# A Comment on "Using Randomization to Break the Curse of Dimensionality"

Robert L. Bray

Kellogg School of Management, Northwestern University

December 22, 2021

**Abstract**

Rust (1997b) discovered a class of dynamic programs that can be solved in polynomial time with a randomized algorithm. For these dynamic programs, the optimal values of a polynomially large sample of states are sufficient statistics for the (near) optimal values everywhere, and the values of this random sample can be bootstrapped from the sample itself. However, I show that this class is limited, as it requires all but a vanishingly small fraction of state variables to behave arbitrarily similarly to *i.i.d.* uniform random variables.

*Keywords*: Markov decision process; dynamic program; curse of dimensionality; random value iteration; empirical process.

## 1 Introduction

Rust (1997b) developed an elegant way to solve high-dimensional dynamic programs: (i) restrict the problem to a random sample of states, (ii) solve this restricted problem with value iteration, and (iii) use the values of this restricted problem to approximate the values of the original problem. Rust explained that this approach can sometimes "break the curse of dimensionality"—precipitating a phase change from exponential-time difficulty to polynomial-time difficulty—because weakening the objective from solving the value function within $\epsilon$ to solving the value function within $\epsilon$ with *high probability* can make some otherwise intractable problems tractable.

The Achilles heel of Rust's approach is the "needle in the haystack problem" that Rust (1997a, p. 33) identified in a companion paper. This problem arises when the state transition function has a spike that is too sharp for the random sampler to detect. Rust (1997a, pp. 35, 47) explained that

this "problem can lead $V_N(s,a)$ to be a poor estimate of the true expected value function" and that it "can be a much more serious problem in higher dimensional problems." Kristensen et al. (2021, p. 329) empirically confirmed this claim, reporting that the "variances of [Rust's] self-approximating method increase dramatically as the number of state variables increases."

I show that Rust's approach can avoid the "needle in the haystack problem"—and hence overcome the curse of dimensionality—in only one special case: when the problem can be recast so that all but a vanishingly small fraction of state variables behave like history-independent uniform random variables. Specifically, only around $\log(d)$ out of $d$ state variables may meaningfully depend on past states and actions; the other state variables must resemble exogenous *i.i.d.* shocks.

## 2 Setup

This section formally defines the dynamic programs under investigation. However, before characterizing these optimization problems we must first establish a few mathematical conventions:

- $\|\cdot\|$, $\|\cdot\|_1$, and $\|\cdot\|_2$ are the $L^\infty$, $L^1$, and $L^2$ norms, respectively.
- A sequence $\{x_d\}_{d\in\mathbb{N}}$ is *polynomially bounded* if and only if there exist $m,n\in\mathbb{N}$ such that $|x_d| < m + d^n$ for all $d\in\mathbb{N}$.
- $\mathbb{b}$ is the set of all polynomially bounded sequences.
- $\mathbb{B}_d$ is the set of bounded functions from $[0,1]^d$ to $\mathbb{R}$.
- $\lambda_d$ represents the Lebesgue measure over the Borel subsets of $\mathbb{R}^d$. For example, $\int_{t\in[0,1]^d} f(t)\lambda_d(dt)$ is the Lebesgue integral of $f$ over $[0,1]^d$.
- If $p_d$ is a probability density function over $[0,1]^d$ then $p_d^{<i}(t_{<i}) \equiv \int_{t_{\geq i}\in[0,1]^{d-i+1}} p_d(t_{<i},t_{\geq i})\lambda_{d-i+1}(dt_{\geq i})$ is the marginal distribution of the first $i-1$ coordinates, $t_{<i}\equiv(t_1,\cdots,t_{i-1})$.
- If $p_d$ is a probability density function over $[0,1]^d$, then

$$p_d^i(t_i|t_{<i}) \equiv \begin{cases} p_d^{<i+1}(t_i) & i=1 \\ p_d^{<i+1}(t_{<i},t_i)/p_d^{<i}(t_{<i}) & 1<i<d \\ p_d(t_{<i},t_i)/p_d^{<i}(t_{<i}) & i=d \end{cases}$$

is the conditional marginal distribution of the $i$th coordinate, $t_i$, given the first $i-1$ coordinates, $t_{<i}$.

With these variables, we can define a sequence of discrete-time, infinite-horizon Markov decision problems, indexed by $d \in \mathbb{N}$. The $d$th problem has $d$-dimensional state space $[0,1]^d$, along with action space $\mathbb{a}$ and discount factor $\beta \in [0,1)$. In the $d$th problem, taking action $a \in \mathbb{a}$ in state $s \in [0,1]^d$ yields utility $u_{da}(s)$ and sets the probability density of the subsequent period's state vector to $p_{da}(t|s)$, for $t \in [0,1]^d$. As always, we can decompose this joint density into a product of conditional marginal densities:

$$p_{da}(t|s) = \prod_{i=1}^{d} p_{da}^{i}(t_i|s, t_{<i}). \tag{1}$$

Now with $u_{da}$ and $p_{da}$ we can define Bellman operator $\Gamma_d : \mathbb{B}_d \to \mathbb{B}_d$, where for $V \in \mathbb{B}_d$ and $s \in [0,1]^d$

$$(\Gamma_d V)(s) \equiv \sup_{a \in \mathbb{a}} u_{da}(s) + \beta \int_{t \in [0,1]^d} V(t) p_{da}(t|s) \lambda_d(dt).$$

This operator has a unique fixed-point solution that satisfies

$$\Gamma_d V_d = V_d. \tag{2}$$

$V_d$ is the dynamic program's *value function*. It characterizes the expected discounted value of all utilities received from a given initial state under the optimal sequence of actions. Deriving value function $V_d$ from primitives $u_{da}$ and $p_{da}$ is our objective. To facilitate this end, we impose some regularity conditions on our primitives.

**Assumption 1.** *The transition density function is bounded by a polynomial function of $d$: there exists $M \in \mathbb{b}$ such that $\sup_{d \in \mathbb{N}} \sup_{a \in \mathbb{a}} \sup_{s,t \in [0,1]^d} p_{da}(t|s)/M_d \leq 1$.*

**Assumption 2.** *The transition density function is strictly positive: $p_{da}(t|s) > 0$.*

**Assumption 3.** *The utility function is bounded by a polynomial function of $d$: there exists $K \in \mathbb{b}$ such that $\sup_{d \in \mathbb{N}} \sup_{a \in \mathbb{a}} \sup_{s \in [0,1]^d} |u_{da}(s)|/K_d \leq 1$.*

**Assumption 4.** *The action space is finite: $|\mathbb{a}| \in \mathbb{N}$.*

**Assumption 5.** *The transition density function is Lipschitz continuous in its second argument: for each* $d \in \mathbb{N}$ *and* $t \in [0,1]^d$*, there exists Lipschitz constant* $\ell_d(t) \in \mathbb{R}_+$ *such that* $\sup_{a \in \mathbb{a}} \sup_{r \in [0,1]^d} \sup_{s \in [0,1]^d \setminus r} \frac{|p_{da}(t|s) - p_{da}(t|r)|}{\ell_d(t) \|s - r\|_2} \leq 1$.

