Dynamic Programming and Dynamic Games

Debraj Ray

Boston University and Instituto de Análisis Económico (CSIC)

Version 1.0, 1995

This is a preliminary set of lecture notes used in Economics 718, Fall 1995, Boston University. Comments most appreciated. I expect these notes to evolve and improve over the coming few years.

Discounted Dynamic Programming under Certainty

1 The Basic Model

The following ingredients are part of the basic model. I will try to keep the notation as close as possible to Stokey and Lucas [1989]:

[1] A set of *states*, to be denoted by X.

[2] A *feasibility correspondence* that maps the current state into a subset of the state space, describing what states are attainable "tomorrow" given the state "today": $\Gamma : X \mapsto X$.

[3] A *one-period return function* F defined on the domain $A \equiv \{(x, y) \in X^2 : y \in \Gamma(x)\}$.

[4] A discount factor $\beta \geq 0$ applied to future payoffs.

Let some initial state x be given. A *feasible program* $\mathbf{x}$ from x is a sequence $\{x_t\}_{t=0}^{\infty}$ such that $x_0 = x$ and $x_{t+1} \in \Gamma(x_t)$ for all $t \geq 0$. The problem is to find a feasible program $\mathbf{x}$ that attains

$$\sum_{t=0}^{\infty} \beta^t F(x_t, x_{t+1}) \tag{1}$$

2 Examples

[1] **Optimal Growth.** Suppose that inputs capital (K) and labor (N) produce a single good (Y), which may be used for consumption (C) or for augmenting the capital stock. Population grows at the rate of n, and there is labor-augmenting technical progress at the rate of γ. If G_t is the gross production function at time t (including the past capital stock, possibly depreciated), then

$$G_t(K, N) = G(K, A_t N)$$

is the production function at time t, where A_t represents labor-augmenting technical progress. Assume G is constant returns to scale.

By assumption, $A_t = A_0(1 + \gamma)^t$ for some $A_0 > 0$ and $\gamma > -1$, and $N_t = N_0(1 + n)^t$ for some $N_0 > 0$ and $n > -1$.

Now, a feasible program from some initial level of output Y is A sequence $\{C_t, K_t, Y_t\}$ such that

$$\begin{aligned} Y_0 &= Y \\ Y_t &= C_t + K_t \text{ for all } t \geq 0 \\ y_{t+1} &= G(K_t, A_t N_t) \end{aligned} \tag{2}$$

Normalize this by dividing through by effective labor AN. Let $c \equiv \frac{C}{AN}$, $k \equiv \frac{K}{AN}$, and $x \equiv \frac{Y}{AN}$. Let $g(k) \equiv G(k, 1)$. We thus obtain the system

$$\begin{aligned} x_0 &= x \\ x_t &= c_t + k_t \text{ for all } t \geq 0 \\ x_{t+1} &= \tfrac{1}{(1+\gamma)(1+n)} g(k_t) \end{aligned} \tag{3}$$

Now consider a constant-elasticity-of-substitution utility function, defined on individual consumption $\hat{c} \equiv \frac{C}{N}$, given by

$$u(\hat{c}) \equiv \frac{1}{1-\sigma}[\hat{c}^{1-\sigma} - 1]$$

The term $\frac{1}{\sigma}$ may be interpreted as the intertemporal elasticity of substitution. Optimal (utilitarian) growth requires the maximization of

$$\sum_{t=0}^{\infty} \delta^t u(\hat{c}) N_t$$

where δ is the planner's discount factor. With the appropriate substitutions made, this can be shown to be equivalent to the maximization of

$$\sum_{t=0}^{\infty} \beta^t \frac{1}{1-\sigma}[c_t^{1-\sigma} - 1] \tag{4}$$

where $\beta \equiv \delta(1+n)(1+\gamma)^{1-\sigma}$. The system (3) and (4) can now be cast in the form of the general model, by simply factoring out the variables c and k.

[2] **Orchards or vintages.** Consider a unit plot of land (say, the unit square of $I\!R^2$) with trees of various ages planted on it. Each tree lives for periods $0, 1, \ldots, T$, after which it is incapable of bearing fruit. The tree yields an amount $R(s)$ in period s, $S = 0, 1, \ldots, t$.

Let X be the nonnegative unit simplex of $I\!R^{T+1}$. The interpretation is that $x \in X$ is a vector which describes the fractions of land devoted to trees of different ages.

Given x at time t, it is possible to move to a set of new vectors at time $t+1$ by eliminating trees of different vintages and replacing them with trees of age 0. Uncleared trees will advance in age by one period. Thus,

$$\begin{aligned} \Gamma(x) &= \{y \in X : \text{ there exist nonnegative reals } (\epsilon(0), \ldots, \epsilon(T)) \text{ with} \\ y(s+1) &= x(s) - \epsilon(s) \text{ for } s = 0, \ldots, T-1, \epsilon(T) = x(T), \text{ and } y(0) = \sum_{s=0}^{T} \epsilon(s)\} \end{aligned}$$

A plot with x yields a harvest of $c(x) \equiv \sum_{s=0}^{T} x(s)R(s)$. If the utility function is $u(c)$, then in the context of our general framework, this means that the payoff function takes the special form on A:

$$F(x, y) = u(x \cdot R).$$

[3] **Exhaustible resources.** A country has exhaustible oil reserves x which it can sell on the world market. If the quantity sold at date t is y, then price is determined by an inverse demand function $p(y)$, so that total profit at date t is given by $yp(y)$.

Clearly, $X = I\!R_+$, and

$$\Gamma(x) = \{x' \geq 0 : \ x \geq x'\}.$$

and

$$F(x, x') = (x - x')p(x - x').$$

For other examples, see Stokey and Lucas [1989], Ch. 5.

3 The functional equation of dynamic programming

The following fundamental assumptions will be maintained throughout. For the sake of comparison, it should be noted that we are using assumptions that are slightly stronger than those in Stokey and Lucas.

[**A.1**] $\Gamma(x)$ is nonempty for all $x \in X$.

[**A.2**] (i) For all feasible programs, $\lim_{T\to\infty} \sum_{t=0}^{T} \beta^t F(x_t, x_{t+1})$ is well-defined and less than $+\infty$. Moreover, for every x, there exists a feasible program $\mathbf{x}$ such that $\lim_{T\to\infty} \sum_{t=0}^{T} \beta^t F(x_t, x_{t+1}) > -\infty$.[1]

It is important to understand that the assumption (A.2) is not the same as the postulate of discounting. We will return to this point below.

For any feasible program $\mathbf{x}$ from x, define $u(\mathbf{x})$ to be the infinite discounted sum of returns generated by this program:

$$u(\mathbf{x}) \equiv \sum_{t=0}^{\infty} \beta^t F(x_t, x_{t+1}).$$

Note that by (A.2), u is well-defined on the space of feasible programs, though it may be $-\infty$ in places. For each initial $x \in X$, we may now define the *supremum function* $v^*(x)$ as

$$v^*(x) \equiv \sup_{\mathbf{x}} u(\mathbf{x}) \tag{5}$$

where the supremum in (5) is taken over all feasible programs $\mathbf{x}$ from x.

By Assumptions (A.1) and (A.2), v^* is well-defined and assumes finite values for each $x \in X$.

Say that a finite-valued function v *satisfies the functional equation* if

$$v(x) = \sup_{y\in\Gamma(x)} [F(x,y) + \beta v(y)] \tag{6}$$

Our purpose in this section is to analyze the connection between the supremum function and the function(s) that satisfy the functional equation.

THEOREM **1** *Suppose that (A.1) and (A.2) are satisfied. Then the supremum function v^* satisfies the functional equation.*

Proof. Fix $x \in X$. Consider, first, any $y \in \Gamma(x)$. Because v^* is a supremum function, it follows that for any $\epsilon > 0$, there is a feasible program $\mathbf{x}'$ from y such that

$$u(\mathbf{x}') \geq v^*(y) - \epsilon \tag{7}$$

Note that $\mathbf{x} \equiv (x, \mathbf{x}')$ is a feasible program from x. Moreover,

$$v^*(x) \geq u(\mathbf{x}) = F(x,y) + \beta u(\mathbf{x}') \geq F(x,y) + \beta v^*(y) - \beta\epsilon,$$

[1] This assumption is a bit restrictive mathematically but does not rule out any cases of economic interest. For a more general treatment, see Stokey and Lucas [1989].

where the last inequality uses (7). Because this holds for any $\epsilon > 0$ and any $y \in \Gamma(x)$, it follows that

$$v^*(x) \geq \sup_{y \in \Gamma(x)} [F(x, y) + \beta v^*(y)] \tag{8}$$

It remains to show that equality must hold in (8). To this end, pick $x \in X$ and any $\epsilon > 0$. Because v^* is a supremum function, it follows that there exists a program $\mathbf{x}$ such that

$$v^*(x) \leq u(\mathbf{x}) + \epsilon = F(x, x') + \beta u(\mathbf{x}') + \epsilon,$$

where x' is the first term following x in the feasible program $\mathbf{x}$, and $\mathbf{x}'$ denotes the feasible program from x' induced by $\mathbf{x}$. Noting that $u(\mathbf{x}') \leq v^*(x')$, it follows that

$$v^*(x) \leq F(x, x') + \beta v^*(x') + \epsilon,$$

and because ϵ is arbitrary, it follows that equality must hold in (8). ■

The converse to Theorem 1 is a deeper issue. There are functions that satisfy the functional equation, but are *not* supremum functions.

EXAMPLE 1. Let $X = \mathbb{R}_+$, $\Gamma(x) = \{y : 0 \leq y \leq \lambda x\}$ for some $\lambda > 1$. $F(x, y)$ is defined by first looking at $c \equiv x - y/\lambda$ for all $(x, y) \in A$, and then defining $u(c) = \frac{c}{1+c}$. Finally, let the discount factor be given by $\beta = 1/\lambda$. Note that (A.1) and (A.2) are satisfied.

By standard arguments (which you should be able to supply), it follows that from every initial stock, it is best to maintain a constant level of c over time. This policy will give us the supremum function, and indeed you can see that it is

$$v^*(x) = \frac{\lambda x}{\lambda + (\lambda - 1)x}$$

Of course, v^* will satisfy the functional equation. It may be a good idea to directly verify this.

However, there is another solution to the functional equation, and this shows that the converse to Theorem 1 is not going to be straightforward. This solution is $v(x) \equiv x$. It will be good practice to check that this indeed satisfies the functional equation.

What is happening in this example? After all, there "should" not be other solutions to the functional equation. Why can't we just recursively unravel (6) to get (5)? Indeed, we can, provided that the tail $\beta^t v(x_t)$ of the unravelled series conveniently vanishes, leaving only the bonafide utility terms behind. If the tail does not vanish, we are in trouble. Re-examining the spurious solution to the functional equation in Example 1, we see that there are unravellings for which the tail does not vanish: $\beta^t v(x_t) = \beta^t x_t$, and any sequence along which x_t grows at an asymptotic rate of $\lambda = 1/\beta$ will create a nonvanishing tail.

Theorem 2 establishes a converse to Theorem 1 by assuming this problem away. Later on, however, we shall see that when the discounting assumption is made, this is a tight assumption, in the sense that it will always be satisfied by the supremum function of the model.

Theorem 2 *Assume (A.1) and (A.2). Let v be a solution to the functional equation satisfying the condition*

$$\lim_{t\to\infty} \beta^t v(x_t) = 0 \tag{9}$$

for every $x \in X$ and every feasible program from x. Then v must be the supremum function.

Proof. Let v satisfy the functional equation (6). We begin by observing that for every $x \in X$ and feasible program $\mathbf{x}$ from x,

$$\begin{aligned} v(x) &\geq F(x_0, x_1) + \beta v(x_1) \\ &\geq F(x_0, x_1) + \beta F(x_1, x_2) + \beta^2 v(x_2) \\ \ldots &\geq \sum_{t=0}^{T} \beta^t F(x_t, x_{t+1}) + \beta^{T+1} v(x_{T+1}). \end{aligned}$$

Passing to the limit as $T \to \infty$ and invoking (9), we see that

$$v(x) \geq u(\mathbf{x}) \tag{10}$$

for each $x \in X$ and feasible program $\mathbf{x}$ from x.

It remains to prove that (10) must hold with equality. To do this, pick any $\epsilon > 0$, and define a strictly positive sequence $\{\epsilon_t\}_{t=0}^{\infty}$ such that $\sum_{t=0}^{\infty} \beta^t \epsilon_t < \epsilon$. Because v satisfies the functional equation, we may define a feasible program $\mathbf{x}$ from x such that for each t,

$$v(x_t) \leq F(x_t, x_{t+1}) + \beta v(x_{t+1}) + \epsilon_t$$

Manipulating this set of inequalities up to time T, we see that

$$v(x) \leq \sum_{t=0}^{T} \beta^t F(x_t, x_{t+1}) + \beta^{T+1} v(x_{T+1}) + \sum_{t=0}^{T} \beta^t \epsilon_t, \tag{11}$$

and passing to the limit in (11) as $T \to \infty$, we see that

$$v(x) \leq u(\mathbf{x}) + \epsilon.$$

Note that ϵ was arbitrary, so that (10) must hold with equality, and the proof is complete. ■

Let us briefly return to an observation made just before the statement of Theorem 2. I claimed that when the discounting assumption is made, the condition of that theorem would be automatically satisfied. We will see this presently, but it is important to note that in general, the condition (9) may be violated by all solutions to the functional equation, *including* the supremum function (though this is not the case in Example 1). To see this, consider a variant of Example 1 with $\lambda = \beta = 1$. Verify that the supremum function now violates the tail condition of Theorem 2.

4 Optimal Programs and Policies

A feasible program $\mathbf{x}^*$ from $x \in X$ is *optimal* if $v^*(x) = u(\mathbf{x}^*)$. Our first Theorem, which runs parallel to Theorem 1, shows that optimal programs attain the supremum of the functional equation (6).

THEOREM **3** *Suppose that (A.1) and (A.2) are satisfied. Let $\mathbf{x}^*$ from x be an optimal program. Then*

$$v^*(x_t^*) = F(x_t^*, x_{t+1}^*) + \beta v^*(x_{t+1}^*) \tag{12}$$

for all $t \geq 0$.

Proof. Denote by $\mathbf{x}^t$ the "tail" of the optimal program from t onwards — with initial stock x_t^*. We claim that each such program is optimal from its corresponding initial stock. For suppose that this assertion were false for some t. Then there exists a program $\mathbf{x}'$ from x_t such that $u(\mathbf{x}') > u(\mathbf{x}^t)$. Now consider the program $\mathbf{x}'' \equiv (x_0, x_1, \ldots, x_t, \mathbf{x}')$: it is certainly feasible from x. However

$$\begin{aligned} u(\mathbf{x}'') &= \sum_{s=0}^{t-1} \beta^s F(x_s^*, x_{s+1}^*) + \beta^t u(\mathbf{x}') > \sum_{s=0}^{t-1} \beta^s F(x_s^*, x_{s+1}^*) + \beta^t u(\mathbf{x}^t) \\ &= u(\mathbf{x}^*) \end{aligned}$$

which contradicts the optimality of $\mathbf{x}^*$. So the claim is true, which means that $v^*(x_t^*) = u(\mathbf{x}^t)$ for each t.

We now use the claim twice to complete the proof. Note that for any t,

$$\begin{aligned} v^*(x_t^*) &= u(\mathbf{x}^t) = F(x_t^*, x_{t+1}^*) + \beta u(\mathbf{x}^{t+1}) \\ &= F(x_t^*, x_{t+1}^*) + \beta v(x_{t+1}^*), \end{aligned}$$

and we are done. ■

The converse to this theorem requires not only that (12) be satisfied, but that an additional boundary condition is met. This is analogous to the tail condition used in Theorem 2.

THEOREM **4** *Suppose that (A.1) and (A.2) are satisfied. Let $\mathbf{x}^*$ be a feasible program from x which satisfies (12) using the supremum function, and meets in addition the boundary condition*

$$\limsup_{t \to \infty} \beta^t v^*(x_t^*) \leq 0. \tag{13}$$

Then $\mathbf{x}^*$ is an optimal program from x.

Proof. Suppose that $\mathbf{x}^*$ satisfies (12). Using this repeatedly for T periods, we see that

$$v^*(x) = \sum_{t=0}^{T} \beta^t F(x_t^*, x_{t+1}^*) + \beta^{T+1} v^*(x_{T+1}^*).$$

Passing to the limit in this expression and applying (13),

$$v^*(x) \leq u(\mathbf{x}^*).$$

The reverse equality also holds since v^* is the supremum function, and we are done. ■

Note well: while I said that this condition (13) is analogous to the tail condition (9), it is far from being the same thing. The present condition is applied to the supremum function itself, in contrast to the tail condition. For this condition to have any bite, then, it must be the case that the tail condition (9) itself fail to apply to the supremum function. So, for instance, Example 1, which illustrated well Theorem 2, will not apply in this case. A modification of that example, however, illustrates the need for (13).

EXAMPLE 2. Put $\beta = \lambda = 1$ in Example 1, and change the utility function to $u(c) = c$. This is just a cake-eating problem with a linear utility function. There are lots of optimal programs from any initial stock x: essentially, any intertemporal division of the cake will do. The supremum function turns out to be (not surprisingly) $v^*(x) = x$ for all $x \geq 0$. Now consider a program that is definitely suboptimal: one that puts $x_t = x$ for all t (so that consumption is always zero). The point is that this program satisfies (12), while it fails (13).

So unimprovability does not necessarily imply optimality. Later, however, we shall see how the assumption of discounting enables to get around this problem.

To finish this section, we define a *policy correspondence*: this is any correspondence $G : X \to X$ with the property that $G(x) \subseteq \Gamma(x)$ for all $x \in X$. If G is single-valued it will be referred to as a policy *function*. A feasible program $\mathbf{x}$ from x is *generated by* a policy correspondence G if $x_0 = x$, and $x_{t+1} \in \Gamma(x_t)$ for all t.

The *optimal* policy correspondence is given by

$$G^*(x) \equiv \{y \in \Gamma(x) : v^*(x) = F(x,y) + \beta v^*(y)\} \quad (14)$$

Up to now (but not for long), this is a bit of a misnomer. While Theorem 3 shows that every optimal plan must be generated by G^*, Theorem 4 argues that plans generated by G^* are optimal *provided* they satisfy (13).

In the next section, we will introduce the notion of discounting and show that this takes away the problems raised in this and the previous section.

5 Insignificant future

Assumption (A.2), with its insistence that utility streams must sum to a number less than plus infinity, looks like discounting. But it isn't. You can see this, for instance, by studying Example 2. There is no discounting in that example. Yet Assumption (A.2) is satisfied. Discounting requires that the future must look "uniformly" insignificant (relative to the present), no matter which feasible program one tries out. Along any (nonwasteful) program in the cake-eating problem, the future looks insignificant compared to the present, because after all "most" of the cake must be eaten in a finite number of periods. Yet observe that unless we have discounting, this property does not hold *uniformly* across all feasible programs. This is the source of the problems raised by Theorems 2 and 4.

I now state the insignificant future assumption. Assume from now on that X is some nicely behaved topological space, say a metric space.

[D] There is a positive sequence $\epsilon(t) \to 0$ such that for *all* feasible programs $\mathbf{x}$ from x and all t,

$$|\sum_{s=t}^{\infty} \beta^s F(x_s, x_{s+1})| \leq h(x)\epsilon(t) \tag{15}$$

where $h(x)$ is a positive continuous function.

Understand this assumption well. It is a *joint* assumption on the discount factor, on the payoff function, and the "technology" (captured by Γ). It is not just discounting, even though we might commonly refer to it as such. It means, more generally, that the discount factor must be powerful enough to swamp any growth effects that are in the technology.

Exercise. [1] Show that in Example 1, Assumption (D) is satisfied. More generally, show that if the one-period return function is bounded, then the condition that the discount factor is less than unity is sufficient for Assumption (D) to hold.

[2] If the payoff function is not bounded, Assumption (D) generally demands more of the discount factor than just being less than unity. Consider a variant of Example 3 where β and λ need not equal unity. What does Condition (D) require in this case?

[3] Finally, consider a case where one-period payoffs, technology, and the discount factor all interact in a nontrivial way to determine the validity of Assumption (D). Take Item [1] in Section 2 (Examples). Find out a necessary and sufficient condition for (D) to be met in that model. [Hint: Assume G is differentiable and concave. You will need the variable $m \equiv \lim_{k\to\infty} g'(k)$ in your characterization.]

The insignificant future condition has two effects on the theorems of the previous section. First, it tightens Theorems 1 and 2 into a complete characterization:

Theorem **5** *Assume (A.1), (A.2), and (D). Then v is a supremum function if and only if it satisfies the functional equation and condition (9):*

$$\lim_{t\to\infty} \beta^t v(x_t) = 0$$

for every $x \in X$ and every feasible program from x.

Proof. In the light of Theorems 1 and 2 it suffices to prove that v^*, the supremum function, must satisfy (9). Take any $x \in X$ and feasible program $\mathbf{x}$ from x. Because v^* is a supremum function, there exists a positive sequence δ_t with $\beta^t \delta_t$ converging to zero such that for each t, there is a feasible program $\mathbf{x}^t$ from x_t with

$$u(\mathbf{x}^t) \leq v(x_t) \leq u(\mathbf{x}^t) + \delta_t.$$

Consequently, using the condition (D) [see how uniformity is being used in the next manipulation], we see that for each t,

$$\begin{aligned} -h(x)\epsilon(t) \leq \beta^t u(\mathbf{x}^t) \leq \beta^t v(x_t) &\leq \beta^t u(\mathbf{x}^t) + \beta^t \delta_t \\ &\leq h(x)\epsilon(t) + \beta^t \delta_t, \end{aligned}$$

so that $\beta^t v(x_t)$ converges to zero by the properties of $\epsilon(x, t)$ and the choice of the sequence δ_t. ■

Note that in the variant of Example 1 discussed at the very end of Section 3, the supremum function does not satisfy the tail condition (9). It follows, of course, that in that variant (D) is not satisfied. Check this directly.

Next, we will see that the insignificant future assumption links the pair of theorems 3 and 4 without any need for a tail condition like (13). Now describing G^* as the optimal policy correspondence is no longer a misnomer.

THEOREM **6** *A feasible program* $\mathbf{x}^*$ *is optimal if and only if (12) is satisfied:*

$$v^*(x_t^*) = F(x_t^*, x_{t+1}^*) + \beta v^*(x_{t+1}^*)$$

Proof. In the light of Theorems 3 and 4, it suffices to prove that (12) is sufficient for optimality. To this end it will be enough to prove that (13) is satisfied. But this restriction is implied by Theorem 5. ■

6 Existence of Optimal Programs

Our next task is to outline conditions under which an optimal program exists. There are two routes to this problem. One is a direct approach, which I shall first outline. The second approach establishes existence by studying the existence of value functions. This approach is really much more suitable for the stochastic case, and I will postpone a discussion of it to the lectures on uncertainty.

We will need additional assumptions, this time on the transformation correspondence and the one-period payoff function. Henceforth, Assumptions (A.1) and (A.2) will always be taken to implicitly hold, and we shall not refer to them anymore.

[T^U] Γ is a compact-valued, upperhemicontinuous (uhc) correspondence.

Remark. A compact-valued correspondence $\Gamma : X \to X$ is *uhc* at $x \in X$ if for all sequences $\{x_n\}$ converging to x and $\{y_n\}$ with $y_n \in \Gamma(x_n)$ for all n, there exists a convergent subsequence $\{y_{n_k}\}$ with limit y, such that $y \in \Gamma(x)$. Observe that this is a bit different from the standard definition that you have seen but can easily be reconciled with it.

