0.1 Optimal Control in Discrete Time

0.1.1 Discrete Time Markov Processes

Let $\{x_n\}$ be a discrete-time stochastic process, that is, a sequence of random variables, with values in a space X and $n = 0, 1, \ldots$. We assume that X is a Borel space.

The process $\{x_n\}$ is called a Markov chain or a discrete-time Markov process if, for every $k \geq 0$ and $B \in \mathcal{B}(X)$,

$$P(x_{k+1} \in B \mid x_0, \ldots, x_k) = P(x_{k+1} \in B \mid x_k). \tag{1}$$

The interpretation of (1) is that k is the present (or current) period, $k+1$ is in the future, and $n \leq k-1$ constitutes the past; then (1) states that the distribution of x_{k+1} given the history of the process up to the current time k depends only on the current state x_k.

The right-hand side of (1) defines the one-step transition probabilities

$$P_k(x, B) := P(x_{k+1} \in B \mid x_k = x)$$

for all $x \in X$, $B \in \mathcal{B}(X)$, and $k = 0, 1, \ldots$. We interpret them as the probability of the state belonging to the set B at time $k+1$ given that the value of the state at time k is x, i.e., the probability of transitioning from x to B. If the transition probabilities are independent of k, that is,

$$P(x, B) := P(x_{k+1} \in B \mid x_k = x) \quad \forall k \geq 0$$

then $\{x_n\}$ is said to be a stationary or time-homogeneous Markov chain. When the state space X is finite (a finite number of states), we can define a Markov process without the machinery of Borel spaces. When X is finite, x_n is a Markov process if, for every $k \geq 0$ and $y \in X$,

$$P(x_{k+1} = y \mid x_0, \ldots, x_k) = P(x_{k+1} = y \mid x_k).$$

which is a restatement of equation (1) for the case with finite X. The corresponding one-step

transition probabilities from state x at time k to state y at $k+1$ are

$$P_k(x,y) := P(x_{k+1} = y \mid x_k = x)$$

or, if the process is stationary,

$$P(x,y) := P(x_{k+1} = y \mid x_k = x) \quad \forall k \geq 0$$

A useful characterization of discrete-time Markov processes is:

Proposition 0.1.1. *Let $\{x_n\}$ and $\{\xi_n\}$ be discrete-time stochastic processes in Borel spaces X and S, respectively. Suppose that $\xi_0, \xi_1, \ldots$ are independent, and also independent of x_0. If there is a measurable function $F : X \times S \to X$ such that*

$$x_{n+1} = F(x_n, \xi_n) \quad \forall n \geq 0, \tag{2}$$

then $\{x_n\}$ is a Markov process. The converse is also true.

Example 0.1.0.1. Let $\{\xi_n\}$ be a sequence of independent random variables, and x_0 a given random variable independent of $\{\xi_n\}$. By Proposition 0.1.1, the following are examples of Markov chains:

(a) A Markov chain that evolves as

$$x_{n+1} = F(x_n) + \xi_n \quad \forall n = 0, 1, \ldots$$

is called a first order autoregressive process, and $\{\xi_n\}$ is said to be an additive noise.

(b) The inventory-production system

$$x_{n+1} = x_n + f(x_n) - \xi_n, \quad n = 0, 1, \ldots,$$

where x_n is the inventory level of a product at time n, $f(x)$ is the production strategy given the stock level x, and ξ_n is the demand of the product in period n.

□

0.1.2 Discrete Time Controlled Markov Processes

When solving discrete-time optimal control problems, there are two formulations. The first one is a "system model" formulation in which the controlled system evolves according to a difference equation of the form

$$x_{t+1} = F(x_t, a_t, \xi_t) \quad \forall t = 0, 1, \ldots, T-1, \tag{3}$$

for $T \leq \infty$, with a given — possibly random — initial condition x_0. Here, the state and control variables x_t, a_t take values in a state space X and an action set A, respectively, both assumed to be Borel spaces. Moreover, the ξ_t are independent random variables with values in a Borel space S, and they denote random perturbations. The idea of (3) is that the evolution of the state process $\{x_t\}$ can be influenced by the sequence of actions $\{a_t\}$.

The second formulation is a "control model" formulation in which the transition probabilities are given by

$$Q(B \mid x, a) := \text{Prob}\left[x_{t+1} \in B \mid x_t = x, a_t = a\right]. \tag{4}$$

where $B \in \mathcal{B}(X)$ and, as in the system model, $x_t \in X$, and $a_t \in A$ for all t. The idea behind (3) is that the sequence of actions $\{a_t\}$ can influence the likelihood of future states, conditional on the current state. Thus, in the control model formulation, the actions influence the evolution of the state process only indirectly through probabilities.

For a large class of models, both formulations are equivalent, so we usually work with whichever one is more convenient.

Example 0.1.0.2. Consider the system model (3), and suppose that the ξ_t are independent and identically distributed (i.i.d.) random variables with a common distribution μ on S. We can go from the system model (3) to the control model (4). From (4), we can see that Q is given by

$$\begin{aligned} Q(B \mid x, a) &= \text{Prob}[F(x, a, \xi) \in B] \\ &= \mu(\{s \in S : F(x, a, s) \in B\}) \\ &= E[I_B(F(x, a, \xi))] \end{aligned}$$

where I_B is the indicator function of the set B, that is,

$$I_B(x) := \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{otherwise} \end{cases}$$

It is also possible to prove that we can go the other way around, from (4) to (3), but the proof is nonconstructive, so not very helpful in practical terms. □

0.1.3 Setup of Control Problem

We now describe a general stochastic discrete-time control problem. Our notation defers the issue of choosing between the system model and the control model until we pick particular functional forms, admissible control sets, etc. Thus, we don't need to make the distinction between the two formulations at this stage. Later on, when we work with specific concrete problems, we will inevitably write each problem in either system model or control model forms.

