

Abstract Dynamic Programming on Partially Ordered Spaces

Chengyuan Peng^{*} John Stachurski[†] Jingni Yang[‡]

June 10, 2026

Abstract

We study abstract dynamic programs on partially ordered spaces, pairing the order-theoretic approach to dynamic programming with topological and metric foundations. We show that readily verifiable forms of topological stability, such as global stability and contractivity of the policy operators, deliver the fundamental optimality properties of dynamic programming together with convergence of value function iteration, Howard policy iteration, and optimistic policy iteration. We also prove that stationary policies dominate nonstationary policy plans under very weak assumptions. Applications include Markov decision processes, structural estimation problems in which maximization and integration are interchanged, optimal stopping without discounting, and Bayesian sequential analysis. For the last two, our results weaken existing assumptions and extend algorithmic guarantees for foundational problems.

Keywords: dynamic programming, Bellman equation, optimal stopping

1 Introduction

Recent years have seen rapid and continuous growth in the study and use of dynamic programming, a methodology for optimization based on recursion. Dynamic programming is routinely applied to existing and new problems in economics, finance

^{*}Research School of Economics, Australian National University, Australia. Email: chengyuanp2022@gmail.com

[†]National Graduate Institute for Policy Studies (GRIPS). Email: j-stachurski@grips.ac.jp

[‡]School of Economics, University of Sydney, Australia. Email: jingni.yang@sydney.edu.au

and operations research, with applications ranging from real options, insurance, firm investment, and portfolio allocation to fiscal policy, job search, economic geography, and production chains (see, e.g., [Carroll \[2009\]](#), [Bäuerle and Rieder \[2011\]](#), [Bommier and Villeneuve \[2012\]](#), [Hsu et al. \[2014\]](#), [Rendahl \[2016\]](#), [Light \[2018\]](#), [Zhu \[2020\]](#), [Cao \[2020\]](#), or [Han et al. \[2026\]](#)). In addition, dynamic programming forms the foundations of reinforcement learning, which is used for tasks as diverse as autonomous driving, hybrid vehicle power management, control of chemical reactors, dynamic pricing, social media feeds, supply chain optimization, and post-training of LLMs (see, e.g., [Kochenderfer et al. \[2022\]](#) or [van Niekerk et al. \[2025\]](#)).

In addition to this rapidly growing range of applications, researchers have modified and extended dynamic programming techniques such that any one of these applications can be viewed or treated in multiple ways. For example, while traditional dynamic programming maximizes expected rewards (or minimizes expected costs), new variations can inject alternative criteria, such as hyperbolic discounting, risk aversion, ambiguity aversion, desire for robustness, adversarial agents, or concern for the entire path of reward distributions (see, e.g., [Liu et al. \[2021\]](#), [de Groot et al. \[2021\]](#), [Jaśkiewicz and Nowak \[2021\]](#), [Dugeon and Wigniolle \[2021\]](#), [Balbus et al. \[2022\]](#), [Bloise et al. \[2024\]](#), [Fadina et al. \[2024\]](#), [Bäuerle and Jaśkiewicz \[2025\]](#), [de Castro et al. \[2025\]](#), [Bäuerle and Vasileiadis \[2026\]](#) [Balbus et al. \[2026\]](#), or [Jaśkiewicz et al. \[2026\]](#)). Finally, across all these applications and variations of the basic model, researchers have steadily increased the set of ways in which dynamic programs can be formulated, using representations such as Q-factors from Q-learning or post-action value functions from structural estimation to transform existing problems in order to obtain computational or analytical advantages (see, e.g., [Kristensen et al. \[2021\]](#) or [Iskhakov et al. \[2020\]](#)).

One of the great challenges for researchers in the field of dynamic programming is to construct fundamental optimality and algorithmic results that can handle the vast range of applications and methodologies stemming from this growth. A crucial contribution in this direction is the monograph of [Bertsekas \[2022\]](#), which began the construction of an “abstract dynamic programming” (ADP) framework that is general enough to handle many of the problems listed above.¹ These ideas were further extended by [Sargent and Stachurski \[2025b\]](#) and [Peng and Stachurski \[2026\]](#), using an order-theoretic approach that can facilitate machine reasoning while also providing the potential to handle additional applications (e.g., recent methods such as distributional dynamic programming, as well as the reformulated dynamic programs found in Q-learning and structural estimation, where maximization and integration are in-

¹The framework of [Bertsekas \[2022\]](#) combines and extends earlier work that includes [Denardo \[1967\]](#), [Bertsekas \[1977\]](#), [Szepesvari \[1998\]](#), [Bertsekas \[2017\]](#), and [Li and Xie \[2021\]](#).

terchanged).

In this paper we help to realize the potential of abstract dynamic programming by pairing these order-theoretic techniques with topological structure. Our motivation is as follows. In the order-theoretic framework, the fundamental optimality and convergence results rest on *order stability* of the policy operators. While order stability can yield simple proofs, it is also abstract and difficult to verify, as well as limited in the range of optimality and convergence outcomes it can deliver. To remedy these issues, we work here with *pospaces* – topological spaces endowed with a closed partial order – in which standard and readily checkable forms of topological stability, such as global stability and contractivity, imply order stability. In this pairing, order theory supplies the optimality structure, while topology supplies sufficient conditions and additional convergence results.

Our main contributions are as follows. First, for regular and globally stable ADPs on pospaces, we show that existence of a fixed point of the Bellman operator delivers the fundamental optimality properties – existence of optimal policies, uniqueness of the value function as a solution to the Bellman equation, and Bellman’s principle of optimality – together with convergence of the three core algorithms of dynamic programming: value function iteration, Howard policy iteration, and optimistic policy iteration. Second, we provide versions of this result in which all conditions are placed on primitives. Third, specializing to partially ordered metric spaces, we show that contractivity of the policy operators yields the same optimality results, along with geometric convergence of value function iteration. Fourth, we prove that, under the same weak conditions, every nonstationary policy plan is weakly dominated by a stationary policy, so that restricting attention to stationary policies entails no loss of generality.

Using these foundations, we study four major applications. The first is a benchmark: for Markov decision processes (MDPs) with weak Feller transition kernels, we recover standard optimality results, confirming that the general theory specializes correctly. The second treats the optimization step of structural estimation in discrete choice models, where maximization and integration are interchanged relative to the standard Bellman formulation. Such problems lie outside the classical framework of [Bellman \[1957\]](#) and [Puterman \[2005\]](#), as well as the generalized contractive frameworks of [Denardo \[1967\]](#) and [Bertsekas \[2022\]](#); nonetheless, they are easily expressed as ADPs, and our optimality and convergence results apply directly. The third and fourth applications concern problems in which discounting is absent and contractivity fails entirely: optimal stopping without discounting and Bayesian sequential analysis. For these classical problems, our framework yields several new results.

Regarding optimal stopping, the discrete-time theory has a long history, beginning

with the sequential analysis of [Wald \[1947\]](#) and the foundational stopping-rule theory of [Chow and Robbins \[1963\]](#); textbook treatments include [Shiryaev \[2007\]](#), while [Peskir and Shiryaev \[2006\]](#) develops the continuous-time theory in detail. The most closely related dynamic-programming framework is the stochastic shortest path (SSP) literature initiated by [Bertsekas and Tsitsiklis \[1991\]](#) and extended in [Yu and Bertsekas \[2013\]](#) and [Bertsekas \[2018\]](#), in which contractivity is replaced by a properness condition that guarantees termination in finite expected time. Our optimal stopping result complements this literature: working on a general measurable state space, and under the single assumption that a canonical “stop-when-cheap-to-exit” policy has uniformly bounded expected hitting time, we obtain the full set of optimality properties together with convergence of all three core algorithms.

Our final application is the classical Bayesian sequential analysis problem introduced by [Wald \[1947\]](#) and [Arrow et al. \[1949\]](#); see Chapter 5 of [Bertsekas \[1976\]](#) for a textbook treatment. This problem has a substantial modern literature. For example, [Naghshvar and Javidi \[2013\]](#) study active sequential hypothesis testing with action-dependent observations and obtain lower bounds and asymptotically optimal heuristic policies under information-theoretic regularity conditions (notably, positive Kullback–Leibler divergence between the hypotheses), while [Ekström and Wang \[2022\]](#) treat a composite-hypothesis variant within exponential families and establish concavity of the cost function and monotonicity of the optimal stopping boundary. Neither framework supplies convergence of Howard or optimistic policy iteration, and both impose stronger assumptions on the observation model than we require. Applying our optimal stopping results, we obtain existence of optimal policies, uniqueness of the value function as a solution to the Bellman equation, and convergence of all three core algorithms, under the minimal condition that the candidate densities differ on a set of positive measure.

The remainder of the paper is structured as follows. Section 2 contains mathematical preliminaries and reviews the ADP framework, including the fundamental optimality properties and the core algorithms. Section 3 states and proves our main optimality and convergence results, treating ADPs on pospaces, ADPs on partially ordered metric spaces, nonstationary policies, and minimization. Section 4 presents the applications and Section 5 concludes. Longer proofs are deferred to the appendix.

2 Preliminaries

This section collects preliminaries from order theory, topology, and metric space theory, and then reviews abstract dynamic programs, their optimality properties, and the

core algorithms used to solve them.

2.1 Mathematical Preliminaries

This section collects the definitions and results from order theory, metric space theory and topology that are used in the rest of the paper.

As usual, a **partially ordered set** is a set V paired with a partial order $\preceq$. An **order interval** in V is a set of the form $\{v \in V : a \preceq v \preceq b\}$ for some $a, b \in V$. When V is understood, we denote this set by $[a, b]$. A set $A \subset V$ is called **order bounded** if there exists an order interval $[a, b] \subset V$ with $A \subset [a, b]$. The set V is called **countably Dedekind complete** if every nonempty and countable subset that is bounded above has a supremum in V , and every nonempty and countable subset that is bounded below has an infimum in V .

A sequence (v_n) in V is called **increasing** if $v_n \preceq v_{n+1}$ for all $n \in \mathbb{N}$. If (v_n) is increasing and $\bigvee_n v_n = v$ for some $v \in V$, then we write $v_n \uparrow v$. (Here and below, $\bigvee$ denotes a supremum and $\bigwedge$ denotes an infimum.) Similarly, we write $v_n \downarrow v$ if (v_n) is decreasing and $\bigwedge_n v_n = v$ for some $v \in V$. A self-map $S : V \rightarrow V$ is called **order preserving** if $u \preceq v$ implies $Su \preceq Sv$. The map S is called **order stable** if S has a unique fixed point $\bar{v}$ in V and the following conditions hold:

- (i) $v \preceq Sv$ implies $v \preceq \bar{v}$ and
- (ii) $Sv \preceq v$ implies $\bar{v} \preceq v$.

Order stability is closely related to more standard forms of stability involving topologies. To clarify this fact and introduce other related concepts, suppose now that V also has a topology and let S be a self-map on V . We call S **globally stable** on V if S has a unique fixed point $u^* \in V$ and $S^k u \rightarrow u^*$ as $k \rightarrow \infty$ for all $u \in V$.

A partial order $\preceq$ on V is called **closed** if, for every directed index set Λ and any nets $(u_\alpha)_{\alpha \in \Lambda}$ and $(v_\alpha)_{\alpha \in \Lambda}$ in V ,

$$u_\alpha \rightarrow u, \quad v_\alpha \rightarrow v \quad \text{and} \quad u_\alpha \preceq v_\alpha \text{ for all } \alpha \in \Lambda \quad \implies \quad u \preceq v.$$

A **partially ordered space**, also called a **pospace**, is a topological space endowed with a closed partial order. We will make use of the following elementary result, which is valid in any pospace.

Lemma 2.1. *If $v \in V$ with $v_\alpha \rightarrow v$ and $v_\alpha \preceq v$ for all $\alpha \in \Lambda$, then $\bigvee_\alpha v_\alpha = v$.*

Proof. Let (v_α) and v be as stated. By assumption, v is an upper bound of (v_α) . If w is any other upper bound, then $v_\alpha \preceq w$ for all α . Since $\preceq$ is closed and $v_\alpha \rightarrow v$, this implies $v \preceq w$. Hence v is the least upper bound of (v_α) . $\square$

Lemma 2.2. *Let V be a pospace and let S be an order preserving self-map on V . If S is globally stable on V , then S is order stable on V .*

Proof. Let S, V have the stated properties and let $\bar{v}$ be the unique fixed point of S in V . If $v \in V$ and $v \preceq S v$, then, iterating on this inequality and using the fact that S is order preserving, we have $v \preceq S^n v$ for all $n \in \mathbb{N}$. Since the partial order is closed and S is globally stable, taking the limit gives $v \preceq \bar{v}$. This proves one direction in the definition of order stability. The proof of the other direction is similar. $\square$

A **partially ordered metric space** is a tuple $(V, \preceq, d)$ where $(V, \preceq)$ is a partially ordered set, d is a metric on V , and $(V, \preceq)$ is a pospace under the topology induced by d . In particular, $\preceq$ is closed with respect to d . The metric d is called **sup-nonexpansive** if, for any index set Λ and any nets $(v_\alpha)_{\alpha \in \Lambda}$ and $(w_\alpha)_{\alpha \in \Lambda}$ in V ,

$$d\left(\bigvee_{\alpha} v_{\alpha}, \bigvee_{\alpha} w_{\alpha}\right) \leq \bigvee_{\alpha} d(v_{\alpha}, w_{\alpha})$$

whenever both suprema on the left-hand side exist.

As usual, a self-map S on V is called a **contraction of modulus $\lambda \in [0, 1)$** if

$$d(Su, Sv) \leq \lambda d(u, v) \quad \text{for all } u, v \in V. \quad (1)$$

The map S is called **asymptotically contracting** if

$$\lim_{n \rightarrow \infty} d(S^n u, S^n v) = 0 \quad \text{for all } u, v \in V.$$

Every contraction map is asymptotically contracting. If, in addition, (V, d) is complete, then Banach's contraction mapping theorem gives S a unique fixed point.

2.2 Background on ADPs

Abstract dynamic programs in the sense of the term used here were introduced in [Sargent and Stachurski \[2025b\]](#). To make the paper somewhat self-contained, we review the key definitions.

