Value Iteration Convergence Proof

Sirjan Kafle
sxkafle12@gmail.com

Contents

1 Introduction

While studying **Reinforcement Learning** by Sutton and Barto, I came across the value iteration algorithm for finite state and action spaces and pondered at the proof that it converges to the optimal value function. This took me on a rich and deep journey featuring a fantastic review of analysis. This writeup derives the proof for the convergence of the value iteration algorithm from barebones first principles[1]. We take definitions of sequences (and notion of bounded sequences), limits, and continuity along with the least upper bound property of $\mathbb{R}$ and basic set theory for granted[2].

1.1 Value Iteration Algorithm

We use the same notation as the Sutton and Barto text. The Value Iteration Algorithm is an iterative algorithm that refines value functions $V_k : \mathcal{S} \to \mathbb{R}$ until they near the optimal value function V_*. The Bellman optimal state-value function states (with discounting $\gamma \in (0, 1)$):

$$V_*(s) = \max_{a \in \mathcal{A}(s)} \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} p(s', r|s, a) \left[r + \gamma V_*(s')\right]; \forall s \in \mathcal{S}$$

The iterative algorithm for value iteration sets:

$$V_{k+1}(s) = \max_{a \in \mathcal{A}(s)} \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} p(s', r|s, a) \left[r + \gamma V_k(s')\right]; \forall s \in \mathcal{S}$$

With $V_0(s)$ defined arbitrarily.

[1] Got a lot of help from Wikipedia along the way, which I used to verify intermediate proofs

[2] **Principles of Mathematical Analysis** by Rudin has more in depth coverage of everything - I reference this a ton

Theorem 1.1. *The Value Iteration Algorithm converges to a unique optimal value function:*

$$\lim_{k\to\infty} V_k(s) = V_*(s); \forall s \in \mathcal{S}$$

This is the main theorem we want to prove.

2 Necessary Analysis Definitions and Theorems

In this section, we provide definitions and theorems (with proof) from analysis that are all necessary to prove **Thm 1.1**.

2.1 Main Definitions

Defn 2.1. *A Metric Space is a pair (X, d) of a set X and "distance" function $d : X \times X \to \mathbb{R}$ with the following properties:*

1. $\forall x, y \in X, d(x, y) > 0 \iff x \neq y, \forall x \in X, d(x, x) = 0$
2. $\forall x, y \in X, d(x, y) = d(y, x)$
3. $\forall x, y, z \in X, d(x, z) \leq d(x, y) + d(y, z)$

Defn 2.2. *Given a metric space (X, d), a Cauchy Sequence is an infinite sequence $\{x_n\}_{n\in\mathbb{Z}_+}$ such that, $\forall \epsilon > 0$, $\exists N \in \mathbb{Z}_+$, $\forall m, n > N$, $d(x_m, x_n) < \epsilon$.*

Defn 2.3. *A Complete Metric Space is a metric space (X, d) if for all Cauchy sequences $\{x_n\}_{n\in\mathbb{Z}_+}$, $\lim_{n\to\infty} x_n$ exists and is in X.*

Defn 2.4. *Given a metric space (X, d), a contraction mapping is a mapping $f : X \to X$ such that $\exists q \in (0, 1), \forall x, y \in X$, $d(f(x), f(y)) \leq qd(x, y)$.*

q is called the Lipschitz constant.

2.2 $\mathbb{R}^k$ with l_∞ Distance Metric is a Complete Metric Space

This is a critical thing to prove to lead us to **Thm 1.1**. We start with some building block theorems to work up to this. When we specify $\mathbb{R}$, we mean the metric space defined by $(\mathbb{R}, d)$ where $d(x, y) = |x - y|$.

