Dynamic optimization

Robert J. Carroll

Last revised: 6 May 2026

---

1 Motivation

Most political-economy choices are made with the future in view. A legislator deciding how much political capital to spend on this session's bills knows that her capital tomorrow depends on what she does today. A regulator imposing a policy this period knows that the constraints she faces next period are partly a function of how this one resolves. A voter forming a posterior over a candidate's competence updates today's beliefs in ways that will color tomorrow's vote. An authoritarian regime planning repression chooses today's intensity knowing it will affect citizen mobilization in subsequent years. In each case, the agent's decision today is one move in an ongoing problem, and the right framework is not the static optimization of the previous handout but its dynamic generalization.

Three structural insights organize the dynamic theory and run through this handout. First, when the problem has a recursive structure — the situation tomorrow depends on today's state and action in a way that respects the same problem-shape next period — the solution can be characterized by a *value function* that satisfies a self-referential equation called the *Bellman equation* (§2–§4). Solving the dynamic problem reduces to solving for the value function. Second, when the horizon is infinite and payoffs are discounted, the Bellman equation is a fixed-point equation in a function space, and the *contraction mapping theorem* (Banach, §5) certifies that the equation has a unique solution and that iteration from any starting point converges to it. This is the structural reason infinite-horizon dynamic programming is computationally tractable, and it is the result behind the *value-function iteration* algorithm that underlies most numerical work in dynamic political-economy modeling. Third, comparative statics in dynamic settings are governed by an *envelope condition* (§6) that is the dynamic analogue of the static envelope theorem from the previous handout: when a state variable shifts, the value function's first-order response is determined by the parameter's direct effect on the one-period reward, with the future-value response captured implicitly through the recursive structure.

The handout closes the optimization cluster. Differentiation supplied gradients, Hessians, the implicit function theorem; convex sets and concave functions supplied the local-implies-global tractability under concavity; static optimization supplied the FOC, SOC, Lagrangian, KKT, Berge, and envelope vocabulary; and now dynamic optimization supplies the recursive-and-fixed-point apparatus. The applied clusters that follow — game theory most immediately, social choice and mechanism design beyond that — will lean on every one of these handouts.

2 Finite-horizon dynamic programming

Some political-economy problems have a natural last period. A legislator serving a four-year term knows the term ends. A candidate running a campaign with a fixed election date sees the campaign as a finite-horizon decision problem. A regulator with a sunset clause on her authority chooses policies over a known stopping date. The right approach in these settings is *backward induction*:

solve the last period first, knowing that no future periods constrain the choice; then work backward, solving each preceding period under the assumption that future-period choices will be optimal. The finite-horizon Bellman recursion is the formalization of this argument, and it is also the natural pedagogical entry point for dynamic programming since it requires no fixed-point machinery.

The setup. There is a finite horizon T, and each period $t = 0, 1, \ldots, T$ the agent observes a *state* s_t in some state space S and chooses an *action* a_t from a feasible set $A(s_t)$. The state evolves according to a transition function $s_{t+1} = f(s_t, a_t)$ (we treat the deterministic case for clarity; the stochastic generalization replaces f with a transition kernel and is structurally identical). The agent receives a one-period reward $r(s_t, a_t)$ each period and an end-of-horizon payoff $r_T(s_T)$. Her total payoff over the horizon is

$$\sum_{t=0}^{T-1} \beta^t r(s_t, a_t) + \beta^T r_T(s_T),$$

with $\beta \in [0, 1]$ a discount factor. (The finite-horizon case allows $\beta = 1$; the infinite-horizon case in §3 will require $\beta < 1$ for convergence.) The agent's task is to choose actions $a_0, a_1, \ldots, a_{T-1}$ to maximize this sum, given the initial state s_0.

Definition 1. The *value function* at period t in the finite-horizon problem is

$$V_t(s) = \max_{a_t, a_{t+1}, \ldots, a_{T-1}} \left[\sum_{\tau=t}^{T-1} \beta^{\tau-t} r(s_\tau, a_\tau) + \beta^{T-t} r_T(s_T) \right]$$

subject to s_τ evolving from $s_t = s$ according to the transition rule and each $a_\tau \in A(s_\tau)$.

