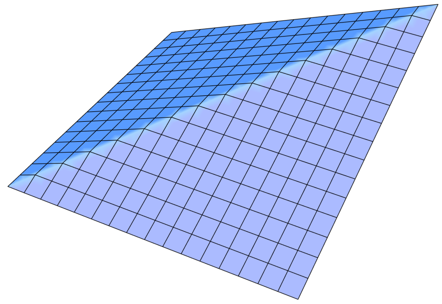

MATH 247: Calculus III (Advanced Level)

Prof. Alexandru Nica • Fall 2013 • University of Waterloo
MFW 11:30AM–12:20PM in MC 4041

Transcribed, LaTeXed, and edited by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is `https://rob-zimmerman.github.io/coursenotes`.

Contents

1. Euclidean space, sequences, and continuity in $\mathbb{R}^n$

Lecture 1 — Monday, September 09, 2013

Euclidean Space

This course is about the structure of Euclidean n-space, denoted $\mathbb{R}^n$. The main topics are sequences, limits, and continuity in $\mathbb{R}^n$; sets and topology in $\mathbb{R}^n$; derivatives of functions of several variables; multivariable integrals; and the connections between derivatives and integrals.

Notation 1.1 We write $\vec{x} = (x^{(1)}, x^{(2)}, \ldots, x^{(n)})$ for a vector $\vec{x} \in \mathbb{R}^n$, where each $x^{(i)} \in \mathbb{R}$. The number $x^{(i)}$ is called the ***i*th component** of the vector $\vec{x}$.

Let $\vec{x} = (x^{(1)}, \ldots, x^{(n)})$ and $\vec{y} = (y^{(1)}, \ldots, y^{(n)})$ be vectors in $\mathbb{R}^n$, and let $\alpha \in \mathbb{R}$. We have the following basic operations on these vectors:

1. **Addition**: define
$$\vec{x} + \vec{y} = (x^{(1)} + y^{(1)}, x^{(2)} + y^{(2)}, \ldots, x^{(n)} + y^{(n)}).$$

2. **Scalar multiplication**:
$$\alpha\vec{x} = (\alpha x^{(1)}, \alpha x^{(2)}, \ldots, \alpha x^{(n)}).$$

3. **Standard inner product**:
$$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{n} x^{(i)} y^{(i)} = x^{(1)}y^{(1)} + x^{(2)}y^{(2)} + \cdots + x^{(n)}y^{(n)}.$$

4. **Norm** (or **length**):
$$\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle} = \sqrt{\sum_{i=1}^{n} (x^{(i)})^2}.$$

   This square root makes sense because it is the square root of a sum of squares. In particular, $\|\vec{x}\| \geq 0$ for every $\vec{x} \in \mathbb{R}^n$, with equality if and only if $\vec{x} = \vec{0}$.

Remark 1.2 Here are some basic properties of the inner product:

1. **Bilinearity**: for $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{R}$,
$$\langle \alpha_1\vec{x}_1 + \alpha_2\vec{x}_2, \vec{y} \rangle = \alpha_1\langle \vec{x}_1, \vec{y} \rangle + \alpha_2\langle \vec{x}_2, \vec{y} \rangle,$$

$$\langle \vec{x}, \beta_1 \vec{y}_1 + \beta_2 \vec{y}_2 \rangle = \beta_1 \langle \vec{x}, \vec{y}_1 \rangle + \beta_2 \langle \vec{x}, \vec{y}_2 \rangle.$$

2. **Symmetry**:

$$\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle \quad \text{for all } \vec{x}, \vec{y} \in \mathbb{R}^n.$$

3. **Positivity**:

$$\langle \vec{x}, \vec{x} \rangle \geq 0 \quad \text{with equality if and only if } \vec{x} = \vec{0}.$$

Proposition 1.3 (Cauchy–Schwarz inequality) For all $\vec{x}, \vec{y} \in \mathbb{R}^n$, we have

$$|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\| \cdot \|\vec{y}\|.$$

Proof. If $\vec{y} = \vec{0}$, then the inequality is immediate, since both sides are equal to 0. So assume $\vec{y} \neq \vec{0}$. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(t) = \langle \vec{x} - t\vec{y}, \vec{x} - t\vec{y} \rangle.$$

By bilinearity and symmetry,

$$\begin{aligned} f(t) &= \langle \vec{x}, \vec{x} \rangle - \langle \vec{x}, t\vec{y} \rangle - \langle t\vec{y}, \vec{x} \rangle + \langle t\vec{y}, t\vec{y} \rangle \\ &= \|\vec{x}\|^2 - 2t\langle \vec{x}, \vec{y} \rangle + t^2 \|\vec{y}\|^2. \end{aligned}$$

So f is a quadratic polynomial in t:

$$f(t) = at^2 + bt + c, \qquad a = \|\vec{y}\|^2, \qquad b = -2\langle \vec{x}, \vec{y} \rangle, \qquad \text{and} \qquad c = \|\vec{x}\|^2.$$

Now $f(t) = \|\vec{x} - t\vec{y}\|^2 \geq 0$ for every $t \in \mathbb{R}$, so the quadratic polynomial f never takes negative values. Hence its discriminant must satisfy

$$\Delta = b^2 - 4ac \leq 0.$$

That is,

$$(-2\langle \vec{x}, \vec{y} \rangle)^2 - 4\|\vec{y}\|^2 \|\vec{x}\|^2 \leq 0,$$

which gives

$$\langle \vec{x}, \vec{y} \rangle^2 \leq \|\vec{x}\|^2 \|\vec{y}\|^2.$$

Taking square roots yields

$$|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\| \cdot \|\vec{y}\|.$$

$\square$

Exercise 1.4 Determine all cases when the Cauchy–Schwarz inequality holds with equality.

Lecture 2 — Wednesday, September 11, 2013

Some Inequalities in $\mathbb{R}^n$

Definition 2.1 The **(Euclidean) distance** on $\mathbb{R}^n$ is the function $d : \mathbb{R}^n \times \mathbb{R}^n \to [0, \infty)$ defined by

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

Proposition 2.2 (Triangle inequality)

1. $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ for all $\vec{x}, \vec{y} \in \mathbb{R}^n$.
2. $d(\vec{x}, \vec{z}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$ for all $\vec{x}, \vec{y}, \vec{z} \in \mathbb{R}^n$.

Proof. For (1), we compute

$$\begin{aligned}
\|\vec{x} + \vec{y}\|^2 &= (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) \\
&= \|\vec{x}\|^2 + 2(\vec{x} \cdot \vec{y}) + \|\vec{y}\|^2 \\
&\leq \|\vec{x}\|^2 + 2\|\vec{x}\|\|\vec{y}\| + \|\vec{y}\|^2 \quad \text{(by the Cauchy–Schwarz inequality)} \\
&= (\|\vec{x}\| + \|\vec{y}\|)^2.
\end{aligned}$$

Since both sides are nonnegative, taking square roots gives

$$\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|.$$

For (2), we have

$$\begin{aligned}
d(\vec{x}, \vec{z}) &= \|\vec{x} - \vec{z}\| \quad \text{(by the definition of distance)} \\
&= \|(\vec{x} - \vec{y}) + (\vec{y} - \vec{z})\| \leq \|\vec{x} - \vec{y}\| + \|\vec{y} - \vec{z}\| \quad \text{(by (1))}
\end{aligned}$$

$$= d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$$

$\square$

Remark 2.3 Both inequalities (1) and (2) extend to the "polygon" inequalities.

Example 2.4 For any $\vec{x}, \vec{y}, \vec{z}, \vec{w} \in \mathbb{R}^n$, we have $d(\vec{x}, \vec{w}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z}) + d(\vec{z}, \vec{w})$.

Figure 1 depicts the corresponding polygonal path from $\vec{x}$ to $\vec{w}$ through $\vec{y}$ and $\vec{z}$.

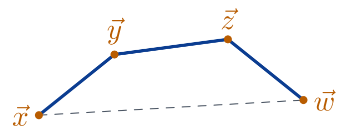

FIGURE 1

Proof. By applying (2) of the triangle inequality twice, we get

$$d(\vec{x}, \vec{w}) \leq d(\vec{x}, \vec{z}) + d(\vec{z}, \vec{w}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z}) + d(\vec{z}, \vec{w}).$$

The general polygon inequality follows by induction. $\square$

Definition 2.5 For $\vec{a} \in \mathbb{R}^n$ and $r > 0$, we denote

$$B(\vec{a}, r) := \{\vec{x} \in \mathbb{R}^n \mid \|\vec{x} - \vec{a}\| < r\},$$

and call it the **open ball** of center $\vec{a}$ and radius r. Equivalently, we may write $d(\vec{x}, \vec{a}) < r$. We also denote

$$\overline{B}(\vec{a}, r) := \{\vec{x} \in \mathbb{R}^n \mid \|\vec{x} - \vec{a}\| \leq r\},$$

and call it the **closed ball** of center $\vec{a}$ and radius r.

Remark 2.6 The difference between the open and closed balls is that the closed ball may contain points $\vec{x}$ satisfying $\|\vec{x} - \vec{a}\| = r$. This set,

$$\{\vec{x} \in \mathbb{R}^n \mid \|\vec{x} - \vec{a}\| = r\},$$

is the **sphere** of center $\vec{a}$ and radius r.

Example 2.7 Let $\vec{a}, \vec{b} \in \mathbb{R}^n$ and $r, s > 0$, and suppose that $d(\vec{a}, \vec{b}) > r + s$. Prove that $B(\vec{a}, r) \cap B(\vec{b}, s) = \emptyset$.

Solution. Assume for a contradiction that $B(\vec{a}, r) \cap B(\vec{b}, s) \neq \emptyset$. Then there exists $\vec{x} \in B(\vec{a}, r) \cap B(\vec{b}, s)$. By part (2) of the triangle inequality,

$$r + s < d(\vec{a}, \vec{b}) \leq d(\vec{a}, \vec{x}) + d(\vec{x}, \vec{b}) < r + s,$$

since $d(\vec{a}, \vec{x}) < r$ and $d(\vec{x}, \vec{b}) < s$. This is impossible. Therefore $B(\vec{a}, r) \cap B(\vec{b}, s) = \emptyset$. □

Example 2.8 Let $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Prove the following:

1. $|x_i| \leq \|\vec{x}\|$ for every $1 \leq i \leq n$;
2. $$\|\vec{x}\| \leq \sum_{i=1}^{n} |x_i|.$$

Solution. For (1), set

$$M := \max\{|x_1|, \ldots, |x_n|\}.$$

Then $|x_i| \leq M$ for every $1 \leq i \leq n$. Also, for each j we have $|x_j| \leq M$, so $x_j^2 \leq M^2$. Since at least one index k satisfies $|x_k| = M$, we get

$$M^2 = x_k^2 \leq x_1^2 + \cdots + x_n^2 = \|\vec{x}\|^2.$$

Taking square roots yields $M \leq \|\vec{x}\|$. Therefore, for every i,

$$|x_i| \leq M \leq \|\vec{x}\|.$$

For (2), observe that

$$\left(\sum_{i=1}^{n} |x_i|\right)^2 = \sum_{i=1}^{n} |x_i|^2 + 2 \sum_{1 \leq i < j \leq n} |x_i||x_j| \geq \sum_{i=1}^{n} |x_i|^2 = \|\vec{x}\|^2.$$

Taking square roots yields

$$\|\vec{x}\| \leq \sum_{i=1}^{n} |x_i|.$$

□

Definition 2.9 For $\vec{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, define the **1-norm** by

$$\|\vec{x}\|_1 \coloneqq \sum_{i=1}^{n} |x_i|,$$

and the **∞-norm** by

$$\|\vec{x}\|_\infty \coloneqq \max(|x_1|, |x_2|, \dots, |x_n|).$$

The statements in Example 2.8 thus combine to say that

$$\|\vec{x}\|_\infty \leq \|\vec{x}\| \leq \|\vec{x}\|_1 \quad \text{for all } \vec{x} \in \mathbb{R}^n.$$

Sequences in $\mathbb{R}^n$

We use the notation $(\vec{x}_k)_{k=1}^\infty$ for a sequence $\vec{x}_1, \vec{x}_2, \dots$ in $\mathbb{R}^n$. We reserve n for the dimension rather than using it as the sequence index. Repetitions are allowed (for example, $\vec{x}_i = \vec{x}_j$ for some $i \neq j$), so constant sequences are allowed as well.

Definition 2.10 Let $(\vec{x}_k)$ be a sequence in $\mathbb{R}^n$ and let $\vec{a} \in \mathbb{R}^n$. Consider the sequence of real numbers $(\|\vec{x}_k - \vec{a}\|)_{k=1}^\infty$. We say that $(\vec{x}_k)$ **converges to $\vec{a}$** in $\mathbb{R}^n$, or equivalently $\vec{x}_k \to \vec{a}$, when $\|\vec{x}_k - \vec{a}\| \to 0$ in $\mathbb{R}$.

Let us spell out the condition $\|\vec{x}_k - \vec{a}\| \to 0$:

$$\begin{aligned} \text{(Lim)} : \quad & \text{for every } \varepsilon > 0, \text{ there exists } k_0 \in \mathbb{N} \\ & \text{such that } \|\vec{x}_k - \vec{a}\| < \varepsilon \text{ for all } k \geq k_0. \end{aligned}$$

We refer to this as (Lim). In (Lim), instead of "$\|\vec{x}_k - \vec{a}\| < \varepsilon$", we could write "$d(\vec{x}_k, \vec{a}) < \varepsilon$" or "$\vec{x}_k \in B(\vec{a}, \varepsilon)$".

So (Lim) can also be phrased in the following way:

$$(\mathrm{Lim})' : \quad \text{for every } \varepsilon > 0, \text{ there exists } k_0 \in \mathbb{N} \text{ such that } \vec{x}_k \in B(\vec{a}, \varepsilon) \text{ for all } k \geq k_0.$$

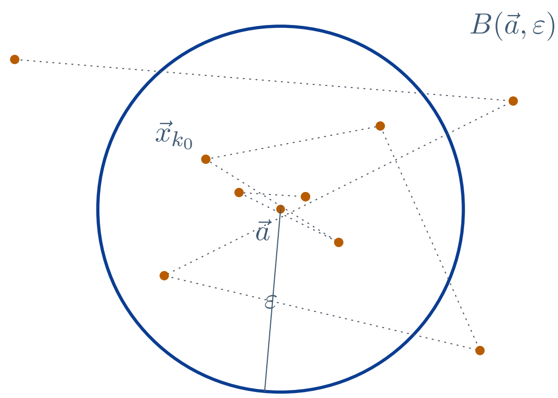

FIGURE 2

In words, once we draw the ball of radius ε around $\vec{a}$, there is a time after which all terms of the sequence lie in that ball, never to leave again. See Figure 2.

Note that in (Lim) or (Lim)', the order of the quantifiers is very important! Compare (Lim) with the following condition:

$$(\text{E-C}) : \quad \text{there exists } k_0 \in \mathbb{N} \text{ such that for every } \varepsilon > 0, \; \|\vec{x}_k - \vec{a}\| < \varepsilon \text{ for all } k \geq k_0.$$

This condition is strictly stronger than (Lim).

Of course, "E-C" means "eventually constant".

Lecture 3 — Friday, September 13, 2013

Definition 3.1 Let $(\vec{x}_k)$ be a sequence in $\mathbb{R}^n$. Write

$$\vec{x}_1 = (x_1^{(1)}, \dots, x_1^{(n)}), \qquad \vec{x}_2 = (x_2^{(1)}, \dots, x_2^{(n)}), \qquad \dots, \qquad \text{and} \qquad \vec{x}_k = (x_k^{(1)}, \dots, x_k^{(n)}).$$

Looking "vertically," we obtain n sequences of real numbers:

$$(x_k^{(1)})_{k=1}^{\infty}, \ \ldots, \ (x_k^{(n)})_{k=1}^{\infty}.$$

These are called the **component sequences** of $(\vec{x}_k)_{k=1}^{\infty}$. Conversely, given n sequences in $\mathbb{R}$, we can assemble them into one sequence in $\mathbb{R}^n$.

Proposition 3.2 Let $(\vec{x}_k)$ be a sequence in $\mathbb{R}^n$, and let $\vec{a} \in \mathbb{R}^n$. Then $(\vec{x}_k)$ converges to $\vec{a}$ in $\mathbb{R}^n$ if and only if each component sequence converges to the corresponding component of $\vec{a}$ in $\mathbb{R}$.

Proof. *Forward implication.* Assume that $\vec{x}_k \to \vec{a}$ in $\mathbb{R}^n$. Fix i, where $1 \le i \le n$. For every $k \ge 1$,

$$|x_k^{(i)} - a^{(i)}| = |(\vec{x}_k - \vec{a})^{(i)}| \le \|\vec{x}_k - \vec{a}\|$$

by Example 2.8. Since the right-hand side tends to 0, the squeeze theorem gives $x_k^{(i)} \to a^{(i)}$.

Reverse implication. Assume that $x_k^{(i)} \to a^{(i)}$ for every $1 \le i \le n$. Then each $|x_k^{(i)} - a^{(i)}| \to 0$, so

$$|x_k^{(1)} - a^{(1)}| + \cdots + |x_k^{(n)} - a^{(n)}| \to 0.$$

By the coordinate inequalities (Example 2.8 again)

$$|v^{(i)}| \le \|\vec{v}\| \le \sum_{j=1}^{n} |v^{(j)}|$$

applied to $\vec{v} = \vec{x}_k - \vec{a}$,

$$0 \le \|\vec{x}_k - \vec{a}\| \le |x_k^{(1)} - a^{(1)}| + \cdots + |x_k^{(n)} - a^{(n)}|.$$

The squeeze theorem therefore gives $\|\vec{x}_k - \vec{a}\| \to 0$, and hence $\vec{x}_k \to \vec{a}$. $\square$

Next we upgrade the Bolzano–Weierstrass theorem. Recall that in $\mathbb{R}$, this theorem states that every bounded sequence has a convergent subsequence.

Definition 3.3 We say that a sequence $(\vec{x}_k)$ in $\mathbb{R}^n$ is **bounded** if there exists $r > 0$ such that $\|\vec{x}_k\| \le r$ for all $k \in \mathbb{N}$. Since $\|\vec{x}_k\| = d(\vec{x}_k, \vec{0})$, this is equivalent to saying that $\vec{x}_k \in \overline{B}(\vec{0}, r)$

for all $k \in \mathbb{N}$.

Proposition 3.4 Let $(\vec{x}_k)$ be a sequence in $\mathbb{R}^n$. Then $(\vec{x}_k)$ is bounded in $\mathbb{R}^n$ if and only if each component sequence is bounded in $\mathbb{R}$.

Proof. *Forward implication.* Assume first that $(\vec{x}_k)$ is bounded in $\mathbb{R}^n$. Then there exists $r > 0$ such that

$$\|\vec{x}_k\| \le r \quad \text{for all } k \in \mathbb{N}.$$

Fix $i \in \{1, \ldots, n\}$. By Example 2.8,

$$|x_k^{(i)}| \le \|\vec{x}_k\| \le r \quad \text{for all } k,$$

so $(x_k^{(i)})$ is bounded in $\mathbb{R}$. Since i was arbitrary, every component sequence is bounded.

Reverse implication. Conversely, assume each component sequence $(x_k^{(i)})$ is bounded in $\mathbb{R}$. Then for each $i \in \{1, \ldots, n\}$ there exists $M_i > 0$ such that

$$|x_k^{(i)}| \le M_i \quad \text{for all } k \in \mathbb{N}.$$

Let $r := M_1 + \cdots + M_n$. For every k, applying Example 2.8 to $\vec{x}_k$ gives

$$\|\vec{x}_k\| \le \sum_{j=1}^{n} |x_k^{(j)}| \le \sum_{j=1}^{n} M_j = r.$$

Hence $(\vec{x}_k)$ is bounded in $\mathbb{R}^n$. $\square$

Definition 3.5 Let $(\vec{x}_k)$ be a sequence in $\mathbb{R}^n$. A **subsequence** of $(\vec{x}_k)$ is a sequence of the form $(\vec{x}_{k(p)})_{p=1}^{\infty}$ for some

$$1 \le k(1) < k(2) < \cdots < k(p) < \ldots$$

in $\mathbb{N}$. Thus the subsequence is

$$\vec{x}_{k(1)}, \vec{x}_{k(2)}, \ldots, \vec{x}_{k(p)}, \ldots.$$

We use the following convention: instead of $(\vec{x}_{k(p)})_{p=1}^{\infty}$, we write $(\vec{x}_k)_{k\in P}$, where

$$P = \{k(1), \dots, k(p), \dots\},$$

an infinite subset of $\mathbb{N}$. This notation is convenient because we do not need separate "sub-subsequence" notation: we can simply write $(\vec{x}_k)_{k\in Q}$ with $Q \subseteq P$.

Example 3.6 Let $P = \{2, 4, 6, \dots, 2m, \dots\}$ and $Q = \{2, 4, 8, 16, \dots, 2^m, \dots\}$. Then for any sequence $(\vec{x}_k)_{k=1}^{\infty}$ in $\mathbb{R}^n$, we can form the subsequence $(\vec{x}_k)_{k\in P}$ which is $\vec{x}_2, \vec{x}_4, \dots, \vec{x}_{2m}, \dots$, and the subsequence $(\vec{x}_k)_{k\in Q}$, which is $\vec{x}_2, \vec{x}_4, \vec{x}_8, \dots, \vec{x}_{2^m}, \dots$.

Lecture 4 — Monday, September 16, 2013

An immediate fact is that if $\vec{x}_k \to \vec{a}$ in $\mathbb{R}^n$, then every subsequence also converges to $\vec{a}$. Indeed, if $\|\vec{x}_k - \vec{a}\| < \varepsilon$ for every $k \geq k_0$, then the same inequality holds along a subsequence once its indices exceed k_0. We may write this as $\vec{x}_k \xrightarrow[k\in P]{k\to\infty} \vec{a}$.

Today we prove Bolzano–Weierstrass in $\mathbb{R}^n$ for every $n \in \mathbb{N}$ by induction on n. We need a lemma first.

Lemma 4.1 Let $(\vec{x}_k)$ be a sequence in $\mathbb{R}^{n+1}$, and for every $k \in \mathbb{N}$ we write $\vec{x}_k = (\vec{y}_k, t_k)$ where $\vec{y}_k = (x_k^{(1)}, \dots, x_k^{(n)}) \in \mathbb{R}^n$ and t_k is $x_k^{(n+1)} \in \mathbb{R}$. Suppose that $\vec{y}_k \to \vec{b} \in \mathbb{R}^n$ and $t_k \to t \in \mathbb{R}$. Then $\vec{x}_k \to \vec{a} \in \mathbb{R}^{n+1}$ where $\vec{a} = (\vec{b}, t)$.

Proof. Since $\vec{y}_k \to \vec{b}$, Proposition 3.2 implies that $x_k^{(i)} \to b^{(i)}$ for every $1 \leq i \leq n$. Also, $t_k \to t$, so $x_k^{(n+1)} \to t = a^{(n+1)}$. Hence each component of $\vec{x}_k$ converges to the corresponding component of $\vec{a} = (\vec{b}, t)$. Applying Proposition 3.2 again, we conclude that $\vec{x}_k \to \vec{a}$ in $\mathbb{R}^{n+1}$. □

Theorem 4.2 (Bolzano–Weierstrass in $\mathbb{R}^n$) Let $(\vec{x}_k)$ be a bounded sequence in $\mathbb{R}^n$. Then we can find indices

$$1 \leq k(1) < k(2) < \cdots < k(p) < \dots$$

such that the subsequence $(\vec{x}_{k(p)})$ is convergent to some $\vec{a} \in \mathbb{R}^n$.

Proof. We argue by induction on the dimension n. The base case $n = 1$ was proved in MATH 147. For the induction step, assume that Bolzano–Weierstrass holds in $\mathbb{R}^n$, and take a bounded sequence $(\vec{x}_k)$ in $\mathbb{R}^{n+1}$. Write $\vec{x}_k = (\vec{y}_k, t_k)$, where $\vec{y}_k = (x_k^{(1)}, \dots, x_k^{(n)}) \in \mathbb{R}^n$ and $t_k = x_k^{(n+1)} \in \mathbb{R}$.

Claim 1: $(\vec{y}_k)$ is a bounded sequence in $\mathbb{R}^n$, and (t_k) forms a bounded sequence in $\mathbb{R}$. *Verification*: Since $(\vec{x}_k)$ is bounded, there exists $r > 0$ such that $\|\vec{x}_k\| \leq r$ for all $k \in \mathbb{N}$. Therefore

$$r^2 \geq \|\vec{x}_k\|^2 = \|\vec{y}_k\|^2 + t_k^2 \quad \text{for all } k \in \mathbb{N}.$$

It follows that $\|\vec{y}_k\| \leq r$ and $t_k \in [-r, r]$ for all $k \in \mathbb{N}$. This proves Claim 1.

Claim 2: We can find an infinite set $M = \{k(1), k(2), \dots, k(m), \dots\} \subseteq \mathbb{N}$ such that the subsequence $(\vec{y}_{k(m)}) = (\vec{y}_k)_{k \in M}$ is convergent to some limit $\vec{b} \in \mathbb{R}^n$. *Verification*: This follows from the induction hypothesis (Bolzano–Weierstrass in $\mathbb{R}^n$) applied to the bounded sequence $(\vec{y}_k) \in \mathbb{R}^n$.

Claim 3: Let M be the same as in Claim 2, and consider the sequence $(t_k)_{k \in M} \in \mathbb{R}$. Then we can find an infinite subset $P \subseteq M$ such that $(t_k)_{k \in P}$ is convergent to some limit $t \in \mathbb{R}$. *Verification*: $(t_k)_{k \in M}$ is a bounded sequence in $\mathbb{R}$ because it is a subsequence of $(t_k)_{k=1}^{\infty}$, which is bounded by Claim 1. Apply the Bolzano–Weierstrass theorem in $\mathbb{R}$. Claim 3 follows.

Claim 4: Let P be as in Claim 3. Then the subsequence $(\vec{x}_k)_{k \in P}$ of $(\vec{x}_k)_{k=1}^{\infty}$ is convergent to the limit $\vec{a} = (\vec{b}, t) \in \mathbb{R}^{n+1}$. This claim finishes the proof. *Verification*: We have infinite sets P and M, with $P \subseteq M \subseteq \mathbb{N}$, such that $\vec{y}_k \xrightarrow{k \in M} \vec{b}$ and $t_k \xrightarrow{k \in P} t$, by Claim 2 and Claim 3. Since $P \subseteq M$, it follows that $\vec{y}_k \xrightarrow{k \in P} \vec{b}$. Together with $t_k \xrightarrow{k \in P} t$, Lemma 4.1 gives

$$\vec{x}_k \xrightarrow{k \in P} (\vec{b}, t).$$

$\square$

Lecture 5 — Wednesday, September 18, 2013

Continuity

Definition 5.1 Let $m, n \in \mathbb{N}$ and $A \subseteq \mathbb{R}^n$ (assume $A \neq \emptyset$). Let $f : A \to \mathbb{R}^m$ be a function and let $\vec{a} \in A$. We say f is **continuous at $\vec{a}$** when the following happens:

$$\text{(continuous at } \vec{a}): \quad \begin{array}{l} \text{for every } \varepsilon > 0, \text{ there exists } \delta > 0 \\ \text{such that } \|f(\vec{x}) - f(\vec{a})\| < \varepsilon \\ \text{whenever } \vec{x} \in A \text{ and } \|\vec{x} - \vec{a}\| < \delta. \end{array}$$

Geometrically, given $\varepsilon > 0$, there exists $\delta > 0$ such that $f(A \cap B(\vec{a};\delta)) \subseteq B(f(\vec{a});\varepsilon)$, as shown in Figure 3.

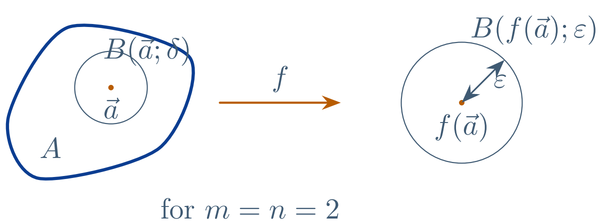

Figure 3

We can think of continuity as an ε-δ game against an adversary called Mr. Opponent. Mr. Opponent chooses $\varepsilon > 0$, and you must respond with $\delta > 0$ such that $f(A \cap B(\vec{a};\delta)) \subseteq B(f(\vec{a});\varepsilon)$. If you can always respond, then f is continuous at $\vec{a}$. If Mr. Opponent wins even once, then f is not continuous there.

