Appendix to Lecture 11, Examples

The following example shows that the value function \( V \) may not be continuous if we do not assume that \( G \) is lower semicontinuous.

**Example:** (Discontinuous value function) Let

\[
X = \{(y_1, y_2) \in \mathbb{R}^2 \mid (y_1 - 3)^2 + (y_2 - 3)^2 \leq 2\}.
\]

As a step in the definition of \( G \), let

\[
\Gamma = \{(y_1, y_2, 2, 2) \mid (y_1, y_2) \in X\} \cup \{(4, 4, c, c) \mid 2 \leq c \leq 4\}.
\]

Let the graph of \( G \) be the convex hull of \( \Gamma \), which is the set of all convex combinations of points in \( \Gamma \). The graph of \( G \) is convex and compact. The correspondence \( G \) is defined by the equation

\[
G(y_1, y_2) = \{(y_3, y_4) \mid (y_1, y_3, y_2, y_4) \in \text{graph of } G\}.
\]

Let \( A = X \times X \) and let \( u : X \times X \to \mathbb{R} \) be defined by the equation

\[
u(x_1, x_2) = x_{21} + x_{22}
\]

and let \( \beta \in (0, 1) \). The set \( G(x) \) is not hard to visualize. Let \( C \) be the circle that is the boundary of \( X \). That is,

\[
C = \{(y_1, y_2) \in \mathbb{R}^2 \mid (y_1 - 3)^2 + (y_2 - 3)^2 = 2\},
\]

so that \( (4, 4) \in C \). If \( x \in X \) and \( x \neq (4, 4) \), there is a unique line segment going from \( (4, 4) \) through \( x \). Let \( a(x) \) be the intersection of \( C \) with this line segment that is distinct from \( (4, 4) \). Then \( x \) is a convex combination of \( a(x) \) and \( (4, 4) \). Let \( \alpha(x) \) be defined by the equation

\[
x = \alpha(x)a(x) + (1 - \alpha(x))(4, 4).
\]

Let \( \alpha(4, 4) = 0 \). Notice that \( \alpha(x) = 1 \), if \( x \in C \) and \( x \neq (4, 4) \). Then

\[
G(x) = \{(r, r) \mid 2 \leq r \leq 4 - 2\alpha(x)\} \subset X.
\]

By lemma 5.4, \( G \) is upper semicontinuous, because its graph is a closed set and for all \( x \), \( G(x) \) is a subset of the compact set \( X \). \( G(x) \) is not empty, for all \( x \). The sets \( X \) and \( A \) are compact and \( u \) is continuous. The sets \( A \) and \( X \) and the graph of \( G \) are convex, \( X \) has non-empty interior, and \( u \) is concave, so that the example satisfies assumptions 1 - 4, 8, and 9. I show that the correspondence \( G \) is not lower semicontinuous at the point \( (4, 4) \). Let \( x^n \) be a sequence in \( X \) such that \( x^n \neq (4, 4) \), for all \( n \), and \( \lim_{n \to \infty} x^n = (4, 4) \). Then \( G(x^n) = \{(2, 2)\} \), for all \( n \), and
$G((4, 4)) = \{(r, r) \mid 2 \leq r \leq 4\}$. Hence there is no sequence $y^n$ such that $y^n \in G(x^n)$, for all $n$, and $\lim n y^n = (4, 4)$, and so $G$ is not lower semicontinuous at $(4, 4)$ and the example does not satisfy assumption 5. The next figure illustrates the definition of $a(x)$.

If $x_0 = (4, 4)$, then it is optimal to choose $x_1 = (4, 4)$ and enjoy utility $8$ in period 0. The optimal choice is then repeated indefinitely and a utility of $8$ is enjoyed in every period. Therefore $V(4, 4) = 8/(1 - \beta)$. If $x_0$ is any point in the circle $C$ other than $(4, 4)$, then the only choice possible is $(2, 2)$, which is in $C$ and not equal to $(4, 4)$, and so $(2, 2)$ must be chosen in period 1 and in every period thereafter. Therefore $V(x_0) = 4/(1 - \beta)$, and so the value function is not continuous at $(4, 4)$. This ends the discussion of the example.

The next example shows that Blackwell's theorem does not apply to every contraction mapping.

**Example:** Let $Z = \{ (\pi_1, \pi_2) \in \mathbb{R}_+^2 \mid \pi_1 + \pi_2 = 1 \}$. That is, $Z$ is the set of probability vectors over two events. Let $T : Z \rightarrow Z$ be defined by
\[
T(\pi_1, \pi_2) = \begin{pmatrix} 3/4 & 1/3 \\ 1/4 & 2/3 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = \begin{pmatrix} 3/4 \pi_1 + 1/3 \pi_2 \\ 1/4 \pi_1 + 2/3 \pi_2 \end{pmatrix}
\]

If \( x = (x_1, x_2) \in \mathbb{R}^2 \), let the norm of \( x \) be \( |x| = \max(|x_1|, |x_2|) \). That is, the norm of \( x \) is the maximum of the absolute values of the components of \( x \). If \( \pi = (\pi_1, \pi_2) \) and \( \eta = (\eta_1, \eta_2) \) belong to \( Z \), then \( x = \eta - \pi \) has the property that \( x_1 + x_2 = 0 \), so that \( x = (M, -M) \), for some number \( M \).

