Math 351 Notes

Instructor Arkady Etkin

Continuity

LIMITS & CONTINUITY

In this section we present a brief refresher course on limits and continuity for real-valued functions. To begin, let f be a real-valued function defined (at least) for all points in some open interval containing the point $a \in \mathbb{R}$ except, possibly, at a itself. We will refer to such a set as a punctured neighborhood of a.

Given a number $L \in \mathbb{R}$, we write $\lim_{x \to a} f(x) = L$ to mean:

$$\begin{cases} \text{for every } \varepsilon > 0, \text{ there is some } \delta > 0 \text{ such that } |f(x) - L| < \varepsilon \\ \text{whenever } x \text{ satisfies } 0 < |x - a| < \delta. \end{cases}$$

We say that $\lim_{x \to a} f(x)$ exists if there is some number $L \in \mathbb{R}$ that satisfies the requirements spelled out above.

• Theorem:

Let f be a real-valued function defined in some punctured neighborhood of $a \in \mathbb{R}$. Then, the following are equivalent:

i) There exists a number L such that $\lim_{x \to a} f(x) = L$ (by the ε-δ definition).

ii) There exists a number L such that $f(x_n) \to L$ whenever $x_n \to a$, where $x_n \neq a$ for all n.

iii) $\{f(x_n)\}_{n=1}^{\infty}$ converges (to something) whenever $x_n \to a$, where $x_n \neq a$ for all n.

(Try to prove this yourself!)

Note: The point to item iii) is that if $\lim_{n \to \infty} f(x_n)$ always exists, then it must actually be independent of the choice of $\{x_n\}$. This is not as mystical as it might sound; indeed, if

$x_n \to a$ and $y_n \to a$, then the sequence $x_1, y_1, x_2, y_2, \ldots$ also converges to a. (How does this help?) This particular phrasing is interesting because it does not refer to L. That is, we can test for the existence of a limit without knowing its value.

Now suppose that f is defined in a neighborhood of a, this time including the point a itself. We say that f is continuous at a if $\lim_{x \to a} f(x) = f(a)$. That is, if:

$$\begin{cases} \text{for every } \varepsilon > 0, \text{ there is a } \delta > 0 \text{ (that depends on } f, a, \text{ and } \varepsilon) \\ \text{such that } |f(x) - f(a)| < \varepsilon \text{ whenever } x \text{ satisfies } |x - a| < \delta. \end{cases}$$

Notice that we replaced L by $f(a)$ and we dropped the requirement that $x \neq a$. The previous theorem has an obvious extension to this case:

• Theorem:

Let f be a real-valued function defined in some neighborhood of $a \in \mathbb{R}$. Then, the following are equivalent:

i) f is continuous at a (by the ε-δ definition).

ii) $f(x_n) \to f(a)$ whenever $x_n \to a$;

iii) $\{f(x_n)\}$ converges (to something) whenever $x_n \to a$.

Notice that we dropped the requirement that $x_n \neq a$. Thus, if $\lim_{n \to \infty} f(x_n)$ always exists, then it must equal $f(a)$ (why?).

You might also recall that we have a notation for left- and right-hand limits and left and right continuity. For example, if we define

$$f(a-) = \lim_{x \to a^-} f(x) \qquad \text{and} \qquad f(a+) = \lim_{x \to a^+} f(x)$$

(provided that these limits exist, of course), then we could add another equivalence to the above theorem:

iv) $f(a-)$ and $f(a+)$ both exist, and both are equal to $f(a)$.

Note: One-sided limits are peculiar to functions defined on $\mathbb{R}$, and they do not generalize very well (because they are tied to the order in $\mathbb{R}$). But they are very good at what they do: They permit the cataloguing of very refined types of discontinuities. For example, we say that f is right-continuous at a if $f(a+)$ exists and equals $f(a)$, and we say that f has a jump discontinuity at a if $f(a-)$ and $f(a+)$ both exist but at least one is different from $f(a)$. A function having only jump discontinuities is not that terrible. In particular, monotone functions are rather well behaved:

• Proposition:

Let $f:(a,b)\longrightarrow\mathbb{R}$ be monotone and let $a<c<b$. Then, $f(c-)$ and $f(c+)$ both exist. Thus, f can have only jump discontinuities.