Definition 5.2 Let $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}^m$, and $\vec{a} \in A$ as before. We say that the function f **respects sequences in A which converge to $\vec{a}$** when the following happens:

$$\text{(R-Seq)}: \quad \begin{array}{l} \text{whenever } (\vec{x}_k)_{k=1}^{\infty} \text{ is a sequence in } A \\ \text{such that } \vec{x}_k \to \vec{a}, \text{ it follows that } f(\vec{x}_k) \to f(\vec{a}). \end{array}$$

Proposition 5.3 Let $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}^m$, and $\vec{a} \in A$. Then f respects sequences in A which converge to $\vec{a}$ if and only if f is continuous at $\vec{a}$.

Proof. *Forward implication.* Assume that (R-Seq) holds. We must prove continuity at $\vec{a}$. Fix $\varepsilon > 0$, and suppose for contradiction that no $\delta > 0$ makes (continuous at $\vec{a}$) true. Then:

$$\delta = 1 \implies \text{there exists } \vec{x}_1 \in A \text{ such that } \|\vec{x}_1 - \vec{a}\| < 1 \text{ and } \|f(\vec{x}_1) - f(\vec{a})\| \geq \varepsilon.$$

Likewise, for $\delta = 1/2$ we obtain $\vec{x}_2 \in A$ with $\|\vec{x}_2 - \vec{a}\| < 1/2$ and $\|f(\vec{x}_2) - f(\vec{a})\| \geq \varepsilon$, and in general for $\delta = 1/k$ we obtain $\vec{x}_k \in A$ such that

$$\|\vec{x}_k - \vec{a}\| < \frac{1}{k} \qquad \text{and} \qquad \|f(\vec{x}_k) - f(\vec{a})\| \geq \varepsilon.$$

Thus $(\vec{x}_k)$ is a sequence in A with $\vec{x}_k \to \vec{a}$ by the squeeze theorem, while $f(\vec{x}_k) \not\to f(\vec{a})$ because the distance from $f(\vec{x}_k)$ to $f(\vec{a})$ is always at least ε. This contradicts (R-Seq), so the required δ must exist.

Reverse implication. Assume that f is continuous at $\vec{a}$, and let $(\vec{x}_k)$ be a sequence in A with $\vec{x}_k \to \vec{a}$. Given $\varepsilon > 0$, continuity supplies $\delta > 0$ such that

$$\vec{x} \in A \text{ and } \|\vec{x} - \vec{a}\| < \delta \implies \|f(\vec{x}) - f(\vec{a})\| < \varepsilon.$$

Since $\vec{x}_k \to \vec{a}$, there exists k_0 such that $\|\vec{x}_k - \vec{a}\| < \delta$ for all $k \geq k_0$. Hence $\|f(\vec{x}_k) - f(\vec{a})\| < \varepsilon$ for all $k \geq k_0$, so $f(\vec{x}_k) \to f(\vec{a})$. □

Definition 5.4 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. For every $\vec{x} \in A$, write explicitly $f(\vec{x}) = (f^{(1)}(\vec{x}), \ldots, f^{(m)}(\vec{x})) \in \mathbb{R}^m$. This gives m functions $f^{(1)}, f^{(2)}, \ldots, f^{(m)} : A \to \mathbb{R}$. These functions are called the **component functions** of f.

Proposition 5.5 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. Let $f^{(1)}, \ldots, f^{(m)} : A \to \mathbb{R}$ be the component functions of f, and let $\vec{a} \in A$. Then f is continuous at $\vec{a}$ if and only if each of $f^{(1)}, \ldots, f^{(m)}$ is continuous at $\vec{a}$.

Proof. *Claim 1*: f respects sequences in A which converge to $\vec{a}$ if and only if each of $f^{(1)}, \ldots, f^{(m)}$ respects such sequences.

Verification: Take a sequence $(\vec{x}_k) \subseteq A$ with $\vec{x}_k \to \vec{a}$. For each k, write

$$f(\vec{x}_k) = (f^{(1)}(\vec{x}_k), \ldots, f^{(m)}(\vec{x}_k)).$$

By Proposition 3.2, the sequence $f(\vec{x}_k)$ converges to $f(\vec{a})$ if and only if each component sequence $f^{(j)}(\vec{x}_k)$ converges to $f^{(j)}(\vec{a})$. This proves Claim 1.

Lecture 6 — Friday, September 20, 2013

Proof continued. *Claim 2*: f is continuous at $\vec{a}$ if and only if each of $f^{(1)}, \ldots, f^{(m)}$ is continuous at $\vec{a}$.

Verification: By Proposition 5.3, f is continuous at $\vec{a}$ if and only if it respects sequences in A converging to $\vec{a}$. By Claim 1, this is equivalent to each component function respecting sequences converging to $\vec{a}$. Applying Proposition 5.3 again to each component function, we obtain the result. $\square$

Some Stronger Forms of Continuity

> **Definition 6.1** Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. For f to be **continuous on $\boldsymbol{A}$** means that f is continuous at every $\vec{a} \in A$.

We have two explicit ways of describing "f is continuous on A": The ε-δ definition can be written as

$$\begin{aligned}(\varepsilon\text{-}\delta \text{ on } A): \quad & \text{given } \vec{a} \in A \text{ and } \varepsilon > 0, \text{ we can find } \delta > 0 \\ & \text{such that } \|f(\vec{x}) - f(\vec{a})\| < \varepsilon \\ & \text{for all } \vec{x} \in A \text{ with } \|\vec{x} - \vec{a}\| < \delta.\end{aligned}$$

Note that δ depends on $\vec{a}$ and ε.

The sequences definition can be written as

$$\text{(R-Seq on } A) : \quad \begin{array}{l}\text{whenever } (\vec{x}_k) \text{ is a sequence in } A \text{ convergent to a limit } \vec{a} \in A, \\ \text{it follows that } f(\vec{x}_k) \to f(\vec{a}) \text{ in } \mathbb{R}^m.\end{array}$$

So f respects all convergent sequences in A with limits still in A.

Definition 6.2 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. We say that f is **uniformly continuous on *A*** when the following happens:

$$\text{(Unif-Cont on } A) : \quad \begin{array}{l}\text{for every } \varepsilon > 0, \text{ there exists } \delta > 0 \\ \text{such that } \|f(\vec{x}_1) - f(\vec{x}_2)\| < \varepsilon \\ \text{whenever } \vec{x}_1, \vec{x}_2 \in A \text{ and } \|\vec{x}_1 - \vec{x}_2\| < \delta.\end{array}$$

Here δ depends *only* on ε.

Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. If f is uniformly continuous on A, then f is continuous on A. Indeed, assume that f satisfies (Unif-Cont on A). Let $\varepsilon > 0$ and $\vec{a} \in A$ be given. By uniform continuity, there exists $\delta > 0$ such that

$$\vec{x}_1, \vec{x}_2 \in A \text{ and } \|\vec{x}_1 - \vec{x}_2\| < \delta \implies \|f(\vec{x}_1) - f(\vec{x}_2)\| < \varepsilon.$$

This same δ works for continuity at $\vec{a}$: simply set $\vec{x}_2 = \vec{a}$ and rename $\vec{x}_1$ as $\vec{x}$. Thus (ε-δ on A) is satisfied.

Example 6.3 "Continuous on A" does not imply "uniformly continuous on A".

Proof. Take $n = 2$, $m = 1$. Let

$$A = (0,1) \times (0,1) = \{(s,t) \mid 0 < s,t < 1\}.$$

Let $f : A \to \mathbb{R}$ be defined by $f(s,t) = \frac{s}{t}$ for $(s,t) \in A$. We claim the following:

Claim 1: f is continuous.

Verification of Claim 1: We use the sequential criterion. Suppose $(\vec{x}_k) \subseteq A$ and $\vec{x}_k \to \vec{a} \in A$. Write $\vec{x}_k = (s_k, t_k)$ and $\vec{a} = (s,t)$. By Proposition 3.2, $\vec{x}_k \to \vec{a}$ if and only if $s_k \to s$ and $t_k \to t$.

Since $t > 0$, the quotient rule for sequences gives

$$f(\vec{x}_k) = \frac{s_k}{t_k} \longrightarrow \frac{s}{t} = f(\vec{a}).$$

Therefore f is continuous on A.

Claim 2: f is not uniformly continuous.

Verification of Claim 2: Take $\varepsilon = 1$. Can we find $\delta > 0$ such that if $\vec{x}_1, \vec{x}_2 \in A$ and $\|\vec{x}_1 - \vec{x}_2\| < \delta$, then $\|f(\vec{x}_1) - f(\vec{x}_2)\| < 1$? $(\diamond)$

Assume by contradiction that we can find a δ that works. Take $k \in \mathbb{N}$ such that $\frac{1}{k} < \delta$. Put $\vec{x}_1 = (\frac{1}{k}, \frac{1}{k})$ and $\vec{x}_2 = (\frac{1}{k}, \frac{1}{3k})$. Then

$$\|\vec{x}_1 - \vec{x}_2\| = \left\|(0, \frac{2}{3k})\right\| = \frac{2}{3k} < \frac{1}{k} < \delta.$$

Also,

$$f(\vec{x}_1) = \frac{1/k}{1/k} = 1 \qquad \text{and} \qquad f(\vec{x}_2) = \frac{1/k}{1/3k} = 3.$$

Hence

$$|f(\vec{x}_1) - f(\vec{x}_2)| = |1 - 3| = 2 > 1,$$

a contradiction. Therefore we cannot find a δ that satisfies $(\diamond)$.

This shows that no single δ can work for all points of A, so f is not uniformly continuous on A. $\square$

Exercise 6.4 Suppose that $(\vec{x}_k)$ is bounded but $\vec{x}_k \not\to \vec{a}$. It may still happen that some subsequences of $(\vec{x}_k)$ converge to $\vec{a}$. Is it possible that *every* convergent subsequence of $(\vec{x}_k)$ converges to $\vec{a}$?

Prove that the answer is no. A useful strategy is to adapt the set notation from the Bolzano–Weierstrass proof and then pass to a further subsequence.

Lecture 7 — Monday, September 23, 2013

For a function $f : A \to \mathbb{R}^m$ with $A \subseteq \mathbb{R}^n$, we have now defined both continuity on A and uniform continuity on A. Later we will see that on compact sets these two notions coincide. For example, when $n = 1$, the interval $[\alpha, \beta]$ is compact because it is closed and bounded.

Definition 7.1 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$.

1. Let $c \geq 0$ be a constant. We say f is ***c*-Lipschitz on *A*** to mean that

$$(c\text{-Lip}) : \quad \|f(\vec{x}_1) - f(\vec{x}_2)\| \leq c \cdot \|\vec{x}_1 - \vec{x}_2\| \quad \text{for all } \vec{x}_1, \vec{x}_2 \in A.$$

2. We say f is **Lipschitz on *A*** if there exists $c \geq 0$ such that f is c-Lipschitz on A.

Proposition 7.2 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. If f is Lipschitz on A, then f is uniformly continuous on A.

Proof. Assume that f satisfies (c-Lip) on A for some $c \geq 0$. We verify (Unif-Cont on A). Let $\varepsilon > 0$, and set

$$\delta = \frac{\varepsilon}{c+1} > 0.$$

If $\vec{x}_1, \vec{x}_2 \in A$ satisfy $\|\vec{x}_1 - \vec{x}_2\| < \delta$, then

$$\|f(\vec{x}_1) - f(\vec{x}_2)\| \leq c\|\vec{x}_1 - \vec{x}_2\| < c\delta = c\frac{\varepsilon}{c+1} < \varepsilon.$$

Hence f is uniformly continuous on A. □

Example 7.3 Let $n = 2, m = 1$ and

$$A = \overline{B}(\vec{0}; 1) = \{\vec{x} \in \mathbb{R}^2 \mid \|\vec{x}\| \leq 1\}.$$

Define $f : A \to \mathbb{R}$ by $f(\vec{x}) = \sqrt{\|\vec{x}\|}$, $\vec{x} \in A$.

Claim 1: f is uniformly continuous on A. To verify this, recall that

$$\left|\sqrt{u}-\sqrt{v}\right| \le \sqrt{|u-v|} \quad \text{for all } u,v \ge 0.$$

The reverse triangle inequality therefore gives

$$|f(\vec{x}_1)-f(\vec{x}_2)| \le \sqrt{\left|\|\vec{x}_1\|-\|\vec{x}_2\|\right|} \le \sqrt{\|\vec{x}_1-\vec{x}_2\|}.$$

Thus $\delta=\varepsilon^2$ works for every $\varepsilon>0$.

Claim 2: f is not Lipschitz on A. Assume by contradiction that there exists $c \ge 0$ such that

$$|f(\vec{x}_1)-f(\vec{x}_2)| \le c \cdot \|\vec{x}_1-\vec{x}_2\| \quad \text{for all } \vec{x}_1,\vec{x}_2 \in A. \tag{$*$}$$

Choose $k \in \mathbb{N}$ with $k>c$, and set $\vec{x}_1=\left(\frac{1}{k^2},0\right)$ and $\vec{x}_2=(0,0)$. Then

$$f(\vec{x}_1)=\sqrt{\|\vec{x}_1\|}=\frac{1}{k} \qquad \text{and} \qquad f(\vec{x}_2)=0,$$

and

$$\|\vec{x}_1-\vec{x}_2\|=\frac{1}{k^2}.$$

Substituting into ($*$) gives

$$\frac{1}{k}=|f(\vec{x}_1)-f(\vec{x}_2)| \le c\cdot\frac{1}{k^2},$$

so $k \le c$, a contradiction. Therefore f is not Lipschitz on A.

Definition 7.4 (Restriction) $A \subseteq \mathbb{R}^n$, $f: A \to \mathbb{R}^m$. Assume $A \ne \emptyset$. If $B \subseteq A$ is nonempty, then $f|_B$ denotes the **restriction** of f to B; this is the function $g: B \to \mathbb{R}^m$ defined by $g(\vec{x})=f(\vec{x})$ for every $\vec{x} \in B$. We say that f is continuous on B when the restriction $f|_B$ is continuous.

For example, the sequential formulation says that whenever $(\vec{y}_k)$ is a sequence in B with $\vec{y}_k \to \vec{b} \in B$, then $f(\vec{y}_k) \to f(\vec{b})$ in $\mathbb{R}^m$.

Example 7.5 For a complete example, let

$$A=[0,1]\times[0,1]=\{(s,t) \mid 0 \le s,t \le 1\} \subseteq \mathbb{R}^2,$$

$$B=\{(0,t) \mid 0 \le t \le 1\}.$$

See Figure 4. Let $f : A \to \mathbb{R}$ be defined by

$$f(s,t) = \begin{cases} \frac{1}{s}, & \text{if } 0 < s \leq 1, \\ 0, & \text{if } s = 0. \end{cases}$$

Then f is continuous on B, because the restriction $g = f|_B$ is the constant zero function on B. However, f is not continuous on all of A: if $\vec{x}_k = \left(\frac{1}{k}, \frac{1}{2}\right)$, then $\vec{x}_k \to \left(0, \frac{1}{2}\right)$ in A, but

$$f(\vec{x}_k) = k \not\to 0 = f\left(0, \frac{1}{2}\right),$$

so f does not respect the sequence $(\vec{x}_k)$ in A.

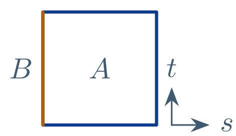

FIGURE 4

Next, we will specialize the general framework $f : A \to \mathbb{R}^m$ in two useful directions, namely scalar-valued functions ($m = 1$) and paths ($n = 1$).

Lecture 8 — Wednesday, September 25, 2013

2. Topology of $\mathbb{R}^n$

C^1-Paths in $\mathbb{R}^m$

In the general framework "$A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}^m$," we now set $n = 1$, so $A \subseteq \mathbb{R}$. In particular, we take A to be an open interval

$$A = \{t \in \mathbb{R} \mid \alpha < t < \beta\} = (\alpha, \beta),$$

where $\alpha \in \mathbb{R} \cup \{-\infty\}$ and $\beta \in \mathbb{R} \cup \{+\infty\}$. We use I or J for intervals of this form and reserve A for general subsets.

Definition 8.1 Let $I \subseteq \mathbb{R}$ be an open interval and let $m \in \mathbb{N}$. A function $f : I \to \mathbb{R}^m$ is called a **path in $\mathbb{R}^m$**; its domain-to-image interpretation is shown in Figure 5.

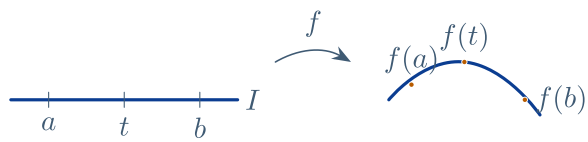

Figure 5

Let $f : I \to \mathbb{R}^m$ be a path, with component functions $f^{(1)}, \dots, f^{(m)}$, so that

$$f(t) = (f^{(1)}(t), \dots, f^{(m)}(t)) \quad (t \in I).$$

By Proposition 5.5, for any fixed $a \in I$, the function f is continuous at a if and only if each component $f^{(1)}, \dots, f^{(m)}$ is continuous at a. Hence f is continuous on I if and only if each component function is continuous on I. This enables differentiability for paths.

Definition 8.2 Let $f : I \to \mathbb{R}^m$ be a path in $\mathbb{R}^m$, with components $f^{(1)}, \dots, f^{(m)} : I \to \mathbb{R}$.

1. We say that f is a **differentiable path** to mean that each of $f^{(1)}, \dots, f^{(m)}$ is a differentiable function (in the sense of MATH 147).
2. Suppose f is differentiable. Then for every $t \in I$, define

$$f'(t) := ((f^{(1)})'(t), \dots, (f^{(m)})'(t)) \in \mathbb{R}^m.$$

   This vector is called the **velocity vector** of f at t. Its length $\|f'(t)\|$ is called the **speed** of f at t.
3. f is said to be a **C^1-path** if it is differentiable as in (1) and if the new function $f' : I \to \mathbb{R}^m$ is continuous.

The derivative $f' : I \to \mathbb{R}^m$ is continuous if and only if each component derivative $(f^{(j)})' : I \to \mathbb{R}$ is continuous by Proposition 5.5. Thus f is a C^1-path if and only if each component $f^{(j)}$ is differentiable and has continuous derivative.

Definition 8.3 Let $a < b$ in I. The **length** of the path between a and b is defined as

$$L = \int_a^b \|f'(t)\| \, \mathrm{d}t.$$

If $f : I \to \mathbb{R}^m$ is a C^1-path, then the length integral is well defined. Indeed, if we set

$$u(t) = \left\| f'(t) \right\| \quad (t \in I),$$

then $u : I \to \mathbb{R}$ is continuous and hence integrable by MATH 148. To check that u is continuous, use respect for sequences, for example: take $t_k \to t \in I$ and so $f'(t_k) \to f'(t) \in \mathbb{R}^m$ because $f' : I \to \mathbb{R}^m$ is continuous. Then taking norms,

$$\left\| f'(t_k) \right\| \to \left\| f'(t) \right\| = u(t).$$

This uses the continuity of the norm itself.

Exercise 8.4 Prove that the norm $\vec{x} \mapsto \|\vec{x}\|$ is continuous.

Example 8.5 Let $I = (-1, 2)$ and $m = 2$. Define $f : I \to \mathbb{R}^2$ by

$$f(t) = (e^t \cos 2\pi t, e^t \sin 2\pi t), \quad t \in I.$$

Equivalently, $f(t) = e^t(\cos 2\pi t, \sin 2\pi t)$. See Figure 6. Compute the length of this path between 0 and 1.

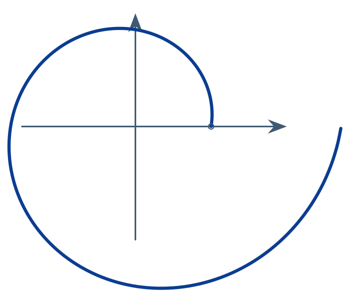

Figure 6

Solution. Some sample values are

$$f(0) = e^0(1,0) = (1,0), \qquad f(1/4) = e^{1/4}(0,1) = (0, e^{1/4}),$$

$$f(1/2) = (-\sqrt{e}, 0) \qquad \text{and} \qquad f(1) = (e, 0).$$

We compute

$$f'(t) = (e^t \cos(2\pi t) - 2\pi e^t \sin(2\pi t), e^t \sin(2\pi t) + 2\pi e^t \cos(2\pi t)).$$

Therefore

$$\|f'(t)\|^2 = e^{2t}\left((\cos(2\pi t) - 2\pi\sin(2\pi t))^2 + (\sin(2\pi t) + 2\pi\cos(2\pi t))^2\right) = e^{2t}(1+4\pi^2)$$

and

$$\|f'(t)\| = e^t\sqrt{1+4\pi^2}$$

so

$$L = \int_0^1 e^t\sqrt{1+4\pi^2}\,\mathrm{d}t = \sqrt{1+4\pi^2}(e-1).$$

$\square$

Lecture 9 — Friday, September 27, 2013

Here is an explanation for the length formula. Let $I = (\alpha, \beta) \subseteq \mathbb{R}$ be an open interval, and let $f : I \to \mathbb{R}^m$ be a C^1-path. Write

$$f(t) = (f^{(1)}(t), \ldots, f^{(m)}(t)) \qquad \text{and} \qquad f'(t) = ((f^{(1)})'(t), \ldots, (f^{(m)})'(t)).$$

If we set $u(t) = \|f'(t)\|$, then u is the speed of the path at time t. For $a < b$ in I, the length of the path between times a and b is defined by

$$L = \int_a^b u(t)\,\mathrm{d}t.$$

To explain this, take a division $\Delta = (t_0, t_1, \ldots, t_p)$ of the interval $[a, b]$, so

$$a = t_0 < t_1 < t_2 < \cdots < t_p = b.$$

The marked subdivision in Figure 7 represents this partition.

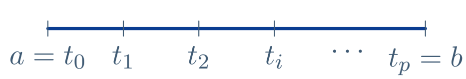

FIGURE 7

Its mesh is

$$\operatorname{mesh}(\Delta) = \max(t_1 - t_0, t_2 - t_1, \ldots, t_p - t_{p-1}).$$

It is natural to approximate L by the polygonal length

$$L_\Delta \coloneqq \sum_{i=1}^{p} \|f(t_i) - f(t_{i-1})\|.$$

To see why this behaves like a Riemann sum, consider $\|f(t) - f(s)\|$ for

$$a \leq s \leq t \leq b.$$

We have

$$\|f(t) - f(s)\| = \sqrt{\sum_{j=1}^{m} (f^{(j)}(t) - f^{(j)}(s))^2}.$$

For each j, the function $f^{(j)}$ is continuous on $[s,t]$ and differentiable on (s,t), so the mean value theorem yields some $c_j \in (s,t)$ such that

$$\frac{f^{(j)}(t) - f^{(j)}(s)}{t-s} = (f^{(j)})'(c_j).$$

Therefore

$$\|f(t) - f(s)\| = (t-s)\sqrt{\sum_{j=1}^{m} ((f^{(j)})'(c_j))^2}.$$

If $(t-s)$ is small, then $(f^{(j)})'(c_j) \approx (f^{(j)})'(t)$ because $(f^{(j)})'$ is continuous. Hence

$$\|f(t) - f(s)\| \approx (t-s)\sqrt{\sum_{j=1}^{m} ((f^{(j)})'(t))^2} = (t-s)\big\|f'(t)\big\| \qquad (\star)$$

Hence, when $(t-s)$ is small, we have

$$\|f(t) - f(s)\| \approx (t-s)\, u(t).$$

Returning to L_Δ, this suggests

$$L_\Delta = \sum_{i=1}^{p} \|f(t_i) - f(t_{i-1})\| \approx \sum_{i=1}^{p} (t_i - t_{i-1}) u(t_i),$$

which is a Riemann sum for u. If we choose a sequence of divisions (Δ_k) with mesh tending to 0, then the corresponding polygonal lengths converge to

$$L = \int_a^b u(t)\,\mathrm{d}t.$$

We now return to the general framework $f : A \to \mathbb{R}^m$ where $A \subseteq \mathbb{R}^n$. We ask what kind of sets A play the role of open intervals.

Interior and Closure for Subsets of $\mathbb{R}^n$

Definition 9.1 Let $A \subseteq \mathbb{R}^n$.

1. A point $\vec{a} \in A$ is said to be an **interior point** of A when there exists $r > 0$ such that $B(\vec{a}; r) \subseteq A$.
2. A point $\vec{b} \in \mathbb{R}^n$ is said to be **adherent** to A when $B(\vec{b}; r) \cap A \neq \emptyset$ for every $r > 0$.

Definition 9.2

1. The set of all interior points of A is called the **interior of $\boldsymbol{A}$**, denoted $\operatorname{int}(A)$.
2. The set of all points $\vec{b} \in \mathbb{R}^n$ that are adherent to A is called the **closure of $\boldsymbol{A}$**, denoted $\operatorname{cl}(A)$.

We will see that $\operatorname{int}(A) \subseteq A \subseteq \operatorname{cl}(A)$.

Lecture 10 — Monday, September 30, 2013

Recall that for $A \subseteq \mathbb{R}^n$,

$$\operatorname{int}(A) = \{\vec{a} \in A \mid \text{there exists } r > 0 \text{ such that } B(\vec{a}; r) \subseteq A\},$$

and

$$\operatorname{cl}(A) = \{\vec{b} \in \mathbb{R}^n \mid B(\vec{b}; r) \cap A \neq \emptyset \text{ for every } r > 0\}.$$

For every $A \subseteq \mathbb{R}^n$, we have $\mathrm{int}(A) \subseteq A \subseteq \mathrm{cl}(A)$. The first inclusion is immediate from the definition (1) of interior. For the second inclusion, given $\vec{a} \in A$ and $r > 0$, we have $\vec{a} \in B(\vec{a}; r) \cap A$, hence $B(\vec{a}; r) \cap A \neq \emptyset$. Therefore $\vec{a} \in \mathrm{cl}(A)$.

Definition 10.1 The set difference $\mathrm{cl}(A) \setminus \mathrm{int}(A)$ is called the **boundary of A**, denoted $\mathrm{bd}(A)$.

Example 10.2 Let $n = 2$. Define

$$A = \{(s,t) \mid t > 0,\ s,t \in \mathbb{R}\} \cup \{(s,0) \mid s \geq 0,\ s \in \mathbb{R}\}.$$

See Figure 8. Let $\vec{a} = (s,t) \in A$ with $t > 0$. Then $\vec{a} \in \mathrm{int}(A)$. To check this, we must find $r > 0$ such that $B(\vec{a}; r) \subseteq A$. For instance, take $r = \frac{t}{2}$.

Now let $\vec{a} = (s,0) \in A$ with $s \geq 0$. Then $\vec{a} \notin \mathrm{int}(A)$. Indeed, no matter what $r > 0$ we pick, we have $B(\vec{a}; r) \not\subseteq A$. (For instance, $(s, -\frac{r}{2}) \in B(\vec{a}; r)$ but $(s, -\frac{r}{2}) \notin A$.) Hence

$$\mathrm{int}(A) = \{(s,t) \mid t > 0, s, t \in \mathbb{R}\}.$$

In general, taking interiors can remove points.

Now consider $\mathrm{cl}(A)$. We already know that $A \subseteq \mathrm{cl}(A)$. If $\vec{b} = (s,t)$ with $t < 0$, then $\vec{b} \notin \mathrm{cl}(A)$, since choosing $r = -\frac{t}{2}$ gives $B(\vec{b}; r) \cap A = \emptyset$. On the other hand, if $\vec{b} = (s,0)$ with $s < 0$, then $\vec{b} \in \mathrm{cl}(A)$ because for every $r > 0$, the point $(s, r/2)$ lies in $B(\vec{b}; r) \cap A$.

Conclusion: $\mathrm{cl}(A) = \{(s,t) \mid t \geq 0, s, t \in \mathbb{R}\}$. Finally,

$$\mathrm{bd}(A) = \mathrm{cl}(A) \setminus \mathrm{int}(A) = \{(s,0) \mid s \in \mathbb{R}\}.$$

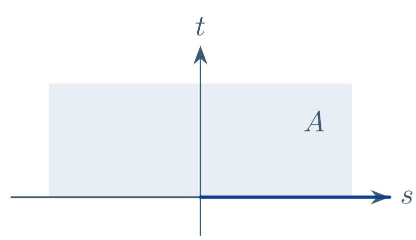

FIGURE 8

Proposition 10.3 Let $A \subseteq \mathbb{R}^n$ and let $\vec{b} \in \mathbb{R}^n$. Then $\vec{b} \in \text{cl}(A)$ if and only if there exists a sequence $(\vec{x}_k) \subseteq A$ such that $\vec{x}_k \to \vec{b}$.

