Linear transformations on normed vector spaces.

Suppose $\mathcal{V}$ and $\mathcal{W}$ are normed vector spaces. Denote norms on both by $\|\cdot\|$. Suppose $T: \mathcal{V} \to \mathcal{W}$ is linear.

Def.: T is bounded if $\exists\, C$ such that $\|T(v)\| \le C\|v\|$ for all $v \in \mathcal{V}$.

In general, say $F: \mathcal{V} \to \mathcal{W}$ is continuous at $v_* \in \mathcal{V}$ if F is defined at v_* and $\lim_{v \to v_*} F(v) = F(v_*)$, i.e., for every $\epsilon > 0$, $\exists\, \delta > 0 \ni \|F(v) - F(v_*)\| < \epsilon$ whenever $\|v - v_*\| < \delta$.

Prop. If T is bounded, then T is continuous everywhere in $\mathcal{V}$.

Pf.: Suppose $v_* \in \mathcal{V}$. Certainly T is defined at v_*. Moreover, $\|T(v) - T(v_*)\| \le C\|v - v_*\|$, and it follows that $\lim_{v \to v_*} T(v) = T(v_*)$.

Theorem If $\mathcal{V}$ is finite-dimensional, then T is bounded.

Pf.: Suppose $\{v_1, -, v_n\}$ is a basis for $\mathcal{V}$.

For $v = \sum_{i=1}^{n} \alpha_i v_i$, define $\|v\|_\alpha = \|a\|_\alpha$, where $a = \begin{pmatrix}\alpha_1 \\ \vdots \\ \alpha_n\end{pmatrix}$.

Then $\|T(v)\| = \|\sum \alpha_i T(v_i)\| \le \sum |\alpha_i| \, |T(v_i)|$

$$\le \left(\sum \|T(v_i)\|\right) \max_i |\alpha_i| = C\|a\|_\alpha = C\|v\|_\alpha,$$

where $C = \sum \|T(v_i)\|$. Letting M be such that $\|v\|_\alpha \le M\|v\|$ for all $v \in \mathcal{V}$, we have

$$\|T(v)\| \le C\|v\|_\alpha \le CM\|v\|$$

for all $v \in \mathcal{V}$.

Note: If V is infinite-dimensional, then T may not be bounded.

Example: On $C^\infty[0,\pi]$ (infinitely differentiable functions on $[0,\pi]$), suppose $\|\cdot\|$ is the 2-norm, i.e.,

$$\|f\| = \left\{ \int_0^\pi |f(x)|^2 \, dx \right\}^{1/2}.$$

Suppose $T: C^\infty[0,\pi] \to C^\infty[0,\pi]$ is defined by $T(f) = f'$. Suppose $f_k(x) = \sin kx$. Then

$$\|f_k\|^2 = \int_0^\pi (\sin kx)^2 \, dx = \int_0^\pi \tfrac{1}{2}(1-\cos 2kx)\, dx$$

$$= \frac{\pi}{2} - \frac{1}{2k} \sin 2kx \Big|_0^\pi = \frac{\pi}{2}$$

$$\|T(f_k)\|^2 = \int_0^\pi (k\cos kx)^2 \, dx = k^2 \int_0^\pi \tfrac{1}{2}(1+\cos 2kx)\, dx$$

$$= k^2\left(\frac{\pi}{2} + \frac{\sin 2kx}{2k}\Big|_0^\pi\right) = \frac{k^2\pi}{2}$$

So $\|f_k\| = \sqrt{\pi/2}$, $\|T(f_k)\| = k\sqrt{\frac{\pi}{2}} = k\|f_k\|$, and it follows that T is unbounded.

From here, assume that V is finite-dimensional. Note that the set of linear transformations from V to W is a vector space with vector addition and scalar multiplication given by

$$(S+T)(v) = S(v) + T(v), \quad (\alpha T)(v) = \alpha T(v)$$

for linear transformations S, T and $\alpha \in F$. We can define a norm on these linear transformations by

$$\|T\| = \max_{\|v\|=1} \|T(v)\|$$

This is the <u>induced norm</u> or <u>operator norm</u> of T with respect to the vector norm $\|\cdot\|$ on V.

