Value-Policy Iterations

Vishal Raman, Sunay Joshi

May 10, 2023

We propose an algorithm called Value-Policy Iterations as an improvement to traditional policy iterations, primarily by focusing on the policy evaluation step. There is a large body of literature focusing on approximating this step leading to a procedures such as modified/optimistic policy iterations. We give an alternate approach that decreases the overall number of policy evaluations. This can also be combined with an approximate policy evaluation step. We present the algorithm and analysis for the discounted infinite horizon problem and the stochastic shortest path problem. This work consists of progress made at the Cornell Mathematics REU 2022 and was supervised by Alexander Vladimirsky.

1 Discounted Infinite-Horizon MDP

1.1 Notation

We follow the notation from [1]. Let $S = \{1, \dots, n\}$ denote the set of states, C the compact set of controls, and $U(i) \subset C$ the set of controls available at state $i \in S$. Let $g(i, u)$ denote the cost of using control $u \in U(i)$ at state i, $J(i)$ denote the total cost function at state i, $P(u) = (p_{ij}(u))$ where $p_{ij}(u)$ is the transition probability from i to j using control u. Let $\alpha \in (0, 1)$ denote the discount factor.

We make the following assumptions:

1. $p_{ij}(\cdot)$ is continuous.
2. $g(i, \cdot)$ is continuous and bounded.

We wish to find a stationary policy $\mu : S \to C$, $\mu(x_k) \in U(x_k)$ for all $x_k \in S$, $k = 0, 1, \dots$ that minimizes the cost function

$$J_\mu(x_0) = \lim_{N\to\infty} E\left[\sum_{k=0}^{N-1} \alpha^k g(x_k, \mu(x_k))\right].$$

If we let $\Pi = U(1) \times U(2) \times \dots U(n)$, we denote the optimal cost vector as

$$J^*(x) = \min_{\mu\in\Pi} J_\mu(x).$$

We say that the stationary policy μ is optimal if $J_\mu = J^*(x)$ for all $x \in S$.

Given a cost vector $J : S \to \mathbb{R}$, we have the Bellman operator

$$TJ(i) = \min_{u\in U(i)} \left[g(i, u) + \alpha \sum_{j=1}^{n} p_{ij}(u)J(j)\right].$$

If we define a stationary policy $\mu : S \to C$, we have corresponding definitions of P_μ, T_μ, and g_μ given by $P_\mu = (p_{ij}(\mu(i)))$, $g_\mu(i) = g(i, \mu(i))$, and

$$T_\mu J(i) = g(i, \mu(i)) + \alpha \sum_{j=1}^{n} p_{ij}(\mu(i))J(j).$$

We will use the vector notation for $J, TJ, T_\mu J, g, g_\mu$ and matrices P, P_μ going forward. The cost

function J_μ corresponding to a stationary policy μ is the solution to the equation $J_\mu = T_\mu J_\mu$.

Finally, note that we can implicitly define a policy μ given by $T_\mu J = TJ$, where

$$\mu(i) = \underset{u \in U(i)}{\operatorname{argmin}} \left[g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J(j) \right],$$

in the case where the minimum is attained by a single control. Otherwise, we can choose any $u \in U(i)$ attaining the minimum.

Remark 1.1. Unless otherwise stated, we will use $\|\cdot\|$ to denote the $\ell^\infty(S;\mathbb{R})$ norm.

We now highlight important properties of the Bellman operator [1] that we will use for the proof.

Proposition 1.2 (Properties of the Bellman Operator)

The Bellman Operator T satisfies the following properties:

- T is **monotone**: given $J_1, J_2 : S \to \mathbb{R}$, if $J_1(i) \leq J_2(i)$ for all $i \in S$, then $TJ_1(i) \leq J_2(i)$ for all $i \in S$.
- T is an α-**contraction**: given $J_1, J_2 : S \to \mathbb{R}$, $\|TJ_1 - TJ_2\| \leq \alpha \|J_1 - J_2\|$.

Similarly, for any policy $\mu : S \to C$, T_μ is monotone and an α-contraction.

Proposition 1.3

For any bounded function $J : S \to \mathbb{R}$, the optimal cost function satisfies

$$J^*(x) = \lim_{N \to \infty} (T^N J)(x).$$

Furthermore, we have

$$J^* = TJ^*.$$

Corollary 1.4

For any stationary policy μ, the associated cost function satisfies

$$J_\mu(x) = \lim_{N \to \infty} (T_\mu^N J)(x).$$

Furthermore, we have

$$J_\mu = T_\mu J_\mu.$$

1.2 Algorithm

The Value-Policy Iteration Algorithm in the discounted infinite-horizon case is as follows:

1. **Initialization:**
   a) Choose $\rho \in (0,1)$, $\delta > 0$.
   b) Start with an arbitrary bounded cost function J_0.

c) Perform a value iteration TJ_0 and let $\epsilon_0 = \|TJ_0 - J_0\|$.

2. **Value Iteration**: Start at $k = 0$.
   a) Perform m_k value iterations where m_k is the first integer so that
   $$\|T^{m_k} J_k - T^{m_k - 1} J_k\| < \rho \cdot \epsilon_k.$$
   b) Set $\epsilon_{k+1} = \|T^{m_k} J_k - T^{m_k - 1} J_k\|$.
   c) Implicitly choose μ^{k+1} so that $T_{\mu^{k+1}}(T^{m_k - 1} J_k) = T^{m_k} J_k$.
   d) If $\|T^{m_k - 1} J_k - T^{m_k} J_k\| < \delta$, terminate the process.
3. **Policy Evaluation**: Using μ^{k+1}, solve for $J_{k+1} = J_{\mu^{k+1}}$ satisfying the equation
   $$(I - \alpha P_{\mu^{k+1}}) J_{\mu^{k+1}} = g_{\mu^{k+1}}.$$

   Then, return to step 2 and repeat the process.

> **Remark 1.5.** Note that the policy evaluation step can be done approximately via an iterative method(such as in the modified policy iteration procedure) or solved directly. We present the analysis where this is done exactly.

1.3 Convergence Bounds

The proof of convergence follows from the following lemma.