Proof. *Forward implication.* Assume that $\vec{b} \in \text{cl}(A)$. Then for every $k \in \mathbb{N}$, the ball $B(\vec{b}; \frac{1}{k})$ meets A, so we may choose

$$\vec{x}_k \in B\left(\vec{b}; \frac{1}{k}\right) \cap A.$$

Hence $\left\|\vec{x}_k - \vec{b}\right\| < \frac{1}{k}$ for every k, and therefore $\vec{x}_k \to \vec{b}$.

Reverse implication. Conversely, suppose there exists a sequence $(\vec{x}_k) \subseteq A$ with $\vec{x}_k \to \vec{b}$. Let $r > 0$. Since $\vec{x}_k \to \vec{b}$, there exists $k_0 \in \mathbb{N}$ such that $\left\|\vec{x}_k - \vec{b}\right\| < r$ for all $k \geq k_0$. In particular, $\vec{x}_{k_0} \in B(\vec{b}; r) \cap A$, so $B(\vec{b}; r) \cap A \neq \emptyset$. Hence $\vec{b} \in \text{cl}(A)$. $\square$

Proposition 10.4 (Duality of interior and closure) For every $A \subseteq \mathbb{R}^n$, we have that

$$\text{int}(\mathbb{R}^n \setminus A) = \mathbb{R}^n \setminus \text{cl}(A)$$

and

$$\text{cl}(\mathbb{R}^n \setminus A) = \mathbb{R}^n \setminus \text{int}(A).$$

Proof. We verify the first identity; the second follows similarly. Suppose $\vec{b} \in \text{int}(\mathbb{R}^n \setminus A)$. Then there exists $r > 0$ such that $B(\vec{b}; r) \subseteq \mathbb{R}^n \setminus A$. Hence $B(\vec{b}; r) \cap A = \emptyset$, so $\vec{b} \notin \text{cl}(A)$. Therefore $\vec{b} \in \mathbb{R}^n \setminus \text{cl}(A)$.

Conversely, suppose $\vec{b} \in \mathbb{R}^n \setminus \text{cl}(A)$. Then $\vec{b} \notin \text{cl}(A)$, so by Definition 9.1 there exists $r > 0$ such that $B(\vec{b}; r) \cap A = \emptyset$. This means $B(\vec{b}; r) \subseteq \mathbb{R}^n \setminus A$, and therefore $\vec{b} \in \text{int}(\mathbb{R}^n \setminus A)$. $\square$

Corollary 10.5 For every $A \subseteq \mathbb{R}^n$, we have $\text{bd}(A) = \text{cl}(A) \cap \text{cl}(\mathbb{R}^n \setminus A)$.

Proof.

$$\text{bd}(A) = \text{cl}(A) \setminus \text{int}(A) = \text{cl}(A) \cap (\mathbb{R}^n \setminus \text{int}(A)) = \text{cl}(A) \cap \text{cl}(\mathbb{R}^n \setminus A).$$

by Proposition 10.4 $\square$

Lecture 11 — Wednesday, October 02, 2013

Last time we proved that for every $A \subseteq \mathbb{R}^n$,

$$\operatorname{bd}(A) = \operatorname{cl}(A) \cap \operatorname{cl}(\mathbb{R}^n \setminus A).$$

Remark 11.1

1. For every $A \subseteq \mathbb{R}^n$, we have $\operatorname{bd}(A) = \operatorname{bd}(\mathbb{R}^n \setminus A)$.
2. For every $A \subseteq \mathbb{R}^n$, the boundary admits the sequential description

$$\operatorname{bd}(A) = \left\{ \vec{b} \in \mathbb{R}^n \;\middle|\; \begin{array}{l} \text{there exists a sequence } (\vec{x}_k) \subseteq A \text{ with } \vec{x}_k \to \vec{b}, \\ \text{and there exists a sequence } (\vec{y}_k) \subseteq \mathbb{R}^n \setminus A \text{ with } \vec{y}_k \to \vec{b} \end{array} \right\}.$$

Open and Closed Sets in $\mathbb{R}^n$

In Lecture 10 we saw that $\operatorname{int}(A) \subseteq A \subseteq \operatorname{cl}(A)$ for every $A \subseteq \mathbb{R}^n$. Usually both inclusions are strict, as in Example 10.2. Special cases in which equality holds lead to the definitions of open and closed sets.

Definition 11.2

1. $A \subseteq \mathbb{R}^n$ is said to be **open** when it satisfies $A = \operatorname{int}(A)$. That is, A is open if and only if every $\vec{a} \in A$ is an interior point of A.
2. $A \subseteq \mathbb{R}^n$ is said to be **closed** when it satisfies $A = \operatorname{cl}(A)$. That is, A is closed if and only if A has no adherent points $\vec{b} \in \mathbb{R}^n \setminus A$.

Warning Most subsets of $\mathbb{R}^n$ are neither open nor closed. In particular, proving that A is not closed does *not* prove that A is open. Heed this warning! To prove that A is open, you must verify that every point of A is an interior point.

Proposition 11.3 For $A \subseteq \mathbb{R}^n$, the set A is closed if and only if $\mathbb{R}^n \setminus A$ is open.

Proof. *Forward implication.* Suppose that A is closed. Then $\mathrm{cl}(A) = A$. By Proposition 10.4,

$$\mathrm{int}(\mathbb{R}^n \setminus A) = \mathbb{R}^n \setminus \mathrm{cl}(A) = \mathbb{R}^n \setminus A.$$

Therefore $\mathbb{R}^n \setminus A$ is open.

Reverse implication. Suppose that $\mathbb{R}^n \setminus A$ is open. Then $\mathbb{R}^n \setminus A = \mathrm{int}(\mathbb{R}^n \setminus A)$ by Definition 11.2. By duality, $\mathbb{R}^n \setminus A = \mathbb{R}^n \setminus \mathrm{cl}(A)$. Taking complements gives $\mathrm{cl}(A) = A$, so A is closed. $\square$

Definition 11.4 We say that a subset $A \subseteq \mathbb{R}^n$ has the "no-escape" property when

$$\text{(NO-Esc)} = \left\{ \text{whenever } (\vec{x}_k) \text{ is a sequence in } A \text{ such that } (\vec{x}_k) \to \vec{b} \in \mathbb{R}^n, \text{ then } \vec{b} \in A \right\}.$$

In words, a convergent sequence in A cannot escape A. This is equivalent to saying that A is closed.

Proposition 11.5 For $A \subseteq \mathbb{R}^n$, A is closed if and only if A has the "no-escape" property.

Proof. *Forward implication.* Assume that A is closed. Then $A = \mathrm{cl}(A)$ by Definition 11.2. Let $(\vec{x}_k)$ be a sequence in A such that $\vec{x}_k \to \vec{b} \in \mathbb{R}^n$. By Proposition 10.3, this implies $\vec{b} \in \mathrm{cl}(A)$. Since $\mathrm{cl}(A) = A$, we get $\vec{b} \in A$. Hence A has the no-escape property.

Reverse implication. Assume that A has the no-escape property. To show that A is closed, by Definition 11.2 it is enough to prove that $A = \mathrm{cl}(A)$. We always have $A \subseteq \mathrm{cl}(A)$, so it remains to show $\mathrm{cl}(A) \subseteq A$. Let $\vec{b} \in \mathrm{cl}(A)$. By Proposition 10.3, there exists a sequence $(\vec{x}_k) \subseteq A$ such that $\vec{x}_k \to \vec{b}$. The no-escape property then gives $\vec{b} \in A$. Therefore $\mathrm{cl}(A) \subseteq A$, and thus $A = \mathrm{cl}(A)$. Hence A is closed. $\square$

For $A \subseteq \mathbb{R}^n$, the set A is open if and only if every point of A is an interior point. For closed sets we now have three equivalent descriptions: A is closed if and only if $A = \mathrm{cl}(A)$, if and only if $\mathbb{R}^n \setminus A$ is open, if and only if A has the no-escape property. Note that different textbooks take any one of those as the original definition and prove the others from it as propositions.

For every $\vec{a} \in \mathbb{R}^n$ and $r > 0$, the open ball $B(\vec{a}; r)$ is an open set and the closed ball $\overline{B}(\vec{a}; r)$ is a closed set. To prove the first assertion, take $\vec{x} \in B(\vec{a}; r)$ and put

$$\rho = r - \|\vec{x} - \vec{a}\| > 0.$$

If $\vec{y} \in B(\vec{x}; \rho)$, then the triangle inequality gives

$$\|\vec{y} - \vec{a}\| \le \|\vec{y} - \vec{x}\| + \|\vec{x} - \vec{a}\| < r,$$

so $B(\vec{x}; \rho) \subseteq B(\vec{a}; r)$. For the closed ball, let $(\vec{x}_k)$ be a sequence in $\overline{B}(\vec{a}; r)$ with $\vec{x}_k \to \vec{x}$. The reverse triangle inequality implies

$$|\|\vec{x}_k - \vec{a}\| - \|\vec{x} - \vec{a}\|| \le \|\vec{x}_k - \vec{x}\| \to 0.$$

Passing to the limit in $\|\vec{x}_k - \vec{a}\| \le r$ gives $\|\vec{x} - \vec{a}\| \le r$. Thus the closed ball has the no-escape property and is closed.

Lecture 12 — Friday, October 04, 2013

Recall that $A \subseteq \mathbb{R}^n$ is open if and only if every point of A is an interior point, that is, if and only if $\text{int}(A) = A$. We also have several equivalent characterizations of closed sets.

We now prove two lemmas that lead to a useful characterization of interiors.

Lemma 12.1 Let $A \subseteq \mathbb{R}^n$. Suppose $D \subseteq A$ such that D is an open set. Then $D \subseteq \text{int}(A)$.

Proof. If $D = \emptyset$, there is nothing to prove. Otherwise, let $\vec{x} \in D$. Since D is open, there exists $r > 0$ such that $B(\vec{x}; r) \subseteq D$. Because $D \subseteq A$, the same radius also satisfies $B(\vec{x}; r) \subseteq A$. Hence $\vec{x} \in \text{int}(A)$. $\square$

Lemma 12.2 Let $A \subseteq \mathbb{R}^n$. Then $\text{int}(A)$ is an open set.

Proof. If $\text{int}(A) = \emptyset$, then it is open. Otherwise, set $T = \text{int}(A)$ and let $\vec{x} \in T$. Since $\vec{x} \in \text{int}(A)$, there exists $r > 0$ such that $B(\vec{x}; r) \subseteq A$. The ball $B(\vec{x}; r)$ is itself an open set, so Lemma 12.1 gives $B(\vec{x}; r) \subseteq \text{int}(A) = T$. Hence T is open. $\square$

Proposition 12.3 For every $A \subseteq \mathbb{R}^n$, the set $\text{int}(A)$ is the largest open set contained in A. That is:

i. $\text{int}(A) \subseteq A$, and $\text{int}(A)$ is open.

ii. if $D \subseteq \mathbb{R}^n$ is an open set such that $D \subseteq A$, then $D \subseteq \text{int}(A)$.

Proof. (i.) follows from Lemma 12.2. (ii.) follows from Lemma 12.1. $\square$

Proposition 12.4 For every $A \subseteq \mathbb{R}^n$, the set $\text{cl}(A)$ is the smallest closed set which contains A. That is:

j. $\text{cl}(A) \supseteq A$, and $\text{cl}(A)$ is closed.

jj. if $F \subseteq \mathbb{R}^n$ is a closed set such that $F \supseteq A$, then $F \supseteq \text{cl}(A)$.

Proof. Let M be the complement of A, $M = \mathbb{R}^n \setminus A$. Apply Proposition 12.3 to M. It says (i.): $\text{int}(M) \subseteq M$ and $\text{int}(M)$ is open, and (ii.): if $D \subseteq M$ and D is open, then $D \subseteq \text{int}(M)$. We will show that (i.) and (ii.) for M imply (12.4) and (12.4) for A.

Verification of (12.4): since $\text{int}(M)$ is open, the set $\mathbb{R}^n \setminus \text{int}(M)$ is closed. But $\text{cl}(A) = \text{cl}(\mathbb{R}^n \setminus M) = \mathbb{R}^n \setminus \text{int}(M)$. Hence $\text{cl}(A)$ is closed. We also have $A \subseteq \text{cl}(A)$ for every $A \subseteq \mathbb{R}^n$, so both assertions in (12.4) hold.

Verification of (12.4): let $F \subseteq \mathbb{R}^n$ be closed and suppose $F \supseteq A$. Set $D = \mathbb{R}^n \setminus F$. Then D is open and $D \subseteq M = \mathbb{R}^n \setminus A$. By (ii.) applied to M, we obtain $D \subseteq \text{int}(M)$. Taking complements yields

$$F = \mathbb{R}^n \setminus D \supseteq \mathbb{R}^n \setminus \text{int}(M) = \text{cl}(A).$$

$\square$

Compact Subsets of $\mathbb{R}^n$

Definition 12.5 A set $A \subseteq \mathbb{R}^n$ is said to be **bounded** if there exists $r > 0$ such that $\|\vec{x}\| \leq r$ for every $\vec{x} \in A$. Equivalently, $A \subseteq \overline{B}(\vec{0}; r)$ for some $r > 0$.

Definition 12.6 $A \subseteq \mathbb{R}^n$ is said to be **compact** when it is closed and bounded.

There are several equivalent descriptions of compactness. Some of them extend to more general spaces than $\mathbb{R}^n$. The one most useful here is sequential compactness.

Definition 12.7 A set $A \subseteq \mathbb{R}^n$ is **sequentially compact** if every sequence $(\vec{x}_k)$ in A has a convergent subsequence $(\vec{x}_{k(p)})$ whose limit still belongs to A.

We will show that Definitions 12.6 and 12.7 are equivalent.

Lecture 13 — Monday, October 07, 2013

Recall that $A \subseteq \mathbb{R}^n$ is compact when it is closed and bounded. Also, $A \subseteq \mathbb{R}^n$ is sequentially compact when the following condition holds: for every sequence $(\vec{x}_k)$ in A, there is a convergent subsequence $(\vec{x}_{k(p)})$ such that $\vec{a} = \lim_{p\to\infty} \vec{x}_{k(p)}$ still belongs to A.

Theorem 13.1 For $A \subseteq \mathbb{R}^n$, A is compact if and only if A is sequentially compact.

Proof. *Forward implication.* Assume that A is closed and bounded, and let $(\vec{x}_k)$ be a sequence in A. Since A is bounded, the sequence $(\vec{x}_k)$ is bounded. By Bolzano–Weierstrass, it has a convergent subsequence $(\vec{x}_{k(p)})$. Write

$$\vec{x}_{k(p)} \to \vec{a}.$$

Because A is closed, it has the no-escape property, so $\vec{a} \in A$. Hence A is sequentially compact.

Reverse implication. Assume that A is sequentially compact. First we show that A is bounded. Suppose not, for a contradiction. Then for every $k \in \mathbb{N}$ we can choose $\vec{x}_k \in A$ such that

$$\|\vec{x}_k\| > k.$$

This gives a sequence $(\vec{x}_k)$ in A. By sequential compactness, there exists a subsequence $(\vec{x}_{k(p)})$ that converges in $\mathbb{R}^n$ (in fact, to a point of A). Every convergent sequence is bounded, so $(\vec{x}_{k(p)})$ is bounded. But

$$\left\|\vec{x}_{k(p)}\right\| > k(p) \geq p,$$

so $(\vec{x}_{k(p)})$ is unbounded, a contradiction. Hence A is bounded.

Next we show that A is closed. Let $(\vec{y}_m)$ be any sequence in A with $\vec{y}_m \to \vec{b} \in \mathbb{R}^n$. By sequential compactness, $(\vec{y}_m)$ has a subsequence $(\vec{y}_{m(q)})$ that converges to some $\vec{a} \in A$. Since $(\vec{y}_m) \to \vec{b}$, every subsequence also converges to $\vec{b}$, so $\vec{y}_{m(q)} \to \vec{b}$. Limits are unique, therefore $\vec{a} = \vec{b}$. Thus $\vec{b} \in A$. So A has the no-escape property, and by Proposition 11.5, A is closed.

Therefore A is closed and bounded, so A is compact. □

The Extreme Value Theorem

Definition 13.2 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$. A point $\vec{a} \in A$ is a point of **global minimum** for f on A if $f(\vec{a}) \leq f(\vec{x})$ for every $\vec{x} \in A$. Similarly, $\vec{b} \in A$ is a point of **global maximum** if $f(\vec{x}) \leq f(\vec{b})$ for every $\vec{x} \in A$.

Points of global minimum or global maximum may or may not exist. When they exist, they may or may not be unique.

Example 13.3

$$A = (0,1] \times (0,1] = \{(s,t) \mid 0 < s,t \leq 1\} \subseteq \mathbb{R}^2.$$

Let $f : A \to \mathbb{R}$ be given by $f(s,t) = |s-t|$. For every $\vec{x} \in A$, we have $0 \leq f(\vec{x}) < 1$. The global minimum is $0 = f(s,s)$. Also,

$$\sup\{f(\vec{x}) \mid \vec{x} \in A\} = 1,$$

which is not achieved, so there are no points of global maximum for f on A.

Example 13.4 Take the same set A as in Example 13.3. Let $g : A \to \mathbb{R}$ be given by $g(s,t) = \frac{s}{t}$. Then $\inf\{g(\vec{x}) \mid \vec{x} \in A\} = 0$ and $\sup\{g(\vec{x}) \mid \vec{x} \in A\} = \infty$. So neither a global maximum nor a global minimum is achieved.

Theorem 13.5 (Extreme value theorem) If $A \subseteq \mathbb{R}^n$ is nonempty and compact and $f : A \to \mathbb{R}$ is continuous, then f has at least one point of global minimum and at least one point of global maximum on A.

The proof will follow from Proposition 13.6 and Remark 13.7 below.

Proposition 13.6 Let $A \subseteq \mathbb{R}^n$ be compact, and let $f : A \to \mathbb{R}^m$ be continuous. Then the **image set**

$$M := f(A) = \{\vec{y} \in \mathbb{R}^m \mid \text{there exists } \vec{x} \in A \text{ such that } f(\vec{x}) = \vec{y}\}$$

is compact.

Remark 13.7 Let K be a nonempty compact subset of $\mathbb{R}$. Boundedness gives finite values $\alpha = \inf(K)$ and $\beta = \sup(K)$. We now show that these bounds belong to K; then $\alpha = \min(K)$ and $\beta = \max(K)$.

But why are α and β in K? Assume $\alpha \notin K$. For every $m \in \mathbb{N}$, we can find $t_m \in K$ such that

$$\alpha \leq t_m < \alpha + \frac{1}{m};$$

otherwise $\alpha + 1/m$ would be a lower bound for K larger than its infimum. The squeeze theorem gives $t_m \to \alpha$. Since K is closed, its no-escape property gives $\alpha \in K$, a contradiction.

The argument for β is symmetric. For every $m \in \mathbb{N}$, choose $s_m \in K$ such that

$$\beta - \frac{1}{m} < s_m \leq \beta;$$

otherwise $\beta - 1/m$ would be an upper bound for K smaller than its supremum. Then $s_m \to \beta$, so closedness gives $\beta \in K$. Therefore $\alpha = \min(K)$ and $\beta = \max(K)$.

Lecture 14 — Wednesday, October 09, 2013

Recall Proposition 13.6: if $A \subseteq \mathbb{R}^n$ is compact and $f : A \to \mathbb{R}^m$ is continuous, then the image set

$$M := f(A) = \{\vec{y} \in \mathbb{R}^m \mid \text{there exists } \vec{x} \in A \text{ such that } f(\vec{x}) = \vec{y}\}$$

is compact.

Proof of Proposition 13.6. We prove that M is sequentially compact. Let $(\vec{y}_k)$ be a sequence in M. Since each $\vec{y}_k$ lies in $f(A)$, there exists $\vec{x}_k \in A$ such that $f(\vec{x}_k) = \vec{y}_k$. Because A is compact, it is sequentially compact, so some subsequence $(\vec{x}_{k(p)})$ converges to a point $\vec{a} \in A$. Continuity of f then gives

$$\vec{y}_{k(p)} = f(\vec{x}_{k(p)}) \to f(\vec{a}) \in M.$$

Thus every sequence in M has a convergent subsequence with limit in M, so M is compact. □

Recall the extreme value theorem: if $A \subseteq \mathbb{R}^n$ is nonempty and compact and $f : A \to \mathbb{R}$ is continuous, then f has at least one point of global minimum and at least one point of global maximum on A.

To prove the theorem, combine Proposition 13.6 and Remark 13.7 from the previous lecture: if $\emptyset \neq K \subseteq \mathbb{R}$ is compact, then there exist $\alpha, \beta \in K$ such that $\alpha \leq t \leq \beta$ for every $t \in K$.

Proof of Theorem 13.5. Let $A \subseteq \mathbb{R}^n$ be nonempty and compact, and let $f : A \to \mathbb{R}$ be continuous. Consider the image set $K = f(A) \subseteq \mathbb{R}$. By Proposition 13.6, the set K is compact and nonempty. By Remark 13.7, it has a minimum α and a maximum β. Since $\alpha, \beta \in f(A)$, there exist $\vec{a}, \vec{b} \in A$ such that $f(\vec{a}) = \alpha$ and $f(\vec{b}) = \beta$. Then for every $\vec{x} \in A$, we have $f(\vec{x}) \in K$, so

$$f(\vec{a}) \leq f(\vec{x}) \leq f(\vec{b}).$$

□

3. Differential calculus in several variables

Partial Derivatives

Definition 14.1 Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}$, and let $\vec{a} = (a^{(1)}, \dots, a^{(n)}) \in \operatorname{int}(A)$. Fix $r > 0$ such that $B(\vec{a}; r) \subseteq A$. For a direction $i \in \{1, \dots, n\}$, the function $u : (a^{(i)} - r, a^{(i)} + r) \to \mathbb{R}$ given by

$$u(t) = f(a^{(1)}, \dots, a^{(i-1)}, t, a^{(i+1)}, \dots, a^{(n)})$$

for $a^{(i)} - r < t < a^{(i)} + r$ is said to be a **partial function** of f around $\vec{a}$, in direction i.

Why does the definition of u make sense? Let $\vec{x} = (a^{(1)}, \dots, a^{(i-1)}, t, a^{(i+1)}, \dots, a^{(n)})$ with $a^{(i)} - r < t < a^{(i)} + r$. Then $\left|a^{(i)} - t\right| < r$, and so

$$\|\vec{x} - \vec{a}\| = \left|t - a^{(i)}\right| < r.$$

Hence $\vec{x} \in B(\vec{a}; r) \subseteq A$, so $f(\vec{x})$ is well-defined.

Example 14.2 Take $n = 2$ and $i = 2$. Define $u : (a^{(2)} - r, a^{(2)} + r) \to \mathbb{R}$ by $u(t) = f(a^{(1)}, t)$. See Figure 9.

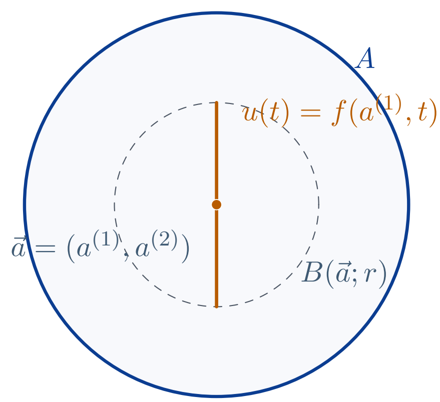

Figure 9

Definition 14.3 Consider the same framework as in Definition 14.1, and let $u : (a^{(i)} - r, a^{(i)} + r) \to \mathbb{R}$ be the partial function of f around $\vec{a}$ in direction i. If u is differentiable at $t = a^{(i)}$, we say that f has a **partial derivative** at $\vec{a}$ in direction i, and we use the notation

$$(\partial_i f)(\vec{a}) \coloneqq u'(a^{(i)}).$$

Lecture 15 — Friday, October 11, 2013

Recall that $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}$, and $\vec{a} = (a^{(1)}, \ldots, a^{(n)}) \in \operatorname{int}(A)$, so there exists $r > 0$ such that $B(\vec{a}; r) \subseteq A$. For a direction $i \in \{1, \ldots, n\}$, define the partial function $u : (a^{(i)} - r, a^{(i)} + r) \to \mathbb{R}$ by

$$u(t) = f(a^{(1)}, \ldots, t, \ldots, a^{(n)})$$

for $a^{(i)} - r < t < a^{(i)} + r$. If u is differentiable at $t = a^{(i)}$, then

$$(\partial_i f)(\vec{a}) \coloneqq u'(a^{(i)}).$$

Example 15.1 Take $n = 3$, $A = \mathbb{R}^3$, and $f(x, y, z) = x^2 y - 5z$ for $(x, y, z) \in \mathbb{R}^3$. Pick $\vec{a} = (1, 2, 3)$. We may choose any $r > 0$; for instance, take $r = 11$ since today is October 11th. Pick $i = 1$. To compute $(\partial_1 f)(\vec{a})$, define

$$u(t) = f(t, 2, 3) = t^2 \cdot 2 - 5(3) = 2t^2 - 15$$

for $1 - r < t < 1 + r$, that is, $-10 < t < 12$. Note that $u(1) = f(1, 2, 3) = -13$. Since u is differentiable, $u'(t) = 4t$, so $u'(1) = 4$. Therefore, $(\partial_1 f)(1, 2, 3) = 4$.

For the function in Example 15.1, we can get a general formula for $\partial_1 f$ at a general point $\vec{a} = (x_0, y_0, z_0)$. As in Example 15.1,

$$u(t) = f(t, y_0, z_0) = t^2 y_0 - 5z_0$$

for $x_0 - r < t < x_0 + r$, where r is arbitrary. This u is differentiable, and $u'(t) = 2t y_0$. Hence $(\partial_1 f)(x_0, y_0, z_0)$ exists and

$$(\partial_1 f)(x_0, y_0, z_0) = 2x_0 y_0.$$

So for $f(x, y, z) = x^2 y - 5z$, the partial derivative $\partial_1 f$ is defined at every point in $\mathbb{R}^3$, and $(\partial_1 f) = 2xy$.

Example 15.2 Let $n = 2$, $A = (0, \infty)^2 \subseteq \mathbb{R}^2$, and define $f : A \to \mathbb{R}$ by $f(p, q) = p^q$ for all $p, q > 0$. Then $\partial_q f$ is defined at every $(p, q) \in A$, and

$$(\partial_q f)(p, q) = p^q \log(p)$$

and

$$(\partial_p f)(p, q) = qp^{q-1}$$

for all $(p, q) \in A$.

Definition 15.3 Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}$, and let $\vec{a} \in \operatorname{int}(A)$. Suppose that f has partial derivatives at $\vec{a}$ in all directions $1 \leq i \leq n$. The vector

$$((\partial_1 f)(\vec{a}), \ldots, (\partial_n f)(\vec{a})) \in \mathbb{R}^n$$

is called the **gradient vector** of f at $\vec{a}$, denoted by $(\nabla f)(\vec{a})$.

Definition 15.4 Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}$.

1. A point $\vec{a} \in A$ is a point of **local minimum** for f if there exists $r > 0$ such that $f(\vec{a}) \leq f(\vec{x})$ for all $\vec{x} \in A \cap B(\vec{a}; r)$.
2. A point $\vec{b} \in A$ is a point of **local maximum** for f if there exists $r > 0$ such that $f(\vec{b}) \geq f(\vec{x})$ for all $\vec{x} \in A \cap B(\vec{b}; r)$.
3. A point of A which is either a local minimum or a local maximum is called a **point of local extremum**.

Proposition 15.5 Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}$, and let $\vec{a} \in \operatorname{int}(A)$ such that f has partial derivatives at $\vec{a}$ in all directions. If $\vec{a}$ is a point of local extremum for f, then

$$(\nabla f)(\vec{a}) = ((\partial_1 f)(\vec{a}), \ldots, (\partial_n f)(\vec{a})) = \vec{0} \in \mathbb{R}^n.$$

The proof is deferred to the next lecture. It may help to review **Fermat's theorem on stationary points** from MATH 147 first.

