Apprentice Linear Algebra, 5th day, 07/11/05

REU 2005

Instructor: László Babai
Scribe: Ben Lee

7 More on Linear Transformations

7.1 Bases and Similarity

Suppose $\varphi : V \to V$ is linear transformation. Recall λ is an eigenvalue if $(\exists x \in V, x \neq 0)(\varphi x = \lambda x)$.

To associate a matrix to a linear transformation: $\varphi \mapsto [\varphi]$, choose $\mathbf{e} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$ a basis of V

$$\mathbf{x} \mapsto [\mathbf{x}]_{\mathbf{e}} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ the column matrix of coordinates, i.e. } \mathbf{x} = \textstyle\sum_{i=1}^n x_i \mathbf{e}_i.$$

Then the matrix $[\varphi]_{\mathbf{e}}$ is $[[\varphi\mathbf{e}_1]_{\mathbf{e}} \cdots [\varphi\mathbf{e}_n]_{\mathbf{e}}]$.

Example 7.1. ρ_α rotation by α in the plane:

$$\begin{aligned}
\rho_\alpha \mathbf{e}_1 &= \cos\alpha \mathbf{e}_1 + \sin\alpha \mathbf{e}_2 \\
\rho_\alpha \mathbf{e}_2 &= -\sin\alpha \mathbf{e}_1 + \cos\alpha \mathbf{e}_2 \\
[\rho_\alpha]_{\mathbf{e}} &= \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}.
\end{aligned}$$

Theorem 7.2. *If A and B are the matrices associated with the same linear transformation in two bases then $A \sim B$ (A, B are similar.)*

Definition 7.3. $A \sim B$ if $(\exists S, S^{-1})(B = S^{-1}AS)$. S is the matrix of change of basis transformation.

If we have the "old" basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and "new" basis $\mathbf{b}_1, \ldots, \mathbf{b}_n$, $\sigma : \mathbf{a}_i \mapsto \mathbf{b}_i$, then

$$S = [\sigma]_{\mathbf{a}} = [\sigma]_{\mathbf{b}} = [[\mathbf{b}_1]_{\mathbf{a}}, \ldots, [\mathbf{b}_n]_{\mathbf{a}}].$$

If you want to classify all linear transformations, you need to classify matrices up to similarity.

Theorem 7.4. *If $A \sim B$ then $f_A(x) = f_B(x)$ ($f_A(x) = \det(xI - A)$ is the characteristic polynomial.)*

Proof: First prove that if $A \sim B$ then $\det A = \det B$: $B = S^{-1}AS, \det B = \det S^{-1}AS = \det S^{-1}SA = \det A$ since $\det CD = \det C \det D$.

Question: is $xI - A \sim xI - B$? Yes: $S^{-1}(xI - A)S = xS^{-1}IS - S^{-1}AS = xI - B$. So we are done. $\square$

7.2 Diagonalizability and Eigenbases

Definition 7.5. An *eigenbasis* is a basis consisting of eigenvectors.

If $\mathbf{e}_1, \ldots, \mathbf{e}_n$ is an eigenbasis then $\varphi\mathbf{e}_i = \lambda_i\mathbf{e}_i$.

$$\varphi\mathbf{e}_1 = \lambda_1\mathbf{e}_1 + 0\mathbf{e}_2 + \cdots + 0\mathbf{e}_n$$

$$[\varphi\mathbf{e}_1]_{\mathbf{e}} = \begin{pmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\varphi\mathbf{e}_2 = 0\mathbf{e}_1 + \lambda_2\mathbf{e}_2 + 0\mathbf{e}_3 + \cdots + 0\mathbf{e}_n$$

$$[\varphi\mathbf{e}_2]_{\mathbf{e}} = \begin{pmatrix} 0 \\ \lambda_2 \\ \vdots \\ 0 \end{pmatrix} \text{ et cetera, so}$$

$$[\varphi]_{\mathbf{e}} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

Corollary 7.6. $\mathbf{e}$ *is an eigenbasis of* $\varphi \Leftrightarrow [\varphi]_{\mathbf{e}}$ *is a diagonal matrix.*

Definition 7.7. A is *diagonalizable* if it is similar to a diagonal matrix.

Corollary 7.8. $[\varphi]_{\mathbf{e}}$ *is diagonalizable iff* φ *has an eigenbasis.*

Remark: the definitions involving eigenvectors only make sense for linear transformations, not linear maps.

Theorem 7.9. *If $f_A(x) = (x - \lambda_1)\cdots(x - \lambda_n)$ with all λ_i distinct, then A is diagonalizable.*

This is sufficient but not necessary – for example the identity matrix is already diagonal.

