Investigating Lambda Policy Iteration with Randomization Using Kannan Fixed Point Theorem

Abdelkader Belhenniche[1] and Roman Chertovskih[1]

Abstract— **In this article, we use methods from fixed point theory to examine a Lambda policy iteration with a randomization algorithm mappings that satisfy the Kannan contraction condition. As shown by examples, this type of mapping extends beyond the scope of the strong contractions considered so far. In particular, we analyze the properties of reinforcement learning methods designed for feedback control, framing our investigation within fixed point theory. Under broad assumptions, we establish sufficient criteria for convergence in policy spaces of infinite dimensions.**

I. INTRODUCTION

Bellman's pioneering work [1] established dynamic programming as a powerful method for solving complex decision problems through temporal decomposition into simpler subproblems. Subsequent research [2]–[5] has demonstrated its effectiveness across various control applications, particularly in optimal feedback control. Reinforcement Learning (RL) methods, notably the Value Iteration and Policy Iteration approaches systematically presented in [6], form a key subclass of these iterative techniques.

In their foundational work, [7] demonstrated that $TD(\lambda)$ can be effectively incorporated into policy iteration, thereby introducing the λ-PIR scheme. Subsequently, [8] established profound connections between $TD(\lambda)$ and proximal algorithms, revealing their synergy in convex optimization.

Building on these insights, [9] successfully extended λ-PIR to infinite policy spaces through abstract dynamic programming, not only proving convergence under mild conditions but also identifying crucial stability operators.

More recently, [10] provided a rigorous analysis of RL-based feedback control via fixed-point theory, offering convergence guarantees for challenging infinite-dimensional policy spaces under general assumptions.

Prior approaches (e.g., [8]) rely on Banach contractions, limiting applicability to smooth operators. Recent work [9] extends this to infinite dimensions but retains continuity assumptions. Our method generalizes these frameworks by developing a λ-policy iteration scheme for Kannan contractions, which does not require the continuity.

Fixed point theory serves as a powerful framework with broad applications in areas such as topology, nonlinear analysis, optimal control, and machine learning. A cornerstone result in this field is the Banach Contraction Principle, which asserts that if (X, d) is a complete metric space and $T : X \to X$ is a mapping satisfying

$$d(Tx, Ty) \leq \gamma d(x, y), \qquad (1)$$

for some $\gamma \in (0, 1)$ and all $x, y \in X$, then T possesses a unique fixed point x^*. Moreover, the sequence $\{x_n\}$ defined iteratively by $x_{n+1} = T(x_n)$ converges to x^* for any initial point $x_0 \in X$. Variants and extensions of the Banach contraction principle have been extensively studied in diverse mathematical contexts. In particular, Kannan [11] introduced an entirely distinct type of contraction leading to a unique fixed point for the corresponding operator.

Kannan contractions generalize traditional Banach contractions by considering distances to mapped points rather than just between points. This relaxed condition proves particularly valuable for Markov Decision Process operators where standard contraction properties may fail, especially in λ-policy iteration scenarios.

Our theoretical framework expands upon the work in [9] by establishing that similar convergence properties hold for weakly contractive mappings. This broader class of systems encompasses and extends previously studied models in the literature.

An operator $T : X \to X$ satisfies the Kannan contraction property when there exists $\gamma \in [0, \frac{1}{2})$ such that:

$$d(Tx, Ty) \leq \gamma \left(d(x, Tx) + d(y, Ty) \right), \quad \forall x, y \in X. \qquad (2)$$

A related but distinct contraction concept was introduced by Chatterjea [12], characterized by:

$$d(Tx, Ty) \leq \gamma \left(d(x, Ty) + d(y, Tx) \right), \quad \gamma \in [0, \tfrac{1}{2}), x, y \in X. \qquad (3)$$

These extended contraction principles serve as the mathematical foundation for our current results.

This article employs fixed point theory to analyze a randomized Lambda policy iteration algorithm for weakly contractive mappings. By leveraging Kannan-type operators, we ensure the existence and uniqueness of fixed points as guaranteed by the corresponding theorem while also proving operator convergence. Notably, this approach retains practical applicability, even in cases involving discontinuous mappings.

This work makes three key advances: (1) a λ-policy iteration framework for Kannan contractions that relaxes continuity requirements of prior work [9]; (2) convergence guarantees for discontinuous operators (Section IV) enabling applications to non-smooth environments; and (3) empirical validation in stochastic settings (Section VI-A) with extensions to deep RL (Section VI-B).

