Part IA — Vectors and Matrices

Theorems with proof

Based on lectures by N. Peake

Notes taken by Dexter Chua

Michaelmas 2014

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.

Complex numbers

Review of complex numbers, including complex conjugate, inverse, modulus, argument and Argand diagram. Informal treatment of complex logarithm, n-th roots and complex powers. de Moivre's theorem. [2]

Vectors

Review of elementary algebra of vectors in $\mathbb{R}^3$, including scalar product. Brief discussion of vectors in $\mathbb{R}^n$ and $\mathbb{C}^n$; scalar product and the Cauchy-Schwarz inequality. Concepts of linear span, linear independence, subspaces, basis and dimension.

Suffix notation: including summation convention, δ_{ij} and ε_{ijk}. Vector product and triple product: definition and geometrical interpretation. Solution of linear vector equations. Applications of vectors to geometry, including equations of lines, planes and spheres. [5]

Matrices

Elementary algebra of 3×3 matrices, including determinants. Extension to $n \times n$ complex matrices. Trace, determinant, non-singular matrices and inverses. Matrices as linear transformations; examples of geometrical actions including rotations, reflections, dilations, shears; kernel and image. [4]

Simultaneous linear equations: matrix formulation; existence and uniqueness of solutions, geometric interpretation; Gaussian elimination. [3]

Symmetric, anti-symmetric, orthogonal, hermitian and unitary matrices. Decomposition of a general matrix into isotropic, symmetric trace-free and antisymmetric parts. [1]

Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors; geometric significance. [2]

Proof that eigenvalues of hermitian matrix are real, and that distinct eigenvalues give an orthogonal basis of eigenvectors. The effect of a general change of basis (similarity transformations). Diagonalization of general matrices: sufficient conditions; examples of matrices that cannot be diagonalized. Canonical forms for 2×2 matrices. [5]

Discussion of quadratic forms, including change of basis. Classification of conics, cartesian and polar forms. [1]

Rotation matrices and Lorentz transformations as transformation groups. [1]

Contents

0 Introduction

1 Complex numbers

1.1 Basic properties

Proposition. $z\bar{z} = a^2 + b^2 = |z|^2$.

Proposition. $z^{-1} = \bar{z}/|z|^2$.

Theorem (Triangle inequality)**.** For all $z_1, z_2 \in \mathbb{C}$, we have

$$|z_1 + z_2| \leq |z_1| + |z_2|.$$

Alternatively, we have $|z_1 - z_2| \geq ||z_1| - |z_2||$.

1.2 Complex exponential function

Lemma.

$$\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} a_{mn} = \sum_{r=0}^{\infty}\sum_{m=0}^{r} a_{r-m,m}$$

Proof.

$$\begin{aligned}
\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} a_{mn} &= a_{00} + a_{01} + a_{02} + \cdots \\
&\quad + a_{10} + a_{11} + a_{12} + \cdots \\
&\quad + a_{20} + a_{21} + a_{22} + \cdots \\
&= (a_{00}) + (a_{10} + a_{01}) + (a_{20} + a_{11} + a_{02}) + \cdots \\
&= \sum_{r=0}^{\infty}\sum_{m=0}^{r} a_{r-m,m}
\end{aligned}$$

□

Theorem. $\exp(z_1)\exp(z_2) = \exp(z_1 + z_2)$

Proof.

$$\begin{aligned}
\exp(z_1)\exp(z_2) &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{z_1^m}{m!}\frac{z_2^n}{n!} \\
&= \sum_{r=0}^{\infty}\sum_{m=0}^{r} \frac{z_1^{r-m}}{(r-m)!}\frac{z_2^m}{m!} \\
&= \sum_{r=0}^{\infty}\frac{1}{r!}\sum_{m=0}^{r} \frac{r!}{(r-m)!m!} z_1^{r-m} z_2^m \\
&= \sum_{r=0}^{\infty}\frac{(z_1 + z_2)^r}{r!}
\end{aligned}$$

□

Theorem. $e^{iz} = \cos z + i \sin z$.

Proof.

$$\begin{aligned} e^{iz} &= \sum_{n=0}^{\infty} \frac{i^n}{n!} z^n \\ &= \sum_{n=0}^{\infty} \frac{i^{2n}}{(2n)!} z^{2n} + \sum_{n=0}^{\infty} \frac{i^{2n+1}}{(2n+1)!} z^{2n+1} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} + i \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1} \\ &= \cos z + i \sin z \end{aligned}$$

□

1.3 Roots of unity

Proposition. If $\omega = \exp\left(\frac{2\pi i}{n}\right)$, then $1 + \omega + \omega^2 + \cdots + \omega^{n-1} = 0$

Proof. Two proofs are provided:

(i) Consider the equation $z^n = 1$. The coefficient of z^{n-1} is the sum of all roots. Since the coefficient of z^{n-1} is 0, then the sum of all roots $= 1 + \omega + \omega^2 + \cdots + \omega^{n-1} = 0$.

(ii) Since $\omega^n - 1 = (\omega - 1)(1 + \omega + \cdots + \omega^{n-1})$ and $\omega \neq 1$, dividing by $(\omega - 1)$, we have $1 + \omega + \cdots + \omega^{n-1} = (\omega^n - 1)/(\omega - 1) = 0$. □

1.4 Complex logarithm and power

1.5 De Moivre's theorem

Theorem (De Moivre's theorem)**.**

$$\cos n\theta + i \sin n\theta = (\cos\theta + i\sin\theta)^n.$$

Proof. First prove for the $n \geq 0$ case by induction. The $n = 0$ case is true since it merely reads $1 = 1$. We then have

$$\begin{aligned} (\cos\theta + i\sin\theta)^{n+1} &= (\cos\theta + i\sin\theta)^n(\cos\theta + i\sin\theta) \\ &= (\cos n\theta + i\sin n\theta)(\cos\theta + i\sin\theta) \\ &= \cos(n+1)\theta + i\sin(n+1)\theta \end{aligned}$$

If $n < 0$, let $m = -n$. Then $m > 0$ and

$$\begin{aligned} (cos\theta + i\sin\theta)^{-m} &= (\cos m\theta + i\sin m\theta)^{-1} \\ &= \frac{\cos m\theta - i\sin m\theta}{(\cos m\theta + i\sin m\theta)(\cos m\theta - i\sin m\theta)} \\ &= \frac{\cos(-m\theta) + i\sin(-m\theta)}{\cos^2 m\theta + \sin^2 m\theta} \\ &= \cos(-m\theta) + i\sin(-m\theta) \\ &= \cos n\theta + i\sin n\theta \end{aligned}$$

□

1.6 Lines and circles in $\mathbb{C}$

Theorem (Equation of straight line)**.** The equation of a straight line through z_0 and parallel to w is given by

$$z\bar{w} - \bar{z}w = z_0\bar{w} - \bar{z}_0 w.$$

Theorem. The general equation of a circle with center $c \in \mathbb{C}$ and radius $\rho \in \mathbb{R}^+$ can be given by

$$z\bar{z} - \bar{c}z - c\bar{z} = \rho^2 - c\bar{c}.$$

2 Vectors

2.1 Definition and basic properties

2.2 Scalar product

2.2.1 Geometric picture ($\mathbb{R}^2$ and $\mathbb{R}^3$ only)

2.2.2 General algebraic definition

2.3 Cauchy-Schwarz inequality

Theorem (Cauchy-Schwarz inequality)**.** For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|.$$

