Functions of matrices.

Recall that in our application to an ODE IVP

$$y' = Ay, \quad y(0) = y_0,$$

with diagonalizable $A = M \Lambda M^{-1}$, $\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$, we found the solution $y = e^{At} y_0$, where

$$e^{At} = M \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \ddots & \\ 0 & & e^{\lambda_n t} \end{pmatrix} M^{-1},$$

and $\sigma(A)$ is critical in determining the behavior of solutions.

Important note: $\lambda \in \sigma(A) \iff e^{\lambda t} \in \sigma(e^{At})$.

We want to use the Jordan form to generalize this to non-diagonalizable matrices and broader classes of functions.

Begin with polynomials in A and prove the iconic Cayley-Hamilton Theorem.

Suppose $A = MJM^{-1}$, where $J = \begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_\ell \end{pmatrix}$ and each J_i is $\ell_i \times \ell_i$ and of the form

$$J_i = \begin{cases} (\lambda_i) & \text{if } \ell_i = 1 \\ \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix} & \text{if } \ell_i > 1 \end{cases}$$

With each J_i, there is a natural invariant subspace of J, viz., $\mathcal{S}_i$ = set of all vectors with only zero components outside the block corresponding to J_i.

Note that $B(\mathcal{S}_i) = \{Bs : s \in \mathcal{S}_i\} = \{0\}$ for $1 \le i \le \ell \iff B = 0$.

See: $A^k = \underbrace{MJM^{-1} \dots MJM^{-1}}_{k \text{ times}} = MJ^kM^{-1} = M\begin{pmatrix} J_1^k & & 0 \\ & \ddots & \\ 0 & & J_\ell^k \end{pmatrix}M^{-1}$,

So if $p(z) = a_0 + a_1 z + \dots + a_q z^q$, then

$$p(A) = a_0 I + a_1 A + \dots + a_q A^q$$
$$= M\begin{pmatrix} p(J_1) & & 0 \\ & \ddots & \\ 0 & & p(J_\ell) \end{pmatrix}M^{-1}.$$

Suppose $p(\lambda) = \det(A - \lambda I)$ is the characteristic polynomial of A. Then

$$p(\lambda) = \det(M^{-1}AM - \lambda M^{-1}M) = \det(J - \lambda I)$$
$$= (-1)^n \prod_{i=1}^{\ell} (\lambda - \lambda_i)^{\ell_i}$$

Consider $p(A)$! We have

$$p(A) = (-1)^n \prod_{i=1}^{\ell} (A - \lambda_i I)^{\ell_i} = (-1)^n \prod_{i=1}^{\ell} [M(J - \lambda_i I)M^{-1}]^{\ell_i}$$

$$= (-1)^n M\left\{\prod_{i=1}^{\ell} (J - \lambda_i I)^{\ell_i}\right\}M^{-1}$$

See: $(\underset{\underset{\ell_i \times \ell_i}{\uparrow}}{J_i - \lambda_i I})^{\ell_i} = \begin{cases} (0) \text{ if } \ell_i = 1 \\ \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}^{\ell_i} = 0 \text{ if } \ell_i > 1 \end{cases}$

So

$$(J - \lambda_i I)^{\ell_i} (\overset{m \times m}{\mathcal{S}_i}) = \{0\}.$$

It follows that $\left\{\prod_{i=1}^{\ell} (J - \lambda_i I)^{\ell_i}\right\}(\mathcal{S}_j) = \{0\}$ for $1 \le j \le \ell$; hence

$$\left\{\prod_{i=1}^{\ell} (J - \lambda_i I)^{\ell_i}\right\} = 0 \implies p(A) = 0.$$

We have proved

Cayley-Hamilton Theorem: If A is a square matrix and $p(\lambda)=\det(A-\lambda I)$ is its characteristic polynomial, then $p(A)=0$.

For $n\times n$ A, this says that the n^2 entries in A concurrently satisfy n^2 polynomial equations!

Now consider $f(A)$ for more general f.

Consider a power series $\sum_{k=0}^{\infty} a_k z^k$.

Define $\rho_* = \dfrac{1}{\overline{\lim_{k\to\infty}} |a_k|^{1/k}}$ = radius of convergence.

Analysis fact: $\sum_{k=0}^{\infty} a_k z^k \begin{cases} \text{converges if } |z|<\rho_* \\ \text{diverges if } |z|>\rho_* \end{cases}$

In general, may have $\rho_*=0$ (converges only for $z=0$) or $\rho_*=\infty$ (converges for all z).

