Module MA2321: Analysis in Several Real Variables
Michaelmas Term 2016
Section 7: Norms on Finite-Dimensional Vector Spaces

D. R. Wilkins

Contents

7 Norms on Finite-Dimensional Vector Spaces

7.1 Norms

Definition A *norm* $\|.\|$ on a real or complex vector space X is a function, associating to each element x of X a corresponding real number $\|x\|$, such that the following conditions are satisfied:—

(i) $\|x\| \geq 0$ for all $x \in X$,

(ii) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$,

(iii) $\|\lambda x\| = |\lambda| \, \|x\|$ for all $x \in X$ and for all scalars λ,

(iv) $\|x\| = 0$ if and only if $x = 0$.

A *normed vector space* $(X, \|.\|)$ consists of a a real or complex vector space X, together with a norm $\|.\|$ on X.

The Euclidean norm $|.|$ is a norm on $\mathbb{R}^n$ defined so that

$$|(x_1, x_2, \ldots, x_n)| = \sqrt{x_1^2 + x_2^2 + \cdots x_n^2}$$

for all $(x_1, x_2, \ldots, x_n)$. There are other useful norms on $\mathbb{R}^n$. These include the norms $\|.\|_1$ and $\|.\|_{\mathrm{sup}}$, where

$$\|(x_1, x_2, \ldots, x_n)\|_1 = |x_1| + |x_2| + \cdots + |x_n|$$

and

$$\|(x_1, x_2, \ldots, x_n)\|_{\mathrm{sup}} = \mathrm{maximum}(|x_1|, |x_2|, \ldots, |x_n|)$$

for all $(x_1, x_2, \ldots, x_n)$.

Definition Let $\|.\|$ and $\|.\|_*$ be norms on a real vector space X. The norms $\|.\|$ and $\|.\|_*$ are said to be *equivalent* if and only if there exist constants c and C, where $0 < c \leq C$, such that

$$c\|x\| \leq \|x\|_* \leq C\|x\|$$

for all $x \in X$.

Lemma 7.1 *If two norms on a real vector space are equivalent to a third norm then they are equivalent to each other.*

Proof let $\|.\|_*$ and $\|.\|_{**}$ be norms on a real vector space X that are both equivalent to a norm $\|.\|$ on X. Then there exist constants c_*, c_{**}, C_* and C_{**}, where $0 < c_* \leq C_*$ and $0 < c_{**} \leq C_{**}$, such that

$$c_*\|x\| \leq \|x\|_* \leq C_*\|x\|$$

and

$$c_{**}\|x\| \leq \|x\|_{**} \leq C_{**}\|x\|$$

for all $x \in X$. But then

$$\frac{c_{**}}{C_*}\|x\|_* \leq \|x\|_{**} \leq \frac{C_{**}}{c_*}\|x\|_*.$$

for all $x \in X$, and thus the norms $\|.\|_*$ and $\|.\|_{**}$ are equivalent to one another. The result follows. ∎

We shall show that all norms on a finite-dimensional real vector space are equivalent.

Lemma 7.2 *Let $\|.\|$ be a norm on $\mathbb{R}^n$. Then there exists a positive real number C with the property that $\|\mathbf{x}\| \leq C|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^n$.*

Proof Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ denote the basis of $\mathbb{R}^n$ given by

$$\mathbf{e}_1 = (1, 0, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \cdots, \quad \mathbf{e}_n = (0, 0, 0, \ldots, 1).$$