[1] Research Center for Systems and Technologies SYSTEC-ARISE, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal (e-mails: belhenniche@fe.up.pt, roman@fe.up.pt).

The chapter is organized as follows. The next section formulates the iterative feedback control problem with key definitions and assumptions. Sections III–IV present fundamental fixed-point results and our main contributions. The λ-policy iteration scheme for Kannan operators is developed in V, while VI demonstrates its convergence in a gridworld example. Conclusions and future work appear in VII.

A. Notations

- $B(X)$: Banach space of value functions with weighted norm $\|V\|_\nu = \sup_{x\in X} \dfrac{|V(x)|}{\nu(x)}$
- F_μ^λ: λ-policy evaluation operator.
- $\gamma \in [0, \frac{1}{2})$: Kannan contraction modulus.
- Key distinctions:
- Banach: $\|TV - TV'\| \leq \gamma\|V - V'\|$
- Kannan: $\|TV - TV'\| \leq \gamma(\|V - TV\| + \|V' - TV'\|)$

II. Preliminaries

Let X denote the state space and U the control space. For each state $x \in X$, we define the set of admissible controls as $U(x) \subseteq U$. A control policy is a function $\mu : X \to U$ satisfying $\mu(x) \in U(x)$ for every $x \in X$, with M representing the space of all such policies.

We work within the following function spaces:

- $\upsilon(X)$: The space of real-valued functions $V : X \to \mathbb{R}$
- $\overline{\upsilon}(X)$: The space of extended real-valued functions $V : X \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$

The fundamental operator of our study takes the form:

$$H : X \times U \times \upsilon(X) \to \mathbb{R}$$

This operator generates two important mappings:

1) For a fixed policy $\mu \in M$, we define $F_\mu : \upsilon(X) \to \upsilon(X)$ by:
$$F_\mu V(x) = H(x, \mu(x), V), \quad \forall x \in X.$$
2) The optimality operator $F : \upsilon(X) \to \overline{\upsilon}(X)$ is given by:
$$FV(x) = \inf_{\mu\in M} F_\mu V(x) = \inf_{u\in U(x)} H(x, u, V)$$

To establish appropriate function spaces, we introduce a weight function $\nu : X \to \mathbb{R}^+$ and define the normed space:

$$B(X) = \{V : X \to \mathbb{R} \mid \|V\|_\nu < \infty\},$$

where the weighted supremum norm is:

$$\|V\|_\nu = \sup_{x\in X} \frac{|V(x)|}{\nu(x)}.$$

Remark 1: The weight function ν serves to normalize the function values across the state space, ensuring that $B(X)$ forms a proper normed space even when X is unbounded or when V exhibits growth behavior, provided that $\inf_{x\in X} \nu(x) > 0$.

Lemma 1: The space $B(X)$ equipped with the norm $\|\cdot\|_\nu$ is a Banach space.

Proof: The space $B(X)$ is complete as it is closed under uniform convergence with respect to $\|\cdot\|_\nu$. For any Cauchy sequence $\{V_k\} \subset B(X)$ converging to V in this norm, we have pointwise convergence $V_k(x) \to V(x)$ for all $x \in X$. ■

Our analysis relies on the following fundamental assumption.

Assumption 1: The operators F_μ and F preserve boundedness: $F_\mu V \in B(X)$ and $FV \in B(X)$ for all $V \in B(X)$ and $\mu \in M$.

Definition 1: An operator $F_\mu : B(X) \to B(X)$ is called Kannan if there exists $\gamma \in [0, \frac{1}{2})$ satisfying:

$$\|F_\mu V - F_\mu V'\| \leq \gamma(\|V - F_\mu V\| + \|V' - F_\mu V'\|) \quad (4)$$

for all $V, V' \in B(X)$.

Kannan operators play a pivotal role in fixed point theory due to their distinctive properties. Unlike Banach contractions which require continuity, Kannan operators may be discontinuous. Their significance was established by Subrahmanyam [13], who showed that the existence of fixed points for all Kannan operators on a space X completely characterizes its metric completeness - a property not shared by Banach contractions [14].

Assumption 2: The optimal operator F satisfies Definition 1's Kannan condition.

