IO Class Notes: Math Preliminaries to Dynamics.

Ariel Pakes.

November 8, 2004

Mathematical Preliminaries.

Some Definitions.

You might want to have these definitions accessible.

Definition 1. (Real Vector Space). A (real) vector space X is a set of elements (vectors) together with two operations, addition and scalar multiplication. Further

1. For any two vectors x, $y \in X$, addition gives a vector $x + y \in X$; and for any vector $x \in X$ and any real number $\alpha \in \mathbf{R}$, scalar multiplication gives a vector $\alpha x \in X$. These operations obey the usual algebraic laws; that is, for all x, y, $z \in X$, and $\alpha, \beta \in \mathbf{R}$:
   - $x + y = y + x$;
   - $(x + y) + z = x + (y + z)$;
   - $\alpha(x + y) = \alpha x + \alpha y$;
   - $(\alpha + \beta)x = \alpha x + \beta x$; and

- $(\alpha\beta)x = \alpha(\beta x)$.

2. Moreover, there is a zero vector $\theta \in X$ that has the following properties
   - $x + \theta = x$; and
   - $\theta x = \theta$
3. Finally, there is a unit vector $1 \in X$ with the property that

$$1x = x$$

Definition 2. (Metric Space). A metric space is a set B, together with a metric (distance function) $\rho : B \times B \to \mathbf{R}$, such that for all $x,\ y,\ z \in B$:

1. $\rho(x, y) \geq 0$, with equality if and only if $x = y$;
2. $\rho(x, y) = \rho(y, x)$; and
3. $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$.

Definition 3. (Normed Vector Space). A normed vector space is a vector space B, together with a norm $\|\cdot\| : B \to \mathbf{R}$, such that for all $x, y \in B$ and $\alpha \in \mathbf{R}$:

1. $\|x\| \geq 0$, with equality if and only if $x = \theta$;
2. $\|\alpha x\| = |\alpha| \cdot \|x\|$; and
3. $\|x + y\| \leq \|x\| + \|y\|$ (the triangle inequality).

A useful exercise is to show that a normed linear space is a metric space with $\rho(x, y) = \|x - y\|$.

Definition 4. (Convergence). A sequence $\{x_n\}_{n=0}^{\infty}$ in B **converges** to $x \in B$, if for each $\epsilon > 0$, there exists N_ϵ such that

$$\rho(x_n, x) < \epsilon, \ \ \forall \ \ n \geq N_\epsilon.$$

Definition 5. (Cauchy sequence). A sequence $\{x_n\}_{n=0}^{\infty}$ in B is a **Cauchy sequence** (satisfies the **Cauchy criterion**) if for each $\epsilon > 0$, there exists $N_\epsilon$such that

$$\rho(x_n, x_m) < \epsilon, \ \ \forall \ \ n, m \geq N_\epsilon.$$

Definition 6. (Complete Metric Space). A metric space (B, ρ) is **complete** if every Cauchy sequence in B converges to an element in B. A *Banach* space is a complete normed linear space.

When working in a complete metric space is often easier to show that a sequence is a Cauchy sequence then to show that the sequence converges (this is particularly true when we do not have a good guess as to what the limit point is).

Optional Question (purpose: show that completeness depends on the norm as well as the definition of B; discuss some limit properties of sequences of functions). Show that $C[1, b]$ (i.e. the space of continuous real valued functions with domain $[1, b]$), with $\|x\| = \int_1^b |x(t)|dt$ (this norm is usually denoted by $L^1[1, b]$) , is a normed vector space but is not a Banach space (i.e. it is not complete).
Hints. You should be able to prove it is a normed vector space without hints. Prove it is not complete by counterexample: i.e. show that there exists a Cauchy sequence of functions that are contained in C but that do not have a limit in C. Consider

$$f_n(x) = \{1/x^n\}, \quad x \in [1, b], \quad n = 1, 2, \ldots.$$

First show that this sequence is Cauchy convergent (use the form $|x^{-n} - x^{-m}|$, note that this difference goes to zero for each point in $[1, b]$, and then apply the Lebsegue Dominated Convergence Theorem to insure the integral converges [1]). Then, to prove the space is not complete it will suffice to show that this sequence of functions converges to a function which is not in the space (is not continuous). Note that $lim_{n \to \infty} f_n(\cdot) = f(x)$ where $f(x = 1) = 1$ and $f(x) = 0$ for $x > 1$.

There are a couple of reasons to be interested in this example.
Note 1. It shows that the limit of continuous functions need not be continuous.
Note 2. It shows that the limit of a sequence of strictly monotone functions is not necessarily strictly monotone, though it *is* weakly monotone. You should remember this.
Note 3. Though $\{f_n(x)\}$ is Cauchy in the $L^1[1, b]$ norm, it cannot be Cauchy in the sup norm; i.e. the norm defined by $\|f_n - f_m\| = sup_{(x \in [1,b])} |f_n(x) - f_m(x)|$ — since it is well known that the space $C[1, b]$ with the sup norm is complete (this is another point you should remember). You might prove that this sequence of functions is not a Cauchy sequence in the sup norm.