**Assumption 6.** *The square integral of the transition density Lipschitz function is bounded by a polynomial function of d: there exists* $L \in \mathbb{b}$ *such that* $\sup_{d \in \mathbb{N}} \int_{t \in [0,1]^d} \ell_d(t)^2 \lambda_d(dt) / L_d \leq 1$.

**Assumption 7.** *At the origin, the transition density function is bounded by a polynomial function of d: there exists* $M \in \mathbb{b}$ *such that* $\sup_{d \in \mathbb{N}} \sup_{a \in \mathbb{a}} \sup_{t \in [0,1]^d} p_{da}(t|0) / M_d \leq 1$.

We will not impose all of these assumptions simultaneously. Instead, we will either use Assumptions 1–4 or 2–7. Rust imposed Assumption 1 but didn't assign it a formal assumption number because he claimed to use it for "notational convenience" only. However, his argument does require this assumption, as his equation (6.11) is incorrect without it. Assumption 2 is not essential. I impose it simply to prevent the denominator in expression (3) from being zero (a problem Rust didn't consider). However, a slight modification of Rust's algorithm would also obviate this issue. Assumption 3 is a weaker version of Rust's Assumption A2: whereas a Lipschitz function over the unit square is always bounded, a bounded function over the unit square is not always Lipschitz. Rust imposes Assumptions 4 and 5 as they appear here, but did not impose Assumption 6. However, Blondel and Tsitsiklis (2000, p. 1268) noted that this assumption must hold for Rust's results to have any bite. Finally, Assumption 7 is a relaxed version of Assumption 1, which Rust used.

## 3 Breaking the Curse of Dimensionality

In general, we cannot exactly solve the system in (2) because it comprises a continuum of equations. However, we can approximate its fixed-point solution to any degree of precision by discretizing the state space, $[0,1]^d$, into a finite grid. Unfortunately, this approach usually requires the number of grid points—and hence the computational difficulty—to grow exponentially with $d$. This is the dynamic program's infamous curse of dimensionality.

Rust (1997b), however, showed that we can sometimes circumvent this curse of dimensionality by representing the state space with a small number of randomly sampled points. Rather than cover $[0,1]^d$ with a grid that grows exponentially with $d$, Rust proposed surveying it with a statistical sample that grows polynomially with $d$. Specifically, he replaced Bellman operator $\Gamma_d$ with random

Bellman operator $\hat{\Gamma}_d^b$, where

$$(\hat{\Gamma}_d^b V)(s) \equiv \sup_{a \in \mathbb{a}} u_{da}(s) + \beta \frac{\sum_{i=1}^{b_d} V(m_d^i) p_{da}(m_d^i|s)}{\sum_{i=1}^{b_d} p_{da}(m_d^i|s)}, \tag{3}$$

$b = \{b_d\}_{d \in \mathbb{N}} \in \mathbb{b}$ is a polynomially bounded sequence, and $m_d = \{m_d^i\}_{i=1}^{b_d}$ is a collection of *i.i.d.* random variables drawn uniformly from $[0,1]^d$ (these values remain fixed for the duration of the algorithm).[1]

Our two Bellman operators share two useful properties. First, both $\Gamma_d$ and $\hat{\Gamma}_d^b$ are contraction mappings with Lipschitz constant $\beta$, so sequences $V_d^{(j+1)} \equiv \Gamma_d V_d^{(j)}$ and $\hat{V}_d^{b(j+1)} \equiv \hat{\Gamma}_d^b \hat{V}_d^{b(j)}$ converge to respective fixed points $V_d$ and $\hat{V}_d^b$ at linear rate $\beta$.[2] I'll refer to the latter fixed point as the *Rust value function*. Second, both $\Gamma_d$ and $\hat{\Gamma}_d^b$ map $\mathbb{V}_d \equiv \{V \in \mathbb{B}_d : \|V\| \leq K_d/(1-\beta)\}$ to $\mathbb{V}_d$, which implies that $V_d \in \mathbb{V}_d$ and $\hat{V}_d^b \in \mathbb{V}_d$ (see Rust's (1997b) Lemma 2.3).

The $\hat{\Gamma}_d^b$ operator has a third useful property: $\{\hat{V}_d^{b(j-1)}(m_d^i)\}_{i=1}^{b_d}$ are sufficient statistics for $(\hat{\Gamma}_d^b \hat{V}_d^{b(j-1)})(s)$, for all $s \in [0,1]^d$. Accordingly, the $j$th iteration of the algorithm need only translate $\{\hat{V}_d^{b(j-1)}(m_d^i)\}_{i=1}^{b_d}$ into $\{\hat{V}_d^{b(j)}(m_d^i)\}_{i=1}^{b_d}$. And with the contraction mapping property, this fact implies that the Rust value function is computable to any degree of precision in polynomial time. Hence, Rust's algorithm breaks the curse of dimensionality when the error between the Rust value function and the true value function is appropriately small. Unfortunately, we can't *guarantee* there won't be a meaningful error, but we can make the probability of the error being meaningful arbitrarily small, as the following definitions and propositions establish.

**Definition 1.** *A sequence of dynamic programs is* strongly Rust solvable *if for all* $\epsilon > 0$ *there exists* $b \in \mathbb{b}$ *such that* $\sup_{d \in \mathbb{N}} \mathrm{E}(\|\hat{V}_d^b - V_d\|) < \epsilon$.

**Definition 2.** *A sequence of dynamic programs is* weakly Rust solvable *if for all* $\epsilon > 0$ *there exists* $b \in \mathbb{b}$ *such that* $\sup_{d \in \mathbb{N}} \mathrm{E}(\|\hat{V}_d^b - V_d\|_1) < \epsilon$.

**Proposition 1.** *A sequence of dynamic programs that satisfies Assumptions 2–7 is strongly Rust solvable.*

**Proposition 2.** *A sequence of dynamic programs that satisfies Assumptions 1–4 is weakly Rust solvable.*

[1] My argument extends to the case in which $m_d$ are drawn from a general density function (see the online appendix).

[2] That is, $\|V_d - V_d^{(j+1)}\|/\|V_d - V_d^{(j)}\| \to \beta$ and $\|\hat{V}_d - \hat{V}_d^{b(j+1)}\|/\|\hat{V}_d - \hat{V}_d^{b(j)}\| \to \beta$ as $j \to \infty$.