The value function records the maximum total discounted payoff achievable starting from state s at period t, with $T - t$ periods to go. The finite-horizon Bellman recursion is the structural recursive characterization of V_t.

Theorem 2 (Finite-horizon Bellman recursion)**.** *The value functions satisfy $V_T(s) = r_T(s)$ and, for $t = 0, 1, \ldots, T-1$,*

$$V_t(s) = \max_{a \in A(s)} \big[r(s, a) + \beta V_{t+1}(f(s, a)) \big].$$

The maximizer in this Bellman equation defines the optimal action at (s, t).

The proof is a standard "one-step optimization plus recursion" argument: the agent's best plan from (s, t) is to pick today's action a to maximize the one-period reward $r(s, a)$ plus the discounted continuation value $V_{t+1}(f(s, a))$, and the latter is itself recursively characterized by the same kind of equation. Backward induction starts at $V_T = r_T$ and works backward to V_0.

Example 3 (Legislative-bargaining over a four-period session)**.** A legislator chooses how much of her per-period political capital c to spend on bills each period $t \in \{0, 1, 2, 3\}$. Spending a_t today yields one-period payoff $r(c_t, a_t) = u(a_t) - \kappa(c_t - a_t)$ (utility from spending plus a holding cost on unspent capital), and capital evolves as $c_{t+1} = c_t - a_t + \delta$ for some replenishment rate δ. The terminal-period payoff is the salvage value $r_T(c_T) = \rho c_T$. Backward induction: at $T = 3$, $V_3(c) = \rho c$. At $T = 2$, $V_2(c) = \max_a [u(a) - \kappa(c - a) + \beta\rho(c - a + \delta)]$, the FOC of which pins down $a_2^*(c)$. Continuing backward to V_1 and V_0 gives the full optimal-policy schedule. The political-economy reading is that the legislator's optimal spending each period depends on her remaining horizon, with the salvage value capturing the longer-run reputation effects of unspent capital.

3 Infinite-horizon problems and discounting

Many political-economy problems have no natural finite horizon. A voter's learning across an indefinitely-long sequence of elections has no built-in end. An authoritarian regime planning policy under stable institutions thinks forward without bound. An optimal-taxation problem under a long-lived government is forward-looking indefinitely. The infinite-horizon version of dynamic programming is the natural setting for these cases. The structural complication — that the total payoff is now an infinite sum — is handled by *discounting*: the agent values future rewards less than present ones, and a strict discount factor $\beta \in [0, 1)$ keeps the infinite sum finite under standard regularity conditions.

The setup. The state space S, action correspondence $A : S \rightrightarrows A$, transition $f : S \times A \to S$, and per-period reward $r : S \times A \to \mathbb{R}$ are now *stationary* (do not depend on t), and the agent chooses an action sequence $\{a_t\}_{t \geq 0}$ to maximize the infinite-horizon discounted payoff

$$\sum_{t=0}^{\infty} \beta^t r(s_t, a_t),$$

subject to $s_{t+1} = f(s_t, a_t)$ and $a_t \in A(s_t)$, with s_0 given. Stationarity is the structurally important condition: the problem looks the same from any starting state, so the optimal policy can be characterized as a stationary function $\pi^* : S \to A$, and the value function as a stationary function $V^* : S \to \mathbb{R}$.

Definition 4. The *value function* of the stationary infinite-horizon problem is

$$V^*(s) = \sup_{\{a_t\}} \sum_{t=0}^{\infty} \beta^t r(s_t, a_t),$$

the supremum taken over feasible action sequences with $s_0 = s$. A *stationary policy* $\pi : S \to A$ is *optimal* if, starting from any $s_0 = s$ and following π, the resulting payoff equals $V^*(s)$.

The boundedness of V^* requires regularity. A sufficient condition the literature usually invokes is that r is bounded on $S \times A$, in which case $V^* \leq \frac{1}{1-\beta} \sup |r| < \infty$. Without bounded rewards, additional conditions (e.g., growth conditions on r along feasible sequences) are needed. Stokey, Lucas, and Prescott (1989, Ch. 3–4) works through the technical hypotheses with care; we proceed under boundedness throughout.