Suppose $T: \mathcal{V} \to \mathcal{V}$.

Let $\rho(T) = \max_{\lambda \in \sigma(T)} |\lambda|$ be the spectral radius of T.

Prop. $\rho(T) \le \|T\|$.

Pf.: Suppose $\lambda \in \sigma(T)$ and $T(v) = \lambda v$ for nonzero $v \in \mathcal{V}$. We can take $\|v\| = 1$. Then

$$|\lambda| = \|\lambda v\| = \|T(v)\| \le \max_{\|v\|=1} \|T(v)\| = \|T\|.$$

Induced matrix norms.

Important and interesting special case when $\mathcal{V} = \mathbb{R}^n$ (or $\mathbb{C}^n$), $\mathcal{W} = \mathbb{R}^m$ (or $\mathbb{C}^m$) and $T(v) = Av$ for $A \in \mathbb{R}^{m\times n}$ (or $\mathbb{C}^{m\times n}$).

Easy to show: If $A \sim m\times p$ and $B \sim p \times n$, then $\|AB\| \le \|A\| \, \|B\|$.

Examples of induced matrix norms:

(1) $\|A\|_1 = \max_{\|v\|_1 = 1} \|Av\|_1$

Claim: $\|A\|_1 = \max_j \sum_i |a_{ij}|$ (max. column sum)

Pf. For v $\ni$ $\|v\|_1 = \sum |v_i| = 1$, we have

$$\|Av\|_1 = \sum_{i=1}^m \Big| \sum_{j=1}^m a_{ij} v_j \Big| \le \sum_i \sum_j |a_{ij}| \, |v_j|$$

$$= \sum_j \Big(\sum_i |a_{ij}| \Big) |v_j| \le \Big(\max_j \sum_i |a_{ij}| \Big) \sum_j |v_j|$$

$$= \max_j \sum_i |a_{ij}| .$$

Follows that $\|A\|_1 \le \max_j \sum_i |a_{ij}|$.

To show $\|A\|_1 = \max_j \sum_i |a_{ij}|$, suppose $\sum_i |a_{ij_*}| = \max_j \sum_i |a_{ij}|$ and take $v = e_{j_*}$. Then $\|v\|_1 = 1$, and

$$\|Av\|_1 = \sum_i |a_{ij_*}| = \max_j \sum_i |a_{ij}|$$

Follows that $\|A\|_1 \ge \max_j \sum_i |a_{ij}|$; hence $\|A\|_1 = \max_j \sum_i |a_{ij}|$.

(2) $\|A\|_\infty = \max_{\|v\|_\infty = 1} \|Av\|_\infty$.

<u>Claim</u>: $\|A\|_\infty = \max_i \sum_j |a_{ij}|$ (max. row sum)

<u>Pf.</u>: For $v \ni \|v\|_\infty = \max_i |v_i| = 1$, we have

$$\|Av\|_\infty = \max_i \left|\sum_j a_{ij} v_j\right| \le \max_i \sum_j |a_{ij}|\,|v_j|$$

$$\le \left(\max_i \sum_j |a_{ij}|\right) \max_j |v_j| = \max_i \sum_j |a_{ij}|$$

$$\Rightarrow \|A\|_\infty \le \max_i \sum_j |a_{ij}|.$$

To show $\|A\|_\infty = \max_i \sum_j |a_{ij}|$, suppose $\sum_j |a_{i_* j}| = \max_i \sum_j |a_{ij}|$ and take v such that $|v_j| = 1$ and $a_{i_* j} v_j = |a_{i_* j}|$ for each j. Then $\|v\|_\infty = 1$ and

$$\|Av\|_\infty = \max_i \left|\sum_j a_{ij} v_j\right| \ge \left|\sum_j a_{i_* j} v_j\right| = \sum_j |a_{i_* j}| = \max_i \sum_j |a_{ij}|$$

Follows that $\|A\|_\infty \ge \max_i \sum_j |a_{ij}|$; hence $\|A\|_\infty = \max_i \sum_j |a_{ij}|$.