It is not hard to calculate that \( T(M, -M) = \frac{5}{12} (M, -M) \). Therefore \( T(\eta - \pi) = \frac{5}{12} (\eta - \pi) \), for \( \eta \) and \( \pi \) in \( Z \). Therefore \( |T(\eta - \pi)| = \frac{5}{12} |(\eta - \pi)| \) and so \( T \) is a contraction on \( C \). It is easy to check algebraically that \( \left( \frac{4}{7}, \frac{3}{7} \right) \) is the unique fixed point of \( T \). Blackwell’s theorem does not apply to \( T \), however. An unimportant reason why it does not apply is that if \( \pi = (\pi_1, \pi_2) \in Z \) and \( a \) is a non-negative number, then \( (\pi_1 + a, \pi_2 + a) \) does not belong to \( Z \). This alone should not matter, however, because \( T(\pi_1 + a, \pi_2 + a) \) is still defined. The real difficulty is that
\[
T(\pi + (a, a)) = T(\pi) + \begin{pmatrix} 1/3 a \\ 1/2 \end{pmatrix}
\]
which is not less than or equal to \( T(\pi) + \beta(a, a) \), for any number \( \beta \) such that \( 0 < \beta < 1 \). That is, \( T \) does not satisfy condition b of Blackwell’s theorem.

This ends the discussion of the example.

The matrix of the linear transformation \( T \) in the previous example is an example of a transition matrix of a Markov process. If the states are 1 and 2, then \( \pi_1 \) and \( \pi_2 \) are the probabilities of states 1 and 2. If the state currently is 1, the probability that in the next period the state will be 1 is 3/4 and the probability that the state will be 2 is 1/4. Similarly if the state currently is 2, the probabilities that in the next period the state will be 1 or 2 are 1/3 and 2/3, respectively. The probability distribution \( \pi = (\pi_1, \pi_2) = (4/7, 3/7) \) is called a stationary distribution, because it is preserved after the random transition.

A Markov transition over \( N \) states, \( n = 1, 2, \ldots, N \), is defined by an \( N \times N \) matrix \( A \) with \( (k, n) \)th entry \( a_{kn} \), where \( a_{kn} \) is the probability that the state will be \( k \) in the next period if it is \( n \) in the current period. Then \( a_{kn} \geq 0 \), for all \( k \) and \( n \), and \( \sum_{k=1}^{N} a_{kn} = 1 \), for all \( n \), so that the columns of \( A \) are probability vectors. If the probability distribution over the states currently is \( \pi = (\pi_1, \pi_2, \ldots, \pi_N) \), then the probability the state is \( k \) in the succeeding period is \( \sum_{n=1}^{N} a_{kn} \pi_n \).

The vector of these probabilities is
\[
\left( \sum_{n=1}^{N} a_{n} \pi_{n}, \ldots, \sum_{n=1}^{N} a_{n} \pi_{n} \right)
\]
or in vertical form

\[
\begin{pmatrix}
\sum_{n=1}^{N} a_{n} \pi_{n} \\
\vdots \\
\sum_{n=1}^{N} a_{n} \pi_{n}
\end{pmatrix}
= A \pi.
\]

The probability distribution over states two periods hence is \( A^{2} \pi \), and the probability distribution \( t \) periods from now is \( A^{t} \pi \), where \( A^{t} \) is \( A \) to the power \( t \). The next theorem asserts that as \( t \) goes to infinity, this distribution converges to a unique stationary distribution \( \eta \) such that \( A \eta = \eta \), provided \( \sum_{n=1}^{N} a_{kn} > 0 \), for all \( k \) and \( n \). I prove the theorem, because the demonstration is a contraction argument somewhat different from that of theorems 11.4 or 11.7.

**Theorem 11.23:** Let \( A \) be a non-negative \( N \times N \) matrix such that \( \sum_{k=1}^{N} a_{kn} = 1 \), for all \( n \), and \( a_{kn} > 0 \), for all \( k \) and \( n \). Then \( B = \lim_{t \to \infty} A^{t} \) exists and is such that if \( b_{k} \) is the \( k \)th row of \( B \), for \( k = 1, \ldots, N \), then \( b_{k} = (\eta_{1}, \eta_{2}, \ldots, \eta_{k}) \), where \( \sum_{k=1}^{N} \eta_{k} = 1 \). If \( \eta = \left( \begin{array}{c}
\eta_{1} \\
\vdots \\
\eta_{N}
\end{array} \right) \), then \( A \eta = \eta \). If \( \pi \) is any probability distribution over the \( N \) states, then \( \lim_{t \to \infty} A^{t} \pi = \eta \).

