Semilinear Dynamic Programming: Analysis, Algorithms, and Certainty Equivalence Properties

Yuchao Li and Dimitri P. Bertsekas

Based on the paper
"Semilinear Dynamic Programming: Analysis, Algorithms, and Certainty Equivalence Properties", by Y. Li and D.P. Bertsekas, arXiv:2501.04668 , Jan. 2025

Outline

An Investment Problem for Farming from 70s [Ber76]

- A farmer annually produces x_k units of a certain crop, stores $(1-u_k)x_k$ and invests $u_k x_k$ for improving production of next year, where $0 \leq u_k \leq 1$
- The production of next year x_{k+1} is given by

$$x_{k+1} = x_k + w_k x_k u_k,$$

  where w_k is an independent random variable with $E\{w_k\} = \bar{w}$ for all k
- The problem is to find the optimal investment policy so that it maximizes the total expected product stored over N years:

$$\underset{\substack{w_k \\ k=0,1,\dots,N-1}}{E}\left\{x_N + \sum_{k=0}^{N-1}(1-u_k)x_k\right\}$$

Key characteristics of the problem

- The system and the cost have bilinear structure: $w_k x_k u_k$ in the system equation and $u_k x_k$ in the cost
- Nonegative states and bounded control $0 \leq u_k \leq 1$

Exact Solution via Dynamic Programming (DP)

- Start by setting $J_N^*(x_N) = c_N^* x_N$, where $c_N^* = 1$
- Suppose $J_{k+1}^*(x_{k+1}) = c_{k+1}^* x_{k+1}$. Going backwards, for $k = N-1, N-2, \ldots, 0$, let

$$
\begin{aligned}
J_k^*(x_k) = & \max_{0 \le u_k \le 1} E\big\{(1-u_k)x_k + c_{k+1}^*(x_k + w_k x_k u_k)\big\} \\
= & (1 + c_{k+1}^*)x_k + \max_{0 \le u_k \le 1} \big(c_{k+1}^* E\{w_k\} - 1\big)x_k u_k \\
= & (1 + c_{k+1}^*)x_k + \max_{0 \le u_k \le 1} (c_{k+1}^* \bar{w} - 1)x_k u_k \\
= & (1 + c_{k+1}^*)x_k + (c_{k+1}^* \bar{w} - 1)u_k^* x_k
\end{aligned}
$$

- We have $J_k^*(x) = c_k^* x_k$, where $c_k^* = c_{k+1}^*(1 + \bar{w})$ and $u_k^* = 1$ if $c_{k+1}^* \bar{w} > 1$, and $c_k^* = 1 + c_{k+1}^*$ and $u_k^* = 0$ otherwise
- Solution property: Linear functions are closed under the DP calculation and optimal policies are of special type.

Focus of this talk

- Can we obtain similarly structured pairs of cost functions and policies for problems with an inifite horizon?
- 1) Deterministic problems; 2) Stochastic problems; 3) Markov jump problems

Deterministic Optimal Control Problems with Nonnegative Costs

- State space $X \subset \Re^n$, control space U, control constraint set at x given by $U(x)$
- Optimal control problem: for a given $x_0 \in X$, solve

$$\min_{\{u_k\}_{k=0}^{\infty}} \sum_{k=0}^{\infty} \alpha^k g(x_k, u_k) \;\; \text{s.t.} \;\; x_{k+1} = f(x_k, u_k),\; u_k \in U(x_k), \quad k = 0, 1, \ldots,$$

where $f : X \times U \mapsto \Re^n$ and $g : X \times U \mapsto \Re$ are the system function and cost per stage, respectively, and $\alpha \in (0, 1]$ is a given scalar.

- Nonnegative cost condition $g(x, u) \geq 0$ for all $x \in X, u \in U(x)$.

