Abstract Dynamic Programming

Dimitri P. Bertsekas

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Overview of the Research Monograph
“Abstract Dynamic Programming"
Athena Scientific, 2013

Main Objective

- Unification of the core theory and algorithms of total cost sequential decision problems
- Simultaneous treatment of a variety of problems: MDP, sequential games, sequential minimax, multiplicative cost, risk-sensitive, etc

Methodology

- Define a problem by its “mathematical signature": the mapping defining the optimality equation
- Structure of this mapping (contraction, monotonicity, etc) determines the analytical and algorithmic theory of the problem
- Fixed point theory: An important connection

Three Main Classes of Total Cost DP Problems

Discounted:

- Discount factor < 1 and bounded cost per stage
- Dates to 50s (Bellman, Shapley)
- Nicest results

Undiscounted (Positive and Negative DP):

- N-step horizon costs are going $\downarrow$ or $\uparrow$ with N
- Dates to 60s (Blackwell, Strauch)
- Not nearly as powerful results compared with the discounted case

Stochastic Shortest Path (SSP):

- Also known as first passage or transient programming
- Aim is to reach a termination state at min expected cost
- Dates to 60s (Eaton-Zadeh, Derman, Pallu de la Barriere)
- Results are almost as strong as for the discounted case (under appropriate conditions)

Corresponding Abstract Models

Contractive:

- Patterned after discounted
- The DP mapping is a sup-norm contraction (Denardo 1967)

Monotone Increasing/Decreasing:

- Patterned after positive and negative DP
- No reliance on contraction properties, just monotonicity (Bertsekas 1977)

Semicontractive:

- Patterned after stochastic shortest path
- Some policies are “regular"/contractive; others are not, but assumptions are imposed so there exist optimal “regular" policies
- New research, inspired by SSP, where “regular" policies are the “proper" ones (the ones that terminate w.p.1)

Outline

- 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 $J : X \mapsto [-\infty, \infty]$

- Mappings T_μ and T

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

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

Stochastic Optimal Control - MDP example:

$(TJ)(x) = \inf_{u \in U(x)} E\big\{g(x, u, w) + \alpha J\big(f(x, u, w)\big)\big\}$

Abstract Problem Formulation

Abstract Optimization 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

Notes

- 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
- We restrict attention (initially) to issues involving only stationary policies

Examples With a Dynamic System $x_{k+1} = f\big(x_k, \mu(x_k), w_k\big)$

Stochastic Optimal Control

$$\bar{J}(x) \equiv 0, \qquad (T_\mu J)(x) = E_w\big\{g(x, \mu(x), w) + \alpha J\big(f(x, \mu(x), w)\big)\big\}$$

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

Minimax - Sequential Games

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

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

Multiplicative Cost Problems

$$\bar{J}(x) \equiv 1, \qquad (T_\mu J)(x) = E_w\big\{g(x, \mu(x), w) J\big(f(x, \mu(x), w)\big)\big\}$$

$$J_\mu(x_0) = \lim_{N\to\infty} E_{w_0, w_1, \ldots} \left\{ \prod_{k=0}^{N} g\big(x_k, \mu(x_k), w_k\big) \right\}$$

Finite-State Markov and Semi-Markov Decision Processes

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

$$J_\mu(i_0) = \limsup_{N\to\infty} E\left\{\sum_{k=0}^{N}\big(\alpha_{i_0}\big(\mu(i_0)\big)\cdots a_{i_k i_{k+1}}\big(\mu(i_k)\big)\big)\, g\big(i_k,\mu(i_k),i_{k+1}\big)\right\}$$

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

Undiscounted Exponential Cost

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

$$J_\mu(x_0) = \limsup_{N\to\infty} E\left\{e^{h\left(i_0,\mu(i_0),i_1\right)}\cdots e^{h\left(i_N,\mu(i_N),i_{N+1}\right)}\right\}$$

Models

Contractive (C)

All T_μ are contractions within 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_μ has "contraction-like" properties for some μ - to be discussed (e.g., SSP problems)

Semicontractive Nonnegative (SC^+)

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

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

(e.g., affine monotonic, exponential/risk-sensitive problems)

Bellman's Equation

Optimality/Bellman's Equation

$J^* = TJ^*$ always holds under our assumptions

Bellman's Equation for Policies: Cases (C), (I), and (D)

$J_\mu = T_\mu J_\mu$ always holds

Bellman's Equation for Policies: Case (SC)

