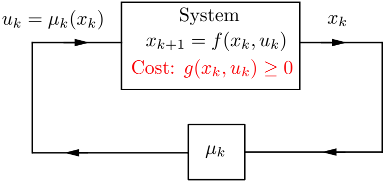
Stable Optimal Control and Semicontractive Dynamic Programming

Dimitri P. Bertsekas

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

May 2017



"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: $\pi = \{\mu_0, \mu_1, \dots\}$, $\mu_k(x) \in U(x),\ \forall\ x$
- Cost $g(x, u) \geq 0$. Absorbing destination: $f(t, u) = t,\ g(t, u) = 0,\ \forall\ u \in U(t)$
- Minimize over policies $\pi = \{\mu_0, \mu_1, \dots\}$

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

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.

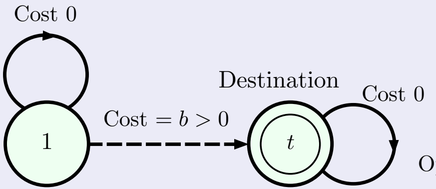
$$J^*(x) = \inf_{u \in U(x)} \{g(x,u) + J^*(f(x,u))\}, \quad x \in X, \qquad J^*(t) = 0,$$

  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

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

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

A Deterministic Shortest Path Problem



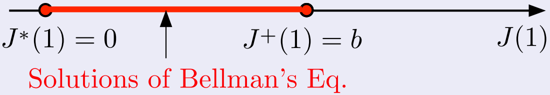
Bellman's equation

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

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

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

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

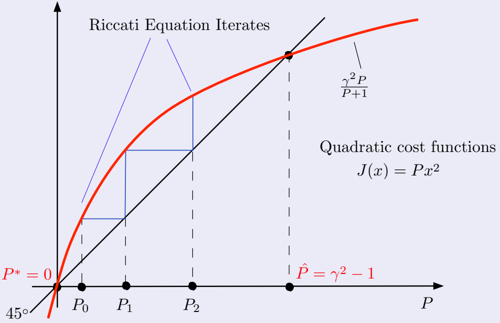


Algorithmic difficulties

- The VI algorithm is attracted to J^+ if started with $J_0(1) \geq J^+(1)$
- The PI algorithm is also 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

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

- A policy π is called *p*-stable if

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

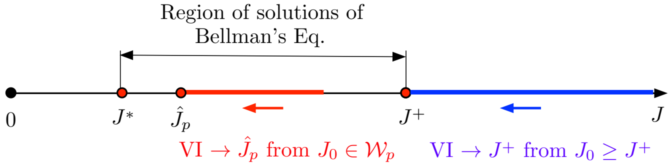
- The role of p:
  - Ensures that p-stable policies drive x_k to t (p-stable implies $p(x_k) \to 0$)
  - Differentiates stable policies by "speed of stability" (e.g., $p(x) = \|x\|$ vs $p(x) = \|x\|^2$)

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

- $\hat{J}_p(x)$: 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

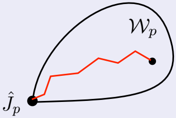


J^*, $\hat{J}_p$, and J^+ are solutions of Bellman's Eq. with $J^* \leq \hat{J}_p \leq J^+$

Favorable case is when $J^* = J^+$. Then:

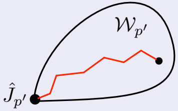
- J^* is the unique solution of Bellman's Eq.
- VI and PI converge to J^* from above

$\mathcal{W}_p$: Functions $J \geq \hat{J}_p$ with $J(x_k) \to 0$ for all p-stable π



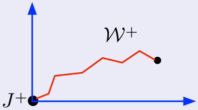
VI converges to $\hat{J}_p$ from within $\mathcal{W}_p$

$\mathcal{W}_{p'}$: Functions $J \geq \hat{J}_{p'}$ with $J(x_k) \to 0$ for all p'-stable π



VI converges to $\hat{J}_p'$ from within $\mathcal{W}_{p'}$

$\mathcal{W}^+ = \{J \mid J \geq J^+,\ J(t) = 0\}$



VI converges to J^+ from within $\mathcal{W}^+$

Case $J^* = J^+$: VI converges to J^* from $J_0 \geq J^*$ (or from $J_0 \geq 0$ under mild conditions)

References

Research Monograph

DPB, Abstract Dynamic Programming, Athena Scientific, 2013; updates on-line.