Recalling the definition from [Sargent and Stachurski \[2025b\]](#), an **abstract dynamic program** (ADP) is a pair $(V, \mathbb{T})$, where

- (i) $V = (V, \preceq)$ is a partially ordered set and
- (ii) $\mathbb{T} = \{T_\sigma : \sigma \in \Sigma\}$ is a nonempty family of order preserving self-maps on V .

In what follows,

- V is called the **value space**.
- Each operator T_σ in $\mathbb{T}$ is called a **policy operator**.
- Σ is an arbitrary index set and elements of Σ will be referred to as **policies**.

In applications we will impose conditions under which each T_σ has a unique fixed point in V . In these settings, the significance of T_σ is that its fixed point represents the lifetime value of following policy σ . We will denote the unique fixed point of T_σ by v_σ and call it the **σ -value function**.

Let $(V, \mathbb{T})$ be an ADP with policy set Σ . Given $v \in V$, we

- say that $\sigma \in \Sigma$ is **v -greedy** if $T_\tau v \preceq T_\sigma v$ for all $\tau \in \Sigma$, and
- set $V_G := \{v \in V : \text{at least one } v\text{-greedy policy exists}\}$.

In many applications, solving for a v -greedy policy is straightforward. Moreover, there exist conditions under which solving the overall problem reduces to solving for a v -greedy policy with a “correct” choice of v . This idea is at the heart of dynamic programming and the details are clarified below.

We now list properties useful for ADP optimality theory. First, we call $(V, \mathbb{T})$

well-posed if each $T_\sigma \in \mathbb{T}$ has a unique fixed point in V .

Well-posedness is a minimal precondition for a coherent dynamic program: in order to maximize over policies, we need well-defined lifetime values. When well-posedness holds we will always use v_σ to denote the unique fixed point of T_σ . Also, we set

$$V_\Sigma := \{v \in V \text{ such that } T_\sigma v = v \text{ for some } T_\sigma \in \mathbb{T}\}.$$

We call $(V, \mathbb{T})$

- **finite** if $\mathbb{T}$ is a finite set,
- **regular** if $V_G = V$; that is, if a v -greedy policy exists for all $v \in V$,
- **order stable** if each $T_\sigma \in \mathbb{T}$ is order stable on V , and
- **order bounded** if there exists a $u \in V$ with $T_\sigma u \preceq u$ for all $T_\sigma \in \mathbb{T}$.

Finiteness holds for many standard models, such as Markov decision processes (MDPs) with finite state and action spaces. Regularity helps us to construct algorithms and to obtain existence of optimal policies. Order stability is important for obtaining optimality properties, as emphasized in [Sargent and Stachurski \[2025b\]](#). Order boundedness is a useful property for optimality and is implied by, but does not imply, the value space V itself being order bounded.

As in [Sargent and Stachurski \[2025b\]](#), we define the **Bellman operator** generated by $(V, \mathbb{T})$ via

$$Tv := \bigvee_{\sigma} T_{\sigma} v \quad \text{whenever the supremum exists.}$$

In addition, we say that $v \in V$ satisfies the **Bellman equation** if

$$v = \bigvee_{\sigma} T_{\sigma} v.$$

Here the supremum is taken over all $\sigma \in \Sigma$. There is, in general, no guarantee that the supremum exists. Evidently $v \in V$ satisfies the Bellman equation if and only if Tv exists and $Tv = v$.

2.3 Optimality

For the remainder of this section, $(V, \mathbb{T})$ is a well-posed ADP. In this subsection we define optimality for ADPs, connect it to the Bellman equation, and state the fundamental optimality properties. We then give sufficient conditions for these properties to hold.

We say that a policy $\sigma \in \Sigma$ is **optimal** for $(V, \mathbb{T})$ if v_{σ} is a greatest element of V_{Σ} . In other words, σ is optimal if it attains the “highest possible” lifetime value.

Perhaps the most important aspect of the theory of dynamic programming is the link between optimality and the Bellman equation. To clarify this link we introduce some terminology and set

$$v^* := \bigvee_{\sigma} v_{\sigma} = \bigvee V_{\Sigma} \quad \text{whenever the supremum exists.}$$

When v^* exists (i.e., when the supremum exists) we call v^* the **value function** of the ADP. The following statements are obvious from the definitions:

- Existence of an optimal policy σ implies that v^* exists and is equal to v_{σ} .
- If $v^* = v_{\sigma}$ for some $\sigma \in \Sigma$, then σ is optimal.

At the same time, existence of v^* does not generally imply existence of a greatest element (and hence an optimal policy).

Definition 2.1. We say that **Bellman's principle of optimality holds** if

$$\{\sigma \in \Sigma : \sigma \text{ is optimal}\} = \{\sigma \in \Sigma : \sigma \text{ is } v^* \text{-greedy}\}. \quad (2)$$

(When v^* does not exist both sets are understood as empty.)

Definition 2.2. We say that the **fundamental optimality properties hold** if

- (B1) at least one optimal policy exists,
- (B2) v^* exists and is the unique solution to the Bellman equation in V_G , and
- (B3) Bellman's principle of optimality holds.

It is important to emphasize here that (B1)–(B3) are not independent. Indeed, (B2) implies both (B1) and (B3), as shown in the next proposition.

Proposition 2.3. *The fundamental optimality properties hold if and only if v^* exists and is the unique solution to the Bellman equation in V_G .*

Proof. Regarding $(\Leftarrow)$, suppose that v^* exists and is the unique solution to the Bellman equation in V_G . (B2) holds by the hypothesis. By Lemma 4.2 of [Sargent and Stachurski \[2025b\]](#), an optimal policy exists and Bellman's principle of optimality holds. This gives (B1) and (B3) respectively. The claim $(\Rightarrow)$ is trivial, so the proof of Proposition 2.3 is complete. $\square$

Despite the fact that (B2) implies both (B1) and (B3), we have stated them together because, in terms of applications, all three parts are significant. (B1) tells us that the problem at hand has a solution. (B3) implies that a solution can be computed by taking a v^* -greedy policy, and that any other optimal policy must also be v^* -greedy. Finally, (B2) provides a restriction that can help us calculate v^* . Provided that we search in V_G , any fixed point of T is equal to v^* .

Proposition 2.3 tells us our main goal is to construct conditions under which v^* exists and is the unique fixed point of T in V_G . We begin this task in Section 3.

2.4 Algorithms

In this section we review the three core algorithms for solving dynamic programs: value function iteration (VFI), Howard policy iteration (HPI), and optimistic policy iteration (OPI). For background on the role and significance of these algorithms, see Bertsekas [2022], Sargent and Stachurski [2025a], or Sargent and Stachurski [2025b].

Let $(V, \mathbb{T})$ be a regular and well-posed ADP. As in Sargent and Stachurski [2025b], we define the **Howard policy operator** corresponding to $(V, \mathbb{T})$ via

$$H: V_G \rightarrow V_\Sigma, \quad Hv = v_\sigma \quad \text{where } \sigma \text{ is } v\text{-greedy.}$$

For each $m \in \mathbb{N}$, we define the **optimistic policy operator**

$$W: V_G \rightarrow V, \quad Wv = T_\sigma^m v \quad \text{where } \sigma \text{ is } v\text{-greedy.} \quad (3)$$

Here and below, the dependence of W on m is often suppressed to simplify notation.

When $(V, \mathbb{T})$ is regular, $V_G = V$, so the map W is well-defined on all of V (just like the Bellman operator T). The Howard policy operator H is well-defined on all of V when, in addition, $(V, \mathbb{T})$ is well-posed. Note that, for both of these maps, we always select the same v -greedy policy when applying them to some fixed v .²

As in Sargent and Stachurski [2025b], we associate VFI, OPI, and HPI with fixing a $v \in V$ and then iteratively applying the operators T , W , and H respectively.

Suppose that $(V, \mathbb{T})$ is a regular ADP with Bellman operator T and the fundamental optimality properties hold. Let v^* denote the value function. We set

- $V_U := \text{all } v \in V \text{ with } v \preceq Tv$.

In this setting, we say that

- **VFI converges** if $T^n v \uparrow v^*$ for all $v \in V_U$,
- **OPI converges** if $W^n v \uparrow v^*$ for all $v \in V_U$, and
- **HPI converges** if $H^n v \uparrow v^*$ for all $v \in V_U$.

For OPI convergence, the meaning is that convergence holds for any choice of the OPI step size $m \in \mathbb{N}$. The following lemma will be useful for our analysis:

²In general, designating a v -greedy policy for all $v \in V$ requires the Axiom of Choice. In practice, many applications induce some structure on the policy set that can be used to produce simple selection mechanisms.

Lemma 2.4. *If $(V, \mathbb{T})$ is regular and order stable, then*

- (i) *$v \preceq v_\sigma$ for every $v \in V_U$ and every v -greedy $\sigma \in \Sigma$. In particular, if v^* exists, then $v \preceq v^*$ for all $v \in V_U$.*
- (ii) *convergence of VFI implies convergence of OPI and HPI.*

Proof. Let $(V, \mathbb{T})$ be as stated, fix $v \in V_U$, and let σ be any v -greedy policy (which exists by regularity). Then $Tv = T_\sigma v$, and since $v \in V_U$ we have $v \preceq T_\sigma v$. This inequality and order stability yield $v \preceq v_\sigma$, where v_σ is the fixed point of T_σ . If v^* exists, then $v_\sigma \preceq v^*$ and hence $v \preceq v^*$. Part (ii) is a direct consequence of Corollary 6.3 of [Sargent and Stachurski \[2025b\]](#), on noting that every order stable operator is upward stable in the sense of that paper. $\square$

3 Optimality Results

In this section we study ADPs in settings where the value space is not only partially ordered but also a pospace. We then consider optimality in settings where the policy operators have topological stability properties, as well as order properties. *Throughout this section, V is always assumed to be a pospace.*

In Section 3.1 we study ADPs on pospace, leveraging global stability to obtain optimality and convergence results. In Section 3.2 we specialize to partially ordered metric spaces, where contraction-based arguments apply. Section 3.3 treats nonstationary policies, and Section 3.4 provides minimization counterparts.

3.1 ADPs on Pospace

Let V be a pospace and let $(V, \mathbb{T})$ be an ADP. We say that

$(V, \mathbb{T})$ is **globally stable** if each $T_\sigma \in \mathbb{T}$ is globally stable on V .

Obviously, if $(V, \mathbb{T})$ is globally stable, then $(V, \mathbb{T})$ is well-posed. The next result shows that regularity and global stability together yield strong optimality properties.

Theorem 3.1. *Let $(V, \mathbb{T})$ be regular and globally stable. If T has a fixed point in V , then*

- (i) *the fundamental optimality properties hold and*
- (ii) *VFI, OPI and HPI all converge.*

Proof. Let $(V, \mathbb{T})$ and T be as stated. Since $(V, \mathbb{T})$ is order stable (Lemma 2.2), part (i) follows from Theorem 5.1 of Sargent and Stachurski [2025b]. Regarding convergence of VFI, fix $v \in V_U$, let σ be an optimal policy and let v^* be the value function. We have $T_\sigma v \preceq Tv \preceq v^*$, where the last inequality is by Lemma 2.4 (since, by that lemma, $v \preceq v^*$ and, therefore, $Tv \preceq Tv^* = v^*$). From this chain of inequalities, combined with the fact that $T_\sigma \preceq T$ on V , we obtain $T_\sigma^n v \preceq T^n v \preceq v^*$ for all n . As σ is optimal and $(V, \mathbb{T})$ is globally stable, the policy operator T_σ is globally stable with unique fixed point v^* . Because $T_\sigma^n v \preceq v^*$ for all n and $T_\sigma^n v \rightarrow v^*$, Lemma 2.1 implies that the supremum of $T_\sigma^n v$ is v^* . From this fact and $T_\sigma^n v \preceq T^n v \preceq v^*$ for all n , the supremum of $T^n v$ is also v^* . Moreover, this sequence is increasing (because $v \in V_U$), which leads us to $T^n v \uparrow v^*$. Hence VFI converges. Since $(V, \mathbb{T})$ is regular and order stable (by Lemma 2.2), convergence of VFI implies convergence of OPI and HPI (Lemma 2.4). $\square$

Theorem 3.1 is a high-level result because a condition (existence of a fixed point) is placed on the derived object T , rather than the primitives V and $\mathbb{T}$. Below, we present a sequence of results that leverage Theorem 3.1 while also placing all assumptions on primitives.

Corollary 3.2. *Let $(V, \mathbb{T})$ be regular and globally stable. If $(V, \mathbb{T})$ is also finite, then*

- (i) *the fundamental optimality properties hold and*
- (ii) *VFI, OPI and HPI all converge.*

Proof. Since $(V, \mathbb{T})$ is globally stable, $(V, \mathbb{T})$ is also order stable (Lemma 2.2). Since it is also regular and finite, Theorem 5.4 of Sargent and Stachurski [2025b] implies that the fundamental optimality properties hold. In particular, T has a fixed point in V , so all the conclusions of Corollary 3.2 hold. $\square$

The next theorem uses countable Dedekind completeness and order boundedness, combined with global stability.