Lecture 16 — Friday, October 18, 2013

Proof of Proposition 15.5. The point $\vec{a}$ is either a local minimum or a local maximum. We treat the case of a local minimum; the other case is analogous. Fix $i \in \{1, \ldots, n\}$. Choose $r_1 > 0$ such that

$$f(\vec{a}) \leq f(\vec{x}) \quad \text{for all } \vec{x} \in B(\vec{a}; r_1) \cap A,$$

and choose $r_2 > 0$ such that $B(\vec{a}; r_2) \subseteq A$. Set $r = \min\{r_1, r_2\}$. Then $B(\vec{a}; r) \subseteq A$, and $f(\vec{a}) \leq f(\vec{x})$ for every $\vec{x} \in B(\vec{a}; r)$. Now consider the partial function

$$u(t) = f(a^{(1)}, \ldots, a^{(i-1)}, t, a^{(i+1)}, \ldots, a^{(n)})$$

for $t \in (a^{(i)} - r, a^{(i)} + r)$. This function is differentiable at $a^{(i)}$, and

$$u'(a^{(i)}) = (\partial_i f)(\vec{a}).$$

Moreover, $u(a^{(i)}) \leq u(t)$ for all t in the interval. Indeed, if $t \in (a^{(i)} - r, a^{(i)} + r)$, then

$$\vec{x} = (a^{(1)}, \ldots, a^{(i-1)}, t, a^{(i+1)}, \ldots, a^{(n)}) \in B(\vec{a}; r),$$

and

$$u(t) = f(\vec{x}) \geq f(\vec{a}) = u(a^{(i)}).$$

By Fermat's theorem, a differentiable function has derivative 0 at an interior local extremum. Therefore $(\partial_i f)(\vec{a}) = u'(a^{(i)}) = 0$. □

Directional Derivatives

Let $A \subseteq \mathbb{R}^n$, $f : A \to \mathbb{R}$, $\vec{a} \in \mathrm{int}(A)$, and let $i \in \{1, \ldots, n\}$. Suppose that $(\partial_i f)(\vec{a})$ exists. We can write it using the **Newton quotient** from MATH 147. Pick $r > 0$ such that $B(\vec{a}; r) \subseteq A$, and let $u : (a^{(i)} - r, a^{(i)} + r) \to \mathbb{R}$ be the partial function. Then

$$(\partial_i f)(\vec{a}) = u'(a^{(i)}) = \lim_{h \to 0} \frac{u(a^{(i)} + h) - u(a^{(i)})}{h}.$$

Recall that $u(a^{(i)}) = f(a^{(1)}, \ldots, a^{(i)}, \ldots, a^{(n)}) = f(\vec{a})$, and

$$u(a^{(i)} + h) = f(a^{(1)}, \ldots, a^{(i)} + h, \ldots, a^{(n)}) = f(\vec{a} + h \cdot \vec{e}_i).$$

Hence we can also write
$$(\partial_i f)(\vec{a}) = \lim_{h\to 0} \frac{f(\vec{a} + h \cdot \vec{e}_i) - f(\vec{a})}{h}.$$
This is the form that generalizes.

Definition 16.1 Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}$, and let $\vec{a} \in \text{int}(A)$. Let $\vec{v} \in \mathbb{R}^n$. If the limit
$$\lim_{h\to 0} \frac{f(\vec{a} + h\vec{v}) - f(\vec{a})}{h}$$
exists, then we say that f has a **directional derivative** in direction $\vec{v}$ at the point $\vec{a}$, and notate it as $(\partial_{\vec{v}} f)(\vec{a})$.

When $\vec{v} = \vec{e}_i$, this gives back $(\partial_{\vec{e}_i} f)(\vec{a}) = (\partial_i f)(\vec{a})$. In the same framework, if $\vec{v} = \vec{0} \in \mathbb{R}^n$, then $(\partial_{\vec{v}} f)(\vec{a})$ always exists and equals 0, since
$$\lim_{h\to 0} \frac{f(\vec{a} + h\vec{0}) - f(\vec{a})}{h} = \lim_{h\to 0} \frac{0}{h} = 0.$$

Proposition 16.2 (Homogeneity of directional derivatives) Let $A \subseteq \mathbb{R}^n$, let $f : A \to \mathbb{R}$, and let $\vec{a} \in \text{int}(A)$. Let $\vec{v} \neq \vec{0} \in \mathbb{R}^n$ and suppose that $(\partial_{\vec{v}} f)(\vec{a})$ exists. Then for every $\alpha \in \mathbb{R}$, the directional derivative $(\partial_{\alpha\vec{v}} f)(\vec{a})$ exists and is equal to
$$\alpha \cdot (\partial_{\vec{v}} f)(\vec{a}).$$

Proof. If $\alpha = 0$, this is immediate from the previous remark. So assume $\alpha \neq 0$, and set $\vec{w} := \alpha\vec{v}$. We must compute
$$\lim_{h\to 0} \frac{f(\vec{a} + h\vec{w}) - f(\vec{a})}{h} = \lim_{h\to 0} \frac{f(\vec{a} + h\alpha\vec{v}) - f(\vec{a})}{h}.$$
Let $s := h\alpha$. Then $h \to 0$ implies $s \to 0$, and
$$\lim_{s\to 0} \frac{f(\vec{a} + s\vec{v}) - f(\vec{a})}{s/\alpha} = \alpha \cdot \lim_{s\to 0} \frac{f(\vec{a} + s\vec{v}) - f(\vec{a})}{s}.$$
The latter limit exists by hypothesis, and equals $\alpha \cdot (\partial_{\vec{v}} f)(\vec{a})$. $\square$

What about additivity? Is it true that
$$(\partial_{\vec{v}+\vec{w}} f)(\vec{a}) = (\partial_{\vec{v}} f)(\vec{a}) + (\partial_{\vec{w}} f)(\vec{a})?$$

In general this need *not* hold.

Example 16.3 Let

$$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(s,t) = \begin{cases} \dfrac{s^2 t}{s^2+t^2}, & (s,t) \neq (0,0), \\ 0, & (s,t) = (0,0). \end{cases}$$

Fix $\vec{a} = (0,0)$. For $\vec{v} = (\alpha, \beta)$,

$$(\partial_{\vec{v}} f)(\vec{a}) = \lim_{h\to 0} \frac{f(h\alpha, h\beta) - f(0,0)}{h} = \lim_{h\to 0} \frac{f(h\alpha, h\beta)}{h}.$$

If $\alpha \neq 0$, then for $h \neq 0$ we have

$$f(h\alpha, h\beta) = \frac{(h\alpha)^2 (h\beta)}{(h\alpha)^2 + (h\beta)^2} = h\,\frac{\alpha^2\beta}{\alpha^2+\beta^2},$$

so

$$(\partial_{(\alpha,\beta)} f)(0,0) = \frac{\alpha^2\beta}{\alpha^2+\beta^2}.$$

If $\alpha = 0$, then $f(0, h\beta) = 0$ for all h, so $(\partial_{(0,\beta)} f)(0,0) = 0$. Thus all directional derivatives at $(0,0)$ exist.

Define

$$L(\vec{v}) \coloneqq (\partial_{\vec{v}} f)(0,0), \quad \vec{v} \in \mathbb{R}^2.$$

To test additivity, choose

$$\vec{v} = (1,1) \qquad \text{and} \qquad \vec{w} = (-1,1).$$

Then

$$L(\vec{v}) = (\partial_{(1,1)} f)(0,0) = \frac{1^2 \cdot 1}{1^2 + 1^2} = \frac{1}{2},$$

$$\text{and} \quad L(\vec{w}) = (\partial_{(-1,1)} f)(0,0) = \frac{(-1)^2 \cdot 1}{(-1)^2 + 1^2} = \frac{1}{2}.$$

But

$$\vec{v} + \vec{w} = (0,2),$$

and by the $\alpha = 0$ case,

$$L(\vec{v} + \vec{w}) = L(0,2) = 0.$$

Therefore
$$L(\vec{v}+\vec{w}) \neq L(\vec{v}) + L(\vec{w}),$$
so L is not additive.

Later we will prove that additivity does hold for sufficiently regular functions, in particular for C^1 functions.

Lecture 17 — Monday, October 21, 2013

The Mean Value Theorem

Definition 17.1 Fix $\vec{x}, \vec{y} \in \mathbb{R}^n$. A vector of the form
$$\vec{v} = (1-t)\vec{x} + t\vec{y}, \quad t \in [0,1]$$
is called a convex combination of $\vec{x}$ and $\vec{y}$.

A subset $A \subseteq \mathbb{R}^n$ is said to be convex when it has the following property:

$$\begin{aligned}(\text{Conv}): \quad & \text{For every } \vec{x}, \vec{y} \in A \text{ and every } t \in [0,1], \\ & \text{it follows that } (1-t)\vec{x} + t\vec{y} \text{ still belongs to } A.\end{aligned}$$

Also define the **line segment** in $\mathbb{R}^n$ connecting $\vec{x}$ and $\vec{y}$ as
$$L(\vec{x}, \vec{y}) \coloneqq \{(1-t)\vec{x} + t\vec{y} \mid t \in [0,1]\}.$$

In words, the condition (Conv) says that all convex combinations made with vectors in A must still belong to A.

Note that
$$(1-t)\vec{x} + t\vec{y} = \vec{x} + t(\vec{y} - \vec{x}).$$
When $t=0$ we get $\vec{x}$, when $t=1$ we get $\vec{y}$, and when $t \in (0,1)$ we get points between them.

Theorem 17.2 (Mean value theorem in direction $\vec{v}$) Let $A \subseteq \mathbb{R}^n$ be open, let $f : A \to \mathbb{R}$, and let $\vec{v} \in \mathbb{R}^n \setminus \{\vec{0}\}$. Suppose that $(\partial_{\vec{v}} f)(\vec{a})$ exists at every $\vec{a} \in A$. Let $\vec{x}, \vec{y} \in A$ satisfy $L(\vec{x}, \vec{y}) \subseteq A$ and $\vec{y} - \vec{x} = \alpha\vec{v}$ for some $\alpha \in \mathbb{R}$. Then there exists $\vec{b} \in L(\vec{x}, \vec{y})$ such that

$$f(\vec{y}) - f(\vec{x}) = \alpha \cdot (\partial_{\vec{v}} f)(\vec{b}) = (\partial_{\alpha\vec{v}} f)(\vec{b}) = (\partial_{\vec{y}-\vec{x}} f)(\vec{b}).$$

The segment and intermediate point in the directional mean value theorem are shown in Figure 10.

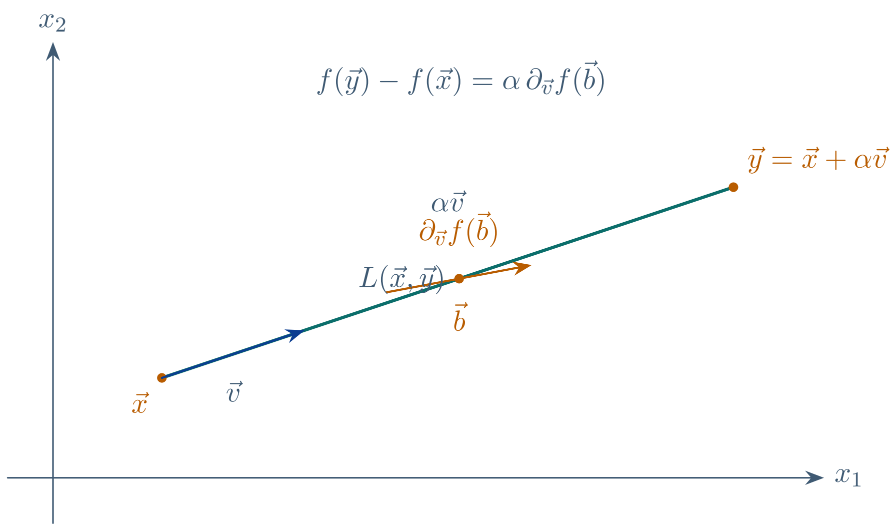

FIGURE 10

Proof. We discuss the case $\alpha \neq 0$; if $\alpha = 0$, then $\vec{x} = \vec{y}$ and the statement is immediate. Define the auxiliary function $\phi : [0, 1] \to \mathbb{R}$ by

$$\phi(t) = f((1 - t)\vec{x} + t\vec{y}).$$

Then $\phi(0) = f(\vec{x})$ and $\phi(1) = f(\vec{y})$, so the idea is to apply the mean value theorem in one variable to ϕ.

Claim 1: The function ϕ is continuous on $[0, 1]$ and differentiable on $(0, 1)$, with

$$\phi'(t) = \alpha \cdot (\partial_{\vec{v}} f)((1 - t)\vec{x} + t\vec{y}).$$

Verification: Fix $t_0 \in [0,1]$, and set $\vec{a} := (1-t_0)\vec{x} + t_0\vec{y}$. Consider

$$\frac{\phi(t_0+h)-\phi(t_0)}{h},$$

with $t_0 + h \in [0,1]$. We have

$$\phi(t_0) = f((1-t_0)\vec{x} + t_0\vec{y}) = f(\vec{a}),$$

and

$$\phi(t_0+h) = f((1-(t_0+h))\vec{x} + (t_0+h)\vec{y}) = f(\vec{a} + h(\vec{y}-\vec{x})) = f(\vec{a} + h\alpha\vec{v}).$$

So

$$\frac{\phi(t_0+h)-\phi(t_0)}{h} = \frac{f(\vec{a}+h\alpha\vec{v}) - f(\vec{a})}{h} = \frac{f(\vec{a}+h\alpha\vec{v}) - f(\vec{a})}{h\alpha} \cdot \alpha.$$

For admissible nonzero h, it follows that

$$\frac{\phi(t_0+h)-\phi(t_0)}{h} \longrightarrow \alpha \cdot (\partial_{\vec{v}} f)(\vec{a})$$

as $h \to 0$ through values for which $t_0 + h \in [0,1]$. Multiplying by h shows that $\phi(t_0+h) \to \phi(t_0)$, including at the endpoints. When $t_0 \in (0,1)$, the limit is two-sided and gives the displayed derivative formula.

Claim 2: There exists $c \in (0,1)$ such that $\phi(1) - \phi(0) = \phi'(c)$. *Verification:* Claim 1 supplies exactly the hypotheses of the mean value theorem in one variable, so there exists $c \in (0,1)$ with

$$\frac{\phi(1)-\phi(0)}{1-0} = \phi'(c).$$

Claim 3: There exists $\vec{b} \in L(\vec{x}, \vec{y})$ such that

$$f(\vec{y}) - f(\vec{x}) = \alpha \cdot (\partial_{\vec{v}} f)(\vec{b}).$$

Verification: Let $c \in (0,1)$ be as in Claim 2, and set $\vec{b} := (1-c)\vec{x} + c\vec{y}$. Then

$$f(\vec{y}) - f(\vec{x}) = \phi(1) - \phi(0) = \phi'(c),$$

and by Claim 1,

$$\phi'(c) = \alpha \cdot (\partial_{\vec{v}} f)((1-c)\vec{x} + c\vec{y}) = \alpha \cdot (\partial_{\vec{v}} f)(\vec{b}).$$

$\square$

Recall from Definition 17.1 that $A \subseteq \mathbb{R}^n$ is **convex** when

$$\vec{x}, \vec{y} \in A \implies L(\vec{x}, \vec{y}) \subseteq A.$$

Example 17.3 Let $A \subseteq \mathbb{R}^n$ be nonempty, open, and convex, and let $f : A \to \mathbb{R}$. Suppose that for every $\vec{a} \in A$ and $\vec{v} \in \mathbb{R}^n$, the derivative $(\partial_{\vec{v}} f)(\vec{a})$ exists and is equal to 0. Prove that f is constant; that is, there exists $\alpha \in \mathbb{R}$ such that $f(\vec{x}) = \alpha$ for all $\vec{x} \in A$.

Solution. Fix $\vec{x}_0 \in A$, and denote $f(\vec{x}_0) =: \alpha$. We show that $f(\vec{x}) = \alpha$ for all $\vec{x} \in A$. Given $\vec{x} \in A$, set $\vec{v} := \vec{x} - \vec{x}_0$. The hypotheses of Theorem 17.2 are satisfied, so

$$f(\vec{x}) - f(\vec{x}_0) = 1 \cdot (\partial_{\vec{v}} f)(\vec{b}) = 0$$

for some $\vec{b} \in L(\vec{x}_0, \vec{x})$. Hence $f(\vec{x}) = f(\vec{x}_0) = \alpha$. $\square$

Lecture 18 — Wednesday, October 23, 2013

C^1 functions

Definition 18.1 Let $A \subseteq \mathbb{R}^n$ be open, and let $f : A \to \mathbb{R}$. We say that f is a $\boldsymbol{C^1}$ **function** when:

1. f is continuous on A,
2. f has partial derivatives $(\partial_i f)(\vec{a})$ for every $\vec{a} \in A$ and every $1 \leq i \leq n$, and
3. the new functions $\partial_i f : A \to \mathbb{R}$ are continuous on A.

The collection of all C^1 functions from A to $\mathbb{R}$ is denoted $C^1(A, \mathbb{R})$. We also use the notation $C^0(A, \mathbb{R})$ to mean the set of all continuous functions from A to $\mathbb{R}$.

We now discuss linear approximations for C^1 functions. First, we review the case of one variable from MATH 147.

Take $n = 1$, let $A = (\alpha, \beta) \subseteq \mathbb{R}$ be an open interval, and fix $a \in A$. If $f : A \to \mathbb{R}$ is differentiable at a, then for $x \in A$ close to a we have the linear approximation

$$f(x) \approx f(a) + f'(a)(x - a).$$

Here $y = f(a) + f'(a)(x - a)$ is the tangent line in **point-slope form**, with slope $f'(a)$. This says more than just

$$\lim_{x \to a} \big(f(x) - f(a) - f'(a)(x - a)\big) = 0.$$

It also says that

$$\lim_{x \to a} \left| \frac{f(x) - f(a) - f'(a)(x - a)}{x - a} \right| = 0.$$

Indeed, the second limit can be rewritten as

$$\lim_{x \to a} \left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| = 0,$$

which is precisely the statement that

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} = f'(a).$$

Theorem 18.2 Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^1(A, \mathbb{R})$, and let $\vec{a} \in A$. Then:

$$\lim_{\vec{x} \to \vec{a}} \frac{|f(\vec{x}) - f(\vec{a}) - \langle \vec{x} - \vec{a}, (\nabla f)(\vec{a}) \rangle|}{\|\vec{x} - \vec{a}\|} = 0.$$

This dictionary might be helpful:

$\mathbb{R}$	$\mathbb{R}^n$
x, a	$\vec{x}, \vec{a}$
$f'(a)$	$(\nabla f)(\vec{a})$
xy	$\langle \vec{x}, \vec{y} \rangle$

The proof of Theorem 18.2 is deferred to the next lecture (because it's *long*). For now, we turn to the geometric interpretation in terms of the tangent hyperplane to the graph of f at the point $(\vec{a}, f(\vec{a}))$.

Definition 18.3 Let $f : A \to \mathbb{R}$. The **graph** of f is the set

$$\Gamma = \{(\vec{x}, t) \in \mathbb{R}^{n+1} \mid \vec{x} \in A,\, t = f(\vec{x})\}.$$

Definition 18.4 A **hyperplane** through a point $\vec{p} \in \mathbb{R}^{n+1}$ is a set of the form

$$H = \left\{ \vec{p} + \sum_{i=1}^{n} \alpha_i \vec{y}_i \;\middle|\; \alpha_1, \ldots, \alpha_n \in \mathbb{R} \right\},$$

where $\vec{y}_1, \ldots, \vec{y}_n \in \mathbb{R}^{n+1}$ is a linearly independent family.

If $\vec{z} \in \mathbb{R}^{n+1}$, $\vec{z} \neq \vec{0}$, and $\vec{z} \perp \vec{y}_i$ for every $1 \leq i \leq n$, then $\vec{z}$ is called **normal** to the hyperplane H.

If $\vec{z}_1, \vec{z}_2$ are both normal vectors to H, then there exists $\alpha \in \mathbb{R}$ such that $\vec{z}_1 = \alpha \vec{z}_2$, so the normal vector is unique up to rescaling.

Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^1(A, \mathbb{R})$, and let $\vec{a} \in A$. Consider the graph

$$\Gamma = \{(\vec{x}, f(\vec{x})) \mid \vec{x} \in A\} \subseteq \mathbb{R}^{n+1}$$

and $\vec{p} = (\vec{a}, f(\vec{a})) \in \Gamma$. The linear approximation says that if $\delta > 0$ is small enough, then $B(\vec{a}; \delta) \subseteq A$, and for $\vec{x} \in B(\vec{a}; \delta)$ we can replace

$$\vec{q} = (\vec{x}, f(\vec{x})) \in \Gamma$$

by

$$\vec{q}_L = (\vec{x}, f(\vec{a}) + \langle \vec{x} - \vec{a}, (\nabla f)(\vec{a}) \rangle).$$

Figure 11 illustrates this construction.

How do we find H so that $\vec{q}_L$ runs in H? Compute

$$\begin{aligned}
\vec{q}_L - \vec{p} &= (\vec{x}, f(\vec{a}) + \langle \vec{x} - \vec{a}, (\nabla f)(\vec{a}) \rangle) - (\vec{a}, f(\vec{a})) \\
&= (\vec{x} - \vec{a}, \langle \vec{x} - \vec{a}, (\nabla f)(\vec{a}) \rangle) \\
&= \left(v^{(1)}, \ldots, v^{(n)}, \sum_{i=1}^{n} v^{(i)} (\partial_i f)(\vec{a}) \right),
\end{aligned}$$

where $\vec{x} - \vec{a} = \vec{v} = (v^{(1)}, \ldots, v^{(n)})$. Thus

$$\vec{q}_L - \vec{p} = v^{(1)}\vec{y}_1 + \cdots + v^{(n)}\vec{y}_n$$

for suitable vectors $\vec{y}_1, \ldots, \vec{y}_n \in \mathbb{R}^{n+1}$.

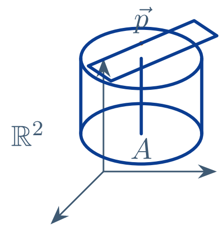

FIGURE 11

Lecture 19 — Friday, October 25, 2013

Recall the tangent hyperplane discussion for the graph of $f \in C^1(A, \mathbb{R})$. Let $A \subseteq \mathbb{R}^n$ be open, let $f : A \to \mathbb{R}$ be C^1, and set

$$\Gamma = \{(\vec{x}, f(\vec{x})) \mid \vec{x} \in A\} \subseteq \mathbb{R}^{n+1}.$$

Fix $\vec{a} \in A$ and $\vec{p} = (\vec{a}, f(\vec{a}))$. Take $\delta > 0$ such that $B(\vec{a}; \delta) \subseteq A$. For $\vec{x} = \vec{a} + \vec{v} \in B(\vec{a}; \delta)$, where $\vec{v} = (v^{(1)}, \ldots, v^{(n)})$ and $\|\vec{v}\| < \delta$, we have

$$\begin{aligned} \vec{q}_L - \vec{p} &= \left(v^{(1)}, \ldots, v^{(n)}, \sum_{i=1}^{n} v^{(i)} (\partial_i f)(\vec{a}) \right) \\ &= v^{(1)}(1, 0, \ldots, 0, \partial_1 f(\vec{a})) + \cdots + v^{(n)}(0, \ldots, 1, \partial_n f(\vec{a})) = v^{(1)}\vec{y}_1 + \cdots + v^{(n)}\vec{y}_n, \end{aligned}$$

where $\vec{y}_i = (0, \ldots, 1, 0, \ldots, (\partial_i f)(\vec{a})) \in \mathbb{R}^{n+1}$. This calculation motivates the following definition.

Definition 19.1 Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^1(A, \mathbb{R})$, and let $\Gamma \subseteq \mathbb{R}^{n+1}$ be the graph of f. Fix $\vec{a} \in A$, and write $\vec{p} = (\vec{a}, f(\vec{a})) \in \Gamma$. Then the hyperplane

$$H = \left\{ \vec{p} + \sum_{i=1}^{n} \alpha_i \vec{y}_i \;\middle|\; \alpha_1, \ldots, \alpha_n \in \mathbb{R} \right\}, \qquad \text{and} \qquad \vec{y}_i = (0, \ldots, 0, 1, 0, \ldots, (\partial_i f)(\vec{a})),$$

is called the **tangent hyperplane to $\boldsymbol{\Gamma}$ at $\vec{\boldsymbol{p}}$.** Scalar multiples of

$$\vec{z} = (-(\nabla f)(\vec{a}), 1)$$

are called **normal vectors to $\boldsymbol{\Gamma}$ at $\vec{\boldsymbol{p}}$.**

We have $\langle \vec{y}_i, \vec{z} \rangle = 0$ for all $1 \leq i \leq n$.

Lemma 19.2 Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^1(A, \mathbb{R})$, and let $\vec{a} \in A$. Let $r > 0$ satisfy $B(\vec{a}; r) \subseteq A$. Then for every $\vec{x} \in B(\vec{a}; r)$, there exist $\vec{b}_1, \ldots, \vec{b}_n \in B(\vec{a}; r)$ such that

$$f(\vec{x}) - f(\vec{a}) = \langle \vec{x} - \vec{a}, \vec{w} \rangle,$$

where

$$\vec{w} = ((\partial_1 f)(\vec{b}_1), \ldots, (\partial_n f)(\vec{b}_n)).$$

So $\vec{w}$ is "almost" the gradient vector.

Proof. Fix $\vec{x} \in B(\vec{a}; r)$. Define

$$\vec{x}_0 = \vec{a} = (a^{(1)}, \ldots, a^{(n)}),$$

$$\vec{x}_1 = (x^{(1)}, a^{(2)}, \ldots, a^{(n)}),$$

and more generally

$$\vec{x}_i = (x^{(1)}, \ldots, x^{(i)}, a^{(i+1)}, \ldots, a^{(n)}), \quad 1 \leq i \leq n,$$

so that $\vec{x}_n = \vec{x}$. Since $\|\vec{x}_i - \vec{a}\| \leq \|\vec{x} - \vec{a}\| < r$, each $\vec{x}_i$ belongs to $B(\vec{a}; r)$. Also,

$$\vec{x}_i - \vec{x}_{i-1} = (x^{(i)} - a^{(i)})\vec{e}_i.$$

Applying the mean value theorem in direction $\vec{e}_i$, we obtain some

$$\vec{b}_i \in L(\vec{x}_{i-1}, \vec{x}_i) \subseteq B(\vec{a}; r)$$

such that

$$f(\vec{x}_i) - f(\vec{x}_{i-1}) = (x^{(i)} - a^{(i)})(\partial_i f)(\vec{b}_i).$$

Summing over i gives

$$f(\vec{x}) - f(\vec{a}) = f(\vec{x}_n) - f(\vec{x}_0)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \big(f(\vec{x}_i) - f(\vec{x}_{i-1})\big) \\
&= \sum_{i=1}^{n} (x^{(i)} - a^{(i)})(\partial_i f)(\vec{b}_i) \\
&= \langle \vec{x} - \vec{a}, \vec{w} \rangle.
\end{aligned}
$$

$\square$

We continue to defer the proof of Theorem 18.2...

