MATHEMATICS 3103 (Functional Analysis)
YEAR 2012–2013, TERM 2

HANDOUT #7: THE BAIRE CATEGORY THEOREM AND ITS CONSEQUENCES

We shall begin this last section of the course by returning to the study of general metric spaces, and proving a fairly deep result called the **Baire category theorem**.[1] We shall then apply the Baire category theorem to prove three fundamental results in functional analysis: the Uniform Boundedness Theorem, the Open Mapping Theorem, and the Closed Graph Theorem.

The Baire category theorem

Let X be a metric space. A subset $A \subseteq X$ is called **nowhere dense** in X if the interior of the closure of A is empty, i.e. $(\overline{A})^\circ = \varnothing$. Otherwise put, A is nowhere dense iff it is contained in a closed set with empty interior. Passing to complements, we can say equivalently that A is nowhere dense iff its complement contains a dense open set (why?).

Proposition 7.1 *Let X be a metric space. Then:*

(a) Any subset of a nowhere dense set is nowhere dense.

(b) The union of finitely many nowhere dense sets is nowhere dense.

(c) The closure of a nowhere dense set is nowhere dense.

(d) If X has no isolated points, then every finite set is nowhere dense.

Proof. (a) and (c) are obvious from the definition and the elementary properties of closure and interior.

To prove (b), it suffices to consider a *pair* of nowhere dense sets A_1 and A_2, and prove that their union is nowhere dense (why?). It is also convenient to pass to complements, and prove that the intersection of two dense open sets V_1 and V_2 is dense and open (why is this equivalent?). It is trivial that $V_1 \cap V_2$ is open, so let us prove that it is dense. Now, a subset is dense iff every nonempty open set intersects it. So fix any nonempty open set $U \subseteq X$. Then $U_1 = U \cap V_1$ is open and nonempty (why?). And by the same reasoning, $U_2 = U_1 \cap V_2 = U \cap (V_1 \cap V_2)$ is open and nonempty as well. Since U was an *arbitrary* nonempty open set, we have proven that $V_1 \cap V_2$ is dense.

To prove (d), it suffices to note that a one-point set $\{x\}$ is open if and only if x is an isolated point of X; then use (b). □

[1]Proved (for $\mathbb{R}^n$) by the French mathematician René-Louis Baire (1874–1932) in his 1899 doctoral thesis. Baire made a number of important contributions to real analysis in addition to the category theorem. However, it turns out that the Baire category theorem for the real line was actually proved two years earlier, in 1897, by the American mathematician William Fogg Osgood (1864–1943)!

(a) and (b) can be summarized by saying that the nowhere dense sets form an **ideal** of sets.

Example. The **Cantor ternary set** C consists of all real numbers in the interval $[0,1]$ that can be written as a ternary (base-3) expansion in which the digit 1 does not occur, i.e. $x = \sum_{n=1}^{\infty} a_n/3^n$ with $a_n \in \{0,2\}$ for all n. Equivalently, C can be constructed from $[0,1]$ by deleting the open middle third of the interval $[0,1]$, then deleting the open middle thirds of each of the intervals $[0,1/3]$ and $[2/3,1]$, and so forth. If C_n denotes the union of the 2^n closed intervals of length $1/3^n$ that remain at the nth stage, then $C = \bigcap_{n=1}^{\infty} C_n$. It follows immediately that C is closed (and indeed compact). Moreover, since C_n contains no open interval of length greater than $1/3^n$, it follows that C contains no open interval at all, i.e. C has empty interior and hence is nowhere dense. □

Although the union of *finitely* many nowhere dense sets is nowhere dense, the union of *countably* many nowhere dense sets need not be nowhere dense: for instance, in $X = \mathbb{R}$, the rationals $\mathbb{Q}$ are the union of countably many nowhere dense sets (why?), but the rationals are certainly not nowhere dense (indeed, they are *everywhere* dense, i.e. $(\overline{\mathbb{Q}})^\circ = \overline{\mathbb{Q}} = \mathbb{R}$).

This observation motivates the introduction of a larger class of sets: A subset $A \subseteq X$ is called **meager** (or of **first category**) in X if it can be written as a countable union of nowhere dense sets. Any set that is not meager is said to be **nonmeager** (or of **second category**). The complement of a meager set is called **residual**.

We then have as an immediate consequence:

Proposition 7.2 *Let X be a metric space. Then:*

(a) Any subset of a meager set is meager.

(b) The union of countably many meager sets is meager.

(c) If X has no isolated points, then every countable set is meager.

(a) and (b) can be summarized by saying that the meager sets form a **σ-ideal** of sets. In a certain topological sense, the meager sets can be considered "small" and even "negligible".

Warnings. 1. Be careful of the terminology, which can be confusing. The meager sets are in some sense "small". The residual sets are in some sense "large" (i.e. their complements are "small"). But the second category sets are not necessarily "large"; they are merely "not small".

2. Note also that "meager", "nonmeager" and "residual" are attributes not of a set A in and of itself, but of a set A *in the metric space* X. Which category a set has depends on the space within which it is considered. For example, a line is residual (and, we will soon show, nonmeager) inside itself, but it is nowhere dense (and hence meager) inside a plane. Similarly, $\mathbb{Z}$ is residual and nonmeager inside itself — indeed, in $\mathbb{Z}$ every set is open (why?), so the only meager set is $\varnothing$ (why?) — but $\mathbb{Z}$ is nowhere dense (and hence meager) inside $\mathbb{R}$. □

Remark. If you have studied Measure Theory, then you have encountered another important σ-ideal of sets in $\mathbb{R}$ or $\mathbb{R}^n$, namely the sets of **Lebesgue measure zero** (also called **null sets**). These sets are "negligible" in the measure-theoretic sense.

The meager sets and the measure-zero sets thus constitute two σ-ideals, each of which properly contains the σ-ideal of countable sets.[2] It is natural to ask whether these properties are related. For instance, does one of the two classes contain the other? It turns out that the answer is no, and that the two notions of "smallness" can in some cases even be diametrically opposed. In fact, it is not difficult to prove that the real line can be decomposed into two complementary sets, one of which is meager and the other of which has measure zero. So a set that is "small" in one of these senses can be very "big" in the other sense, and vice versa. There is of course nothing intrinsically paradoxical about the fact that a set that is small in one sense may be large in another sense.

But despite the fact that there is no necessary *relation* between the properties of being meager and measure zero, it turns out that there is an striking *analogy* between these properties. Many (but not all) theorems about meager sets have analogues (albeit sometimes with quite different proofs) for sets of measure zero, and conversely. A beautiful (and very readable) book about this analogy is John C. Oxtoby, *Measure and Category*. □

We are now ready to state the Baire category theorem:

Theorem 7.3 (Baire category theorem) *Let X be a* complete *metric space. Then:*

(a) A meager set has empty interior.

(b) The complement of a meager set is dense. (That is, a residual set is dense.)

(c) A countable intersection of dense open sets is dense.

You should carefully verify that (a), (b) and (c) are equivalent statements, obtained by taking complements.

In applications we frequently need only the weak form of the Baire category theorem that is obtained by weakening "is dense" in (b,c) to "is nonempty" (which is valid whenever X is itself nonempty):

Corollary 7.4 (weak form of the Baire category theorem) *Let X be a nonempty* complete *metric space. Then:*

(b) X cannot be written as a countable union of nowhere dense sets. (In other words, X is nonmeager in itself.)

(b′) If X is written as a countable union of closed sets, then at least one of those closed sets has nonempty interior.

(c) A countable intersection of dense open sets is nonempty.

---

[2] To see that the containment is proper, it suffices to observe that the Cantor ternary set is meager, measure-zero (why?) and uncountable (why?).

