COMPLETELY ABSTRACT DYNAMIC PROGRAMMING

THOMAS J. SARGENT AND JOHN STACHURSKI

Abstract. We introduce a "completely abstract" dynamic programming framework in which dynamic programs are sets of policy operators acting on a partially ordered space. We provide an optimality theory based on high-level assumptions. We then study symmetric and asymmetric relationships between dynamic programs, and show how these relationships transmit optimality properties. Our formulation includes and extends applications of dynamic programming across many fields.

1. Introduction

Dynamic programming is a form of optimization with a vast array of applications, from finance and artificial intelligence to the sequencing of DNA (Powell et al. (2012), Bertsekas (2021), Kochenderfer et al. (2022)). Dynamic programs control aircraft, route shipping, test products, recommend information on media platforms, and solve frontier research problems. Within economics, dynamic programming is applied to topics ranging from unemployment, monetary policy, and fiscal policy to asset pricing, firm investment, wealth dynamics, inventory control, commodity pricing, sovereign default, the division of labor, natural resource extraction, human capital accumulation, retirement decisions, portfolio choice, and dynamic pricing.

One impediment to progress in the field of dynamic programming has been absence of a common framework that can span the full range of traditional and modern applications. Abstract dynamic programming (ADP) is a major step towards removing this obstruction. The ADP framework has been driven primarily by Bertsekas (2022), although early ideas date back to Denardo (1967). ADP starts with a generalization of the Bellman equation, which is a recursive statement of optimality that exploits a temporal structure at the heart of dynamic programming. By abstracting the Bellman equation, one can closely analyze sets of properties required for standard optimality results. This idea has already been exploited in numerous applications (see, e.g., Ren and Stachurski (2021) and Bloise et al. (2023)).

Date: August 9, 2023.

The authors thank Jingni Yang and Shu Hu for valuable comments and suggestions. Financial support from Schmidt Futures and ARC grant FT160100423 is gratefully acknowledged.

In this paper we raise the level of abstraction over and beyond that found in existing ADP. While in Bertsekas (2022) value functions are functions and the Bellman equation is a relationship between elements of a function space, here we replace functions with elements of an abstract partially ordered set V. Policies are just indices over a family of "policy operators". The Bellman equation in Bertsekas (2022) is replaced by a fixed point problem for an order-preserving self-map on V.

This approach brings two main benefits. One is additional generality. Too see why this extra generality is valuable, consider the many modifications that have been made to standard Bellman equations across the spectrum of applications of dynamic programming. These include the "action-value" or "Q-factor" form of the Bellman equation (see, e.g., Kochenderfer et al. (2022)), the "expected value functions" and their corresponding Bellman equations in the literature on dynamic structural estimation (see, e.g., Rust (1994)), the "integrated value function" approach discussed in (Mogensen, 2018), and the "exponential Q-factor risk-sensitive" Bellman equations in Fei et al. (2021). Further examples can be found in Ma and Stachurski (2021) and many other sources. All of these variations can be viewed as dynamic programs in their own right under the "completely abstract" framework introduced here.

A second benefit of the "completely abstract" policy operator approach is that high level abstraction clarifies relationships between different dynamic programs. To facilitate analysis, we introduce two relationships: *isomorphic* (and anti-isomorphic) dynamic programs and *subordinate* dynamic programs. For example, we show that Q-factor dynamic programs are subordinate to their traditional counterparts and exponential Q factor dynamic programs are isomorphic to Q-factor dynamic programs. By classifying and illuminating relationships between dynamic programs, we can closely monitor how optimality in one setting translates into optimality in another, or whether convergence of a particular algorithm for one variation of a dynamic program implies convergence for another.

The framework developed below focuses on dynamics associated with policy operators. We study relationships between dynamic programs using a concept similar to topological conjugacy—a standard method for understanding connections between discrete and continuous time dynamical systems. Because order, rather than topology, is the essential structure that our analysis requires, we replace topological conjugacy with what we call "order conjugacy." This allows us to state straightforward optimality results and then to explore how these results transfer across closely related dynamic programs.

Section 2 introduces order conjugacy and discusses its properties. Sections 3–5 introduce "completely abstract" dynamic programs and prove various optimality properties. Section 6 introduces isomorphisms between abstract dynamic programs and shows how isomorphisms preserve optimality properties. Section 7 does the same for "subordinate" dynamic programs. Section 8 provides applications. Section 9 concludes.

2. Preliminaries

A *dynamical system* is a pair (V, S) where V is a set and S is a self-map on V. A system (V, S) is called *asymptotically stable* whenever V has a topology and S has a unique fixed point $\bar{v}$ in V and $\lim_{k\to\infty} S^k v = \bar{v}$ for all $v \in S$. Two dynamical systems (V, S) and $(\hat{V}, \hat{S})$ are called *topologically conjugate* under F if both V and $\hat{V}$ have topologies and there exists a homeomorphism F from V to $\hat{V}$ such that $F \circ S = \hat{S} \circ F$. The following results are elementary (see, e.g., Sternberg (2014) or Layek (2015)).

Lemma 2.1. *If (V, S) and $(\hat{V}, \hat{S})$ are topologically conjugate under F and S has a unique fixed point in $v \in V$, then Fv is the unique fixed point of $\hat{S}$ in $\hat{V}$. Moreover, (V, S) is asymptotically stable if and only if $(\hat{V}, \hat{S})$ is asymptotically stable.*

While some of the ideas in Lemma 2.1 are central to our methodology, the foundational aspects of dynamic programming are driven by order, not topology. For this reason, we now drop the assumption that $V, \hat{V}$ have topologies and propose order-theoretic counterparts of these ideas.[1]

Let (V, S) be a dynamical system where V is partially ordered by $\preceq$ and S has a unique fixed point $\bar{v}$ in V. In this setting, we call (V, S)

(i) *upward stable* if $v \in V$ and $v \preceq S\,v$ implies $v \preceq \bar{v}$,
(ii) *downward stable* if $v \in V$ and $S\,v \preceq v$ implies $\bar{v} \preceq v$, and
(iii) *order stable* if upward and downward stability both hold.

(Unless otherwise specified, in what follows, we use $\preceq$ to represent the partial order on all partially ordered spaces.)

Order stability can be thought of as an extension of asymptotic stability to dynamics on partially ordered sets. It is implied by asymptotic stability when S is order-preserving and the space V has both a topology and a partial order:

[1]For definitions of basic order-theoretic concepts used below, see, e.g., Davey and Priestley (2002).

Example 2.1. Let $(V, \preceq, \rho)$ be a partially ordered space (i.e., a metric space (V, ρ) where $\preceq$ is closed under limits). If (V, S) is asymptotically stable and S is order-preserving, then (V, S) is order stable. To see this let $\bar{v}$ be the unique fixed point of S in V. Upward stability holds because 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 $v \preceq S^k\,v$ for all $k \in \mathbb{N}$. Applying asymptotic stability and taking the limit gives $v \preceq \bar{v}$. Hence upward stability holds. The proof of downward stability is similar.

Given partially ordered set $V = (V, \preceq)$, let $V^\partial = (V, \preceq^\partial)$ be the order dual, so that, for $u, v \in V$, we have $u \preceq^\partial v$ if and only if $v \preceq u$.

Lemma 2.2. *(V, S) is order stable if and only if (V^∂, S) is order stable.*

Proof. Let (V, S) be order stable. By definition, S has a unique fixed point $\bar{v} \in V$. We claim that (V^∂, S) is upward and downward stable. Regarding upward stability, suppose $v \in V$ and $v \preceq^\partial Sv$. Then $Sv \preceq v$ and hence $\bar{v} \preceq v$, by downward stability of (V, S). But then $v \preceq^\partial \bar{v}$, so (V^∂, S) is upward stable. The proof of downward stability is similar. Hence order stability of (V, S) is sufficient for order stability of (V^∂, S). Necessity follows from sufficiency, since the dual of (V^∂, S) is (V, S). □

A bijective map $F \colon V \to \hat{V}$ is called an *order isomorphism* if $u \preceq v$ if and only if $Fu \preceq Fv$. Let (V, S) and $(\hat{V}, \hat{S})$ be two dynamical systems where V and $\hat{V}$ are partially ordered. We call (V, S) and $(\hat{V}, \hat{S})$ *order conjugate* under F when there exists an order isomorphism F from V to $\hat{V}$ such that $F \circ S = \hat{S} \circ F$. It is easy to verify that order conjugacy is an equivalence relation on the set of dynamical systems over partially ordered sets. The next lemma modifies Lemma 2.1 to an order setting.

Lemma 2.3. *If (V, S) and $(\hat{V}, \hat{S})$ are order conjugate under F and S has a unique fixed point in $v \in V$, then Fv is the unique fixed of $\hat{S}$ in $\hat{V}$. Moreover, (V, S) is order stable if and only if $(\hat{V}, \hat{S})$ is order stable.*

Proof. Let (V, S) and $(\hat{V}, \hat{S})$ be order conjugate under F, let $\mathrm{fix}(S)$, the set of fixed points of S, be equal to $\{v\}$ and let $\hat{v} = Fv$. Clearly $\hat{v} \in \mathrm{fix}(\hat{S})$ because $\hat{S}Fv = FSv = Fv$. Also, if $\hat{w}$ is any fixed point of $\hat{S}$ in $\hat{V}$, then $F^{-1}\hat{w} \in \mathrm{fix}(S)$, since $SF^{-1}\hat{w} = F^{-1}\hat{S}\hat{w} = F^{-1}\hat{w}$. Hence $F^{-1}\hat{w} = v$ and, therefore, $\hat{w} = Fv = \hat{v}$. It follows that $\mathrm{fix}(\hat{S}) = \{\hat{v}\}$.

Regarding stability, let (V, S) be order stable and let $\hat{w}$ be an element of $\hat{V}$ satisfying $\hat{S}\hat{w} \preceq \hat{w}$. Then $F^{-1}\hat{S}\hat{w} \preceq F^{-1}\hat{w}$ and hence $SF^{-1}\hat{w} \preceq F^{-1}\hat{w}$. But then $v \preceq F^{-1}\hat{w}$, by downward stability of (V, S). Applying F gives $\hat{v} \preceq \hat{w}$. Hence $(\hat{V}, \hat{S})$ is downward stable. Similarly, if $\hat{w}$ is an element of $\hat{V}$ satisfying $\hat{w} \preceq \hat{S}\hat{w}$, then $F^{-1}\hat{w} \preceq F^{-1}\hat{S}\hat{w} =$

$SF^{-1}\hat{w}$. By upward stability of (V, S), we have $F^{-1}\hat{w} \preceq v$. Applying F gives $\hat{w} \preceq \hat{v}$, so $(\hat{V}, \hat{S})$ is upward stable. Together, these results show that $(\hat{V}, \hat{S})$ is order stable. □

Let (V, S) and $(\hat{V}, \hat{S})$ be a pair of dynamical systems, where V and $\hat{V}$ are partially ordered sets. We call (V, S) and $(\hat{V}, \hat{S})$ *mutually (order) semiconjugate* if there exist order-preserving maps $F \colon V \to \hat{V}$ and $G \colon \hat{V} \to V$ such that

$$S = G \circ F \quad \text{and} \quad \hat{S} = F \circ G \tag{1}$$

The "order semiconjugate" terminology comes from the fact that, when (1) holds,

$$F \circ S = \hat{S} \circ F \quad \text{and} \quad G \circ \hat{S} = S \circ G. \tag{2}$$

(Evidently, if either F or G is also an order isomorphism, then (V, S) and $(\hat{V}, \hat{S})$ are order conjugate.)

