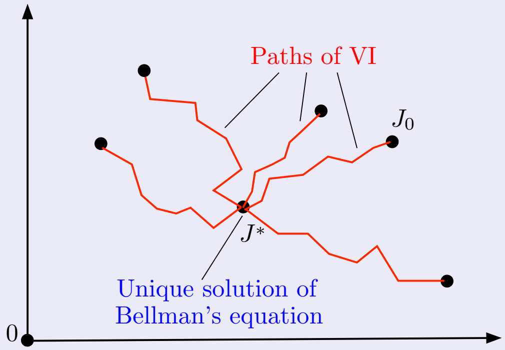
Why is $\hat{J}_p$ a solution of Bellman's Eq.?

- p-unstable π cannot be optimal in the δ-perturbed problem, so $\hat{J}_{p,\delta} \downarrow \hat{J}_p$ as $\delta \downarrow 0$
- Take limit as $\delta \downarrow 0$ in the (p,δ)-perturbed Bellman Eq. (which is satisfied by $\hat{J}_{p,\delta}$)

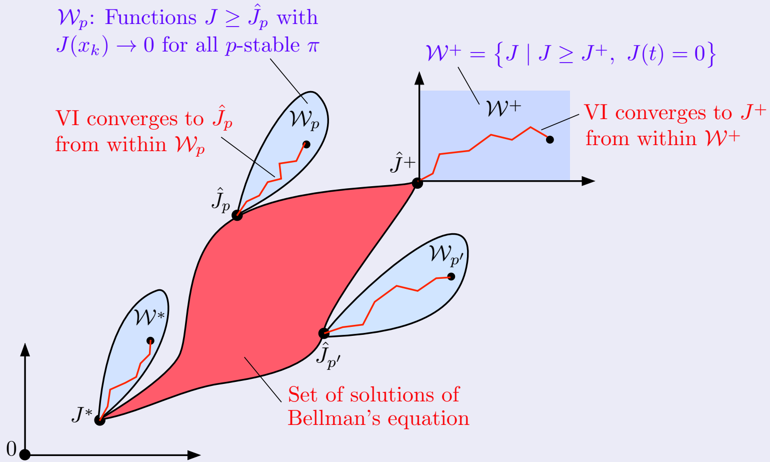
Favorable case is when $J^* = J^+$ (often holds). Then:

- J^* is the unique solution of Bellman's Eq.; optimal policy is p-stable
- VI and PI converge to J^* from above

Summary of Analysis III: Favorable Case $J^* = J^+$

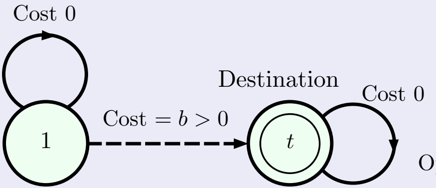


- J^* is the unique nonnegative solution of Bellman's Eq. [with $J^*(t) = 0$]
- VI converges to J^* from $J_0 \geq J^*$ (or from $J_0 \geq 0$ under mild conditions)
- Optimal policies are p-stable
- A "linear programming" approach works [J^* is the "largest" J satisfying $J(x) \leq g(x, u) + J\big(f(x, u)\big)$ for all (x, u)]



- Region of VI convergence to $\hat{J}_p$ is $\mathcal{W}_p$
- $\mathcal{W}_p$ can be viewed as a set of "Lyapounov functions" for the p-stable policies

Another Example: A Deterministic Shortest Path Problem



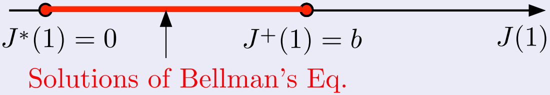
Bellman's equation

$J(1) = \min\{b,\ J(1)\},\ \ J(t) = 0$

Optimal cost $J^*(1) = 0$

Optimal cost over the stable policies $J^+(1) = b$

Set of solutions ≥ 0 of Bellman's Eq. with $J(t) = 0$



The VI algorithm

- It is attracted to J^+ if started with $J_0(1) \geq J^+(1)$

Stochastic Shortest Path (SSP) Problems

Bellman's Eq.: $J(x) = \inf_{u \in U(x)} \Big\{ g(x,u) + E\big\{J(f(x,u,w))\big\}\Big\}, \quad J(t) = 0$