Theorem 2.1. *Monotone Convergence Theorem: Any infinite sequence of $\mathbb{R}$ that is monotonic and bounded converges.*

Proof. Assume $\{x_n\}_{n\in\mathbb{Z}_+}$ is monotonically non-decreasing. By the least upper bound property of $\mathbb{R}$ and since $\{x_n\}$ is bounded, $x^* = \sup\{x_n\}$ exists, $x^* \in \mathbb{R}$. See that, $\forall \epsilon > 0, \exists x_N, x_N > x^* - \epsilon$. Now, $\forall n > N$, we know $x_n \geq x_N$, so $|x_n - x^*| = x^* - x_n \leq x^* - x_N < \epsilon$. Therefore, $\lim_{n\to\infty} x_n = x^*$.

Similarly, now assume that $\{x_n\}_{n\in\mathbb{Z}_+}$ is monotonically non-increasing. We know $x^* = \inf\{x_n\}$ exists, $x^* \in \mathbb{R}$. See that, $\forall \epsilon > 0, \exists x_N, x_N < x^* + \epsilon$. Now, $\forall n > N$, we know $x_n \leq x_N$, so $|x_n - x^*| = x_n - x^* \leq x^N - x^* < \epsilon$. Therefore, $\lim_{n\to\infty} x_n = x^*$.

Therefore, in both cases, we have the monotonic and bounded sequences converge. □

Theorem 2.2. *Bolzano-Weierstrass Theorem for $\mathbb{R}$: Any bounded sequence in $\mathbb{R}$ has a convergent subsequence.*

Proof. We first state and prove this Lemma:

Lemma 2.3. *Any infinite sequence in $\mathbb{R}$ has an infinite monotonic subsequence.*

Proof. Define "peak" index n be such that $\forall m > n$, $x_m \leq x_n$. Define $\mathcal{N} = \{n_1, n_2, \ldots\} \subseteq \mathbb{Z}_+$ be the set of all peak indices.

Case 1: $\mathcal{N}$ is infinite. In that case, we can order $x_{n_1} \geq x_{n_2} \geq x_{n_3} \cdots$ for an infinite $n_i \in \mathcal{N}$. Therefore, we have an infinite monotonically non-increasing subsequence.

Case 2: $\mathcal{N}$ is finite. In that case, let $N = \begin{cases} 0 & \mathcal{N} = \emptyset \\ \max(\mathcal{N}) & \mathcal{N} \neq \emptyset \end{cases}$. Repeat the following process: choose x_{N+1}. Since $N+1$ is not a peak index, $\exists n > N+1, x_n \geq x_{N+1}$. Add that x_n. n is not a peak index, so find the next max $m > n, x_m \geq x_n$. We can continue this infinitely many times to form an infinite monotonically non-decreasing subsequence. □

Now, by **Lemma 2.3**, any bounded sequence in $\mathbb{R}$ has an infinite monotonic subsequence. By **Thm 2.1**, this infinite monotonic subsequence (which is also bounded) converges. □

We're now set to show:

Theorem 2.4. *$\mathbb{R}$ is a complete metric space.*

Proof. We first prove this Lemma:

Lemma 2.5. *A Cauchy sequence on any metric space is bounded.*

Proof. Let (X, d) be a metric space, and $\{x_n\}_{n\in\mathbb{Z}_+}$ be a Cauchy sequence. We will show that $\{x_n\}$ is bounded.

Let $\epsilon = 1$. $\exists N \in \mathbb{Z}_+, \forall m, n > N, d(x_m, x_n) < 1$. Define $C = \max_{m,n \leq M} d(x_m, x_n)$ which is defined and finite. Let $B = \max\{C, 1\}$. It is thus true that, $\forall m, n, d(x_m, x_n) < B$, therefore the Cauchy sequence is bounded. □

Let $\{x_n\}_{n\in\mathbb{Z}_+}$ be a Cauchy sequence. By **Lemma 2.5** and **Thm 2.2**, the sequence has a convergent subsequence $\{x_{n_i}\}_{i\in\mathbb{Z}_+}$, $\lim_{i\to\infty} x_{n_i} = L \in \mathbb{R}$. We argue $\lim_{n\to\infty} x_n = L$.