(3) $\|A\|_2 = \max_{\|v\|_2 = 1} \|Av\|_2$.

Useful for some theoretical purposes but not easy to evaluate. Return to this norm later to provide some useful characterizations.

Suppose $A \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$). From the general proposition above, we have $\rho(A) \le \|A\|$ for every induced norm $\|\cdot\|$. We can say more.

Theorem. If A is diagonalizable, then there is an induced norm for which $\|A\| = \rho(A)$.

Pf.: Suppose $A = M\Lambda M^{-1}$ for diagonal Λ with diagonal entries in $\sigma(A)$. Let $\|\cdot\|$ be any vector norm for which $\|\Lambda v\| \le (\max_{\lambda\in\sigma(A)} |\lambda|)\|v\| = \rho(A)\|v\|$. ($\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ are examples.) Define $\|\cdot\|_{M^{-1}}$ by $\|v\|_{M^{-1}} = \|M^{-1}v\|$ for $v \in \mathbb{R}^n$ (or $\mathbb{C}^n$). This induces a matrix norm $\|\cdot\|_{M^{-1}}$, and

$$\|A\|_{M^{-1}} = \max_{\|v\|_{M^{-1}}=1} \|Av\|_{M^{-1}} = \max_{\|v\|_{M^{-1}}=1} \|M^{-1}M\Lambda M^{-1}v\|$$

$$= \max_{\|v\|_{M^{-1}}=1} \|\Lambda M^{-1}v\| \le \rho(A) \max_{\|v\|_{M^{-1}}=1} \|M^{-1}v\|$$

$$= \rho(A).$$

Since we already have $\|A\|_{M^{-1}} \ge \rho(A)$, it follows that $\|A\|_{M^{-1}} = \rho(A)$.

In general, we have the following.

Theorem. For any $\epsilon > 0$, there is an induced norm for which $\|A\| \le \rho(A) + \epsilon$.

Pf.: We have the Jordan form $A = MJM^{-1}$, where

$$J = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_\ell \end{pmatrix} \text{ with } J_i = \begin{pmatrix} \lambda_i & \epsilon & & 0 \\ & \ddots & \ddots & \\ & & \ddots & \epsilon \\ 0 & & & \lambda_i \end{pmatrix}.$$

Define $\|\cdot\|_{M^{-1}}$ by $\|v\|_{M^{-1}} = \|M^{-1}v\|_\infty$ for $v \in \mathbb{R}^n$ (or $\mathbb{C}^n$). This induces a matrix norm $\|\cdot\|_{M^{-1}}$, and

$$\|A\|_{M^{-1}} = \max_{\|v\|_{M^{-1}}=1} \|M^{-1} M J M^{-1} v\|_\infty = \max_{\|v\|_{M^{-1}}=1} \|J M^{-1} v\|_\infty$$

$$\leq \max_{1\leq i\leq \ell} (|\delta_i|+\epsilon) \max_{\|v\|_{M^{-1}}=1} \|M^{-1} v\|_\infty = \rho(A)+\epsilon.$$

<u>Inner products</u>.

Norms introduced <u>topology</u>. Inner products introduce <u>geometry</u> (notion of angle, etc.).

<u>Def.</u>: An <u>inner product</u> on a vector space $\mathcal{V}$ over $\mathcal{F}$ is a function $\langle \cdot, \cdot \rangle$ with values in $\mathcal{F}$ satisfying

(1) $\langle v, v\rangle \geq 0$ for all $v \in \mathcal{V}$, with $\langle v,v\rangle = 0 \Leftrightarrow v = 0$.

(2) $\langle v, w\rangle = \begin{cases} \langle w, v\rangle & \text{if } \mathcal{F}=\mathbb{R} \\ \overline{\langle w, v\rangle} & \text{if } \mathcal{F}=\mathbb{C} \end{cases}$ for all $v, w \in \mathcal{V}$.