Existing results for the problem

- The optimal cost function $J^*(x)$ satisfies Bellman's equation

$$J^*(x) = \min_{u \in U(x)} \left\{ g(x, u) + \alpha J^*\big(f(x, u)\big) \right\}, \qquad \text{for all } x.$$

- Given a policy $\mu : X \mapsto U$ so that $\mu(x) \in U(x)$ for all x, its cost function J_μ satisfies

$$J_\mu(x) = g\big(x, \mu(x)\big) + \alpha J_\mu\Big(f\big(x, \mu(x)\big)\Big), \qquad \text{for all } x.$$

- Optimal policy μ^* attains the min. in Bellman's equation and satisfies $J^* = J_{\mu^*}$
- However, computation of J^* and μ^* is intractable

Semilinear Structure for Deterministic Problems

- **State remains in the positive orthant**: $X \subset \Re^n_+$ and $f(x,u) \in X$ for all $x \in X$ and $u \in U(x)$
- **Favorable class of policies**: a set of policies $\widehat{\mathcal{M}}$ such that for every $\mu \in \widehat{\mathcal{M}}$

$$f\big(x,\mu(x)\big) = A_\mu x, \qquad g\big(x,\mu(x)\big) = q_\mu' x,$$

  where A_μ is $n \times n$ nonneg. matrix, and q_μ is n-dimensional nonneg. vector.

Critical conditions enabling exact solutions (in addition to other conditions)

- **The set of nonnegative linear functions $\widehat{\mathcal{J}}$ is closed under value iteration (VI)** in the sense that for every $c \in \Re^n_+$, the function

$$\min_{u \in U(x)} \Big[g(x,u) + \alpha c' f(x,u) \Big]$$

  belongs to $\widehat{\mathcal{J}}$, i.e., it has the form $\hat{c}' x$ for some $\hat{c} \geq 0$. Furthermore, **$\hat{c}$ depends continuously on c**.
- **There is a policy $\mu \in \widehat{\mathcal{M}}$ that attains the minimum above**, in the sense that

$$\mu(x) \in \arg \min_{u \in U(x)} \Big[g(x,u) + \alpha c' f(x,u) \Big], \qquad \text{for all } x \in X.$$

Example I: Positive Linear Systems with Control Constraint [Ran22, LR24]

- The state equation is by $x_{k+1} = Ax_k + Bu_k$, where $A \in \Re^{n\times n}$ and $B \in \Re^{n\times m}$
- The cost of stage k is $q'x_k + r'u_k$, where $q \in \Re^n_+$ and $r \in \Re^m_+$
- The control constraint set $U(x) = \{u \in \Re^m \mid |u| \leq Hx\}$, where $H \in \Re^{n\times m}_+$
- In this problem, $\widehat{\mathcal{M}}$ is the set of feasible linear policies:

$$\widehat{\mathcal{M}} = \{\mu \mid \mu(x) = Lx,\ \text{where } L \in \Re^{n\times m} \text{ and } |Lx| \leq Hx \text{ for all } x.\}.$$

- Starting with $J(x) = c'x$, $c \geq 0$, the VI operation produces the function

$$\begin{aligned}\hat{J}(x) = & \min_{|u|\leq Hx} \{q'x + r'u + c'(Ax + Bu)\} \\ =&(q + A'c)'x + \min_{|u|\leq Hx} (r + B'c)'u \\ =&(q + A'c)'x - |r + B'c|'Hx,\end{aligned}$$

  where $|\cdot|$ takes absolute values of each component.
- Let $\hat{J}(x) = \hat{c}'x$. It can be seen that $\hat{c}$ depends continuously on c.
- The minimum is attained at Lx, where L depends on r, B, H, and c.

Example II: Markov Decision Problems with Distributions as States [BS78]

- Each state is a probability distribution over n points: $x = (x^1, \ldots, x^n)$
- Each control u has n scalar components $u^1, \ldots, u^n$, with $u^i \in U^i$, $i = 1, \ldots, n$
- The system equation is

$$x_{k+1} = \sum_{i=1}^{n} p_i(u_k^i) x_k^i,$$

where the function p_i maps each u^i to a probability distribution over n points

- The objective is to minimize the total cost $\sum_{k=0}^{\infty} \alpha^k \sum_{i=1}^{n} g_i(u_k^i) x_k^i$, where $\alpha \in (0, 1)$, $g_i : U^i \mapsto \Re_+$, $i = 1, \ldots, n$
- In this problem, $\widehat{\mathcal{M}}$ is the set of constant control policies:

$$\widehat{\mathcal{M}} = \{\mu \,|\, \mu(x) = (u^1, \ldots, u^n) \text{ for all } x, \text{ where } u^i \in U^i,\ i = 1, \ldots, n.\}.$$