Proposition 1 is a slight generalization of Rust's primary finding.[3] Proposition 2, however, is an entirely new result. This latter proposition has two incremental benefits. First, it relies on simpler assumptions that are easier to establish. Specifically, it is more straightforward to verify Assumption 1, which directly bounds the density function, than to verify Assumption 6, which bounds the square integral of a function of Lipschitz constants that restrict the growth rate of the density function. Second, this proposition has a much simpler logical underpinning. For example, the proof of Proposition 1 hinges on several abstruse empirical process results, whereas the proof of Proposition 2 uses nothing more complicated than Chebyshev's inequality.

In exchange for these two benefits, we must downgrade strong Rust solvability to weak Rust solvability. The former establishes a high probability of a good value function estimate across the entire state space (i.e., closeness under the $L^\infty$ norm), whereas the latter establishes a high probability of a good value function estimate across all but an arbitrarily small region of the state space (i.e., closeness under the $L^1$ norm). In practice, however, there's no meaningful difference between a good estimate for 100% of states and a good estimate for 99.9999% of states.

# 4 Limitation of Randomization

The following definitions and proposition illustrate the limited scope of the preceding results.

**Definition 3.** *A dynamic program's ith state variable is* $\epsilon$-dependent *if* $\|V_d^{-i} - V_d\|_1 < \epsilon$*, where* $V_d^{-i}$ *is the value function under density function* $p_{da}^{-i}(t|s) \equiv p_{da}(t|s)/p_{da}^i(t_i|s, t_{<i})$.

**Definition 4.** *A sequence of dynamic programs is* nearly memoryless *if for all* $\epsilon > 0$ *the number of state variables that are not* $\epsilon$*-dependent is* $O(\log(d))$ *as* $d \to \infty$.

**Definition 5.** *A sequence of dynamic programs with Rust value functions* $\{\underline{\hat{V}}_d^b\}_{d\in\mathbb{N}}$ $\epsilon$-approximates *a sequence of dynamic programs with value functions* $\{V_d\}_{d\in\mathbb{N}}$ *if there exists* $b \in \mathbb{b}$ *for which* $\sup_{d\in\mathbb{N}} \mathrm{E}(\|\underline{\hat{V}}_d^b - V_d\|_1) < \epsilon$.

**Proposition 3.** *Any weakly or strongly Rust solvable sequence of dynamic programs that satisfies Assumptions 2–4 can be* $\epsilon$*-approximated by a nearly memoryless sequence of dynamic programs, for all* $\epsilon > 0$.

Density function $p_{da}^{-i}$ equals density function $p_{da}$, but with the $i$th state variable, $t_i$, set to a uniformly distributed random variable, independent of $s$, $a$, and $t_{<i}$. Hence, if the $i$th state

[3] Whereas Rust bounded the transition density function uniformly across $s \in [0, 1]^d$, I bound it only at $s = 0$, and whereas Rust required the utility function to be Lipschitz and bounded, I require it only to be bounded.

variable is $\epsilon$-dependent then it can be replaced with a sequence of $i.i.d.$ exogenous shocks without changing the value function by more than $\epsilon$. And if the dynamic program is nearly memoryless then all but an $O(\log(d)/d)$ fraction of its state variables can be replaced in this fashion with history-independent uniform random shocks. To put it differently, density function $p_{da}(t|s)$ has $d$ channels of communication through which the current period's variables, $(s, a)$, can influence the next period's state, $t$: namely, the $d$ conditional marginal distributions, $\{p^i_{da}(t_i|s, t_{<i})\}_{i=1}^d$ (see (1)). Density function $p^{-i}_{da}(t|s) \equiv \prod_{j\in\{1,\cdots,d\}\setminus i} p^j_{da}(t_j|s, t_{<j})$ equals $p_{da}(t|s)$, with the $i$th channel "turned off." A sequence of dynamic programs is nearly memoryless if all but $O(\log(d))$ of these communication channels can be thus nullified.

Proposition 3 establishes that a dynamic program with a positive transition density function, a polynomially bounded transition density function, and a finite action space is not Rust solvable unless it has a nearly memoryless analog that approximates it arbitrarily well. Hence, we find that Rust's algorithm does not break the curse of dimensionality unless the given dynamic program can be reworked into a nearly memoryless form.

## Acknowledgments

I would like to thank John Rust for his tireless and selfless efforts over many iterations of feedback, the peer review team for their very constructive criticism, and Øystein Daljord for his insight, advice, and inspiration.

## Proofs

*Proposition 1 Proof.* For $d \in \mathbb{N}$, $V \in \mathbb{V}_d$, $a \in \mathbb{a}$, and $b \in \mathbb{b}$ define

$$\begin{aligned}
(\Gamma_{da}V)(s) \equiv& u_{da}(s) + \beta \int_{t\in[0,1]^d} V(t)p_{da}(t|s)\lambda_d(dt),\\
(\hat{\Gamma}^b_{da}V)(s) \equiv& u_{da}(s) + \beta \frac{\sum_{i=1}^{b_d} V(m^i_d)p_{da}(m^i_d|s)}{\sum_{i=1}^{b_d} p_{da}(m^i_d|s)},\\
(\tilde{\Gamma}^b_{da}V)(s) \equiv& u_{da}(s) + \beta \sum_{i=1}^{b_d} V(m^i_d)p_{da}(m^i_d|s)/b_d,\\
\text{and} \quad Z^b_{da}(s) \equiv& \beta \sum_{i=1}^{b_d} g^i_d V(m^i_d)p_{da}(m^i_d|s)/\sqrt{b_d},
\end{aligned}$$

where $\{g_d^i\}_{i=1}^{b_d}$ is a set of independent standard normal random variables. Now since $\mathrm{E}((\tilde{\Gamma}_{da}^b V)(s)) = (\Gamma_{da} V)(s)$, Pollard's (1989) seventh equation implies that

$$\mathrm{E}\left(\|\tilde{\Gamma}_{da}^b V - \Gamma_{da} V\|^2\right) \leq \frac{2\pi}{\sqrt{b_d}} \mathrm{E}\Big(\sup_{s\in[0,1]^d} Z_{da}^b(s)^2\Big). \tag{4}$$

We will now bound the expectation on the right. First, note that

$$\begin{aligned}\mathrm{E}(|Z_{da}^b(t) - Z_{da}^b(s)|^2 : m_d) =& \beta^2 \sum_{i=1}^{b_d} |V(m_d^i)(p_{da}(m_d^i|t) - p_{da}(m_d^i|s))|^2 / b_d \\ \leq& \left(\frac{\beta K_d \|t-s\|_2}{1-\beta}\right)^2 \sum_{i=1}^{b_d} \ell_d(m_d^i)^2 / b_d. \end{aligned} \tag{5}$$

This expression implies that

$$\begin{aligned}\delta_d^m \equiv& \frac{\beta K_d}{1-\beta} \sqrt{\sum_{i=1}^{b_d} d\ell_d(m_d^i)^2 / b_d} \\ \geq& \sup_{s,t\in[0,1]^d} \sqrt{\mathrm{E}(|Z_{da}^b(t) - Z_{da}^b(s)|^2 : m_d)}. \end{aligned} \tag{6}$$