Finite-state SSP (A long history - many applications)

- Analog of terminating policy is a proper policy: Leads to t with prob. 1 from all x
- J^+: Optimal cost over just the proper policies
- Case $J^* = J^+$ (Bertsekas and Tsitsiklis, 1991): If each improper policy has ∞ cost from some x, J^* solves uniquely Bellman's Eq.; VI converges to J^* from any $J \geq 0$
- Case $J^* \neq J^+$ (Bertsekas and Yu, 2016): J^* and J^+ are the smallest and largest solutions of Bellman's Eq.; VI converges to J^+ from any $J \geq J^+$

Infinite-State SSP with $g \geq 0$ and g: bounded (Bertsekas, 2017)

- Definition: π is a proper policy if π reaches t in bounded $E\{\text{Number of steps}\}$
- J^+: Optimal cost over just the proper policies
- J^* and J^+ are the smallest and largest solutions of Bellman's Eq. within the class of bounded functions
- VI converges to J^+ from any bounded $J \geq J^+$

Generalizing the Analysis: Abstract DP

Abstraction in mathematics (according to Wikipedia)

“Abstraction in mathematics is the process of extracting the underlying essence of a mathematical concept, removing any dependence on real world objects with which it might originally have been connected, and generalizing it so that it has wider applications or matching among other abstract descriptions of equivalent phenomena."

“The advantages of abstraction are:

- It reveals deep connections between different areas of mathematics.
- Known results in one area can suggest conjectures in a related area.
- Techniques and methods from one area can be applied to prove results in a related area."

ELIMINATE THE CLUTTER ... LET THE FUNDAMENTALS STAND OUT

What is Fundamental in DP? Answer: The Bellman Eq. Operator

Define a general model in terms of an abstract mapping $H(x, u, J)$

- Bellman's Eq. for optimal cost:

$$J(x) = \inf_{u \in U(x)} H(x, u, J)$$

- For the deterministic optimal control problem

$$H(x, u, J) = g(x, u) + J\big(f(x, u)\big)$$

- Another example: Discounted and undiscounted stochastic optimal control

$$H(x, u, J) = g(x, u) + \alpha E\big\{J(f(x, u, w))\big\}, \qquad \alpha \in (0, 1]$$

- Other examples: Minimax/games, semi-Markov, multiplicative/exponential cost, etc
- Key premise: *H is the "math signature" of the problem*
- Important structure of H: monotonicity (always true) and contraction (may be true)
- Top down development:

  Math Signature –> Analysis and Methods –> Special Cases

Abstract DP Problem

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

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

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

- Define for each admissible control function of state μ

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

and also define

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

Abstract Optimization Problem

- Introduce an initial function $\bar{J} \in E(X)$ and the cost function of a policy $\pi = \{\mu_0, \mu_1, \ldots\}$:

$$J_\pi(x) = \limsup_{N\to\infty} (T_{\mu_0} \cdots T_{\mu_N} \bar{J})(x), \qquad x \in X$$

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

Notes

- Deterministic optimal control interpretation: $(T_{\mu_0} \cdots T_{\mu_N} \bar{J})(x_0)$ is the cost of starting from x_0, using π for N stages, and incurring terminal cost $\bar{J}(x_N)$
- Theory revolves around fixed point properties of mappings T_μ and T:

$$J_\mu = T_\mu J_\mu, \qquad J^* = TJ^*$$

  These are generalized forms of Bellman's equation
- Algorithms are special cases of fixed point algorithms

Principal Types of Abstract Models

Contractive:

- Patterned after discounted optimal control w/ bounded cost per stage
- The DP mappings T_μ are weighted sup-norm contractions (Denardo 1967)

Monotone Increasing/Decreasing:

- Patterned after nonpositive and nonnegative cost DP problems
- No reliance on contraction properties, just monotonicity of T_μ (Bertsekas 1977, Bertsekas and Shreve 1978)

Semicontractive:

- Patterned after control problems with a goal state/destination
- Some policies μ are “well-behaved" (T_μ is contractive-like); others are not, but focus is on optimization over just the “well-behaved" policies
- Examples of “well-behaved" policies: Stable policies in det. optimal control; proper policies in SSP