Lemma 1.6 (Strict Improvement Lemma)
Let μ, ν be stationary policies such that $T_\nu(T^{m-1} J_\mu) = T^m J_\mu$ for some $m \in \mathbb{N}$. Then we have $J_\nu(i) \leq J_\mu(i)$ for $i \in S$ with strict inequality for at least one i if μ is not an optimal policy ($J_\mu = TJ_\mu = J^*$).

Proof. Suppose we have $T_\nu(T^{m-1} J_\mu) = T^m J_\mu$. Then,

$$T^{m-1} J_\mu = T^{m-1}(T_\mu J_\mu) \geq T^m J_\mu = T_\nu(T^{m-1} J_\mu).$$

From monotonicity, we have

$$T^{m-1} J_\mu \geq T_\nu(T^{m-1} J_\mu) \geq \dots T_\nu^\ell(T^{m-1} J_\mu) \geq \dots \geq \lim_{\ell \to \infty} T_\nu^\ell(T^{m-1} J_\mu) = J_\nu.$$

It follows that $J_\nu \leq T^{m-1} J_\mu \leq T_\mu^{m-1} J_\mu = J_\mu$. If $J_\nu = J_\mu$, then we must also have that $J_\nu = T^{m-1} J_\mu$ since all the above inequalities must be equalities.

Therefore, we must have $J_\mu = T^{m-1} J_\mu = \lim_{\ell \to \infty} T^{\ell(m-1)} J_\mu = J^*$, so μ must be the optimal stationary policy. It follows that if μ is not optimal, then $J_\nu(i) < J_\mu(i)$ for some $i \in S$.

□

Using the improvement lemma, we show that VPI has at least a linear convergence rate. This can be improved to a superlinear rate, which we demonstrate in Section 4.

Corollary 1.7 (Convergence Rate)
Given $\{\mu^{k+1}\}$ generated by VPI, we have

$$\|J_{\mu^{k+1}} - J^*\| \leq \alpha^{m_k - 1} \|J_{\mu^k} - J^*\|.$$

Proof. By the improvement lemma, $J_{\mu^{k+1}} \leq T^{m_k-1}J_{\mu^k}$, so it follows that

$$\|J_{\mu^{k+1}} - J^*\| \leq \|T^{m_k-1}J_{\mu^k} - J^*\| = \|T^{m_k-1}J_{\mu^k} - T^{m_k}J^*\| \leq \alpha^{m_k-1}\|J_{\mu^k} - J^*\|.$$

□

2 Stochastic Shortest Path

2.1 Notation

We study the stochastic shortest path problem [2]: we have a graph with nodes $S = \{1, 2, \ldots, n, t\}$ where t is the termination state that is absorbing. At each node i, we select a probability distribution of successor nodes j parameterized by a control $u \in U$, written as $p_{ij}(u)$. A cost $g(i, u)$ is incurred for selecting $u \in U(i)$ at state i. We assume that $g(t, u) = 0$ for $u \in U(t)$. We seek to select probability distribution of successor nodes in order to reach the termination node with minimum expected cost.

As before, we define the Bellman operator

$$TJ(i) = \min_{u \in U(i)} \left[g(i, u) + \sum_{j=1}^{n} p_{ij}(u)J(j) \right].$$

If we define a stationary policy $\mu : S \to C$, we have corresponding definitions of P_μ, T_μ, and g_μ given by $P_\mu = (p_{ij}(\mu(i)))$, $g_\mu(i) = g(i, \mu(i))$, and

$$T_\mu J(i) = g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i))J(j).$$

Moreover, we say that a stationary policy μ is proper if, when using this policy, there is a positive probability that the destination will be reached after at most n stages, regardless of the initial state. Equivalently, if we let x_k denote the state after k stages, then we have

$$\gamma_\mu = \max_{i=1,\ldots,n} P(x_n \neq t | x_0 = i, \mu) < 1.$$

For convenience, we use the compact vector/matrix notation for P_μ, g_μ, J given by

$$P_\mu = \begin{bmatrix} p_{11}(\mu(1)) & \ldots & p_{1n}(\mu(1)) \\ \vdots & \ddots & \vdots \\ p_{n1}(\mu(1)) & \ldots & p_{nn}(\mu(1)) \end{bmatrix},$$

$$g_\mu = \begin{bmatrix} g(1, \mu(1)) \\ g(2, \mu(2)) \\ \vdots \\ g(n, \mu(n)) \end{bmatrix},$$

$$J = \begin{bmatrix} J(1) \\ J(2) \\ \vdots \\ J(n) \end{bmatrix}.$$

We make the following assumptions [1]:

1. There exists at least one proper policy.
2. For every improper policy μ, the corresponding cost $J_\mu(i)$ is ∞ for at least one state i.

Given the above assumptions, we have the following propositions:

Proposition 2.1 (a) For a proper policy μ, the associated cost vector J_μ satisfies

$$\lim_{k\to\infty}(T_\mu^k J)(i) = J_\mu(i).$$

Furthermore, $J_\mu = T_\mu J_\mu$ and J_μ is the unique solution to the equation.

(b) A stationary policy μ satisfying $J(i) \geq (T_\mu J)(i)$ for $i = 1, \ldots, n$ is proper.

Proposition 2.2 (a) The optimal cost vector J^* is the unique solution to the equation $J^* = TJ^*$.

(b) We have $\lim_{k\to\infty}(T^k J)(i) = J^*(i)$ for $i = 1, \ldots, n$.

(c) A stationary policy is optimal if and only if $T_\mu J^* = TJ^*$.