Let $\mathbf{x}$ be a point of $\mathbb{R}^n$, where

$$\mathbf{x} = (x_1, x_2, \ldots, x_n).$$

Using Schwarz's Inequality, we see that

$$\begin{aligned} \|\mathbf{x}\| &= \left\| \sum_{j=1}^n x_j \mathbf{e}_j \right\| \leq \sum_{j=1}^n |x_j| \, \|\mathbf{e}_j\| \\ &\leq \left(\sum_{j=1}^n x_j^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^n \|\mathbf{e}_j\|^2 \right)^{\frac{1}{2}} = C|\mathbf{x}|, \end{aligned}$$

where

$$C^2 = \|\mathbf{e}_1\|^2 + \|\mathbf{e}_2\|^2 + \cdots + \|\mathbf{e}_n\|^2$$

and

$$|\mathbf{x}| = \left(\sum_{j=1}^n x_j^2 \right)^{\frac{1}{2}}$$

for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. The result follows. ∎

Lemma 7.3 *Let $\|.\|$ be a norm on $\mathbb{R}^n$. Then there exists a positive constant C such that*

$$\big|\|\mathbf{x}\| - \|\mathbf{y}\|\big| \leq \|\mathbf{x} - \mathbf{y}\| \leq C|\mathbf{x} - \mathbf{y}|$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Proof Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then

$$\|\mathbf{x}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y}\|, \qquad \|\mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{x}\|.$$

It follows that

$$\|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\|$$

and

$$\|\mathbf{y}\| - \|\mathbf{x}\| \leq \|\mathbf{x} - \mathbf{y}\|,$$

and therefore

$$\Big|\|\mathbf{y}\| - \|\mathbf{x}\|\Big| \leq \|\mathbf{x} - \mathbf{y}\|$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. The result therefore follows from Lemma 7.2. ∎

Theorem 7.4 *Any two norms on $\mathbb{R}^n$ are equivalent.*

Proof Let $\|.\|$ be any norm on $\mathbb{R}^n$. We show that $\|.\|$ is equivalent to the Euclidean norm $|.|$. Let S^{n-1} denote the unit sphere in $\mathbb{R}^n$, defined by

$$S^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| = 1\}.$$

Now it follows from Lemma 7.3 that the function $\mathbf{x} \mapsto \|\mathbf{x}\|$ is continuous. Also S^{n-1} is a compact subset of $\mathbb{R}^n$, since it is both closed and bounded. It therefore follows from the Extreme Value Theorem (Theorem 5.5) that there exist points $\mathbf{u}$ and $\mathbf{v}$ of S^{n-1} such that $\|\mathbf{u}\| \leq \|\mathbf{x}\| \leq \|\mathbf{v}\|$ for all $\mathbf{x} \in S^{n-1}$. Set $c = \|\mathbf{u}\|$ and $C = \|\mathbf{v}\|$. Then $0 < c \leq C$ (since it follows from the definition of norms that the norm of any non-zero element of $\mathbb{R}^n$ is necessarily non-zero).

If $\mathbf{x}$ is any non-zero element of $\mathbb{R}^n$ then $\lambda\mathbf{x} \in S^{n-1}$, where $\lambda = 1/|\mathbf{x}|$. But $\|\lambda\mathbf{x}\| = |\lambda|\,\|\mathbf{x}\|$ (see the the definition of norms). Therefore $c \leq |\lambda|\,\|\mathbf{x}\| \leq C$, and hence $c|\mathbf{x}| \leq \|\mathbf{x}\| \leq C|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^n$, showing that the norm $\|.\|$ is equivalent to the Euclidean norm $|.|$ on $\mathbb{R}^n$. If two norms on a vector space are equivalent to a third norm, then they are equivalent to each other. It follows that any two norms on $\mathbb{R}^n$ are equivalent, as required. ∎

7.2 Linear Transformations

The space $\mathbb{R}^n$ consisting of all n-tuples $(x_1, x_2, \ldots, x_n)$ of real numbers is a vector space over the field $\mathbb{R}$ of real numbers, where addition and multiplication by scalars are defined by

$$\begin{aligned}(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n), \\ \lambda(x_1, x_2, \ldots, x_n) &= (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)\end{aligned}$$

for all $(x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

Definition A map $T \colon \mathbb{R}^m \to \mathbb{R}^n$ is said to be a *linear transformation* if

$$T(\mathbf{x} + \mathbf{y}) = T\mathbf{x} + T\mathbf{y}, \qquad T(\lambda\mathbf{x}) = \lambda T\mathbf{x}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ and $\lambda \in \mathbb{R}$.