[**F**] F is continuous on $X \times X$.

THEOREM **7** *Assume (D), (T^U), and (F). From each* $x \in X$*, an optimal program exists.*

Proof. The following lemma will be useful in the proof of this and other results:

LEMMA **1** *Let* $\mathbf{x}^n$ *be a sequence of feasible programs such that for each* t*,*

$$x_t^n \to x_t$$

as $n \to \infty$*. Then* $\mathbf{x} \equiv \{x_t\}$ *is a feasible program from* x_0*, and* $u(\mathbf{x}^n) \to u(\mathbf{x})$ *as* $n \to \infty$*.*

Proof. Our first task is to establish that $\mathbf{x}$ is feasible. This follows directly from the assumption that Γ is uhc; you can fill in the details.

Now, suppose that the rest of the assertion in the lemma is false. Then there exists a subsequence $\{r\}$ and $\epsilon > 0$ such that

$$|u(\mathbf{x}^r) - u(\mathbf{x})| \geq \epsilon$$

for all r. We shall only study the subcase

$$u(\mathbf{x}^r) \geq u(\mathbf{x}) + \epsilon; \tag{16}$$

the other case uses a parallel argument and is omitted.

By assumption (D), there exists some time T such that for all $t \geq T$ and all r,

$$h(x_0^r)\epsilon(t) < \epsilon/3 \text{ and } h(x_0)\epsilon(t) < \epsilon/3.$$

[This uses the continuity of $h(.)$.] Using this choice of T and (16),

$$\begin{aligned}
\sum_{t=0}^{T} \beta^t F(x_t^r, x_{t+1}^r) &\geq u(\mathbf{x}^r) - \epsilon/3 \\
&\geq u(\mathbf{x}) + 2\epsilon/3 \\
&= \sum_{t=0}^{T} \beta^t F(x_t, x_{t+1}) + \sum_{t=T}^{\infty} \beta^t F(x_t, x_{t+1}) + 2\epsilon/3 \\
&\geq \sum_{t=0}^{T} \beta^t F(x_t, x_{t+1}) + \epsilon/3
\end{aligned}$$

for *all* r. But this is impossible by (F). ■

Remark. Appreciate in the proof above the difference between the continuity of a *finite* sum of continuous functions, which is trivial, and the infinite sum, which is not.

Exercise. Show that in the cake-eating problem with any increasing, differentiable, utility function (with positive derivative at zero), and with discount factor equal to one, the assertion of Lemma 1 fails.

We now return to the main proof of existence. Fix $x \in X$ and let $M \equiv v^*(x)$. Because v^* is a supremum function, there exists a sequence of feasible programs $\mathbf{x}^n$ from x such that $u(\mathbf{x}^n) \to M$. Consider period 0: $x_1^n \in \Gamma(x)$ for all n. By uhc of Γ, there exists a subsequence $\{n_k\}$ such that $x_1^{n_k}$ converges to some $x_1^* \in X$. Recursively, suppose that at date t, there is a sequence $\{x_t^m\}$ converging to some x_t^*. Then there is a subsequence $\{m_k\}$ such that $x_{t+1}^{m_k}$ converges to some $x_{t+1}^* \in \Gamma(x_t^*)$. The problem is to extract out of all this *one* subsequence (call it $\{r\}$) such that

$$x_t^r \to x_t^*$$

for each t, where (by uhc we already know that) $\mathbf{x}^*$ is a feasible program from x. A diagonal argument due to Cantor allows us to do just that.

We now claim that $\mathbf{x}^*$ is indeed an optimal program. To see this, just use the lemma, which claims that $u(\mathbf{x}^r)$ must converge to $u(\mathbf{x}^*)$ as $r \to \infty$. But it is also true by construction that $u(\mathbf{x}^r) \to M$. It follows that $u(\mathbf{x}^*) = M$, and the proof is complete. ■

7 Value functions and policy correspondences

7.1 Continuity of v^*

Once an optimal program exists, we may rename the supremum function as the *value function*, the supremum being attained.

A fundamental property of the value function that we hope to obtain in many cases is its continuity. However, the assumptions made so far do not guarantee this.

EXAMPLE 3. Suppose that $X = I\!R_+$, and that the transformation correspondence is given by

$$\begin{aligned}\Gamma(x) &= \{y : 0 \leq y \leq x\} \text{ for } 0 \leq x < 1 \\ &= \{y : 0 \leq y \leq 4x\} \text{ for } x \geq 1\end{aligned}$$

Let the utility function be defined on consumption just as in Example 1: $u(c) = \frac{c}{1+c}$, and let the discount factor equal 1/4. Now verify that all the assumptions made so far are satisfied.

Take any initial $x \in [0, 1)$. Look at the marginal utility of consumption: it is $\frac{1}{(1+c)^2}$ which is bounded above by 1/4. So it is optimal for these initial stocks to consume everything in the first period. It follows that

$$v^*(x) = \frac{x}{1+x}$$

for all $x \in [0, 1)$. Now consider any initial stock $x \geq 1$. Then by the same argument as in Example 1, it is optimal to maintain constant consumption. From Example 1, it follows that the value of doing this,

$$v^*(x) = \frac{4x}{4+3x}$$

for all $x \geq 1$. Clearly, we do not have continuity of the value function at $x = 1$.

An assumption that guarantees continuity is an additional restriction on the transformation correspondence:

$[T^L]$ Γ is lower hemicontinuous (lhc) on X.

To recall the definition: A nonempty-valued correspondence Γ is *lhc* at $x \in X$ if for every $y \in \Gamma(x)$ and every sequence $x^n \to x$, there exists a sequence $\{y^n\}$ with $y^n \in \Gamma(x^n)$ for all n, such that $y^n \to y$.

THEOREM **8** *Suppose that assumptions (D), (F), (T^U), and (T^L) are satisfied. Then the value function v^* is continuous on X.*

Proof. Suppose that $x^n \to x$. We may assume wlog that $v^*(x^n)$ converges as well (why?). We wish to show that

$$\lim_{n\to\infty} v^*(x^n) = v^*(x).$$

For each n, pick an optimal program $\mathbf{x}^n$ from initial stock x^n, so that $u(\mathbf{x}^n) = v^*(x^n)$ for all n. By a diagonal argument similar to that used in the proof of Theorem 7, there exists a subsequence $\{r\}$ of $\{n\}$ such that for each t,

$$x_t^r \to x_t$$

as $r \to \infty$, where $\mathbf{x} = \{x_t\}$ is *some* feasible program from x. By Lemma 1,

$$u(\mathbf{x}^r) \to u(\mathbf{x})$$

as $r \to \infty$. This shows right away that

$$\lim_{r\to\infty} v^*(x^r) \le v^*(x).$$

Remark. So far we have showed, without invoking lhc of Γ, that the function v^* is uppersemicontinuous (review the definition of an uppersemicontinuous function and see why we have proved this).

Now suppose that equality failed to hold in the expression above. Then there exists a feasible program $\mathbf{y}$ from x such that

$$u(\mathbf{x}) < u(\mathbf{y}) \tag{17}$$

Using (T^L) repeatedly, we may now construct a sequence of feasible programs $\mathbf{y}^r$ from x^r, such that

$$y_t^r \to x_t$$

for each t. So by Lemma 1,

$$u(\mathbf{y}^r) \to u(\mathbf{y})$$

while we already know that

$$u(\mathbf{x}^r) \to u(\mathbf{x}).$$

But these two expressions, together with (17), contradict the assumption that for all r, $\mathbf{x}^r$ is an optimal program from x^r. ■

So under the assumptions so far, v^* is continuous. We already know that

$$v^*(x) = \max_{y\in\Gamma(x)} F(x,y) + \beta v^*(y). \tag{18}$$

Recall that for each $x \in X$,

$$G^*(x) \equiv \{y \in \Gamma(x) : v^*(x) = F(x,y) + \beta v^*(y)\}$$

describes the optimal policy correspondence. Recall, too, by Theorem 6, that G^* generates only (and all the) optimal programs from any initial state x. The continuity properties of G^* are are summarized in

THEOREM **9** *Under (F), (D), (T^U), and (T^L), G^* is an uhc correspondence.*

The proof of this theorem directly follows from the

MAXIMUM THEOREM *Consider the problem*

$$\max_{y\in\Gamma(x)} D(x,y)$$

where Γ *is a nonempty, compact-valued, continuous (uhc and lhc) correspondence and* D *is a continuous function. Then* $G(x)$*, which describes for each* x*, the set of* y*-solutions to this problem, is a nonempty, compact-valued, uhc correspondence.*[2]

Proof. Suppose that $\{x^n\} \to x$ and $y^n \in G(x^n)$ for every n. By the uhc of Γ the fact that $G(x) \subseteq \Gamma(x)$ for all x, there exists a subsequence $\{y^r\} \to y \in \Gamma(x)$. We must show that $y \in G(x)$. Suppose not. Then there exists $y' \in \Gamma(x)$ such that

$$D(x, y') > D(x, y) \tag{19}$$

Because Γ is lhc, there exists a sequence $y'^n \in \Gamma(x^n)$ such that $y'^n \to y'$. Because D is continuous, we know that

$$D(x^n, y^n) \to D(x, y)$$

and

$$D(x^n, y'^n) \to D(x, y').$$

But these two assertions, along with (19), imply that

$$D(x^n, y'^n) > D(x^n, y^n)$$

which contradicts the fact that $y^n \in G(x^n)$ for all n. ■

Exercise. [1] Find an example where all the assumptions of the maximum theorem are satisfied, except for the lhc of Γ, and where the conclusion of the theorem is false.

[2] Find an example where under the conditions of the maximum theorem, G is not lhc.

7.2 Concavity of v^*

In economics, the assumption of convexity properties typically allow us to establish further properties of value functions and policies. In what follows, then, we take X to be some metric vector space.

Two kinds of convexity assumptions are often invoked. One has to do with the transformation possibilities: if (x, y) and (x', y') are both feasible, then so is any convex combination of them; this is equivalent to the assumption

[T^{conv}] $A = \{(x, y) : y \in \Gamma(x)\}$ is a convex set.

The other kind of convexity assumption describes the convexity of preferences:

[F^{conc}] F is strictly concave on A.

Exercise. Consider a single-person intertemporal allocation problem, such as the cake-eating problem, with strictly concave utility function defined on consumption. Prove that the induced F function on A is strictly concave.

[2] The Maximum Theorem also asserts the continuity of the maximized function in x, something that can be proved along the lines of the proof of Theorem 8.

THEOREM **10** *Assume (D), (T^U), (T^L), (T^{conv}), (F) and (F^{conc}). Then v^* is a strictly concave function on X, and the optimal policy correspondence is a continuous function.*

Proof. Let x and y be two initial states in X, with $x \neq y$. Let $\mathbf{x}$ and $\mathbf{y}$ be optimal programs from x and y respectively. Fix and $\lambda \in (0, 1)$. Let $\mathbf{z}$ be the program formed by defining

$$z_t \equiv \lambda x_t + (1 - \lambda) y_t$$

for all t. Using the assumption (T^{conv}), it follows that $\mathbf{z}$ thus defined is a feasible program from z_0. Moreover, by the strict concavity of F, we have

$$F(\lambda x_t + (1 - \lambda) y_t, \lambda x_{t+1} + (1 - \lambda) y_{t+1}) \geq \lambda F(x_t, x_{t+1}) + (1 - \lambda) F(y_t, y_{t+1})$$

for all t, with strict inequality holding whenever $(x_t, x_{t+1}) \neq (y_t, y_{t+1})$, which is the case, for instance, at $t = 0$. So because $v^*(x) = \sum_{t=0}^{\infty} \beta^t F(x_t, x_{t+1})$, $v^*(y) = \sum_{t=0}^{\infty} \beta^t F(y_t, y_{t+1})$ and

$$v^*(\lambda x + (1 - \lambda) y) \geq \sum_{t=0}^{\infty} \beta^t F(\lambda x_t + (1 - \lambda) y_t, \lambda x_{t+1} + (1 - \lambda) y_{t+1}),$$

the strict concavity of v^* is established.

To complete the proof we observe that the optimal policy correspondence G^* must be a singleton at each x. This is because the solution to a strictly concave maximization problem like (18) must be unique. But now we are done, because a single-valued uhc correspondence must in fact be a continuous function [check this]. ■

7.3 Differentiability of v^*

Under still stronger conditions, one obtains a differentiable value function. This result, as we shall see, lies at the heart of a classical approach involving Euler equations.

In what follows, we restrict our attention to state spaces X that are convex subsets of $\mathbb{R}^n$.

[F^{diff}] F is differentiable on the interior of A.

THEOREM **11** *Assume (D), (T^U), (T^L), (T^{conv}), (F), (F^{conc}), and (F^{diff}). Then, if $x \in$ int X and the optimal policy $G^*(x) \in$ int $\Gamma(x)$, then v^* is differentiable at x, with derivatives given by*

$$v_i^*(x) = F_{1i}(x, G^*(x)) \tag{20}$$

[The notation v_i^* means the partial derivative with respect to the ith component of x, while for F, the notation F_{ji} stands for the derivative of the ith component of the jth vector entry in F, for $j = 1, 2$.]

Proof. The following lemma, which I will not prove, is at the heart of this result.

LEMMA **2** *Let $X \subseteq I\!R^n$ be a convex set, let $v : X \to I\!R$ be concave, let $x \in$ int X, and let D be an open neighborhood of x contained in X. If there is a differentiable function $w : D \to I\!R$ with $w(x) = v(x)$, and $w(y) \leq v(y)$ for all $y \in D$, then v is differentiable at x, and*

$$v_i(x) = w_i(x).$$

The proof is omitted. It uses natural gradient properties of concave functions. See the diagram in class for an intuitive explanation. See Stokey and Lucas [1989, p.84] for a proof (they assume the concavity of w but this is not needed).

Now let us return to the main proof, armed with Lemma 2. Since, by assumption, $G^*(x) \in$ int $\Gamma(x)$ and Γ is a continuous correspondence, $G^*(x)$ continues to lie in $\Gamma(y)$ for all y in some open neighborhood D of x [draw a diagram to understand this, then check it formally]. Define a function $w : D \to I\!R$ by

$$w(y) \equiv F(y, G^*(x)) + \beta v^*(G^*(x)).$$

Because of (F^{diff}) it follows that w is differentiable on D. Moreover, because $G^*(x) \in \Gamma(y)$ for all $y \in D$,

$$w(y) \leq \max_{y' \in \Gamma(y)} F(y, y') + \beta v^*(y') = v^*(y)$$

for all $y \in D$. But these properties establish the conditions of Lemma 2, and show that v^* must be differentiable at x. Moreover, by construction of the function w in this proof and Lemma 2, (20) must hold. ■

8 Some Applications

8.1 The Euler-Ramsey Equation for Differentiable Models

In this subsection, we maintain all the assumptions needed to establish the differentiability theorem (Theorem 11).

Recall the maximization problem implied by the functional equation:

$$\max_{y \in \Gamma(x)} F(x, y) + \beta v^*(y)$$

for each $x \in X$.

Suppose that for each x, the (unique) solution y to this problem is interior; i.e., $y \in$ int $\Gamma(x)$. Then the first-order conditions of this maximization problem yield the following set of necessary conditions:

$$F_{2i}(x, y) + \beta v_i^*(y) = 0$$

for each i. Using Theorem 11, we may deduce that

$$F_{2i}(x, y) + \beta F_{1i}(y, G^*(y)) = 0$$

for each i. This yields the *Euler-Ramsey equations* along any interior optimal program. If an optimal program $\mathbf{x}^*$ from x is interior, i.e., if $x^*_{t+1} \in \text{ int } \Gamma(x^*_t)$ for all t, then

$$F_{2i}(x^*_t, x^*_{t+1}) + \beta F_{1i}(x^*_{t+1}, x^*_{t+2}) = 0 \tag{21}$$

for each i and every t. This is a second-order difference equation characterizing the necessary condition for an optimum.

This classical approach gives rise to a line of argument that merits some discussion. Observe that the difference equation above is pinned down partially by the boundary conditions defining the initial state, but this is not enough to determine the fate of the equation. The remaining restrictions are supplied by the so-called *transversality conditions*. This is something that only comes seriously into play when an infinite time horizon is involved. Finite horizon models where certain terminal states must be achieved supply the transversality condition in this fashion, by determining initial as well as final conditions. So, for instance, in convex finite horizon models, every interior optimum must satisfy the Euler-Ramsey equations, and every interior feasible program satisfying the Euler-Ramsey equations must yield an optimum. This is not true in infinite horizon models. To understand this fully, let us start by considering an example.

Example 4. Consider the one-person, one-sector optimal growth model. In each period, x_t is to be interpreted as (gross) output, which is divided into consumption c_t and new capital stock k_t:

$$x_t = c_t + k_t.$$

Utility is achieved from consumption, according to some function u that is increasing, continuous, strictly concave, and differentiable, and satisfying the end-point restrictions $u'(0) = \infty$ and $u'(\infty) = 0$. Let the discount factor β lie between 0 and 1.

Let g be the production function of fresh output: assume that g is increasing, continuous, strictly concave, and differentiable, satisfying the end-point restrictions $g'(0) = \infty$ and $g'(\infty) = 0$. Gross output is the sum of current output and the depreciated past capital stock (modulo free disposal):

$$x_{t+1} \leq f(k_t) \equiv g(k_t) + (1-d)k_t,$$

where d is the depreciation rate. It is pretty obvious that for all $(x, y) \in A$,

$$F(x, y) = u(x - f^{-1}(y)).$$

Exercise. Verify that all the assumptions made so far hold for this model, and by applying the Euler Ramsey equations (21), show that a necessary condition for any interior program $\{x^*_t, k^*_t, c^*_t\}$ to be optimal is

$$u'(c^*_t) = \beta f'(k^*_t) u'(c^*_{t+1}) \tag{22}$$

Now we will examine (22) from a particular initial stock, given by the value $x^* \equiv f(k^*)$, where k^* is the unique solution to the equation

$$f'(k^*) = 1.$$

We are going to construct an interior program which satisfies Euler-Ramsey, but is nevertheless suboptimal. To do this, start by choosing $k_0 \in (k^*, x^*)$, and $c_0 \equiv x^* - k_0$. Then

$f'(k_0) < 1$, so that $\beta f'(k_0) < 1$ as well. It follows from (22) that in the next period, we must choose c_1 such that $u'(c_1) > u'(c_0)$, which implies that $c_1 < c_0$. Also note that by the choice of k_0, we must have $x_1 > x_0 = x^*$ (provided we throw no output away). It follows that $k_1 > k_0$, and we can repeat the same story over and over again. [Note, in particular, that there is no problem with meeting the Euler-Ramsey equations with equality in every period, given the end-point conditions (check this).] This construction gives us an interior feasible program satisfying the Euler-Ramsey equations at all dates. Nevertheless, this program *cannot* be optimal, from direct inspection (why?).

EXERCISE. The argument above also shows that from the initial stock x^*, there is more than one solution to the difference equation defined by the Euler-Ramsey equations. How would you prove this?

The problem, as usual, lies in the tail of the difference equation: the way in which it behaves far enough out on the program. It turns out the intertemporal behavior of the (implicit) valuation of the capital stock is what drives the appropriate tail or transversality condition. The resulting theorem is general enough to be stated in terms of the general model.

THEOREM **12** *Suppose that $X \subseteq I\!R^n_+$, and that $0 \in X$. Suppose, moreover, that F is nondecreasing in its first n arguments, and that the discount factor is strictly less than unity. Then under the assumptions of Theorem 11, an interior feasible program $\mathbf{x}^*$ is optimal if and only if the Euler-Ramsey conditions (21) hold for all dates, and*

$$\lim_{t \to \infty} \beta^t F_1(x_t^*, x_{t+1}^*) \cdot x_t^* = 0 \tag{23}$$

Proof. (Sufficiency) Suppose that an interior program $\mathbf{x}^*$ satisfies (21) and (23). Let $\mathbf{x}$ be any other feasible program from the same initial state.

Note that if a differentiable function H is concave on some finite-dimensional convex domain Z,

$$H(z^*) - H(z) \geq H_z(z^*) \cdot (z^* - z)$$

for all z, z^* in Z with $z^* \in \text{int}Z$. Applying this inequality to the function F and the domain A,

$$F(x_t^*, x_{t+1}^*) - F(x_t, x_{t+1}) \geq F_1(x_t^*, x_{t+1}^*) \cdot (x_t^* - x_t) + F_2(x_t^*, x_{t+1}^*) \cdot (x_{t+1}^* - x_{t+1}),$$

so that

$$\begin{aligned}
D^T &\equiv \sum_{t=0}^{T} \beta^t [F(x_t^*, x_{t+1}^*) - F(x_t, x_{t+1})] \\
&\geq \sum_{t=0}^{T} [\beta^t F_1(x_t^*, x_{t+1}^*) \cdot (x_t^* - x_t) + \beta^t F_2(x_t^*, x_{t+1}^*) \cdot (x_{t+1}^* - x_{t+1})] \\
&= \sum_{t=0}^{T-1} \beta^t \{F_2(x_t^*, x_{t+1}^*) + \beta F_1(x_{t+1}^*, x_{t+2}^*)\} \cdot (x_{t+1}^* - x_{t+1}) + \beta^T F_2(x_T^*, x_{T+1}^*) \cdot (x_{T+1}^* - x_{T+1})
\end{aligned}$$

(using the fact that $x_0^* - x_0 = 0$). Applying now the Euler-Ramsey equations (twice in what follows), and the assumption that $X \subseteq \mathbb{R}_+^n$, we may conclude that

$$\begin{aligned} D^T &\geq \beta^T F_2(x_T^*, x_{T+1}^*) \cdot (x_{T+1}^* - x_{T+1}) \\ &= -\beta^{T+1} F_1(x_{T+1}^*, x_{T+2}^*) \cdot (x_{T+1}^* - x_{T+1}) \\ &\geq -\beta^{T+1} F_1(x_{T+1}^*, x_{T+2}^*) \cdot x_{T+1}^* \end{aligned}$$

It follows, using the transversality condition, that $\lim_{T\to\infty} D^T \geq 0$, and the proof in this direction is complete.

(Necessity) Recall the value function v^*. By Theorem 10, it is concave, and by Theorem 11, it is differentiable. By the familiar inequality for concave functions that we discussed above, we see that for any x and x' with $x \in \text{ int } X$,

$$v^*(x) - v^*(x') \geq v_x^*(x) \cdot (x - x').$$

Putting $x = x_t^*$ and $x' = 0$ in the above inequality, we observe that

$$\begin{aligned} v^*(x_t^*) - v^*(0) &\geq v_x^*(x_t^*) \cdot x_t^* \\ &= F_1(x_t^*, x_{t+1}^*) \cdot x_t^*, \end{aligned}$$

using (20). It follows that

$$\beta^t v^*(x_t^*) - \beta^t v^*(0) \geq \beta^t F_1(x_t^*, x_{t+1}^*) \cdot x_t^* \geq 0, \tag{24}$$

the last inequality following from the assumption that F is nondecreasing in its first n arguments.

Because $\beta < 1$ by assumption, $\beta^t v^*(0) \to 0$ as $t \to \infty$, and by Theorem 5, $\beta^t v^*(x_t^*) \to 0$ as $t \to \infty$. It follows from these observations and (24) that the transversality condition is satisfied. ■

Exercise. In some situations, the transversality condition is automatically satisfied. In this exercise, we will consider one such situation. Consider the optimal growth model introduced in Section 2 of these notes, and suppose that the production function g is *linear* in k, the capital stock. Thus $g(k) = Ak$, where the A term includes all the other terms for population growth and technical progress.

