Fixed Point Theorems

Definition: Let X be a set and let $T : X \to X$ be a function that maps X into itself. (Such a function is often called an **operator**, a **transformation**, or a **transform** on X, and the notation Tx is often used in place of $T(x)$.) A **fixed point** of T is an element $x \in X$ for which $T(x) = x$.

Examples:

1. Let X be the two-element set $\{a, b\}$. The function $f : X \to X$ defined by $f(a) = b$ and $f(b) = a$ has no fixed point, but the other three functions that map X into itself each have one or two fixed points.

2. Let X be the unit interval $[0, 1]$ in $\mathbb{R}$. The graph of a function $f : X \to X$ is a subset of the unit square $X \times X$. If f is continuous, then its graph is a curve from the left edge of the square to the right edge. A fixed point of f is an element of $[0, 1]$ at which the graph of f intersects the 45°-line. Intuitively, it seems clear that if f is continuous then it must have a fixed point (its graph must cross or touch the 45°-line), and also that discontinuous functions f may not have a fixed point.

Remark: Note that the definition of a fixed point requires no structure on either the set X or the function T.

The Banach Fixed Point Theorem

Definition: Let (X, d) be a metric space. A **contraction** of X (also called a **contraction mapping** on X) is a function $f : X \to X$ that satisfies

$$\forall x, x' \in X : \; d(f(x'), f(x)) \leqq \beta d(x', x)$$

for some real number $\beta < 1$. Such a β is called a **contraction modulus** of f.

Example: Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable real function. If there is a real number $\beta < 1$ for which the derivative f' satisfies $|f'(x)| \leqq \beta$ for all $x \in \mathbb{R}$, then f is a contraction with respect to the usual metric on $\mathbb{R}$ and β is a modulus of contraction for f. This is a straightforward consequence of the Mean Value Theorem: let $x, x' \in \mathbb{R}$ and wlog assume $x < x'$; the MVT tells us there is a number $\xi \in (x, x')$ such that $f(x') - f(x) = f'(\xi)(x' - x)$ and therefore $|f(x') - f(x)| = |f'(\xi)||(x' - x)| \leqq \beta|(x' - x)|$. The same MVT argument establishes that if $\beta < 1$ and $f : (a, b) \to (a, b)$ satisfies $|f'(x)| \leqq \beta$ for all $x \in (a, b)$, then f is a contraction of (a, b).

Theorem: Every contraction mapping is continuous.

Proof: Let $T : X \to X$ be a contraction on a metric space (X, d), with modulus β, and let $\overline{x} \in X$. Let $\epsilon > 0$, and let $\delta = \epsilon$. Then $d(x, \overline{x}) < \delta \Rightarrow d(Tx, T\overline{x}) \leqq \beta\delta < \epsilon$. Therefore T is continuous at $\overline{x}$. Since $\overline{x}$ was arbitrary, T is continuous on X. $\|$

The above proof actually establishes that a contraction mapping is *uniformly* continuous:

Definition: Let (X, d_X) and (Y, d_Y) be metric spaces. A function $f : X \to Y$ is **uniformly continuous** if for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$\forall x, x' \in X : d_X(x, x') < \delta \Rightarrow d_Y(f(x), f(x')) < \epsilon.$$

Notice how this definition differs from the definition of continuity: uniform continuity requires that, for a given ϵ, a single δ will work across the entire domain of f. Continuity allows that the δ may depend upon the point x at which continuity of f is being evaluated; uniform continuity requires that (for a given ϵ) being within δ of *any* $x \in X$ guarantees that the image under f is within ϵ of $f(x)$.

Theorem: Every contraction mapping is uniformly continuous.

Banach Fixed Point Theorem: Every contraction mapping on a complete metric space has a unique fixed point. (This is also called the Contraction Mapping Theorem.)