2.2 Algorithm

The Value-Policy Iterations algorithm in the Stochastic Shortest Path case is as follows:

1. **Initialization:**
   a) Choose $\rho \in (0, 1)$, $\delta > 0$.
   b) Start with an initial proper policy μ_0 and let $J_0 = J_{\mu^0}$, which is obtained from solving
   $$(I - P_{\mu^0})J_{\mu^0} = g_{\mu^0}.$$
   c) Perform a value iteration TJ_0 and let $\epsilon_0 = \|TJ_0 - J_0\|$.
2. **Value Iteration**: Start at $k = 0$.
   a) Perform m_k value iterations where m_k is the first integer is so that
   $$\|T^{m_k}J_k - T^{m_k-1}J_k\| < \rho \cdot \epsilon_k.$$
   b) Set $\epsilon_{k+1} = \|T^{m_k}J_k - T^{m_k-1}J_k\|$.
   c) Implicitly choose μ^{k+1} so that $T_{\mu^{k+1}}(T^{m_k-1}J_k) = T^{m_k}J_k$.
   d) If $\|T^{m_k-1}J_k - T^{m_k}J_k\| < \delta$, terminate the process.
3. **Policy Evaluation**: Using μ^{k+1}, solve for $J_{k+1} = J_{\mu^{k+1}}$ satisfying the equation
   $$(I - P_{\mu^{k+1}})J_{\mu^{k+1}} = g_{\mu^{k+1}}.$$
   Then, return to step 2 and repeat the process.

Note that the algorithm is almost the same as in the discounted infinite-horizon case, but we choose J_0 corresponding to a proper policy rather than an arbitrary bounded initial cost.

2.3 Convergence Bounds

Proposition 2.3

The value-policy iterations algorithm converges.

Proof. Note that Lemma 2.6 holds in the case of SSP when μ^k is proper for all k. But this is clear because

$$T_\nu(T^{m-1}J_\mu) = T(T^{m-1}J_\mu) \leq T^{m-1}(T_\mu J_\mu) = T^{m-1}J_\mu,$$

which implies that ν is proper by propsition 3.1b. □

As before, we have the following corollary:

Corollary 2.4

Suppose there exists a proper optimal stationary policy μ^* corresponding to J^*. Given $\{\mu^k\}$ generated by VPI, we have

$$\|J_{\mu^{k+1}} - J^*\| \leq \gamma_{\mu^*}^{\lfloor (m_k-1)/n \rfloor} \|J_{\mu^k} - J^*\|.$$

Proof. By the improvement lemma, $J_{\mu^{k+1}} \leq T^{m_k-1}J_{\mu^k}$, so it follows that

$$\begin{aligned}
\|J_{\mu^{k+1}} - J^*\| &\leq \|T^{m_k-1}J_{\mu^k} - J^*\| \\
&= \|T^{m_k-1}J_{\mu^k} - T_{\mu^*}^{m_k-1}J^*\| \\
&\leq \|T_{\mu^*}^{m_k-1}J_{\mu^k} - T_{\mu^*}^{m_k-1}J^*\| \\
&= \|P_{\mu^*}^{m_k-1}(J_{\mu^k} - J^*)\| \\
&\leq \gamma_{\mu^*}^{\lfloor (m_k-1)/n \rfloor} \|J_{\mu^k} - J^*\|.
\end{aligned}$$

□

3 Improved Rates of Convergence

3.1 Generalized Jacobians and Semismoothness

We follow the notation from [7] and [8]. Let $r : \mathbb{R}^d \to \mathbb{R}^d$ be locally Lipschitz-continuous and consider the root-finding problem $r(\theta) = 0$.

Note the Rademacher Theorem [6]:

Theorem 3.1 (Rademacher)

If $U \subset \mathbb{R}^n$ is open and $f : U \to \mathbb{R}^m$ is Lipschitz continuous, then f is differentiable almost everywhere.

By the Rademacher Theorem, r is differentiable almost everywhere, so we denote $\mathcal{M}_r$ as the set of points where r is differentiable.

We also have the following definition of the B-differential of r following from the Rademacher Theorem:

Definition 3.2 (B-differential of r)**.** The B-differential of r at $\theta \in \mathbb{R}^d$ is the set

$$\partial_B r(\theta) = \{J \in \mathbb{R}^{d\times d} : \text{exists } \{\theta_k \subset \mathcal{M}_r\} \text{ such that } \{\theta_k\} \to \theta, \{r'(\theta_k)\} \to J\}.$$

From this, we defined the generalized Jacobian:

Definition 3.3 (Generalized Jacobian)**.** The generalized Jacobian of r at $\theta \in \mathbb{R}^d$ is defined as

$$\partial r(\theta) = \text{conv}\left(\partial r_B(\theta)\right)$$

Definition 3.4. (CD-regularity) A function $f : \mathbb{R}^d \to \mathbb{R}^o$ is CD-regular at $\theta \in \mathbb{R}^d$ if each matrix $J \in \partial f(\theta)$ is nonsingular.

Definition 3.5. (BD-regularity) A function $f : \mathbb{R}^d \to \mathbb{R}^o$ is BD-regular at $\theta \in \mathbb{R}^d$ if each matrix $J \in \partial_B f(\theta)$ is nonsingular.

Proposition 3.6

Let $r : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz-continuous at $x \in \mathbb{R}^n$ with constant $L > 0$. Then, $\partial_B r(x)$ and $\partial r(x)$ are nonempty compact sets, and

$$\|J\| \leq L \quad \forall J \in \partial r(x).$$

Moreover, for any $\epsilon > 0$, there exists a neighborhood U of 0 such that

$$\text{dist}(J, \partial_B r(x)) < \epsilon \quad \forall J \in \partial_B r(x+\xi), \quad \forall \xi \in U,$$

$$\text{dist}(J, \partial r(x)) < \epsilon \quad \forall J \in \partial r(x+\xi), \quad \forall \xi \in U.$$

For a proof, see proposition 1.51 from [5]

Definition 3.7 (Directionally Differentiable)**.** A mapping Φ is called *directionally differentiable* at x in a direction ξ if the limit

$$\Phi'(x;\xi) = \lim_{t \to 0^+} \frac{\Phi(x+t\xi) - \Phi(x)}{t}$$

exists and is finite.

Definition 3.8 (Generalized Directional Derivative)**.** The generalized directional derivative of Φ at x in direction ξ is given by

$$\Phi^\circ(x;\xi) = \limsup_{h \to 0; t \to 0^+} \frac{\Phi(x+h+t\xi) - \Phi(x+h)}{t}.$$

Proposition 3.9 (Support Function)

For $f : \mathbb{R}^n \to \mathbb{R}$, the generalized directional derivative $f^\circ(x;\cdot)$ is the support function of the set $\partial f(x)$; that is

$$f^\circ(x,d) = \max_{s \in \partial f(x)} s^\intercal d,$$

$$\partial f(x) = \{s \in \mathbb{R}^n; s^\intercal d \leq f^\circ(x,d) \quad \forall d \in \mathbb{R}^n\}.$$

For a proof, see [4], Proposition 1.4.