As preparation for the proof of the Baire category theorem, let us prove a useful lemma due to Georg Cantor.[3] Recall first that if X is a *compact* metric space, then any decreasing sequence $F_1 \supseteq F_2 \supseteq F_3 \supseteq \ldots$ of nonempty closed sets has a nonempty intersection (why?). This is *not* true in general in a noncompact metric space (even one that is complete and separable): for instance, in $X = \mathbb{R}$ (or $\mathbb{N}$) consider $F_n = [n, \infty)$. But if we add the additional hypothesis that the diameters of the sets F_n tend to zero, then it is true, provided only that X is complete:

Lemma 7.5 (Cantor) *Let X be a complete metric space, and let $F_1 \supseteq F_2 \supseteq F_3 \supseteq \ldots$ be a decreasing sequence of nonempty closed subsets of X, with* $\operatorname{diam} F_n \to 0$. *Then there exists a point $x \in X$ such that $\bigcap_{n=1}^{\infty} F_n = \{x\}$. In particular, $\bigcap_{n=1}^{\infty} F_n \neq \varnothing$.*

Proof. In each set F_n choose a point x_n. Then the sequence (x_n) is Cauchy: for if $m, n \geq N$ we have $d(x_m, x_n) \leq \operatorname{diam} F_N$ (why?), which tends to zero as $N \to \infty$. Since X is complete, the sequence (x_n) has a limit x. But since $x_n \in F_N$ for all $n \geq N$, and F_N is closed, we have $x \in F_N$. Since this holds for all N, we have $x \in \bigcap_{N=1}^{\infty} F_N$. But since $\operatorname{diam}\Big(\bigcap_{N=1}^{\infty} F_N\Big) \leq \inf_{N \geq 1} \operatorname{diam}(F_N) = 0$, we must have $\bigcap_{n=1}^{\infty} F_n = \{x\}$. $\square$

Proof of the Baire category theorem. It is easily seen that (a) and (b) are equivalent statements. Now, if $A_1, A_2, \ldots$ is a sequence of nowhere dense sets, then $\overline{A_1}, \overline{A_2}, \ldots$ is a sequence of nowhere dense closed sets (why?) satisfying $\cup A_n \subseteq \cup \overline{A_n}$; so to prove (b) it clearly suffices to prove that a countable union of nowhere dense *closed* sets has a dense complement. But this is equivalent, by complementation, to the assertion that a countable intersection of dense open sets is dense, i.e. statement (c). So this is what we shall prove.

The proof follows the same idea as in the proof of Proposition 7.1, but with infinitely many steps rather than just two, and with some minor modifications to allow us to exploit Cantor's lemma. Let $V_1, V_2, \ldots$ be the given sequence of dense open sets, and fix any nonempty open set $U \subseteq X$. Fix also a sequence (a_n) of positive numbers tending to zero (say, $a_n = 1/n$). The set $U \cap V_1$ is open and nonempty (why?), so we can choose a closed ball $\overline{B}(x_1, r_1)$ inside it, with radius $0 < r_1 < a_1$. Then $B(x_1, r_1) \cap V_2$ is open and nonempty (why?), so we can choose a closed ball $\overline{B}(x_2, r_2)$ inside it, with radius $0 < r_2 < a_2$. And so forth: at the nth stage we observe that $B(x_{n-1}, r_{n-1}) \cap V_n$ is open and nonempty, so we can choose a closed ball $\overline{B}(x_n, r_n)$ inside it, with radius $0 < r_n < a_n$. Then Cantor's lemma tells us that the sequence of closed balls $\overline{B}(x_n, r_n)$ has nonempty intersection (why?). But $\overline{B}(x_n, r_n) \subseteq U \cap V_n$ (why?), so it follows that $U \cap \Big(\bigcap_{n=1}^{\infty} V_n\Big)$ is nonempty. Since U was an arbitrary nonempty open set, we have proven that $\bigcap_{n=1}^{\infty} V_n$ is dense.[4] $\square$

---

[3]German mathematician Georg Cantor (1845–1918) revolutionized mathematics by founding, almost single-handedly, the modern theory of infinite sets. He also made important contributions to real analysis and number theory.

[4]If you worry about set-theoretic questions, you may have observed that some form of the axiom of choice was invoked implicitly in this proof, in order to choose the closed ball $\overline{B}(x_n, r_n)$. It turns out that

Example. We have seen that the rationals $\mathbb{Q}$ form a meager subset of the real line. What about the irrationals $\mathbb{R} \setminus \mathbb{Q}$? Are they meager? The answer is not immediately obvious from the definition of "meager" (we would have to consider all possible ways of writing $\mathbb{R} \setminus \mathbb{Q}$ as a countable union). But the answer follows immediately from the Baire category theorem: for if $\mathbb{R} \setminus \mathbb{Q}$ were meager, then so would be $\mathbb{Q} \cup (\mathbb{R} \setminus \mathbb{Q}) = \mathbb{R}$, contradicting the (weak form of the) Baire category theorem. $\square$

Remark. When I said at the beginning of these notes that the Baire category theorem is a "fairly deep result", I did not mean that its proof is especially difficult — as you have seen, it is not (the proof of the equivalence of the three notions of compactness was more difficult in my opinion). But what is deep is the mere *idea* to *consider* the countable unions of nowhere dense sets. This was a stroke of genius on Baire's (and Osgood's) part, and it has had enormously powerful consequences in both real analysis and functional analysis — of which we will see merely a few examples.

Some applications of the Baire category theorem

Before beginning the applications to functional analysis, I would like to give just a little bit of the flavor of the results that can be obtained in real analysis using the Baire category theorem.

Example 1: $\boldsymbol{F_\sigma}$ and $\boldsymbol{G_\delta}$ sets. Let X be a metric space. A subset $A \subseteq X$ is called an **$\boldsymbol{F_\sigma}$ set** if it can be written as a countable union of closed sets. It is called a **$\boldsymbol{G_\delta}$ set** if it can be written as a countable intersection of open sets. Clearly, a set is F_σ if and only if its complement is G_δ. Also, in a metric space (though not in a general topological space) every closed set is a G_δ: it suffices to observe that if A is closed, then $A = \bigcap_{n=1}^{\infty} \{x \colon d(x, A) < 1/n\}$ (why?) and that the latter sets are open (why?). By the same token, every open set is an F_σ.

Now let $X = \mathbb{R}$. The rationals $\mathbb{Q}$ are clearly an F_σ (why?) and hence the irrationals $\mathbb{R} \setminus \mathbb{Q}$ are a G_δ. What about the reverse? Are the rationals also a G_δ (and hence the irrationals also an F_σ)?

It is not so easy to answer these questions by elementary methods. But using the Baire category theorem one can give an easy answer "no". We know that the rationals are an F_σ;

what was needed is a weak form of the axiom of choice called "the axiom of dependent choices" (see e.g. `http://en.wikipedia.org/wiki/Baire_category_theorem`). Indeed, logicians have proven that the axiom of dependent choices is *equivalent* in Zermelo–Fraenkel (ZF) set theory to the Baire category theorem.

On the other hand, the Baire category theorem for *separable* metric spaces can be proven in ZF set theory *without* any choice axiom: it suffices to fix a dense sequence (y_n) in X and an enumeration (q_n) of the positive rational numbers, and then choose (x_n, r_n) to be the *first* (in lexicographic order) pair (y_k, q_k) that has the desired property.

So the oft-made statement that the Baire category theorem provides *nonconstructive* existence proofs is not quite right: when applied to *separable* metric spaces (which are indeed the principal applications), the Baire category method provides *in principle* a construction by successive approximation of the elements of the space that are claimed to exist. But this construction may be too complicated to be useful, so *for all practical purposes* the method may be treated as nonconstructive.

so if also the irrationals were an F_σ, then the whole real line could be written as a countable union of closed sets, each one of which is entirely contained either in $\mathbb{Q}$ or in $\mathbb{R} \setminus \mathbb{Q}$. But Corollary 7.4b′ to the Baire category theorem tells us that one of these closed sets would have to have nonempty interior, which contradicts the fact that it is contained in $\mathbb{Q}$ or $\mathbb{R} \setminus \mathbb{Q}$.