Subsequent Papers

- DPB, “Stable Optimal Control and Semicontractive Dynamic Programming," Report LIDS-P-3506, MIT, May 2017.
- DPB, “Proper Policies in Infinite-State Stochastic Shortest Path Problems," Report LIDS-P-3507, MIT, May 2017.
- DPB, “Value and Policy Iteration in Optimal Control and Adaptive Dynamic Programming," IEEE Trans. on Neural Networks and Learning Systems, 2015.
- DPB, “Regular Policies in Abstract Dynamic Programming," Report LIDS-P-3173, MIT, May 2015; to appear in SIAM J. Control and Opt.
- DPB, “Affine Monotonic and Risk-Sensitive Models in Dynamic Programming," Report LIDS-3204, MIT, June 2016.
- DPB, “Robust Shortest Path Planning and Semicontractive Dynamic Programming," Naval Research Logistics J., 2016.
- DPB and H. Yu, “Stochastic Shortest Path Problems Under Weak Conditions," Report LIDS-P-2909, MIT, January 2016.

Outline

Original Optimal Control Problem

- System: $x_{k+1} = f(x_k, u_k), \; k \geq 0$, where $x_k \in X, \; u_k \in U(x_k) \subset U$
- Cost per stage $g(x, u) \geq 0$
- Destination t: $f(t, u) = t, \; g(t, u) = 0, \; \forall \, u \in U(t)$ (absorbing, cost free)
- Policies: $\pi = \{\mu_0, \mu_1, \ldots\}$, $\mu_k(x) \in U(x), \; \forall \, x$
- Minimize over π

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

Composite Optimization with an Added Stability Objective

We introduce a forcing function p with

$$p(x) > 0, \quad \forall\, x \neq t, \qquad p(t) = 0$$

The δ-perturbed problem ($\delta > 0$) for a given p

- This is the same problem as the original, except the cost per stage is

$$g(x,u) + \delta p(x)$$

- Composite/perturbed objective

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

- J^*_δ: the optimal cost function of the δ-perturbed problem
- We have that J^*_δ solves the δ-perturbed Bellman Eq.:

$$J(x) = \inf_{u \in U(x)} \big\{ g(x,u) + \delta p(x) + J\big(f(x,u)\big)\big\}, \qquad x \in X$$

p-Stable Policies

- A policy π is called *p-stable* if

$$J_{\pi,\delta}(x) < \infty, \qquad \forall\, x \text{ with } J^*(x) < \infty$$

- $\hat{J}_p(x)$: optimal cost starting from x and using a p-stable policy

Line of analysis:

- p-unstable policies are "ignored" in the δ-perturbed problem
- J_δ^* is the optimal cost over stable policies plus $O(\delta)$ perturbation, so

$$\lim_{\delta \downarrow 0} J_\delta^* = \hat{J}_p$$

- J_δ^* can be used to approximate $\hat{J}_p$
- $\hat{J}_p$ solves the unperturbed Bellman Eq. (since J_δ^* solves the perturbed version)