Def.: $f(z)$ is analytic near $z=0$ if

$$f(z)=\sum_{k=0}^{\infty} a_k z^k \quad \text{for } 0\le |z|<\rho_* . \qquad (*)$$

In this case, f is infinitely differentiable, and $a_k = f^{(k)}(0)/k!$, i.e., f can be expressed as a convergent Taylor series near $z=0$.

Suppose $(*)$ holds. What can we say about $f(A)=\sum_{k=0}^{\infty} a_k A^k$?

Suppose $A = MJM^{-1}$, $J=\begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_\ell \end{pmatrix}$ with $J_i \sim l_i\times l_i$

and $J_i = \begin{cases} (\lambda_i) & \text{if } l_i = 1, \\ \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix} & \text{if } l_i > 1. \end{cases}$

As before, $A^k = MJ^kM^{-1} = M\begin{pmatrix} J_1^k & & 0 \\ & \ddots & \\ 0 & & J_\ell^k \end{pmatrix}$.

Look closely at J_i^k:

$$J_i^2 = \begin{pmatrix} \lambda_i^2 & 2\lambda_i & 1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & \ddots & 2\lambda_i \\ 0 & & & & \lambda_i^2 \end{pmatrix}$$

$$J_i^3 = \begin{pmatrix} \lambda_i^3 & 3\lambda_i^2 & 3\lambda_i & 1 & & 0 \\ & \ddots & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & 1 \\ & & & \ddots & \ddots & 3\lambda_i \\ & & & & \ddots & 3\lambda_i^2 \\ 0 & & & & & \lambda_i^3 \end{pmatrix}$$

and for large k,

$$J_i^k = \begin{pmatrix} \lambda_i^k & b_{1k}\lambda_i^{k-1} & b_{2k}\lambda_i^{k-1} & \cdots & b_{l_i-1,k}\lambda_i^{k-l_i+1} \\ & \ddots & \ddots & \ddots & \vdots \\ & & \ddots & \ddots & b_{2k}\lambda_i^{k-1} \\ & & & \ddots & b_{1k}\lambda_i^{k-1} \\ 0 & & & & \lambda_i^k \end{pmatrix}$$

where

$$b_{jk} = \frac{k!}{j!(k-j)!}$$

Easy to see that $1 \le b_{jk} \le k^j$ for large j, so

$$1 \le \varliminf_{k\to\infty} b_{jk}^{1/k} \le \varlimsup_{k\to\infty} b_{jk}^{1/k} \le \varlimsup_{k\to\infty} (k^j)^{1/k} = 1$$

$$\Rightarrow \lim_{k\to\infty} b_{jk}^{1/k} = 1.$$

Then $f(A) = \sum_{k=0}^{\infty} a_k A^k = M\left(\sum_{k=0}^{\infty} a_k J^k\right) M^{-1}$

$$= M \begin{pmatrix} f(J_1) & & \\ & \ddots & \\ & & f(J_\ell) \end{pmatrix} M^{-1}$$

where $f(J_i) = \sum_{k=0}^{\infty} a_k J_i^k$. The entries in the upper triangle of $f(J_i)$ are $\sum_{k=0}^{\infty} a_k b_{jk} \lambda_i^{k-j}$, and

$$\overline{\lim_{k\to\infty}} |a_k b_{jk} \lambda_i^{k-j}|^{1/k} = \overline{\lim_{k\to\infty}} |a_k|^{1/k} |b_{jk}|^{1/k} |\lambda_i^{k-j}|^{1/k}$$

$$= \frac{|\lambda_i|}{\rho_*}.$$

So the power series for $f(A)$ converges

$$\Leftrightarrow \frac{|\lambda_i|}{\rho_*} < 1 \text{ for each } i \Leftrightarrow \rho(A) < \rho_*.$$

Note that if the series converges, then the eigenvalues of $f(A)$ are just $f(\lambda_1), \dots, f(\lambda_\ell)$, i.e., $\sigma(f(A)) = f(\sigma(A))$.

Summarize all this in the

<u>Spectral Mapping Theorem</u>. Suppose that $f(z) = \sum^{\infty} a_k z^k$ is such that

$$\rho_* = \frac{1}{\overline{\lim_{k\to\infty}} |a_k|^{1/k}} > 0.$$

If $A \in \mathbb{F}^{n\times n}$ is such that $\rho(A) < \rho_*$, then

$$f(A) = \sum_{k=0}^{\infty} a_k A^k$$

converges, and $\sigma(f(A)) = f(\sigma(A))$.