Now for a given $x > 0$, divide $[0,1]^d$ into $f(x) \equiv \lceil \delta_d^m/(2x) \rceil^d$ equally sized cubes, and define $\{c_x^i\}_{i=1}^{f(x)}$ as the center points of these cubes. By design, these center points satisfy

$$\sup_{s\in[0,1]^d} \min_{i\in\{1,\cdots,f(x)\}} \|s - c_x^i\|_2 \leq \frac{x(1-\beta)}{\beta K_d \sqrt{\sum_{i=1}^{b_d} \ell_d(m_d^i)^2 / b_d}},$$

which with (5) implies that

$$\sup_{s\in[0,1]^d} \min_{i\in\{1,\cdots,f(x)\}} \mathrm{E}(|Z_{da}^b(s) - Z_{da}^b(c_x^i)|^2 : m_d) \leq x^2. \tag{7}$$

Now combining (6) and (7) with Pollard's (1989) eighth equation yields

$$\begin{aligned}\sqrt{\mathrm{E}\Big(\sup_{s\in[0,1]^d} Z_{da}^b(s)^2 : m_d\Big)} - \sqrt{\mathrm{E}(Z_{da}^b(0)^2 : m_d)} \\ \leq & C\int_0^{\delta_d^m} \sqrt{\log(f(x))}dx \\ = & C\delta_d^m\sqrt{d}\int_0^1 \sqrt{\log\big(1/u\big)}du \\ = & C\delta_d^m\sqrt{d\pi}/2,\end{aligned}$$

where $C \geq 1$ is a universal constant that is independent of all model parameters. And since $x - y \leq z$ implies $x^2 \leq 2y^2 + 2z^2$ for all $x, y, z \in \mathbb{R}$, this implies that

$$\begin{aligned}\mathrm{E}\Big(\sup_{s\in[0,1]^d} Z_{da}^b(s)^2\Big) = & \mathrm{E}\Big(\mathrm{E}\Big(\sup_{s\in[0,1]^d} Z_{da}^b(s)^2 : m_d\Big)\Big) \\ \leq & \mathrm{E}\Big(2\,\mathrm{E}(Z_{da}^b(0)^2 : m_d) + (C\delta_d^m)^2 d\pi\Big) \\ \leq & 2\Big(\frac{\beta K_d M_d}{1-\beta}\Big)^2 + \pi\Big(\frac{d\beta C K_d}{1-\beta}\Big)^2 \mathrm{E}(\ell_d(m_d^i)^2) \\ \leq & \Big(\frac{\beta K_d}{1-\beta}\Big)^2 (2M_d^2 + \pi d^2 C^2 L_d).\end{aligned}$$

Now combining this with (4) and applying Jensen's inequality yields for all $V \in \mathbb{V}_d$

$$\begin{aligned}\mathrm{E}(\|\tilde{\Gamma}_{da}^b V - \Gamma_{da} V\|) \leq & \sqrt{\mathrm{E}\left(\|\tilde{\Gamma}_{da}^b V - \Gamma_{da} V\|^2\right)} \\ \leq & \sqrt{\frac{2\pi}{\sqrt{b_d}}\,\mathrm{E}\Big(\sup_{s\in[0,1]^d} Z_{da}^b(s)^2\Big)} \\ \leq & \frac{2\pi\beta K_d}{b_d^{1/4}(1-\beta)}\sqrt{M_d^2 + d^2C^2L_d}. \qquad (8)\end{aligned}$$

Next, define constant function $\iota_d \in \mathbb{V}_d$, where $\iota_d(t) = K_d/(1-\beta)$ for all $t \in [0,1]^d$. With this, (8) yields

$$\begin{aligned}
\mathrm{E}(\|\hat{\Gamma}^b_{da}V_d - \tilde{\Gamma}^b_{da}V_d\|) =& \mathrm{E}\left(\sup_{s\in[0,1]^d}\left|\Big(1-\sum_{i=1}^{b_d} p_{da}(m_d^i|s)/b_d\Big)\beta\frac{\sum_{i=1}^{b_d} V(m_d^i)p_{da}(m_d^i|s)}{\sum_{i=1}^{b_d} p_{da}(m_d^i|s)}\right|\right)\\
\leq& \frac{\beta K_d}{1-\beta}\,\mathrm{E}\left(\sup_{s\in[0,1]^d}\left|1-\sum_{i=1}^{b_d} p_{da}(m_d^i|s)/b_d\right|\right)\\
=& \mathrm{E}\left(\|\tilde{\Gamma}^b_{da}\iota_d - \Gamma_{da}\iota_d\|\right)\\
\leq& \frac{2\pi\beta K_d}{b_d^{1/4}(1-\beta)}\sqrt{M_d^2 + d^2C^2L_d}.
\end{aligned}$$

Combining this with (8) yields

$$\begin{aligned}
\mathrm{E}\left(\|\hat{\Gamma}^b_d V_d - V_d\|\right) =& \mathrm{E}\left(\|\hat{\Gamma}^b_d V_d - \Gamma_d V_d\|\right)\\
\leq& \sum_{a\in\mathbb{a}} \mathrm{E}\left(\|\hat{\Gamma}^b_{da}V_d - \Gamma_{da}V_d\|\right)\\
\leq& \sum_{a\in\mathbb{a}} \mathrm{E}\left(\|\hat{\Gamma}^b_{da}V_d - \tilde{\Gamma}^b_{da}V_d\|\right) + \mathrm{E}\left(\|\tilde{\Gamma}^b_{da}V_d - \Gamma_{da}V_d\|\right)\\
\leq& \frac{4|\mathbb{a}|\pi\beta K_d}{b_d^{1/4}(1-\beta)}\sqrt{M_d^2 + d^2C^2L_d}.
\end{aligned}$$

And with this, Rust's (1997b) Lemma 2.2 establishes that

$$\begin{aligned}
\mathrm{E}\left(\|\hat{V}^b_d - V_d\|\right) \leq& \mathrm{E}\left(\|\hat{\Gamma}^b_d V_d - V_d\|\right)/(1-\beta)\\
\leq& \frac{4|\mathbb{a}|\pi\beta K_d}{b_d^{1/4}(1-\beta)^2}\sqrt{M_d^2 + d^2C^2L_d},
\end{aligned}$$

which is smaller than $\epsilon$ when $b_d$ is larger than $\left(\frac{4|\mathbb{a}|\pi\beta K_d}{\epsilon(1-\beta)^2}\sqrt{M_d^2 + d^2C^2L_d}\right)^4$. □

