Course notes for PMATH 990

Christa Hawthorne

Lectures by Vern Paulsen, Fall 2015

Contents

1 Preliminaries **2**
1.1 Determinants . . . 4
1.1.1 Laplace expansion . . . 4
1.1.2 Permutations . . . 4
1.1.3 Cramer's rule and the adjugate . . . 5
1.2 Row-reduced echelon forms and elementary matrices . . . 6
1.3 Rank . . . 7
1.4 Submatrices . . . 8
1.5 Sums of subspaces . . . 8
1.6 Partitioned matrices . . . 9
1.7 Euclidean norms and inner products . . . 10
1.8 Gram-Schmidt orthogonalization process . . . 11
1.9 Lowden orthogonalization . . . 11

2 Eigenvalues, eigenvectors, and spectra **12**
2.1 The characteristic polynomial . . . 13
2.2 The elementary symmetric functions . . . 14
2.3 Moments and Newton's identities . . . 17
2.4 Right multiplication . . . 17
2.5 Similarity . . . 18
2.6 Persistence of eigenvalues . . . 23

3 Unitaries and isometries **24**
3.1 Householder transformations . . . 29
3.2 Unitary equivalence . . . 30
3.3 Specht's invariants . . . 31

4 Linear maps on matrices **37**

5 QR factorization and Gram-Schmidt **39**

6 Normal and Hermitian **40**
6.1 Hermitian matrices . . . 42
6.2 Majorization . . . 53

7 Positive semidefinite **55**
7.1 Factorization and decomposition . . . 59
7.1.1 Cholesky factorization . . . 60
7.2 Subspaces, orthogonal complements, and projections . . . 62
7.3 Gram matrices . . . 63
7.4 Polar form and singular valued decomposition . . . 64
7.5 Schur products . . . 67

7.6 The positive semidefinite ordering 68

8 Matrix norms **69**
8.1 Power series 72
8.2 Gersgorin disks 73

9 Non-negative matrices **76**
9.1 Irreducible non-negative matrices 83
9.2 Stochastic and doubly stochastic matrices 85

10 Matroids **87**

1 Preliminaries

Vern Paulsen, MC5318, vpaulsen@uwaterloo.ca

To get in touch: send email, and set up a time. After seminars are decided, he'll post office hours.

Recommended book: Matrix Analysis, Horn and Johnson

Outline:

- General matrix theory
  - Unitary equivalence, similarity
  - QR factorization, which we'll use to prove Jordan canonical form
  - Cholesky factorization, Specht invariants
  - Partitioned matrices
- Special families
  - Hermitian, normal, unitary, positive semidefinite, non-negative matrices
  - Circulant matrices
  - Majorization
  - Eigenvalue interlacing theorems
  - Estimates about eigenvalues of sums of Hermitian matrices

Weekly homework assignments of 5-10 problems. (Probably closer to 5.) A bit of discussion is okay, but the proofs should not all be identical. (He'll clarify as time goes on.)

Linear algebra	Matrix analysis
General fields	$\mathbb{R}$ or $\mathbb{C}$
Rings and modules	Limits, continuity, power series
Basis independent	Basis dependent
	Inner products, geometry

Assume you know:

- Fields
- Basic properties of $\mathbb{R}$ and $\mathbb{C}$
- General theory of vector spaces (bases, dimension, matrix of a linear map, determinants and their computations, matrix inverses, etc.)
- Analysis: sequences and series, Heine-Borel

We use $\mathbb{F}$ to denote $\mathbb{R}$ or $\mathbb{C}$. We use $\mathbb{F}^m$ to denote the vector space of m-tuples over $\mathbb{F}$. The canonical basis of $\mathbb{F}^m$ is $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ with a 1 in the i^{th} position. So

$$v = \sum_{i=1}^{m} x_i e_i$$

Dot product is given by, for $v = (x_1, \dots, x_m)$, $w = (y_1, \dots, y_m)$, we have

$$v \cdot w = x_1 y_1 + \cdots + x_m y_m$$

The inner product is given by

$$\langle v, w \rangle = x_1 \overline{y_1} + \cdots + x_m \overline{y_m}$$

When $\mathbb{F} = \mathbb{R}$, they coincide.

We use $M_{m,n}$ to denote the set of $m \times n$ matrices. If we need to specify, we will write $M_{m,n}(\mathbb{R})$ or $M_{m,n}(\mathbb{C})$. In particular, for $A \in M_{m,n}$, we write

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = (a_{i,j})$$

This forms a vector space in the natural way: $M_{m,n} \cong \mathbb{F}^{mn}$. By the canonical basis for $M_{m,n}$, we mean the matrices $E_{i,j}$ containing a 1 in the (i, j) entry and a 0 elsewhere.

Every $A \in M_{m,n}$ defines a linear map $L_A \colon \mathbb{F}^n \to \mathbb{F}^m$ given by

$$L_A((x_1, \dots, x_n)) = \left(\sum_{j=1}^{n} a_{1j} x_j, \dots, \sum_{j=1}^{n} a_{mj} x_j \right)$$

For $A \in M_{m,n}$, $A = (a_{ij})$, then $\overline{A} = (\overline{a_{ij}}) \in M_{m,n}$. Also $A^t = (a_{ji}) \in M_{n,m}$. Also $A^* = \overline{A^t} = \overline{A}^t$ is the *adjoint* or *conjugate transpose*.

Suppose $A \in M_{m,p}$, $B \in M_{p,n}$. Define their product to be $(c_{ij}) \in M_{m,n}$ given by

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$$

Remark 1.

1. $(AB)C = A(BC)$
2. $(A_1 + A_2)B = A_1 B + A_2 B$
3. $A(B_1 + B_2) = AB_1 + AB_2$
4. $\lambda(AB) = (\lambda A)B = A(\lambda B)$
5. Using the association $\mathbb{F}^m \cong M_{m,1}$, we have $L_A(v) = Av$.
6. Using the association $\mathbb{F}^m \cong M_{m,1}$, we have $v \cdot w = w^t v$ and $\langle v, w \rangle = w^* v$.

Remark 2.

1. If we write $A = [C_1 \mid \cdots \mid C_n]$ for $C_i \in M_{m,1}$, and if $v = (x_1, \dots, x_n)$, then

$$L_A(v) = L_A\left(\sum_{j=1}^{n} x_j e_j \right) = \sum_{j=1}^{n} x_j L_A(e_j) \cong \sum_{j=1}^{n} x_j C_j$$

2. Thus $\mathrm{range}(L_A) \cong \mathrm{span}(C_1, \ldots, C_n)$.

Remark 3. For $A \in M_{m,p}$, $B \in M_{p,n}$, we have

1. Writing

$$A = \begin{pmatrix} R_1 \\ \hline \vdots \\ \hline R_m \end{pmatrix}$$

$$B = [C_1 \mid \cdots \mid C_n]$$

   Then $AB = (R_i C_j)$.

2. $AB = [A \cdot C_1 \mid \cdots \mid A \cdot C_n]$

3. Writing

$$A = [W_1 \mid \cdots \mid W_p]$$

$$B = \begin{pmatrix} V_1 \\ \hline \vdots \\ \hline V_p \end{pmatrix}$$

   For $W_k \in \mathbb{F}^m \cong M_{m,1}$, $V_k \in \mathbb{F}^n \cong M_{1,n}$. Then

$$AB = \sum_{k=1}^{p} w_k \cdot v_k$$

   called the "outer product".

Notation 4. We write M_m for $M_{m,m}$.

1.1 Determinants

1.1.1 Laplace expansion

For $A \in M_n$, we put $\det(A) \in \mathbb{F}$. Let $A_{i,j} \in M_{n-1}$ be obtained by eliminating the ith row and jth column. The Laplace expansion is then given by

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det(A_{i,j}) = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \det(A_{i,j})$$

for any choice of i, j respectively.

1.1.2 Permutations

$\sigma \colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ bijective. The sign of a permutation: find a way to express σ as a product of transpositions. The parity modulo 2 of the number of transpositions turns out to be independent of the expression. We set

$$\mathrm{sgn}(\sigma) = \begin{cases} 1 & \text{the parity is even} \\ -1 & \text{else} \end{cases}$$

i.e. $(-1)^k$ where k is the number of transpositions. We can then write

$$\det(A) = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}$$

We also define the *permanent* of a matrix A is

$$\operatorname{perm}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} a_{i,\sigma(i)}$$

This has applications in graph theory, order statistic, symmetric tensor products. In particular,

$$\langle x_1 \vee \cdots \vee x_n, y_1 \vee \cdots \vee y_n \rangle = \operatorname{perm}(\langle x_i, y_j \rangle)$$

Proposition 5.

1. $\det(A^t) = \det(A)$
2. $\det(AB) = \det(A)\det(B)$
3. *A is invertible if and only if* $\det(A) \neq 0$

Definition 6. Suppose $V_1, \ldots, V_n, W$ are vector spaces over $\mathbb{F}$. Then

$$L\colon \bigtimes_{i=1}^{n} V_i \to W$$

is *multilinear* if it satisfies

1. for all j, all $v_j, v_j' \in V_j$, we have
$$L(v_1, \ldots, v_{j-1}, v_j + v_j', v_{j+1}, \ldots, v_n) = L(v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_n) + L(v_1, \ldots, v_{j-1}, v_j', v_{j+1}, \ldots, v_n)$$
2. for all j, all $v_j \in V_j$, and all $\lambda \in \mathbb{F}$, we have
$$L(v_1, \ldots, v_{j-1}, \lambda v_j, v_{j+1}, \ldots, v_n) = \lambda L(v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_n)$$

If we regard matrices as tuples $[C_1 \mid \ldots, \mid C_n]$ with $C_1, \ldots, C_n \in \mathbb{F}^n$, then

$$\det\colon \bigtimes_{i=1}^{n} \mathbb{F}^n \to \mathbb{F}$$

satisfies

1. det is multilinear
2. If B is obtained from A by transposing two columns then $\det(B) = (-1)\det(A)$ (alternation)
3. $\det(I_n) = 1$ (normalization)

Theorem 7. *If*

$$L\colon \bigtimes_{i=1}^{n} \mathbb{F}^n \to \mathbb{F}$$

is alternating, multilinear, and normalized, then

$$L(C_1, \ldots, C_n) = \det([C_1 \mid \cdots \mid C_n])$$

A similar result holds for rows.

1.1.3 Cramer's rule and the adjugate

Suppose $A \in M_n$. Then $A_{i,j} \in M_{n-1}$, as above. Let

$$b_{i,j} = (-1)^{i+j} \det(A_{j,i})$$

Then $B = (b_{i,j}) \in M_n$ is called the *adjugate* of A.

Theorem 8 (Cramer)**.** $BA = AB = \det(A)I$.

1.2 Row-reduced echelon forms and elementary matrices

Definition 9. Suppose $B \in M_{m,n}$. We say B is in RREF if

1. The first non-zero entry in each row is 1. (These are called *leading ones.*)
2. The first non-zero entry of the $(i+1)$th row is to the right of the first non-zero entry of the ith row.
3. All other entries in a column with a leading one are zero.
4. Rows of all zeroes are at the bottom.

Example 10.

$$\begin{pmatrix} 0 & 1 & * & 0 & * & 0 \\ 0 & 0 & 0 & 1 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

is in RREF.

Definition 11. If we omit Item 3, we get the definition of *row echelon form.*

Definition 12. The *elementary operations* are:

Type I Row interchange

Type II Multiply a row by a scalar

Type III Add a multiple of a row to another row

Theorem 13. *Given $A \in M_{m,n}$, there is a sequence of elementary operations to perform on A yielding B in RREF. Moreover, this B is unique. The process $A \to B$ is referred to as Gauss-Jordan elimination. The process of reducing to something in REF is referred to ass Gaussian elimination.*

Proof. See Hoffman and Kunze. □ Theorem 13

Definition 14. The elementary matrices are

Type 1 For $k \neq \ell$, let

$$U(k,\ell) = \sum_{i \neq k,\ell} E_{i,i} + E_{k,\ell} + E_{\ell,k}$$

Then $U(k,\ell)^{-1} = U(k,\ell) = U(\ell,k)$, and $U(k,\ell)A$ corresponds to interchanging the k and ℓ row of A.

Type 2 For $\lambda \neq 0$, let

$$D(k,\lambda) = \sum_{i \neq k} E_{ii} + \lambda E_{kk}$$

Then $D(k,\lambda)^{-1} = D(k,\lambda^{01})$ and $D(k,\lambda)A$ corresponds to scaling row k by λ.

Type 3 For $k \neq \ell$ and $\lambda \in \mathbb{F}$, let

$$S(k,\ell,\lambda) = I + \lambda E_{\ell,k}$$

Then $S(k,\ell,\lambda)^{-1} = S(k,\ell,-\lambda)$ and $S(k,\ell,\lambda)A$ corresponds to adding the k row of A, scaled by λ, to the ℓ row of A.

Corollary 15. *Given $A \in M_{m,n}$, there is $W \in M_m$ that is a product of elementary matrices such that WA is in RREF; furthermore, this W is unique.*

Remark 16. If $WA = B$ as above, then W is invertible, and $A = W^{-1}B$ where W^{-1} is also a product of elementary matrices.

1.3 Rank

Definition 17. For $A \in M_{m,n}$, we define the *column rank* of A, denoted $\operatorname{rank}_c(A)$ is the dimension of the subspace of $\mathbb{F}^m$ spanned by the columns. The *row rank*, denoted $\operatorname{rank}_r(A)$, is the dimension of the subspace of $\mathbb{F}^n$ spanned by the rows.

Theorem 18. *Let $A \in M_{m,n}$; let B be the RREF of A. Then*

1. $\operatorname{rank}_r(A) = \operatorname{rank}_r(B)$
2. $\operatorname{rank}_c(A) = \operatorname{rank}_c(B)$
3. $\operatorname{rank}_r(B) = \operatorname{rank}_c(B)$ *are both equal to the number of leading 1s in B.*

Corollary 19. $\operatorname{rank}_c(A) = \operatorname{rank}_r(A)$, *and we henceforth refer to it as* $\operatorname{rank}(A)$.

Definition 20. Let $L \colon V \to W$ be linear. We define the *range of L* by

$$\mathcal{R}(L) = \{\, L(v) : v \in V \,\}$$

Then $\mathcal{R}(L)$ is a subspace, since L is linear.

Proof of Theorem 18.

1. Recall $B = WA$ where W is a product of elementary matrices. Thus each row of B is a linear combination of rows of B. Thus

   $$\operatorname{rank}_r(B) = \dim(\operatorname{span}(\text{rows of } B)) \leq \dim(\operatorname{span}(\text{rows of } A)) = \operatorname{rank}_r(A)$$

   But we can apply the same argument to $A = W^{-1}B$ to get $\operatorname{rank}_r(B) \geq \operatorname{rank}_r(A)$. Thus $\operatorname{rank}_r(A) = \operatorname{rank}_r(B)$.

2. Look at $L_A, L_B \colon \mathbb{F}^n \to \mathbb{F}^m$. Write

   $$A = [C_1 | \dots | C_n]$$

   Then

   $$A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \lambda_1 C_1 + \cdots + \lambda_n C_n$$

   Thus $\mathcal{R}(L_A) = \operatorname{span}\{\, C_1, \dots, C_n \,\}$. But $B = WA$; so $\mathcal{R}(B) = \mathcal{R}(WA) = W(\mathcal{R}(A))$. But W is invertible; so we get

   $$\operatorname{rank}_c(B) = \dim(\mathcal{R}(B)) = \dim(W(\mathcal{R}(A))) = \dim(\mathcal{R}(A)) = \operatorname{rank}_c(A)$$

3. Picture

   $$B = \begin{pmatrix} 0 & \dots & 0 & 1 & * & * & 0 & * & * & 0 \\ 0 & \dots & 0 & 0 & 0 & 0 & 1 & * & * & 0 \\ 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

   The columns for leading 1's span the column space and are linearly independent. Thus the dimension of the column space of B is just the number of leading 1's.

   The non-zero rows are exactly the rows with leading 1's; these thus span the row space. Furthermore, since each column containing a leading 1 has exactly one non-zero entry, we have that the rows with leading 1's are independent. So they form a basis, and the dimension of the row space is the number of leading 1's.

□ Theorem 18

Remark 21. In general it holds that the rows of CD are linear combinations of the rows of D.

1.4 Submatrices

Definition 22. Suppose $A \in M_{m,n}$. Suppose

$$\alpha = \{ \alpha_1, \dots, \alpha_k \}$$
$$\beta = \{ \beta_1, \dots, \beta_j \}$$

with

$$1 \leq \alpha_1 < \cdots < \alpha_k \leq m$$

and

$$1 \leq \beta_1 < \cdots < \beta_j \leq n$$

We define $A[\alpha, \beta] \in M_{k,j}$ by

$$A[\alpha, \beta] = (a_{\alpha_i, \beta_j})$$

(where $A = (a_{ij})$). These are the *submatrices of* A.

Any matrix of the form $A[\alpha, \alpha]$ is said to be a *principal submatrix* of A.

For $|\alpha| = |\beta|$, we call $\det(A[\alpha, \beta])$ a *minor of* A.

When $\alpha = \{ 1, \dots, k \}$, then $A[\alpha, \alpha]$ is called a *leading principal submatrix.*

When $n = m$ and $\alpha = \{ k, k+1, \dots, n \}$, then $A[\alpha, \alpha]$ is a *trailing principal submatrix.*

The determinant of a leading or trailing principal submatrix is called a *leading* or *trailing principal minor*, respectively.

1.5 Sums of subspaces

Suppose V, W are vector spaces. Consider their Cartesian product $V \times W$. We can regard it as a vector space by taking

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2)$$
$$\lambda(v_1, w_1) = (\lambda v_1, \lambda w_1)$$

This is denoted $V \oplus W$ and is called the *direct sum* of V and W.

Proposition 23. *If* $\{ v_\alpha : \alpha \in A \}$ *is a basis for* V*,* $\{ w_\beta : \beta \in B \}$ *is a basis for* W *(neither necessarily finite), then*

$$\{ (v_\alpha, 0) : \alpha \in A \} \cup \{ (0, w_\beta) : \beta \in B \}$$

is a basis for $V \oplus W$*. Thus* $\dim(V \oplus W) = \dim(V) \oplus \dim(W)$*.*

Proof. Suppose $(v, w) \in V \oplus W$. Write

$$v = \sum_\alpha \lambda_\alpha v_\alpha$$
$$w = \sum_\beta \mu_\beta w_\beta$$

Then

$$(v, w) = \sum_\alpha \lambda_\alpha (v_\alpha, 0) + \sum_\beta \mu_\beta (0, w_\beta)$$

Suppose

$$\sum_\alpha \lambda_\alpha (v_\alpha, 0) + \sum_\beta \mu_\beta (0, w_\beta) = 0$$

Then

$$\sum_\alpha \lambda_\alpha v_\alpha = 0$$
$$\sum_\beta \mu_\beta w_\beta = 0$$

and thus each λ_α and μ_β is 0. □ Proposition 23

Example 24. $\mathbb{F}^m \oplus \mathbb{F}^p$ contains vectors of the form

$$((x_1, \dots, x_m), (y_1, \dots, y_p)) \approx (x_1, \dots, x_m, y_1, \dots, y_p) \in \mathbb{F}^{m+p}$$

This yields an isomorphism $\mathbb{F}^m \oplus \mathbb{F}^p \cong \mathbb{F}^{m+p}$.

Conversely, given $(x_1, \dots, x_m, x_{m+1}, \dots, x_{m+p}) \in \mathbb{F}^{m+p}$, we can partition it after the x_m, and regard it as an element of $\mathbb{F}^m \oplus \mathbb{F}^p$.

1.6 Partitioned matrices

Suppose $A \in M_{m_1+m_2, n_1+n_2}$. We can then define $A_{ij} \in M_{m_i, n_j}$ in the natural way. These can be regarded as

$$A_{ij} \colon \mathbb{F}^{n_j} \to \mathbb{F}^{m_i}$$

Recall that A can be regarded as

$$A \colon \mathbb{F}^{n_1+n_2} \to \mathbb{F}^{m_1+m_2}$$

Then

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_{n_1} \\ x_{n_1+1} \\ \vdots \\ x_{n_1+n_2} \end{pmatrix} = \begin{pmatrix} A_{11}x + A_{12}y \\ \hline A_{21}x + A_{22}y \end{pmatrix}$$

$$\begin{tikzcd} \mathbb{F}^{n_1+n_2} \arrow[r, "A"] \arrow[d, "\cong"] & \mathbb{F}^{m_1+m_2} \arrow[d, "\cong"] \\ \mathbb{F}^{n_1} \oplus \mathbb{F}^{n_2} \arrow[r] & \mathbb{F}^{m_1} \oplus \mathbb{F}^{m_2} \end{tikzcd}$$

where the bottom map is given by

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

Check: for $A \in M_{m_1+m_2, p_1+p_2}$, $B \in M_{p_1+p_2, n_1+n_2}$, if

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$
$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

where $A_{i,j} \in M_{m_i, p_j}$ and $B_{i,j} \in M_{p_i, n_j}$, then

$$AB = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}$$

On the assignment, we may assume we never deal with 1×1 matrices. We may use anything asserted in class.

Definition 25. If we say $A \in M_{m_1+\dots+m_k, n_1+\dots+n_k}$ is *partitioned*, we mean that we can partition A as $A_{i,j} \in M_{m_i, n_j}$. We say A is *block-diagonal* to mean $A_{i,j} = 0$ for $i \neq j$.

Proposition 26. *Suppose $A \in M_{n_1+\dots+n_k, n_1+\dots+n_k}$ is block-diagona. Then*

$$\det(A) = \det(A_{11}) \dots \det(A_{nn})$$

Proof. Expand by Laplace. □ Proposition 26

Suppose

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

for $a \neq 0$. Recall that

$$\det(A) = ad - bc = a(d - ba^{-1}c)$$

Proposition 27. *Suppose $A \in M_{n_1+n_2,n_1+n_2}$. If A_{11} is invertible, then*

$$\det(A) = \det(A_{11})\det(A_{22} - A_{21}A_{11}^{-1}A_{12})$$

If A_{22} is invertible, then

$$\det(A) = \det(A_{22})\det(A_{11} - A_{12}A_{22}^{-1}A_{21})$$

Proof.

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ A_{21} & I_{n_2} \end{pmatrix} \begin{pmatrix} I_{n_1} & A_{11}^{-1}A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{pmatrix}$$

and thus

$$\det \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \det \begin{pmatrix} A_{11} & 0 \\ A_{21} & I_{n_2} \end{pmatrix} \det \begin{pmatrix} I_{n_1} & A_{11}^{-1}A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{pmatrix} = \det(A_{11})\det(A_{22} - A_{21}A_{11}^{-1}A_{12})$$

by taking the Laplace expansion along the identity matrices. Proof of the second fact is similar, using the factorization

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} I_{n_1} & A_{12} \\ 0 & A_{22} \end{pmatrix} \begin{pmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & 0 \\ A_{22}^{-1}A_{21} & I_{n_2} \end{pmatrix}$$

□ Proposition 27

1.7 Euclidean norms and inner products

Suppose $x, y \in \mathbb{F}^n$ with

$$\begin{aligned} x &= (x_1, \ldots, x_n)^c \\ y &= (y_1, \ldots, y_n)^c \end{aligned}$$

we set

$$\langle x, y \rangle = x_1\overline{y_1} + \cdots + x_n\overline{y_n}$$

Remark 28. $\langle \cdot, \cdot \rangle \colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ satisfies

$$\begin{aligned} \langle x + x', y \rangle &= \langle x, y \rangle + \langle x', y \rangle \\ \langle x, y + y' \rangle &= \langle x, y \rangle + \langle x, y' \rangle \\ \langle \lambda x, y \rangle &= \lambda \langle x, y \rangle \\ \langle x, \lambda y \rangle &= \overline{\lambda} \langle x, y \rangle \end{aligned}$$

i.e. $\langle \cdot, \cdot \rangle$ is *sesquilinear.*

Remark 29. In the case $\mathbb{F} = \mathbb{R}$, we have $\langle \cdot, \cdot \rangle$ is *bilinear.*

Definition 30. Suppose $x, y \in \mathbb{F}^n$. We say x and y are *orthogonal* (written $x \perp y$) when $\langle x, y \rangle = 0$.

Definition 31. Suppose $x \in \mathbb{F}^n$. We define the *Euclidean 2-norm* of x to be

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

Theorem 32 (Cauchy-Schwarz inequality)**.** $|\langle x, y \rangle| \leq \|x\|_2 \|y\|_2$.

Proof. Well

$$\begin{aligned}0 &\le \langle t\exp(i\theta)x + y, t\exp(i\theta)x + y\rangle \\ &= t^2\langle x,x\rangle + t\exp(i\theta)\langle x,y\rangle + t\exp(-i\theta)\langle y,x\rangle + \langle y,y\rangle \\ &= t^2\|x\|_2^2 + 2t|\langle x,y\rangle| + \|y\|_2^2\end{aligned}$$

thus the discriminant is non-positive, and

$$4|\langle x,y\rangle|^2 - 4\|x\|_2^2\|y\|_2^2 \le 0$$

□ Theorem 32

Theorem 33 (Triangle inequality)**.** $\|x+y\|_2 \le \|x\|_2 + \|y\|_2$.

Proof. Well,

$$\begin{aligned}0 &\le \|x+y\|_2^2 \\ &= \langle x+y, x+y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\ &\le \|x\|_2^2 + 2\|x\|_2\|y\|_2 + \|y\|_2^2 \\ &= (\|x\|_2 + \|y\|_2)^2\end{aligned}$$

□ Theorem 33

Fact 34. $\|\lambda x\|_2 = |\lambda|\|x\|_2$.

1.8 Gram-Schmidt orthogonalization process

Definition 35. A set S is *orthonormal* (o.n.) if it satisfies the following:

- For all $u \in S$ we have $\|u\|_2 = 1$
- For all $u \ne v$ in S, we have $u \perp v$

Given independent $\{ x_1, \dots, x_m \} \subseteq \mathbb{F}^n$, Gram-Schmidt yields orthonormal $\{ u_1, \dots, u_m \} \subseteq \mathbb{F}^n$ such that

$$\operatorname{span} u_1, \dots, u_k = \operatorname{span} x_1, \dots, x_k$$

for all $1 \le k \le n$. In particular, we set

$$\begin{aligned}u_1 &= \frac{x_1}{\|x_1\|_2} \\ u_{n+1} &= \frac{x_{n+1} - \sum_{i=1}^n \langle x_{n+1}, u_i\rangle u_i}{\left\|x_{n+1} - \sum_{i=1}^n \langle x_{n+1}, u_i\rangle u_i\right\|_2}\end{aligned}$$

It's not hard to see that the u_i satisfy the desired properties.

1.9 Lowden orthogonalization

Gram-Schmidt is numerically unstable; an alternative is Lowden orthogonalization.

Given independent $\{ x_1, \dots, x_m \} \subseteq \mathbb{F}^m$, consider

$$\inf\left\{ \sum_{j=1}^m \|x_j - u_j\|_2^2 : \{ u_1, \dots, u_m \} \text{ orthonormal} \right\}$$

Theorem 36 (Lowden)**.** *There is a unique orthonormal* $\{ u_1, \dots, u_m \}$ *attaining this infimum.*

Theorem 37. *This unique set* $\{ u_1, \dots, u_m \}$ *is called the Lowden orthogonalization of* $\{ x_1, \dots, x_m \}$.

Here marks the end of the review.

2 Eigenvalues, eigenvectors, and spectra

Definition 38. Suppose S is a linear map. Define the *nullspace* or *kernel* of S to be

$$\mathcal{N}(S) = \{ x : Sx = 0 \}$$

Proposition 39. *Suppose $S \in M_n$. Then the following are equivalent:*

1. *There is T such that $ST = TS = I_n$*
2. $\det(S) \neq 0$
3. *$L_S \colon \mathbb{F}^n \to \mathbb{F}^n$ is injective*
4. *$L_S \colon \mathbb{F}^n \to \mathbb{F}^n$ is surjective*

Proof.

(1) $\iff$ (2) Cramer's rule.

(1) $\implies$ (3) Assume S is not injective. Then $\mathcal{N}(S) \neq 0$. Thus there is $x \neq 0$ such that $Sx = 0$; thus $TS \neq I$.

(3) $\iff$ (4) By the rank-nullity theorem: that

$$\dim(\mathcal{R}(S)) + \dim(\mathcal{N}(S)) = n$$

(3) $\implies$ (1) By ((3) $\iff$ (4)), we have S is bijective, and is thus invertible.

$\square$ Proposition 39

We let $M_n^{-1} = \{ S \in M_n : S \text{ invertible} \}$.

Definition 40. Suppose $A \in M_n$, $x \in \mathbb{C}^n$, $x \neq 0$, and $Ax = \lambda x$. We call such an x an *eigenvector*, and we call such a λ an *eigenvalue*. The *spectrum* of A, denoted $\sigma(A)$, is the set of all eigenvalues of A. The *spectral radius* of A is

$$\rho(A) = \sup\{ |\lambda| : \lambda \in \sigma(A) \}$$

Proposition 41. *Suppose $A \in M_n$. Then*

$$\sigma(A) = \{ \lambda : (A - \lambda I) \notin M_n^{-1} \}$$

Proof. $(A - \lambda I)$ is not invertible if and only if $\mathcal{N}(A - \lambda I) \neq \{ 0 \}$, which holds if and only if there is a non-zero x such that $Ax = \lambda x$. $\square$ Proposition 41

Given $A \in M_n$, $p(t) = a_k t^k + \cdots + a_1 t + a_0$, we can define

$$p(A) = a_k A^k + \cdots + a_1 A + a_0 I$$

Remark 42. $(pq)(A) = p(A)q(A)$.

Theorem 43 (Spectral mapping theorem)**.** *Suppose $A \in M_n$, p a polynomial. Then*

$$\sigma(p(A)) = \{ p(\lambda) : \lambda \in \sigma(A) \} = p(\sigma(A))$$

Proof. $\supseteq$ Let $\lambda \in \sigma(A)$. Then there is $x \neq 0$ such that $Ax = \lambda x$. We then have that $A^j x = \lambda^j x$. Then

$$\begin{aligned}(a_k A^k + \cdots + a_1 A + a_0 I)x &= (a_k \lambda^k + \cdots + a_1 \lambda + a_0)x \\ &= p(\lambda)x\end{aligned}$$

so $p(A)x = p(\lambda)x$, and $p(\lambda) \in \sigma(p(A))$.