Proof. Consider the expression $|\mathbf{x} - \lambda\mathbf{y}|^2$. We must have

$$\begin{aligned} |\mathbf{x} - \lambda\mathbf{y}|^2 &\geq 0 \\ (\mathbf{x} - \lambda\mathbf{y}) \cdot (\mathbf{x} - \lambda\mathbf{y}) &\geq 0 \\ \lambda^2|\mathbf{y}|^2 - \lambda(2\mathbf{x} \cdot \mathbf{y}) + |\mathbf{x}|^2 &\geq 0. \end{aligned}$$

Viewing this as a quadratic in λ, we see that the quadratic is non-negative and thus cannot have 2 real roots. Thus the discriminant $\Delta \leq 0$. So

$$\begin{aligned} 4(\mathbf{x} \cdot \mathbf{y})^2 &\leq 4|\mathbf{y}|^2|\mathbf{x}|^2 \\ (\mathbf{x} \cdot \mathbf{y})^2 &\leq |\mathbf{x}|^2|\mathbf{y}|^2 \\ |\mathbf{x} \cdot \mathbf{y}| &\leq |\mathbf{x}||\mathbf{y}|. \end{aligned}$$ □

Corollary (Triangle inequality)**.**

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|.$$

Proof.

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &= (\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) \\ &= |\mathbf{x}|^2 + 2\mathbf{x} \cdot \mathbf{y} + |\mathbf{y}|^2 \\ &\leq |\mathbf{x}|^2 + 2|\mathbf{x}||\mathbf{y}| + |\mathbf{y}|^2 \\ &= (|\mathbf{x}| + |\mathbf{y}|)^2. \end{aligned}$$

So

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|.$$ □

2.4 Vector product

Proposition.

$$\begin{aligned} \mathbf{a} \times \mathbf{b} &= (a_1\hat{\mathbf{i}} + a_2\hat{\mathbf{j}} + a_3\hat{\mathbf{k}}) \times (b_1\hat{\mathbf{i}} + b_2\hat{\mathbf{j}} + b_3\hat{\mathbf{k}}) \\ &= (a_2b_3 - a_3b_2)\hat{\mathbf{i}} + \cdots \\ &= \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \end{aligned}$$

2.5 Scalar triple product

Proposition. If a parallelepiped has sides represented by vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ that form a right-handed system, then the volume of the parallelepiped is given by $[\mathbf{a}, \mathbf{b}, \mathbf{c}]$.

Proof. The area of the base of the parallelepiped is given by $|\mathbf{b}||\mathbf{c}|\sin\theta = |\mathbf{b}\times\mathbf{c}|$. Thus the volume$= |\mathbf{b}\times\mathbf{c}||\mathbf{a}|\cos\phi = |\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})|$, where ϕ is the angle between $\mathbf{a}$ and the normal to $\mathbf{b}$ and $\mathbf{c}$. However, since $\mathbf{a}, \mathbf{b}, \mathbf{c}$ form a right-handed system, we have $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) \geq 0$. Therefore the volume is $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$. □

Theorem. $\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \mathbf{a}\times\mathbf{b} + \mathbf{a}\times\mathbf{c}$.

Proof. Let $\mathbf{d} = \mathbf{a}\times(\mathbf{b}+\mathbf{c}) - \mathbf{a}\times\mathbf{b} - \mathbf{a}\times\mathbf{c}$. We have

$$\begin{aligned}\mathbf{d}\cdot\mathbf{d} &= \mathbf{d}\cdot[\mathbf{a}\times(\mathbf{b}+\mathbf{c})] - \mathbf{d}\cdot(\mathbf{a}\times\mathbf{b}) - \mathbf{d}\cdot(\mathbf{a}\times\mathbf{c})\\ &= (\mathbf{b}+\mathbf{c})\cdot(\mathbf{d}\times\mathbf{a}) - \mathbf{b}\cdot(\mathbf{d}\times\mathbf{a}) - \mathbf{c}\cdot(\mathbf{d}\times\mathbf{a})\\ &= 0\end{aligned}$$

Thus $\mathbf{d} = \mathbf{0}$. □

2.6 Spanning sets and bases

2.6.1 2D space

Theorem. The coefficients λ, μ are unique.

Proof. Suppose that $\mathbf{r} = \lambda\mathbf{a} + \mu\mathbf{b} = \lambda'\mathbf{a} + \mu'\mathbf{b}$. Take the vector product with $\mathbf{a}$ on both sides to get $(\mu - \mu')\mathbf{a}\times\mathbf{b} = \mathbf{0}$. Since $\mathbf{a}\times\mathbf{b} \neq \mathbf{0}$, then $\mu = \mu'$. Similarly, $\lambda = \lambda'$. □

2.6.2 3D space

Theorem. If $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ are non-coplanar, i.e. $\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) \neq 0$, then they form a basis of $\mathbb{R}^3$.

Proof. For any $\mathbf{r}$, write $\mathbf{r} = \lambda\mathbf{a} + \mu\mathbf{b} + \nu\mathbf{c}$. Performing the scalar product with $\mathbf{b}\times\mathbf{c}$ on both sides, one obtains $\mathbf{r}\cdot(\mathbf{b}\times\mathbf{c}) = \lambda\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) + \mu\mathbf{b}\cdot(\mathbf{b}\times\mathbf{c}) + \nu\mathbf{c}\cdot(\mathbf{b}\times\mathbf{c}) = \lambda[\mathbf{a}, \mathbf{b}, \mathbf{c}]$. Thus $\lambda = [\mathbf{r}, \mathbf{b}, \mathbf{c}]/[\mathbf{a}, \mathbf{b}, \mathbf{c}]$. The values of μ and ν can be found similarly. Thus each $\mathbf{r}$ can be written as a linear combination of $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$.

By the formula derived above, it follows that if $\alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c} = \mathbf{0}$, then $\alpha = \beta = \gamma = 0$. Thus they are linearly independent. □

2.6.3 $\mathbb{R}^n$ space

2.6.4 $\mathbb{C}^n$ space

2.7 Vector subspaces

2.8 Suffix notation

Proposition. $(\mathbf{a}\times\mathbf{b})_i = \varepsilon_{ijk}a_jb_k$

Proof. By expansion of formula □

Theorem. $\varepsilon_{ijk}\varepsilon_{ipq} = \delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp}$

Proof. Proof by exhaustion:

$$\text{RHS} = \begin{cases} +1 & \text{if } j = p \text{ and } k = q \\ -1 & \text{if } j = q \text{ and } k = p \\ 0 & \text{otherwise} \end{cases}$$

LHS: Summing over i, the only non-zero terms are when $j, k \neq i$ and $p, q \neq i$. If $j = p$ and $k = q$, LHS is $(-1)^2$ or $(+1)^2 = 1$. If $j = q$ and $k = p$, LHS is $(+1)(-1)$ or $(-1)(+1) = -1$. All other possibilities result in 0. □

Proposition.