We can also make the following trivial but useful observation:

Proposition 7.6 *Let X be a metric space. Then an F_σ subset $A \subseteq X$ is either meager or else has nonempty interior.*

PROOF. By hypothesis A can be written as a countable union of closed sets. If all of these closed sets have empty interior, then A is meager; if at least one of them has nonempty interior, then so does A. □

Remark. Proposition 7.6 holds whether or not X is complete (all we used was the *definition* of "meager"). But if X is not complete, then a meager set need not be "small"; indeed, it could be all of X! (E.g. if $X = \mathbb{Q}$ considered with the metric inherited from $\mathbb{R}$.) In particular, when X is not complete, the two alternatives given in Proposition 7.6 need not be mutually exclusive: a set could be meager *and* have nonempty interior. On the other hand, if X is complete, then the meager sets really are "small", i.e. they have empty interior (or equivalently, dense complement).

Example 2: Sets of continuity and discontinuity. On $X = \mathbb{R}$, consider the function f defined by

$$f(x) \;=\; \begin{cases} 1/q & \text{if } x = p/q \text{ in lowest terms} \\ 0 & \text{if } x \text{ is irrational} \end{cases} \tag{7.1}$$

It is easy to see that f is continuous at the irrationals (why?) and discontinuous at the rationals (why?). Does there exist a function with the reverse property, i.e. one that is continuous at the rationals and discontinuous at the irrationals?

After spending an hour or two trying to construct such a function, one begins to suspect that no such function can exist. But how to prove it? Once again, the Baire category theorem comes to the rescue.

In order to measure quantitatively the continuity or discontinuity of a function, we introduce the concept of *oscillation*. Suppose that f is a map from a metric space X into another metric space Y. For any subset $S \subseteq X$, we define the **oscillation of f on S** as

$$\omega_f(S) \;=\; \operatorname{diam}(f[S]) \;=\; \sup_{x,x' \in S} d_Y(f(x), f(x')) \tag{7.2}$$

(its value is a nonnegative real number or $+\infty$). Then, for any $x \in X$, we define the **oscillation of f at x** as

$$\omega_f(x) \;=\; \inf_{\epsilon > 0} \omega_f(B(x, \epsilon)) \tag{7.3}$$

(its value is again a nonnegative real number or $+\infty$). Clearly f is continuous at x_0 if and only if $\omega_f(x_0) = 0$ (why?). When f is discontinuous at x_0, $\omega_f(x_0)$ gives a quantitative measure of the size of the discontinuity.

Note that if $\omega_f(x_0) < c$ (where c is some real number), then $\omega_f(x) < c$ for all x in a sufficiently small neighborhood of x_0 (why?).[5] So the set $\{x\colon \omega_f(x) < c\}$ is open. But then the set C of points of continuity of f can be written as

$$C = \bigcap_{n=1}^{\infty} \{x\colon \omega_f(x) < 1/n\} \tag{7.4}$$

and hence is a G_δ. We have therefore proven:

Theorem 7.7 *Let f be any map from a metric space X into another metric space Y. Then the set of points of continuity of f is a G_δ, and the set of points of discontinuity is an F_σ.*

Putting this together with the result of Example 1 that the rationals are *not* a G_δ subset of $\mathbb{R}$, we conclude that there does *not* exist a map from $\mathbb{R}$ into $\mathbb{R}$ (or indeed into any metric space) that is continuous at the rationals and discontinuous at the irrationals.

We also have the following corollary showing that the set of points of discontinuity of a function f is either "small" (i.e. meager) or else "somewhat large" (i.e. contains a nonempty open set):

Corollary 7.8 *Let f be any map from a metric space X into another metric space Y. Then the set of points of discontinuity of f is either meager or else has nonempty interior.*

PROOF. This follows immediately from Theorem 7.7 and Proposition 7.6. □

Remarks. 1. The nonexistence of a function that is continuous at the rationals and discontinuous at the irrationals can alternatively be deduced from Corollary 7.8, by observing that the irrationals have empty interior but are not meager.

2. Theorem 7.7 and Corollary 7.8 hold whether or not X is complete. The remarks given after Proposition 7.6 apply also here. □

Example 3: Nowhere-differentiable continuous functions. Most of the continuous functions studied in elementary calculus are differentiable everywhere — or at worst, differentiable everywhere except at some finite set of points (e.g. the function $|x|$ is nondifferentiable at $x = 0$). So it was something of a surprise when Weierstrass published in 1872 his famous example of a function that is continuous everywhere and differentiable nowhere.[6]

---

[5] In other words, ω_f is an *upper semicontinuous* function on X with values in $[0, +\infty]$.

[6] Weierstrass' example was

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(b^n x)$$

where $0 < a < 1$, b is an odd positive integer, and $ab > 1 + (3\pi/2)$. The condition $0 < a < 1$ guarantees that the series is uniformly convergent and hence that f is continuous; the remaining two conditions on b are used by Weierstrass to prove that f is nowhere differentiable. But these limitations on the allowed values of b look a bit unnatural, since intuitively one expects that taking b to be any *real* number with $ab > 1$ should suffice to make f nowhere differentiable, because the formal series for $f'(x)$ then looks likely to diverge for all x (or maybe only for most x? that is the trouble). This intuitive guess does turn out to be correct, but it took over four decades after Weierstrass' work for this to be proven. Finally, in 1916 the celebrated English mathematician G.H. Hardy (1877–1947) proved that f is nowhere differentiable whenever $ab \geq 1$.

But applying the Baire category theorem provides an even bigger surprise: it turns out that "most" continuous functions are nowhere differentiable! Here, of course, "most" has to be understood in the sense of Baire category: the precise claim is that the functions that are differentiable at at least one point form a meager set in the space $\mathcal{C}[0,1]$ of continuous functions. Indeed, the functions that have a finite one-sided derivative at at least one point, or even a bounded difference quotient on at least one side at at least one point, form a meager set in $\mathcal{C}[0,1]$. Let us prove this, as follows:

First let us be precise about what we are asserting. We say that a function $f \in \mathcal{C}[0,1]$ has a **bounded right difference quotient** at $x \in [0,1)$ in case

$$\limsup_{h \downarrow 0} \left| \frac{f(x+h)-f(x)}{h} \right| \;<\; \infty \;, \tag{7.5}$$

or in other words

$$\text{there exists } n < \infty \text{ and } \epsilon > 0 \text{ such that } \left| \frac{f(x+h)-f(x)}{h} \right| \;\le\; n \text{ for all } h \in (0,\epsilon] \;. \tag{7.6}$$

So, for $x \in [0,1)$, $h \in (0, 1-x]$ and $n \in \mathbb{N}$, let us define

$$A_{x,h,n} \;=\; \left\{ f \in \mathcal{C}[0,1] \colon \left| \frac{f(x+h)-f(x)}{h} \right| \le n \right\} \;. \tag{7.7}$$

This is the set of continuous functions on $[0,1]$ whose difference quotient at x with step $+h$ is bounded in absolute value by n. The set of continuous functions on $[0,1]$ that have a bounded right difference quotient at at least one point is therefore

$$A \;=\; \bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcup_{0 \le x < 1} \bigcap_{0 < h \le \min(1-x, 1/m)} A_{x,h,n} \tag{7.8a}$$

$$\;=\; \bigcup_{n=1}^{\infty} \bigcup_{0 \le x \le 1-1/n} \bigcap_{0 < h \le 1/n} A_{x,h,n} \tag{7.8b}$$

(why is the second form equivalent to the first?). So let us define

$$A_n \;=\; \bigcup_{0 \le x \le 1-1/n} \bigcap_{0 < h \le 1/n} A_{x,h,n} \;. \tag{7.9}$$

We will prove that A_n is closed and nowhere dense, and hence that $A = \bigcup_{n=1}^{\infty} A_n$ is meager.