Lemma 2.4. *Let (V, S) and $(\hat{V}, \hat{S})$ be mutually semiconjugate under F, G in (1). If, in this setting, S has a unique fixed point in $v \in V$, then Fv is the unique fixed point of $\hat{S}$ in $\hat{V}$. If, in addition, (V, S) is order stable, then $(\hat{V}, \hat{S})$ is order stable.*

Proof. Suppose $\text{fix}(S) = \{v\}$. Then $\hat{S}Fv = FSv = Fv$, so $Fv \in \text{fix}(\hat{S})$. Now suppose $\hat{v} \in \text{fix}(\hat{S})$. Then $FG\hat{v} = \hat{v}$ and hence $GFG\hat{v} = G\hat{v}$, or $SG\hat{v} = G\hat{v}$. But $\text{fix}(S) = \{v\}$, so then $G\hat{v} = v$. Applying F gives $\hat{S}\hat{v} = Fv$. But $\hat{v} \in \text{fix}(\hat{S})$, so $\hat{v} = Fv$. This proves the first claim in Lemma 2.4.

Now suppose that (V, S) is order stable, with unique fixed point $v \in V$. Then, by the preceding argument, Fv is the unique fixed point of $\hat{S}$ in $\hat{V}$. The pair $(\hat{V}, \hat{S})$ is upward stable because if $\hat{v} \in \hat{V}$ and $\hat{v} \preceq \hat{S}\hat{v}$, then $G\hat{v} \preceq G\hat{S}\hat{v} = SG\hat{v}$ and so, by upward stability of (V, S), $G\hat{v} \preceq v$. Applying F gives $\hat{S}\hat{v} \preceq Fv$, so $(\hat{V}, \hat{S})$ is upward stable. The proof of downward stability is similar. □

3. Abstract Dynamic Programs

In this section we represent dynamic programs in terms of their policy operators. This abstract representation allows us to apply the notion of (order) conjugacy to study relationships between dynamic programs.

3.1. Definition and Examples.

We define an *abstract dynamic program* (*ADP*) to be a pair $\mathscr{A} = (V, \{T_\sigma\}_{\sigma \in \Sigma})$, where

(i) $V = (V, \preceq)$ is a partially ordered set and
(ii) $\{T_\sigma\}_{\sigma \in \Sigma}$ is a family of self-maps on V, indexed by $\sigma \in \Sigma$.

Elements of the index set Σ will be referred to as *policies* and elements of $\{T_\sigma\}$ are called *policy operators*. When Σ is understood, we often write $\{T_\sigma\}_{\sigma \in \Sigma}$ as $\{T_\sigma\}$.

In all applications, the significance of policy operator T_σ is that its fixed point represents the lifetime value (or cost) of following policy σ. (Below we impose conditions to ensure existence and uniqueness of fixed points.) The objective of the ADP is to maximize (or minimize) this lifetime value. This point is clarified in Section 3.2.

Example 3.1 (MDPs)**.** Consider a *Markov decision process* (MDP; see, e.g., Puterman (2005)) where the objective is to maximize $\mathbb{E} \sum_{t \geqslant 0} \beta^t r(X_t, A_t)$ when

- X_t takes values in finite set X (the state space),
- A_t takes values in finite set A (the action space),
- Γ is a nonempty correspondence from X to A (feasible correspondence),
- $\mathsf{G} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\}$ denotes the feasible state-action pairs,
- r is a reward function defined on G,
- $\beta \in (0, 1)$ is a discount factor, and
- $P \colon \mathsf{G} \times \mathsf{X} \to [0, 1]$ provides transition probabilities and $\sum_{x'} P(x, a, x') = 1$.

The Bellman equation for this problem is

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \sum_{x'} v(x') P(x, a, x') \right\} \qquad (x \in \mathsf{X}). \tag{3}$$

We define the set of *feasible policies* to be $\Sigma := \{\sigma \in \mathsf{A}^\mathsf{X} : \sigma(x) \in \Gamma(x) \text{ for all } x \in \mathsf{X}\}$. We combine $\mathbb{R}^\mathsf{X}$ (the set of all real-valued functions on X) with the pointwise partial order $\leqslant$ and, for $\sigma \in \Sigma$ and $v \in \mathbb{R}^\mathsf{X}$, define the MDP policy operator

$$(T_\sigma \, v)(x) = r(x, \sigma(x)) + \beta \sum_{x'} v(x') P(x, \sigma(x), x') \tag{4}$$

The pair $(\mathbb{R}^\mathsf{X}, \{T_\sigma\})$ is an ADP. The main significance of T_σ is that the lifetime value of the policy σ is the unique fixed point of this operator, a point we return to below.

Example 3.2 (MDPs with minimization)**.** If $r(x, a)$ is instead interpreted as a cost function and our aim is to minimize lifetime costs, we still have an identical expression for the lifetime value of a policy and the policy operators are still given by (4). Hence, for this MDP minimization problem, we work with the same ADP $(\mathbb{R}^\mathsf{X}, \{T_\sigma\})$.

Example 3.3 (Q-learning)**.** Continuing with the MDP setting of Example 3.1, the Q-learning literature studies the Q-factor Bellman equation, which is given by

$$f(x, a) = r(x, a) + \beta \sum_{x'} \max_{a' \in \Gamma(x')} f(x', a') P(x, a, x') \qquad ((x, a) \in \mathsf{G}). \tag{5}$$

Here $f \in \mathbb{R}^{\mathsf{G}}$ is called the *Q-factor*. The policy operators over Q-factors take the form

$$(Q_\sigma f)(x,a) = r(x,a) + \beta \sum_{x'} f(x', \sigma(x')) P(x,a,x'), \tag{6}$$

where $f \in \mathbb{R}^{\mathsf{G}}$ and $\sigma \in \Sigma$. If we pair $\mathbb{R}^{\mathsf{G}}$ (the set of all real-valued functions on G) with the pointwise partial order $\leqslant$ and Q_σ as in (6), then $(\mathbb{R}^{\mathsf{G}}, \{Q_\sigma\})$ is an ADP.

Example 3.4 (Risk-sensitive MDPs)**.** Risk-sensitive MDPs (Howard and Matheson, 1972) modify the MDP model in Example 3.1 so that the policy operators take the form

$$(T_\sigma^\theta v)(x) = r(x, \sigma(x)) + \frac{\beta}{\theta} \ln \left[\sum_{x'} \exp(\theta v(x')) P(x, \sigma(x), x') \right]$$

where θ is some fixed value in $\Theta := \mathbb{R} \setminus \{0\}$. The pair $(\mathbb{R}^{\mathsf{X}}, \{T_\sigma^\theta\})$ is an ADP.

Example 3.5 (Risk-sensitive Q-learning)**.** It has become popular in reinforcement learning and related fields to extend the Q-factor approach from Example 3.3 to risk-sensitive decision processes (see, e.g., Fei et al. (2021). The corresponding Q-factor Bellman equation is given by

$$f(x,a) = r(x,a) + \frac{\beta}{\theta} \ln \left\{ \sum_{x'} \exp \left[\theta \max_{a' \in \Gamma(x')} f(x', a') \right] P(x,a,x') \right\} \qquad ((x,a) \in \mathsf{G}). \tag{7}$$

The policy operators over risk-sensitive Q-factors take the form

$$(Q_\sigma^\theta f)(x,a) = r(x,a) + \frac{\beta}{\theta} \ln \left[\sum_{x'} \exp \left[\theta f(x', \sigma(x')) \right] P(x,a,x') \right] \tag{8}$$

where $f \in \mathbb{R}^{\mathsf{G}}$ and $\sigma \in \Sigma$. The pair $(\mathbb{R}^{\mathsf{G}}, \{Q_\sigma^\theta\})$ is an ADP.

Example 3.6 (MDPs via expected values)**.** Returning again to the MDP setting of Example 3.1, the literature on dynamic structural estimation often works with the expected value function $g(x,a) := \sum_{x'} v(x') P(x,a,x')$ rather than the value function v, partly because the corresponding Bellman equation

$$g(x,a) = \sum_{x'} \left\{ \max_{a' \in \Gamma(x')} \left[r(x',a') + \beta g(x',a') \right] \right\} P(x,a,x') \qquad ((x,a) \in \mathsf{G})$$

shifts the expectation outside the maximum (which facilitates the "Gumbel max trick"—see, e.g., Rust (1987, 1994)). We can set this problem up as an ADP by introducing the expected value policy operator

$$(R_\sigma g)(x,a) = \sum_{x'} \left\{ r(x', \sigma(x')) + \beta g(x', \sigma(x')) \right\} P(x,a,x') \tag{9}$$

The pair $(\mathbb{R}^{\mathsf{G}}, \{R_\sigma\})$ is an ADP.

3.2. **Lifetime Values.** The objective of dynamic programming is to maximize or minimize lifetime value. In the present setting, we identify lifetime value of policy σ as the unique fixed point of T_σ whenever it exists. When it does exist, we denote it by v_σ and call it the *σ-value function*.

To illustrate, consider the MDP setting of Example 3.1. Let r_σ and P_σ be defined by

$$P_\sigma(x, x') := P(x, \sigma(x), x') \quad \text{and} \quad r_\sigma(x) := r(x, \sigma(x)). \tag{10}$$

The lifetime value of policy σ given $X_0 = x$ is $v_\sigma(x) = \mathbb{E} \sum_{t \geqslant 0} \beta^t r(X_t, \sigma(X_t))$, where $(X_t)_{t \geqslant 0}$ is a Markov chain generated by P_σ with initial condition $X_0 = x \in \mathsf{X}$. Pointwise on X, we can express v_σ as

$$v_\sigma = \sum_{t \geqslant 0} (\beta P_\sigma)^t r_\sigma = (I - \beta P_\sigma)^{-1} r_\sigma \tag{11}$$

(see, e.g., Puterman (2005), Theorem 6.1.1). Equivalently, v_σ is the unique solution to the equation $v = r_\sigma + \beta P_\sigma v$. Inspecting the definition of T_σ in (4), we see this is also equivalent to the statement that v_σ is the unique fixed point of T_σ.

We discuss other examples below.

3.3. **Greedy Policies.** Given $v \in V$, a policy σ in Σ is called

- *v-min-greedy* if $T_\sigma v \preceq T_\tau v$ for all $\tau \in \Sigma$, and
- *v-max-greedy* if $T_\sigma v \succeq T_\tau v$ for all $\tau \in \Sigma$.

In the context of Example 3.1, a max-greedy policy can be constructed as follows: Given $v \in V$, let σ be any policy satisfying

$$\sigma(x) \in \operatorname{arg\,max} \left\{ r(x, a) + \beta \sum_{x'} v(x') P(x, a, x') \right\} \tag{12}$$

for all $x \in X$. This policy satisfies $T_\sigma v \geqslant T_\tau v$ for all $\tau \in \Sigma$, and hence is v-max-greedy. A v-min-greedy policy can be constructed by replacing arg max with arg min.

3.4. **Bellman Operators.** For a generic ADP $(V, \{T_\sigma\})$, we define the *Bellman min-operator* and the *Bellman max-operator* via

$$T_\perp v := \bigwedge_\sigma T_\sigma v \quad \text{and} \quad T_\top v := \bigvee_\sigma T_\sigma v \tag{13}$$

whenever the infimum (resp., supremum) exists. We say that $v \in V$ satisfies the *Bellman max-equation* (resp., the *Bellman min-equation*) if it is a fixed point of $T_\top$ (resp., $T_\perp$). Notice that $\sigma \in \Sigma$ is

(i) v-max-greedy if and only if $T_\sigma \, v = T_\top \, v$, and
(ii) v-min-greedy if and only if $T_\sigma \, v = T_\perp \, v$.