Theorem 3.3. *Let $(V, \mathbb{T})$ be regular and globally stable. If, in addition, $(V, \mathbb{T})$ is order bounded and V is countably Dedekind complete, then*

- (i) *the fundamental optimality properties hold and*
- (ii) *VFI, OPI and HPI all converge.*

Proof. In view of Theorem 3.1, we need only show that T has a fixed point in V . To see this, we use the order bounded property to take a $u \in V$ such that $T_\sigma u \preceq u$ for all $\sigma \in \Sigma$. Now let v be any element of V_Σ . Global stability implies order stability

(Lemma 2.2), so we have $v \preceq u$. Moreover, $v \in V_U$: by regularity, $v \in V_G$, so taking σ with $T_\sigma v = v$ gives $v = T_\sigma v \preceq T v$. Hence T is a self-map on the order interval $[v, u]$. Letting $v_n := T^n v$ and applying countably Dedekind completeness, we have $v_n \uparrow \bar{v}$ for some $\bar{v} \in [v, u]$. We claim that $T\bar{v} = \bar{v}$. To see this, first observe that $v_{n+1} = T v_n \preceq T\bar{v}$ for all n , so $\bar{v} \preceq T\bar{v}$. Letting σ be $\bar{v}$ -greedy, we have $\bar{v} \preceq T\bar{v} = T_\sigma \bar{v}$, so, by order stability, $\bar{v} \preceq v_\sigma$. Also, letting $w_n := T_\sigma^n v$, induction on n gives $w_n \preceq v_n$ for all n , since $w_n = T_\sigma w_{n-1} \preceq T_\sigma v_{n-1} \preceq T v_{n-1} = v_n$ by order preservation of T_σ and $T_\sigma \preceq T$. Combined with $v_n \preceq \bar{v} \preceq v_\sigma$, we have $w_n \preceq v_n \preceq \bar{v} \preceq v_\sigma$ for all $n \in \mathbb{N}$. By global stability, $w_n \rightarrow v_\sigma$. Applying Lemma 2.1, we also have $\bigvee_n w_n = v_\sigma$. From this fact and the previous chain of inequalities, it must be that $\bar{v} = v_\sigma$. Hence $T\bar{v} = T_\sigma \bar{v} = T_\sigma v_\sigma = v_\sigma = \bar{v}$. This completes the proof that $\bar{v}$ is a fixed point of T . $\square$

3.2 ADPs on Metric Space

In this section we add more structure by assuming that our value space V is in fact a partially ordered metric space, as defined in Section 2.1. By construction, this specializes the earlier setting of Section 3.1, where $(V, \preceq)$ is just a pospace.

Throughout Section 3.2,

- $(V, \mathbb{T})$ is an ADP and
- $V = (V, \preceq, d)$ is a partially ordered metric space.

Given a closed subset V_0 of V , we will say that the ADP $(V, \mathbb{T})$ is **V_0 -semi-regular** if $V_0 \subset V_G$ and $TV_0 \subset V_0$. In addition, we will say that VFI is **geometrically convergent on V_0** if v^* exists and there is a $\beta \in (0, 1)$ such that

$$d(T^n v, v^*) = \mathcal{O}(\beta^n) \text{ as } n \rightarrow \infty \text{ for each } v \in V_0.$$

Here $f(n) = \mathcal{O}(\beta^n)$ means that there exists a $C < \infty$ with $f(n) \leq C\beta^n$ for all $n \in \mathbb{N}$.

In stating the next theorem, we use the notion of sup-nonexpansiveness from Section 2.1.

Theorem 3.4. *Let $(V, \mathbb{T})$ be V_0 -semi-regular for some closed $V_0 \subset V$, and let d be complete and sup-nonexpansive. If each $T_\sigma \in \mathbb{T}$ is a contraction of modulus β on V , then*

- (i) *the fundamental optimality properties hold,*
- (ii) *the value function v^* lies in V_0 , and*
- (iii) *VFI is geometrically convergent on V_0 .*

If, in addition, $(V, \mathbb{T})$ is regular, then OPI and HPI also converge.

To prove Theorem 3.4, we use the following elementary lemma. In the statement of the lemma, $(V, \mathbb{T})$ is an ADP, V_0 is a subset of V and $Tv := \bigvee_{\sigma} T_{\sigma} v$ exists at each $v \in V_0$.

Lemma 3.5. *Let d be sup-nonexpansive. If $TV_0 \subset V_0$ and each $T_{\sigma} \in \mathbb{T}$ is a contraction of modulus β on V , then T is a contraction of modulus β on V_0 .*

Proof. Let the stated conditions hold, so that T is a self-map on V_0 . For any $v, w \in V_0$, we have

$$d(Tv, Tw) = d\left(\bigvee_{\sigma} T_{\sigma} v, \bigvee_{\sigma} T_{\sigma} w\right) \leq \bigvee_{\sigma} d(T_{\sigma} v, T_{\sigma} w) \leq \beta d(v, w).$$

This shows that T is a contraction of modulus β on V_0 . $\square$

Proof of Theorem 3.4. Let V and $\mathbb{T}$ have the stated properties. By completeness and Banach's fixed point theorem, the ADP $(V, \mathbb{T})$ is globally stable and hence order stable (Lemma 2.2). Moreover, each T_{σ} is a contraction of modulus β on V and T is a well-defined self-map on V_0 . Lemma 3.5 now implies that T is a contraction of modulus β on V_0 . As V_0 is closed and V is complete, this implies that T has a fixed point in $V_0 \subset V_G$. Hence, by Theorem 5.1 of Sargent and Stachurski [2025b], the fundamental optimality properties hold. It follows that v^* exists and is the unique fixed point of T in V_G . Since T has a fixed point in V_0 , this also implies that $v^* \in V_0$. Geometric convergence of VFI now follows from contractivity of T on V_0 .

If $(V, \mathbb{T})$ is also regular, then convergence of OPI and HPI follows from Theorem 3.1. $\square$

3.3 Nonstationary Policies

Consider an ADP $(V, \mathbb{T})$ where $V = (V, \preceq, d)$ is a partially ordered metric space. In all of the preceding discussion, when discussing optimality of this ADP, we focused on stationary policies, assuming that a fixed σ is applied at each point in time. But is this focus on stationary policies justified? Could it be that higher lifetime value is available when we allow a change of policy in each period?

To address this question, suppose that we can select a **policy plan** $\bar{\sigma} := (\sigma_t)_{t \geq 0}$ in the infinite Cartesian product $\prod_{t \geq 0} \Sigma$ and apply the t -th element σ_t at time t . The lifetime value of $\bar{\sigma}$ can be defined by

$$v_{\bar{\sigma}} = \lim_{n \rightarrow \infty} T_{\sigma_0} T_{\sigma_1} \cdots T_{\sigma_n} v. \quad (4)$$

To ensure that (4) exists and is independent of v , we require the following:

Assumption 3.1. The metric d is complete and sup-nonexpansive. In addition, there exists a positive constant λ with $\lambda < 1$ and

$$d(T_\sigma v, T_\sigma w) \leq \lambda d(v, w) \quad \text{for all } v, w \in V \text{ and all } \sigma \in \Sigma.$$

In addition, for all $v \in V$ we have $\sup_{\sigma \in \Sigma} d(v, T_\sigma v) < \infty$.

We can now state the main result of this section, which shows that, under the stated assumptions, any policy plan is (weakly) dominated in value by a stationary policy.

Theorem 3.6. *If $(V, \mathbb{T})$ is regular and Assumption 3.1 holds, then the fundamental optimality properties hold. In addition, given any policy plan $\bar{\sigma}$, there exists a stationary policy $\sigma \in \Sigma$ such that $v_{\bar{\sigma}} \preceq v_\sigma$.*

The proof of Theorem 3.6 is provided in Section 6.2 of the appendix.

3.4 Minimization

Now let's turn to the minimization case. In the definitions we simply replace suprema with infima. Then we restate Theorem 3.1, Theorem 3.3 and Theorem 3.4 for minimization problems. These minimization results are obtained from the maximization results using order duality. The arguments are relatively mechanical, so all proofs have been deferred to Section 6.3 in the appendix.

Let $(V, \mathbb{T})$ be an ADP with policy set Σ . We define the **Bellman min-operator** T corresponding to $(V, \mathbb{T})$ by

$$Tv = \bigwedge_{\sigma} T_\sigma v \quad \text{whenever the infimum exists.} \tag{5}$$

We say that $v \in V$ satisfies the **Bellman min-equation** if $Tv = v$.

Remark 3.1. Note the abuse of notation here: T is used both for the original Bellman max-operator from Section 2.2 and the new Bellman min-operator in (5). Below, when introducing min-versions of V_G , v^* , H , and W , we will again recycle the symbols, also shadowing the max case. We follow this convention to avoid tedious and complex notation throughout. In the theorem statements and applications below, we will always be considering one case or another (max or min) and no confusion will result. In Section 6.3, when we provide proofs that involve max and min arguments together, we will use more precise notation.

Paralleling the maximization terminology, we say that

$\sigma \in \Sigma$ is **ν -min-greedy** if $T_\sigma \nu \preceq T_\tau \nu$ for all $\tau \in \Sigma$.

Analogous to the max case, σ is ν -min-greedy if and only if $T_\sigma \nu = T\nu$. In addition, we say that $(V, \mathbb{T})$ is

- **min-order bounded** if there exists a $b \in V$ with $b \preceq T_\sigma b$ for all $T_\sigma \in \mathbb{T}$, and
- **min-regular** if, for each $\nu \in V$, at least one ν -min-greedy policy exists.

We let

$$V_G = \{\nu \in V : \text{at least one } \nu\text{-min-greedy policy exists}\}.$$

Now suppose $(V, \mathbb{T})$ is well-posed and let V_Σ be defined as before (i.e., the set of σ -value functions). In this setting we set $\nu^* = \bigwedge_\sigma \nu_\sigma$ and call it the **min-value function** whenever the infimum exists. Also, we say that

- $\sigma \in \Sigma$ is **min-optimal** for $(V, \mathbb{T})$ if $\nu_\sigma = \nu^*$ (equivalently, if ν_σ is a least element of V_Σ), and
- $(V, \mathbb{T})$ obeys **Bellman's principle of min-optimality** if

$$\sigma \in \Sigma \text{ is min-optimal for } (V, \mathbb{T}) \iff \sigma \text{ is } \nu^*\text{-min-greedy}.$$

When $(V, \mathbb{T})$ is min-regular and well-posed, we define the **Howard policy min-operator** corresponding to $(V, \mathbb{T})$ via

$$H: V_G \rightarrow V_\Sigma, \quad Hv = \nu_\sigma \quad \text{where } \sigma \text{ is } \nu\text{-min-greedy},$$

as well as, for each $m \in \mathbb{N}$, the **optimistic policy min-operator** via

$$W: V_G \rightarrow V, \quad W\nu := T_\sigma^m \nu \quad \text{where } \sigma \text{ is } \nu\text{-min-greedy}. \quad (6)$$

We say that the **fundamental min-optimality properties hold** if

- at least one min-optimal policy exists,
- ν^* exists and is the unique solution to the Bellman min-equation in V_G , and
- Bellman's principle of min-optimality holds.

Let V_D be all $\nu \in V$ with $T\nu \preceq \nu$. We say that

- **min-VFI converges** if $T^n \nu \downarrow \nu^*$ for all $\nu \in V_D$,

- **min-OPI converges** if $W^n v \downarrow v^*$ for all $v \in V_D$ and all $m \in \mathbb{N}$, and
- **min-HPI converges** if $H^n v \downarrow v^*$ for all $v \in V_D$.

We can now state min-versions of two of our previous max-based optimality results. Our first is a min-version of Theorem 3.1.

Theorem 3.7. *Let $(V, \mathbb{T})$ be min-regular and globally stable. Let T be the Bellman min-operator for $(V, \mathbb{T})$. If T has a fixed point in V , then*

- (i) *the fundamental min-optimality properties hold and*
- (ii) *min-VFI, min-OPI and min-HPI all converge.*

Next we state a min-version of Theorem 3.3.

Theorem 3.8. *Let $(V, \mathbb{T})$ be min-regular and globally stable. If, in addition, $(V, \mathbb{T})$ is min-order bounded and V is countably Dedekind complete, then*

- (i) *the fundamental min-optimality properties hold and*
- (ii) *min-VFI, min-OPI and min-HPI all converge.*

A min-version of Theorem 3.4 can also be stated, but we omit it for brevity.

Below we show several applications of these results.

4 Applications

In this section we apply the preceding results to the four applications described in Section 1: weak Feller MDPs, structural estimation, no-discount optimal stopping, and Bayesian sequential analysis.

4.1 Contractive Problems

This subsection presents two contractive applications to demonstrate that our theory captures these cases. The first is very standard, involving MDPs with constant discount factors and useful continuity properties. The second is contractive but nonstandard. It considers the optimization component (but not the estimation component) of a structural estimation problem, where the maximization operator is shifted inside the expectation operator.

4.1.1 Weak Feller MDPs

We begin with a simple example involving MDPs (see, e.g., [Hernández-Lerma and Lasserre \[2012\]](#), [Bäuerle and Rieder \[2011\]](#)). We maximize $\mathbb{E} \sum_{t \geq 0} \beta^t r(X_t, A_t)$ where X_t takes values in state space X , A_t takes values in action space A , r is a reward function defined on $X \times A$ and $\beta \in (0, 1)$ is a discount factor. Here we consider a general case where X is a separable metric space with Borel sets $\mathcal{B}$, and A is a separable metric space. The set of feasible actions at state x is denoted by $\Gamma(x)$, where Γ is a nonempty correspondence from X to A . The set $G := \{(x, a) \in X \times A : a \in \Gamma(x)\}$ denotes the feasible state-action pairs. Let $P: G \times \mathcal{B} \rightarrow [0, 1]$ provide transition probabilities for the next period states given current state and action. The Bellman equation is

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \int v(x') P(x, a, dx') \right\} \quad (x \in X). \quad (7)$$

A feasible policy is a Borel measurable map $\sigma: X \rightarrow A$ such that $\sigma(x) \in \Gamma(x)$ for all $x \in X$. Let Σ be the set of all feasible policies. We combine bX (the set of all bounded and Borel measurable real-valued functions on X) with the pointwise partial order $\leq$ and the supremum norm. Since the pointwise partial order is closed with respect to the topology of the supremum norm, bX is a partially ordered metric space in the sense of Section 2.1. (The same is true for the other value spaces used in the remainder of Section 4, all of which combine a set of bounded measurable functions with the pointwise partial order and the supremum norm.) For $\sigma \in \Sigma$ and $v \in bX$, we define the MDP policy operator

$$(T_\sigma v)(x) = r(x, \sigma(x)) + \beta \int v(x') P(x, \sigma(x), dx') \quad (x \in X).$$

We impose some regularity conditions.

Assumption 4.1. The following conditions hold:

- (i) The reward function $r: X \times A \rightarrow \mathbb{R}$ is bounded and continuous on G ,
- (ii) The feasible correspondence $\Gamma: X \rightarrow A$ is compact-valued and continuous, and
- (iii) The transition kernel P is weak Feller, i.e., the map

$$(x, a) \mapsto \int h(x') P(x, a, dx')$$

is bounded and continuous whenever function h is bounded and continuous on X .