Let $(\Omega, \mathcal{F}, P, \mathbf{F})$ be a filtered probability space, where $\mathbf{F} = \{\mathcal{F}_n\}_{n=0}^{\infty}$. On this space we consider a discrete-time controlled Markov process X living on the state space $\mathbf{X}$, with controls u in some control space $\mathcal{U}$. Time is indexed by non-negative integers $\mathcal{T} = \{0, 1, 2, \dots T\}$, and we use the notation $[n, m]$ to denote a discrete time interval, so $[n, m] = \{n, n+1, \dots, m-1, m\}$, where $n < m$. We also have some exogenously given objects:

- A reward functional of the form
$$E\left[\sum_{n=0}^{T-1} H_n(X_n, u_n) + F(X_T)\right]$$
- An indexed family $\{U_n(x) : x \in \mathbf{X}, n \in \mathcal{T}\}$ of subsets of $\mathcal{U}$, so $U_n(x) \subseteq \mathcal{U}$ for all $x \in \mathbf{X}$ and all $n = 0, 1, 2, \dots, T$. This family provides us with control restrictions in the sense that if $X_n = x$ then we must choose the control u_n such that $u_n \in U_n(x)$.

The problem to be solved is to choose an adapted control process u that maximizes

$$E\left[\sum_{n=0}^{T-1} H_n(X_n, u_n) + F(X_T)\right]$$

subject to the constraints

$$u_n \in U_n(X_n), \quad n = 0, 1, 2, \ldots$$

We can accommodate infinite horizon problems by letting $T = \infty$ and $F(\cdot) \equiv 0$. In principle the control process u is allowed to be any adapted process satisfying the constraints above, but we will restrict ourselves to the case of feedback control laws.

Definition 0.1.1 (A feedback control law)**.** A feedback control law is a mapping $\mathbf{u} : \mathcal{T} \times \mathbf{X} \to \mathcal{U}$.

The interpretation is that, given the control law $\mathbf{u}$, the control process u will be of the form

$$u_n = \mathbf{u}_n(X_n).$$

The class of feedback control laws is of course smaller than the class of adapted controls. However, it is possible to prove that the optimal control is always realized by a feedback law, so from an optimality point of view there is no restriction to limiting ourselves to feedback laws. On the other hand, the economic implications of following a feedback control law or an "open loop" control law (i.e., a control law that depends on time and exogenous disturbances but not on the state) can be quite important *for equilibrium determination*. For example, in standard New Keynesian models, if the central bank sets interest rates according to an open loop control law, then there always is indeterminacy (multiple equilibria). In contrast, with a feedback control law that depends on inflation, it is possible for the central bank to guarantee determinacy (a unique equilibrium). In either case, however, the optimal path that solves the control problem is the same.

The class $\mathbf{U}$ of admissible feedback laws is defined as the class of feedback laws $\mathbf{u}$ satisfying the constraints

$$u_n \in U_n(X_n), \quad n = 0, 1, 2, \ldots$$

0.1.4 The Bellman Equation

The Value Function

The way to approach our optimization problem is construct a family of problems indexed by time and initial conditions. Then we can connect all these problems by a recursive equation, the Bellman equation. Solving the Bellman equation is equivalent to solving the optimal control problem.

Definition 0.1.2. For each fixed initial point (n, x) we define the problem $\mathcal{P}_{n,x}$ as the problem of maximizing

$$E_{n,x}\left[\sum_{k=n}^{T-1} H_k(X_k, \mathbf{u}_k(X_k)) + F(X_T)\right]$$

over the class of feedback laws $\mathbf{u}$ satisfying the constraints

$$\mathbf{u}_k(x) \in U_k(x), \quad \text{for all } k \geq n, \quad x \in \mathbf{X}$$

We will need the following:

Definition 0.1.3. The value function

$$J : \mathcal{T} \times \mathbf{X} \times \mathbf{U} \to \mathbb{R}$$

is defined by

$$J_n(x, \mathbf{u}) = E_{n,x}\left[\sum_{k=n}^{T-1} H_k(X_k, \mathbf{u}_k(X_k)) + F(X_T)\right]$$

The optimal value function

$$V : \mathcal{T} \times \mathbf{X} \to \mathbb{R}$$

is defined by

$$V_n(x) = \sup_{\mathbf{u} \in \mathbf{U}} J_n(x, \mathbf{u})$$

The interpretation is that $J_n(x, \mathbf{u})$ yields the expected utility of employing the control law $\mathbf{u}$ for the time interval $[n, T]$ if you start in state x at time n. The optimal value function $V_n(x)$

gives you the optimal utility over $[n, T]$ if you start in state x at time n.

Remark 0.1.1. It is common to use the term value function for both the value function and the *optimal* value function, and we will sometimes follow that practice when the context does not allow for confusion.

Time Consistency and the Bellman Principle

We assume that for every initial point (n, x) there exists an optimal control law for problem $\mathcal{P}_{n,x}$. This control law is denoted by $\hat{\mathbf{u}}^{n,x}$.

The object $\hat{\mathbf{u}}^{n,x}$ is a mapping $\hat{\mathbf{u}}^{n,x} : [n, T] \times \mathbf{X} \to \mathbb{R}$, where the upper index (n, x) denotes the fixed initial point for problem $\mathcal{P}_{n,x}$. Consequently, the control applied at some time $k \geq n$ will be given by

$$\hat{\mathbf{u}}_k^{n,x}(X_k).$$

The optimal law for the problem $\mathcal{P}_{n,x}$ could very well depend on the choice of the starting point (n, x). However, it turns out that the optimal law is independent of this choice:

Theorem 0.1.1 (The Bellman Optimality Principle)**.** *Fix an initial point (n, x) and consider the corresponding optimal law $\hat{\mathbf{u}}^{n,x}$. Then the law $\hat{\mathbf{u}}^{n,x}$ is also optimal for any subinterval of the form $[m, T]$ where $m \geq n$. In other words,*

$$\hat{\mathbf{u}}_k^{n,x}(y) = \hat{\mathbf{u}}_k^{m,X_m}(y)$$

for all $k \geq m$ and all $y \in \mathbf{X}$. In particular, the optimal law for the initial point $n = 0$ will be optimal for all subintervals. This law will be denoted by $\hat{\mathbf{u}}$.