(3) $\langle \alpha u + \beta v, w\rangle = \alpha \langle u, w\rangle + \beta \langle v, w\rangle$ for all $u, v, w \in \mathcal{V}$ and all $\alpha, \beta \in \mathcal{F}$.

<u>Note</u>: when $\mathcal{F} = \mathbb{C}$, we must have $\langle v,w\rangle = \overline{\langle w,v\rangle}$ (and not $\langle v,w\rangle = \langle w,v\rangle$) for consistency with (1) and (3). Indeed, if $\mathcal{F} = \mathbb{C}$ and $\langle v,w\rangle = \langle w,v\rangle$, then (3) would give $\langle iv, iv\rangle = -\langle v,v\rangle$ for all v, which contradicts (1).

<u>Note</u>: An inner product $\langle\cdot,\cdot\rangle$ on $\mathcal{V}$ induces a norm on $\mathcal{V}$ by $\|v\| = \langle v,v\rangle^{\frac{1}{2}}$. First two norm properties clearly hold. For triangle inequality, need <u>Schwarz inequality</u> (later).

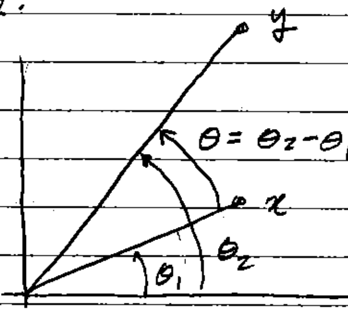
<u>Examples.</u>

(1) The "dot" product on $\mathbb{R}^n$ or $\mathbb{C}^n$.

$$\langle x, y\rangle = \begin{cases} \sum_{i=1}^n x_i y_i & \text{on } \mathbb{R}^n \\ \sum_{i=1}^n x_i \bar{y}_i & \text{on } \mathbb{C}^n \end{cases}$$

See: $\|x\|_2 = \langle x, x\rangle^{\frac{1}{2}}$.

<u>Claim</u>: In $\mathbb{R}^2$, $\langle x, y\rangle = \|x\|\,\|y\| \cos\theta$, where θ is the angle between x and y.



Write $x = \|x\| \begin{pmatrix} \cos\theta_1 \\ \sin\theta_1 \end{pmatrix}$

$y = \|y\| \begin{pmatrix} \cos\theta_2 \\ \sin\theta_2 \end{pmatrix}$

Then

$$\langle x, y\rangle = \|x\|\,\|y\| \left[\cos\theta_1\cos\theta_2 + \sin\theta_1 \sin\theta_2\right]$$

$$= \|x\|\,\|y\| \cos(\theta_2 - \theta_1) = \|x\|\,\|y\| \cos\theta$$

Can extend this to $\mathbb{R}^n$, e.g., by using coordinates in the plane containing x and y. (See also note: That can define $\theta = \cos^{-1}\left(\frac{\langle v, w\rangle}{\|v\|\|w\|}\right)$.)

(2) On $C[0,1]$, $\langle f, g\rangle = \begin{cases} \int_0^1 f(x) g(x)\,dx & \text{if } \mathcal{F} = \mathbb{R} \\ \int_0^1 f(x)\overline{g(x)}\,dx & \text{if } \mathcal{F} = \mathbb{C} \end{cases}$

(Can define $\langle \cdot, \cdot \rangle$ similarly on other function spaces.)

<u>See</u>: $\|f\|_2 = \langle f, f\rangle^{1/2}$.

(3) The <u>Frobenius</u> inner product on $\mathbb{R}^{n\times n}$ or $\mathbb{C}^{n\times n}$.

$$\langle A, B\rangle_F = \begin{cases} \sum_{i,j} a_{ij} b_{ij} & \text{on } \mathbb{R}^{n\times n} \\ \sum_{i,j} a_{ij} \overline{b_{ij}} & \text{on } \mathbb{C}^{n\times n} \end{cases}$$

$$= \begin{cases} \operatorname{trace}\{AB^T\} & \text{on } \mathbb{R}^{n\times n} \\ \operatorname{trace}\{AB^*\} & \text{on } \mathbb{R}^{n\times n} \end{cases}$$

Here, $B^* = \overline{B}^T$ is the <u>Hermitian transpose</u> of B.