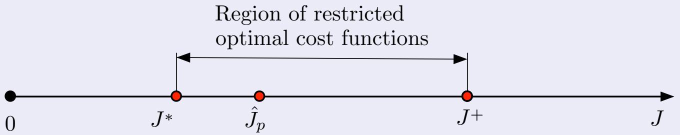
Terminating Policies

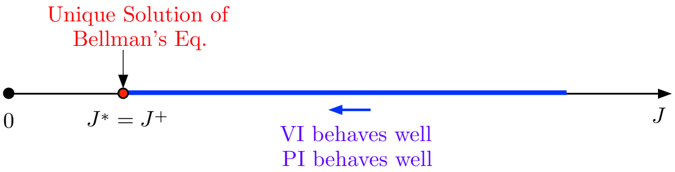
The forcing function $\bar{p}(x) = 1$ for all $x \neq t$ is special

- Then the $\bar{p}$-stable policies are the terminating policies (reach t in a finite number of steps for all relevant x_0)
- A terminating policy is p-stable with respect to every p

A hierarchy of policies and restricted optimal cost functions

- $J^*(x)$: optimal cost starting from x
- $\hat{J}_p(x)$: optimal cost starting from x and using a p-stable policy
- $J^+(x) = \hat{J}_{\bar{p}}(x)$: optimal cost starting from x and using a terminating policy





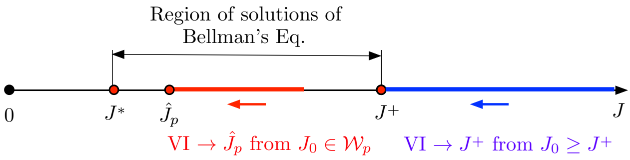
- True in the linear quadratic case under the classical controllability/observability conditions (even though there is no optimal terminating policy)
- Generally, for $J^* = J^+$ there must exist at least one terminating policy (a form of controllability)

Main Result (DPB 2015)

Let $\mathcal{J} = \{J \geq 0 \mid J(t) = 0\}$

- J^* is the unique solution of Bellman's Eq. within $\mathcal{J}$
- A sequence $\{J_k\}$ generated by VI starting from $J_0 \in \mathcal{J}$ and $J_0 \geq J^*$ converges to J^*. (Under a "compactness condition" converges to J^* starting from every $J_0 \in \mathcal{J}$.)
- A sequence $\{J_{\mu^k}\}$ generated by PI converges to J^*. (An optimistic version of PI also works.)

Result for the Unfavorable Case: $J^* \neq J^+$



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

Assumption: $\hat{J}_p(x) < \infty$ for all x with $J^*(x) < \infty$ (true if there exists a p-stable policy)

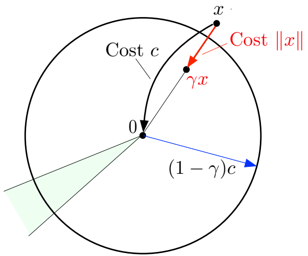
Main result (DPB 2017)

Let

$\mathcal{W}_p = \{J \geq \hat{J}_p \mid J(x_k) \to 0, \ \forall \ \{x_k\}$ generated from (π, x_0) w/ π: p-stable, $J^*(x_0) < \infty\}$

- $\mathcal{W}_p$ can be viewed as the set of Lyapounov functions for the p-stable policies
- $\hat{J}_p$ is the unique solution of Bellman's Eq. within $\mathcal{W}_p$
- J^+ is the unique solution of Bellman's Eq. within $\mathcal{W}^+ = \{J \geq J^+ \mid \ J(t) = 0\}$
- A sequence $\{J_k\}$ generated by VI starting from $J_0 \in \mathcal{W}_p$ converges to $\hat{J}_p$
- There are versions of PI that converge to $\hat{J}_p$