The Bellman optimality principle says that a plan for the future deemed optimal at an earlier point in time will also remain optimal. Suppose that you optimize at time $n = 0$ and follow control law $\hat{\mathbf{u}}$ up to time n, where you now have reached the state X_n. At time n you reconsider, and now decide to forget your original problem and instead solve problem $\mathcal{P}_{n,X_n}$. What the Bellman Principle tells you is that the law $\hat{\mathbf{u}}$ (restricted to the time interval $[n, T]$) is optimal, not only for your original problem, but also for your new problem. We could say that our family of problems is time consistent.

As a consequence of the Bellman Principle, we can obtain the Bellman equation, which is the recursive relation for the optimal value function.

Theorem 0.1.2 (The Bellman Equation)**.** *The optimal value function satisfies the recursive equation*

$$V_n(x) = \sup_{u \in U_n(x)} \{H_n(x,u) + E_{n,x}\left[V_{n+1}(X_{n+1}^u)\right]\},$$
$$V_T(x) = F(x).$$

Furthermore, the supremum in the equation is realized by the optimal control law $\hat{\mathbf{u}}_n(x)$.

Optional

*0.1.5 Proofs of Theorems 0.1.1 and 0.1.2

Proof of Theorem 0.1.1. The proof is by contradiction. Let us assume that for some $n > 0$ there exists a law $\overline{\mathbf{u}}$ on the interval $[n, T]$ such that

$$E_{n,x}\left[\sum_{k=n}^{T-1} H_k(X_k, \overline{\mathbf{u}}_k(X_k)) + F(X_T)\right] \geq E_{n,x}\left[\sum_{k=n}^{T-1} H_k(X_k, \hat{\mathbf{u}}_k(X_k)) + F(X_T)\right]$$

for all $x \in \mathbf{X}$ with strict inequality for some $x \in \mathbf{X}$. We can then construct a new law $\mathbf{u}^\star$ on $[0, T]$ by the following formula

$$\mathbf{u}_k^\star(y) = \begin{cases} \hat{\mathbf{u}}_k(y) \text{ for } 0 \leq k < n-1 \\ \overline{\mathbf{u}}_k(y) \text{ for } n \leq k < T-1 \end{cases}$$

We then have

$$\begin{aligned} J_0(x_0, \mathbf{u}^\star) =& E_{0,x_0}\left[\sum_{k=0}^{T-1} H_k(X_k, \mathbf{u}_k^\star) + F(X_T)\right] \\ =& E_{0,x_0}\left[\sum_{k=0}^{n-1} H_k(X_k, \hat{\mathbf{u}}_k)\right] + E_{0,x_0}\left[\sum_{k=n}^{T-1} H_k(X_k, \overline{\mathbf{u}}_k) + F(X_T)\right] \\ =& E_{0,x_0}\left[\sum_{k=0}^{n-1} H_k(X_k, \hat{\mathbf{u}}_k)\right] + E_{0,x_0}\left[E_{n,X_n}\left[\sum_{k=n}^{T-1} H_k(X_k, \overline{\mathbf{u}}_k) + F(X_T)\right]\right], \end{aligned}$$

where we have used the law of iterated expectations and the Markov property to obtain the last term. It now follows from the assumption concerning $\overline{\mathbf{u}}$ that we have

$$E_{n,X_n}\left[\sum_{k=n}^{T-1} H_k(X_k, \overline{\mathbf{u}}_k) + F(X_T)\right] \geq E_{n,X_n}\left[\sum_{k=n}^{T-1} H_k(X_k, \hat{\mathbf{u}}_k) + F(X_T)\right].$$

with strict inequality with positive probability so, again using iterated expectations and the Markov property, we obtain

$$\begin{aligned} J_0(x_0, \mathbf{u}^\star) &> E_{0,x_0}\left[\sum_{k=0}^{n-1} H_k(X_k, \hat{\mathbf{u}}_k)\right] + E_{0,x_0}\left[E_{n,X_n}\left[\sum_{k=n}^{T-1} H_k(X_k, \hat{\mathbf{u}}_k)\right] + F(X_T)\right] \\ &= E_{0,x_0}\left[\sum_{k=0}^{n-1} H_k(X_k, \hat{\mathbf{u}}_k)\right] + E_{0,x_0}\left[\sum_{k=n}^{T} H_k(X_k, \hat{\mathbf{u}}_k)\right] \\ &= E_{0,x_0}\left[\sum_{k=0}^{T} H_k(X_k, \hat{\mathbf{u}}_k) + F(X_T)\right] = J_0(x_0, \hat{\mathbf{u}}). \end{aligned}$$

We have thus obtained the inequality

$$J_0(x_0, \mathbf{u}^\star) > J_0(x_0, \hat{\mathbf{u}}),$$

which contradicts the optimality of $\hat{\mathbf{u}}$ on the interval $[0, T]$. □

Proof of Theorem 0.1.2. We fix an arbitrary initial point (n, x) and consider the control law $\mathbf{u}^\star$ that deviates from the optimal control only at time n. That is, at time n we use an arbitrary control value $u \in U_n(x)$, and from time $n+1$ onwards we use the optimal control $\hat{\mathbf{u}}$. Formally, the law $\mathbf{u}^\star$ on $[n, T]$ is defined by

$$\begin{aligned} &\mathbf{u}^\star_n(x) = u, \\ &\mathbf{u}^\star_k(y) = \hat{\mathbf{u}}_k(y), \qquad \text{for all } k \in [n+1, T] \text{ and for all } y \in \mathbf{X}. \end{aligned}$$

The idea now is, given the initial point (n, x), to compare the following two control laws on $[n, T]$:

1. The optimal law $\hat{\mathbf{u}}$.
2. The law $\mathbf{u}^\star$ defined above.

In order to do this, we will compute the expected utilities generated by the two laws.

Using the fact that the utility from $\hat{\mathbf{u}}$ must be greater than or equal to the utility from $\mathbf{u}^\star$ will allow us to obtain our recursive relation.