To illustrate, consider the MDP setting of Example 3.1. Traditionally, the Bellman operator for this model is given by

$$(Tv)(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{x'} v(x') P(x,a,x') \right\} \qquad (x \in \mathsf{X}). \tag{14}$$

For the ADP $(\mathbb{R}^{\mathsf{X}}, \{T_\sigma\})$ generated by this MDP, this Bellman operator exactly coincides with the Bellman max-operator $T_\top$ in (13). Replacing max with min in (14) produces the Bellman min-operator from (13).

4. Properties

To obtain optimality results, we need to place structure on ADPs. Here and below, we call an ADP $\mathscr{A} = (V, \{T_\sigma\})$

- *well-posed* if T_σ has a unique fixed point v_σ in V for each $\sigma \in \Sigma$,
- *order stable* if (V, T_σ) is order stable for each $\sigma \in \Sigma$,
- *max-stable* if $\mathscr{A}$ is order stable, each $v \in V$ has at least one max-greedy policy, and $T_\top$ has at least one fixed point in V, and
- *min-stable* if $\mathscr{A}$ is order stable, each $v \in V$ has at least one min-greedy policy, and $T_\perp$ has at least one fixed point in V.

Well-posedness is a minimum regularity condition for ADPs. Without it, we cannot be sure that policies have well-defined lifetime values. Well-defined lifetime values are essential because maximizing (or minimizing) lifetime value over the set of all policies is the objective of dynamic programming.

Example 4.1 (Continuous time MDPs). We return to the MDP setting of Example 3.1 but replacing the discount factor β with a discount rate $\delta > 0$ and the stochastic kernel P with an intensity kernel Q from $\mathsf{G} \times \mathsf{X} \to \mathbb{R}$ satisfying $\sum_{x' \in \mathsf{X}} Q(x,a,x') = 0$ for all (x,a) in G and $Q(x,a,x') \geqslant 0$ whenever $x \neq x'$. Lifetime value of policy σ is given by

$$v_\sigma(x) = \mathbb{E}_x \int_0^\infty \exp(-\delta t) r(X_t, \sigma(X_t)) \, \mathrm{d}t. \tag{15}$$

where $Q_\sigma(x,x') := Q(x, \sigma(x), x')$ is the infinitesimal generator of the continuous time Markov chain $(X_t)_{t \geqslant 0}$ and $r_\sigma(x) := r(x, \sigma(x))$. The function v_σ can alternatively be written as $v_\sigma = (\delta I - Q_\sigma)^{-1} r_\sigma$ (see, e.g., Guo and Hernández-Lerma (2009), Lemma 4.16). Rearranging this expression and using the fact that $\delta I - Q_\sigma$ is bijective shows that v_σ is the unique fixed point of $T_\sigma v = r_\sigma + (Q_\sigma + (1-\delta)I)v$ in $\mathbb{R}^\mathsf{X}$. Hence $(\mathbb{R}^\mathsf{X}, \{T_\sigma\})$ is a well-posed ADP.

Order stability is a natural regularity assumption for ADPs. To understand this, suppose $\mathscr{A}$ is well-posed and consider upward stability for an arbitrary policy operator T_σ with fixed point v_σ. If $v \preceq T_\sigma v$, then following policy σ for one period offers an improvement in value. Since the problem is stationary, this suggests that following the policy forever will also be an improvement. Thus, we expect $v \preceq v_\sigma$, in which case upward stability holds. Intuition for downward stability is similar.

Example 4.2. Consider the MDP setting of Example 3.1 and fix $\sigma \in \Sigma$. Recall that the policy operator T_σ has unique fixed point $v_\sigma := (I - \beta P_\sigma)^{-1} r_\sigma$. If $T_\sigma v \geqslant v$, then $r_\sigma + \beta P_\sigma v \geqslant v$ and hence $(I - \beta P_\sigma) v \leqslant r_\sigma$. Since $(I - \beta P_\sigma)^{-1}$ is positive, we get $v \leqslant v_\sigma$. Hence (V, T_σ) is upward stable. A similar proof shows that (V, T_σ) is downward stable, and therefore order stable.

Example 4.3. Consider the ADP $(\mathbb{R}^\mathsf{X}, \{T_\sigma\})$ from the MDP setting of Example 3.1. For this case $T_\top$ is given by (14). Since $T_\top$ is a contraction on $\mathbb{R}^\mathsf{X}$, it has a unique fixed point in $\mathbb{R}^\mathsf{X}$. Since max-greedy policies always exist, and since $(\mathbb{R}^\mathsf{X}, \{T_\sigma\})$ is order stable by Example 4.2, we see that $(\mathbb{R}^\mathsf{X}, \{T_\sigma\})$ is max-stable.

Regarding the definition of max-stability (resp., min-stability), the Bellman min- and max-operators are often contraction maps and existence of a fixed point is easily verified (see, e.g, (Denardo, 1967) or Chapter 2 of Bertsekas (2022)). Here is another useful condition, which covers the case where state and action spaces are finite.

Proposition 4.1. *Let $\mathscr{A}$ be order stable and suppose the set of policies is finite. In this setting,*

(i) *if max-greedy always exist, then $\mathscr{A}$ is max-stable, and*
(ii) *if min-greedy always exist, then $\mathscr{A}$ is min-stable.*

Proposition 4.1 is proved in the appendix (page 26).

Example 4.4. The risk-sensitive MDP from Example 3.4 is max-stable. Evidently T_σ is order-preserving on $V = \mathbb{R}^\mathsf{X}$. Moreover, (V, T_σ) is asymptotically stable (see, e.g., Bäuerle and Jaśkiewicz (2018)) and hence order stable, by Example 2.1. This shows that $(V, \{T_\sigma\})$ is an order stable ADP. Given $v \in V$, we can construct a v-max-greedy policy σ by setting

$$\sigma(x) \in \arg\max \left\{ r(x,a) + \frac{\beta}{\theta} \ln \left[\sum_{x'} \exp(\theta v(x')) P(x,a,x') \right] \right\}$$

for all $x \in X$. As the policy set Σ is finite (since X and the choice sets are all finite), Proposition 4.1 implies that $(V, \{T_\sigma\})$ is max-stable.

5. Optimality

In this section we provide conditions for optimality in our abstract setting. While these conditions are interesting in their own right, our main purpose is to study how optimality properties are preserved under transformations (which is the task of Sections 6–7).

5.1. Max-Optimality.

Let $\mathscr{A}$ be a well-posed ADP. We let $V_\Sigma := \{v_\sigma\}_{\sigma \in \Sigma}$ denote the set of σ-value functions. If V_Σ has a greatest element, then we denote it by $v_\top$ and call it the *max-value function*. A policy $\sigma \in \Sigma$ is called *max-optimal* for $\mathscr{A}$ if v_σ is a greatest element of V_Σ; that is, if $v_\top$ exists and $v_\sigma = v_\top$.

We define a map $H_\top$ from V to $\{v_\sigma\}$ via $H_\top\, v = v_\sigma$ where σ is v-max-greedy. Iterating with $H_\top$ is an abstraction of Howard policy iteration.[2] In what follows, we call $H_\top$ the *Howard max-operator* generated by the ADP.

In the following result, we take $\mathscr{A}$ to be an ADP with Bellman operator $T_\top$.

Theorem 5.1 (Max-optimality)**.** *If $\mathscr{A}$ is max-stable, then*

(i) *the max-value function $v_\top$ exists in V,*
(ii) *$v_\top$ is the unique solution to the Bellman max-equation in V,*
(iii) *a policy is max-optimal if and only if it is $v_\top$-max-greedy.*
(iv) *at least one max-optimal policy exists.*

If, in addition, Σ is finite, then Howard max-policy iteration converges to $v_\top$ in finitely many steps.

The last statement means that, for all $v \in V$, there exists a $K \in \mathbb{N}$ such that $k \geqslant K$ implies $H_\top^k v = v_\top$. The proof of Theorem 5.1 is given in the appendix.

[2]For $H_\top$ to be well-defined, we must always select the same v-greedy policy when the operator is applied to v. We can use the axiom of choice to assign to each v a designated v-greedy policy, although, in applications, a simple rule usually suffices. For example, if Σ is finite, we can enumerate the policy set Σ and choose the first v-greedy policy.

5.2. **Min-Optimality.** In the present abstract setting, minimization results are readily recovered from maximization results by order duality. This section illustrates.

We call a policy $\sigma \in \Sigma$ *min-optimal* for $\mathscr{A}$ if v_σ is a least element of V_Σ. When V_Σ has a least element we denote it by $v_\perp$ and call it the *min-value function* We define $H_\perp$ from V to $\{v_\sigma\}$ via $H_\perp\, v = v_\sigma$ where σ is v-min-greedy and call $H_\perp$ the *Howard min-operator* generated by $\mathscr{A}$.

Below, if $\mathscr{A} := (V, \{T_\sigma\})$ is an ADP then its *dual* $\mathscr{A}^\partial$ is the ADP $(V^\partial, \{T_\sigma\})$ where the partial order $\preceq$ on V is replaced with its dual $\preceq^\partial$. In this setting, we let $T^\partial_\top$ be the Bellman max-operator for $\mathscr{A}^\partial$, $v^\partial_\top$ be the max-value function for $\mathscr{A}^\partial$, and so on.

Lemma 5.2. *$\mathscr{A}$ is min-stable if and only if $\mathscr{A}^\partial$ is max-stable, in which case $T_\perp = T^\partial_\top$ and $H_\perp = H^\partial_\top$. A policy σ is max-optimal for $\mathscr{A}$ if and only if σ is min-optimal for $\mathscr{A}^\partial$.*

Proof. Let $\mathscr{A}$ be min-stable. Then $\mathscr{A}^\partial$ is order stable, by Lemma 2.2. Now fix $v \in V$ and suppose that σ is min-greedy for $\mathscr{A}$, so that $T_\sigma\, v \preceq T_\tau\, v$ for all $\tau \in \Sigma$. Then $T_\sigma\, v \succeq^\partial T_\tau\, v$ for all $\tau \in \Sigma$, so σ is v-max-greedy for $\mathscr{A}^\partial$ and $T^\partial_\top v = T_\sigma v = T_\perp v$. We have proved that $\mathscr{A}^\partial$ is max-stable and $T^\partial_\top = T_\perp$. The remaining steps follow easily from the definitions. □

Results analogous to Theorem 5.1 hold for minimization.

Theorem 5.3 (Min-optimality)**.** *If $\mathscr{A}$ is min-stable, then*

(i) *the min-value function $v_\perp$ exists in V,*
(ii) *$v_\perp$ is the unique solution to the Bellman min-equation in V,*
(iii) *a policy is min-optimal if and only if it is $v_\perp$-min-greedy.*
(iv) *at least one min-optimal policy exists.*

If, in addition, Σ is finite, then Howard min-policy iteration converges to $v_\perp$ in finitely many steps.

Proof. Let $\mathscr{A}$ be min-stable. By Lemma 5.2, the dual $\mathscr{A}^\partial$ is max-stable. Hence, by Theorem 5.1, $v^\partial_\top$ exists in V. But then $v_\perp$ exists in V and is equal to $v^\partial_\top$, since $v_\perp = \bigwedge_\sigma v_\sigma = \bigvee^\partial_\sigma v_\sigma = v^\partial_\top$. Also, by Theorem 5.1, $v^\partial_\top$ is the unique solution to $T^\partial_\top v^\partial_\top = v^\partial_\top$. Applying Lemma 5.2, we see that $T_\perp\, v_\perp = v_\perp$. The result of the proof is similar. □

6. Isomorphic ADPs

In this section we introduce isomorphic relationships between ADPs and explore their implications for optimality. True to their name, isomorphic relationships are symmetric, transitive and reflexive. We show that isomorphic ADPs have identical optimality properties.