Proposition 3.10

For a convex function $f : \mathbb{R}^n \to \mathbb{R}$, the directional derivative coincides with the generalized directional derivative; that is for all $x, d \in \mathbb{R}^n$.

$$f'(x;d) = f^\circ(x;d).$$

For a proof, see [9]

Definition 3.11 ((Strong) Semismoothness)**.** A function $f : \mathbb{R}^d \to \mathbb{R}^o$ is *semismooth* at $\theta \in \mathbb{R}^d$ if it is locally Lipschitz-continuous at θ, directionally differentiable at θ in every direction, and the following estimate holds as $\xi \in \mathbb{R}^d$ tends to zero

$$\sup_{J \in \partial f(\theta+\xi)} \|f(\theta+\xi) - f(\theta) - J\xi\| = o(\|\xi\|).$$

If instead, we have the stronger estimate

$$\sup_{J \in \partial f(\theta+\xi)} \|f(\theta+\xi) - f(\theta) - J\xi\| = O(\|\xi\|^2),$$

then f is *strongly semismooth* at x.

Proposition 3.12

For given $\Phi_1 : \mathbb{R}^n \to \mathbb{R}^m$, $\Phi_2 : \mathbb{R}^n \to \mathbb{R}^\ell$, the mapping $\Phi(\cdot) = (\Phi_1(\cdot), \Phi_2(\cdot))$ is (strongly) semismooth at $x \in \mathbb{R}^n$ if and only if Φ_1 and Φ_2 are (strongly) semismooth at x.

The following results come from [10], which proves semismoothness of a pointwise minimum over a compact family of C^1 functions.

Theorem 3.13

If $\phi : \mathbb{R}^n \to \mathbb{R}$ is convex on a convex neighborhood of $x \in \mathbb{R}^n$, then ϕ is semismooth at x.

Theorem 3.14

Let $f : \mathbb{R}^n \times T \to \mathbb{R}$ be a where T is a topological space, and U a sequentially compact subspace $U \subset T$. Define $E : \mathbb{R}^n \to \mathbb{R}$ be defined on an open subset $B \subset \mathbb{R}^n$, where

$$E(x) = \min\{f(x,u) : u \in U\}.$$

If $f(x,u)$ is continuous for $(x,u) \in B \times U$, $f(\cdot,u)$ is differentiable on B for each $u \in U$ and $\nabla_X f(\cdot,\cdot)$ is continuous and bounded on $B \times U$, then E is semismooth on B.

3.2 Semismooth Newton-type Algorithm

Consider the root-finding problem $r(x) = 0$ where $r : \mathbb{R}^n \to \mathbb{R}^n$ is not assumed to be smooth.

The Semismooth Newton Algorithm is as follows:

Choose $x^0 \in \mathbb{R}^n$ and set $k = 0$.

1. If $r(x^k) = 0$, stop.
2. Compute $J_k \in \partial r(x^k)$. Compute x^{k+1} as the solution of
$$r(x^k) + J_k(x^{k+1} - x^k) = 0.$$
3. Increase k by 1 and go to step 1.

Theorem 3.15

Let $r : \mathbb{R}^d \to \mathbb{R}^d$ be semismooth at $\theta^* \in \mathbb{R}^d$ where $r(\theta^*) = 0$, $L > 0$ and $\kappa \in [0, 1)$. Then, we have the following results:

1. For any nonsingular $B \in \mathbb{R}^{d\times d}$ such that $\|B^{-1}\| \leq L$ and there exists $J \in \partial r(\theta)$ with $\|B^{-1}(B - J)\| \leq \kappa$, we have
$$\|\theta - B^{-1}r(\theta) - \theta^*\| \leq \kappa\|\theta - \theta^*\| + o(\|\theta - \theta^*\|).$$

2. There exist an open neighborhood of θ^* such that for any θ in the neighborhood and any sequence of nonsingular matrices $\{B_k\} \subset \mathbb{R}^{d\times d}$ such that, for all k, $\|B_k^{-1}\| \leq L$ there exist $J_k \in \partial r(\theta_k)$ such that
$$\|B_k^{-1}(B_k - J_k)\| \leq \kappa_k \leq \kappa,$$
the sequence $\{\theta_k\}$ generated by the Semismooth Newton Algorithm converges to θ^* with
$$\|\theta_{k+1} - \theta^*\| \leq \kappa_k\|\theta_k - \theta^*\| + o(\|\theta_k - \theta^*\|).$$

Proof. Let $J \in \partial r(\theta)$ with $\|B^{-1}(B - J)\| \leq \kappa$. Since $r(\theta^*) = 0$,

$$\begin{aligned}
\|\theta - B^{-1}r(\theta) - \theta^*\| &= \|B^{-1}(B(\theta - \theta^*)) - B^{-1}(r(\theta) - r(\theta^*))\| \\
&= \|B^{-1}(B - J)(\theta - \theta^*)) - B^{-1}(r(\theta) - r(\theta^*) - J(\theta - \theta^*))\| \\
&\leq \|B^{-1}(B - J)(\theta - \theta^*)\| + \|B^{-1}(r(\theta) - r(\theta^*) - J(\theta - \theta^*))\| \\
&\leq \|B^{-1}(B - J)\|\|\theta - \theta^*\| + \|B^{-1}\|\|(r(\theta) - r(\theta^*) - J(\theta - \theta^*))\| \\
&\leq \kappa\|\theta - \theta^*\| + L\|(r(\theta) - r(\theta^*) - J(\theta - \theta^*))\| \\
&= \kappa\|\theta - \theta^*\| + o(\|\theta - \theta^*\|),
\end{aligned}$$

proving the first assertion.