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a})$$

Proof. In suffix notation, we have

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = a_i(\mathbf{b} \times \mathbf{c})_i = \varepsilon_{ijk}b_jc_ka_i = \varepsilon_{jki}b_jc_ka_i = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}).$$ □

Theorem (Vector triple product)**.**

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}.$$

Proof.

$$\begin{aligned} [\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_i &= \varepsilon_{ijk}a_j(b \times c)_k \\ &= \varepsilon_{ijk}\varepsilon_{kpq}a_jb_pc_q \\ &= \varepsilon_{ijk}\varepsilon_{pqk}a_jb_pc_q \\ &= (\delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp})a_jb_pc_q \\ &= a_jb_ic_j - a_jc_ib_j \\ &= (\mathbf{a} \cdot \mathbf{c})b_i - (\mathbf{a} \cdot \mathbf{b})c_i \end{aligned}$$ □

Proposition. $(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{a} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{c}) - (\mathbf{a} \cdot \mathbf{b})(\mathbf{a} \cdot \mathbf{c})$.

Proof.

$$\begin{aligned} \text{LHS} &= (\mathbf{a} \times \mathbf{b})_i(\mathbf{a} \times \mathbf{c})_i \\ &= \varepsilon_{ijk}a_jb_k\varepsilon_{ipq}a_pc_q \\ &= (\delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp})a_jb_ka_pc_q \\ &= a_jb_ka_jc_k - a_jb_ka_kc_j \\ &= (\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{c}) - (\mathbf{a} \cdot \mathbf{b})(\mathbf{a} \cdot \mathbf{c}) \end{aligned}$$ □

2.9 Geometry

2.9.1 Lines

Theorem. The equation of a straight line through $\mathbf{a}$ and parallel to $\mathbf{t}$ is

$$(\mathbf{x} - \mathbf{a}) \times \mathbf{t} = \mathbf{0} \text{ or } \mathbf{x} \times \mathbf{t} = \mathbf{a} \times \mathbf{t}.$$

2.9.2 Plane

Theorem. The equation of a plane through $\mathbf{b}$ with normal $\mathbf{n}$ is given by

$$\mathbf{x} \cdot \mathbf{n} = \mathbf{b} \cdot \mathbf{n}.$$

2.10 Vector equations

3 Linear maps

3.1 Examples

3.1.1 Rotation in $\mathbb{R}^3$

3.1.2 Reflection in $\mathbb{R}^3$

3.2 Linear Maps

Theorem. Consider a linear map $f : U \to V$, where U, V are vector spaces. Then $\text{im}(f)$ is a subspace of V, and $\ker(f)$ is a subspace of U.

Proof. Both are non-empty since $f(\mathbf{0}) = \mathbf{0}$.

If $\mathbf{x}, \mathbf{y} \in \text{im}(f)$, then $\exists \mathbf{a}, \mathbf{b} \in U$ such that $\mathbf{x} = f(\mathbf{a}), \mathbf{y} = f(\mathbf{b})$. Then $\lambda\mathbf{x} + \mu\mathbf{y} = \lambda f(\mathbf{a}) + \mu f(\mathbf{b}) = f(\lambda\mathbf{a} + \mu\mathbf{b})$. Now $\lambda\mathbf{a} + \mu\mathbf{b} \in U$ since U is a vector space, so there is an element in U that maps to $\lambda\mathbf{x} + \mu\mathbf{y}$. So $\lambda\mathbf{x} + \mu\mathbf{y} \in \text{im}(f)$ and $\text{im}(f)$ is a subspace of V.

Suppose $\mathbf{x}, \mathbf{y} \in \ker(f)$, i.e. $f(\mathbf{x}) = f(\mathbf{y}) = \mathbf{0}$. Then $f(\lambda\mathbf{x} + \mu\mathbf{y}) = \lambda f(\mathbf{x}) + \mu f(\mathbf{y}) = \lambda\mathbf{0} + \mu\mathbf{0} = \mathbf{0}$. Therefore $\lambda\mathbf{x} + \mu\mathbf{y} \in \ker(f)$. □

3.3 Rank and nullity

Theorem (Rank-nullity theorem)**.** For a linear map $f : U \to V$,

$$r(f) + n(f) = \dim(U).$$

Proof. (Non-examinable) Write $\dim(U) = n$ and $n(f) = m$. If $m = n$, then f is the zero map, and the proof is trivial, since $r(f) = 0$. Otherwise, assume $m < n$.

Suppose $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m\}$ is a basis of $\ker f$, Extend this to a basis of the whole of U to get $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_m, \mathbf{e}_{m+1}, \cdots, \mathbf{e}_n\}$. To prove the theorem, we need to prove that $\{f(\mathbf{e}_{m+1}), f(\mathbf{e}_{m+2}), \cdots f(\mathbf{e}_n)\}$ is a basis of $\text{im}(f)$.

(i) First show that it spans $\text{im}(f)$. Take $\mathbf{y} \in \text{im}(f)$. Thus $\exists \mathbf{x} \in U$ such that $\mathbf{y} = f(\mathbf{x})$. Then

$$\mathbf{y} = f(\alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2 + \cdots + \alpha_n\mathbf{e}_n),$$

since $\mathbf{e}_1, \cdots \mathbf{e}_n$ is a basis of U. Thus

$$\mathbf{y} = \alpha_1 f(\mathbf{e}_1) + \alpha_2 f(\mathbf{e}_2) + \cdots + \alpha_m f(\mathbf{e}_m) + \alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n).$$

The first m terms map to $\mathbf{0}$, since $\mathbf{e_1}, \cdots \mathbf{e_m}$ is the basis of the kernel of f. Thus

$$\mathbf{y} = \alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n).$$

(ii) To show that they are linearly independent, suppose

$$\alpha_{m+1} f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n) = \mathbf{0}.$$

Then

$$f(\alpha_{m+1}\mathbf{e}_{m+1} + \cdots + \alpha_n\mathbf{e}_n) = \mathbf{0}.$$

Thus $\alpha_{m+1}\mathbf{e}_{m+1} + \cdots + \alpha_n\mathbf{e}_n \in \ker(f)$. Since $\{\mathbf{e}_1, \cdots, \mathbf{e}_m\}$ span $\ker(f)$, there exist some $\alpha_1, \alpha_2, \cdots \alpha_m$ such that

$$\alpha_{m+1}\mathbf{e}_{m+1} + \cdots + \alpha_n\mathbf{e}_n = \alpha_1\mathbf{e_1} + \cdots + \alpha_m\mathbf{e}_m.$$

But $\mathbf{e}_1 \cdots \mathbf{e}_n$ is a basis of U and are linearly independent. So $\alpha_i = 0$ for all i. Then the only solution to the equation $\alpha_{m+1}f(\mathbf{e}_{m+1}) + \cdots + \alpha_n f(\mathbf{e}_n) = \mathbf{0}$ is $\alpha_i = 0$, and they are linearly independent by definition. □

3.4 Matrices

3.4.1 Examples

3.4.2 Matrix Algebra

Proposition.