[1] Write down the condition that is equivalent to (D) in this model.

[2] Show that every optimal program from some $x > 0$ *must* be interior, and write down the Euler-Ramsey equations in this case.

[3] Show that any interior program that satisfies the Euler-Ramsey equations must be optimal.

[4] Use the Euler-Ramsey equations to calculate the policy and value functions of this problem.

8.2 Intertemporal Behavior in Aggregative Models

The functional equation approach is also very useful in deriving some intertemporal properties of the state variable in aggregative models. To see this, we will assume, in this section, that $X \subset \mathbb{R}_+$.

Moreover, we will need to put some order structure on the problem. Specifically, suppose that

$[T^I]$ Γ is *isotone*; i.e., if $x' \geq x$ and $y' \leq y$, and $y \in \Gamma(x)$, then $y' \in \Gamma(x')$.

The other assumptions that we retain in this section are (D), (F) and (T^U). Under these assumptions, it will be recalled, an optimal program exists from every initial stock. It will also be recalled (see the proof of Theorem 8) that under these assumptions, v^* is uppersemicontinuous on X.

Exercise. Show that under these conditions, the optimal policy correspondence derived from solving the functional equation is nonempty-valued.

It will be convenient to start this section with a digression that introduces the idea of supermodularity.

Consider a function $H(x, y)$ defined on the domain $A = \{(x, y) : y \in \Gamma(x)\}$. Say that H is *supermodular* if for all pairs (x_1, y_1) and (x_2, y_2), such that $x_1 > x_2$ and $y_1 > y_2$, and such that $(x_i, y_j) \in A$ for all i and all j, we have

$$H(x_1, y_1) + H(x_2, y_2) > H(x_1, y_2) + H(x_2, y_1). \tag{25}$$

If the opposite inequality holds for all such pairs, we will say that H is *submodular*.[3]

Now consider the problem

$$\max_{y \in \Gamma(x)} H(x, y).$$

Suppose that H is an uppersemicontinuous function. Then, by the same argument as in the exercise above, the set of solutions $G(x)$ is nonempty for each $x \in X$. We are interested in establishing the following result:

Theorem **13** *Suppose that $x_1 > x_2$, and $z_i \in G(x_i)$ for $i = 1, 2$. Then if H is supermodular, $z_1 \geq z_2$.*

Proof. Suppose that the result is false for some (x_1, x_2, z_1, z_2) satisfying the conditions of the theorem. Then it must be the case that $z_1 < z_2$. Define $y_1 \equiv z_2$ and $y_2 \equiv z_1$. Then $y_1 > y_2$.

Note that $y_1 \in \Gamma(x_1)$, because $y_1 = z_2 \in \Gamma(x_2)$, and $x_1 > x_2$, and assumption (T^I) holds. Moreover, $y_2 \in \Gamma(x_2)$, because $y_2 < y_1$, and $y_1 \in \Gamma(x_2)$.

Thus supermodularity applies to the collection $(x_1, y_1; x_2, y_2)$, and (25) holds.

On the other hand, note that because $y_i \in \Gamma(x_i)$ and y_j is optimal for x_i, for each $j \neq i$ and $i = 1, 2$,

$$H(x_i, y_j) \geq H(x_i, y_i).$$

[3] Strictly speaking, this would correspond to *strict supermodularity* (or submodularity) of H.

Adding the two implied inequalities, we see that

$$H(x_1, y_1) + H(x_2, y_2) \leq H(x_1, y_2) + H(x_2, y_1),$$

a contradiction to (25). ■

Because of the value function approach, this result for domains in $\mathbb{R}^2$ can be very easily applied to optimal policies in an infinite horizon problem. Specifically,

THEOREM **14** *Consider a dynamic programming problem satisfying the assumptions of this section, and suppose that the one-period payoff function is supermodular on A. Then every selection from the optimal policy correspondence must be a nondecreasing function.*

Proof. The optimal policy correspondence is the set of the solutions, for each x, to the problem

$$\max_{y \in \Gamma(x)} F(x, y) + \beta v^*(y).$$

Define $H(x, y) \equiv F(x, y) + \beta v^*(y)$. Because F is continuous and v^* is uppersemicontinuous, it follows that H is uppersemicontinuous. Moreover (and this is the simple observation that drives everything, so check it well), H as defined is supermodular if and only if F is supermodular (no assumption on v^* is needed). So Theorem 13 applies, and we are done. ■

COROLLARY. **Turnpike Theorem for Aggregative Systems**. *Under the assumptions of the preceding theorem, optimal programs under any optimal policy are monotone, so that if X is compact, they must converge to a steady state.*

Proof. Left as an exercise.

If the payoff function F is twice continuously differentiable, there is an easy check for the supermodularity of F.

THEOREM **15** *Suppose that F is twice continuously differentiable, and that $F_{12}(x, y) > 0$ for all $(x, y) \in A$. Then F is supermodular.*

Proof. [Tech. Note (ignore if you like) In what follows we use the appropriate directional derivatives on the boundary of the set A whenever needed. These can be found simply by taking the unique continuous extensions of the first and second derivatives from the interior of A to A.]

Pick (x_1, y_1) and (x_2, y_2), such that $x_1 > x_2$ and $y_1 > y_2$, and such that $(x_i, y_j) \in A$ for all i and all j. It follows from assumption (T^I) that for all $x \in [x_1, x_2]$ and all $y \in [y_1, y_2]$, $(x, y) \in A$ (check this). Consequently,

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} F_{12}(x, y) dy dx$$

is well-defined, and

$$\begin{aligned} 0 < \int_{x_1}^{x_2} \int_{y_1}^{y_2} F_{12}(x, y) dy dx &= \int_{y_1}^{y_2} [F_2(x_1, y) - F_2(x_2, y)] dy \\ &= \int_{y_1}^{y_2} F_2(x_1, y) dy - \int_{y_1}^{y_2} F_2(x_2, y) dy \\ &= F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1) + F(x_2, y_2), \end{aligned}$$

which verifies supermodularity. ■

Look at the optimal growth interpretation again, with utility a function of consumption. Then

$$F(x, y) = u(x - f^{-1}(y))$$

for all $(x, y) \in A$, where u is the utility function and f is a strictly increasing production function. Assume for the moment that u, f and f^{-1} are twice continuously differentiable, so that we are in the realm of Theorem 15. Then

$$F_1(x, y) = u'(x - f^{-1}(y)),$$

and so

$$F_{12}(x, y) = -u''(x - f^{-1}(y)) f^{-1'}(y)$$

which means that if $u''(c) < 0$ for all c, the implied function F is supermodular. This means that we get monotonicity of all policy functions, as well as of the optimal path, without making any curvature assumptions on the technology.

EXERCISE. Suppose that u and f are not differentiable, but that u and f are continuous and strictly increasing. Prove, by a direct argument that sidesteps Theorem 15, that if u is strictly concave, then the implied function F is supermodular.

We end this section with a remark on submodularity. Let me consider the situation in the restrictive setting where G^*, the optimal policy, is a continuous function. This is implied by the following setting:

F satisfies all assumptions, except possibly (F^{diff}). Γ satisfies all the assumptions made so far. Finally, the optimal policy function is interior: $G^*(x) \in \text{ int } \Gamma(x)$ for all $x > 0$.

Now begin by considering some point $x_1 > 0$. By assumption, $y_1 \equiv G^*(x_1) \in \text{ int } \Gamma(x_1)$. Therefore for $x_2 < x_1$ but sufficiently close to it, $y_1 \in \text{ int } \Gamma(x_2)$.

Let $y_2 \in G^*(x_2)$. I am going to show that in stark contrast to the supermodularity case, $y_2 > y_1$ if F is submodular.

To see this, define, just as before,

$$H(x, y) \equiv F(x, y) + \beta v^*(y)$$

and suppose on the contrary that $y_1 \geq y_2$. Note that y_2 is feasible for x_2, so by (T^I) it is feasible for x_1. [Or another route: y_1 is feasible for x_1, so $y_2 \leq y_1$ must be feasible for x_1 as well.]

And by construction, y_1 is interior feasible under both x_1 and x_2. Therefore, there is a y_1' close to y_1 but bigger than it such that $y_1' \in \Gamma(x_1)$ and $y_1' \in \Gamma(x_2)$. So we are ready to apply submodularity to $\{(x_1, y_1'), (x_2, y_2)\}$:

$$H(x_1, y_1') + H(x_2, y_2) < H(x_1, y_2) + H(x_2, y_1').$$

Taking an appropriate sequence $y_1' \to y_1$ and passing to the limit above,

$$H(x_1, y_1) + H(x_2, y_2) \leq H(x_1, y_2) + H(x_2, y_1).$$

But this inequality contradicts the unique optimality conditions $H(x_i, y_i) > H(x_i, y_j)$ for all i and $j \neq i$.

This finding implies that in submodular cases which admit a steady state, optimal programs must oscillate. The oscillations can be damped, creating convergence to the steady state, or might create perpetual cycles. [Draw a diagram to see this.]

Discounted Dynamic Programming under Uncertainty

9 Preliminaries

The case of uncertainty runs parallel, in large part, to the case of certainty. In the interests of keeping technicalities to a manageable minimum, I am going to consider in these notes only a restricted class of models (the details are, of course, given below).

The main additional feature in the case of uncertainty is the explicit consideration of a space of *actions*, which we shall denote by Y. The state will evolve partially randomly, and partially conditioned by the previous state and the action taken, in a manner to be described below. Given a state $x \in X$, the set of possible actions is captured by a *feasibility correspondence* $\Gamma : X \to Y$.

Throughout these notes, I will maintain the following assumption on the state space and the feasibility correspondence.

[**B.1**] X and Y are subsets of Euclidean space, and for each $x \in X$, $\Gamma(x)$ is a nonempty, compact-valued, continuous correspondence.

You will recognize this as an amalgamation of various earlier assumptions. We restrict ourselves to Euclidean space to avoid serious technical issues; I will point out where the restriction is needed in our discussions. Similarly, the continuity assumption on Γ can be dropped, but not free of charge. I will try to point out below where it is "really" needed.

Just as before, we denote by A the graph of Γ:

$$A \equiv \{(x, y) \in X \times Y : y \in \Gamma(x)\}. \tag{26}$$

A one-period payoff function F is defined on A, and discounting of future payoffs is carried out by a discount factor β. We will assume

[**B.2**] F is continuous and bounded on A, and $\beta \in (0, 1)$.

Again, (B.2) is stronger than we need, but it will help us in getting a good introduction to the problem.

Finally, we need to describe the evolution of states over time. This is captured by the postulate of a probability measure over the state space. We permit dependence of this measure on the past action and the state, so that the probability measure is written as a measure $Q(x, y, .)$ on X, where $(x, y) \in A$.

We can easily reconcile this setup with the certainty model studied earlier. To induce this model, assume that $X = Y$, and that for each $(x, y) \in A$, $Q(x, y, .)$ is the degenerate measure assigning probability one to y. However, for the purpose of easy interpretation, it will be convenient to change a little bit the interpretation of y in the examples to be discussed.

We now need to discuss feasible (and optimal) programs. It turns out that the sequence definition used so far is not very useful any more. This is for the obvious reason that we cannot write down deterministic sequences of states in the stochastic case. We will therefore introduce policies as our fundamental notion, in contrast to the harmony between sequences and policies emphasized earlier.

10 Histories and Policies

Consider any time t. A *t-history* (sometimes called a *partial history* when no particular date is emphasized) is a complete listing of all states and actions taken up to date $t-1$, together with a listing of the state at date t.

Let H_t denote the set of all t-histories. For each $h_t \in H_t$, let us keep track of the last item, the state at date t, by the notation $x(h_t)$. A *policy* π is a sequence of (measurable) mappings $\{\pi_t\}$ such that for each t, $\pi_t : H_t \to Y$, and $\pi_t(h_t) \in \Gamma(x(h_t))$ for all $h_t \in H_t$.

A policy is *Markov* if there exists a map $\pi : X \to Y$ with $\pi(x) \in \Gamma(x)$ for all $x \in X$ such that for every t and every $h_t \in H_t$, $\pi_t(h_t) = \pi(x(h_t))$. I have purposely used the same notation for a policy in general and the function that defines a Markov policy: we shall simply refer to this function, instead of to the cumbersome sequence that is compatible with it, as the Markov policy.

Begin with some initial state and a policy. It is easy to see that this pair of objects, via the measure Q, defines a probability measure on the space of sequences $\{x_t, y_t\}$ (with $(x_t, y_t) \in A$ for all t). Thus A^∞ is the relevant space of sequences. For each sequence $(\mathbf{x}, \mathbf{y}) \equiv \{x_t, y_t\} \in A^\infty$, we may evaluate the infinite sum of discounted payoffs as

$$w(\mathbf{x}, \mathbf{y}) \equiv \sum_{t=0}^{\infty} \beta^t F(x_t, y_t)$$

Consequently, the expected payoff of following the policy π, and starting from $x \in X$, is given by

$$u(\pi, x) \equiv \mathbb{E}_\pi w(\mathbf{x}, \mathbf{y}), \tag{27}$$

where the notation $\mathbb{E}_\pi$ refers, of course, to the mathematical expectation of the function w over the probability measure induced on A^∞ by the policy π.

In complete analogy to the certainty case, this defines a supremum function v^* on X as

$$v^*(x) \equiv \sup_\pi u(\pi, x) \tag{28}$$

for each $x \in X$.

Remark. There is a technical point in the definition of v^* which merits attention. It is that we are allowing the policy to vary with x. At the same time, it should be the case that an optimal policy (if one exists) need not have to vary with the initial state. This point is related to subtle measurability issues that we bypass for now.

A policy π is *optimal* if it attains the supremum in (28) for every $x \in X$.

The objective of stochastic dynamic programming is to develop methods of analysis that will help us to solve the maximization problem implicit in (28).

11 Examples

[1] **Another look at optimal growth.** Consider the standard optimal growth model, with uncertainty in the production function. To be concrete, suppose that the production

function is given by some continuous $f(y, z)$, where y is the capital input in production and z represents the realization of some i.i.d random variable taking values in some compact interval. As before, output x is divided every period into consumption and capital input (y), and there is some continuous utility function defined on consumption, as well as some factor $\beta \in (0, 1)$. Assume that the family of production functions induced by each z satisfies the Inada conditions. Then without loss of generality, we may restrict all stocks to lie in some compact interval, so that the payoff function is bounded. This is both the space X and Y, and $\Gamma(x) = \{y \in Y : 0 \leq y \leq x\}$. Note that any utility function over consumption can be viewed as a special case of a payoff function defined on $X \times Y$.

It remains to understand the transition probability Q. To see this, observe that each choice of y leads, effectively to a probabilistic outcome over the space of next period's x: this is what is induced by the i.i.d. noise z and the production function f. This, then, is a case where Q need not depend on x. Thus far, this description incorporates both the fundamental one-sector growth model under uncertainty (as described by Brock and Mirman [*Journal of Economic Theory* (1976)]), as well as the model of income distribution with bequests studied by Loury [*Econometrica* (1981)]. These models differ technically in the assumptions specifying z.[4]

The dependence of Q on x in this example will be needed if, for instance, the stochastic process describing z is Markov. In that case, the current output is an insufficient description of the state space, because we will need separate information regarding the current realization of the stochastic shock.

Exercise. By suitably redefining the state space in this example, show precisely how the extension described in the last paragraph can be viewed as a special case of our general model.

[2] **Search.** I am looking for a job. In each period, I get a wage offer from a known distribution G. If I accept an offer w in any particular period, I must keep it for life, obtaining utility $u(w)$ in each period. If I do not accept the job, I incur a search cost $c \geq 0$ and draw again from the same distribution in the next period. Future returns are discounted by β.

To cast this in the general framework, let X be the set of pairs, where the first entry is a wage offer, and the second entry is a 0 or 1, where 0 means a current state where search is ongoing, and 1 denotes a current state where search has ended. Let Y be the two point set $\{0, 1\}$, where 0 means a current decision to keep searching, and where 1 means an acceptance decision.

Exercise. [1] There is a natural description of the feasibility correspondence Γ. Any wage offer while received in a state of 0 can be accepted, ending the search, or rejected, continuing the search. Show that $\Gamma(x) = \{0, 1\}$ if $x = (w, 0)$, and $\Gamma(x) = \{1\}$ captures this description.

[2] Show that the probability measure described as follows captures the transition possibilities: $Q(x, y, .)$ assigns probability one to states ending in a 1, using any distribution function over the first component of the states, provided that either the second component of x equals 1, or $y = 1$. Otherwise, the measure $Q(x, y, .)$ assigns probability one to states ending in a 0, using the distribution function G over the first component of the states.

[4] Of course, there are substantial conceptual and interpretational differences as well, but that is another matter.

[3] Write down the function F on A that captures one-period payoffs. [Hint: One-period payoffs in the *formal* description need not really correspond to one-period payoffs in the model.]

There isn't a unique answer to [2] and [3] taken together. Any answer that is consistent with the description of the problem will do.

[3] **A Replacement Problem.** You have bought a machine. It is either a good machine or a bad machine, but you can't observe this directly. A machine produces one widget in any period. The widget is either okay or a lemon, and this you *can* observe. You also know that good machines produce lemons with probability g and bad machines do so with probability b. Assume $0 < g < b < 1$. An okay widget gives you a zero return. A lemon imposes a cost of c on you.

At each date, you decide, after looking at the quality of the produced item, whether to scrap the machine or not. If you do so, you buy a new machine at cost P. The prior probability that a machine is bad is $q \in (0, 1)$.

EXERCISE. Set Example 3 up as a stochastic dynamic program.

12 Continuity of Probability Measures

Let X be a (measurable) space, and let $\mathcal{P}$ be a collection of probability measures on X. How do we understand when one measure is "close" to another?

One can think of all kinds of notions of closeness. Possibly the strongest notion that one can conjure up is the idea of *strong convergence*, or *convergence in the total variation norm* (to be discussed in more detail later). Say that a sequence P^n of probability measures converges *strongly* if

$$\sup_B |P^n(B) - P(B)|$$

converges to zero, where the sup is taken over all (measurable) subsets of X. This is like asking for uniform closeness over the probabilities over all conceivable events. More on this later. For now, one can think of the usual weakening: drop uniform convergence to just pointwise convergence over events. It turns out, however, that this by itself is not very useful, and there is a far more fundamental criticism to worry about.

The easiest way to start thinking about this problem is to imagine that X is a subinterval of the real line and that $\mathcal{P}$ consists of those probability measures which have distribution functions on X. Now one might desire a notion of convergence that also yields, as a special case, the obvious notion of convergence in the deterministic case (for $\{x^n\} \in X$ converging to some point $x \in X$). How would we translate such a notion in probabilistic terms? We can identify x^n with the degenerate distribution function P^n that assigns probability one to x^n, and likewise x can be replaced by the degenerate distribution P on X. But now it should be pretty obvious that the idea "$x^n \to x$" is *not* captured by strong or even pointwise convergence, none of which is occurring along the sequence P^n, as long as $x^n \neq x$. This poses a dilemma: if we want to carry the deterministic model as a special case, we need a weaker notion of convergence.

This motivates the idea of weak convergence. Basically, what it asks for is to drop any requirement on limit points in the support of P which are carrying positive probability under P. For such points no convergence is required. On the rest, ask for pointwise continuity. In the space of distribution functions, say that F^n converges *weakly* to F (written $F^n \Rightarrow F$) if $F^n(x) \to F(x)$ at all points of continuity of $F(x)$. It's called "weak" convergence because it is a very mild definition of continuity. Notice very carefully that while the definitions of strong or pointwise convergence do not require any notion of "closeness" of elements of X, the definition of weak convergence relies heavily on this.

Here is a more formal (though not the standard) definition for probability measures on a general metric space[5] X. For any subset B of X, let ∂B denote its *boundary*; i.e., the set of points that are limit points of sequences both in B and in B^C. Than say that P^n *converges weakly* to P if for every (measurable) set B such that $P(\partial B) = 0$, $P^n(B) \to P(B)$.

The following lemma establishes a basic property of weak convergence:

LEMMA **3** *If $P^n \Rightarrow P$, then for every bounded, continuous function $f : X \to I\!R$,*

$$\int_X f(x)dP^n(x) \to \int_X f(x)dP(x). \tag{29}$$

Remark. In fact this implication is two-way. The expression (29) is actually used as the *definition* of weak convergence in many expositions. We skip the opposite implication here.

Proof. We will need to establish the following

CLAIM. *Let f be a bounded continuous function, taking values between 0 and 1, on some space X. Let P be a probability measure on that space. Then for each $\epsilon > 0$, there is an integer $k \geq 1$ and a family of sets $B_1, B_2, \ldots, B_k$, such that $P(\partial B_i) = 0$ for all i, and*

$$\sum_{i=1}^{k} \frac{i-1}{k} \hat{P}(B_i) \leq \int_X f d\hat{P}(x) \leq \sum_{i=1}^{k} \frac{i-1}{k} \hat{P}(B_i) + \epsilon$$

for all probability measures $\hat{P}$ on X.

Proof of the claim. We first note that if for any a and b with $a < b$, we define $B \equiv \{x \in X : a \leq f(x) < b\}$, then $\partial B \subseteq Z \equiv \{f(x) = a \text{ or } f(x) = b\}$, and Z is a closed set.

EXERCISE. [1] Prove the assertion above. You will have to use the assumption that f is continuous. Provide an example to illustrate why only the subset relationship, and not the equality relationship, can be asserted.

[2] Let $X = I\!R$, and suppose that for some x^*, $f(x) = 1$ for all $x \leq x^*$, while $f(x) = 0$ for all $x > x^*$. Using this (discontinuous) f, find a and b such that the assertion above is false. This is the clue to why continuity is needed in proving the lemma. We will return to this point below.

[5]Of course, this is also a measure space equipped with the Borel σ-algebra.

Pick $\epsilon > 0$, and choose any integer $k \geq 1$ and a real number $\eta > 0$ such that $\eta + \frac{1}{k} < \epsilon$. Take the interval $[0, 1]$, and find k equally spaced points $a_0, a_1, \ldots, a_k$ such that

[i] $a_0 = \delta \in (0, \eta)$,

[ii] $P(f(x) = a_i) = 0$ for $i = 0, 1, \ldots, k$, and

[iii] $a_k - a_0 = 1$.

The only (relatively) tough part is to establish [ii]. To see this, define the set $C(\delta) \equiv \{x : f(x) = \frac{i}{k} + \delta \text{ for some } i\}$. By the exercise above, $C(\delta)$ is a closed set, and it is obvious that for each $\delta \neq \delta'$, $C(\delta) \cap C(\delta') = \emptyset$. Only a countable number of such sets can be assigned positive probability, so this allows us to comfortably choose a δ in the interval given by [i], so that [ii] is satisfied.

Define $B_i \equiv \{x \in X : a_{i-1} \leq f(x) < a_i\}$, for each $i = 1, \ldots, k$. Then using the assertion proved in the exercise and part [ii] above, we see that $P(\partial B_i) = 0$ for each i.

Digression. Now let us see where the continuity of f is needed. Go back to that example in part [2] of the exercise above. Suppose that we are in a situation where x^* is always between two points of the form $\frac{i-1}{k} + \delta$ and $\frac{i}{k} + \delta$, no matter what value δ takes in the (cramped) interval $[0, \eta]$. Check that no matter what δ you choose, the set B_i will always be of the form $(-\infty, x^*]$. It will not "move" as δ moves. Worse still, its boundary, the set $\{x^*\}$, will always be assigned positive probability under P, if P happens to be the distribution that assigns probability one to x^*!