Now, for each θ_k, θ_{k+1} is solved uniquely by the equation $\theta_{k+1} = \theta_k - B_k^{-1}r(\theta_k)$it follows that for any $q \in (\kappa, 1)$, there exists δ such that $\theta_k \in B(\theta^*, \delta)$ implies that $\|\theta_{k+1} - \theta^*\| \leq q\|\theta_k - \theta^*\|$ so it follows that $\theta_{k+1} \in B(\theta^*, \delta)$. Therefore, for $\theta_0 \in B(\theta^*, \delta)$, the sequence $\{\theta^k\} \subset B(\theta^*, \delta)$ and converges to θ^*.

Finally, from the bound $\|B_k^{-1}(B_k - J_k)\| \leq \kappa_k \leq \kappa$, we obtain the desired inequality

$$\|\theta_{k+1} - \theta^*\| \leq \kappa_k\|\theta_k - \theta^*\| + o(\|\theta_k - \theta^*\|).$$

□

A analogous result holds when r is strongly semismooth.

Corollary 3.16

Let $r : \mathbb{R}^d \to \mathbb{R}^d$ be strongly semismooth at root $\theta^* \in \mathbb{R}^d$, $L > 0$ and $\kappa \in [0, 1)$. Then, we have the following results:

1. For any nonsingular $B \in \mathbb{R}^{d\times d}$ such that $\|B^{-1}\| \leq L$ and there exists $J \in \partial r(\theta)$ with $\|B^{-1}(B - J)\| \leq \kappa$, we have
$$\|\theta - B^{-1}r(\theta) - \theta^*\| \leq \kappa\|\theta - \theta^*\| + \mathcal{O}(\|\theta - \theta^*\|^2).$$
2. There exist an open neighborhood of θ^* such that for any θ in the neighborhood and any sequence of nonsingular matrices $\{B_k\} \subset \mathbb{R}^{d\times d}$ such that, for all k, $\|B_k^{-1}\| \leq L$ there exist $J_k \in \partial r(\theta_k)$ such that
$$\|B_k^{-1}(B_k - J_k)\| \leq \kappa_k \leq \kappa,$$
the sequence $\{\theta_k\}$ generated by the Semismooth Newton Algorithm converges to θ^* with
$$\|\theta_{k+1} - \theta^*\| \leq \kappa_k\|\theta_k - \theta^*\| + \mathcal{O}(\|\theta_k - \theta^*\|^2).$$

From these, we obtain the following corollary which gives us the local convergence rate of semismooth Newton.

Corollary 3.17 (Semismooth Newton; Rate of Convergence)

Let r be semismooth and CD-regular at θ^*. Provided that θ_0 is close to θ^*, the sequence $\{\theta_k\}$ generated by semismooth Newton method with starting point θ_0 converges to θ^* superlinearly according to
$$\|\theta_{k+1} - \theta^*\| = o(\|\theta_k - \theta^*\|).$$
If instead, r is strongly semismooth at θ^*, we have the quadratic convergence rate
$$\|\theta_{k+1} - \theta^*\| = \mathcal{O}(\|\theta_k - \theta^*\|^2).$$

Proof. First, note the following lemma:

Lemma 3.18

Let $A \in \mathbb{R}^{n\times n}$ be nonsingular. Then, any matrix $B \in \mathbb{R}^{n\times n}$ satisfying $\|B - A\| \leq \frac{1}{\|A^{-1}\|}$ is nonsingular and
$$\|B^{-1} - A^{-1}\| \leq \frac{\|A^{-1}\|^2\|B - A\|}{1 - \|A^{-1}\|\|B - A\|}.$$

By the above lemma and proposition 4.6, there exists a neighborhood U of θ^* and $L > 0$ such that for all $\theta \in U$ and $J \in \partial r(\theta)$, J is nonsingular and $\|J^{-1}\| \leq L$. Setting $B_k = J_k$ in the proof of Theorem 4.6 and taking δ sufficiently small so that $B(\theta^*, \delta) \subset U$ gives the desired result. □

Remark 3.19. The assumption of CD-regularity can be replaced with BD-regularity if we select $J_k \in \partial_B r(\theta_k)$ in Theorem 4.15 and Corollary 4.16, 4.17.

3.2.1 Application to the Bellman Operator

We can consider the Bellman equation as a nonlinear root finding problem for the Bellman residual $r : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$r(\theta) = \theta - T\theta = \theta - \min_{\mu \in \tilde{\Pi}_\theta} \{T_\mu \theta\}$$

with the i-th component given by

$$\theta_i - \min_{a \in U(i)} \{g(i,a) + \alpha \sum_{j=1}^{n} p_{ij}(a)\theta_j\},$$

where $\Pi = U(1) \times U(2) \times \cdots \times U(n)$ denotes the set of policies and $\tilde{\Pi}_\theta$ denotes the set of greedy policies with respect to θ. Note that for $\mu \in \tilde{\Pi}_\theta$, $r(\theta) = \theta - T_\mu \theta$ from the definition of a greedy policy.

Unless otherwise stated, we will $\alpha \in (0, 1]$ and P_μ substochastic; $\|P_\mu\| \leq 1$. The case where $\alpha = 1$ and $\|P_\mu\| < 1$ corresponds to the stochastic shortest path problem and $\alpha \in (0, 1)$ with $\|P_\mu\| = 1$ corresponds to the discounted infinite-horizon problem.

Definition 3.20. (Spurious Greedy Policy) A policy $\mu \in \tilde{\Pi}_\theta$ is a spurious greedy policy for θ if $\text{int}(\{\tilde{\theta} : r(\tilde{\theta}) = \tilde{\theta} - T_\mu \tilde{\theta}\}) = \emptyset$. We denote the set of spurious greedy policies for θ as $\tilde{\Pi}_\theta^S$.

We also make the following additional assumptions:

1. $g(i, \cdot)$ is continuous for all $i \in S$.
2. $p_{ij}(\cdot)$ is continuous for all $i, j \in S$.
3. $U(i)$ is compact for all $i \in S$.
4. $\tilde{\Pi}_\theta^S = \emptyset$ for all θ.

We first demonstrate properties of r.

Proposition 3.21

Let $r : \mathbb{R}^n \to \mathbb{R}^n$ denote the Bellman residual. Let $\pi_s : \mathbb{R}^n \to \mathbb{R}$ denote the projection onto the s-th component and define $r^s = \pi_s \circ r$. The Bellman residual satisfies the following properties:

(a) r^s is convex for all $s \in S$.