1. Since $\hat{\mathbf{u}}$ is the optimal law we know that the expected utility for $\hat{\mathbf{u}}$ is given by

$$J_n(x, \hat{\mathbf{u}}) = V_n(x).$$

2. From the definition of $\mathbf{u}^\star$ we also know that the expected utility for $\mathbf{u}^\star$ is

$$\begin{aligned} J_n(x, \mathbf{u}^\star) &= E_{n,x}\left[\sum_{k=n}^{T-1} H_k(X_k^{\mathbf{u}^\star}, \mathbf{u}_k^\star(X_k)) + F(X_T)\right] \\ &= H_n(x,u) + E_{n,x}\left[\sum_{k=n+1}^{T-1} H_k(X_k^{\mathbf{u}^\star}, \mathbf{u}_k^\star(X_k)) + F(X_T)\right], \end{aligned}$$

where the notation $X_k^{\mathbf{u}^\star}$ emphasizes the dependence of the distribution of the X-process on the control law $\mathbf{u}^\star$. Using the law of iterated expectations and the Markov property gives

$$\begin{aligned} &E_{n,x}\left[\sum_{k=n+1}^{T-1} H_k(X_k^{\mathbf{u}^\star}, \mathbf{u}_k^\star(X_k)) + F(X_T)\right] \\ &= E_{n,x}\left[E_{n+1,X_{n+1}^u}\left[\sum_{k=n+1}^{T-1} H_k(X_k^{\mathbf{u}^\star}, \mathbf{u}_k^\star(X_k)) + F(X_T)\right]\right], \end{aligned}$$

where we note that the distribution of X_{n+1} depends only on the chosen control value u at time n, which is captured by the notation X_{n+1}^u. Using the fact that $\mathbf{u}^\star = \hat{\mathbf{u}}$ on the time interval $[n+1, T]$ we obtain

$$E_{n+1,X_{n+1}^u}\left[\sum_{k=n+1}^{T} H_k(X_k^{\mathbf{u}^\star}, \mathbf{u}_k^\star(X_k)) + F(X_T)\right] = V_{n+1}(X_{n+1}^u).$$

This gives us the following result for the expected utility from the "deviation" law $\mathbf{u}^\star$:

$$J_n(x, \mathbf{u}^\star) = H_n(x,u) + E_{n,x}\left[V_{n+1}(X_{n+1}^u)\right].$$

Finally, when comparing the two control laws $\hat{\mathbf{u}}$ and $\mathbf{u}^\star$, a clear ranking of the expected

utilities emerges,

$$J_n(x, \hat{\mathbf{u}}) \geq J_n(x, \mathbf{u}^\star), \qquad \text{for all } u \in U_n(x)$$

with equality when $u = \hat{\mathbf{u}}_n(x)$. Plugging in the results from above gives us the following inequality:

$$V_n(x) \geq H_n(x, u) + E_{n,x}\left[V_{n+1}(X_{n+1}^u)\right]$$

This relation holds for all $u \in U_n(x)$ (recall that u was chosen arbitrarily), with equality for $u = \hat{\mathbf{u}}_n(x)$, which completes the proof. □

0.1.6 On State Variables

We do not give a completely formal definition of a state variable, but for our purposes the following imprecise heuristic definition will be enough.

Definition 0.1.4. Consider a process X (deterministic or stochastic). A state variable Z is a process such that knowledge of Z_n at time n is sufficient to calculate the probability distribution of X_m at any time $m \geq n$.

The concept of a state variable is thus very close to the concept of a sufficient statistic and to the Markov property. We usually encounter the following:

- X is a Markov process itself, which implies that X is a state variable.
- X is not Markovian, but if we enlarge the state space by adding a process Y such that the process $Z = (X, Y)$ is Markov, then Z acts a state variable.

Given a fixed feedback law $\mathbf{u}$, the process $X^{\mathbf{u}}$ is a state variable. Given knowledge of $X_n^{\mathbf{u}}$ we can, by the Markov property of $X^{\mathbf{u}}$, calculate the distribution of $X_m^{\mathbf{u}}$ for any $m \geq n$.

0.1.7 Deterministic Infinite Horizon Problems in Economics[1]

We now consider the type of deterministic infinite-horizon problem usually encountered in economics and show how to use the Bellman equation to solve it step by step. This problem fits the framework described above, but the notation is somewhat different to more closely align with how economists write these kinds of problems.

Given initial condition a_0, choose $\{c_t\}_{t=0}^{\infty}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t u(a_t, c_t)$$

subject to the dynamic system

$$a_{t+1} = g(a_t, c_t). \tag{5}$$

where $u(a_t, c_t)$ is a twice-differentiable concave period utility function. In economics, it is common to refer to equation (5) as the law of motion for a. Also note that here a_t stands for "asset"; the variable a_t is the state variable and *not* the control variable. The control variable is c_t, which stands for "consumption".

We seek to find a policy function h that maps the state a_t into the control c_t such that the sequence $\{c_t\}_{t=0}^{\infty}$ generated by iterating the two functions

$$c_t = h(a_t)$$

$$a_{t+1} = g(a_t, c_t)$$

solves the original problem.

Step 1: Write the Bellman Equation

The Bellman equation for this problem is:

$$\tilde{V}_t(a) = \max_c \left\{ \beta^t u(a, c) + \tilde{V}_{t+1}(g(a, c)) \right\}. \tag{6}$$

where $\tilde{V}_t$ is the optimal value function. Since the utility function (the stage cost) depends

1. The content of this section is an adaptation of notes created by Pascal Michaillat that are licensed under the Creative Commons Attribution 4.0 International License. The original notes can be found at https://github.com/pmichaillat/math-for-macro. All errors are mine.

on time only through the β^t term, we can simplify the Bellman equation by guessing that the value function is of the form

$$\tilde{V}_t(a) = \beta^t V(a)$$

with $V(\,\cdot\,)$ a time-invariant function. Using this guess in equation (6) gives

$$V(a) = \max_c \{u(a,c) + \beta V(g(a,c))\}. \tag{7}$$

The optimal consumption is given by the policy function $c = h(a)$. Another representation of the Bellman equation is

$$V(a) = u(a, h(a)) + \beta V(g(a, h(a))). \tag{8}$$

To highlight the recursive structure of the problem, we can write the symbolic representation of the Bellman equation as:

$$V(\text{state(t)}) = \max_{\text{control(t)}} \{u(\text{control(t)}, \text{state(t)}) + \beta V(\text{state(t+1)})\}$$

subject to

$$\text{state(t+1)} = g(\text{control(t)}, \text{state(t)}),$$

which is equivalent to

$$V(\text{state(t)}) = \max_{\text{control(t)}} \{u(\text{control(t)}, \text{state(t)}) + \beta V(g(\text{control(t)}, \text{state(t)}))\},$$

where control(t) and state(t) are vectors of control variables and state variables.

