# LECTURE: BANACH SPACES AND DERIVATIVES

## 1. Normed Vector Spaces and Banach Spaces

**Definition:**

A **norm** on a vector space $V$ is a function $\|\cdot\| : V \to \mathbb{R}$ with the following properties:

(1) $\|x\| \geq 0$

(2) $\|x\| = 0 \iff x = 0$

(3) $\|cx\| = |c|\|x\|$ for all scalars $c$

(4) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)

A vector space paired with a norm is a **normed vector space**.

Intuitively, a norm measures the length of a vector.

Every normed vector space is a metric space, since a norm induces a metric, which is given by

$$d(x, y) = \|x - y\|.$$

The converse, however, is not true. There are vector spaces on which there is a metric, but no norm can be found.

Next, we define a bounded linear map between normed vector spaces.

**Definition:**

Let $X$ and $Y$ be normed vector spaces and let $L : X \to Y$ be a linear map.

Then $L$ is **bounded** if there exists a constant $C \geq 0$ such that for all $u \in X$,

$$\|Lu\|_Y \leq C\|u\|_X.$$

Here $\|u\|_X$ is the norm in $X$ and $\|u\|_Y$ is the norm in $Y$

The miracle of functional analysis is that for a linear transformation, continuity is equivalent to boundedness:

**Theorem:**

A linear transformation $L : X \to Y$ is bounded if and only if $L$ is continuous.

**Proof:** ($\Rightarrow$) If $L$ is bounded, then

$$\|Lu - Lv\|_Y = \|L(u - v)\|_Y \leq L\|u - v\|_X,$$

and so $L$ is Lipschitz, thus continuous.

($\Leftarrow$) Assume $L$ is continuous. Taking $\epsilon = 1$, since $L$ is continuous at 0, we can find $\delta > 0$ such that, for all $u \in X$ with $\|u\|_X \leq \delta$, $\|Lu\|_Y \leq 1$

Let $x \in X$ with $x \neq 0$. Then

$$\left\| \frac{\delta}{\|x\|} x \right\|_X = \delta \left\| \frac{x}{\|x\|} \right\|_X = \delta,$$

from which it follows that

$$\left\| L\left(\frac{\delta}{\|x\|}x\right)\right\|_Y \leq 1.$$

Therefore

$$\|Lx\|_Y \leq \frac{1}{\delta}\|x\|_X,$$

thus $L$ is a bounded linear operator with $C = 1/\delta$ □

Let $\mathcal{L}(X, Y)$ be the space of bounded linear maps from $X$ to $Y$. If $X = Y$, we usually denote this $\mathcal{L}(X)$. We define the operator norm of $L \in \mathcal{L}$ as follows:

**Definition:**

Let $L : X \to Y$ be a bounded linear operator. Then the **operator norm** of $L$ is defined as one of the following, all of which are equivalent:

$$\|L\| = \sup_{\|u\|_X \leq 1} \|Lu\|_Y$$
$$\|L\| = \sup_{\|u\|_X = 1} \|Lu\|_Y$$
$$\|L\| = \inf\{C \geq 0 : \|Lu\|_Y \leq C\|u\|_X \text{ for all } u \in X\}$$

Think of $\|L\|$ as the maximum possible spread of $L$. For example, if $\|L\| = 2$ then for all $x$, $\|Lx\|_Y \leq 2\,\|x\|_X$, so $\|Lx\|_Y$ is never more than twice as big as $\|x\|_X$.

**Definition:**

A **Banach space** $X$ is a complete normed vector space.

Here completeness means $(X, d)$ is complete, where $d(x, y) = \|x - y\|$, so Cauchy sequences in $\|\cdot\|$ converge.

$\mathcal{L}(X, Y)$ becomes a normed vector space with the norm $\|L\|$ defined above. Moreover:

**Definition:**

If $Y$ is a Banach space, then $\mathcal{L}(X, Y)$ is also a Banach space.

The proof is left as an exercise. The proof that $\mathcal{L}(X, Y)$ is complete is similar to the proof of the completeness of $C([a, b])$.