[1] The Lebesgue Dominated Convergence Theorem states that if $g_n(x) \to g(x)$ for every $x \in [a, b]$ except possibly on a finite subset of $[a, b]$, and $|g_n(x)| \leq g(x)$ for all n and x, then, provided $\int_a^b g(x)dx \leq K < \infty$, $lim_{n \to \infty} \int_a^b g_n(x)dx = \int_a^b g(x)dx$.

Relevant Theorems

If you have not seen this material before, I suggest you look over these notes before going to class, and then refer to them in more detail after you have seen the examples. The notes are more detailed than what we will need in class, and contain very little in the way of verbal exposition. The extra detail in the notes is designed as a stepping stone to using these techniques in your own research if that need arises.

The theorems that I will use in class are Theorem 1 and its corollary, and Theorem 2. Theorem 1 is the contraction mapping theorem. This will generally be used to find the value functions for our problems (though it is sometimes used to prove uniqueness of solutions to other fixed point problems in both theoretical and empirical I.O.). Its corollary establishes a way of proving properties of the limit functions; generally the value function and policy functions in our examples. Theorem 2 is Blackwell's theorem (it provides easy to use sufficient conditions for a contraction).

Operators in Banach spaces of functions are defined pointwise. That is if we say that T is an operator which takes the space of bounded functions on $S \subset R^l$, say $B(S)$, into itself it means that T gives us a rule which takes any element of $B(S)$, say $f(\cdot) : S \to R$, and produces another element of $B(S)$, say $g(\cdot) : S \to R$, by telling us how T transforms the number $f(s)$ into the number $g(s)$ for each $s \in S$. Thus if c is a constant, and $Tf = f + c$ we have $Tf_{(s)} = f(s) + c$, $\forall s \in S \ldots$.

Definition 7. (contraction mapping). Let (B, ρ) be a complete metric space and $T : B \rightarrow B$ be a function mapping B into itself. T is a **contraction mapping** (with **modulus** β) if for some $\beta \in (0, 1)$, $\rho(Tf, Tg) \leq \beta\rho(f, g)$, for all $f, g \in B$.•

In our examples B will typically be the set of bounded function on R^l (or some subset thereof), and $\rho(\cdot)$ will be the sup norm, i.e. $\rho(f, g) = sup_{x \in R^l}|f(x) - g(x)|$.

Theorem 1. (Contraction Mapping Theorem): If (B, ρ) is a complete metric space and $T : B \rightarrow B$ is a contraction mapping with modulus β then

- T has exactly one fixed point, say $\upsilon* \in B$
- for any $\upsilon_o \in B$, $\rho(T^n\upsilon_o, \upsilon*) \leq \beta^n\rho(\upsilon_o, \upsilon*)$, and hence by completeness $lim_{n\rightarrow\infty}T^n\upsilon_0 = \upsilon*$.
  - Here T^n is the n^{th} iterate of T so $T^2(\upsilon) = T(T(\upsilon))$ and $T^n(\upsilon) = T(T^{n-1})(\upsilon)$.
- $\rho(T^n\upsilon_o, \upsilon*) \leq (1 - \beta)^{-1}\rho(\upsilon_n, \upsilon_{n-1})$ •

The first statement insures that there is one and only one value function. For computational purposes, it is easiest to think of there being a finite set of possible $s \in S$, as then we can compute the value function exactly (if there are only a finite set of elements in S we say the cardinality of S is finite, or $\#S$ is finite). Actually often we can start out with S infinite and then prove that we will only observe a finite subset of those S.

The second statement gives you a surefire way to compute that value function (at least if we are willing to iterate indefinitely). Note that when we do this we might think a little about how to start off the iterations.

The third statement gives you a way of determining a greatest upper bound for the error for any given computation. That is the *l.h.s.* of the last inequality is just $\rho(\upsilon_n, \upsilon*)$, a quantity we would like to know in order to determine how far away we are from the $\upsilon*$ we are trying to approximate. The *r.h.s.* is a quantity we can measure, and hence allows us to bound how far we are away from the $\upsilon*$.
A simple proof of the last point follows.

$$\rho(\upsilon_m, \upsilon_n) \leq \rho(\upsilon_m, \upsilon_{m+1}) + \ldots + \rho(\upsilon_{n+1}, \upsilon_n)$$

(use the triangle inequality to prove this).

$$\leq (\beta^{m-1} + ... + \beta^n)\rho(\upsilon_1, \upsilon_o)$$

(from Theorem 1, since it insures $\rho(\upsilon_q, \upsilon_{q-1}) \leq \beta^{q-1}\rho(\upsilon_1, \upsilon_0)$)

$$= \frac{\beta^n - \beta^m}{1-\beta}\rho(\upsilon_1, \upsilon_o)$$

(wll we have done is summed the series). Now take limits as $m \to \infty$

$$\rho(\upsilon*, \upsilon_n) = lim_{m\to\infty}\rho(\upsilon_m, \upsilon_n) \leq (1-\beta)^{-1}\beta^n\rho(\upsilon_1, \upsilon_o)$$