(i) $(A^T)^T = A$.

(ii) If $\mathbf{x}$ is a column vector $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $\mathbf{x}^T$ is a row vector $(x_1\ x_2 \cdots x_n)$.

(iii) $(AB)^T = B^TA^T$ since $(AB)^T_{ij} = (AB)_{ji} = A_{jk}B_{ki} = B_{ki}A_{jk}$ $= (B^T)_{ik}(A^T)_{kj} = (B^TA^T)_{ij}$.

Proposition. $\operatorname{tr}(BC) = \operatorname{tr}(CB)$

Proof. $\operatorname{tr}(BC) = B_{ik}C_{ki} = C_{ki}B_{ik} = (CB)_{kk} = \operatorname{tr}(CB)$ □

3.4.3 Decomposition of an $n \times n$ matrix

3.4.4 Matrix inverse

Proposition. $(AB)^{-1} = B^{-1}A^{-1}$

Proof. $(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}B = I.$ □

3.5 Determinants

3.5.1 Permutations

Proposition. Any q-cycle can be written as a product of 2-cycles.

Proof. $(1\ 2\ 3\ \cdots\ n) = (1\ 2)(2\ 3)(3\ 4)\cdots(n-1\ n)$. □

Proposition.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

3.5.2 Properties of determinants

Proposition. $\det(A) = \det(A^T)$.

Proof. Take a single term $A_{\sigma(1)1}A_{\sigma(2)2}\cdots A_{\sigma(n)n}$ and let ρ be another permutation in S_n. We have

$$A_{\sigma(1)1}A_{\sigma(2)2}\cdots A_{\sigma(n)n} = A_{\sigma(\rho(1))\rho(1)}A_{\sigma(\rho(2))\rho(2)}\cdots A_{\sigma(\rho(n))\rho(n)}$$

since the right hand side is just re-ordering the order of multiplication. Choose $\rho = \sigma^{-1}$ and note that $\varepsilon(\sigma) = \varepsilon(\rho)$. Then

$$\det(A) = \sum_{\rho\in S_n} \varepsilon(\rho)A_{1\rho(1)}A_{2\rho(2)}\cdots A_{n\rho(n)} = \det(A^T). \qquad \square$$

Proposition. If matrix B is formed by multiplying every element in a single row of A by a scalar λ, then $\det(B) = \lambda\det(A)$. Consequently, $\det(\lambda A) = \lambda^n\det(A)$.

Proof. Each term in the sum is multiplied by λ, so the whole sum is multiplied by λ^n. $\square$

Proposition. If 2 rows (or 2 columns) of A are identical, the determinant is 0.

Proof. wlog, suppose columns 1 and 2 are the same. Then

$$\det(A) = \sum_{\sigma\in S_n} \varepsilon(\sigma)A_{\sigma(1)1}A_{\sigma(2)2}\cdots A_{\sigma(n)n}.$$

Now write an arbitrary σ in the form $\sigma = \rho(1\ 2)$. Then $\varepsilon(\sigma) = \varepsilon(\rho)\varepsilon((1\ 2)) = -\varepsilon(\rho)$. So

$$\det(A) = \sum_{\rho\in S_n} -\varepsilon(\rho)A_{\rho(2)1}A_{\rho(1)2}A_{\rho(3)3}\cdots A_{\rho(n)n}.$$

But columns 1 and 2 are identical, so $A_{\rho(2)1} = A_{\rho(2)2}$ and $A_{\rho(1)2} = A_{\rho(1)1}$. So $\det(A) = -\det(A)$ and $\det(A) = 0$. $\square$

Proposition. If 2 rows or 2 columns of a matrix are linearly dependent, then the determinant is zero.

Proof. Suppose in A, (column r) + λ(column s) = 0. Define

$$B_{ij} = \begin{cases} A_{ij} & j \neq r \\ A_{ij} + \lambda A_{is} & j = r \end{cases}.$$

Then $\det(B) = \det(A) + \lambda\det(\text{matrix with column } r = \text{column } s) = \det(A)$. Then we can see that the rth column of B is all zeroes. So each term in the sum contains one zero and $\det(A) = \det(B) = 0$. $\square$

Proposition. Given a matrix A, if B is a matrix obtained by adding a multiple of a column (or row) of A to another column (or row) of A, then $\det A = \det B$.

Corollary. Swapping two rows or columns of a matrix negates the determinant.

Proof. We do the column case only. Let $A = (\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n)$. Then

$$\begin{aligned}\det(\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n) &= \det(\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{a}_j \cdots \mathbf{a}_j \cdots \mathbf{a}_n)\\ &= \det(\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{a}_j \cdots \mathbf{a}_j - (\mathbf{a}_i + \mathbf{a}_j) \cdots \mathbf{a}_n)\\ &= \det(\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{a}_j \cdots - \mathbf{a}_i \cdots \mathbf{a}_n)\\ &= \det(\mathbf{a}_1 \cdots \mathbf{a}_j \cdots - \mathbf{a}_i \cdots \mathbf{a}_n)\\ &= -\det(\mathbf{a}_1 \cdots \mathbf{a}_j \cdots \mathbf{a}_i \cdots \mathbf{a}_n)\end{aligned}$$

Alternatively, we can prove this from the definition directly, using the fact that the sign of a transposition is -1 (and that the sign is multiplicative). □

Proposition. $\det(AB) = \det(A)\det(B)$.

Proof. First note that $\sum_\sigma \varepsilon(\sigma) A_{\sigma(1)\rho(1)} A_{\sigma(2)\rho(2)} = \varepsilon(\rho)\det(A)$, i.e. swapping columns (or rows) an even/odd number of times gives a factor ± 1 respectively. We can prove this by writing $\sigma = \mu\rho$.

Now

$$\begin{aligned}\det AB &= \sum_\sigma \varepsilon(\sigma)(AB)_{\sigma(1)1}(AB)_{\sigma(2)2}\cdots(AB)_{\sigma(n)n}\\ &= \sum_\sigma \varepsilon(\sigma) \sum_{k_1, k_2, \cdots, k_n}^{n} A_{\sigma(1)k_1} B_{k_1 1} \cdots A_{\sigma(n)k_n} B_{k_n n}\\ &= \sum_{k_1, \cdots, k_n} B_{k_1 1} \cdots B_{k_n n} \underbrace{\sum_\sigma \varepsilon(\sigma) A_{\sigma(1)k_1} A_{\sigma(2)k_2} \cdots A_{\sigma(n)k_n}}_{S}\end{aligned}$$

Now consider the many different S's. If in S, two of k_1 and k_n are equal, then S is a determinant of a matrix with two columns the same, i.e. $S = 0$. So we only have to consider the sum over distinct k_is. Thus the k_is are are a permutation of $1, \cdots n$, say $k_i = \rho(i)$. Then we can write

$$\begin{aligned}\det AB &= \sum_\rho B_{\rho(1)1} \cdots B_{\rho(n)n} \sum_\sigma \varepsilon(\sigma) A_{\sigma(1)\rho(1)} \cdots A_{\sigma(n)\rho(n)}\\ &= \sum_\rho B_{\rho(1)1} \cdots B_{\rho(n)n} (\varepsilon(\rho)\det A)\\ &= \det A \sum_\rho \varepsilon(\rho) B_{\rho(1)1} \cdots B_{\rho(n)n}\\ &= \det A \det B\end{aligned}$$

□

Corollary. If A is orthogonal, $\det A = \pm 1$.