The next lemma is incredibly useful and gives a criterion for a specific linear operator on a Banach space to be invertible with bounded inverse:

**Neumann Series Theorem:**

Let $X$ be a Banach space, and let $S \in \mathcal{L}(X)$ with $\|S\| < 1$. Then $I - S$ is invertible, and $(I - S)^{-1} \in L(X)$, where $I$ is the identity operator on $X$.

**Proof:** Define the *Neumann series* for $S$ as

$$L = \sum_{n=0}^{\infty} S^n = I + S + S^2 + \ldots$$

which is the operator analogue of the ordinary geometric series.

To show that this is well-defined, we note that the sequence of partial sums of $L$ is a Cauchy sequence, thus the sum converges since $X$ is complete.

In addition, $L$ is bounded, with

$$\|L\| \leq \sum_{n=0}^{\infty} \|S\|^n \leq \frac{1}{1 - \|S\|}.$$

Finally, since

$$(I - S)L = (I - S) \underbrace{\sum_{n=0}^{N} S^n}_{L} = \sum_{n=0}^{N} (S^n - S^{n+1}) = \underbrace{I - S^{N+1}}_{\to I},$$

$(I - S)L = I$. Similarly, $L(I - S) = I$. □

## 2. Differentiation in Banach Spaces

**Goal:** If $f : \mathbb{R}^n \to \mathbb{R}^m$, how to define the derivative $f'(x)$?

**Mnemonic:** **In**put to **M**outhput

**First guess:** By analogy with the scalar case, if $x \in \mathbb{R}^n$

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

But here $h$ is a vector, so it makes no sense to divide by $h$

**Analogy:** $(n = 1)$ Note that if $h$ is small, then

$$f(x+h) = f(x) + f'(x)h + \text{ Smaller terms}$$

**Definition:**

Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ and $x \in \mathbb{R}^n$.

*If* there is a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$f(x+h) = f(x) + Lh + o(h)$$

$$\text{Where } \lim_{h \to 0} \frac{|o(h)|}{|h|} = 0$$

Then we say $f$ is **differentiable at** $x$ and $f'(x) = L$

And $f$ is **differentiable** if $f$ is differentiable at all $x$

In other words, if you can expand $f(x+h)$ out with a small remainder, then the linear part is the derivative of $f$.

Before, $f'(x)$ was just a number, but now it's something more dynamic, it's a linear transformation. Intuitively, if $f$ distorts space, then $f'(x)$ describes the linear part of the distortion.

This is useful for theoretical purposes like showing the chain rule in higher dimensions, but in practice, we have the following shortcut:

If all partial derivatives of $f$ exist and are continuous in a neighborhood of $a$, then $f$ is differentiable, and the derivative is given by the Jacobian matrix

$$DF(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_m} & \cdots & \frac{\partial f_1}{\partial x_n} \end{pmatrix}_{x=a}.$$

Finally, we extend this definition to arbitrary Banach spaces.

**Definition:**

Let $X$ and $Y$ be Banach spaces, $U \subset X$ open, and $f\colon U \to Y$.

Then $f$ is **differentiable** at $u \in U$ if there exists a bounded linear transformation $L \in \mathcal{L}(X, Y)$ such that

$$\lim_{h \to 0} \frac{\|f(u+h) - f(u) - Lh\|_Y}{\|h\|_X} = 0.$$

The map $L$ is sometimes called the Fréchet derivative.

**Note:** If f is differentiable at $u_0 \in U$, we use the notation $Df(u_0)$ or $f_u(u_0)$ for the derivative.

**Remarks:**

(1) That the Fréchet derivative, if it exists, is unique.

(2) To show that the Fréchet derivative exists, we usually find a guess for the derivative, and then use the definition about to show that that guess works.

(3) If $f$ is differentiable for all $u \in U$, then the map $Df\colon U \to \mathcal{L}(X, Y)$ defined by $u \mapsto Df(u)$ is well-defined.

(4) A function $f$ is $C^1$ if this map is continuous.

(5) The chain rule remains valid in Banach spaces (provided the appropriate derivatives exist)

(6) Higher order derivatives can be defined by considering the differentiability of $Df : U \to \mathcal{L}(X, Y)$, etc.

