MA2321—Analysis in Several Variables School of Mathematics, Trinity College Michaelmas Term 2017 Lecture 18 (November 16, 2017)

David R. Wilkins

8. Norms of Linear Transformations

8.1. 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

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

8.2. 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 8.1

Let $T: \mathbb{R}^m \to \mathbb{R}^n$ and $U: \mathbb{R}^m \to \mathbb{R}^n$ 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 $\|T\|_{\mathrm{op}} \leq C$. Thus $\|T\|_{\mathrm{op}}$ 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

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

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

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

It follows that

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

This completes the proof. $\blacksquare$

8.3. 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* $\|T\|_{\mathrm{HS}}$ of the linear transformation is then defined so that

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

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 2.2) that

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

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

Lemma 8.2

Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ 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 2.1) 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 $T\colon \mathbb{R}^m \to \mathbb{R}^n$ is uniformly continuous on $\mathbb{R}^m$, as required. ∎

Lemma 8.3

Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ and let $S\colon \mathbb{R}^n \to \mathbb{R}^p$ 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

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

where $S_{k,i}$ and $T_{i,j}$ denote the components in the relevant rows and columns of the matrices representing the linear transformations S and T respectively.

It follows from Schwarz's Inequality (Lemma 2.1) 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. ∎

Proposition 8.4

Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. Then the operator norm and the Hilbert-Schmidt norm of the linear operator T satisfies the inequalities linear operators T and S satisfies the inequality

$$\|T\|_{\mathrm{op}} \leq \|T\|_{\mathrm{HS}} \leq \sqrt{n}\,\|T\|_{\mathrm{op}}.$$

Proof

It follows from Lemma 8.2 that that $|T\mathbf{x}| \leq \|T\|_{\mathrm{HS}}|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$. Now the operator norm $\|T\|_{\mathrm{op}}$ of T is by definition the smallest non-negative real number with the property that $|T\mathbf{x}| \leq \|T\|_{\mathrm{op}}|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$. It follows that $\|T\|_{\mathrm{op}} \leq \|T\|_{\mathrm{HS}}$.

We denote by $T_{i,j}$ the coefficient in the ith row and jth column of the matrix representing the linear transformation T for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.
Let i be an integer between 1 and n, and let

$$\mathbf{x} = (T_{i,1}, T_{i,2}, \ldots, T_{i,m}).$$

Then the ith component $(T\mathbf{x})_i$ of the vector $T\mathbf{x}$ satisfies the equation

$$(T\mathbf{x})_i = \sum_{j=1}^{m} T_{i,j}^2.$$

It follows that

$$\sum_{j=1}^{m} T_{i,j}^2 \leq |T\mathbf{x}| \leq \|T\|_{\mathrm{op}} |\mathbf{x}| = \|T\|_{\mathrm{op}} \sqrt{\sum_{j=1}^{m} T_{i,j}^2}.$$

Therefore

$$\sum_{j=1}^{m} T_{i,j}^2 \leq \|T\|_{\mathrm{op}}^2,$$

and hence

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

The result follows. ∎

The Hilbert-Schmidt norm on the real vector space $L(\mathbb{R}^m, \mathbb{R}^n)$ of linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$ is just the Euclidean norm on that vector space obtained on identifying that vector space with $\mathbb{R}^{mn}$ by means of its natural basis. The definition of convergence in a Euclidean space then ensures that an infinite sequence $T_1, T_2, T_3, \ldots,$ of linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$ converges to some linear transformation T if and only if, given any positive real number η, there exists some positive integer N such that $\|T_j - T\|_{\mathrm{HS}} < \eta$ whenever $j \geq N$. Similarly a subset V of $L(\mathbb{R}^m, \mathbb{R}^n)$ is open in $L(\mathbb{R}^m, \mathbb{R}^n)$ if and only if, given any linear transformation $S\colon \mathbb{R}^m \to \mathbb{R}^n$ that belongs to V, there exists some strictly positive real number η such that

$$\{T \in L(\mathbb{R}^m, \mathbb{R}^n) : \|T - S\|_{\mathrm{HS}} < \eta\} \subset V.$$

Lemma 8.5

Let $T_1, T_2, T_3, \ldots$ be an infinite sequence of linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let T be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. Then the infinite sequence $T_1, T_2, T_3, \ldots$ of linear transformations converges to the linear transformation T if and only if, given any positive real number ε, there exists some positive integer N such that positive integer N such that $\|T_j - T\|_{\mathrm{op}} < \varepsilon$ whenever $j \geq N$.

Proof

First suppose that the infinite sequence $T_1, T_2, T_3, \ldots$ converges to T. Then, given any positive real number ε, there exists some positive integer N such that $\|T_j - T\|_{\mathrm{HS}} < \varepsilon$ whenever $j \geq N$. It then follows from Proposition 8.4 that

$$\|T_j - T\|_{\mathrm{op}} \leq \|T_j - T\|_{\mathrm{HS}} < \varepsilon$$

whenever $j \geq N$.

Conversely, suppose that, given any positive real number η, there exists some positive integer N such that $\|T_j - T\|_{\mathrm{op}} < \eta$ whenever $j \geq N$. Let some positive real number ε be given. Then there exists some positive integer N such that $\|T_j - \|_{\mathrm{op}} < \varepsilon/\sqrt{n}$ whenever $j \geq N$. It then follows from Proposition 8.4 that

$$\|T_j - T\|_{\mathrm{HS}} \leq \sqrt{n}\,\|T_j - T\|_{\mathrm{op}} < \varepsilon$$

whenever $j \geq N$, and thus T_j converges to T as $j \to +\infty$. ∎

Lemma 8.6

Let V be a subset of the set $L(\mathbb{R}^m, \mathbb{R}^n)$ of linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$. Then V is open in $L(\mathbb{R}^m, \mathbb{R}^n)$ if and only if, given any element S of V, there exists some positive real number ε such that

$$\{T \in L(\mathbb{R}^m, \mathbb{R}^n) : \|T - S\|_{\mathrm{op}} < \varepsilon\} \subset V$$

Proof

The set V is open in $L(\mathbb{R}^m, \mathbb{R}^n)$ if and only if, given any element S of V, there exists some positive real number η such that

$$\{T \in L(\mathbb{R}^m, \mathbb{R}^n) : \|T - S\|_{\mathrm{HS}} < \eta\} \subset V.$$

Let some positive real number ε be given. It follows from Proposition 8.4 that if the set V contains all linear transformations T from $\mathbb{R}^m$ to $\mathbb{R}^n$ that satisfy $\|T - S\|_{\mathrm{op}} < \varepsilon$ then it contains all linear transformations T that satisfy $\|T - S\|_{\mathrm{HS}} < \varepsilon$, because $\|T - S\|_{\mathrm{op}} \leq \|T - S\|_{\mathrm{HS}}$. In the other direction, if the set V contains all linear transformations that satisfy $\|T - S\|_{\mathrm{HS}} < \varepsilon$ then it contains all linear transformations T that satisfy $\|T - S\|_{\mathrm{op}} < \varepsilon/\sqrt{n}$, because $\|T - S\|_{\mathrm{HS}} \leq \sqrt{n}\,\|T - S\|_{\mathrm{op}}$. The result follows. ∎