Norms on Operators

If V, W are vector spaces then so is the space of linear transformations from V to W denoted $\mathcal{L}(V,W)$. We now consider norms on $\mathcal{L}(V,W)$. When $V = W$, $\mathcal{L}(V,V) = \mathcal{L}(V)$ is an algebra with composition as multiplication; norms on $\mathcal{L}(V)$ which have a relationship to composition are particularly useful. A norm on $\mathcal{L}(V)$ is said to be *submultiplicative* if $\|A \circ B\| \leq \|A\| \cdot \|B\|$. (H-J calls this a matrix norm in finite dimensions.)

Example. For $A \in \mathbb{C}^{n\times n}$, define $\|A\| = \sup_{1\leq i,j\leq n} |a_{ij}|$. This norm is *not* submultiplicative: if $A = B = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$, then $\|A\| = \|B\| = 1$, but $AB = A^2 = nA$ so $\|AB\| = n$.

Exercise. Show that although the norm $\|A\| = \sup_{1\leq i,j\leq n} |a_{ij}|$ on $\mathbb{C}^{n\times n}$ is not submultiplicative, the norm $A \mapsto n \sup_{1\leq i,j\leq n} |a_{ij}|$ is submultiplicative.

Bounded Linear Operators and Operator Norms

Let $(V, \|\cdot\|_V)$ and $(W, \|\cdot\|_W)$ be normed linear spaces. An $L \in \mathcal{L}(V,W)$ is called a *bounded linear operator* if $\sup_{\|v\|_V=1} \|Lv\|_W < \infty$. Let $\mathcal{B}(V,W)$ denote the set of all bounded linear operators from V to W. In the special case $W = \mathbb{F}$ we have *bounded linear functionals*, and we set $V^* = \mathcal{B}(V,\mathbb{F})$. If $\dim V < \infty$, then $\mathcal{L}(V,W) = \mathcal{B}(V,W)$, so also $V^* = V'$. In fact, if we choose a basis $\{v_1,\ldots,v_n\}$ for V and let $\{f_1,\ldots,f_n\}$ be the dual basis, then $\sum_{i=1}^n |f_i(v)|$ is a norm on V (see exercise below), so by the Norm Equivalence Theorem, $\exists\, M \ni \sum_{i=1}^n |f_i(v)| \leq M\|v\|_V$; then

$$\begin{aligned}
\|Lv\|_W &= \left\| L\left(\sum_{i=1}^n f_i(v)v_i\right)\right\|_W \\
&\leq \sum_{i=1}^n |f_i(v)| \cdot \|Lv_i\|_W \\
&\leq \left(\max_{1\leq i\leq n} \|Lv_i\|_W\right) \sum_{i=1}^n |f_i(v)| \\
&\leq \left(\max_{1\leq i\leq n} \|Lv_i\|_W\right) M\|v\|_V,
\end{aligned}$$

so

$$\sup_{v\neq 0} \left(\|Lv\|_W / \|v\|_V\right) \leq \left(\max_{1\leq i\leq n} \|Lv_i\|_W\right) \cdot M < \infty.$$

(Recall that if $v = \sum_{i=1}^n x_i v_i$, then $x_i = f_i(v)$.)

Caution. A bounded linear operator doesn't necessarily have $\{\|Lv\|_W : v \in V\}$ being a bounded set of $\mathbb{R}$: in fact, if it is, then $L \equiv 0$. Similarly, if a linear functional is a bounded linear functional, it does *not* mean that there is an M for which $(\forall\, v \in V)\ |f(v)| \leq M$.

Exercise.

(1) Suppose V is a finite dimensional vector space and let $\{v_1, \ldots, v_n\}$ be a basis for V with associated dual basis $\{f_1, \ldots, f_n\}$. Show that the mapping $v \mapsto \sum_{i=1}^n |f_i(v)|$ defines a norm on V.

(2) Let $L \in \mathcal{L}(V, W)$ and show that $\sup_{\|v\|_V=1} \|Lv\|_W = \sup_{\|v\|_V \leq 1} \|Lv\|_W = \sup_{v \neq 0}(\|Lv\|_W / \|v\|_V)$.

Examples.

(1) Let $V = \mathcal{P}$ be the space of polynomials with norm $\|p\| = \sup_{0 \leq x \leq 1} |p(x)|$. The differentiation operator $\frac{d}{dx} : \mathcal{P} \to \mathcal{P}$ is not a bounded linear operator: $\|x^n\| = 1$ for all $n \geq 1$; but $\left\|\frac{d}{dx} x^n\right\| = \|nx^{n-1}\| = n$.

(2) Let $V = \mathbb{F}_0^\infty$ with ℓ^p-norm for some p, $1 \leq p \leq \infty$. Let L be diagonal, so $Lx = (\lambda_1 x_1, \lambda_2 x_2, \lambda_3 x_3, \ldots)^T$ for $x \in \mathbb{F}_0^\infty$, where $\lambda_i \in \mathbb{C}$, $i \geq 1$. Then L is a bounded linear operator iff $\sup_i |\lambda_i| < \infty$.

Exercise. Verify the claim in example (2) above.

We have already proved:

Proposition. Let $L : V \to W$ be a linear transformation between normed vector spaces. Then L is bounded iff L is continuous iff L is uniformly continuous.

Definition. Let $L : V \to W$ be a bounded linear operator between normed linear spaces, i.e., $L \in \mathcal{B}(V, W)$. Define the operator norm of L to be

$$\|L\| = \sup_{\|v\|_V \leq 1} \|Lv\|_W \left(= \sup_{\|v\|_V = 1} \|Lv\|_W = \sup_{v \neq 0} (\|Lv\|_W / \|v\|_V) \right).$$

Remark. $(\forall\, v \in V)$ $\|Lv\|_W \leq \|L\| \cdot \|v\|_V$. In fact, $\|L\|$ is the smallest constant with this property: $\|L\| = \min\{C \geq 0 : (\forall\, v \in V)\ \|Lv\|_W \leq C\|v\|_V\}$.