## 3. Fixed Point Theorems

This section covers fixed point theorems, which guarantee the existence of a unique fixed point of a function, i.e. a unique $x$ such that $f(x) = x$. Fixed point methods are a powerful tool in analysis, especially numerical analysis, as we'll soon discover

**Definition:**

If $X$ is any nonempty set and $f : X \to X$, then $p$ is a fixed point of $f$ if

$$f(p) = p$$

Our next goal is to state the Banach fixed point theorem, which gives a fairly simple sufficient condition for a map to have a fixed point.

**Definition:**

Let $X$ be metric any metric space then $f : X \to X$ is a **contraction** if there is $k < 1$ such that

$$d(f(x), f(y)) \leq kd(x, y)$$

For any $x$ and $y$ in $X$

In other words, $f$ is Lipschitz continuous with Lipschitz constant $L < 1$.

Intuitively, contractions shrink distances between points

**Banach Fixed Point Theorem:**

If $X$ is complete and $f$ is a contraction, then $f$ has a unique fixed point $p$.

**Analogy:** You may have noticed this phenomenon when you start with a number on a calculator, and repeatedly apply cos or $\sqrt{x}$ on it. Eventually the number stays the same!

**Proof:**[1]

**STEP 1:** Let $x_0 \in X$ and define $x_n = f^n(x_0)$ ($f$ applied $n$ times)

Notice $d(x_1, x_2) = d(f(x_0), f(x_1)) \leq k d(x_0, x_1)$ and

And more generally you can show that

$$d(x_n, x_{n+1}) \leq k^n d(x_0, x_1)$$

**STEP 2:** **Claim:** $(x_n)$ is Cauchy

**Why?** Let $\epsilon > 0$ be given and $N$ be TBA, then if $m, n > N$ (WLOG assume $n \geq m$), then

[1] The proof is from Pugh's book, Theorem 23 in Chapter 4

$$
\begin{aligned}
d(x_m, x_n) \leq & d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n) \\
\leq & k^m d(x_0, x_1) + k^{m+1} d(x_0, x_1) + \cdots + k^{n-1} d(x_1, x_0) \qquad \text{(By } \textbf{STEP 1}\text{)} \\
\leq & \left(k^m + k^{m+1} + \cdots + k^{n-1}\right) d(x_1, x_0) \\
= & k^m \left(1 + k + \cdots + k^{n-m-1}\right) d(x_0, x_1) \\
\leq & k^m \left(1 + k + k^2 + \cdots\right) d(x_0, x_1) \\
= & k^m \left(\frac{1}{1-k}\right) d(x_0, x_1) \\
\leq & \frac{k^N}{1-k} d(x_0, x_1) \quad \text{Since } m > N \text{ and } k < 1
\end{aligned}
$$

But since $k < 1$ we have $\lim_{n\to\infty} k^n = 0$, so we can choose $N$ large enough so that $\frac{k^N}{1-k} d(x_0, x_1) < \epsilon$, which in turn implies $d(x_m, x_n) < \epsilon$ ✓

**STEP 3:** Since $(x_n)$ is Cauchy and $X$ is complete, $(x_n)$ converges to some $p$

**Claim:** $p$ is a fixed point of $f$.

This follows because

$$
\begin{aligned}
x_{n+1} = & f(x_n) \\
\lim_{n\to\infty} x_{n+1} = & \lim_{n\to\infty} f(x_n) \\
p = & f\left(\lim_{n\to\infty} x_n\right) \qquad \text{(continuity)} \\
p = & f(p) \checkmark
\end{aligned}
$$

**STEP 4:** **Uniqueness:** Suppose there are two fixed points $p \neq q$, then

$$
d(p, q) = d(f(p), f(q)) \leq k d(p, q) < d(p, q) \Rightarrow\Leftarrow
$$

Applications of this include proving the ODE existence uniqueness theorem and proving the Inverse Function Theorem (see next time)

Here we give a nice application to numerical analysis, more precisely let's prove that Newton's method converges.