Example 5 (Voter learning over infinitely many elections)**.** A voter chooses each period whether to vote for an incumbent candidate or a challenger; her belief about the incumbent's competence is the state $s_t \in [0, 1]$, evolving via Bayesian updating after each period's electoral outcome. Her per-period reward is the expected utility from the chosen candidate given current belief; her discounted payoff aggregates across an indefinite electoral horizon. The infinite-horizon Bellman framework is the right home for this problem: stationarity is plausible (the underlying institutions don't change period-to-period), the state space is bounded ($s \in [0, 1]$), and the per-period reward is bounded if utilities are.

4 The Bellman equation as a fixed-point equation

The structural insight that makes infinite-horizon dynamic programming tractable is that the value function satisfies a self-referential equation. The optimal value at state s is the best one-period reward plus the optimally-continued value from next period's state, where the continuation value is the same value function V^* evaluated at the new state. This recursion gives the *Bellman equation*, and the dynamic-programming theory's central problem is to solve it.

Theorem 6 (Bellman equation, infinite horizon)**.** *Under the regularity conditions of §3, the value function V^* satisfies*

$$V^*(s) = \max_{a \in A(s)} \big[r(s,a) + \beta V^*(f(s,a))\big],$$

and a stationary policy π^* is optimal iff for every s,

$$\pi^*(s) \in \arg\max_{a \in A(s)} \big[r(s,a) + \beta V^*(f(s,a))\big].$$

The conditions for the maximum to be attained (so that max rather than sup is appropriate) are the Weierstrass / Berge conditions from the previous handout: $A(s)$ compact, r and f continuous in a for each fixed s. Stokey, Lucas, and Prescott (1989, Ch. 4) works through the technical conditions; under them, the Bellman equation is the right characterization of V^*.

The political-economy reading: the agent's decision today (choice of a) trades off today's reward $r(s,a)$ against tomorrow's continuation value $V^*(f(s,a))$, weighted by the discount factor. The optimal policy is the maximizer; the value function is the maximized objective. The two quantities are mutually defined — once you know V^*, the optimal policy is the maximizer; once you know the policy, V^* is the policy's discounted return — and the structural problem is to solve for the pair simultaneously.

The Bellman equation can be re-cast in operator form. Define the *Bellman operator* T on the space of bounded real-valued functions on S by

$$(TV)(s) = \max_{a \in A(s)} \big[r(s,a) + \beta V(f(s,a))\big].$$

Then the Bellman equation says exactly that $V^* = TV^*$ — the optimal value function is a fixed point of T. The next section's contraction-mapping argument shows the fixed point exists, is unique, and can be computed by iteration.

5 The contraction mapping theorem

The structural workhorse for infinite-horizon dynamic programming is the contraction mapping theorem of Banach: a contraction on a complete metric space has a unique fixed point, and iteration from any starting point converges to it geometrically. Applied to the Bellman operator on bounded functions with the sup norm, the theorem certifies that the Bellman equation has a unique solution, that the value function V^* exists and is well-defined, and that value-function iteration — the algorithm $V_{n+1} = TV_n$ — converges to it from any starting guess. The result is the structural reason that dynamic programming is computationally tractable, and it is the formal underpinning of essentially all numerical work in dynamic political-economy modeling.

Definition 7. Let (X,d) be a metric space. A function $T : X \to X$ is a *contraction* (with modulus $\gamma \in [0,1)$) if for every $x, y \in X$,

$$d(Tx, Ty) \le \gamma\, d(x,y).$$

Theorem 8 (Banach fixed-point theorem)**.** *Let (X, d) be a non-empty complete metric space and let $T : X \to X$ be a contraction with modulus γ. Then T has a unique fixed point $x^* \in X$, and for every $x_0 \in X$, the iterates $x_n = T^n(x_0)$ satisfy*

$$d(x_n, x^*) \leq \gamma^n \, d(x_0, x^*),$$

so $x_n \to x^*$ at the geometric rate γ.