The Line of Analysis of Semicontractive DP

- Introduce a class of well-behaved policies (formally called regular)
- Define a restricted optimization problem over just the regular policies
- Show that the restricted problem has nice theoretical and algorithmic properties
- Relate the restricted problem to the original
- Under reasonable conditions: Obtain interesting theoretical and algorithmic results
- Under favorable conditions: Obtain powerful analytical and algorithmic results (comparable to those for contractive models)

Regular Collections of Policy-State Pairs

Definition: For a set of functions $S \subset E(X)$ (the set of extended real-valued functions on X), we say that a collection $\mathcal{C}$ of policy-state pairs (π, x_0) is *S-regular* if

$$J_\pi(x) = \limsup_{N\to\infty} (T_{\mu_0} \cdots T_{\mu_N} J)(x), \qquad \forall\, (\pi, x_0) \in \mathcal{C},\ J \in S$$

Interpretation:

Changing the terminal cost function from $\bar{J}$ to any $J \in S$ does not matter in the definition of $J_\pi(x_0)$

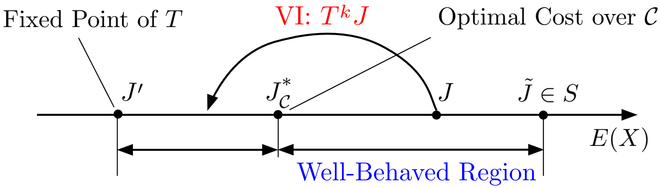
Optimal control example: Let $S = \{J \geq 0 \mid J(t) = 0\}$

The set of all (π, x) such that π is terminating starting from x is S-regular

Restricted optimal cost function with respect to $\mathcal{C}$

$$J^*_{\mathcal{C}}(x) = \inf_{\{\pi \mid (\pi,x)\in\mathcal{C}\}} J_\pi(x), \qquad x \in X$$

A Basic Theorem



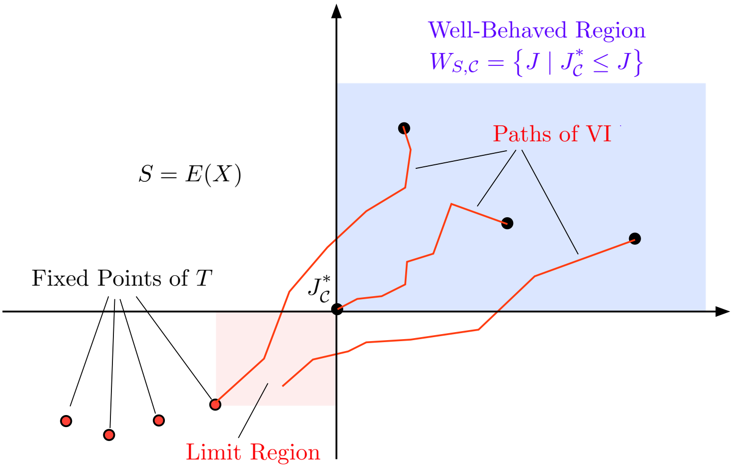
Well-behaved region

Let $\mathcal{C}$ be a collection of policy-state pairs (π, x) that is S-regular. The well-behaved region is the set

$$W_{S,\mathcal{C}} = \{J \mid J^*_{\mathcal{C}} \leq J \leq \tilde{J} \text{ for some } \tilde{J} \in S\}$$

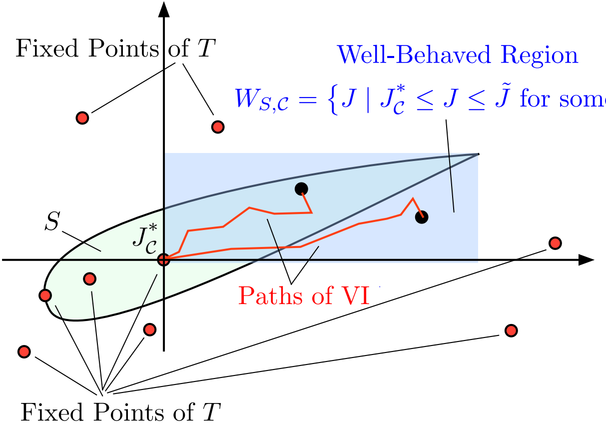
Key result: The limits of VI starting from $W_{S,\mathcal{C}}$ lie below $J^*_{\mathcal{C}}$ and above all fixed points of T

$$J' \leq \liminf_{k\to\infty} T^kJ \leq \limsup_{k\to\infty} T^kJ \leq J^*_{\mathcal{C}}, \qquad \forall\, J \in W_{S,\mathcal{C}} \text{ and fixed points } J' \text{ of } T$$