Since this is true regardless of the initial condition (υ_o)

$$\rho(\upsilon*, \upsilon_n) \leq min_{j\leq n}(1-\beta)^{-1}\beta^{n-j}\rho(\upsilon_{j+1}, \upsilon_j) = (1-\beta)^{-1}\rho(\upsilon_{n+1}, \upsilon_n).\bullet$$

Corollary 1.
Let (B, ρ) be a complete metric space, and let $T : B \to B$ be a contraction mapping with fixed point $\nu \in S$. Then

1. if $B^{'} is$ a closed nonempty subset of B and $T(B^{'}) \subseteq B^{'}$, then $\nu \in B^{'}$.

2. if in addition $T(B^{'}) \subseteq B" \subseteq B^{'}$, then $v \in B"$ •.

The first part of the corollary is usually used as follows. Say we take a function which is in a closed subset of B (eg. the set of weakly increasing functions is a subset of the set of functions). Further assume that if, we take any member of the subset and apply T to it we obtain another member of the subset. Then the fixed point is a member of the subset (eg. it is weakly increasing). The second part assumes that anytime we apply the operator T to a member of the inial closed subset, we obtain a member of an open subset of the closed subset (eg. a member of the set of strictly increasing functions). If this is also true than the fixed point is also a member of the open subset.

Theorem 2. (Blackwell's sufficient conditions for a contraction)

Let $S \subseteq \mathbf{R}^l$, and let $B(S)$ be a space of bounded functions $f : S \to \mathbf{R}$, with the sup norm. Let $T : B(S) \to B(S)$ be an operator satisfying

1. Monotonicity. $f, g \in B(S)$ and $f(s) \leq g(s)$, for all $s \in S$, implies $(Tf)(s) \leq (Tg)(s)$, for all $s \in S$, and
2. Discounting. There exists some $\beta \in (0, 1)$ such that
   $[T(f + c)](s) \leq (Tf)(s) + \beta c$, all $f \in B(S)$, $c \geq 0$, $s \in S$. Here $(f + a)(s)$ is the function defined by $(f + a)(s) = f(s) + a$.

Then T is a contraction with modulus β . •

The rest of the theorems in this section are less central to the examples that come thereafter. Nevertheless they are often useful.

Sometimes the operator T that we will be applying to a function will involve choosing a policy (an x) from a correspondence of possible policies ($\Gamma(k)$) to maximize the value of the function we apply T to (in this case $f(\cdot)$). Moreover we will often work with the space of continuous bounded functions equipped with the sup norm. In such a case we will want to know whether our operator (our T), an operator which involves choosing a policy from this correspondence, takes continuous into continuous functions. We use the theorem of the maximum to insure that this is indeed the case. For the most part I will just assume the conditions of this theorem are met in applications. S.L.&P go through quite a bit of detail about when this is true, and you can read those details if you are concerned about a given application.

Theorem 3. (Theorem of the Maximum).
Let $K \subset \mathbf{R}^l$, $X \subset \mathbf{R}^m$, $f : K \times X \to \mathbf{R}$ be continuous, and $\Gamma : K \to X$ be a compact — valued and continous correspondence. Then the function $h : K \to \mathbf{R}$ defined in (*) is continuous, and the correspondence $G : K \to X$ in (**) is nonempty, compact valued, and upper hemi-continuous (an upper hemi-continuous correspondence that is single valued is continuous).

$$(*) : h(k) = max_{x \in \Gamma(k)} f(x, k).$$

$$(**) : G(k) = \{x \in \Gamma(k) : h(k) = f(x, k)\}.\bullet$$

Finally we are as concerned with our ability to compute the policy as with our ability to compute the value functions. Hence we want to insure that the policy derived from the appropriate iterate of the value function converges to the optimal policy. Here are some conditions which insure this is so. Again, S.L. &P provide quite a bit more detail.

Theorem 4. (Convergence of the Policy Function)
Let $K \subset \mathbf{R}^l$, $X \subset \mathbf{R}^m$ and assume that the correspondence $\Gamma : K \to X$ is nonempty, compact and convex — valued, and continuous. Assume that for each n and each $k \in K$, $f_n(k, x)$ is strictly concave in its 2nd argument and that f has the same properties, where $f_n \to f$ in the sup norm. (Recall that this implies that f_n and f are continuous.) Define g_n, and g, by

$$g_n(k) = argmax_{x \in \Gamma(k)} f_n(k, x), \quad n = 1, 2, ...$$

$$g(k) = argmax_{x \in \Gamma(k)} f(k, x)$$

Then $g_n \to g$ pointwise. If K is compact, $g_n \to g$, uniformly. •

END MATH PRELIMINARIES.