Lecture 20 — Monday, October 28, 2013

Proof of Theorem 18.2. We have $A \subseteq \mathbb{R}^n$ open, $f \in C^1(A, \mathbb{R})$, and $\vec{a} \in A$. Let us also fix $r_0 > 0$ such that $B(\vec{a}; r_0) \subseteq A$. (We work with balls $B(\vec{a}; r)$ with $r \leq r_0$ to ensure that $B(\vec{a}; r) \subseteq A$.) Given $\varepsilon > 0$, we want to find $0 < \delta \leq r_0$ such that

$$\frac{|f(\vec{x}) - f(\vec{a}) - \langle \vec{x} - \vec{a}, (\nabla f)(\vec{a}) \rangle|}{\|\vec{x} - \vec{a}\|} < \varepsilon, \text{ for all } \vec{x} \in B(\vec{a}; \delta) \setminus \{\vec{a}\}. \tag{20.1}$$

For every $1 \leq i \leq n$, we know that $(\partial_i f)(\vec{x})$ is defined and continuous for all $\vec{x} \in A$. Hence it is continuous at $\vec{x} = \vec{a}$. Therefore there exists r_i such that $0 < r_i \leq r_0$ and

$$\vec{x} \in B(\vec{a}; r_i) \implies$$

$$|(\partial_i f)(\vec{x}) - (\partial_i f)(\vec{a})| < \frac{\varepsilon}{n}. \tag{20.2}$$

Put $\delta = \min\{r_1, r_2, \ldots, r_n\}$, so $0 < \delta \leq r_0$. We claim that this δ is good for (20.1). Pick $\vec{x} \in B(\vec{a}; \delta) \setminus \{\vec{a}\}$, and we verify that

$$|f(\vec{x}) - f(\vec{a}) - \langle \vec{x} - \vec{a}, (\nabla f)(\vec{a}) \rangle| < \varepsilon \|\vec{x} - \vec{a}\|,$$

which is equivalent to (20.1). We use Lemma 19.2, since $\vec{x} \in B(\vec{a}; \delta) \subseteq A$. That lemma says that there exist $\vec{b}_1, \ldots, \vec{b}_n \in B(\vec{a}; \delta)$ such that $f(\vec{x}) - f(\vec{a}) = \langle \vec{x} - \vec{a}, \vec{w} \rangle$, where $\vec{w} =$

$((\partial_1 f)(\vec{b}_1), \ldots, (\partial_n f)(\vec{b}_n))$. Now we roll up our sleeves and calculate:

$$\begin{aligned}
&|f(\vec{x}) - f(\vec{a}) - \langle \vec{x} - \vec{a}, (\nabla f)(\vec{a})\rangle| \\
&= |\langle \vec{x} - \vec{a}, \vec{w}\rangle - \langle \vec{x} - \vec{a}, (\nabla f)(\vec{a})\rangle| \\
&= |\langle \vec{x} - \vec{a}, \vec{w} - (\nabla f)(\vec{a})\rangle| \\
&\leq \|\vec{x} - \vec{a}\| \cdot \|\vec{w} - (\nabla f)(\vec{a})\| \quad \text{(by the Cauchy–Schwarz inequality)} \\
&\leq \|\vec{x} - \vec{a}\| \cdot \|\vec{w} - (\nabla f)(\vec{a})\|_1 \quad \text{(by Example 2.8)} \\
&= \|\vec{x} - \vec{a}\| \cdot \sum_{i=1}^{n} |(\partial_i f)(\vec{b}_i) - (\partial_i f)(\vec{a})| < \|\vec{x} - \vec{a}\| \cdot \sum_{i=1}^{n} \frac{\varepsilon}{n} \quad (\text{since } \vec{b}_i \in B(\vec{a};\delta) \subseteq B(\vec{a};r_i)) \\
&= \|\vec{x} - \vec{a}\| \cdot \varepsilon.
\end{aligned} \tag{20.3}$$

□

Here is an application of linear approximation:

Corollary 20.1 Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^1(A, \mathbb{R})$, and let $\vec{a} \in A$. Then for every $\vec{v} \in \mathbb{R}^n$, the directional derivative $(\partial_{\vec{v}} f)(\vec{a})$ exists, and

$$(\partial_{\vec{v}} f)(\vec{a}) = \langle \vec{v}, (\nabla f)(\vec{a})\rangle.$$

In particular, for $f \in C^1(A, \mathbb{R})$ we have

$$\partial_{\vec{v}_1 + \vec{v}_2} f(\vec{a}) = \partial_{\vec{v}_1} f(\vec{a}) + \partial_{\vec{v}_2} f(\vec{a}),$$

since the inner product is bilinear.

Proof. Assume $\vec{v} \neq \vec{0}$; if $\vec{v} = \vec{0}$, the statement is immediate. In linear approximation, choose

$$\vec{x} = \vec{a} + h\vec{v}.$$

Then $\vec{x} \to \vec{a}$ with $\vec{x} \neq \vec{a}$ corresponds to $h \to 0$ with $h \neq 0$. Multiplying by $\|\vec{v}\|$, we obtain

$$\begin{aligned}
&\lim_{h\to 0} \frac{|f(\vec{a} + h\vec{v}) - f(\vec{a}) - \langle h\vec{v}, (\nabla f)(\vec{a})\rangle|}{\|(\vec{a} + h\vec{v}) - \vec{a}\|} \cdot \|\vec{v}\| = 0 \\
&\implies \lim_{h\to 0} \frac{|f(\vec{a} + h\vec{v}) - f(\vec{a}) - h\langle \vec{v}, (\nabla f)(\vec{a})\rangle|}{|h|} = 0.
\end{aligned}$$

Equivalently,

$$\lim_{h\to 0}\left|\frac{f(\vec{a}+h\vec{v})-f(\vec{a})}{h}-\langle\vec{v},(\nabla f)(\vec{a})\rangle\right|=0.$$

Thus the directional derivative exists and equals $\langle\vec{v},(\nabla f)(\vec{a})\rangle$. □

Lecture 21 — Wednesday, October 30, 2013

The Chain Rule

Let $A\subseteq\mathbb{R}^n$ be open and let $f,g\in C^1(A,\mathbb{R})$. We can perform the usual operations with f and g. For example, if $h=fg$, meaning $h(\vec{x})=f(\vec{x})g(\vec{x})$ for all $\vec{x}\in A$, then $h\in C^1(A,\mathbb{R})$ and

$$(\partial_i(fg))=(\partial_i f)\,g+f\,(\partial_i g),\quad 1\le i\le n.$$

Also, for $\alpha,\beta\in\mathbb{R}$, the linear combination $(\alpha f+\beta g):A\to\mathbb{R}$ given by

$$(\alpha f+\beta g)(\vec{x})=\alpha f(\vec{x})+\beta g(\vec{x})$$

satisfies

$$\partial_i(\alpha f+\beta g)=\alpha(\partial_i f)+\beta(\partial_i g).$$

The chain rule concerns composition of C^1 functions. We first prove a special case.

Theorem 21.1 (Chain rule) Let $A\subseteq\mathbb{R}^n$ be open and let $f\in C^1(A,\mathbb{R})$. Let $I\subseteq\mathbb{R}$ be open, and let $\gamma:I\to\mathbb{R}^n$ be a C^1 path such that $\gamma(t)\in A$ for all $t\in I$. Define $u:I\to\mathbb{R}$ by

$$u(t)=f(\gamma(t))\quad(u=f\circ\gamma).$$

Then u is differentiable and

$$u'(t)=\langle(\nabla f)(\gamma(t)),\gamma'(t)\rangle,\quad t\in I.$$

Proof. Figure 12 summarizes the path γ through $\vec{a}=\gamma(t_0)$ and the induced function of one variable $u=f\circ\gamma$.

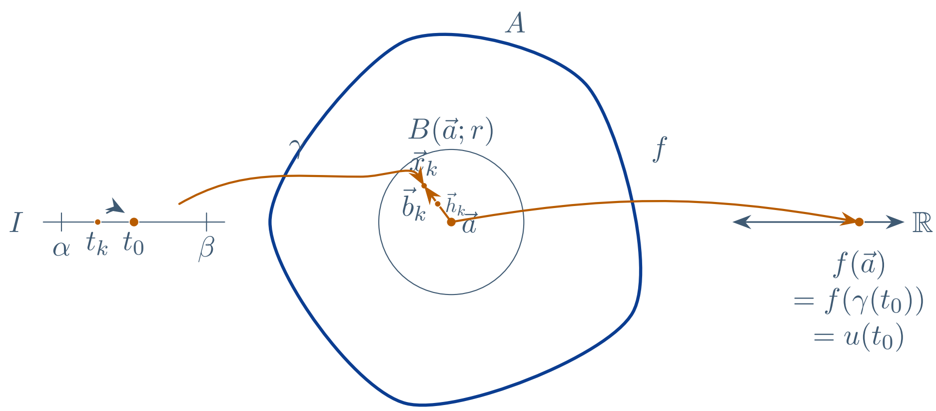

FIGURE 12

Fix $t_0 \in I$. We must show

$$\lim_{t \to t_0} \frac{u(t) - u(t_0)}{t - t_0} = \langle (\nabla f)(\gamma(t_0)), \gamma'(t_0) \rangle.$$

Use a sequential argument: let (t_k) be a sequence in I with $t_k \to t_0$ and $t_k \neq t_0$. We show

$$\lim_{k \to \infty} \frac{u(t_k) - u(t_0)}{t_k - t_0} = \langle (\nabla f)(\gamma(t_0)), \gamma'(t_0) \rangle.$$

Set $\vec{a} \coloneqq \gamma(t_0)$ and $\vec{x}_k \coloneqq \gamma(t_k)$. Since γ is continuous, $\vec{x}_k \to \vec{a}$. Choose $r > 0$ with $B(\vec{a}; r) \subseteq A$. Then there exists k_0 such that $\vec{x}_k \in B(\vec{a}; r)$ for all $k \geq k_0$. For each $k \geq k_0$, let $\vec{v}_k \coloneqq \vec{x}_k - \vec{a}$. Because balls are convex, $L(\vec{a}, \vec{x}_k) \subseteq B(\vec{a}; r) \subseteq A$. If $\vec{v}_k \neq \vec{0}$, the mean value theorem in direction $\vec{v}_k$ supplies $\vec{b}_k \in L(\vec{a}, \vec{x}_k)$ such that

$$f(\vec{x}_k) - f(\vec{a}) = (\partial_{\vec{v}_k} f)(\vec{b}_k) = \langle \vec{v}_k, (\nabla f)(\vec{b}_k) \rangle = \langle \vec{x}_k - \vec{a}, (\nabla f)(\vec{b}_k) \rangle,$$

where we used Corollary 20.1. If $\vec{v}_k = \vec{0}$, take $\vec{b}_k = \vec{a}$; the same identity then reads $0 = 0$. In either case, $\vec{b}_k \in L(\vec{a}, \vec{x}_k)$ implies

$$\left\| \vec{b}_k - \vec{a} \right\| \leq \| \vec{x}_k - \vec{a} \|,$$

so $\vec{b}_k \to \vec{a}$ by the squeeze theorem. Thus we have a sequence $(\vec{b}_k)$ in A with $\vec{b}_k \to \vec{a}$ and

$$(\heartsuit) \qquad f(\vec{x}_k) - f(\vec{a}) = \langle \vec{x}_k - \vec{a}, (\nabla f)(\vec{b}_k) \rangle$$

for all $k \geq k_0$.

Claim 1: for all $k \geq k_0$, we have

$$\frac{u(t_k) - u(t_0)}{t_k - t_0} = \left\langle \frac{\gamma(t_k) - \gamma(t_0)}{t_k - t_0}, (\nabla f)(\vec{b}_k) \right\rangle.$$

Verification:

$$\begin{aligned} \frac{u(t_k) - u(t_0)}{t_k - t_0} &= \frac{f(\gamma(t_k)) - f(\gamma(t_0))}{t_k - t_0} \\ &= \frac{f(\vec{x}_k) - f(\vec{a})}{t_k - t_0} \\ &= \frac{\langle \vec{x}_k - \vec{a}, (\nabla f)(\vec{b}_k)\rangle}{t_k - t_0} \\ &= \left\langle \frac{\gamma(t_k) - \gamma(t_0)}{t_k - t_0}, (\nabla f)(\vec{b}_k) \right\rangle \end{aligned}$$

by ($\heartsuit$).

Claim 2:

$$\lim_{k\to\infty} \frac{\gamma(t_k) - \gamma(t_0)}{t_k - t_0} = \gamma'(t_0).$$

Verification: This is the definition of the velocity vector $\gamma'(t_0)$.

Claim 3: $\lim_{k\to\infty}(\nabla f)(\vec{b}_k) = (\nabla f)(\vec{a})$.

Verification: We know that $\vec{b}_k \to \vec{a}$. For every $1 \leq i \leq n$, it follows that $(\partial_i f)(\vec{b}_k) \to (\partial_i f)(\vec{a})$ since $\partial_i f$ is continuous on A. Thus we obtain n convergent sequences in $\mathbb{R}$, and combining them gives $(\nabla f)(\vec{b}_k) \to (\nabla f)(\vec{a})$.

Claim 4:

$$\lim_{k\to\infty} \left\langle \frac{\gamma(t_k) - \gamma(t_0)}{t_k - t_0}, (\nabla f)(\vec{b}_k) \right\rangle = \langle \gamma'(t_0), (\nabla f)(\vec{a})\rangle.$$

Verification: We have the following easy fact: if $\vec{p}_k \to \vec{p}$ and $\vec{q}_k \to \vec{q}$ in $\mathbb{R}^n$, then

$$\langle \vec{p}_k, \vec{q}_k\rangle \to \langle \vec{p}, \vec{q}\rangle.$$

Taking

$$\vec{p}_k = \frac{\gamma(t_k) - \gamma(t_0)}{t_k - t_0} \qquad \text{and} \qquad \vec{q}_k = (\nabla f)(\vec{b}_k)$$

in the fact and invoking Claims 2 and 3 finishes the proof. □

The chain rule allows us to compute the derivative of a composition $f \circ \gamma$ by taking the dot product of the gradient of the outer function and the velocity of the inner path.

Lecture 22 — Friday, November 01, 2013

Recall Theorem 21.1: if $u : I \to \mathbb{R}$ is defined by

$$u(t) = f(\gamma(t)),$$

then u is differentiable on I, with

$$u'(t) = \langle (\nabla f)(\gamma(t)), \gamma'(t) \rangle.$$

Example 22.1 Compute $(t^t)'$ using the chain rule.

Solution. Let $A = (0, \infty)^2 \subseteq \mathbb{R}^2$, and define $f : A \to \mathbb{R}$ by $f(s,t) = s^t$. Then f is C^1, with

$$(\partial_1 f)(s,t) = ts^{t-1}, \qquad \text{and} \qquad (\partial_2 f)(s,t) = s^t \log(s).$$

Take $I = (0, \infty)$ and define $\gamma : I \to \mathbb{R}^2$ by $\gamma(t) = (t,t)$. The composed function is

$$u(t) = f(\gamma(t)) = f(t,t) = t^t \in (0, \infty).$$

By the chain rule:

$$\begin{aligned} u'(t) &= \langle (\nabla f)(t,t), \gamma'(t) \rangle \\ &= \langle (t \cdot t^{t-1}, t^t \log(t)), (1,1) \rangle \\ &= t \cdot t^{t-1} + t^t \log(t) \\ &= t^t (1 + \log(t)). \end{aligned}$$

□

C^2 functions

Definition 22.2 Let $A \subseteq \mathbb{R}^n$ be open and let $f \in C^1(A, \mathbb{R})$. We say that f is a $\boldsymbol{C^2}$ **function** if

$$\partial_i f \in C^1(A, \mathbb{R}), \quad 1 \le i \le n.$$

We denote

$$C^2(A, \mathbb{R}) = \{f : A \to \mathbb{R} \mid f \text{ is a } C^2 \text{ function}\}.$$

So, for f to be C^2, each $\partial_i f$ must have partial derivatives $\partial_j(\partial_i f)$ that are continuous on A. By convention, we write $\partial_j \partial_i f$ without parentheses. When $i = j$, we write $\partial_i^2 f$.

Remark 22.3 More generally, for every $p \in \mathbb{N}$, define

$$C^p(A, \mathbb{R}) = \{f : A \to \mathbb{R} \mid f \text{ has continuous partial derivatives up to order } p\}.$$

Formally, for every $p \ge 2$,

$$C^p(A, \mathbb{R}) \coloneqq \{f \in C^1(A, \mathbb{R}) \mid \partial_1 f, \ldots, \partial_n f \in C^{p-1}(A, \mathbb{R})\}.$$

If $f \in C^p(A, \mathbb{R})$, then we can consider $\partial_{i_1} \partial_{i_2} \cdots \partial_{i_p} f$ for all $1 \le i_1, \ldots, i_p \le n$. Also,

$$C^1(A, \mathbb{R}) \supseteq C^2(A, \mathbb{R}) \supseteq \cdots \supseteq C^p(A, \mathbb{R}) \supseteq \cdots.$$

We write

$$C^\infty(A, \mathbb{R}) = \bigcap_{p \ge 1} C^p(A, \mathbb{R}).$$

Equivalently, $C^\infty(A, \mathbb{R})$ consists of the functions $f : A \to \mathbb{R}$ with continuous partial derivatives of all orders. Such a function is called **smooth**.

Consider mixed derivatives. For simple examples such as $f(s, t) = u(s)v(t)$, we get

$$(\partial_1 f) = u'(s)v(t) \qquad \text{and} \qquad (\partial_2 \partial_1 f) = u'(s)v'(t),$$

and also

$$(\partial_2 f) = u(s)v'(t) \qquad \text{and} \qquad (\partial_1 \partial_2 f) = u'(s)v'(t).$$

Hence $\partial_1 \partial_2 f = \partial_2 \partial_1 f$ in this case.

For $f(s,t) = s^t$ as in Example 22.1, we have

$$(\partial_2 f)(s,t) = s^t \log(s),$$

and

$$(\partial_1\partial_2 f)(s,t) = ts^{t-1}\log(s) + s^t \cdot \frac{1}{s} = s^{t-1}(t\log(s)+1).$$

Computing $\partial_2\partial_1 f$ gives the same expression.

Theorem 22.4 Let $A \subseteq \mathbb{R}^n$ be open and let $f \in C^2(A,\mathbb{R})$. Then for every $1 \leq i, j \leq n$,

$$\partial_i\partial_j f = \partial_j\partial_i f.$$

The proof uses a carefully chosen auxiliary function together a *smart* application of the mean value theorem. We will simply accept this because we've had enough practice with the mean value theorem already. We must move onto new horizons.

Definition 22.5 Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^2(A,\mathbb{R})$, and let $\vec{a} \in A$. Then

$$(Hf)(\vec{a}) = \begin{bmatrix} \partial_1^2 f(\vec{a}) & \dots & (\partial_1\partial_n f)(\vec{a}) \\ \vdots & \ddots & \vdots \\ (\partial_n\partial_1 f)(\vec{a}) & \dots & (\partial_n^2 f)(\vec{a}) \end{bmatrix}$$

is called the **Hessian matrix** of f at $\vec{a}$.

By Theorem 22.4, the Hessian matrix is symmetric. Therefore all its eigenvalues are real, and it can be diagonalized by an orthogonal matrix.

Lecture 23 — Monday, November 04, 2013

Recall that $A \subseteq \mathbb{R}^n$ open and $f \in C^2(A,\mathbb{R})$ means $\partial_i\partial_j f$ exists and is continuous on A for all $1 \leq i, j \leq n$. For each $\vec{a} \in A$, the Hessian matrix is

$$(Hf)(\vec{a}) = \begin{bmatrix} \partial_1^2 f(\vec{a}) & \dots & (\partial_1\partial_n f)(\vec{a}) \\ \vdots & \ddots & \vdots \\ (\partial_n\partial_1 f)(\vec{a}) & \dots & \partial_n^2 f(\vec{a}) \end{bmatrix}.$$

By Theorem 22.4, this matrix is symmetric.

It is time to review some linear algebra.

If $H \in M_{n\times n}(\mathbb{R})$ is symmetric, then all eigenvalues are real. Writing them as $\lambda_1, \ldots, \lambda_n$ (with multiplicity), there exists an orthogonal matrix V such that

$$H = VDV^T \qquad \text{and} \qquad D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

The matrix is positive definite when all eigenvalues are strictly positive (and similarly for negative definite). If $\lambda_1, \ldots, \lambda_n > 0$, then

$$\langle L_H(\vec{v}), \vec{v}\rangle > 0$$

for all $\vec{v} \in \mathbb{R}^n \setminus \{\vec{0}\}$, where $L_H(\vec{v}) = H\vec{v}$. Indeed, in an orthonormal eigenbasis $\vec{w}_1, \ldots, \vec{w}_n$ with $L_H(\vec{w}_i) = \lambda_i \vec{w}_i$, writing

$$\vec{v} = \sum_i \alpha_i \vec{w}_i$$

gives

$$\langle L_H(\vec{v}), \vec{v}\rangle = \sum_{i=1}^n \alpha_i^2 \lambda_i > 0.$$

Recall from MATH 147 (or high school) that if $f : (\alpha, \beta) \to \mathbb{R}$ is twice differentiable and if $t \in (\alpha, \beta)$ satisfies $f'(t) = 0$ and $f''(t) > 0$, then t is a point of local minimum. If $f''(t) < 0$, then t is a point of local maximum.

Definition 23.1 Let $A \subseteq \mathbb{R}^n$ be open and let $f \in C^1(A, \mathbb{R})$. A point $\vec{a} \in A$ such that $(\nabla f)(\vec{a}) = \vec{0}$ is called a **stationary point** of f.

By Proposition 15.5, if $\vec{a}$ is a point of local extremum, then $\vec{a}$ is a stationary point.

Theorem 23.2 Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^2(A, \mathbb{R})$, and let $\vec{a} \in A$ be stationary. Consider $(Hf)(\vec{a})$.

1. If $(Hf)(\vec{a})$ is positive definite, then $\vec{a}$ is a point of local minimum for f.
2. If $(Hf)(\vec{a})$ is negative definite, then $\vec{a}$ is a point of local maximum for f.

If the Hessian is not definite, several outcomes are possible. For instance, if there exist i, j with $\lambda_i < 0 < \lambda_j$, then $\vec{a}$ is a **saddle point** for f, so it is surely neither a local minimum nor a local maximum. If some eigenvalues are 0, the second derivative test must be handled case by case; for example, if $(Hf)(\vec{a}) = 0_{n\times n}$, the test fails.

The proof of Theorem 23.2 is deferred to the next lecture. The following notation is an ingredient:

Definition 23.3 Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^2(A,\mathbb{R})$, and let $\vec{v} = (v^{(1)},\ldots,v^{(n)}) \in \mathbb{R}^n$. Consider

$$\partial_{\vec{v}} f : A \to \mathbb{R}.$$

Since

$$(\partial_{\vec{v}} f)(\vec{x}) = \langle \vec{v}, (\nabla f)(\vec{x})\rangle = v^{(1)}(\partial_1 f)(\vec{x}) + \cdots + v^{(n)}(\partial_n f)(\vec{x}),$$

it is a linear combination of $\partial_1 f,\ldots,\partial_n f \in C^1(A,\mathbb{R})$, hence $\partial_{\vec{v}} f \in C^1(A,\mathbb{R})$. Therefore we can define

$$\partial_{\vec{v}}(\partial_{\vec{v}} f) : A \to \mathbb{R},$$

denoted $\partial_{\vec{v}}^2 f$.

Lecture 24 — Wednesday, November 06, 2013

Recall that $A \subseteq \mathbb{R}^n$ is open and $f \in C^2(A,\mathbb{R})$, that is, $\partial_i\partial_j f$ exists and is continuous on A for all $1 \le i,j \le n$. Given $\vec{v} \in \mathbb{R}^n$, define

$$g = \partial_{\vec{v}} f : A \to \mathbb{R}.$$

Then $g \in C^1(A,\mathbb{R})$, so we may again differentiate in direction $\vec{v}$:

$$\partial_{\vec{v}} g = \partial_{\vec{v}}(\partial_{\vec{v}} f) = \partial_{\vec{v}}^2 f.$$

Lemma 24.1 Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^2(A,\mathbb{R})$, and let $\vec{v} \in \mathbb{R}^n$. Then we have

$$(\partial_{\vec{v}}^2 f)(\vec{x}) = \langle (Hf)(\vec{x}) \cdot \vec{v}, \vec{v}\rangle,$$

for all $\vec{x} \in A$.

Proof. Denote $g \coloneqq \partial_{\vec{v}} f : A \to \mathbb{R}$. For every $\vec{x} \in A$,

$$g(\vec{x}) = (\partial_{\vec{v}} f)(\vec{x}) = \langle \vec{v}, (\nabla f)(\vec{x})\rangle = v^{(1)}\partial_1 f(\vec{x}) + \cdots + v^{(n)}\partial_n f(\vec{x}),$$

by Corollary 20.1. Fix $i \in \{1, \dots, n\}$ and apply ∂_i:

$$\partial_i g(\vec{x}) = v^{(1)}(\partial_i \partial_1 f)(\vec{x}) + \cdots + v^{(n)}(\partial_i \partial_n f)(\vec{x}) = \langle \vec{v}, \text{row } i \text{ of } (Hf)(\vec{x}) \rangle.$$

Collecting these identities for $i = 1, \dots, n$ gives

$$(\nabla g)(\vec{x}) = (Hf)(\vec{x}) \cdot \vec{v},$$

which could also be denoted by $L_{(Hf)(\vec{x})}(\vec{v})$. Therefore

$$\partial_{\vec{v}}^2 f(\vec{x}) = \partial_{\vec{v}} g(\vec{x}) = \langle (\nabla g)(\vec{x}), \vec{v} \rangle = \langle (Hf)(\vec{x}) \cdot \vec{v}, \vec{v} \rangle.$$

□

Lemma 24.2 Let $A \subseteq \mathbb{R}^n$ be open and let $f \in C^2(A, \mathbb{R})$. Let $\vec{a}$ be a stationary point for f. Suppose $\vec{x} \in A$ and $L(\vec{a}, \vec{x}) \subseteq A$. Set $\vec{v} := \vec{x} - \vec{a}$. Then there exist $\beta \in [0, 1]$ and $\vec{z} \in L(\vec{a}, \vec{x})$ such that $f(\vec{x}) - f(\vec{a}) = \beta \cdot (\partial_{\vec{v}}^2 f)(\vec{z})$.

This is a mean-value-theorem-type statement.

Proof. If $\vec{x} = \vec{a}$, take $\beta = 0$ and $\vec{z} = \vec{a}$. Assume henceforth that $\vec{x} \neq \vec{a}$, so $\vec{v} \neq \vec{0}$. Set $g := \partial_{\vec{v}} f$. Then $g \in C^1(A, \mathbb{R})$ and $\partial_{\vec{v}} g = \partial_{\vec{v}}^2 f$. Also

$$g(\vec{a}) = (\partial_{\vec{v}} f)(\vec{a}) = \langle \vec{v}, (\nabla f)(\vec{a}) \rangle = 0$$

because $\vec{a}$ is stationary. Apply the mean value theorem in direction $\vec{v}$ to f: there exists $\vec{b} \in L(\vec{a}, \vec{x})$ such that

$$f(\vec{x}) - f(\vec{a}) = (\partial_{\vec{v}} f)(\vec{b}) = g(\vec{b}) = g(\vec{b}) - g(\vec{a}).$$

The one-variable mean value theorem used in the proof of Theorem 17.2 places $\vec{b}$ in the interior of the segment, so

$$\vec{b} - \vec{a} = \beta \vec{v} \quad \text{for some } \beta \in (0, 1).$$

Apply Theorem 17.2 to g on the segment from $\vec{a}$ to $\vec{b}$. Since its displacement is $\beta \vec{v}$, there exists $\vec{z} \in L(\vec{a}, \vec{b}) \subseteq L(\vec{a}, \vec{x})$ such that

$$g(\vec{b}) - g(\vec{a}) = \beta (\partial_{\vec{v}} g)(\vec{z}) = \beta (\partial_{\vec{v}}^2 f)(\vec{z}).$$

Combining the two mean-value identities proves the formula. □

Lemma 24.3 Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^2(A, \mathbb{R})$, and let $\vec{a} \in A$. Suppose that $(Hf)(\vec{a})$ is positive definite. Then there exists $r > 0$ such that $B(\vec{a}; r) \subseteq A$ and such that $(Hf)(\vec{x})$ is still positive definite for every $\vec{x} \in B(\vec{a}; r)$.

We simply accept this fact. The idea is that $(Hf)(\vec{x})$ depends continuously on $\vec{x}$. If matrix entries vary continuously with respect to $\vec{x}$, then so too do the eigenvalues of the matrix. Hence, if all eigenvalues are positive at $\vec{x} = \vec{a}$, they remain positive on a sufficiently small ball around $\vec{a}$. For the passage from entries to positive definiteness, we may use Sylvester's criterion.

Theorem 24.4 Let $A \subseteq \mathbb{R}^n$ be open, let $f \in C^2(A, \mathbb{R})$, and suppose $\vec{a} \in A$ is stationary for f. If $(Hf)(\vec{a})$ is positive definite, then $\vec{a}$ is a local minimum for f. If $(Hf)(\vec{a})$ is negative definite, then $\vec{a}$ is a local maximum for f.

Proof. The two cases are analogous, so we prove the positive-definite case.