- VI behavior: Well-behaved region $\{J \mid J \geq J^*_{\mathcal{C}}\}$ –> Limit region $\{J \mid J \leq J^*_{\mathcal{C}}\}$
- All fixed points J' of T lie below $J^*_{\mathcal{C}}$

Visualization when $J^*_{\mathcal{C}}$ is a Fixed Point of T and $S \subset E(X)$



- If J' is a fixed point of T with $J' \le \tilde{J}$ for some $\tilde{J} \in S$, then $J' \le J^*_{\mathcal{C}}$
- If $W_{S,\mathcal{C}}$ is unbounded above [e.g., if $S = E(X)$], $J^*_{\mathcal{C}}$ is a maximal fixed point of T
- VI converges to $J^*_{\mathcal{C}}$ starting from any $J \in W_{S,\mathcal{C}}$

Application to Deterministic Optimal Control

Let

$$S = \{J \mid J \geq 0,\ J(0) = 0\}$$

Consider collection

$$\mathcal{C} = \big\{(\pi, x) \mid \pi \text{ terminates starting from } x\big\}$$

Then:

- $\mathcal{C}$ is S-regular (since the terminal cost function $\bar{J}$ does not matter for terminating policies)
- General theory yields:
  - J^* and $J^*_{\mathcal{C}} = J^+$ are the smallest and largest solution of Bellman's Eq.
  - VI converges to J^+ starting from $J \geq J^+$
  - Etc

Refinements relating to p-stability

Consider collection

$$\mathcal{C} = \big\{(\pi, x) \mid \pi \text{ is } p\text{-stable from } x\big\}$$

$\mathcal{C}$ is S-regular for S equal to the set of "Lyapounov functions" of the p-stable policies:

$$S = \big\{J \mid J(t) = 0,\ J(x_k) \to 0,\ \forall\, (\pi, x_0) \text{ s.t. } \pi \text{ is } p\text{-stable from } x_0\big\}$$

Similar Applications to Various Types of DP Problems

Abstract and semicontractive analyses apply

- To discounted and undiscounted stochastic optimal control

$$H(x,u,J) = E\{g(x,u,w) + \alpha J\big(f(x,u,w)\big)\}, \qquad \bar{J}(x) \equiv 0$$

- To minimax problems (also zero sum games); e.g.,

$$H(x,u,J) = \sup_{w \in W} \left\{g(x,u,w) + \alpha J\big(f(x,u,w)\big)\right\}, \qquad \bar{J}(x) \equiv 0$$

- To robust shortest path planning (minimax with a termination state)
- To multiplicative and exponential/risk-sensitive cost functions

$$H(x,u,J) = E\left\{g(x,u,w) J\big(f(x,u,w)\big)\right\}, \qquad \bar{J}(x) \equiv 1$$

or

$$H(x,u,J) = E\left\{e^{g(x,u,w)} J\big(f(x,u,w)\big)\right\}, \qquad \bar{J}(x) \equiv 1$$

- More ... see the references

Concluding Remarks

Highlights of results for optimal control

- Connection of stability and optimality through forcing functions, perturbed optimization, and p-stable policies
- Connection of solutions of Bellman's Eq., p-Lyapounov functions, and p-regions of convergence of VI
- VI and PI algorithms for computing the restricted optimum (over p-stable policies)

Highlights of abstract and semicontractive analysis

- Streamlining the theory through abstraction
- S-regularity is fundamental in semicontractive models
- Restricted optimization over the S-regular policy-state pairs
- Localization of the solutions of Bellman's equation
- Localization of the limits of VI and PI
- "Favorable" and "not so favorable" cases
- Broad range of applications

Thank you!