$\subseteq$ Suppose $\mu \in \sigma(p(A))$. Then there is $x \neq 0$ such that $p(A)x = \mu x$. Let

$$q(t) = p(t) - \mu = a_k \prod_{j=1}^{k} (t - \mu_j)$$

Then

$$q(A) = a_k(A - \mu_1 I) \dots (A - \mu_k I)$$

and

$$p(A) - \mu I = q(A) = a_k(A - \mu_1 I) \dots (A - \mu_k I)$$

But $p(A) - \mu I$ is not invertible because $\mu \in \sigma(p(A))$. So there is some j_0 such that $A - \mu_{j_0}$ is not invertible, and thus $\mu_{j_0} \in \sigma(A)$. Thus

$$p(\mu_{j_0}) - \mu = q(\mu_{j_0}) = 0$$

and thus

$$\mu = p(\mu_{j_0}) \in \{ p(\lambda) : \lambda \in \sigma(A) \}$$

$\square$ Theorem 43

2.1 The characteristic polynomial

Definition 44. Suppose $A \in M_n$. Then the *characteristic polynomial* of A is

$$p_A(t) = \det(tI - A)$$

which is then monic of degree n.

Proposition 45. $\lambda \in \sigma(A) \iff p_A(\lambda) = 0$.

Proof.

$$\begin{aligned} \lambda \in \sigma(A) &\iff \lambda I - A \text{ is singular} \\ &\iff \det(\lambda I - A) = 0 \\ &\iff p_A(\lambda) = 0 \end{aligned}$$

$\square$ Proposition 45

Note that $\lambda \in \sigma(A)$ if and only if $\mathcal{N}(A - \lambda I) \neq \{ 0 \}$. This space is called the *eigenspace for* λ.

Definition 46. For $\lambda \in \sigma(A)$, the *geometric multiplicity of* λ is the dimension of $\mathcal{N}(A - \lambda I)$. The *algebraic multiplicity of* λ is the number of times $(t - \lambda)$ is a factor of $p_A(t)$.

Example 47.

$$A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

has geometric multiplicity 1, but

$$\det(tI - A) = (t - \lambda)^2$$

so it has algebraic multiplicity 2.

Proposition 48. *Suppose $A \in M_n$, $\lambda \in \sigma(A)$. Then the geometric multiplicity of λ is no more than the algebraic multiplicity of λ.*

Remark 49. If $B = (b_{ij})$, then $b_{i,j} = \langle Be_j, e_i \rangle$.

Proof of Proposition 48. Let k be the geometric multiplicity of λ. Pick $v_1, \ldots, v_k$ a basis for $\mathcal{N}(\lambda I - A)$. Extend to a basis $\{ v_1, \ldots, v_n \}$ for $\mathbb{C}^n$. Let

$$S = [v_1 \mid \cdots \mid v_n] \in M_n$$

Then $\mathcal{R}(S) = \operatorname{span}\{ v_1, \ldots, v_n \} = \mathbb{C}^n$, so S is surjective, and thus invertible. Let $B = S^{-1}AS$. For $1 \le j \le k$, and for $1 \le i \le n$, note that

$$\begin{aligned} b_{i,j} &= \langle S^{-1}ASe_j, e_i \rangle \\ &= \langle S^{-1}Av_j, e_j \rangle \\ &= \langle S^{-1}(\lambda v_j), e_i \rangle \\ &= \lambda \langle e_j, e_i \rangle \\ &= \begin{cases} \lambda & i = j \\ 0 & \text{else} \end{cases} \end{aligned}$$

Then

$$B = \left(\begin{array}{ccccc|c} \lambda & 0 & \ldots & 0 & | & B_{12} \\ 0 & \lambda & \ldots & 0 & | & \\ \vdots & \vdots & \ddots & \vdots & | & \\ 0 & 0 & \ldots & \lambda & | & \\ \hline 0 & & & & | & B_{22} \end{array}\right)$$

Then

$$p_B(t) = (t - \lambda)^k \det(tI - B_{22})$$

so $(t - \lambda)^k$ divides $p_B(t)$. But

$$p_B(t) = \det(tI - B) = \det(S^{-1}(tI - A)S) = \det(tI - A) = p_A(t)$$

So $(t - \lambda)^k$ divides $p_A(t)$, and the algebraic multiplicity is at least the geometric multiplicity.

□ Proposition 48

2.2 The elementary symmetric functions

Observe that

$$(t - \lambda_1) \ldots (t - \lambda_n) = t^n - (\lambda_1 + \cdots + \lambda_n)t^{n-1} + (\lambda_1\lambda_2 + \lambda_1\lambda_3 + \cdots + \lambda_{n-1}\lambda_n) + \cdots + (-1)^n \lambda_1 \ldots \lambda_n$$

Definition 50. Given $n \in \mathbb{N}$ and $1 \le k \le n$, we define the k^{th} *elementary symmetric function* by

$$S_k(\lambda_1, \ldots, \lambda_n) = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} \prod_{j=1}^{k} \lambda_{i_j}$$

Example 51.

$$\begin{aligned} S_1(\lambda_1, \ldots, \lambda_n) &= \lambda_1 + \cdots + \lambda_n \\ S_2(\lambda_2, \ldots, \lambda_n) &= \sum_{1 \le i < j \le n} \lambda_i \lambda_j \\ &\vdots \\ S_n(\lambda_1, \ldots, \lambda_n) &= \lambda_1 \ldots \lambda_n \end{aligned}$$

Fact 52.

$$(t - \lambda_1) \ldots (t - \lambda_n) = \sum_{k=0}^{n} (-1)^k S_k(\lambda_1, \ldots, \lambda_n) t^{n-k}$$

Definition 53. Let $A \in M_n$. Let

$$E_k(A) = \sum_{J \subseteq \{ 1,\dots,n \}, |J|=k} \det(A(J,J))$$

i.e. $E_k(A)$ is the sum of all the $k \times k$ principal minors of A.

Definition 54. For $A \in M_n$, we define the *trace* of A to be

$$\operatorname{tr}(A) = \sum_{j=1}^{n} a_{jj}$$

Example 55. $E_1(A) = \operatorname{tr}(A)$ and $E_n(A) = \det(A)$.

Theorem 56. *Let $A \in M_n$. Let $(\lambda_1, \dots, \lambda_n)$ be the roots of $p_A(t)$ repeated according to their algebraic multiplicity. Then for $1 \le k \le n$, we have $E_k(A) = S_k(\lambda_1, \dots, \lambda_n)$.*

Corollary 57. *$\{ E_1(A), \dots, E_n(A) \}$ uniquely determines the roots of $p_A(t)$.*

Notation 58. Given $f \colon (a,b) \to \mathbb{C}$, we can write $f(t) = f_1(t) + i f_2(t)$ for $f_i \colon (a,b) \to \mathbb{R}$. We say f is *differentiable* at t if f_1, f_2 are differentiable at t, and we write $f'(t) = f_1'(t) + i f_2'(t)$.

Notation 59. For $f \colon (a,b) \to \mathbb{C}^n$, we write

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$$

is *differentiable* at t if $f_1, \dots, f_n$ are differentiable at t. In this case, we write

$$f'(t) = \begin{pmatrix} f_1'(t) \\ \vdots \\ f_n'(t) \end{pmatrix}$$

Remark 60.

$$(f_1, \dots, f_n)' = f_1' f_2 \dots f_n + f_1 f_2' \dots f_n + \dots + f_1 \dots f_{n-1} f_n'$$

Theorem 61. *Let*

$$f_j = \begin{pmatrix} f_{ij} \\ \vdots \\ f_{nj} \end{pmatrix} : (a,b) \to \mathbb{C}^n$$

for $1 \le j \le n$. Assume each f_{ij} is differentiable at t for $a < t < b$. Set $g(t) = \det((f_1 \mid \dots \mid f_n))$. Then g is differentiable at t and

$$g'(t) = \sum_{j=1}^{n} \det((f_1 \mid \dots \mid f_{j-1} \mid f_j' \mid f_{j+1} \mid \dots \mid f_n))$$

Proof.

$$\begin{aligned}
g(t+h) - g(t) &= \det((f_1(t+h) \mid f_2(t+h) \mid \dots \mid f_n(t+h))) - \det(f_1(t) \mid \dots \mid f_n(t)) \\
&= \det(f_1(t+h) \mid \dots \mid f_n(t+h)) - \det(f_1(t) \mid f_2(t+h) \mid \dots \mid f_n(t+h)) \\
&\quad + \det(f_1(t) \mid f_2(t+h) \mid \dots \mid f_n(t+h)) - \det(f_1(t) \mid f_2(t) \mid f_3(t+h) \mid \dots \mid f_n(t+h)) \\
&\quad + \det(f_1(t) \mid f_2(t) \mid f_3(t+h) \mid \dots \mid f_n(t+h)) - \dots \\
&= \det(f_1(t+h) - f_1(t) \mid f_2(t+h) \mid \dots f_n(t+h)) \\
&\quad + \det(f_1(t) \mid f_2(t+h) - f_2(t) \mid f_3(t+h) \mid \dots \mid f_n(t+h) + \dots \\
&= \sum_{j=1}^{n} \det(f_1(t) \mid \dots \mid f_{j-1}(t) \mid f_j(t+h) - f_j(t) \mid f_{j+1}(t+h) \mid \dots \mid f_n(t+h))
\end{aligned}$$

Thus

$$\lim_{h\to 0}\frac{g(t+h)-g(t)}{h} = \lim_{h\to 0}\sum_{j=1}^{n}\det(f_1(t)\mid\cdots\mid f_{j-1}(t)\mid f_j(t+h)-f_j(t)\mid f_{j+1}(t+h)\mid\cdots\mid f_n(t+h))$$
$$= \sum_{j=1}^{n}\det(f_1(t)\mid\cdots\mid f_{j-1}(t)\mid f_j'(t)\mid f_{j+1}(t)\mid\ldots f_n(t))$$

since det is continuous in its entries. □ Theorem 61

Proof of Theorem 56.

Notation 62. For $J\subseteq\{1,\ldots,n\}$, we set $A_J = A(J^c, J^c)$.

Well,

$$\begin{aligned}(-1)^n S_n(\lambda_1,\ldots,\lambda_n) &= a_0\\ &= p_A(0)\\ &= \det(-A)\\ &= \det((-I)A)\\ &= \det(-I)\det(A)\\ &= (-1)^n\det(A)\end{aligned}$$

So

$$S_n(\lambda_1,\ldots,\lambda_n) = \lambda_1\ldots\lambda_n = a_0 = \det(A) = E_n(A)$$

And we have the case $k=n$.

Also

$$\begin{aligned}(01)^{n-1}S_{n-1}(\lambda_1,\ldots,\lambda_n) &= a_1\\ &= p_A'(0)\\ &= \det(tI-A)'|_{t=0}\end{aligned}$$

But

$$\begin{aligned}\det(tI-A)' &= \det\begin{pmatrix}1 & -a_{12} & -a_{13} & \ldots\\ 0 & t-a_{22} & -a_{23} & \ldots\\ \vdots & \vdots & \vdots & \ldots\\ 0 & -a_{n2} & -a_{n3} & \ldots\end{pmatrix} + \det\begin{pmatrix}-a_{11} & 0 & -a_{13} & \ldots\\ t-a_{21} & 1 & -a_{23} & \ldots\\ \vdots & \vdots & \vdots & \ldots\\ -a_{n1} & 0 & -a_{n3} & \ldots\end{pmatrix} + \ldots\\ &= \det(tI_{n-1}-A_{\{1\}}) + \det(tI_{n-1}-A_{\{2\}})\ldots\end{aligned}$$

Thus

$$\det(tI_n - A)' = \sum_{i=1}^{n}\det(tI_{n-1}-A_{\{i\}}$$

and

$$\begin{aligned}(-1)^{n-1}S_{n-1}(\lambda_1,\ldots,\lambda_n) &= p_A'(0)\\ &= \sum_{i=1}^{n}\det(-A_{\{i\}})\\ &= \sum_{i=1}^{n}(-1)^{n-1}\det(A_{\{i\}})\\ &= (-1)^{n-1}E_{n-1}(A)\end{aligned}$$

So $S_{n-1}(\lambda_1,\ldots,\lambda_n) = E_{n-1}(A)$. And we have the case $k=n-1$.

But now

$$p_A''(0) = 2a_2 = 2(-1)^{n-1}S_{n-2}(\lambda_1, \dots, \lambda_n)$$

and

$$p_A''(t) = \sum_{i=1}^{n} \det(tI_{n-1} - A_{\{i\}})' = \sum_{i=1}^{n} \sum_{j \neq i} \det(tI_{n-2} - A_{\{i,j\}})$$

So

$$p_A''(0) = \sum_{i=1}^{n} \sum_{j \neq i} \det(-A_{\{i,j\}}) = \sum_{i=1}^{n} \sum_{j \neq i} (-1)^{n-2} \det(A_{\{i,j\}}) = (-1)^{n-2} \cdot 2 \cdot E_{n-2}(A)$$

So $S_{n-2}(\lambda_1, \dots, \lambda_n) = E_{n-2}(\lambda_1, \dots, \lambda_n)$, and we have the case $k = n-2$.

In the general case, we have $k!a_k = p_A^{(k)}(0)$. But

$$p_A^{(k)}(t) = (k!) \sum_{|J|=k} \det(tI - A_J)$$

so

$$p_A^{(k)}(0) = (k!)E_{n-k}(A)(-1)^{n-k}$$

Thus $S_{n-k}(\lambda_1, \dots, \lambda_n) = E_{n-k}(A)$. □ Theorem 56

2.3 Moments and Newton's identities

Definition 63. The k^{th} *moment* is given by $\mu_k = M_k(\lambda_1, \dots, \lambda_k) = \lambda_1^k + \dots + \lambda_n^k$.

Remark 64.

$$\begin{aligned} S_1(\lambda_1, \dots, \lambda_n) &= \lambda_1 + \dots + \lambda_n \\ &= M_1(\lambda_1, \dots, \lambda_n) \\ &= \mu_1 \end{aligned}$$

$$\begin{aligned} S_1(\lambda_1, \dots, \lambda_n)^2 &= (\lambda_1 + \dots + \lambda_n)^2 \\ &= \lambda_1^2 + \lambda_2^2 + \dots + \lambda_n^2 + \sum_{i \neq j} \lambda_i \lambda_j \\ &= \mu_2 + 2S_2(\lambda_1, \dots, \lambda_n) \\ \implies \mu_2 &= S_1^2 - 2S_2 \end{aligned}$$

Theorem 65 (Newton's identities)**.** *Suppose $p(t) = t^n + a_{n-1}t^{n-1} + \dots + a_1 t + a_0$. (Note that $a_k = (-1)^{n-k}S_{n-k}$.) Then*

$$ka_{n-k} + \mu_1 a_{n-k+1} + \dots + \mu_k a_n = 0$$

for all $k \in \{1, \dots, n\}$.

Solving these can get μ in terms of S and vice versa.

2.4 Right multiplication

Suppose $A \in M_{m,n}$. We typically regard $\mathbb{F}^k \cong M_{k,1}$. Then

$$L_A \colon \mathbb{F}^n \to \mathbb{F}^m$$

is given by $L_A x = Ax$. If we instead regard $\mathbb{F}^k \cong M_{1,k}$, then for $y = (y_1, \dots, y_n)$, we have

$$yA \in M_{1,n} \cong \mathbb{F}^n$$

So $R_A \colon \mathbb{F}^m \to \mathbb{F}^n$. Thus $A \in M_n$ induces two linear maps

$$L_A, R_A \colon \mathbb{F}^n \to \mathbb{F}^n$$

Question 66. What is the deal with σ, eigenvalues, etc.?

Well, observe $(yA)^t = A^t y^t$. Thus $R_A \cong L_{A^t}$.

Theorem 67. *For $A \in M_n$. Then*

1. *$p_A(t) = p_{A^t}(t)$.*
2. *$\sigma(A) = \sigma(A^t)$.*
3. *For $\lambda \in \sigma(A) = \sigma(A^t)$, the algebraic multiplicities coincide.*
4. *For $\lambda \in \sigma(A) = \sigma(A^t)$, the geometric multiplicities coincide.*

Proof.

1. $p_{A^t}(t) = \det(tI - A^t) = \det((tI - A)^t) = \det(tI - A) = p_A(t)$.
2. $\sigma(A) = \{ x \in \mathbb{F} : p_A(x) = 0 \} = \{ x \in \mathbb{F}; p_{A^t}(x) = 0 \} = \sigma(A^t)$.
3. Follows as algebraic multiplicity is the number of times $t - \lambda$ appears as a factor of $p_A(t) = p_{A^t}(t)$.
4. The geometric multiplicity of λ in A is given by

$$\begin{aligned} \dim(\mathcal{N}(\lambda I - A)) &= n - \dim(\mathcal{R}(\lambda I - A)) \\ &= n - \operatorname{rank}_c(\lambda I - A) \\ &= n - \operatorname{rank}_r(\lambda I - A^t) \\ &= n - \dim(\mathcal{R}(\lambda I - A^t)) \\ &= \dim(\mathcal{N}(\lambda I - A^t)) \end{aligned}$$

which is just the geometric multiplicity of λ in A^t.

□ Theorem 67

2.5 Similarity

Let $L \colon V \to V$ be linear with $\dim(V) = n$. Let $\mathcal{B} = \{ v_1, \dots v_n \}$ be a basis for V. Recall

$$L(v_j) = \sum_{i=1}^{n} b_{i,j} v_i$$

where $B = (b_{i,j}) \in M_n$ is the *matrix for* L with respect to $\mathcal{B}$, denoted $B = \operatorname{mat}_{\mathcal{B}}(L)$. Define $S \colon \mathbb{F}^n \to V$ be given by $Se_j = v_j$. Then the following diagram commutes:

$$\begin{tikzcd} V \arrow[r, "L"] \arrow[d, "S^{-1}"] & V \\ \mathbb{F}^n \arrow[r, "L_B"] & \mathbb{F}^n \arrow[u, "S"'] \end{tikzcd}$$

For $A = (a_{ij}) \in M_n$, a new basis $\{ v_1, \dots, v_n \}$ for $\mathbb{C}^n$, we have

$$\begin{tikzcd} V \arrow[r, "L_A"] \arrow[d, "S^{-1}"] & V \\ \mathbb{F}^n \arrow[r, "L_B"] & \mathbb{F}^n \arrow[u, "S"'] \end{tikzcd}$$

where $Se_j = v_j$; we set $\operatorname{mat}_{\mathcal{B}}(L_A) = L_B$; then $L_A = S L_B S^{-1}$. i.e.

$$\begin{aligned} A &= S^{-1} B S \\ B &= S A S^{-1} \end{aligned}$$

Then

$$\operatorname{mat}_{\mathcal{B}}(L_A) = S A S^{-1}$$

where $S = [v_1 \mid \cdots \mid v_n]$.

Definition 68. Suppose $A, B \in M_n$. We say B is *similar* to A if there is $S \in M_n^{-1}$ such that $B = SAS^{-1}$. We then write $B \sim A$.

Remark 69.

1. $A = I^{-1}AI$. So $A \sim A$.
2. If $B \sim A$, say $B = SAS^{-1}$, then $A = (S^{-1})^{-1}B(S^{-1})$, so $A \sim B$.
3. If $C \sim B$ and $B \sim A$, then $C = RBR^{-1}$, $B = SAS^{-1}$, then $C = (SR)^{-1}A(SR)$.

So $\sim$ is an *equivalence relation*.

Proposition 70. *Suppose $B \sim A$. Then*

1. *$p_A(t) = p_B(t)$.*
2. *$\sigma(A) = \sigma(B)$.*
3. *The geometric and algebraic multiplicities of $\lambda \in \sigma(A) = \sigma(B)$ coincide.*
4. *$E_k(A) = E_k(B)$ for $1 \leq k \leq n$.*

Proof.

1. $p_B(t) = \det(tI - B) = \det(S(tI - A)S^{-1}) = \det(S)\det(S^{-1})\det(tI - A) = p_A(t)$.
2. $\sigma(A)$ is the roots of $p_A(t) = p_B(t)$, and is thus the roots of $p_B(t)$.
3. The algebriac multiplicities are clear, as they have the same characteristic polynomial. For the geometric multiplicity, note that
$$\dim(\mathcal{N}(\lambda I - A)) = \dim(\mathcal{N}(S(\lambda I - A)S^{-1}))$$
4. Because $E_k(A)$ is determined by the k^{th} coefficient of $p_A(t) = p_B(t)$.

□ Proposition 70

Example 71. $A, B \in M_7$ are given by

$$A = \begin{pmatrix} 0 & 1 & & & & & \\ 0 & 0 & & & & & \\ & & 0 & 1 & & & \\ & & 0 & 0 & & & \\ & & & & 0 & 1 & 0 \\ & & & & 0 & 0 & 1 \\ & & & & 0 & 0 & 0 \end{pmatrix}$$

$$B = \begin{pmatrix} 0 & & & & & & \\ & 0 & 1 & 0 & & & \\ & 0 & 0 & 1 & & & \\ & 0 & 0 & 0 & & & \\ & & & & 0 & 1 & 0 \\ & & & & 0 & 0 & 1 \\ & & & & 0 & 0 & 0 \end{pmatrix}$$

Then A and B are not similar but $p_A(t) = p_B(t) = t^7$, $\sigma(A) = \sigma(B) = \{0\}$, the algebraic multiplicities of 0 are both 7, and the geometric multiplicities of 0 are both 3.

Definition 72. $D = (d_{ij}) \in M_n$ is *diagonal* if $d_{ij} = 0$ for all $i \neq j$. We write $D = \operatorname{diag}(d_{11}, \ldots, d_{nn})$. We say $A \in M_n$ is *diagonalizable* if there is diagonal D and $S \in M_n^{-1}$ such that $D = SAS^{-1}$.

Proposition 73. *$A \in M_n$ is diagonalizable if and only if there is a basis for $\mathbb{C}^n$ of eigenvectors of A.*

Proof.

($\Longrightarrow$) Suppose $D = SAS^{-1}$, where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then $AS^{-1} = S^{-1}D$. Hence $AS^{-1}e_j = S^{-1}De_j = \lambda_j S^{-1}e_j$, and $\{ S^{-1}e_1, \ldots, S^{-1}e_n \}$ is a basis for $\mathbb{C}^n$ of eigenvectors of A.

($\Longleftarrow$) Suppose $\mathcal{B} = \{ v_1, \ldots, v_n \}$ is a basis with $Av_j = \lambda_j v_j$. Then $R = [v_1 \mid \cdots \mid v_n]$ is invertible, and $ARe_j = Av_j = \lambda_j v_j = \lambda_j v_j = \lambda_j Re_j = RDe_j$. So, if $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, then $R^{-1}AR = D$.

☐ Proposition 73

Lemma 74. *Suppose $A \in M_n$, $B \in M_m$. Let*

$$C = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \in M_{n+m}$$

Then C is diagonalizable if and only if A and B are.

Proof.

($\Longleftarrow$) Let $D_1 = SAS^{-1}$, $D_2 = RBR^{-1}$. Then

$$\begin{pmatrix} S & 0 \\ 0 & R \end{pmatrix}$$

diagonalizes C.

($\Longrightarrow$) Suppose

$$S^{-1}CS = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}$$

where

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

Then

$$CSe_j = S \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix} e_j = S\lambda_j e_j = \lambda_j Se_j$$

where

$$Se_j = \begin{pmatrix} x_j \\ y_j \end{pmatrix} \in \mathbb{F}^{n+m}$$

Then

$$\begin{pmatrix} Ax_j \\ By_j \end{pmatrix} = C \begin{pmatrix} x_j \\ y_j \end{pmatrix} = \begin{pmatrix} \lambda x_j \\ \lambda y_j \end{pmatrix}$$

So the $x_i \in \mathbb{F}^n$ are all eigenvectors of A, and the $y_i \in \mathbb{F}^m$ are all eigenvectors of B.

Notice, though, that

$$S = \begin{pmatrix} x_1 & \cdots & x_{n+m} \\ y_1 & \cdots & y_{n+m} \end{pmatrix}$$

and S is invertible; so $\operatorname{rank}(S) = n + m$. Thus

$$\operatorname{rank}_r[x_1 \mid \cdots \mid x_{n+m}] = n$$
$$\operatorname{rank}_r[y_1 \mid \cdots \mid y_{n+m}] = m$$

and in particular

$$\operatorname{rank}_c[x_1 \mid \cdots \mid x_{n+m}] = n$$
$$\operatorname{rank}_c[y_1 \mid \cdots \mid y_{n+m}] = m$$

Thus some subset of n of the vectors $\{ x_1, \ldots, x_{n+m} \}$ are linearly independent. This set of n is a basis for $\mathbb{F}^n$; thus A has a basis of eigenvectors, and A is diagonalizable. Proof that R is diagonalizable is mutatis mutandis.

$\square$ Lemma 74

Proposition 75. *Suppose*

$$B = \begin{pmatrix} B_1 & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & B_k \end{pmatrix}$$

with each $B_j \in M_{n_j}$. Then B is diagonalizable if and only if each B_j is diagonalizable.

Proof. $b = 2$ was done above. Assume the result holds for $k \in \mathbb{N}$; we show the result for $k+1$. Assume

$$B = \begin{pmatrix} B_1 & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & B_{k+1} \end{pmatrix} = \begin{pmatrix} C & 0 \\ 0 & B_{k+1} \end{pmatrix}$$

Then B is diagonalizable if and only if C and B_{k+1} are, which holds if and only if $B_1, \dots, B_k$ and B_{k+1} are by induction. $\square$ Proposition 75

Definition 76. A set $\mathcal{F} \subseteq M_n$ is called *simultaneously diagonalizable* if there is $S \in M_n^{-1}$ such that SAS^{-1} is diagonal for all $A \in \mathcal{F}$. We say $\mathcal{F}$ is *commuting* if for all $A, B \in \mathcal{F}$, we have $AB = BA$.

Theorem 77. *Let $\mathcal{F} \subseteq M_n$. Then $\mathcal{F}$ is simultaneously diagonalizable if and only if*

1. *Each $A \in \mathcal{F}$ is diagonalizable.*
2. *$\mathcal{F}$ is commuting.*

Proof. We do the case $\mathcal{F} = \{ A, B \}$.

($\Longrightarrow$) Suppose there is $S \in M_n^{-1}$ such that $S^{-1}AS$ and $S^{-1}BS$ are both diagonal. Then

1. A and B were both diagonalizable.
2. Since diagonal matrices commute, we have

$$S^{-1}ABS = S^{-1}ASS^{-1}BS = S^{-1}BSS^{-1}AS = S^{-1}BAS$$

and thus

$$AB = BA$$

($\Longleftarrow$) Pick S such that $S^{-1}AS$ is diagonal. Write

$$S^{-1}AS = \begin{pmatrix} \lambda_1 I_{n_1} & & & 0 \\ & \lambda_2 I_{n_2} & & \\ & & \ddots & \\ 0 & & & \lambda_k I_{n_k} \end{pmatrix}$$

Then $AB = BA$ implies $(S^{-1}AS)(S^{-1}BS) = (S^{-1}BS)(S^{-1}AS)$. Thus, blocking $S^{-1}BS \in M_{n_1+\dots+n_k}$ with $R_{ij} \in M_{n_i,n_j}$, we have

$$(\lambda_i R_{ij}) = (R_{ij}\lambda_j)$$

So $R_{ij} = 0$ for $i \neq j$. So

$$S^{-1}BS = \begin{pmatrix} R_{11} & & 0 \\ & \ddots & \\ 0 & & R_{kk} \end{pmatrix}$$

But B is diagonalizable. So $S^{-1}BS$ is diagonalizable. So R_{ii} is diagonalizable for each $1 \leq i \leq k$. Pick $T_i \in M_{n_i}^{-1}$ such that $T^{-1}R_{ii}T_i$ is diagonal. Let

$$T = \begin{pmatrix} T_1 & & 0 \\ & \ddots & \\ 0 & & T_k \end{pmatrix}$$

Then $T(S^{-1}BS)T$ is diagonal, and

$$T^{-1}(S^{-1}AS)T = \begin{pmatrix} T_1^{-1} & & 0 \\ & \ddots & \\ 0 & & T_k^{-1} \end{pmatrix} \begin{pmatrix} \lambda_1 I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \lambda_k I_{n_k} \end{pmatrix} \begin{pmatrix} T_1^{-1} & & 0 \\ & \ddots & \\ 0 & & T_k^{-1} \end{pmatrix} = S^{-1}AS$$

is diagonal. So $(ST)^{-1}A(ST)$ and $(ST)^{-1}B(ST)$ are both diagonal.

General case was mumbled about.

□ Theorem 77

How does AB compare to BA?

Well, if A is invertible, then $A^{-1}(AB)A = BA$, so $AB \sim BA$. So $p_{AB}(t) = p_{BA(t)}$ and $\sigma(AB) = \sigma(BA)$.

Example 78.

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$
$$B = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$$

Then

$$AB = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$
$$BA = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$$

Then $AB \not\sim BA$. Also $p_{AB}(t) = t^2 = p_{BA}(t)$. So $\sigma(AB) = \sigma(BA)$.

Theorem 79. *For $A \in M_{m,n}$, $B \in M_{n,m}$, we have $t^n p_{AB}(t) = t^m p_{BA}(t)$.*

Corollary 80. *If $A \in M_{m,n}$, $B \in M_{n,m}$. Then $\sigma(AB) \cup \{0\} = \{0\} \cup \sigma(BA)$.*

Proof of Theorem 79. Observe

$$\begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix} \begin{pmatrix} I_m & -A \\ 0 & I_n \end{pmatrix} = I_{m+n}$$

Now

$$\begin{aligned} \begin{pmatrix} I_m & -A \\ 0 & I_n \end{pmatrix} \begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix} &= \begin{pmatrix} AB - AB & 0 \\ B & 0 \end{pmatrix} \begin{pmatrix} I_m & A \\ 0 & I_n \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix} \end{aligned}$$

Thus

$$\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \sim \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$$

So

$$t^n p_{AB}(t) = \det \begin{pmatrix} tI_m - AB & 0 \\ -B & tI_n \end{pmatrix} = \det \begin{pmatrix} tI_m & 0 \\ -B & tI_n - BA \end{pmatrix} = t^m p_{BA}(t)$$

□ Theorem 79

2.6 Persistence of eigenvalues

Theorem 81. *Suppose $A \in M_n$ $\lambda \in \mathbb{C}$, and $1 \le k \le n$. Consider the statements*

(a) *λ is an eigenvalue of A of geometric multiplicity $\ge k$.*

(b) *For all $m > n - k$ and for all $S \subseteq \{1, \dots, n\}$ with $|S| = m$, we have that λ is an eigenvalue of $\widehat{A} = A[S, S]$.*

(c) *λ is n eigenvalue of algebraic multiplicity $\ge k$.*

Then (a) $\implies$ (b) $\implies$ (c).

Proof.