Proof: Let $T : X \to X$ be a contraction on the complete metric space (X, d), and let β be a contraction modulus of T. First we show that T can have at most one fixed point. Then we construct a sequence which converges and show that its limit is a fixed point of T.

(a) Suppose x and x' are fixed points of T. Then $d(x, x') = d(Tx, Tx') \leqq \beta d(x, x')$; since $\beta < 1$, this implies that $d(x, x') = 0$, *i.e.*, $x = x'$.

(b) Let $x_0 \in X$, and define a sequence $\{x_n\}$ as follows:

$$x_1 = Tx_0, \quad x_2 = Tx_1 = T^2 x_0, \quad \ldots \quad , x_n = Tx_{n-1} = T^n x_0, \quad \ldots$$

We first show that adjacent terms of $\{x_n\}$ grow arbitrarily close to one another — specifically, that $d(x_n, x_{n+1}) \leqq \beta^n d(x_0, x_1)$:

$$\begin{aligned} d(x_1, x_2) &\leqq \beta d(x_0, x_1) \\ d(x_2, x_3) &\leqq \beta d(x_1, x_2) \leqq \beta^2 d(x_0, x_1) \\ &\cdots \\ d(x_n, x_{n+1}) &\leqq \beta d(x_{n-1}, x_n) \leqq \beta^n d(x_0, x_1). \end{aligned}$$

Next we show that if $n < m$ then $d(x_n, x_m) < \beta^n \frac{1}{1-\beta} d(x_0, x_1)$:

$$\begin{aligned}
d(x_n, x_{n+1}) &\leqq \beta^n d(x_0, x_1) \\
d(x_n, x_{n+2}) &\leqq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) \\
&\leqq \beta^n d(x_0, x_1) + \beta^{n+1} d(x_0, x_1) = (\beta^n + \beta^{n+1}) d(x_0, x_1) \\
&\cdots \\
d(x_n, x_m) &\leqq (\beta^n + \beta^{n+1} + \cdots + \beta^{m-1}) d(x_0, x_1) \\
&= \beta^n (1 + \beta + \beta^2 + \cdots + \beta^{m-1-n}) d(x_0, x_1) \\
&< \beta^n (1 + \beta + \beta^2 + \cdots) d(x_0, x_1) \\
&= \beta^n \frac{1}{1-\beta} d(x_0, x_1).
\end{aligned}$$

Therefore $\{x_n\}$ is Cauchy: for $\epsilon > 0$, let N be large enough that $\beta^N \frac{1}{1-\beta} d(x_0, x_1) < \epsilon$, which ensures that $n, m > N \Rightarrow d(x_n, x_m) < \epsilon$. Since the metric space (X, d) is complete, the Cauchy sequence $\{x_n\}$ converges to a point $x^* \in X$. We show that x^* is a fixed point of T: since $x_n \to x^*$ and T is continuous, we have $Tx_n \to Tx^*$ — *i.e.*, $x_{n+1} \to Tx^*$. Since $x_{n+1} \to x^*$ and $x_{n+1} \to Tx^*$, we have $Tx^* = x^*$. ‖

First Cournot Equilibrium Example: Two firms compete in a market, producing at output levels q_1 and q_2. Each firm responds to the other firm's production level when choosing its own level of output. Specifically (with a_1, a_2, b_1, b_2 all positive),

$$\begin{aligned}
q_1 &= r_1(q_2) = a_1 - b_1 q_2 \\
q_2 &= r_2(q_1) = a_2 - b_2 q_1
\end{aligned}$$

but $q_i = 0$ if the above expression for q_i is negative. The function $r_i : \mathbb{R}_+ \to \mathbb{R}_+$ is firm i's *reaction function*. Define $\overline{r} : \mathbb{R}^2_+ \to \mathbb{R}^2_+$ by $\overline{r}(q_1, q_2) = (r_1(q_2), r_2(q_1))$. The function $\overline{r}$ is a contraction with respect to the city-block metric if $b_1, b_2 < 1$:

$$\begin{aligned}
d(\overline{r}(q), \overline{r}(q')) &= |\overline{r}_1(q) - \overline{r}_1(q')| + |\overline{r}_2(q) - \overline{r}_2(q')| \\
&= |(a_1 - b_1 q_2) - (a_1 - b_1 q_2')| + |(a_2 - b_2 q_1) - (a_2 - b_2 q_1')| \\
&= b_1 |q_2' - q_2| + b_2 |q_1' - q_1| \\
&\leqq \max\{b_1, b_2\} (|q_1 - q_1'| + |q_2 - q_2'|) \\
&= \max\{b_1, b_2\}\, d(q, q').
\end{aligned}$$

Therefore we have an "existence and uniqueness result" for Cournot equilibrium in this example: $\overline{r}$ has a unique fixed point q^* — a unique Cournot equilibrium — if each $b_i < 1$.

There are several things to note about this example. First, note that while the condition $b_1, b_2 < 1$ is *sufficient* to guarantee the existence of an equilibrium, it is not necessary. Second, note that we could have obtained the same result, and actually calculated the equilibrium production levels, by simply solving the two "response" equations simultaneously. The condition $b_1, b_2 < 1$ is easily seen to be sufficient (and again, not necessary) to guarantee that the two-equation system has a solution. (Note that $b_1 b_2 \neq 1$ is in fact sufficient.) However, things might not be so simple if the response functions are not linear, or if there are more firms (and therefore more equations). We'll consider a second, nonlinear example shortly, and we'll use the Method of Successive Approximations to compute the equilibrium.

A third thing to note about the example is that we used the city-block metric instead of the Euclidean metric. This highlights an important and useful fact: a given function mapping a set X into itself may be a contraction according to one metric on X but not be a contraction according to other metrics. Recall that the definition of a fixed point, and therefore whether a given point in the function's domain *is* a fixed point, does not depend on the metric we're using, and in fact does not even require that X be endowed with any metric structure.

The Method of Successive Approximations: For any "initial" $x_0 \in X$, the sequence $x_n = Tx_{n-1}$ that we constructed in the proof of the Banach Fixed Point Theorem not only converges to the fixed point x^*, it converges to x^* *monotonically*: each successive term is closer to x^* than its predecessor was, as we show in the following proof. Consequently, starting from any arbitrary point in X, we can repeatedly apply the function T to the current "approximation" of x^*, obtaining a better approximation, and the approximations converge to x^*. This often provides a straightforward method for computing x^*.

Theorem: If T is a contraction on a complete metric space (X, d) and β is a contraction modulus of T, then for any $x \in X$,

$$\forall n \in \mathbb{N} : d(T^n x, x^*) \leqq \beta^n d(x, x^*),$$

where x^* is the unique fixed point of T.

Proof:

$$\begin{aligned} d(T^n x, x^*) &= d(TT^{n-1}x, Tx^*), \quad \text{because } x^* \text{ is a fixed point of } T \\ &\leqq \beta d(T^{n-1}x, x^*), \quad \text{because } T \text{ is a contraction} \\ &\leqq \beta^2 d(T^{n-2}x, x^*) \\ &\leqq \quad \cdots \quad \leqq \beta^n d(T^0 x, x^*) \\ &= \beta^n d(x, x^*). \ \| \end{aligned}$$