Proof sketch. The sequence $\{x_n\}$ is Cauchy: $d(x_{n+1}, x_n) \leq \gamma^n d(x_1, x_0)$, and the triangle inequality gives $d(x_{n+k}, x_n) \leq \gamma^n d(x_1, x_0)/(1-\gamma)$, which $\to 0$ as $n \to \infty$. By completeness, x_n converges to some $x^* \in X$. By continuity of T (which contractions enjoy automatically), $Tx_n \to Tx^*$, so $x^* = Tx^*$. Uniqueness: if $Ty^* = y^*$, then $d(x^*, y^*) = d(Tx^*, Ty^*) \leq \gamma d(x^*, y^*)$, forcing $d(x^*, y^*) = 0$. $\square$

The connection to the Bellman operator is direct.

Theorem 9 (Bellman operator is a contraction)**.** *Suppose the per-period reward r is bounded on $S \times A$. Then the Bellman operator T on the space $\mathcal{B}(S)$ of bounded measurable functions $V : S \to \mathbb{R}$, endowed with the sup norm $\|V\|_\infty = \sup_{s \in S} |V(s)|$, is a contraction with modulus β.*

Proof. For any $V_1, V_2 \in \mathcal{B}(S)$ and any s, let $a_1^* \in \arg\max_a [r(s,a) + \beta V_1(f(s,a))]$. Then $(TV_1)(s) - (TV_2)(s) \leq r(s, a_1^*) + \beta V_1(f(s, a_1^*)) - [r(s, a_1^*) + \beta V_2(f(s, a_1^*))] = \beta[V_1(f(s, a_1^*)) - V_2(f(s, a_1^*))] \leq \beta \|V_1 - V_2\|_\infty$. By symmetry, $|(TV_1)(s) - (TV_2)(s)| \leq \beta \|V_1 - V_2\|_\infty$ for every s, so $\|TV_1 - TV_2\|_\infty \leq \beta \|V_1 - V_2\|_\infty$. $\square$

Combining Theorems 8 and 9, the Bellman equation $V = TV$ has a unique solution V^* on $\mathcal{B}(S)$, and iteration from any bounded starting function V_0 converges to V^* geometrically at rate β. The slogan: dynamic programming works because the discount factor turns the Bellman operator into a contraction.[1]

The political-economy reading. The discount factor β is doing two pieces of work simultaneously. Substantively, it reflects the agent's patience or the institution's longevity — a higher β means future rewards matter more. Structurally, $\beta < 1$ is what makes the Bellman operator a contraction — a higher β means slower convergence of value-function iteration, but still convergence. When β is too close to 1 the iteration is slow but valid; when $\beta = 1$ (no discounting) the contraction property fails and additional structure (e.g., terminal conditions, undiscounted-average-reward formulations) is needed.

[1]Banach's fixed-point theorem is one of the most-used theorems in mathematical economics, with applications well beyond dynamic programming. The implicit function theorem from the differentiation handout has a standard proof via Banach (the IFT's local existence claim is the unique fixed point of a contraction defined from the linearization). Existence-and-uniqueness theorems for ordinary differential equations (Picard–Lindelöf) are Banach applications. Existence of equilibria in models with strategic complementarities can sometimes be proved by exhibiting a contraction on the strategy space. The structural reason Banach is so widely useful: many existence questions in mathematical economics can be recast as fixed-point problems, and the contraction-mapping route gives existence *and* uniqueness *and* a constructive computational method, all from one structural condition. The standard graduate references treating the theorem and its applications include Rudin (1976, Ch. 9) for the topology, Stokey, Lucas, and Prescott (1989, Ch. 3) for the DP application, Boyd and Vandenberghe (2004, Ch. 7) for the optimization-theoretic framing, and Aliprantis and Border (2006, Ch. 3) for the general functional-analytic setting.