6.1. Definition and Properties.

Let $\mathscr{A} = (V, \{T_\sigma\})$ and $\hat{\mathscr{A}} = (\hat{V}, \{\hat{T}_\sigma\})$ be two ADPs. We call $\mathscr{A}$ and $\hat{\mathscr{A}}$ *isomorphic* under F if these two ADPs have the same policy set Σ and F is an order isomorphism from V to $\hat{V}$ such that

$$F \circ T_\sigma = \hat{T}_\sigma \circ F \quad \text{on } V \text{ for all } \sigma \in \Sigma. \tag{16}$$

In other words, (V, T_σ) and $(\hat{V}, \hat{T}_\sigma)$ are order conjugate under F for all $\sigma \in \Sigma$.[3]

Example 6.1. Fei et al. (2021) work with an "exponential" risk-sensitive Q-factor Bellman equation that has policy operator

$$(M_\sigma\, h)(x,a) = \exp\left\{\theta r(x,a) + \beta \ln\left[\sum_{x'} h(x', \sigma(x'))P(x,a,x')\right]\right\}$$

that maps $\mathbb{R}^{\mathsf{G}}_{++}$, the set of strictly positive functions in $\mathbb{R}^{\mathsf{G}}$, into itself. All primitives are as in the risk-sensitive Q-factor ADP $\mathscr{A} := (\mathbb{R}^{\mathsf{G}}, \{Q_\sigma\})$ in Example 3.5. Let F be the order isomorphism from $\mathbb{R}^{\mathsf{G}}$ to $\mathbb{R}^{\mathsf{G}}_{++}$ defined by $(Fh)(x,a) = \exp(\theta h(x,a))$. Then, for Q_σ defined in (8), $h \in \mathbb{R}^{\mathsf{G}}$ and $(x,a) \in \mathsf{G}$,

$$(FQ_\sigma\, h)(x,a) = \exp\left\{\theta r(x,a) + \beta \ln\left[\sum_{x'} \exp\left[\theta h(x', \sigma(x'))\right] P(x,a,x')\right]\right\},$$

which is equal to $(M_\sigma F h)(x,a)$. Thus, $F \circ Q_\sigma = M_\sigma \circ F$ on $\mathbb{R}^{\mathsf{G}}$. Hence $\hat{\mathscr{A}} := (\mathbb{R}^{\mathsf{G}}_{++}, \{M_\sigma\})$ is an ADP and $\mathscr{A}$ and $\hat{\mathscr{A}}$ are isomorphic.

For ADPs $\mathscr{A}, \hat{\mathscr{A}}$, let $\mathscr{A} \sim \hat{\mathscr{A}}$ indicate that $\mathscr{A}$ and $\hat{\mathscr{A}}$ are isomorphic. It is elementary to show that the relation $\sim$ is reflexive, symmetric and transitive. Hence $\sim$ is an equivalence relation on the set of all ADPs.

---

[3] While the definition require that the two ADPs have the same policy set Σ, it suffices that the policy sets can be put in one-to-one correspondence with each other.

6.2. **Isomorphisms and Optimality.** We seek a connection between value functions and optimality properties of isomorphic ADPs. The theory below provides this relationship. For all of this section, we take $\mathscr{A} = (V, \{T_\sigma\})$ and $\hat{\mathscr{A}} = (\hat{V}, \{\hat{T}_\sigma\})$ to be two ADPs with the same policy set. When they exist, we let

- v_σ (resp., $\hat{v}_\sigma$) be the unique fixed point of T_σ (resp., $\hat{T}_\sigma$)
- $T_\top$ (resp., $\hat{T}_\top$) be the max-Bellman operator of $\mathscr{A}$ (resp., $\hat{\mathscr{A}}$)
- $T_\bot$ (resp., $\hat{T}_\bot$) be the min-Bellman operator of $\mathscr{A}$ (resp., $\hat{\mathscr{A}}$)
- $v_\top$ (resp., $\hat{v}_\top$) be the max-value function of $\mathscr{A}$ (resp., $\hat{\mathscr{A}}$)
- $v_\bot$ (resp., $\hat{v}_\bot$) be the min-value function of $\mathscr{A}$ (resp., $\hat{\mathscr{A}}$)

Isomorphic ADPs share the same regularity properties:

Theorem 6.1. *If $\mathscr{A}$ and $\hat{\mathscr{A}}$ are isomorphic under F, then*

(i) *$\mathscr{A}$ is well-posed if and only if $\hat{\mathscr{A}}$ is well-posed.*
(ii) *$\mathscr{A}$ is order stable if and only if $\hat{\mathscr{A}}$ is order stable.*
(iii) *$\mathscr{A}$ is max-stable if and only if $\hat{\mathscr{A}}$ is max-stable. In this case,*

$$F \circ T_\top = \hat{T}_\top \circ F \quad \text{and} \quad \hat{v}_\top = F\, v_\top. \tag{17}$$

Moreover, $\mathscr{A}$ and $\hat{\mathscr{A}}$ have the same max-optimal policies.

(iv) *$\mathscr{A}$ is min-stable if and only if $\hat{\mathscr{A}}$ is min-stable. In this case,*

$$F \circ T_\bot = \hat{T}_\bot \circ F \quad \text{and} \quad \hat{v}_\bot = F\, v_\bot. \tag{18}$$

Moreover, $\mathscr{A}$ and $\hat{\mathscr{A}}$ have the same min-optimal policies.

Proof. Claims (i)–(ii) follow directly from Lemma 2.3. Regarding (iii), suppose $\mathscr{A}$ is max-stable. We claim that, for $\hat{\mathscr{A}}$, max-greedy policies always exist. To see this, fix $\hat{v} \in \hat{V}$. Since $\mathscr{A}$ is max-stable, we can choose σ to be $F^{-1}\hat{v}$-max-greedy, so that $T_\tau\, F^{-1}\hat{v} \preceq T_\sigma\, F^{-1}\hat{v}$ for all $\tau \in \Sigma$. Then $F^{-1}\hat{T}_\tau\, \hat{v} \preceq F^{-1}\hat{T}_\sigma\, \hat{v}$ and hence $\hat{T}_\tau\, v \preceq \hat{T}_\sigma\, v$ for all $\tau \in \Sigma$. In particular, σ is $\hat{v}$-max-greedy.

Continuing to assume that $\mathscr{A}$ is max-stable, we now prove (17). For given $v \in V$, applying the order conjugacy (16) yields

$$T_\top v = \bigvee_\sigma T_\sigma\, v = \bigvee_\sigma F^{-1}\hat{T}_\sigma\, F\, v = F^{-1} \bigvee_\sigma \hat{T}_\sigma\, F\, v = F^{-1}\hat{T}_\top\, F\, v,$$

which is equivalent to $F \circ T_\top = \hat{T}_\top \circ F$ from (17), and implies that $(V, T_\top)$ and $(\hat{V}, \hat{T}_\top)$ are order conjugate under F. By max-stability of $\mathscr{A}$ and Theorem 5.1, the operator $T_\top$ has unique fixed point $v_\top$ in V. Lemma 2.3 then implies that $\hat{T}_\top$ has unique fixed point

$F v_\top$ in $\hat{V}$. This completes the proof that $\hat{\mathscr{A}}$ is max-stable. By max-stability of $\hat{\mathscr{A}}$, the unique fixed point of $\hat{T}_\top$ in $\hat{V}$ is $\hat{v}_\top$, so $F v_\top = \hat{v}_\top$ and both claims in (17) are verified. Finally, note that σ is max-optimal for $\mathscr{A}$ if and only if $T_\sigma v_\top = v_\top$, which, by the bijection property of F, is also equivalent to $F T_\sigma v_\top = \hat{v}_\top$. Using $F \circ T_\sigma = \hat{T}_\sigma \circ F$, we can write this as $\hat{T}_\sigma \hat{v}_\top = \hat{v}_\top$, which is equivalent to the statement that σ is max-optimal for $\hat{\mathscr{A}}$. We have now confirmed all the claims in (iii).

The proof of (iv) is identical after replacing max with min and $\bigvee$ with $\bigwedge$. (Alternatively, the proof can be derived from (v) and duality.) ◻

6.3. **Anti-Isomorphic ADPs.** Let $\mathscr{A} = (V, \{T_\sigma\})$ and $\hat{\mathscr{A}} = (\hat{V}, \{\hat{T}_\sigma\})$ be two ADPs. We call $\mathscr{A}$ and $\hat{\mathscr{A}}$ *anti-isomorphic* under F if they have the same policy set Σ and, in addition, there exists an order anti-isomorphism F from V to $\hat{V}$ such that (16) holds. Equivalently, $\mathscr{A}$ and $\hat{\mathscr{A}}$ are anti-isomorphic if $\mathscr{A}$ is isomorphic to $\hat{\mathscr{A}}^\partial$, the dual of $\hat{\mathscr{A}}$.

If $\mathscr{A} \overset{a}{\sim} \hat{\mathscr{A}}$ indicates that $\mathscr{A}$ and $\hat{\mathscr{A}}$ are anti-isomorphic, then $\overset{a}{\sim}$ is symmetric and transitive but, in general, not reflexive.

Here is an optimality result for anti-isomorphic ADPs that parallels Theorem 6.1.

Theorem 6.2. *If $\mathscr{A}$ and $\hat{\mathscr{A}}$ are anti-isomorphic under F, then*

(i) *$\mathscr{A}$ is well-posed if and only if $\hat{\mathscr{A}}$ is well-posed.*
(ii) *$\mathscr{A}$ is order stable if and only if $\hat{\mathscr{A}}$ is order stable.*
(iii) *$\mathscr{A}$ is max-stable if and only if $\hat{\mathscr{A}}$ is min-stable. In this case,*

$$F \circ T_\top = \hat{T}_\perp \circ F \quad \text{and} \quad \hat{v}_\perp = F v_\top. \tag{19}$$

Moreover, $\sigma \in \Sigma$ is max-optimal for $\mathscr{A}$ if and only if σ is min-optimal for $\hat{\mathscr{A}}$.

Proof. Let $\mathscr{A}$ and $\hat{\mathscr{A}}$ be anti-isomorphic, so that $\mathscr{A}$ is isomorphic to $\hat{\mathscr{A}}^\partial$. If $\mathscr{A}$ is well-posed, then, by Theorem 6.1, $\hat{\mathscr{A}}^\partial$ is well-posed, so $\hat{T}_\sigma$ has a unique fixed point in $\hat{V}$ for all $\sigma \in \Sigma$. This implies that $\hat{\mathscr{A}}$ is likewise well-posed, completing the proof of (i). Similarly, If $\mathscr{A}$ is order stable, then, by Theorem 6.1, $\hat{\mathscr{A}}^\partial$ is order stable, in which case $\hat{\mathscr{A}}$ is order stable, by Lemma 2.2. This proves (ii).

Now suppose $\mathscr{A}$ is max-stable. Then, by $\mathscr{A} \sim \hat{\mathscr{A}}^\partial$ and Theorem 6.1, $\hat{\mathscr{A}}^\partial$ is max-stable with $F \circ T_\top = \hat{T}^\partial_\top \circ F$ and $\hat{v}^\partial_\top = F v_\top$. As with our discussion of duality in Section 5.2, this is equivalent to $F \circ T_\top = \hat{T}_\perp \circ F$ and $\hat{v}_\perp = F v_\top$, which proves (19).