(a) $\implies$ (b) Suffices to do the case $S = \{1, \dots, m\}$. Write

$$A = \begin{pmatrix} \widehat{A} & B \\ C & D \end{pmatrix}$$

Pick $v_1, \dots, v_k$ linearly independent eigenvectors with $Av_i = \lambda v_i$. Say

$$v_i = \begin{pmatrix} u_i \\ w_i \end{pmatrix}$$

where $u_i \in \mathbb{C}^m$ and $w_i \in \mathbb{C}^{n-m}$. Since $k > n - m$, we have $\{w_1, \dots, w_k\}$ is linearly dependent. So there are α_i not all 0 such that

$$\alpha_1 w_1 + \cdots + \alpha_k w_k = 0$$

Thus

$$\alpha_1 v_1 + \cdots + \alpha_k v_k = \begin{pmatrix} u \\ 0 \end{pmatrix}$$

and $u \neq 0$ since $\{v_1, \dots, v_k\}$ is linearly independent. Then

$$\begin{aligned} Av &= \sum_i A(\alpha_i v_i) \\ &= \sum_i \alpha_i A v_i \\ &= \sum_i \alpha_i \lambda v_i \\ &= \lambda v \end{aligned}$$

So

$$\begin{pmatrix} \widehat{A} & B \\ C & D \end{pmatrix} \begin{pmatrix} u \\ 0 \end{pmatrix} = \begin{pmatrix} \widehat{A}u \\ Cu \end{pmatrix} = \lambda \begin{pmatrix} u \\ 0 \end{pmatrix}$$

and $\widehat{A}u = \lambda u$.

(b) $\implies$ (c)

Lemma 82. *Given $\mu_1, \dots, \mu_n$ such that $S_m(\mu_1, \dots, \mu_n) = 0$ for all $m > n - k$, then at least k of the μ_i are 0.*

Proof. Well

$$\mu_1 \dots \mu_n = S_n(\mu_1, \dots, \mu_n) = 0$$

So some $\mu_i = 0$; say $\mu_1 = 0$. Also

$$\begin{aligned} 0 &= S_{n-1}(\mu_1, \dots, \mu_n) \\ &= \mu_2 \dots \mu_n + \mu_1(\mu_3 \dots \mu_n + \mu_2\mu_4 \dots \mu_n + \cdots + \mu_2 \dots \mu_{n-1}) \\ &= \mu_2 \end{aligned}$$

So there is $2 \le i \le n$ such that $\mu_i = 0$; say $\mu_2 = 0$. Continuing as above, we find k distinct i such that $\mu_i = 0$. ☐ Lemma 82

Suppose λ is an eigenvalue of $A[S,S]$ for all $|S| = m$, all $m > n-k$. Then $\det(\lambda I - A[S,S]) = 0$ for all such S, m. Look at $\lambda I - A$. We then have

$$\begin{aligned} E_m(\lambda I - A) &= \sum_{|S|=m} \det(\lambda I - A[S,S]) \\ &= 0 \end{aligned}$$

So $S_m(\lambda - \lambda_1, \ldots, \lambda - \lambda_n) = 0$ for all $m > n-k$. So there is k distinct i such that $\lambda = \lambda_i$. So $(t-\lambda)^k$ is a factor of

$$\prod_{i=1}^{n} (t - \lambda_i) = p_A(t)$$

$\square$ Theorem 81

Corollary 83. *Suppose $A \in M_n$ has $\dim(\mathcal{N}(A)) = k$. Then for all $m > n-k$ and all $|S| = m$, we have that $A[S,S]$ is not invertible.*

Proof. Since $\dim(\mathcal{N}(A)) = k$, we have that 0 has geometric multiplicity k; then apply (b).

$\square$ Corollary 83

3 Unitaries and isometries

Definition 84. A map $L \colon \mathbb{F}^n \to \mathbb{F}^m$ is called a *isometry* if for all $x \in \mathbb{F}^n$, we have

$$\|x\|_2 = \|Lx\|_2$$

Remark 85. Suppose $L \colon \mathbb{F}^n \to \mathbb{F}^m$ is an isometry. Then $\mathrm{dist}(x,y) = \|x-y\|_2 = \|Lx - Ly\|_2 = \mathrm{dist}(Lx, Ly)$. We also have $\mathcal{N}(L) = \{0\}$, so L is injective, and $m \geq n$.

Theorem 86. *Suppose $V \in M_{m,n}$ with $m \geq n$. Then the following are equivalent:*

1. *L_V is an isometry.*
2. *The columns of V are orthonormal.*
3. *$V^*V = I_n$.*

Proof.

(1) $\implies$ (2) Let $V = [v_1 \mid \cdots \mid v_n]$. Then

$$\|v_j\|_2 = \|L_v e_j\|_2 = \|e_j\|_2 = 1$$

Take any $i \neq j$, and $|\alpha| = 1$. Then

$$\|v_i + \alpha v_j\|_2 = \|L_V(e_i + \alpha e_j)\|_2 = \|e_i + \alpha e_j\|_2 = \sqrt{1 + |\alpha|^2} = \sqrt{2}$$

So

$$\begin{aligned} \langle v_i + \alpha v_j, v_i + \alpha v_j \rangle &= 2 \\ \implies \langle v_i, v_i \rangle + \overline{\alpha} \langle v_i, v_j \rangle + \alpha \langle v_j, v_i \rangle + \langle v_j, v_j \rangle &= 2 \\ \implies \overline{\alpha} \langle v_i, v_j \rangle + \overline{\overline{\alpha} \langle v_i, v_j \rangle} &= 0 \end{aligned}$$

Pick $|\alpha| = 1$ so that $\overline{\alpha}\langle v_i, v_j \rangle = |\langle v_i, v_j \rangle|$. Then $2|\langle v_i, v_j \rangle| = 0$, and $v_i \perp v_j$.

(2) $\implies$ (3) Write $V = [v_1 \mid \cdots \mid v_n]$. Then

$$V^* = \begin{pmatrix} v_1^* \\ \vdots \\ v_n^* \end{pmatrix}$$

Then $V^*V = (v_i^* v_j) = (\langle v_j, v_i \rangle) = I_n$.

(3) $\implies$ (1) Note that

$$\begin{aligned}
\|L_V(x)\|_2^2 &= \|Vx\|_2^2 \\
&= \langle Vx, Vx\rangle \\
&= (Vx)^*(Vx) \\
&= x^2 V^* Vx \\
&= x^* I x \\
&= x^* x \\
&= \|x\|_2^2
\end{aligned}$$

□ Theorem 86

What can be said about the rows of an isometry?

Well, write $V^* = [r_1 \mid \dots r_n]$. Then

$$V = \begin{pmatrix} r_1^* \\ \vdots \\ r_n^* \end{pmatrix}$$

and

$$Vx = \begin{pmatrix} r_1^* x \\ \vdots \\ r_n^* x \end{pmatrix} = \begin{pmatrix} \langle x, r_1\rangle \\ \vdots \\ \langle x, r_m\rangle \end{pmatrix}$$

Definition 87. A set of vectors $\{ r_1, \dots, r_m \} \subseteq \mathbb{F}^n$ is called a *Parseval frame* for $\mathbb{F}^n$ if

$$\|x\|_2^2 = \sum_{i=1}^{m} |\langle x, r_i\rangle|^2$$

for all $x \in \mathbb{F}^n$.

Proposition 88. *Suppose $\{ r_1, \dots, r_m \} \subseteq \mathbb{F}^n$ is a Parseval frame if and only if*

$$V = \begin{pmatrix} r_1^* \\ \vdots \\ r_m^* \end{pmatrix}$$

is an isometry.

Theorem 89. *Let $\{ r_1, \dots, r_m \} \subseteq \mathbb{F}^n$. Then $\{ r_1, \dots, r_m \}$ are a Parseval frame if and only if for all $x \in \mathbb{F}^n$, we have that*

$$x = \sum_{i=1}^{m} \langle x, r_i\rangle r_i$$

Proof.

($\implies$) Let

$$V = \begin{pmatrix} r_1^* \\ \vdots \\ r_m^* \end{pmatrix}$$

Then $V \colon \mathbb{F}^n \to \mathbb{F}^m$ is an isometry. Then

$$V^* = [r_1 \mid \cdots \mid r_n] \colon \mathbb{F}^m \to \mathbb{F}^n$$

and $V^* e_j = r_j$. Since V is an isometry, we have that

$$\begin{aligned} x &= I_n x \\ &= V^* V(x) \\ &= V^* \begin{pmatrix} \langle x, r_1 \rangle \\ \vdots \\ \langle x, r_m \rangle \end{pmatrix} \\ &= V^* \left(\sum_{j=1}^{m} \langle x, r_j \rangle e_j \right) \\ &= \sum_{j=1}^{m} \langle x, r_j \rangle r_j \end{aligned}$$

($\Longleftarrow$) Let

$$V = \begin{pmatrix} r_1^* \\ \vdots \\ r_m^* \end{pmatrix}$$

Then

$$\begin{aligned} V^* V x &= V^* \begin{pmatrix} \langle x, r_1 \rangle \\ \vdots \\ \langle x, r_m \rangle \end{pmatrix} \\ &= \sum_{j=1}^{m} \langle x, r_j \rangle r_j \\ &= x \end{aligned}$$

So $V^* V x = x$ for all x. So $V^* V = I_n$ and V is an isometry. So $\{ r_1, \ldots, r_m \}$ is a Parseval frame.

$\square$ Theorem 89

Definition 90. A Parseval frame $\{ r_1, \ldots, r_m \} \subseteq \mathbb{F}^n$ is called *uniform* if $\|r_i\| = \|r_j\|$ for all $i, j \in \{ 1, \ldots, m \}$. It is called *equiangular* if it is uniform and $|\langle r_i, r_j \rangle|$ is constant for all $i \neq j$.

Fact 91.

1. *For all $m \geq n$ there exist uniform Parseval frames of size m.*
2. *Finding the pairs (m, n) such that there is exist equiangular Parseval frames is an area of current research.*

Fact 92 (The case of $\mathbb{R}$)**.**

1. *If there is an equiangular Parseval frame then $m \leq \frac{n(n+1}{2}$.*
2. *There are many n for which there does not exist an $\left(\frac{n(n+1)}{2}, n\right)$ equiangular Parseval frame. (i.e. ambient dimension is n, frame size $\frac{n(n+1)}{2}$.)*
3. *(m, n) equiangular Parseval frames exist if and only if there is a completely regular graph on m vertices with certin parameters that tell the value of n. (I believe Wikipedia knows these as "strongly regular graphs".)*

Fact 93 (The case of $\mathbb{C}$)**.**

1. *If there is an equiangular Parseval frame, then $m \leq n^2$.*

2. *Zauner's conjecture: for all n there is $\{ r_1, \ldots, r_{n^2} \}$ an equiangular Parseval frame for $\mathbb{C}^n$.*

3. *What pairs (m, n) have equiangular Parseval frames? Very little is known.*

A closely related problem:

Example 94. If $\{ u_1, \ldots, u_n \}$ and $\{ v_1, \ldots, v_n \}$ are both orthonormal bases for $\mathbb{C}^n$, then

$$\left\{ \frac{u_1}{\sqrt{2}}, \ldots, \frac{u_n}{\sqrt{2}}, \frac{v_1}{\sqrt{2}}, \ldots, \frac{v_n}{\sqrt{2}} \right\}$$

is a uniform Parseval frame, since

$$\|x\|^2 = \sum_{i=1}^{n} |\langle x, u_i \rangle|^2 = \sum_{i=1}^{n} |\langle x, v_i \rangle|^2$$

so

$$\|x\|^2 = \sum_{i=1}^{n} \left| \left\langle x, \frac{u_i}{\sqrt{2}} \right\rangle \right|^2 + \sum_{i=1}^{n} \left| \left\langle x, \frac{v_i}{\sqrt{2}} \right\rangle \right|^2$$

Definition 95. Two orthonormal bases $\{ u_1, \ldots, u_n \}$ and $\{ v_1, \ldots, v_n \}$ are called *mutually unbiased* if $|\langle u_i, v_j \rangle|$ is constant for $i, j \in \{ 1, \ldots, n \}$. It's not hard to see that the constant is $\frac{1}{\sqrt{n}}$.

Question 96. How many orthonormal bases can there be such that each pair of bases is mutually unbiased?

It is conjectured that the answer is $n + 1$; this is known when $n = p^k$ for p prime. The question is still open for $n = 6$.

Definition 97. A matrix $U \in M_n$ is *unitary* if $U^* U = I_n$.

Theorem 98. *Suppose $U \in M_n$. Then the following are equivalent:*

(a) *U is unitary.*

(b) *U is invertible and $U^{-1} = U^*$.*

(c) *$UU^* = I_n$.*

(d) *U^* is unitary.*

(e) *The columns of U are orthonormal.*

(f) *The rows of U are orthonormal.*

(g) *U is an isometry.*

Proof.

(a) $\iff$ **(g)** $\iff$ **(e)** By isometry theorem.

(a) $\implies$ **(b)** U^* is a left inverse of U implies that U^* is a right inverse of U.

(b) $\implies$ **(c)** If $U^* = U^{-1}$ then $I = UU^{-1} = UU^*$.

(c) $\implies$ **(d)** $I = UU^* = (U^*)^* U^*$, so U^* is unitary.

(d) $\implies$ **(f)** Since U^* is unitary, we have that U^* is an isometry. So the columns of U^* are orthonormal. So the rows of U are orthonormal.

(f) $\implies$ **(d)** If the rows of U are orthonormal, then $UU^* = I$, and U^* is unitary.

□ Theorem 98

Proposition 99. *If $\{ u_1, \ldots, u_n \}$ is an orthonormal set of vectors, then they are linearly independent.*

Proof. Suppose

$$\alpha_1 u_1 + \cdots + \alpha_n u_n = 0$$

Then

$$\begin{aligned} 0 &= \langle \alpha_1 u_1 + \cdots + \alpha_n u_n, \alpha_1 u_1 + \cdots + \alpha_n u_n \rangle \\ &= \sum_{i,j=1}^{n} \alpha_i \overline{\alpha_j} \langle u_i, u_j \rangle \\ &= \sum_{i=1}^{n} |\alpha_i|^2 \end{aligned}$$

so $\alpha_i = 0$ for all $i \in \{1, \dots, n\}$. So $\{u_1, \dots, u_n\}$ is linearly independent. □ Proposition 99

Hence an orthonormal set of n vectors in $\mathbb{C}^n$ is a basis for $\mathbb{C}^n$.

Remark 100.

1. The set of $n \times n$ invertible matrices M_n^{-1} is a grape, usually denoted $\mathrm{GL}(n; \mathbb{F})$.
2. The subset $\mathcal{U}(n) \subseteq M_n^{-1}$ of unitary matrices is a subgrape, since if $U, V \in \mathcal{U}(n)$, then

$$(UV)^*(UV) = V^*U^*UV = V^*V = I_n$$

and if $U \in \mathcal{U}(n)$, then $U^{-1} = U^* \in \mathcal{U}(n)$.

When $\mathbb{F} = \mathbb{R}$, the unitaries are often called the *orthogonal* matrices and denoted $\mathcal{O}(n)$.

Definition 101. Given matrices $A = (a_{i,j})$ and $B = (b_{i,j})$ in $M_{m,n} \cong \mathbb{C}^{mn}$, we can think of

$$\mathrm{dist}(A, B) = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{i,j} - b_{i,j}|^2 \right)$$

In this metric, given a sequence of matrices $A_k = (a_{i,j}(k)) \to A = (a_{i,j})$ as $k \to \infty$ if and only if each $a_{i,j}(k) \to a_{i,j}$ as $k \to \infty$.

Remark 102. $\frac{1}{k} I_n \in \mathrm{GL}(n, \mathbb{F}) = M_n^{-1}$ but $\frac{1}{k} I_n \to 0 \notin M_n^{-1}$. So $\mathrm{GL}(n, \mathbb{F})$ is not closed.

Lemma 103.

1. *Suppose $U_k \in M_n$ are unitary an $(U_k : k \in \mathbb{N}) \to U$. Then U is unitary. i.e. $\mathcal{U}(n)$ is a closed subset of M_n.*
2. *$\mathcal{U}(n) \subseteq M_n$ is compact.*
3. *Suppose $U_k \in M_n$ are unitary. Then there is a subsequence $(U_{k_j} : j \in \mathbb{N})$ that converges to some $U \in M_n$ (which must then be unitary by part (1)).*

Proof.

1. Let $U_k = (u_{ij}(k))$; let $U = (u_{ij})$. We then have that

$$\lim_{k \to \infty} u_{ij}(k) = u_{ij}$$

for all i, j. Note also that $U^* = (\overline{u_{ji}})$, and that

$$U^*U = \left(\sum_{\ell=1}^{n} \overline{u_{\ell,i}} u_{\ell,j} \right) = \left(\lim_{k \to \infty} \sum_{\ell=1}^{n} \overline{u_{\ell,i}(k)} u_{\ell,j}(k) \right) = I_n$$

So $U \in \mathcal{U}(n)$.

2. We just showed that $\mathcal{U}(n)$ is closed. Also $U \in \mathcal{U}(n)$ means that the columns are orthonormal. So

$$\|U\|_2^2 = \sum_{i=1}^{n} |u_{ij}|^2 = n$$

   So $\mathcal{U}(n)$ is closed and bounded. So it is compact.

3. Heine-Borel.

□ Lemma 103

Remark 104.

1. Suppose $U \in \mathcal{U}(n)$. Then $U^*U = I$. So

$$\begin{aligned} 1 &= \det(U^*U) \\ &= \det(U^*)\det(U) \\ &= \overline{\det(U)}\det(U) \\ &= |\det(U)| \end{aligned}$$

2. If $U \in \mathcal{O}(n)$ then $\det(U) \in \{ \pm 1 \}$.

3. We look at $\mathbb{R}^2$. The matrix representing a rotation by θ is then

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

   Then $R(\theta)$ is unitary, and $\det(R(\theta)) = 1$. Also, if we have a unitary U with

$$Ue_1 = \begin{pmatrix} c \\ s \end{pmatrix}$$

   for some c, s, we know that

$$Ue_2 \perp \begin{pmatrix} c \\ s \end{pmatrix}$$

   and $\|Ue_2\|_2 = 1$; so

$$Ue_2 = \begin{pmatrix} \pm s \\ \mp c \end{pmatrix}$$

   We then conclude that

$$\mathcal{O}(2) = \{ R(\theta) : 0 \le \theta < 2\pi \} \cup \left\{ R(\theta) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} : 0 \le \theta < 2\pi \right\}$$

   and further that the former have determinant 1 while the latter have determinant -1. Finally, note that

$$R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2) = R(\theta_2)R(\theta_1)$$

3.1 Householder transformations

Suppose $w \in \mathbb{C}^n \setminus \{ 0 \}$. Define

$$U_w x = x - \frac{2}{\|w\|_2^2}(ww^*)x$$

i.e.

$$U_w = I - \frac{2}{\|w\|_2^2}(ww^*)$$

to be the *Householder transformation* of w. Note that

$$U_w = I - 2\left(\left(\frac{w}{\|w\|}\right)\left(\frac{w}{\|w\|}\right)^*\right)$$

In practice, these are normalized for unit vectors.

Given $v \in \mathbb{C}^n$, write $v = v_1 + v_2$ where $v_1 \perp w$ and $v_1 = \alpha w$. Then

$$\begin{aligned}
U_w v &= v - 2\left(\frac{w}{\|w\|}\frac{w^*}{\|w\|}\right)v \\
&= v - 2\left\langle v, \frac{w}{\|w\|}\frac{w}{\|w\|}\right\rangle \\
&= v_1 + v_2 - 2\left\langle v_1, \frac{w}{\|w\|}\right\rangle\frac{w}{\|w\|} \\
&= v_1 + v_2 - 2\alpha\left\langle w, \frac{w}{\|w\|}\right\rangle\frac{w}{\|w\|} \\
&= v_1 + v_2 - 2\alpha w \\
&= v_1 + v_2 - 2v_1 \\
&= v_2 - v_1
\end{aligned}$$

Geometrically, this corresponds to negating the w component.

Remark 105. $U_w^* = U_w$; then $U_w^* U_w = U_w^2 = I$. So U_w is unitary.

3.2 Unitary equivalence

Definition 106. We say $A, B \in M_n$ are *unitarily equivalent* (written $A \sim_{\text{u.e.}} B$) if there is $U \in \mathcal{U}(n)$ such that $B = U^*AU$.

Remark 107.

1. This is an equivalence relation.
2. Since $U^* = U^{-1}$, we have that $A \sim_{\text{u.e.}} B \implies A \sim B$.

Proposition 108. *Let $A, B \in M_n$. If $A \sim_{\text{u.e.}} B$ then*

$$\sum_{i,j=1}^{n} |a_{ij}|^2 = \sum_{i,j=1}^{n} |b_{ij}|^2$$

We prove this using traces.

Proposition 109. *Suppose $A \in M_{m,n}$; suppose $B \in M_{n,m}$. Then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.*

Proof. Well,

$$\begin{aligned}
\operatorname{tr}(AB) &= \sum_{i=1}^{m}\sum_{k=1}^{n} a_{ik}b_{ki} \\
&= \sum_{q=1}^{n}\sum_{\ell=1}^{m} b_{q\ell}q_{\ell q} \\
&= \operatorname{tr}(BA)
\end{aligned}$$

□ Proposition 109

Proposition 110. *Suppose $A \in M_{m,n}$. Then*

$$\operatorname{tr}(A^*A) = \operatorname{tr}(AA^*) = \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2$$

Proof. Well, $A^* = (\overline{a_{j,i}})$. Then

$$\begin{aligned}\operatorname{tr}(A^*A) &= \sum_i \sum_k \overline{a_{ki}} a_{ki} \\ &= \sum |a_{i,j}|^2\end{aligned}$$

☐ Proposition 110

Proof of Proposition 108. Suppose $B = U^*AU$. Then

$$\begin{aligned}\sum_{i,j} |b_{ij}|^2 &= \operatorname{tr}(B^*B) \\ &= \operatorname{tr}((U^*A^*U)(U^*AU)) \\ &= \operatorname{tr}(U^*(A^*AU)) \\ &= \operatorname{tr}((A^*AU)U^*) \\ &= \operatorname{tr}(A^*A) \\ &= \sum_{i,j} |a_{i,j}|^2\end{aligned}$$

☐ Proposition 108

Example 111. Let

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix}$$
$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Then A has 3 distinct eigenvalues; so we have a basis of eigenvectors. So A is diagonalizable. So $A \sim B$. But

$$\sum |a_{i,j}|^2 > \sum |b_{i,j}|^2$$

so $A \not\sim_{\text{u.e.}} B$.

Example 112 (H. Radjavi). Suppose

$$A = \begin{pmatrix} \lambda_1 & p_1 & & & \\ & \lambda_2 & p_2 & & \\ & & \ddots & \ddots & \\ & & & \lambda_{n-1} & p_{n-1} \\ & & & & \lambda_n \end{pmatrix}$$

If $\lambda_i \neq \lambda_j$ for all $i \neq j$ and $p_i > 0$ for all i and A' is another usch matrix with $\lambda_i = \lambda_i'$, then $A \sim_{\text{u.e.}} A'$ if and only if $A = A'$.

3.3 Specht's invariants

Let s, t be free non-commuting variables. Given

$$w(s,t) = s^{n_1} t^{m_1} \dots s^{n_k} t_{m_k}$$

where all $m_i \geq 0$, all $n_i \geq 0$, and given $A, B \in M_n$, we can set

$$w(A,B) = A^{n_1} B^{m_1} \dots A^{n_k} B^{m_k}$$

We also set $|w|$ to be the length of w; i.e.

$$n_1 + m_1 + \dots + n_k + m_k$$

Theorem 113 (Specht, 1940)**.** *Suppose $A, B \in M_n$. Then $A \sim_{\text{u.e.}} B$ if and only if* $\operatorname{tr}(w(A, A^*)) = \operatorname{tr}(w(B, B^*))$ *for all words w.*

Proof.

($\implies$) Suppose $A \sim_{\text{u.e.}} B$; say $B = U^*AU$. Then $B^k = U^*A^kU$. Also $B^* = U^*AU$; so $(B^*)^k = U^*(A^*)^kU$. Then

$$\begin{aligned} w(B, B^*) &= B^{n_1}(B^*)^{m_1} \ldots B^{n_k}(B^*)^{m_k} \\ &= (U^*A^{n_1}U)(U^*(A^*)^{m_1}U) \ldots (U^*A^{n_k}U)(U^*(A^*)^{m_k}U) \\ &= U^*W(A, A^*)U \end{aligned}$$

So

$$\operatorname{tr}(w(B, B)) = \operatorname{tr}(U^*w(A, A^*)U) = \operatorname{tr}(w(A, A^*))$$

($\impliedby$)

Lemma 114. *Suppose V, W are vector spaces with* $\operatorname{span}\{ v_\alpha : \alpha \in I \} = V$ *and* $\operatorname{span}\{ w_\alpha : \alpha \in I \} = W$. *Then there is linear $L \colon V \to W$ with $L(v_\alpha) = w_\alpha$ for all $\alpha \in I$ if and only if whenever*

$$\sum_{i=1}^{n} \lambda_i v_{\alpha_i} = 0$$

we also have

$$\sum_{i=1}^{n} \lambda_i w_{\alpha_i} = 0$$

Definition 115. Let $\mathcal{A} \subseteq M_n$; then $\mathcal{A}$ is an *algebra* if

1. $\mathcal{A}$ is a vector subspace
2. If $x, y \in \mathcal{A}$ then $xy \in \mathcal{A}$.

It is a $*$-*algebra* if $X \in \mathcal{A}$ implies that $X^* \in \mathcal{A}$. If $\mathcal{A}, \mathcal{B}$ are $*$-algebras, then a map $\pi \colon \mathcal{A} \to \mathcal{B}$ is called a $*$-*homomorphism* if

1. π is linear
2. $\pi(xy) = \pi(x)\pi(y)$
3. $\pi(x^*) = \pi(x)^*$

Proposition 116. *Let $A, B \in M_n$. Let*

$$\begin{aligned} \mathcal{A} &= \operatorname{span}\{ w(A, A^*) : w \text{ word} \} \\ \mathcal{B} &= \operatorname{span}\{ w(B, B^*) : w \text{ word} \} \end{aligned}$$

(Then these are $*$-algebras.) If for all words w we have $\operatorname{tr}(w(A, A^*)) = \operatorname{tr}(w(B, B^*))$ *then there is a $*$-isomorphism $\pi \colon \mathcal{A} \to \mathcal{B}$ with $\pi(w(A, A^*)) = w(B, B^*)$.*

Proof. By lemma, there is a *linear* map $\pi \colon \mathcal{A} \to \mathcal{B}$ satisfying the above if and only if

$$\sum_{i=1}^{n} w_i(A, A^*) = 0 \implies \sum_{i=1}^{n} \lambda_i w_i(B, B^*) = 0$$

Well, let

$$\begin{aligned} X &= \sum_{i=1}^{n} \lambda_i w_i(A, A^*) \\ Y &= \sum_{i=1}^{n} \lambda_i w_i(B, B^*) \end{aligned}$$

Let $X = (x_{ij})$. Then

$$X = 0 \iff \sum |x_{ij}| = 0 \iff \operatorname{tr}(X^*X) = 0$$

But

$$X^*X = \sum \overline{\lambda_j} \lambda_i w_j(A, A^*)^* w_i(A, A^*)$$

So

$$\operatorname{tr}(X^*X) = \sum \lambda_j \lambda_i \operatorname{tr}(w_j(A, A^*)^* w_i(A, A^*)) = \sum \lambda_j \lambda_i \operatorname{tr}(w_j(B, B^*)^* w_i(B, B^*)) = \operatorname{tr}(Y^*Y)$$

So

$$X = 0 \implies \operatorname{tr}(X^*X) = 0 \implies \operatorname{tr}(Y^*Y) = 0 \implies Y = 0$$

So there is a well-defined linear map π with $\pi(w(A, A^*)) = w(B, B^*)$.

Claim 117. *π is a $*$-homomorphism.*

Proof. Let

$$\begin{aligned} X_1 &= \sum \lambda_i w_i(A, A^*) \\ X_2 &= \sum \mu_\ell \widetilde{w}_\ell(A, A^*) \end{aligned}$$

Then

$$\begin{aligned} \pi(X_1 X_2) &= \pi\Big(\sum \lambda_i \mu_\ell w_i(A, A^*) \widetilde{w}_\ell(A, A^*)\Big) \\ &= \sum \lambda_i \mu_\ell \pi(w_i(A, A^*) \widetilde{w}_\ell(A, A^*)) \\ &= \sum \lambda_i \mu_\ell w_i(B, B^*) \widetilde{w}_\ell(B, B^*) \\ &= \pi(X_1)\pi(X_2) \end{aligned}$$

Similarly, it is a $*$-homomorphism. ☐ Claim 117

Note that the same proof shows there is $\rho \colon \mathcal{B} \to \mathcal{A}$ such that $\rho(w(B, B^*)) = w(A, A^*)$; then $\rho = \pi^{-1}$. ☐ Proposition 116

Then the $*$-algebra generated by $\mathcal{A}$ is $*$-isomorphic to the $*$-algebra generated by $\mathcal{B}$. Wedderburn's theorem then yields

$$\mathcal{A} \cong M_{n_1} \oplus \cdots \oplus M_{n_k} \cong \mathcal{B}$$

It also yields that since $\mathcal{A} \subseteq M_n$ there are multiplicities $m_1, \dots, m_k$ such that

$$\mathcal{A} = \left\{ \begin{pmatrix} A_1 & & & & & & \\ & \ddots & & & & & \\ & & A_1 & & & & \\ & & & \ddots & & & \\ & & & & A_k & & \\ & & & & & \ddots & \\ & & & & & & A_k \end{pmatrix} \right\}$$

where A_i shows up m_i times. Similarly for $\mathcal{B}$, we get multiplicities $\widetilde{m}_1, \dots, \widetilde{m}_k$. Since the traces are equal, we have $m_i = \widetilde{m}_i$; so the $*$-isomorphisms are implemented by a unitary.

☐ Theorem 113

Theorem 118 (Pearcy 1968)**.** *Suppose $A, B \in M_n$. Then $A \sim_{\text{u.e.}} B$ if and only if* $\operatorname{tr}(w(A, A^*)) = \operatorname{tr}(w(B, B^*))$ *for all $|w| \leq 2n^2$. (This is 4^{n^2} words.)*

We use two lemmata.

Lemma 119. *Let* $\mathcal{L}_A(d) = \operatorname{span}\{ w(A, A^*) : |w| \leq d \}$. *Suppose* $\mathcal{L}_A(d) = \mathcal{L}_A(d+1)$. *Then* $\mathcal{L}_A(d) = \mathcal{A}$.