Step 2: Derive the First-Order Condition

Taking the first-order condition (FOC) with respect to c in the optimization problem (7) yields

$$\frac{\partial u}{\partial c}(a,c) + \beta \frac{\partial g}{\partial c}(a,c) V'(g(a,c)) = 0, \tag{9}$$

where

$$\frac{\partial u}{\partial c}(a,c)$$

is the partial derivative of the function $u(a, c)$ with respect to its second argument, evaluated at the pair (a, c);

$$\frac{\partial g}{\partial c}(a, c)$$

is the partial derivative of the function $g(a, c)$ with respect to its second argument, evaluated at the pair (a, c); and

$$V'(g(a, c))$$

is the derivative of the function $V(a)$ (with respect to its only argument), evaluated at $g(a, c)$.

Step 3: Benveniste-Scheinkman Equation

In the FOC (9), we do not know the derivative V' of the value function (because we do not yet know the value function). Hence, the next step is to determine what the derivative V' of the value function is.

To do so, we apply the Benveniste-Scheinkman theorem. This theorem says that under some regularity conditions,

$$V'(a) = \frac{\partial u}{\partial a}(a, c) + \beta \frac{\partial g}{\partial a}(a, c) V'(g(a, c)). \tag{10}$$

The theorem is a version of the envelope theorem applied to the Bellman equation (7). Roughly speaking, the Benveniste-Scheinkman theorem says we can take the derivative $V'(a)$ by taking the derivative of the right-hand side of equation (7) as if c did not depend on a.

Optional

Intuition for the Envelop Theorem

Consider the following simple maximization problem. There are two choice variables x and y, and one parameter, α. The problem is to maximize

$$U = f(x, y, \alpha).$$

The first order necessary conditions are

$$f_x(x, y, \alpha) = f_y(x, y, \alpha) = 0.$$

If second-order conditions are met, these two equations implicitly define the solution

$$x = x^*(\alpha) \quad y = x^*(\alpha).$$

If we substitute these solutions into the objective function, we get a new function

$$V(\alpha) = f(x^*(\alpha), y^*(\alpha), \alpha), \tag{11}$$

where this new function is the value of f when the values of x and y are those that maximize $f(x, y, \alpha)$. Therefore, $V(\alpha)$ is the optimal value function. If we differentiate V with respect to α, we get

$$\frac{\partial V}{\partial \alpha} = f_x \frac{\partial x^*}{\partial \alpha} + f_y \frac{\partial y^*}{\partial \alpha} + f_\alpha. \tag{12}$$

However, from the first order conditions we know $f_x = f_y = 0$. Therefore, the first two terms disappear and the result becomes

$$\frac{\partial V}{\partial \alpha} = f_\alpha. \tag{13}$$

The last equation says that, the result of varying α when x^* and y^* allowed to adjust optimally is the same as if x^* and y^* were held constant! Note that α enters the value function in equation (11) in three places: one direct and two indirect (through x^* and y^*). Equations (12) and (13) show that, at the optimum, only the direct effect of α on the objective function survives. This is the essence of the envelope theorem. The envelope theorem says that, at the optimum, only the direct effect of a change in a parameter or an exogenous variable need be considered, even though the parameter or exogenous variable may enter the maximum value function indirectly as part of the solution to the endogenous choice variables.

Step 3': Combine the FOC with the Benveniste-Scheinkman equation

Use the the first-order condition (9) and the Benveniste-Scheinkman equation (10) to

eliminate $V'(a')$ (e.g., solve (9) for $V'(a')$ and plug it into (10)) to get:

$$V'(a) = \frac{\partial u}{\partial a}(a,c) - \frac{\frac{\partial g}{\partial a}(a,c)\frac{\partial u}{\partial c}(a,c)}{\frac{\partial g}{\partial c}(a,c)} \tag{14}$$

Equation (14) gives V' as a function of the known functions u and g.

This step is necessary when

$$\frac{\partial g(a,c)}{\partial a} \neq 0$$

and can be bypassed otherwise.

Step 4: One Step Forward

Equation (14) is true for all time periods. In particular, it is true for the next period. Therefore, using the notation $a' := g(a, h(a))$

$$V'(a') = \frac{\partial u}{\partial a}(a',c') - \frac{\frac{\partial g}{\partial a}(a',c')\frac{\partial u}{\partial c}(a',c')}{\frac{\partial g}{\partial c}(a',c')}. \tag{15}$$

Step 5: Euler Equation

We plug equation (15) into equation (9) to get the Euler equation

$$\frac{\partial u}{\partial c}(a,c) + \beta\frac{\partial g}{\partial c}(a,c)\left\{\frac{\partial u}{\partial a}(a',c') - \frac{\partial u}{\partial c}(a',c')\frac{\partial g(a',c')/\partial a}{\partial g(a',c')/\partial c}\right\} = 0.$$

Using that $a' = g(a,c)$, we can see that the Euler equation is an equation that, given the current state a and the current control c, determines the optimal control c' for the next period. In other words, it gives the dynamics of the control variable.

Step 6: Find the Initial Control

The Euler equation allows us to find the control in the next period, c', given the current state a and the current control c. However, it does not allow us to find the initial control c_0, since there is no Euler equation for the period before $t = 0$ (which is a period that does not exist!).