Optimal Stopping with State Space $\Re^n$, $t = 0$



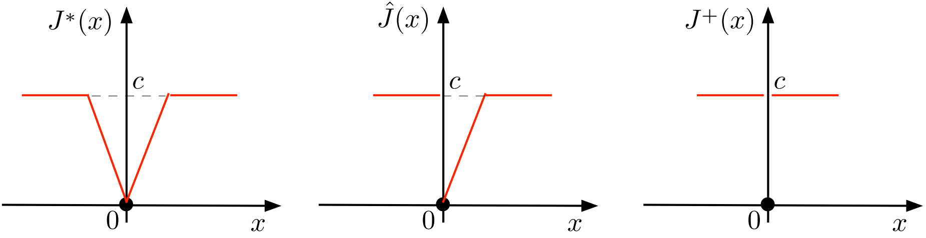
At state $x \neq 0$ we have two choices

- Stop (cost $c > 0$, move to 0)
- Continue [cost $\|x\|$, move to γx, where $\gamma \in (0, 1)$]
- Bellman's Eq.: $J(x) = \min\{c, \|x\| + J(\gamma x)\}, \quad x \neq 0$

All policies are stable! The solutions of Bellman's equation are:

- $J^*(x) = \min\left\{c, \frac{1}{1-\gamma}\|x\|\right\}$ and $J^+(x) = c$ for all $x \neq 0$
- An infinity of solutions in between, such as $J(x) = J^*(x)$ for x in some cone and $J(x) = J^+(x)$ for x in the complementary cone

Case $X = \Re$: Four Solutions of Bellman's Eq (J^*, J^+, two symmetric versions of $\hat{J}$)



Regions of Convergence of VI

- If $\lim_{x \to 0} J_0(x) = 0$ and $J_0 \geq J^*$, VI converges to J^* (also if $0 \leq J_0 \leq J^*$)
- If $J_0(0) = 0$, for all $x \neq 0$, and $J_0 \geq J^+$, VI converges to J^+
- If $\lim_{x \downarrow 0} J_0(x) = 0$ and $J_0 \geq \hat{J}$, VI converges to $\hat{J}$
- For dimensions $n \geq 2$, there is an infinity of regions of convergence of VI

Extension to Stochastic Shortest Path (SSP) Problems

Bellman's equation: $J(x) = \inf_{u \in U(x)} \Big\{ g(x,u) + E\big\{ J(f(x,u,w)) \big\} \Big\}$

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 proper policies (assumed real-valued)
- Result for case $J^* = J^+$ (BT, 1991): Assuming each improper policy has ∞ cost from some x, J^* solves uniquely Bellman's Eq. and VI works starting from any real-valued $J \geq 0$
- Result for case $J^* \neq J^+$ (BY, 2016): J^+ solves Bellman's Eq. and VI converges to J^+ starting from any real-valued $J \geq J^+$

Infinite-State SSP with $g \geq 0$

- π is a proper policy if J_π is bounded and π reaches t in bounded $E\{\text{No of steps}\}$ (over the initial x). Optimal cost over proper policies: J^+ (assumed bounded)
- Main result: J^+ solves Bellman's Eq. and VI converges to J^+ starting from any bounded $J \geq J^+$

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 of this lecture

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

- 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, semi-Markov, exponential risk-sensitive 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

Semicontractive Abstract DP Models

- Some policies are “well-behaved" and some are not
- Example of “well-behaved" policy: A μ whose $H\big(x, \mu(x), J\big)$ is a contraction (in J), e.g., a “stable" policy (or “proper" in the context of SSP)
- Generally, “unusual" behaviors are due to policies that are not “well-behaved"

The Line of Analysis of Semicontractive DP

- Introduce a class of well-behaved policies (formally called regular)
- 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 interesting theoretical and algorithmic results
- Under favorable conditions: Obtain powerful analytical and algorithmic results (comparable to those for contractive models)

Concluding Remarks

Highlights of results

- Connection of stability and optimization 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)

Outstanding Issues and Extensions

- How do we compute an optimal p-stable policy for a continuous-state problem (in practice, using discretization and approximation)?
- How do we check the existence of a p-stable policy (finiteness of $\hat{J}_p$)?
- Extensions to problems with both positive and negative costs per stage? If $J^* \neq J^+$, then J^* may not satisfy Bellman's Eq. for finite-state stochastic problems (J^+ does).

Thank you!