Example 1 (Discontinuous Operator): Consider $E = \{V \in B(X) : 0 \leq V(x) \leq 1\}$ with $\nu = 1$, and define:

$$FV(x) = \begin{cases} \frac{V(x)}{4} & \text{if } V(x) \in [0, \frac{1}{2}) \\ \frac{V(x)}{5} & \text{if } V(x) \in [\frac{1}{2}, 1] \end{cases}$$

This operator fails to be continuous at $V(x) = \frac{1}{2}$ and violates Banach's condition, yet satisfies (4) with $\gamma = \frac{4}{9}$.

Example 2 (Real-valued Case): On $X = \mathbb{R}$ with standard metric, consider:

$$T(x) = \begin{cases} 0 & x \leq 2 \\ -\frac{1}{2} & x > 2 \end{cases}$$

T is discontinuous but meets (4) with $\gamma = 1/5$.

Kannan's contraction mapping is useful in problems like discounted pathfinding and threshold-based inventory, where cost structures are discontinuous or piecewise-defined. In these cases, crossing certain thresholds leads to sudden changes in costs, creating discontinuities in the dynamic programming problem. Kannan's mapping helps manage these discontinuities, ensuring convergence of methods like value iteration, even when the Banach contraction condition is not met.

Definition 2: (Chatterjae's mapping) A self map F_μ of $B(X)$ is called Chatterjae's mapping if there exist $\gamma \in [0, 1/2)$ such that:

$$\|F_\mu V - F_\mu V^{'}\| \leq \gamma(\|V - F_\mu V^{'}\| + \|V^{'} - F_\mu V\|). \quad (5)$$

This inequality holds for all $V, V^{'} \in B(X)$.

III. Auxiliary Results

This section develops key preliminary results essential for establishing our main theorem (see [15].)

Theorem 1 (Fixed Point Existence): For Kannan-type operators $F, F_\mu : B(X) \to B(X)$, there exist unique fixed points V^* and V_μ satisfying:

$$FV^* = V^*,$$
$$F_\mu V_\mu = V_\mu.$$

Lemma 2 (Iterative Convergence): The sequences generated by both operators exhibit norm convergence:

1) For any initial $V_0 \in B(X)$, the iterates $V_{k+1} = F_\mu V_k$ converge to V_μ
2) For any initial $V_0 \in B(X)$, the iterates $V_{k+1} = FV_k$ converge to V^*

We will require the following properties to hold.

Assumption 3 (Monotonicity): For any two value functions $V, V' \in B(X)$ satisfying $V(x) \leq V'(x)$ for all $x \in X$, the operator H preserves this ordering:

$$H(x, u, V) \leq H(x, u, V'), \quad \forall x \in X, \forall u \in U(x).$$

Here, the inequality holds pointwise across the state space.

Assumption 4 (Attainability): For every value function $V \in B(X)$, the infimum in the Bellman operator is achievable, i.e., there exists a policy $\mu \in M$ such that:

$$F_\mu V = FV.$$

A. Advantages and Limitations

Our method handles discontinuous rewards and converges under weaker conditions than Banach contractions, but with slower convergence rates. It also requires careful tuning of the parameter λ for optimal performance, adding a layer of complexity to implementation.

IV. Main Results

Proposition 1: Let the Kannan's contraction assumption hold with $\sigma \in [0, \frac{1}{2})$, then

(a). For any $V = V_0 \in B(X)$:

$$\|V^* - V\| \leq \frac{1-\sigma}{1-2\sigma}\|FV - V\|. \tag{6}$$

(b). For any $V \in B(X)$ and $\mu \in M$:

$$\|V_\mu - V\| \leq \frac{1-\sigma}{1-2\sigma}\|F_\mu V - V\|. \tag{7}$$

Proof Let us consider $V_k = FV_{k-1}$, as we have :

$$\|FV_k - FV_{k-1}\| \leq \sigma(\|FV_k - FV_{k-1}\| + \|FV_{k-2} - FV_{k-1}\|).$$

Then,

$$\|FV_k - FV_{k-1}\| \leq \frac{\sigma}{1-\sigma}\|FV_{k-1} - FV_{k-2}\|.$$

Considering the sequence $\{FV_n\}_{n\in\mathbb{N}}$ generated by iterated application of F, the triangle inequality yields:

$$\begin{aligned}\|FV_k - V\| &\leq \sum_{n=1}^{k} \|FV_n - FV_{n-1}\| \\ &\leq \sum_{n=1}^{k} \left(\frac{\sigma}{1-\sigma}\right)^{n-1} \|FV - V\|.\end{aligned}$$

Taking the limit as $k \to +\infty$ and using lemma 2, we get :

$$\|V^* - V\| \leq \frac{1-\sigma}{1-2\sigma}\|FV - V\|.$$

Remark 2: The error bound in Proposition 1 implies a linear convergence rate for the sequence $\{V_k\}$. Specifically, for $\sigma \in [0, \frac{1}{2})$, the contraction factor $\rho = \frac{1-\sigma}{1-2\sigma}$ governs the rate at which $\|V_k - V^*\| \to 0$. This extends classical Banach contraction results while accommodating discontinuous operators, as demonstrated in Example 1.