6 The envelope condition

Comparative statics in dynamic problems — how does the value function shift when a state variable shifts? — is governed by an envelope condition that is the dynamic analogue of the static envelope theorem from the previous handout. The result is sometimes called the *Benveniste–Scheinkman theorem* after its originators, and it is the structural backbone of comparative statics in dynamic political-economy models. The political-economy substance is the same as in the static envelope: when computing how the value function responds to a state shift, the shift in the optimal action does not enter to first order. Only the parameter's direct effect on the one-period reward and the recursive continuation matters.

Theorem 10 (Envelope condition)**.** *Suppose $r(s,a)$ and $f(s,a)$ are continuously differentiable in s, and the optimal stationary policy $\pi^*(s)$ is interior in $A(s)$ and continuously differentiable in s on an open subset of S. Then on that subset,*

$$\frac{\partial V^*}{\partial s}(s) = \frac{\partial r}{\partial s}(s,\pi^*(s)) + \beta\,\frac{\partial V^*}{\partial s}(f(s,\pi^*(s)))\cdot\frac{\partial f}{\partial s}(s,\pi^*(s)).$$

*That is, the value function's partial derivative with respect to the state is the partial derivative of the one-period Bellman objective with respect to s, evaluated at $a=\pi^*(s)$ and treating π^* as fixed (*not *differentiating through it).*

Proof sketch. Differentiate the Bellman equation $V^*(s) = r(s,\pi^*(s)) + \beta V^*(f(s,\pi^*(s)))$ with respect to s via the chain rule. The terms $\partial r/\partial a\cdot\partial\pi^*/\partial s$ and $\beta(\partial V^*/\partial s)(f(s,\pi^*(s)))(\partial f/\partial a)(\partial\pi^*/\partial s)$ that arise from differentiating through π^* combine and vanish by the FOC of the inner maximization, $\partial r/\partial a + \beta(\partial V^*/\partial s)(f)(\partial f/\partial a) = 0$. What remains is the identity in the theorem. □

The reading: the substantive comparative-static effect of the state on the value function decomposes into a direct effect (the shift in today's reward) and an indirect effect (the shift in the next-period state, propagated through tomorrow's value function). The optimal-action adjustment cancels by the FOC, exactly as in the static envelope theorem. The dynamic analogue thus gives the same kind of comparative-static identities, with one extra layer of recursion.[2]

Example 11 (Comparative statics under campaign-effectiveness shifts)**.** A long-lived candidate's per-period vote-share gain depends on her current popularity (the state s) and her current effort (the action a): $r(s,a,\theta) = s\cdot g(a,\theta) - c(a)$, where θ is a campaign-effectiveness parameter. Popularity

[2]The dynamic envelope condition is the structural starting point for a substantial literature on *time consistency* in dynamic policy problems — the question of whether an agent who optimizes at $t=0$ continues to find the same $t=0$ plan optimal at later periods. The infinite-horizon Bellman equation built into this handout has stationary policies π^* that are *automatically* time-consistent (the maximization is the same at every period, so the policy doesn't change), but a substantial political-economy literature studies environments where time consistency *fails*. Strotz (1955) originally identified the problem in the consumption-and-saving setting with non-geometric (e.g., quasi-hyperbolic) discounting: an agent's $t=0$ plan for $t=5$ consumption may be revised when $t=5$ arrives, because the discounting shape changes how future rewards are weighted from the new vantage point. Kydland and Prescott (1977) extended the analysis to optimal monetary and fiscal policy: a government's announced inflation rate or tax schedule may be unsustainable once the announcement period passes, leading to a commitment problem the modern central-banking-and-fiscal literature addresses through institutional design (independent central banks, fiscal rules). Dynamic-inconsistency questions of this kind are central to the political-economy of policy commitment and to the analysis of authoritarian-regime promises, election-cycle policy, and any setting where the principal cannot bind herself credibly to a future course of action.

evolves as $s_{t+1} = \rho s_t + a_t$, decaying at rate $\rho \in (0, 1)$ and reinforced by current effort. The infinite-horizon Bellman equation gives a value function $V^*(s, \theta)$. The envelope condition gives

$$\frac{\partial V^*}{\partial \theta}(s) = s \cdot \frac{\partial g}{\partial \theta}(\pi^*(s, \theta), \theta) + \beta \frac{\partial V^*}{\partial \theta}(\rho s + \pi^*(s, \theta)),$$

without differentiating through π^*. The political-economy reading: an increase in campaign effectiveness raises the candidate's value, with the per-period direct effect $s \cdot \partial g / \partial \theta$ propagated forward through the state-dynamics chain. The result holds without re-solving the optimization at the new θ, by the standard envelope argument.