Proof that A_n is closed. Consider a sequence (f_k) in A_n converging (in sup norm) to $f \in \mathcal{C}[0,1]$. For each k there exists a point $x_k \in [0, 1-1/n]$ such that

$$|f_k(x_k+h) - f_k(x_k)| \;\le\; nh \quad \text{for all } h \in (0, 1/n] \;. \tag{7.10}$$

Then, by compactness, there exists a subsequence of (x_k) converging to some point $x \in [0, 1-1/n]$; so, replacing (f_k) and (x_k) by this subsequence, we may assume for simplicity of notation that the original sequence (x_k) tends to x. Taking $k \to \infty$ in (7.10), we obtain[7]

$$|f(x+h) - f(x)| \;\le\; nh \quad \text{for all } h \in (0, 1/n]\,, \tag{7.11}$$

which shows that $f \in A_n$.

PROOF THAT A_n HAS EMPTY INTERIOR. Why does A_n not contain any open ball? The reason is that near (in sup norm) to any continuous function there is another continuous function whose slope is very badly behaved. To prove this, we shall first observe that near any continuous function there is another continuous function whose slope is *well* behaved, and then modify this latter function so that the slope is badly behaved. (This may seem a rather contorted way of doing things, but you will see why it is necessary.)

So, the first step is to observe that the Lipschitz functions — i.e. those for which there exists $M < \infty$ such that $|g(x) - g(y)| \le M|x-y|$ for all $x, y \in [0,1]$ — are dense in $\mathcal{C}[0,1]$. This follows immediately from the Weierstrass approximation theorem, because the polynomials are obviously Lipschitz (why?); or it follows in a more elementary way by using the fact that every continuous function on $[0,1]$ is uniformly continuous, so we can uniformly approximate it by a *piecewise linear* function (you should supply the details of this proof). Either way, we conclude that given any $f \in \mathcal{C}[0,1]$ and any $\epsilon > 0$, we can find $g \in \mathcal{C}[0,1]$ and $M < \infty$ such that $\|f - g\|_\infty < \epsilon$ and $|g(x) - g(y)| \le M|x-y|$ for all $x, y \in [0,1]$.

Now, for each positive integer N, let $T_N \in \mathcal{C}[0,1]$ be a triangular wave of amplitude 1 and half-period $1/N$, i.e. the function that takes the values $T_N(j/N) = (-1)^j$ for j integer and that is defined to be linear on each interval $[j/N, (j+1)/N]$. On each such interval, T_N has slope $\pm 2N$; so given any point $x \in [0,1)$ we have

$$\left| \frac{T_N(x+h) - T_N(x)}{h} \right| \;=\; 2N \tag{7.12}$$

for all sufficiently small $h > 0$ (why? can you see what determines how small h has to be?). It follows that the function $g + \epsilon T_N$ has right difference quotient at least $2\epsilon N - M$ in magnitude at every point of $[0,1)$ (why?). So if we choose $N > (M+n)/(2\epsilon)$, we conclude that $g + \epsilon T_N \notin A_n$. Since $\|f - (g + \epsilon T_N)\|_\infty < 2\epsilon$, we have proven that A_n does not contain the open ball $B(f, 2\epsilon)$. But since $f \in \mathcal{C}[0,1]$ and $\epsilon > 0$ were arbitrary, we have proven that A_n does not contain any open ball, i.e. A_n has empty interior.

This completes the proof that $A = \bigcup\limits_{n=1}^{\infty} A_n$ is meager.

Of course, a completely analogous proof works for left rather than right difference quotients. Since the union of two meager sets is meager, we conclude that the functions that have a bounded difference quotient on at least one side at at least one point form a meager set in $\mathcal{C}[0,1]$. $\square$

---

[7] Here we use the following simple lemma: If f_k converges *uniformly* (i.e. in sup norm) to f, and x_k converges to x, then $f_k(x_k)$ converges to $f(x)$. This follows from

$$\begin{aligned} |f(x) - f_k(x_k)| \;&\le\; |f(x) - f(x_k)| \;+\; |f(x_k) - f_k(x_k)| \\ &\le\; |f(x) - f(x_k)| \;+\; \|f - f_k\|_\infty \end{aligned}$$

and the continuity of the limiting function f.

Several other interesting applications of the Baire category theorem can be found in Giles, *Introduction to the Analysis of Normed Linear Spaces*, Section 9; and many fascinating applications can be found in Oxtoby, *Measure and Category*.

The uniform boundedness theorem

Let X and Y be normed linear spaces, and let $\mathcal{F} \subseteq \mathcal{B}(X,Y)$ be a family of bounded (i.e. continuous) linear maps from X to Y. We say that the family $\mathcal{F}$ is **bounded** (or **uniformly bounded**) if it is bounded as a subset of the normed linear space $\mathcal{B}(X,Y)$, i.e. if $\{\|T\|_{X\to Y}\colon T \in \mathcal{F}\}$ is a bounded set of real numbers. We say that the family $\mathcal{F}$ is **pointwise bounded** if, for each $x \in X$, the set $\{Tx\colon T \in \mathcal{F}\}$ is bounded as a subset of the normed linear space Y, i.e. if $\{\|Tx\|_Y\colon T \in \mathcal{F}\}$ is a bounded set of real numbers. Clearly a bounded family is pointwise bounded (why?). The rather surprising fact is that the converse is also true, provided that X is complete. This important result is known as the **uniform boundedness theorem** (or **principle of uniform boundedness** or **Banach–Steinhaus theorem**):[8]

Theorem 7.9 (uniform boundedness theorem) *Let X be a Banach space and Y a normed linear space, and let $\mathcal{F}$ be a family of bounded linear maps from X to Y. If $\mathcal{F}$ is pointwise bounded, then it is bounded.*

The key step in the proof of the uniform boundedness theorem is the following lemma, which is based on the Baire category theorem. We say that a subset A of a normed linear space is **symmetric** if $x \in A$ implies $-x \in A$.

Lemma 7.10 *Let X be a Banach space, and let C be a closed convex symmetric subset of X satisfying $\bigcup\limits_{n=1}^{\infty} nC = X$. Then C is a neighborhood of 0 [that is, C contains an open ball $B(0,\epsilon)$ for some $\epsilon > 0$].*

Proof. By the weak form of the Baire category theorem (Corollary 7.4b′), one of the sets nC must have nonempty interior, hence C must have nonempty interior (why?). So C contains some open ball $B(x,\epsilon)$ with $\epsilon > 0$. Since C is symmetric, C also contains the open ball $B(-x,\epsilon)$. And then, since C is convex, C also contains the open ball $B(0,\epsilon)$ (why?). □

Remark. Lemma 7.10 actually holds without the hypothesis that C is symmetric (though we will not need this fact). To see this, it suffices to apply Lemma 7.10 to the set $D = C \cap (-C)$. Clearly D is closed, convex and symmetric (why?). To see that $\bigcup\limits_{n=1}^{\infty} nD = X$, observe first that since $0 \in C$ (why?) and C is convex, for each $x \in X$ there exists an integer n_x such that $x \in nC$ for *all* $n \geq n_x$ (why?); so if $n \geq \max(n_x, n_{-x})$ we have $x \in nD$. □

[8]The uniform boundedness theorem was first proven in 1922, independently by the Austrian mathematician Hans Hahn (1879–1934) and the Polish mathematician Stefan Banach (1892–1945). The proof using the Baire category theorem was published in 1927 by Banach together with his Polish colleague Hugo Steinhaus (1887–1972); the idea had been suggested by another Polish mathematician, Stanisław Saks (1897–1942).