*Proposition 2 Proof.* Assumptions 1 and 3 ensure that $|V(t)p_{da}(t|s)| \leq K_d M_d/(1-\beta)$ for all $V \in \mathbb{V}$. Hence, for $V \in \mathbb{V}$, Popoviciu's inequality implies that $\mathrm{Var}(V(m_d^i)p_{da}(m_d^i|s)) \leq K_d^2 M_d^2/(1-\beta)^2$ and thus that

$$\mathrm{Var}\Big(\sum_{i=1}^{b_d} V(m_d^i)p_{da}(m_d^i|s)/b_d\Big) \leq \frac{K_d^2 M_d^2}{b_d(1-\beta)^2}.$$

Accordingly, Chebyshev's inequality establishes, for a given $\delta > 0$ and $V \in \mathbb{V}$, that

$$\begin{aligned}
&\Pr(y_d^b(s) = 1) \leq \frac{K_d^2 M_d^2}{\delta^2 b_d (1-\beta)^2}, \\
&\text{where} \quad y_d^b(s) \equiv \mathbb{1}\left\{ \left| \sum_{i=1}^{b_d} V(m_d^i) p_{da}(m_d^i|s)/b_d - \int_{t\in[0,1]^d} V_d(t) p_{da}(t|s) \lambda_d(dt) \right| > \delta \right\}.
\end{aligned}$$

And this implies for $V \in \mathbb{V}$ that

$$\begin{aligned}
&\mathrm{E}(|(\tilde{\Gamma}_{da}^b V)(s) - (\Gamma_{da} V)(s)|) \\
&= \beta \, \mathrm{E}\left( \left| \sum_{i=1}^{b_d} V(m_d^i) p_{da}(m_d^i|s)/b_d - \int_{t\in[0,1]^d} V_d(t) p_{da}(t|s) \lambda_d(dt) \right| \right) \\
&\leq \Pr(y_d^b(s) = 0) \, \mathrm{E}\left( \left| \sum_{i=1}^{b_d} V(m_d^i) p_{da}(m_d^i|s)/b_d - \int_{t\in[0,1]^d} V_d(t) p_{da}(t|s) \lambda_d(dt) \right| : y_d^b(s) = 0 \right) \\
&\qquad + \Pr(y_d^b(s) = 1) \, \mathrm{E}\left( \left| \sum_{i=1}^{b_d} V(m_d^i) p_{da}(m_d^i|s)/b_d - \int_{t\in[0,1]^d} V_d(t) p_{da}(t|s) \lambda_d(dt) \right| : y_d^b(s) = 1 \right) \\
&\leq 1 \cdot \delta + \frac{K_d^2 M_d^2}{\delta^2 b_d (1-\beta)^2} \cdot 2 K_d M_d / (1-\beta) \\
&= \delta + \frac{2 K_d^3 M_d^3}{\delta^2 b_d (1-\beta)^3},
\end{aligned}$$

where $\tilde{\Gamma}_{da}^b$ and $\Gamma_{da}$ are defined in the proof of Proposition 1. Hence, for $V \in \mathbb{V}$ we have

$$\mathrm{E}(\|\tilde{\Gamma}_{da}^b V - \Gamma_{da} V\|_1) \leq \delta + \frac{2 K_d^3 M_d^3}{\delta^2 b_d (1-\beta)^3}. \tag{9}$$

And this, in turn, implies that

$$\begin{aligned}
\mathrm{E}(\|\hat{\Gamma}_{da}^b V_d - \tilde{\Gamma}_{da}^b V_d\|_1) &= \mathrm{E}\left( \int_{s\in[0,1]^d} \left| \left(1 - \sum_{i=1}^{b_d} p_{da}(m_d^i|s)/b_d\right) \beta \frac{\sum_{i=1}^{b_d} V_d(m_d^i) p_{da}(m_d^i|s)}{\sum_{i=1}^{b_d} p_{da}(m_d^i|s)} \right| \lambda_d(ds) \right) \\
&\leq \frac{\beta K_d}{1-\beta} \, \mathrm{E}\left( \int_{s\in[0,1]^d} \left| 1 - \sum_{i=1}^{b_d} p_{da}(m_d^i|s)/b_d \right| \lambda_d(ds) \right) \\
&= \mathrm{E}\left( \|\tilde{\Gamma}_{da}^b \iota_d - \Gamma_{da} \iota_d\|_1 \right) \\
&\leq \delta + \frac{2 K_d^3 M_d^3}{\delta^2 b_d (1-\beta)^3},
\end{aligned} \tag{10}$$

where $\hat{\Gamma}^b_{da}$ and $\iota_d$ are defined in the proof of Proposition 1. Now combining (9) and (10) yields

$$\begin{aligned}
\mathrm{E}\left(\|\hat{\Gamma}^b_d V_d - V_d\|_1\right) =& \mathrm{E}\left(\|\hat{\Gamma}^b_d V_d - \Gamma_d V_d\|_1\right) \\
\leq& \sum_{a\in\mathbb{a}} \mathrm{E}\left(\|\hat{\Gamma}^b_{da} V_d - \Gamma_{da} V_d\|_1\right) \\
\leq& \sum_{a\in\mathbb{a}} \mathrm{E}\left(\|\hat{\Gamma}^b_{da} V_d - \tilde{\Gamma}^b_{da} V_d\|_1\right) + \mathrm{E}\left(\|\tilde{\Gamma}^b_{da} V_d - \Gamma_{da} V_d\|_1\right) \\
\leq& 2|\mathbb{a}|\delta + \frac{4|\mathbb{a}|K_d^3 M_d^3}{\delta^2 b_d (1-\beta)^3}.
\end{aligned}$$

And with this, Rust's (1997b) Lemma 2.2 establishes that

$$\begin{aligned}
\mathrm{E}\left(\|\hat{V}^b_d - V_d\|_1\right) \leq& \mathrm{E}\left(\|\hat{\Gamma}^b_d V_d - V_d\|_1\right)/(1-\beta) \\
\leq& 2|\mathbb{a}|\delta/(1-\beta) + \frac{4|\mathbb{a}|K_d^3 M_d^3}{\delta^2 b_d (1-\beta)^4},
\end{aligned}$$

which is less than $\epsilon$ when $\delta < \frac{\epsilon(1-\beta)}{4|\mathbb{a}|}$ and $b_d > \frac{8|\mathbb{a}|K_d^3 M_d^3}{\delta^2 \epsilon (1-\beta)^4}$. □

**Definition 6.** *$\kappa(p,q) \equiv \int_{t\in[0,1]} p(t)\log\big(p(t)/q(t)\big)\lambda_1(dt)$ is the Kullback-Leibler divergence between densities $p$ and $q$, which are defined with full support over $[0,1]$.*

**Lemma 1.** *For Assumption 1 to hold, there must exist $m, n \in \mathbb{N}$ such that* $\sup_{d\in\mathbb{N}} \sup_{a\in\mathbb{a}} \sup_{s\in[0,1]^d} \frac{\sum_{i=1}^d \mathrm{E}(\kappa(p^i_{da}(\cdot|t_{<i},s),\lambda_1))}{m+n\log(d)} < 1$.