Every linear transformation $T \colon \mathbb{R}^m \to \mathbb{R}^n$ is represented by an $n \times m$ matrix $(T_{i,j})$. Indeed let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m$ be the standard basis vectors of $\mathbb{R}^m$ defined by

$$\mathbf{e}_1 = (1, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, \ldots, 0), \ldots, \mathbf{e}_m = (0, 0, \ldots, 1).$$

Thus if $\mathbf{x} \in \mathbb{R}^m$ is represented by the m-tuple $(x_1, x_2, \ldots, x_m)$ then

$$\mathbf{x} = \sum_{j=1}^{m} x_j \mathbf{e}_j.$$

Similarly let $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ be the standard basis vectors of $\mathbb{R}^n$ defined by

$$\mathbf{f}_1 = (1, 0, \ldots, 0), \quad \mathbf{f}_2 = (0, 1, \ldots, 0), \ldots, \mathbf{f}_n = (0, 0, \ldots, 1).$$

Thus if $\mathbf{v} \in \mathbb{R}^n$ is represented by the n-tuple $(v_1, v_2, \ldots, v_n)$ then

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{f}_i.$$

Let $T \colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Define $T_{i,j}$ for all integers i between 1 and n and for all integers j between 1 and m such that

$$T\mathbf{e}_j = \sum_{i=1}^{n} T_{i,j} \mathbf{f}_i.$$

Using the linearity of T, we see that if $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ then

$$T\mathbf{x} = T\left(\sum_{j=1}^{m} x_j \mathbf{e}_j\right) = \sum_{j=1}^{m} (x_j\, T\mathbf{e}_j) = \sum_{i=1}^{n}\left(\sum_{j=1}^{m} T_{i,j} x_j\right)\mathbf{f}_i.$$

Thus the ith component of $T\mathbf{x}$ is

$$T_{i,1}x_1 + T_{i,2}x_2 + \cdots + T_{i,m}x_m.$$

Writing out this identity in matrix notation, we see that if $T\mathbf{x} = \mathbf{v}$, where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix},$$

then

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} T_{1,1} & T_{1,2} & \ldots & T_{1,m} \\ T_{2,1} & T_{2,2} & \ldots & T_{2,m} \\ \vdots & \vdots & & \vdots \\ T_{n,1} & T_{n,2} & \ldots & T_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}.$$

7.3 The Operator Norm of a Linear Transformation

Definition Given $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. The *operator norm* $\|T\|_{\mathrm{op}}$ of T is defined such that

$$\|T\|_{\mathrm{op}} = \sup\{|T\mathbf{x}| : \mathbf{x} \in \mathbb{R}^m \quad \text{and} \quad |\mathbf{x}| = 1\}.$$

Lemma 7.5 *Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ and $U\colon \mathbb{R}^m \to \mathbb{R}^n$ be linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let λ be a real number. Then $\|T\|_{\mathrm{op}}$ is the smallest non-negative real number with the property that $|T\mathbf{x}| \leq \|T\|_{\mathrm{op}}|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$. Moreover*

$$\|\lambda T\|_{\mathrm{op}} = |\lambda|\,\|T\|_{\mathrm{op}} \quad \textit{and} \quad \|T + U\|_{\mathrm{op}} \leq \|T\|_{\mathrm{op}} + \|U\|_{\mathrm{op}}.$$

Proof Let $\mathbf{x}$ be an element of $\mathbb{R}^m$. Then we can express $\mathbf{x}$ in the form $\mathbf{x} = \mu\mathbf{z}$, where $\mu = |\mathbf{x}|$ and $\mathbf{z} \in \mathbb{R}^m$ satisfies $|\mathbf{z}| = 1$. Then

$$|T\mathbf{x}| = |T(\mu\mathbf{z})| = |\mu T\mathbf{z}| = |\mu|\,|T\mathbf{z}| = |\mathbf{x}|\,|T\mathbf{z}| \leq \|T\|_{\mathrm{op}}|\mathbf{x}|.$$