Proof. Suppose $m = d + 1 + k$. Suppose $|w| = m$. Write $w = w_1 w_2$ where $|w_1| = d + 1$ and $|w_2| = k$. Then

$$w(A, A^*) = w_1(A, A^*) w_2(A, A^*)$$

But

$$w_1(A, A^*) = \sum \lambda_\ell w_\ell(A, A^*)$$

where $|w_\ell| = d$. So $w(A, A^*)$ is a linear combination of things of length $d + k$. So $\mathcal{L}_A(d + 1 + k) = \mathcal{L}_A(d + k)$. By induction, we have $\mathcal{L}_A(d + k) = \mathcal{L}_A(d)$ for all $k \in \mathbb{N}$, and $\mathcal{A} = \mathcal{L}_A(d)$. □ Lemma 119

Lemma 120. $\mathcal{A} = \mathcal{L}_A(n^2)$.

Proof. Suppose there does not exist $d \leq n^2$ with $\mathcal{L}_A(d) = \mathcal{L}_A(d+1)$. Then $\{ 0 \} \subsetneqq \mathcal{L}_A(1) \subsetneqq \mathcal{L}_A(2) \subsetneqq \ldots \subsetneqq \mathcal{L}_A(n^2)$. So $\dim(\mathcal{L}_A(n^2)) \geq n^2$. But $\dim(\mathcal{A}) \leq n^2$, as $\mathcal{A} \subseteq M_n$. So $\mathcal{L}_A(n^2) = \mathcal{A}$. □ Lemma 120

Proof of Theorem 118.

($\implies$) Easy.

($\impliedby$) Again, want to show that there is $\pi \colon \mathcal{A} \to \mathcal{B}$ with $\pi(w(A, A^*)) = w(B, B^*)$ well-defined. We know that $\mathcal{A} = \mathcal{L}_A(n^2)$; for

$$X = \sum \lambda_i w_i(A, A^*)$$

and

$$Y = \sum \lambda_i w_i(B, B^*)$$

we now need $X = 0 \implies Y = 0$. But

$$\begin{aligned} X = 0 &\iff \operatorname{tr}(X^* X) = 0 \\ &\iff \sum \lambda_i \overline{\lambda_j} \operatorname{tr}(w_j(A, A^*)^* w_i(A, A^*)) \\ &\iff \sum \lambda_i \overline{\lambda_j} \operatorname{tr}(w_j(B, B^*)^* w_i(B, B^*)) \\ &\iff \operatorname{tr}(Y^* Y) = 0 \\ &\iff Y = 0 \end{aligned}$$

since $|w_j w_i| \leq 2n^2$.

□ Theorem 118

Theorem 121 (Djokovic-Johnson 2007)**.** *Suppose* $A, B \in M_n$. *Then* $A \sim_{\text{u.e.}} B$ *if and only if* $\operatorname{tr}(w(A, A^*)) = \operatorname{tr}(w(B, B^*))$ *for some set of at most*

$$n\sqrt{\frac{2n^2}{n-1} + \frac{1}{4}} + \frac{n}{2} - 2$$

words.

Fact 122. *In the case* $n = 2$, *it suffices to check the words* $\{ s, s^2, st \}$. *i.e.*

$$A \sim_{\text{u.e.}} B \iff \operatorname{tr}(A) = \operatorname{tr}(B), \operatorname{tr}(A^2) = \operatorname{tr}(B^2)), \operatorname{tr}(AA^*) = \operatorname{tr}(BB^*)$$

In the case $n = 3$, *it suffices to check*

$$\{ s, s^2, ts, s^3, s^2 t, s^2 t^2, s^2 t^2 st \}$$

In the case $n = 4$, *the paper exhibits* 20 *words.*

Recall now the Householder transformations U_w for $w \neq 0$.

Lemma 123. *Let $\|x\| = \|u\| = 1$ with $\langle x, u\rangle \geq 0$ and $w = u - x \neq 0$. Then $U_w u = x$ and $U_w x = u$.*

Proof. Draw a picture? □ Lemma 123

Theorem 124 (Schur)**.** *Suppose $A \in M_n$. Suppose $\lambda_1, \dots, \lambda_n$ are the roots of $p_A(t)$ (in some order). Then there is a unitary U such that $U^*AU = T$ is upper triangular with diagonal entries $t_{ii} = \lambda_i$. Moreover, U can be taken to be a product of Householder transformations.*

Proof. Since $\lambda_1 \in \sigma(A)$, we have that there is $u_1 \neq 0$ such that $Au_1 = \lambda_1 u_1$ with $\|u_1\| = 1$ and $\langle u_1, e_1\rangle \geq 0$. By the lemma, there is w such that $U_w u_1 = e_1$ and $U_w e_1 = u_1$. But then

$$\begin{aligned}
\langle U_w^* A U_w e_1, e_i\rangle &= \langle U_w^* A u_1, e_i\rangle \\
&= \langle U_w^* \lambda_1 u_1, e_i\rangle \\
&= \lambda_1 \langle U_w^* u_1, e_i\rangle \\
&= \lambda_1 \langle e_1, e_i\rangle \\
&= \begin{cases} 0 & i > 1 \\ \lambda_1 & \text{else} \end{cases}
\end{aligned}$$

Then

$$U_w^* A U_w = \begin{pmatrix} \lambda_1 & * \\ 0 & A_1 \end{pmatrix}$$

for $A_1 \in M_{n-1}$. But also

$$(t - \lambda_1)\dots(t - \lambda_n) = p_A(t) = p_{U_w^* A U_w}(t) = (t - \lambda_1) p_{A_1}(t)$$

So $p_{A_1}(t) = (t - \lambda_2)\dots(t - \lambda_n)$. We then repeat for $A_1 \in M_{n-1}$; by induction, we get some product of unitaries such that U^*AU has the desired form.

For the "moreover", recall that Householder unitaries are given by

$$U_w = I - \frac{2}{\|w\|^2} ww^*$$

Observe, however, that if $w \in \mathbb{C}^{n-1}$ and

$$\widetilde{w} = \begin{pmatrix} 0 \\ w_1 \end{pmatrix}$$

then

$$\begin{aligned}
U_{\widetilde{w}} &= I_n - \frac{2}{\|\widetilde{w}\|^2} \widetilde{w}\widetilde{w}^* \\
&= I_n - \frac{2}{\|w\|^2} \begin{pmatrix} 0 & 0 \\ 0 & w_1 w_1^* \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & I_{n-1} - \frac{2}{\|w\|^2} ww^* \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & U_w \end{pmatrix}
\end{aligned}$$

So we in fact get that the unitaries we conjugated by were Householder transformations. □ Theorem 124

Corollary 125. *Suppose $A \in M_n$; suppose $\lambda_1, \dots, \lambda_n$ are the roots of $p_A(t)$. Then $\operatorname{tr}(A^k) = \lambda_1^k + \dots + \lambda_n^k$.*

Proof. Pick a unitary U such that

$$U^*AU = \begin{pmatrix} \lambda_1 & * & * \\ 0 & \ddots & * \\ 0 & \dots & \lambda_n \end{pmatrix}$$

then

$$(U^*AU)^k = \begin{pmatrix} \lambda_1^k & * & * \\ 0 & \ddots & * \\ 0 & \dots & \lambda_n^k \end{pmatrix}$$

so

$$\operatorname{tr}(A^k = \operatorname{tr}(U^*A^kU) = \operatorname{tr}((U^*AU)^k) = \lambda_1^k + \cdots + \lambda^k$$

☐ Corollary 125

Remark 126. By Newton's identities, we can mutually solve for $S_1(\lambda_1, \dots, \lambda_n), \dots, S_n(\lambda_1, \dots, \lambda_n)$ in terms of $\mu_1, \dots, \mu_n$. But also recall that $-1)^k S_k(\lambda_1, \dots, \lambda_n)$ is the k^{th} coefficient of $p_A(t)$. So the coefficients of $p_A(t)$ are uniquely determined by the $\mu_k = \operatorname{tr}(A^k)$. Hence $\operatorname{tr}(A), \dots, \operatorname{tr}(A^n)$ determines $p_A(t)$, and thus $\lambda_1, \dots, \lambda_n$.

Theorem 127. *Suppose $A \in M_n$; suppose $\varepsilon > 0$. Then there is $B \in M_n$ such that $\|A - B\|_2 < \varepsilon$ such that B is invertible and diagonalizable.*

Proof. Suppose $\lambda_1, \dots, \lambda_n$ are the roots of $p_A(t)$. Pick

$$|\varepsilon_i| < \frac{\varepsilon}{\sqrt{n}}$$

such that $\lambda_1 + \varepsilon_1, \dots, \lambda_n + \varepsilon_n$ all distinct and non-zero. Then, by Schur's theorem, we have some U such that

$$U^*AU = T = \begin{pmatrix} \lambda_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_n \end{pmatrix}$$

Let $D = \operatorname{diag}(\varepsilon_1, \dots, \varepsilon_n)$; let $R = T + D$. Then the roots of $p_R(t)$ are $\lambda_1 + \varepsilon_1, \dots, \lambda_n + \varepsilon_n$. So R is invertible and diagonal. Let $B = URU^*$. Then

$$\begin{aligned}
\|A - B\|_2^2 &= \operatorname{tr}((A - B)^*(A - B)) \\
&= \operatorname{tr}((UTU^* - URU^*)^*(UTU^* - URU^*)) \\
&= \operatorname{tr}((U(T - R)U^*)^*(U(T - R)U^*)) \\
&= \operatorname{tr}(UD^*DU^*) \\
&= \operatorname{tr}(D^*D) \\
&= \sum_{i=1}^n |\varepsilon_i|^2 < \varepsilon^2
\end{aligned}$$

and B is invertible and diagonalizable. ☐ Theorem 127

Lemma 128. *Suppose*

$$R = \begin{pmatrix} 0 & * \\ 0 & R_1 \end{pmatrix}$$

$$S = \begin{pmatrix} S_1 & * \\ 0 & S_2 \end{pmatrix}$$

where the blocking is $n = k + (n - k)$ (i.e. $R, S_1 \in M_k$) and $S_2(1,1) = 0$ and S_2 is upper triangular. Then

$$RS = \begin{pmatrix} 0 & * \\ 0 & x \end{pmatrix}$$

is upper triangular, where the blocking is now $n = (k + 1) + (n - (k + 1))$.

Remark 129. If $q(t) = q_n t^n + \cdots + q_0$ and $U^*AU = T$, then $U^*AU = T$, so $q(T) = U^*q(A)U$.

Theorem 130 (Cayley-Hamilton)**.** *Suppose $A \in M_n$. Then $p_A(A) = 0$.*

Proof. Let $p_A(t) = (t - \lambda_1) \dots (t - \lambda_n)$; apply Schur's theorem to get $U^*AU = T$ upper triangular with $t_{ii} = \lambda_i$. Then $T - \lambda_1 I$ has a 1×1 block of 0 in the upper-left corner, and $T - \lambda_2 I$ has a 0 in the $(2, 2)$ entry. So, by lemma, we have that $(T - \lambda_1 I)(T - \lambda_2 I)$ has a 2×2 block of 0 in the top-left corner. Proceeding inductively, we get that $(T - \lambda_1 I) \dots (T - \lambda_k I)$ has a $k \times k$ block of 0 in the top-left corner, and $p_A(T) = 0$. But then $p_A(A) = U^* p_A(T) U = 0$. □ Theorem 130

Corollary 131. *Suppose $A \in M_n^{-1}$. Then $A^{-1} \in \operatorname{span}\{ I, A, \dots, A^{n-1} \}$.*

Proof. Well,

$$p_A(t) = t^n + a_{n-1}t^{n-1} + \dots + a_1 t + a_0$$

where $a_0 = \det(A) \neq 0$. Then, by Cayley-Hamilton, we have

$$0 = (A^n + a_{n-1}A^{n-1} + \dots + a_1 A + a_0 I)A^{-1}$$

so

$$0 = A^{n-1} + a_{n-1}A^{n-2} + \dots + a_1 I + a_0 A^{-1}$$

at which point we can solve for A^{-1}. □ Corollary 131

4 Linear maps on matrices

Suppose $A_1, \dots, A_r \in M_n$; suppose $B_1, \dots, B_r \in M_m$. Then there is $L \to M_{n,m} \to M_{n,m}$ defined by

$$L(Y) = A_1 Y B_1 + \dots + A_r Y B_r$$

These are called *elementary linear maps.*

Proposition 132. *Suppose $A \in M_n$ and $B \in M_m$. Let $L(Y) = AY - YB$. If $\sigma(A) \cap \sigma(B) = \emptyset$, then $L\colon M_{n,m} \to M_{n,m}$ is invertible.*

Proof. Suppose $Y \in \ker(L)$; then $AY = YB$. So

$$A^2 Y = A(AY) = A(YB) = YB^2$$

and inductively $A^k Y = YB^k$. Then for any polynomial $p(t)$, we have $p(A)Y = Yp(B)$. By Cayley-Hamilton, we have $0 = p_A(t)Y = Yp_A(B)$. So

$$\sigma(p_A(B)) = \{ p_A(\lambda) : \lambda \in \sigma(B) \}$$

So $0 \notin \sigma(p_A(B)$, and $p_A(B)$ is invertible. But $0 = Yp_A(B)$; so $0 = Y$. So $\ker(L) = \{ 0 \}$, and L is invertible. □ Proposition 132

Corollary 133. *Suppose $A \in M_n$, $B \in M_m$, and $X \in M_{n,m}$. If $\sigma(A) \cap \sigma(B) = \emptyset$, then*

$$\begin{pmatrix} A & X \\ 0 & B \end{pmatrix} \sim \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

Proof. If $Y \in M_{n,m}$, then

$$\begin{pmatrix} I_n & Y \\ 0 & I_m \end{pmatrix}^{-1} = \begin{pmatrix} I_n & -Y \\ 0 & I_m \end{pmatrix}$$

By proposition, we have $L(Y) = AY - YB$ is surjective; so there is Y such that $AY - YB = X$. But then

$$\begin{aligned}
\begin{pmatrix} I & Y \\ 0 & I \end{pmatrix} \begin{pmatrix} A & X \\ 0 & B \end{pmatrix} \begin{pmatrix} I & Y \\ 0 & I \end{pmatrix}^{-1} &= \begin{pmatrix} A & X + YB \\ 0 & B \end{pmatrix} \begin{pmatrix} I & -Y \\ 0 & I \end{pmatrix} \\
&= \begin{pmatrix} A & X + YB - AY \\ 0 & B \end{pmatrix} \\
&= \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}
\end{aligned}$$

□ Corollary 133

Theorem 134. *Suppose $A \in M_n$; let $p_A(t) = (t-\lambda_1)^{n_1} \dots (t-\lambda_k)^{n_k}$. Then*

$$A \sim \begin{pmatrix} T_1 & & 0 \\ & \ddots & \\ 0 & & T_k \end{pmatrix}$$

where each T_i is upper triangular with diagonal entries all equal to λ_i.

Proof. List eigenvalues as $\lambda_1, \dots, \lambda_1, \dots, \lambda_k, \dots, \lambda_k$ where λ_i appears n_i times. By Schur, there is a unitary U such that U^*AU is upper triangular with diagonal entries equal to the above list; then

$$U^*AU = \begin{pmatrix} T_{11} & T_{12} & \dots & T_{1k} \\ 0 & T_{22} & \dots & T_{2k} \\ \vdots & & \ddots & \\ 0 & \dots & T_{kk} & \end{pmatrix}$$

with each T_{ii} is $n_i \times n_i$ and upper triangular with diagonal entries λ_i. Blocking off the upper-left block, i.e. with $A = T_{11}$ and

$$B = \begin{pmatrix} T_{22} & & \\ & \ddots & \\ & & T_{kk} \end{pmatrix}$$

the corollary then yields that U^*AU is similar to

$$\begin{pmatrix} T_{11} & 0 & \dots & 0 \\ 0 & T_{12} & & \\ \vdots & & \ddots & \\ 0 & 0 & \dots & T_{kk} \end{pmatrix}$$

We then proceed by induction. □ Theorem 134

Definition 135. Let $J_k(\lambda) \in M_k$ be given by

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & & & 0 \\ 0 & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ 0 & & & & \lambda \end{pmatrix}$$

This is called the *elementary Jordan block*.

If we want to prove that each $A \in M_n$ is similar to a block diagonal matri each of whose blocks is of the form $J_k(\lambda)$ then it suffices to prove it for matrices of the form

$$T_i = \begin{pmatrix} \lambda_i & & * \\ & \ddots & \\ 0 & & \lambda_i \end{pmatrix}$$

i.e. $T_i = \lambda_i I + N_i$ where N_i is strictly upper triangular. If N is strictly upper triangular, we can prove N is similar to a block diagonal with blocks $J_k(0)$. So $T = \lambda_i I + N$ is similar to a block diagonal matrix with blocks $J_k(\lambda_i)$.

Remark 136. If $N \in M_m$ is strictly upper triangular, then $N^m = 0$.

Definition 137. We say $N \in M_n$ is *nilpotent* of there is k such that $N^k = 0$. The least such k is called the *order of nilpotency*.

Theorem 138. *Let $N \in M_n$ be nilpotent of order k. Then*

$$N \sim \begin{pmatrix} J_{m_1}(0) & & 0 \\ & \ddots & \\ 0 & & J_{m_r}(0) \end{pmatrix}$$

Furthermore, let ℓ_i be the number of Jordan blocks of size i (for $1 \leq i \leq k$); let $d_i = \dim(\mathcal{N}(N^i))$. Then

$$(\min\{\, i, j \,\}) \begin{pmatrix} \ell_1 \\ \vdots \\ \ell_k \end{pmatrix} = \begin{pmatrix} d_1 \\ \vdots \\ d_k \end{pmatrix}$$

Remark 139. Since we know that $(\min\{\, i, j \,\})$ is invertible, we get that

$$\begin{pmatrix} \ell_1 \\ \vdots \\ \ell_k \end{pmatrix} = (\min\{\, i, j \,\})^{-1} \begin{pmatrix} d_1 \\ \vdots \\ d_k \end{pmatrix}$$

So the d_i determines the Jordan structure.

5 QR factorization and Gram-Schmidt

Recall that if

$$B = (\overrightarrow{b_1} \mid \cdots \mid \overrightarrow{b_m})$$

and $R = (r_{ij})$, then

$$BR = \left(\sum_{j=1}^{m} r_{1j} \overrightarrow{b_j} \mid \ldots \right)$$

Recall $R \in M_{m,n}$ is called upper triangular when $r_{ij} = 0$ for all $i > j$.

Theorem 140 (QR)**.** *Let $A \in M_{m,n}$. Then there is upper triangular $R \in M_{m,n}$ and unitary $Q \in M_m$ such that $A = QR$.*

Proof.

Case 1. Suppose $m = n$ and A is invertible. Then

$$A = (\overrightarrow{a_1} \mid \cdots \mid \overrightarrow{a_m})$$

with $\{\, \overrightarrow{a_1}, \ldots, \overrightarrow{a_m} \,\}$ linearly independent and spanning $\mathbb{C}^m$. Recall that Gram-Schmidt gives us $\{\, \overrightarrow{u_1}, \ldots, \overrightarrow{u_m} \,\}$. orthonormal such that

$$\operatorname{span}\{\, \overrightarrow{a_1}, \ldots, \overrightarrow{a_j} \,\} = \operatorname{span}\{\, \overrightarrow{u_1}, \ldots, \overrightarrow{u_j} \,\}$$

So $\overrightarrow{a_j} \in \operatorname{span}\{\, \overrightarrow{u_1}, \ldots, \overrightarrow{u_j} \,\}$, and

$$\overrightarrow{a_j} = \sum_{i=1}^{j} r_{ij} \overrightarrow{u_j}$$

Set $r_{i,j} = 0$ for $i > j$. Since $\{\, \overrightarrow{u_1}, \ldots, \overrightarrow{u_m} \,\}$ is orthonormal, we have that $Q = (\overrightarrow{u_1} \mid \cdots \mid \overrightarrow{u_m})$ is unitary. Let $R = (r_{ij})$. Then $QR = A$.

Case 2. Suppose $m = n$ and A is singular.Then $0 \in \sigma(A)$. Let

$$t = \min\{\, |\lambda| : \lambda \in \sigma(A), \lambda \neq 0 \,\} > 0$$

Let $|\varepsilon_n| < t$ with $\varepsilon_n \to 0$ and $A(n) = A - \varepsilon_n I_n$ is invertible. So $A(n) = U(n)R(n)$. Note that

$$\|R(n)\|_2^2 = \operatorname{tr}(R_n^* R_n) = \operatorname{tr}(R(n)^* U(n)^* U(n) R(n)) = \operatorname{tr}(A(n)^* A(n))$$

is bounded. So $R(n)$ is bounded. Pick a subsequence such that $U(n_k) \to Q$ unitary and $R(n_k) \to R$ upper triangular. Then

$$A = \lim_k A(n_k) = \lim_k U(n_k) R(n_k) = QR$$

Case 3. Suppose $m < n$. Write $A = [A_1 \mid A_2]$ where $A_1 \in M_m$, $A_2 \in M_{m,n-m}$. By earlier case $A_1 = Q_1 R_1$ with $R_1 \in M_{m,n}$. Set $R_2 = Q^* A_2$. Check

$$Q[R_1 \mid R_2] = [QR_1 \mid QR_2] = A$$

Case 4. Suppose $m > n$. Let $\widetilde{A} = [A \mid 0] \in M_m$. Then $\widetilde{A} = QR$ where $R \in M_m$. Write $R = [R_1 \mid *]$ where $R_1 \in M_{m,n}$. Check that $A = QR_1$.

$\square$ Theorem 140

6 Normal and Hermitian

Definition 141. We say a matrix H is Hermitian if $H = H^*$. We let $(M_n)_h = \{\, H \in M_n : H = H^* \,\}$.

Remark 142. Suppose $A \in M_n$; set

$$\operatorname{Re}(A) = \frac{A + A^*}{2}$$
$$\operatorname{Im}(A) = \frac{A - A^*}{2i}$$

Then

1. $\operatorname{Re}(A), \operatorname{Im}(A) \in (M_n)_h$
2. $A = \operatorname{Re}(A) + i \operatorname{Im}(A)$
3. If $A = H + iK$ for $H, K \in (M_n)_h$, then $H = \operatorname{Re}(A)$ and $K = \operatorname{Im}(A)$.

Proof.

1. $$\operatorname{Re}(A)^* = \left(\frac{A + A^*}{2}\right)^* = \frac{A^* + A}{2} = \operatorname{Re}(A)$$

   Similarly we have $\operatorname{Im}(A)^* = \operatorname{Im}(A)$.
2. Easy.
3. If $A = H + iK$ then

   $$\operatorname{Re}(A) = \frac{A + A^*}{2} = \frac{H + iK + (H + iK)^*}{2} = \frac{H + iK + H - iK}{2} = H$$

   and similarly $\operatorname{Im}(A) = K$.

$\square$ Remark 142

Definition 143. The *commutator* of X and Y is $[X, Y] = XY - YX$.

Remark 144. $[X, Y] = 0$ if and only if X and Y are commuting.

Definition 145. We say $A \in M_n$ is *normal* if $[A, A^*] = 0$.

Remark 146. Hermitian and unitary matrices are normal.

Proposition 147. *If A is normal an U is unitary, then U^*AU is normal.*

Proof.

$$(U^*AU)^*(U^*AU) = U^*A^*UU^*AU = U^*AA^*U = (U^*AU)(U^*A^*U) = (U^*AU)(U^*AU)^*$$

□ Proposition 147

Theorem 148. *Suppose $A = (a_{ij}) \in M_n$ with $\lambda_1, \dots, \lambda_n$ the roots of $p_A(t)$. Then the following are equivalent:*

1. *A is normal.*
2. *$[\mathrm{Re}(A), \mathrm{Im}(A)] = 0$.*
3. *A is unitarily diagonalizable.*
4. $$\sum_{i,j=1}^{n} |a_{ij}|^2 = \sum_{i=1}^{n} |\lambda_i|^2$$
5. *There is an orthonormal basis for $\mathbb{C}^n$ of eigenvectors of A.*

Proof.

(1) $\iff$ (2) Well,

$$\begin{aligned} A \text{ normal} &\iff A^*A = AA^* \\ &\iff (\mathrm{Re}(A) - i\,\mathrm{Im}(A))(\mathrm{Re}(A) + i\,\mathrm{Im}(A)) = (\mathrm{Re}(A) + i\,\mathrm{Im}(A))(\mathrm{Re}(A) - i\,\mathrm{Im}(A)) \\ &\iff \mathrm{Re}(A)^2 + \mathrm{Im}(A)^2 + i(\mathrm{Re}(A)\,\mathrm{Im}(A) - \mathrm{Im}(A)\,\mathrm{Re}(A)) = \mathrm{Re}(A)^2 + \mathrm{Im}(A)^2 + i[\mathrm{Im}(A)\,\mathrm{Re}(A) - \mathrm{Re}(A)\,\mathrm{Im}(A \\ &\iff \mathrm{Re}(A)\,\mathrm{Im}(A) = \mathrm{Im}(A)\,\mathrm{Re}(A) \end{aligned}$$

(3) $\implies$ (1) Suppose there is unitary U such that $U^*AU = D$ where D is diagonal. Then $A = UDU^*$ and $D^*D = DD^* = \mathrm{diag}(|\lambda_1|^2, \dots, |\lambda_n|^2)$. So, by previous proposition, we have that A is normal.

(1) $\implies$ (3) By Schur there is a unitary U such that $U^*AU = T = (t_{ij})$ is upper triangular. By proposition, we have that T is normal. So $T^*T = TT^*$. But

$$\begin{aligned} (T^*T)_{i,i} &= |t_{i,i}|^2 \\ (TT^*)_{i,i} &= \sum_{j=1}^{n} |t_{i,j}|^2 \end{aligned}$$

So

$$\sum_{j>i} |t_{ij}|^2 = 0$$

and T is diagonal. So $T = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$.

(3) $\implies$ (4) Suppose $U^*AU = D = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$. Then

$$\begin{aligned} \sum_{i,j=1}^{n} |a_{ij}|^2 &= \mathrm{tr}(A^*A) \\ &= \mathrm{tr}((UDU^*)^*(UDU^*)) \\ &= \mathrm{tr}(UDD^*U^*) \\ &= \mathrm{tr}(DD^*) \\ &= \sum_{i=1}^{n} |\lambda_i|^2 \end{aligned}$$

(4) $\implies$ (3) By Schur, we have unitary U such that $U^*AU = T$ is upper triangular with $t_{ii} = \lambda_i$. Then

$$\sum |t_{ij}|^2 = \operatorname{tr}(T^*T) = \operatorname{tr}(A^*A) = \sum |a_{ij}|^2 = \sum |\lambda_i|^2$$

So

$$\sum_{i \neq j} |t_{ij}|^2 = 0$$

So T is diagonal.

(3) $\iff$ (5) If U is unitary, then $Ue_i = u_i$ is an orthonormal basis of eigenvectors.

□ Theorem 148

Corollary 149. *The following are equivalent.*

1. $H = H^*$.
2. *There is unitary U such that $U^*HU = D$ with real diagonal entries.*
3. *There is an orthonormal basis of eigenvectors for H with real eigenvalues.*
4. *H is normal and $\{ \lambda_1, \dots, \lambda_n \} \subseteq \mathbb{R}$.*

6.1 Hermitian matrices

These show up in a *lot* of places.

1. Suppose $D \subseteq \mathbb{R}^n$ is a domain; suppose $f \colon D \to \mathbb{R}$ is C^2. Then

$$\frac{\partial^2}{\partial x_i \partial x_j} f = \frac{\partial^2}{\partial x_j \partial x_i} f$$

Hence the *Hessian*

$$H_f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right) \in M_n$$

has $H_f(x)^* = H_f(x)^t = H_f(x)$.

2. Let G be a graph on n vertices. Consider the adjacency matrix $A_G = (a_{ij})$ where

$$a_{i,j} = \begin{cases} 1 & (i,j) \text{ an edge} \\ 0 & \text{else} \end{cases}$$

Then $A_G^* = A_G^t = A_G$.

We stick to $\mathbb{C}$ instead of $\mathbb{R}$. The key difference: given

$$A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

for $a, b \in \mathbb{R}$, we note that

$$\left\langle A \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle = ax^2 + by^2 + ab(xy - yx)$$

and $A \neq A^t$. On the other hand, in the case of $\mathbb{C}$, we have the following:

Theorem 150. *Suppose $A \in M_n$. Then $A = A^* \iff \langle Av, v \rangle \in \mathbb{R}$ for all $v \in \mathbb{C}^n$.*

Proof.

($\implies$) Well,

$$\begin{aligned}
\overline{\langle Av, v\rangle} &= \overline{v^* Av} \\
&= (v^* Av)^* \\
&= v^* A^* v \\
&= v^* Av \\
&= \langle Av, v\rangle
\end{aligned}$$

So $\langle Av, v\rangle \in \mathbb{R}$.

($\impliedby$) Note that

$$\langle Ae_j, e_i\rangle = a_{i,j}$$

Then

$$a_{ii} = \langle Ae_i, e_i\rangle \in \mathbb{R}$$

Let $z \in \mathbb{C}$. Then

$$\mathbb{R} \ni \langle A(e_k + ze_\ell), (e_k + ze_\ell)\rangle = a_{k,k} + \overline{z}a_{\ell,k} + za_{k,\ell} + |z|^2 a_{\ell,\ell}$$

So $\overline{z}a_{\ell,k} + za_{k,\ell} \in \mathbb{R}$ for all $z \in \mathbb{C}$. Let $a_{\ell,k} = x_1 + iy_1$; let $a_{k,\ell} = x_2 + iy_2$.
Taking $z = 1$ then yields $x_1+x_2+i(y_1+y_2) \in \mathbb{R}$. So $y_2 = -y_1$. Taking $z = i$ yields $-ix_1+y_1+ix_2-y_2 \in \mathbb{R}$, and $x_1 = x_2$. So $a_{k,\ell} = x_1 - iy_1 = \overline{a_{\ell,k}}$. So $A = A^*$.

□ Theorem 150

Remark 151. Suppose $H = H^*$. Then $\lambda_1, \dots, \lambda_n$ are real. We always order $\lambda_1 \leq \cdots \leq \lambda_n$.

Theorem 152 (Rayleigh-Ritz)**.** *Suppose $A = A^* \in M_n$. Then*

1. $\lambda_1 \|x\|_2^2 \leq \langle Ax, x\rangle \leq \lambda_n \|x\|_2^2$ *for all* $x \in \mathbb{C}$.
2. $$\lambda_n = \max_{x \neq 0} \frac{\langle Ax, x\rangle}{\|x\|_2^2} = \max_{\|x\|_2 = 1} \langle Ax, x\rangle$$
3. $$\lambda_1 = \min_{x \neq 0} \frac{\langle Ax, x\rangle}{\|x\|_2^2} = \min_{\|x\|_2 = 1} \langle Ax, x\rangle$$

Proof.

