Mathematics 522

Bounded Linear Operators and the Definition of Derivatives

Definition. Let V, W be normed vector spaces (both over $\mathbb{R}$ or over $\mathbb{C}$). A *linear transformation* or *linear operator* $T : V \to W$ is *bounded* if there is a constant C such that

$$\|Tx\|_W \le C\|x\|_V \quad \text{for all } x \in V. \tag{1}$$

Remark: We use the linearity of T and the homogeneity of the norm in W to see that

$$\Big\|T\big(\frac{x}{\|x\|_V}\big)\Big\|_W = \Big\|\frac{T(x)}{\|x\|_V}\Big\|_W = \frac{\|T(x)\|_W}{\|x\|_V}$$

we see that T is bounded, satisfying (1), if and only if

$$\sup_{\|x\|_V=1} \|T(x)\|_W \le C.$$

Theorem. *Let V, W be normed vector spaces and let $T : V \to W$ be a linear transformation. The following statements are equivalent.*
(i) T is a bounded linear transformation.
(ii) T is continuous everwhere in V.
(iii) T is continuous at 0 in V.

Proof. (i) $\Longrightarrow$ (ii). Let C as in the definition of bounded linear transformation. By linearity of T we have

$$\|T(v) - T(\tilde{v})\|_W = \|T(v - \tilde{v})\|_W \le C\|v - \tilde{v}\|_V$$

which implies (ii).
(ii) $\Longrightarrow$ (iii) is trivial.
(iii) $\Longrightarrow$ (i): If T is continuous at 0 there exists $\delta > 0$ such that for all $v \in V$ with $\|v\| < \delta$ we have $\|Tv\| < 1$. Now let $x \in V$ and $x \neq 0$. Then

$$\Big\|\delta\frac{x}{2\|x\|_V}\Big\|_V = \delta/2 \text{ and thus } \Big\|T\big(\delta\frac{x}{\|x\|_V}\big)\Big\|_W < 1.$$

But by the linearity of T and the homogeneity of the norm we get

$$1 \ge \Big\|T\big(\delta\frac{x}{\|x\|_V}\big)\Big\|_W = \Big\|\delta\frac{T(x)}{2\|x\|_V}\Big\|_W = \frac{\delta}{2\|x\|_V}\|Tx\|_W$$

and therefore $\|Tx\|_W \le C\|x\|_V$ with $C = 2/\delta$. □

Notation: If $T : V \to W$ is linear one often writes Tx for $T(x)$.

Definition. We denote by $L(V, W)$ the set of all bounded linear transformations $T : V \to W$.

$L(V,W)$ form a vector space. $S+T$ is the transformation with $(S+T)(x) = S(x)+T(x)$ and cT is the operator $x \mapsto cT(x)$. On $L(V,W)$ we define the *operator norm* (depending on the norms on V and W) by

$$\|T\|_{L(V,W)} \equiv \|T\|_{op} = \sup_{v\neq 0} \frac{\|Tv\|_W}{\|v\|_V}.$$

We can think of $\|T\|_{L(V,W)}$ as the best constant for which (1) holds. Note that

$$\|Tx\|_W \le \|T\|_{L(V,W)} \|x\|_V \,.$$

Using the homogeneity of the W-norm we also can write

$$\|T\|_{L(V,W)} = \sup_{\|x\|_V=1} \|Tx\|_W.$$

We use the $\|\cdot\|_{op}$ notation if the choice of V, W and the norms are clear from the context. In the textbook, Rudin considers $V=\mathbb{R}^n$, $W=\mathbb{R}^m$ with the standard Euclidean norms and simply writes $\|T\|$ for the operator norm.