Finally, Theorem 6.1 tells us that $\mathscr{A}$ and $\hat{\mathscr{A}}^\partial$ have the same max-optimal policies. Applying Lemma 5.2, we see that the max-optimal policies of $\mathscr{A}$ are the same as the min-optimal policies of $\hat{\mathscr{A}}$. ◻

7. Subordinate ADPs

Next we introduce an asymmetric relationship called subordination between ADPs. In essence, a subordinate ADP is an ADP that is derived from another ADP, typically via some kind of rearrangement. Often the associated transformations are not bijective, since one dynamic program evolves in a higher dimensional space than another. Nonetheless, subordination provides valuable optimality connections between ADPs.

7.1. Definition and Properties.

Let $\mathscr{A} := (V, \{T_\sigma\})$ and $\hat{\mathscr{A}} := (\hat{V}, \{\hat{T}_\sigma\})$ be ADPs. We say that $\hat{\mathscr{A}}$ is *subordinate* to $\mathscr{A}$ if there exists an order-preserving map F from V onto $\hat{V}$ and a family of order-preserving maps $\{G_\sigma\}_{\sigma \in \Sigma}$ from $\hat{V}$ to V such that

$$T_\sigma = G_\sigma \circ F \quad \text{and} \quad \hat{T}_\sigma = F \circ G_\sigma \qquad \text{for all } \sigma \in \Sigma. \tag{20}$$

In other words, (V, T_σ) and $(\hat{V}, \hat{T}_\sigma)$ are mutually semiconjugate for all $\sigma \in \Sigma$.

Many dynamic programs investigated in the recent literature are subordinate to a more traditional dynamic program.

Example 7.1. Let $\mathscr{A} = (\mathbb{R}^{\mathsf{X}}, \{T_\sigma\})$ be the ADP associated with the MDP from Example 3.1 and let F be defined by

$$(Fv)(x, a) = r(x, a) + \beta \sum_{x'} v(x') P(x, a, x') \qquad (v \in \mathbb{R}^{\mathsf{X}}) \tag{21}$$

If $\hat{V} := F(\mathbb{R}^{\mathsf{X}}) \subset \mathbb{R}^{\mathsf{G}}$ and $\{Q_\sigma\}$ is the policy operators from the Q-factor MDP defined in (6), then $\hat{\mathscr{A}} := (\hat{V}, \{Q_\sigma\})$ is an ADP. Note that F is the order preserving and, by construction, maps $V := \mathbb{R}^{\mathsf{X}}$ onto $\hat{V}$. Also, the maps $\{G_\sigma\}$ defined by

$$(G_\sigma f)(x) = f(x, \sigma(x)) \qquad (f \in \mathbb{R}^{\mathsf{G}})$$

are order-preserving and, for each σ, the operator T_σ satisfies

$$(T_\sigma \, v)(x) = r(x, \sigma(x)) + \beta \sum_{x'} v(x') P(x, \sigma(x), x') = (G_\sigma \, F \, v)(x),$$

for all $v \in \mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$, while the Q-factor policy operator satisfies

$$(Q_\sigma \, f)(x, a) = r(x, a) + \beta \sum_{x'} f(x', \sigma(x')) P(x, a, x') = (F \, G_\sigma \, f)(x, a)$$

for all $f \in \hat{V}$. Hence $T_\sigma = G_\sigma \circ F$ and $Q_\sigma = F \circ G_\sigma$, so $\hat{\mathscr{A}}$ is subordinate to $\mathscr{A}$.

Example 7.2. The risk-sensitive Q-factor ADP in Example 3.5 is subordinate to the standard risk-sensitive ADP in Example 3.4. The proof is almost identical to the argument in Example 7.1, after replacing F in (21) with

$$(Fv)(x,a) = r(x,a) + \frac{\beta}{\theta} \ln \left[\sum_{x'} \exp(\theta v(x')) P(x,a,x') \right].$$

Example 7.3. Let $\mathscr{A} = (\mathbb{R}^{\mathsf{X}}, \{T_\sigma\})$ be an ADP generated by the MDP model from Example 3.1 and let F be the order-preserving map defined by

$$(Fv)(x,a) = \sum_{x'} v(x') P(x,a,x') \qquad (v \in \mathbb{R}^{\mathsf{X}}). \tag{22}$$

If $\hat{V} := F(\mathbb{R}^{\mathsf{X}}) \subset \mathbb{R}^{\mathsf{G}}$ and $\{R_\sigma\}$ is the policy operators from the expected value MDP described in Example 3.6, then $\hat{\mathscr{A}} := (\hat{V}, \{R_\sigma\})$ is subordinate to $\mathscr{A}$. Indeed, F is order-preserving and, by construction, maps $V := \mathbb{R}^{\mathsf{X}}$ onto $\hat{V}$. Moreover, the maps $\{G_\sigma\}$ defined by

$$(G_\sigma g)(x) = r(x, \sigma(x)) + \beta g(x, \sigma(x)) \qquad (g \in \mathbb{R}^{\mathsf{G}}). \tag{23}$$

are order-preserving and the MDP policy operator T_σ satisfies

$$(T_\sigma v)(x) = r(x, \sigma(x)) + \beta \sum_{x'} v(x') P(x, \sigma(x), x') = (G_\sigma F v)(x),$$

while the expected value policy operator R_σ from (9) satisfies

$$(R_\sigma g)(x,a) = \sum_{x'} \{ r(x', \sigma(x')) + \beta g(x', \sigma(x')) \} P(x,a,x') = (F G_\sigma g)(x,a).$$

7.2. **Stability.** Let $\mathscr{A}$ and $\hat{\mathscr{A}}$ represent ADPs with with respective σ-value functions $\{v_\sigma\}$ and $\{\hat{v}_\sigma\}$. Below, if $\hat{\mathscr{A}}$ is subordinate to $\mathscr{A}$, then F and $\{G_\sigma\}$ always represent the order-preserving maps in (20).

Proposition 7.1. *If $\hat{\mathscr{A}}$ is subordinate to $\mathscr{A}$, then*

(i) *$\hat{\mathscr{A}}$ is well-posed if and only if $\mathscr{A}$ is well-posed, and*
(ii) *$\hat{\mathscr{A}}$ is order stable if and only if $\mathscr{A}$ is order stable.*

In either case, the σ-value functions are linked by

$$\hat{v}_\sigma = F v_\sigma \quad \textit{and} \quad v_\sigma = G_\sigma \hat{v}_\sigma \quad \textit{for all} \quad \sigma \in \Sigma.$$

Proof. All claims follow from Lemma 2.4 and the observation that (V, T_σ) and $(\hat{V}, \hat{T}_\sigma)$ are mutually semiconjugate at every $\sigma \in \Sigma$. □

7.3. **Optimality.** In this section we consider whether optimality properties in Theorem 5.1 can be inferred under a subordinate relationship. Below, when $\mathscr{A}$ and $\hat{\mathscr{A}}$ are max-stable, $v_\top$ and $\hat{v}_\top$ represent their max-value functions, while $T_\top$ (resp., $\hat{T}_\top$) denotes the Bellman operator for $\mathscr{A}$ (resp., $\hat{\mathscr{A}}$).

Note that, when $\mathscr{A}$ is max-stable and $T_\sigma = G_\sigma \circ F$, as in (20), then $\{G_\sigma \hat{v}\}_{\sigma \in \Sigma}$ has a greatest element for every $\hat{v} \in \hat{V}$. Indeed, if $\hat{v} \in \hat{V}$, then, since F is onto, there exists a $v \in V$ with $Fv = \hat{v}$. Moreover, by max-stability, there exists a policy $\sigma \in \Sigma$ such that $T_\sigma v = G_\sigma \hat{v}$ dominates $T_\tau v = G_\tau \hat{v}$ for all $\tau \in \Sigma$. This confirms that $\{G_\sigma \hat{v}\}_{\sigma \in \Sigma}$ has a greatest element.

In the setting where $\mathscr{A}$ is max-stable, which we consider below, we take $G_\top$ to be the order-preserving map from $\hat{V}$ to V defined by

$$G_\top \hat{v} := \bigvee_\sigma G_\sigma \hat{v}. \tag{24}$$

Theorem 7.2. *If $\mathscr{A}$ is max-stable and $\hat{\mathscr{A}}$ is subordinate to $\mathscr{A}$, then $\hat{\mathscr{A}}$ is also max-stable and the Bellman max-operators are related by*

$$T_\top = G_\top \circ F \quad \textit{and} \quad \hat{T}_\top = F \circ G_\top, \tag{25}$$

while the max-value functions are related by

$$v_\top = G_\top \hat{v}_\top \quad \textit{and} \quad \hat{v}_\top = F v_\top. \tag{26}$$

Moreover,

(i) *if σ is max-optimal for $\mathscr{A}$, then σ is max-optimal for $\hat{\mathscr{A}}$, and*
(ii) *if $G_\sigma \hat{v}_\top = G_\top \hat{v}_\top$, then σ is max-optimal for $\mathscr{A}$.*

Note the asymmetry in (i) of Theorem 7.2. This implication cannot be reversed without additional conditions, as confirmed by Example A.1 in the appendix. Still, if we want to use $\hat{\mathscr{A}}$ to find optimal policies for $\mathscr{A}$, we can do so via (ii) of Theorem 7.2.

Proof. Fix $\hat{v} \in \hat{V}$ and let $\sigma \in \Sigma$ be such that $G_\sigma \hat{v} \geq G_\tau \hat{v}$ for all $\tau \in \Sigma$. Applying F gives $\hat{T}_\sigma \hat{v} \geq \hat{T}_\tau \hat{v}$ for all $\tau \in \Sigma$, so a $\hat{v}$-max-greedy policy exists. Also, given $v \in V$, we have $T_\top v = \bigvee_\sigma G_\sigma F v = G_\top F v$. This proves the first claim in (25). The second claim in (25) follows from

$$\hat{T}_\top \hat{v} = \bigvee_\sigma F G_\sigma \hat{v} = F \bigvee_\sigma G_\sigma \hat{v} = F G_\top \hat{v},$$

where the second equality holds because $\{G_\sigma \hat{v}\}_{\sigma \in \Sigma}$ has a greatest element. This proves (25), so $(V, T_\top)$ and $(\hat{V}, \hat{T}_\top)$ are mutually semiconjugate under $F, G_\top$.

In view of Lemma 2.4, combined with the fact that $v_\top$ is a fixed point of $T_\top$ (by Theorem 5.1), we see that $\hat{T}_\top$ has a fixed point in $\hat{V}$. This proves that $\hat{\mathscr{A}}$ is max-stable. Therefore, applying Theorem 5.1 to $\hat{\mathscr{A}}$, the supremum $\hat{v}_\top := \bigvee_\sigma \hat{v}_\sigma$ exists and is the unique fixed point of $\hat{T}_\top$ in $\hat{V}$. The equalities in (26) now follow from mutual semiconjugacy of $(V, T_\top)$ and $(\hat{V}, \hat{T}_\top)$ under $F, G_\top$.

Regarding (i) of the last part of Theorem 7.2, let σ be max-optimal for $\mathscr{A}$. Since $\mathscr{A}$ is max-stable, Theorem 5.1 implies that σ is $v_\top$-max-greedy for $\mathscr{A}$ ($T_\sigma v_\top = T_\top v_\top$) and $v_\top = T_\top v_\top$. Also, by (26), we have $\hat{v}_\top = F v_\top$. Therefore,

$$\hat{T}_\sigma \hat{v}_\top = \hat{T}_\sigma F v_\top = F\, T_\sigma v_\top = F v_\top = \hat{v}_\top = \hat{T}_\top \hat{v}_\top.$$

(The last equality uses the fact that stability of $\mathscr{A}$ implies stability of $\hat{\mathscr{A}}$ combined with Theorem 5.1.) In other words, σ is $\hat{v}_\top$-max-greedy for $\hat{\mathscr{A}}$. But $\hat{\mathscr{A}}$ is max-stable, so another application of Theorem 5.1 confirms that σ is max-optimal for $\hat{\mathscr{A}}$.