$\forall\epsilon > 0$, $\exists I \in \mathbb{Z}_+$ such that $\forall i > I$, $|x_{n_i} - L| < \frac{\epsilon}{2}$. Furthermore, $\exists N \in \mathbb{Z}_+$ such that $\forall m, n > N, |x_n - x_m| < \frac{\epsilon}{2}$. Take $N' = \max\{n_I, N\}$. We know $\exists n_j > N'$. Therefore, $\forall n > N', |x_n - L| \leq |x_n - x_{n_j}| + |x_{n_j} - L| < \epsilon$.

Therefore, the Cauchy sequence $\{x_n\}$ converges to a value in $\mathbb{R}$. □

We are now finally ready to prove our main theorem of this section:

Theorem 2.6. *The metric space $(\mathbb{R}^k, d)$ with $d(x, y) = \|x - y\|_\infty$ is a complete metric space.*[3]

Proof. Let $\{v_n\}_{n\in\mathbb{Z}_+}$ be a Cauchy sequence. Define k sequences in $\mathbb{R}$ as follows: for $i = 1, \ldots, k$, $\{v_n^{(i)}\}_{n\in\mathbb{Z}_+}$ is the sequence such that $v_n^{(i)} = (v_n)_i$ (the i'th element of v_n). We argue all k of these sequences in $\mathbb{R}$ are Cauchy as well.

$\forall\epsilon > 0, \exists N \in \mathbb{Z}_+, \forall m, n > N, \|v_m - v_n\|_\infty < \epsilon$. $\|v_m - v_n\|_\infty = \max_i |(v_m)_i - (v_n)_i| < \epsilon$, which implies $\forall i = 1, \ldots, k$, $|(v_m)_i - (v_n)_i| < \epsilon$.

Therefore, each $\{v_n^{(i)}\}_{n\in\mathbb{Z}_+}$ is Cauchy. By **Thm 2.4**, $\forall i, \lim_{n\to\infty}\{v_n^{(i)}\}$ exists - let's set it to v_i^*.

Thus, $\lim_{n\to\infty} v_n = v^* \in \mathbb{R}^k$. Therefore, $(\mathbb{R}^k, d)$ is a complete metric space. □

2.3 Banach Fixed Point Theorem

We dedicate this section to the Banach Fixed Point Theorem which serves as the essence to proving Value Iteration convergence. First, we show:

Theorem 2.7. *Every contraction mapping is uniformly continuous.*

[3] We've taken for granted earlier and here that norm distances are valid for metric spaces. This is very easy to show.

Proof. Let (X, d) be a metric space, and $f : X \to X$ be a contraction mapping with Lipschitz constant $q \in (0, 1)$. $\forall \epsilon > 0$, let $\delta = \frac{\epsilon}{q} > 0$. Then, $\forall x, y \in X$ such that $d(x, y) < \delta$, $d(f(x), f(y)) \leq qd(x, y) < q\delta = \epsilon$. Therefore, f is uniformly continuous. $\square$

Now, we define and prove:

Theorem 2.8. *Banach Fixed Point Theorem: Given a complete metric space (X, d), any contraction mapping $T : X \to X$ will have a unique fixed point $x^* \in X$ such that $T(x^*) = x^*$. Furthermore, let $x_0 \in X$ an arbitrary point. The sequence $\{x_n\}_{n \in \mathbb{Z}_+}$ such that $x_n = T(x_{n-1})$ converges to x^*.*

Proof. Let the Lipschitz constant for T be $q \in (0, 1)$. We will first argue that $\{x_n\}$ is a Cauchy sequence.