**Newton's Method:** The goal is to find zeros of $f$, that is $x$ such that $f(x) = 0$

For this, start with any $x_0$ such that $f'(x_0) \neq 0$ and then iterate the algorithm

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

Geometrically, the new value of $x$ is where the tangent line of $f$ at $(x_n, f(x_n))$ hits the $x$-axis.

And the hope is that this algorithm converges to a zero of $f$

**Convergence of Newton's Method:**

Let $f : [a, b] \to \mathbb{R}$ be $C^2$. Suppose for some $x \in [a, b]$ that $f(x) = 0$ and $f'(x) \neq 0$.

Then there exists an interval $I = [x - \delta, x + \delta] \subset [a, b]$ such that Newton's method converges to $x$ starting at any $x_0 \in I$.

**Proof:**

**STEP 1:** Define the "Newton function"

$$g(x) = x - \frac{f(x)}{f'(x)}.$$

This corresponds to the right-hand-side of Newton's method

We will show that there is an interval containing $x$ on which $g$ is a contraction.

**STEP 2:** Since $f''$ is continuous $[a, b]$, there is $M > 0$ such that $|f''(x)| \leq M$ for all $x \in [a, b]$.

Given $\delta > 0$ TBA, let $I = [x - \delta, x + \delta] \subset [a, b]$.

Then for any $y_1, y_2 \in I$, since $g$ is continuously differentiable, by the mean-value theorem, we have

$$|g(y_1) - g(y_2)| \leq \sup_{y \in I} |g'(y)| \, |y_1 - y_2|.$$

Our goal is to choose $\delta$ sufficiently small to control $g'(y)$.

Differentiating $g(y)$, we obtain

$$g'(y) = 1 - \frac{f'(y)^2 - f(y)f''(y)}{f'(y)^2} = \frac{f(y)f''(y)}{f'(y)^2}$$

**STEP 3:** Since $f(x) = 0$, and $f'(x) \neq 0$, choose $\delta$ sufficiently small so that $I \subset [a, b]$, and, for all $y \in I$ the following two things hold:

(1) $|f'(y)| \geq \frac{1}{2}|f'(x)|$

(2) $|f(y)| \leq \frac{|f'(x)|^2}{8M}$

For all $y \in I$, using the expression for $g'(y)$ from the previous step,

$$|g(y_1) - g(y_2)| \leq \frac{|f(y)||f''(y)|}{|f'(y)|^2}|y_1 - y_2| \leq \frac{|f'(x)|^2}{8M} M \frac{4}{|f'(x)|^2} \leq \frac{1}{2}|y_1 - y_2|.$$

Hence $g$ is a contraction on $I$

Therefore, by the Banach Fixed Point Theorem $g$ has a unique fixed point $x^\star$ in $I$, that is Newton's method converges to some $x^\star$

But by assumption, since $f(x) = 0$ we then have $g(x) = x - \frac{0}{f'(x)} = x$ and so by uniqueness of fixed points, $x^\star = x$

That is, Newton's method indeed converges to a zero of $f$ $\square$

## 4. Inverse Function Theorem

As another consequence, we can prove the celebrated Inverse Function Theorem in Analysis.

**Goal:** If $y = f(x)$, when can we solve for $x$ in terms of $y$? That is, when can we write $x = g(y)$ where $g$ is a **smooth** function?

**Example 1:** If $f(x) = x^3$ then $g(y) = y^{\frac{1}{3}}$. Notice $g$ is differentiable **except** at 0, and 0 is *precisely* the point where $f'(x) = 0$

**Example 2:** If $f(x) = x^2$ then we can't find a *global* inverse (valid for all $x$) since $f$ isn't one-to-one, but our hope is to do this locally, around a point. Once again there is no inverse when $f'(x) = 0$.

In short, we would like to say "As long as $f'(x) \neq 0$, we can solve for $x$ in terms of $y$, at least locally"

Moreover, if $n = 1$ if $f(g(x)) = x$ then differentiating this, we get

$$f'(g(x))g'(x) = 1 \Rightarrow g'(x) = \frac{1}{f'(g(x))}$$

This was used in Calculus to get the derivatives of $\ln(x)$ or $\sin^{-1}(x)$ for example.