And indeed, the lemma will fail for such an f. Take P to be the distribution just described, and take a sequence of point masses P^n centered on x^n, where $x^n \downarrow x^*$. We know that $P^n \Rightarrow P$. But alas, $\int_X f(x)dP^n(x)$ does *not* converge to $\int_X f(x)dP(x)$.

Now let us continue with the proof of the claim. Because $a_i = \frac{i}{k} + \delta$ for all i, we see that for any probability measure $\hat{P}$ on X,

$$\begin{aligned}\int_X f(x)d\hat{P}(x) &= \sum_{i=1}^{k} \int_{B_i} f d\hat{P}(x) + \int_{X \setminus \cup_i B_i} f d\hat{P}(x) \\ &\geq \sum_{i=1}^{k} \frac{i-1}{k} \hat{P}(B_i).\end{aligned}$$

Likewise,

$$\begin{aligned}\int_X f(x)d\hat{P}(x) &= \sum_{i=1}^{k} \int_{B_i} f d\hat{P}(x) + \int_{X \setminus \cup_i B_i} f d\hat{P}(x) \\ &\leq \sum_{i=1}^{k} \frac{i}{k} \hat{P}(B_i) + \eta \\ &= \sum_{i=1}^{k} \frac{i-1}{k} \hat{P}(B_i) + \frac{1}{k} \sum_{i=1}^{k} \hat{P}(B_i) + \eta \\ &\leq \sum_{i=1}^{k} \frac{i-1}{k} \hat{P}(B_i) + \frac{1}{k} + \eta\end{aligned}$$

$$\leq \sum_{i=1}^{k} \frac{i-1}{k} \hat{P}(B_i) + \epsilon.$$

This completes the proof of the claim.

Now we return to the proof of the lemma. Fix any $\epsilon > 0$. let k and the sets $B_1, \ldots, B_k$ be given by the claim. We see that for each n,

$$\sum_{i=1}^{k} \frac{i-1}{k} P^n(B_i) \leq \int_X f dP^n(x) \leq \sum_{i=1}^{k} \frac{i-1}{k} P^n(B_i) + \epsilon.$$

Passing to the limit in the expression above and using weak convergence,

$$\sum_{i=1}^{k} \frac{i-1}{k} P(B_i) \leq \liminf_n \int_X f dP^n(x) \leq \limsup_n \int_X f dP^n(x) \leq \sum_{i=1}^{k} \frac{i-1}{k} (B_i) + \epsilon.$$

Using the claim and the inequalities above we may conclude that

$$\int_X f dP(x) - \epsilon \leq \liminf_n \int_X f dP^n(x) \leq \limsup_n \int_X f dP^n(x) \leq \int_X f dP(x) + \epsilon.$$

But since ϵ is arbitrary, we may infer that

$$\liminf_n \int_X f dP^n(x) = \limsup_n \int_X f dP^n(x) = \int_X f dP(x),$$

which completes the proof of the lemma. ■

We shall make the following assumption on the conditional probability Q:

[Q] Suppose that a sequence $\{x^n, y^n\}$ in A converges to $(x, y) \in A$. Then $Q(x^n, y^n, .) \Rightarrow Q(x, y, .)$.

13 The Functional Equation

The basic functional equation of stochastic dynamic programming is, as you may have guessed, analogous to the functional equation studied earlier. A function $v : X \to I\!R$ will be said to satisfy the functional equation if for all $x \in X$,

$$v(x) = \sup_{y \in \Gamma(x)} F(x, y) + \beta \int_X v(x') Q(x, y, dx'). \tag{30}$$

Here is the route to be followed. First we show that a bounded continuous solution to the functional equation *must* be the supremum function. Second, we show that such a solution exists. The method of proof for this will show that the solution is unique, though the first result already tells us that this must be the case.

Technical Remark. We could continue to follow the path set out earlier in the deterministic case. However, with the assumptions already made, we can move much quicker than that. On the other hand, without these assumptions we would not, in general, be able to prove that v^*, the supremum function, is measurable: see Blackwell's example in Stokey and Lucas [1989, 253–254].

THEOREM **16** *Make assumptions (B.1), (B.2) and (Q). Then if v is a bounded, continuous solution to the functional equation (30), it must be the supremum function v^*.*

Moreover, there exists a policy $\pi^*(x)$ that solves (30) for each $x \in X$, and the Markov policy corresponding to this must be an optimal policy for the stochastic dynamic programming problem.

Proof. The proof of this theorem will require some background results from probability theory, which we shall develop as we go along.

Recall the definition of π and $u(\pi, x)$ from two sections ago. There is another way of writing out $u(\pi, x)$ that will be useful for our present concerns. To do so, we will first derive, for any plan and any initial state, a sequence of conditional probability measures over partial histories. Fix a policy π, and meditate on the conditional probability $Q(x, \pi_0(x), .)$. It is a probability over tomorrow's states. But a trivial reinterpretation allows us to also think of this as a probability measure over the set of partial histories H_1, where the first term of the history will be restricted, perforce, to equal x. Let's call this reinterpretation $Q^1(x, \pi, .)$ where for any (measurable) set B_1 in H_1 of the form $B_1 = \{x\} \times \{\pi_0(x)\} \times B$ for some $B \subseteq X$,

$$Q^1(x, \pi, B_1) = \int_B Q(x, \pi_0(x), dx').$$

(and $Q^1(x, \pi, B_1) = 0$ for all events not satisfying the above restriction).

Recursively, suppose that we have a probability measure already defined over the set of partial histories H_t, starting from x. Call this $Q^t(x, \pi, .)$. We can easily define the measure for the next period by first defining it on "product sets" of the form $B_{t+1} = D_t \times B$, where D_t is a subset of $H_t \times Y$ (restricted by the requirements that only x be in the first term, and that $y = \pi_t(h_t)$ for any $(h_t, y) \in B_t$), and B is a subset of X. Then

$$Q^{t+1}(x, \pi, B_{t+1}) = \int_B \int_{B_t} Q^t(x, \pi, dh_t).Q(x(h_t), \pi_t(h_t), dx')$$

where B_t is the restriction of D_t to H_t, describes the conditional probability on $(t+1)$-histories on these "cylinder" sets. For other cylinder sets B_{t+1} where these initial restrictions are not satisfied, define $Q^{t+1}(x, \pi, B_{t+1}) = 0$. By well-known arguments, we can then extend this probability uniquely to arbitrary sets in the product space.

Now we can write another representation of $u(\pi, x)$. It is just the expected utility taken over all partial histories thus:

$$u(\pi, x) = \lim_{T \to \infty} \sum_{t=0}^{T} \beta^t \int_{H_t} F(x(h_t), \pi_t(h_t)) Q^t(x, \pi, dh_t). \tag{31}$$

Now we begin our main proof by establishing that for any policy π and for each $x \in X$,

$$v(x) \geq u(\pi, x) \tag{32}$$

So fix an initial state $x_0 = x \in X$ and a policy $\pi = \{\pi_t\}$. Because v solves the functional equation,

$$v(x) \quad \geq \quad F(x_0, \pi_0(x_0)) + \beta \int_X v(x_1) Q(x_0, \pi_0(x_0), dx_1)$$

$$
\begin{aligned}
&= F(x_0,\pi_0(x_0)) + \beta \int_{H_1} v(x(h_1)) Q^1(x_0,\pi,dh_1) \\
&= F(x_0,\pi_0(x_0)) + \beta \int_{H_1} \sup_{y\in\Gamma(x(h_1))} [F(x(h_1),y) + \beta \int_X v(x_2) Q(x(h_1),y,dx_2)] Q^1(x_0,\pi,dh_1) \\
&\geq F(x_0,\pi_0(x_0)) + \beta \int_{H_1} F(x(h_1),\pi_1(h_1)) Q^1(x_0,\pi,dh_1) + \beta^2 \int_{H_2} v(x(h_2)) Q^2(x_0,\pi,dh_2) \\
&\dots \\
&\geq \sum_{t=0}^{T} \beta^t \int_{H_t} F(x(h_t),\pi_t(h_t)) Q^t(x_0,\pi,dh_t) + \beta^{T+1} \int_{H_{T+1}} v(x(h_{T+1})) Q^{T+1}(x_0,\pi,dh_{T+1}),
\end{aligned}
$$

where the various steps in the derivation above draw heavily on the way the family $\{Q^t\}$ has been constructed. Passing to the limit in the above expression and using (31) as well as (B.2), we obtain (32).

To complete the proof it will suffice to exhibit a Markov policy π^* such that for every $x \in X$,

$$v(x) = u(\pi^*, x) \tag{33}$$

To prove this we need to use Lemma 3. Look at the function

$$H(x,y) \equiv F(x,y) + \beta \int_X v(x') Q(x,y,dx').$$

By Lemma 3 and (B.2), H is a continuous function on A. Define

$$G(x) \equiv \{y \in \Gamma(x) : H(x,y) \geq H(x,y') \text{ for all } y' \in \Gamma(x)\}.$$

By (B.1) and the maximum theorem, it follows that G is a nonempty, compact-valued, uhc correspondence. Take any selection from this correspondence; call it π^*.

Technical Note. The problem is to guarantee the measurability of π^*. This is where we use the assumptions that X and Y are Euclidean space. Under these assumptions and given that G is uhc, there exists a measurable selection from G (see Stokey and Lucas [1989, p.184] for a precise statement of this result).

Now consider this plan. The choice $\pi^*(x)$ attains the sup each time in (30), so that we can go through the same chain of reasoning as above, *with equalities everywhere*. It follows that (33) holds, and the proof is complete. ■

This result is the unimprovability theorem in full force. It was implicit in what we did earlier for the deterministic case. Let us take a moment to understand the connection. Recall that in Theorem 6, a feasible program which solved the functional equation *under the supremum function* was optimal. But we also know from Theorem 2 that any solution to the functional equation (satisfying a tail condition) must be the supremum function. In a sense, the present result is a combination of these two observations for the stochastic case. It says that any policy which is unimprovable (in one step), not with respect to the supremum function but with respect to the "value function" which it *itself* generates, must be the optimal policy. As a byproduct we also obtain the result that under the conditions of this theorem, the optimal policy can be taken to be Markov.

EXERCISE. How did this theorem manage to get around the problem raised by Example 1 in Section 3?

Our next step complements Theorem 16. We will show that under the maintained assumptions, there is indeed a continuous, bounded solution to the functional equation.

We begin with a digression that explains the contraction mapping theorem, due to Banach. Let X be a metric space with a distance function ρ. Say that a sequence $\{x_n\}$ is *Cauchy* if for every $\epsilon > 0$, there exists an integer N such that if n, m are at least as great as N,

$$\rho(x^n, x^m) < \epsilon.$$

EXERCISE. [1] Understand that this requirement is stronger than the statement: $\rho(x^n, x^{n+1}) \to 0$. Prove that all convergent sequences must be Cauchy.

[2] Give an example of a sequence that satisfies the weaker statement in part (1), and diverges.

A metric space is *complete* if every Cauchy sequence converges to some point in that space.

Examples of spaces that are *not* complete are the open interval $(0, 1)$, or (more subtly), the space of bounded continuous functions on $[0, 1]$ when equipped with any metric induced by pointwise convergence.

All Euclidean spaces are complete.

The following lemma is basic.

LEMMA **4** *Let X be a metric space with metric d, and let $\mathcal{C}$ be the space of all bounded, continuous, real valued functions on X. Define on $\mathcal{C}$ a distance function by*

$$\rho(f, g) \equiv \sup_{x \in X} |f(x) - g(x)|$$

for all f, g in $\mathcal{C}$. Then equipped with this metric, $\mathcal{C}$ is complete.

Proof. We must show that every Cauchy sequence in $\mathcal{C}$ converges to some element of $\mathcal{C}$. To this end, let f^n be a Cauchy sequence in $\mathcal{C}$. Then it should be obvious that for each $x \in x$, the sequence of *numbers* $\{f^n(x)\}$ is Cauchy. Since the space of real numbers is complete, there is a number — call it $f(x)$ — such that $f^n(x) \to f(x)$. This collection of numbers, one for each x, defines a function $f : X \to \mathbb{R}$. We must show that $f \in \mathcal{C}$, and that f^n converges to f under the metric ρ.

Begin with the second implication. Fix any $\epsilon > 0$. Then there exists an integer N such that if n, m are at least as great as N, $\rho(f^n, f^m) < \epsilon/2$. Now, for any $x \in X$,

$$\begin{aligned} |f^n(x) - f(x)| &\leq |f^n(x) - f^m(x)| + |f^m(x) - f(x)| \\ &\leq \rho(f^n, f^m) + |f^m(x) - f(x)| \\ &\leq \epsilon/2 + |f^m(x) - f(x)|. \end{aligned}$$

Now choose m large enough so that $|f^m(x) - f(x)| < \epsilon/2$. Then the above expression shows that

$$|f^n(x) - f(x)| < \epsilon$$

for all $n \geq N$, and note that this is independent of x. Consequently,

$$\rho(f^n, f) \to 0,$$

which establishes the second feature.

It remains to show that f is a continuous, bounded function. Boundedness is obvious (but why?). To prove continuity, fix any $x \in X$. We must show that for every $\epsilon > 0$, there exists $\delta > 0$ such that if $d(x, x') < \delta$, then $|f(x) - f(x')| < \epsilon$.

To establish this, start with any $\epsilon > 0$. Choose any integer n such that $\rho(f^n, f) < \epsilon/3$, and choose $\delta > 0$ such that $d(x, x') < \delta$ implies that $|f^n(x) - f^n(x')| < \epsilon/3$ (this last is possible because by assumption, f^n is continuous). Then for any x' such that $d(x, x') < \delta$,

$$|f(x) - f(x')| \leq |f(x) - f^n(x)| + |f^n(x) - f^n(x')| + |f^n(x') - f(x')| < \epsilon,$$

completing the proof. ∎

Our next job is to introduce the idea of a *contraction mapping.* Let (Z, δ) and (W, ρ) be two metric spaces. A function $T : Z \to W$ is a *contraction* if there exists $\lambda \in (0, 1)$ such that

$$\rho(T(z), T(z')) \leq \lambda \delta(z, z')$$

for all z, z' in Z. λ is called the *modulus of contraction.*

Exercise. [1] Note that λ has to be uniform: it cannot depend on z and z'. Give an example of a function on $I\!R$ which is strictly flatter than the 45^0 line throughout, but is *not* a contraction.

[2] Prove that a contraction must be continuous.

The following lemma is also basic, and it is the centerpiece of computational arguments with value functions.

Lemma **5 Banach Fixed Point Theorem.** *Let (Z, δ) be a complete metric space, and suppose that $T : Z \to Z$ is a contraction. Then there exists a unique point $z^* \in Z$ such that $T(z^*) = z^*$. Moreover, starting from any initial $z \in Z$, the iterates $T^{(n)}(z)$ converge to z^* at a geometric rate determined by the modulus of contraction.*

Proof. Begin with any $z \in Z$, and consider the sequence in Z defined recursively by $z_0 \equiv z$, and $z_{n+1} \equiv T(z_n)$ for all $n \geq 0$. Because T is a contraction (of modulus λ, say),

$$\rho(z_{n+2}, z_{n+1}) \leq \lambda \rho(z_{n+1}, z_n) \tag{34}$$

which implies, in particular, that the sequence $\{z_n\}$ is Cauchy (prove this). By completeness, there exists z^* such that $z_n \to z^*$ as $n \to \infty$. By part [2] of the exercise above, T is continuous, which implies that $T(z^*) = z^*$.

To prove the uniqueness of the fixed point z^*, suppose on the contrary that there is another fixed point $z' \neq z^*$. Then $T(z^*) = z^*$, and $T(z') = z'$. Therefore $\rho(T(z^*), T(z')) = \rho(z^*, z') > 0$, which contradicts the fact that T is a contraction.

The expression (34) also reveals the geometric convergence (at rate λ) of the iterates $T^{(n)})(z)$ to (the same) z^* from any initial point z. This completes the proof. ■

We are now in a position to state and prove the second main result of this section. Together with Theorem 16, this proves that an optimal policy always exists, giving rise to a bounded, continuous value function, and that the optimum is achieved by following some Markov policy.

THEOREM **17** *assume (B.1), (B.2), and (Q). Then there exists a unique bounded and continuous solution v^* to the functional equation.*

Let $\mathcal{C}$ be the space of all bounded continuous functions on X. Define the operator $T : X \to X$ by

$$Tv(x) \equiv \max_{y\in\Gamma(x)} F(x,y) + \beta \int_X v(x')Q(x,y,dx') \tag{35}$$

for all $x \in X$, for each $v \in \mathcal{C}$. Then the iterates $T^{(n)}(v)$ converge at a geometric rate (of β) to v^* in the sup norm.

Proof. Let T be given by (35). We claim that T is a contraction of modulus β. To see this, pick $x \in X$ and $v \in \mathcal{C}$. By Lemma 3, the maximand on the right-hand side of (35) is a continuous function of y. Let y^* solve the problem.

Now pick any $\hat{v} \in \mathcal{C}$. Then

$$Tv(x) = F(x,y^*) + \beta \int_X v(x')Q(x,y^*,dx'),$$

while

$$T\hat{v}(x) \geq F(x,y^*) + \beta \int_X \hat{v}(x')Q(x,y^*,dx').$$

Combining these two expressions, we see that

$$\begin{aligned} Tv(x) - T\hat{v}(x) &\leq \beta \int_X \{v(x') - \hat{v}(x')\}Q(x,y^*,dx') \\ &\leq \beta\rho(v,\hat{v}). \end{aligned}$$

By similar reasoning,

$$T\hat{v}(x) - Tv(x) \leq \beta\rho(v,\hat{v}),$$

so that

$$|T\hat{v}(x) - Tv(x)| \leq \beta\rho(v,\hat{v}).$$

Because the RHS of the above expression is independent of x, we may conclude that

$$\rho(Tv, T\hat{v}) \leq \beta\rho(v,\hat{v}),$$

which shows that T is a contraction.

A direct application of the Banach fixed point theorem (Lemma 5) completes the proof. ■

14 Extension to Unbounded Returns

The results of the previous section can be pretty easily extended to the case of unbounded returns, under some assumptions. This extension would allow us to handle, for instance, the theory of optimal growth with an unbounded state space and a utility function that is constant-elasticity.

To this end, let $S(x, y)$ denote the support of next period's state, i.e., supp $Q(x, y, .)$. Now, for any $x \in X$, define $S^0(x) \equiv \{x\}$, and recursively,

$$S^{t+1}(x) \equiv \{x'' \in X : \text{ there is } x' \in S^t(x) \text{ and } y' \in \Gamma(x') \text{ with } x'' \in S(x', y')\}.$$

Next, we define a sequence of maximal one-period utilities starting from x by

$$F^t(x) \equiv \sup_{x' \in S^t(x),\ y' \in \Gamma(x')} F(x', y')$$

for all $t \geq 0$. We will weaken assumption (B.2) to

[B.2′] The payoff function F is continuous and bounded below on A, $\beta \in (0, 1)$, and

$$M(x) \equiv \sum_{t=0}^{\infty} \beta^t F^t(x) < \infty \tag{36}$$

for all $x \in X$.

Without loss of generality (why?), let 0 be the lower bound on one-period payoffs assumed in [B.2′]. Let $\mathcal{C}$ be the space of all continuous functions on X, with the property that $0 \leq v(x) \leq M(x)$ for any $v \in \mathcal{C}$. Note that this is the appropriate space for finding value functions, because it must be the case that for any initial state x and any policy π,

$$u(\pi, x) \leq M(x)$$

(check this). Define a new distance between any two elements of $\mathcal{C}$, $\rho(v, v')$, by

$$\rho(v, v') \equiv \frac{1}{\max\{1, M(x)\}} |v(x) - v'(x)|.$$

Lemma **6** *Under assumption [B.2′], $(\mathcal{C}, \rho)$ is a complete metric space.*

Proof. The first part of the proof is the following step:

Exercise. Show that ρ is indeed a metric on $\mathcal{C}$.

For the rest of the proof, we will need to modify the proof of Lemma 4. This is also left as an exercise.

Exercise. Take a Cauchy sequence $\{v^n\}$ in $\mathcal{C}$. Note that it must converge pointwise: let v denote its pointwise limit. Show that (i) convergence to v is also in the metric ρ, and (ii) that v is a continuous function on X with $0 \leq v(x) \leq M(x)$ for all x. Observe that in your proof, you nowhere need the continuity of the function M.

These two exercises prove the lemma. ■

Theorem **18** *Assume [B.1], [B.2′], and [Q]. Then any solution to the functional equation that comes from the class $\mathcal{C}$ is the supremum function, and the associated policy is optimal. Moreover, there exists a (unique) solution to the functional equation in the class $\mathcal{C}$. This solution is a fixed point of the operator defined in (35), and the iterates of the operator starting from any function in $\mathcal{C}$ converge geometrically, just as before.*

Proof. This theorem is clearly a mixture of Theorems 16 and 7, and the proof runs very close to the corresponding proofs in the two theorems. Some steps require change, of course. The first step to change is the last part of the proof of (32). We need to show that for any initial x and any policy π,

$$\lim_{T\to\infty} \beta^T \int_{H_T} v(x(h_T))Q^T(x,\pi,dh_T) = 0.$$

[Earlier, this was trivial given [B.2].] Note that for any x_0 and any T-history $h_T \in H_T$ starting from x,

$$x(h_T) \in S^T(x),$$

so that

$$v(x(h_T)) \le M(x(h_T)) \le \sum_{t=T}^{\infty} \beta^{t-T} F^t(x).$$

[Question: Why is the second inequality true in the expression above?] Consequently,

$$\beta^T \int_{H_T} v(x(h_T))Q^T(x,\pi,dh_T) \le \sum_{t=T}^{\infty} \beta^t F^t(x)$$

and the right-hand side of the expression above must converge to zero as $T \to \infty$, given [B.2′].

This proves the first part of the theorem. To prove the analogue of Theorem 7, it suffices to show that the operator defined in (35) is a contraction. To do so, follow *exactly* the argument in the proof of Theorem 7, taking care to normalize by $\max\{1, M(x)\}$ in the relevant steps. The details are left as an exercise. ■

15 Applications

15.1 Optimal Search

Reconsider example [2] in Section 10. Let G be the i.i.d. distribution of wage offers. Assume that it is continuous. Recall our definition of a state: it is $(w,0)$ if search is ongoing and $(w,1)$ if search has been completed. There is a nontrivial action only in the former type of state: $y = 1$ for "accept offer" and $y = 0$ for "continue searching". From the exercise in that section (and recalling that the acceptance of a wage offer is assumed to be binding for life), we may write one-period payoffs as

$$\begin{aligned} F((w,1),y) &= 0 \text{ for all } (w,y) \text{ (search is over)} \\ F((w,0),0) &= -c \text{ for all } w \text{ (search continues)} \\ F((w,0),1) &= \frac{w}{1-\beta} \text{ for all } (w) \text{ (accepts offer)} \end{aligned}$$

Let v^* be the value function. Assumptions (B.1), (B.2) and (Q) are all satisfied, so v^* is the unique (bounded and continuous) solution to the functional equation. Given the specification, we only need to worry about the value function on states of the form $(w, 0)$:

$$v^*(w,0) = \max\{-c + \beta \int v^*(w',0)dG(w'), \frac{w}{1-\beta}\}.$$

It is immediate from the above expression that an optimal policy must take the form of a cut-off: define w^* by the equality

$$\frac{w^*}{1-\beta} \equiv -c + \beta \int v^*(w',0)dG(w'), \tag{37}$$

and accept any offer with wage at least w^* (or strictly greater: it doesn't matter in this case).