Assume $(Hf)(\vec{a})$ is positive definite. By Lemma 24.3, there exists $r > 0$ such that $B(\vec{a}; r) \subseteq A$ and $(Hf)(\vec{x})$ is positive definite for every $\vec{x} \in B(\vec{a}; r)$. Claim: $f(\vec{x}) \geq f(\vec{a})$ for all $\vec{x} \in B(\vec{a}; r)$. Take $\vec{x} \in B(\vec{a}; r)$, and set $\vec{v} := \vec{x} - \vec{a}$. If $\vec{x} = \vec{a}$ the inequality is immediate, so assume $\vec{x} \neq \vec{a}$ and hence $\vec{v} \neq \vec{0}$. By Lemma 24.2, there exist $\beta \in [0, 1]$ and $\vec{z} \in L(\vec{a}, \vec{x})$ such that

$$f(\vec{x}) - f(\vec{a}) = \beta \cdot (\partial_{\vec{v}}^2 f)(\vec{z}).$$

Using Lemma 24.1,

$$f(\vec{x}) - f(\vec{a}) = \beta \cdot \langle (Hf)(\vec{z}) \cdot \vec{v}, \vec{v} \rangle.$$

Now $\beta \geq 0$, and $\vec{z} \in B(\vec{a}; r)$ implies $(Hf)(\vec{z})$ is positive definite, so $\langle (Hf)(\vec{z}) \cdot \vec{v}, \vec{v} \rangle \geq 0$. Therefore $f(\vec{x}) \geq f(\vec{a})$. $\square$

The intermediate point $\vec{z}$ and the relation $\vec{z} - \vec{a} = \beta(\vec{x} - \vec{a})$ are shown in Figure 13.

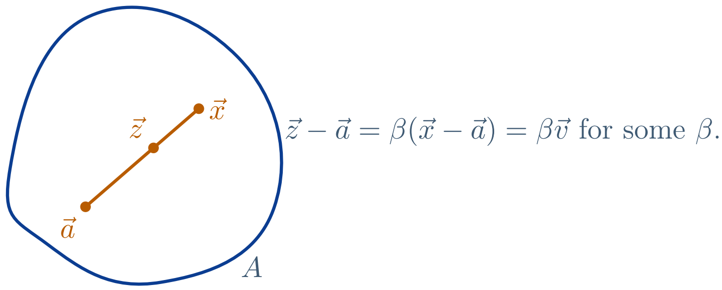

FIGURE 13

Lecture 25 — Friday, November 08, 2013

4. Riemann integration in several variables

Integrable Functions

Definition 25.1 We will use the name **half-open rectangle in $\mathbb{R}^n$** in reference to sets of the form $P=(a_1,b_1]\times\cdots\times(a_n,b_n]\subseteq\mathbb{R}^n$, with $a_1<b_1,\ldots,a_n<b_n$ in $\mathbb{R}$. The collection of all half-open rectangles in $\mathbb{R}^n$ will be denoted $\mathscr{P}_n$. We define a **volume function** $\mathrm{vol}_o:\mathscr{P}_n\to[0,\infty)$, where for $P=(a_1,b_1]\times\cdots\times(a_n,b_n]$ we put $\mathrm{vol}_o(P):=(b_1-a_1)\cdots(b_n-a_n)$.

Definition 25.2 A set $S\subseteq\mathbb{R}^n$ is said to be a **set of content zero** when it has the following "(ε-Cover)" property:

$$(\varepsilon\text{-Cover})\qquad \begin{cases}\text{For every } \varepsilon>0 \text{ we can find } P_1,\ldots,P_\ell\in\mathscr{P}_n \text{ (for some } \ell\in\mathbb{N})\\ \text{such that } P_1\cup\cdots\cup P_\ell\supseteq S \text{ and } \sum_{i=1}^{\ell}\mathrm{vol}_o(P_i)<\varepsilon,\end{cases}$$

Definition 25.3 We say that a set $A\subseteq\mathbb{R}^n$ is **Jordan measurable** when it satisfies the following two conditions:

1. A is bounded;

2. the boundary $\mathrm{bd}(A)$ has content zero.

The collection of all Jordan measurable subsets of $\mathbb{R}^n$ is denoted by $\mathscr{M}_n$.

Definition 25.4 (and Proposition) Let $A \subseteq \mathbb{R}^n$ be a Jordan measurable set. Consider the real nonnegative numbers $V^-(A)$ and $V^+(A)$ defined as follows:

$$V^-(A) := \sup\Big\{ \sum_{i=1}^{r} \mathrm{vol}_o(P_i) \;\Big|\; \begin{array}{c} P_1, \dots, P_r \in \mathscr{P}_n, \text{ with } \cup_{i=1}^r P_i \subseteq A \\ \text{and with } P_i \cap P_j = \emptyset \text{ for } i \neq j \end{array} \Big\}.$$

$$V^+(A) := \inf\Big\{ \sum_{j=1}^{q} \mathrm{vol}_o(Q_j) \mid Q_1, \dots, Q_q \in \mathscr{P}_n, \text{ with } \cup_{j=1}^q Q_j \supseteq A \Big\}.$$

Then $V^-(A) = V^+(A)$. The **volume** of A is defined as

$$\mathrm{vol}(A) := V^-(A) = V^+(A).$$

If no half-open rectangle $P \in \mathscr{P}_n$ is contained in A, we set $V^-(A) := 0$ by convention. The proposition in Definition 25.4 then gives $V^+(A) = 0$ as well, so $\mathrm{vol}(A) = 0$.

Definition 25.5 Let $\mathscr{A}$ be a set of subsets of $\mathbb{R}^n$. We say that $\mathscr{A}$ is a **field of subsets** of $\mathbb{R}^n$ when it has the following properties:

(FS-1) The empty set $\emptyset$ belongs to $\mathscr{A}$.

(FS-2) If $A, B \in \mathscr{A}$, then the set difference $A \setminus B$ still belongs to $\mathscr{A}$.

(FS-3) If $A_1, \dots, A_k \in \mathscr{A}$, then the union $A_1 \cup \dots \cup A_k$ and the intersection $A_1 \cap \dots \cap A_k$ still belong to $\mathscr{A}$.

Definition 25.6 Let $\mathscr{A}$ be a field of subsets of $\mathbb{R}^n$. A function $\mu : \mathscr{A} \to [0, \infty)$ is said to be **additive** when it has the property that:

$$\left(\begin{array}{c} A_1, \dots, A_k \in \mathscr{A} \\ \text{with } A_i \cap A_j = \emptyset \text{ for } i \neq j \end{array} \right) \implies \left(\mu(A_1 \cup \dots \cup A_k) = \sum_{i=1}^{k} \mu(A_i) \right).$$

Proposition 25.7 Consider the collection $\mathscr{M}_n$ of Jordan measurable subsets of $\mathbb{R}^n$. The following hold:

1. $\mathscr{M}_n \supseteq \mathscr{P}_n$.
2. $\mathscr{M}_n$ is a field of subsets of $\mathbb{R}^n$.

Proposition 25.8 The volume function $\text{vol} : \mathscr{M}_n \to [0, \infty)$ satisfies the following:

1. $\text{vol}(P) = \text{vol}_o(P)$, for every $P \in \mathscr{P}_n$. (That is: the function $\text{vol} : \mathscr{M}_n \to [0, \infty)$ is an extension of the function $\text{vol}_o : \mathscr{P}_n \to [0, \infty)$.)
2. vol is an additive function.
3. Suppose $\mu : \mathscr{M}_n \to [0, \infty)$ is an additive function, such that $\mu(P) = \text{vol}_o(P)$ for every $P \in \mathscr{P}_n$. Then it follows that $\mu(A) = \text{vol}(A)$ for every $A \in \mathscr{M}_n$.

Proposition 25.9 Consider the collection $\mathscr{M}_n$ of Jordan measurable subsets of $\mathbb{R}^n$. The following hold:

1. If A is a subset of $\mathbb{R}^n$ with content zero, then it follows that $A \in \mathscr{M}_n$ and that $\text{vol}(A) = 0$.
2. Conversely, if $A \in \mathscr{M}_n$ and if $\text{vol}(A) = 0$, then it follows that A is a subset of $\mathbb{R}^n$ with content zero.

The notions of Jordan measure and sets of content zero are predecessors of Lebesgue measure and null sets. The latter strictly generalize the former: a bounded set is Jordan measurable if and only if its boundary has Lebesgue measure zero, so every Jordan measurable set is Lebesgue measurable, but not conversely. Jordan theory is based on finite coverings (i.e., finite unions of rectangles), whereas Lebesgue theory allows countable coverings; for example, $\mathbb{Q} \cap [0, 1]$ is Lebesgue null but not Jordan measurable since its boundary is all of $[0, 1]$, and it is not a set of content zero because any finite cover must have total length at least 1. Thus Jordan theory is not stable under countable operations and breaks down in the presence of limits.

Today, we will see how to construct a concept of the Riemann integral based only on Propositions 25.7 to 25.9.

Definition 25.10 Let A be a nonempty set in $\mathscr{M}_n$. By a **division** of A, we understand a set $\Delta = \{A_1, \ldots, A_p\}$ where $A_1, \ldots, A_p$ are nonempty, disjoint Jordan measurable sets whose union is A.

Definition 25.11 Let $A \in \mathscr{M}_n$ be nonempty, and let $\Delta = \{A_1, \ldots, A_p\}$ and $\Gamma = \{B_1, \ldots, B_q\}$ be two divisions of A. We say that Γ **refines** Δ, denoted by "$\Gamma \prec \Delta$", when for every $1 \leq j \leq q$, there exists $1 \leq i \leq p$ such that $B_j \subseteq A_i$.

In other words, refinement breaks the sets in Δ into smaller disjoint pieces. The condition in Definition 25.11 forces every set from Δ to be a disjoint union of sets from Γ. That is, we can rewrite

$$\Delta = \{A_1, \ldots, A_p\}$$

and

$$\Gamma = \{B_{1,1}, \ldots, B_{1,q_1}, \ldots, B_{p,1}, \ldots, B_{p,q_p}\}$$

where $\{B_{i,1}, \ldots, B_{i,q_i}\}$ is a division of A_i. Figure 14 depicts such a refinement.

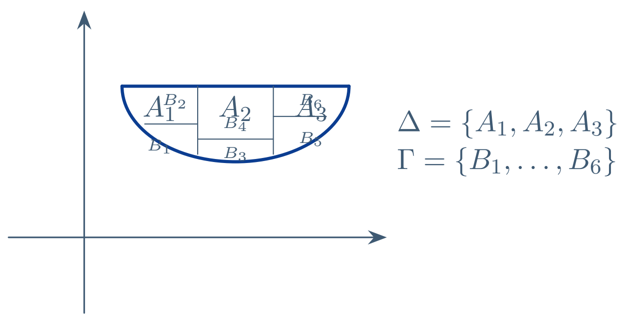

Figure 14

Proposition 25.12 Let $A \in \mathscr{M}_n$ be nonempty. If Δ' and Δ'' are two divisions of A, then there exists a division Γ of A such that $\Gamma \prec \Delta'$ and $\Gamma \prec \Delta''$.

Proof. Write $\Delta' = \{A'_1, \ldots, A'_r\}$ and $\Delta'' = \{A''_1, \ldots, A''_s\}$. Put

$$\Gamma = \{A'_i \cap A''_j \mid 1 \leq i \leq r, 1 \leq j \leq s, A'_i \cap A''_j \neq \emptyset\}.$$

Note that every $A'_i \cap A''_j \in \mathscr{M}_n$ by one of the field axioms. So Γ consists of some number q of sets from $\mathscr{M}_n$, where $q \leq rs$.

Claim 1: Γ is a division of A.

Verification: This follows from the field properties and immediate Boolean algebra.

Claim 2: $\Gamma \prec \Delta'$ and $\Gamma \prec \Delta''$.

Verification: For every set $A'_i \cap A''_j$ in Γ, we have $A'_i \cap A''_j \subseteq A'_i$, where $A'_i \in \Delta'$, so $\Gamma \prec \Delta'$. Likewise, $A'_i \cap A''_j \subseteq A''_j$, so $\Gamma \prec \Delta''$. $\square$

Definition 25.13 Consider a function $f : A \to \mathbb{R}$ with $A \in \mathscr{M}_n$. We assume f is bounded, meaning there exists $c \geq 0$ such that

$$-c \leq f(\vec{x}) \leq c \quad \text{for all } \vec{x} \in A.$$

For $\emptyset \neq B \subseteq A$, define the **supremum and infimum of f over B**,

$$\sup_B(f) := \sup\{f(\vec{x}) \mid \vec{x} \in B\}, \qquad \text{and} \qquad \inf_B(f) := \inf\{f(\vec{x}) \mid \vec{x} \in B\},$$

and the **oscillation of f over B**,

$$\operatorname{osc}_B(f) = \sup_B(f) - \inf_B(f) = \sup\{|f(\vec{x}) - f(\vec{y})| \mid \vec{x}, \vec{y} \in B\}.$$

Definition 25.14 Let $f \colon A \to \mathbb{R}$ be a bounded function, where $A \in \mathscr{M}_n$ is nonempty. Let $\Delta = \{A_1, \ldots, A_p\}$ be a division of A. Then

$$U(f, \Delta) = \sum_{i=1}^{p} \operatorname{vol}(A_i) \cdot \sup_{A_i}(f)$$

is called the **upper Darboux sum** for f and Δ and

$$L(f, \Delta) = \sum_{i=1}^{p} \operatorname{vol}(A_i) \cdot \inf_{A_i}(f)$$

is called the **lower Darboux sum** for f and Δ.

Remark 25.15 Observe that

$$U(f, \Delta) - L(f, \Delta) = \sum_{i=1}^{p} \operatorname{vol}(A_i) \cdot \operatorname{osc}_{A_i}(f).$$

$U(f,\Delta)$ and $L(f,\Delta)$ have interpretations as volumes in $\mathbb{R}^{n+1}$, which we will return to in the next lecture.

Lecture 26 — Monday, November 11, 2013

Lemma 26.1 Let $f\colon A \to \mathbb{R}$ be a bounded function, where $A \in \mathscr{M}_n$ is nonempty. Let Δ and Γ be two divisions of A such that $\Gamma \prec \Delta$. Then $L(f,\Delta) \le L(f,\Gamma)$ and $U(f,\Gamma) \le U(f,\Delta)$.

Proof. Let $\Delta = \{A_1, \ldots, A_p\}$. Since $\Gamma \prec \Delta$, we can write

$$\Gamma = \{B_{i,j} \mid 1 \le i \le p, 1 \le j \le q_i\}$$

where $\{B_{i,1}, \ldots, B_{i,q_i}\}$ is a division of A_i. Then:

$$\begin{aligned} U(f,\Delta) &= \sum_{i=1}^{p} \operatorname{vol}(A_i) \sup_{A_i}(f) \\ &= \sum_{i=1}^{p} \left(\sum_{j=1}^{q_i} \operatorname{vol}(B_{i,j}) \right) \sup_{A_i}(f) \\ &= \sum_{i=1}^{p} \sum_{j=1}^{q_i} \operatorname{vol}(B_{i,j}) \sup_{A_i}(f) \\ &\ge \sum_{i=1}^{p} \sum_{j=1}^{q_i} \operatorname{vol}(B_{i,j}) \sup_{B_{i,j}}(f) = U(f,\Gamma) \end{aligned}$$

since $B_{i,j} \subseteq A_i$ implies that $\sup_{B_{i,j}}(f) \le \sup_{A_i}(f)$. The proof for $L(f,\Delta) \le L(f,\Gamma)$ is analogous. $\square$

Proposition 26.2 Let $f\colon A \to \mathbb{R}$ be bounded, where $A \in \mathscr{M}_n$ is nonempty. Let Δ_1, Δ_2 be any two divisions of A. Then $L(f,\Delta_1) \le U(f,\Delta_2)$.

Proof. By Proposition 25.12, there exists a common refinement Γ such that $\Gamma \prec \Delta_1$ and $\Gamma \prec \Delta_2$.

Then by Lemma 26.1 applied twice,

$$L(f,\Delta_1) \le L(f,\Gamma) \le U(f,\Gamma) \le U(f,\Delta_2)$$

where the middle inequality $L(f,\Gamma) \le U(f,\Gamma)$ follows directly from Definition 25.14, since $\inf_B(f) \le \sup_B(f)$ for any set B. $\square$

Definition 26.3 Let $f\colon A \to \mathbb{R}$ be bounded, where $A \in \mathscr{M}_n$ is nonempty.

1) The **lower integral** of f over A is

$$\underline{\int_A} f = \underline{\int_A} f(\vec{x})\,\mathrm{d}\vec{x} := \sup\{L(f,\Delta) \mid \Delta \text{ is a division of } A\}.$$

2) The **upper integral** of f over A is

$$\overline{\int_A} f = \overline{\int_A} f(\vec{x})\,\mathrm{d}\vec{x} := \inf\{U(f,\Delta) \mid \Delta \text{ is a division of } A\}.$$

3) If it so happens that

$$\underline{\int_A} f = \overline{\int_A} f,$$

we say that f is **integrable** on A and that its **integral** is

$$\int_A f := \underline{\int_A} f = \overline{\int_A} f.$$

Theorem 26.4 Let $f\colon A \to \mathbb{R}$ be bounded, where $A \in \mathscr{M}_n$ is nonempty.

1. The lower and upper integrals of f over A are finite numbers, and

$$\underline{\int_A} f \le \overline{\int_A} f$$

regardless of whether or not f is integrable.

2. The following statements are equivalent:

   (I) f is integrable on A.

(I') For every $\varepsilon > 0$, there exists a division Δ of A such that

$$U(f,\Delta) - L(f,\Delta) < \varepsilon.$$

(I'') There exists a sequence (Δ_k) of divisions of A such that

$$U(f,\Delta_k) - L(f,\Delta_k) \to 0.$$

The proof is deferred. We will see that the proof only depends on the sets

$$S = \{s \in \mathbb{R} : \text{ there exists a division } \Delta' \text{ of } A \text{ such that } L(f,\Delta') = s\}$$

and

$$T = \{t \in \mathbb{R} : \text{ there exists a division } \Delta'' \text{ of } A \text{ such that } U(f,\Delta'') = t\}.$$

Then

$$\underline{\int_A} f = \sup(S) \qquad \text{and} \qquad \overline{\int_A} f = \inf(T).$$

Proposition 26.2 says that $s \le t$ for all $s \in S$ and $t \in T$.

Lecture 27 — Monday, November 11, 2013

Figure 15 shows the upper and lower "skyscraper" heights over a piece A_i when $n = 2$.

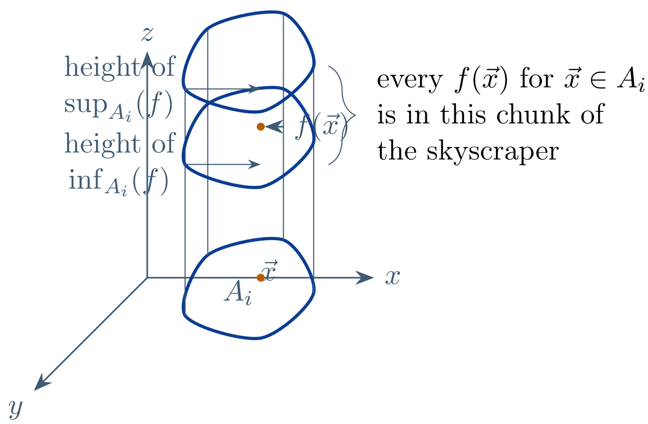

FIGURE 15

Proof of Theorem 26.4. Let $A \in \mathcal{M}_n$ be nonempty, and let $f : A \to \mathbb{R}$ be bounded. Consider the sets

$$S = \{s \in \mathbb{R} \mid \text{there exists a division } \Delta' \text{ of } A \text{ such that } L(f, \Delta') = s\},$$

and

$$T = \{t \in \mathbb{R} \mid \text{there exists a division } \Delta'' \text{ of } A \text{ such that } U(f, \Delta'') = t\}.$$

Then

$$\underline{\int_A} f = \sup(S), \qquad \text{and} \qquad \overline{\int_A} f = \inf(T).$$

Choose $c \geq 0$ such that $|f| \leq c$. The one-piece division $\{A\}$ shows that S and T are nonempty, and every Darboux sum lies between $-c \operatorname{vol}(A)$ and $c \operatorname{vol}(A)$. Thus $\sup(S)$ and $\inf(T)$ are finite. It remains to show that

$$\underline{\int_A} f \leq \overline{\int_A} f.$$

By Proposition 26.2, whenever $s \in S$ and $t \in T$, we have $s \leq t$. Fixing $t \in T$ shows that every $s \in S$ satisfies $s \leq t$, so S is bounded above and $\sup(S) \leq t$. Since this holds for every $t \in T$, the set T is bounded below and

$$\inf(T) \geq \sup(S).$$

Hence $\overline{\int_A} f \geq \underline{\int_A} f$, which proves part (1).

Part (2) says that conditions (I), (I'), and (I'') are equivalent. We prove that (I) implies (I'), that (I') implies (I''), and that (I'') implies (I).

(I) implies (I'). Assume (I). We have $\inf(T) = \sup(S)$. Call this common value λ; then

$$\lambda = \int_A f.$$

Given $\varepsilon > 0$, by definition of supremum we can choose $s \in S$ such that

$$\lambda - \frac{\varepsilon}{2} < s \leq \lambda.$$

Write $s = L(f, \Delta')$ for some division Δ' of A. Similarly, choose $t \in T$ with

$$\lambda \leq t < \lambda + \frac{\varepsilon}{2},$$

where $t = U(f, \Delta'')$ for some division Δ'' of A. Let Δ refine both Δ' and Δ'' (Proposition 25.12).

By Lemma 26.1,

$$\lambda - \frac{\varepsilon}{2} < s = L(f, \Delta') \leq L(f, \Delta) \leq U(f, \Delta) \leq U(f, \Delta'') = t < \lambda + \frac{\varepsilon}{2}.$$

Hence

$$U(f, \Delta) - L(f, \Delta) < \left(\lambda + \frac{\varepsilon}{2}\right) - \left(\lambda - \frac{\varepsilon}{2}\right) = \varepsilon.$$

Thus (I) implies (I').

(I') implies (I'') Apply (I') with $\varepsilon = 1/k$ for each $k \in \mathbb{N}$. This gives a division Δ_k of A such that

$$0 \leq U(f, \Delta_k) - L(f, \Delta_k) < \frac{1}{k} \to 0.$$

By the squeeze theorem, $U(f, \Delta_k) - L(f, \Delta_k) \to 0$. Thus (I') implies (I'').

(I'') implies (I). Assume there exists (Δ_k) as in (I''). For every k,

$$L(f, \Delta_k) \leq \underline{\int_A} f \leq \overline{\int_A} f \leq U(f, \Delta_k),$$

because $L(f, \Delta_k) \in S$ and $U(f, \Delta_k) \in T$. Therefore

$$0 \leq \overline{\int_A} f - \underline{\int_A} f \leq U(f, \Delta_k) - L(f, \Delta_k) \to 0.$$

By the squeeze theorem,

$$\overline{\int_A} f - \underline{\int_A} f = 0,$$

so

$$\overline{\int_A} f = \underline{\int_A} f.$$

□

Proposition 27.1 $A \in \mathcal{M}_n$, $f : A \to \mathbb{R}$ bounded. Suppose f is integrable, and let (Δ_k) be as in (I'') of Theorem 26.4. Then

$$\lim_{k\to\infty} U(f, \Delta_k) = \lim_{k\to\infty} L(f, \Delta_k) = \int_A f.$$

Proof. For every $k \in \mathbb{N}$, we have

$$L(f, \Delta_k) \leq \underline{\int_A} f = \int_A f = \overline{\int_A} f \leq U(f, \Delta_k),$$

so

$$0 \leq \int_A f - L(f, \Delta_k) \leq U(f, \Delta_k) - L(f, \Delta_k) \to 0.$$

Hence

$$L(f, \Delta_k) \to \int_A f.$$

The convergence

$$U(f, \Delta_k) \to \int_A f$$

follows analogously. ☐

Lecture 28 — Friday, November 15, 2013

Operations with Integrable Functions

Definition 28.1 Let $A \in \mathcal{M}_n$ with $A \neq \emptyset$. Denote

$$\mathrm{Int}_b(A, \mathbb{R}) := \{f : A \to \mathbb{R} \mid f \text{ is bounded and integrable}\}.$$

Theorem 28.2 Fix $A \in \mathcal{M}_n$, $A \neq \emptyset$. Then the following hold:

1. Every constant function is in $\mathrm{Int}_b(A, \mathbb{R})$.
2. If $f, g \in \mathrm{Int}_b(A, \mathbb{R})$, then $f + g \in \mathrm{Int}_b(A, \mathbb{R})$, with

$$\int_A (f + g) = \int_A f + \int_A g.$$

3. If $f \in \mathrm{Int}_b(A, \mathbb{R})$ and $\alpha \in \mathbb{R}$, then $\alpha f \in \mathrm{Int}_b(A, \mathbb{R})$, with
$$\int_A (\alpha f) = \alpha \int_A f.$$
4. If $f, g \in \mathrm{Int}_b(A, \mathbb{R})$, then $f \cdot g \in \mathrm{Int}_b(A, \mathbb{R})$ (however, there is no general formula for $\int_A (fg)$ in terms of $\int_A f$ and $\int_A g$).

A set of functions with properties (1)–(4) is an **algebra of functions**. (2) and (3) give a **vector space**. (1), (2), and (4) give a **ring**.

We showed (1) last time: if $f(\vec{x}) = c$ for all $\vec{x} \in A$, then
$$\int_A f = c \cdot \mathrm{vol}(A).$$
(2) and (3) follow from a general statement about upper and lower integrals:

Proposition 28.3 $A \in \mathcal{M}_n$, $A \neq \emptyset$. Let $f, g : A \to \mathbb{R}$ bounded but not necessarily integrable. Let $\alpha \in \mathbb{R}$. Then
$$\overline{\int_A} (f+g) \le \overline{\int_A} f + \overline{\int_A} g \quad \text{and} \quad \underline{\int_A} (f+g) \ge \underline{\int_A} f + \underline{\int_A} g, \tag{28.1}$$
and
$$\begin{cases} \overline{\int_A} (\alpha f) = \alpha \overline{\int_A} f, \ \underline{\int_A} (\alpha f) = \alpha \underline{\int_A} f, & \alpha \ge 0, \\ \overline{\int_A} (\alpha f) = \alpha \underline{\int_A} f, \ \underline{\int_A} (\alpha f) = \alpha \overline{\int_A} f, & \alpha < 0. \end{cases} \tag{28.2}$$

How do (2) and (3) follow from (28.1) and (28.2)? For (2), if $f, g \in \mathrm{Int}_b(A, \mathbb{R})$, then
$$\overline{\int_A} (f+g) \le \overline{\int_A} f + \overline{\int_A} g = \int_A f + \int_A g = \underline{\int_A} f + \underline{\int_A} g \le \underline{\int_A} (f+g).$$
Hence equality holds throughout, so $f+g$ is integrable and
$$\int_A (f+g) = \int_A f + \int_A g.$$

Proof of Proposition 28.3. We prove (28.1) for upper integrals: if $A \in \mathcal{M}_n$, $A \neq \emptyset$, and f, g are bounded, then

$$\overline{\int_A}(f+g) \leq \overline{\int_A} f + \overline{\int_A} g.$$

Claim 1: for every nonempty $B \subseteq A$, we have $\sup_B(f+g) \leq \sup_B(f) + \sup_B(g)$.

Verification: denote $\sup_B(f) =: \sigma$ and $\sup_B(g) =: \tau$. Then for any $\vec{x} \in B$,

$$(f+g)(\vec{x}) = f(\vec{x}) + g(\vec{x}) \leq \sigma + \tau.$$

Thus $\sigma + \tau$ is an upper bound for $\{(f+g)(\vec{x}) \mid \vec{x} \in B\}$, so $\sup_B(f+g) \leq \sigma + \tau$.

Claim 2: for every division $\Delta = \{A_1, \ldots, A_r\}$ of A, we have $U(f+g, \Delta) \leq U(f, \Delta) + U(g, \Delta)$.