Regarding (ii) of the last part of Theorem 7.2, let $\sigma \in \Sigma$ be such that $G_\sigma \hat{v}_\top = G_\top \hat{v}_\top$. Applying (26) yields $G_\sigma F v_\top = G_\top F v_\top$, or $T_\sigma v_\top = T_\top v_\top$. Thus, σ is $v_\top$-max-greedy for $\mathscr{A}$. Since $\mathscr{A}$ is max-stable, Theorem 5.1 implies that σ is max-optimal for $\mathscr{A}$. □

Example 7.4. To find optimal policies for the standard MDP model in Example 3.1, with associated ADP $\mathscr{A} := (\mathbb{R}^{\mathsf{X}}, \{T_\sigma\})$, we can study instead the expected value variation in Example 3.6, with subordinate ADP $\hat{\mathscr{A}} := (\hat{V}, \{R_\sigma\})$ for $\hat{V} = F\mathbb{R}^{\mathsf{G}}$ with F defined in (22) and R_σ defined in (9). (Subordination is established in Example 7.3.) In view of Theorem 7.2, we can do this by computing the fixed point $\bar{g}$ of the corresponding Bellman operator $R := \bigvee_\sigma R_\sigma$ and then finding a policy σ obeying $G_\sigma \bar{g} = G_\top \bar{g}$. By the definition of G_σ in (23), this means that we solve for σ satisfying

$$\sigma(x) \in \operatorname*{arg\,max}_{a \in \Gamma(x)} \{r(x,a) + \beta \bar{g}(x,a)\} \qquad (x \in \mathsf{X}).$$

7.4. **Minimization and Subordination.** The minimization case is identical to the maximization setting after the obvious modifications. In the setting where $\mathscr{A}$ is min-stable, we take $G_\perp$ to the order-preserving map from $\hat{V}$ to V defined by

$$G_\perp \hat{v} := \bigwedge_\sigma G_\sigma \hat{v}. \tag{27}$$

We can then state the following minimization version of Theorem 7.2.

Theorem 7.3. *If $\mathscr{A}$ is min-stable and $\hat{\mathscr{A}}$ is subordinate to $\mathscr{A}$, then $\hat{\mathscr{A}}$ is also min-stable and the Bellman min-operators are related by*

$$T_\perp = G_\perp \circ F \quad \textit{and} \quad \hat{T}_\perp = F \circ G_\perp, \tag{28}$$

while the min-value functions are related by

$$v_\perp = G_\perp \hat{v}_\perp \quad \text{and} \quad \hat{v}_\perp = F v_\perp. \tag{29}$$

Moreover,

(i) *if σ is min-optimal for $\mathscr{A}$, then σ is min-optimal for $\hat{\mathscr{A}}$, and*
(ii) *if $G_\sigma \hat{v}_\perp = G_\perp \hat{v}_\perp$, then σ is min-optimal for $\mathscr{A}$.*

The proof of Theorem 7.3 is the same as that of Theorem 7.2, after replacing max with min and $\vee$ with $\wedge$. It can also be obtained by applying Theorem 7.2 to the duals of $\mathscr{A}$ and $\hat{\mathscr{A}}$.

8. Applications

In this section we show how isomorphic and subordinate relationships can simplify or illuminate dynamic programming problems. We begin in Section 8.1 with a specific dynamic program and apply a series of transformations. We connection these transformations to isomorphic and subordinate relationships and deduce some new results on optimality.

8.1. **Modified Epstein–Zin Equations.**

Consider an Epstein–Zin version of the MDP in Example 3.1 (see, e.g., Epstein and Zin (1989) or Weil (1990)), in which a Bellman max-equation takes the form

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a)^\alpha + \beta(x) \left(\sum_{x'} v(x')^\gamma P(x,a,x') \right)^{\alpha/\gamma} \right\}^{1/\alpha}.$$

Following much of the recent literature, we allow the rate of time preference β to depend on the state (see, e.g., Albuquerque et al. (2016), Schorfheide et al. (2018), de Groot et al. (2018), Gomez-Cram and Yaron (2020)). The function β maps X to $\mathbb{R}_+$ and γ and α are nonzero parameters. Other details are as in Example 3.1. Using the symbols r_σ and P_σ from (10), the policy operator T_σ can be written as

$$T_\sigma v = \left\{ r_\sigma^\alpha + \beta \, (P_\sigma v^\gamma)^{\alpha/\gamma} \right\}^{1/\alpha}, \tag{30}$$

where powers are taken pointwise. Since γ and α can be negative, we assume r is positive. We also suppose that P_σ is irreducible for all $\sigma \in \Sigma$. Under these assumptions, T_σ is an order-preserving self-map on $(V, \leqslant)$, the set of all strictly positive functions on X paired with the pointwise partial order, and $\mathscr{A} := (V, \{T_\sigma\})$ is an ADP.

Now set

$$\theta := \frac{\gamma}{\alpha} \quad \text{and} \quad \hat{T}_\sigma\, v := \left\{ r_\sigma^\alpha + \beta\, (P_\sigma v)^{1/\theta} \right\}^\theta. \tag{31}$$

The pair $\hat{\mathscr{A}} = (V, \{\hat{T}_\sigma\})$ is also an ADP.

Lemma 8.1. *The following relationships hold:*

(i) *If $\gamma > 0$, then $\mathscr{A}$ and $\hat{\mathscr{A}}$ are isomorophic.*
(ii) *If $\gamma < 0$, then $\mathscr{A}$ and $\hat{\mathscr{A}}$ are anti-isomorophic.*

Proof. Let $F \colon V \to V$ be the bijective map $Fv = v^\gamma$. Fixing σ and applying (30) yields

$$F\, T_\sigma\, v = (T_\sigma\, v)^\gamma = \left\{ r_\sigma^\alpha + \beta\, (P_\sigma\, v^\gamma)^{\alpha/\gamma} \right\}^{\gamma/\alpha} = \left\{ r_\sigma^\alpha + \beta\, (P_\sigma\, v^\gamma)^{1/\theta} \right\}^\theta.$$

Inspection of (31) shows that $\hat{T}_\sigma\, Fv$ is identical to the last expression. Hence $F \circ T_\sigma = \hat{T}_\sigma \circ F$ on V. If $\gamma > 0$, then F is order-preserving, so $\mathscr{A}$ and $\hat{\mathscr{A}}$ are isomorophic. If $\gamma < 0$, then F is order-reversing, so $\mathscr{A}$ and $\hat{\mathscr{A}}$ are anti-isomorophic. □

Lemma 8.1 gives us a way to solve for max-optimal policies of $\mathscr{A}$ by studying $\hat{\mathscr{A}}$ (and applying either Theorem 6.1 or Theorem 6.2). This is convenient because $\hat{\mathscr{A}}$ is easier to analyze. The next section illustrates.

8.2. **Characterizing Optimality.** Suppose that β depends on x through a purely exogenous state component (as in, say, de Groot et al. (2018) and Schorfheide et al. (2018)). Specifically, $\mathsf{X} = \mathsf{Y} \times \mathsf{Z}$ and $x = (y, z)$, where

$$P(x, a, x') = R(y, a, y')Q(z, z') \quad \text{and} \quad \beta(x) = \beta(z).$$

Here $R(y, a, \cdot)$ is a distribution over y for each feasible (y, a) pair and Q is a stochastic matrix over Z. For each $z \in \mathsf{Z}$, let $(Z_t(z))_{t \geqslant 0}$ be a Markov chain on Z generated by Q and starting at z. Define

$$\mathcal{E}(\beta, Q, \theta) := \lim_{k \to \infty} \left\{ \sup_{z \in \mathsf{Z}} \mathbb{E} \prod_{t=0}^{k-1} \beta(Z_t(z))^\theta \right\}^{1/k}. \tag{32}$$

We can now state the following exact result.

Theorem 8.2. *The Epstein–Zin ADP $\mathscr{A}$ is max-stable if and only if*

$$\mathcal{E}(\beta, Q, \theta)^{1/\theta} < 1 \tag{33}$$

Hence, under (33), all of the optimality results in Theorem 5.1 apply. Conversely, if (33) fails, then $\mathscr{A}$ is not well-posed and optimality is undefined.

The sufficiency of (33) for optimality properties is related to an earlier result in Stachurski and Zhang (2021). The converse result is new. The proof of Theorem 8.2 proceeds by studying the simpler ADP $\hat{\mathscr{A}}$ introduced in Section 8.1, which we show is either isomorphic or anti-isomorphic, depending on the sign of γ. These relationships imply that one can solve $\mathscr{A}$ by, say, applying Howard policy iteration to $\hat{\mathscr{A}}$ rather than $\mathscr{A}$, with the better choice depending on relative numerical stability.

The proof of Theorem 8.2 can be found in Section A.2 of the appendix.

8.3. **Efficiency from Subordination.** Consider a specific Epstein–Zin dynamic program with Bellman equation

$$v(w, e) = \max_{0 \leqslant s \leqslant w} \left\{ r(w, s, e)^\alpha + \beta \left(\sum_{e'} v(s, e')^\gamma \varphi(e') \right)^{\alpha/\gamma} \right\}^{1/\alpha}. \tag{34}$$

Here w is current wealth, s is savings and e is an endowment shock that we take to be IID with common distribution φ and range E. We take β to be constant in $(0, 1)$ and restrict w, s to a finite set W. To ensure that (34) is well-defined at all parameter values, we assume $r > 0$. The policy operator corresponding to $\sigma \in \Sigma$ is

$$(T_\sigma v)(w, e) = \left\{ r(w, \sigma(w), e)^\alpha + \beta \left(\sum_{e'} v(\sigma(w), e')^\gamma \varphi(e') \right)^{\alpha/\gamma} \right\}^{1/\alpha}. \tag{35}$$

If $\mathsf{X} := \mathsf{W} \times \mathsf{E}$ and $V := (0, \infty)^{\mathsf{X}}$, then $\mathscr{A} = (V, \{T_\sigma\})$ is a special case of the ADP introduced in Section 8.1. Since β is constant, we have $\mathcal{E}(\beta, Q, \theta)^{1/\theta} = \beta < 1$. Hence $\mathscr{A}$ is max-stable (Theorem 8.2).

Next consider the operator

$$(B_\sigma h)(w) = \left\{ \sum_e \{ r(w, \sigma(w), e)^\alpha + \beta h(\sigma(w))^\alpha \}^{\gamma/\alpha} \varphi(e) \right\}^{1/\gamma}, \tag{36}$$

where h is an element of $(0, \infty)^{\mathsf{W}}$. If F is defined at $v \in V$ by

$$(Fv)(w) = \left\{ \sum_e v(w, e)^\gamma \varphi(e) \right\}^{1/\gamma} \qquad (w \in \mathsf{W})$$

and $H := F(V) \subset (0, \infty)^{\mathsf{W}}$, then $\mathscr{B} = (H, \{B_\sigma\})$ is an ADP. Moreover, $\mathscr{B}$ is subordinate to $\mathscr{A}$, since, with G_σ defined at $h \in H$ by

$$(G_\sigma h)(w, e) = \{ r(w, \sigma(w), e)^\alpha + \beta h(\sigma(w))^\alpha \}^{1/\alpha} \qquad ((w, e) \in \mathsf{X}),$$

we can see that F and G_σ are both order-preserving, that T_σ in (35) is equal to $G_\sigma \circ F$, and that B_σ in (36) is equal to $F \circ G_\sigma$.