Note that we are still far from describing the optimal policy in terms of the parameters of the system (v^* itself appears in the expression above). So let us keep going. It should be clear that

$$v^*(w,0) = \frac{w}{1-\beta}$$

for any $w \geq w^*$. For $w < w^*$, $v^*(w,0)$ is clearly independent of the offer w: call the value a. But then a must solve the equation

$$a = -c + \beta[1 - G(w^*)]\frac{\mathbb{E}(w|w \geq w^*)}{1-\beta} + \beta G(w^*)a.$$

which solves for a as

$$a = \frac{\beta[1 - G(w^*)]\frac{\mathbb{E}(w|w \geq w^*)}{1-\beta} - c}{1 - \beta G(w^*)} \tag{38}$$

Now, it's obvious that the right-hand side of (37) is a, so that combining equations (37) and (38),

$$\frac{w^*}{1-\beta} = \frac{\beta[1 - G(w^*)]\frac{\mathbb{E}(w|w \geq w^*)}{1-\beta} - c}{1 - \beta G(w^*)}.$$

the solutions to which correspond to the optimal policies.

15.2 Replacement

This example is taken from Rust [1995]. Suppose that a durable good is bought, then used for some time, then replaced. The problem is to determine the optimal replacement time. Let x_t be a number at time t which represents the accumulated use of the durable up to time t (such as the odometer reading on a car). At each date, Y consists of two options $\{0, 1\}$ where 0 means "keep the good", and 1 means "throw away the good and buy another at price $p > 0$".

Suppose that in each period, the level of utilization of the good has an exponential distribution (with parameter λ), given exogenously. Then we may write the transition probability

Q as

$$\begin{aligned} Q(x, y, dx') &= 1 - \exp\{-\lambda d(x_{t+1} - x_t)\} \text{ if } y_t = 0 \text{ and } x_{t+1} \geq x_t \\ &= 1 - \exp\{-\lambda dx_{t+1}\} \text{ if } y_t = 1 \text{ and } x_{t+1} \geq 0 \\ &= 0 \text{ otherwise.} \end{aligned}$$

Assume that the per-period cost of using the durable good in state x is given by an increasing, differentiable, continuous function $c(x)$, with $c(0) = 0$. We may then define the payoff function by

$$\begin{aligned} F(x, y) &= -c(x) \text{ if } y = 0 \\ &= -p - c(0) \text{ if } y = 1. \end{aligned}$$

It is obvious that if at any date, the state of the good is such that $c(x) > p + c(0)$, it pays to immediately scrap the good. Therefore, we may truncate all costs above this value by a constant, so that [B.1] and [B.2] are automatically seen to be satisfied.

Exercise. Verify that [Q] is satisfied.

Using the density implied by the transition probability, the functional equation solving for v takes the form

$$v(x) = \max\{-c(x) + \beta \int_x^\infty v(x')\lambda e^{-\lambda(x'-x)}dx', -p - c(0) + \beta \int_x^\infty v(x')\lambda e^{-\lambda x'}dx'\}.$$

By Theorems 16 and 17, the solution describing the optimal value v is bounded and continuous. Also note that it must be nonincreasing in x (check this). Let's see what kind of extra information the functional equation gives us. Of the two terms on the RHS, the second is evidently a constant (once v is known), while the first changes with the state x. A little inspection will convince you that the first term must steadily *decline* in x, starting (when $x = 0$) at a value higher than the second term, and ultimately (for large x) declining below the value of the second term ($c(x) > p + c(0)$ is sufficient). There must be some unique value x^* such that the two terms are equal, and the optimal policy (modulo indifference) is given by: replace as soon as the state crosses x^*.

It remains to compute x^* and the value function. To do so, let us differentiate the functional equation in the region $[0, x^*]$:

$$\begin{aligned} v'(x) &= -c'(x) - \beta\lambda v(x) + \lambda\beta \int_x^\infty v(x')\lambda e^{-\lambda(x'-x)}dx' \\ &= -c'(x) + \lambda(1-\beta)v(x) + \lambda c(x). \end{aligned}$$

The boundary condition determining this differential equation is determined as follows. Note that the value $v(x)$ must flatten out once the state x crosses x^*: i.e., $v(x) = v(x^*)$ for all $x \geq x^*$. Consequently, using the functional equation at the point x^*, we see that

$$\begin{aligned} v(x^*) &= -c(x^*) + \beta \int_{x^*}^\infty v(x')\lambda e^{-\lambda(x'-x^*)}dx' \\ &= -c(x^*) + \beta v(x^*), \end{aligned}$$

so that

$$v(x^*) = -\frac{c(x^*)}{1-\beta}. \tag{39}$$

Using the boundary condition (39), we may solve the differential equation. Simple integration shows that

$$v(x) = \max\{\frac{-c(x^*)}{1-\beta}, \frac{-c(x^*)}{1-\beta} + \int_x^{x^*} \frac{c'(z)}{1-\beta}[1-\beta e^{-\lambda(1-\beta)(z-x)}]dz\}. \tag{40}$$

It remains to solve out for the value of x^* in terms of the parameters of the model. Note also that at the point x^*, we are indifferent between continuing, and scrapping and starting again, so that

$$-p + v(0) = -c(x^*) + \beta v(x^*) = \frac{-c(x^*)}{1-\beta},$$

so that, using the formula for $v(0)$ implied by (40),

$$\int_0^{x^*} \frac{c'(z)}{1-\beta}[1-\beta e^{-\lambda(1-\beta)z}]dz = p$$

which uniquely solves for the value x^*.

15.3 Optimal Growth Under Uncertainty

Consider the one-person optimal growth problem with utility function of the form $u(c) = c^{1-\sigma}$, for $\sigma \in (0,1)$, and discount factor $\beta \in (0,1)$. Let x represent output and y the choice of capital for the next period, so that

$$x = y + c.$$

Suppose that the production function at any date is linear, and subject to multiplicative uncertainty, so that

$$x_{t+1} = z_t y_t,$$

where $\{z_t\}$ is a sequence of positive, i.i.d. random variables on some compact support $[a,b]$, with $0 < a < b < \infty$.

Check again that [B.1] and [B.2] are satisfied, but note that [B.2] is not. We are now going to impose conditions on the model that ensure that [B.2′] is satisfied instead, and I'd like you to note that this is analogous to the generalized notions of discounting pursued in the deterministic case.

Recall the definition of $\{F^t\}$ before the statement of [B.2′]. For any x, note that $F^t(x)$ is simply the utility value of the largest income that can be generated in t periods from x, which is just $b^t x$. So

$$F^t(x) = b^{t(1-\sigma)} x^{1-\sigma}$$

for all $x \geq 0$. It follows that

$$M(x) = \sum_{t=0}^{\infty} \beta^t b^{t(1-\sigma)} x^{1-\sigma}$$

so that $M(x) < \infty$ for all x if and only if $\beta b^{1-\sigma} < 1$. The similarity between this and the discounting conditions explored earlier should now be clear.

Let us proceed to analyze the model under this condition. Write down the functional equation that might describe the true value function, if an appropriate solution can be found to it:

$$v(x) = \max_{0 \leq y \leq x} (y-x)^{1-\sigma} + \beta \int_a^b v(zy) dH(z),$$

where H is the measure describing z. Now we have to guess at a solution; needless to say, the method doesn't always work! The point is that when it *does* work, our theorems tell us that we have found the true value function.

Let us try a solution of the form $v(x) = Dx^{1-\sigma}$, for some $D > 0$. Then, if this is correct, it must be the case that

$$Dx^{1-\sigma} = \max_{0 \leq y \leq x} (y-x)^{1-\sigma} + \beta D \int_a^b z^{1-\sigma} y^{1-\sigma} dH(z).$$

Let $k \equiv \int_a^b z^{1-\sigma} dH(z)$. Substituting this in the above expression and maximizing, we obtain the necessary and sufficient first-order condition

$$\frac{y}{x} = \frac{(D\beta k)^{1/\sigma}}{(D\beta k)^{1/\sigma} + 1}.$$

Substituting this into the functional equation and simplifying, we see that

$$Dx^{1-\sigma} = \{\frac{1}{(D\beta k)^{1/\sigma} + 1}\}^{1-\sigma} x^{1-\sigma} + (A\beta k) \frac{(A\beta k)^{\frac{1-\sigma}{\sigma}}}{\{(D\beta k)^{1/\sigma} + 1\}^{1-\sigma}} x^{1-\sigma}$$

and solving this out for D, we obtain

$$D = \{\frac{1}{1 - (\beta k)^{1/\sigma}}\}^{\sigma}.$$

This solution only makes sense if $(\beta k)^{1/\sigma} < 1$. But $k = \int_a^b z^{1-\sigma} dH(z) \leq b^{1-\sigma}$, and condition [B.2′] guarantees that $\beta b^{1-\sigma} < 1$.

To complete the proof that we have found the true value function, one more step is required.

Exercise. Show that $v(x) = Dx^{1-\sigma}$ belongs to the space $\mathcal{C}$; i.e., that $Dx^{1-\sigma} \leq M(x)$, where $M(x)$ has been described above.

16 Markov Processes

The dynamic programming problem discussed in the previous sections gives rise to a Markov process on the state space. This comes about very naturally. Consider an optimal Markov policy π. This defines a process on X by some initial conditions, and the transition probability

$$P(x, .) \equiv Q(x, \pi(x), .) \tag{41}$$

In many problems, it is of interest to examine the process so generated. To study this, it will be useful to begin with some mathematical background on Markov processes, and this is the subject of the section.

16.1 Markov processes on a finite state space

It will be convenient to begin our study by considering situations where the state space X is finite. A transition probability can then be simply thought of as a matrix of nonnegative elements, with each row sum equal to unity (such a matrix is also referred to as a *stochastic matrix*). Labelling the states as $\{1, 2, \ldots, n\}$, the ijth element is

$$\pi_{ij} \equiv P(i, \{j\})$$

and represents the probability that starting from state i, the system moves to state j in a one-period transition. It follows that if we are given some probability $\mu_t = \{\mu_t(1), \ldots, \mu_t(n)\}$ on X, representing the distribution over the states "today", then the situation "tomorrow" is expressed by a probability measure μ_{t+1}, where for every $j \in X$,

$$\mu_{t+1}(j) = \sum_{i+1}^{n} \mu_t(i)\pi_{ij},$$

or putting this more compactly in matrix notation,

$$\mu_{t+1} = \mu_t \Pi, \tag{42}$$

where Π is the stochastic matrix of transition probabilities.

Iterating (42) from any initial distribution μ_0 on X, we see that

$$\mu_t = \mu_0 \Pi^t. \tag{43}$$

So to understand the long-run behavior of μ_t, we must study the properties of the sequence $\{\Pi^t\}$. Note in particular that if we start with a deterministic distribution placing all the weight on state i, the probability distribution on X generated after t periods is just the ith row of Π^t (why?). Therefore, our interest in convergence leads us to questions of the form:

[1] Does each row of Π^t converge as $t \to \infty$?

[2] If the answer to [1] is "yes", then does each row converge to the *same* vector?

Observe that if μ_t converges to μ from some initial condition, then

$$\mu = \mu \Pi. \tag{44}$$

A distribution that satisfies (44) will be called *invariant*.

THEOREM **19** *An invariant distribution exists. The set of invariant distributions is compact and convex.*

Proof. Let $\mathcal{M}$ be the set of all probability distributions on X. Then it is obvious that $\mathcal{M}$ is a nonempty, compact, convex subset of $\mathbb{R}^n$. Define a function $f : \mathcal{M} \to \mathcal{M}$ by

$$f(\mu) = \mu\Pi.$$

Then f is continuous. By Brouwer's fixed point theorem, there exists μ^* such that $f(\mu^*) = \mu^*$, which is just saying that μ^* is an invariant distribution. The compactness and convexity of the set of invariant distributions follows immediately from the observation that (44) defines a compact, convex set of solutions. ■

There are several examples in Stokey and Lucas [1989, Chapter 11] which are worth studying before proceeding further with these notes. They illustrate the various patterns that Markov processes on a finite state space might display, and they set the stage for the results to follow below.

A subset S of X is *nice* if $P(i, S) = 1$ for every $i \in S$. A subset E of X is *ergodic* if it is nice and there isn't a subset of E which is nice.

Now some definitions for states. Call j a *successor* of i if $\pi_{ji}^{(n)} > 0$ for some $n \geq 1$. Say that i is *recurrent* if it is a successor of every state j that is a successor of i. Finally, a state $i \in X$ is *transient* if there is a positive probability of leaving that state and never coming back: i.e., if i has a successor j such that $\pi_{ji}^{(n)} = 0$ for all n.

THEOREM **20** *Let X be a finite set of the form $\{1, 2, \ldots, n\}$ and Π be a transition probability. Then*

1. *X can be partitioned into $M \geq 1$ ergodic sets and a (possibly empty) set of transient states.*
2. *The sequence $\frac{1}{T+1}\sum_{t=0}^{T}\Pi^t$ converges to a stochastic matrix Q. In other words, for any initial μ_0 on X,*
$$\frac{1}{T+1}\sum_{t=0}^{T}\mu_t \to \mu_0 Q.$$
3. *The set of invariant distributions is given by the convex hull of Q.In particular, the entries in Q under each of the columns corresponding to the transient states must be zero.*

Proof. We begin by showing that X possesses at least one recurrent state. Suppose not. Then because state one is transient, there is another state (call it 2) such that $\pi_{12}^{(N)} > 0$ for some N but with $\pi_{21}^{(m)} = 0$ for all m. Because state 2 is transient as well, there is a state (call it 3) such that $\pi_{23}^{(N')} > 0$ for some N' but with $\pi_{32}^{(m)} = 0$ for all m. Indeed, it must also be the case that $\pi_{31}^{(m)} = 0$ for all m (why?). Continuing in this vein, we see that $\pi_{nn} < 1$ and $\pi_{ni} = 0$ for all $i \neq n$, which contradicts the fact that Π is a transition probability.

So there exists a recurrent state. Next we create an equivalence relation using succession and recurrence. Note that if i is recurrent and j is a successor of i, then j must be recurrent and i must be a successor of j. To see this observe that if i is recurrent and j is a successor, then $\pi_{ji}^{(m)} > 0$ for some i, which means that i is a successor of j. To show that j is recurrent as well, let k be any successor of j. Then, of course, k must be a successor of i. Because i is recurrent, it follows that $\pi_{ki}^{(m)} > 0$ for some m. Also, because j is a successor of i, $\pi_{ij}^{(m')} > 0$ for some m'. Consequently, $\pi_{kj}^{(m+m')} \geq \pi_{ki}^{(m)}\pi_{ij}^{(m')} > 0$, so j is also a successor of i.

Hence the set X can be partitioned as follows. Let T be the set of all transient states. Create ergodic sets $E_1, \ldots, E_M$ by assigning two recurrent states to the same set E_i if and only if they are successors of each other. By our earlier argument that there exists at least one recurrent state, $M \geq 1$. Note that once the state enters some E_i, it stays there forever.

Now we show convergence of $A(T) \equiv \frac{1}{T+1}\sum_{t=0}^{T}\Pi^t$. Note that $A(T)$ is itself a stochastic matrix for all T, being an average of stochastic matrices. Clearly, there is a subsequence T_s of T for which $A(T_s)$ converges; call this limit Q. In other words,

$$\lim_{s\to\infty}\frac{1}{T_s+1}\sum_{t=0}^{T_s}\Pi^t = Q.$$

Pre- and post-multiplying by Π, we see that

$$\lim_{s\to\infty}\frac{1}{T_s+1}\sum_{t=1}^{T_s+1}\Pi^t = \Pi Q = Q\Pi.$$

But observe that

$$\frac{1}{T_s+1}\sum_{t=1}^{T_s+1}\Pi^t - \frac{1}{T_s+1}\sum_{t=0}^{T_s}\Pi^t = \frac{1}{T_s+1}[\Pi^{T_s+1} - \Pi^0] \to 0$$

as $s \to \infty$. using this information in the two expressions that precede it,

$$Q = Q\Pi = \Pi Q, \tag{45}$$

so that in particular,

$$Q = Q\Pi^t = \Pi^t Q \tag{46}$$

for all t. Now let A be some other limit point of the sequence $A(T)$. Then, using (46), we see that $Q = QA = AQ$, and with the roles of A and Q reversed, we have that $A = AQ = QA$. Consequently, $A = Q$, which establishes convergence.

Now for the last part. That each row of Q is an invariant distribution follows right away from (45) (why?). Conversely, if μ is an invariant distribution, then

$$\mu = \mu\Pi^t$$

for all t, so that taking combinations with equal weight of the above equation for $t = 0, 1, \ldots, T$,

$$\mu = \mu\frac{1}{T+1}\sum_{t=0}^{T}\Pi^t.$$

Passing to the limit in the above expression, we see that $\mu = \mu Q$, which means that for each j,

$$\mu(j) = \sum_{i=1}^{n} \mu_i q_{ij},$$

so that μ is a convex combination of the rows of Q. ■

Theorem 20 provides us with an interesting way of rewriting the state space, and in this way we can connect the first part of the theorem with the second and third parts. Partition X into the sets $(T; E_1, \ldots, E_M)$, where as above, T is the set of all transient states and E_i, for each i, is an ergodic set. By Theorem 20, this is indeed a partition.

Now write the transition probability in the following way:

$$\Pi = \begin{array}{c|cccccc} & T & E_1 & E_2 & \ldots & E_M \\ \hline T & R_{00} & R_{01} & R_{02} & \ldots & R_{0M} \\ E_1 & 0 & R_{11} & 0 & \ldots & 0 \\ E_2 & 0 & 0 & R_{22} & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ E_M & 0 & 0 & 0 & \ldots & R_{MM} \end{array}$$

where R_{ii}, $i \geq 1$, is a stochastic matrix but R_{00} is not; i.e., some row sum of R_{00} must be strictly less than unity. As before, let $Q = \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T-1} \Pi^t$. Then Q must have the form

$$Q = \begin{array}{c|cccccc} & T & E_1 & E_2 & \ldots & E_M \\ \hline T & 0 & A_1 & A_2 & \ldots & A_M \\ E_1 & 0 & Q_1 & 0 & \ldots & 0 \\ E_2 & 0 & 0 & Q_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ E_M & 0 & 0 & 0 & \ldots & Q_M \end{array}$$

where for each $i \geq 1$, $Q_i \equiv \ lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{T-1} R_{ii}^t$, and where $\{A_j\}$ will be described presently. The important observation is

Theorem **21** *For each ergodic set E_i, the limit matrix Q_i has the property that all its rows are the same.*

Proof. Consider two states j and k, and let a_j and a_k be the limit of the time-averaged probability of being in some given state, say m, starting from j and k respectively.

Let $\Psi(t)$ denote the probability that starting from the state j, the system hits k for the *first* time at date t. Because j and k belong to the same ergodic class, $\sum_{t=1}^{\infty} \Psi(t) = 1$. Denote by $\mu_s^j(m/t)$ the probability of the system being in state m at date s, starting from state j, conditional on the event that k is reached for the first time at date t. Note that if $t \leq s$, $\mu_s^j(m/t)$ is simply equal to $\mu_{s-t}^k(m)$, the probability of being in state m starting from k after $s - t$ dates. Using this information, we see that

$$a_j \quad = \quad \sum_{t=1}^{\infty} \Psi(t) \lim_{N\to\infty} \{ \frac{\sum_{s=0}^{t-1} \mu_s^j(m/t)}{N} + \frac{1}{N} \sum_{s=t}^{N} \mu_s^j(m/t) \}$$

$$
\begin{aligned}
&= \sum_{t=1}^{\infty} \Psi(t) \lim_{N\to\infty} \{ \frac{\sum_{s=0}^{t-1} \mu_s^j(m/t)}{N} + \frac{N-t}{N} \frac{1}{N-t} \sum_{s=0}^{N-t} \mu_s^k(m) \} \\
&= \sum_{t=1}^{\infty} a_k = a_k.
\end{aligned}
$$

this establishes that each row of Q_i must be the same (why?), or equivalently, that each ergodic class can have at most *one* invariant distribution. ■

To complete the description of Q, it will be useful to describe the matrices A_j. Each row i of A_j must be a scalar multiple w_{ij} of the invariant distribution corresponding to the ergodic class E_j, i.e., of the unique row of Q_j. The term w_{ij} thus reflects the eventual probability of moving from the transient state i to the ergodic class E_j. Of course, for each i, $\sum_{j=1}^{M} w_{ij} = 1$.

The ideas in the preceding theorem quickly lead to

Theorem **22** *The Markov process (X, Π) possesses a unique ergodic set if and only if there exists a state i such that for each $j \in X$, $\pi_{ji}^{(m)} > 0$ for some $m \geq 1$. In such a case, Π has a unique invariant distribution μ^*. Equivalently, each row of Q equals μ^*.*

Proof. By virtue of the previous argument it suffices to prove the first part of this theorem. Suppose, first, that there exists i with the property in the statement of the theorem. This means that i is a successor of every $j \in X$, so that i must belong to the intersection of all the ergodic sets. Since ergodic sets are disjoint (Theorem 20), there can be only one such set.

Conversely, suppose that (X, Π) has a unique ergodic set E. Pick any $i \in E$. Then for each $j \in E$, i is a successor of j, so that $\pi_{ji}^{(m)} > 0$ for some m for each such j. Now take $j \notin E$. Then because j is transient, there exists $k \in E$ such that for some m_1, $\pi_{jk}^{(m_1)} > 0$. Also, $\pi_{ki}^{(m_2)} > 0$ for some m_2. So $\pi_{ji}^{(m_1+m_2)} \geq \pi_{jk}^{(m_1)} \pi_{ki}^{(m_2)} > 0$, and we are done. ■

The uniqueness of the ergodic set still does not guarantee convergence of the distributions μ_t from any initial condition (why?). Our last theorem provides necessary and sufficient conditions for such convergence to occur.

Some notation will be useful. For each m and j, let $\epsilon_j^{(m)} \equiv \min_i \pi_{ij}^{(m)}$, and let $\epsilon^{(m)} \equiv \sum_j \epsilon_j^{(m)}$. Define, too, a distance between probability measures on X by $||\mu, \mu'|| \equiv \sum_i |\mu(i) - \mu'(i)|$.

Theorem **23** *(X, Π) has a unique ergodic set with no cyclically moving subsets if and only if $\epsilon^{(m)} > 0$ for some $m \geq 1$. In this case, starting from any μ_0 on X, $\mu_t = \mu_0 \Pi^t$ converges to a unique limit μ^*, and $||\mu_t, \mu^*|| \to 0$ at a geometric rate that is independent of μ_0.*

Proof. First we assume that $\epsilon \equiv \epsilon^{(m)} > 0$ for some $m \geq 1$. Consider the m-step Markov process on X with transition probability given by $R \equiv \Pi^m$. Let r_{ij} be a typical element of R. Let $\mathcal{M}$ be the set of all probability measures on X.

Claim. The map $g : \mathcal{M} \to \mathcal{M}$ given by $g(\mu) = \mu R$ is a contraction of modulus $1 - \epsilon$.