7 What's next

This handout closes the optimization cluster (#20–#23). Three strands extend it.

Game theory is the immediate next applied cluster. The static-optimization handout's Berge theorem is what guarantees best-response correspondences are upper-hemicontinuous and non-empty-valued; the contraction-mapping theorem of this handout is one of the standard routes to existence-and-uniqueness of equilibrium in supermodular and other special settings. The envelope theorem (static and dynamic) supplies the comparative-statics machinery for game-theoretic analysis. The dynamic generalization of game theory — repeated games, dynamic games of incomplete information, Markov-perfect equilibria — uses the dynamic-programming apparatus of this handout directly.

Markov chains as a topic in their own right: many of the dynamic problems considered here have stochastic state transitions, in which case the transition function f is replaced by a Markov-chain transition kernel and the contraction-mapping argument extends with minimal modification. The Perron–Frobenius theorem of the eigenvalue handout is the structural input on the kernel side; the dynamic-programming machinery here is the choice-theoretic complement.

Optimal control and continuous-time methods, beyond this cluster's discrete-time scope, are the natural successor framework for problems where time is treated as a continuum. Hamilton–Jacobi–Bellman equations are the continuous-time analogues of the Bellman equation; Pontryagin's maximum principle is the analogue of the FOC-and-multiplier framework of the static handout. Macroeconomic and growth-theory applications use these heavily; the political-economy applications are scattered but real.

For graduate-level treatments at this handout's level of abstraction, Stokey, Lucas, and Prescott (1989) is the canonical economic-dynamics reference and treats both finite- and infinite-horizon discounted DP with care. Bertsekas (2017) is the standard engineering / operations-research treatment, with the algorithmic and computational sides better developed. Sundaram (1996, Ch. 12) covers the basic dynamic-programming theorems with the political-economy-flavored presentation this handout has aimed at.

8 Exercises

Exercise 12. *Backward induction by hand.* A two-period problem has state $s \in \{0, 1\}$, action $a \in \{0, 1\}$, deterministic transition $s_{t+1} = a_t$, per-period reward $r(s, a) = s + 2a$, terminal payoff $r_T(s) = 3s$, discount $\beta = 0.5$. Compute V_2, V_1, V_0 by backward induction and identify the optimal action at each (s, t).

Exercise 13. *Banach fixed-point on $\mathbb{R}$.* The function $T : \mathbb{R} \to \mathbb{R}$ given by $T(x) = \frac{1}{2}x + 1$ is a contraction on $(\mathbb{R}, |\cdot|)$ with modulus 1/2. (a) Find its fixed point. (b) Verify the geometric-convergence bound on the iterates from $x_0 = 0$.

Exercise 14. *Value-function iteration.* Consider the infinite-horizon problem on $S = \{1, 2, 3\}$ with two actions $A = \{a, b\}$, discount $\beta = 0.5$, and rewards / transitions specified by a small table (write your own simple tabular example, or use a textbook one). Starting from $V_0 \equiv 0$, perform three iterations of $V_{n+1} = TV_n$ and observe the convergence.

Exercise 15. Verify the envelope condition (Theorem 10) directly on a simple example. Let $r(s, a) = 2as - a^2$, $f(s, a) = s + a$, $\beta \in (0, 1)$. (a) Solve the Bellman equation by guessing $V^*(s) = \alpha s^2 + \gamma$ and matching coefficients. (b) Compute $\partial V^*/\partial s$ directly from the closed form. (c) Verify it matches the envelope-condition expression.