We can now show that $\mathcal{B}(V, W)$ is a vector space (a subspace of $\mathcal{L}(V, W)$). If $L \in \mathcal{B}(V, W)$ and $\alpha \in \mathbb{F}$, clearly $\alpha L \in \mathcal{B}(V, W)$ and $\|\alpha L\| = |\alpha| \cdot \|L\|$. If $L_1, L_2 \in \mathcal{B}(V, W)$, then $\|(L_1 + L_2)v\|_W \leq \|L_1 v\|_W + \|L_2 v\|_W \leq (\|L_1\| + \|L_2\|)\|v\|_V$, so $L_1 + L_2 \in \mathcal{B}(V, W)$, and $\|L_1 + L_2\| \leq \|L_1\| + \|L_2\|$. It follows that the operator norm is indeed a norm on $\mathcal{B}(V, W)$. $\|\cdot\|$ is sometimes called the operator norm on $\mathcal{B}(V, W)$ induced by the norms $\|\cdot\|_V$ and $\|\cdot\|_W$ (as it clearly depends on both $\|\cdot\|_V$ and $\|\cdot\|_W$).

In the special case $W = \mathbb{F}$, the norm $\|f\| = \sup_{\|v\|_V \leq 1} |f(v)|$ on V^* is called the *dual norm* to that on V. If $\dim V < \infty$, then we can choose bases and identify V and V^* with $\mathbb{F}^n$. Thus every norm on $\mathbb{F}^n$ has a dual norm on $\mathbb{F}^n$. We sometimes write F^{n^*} for $\mathbb{F}^n$ when it is being identified with V^*. Consider some examples.

Examples.

(1) If $\mathbb{F}^n$ is given the ℓ^1-norm, then the dual norm is $\|f\| = \max_{\|x\|_1 \leq 1} \left|\sum_{i=1}^n f_i x_i\right|$ for $f = (f_1, \ldots, f_n) \in \mathbb{F}^{n^*}$, which is easily seen to be the ℓ^∞-norm $\|f\|_\infty$ (exercise).

(2) If $\mathbb{F}^n$ is given the ℓ^∞-norm, then the dual norm is $\|f\| = \max_{\|x\|_\infty \le 1} |\sum_{i=1}^n f_i x_i|$ for $f = (f_1, \ldots, f_n) \in {\mathbb{F}^n}^*$, which is easily seen to be the ℓ^1-norm $\|f\|_1$ (exercise).

(3) The dual norm to the ℓ^2-norm on $\mathbb{F}^n$ is again the ℓ^2-norm; this follows easily from the Schwarz inequality (exercise). The ℓ^2-norm is the only norm on $\mathbb{F}^n$ which equals its own dual norm; see the homework.

(4) Let $1 < p < \infty$. The dual norm to the ℓ^p-norm on $\mathbb{F}^n$ is the ℓ^q-norm, where $\frac{1}{p} + \frac{1}{q} = 1$. The key inequality is Hölder's inequality: $|\sum_{i=1}^n f_i x_i| \le \|f\|_q \cdot \|x\|_p$. We will be primarily interested in the cases $p = 1, 2, \infty$. (Note: $\frac{1}{p} + \frac{1}{q} = 1$ in an extended sense when $p = 1$ and $q = \infty$, or when $p = \infty$ and $q = 1$; Hölder's inequality is trivial in these cases.)

It is instructive to consider linear functionals and the dual norm geometrically. Recall that a norm on $\mathbb{F}^n$ can be described geometrically by its closed unit ball B, a compact convex set. The geometric realization of a linear functional (excluding the zero functional) is a hyperplane. (A hyperplane in $\mathbb{F}^n$ is a set of the form $\{x \in \mathbb{F}^n : \sum_{i=1}^n f_i x_i = c\}$, where $f_i \in \mathbb{F}$ and not all $f_i = 0$; sets of this form are sometimes called *affine* hyperplanes if the term "hyperplane" is being reserved for a subspace of $\mathbb{F}^n$ of dimension $n-1$.) In fact, there is a natural $1-1$ correspondence between ${\mathbb{F}^n}^* \backslash \{0\}$ and the set of hyperplanes in $\mathbb{F}^n$ which do not contain the origin: to $f = (f_1, \ldots, f_n) \in {\mathbb{F}^n}^*$, associate the hyperplane $\{x \in \mathbb{F}^n : f(x) = f_1 x_1 + \cdots + f_n x_n = 1\}$; since every hyperplane not containing 0 has a unique equation of this form, this is a $1-1$ correspondence as claimed.

If $\mathbb{F} = \mathbb{C}$ it is often more appropriate to use real hyperplanes in $\mathbb{C}^n = \mathbb{R}^{2n}$; if $z \in \mathbb{C}^n$ and we write $z_j = x_j + i y_j$, then a real hyperplane not containing $\{0\}$ has a unique equation of the form $\sum_{j=1}^n (a_j x_j + b_j y_j) = 1$ where $a_j, b_j \in \mathbb{R}$, and not all of the a_j's and b_j's vanish. Observe that this equation is of the form $\mathcal{R}e\left(\sum_{j=1}^n f_j z_j\right) = 1$ where $f_j = a_j - i b_j$ is uniquely determined. Thus the real hyperplanes in $\mathbb{C}^n$ not containing $\{0\}$ are all of the form $\mathcal{R}e f(z) = 1$ for a unique $f \in {\mathbb{C}^n}^* \backslash \{0\}$.

Proposition. If $(V, \|\cdot\|)$ is a normed linear space and $f \in V^*$, then the dual norm of f satisfies $\|f\| = \sup_{\|v\| \le 1} \mathcal{R}e f(v)$.