**Proof:** Let \( x^{t}_{k} = (x^{t}_{k}, \ldots, x^{t}_{1}) \) be the \( k \)th row of \( A^{t} \). Then \( x^{t}_{k} \geq 0 \), for all \( t, k \), and \( n \). Let \( M^{t}_{k} = \max(x^{t}_{k}, \ldots, x^{t}_{1}) \) and let \( m^{t}_{k} = \min(x^{t}_{k}, \ldots, x^{t}_{1}) \). Let \( \epsilon = \min_{k=1, \ldots, N} a_{kn} \), which by assumption is a positive number. Without loss of generality, assume that \( m^{t}_{k} = x^{t}_{k} \). Then

\[
\begin{align*}
x^{t+1}_{kn} &= \sum_{m=1}^{N} x^{t}_{km} a_{mn} + \sum_{m=2}^{N} x^{t}_{km} a_{mn} + \sum_{m=2}^{N} x^{t}_{km} a_{mn} \\
&\leq m^{t}_{k}a_{kn} + M^{t}_{k}\sum_{m=2}^{N} a_{mn} = m^{t}_{k}a_{kn} + M^{t}_{k}(1-a_{kn}) \leq \epsilon m^{t}_{k} + (1-\epsilon) M^{t}_{k}.
\end{align*}
\]
Since \( M_{k}^{t+1} = \max(x_{k1}^{t+1}, \ldots, x_{kN}^{t+1}) \), we see that
\[
M_{k}^{t+1} \leq \varepsilon m_{k}^{t} + (1 - \varepsilon) M_{k}^{t}.
\]

A similar argument proves that
\[
m_{k}^{t+1} \geq \varepsilon M_{k}^{t} + (1 - \varepsilon) m_{k}^{t}.
\]

Hence
\[
M_{k}^{t+1} - m_{k}^{t+1} \leq \varepsilon m_{k}^{t} + (1 - \varepsilon) M_{k}^{t} - \varepsilon M_{k}^{t} - (1 - \varepsilon) m_{k}^{t} = (1 - 2\varepsilon) (M_{k}^{t} - m_{k}^{t})
\]
so that
\[
M_{k}^{t} - m_{k}^{t} \leq (1 - 2\varepsilon)^{k} (M_{k}^{1} - m_{k}^{1})
\]
for all \( t \), and hence
\[
\lim_{t \to \infty} (M_{k}^{t} - m_{k}^{t}) = 0. \tag{11.17}
\]

The sequence \( M_{k}^{t} \) is non-negative and non-increasing in \( t \), and \( m_{k}^{t} \) is bounded above by \( M_{k}^{1} \) and is non-decreasing in \( t \), so that both sequences converge by the monotone convergence theorem 4.17. Inequality 11.17 implies that they must converge to the same limit. Call that limit \( \eta_{k} \).

Since \( m_{k}^{1} = \min_{n=1, \ldots, N} a_{kn} > 0 \) and \( \eta_{k} \geq m_{k}^{1} \), we know that \( \eta_{k} > 0 \). So every entry of the \( k^{th} \) row of \( B = \lim_{t \to \infty} A^{t} \) equals the positive number \( \eta_{k} \).

I show that \( \sum_{k=1}^{N} \eta_{k} = 1 \). Let \( e \) be the \( N \)-vector all of the components of which are 1. Since the columns of \( A \) are probability vectors, \( e^{T}A = e^{T} \). Therefore \( e^{T}A^{2} = e^{T}AA = e^{T}A = e^{T} \).

Continuing by induction on \( t \), we see that \( e^{T}A^{t} = e^{T} \), for all \( t \). Let \( \eta = \begin{pmatrix} \eta_{1} \\ \vdots \\ \eta_{N} \end{pmatrix} \). Then
\[
e^{T} = e^{T} \left( \lim_{t \to \infty} A^{t} \right) = e^{T} (\eta, \eta, \ldots, \eta) \quad \text{and so} \quad 1 = e^{T} \eta = e \eta. \quad \text{That is,} \quad \sum_{k=1}^{N} \eta_{k} = 1.
\]

I show that \( A\eta = \eta \). Notice that \( AB = A \left( \lim_{t \to \infty} A^{t} \right) = \lim_{t \to \infty} A^{t+1} = B \). Hence if \( b^{1} \) is the first column of \( B \), \( Ab^{1} = b^{1} \). Since \( b^{1} = \eta \), \( A\eta = \eta \).
I show that \( \lim_{t \to \infty} A^t \pi = \eta \), if \( \pi \) is any probability distribution over the \( N \) states. Let \( b_{kn} \) be the \((k, n)\)th entry of \( B \). Since \( b_{kn} = \eta_k \), for all \( k \) and \( n \), 
\[
\sum_{n=1}^{N} b_{kn} \pi_n = \sum_{n=1}^{N} \eta_k \pi_n = \eta_k \sum_{n=1}^{N} \pi_n = \eta_k.
\]
Hence \( B \pi = \eta \) and so \( \lim_{t \to \infty} A^t \pi = B \pi = \eta \). \( \blacksquare \)