Lemma 7.10. *If* $\mathbf{e}_1, \ldots, \mathbf{e}_k$ *are eigenvectors to distinct eigenvalues, then they are linearly independent.*

Proof: [of the theorem from the lemma] Let $A = [\varphi]_{\mathbf{f}}$. Take an eigenvector for each eigenvalue: $\varphi \mathbf{e}_i = \lambda_i \mathbf{e}_i$. By the lemma the $\mathbf{e}_i$ are linearly independent. There are n of them $= \dim V$ so they are a basis. So φ has an eigenbasis $\Rightarrow A$ is diagonalizable. $\square$

Proof: [of lemma] Assumption: $\varphi \mathbf{e}_i = \lambda_i \mathbf{e}_i, i = 1, \ldots, k, \lambda_i$ distinct. Desired conclusion: $\mathbf{e}_1, \ldots, \mathbf{e}_k$ are linearly independent.

Suppose $\sum_{i=1}^k \alpha_i \mathbf{e}_i = 0$. Desired conclusion: all $\alpha_i = 0$. Apply φ to both sides: $\varphi(\sum_{i=1}^k \alpha_i \mathbf{e}_i) = \sum_{i=1}^k \alpha_i \varphi \mathbf{e}_i = \sum_{i=1}^k \alpha_i \lambda_i \mathbf{e}_i = 0$. Would like to use induction.

Base case $k = 1$: single vector is linearly independent only if it is not zero. An eigenvector is always not zero.

Case $k = 2$ (instructive, not needed for induction.) $\varphi \mathbf{e}_1 = \lambda_1 \mathbf{e1}, \varphi \mathbf{e}_2 = \lambda_2 \mathbf{e2}$. Assumption: $\alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 = 0, \alpha_1 \lambda_1 \mathbf{e}_1 + \alpha_2 \lambda_2 \mathbf{e}_2 = 0$. Multiply first equation by $\lambda_1 : \alpha_1 \lambda_1 \mathbf{e}_1 + \alpha_2 \lambda_1 \mathbf{e}_2 = 0$. Eliminate $\mathbf{e}_1 : \alpha_2 \lambda_2 \mathbf{e}_2 - \alpha_2 \lambda_1 \mathbf{e}_2 = 0 \Rightarrow \alpha_2(\lambda_2 - \lambda_1)\mathbf{e}_2 = 0 \Rightarrow \alpha_2(\lambda_2 - \lambda_1) = 0 \Rightarrow \alpha_2 = 0$ since $\lambda_1 \neq \lambda_2$. But this can't happen.

Inductive step: assume lemma is true for $k-1$. Need to prove it for k. Multiply $\sum \alpha_i \mathbf{e}_i = 0$ by λ_k: $\sum_{i=1}^k \alpha_i \lambda_k \mathbf{e}_i$. Subtract: $\sum_{i=1}^{k-1} \alpha_i(\lambda_i - \lambda_k)\mathbf{e}_i = 0$ is a linear relation between $\mathbf{e}_1, \ldots, \mathbf{e}_{k-1}$. By the inductive hypothesis $\mathbf{e}_1, \ldots, \mathbf{e}_{k-1}$ are linearly independent, so $\alpha_i(\lambda_i - \lambda_{k-1}) = 0$ for $i = 1, \ldots, k-1$. $\lambda_i - \lambda_k \neq 0 \Rightarrow \alpha_i = 0$ for $i = 1, \ldots, k-1$. We still need: $\alpha_k = 0$. But $\alpha_1 \mathbf{e}_1 + \cdots + \alpha_k \mathbf{e}_k = 0 \Rightarrow \alpha_k \mathbf{e}_k = 0 \Rightarrow \alpha_k = 0$. $\square$

Alternative proof (illustrated for $k = 3$): we have

$$\begin{aligned} & \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \alpha_3 \mathbf{e}_3 = 0. \\ \text{Apply } \varphi : \quad & \lambda_1 \alpha_1 \mathbf{e}_1 + \lambda_2 \alpha_2 \mathbf{e}_2 + \lambda_3 \alpha_3 \mathbf{e}_3 = 0. \\ \text{Apply } \varphi \text{ again:} \quad & \lambda_1^2 \alpha_1 \mathbf{e}_1 + \lambda_2^2 \alpha_2 \mathbf{e}_2 + \lambda_3^2 \alpha_3 \mathbf{e}_3 = 0. \end{aligned}$$

Call $\alpha_i \mathbf{e}_i = x_i$. We get

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 &= 0 \\ \lambda_1^2 x_1 + \lambda_2^2 x_2 + \lambda_3^2 x_3 &= 0. \end{aligned}$$

This is a system of three equations in three unknowns. If the x_i were numbers (as opposed to vectors) then the system has no nontrivial solutions because the matrix

$$\begin{bmatrix} 1 & 1 & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 \end{bmatrix}$$

is a Vandermonde matrix, with determinant $(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)$. (In general there are $\binom{n}{2}$ terms.) This is only zero if two of the λ_i are equal.