Proof. Since $\mathcal{R}e f(v) \le |f(v)|$, $\sup_{\|v\| \le 1} \mathcal{R}e f(v) \le \sup_{\|v\| \le 1} |f(v)| = \|f\|$. For the other direction, choose a sequence $\{v_j\}$ from V with $\|v_j\| = 1$ and $|f(v_j)| \to \|f\|$. Taking $\theta_j = -\arg f(v_j)$ and setting $w_j = e^{i\theta_j} v_j$, we have $\|w_j\| = 1$ and $f(w_j) = |f(v_j)| \to \|f\|$, so $\sup_{\|v\| \le 1} \mathcal{R}e f(v) \ge \|f\|$. □

With these observations, we can give a description of the dual unit ball in terms of the geometry of the hyperplanes and the unit ball in the original norm. By the above, $f \in {\mathbb{F}^n}^*$ satisfies $\|f\| \le 1$ iff $\sup_{\|v\| \le 1} \mathcal{R}e f(v) \le 1$, i.e., iff the unit ball $B \subset \mathbb{F}^n$ is contained in the closed half-space $\mathcal{R}e f(v) \le 1$ (the real hyperplane $\{\mathcal{R}e f(v) = 1\}$ divides $\mathbb{F}^n$ into two half-spaces; this is the one containing the origin). Moreover, by linearity, if $\|f\| \le 1$ and $\|v\| = p < 1$, then $\mathcal{R}e f(v) \le p < 1$, so the open unit ball $B^0 \subset \{f : \mathcal{R}e f(v) < 1 \ \forall v \in B\}$. So we have a description of the dual unit ball on those functionals corresponding to hyperplanes lying outside the open unit ball $B^0 = \{f : \|f\| < 1\}$. It is interesting to translate this into a geometric dual unit ball in specific examples; see the homework.

Proposition. If $(V, \|\cdot\|)$ is a normed linear space and $v \in V$, then

$$(\exists\, f \in V^*) \ni \|f\| = 1 \qquad \text{and} \qquad f(v) = \|v\|.$$

In general, this is an immediate consequence of the Hahn-Banach theorem (see, e.g., Royden *Real Analysis* or Folland *Real Analysis*), and for convenience we will refer to it here as the Hahn-Banach theorem. In finite dimensions, there are more geometric proofs based on relating hyperplanes to the closed unit ball. See, e.g., Corollary 5.5.15 in H-J (see also Appendix B in H-J).

Consequences of the Hahn-Banach theorem

The Second Dual

Let $(V, \|\cdot\|)$ be a normed linear space, V^* be its dual equipped with the dual norm, and V^{**} be the dual of V^* with the norm dual to that on V^*. Given $v \in V$, define $v^{**} \in V^{**}$ by $v^{**}(f) = f(v)$; since $|v^{**}(f)| \leq \|f\| \cdot \|v\|$, $v^{**} \in V^{**}$ and $\|v^{**}\| \leq \|v\|$. By the Hahn-Banach theorem, $\exists\, f \in V^*$ with $\|f\| = 1$ and $f(v) = \|v\|$, i.e., $v^{**}(f) = \|v\|$, so $\|v^{**}\| = \sup_{\|f\|=1} |v^{**}(f)| \geq \|v\|$. Hence $\|v^{**}\| = \|v\|$, so the mapping $v \mapsto v^{**}$ from V into V^{**} is an isometry of V onto the range of this map. In general, this embedding is not surjective; if it is, then $(V, \|\cdot\|)$ is called *reflexive*

In finite dimensions, dimension arguments imply this map is surjective. Thus the dual norm to the dual norm is just the original norm on V.

Adjoint Transformations

Recall that if $L \in \mathcal{L}(V, W)$, the adjoint transformation $L^* : W' \to V'$ is given by $(L^*g)(v) = g(Lv)$.

Proposition. Let V, W be normed linear spaces. If $L \in \mathcal{B}(V, W)$, then $L^*[W^*] \subset V^*$. Moreover, $L^* \in \mathcal{B}(W^*, V^*)$ and $\|L^*\| = \|L\|$.

Proof. For $g \in W^*$, $|(L^*g)(v)| = |g(Lv)| \leq \|g\| \cdot \|L\| \cdot \|v\|$, so $L^*g \in V^*$, and $\|L^*g\| \leq \|g\| \cdot \|L\|$. Thus $L^* \in \mathcal{B}(W^*, V^*)$ and $\|L^*\| \leq \|L\|$. Now given $v \in V$, apply the Hahn-Banach theorem to Lv to conclude that $\exists\, g \in W^*$ with $\|g\| = 1$ and $(L^*g)(v) = g(Lv) = \|Lv\|$. So $\|L^*\| = \sup_{\|g\|\leq 1} \|L^*g\| = \sup_{\|g\|\leq 1} \sup_{\|v\|\leq 1} |(L^*g)(v)| \geq \sup_{\|v\|\leq 1} \|Lv\| = \|L\|$. Hence $\|L^*\| = \|L\|$. □

Completeness of $\mathcal{B}(V, W)$ when W is complete

Proposition. If W is complete, the $\mathcal{B}(V, W)$ is complete. In particular, V^* is always complete (since $\mathbb{F}$ is), whether or not V is.

Proof. If $\{L_n\}$ is Cauchy in $\mathcal{B}(V, W)$, then $(\forall\, v \in V)\{L_n v\}$ is Cauchy in W, so the limit $\lim_{n\to\infty} L_n v \equiv Lv$ exists in W. Clearly $L : V \to W$ is linear, and it is easy to see that $L \in \mathcal{B}(V, W)$ and $\|L_n - L\| \to 0$. □

Analysis with Operators

Throughout this discussion, let V be a Banach space. Since V is complete, $\mathcal{B}(V) = \mathcal{B}(V, V)$ is also complete (in the operator norm).

Fact. Operator norms are always submultiplicative.