$J_\mu = T_\mu J_\mu$ holds only for μ: "regular"

J_μ may take ∞ values for "irregular" μ

Uniqueness of Solution of Bellman's Equations

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)

J^* is the unique fixed point of T within a subset of $J \in R(X)$ with “regular" behavior

Optimality Conditions (A Complicated Story)

Cases (C), (I), and (SC - under one set of assumptions)

μ^* is optimal if and only if $T_{\mu^*} J^* = TJ^*$

Case (SC - under another set of assumptions)

A “regular" μ^* is optimal if and only if $T_{\mu^*} J^* = TJ^*$

Case (D)

μ^* is optimal if and only if $T_{\mu^*} J_{\mu^*} = TJ_{\mu^*}$

Asynchronous 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)

$T^k J \to J^*$ for all $J \in R(X)$ within a set of “regular" behavior

Policy Iteration: $T_{\mu^{k+1}} J_{\mu^k} = T J_{\mu^k}$ (A Complicated Story)

Classical Form of Exact PI

- (C): Convergence starting with any μ
- (SC): Convergence starting with a “regular" μ (not if “irregular" μ arise)
- (I), (D): Convergence fails

Optimistic/Modified PI (Combination of VI and PI)

- (C): Convergence starting with any μ
- (SC): Convergence starting with any μ after a major 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 major modification
- (SC): Works after a major modification
- (D), (I): Convergence may fail (special conditions required)

Results Overview: Approximate DP

Approximate J_μ and J^* within a subspace spanned by basis functions

- Aim for approximate versions of value iteration, policy iteration, and linear programming
- Simulation-based algorithms are common
- No mathematical model is necessary (a computer simulator of the controller system is sufficient)
- Very large and complex problems has been addressed

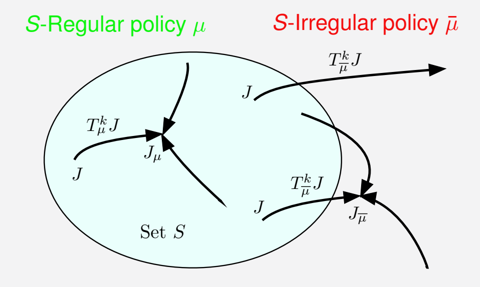
Case (C)

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

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

Hardly any results available

Key idea: Introduce a “domain of regularity," $S \subset E(X)$



Definition: A policy μ is S-regular if

- $J_\mu \in S$ and is the only fixed point of T_μ within S
- Starting function $\bar{J}$ does not affect J_μ, i.e.

$$T_\mu^k J \to J_\mu \qquad \forall\, J \in S$$

Typical Assumptions in Semicontractive Models

1st Set of Assumptions (Plus Additional Technicalities)

- There exists an S-regular policy and irregular policies are “bad": For each irregular μ and $J \in S$, there is at least one $x \in X$ such that

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

2nd Set of Assumptions (Plus Additional Technicalities)

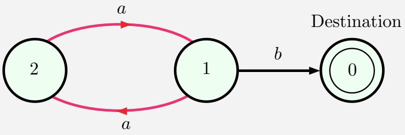
- There exists an *optimal* S-regular policy

Perturbation-Type Assumptions (Plus Additional Technicalities)

- There exists an *optimal* S-regular policy μ^*
- If H is perturbed by an additive $\delta > 0$, each S-regular policy is also δ-S-regular (i.e., regular for the δ-perturbed problem), and every δ-S-irregular policy μ is “bad", i.e., there is at least one $x \in X$ such that

$$\limsup_{k \to \infty} \big(T_{\mu,\delta}^k J_{\mu^*,\delta}\big)(x) = \infty$$

Semicontractive Example: Shortest Paths with Exponential Cost



Bellman equation

$$J(1) = \min\left\{\exp(b),\ \exp(a)J(2)\right\}$$

$$J(2) = \exp(a)J(1)$$

Two policies: $\bar{J} \equiv 1$; $S = \{J \mid J \geq 0\}$ or $S = \{J \mid J > 0\}$ or $S = \{J \mid J \geq \bar{J}\}$

- Noncyclic μ: $2 \to 1 \to 0$ (S-regular except when $S = \{J \mid J \geq \bar{J}\}$ and $b < 0$)

$$(T_\mu J)(1) = \exp(b), \qquad (T_\mu J)(2) = \exp(a)J(1)$$

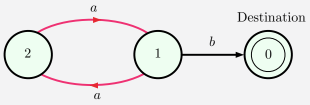
$$J_\mu(1) = \exp(b), \qquad J_\mu(2) = \exp(a+b)$$