To prove this claim, pick μ and μ' in $\mathcal{M}$. Then

$$\begin{aligned}
||\mu R, \mu' R|| &= \sum_{j=1}^{n} |\sum_{i=1}^{n} \mu(i) r_{ij} - \sum_{i=1}^{n} \mu'(i) r_{ij}| \\
&= \sum_{j=1}^{n} |\sum_{i=1}^{n} (\mu(i) - \mu'(i)) r_{ij}| \\
&= \sum_{j=1}^{n} |\sum_{i=1}^{n} (\mu(i) - \mu'(i))(r_{ij} - \epsilon_j^{(m)}) + \sum_{i=1}^{n} (\mu(i) - \mu'(i)) \epsilon_j^{(m)}| \\
&\leq \sum_{j=1}^{n} \{\sum_{i=1}^{n} |\mu(i) - \mu'(i)|(r_{ij} - \epsilon_j^{(m)}) + |\sum_{i=1}^{n} (\mu(i) - \mu'(i))| \epsilon_j^{(m)}\} \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} |\mu(i) - \mu'(i)|(r_{ij} - \epsilon_j^{(m)}) + |\sum_{i=1}^{n} (\mu(i) - \mu'(i))| \sum_{j=1}^{n} \epsilon_j^{(m)} \\
&= \sum_{i=1}^{n} |\mu(i) - \mu'(i)| \sum_{j=1}^{n} (r_{ij} - \epsilon_j^{(m)}) + 0 \\
&= (1 - \epsilon) ||\mu, \mu'||.
\end{aligned}$$

With this claim in hand, we may appeal to the contraction mapping theorem to assert that g has a unique fixed point μ^*, and moreover, that (geometric) convergence of the iterates $g^{(t)}(\mu_0)$ is assured from any initial μ_0. But this is only for the m-step problem. To complete the proof we must show that this is the case for the original problem as well. To do this, take any μ_0 and consider the m subsequences, of m steps each, starting off from the initial conditions $(\mu_0, \mu_1, \ldots \mu_{m-1})$. With the result already established, we know that each of the subsequences converge geometrically to μ^*. The fact that testing the convergence of each of these subsequences is sufficient for establishing the convergence of the entire sequence is left as an exercise.

To complete the proof, we prove the converse. Suppose that there is indeed a unique μ^* such that for all $\mu_0 \in \mathcal{M}$, $\mu_0 \Pi^t \to \mu^*$. Then we know that Π^t converges to the matrix Q with every row given by μ^*. Take any j such that $\mu_j^* > 0$. Then $\pi_{ij}^{(m)} \to \mu_j^*$ as $m \to \infty$ for each i, so that $\epsilon_j^{(m)} = \min_i \pi_{ij}^{(m)} > 0$ for large m. Consequently, $\epsilon^{(m)} > 0$. ■

16.2 Markov processes on an infinite state space

Now return to the case of an arbitrary state space X, not necessarily finite. We will restrict ourselves in this section to a result on the *strong convergence* of the process to a unique ergodic distribution. As one can imagine, there are other results that can be proved under weaker conditions. For instance, in line with what transpired in the previous section, we might be interested in the question of convergence of the time averages of the distributions, instead of the distributions themselves. Or we may be interested in studying the more delicate conditions for weak convergence.

We begin by reviewing the basic definitions of a Markov process for the case of an infinite state space. The fundamental ingredient is a *transition probability* $P(x, .)$, which is a probability measure on the state space X for each $x \in X$. The interpretation is that $P(x, A)$ represents the probability of the system being in the subset A "tomorrow", given that it is in the state x today.[6]

In addition to the transition probability, we need a description of initial conditions to get the system started. This initial description may be a deterministic or stochastic. Saying that there is an initial probability measure μ_0 on X maintains full generality.

We may now iterate the process to get a sequence of probability measures on X. Specifically, if μ_t denotes the probability measure on X at date t, then at date $t+1$

$$\mu_{t+1}(A) = \int_X P(x, A)\mu_t(dx)$$

for every event A. Convince yourself that this integral representation faithfully mimics the sum in the case of a finite state space.

With this iteration in place and given μ_0, the Markov process generates an infinite sequence of probability measures $\{\mu_t\}_{t=0}^{\infty}$. We are interested in the special question:

Do the μ_t's converge strongly to some limit distribution μ^* as $t \to \infty$?

Why is the question special? This is because such a specification cannot hope to capture the notion of convergence in the deterministic case, as we have already seen. Indeed, as we shall see, this notion applies well to transition probabilities that admit densities. While this representation is not general enough to retain the deterministic model as a special case, it is nevertheless general enough to encompass several models of economic interest.

We begin by defining strong convergence precisely. Let $\mathcal{M}$ be the space of all probability measures on X. Define a metric on $\mathcal{M}$ by

$$||\mu, \mu'|| \equiv \sup_A |\mu(A) - \mu'(A)|.$$

for every μ and μ' in $\mathcal{M}$.

Remark. This metric is sometimes referred to as (one generated by) the total variation norm. The reason is that the metric as defined above is a scalar multiple of the total variation metric, defined as

$$||\mu, \mu'||_{TV} \equiv \sup \sum_{i=1}^{N} |\mu(S_i) - \mu'(S_i)|,$$

where the sup above is taken over all finite partitions of S into $(S_1, \ldots, S_N)$. In fact you can check that $||\mu, \mu'||_{TV} = 2||\mu, \mu'||$.

EXERCISE [1] By playing with different scenarios where sequences of continuous density functions converge pointwise to a continuous density function, try and relate strong convergence of probability measures to the pointwise convergence of continuous density functions.

[2] Is the space of all distributions on an infinite (but compact) state space compact in the total variation norm?

[6] The definition of a transition probability also includes other technical restrictions, notably the measurability of $P(x, A)$ in x for each fixed A.

It will be useful to have notation for the m-step transition probability generated by the transition probability P. This is, intuitively, the probability of the system being in the subset A after m periods, starting from some given state x "today". Clearly, this is given precisely by the measure μ_m, starting from the case where μ_0 assigns probability one to x. This measure we will denote in transition probability form as $P^m(x, A)$.

The fundamental condition to be investigated in these notes is

Condition M (Stokey and Lucas [1989]) There exist $\epsilon > 0$ and an integer $M \geq 1$ such that for any event A, either (i) $P^M(x, A) \geq \epsilon$ for all $x \in X$, or (ii) $P^M(x, A^C) \geq \epsilon$ for all $x \in X$.

To appreciate condition M, let's look at a case when it is *not* satisfied. This is our familiar two-state Markov process where $\pi_{ij} = 1$ if and only if $i \neq j$, for $i, j = 1, 2$. Pick a set $A = \{1\}$. Then for any positive integer M, $P^M(x, A) = 1$ either if M is even and $x = 1$, or if M is odd and $x = 2$. Otherwise, $P^M(x, A) = 0$. This means that condition M fails. We see therefore, that the real bite of condition M is in the postulated *uniformity* with which all states hit particular events.

Exercise. Prove that condition M is equivalent to the condition: $\epsilon^{(m)} > 0$ for some $m \geq 1$, in the statement of Theorem 23, provided that the underlying state space is finite.

The main theorem of this section is

Theorem **24** *Under condition M, there exists a unique invariant probability measure $\mu^* \in \mathcal{M}$ such that for any initial μ_0 on X, the generated sequence $\{\mu_t\}$ converges strongly to μ^*.*

Proof. We will follow the finite horizon case exactly. That is, we will show that

1. $\mathcal{M}$ equipped with the total variation metric is a complete metric space.

2. The operator $T^M : \mathcal{M} \to \mathcal{M}$ given by

$$T^M(\mu)(A) \equiv \int_X P^M(x, A)\mu(dx),$$

for all A, is a contraction.

3. Thus T^M has a unique fixed point μ^* and the M-step iterates of any initial probability measure must converge to μ^*.

4. The convergence of the entire sequence of measures, and not just this particular subsequence, can then be established by a subsequence argument identical to that used in the finite horizon case.

All the new stuff is in the first two items. To these we now proceed.

First we establish the completeness of $\mathcal{M}$. To this end, suppose that $\{\mu^n\}$ is a Cauchy sequence in $\mathcal{M}$. then from the definition of the total variation metric, it follows that for each event A, $\mu^n(A)$ is a Cauchy sequence of *numbers*. By the completeness of the real line, $\mu^n(A)$ converges to some $\mu(A)$ for each A. We will show that μ is a probability measure and that μ^n converges strongly to μ.

It is obvious that $\mu(A) \in [0,1]$ for all A, that $\mu(X) = 1$, and that $\mu(\emptyset) = 0$. It remains to prove countable additivity to establish that μ is indeed a probability measure. To this end, let $\{A_i\}$ be a countable collection of disjoint events in X. Then

$$\mu(\cup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mu^n(\cup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \sum_{i=1}^{\infty} \mu^n(A_i) = \sum_{i=1}^{\infty} \lim_{n\to\infty} \mu^n(A_i) = sum_{i=1}^{\infty} \mu(A_i)$$

where the second-last equality follows from the dominated convergence theorem. This proves that μ is indeed a bonafide probability measure. What's left to do is to show that $||\mu^n, \mu|| \to 0$ as $n \to \infty$. Note that because $\{\mu^n\}$ is Cauchy, for all $\epsilon > 0$, there is N such that if $n, m \geq N$,

$$|\mu^n(A) - \mu^m(A)| \leq \epsilon$$

for all sets A. Taking limits in m, it follows that

$$|\mu^n(A) - \mu(A)| \leq \epsilon$$

for all $n \geq N$, and for all A, which implies that $||\mu^n, \mu|| \to 0$ as $n \to \infty$.

This establishes the completeness of $\mathcal{M}$.

Our next task is to show that T^M is a contraction of modulus $1 - \epsilon$. To this end, pick μ and μ' in $\mathcal{M}$. Then there is a "common" part γ and "idiosyncratic" parts μ^1 and μ^2, so that $\mu = \mu^1 + \gamma$, $\mu' = \mu^2 + \gamma$, and μ^1 and μ^2 have disjoint support.

Digression. The technical details of this assertion rely on the Radon-Nikodym Theorem (see Stokey and Lucas [1989, Lemma 7.12]). But one can informally illustrate how the common part is obtained in the case where μ and μ' have densities f and f' on the real line. In this case, simply define $g(x) \equiv \min\{f(x), f'(x)\}$ for each $x \in I\!R$, and integrate this as you would to get a cdf, to arrive at the common part of the measure γ. This measure γ is, of course, not a probability measure. The idiosyncratic residuals μ^1 and μ^2 are then defined by inserting the rest of the probability, event by event, to bring total probability up to μ and μ'. In this way you see that the measures μ^1 and μ^2 are not probability measures either. Now study the values $\mu^1(X)$ and $\mu^2(X)$, which are the values assumed by the measures on the entire state space. Note, first, that these must be equal to each other. Note, also, that their common value must be equal to the supremum difference between the probability measures μ and μ' over any event. To see this, note that for any event A

$$\begin{aligned} |\mu(A) - \mu'(A)| &= |\mu^1(A) - \mu^2(A)| \\ &\leq \max\{\mu^1(A), \mu^2(A)\} \\ &\leq \max\{\mu^1(X), \mu^2(X)\}. \end{aligned}$$

This ends our digression.

Returning to the main argument, we see that

$$\begin{aligned} ||T^M\mu, T^M\mu'|| &= \sup_A |\int P^M(x,A)\mu(dx) - \int P^M(x,A)\mu'(dx)| \\ &= \sup_A |\int P^M(x,A)\mu^1(dx) - \int P^M(x,A)\mu^2(dx)|. \end{aligned}$$

Now consider any event A and its complement A^C. Without loss of generality suppose that $P^M(x, A) \geq \epsilon$ for all $x \in X$. If K denotes the common value of $\mu^1(X)$ and $\mu^2(X)$ (see digression above), then it must be the case that

$$|\int P^M(x, A)\mu^1(dx) - \int P^M(x, A)\mu^2(dx)| \leq (1 - \epsilon)K.$$

Combining these last two observations, and the observation in the digression, we see that

$$||T^M\mu, T^M\mu'|| \leq (1 - \epsilon)||\mu, \mu'||,$$

which completes the proof that T^M is a contraction. ■

Dynamic Games

17 Discounted Repeated Games Under Certainty

17.1 Preliminaries

A *one-shot* or *stage game* is denoted by $G = (\{A_i\}_{i=1}^n, \{f_i\}_{i=1}^n)$, where $N \equiv \{1, 2, \ldots, n\}$ is the set of players, and for each player i, A_i is a set of pure actions[7] and $f_i A_i \to I\!R$ a bounded payoff function (where $A \equiv \prod_{i \in N} A_i$.

Let β be a discount factor common to all players. Denote by G^∞ the game obtained by repeating G infinitely many times, and evaluating overall payoffs by the sum of one-period payoffs, discounted using β.

A *path* (or *punishment*) is given by a sequence $\mathbf{a} \equiv \{a(t)\}_{t=0}^\infty$, where $a(t) \in A$ for all t. Along a path $\mathbf{a}$, the payoff to player i is given by

$$F_i(\mathbf{a}) \equiv \sum_{t=0}^{\infty} \beta^t f_i(a(t)).$$

For $t \geq 1$, a *t-history* is a complete specification of all that has transpired up to and including date $t-1$. Thus a t-history is a vector $h(t) \equiv (a(0), a(1), \ldots, a(t-1))$ of all the actions taken by all the players up to date $t-1$. Clearly, A^t is the set of all t-histories.

A *strategy* (or *policy*) for player i is a specification of an action at date 0, and thereafter an action conditioned on every t-history. Formally, a strategy is given by the sequence of functions $\sigma_i i \equiv \{\sigma_i(t)\}_{t=0}^\infty$, where $\sigma_i(0) \in A_i$, and $\sigma_i(t) : A^t \to A_i$ for all $t \geq 1$. A *strategy profile* is a collection σ of strategies, one for each player. In the obvious way, a strategy profile generates a path $\mathbf{a}(\sigma)$, as well as a path $\mathbf{a}(\sigma, h(t))$ conditional on every t-history. Observe that as defined, the path contains the given t-history in its first t terms.

We will subscript a vector using the notation $-i$ to denote the same vector but with its ith component removed.

[7] There is formally no loss in interpreting A_i to be a set of mixed strategies. But there are conceptual problems with this, as it requires that the strategies themselves (and not their realizations) be observed by all players.

A strategy profile σ is a *Nash equilibrium* if for every player i and every strategy σ'_i,

$$F_i(\mathbf{a}(\sigma)) \geq F_i(\mathbf{a}(\sigma_{-i}, \sigma'_i)).$$

In class, we discussed with the help of an example that a Nash equilibrium may be too broad a concept in the context of a dynamic game. Specifically, we may want σ to induce a Nash equilibrium following every t-history. Formally, σ is a *subgame perfect Nash equilibrium* (to be written SGPE) if it is a Nash equilibrium, and for every t-history $h(t)$, every player i, and every alternative strategy σ'_i,

$$F_i(\mathbf{a}(\sigma, h(t))) \geq F_i(\mathbf{a}(\sigma_{-i}, \sigma'_i, h(t))).$$

Exercise. [1] Formalize the notion (you will need new notation) that SGPE means that σ should "induce a Nash equilibrium following every t-history", by defining what the phrase in quotes means, and showing equivalence with the definition here.

[2] For a game that is repeated just *once*, provide an example of a Nash equilibrium that is not subgame perfect.

[3] Prove that if a stage game has a unique Nash equilibrium, then every *finite* repetition of that game has a unique SGPE.

[4] Is (2) true if the stage game possesses more than one Nash equilibrium?

Note that in principle, subgame perfection may be a difficult notion to check as there are an infinite variety of possible histories, and all sorts of different paths may emanate from them. The contribution here is that using the techniques of discounted dynamic programming, we can characterize quite easily those paths which are "supportable as SGPE" (see below for a formal definition).

17.2 Strategies and Paths

By retaining the connections between feasible programs and policies in dynamic programming (under certainty), we can think about strategies in terms of the paths that they induce following every history. Informally, these paths may be pieced back together to form a strategy, provided that their description satisfies some minimal consistency requirements.

Thus think of a strategy as specifying (i) an "initial path" $\mathbf{a}$, and (ii) paths $\mathbf{a}'$ following each t-history. We may think of the initial path as the "desired" outcome of the game, and all of all other (noninitial) paths as "punishments".

Example. A discrete Cournot game (Abreu [1988]). Consider the following 3×3 game where L, M, and H may be interpreted as choices of low, medium, and high output respectively, in a Cournot duopoly. The payoff matrix is described as

	L	M	H
L	$10, 10$	$3, 15$	$0, 7$
M	$15, 3$	$7, 7^*$	$-4, 5$
H	$7, 0$	$5, -4$	$-15, -15$

Consider the following description. Begin by playing the *path* given by $\mathbf{a}^1 \equiv \{(L,L),(L,L),\ldots\}$. For any t-history of the form $h(t) = \{(L,L),(L,L),\ldots(L,L)\}$, continue to play this path. For any *other* t-history, start up the path $\mathbf{a}^2 \equiv \{(M,M),(M,M),\ldots\}$. It is obvious that the paths $\mathbf{a}^1$ and $\mathbf{a}^2$, and the attendant rules that describe when to play which path, have an equivalent description in terms of strategies. In this case, it would be the strategy σ common to all players: $\sigma(0) = L$, and for each $t \geq 1$, $\sigma(t)[h(t)] = L$ if $h(t) = \{(L,L),(L,L),\ldots(L,L)\}$, and $\sigma(t)[h(t)] = M$ otherwise.

Exercise. [1] Prove that in any repeated game, the strategy given by the play of a a given one-shot Nash equilibrium of the stage game is SGPE.

[2] Using [1], prove that the strategy described in the example above is SGPE if $\beta \geq \frac{5}{8}$.

17.3 Simple Strategy Profiles

Consider strategy profiles that have the good fortune to be completely described by an $(n+1)$-vector of paths $(\mathbf{a}^0,\mathbf{a}^1,\ldots,\mathbf{a}^n)$, and a simple rule that describes when each path is to be in effect. Think of $\mathbf{a}^0$ as the initial or "desired" path and of $\mathbf{a}^i$ as the punishment for player i. That is, any unilateral deviation of player i from any path will be followed by starting up the path $\mathbf{a}^i$. Formally: the *simple strategy profile* $\sigma(\mathbf{a}^0,\mathbf{a}^1,\ldots,\mathbf{a}^n)$ specifies (i) the play of $\mathbf{a}^0$ until some player deviates unilaterally from this path, (ii) for any $i \in N$, the play of the path $\mathbf{a}^i$ if the ith player deviates unilaterally from the path $\mathbf{a}^j$, $j = 0,1,\ldots,n$, which is the ongoing path, (iii) the continuation of any ongoing path if no deviations occur or if two or more deviations occur simultaneously.

Exercise. Using an inductive argument on time periods, or otherwise, prove that the above description pins down a unique strategy profile.

We will introduce one more piece of notation to describe the utility received from the tail of a path. Thus let $F_i(\mathbf{a},t) \equiv \sum_{s=0}^{\infty} \beta^s f_i(a(t+s))$ for all paths $\mathbf{a}$, all t, and all i.

Theorem **25** *The simple strategy profile* $\sigma(\mathbf{a}^0,\mathbf{a}^1,\ldots,\mathbf{a}^n)$ *is a SGPE if and only if*

$$f_i(a_i', a_{-i}^j(t)) - f_i(a^j(t)) \leq \beta \left[F_i(\mathbf{a}^j, t+1) - F_i(\mathbf{a}^i)\right] \tag{47}$$

for all $j = 0,1,\ldots,n$, $i = 1,2,\ldots,n$, $t = 0,1,2,\ldots$, *and* $a_i' \in A_i$.

Proof. [Sufficiency.] Look at the problem faced by an arbitrary player i, assuming that all other players are sticking to the strategy suggested by the simple strategy profile. Now construct the following stochastic dynamic programming problem.

Collect as elements of the state space X all pairs of the following kind: the first element of the pair is one of the paths, and the second element of the pair is a *date*, signifying at which stage of that path we are in. Thus

$$X = \{\mathbf{a}^j, t\}_{j=0,1,\ldots,n;t=0,1,2,\ldots}$$

This is a deterministic dynamic programming problem and we are going to allow player i to choose next period's state, subject to feasibility. If the state is $(\mathbf{a}^j, t)$, the player can choose to "continue" by choosing $(\mathbf{a}^j, t+1)$ (the interpretation is that he has played his piece of the required vector, $a_i^j(t)$), or he can "deviate" by choosing the tomorrow's state to be $(\mathbf{a}^i, 0)$ (the implication being that he has not played his part of the required action vector). These are the only two choices, and this defines the feasibility correspondence for all states.

Let $w(x)$ be the value function obtained by the policy: choose the next state to be $(\mathbf{a}^j, t+1)$ whenever the state is $(\mathbf{a}^j, t)$. This function is certainly bounded. Consider the problem:

$$\max_{x' \in \Gamma(x)} f((\mathbf{a}^j, t), x') + \beta w(x').$$

Note that w satisfies all the conditions of theorem 2. It follows that w is the supremum function. By Theorem 4, the policy that generates w is an optimal strategy.

Necessity is trivial. ■

To proceed further, the following definitions will be useful. Say that a path $\mathbf{a}$ is a *perfect equilibrium path* (or is *supportable as a SGPE*) if there exists a SGPE σ such that $\mathbf{a} = \mathbf{a}(\sigma)$. A payoff vector v is a *perfect equilibrium payoff vector* if $v_i = F_i(\mathbf{a}(\sigma))$ for all i, for some SGPE σ. Let V be the set of all perfect equilibrium payoff vectors.

Assumption (G.1) There exists a subgame perfect equilibrium of the game G^∞.

Exercise. By a previous exercise, (G.1) is satisfied if the stage game admits a one-shot Nash equilibrium. Construct an example to show that (G.1) may be satisfied even if the one-shot game has no Nash equilibrium. you may want to return to this problem after finishing the study of the main theorems.

Under (G.1), V is nonempty. Define for each i the infimum payoff within the set V. That is,

$$\underline{v}_i \equiv \inf\{v_i | v \in V\}.$$

Lemma **7** *If* $\mathbf{a}$ *is a perfect equilibrium path, then for all* $t \geq 0$ *and all* i,

$$f_i(a_i', a_{-i}(t)) - f_i(a(t)) \leq \beta[F_i(\mathbf{a}, t+1) - \underline{v}_i] \tag{48}$$

Proof. Let σ support the path $\mathbf{a}$. Then it must be true that along the path, no deviation is possible. Because σ is a SGPE, it cannot prescribe payoffs following the deviation that fall short of $\underline{v}_i$. Therefore (48) must hold (make this precise). ■

Assumption (G.1) A is compact, and $f_i : A \to I\!R$ is continuous for all i.

Theorem **26** *There are paths* $(\tilde{\mathbf{a}}^i)_{i \in N}$ *such that for each* i,

$$F_i(\tilde{\mathbf{a}}^i) = \underline{v}_i.$$

Moreover, for each j*, the simple strategy profile generated by* $(\tilde{\mathbf{a}}^j; \tilde{\mathbf{a}}^1, \ldots, \tilde{\mathbf{a}}^n)$ *is a SGPE.*

Proof. Give A^∞ the topology of pointwise convergence. Because f_i is bounded on A and $\beta \in (0,1)$, the function $F_i(\mathbf{a})$ is continuous on A^∞ in this topology (Lemma 1).

Now for each i, because $\underline{v}_i \equiv \inf\{v_i | v \in V\}$, there exists a sequence of paths $\mathbf{a}^{(k)}$ such that

$$F_i(\mathbf{a}^{(k)}) \to \underline{v}_i \tag{49}$$

as $k \to \infty$. Because A is compact, we can use a diagonal argument to establish the existence of a subsequence m of k such that $a^{(m)}(t) \to \tilde{a}^i(t)$ for each $t \geq 0$. This gives us, in particular, a path $\tilde{\mathbf{a}}^i$ corresponding to each infimum value $\underline{v}_i$. Using (49) and the continuity of F_i, it follows that $F_i(\tilde{\mathbf{a}}^i) = \underline{v}_i$ for each i, as desired.