*Proof.* Define $H_{da}(s) \equiv -\int_{t\in[0,1]^d} \log(p_{da}(t|s)) p_{da}(t|s)\lambda_d(dt)$ as the differential entropy of distribution $p_{da}(\cdot|s)$, and define $H^i_{da}(s,t_{<i}) \equiv -\int_{t_i\in[0,1]} \log\big(p^i_{da}(t_i|s,t_{<i})\big) p^i_{da}(t_i|s,t_{<i})\lambda_1(dt_i)$ as the differential entropy of conditional marginal distribution $p^i_{da}(\cdot|s,t_{<i})$. The differentiable entropy is zero for the uniform distribution and negative for all other distributions (see Marsh, 2013, p. 9).

The entropy chain rule states that the total entropy equals the sum of the expected conditional marginal entropies: $H_{da}(s) = \sum_{i=1}^d \mathrm{E}(H^i_{da}(s,t_{<i}))$. And the conditional marginal entropy equals the negative Kullback-Leibler divergence between the conditional marginal distribution and the standard uniform distribution: $H^i_{da}(s,t_{<i}) = -\kappa(p^i_{da}(\cdot|s,t_{<i}),\lambda_1)$. Together, these two properties

imply that

$$\begin{aligned}
\sup_{t\in[0,1]^d} p_{da}(t|s) &= \exp\Big(\sup_{t\in[0,1]^d} \log(p_{da}(t|s))\Big) \\
&\geq \exp\Big(\int_{t\in[0,1]^d} \log(p_{da}(t|s))p_{da}(t|s)\lambda_d(dt)\Big) \\
&= \exp(-H_{da}(s)) \\
&= \exp\Big(-\sum_{i=1}^{d} \mathrm{E}\left(H^i_{da}(s,t_{<i})\right)\Big) \\
&= \exp\Big(\sum_{i=1}^{d} \mathrm{E}\left(\kappa(p^i_{da}(\cdot|s,t_{<i}),\lambda_1)\right)\Big),
\end{aligned}$$

which implies the result. □

**Lemma 2.** *Any sequence of dynamic programs that satisfies Assumptions 1–4 is nearly memoryless.*

*Proof.* To outline the proof, I first use Pinsker's inequality to bound the distance between $p^{-i}_{da}(\cdot|s)$ and $p_{da}(\cdot|s)$ with the expected distance between $p^i_{da}(\cdot|s,t_{<i})$ and $\lambda_1$. Intuitively, if the expected distance between $p^i_{da}(\cdot|s,t_{<i})$ and $\lambda_1$ is small then $p^i_{da}(t_i|s,t_{<i})$ must be close to one with high probability, which means that $p^{-i}_{da}(t|s)$ must be close to $p_{da}(t|s)$ with high probability. Second, I use Lemma 1 to establish that the expected distance between $p^i_{da}(\cdot|s,t_{<i})$ and $\lambda_1$ is small for all but $O(\log(d))$ values of $i$, which means that the distance between $p^{-i}_{da}(\cdot|s)$ and $p_{da}(\cdot|s)$ is small for all but $O(\log(d))$ values of $i$. This fact holds for all $s\in[0,1]^d$, which means that the distance between $p^{-i}_{da}(\cdot|s)$ and $p_{da}(\cdot|s)$ is small for almost all $s$, for all but $O(\log(d))$ values of $i$. Third, I use this fact to establish that $\int_{s\in[0,1]^d} \|p^{-i}_{da}(\cdot|s) - p_{da}(\cdot|s)\|_1 \lambda_d(ds)$ is small for all but $O(\log(d))$ values of $i$. And after this integral is bounded, the rest is fairly straightforward.

Pinsker's inequality establishes that $\|\lambda_1 - p^i_{da}(\cdot|s,t_{<i})\|_1 \leq \sqrt{2\kappa(p^i_{da}(\cdot|s,t_{<i}),\lambda_1)}$, which in turn

implies that

$$\begin{aligned}
&\int_{t_{\geq i}\in[0,1]^{d-i+1}} \left|p_{da}^{-i}(t_{<i}, t_{\geq i}|s) - p_{da}(t_{<i}, t_{\geq i}|s)\right| \lambda_{d-i+1}(dt_{\geq i}) \\
&\qquad = \int_{t_{\geq i}\in[0,1]^{d-i+1}} \left|1/p_{da}^{i}(t_i|s, t_{<i}) - 1\right| p_{da}(t_{<i}, t_i, t_{\geq i+1}|s) \lambda_{d-i+1}(dt_{\geq i}) \\
&\qquad = \int_{t_i\in[0,1]} \left|1/p_{da}^{i}(t_i|s, t_{<i}) - 1\right| p_{d}^{<i+1}(t_{<i}, t_i) \lambda_1(dt_i) \\
&\qquad = p_{d}^{<i}(t_{<i}) \int_{t_i\in[0,1]} \left|1 - p_{da}^{i}(t_i|s, t_{<i})\right| \lambda_1(dt_i) \\
&\qquad = p_{d}^{<i}(t_{<i}) \left\|\lambda_1 - p_{da}^{i}(\cdot|s, t_{<i})\right\|_1 \\
&\qquad \leq p_{d}^{<i}(t_{<i}) \sqrt{2\kappa(p_{da}^{i}(\cdot|s, t_{<i}), \lambda_1)}.
\end{aligned}$$

With this, Jensen's inequality yields

$$\begin{aligned}
\left\|p_{da}^{-i}(\cdot|s) - p_{da}(\cdot|s)\right\|_1 &= \int_{t\in[0,1]^d} \left|p_{da}^{-i}(t|s) - p_{da}(t|s)\right| \lambda_d(dt) \\
&\leq \int_{t_{<i}\in[0,1]^{i-1}} p_{d}^{<i}(t_{<i}) \sqrt{2\kappa(p_{da}^{i}(\cdot|s, t_{<i}), \lambda_1)} \lambda_{i-1}(dt_{<i}) \\
&\leq \sqrt{2 \int_{t_{<i}\in[0,1]^{i-1}} p_{d}^{<i}(t_{<i}) \kappa(p_{da}^{i}(\cdot|s, t_{<i}), \lambda_1) \lambda_{i-1}(dt_{<i})} \\
&= \sqrt{2\,\mathrm{E}\left(\kappa(p_{da}^{i}(\cdot|s, t_{<i}), \lambda_1)\right)}.
\end{aligned}$$