1. We know there is $\{ u_1, \dots, u_n \}$ an orthonormal basis for $\mathbb{C}^n$ such that $Au_i = \lambda_i u_i$. Let

$$x = \sum_{i=1}^n \alpha_i u_i$$

Then

$$\|x\|_2^2 = \sum_{i=1}^n |\alpha_i|^2$$

So

$$\begin{aligned}
\langle Ax, x\rangle &= \left\langle \sum \alpha_i A u_i, \sum \alpha_i u_i \right\rangle \\
&= \left\langle \sum \alpha_i \lambda_i u_i, \sum \alpha_i u_i \right\rangle \\
&= \sum |\alpha_i|^2 \lambda_i
\end{aligned}$$

so $\lambda_1 \|x\|_2^2 \leq \langle Ax, x\rangle \leq \lambda_n \|x\|_2^2$.

2. Similar.

3. Similar.

☐ Theorem 152

Corollary 153. *Suppose $A = A^* \in M_n$; suppose $x \in \mathbb{C}^n$ with $\|x\| = 1$. Let $\alpha = \langle Ax, x\rangle$. Then A has an eigenvalue in $[\alpha, +\infty)$ and in $(-\infty, \alpha]$.*

Proof. By Rayleigh-Ritz, we know $\lambda_1 \leq \alpha \leq \lambda_n$. ☐ Corollary 153

Lemma 154 (Subspace intersection lemma)**.** *Suppose V_1, V_2 are subspaces of W. Then $\dim(V_1 + V_2) + \dim(V_1 \cap V_2) = \dim(V_1) + \dim(V_2)$.*

Proof. Consider $L\colon V_1 \oplus V_2 \to V_1 + V_2$ given by $L(v_1, v_2) = v_1 - v_2$. Then $\mathcal{R}(L) = V_1 + V_2$ and $\mathcal{N}(L) = \{\, (v, v) : v \in V_1 \cap V_2 \,\} \cong V_1 \cap V_2$. By rank-nullity, we have

$$\dim(V_1 + V_2) + \dim(V_1 \cap V_2) = \dim(\mathcal{R}(L)) + \dim(\mathcal{N}(L)) = \dim(V_1 \oplus V_2) = \dim(V_1) + \dim(V_2)$$

☐ Lemma 154

Corollary 155. *If $\dim(V_1) + \dim(V_2) - \dim(V_1 + V_2) \geq 1$, then $V_1 \cap V_2 \neq \{\, 0 \,\}$.*

Theorem 156 (Courant-Fischer)**.** *Let $A = A^* \in M_n$. Let $\lambda_1 \leq \cdots \leq \lambda_n$ be the roots of $p_A(t)$. Suppose $1 \leq k \leq n$; suppose $S \subseteq \mathbb{C}^n$. Then*

1. $$\lambda_k = \min_{\dim(S)=k} \max_{\substack{x \in S \\ \|x\|=1}} \langle Ax, x\rangle$$

2. $$\lambda_k = \max_{\dim(S)=n-k+1} \min_{\substack{x \in S \\ \|x\|=1}} \langle Ax, x\rangle$$

Proof. Let $Au_i = \lambda_i u_i$ where $\{\, u_1, \ldots, u_n \,\}$ is orthonormal.

1. Let $S_0 = \operatorname{span}\{\, u_1, \ldots, u_k \,\}$. Suppose $x \in S_0$ with $\|x\| = 1$. Then

$$x = \sum_{i=1}^{k} \alpha_i u_i$$

so

$$\sum_{i=1}^{k} |\alpha_i|^2 = 1$$

Thus

$$\langle Ax, x\rangle = \left\langle \sum_{i=1}^{k} \alpha_i \lambda_i u_i, \sum_{ij=1}^{k} \alpha_j u_j \right\rangle = \sum_{i=1}^{k} |\alpha_i|^2 \lambda_i \leq \sum_{i=1}^{k} |\alpha_i|^2 \lambda_k = \lambda_k$$

So

$$\lambda_k \leq \min_{\dim(S)=k} \max_{\substack{x \in S \\ \|x\|=1}} \langle Ax, x\rangle$$

For the other direction, suppose $S \subseteq \mathbb{C}^n$ has $\dim(S) = k$. Let $S' = \operatorname{span}\{\, u_k, \ldots, u_n \,\}$. Then $\dim(S') = n - (k-1) = n - k + 1$. So $\dim(S) + \dim(S') = n + 1 > \dim(\mathbb{C}^n) \geq \dim(S + S')$, and

$$\dim(S) + \dim(S') - \dim(S + S') \geq 1$$

So $S \cap S' \neq \{ 0 \}$. Pick $x \in S \cap S'$ with $\|x\| = 1$. Then

$$x = \sum_{j=k}^{n} \alpha_j u_j$$

with

$$\sum_{j=k}^{n} |\alpha_j|^2 = 1$$

So

$$\langle Ax, x \rangle = \sum_{j=k}^{n} |\alpha_j|^2 \lambda_j \geq \lambda_k$$

So

$$\lambda_k \leq \max_{\substack{x \in S \\ \|x\|=1}} \langle Ax, x \rangle$$

for all S with $\dim(S) = k$. So

$$\lambda_k \leq \inf_{\dim(S)=k} \max_{\substack{x \in S \\ \|x\|=1}} \langle Ax, x \rangle$$

But for $S_0 = \operatorname{span}\{ u_1, \ldots, u_k \}$, we can show that

$$\max_{\substack{x \in S \\ \|x\| \leq 1}} \langle Ax, x \rangle = \lambda_k$$

So the infimum is attained at S_0; so we have a minimum.

2. Proof similar; start with $S_0 = \operatorname{span}\{ u_k, \ldots, u_n \}$.

$\square$ Theorem 156

Theorem 157. *Suppose $A = A^* \in M_n$; let $\lambda_1 \leq \cdots \leq \lambda_n$ be the roots of $p_A(t)$. Let $S \subseteq \mathbb{C}^n$ have $\dim(S) = k$. Suppose $c \in \mathbb{R}$ satisfies*

(a) *$c \leq \langle Ax, x \rangle$ for all $x \in S$ with $\|x\| = 1$. Then $c \leq \lambda_{n-k+1}$.*

(a') *$c < \langle Ax, x \rangle$ for all $x \in S$ with $\|x\| = 1$. Then $c < \lambda_{n-k+1}$.*

(b) *$\langle Ax, x \rangle \leq c$ for all $x \in S$ with $\|x\| = 1$. Then $\lambda_k \leq c$.*

(b') *$\langle Ax, x \rangle < c$ for all $x \in S$ with $\|x\| = 1$. Then $\lambda_k < c$.*

Proof. Let $Au_i = \lambda_i u_i$ for $\{ u_1, \ldots, u_n \}$ orthonormal.

(a) Let $S_1 = \operatorname{span}\{ u_1, \ldots, u_{n-k+1} \}$. Then

$$\dim(S) + \dim(S_1) - \dim(S + S_1) \geq k + n - k + 1 - n \geq 1$$

So $S \cap S_1 \neq \{ 0 \}$. Pick $x \in S \cap S_1$ with $\|x\| = 1$. Then

$$x = \sum_{j=1}^{n-k+1} \alpha_j u_j$$

So

$$c \leq \langle Ax, x \rangle = \sum_{j=1}^{n-k+1} |\alpha_j|^2 \lambda_j \leq \lambda_{n-k+1}$$

(a') Identical except for the last line.

(b) Look at $-A$. This has eigenvalues $-\lambda_n \leq \cdots \leq -\lambda_1$. Then since $\langle Ax, x\rangle \leq c$, we have $-c \leq \langle -Ax, x\rangle$. Apply (a) to this: so

$$-c \leq -\lambda_{n-k+1}(-A) = -\lambda_k$$

So $\lambda_k \leq c$.

(b') Similar.

$\square$ Theorem 157

Notation 158. For $A = A^* \in M_n$, we let $\lambda_1(A) \leq \cdots \leq \lambda_n(A)$ be the eigenvalues.

Theorem 159 (Weyl)**.** *Suppose $A = A^* \in M_n$ and $B = B^* \in M_n$. Then*

1. *$\lambda_i(A + B) \leq \lambda_{i+j}(A) + \lambda_{n-j}(B)$ for $0 \leq j \leq n - 1$ and $1 \leq i \leq n$, with equality if and only if there is $x \neq 0$ such that*

$$\begin{aligned} Ax &= \lambda_{i+j}(A)x \\ Bx &= \lambda_{n-j}(B)x \\ (A + B)x &= \lambda_i(A + B)x \end{aligned}$$

2. *$\lambda_{i-j+1}(A) + \lambda_j(B) \leq \lambda_i(A + B)$ for $1 \leq j \leq i$ with equality if and only if there is $x \neq 0$ such that*

$$\begin{aligned} Ax &= \lambda_{i-j+1}(A)x \\ Bx &= \lambda_j(B)x \\ (A + B)x &= \lambda_i(A + B)x \end{aligned}$$

Lemma 160 (Second subspace lemma)**.** *Suppose $S_1, \ldots, S_k \subseteq \mathbb{C}^n$ are subspaces. Let*

$$\delta = \dim(S_1) + \cdots + \dim(S_k) - (k - 1)n$$

Then $\dim(S_1 \cap \cdots \cap S_k) \geq \delta$.

Proof. Let

$$L \colon \bigoplus_{i=1}^{k} S_i \to \bigoplus_{i=1}^{k-1} \mathbb{C}^n$$

be $L(v_1, \ldots, v_k) = (v_1 - v_2, v_2 - v_3, \ldots, v_{k-1} - v_k)$. Then

$$\mathcal{N}(L) = \{\, (v, v, \ldots, v) : v \in S_1 \cap \cdots \cap S_k \,\}$$

Then

$$\dim(\mathcal{R}(L)) + \dim(\mathcal{N}(L)) = \dim\left(\bigoplus_{i=1}^{k} S_i\right) = \sum_{i=1}^{k} \dim(S_i)$$

Hence

$$\dim(\mathcal{N}(L)) \geq \left(\sum_{i=1}^{k} \dim(S_i)\right) - \dim(\mathcal{R}(L)) \geq \sum_{i=1}^{k} -(k - 1)n = \delta$$

$\square$ Lemma 160

Proof of Theorem 159.

1. Let ; let $Bv_i = \lambda_i(B)v_i$.

$$\begin{aligned} Au_i &= \lambda_i(A)u_i \\ Bv_i &= \lambda_i(B)v_i \\ (A+B)z_i &= \lambda_i(A+B)z_i \\ S_1 &= \operatorname{span}\{\, u_1, \ldots, u_{i+j} \,\} \\ S_2 &= \operatorname{span}\{\, v_1, \ldots, v_{n-j} \,\} \\ S_3 &= \operatorname{span}\{\, z_i, \ldots, z_n \,\} \end{aligned}$$

Then

$$\begin{aligned} \dim(S_1) + \dim(S_2) + \dim(S_3) - (3-1)n &= i + j + n - j + n - (i-1) - 2n \\ &= 2n + 1 - 2n \\ &\geq 1 \end{aligned}$$

So $\dim(S_1 \cap S_2 \cap S_3) \geq 1$, and there is $x \in S_1 \cap S_2 \cap S_3$ with $\|x\| = 1$. Write

$$x = \sum_{\ell=i}^{n} \alpha_\ell z_\ell$$

Then $\lambda_i(A+B) \leq \langle (A+B)x, x \rangle$ since

$$(A+B)x = \sum_{\ell=i}^{n} \alpha_\ell \lambda_\ell(A+B) z_\ell$$

and thus

$$\langle (A+B)x, x \rangle = \sum_{\ell=1}^{n} |\alpha_\ell|^2 \lambda_\ell(A+B) \geq \lambda_i(A+B)$$

Now, $x \in S_1$, so we may write

$$x = \sum_{\ell=1}^{i+j} \beta_\ell u_\ell$$

Then

$$\langle Ax, x \rangle = \sum_{\ell=1}^{i+j} |\beta_\ell|^2 \lambda_\ell(A) \leq \lambda_{i+j}(A)$$

Similarly, $x \in S_2$, so we may write

$$x = \sum_{\ell=1}^{n-j} \gamma_\ell v_\ell$$

So

$$\langle Bx, x \rangle \leq \lambda_{n-j}(B)$$

Putting it all together, we find

$$\begin{aligned} \lambda_i(A+B) &\leq \langle (A+B)x, x \rangle \\ &= \langle Ax, x \rangle + \langle Bx, x \rangle \\ &\leq \lambda_{i+j}(A) + \lambda_{n-j}(B) \end{aligned}$$

If equality holds then

$$\lambda_i(A+B) = \langle (A+B)x, x \rangle = \langle Ax, x \rangle + \langle Bx, x \rangle = \lambda_{i+j}(A) + \lambda_{n-j}(B)$$

and thus

$$\begin{aligned}
Ax &= \lambda_i(A)x \\
Bx &= \lambda_{n-j}(B)x \\
(A+B)x &= \lambda_i(A+B)x
\end{aligned}$$

(because $S_1 \cap S_2 \cap S_3 \neq \{0\}$, we have that any $x \in S_1 \cap S_2 \cap S_3$ is simultaneously an eigenvector for A, B, and $A+B$ with the appropriate eigenvalue.)

Conversely, if

$$\begin{aligned}
Ax &= \lambda_i(A)x \\
Bx &= \lambda_{n-j}(B)x \\
(A+B)x &= \lambda_i(A+B)x
\end{aligned}$$

then

$$\begin{aligned}
\lambda_i(A+B) &= \langle (A+B)x, x\rangle \\
&= \langle Ax, x\rangle + \langle Bx, x\rangle \\
&= \lambda_{i+j}(A) + \lambda_{n-j}(B)
\end{aligned}$$

2. Substitute $-A$, $-B$, $-(A+B)$: let

$$\begin{aligned}
\widehat{i} &= n-i+1 \\
\widehat{j} &= j-1
\end{aligned}$$

Then

$$\begin{aligned}
-\lambda_i(A+B) &= \lambda_{n-i+1}(-A-B) \\
&\leq \lambda_{\widehat{i}+\widehat{j}}(-A) + \lambda_{n-\widehat{j}}(-B) \\
&= \lambda_{n-i+j}(-A) + \lambda_{n-j+1}(-B) \\
&= -\lambda_{i-j+1}(A) - \lambda_j(B)
\end{aligned}$$

So

$$\lambda_i(A+B) \geq \lambda_{i-j+1}(A) + \lambda_j(B)$$

□ Theorem 159

Theorem 161 (Cauchy's eigenvalue interlacing theorem)**.** *Suppose $A = A^* \in M_n$; let $\lambda_i = \lambda_i(A)$. Suppose $y \in \mathbb{C}^n$, $a \in \mathbb{R}$. Set*

$$\widehat{A} = \begin{pmatrix} A & y \\ y^* & a \end{pmatrix} = \widehat{A}^* \in M_{n+1}$$

Let $\widehat{\lambda_i} = \lambda_i(\widehat{A})$. Then

$$\widehat{\lambda_1} \leq \lambda_1 \leq \widehat{\lambda_2} \leq \lambda_2 \leq \cdots \leq \widehat{\lambda_n} \leq \lambda_n \leq \widehat{\lambda_{n+1}}$$

Proof. Let $1 \leq k \leq n$. We show $\widehat{\lambda_k} \leq \lambda_k \leq \widehat{\lambda_{k+1}}$. We identify

$$\mathbb{C}^n = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} : x \in \mathbb{C}^n \right\} \subseteq \mathbb{C}^{n+1}$$

Then

$$
\begin{aligned}
\widehat{\lambda_k} &= \min_{\substack{S\subseteq\mathbb{C}^{n+1}\\ \dim(S)=k}} \max_{\substack{\|\widehat{x}\|=1\\ \widehat{x}\in S}} \langle \widehat{A}\widehat{x}, \widehat{x}\rangle \\
&\leq \min_{\substack{S\subseteq\mathbb{C}^{n}\\ \dim(S)=k}} \max_{\substack{\|\widehat{x}\|=1\\ \widehat{x}\in S\\ \widehat{x}=[x|0]^t}} \left\langle \widehat{A}\begin{pmatrix} x\\ 0\end{pmatrix}, \begin{pmatrix} x\\ 0\end{pmatrix}\right\rangle \\
&= \min_{\substack{S\subseteq\mathbb{C}^{n}\\ \dim(S)=k}} \max_{[x|0]^t\in S} \langle Ax, x\rangle \\
&= \lambda_k \\
&= \max_{\substack{S\subseteq\mathbb{C}^{n}\\ \dim(S)=n-k+1}} \min_{\substack{x\in S\\ \|x\|=1}} \langle Ax, x\rangle \\
&= \max_{\substack{S\subseteq\mathbb{C}^{n}\\ \dim(S)=n-k+1}} \min_{\substack{[x|0]^t\in S\\ \|[x|0]\|=1}} \left\langle \widehat{A}\begin{pmatrix} x\\ 0\end{pmatrix}, \begin{pmatrix} x\\ 0\end{pmatrix}\right\rangle \\
&\leq \max_{\substack{S\subseteq\mathbb{C}^{n+1}\\ \dim(S)=(n+1)-(k+1)+1}} \min_{\substack{\widehat{x}\in S\\ \|\widehat{x}\|=1}} \langle \widehat{A}\widehat{x}, \widehat{x}\rangle \\
&= \widehat{\lambda_{k+1}}
\end{aligned}
$$

□ Theorem 161

As a corollary, we get another persistence theorem:

Corollary 162. *Suppose $A = A^* \in M_n$; suppose λ is an eigenvalue of A of geometric multiplicity k. Let*

$$B = \begin{pmatrix} A & C\\ C^* & D\end{pmatrix} = B^* \in M_{n+(k-1)}$$

Then λ is an eigenvalue of B.

Proof.

Case 1. Suppose $k = 2$; say $\lambda = \lambda_i(A) = \lambda_{i+1}(A)$. When we go to $\widehat{A}$ of size $(n+1)\times(n+1)$. Then

$$\widehat{\lambda_i} \leq \lambda_i \leq \widehat{\lambda_{i+1}} \leq \lambda_{i+1}$$

So $\lambda_i = \widehat{\lambda_{i+1}} = \lambda_{i+1} = \lambda$, and λ is an eigenvalue of $\widehat{A}$.

Case 2. Suppose $k = 3$; say $\lambda = \lambda_i = \lambda_{i+1} = \lambda_{i+2}$. For $\widehat{A}$, λ is now an eigenvalue of geometric multiplicity at least 2. So when we go to $\widehat{\widehat{A}}$, we have that λ is still an eigenvalue.

The rest follows by induction. □ Corollary 162

Theorem 163. *Let*

$$\mu_1 \leq \lambda_1 \leq \mu_2 \leq \lambda_2 \leq \cdots \leq \mu_n \leq \lambda_n \leq \mu_{n+1}$$

Then there is $A = A^* \in M_n(\mathbb{R})$, $a \in \mathbb{R}$, and $y_k \geq 0$ for $1 \leq k \leq n$ such that if

$$\widehat{A} = \begin{pmatrix} A & y\\ y^* & a\end{pmatrix}$$

then $\lambda_i(\widehat{A}) = \mu_i$ and $\lambda_i(A) = \lambda_i$.

Proof.

Case 1. Suppose $\lambda_1 < \lambda_2 < \cdots < \lambda_n$. We then let $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. It remains to construct a and the y_k. Well, we need

$$\lambda_1 + \cdots + \lambda_n + a = \operatorname{tr}(\widehat{A}) = \mu_1 + \cdots + \mu_{n+1}$$

We then set $a = \mu_1 + \cdots + \mu_{n+1} - \lambda_1 - \cdots - \lambda_n$. It remains to find the y_k. We think of $\widehat{A}$ as a function of the y_k and compute

$$\begin{aligned}
p_{\widehat{A}}(t) &= \det \begin{pmatrix} tI - A & -y \\ -y^* & t-a \end{pmatrix} \\
&= \det(tI - A)\det((t-a) - (-y)^*(tI-A)^{-1}(-y)) \\
&= (t-\lambda_1)\ldots(t-\lambda_n)\det\left(t - a - (y_1, \ldots, y_n)\begin{pmatrix} (t-\lambda_1)^{-1} & & \\ & \ddots & \\ & & (t-\lambda_n)^{-1} \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}\right) \\
&= (t-\lambda_1)\ldots(t-\lambda_n)\left(t - a - \sum_{k=1}^n y_k^2 (t-\lambda_k)^{-1} \right) \\
&= (t-\lambda_1)\ldots(t-\lambda_n)(t-a) - \sum_{k=1}^n y_k^2 \prod_{j\neq k}(t-\lambda_j)
\end{aligned}$$

(with some conditions on invertibility, but since we're working with polynomials, equality at all but finitely many points is equivalent to equality.) Set $q(t) = (t-\mu_1)\ldots(t-\mu_{n+1})$. We want $p_{\widehat{A}}(t) = q(t)$. Evaluate at $\lambda_1, \ldots, \lambda_n$:

$$\begin{aligned}
p_{\widehat{A}}(\lambda_k) &= -y_k^2 \prod_{j\neq k}(\lambda_k - \lambda_j) \\
q(\lambda_k) &= (\lambda_k - \mu_1)\ldots(\lambda_k-\mu_k)(\lambda_k-\mu_{k+1})\ldots(\lambda_k-\mu_{n+1})
\end{aligned}$$

if

$$y_k^2 = -\frac{q(\lambda_k)}{\prod_{j\neq k}(\lambda_k-\lambda_j)}$$

then $p_{\widehat{A}}(\lambda_k) = q(\lambda_k)$ for $1 \leq k \leq n$.

Claim 164.

$$\frac{q(\lambda_k)}{\prod_{j\neq k}(\lambda_k-\lambda_j)} \leq 0$$

Hence we can pick

$$y_k = \sqrt{-\frac{q(\lambda_k)}{\prod_{j\neq k}(\lambda_k-\lambda_j)}}$$

Proof. Computing signs on the product based on $\lambda_1 < \cdots < \lambda_n$, we find

$$\operatorname{sgn}\left(\prod_{j\neq k}(\lambda_k - \lambda_i)\right) = (-1)^{n-k}$$

Making a similar computation, we find

$$\operatorname{sgn}(q(\lambda_k)) = \operatorname{sgn}\left(\prod_{i=1}^{n+1}(\lambda_k-\mu_i)\right) = (-1)^{n+1-k}$$

Taking the quotient, we find that the claim holds. □ Claim 164

Picking y_k as in the claim, we get $p_{\widehat{A}}(\lambda_k) = q(\lambda_k)$ for all $1 \leq k \leq n$. Recall that p_A and q are both monic of degree $n+1$. Also

$$\begin{aligned} p_{\widehat{A}}(t) &= t^{n+1} - (\lambda_1 + \cdots + \lambda_n + a)t^n + \ldots \\ q(t) &= t^{n+1} - (\mu_1 + \cdots + \mu_{n+1})t^n + \ldots \end{aligned}$$

Hence $p_{\widehat{A}}(t) - q(t)$ has degree $n-1$ and is 0 at n distinct points. (The distinctness is where we use that the λ_i are distinct.) So $p_{\widehat{A}}(t) = q(t)$.

Case 2. We now consider case where the λ_i are not necessarily distinct.

Pick $\lambda_1(m) < \cdots < \lambda_n(m)$ a sequence and $\mu_1(m) \leq \lambda_1(m) \leq \mu_2(m) < \cdots \leq \mu_{n+1}(m)$ such that

$$\begin{aligned} \lim_{m\to\infty} \lambda_i(m) &= \lambda_i \\ \lim_{m\to\infty} \mu_i(m) &= \mu_i \end{aligned}$$

By the previous case, for each m there is

$$\widehat{A}(m) = \begin{pmatrix} \lambda_1(m) & & & \\ & \ddots & & y(m) \\ & & \lambda_n(m) & \\ & y(m)^* & & a_m \end{pmatrix}$$

with

$$a_m = \mu_1(m) + \cdots + \mu_{n+1}(m) - \lambda_1(m) - \cdots - \lambda_n(m)$$

such that $p_{\widehat{A}(m)}(t) = (t - \mu_1(m)) \ldots (t - \mu_{n+1}(m))$. Now, since

$$\operatorname{tr}(\widehat{A}(m)^* \widehat{A}(m)) = \sum_{i=1}^{n+1} \mu_i(m)^2 \leq \sum_{i=1}^{n+1} \mu_i^2 + \varepsilon$$

So the $\widehat{A}(m)$ is bounded; hence we can pick a convergent subsequence with

$$\lim_{j\to\infty} \widehat{A}(m_j) = \widehat{A}$$

Then

$$\widehat{A} = \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & y \\ & & \lambda_n & \\ & y^* & & a \end{pmatrix}$$

Finally, we have

$$\begin{aligned} p_{\widehat{A}}(t) &= \lim_{j\to\infty} \det(tI - A(m_j)) \\ &= \lim_{j\to\infty} (t - \mu_1(m_j)) \ldots (t - \mu_{n+1}) \\ &= (t - \mu_1) \ldots (t - \mu_{n+1}) \end{aligned}$$

□ Theorem 163

Theorem 165. *Let $A = A^* = [a_{ij}] \in M_n$ with $1 \leq m \leq n$. Then*

$$\sum_{i=1}^{m} \lambda_i(A) \leq \sum_{i=1}^{m} a_{ii}$$

Proof. Write

$$A = \begin{pmatrix} B & C \\ C^* & D \end{pmatrix}$$

where $B \in M_m$. Let $A_1 \in M_{n+1}$ be

$$A_1 = \begin{pmatrix} & a_{1,m+1} \\ B & \vdots \\ \dots & a_{m+1,m+1} \end{pmatrix}$$

Let $A_2 \in M_{n+2}$ be the next 2 rows and columns, and so on; then $A_{n-m} = A$. We know

$$\lambda_1(A_1) \le \lambda_1(B) \le \lambda_2(A_1) \le \cdots \le \lambda_m(A_1) \le \lambda_m(B) \le \lambda_{m+1}(A_1)$$

So

$$\lambda_1(A_1) + \cdots + \lambda_m(A_1) \le \lambda_1(B) + \cdots + \lambda_m(B) = \sum_{i=1}^{m} a_{ii}$$

Similarly

$$A_2 = \begin{pmatrix} A_1 & \vdots \\ \dots & a_{m+2,m+2} \end{pmatrix}$$

yields that

$$\lambda_1(A_2) + \cdots + \lambda_m(A_2) \le \lambda_1(A_1) + \cdots + \lambda_m(A_1) \le \sum_{i=1}^{m} a_{ii}$$

By induction this holds for all A_k, and in particular for $A_{n-m} = A$. □ Theorem 165

Corollary 166. *Suppose $A = A^* = [a_{ij}] \in M_n$; suppose $1 \le m \le n$. Then*

$$\lambda_1(A) + \cdots + \lambda_m(A) \le \min_{1 \le i_1 < \cdots < i_m \le n} \sum_{j=1}^{m} a_{i_j,i_j}$$

Proof. Do a permutation P_σ; then $P_\sigma^* A P_\sigma = [b_{i,j}]$ where $b_{i,j} = a_{\sigma(i),\sigma(j)}$. Then

$$\lambda_1(A) + \cdots + \lambda_m(A) = \lambda_1(P_\sigma^* A P_\sigma) + \cdots + \lambda_m(P_\sigma^* A P_\sigma) \le \sum_{i=1}^{m} b_{i,i} = \sum_{i=1}^{m} a_{\sigma(i),\sigma(i)}$$

□ Corollary 166

Corollary 167. *Suppose $A = A^* = [a_{i,j}] \in M_n$; suppose $1 \le m \le n$. Then*

$$\lambda_1(A) + \cdots + \lambda_m(A) = \min_{\{\, u_1,\dots,u_m \,\} \text{ orthonormal}} \sum_{i=1}^{m} \langle Au_i, u_i \rangle$$

Proof. Given any $\{\, u_1, \dots, u_m \,\}$ orthornormal, pick $\{\, u_{m+1}, \dots, u_n \,\}$ such that $\{\, u_1, \dots, u_n \,\}$ is an orthonormal basis for $\mathbb{C}^n$. Let $U = [u_1 \mid \cdots \mid u_n]$ be unitary. Then

$$(U^*AU)_{i,i} = \langle (U^*AU)e_i, e_i \rangle = (e_i^* U^*)AUe_i = u_i^* A u_i = \langle Au_i, u_i \rangle$$

So

$$\lambda_1(A) + \cdots + \lambda_m(A) = \lambda_1(U^*AU) + \cdots + \lambda_m(U^*AU) \le \sum_{i=1}^{m} (U^*AU)_{i,i} = \sum_{i=1}^{m} \langle Au_i, u_i \rangle$$

Thus

$$\lambda_1(A) + \cdots + \lambda_m(A) \le \inf_{\{\, u_1,\dots,u_n \,\} \text{ orthonormal}} \sum_{i=1}^{m} \langle Au_i, u_i \rangle$$

Pick $\{ u_1, \ldots, u_n \}$ an orthonormal basis such that $Au_i = \lambda_i(A)u_i$. For this orthonormal basis , we have

$$\sum_{i=1}^{m} \langle Au_i, u_i \rangle = \sum_{i=1}^{m} \langle \lambda_i(A)u_i, u_i \rangle = \sum_{i=1}^{m} \lambda_i(A)$$

□ Corollary 167

Corollary 168. *Suppose $A = A^* = [a_{i,j}] \in M_n$; suppose $1 \leq k \leq n$. Then*

$$\lambda_n(A) + \lambda_{n-1}(A) + \cdots + \lambda_{n-k+1}(A) = \max_{\{ u_1,\ldots,u_n \} \text{ orthonormal}} \sum_{i=1}^{k} \langle Au_i, u_i \rangle$$

Proof. Well,

$$\begin{aligned}
\lambda_n(A) + \cdots + \lambda_{n-k+1}(A) &= -(\lambda_1(-A) + \cdots + \lambda_k(-A)) \\
&= -\left(\min_{\{ u_1,\ldots,u_k \} \text{ orthonormal}} \sum_{i=1}^{k} \langle -Au_i, u_i \rangle \right) \\
&= -\left(- \max_{\{ u_1,\ldots,u_k \} \text{ orthonormal}} \sum_{i=1}^{k} \langle Au_i, u_i \rangle \right)
\end{aligned}$$

□ Corollary 168

6.2 Majorization

Definition 169. Suppose $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Re-order from smallest to largest:

$$x_1^{\uparrow} \leq \cdots \leq x_n^{\uparrow}$$

Re-order from largest to smallest:

$$x_1^{\downarrow} \geq \cdots \geq x_n^{\downarrow}$$

Example 170. Suppose $x = (1, 2, 1)$. Then

$$\begin{aligned}
x_1^{\uparrow} &= 1 \\
x_2^{\uparrow} &= 1 \\
x_3^{\uparrow} &= 2 \\
x_1^{\downarrow} &= 2 \\
x_2^{\downarrow} &= 1 \\
x_3^{\downarrow} &= 1
\end{aligned}$$

Example 171. Suppose $A = A^*$ and $\lambda_1, \ldots, \lambda_n$ are the roots of $p_A(t)$. Then $\lambda_i(A) = \lambda_i^{\uparrow}$.