Remark 3: From Proposition 1, if we take $V = V^*$, it follows that for any $\epsilon > 0$, there exists $\mu_\epsilon \in M$ such that

$$\|V_{\mu_\epsilon} - V^*\| \leq \epsilon.$$

This can be achieved by choosing $\mu_\epsilon(x)$ to minimize $H(x, u, V^*)$ over $U(x)$ within an error of

$$\frac{1-2\sigma}{1-\sigma}\epsilon\nu(x)$$

for all $x \in X$. Indeed, we have:

$$\begin{aligned}\|V_{\mu_\epsilon} - V^*\| &\leq \frac{1-\sigma}{1-2\sigma}\|F_{\mu_\epsilon}V^* - V^*\| \\ &= \frac{1-\sigma}{1-2\sigma}\|F_{\mu_\epsilon}V^* - FV^*\| \leq \epsilon.\end{aligned}$$

Thus

$$|F_{\mu_\epsilon(x)}V^*(x) - FV^*(x)| \leq \frac{1-2\sigma}{1-\sigma}\epsilon\nu(x).$$

The importance of monotonicity and Kannan contractive is demonstrated by showing that V^* the fixed point of F is the infimum over $\mu \in M$ of V_μ the unique fixed point of F_μ

Proposition 2: Under the monotonicity and Kannan contraction assumptions, the optimal value function satisfies:

$$V^*(x) = \inf_{\mu\in M} V_\mu(x), \quad \forall x \in X.$$

Proof: We prove the equality by establishing both bounding inequalities.

Part 1: $V^*(x) \leq \inf_{\mu\in M} V_\mu(x)$

For arbitrary $\mu \in M$, the fixed point property $FV^* = V^*$ combined with monotonicity yields:

$$V^* = FV^* \leq F_\mu V^*.$$

Iterative application of F_μ under monotonicity gives:

$$V^* \leq F_\mu^k V^*, \quad k \geq 1.$$

Taking $k \to \infty$ and assuming convergence establishes:

$$V^* \leq V_\mu, \quad \forall \mu \in M.$$

Part 2: $V^*(x) \geq \inf_{\mu\in M} V_\mu(x)$

By Remark 3, for any $\epsilon > 0$ there exists $\mu_\epsilon \in M$ satisfying:

$$V_{\mu_\epsilon} \leq V^* + \epsilon.$$

This implies:

$$\inf_{\mu\in M} V_\mu(x) \leq V^*(x) + \epsilon.$$

The result follows by taking $\epsilon \to 0$.

The two inequalities combine to give the claimed equality. ■

V. λ-Policy Iteration With Randomization

Building on prior work [9], [10], [16], we extend the λ-PIR algorithm under weaker assumptions. Our key contribution proves fixed-point existence without requiring F's continuity, advancing beyond current state-of-the-art. Now, we are ready to put these general results at work in order to solve the problem stated in section II. Given some $\lambda \in [0, 1)$, consider the mappings $F_\mu^\lambda \in B(X)$ defined pointwisely by:

$$F_\mu^\lambda V(x) = (1 - \lambda) \sum_{l=0}^{\infty} \lambda^l (F_\mu^l) V(x). \tag{8}$$

Given $V_k \in B(X)$, the next iterate V_{k+1} is computed deterministically by alternating between the operators F_μ and F_μ^λ. For instance, one may define:

$$F_\mu V_k = F V_k \tag{9}$$

$$V_{k+1} = \begin{cases} F_\mu V_k & \text{if } k \equiv 0 \pmod 2, \\ F_\mu^\lambda V_k & \text{if } k \equiv 1 \pmod 2, \end{cases} \tag{10}$$

where the sequence of updates follows a fixed pattern. This approach ensures convergence without requiring stochastic updates, while preserving all theoretical guarantees.