The Cournot Equilibrium Example Again: Suppose the current "state" of the market is $q(0) = (q_1(0), q_2(0))$ — Firm i is producing $q_i(0)$ units. Suppose further that "tomorrow" (at time $t = 1$) each firm chooses its production level $q_i(1)$ by responding to the amount its rival firm produced today (at time $t = 0$), and similarly at each subsequent date:

$$q_1(t+1) = r_1(q_2(t)) = a_1 - b_1 q_2(t) \qquad \text{and} \qquad q_2(t+1) = r_2(q_1(t)) = a_2 - b_2 q_1(t),$$

or more concisely, $q(t+1) = \overline{r}(q(t))$. This construction, called the "best reply dynamic," defines a sequence $\{q(t)\}$ of states. This is exactly the same construction we used in the proof of the Banach Theorem and the Method of Successive Approximations. The sequence will therefore converge monotonically (in distance) to the Cournot equilibrium q^*, because $\overline{r}$ is a contraction. Try it yourself by choosing values for the parameters a_1, a_2, b_1, b_2 — for example, $a_1 = 25, a_2 = 30, b_1 = 3/5, b_2 = 1/2$ — and any starting state $q(0)$. Plot the sequence along with the reaction curves in the q_1-q_2-space.

Second Cournot Equilibrium Example: Suppose the response functions in our Cournot example are

$$q_1 = r_1(q_2) = \frac{1}{2(1+q_2)} \qquad \text{and} \qquad q_2 = r_2(q_1) = \frac{1}{2}e^{-q_1}.$$

Now it's not so easy to calculate the equilibrium, or even to tell whether an equilibrium exists, as it was in the linear example, where we could simply solve the two response equations simultaneously. But it's possible to show that the function $\overline{r} : \mathbb{R}^2_+ \to \mathbb{R}^2_+$ defined as before is a contraction (so a Cournot equilibrium exists and is unique), and one can use the Method of Successive Approximations to compute an approximation to the equilibrium.

Exercise: Use Excel and the Method of Successive Approximations to compute the Cournot equilibrium to three decimal places in the Second Cournot Example.

The Brouwer Fixed Point Theorem

In the unit-interval example in the preceding section, the Banach Theorem seems somewhat limited. It seems intuitively clear that any continuous function mapping the unit interval into itself will have a fixed point, but the Banach Theorem applies only to functions f that satisfy $|f'(x)| \leqq \beta$ for some $\beta < 1$. An elementary example of this is the function $f(x) = 1-x$, which has an obvious fixed point at $x = 1/2$, but whose derivative satisfies $|f'(x)| = 1$ everywhere — therefore $d(f(x), f(x')) = d(x, x')$ — so f is not a contraction and the Banach Fixed Point Theorem doesn't apply to f. The fixed point theorem due to Brouwer covers this case as well as a great many others that the Banach Theorem fails to cover because the relevant functions aren't contractions

Brouwer Fixed Point Theorem: Let S be a nonempty, compact, convex subset of $\mathbb{R}^n$. Every continuous function $f : S \to S$ mapping S into itself has a fixed point.

The Brouwer Theorem requires only that f be continuous, not that it be a contraction, so there are lots of situations in which the Brouwer Theorem applies but the Banach Theorem doesn't. In particular, Brouwer's Theorem confirms our intuition that any continuous function mapping $[0,1]$ into itself has a fixed point, not just the functions that satisfy $|f'(x)| \leqq \beta$ for some $\beta < 1$. But conversely, the Banach Theorem doesn't require compactness or convexity — in fact, it doesn't require that the domain of f be a subset of a vector space, as this version of Brouwer's Theorem does. So there are also lots of situations where Banach's Theorem applies and Brouwer's doesn't.

Proofs of Brouwer's Theorem require some highly specialized mathematical ideas. For us, the benefit of developing these ideas in order to work through a proof of the theorem doesn't come close to justifying the time cost it would require, so we won't go there.

Here's a generalization of Brouwer's Theorem to normed vector spaces (which, in particular, don't have to be finite-dimensional, as Brouwer's Theorem requires):

Schauder Fixed Point Theorem: Let S be a nonempty, compact, convex subset of a normed vector space. Every continuous function $f : S \to S$ mapping S into itself has a fixed point.

This theorem would apply, for example, to any compact convex subset of $C[0,1]$, the vector space of continuous functions on the unit interval, with the max norm.