Proposition 172. *Suppose $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$ with*

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

Then the following are equivalent:

1. *For all $1 \leq k \leq n$, we have*

$$\max_{1 \leq i_1 < \cdots < i_k \leq n} \sum_{\ell=1}^{k} x_{i_\ell} \geq \max_{1 \leq i_1 < \cdots < i_k \leq n} \sum_{\ell=1}^{k} y_{i_\ell}$$

2. *For all* $1 \le k \le n$, *we have*
$$\sum_{i=1}^{k} x_i^{\downarrow} \ge \sum_{i=1}^{k} y_i^{\downarrow}$$

3. *For all* $1 \le k \le n$, *we have*
$$\sum_{i=1}^{k} x_i^{\uparrow} \le \sum_{i=1}^{k} y_i^{\uparrow}$$

4. *For all* $1 \le k \le n$, *we have*
$$\min_{1 \le i_1 < \cdots < i_k \le n} \sum_{\ell=1}^{k} x_{i_\ell} \le \min_{1 \le i_1 < \cdots < i_k \le n} \sum_{\ell=1}^{k} y_{i_\ell}$$

Proof. Well
$$\max_{1 \le i_1 < \cdots < i_k \le n} \sum_{\ell=1}^{k} x_{i_\ell} = \sum_{\ell=1}^{k} x_\ell^{\downarrow}$$
So (1) $\iff$ (2). Similarly
$$\min_{1 \le i_1 < \cdots < i_k \le n} \sum_{\ell=1}^{k} x_{i_\ell} = \sum_{\ell=1}^{k} x_\ell^{\uparrow}$$
So (3) $\iff$ (4). Let
$$S = \sum_{\ell=1}^{n} x_\ell = \sum_{i=1}^{n} y_\ell$$
Then
$$\begin{aligned} \sum_{i=1}^{k} x_i^{\downarrow} &= S - \sum_{\ell=1}^{n-k+1} x_\ell^{\uparrow} \\ &\ge \sum_{i=1}^{k} y_i^{\downarrow} \\ &= S - \sum_{\ell=1}^{n-k+1} y_i^{\uparrow} \end{aligned}$$
So
$$\sum_{i=1}^{k} x_i^{\downarrow} \ge \sum_{i=1}^{k} y_i^{\downarrow} \iff \sum_{\ell=1}^{n-k+1} x_\ell^{\uparrow} \le \sum_{\ell=1}^{n-k+1} y_\ell^{\uparrow}$$
□ Proposition 172

Definition 173. Given $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n$, we say that x *majorizes* y if
$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$
and any of the 4 equivalent properties occurs.

Theorem 174 (Schur)**.** *Suppose* $A = A^* = [a_{i,j}] \in M_n$. *Let* $\lambda(A) = (\lambda_1, \ldots, \lambda_n)$ *be the roots of* $p_A(t)$; *let* $d(A) = (a_{11}, \ldots, a_{nn})$. *Then* $\lambda(A)$ *majorizes* $d(A)$.

Proof. Note that

$$\lambda_1 + \cdots + \lambda_n = \operatorname{tr}(A) = a_{11} + \cdots + a_{nn}$$

By the last theorem, for all $1 \le m \le k$ we have

$$\begin{aligned}
\lambda_1^{\uparrow} + \cdots + \lambda_m^{\uparrow} &= \lambda_1(A) + \cdots + \lambda_m(A) \\
&\le \min_{1 \le i_1 < \cdots < i_m \le n} \sum_{\ell=1}^{m} a_{i_\ell, i_\ell} \\
&= \sum_{\ell=1}^{m} a_{\ell,\ell}^{\uparrow}
\end{aligned}$$

□ Theorem 174

Hence

$$\sum_{\ell=1}^{k} \lambda_\ell^{\downarrow} \ge \sum_{\ell=1}^{k} a_{\ell,\ell}^{\downarrow}$$

7 Positive semidefinite

Definition 175. We say $A \in M_n$ is *positive semidefinite* (written $A \ge 0$) if for all $x \in \mathbb{C}^n$ we have $\langle Ax, x \rangle \ge 0$. We say A is *positive definite* (written $A > 0$) if for all non-zero $x \in \mathbb{C}^n$ we have $\langle Ax, x \rangle > 0$.

Proposition 176.

1. *$A \ge 0$ if and only if $A = A^*$ and each $\lambda_i(A) \ge 0$.*
2. *$A > 0$ if and only if $A = A^*$ and each $\lambda_i(A) > 0$; this occurs if and only if there is $\delta > 0$ such that $\langle Ax, x \rangle \ge \delta$ whenever $\|x\| = 1$.*

Proof.

1. ($\implies$) Suppose $A \ge 0$. Then $\langle Ax, x \rangle \in \mathbb{R}$ for all x; so $A = A^*$. Fix an eigenvalue $\lambda_i(A)$ with eigenvector u_i of unit length. Then $Au_i = \lambda_i(A)u_i$; so $\lambda_i(A) = \langle Au_i, u_i \rangle \ge 0$.

   ($\impliedby$) We know there is $\{ u_1, \ldots, u_n \}$ an orthonormal basis for $\mathbb{C}^n$ of eigenvectors with $Au_i = \lambda_i(A)u_i$. Suppose

   $$x = \sum_{i=1}^{n} \alpha_i u_i$$

   Then

   $$\begin{aligned}
   \langle Ax, x \rangle &= \left\langle \sum_i \alpha_i \lambda_i(A) u_i, \sum_j \alpha_j u_j \right\rangle \\
   &= \sum_{i,j} \lambda_i(A) \alpha_i \overline{\alpha_J} \\
   &= \sum_i \lambda_i(A) |\alpha_i|^2 \ge 0
   \end{aligned}$$

2. Suppose $A > 0$. Then $A \ge 0$, so $A = A^*$ and each $\lambda_i(A) \ge 0$. But if some $\lambda_i(A) = 0$, then

   $$\langle Au_i, u_i \rangle = \lambda_i(A) \langle u_i, u_i \rangle = 0$$

   a contradiction. So each $\lambda_i(A) > 0$.

   Suppose $A = A^*$ and each $\lambda_i(A) > 0$. Then, as above, we take

   $$x = \sum_i \alpha_i u_i$$

with $\|x\| = 1$; then

$$\sum_i |\alpha_i|^2 = 1$$

Then

$$\begin{aligned}\langle Ax, x\rangle &= \sum_i \lambda_i(A)|\alpha_i|^2 \\ &\geq \lambda_1(A)\sum_i |\alpha_i|^2 \\ &= \lambda_1(A) \\ &> 0\end{aligned}$$

In particular, the minimum value of $\langle Ax, x\rangle$ is $\delta = \lambda_1(A)$.

$\square$ Proposition 176

Lemma 177. *Suppose $A = A^* \in M_n^{-1}$. Then $A^{-1} = (A^{-1})^*$ and the roots of $p_{A^{-1}}(t)$ are $\lambda_1(A)^{-1}, \ldots, \lambda_n(A)^{-1}$.*

Proof. Note that

$$I^* = (A^{-1}A)^* = A^*(A^{-1})^* = A(A^{-1})^*$$

So $(A^{-1})^* = A^{-1}$.

If $Au_i = \lambda_i(A)u_i$, then $A^{-1}u_i = \lambda_i(A)^{-1}u_i$; thus if $\{ u_1, \ldots, u_n \}$ is an orthonormal basis of eigenvectors of A, then it is also an orthonormal basis of eigenvectors of A^{-1} with eigenvalues $\lambda_i(A)^{-1}$. $\square$ Lemma 177

Proposition 178.

1. *If $A \geq 0$, then $\overline{A} = A^t \geq 0$.*
2. *If $A > 0$, then $\overline{A} = A^t > 0$ and $A^{-1} > 0$.*
3. *If $A \geq 0$ and $S \subseteq \{ 1, \ldots, n \}$, then $A[S, S] \geq 0$.*
4. *If $A > 0$ and $S \subseteq \{ 1, \ldots, n \}$, then $A[S, S] > 0$.*

Proof.

1. Note that

$$\begin{aligned}\langle \overline{A}x, x\rangle &= \sum_{i,j=1}^n \overline{a_{ij}} x_j \overline{x_i} \\ &= \overline{\sum_{i,j=1}^n a_{ij}\overline{x_j} x_i} \\ &= \overline{\langle A\overline{x}, \overline{x}\rangle} \\ &\in \mathbb{R}\end{aligned}$$

So $(\overline{A})^* = \overline{A}$. If $\{ u_1, \ldots, u_n \}$ is an orthonormal basis of eigenvectors for A with $Au_i = \lambda_i(A)u_i$ and

$x = (x_1, \ldots, x_n)$, then

$$
\begin{aligned}
Ax &= \begin{pmatrix} \sum_{j=1}^n a_{ij}x_j \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j \end{pmatrix} \\
&= \lambda \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\
\implies \overline{A}\overline{x} &= \begin{pmatrix} \sum_{j=1}^n \overline{a_{1j}x_j} \\ \vdots \\ \sum_{j=1}^n \overline{a_{nj}x_j} \end{pmatrix} \\
&= \lambda \begin{pmatrix} \overline{x_1} \\ \vdots \\ \overline{x_n} \end{pmatrix}
\end{aligned}
$$

So if $\{ u_1, \ldots, u_n \}$ are eigenvectors with real eigenvalues $\lambda_1, \ldots, \lambda_n$, then $\{ \overline{u_1}, \ldots, \overline{u_n} \}$ is a set of eigenvectors for $\overline{A}$ with the same eigenvalues. Thus $\lambda_i(\overline{A}) = \lambda_i(A) \geq 0$, and $\overline{A} \geq 0$.

2. Same: we get $\lambda_i(\overline{A}) = \lambda_i(A) > 0$ and $A^{-1} = (A^{-1})^*$.

3. Without loss of generality, we have $S = \{ 1, \ldots, k \}$. So

$$A = \begin{pmatrix} A[S,S] & B \\ B^* & C \end{pmatrix} \geq 0$$

Let $x \in \mathbb{C}^k$; let

$$\widehat{x} = \begin{pmatrix} x \\ 0 \end{pmatrix}$$

Then

$$0 \leq \langle A\widehat{x}, \widehat{x} \rangle = \langle A[S,S]x, x \rangle$$

So $A[S,S] \geq 0$.

4. As above, noting that $x \neq 0$ implies that $\widehat{x} \neq 0$; thus

$$0 < \langle A\widehat{x}, \widehat{x} \rangle = \langle A[S,S]x, x \rangle$$

and $A[S,S] > 0$.

$\square$ Proposition 178

Proposition 179. *Suppose $A \geq 0$ ($A > 0$). Then $A^k \geq 0$ ($A^k > 0$) for all $k \in \mathbb{N}$.*

Proof. Note that $(A^k)^* = A^k$, and that the eigenvaluse are $\lambda_i(A)^k \geq 0$ (> 0). $\square$ Proposition 179

Set $A_i = A[\{ 1, \ldots, i \}, \{ 1, \ldots, i \}]$.

Theorem 180. *Suppose $A = A^* \in M_n$. Then $A > 0$ if and only if $\det(A_i) > 0$ for all $1 \leq i \leq n$.*

Example 181. Consider

$$A = A^* = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 0 & 0 \\ 2 & 0 & 1 \end{pmatrix}$$

Then

$$\det(A_1) = 1$$
$$\det(A_2) = 0$$
$$\det(A_3) = 0$$

But

$$\left\langle A\begin{pmatrix}1\\0\\-1\end{pmatrix}, \begin{pmatrix}1\\0\\-1\end{pmatrix}\right\rangle = -2$$

So $A \not\geq 0$.

Proof of Theorem 180.

$(\implies)$ Suppose $A > 0$. Then all eigenvalues are positive; so $\det(A_i) > 0$.

$(\impliedby)$ Note that $\det(A_1) = a_{11} > 0$. The

$$A_2 = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

We now use the eigenvalue interlacing theorem:

$$\lambda_1(A_2) \leq \lambda_1(A_1) a_{11} \leq \lambda_2(A_2)$$

So $\lambda_2(A_2) > 0$. Bu $\operatorname{tdet}(A_2) = \lambda_1(A_1)\lambda_2(A_2) > 0$. So $\lambda_1(A_2) > 0$.

Assume $0 < \lambda_1(A_k) < \cdots < \lambda_k(A_k)$. Then, by eigenvalue interlacing, we have

$$\lambda_1(A_{k+1}) \leq \lambda_1(A_k) \leq \lambda_2(A_{k+1})$$
$$\vdots$$
$$\lambda_k(A_{k+1}) \leq \lambda_k(A_k) \leq \lambda_{k+1}(A_{k+1})$$

So $\lambda_2(A_{k+1}), \ldots, \lambda_{k+1}(A_{k+1}) > 0$. But $0 < \det(A_{k+1}) = \lambda_i(A_{k+1}) \ldots \lambda_{k+1}(A_{k+1})$. So $\lambda_1(A_{k+1}) > 0$.

☐ Theorem 180

Aside 182. Suppose $D \subseteq \mathbb{R}^n$ and $f\colon D \to \mathbb{R}$ is C^2. Suppose $x_0 \in D$ with

$$f'(x_0) = \left(\frac{\partial f}{\partial x_1}(x_0), \ldots, \frac{\partial f}{\partial x_n}(x_0)\right) = 0$$

We set

$$H_f(x_0) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)\right)$$

If $u \in \mathbb{R}^n$ is a unit vector and $g_u(t) = f(x_0 + tu)$, then $g_u''(0) = \langle H_f(x_0)u, u\rangle$.

Theorem 183. *Suppose $f \in C^2(D)$ and $f'(x_0) = 0$. If $H_f(x_0) > 0$, then x_0 is a local minimum. If $-H_f(x_0) > 0$, then x_0 is a local maximum.*

Sketch. If $H_f(x_0) > 0$, then $g_u''(0) > 0$ for all u; roughly speaking, we then have that x_0 is a local minimum in all directions, we have that x_0 is a local minimum. ☐ Theorem 183

Since

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

we have that $H_f(x_0) = H_f(x_0)^*$.

Corollary 184. *If $\det(H_f(x_0)_i) > 0$ for all $1 \leq i \leq n$, then x_0 is a local minimum. If $(-1)^i \det(H_f(x_0)_i) > 0$ for all $1 \leq i \leq n$, then x_0 is a local maximum.*

Proposition 185. *Suppose $A \in M_n$, $C \in M_{n,m}$. If $A \geq 0$ then $C^*AC \geq 0$. In particular, we have $C^*C \geq 0$.*

Proof. Note that

$$\begin{aligned}
\langle C^*ACx, x\rangle &= (x^*C^*)A(Cx) \\
&= (Cx)^*A(Cx) \\
&= \langle A(Cx), (Cx)\rangle \\
&\geq 0
\end{aligned}$$

□ Proposition 185

Proposition 186. *Suppose $A \geq 0$ and $k \in \mathbb{N}$. Then there is $B \geq 0$ such that $A = B^k$.*

Proof. Write $A = U^*DU$ where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ where $\lambda_i = \lambda_i(A)$. Let

$$\widetilde{D} = \operatorname{diag}\left(\lambda_1^{\frac{1}{k}}, \ldots, \lambda_n^{\frac{1}{k}}\right)$$

Let $B = U^*\widetilde{D}U$. Then $B^k = U^*\widetilde{D}^kU = U^*DU = A$. □ Proposition 186

We write $B = A^{\frac{1}{k}}$.

Proposition 187. *Suppose $A \geq 0$. Then $Ax = 0$ if and only if $\langle Ax, x\rangle = 0$.*

Proof.

($\Longrightarrow$) Clear.

($\Longleftarrow$) Write $A = B^2$ where $B \geq 0$.

$$\begin{aligned}
0 &= \langle Ax, x\rangle \\
&= x^*B^2x \\
&= (Bx)^*(Bx) \\
&= \|Bx\|^2
\end{aligned}$$

so $Bx = 0$ and $Ax = B(Bx) = 0$.

□ Proposition 187

7.1 Factorization and decomposition

We already saw that if $A \geq 0$, there is $B \geq 0$ such that $A = B^2$; then $B^* = B$, and $A = B^2B$. This is one factorization.

Proposition 188. *Let $A \geq 0$; then $A = R^*R$ with R upper triangular.*

Proof. Write $A = B^*B$ as above. Apply the QR theorem to write $B = QR$ where Q is unitary and R is upper triangular. Then

$$A = B^*B = (QR)^*(QR) = R^*Q^*QR = R^*R$$

□ Proposition 188

7.1.1 Cholesky factorization

Lemma 189 (Cholesky's lemma)**.** *Let*

$$P = \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} = P^*$$

with $A > 0$. Then the following are equivalent:

1. $P \geq 0$
2. $$P - \begin{pmatrix} A^{\frac{1}{2}} \\ B^* A^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} A^{\frac{1}{2}} & A^{-\frac{1}{2}} B \end{pmatrix} \geq 0$$
3. $C - B^* A^{-1} B \geq 0$

Proof.

(2) $\iff$ **(3)** Note that

$$\begin{aligned} P - \begin{pmatrix} A^{\frac{1}{2}} \\ B^* A^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} A^{\frac{1}{2}} & A^{-\frac{1}{2}} B \end{pmatrix} &= P - \begin{pmatrix} A & B \\ B^* & B^* A^{-1} B \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & C - B^* A^{-1} B \end{pmatrix} \end{aligned}$$

So

$$P - \begin{pmatrix} A^{\frac{1}{2}} \\ B^* A^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} A^{\frac{1}{2}} & A^{-\frac{1}{2}} B \end{pmatrix} \geq 0$$

if and only if $C - B^* A^{-1} B \geq 0$.

(1) $\implies$ **(3)** Let $X = -A^{-1}B$. Then since $P \geq 0$, we have

$$\begin{aligned} 0 &\leq \begin{pmatrix} I & 0 \\ X^* & I \end{pmatrix} \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \\ &= \begin{pmatrix} I & 0 \\ -B^* A^{-1} & I \end{pmatrix} \begin{pmatrix} A & AX + B \\ B^* & B^* X + C \end{pmatrix} \\ &= \begin{pmatrix} I & 0 \\ -B^* A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ B^* & C - B^* A^{-1} B \end{pmatrix} \\ &= \begin{pmatrix} A & 0 \\ 0 & C - B^* A^{-1} B \end{pmatrix} \end{aligned}$$

So $C - B^* A^{-1} B \geq 0$.

(3) $\implies$ **(1)** Suppose $C - B^* A^{-1} B \geq 0$. Then

$$\begin{pmatrix} A & 0 \\ 0 & C - B^* A^{-1} B \end{pmatrix} \geq 0$$

So

$$0 \leq \begin{pmatrix} I & 0 \\ -X^* & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & C - B^* A^{-1} B \end{pmatrix} \begin{pmatrix} I & -X \\ 0 & I \end{pmatrix} = P$$

□ Lemma 189

Lemma 190. *Suppose $P = P^* = (p_{ij}) \in M_n$ satisfies $P \geq 0$. If $p_{ii} = 0$ then $p_{ij} = p_{ji} = 0$ for all j.*

Proof. Let $S = \{ i, j \}$. Then since $P \geq 0$ we have $P[S, S] \geq 0$. But

$$P[S,S] = \begin{pmatrix} p_{ii} & p_{ij} \\ p_{ji} & p_{jj} \end{pmatrix} = \begin{pmatrix} 0 & p_{ij} \\ p_{ji} & p_{jj} \end{pmatrix}$$

so $0 \leq \det(P[S,S]) = -|p_{ij}|^2$; so $p_{ij} = 0$. □ Lemma 190

This yields *Cholesky's algorithm*, a fast way to tell if a Hermitian matrix P is positive semidefinite and, if it is, find upper triangular T such that $P = T^*T$.

Example 191. Let

$$P = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 8 & 7 \\ 3 & 7 & 11 \end{pmatrix}$$

Decompose P as in Cholesky's lemma

$$\begin{aligned} A &= (1) \\ B &= \begin{pmatrix} 2 & 3 \end{pmatrix} \\ C &= \begin{pmatrix} 8 & 7 \\ 7 & 11 \end{pmatrix} \end{aligned}$$

If $P \geq 0$, then

$$\begin{aligned} P - \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} &= P - \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 1 & 2 \end{pmatrix} \end{aligned}$$

Then $P \geq 0$ if and only if

$$\begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} \geq 0$$

Again, using Cholesky, this holds if and only if

$$\begin{aligned} 0 &\leq \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} - \begin{pmatrix} 4 \\ 1 \end{pmatrix} 4^{-1} \begin{pmatrix} 4 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix} - \begin{pmatrix} 4 & 1 \\ 1 & \frac{1}{4} \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & \frac{7}{4} \end{pmatrix} \end{aligned}$$

But $\frac{7}{4} = \left(\frac{\sqrt{7}}{2}\right)^2$; so $P \geq 0$. In general, we get that $P \geq 0$ unless at some step the "new" $(1,1)$-entry either is negative or is 0 and some entries in that row and column are non-zero.

Finally, to get T, save the scaled row vectors we subtract by (where we distribute the square root of middle scalar to both sides), and get

$$T = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & \frac{1}{2} \\ 0 & 0 & \frac{\sqrt{7}}{2} \end{pmatrix}$$

and $T^*T = P$. To see that this works in general, recall that if

$$W = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}$$

Then $W^*W = r_1^* r_1 + \cdots + r_n^* r_n$. So

$$T^*T = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \\ \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 & 2 & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \frac{\sqrt{7}}{2} \end{pmatrix} \begin{pmatrix} 0 & 0 & \frac{\sqrt{7}}{2} \end{pmatrix}$$

Remark 192. We saw that if $A \in M_n$ has $A \geq 0$ then $A = R^*R$ for some R. Let $C = R^*$; then $A = CC^*$. Let $C = [C_1 \mid \cdots \mid C_n]$; then $A = CC^* = C_1C_1^* + \cdots + C_nC_n^*$, and we can write A as a sum of positive, rank 1 matrices. The moral is that a factorization $A = R^*R$ corresponds to a decomposition of A into a sum of poistive, rank 1 matrices.

7.2 Subspaces, orthogonal complements, and projections

Definition 193. Suppose $V \subseteq \mathbb{C}^n$ is a subspace. We set $V^\perp = \{ w \in \mathbb{C}^n : \langle w, v \rangle = 0 \text{ for all } v \in V \}$.

Proposition 194. *$V^\perp$ is a subspace and $V \cap V^\perp = \{ 0 \}$.*

Proof. If $w_1, w_2 \in V^\perp$ and $\lambda \in \mathbb{C}$, then $\langle \lambda w_1 + w_2, v \rangle = \lambda \langle w_1, v \rangle + \langle w_2, v \rangle = 0$ for all $v \in V$; so $\lambda w_1 + w_2 \in V^\perp$, and $V^\perp$ is a subspace. Furthermore, if $w \in V \cap V^\perp$, then $\|w\|^2 = \langle w, w \rangle = 0$, and $w = 0$. ☐ Proposition 194

Theorem 195. *Let $V \subseteq \mathbb{C}^n$ be a subspace with $\dim(V) = d$. Let $\{ v_1, \ldots, v_d \}$ be an orthonormal basis for V. Set*

$$P = \sum_{i=1}^{d} v_i v_i^*$$

Then

1. *$P = P^2 = P^*$.*
2. *$\mathcal{R}(P) = V$.*
3. *$Pv = v$ for all $v \in V$.*
4. *$Pw = 0$ for all $w \in V^\perp$.*
5. *$(I - P)^2 = (I - P)^* = (I - P)$.*
6. *$\mathcal{R}(I - P) = V^\perp$.*
7. *$V + V^\perp = \mathbb{C}^n$ and if $w = v_1 + v_2 = v_1' + v_2'$ where $v_1, v_1' \in V$ and $v_2, v_2' \in V^\perp$, then $v_1 = v_1'$ and $v_2 = v_2'$.*
8. *If $\{ w_1, \ldots, w_d \}$ is another orthonormal basis for V, then*

$$\sum_{i=1}^{d} w_i w_i^* P$$

Proof.

1. Note that

$$P^* = \sum_{i=1}^{d} (v_i v_i^*)^* = \sum_{i=1}^{d} v_i^{**} v_i^* = P$$

and

$$P^2 = \sum_{i,j=1}^{d} v_i v_i^* v_j v_j^* = \sum_{i,j=1}^{d} \langle v_j, v_i \rangle v_i v_j^* = \sum_{i=1}^{d} v_i v_i^* = P$$

2. Note that
$$Pw = \sum_{i=1}^{d} v_i v_i^* w = \sum_{i=1}^{d} \langle w, v_i \rangle v_i \in \operatorname{span}\{ v_1, \dots, v_d \}$$
So $\mathcal{R}(P) \subseteq V$. But
$$Pv_j = \sum_{i=1}^{d} v_i v_i^* v_j = v_j$$
So $v_j \in \mathcal{R}(P)$, and $V = \operatorname{span}\{ v_1, \dots, v_d \} \subseteq \mathcal{R}(P)$. So $\mathcal{R}(P) = V$.

3.

4. If $w \in V^\perp$ then
$$Pw = \sum_{i=1}^{d} v_i v_i^* w = 0$$

5. $(I-P)^* = I^* - P^* = I - P$, and $(I-P)^2 = I - P - P + P^2 = I - P$.

6. If $w \in V^\perp$, then $(I-P)w = w - Pw = w$; so if $w \in V^\perp$ then $w \in \mathcal{R}(I-P)$. Conversely, if $w \in \mathcal{R}(I-P)$, say $w = (I-P)z$, then
$$w = z - Pz = z - \sum_{i=1}^{d} \langle z, v_i \rangle v_i$$
so
$$\langle w, v_j \rangle = \langle z, v_j \rangle - \sum_{i=1}^{d} \langle z, v_i \rangle \langle v_i, v_j \rangle = \langle z, v_j \rangle - \langle z, v_j \rangle = 0$$
So $w \in V^\perp$. So $\mathcal{R}(I-P) \subseteq V^\perp$. So $\mathcal{R}(I-P) = V^\perp$.

7. Note that
$$w = (P + (I-P))w = Pw + (I-P)w \in V + V^\perp$$
Suppose now that $w = v_1 + v_2 = v_1' + v_2'$ for $v_1, v_1' \in V$ and $v_2, v_2' \in V^\perp$. Then $V \ni v_1 - v' = v_2' - v_2 \in V^\perp$. So $v_1 - v_1' = v_2' - v_2 = 0$, as $V \cap V^\perp = \{ 0 \}$.

8. Set
$$Q = \sum_{i=1}^{d} w_i w_i^*$$
Then (1) through (6) hold for Q as well. Thus if $w = v_1 + v_2$ for $v_1 \in V$ and $v_2 \in V^\perp$, we have $Pw = v_1$ and $Qw = v_1$; so $P = Q$.

□ Theorem 195

Definition 196. The matrix P is called the *orthogonal projection* onto V.

7.3 Gram matrices

Definition 197. Given $w_1, \dots, w_n \in \mathbb{C}^k$, we define the *Gram matrix* or *Grammian* of the vectors is the $n \times n$ matrix $G = (g_{i,j})$ with $g_{i,j} = \langle w_j, w_i \rangle = w_i^* w_j$.

Theorem 198. *Suppose $w_1, \dots, w_n \in \mathbb{C}^k$; let $W = [w_1 \mid \cdots \mid w_n] \in M_{k,n}$. Then*

1. *$G = W^*W$, and hence $G \geq 0$.*

2. *$G > 0$ if and only if $\{ w_1, \dots, w_n \}$ are linearly independent.*

3. $\operatorname{rank}(G) = \operatorname{rank}(W) = \dim(\operatorname{span}\{ w_1, \dots, w_n \})$.

Proof.

1. Clear.
2. Let
$$x = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$
Then
$$\begin{aligned} \langle Gx, x \rangle &= \langle W^*Wx, x \rangle \\ &= x^*W^*Wx \\ &= (Wx)^*(Wx) \\ &= \|Wx\|^2 \\ &= \left\| \sum_{i=1}^n \lambda_i w_i \right\|^2 \end{aligned}$$
So $\langle Gx, x \rangle > 0$ if and only if
$$\sum_{i=1}^n \lambda_i w_i \neq 0$$
3. Suppose $x \in \mathcal{N}(G)$. Then $Gx = 0$, so $\langle Gx, x \rangle = 0$, and
$$\left\| \sum_{i=1}^n \lambda_i w_i \right\| = 0$$
So
$$Wx = \sum_{i=1}^n \lambda_i w_i = 0$$
So $\mathcal{N}(G) \subseteq \mathcal{N}(W)$.
Conversely, if $x \in \mathcal{N}(W)$, then $Gx = W^*Wx = W^*0 = 0$.
So $\mathcal{N}(G) = \mathcal{N}(W)$. Then
$$\operatorname{rank}(G) = n - \dim(\mathcal{N}(G)) = n - \dim(\mathcal{N}(W)) = \operatorname{rank}(W) = \dim(\operatorname{span}\{ w_1, \ldots, w_n \})$$

☐ Theorem 198

7.4 Polar form and singular valued decomposition

Definition 199. Suppose $A \in M_{m,n}$. We set $|A| = (A^*A)^{\frac{1}{2}}$; this is called the *absolute value of* A. The *singular values of* A are $S_i(A) = \lambda_i^{\downarrow}(|A|)$ for $1 \leq i \leq n$.

Lemma 200. *Suppose* $A \in M_{m,n}$*; suppose* $x \in \mathbb{C}^n$*. Then*

1. $\left\| |A|x \right\|_2 = \|Ax\|_2$.
2. $Ax = 0$ *if and only if* $|A|x = 0$.

Proof.