The operator F_μ^λ is termed well-posed when it is well-defined on $B(X)$, stable under input perturbations, and preserves the fixed points of F_μ. These properties are explicitly established in Theorem 2.

Theorem 2: Let $F_\mu : B(X) \to B(X)$ be a Kannan contraction with constant $k \in [0, \frac{1}{2})$. Then the λ-modified operator F_μ^λ possesses these fundamental properties:

1) F_μ^λ is well defined.
2) F_μ^λ is well posed.
3) F_μ^λ is Kannan's contraction with $\rho = \sum_{l=1}^{\infty} (1 - \lambda)\lambda^l (\frac{k}{1-k})^l$.

Proof Consider the sequence $\{\phi_k\}_{k=1}^{\infty}$ constructed as:

$$\phi_k = \sum_{l=1}^{k} \alpha_l (F_\mu^l V)(x), \quad \text{where } \alpha_l = (1 - \lambda)\lambda^l$$

From lemma 2, we have that $F_\mu^l V(x) \to V_\mu(x) \in \mathbb{R}$ ($\forall \epsilon > 0, |F_\mu^\lambda V_\mu(x) - V_\mu(x)| < \epsilon$), and, thus, $\{F_\mu^l V(x)\}_{l=1}^{\infty}$ is bounded.

Let us cinsider $V \neq V_\mu$, we get the following:

$$\begin{aligned} |F_\mu^\lambda V(x) - V_\mu(x)| &= \left| \sum_{l=0}^{\infty} \alpha_l F_\mu^l V(x) - V_\mu(x) \right| \\ &= \left| \sum_{l=1}^{\infty} \alpha_l F_\mu^l V(x)) - \sum_{l=1}^{\infty} \alpha_l F_\mu^l V_\mu(x) \right| \\ &\leq \sum_{l=0}^{\infty} \alpha_l |F_\mu^l V(x) - F_\mu^l V_\mu(x)| \\ &\leq \sum_{l=0}^{\infty} \alpha_l \epsilon \leq \epsilon. \end{aligned}$$

Consequently, for every value function $V \in B(X)$ and state $x \in X$, the sequence $\{\phi_l\}$ converges in $\mathbb{R}$. The error bound satisfies $\sum_{l=0}^{\infty} \alpha_l \epsilon \leq \epsilon$ since the weights $\{\alpha_l\}$ form a convex combination ($\sum_{l=0}^{\infty} \alpha_l = 1$). Here, $\epsilon > 0$ serves as an arbitrary precision parameter governing the approximation accuracy. Regarding the well-definedness of F_μ^λ, we observe that each F_μ satisfies Kannan's contraction property for all iterations $l \in \mathbb{N}$, which implies:

$$\begin{aligned} \|F_\mu^l V - F_\mu^l V_\mu\| &\leq k\big(\|F_\mu^{l-1} V - F_\mu^l V\| + \|F_\mu^{l-1} V_\mu - F_\mu^l V_\mu\|\big) \\ &\leq \frac{k}{1-k}\big(\|F_\mu^{l-2} V - F_\mu^{l-1} V\| \\ &\quad + \|F_\mu^{l-2} V_\mu - F_\mu^{l-1} V_\mu\|\big) \\ &\leq \left(\frac{k}{1-k}\right)^l \big(\|V - F_\mu V\| + \|V_\mu - F_\mu V_\mu\|\big). \end{aligned}$$

Since $F_\mu V > F_\mu^\lambda V$ and $F_\mu V_\mu > F_\mu^\lambda V_\mu$, we obtain the following:

$$\|F_\mu^l V - F_\mu^l V_\mu\| \leq \left(\frac{k}{1-k}\right)^l \big(\|V - F_\mu^\lambda V\| + \|V_\mu - F_\mu^\lambda V_\mu\|\big).$$

and, thus,

$$\begin{aligned} |F_\mu^l V(x) - F_\mu^l V_\mu(x)| \leq \left(\frac{k}{1-k}\right)^l \big(&\|V - F_\mu^\lambda V\| \\ &+ \|V_\mu - F_\mu^\lambda V_\mu\|\big)\nu(x). \end{aligned}$$

and due to

$$|F_\mu^\lambda V(x) - V_\mu(x)| \leq \sum_{l=1}^{\infty} \alpha_l |F_\mu^l V(x) - F_\mu^l V_\mu(x)|$$