Norms and Inner Products.

Until now, everything we've done has been strictly algebraic. Now, will introduce topology on vector spaces (convergence, open & closed sets, continuity of functions, etc.). This will be done through norms.

Suppose $\mathcal{V}$ is a vector space over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$ as always).

Def. A norm on $\mathcal{V}$ is a function $\|\cdot\|: \mathcal{V} \to \mathbb{R}$ such that

(1) $\|x\| \geq 0 \;\forall x \in \mathcal{V}$, and $\|x\| = 0 \iff x = 0$.

(2) $\|\alpha x\| = |\alpha| \|x\| \;\forall x \in \mathcal{V},\ \alpha \in \mathbb{F}$.

(3) $\|x+y\| \leq \|x\| + \|y\| \;\forall x, y \in \mathcal{V}$. (triangle inequality)

Examples: For $x \in \mathbb{R}^n$ or $\mathbb{C}^n$,

(1) $\|x\|_1 = \sum_i |x_i|$, (2) $\|x\|_\infty = \max_i |x_i|$, (3) $\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$, $1 \leq p < \infty$.

(1) and (2) are clearly norms. (3) clearly satisfies the first two norm properties; however, the triangle inequality is not obvious. For general p, proving the triangle inequality for $\|\cdot\|_p$ requires the Hölder and Minkowski inequalities. For the special case $p = 2$ (the Euclidean norm), the triangle inequality follows from Schwarz's Inequality. We will show this later when we introduce inner products.

Examples: Have analogs of these norms on $C[0,1]$.

(1) $\|f\|_1 = \int_0^1 |f(x)|dx$, (2) $\|f\|_\infty = \max_{0 \leq x \leq 1} |f(x)|$, (3) $\|f\|_p = \left(\int_0^1 |f(x)|^p dx\right)^{1/p}$, $1 \leq p < \infty$.

With a norm $\|\cdot\|$ on $\mathcal{V}$, we can discuss convergence.

Def: $\{x_k\} \subseteq \mathcal{V}$ converges to $x_* \in \mathcal{V} \iff \lim_{k \to \infty} \|x_k - x_*\| = 0$.

Denote this by $\lim_{k \to \infty} x_k = x_*$ and say $x_k \to x_*$ as $k \to \infty$.

Q.: Does convergence in one norm imply convergence in another?

In general, no.

Example: Introduce $l_1 = \{(a_1, a_2, \dots) : \sum_{i=1}^{\infty} |a_i| < \infty\}$

For $a = (a_1, a_2, \dots) \in l_1$, define $\|a\|_1 = \sum_{i=1}^{\infty} |a_i|$, $\|a\| = \sum_{i=1}^{\infty} \frac{1}{i} |a_i|$.

Set $e_k = (0, \dots, 0, \underset{k^{th}}{1}, 0, \dots)$ for $k = 1, 2, \dots$

Then $\|e_k\| = \frac{1}{k} \to 0$, so $e_k \to 0$ in the norm $\|\cdot\|$.
However, $\|e_k\|_1 = 1$ for all k, so $e_k \not\to 0$ in the norm $\|\cdot\|_1$.

But in a finite dimensional vector space, convergence in any norm implies convergence in every norm. Will work toward proving this.

Def.: Two norms $\|\cdot\|$ and $|||\cdot|||$ are equivalent if $\exists\, m > 0,\ M > 0$ such that $m\|x\| \le |||x||| \le M\|x\|$ $\forall\, x \in \mathcal{V}$.

See: $m\|x\| \le |||x||| \le M\|x\| \iff \frac{1}{M}|||x||| \le \|x\| \le \frac{1}{m}|||x|||$, so the definition is symmetric. It is also transitive, i.e., if two norms are equivalent and one is equivalent to a third, then the other is also equivalent to the third.

Suppose that $\mathcal{V}$ is finite-dimensional with $\dim \mathcal{V} = n$, and let $\{v_1, \dots, v_n\}$ be a basis for $\mathcal{V}$.

Lemma 8.1 Suppose $\|\cdot\|$ is a norm on $\mathcal{V}$ and $\{x_k = \sum \alpha_{ik} v_i\} \subseteq \mathcal{V}$. If $\lim_{k\to\infty} \alpha_{ik} = \alpha_i^*$ for $1 \le i \le n$, then for $x_* = \sum_{i=1}^{n} \alpha_i^* v_i$,

$\lim_{k\to\infty} \|x_k - x_*\| = 0$, i.e., $\lim_{k\to\infty} x_k = x_*$.