This system of equations works for numbers. To make it work for vectors, take a vector $\mathbf{y}$. Take the inner product of each vector equation with $\mathbf{y}$:

$$\begin{aligned} \alpha_1 \mathbf{e}_1 \cdot \mathbf{y} + \alpha_2 \mathbf{e}_2 \cdot \mathbf{y} + \alpha_3 \mathbf{e}_3 \cdot \mathbf{y} &= 0 \\ \lambda_1 \alpha_1 \mathbf{e}_1 \cdot \mathbf{y} + \lambda_2 \alpha_2 \mathbf{e}_2 \cdot \mathbf{y} + \lambda_3 \alpha_3 \mathbf{e}_3 \cdot \mathbf{y} &= 0 \\ \lambda_1^2 \alpha_1 \mathbf{e}_1 \cdot \mathbf{y} + \lambda_2^2 \alpha_2 \mathbf{e}_2 \cdot \mathbf{y} + \lambda_3^2 \alpha_3 \mathbf{e}_3 \cdot \mathbf{y} &= 0 \end{aligned}$$

Set $x_i = \alpha_i \mathbf{e}_i \cdot \mathbf{y}$. Then all x_i must be zero by our Vandermonde argument. Therefore $(\forall \mathbf{y})(\alpha_i \mathbf{e}_i \mathbf{y} = 0)$. In particular over the real numbers $\mathbf{y} = \mathbf{e}_i, \mathbf{e}_i \cdot \mathbf{e}_i \neq 0 \Rightarrow \alpha_i = 0$. In general choose $\mathbf{y} \notin \mathbf{e}_i^{\perp}$.

Problem 7.11. *Find a curve in $\mathbb{R}^n$ such that any n points on the curve are linearly independent. $f : \mathbb{R} \to \mathbb{R}^n$. $f(t) := (f_1(t), \dots, f_n(t))$ where $f_i(t) : \mathbb{R} \to \mathbb{R}$.*

Solution. Set $f(t) = (1, t, t^2, \dots, t^n)$ is the moment curve.

$$\begin{pmatrix} f(t_1) \\ \vdots \\ f(t_n) \end{pmatrix} = \begin{pmatrix} 1 & t_1 & \cdots & t_1^{n-1} \\ \vdots & & & \vdots \\ 1 & t_n & \cdots & t_n^{n-1} \end{pmatrix}$$

has nonzero Vandermonde determinant.

Aside: to make this bounded we can use the arctangent function – it will rescale $\mathbb{R}$ a bounded open interval.

7.3 Relations between the roots and the coefficients of a polynomial

$$
\begin{aligned}
f(x) &= (x-\lambda_1)\cdots(x-\lambda_n) = \sum_{i=0}^{n} a_i x^i \\
a_n &= 1 \\
a_{n-1} &= -\sum_{i=1}^{n} \lambda_i \\
a_{n-2} &= \sum_{1\le i<j\le n} \lambda_i\lambda_j \text{ (there are } \binom{n}{2} \text{ terms)} \\
&\vdots \\
a_{n-i} &= (-1)^i \sum_{1\le j_1<\cdots<j_i\le n} \prod \lambda_{j_\ell} \text{ (there are } \binom{n}{i} \text{ terms.)} \\
&\vdots \\
a_0 &= f(0) = (-1)^n \prod_{i=1}^{n} \lambda_i
\end{aligned}
$$

Notation 7.12. *The elementary symmetric polynomials:*

$$
\begin{aligned}
\sigma_1(x_1,\ldots,x_n) &= x_1+\cdots+x_n \\
\sigma_2(x_1,\ldots,x_n) &= x_1x_2+x_1x_3+\cdots+x_{n-1}x_n = \sum_{1\le i<j\le n} x_ix_j \\
\sigma_i(x_1,\ldots,x_n) &= \sum \text{all } i\text{-wise products of the } x_i \text{ (} \binom{n}{i} \text{ terms.)}
\end{aligned}
$$

So $a_{n-i} = (-1)^i\sigma_i(\lambda_1,\ldots,\lambda_n)$. This is the relationship between roots and coefficients of a polynomial.

We apply this to the characteristic polynomial.