1. Note that

$$\begin{aligned}\Big\||A|x\Big\|^2 &= (|Ax|)^*(|A|x)\\ &= x^*|A|^*|A|x\\ &= x^*A^*Ax\\ &= \|Ax\|^2\end{aligned}$$

2. Follows easily from (1).

□ Lemma 200

Theorem 201 (Polar decomposition I)**.** *Suppose $A \in M_{m,n}$. Then there is a unique isometry $W : \mathcal{R}(|A|) \to \mathcal{R}(A)$ such that $A = W|A|$.*

Proof. Note that $Ax = W|A|x$ if and only if $W(|A|x) = Ax$. We check that this is well-defined. By Lemma 114, there is linear W with $W|A|x = Ax$ for all x if and only if whenever we have

$$\sum \lambda_i(|A|x_i) = 0$$

we also have

$$\sum \lambda_i(Ax_i) = 0$$

But

$$\sum \lambda_i|A|x_i = |A|\Big(\sum \lambda_i x_i\Big) = 0$$

so by the lemma we have that

$$\sum \lambda_i(Ax_i) = A\Big(\sum \lambda_i x_i\Big) = 0$$

So W exists. But then

$$\Big\|W(|A|x)\Big\| = \|Ax\| = \Big\||A|x\Big\|$$

So W is an isometry. To check uniqueness, suppose we had V also satisfying the desired properties. Then

$$V(|A|x) = Ax = W(|A|x)$$

for all $v \in \mathcal{R}(|A|)$. So $V = W$. □ Theorem 201

Aside 202 (Vandermonde matrices)*.* Suppose $\lambda_1, \ldots, \lambda_k$ are distinct. Set

$$V = \begin{pmatrix} 1 & \ldots & 1\\ \lambda_1 & \ldots & \lambda_k\\ \lambda_1^2 & \ldots & \lambda_k^2\\ \vdots & \ddots & \vdots\\ \lambda_1^{k-1} & \ldots & \lambda_k^{k-1}\end{pmatrix}$$

Claim 203*. V is invertible.*

Proof. Note that $\sigma(V) = \sigma(V^t)$. It suffices to show that $\mathcal{N}(V^t) = \{0\}$. But

$$V^t\begin{pmatrix} p_0\\ \vdots\\ p_{k-1}\end{pmatrix} = \begin{pmatrix} p_0 + p_1\lambda_1 + \cdots + p_{k-1}\lambda_1^{k-1}\\ \vdots\\ p_0 + p_1\lambda_k + \cdots + p_{k-1}\lambda_k^{k-1}\end{pmatrix} = \begin{pmatrix} p(\lambda_1)\\ \vdots\\ p(\lambda_k)\end{pmatrix}$$

So if all the entries are 0, we have $p(t)$ is degree k with k distinct zeroes; so $p_i = 0$ for all i. □ Claim 203

Problem: given $Av_i = \lambda_i v_i$ with $v_i \neq 0$ and $\lambda_1, \ldots, \lambda_k$ distinct, show the v_i are linearly independent. Suppose now that $a_1 v_1 + \cdots + a_k v_k = 0$. Applying A^i, we get $a_1 \lambda_1^i v_1 + \cdots + a_k \lambda_k^i v_k = 0$. Take any $w \in \mathbb{C}^n$; then

$$\begin{aligned} w^*(a_1 v_1) + \cdots + w^*(a_k v_k) &= 0 \\ &\vdots \\ \lambda_1^{k-1} w^*(a_1 v_1) + \cdots + \lambda_k^{k-1} w^*(a_k v_k) &= 0 \end{aligned}$$

So

$$\begin{pmatrix} 1 & \cdots & 1 \\ \lambda_1 & \cdots & \lambda_k \\ \vdots & \cdots & \vdots \\ \lambda_1^{k-1} & \cdots & \lambda_k^{k-1} \end{pmatrix} \begin{pmatrix} w^*(a_1 v_1) \\ \vdots \\ w^*(a_k v_k) \end{pmatrix} = 0$$

So $w^*(a_j v_j) = 0$ for all j and w. So $a_j v_j = 0$ for all j, and all $a_j = 0$.

Recall that $\|Ax\| = \left\| |A| x \right\|$, so $Ax = 0$ if and only if $|A|x = 0$; i.e. $\mathcal{N}(A) = \mathcal{N}(|A|)$. We showed there is a unique isometry $W : \mathcal{R}(|A|) \to \mathcal{R}(A)$ such that $A = W|A|$. (This is the polar decomposition, analogous to $z = \exp(i\theta)|z|$.)

Theorem 204 (Polar Decomposition II)**.** *Suppose $A \in M_n$. Then there is a unitary U such that $A = U|A|$.*

Proof. We know there is a unique isometry $W : \mathcal{R}(|A|) \to \mathcal{R}(A)$ such that $A = W|A|$; note that both the domain and the codomain are subset of $\mathbb{C}^n$. So $\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(|A|))$. We also saw that for any subspace $V \subseteq \mathbb{C}^n$, we have $V + V^\perp = \mathbb{C}^n$ and $V \cap V^\perp = \{ 0 \}$. So $\dim(V^\perp) = n - \dim(V)$, and in particular we have $\dim(\mathcal{R}(A)^\perp) = \dim(\mathcal{R}(|A|)^\perp)$. Let $\dim(\mathcal{R}(A)^\perp) = d$. Pick $\{ z_1, \ldots, z_d \}$ an orthonormal basis for $\mathcal{R}(|A|)^\perp$; pick $\{ \widetilde{z}_1, \ldots, \widetilde{z}_d \}$ an orthonormal basis for $\mathcal{R}(A)^\perp$. Every vector in $\mathbb{C}^n$ has a unique decomposition $|A|x + z$ where $|A|x \in \mathcal{R}(|A|)$ and $z \in \mathcal{R}(|A|)^\perp$. We may then find $\alpha_1, \ldots, \alpha_d$ such that $z = \alpha_1 z_1 + \cdots + \alpha_d z_d$. Define $U : \mathbb{C}^n \to \mathbb{C}^n$ by $U(|A|x + z) = Ax + \alpha_1 \widetilde{z}_1 + \cdots + \alpha_d \widetilde{z}_d$. Then

$$\begin{aligned} \|U(|A|x + z)\|^2 &= \|Ax + \alpha_1 \widetilde{z}_1 + \cdots + \alpha_d \widetilde{z}_d\|^2 \\ &= \|Ax\|^2 + |\alpha_1|^2 + \cdots + |\alpha_d|^2 \\ &= \left\| |A|x \right\|^2 + \|\alpha_1 z_1 + \cdots + \alpha_d z_d\|^2 \\ &= \left\| |A|x + \alpha_1 z_1 + \cdots + \alpha_d z_d \right\|^2 \end{aligned}$$

So $U : \mathbb{C}^n \to \mathbb{C}^n$ is an isometry; so U is a unitary. But $U|A|x = U(|Ax|) = Ax$; so $U|A| = A$. So U is our desired unitary. ☐ Theorem 204

Corollary 205. *Suppose $A \in M_n$; then there is unitary V such that $A = |A^*|V$.*

Proof. Note that $A^* = U|A^*|$; so $A = |A^*|U^*$. We then set $V = U^*$. ☐ Corollary 205

Corollary 206 (Singular value decomposition I)**.** *Suppose $A \in M_n$; let $S = \operatorname{diag}(S_1(A), \ldots, S_n(A))$. Then there are isometries U, V such that $A = USV$.*

Proof. Write $A = U_1|A|$ as above. Now, $|A| \geq 0$, so there is unitary V such that $|A| = V^* S V$ for some unitary V. Then $A = (U_1 V^*) SV$, and we have our $U = U_1 V^*$. ☐ Corollary 206

Corollary 207 (Singular value decomposition II)**.** *Suppose $A \in M_n$. Then there is $\{ u_1, \ldots, u_n \}$ and $\{ v_1, \ldots, v_n \}$ orthonormal such that*

$$A = \sum_{k=1}^{n} S_k(A) u_k v_k^*$$

Proof. Write $A = USV$ as above. Write $U = [u_1 \mid \cdots \mid u_n]$; then $\{ u_1, \dots, u_n \}$ is orthonormal. Write $V^* = [w_1 \mid \cdots \mid w_n]$; then $\{ w_1, \dots, w_n \}$ is orthonormal, and

$$V = \begin{pmatrix} w_1^* \\ \vdots \\ w_n^* \end{pmatrix}$$

So

$$\begin{aligned} A &= USV \\ &= [u_1 \mid \cdots \mid u_n] \begin{pmatrix} s_1 w_1^* \\ \vdots \\ s_n w_n^* \end{pmatrix} \\ &= \sum_{k=1}^n u_k (s_k w_k^*) \\ &= \sum_{k=1}^n s_k u_k w_k^* \end{aligned}$$

□ Corollary 207

7.5 Schur products

Definition 208. Suppose $A = [a_{i,j}]$ and $B = [b_{i,j}]$ in $M_{m,n}$. We define the *Schur (or Hadamard or freshman) product* is $A \circ B = [a_{i,j} b_{i,j}]$.

Proposition 209.

1. *(a)* $A \circ B = B \circ A$
   (b) $(A \circ B) \circ C = A \circ (B \circ C)$
   (c) $A \circ (B + C) = A \circ B + A \circ C$
   (d) $\lambda(A \circ B) = (\lambda A) \circ B = A \circ (\lambda B)$

   So $\circ \colon M_{m,n} \times M_{m,n} \to M_{m,n}$ *is bilinear.*

2. *If* $J_{m,n}$ *is the matrix of all* 1*, then* $J_{m,n} \circ A = A = A \circ J_{m,n}$.

Theorem 210. *Suppose* $A, B \in M_n$ *with* $A \geq 0$ *and* $B \geq 0$*. Then* $A \circ B \geq 0$.

Proof. Write

$$\begin{aligned} A &= \sum x_i x_i^* \\ B &= \sum y_i y_i^* \end{aligned}$$

Then

$$A \circ B = \sum_{i,j} (x_i x_i^*) \circ (y_i y_i^*)$$

So it suffices to prove that given

$$x = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

$$y = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

we have $(xx^*) \circ (yy^*) \geq 0$. Let

$$z = \begin{pmatrix} \alpha_1\beta_1 \\ \vdots \\ \alpha_n\beta_n \end{pmatrix}$$

Then $(xx^*) \circ (yy^*) = (\alpha_i\overline{\alpha_j}) \circ (\beta_i\overline{\beta_j}) = (\alpha_i\beta\overline{\alpha_j\beta_j}) = zz^* \geq 0$. □ Theorem 210

7.6 The positive semidefinite ordering

Given $A = A^*$ and $B = B^*$ in M_n, we write $A \geq B$ (or $B \leq A$) if and only if $A - B$ is positive semidefinite. Similarly, we write $A > B$ or $B < A$ if and only if $a - B$ is positive definite.

Proposition 211. *Suppose $A = A^*$, $B = B^*$, and $C = C^*$ in M_n. Then*

1. *If $A \leq B$ and $B \leq A$, then $A = B$.*
2. *If $A \leq B$ and $B \leq C$, then $A \leq C$.*
3. *If $A \leq B$ then $A + C \leq B + C$.*
4. *If $A \leq B$ and $X \in M_{n,m}$, then $X^*AX \leq X^*BX$.*

Lemma 212. *Suppose $A \geq I$. Then $I \geq A^{-1} > 0$.*

Proof. Since $A = A^*$, we may let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues, and find orthonormal $\{ v_1, \ldots, v_n \}$ such that $Av_i = \lambda_i v_i$. Then since $A \geq I$ we have $A - I \geq 0$; so $\lambda_i - 1 = \langle (\lambda_i - 1)v_i, v_i \rangle = \langle (A - I)v_i, v_i \rangle \geq 0$ for all i. So $A^{-1}v_i = \lambda_i^{-1}v_i$ and $\lambda_i^{-1} \leq 1$ for all i. So if

$$x = \sum \alpha_i v_i$$

then

$$\langle (I - A^{-1})x, x \rangle = \left\langle \sum_i (1 - \lambda_i)^{-1}\alpha_i v_i, \sum_j \alpha_j v_j \right\rangle = \sum (1 - \lambda_i)^{-1}|\alpha_i|^2 \geq 0$$

□ Lemma 212

Theorem 213. *Suppose $A = A^*$ and $B = B^*$ are in M_n.*

1. *If $A \geq B > 0$ then $A^{-1} \leq B^{-1}$.*
2. *If $A \geq B \geq 0$ then $\det(A) \geq \det(B)$ and $\operatorname{tr}(A) \geq \operatorname{tr}(B)$.*
3. *If $A \geq B$ then $\lambda_k(A) \geq \lambda_k(B)$.*

Proof.

1. Suppose $A \geq B > 0$. Then $B^{-\frac{1}{2}}AB^{-\frac{1}{2}} \geq B^{-\frac{1}{2}}BB^{-\frac{1}{2}} = I$; so, by the lemma, we have $(B^{-\frac{1}{2}}A^{-1}B^{-\frac{1}{2}})^{-1} \leq I$. So $A^{-1} \leq B^{-\frac{1}{2}}IB^{-\frac{1}{2}} = B^{-1}$.
2. Follows from (3).
3. Recall that b Courant-Fischer, we have

$$\begin{aligned} \lambda_k(A) &= \min_{\dim(S)=k} \max_{\substack{x \in S \\ \|x\|=1}} \langle Ax, x \rangle \\ &\geq \min_{\dim(S)=k} \max_{\substack{x \in S \\ \|x\|=1}} \langle Bx, x \rangle \\ &= \lambda_k(B) \end{aligned}$$

□ Theorem 213

Theorem 214. *Suppose $P \in M_n$ and $S \subseteq \{1, \ldots, n\}$. If $P > 0$, then $P^{-1}[S, S] \geq (P[S, S])^{-1}$.*

Proof. By permuting, it suffices to check the case $S = \{1, \ldots, k\}$. Partition

$$P = \begin{pmatrix} A & B \\ B^* & C \end{pmatrix}$$

where $A \in M_k$. Likewise partition

$$P^{-1} = \begin{pmatrix} D & E \\ E^* & F \end{pmatrix}$$

Then $P^{-1}[S, S] = D$. We wish to show that $D \geq A^{-1}$. Recall that in Cholesky we showed that if $X = -A^{-1}B$, then

$$P = \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} = \begin{pmatrix} I & 0 \\ -X^* & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & C - B^* A^{-1} B \end{pmatrix} \begin{pmatrix} I & -X \\ 0 & I \end{pmatrix}$$

So

$$\begin{aligned} P^{-1} &= \begin{pmatrix} I & -X \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} A & 0 \\ 0 & C - B^* A^{-1} B \end{pmatrix}^{-1} \begin{pmatrix} I & 0 \\ -X^* & I \end{pmatrix}^{-1} \\ &= \begin{pmatrix} I & X \\ 0xI & \end{pmatrix} \begin{pmatrix} A^{-1} & 0 \\ 0 & (C - B^* A^{-1} B)^{-1} \end{pmatrix} \begin{pmatrix} I & 0 \\ X^* & I \end{pmatrix} \\ &= \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} \begin{pmatrix} A^{-1} & 0 \\ (C - B^* A^{-1} B)^{-1} X^* & (C - B^* A^{-1} B)^{-1} \end{pmatrix} \\ &= \begin{pmatrix} A^{-1} + X(C - B^* A^{-1} B)^{-1} X^* & * \\ * & * \end{pmatrix} \end{aligned}$$

So $D = A^{-1} + X(C - B^* A^{-1} B)^{-1} X^*$. But by Cholesky, we have $C - B^* AB > 0$; so $X(C - B^* A^{-1} B)^{-1} X^* \geq 0$, and $D - A^{-1} \geq 0$; so $D \geq A^{-1}$. □ Theorem 214

8 Matrix norms

Definition 215. Suppose V is a vector space. We say $\|\cdot\| : V \to \mathbb{R}$ is a *norm* provided the following hold:

1. $\|v\| \geq 0$ for all $v \in V$.
2. $\|v\| = 0$ if and only if $v = 0$.
3. $\|\lambda v\| = |\lambda| v$
4. $\|v + w\| \leq \|v\| + \|w\|$.

Given a norm, the function $d(v, w) = \|v - w\|$ defines a metric on V.

Example 216.

$$\begin{aligned} \|v\|_w &= \left(\sum_{i=1}^{n} |v_i|^2 \right)^{\frac{1}{2}} \\ \|v\|_1 &= \sum_{i=1}^{n} |v_i| \\ \|v\|_\infty &= \max\{ |v_1|, \ldots, |v_n| \} \end{aligned}$$

Fact 217. *When $\dim(V) < \infty$, all norms are equivalent. i.e. given any two norms $\|\cdot\|_1$ and $\|\cdot\|_2$, there are c_1, c_2 such that for all $v \in V$ we have $\|v\|_1 \leq c_2 \|v\|_2$ and $\|v\|_2 \leq c_1 \|v\|_1$. Consequently, for any norm, we have $d(v_n, 0) = \|v_n\| \to 0$ if and only if all components tend to 0.*

Definition 218. A norm on M_n is called a *matrix norm* provided $\|AB\| \le \|A\|\|B\|$ for all $A, B \in M_n$ (submultiplicative).

Remark 219. We don't require that $\|I\| = 1$.

Example 220. For $A \in M_n$, set

$$\|A\|_2 = \left(\sum |a_{i,j}|^2\right)^{\frac{1}{2}}$$

is a matrix norm. To see this, suppose $C = AB$. Then

$$|c_{i,j}| = \left(\sum_k a_{ik} b_{kj}\right)^2 \le \left(\sum_{k=1}^n |a_{ik}^2|\right)^{\frac{1}{2}} \left(\sum_{\ell=1}^n |b_{\ell j}^2|\right)^{\frac{1}{2}}$$

So

$$\begin{aligned}
\|C\|_2^2 &= \sum_{i,j} |c_{i,j}|^2 \\
&\le \sum_{i,j} \left(\sum_k |a_{i,k}|^2\right)\left(\sum_\ell |b_{\ell,j}|^2\right) \\
&= \left(\sum_{i,k} |a_{ik}^2|\right)\left(\sum_{j,\ell} |b_{\ell j}|^2\right) \\
&= \|A\|_2^2 \|B\|_2^2
\end{aligned}$$

Note, however that $\|I_n\|_2 = \sqrt{n}$.

Example 221. Let

$$\|A\|_1 = \sum_{i,j=1}^n |a_{ij}|$$

This too is a matrix norm; here we have $\|I_n\|_1 = n$.

Example 222. Set

$$\|A\|_\infty = \max_{i,j} |a_{i,j}|$$

Let J_n be the $n \times n$ matrix of all 1. Then $J_n^2 = nJ_n$. But $\|J_n\|_\infty = 1$ and $n = \|J_n^2\|_\infty \not\le \|J_n\|_\infty \|J_n\|_\infty$.

Example 223. Given any norm $\|\cdot\|$ on $\mathbb{C}^n$, we define the *induced operator norm* on M_n by

$$|||A||| = \sup\{\, \|Ax\| : \|x\| = 1 \,\}$$

This is always a matrix norm; here $|||I||| = 1$.

Example 224. Start with $\|\cdot\|_2$ on $\mathbb{C}^n$. Set $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then the induced norm of D is

$$|||D|||_2 = \sup\{\, \|(\lambda_1 x_1, \ldots, \lambda_n x_n\| : \|(x_1, \ldots, x_n)\|_2 = 1 \,\}$$

So $|||D|||_2 \ge |\lambda_j|$ for all j; so $|||D|||_2 \ge \max\{\, |\lambda_1|, \ldots, |\lambda_n| \,\}$. Conversely, we have

$$\|(\lambda_1 x_1, \ldots, \lambda_n x_n)\|_2^2 = \sum |\lambda_j|^2 |x_j|^2 \le \max\{\, |\lambda_1|, \ldots, |\lambda_n| \,\} \sum |x_j|^2$$

So $|||D|||_2 \le \max\{\, |\lambda_1|, \ldots, |\lambda_n| \,\}$. So $|||D|||_2 = \max\{\, |\lambda_1|, \ldots, |\lambda_n| \,\}$.

Proposition 225. *Suppose $\|\cdot\|$ is any matrix norm on M_n. Then*

1. $\|I\| \ge 1$.
2. *If $A \in M_n^{-1}$, then* $\|A\| \ge \frac{\|I\|}{\|A\|}$.

3. *For all $\lambda \in \sigma(A)$, we have $|\lambda| \le \rho(A) \le \|A\|$.*

Proof.

1. $I^2 = I$ so $\|I\| = \|I^2\| \le \|I\|^2$; so $1 \le \|I\|$.
2. $\|I\| = \|AA^{-1}\| \le \|A\|\|A^{-1}\|$.
3. Let $\lambda \in \sigma(A)$. Then there is non-zero x such that $Ax = \lambda x$. Pick any $y \ne 0$; let $X = \lambda y^*$. Then $AX = (Ax)y^* = \lambda X$. So
$$|\lambda|\|X\| = \|\lambda X\| = \|AX\| \le \|A\|\|X\|$$
So $|\lambda| \le \|A\|$. But this holds for all $\lambda \in \sigma(A)$. So $\rho(A) \le \|A\|$.

☐ Proposition 225

Theorem 226. *Suppose $A \in M_n$. Then $\rho(A) = \inf\{\, \|A\| : \|\cdot\| \text{ a matrix norm}\,\}$.*

Proof. By (3) we have $\rho(A) = \inf\{\, \|A\| : \|\cdot\| \text{ a matrix norm}\,\}$. Given any matrix norm $\|\cdot\|$ and $S \in M_n^{-1}$, define $\|A\|_S = \|S^{-1}AS\|$; this is a matrix norm. By Schur there is U unitary such that $U^*AU = T = (t_{i,j})$ is upper triangular and $t_{ii} = \lambda_i$. Fix $r > 0$. Let $D_r = \operatorname{diag}(1, r, \dots, r^{n-1}$. Then

$$D_r^{-1}TD_r = (t_{i,j}r^{-i+j}) = \begin{pmatrix} \lambda_1 & rt_{12} & \dots & r^{n+1}t_{1,n} \\ & \ddots & \ddots & \\ & & \lambda_{n-1} & rt_{n-1,n} \\ 0 & & & \lambda_n \end{pmatrix}$$

So $S = UD_r$. then

$$\begin{aligned} \|A\|_S &= \|D + rT_1 + \dots + r^{n-1}T_{n-1}\| \\ &\le \|D\| + r\|T_1\| + \dots + r^{n-1}\|T_{n-1}\| \end{aligned}$$

Given $\varepsilon > 0$, for r small enough, this is $\le \|D\| + \varepsilon$. But we can do this starting with any matrix norm; starting with operator norm induced by $\|\cdot\|_2$, we get $\|D\| = \max\{\, |\lambda_1|, \dots, |\lambda_n|\,\} = \rho(A)$. So $\inf\{\, \|A\| : \|\cdot\| \text{ a matrix norm}\,\} \le \rho(A) + \varepsilon$. ☐ Theorem 226

Corollary 227. *Suppose $A \in M_n$ with $\rho(A) < 1$. Then $A^k \to 0$ as $k \to \infty$.*

Proof. Pick $|||\cdot|||$ a matrix norm such that $\rho(A) < |||A||| < 1$. Then $|||A^k||| \le |||A|||^k \to 0$. So $|||A^k||| \to 0$, and $A^k \to 0$. ☐ Corollary 227

Theorem 228 (Gelfand)**.** *Suppose $|||\cdot|||$ is any matrix norm on M_n; suppose $A \in M_n$. Then*

$$\rho(A) = \lim_n |||A^n|||^{\frac{1}{n}}$$

Proof. Recall that $\sigma(A^k) = \{\, \lambda^k : \lambda \in \sigma(A)\,\}$; so $\rho(A^k) = \rho(A)^k$. So $\rho(A)^k = \rho(A^k) \le |||A^k|||$, and $\rho(A) \le |||A^k|||^{\frac{1}{k}}$. Let $\varepsilon > 0$; let $r = \rho(A) + \varepsilon$. Then $\rho(\frac{A}{r}) < 1$. So, by our corollary, we have $|||(\frac{A}{r})^k||| \to 0$. So there is k_0 such that $|||(\frac{A}{r})^k||| < 1$ for all $k \ge k_0$. So $|||A^k||| < r^k$ for all $k \ge k_0$. So $|||A^k|||^{\frac{1}{k}} < \rho(A) + \varepsilon$ for all $k \ge k_0$. So for all $k \ge k_0$, we have

$$\left| |||A^k|||^{\frac{1}{k}} - \rho(A) \right| < \varepsilon$$

So

$$\lim_k |||A^k|||^{\frac{1}{k}} = \rho(A)$$

☐ Theorem 228

8.1 Power series

If

$$p(z) = \sum_{k=0}^{\infty} p_k z^k$$

we set

$$\limsup_k |p_k|^{\frac{1}{k}} = \frac{1}{R}$$

where R is the radius of convergence. Given a matrix $A \in M_n$, consider

$$p(A) = \sum_{k=0}^{\infty} p_k A^k$$

Let

$$B_n = \sum_{k=0}^{n} p_k A^k$$

If $B_n \to B$, then we write $B = p(A)$. Recall that $(B_n : n \in \mathbb{N})$ converges if and only if $\{ B_n \}$ is Cauchy.

Theorem 229. *Suppose*

$$p(z) = \sum_{p=0}^{\infty} p_k z^k$$

with radius of convergence $R > 0$. *Suppose* $A \in M_n$ *satisfies* $\rho(A) < R$. *Then*

$$\sum_{k=0}^{\infty} p_k A^k$$

converges.

Proof. Pick r_1, r_2 such that $\rho(A) < r_1 < r_2 < R$. Then there is k_1 such that $|||A^k|||^{\frac{1}{k}} < r_1$ for all $k \geq k_1$. Since $\frac{1}{R} < \frac{1}{r_2}$, we have that there is k_2 such that

$$\sup\{ |p_k|^{\frac{1}{k}} : k \geq k_2 \} < \frac{1}{r_2}$$

So $|p_k| < \frac{1}{r_2^k}$ for all $k \geq k_2$. Let $k_0 = \max\{ k_1, k_2 \}$. Then for all $k \geq k_0$ we have

$$|||p_k A^k||| = |p_k| |||A^k||| \leq \left(\frac{r_1}{r_2} \right)^k < 1$$

Thus for all $n, m > k_0$ we have

$$|||B_n - B_m||| = \left|\left|\left| \sum_{k=m+1}^{n} p_k A^k \right|\right|\right| \leq \sum_{k=m+1}^{n} \left(\frac{r_1}{r_2} \right)^k$$

Thus $(B_n : n \in \mathbb{N})$ is Cauchy; so $(B_n : n \in \mathbb{N})$ converges. □ Theorem 229

Corollary 230. *Suppose* $A \in M_n$. *If there is a matrix norm* $|||\cdot|||$ *such that* $|||I - A||| < 1$, *then* A *is invertible.*

Proof. We know that

$$\sum_{k=0}^{\infty} (I - A)^k$$

But

$$B_n = \sum_{k=0}^{n} (I - A)^k$$

so

$$\begin{aligned} AB_n &= (I - (I - A))B_n \\ &= \sum_{k=0}^{n} (I - A)^k - \sum_{k=1}^{n+1} (I - A)^k \\ &= I - (I - A)^{n+1} \\ &\to I \end{aligned}$$

So if

$$B = \sum_{k=0}^{\infty} (I - A)^k = \lim_n B_n$$

then

$$AB = \lim_n AB_n = I$$

□ Corollary 230

8.2 Gersgorin disks

Suppose $A \in M_n$; write $A = D + B$ with $D = \operatorname{diag}(a_{11}, \ldots, a_{nn})$ and B has 0 on the diagonal. For each i, define

$$R_i'(A) = \sum_{j \neq i} |a_{ij}|$$

Let

$$\Omega_i = \{\, z \in \mathbb{C} : |z - a_{ii}| \leq R_i'(A) \,\}$$

These are the *Gersgorin disks*.

Theorem 231. *Suppose $A = (a_{ij}) \in M_n$. Then*

$$\sigma(A) \subseteq \bigcup_{i=1}^{n} \Omega_i$$

Proof. Suppose $\lambda \in \sigma(A)$; say $Ax = \lambda x$ for $x = (x_1, \ldots, x_n) \neq 0$. Pick p such that $|x_p| = \max\{\, |x_1|, \ldots, |x_n| \,\} = \|x\|_\infty \neq 0$. Then

$$\lambda x_p = \sum_{j=1}^{n} a_{p,j} x_j$$

and

$$(\lambda - a_{p,p}) x_p = \sum_{j \neq p} a_{p,j} x_j$$

So

$$|\lambda - a_{pp}||x_p| \leq \sum_{j \neq p} |a_{p,j}||x_p|$$

and

$$|\lambda - a_{pp}| \leq R_p'(A)$$

So $\lambda \in \Omega_p = \{\, z : |z - a_{pp}| \leq R_p'(A) \,\}$, and

$$\lambda \in \bigcup_{i=1}^{n} \Omega_i$$

So

$$\sigma(A) \subseteq \bigcup_{i=1}^{n} \Omega_i$$

□ Theorem 231

Let

$$C'_j(A) = \sum_{i \neq j} |a_{ij}|$$

Let $\widetilde{\Omega_j} = \{ z : |z - a_{jj}| \leq C'_j(A) \}$.

Corollary 232. *Suppose $A \in M_n$. Then*

$$\sigma(A) \subseteq \left(\bigcup_{i=1}^{n} \Omega_i \right) \cap \left(\bigcup_{j=1}^{n} \widetilde{\Omega_j} \right)$$

Proof. Simply observe that $\sigma(A) = \sigma(A^t)$. □ Corollary 232

Definition 233. We define the Gersgorin set of $A \in M_n$ to be

$$G(A) = \bigcup_{i=1}^{n} \Omega_i$$

Then the above theorem states that $\sigma(A) \subseteq G(A)$.

Aside 234 (Some complex analysis). Suppose $\gamma\colon [0,1] \to \mathbb{C}$ is smooth with $\gamma(0) = \gamma(1)$. Let $\Gamma = \{ \gamma(t) : 0 \leq t \leq 1 \}$. If Γ does not intersect itself, we can sensibly define $\int(\Gamma)$ and $\operatorname{ext}(\Gamma)$. We then set

$$\int_\Gamma f(z) dz = \int_0^1 f(\gamma(t))\gamma'(t) dt$$

Recall from complex analysis that if $p(z)$ is a polynomial with $p(\gamma(t)) \neq 0$ for all $t \in [0,1]$, then

$$\frac{1}{2\pi i} \int_\Gamma \frac{p'(z)}{p(z)} dz$$

is the number of roots of p inside Γ with multiplicities.

Theorem 235. *Suppose $A = (a_{ij}) \in M_n$; suppose $\Omega_1, \ldots, \Omega_n$ are the Gersgorin disks. Suppose*

$$(\Omega_{i_1} \cup \cdots \cup \Omega_{i_k}) \cap \left(\bigcup_{i \notin \{ i_1, \ldots, i_k \}} \Omega_i \right) = \emptyset$$

Then there are k roots of $p_A(t)$ inside $\Omega_{i_1} \cup \cdots \cup \Omega_{i_k}$.