With $\mathbb{T}_M := \{T_\sigma : \sigma \in \Sigma\}$, the pair $(bX, \mathbb{T}_M)$ is an ADP, since each T_σ is an order preserving self-map on bX . If v_σ is the (unique) fixed point of T_σ , then $v_\sigma(x)$ represents the lifetime value of using policy σ in every period, conditional on starting in state x (see [Sargent and Stachurski \[2025a\]](#), Section 3.1, or [Puterman \[2005\]](#)). Moreover, v satisfies the ADP Bellman equation $Tv = v$ if and only if the traditional Bellman equation (7) holds (see, e.g., Example 3.1 of [Sargent and Stachurski \[2025b\]](#)).

Let bcX be the subset of bX consisting of all continuous functions in bX . We can now state the following result.

Theorem 4.1. *If Assumption 4.1 holds, then for the ADP $(bX, \mathbb{T}_M)$,*

- (i) *the fundamental optimality properties hold,*
- (ii) *the value function v^* lies in bcX , and*
- (iii) *VFI is geometrically convergent on bcX .*

Proof. It is trivial that $bcX \subset bX$ is closed, the supremum norm is complete and sup-nonexpansive on bX , and each T_σ is a contraction of modulus β on bX . In view of Theorem 3.4, we only need to show that $(bX, \mathbb{T}_M)$ is bcX -semi-regular. Fix $v \in bcX$. By (i) and (iii) of Assumption 4.1, the map

$$(x, a) \mapsto r(x, a) + \beta \int v(x')P(x, a, dx')$$

is continuous, so, by (ii) of Assumption 4.1 and the Berge maximum theorem (see Theorem 17.31 of [Aliprantis and Border \[2006\]](#)), the function Tv given by

$$(Tv)(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \int v(x')P(x, a, dx') \right\}$$

is well-defined and continuous on X . Hence $T(bcX) \subset bcX$. By the measurable maximum theorem (see Theorem 18.19 of [Aliprantis and Border \[2006\]](#)), there exists a measurable function $\sigma: X \rightarrow A$ such that

$$\sigma(x) \in \operatorname{argmax}_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \int v(x')P(x, a, dx') \right\} \text{ for all } x \in X.$$

For this σ and any other $\tau \in \Sigma$, we have $(T_\tau v)(x) \leq (T_\sigma v)(x)$ for all $x \in X$. In particular, σ is v -greedy. $\square$

4.1.2 Structural Estimation

As discussed in Section 1, some dynamic programs reverse the order of maximization and expectation, placing them outside both standard and generalized contractive frameworks. At the same time, they are easily expressed as ADPs in the sense defined above, and our theory is straightforward to apply.

The set up we consider here involves the optimization step of a structural estimation problem for discrete choice models (see, e.g., Rust [1987, 1994], Kristensen et al. [2021] and Example 3.6 of Sargent and Stachurski [2025b]). The Bellman equations take the form

$$g(x, a) = \int \int \left\{ \max_{a' \in A} [r(x', a', e') + \beta g(x', a')] \right\} \nu(de') P(x, a, dx'). \quad (8)$$

The setting is similar to Section 4.1.1. Here the action space A is finite (hence discrete choice), and the component e in the reward function takes values in some measurable space E and is IID with distribution ν . Since the current shock is observed before the current action is chosen, policies are permitted to depend on it. Hence we take Σ to be the set of measurable maps from $X \times E$ to A and, for each $\sigma \in \Sigma$, we introduce the policy operator

$$(T_\sigma g)(x, a) = \int \int [r(x', \sigma(x', e'), e') + \beta g(x', \sigma(x', e'))] \nu(de') P(x, a, dx').$$

Assume that the reward function r is bounded and measurable. We take the value space to be bG , the set of bounded and Borel measurable real-valued functions on G , paired with the pointwise partial order and the supremum norm. (Every action is feasible in the present setting, so the set of feasible state-action pairs is $G = X \times A$.)

Lemma 4.2. *The pair $(bG, \mathbb{T}_\Sigma)$ is a regular ADP.*

Proof. The arithmetic and integral operations in the definition of T_σ preserve boundedness and measurability, so each T_σ is a self-map on bG . Clearly, T_σ is order preserving. Hence, with $\mathbb{T}_\Sigma := \{T_\sigma : \sigma \in \Sigma\}$, the pair $(bG, \mathbb{T}_\Sigma)$ is an ADP.

Regarding regularity, fix $g \in bG$. Let

$$\hat{r}(x, e, a) := r(x, a, e) + \beta g(x, a).$$

Clearly, the function $\hat{r}$ is bounded and measurable. Since A is finite, $\hat{r}$ is also continuous on A . By the measurable maximum theorem (see Theorem 18.19 of Aliprantis and Border [2006]), there exists a measurable function $\sigma : X \times E \rightarrow A$ such that $\sigma(x, e) \in$

$\operatorname{argmax}_{a \in A} \hat{r}(x, e, a)$ for all $(x, e) \in X \times E$. For this σ and any other $\tau \in \Sigma$, we have $(T_\tau g)(x, a) \leq (T_\sigma g)(x, a)$ for each $(x, a) \in G$. In particular, σ is g -greedy. Hence, $(bG, \mathbb{T}_S)$ is regular. $\square$

When σ is the g -greedy policy constructed in the proof of Lemma 4.2, the maximization in the definition of σ yields

$$(Tg)(x, a) = (T_\sigma g)(x, a) = \int \int \left\{ \max_{a' \in A} [r(x', a', e') + \beta g(x', a')] \right\} \nu(de') P(x, a, dx').$$

In particular, $g \in bG$ satisfies the ADP Bellman equation $Tg = g$ if and only if g solves the Bellman equation (8).

We can now state the following result.

Theorem 4.3. *For the ADP $(bG, \mathbb{T}_S)$, the fundamental optimality properties hold, VFI is geometrically convergent on bG , and OPI and HPI also converge.*

Proof. By Lemma 4.2, $(bG, \mathbb{T}_S)$ is a regular ADP. It is trivial that the supremum norm is complete and sup-nonexpansive, and each T_σ is a contraction of modulus β . All claims follow from Theorem 3.4. $\square$

4.2 No-Discount Optimal Stopping

Many important applications—including sampling problems, shortest path and routing problems, bandit problems, and reinforcement learning tasks—involve no discounting. In the absence of a discount factor, contractivity of the policy operators typically fails. In this section we study one example of a non-contractive model: optimal stopping without discounting. (For background and connections to the related literature, see the introduction.) Later, in Section 4.3, we apply the optimal stopping results obtained below to a classical problem in sequential analysis.

4.2.1 Setup

Let $(X, \mathcal{B})$ be a measurable space and let P be a stochastic kernel on $(X, \mathcal{B})$. In what follows, a discrete time X -valued stochastic process $(X_t)_{t \geq 0}$ on probability space $(\Omega, \mathcal{F}, \mathbb{P}_x)$ will be called **P -Markov** if

$$\mathbb{P}\{X_{t+1} \in B \mid \mathcal{F}_t\} = P(X_t, B) \quad \mathbb{P}\text{-a.s. for all } t \geq 0 \text{ and } B \in \mathcal{B}.$$

We write $\mathbb{P}_x$ and $\mathbb{E}_x$ for probabilities and expectations when conditioning on $X_0 = x$. We will also use the time-shift form of the Markov property: for any $\mathcal{B}^{\otimes \mathbb{N}}$ -measurable random variable H depending on the trajectory $(X_0, X_1, \dots)$ and any $t \geq 0$,

$$\mathbb{E}_x [H \circ \theta^t \mid \mathcal{F}_t] = \mathbb{E}_{X_t} [H], \quad (9)$$

where $\theta: (x_0, x_1, \dots) \mapsto (x_1, x_2, \dots)$ is the shift operator on $X^{\mathbb{N}}$ (see, e.g., Chapter 3 of [Meyn and Tweedie \[2009\]](#)).

Let bX denote the set of bounded $\mathcal{B}$ -measurable functions from X to $\mathbb{R}$, and let bX_+ be all $g \in bX$ taking only nonnegative values. We consider a cost minimization problem with Bellman equation

$$g(x) = \min \left\{ e(x), c(x) + \int g(x') P(x, dx') \right\}, \quad (10)$$

where e, c are functions in bX_+ , while P is a stochastic kernel on X . The function e is called the **exit cost function** and c is called the **flow cost function**. The Bellman equation corresponds to a setting where a controller observes a P -Markov state process $(X_t)_{t \geq 0}$ and decides when to stop. Stopping at time t incurs the one-off penalty $e(X_t)$. Continuing incurs the flow cost $c(X_t)$, followed by transition to the new state X_{t+1} and the opportunity to decide again.

4.2.2 Policies

A policy is a $\mathcal{B}$ -measurable map σ from X to $\{0, 1\}$, with $\sigma(x) = 1$ indicating the decision to stop in state x . Given a policy σ , we call

$$E_\sigma := \{x \in X : \sigma(x) = 1\}$$

the **exit region** for σ . We call its complement E_σ^c the **continuation region**.

To each policy σ , we associate the **stopping time**

$$\tau^\sigma := \inf\{t \geq 0 : X_t \in E_\sigma\} = \inf\{t \geq 0 : \sigma(X_t) = 1\}. \quad (11)$$

Here and below, the convention for the infimum is that $\inf \emptyset := \infty$. Also, given σ , we define the **σ -loss function** via

$$g_\sigma(x) := \mathbb{E}_x \left[\sum_{t=0}^{\tau^\sigma-1} c(X_t) + e(X_{\tau^\sigma}) \right]. \quad (12)$$

Here $\sum_{t=0}^{-1} c(X_t)$ is understood as 0, so that $g_\sigma(x) = e(x)$ when $x \in E_\sigma$. The function g_σ takes values in $[0, \infty]$ and $g_\sigma(x)$ represents the total expected cost when applying σ in every period, conditional on starting in state x .

4.2.3 The Lower Bound Policy

One policy of particular interest is $\bar{\sigma} = \mathbb{1}\{e \leq c\}$. To simplify notation, the exit region for this policy is denoted

$$\bar{E} := E_{\bar{\sigma}} = \{x \in X : e(x) \leq c(x)\} \quad (13)$$

and the stopping time is denoted

$$\bar{\tau} := \tau^{\bar{\sigma}} = \inf\{t \geq 0 : \bar{\sigma}(X_t) = 1\}. \quad (14)$$

We call $\bar{E}$ the **certain exit region**. Under the innocuous convention that the controller always stops when indifferent between stopping and continuing, any optimal policy will choose to stop when $x \in \bar{E}$. The reason is that the controller has the opportunity to exit at cost $e(x) \leq c(x)$, and continuing incurs $c(x)$ plus additional costs in subsequent stages.

The fact that the controller always stops when $x \in \bar{E}$ allows us to shrink the policy space. Specifically, we consider only policies where $\sigma(x) = 1$ for all $x \in \bar{E}$. Let Σ be the set of all such policies. We can also express this set via

$$\Sigma := \{\text{all } \mathcal{B}\text{-measurable } \sigma: X \rightarrow \{0, 1\} \text{ with } \bar{\sigma} \leq \sigma\}. \quad (15)$$

Since $\bar{\sigma} \leq \sigma$ for all $\sigma \in \Sigma$, we refer to $\bar{\sigma}$ as the **lower bound policy**. Also, since

$$\bar{\sigma} \leq \sigma \implies \tau^\sigma \leq \bar{\tau} \text{ } \mathbb{P}\text{-a.s.}, \quad (16)$$

we refer to $\bar{\tau}$ as the **upper bound stopping time**.

Our key assumption is as follows.

Assumption 4.2. The upper bound stopping time $\bar{\tau}$ obeys

$$\sup_{x \in X} \mathbb{E}_x \bar{\tau} < \infty.$$

For Assumption 4.2, it suffices to check that $\sup_{x \in \bar{E}^c} \mathbb{E}_x \bar{\tau}$ is finite, since $\bar{\tau} = 0$ with probability one when $x \in \bar{E}$. Below we show that Assumption 4.2 is sufficient for all of the major optimality results associated with dynamic programming.

Lemma 4.4. *When Assumption 4.2 holds, the σ -loss function (12) is finite and bounded on X for all $\sigma \in \Sigma$.*

Proof. Fix $\sigma \in \Sigma$. By Assumption 4.2, there exists a constant $M < \infty$ with $\mathbb{E}_x \bar{\tau} \leq M$ for all $x \in X$. In addition, the functions c and e are bounded, so we can take constants N_c, N_e with $c \leq N_c$ and $e \leq N_e$. As a result,

$$g_\sigma(x) \leq N_c \mathbb{E}_x \tau^\sigma + N_e \leq N_c \mathbb{E}_x \bar{\tau} + N_e \leq N_c M + N_e,$$

where, in the second inequality, $\mathbb{E}_x \tau^\sigma \leq \mathbb{E}_x \bar{\tau}$ by (16). $\square$

4.2.4 Policy Operators

For each $\sigma \in \Sigma$, we define a **policy operator** T_σ via

$$(T_\sigma g)(x) = \sigma(x)e(x) + (1 - \sigma(x)) \left[c(x) + \int g(x')P(x, dx') \right], \quad (17)$$

or, in operator notation, as

$$T_\sigma g = \sigma e + (1 - \sigma)c + K_\sigma g \quad (g \in bX), \quad (18)$$

where

$$(K_\sigma g)(x) := (1 - \sigma(x)) \int g(x')P(x, dx') \quad (x \in X). \quad (19)$$

In (18), expressions such as σe are understood as pointwise products.

As usual, the policy operator associated with σ is introduced with the idea that its fixed point gives the lifetime value – in this case, the lifetime cost – generated by σ . This turns out to be true here as well, although the proof is not entirely trivial. It requires some familiarity with shift operators and the Markov property in its general form. An introduction to these topics can be found in Chapter 3 of [Meyn and Tweedie \[2009\]](#).