The induced norm is $\|A\|_F = \langle A, A\rangle_F^{1/2}$.

<u>Note</u>: This is <u>not</u> an induced norm.

In a general inner product space, we have the fundamentally important

Schwarz Inequality: $|\langle v,w\rangle| \le \|v\|\,\|w\|$ for all v,w, and $|\langle v,w\rangle| = \|v\|\,\|w\| \Leftrightarrow v$ and w are linearly dependent.

Note: AKA Cauchy-Schwarz, Cauchy-Schwarz-Bunyakovskii

Pf.: If $v=w=0$, then the result is trivial.

Suppose that, say, $v\neq 0$. Then

$$0 \le \left\| \frac{\overline{\langle v,w\rangle}}{\|v\|^2} v - w \right\|^2$$

$$= \frac{|\langle v,w\rangle|^2}{\|v\|^4}\|v\|^2 - \left\langle \frac{\overline{\langle v,w\rangle}}{\|v\|^2} v, w\right\rangle - \left\langle w, \frac{\overline{\langle v,w\rangle}}{\|v\|^2} v\right\rangle + \|w\|^2$$

$$= \frac{|\langle v,w\rangle|^2}{\|v\|^2} - 2\frac{|\langle v,w\rangle|^2}{\|v\|^2} + \|w\|^2$$

$$= \|w\|^2 - \frac{|\langle v,w\rangle|^2}{\|v\|^2}$$

$$\Longleftrightarrow \quad |\langle v,w\rangle|^2 \le \|w\|^2\,\|v\|^2$$

See: Equality holds $\Leftrightarrow$ $\frac{\overline{\langle v,w\rangle}}{\|v\|^2} v - w = 0$, i.e., v and w are linearly dependent.

Prop. If $\|\cdot\|$ is defined by $\|v\| = \langle v,v\rangle^{1/2}$, then $\|v+w\| \le \|v\| + \|w\|$ for all v,w.

Pf.: $\|v+w\|^2 = \|v\|^2 + \langle v,w\rangle + \langle w,v\rangle + \|w\|^2$
$\le \|v\|^2 + 2\|v\|\,\|w\| + \|w\|^2 = (\|v\|+\|w\|)^2$.

Note: In a general inner product space $\mathcal{V}$ over $\mathbb{R}$, we can write $\langle v,w\rangle = \|v\|\,\|w\|\cos\theta$, where θ is defined by $\theta = \cos^{-1}\left(\frac{\langle v,w\rangle}{\|v\|\,\|w\|}\right)$.

Orthogonality

In an inner-product space $\mathcal{V}$, vectors v and w are orthogonal if $\langle v, w\rangle = 0$. This is consistent with $\langle v, w\rangle = \|v\|\,\|w\|\cos\theta$.

Extremely useful notion. Much to say. Begin with

Def.: $\{v_1, -, v_k\} \subseteq \mathcal{V}$ is orthonormal if $\langle v_i, v_j\rangle = \delta_{ij}$.

(Here, $\delta_{ij} = \begin{cases} 1, & i=j \\ 0, & i\neq j \end{cases}$.)

Prop. An orthonormal set is linearly independent.

Pf.: If $\sum \alpha_i v_i = 0$, then for $1 \le j \le k$,

$$0 = \langle v_j, \sum \alpha_i v_i\rangle = \sum \alpha_i \overset{\delta_{ij}}{\overset{\shortparallel}{\langle v_j, v_i\rangle}} = \alpha_j.$$

If $\dim \mathcal{V} = n$, then an orthonormal set $\{v_1, -, v_n\}$ is an orthonormal basis. Especially useful: For $v \in \mathcal{V}$, we have

$$v = \sum_{i=1}^{n} \alpha_i v_i \Rightarrow \langle v, v_j\rangle = \sum \alpha_i \overset{\delta_{ij}}{\overset{\shortparallel}{\langle v_i, v_j\rangle}} = \alpha_j.$$

So $v \in \mathcal{V} \Rightarrow v = \sum_{i=1}^{n} \langle v, v_i\rangle v_i$.