Next, Lemma 1 implies that there exists $m, n \in \mathbb{N}$ such that $\sum_{i=1}^{d} \mathrm{E}\left(\kappa(p_{da}^{i}(\cdot|s, t_{<i}), \lambda_1)\right) < m + n\log(d)$, for all $d \in \mathbb{N}$. Since $\kappa(p_{da}^{i}(\cdot|s, t_{<i}), \lambda_1) \geq 0$, this implies that for a given $\gamma > 0$ the inequality $\sqrt{2\,\mathrm{E}\left(\kappa(p_{da}^{i}(\cdot|s, t_{<i}), \lambda_1)\right)} \leq \gamma$ holds for at least $d - 2(m + n\log(d))/\gamma^2$ values of $i$. And with the result above, this implies that $\left\|p_{da}^{-i}(\cdot|s) - p_{da}(\cdot|s)\right\|_1 < \gamma$ holds for at least $d - 2(m + n\log(d))/\gamma^2$ values of $i$. Now define $\Omega_{da}^{i} = \{s \in [0,1]^d : \left\|p_{da}^{-i}(\cdot|s) - p_{da}(\cdot|s)\right\|_1 < \gamma\}$ as the set points that satisfy

this inequality for a given $i \in \{1, \cdots, d\}$. The Lebesgue measures of this set satisfies

$$\begin{aligned}
\sum_{i=1}^{d} \lambda_d(\Omega_{da}^i) =& \sum_{i=1}^{d} \int_{s\in[0,1]^d} \mathbb{1}\{s \in \Omega_{da}^i\} \lambda_d(ds) \\
=& \int_{s\in[0,1]^d} \sum_{i=1}^{d} \mathbb{1}\{s \in \Omega_{da}^i\} \lambda_d(ds) \\
\geq& \int_{s\in[0,1]^d} \big(d - 2(m + n\log(d))/\gamma^2\big) \lambda_d(ds) \\
=& d - 2(m + n\log(d))/\gamma^2.
\end{aligned}$$

And since $\lambda_d(\Omega_{da}^i) \leq 1$, this implies that for a given $\delta > 0$ there are at least $d - 2(m + n\log(d))/(\delta\gamma^2)$ values of $i \in \{1, \cdots, d\}$ for which we have

$$\lambda_d(\Omega_{da}^i) \geq 1 - \delta. \tag{11}$$

Now let $\Gamma_{da}$ be the action-$a$ Bellman operator defined in the proof of Proposition 1, and let $\Gamma_{da}^{-i}$ be the analogous operator under $p_{da}^{-i}$. If $i$ satisfies (11) then we have

$$\begin{aligned}
\left\|\Gamma_{da}^{-i} V_d - \Gamma_{da} V_d\right\|_1 =& \int_{s\in[0,1]^d} \Bigg| u_{da}(s) + \beta \int_{t\in[0,1]^d} V(t) p_{da}^{-i}(t|s) \lambda_d(dt) \\
& \qquad - u_{da}(s) - \beta \int_{t\in[0,1]^d} V(t) p_{da}(t|s) \lambda_d(dt) \Bigg| \lambda_d(ds) \\
=& \beta \int_{s\in[0,1]^d} \left| \int_{t\in[0,1]^d} V_d(t) (p_{da}^{-i}(t|s) - p_{da}(t|s)) \lambda_d(dt) \right| \lambda_d(ds) \\
\leq& \beta \|V_d\| \int_{s\in[0,1]^d} \left\|p_{da}^{-i}(\cdot|s) - p_{da}(\cdot|s)\right\|_1 \lambda_d(ds) \\
\leq& \beta \|V_d\| \Bigg( \int_{s\in\Omega_{da}^i} \left\|p_{da}^{-i}(\cdot|s) - p_{da}(\cdot|s)\right\|_1 \lambda_d(ds) \\
& \qquad + \int_{s\in[0,1]^d\setminus\Omega_{da}^i} \left(\left\|p_{da}^{-i}(\cdot|s)\right\|_1 + \left\|p_{da}(\cdot|s)\right\|_1\right) \lambda_d(ds) \Bigg) \\
\leq& \beta \|V_d\| \left( \int_{s\in\Omega_{da}^i} \gamma \lambda_d(ds) + \int_{s\in[0,1]^d\setminus\Omega_{da}^i} 2 \lambda_d(ds) \right) \\
=& \beta \|V_d\| \left(\gamma \lambda_d(\Omega_{da}^i) + 2(1 - \lambda_d(\Omega_{da}^i))\right) \\
\leq& \beta (K_d/(1-\beta))(\gamma + 2\delta).
\end{aligned}$$

And thus if $i$ satisfies (11) then we have

$$\begin{aligned}\left\|\Gamma_d^{-i}V_d - V_d\right\|_1 =& \left\|\Gamma_d^{-i}V_d - \Gamma_d V_d\right\|_1 \\ \leq& \sum_{a\in\mathbb{a}} \left\|\Gamma_{da}^{-i}V_d - \Gamma_{da}V_d\right\|_1 \\ \leq& |\mathbb{a}|\beta(K_d/(1-\beta))(\gamma+2\delta),\end{aligned}$$

where $\Gamma_d^{-i}$ is the analogue of $\Gamma_d$ under $p_d^{-i}$. With this, Rust's (1997b) Lemma 2.2 implies that $\left\|V_d^{-i} - V_d\right\|_1 < |\mathbb{a}|\beta K_d(\gamma+2\delta)/(1-\beta)^2$ holds for all $i$ that satisfy (11). Hence, setting $\gamma = \delta = \frac{\epsilon(1-\beta)^2}{4|\mathbb{a}|\beta K_d}$, we find that $\left\|V_d^{-i} - V_d\right\|_1 < \epsilon$ holds for at least $d - 2(m+n\log(d))/(\delta\gamma^2) = d - 128\left(\frac{|\mathbb{a}|\beta K_d}{\epsilon(1-\beta)^2}\right)^3(m+n\log(d))$ values of $i$. □

**Lemma 3.** *Any strongly Rust solvable sequence of dynamic programs is also weakly Rust solvable.*

*Proof.* This holds because the state space, $[0,1]^d$, has Lebesgue measure 1. □

*Proposition 3 Proof.* Define the following probability transition density function:

$$\begin{aligned}\underline{p}_{da}^b(t|s) \equiv& \mathbb{1}\{p_{da}(t|s) \leq b_d^2\}p_{da}(t|s) + b_d^2\mathbb{1}\{\|t\| \leq \exp(-2\log(b_d)/d)\}\overline{p}_{da}(s), \\ \text{where}\quad \overline{p}_{da}(s) \equiv& \int_{r\in[0,1]^d} \mathbb{1}\{p_{da}(r|s) > b_d^2\}p_{da}(r|s)\lambda_d(dr).\end{aligned}$$

This density function funnels all the mass that exceeds $b_d^2$ into a cube with Lebesgue measure $\exp(-2\log(b_d)/d)^d = b_d^{-2}$. Since it never exceeds $2b_d^2$, this density function satisfies Assumption 1. Therefore, Lemma 2 establishes that this density function corresponds with a nearly memoryless sequence of dynamic programs (given Assumptions 2–4).