Lemma. *Let V and W be normed spaces. If V is finite dimensional then all linear transformations from V to W are bounded.*

Proof. Let $v_1,\dots,v_n$ be a basis of V. Then for $v=\sum_{j=1}^n \alpha_j v_j$ we have

$$\|Tv\|_W = \Big\| \sum_{j=1}^n \alpha_j T v_j \Big\|_W \le \sum_{j=1}^n |\alpha_j| \|Tv_j\|_W \le \sum_{j=1}^n \|Tv_j\|_W \max_{k=1,\dots,n} |\alpha_k|.$$

The expression $\max_{k=1,\dots,n} |\alpha_k|$ defines a norm on V. Since all norms on V are equivalent, there is a constant C_1 such that

$$\max_{j=1,\dots,n} |\alpha_j| \le C \Big\| \sum_{j=1}^n \alpha_j v_j \Big\|_V$$

for all choices of $\alpha_1,\dots,\alpha_n$. Thus we get $\|Tv\|_W \le C\|v\|_V$ for all $v\in V$, where the constant C is given by $C = C_1 \sum_{j=1}^n \|Tv_j\|$. □

Lemma. *On $\mathbb{R}^n$, $\mathbb{R}^m$ use the Euclidean norms $\|x\|_2 = (\sum_{j=1}^n |x_j|^2)^{1/2}$, $\|y\|_2 = (\sum_{i=1}^m |y_i|^2)^{1/2}$. Let A be an $m\times n$ matrix and consider the linear operator $T:\mathbb{R}^n \to \mathbb{R}^m$ defined by $T(x) = Ax$. Let*

$$\|A\|_{HS} := (\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2)^{1/2}.$$

Then

$$\|T\|_{op} \le \|A\|_{HS}.$$

Proof. $\|Ax\|_2^2 = \sum_{i=1}^m |\sum_{j=1}^n a_{ij}x_j|^2$. By the Cauchy-Schwarz inequality $|\sum_{j=1}^n a_{ij}x_j|^2 \le \sum_{j=1}^n |a_{ij}|^2 \|x\|_2^2$ and hence $\|Ax\|_2 \le \|A\|_{HS}\|x\|_2$ for all $x\in\mathbb{R}^n$. Thus $\|T\|_{op} \le \|A\|_{HS}$. □

Remark. We often write A for both the matrix and the linear operator $x \to Ax$. The expression $\|A\|_{HS}$ is a norm on the space of $m \times n$ matrices called the Hilbert-Schmidt norm of A. In many cases it is substantially larger then the operator norm (and so the estimate in the Lemma is rather inefficient). To illustrate this let D an $n \times n$ diagonal matrix (with entries λ_1,...,λ_n on the diagonal and zeroes elsewhere. Then it is easy to verify the inequality

$$\|Dx\|_2 \le \max_{j=1,\dots,n} |\lambda_j| \|x\|_2$$

and this is seen to be sharp by testing on the unit vectors $x = e_j$. Thus $\|D\|_{op} = \max_{j=1,\dots,n} |\lambda_j|$ while $\|D\|_{HS} = (\sum_{j=1}^n |\lambda_j|^2)^{1/2}$. In particular for the identity operator I we have $\|I\|_{op} = 1$ but $\|I\|_{HS} = \sqrt{n}$.

Normed vector spaces and compleness. Let V be a normed space. Then V is a metric space with respect to the metric $d(x, y) = \|x-y\|$. Recall that a sequence $\{x_n\}_{n=1}^\infty$ is a Cauchy-sequence in V if for every $\varepsilon > 0$ there is K such that $\|x_n - x_m\|_V < \varepsilon$ for $m, n \ge K$.

Definition. A *Banach space* is a normed space which is complete (i.e. every Cauchy sequence converges).

Lemma: *A finite dimensional normed space over* $\mathbb{R}$ *or* $\mathbb{C}$ *is complete.*

Proof. Let $v_1, \dots, v_N$ be a basis of V and for $x \in V$ let $\beta_1(x), \dots, \beta_N(x)$ be the coordinates of f with respect to the basis $v_1, \dots, v_N$.

We need to show that $\{x_n\}$ has a limit in V. For every n we can write

$$x_n = \sum_{i=1}^N \beta_i(x_n) v_i.$$

The expression

$$\|x\|_\infty := \max_{i=1,\dots,N} |\beta_i(x)|$$

defines a norm on V which is equivalent to the given norm. In particular we have

$$|\beta_i(x_n) - \beta_i(x_m)| \le C\|x_n - x_m\|_V, \quad i = 1, \dots, N\,.$$

This shows that for each $i = 1, \dots, N$ the numbers $\beta_i(x_n)$ form a Cauchy-sequence of scalars and thus converge to a scalar β_i. Define $x = \sum_{i=1}^N \beta_i v_i$ (so that $\beta_i = \beta_i(x)$). Then

$$\|x_n - x\|_V = \|\sum_{i=1}^N \big(\beta_i(x_n) - \beta_i\big) v_i\|_V \le \sum_{i=1}^n |\beta_i(x_n) - \beta_i| \|v_i\|_V$$

which converges to 0 as $n \to \infty$. Hence $x_n \to x$ in V. □

Theorem: *Let V and W be normed spaces and assume that W is complete. Then the space $L(V, W)$ of bounded linear operators from V to W is a Banach space.*