How to obtain an orthonormal basis?
Suppose $\{w_1, \dots, w_n\}$ is any basis.
Apply Gram-Schmidt orthogonalization.

Gram-Schmidt Orthogonalization.

Given $\{w_1, \dots, w_k\}$.

Set $v_1 = w_1 / \|w_1\|$.

For $j = 1, \dots, k-1$

$$\tilde{v}_{j+1} = w_{j+1} - \sum_{i=1}^{j} \langle w_{j+1}, v_i \rangle v_i .$$

$$v_{j+1} = \tilde{v}_{j+1} / \|\tilde{v}_{j+1}\|$$

Lemma. If $w_1, \dots, w_k$ are linearly independent, then for $1 \le j \le k$,

(a) v_j is well-defined;

(b) $\operatorname{span}\{v_1, \dots, v_j\} = \operatorname{span}\{w_1, \dots, w_j\}$;

(c) $\{v_1, \dots, v_j\}$ is orthonormal.

Pf: (a)-(c) clearly hold for $j=1$.

Suppose they hold for some j, $1 \le j < k$.

Since $w_1, \dots, w_k$ are linearly indendent, we have that

$$w_{j+1} \notin \operatorname{span}\{w_1, \dots, w_j\} \Rightarrow w_{j+1} \notin \operatorname{span}\{v_1, \dots, v_j\}$$
$$\Rightarrow \tilde{v}_{j+1} \neq 0 \Rightarrow v_{j+1} \text{ is well-defined.}$$

Also, $\operatorname{span}\{v_1, \dots, v_{j+1}\} = \operatorname{span}\{v_1, \dots, v_j, w_{j+1}\} = \operatorname{span}\{w_1, \dots, w_j, w_{j+1}\}$.

Finally, we have $\|v_{j+1}\| = 1$, and for $1 \le l \le j$,

$$\langle v_{j+1}, v_l \rangle = \frac{1}{\|\tilde{v}_{j+1}\|} \langle w_{j+1} - \sum_{i=1}^{j} \langle w_{j+1}, v_i \rangle, v_l \rangle$$

$$= \frac{1}{\|\tilde{v}_{j+1}\|} \left\{ \langle w_{j+1}, v_l \rangle - \sum_{i=1}^{j} \langle w_{j+1}, v_i \rangle \underbrace{\langle v_i, v_l \rangle}_{\delta_{il}} \right\}$$

$$= \frac{1}{\|\tilde{v}_{j+1}\|} \left\{ \langle w_{j+1}, v_l \rangle - \langle w_{j+1}, v_l \rangle \right\} = 0 .$$

So $\{v_1, \dots, v_{j+1}\}$ is orthonormal.

Follows that if $w_1, \ldots, w_k$ are linearly independent, then Gram-Schmidt produces an orthonormal $\{v_1, \ldots, v_k\}$ with $\operatorname{span}\{v_1, \ldots, v_k\} = \operatorname{span}\{w_1, \ldots, w_k\}$. In particular, if $\{w_1, \ldots, w_n\}$ is a basis for $\mathcal{V}$, then $\{v_1, \ldots, v_n\}$ is an orthonormal basis for $\mathcal{V}$.

Note, the algorithm as given is unstable in numerical computations. A stable algorithm that (barring roundoff) produces the same vectors is

Modified Gram-Schmidt.

Given $\{w_1, \ldots, w_k\}$.
Set $v_1 = w_1 / \|w_1\|$.
For $j = 1, \ldots, k-1$
  Set $v_{j+1} = w_{j+1}$
  For $i = 1, \ldots, j$
    $v_{j+1} \leftarrow v_{j+1} - \langle v_{j+1}, v_i \rangle v_i$
  $v_{j+1} \leftarrow v_{j+1} / \|v_{j+1}\|$

Here, the "$\leftarrow$" notation means to replace the object on the left with the object on the right ("overwrite" in computerese).