<u>Proof:</u> We have

$$\|x_k - x_*\| = \left\| \sum_{i=1}^{n} (\alpha_{ik} - \alpha_i^*) v_i \right\| \leq \sum_{i=1}^{n} |\alpha_{ik} - \alpha_i^*| \, \|v_i\|$$

$$\to 0 \text{ as } k \to \infty.$$

Recall: For $a = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{F}^n$, we have $\|a\|_\infty = \max_i |\alpha_i|$

Note that $\|\cdot\|_\infty$ on $\mathbb{F}^n$ induces a norm $\|\cdot\|_\infty$ on $\mathcal{V}$ by $\|x\|_\infty = \|a\|_\infty$ for $x = \sum_{i=1}^{n} \alpha_i v_i \in \mathcal{V}$ and $a = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{F}^n$.

Suppose $\|\cdot\|$ is any norm on $\mathcal{V}$. Will show $\|\cdot\|$ and $\|\cdot\|_\infty$ are equivalent.

<u>Claim 1:</u> $\exists$ M such that $\|x\| \leq M \|x\|_\infty$ $\forall$ $x \in \mathcal{V}$.

<u>Proof:</u> Suppose not. Then for $k = 1, 2, \ldots$, $\exists$ nonzero $x_k \in \mathcal{V}$ such that $\|x_k\| \geq k \|x_k\|_\infty$. We can assume that $\|x_k\| = 1$ (divide x_k by $\|x_k\|$ if necessary). Then $1/k \geq \|x_k\|_\infty \Rightarrow \|x_k\|_\infty \to 0$. So if $x_k = \sum \alpha_{ik} v_i$, then $\max_i |\alpha_{ik}| \to 0$, and it follows from Lemma 8.1 that $\|x_k\| \to 0$, which is a contradiction.

<u>Claim 2:</u> $\exists$ m such that $\|x\| \geq m \|x\|_\infty$ $\forall$ $x \in \mathcal{V}$.

<u>Proof:</u> Consider the map $a = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \in \mathbb{F}^n \to \left\| \sum_{i=1}^{n} \alpha_i v_i \right\| \in \mathbb{R}$. It follows from Lemma 8.1 that this is continuous. We know that $S = \{a \in \mathbb{F}^n : \|a\|_\infty = 1\}$ is closed and bounded, so this function has a minimum on S, i.e., $\exists$ $a_* = \begin{pmatrix} \alpha_1^* \\ \vdots \\ \alpha_n^* \end{pmatrix}$ such that $\|a_*\|_\infty = 1$ and

$$\left\| \sum_{i=1}^{n} \alpha_i^* v_i \right\| = \min_{a \in S} \left\| \sum_{i=1}^{n} \alpha_i v_i \right\|.$$

Set $m = \left\| \sum_{i=1}^{n} \alpha_i^* v_i \right\|$. Then for $x = \sum_{i=1}^{n} \alpha_i v_i \in \mathcal{V}$, we have

$$\|x\| = \left\|\sum_{i=1}^{m} \alpha_i v_i\right\| = \|a\|_\infty \left\|\sum_{i=1}^{m} \frac{\alpha_i}{\|a\|_\infty} v_i\right\| \quad \text{where } a = \begin{pmatrix}\alpha_1 \\ \vdots \\ \alpha_m\end{pmatrix}$$

$$\geq \|a\|_\infty \cdot m = m\|x\|_\infty .$$

Since $\|\cdot\|$ was an arbitrary norm on $\mathcal{V}$, we have just shown that every norm is equivalent to $\|\cdot\|_\infty$. Since norm equivalence is transitive, we have

Theorem 8.2. If $\|\cdot\|$ and $|||\cdot|||$ are two norms on a finite-dimensional vector space, then $\|\cdot\|$ and $|||\cdot|||$ are equivalent.

Note the important role in this of Lemma 8.1. This allowed us to "transfer" topological properties of $\mathbb{F}^m$ (derived from those of $\mathbb{R}$ or $\mathbb{C}$) to $\mathcal{V}$ via the norm $\|\cdot\|_\infty$ on both spaces. Continue in this vein.

Still assume that $\mathcal{V}$ is a finite-dimensional vector space with basis $\{v_1, \ldots, v_m\}$ and norm $\|\cdot\|$.

It follows from Theorem 8.2 that these are norm-independent.