- Starting with $J(x) = c'x$, $c \geq 0$, the VI operation produces the function

$$\hat{J}(x) = \sum_{i=1}^{n} \min_{u^i \in U^i} [g_i(u^i)x^i + \alpha c' p_i(u^i) x^i] = \sum_{i=1}^{n} \min_{u^i \in U^i} [g_i(u^i) + \alpha c' p_i(u^i)] x^i.$$

Summary of Analytical Results

- The optimal cost function J^* satisfies $J^*(x) = (c^*)'x$, with $c^* \geq 0$
- Bellman's equation can be expressed in terms of coefficients:

$$c = G(c),$$

where $G : \Re^n_+ \mapsto \Re^n_+$ is defined uniquely through the equations

$$G(c)'x = \min_{u \in U(x)} \left[g(x,u) + \alpha c' f(x,u) \right] = \min_{\mu \in \widehat{\mathcal{M}}} G_\mu(c)'x, \qquad \text{for all } x \in X,$$

where $G_\mu = q_\mu + \alpha A'_\mu c$.

- The coefficient c^* is the unique solution of the equation $c = G(c)$ within $\Re^n_+$
- There exists an optimal policy μ^* that belongs to $\widehat{\mathcal{M}}$. Moreover, αA_{μ^*} is stable

Analytical approach applied to the problem

- The sequence $\{J_k\}$ generated via VI with $0 \leq J_0 \leq J^*$ typically converges to J^*
- Linear functions closed under VI implies that the limit J^* is also a linear function
- Uniqueness is due to 1) observability condition 2) uniqueness of the solution within the interval $[J^*, sJ^*]$ for any $s > 1$ [YB15]
- Stability of αA_{μ^*}: Perron-Frobenius theorem and X containing x of all "directions"

Summary of Computational Approaches

Synchronous and Asynchronous VI

- Starting with $J_0(x) = c_0'x$ with $c_0 \geq 0$, VI generates $\{J_k\}$ that satisfies $J_k(x) = c_k'x$, where $c_{k+1} = G(c_k)$, $k = 0, 1, \ldots$
- For every $c_0 \in \Re^n_+$, the sequence $\{c_k\}$ with $c_{k+1} = G(c_k)$ converges to c^*
- The VI algorithm for coefficients can be implemented in asynchronous and distributed fashion; see [Ber82, Ber83]

Policy Iteration and Its Variants

- Starting with $\mu_0(x) \in \widehat{\mathcal{M}}$ so that αA_{μ^0} is stable, policy iteration (PI) generates a sequence of policies $\{\mu^k\} \subset \widehat{\mathcal{M}}$ such that αA_{μ^k}, $k = 1, 2, \ldots,$ are stable
- For every μ^k, its cost function $J_{\mu^k}(x)$ satisfies $J_{\mu^k}(x) = c_{\mu^k}'x$, where $c_{\mu^k} \geq 0$
- Policy evaluation is simplified as solving linear equation $c = G_{\mu^k}(c)$; the improved policy μ^{k+1} satisfies $G_{\mu^{k+1}}(c_{\mu^k}) = G(c_{\mu^k})$
- The sequence of policies $\{\mu^k\}$ generated by PI (or its variants) satisfies $c_{\mu^k} \to c^*$ as $k \to \infty$.

Mathematical programming approach

- The coefficient c^* can be obtained by solving a convex program

Stochastic Problems with Semilinear Structure

- The stochastic version of the nonnegative cost problem is

$$\min_{\{\mu_k\}_{k=0}^{\infty}} \quad \lim_{N\to\infty} \underset{\substack{\theta_k \\ k=0,\ldots,N-1}}{E} \left\{ \sum_{k=0}^{N-1} \alpha^k g\big(x_k, \mu_k(x_k), \theta_k\big) \right\}$$

$$\text{s. t.} \qquad x_{k+1} = f\big(x_k, \mu_k(x_k), \theta_k\big), \quad \mu_k(x_k) \in U(x_k), \quad k = 0, 1, \ldots,$$

  where $\theta_k \in \Theta$ is generated according to a known stationary distribution
- For all $x \in X$, $u \in U(x)$, $\theta \in \Theta$, f satisfies $f(x, u, \theta) \in X$, $E_\theta\big\{f(x, u, \theta)\big\} \in X$
- Nonnegative cost condition $g(x, u, \theta) \geq 0$ for all $x \in X, u \in U(x), \theta \in \Theta$.