Next let C be a non-negative real number with the property that $|T\mathbf{x}| \leq C|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$. Then C is an upper bound for the set

$$\{|T\mathbf{x}| : \mathbf{x} \in \mathbb{R}^m \quad \text{and} \quad |\mathbf{x}| = 1\},$$

and thus $\|T\|_{\rm op} \leq C$. Thus $\|T\|_{\rm op}$ is the smallest non-negative real number C with the property that $|T\mathbf{x}| \leq C|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$.

Next we note that

$$\begin{aligned} \|\lambda T\|_{\rm op} &= \sup\{|\lambda T\mathbf{x}| : \mathbf{x} \in \mathbb{R}^m \quad \text{and} \quad |\mathbf{x}| = 1\} \\ &= \sup\{|\lambda|\,|T\mathbf{x}| : \mathbf{x} \in \mathbb{R}^m \quad \text{and} \quad |\mathbf{x}| = 1\} \\ &= |\lambda| \sup\{|T\mathbf{x}| : \mathbf{x} \in \mathbb{R}^m \quad \text{and} \quad |\mathbf{x}| = 1\} \\ &= |\lambda|\,\|T\|_{\rm op}. \end{aligned}$$

Let $\mathbf{x} \in \mathbb{R}^m$. Then

$$\begin{aligned} |(T+U)\mathbf{x}| &\leq |T\mathbf{x}| + |U\mathbf{x}| \leq \|T\|_{\rm op}|\mathbf{x}| + \|U\|_{\rm op}|\mathbf{x}| \\ &\leq (\|T\|_{\rm op} + \|U\|_{\rm op})|\mathbf{x}| \end{aligned}$$

It follows that

$$\|(T+U)\|_{\rm op} \leq \|T\|_{\rm op} + \|U\|_{\rm op}.$$

This completes the proof. ∎

7.4 The Hilbert-Schmidt Norm of a Linear Transformation

Recall that the *length* (or *norm*) of an element $\mathbf{x} \in \mathbb{R}^n$ is defined such that

$$|\mathbf{x}|^2 = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Definition Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let $(T_{i,j})$ be the $n \times m$ matrix representing this linear transformation with respect to the standard bases of $\mathbb{R}^m$ and $\mathbb{R}^n$. The *Hilbert-Schmidt norm* $\|T\|_{\rm HS}$ of the linear transformation is then defined so that

$$\|T\|_{\rm HS} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} T_{i,j}^2}.$$

Note that the Hilbert-Schmidt norm is just the Euclidean norm on the real vector space of dimension mn whose elements are $n \times m$ matrices representing linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$ with respect to the standard bases of these vector spaces. Therefore it has the standard properties of the Euclidean norm. In particular it follows from the Triangle Inequality (Lemma 4.2) that

$$\|T+U\|_{\rm HS} \leq \|T\|_{\rm HS} + \|U\|_{\rm HS} \quad \text{and} \quad \|\lambda T\|_{\rm HS} = |\lambda|\,\|T\|_{\rm HS}$$

for all linear transformations T and U from $\mathbb{R}^m$ to $\mathbb{R}^n$ and for all real numbers λ.

Lemma 7.6 *Let $T: \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. Then T is uniformly continuous on $\mathbb{R}^n$. Moreover*

$$|T\mathbf{x} - T\mathbf{y}| \leq \|T\|_{\mathrm{HS}} |\mathbf{x} - \mathbf{y}|$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$, where $\|T\|_{HS}$ is the Hilbert-Schmidt norm of the linear transformation T.