To find the initial c_0, we use the condition that the value function must be finite. Without further assumptions, it is difficult to find c_0. Additionally, neither the existence nor the uniqueness of c_0 are guaranteed. In most cases, the best way to go is to make further assumptions that ensure the existence of a unique c_0 and that allow us to find workable

conditions to actually find such a c_0. For example, we could make assumptions on u and g (e.g., marginal utility goes to $-\infty$ as c goes to 0), on the behavior of V', or by adding further constraints in the original problem (such as a "no Ponzi-scheme" condition, i.e., a condition that precludes accumulation of negative a at a fast enough rate).

0.1.8 Stochastic Infinite Horizon Problems in Economics[2]

We now add some randomness to the deterministic problem of the last section and characterize its solution following the same steps.

Given initial condition a_0, choose $\{c_t\}_{t=0}^{\infty}$ to maximize

$$\sum_{t=0}^{\infty} \beta^t \theta_t u(a_t, c_t)$$

subject to the dynamic system

$$a_{t+1} = g(a_t, c_t). \tag{16}$$

where, as before, $u(a_t, c_t)$ is twice-differentiable concave period utility function, a_t is the state variable, and c_t is the control variable. The only new element is the preference shock θ_t that multiplies $u(a_t, c_t)$. We assume θ_t is determined at time t and observed before the consumption decision. The shock can take only two values: θ_h or θ_l with $\theta_h > \theta_l > 0$. We also assume the preference shock follows a Markov process; therefore, the distribution of θ_t only depends on the realization θ_{t-1} of θ in the previous period. The main difference compared to the deterministic version of the problem considered earlier is that the value function is not only a function of a, but also a function of the current realization of the taste shock, θ_t. In other words, there are two state variables: a and θ.

As before, we seek to find a policy function h that maps the state (a_t, θ_t) into the control c_t

2. The content of this section is an adaptation of notes created by Pascal Michaillat that are licensed under the Creative Commons Attribution 4.0 International License. The original notes can be found at https://github.com/pmichaillat/math-for-macro. All errors are mine.

such that the sequence $\{c_t\}_{t=0}^{\infty}$ generated by iterating the two functions

$$c_t = h(a_t, \theta_t)$$
$$a_{t+1} = g(a_t, c_t)$$

solves the original problem.

Step 1: Bellman Equation

The Bellman equation for this problem is:

$$V(a, \theta) = \max_c \{\theta u(a, c) + \beta E_t[V(a', \theta')]\}. \tag{17}$$

Compared to the Bellman equation for the deterministic case in equation (7), there are three differences:

1. The value function is a function of a and θ instead of being a function of a only; on the right-hand side of the Bellman equation, a' is the value of assets in the next period, and θ' denotes the (random) value of the preference shock next period.

2. The period utility $\theta u(a, c)$ has the preference shock θ, which was absent in the deterministic case,

3. Since the value function in the next period $V(a', \theta')$ is random (because it depends on θ') there is an expectation E_t in front of it. The subscript t in E_t denotes that we are taking expectations conditional on all information available at time t. In the problem we are considering, the information available is the current value θ of the preference shock. Because the preference shock follows a Markov process, the current value of the preference shock contains information about the probability distribution of tomorrow's value, θ'. In addition, because of the Markov property, the current value of the preference shock is the only relevant information.

Using that $a_{t+1} = g(a_t, c_t)$, we can re-write equation (17) as

$$V(a, \theta) = \max_c \{\theta u(a, c) + \beta E_t[V(g(a, c), \theta')]\}. \tag{18}$$

Step 2: First-Order Condition

Taking the first-order condition with respect to c in the optimization problem (18) yields

$$\theta \frac{\partial u}{\partial c}(a,c) + \beta E_t \left[\frac{\partial g}{\partial c}(a,c) \frac{dV}{da}(g(a,c),\theta') \right] = 0, \tag{19}$$

where $dV/da(g(a,c),\theta')$ is the derivative of V with respect to its first argument, evaluated at $(g(a,c),\theta')$. Using that $g(a,c) = a'$ and that $\partial g(a,c)/\partial c$ is known at t, we get

$$\theta \frac{\partial u}{\partial c}(a,c) + \frac{\partial g}{\partial c}(a,c) \beta E_t \left[\frac{dV}{da}(a',\theta') \right] = 0, \tag{20}$$

Step 3: Benveniste-Scheinkman Equation

By the Benveniste-Scheinkman theorem,

$$\frac{dV}{da}(a,\theta) = \theta \frac{\partial u}{\partial a}(a,c) + \beta E_t \left[\frac{\partial g}{\partial a}(a,c) \frac{dV}{da}(a',\theta') \right]. \tag{21}$$

Using that $\partial g(a,c)/\partial a$ is known at t, we get

$$\frac{dV}{da}(a,\theta) = \theta \frac{\partial u}{\partial a}(a,c) + \frac{\partial g}{\partial a}(a,c) \beta E_t \left[\frac{dV}{da}(a',\theta') \right]. \tag{22}$$

Step 3': A Combination

Eliminating $E_t \left[\frac{dV}{da}(a',\theta') \right]$ from the FOC (20) and the Benveniste-Scheinkman equation (22) gives

$$\frac{\partial V}{\partial a}(a,\theta) = \theta \frac{\partial u}{\partial a}(a,c) - \frac{\theta \frac{\partial g}{\partial a}(a,c) \frac{\partial u}{\partial c}(a,c)}{\frac{\partial g}{\partial c}(a,c)}. \tag{23}$$

Step 4: One Step Forward

Equation (23) is true for any value of the state variable a. In particular, it is true for $a' = g(a,h(a))$. Therefore, using that $c = h(a,\theta)$, we have

$$\frac{\partial V}{\partial a}(a',\theta') = \theta \frac{\partial u}{\partial a}(a',c') - \frac{\theta \frac{\partial g}{\partial a}(a',c') \frac{\partial u}{\partial c}(a',c')}{\frac{\partial g}{\partial c}(a',c')}. \tag{24}$$