Proof. We need to show that $L(V, W)$ is complete. Let T_n be a Cauchy sequence in $L(V, W)$ (with respect to the norm $\|T\|_{op} = \sup_{\|x\|_V=1} \|Tx\|_W$). Then for each $x \in V$, $T_n x$ is Cauchy in W and by assumption $T_n x$ converge to some vector Tx. Check, using limiting arguments, that $T : V \to W$is linear. Since T_n is Cauchy the sequence T_N is bounded in $L(V, W)$ and thus there exists M with $\|T_n\|_{op} \le M$.

We need to show that T is a bounded operator. We have for all n and all $x \in V$

$$\|Tx\|_W \le \|Tx - T_n x\|_W + \|T_n x\|_W \le \|Tx - T_n x\|_W + \|T_n\|_{op} \|x\|_W .$$

Letting $n \to \infty$ we see that $\|Tx\|_W \le M|x\|_V$. Hence $T \in L(V, W)$ with $\|T\|_{op} \le M$.

Finally we need to show that $T_n \to T$ in $L(V, W)$. Given $\varepsilon > 0$ we choose N such that $\|T_n - T_m\|_{op} < \varepsilon/2$ for $m, n \ge N$.

We prove the following *Claim:*

For each $x \in V$, $x \neq 0$, and $n \ge N$ we have $\|T_n x - Tx\|_W < \varepsilon \|x\|_V$.

Let $n \ge N$. Since $T_m x \to Tx$ in W we can choose $m(x) \ge N$ such that $\|T_{m(x)} x - Tx\|_W \le \varepsilon/4 \|x\|_V$. We estimate we estimate (with $m \ge N$)

$$\begin{aligned}
\|Tx - T_n x\|_W &\le \|Tx - T_{m(x)} x + T_{m(x)} x - T_n(x)\|_W \\
&\le \|Tx - T_{m(x)} x\|_W + \|T_{m(x)} x - T_n x\|_W \\
&\le \frac{\varepsilon}{4} \|x\|_V + \|T_{m(x)} - T_n\|_{op} \|x\|_V \\
&\le \frac{\varepsilon}{4} \|x\|_V + \frac{\varepsilon}{2} \|x\|_V < \varepsilon \|x\|_V .
\end{aligned}$$

Hence the claim is established and thus $\|T_n - T\|_{op} < \varepsilon$ for $n \ge N$. This shows $T_n \to T$ with respect to the operator norm. □

Definition. (i) Let V be a normed vector space over $\mathbb{R}$. Bounded linear transformations from V to $\mathbb{R}$ are called bounded linear functionals. The space $L(V, \mathbb{R})$ with the operator norm is called the dual space to V, or V^*.

(ii) Similarly if V be a normed vector space over $\mathbb{C}$ we call the bounded linear transformations from V to $\mathbb{R}$ bounded linear functionals and refer to the space $L(V, \mathbb{C})$ with the operator norm as the dual space V^*.

Since $\mathbb{R}$ and $\mathbb{C}$ are complete the above theorem about completeness of $L(V, W)$ immediately yields the

Corollary. *Dual spaces of normed vector spaces are Banach spaces.*

Exercise: The space $L(\mathbb{R}^n, \mathbb{R})$ of linear functionals $\ell : \mathbb{R}^n \to \mathbb{R}$ can be identified with $\mathbb{R}$, since for any such ℓ there is a unique vector $y \in \mathbb{R}^n$ such that $\ell(x) = \sum_{j=1}^n x_j y_j$ for all $x \in \mathbb{R}^n$. We consider the p-norm on $\mathbb{R}^n$, $\|x\|_p = (\sum_{j=1}^n |x_j|^p)^{1/p}$, $1 \le p < \infty$. Also set $\|x\|_\infty = \max_{j=1,\dots,n} |x_j|$. The norm on $\mathbb{R}$ is just the absolute value $|\cdot|$.