(b) r^s is continuous for all $s \in S$

(c) r is globally Lipschitz-continuous.

(d) r is semismooth.

Proof. We first prove (a). Recall [3] that a function $f : S \to \mathbb{R}$ is concave if and only if its hypograph is a convex subset of $\mathbb{R}^n$, where the hypograph is given by

$$\text{hyp}(f) = \{(x, y) \in S \times \mathbb{R} : y \leq f(x)\}.$$

Define $f_s^a(\theta) := f_s(\theta, a) = g(s, a) + \alpha \sum_{j=1}^{n} p_{sj}(a)\theta_j$. Then $\text{hyp}(f_s^a)$ is an n-slab, the space on one side of an n-hyperplane. It is clear that a slab is convex, so it follows that

$$\text{hyp}(\min_a \{f_s^a\}) = \bigcap_{a \in U(s)} \text{hyp}(f_s^a)$$

is convex as the intersection of convex sets. Therefore, $\min_{a\in U(s)}\{f_s^a\}$ is concave, so it follows that $r^s(\theta) = \theta_s - \min_{a\in U(s)}\{f_s^a(\theta)\}$ is convex.

Next, we prove (b). Note that we can alternatively write

$$r(\theta) = \theta + \max_{\mu}\{-g_\mu - \alpha P_\mu \theta\}.$$

Since $g(s,\cdot)$, $p_{ij}(\cdot)$ are continuous, it follows that f_s is continuous. Now, recall the Maximum Theorem: [11]

Theorem 3.22 (Berge, Maximum Theorem)

Let X, Θ be topological spaces, $f : X \times \Theta \to \mathbb{R}$ be a continuous function on the product $X \times \Theta$ and $c : \Theta \to X$ be a compact-valued correspondence such that $c(\theta) \neq \emptyset$ for all $\theta \in \Theta$. Define the function $f^* : \Theta \to \mathbb{R}$ by

$$f^*(\theta) = \sup_{x\in c(\theta)}\{f(x,\theta)\}.$$

If c is continuous at θ, then f^* is continuous.

Note that we can apply the theorem to $f_s(a,\theta)$, where the correspondence $c_s : \mathbb{R}^n \to C$ is the trivial mapping $\theta \mapsto U(s)$, a compact set. Therefore, $\pi_s \circ r$ is continuous as the difference of continuous functions.

Now, we prove (c). Define $f_\mu(\theta) := g_\mu + \alpha P_\mu \theta$. It is clear that $\|\partial f_\mu(\theta)\| = \|\alpha P_\mu\| = \alpha$. It follows from [10] that $\min_{\mu\in\tilde{\Pi}_\theta}\{f_\mu\}$ is α-Lipschitz everywhere. Therefore, r is globally $(1+\alpha)$-Lipschitz.

Finally, we prove (d). We showed that $\pi_s \circ r$ is convex for each s in a, so $\pi_s \circ r$ is semismooth for each $s \in S$. It follows from Proposition 4.13 that r is semismooth as desired. □

Conjecture 3.23. r is strongly semismooth.

Remark 3.24. A roadmap to proving this is studying Theorem 4.14 from Mifflin - we have much stronger regularity properties than the one presented in this theorem(smooth functions, affine). This will strengthen the convergence rates from superlinear to quadratic.

Proposition 3.25

Let $r : \mathbb{R}^n \to \mathbb{R}^n$ be the Bellman residual. Then, r is globally CD-regular and

$$\partial r(\theta) = \operatorname{conv}\left(\{I - \alpha P_\mu : \mu \in \tilde{\Pi}_\theta\}\right).$$

Proof. Let $\theta_0 \in \mathbb{R}^n$. Define $r_\mu(\theta) = \theta - T_\mu\theta$. Note that since we assumed $\tilde{\Pi}_\theta^S = \emptyset$ for all θ, it follows that for each $\mu \in \tilde{\Pi}_{\theta_0}$, there exists a sequence of points $\theta_n \in \operatorname{int}\{\theta \in \mathbb{R}^n : r(\theta) = r_\mu(\theta)\}$ with $\theta_n \to \theta_0$. Hence, at θ_n, we have $r'(\theta_n) = r'_\mu(\theta_n) = I - \alpha P_\mu$. Hence, from the continuity of r'_μ, it follows that $I - \alpha P_\mu \in \partial_B r(\theta_0)$. Since $\partial r(\theta) = \operatorname{conv}(\partial_B r(\theta))$, it follows that

$$\operatorname{conv}\left(\{I - \alpha P_\mu : \mu \in \tilde{\Pi}_\theta\}\right) \subset \partial r(\theta).$$

Now, we prove the other inclusion. Since r^s, r^s_μ are convex, by proposition 4.10, it suffices to show that for each $s \in S$

$$\partial_S r^s(\theta_0) \subset \operatorname{conv}\{\partial_S r^s_\mu(\theta_0) : \mu \in \tilde{\Pi}_{\theta_0}\},$$

where ∂_S denotes the classical subdifferential. This follows from Theorem 4.4.2 in [8].

We first show the second part holds in the discounted infinite horizon case. Note that for $\alpha \in (0,1)$, $I - \alpha P$ is invertible for any probability matrix P and for any $\lambda \in [0,1]$ and probability matrices P_1, P_2, we have

$$\lambda(I - \alpha P_1) + (1-\lambda)(I - \alpha P_2) = I - \alpha(\lambda P_1 + (1-\lambda)P_2),$$

which is invertible since $\lambda P_1 + (1-\lambda)P_2$ is a probability matrix.

In the stochastic shortest path case, we proved that the corresponding matrix $(I - P_\mu)$ is invertible when μ is proper.