Define the following as the set of points for which the new density function equals the old density function:

$$\Omega_d^b(s) \equiv \left\{t \in [0,1]^d : \underline{p}_{da}^b(t|s) = p_{da}(t|s)\ \forall\ a \in \mathbb{a}\right\}.$$

The Lebesgue measure of this set satisfies

$$\begin{aligned}
\lambda_d(\Omega_d^b(s)) \geq & 1 - \int_{t\in[0,1]^d} \mathbb{1}\{\|t\| \leq \exp(-2\log(b_d)/d)\}\lambda_d(dt) - \sum_{a\in\mathbb{a}} \int_{t\in[0,1]^d} \mathbb{1}\{p_{da}(r|s) > b_d^2\}\big)\lambda_d(dt) \\
\geq & 1 - b_d^{-2} - b_d^{-2} \sum_{a\in\mathbb{a}} \int_{t\in[0,1]^d} \mathbb{1}\{p_{da}(r|s) > b_d^2\} p_{da}(r|s)\lambda_d(dt) \\
\geq & 1 - b_d^{-2} - b_d^{-2} \sum_{a\in\mathbb{a}} \int_{t\in[0,1]^d} p_{da}(r|s)\lambda_d(dt) \\
\geq & 1 - (1+|\mathbb{a}|)/b_d^2.
\end{aligned}$$

The probability that this set contains all $m_d^i$ values satisfies

$$\begin{aligned}
\Pr\big(\cup_{i=1}^{b_d} m_d^i \subset \Omega_d^b(s)\big) = & \lambda_d(\Omega_d^b(s))^{b_d} \\
\geq & (1 - (1+|\mathbb{a}|)/b_d^2)^{b_d} \\
= & \exp\big(b_d \log(1 - (1+|\mathbb{a}|)/b_d^2)\big) \\
= & \exp\big(b_d(-(1+|\mathbb{a}|)/b_d^2 - (1+|\mathbb{a}|)^2/(2b_d^4) - (1+|\mathbb{a}|)^3/(3b_d^6) - \cdots)\big) \\
> & \exp(-(1+|\mathbb{a}|)/b_d) \\
> & 1 - (1+|\mathbb{a}|)/b_d. \qquad (12)
\end{aligned}$$

Next, define $\hat{\underline{\Gamma}}_d^b$ as Rust's random Bellman operator evaluated under density function $\underline{p}_{da}^b$. Since $\hat{V}_d^b \in \mathbb{V}_d$, we have

$$|(\hat{\underline{\Gamma}}_d^b \hat{V}_d^b)(s) - (\hat{\Gamma}_d^b \hat{V}_d^b)(s)| \leq \beta K_d/(1-\beta). \qquad (13)$$

If $\cup_{i=1}^{b_d} m_d^i \subset \Omega_d^b(s)$, then

$$|(\hat{\underline{\Gamma}}_d^b \hat{V}_d^b)(s) - (\hat{\Gamma}_d^b \hat{V}_d^b)(s)| = 0. \qquad (14)$$

Now combining (12)–(14) yields

$$\begin{aligned}
\mathrm{E}\left(\|\hat{\underline{\Gamma}}_d^b \hat{V}_d^b - \hat{V}_d^b\|_1\right) &= \mathrm{E}\left(\|\hat{\underline{\Gamma}}_d^b \hat{V}_d^b - \hat{\Gamma}_d^b \hat{V}_d^b\|_1\right) \\
&= \int_{s\in[0,1]^d} \mathrm{E}\left(|(\hat{\underline{\Gamma}}_d^b \hat{V}_d^b)(s) - (\hat{\Gamma}_d^b \hat{V}_d^b)(s)|\right)\lambda_d(ds) \\
&\leq \int_{s\in[0,1]^d} \beta K_d/(1-\beta)\big(1 - \Pr\left(\cup_{i=1}^{b_d} m_d^i \subset \Omega_d^b(s)\right)\big)\lambda_d(ds) \\
&< \int_{s\in[0,1]^d} \frac{\beta K_d(1+|\mathbb{a}|)}{b_d(1-\beta)}\lambda_d(ds) \\
&= \frac{\beta K_d(1+|\mathbb{a}|)}{b_d(1-\beta)}.
\end{aligned}$$

With this, we can use Rust's (1997b) Lemma 2.2 to establish that

$$\mathrm{E}\left(\|\hat{\underline{V}}_d^b - \hat{V}_d^b\|_1\right) \leq \mathrm{E}\left(\|\hat{\underline{\Gamma}}_d^b \hat{V}_d^b - \hat{V}_d^b\|_1\right)/(1-\beta) < \frac{\beta K_d(1+|\mathbb{a}|)}{b_d(1-\beta)^2}. \tag{15}$$

Finally, Lemma 3 establishes that our sequence of dynamic programs is weakly Rust solvable. Thus, we can set $b \in \mathbb{b}$ to satisfy $\mathrm{E}(\|\hat{V}_d^b - V_d\|_1) < \epsilon/2$ and $\frac{\beta K_d(1+|\mathbb{a}|)}{b_d(1-\beta)^2} < \epsilon/2$. And, with this, (15) yields $\mathrm{E}\left(\|\hat{\underline{V}}_d^b - V_d\|_1\right) \leq \mathrm{E}\left(\|\hat{\underline{V}}_d^b - \hat{V}_d^b\|_1\right) + \mathrm{E}(\|\hat{V}_d^b - V_d\|_1) < \epsilon$. □

## References


Blondel, Vincent D., John N. Tsitsiklis. 2000. A survey of computational complexity results in systems and control. *Automatica* **36** 1249–1274.

Chen, Yichen, Mengdi Wang. 2017. Lower Bound On the Computational Complexity of Discounted Markov Decision Problems. *Working Paper* 1–13.

Chow, Chee-seng, John N. Tsitsiklis. 1989. The complexity of dynamic programming. *Journal of Complexity* **5** 466–488.

Kristensen, Dennis, Patrick K. Mogensen, Jong Myun Moon, Bertel Schjerning. 2021. Solving dynamic discrete choice models using smoothing and sieve methods. *Journal of Econometrics* **223**(2) 328–360.

Marsh, Charles. 2013. Introduction to Continuous Entropy. Tech. rep., Princeton University.

Pollard, David. 1989. Asymptotics via empirical processes. *Statistical Science* **4**(4) 341–354.

Rust, John. 1997a. A Comparison of Policy Iteration Methods for Solving Continuous-State, Infinite-Horizon Markovian Decision Problems Using Random, Quasi-random, and Deterministic Discretizations. *Available at SSRN* 1–51.

Rust, John. 1997b. Using randomization to break the curse of dimensionality. *Econometrica* **65**(3) 487–516.

Ye, Yinyu. 2011. The Simplex and Policy-Iteration Methods are Strongly Polynomial for the Markov Decision Problem with a Fixed Discount Rate. *Mathematics of Operations Research* **36**(4) 1–12.