Proof.

$$\begin{aligned}AA^T &= I\\ \det AA^T &= \det I\\ \det A \det A^T &= 1\\ (\det A)^2 &= 1\\ \det A &= \pm 1\end{aligned}$$

□

Corollary. If U is unitary, $|\det U| = 1$.

Proof. We have $\det U^\dagger = (\det U^T)^* = \det(U)^*$. Since $UU^\dagger = I$, we have $\det(U)\det(U)^* = 1$. □

Proposition. In $\mathbb{R}^3$, orthogonal matrices represent either a rotation ($\det = 1$) or a reflection ($\det = -1$).

3.5.3 Minors and Cofactors

Theorem (Laplace expansion formula)**.** For any particular fixed i,

$$\det A = \sum_{j=1}^{n} A_{ji}\Delta_{ji}.$$

Proof.

$$\det A = \sum_{j_i=1}^{n} A_{j_i i} \sum_{j_1,\cdots,\overline{j_i},\cdots j_n}^{n} \varepsilon_{j_1 j_2\cdots j_n} A_{j_1 1}A_{j_2 2}\cdots\overline{A_{j_i i}}\cdots A_{j_n n}$$

Let $\sigma \in S_n$ be the permutation which moves j_i to the ith position, and leave everything else in its natural order, i.e.

$$\sigma = \begin{pmatrix} 1 & \cdots & i & i+1 & i+2 & \cdots & j_i - 1 & j_i & j_i + 1 & \cdots & n \\ 1 & \cdots & j_i & i & i+1 & \cdots & j_i - 2 & j_i - 1 & j_i + 1 & \cdots & n \end{pmatrix}$$

if $j_i > i$, and similarly for other cases. To perform this permutation, $|i - j_i|$ transpositions are made. So $\varepsilon(\sigma) = (-1)^{i - j_i}$.

Now consider the permutation $\rho \in S_n$

$$\rho = \begin{pmatrix} 1 & \cdots & \cdots & \bar{j}_i & \cdots & n \\ j_1 & \cdots & \bar{j}_i & \cdots & \cdots & j_n \end{pmatrix}$$

The composition $\rho\sigma$ reorders $(1, \cdots, n)$ to $(j_1, j_2, \cdots, j_n)$. So $\varepsilon(\rho\sigma) = \varepsilon_{j_1\cdots j_n} = \varepsilon(\rho)\varepsilon(\sigma) = (-1)^{i-j_i}\varepsilon_{j_1\cdots\bar{j}_i\cdots j_n}$. Hence the original equation becomes

$$\begin{aligned} \det A &= \sum_{j_i=1}^{n} A_{j_i i} \sum_{j_1\cdots\bar{j}_i\cdots j_n} (-1)^{i-j_i}\varepsilon_{j_1\cdots\bar{j}_i\cdots j_n} A_{j_1 1}\cdots\overline{A_{j_i i}}\cdots A_{j_n n} \\ &= \sum_{j_i=1}^{n} A_{j_i i}(-1)^{i-j_i} M_{j_i i} \\ &= \sum_{j_i=1}^{n} A_{j_i i}\Delta_{j_i i} \\ &= \sum_{j=1}^{n} A_{ji}\Delta_{ji} \end{aligned}$$

□

4 Matrices and linear equations

4.1 Simple example, 2×2

4.2 Inverse of an $n \times n$ matrix

Lemma. $\sum A_{ik}\Delta_{jk} = \delta_{ij} \det A$.

Proof. If $i \neq j$, then consider an $n \times n$ matrix B, which is identical to A except the jth row is replaced by the ith row of A. So Δ_{jk} of $B = \Delta_{jk}$ of A, since Δ_{jk} does not depend on the elements in row j. Since B has a duplicate row, we know that

$$0 = \det B = \sum_{k=1}^{n} B_{jk}\Delta_{jk} = \sum_{k=1}^{n} A_{ik}\Delta_{jk}.$$

If $i = j$, then the expression is $\det A$ by the Laplace expansion formula. □

Theorem. If $\det A \neq 0$, then A^{-1} exists and is given by

$$(A^{-1})_{ij} = \frac{\Delta_{ji}}{\det A}.$$

Proof.

$$(A^{-1})_{ik}A_{kj} = \frac{\Delta_{ki}}{\det A}A_{kj} = \frac{\delta_{ij}\det A}{\det A} = \delta_{ij}.$$

So $A^{-1}A = I$. □

4.3 Homogeneous and inhomogeneous equations

4.3.1 Gaussian elimination

4.4 Matrix rank

Theorem. The column rank and row rank are equal for any $m \times n$ matrix.

Proof. Let r be the row rank of A. Write the biggest set of linearly independent rows as $\mathbf{v}_1^T, \mathbf{v}_2^T, \cdots \mathbf{v}_r^T$ or in component form $\mathbf{v}_k^T = (v_{k1}, v_{k2}, \cdots, v_{kn})$ for $k = 1, 2, \cdots, r$.

Now denote the ith row of A as $\mathbf{r}_i^T = (A_{i1}, A_{i2}, \cdots A_{in})$.

Note that every row of A can be written as a linear combination of the $\mathbf{v}$'s. (If $\mathbf{r_i}$ cannot be written as a linear combination of the $\mathbf{v}$'s, then it is independent of the $\mathbf{v}$'s and $\mathbf{v}$ is not the maximum collection of linearly independent rows) Write

$$\mathbf{r}_i^T = \sum_{k=1}^{r} C_{ik}\mathbf{v}_k^T.$$

For some coefficients C_{ik} with $1 \leq i \leq m$ and $1 \leq k \leq r$.

Now the elements of A are

$$A_{ij} = (\mathbf{r}_i)_j^T = \sum_{k=1}^{r} C_{ik}(\mathbf{v}_k)_j,$$

or

$$\begin{pmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{mj} \end{pmatrix} = \sum_{k=1}^{r} \mathbf{v}_{kj} \begin{pmatrix} C_{1k} \\ C_{2k} \\ \vdots \\ C_{mk} \end{pmatrix}$$

So every column of A can be written as a linear combination of the r column vectors $\mathbf{c}_k$. Then the column rank of $A \leq r$, the row rank of A.