Proof Let $\mathbf{v} = T\mathbf{x} - T\mathbf{y}$, where $\mathbf{v} \in \mathbb{R}^n$ is represented by the n-tuple $(v_1, v_2, \ldots, v_n)$. Then

$$v_i = T_{i,1}(x_1 - y_1) + T_{i,2}(x_2 - y_2) + \cdots + T_{i,m}(x_m - y_m)$$

for all integers i between 1 and n. It follows from Schwarz's Inequality (Lemma 4.1) that

$$v_i^2 \leq \left(\sum_{j=1}^m T_{i,j}^2 \right) \left(\sum_{j=1}^m (x_j - y_j)^2 \right) = \left(\sum_{j=1}^m T_{i,j}^2 \right) |\mathbf{x} - \mathbf{y}|^2.$$

Hence

$$|\mathbf{v}|^2 = \sum_{i=1}^n v_i^2 \leq \left(\sum_{i=1}^n \sum_{j=1}^m T_{i,j}^2 \right) |\mathbf{x} - \mathbf{y}|^2 = \|T\|_{\mathrm{HS}}^2 |\mathbf{x} - \mathbf{y}|^2.$$

Thus $|T\mathbf{x} - T\mathbf{y}| \leq \|T\|_{\mathrm{HS}} |\mathbf{x} - \mathbf{y}|$. It follows from this that T is uniformly continuous. Indeed let some positive real number ε be given. We can then choose δ so that $\|T\|_{\mathrm{HS}}\, \delta < \varepsilon$. If $\mathbf{x}$ and $\mathbf{y}$ are elements of $\mathbb{R}^n$ which satisfy the condition $|\mathbf{x} - \mathbf{y}| < \delta$ then $|T\mathbf{x} - T\mathbf{y}| < \varepsilon$. This shows that $T: \mathbb{R}^m \to \mathbb{R}^n$ is uniformly continuous on $\mathbb{R}^m$, as required. ∎

Lemma 7.7 *Let $T: \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ and let $S: \mathbb{R}^n \to \mathbb{R}^p$ be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^p$. Then the Hilbert-Schmidt norm of the composition of the linear operators T and S satisfies the inequality $\|ST\|_{\mathrm{HS}} \leq \|S\|_{\mathrm{HS}}\, \|T\|_{\mathrm{HS}}$.*

Proof The composition ST of the linear operators is represented by the product of the corresponding matrices. Thus the component $(ST)_{k,j}$ in the kth row and the jth column of the $p \times m$ matrix representing the linear transformation ST satisfies

$$(ST)_{k,j} = \sum_{i=1}^n S_{k,i} T_{i,j}.$$

It follows from Schwarz's Inequality (Lemma 4.1) that

$$(ST)_{k,j}^2 \leq \left(\sum_{i=1}^n S_{k,i}^2 \right) \left(\sum_{i=1}^n T_{i,j}^2 \right).$$

Summing over k, we find that

$$\sum_{k=1}^p (ST)_{k,j}^2 \leq \left(\sum_{k=1}^p \sum_{i=1}^n S_{k,i}^2 \right) \left(\sum_{i=1}^n T_{i,j}^2 \right) = \|S\|_{\mathrm{HS}}^2 \left(\sum_{i=1}^n T_{i,j}^2 \right).$$

Then summing over j, we find that

$$\begin{aligned} \|ST\|_{\mathrm{HS}}^2 &= \sum_{k=1}^p \sum_{j=1}^m (ST)_{k,j}^2 \leq \|S\|_{\mathrm{HS}}^2 \left(\sum_{i=1}^n \sum_{j=1}^m T_{i,j}^2 \right) \\ &\leq \|S\|_{\mathrm{HS}}^2 \|T\|_{\mathrm{HS}}^2. \end{aligned}$$

On taking square roots, we find that $\|ST\|_{\mathrm{HS}} \leq \|S\|_{\mathrm{HS}} \, \|T\|_{\mathrm{HS}}$, as required. ∎