It remains to prove that for each j, the simple strategy profile generated by $(\tilde{\mathbf{a}}^j; \tilde{\mathbf{a}}^1, \ldots, \tilde{\mathbf{a}}^n)$ is a SGPE. Suppose not. Then by Theorem 25, there exist t, a player i, and $k \in \{1, \ldots, n\}$ such that for some $a'_i \in A_i$,

$$\begin{aligned} f_i(a'_i, \tilde{a}^k_{-i}(t)) - f_i(\tilde{a}^k(t)) &> \beta[F_i(\tilde{\mathbf{a}}^k, t+1) - F_i(\tilde{\mathbf{a}}^i)] \\ &= \beta[F_i(\tilde{\mathbf{a}}^k, t+1) - \underline{v}_i], \end{aligned}$$

using the definition of $\underline{v}_i$. Now recall a sequence of perfect equilibrium paths $\mathbf{a}^{(m)}$ that converged pointwise to the limit path $\tilde{\mathbf{a}}^k$. It follows from the above inequality that for m large enough,

$$f_i(a'_i, a^{(m)}_{-i}(t)) - f_i(a^{(m)}(t)) > \beta[F_i(\mathbf{a}^{(m)}, t+1) - \underline{v}_i],$$

which contradicts Lemma 7 applied to the perfect equilibrium path $\mathbf{a}^{(m)}$. ■

The characterization above leads easily to the following fundamental theorem for discounted repeated games.

Theorem **27** (Abreu [1988]). *A path* $\mathbf{a}$ *is supportable as a SGPE if and only if the simple strategy profile generated by* $(\mathbf{a}; \tilde{\mathbf{a}}^1, \ldots, \tilde{\mathbf{a}}^n)$ *is a SGPE.*

Proof. The "if" direction is trivial. To establish necessity, suppose on the contrary that $\sigma(\mathbf{a}; \tilde{\mathbf{a}}^1, \ldots, \tilde{\mathbf{a}}^n)$ is not a SGPE. For ease in writing, let $\mathbf{a} \equiv \tilde{\mathbf{a}}^0$. Then by Theorem 25, there exist t, i and $a'_i \in A_i$ such that for some path $\tilde{\mathbf{a}}^j$, $j = 0, 1, \ldots, n$,

$$\begin{aligned} f_i(a'_i, \tilde{a}^j_{-i}(t)) - f_i(\tilde{a}^j(t)) &> \beta[F_i(\tilde{\mathbf{a}}^j, t+1) - F_i(\tilde{\mathbf{a}}^i)] \\ &= \beta[F_i(\tilde{\mathbf{a}}^j, t+1) - \underline{v}_i]. \end{aligned}$$

But using Lemma 7, this contradicts the supposition that $\tilde{\mathbf{a}}^j$ is a perfect equilibrium path for every $j = 0, 1, \ldots, n$. ■

The collection $(\tilde{\mathbf{a}}^1, \ldots, \tilde{\mathbf{a}}^n)$ is called an *optimal penal code*, or sometimes *Abreu punishments.* The insight underlying an optimal penal code is that unless there is some extraneous reason to make the punishment fit the crime, a discounted repeated game sees no reason to use such tailored punishments.[8] It should be mentioned, however, that while these punishments appear to be "simple" in principle, they may be hard to compute in actual applications.

[8] Be warned: this result is *not* true of undiscounted repeated games.

Later on we shall specialize to symmetric games to obtain some additional insight into how these punishments work.

For now, let's return to the example introduced earlier to see how optimal penal codes may be found under some circumstances, and to reiterate some special features of these punishments. Recall the discrete Cournot game:

	L	M	H
L	$10, 10$	$3, 15$	$0, 7$
M	$15, 3$	$7, 7^*$	$-4, 5$
H	$7, 0$	$5, -4$	$-15, -15$

Suppose that $\delta = \frac{4}{7} < \frac{5}{8}$. Consider the two paths given by

$$\tilde{\mathbf{a}}^1 \equiv \{(M, H); (L, M), (L, M), (L, M), \ldots\}$$

and

$$\tilde{\mathbf{a}}^2 \equiv \{(H, M); (M, L), (M, L), (M, L), \ldots\}.$$

Begging the question for the moment of whether these generate a SGPE, observe that these punishments do lead to lower payoffs than Cournot-Nash reversion. With $\delta = \frac{4}{7}$, player 1, for instance receives

$$-4 + \frac{4}{7}\frac{3}{1 - \frac{4}{7}} = 0$$

if the punishment is first imposed on him. But a payoff of zero is special in this case. To see this, define the *security level* for a player as the lowest payoff that the player can conceivably be pushed to, if he is permitted to play a best response to the strategy vector followed by the others (the others are not necessarily playing a best response).

Exercise. Verify that the security level of each player in this example is zero.

Observation 1. Because the punishments push down the player to his security level, we know that we have found an optimal penal code (provided that the equilibrium properties are satisfied). This is a general principle which is of use in many applications.

Now let us check that these punishments indeed form an optimal penal code. To do so:

Exercise. Check that player 1 has no incentive to deviate from the first path, and nor does player 2, under the usual rules. Symmetry then takes care of the second path. This exercise leads to the following general observations:

Observation 2. A player may need to cooperate in his own punishment.

Observation 3. A player may need to cooperate in the punishment of another player.

Exercise. Let $(\tilde{\mathbf{a}}^1, \ldots, \tilde{\mathbf{a}}^n)$ be an optimal penal code for a game. Prove that

$$F_i(\tilde{\mathbf{a}}^i, 1) \geq \underline{v}_i$$

with strict inequality holding whenever the collection of actions $\{\tilde{a}^i_j(0)\}$ is not a one-shot Nash equilibrium.

17.4 Symmetric Games: A Special Case

Finding individual-specific punishments may be a very complicated exercise in actual applications. See Abreu [1986] for just how difficult this exercise can get, even in the context of a simple game such as Cournot oligopoly. The purpose of this section is to identify the worst punishments in a subclass of cases when we restrict strategies to be symmetric in a strong sense.

A game G is *symmetric* if $A_i = A_j$ for all players i and j, and the payoff functions are symmetric in the sense that for every permutation p of the set of players $\{1, \ldots, n\}$,

$$f_i(a) = f_{p(i)}(a_p)$$

for all i and action vectors a, where a_p denotes the action vector obtained by permuting the indices of a according to the permutation p.

A strategy profile σ is *strongly symmetric* if $\sigma_i(0) = \sigma_j(0)$ for all i and j, and if for every $t \geq 1$ and t-history $h(t)$, $\sigma_i(t)[h(t)] = \sigma_j(t)[h(t)]$ for all i and j. Note that the symmetry is "strong" in the sense that players take the same actions after all histories, including asymmetric ones.

Now for some special assumptions. We will suppose that each A_i is the (same) interval of real numbers, unbounded above. Assume that payoffs are continuous and bounded above, and

Condition 1. The payoff to symmetric action vectors (captured by the scalar a), denoted $f(a)$, is quasiconcave, with $f(a) \to -\infty$ as $a \to \infty$.

Condition 2. The best payoff to any player when all other players take the symmetric action a, denoted by $d(a)$, is nonincreasing in a, but bounded below.

It will be convenient in what follows, and in later sections, to *normalize* all payoffs using the discount factor. Thus for an action path $\mathbf{a}$, we will write the payoff function to any player i as

$$(1 - \beta) \sum_{t=0}^{\infty} \beta^t f_i(a(t)).$$

In this way, we can think of any infinite-horizon payoff as a convex combination of one-shot payoffs. This interpretation will also be useful when we do comparative statics by varying the discount factor.

Exercise. As you can tell, Conditions 1 and 2 are set up to handle something like the case of Cournot oligopoly. Even though the action sets do not satisfy the compactness assumption (G.2), the equilibrium payoff set is nevertheless compact. How do we prove this?

[1] First note that a one-shot equilibrium exists. To prove this, use Condition 1 and Kakutani's fixed point theorem.

[2] This means that the set of perfect equilibrium payoffs V is nonempty. Now, look at the worst perfect equilibrium payoff. Show that it is bounded below, using Condition 2. Using Condition 1, show that the best perfect equilibrium payoff is bounded above.

[3] Now show that the paths supporting infimum punishments indeed are well-defined, and together they form a simple strategy profile which is a SGPE. To show this, you will have

to prove that the action vectors along any perfect equilibrium path lie in a bounded set, so that the diagonal argument of Theorem 26 works. Prove this.

[4] Finally, prove the compactness of V by using part [3].

With the above exercise worked out, we can claim that there exists best and worst symmetric payoffs v^* and v_* respectively, in the class of all strongly symmetric SGPE. The following theorem then applies to these payoffs.

THEOREM **28** *Consider a symmetric game satisfying Conditions 1 and 2. Let v^* and v_* denote the highest and lowest payoff respectively in the class of all strongly symmetric SGPE. Then*

[a] The payoff v_* can be supported as a SGPE in the following way: Begin in phase I, where all players take an action a_* such that

$$(1-\beta)f(a_*) + \beta v^* = v_*.$$

If there are any defections, start up phase I again. Otherwise, switch to a perfect equilibrium with payoffs v^*.

[b] The payoff v^* can be supported as a SGPE using strategies that play a constant action a^* as long as there are no deviations, and by switching to phase 1 (with attendant payoffs v_*) if there are any deviations.

Proof. Part [a]. Fix some strongly symmetric equilibrium $\hat{\sigma}$ with payoff v_*. Because the continuation payoff can be no more than v^*, the first period action along this equilibrium must satisfy

$$f(a) \geq \frac{-\beta v^* + v_*}{1-\beta}.$$

Using Condition 1, it is easy to see that there exists a_* such that $f(a_*) = \frac{-\beta v^* + v_*}{1-\beta}$. By Condition 2, it follows that $d(a_*) \leq d(a)$. Now, because $\hat{\sigma}$ is an equilibrium, it must be the case that

$$v_* \geq (1-\beta)d(a) + \beta v_* \geq 1-\beta)d(a_*) + \beta v_*,$$

so that the proposed strategy is immune to deviation in Phase I. If there are no deviations, we apply some SGPE creating v^*, so it follows that this entire strategy as described constitutes a SGPE.

Part [b]. Let $\tilde{\sigma}$ be a strongly symmetric equilibrium which attains the equilibrium payoff v^*. Let $\mathbf{a} \equiv \mathbf{a}(\tilde{\sigma})$ be the path generated. Then $\mathbf{a}$ has symmetric actions $a(t)$ at each date, and

$$v^* = (1-\beta)\sum_{t=0}^{\infty} \beta^t f(a_t).$$

Clearly, for the above equality to hold, there must exist some date T such that $f(a_T) \geq v^*$. Using Condition 1, pick $a^* \geq a_T$ such that $f(a^*) = v^*$. By Condition 2, $d(a^*) \leq d(a_T)$. Now consider the strategy profile that dictates the play of a^* forever, switching to Phase I if there are any deviations. Because $\tilde{\sigma}$ is an equilibrium, because v_* is the worst strongly symmetric continuation payoff, and because v^* is the largest continuation payoff along the equilibrium path at any date, we know that

$$v^* \geq (1-\beta)d(a_T) + \beta v_*.$$

Because $d(a_T) \geq d(a^*)$,

$$v^* \geq (1-\beta)d(a^*) + \beta v_*$$

as well, and we are done. ■

The problem of finding the best strongly symmetric equilibrium therefore reduces, in this case, to that of finding two numbers, representing the actions to be taken in two phases.

Something more can be said about the punishment phase, under the assumptions made here.

THEOREM **29** *Consider a symmetric game satisfying Conditions 1 and 2, and let (a_*, a^*) be the actions constructed to support v_* and v^* (see statement of Theorem 28). Then*

$$d(a_*) = v_*$$

Proof. We know that in the punishment phase,

$$v_* \geq (1-\beta)d(a_*) + \beta v_*, \tag{50}$$

while along the equilibrium path,

$$v_* = (1-\beta)f(a_*) + \beta v^*. \tag{51}$$

Suppose that strict inequality were to hold in (50), so that there exists a number $v < v_*$ such that

$$v \geq (1-\beta)d(a_*) + \beta v. \tag{52}$$

Using Condition 1, pick $a \geq a_*$ such that

$$v = (1-\beta)f(a) + \beta v^*. \tag{53}$$

[To see that this is possible, use Condition 1, (51), and the fact that $v < v_*$.] Note that $d(a) \leq d(a_*)$, by Condition 2. Using this information in (52), we may conclude that

$$v \geq (1-\beta)d(a) + \beta v. \tag{54}$$

Combining (53) and (54), we see from standard arguments (check) that v must be a strongly symmetric equilibrium payoff, which contradicts the definition of v_*. ■

17.5 The Set of Equilibrium Payoffs of a Repeated Game

The set of equilibrium payoffs of a repeated game is often a simpler object to deal with than the strategies themselves. There are also interesting properties of this set that are related to the functional equation of dynamic programming.

The set of feasible payoff vectors of G is given by the set

$$F \equiv \{p \in I\!R^n | f(a) = p \text{ for some } a \in A\}.$$

Let F^* be the convex hull of the set of feasible payoffs. It should be clear that any normalized payoff in the repeated game must lie in this set. Define the class of sets $\mathcal{F}$ by collecting all nonempty subsets of F^*. Formally,

$$\mathcal{F} \equiv \{E | E \neq \emptyset \text{ and } E \subseteq F^*\}.$$

Pick any $E \in \mathcal{F}$ and $p \in F^*$. Say that p is *supported by* E if there exist $n+1$ vectors (not necessarily distinct) $(p', p^1, \ldots, p^n) \in E$ and an action vector $a \in A$ such that

$$p = (1-\beta)f(a) + \beta p', \tag{55}$$

and for each i and action $a_i' \in A_i$,

$$p \geq (1-\beta)f(a_i', a_{-i}) + \beta p_i^i. \tag{56}$$

We may think of a as the *supporting action* of p, of p' as the *supporting continuation payoff* of p, and so on. Observe that part 2 of the definition already incorporates the idea that punishments are chosen independently of the crime. Under our assumptions (G.1) and (G.2), this will be all that we will need.

Now define a map $\phi : \mathcal{F} \to \mathcal{F}$ by

$$\phi(E) \equiv \{p \in F^* | p \text{ is supported by } E\}.$$

We will study several properties of this map. Begin with a fundamental notion: a set $W \in \mathcal{F}$ is *self-generating* (Abreu, Pearce and Stacchetti [1990]) if $W \subseteq \phi(W)$.

Theorem **30** *If W is self-generating, then $W \subseteq V$, where V is the set of all normalized perfect equilibrium payoffs.*

Proof. Pick $p \in W$. To show that p is a perfect equilibrium payoff, we must find a SGPE that supports it. We will define such a strategy profile σ by induction on the length of all t-histories. At $t = 0$ let $h(0)$ be an arbitrary singleton that stands for the only 0-history. Let $\sigma(0)[h(0)] = a(h(0))$, where $a(h(0))$ is the supporting action of p. Also, use the notation $p(h(0))$ to denote p.

Recursively, suppose that we have defined an action $a(h(s))$ as well as an equilibrium payoff vector $p(h(s))$ for every s-history $h(s)$ and all $0 \leq s \leq t$. Now consider a $t+1$ history h_{t+1}, which we can write in the obvious way as $h_{t+1} = (h(t), a(t))$ for some t-history $h(t)$ and some action vector $a(t) \in A$.

Let a be a supporting action for $p(h(t))$. If $a = a(t)$, or if a differs from $a(t)$ in at least two components, define $p(h(t+1))$ to be p', where p' is the supporting continuation payoff for $p(h_t)$, and define $a(h(t+1))$ to be the action that supports $p(h(t+1))$. If $a \neq a(t)$ in precisely one component i, then define $p(h(t+1))$ to be the ith supporting punishment for $p(h(t))$, and $a(h(t+1))$ to be the supporting action for $p(h(t+1))$.

Having completed this recursion, define a strategy profile by $\sigma(t)[h(t)] = a(h(t))$ for every t and every t-history. We must check that this profile indeed constitutes a Nash equilibrium.

To show this we proceed along lines similar to those in the proof of Theorem 25. Look at the problem faced by an arbitrary player i, assuming that all other players are sticking to the strategy suggested by the simple strategy profile. Now construct the following stochastic dynamic programming problem.

Collect as elements of the state space X t-histories, for $t \geq 0$ (remember that we constructed an artificial singleton 0-history). Thus if H^t is the space of all t-histories,

$$X = \cup_{t=0}^{\infty} H^t$$

This is a deterministic problem, so we allow player i to choose next period's state, subject to feasibility. If the state is $h(t)$, the player can choose any $t+1$ history which is consistent with the other players sticking to the strategy profile σ. The feasibility correspondence is therefore given by

$$\Gamma(h(t)) = \{h(t+1) | h(t+1) = (h(t), a(t)) \text{ for some } a(t) \in A(t) \text{ with } a_{-i}(t) = \sigma_{-i}(t)[h(t)]\}.$$

Finally, the one-period payoff, call it $\pi_i(h(t), h(t+1))$, is given simply by the action vector $a(t)$ implied by $h(t)$ and $h(t+1)$.

Let $w(x)$ be the value function obtained by the policy: choose the next state to be $(h(t), \sigma(t)[h(t)])$ whenever the state is $h(t)$. This function is certainly bounded. Consider the problem:

$$\begin{aligned} & \max_{h(t+1)\in\Gamma(h(t))} \pi_i(h(t), h(t+1)) + \beta w(h(t+1)) \\ = & \max_{a_i \in A_i} f_i(a_i, \sigma_{-i}(t)[h(t)]) + \beta w((h(t), a_i, \sigma_{-i}(t)[h(t)]). \end{aligned}$$

Note that w satisfies all the conditions of Theorem 2. It follows that w is the supremum function.

Using the definitions of a supporter (see (55) and (56), it follows that this functional equation is solved at all histories by picking precisely the action chosen by σ_i, namely, $\sigma_i(t)[h(t)]$ for every t and $h(t)$. By Theorem 4, the policy that generates w is an optimal strategy. ∎

Exercise. Establish the following properties of the mapping ϕ.

[1] ϕ is isotone in the sense that if $E \subseteq E'$, then $\phi(E) \subseteq \phi(E')$.

[2] Under assumptions (G.1) and (G.2), ϕ maps compact sets to compact sets: that is, if E is a compact subset of F^*, then $\phi(E)$ is compact as well.

Our next theorem is an old result: the set of perfect equilibrium payoffs is compact. But the proof is new.

Theorem **31** *Under assumptions (G.1) and (G.2), the set of perfect equilibrium payoffs V is compact.*

Proof. We begin by showing that $V \subseteq \phi(\text{cl } V)$, where cl V denotes the closure of V. To this end, take any perfect equilibrium payoff p. Then there is a SGPE supporting p. Consider the action a vector prescribed in the first date of this equilibrium, as well as the prescribed paths and payoff vectors following every 1-history. These may be partitioned in the following way: (i) the payoff p' assigned to the continuation of the initial path, (ii) for each i, a function that assigns a payoff vector $p^i(a'_i)$ following each choice of a'_i at time period zero, assuming that others are sticking to the prescription of a, and (iii) payoff vectors that follow upon multiple simultaneous deviations of players from the prescribed initial action vector i.

Ignore (iii) in what follows.

Consider (ii). Note that $p^i(a'_i) \in V$ for all a'_i, and that by the notion of SGPE,

$$p_i \geq (1-\beta) f_i(a'_i, a_{-i}) + \beta p^i_i(a'_i)$$

for all choices $a'_i \in A_i$. Replacing each $p^i(a'_i)$ by a payoff vector p^i in cl V that minimizes i's payoff (why is this possible?), we see that

$$p_i \geq (1-\beta) f_i(a'_i, a_{-i}) + \beta p^i_i$$

for every action a'_i. Do this for every i, and combine with (i) to conclude that $(a, p', p^1, \ldots, p^n)$ is a supporter of p. This proves that $p \in \phi(\text{cl } V)$, so that $V \subseteq \phi(\text{cl } V)$.

Next, observe that because V is bounded, cl V is compact. Consequently, by (2) of the exercise above, $\phi(\text{cl } V)$ is compact as well. It follows from this and the claim of the previous paragraph that cl $V \subseteq \phi(\text{cl } V)$. But then by Theorem 30, cl $V \subseteq V$. This means that V is closed (why?). Since V is bounded, V is compact. ■

These results permit the following characterization of V (note the analogy with the functional equation of dynamic programming).

Theorem **32** *Under assumptions (G.1) and (G.2), V is the largest fixed point of ϕ.*

Proof. First we show that V is indeed a fixed point of ϕ. Since $V \subseteq \phi(\text{cl } V)$ (see proof of Theorem 31) and since V is compact, it follows that $V \subseteq \phi(V)$. Let $W \equiv \phi(V)$, then $V \subseteq W$. By the exercise (1) above, it follows that $W = \phi(V) \subseteq \phi(W)$. Therefore W is self-generating, and so by Theorem 30, $W \subseteq V$. Combining, we see that $W = V$, which just means that V is a fixed point of ϕ.

To complete the proof, let W be any other fixed point of ϕ. Then W is self-generating. By Theorem 30, $W \subseteq V$, and we are done. ■

18 The Folk Theorem

This section of the notes relies on Fudenberg and Maskin [1986], and may be seen as another application of the concept of simple strategy profiles. The folk theorem reaches a negative conclusion regarding repeated games. Repeated games came into being as a way of reconciling the observation of collusive (non-Nash) behavior with some notion of individual rationality. The folk theorem tells us that in "explaining" such behavior, we run into a dilemma: we end up explaining too much. Roughly speaking, *every* individually rational payoff is supportable as a SGPE, provided that the discount factor is sufficiently close to unity.

As a preliminary definition, define the *security level* of each player i by the value

$$\hat{v}_i \equiv \min_{a \in A} d_i(a).$$

Let $\hat{a}^i$ be an action vector such that $\hat{v}_i$ is exactly attained. Let $\hat{v}^i_j$ be the payoff to j when this is happening. In other words, $\hat{v}_i = \hat{v}^i_i$. Normalize the security level to equal zero for each player.

For each β, denote by $V(\beta)$ the set of all (normalized) perfect equilibrium payoffs.

Finally , define $M_i \equiv \max_{a \in A} |f_i(a)|$, and $M \equiv \max_{i \in N} M_i$.

EXERCISE. Prove that for each i, there exists an action vector that attains i's security level, provided that (G.2) is satisfied.

THEOREM **33** *Define F^* to be the set of all individually rational feasible payoffs, i.e.,*

$$F^* \equiv F \cap \{v \in I\!R^n | v \geq 0 \text{ for all } i\},$$

and assume that F^* is n-dimensional. Then for each $\tilde{p}$ in F^* and each $\epsilon > 0$, there exists a payoff vector p in F^* and also in the ϵ-neighborhood of $\tilde{p}$ such that $p \in V(\beta)$ for all β sufficiently close to unity.

Proof. For simplicity we shall assume in this proof that any point in F, the convex hull of the set of feasible payoffs, can be attained by some pure strategy combination. Later, we indicate how the proof can be extended when this is not the case.

Pick any $\tilde{p} \in F^*$, and $\epsilon > 0$. Because F^* has full dimension, it is possible to find p in the ϵ-neighborhood of $\tilde{p}$ such that $p \in$ int F^*. now pick n payoff vectors $\{\bar{p}^i\}_{i \in N}$ (each in F^*) "around" p as follows:

$$\bar{p}^i_i = p_i,$$

$$\bar{p}^i_j = p_j + \gamma \text{ for } j \neq i,$$

for some $\gamma > 0$. These vectors will be fixed throughout the proof. By our simplifying assumption, there are action vectors $a, \bar{a}^1, \ldots, \bar{a}^n$ such that $f(a) = p$ and $f(\bar{a}^i) = \bar{p}^i$ for each $i = 1, \ldots, n$.