Semilinear structure

- There exists a set of policies $\widehat{\mathcal{M}}$ such that for every $\mu \in \widehat{\mathcal{M}}$ and $\theta \in \Theta$,

$$f\big(x, \mu(x), \theta\big) = A_\mu^\theta x, \qquad g\big(x, \mu(x), \theta\big) = (q_\mu^\theta)' x$$

  where $A_\mu^\theta \in \Re_+^{n\times n}$ and $q_\mu^\theta \in \Re_+^n$
- Other conditions are similar to those of deterministic problems

Certainty Equivalence Principle for Stochastic Problems

Formulation of a deterministic problem

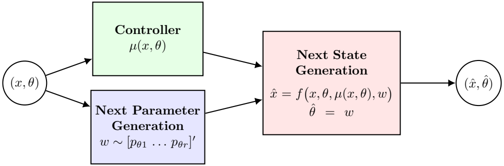
- For every policy $\mu \in \widehat{\mathcal{M}}$, we define matrix A_μ and vector q_μ as $A_\mu = E\{A_\mu^\theta\}$, $q_\mu = E\{q_\mu^\theta\}$
- We also introduce functions $\hat{f}$ and $\hat{g}$ defined as $\hat{f}(x,u) = E\{f(x,u,\theta)\}$, $\hat{g}(x,u) = E\{g(x,u,\theta)\}$
- We obtain a deterministic problem: For every $x_0 \in X$, solve

$$\min_{\{u_k\}_{k=0}^{\infty}} \sum_{k=0}^{\infty} \alpha^k \hat{g}(x_k, u_k) \quad \text{s. t. } x_{k+1} = \hat{f}(x_k, u_k),\ u_k \in U(x_k),\ k = 0, 1, \ldots,$$

Certainty equivalence principle

- The deterministic problem satisfies the semilinear conditions of previous section
- The optimal cost $\hat{J}^*(x_0)$ of the deterministic problem and the optimal cost $J^*(x_0)$ of the stochastic problem are equal
- A policy $\mu \in \widehat{\mathcal{M}}$ is optimal for the deterministic problem if and only if it is optimal for the stochastic problem

Markov Jump Problems with Nonnegative Costs



- The Markov jump problems involves a parameter set $\Theta = \{1, 2, \dots, r\}$
- The probability of $\theta_{k+1} = j$ given that $\theta_k = i$ is p_{ij}
- Control u is selected based on (x, θ) from the constraint set $U(x, \theta)$

$$
\begin{aligned}
\min_{\{\mu_k\}_{k=0}^{\infty}} \quad & \lim_{N \to \infty} \underset{\substack{w_k \\ k=0,\dots,N-1}}{E} \left\{ \sum_{k=0}^{N-1} g\big(x_k, \theta_k, \mu_k(x_k, \theta_k), w_k\big) \right\} \\
\text{s.t.} \quad & x_{k+1} = f\big(x_k, \theta_k, \mu_k(x_k, \theta_k), w_k\big), \quad k = 0, 1, \dots, \\
& \theta_{k+1} = w_k, \quad k = 0, 1, \dots, \\
& \mu_k(x_k, \theta_k) \in U(x_k, \theta_k), \quad k = 0, 1, \dots
\end{aligned}
$$

- For all $x \in X$, $u \in U(x, \theta)$, θ, $w \in \Theta$, the function f satisfies $f(x, \theta, u, w) \in X$, $E_w\{f(x, \theta, u, w) \mid \theta\} \in X$
- Nonnegative cost condition $g(x, \theta, u, w) \geq 0$ for all $x \in X, u \in U(x), \theta, w \in \Theta$.