Proof of the uniform boundedness theorem. Let

$$C = \{x \in X\colon \|Tx\|_Y \le 1 \text{ for all } T \in \mathcal{F}\} . \tag{7.13}$$

C is closed because each $T \in \mathcal{F}$ is continuous (how is the “for all” handled?). C is symmetric (why?) and convex (why?). Finally, the pointwise boundedness of $\mathcal{F}$ tells us that for each $x \in X$ there exists an integer n (depending on x) such that $\|Tx\|_Y \le n$ for all $T \in \mathcal{F}$, or in other words $x \in nC$. This means that $\bigcup\limits_{n=1}^{\infty} nC = X$. Lemma 7.10 then tells us that C contains an open ball $B(0, \epsilon)$ for some $\epsilon > 0$. It follows that $\|Tx\|_Y \le 1$ for all $T \in \mathcal{F}$ whenever $\|x\|_X < \epsilon$ [and hence also whenever $\|x\|_X \le \epsilon$ — why?], or in other words $\|Tx\|_Y \le \epsilon^{-1}\|x\|_X$ for all $T \in \mathcal{F}$ and all $x \in X$, or in other words $\|T\|_{X\to Y} \le \epsilon^{-1}$ for all $T \in \mathcal{F}$. This shows that the family $\mathcal{F}$ is bounded. $\square$

Example. Here is a simple example showing that the conclusion of the uniform boundedness theorem can fail if X is incomplete: Let X be the space c_{00} of sequences $x = (x_1, x_2, \ldots)$ with at most finitely many nonzero entries, equipped with the sup norm (i.e. considered as a linear subspace of ℓ^∞), and take $Y = \mathbb{R}$. For $n = 1, 2, \ldots$, let $\ell_n \in X^*$ be defined by $\ell_n(x) = nx_n$. Then for any $x \in X$ there is an integer N_x such that $x_n = 0$ for $n > N_x$, so we have

$$|\ell_n(x)| \le N_x\|x\|_\infty \quad \text{for all } n . \tag{7.14}$$

This shows that the family $\{\ell_n\}_{n=1}^\infty$ is pointwise bounded. But

$$\|\ell_n\| = n \quad \text{for each } n \tag{7.15}$$

(why?), so the family $\{\ell_n\}_{n=1}^\infty$ is not bounded. $\square$

Remark. In the proof of Lemma 7.10 and hence the uniform boundedness theorem, we did not really use as such the assumption that X is complete; all we used is the fact that it is nonmeager in itself (where did we use this?). So the uniform boundedness theorem holds under this weaker hypothesis. There do exist incomplete normed linear spaces that are nonmeager in themselves, so this extra generality is not vacuous; but such spaces rarely arise in applications. Please observe that the example space c_{00} is indeed meager in itself (why?). $\square$

The standard textbook proof of the uniform boundedness theorem is the one I have just presented, using the Baire category theorem. As you have just seen, this proof is quite simple; but its reliance on the Baire category theorem makes it not completely elementary.

By contrast, the original proofs of the uniform boundedness theorem given by Hans Hahn and Stefan Banach in 1922 were quite different: they began from the assumption that $\sup\limits_{T\in\mathcal{F}} \|T\| = \infty$ and used a “gliding hump” technique to construct a sequence (T_n) in $\mathcal{F}$ and a point $x \in X$ such that $\lim\limits_{n\to\infty} \|T_n x\| = \infty$. These proofs are elementary, but the details are a bit fiddly. Here is a *really* simple proof along similar lines, which I discovered three years ago while preparing this course[9]:

Lemma. Let T be a bounded linear operator from a normed linear space X to a normed linear space Y. Then for any $x \in X$ and $r > 0$, we have

$$\sup_{x' \in B(x,r)} \|Tx'\| \;\geq\; \|T\| r \;. \tag{7.16}$$

Proof. For $\xi \in X$ we have

$$\max\{\|T(x+\xi)\|, \|T(x-\xi)\|\} \;\geq\; \tfrac{1}{2}[\|T(x+\xi)\| + \|T(x-\xi)\|] \;\geq\; \|T\xi\| \;, \tag{7.17}$$

where the second $\geq$ uses the triangle inequality in the form $\|\alpha - \beta\| \leq \|\alpha\| + \|\beta\|$. Now take the supremum over $\xi \in B(0,r)$. $\square$

Proof of the uniform boundedness theorem. Suppose that $\sup_{T \in \mathcal{F}} \|T\| = \infty$, and choose $(T_n)_{n=1}^\infty$ in $\mathcal{F}$ such that $\|T_n\| \geq 4^n$. Then set $x_0 = 0$, and for $n \geq 1$ use the lemma to choose inductively $x_n \in X$ such that $\|x_n - x_{n-1}\| \leq 3^{-n}$ and $\|T_n x_n\| \geq \frac{2}{3} 3^{-n} \|T_n\|$. The sequence (x_n) is Cauchy, hence convergent to some $x \in X$; and it is easy to see that $\|x - x_n\| \leq \frac{1}{2} 3^{-n}$ and hence $\|T_n x\| \geq \frac{1}{6} 3^{-n} \|T_n\| \geq \frac{1}{6}(4/3)^n \to \infty$. $\square$

One important application of the uniform boundedness theorem concerns pointwise convergent sequences of linear mappings. So let X and Y again be normed linear spaces, and let (T_n) be a sequence of continuous linear maps from X to Y. We say that the sequence (T_n) is **pointwise convergent** to a mapping T if, for each $x \in X$, $T_n x$ is convergent to Tx in Y. In such a situation the limiting mapping T is obviously linear, but need it be continuous? In general the answer is no:

Example. Consider once again $X = c_{00}$, equipped with the sup norm, and $Y = \mathbb{R}$. This time let $\ell_n \in X^*$ be defined by

$$\ell_n(x) \;=\; \sum_{i=1}^{n} x_i \;. \tag{7.18}$$

Then the sequence (ℓ_n) is pointwise convergent to the linear mapping ℓ_∞ defined by

$$\ell_\infty(x) \;=\; \sum_{i=1}^{\infty} x_i \tag{7.19}$$

(why is ℓ_∞ well-defined? why is (ℓ_n) pointwise convergent to ℓ_∞?). But ℓ_∞ is unbounded, because the vector $x^{(n)} = (1, 1, \ldots, 1, 0, 0, \ldots)$ consisting of n ones followed by zeros has $\|x^{(n)}\|_\infty = 1$ but $\ell_\infty(x^{(n)}) = n$. $\square$

But once again, you will observe in this example that X is incomplete (and indeed meager in itself). If X is complete (or more generally if X is nonmeager in itself), then the limiting mapping T *does* have to be continuous:

[9]See A.D. Sokal, A really simple elementary proof of the uniform boundedness theorem, Amer. Math. Monthly **118**, 450–452 (2011), also available at `http://arxiv.org/abs/1005.1585`.

Corollary 7.11 *Let X be a Banach space and Y a normed linear space, and let (T_n) be a sequence of continuous linear maps from X to Y that is pointwise convergent to a mapping T. Then T is continuous.*

PROOF. Since the sequence (T_n) is pointwise convergent, it is pointwise bounded. Therefore, by the uniform boundedness theorem, the sequence (T_n) is bounded, i.e. there exists $M < \infty$ such that $\|T_n\|_{X\to Y} \le M$ for all n. This can be rephrased as saying that $\|T_n x\|_Y \le M\|x\|_X$ for all n and all $x \in X$ (why?). Taking $n \to \infty$, we have $T_n x \to Tx$ in Y for all $x \in X$ (why?), hence $\|Tx\| \le M\|x\|_X$ for all $x \in X$ (why?). This proves that T is bounded, i.e. continuous. □

Here is another important application of the uniform boundedness theorem. Let X, Y, Z be metric spaces. Then a map $f\colon X \times Y \to Z$ is said to be **separately continuous** if $f(x, y)$ is continuous as a function of x for each fixed value of y, and vice versa; or in more detail

$$\begin{aligned} x_n \to x,\ y \in Y &\implies f(x_n, y) \to f(x, y) \\ y_n \to y,\ x \in X &\implies f(x, y_n) \to f(x, y) \end{aligned}$$

The map f is said to be **jointly continuous** if it is continuous as a map from the product metric space $X \times Y$ (with any of its standard equivalent metrics) to Z; or in more detail

$$x_n \to x,\ y_n \to y \implies f(x_n, y_n) \to f(x, y)$$

Clearly joint continuity implies separate continuity (why?), but is the reverse true? Cauchy wrote in his 1821 *Cours d'Analyse* that a separately continuous function of two real variables is jointly continuous, but this is *false*! (Even great mathematicians can make mistakes.) You perhaps saw in your real analysis course the following counterexample[10]:

$$f(x, y) = \begin{cases} \dfrac{xy}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{if } (x, y) = (0, 0) \end{cases} \tag{7.20}$$

(Why is this function separately continuous? Why is it *not* jointly continuous?)