Observe, for $n \geq 1$:

$$\begin{aligned} d(x_{n+1}, x_n) &= d(T(x_{n+1}), T(x_n)) \\ &\leq qd(x_n, x_{n-1}) \\ &\leq q^2 d(x_{n-1}, x_{n-2}) \\ &\cdots \leq q^n d(x_1, x_0) \end{aligned}$$

Now, take arbitrary $m < n \in \mathbb{Z}_+$:

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n) \\ &= \sum_{i=1}^{n-m} d(x_{m+i-1}, x_{m+i}) \\ &\leq \sum_{i=1}^{n-m} q^{m+i} d(x_1, x_0) \\ &= q^m d(x_1, x_0) \sum_{i=1}^{n-m} q^i \\ &\leq q^m d(x_1, x_0) \sum_{i=0}^{\infty} q^i \\ &= \frac{q^m d(x_1, x_0)}{1 - q} \end{aligned}$$

Now, $\forall \epsilon > 0$, $q \in (0, 1)$ implies $\exists N \in \mathbb{Z}_+$ such that $q^N < \frac{\epsilon(1-q)}{d(x_1, x_0)}$. $\forall n > m > N$, observe:

$$d(x_n, x_m) \leq \frac{q^m d(x_1, x_0)}{1 - q} < \frac{q^N d(x_1, x_0)}{1 - q} < \epsilon$$

Therefore, $\{x_n\}$ is Cauchy.

Thus, by **Defn 2.3**, $\{x_n\}$ converges: $\lim_{n \to \infty} x_n = x^*$.

By **Thm 2.7**, T is continuous, therefore $T(x^*) = \lim_{n \to \infty} T(x_n) = \lim_{n \to \infty} x_n = x^*$, showing x^* is a fixed point.

Finally, we argue x^* is a unique fixed point. Assume on the contrary $x^* \neq y^*$ were both fixed points such that $T(x^*) = x^*, T(y^*) = y^*$. Then, $d(T(x^*), T(y^*)) = d(x^*, y^*) > qd(x^*, y^*)$ which is a contradiction. Therefore, x^* is the sole fixed point.

Therefore, T has a unique fixed point $x^* = T(x^*)$ and any sequence with arbitrary $x_0 \in X$, $x_n = T(x_{n-1})$ will converge to x^*. $\square$

3 Value Iteration Convergence

We use all the previous results to prove that the Value Iteration Algorithm converges to the optimal value function.

3.1 Complete Metric Space of Value Functions

Defn 3.1. *We define a value function metric space for a given state space $\mathcal{S}$ as the set $\mathcal{V} = \{V : \mathcal{S} \to \mathbb{R}\}$ with $|\mathcal{S}| = k$ finite along with the distance function $d(V, W) = \max_{s\in\mathcal{S}} |V(s) - W(s)|$.*[4]

We show the following:

Theorem 3.1. *$(\mathcal{V}, d)$ is isometric to the metric space $(\mathbb{R}^k, d_\infty)$ where $d_\infty(v, w) = \|v - w\|_\infty$.*

Proof. We first construct the bijective mapping $f : \mathcal{V} \to \mathbb{R}$. Enumerate the states as $\mathcal{S} = \{s_1, s_2, \ldots, s_k\}$. Let $v = f(V)$ be such that $v_i = V(s_i)$. f admits an inverse $f^{-1}(v)$ for any $v \in \mathbb{R}$ as the value function $V(s_i) = v_i$ for all i.

We then show that:

$$d(V, W) = \max_{s\in\mathcal{S}} |V(s) - W(s)| = \max_{i\in\{1,\ldots,k\}} |V(s_i) - W(s_i)| = \max_{i\in\{1,\ldots,k\}} |v_i - w_i| = \|v - w\|_\infty$$

Where $v = f(V)$, $w = f(W)$.

Therefore, $(\mathcal{V}, d)$ is isometric to the metric space $(\mathbb{R}^k, d_\infty)$. □

Corollary 3.1.1. *$(\mathcal{V}, d)$ is a complete metric space.*

This corollary follows from **Thm 2.6**.