Suppose $S \subseteq \mathcal{V}$. Then

- S is bounded if $\sup_{x \in S} \|x\| < \infty$
- x is a limit point of S if $\exists \{x_k\} \subseteq S$ such that $\lim_{k\to\infty} x_k = x$
- S is closed if it contains all of its limit points
- S is compact if every $\{x_k\} \subseteq S$ has a subsequence $\{x_{k_i}\}$ that converges to a point in S.

Heine-Borel Theorem: $S \subseteq \mathcal{V}$ is compact $\Leftrightarrow$ S is closed and bounded.

Proof: $\Rightarrow$: If S is not closed, then $\exists \{x_k\} \subseteq S$ and $x \notin S$ such that $x_k \to x$. Since every subsequence must also converge to x, there can be no subsequence that converges to a point of S, and it follows that S is not compact.

If S is not bounded, then for $k = 1, 2, \ldots$, $\exists\, x_k$ such that $\|x_k\| \geq k$. Then $\{x_k\}$ cannot have a convergent subsequence, and it follows that S is not compact.

$\Leftarrow$: Suppose S is closed and bounded in $\|\cdot\|$. Then S is also closed and bounded in $\|\cdot\|_\alpha$. Suppose $\{x_k = \sum \alpha_{ik} v_i\} \subseteq S$. Then for $1 \leq i \leq n$,

$\left\{ a_k = \begin{pmatrix} \alpha_{1k} \\ \vdots \\ \alpha_{nk} \end{pmatrix} \right\}$ is such that $\left\{ \|a_k\|_\infty = \max_i |\alpha_{ik}| \right\}$ is bounded.

Choose a subsequence of $\{a_k\}$ for which the first components converge.

Choose a subsequence of that subsequence so that the second components also converge.

Continue to obtain $\{a_{k_j}\} \subseteq \{a_k\}$ for which $\alpha_{ik_j} \to \alpha_i^*$ for $1 \leq i \leq n$. Then for $a_* = \begin{pmatrix} \alpha_1^* \\ \vdots \\ \alpha_n^* \end{pmatrix}$,

$\|a_{k_j} - a_*\|_\infty \to 0$, and for $x_* = \sum_i \alpha_i^* v_i$,

$\|x_{k_j} - x_*\|_\alpha \to 0$ and, hence, $\|x_{k_j} - x_*\| \to 0$.

This is an extension of the Heine-Borel Theorem. Have extensions of other familiar results.

Bolzano-Weierstrass Theorem. If $\{x_k\} \subseteq \mathcal{V}$ is bounded, i.e., $\exists M$ such that $\|x_k\| \leq M$ for all k, then $\{x_k\}$ has a convergent subsequence.

Def.: $\{x_k\} \subseteq \mathcal{V}$ is Cauchy if for every $\epsilon > 0$, $\exists N$ such that $\|x_j - x_k\| < \epsilon$ whenever $j, k \geq N$.

Def.: $S \subseteq \mathcal{V}$ is **complete** if every Cauchy sequence in S converges to a point in S.

Note: These are norm independent by Theorem 8.2.

<u>Completeness Theorem</u>. V is complete.

The proofs of the Bolzano-Weierstrass and Completeness Theorems are straightforward and use the $\|\cdot\|_\infty$ norm to "transfer" to V the counterpart results in $\mathbb{R}$ and $\mathbb{C}$.

<u>Note</u>: If V is infinite-dimensional, then none of these theorems holds.

<u>Example</u>: In ℓ_1 introduced earlier, $\{e_k\}$ is bounded in $\|\cdot\|_1$ but has no convergent subsequence, so the Bolzano-Weierstrass Theorem doesn't hold. Moreover, since $\{e_k\}$ is contained in the closed and bounded set $S' = \{a \in \ell_1 : \|a\|_1 = 1\}$, S' is not compact, and the Heine-Borel Theorem doesn't hold.

Finally, consider $\{x_k = (1, \frac{1}{2}, \dots, \frac{1}{k}, 0, 0, \dots)\} \subseteq \ell_1$ and recall the norm $\|a\| = \sum_{j=1}^{\infty} |a_j|/j$ for $a = (a_1, a_2, \dots) \in \ell_1$. Since $\sum_{j=0}^{\infty} 1/j^2 < \infty$, it is easy to see that $\{x_k\}$ is Cauchy <u>in the norm</u> $\|\cdot\|$. However, the limit in the norm $\|\cdot\|$ is $x_* = (1, 1/2, 1/3, \dots)$. Since $x_* \notin \ell_1$, ℓ_1 is not complete in the norm $\|\cdot\|$.