So a separately continuous function of two variables need not be jointly continuous. It is therefore somewhat surprising that a separately continuous *bilinear* mapping defined on a pair of *Banach* spaces is always jointly continuous:

Proposition 7.12 (joint continuity of separately continuous bilinear mappings) *Let X, Y, Z be normed linear spaces, with either X or Y (or both) complete, and let $T\colon X \times Y \to Z$ be a separately continuous bilinear mapping. Then T is jointly continuous.*

[10] Due to the German mathematician Hermann Schwarz (1843–1921) in 1872. A similar but slightly more complicated example was given two years earlier by the German mathematician Johannes Thomae (1840–1921), crediting his colleague Eduard Heine (1821–1881).

PROOF. Suppose that X is complete. For each $x \in X$, consider the linear mapping $T(x, \cdot)\colon Y \to Z$. This is a bounded linear mapping (why?), so there exists a constant $K_x < \infty$ such that

$$\|T(x,y)\|_Z \;\le\; K_x \|y\|_Y \quad \text{for all } y \in Y\,. \tag{7.21}$$

Next consider, for each $y \in Y$ with $\|y\|_Y \le 1$, the linear mapping $T(\,\cdot\,, y)\colon X \to Z$. This is a bounded linear mapping (why?), and it follows from (7.21) that the collection of these mappings, $\{T(\,\cdot\,, y)\colon \|y\|_Y \le 1\}$, is pointwise bounded on X (why?). Since X is complete, the uniform boundedness theorem tells us that this collection of mappings is bounded, i.e. there exists a constant $K < \infty$ such that

$$\|T(\,\cdot\,, y)\|_{X\to Z} \;\le\; K \quad \text{for all } y \in Y \text{ with } \|y\|_Y \le 1\,, \tag{7.22}$$

or equivalently

$$\|T(x,y)\|_Z \;\le\; K\|x\|_X \quad \text{for all } x \in X \text{ and } y \in Y \text{ with } \|y\|_Y \le 1\,, \tag{7.23}$$

or equivalently

$$\|T(x,y)\|_Z \;\le\; K\|x\|_X \|y\|_Y \quad \text{for all } x \in X \text{ and } y \in Y\,. \tag{7.24}$$

This shows that T is jointly continuous. $\square$

Example. Here is a simple example showing that we need *at least one* of the spaces X and Y to be complete, otherwise the conclusion can fail: Let $X = Y = c_{00}$ equipped with the sup norm, and $Z = \mathbb{R}$; then let $T(x,y) = \sum\limits_{n=1}^{\infty} x_n y_n$. Now T is well-defined (why?); and it is separately continuous, because $\|T(x, \cdot\,)\|_{Y\to Z} = \|x\|_1 < \infty$ (why?) and likewise for $T(\,\cdot\,, y)$. But if $x^{(n)} = (1, 1, \ldots, 1, 0, 0, \ldots)$ is the vector consisting of n ones followed by zeros, we have $\|x^{(n)}\|_\infty = 1$ but $T(x^{(n)}, x^{(n)}) = n$; so the sequence $(x^{(n)}/\sqrt{n})$ converges to zero but $T(x^{(n)}/\sqrt{n}, x^{(n)}/\sqrt{n}) = 1$ does not converge to zero; so T is not jointly continuous.

What would happen in this example if we took one of the two domain spaces (say, X) to be ℓ^∞ rather than c_{00}? Then we would satisfy the hypothesis of Proposition 7.12 that at least one of the domain spaces be complete, and T would remain well-defined (why?), but T would fail to be separately continuous! Indeed, $T(x, \cdot\,)$ would be an *unbounded* linear functional on $Y = c_{00}$ whenever $x \in \ell^\infty \setminus \ell^1$ (why?). $\square$

The uniform boundedness theorem and its corollaries are extremely useful in real and functional analysis. In Problem Set #7 I will give you some typical applications.

The open mapping theorem

Let X and Y be metric spaces, and let f be a map from X to Y. We have frequently used the elementary fact that f is continuous if and only if the inverse image of every open set is open, i.e. U open in Y implies $f^{-1}[U]$ open in X. Let us now say that the map f is an **open mapping** if the *direct* image of every open set is open, i.e. U open in X implies $f[U]$ open in Y.

A continuous mapping need not be open, and an open mapping need not be continuous. For instance, if X is the real line with the usual metric, Y is the real line with the discrete metric, and $f\colon X \to Y$ is the identity map, then f is open but not continuous (why?), and f^{-1} is continuous but not open (why?).[11] However, if $f\colon X \to Y$ is a *bijection*, then f^{-1} is continuous if and only if f is open (why?). This is the major reason that we are interested in open mappings: when we encounter a bijective map that is continuous, it is natural to ask whether its inverse is continuous as well. In particular we now want to pose questions of this type for *linear* operators from one Banach space to another.

You will recall (Example 6 of Handout #3) the following simple example of a bounded (i.e. continuous) linear map with an unbounded (i.e. discontinuous) inverse: Consider $T\colon \ell^2 \to \ell^2$ defined by

$$(Tx)_j \;=\; \frac{1}{j}\, x_j \; . \tag{7.25}$$

It is easy to see that T is bounded (of operator norm $\|T\|_{\ell^2 \to \ell^2} = 1$). Furthermore, T is an injection (why?). But T is not a surjection; rather, its image is the proper dense linear subspace

$$M \;=\; \{x \in \ell^2 \colon \sum_{k=1}^{\infty} k^2\, |x_k|^2 < \infty\} \tag{7.26}$$

(why?). Since T is a linear bijection from ℓ^2 to M, the inverse map T^{-1} is well-defined as a map from M to ℓ^2; it is obviously given by

$$(T^{-1}x)_j \;=\; j\, x_j \; , \tag{7.27}$$

which is *unbounded* (why?). So a bounded linear bijection $T\colon \ell^2 \to M$ can have an unbounded inverse $T^{-1}\colon M \to \ell^2$.

As remarked already in Handout #3, this pathology arises from the fact that the range space M is not complete. In fact, we have the following amazing results, due principally to Banach in the late 1920s:

Theorem 7.13 (open mapping theorem) *A bounded linear surjection of a* Banach *space X onto a* Banach *space Y is an open mapping.*

Corollary 7.14 (inverse mapping theorem) *A bounded linear bijection of a* Banach *space X onto a* Banach *space Y has a bounded inverse (and hence is a topological isomorphism).*

Do you see why the inverse mapping theorem is an immediate consequence of the open mapping theorem?

We will actually prove the following strong version of the open mapping theorem:

Theorem 7.15 (open mapping theorem, strong version) *Let T be a bounded linear map from a* Banach *space X into a normed linear space Y. If the image $T[X]$ is nonmeager in Y, then T is surjective (i.e. $T[X] = Y$) and an open mapping.*

[11] The fact that a continuous mapping need not be open was remarked already in Handout #1, using the following example: the map $f(x) = x^2$ is continuous from $\mathbb{R}$ to $\mathbb{R}$, but the image of the open set $(-1, 1)$ is the non-open set $[0, 1)$.

Do you see why the usual form of the open mapping theorem follows immediately from this version together with the (weak form of the) Baire category theorem?