Lemma 4.5. *For every $\sigma \in \Sigma$, the σ -loss function g_σ is a fixed point of T_σ .*

Proof. Fix $\sigma \in \Sigma$. To simplify notation, for the duration of this proof we set $\tau := \tau^\sigma$ and $E := E_\sigma$. For $x \in E$, we have $\sigma(x) = 1$ and hence $(T_\sigma g_\sigma)(x) = e(x)$. At the same time, (12) implies that $g_\sigma(x) = e(x)$ also holds for such x . In particular, $T_\sigma g_\sigma = g_\sigma$ on E . Hence, to complete the proof, we only need to show that $T_\sigma g_\sigma = g_\sigma$ on E^c .

To this end, fix $x \notin E$ and define the random variable

$$H := \sum_{t=0}^{\tau-1} c(X_t) + e(X_\tau),$$

so that

$$g_\sigma(x) = \mathbb{E}_x H \quad \text{and} \quad g_\sigma(X_1) = \mathbb{E}_{X_1} H = \mathbb{E}_x [H \circ \theta \mid X_1].$$

On the right-hand side, $\theta: (x_0, x_1, \dots) \mapsto (x_1, x_2, \dots)$ is the shift operator on the sequence space $X^{\mathbb{N}}$, and the last equality is by the Markov property (9). We can write $g_\sigma(X_1)$ more explicitly by expanding $H \circ \theta$ to get

$$g_\sigma(X_1) = \mathbb{E}_x \left[\sum_{t=0}^{\tau \circ \theta - 1} c(X_{t+1}) + e(X_\tau \circ \theta) \mid X_1 \right] = \mathbb{E}_x \left[\sum_{t=1}^{\tau \circ \theta} c(X_t) + e(X_\tau \circ \theta) \mid X_1 \right].$$

From the law of iterated expectations, this yields

$$\mathbb{E}_x g_\sigma(X_1) = \mathbb{E}_x \left[\sum_{t=1}^{\tau \circ \theta} c(X_t) + e(X_\tau \circ \theta) \right]. \quad (20)$$

Recalling that $x \in E^c$, so that $\sigma(x) = 0$, we have

$$(T_\sigma g_\sigma)(x) = c(x) + \mathbb{E}_x g_\sigma(X_1) = c(x) + \mathbb{E}_x \left[\sum_{t=1}^{\tau \circ \theta} c(X_t) + e(X_\tau \circ \theta) \right].$$

Using the fact that $x \in E^c$, so that the first visit to E occurs at some $t \geq 1$, we obtain $\tau \circ \theta = \tau - 1$ and $X_\tau \circ \theta = X_\tau$. (The stopping time $\tau \circ \theta$ returns the original index minus 1 because it counts time using the shifted sequence $(X_1, X_2, \dots)$. At the same time, $X_\tau \circ \theta$ – i.e., the stopped value X_τ composed with the shift – evaluates to $X_{(\tau \circ \theta) + 1} = X_\tau$, since the original-sequence and shifted-sequence stopping times pick out the same physical state.) Applying these facts to the last display yields

$$(T_\sigma g_\sigma)(x) = c(x) + \mathbb{E}_x \left[\sum_{t=1}^{\tau-1} c(X_t) + e(X_\tau) \right] = \mathbb{E}_x \left[\sum_{t=0}^{\tau-1} c(X_t) + e(X_\tau) \right] = g_\sigma(x).$$

This confirms that $T_\sigma g_\sigma = g_\sigma$ on E^c , so the proof of Lemma 4.5 is done. $\square$

4.2.5 ADP Formulation

Let $\mathbb{T}$ be the set of all policy operators, as defined in (17), indexed over the restricted policy set Σ defined in (15). Since each T_σ is order preserving, the pair $(bX_+, \mathbb{T})$ forms an ADP. (As elsewhere, bX_+ is paired with the pointwise partial order and the supremum norm.) For this ADP and given $g \in bX_+$, a policy $\sigma \in \Sigma$ is g -min-greedy when $T_\sigma g \leq T_s g$ for all $s \in \Sigma$. It is easy to check that one such policy always exists. Indeed, such a policy can be found by setting

$$\sigma(x) = \mathbb{1} \left\{ e(x) \leq c(x) + \int g(x') P(x, dx') \right\} \quad (x \in X)$$

This policy is in Σ , since σ is $\mathcal{B}$ -measurable and, in addition,

$$e(x) \leq c(x) \implies e(x) \leq c(x) + \int g(x') P(x, dx'),$$

so that $\bar{\sigma} \leq \sigma$. Moreover, it is clear that

$$(T_\sigma g)(x) = \min \left\{ e(x), c(x) + \int g(x') P(x, dx') \right\} \leq (T_s g)(x)$$

for all $s \in \Sigma$.

This argument also proves that the ADP $(bX_+, \mathbb{T})$ is min-regular.

In essence, a g -min-greedy policy treats g as a loss function, using it to assign a total expected cost to each state, and makes the current best choice accordingly.

4.2.6 Stability of the Policy Operators

In this section we prove that $(bX_+, \mathbb{T})$ is globally stable.

Proposition 4.6. *If Assumption 4.2 holds, then every T_σ is globally stable on bX_+ .*

We prove the proposition using a sequence of lemmas. In the statement of the next lemma, σ is a fixed policy, K_σ is as defined in (19), and τ^σ is as defined in (11).

Lemma 4.7. *For all $n \in \mathbb{N}$ and f in bX , we have*

$$\|K_\sigma^n f\| \leq \|f\| \cdot \sup_{x \in X} \mathbb{P}_x \{ \tau^\sigma \geq n \}.$$

Proof. Fixing $f \in bX$ and $x_0 \in X$, we iterate with K_σ and apply the triangle inequality to obtain

$$|(K_\sigma^n f)(x_0)| \leq (1 - \sigma(x_0)) \int (1 - \sigma(x_1)) \cdots \int (1 - \sigma(x_{n-1})) \int |f(x_n)| P(x_{n-1}, dx_n) P(x_{n-2}, dx_{n-1}) \cdots P(x_0, dx_1).$$

Hence, if (X_t) is P -Markov and starts at x_0 , then

$$\begin{aligned} |(K_\sigma^n f)(x_0)| &\leq \|f\| \int \cdots \int \prod_{t=0}^{n-1} (1 - \sigma(x_t)) P(x_{n-1}, dx_n) P(x_{n-2}, dx_{n-1}) \cdots P(x_0, dx_1) \\ &= \|f\| \cdot \mathbb{P}_{x_0} \bigcap_{t=0}^{n-1} \{\sigma(X_t) = 0\} = \|f\| \cdot \mathbb{P}_{x_0} \{\tau^\sigma \geq n\}. \end{aligned}$$

Taking the supremum on the right and then the left produces the bound in Lemma 4.7. $\square$

We will also use the following result regarding stopping times.

Lemma 4.8. *If Assumption 4.2 holds, then, for all $\sigma \in \Sigma$,*

$$\lim_{n \rightarrow \infty} \sup_{x \in X} \mathbb{P}_x \{\tau^\sigma \geq n\} = 0. \quad (21)$$

Proof. Let $\bar{\tau}$ be as in (14). Under Assumption 4.2, there exists an $M < \infty$ with $\mathbb{E}_x \bar{\tau} \leq M$ for all x . In addition, for any $x \in X$ and any $n > 0$, we have:

$$\mathbb{E}_x \bar{\tau} = \sum_{s=0}^{\infty} \mathbb{P}_x \{\bar{\tau} > s\} \geq \sum_{s=0}^n \mathbb{P}_x \{\bar{\tau} > s\} \geq n \cdot \mathbb{P}_x \{\bar{\tau} > n\},$$

where the last inequality holds because $\bar{\tau} > n$ implies $\bar{\tau} > s$ for all $s \leq n$. Rearranging yields

$$\mathbb{P}_x \{\bar{\tau} > n\} \leq \frac{\mathbb{E}_x \bar{\tau}}{n} \leq \frac{M}{n} \quad \text{for all } x \in X.$$

As a consequence,

$$\lim_{n \rightarrow \infty} \sup_{x \in X} \mathbb{P}_x \{\bar{\tau} > n\} = 0. \quad (22)$$

Now fix $\sigma \in \Sigma$. By (16) we have $\tau^\sigma \leq \bar{\tau}$ and hence $\{\tau^\sigma \geq n\} = \{\tau^\sigma > n-1\} \subset \{\bar{\tau} > n-1\}$. Combining this with (22) yields (21). $\square$

Lemma 4.9. *If Assumption 4.2 holds, then, for every $\sigma \in \Sigma$, the policy operator T_σ is asymptotically contracting on bX .*

Proof. Fixing $f, g \in bX$, using the definition of T_σ in (18) and the bound in Lemma 4.7 produces

$$\|T_\sigma^n f - T_\sigma^n g\| \leq \|K_\sigma^n(f - g)\| \leq \|f - g\| \cdot \sup_{x \in X} \mathbb{P}_x \{\tau^\sigma \geq n\}$$

for all $n \in \mathbb{N}$. The claim in Lemma 4.9 now follows from Lemma 4.8. $\square$

Proof of Proposition 4.6. Global stability of T_σ follows from asymptotic contractiveness (Lemma 4.9) and existence of a fixed point (Lemma 4.5). $\square$

4.2.7 Optimality

We define the **minimum loss function** g_∇ via

$$g_\nabla(x) := \inf_{\sigma \in \Sigma} g_\sigma(x) \quad (x \in X). \quad (23)$$

The function g_∇ also takes values in $[0, \infty)$ and is well-defined everywhere on X . Below (see the discussion after Theorem 4.10) we show that g_∇ is equal to the min-value function v_∇ of the ADP, as defined in Section 3.4.

A policy $\sigma \in \Sigma$ is called **optimal** if $g_\sigma \leq g_s$ for all $s \in \Sigma$. This is equivalent to the statement that σ attains the minimum possible cost from every state, and is equivalent to the ADP definition in Section 3.4.

We can now state the following result.

Theorem 4.10. *If Assumption 4.2 holds, then, for the no-discount optimal stopping ADP $(bX_+, \mathbb{T})$,*

- (i) *the fundamental min-optimality properties hold and*
- (ii) *the min-VFI, min-OPI, and min-HPI algorithms all converge.*

Proof. We saw in Section 4.2.5 that $(bX_+, \mathbb{T})$ is min-regular, and in Proposition 4.6 that $(bX_+, \mathbb{T})$ is globally stable. The ADP is min-order bounded, as well, since $T_\sigma 0 \geq 0$ for all σ . Since bX_+ is countably Dedekind complete, all of the conditions of Theorem 3.8 are satisfied. This implies the conclusions of Theorem 4.10. $\square$

With Theorem 4.10 in hand, we can confirm the claim made after (23) that $g_\nabla = v_\nabla$. (The only potential difficulty is that the pointwise infimum in (23) is over an uncountable family, so measurability of g_∇ is not automatic.) By the fundamental min-optimality properties, there exists a min-optimal policy σ , and, for this policy, $v_\sigma = v_\nabla$. Since $v_\tau = g_\tau$ for all $\tau \in \Sigma$ (Lemma 4.5 and Proposition 4.6), we have $v_\nabla = g_\sigma \geq g_\nabla$ pointwise. At the same time, v_∇ is a lower bound of $\{g_\tau : \tau \in \Sigma\}$, so $v_\nabla \leq g_\nabla$ pointwise. Hence $g_\nabla = v_\nabla$; in particular, $g_\nabla \in bX_+$.

4.3 Sequential Analysis

In this section we study the classical Bayesian sequential analysis problem introduced by Wald [1947] and Arrow et al. [1949] (see the introduction for background and related literature). The problem involves no discounting, so the Bellman operator is not a contraction and standard contraction-based theory does not apply. Using Theorem 4.10, our optimality result for no-discount optimal stopping, we show that the fundamental min-optimality properties hold and that min-VFI, min-OPI, and min-HPI all converge under the weak assumption that the candidate densities f_0 and f_1 are distinct.

We proceed by first reducing the sequential analysis problem to a special case of the no-discount optimal stopping problem treated in Theorem 4.10, then verifying the conditions of that theorem.

4.3.1 Set Up

A decision-maker observes a sequence $Z_1, Z_2, \dots$ of IID samples drawn from an unknown probability density f on $\mathbb{R}$. It is known that f is either f_0 or f_1 , where f_0 and f_1 are given probability densities. After each observation the decision-maker chooses one of three actions: accept the hypothesis $f = f_0$ and stop, accept the hypothesis $f = f_1$ and stop, or draw an additional sample at cost $c > 0$. Accepting the wrong hypothesis incurs a strictly positive loss: L_0 when f_0 is accepted but $f = f_1$, and L_1 when f_1 is accepted but $f = f_0$. The decision-maker holds a prior $\pi_0 \in (0, 1)$ that $f = f_1$, and updates to the posterior π_n after n observations via Bayes' rule. This posterior is the natural state variable for the dynamic program below.

The state space for the belief state π is $X = (0, 1)$ and action space is $A = \{0, 1, 2\}$, where action 0 represents accepting f_0 , action 1 represents accepting f_1 , and action 2 represents continuing to sample. We assume that the two densities are defined on $\mathbb{R}$.

The Bellman equation has the form

$$g(\pi) = \min \left\{ \pi L_0, (1 - \pi) L_1, c + \int g(\pi') P(\pi, d\pi') \right\} \quad (24)$$

for $\pi \in (0, 1)$, where the stochastic kernel P obeys

$$(Pg)(\pi) := \int g(\kappa(\pi, z)) \psi(\pi, z) dz. \quad (25)$$

Here

$$\psi(\pi, z) = (1 - \pi) f_0(z) + \pi f_1(z) \quad (26)$$

is the predictive density and

$$\kappa(\pi, z) = \frac{\pi f_1(z)}{(1 - \pi) f_0(z) + \pi f_1(z)} = \frac{\pi f_1(z)}{\psi(\pi, z)}$$

is the Bayesian update rule. Together, ψ and κ define the stochastic kernel P governing the belief state $(\pi_n)_{n \geq 0}$, describing how beliefs evolve from the perspective of the controller when the observation sequence $(Z_n)_{n \geq 1}$ is forecast using the predictive density.