Exercise 16. *Voter learning under Bayesian updating.* Continuing Example 5: model the voter's belief $s_t \in [0, 1]$ as her posterior probability that the incumbent is competent, updated each period by Bayes' rule given a binary signal of competence. Write down the per-period reward $r(s, a)$ for action $a \in \{\text{vote-incumbent}, \text{vote-challenger}\}$ and the transition $s_{t+1} = f(s, a, \text{signal})$ (stochastic). State the Bellman equation for the voter's optimal-voting problem. Discuss in one or two sentences why the Bellman framework is the right setting for this problem (rather than period-by-period myopic decisionmaking).

Exercise 17. *Infinite-horizon legislative bargaining.* A legislator chooses how much political capital $a \in [0, c]$ to spend each period, with capital evolving as $c_{t+1} = c_t - a_t + \delta$ (replenishment $\delta > 0$). Per-period payoff is $r(c, a) = u(a)$ for some increasing concave u, with discount $\beta \in (0, 1)$. (a) State the Bellman equation. (b) Argue that the optimal policy is stationary: $\pi^*(c)$ depends only on the current state c, not on t. (c) Identify (heuristically, without solving) how the optimal spending $\pi^*(c)$ should depend on the discount factor β and the replenishment rate δ.

Exercise 18. *Policy iteration vs. value iteration.* An alternative to the value-function iteration $V_{n+1} = TV_n$ is *policy iteration*: starting from any policy π_0, compute its value V_{π_0} (by solving the linear system $V_{\pi_0}(s) = r(s, \pi_0(s)) + \beta V_{\pi_0}(f(s, \pi_0(s)))$), then update π_1 to be the maximizer of $r(s, a) + \beta V_{\pi_0}(f(s, a))$, and iterate. Argue informally that policy iteration converges in finitely many iterations on a finite state space. (Hint: the policy improves monotonically and there are only finitely many policies.)

Exercise 19. *Time consistency under geometric discounting.* Show that if the agent's discounting is geometric (β^t for some constant $\beta \in [0, 1)$), her infinite-horizon optimum is time-consistent: at any later period t, her remaining-horizon optimal plan coincides with what was originally planned at $t = 0$. (Hint: the Bellman equation is stationary in t under geometric discounting, so the policy is too.)

Exercise 20. The stochastic generalization of the Bellman equation replaces the deterministic transition $s_{t+1} = f(s_t, a_t)$ with a transition kernel $P(s'|s, a)$ (a probability distribution over next-period states). Write down the corresponding Bellman equation. Argue that the contraction-mapping theorem applies (with the same modulus β) under the same boundedness conditions.

Exercise 21. *Authoritarian-regime survival.* A regime chooses each period a level of repression $a \in [0, 1]$ to suppress mobilization. The regime's survival probability next period is $p(s, a) = s(1 - q(a))$, where s is current legitimacy stock and q is increasing in a. Per-period payoff (conditional on survival) is $r(s, a) = s - \kappa a$. (a) State the (stochastic) Bellman equation for the regime's value function. (b) Discuss in two or three sentences how the discount factor β shapes the regime's repression intensity, and how that interpretation depends on whether β captures the regime's patience or its expected longevity.

References

Aliprantis, Charalambos D. and Kim C. Border (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide*. 3rd ed. Berlin: Springer.

Bertsekas, Dimitri P. (2017). *Dynamic Programming and Optimal Control*. 4th ed. Belmont, MA: Athena Scientific.

Boyd, Stephen and Lieven Vandenberghe (2004). *Convex Optimization*. Cambridge: Cambridge University Press.

Kydland, Finn E. and Edward C. Prescott (1977). "Rules Rather than Discretion: The Inconsistency of Optimal Plans". In: *Journal of Political Economy* 85.3, pp. 473–491.

Rudin, Walter (1976). *Principles of Mathematical Analysis*. 3rd ed. New York: McGraw-Hill.

Stokey, Nancy L., Robert E. Lucas, and Edward C. Prescott (1989). *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.

Strotz, R. H. (1955). "Myopia and Inconsistency in Dynamic Utility Maximization". In: *Review of Economic Studies* 23.3, pp. 165–180.

Sundaram, Rangarajan K. (1996). *A First Course in Optimization Theory*. Cambridge: Cambridge University Press.