Proof:

We might as well suppose that f is increasing (otherwise, consider $-f$). In that case, $f(c)$ is an upper bound for $\{f(t):a<t<c\}$ and a lower bound for $\{f(t):c<t<b\}$. All that remains is to check that

$$\sup\{f(t):a<t<c\}=\lim_{x\to c^-} f(x) \quad \text{and} \quad \inf\{f(t):c<t<b\}=\lim_{x\to c^+} f(x)\,.$$

We will sketch the proof of the first of these.

Given $\varepsilon>0$, there is some x_0 with $a<x_0<c$ such that $\sup\limits_{t<c} f(t)-\varepsilon<f(x_0)\leq\sup\limits_{t<c} f(t)$. Now let $\delta=c-x_0>0$. Then, if $c-\delta<x<c$, we get $x_0<x<c$, and so $f(x_0)\leq f(x)\leq\sup\limits_{t<c} f(t)$.

Thus, $\left|f(x)-\sup\limits_{t<c} f(t)\right|<\varepsilon$. ■

• Theorem:

If $f:(a,b)\longrightarrow\mathbb{R}$ is monotone, then f has at most countably many points of discontinuity in (a,b), all of which are jump discontinuities.

Proof:

That f has only jump discontinuities follows from the proposition we have just proved above. Now we just need to count the points of discontinuity.

Let's reflect on the situation. If $f:(a,b)\longrightarrow\mathbb{R}$ is, say, increasing, and if $c\in(a,b)$, then the left-and right-hand limits of f at c satisfy $f(c-)\leq f(c)\leq f(c+)$. In particular, f is discontinuous at c iff $f(c-)<f(c+)$. Consequently, if c and d are two different points of discontinuity for f, then the intervals $(f(c-),\ f(c+))$ and $(f(d-),\ f(d+))$ are nonempty and disjoint.

Thus,

$$\left\{(f(c-),\ f(c+)) : c \text{ is } a \text{ point of discontinuity for } f\right\}$$

is a collection of nonempty, disjoint open intervals in $\mathbb{R}$, and any such collection must be countable. ■

• Corollary:

If $f:[a,b]\longrightarrow[c,d]$ is both monotone and onto, then f is continuous.

We can put this corollary to good use. Recall that the Cantor function $f:\Delta\longrightarrow[0,1]$ is monotone and onto. Indeed, if $x\in\Delta$, then $x=0.2\,a_1\,2\,a_2\ldots$ (base 3), where each $a_i=0$ or 1 and $f(x)=\sum_{i=1}^{\infty}\frac{a_i}{2^i}$. Since $\{a_n\}_{n=1}^{\infty}$ can be any sequence of 0's and 1's f is clearly onto.

We can extend the definition of the Cantor function f to all of [0, 1] in an obvious way: We take f to be an appropriate constant on each of the open intervals that make up $[0,1]\backslash\Delta$. For example, we would set $f(x)=f\left(\frac{1}{3}\right)=\frac{1}{2}$ for each $x\in\left(\frac{1}{3},\frac{2}{3}\right)$ and $f(x)=f\left(\frac{1}{9}\right)=\frac{1}{4}$ for each $x\in\left(\frac{1}{9},\frac{2}{9}\right)$.

Formally, we define

$$f(x)=\sup\{f(y):y\in\Delta,\ y\le x\}\text{ for }x\in[0,1]\backslash\Delta\ .$$

The new function $f:[0,1]\longrightarrow[0,1]$ is still increasing (why?) and is actually continuous! (because it is onto). It is called a singular function because $f'=0$ at almost every point in [0, 1]. That is $f'=0$ on $[0,1]\backslash\Delta$, a set of measure 1.

The theorem above has a converse. Given any countable set $\mathcal{D}$ in $\mathbb{R}$, we can construct an increasing function $f:\mathbb{R}\longrightarrow\mathbb{R}$ that is discontinuous precisely at the points of $\mathcal{D}$.

Here is a brief sketch:

Let $\mathcal{D}=\{x_1,x_2,\ \ldots\}$ and let $\{\varepsilon_n\}_{n=1}^{\infty}$ be a sequence of positive numbers with $\sum_{n=1}^{\infty}\varepsilon_n<\infty$. We define $f(x)=\sum_{x_n\le x}\varepsilon_n$, where the sum is over the set $\{n:x_n\le x\}$ and where $f(x)=0$ if the set is empty. Notice that $0\le f(x)\le\sum_{n=1}^{\infty}\varepsilon_n<\infty$ in any case.