Semilinear Structure and Equivalence Principle

Semilinear structure

- There exists a set of policies $\widehat{\mathcal{M}}$ such that for every $\mu \in \widehat{\mathcal{M}}$, θ, $w \in \Theta$ such that

$$f(x,\theta,\mu(x,\theta),w) = A_\mu^{\theta w} x, \quad g\Big(x,\theta,\mu(x,\theta),w\Big) = (q_\mu^{\theta w})' x, \quad \text{for all } x$$

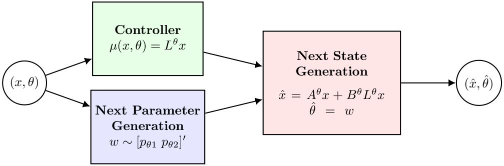
  where $A_\mu^{\theta w} \in \Re_+^{n\times n}$ and $q_\mu^{\theta w} \in \Re_+^n$
- Other conditions of stochastic problems are extended to Markov jump problems

Certainty equivalence principle

- We construct a deterministic problem involving state $\bar{x}$ whose dimension is $n \times r$
- Given $\mu \in \widehat{\mathcal{M}}$, the dynamics and stage cost are both linear with coefficients

$$\bar{A}_\mu = \begin{bmatrix} p_{11}A_\mu^{11} & p_{21}A_\mu^{21} & \cdots & p_{r1}A_\mu^{r1} \\ p_{12}A_\mu^{12} & p_{22}A_\mu^{22} & \cdots & p_{r2}A_\mu^{r2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1r}A_\mu^{1r} & p_{2r}A_\mu^{2r} & \cdots & p_{rr}A_\mu^{rr} \end{bmatrix}, \qquad \bar{q}_\mu = \begin{bmatrix} E_w\Big\{q_\mu^{1w} \mid \theta = 1\Big\} \\ E_w\Big\{q_\mu^{2w} \mid \theta = 2\Big\} \\ \vdots \\ E_w\Big\{q_\mu^{rw} \mid \theta = r\Big\} \end{bmatrix}$$

- The original Markov jump problem can be addressed by solving this deterministic problem with a higher dimension



Switched Linear Policy

- The param. set is $\Theta = \{1,2\}$ whose transition probabilities p_{ij}, $i,j = 1,2$ are given
- Given the current state (x_k, θ_k), the state equation is given by

$$x_{k+1} = A^{\theta_k} x_k + B^{\theta_k} u_k, \quad \theta_{k+1} \sim p_{\theta_k j},$$

  with stage cost $q' x_k + r' u_k$ and control constraint $U(x) = \{u \in \Re^m \mid |u| \leq Hx\}$
- The set $\widehat{\mathcal{M}}$ consists of feasible linear policies with gain matrix dependent on θ:

$$\widehat{\mathcal{M}} = \{\mu \mid \mu(x,\theta) = L^\theta x, \text{ where } L^\theta \in \Re^{n\times m} \text{ and } |L^\theta x| \leq Hx \text{ for all } x, \theta\}.$$

- The gain matrices L^θ, $\theta = 1,2$, are computed by solving a deterministic problem of the type given in Example I
- Optimal policy: linear in x with gain matrix dependent on θ, in full analogy to linear quadratic problems; see, e.g., [CWC86, CFM05]

Thank you!

References I

Dimitri P. Bertsekas.
Dynamic Programming and Stochastic Control.
Academic Press, 1976.

Dimitri P. Bertsekas.
Distributed dynamic programming.
IEEE transactions on Automatic Control, 27(3):610–616, 1982.

Dimitri P. Bertsekas.
Distributed asynchronous computation of fixed points.
Mathematical Programming, 27(1):107–120, 1983.

Dimitir P. Bertsekas and Steven Shreve.
Stochastic Optimal Control: The Discrete-Time Case.
Academic Press, 1978.

Oswaldo Luiz Valle Costa, Marcelo Dutra Fragoso, and Ricardo Paulino Marques.
Discrete-Time Markov Jump Linear Systems.
Springer Science & Business Media, 2005.

References II

Howard J. Chizeck, Alan S. Willsky, and David Castanon.
Discrete-time Markovian-jump linear quadratic optimal control.
International Journal of Control, 43(1):213–231, 1986.

Yuchao Li and Anders Rantzer.
Exact dynamic programming for positive systems with linear optimal cost.
IEEE Transactions on Automatic Control, 69(12):8738 – 8750, 2024.

Anders Rantzer.
Explicit solution to Bellman equation for positive systems with linear cost.
In *2022 IEEE 61st Conference on Decision and Control (CDC)*, pages 6154–6155. IEEE, 2022.

Huizhen Yu and Dimitri P. Bertsekas.
A mixed value and policy iteration method for stochastic control with universally measurable policies.
Mathematics of Operations Research, 40(4):926–968, 2015.