Step 5: Euler Equation

We plug equation (24) into equation (20) to get the Euler equation

$$0 = \theta \frac{\partial u}{\partial c}(a,c) + \beta E_t \left[\frac{\partial g}{\partial c}(a,c) \left[\theta \frac{\partial u}{\partial a}(a',c') - \frac{\theta \frac{\partial g}{\partial a}(a',c') \frac{\partial u}{\partial c}(a',c')}{\frac{\partial g}{\partial c}(a',c')} \right] \right]$$

or, after some manipulations,

$$1 = \beta E_t \left[\frac{\frac{\partial u}{\partial c}(a', c')}{\frac{\partial u}{\partial c}(a, c)} \frac{\frac{\partial g}{\partial c}(a, c)}{\frac{\partial g}{\partial c}(a', c')} \frac{\partial g}{\partial a}(a', c') - \frac{\frac{\partial g}{\partial c}(a, c)}{\frac{\partial u}{\partial c}(a, c)} \frac{\partial u}{\partial a}(a', c') \right]$$

The last equation, together with $a' = g(a, h(a))$, characterizes the dynamics of c.

Step 6: Find the Initial Control

As in the deterministic case, the Euler equation does not pin down c_0.

Example 0.1.1.1. If the utility function is CRRA, we have

$$\begin{aligned} u(a, c) &= \frac{c^{1-\gamma}}{1-\gamma} \\ \frac{\partial u}{\partial c}(a, c) &= c^{-\gamma} \\ \frac{\partial u}{\partial a}(a, c) &= 0 \end{aligned}$$

In addition, if the function g is a budget constraint that represents investment in a single asset with returns R,

$$\begin{aligned} g(a, c) &= Ra_t - c_t \\ \frac{\partial g}{\partial c}(a, c) &= -1 \\ \frac{\partial g}{\partial a}(a, c) &= R \end{aligned}$$

Plugging the above expressions into the Euler equation (0.1.8) gives

$$1 = \beta E_t \left[\left(\frac{c'}{c} \right)^{-\gamma} R \right]$$

or

$$\frac{c^{-\gamma}}{\beta R} = \beta E_t \left[(c')^{-\gamma} \right].$$

We can make the dependence of c on the two-state Markov process θ explicit and write

$$\frac{c(\theta)^{-\gamma}}{(1+r)\beta} = p(\theta_h \mid \theta) c'(\theta_h)^{-\gamma} + (1 - p(\theta_l \mid \theta)) c'(\theta_l)^{-\gamma}$$

Writing the last equation for $\theta = \theta_l$ and $\theta = \theta_h$ gives

$$\frac{c(\theta_l)^{-\gamma}}{(1+r)\beta} = p(\theta_h \mid \theta_l)c'(\theta_h)^{-\gamma} + (1 - p(\theta_l \mid \theta_l))c'(\theta_l)^{-\gamma}$$
$$\frac{c(\theta_h)^{-\gamma}}{(1+r)\beta} = p(\theta_h \mid \theta_h)c'(\theta_h)^{-\gamma} + (1 - p(\theta_l \mid \theta_h))c'(\theta_l)^{-\gamma}$$

which is a system of two equations in the two unknowns $c'(\theta_l)^{-\gamma}$ and $c'(\theta_h)^{-\gamma}$. □

0.2 Solution Algorithms

To develop solution algorithms, we introduce some vector notation and operations. We focus on the case with a finite number of states and a finite number of possible actions. Assume that the states $S = \{1, 2, \ldots, n\}$ and actions $X = \{1, 2, \ldots, m\}$ are indexed by the first n and m integers, respectively. Let $v \in \mathbb{R}^n$ denote an arbitrary value vector:

$$v_i \in \mathbb{R} = \text{ value in state } i;$$

and let $x \in X^n$ denote an arbitrary policy vector:

$$x_i \in X = \text{ action in state } i.$$

Also, for each policy $x \in X^n$, let $f(x) \in \mathbb{R}^n$ denote the n-vector of rewards earned in each state when one follows the prescribed policy:

$$f_i(x) = \text{ reward in state } i, \text{ given action } x_i \text{ taken;}$$

and let $P(x) \in \mathbb{R}^{n \times n}$ denote the n-by-n state transition probabilities when one follows the prescribed policy:

$$P_{ij}(x) = \text{ probability of jump from state } i \text{ to } j, \text{ given action } x_i \text{ is taken.}$$

Given this notation, it is possible to express Bellman's equation for the finite horizon model succinctly as a recursive vector equation. Specifically, if $v_t \in \mathbb{R}^n$ denotes the value function in period t, then

$$v_t = \max_x \{f(x) + \beta P(x) v_{t+1}\},$$

were the maximization is the vector operation induced by maximizing each row individually.

Given the recursive nature of the finite horizon Bellman equation, one may compute the optimal value and policy functions v_t and x_t using backward recursion:

Algorithm: Backward Recursion

1. Initialization: Specify the rewards f, transition probabilities P, discount factor β, terminal period T, and post-terminal value function v_{T+1}; set $t \leftarrow T$.
2. Recursion Step: Given v_{t+1}, compute v_t and x_t :

$$v_t \leftarrow \max_x \{f(x) + \beta P(x) v_{t+1}\}$$

$$x_t \leftarrow \operatorname*{argmax}_x \{f(x) + \beta P(x) v_{t+1}\}.$$

3. Termination Check: If $t = 1$, stop; otherwise set $t \leftarrow t - 1$ and return to step 1 .

Each recursive step involves a finite number of matrix-vector operations, implying that the finite horizon value functions are well-defined for every period. Note however, that it may be possible to have more than one sequence of optimal policies if ties occur in Bellman's equation. Since the algorithm requires exactly T iterations, it terminates in finite time with the value functions precisely computed and at least one optimal policy obtained.

Consider now the infinite horizon Markov decision model. Given the notation above, it is also possible to express the infinite horizon Bellman equation as a vector fixed-point equation

$$v = \max_x \{f(x) + \beta P(x) v\}.$$

This vector equation may be solved using standard value function iteration methods:

Algorithm: Value Function Iteration

1. Initialization: Specify the rewards f, transition probabilities P, discount factor β, convergence tolerance τ, and initial guess for the value function v.
2. Function Iteration: Update the value function v:

$$v \leftarrow \max_x \{f(x) + \beta P(x) v\}.$$

3. Termination Check: If $\|\Delta v\| < \tau$, set

$$x \leftarrow \underset{x}{\text{argmax}}\{f(x) + \beta P(x)v\}$$

and stop; otherwise return to step 1.