In fact, if U, V, W are normed linear spaces and $L \in \mathcal{B}(U, V)$ and $M \in \mathcal{B}(V, W)$, then for $u \in U$,

$$\|(M \circ L)(u)\|_W = \|M(Lu)\|_W \leq \|M\| \cdot \|Lu\|_V \leq \|M\| \cdot \|L\| \cdot \|u\|_U,$$

so $M \circ L \in \mathcal{B}(U, W)$ and $\|M \circ L\| \leq \|M\| \cdot \|L\|$. The special case $U = V = W$ shows that the operator norm on $\mathcal{B}(V)$ is submultiplicative (and $L, M \in \mathcal{B}(V) \Rightarrow M \circ L \in \mathcal{B}(V)$). We want to define functions of an operator $L \in \mathcal{B}(V)$. We can compose L with itself, so we can form powers $L^k = L \circ \cdots \circ L$, and thus we can define polynomials in L: if $p(z) = a_0 + a_1 z + \cdots + a_n z^n$, then $p(L) \equiv a_0 I + a_1 L + \cdots + a_n L^n$. By taking limits, we can form power series, and thus analytic functions of L. For example, consider the series $e^L = \sum_{k=0}^{\infty} \frac{1}{k!} L^k = I + L + \frac{1}{2} L^2 + \cdots$ (note L^0 is the identity I by definition). This series converges in the operator norm on $\mathcal{B}(V)$: by submultiplicativity, $\|L^k\| \leq \|L\|^k$, so $\sum_{k=0}^{\infty} \frac{1}{k!} \|L^k\| \leq \sum_{k=0}^{\infty} \frac{1}{k!} \|L\|^k = e^{\|L\|} < \infty$; since the series converges absolutely and $\mathcal{B}(V)$ is complete (recall V is a Banach space), it converges in the operator norm to an operator in $\mathcal{B}(V)$ which we call e^L (note that $\|e^L\| \leq e^{\|L\|}$). In the finite dimensional case, this says that for $A \in \mathbb{F}^{n \times n}$, each component of the partial sum $\sum_{k=0}^{N} \frac{1}{k!} A^k$ converges as $N \to \infty$; the limiting matrix is e^A.

Another fundamental example is the Neumann series.

Proposition. If $L \in \mathcal{B}(V)$ and $\|L\| < 1$, then $I - L$ is invertible, and the Neumann series $\sum_{k=0}^{\infty} L^k$ converges in the operator norm to $(I - L)^{-1}$.

Remark. Formally we can guess this result since the power series of $\frac{1}{1-z}$ centered at $z = 0$ is $\sum_{k=0}^{\infty} z^k$ with radius of convergence 1.

Proof. If $\|L\| < 1$, then $\sum_{k=0}^{\infty} \|L^k\| \leq \sum_{k=0}^{\infty} \|L\|^k = \frac{1}{1 - \|L\|} < \infty$, so the Neumann series $\sum_{k=0}^{\infty} L^k$ converges to an operator in $\mathcal{B}(V)$. Now if $S_j, S, T \in \mathcal{B}(V)$ and $S_j \to S$ in $\mathcal{B}(V)$, then $\|S_j - S\| \to 0$, so $\|S_j T - ST\| \leq \|S_j - S\| \cdot \|T\| \to 0$ and $\|TS_j - TS\| \leq \|T\| \cdot \|S_j - S\| \to 0$, and thus $S_j T \to ST$ and $TS_j \to TS$ in $\mathcal{B}(V)$. Thus $(I - L)\left(\sum_{k=0}^{\infty} L^k\right) = \lim_{N \to \infty} (I - L) \sum_{k=0}^{N} L^k = \lim_{N \to \infty} (I - L^{N+1}) = I$ (as $\|L^{N+1}\| \leq \|L\|^{N+1} \to 0$), and similarly $\left(\sum_{k=0}^{\infty} L^k\right)(I - L) = I$. So $I - L$ is invertible and $(I - L)^{-1} = \sum_{k=0}^{\infty} L^k$. □

This is a very useful fact: a perturbation of I by an operator of norm < 1 is invertible. This implies, among other things, that the set of invertible operators in $\mathcal{B}(V)$ is an open subset of $\mathcal{B}(V)$ (in the operator norm).

Our terminology above is that an operator in $\mathcal{B}(V)$ is called invertible if it is bijective (i.e., invertible as a point-set mapping from V onto V, which implies that the inverse map is well-defined and linear) *and* that its inverse is also in $\mathcal{B}(V)$.

Note: $\mathcal{B}(V)$ has a ring structure using the addition of operators, and composition of operators as the multiplication; the identity of multiplication is just the identity operator I. Our concept of invertibility is equivalent to invertibility in this ring: if $L \in \mathcal{B}(V)$ and $\exists\, M \in$

$\mathcal{B}(V) \ni LM = ML = I$, then $ML = I \Rightarrow L$ injective and $LM = I \Rightarrow L$ surjective. Note that this ring in general is *not* commutative.

It is a consequence of the closed graph theorem (see Royden or Folland) that if $L \in \mathcal{B}(V)$ is bijective (and V is a Banach space), then its inverse map L^{-1} is also in $\mathcal{B}(V)$.

Clearly the power series arguments used above can be generalized. Let $f(z)$ be analytic on the disk $\{|z| < R\} \subset \mathbb{C}$, with power series $f(z) = \sum_{k=0}^{\infty} a_k z^k$ (which has radius of convergence at least R). If $L \in \mathcal{B}(V)$ and $\|L\| < R$, then the series $\sum_{k=0}^{\infty} a_k L^k$ converges absolutely, and thus converges to an element of $\mathcal{B}(V)$ which we call $f(L)$ (recall V is a Banach space). It is easy to check that usual operational properties hold, for example $(fg)(L) = f(L)g(L) = g(L)f(L)$. However, one must be careful to remember that operators do not commute in general. So, for example, $e^{L+M} \neq e^L e^M$ in general, although if L and M commute (i.e. $LM = ML$), then $e^{L+M} = e^L e^M$.