Remark. With a bit more work one can prove not only that T is surjective and open, but also that Y is *complete* and topologically isomorphic to the quotient space $X/(\ker T)$ under the map $\widetilde{T}\colon X/(\ker T) \to Y$ given by $\widetilde{T}(x + \ker T) = T(x)$. $\square$

We will prove the strong version of the open mapping theorem by a series of lemmas. Let us denote by B (resp. B') the open unit ball in X (resp. Y).

The first lemma is almost trivial, but worth stating explicitly:

Lemma 7.16 *Let X and Y be normed linear spaces, and let $T\colon X \to Y$ be a linear mapping (not necessarily even continuous!). Then the following are equivalent:*

(a) T is an open mapping.

(b) $T[B]$ is an open neighborhood of 0 in Y.

(c) $T[B]$ is a neighborhood of 0 in Y.

Proof. If T is an open mapping, then clearly $T[B]$ is an open neighborhood of 0 in Y (why?). Conversely, suppose that $T[B]$ is a neighborhood of 0 in Y, and let $U \subseteq X$ be an open set. Given any $y \in T[U]$, choose $x \in U$ such that $y = Tx$; then there exists $\epsilon > 0$ such that $B(x, \epsilon) \subseteq U$ (why?). But it follows from this that $y + \epsilon T[B] \subseteq T[U]$, which proves that $T[U]$ is open (why?). $\square$

So, to prove that T is an open mapping, we will prove that $T[B]$ is a neighborhood of 0 in Y. We begin with a lemma asserting something slightly weaker, namely that the *closure* of $T[B]$ is a neighborhood of 0 in Y:

Lemma 7.17 *Let X and Y be normed linear spaces, and let $T\colon X \to Y$ be a linear mapping (not necessarily even continuous!). If $T[X]$ is nonmeager in Y, then $\overline{T[B]}$ is a neighborhood of 0 in Y.*

Proof. We have $T[X] = \bigcup_{n=1}^{\infty} nT[B]$. Since $T[X]$ is nonmeager, the sets $nT[B]$ cannot all be nowhere dense, hence $T[B]$ cannot be nowhere dense (why?), or in other words $\overline{T[B]}$ must have nonempty interior. Since the set $\overline{T[B]}$ is convex (why?) and symmetric (why?), the argument used in the proof of Lemma 7.10 tells us that $\overline{T[B]}$ is a neighborhood of 0 in Y. $\square$

And finally, we show how to infer from this weaker statement that $T[B]$ is itself a neighborhood of 0 in Y:

Lemma 7.18 *Let X be a* Banach *space and Y a normed linear space, and let $T\colon X \to Y$ be a* continuous *linear mapping. If $\overline{T[B]}$ is a neighborhood of 0 in Y, then $T[B]$ is also a neighborhood of 0 in Y.*

Proof. Suppose that $\overline{T[B]}$ is a neighborhood of 0 in Y, so that there exists $\delta > 0$ such that $\delta B' \subseteq \overline{T[B]}$. Then for $y \in \delta B'$ we have $y \in \overline{T[B]}$, so we can choose $y_1 \in T[B]$ such that $\|y - y_1\| < \delta/2$ (why?); and we can choose $x_1 \in B$ such that $Tx_1 = y_1$ (why?). Then $y - y_1 \in (\delta/2)B' \subseteq \frac{1}{2}\overline{T[B]}$, so we can choose $y_2 \in \frac{1}{2}T[B]$ such that $\|(y - y_1) - y_2\| < \delta/4$ (why?); and we can choose $x_2 \in \frac{1}{2}B$ such that $Tx_2 = y_2$. Continuing inductively, we obtain sequences (x_n) in X and (y_n) in Y such that $x_n \in 2^{-(n-1)}B$, $y_n = Tx_n$ and

$$\|y - (y_1 + y_2 + \ldots + y_n)\| \;<\; \delta/2^n \;. \tag{7.28}$$

Now write $s_n = \sum\limits_{k=1}^{n} x_k$. Since $\|x_n\| < 1/2^{n-1}$, we have

$$\|s_m - s_n\| \;\le\; \sum_{k=n+1}^{m} \|x_k\| \;<\; \frac{1}{2^{n-1}} \quad \text{whenever } m > n \;, \tag{7.29}$$

hence (s_n) is a Cauchy sequence in X. Since X is complete, (s_n) is convergent to some point $s \in X$. Moreover, we have $\|s\| \le \sum\limits_{k=1}^{\infty} \|x_k\| < 2$. Since $\|y - Ts_n\| < \delta/2^n$ (why?) and $Ts_n \to Ts$ (why? what hypothesis about T is being used here?), we have $y = Ts$. This shows that $y \in 2T[B]$. Since y was an arbitrary element of $\delta B'$, we have shown that $\delta B' \subseteq 2T[B]$, i.e. $(\delta/2)B' \subseteq T[B]$. Hence $T[B]$ is a neighborhood of 0 in Y. □

Proof of the strong version of the open mapping theorem. Let T be a bounded linear map from a Banach space X into a normed linear space Y. If $T[X]$ is nonmeager in Y, then Lemma 7.17 implies that $\overline{T[B]}$ is a neighborhood of 0 in Y. But then Lemma 7.18 implies that $T[B]$ is a neighborhood of 0 in Y. And this means, by Lemma 7.16, that T is an open mapping. And a *linear* map that is an open mapping is obviously surjective (why?). □

The closed graph theorem

Let X and Y be metric spaces. Then a function $f\colon X \to Y$ is said to have a **closed graph** if

$$\mathrm{graph}(f) \;=\; \{(x, f(x))\colon\, x \in X\} \tag{7.30}$$

is a closed subset of the product space $X \times Y$ (equipped with any one of its standard metrics). Note in particular that if f is bijective, then f has a closed graph if and only if f^{-1} does (why?).

How does having a closed graph compare to the related property of *continuity*? Consider the following three statements concerning a sequence (x_n) in X and elements $x \in X$, $y \in Y$:

(a) $x_n \to x$

(b) $f(x_n) \to y$

(c) $y = f(x)$

Continuity is the statement that (a) implies (b) and (c) [for the appropriately chosen y]. Having a closed graph is the statement that (a) and (b) together imply (c).

So a continuous function always has a closed graph, but the converse is false: for instance, the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) \;=\; \begin{cases} 1/x & \text{if } x \neq 0 \\ a & \text{if } x = 0 \end{cases} \tag{7.31}$$

(for any chosen $a \in \mathbb{R}$) has a closed graph (why?) but is not continuous (why?).

It is therefore somewhat surprising that for *linear* operators from one *Banach* space into another, having a closed graph automatically implies continuity:

Theorem 7.19 (closed graph theorem) *Let X and Y be* Banach *spaces, and let $T\colon X \to Y$ be a linear map. Then T is continuous if and only if it has a closed graph.*

PROOF. We have already proved that every continuous map has a closed graph, so let us prove the converse. This will be a fairly simple application of the open mapping theorem.