(i) Prove

$$\sup_{\|x\|_1 \neq 0} \frac{\left|\sum_{j=1}^n x_j y_j\right|}{\|x\|_1} = \|y\|_\infty$$

and

$$\sup_{\|x\|_\infty \neq 0} \frac{\left|\sum_{j=1}^n x_j y_j\right|}{\|x\|_\infty} = \|y\|_1 .$$

(ii) Prove that if $1 < p < \infty$ then

$$\|\ell\|_{op} \equiv \sup_{\|x\|_p \neq 0} \frac{\left|\sum_{j=1}^n x_j y_j\right|}{\|x\|_p} = \|y\|_{p'} \text{ where } p' = \frac{p}{p-1} .$$

Hint: Hölder's inequality should be used here.

Finally we consider compositions of linear transformations. The operator norms are submultiplicative in the sense of the following lemma.

Lemma. *Let V_1, V_2, V_3 be nomed vector spaces and let $T \in L(V_1, V_2)$ and $S \in L(V_2, V_3)$. Define the composition $ST : V_1 \to V_3$ by $ST(x) = S(T(x))$. Then $ST \in L(V_1, V_3)$ and we have*

$$\|ST\|_{L(V_1,V_3)} \le \|S\|_{L(V_2,V_3)} \|T\|_{L(V_1,V_2)} .$$

Proof: For $x \in V_1$,

$$\|ST(x)\|_{V_3} \le \|S\|_{L(V_2,V_3)} \|T(x)\|_{V_2} \le \|S\|_{L(V_2,V_3)} \|T\|_{L(V_1,V_2)} \|x\|_{V_1} .$$

This implies $ST \in L(V_1, V_3)$ and the asserted submultiplicativity property. □

Differentiation

Definition. Let V and W be normed vector spaces. Let $U \subset V$ be an open subset of V and let $F : U \to W$ be a function. Let $a \in U$. We say that F is differentiable at a if there exists a bounded linear transformation $T : V \to W$ (usually denoted by Df_a or by $f'(a)$) such that

$$\lim_{h \to 0} \frac{\|f(a+h) - f(a) - Th\|_W}{\|h\|_V} = 0$$

Uniqueness: For the derivative of f at a to be well defined we must prove uniqueness of T. Suppose there are two linear bounded transformations T and $\tilde{T}$ with $\frac{\|f(a+h)-f(a)-Th\|_W}{\|h\|_V} \to 0$ and $\frac{\|f(a+h)-f(a)-\tilde{T}h\|_W}{\|h\|_V} \to 0$ as $\|h\|_V \to$

0. The triangle inequality implies that $\lim_{h\to 0}\frac{\|Th-\tilde{T}h\|_W}{\|h\|_V}=0$. I.e. for every $\varepsilon>0$ there is a $\delta>0$ such that $\|Th-\tilde{T}h\|_W\leq\varepsilon\|h\|_V$ for all $\|h\|\leq\delta$. However since $T,\tilde{T}$ are linear this inequality also holds with h replaced by ch for all scalars c, and thus we have

$$\|Th-\tilde{T}h\|_W\leq\varepsilon\|h\|_V \text{ for } all\ h\in V.$$

Hence $\|T-\tilde{T}\|_{op}\leq\varepsilon$ for all $\varepsilon>0$ which implies $T=\tilde{T}$. Thus if the derivative of f at a exists it is uniquely defined and we shall henceforth denote it by Df_a or by $f'(a)$.

Remark: Other terns for this derivative Df_a are "total derivative of f at a" or "Fréchet derivative of f at a".

Exercise: Prove that if f is differentiable at a then f is continuous at a.