Let $L(t)$ be a 1-parameter family of operators in $\mathcal{B}(V)$, where $t \in (a, b)$. Since $\mathcal{B}(V)$ is a metric space, we know what it means for $L(t)$ to be a continuous function of t. We can define differentiability as well: $L(t)$ is differentiable at $t = t_0 \in (a, b)$ if $L'(t_0) = \lim_{t \to t_0} \frac{L(t) - L(t_0)}{t - t_0}$ exists in the norm on $\mathcal{B}(V)$. For example, it is easily checked that for $L \in \mathcal{B}(V)$, e^{tL} is differentiable in t for all $t \in \mathbb{R}$, and $\frac{d}{dt} e^{tL} = L e^{tL} = e^{tL} L$.

We can similarly consider families of operators in $\mathcal{B}(V)$ depending on several real parameters or on complex parameter(s). A family $L(z)$ where $z = x + iy \in \Omega^{\text{open}} \subset \mathbb{C}$ $(x, y \in \mathbb{R})$ is said to be holomorphic in Ω if the partial derivatives $\frac{\partial}{\partial x} L(z)$, $\frac{\partial}{\partial y} L(z)$ exist and are continuous in Ω, and $L(z)$ satisfies the Cauchy-Riemann equation $\left(\frac{\partial}{\partial x} + i \frac{\partial}{\partial y}\right) L(z) = 0$ in Ω. As in complex analysis, this is equivalent to the assumption that in a neighborhood of each point $z_0 \in \Omega$, $L(z)$ is given by the $\mathcal{B}(V)$-norm convergent power series $L(z) = \sum_{k=0}^{\infty} \frac{1}{k!} (z - z_0)^k \left(\frac{d}{dz}\right)^k L(z_0)$.

One can also integrate families of operators. If $L(t)$ depends continuously on $t \in [a, b]$, then it can be shown using the same estimates as for $\mathbb{F}$-valued functions (and the uniform continuity of $L(t)$ since $[a, b]$ is compact) that the Riemann sums $\frac{b-a}{N} \sum_{k=0}^{N-1} L\left(a + \frac{k}{N}(b - a)\right)$ converge in $\mathcal{B}(V)$-norm (recall V is a Banach space) as $N \to \infty$ to an operator in $\mathcal{B}(V)$, denoted $\int_a^b L(t)dt$. (More general Riemann sums than just the left-hand "rectangular rule" with equally spaced points can be used.) Many results from standard calculus carry over, including $\left\| \int_a^b L(t)dt \right\| \leq \int_a^b \|L(t)\| dt$ which follows directly from $\left\| \frac{b-a}{N} \sum_{k=0}^{N-1} L\left(a + \frac{k}{N}(b - a)\right) \right\| \leq \frac{b-a}{N} \sum_{k=0}^{N-1} \left\| L\left(a + \frac{k}{N}(b - a)\right) \right\|$. By parameterizing paths in $\mathbb{C}$, one can define line integrals of holomorphic families of operators. We will discuss such constructions further as we need them.

Operators in Finite Dimensions

In the next part of the course we will study in greater detail operators in finite dimensions and the matrices which represent them.

Transposes and Adjoints

If $A \in \mathbb{C}^{m\times n}$ we denote by $A^T \in \mathbb{C}^{n\times m}$ the transpose of A, and by $A^H = \bar{A}^T$ the conjugate-transpose (or hermitian transpose) of A. (Many books, including H-J, use the notation A^* for A^H.) If $x, y \in \mathbb{C}^n$ are represented in terms of matrix multiplication as $\langle x, y\rangle = y^H x$, then for $A \in \mathbb{C}^{n\times n}$, we then have $\langle Ax, y\rangle = \langle x, A^H y\rangle$ since $y^H Ax = (A^H y)^H x$.

Caution: The notation A^*, or L^* for a linear transformation, is used with two different, sometimes contradictory meanings, particularly if $\mathbb{F} = \mathbb{C}$. Recall that if $L \in \mathcal{B}(V, W)$ then $L^* \in \mathcal{B}(W^*, V^*)$ and in the finite dimensional case, we saw that if L corresponds to matrix multiplication on column vectors from the left by the matrix T, then L^* corresponds to matrix multiplication on row vectors from the right by the matrix T, or equivalently by transposition to left-multiplication by the transpose matrix T^T on column vectors. On the other hand, in the presence of an inner product, the usual definition $\langle Lx, y\rangle = \langle x, L^*y\rangle$ identifies L^* with left-multiplication by the conjugate-transpose matrix. These two definitions are related by the identification $V \cong V^*$ induced by the inner product, but the conjugation in this identification gives rise to the two inequivalent definitions of L^*. So you must be careful to be sure which is meant in a given context. (Some authors use the notation V' for $V^* = \mathcal{B}(V, \mathbb{F})$ and the notation $L' \in \mathcal{B}(W', V')$ for the transpose operator, reserving the notation L^* for use with inner products.)

Norms on Matrices

Commonly used norms on $\mathbb{C}^{n\times n}$ are the following. (We use the notation of H-J.)

$$\|A\|_1 = \sum_{i,j=1}^n |a_{ij}| \qquad \text{(the } \ell^1\text{-norm on } A \text{ as if it were in } \mathbb{C}^{n^2})$$

$$\|A\|_\infty = \max_{i,j} |a_{ij}| \qquad \text{(the } \ell^\infty\text{-norm on } A \text{ as if it were in } \mathbb{C}^{n^2})$$

$$\|A\|_2 = \left(\sum_{i,j=1}^n |a_{ij}|^2\right)^{\frac{1}{2}} \qquad \text{(the } \ell^2\text{-norm on } A \text{ as if it were in } \mathbb{C}^{n^2})$$

The norm $\|A\|_2$ is called the Hilbert-Schmidt norm of A, or the Frobenius norm of A, and is often denoted $\|A\|_F$. It is sometimes called the Euclidean norm of A. This norm comes from an inner product $\langle A, B\rangle = \sum_{i,j=1}^n a_{ij}\overline{b_{ij}} = \mathrm{tr}\,(B^*A)$.