3.2 Value Iteration as a Contraction Operator on Value Functions

Defn 3.2. *The iteration operator $T : \mathcal{V} \to \mathcal{V}$ formalizes a step in the value iteration algorithm as:*

$$(T(V))(s) = \max_{a\in\mathcal{A}(s)} \sum_{s'\in\mathcal{S}, r\in\mathcal{R}} p(s', r|s, a) \left[r + \gamma V(s')\right]$$

Theorem 3.2. *The iteration operator T is a contraction mapping on the metric space $(\mathcal{V}, d)$.*

Proof. Let $V, W \in \mathcal{V}$ be arbitrary value functions. For ease of notation, define:

$$Q_V(s, a) = \sum_{s'\in\mathcal{S}, r\in\mathcal{R}} p(s', r|s, a) \left[r + \gamma V(s')\right]$$

Thus, we show:

$$\begin{aligned}
d(T(V), T(W)) &= \max_{s\in\mathcal{S}} |(T(V))(s) - (T(W))(s)| \\
&= \max_{s\in\mathcal{S}} \left| \max_{a\in\mathcal{A}(s)} Q_V(s, a) - \max_{a\in\mathcal{A}(s)} Q_W(s, a) \right| \\
&\leq \max_{s\in\mathcal{S}} \max_{a\in\mathcal{A}(s)} |Q_V(s, a) - Q_W(s, a)| \\
&= \max_{s\in\mathcal{S}} \max_{a\in\mathcal{A}(s)} \left| \sum_{s'\in\mathcal{S}, r\in\mathcal{R}} p(s', r|s, a) \left[r + \gamma V(s')\right] - p(s', r|s, a) \left[r + \gamma W(s')\right] \right| \\
&= \gamma \max_{s\in\mathcal{S}} \max_{a\in\mathcal{A}(s)} \left| \sum_{s'\in\mathcal{S}, r\in\mathcal{R}} p(s', r|s, a)(V(s') - W(s')) \right| \\
&\leq \gamma \max_{s'\in\mathcal{S}} |V(s') - W(s')| \\
&= \gamma d(V, W)
\end{aligned}$$

The second to last line follows from noting $\sum_{s'\in\mathcal{S}, r\in\mathcal{R}} p(s', r|s, a) = 1$, so the weighted average is less than the maximum.

Therefore, since $d(T(V), T(W)) \leq \gamma d(V, W)$, T is a contraction mapping with Lipschitz constant $\gamma \in (0, 1)$. □

[4] It is trivial to see this distance function defines a metric space by going through the properties in **Defn 2.1**

3.3 Putting it Altogether: Final Proof

We now prove **Thm 1.1**.

Proof. Recall $V_k = T(V_{k-1})$ as we defined. By **corollary 3.1.1**, $(\mathcal{V}, d)$ is a complete metric space. By **Thm 3.2**, T is a contraction mapping in that metric space. Therefore, by the Banach Fixed Point Theorem 2.8,

$$\lim_{k \to \infty} V_k = V_*$$

Where V_* is a unique fixed point:

$$T(V_*) = V_*$$

Expanding this, we get $\forall s \in \mathcal{S}$:

$$\begin{aligned} V_*(s) &= (T(V_*))(s) \\ &= \max_{a \in \mathcal{A}(s)} \sum_{s' \in \mathcal{S}, r \in \mathcal{R}} p(s', r | s, a) \left[r + \gamma V_*(s') \right] \end{aligned}$$

Which is exactly the optimal state-value Bellman equation!

Therefore, the Value Iteration Algorithm as defined chooses an arbitrary V_0 and constructs a sequence $V_k = T(V_{k-1})$ which we've shown converges to the unique optimal value Bellman equation V_*. $\square$

In summary, because we can write the Value Iteration Algorithm update (for discounted finite MDP setup) using an operator $T : \mathcal{V} \to \mathcal{V}$: $V_{k+1} = T(V_k)$, and since T is a contraction mapping on a complete metric space, the Value Iteration Algorithm will converge to a unique optimal V_*.