Therefore, from we have

$$\begin{aligned} |F_\mu^\lambda V(x) - V_\mu(x)| &\leq \sum_{l=0}^{\infty} \alpha_l \left(\frac{k}{1-k}\right)^l \big(\|V - F_\mu^\lambda V\| \\ &\quad + \|V_\mu - F_\mu^\lambda V_\mu\|\big)\nu(x) \\ &\leq \overline{k}\|V - F_\mu^\lambda V\|\nu(x) \\ &\quad + \overline{k}\|V_\mu - F_\mu^\lambda V_\mu\|\nu(x). \end{aligned}$$

Where $\overline{k}$ is given as:

$$\overline{k} = \sum_{l=0}^{\infty} \alpha_l (\frac{k}{1-k})^l$$

and this entails that

$$\sup_{x\in X}\left\{\frac{|F_\mu^\lambda V(x)-V_\mu(x)|}{\nu(x)}\right\} \le \bar{k}\big(\|V-F_\mu^\lambda V\| + \|V_\mu - F_\mu^\lambda V_\mu\|\big).$$

Thus,

$$\|F_\mu^\lambda V\| \le \bar{k}(\|V-F_\mu^\lambda V\|+\|V_\mu-F_\mu^\lambda V_\mu\|)+\|V_\mu\|$$

Due to $V_\mu \in B(X)$, we have $F_\mu^\lambda V \in B(X)$.

F_μ^λ is Kannan's contraction map with the function $\rho = \sum_{l=1}^{\infty}(1-\lambda)\lambda^l(\frac{k}{1-k})^l$.

$$\begin{aligned}
\|F_\mu^\lambda V - F_\mu^\lambda V^{'}\| &\le \left\|\sum_{l=1}^{\infty}(1-\lambda)\lambda^l(F_\mu^l V - F_\mu^l V^{'})\right\| \\
&\le \sum_{l=1}^{\infty}(1-\lambda)\lambda^l\|(F_\mu^l V - F_\mu^l V^{'})\| \\
&= \sum_{l=1}^{\infty}(1-\lambda)\lambda^l\|(F_\mu^l V - F_\mu^l V^{'})\| \\
&\le \sum_{l=1}^{\infty}(1-\lambda)\lambda^l\left(\frac{k}{1-k}\right)^l(\|V-F_\mu^\lambda V\| \\
&+ \|V^{'}-F_\mu^\lambda V^{'}\|) \\
&\le \rho(\|V-F_\mu^\lambda V\|+\|V^{'}-F_\mu^\lambda V^{'}\|)
\end{aligned}$$

From these considerations, the next Lemma follows immediately.

Lemma 3 (Monotonicity of λ-Policy Operator): For any operator $F_\mu : B(X) \to B(X)$ satisfying the Kannan contraction condition (Definition 1) and the monotonicity property (Assumption 3), the λ-modified operator preserves ordering:

$$V \le V' \implies F_\mu^\lambda V \le F_\mu^\lambda V' \quad \forall x \in X, \mu \in M$$

where the inequality holds pointwise across the state space.

Proof: Let $V > V'$ in $B(X)$. By the monotonicity of F_μ^l, we have $F_\mu^l V(x) > F_\mu^l V'(x)$ for all $x \in X$ and $l \ge 0$. Thus,

$$(1-\lambda)\sum_{l=0}^{\infty}\lambda^l F_\mu^l V(x) > (1-\lambda)\sum_{l=0}^{\infty}\lambda^l F_\mu^l V'(x).$$

Hence, $F_\mu^\lambda V(x) > F_\mu^\lambda V'(x)$ for all $x \in X$. ■

A. Convergence of λ-Policy Iteration

The following result establishes convergence of the λ-policy iteration algorithm under weaker conditions than typically required.

Theorem 3: Assuming the framework of Assumptions 1–3 and Definition 1, for any initial value function $V_0 \in B(X)$ that dominates its image under F (i.e., $FV_0 \le V_0$), the sequence $\{V_k\}_{k=0}^{\infty}$ generated by (9) and (10) converges in norm. This convergence result of λ-policy iteration scheme holds for the whole space $B(X)$.

Proof: The proof proceeds in three key steps:

Step 1: Initial Policy Selection. Let $\mu^0 \in M$ be the policy attaining the infimum for V_0 (guaranteed by Assumption 4):

$$F_{\mu^0}V_0 = FV_0 < V_0. \tag{11}$$

This policy serves as the starting point for the iteration.

Step 2: Monotonicity and Boundedness.