Apply the same argument to A^T to see that the row rank is $\leq$ the column rank. □

4.5 Homogeneous problem $A\mathbf{x} = \mathbf{0}$

4.5.1 Geometrical interpretation

4.5.2 Linear mapping view of $A\mathbf{x} = \mathbf{0}$

4.6 General solution of $A\mathbf{x} = \mathbf{d}$

5 Eigenvalues and eigenvectors

5.1 Preliminaries and definitions

Theorem (Fundamental theorem of algebra)**.** Let $p(z)$ be a polynomial of degree $m \geq 1$, i.e.

$$p(z) = \sum_{j=0}^{m} c_j z^j,$$

where $c_j \in \mathbb{C}$ and $c_m \neq 0$.

Then $p(z) = 0$ has precisely m (not necessarily distinct) roots in the complex plane, accounting for multiplicity.

Theorem. λ is an eigenvalue of A iff

$$\det(A - \lambda I) = 0.$$

Proof. ($\Rightarrow$) Suppose that λ is an eigenvalue and $\mathbf{x}$ is the associated eigenvector. We can rearrange the equation in the definition above to

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

and thus

$$\mathbf{x} \in \ker(A - \lambda I)$$

But $\mathbf{x} \neq \mathbf{0}$. So $\ker(A - \lambda I)$ is non-trivial and $\det(A - \lambda I) = 0$. The ($\Leftarrow$) direction is similar. $\square$

5.2 Linearly independent eigenvectors

Theorem. Suppose $n \times n$ matrix A has *distinct* eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. Then the corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ are linearly independent.

Proof. Proof by contradiction: Suppose $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ are linearly dependent. Then we can find non-zero constants d_i for $i = 1, 2, \cdots, r$, such that

$$d_1\mathbf{x}_1 + d_2\mathbf{x}_2 + \cdots + d_r\mathbf{x}_r = \mathbf{0}.$$

Suppose that this is the shortest non-trivial linear combination that gives $\mathbf{0}$ (we may need to re-order $\mathbf{x}_i$).

Now apply $(A - \lambda_1 I)$ to the whole equation to obtain

$$d_1(\lambda_1 - \lambda_1)\mathbf{x}_1 + d_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + \cdots + d_r(\lambda_r - \lambda_1)\mathbf{x}_r = \mathbf{0}.$$

We know that the first term is $\mathbf{0}$, while the others are not (since we assumed $\lambda_i \neq \lambda_j$ for $i \neq j$). So

$$d_2(\lambda_2 - \lambda_1)\mathbf{x}_2 + \cdots + d_r(\lambda_r - \lambda_1)\mathbf{x}_r = \mathbf{0},$$

and we have found a shorter linear combination that gives $\mathbf{0}$. Contradiction. $\square$

5.3 Transformation matrices

5.3.1 Transformation law for vectors

Theorem. Denote vector as $\mathbf{u}$ with respect to $\{\mathbf{e}_i\}$ and $\tilde{\mathbf{u}}$ with respect to $\{\tilde{\mathbf{e}}_i\}$. Then

$$\mathbf{u} = P\tilde{\mathbf{u}} \text{ and } \tilde{\mathbf{u}} = P^{-1}\mathbf{u}$$

5.3.2 Transformation law for matrix

Theorem.

$$\tilde{A} = P^{-1}AP.$$

5.4 Similar matrices

Proposition. Similar matrices have the following properties:

(i) Similar matrices have the same determinant.

(ii) Similar matrices have the same trace.

(iii) Similar matrices have the same characteristic polynomial.

Proof. They are proven as follows:

(i) $\det B = \det(P^{-1}AP) = (\det A)(\det P)^{-1}(\det P) = \det A$

(ii)

$$\begin{aligned}\operatorname{tr} B &= B_{ii}\\ &= P_{ij}^{-1}A_{jk}P_{ki}\\ &= A_{jk}P_{ki}P_{ij}^{-1}\\ &= A_{jk}(PP^{-1})_{kj}\\ &= A_{jk}\delta_{kj}\\ &= A_{jj}\\ &= \operatorname{tr} A\end{aligned}$$

(iii)

$$\begin{aligned}p_B(\lambda) &= \det(B - \lambda I)\\ &= \det(P^{-1}AP - \lambda I)\\ &= \det(P^{-1}AP - \lambda P^{-1}IP)\\ &= \det(P^{-1}(A - \lambda I)P)\\ &= \det(A - \lambda I)\\ &= p_A(\lambda)\end{aligned}$$

□

5.5 Diagonalizable matrices

Theorem. Let $\lambda_1, \lambda_2, \cdots, \lambda_r$, with $r \leq n$ be the distinct eigenvalues of A. Let $B_1, B_2, \cdots B_r$ be the bases of the eigenspaces $E_{\lambda_1}, E_{\lambda_2}, \cdots, E_{\lambda_r}$ correspondingly. Then the set $B = \bigcup_{i=1}^{r} B_i$ is linearly independent.

Proof. Write $B_1 = \{\mathbf{x}_1^{(1)}, \mathbf{x}_2^{(1)}, \cdots \mathbf{x}_{m(\lambda_1)}^{(1)}\}$. Then $m(\lambda_1) = \dim(E_{\lambda_1})$, and similarly for all B_i.

Consider the following general linear combination of all elements in B. Consider the equation

$$\sum_{i=1}^{r}\sum_{j=1}^{m(\lambda_i)} \alpha_{ij}\mathbf{x}_j^{(i)} = 0.$$

The first sum is summing over all eigenspaces, and the second sum sums over the basis vectors in B_i. Now apply the matrix

$$\prod_{k=1,2,\cdots,\bar{K},\cdots,r} (A - \lambda_k I)$$

to the above sum, for some arbitrary K. We obtain

$$\sum_{j=1}^{m(\lambda_K)} \alpha_{Kj}\left[\prod_{k=1,2,\cdots,\bar{K},\cdots,r} (\lambda_K - \lambda_k)\right]\mathbf{x}_j^{(K)} = 0.$$

Since the $\mathbf{x}_j^{(K)}$ are linearly independent (B_K is a basis), $\alpha_{Kj} = 0$ for all j. Since K was arbitrary, all α_{ij} must be zero. So B is linearly independent. $\square$

Proposition. A is diagonalizable iff all its eigenvalues have zero defect.

5.6 Canonical (Jordan normal) form

Theorem. Any 2×2 complex matrix A is similar to exactly one of

$$\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}, \begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix}, \begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix}$$

Proof. For each case:

(i) If A has two distinct eigenvalues, then eigenvectors are linearly independent. Then we can use P formed from eigenvectors as its columns

(ii) If $\lambda_1 = \lambda_2 = \lambda$ and $\dim E_\lambda = 2$, then write $E_\lambda = \operatorname{span}\{\mathbf{u}, \mathbf{v}\}$, with $\mathbf{u}, \mathbf{v}$ linearly independent. Now use $\{\mathbf{u}, \mathbf{v}\}$ as a new basis of $\mathbb{C}^2$ and $\tilde{A} = P^{-1}AP = \begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix} = \lambda I$

Note that since $P^{-1}AP = \lambda I$, we have $A = P(\lambda I)P^{-1} = \lambda I$. So A is *isotropic*, i.e. the same with respect to any basis.