Example: Let $M(n,n)$ be the space of $n\times n$ matrices, with any norm. Define $F(A)=A^2$. Then F is differentiable at any A and its derivative DF_A is given by

$$DF_A(H)=AH+HA\,.$$

For the proof observe that $F(A+H)-F(A)=(A+H)^2-A^2=AH+HA+H^2$ and check that $\lim_{H\to 0}\|H^2\|/\|H\|=0$. This is accomplished by showing that $\|H^2\|\leq C\|H\|_2$ and if one uses a submultiplicative norm on $M(n,n)$ one gets this even with the constant 1.

Example: Let $F:\mathbb{R}^2\to\mathbb{R}$ be given by $F(x_1,x_2)=x_1^2+\sin(\pi x_2)$. You are being asked to check from the definition whether F differentiable at $(2,1)$ and to determine the derivative $DF_{(2,1)}$, as a linear transformation $\mathbb{R}^2\to\mathbb{R}$. Observe that

$$\begin{aligned}F(2+h_1,1+h_2)-F(2,1)&=(2+h_1)^2-2^2+(\sin(\pi(1+h_2))-\sin\pi\\&=4h_1+\pi\cos(\pi)h_2+O(h_1^2)+O(h_2^2).\end{aligned}$$

Hence $DF_{(2,1)}:\mathbb{R}^2\to\mathbb{R}$ is the linear transformation given by

$$DF_{(2,1)}(h)=4h_1-\pi h_2\,.$$

Example: Let $C([0,1]$ be the space of continuous functions with the usual max-norm $\|f\|_\infty=\max_{x\in[0,1]}|f(x)|$. We define a function

$$\Gamma:C([0,1])\to C([0,1])$$

by saying that for $f\in C([0,1])$, $\Gamma[f]$ is the function whose values at $x\in[0,1]$ are given by

$$\Gamma[f](x)=\int_0^x f(t)^2\cos t\,dt\,.$$

Show that Γ is differentiable at every $g\in C([0,1])$ and determine $D\Gamma_g$ for all $g\in C([0,1])$.

For $h \in C([0,1])$ we compute

$$\begin{aligned}\Gamma[g+h](x) - \Gamma[g](x) &= \int_0^x (g(t)+h(t))^2 \cos t \, dt - \int_0^x (g(t))^2 \cos t \, dt \\ &= \int_0^x h(t)\, 2g(t) \cos t \, dt + \int_0^x (h(t))^2 \cos t \, dt\end{aligned}$$

We claim that $D\Gamma_g : C([0,1]) \to C([0,1])$ is the linear transformation which to each $h \in C([0,1])$ assigns the function $T[h]$ whose values at x are given by

$$T[h](x) = \int_0^x h(t)\, 2g(t) \cos t \, dt \,.$$

We must show that T is a bounded linear transformation and that it is really the derivative of Γ at g. Clearly T is linear and maps $C([0,1])$ to itself.

It is bounded since

$$\max_{0\le x\le 1} \left| \int_0^x h(t)\, 2g(t) \cos t \, dt \right| \le 2 \max_{[0,1]} |g(x)| \max_{x\in[0,1]} |h(x)|;$$

this inequality shows that the operator norm of T is at most $2\|g\|_\infty$.

To verify that T is really the derivative of Γ at g we must analyze the error term and show that

$$\frac{\max_{x\in[0,1]} \left| \int_0^x h(t)^2 \cos t \, dt \right|}{\|h\|_\infty} \to 0 \text{ as } \|h\|_\infty \to 0.$$

The numerator is bounded by $\max_{t\in[0,1]} |h(t)^2| = \|h\|_\infty^2$ and hence the displayed expression tends to 0 as $\|h\|_\infty \to 0$. Thus we have shown that Γ is differentiable at g and $D\Gamma_g = T$.

Example: Let $C([0,1]$ be the space of continuous functions with the usual max-norm $\|f\|_\infty = \max_{x\in[0,1]} |f(x)|$. Define $\Lambda : C([0,1]) \to \mathbb{R}$ by

$$\Lambda[f] = \int_0^1 f(t)^2 \cos t \, dt \,.$$

Show that Λ is differentiable at every $g \in C([0,1])$ and that

$$D\Lambda_g : C([0,1]) \to \mathbb{R}$$

is the bounded linear functional defined by

$$D\Lambda_g[h] = \int_0^1 h(t)\, 2g(t) \cos t \, dt \,.$$