Now, if $x<y$, then

$$f(y)=\sum_{x_n\le y}\varepsilon_n=\sum_{x_n\le x}\varepsilon_n+\sum_{x<x_n\le y}\varepsilon_n=f(x)+\sum_{x<x_n\le y}\varepsilon_n\ge f(x)\ .$$

Thus, f is increasing.

Next we consider this formula in each of the cases $x=x_k$ and $y=x_k$.

Case 1:

$$x=x_k<y\Longrightarrow f(y)=f(x_k)+\sum_{x_k<x_n\le y}\varepsilon_n\ .$$

Claim: $f(x_k+)=f(x_k)$.

$$\lim_{y \to x_k^+} \sum_{x_k < x_n \leq y} \varepsilon_n = 0 \qquad \text{because} \qquad \sum_{n=N}^{\infty} \varepsilon_n \to 0 \quad \text{as} \quad N \to \infty$$

Case 2:

$$x < x_k = y \Longrightarrow f(x_k) = f(x) + \sum_{x < x_n \leq x_k} \varepsilon_n \geq f(x) + \varepsilon_k \ .$$

<u>Claim:</u> $f(x_k -) = f(x_k) - \varepsilon_k$, i.e.

$$\lim_{x \to x_k^-} \sum_{x < x_n \leq x_k} \varepsilon_n = \varepsilon_k$$

Putting this all together,

$$f(x_k -) + \varepsilon_k = f(x_k) = f(x_k +) \qquad \text{and} \qquad f(x_k +) - f(x_k -) = \varepsilon_k \ .$$

The proof that f is continuous at each $x \in \mathbb{R} \backslash \mathcal{D}$ is similar.

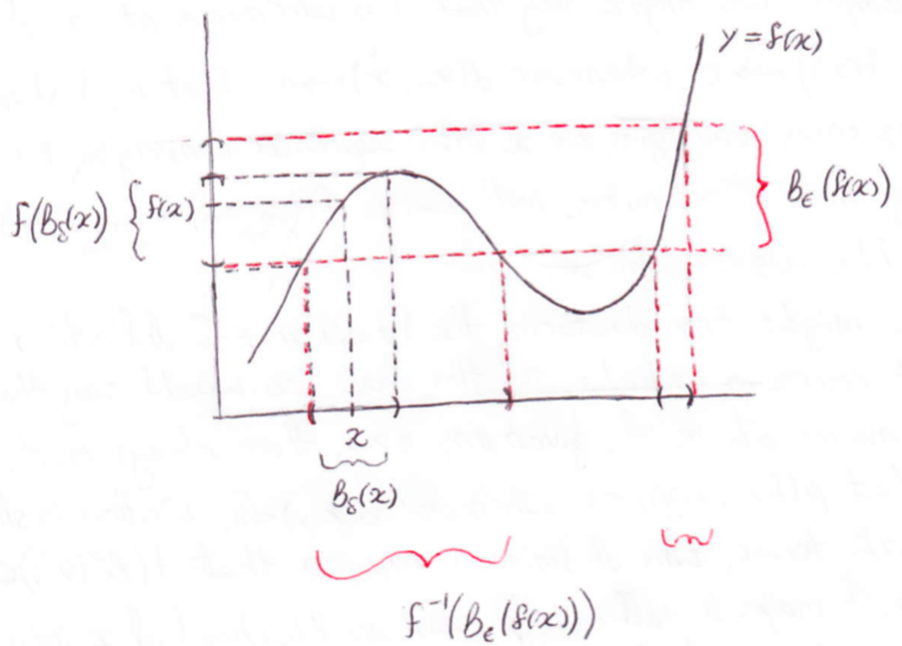
CONTINUITY ON ABSTRACT METRIC SPACES

Given a function $f : (M, d) \longrightarrow (N, \rho)$ (where M, N are arbitrary vector spaces), and given a point $x \in M$, we have at least two plausible definitions for the continuity of f at x. Each definition is derived from its obvious counterpart for real-valued functions by replacing absolute values with an appropriate metric.
For example, we might say that f is continuous at x if $\rho(f(x_n), f(x)) \to 0$ whenever $d(x_n, x) \to 0$. That is, f should send sequences converging to x into sequences converging to $f(x)$. This says that f "commutes" with limits: $f\left(\lim_{n\to\infty} (x_n)\right) = \lim_{n\to\infty} (f(x_n))$.