The benefit of working with $\mathscr{B}$ is that B_σ acts on functions that depending only on w, rather than both w and e (as is the case for T_σ). These lower dimensional operations are significantly more efficient, even when the range E of e is relatively small.

Since $\mathscr{B}$ is subordinate to $\mathscr{A}$, Theorem 7.2 implies that $\mathscr{B}$ is max-stable and we can obtain a max-optimal policy for $\mathscr{A}$ by finding the max-value function $h_\top$ for $\mathscr{B}$ and then calculating a policy σ obeying $G_\sigma h_\top = G_\top h_\top$ (see (ii) in Theorem 7.2). By the definition of G_σ in (23), this means that we solve for σ satisfying

$$\sigma(w,e) \in \operatorname*{arg\,max}_{0 \leqslant s \leqslant w} \{r(w,s,e)^\alpha + \beta h(s)^\alpha\}^{1/\alpha} \tag{37}$$

with $h = h_\top$ at each $(w,e) \in \mathsf{X}$. To compute $h_\top$, we can use Theorem 5.1, which tells us that Howard max-policy iteration converges to $h_\top$ in finitely many steps. Summarizing this analysis, an optimal policy for $\mathscr{A}$ can be computed via Algorithm 1.

Algorithm 1: Solving $\mathscr{A}$ via $\mathscr{B}$

input $\sigma_0 \in \Sigma$, set $k \leftarrow 0$ and $\varepsilon \leftarrow 1$

while $\varepsilon > 0$ **do**

  $h_k \leftarrow$ the fixed point of B_{σ_k}

  $\sigma_{k+1} \leftarrow$ an h_k-max-greedy policy, satisfying

$$\sigma_{k+1}(w) \in \operatorname*{arg\,max}_{0 \leqslant s \leqslant w} \left\{ \sum_e \{r(w,s,e)^\alpha + \beta h(s)^\alpha\}^{\gamma/\alpha} \varphi(e) \right\}^{1/\gamma}$$

  $\varepsilon \leftarrow \mathbb{1}\{\sigma_k \neq \sigma_{k+1}\}$

  $k \leftarrow k+1$

end

return the σ in (37) with $h = h_k$

Figure 1 shows $w \mapsto \sigma^*(w,e)$ for two values of e (smallest and largest) when σ^* is the optimal policy, calculated using Algorithm 1. In the figure we set $r(w,s,e) = (w-s+e)$ and choose α and γ to match values used in Schorfheide et al. (2018). The full parameterization is at `https://github.com/jstac/adps_public`.

In Figure 2 we display the relative speed gain from using the lower-dimensional model $\mathscr{B}$ instead of $\mathscr{A}$ across multiple choices of $|\mathsf{W}|$ and $|\mathsf{E}|$. The speed gain is

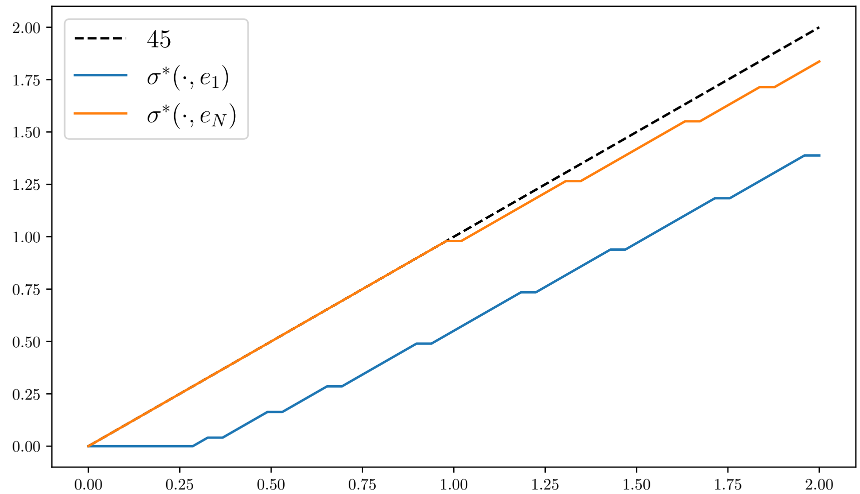

FIGURE 1. Optimal savings policy with Epstein–Zin preference

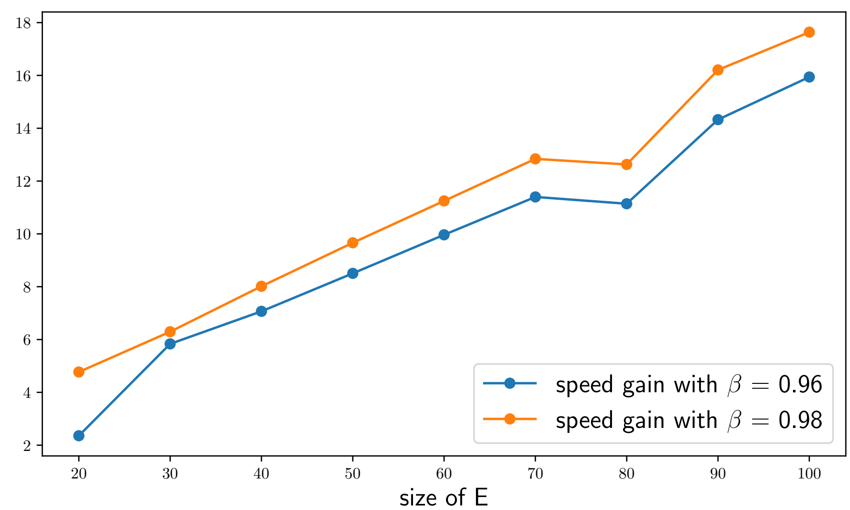

FIGURE 2. Speed gain from replacing $\mathscr{A}$ with subordinate model $\mathscr{B}$

the time required to solve an optimal policy for $\mathscr{A}$ using HPI applied to $\mathscr{A}$ (as in Theorem 5.1), divided by the time required to solve for the same optimal policy via Algorithm 1. The speed gain increases linearly in the size of E. Code (in Julia) and full details can be found at `https://github.com/jstac/adps_public`.

9. Conclusion

We have studied dynamic programming in a setting that abstracts further than Bertsekas (2022). We formulated a dynamic program as a family of self-maps on a partially ordered space. The framework is sufficient to define and study optimality and includes all dynamic programming applications with which we are familiar, ranging from discrete to continuous time and from finite to infinite horizons.

After constructing and defining abstract dynamic programs, we studied relationships between dynamic programs, including isomorphic dynamic programs and subordinate dynamic programs. We showed how optimality transmits across dynamic programs that are paired by these relationships, and how these ideas simplify analysis and open efficient new angles of attack.

Many interesting questions remain unanswered. For example, it would be helpful to uncover more conditions under which max- and min-stability hold across particular classes of applications. Other useful relationships may wait to be uncovered, some of which can, we hope, be seen more clearly from our abstract perspective.

Appendix A. Remaining Proofs

A.1. **Proof of Optimality Results.** Let $\mathscr{A} = (V, \{T_\sigma\})$ be an ADP. When max-greedy policies exist, we let $T_\top$ be the Bellman max-operator and $H_\top$ be the Howard max-operator. As above, we denote the greatest element of V_Σ by $v_\top$ whenever it exists.

Lemma A.1. *If every $v \in V$ has at least one max-greedy policy, then the following statements are true:*

(i) *$H_\top$ obeys $v_\sigma \preceq H_\top v_\sigma$ for all $\sigma \in \Sigma$.*
(ii) *If $\sigma \in \Sigma$ and $T v_\sigma = v_\sigma$, then $v_\top$ exists in V and $v_\sigma = v_\top$.*
(iii) *If $v \in V$ and $H_\top v = v$, then $v = v_\top$ and $T_\top v_\top = v_\top$.*
(iv) *If $v \in V$ and Σ is finite, then $v_\top$ exists, $H_\top v_\top = v_\top$ and and $(H_\top^k v)_{k \geqslant 0}$ converges to $v_\top$ in finitely many steps.*

Proof. As for (i), fix $\sigma \in \Sigma$ and let τ be such that $H_\top v_\sigma = v_\tau$. Since τ is v_σ-greedy, we have $v_\sigma = T_\sigma v_\sigma \preceq T_\top v_\sigma = T_\tau v_\sigma$. Upward stability of T_τ gives $v_\sigma \preceq v_\tau = H_\top v_\sigma$.

As for (ii), suppose $\sigma \in \Sigma$ and $T_\top v_\sigma = v_\sigma$. Fix $\tau \in \Sigma$ and note that $v_\sigma = T_\top v_\sigma \succeq T_\tau v_\sigma$. Downward stability of T_τ implies $v_\sigma \succeq v_\tau$. Since $\tau \in \Sigma$ was arbitrary, $v_\sigma = v_\top$.

As for (iii), fix $v \in V$ with $H_\top v = v$ and let σ be such that $H_\top v = v_\sigma$. Then $v_\sigma = v$, and, since σ is v-max-greedy, $T_\sigma v = T_\top v$. But then $T_\sigma v_\sigma = T_\top v_\sigma$, and, since $v_\sigma = T_\sigma v_\sigma$, we have $v_\sigma = T_\top v_\sigma$. Part (ii) now implies $v = v_\sigma = v_\top$. This proves the first claim. Regarding the second, substituting $v_\sigma = v_\top$ into $v_\sigma = T_\top v_\sigma$ yields $v_\top = T_\top v_\top$.

For (iv), it suffices to show that $H_\top v_\top = v_\top$ and there exists a $K \in \mathbb{N}$ such that $H_\top^K v = v_\top$. To this end, let $v_k = H_\top^k v$ and note that $v_k \in V_\Sigma$ for all $k \geqslant 1$. Part (i) implies that $v_{k+1} \succeq v_k$ for all $k \in \mathbb{N}$. Since the sequence (v_k) is contained in the finite set V_Σ, it must be that $v_{K+1} = v_K$ for some $K \in \mathbb{N}$ (since otherwise V_Σ contains an infinite sequence of distinct points). But then $H_\top v_K = v_{K+1} = v_K$, so v_K is a fixed point of $H_\top$. Part (iii) now implies that $v_K = v_\top$. □

Proof of Proposition 4.1. If $\mathscr{A}$ is an ADP such that max-greedy policies exist and Σ is finite, then, by (iii)–(iv) of Lemma A.1, the point $v_\top$ is a fixed point of $T_\top$. This proves the max version of Proposition 4.1. The proof of the min version is analogous. □

Lemma A.2. *If $\mathscr{A}$ is max-stable, then the following statements hold.*

(i) *V_Σ has a greatest element $v_\top$ and*
(ii) *$v_\top$ is the unique fixed point of $T_\top$ in V.*
(iii) *a policy is max-optimal if and only if it is $v_\top$-max-greedy.*
(iv) *at least one optimal policy exists.*

Proof. As for parts (i)–(ii), we observe that, by max-stability, $T_\top$ has a fixed point $\bar{v}$ in V. By existence of max-greedy policies, we can find a $\sigma \in \Sigma$ such that $\bar{v} = T_\top \bar{v} = T_\sigma \bar{v}$. But T_σ has a unique fixed point in V, equal to v_σ, so $\bar{v} = v_\sigma$. Moreover, if τ is any policy, then $T_\tau \bar{v} \preceq T_\top \bar{v} = \bar{v}$ and hence, by downward stability, $v_\tau \preceq \bar{v}$. These facts imply that $v_\top := \bar{v}$ is the greatest element of V_Σ and a fixed point of $T_\top$. Since greatest elements are unique, $v_\top$ is the only fixed point of $T_\top$ in V.