In all of what follows, the cost c and the losses L_0 and L_1 are assumed to be positive constants. Our aim is to characterize and solve for optimal policies.

In terms of dynamic programming, we can simplify the sequential analysis to a binary stopping problem. Our first step is to set

$$e(\pi) := \min\{\pi L_0, (1 - \pi) L_1\}. \quad (27)$$

Now consider the Bellman equation

$$g(\pi) = \min \{e(\pi), c + (Pg)(\pi)\}. \quad (28)$$

We claim that solving this dynamic program is sufficient for solving the sequential analysis problem. To see this, suppose we are able to show that the fundamental min-optimality properties hold for this dynamic program. Let the min-value function be denoted by g_∇ . Continuing the convention that the controller always stops when indifferent between stopping and continuing, Bellman's principle of min-optimality tells the controller to stop if and only if $e(\pi) \leq c + (Pg_\nabla)(\pi)$.

In the present setting, this means that the controller stops if and only if at least one of the stopping losses πL_0 and $(1 - \pi) L_1$ is less than or equal to the continuation loss

$c + (Pg_{\nabla})(\pi)$. When this stop occurs, the controller then makes the static choice over the two density options (selecting f_0 or f_1) depending on which of πL_0 and $(1 - \pi)L_1$ is smaller. Since this static problem is trivial, we can concentrate on solving the dynamic problem represented by the Bellman equation (28).

The stopping problem just described is a special case of the no-discount optimal stopping problem from Section 4.2, with e as the exit cost function in (27) and the flow cost c constant over the state space $X = (0, 1)$. (To confirm this, compare the Bellman equation (28) with the general case in (10).) As a result, Theorem 4.10 applies. All we need to do is check that Assumption 4.2 is valid in the current setting. This turns out to be true whenever f_0 and f_1 are distinct. Our proof will rely on a bound for martingale stopping times in Theorem 6.1.

4.3.2 Verifying Assumption 4.2

Let's consider verification of Assumption 4.2 in the present setting. Let $\bar{\tau}$ be the upper bound stopping time for the belief state (i.e., the stopping time in (14) specialized to the current setting). This stopping time is defined in terms of the certain exit region (see (13)), which, for our problem, is

$$\bar{E} := \{\pi \in (0, 1) : \min\{\pi L_0, (1 - \pi)L_1\} \leq c\}.$$

Equivalently,

$$\pi \in \bar{E} \iff \pi \leq \frac{c}{L_0} \quad \text{or} \quad \pi \geq 1 - \frac{c}{L_1}.$$

The lower bound policy $\bar{\sigma}$ is (recalling its definition from Section 4.2.3) the indicator function for $\bar{E}$, stopping the process whenever the belief state enters the certain exit region. The upper bound stopping time is, therefore,

$$\bar{\tau} = \inf \left\{ n \geq 0 : \pi_n \leq \frac{c}{L_0} \text{ or } \pi_n \geq 1 - \frac{c}{L_1} \right\}. \quad (29)$$

Here the (π_n) process evolves according to the kernel P from (25): given π_n , we draw Z_{n+1} independently from $\psi(\pi_n, \cdot)$, and then set

$$\pi_{n+1} = \kappa(\pi_n, Z_{n+1}) = \frac{\pi_n f_1(Z_{n+1})}{\psi(\pi_n, Z_{n+1})}. \quad (30)$$

(Note here that division by zero is not a concern: For $\pi \in (0, 1)$, we have $\psi(\pi, z) > 0$ if and only if $f_0(z) + f_1(z) > 0$. Since Z_{n+1} is drawn from $\psi(\pi, \cdot)$, we have $\psi(\pi_n, Z_{n+1}) > 0$ almost surely.)

Matching (15), the policy set Σ under consideration is all $\mathcal{B}$ -measurable $\sigma: X \rightarrow \{0, 1\}$ with $\bar{\sigma} \leq \sigma$. For σ in this set, the policy operator takes the form

$$(T_\sigma g)(\pi) = \sigma(\pi)e(\pi) + (1 - \sigma(\pi)) \left[c + \int g(\pi')P(\pi, d\pi') \right]. \quad (31)$$

Let V be all bounded Borel measurable functions from $(0, 1)$ to $\mathbb{R}_+$. With $\mathbb{T}_{\text{SA}}$ as the family of operators described by (31), indexed by $\sigma \in \Sigma$, the pair $(V, \mathbb{T}_{\text{SA}})$ forms an ADP. For this ADP, we can state the following result. In the statement, the densities f_0 and f_1 are called **distinct** if they differ on a set of positive Lebesgue measure.

Proposition 4.11. *If f_0 and f_1 are distinct, then for the sequential analysis ADP $(V, \mathbb{T}_{\text{SA}})$, the fundamental min-optimality properties hold and min-VFI, min-OPI, and min-HPI all converge.*

To prove Proposition 4.11, we recognize $(V, \mathbb{T}_{\text{SA}})$ as a special case of the no-discount optimal stopping ADP treated in Theorem 4.10. As such, we only need to verify Assumption 4.2, which amounts to showing that $\sup_{0 < \pi < 1} \mathbb{E}_\pi \bar{\tau}$ is finite.

As a first step, we recall that, given two probability densities f_0 and f_1 on $\mathbb{R}$, the **triangular discrimination** is

$$\Delta(f_0, f_1) := \int \frac{[f_1(z) - f_0(z)]^2}{f_0(z) + f_1(z)} dz,$$

where the integrand is defined to be zero when $f_0(z) + f_1(z) = 0$.

Lemma 4.12. *If f_0 and f_1 are distinct, then $\Delta(f_0, f_1) > 0$.*

Proof. If f_0 and f_1 are distinct, then they differ on a set E of positive measure. For $z \in E$ we have $f_0(z) + f_1(z) > 0$ (since otherwise both are zero, which contradicts the definition of E). This means that the integrand in the definition of $\Delta(f_0, f_1)$ is positive on E . Integration of a positive function over a set of positive measure yields a positive value. $\square$

Now let's investigate the properties of stopping times for the process (π_n) from (30). We will begin with a generic stopping time

$$\tau = \inf\{n \geq 0 : \pi_n \notin (a, b)\},$$

where a, b are numbers satisfying $0 < a < b < 1$.

Lemma 4.13. *If f_0 and f_1 are distinct, then, for any initial condition $\pi_0 \in (0, 1)$, we have*

$$\mathbb{E}_{\pi_0}[\tau] \leq \frac{1}{\delta}, \quad \text{where } \delta := a^2(1-b)^2\Delta(f_0, f_1).$$

Proof. Fix $\pi_0 \in (0, 1)$. If π_0 is not in (a, b) , then $\tau = 0$, in which case the claim is trivial. Hence, from now on, we assume that π_0 is in (a, b) . Note that $(\pi_n)_{n \geq 0}$ is a martingale, since

$$\mathbb{E}[\pi_{n+1} \mid \pi_n = \pi] = \int \frac{\pi f_1(z)}{\psi(\pi, z)} \cdot \psi(\pi, z) \, dz = \int \pi f_1(z) \, dz = \pi.$$

We apply Theorem 6.1 to this bounded martingale. Our main task is to obtain δ in (34). For fixed π , the law of motion (30) yields

$$\mathbb{E}[\pi_{n+1}^2 \mid \pi_n = \pi] = \int \frac{\pi^2 f_1^2(z)}{\psi(\pi, z)} \, dz = \pi^2 \int \frac{f_1^2(z)}{\psi(\pi, z)} \, dz.$$

(As for the triangular discrimination, the integrand is defined to be zero when $\psi(\pi, z) = 0$, which occurs only when $f_0(z) = f_1(z) = 0$). Using the fact that (π_n) is a martingale, we get

$$v(\pi) := \mathbb{E}[(\pi_{n+1} - \pi_n)^2 \mid \pi_n = \pi] = \pi^2 \left[\int \frac{f_1^2(z)}{\psi(\pi, z)} \, dz - 1 \right].$$

To simplify the bracketed term, we observe that

$$\begin{aligned} \int \frac{[f_1(z) - \psi(\pi, z)]^2}{\psi(\pi, z)} \, dz &= \int \frac{f_1(z)^2}{\psi(\pi, z)} \, dz - 2 \int f_1(z) \, dz + \int \psi(\pi, z) \, dz \\ &= \int \frac{f_1(z)^2}{\psi(\pi, z)} \, dz - 1. \end{aligned}$$

Using this fact, combined with $f_1(z) - \psi(\pi, z) = (1 - \pi)[f_1(z) - f_0(z)]$ and the definition of $v(\pi)$, gives

$$v(\pi) = \pi^2(1 - \pi)^2 \int \frac{[f_1(z) - f_0(z)]^2}{\psi(\pi, z)} \, dz. \quad (32)$$

Since $\psi(\pi, z) \leq f_0(z) + f_1(z)$, we have

$$\int \frac{[f_1(z) - f_0(z)]^2}{\psi(\pi, z)} \, dz \geq \int \frac{[f_1(z) - f_0(z)]^2}{f_0(z) + f_1(z)} \, dz = \Delta(f_0, f_1). \quad (33)$$

Combining (32) and (33), for $\pi \in [a, b]$:

$$v(\pi) \geq \pi^2(1 - \pi)^2\Delta(f_0, f_1) \geq a^2(1 - b)^2\Delta(f_0, f_1) = \delta.$$

By Lemma 4.12, the triangular discrimination is strictly positive, so $\delta > 0$. Since $v(\pi) = \mathbb{E}[(\pi_{n+1} - \pi_n)^2 \mid \pi_n = \pi]$, this verifies condition (34) for $M_n = \pi_n$ and this choice of δ . Applying Theorem 6.1 and using $(\pi_\tau - \pi_0)^2 \leq 1$, we get

$$\mathbb{E}_{\pi_0}[\tau] \leq \frac{\mathbb{E}_{\pi_0}[(\pi_\tau - \pi_0)^2]}{\delta} \leq \frac{1}{\delta}.$$

This verifies the claim in Lemma 4.13. $\square$

Proof of Proposition 4.11. Let f_0 and f_1 be distinct. Since $(V, \mathbb{T}_{\text{SA}})$ is a special case of the no-discount optimal stopping ADP treated in Theorem 4.10, we only need to show that $\sup_{\pi \in (0,1)} \mathbb{E}_\pi \bar{\tau}$ is finite, where $\bar{\tau}$ is as defined in (29).

Consider first the degenerate case $c/L_0 + c/L_1 \geq 1$. In this case the two intervals $(0, c/L_0]$ and $[1 - c/L_1, 1)$ cover $(0, 1)$, so $\bar{E} = (0, 1)$ and $\bar{\tau} = 0$ $\mathbb{P}_\pi$ -a.s. for every $\pi \in (0, 1)$. Hence $\sup_{\pi \in (0,1)} \mathbb{E}_\pi \bar{\tau} = 0 < \infty$.

Now consider the non-degenerate case $c/L_0 + c/L_1 < 1$. Setting $a := c/L_0$ and $b := 1 - c/L_1$, we have $0 < a < b < 1$. By construction, the stopping time $\bar{\tau}$ in (29) is the first exit time of the belief process from (a, b) , so Lemma 4.13 applies and yields $\mathbb{E}_\pi[\bar{\tau}] \leq 1/\delta$, where

$$\delta := a^2(1-b)^2 \Delta(f_0, f_1) = \left(\frac{c}{L_0}\right)^2 \left(\frac{c}{L_1}\right)^2 \Delta(f_0, f_1) = \left(\frac{c^2}{L_0 L_1}\right)^2 \Delta(f_0, f_1).$$

Since f_0 and f_1 are distinct, Lemma 4.12 implies that $\Delta(f_0, f_1) > 0$, so $\delta > 0$. The bound $\mathbb{E}_\pi[\bar{\tau}] \leq 1/\delta$ is valid for all $\pi \in (0, 1)$, confirming that Assumption 4.2 holds. As a result, all the claims in Theorem 4.10 are valid. $\square$

5 Conclusion

In this paper we extended the theory of abstract dynamic programming by pairing order-theoretic structure with topological and metric foundations. Working on pospaces, we showed that global stability of the policy operators, combined with regularity and existence of a fixed point of the Bellman operator (or primitive conditions implying it), yields the fundamental optimality properties together with convergence of value function iteration, Howard policy iteration, and optimistic policy iteration. On partially ordered metric spaces, contractivity delivers the same conclusions along with geometric rates of convergence. We further showed that stationary policies weakly dominate nonstationary policy plans, and we obtained minimization versions of our

main results through order duality. The applications illustrate the reach of the framework: it recovers standard results for weak Feller MDPs, accommodates structural estimation problems that lie outside existing frameworks, and yields new results for no-discount optimal stopping and Bayesian sequential analysis.

Several directions for future work suggest themselves. One is to extend the present results to unbounded settings, where value spaces fail to be order bounded and contractivity holds only with respect to weighted norms. Another is to apply the framework to additional problem classes mentioned in the introduction, including distributional dynamic programming and recursive preference models. Finally, the convergence results for Howard and optimistic policy iteration obtained here suggest investigating sample-based counterparts of these algorithms, connecting the optimality theory developed above to the convergence theory of reinforcement learning.

6 Appendix

This appendix contains supporting results and longer proofs omitted from the main text. Section 6.1 establishes a bound on exit times for bounded martingales, which is used in the sequential analysis application of Section 4.3. Section 6.2 proves Theorem 3.6, on nonstationary policies. Section 6.3 provides proofs of the minimization results stated in Section 3.4.

6.1 A Stopping Time Bound

Here we state a result on exit times for bounded martingales that is used in the proof of Proposition 4.11. In the statement of the theorem, $(M_t, \mathcal{F}_t)$ is a discrete-time martingale with $M_0 = m$, and

$$\tau = \inf\{t \geq 0 : M_t \notin (a, b)\}$$

for some a, b in $\mathbb{R}$ with $a < b$.