(iii) If $\lambda_1 = \lambda_2 = \lambda$ and $\dim(E_\lambda) = 1$, then $E_\lambda = \operatorname{span}\{\mathbf{v}\}$. Now choose basis of $\mathbb{C}^2$ as $\{\mathbf{v}, \mathbf{w}\}$, where $\mathbf{w} \in \mathbb{C}^2 \setminus E_\lambda$.

We know that $A\mathbf{w} \in \mathbb{C}^2$. So $A\mathbf{w} = \alpha\mathbf{v} + \beta\mathbf{w}$. Hence, if we change basis to $\{\mathbf{v}, \mathbf{w}\}$, then $\tilde{A} = P^{-1}AP = \begin{pmatrix}\lambda & \alpha\\ 0 & \beta\end{pmatrix}$.

However, A and $\tilde{A}$ both have eigenvalue λ with algebraic multiplicity 2. So we must have $\beta = \lambda$. To make $\alpha = 1$, let $\mathbf{u} = (\tilde{A} - \lambda I)\mathbf{w}$. We know $\mathbf{u} \neq \mathbf{0}$ since $\mathbf{w}$ is not in the eigenspace. Then

$$(\tilde{A} - \lambda I)\mathbf{u} = (\tilde{A} - \lambda I)^2\mathbf{w} = \begin{pmatrix}0 & \alpha\\ 0 & 0\end{pmatrix}\begin{pmatrix}0 & \alpha\\ 0 & 0\end{pmatrix}\mathbf{w} = \mathbf{0}.$$

So $\mathbf{u}$ is an eigenvector of $\tilde{A}$ with eigenvalue λ.

We have $\mathbf{u} = \tilde{A}\mathbf{w} - \lambda\mathbf{w}$. So $\tilde{A}\mathbf{w} = \mathbf{u} + \lambda\mathbf{w}$.

Change basis to $\{\mathbf{u}, \mathbf{w}\}$. Then A with respect to this basis is $\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix}$.

This is a two-stage process: P sends basis to $\{\mathbf{v}, \mathbf{w}\}$ and then matrix Q sends to basis $\{\mathbf{u}, \mathbf{w}\}$. So the similarity transformation is $Q^{-1}(P^{-1}AP)Q = (PQ)^{-1}A(PQ)$. □

Proposition. (Without proof) The canonical form, or Jordan normal form, exists for any $n \times n$ matrix A. Specifically, there exists a similarity transform such that A is similar to a matrix to $\tilde{A}$ that satisfies the following properties:

(i) $\tilde{A}_{\alpha\alpha} = \lambda_\alpha$, i.e. the diagonal composes of the eigenvalues.

(ii) $\tilde{A}_{\alpha,\alpha+1} = 0$ or 1.

(iii) $\tilde{A}_{ij} = 0$ otherwise.

5.7 Cayley-Hamilton Theorem

Theorem (Cayley-Hamilton theorem)**.** Every $n \times n$ complex matrix satisfies its own characteristic equation.

Proof. We will only prove for diagonalizable matrices here. So suppose for our matrix A, there is some P such that $D = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = P^{-1}AP$. Note that

$$D^i = (P^{-1}AP)(P^{-1}AP)\cdots(P^{-1}AP) = P^{-1}A^iP.$$

Hence

$$p_D(D) = p_D(P^{-1}AP) = P^{-1}[p_D(A)]P.$$

Since similar matrices have the same characteristic polynomial. So

$$p_A(D) = P^{-1}[p_A(A)]P.$$

However, we also know that $D^i = \operatorname{diag}(\lambda_1^i, \lambda_2^i, \cdots \lambda_n^i)$. So

$$p_A(D) = \operatorname{diag}(p_A(\lambda_1), p_A(\lambda_2), \cdots, p_A(\lambda_n)) = \operatorname{diag}(0, 0, \cdots, 0)$$

since the eigenvalues are roots of $p_A(\lambda) = 0$. So $0 = p_A(D) = P^{-1}p_A(A)P$ and thus $p_A(A) = 0$. □

5.8 Eigenvalues and eigenvectors of a Hermitian matrix

5.8.1 Eigenvalues and eigenvectors

Theorem. The eigenvalues of a Hermitian matrix H are real.

Proof. Suppose that H has eigenvalue λ with eigenvector $\mathbf{v} \neq 0$. Then

$$H\mathbf{v} = \lambda\mathbf{v}.$$

We pre-multiply by $\mathbf{v}^\dagger$, a $1 \times n$ row vector, to obtain

$$\mathbf{v}^\dagger H\mathbf{v} = \lambda\mathbf{v}^\dagger\mathbf{v} \qquad (*)$$

We take the Hermitian conjugate of both sides. The left hand side is

$$(\mathbf{v}^\dagger H\mathbf{v})^\dagger = \mathbf{v}^\dagger H^\dagger \mathbf{v} = \mathbf{v}^\dagger H\mathbf{v}$$

since H is Hermitian. The right hand side is

$$(\lambda \mathbf{v}^\dagger \mathbf{v})^\dagger = \lambda^* \mathbf{v}^\dagger \mathbf{v}$$

So we have

$$\mathbf{v}^\dagger H\mathbf{v} = \lambda^* \mathbf{v}^\dagger \mathbf{v}.$$

From $(*)$, we know that $\lambda \mathbf{v}^\dagger \mathbf{v} = \lambda^* \mathbf{v}^\dagger \mathbf{v}$. Since $\mathbf{v} \neq 0$, we know that $\mathbf{v}^\dagger \mathbf{v} = \mathbf{v} \cdot \mathbf{v} \neq 0$. So $\lambda = \lambda^*$ and λ is real. □

Theorem. The eigenvectors of a Hermitian matrix H corresponding to distinct eigenvalues are orthogonal.

Proof. Let

$$H\mathbf{v}_i = \lambda_i \mathbf{v}_i \tag{i}$$
$$H\mathbf{v}_j = \lambda_j \mathbf{v}_j. \tag{ii}$$

Pre-multiply (i) by $\mathbf{v}_j^\dagger$ to obtain

$$\mathbf{v}_j^\dagger H\mathbf{v}_i = \lambda_i \mathbf{v}_j^\dagger \mathbf{v}_i. \tag{iii}$$

Pre-multiply (ii) by $\mathbf{v}_i^\dagger$ and take the Hermitian conjugate to obtain

$$\mathbf{v}_j^\dagger H\mathbf{v}_i = \lambda_j \mathbf{v}_j^\dagger \mathbf{v}_i. \tag{iv}$$

Equating (iii) and (iv) yields

$$\lambda_i \mathbf{v}_j^\dagger \mathbf{v}_i = \lambda_j \mathbf{v}_j^\dagger \mathbf{v}_i.$$

Since $\lambda_i \neq \lambda_j$, we must have $\mathbf{v}_j^\dagger \mathbf{v}_i = 0$. So their inner product is zero and are orthogonal. □

5.8.2 Gram-Schmidt orthogonalization (non-examinable)

5.8.3 Unitary transformation

5.8.4 Diagonalization of $n \times n$ Hermitian matrices

Theorem. An $n \times n$ Hermitian matrix has precisely n orthogonal eigenvectors.