For $A \in M_n(F)$ we have $f_A(x) = \prod_{i=1}^n (x - \lambda_i) = \sum_{i=0}^n a_i x^i$. Then

$$\begin{aligned}
a_n &= 1 \\
a_{n-1} &= -\operatorname{tr}(A) \\
&\vdots \\
a_{n-i} &= (-1)^i \sum \det(i \times i \text{ symmetric submatrices}) \\
&\vdots \\
a_0 &= (-1)^n \det(A).
\end{aligned}$$

Hence,

$$\begin{aligned}
\operatorname{tr}(A) &= \sum_{i=1}^n \lambda_i = \sigma_1(\lambda_1, \dots, \lambda_n) \\
\sum \det(2 \times 2 \text{ symmetric submatrices}) &= \sum_{i<j} \lambda_i \lambda_j = \sigma_2(\lambda_1, \dots, \lambda_n) \\
\det(A) &= \prod_{i=1}^n \lambda_i.
\end{aligned}$$

7.4 Real roots of polynomials

Exercise 7.13. $x^{100} + 5x^{99} + 13x^{98} + \cdots = 0$ all remaining coefficients real. Prove this polynomial must have roots that are not real, whatever the remaining coefficients are.

If $f(x) \in \mathbb{R}[x], a_n = 1$, write $f(x) = \prod(x - \lambda_i)\lambda_i \in \mathbb{C}$.

Exercise 7.14. If λ is a root, then $\overline{\lambda}$ (the complex conjugate: $\overline{a + ib} = a - ib$) is also a root with the same multiplicity.

$(x - \lambda)(x - \overline{\lambda}) = x^2 - (\lambda + \overline{\lambda})x + \lambda\overline{\lambda} = x^2 - 2\Re(\lambda) + |\lambda|^2$ (where $\Re z =$ the real part of z.) This implies

$$f(x) = \text{ product of real polynomials of degree 2 without real roots } \cdot \prod_{\lambda_i \in \mathbb{R}} (x - \lambda_i).$$

Corollary 7.15. *A real polynomial of odd degree has a real root.*

A calculus proof uses the intermediate value theorem.

7.5 Orthonormality and Sense-preservation

Definition 7.16. An *othornomal basis* is a basis of pairwise orthogonal unit vectors.

An orthonormal basis for $\mathbb{R}^n$ has $e_i \cdot e_j = 1$ if $i = j, 0$ otherwise.

Definition 7.17. A *congruence* is a transformation that preserves orthonomality.

If φ is a congruence, and $A = [\varphi]_{\mathbf{e}} = [[\varphi \mathbf{e}_1]_{\mathbf{e}} \cdots [\varphi \mathbf{e}_n]_{\mathbf{e}}] = [\mathbf{a}_1 \ldots \mathbf{a}_n]$ then $\mathbf{a}_i \cdot \mathbf{a}_j = 1$ if $i = j, 0$ if $i \neq j$. This is the same as $A^T \cdot A = I$.

Definition 7.18. A real $n \times n$ matrix A is called an orthogonal matrix if $A^T A = I$.

Question: what is $\det A$?

$\det(A^T A) = \det I = 1$. So $\det(A^T)\det(A) = 1$. But $\det(A^T) = \det(A)$, so $\det(A)^2 = 1 \Rightarrow \det(A) = \pm 1$.

A "sense preserving" transformation is a congruence that can be deformed little by little into the identity matrix. Therefore a "sense preserving" matrix must have the same determinant as the identity matrix (by continuity), i.e. 1.

Claim: in $\mathbb{R}^3$ every sense preserving congruence is a rotation. (We are excluding translations, i.e. the origin is fixed.)

Lemma 7.19 (Key Lemma). *For every sense preserving congruence in three dimensions, there is an eigenvector with eigenvalue 1, i.e. $\lambda = 1$ is an eigenvalue.*

Proof: $\deg = 3 \Rightarrow \exists$ a real root λ. Because it is a congruence, $\lambda = \pm 1$ because length of $x =$ length of λx.

Write $f_A(x) = (x - \lambda_1)(x - \lambda_2)(x - \lambda_3)$. If all λ_i are real, then they are all ± 1. If all of them are -1 then $\det A = -1$. So one must be 1. If not all are real, then $\lambda_1 = \pm 1, \lambda_2 = \overline{\lambda_3}$. Then $\lambda_2 \cdot \lambda_3 = \lambda_2 \cdot \overline{\lambda_2} = |\lambda|^2 > 0$. But $\lambda_1 \cdot |\lambda_2|^2 > 0 \Rightarrow \lambda_1 > 0$. $\square$

Exercise 7.20. Finish proof that every sense-preserving congruence in $\mathbb{R}^3$ is a rotation.