- Cyclic $\bar{\mu}$: $2 \to 1 \to 2$ (S-irregular except when $S = \{J \mid J \geq 0\}$ and $a < 0$)

$$(T_{\bar{\mu}} J)(1) = \exp(a)J(2), \qquad (T_{\bar{\mu}} J)(2) = \exp(a)J(1)$$

$$J_{\bar{\mu}}(1) = J_{\bar{\mu}}(2) = \lim_{k\to\infty} \big(\exp(a)\big)^k$$

Five Special Cases (Each Covered by a Different Theorem!)



$a > 0$: $J^*(1) = \exp(b)$, $J^*(2) = \exp(a+b)$, is the unique fixed point w/ $J > 0$ (1st set of assumptions applies with $S = \{J \mid J > 0\}$)

- Set of fixed points of T is $\{J \mid J(1) = J(2) \leq 0\}$

$a = 0,\ b > 0$: $J^*(1) = J^*(2) = 1$ (perturbation assumptions apply)

- Set of fixed points of T is $\{J \mid J(1) = J(2) \leq \exp(b)\}$

$a = 0,\ b = 0$: $J^*(1) = J^*(2) = 1$ (2nd set of assumptions applies with $S = \{J \mid J \geq \bar{J}\}$)

- Set of fixed points of T is $\{J \mid J(1) = J(2) \leq 1\}$

$a = 0,\ b < 0$: $J^*(1) = J^*(2) = \exp(b)$ (perturbation assumptions apply)

- Set of fixed points of T is $\{J \mid J(1) = J(2) \leq \exp(b)\}$

$a < 0$: $J^*(1) = J^*(2) = 0$ is the unique fixed point of T (contractive case)

An Example: Affine Monotonic/Risk-Sensitive Models

T_μ is linear of the form $T_\mu J = A_\mu J + b_\mu$ with $b_\mu \geq 0$ and

$$J \geq 0 \qquad \Longrightarrow \qquad A_\mu J \geq 0$$

$S = \{J \mid 0 \leq J\}$ or $S = \{J \mid 0 < J\}$ or S: J bounded above and away from 0

Special case I: Negative DP model, $\bar{J}(x) \equiv 0$, A_μ: Transition prob. matrix

Special case II: Multiplicative model w/ termination state 0, $\bar{J}(x) \equiv 1$

$$H(x,u,J) = p_{x0}(u)g(x,u,0) + \sum_{y \in X} p_{xy}(u)g(x,u,y)J(y)$$

$$A_\mu(x,y) = p_{xy}\big(\mu(x)\big)g\big(x,\mu(x),y\big), \qquad b_\mu(x) = p_{x0}(u)g(x,u,0)$$

Special case III: Exponential cost w/ termination state 0, $\bar{J}(x) \equiv 1$

$$A_\mu(x,y) = p_{xy}\big(\mu(x)\big)\exp\big(h(x,\mu(x),y)\big), \quad b_\mu(x) = p_{x0}\big(\mu(x)\big)\exp\big(h(x,\mu(x),0)\big)$$

SC Assumptions Translated to Affine Monotonic

μ is S-regular if and only if

$$\lim_{k\to\infty}(A_\mu^k J)(x) = 0, \qquad \sum_{m=0}^{\infty}(A_\mu^m b_\mu)(x) < \infty, \qquad \forall\, x \in X,\ J \in S$$

The 1st Set of Assumptions

- There exists an S-regular policy; also $\inf_{\mu: S-regular} J_\mu \in S$
- If μ: S-irregular, there is at least one $x \in X$ such that

$$\sum_{m=0}^{\infty}(A_\mu^m b_\mu)(x) = \infty$$

- Compactness and continuity conditions hold

Notes:

- Value and (modified) policy iteration algorithms are valid
- State and control spaces need not be finite
- Related (but different) results are possible under alternative conditions

Concluding Remarks

- Abstract DP is based on the connections of DP with fixed point theory
- Aims at unification and insight through abstraction
- Semicontractive models fill a conspicuous gap in the theory from the 60s-70s
- Affine monotonic is a natural and useful model
- Abstract DP models with approximations require more research
- Abstract DP models with restrictions, such as measurability of policies, require more research

Thank you!