The Bellman vector fixed-point equation for an infinite horizon model may alternatively be recast at a root finding problem

$$v - \max_x\{f(x) + \beta P(x)v\} = 0$$

and solved using Newton's method. By the Envelope Theorem, the derivative of the left-hand-side with respect to v is $I - \beta P(x)$ where x is optimal for the embedded maximization problem. As such, the Newton iteration rule is

$$v \leftarrow v - (I - \beta P(x))^{-1}(v - f(x) - \beta P(x)v)$$

where P and f are evaluated at the optimal x. After algebraic simplification the update rule may be written

$$v \leftarrow (I - \beta P(x))^{-1} f(x).$$

Newton's method applied to Bellman's equation traditionally has been referred to as policy iteration:

Algorithm: Policy Iteration

1. Initialization: Specify the rewards f, transition probabilities P, discount factor β, and an initial guess for v.

2. Policy Iteration: Given the current value approximant v, update the policy x:

$$x \leftarrow \underset{x}{\text{argmax}}\{f(x) + \beta P(x)v\}$$

and then update the value by setting

$$v \leftarrow (I - \beta P(x))^{-1} f(x)$$

3. Termination Check: If $\Delta v = 0$, stop; otherwise return to step 1.

At each iteration, policy iteration either finds the optimal policy or offers a strict improvement in the value function. Because the total number of states and actions is finite, the total number of admissible policies is also finite, guaranteeing that policy iteration will terminate after finitely many iterations with an exact optimal solution. Policy iteration, however, requires the solution of a linear equation system. If $P(x)$ is large and dense, the linear equation could be expensive to solve, making policy iteration slow and possibly impracticable. In these instances, the value function iteration algorithm may be the better choice.

Optional

*0.2.1 Dynamic Analysis

The path followed by a controlled, finite horizon, deterministic, discrete, Markov decision process is easily computed. Given the state transition function g and the optimal policy functions x_t^*, the path taken by the state from an initial point s_1 can be computed as follows:

$$\begin{aligned} s_2 &= g\left(s_1, x_1^*\left(s_1\right)\right) \\ s_3 &= g\left(s_2, x_2^*\left(s_2\right)\right) \\ s_4 &= g\left(s_3, x_3^*\left(s_3\right)\right) \\ &\vdots \\ s_{T+1} &= g\left(s_T, x_T^*\left(s_T\right)\right). \end{aligned}$$

Given the path of the controlled state, it is straightforward to derive the path of actions through the relationship $x_t = x_t^*\left(s_t\right)$. Similarly, given the path taken by the controlled state and action allows one to derive the path taken by any function of the state and action.

A controlled, infinite horizon, deterministic, discrete Markov decision process can be analyzed similarly. Given the state transition function g and optimal policy function x^*, the path taken by the controlled state from an initial point s_1 can be computed from the iteration rule:

$$s_{t+1} = g\left(s_t, x^*\left(s_t\right)\right).$$

The steady-state of the controlled process can be computed by continuing to form iterates until they converge. The path and steady-state values of other endogenous variables, including the action variable, can then be computed from the path and steady-state of the controlled state.

Analysis of controlled, stochastic, discrete Markov decision processes is a bit more complicated because such processes follow a random, not a deterministic, path. Consider a finite horizon process whose optimal policy x_t^* has been derived for each period t. Under the optimal policy, the controlled state will be a finite horizon Markov chain with nonstationary transition probability matrices P_t^*, whose row i, column j element is the probability of jumping from state i in period t to state j in period $t+1$, given that the optimal policy $x_t^*(i)$ is followed in period t:

$$P_{tij}^* = \Pr\left(s_{t+1} = j \mid x_t = x_t^*(i), s_t = i\right)$$

The controlled state of an infinite horizon, stochastic, discrete Markov decision model with optimal policy x^* will be an infinite horizon stationary Markov chain with transition probability matrix P^* whose row i, column j element is the probability of jumping from state i in one period t to state j in the following period, given that the optimal policy $x^*(i)$ is followed:

$$P_{ij}^* = \Pr\left(s_{t+1} = j \mid x_t = x^*(i), s_t = i\right)$$

Given the transition probability matrix P^* for the controlled state it is possible to simulate a representative state path, or, for that matter, many representative state paths, by performing Monte Carlo simulation. To perform Monte Carlo simulation, one picks an initial state, say s_1. Having the simulated state $s_t = i$, one may simulate a jump to s_{t+1} by randomly picking a new state j with probability P_{ij}^*.

The path taken by the controlled state of an infinite horizon, stochastic, discrete Markov model may also be described probabilistically. To this end, let Q_t denote the matrix whose row i, column j entry gives the probability that the process will be in state j in period t, given that it is in state i in period 0 . Then the t-period transition probability

matrices Q_t are simply the matrix powers of P:

$$Q_t = P^t$$

where $Q_0 = I$. Given the t-period transition probability matrices Q_t, one can fully describe, in a probabilistic sense, the path taken by the controlled process from any initial state $s_0 = i$ by looking at the i^{th} rows of the matrices Q_t.

In most economic applications, the multiperiod transition matrices Q_t will converge to a matrix Q as t goes to infinity. In such cases, each entry of Q will indicate the relative frequency with which the controlled decision process will visit a given state in the longrun, when starting from given initial state. In the event that all the columns of Q are identical and the longrun probability of visiting a given state is independent of initial state, then we say that the controlled state process possesses a steady-state distribution. The steady state distribution is given by the probability vector π that is the common row of the matrix Q. Given the steady-state distribution of the controlled state process, it becomes possible to compute summary measures about the longrun behavior of the controlled process, such as its longrun mean or variance. Also, it is possible to derive the longrun probability distribution of the optimal action variable or the longrun distribution of any other variables that are functions of the state and action.