The first of these action vectors is, of course, going to support the desired payoff, and the latter are going to serve as "rewards" to people who carry out punishments that may not be in their own short-term interests. The punishments, in turn, are going to be derived from the actions $\{\hat{a}^i\}$ that minimax particular players and drive them down to their security levels. Now for a precise statement. For each $i = 0, 1, \ldots, n$, consider the paths

$$\mathbf{a}^0 \equiv (a, a, a, \ldots),$$

$$\begin{aligned} \mathbf{a}^i &= (\hat{a}^i, \ldots \hat{a}^i) \text{ for } T \text{ periods,} \\ &= (\bar{a}^i, \bar{a}^i, \ldots) \text{ thereafter,} \end{aligned}$$

where T is soon going to be cunningly chosen (see below).

Consider the simple strategy profile $\sigma \equiv \sigma(\mathbf{a}^0, \mathbf{a}^1, \ldots, \mathbf{a}^n)$. We claim that there exists an integer T and a $\beta^* \in (0,1)$ such that for all $\beta \in (\beta^*, 1)$, σ is a perfect equilibrium.

For convenience, let is record the normalized payoff to palyer i along each of the paths. Note that

$$F_i(\mathbf{a}^0, \beta) = p_i$$

for all i, and that for all $0 \leq t \leq T$,

$$F_i(\mathbf{a}^j, \beta, t) = \hat{v}_i^j(1 - \beta^{T+1-t} + \beta^{T+1-t} p_i^j(\gamma),$$

where $p_i^j(\gamma) = p_i + \gamma$ if $i \neq j$, and $p_i^j(\gamma) = p_i$ if $i = j$. Of course, for all $t \geq T+1$,

$$F_i(\mathbf{a}^j, \beta, t) = p_i^j(\gamma).$$

We must check the no-deviation constraints from each path. By Theorem 27, this is necessary and sufficient to check that σ is indeed a perfect equilibrium.

Deviations from $\mathbf{a}^0$. Suppose that player i were to deviate from the path $\mathbf{a}^0$. Then he gets minimaxed for $T+1$ periods, with 0 return, after which he gets p_i again forever. The best deviation along the path is bounded above by M, so that the no-deviation condition is surely satisfied if

$$p_i \geq (1-\beta)M + \beta^{T+1} p_i.$$

This inequality holds if

$$\frac{1-\beta^{T+1}}{1-\beta} \geq \frac{M}{p_i}. \tag{57}$$

[Note that $p_i > 0$ (why?) so that this inequality makes perfect sense.] Now look at (57). As $\beta \to 1$, the LHS goes to $T+1$ (why?). So if we take

$$T \geq \max_i \frac{M}{p_i}, \tag{58}$$

then (57) is automatically satisfied for all β sufficiently close to unity.

Deviations from $\mathbf{a}^j$. First check player j's deviation. If (57) is satisfied (for $i = j$), the player j will never deviate from the second phase of his own punishment, because he just goes back to getting p_j. The deviations may be different, of course, but we have bounded these above by M anyway to arrive at (57). In the first phase, note that by construction, player j is playing a one-shot best response. So there is no point in deviating, as there is no short-term gain and it will be followed by restarting his punishment.

It remains to check player i's deviation from the path $\mathbf{a}^j$ when $i \neq j$. By the same argument as in the previous paragraph, a deviation in the second phase is not worthwhile, provided that (57) is satisfied. We need to check, then, that player i will cooperate with j's punishment in the first phase. He will do so if for each integer t that records the number of periods left in the first phase,

$$(1-\beta^t)\hat{v}_i^j + \beta^t(p_i + \gamma) \geq (1-\beta)M + \beta^{T+1}p_i.$$

Replace $\hat{v}_i^j$ by $-M$ on the LHS. On the RHS, replace $(1-\beta)M$ by $(1-\beta^T)M$ and $\beta^{T+1}p_i$ by $\beta^T p_i$. Then noting that M and p_i are both positive, it is clear that the above inequality holds if

$$-(1-\beta^t)M + \beta^t(p_i + \gamma) \geq (1-\beta^T)M + \beta^T p_i,$$

or if

$$\frac{\beta^T}{1-\beta^T} \geq \frac{2M}{\gamma}. \tag{59}$$

Now it should be clear that for any T satisfying (58), both conditions (57) and (59) are satisfied for all β sufficiently close to unity. So we are done.

It remains to remark on the case where the payoffs p and its 'constructed" neighbors are not exactly generated by pure action vectors. The proof then has to be slightly modified. First we perturb p a tiny bit if necessary and the choose β close enough to unity so that in a sufficiently large number of finite periods, we can get p as the convex combination of payoffs from various pure actions (where the convexification is being carried out intertemporally). Then all we have to do is to use a nonstationary path (with a finite periodicity) to generate p. We do the same for each of the payoff vectors $\bar{p}^i$ as well. The proof then goes through just the same way as before. The restrictions created by the choice of β go "in the same direction" anyway. ■

The full-dimensionality of F^* is needed in general (though not for two-player games). See Fudenberg and Maskin [1986] for more details.

18.1 Renegotiation-Proof Equilibrium

The Folk Theorem raises a conceptual puzzle. If, in searching for a theory of how collusive or cooperative outcomes are sustainable in an essentially noncooperative setting, we end up with a theory that "explains" not only such outcomes but many others besides, what are we to make of such a theory? One way to deal with this problem is that despite multiplicity, we take the "most collusive" outcomes permitted by such multiplicity as our solution concept. Put another way, agents seek a point on the Pareto frontier of the set of perfect equilibrium payoffs.

But this procedure leads naturally to an accusation of asymmetry. If we are so concerned that agents seek to find the Pareto frontier of whatever is supportable in the original game, why do they not act this way in *subgames*? At the very least, it is worth exploring the implications of the hypothesis that if agents can negotiate "at the beginning" to collude, they can renegotiate again in all subgames. This is the objective of the current section, which is based on Bernheim and Ray [1989] and Farrell and Maskin [1989].

We begin with finitely repeated games, where the definition is reasonably free of difficult conceptual problems. As a motivating example, consider the following 3×3 game:

$4, 4$	$0, 5$	$0, 0$
$5, 0$	$3, 3^*$	$0, 0$
$0, 0$	$0, 0$	$1, 1^*$

This game has two stage equilibria, which are the cells marked by an asterisk above. Repeat this game once without disccounting. In this two-period formulation, observe that the payoff vector $(4, 4)$ can be sustained in the first period by the promise of playing the "good" stage equilibrium $(3, 3)$ if there are no deviations, and the threat of playing the "bad" equilibrium $(1, 1)$ if there are any deviations.

Note that this equilibrium, though subgame perfect, looks problematic if agents are allowed to communicate, not just in the first period when they chose the collusive strategy, but also in the second period. If the latter is also true, then it seems possible that the two players would let "bygones be bygones" and play the "good" stage equilibrium $(3, 3)$ *regardless* of past history. But, of course, if it is commonly known that this is going to happen in period 2, no collusion is possible in period 1.

This example motivates the following definition for renegotiation-proof equilibrium in finitely repeated games. We shall use the familiar device of supporting payoffs to define the concept; it can easily be translated into a definition using strategies (see Bernheim and Ray [1989] for such a definition). We will also restrict ourselves to two-player games from this point on: while the definition of renegotiation proof equilibrium applies, without alterations, to n-player situations, issues of coalitional deviations also become important in such contexts.

To prepare for the definition, denote by V^T the set of all subgame perfect payoffs in a game that is repeated T times, i.e., played $T + 1$ times. Use the notation $\omega(A)$ to denote the weak Pareto frontier of some set A: i.e., $\omega(A) \equiv \{p \in A|$ there is no $p' \in A$ such that $p' \gg p\}$.

Begin with V^0: this is just the set of stage payoffs, and let $R^0 \equiv \omega(V^0)$. This is the set of renegotiation-proof payoffs in the one-shot game. There isn't much action here: simply pick the weak Pareto-frontier.[9]

Inductively, suppose that we have defined the set R^t as the set of renegotiation-proof payoffs in the t-repeated game. To figure out the R^{t+1}, first consider the set of all payoffs S^{t+1} that can be supported with R^t,

$$S^{t+1} \equiv \phi(R^t),$$

and then define the set of renegotiation-proof payoffs for the $(t+1)$-repeated game as

$$R^{t+1} \equiv \omega(S^{t+1}).$$

This completes the recursive definition.

The structure of the sets $\{R^t\}$ can be quite interesting, as the following example demonstrates.

Example. Consider a lender and borrower. The lender lends to the borrower at some fixed rate of interest $r > 0$. There are two projects A and B that the borrower can invest in, yielding net rates of return to the borrower of $\alpha > \beta > 0$. These projects are a matter of complete indifference to the lender, but in any case the lender can dictate the choice of project. At any date, there is a fixed exogenous penalty π for a default of any size on an ongoing loan. Finally, assume that there is an upper bound on the "bad" project B, given by a loan size of $\bar{L}$. On project A, assume no such bound (or a sufficiently larger bound, as the computations below will make clear).

Begin with the stage game. It is clear that there are only two equilibria. To find them, define

$$\ell_0 \equiv \frac{\pi}{1+r}.$$

Now observe that all loan sizes below ℓ_0 will be repaid by the borrower in the stage game. Consequently,

$$V^0 = \{(r\ell_0, \alpha\ell_0), (r\ell_0, \beta\ell_0)\},$$

and

$$R^0 = \omega(V^0) = V^0.$$

Now turn to the set of all payoffs that can be supported in the game repeated once. Let the discount factor be β. It should then be clear that apart from ℓ_0, an extra amount of loan can be sustained without fear of default, simply by the lender (credibly) threatening to revert to the bad project in the last period in the case of default in the first period. The (present value) loss to the borrower in that case is given by $(\alpha - \beta)\ell_0$, so that the maximum loan size in the first period of the two-period game is given by

$$\ell_1 \equiv \ell_0 + \frac{(\alpha - \beta)\ell_0}{1+r}.$$

Exercise. If $\ell_1 \leq \bar{L}$, carefully find the value of V^1, and then show that

$$R^1 = \{(r\ell_1, \alpha\ell_1), (r\ell_1, \beta\ell_1)\}.$$

[9] We use the weak Pareto frontier as our criterion in keeping with the idea that every player must strictly wish to renegotiate. This point is brought out clearly in the example below.

Recursively, as long as $\ell_t \leq \bar{L}$, we may define in exactly the same way,

$$\ell_{t+1} \equiv \ell_t + \frac{(\alpha - \beta)\ell_t}{1+r},$$

and then deduce that

$$R^{t+1} = \{(r\ell_{t+1}, \alpha\ell_{t+1}), (r\ell_{t+1}, \beta\ell_{t+1})\},$$

provided that $\ell_{t+1} \leq \bar{L}$. This recursion continues until we reach *first* date T (as we certainly must) such that

$$\ell_T > \bar{L}.$$

At this date, check that R^T must be the singleton set given by

$$R^T = \{(r\ell_T, \alpha\ell_T)\}.$$

If the finite-horizon game has a horizon longer than this, the entire process must build up again from this point! The idea is that in the game repeated T periods, there is exactly *one* renegotiation proof payoff. Consequently, in the game repeated $T+1$ times, all that can be sustained at the initial date is the original loan size ℓ_0! For longer games, the cyclical path builds itself up again, just as outlined above.

It is therefore possible for renegotiation-proof equilibria to exhibit "periodic breakdowns" of cooperation *on* the equilibrium path, and indeed, to select such paths as the unique outcome. This illustrates well the consequences of applying the same selection criteria (in this case, Pareto-optimality) to all subgames as well as on the initial equilibrium path.

Now we turn to a definition of the concept for infintely repeated games. Say that a strategy profile σ is *weakly renegotiation proof* (WRP) if it is a SGPE and for all pairs of histories (h_t, h'_s) (where $s = t$ is allowed), the payoff vectors $F(\mathbf{a}(\sigma, h_t))$ and $F(\mathbf{a}(\sigma, h'_s))$ are mutually Pareto-incomparable.

This is easily seen to imply the following feature. Let $P(\sigma)$ be the set of all payoff vectors generated by σ, following all histories. Then $P(\sigma)$ is self-generating, and $\omega(P(\sigma)) = P(\sigma)$. The following observations are relevant.

[1] We could alternatively have taken the feature in the paragraph above to be the defining feature of a WRP *set*. A WRP set need not be associated with a single equilibrium: it is more like a set of payoffs that has a self-referential consistency property.

[2] This self-referentiality leads to conceptual problems. Observe that the singleton set consisting of the payoff vector generated by any Nash equilibrium of the stage game is WRP. WRP sets are by no means unique.

Exercise. Consider the Prisoner's Dilemma given by

$2, 2$	$0, 3$
$3, 0$	$1, 1^*$

Observe that (as discussed) $\{(1, 1)\}$ is a WRP set. Find another equilibrium that is WRP, and nowhere makes use of the mutual defection cell. Describe precisely the WRP set that it generates.

[3] Thus there is a tension in the "choice" of WRP sets: what is the appropriate theory of the game that players should adopt? One obvious answer is to choose the "best" WRP set: one that is not Pareto-dominated by any other point on any other WRP set. This is a requirement of *external consistency*, as you can tell. The WRP set itself is not just the only criterion that is being used, but a comparison *across* WRP sets is being made.

Unfortunately, the external consistency requirement is not ingeneral met in such a straightforward way. There may not exist *any* WRP set with the required property described in the preceding paragraph. The issue of external consistency then becomes problematic. This is as far as we need to go in this course: see Bernheim and Ray [1989] for a detailed discussion of this and related points.

[4] However, even on the grounds of internal consistency alone, WRP sets are suspect. To see this, consider the following example:

$8,8$	$0,0$	$0,0$
$0,0$	$0,0$	$1,2^*$
$0,0$	$2,1^*$	$0,0$

Consider the set of payoffs $W \equiv \{(1,2),(2,1)\}$. Check that this is indeed WRP.

Now suppose that players indeed hold to W as a theory of how the game will be played from "tomorrow" onwards. In that case, observe that W supports more than W itself: in stages, we see that it covers all the combinations in which $(1,2)$ and $(2,1)$ can be played, at the very least. This covers the line segment joining $(1,2)$ to $(2,1)$; at least, all the rational convex combinations (weighted by the discount factor) of the two. Thus points approximately halfway between the two extremes become available. But now observe that with such a set, it is possible to sustain the collusive payoff $(8,8)$ in the first period. This gets us into trouble, because such payoffs Pareto-dominate segments of W, which will be eliminated as we finally apply the map ω. Thus a truly *internally consistent* renegotiation-proof set may lie pretty far away from W. For more on these issues, see Ray [1994].

With these qualifications in mind, let us return to the study of WRP equilibria. First,a definition. Say that a payoff vector v can be *sustained as a WRP payoff* if there is some WRP equilibrium σ such that $v \in P(\sigma)$.

Following Farrell and Maskin [1989], it is possible to obtain a characterization of WRP payoffs, under some assumptions. Let F^{**} be the convex hull of the set of all feasible, *strictly* individually rational payoffs; i.e.,

$$F^{**} \equiv \{v \in F | v \gg 0\}.$$

In what follows, we shall assume that for every $v \in F^{**}$, there is an action vector $a \in A$ such that $f(a) = v$. Recall that for the folk theorem, we also made an assumption like this at the beginning, and then argued after the theorem that such an assumption can be dropped costlessly. This assumption cannot be dropped with equal facility. I will return to this point below.

THEOREM **34** *Assume (G.2) and the assumption in the previous paragraph. Let $v \in F^{**}$. Suppose that there are action vectors $a^i \in A$, for $i = 1,2$ such that*

$$d_i(a^i) \;<\; v_i \text{ for } i = 1,2,$$

$$f_j(a^i) \geq v_j \text{ for } j \neq i. \quad (60)$$

Then v is a WRP payoff for all β sufficiently close to unity.

Moreover, if $v \in F^{**}$ is a WRP payoff for any β, then there exist $a^i \in A$, for $i = 1, 2$ such that

$$\begin{aligned} d_i(a^i) &\leq v_i \text{ for } i = 1, 2, \\ f_j(a^i) &\geq v_j \text{ for } j \neq i. \end{aligned} \quad (61)$$

Proof. Sufficiency. Let a be an action vector that attains the payoff v. This is going to be the initial path, while the action vectors a^1 and a^2 are going to serve as punishments.

Begin by observing that there exists $\beta^1 \in (0, 1)$ such that if $\beta \in (\beta^1, 1)$,

$$(1 - \beta)M + \beta d_i(a^i) < v_i, \quad (62)$$

for $i = 1, 2$, where M, it will be recalled, is the value of the maximum absolute payoff in the stage game. Because (62) is strict, there exists a vector p such that for each i,

$$p_i > d_i(a^i) \quad (63)$$

and

$$(1 - \beta)M + \beta p_i < v_i \quad (64)$$

for all $\beta \in (\beta^1, 1)$.

The idea, now, will be to replicate the value of p_i by playing a^i T times (where T is an integer to be determined), and then go back to the normal phase of playing a. Thus what we want is

$$p_i = (1 - \beta^T) f_i(a^i) + \beta^T v_i \quad (65)$$

for some integer T.

To go about this, let us substitute the RHS of (65) into the inequalities (62) and (63), and see what we need. Note that once we settle on a T, (63) is not going to be a problem for β close enough to unity, because $v_i > d_i(a^i)$ by assumption. For inequality (64) to hold, it must be the case that

$$(1 - \beta)M + \beta[(1 - \beta^T) f_i(a^i) + \beta^T v_i] < v_i$$

must hold for all β close enough to unity. Note that the LHS of the expression above equals v_i at $\beta = 1$. So for the desired result, we need the derivative of the LHS with respect to β to be positive, evaluated at $\beta = 1$. Taking the derivative, we obtain the expression

$$_M + f_i(a^i)[1 - (T + 1)\beta^T] + (T + 1)\beta^T v_i,$$

and evaluating this at $\beta = 1$, we get

$$-M - f_i(a^i)T + (T + 1)v_i$$

so that the required condition is

$$(T+1)v_i > f_i(a^i)T + M. \tag{66}$$

This can be guaranteed for large T, because $v_i > d_i(a^i) \geq f_i(a^i)$. Choose T satisfying (66). Then there is some $\beta^* \in (\beta^1, 1)$ such that if $\beta \in (\beta^*, 1)$, conditions (63), (64) and (65) all hold.

Now define three paths as follows:

$$\mathbf{a}^0 \equiv (a, a, a, \ldots), \text{ and}$$

$$\begin{aligned} \mathbf{a}^i &= (a^i, a^i, \ldots, a^i) \ T \text{ times} \\ &= (a, a, a, \ldots) \text{ thereafter,} \end{aligned}$$

for $i = 1, 2$.

We claim that $\sigma(\mathbf{a}^0, \mathbf{a}^1, \mathbf{a}^2)$ is a WRP equilibrium. To establish this, first let's check for subgame perfection.

Deviations from $\mathbf{a}^0$. If i deviates from $\mathbf{a}^0$, he gets *at most*

$$(1-\beta)M + \beta p_i,$$

by construction. By (64), this is less than v_i.

Deviations from $\mathbf{a}^i$. Suppose, first, that i deviates. The most tempting deviation is in the first period, by the construction of the punishment path. In this case, the total payoff is

$$(1-\beta)d_i(a^i) + \beta p_i.$$

Using (63), this is less than p_i, so that deviations by i from $\mathbf{a}^i$ are not profitable.

Likewise, j will not deviate from $\mathbf{a}^i$, because by the assumption that $f_j(a^i) \geq v_j$ and the nature of the path $\mathbf{a}^i$, (64) applies right away to prevent deviations (check this).

To complete the proof of sufficiency, all we have to do is check that no two payoff vectors generated by $\sigma(\mathbf{a}^0, \mathbf{a}^1, \mathbf{a}^2)$ ever Pareto-dominate each other. This follows directly from the properties of a^1 and a^2 relative to the payoff vector v.

Remark on the assumption. As in the folk theorem, we can relax the assumption that the payoff vector v is achieved by a pure action. The problem that we now have is to make sure that the various actions which we shall use to intertemporally simulate v do not end up Pareto-dominating each other, or indeed, Pareto-dominating the payoffs. This is not trivial. For an indication of how this task is accomplished in the case where mixed strategies can be observed, see Farrell and Maskin [1989].

Necessity. Suppose that $v \in F^{**}$ is a WRP payoff; i.e., $v \in P(\sigma)$ for some WRP equilibrium σ. We will prove that an action pair a^1 satisfying the required conditions (61) exists. [The proof for a^2 is completely analogous.]

Exercise. Assume (G.2). Prove that if v can be supported as a WRP equilibrium, then there exists a WRP σ such that $P(\sigma)$ is compact. [The idea is, as usual, to use a sequential compactness argument. In case you have problems, look at Farrell and Maskin [1989], Lemma 2, 356–357.]

By the exercise, we may assume without loss of generality that σ has a worst continuation equilibrium for player 1. *Choose, among these, the best for player 2.* Let a^1 be the first period action vector on this worst equilibrium path, and let σ^1 denote the continuation equilibrium starting the period after a^1. Finally, let v^* be the payoff vector when the worst punishment begins.

We will show that a^1 satisfies all the needed conditions.

Clearly, $v_1^* \leq v_1$. We claim that $v_2^* \geq v_2$. Suppose not. Then $v_2^* < v_2$. If $v_1^* < v_1$ as well, we have a contradiction to WRP. So this must mean that $v_1^* = v_1$. But in this case we contradict the choice of the worst equilibrium (it has to maximize player 2's payoff in the class of all equilibria that are worst for player 1). So $v_2^* \geq v_2$, as claimed.

Our next claim is that $f_2(a^1) \geq v_2^*$. Suppose not. Then $f_2(a^1) < v_2^*$. But then $F_2(\sigma^1) > v_2^*$, because σ^1 is the continuation equilibrium. By the WRP requirement, it follows that $F_1(\sigma^1) \leq v_1^*$. But this contradicts, again, our choice of v^*.

We complete the proof by showing that $d_1(a^1) \leq v_1^* \leq v_1$. As noted, $v_1^* \leq v_1$ by construction. To see that $d_1(a^1) \leq v_1^*$, observe that if this were not the case, player 1 could deviate from his punishment by getting $d_1(a^1)$ in the first period, followed by no less than v_1^*. Therefore $d_1(a^1) \leq v_1^* \leq v_1$, and we are done. ■

References

D. Abreu (1986), "Extremal equilibria of oligopolistic supergames," *Journal of Economic Theory* **39**, 191–225.

D. Abreu (1988), "On the theory of infinitely repeated games with discounting," *Econmometrica* **56**, 383–396.

D. Abreu, D. Pearce and E. Stacchetti (1990), "Toward a theory of discounted repeated games wth imperfect monitoring," *Econometrica* **58**, 1041–1063.

D. Bernheim and D. Ray (1989), "Collective dynamic consistency in repeated games," *Games and Economic Behavior* **1**, 295–326.

J. Farrell and E. Maskin (1989), "Renegotiation in repeated games," *Games and Economic Behavior* **1**, 327–369.

D. Fudenberg and E. Maskin (1986), "Folk theorems for repeated ga,es with discounting and incomplete information," *Econometrica* **54**, 533–554.

D. Ray (1994), "Internally renegotiation-proof equilibrium sets: limit behavior with low discounting," *Games and Economic Behavior* **6**, 162–177.

J. Rust (1995), "Numerical dynamic programming in economics," forthcoming, *Handbook of Computational Economics*, edited by H. Amman, D. Kendrick and J. Rust.

N. Stokey and R. Lucas (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press.