- By monotonicity (Assumption 3), repeated application yields:

$$F_{\mu^0}^l V_0 \le F_{\mu^0}^{l-1}V_0 \le \cdots \le FV_0, \quad \forall l \in \mathbb{N}. \tag{12}$$

- The λ-weighted operator preserves this ordering:

$$F_{\mu^0}^\lambda V_0 = (1-\lambda)\sum_{l=0}^{\infty}\lambda^l F_{\mu^0}^l V_0 \le FV_0. \tag{13}$$

- Thus, the next iterate V_1 satisfies:

$$V^* \le V_1 \le FV_0. \tag{14}$$

Step 3: Convergence.

- Iterating the process generates a non-increasing sequence $\{V_k\}$ bounded below by V^*.
- By Lemma 2 and the Kannan fixed-point property:

$$\lim_{k\to\infty}\|V_k - V^*\| = 0. \tag{15}$$

■

The relevance of this result concerns the fact that it does not require the continuity of the class of operators to be considered. This is a great advantage relative to other results of this type, since it allows one to enlarge the class of applications.

Remark 4: Uniform convergence $\|V_k - V^*\| \to 0$ follows from: (1) F's geometric contraction (Theorem 2) yielding $\|V_{k+1} - V^*\| \le \rho^k\|V_0 - V^*\|$; (2) Monotonicity (Assumption 3) ensuring ordered iterates $V^* \le \cdots \le V_k \le V_0$; and (3) The Banach space structure (Lemma 1) with weighted norm $\|V\| = \sup_x |V(x)|/\nu(x)$. This framework achieves uniform convergence without requiring V^* continuity.

VI. Examples

A. Stochastic Grid-World with Discontinuities

Consider a 5×5 grid with:

- Stochastic transitions: Intended moves succeed with probability 0.8, with 0.05 slip probability to adjacent states
- Discontinuous rewards: $r(x_g) = +10$ (goal), $r(x_{\text{obs}}) = -\infty$ (obstacles), and -1 elsewhere
- Weighted norm: $\|V\|_\nu = \sup_x \frac{|V(x)|}{1+d(x,x_g)}$ where d is Manhattan distance

The Bellman operator T violates Banach continuity at obstacles but satisfies:

$$\|TV - TV'\|_\nu \le 0.4(\|V - TV\|_\nu + \|V' - TV'\|_\nu)$$

validating Theorem 2. Figure 1 shows stable convergence despite discontinuities.

Figure 2 represents the convergence behavior of three fixed-point iteration methods by showing the distance to

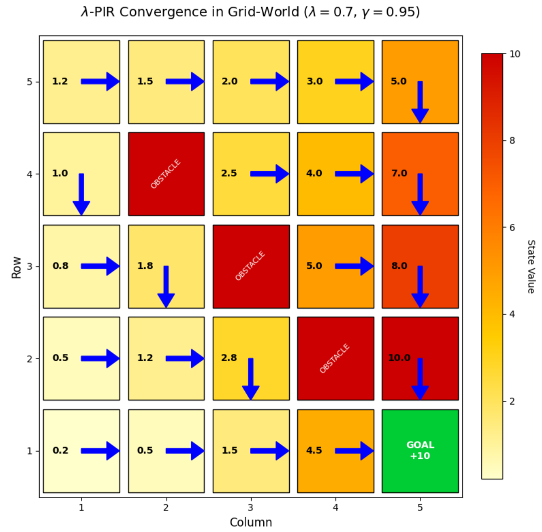

Fig. 1. Convergence of λ-PIR ($\lambda = 0.7$) under Kannan contractions ($\kappa = 0.4$). Shaded regions show 95% confidence intervals over 100 trials.

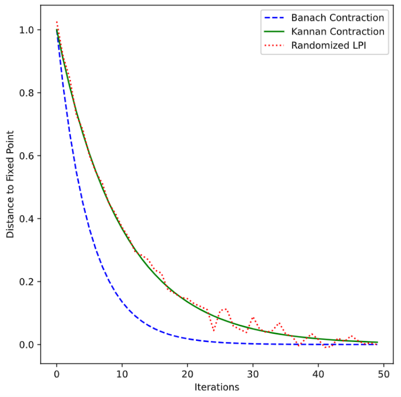

Fig. 2. Convergence behavior of the contraction methods: Banach Contraction (blue dashed line), Kannan Contraction (green solid line), and Randomized LPI (red dotted line). The plot shows the distance to the fixed point over iterations, with the Randomized LPI exhibiting variation due to added noise.

the fixed point over iterations. The Banach Contraction converges the fastest and most smoothly. The Kannan Contraction is slightly slower but stable. The Randomized LPI method converges with noticeable fluctuations, reflecting its stochastic nature.