We also have the following p-norms for matrices: let $1 \le p \le \infty$, then

$$|||A|||_p = \max_{\|x\|_p=1} \|Ax\|_p \qquad \left(= \max_{\|x\|_p\le 1} \|Ax\|_p = \max_{x\ne 0}(\|Ax\|_p/\|x\|_p)\right).$$

Caution: $|||A|||_p$ is a quite non-standard notation; the standard notation is $\|A\|_p$, and a more standard notation for the Frobenius norm is $\|A\|_F$, particularly in numerical analysis. We will, however, go ahead and use the notation of H-J.

Using arguments similar to those identifying the dual norms to the ℓ^1- and ℓ^∞-norms on $\mathbb{C}^n$, it can be easily shown that

$$|||A|||_1 = \max_{1\le j\le n} \sum_{i=1}^n |a_{ij}| \qquad \text{(maximum (absolute) column sum)}$$

$$|||A|||_\infty = \max_{1\le i\le n} \sum_{j=1}^n |a_{ij}| \qquad \text{(maximum (absolute) row sum)}$$

$|||A|||_2$ is often called the spectral norm (we will show later that it equals the square root of the largest eigenvalues of $A^H A$.)

All of the above norms are submultiplicative except for $\|\cdot\|_\infty$, which we have previously discussed.

Consistent Matrix Norms

The concept of submultiplicativity can be extended to rectangular matrices.

Definition. Let $\mu : \mathbb{C}^{m\times n} \to \mathbb{R}$, $\nu : \mathbb{C}^{n\times k} \to \mathbb{R}$, $\rho : \mathbb{C}^{m\times k} \to \mathbb{R}$ be norms. We say that μ, ν, ρ are *consistent* if $\forall\, A \in \mathbb{C}^{m\times n}$ and $\forall\, B \in \mathbb{C}^{n\times k}$,

$$\rho(AB) \le \mu(A)\nu(B)$$

Definition. A norm on $\mathbb{F}^{n\times n}$ is called consistent if it is consistent with itself, i.e., the definition above with $m = n = k$ and $\rho = \mu = \nu$. So by definition a norm on $\mathbb{F}^{n\times n}$ is consistent iff it is submultiplicative.

In this discussion of consistent matrix norms, we identify $\mathbb{F}^n$ with $\mathbb{F}^{n\times 1}$ (i.e., $n\times 1$ matrices or column vectors).
Examples.

(1) Let $k = 1$. Then ρ is a norm on $\mathbb{F}^m$ $(\cong \mathbb{F}^{m\times 1})$, ν is a norm on $\mathbb{F}^n$ $(\cong \mathbb{F}^{n\times 1})$, and μ is a norm on $\mathbb{F}^{m\times n}$. If μ_0 is the operator norm induced by ν and ρ, then $\forall\, A \in \mathbb{F}^{m\times n}$ and $\forall\, x \in \mathbb{F}^n$, $\rho(Ax) \le \mu_0(A)\nu(x)$, so μ_0, ν, and ρ are consistent.

(2) Again, let $k = 1$, and ρ and ν be norms on $\mathbb{F}^m$ and $\mathbb{F}^n$, respectively. Let μ be a norm on $\mathbb{F}^{m\times n}$. Then μ, ν, ρ are consistent iff $\mu \ge \mu_0$ where μ_0 is the operator norm on $\mathbb{F}^{m\times n}$ induced by ν and ρ. (For each $A \in \mathbb{F}^{m\times n}$, $(\forall\, x \in \mathbb{F}^n)$ $\rho(Ax) \le \mu(A)\nu(x)$ iff $(\forall\, x \ne 0)\rho(Ax)/\nu(x) \le \mu(A)$ iff $\mu_0(A) \le \mu(A)$.)

Families of Matrix Norms

A collection $\{\nu_{m,n} : m \ge 1, n \ge 1\}$, where $\nu_{m,n} : \mathbb{F}^{m\times n} \to \mathbb{R}$ is a norm on $\mathbb{F}^{m\times n}$, is called a *family of matrix norms* (we temporarily discard the H-J assumption of submultiplicativity on the "matrix norms" $\nu_{n,n}$).

Definition. A family $\{\nu_{m,n} : m \ge 1, n \ge 1\}$ of matrix norms is called *consistent* if

$$(\forall\, m, n, k \ge 1)(\forall\, A \in \mathbb{F}^{m\times n})(\forall\, B \in \mathbb{F}^{n\times k})\ \nu_{m,k}(AB) \le \nu_{m,n}(A)\nu_{n,k}(B).$$

Facts: Let $\{\nu_{m,n}\}$ be a consistent family of matrix norms. Then

(1) $(\forall\, n \ge 1)$ $\nu_{n,n}$ is submultiplicative.

(2) $(\forall\, m,n \geq 1)$ $(\forall\, A \in \mathbb{F}^{m\times n})$ $\nu_{m,n}(A) \geq \mu_{m,n}(A)$, where $\mu_{m,n}$ is the operator norm on $\mathbb{F}^{m\times n}$ induced by $\nu_{n,1}$ and $\nu_{m,1}$.

Examples.

(1) For $m \geq 1$, let $\nu_{m,1}$ be any norm on $\mathbb{F}^m$. For $m,n \geq 1$, let $\nu_{m,n}$ be the operator norm on $\mathbb{F}^{m\times n}$ induced by $\nu_{n,1}$ and $\nu_{m,1}$ (to avoid contradicting definitions of $\nu_{m,1}$, we take $\nu_{1,1}$ to be the usual absolute value on $\mathbb{F}$). Then $\{\nu_{m,n}\}$ is a consistent family of matrix norms.