Proof. (Non-examinable) Let $\lambda_1, \lambda_2, \cdots, \lambda_r$ be the distinct eigenvalues of H ($r \leq n$), with a set of corresponding orthonormal eigenvectors $B = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$. Extend to a basis of the whole of $\mathbb{C}^n$

$$B' = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r, \mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_{n-r}\}$$

Now use Gram-Schmidt to create an orthonormal basis

$$\tilde{B} = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r, \mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{n-r}\}.$$

Now write

$$P = \begin{pmatrix} \uparrow & \uparrow & & \uparrow & \uparrow & & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r & \mathbf{u}_1 & \cdots & \mathbf{u}_{n-r} \\ \downarrow & \downarrow & & \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

We have shown above that this is a unitary matrix, i.e. $P^{-1} = P^\dagger$. So if we change basis, we have

$$\begin{aligned} P^{-1}HP &= P^\dagger HP \\ &= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \lambda_r & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & c_{11} & c_{12} & \cdots & c_{1,n-r} \\ 0 & 0 & \cdots & 0 & c_{21} & c_{22} & \cdots & c_{2,n-r} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & c_{n-r,1} & c_{n-r,2} & \cdots & c_{n-r,n-r} \end{pmatrix} \end{aligned}$$

Here C is an $(n-r) \times (n-r)$ Hermitian matrix. The eigenvalues of C are also eigenvalues of H because $\det(H - \lambda I) = \det(P^\dagger HP - \lambda I) = (\lambda_1 - \lambda) \cdots (\lambda_r - \lambda) \det(C - \lambda I)$. So the eigenvalues of C are the eigenvalues of H.

We can keep repeating the process on C until we finish all rows. For example, if the eigenvalues of C are all distinct, there are $n - r$ orthonormal eigenvectors $\mathbf{w}_j$ (for $j = r + 1, \cdots, n$) of C. Let

$$Q = \begin{pmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & 1 & & & & \\ & & & & \uparrow & \uparrow & & \uparrow \\ & & & & \mathbf{w}_{r+1} & \mathbf{w}_{r+2} & \cdots & \mathbf{w}_n \\ & & & & \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

with other entries 0. (where we have a $r \times r$ identity matrix block on the top left corner and a $(n-r) \times (n-r)$ with columns formed by $\mathbf{w}_j$)

Since the columns of Q are orthonormal, Q is unitary. So $Q^\dagger P^\dagger HPQ = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r, \lambda_{r+1}, \cdots, \lambda_n)$, where the first r λs are distinct and the remaining ones are copies of previous ones.

The n linearly-independent eigenvectors are the columns of PQ.

□

5.8.5 Normal matrices

Proposition.

(i) If λ is an eigenvalue of N, then λ^* is an eigenvalue of $N^\dagger$.

(ii) The eigenvectors of distinct eigenvalues are orthogonal.

(iii) A normal matrix can always be diagonalized with an orthonormal basis of eigenvectors.

6 Quadratic forms and conics

Theorem. Hermitian forms are real.

Proof. $(\mathbf{x}^\dagger H\mathbf{x})^* = (\mathbf{x}^\dagger H\mathbf{x})^\dagger = \mathbf{x}^\dagger H^\dagger \mathbf{x} = \mathbf{x}^\dagger H\mathbf{x}$. So $(\mathbf{x}^\dagger H\mathbf{x})^* = \mathbf{x}^\dagger H\mathbf{x}$ and it is real. □

6.1 Quadrics and conics

6.1.1 Quadrics

6.1.2 Conic sections ($n = 2$)

6.2 Focus-directrix property

7 Transformation groups

7.1 Groups of orthogonal matrices

Proposition. The set of all $n \times n$ orthogonal matrices P forms a group under matrix multiplication.

Proof.

0. If P, Q are orthogonal, then consider $R = PQ$. $RR^T = (PQ)(PQ)^T = P(QQ^T)P^T = PP^T = I$. So R is orthogonal.
1. I satisfies $II^T = I$. So I is orthogonal and is an identity of the group.
2. Inverse: if P is orthogonal, then $P^{-1} = P^T$ by definition, which is also orthogonal.
3. Matrix multiplication is associative since function composition is associative. □

7.2 Length preserving matrices

Theorem. Let $P \in O(n)$. Then the following are equivalent:

(i) P is orthogonal

(ii) $|P\mathbf{x}| = |\mathbf{x}|$

(iii) $(P\mathbf{x})^T(P\mathbf{y}) = \mathbf{x}^T\mathbf{y}$, i.e. $(P\mathbf{x}) \cdot (P\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$.

(iv) If $(\mathbf{v}_1, \mathbf{v_2}, \cdots, \mathbf{v}_n)$ are orthonormal, so are $(P\mathbf{v}_1, P\mathbf{v}_2, \cdots, P\mathbf{v}_n)$

(v) The columns of P are orthonormal.

Proof. We do them one by one:

(i) $\Rightarrow$ (ii): $|P\mathbf{x}|^2 = (P\mathbf{x})^T(P\mathbf{x}) = \mathbf{x}^T P^T P\mathbf{x} = \mathbf{x}^T\mathbf{x} = |\mathbf{x}|^2$

(ii) $\Rightarrow$ (iii): $|P(\mathbf{x}+\mathbf{y})|^2 = |\mathbf{x}+\mathbf{y}|^2$. The right hand side is

$$(\mathbf{x}^T + \mathbf{y}^T)(\mathbf{x}+\mathbf{y}) = \mathbf{x}^T\mathbf{x} + y^T\mathbf{y} + \mathbf{y}^T\mathbf{x} + \mathbf{x}^T\mathbf{y} = |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2\mathbf{x}^T\mathbf{y}.$$

Similarly, the left hand side is

$$|P\mathbf{x} + P\mathbf{y}|^2 = |P\mathbf{x}|^2 + |P\mathbf{y}| + 2(P\mathbf{x})^T P\mathbf{y} = |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2(P\mathbf{x})^T P\mathbf{y}.$$

So $(P\mathbf{x})^T P\mathbf{y} = \mathbf{x}^T\mathbf{y}$.

(iii) $\Rightarrow$ (iv): $(P\mathbf{v}_i)^T P\mathbf{v}_j = \mathbf{v}_i^T\mathbf{v}_j = \delta_{ij}$. So $P\mathbf{v}_i$'s are also orthonormal.

(iv) $\Rightarrow$ (v): Take the $\mathbf{v}_i$'s to be the standard basis. So the columns of P, being $P\mathbf{e}_i$, are orthonormal.

(v) $\Rightarrow$ (i): The columns of P are orthonormal. Then $(PP^T)_{ij} = P_{ik}P_{jk} = (P_i) \cdot (P_j) = \delta_{ij}$, viewing P_i as the ith column of P. So $PP^T = I$. □

7.3 Lorentz transformations