Theorem 6.1. *Let $(M_t, \mathcal{F}_t)$ be bounded, in the sense that there exists a $K \in \mathbb{R}$ with $|M_t| \leq K$ with probability one for all t . If, in addition, there exists a $\delta > 0$ with*

$$\delta \leq \mathbb{E}[(M_{t+1} - M_t)^2 \mid \mathcal{F}_t] \quad \text{for all } t < \tau, \tag{34}$$

then

$$\mathbb{E}_m[\tau] \leq \frac{\mathbb{E}_m[(M_\tau - m)^2]}{\delta}.$$

Proof. Throughout the proof, all expectations are conditional on $M_0 = m$, and we write $\mathbb{E}$ for $\mathbb{E}_m$. Define the process $\langle M \rangle_t := \sum_{s=0}^{t-1} (M_{s+1} - M_s)^2$, with $\langle M \rangle_0 := 0$. We claim that $(M_t - M_0)^2 - \langle M \rangle_t$ is a martingale. To see this we write $M_{t+1} - M_0 = (M_{t+1} - M_t) + (M_t - M_0)$ and square it to get

$$(M_{t+1} - M_0)^2 = (M_t - M_0)^2 + 2(M_t - M_0)(M_{t+1} - M_t) + (M_{t+1} - M_t)^2.$$

Taking conditional expectations and using the martingale property,

$$\mathbb{E}[(M_{t+1} - M_0)^2 \mid \mathcal{F}_t] = (M_t - M_0)^2 + \mathbb{E}[(M_{t+1} - M_t)^2 \mid \mathcal{F}_t]. \quad (35)$$

By the definition of $\langle M \rangle$, $\langle M \rangle_{t+1} = \langle M \rangle_t + (M_{t+1} - M_t)^2$, so

$$\mathbb{E}[\langle M \rangle_{t+1} \mid \mathcal{F}_t] = \langle M \rangle_t + \mathbb{E}[(M_{t+1} - M_t)^2 \mid \mathcal{F}_t]. \quad (36)$$

Subtracting (36) from (35), we get

$$\mathbb{E}[(M_{t+1} - M_0)^2 - \langle M \rangle_{t+1} \mid \mathcal{F}_t] = (M_t - M_0)^2 - \langle M \rangle_t.$$

This is the martingale property for $(M_t - M_0)^2 - \langle M \rangle_t$. Since (M_t) is bounded, this process is integrable for each t , and hence is a martingale. Since $\tau \wedge n \leq n$ is a bounded stopping time, the optional stopping theorem gives

$$\mathbb{E}[(M_{\tau \wedge n} - M_0)^2 - \langle M \rangle_{\tau \wedge n}] = \mathbb{E}[(M_0 - M_0)^2 - \langle M \rangle_0] = 0,$$

and hence

$$\mathbb{E}[(M_{\tau \wedge n} - M_0)^2] = \mathbb{E}[\langle M \rangle_{\tau \wedge n}]. \quad (37)$$

In addition, for $t < \tau$, we have

$$\mathbb{E}[\langle M \rangle_{t+1} - \langle M \rangle_t \mid \mathcal{F}_t] = \mathbb{E}[(M_{t+1} - M_t)^2 \mid \mathcal{F}_t] \geq \delta,$$

so

$$\mathbb{E}[\langle M \rangle_{\tau \wedge n}] = \mathbb{E} \left[\sum_{t=0}^{\tau \wedge n - 1} (M_{t+1} - M_t)^2 \right] \geq \delta \cdot \mathbb{E}[\tau \wedge n]. \quad (38)$$

Combining (38) and (37) yields

$$\mathbb{E}[\tau \wedge n] \leq \frac{\mathbb{E}[(M_{\tau \wedge n} - M_0)^2]}{\delta}.$$

Taking $n \rightarrow \infty$ and applying monotone convergence on the left and dominated convergence on the right (using boundedness of the martingale) gives the bound claimed

in Theorem 6.1. □

6.2 Proof of Theorem 3.6

In proving Theorem 3.6, we use the following preliminary result.

Lemma 6.2. *If Assumption 3.1 holds, then, for each $v \in V$ and policy plan $\bar{\sigma} := (\sigma_t)_{t \geq 0}$, the limit*

$$v_{\bar{\sigma}} := \lim_{n \rightarrow \infty} T_{\sigma_0} \cdots T_{\sigma_n} v$$

exists in V and is independent of v .

Proof. Fix $v \in V$ and policy plan $\bar{\sigma} := (\sigma_t)_{t \geq 0}$. Given the policy plan above and $m \leq n$, we adopt the following simplified notation:

$$T_{m,n} := T_{\sigma_m} \circ T_{\sigma_{m+1}} \circ \cdots \circ T_{\sigma_n}.$$

Let $v_n = T_{0,n}v$. Our claim is that $\lim_n v_n$ exists in V and is independent of v . To see this, observe first that (v_n) is Cauchy, since, fixing $m, j \in \mathbb{N}$,

$$d(v_m, v_{m+j}) \leq \lambda^{m+1} d(v, T_{m+1, m+j} v),$$

and, by repeatedly applying the triangle inequality,

$$\begin{aligned} d(v, T_{m+1, m+j} v) &\leq d(v, T_{m+1, m+1} v) + d(T_{m+1, m+1} v, T_{m+1, m+2} v) \\ &\quad + \cdots + d(T_{m+1, m+j-1} v, T_{m+1, m+j} v). \end{aligned}$$

By Assumption 3.1, there exists a finite constant b satisfying $d(v, T_{\sigma_k} v) \leq b$ for all k . From this and the last bound (with the convention that $T_{m+1, m}$ is the identity, the k -th telescoping term satisfies $d(T_{m+1, m+k-1} v, T_{m+1, m+k} v) \leq \lambda^{k-1} d(v, T_{\sigma_{m+k}} v) \leq \lambda^{k-1} b$), we obtain

$$d(v, T_{m+1, m+j} v) \leq b + \lambda b + \cdots + \lambda^{j-1} b \leq \frac{b}{1 - \lambda}.$$

As a result, $d(v_m, v_{m+j}) \leq \lambda^{m+1} b / (1 - \lambda)$. This shows that (v_n) is Cauchy. Using completeness of V and letting $\bar{v}$ be the limit of this sequence, we argue that $\bar{v}$ is independent of v . Indeed, if $w_n := T_{0,n} w$ for some $w \in V$, then $d(v_n, w_n) \leq \lambda^{n+1} d(v, w)$ for all n , so that (v_n) and (w_n) have the same limit. Hence $\lim_{n \rightarrow \infty} T_{0,n} v$ exists in V and is independent of the initial condition v . □

Proof of Theorem 3.6. Under the hypotheses of Theorem 3.6, $(V, \mathbb{T})$ is V -semi-regular (since regularity gives $V_G = V$, with $TV \subset V$ trivially), so the first claim is immediate from Theorem 3.4.

Regarding the second, let $\bar{\sigma} := (\sigma_t)_{t \geq 0}$ be an arbitrary policy plan and let σ be an optimal policy. Since v_σ is a fixed point of T and $T_{\sigma_t} w \preceq Tw$ for every $w \in V$ and every t , induction on j yields

$$T_{\sigma_0} T_{\sigma_1} \cdots T_{\sigma_j} v_\sigma \preceq T^{j+1} v_\sigma = v_\sigma \quad \text{for all } j \in \mathbb{N}.$$

By Lemma 6.2, the left-hand side converges to $v_{\bar{\sigma}}$ as $j \rightarrow \infty$. Since the partial order is closed, taking the limit in j yields $v_{\bar{\sigma}} \preceq v_\sigma$. This proves that every policy plan is weakly dominated by a stationary policy. $\square$

6.3 Proofs for the Minimization Cases

Here we provide proofs for the results in Section 3.4. One option is to just replicate all proofs by reversing inequalities and replacing sup with inf. Another is to use order duality, converting minimization problems into maximization problems, calling on the maximization results, and converting back. We use the second strategy.

To increase clarity, when discussing maximization and minimization in the same section, we add a “max-” prefix to the previously introduced definitions that pertain to maximization. For example,

- “ v -greedy policies” will be referred to as **v -max-greedy policies**,
- “optimal policies” will be referred to as **max-optimal policies**,
- “the Bellman equation” will be referred to as the **Bellman max-equation**,

and so on. In addition, rather than using generic T for the Bellman operator, we use T_Δ for the max case and T_∇ for the min case. Likewise, instead of using generic v^* for the value function, we use v_Δ for the max case and v_∇ for the min case. Similar conventions are used for the operators H and W .

If $V := (V, \preceq)$ is a partially ordered set, then its **order dual** $V^\partial := (V, \preceq^\partial)$ is the set V paired with the partial order $\preceq^\partial$ obtained by setting $u \preceq^\partial v$ if and only if $v \preceq u$. If $(V, \mathbb{T})$ is any ADP, then we call $(V, \mathbb{T})^\partial := (V^\partial, \mathbb{T})$ the **dual** of $(V, \mathbb{T})$. In other words, the dual $(V, \mathbb{T})^\partial$ of $(V, \mathbb{T})$ is the ADP created by maintaining the same family of policy operators $\mathbb{T}$ while replacing the partially ordered set V with its order dual V^∂ .

It is easy to confirm that $(V, \mathbb{T})^\partial$ is an ADP.

Regarding notation for $(V, \mathbb{T})^\partial$,

- the Bellman max-operator will be denoted by T_{Δ}^{∂} ,
- the Bellman min-operator will be denoted by T_{∇}^{∂} ,
- the max-value function will be denoted by v_{Δ}^{∂} ,
- V_G^{∂} is all $v \in V$ such that at least one v -max greedy policy exists for $(V, \mathbb{T})^{\partial}$,
- etc.

Each ADP is self-dual, in the sense that $((V, \mathbb{T})^{\partial})^{\partial} = (V, \mathbb{T})$. This follows from the fact that all partially ordered sets are self-dual.

The following elementary observation will be used repeatedly. It says that suprema and infima reverse roles under the order dual.

Lemma 6.3. *Let $A \subset V$. If $\bigvee A$ exists in V , then $\bigwedge^{\partial} A$ exists in V^{∂} and $\bigwedge^{\partial} A = \bigvee A$. Conversely, if $\bigwedge A$ exists in V , then $\bigvee^{\partial} A$ exists in V^{∂} and $\bigvee^{\partial} A = \bigwedge A$.*

Proof. Immediate from the definition $u \preceq^{\partial} v \iff v \preceq u$: an upper bound for A under $\preceq^{\partial}$ is a lower bound under $\preceq$, and the least upper bound under $\preceq^{\partial}$ is the greatest lower bound under $\preceq$. The second statement follows by self-duality. $\square$

Lemma 6.4. *If $(V, \mathbb{T})$ is a well-posed ADP with dual $(V, \mathbb{T})^{\partial}$, then, fixing $v \in V$, the following statements are valid:*

- (i) σ is v -min-greedy for $(V, \mathbb{T})$ if and only if σ is v -max-greedy for $(V, \mathbb{T})^{\partial}$.
- (ii) $(V, \mathbb{T})$ is min-regular if and only if $(V, \mathbb{T})^{\partial}$ is max-regular,
- (iii) $(V, \mathbb{T})$ is min-order bounded if and only if $(V, \mathbb{T})^{\partial}$ is max-order bounded,
- (iv) If $T_{\Delta}^{\partial} v$ exists then so does $T_{\nabla} v$, and, moreover, $T_{\nabla} v = T_{\Delta}^{\partial} v$.
- (v) If $W^{\partial} v$ exists then so does $W_{\nabla} v$, and, moreover, $W_{\nabla} v = W^{\partial} v$.
- (vi) If $H^{\partial} v$ exists then so does $H_{\nabla} v$, and, moreover, $H_{\nabla} v = H^{\partial} v$.
- (vii) If v_{Δ}^{∂} exists for $(V, \mathbb{T})^{\partial}$, then v_{∇} exists for $(V, \mathbb{T})$ and $v_{\nabla} = v_{\Delta}^{\partial}$.
- (viii) $\sigma \in \Sigma$ is min-optimal for $(V, \mathbb{T})$ if and only if σ is max-optimal for $(V, \mathbb{T})^{\partial}$.
- (ix) $V_{\nabla}^G = V_G^{\partial}$.

Proof. Regarding (i), fix $v \in V$. Policy σ is v -min-greedy for $(V, \mathbb{T})$ if and only if $T_{\sigma} v \preceq T_{\tau} v$ for all $\tau \in \Sigma$, which is equivalent to $T_{\tau} v \preceq^{\partial} T_{\sigma} v$ for all $\tau \in \Sigma$. Hence σ is v -min-greedy for $(V, \mathbb{T})$ if and only if σ is v -max-greedy for $(V, \mathbb{T})^{\partial}$.

Claim (ii) follows directly from (i): $(V, \mathbb{T})$ is min-regular iff a ν -min-greedy policy exists for every ν , iff a ν -max-greedy policy exists for $(V, \mathbb{T})^\partial$ for every ν , iff $(V, \mathbb{T})^\partial$ is max-regular.

Claim (iii) holds because $b \preceq T_\sigma b$ for all σ is equivalent to $T_\sigma b \preceq^\partial b$ for all σ by the definition of $\preceq^\partial$.

Claim (iv) is immediate from Lemma 6.3: when $T_\Delta^\partial \nu = \bigvee_\sigma^\partial T_\sigma \nu$ exists in V^∂ , the infimum $\bigwedge_\sigma T_\sigma \nu$ exists in V and equals $T_\Delta^\partial \nu$; the former is precisely $T_\nabla \nu$.

For claim (v), by (i) the set of ν -min-greedy policies for $(V, \mathbb{T})$ coincides with the set of ν -max-greedy policies for $(V, \mathbb{T})^\partial$, and *the same selection rule* fixes a common σ . Since both W_∇ and W^∂ are defined as $T_\sigma^m \nu$ for that common σ , they agree wherever defined. Claim (vi) is analogous: both $H_\nabla \nu$ and $H^\partial \nu$ equal ν_σ – the unique fixed point of T_σ in V , which does not depend on which partial order we equip V with – for the common ν -(min/max)-greedy σ .

For claim (vii), recall $\nu_\nabla := \bigwedge_\sigma \nu_\sigma$ and $\nu_\Delta^\partial := \bigvee_\sigma^\partial \nu_\sigma$. By Lemma 6.3, if ν_Δ^∂ exists in V^∂ then $\nu_\nabla = \bigwedge_\sigma \nu_\sigma$ exists in V and equals ν_Δ^∂ .