(2) (maximum (absolute) row sum norm) For $m,n \geq 1$ and $A \in \mathbb{F}^{m\times n}$, let $\nu_{m,n}(A) = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|$. Then $\nu_{n,1}$ is the ℓ^∞-norm on $\mathbb{F}^n$, and $\nu_{m,n}(A)$ is the operator norm induced by the ℓ^∞-norms on $\mathbb{F}^n$ and $\mathbb{F}^m$ (exercise), which we denote by $|||A|||_\infty$ (even for $m \neq n$). This is a special case of example (1), so it is a consistent family of matrix norms.

(3) (maximum (absolute) column sum norm) For $m,n \geq 1$ and $A \in \mathbb{F}^{m\times n}$, let $\nu_{m,n}(A) = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|$. Then $\nu_{n,1}$ is the ℓ^1-norm on $\mathbb{F}^n$, and $\nu_{m,n}(\cdot)$ is the operator norm induced by the ℓ^1-norms on $\mathbb{F}^n$ and $\mathbb{F}^m$ (exercise), which we denote by $|||A|||_1$ (even for $m \neq n$). This again is a special case of example (1), so it is a consistent family of matrix norms.

(4) (ℓ^1-norm on $\mathbb{F}^{m\times n}$ as if it were $\mathbb{F}^{mn}$) For $m,n \geq 1$ and $A \in \mathbb{F}^{m\times n}$, let $\nu_{m,n}(A) = \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|$. Then $\{\nu_{m,n}\}$ is a consistent family of matrix norms (exercise). We denote $\nu_{m,n}(A)$ by $\|A\|_1$ (even for $m \neq n$). Note that $\nu_{n,1}$ is the ℓ^1-norm on $\mathbb{F}^n$. This is *not* a special case of example (1). Note also that the obvious fact $\|A\|_1 \geq |||A|||_1$ agrees with Fact (2) above.

(5) (ℓ^2-norm on $\mathbb{F}^{m\times n}$ as if it were $\mathbb{F}^{mn}$, i.e., the Hilbert-Schmidt or Frobenius norm) For $m,n \geq 1$ and $A \in \mathbb{F}^{m\times n}$, let $\nu_{m,n}(A) = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2\right)^{\frac{1}{2}}$. Then $\nu_{n,1}$ is the ℓ^2-norm on $\mathbb{F}^n$. If $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times k}$, then by the Schwarz inequality, $(\nu_{m,k}(AB))^2 = \sum_{i=1}^{m}\sum_{j=1}^{k} \left|\sum_{\ell=1}^{n} a_{i\ell}b_{\ell j}\right|^2 \leq \sum_{i=1}^{m}\sum_{j=1}^{k} \left(\sum_{\ell=1}^{n} |a_{i\ell}|^2\right)\left(\sum_{r=1}^{n} |b_{rj}|^2\right) = (\nu_{m,n}(A)\nu_{n,k}(B))^2$, so $\{v_{m,n}\}$ is a consistent family of matrix norms. This is *not* a special case of example (1): for example, for $n > 1$, $\nu_{n,n}(I) = \sqrt{n}$ but the operator norm of I is 1. We denote $\nu_{m,n}(A)$ by $\|A\|_2$ (even for $m \geq n$) (although most authors use $\|A\|_F$ for the Frobenius norm). For $A \in \mathbb{F}^{m\times n}$ and $x \in \mathbb{F}^n$, we have the inequality $\|Ax\|_2 \leq \|A\|_2 \cdot \|x\|_2$. For $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times k}$, $\|AB\|_2 \leq \|A\|_2 \cdot \|B\|_2$. Fact (2) above gives the important inequality: for $A \in \mathbb{F}^{m\times n}$, $|||A|||_2 \leq \|A\|_2$. Thus the operator norm induced by the ℓ^2-norms on $\mathbb{F}^m$ and $\mathbb{F}^n$, which is not trivial to compute, is dominated by the Frobenius norm, which *is* easy to compute.

Condition Number and Error Sensitivity

Throughout this discussion $A \in \mathbb{C}^{n\times n}$ will be assumed to be invertible. We are interested in determining the sensitivity of the solution of the linear system $Ax = b$ (for a given

$b \in \mathbb{C}^n$) to perturbations in the right-hand-side (RHS) vector b or to perturbations in A. One can think of such perturbations as arising from errors in measured data in computational problems, as often occurs when the entries in A and/or b are measured. As we will see, the fundamental quantity is the *condition number* $\kappa(A) = \|A\| \cdot \|A^{-1}\|$ of A, relative to a submultiplicative norm $\|\cdot\|$ on $\mathbb{C}^{n\times n}$. Since $\|I\| \geq 1$ in any submultiplicative norm ($\|I\| = \|I^2\| \leq \|I\|^2 \Rightarrow \|I\| \geq 1$), $\kappa(A) = \|A\| \cdot \|A^{-1}\| \geq \|A \cdot A^{-1}\| = \|I\| \geq 1$.

Suppose $\|\cdot\|$ is a norm on $\mathbb{C}^{n\times n}$ consistent with a norm $\|\cdot\|$ on $\mathbb{C}^n$ (i.e. $\|Ax\| \leq \|A\|\cdot\|x\|$ as defined previously). Suppose first that the RHS vector b is subject to error, but the matrix A is not. Then one actually solves the system $A\widehat{x} = \widehat{b}$ for $\widehat{x}$, where $\widehat{b}$ is presumably close to b, instead of the system $Ax = b$ for x. Let x, $\widehat{x}$ be the solutions of $Ax = b$, $A\widehat{x} = \widehat{b}$, respectively. Define the *error vector* $e = x - \widehat{x}$, and the *residual vector* $r = b - \widehat{b} = b - A\widehat{x}$ (the amount by which $A\widehat{x}$ fails to match b). Then $Ae = A(x - \widehat{x}) = b - \widehat{b} = r$, so $e = A^{-1}r$. Thus $\|e\| \leq \|A^{-1}\| \cdot \|r\|$. Since $Ax = b$, $\|b\| \leq \|A\| \cdot \|x\|$. Multiplying these two inequalities gives $\|e\| \cdot \|b\| \leq \|A\| \cdot \|A^{-1}\| \cdot \|x\| \cdot \|r\|$, i.e. $\frac{\|e\|}{\|x\|} \leq \kappa(A)\frac{\|r\|}{\|b\|}$. So the *relative error* $\frac{\|e\|}{\|x\|}$ is bounded by the condition number $\kappa(A)$ times the *relative residual* $\frac{\|r\|}{\|b\|}$.