Since X and Y are Banach spaces, the product space $X \times Y$ is a Banach space when equipped with the norm $\|(x,y)\|_{X\times Y} = \|x\|_X + \|y\|_Y$. Since graph$(T)$ is a closed linear subspace of $X \times Y$, it is a Banach space in its own right. Now consider the projection operator $\Pi_1\colon \text{graph}(T) \to X$ defined by $\Pi_1(x,y) = x$. Clearly Π_1 is a bijection of graph(T) onto X (why?). Moreover, since $\|\Pi_1(x,y)\|_X = \|x\|_X \le \|(x,y)\|_{X\times Y}$, the map Π_1 is continuous (of norm ≤ 1). The open mapping theorem then implies that Π_1^{-1} is a continuous linear map of X into (in fact onto) graph(T). On the other hand, the projection operator $\Pi_2\colon \text{graph}(T) \to Y$ defined by $\Pi_2(x,y) = y$ is also continuous (of norm ≤ 1) since $\|\Pi_2(x,y)\|_X = \|y\|_Y \le \|(x,y)\|_{X\times Y}$. It follows that $T = \Pi_2\Pi_1^{-1}$ is a continuous map of X into Y. $\square$

The completeness of both X and Y is crucial in the closed graph theorem; otherwise the conclusion can fail:

Example 1: An unbounded multiplication operator. Suppose that we try to define an unbounded linear operator $T\colon \ell^1 \to \ell^1$ by

$$(Tx_k) \;=\; kx_k \;. \tag{7.32}$$

The trouble is that this formula does not map ℓ^1 into itself; rather, we must restrict attention to the proper dense linear subspace $M \subsetneq \ell^1$ defined by

$$M \;=\; \{x \in \ell^1\colon \sum_{k=1}^{\infty} k|x_k| < \infty\} \;. \tag{7.33}$$

(Why is M dense in ℓ^1?) Then $T\colon M \to \ell^1$ is a well-defined linear map. Furthermore, T has a closed graph: for if $(x^{(n)})$ is a sequence in M converging in ℓ^1 norm to $x \in M$, *and we also know that $(Tx^{(n)})$ converges in ℓ^1 norm to some $y \in \ell^1$*, then it follows that $Tx = y$ (why?). [Alternatively, it suffices to observe that T is a bijection from M onto ℓ^1 (why?), and that the inverse map T^{-1} is bounded (why?) and hence has a closed graph; so T must have a closed graph as well.] But T is not continuous. This does not contradict the closed graph theorem, because the domain space M is not complete.

This example could equally well be carried out on any of the spaces ℓ^p with $1 \le p < \infty$, or in c_0. (What changes if we try it in ℓ^∞?)

Example 2: A differential operator. In the space $\mathcal{C}[0,1]$, let us try to define the operator T of differentiation. The trouble (as discussed already in Example 5 of Handout #3) is that differentiation makes no sense for arbitrary functions $f \in \mathcal{C}[0,1]$ (at least, not if we want the derivative f' to also belong to $\mathcal{C}[0,1]$). Rather, we need to restrict ourselves to (for example) the space $\mathcal{C}^1[0,1]$ consisting of functions that are once continuously differentiable on $[0,1]$ (where we insist on having also one-sided derivatives at the two endpoints, and the derivative f' is supposed to be continuous on all of $[0,1]$, including at the endpoints). This is a proper dense linear subspace of $\mathcal{C}[0,1]$ (why is it dense? why is it not closed?). Then the map $T\colon f \mapsto f'$ is a well-defined linear map from $\mathcal{C}^1[0,1]$ into $\mathcal{C}[0,1]$. Furthermore, T has a closed graph: for if (f_n) is a sequence in $\mathcal{C}^1[0,1]$ converging in sup norm to $f \in \mathcal{C}^1[0,1]$, *and we also know that* (f'_n) *converges in sup norm to some* $g \in \mathcal{C}[0,1]$, then it follows that $f' = g$ (why?). On the other hand, T is not continuous (why?). Once again, this does not contradict the closed graph theorem, because the domain space $\mathcal{C}^1[0,1]$ with the sup norm is not complete.

Example 3. A more complicated example shows that the completeness of Y is also essential in the closed graph theorem: see e.g. John B. Conway, *A Course in Functional Analysis*, p. 92.

Some final remarks. In this course we have concentrated our attention on *bounded* (i.e. continuous) linear operators; but unbounded linear operators do play a central role in many applications of functional analysis. For instance, differential operators are invariably unbounded (as Example 2 makes clear) whenever they are considered as acting on the usual normed function spaces such as $\mathcal{C}(X)$ or $L^p(X)$; so applications of functional analysis to the study of partial differential equations (PDE) cannot avoid dealing with unbounded operators. Likewise, most of the operators arising in quantum mechanics (position, momentum, Hamiltonians, ...) are unbounded operators on a Hilbert space.

What the closed graph theorem shows is that unbounded linear operators cannot be defined in any sensible way (e.g. any way that gives them a closed graph) on the *whole* of a Banach space. Rather, as Examples 1 and 2 make clear, unbounded linear operators are most naturally defined on a domain D that is a *proper dense subspace* of a Banach space.

To address this situation, it makes sense to generalize slightly the concept of a closed graph, as follows: Let X and Y be metric spaces, *and let* D *be a subset of* X (the "domain"). Then a function $f\colon D \to Y$ is said to have a **closed graph** (relative to X and Y) if

$$\operatorname{graph}(f) \;=\; \{(x, f(x))\colon x \in D\} \tag{7.34}$$

is a closed subset of the product space $X \times Y$.

Please note that we are now considering the domain D as a subset of a possibly bigger space X, and that whether f has a closed graph or not *can depend on the choice of* X. For instance, the identity map $f(x) = x$ with $D = (0,1)$ has a closed graph if we take $X = (0,1)$ and $Y \supseteq (0,1)$, or $Y = (0,1)$ and $X \supseteq (0,1)$, but not if we take e.g. $X = Y = [0,1]$.

In this new context, we can once again rephrase the closed-graph condition in terms of sequences: a map f has a closed graph (in the new sense) if and only if the situation

$$x_n \in D, \quad x_n \to x \quad \text{and} \quad f(x_n) \to y \tag{7.35}$$

implies that

$$x \in D \quad \text{and} \quad f(x) = y \; . \tag{7.36}$$

The trouble is that continuity no longer automatically implies the closed-graph property (as the preceding example shows). But it is fairly easy to prove the following facts (I leave the proofs as an exercise for you):

Proposition 7.20 *Let X and Y be metric spaces, and let D be a* closed *subset of X. Then a continuous function $f\colon D \to Y$ has a closed graph.*

Proposition 7.21 *Let X and Y be normed linear spaces, with Y complete, and let D be a linear subspace of X. Then a continuous linear map $T\colon D \to Y$ has a closed graph if and only if D is closed.*

The important point is now that the typical unbounded operators, such as those in Examples 1 and 2, are closed also in the new sense (which is *stronger* than the old sense):

Example 1, revisited. We previously considered T as an operator from M to ℓ^1, and we observed that it has a closed graph in the sense that if $(x^{(n)})$ is a sequence in M converging in ℓ^1 norm to $x \in M$, and $(Tx^{(n)})$ converges in ℓ^1 norm to some $y \in \ell^1$, then it follows that $Tx = y$. But let us now consider T as an operator with domain $D = M$ *lying inside the space* $X = \ell^1$. Then, to verify that T has a closed graph, we need to prove the stronger statement that if $(x^{(n)})$ is a sequence in M converging in ℓ^1 norm to $x \in \ell^1$ (not necessarily in M!), and $(Tx^{(n)})$ converges in ℓ^1 norm to some $y \in \ell^1$, then it follows that $x \in M$ and $Tx = y$. Do you see why this is true?

Example 2, revisited. We previously considered T as an operator from $\mathcal{C}^1[0,1]$ to $\mathcal{C}[0,1]$, and we observed that it has a closed graph in the sense that if (f_n) is a sequence in $\mathcal{C}^1[0,1]$ converging in sup norm to $f \in \mathcal{C}^1[0,1]$, and (f_n') converges in sup norm to some $g \in \mathcal{C}[0,1]$, then it follows that $f' = g$. But let us now consider T as an operator with domain $D = \mathcal{C}^1[0,1]$ *lying inside the space* $X = \mathcal{C}[0,1]$. Then, to verify that T has a closed graph, we need to prove the stronger statement that if (f_n) is a sequence in $\mathcal{C}^1[0,1]$ converging in sup norm to $f \in \mathcal{C}[0,1]$ (not necessarily in $\mathcal{C}^1[0,1]$!), and (f_n') converges in sup norm to some $g \in \mathcal{C}[0,1]$, then it follows that $f \in \mathcal{C}^1[0,1]$ and $f' = g$. Do you see why this is true?

Many (but not all) of the results concerning bounded linear operators can be carried over, with some modifications and some extra work, to linear operators having a closed graph (usually called *closed linear operators* for short). But that topic belongs to a more advanced course in functional analysis.