Verification:

$$\begin{aligned}
U(f+g,\Delta) &= \sum_{i=1}^{r} \operatorname{vol}(A_i) \cdot \sup_{A_i}(f+g) \leq \sum_{i=1}^{r} \operatorname{vol}(A_i) \cdot (\sup_{A_i}(f) + \sup_{A_i}(g)) \\
&= \sum_{i=1}^{r} \operatorname{vol}(A_i) \cdot \sup_{A_i}(f) + \sum_{i=1}^{r} \operatorname{vol}(A_i) \cdot \sup_{A_i}(g) \\
&= U(f,\Delta) + U(g,\Delta).
\end{aligned}$$

Claim 3:

$$\overline{\int_A}(f+g) \leq \overline{\int_A} f + \overline{\int_A} g.$$

Verification: denote $\overline{\int_A} f =: \alpha$ and $\overline{\int_A} g =: \beta$, and let $k \in \mathbb{N}$. By the definition of the upper integral as an infimum, we can find a division Δ_k' of A such that $\alpha \leq U(f, \Delta_k') < \alpha + \frac{1}{k}$. Likewise, we can find a division Δ_k'' of A such that $\beta \leq U(g, \Delta_k'') < \beta + \frac{1}{k}$. For each k, let Δ_k be a common refinement of Δ_k' and Δ_k''. Then

$$\overline{\int_A}(f+g) \leq U(f+g, \Delta_k) \leq U(f, \Delta_k) + U(g, \Delta_k) \leq U(f, \Delta_k') + U(g, \Delta_k'') < \alpha + \beta + \frac{2}{k}.$$

Let $k \to \infty$. We obtain

$$\overline{\int_A}(f+g) \leq \alpha + \beta = \overline{\int_A} f + \overline{\int_A} g.$$

The inequality for lower integrals follows by the same argument with infima in place of suprema. The formulas for αf follow directly from the corresponding formulas for suprema and infima; when $\alpha < 0$, multiplication reverses the order and swaps them. $\square$

Finally, we do a bit of work on (4). Suppose we have only the following weaker statement: "If $A \in \mathcal{M}_n$, $A \neq \emptyset$, and $f \in \mathrm{Int}_b(A, \mathbb{R})$, then $f^2 \in \mathrm{Int}_b(A, \mathbb{R})$." If we know this, can we derive

$$f, g \in \mathrm{Int}_b(A, \mathbb{R}) \implies f \cdot g \in \mathrm{Int}_b(A, \mathbb{R})?$$

Lecture 29 — Monday, November 18, 2013

Lemma 29.1 Let $A \subseteq \mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be bounded. Let $c \geq 0$ be such that $|f(\vec{x})| \leq c$, for all $\vec{x} \in A$. Consider the new function $f^2 : A \to \mathbb{R}$, $f^2(\vec{x}) = (f(\vec{x}))^2$, for all $\vec{x} \in A$. Then for every $\emptyset \neq B \subseteq A$, we have

$$\mathrm{osc}_B(f^2) \leq 2c \cdot \mathrm{osc}_B(f).$$

Proof. Denote $\mathrm{osc}_B(f) =: \omega$. For any $\vec{x}, \vec{y} \in B$,

$$\begin{aligned} \left|f^2(\vec{x}) - f^2(\vec{y})\right| &= |(f(\vec{x}) - f(\vec{y}))(f(\vec{x}) + f(\vec{y}))| \\ &\leq |f(\vec{x}) - f(\vec{y})| \left(|f(\vec{x})| + |f(\vec{y})|\right) \\ &\leq \omega(c + c) \\ &= 2c\omega. \end{aligned}$$

Hence $2c\omega$ is an upper bound for the set

$$\{\left|f^2(\vec{x}) - f^2(\vec{y})\right| \mid \vec{x}, \vec{y} \in B\},$$

so $\mathrm{osc}_B(f^2) \leq 2c\omega$. $\square$

Lemma 29.2 Let $A \in \mathcal{M}_n$ with $A \neq \emptyset$. If $f \in \mathrm{Int}_b(A, \mathbb{R})$, then $f^2 \in \mathrm{Int}_b(A, \mathbb{R})$ as well.

Proof. f is bounded, so we can pick $c \geq 0$ such that $|f(\vec{x})| \leq c$ for all $\vec{x} \in A$. If $c = 0$, then f^2 is the zero function and the result is immediate. Assume therefore that $c > 0$.

Then $0 \leq f^2(\vec{x}) \leq c^2$ for all $\vec{x} \in A$, so f^2 is bounded. Let $\varepsilon > 0$. We verify that f^2 satisfies the "ε-Δ" criterion from Theorem 26.4. Since f itself is integrable, we can find a division $\Delta = \{A_1, \ldots, A_r\}$ of A such that $U(f, \Delta) - L(f, \Delta) < \frac{\varepsilon}{2c}$. For this Δ, we write

$$\begin{aligned} U(f^2, \Delta) - L(f^2, \Delta) &= \sum_{i=1}^{r} \mathrm{vol}(A_i) \cdot \mathrm{osc}_{A_i}(f^2) \quad \text{by Remark 25.15} \\ &\leq \sum_{i=1}^{r} \mathrm{vol}(A_i) \cdot 2c \cdot \mathrm{osc}_{A_i}(f) \quad \text{by Lemma 29.1} \\ &= 2c \cdot [U(f, \Delta) - L(f, \Delta)] \\ &< 2c \cdot \frac{\varepsilon}{2c} \\ &= \varepsilon. \end{aligned}$$

The criterion in Theorem 26.4 now shows that f^2 is integrable. □

Proof of Part (4) of Theorem 28.2. We must show that if $A \in \mathcal{M}_n$, $A \neq \emptyset$, and $f, g \in \mathrm{Int}_b(A, \mathbb{R})$, then $fg \in \mathrm{Int}_b(A, \mathbb{R})$. By Part (2) and Part (3) of Theorem 28.2, we have $f + g, f - g \in \mathrm{Int}_b(A, \mathbb{R})$. By Lemma 29.2, this implies

$$(f+g)^2, (f-g)^2 \in \mathrm{Int}_b(A, \mathbb{R}).$$

Therefore

$$\frac{1}{4}(f+g)^2 - \frac{1}{4}(f-g)^2 \in \mathrm{Int}_b(A, \mathbb{R}).$$

But

$$\frac{1}{4}(f+g)^2 - \frac{1}{4}(f-g)^2 = fg,$$

so $fg \in \mathrm{Int}_b(A, \mathbb{R})$. □

Remember this trick! Use linear combinations whenever you can.

Some Consequences of Theorem 28.2

Corollary 29.3 $A \in \mathcal{M}_n, A \neq \emptyset$. Let $f, g \in \mathrm{Int}_b(A, \mathbb{R})$ be such that $f \leq g$ (in the sense that $f(\vec{x}) \leq g(\vec{x})$ for all $\vec{x} \in A$). Then

$$\int_A f \leq \int_A g.$$

Proof. Consider the difference function $h = g - f \in \mathrm{Int}_b(A, \mathbb{R})$. We have $h(\vec{x}) \geq 0$ for all $\vec{x} \in A$. Hence

$$\int_A h = \underline{\int_A} h = \sup\{L(h, \Delta) \mid \Delta \text{ is a division of } A\} \geq L(h, \{A\}) = \mathrm{vol}(A) \cdot \inf_A(h) \geq 0.$$

By linearity, $\int_A g - \int_A f = \int_A h \geq 0$, and therefore $\int_A f \leq \int_A g$. □

Definition 29.4 (and Remark) Let $A \in \mathcal{M}_n$ with $A \neq \emptyset$. For $f, g \in \mathrm{Int}_b(A, \mathbb{R})$, define

$$\langle f, g \rangle \coloneqq \int_A fg \in \mathbb{R}.$$

This is well-defined by Part (4) of Theorem 28.2, and satisfies the following properties:

1. **Bilinearity**:

$$\begin{aligned}\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle &= \int_A (\alpha_1 f_1 + \alpha_2 f_2) g \\ &= \int_A \alpha_1 (f_1 g) + \alpha_2 (f_2 g) \\ &= \alpha_1 \int_A f_1 g + \alpha_2 \int_A f_2 g \\ &= \alpha_1 \langle f_1, g \rangle + \alpha_2 \langle f_2, g \rangle\end{aligned}$$

2. **Symmetry**:

$$\langle f, g \rangle = \int_A fg = \int_A gf = \langle g, f \rangle.$$

3. **Positive semidefiniteness**:

$$\langle f, f\rangle = \int_A f^2 \geq 0 \quad \text{for all } f \in \mathrm{Int}_b(A, \mathbb{R}).$$

However, $\langle f, f\rangle = 0$ does *not* necessarily imply $f = 0$.

Proposition 29.5 (Cauchy–Schwarz inequality for integrals) Let $A \in \mathcal{M}_n$ with $A \neq \emptyset$. For every $f, g \in \mathrm{Int}_b(A, \mathbb{R})$, we have

$$\left|\int_A fg\right| \leq \sqrt{\int_A f^2} \cdot \sqrt{\int_A g^2}.$$

Proof. Set $A_0 = \langle f, f\rangle$, $B_0 = \langle f, g\rangle$, and $C_0 = \langle g, g\rangle$. Positive semidefiniteness gives

$$0 \leq \langle f - tg, f - tg\rangle = A_0 - 2tB_0 + t^2 C_0 \quad \text{for every } t \in \mathbb{R}.$$

If $C_0 > 0$, the discriminant is nonpositive, so $B_0^2 \leq A_0 C_0$. If $C_0 = 0$ and $B_0 \neq 0$, the affine function $A_0 - 2tB_0$ is negative for a suitable t, a contradiction. Hence $B_0 = 0$ in this case, and $B_0^2 \leq A_0 C_0$ again. Taking square roots gives

$$|\langle f, g\rangle| \leq \sqrt{\langle f, f\rangle}\,\sqrt{\langle g, g\rangle}.$$

Equivalently,

$$\left|\int_A fg\right| \leq \sqrt{\int_A f^2}\,\sqrt{\int_A g^2}.$$

$\square$

Remark 29.6 We proved that if $f \in \mathrm{Int}_b(A, \mathbb{R})$, then $f^2 \in \mathrm{Int}_b(A, \mathbb{R})$, via the oscillation trick. A similar oscillation argument gives this:

$$\text{(II)} \quad \text{if } f \in \mathrm{Int}_b(A, \mathbb{R}), \text{ then } |f| \in \mathrm{Int}_b(A, \mathbb{R}),$$

where $|f| : A \to \mathbb{R}$ is defined by $|f|(\vec{x}) = |f(\vec{x})|$ for all $\vec{x} \in A$.

Proposition 29.7 $A \in \mathcal{M}_n$, $A \neq \emptyset$, $f \in \mathrm{Int}_b(A, \mathbb{R})$. Then

$$\left| \int_A f \right| \leq \int_A |f|.$$

Proof. Since $-|f| \leq f \leq |f|$, monotonicity of the integral gives

$$\int_A (-|f|) \leq \int_A f \leq \int_A |f|.$$

Let

$$c := \int_A |f|.$$

Then

$$-c \leq \int_A f \leq c,$$

so

$$\left| \int_A f \right| \leq c = \int_A |f|.$$

$\square$

Lecture 30 — Wednesday, November 20, 2013

Tinkering with the Domain of an Integrable Function

Proposition 30.1 (Extension by 0) Let $A \in \mathcal{M}_n$, $A \neq \emptyset$, and $f \in \mathrm{Int}_b(A, \mathbb{R})$. Let $B \in \mathcal{M}_n$ with $B \supseteq A$, and define $g : B \to \mathbb{R}$ by

$$g(\vec{x}) = \begin{cases} f(\vec{x}), & \vec{x} \in A, \\ 0, & \vec{x} \notin A. \end{cases}$$

Then $g \in \mathrm{Int}_b(B, \mathbb{R})$, and

$$\int_B g = \int_A f.$$

Proof. Since f is bounded and integrable, there exists a sequence (Δ_k) of divisions of A such that

$$U(f, \Delta_k) - L(f, \Delta_k) \to 0.$$

For each k, define a division $\widetilde{\Delta}_k$ of B by adjoining $B \setminus A$ to Δ_k when this set is nonempty. Thus, if $\Delta_k = \{A_1, \ldots, A_r\}$, then

$$\widetilde{\Delta}_k = \begin{cases} \{A_1, \ldots, A_r, B \setminus A\}, & B \setminus A \neq \emptyset, \\ \Delta_k, & B = A. \end{cases}$$

Because $g = 0$ on $B \setminus A$, we have

$$U(g, \widetilde{\Delta}_k) = U(f, \Delta_k), \qquad \text{and} \qquad L(g, \widetilde{\Delta}_k) = L(f, \Delta_k).$$

Hence

$$U(g, \widetilde{\Delta}_k) - L(g, \widetilde{\Delta}_k) = U(f, \Delta_k) - L(f, \Delta_k) \to 0.$$

So g is integrable on B. Also, since the upper and lower sums agree term-by-term, the integral values coincide:

$$\int_B g = \int_A f.$$

□

Corollary 30.2 Let $A, B \in \mathcal{M}_n$ such that $\emptyset \neq A \subseteq B$. Consider the indicator function $\mathbb{1}_A : B \to \mathbb{R}$. Then $\mathbb{1}_A \in \mathrm{Int}_b(B, \mathbb{R})$ with

$$\int_B \mathbb{1}_A = \mathrm{vol}(A).$$

Proof. In Proposition 30.1, take $f : A \to \mathbb{R}$ given by $f(\vec{x}) = 1$ for all $\vec{x} \in A$. By Theorem 28.2, specifically Part (1), we have $f \in \mathrm{Int}_b(A, \mathbb{R})$ and

$$\int_A f = \mathrm{vol}(A).$$

The extension of f by the value 0 from A to B is exactly $g = \mathbb{1}_A$, so

$$\int_B \mathbb{1}_A = \mathrm{vol}(A).$$

□

Corollary 30.3 Let $A_1, \ldots, A_p \in \mathcal{M}_n$ be nonempty and disjoint. Suppose we have $f_i \in \mathrm{Int}_b(A_i, \mathbb{R})$, $1 \le i \le p$. Put $A = \bigcup_{i=1}^p A_i$. Let $f : A \to \mathbb{R}$ be defined by

$$f(\vec{x}) = \begin{cases} f_1(\vec{x}), & \vec{x} \in A_1 \\ \vdots & \\ f_p(\vec{x}), & \vec{x} \in A_p \end{cases}.$$

Then $f \in \mathrm{Int}_b(A, \mathbb{R})$ and

$$\int_A f = \int_{A_1} f_1 + \cdots + \int_{A_p} f_p.$$

Proof. For $1 \le i \le p$, let $g_i : A \to \mathbb{R}$ be the extension of f_i by the value 0. Thus

$$g_i(\vec{x}) = \begin{cases} f_i(\vec{x}), & \vec{x} \in A_i, \\ 0, & \vec{x} \notin A_i. \end{cases}$$

By Proposition 30.1, we have $g_i \in \mathrm{Int}_b(A, \mathbb{R})$ with

$$\int_A g_i = \int_{A_i} f_i$$

for $1 \le i \le p$. Also, $g_1 + \cdots + g_p = f$ (the function in the corollary). Therefore, Theorem 28.2 implies that $f \in \mathrm{Int}_b(A, \mathbb{R})$, and

$$\int_A f = \sum_{i=1}^p \int_A g_i = \sum_{i=1}^p \int_{A_i} f_i.$$

$\square$

Proposition 30.4 Let $S \subseteq \mathbb{R}^n$ be a set of content zero with $S \neq \emptyset$. Then any bounded function $f : S \to \mathbb{R}$ is integrable, with

$$\int_S f = 0.$$

Proof. Fix a bounded function $f : S \to \mathbb{R}$. Consider the division $\Delta_1 = \{S\}$ of S with one piece.

Then

$$U(f,\Delta_1) = \mathrm{vol}(S)\cdot \sup_S(f) = 0, \qquad \text{and} \qquad L(f,\Delta_1) = \mathrm{vol}(S)\cdot \inf_S(f) = 0.$$

Since $\overline{\int_S} f$ is the infimum of upper sums and $\underline{\int_S} f$ is the supremum of lower sums, we get

$$\overline{\int_S} f \le U(f,\Delta_1) = 0 \qquad \text{and} \qquad L(f,\Delta_1) = 0 \le \underline{\int_S} f.$$

Also,

$$\underline{\int_S} f \le \overline{\int_S} f.$$

Therefore all inequalities are equalities, so

$$\overline{\int_S} f = 0 = \underline{\int_S} f.$$

Hence $f \in \mathrm{Int}_b(S,\mathbb{R})$ and $\displaystyle\int_S f = 0$. □

Proposition 30.5 Let $A \in \mathcal{M}_n$, $A \neq \emptyset$. Let $f_1, f_2 : A \to \mathbb{R}$ be bounded functions, and suppose there exists a set $S \subseteq A$ of content zero such that $f_1(\vec{x}) = f_2(\vec{x})$ for all $\vec{x} \in A \setminus S$ (possibly $\vec{x} \in S$ too but possibly not). If $f_1 \in \mathrm{Int}_b(A,\mathbb{R})$, then $f_2 \in \mathrm{Int}_b(A,\mathbb{R})$ and

$$\int_A f_2 = \int_A f_1.$$

Proof. If $S = \emptyset$, then $f_1 = f_2$, and there is nothing to prove. So assume that $S \neq \emptyset$. Define $u : A \to \mathbb{R}$ by $u = f_2 - f_1$, and let $v : S \to \mathbb{R}$ be its restriction to S:

$$v = u|_S.$$

Both u and v are bounded because f_1 and f_2 are bounded. Since S has zero content and v is bounded, we have $v \in \mathrm{Int}_b(S,\mathbb{R})$ and

$$\int_S v = 0.$$

Moreover, u is the extension of v by the value 0, because $f_2(\vec{x}) - f_1(\vec{x}) = 0$ for every $\vec{x} \in A \setminus S$.

By Proposition 30.1, it follows that $u \in \mathrm{Int}_b(A, \mathbb{R})$ and

$$\int_A u = \int_S v = 0.$$

Finally, since $u = f_2 - f_1$, we have $f_2 = f_1 + u$. Applying part (2) of Theorem 28.2, we obtain $f_2 \in \mathrm{Int}_b(A, \mathbb{R})$ and

$$\int_A f_2 = \int_A f_1 + \int_A u = \int_A f_1.$$

□

Lecture 31 — Friday, November 22, 2013

How to Calculate Integrals

Proposition 31.1 Let $M \subseteq \mathbb{R}^m$ be bounded and nonempty, and let $f : M \to \mathbb{R}^n$ be Lipschitz, where $m < n$. Then the image set

$$f(M) = \{\vec{x} \in \mathbb{R}^n \mid \text{there exists } \vec{z} \in M \text{ such that } f(\vec{z}) = \vec{x}\}$$

is a subset of $\mathbb{R}^n$ with content zero.

Proof. Choose $R > 0$ such that $M \subseteq [-R, R]^m$, and let $L \geq 0$ be a Lipschitz constant for f. If $L = 0$, then f is constant on M, so $f(M)$ is a singleton and can be covered by a half-open n-rectangle of arbitrarily small volume.

Assume $L > 0$. For $0 < \delta \leq 1$, cover $[-R, R]^m$ by a regular grid of cubes of side length at most δ. There is a constant C, independent of δ, such that the grid contains at most $C\delta^{-m}$ cubes. For each cube Q meeting M, choose a point $\vec{z}_Q \in Q \cap M$. If $\vec{z} \in Q \cap M$, then

$$\|f(\vec{z}) - f(\vec{z}_Q)\| \leq L\|\vec{z} - \vec{z}_Q\| \leq L\sqrt{m}\,\delta.$$

Consequently, $f(Q \cap M)$ is contained in the half-open rectangle

$$P_Q = \prod_{j=1}^{n} \big(f^{(j)}(\vec{z}_Q) - 2L\sqrt{m}\,\delta, f^{(j)}(\vec{z}_Q) + 2L\sqrt{m}\,\delta\big].$$

The rectangles P_Q cover $f(M)$, and their total volume is at most

$$C\delta^{-m}\big(4L\sqrt{m}\,\delta\big)^n = C\big(4L\sqrt{m}\big)^n\delta^{n-m}.$$

Since $n-m>0$, this bound tends to 0 as $\delta\to 0$. Thus, for every $\varepsilon>0$, the image $f(M)$ has a finite half-open rectangular cover of total volume less than ε. Therefore $f(M)$ has content zero. $\square$

Corollary 31.2 Let $A\subseteq\mathbb{R}^n$ be bounded and nonempty. Suppose that we can find $m<n$, bounded sets $M_1,\dots,M_p\subseteq\mathbb{R}^m$, and Lipschitz functions $f_i : M_i\to\mathbb{R}^n$ such that

$$f_1(M_1)\cup\cdots\cup f_p(M_p)\supseteq \operatorname{bd}(A).$$

Then $A\in\mathcal{M}_n$.

Proof. Denote $S_i=f_i(M_i)$ for $1\le i\le p$. By Proposition 31.1, each S_i has zero content. Hence $S_1\cup\cdots\cup S_p$ has zero content, and therefore so does $\operatorname{bd}(A)\subseteq S_1\cup\cdots\cup S_p$. Since A is bounded and $\operatorname{bd}(A)$ has zero content, it follows that $A\in\mathcal{M}_n$. $\square$

Example 31.3 A model example is

$$A=\{\vec{x}\in\mathbb{R}^3 \mid \|\vec{x}\|<1\},$$

whose boundary is the sphere

$$S=\{(x,y,z)\in\mathbb{R}^3 \mid x^2+y^2+z^2=1\}.$$

To apply Corollary 31.2 cover S by finitely many patches that are images of Lipschitz maps from subsets of $\mathbb{R}^2$. For instance, on the patch $Z^+=\{(x,y,z)\in S \mid z\ge 1/2\}$ we have $x^2+y^2\le 3/4$, so with

$$M=\{(x,y) \mid x^2+y^2\le 3/4\},\qquad \text{and}\qquad f(x,y)=(x,y,\sqrt{1-x^2-y^2}),$$

the patch Z^+ is the image of M. It remains to verify that f is Lipschitz. Set $g(x,y)=\sqrt{1-x^2-y^2}$ on M, so $f(x,y)=(x,y,g(x,y))$. For $(x,y)\in M$ we have

$$1-x^2-y^2\ge\frac{1}{4},$$

hence

$$\left|\frac{\partial g}{\partial x}(x,y)\right| = \frac{|x|}{\sqrt{1-x^2-y^2}} \le 2|x|, \qquad \text{and} \qquad \left|\frac{\partial g}{\partial y}(x,y)\right| = \frac{|y|}{\sqrt{1-x^2-y^2}} \le 2|y|.$$

Since $x^2+y^2 \le 3/4$, we get

$$\|\nabla g(x,y)\| \le 2\sqrt{x^2+y^2} \le \sqrt{3} \quad \text{for all } (x,y) \in M.$$

Now let $u, v \in M$ and define $\phi(t) = g(v + t(u-v))$ on $[0,1]$. By the mean value theorem in one variable,

$$|g(u)-g(v)| = |\phi(1)-\phi(0)| \le \sup_{t\in[0,1]} |\phi'(t)| \le \sup_M \|\nabla g\| \, \|u-v\| \le \sqrt{3}\, \|u-v\|.$$

Therefore,

$$\|f(u)-f(v)\|^2 = \|u-v\|^2 + |g(u)-g(v)|^2 \le (1+3)\|u-v\|^2,$$

so

$$\|f(u)-f(v)\| \le 2\|u-v\|.$$

Thus f is 2-Lipschitz on M.

The upper spherical patch represented by f is shown in Figure 16.

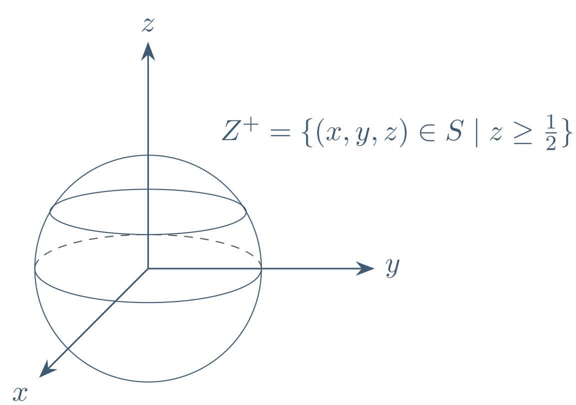

FIGURE 16

Which Functions Are Integrable?

Theorem 31.4 Let $A \in \mathcal{M}_n$ with $A \neq \emptyset$, and let $f : A \to \mathbb{R}$ be bounded. If f is continuous on A, then f is integrable.

Proof idea. Fix $\varepsilon > 0$, and choose M with $|f| \leq M$. The boundary of A has content zero, so we can cover it by finitely many rectangles whose total volume is small enough that their contribution to the gap between the Darboux sums is less than $\varepsilon/2$, even using the crude bound $\operatorname{osc}(f) \leq 2M$.

What remains after removing a small neighbourhood of the boundary is a compact subset of the interior of A. On this compact core, continuity gives a finite collection of neighbourhoods on each of which the oscillation of f is less than

$$\frac{\varepsilon}{2(1 + \operatorname{vol}(A))}.$$

Refining the boundary rectangles and these interior neighbourhoods produces a division Δ of A. The boundary pieces contribute less than $\varepsilon/2$ to $U(f, \Delta) - L(f, \Delta)$, and the interior pieces contribute less than

$$\operatorname{vol}(A)\frac{\varepsilon}{2(1 + \operatorname{vol}(A))} < \frac{\varepsilon}{2}.$$

Hence $U(f, \Delta) - L(f, \Delta) < \varepsilon$, which is the ε-Δ criterion. □

Here is a little twist on that:

Corollary 31.5 Let $A \in \mathcal{M}_n, A \neq \emptyset$ and let $f : A \to \mathbb{R}$ be bounded. Suppose we find $B \subseteq A$ with content zero such that f is continuous at every $\vec{a} \in A \setminus B$. Then $f \in \operatorname{Int}_b(A, \mathbb{R})$.

Proof. If $B = \emptyset$, then f is continuous on all of A, so Theorem 31.4 applies directly. If $A \setminus B = \emptyset$, then A has content zero and Proposition 30.4 applies. Assume from now on that both B and $G := A \setminus B$ are nonempty.

Let $u : B \to \mathbb{R}$ and $v : G \to \mathbb{R}$ be the restrictions of f to B and G, respectively. By Proposition 25.9, the set B is Jordan measurable; hence $G = A \setminus B$ is Jordan measurable by the field property in Proposition 25.7. Then $u \in \operatorname{Int}_b(B, \mathbb{R})$ because B has zero content. Also, $v \in \operatorname{Int}_b(G, \mathbb{R})$ because v is bounded and continuous on G, by Theorem 31.4. Let $\tilde{u} : A \to \mathbb{R}$ and $\tilde{v} : A \to \mathbb{R}$ be the extensions by 0 of u and v. By Proposition 30.1, $\tilde{u}, \tilde{v} \in \mathrm{Int}_b(A, \mathbb{R})$. Finally,

$$f = \tilde{u} + \tilde{v},$$

so $f \in \mathrm{Int}_b(A, \mathbb{R})$ by Theorem 28.2. □

Note that $\tilde{u} = f \cdot \mathbb{1}_B$ and $\tilde{v} = f \cdot \mathbb{1}_G$ in the proof above.

Lecture 32 — Monday, November 25, 2013

Iterated integrals and Fubini's theorem

Definition 32.1 (and Remark) Let $\tilde{\mathcal{P}}_n$ denote the collection of all rectangles in $\mathbb{R}^n$, including open, closed, half-open, and mixed types. For example,

$$(p_1, q_1) \times [p_2, q_2] \in \tilde{\mathcal{P}}_2.$$

By definition, $\tilde{\mathcal{P}}_n \supseteq \mathcal{P}_n$. Given $A \in \mathcal{M}_n$ and $f \in \mathrm{Int}_b(A, \mathbb{R})$, we can choose $C \in \tilde{\mathcal{P}}_n$ with $A \subseteq C$ and compute

$$\int_A f$$

instead as

$$\int_C \tilde{f},$$

where $\tilde{f} : C \to \mathbb{R}$ is the extension of f by 0.

Definition 32.2 (and Remarks) Suppose $n = p + q$, with $p, q \in \mathbb{N}$.

1. For $A \subseteq \mathbb{R}^p, B \subseteq \mathbb{R}^q$, we denote $A \times B = \{(\vec{a}, \vec{b}) \mid \vec{a} \in A, \vec{b} \in B\} \subseteq \mathbb{R}^{p+q} = \mathbb{R}^n$.
2. If $A \in \tilde{\mathcal{P}}_p$ and $B \in \tilde{\mathcal{P}}_q$, then $A \times B \in \tilde{\mathcal{P}}_n$. Conversely, every $C \in \tilde{\mathcal{P}}_n$ can be written as $C = A \times B$.
3. Suppose $A \in \tilde{\mathcal{P}}_p$, $B \in \tilde{\mathcal{P}}_q$, and $C = A \times B \in \tilde{\mathcal{P}}_n$. If $f : C \to \mathbb{R}$, then for each $\vec{a} \in A$ we

define the partial function $f_{\vec{a}} : B \to \mathbb{R}$ by

$$f_{\vec{a}}(\vec{b}) = f(\vec{a}, \vec{b}), \quad \vec{b} \in B.$$

Example 32.3 For example, take $n = 2, p = q = 1$, $A = (p_1, q_1)$, $B = [p_2, q_2]$, and $C = A \times B \subseteq \mathbb{R}^2$ with $f : A \times B \to \mathbb{R}$. Here $f_{\vec{a}}(\vec{b}) = f(\vec{a}, \vec{b})$ for all $\vec{b} \in [p_2, q_2]$, and Figure 17 shows the corresponding slice of C.