Another alternative is to use the familiar $\varepsilon - \delta$ definition from elementary calculus. In this case we would say that f is continuous at x if, given any $\varepsilon > 0$, there always exists a $\delta > 0$ such that $\rho(f(x), f(y)) < \varepsilon$ whenever $d(x, y) < \delta$. Written in slightly different terms, this definition requires that $f\left(B_\delta^d(x)\right) \subset B_\varepsilon^\rho(f(x))$. That is, f maps a sufficiently small neighborhood of x into a given neighborhood of $f(x)$.

We will rewrite this last definition once more, but this time we will use an inverse image. Recall that the inverse image of a set $A \subseteq Y$, under a function $f : X \longrightarrow Y$, is defined to be the set $\{x \in X : f(x) \in A\}$ and is usually written $f^{-1}(A)$ (the inverse image of any set under any function always makes sense. Although the notation is similar, inverse image have nothing whatsoever to do with inverse functions, which don't always make sense).

Stated in terms of an inverse image, our condition reads: $B_\delta^d(x) \subset f^{-1}\left(B_\varepsilon^\rho(f(x))\right)$. Looks a bit imposing? Well, it actually tells us quite a bit. It says that the inverse image of an open set containing $f(x)$ must still be open near x. Curious. The figure below will help you visualize these new definitions:



If f is continuous at every point of M, we simply say that f is continuous on M, or often just that f is continuous.

By now it should be clear that any statement concerning arbitrary open balls will translate into a statement concerning arbitrary open sets. Thus, there is undoubtedly a characterization of continuity available that may be stated exclusively in terms of open sets. Of course, any statement concerning open sets probably has a counterpart using closed sets. And don't forget sequences! Open sets and closed sets can each be characterized in terms of convergent sequences, and so we would expect to find a characterization of continuity in terms of convergent sequences, too. At any rate, we've done enough hinting around about reformulations of the definition of continuity. It's time to put our cards on the table.

• Theorem:

Given $f:(M,d)\longrightarrow(N,\rho)$, the following are equivalent:

i) f is continuous on M (by the ε-δ definition).

ii) For any $x \in M$, if $x_n \to x$ in M, then $f(x_n) \to f(x)$ in N.

iii) If E is closed in N, then $f^{-1}(E)$ is closed in M.

iv) If V is open in N, then $f^{-1}(V)$ is open in M.

Proof:

i) $\Longleftrightarrow$ ii) : (Compare this with the case $f:\mathbb{R}\longrightarrow\mathbb{R}$.) Suppose that $x_n \overset{d}{\to} x$. Given $\varepsilon > 0$, let $\delta > 0$ be such that $f\left(B_\delta^d(x)\right) \subset B_\varepsilon^\rho(f(x))$. Then, since $x_n \overset{d}{\to} x$, we have that $\{x_n\}$ is eventually in $B_\delta^d(x)$. But this implies that $\{f(x_n)\}$ is eventually in $B_\varepsilon^\rho(f(x))$. Since ε is arbitrary, this means that $f(x_n) \overset{\rho}{\to} f(x)$.

ii) $\Longleftrightarrow$ iii) : Let E be closed in (N, ρ). Given $\{x_n\} \subset f^{-1}(E)$ such that $x_n \overset{d}{\to} x \in M$, we need to show that $x \in f^{-1}(E)$. But $\{x_n\} \subset f^{-1}(E)$ implies that $\{f(x_n)\} \subset (E)$, while $x_n \overset{d}{\to} x \in M$ tells us that $f(x_n) \overset{\rho}{\to} f(x)$ from ii). Thus, since E is closed, we have that $f(x) \in E$ or $x \in f^{-1}(E)$.

iii) $\Longleftrightarrow$ iv) is obvious, since $f^{-1}(A^c) = \left(f^{-1}(A)\right)^c$.

iv) $\Longleftrightarrow$ i) : Given $x \in M$ and $\varepsilon > 0$, the set $B_\varepsilon^\rho(f(x))$ is open in (N, ρ) and so, by iv), the set $f^{-1}\left(B_\varepsilon^\rho(f(x))\right)$ is open in (M, d). But then $B_\delta^d(x) \subset f^{-1}\left(B_\varepsilon^\rho(f(x))\right)$, for some $\delta > 0$, because $x \in f^{-1}\left(B_\varepsilon^\rho(f(x))\right)$. ■

Example:

a) Define $X_\mathbb{Q}:\mathbb{R}\longrightarrow\mathbb{R}$ by $X_\mathbb{Q}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}$.