Exercise. Let A, b, x, e, and r be as given above and show that $\frac{\|e\|}{\|x\|} \geq \frac{1}{\kappa(A)}\frac{\|r\|}{\|b\|}$.

Matrices for which $\kappa(A)$ is large are called *ill-conditioned* (relative to the norm $\|\cdot\|$); those for which $\kappa(A)$ is closed to $\|I\|$ (which is 1 if $\|\cdot\|$ is the operator norm) are called *well-conditioned* (and perfectly conditioned if $\kappa(A) = \|I\|$). If A is ill-conditioned, small relative errors in the data (RHS vector b) *can* result in large relative errors in the solution.

If $\widehat{x}$ is the result of a numerical algorithm (with round-off error) for solving $Ax = b$, then the error $e = x - \widehat{x}$ is not computable, but the residual $r = b - A\widehat{x}$ is computable, so we obtain an upper bound on the relative error $\frac{\|e\|}{\|x\|} \leq \kappa(A)\frac{\|r\|}{\|b\|}$. In practice, we don't know $\kappa(A)$ (although we may be able to estimate it), and this upper bound may be much larger than the actual relative error.

Suppose now that A is subject to error, but b is not. Then $\widehat{x}$ is now the solution of $(A + E)\widehat{x} = b$, where we assume that the error $E \in \mathbb{C}^{n\times n}$ in the matrix is small enough that $\|A^{-1}E\| < 1$, so $(I + A^{-1}E)^{-1}$ exists and can be computed by a Neumann series; then $A + E$ is invertible and $(A+E)^{-1} = (I + A^{-1}E)^{-1}A^{-1}$. The simplest inequality bounds $\frac{\|e\|}{\|\widehat{x}\|}$, the error relative to $\widehat{x}$, in terms of the relative error $\frac{\|E\|}{\|A\|}$ in A: the equations $Ax = b$ and $(A + E)\widehat{x} = b$ imply $A(x - \widehat{x}) = E\widehat{x}$, $x - \widehat{x} = A^{-1}E\widehat{x}$, and thus $\|x - \widehat{x}\| \leq \|A^{-1}\| \cdot \|E\| \cdot \|\widehat{x}\|$, so that

$$\frac{\|e\|}{\|\widehat{x}\|} \leq \kappa(A)\frac{\|E\|}{\|A\|}.$$

To estimate the error relative to x is more involved and is similar to the estimate derived below.

One can show that if $\widehat{x}$ is the solution of $(A + E)\widehat{x} = \widehat{b}$ with both A and b perturbed, then

$$\frac{\|e\|}{\|x\|} \leq \frac{\kappa(A)}{1 - \kappa(A)\|E\|/\|A\|}\left(\frac{\|E\|}{\|A\|} + \frac{\|r\|}{\|b\|}\right).$$

To establish this relationship use $(A + E)x = b + Ex$ and $(A + E)\widehat{x} = \widehat{b}$ to show $x - \widehat{x} = (A + E)^{-1}(Ex + r)$, and also use $\|r\| \leq \frac{\|r\|}{\|b\|}\|A\| \cdot \|x\|$. Note that if $\kappa(A)\frac{\|E\|}{\|A\|} = \|A^{-1}\| \cdot \|E\|$ is small, then $\frac{\kappa(A)}{1-\kappa(A)\|E\|/\|A\|} \approx \kappa(A)$.

We conclude this discussion by estimating the change in A^{-1} due to a perturbation in A. Suppose $\|A^{-1}\| \cdot \|E\| < 1$. Then as above $A+E$ is invertible, and

$$\begin{aligned} A^{-1} - (A+E)^{-1} &= A^{-1} - \sum_{k=0}^{\infty} (-1)^k (A^{-1}E)^k A^{-1} \\ &= \sum_{k=1}^{\infty} (-1)^{k+1} (A^{-1}E)^k A^{-1}, \end{aligned}$$

so

$$\begin{aligned} \|A^{-1} - (A+E)^{-1}\| &\le \sum_{k=1}^{\infty} \|A^{-1}E\|^k \cdot \|A^{-1}\| \\ &= \frac{\|A^{-1}E\|}{1 - \|A^{-1}E\|} \|A^{-1}\| \\ &\le \frac{\|A^{-1}\| \cdot \|E\|}{1 - \|A^{-1}\| \cdot \|E\|} \|A^{-1}\| \\ &= \frac{\kappa(A)}{1 - \kappa(A)\|E\|/\|A\|} \frac{\|E\|}{\|A\|} \|A^{-1}\|. \end{aligned}$$

So the relative error in the inverse satisfies

$$\frac{\|A^{-1} - (A+E)^{-1}\|}{\|A^{-1}\|} \le \frac{\kappa(A)}{1 - \kappa(A)\|E\|/\|A\|} \frac{\|E\|}{\|A\|}.$$

Again, if $\kappa(A)\frac{\|E\|}{\|A\|}$ is small, then the relative error in the inverse is bounded (approximately) by the condition number $\kappa(A)$ of A times the relative error $\frac{\|E\|}{\|A\|}$ in the matrix A.