Let μ^1, μ^2 be proper policies. For $\lambda \in [0,1]$, it suffices to show that the following matrix is invertible:

$$\lambda(I - P_{\mu^1}) + (1-\lambda)(I - P_{\mu^2}) = I - (\lambda P_{\mu^1} + (1-\lambda)P_{\mu^2})$$

We show that $I - (\lambda P_{\mu^1} + (1-\lambda)P_{\mu^2})$ is invertible by proving that the corresponding Neumann series converges. In particular, if we let e denote the column vector of 1's,

$$\begin{aligned}
\sum_{k\geq 0}(\lambda P_{\mu^1} + (1-\lambda)P_{\mu^2})^k e &= \sum_{k\geq 0}\sum_{j=0}^{k}\binom{k}{j}\lambda^j(1-\lambda)^{k-j}P_{\mu^1}^j P_{\mu^2}^{k-j}e \\
&\leq \sum_{k\geq 0}\sum_{j=0}^{k}\binom{k}{j}\lambda^j(1-\lambda)^{k-j}\gamma_{\mu^1}^{\lfloor j/n\rfloor}\gamma_{\mu^2}^{\lfloor (k-j)/n\rfloor}e \\
&\leq 2ne + \sum_{k\geq 2n}\sum_{j=0}^{k}\binom{k}{j}\lambda^j(1-\lambda)^{k-j}\gamma_{\mu^1}^{\lfloor j/n\rfloor}\gamma_{\mu^2}^{\lfloor (k-j)/n\rfloor}e \\
&\leq 2ne + \sum_{k\geq 2n}\sum_{j=0}^{k}\binom{k}{j}\lambda^j(1-\lambda)^{k-j}\gamma_{\mu^1}^{\lfloor j/n\rfloor}\gamma_{\mu^2}^{\lfloor (k-j)/n\rfloor}e.
\end{aligned}$$

where we used the fact that

$$\sum_{k=0}^{2n-1}\sum_{j=0}^{k}\binom{k}{j}\lambda^j(1-\lambda)^{k-j}\gamma_{\mu^1}^{\lfloor j/n\rfloor}\gamma_{\mu^2}^{\lfloor (k-j)/n\rfloor}e \leq \sum_{k=0}^{2n-1}1^k = 2n.$$

Let $\gamma = \max\{\gamma_{\mu^1}, \gamma_{\mu^2}\} < 1$. Note the following lemma:

Lemma 3.26

For $k \geq 2n$, $j = 0, 1, \ldots, k$, we have

$$\gamma_{\mu^1}^{\lfloor j/n\rfloor}\gamma_{\mu^2}^{\lfloor (k-j)/n\rfloor} \leq \gamma^{\lfloor k/n\rfloor - 1}.$$

Proof. First, we can express $k = \ell n + r$ with $\ell \geq 2$ and $0 \leq r < n$. Then, for $j = 0, \ldots, r$, we have

$$\gamma_{\mu^1}^{\lfloor j/n\rfloor}\gamma_{\mu^2}^{\lfloor (k-j)/n\rfloor} = \gamma_{\mu^2}^{\ell} \leq \gamma^{\ell} \leq \gamma^{\ell-1}.$$

For $j = r+1, \ldots, n-1$, we have

$$\gamma_{\mu^1}^{\lfloor j/n\rfloor}\gamma_{\mu^2}^{\lfloor (k-j)/n\rfloor} = \gamma_{\mu^2}^{\ell-1} \leq \gamma^{\ell-1}.$$

For $j = n, \ldots, 2n-1$, we have

$$\gamma_{\mu^1}^{\lfloor j/n \rfloor} \gamma_{\mu^2}^{\lfloor (k-j)/n \rfloor} = \gamma_{\mu^1} \gamma_{\mu^2}^{\ell-2} \leq \gamma^{\ell-1-1}.$$

Similarly, it follows that for $j = in, in+1, \ldots, (i+1)n-1$, we have

$$\gamma_{\mu^1}^{\lfloor j/n \rfloor} \gamma_{\mu^2}^{\lfloor (k-j)/n \rfloor} = \gamma_{\mu^1}^{i} \gamma_{\mu^2}^{\ell-i-1} \leq \gamma^{\ell-1}.$$

Therefore, we have the desired claim that for all $j = 0, 1, \ldots, k$

$$\gamma_{\mu^1}^{\lfloor j/n \rfloor} \gamma_{\mu^2}^{\lfloor (k-j)/n \rfloor} \leq \gamma^{\ell-1} = \gamma^{\lfloor k/n \rfloor - 1}.$$

□

Using the above lemma, we have

$$\begin{aligned}
\sum_{k \geq 0} (\lambda P_{\mu^1} + (1-\lambda) P_{\mu^2})^k e &\leq 2ne + \sum_{k \geq 2n} \sum_{j=0}^{k} \binom{k}{j} \lambda^j (1-\lambda)^{k-j} \gamma_{\mu^1}^{\lfloor j/n \rfloor} \gamma_{\mu^2}^{\lfloor (k-j)/n \rfloor} e \\
&\leq 2ne + \sum_{k \geq 2n} \gamma^{\lfloor k/n \rfloor - 1} \sum_{j=0}^{k} \binom{k}{j} \lambda^j (1-\lambda)^{k-j} e \\
&= 2ne + \sum_{k \geq 2n} \gamma^{\lfloor k/n \rfloor - 1} 1^k e \\
&= 2ne + \frac{n\gamma}{1-\gamma} e < \infty.
\end{aligned}$$

Therefore, r is globally CD-regular as desired. □

Using the above proposition, we can now prove the superlinear convergence of VPI. We begin by proving this for the discounted infinite-horizon problem.