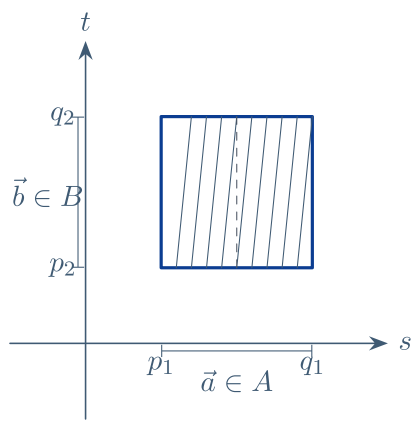

FIGURE 17

Theorem 32.4 (Fubini's theorem) Let $C = A \times B \in \tilde{\mathcal{P}}_n$ and let $f : C \to \mathbb{R}$. Assume that $f \in \operatorname{Int}_b(C, \mathbb{R})$, and that for every $\vec{a} \in A$, the partial function $f_{\vec{a}} : B \to \mathbb{R}$ belongs to $\operatorname{Int}_b(B, \mathbb{R})$. Define $F : A \to \mathbb{R}$ by

$$F(\vec{a}) = \int_B f_{\vec{a}}.$$

Then $F \in \operatorname{Int}_b(A, \mathbb{R})$ and

$$\int_C f = \int_A F.$$

The proof is too boring. We don't want to end the course on a boring note.

With the notation above, write $\vec{x} = (\vec{u}, \vec{v})$ with $\vec{u} \in A$ and $\vec{v} \in B$. Then Fubini's theorem can be written formally as

$$\int_{A \times B} f(\vec{u}, \vec{v})\, \mathrm{d}(\vec{u}, \vec{v}) = \int_A \left(\int_B f(\vec{u}, \vec{v})\, \mathrm{d}\vec{v} \right) \mathrm{d}\vec{u}.$$

Thus, under the stated hypotheses, an integral over a product set can be computed by iterating one-dimensional or lower-dimensional integrals.

Remark 32.5 In Fubini's theorem, the n coordinates need not be grouped in their original order. For example, if

$$C = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3] \subseteq \mathbb{R}^3$$

and if we fix $y \in [a_2, b_2]$, then we may define

$$f_y(x, z) = f(x, y, z)$$

on $[a_1, b_1] \times [a_3, b_3]$. Under the appropriate hypotheses,

$$\int_C f(x, y, z)\, \mathrm{d}(x, y, z) = \int_{a_2}^{b_2} \left(\int_{[a_1,b_1]\times[a_3,b_3]} f_y(x, z)\, \mathrm{d}(x, z) \right) \mathrm{d}y.$$

Example 32.6 Let $n = 2$, $p = q = 1$,

$$E = \left\{ (s, t) \in \mathbb{R}^2 \;\middle|\; \left(\frac{s}{2}\right)^2 + t^2 \leq 1 \right\},$$

and Figure 18 places this ellipse inside the rectangle used for its zero extension.

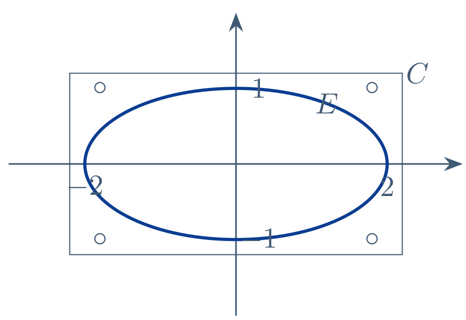

FIGURE 18

and define $f : E \to \mathbb{R}$ by

$$f(s, t) = \sqrt{1 - \left(\frac{s}{2}\right)^2 - t^2}.$$

Then $E \in \mathcal{M}_2$, and f is continuous, hence integrable by Theorem 31.4. To compute

$$\int_E f,$$

let $C = [-2,2] \times [-1,1]$ and let $g : C \to \mathbb{R}$ be the extension of f by 0. For each fixed s, the function $t \mapsto g(s,t)$ is continuous except possibly at the two boundary points of the corresponding vertical slice of E; hence it is integrable by Corollary 31.5. Fubini's theorem now gives

$$\begin{aligned}
\int_E f = \int_C g &= \int_{-2}^{2} \int_{-1}^{1} g(s,t)\,\mathrm{d}t\,\mathrm{d}s \\
&= \int_{-2}^{2} \int_{-\sqrt{1-s^2/4}}^{\sqrt{1-s^2/4}} \sqrt{1 - \frac{s^2}{4} - t^2}\,\mathrm{d}t\,\mathrm{d}s \\
&= \int_{-2}^{2} \frac{\pi}{2} \left(1 - \frac{s^2}{4}\right) \mathrm{d}s \\
&= \frac{\pi}{2} \left[s - \frac{s^3}{12} \right]_{-2}^{2} \\
&= \frac{4\pi}{3}.
\end{aligned}$$

Lecture 33 — Wednesday, November 27, 2013

$C^1(A, \mathbb{R}^m)$ and Change of Variables

Definition 33.1 Let $m, n \in \mathbb{N}$, let $A \subseteq \mathbb{R}^n$ be open, and let $f : A \to \mathbb{R}^m$ have component functions $f^{(1)}, \ldots, f^{(m)} : A \to \mathbb{R}$ so that $f(\vec{x}) = (f^{(1)}(\vec{x}), \ldots, f^{(m)}(\vec{x}))$ for $\vec{x} \in A$.

1. We say that f is a $\boldsymbol{C^1}$ **function** from A to $\mathbb{R}^m$ (written $f \in C^1(A, \mathbb{R}^m)$) if each component $f^{(i)}$ belongs to $C^1(A, \mathbb{R})$.
2. If $f \in C^1(A, \mathbb{R}^m)$, then for every $\vec{a} \in A$ we form the gradient vectors $(\nabla f^{(i)})(\vec{a})$,

$1 \le i \le m$. The $m \times n$ matrix

$$(Jf)(\vec{a}) := \begin{bmatrix} (\nabla f^{(1)})(\vec{a}) \\ \vdots \\ (\nabla f^{(m)})(\vec{a}) \end{bmatrix}$$

is called the **Jacobian matrix** of f at $\vec{a}$.

Some special cases of Definition 33.1:

1. $m = 1$, general n: here $A \subseteq \mathbb{R}^n$ is open and $f \in C^1(A, \mathbb{R})$ has one component. For $\vec{a} \in A$,

$$(Jf)(\vec{a}) = \begin{bmatrix} (\partial_1 f)(\vec{a}) & \dots & (\partial_n f)(\vec{a}) \end{bmatrix},$$

   so the Jacobian is the gradient viewed as a row matrix.

2. General m, $n = 1$: let $A = I$ be an open interval. A map $f \in C^1(I, \mathbb{R}^m)$ is a C^1 path in $\mathbb{R}^m$,

$$f = (f^{(1)}, \dots, f^{(m)}), \quad f^{(i)} \in C^1(I, \mathbb{R}).$$

   Then the Jacobian is

$$(Jf)(a) = \begin{bmatrix} (f^{(1)})'(a) \\ \vdots \\ (f^{(m)})'(a) \end{bmatrix},$$

   which is another old friend, the velocity vector.

Definition 33.2 Let $A \subseteq \mathbb{R}^n$ be open and let $f \in C^1(A, \mathbb{R}^n)$. For every $\vec{a} \in A$, we have $(Jf)(\vec{a}) \in M_{n\times n}(\mathbb{R})$. Define

$$|J|_f(\vec{a}) := |\det((Jf)(\vec{a}))|.$$

This is called the **Jacobian (or Jacobian determinant) of f at $\vec{a}$**. Hence $|J|_f$ is a function $A \to [0, \infty)$.

If $f \in C^1(A, \mathbb{R}^n)$, then $|J|_f : A \to \mathbb{R}$ is continuous because the determinant is a polynomial expression in the matrix entries. For example, when $n = 2$,

$$\begin{aligned} |J|_f(\vec{a}) &= \left| \det \begin{bmatrix} (\partial_1 f^{(1)})(\vec{a}) & (\partial_2 f^{(1)})(\vec{a}) \\ (\partial_1 f^{(2)})(\vec{a}) & (\partial_2 f^{(2)})(\vec{a}) \end{bmatrix} \right| \\ &= \left| (\partial_1 f^{(1)})(\vec{a}) \cdot (\partial_2 f^{(2)})(\vec{a}) - (\partial_2 f^{(1)})(\vec{a}) \cdot (\partial_1 f^{(2)})(\vec{a}) \right|. \end{aligned}$$

and the right-hand side is continuous because each partial derivative $\partial_j f^{(i)}$ is continuous.

Example 33.3 Let $n = 2$. Fix some radii $r_2 > r_1 \geq 0$ and take $A = (r_1, r_2) \times (0, 2\pi)$. Let $f : A \to \mathbb{R}^2$ be defined by $f(r, \theta) = (r\cos\theta, r\sin\theta)$. Find the Jacobian of f at $\vec{a}$.

Solution. We calculate

$$f^{(1)}(r,\theta) = r\cos\theta \implies (\nabla f^{(1)})(r,\theta) = (\cos\theta, -r\sin\theta)$$

and

$$f^{(2)}(r,\theta) = r\sin\theta \implies (\nabla f^{(2)})(r,\theta) = (\sin\theta, r\cos\theta).$$

Hence

$$(Jf)(r,\theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

and

$$\det((Jf)(r,\theta)) = r\cos^2\theta - (-r\sin^2\theta) = r.$$

The conclusion is that for $\vec{a} = (r, \theta) \in A$, we have $|J|_f(\vec{a}) = r$. $\square$

Definition 33.4 $A, B \subseteq \mathbb{R}^n$ open, $f : A \to B$ bijection. Let $g = f^{-1} : B \to A$. If both f and g are C^1 functions (when viewed as $f : A \to \mathbb{R}^n$ and $g : B \to \mathbb{R}^n$), then we say that f is a $\boldsymbol{C^1}$ **diffeomorphism**.

Example 33.5 Again take $A = (r_1, r_2) \times (0, 2\pi) \subseteq \mathbb{R}^2$ and

$$f(r,\theta) = (r\cos\theta, r\sin\theta).$$

Then

$$f(A) = B = \{(s,t) \in \mathbb{R}^2 \mid r_1 < \sqrt{s^2+t^2} < r_2\} \setminus \{(s,0) \in \mathbb{R}^2 \mid r_1 < s < r_2\}.$$

The map f is a bijection from A to B. Its inverse is given by the radius $\sqrt{s^2+t^2}$ together with the unique angle in $(0, 2\pi)$; both components are C^1 on the slit annulus B. Thus f is a C^1 diffeomorphism.

Lecture 34 — Friday, November 29, 2013

Recall that for $f \in C^1(A, \mathbb{R}^m)$, the Jacobian matrix at $\vec{a} \in A$ is

$$(Jf)(\vec{a}) = \begin{bmatrix} (\partial_1 f^{(1)})(\vec{a}) & \dots & (\partial_n f^{(1)})(\vec{a}) \\ \vdots & \ddots & \vdots \\ (\partial_1 f^{(m)})(\vec{a}) & \dots & (\partial_n f^{(m)})(\vec{a}) \end{bmatrix}$$

and $(Jf)(\vec{a}) \in M_{m,n}(\mathbb{R})$. When $m = n$, the Jacobian determinant is

$$|J|_f(\vec{a}) = |\det(Jf(\vec{a}))|,$$

and $|J|_f : A \to \mathbb{R}$ is continuous. Also recall that a C^1 diffeomorphism between open sets $A, B \subseteq \mathbb{R}^n$ is a bijection $\phi : A \to B$ such that both ϕ and $\phi^{-1} : B \to A$ are C^1.

The next formula concerns a C^1 diffeomorphism $\phi : A \to B$. Given $g : B \to \mathbb{R}$, define $f = g \circ \phi : A \to \mathbb{R}$. We would like to relate

$$\int_B g(\vec{y})\,\mathrm{d}\vec{y} \quad \text{and} \quad \int_A f(\vec{x})\,\mathrm{d}\vec{x}.$$

Theorem 34.1 Let $A, B \subseteq \mathbb{R}^n$ be open, bounded, and Jordan measurable, and let $\phi : A \to B$ be a C^1 diffeomorphism. Assume that the first derivatives of ϕ and ϕ^{-1} are bounded. Then for every $g \in \mathrm{Int}_b(B, \mathbb{R})$, the function $f = g \circ \phi$ belongs to $\mathrm{Int}_b(A, \mathbb{R})$, and the change-of-variables formula holds:

$$\text{(Ch-Var)} \qquad \int_B g(\vec{y})\,\mathrm{d}\vec{y} = \int_A f(\vec{x}) \cdot |J|_\phi(\vec{x})\,\mathrm{d}\vec{x}.$$

The commuting maps behind the change-of-variables formula are summarized in Figure 19.

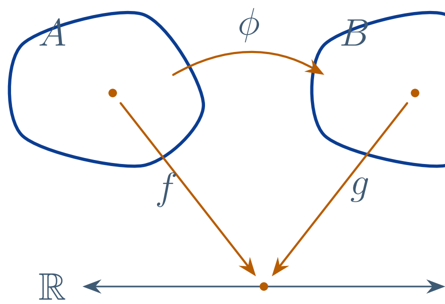

FIGURE 19

Formula (Ch-Var) can be remembered as the substitution $\vec{y} = \phi(\vec{x})$, where $\mathrm{d}\vec{y} = |J|_\phi(\vec{x})\,\mathrm{d}\vec{x}$. Indeed, we can write formally

$$\int_B g(\vec{y})\,\mathrm{d}\vec{y} = \int_A g(\phi(\vec{x})) \cdot |J|_\phi(\vec{x})\,\mathrm{d}\vec{x} = \int_A f(\vec{x}) \cdot |J|_\phi(\vec{x})\,\mathrm{d}\vec{x}.$$

Note the analogy with the u-substitution formula for one variable. There, we have $y = \phi(x)$, with $\mathrm{d}y = \phi'(x)\,\mathrm{d}x$.

Example 34.2 (Polar coordinates) Take $n = 2$ and $0 < r_1 < r_2$. Use the diffeomorphism from Example 33.5:

$$A = (r_1, r_2) \times (0, 2\pi) \qquad \text{and} \qquad \phi(r, \theta) = (r\cos\theta, r\sin\theta),$$

with

$$B = \{\vec{x} \in \mathbb{R}^2 \mid r_1 < \|\vec{x}\| < r_2\} \setminus \{(s, 0) \mid r_1 \le s \le r_2\}.$$

We computed $|J|_\phi(r, \theta) = r$. If $g \in \mathrm{Int}_b(B, \mathbb{R})$ and $f = g \circ \phi$, then by change of variables

$$\int_B g(\vec{y})\,\mathrm{d}\vec{y} = \int_A f(r, \theta)\,|J|_\phi(r, \theta)\,\mathrm{d}(r, \theta) = \int_A f(r, \theta)\,r\,\mathrm{d}(r, \theta).$$

Figure 20 shows how the polar rectangle maps onto the slit annulus.

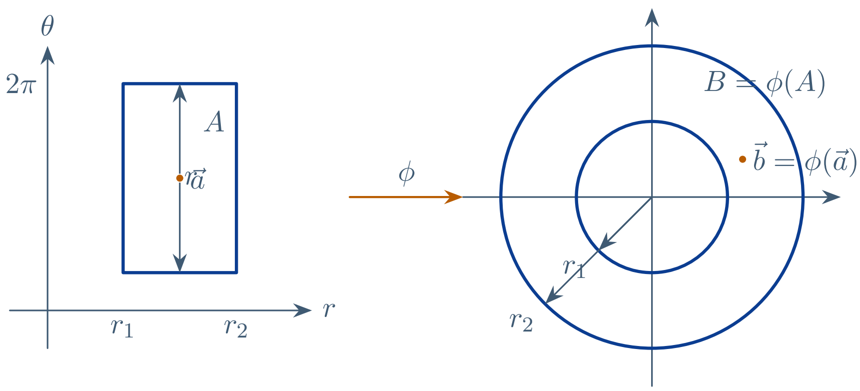

FIGURE 20

Example 34.3 Compute

$$\int_{D_R} e^{-(s^2+t^2)}\,\mathrm{d}(s,t), \quad D_R = B(\vec{0}; R),\ R > 0.$$

Solution. This is a **Gaussian integral**. Let

$$0 < \rho < R, \qquad A_{\rho,R} = (\rho, R) \times (0, 2\pi),$$

and let

$$B_{\rho,R} = \{(s,t) \mid \rho < \sqrt{s^2+t^2} < R\} \setminus \{(s,0) \mid \rho < s < R\}.$$

Define $\phi : A_{\rho,R} \to B_{\rho,R}$ by $\phi(r,\theta) = (r\cos\theta, r\sin\theta)$. On these domains, both ϕ and ϕ^{-1} have bounded first derivatives, so the stated change-of-variables theorem applies. Set $g(s,t) = e^{-(s^2+t^2)}$ and let $f = g \circ \phi$. Then

$$f(r,\theta) = g(r\cos\theta, r\sin\theta) = e^{-(r^2\cos^2\theta + r^2\sin^2\theta)} = e^{-r^2}.$$

Therefore,

$$\begin{aligned}\int_{B_{\rho,R}} e^{-(s^2+t^2)}\,\mathrm{d}(s,t) &= \int_{A_{\rho,R}} e^{-r^2}\, r\,\mathrm{d}(r,\theta) \\ &= \int_0^{2\pi}\int_\rho^R e^{-r^2}\, r\,\mathrm{d}r\,\mathrm{d}\theta \\ &= \pi(e^{-\rho^2} - e^{-R^2}).\end{aligned}$$

Removing the radial slit does not change the integral because the slit has content zero. After ignoring that slit, the part of D_R not accounted for by the annulus is contained in $[-\rho, \rho]^2$, and the integrand is at most 1. Therefore

$$0 \leq \int_{D_R} e^{-(s^2+t^2)}\,\mathrm{d}(s,t) - \pi(e^{-\rho^2} - e^{-R^2}) \leq 4\rho^2.$$

Letting $\rho \downarrow 0$ gives

$$\int_{D_R} e^{-(s^2+t^2)}\,\mathrm{d}(s,t) = \pi(1 - e^{-R^2}).$$

□

Example 34.4 Let $n = 2$. Consider the linear map $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $\phi(x,y) = (ax+by, cx+dy)$. This can be written as $\phi(\vec{x}) = M\vec{x}$ where $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The Jacobian matrix is $(J\phi)(x,y) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = M$. Thus, the Jacobian determinant is $|J|_\phi(x,y) = |\det(M)| = |ad - bc|$.

If $\det(M) \neq 0$, then ϕ is a C^1 diffeomorphism from $\mathbb{R}^2$ to $\mathbb{R}^2$. For any open, bounded, Jordan measurable set $A \subseteq \mathbb{R}^2$, let $B = \phi(A)$. Then for a bounded continuous function $g : B \to \mathbb{R}$, the change-of-variables theorem gives:

$$\iint_B g(u,v)\,\mathrm{d}u\,\mathrm{d}v = \iint_A g(\phi(x,y)) \cdot |ad - bc|\,\mathrm{d}x\,\mathrm{d}y.$$

In particular, if $g \equiv 1$, we find that $\text{Area}(B) = |ad - bc| \cdot \text{Area}(A)$. This shows that the determinant of a linear map represents the factor by which the map scales areas (or volumes in higher dimensions).

Example 34.5 (Spherical coordinates in $\mathbb{R}^3$) Let $A = (0,\infty) \times (0,\pi) \times (0, 2\pi)$. Define $\Phi : A \to \mathbb{R}^3$ by

$$\Phi(r,\varphi,\theta) = (r\sin\varphi\cos\theta, r\sin\varphi\sin\theta, r\cos\varphi).$$

The Jacobian matrix is

$$J\Phi(r,\varphi,\theta) = \begin{pmatrix} \sin\varphi\cos\theta & r\cos\varphi\cos\theta & -r\sin\varphi\sin\theta \\ \sin\varphi\sin\theta & r\cos\varphi\sin\theta & r\sin\varphi\cos\theta \\ \cos\varphi & -r\sin\varphi & 0 \end{pmatrix}.$$

Calculating the determinant (for example, by expanding along the third row), we get

$$\begin{aligned}\det(J\Phi) &= \cos\varphi\big(r^2\sin\varphi\cos\varphi\cos^2\theta + r^2\sin\varphi\cos\varphi\sin^2\theta\big)\\ &\quad + r\sin\varphi\big(r\sin^2\varphi\cos^2\theta + r\sin^2\varphi\sin^2\theta\big)\\ &= r^2\sin\varphi\cos^2\varphi + r^2\sin^3\varphi\\ &= r^2\sin\varphi(\cos^2\varphi+\sin^2\varphi)\\ &= r^2\sin\varphi.\end{aligned}$$

Since $\varphi\in(0,\pi)$, $\sin\varphi>0$, so $|J|_\Phi(r,\varphi,\theta) = r^2\sin\varphi$.

Lecture 35 — Monday, December 02, 2013

Example 35.1 A standard Gaussian integral computation shows that although there is no elementary antiderivative for e^{-t^2}, we can still prove

$$\int_{-\infty}^{\infty} e^{-t^2}\,\mathrm{d}t = \sqrt{\pi}.$$

Proof. *Claim.*

$$\lim_{k\to\infty}\int_{-k}^{k} e^{-t^2}\,\mathrm{d}t = \sqrt{\pi}.$$

Verification. denote

$$a_k = \int_{-k}^{k} e^{-t^2}\,\mathrm{d}t \quad (k\in\mathbb{N}).$$

Then

$$a_k^2 = \left(\int_{-k}^{k} e^{-s^2}\,\mathrm{d}s\right)\left(\int_{-k}^{k} e^{-t^2}\,\mathrm{d}t\right) = \int_{[-k,k]^2} e^{-(s^2+t^2)}\,\mathrm{d}(s,t).$$

Observe: denote $D_R = \{\vec{x}\in\mathbb{R}^2 \mid \|\vec{x}\| < R\}$ for $R>0$. Then $D_k\subseteq[-k,k]^2\subseteq D_{2k}$. Since $e^{-(s^2+t^2)}$ is a positive function, we get

$$\int_{D_k} e^{-(s^2+t^2)}\,\mathrm{d}(s,t) \le \int_{[-k,k]^2} e^{-(s^2+t^2)}\,\mathrm{d}(s,t) \le \int_{D_{2k}} e^{-(s^2+t^2)}\,\mathrm{d}(s,t).$$

Figure 21 depicts the disk–square inclusions used in this squeeze.

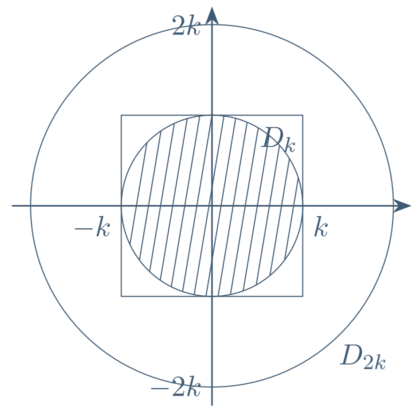

FIGURE 21

(In general, if $g \geq 0$ is continuous and $A \subseteq B$ are measurable, then

$$\int_B g \geq \int_A g$$

.)

Now, the middle integral is a_k^2. The left side is $\pi(1 - e^{-k^2})$, and by the same computation the right side is $\pi(1 - e^{-4k^2})$. Therefore,

$$\pi(1 - e^{-k^2}) \leq a_k^2 \leq \pi(1 - e^{-4k^2}), \quad k \in \mathbb{N}.$$

Letting $k \to \infty$ and applying the squeeze theorem gives $a_k^2 \to \pi$. Since $a_k \geq 0$, it follows that $a_k \to \sqrt{\pi}$. $\square$

Why does the polar coordinate formula contain the factor r? Recall that

$$\phi(r, \theta) = (r \cos\theta, r \sin\theta)$$

maps

$$A = (r_1, r_2) \times (0, 2\pi)$$

onto

$$B = \{\vec{x} \in \mathbb{R}^2 \mid r_1 < \|\vec{x}\| < r_2\} \setminus \{(s, 0) \mid r_1 \leq s \leq r_2\}.$$

For $g \in \mathrm{Int}_b(B, \mathbb{R})$, if $f = g \circ \phi$, then

$$\int_B g(s,t)\,\mathrm{d}(s,t) = \int_A f(r,\theta)\,r\,\mathrm{d}(r,\theta).$$

The reason is geometric. Figure 22 shows how a rectangular polar partition maps to annular sectors, first globally and then for a single corresponding pair P_i, Q_i.

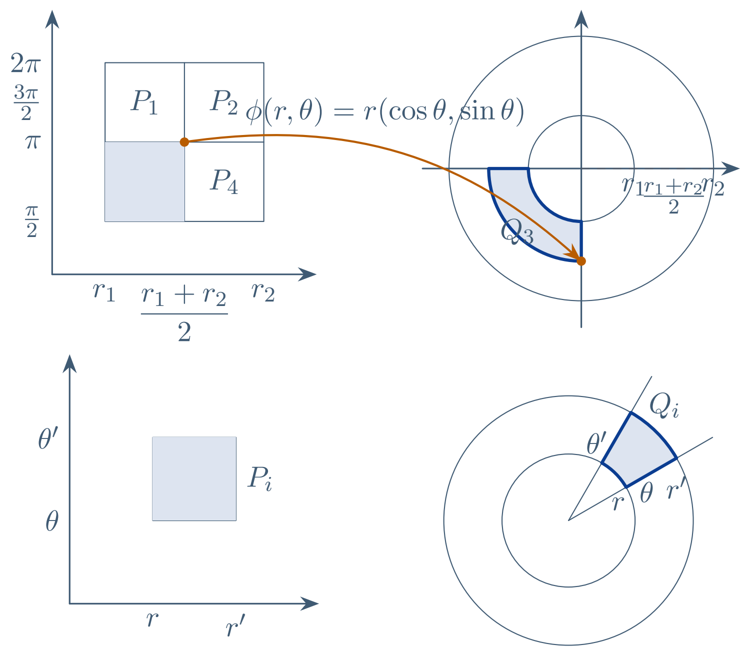

FIGURE 22

Take a division $\Delta = \{P_1, \ldots, P_l\}$ of A, and let $\Delta' = \{Q_1, \ldots, Q_l\}$ be its image in B, where $Q_i = \phi(P_i)$. Each Q_i is an annular sector. Compare the corresponding upper Darboux sums:

$$\int_A f \approx U(f, \Delta) = \sum_{i=1}^{l} \mathrm{vol}(P_i) \cdot \sup_{P_i}(f), \tag{$*$}$$

$$\int_B g \approx U(g, \Delta') = \sum_{i=1}^{l} \mathrm{vol}(Q_i) \cdot \sup_{Q_i}(g). \tag{$**$}$$

The supremum terms match well:

$$\sup_{Q_i}(g) = \sup_{P_i}(f).$$

Thus the essential point is to understand the area ratio $\frac{\mathrm{vol}(Q_i)}{\mathrm{vol}(P_i)}$.

If P_i is a very small rectangle $[r, r'] \times [\theta, \theta']$, then

$$\mathrm{vol}(P_i) = (r' - r)(\theta' - \theta),$$

while

$$\mathrm{vol}(Q_i) = \frac{(r')^2 - r^2}{2}\,(\theta' - \theta) = \frac{r' + r}{2}(r' - r)(\theta' - \theta).$$

Hence

$$\frac{\mathrm{vol}(Q_i)}{\mathrm{vol}(P_i)} = \frac{r' + r}{2} \approx r,$$

which explains the extra factor. So

$$\int_B g \approx \sum_{i=1}^{l} \mathrm{vol}(Q_i) \cdot \sup_{Q_i}(g) = \sum_{i=1}^{l} \mathrm{vol}(P_i) \cdot \frac{\mathrm{vol}(Q_i)}{\mathrm{vol}(P_i)} \cdot \sup_{P_i}(f) \approx \int_A f(r, \theta)\, r \, \mathrm{d}(r, \theta).$$

Appendix: Lecture and Tutorial Index