For (iii), parts (i)–(ii) give $v_\top \in V$ and $T_\top v_\top = v_\top$. Now recall that σ is optimal if and only if $v_\sigma = v_\top$. Since v_σ is the unique fixed point of T_σ, this is equivalent to $T_\sigma v_\top = v_\top$. Since $T_\top v_\top = v_\top$, the last statement is equivalent to $T_\sigma v_\top = T_\top v_\top$, which is, in turn equivalent to the statement that σ is $v_\top$-greedy. This proves the first claim in the proposition.

Part (iv) follows from part (iii) and existence of a $v_\top$-greedy policy. □

Proof of Theorem 5.1. Parts (i)–(iv) of Theorem 5.1 follow from Lemma A.2. The last claim follows from Lemma A.1. □

Example A.1. Let $V = \{a, b\}$ and $\hat{V} = \{\hat{a}, \hat{b}\}$ where $a \leq b$ and $\hat{a} \leq \hat{b}$. Let

$$F\,a = F\,b = \hat{a}, \quad G_1\,\hat{a} = G_1\,\hat{b} = b \quad \text{and} \quad G_2\,\hat{a} = G_2\,\hat{b} = a.$$

Being constant functions, F, G_1 and G_2 are order-preserving. Setting $T_i := G_i \circ F$ and $\hat{T}_i = F \circ G_i$ forms ADPs $\mathscr{A} = (V, \{T_i\})$ and $\hat{\mathscr{A}} = (\hat{V}, \{\hat{T}_i\})$. By construction, $\hat{\mathscr{A}}$ is subordinate to $\mathscr{A}$. For $\mathscr{A}$ we have $T_1 a = T_1 b = b$, so T_1 is order stable and $v_1 = b$. Similarly, T_2 is order stable and $v_2 = a$. Hence $\mathscr{A}$ is order stable and $v_\top = v_1 = b$. Letting Σ^* be the set of optimal policies, we have $\Sigma^* = \{1\}$. For $\hat{\mathscr{A}}$ we have $\hat{T}_i\,\hat{a} = \hat{T}_i\,\hat{b} = \hat{a}$ for $i = 1, 2$, so $\hat{v}_1 = \hat{v}_2 = \hat{a}$. Hence $\hat{v}_\top = \hat{a}$ and the set of optimal policies is $\hat{\Sigma}^* = \{1, 2\}$. This shows that the reverse implication in the final part of Theorem 7.2 is invalid: the fact that σ is max-optimal for $\hat{\mathscr{A}}$ does not imply that if σ is max-optimal for $\mathscr{A}$.

A.2. **Proofs of Epstein–Zin Optimality Results.** This section offers a proof of Theorem 8.2. To do so, we establish the following.

(C1) If $\mathcal{E}(\beta, Q, \theta)^{1/\theta} < 1$ and $\gamma < 0$, then $\hat{\mathscr{A}}$ is min-stable.
(C2) If $\mathcal{E}(\beta, Q, \theta)^{1/\theta} < 1$ and $\gamma > 0$, then $\hat{\mathscr{A}}$ is max-stable.
(C3) If $\mathcal{E}(\beta, Q, \theta)^{1/\theta} \geqslant 1$, then $\hat{\mathscr{A}}$ is not well-posed.

Together these facts establish Theorem 8.2. Indeed, if (C1) holds, then, since $\mathscr{A}$ and $\hat{\mathscr{A}}$ are anti-isomorphic (see Lemma 8.1), it follows that $\mathscr{A}$ is max-stable (Theorem 6.2). If (C2) holds, then, since $\mathscr{A}$ and $\hat{\mathscr{A}}$ are isomorphic (see Lemma 8.1), it follows that $\mathscr{A}$ is max-stable (Theorem 6.1). Finally, if (C3) holds, then $\mathscr{A}$ is also not well-posed (by Theorem 6.1 or Theorem 6.2, depending on whether $\gamma > 0$ or $\gamma < 0$.) When $\mathscr{A}$ is not well-posed, recursive utility does not exist, so the dynamic program is undefined.

In what follows, given $\sigma \in \Sigma$ and defining P_σ as in (10), we set

$$A_\sigma(x, x') := \beta(x)^\theta P_\sigma(x, x') \qquad (x, x' \in \mathsf{X}).$$

Also, for any linear operator B, the symbol $\rho(B)$ represents the spectral radius.

Lemma A.3. *For all $\sigma \in \Sigma$, we have $\rho(A_\sigma) = \mathcal{E}(\beta, Q, \theta)$.*

Proof. Fix $z \in \mathsf{Z}$ and let $\mathbb{1}$ be a vector of ones. An inductive argument shows that

$$(A_\sigma^k \mathbb{1})(x) = (A_\sigma^k \mathbb{1})(z) = \mathbb{E} \prod_{t=0}^{k-1} \beta(Z_t(z))^\theta. \tag{38}$$

Combining (38) with Theorem 9.1 of Krasnosel'skii et al. (1972), we have

$$\rho(A_\sigma) = \lim_{k\to\infty} \left\{ \sup_z (A_\sigma^k \mathbb{1})(z) \right\}^{1/k} = \lim_{k\to\infty} \left\{ \sup_{z\in Z} \mathbb{E} \prod_{t=0}^{k-1} \beta(Z_t(z))^\theta \right\}^{1/k},$$

as was to be shown. □

Lemma A.4. *The ADP $\hat{\mathscr{A}}$ is order stable if and only if $\mathcal{E}(\beta, Q, \theta)^{1/\theta} < 1$. Moreover, if this condition fails, then $\hat{\mathscr{A}}$ is not well posed.*

Proof. Fix $\sigma \in \Sigma$ and let V and $\hat{T}_\sigma$ be as defined in Section 8.1. By Theorem 3.1 of Stachurski et al. (2022),

(i) $\rho(A_\sigma)^{1/\theta} < 1 \implies (V, \hat{T}_\sigma)$ is asymptotically stable on V, and
(ii) $\rho(A_\sigma)^{1/\theta} \geqslant 1 \implies \hat{T}_\sigma$ has no fixed point in V.

We saw in Lemma A.4 that $\rho(A_\sigma) = \mathcal{E}(\beta, Q, \theta)$, so $\mathcal{E}(\beta, Q, \theta)^{1/\theta} < 1$ if and only if (i) holds. In this case, $(V, \hat{T}_\sigma)$ is asymptotically stable and hence order stable (by Example 2.1). Therefore $\hat{\mathscr{A}}$ is order stable.

If, on the other hand, $\mathcal{E}(\beta, Q, \theta)^{1/\theta} \geqslant 1$, then (ii) holds and $\hat{\mathscr{A}}$ is not well-posed (and therefore not order stable). □

Now we return to (C1)–(C3) above. Assume the conditions in (C1). Then $\hat{\mathscr{A}}$ is order stable by Lemma A.4. Also, $v \in V$, we construct a v-min-greedy policy σ by taking

$$\sigma(x) \in \arg\min \left\{ r(x,a)^\alpha + \beta(x) \left[\sum_{x'\in \mathsf{X}} v(x') P(x, a, x') \right]^{1/\theta} \right\}^\theta$$

for all $x \in \mathsf{X}$. Since the policy set is finite, Proposition 4.1 implies that $\hat{\mathscr{A}}$ is min-stable. Hence (C1) holds. The proof of (C2) is analogous. Finally, (C3) follows directly from Lemma A.4.

References

Albuquerque, R., Eichenbaum, M., Luo, V. X., and Rebelo, S. (2016). Valuation risk and asset pricing. *The Journal of Finance*, 71(6):2861–2904.

Bäuerle, N. and Jaśkiewicz, A. (2018). Stochastic optimal growth model with risk sensitive preferences. *Journal of Economic Theory*, 173:181–200.

Bertsekas, D. (2021). *Rollout, policy iteration, and distributed reinforcement learning.* Athena Scientific.

Bertsekas, D. P. (2022). *Abstract dynamic programming*. Athena Scientific, 3 edition.

Bloise, G., Le Van, C., and Vailakis, Y. (2023). Do not blame bellman: It is koopmans' fault. *Econometrica*, in press.

Davey, B. A. and Priestley, H. A. (2002). *Introduction to lattices and order*. Cambridge University Press.

de Groot, O., Richter, A. W., and Throckmorton, N. A. (2018). Uncertainty shocks in a model of effective demand: Comment. *Econometrica*, 86(4):1513–1526.

Denardo, E. V. (1967). Contraction mappings in the theory underlying dynamic programming. *Siam Review*, 9(2):165–177.

Epstein, L. G. and Zin, S. E. (1989). Risk aversion and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969.

Fei, Y., Yang, Z., Chen, Y., and Wang, Z. (2021). Exponential bellman equation and improved regret bounds for risk-sensitive reinforcement learning. *Advances in Neural Information Processing Systems*, 34:20436–20446.

Gomez-Cram, R. and Yaron, A. (2020). How Important Are Inflation Expectations for the Nominal Yield Curve? *The Review of Financial Studies*, 34(2):985–1045.

Guo, X. and Hernández-Lerma, O. (2009). *Continuous-time Markov decision processes*. Springer.

Howard, R. A. and Matheson, J. E. (1972). Risk-sensitive Markov decision processes. *Management science*, 18(7):356–369.

Kochenderfer, M. J., Wheeler, T. A., and Wray, K. H. (2022). *Algorithms for decision making*. The MIT Press.

Krasnosel'skii, M. A., Vainikko, G. M., Zabreiko, P. P., Rutitskii, Y. B., and Stetsenko, V. Y. (1972). *Approximate Solution of Operator Equations*. Springer Netherlands.

Layek, G. (2015). *An Introduction to Dynamical Systems and Chaos*. Springer.

Ma, Q. and Stachurski, J. (2021). Dynamic programming deconstructed: Transformations of the bellman equation and computational efficiency. *Operations Research*, 69(5):1591–1607.

Mogensen, P. K. (2018). Solving dynamic discrete choice models: Integrated or expected value function? *arXiv preprint arXiv:1801.03978*.

Powell, W. B., Simao, H. P., and Bouzaiene-Ayari, B. (2012). Approximate dynamic programming in transportation and logistics: a unified framework. *EURO Journal on Transportation and Logistics*, 1(3):237–284.

Puterman, M. L. (2005). *Markov decision processes: discrete stochastic dynamic programming*. Wiley Interscience.

Ren, G. and Stachurski, J. (2021). Dynamic programming with value convexity. *Automatica*, 130:109641.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, pages 999–1033.

Rust, J. (1994). Structural estimation of markov decision processes. *Handbook of econometrics*, 4:3081–3143.

Schorfheide, F., Song, D., and Yaron, A. (2018). Identifying long-run risks: A bayesian mixed-frequency approach. *Econometrica*, 86(2):617–654.

Stachurski, J., Wilms, O., and Zhang, J. (2022). Unique solutions to power-transformed affine systems. *arXiv preprint arXiv:2212.00275*.

Stachurski, J. and Zhang, J. (2021). Dynamic programming with state-dependent discounting. *Journal of Economic Theory*, 192:105190.

Sternberg, S. (2014). *Dynamical Systems*. Dover Publications.

Weil, P. (1990). Nonexpected utility in macroeconomics. *The Quarterly Journal of Economics*, 105(1):29–42.

Department of Economics, New York University

Email address: `ts43@nyu.edu`

Research School of Economics, Australian National University

Email address: `john.stachurski@anu.edu.au`