Then $X_\mathbb{Q}^{-1}(B_{1/3}(1)) = \mathbb{Q}$ and $X_\mathbb{Q}^{-1}(B_{1/3}(0)) = \mathbb{R}\backslash\mathbb{Q}$. Thus $X_\mathbb{Q}$ cannot be continuous at any point of $\mathbb{R}$ because neither $\mathbb{Q}$ nor $\mathbb{R}\backslash\mathbb{Q}$ contains an interval.

b) A function $f: M \longrightarrow N$ between metric spaces is called an isometry (into) if f preserves distances, that is, if $\rho(f(x), f(y)) = d(x, y)$ for all $x, y \in M$.
Obviously, an isometry is continuous. The natural inclusions from $\mathbb{R}$ into $\mathbb{R}^2$ (i.e. $x \longmapsto (x, 0)$) and from $\mathbb{R}^2$ into $\mathbb{R}^3$ (this time $(x, y) \longmapsto (x, y, 0)$) are isometries.

c) Let $f:\mathbb{N}\longrightarrow\mathbb{R}$ be any function. Then f is continuous! Why? Because $\{n\}$ is an open ball in $\mathbb{N}$. Specifically, $\{n\} = B_{1/2}(n) \subset f^{-1}(B_\varepsilon(f(n)))$ for any $\varepsilon > 0$.

d) $f:\mathbb{R}\longrightarrow\mathbb{N}$ is continuous iff f is constant! Why? [Hint: Recall that $\mathbb{R}$ has no nontrivial clopen sets.]

e) Relative continuity can sometimes be counterintuitive. From a) we know that $X_{\mathbb{Q}}$ has no points of continuity relative to $\mathbb{R}$, but the restriction of $X_{\mathbb{Q}}$ to $\mathbb{Q}$ is everywhere continuous relative to $\mathbb{Q}$! Why?

f) If y is any fixed element of (M, d), then the real-valued function $f(x) = d(x, y)$ is continuous on M.

Note: The theorem stated prior to the above examples characterizes continuous functions in terms of open sets and closed sets. As it happens, we can use these characterizations "in reverse" to derive information about open and closed sets. In particular, we can characterize closures in terms of certain continuous functions.

Definition: Given a nonempty set A and a point $x \in M$, we define the distance from x to A by:

$$d(x, A) = \inf\{d(x, a) : a \in A\}.$$

Clearly, $0 \leq d(x, A) < \infty$ for any x and any A, but it is not necessarily true that $d(x, A) > 0$ when $x \notin A$. For example, $d(x, \mathbb{Q}) = 0$ for any $x \in \mathbb{R}$.

• Proposition:

$d(x, A) = 0$ iff $x \in \overline{A}$.

Proof:

$d(x, A) = 0$ iff there is a sequence of points $\{a_n\}_{n=1}^{\infty}$ in A such that $d(x, a_n) \to 0$. But this means that $a_n \to x$ and, hence, $x \in \overline{A}$. ■

Note that this proposition has given us another connection between limits in M and limits in $\mathbb{R}$. Loosely speaking, this proposition shows that 0 is a limit point of $\{d(x, a) : a \in A\}$ iff x is a limit point of A. We can get even more mileage out of this observation by checking that the map $x \longmapsto d(x, A)$ is actually continuous. For this it suffices to establish the following inequality:

• Proposition:

$|d(x, A) - d(y, A)| \leq d(x, y)$.

Proof:

It is true by triangle inequality that $d(x, a) \leq d(x, y) + d(y, a)$ for any $a \in A$. But $d(x, A)$ is

a lower bound for $d(x, a)$; hence $d(x, A) \leq d(x, y) + d(y, a)$. Now, by taking the infimum over $a \in A$, we get $d(x, A) \leq d(x, y) + d(y, A)$.
Since the roles of x and y are interchangeable, we're done. ■

To appreciate what this has done for us, let's make two simple observations. First, if $f : M \longrightarrow \mathbb{R}$ is a continuous function, then the set $E = \{x \in M : f(x) = 0\}$ is closed (why?). Conversely, if E is a closed set in M, then E is the "zero set" of some continuous real-valued function on M; in particular, $E = \{x \in M : d(x, E) = 0\}$. Thus a set E is closed iff $E = f^{-1}(\{0\})$ for some continuous function $f : M \longrightarrow \mathbb{R}$.

Conclusion: If you know all of the closed (or open) sets in a metric space M, then you know all of the continuous real-valued functions on M. Conversely, if you know all of the continuous real-valued functions on M, then you know all of the closed (or open) sets in M.