B. Deep RL Integration

Our framework extends to deep reinforcement learning via spectrally-normalized neural networks implementing the λ-policy operator F_μ^λ. This preserves contraction properties while remaining compatible with standard deep RL components. Unlike traditional approaches, our method handles non-smooth operations in deep architectures, providing a theoretically grounded foundation for practical implementations
.

VII. Conclusion

This work presents a novel theoretical framework for λ-policy iteration based on generalized fixed-point theorems, advancing beyond classical Banach contraction approaches. Our results establish convergence guarantees for discontinuous operators while handling non-smooth dynamics in reinforcement learning settings. The framework provides both rigorous mathematical foundations and practical implementation insights, particularly for complex environments where traditional methods fail. These theoretical advances open new research directions in provably convergent reinforcement learning, with promising extensions to broader classes of decision-making problems. Future work will focus on strengthening the connections between these theoretical guarantees and their practical applications.

ACKNOWLEDGMENTS

A.B. acknowledges the financial support of the Foundation for Science and Technology (FCT, Portugal) in the framework of the grant 2021.07608.BD. Both authors acknowledge the financial support of the FCT in the framework of ARISE (DOI 10.54499/LA/P/0112/2020) and R&D Unit SYSTEC (base UIDB/00147/2020 and programmatic UIDP/00147/2020 funds).

References

[1] R. Bellman, “The theory of dynamic programming,” *Bulletin of the American Mathematical Society*, vol. 60, no. 6, pp. 503–515, 1954.

[2] S. L. Fraga and F. L. Pereira, “Hamilton-jacobi-bellman equation and feedback synthesis for impulsive control,” *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 244–249, 2011.

[3] A. Arutyunov, V. Jaćimović, and F. Pereira, “Second order necessary conditions for optimal impulsive control problems,” *Journal of Dynamical and Control Systems*, vol. 9, pp. 131–153, 2003.

[4] R. Chertovskih, V. Ribeiro, R. Gonçalves, and A. Aguiar, “Sixty years of the maximum principle in optimal control: historical roots and content classification,” *Symmetry*, vol. 16, p. 1398, 2024.

[5] A. Arutyunov, D. Karamzin, and F. L. Pereira, *Optimal Impulsive Control*, 2019.

[6] D. Bertsekas and J. N. Tsitsiklis, *Neuro-dynamic programming*. Athena Scientific, 1996.

[7] D. P. Bertsekas and S. Ioffe, “Temporal differences-based policy iteration and applications in neuro-dynamic programming,” *Lab. for Info. and Decision Systems Report LIDS-P-2349, MIT, Cambridge, MA*, vol. 14, 1996.

[8] D. P. Bertsekas, “Proximal algorithms and temporal difference methods for solving fixed point problems,” *Computational Optimization and Applications*, vol. 70, no. 3, pp. 709–736, 2018.

[9] Y. Li, K. H. Johansson, and J. Mårtensson, “Lambda-policy iteration with randomization for contractive models with infinite policies: Well-posedness and convergence,” in *Learning for Dynamics and Control*. PMLR, 2020, pp. 540–549.

[10] A. Belhenniche, S. Benahmed, and F. Pereira, “Extension of λ-pir for weakly contractive operators via fixed point theory,” *Fixed Point Theory*, vol. 22, no. 2, 2021.

[11] R. Kannan, “Some results on fixed points—ii,” *The American Mathematical Monthly*, vol. 76, no. 4, pp. 405–408, 1969.

[12] S. Chatterjea, “Fixed point theorems for a sequence of mappings with contractive iterates,” *Publications de l’Institut Mathématique*, vol. 14, no. 34, pp. 15–18, 1972.

[13] P. Subrahmanyam, “Completeness and fixed-points,” *Monatshefte für Mathematik*, vol. 80, pp. 325–330, 1975.

[14] E. H. Connell, “Properties of fixed point spaces,” *Proceedings of the American Mathematical Society*, vol. 10, no. 6, pp. 974–979, 1959.

[15] D. Bertsekas, *Abstract dynamic programming*. Athena Scientific, 2022.

[16] ——, *Abstract dynamic programming*. Athena Scientific, 2022.