Theorem 3.27 (Discounted Convergence Rate)
Under VPI with sufficiently small ρ, there exists $\bar{k} \in \mathbb{N}$ such that for $k \geq \bar{k}$,

$$\|\theta_{k+1} - \theta^*\| = o(\|\theta_k - \theta^*\|).$$

Proof. First, note that $\theta_{k+1} = (I - \alpha P_{\mu^{k+1}})^{-1} g_{\mu^{k+1}}$. From Proposition 4.25, we also have that $I - \alpha P_{\mu^{k+1}} \in \partial_B r(T^{m_k - 1}\theta_k)$ since $\mu^{k+1} \in \tilde{\Pi}_{T^{m_k-1}\theta_k}$, i. e. $T_{\mu^{k+1}}(T^{m_k-1}\theta_k) = T^{m_k}\theta_k$.
Now, note that

$$\begin{aligned}
\theta_{k+1} &= (I - \alpha P_{\mu^{k+1}})^{-1} g_{\mu^{k+1}} \\
&= T^{m_k-1}\theta_k - (I - \alpha P_{\mu^{k+1}})^{-1}((I - \alpha P_{\mu^{k+1}}) T^{m_k-1}\theta_k - g_{\mu^{k+1}}) \\
&= T^{m_k-1}\theta_k - (I - \alpha P_{\mu^{k+1}})^{-1}(T^{m_k-1}\theta_k - g_{\mu^{k+1}} - \alpha P_{\mu^{k+1}} T^{m_k-1}\theta_k) \\
&= T^{m_k-1}\theta_k - (I - \alpha P_{\mu^{k+1}})^{-1} r(T^{m_k-1}\theta_k).
\end{aligned}$$

Suppose $\theta_k \in B(\theta^*, \delta)$, where δ is given from Theorem 4.5. From the semismooth Newton bound, we have

$$\|\theta_{k+1} - \theta^*\| = o(\|T^{m_k-1}\theta_k - \theta^*\|) = o(\|\theta_k - \theta^*\|),$$

so it follows that for $m \geq k$, $\theta_m \in B(\theta^*, \delta)$.

It suffices to show that there exists some k so that $\theta_k \in B(\theta^*, \delta)$. Note that $\rho^k \leq \rho$, and $\alpha_{m_k-1} \leq \alpha$ since we assert that $m_k \geq 2$. So it follows from corollary 2.8 that

$$\|\theta_{k+1} - \theta^*\| \leq \alpha\|\theta_k - \theta^*\| \leq \alpha^2\|\theta_{k-1} - \theta^*\| \leq \dots \alpha^k\|\theta_1 - \theta^*\|.$$

It suffices to choose $\bar{k}$ so that for $k \geq \bar{k}$, $\alpha^k\|\theta_1 - \theta^*\| \leq \delta$. This is given by

$$k \geq \bar{k} = \left\lceil \frac{\log(\|\theta_1 - \theta^*\|) - \log(\delta)}{1 - \log(\alpha)} \right\rceil.$$

□

Finally, we prove the superlinear convergence rate for the stochastic shortest path problem.

Corollary 3.28 (SSP Convergence Rate)

Assume there exists $\gamma \in (0, 1)$ such that $\gamma_\mu \leq \gamma$ for all proper policies μ. Furthermore, assume there exists an proper optimal stationary policy μ^* corresponding to θ^*. Under VPI with sufficiently small ρ, there exists $\bar{k}$ such that for $k \geq \bar{k}$,

$$\|\theta_{k+1} - \theta^*\| = o(\|\theta_k - \theta^*\|).$$

Proof. As above, note that

$$\theta_{k+1} = T^{m_k-1}\theta_k - (I - P_{\mu^{k+1}})^{-1}r(T^{m_k-1}\theta_k).$$

Suppose $\theta_k \in B(\theta^*, \delta)$, where δ is given from Theorem 4.5. From the semismooth Newton bound, we have

$$\|\theta_{k+1} - \theta^*\| = o(\|T^{m_k-1}\theta_k - \theta^*\|) = o(\|P_{\mu^*}^{m_k-1}(\theta_k - \theta^*)\|) = o(\|\theta_k - \theta^*\|)$$

so it follows that for $m \geq k$, $\theta_m \in B(\theta^*, \delta)$.

It suffices to show that there exists some k so that $\theta_k \in B(\theta^*, \delta)$. Note that $\rho^k \leq \rho$, and $\gamma_{\mu^*}^{\lfloor (m_k-1)/n \rfloor} \leq \gamma_{\mu^*}$, since we assert that $m_k > n$. So it follows from corollary 3.5 that

$$\|\theta_{k+1} - \theta^*\| \leq \gamma_{\mu^*}\|\theta_k - \theta^*\| \leq \gamma_{\mu^*}^2\|\theta_{k-1} - \theta^*\| \leq \dots \leq \gamma_{\mu^*}^k\|\theta_1 - \theta^*\|.$$

It suffices to choose $\bar{k}$ so that for $k \geq \bar{k}$, $\gamma_{\mu^*}^k\|\theta_1 - \theta^*\| \leq \delta$. This is given by

$$k \geq \bar{k} = \left\lceil \frac{\log(\|\theta_1 - \theta^*\|) - \log(\delta)}{1 - \log(\gamma_{\mu^*})} \right\rceil.$$

□

References

[1] Dimitri Bertsekas. *Dynamic programming and optimal control 2, 2. ed.* Athena Scientific, 2000.

[2] Dimitri P. Bertsekas and John N. Tsitsiklis. An analysis of stochastic shortest path problems. *Mathematics of Operations Research*, 16(3):580–595, aug 1991.

[3] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[4] Frank H Clarke. Generalized gradients and applications. *Transactions of the American Mathematical Society*, 205:247–262, 1975.

[5] Francisco Facchinei and Jong-Shi Pang, editors. *Finite-Dimensional Variational Inequalities and Complementarity Problems Volume II*. Springer New York, 2004.

[6] Herbert Federer. *Geometric measure theory*. Classics in mathematics. Springer, Berlin, Germany, January 1996.

[7] M. Gargiani, A. Zanelli, D. Liao-McPherson, T. H. Summers, and J. Lygeros. Dynamic programming through the lens of semismooth newton-type methods. *IEEE Control Systems Letters*, 6:2996–3001, 2022.

[8] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex Analysis and Minimization Algorithms I*. Springer Berlin Heidelberg, 1993.

[9] Jiajin Li, Anthony Man-Cho So, and Wing-Kin Ma. Understanding notions of stationarity in nonsmooth optimization: A guided tour of various constructions of subdifferential for nonsmooth functions. *IEEE Signal Processing Magazine*, 37(5):18–31, 2020.

[10] Robert Mifflin. Semismooth and semiconvex functions in constrained optimization. *SIAM Journal on Control and Optimization*, 15(6):959–972, nov 1977.

[11] Efe A. Ok. *Real Analysis with Economic Applications*. Princeton University Press, 2007.