Claim (viii) follows from (vii): σ is min-optimal for $(V, \mathbb{T})$ iff $\nu_\sigma = \nu_\nabla$, which by (vii) holds iff $\nu_\sigma = \nu_\Delta^\partial$, iff σ is max-optimal for $(V, \mathbb{T})^\partial$.

Claim (ix) follows from (i): $\nu \in V_G^G$ iff a ν -min-greedy policy for $(V, \mathbb{T})$ exists, iff a ν -max-greedy policy for $(V, \mathbb{T})^\partial$ exists, iff $\nu \in V_G^\partial$. $\square$

Self-duality implies corollaries to Lemma 6.4 that we treat as self-evident. For example, $\sigma \in \Sigma$ is min-optimal for $(V, \mathbb{T})^\partial$ if and only if σ is max-optimal for $(V, \mathbb{T})$.

Part (viii) of Lemma 6.4 tells us that we can solve an ADP for a min-optimal policy by switching to the dual ADP and maximizing.

The following result is useful when we consider minimization problems.

Lemma 6.5. *S is order stable on V if and only if S is order stable on V^∂ .*

Proof. Suppose that S is order stable on V . By definition, S has a unique fixed point $\bar{\nu} \in V$. Hence it remains only to check that conditions (i)–(ii) in the definition of order stability hold on V^∂ . Regarding (i), suppose $\nu \in V$ and $\nu \preceq^\partial S\nu$. Then $S\nu \preceq \nu$ and hence $\bar{\nu} \preceq \nu$, by condition (ii) applied on V . But then $\nu \preceq^\partial \bar{\nu}$, so condition (i) holds on V^∂ . The proof of condition (ii) on V^∂ is similar.

We have shown that S is order stable on V^∂ whenever S is order stable on V . The reverse implication holds because the dual of V^∂ is V . $\square$

Theorem 6.6. *The fundamental max-optimality properties hold for $(V, \mathbb{T})^\partial$ if and only if the fundamental min-optimality properties hold for $(V, \mathbb{T})$. Moreover,*

- (i) *max-VFI converges for $(V, \mathbb{T})^\partial$ if and only if min-VFI converges for $(V, \mathbb{T})$,*
- (ii) *max-OPI converges for $(V, \mathbb{T})^\partial$ if and only if min-OPI converges for $(V, \mathbb{T})$, and*
- (iii) *max-HPI converges for $(V, \mathbb{T})^\partial$ if and only if min-HPI converges for $(V, \mathbb{T})$.*

Proof. Immediate from Lemma 6.4 and Lemma 6.3: each ingredient of the min-optimality properties for $(V, \mathbb{T})$ – existence of a min-optimal policy, v_∇ existing in V_∇^G and satisfying the Bellman min-equation, Bellman’s principle of min-optimality – coincides with the corresponding max-side ingredient for $(V, \mathbb{T})^\partial$. The convergence claims follow analogously, using that $T^n v \downarrow v^*$ in V if and only if $T^n v \uparrow v^*$ in V^∂ . $\square$

Now we can prove the main results from Section 3.4.

Proof of Theorem 3.7. Let $(V, \mathbb{T})$ be as stated and let $(V, \mathbb{T})^\partial$ be the dual ADP. Since $(V, \mathbb{T})$ is min-regular, $(V, \mathbb{T})^\partial$ is max-regular (Lemma 6.4). Since $(V, \mathbb{T})$ is globally stable, $(V, \mathbb{T})^\partial$ is likewise globally stable. By assumption, the Bellman min-operator T_∇ for $(V, \mathbb{T})$ has a fixed point $\bar{v}$ in V . Since the Bellman max-operator T^∂ for $(V, \mathbb{T})^\partial$ satisfies $T_\Delta^\partial = T_\nabla$ (the supremum under $\preceq^\partial$ equals the infimum under $\preceq$), the element $\bar{v}$ is also a fixed point of T_Δ^∂ . As a result, Theorem 3.1 implies that, for $(V, \mathbb{T})^\partial$, the fundamental max-optimality properties hold and max-VFI, max-OPI and max-HPI all converge. The conclusions of Theorem 3.7 now follow from Theorem 6.6. $\square$

Lemma 6.7. *If V is any partially ordered set, then V^∂ is countably Dedekind complete whenever V is countably Dedekind complete.*

Proof. Let V be countably Dedekind complete and let $A \subset V$ be finite or countable and, for the partial order $\preceq^\partial$, bounded above by some $b \in V$. This means that $a \preceq^\partial b$ for all $a \in A$, so $b \preceq a$ for all $a \in A$. In other words, b is a lower bound for A in V . Thus, by countable Dedekind completeness, the infimum of A exists in V . It follows that the supremum of A exists in V^∂ . The proof for the bounded below case is similar. $\square$

Proof of Theorem 3.8. Let $(V, \mathbb{T})$ be as stated and let $(V, \mathbb{T})^\partial$ be the dual ADP. The set V^∂ is countably Dedekind complete by Lemma 6.7. Since $(V, \mathbb{T})$ is min-regular and min-order bounded, the dual $(V, \mathbb{T})^\partial$ is max-regular and max-order bounded (Lemma 6.4). Since global stability of $(V, \mathbb{T})$ is equivalent to global stability of $(V, \mathbb{T})^\partial$, the conditions of Theorem 3.3 all hold for $(V, \mathbb{T})^\partial$. As a result, for $(V, \mathbb{T})^\partial$, the fundamental max-optimality properties hold and max-VFI, max-OPI and max-HPI all converge. The conclusions of Theorem 3.8 now follow from Theorem 6.6. $\square$

References

- Charalambos D Aliprantis and C Border, Kim. *Infinite dimensional analysis: a hitchhiker's guide*. Springer-Verlag, New York, 3 edition, 2006.
- Kenneth J. Arrow, David Blackwell, and Meyer A. Girshick. Bayes and minimax solutions of sequential decision problems. *Econometrica*, 17(3/4):213–244, 1949. doi: 10.2307/1905525.
- Łukasz Balbus, Kevin Reffett, and Łukasz Woźny. Time-consistent equilibria in dynamic models with recursive payoffs and behavioral discounting. *Journal of Economic Theory*, page 105493, 2022.
- Łukasz Balbus, Anna Jaśkiewicz, Andrzej S Nowak, and Łukasz Woźny. Markov perfect equilibria in stochastic growth models with quasi-hyperbolic discounting and risk-sensitive preferences. *Dynamic Games and Applications*, 16(1):107–125, 2026.
- Nicole Bäuerle and Anna Jaśkiewicz. Time-consistency in the mean-variance problem: A new perspective. *IEEE Transactions on Automatic Control*, 70(1), 2025.
- Nicole Bäuerle and Ulrich Rieder. *Markov decision processes with applications to finance*. Springer Science & Business Media, 2011.
- Richard Bellman. *Dynamic programming*. American Association for the Advancement of Science, 1957.
- Dimitri P. Bertsekas. *Dynamic Programming and Stochastic Control*, volume 125 of *Mathematics in Science and Engineering*. Academic Press, New York, 1976.
- Dimitri P Bertsekas. Monotone mappings with application in dynamic programming. *SIAM Journal on Control and Optimization*, 15(3):438–464, 1977.
- Dimitri P Bertsekas. Regular policies in abstract dynamic programming. *SIAM Journal on Optimization*, 27(3):1694–1727, 2017.
- Dimitri P Bertsekas. Proper policies in infinite-state stochastic shortest path problems. *IEEE Transactions on Automatic Control*, 63(11):3787–3792, 2018.
- Dimitri P Bertsekas. *Abstract dynamic programming*. Athena Scientific, 3 edition, 2022.
- Dimitri P. Bertsekas and John N. Tsitsiklis. An analysis of stochastic shortest path problems. *Mathematics of Operations Research*, 16(3):580–595, 1991.

- Gaetano Bloise, Cuong Le Van, and Yiannis Vailakis. Do not blame Bellman: It is Koopmans' fault. *Econometrica*, 92:111–140, 2024.
- Antoine Bommier and Bertrand Villeneuve. Risk aversion and the value of risk to life. *Journal of Risk and Insurance*, 79(1):77–104, 2012.
- Nicole Bäuerle and Athanasios Vasileiadis. Markov decision processes of the third kind: Learning distributions by policy gradient descent, 2026. URL <https://arxiv.org/abs/2602.06567>.
- Dan Cao. Recursive equilibrium in Krusell and Smith (1998). *Journal of Economic Theory*, 186, 2020.
- Christopher D Carroll. Precautionary saving and the marginal propensity to consume out of permanent income. *Journal of Monetary Economics*, 56(6):780–790, 2009.
- Y. S. Chow and H. Robbins. On optimal stopping rules. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 2(1):33–49, 1963. doi: 10.1007/BF00535296.
- Luciano de Castro, Antonio F Galvao, and Daniel Nunes. Dynamic economics with quantile preferences. *Theoretical Economics*, 20(1):353–425, 2025.
- Oliver de Groot, Alexander W Richter, and Nathaniel A Throckmorton. Valuation risk revalued. *Quantitative Economics*, in press, 2021.
- Eric V. Denardo. Contraction mappings in the theory underlying dynamic programming. *SIAM Review*, 9(2):165–177, 1967.
- Jean-Pierre Drugeon and Bertrand Wigniolle. On Markovian collective choice with heterogeneous quasi-hyperbolic discounting. *Economic Theory*, 72(4):1257–1296, 2021.
- Erik Ekström and Yuqiong Wang. Bayesian sequential composite hypothesis testing in discrete time. *ESAIM: Probability and Statistics*, 26:265–282, 2022.
- Tolulope Fadina, Yang Liu, and Ruodu Wang. A framework for measures of risk under uncertainty. *Finance and Stochastics*, 28:363–390, 2024. doi: 10.1007/s00780-024-00528-2.
- Jiequn Han, E Weinan, and Yucheng Yang. Deepham: A global solution method for heterogeneous agent models with aggregate shocks. *Quantitative Economics*, 17: 297–341, 2026.

- Onésimo Hernández-Lerma and Jean B Lasserre. *Discrete-time Markov control processes: basic optimality criteria*, volume 30. Springer Science & Business Media, 2012.
- Wen-Tai Hsu, Thomas J Holmes, and Frank Morgan. Optimal city hierarchy: a dynamic programming approach to central place theory. *Journal of Economic Theory*, 154:245–273, 2014.
- Fedor Iskhakov, John Rust, and Bertel Schjerning. Machine learning and structural econometrics: contrasts and synergies. *The Econometrics Journal*, 23(3):S81–S124, 2020.
- Anna Jaśkiewicz and Andrzej S Nowak. Markov decision processes with quasi-hyperbolic discounting. *Finance and Stochastics*, 25(2):189–229, 2021.
- Anna Jaśkiewicz, Andrzej S Nowak, and Łukasz Woźny. Stochastic games of risk-sensitive players with quasi-hyperbolic discounting. *Automatica*, 183:112656, 2026.
- Mykel J Kochenderfer, Tim A Wheeler, and Kyle H Wray. *Algorithms for decision making*. The MIT Press, 2022.
- Dennis Kristensen, Patrick K Mogensen, Jong Myun Moon, and Bertel Schjerning. Solving dynamic discrete choice models using smoothing and sieve methods. *Journal of Econometrics*, 223(2):328–360, 2021.
- Xiuxian Li and Lihua Xie. Online abstract dynamic programming with contractive models, 2021.
- Bar Light. Precautionary saving in a Markovian earnings environment. *Review of Economic Dynamics*, 29:138–147, 2018.
- Peng Liu, Alexander Schied, and Ruodu Wang. Distributional transforms, probability distortions, and their applications. *Mathematics of Operations Research*, 46(4): 1490–1512, 2021. doi: 10.1287/moor.2020.1090.
- Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, 2009.
- Mohammad Naghshvar and Tara Javidi. Active sequential hypothesis testing. *Annals of Statistics*, 41(6):2703–2738, 2013.

- Chengyuan Peng and John Stachurski. Dynamic programming in ordered vector space. *Operations Research*, 2026. doi: 10.1287/opre.2025.1971. Published online, 30 March 2026.
- Goran Peskir and Albert N. Shiryaev. *Optimal Stopping and Free-Boundary Problems*. Lectures in Mathematics. ETH Zürich. Birkhäuser, Basel, 2006.
- Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. Wiley Interscience, 2005.
- Pontus Rendahl. Fiscal policy in an unemployment crisis. *The Review of Economic Studies*, 83(3):1189–1224, 2016.
- John Rust. Optimal replacement of gmc bus engines: An empirical model of Harold Zurcher. *Econometrica*, pages 999–1033, 1987.
- John Rust. Structural estimation of Markov decision processes. *Handbook of Econometrics*, 4:3081–3143, 1994.
- Thomas J. Sargent and John Stachurski. *Dynamic Programming: Finite States*. Cambridge University Press, 2025a.
- Thomas J Sargent and John Stachurski. Dynamic programs on partially ordered sets. *SIAM Journal on Control and Optimization*, 63(2):778–795, 2025b.
- Albert N Shiryaev. *Optimal stopping rules*, volume 8. Springer Science & Business Media, 2007.
- Csaba Szepesvari. Non-markovian policies in sequential decision problems. *Acta Cybernetica*, 13(3):305–318, 1998.
- Carel van Niekerk, Renato Vukovic, Benjamin Matthias Ruppik, Hsien-chin Lin, and Milica Gašić. Post-training large language models via reinforcement learning from self-feedback. *arXiv preprint arXiv:2507.21931*, 2025.
- Abraham Wald. *Sequential Analysis*. John Wiley and Sons, New York, 1947.
- Huizhen Yu and Dimitri P. Bertsekas. Q-learning and policy iteration algorithms for stochastic shortest path problems. *Annals of Operations Research*, 208(1):95–132, 2013.
- Shenghao Zhu. Existence of stationary equilibrium in an incomplete-market model with endogenous labor supply. *International Economic Review*, 61(3):1115–1138, 2020.