Proof. Write $A = D + B$ where $D = \operatorname{diag}(a_{11}, \ldots, a_{nn})$. Define $A_s = D + sB$; so $A_0 = D$ and $A_1 = A$. Take a Γ such that $\Omega_{i_1} \cup \cdots \cup \Omega_{i_k} \subseteq \int(\Gamma)$ and

$$\bigcup_{i \notin \{ i_1, \ldots, i_k \}} \Omega_i \subseteq \operatorname{ext}(\Gamma)$$

Let $p_s(z) = \det(zI - A_s) = p_{A_s}(z)$. Let

$$N(s) = \frac{1}{2\pi i} \int_\Gamma \frac{p'_s(z)}{p_s(z)} dz$$

(Since $G(A_s) \subseteq G(A)$, we know $p_s(z) \neq 0$ on Γ.) Then $N(s)$ is the number of roots of $p_s(t)$ inside Γ. Note, however, that p_s varies continuously with s; so $N(s)$ is a continuous function of s. But $N(s)$ is integer-valued; so $N(s)$ is constant. So $N(0) = N(1)$. But $A_0 = D$, and $p_D(z) = (z - a_{11}) \ldots (z - a_{nn})$ has k roots inside Γ. So $N(0) = k$. So $N(1) = k$. But $N(1)$ is the number of roots of $p_A(z)$ inside Γ. Thus, by Gersgorin, we're done. □ Theorem 235

Let $D = \operatorname{diag}(p_1, \ldots, p_n)$ with all $p_i > 0$. Then

$$D^{-1}AD = (a_{ij}p_i^{-1}p_j)$$

and $\sigma(A) = \sigma(D^{-1}AD) \subseteq G(D^{-1}AD)$. Then

$$R_i' = \sum_{j \neq i} p_i^{-1}|a_{ij}|v_j = p_i^{-1}\left(\sum_{j \neq i}|a_{ij}|p_j\right)$$

This yields the following corollary:

Corollary 236. *Suppose $A \in M_n$. Then*

$$\sigma(A) \subseteq \bigcap_{p_1,\ldots,p_n>0}\left(\bigcup_{i=1}^{n}\left\{ z : |z - a_{ii} \leq p_i^{-1}\left(\sum_{j \neq i}|a_{ij}|p_j\right)| \right\}\right)$$

Proof. RHS is

$$\bigcap_{p_1,\ldots,p_n} G(D_p^{-1}AD_p)$$

$\square$ Corollary 236

Aside 237. An *ellipse* is given by r and foci a, b; it is then

$$\{ z : |z - a| + |z - b| = r \}$$

An *oval of Cassini* is analogous:

$$\{ z : |z - a||z - b| = r \}$$

Theorem 238 (Brauer)**.** *Suppose $A = (a_{ij}) \in M_n$ for $n \geq 2$. Then*

$$\sigma(A) \subseteq \bigcup_{i \neq j}\{ z : |z - a_{ii}||z - a_{jj}| \leq r_i' R_j' \}$$

where

$$R_i' = \sum_{j \neq i}|a_{ij}|$$

Fact 239. *The union of these ovals is contained in $G(A)$.*

Proof of Theorem 238. Suppose $\lambda \in \sigma(A)$; say $Ax = \lambda x$ for $x \neq 0$. Let $|x_p| = \max\{ |x_1|, \ldots, |x_n| \} \neq 0$. If

$$\max_{i \neq p}|x_i| = 0$$

then $x = x_p e_p$. So $(Ax)_p = a_{pp}x_p e_p = \lambda(x_p e_p)$; so $\lambda = a_{pp}$, and $\lambda \in$ RHS.

Assume then that

$$|x_q| = \max_{i \neq p}|x_i| \neq 0$$

Then

$$\lambda x_p = \sum_{j=1}^{n} a_{pj}x_j$$

So

$$|(\lambda - a_{pp})x_p| = \left|\sum_{j \neq p} a_{pj}x_j\right| \leq \sum_{j \neq p}|a_{pj}||x_q| = |x_q|R_p'$$

So

$$|\lambda - a_{pp}| \le \frac{|x_q|}{|x_p|} R'_p$$

But we also have

$$\lambda x_q = \sum_{j=1}^{n} a_{qj} x_j$$

So

$$|(\lambda - a_{qq}) x_q| = \left| \sum_{j \ne q} a_{qj} x_j \right| \le |x_p| R'_q$$

So

$$|\lambda - a_{qq}| \le \frac{|x_p|}{|x_q|} R'_q$$

Putting it all together, we have

$$|\lambda - a_{pp}||\lambda - a_{qq}| \le R'_p R'_q$$

so $\lambda \in$ RHS. ☐ Theorem 238

Definition 240. $A = (a_{ij})$ is *(strictly) diagonally dominant* provided

$$|a_{ii}| \ge \sum_{j \ne i} |a_{ij}| = R'_i$$

for diagonally dominant, and

$$|a_{ii}| > \sum_{j \ne i} |a_{ij}| = R'_i$$

for strictly diagonally dominant.

Theorem 241. *Suppose $A = (a_{ij})$ is strictly diagonally dominant. Then*

1. *$A \in M_n^{-1}$.*
2. *If $a_{ii} > 0$, then $\sigma(A) \subseteq \{ \lambda : \mathrm{Re}(\lambda) > 0 \}$; we call this latter set the right half-plane (RHP).*
3. *If $A = A^*$ and al $a_{ii} > 0$, then A is positive definite.*

Proof.

1. $0 \notin G(A)$ implies $0 \notin \sigma(A)$, since $\sigma(A) \subseteq G(A)$.
2. Note that $G(A) \subseteq$ RHP.
3. By (1), we have $A \in M_n^{-1}$. Also $\sigma(A) \subseteq$ RHP. But $A = A^*$, so $\sigma(A) \subseteq \mathbb{R}$. So $\sigma(A) \subseteq (0, \infty)$, and A is positive definite.

☐ Theorem 241

9 Non-negative matrices

Used in combinatorics, probability, and Markov chains.

An application: suppose $A = (a_{ij})$ with $a_{ij} \in \mathbb{N} \cup \{ 0 \}$. We interpret this as a directed graph, where a_{ij} is the number of edges from vertex j to vertex i. Then $(A^k)_{i,j}$ is the number of paths of length k from j to i.

Another application: imagine you have states $\{ 1, \dots, n \}$ with $p(i \mid j)$ the probability of going from state j to state i. Let $P = (p(i \mid j))$. Let

$$q = \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix}$$

where $q_i \ge 0$ and $q_1 + \cdots + q_n = 1$; we interpret q_i as the probability of initially being in state i. Then $(Pq)_i$ is the probability that after one event we are in state i. We can ask whether $P^k q$ converges as $k \to \infty$.

Definition 242. Suppose $A = (a_{ij}) \in M_{m,n}$. We say A is *non-negative* ($0 \preceq A$)) if $a_{ij} \geq 0$ for all i, j. We say A is *positive* ($0 \prec A$) if $a_{ij} > 0$ for all i, j. If $A, B \in M_{m,n}(\mathbb{R})$, we write $A \preceq B$ if $b_{ij} \geq a_{ij}$ for all i, j; we say $A \prec B$ if $b_{ij} > a_{ij}$ for all i, j. We write $|A|_e = (|a_{ij}|)$.

Proposition 243.

1. *$|aA|_e = |a||A|_e$.*
2. *$|A + B|_e \preceq |A|_e + |B|_e$.*
3. *If $0 \preceq A$, $0 \preceq B$, $a \geq 0$, and $b \geq 0$, then $0 \preceq aA + bB$.*
4. *If $A \preceq B$ and $C \preceq D$, then $A + C \preceq B + D$.*
5. *$|AB|_e \leq |A|_e|B|_e$.*
6. *Suppose $0 \preceq A \preceq B$ and $0 \preceq C \preceq D$. Then $AC \preceq BD$.*

Proof.

5. Note that

$$\begin{aligned} |AB|_e &= \left(\left|\sum_{k=1}^{n} a_{ik}b_{kj}\right|\right) \\ &\preceq \left(\sum_{k=1}^{n}|a_{ik}||b_{kj}|\right) \\ &= |A|_e|B|_e \end{aligned}$$

□ Proposition 243

Proposition 244. *Say $A \in M_n$. Then $\rho(A) \leq \rho(|A|_e)$.*

Proof. Take

$$\|B\|_2 = \left(\sum |b_{ij}|^2\right)^{\frac{1}{2}}$$

Then this is a matrix norm, and $\left\||B|_e\right\|_2 = \|B\|_2$. Also $|A^n|_e \preceq |A|_e^n$. So

$$\|A^n\|_2 = \left\||A^n|_e\right\|_2 \leq \left\||A|_e^n\right\|_2$$

So

$$\rho(A) = \lim_n \|A^n\|_2^{\frac{1}{n}} \leq \lim_n \left\||A|_e^n\right\|_2^{\frac{1}{n}} = \rho(|A|_e)$$

□ Proposition 244

Theorem 245. *Suppose $A, B \in M_n$. Suppose $|A|_e \preceq B$. Then $\rho(A) \leq \rho(|A|_e) \leq \rho(B)$.*

Proof. We showed that $\rho(A) \leq \rho(|A|_e)$. But $|A|_e^n \preceq B^n$ implies that $\left\||A|_e^n\right\|_2 \leq \|B^n\|_2$. So

$$\rho(|A|_e) = \lim_n \left\||A|_e^n\right\|_2^{\frac{1}{n}} \leq \lim_n \|B^n\|_2^{\frac{1}{n}} = \rho(B)$$

□ Theorem 245

Corollary 246. *Suppose $A \in M_n$ with $0 \preceq A$. Then*

1. *For all $S \subseteq \{1, \ldots, n\}$ we have $\rho(A[S, S]) \leq \rho(A)$.*
2. *$\max\{a_{11}, \ldots, a_{nn}\} \leq \rho(A)$.*

3. *If there is i such that $a_{ii} \neq 0$, then $\rho(A) > 0$.*

Proof.

1. We may assume A takes the form
$$A = \begin{pmatrix} A[S,S] & B \\ C & D \end{pmatrix}$$
But then
$$\begin{pmatrix} A[S,S] & 0 \\ 0 & 0 \end{pmatrix} \preceq A$$
So
$$\rho(A[S,S]) = \rho \begin{pmatrix} A[S,S] & 0 \\ 0 & 0 \end{pmatrix} \leq \rho(A)$$

2. Note that $a_{ii} = A[\{ i \}, \{ i \}]$. So $a_{ii} = |a_{ii}| = \rho(A[\{ i \}, \{ i \}]) \leq \rho(A)$. So $\max\{ a_{11}, \dots, a_{nn} \} \leq \rho(A)$.

3. $0 < \max\{ a_{11}, \dots, a_{nn} \} \leq \rho(A)$.

☐ Corollary 246

Proposition 247. *For $x \in \mathbb{C}^n$, let*
$$\|x\|_1 = \sum_{j=1}^{n} |x_j|$$
$$\|x\|_\infty = \max\{ |x_1|, \dots, |x_n| \}$$
Then for $A \in M_n$, the induced operator norms are
$$|||A|||_1 = \max_j \left\{ \sum_{i=1}^{n} |a_{ij}| \right\}$$
$$|||A|||_\infty = \max_i \left\{ \sum_{j=1}^{n} |a_{ij}| \right\}$$
Both of these are matrix norms.

Proof. Let $\|x\|_1 = 1$. Let $A = [C_1 \mid \cdots \mid C_n]$. Let C_k be a column with $\|C_k\|_1$ maximum. Then
$$\begin{aligned} \|Ax\|_1 &= \|x_1 C_1 + \cdots + x_n C_n\| \\ &\leq |x_1| \|C_1\|_1 + \cdots + |x_n| \|C_n\| \\ &\leq |x_1| \|C_k\|_1 + \cdots + |x_n| \|C_k\|_1 \\ &= \|C_k\|_1 \end{aligned}$$
So $|||A|||_1 \leq \max\{ \|C_1\|_1, \dots, \|C_n\|_1 \}$. For equality, let $x = e_k$. Then $\|x\|_1 = 1$ and $Ax = C_k$. So $|||A|||_1 \geq \|C_k\|_1$ for all k. So $|||A|||_1 \geq \max\{ \|C_1\|_1, \dots, \|C_n\|_1 \}$. So$|||A|||_1 = \max\{ \|C_1\|_1, \dots, \|C_n\|_1 \}$.

The case $|||\cdot|||_\infty$ is similar. ☐ Proposition 247

Corollary 248. *Suppose $A \in M_n$. Then $\rho(A) \leq \min\{ |||A|||_1, |||A|||_\infty \}$.*

Corollary 249. *Suppose $A \in M_n$ with $0 \preceq A$.*

1. *If*
$$\sum_j a_{ij} = |||A|||_\infty$$
for all i, then $\rho(A) = |||A|||_\infty$.

2. *If*

$$\sum_i a_{ij} = \|\|A\|\|_1$$

for all j, then $\rho(A) = \|\|A\|\|_1$.

Proof.

1. Let $e = (1, \ldots, 1)^t$. Then

$$(Ae)_i = \sum_{j=1}^n a_{ij} = \|\|A\|\|_\infty$$

So $Ae = \|\|A\|\|_\infty e$. So $\|\|A\|\|_\infty \in \sigma(A)$. So $\|\|A\|\|_\infty \le \rho(A) \le \|\|A\|\|_\infty$. So $\|\|A\|\|_\infty = \rho(A)$.

Remark 250. In this case we have $\|\|A\|\|_\infty \le \|\|A\|\|_1$.

2. Similar: note that $e^t A = \|\|A\|\|_1 d^t$. So $\|\|A\|\|_1 \in \sigma(A^t) = \sigma(A)$. So $\|\|A\|\| \le \rho(A) \le \|\|A\|\|_1$.

Remark 251. In this case we have $\|\|A\|\|_1 \le \|\|A\|\|_\infty$.

□ Corollary 249

Theorem 252. *Suppose $A \in M_n$ with $0 \preceq A$. Then*

$$\min_i \sum_{j=1}^n a_{ij} \le \rho(A) \le \max_i \sum_{j=1}^n a_{ij}$$

and

$$\min_j \sum_{i=1}^n a_{ij} \le \rho(A) \le \max_j \sum_{i=1}^n a_{ij}$$

Proof. The second pair of inequalities follows from the first pair applied to A^t, since $\rho(A) = \rho(A^t)$. It remains to prove the first pair.

Right-hand inequality By Proposition 247, we have

$$\rho(A) \le \|\|A\|\|_1 = \max_i \sum_{j=1}^n |a_{ij}|$$

Alternatively, we can use Gersgorin disks.

Left-hand inequality Let

$$\alpha = \min_i \sum_{j=1}^n a_{i,j}$$

If $\alpha = 0$ then there is nothing to prove. Suppose then that $\alpha \neq 0$. Define $B \in M_n$ by

$$b_{i,j} = \frac{\alpha}{\sum_{j=1}^n a_{ij}} a_{i,j}$$

Then $0 \le b_{i,j} \le a_{i,j}$; so $0 \preceq B \preceq A$. So $\rho(B) \le \rho(A)$. But

$$\sum_{j=1}^n b_{i,j} = \alpha$$

for all i. So, letting $e = (1, \ldots, 1)$, we see $Be = \alpha e$; so $\alpha \le \rho(B)$. So $\alpha \le \rho(A)$.

□ Theorem 252

Corollary 253. *Suppose $A \in M_n$ with $0 \preceq A$. Then*

$$\sup_{0 \prec x} \min_i \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j \le \rho(A) \le \inf_{0 \prec x} \max_i \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j$$

and

$$\sup_{0 \prec x} \min_j \frac{1}{x_i} \sum_{i=1}^n a_{ij} x_i \le \rho(A) \le \inf_{0 \prec x} \max_j \frac{1}{x_j} \sum_{i=1}^n a_{ij} x_i$$

Proof. Again, suffices to do the first pair of inequalities. Let $D = \operatorname{diag}(x_1, \ldots, x_n)$. Then $\rho(D^{-1}AD) = \rho(A)$. Apply the above theorem to $D^{-1}AD = (\frac{1}{x_i} a_{ij} x_j)$. □ Corollary 253

Corollary 254. *Suppose $0 \preceq A$ and $0 \prec x$. If $\alpha, \beta \ge 0$ satisfy $\alpha x \preceq Ax \preceq \beta x$, then $\alpha \le \rho(A) \le \beta$. Furthermore, if $\alpha x \prec Ax \prec \beta x$, then $\alpha < \rho(A) < \beta$.*

Proof. For each i we have

$$\alpha x_i \le \sum_{j=1}^n a_{ij} x_j \le \beta x_i$$

and thus

$$\alpha \le \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j \le \beta$$

The result then follows by the previous corollary. □ Corollary 254

Corollary 255. *Suppose $0 \preceq A$ and $0 \prec x$ with $Ax = \lambda x$. Then*

1. *$\lambda = \rho(A)$.*

2. $$\max_{x \succ 0} \min_i \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j = \rho(A) = \min_{x \succ 0} \max_i \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j$$

Proof. 1. $\lambda x = Ax$, so $\lambda x \preceq Ax \preceq \lambda x$. So, by the previous corollary, we have $\lambda \le \rho(A) \le \lambda$, and $\lambda = \rho(A)$.

2. For all i we have

$$\lambda x_i = \sum_{j=1}^n a_{ij} x_j$$

So

$$\lambda = \frac{1}{x_i} \sum_{j=1}^n a_{ij} x_j$$

for all i, The result then follows.

□ Corollary 255

Theorem 256. *Suppose $0 \preceq A$. Then there is $0 \prec x$ such that $Ax = \rho(A)x$.*

Proof. We know $\rho(A) > 0$. Also, if $0 \preceq x$ with $x \ne 0$, then $0 \prec Ax$. So if $0 \preceq x$ and $Ax = \rho(A)x$, then $0 \prec x$.
Let $\lambda \in \sigma(A)$ satisfy $|\lambda| = \rho(A)$. So there is $y \ne 0$ such that $Ay = \lambda y$. Let $x = |y|_e$. Then

$$(Ax)_i = \sum_{j=1}^n a_{ij} |y_j| > 0$$

So $0 \prec Ax$. But we also have

$$(Ax)_i = (A|y|_e)_i = \sum_{j=1}^n a_{ij} |y_j| \ge \left| \sum_{j=1}^n a_{ij} y_j \right| = |\lambda y_i| = |\lambda||y_i| = |\lambda| x_i = \rho(A) x_i$$

So $|\lambda|x \preceq Ax$, and $\rho(A)x \preceq Ax$. Now, let $w = Ax - \rho(A)x$. If $w = 0$, the theorem is proven. Assume then that $w \neq 0$. But $0 \prec Aw = A(Ax) - \rho(A)(Ax)$; so $\rho(A)(Ax) \prec A(Ax)$. If we then let $\alpha = \rho(A)$ and apply Corollary 254, we get that $\alpha < \rho(A)$, a contradiction. So $w = 0$, and $Ax = \rho(A)x$. ☐ Theorem 256

Remark 257. The proof showed that if $Ay = \lambda y$ with $|\lambda| = \rho(A)$ and $x = |y|_e$, then $Ax = \rho(A)x$.

Theorem 258 (Perron)**.** *Suppose $A \in M_n$ with $0 \prec A$. Then there is a unique $x = (x_1, \dots, x_n)$ such that*

- *Each $x_i \geq 0$.*
- *$x_1 + \cdots + x_n = 1$.*
- *$Ax = \rho(A)x$.*

Proof. By above there is $\widetilde{x} = (\widetilde{x}_1, \dots, \widetilde{x}_n)$ with each $\widetilde{x}_i \geq 0$ and $A\widetilde{x} = \rho(A)\widetilde{x}$ such that if $r = \widetilde{x}_1 + \cdots + \widetilde{x}_n$ and $x = \frac{1}{r}\widetilde{x}$ then $Ax = \rho(A)x$. So there exists one such vector.

Suppose now that $y = (y_1, \dots, y_n)$ satisfies each $y_i \geq 0$, $y_1 + \cdots + y_n = 1$, and $Ay = \rho(A)y$. Let

$$\beta = \min_i y_i x_i^{-1}$$

Then each $y_i - \beta x_i \geq 0$; so, if $w = y - \beta x$, then $w \geq 0$, and one entry of w is 0. Then, if we had $w \neq 0$, we have $0 \prec Aw = Ay - \beta Ax = \rho(A)(y - \beta x) = \rho(A)w$, contradicting our statement that w has 0 as an entry. So $w = 0$, and $y = \beta x$. So $1 = y_1 + \cdots + y_n = \beta(x_1 + \cdots + x_n) = \beta$. So $y = x$. ☐ Theorem 258

Corollary 259. *Suppose $0 \prec A$. Then $\rho(A)$ is a root of $p_A(t)$ of geometric multiplicity exactly 1.*

Proof. Follows from theorem above. ☐ Corollary 259

Question 260 (Challenge). What about the algebraic multiplicity?

Solution:

Theorem 261 (1.4.12)**.** *Suppose $A \in M_n$, $\lambda \in \mathbb{C}$, $x \neq 0$, $y \neq 0$, $Ax = \lambda x$, and $y^*A = \lambda y^*$.*

1. *If λ has algebraic multiplicity 1, then $y^*x \neq 0$.*
2. *Assume λ has geometric mult 1. Then it has algebraic multiplicity 1 if and only if $y^*x \neq 0$.*

Proof.

1. By Schur, we have
$$\widetilde{A} = U^*AU = \begin{pmatrix} \lambda & * \\ 0 & B \end{pmatrix}$$
and $x = e_1$. Suppose $y^*x = y^*e_1 = 0$; then $y^* = (0, \widetilde{y}^*)$. Then
$$(0, \lambda\widetilde{y}) = \lambda y^* = y^*\widetilde{A} = (0, \widetilde{y}) \begin{pmatrix} \lambda & * \\ 0 & B \end{pmatrix} = (0, \widetilde{y}B)$$
so $\lambda \in \sigma(B)$. So $(t - \lambda) \mid p_B(t)$. So $(t - \lambda)^2 \mid p_A(t) = (t - \lambda)p_B(t)$. So λ has algebraic multiplicity > 1.

2.

($\implies$) By (1).

($\impliedby$) Show $p_A'(\lambda) = \gamma y^*x$ for $\gamma \neq 0$; this uses the adjugate. Thus, if $p_A(t) = (t - \lambda)^2 q(t)$, then $p_A'(\lambda) = 0$; so $(t - \lambda)$ only occurs to the first power in $p_A(t)$.

☐ Theorem 261

Definition 262. Suppose $0 \prec A$. We proved that there is a unique $x = (x_1, \dots, x_n)$ with each $x_i > 0$ and $x_1 + \cdots + x_n = 1$ such that $Ax = \rho(A)x$. This vector is called the *right Perron vector* for A. Since $y^tA = \rho(A)y^t$ if and only if $A^ty = \rho(A)y^t$, we know there is $y = (y_1, \dots, y_n)$ with each $y_i > 0$ such that $y^tA = \rho(A)y^t$. Thus there is a unique $y = (y_1, \dots, y_n)$ such that

1. $y^t A = \rho(A) y^t$
2. $y_i > 0$ for all i
3. $y_1 x_1 + \cdots + y_n x_n = 1$

This is called the *left Perron vector* for A.

Corollary 263. *Suppose $0 \prec A$. Then $\rho(A)$ is of algebraic multiplicity 1.*

Proof. We know $\rho(A)$ is of geometric multiplicity 1. Apply 1.4.12, part (2). ☐ Corollary 263

This solves Question 260.

Theorem 264 (Perron)**.** *Suppose $A \in M_n$ with $0 \prec A$. Then*

1. *$\rho(A) > 0$*
2. *$\rho(A)$ is an eigenvalue of geometric multiplicity one.*
3. *There is a unique $x = (x_1, \ldots, x_n)$ with each $x_i > 0$ and $x_1 + \cdots + x_m = 1$ such that $Ax = \rho(A)x$.*
4. *There is a unique $y = (y_1, \ldots, y_n)$ with each $y_i > 0$ and $y_1 + \cdots + y_n = 1$ such that $y^t A = \rho(A) y^t$.*
5. *For all $\lambda \in \sigma(A)$ with $\lambda \neq \rho(A)$ we have $|\lambda| < \rho(A)$.*

Proof.

(1)-(3) Done already.

(4) Apply (3) to A^t.

(5) Suppose $\lambda \in \sigma(A)$ with $|\lambda| = \rho(A)$ and $\lambda \neq \rho(A)$. Pick $x \neq 0$ such that $Ax = \lambda x$. Then $\rho(A)|x|_e = |\lambda x|_e = |Ax|_e \preceq A|x|_e$. Let $w = A|x|_e - \rho(A)|x|_e \succeq 0$. If $w \neq 0$ then $0 \prec Aw = A(A|x|_e) - \rho(A)A|x|_e$; so $\rho(A)(A|x|_e) \prec A(A|x|_e)$; so $\rho(A) < \rho(A)$ by a previous theorem, a contradiction. So $w = 0$ and $A|x|_e = \rho(A)|x|_e$. Returning to the inequality, we find

$$\rho(A)|x|_e = |\lambda x|_e = |Ax|_e \leq A|x|_e = \rho(A)|x|_e$$

So $|Ax| = A|x|$, and

$$\left| \sum_{j=1}^{n} a_{ij} x_j \right| = \sum_{j=1}^{n} a_{ij} |x_j|$$

So the x_j are all collinear in $\mathbb{C}$; so there is θ such that $\exp(i\theta)x_j = |x_j|$ for all j. But then

$$\begin{aligned}
\rho(A)\exp(i\theta)x &= \rho(A)|x| \\
&= A|x| \\
&= A(\exp(i\theta)x) \\
&= \exp(i\theta)Ax \\
&= \exp(i\theta)\lambda x
\end{aligned}$$

So $\rho(A) = \lambda$, contradicting our assumption that $\rho(A) \neq \lambda$. So $\rho(A) = \lambda$.

☐ Theorem 264

Theorem 265. *Suppose $A \in M_n$ with $0 \preceq A$. Then $\rho(A)$ is an eigenvalue of A with a non-negative eigenvector.*

Proof. Let $J_n \in M_n$ be the matrix of all ones; set $A_k = A + \frac{1}{k}J_n$. So $0 \prec A_k$. So, by Perron's theorem there is $x_k = (x_{1k}, \dots, x_{nk})$ with $x_{jk} \geq 0$ and $x_{1k} + \cdots + x_{nk} = 1$ such that $A_k x_k = \rho(A_k)x_k$. These come from a bounded set; thus there is k_ℓ such that $(x_{k_\ell} : \ell \in \mathbb{N}) \to y = (y_1, \dots, y_n)$. Then each $y_j \geq 0$ and $y_1 + \cdots + y_n = 1$. Furthermore, we have $(A_{k_\ell} : \ell \in \mathbb{N}) \to A$; so

$$Ay = \lim_{\ell\to\infty} A_{k_\ell} x_{k_\ell} = \lim_{\ell\to\infty} \rho(A_{k_\ell}) x_{k_\ell}$$

So

$$\lim_{\ell\to\infty} \rho(A_{k_\ell}) = \mu \geq 0$$

and $Ay = \mu y$. So $0 \leq \mu \leq \rho(A)$. But $A \preceq A_k$; so $\rho(A) \leq \rho(A_k)$, and

$$\rho(A) \leq \lim_{\ell\to\infty} \rho(A_{k_\ell}) = \mu$$

So $\mu = \rho(A)$, and $Ay = \mu y$ for $y \succeq 0$. □ Theorem 265

9.1 Irreducible non-negative matrices

Definition 266. A matrix $A \in M_n$ is *reducible* if there is a permutation matrix P such that

$$P^t AP = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$$

where $B \in M_r$. We say A is *irreducible* if A is not reducible.

Definition 267. Given a vertex set $\{1, \dots, n\}$ and a non-empty $E \subseteq \{1, \dots, n\} \times \{1, \dots, n\}$, we think of there being a *path of length* 1 from i to j if and only if $(i, j) \in E$. This is what we mean by a *directed graph* (with loops).

Definition 268. Suppose $A \in M_n$ with $0 \preceq A$. We define $\Gamma(A)$, the *directed graph of* A, to have n vertices set and edge set $E = \{(i, j) : a_{i,j} \neq 0\}$.

Definition 269. Given a directed graph Γ with edge set E, we say that there is a *path of length* k from i to j if there is $(i, i_1), (i_1, i_2), \dots, (i_{k-2}, i_{k-1}), (i_{k-1}, j) \in E$. We say that Γ is *path-connected* if given any i and j there is a path from i to j.

Proposition 270. *Suppose $0 \preceq A$. Then $(A^m)_{i,j} \neq 0$ if and only if there is a path of length m from i to j in $\Gamma(A)$.*

Proof. We apply induction on m. Assume the proposition holds for m. Then $(A^{m+1})_{i,j} \neq 0$ if and only if there is k such that $(A^m)_{i,k} \neq 0$ and $A_{k,j} \neq 0$. By the induction hypothesis, this holds if and only if there is k and a path of length m from i to k and a path of length 1 from k to j; but this is just the definition of there being a path of length m from i to j.

So, by induction, the proposition holds. □ Proposition 270

Proposition 271. *Suppose Γ is a directed graph on n vertices. If there is a path from i to j with $i \neq j$, then there is a directed path of length $\leq n - 1$ from i to j.*

Proof. Suppose you have a longer path. Then it passes through some vertex twice. Then you can shorten it. □ Proposition 271

Theorem 272. *Suppose $A \in M_n$ with $0 \preceq A$. Then the following are equivalent:*

1. *A is irreducible.*
2. *$\Gamma(A)$ is path-connected.*
3. *$0 \prec (I + A)^{n-1}$.*

Proof.

(3) $\implies$ (1) Suppose A is reducible. Then, after permutation, we have

$$A = \begin{pmatrix} B & X \\ 0 & C \end{pmatrix}$$

So

$$(I+A)^{n-1} = \begin{pmatrix} I+B & X \\ 0 & I+C)^{n-1} \end{pmatrix} = \begin{pmatrix} (I+B)^{n-1} & * \\ 0 & (I+C)^{n-1} \end{pmatrix}$$

and (3) fails.

(1) $\implies$ (2) Suppose $\Gamma(A)$ is not path-connected. After renumbering, we can say that there is no path from n to 1. Renumbering, let $\{2,\dots,r\}$ be the vertices with a path to 1; let $\{r+1,\dots,n\}$ be the vertices with no path to 1. Suppose for contradiction that $a_{i,j} \neq 0$ for some $r+1 \le i \le n$ and some $1 \le j \le r$. Then there is a path from i to j. But there is a path from j to 1. So there is a path from i to 1, a contradiction. So $a_{i,j} = 0$ whenever $r+1 \le i \le n$ and $1 \le j \le r$. So

$$P^t A P = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}$$

so A is reducible.

(2) $\implies$ (3) For all $i \neq j$ we know there is $m \le n-1$ and a path of length m from i to j. So $(A^m)_{i,j} \neq 0$. So $I + A + A^2 + \cdots + A^{n-1} \succ 0$. So $(I+A)^{n-1} = I + \binom{n}{1}A + \binom{n}{2}A^2 + \cdots + A^{n-1} \succ 0$.

☐ Theorem 272

Theorem 273 (Perron-Frobenius)**.** *Suppose $n \ge 2$; suppose $A \in M_n$ is irreducible with $0 \preceq A$. Then:*

1. *$\rho(A) > 0$.*
2. *$\rho(A)$ is an eigenvalue of algebraic multiplicity 1.*
3. *There is a unique $x = (x_1,\dots,x_n) \in \mathbb{R}^n$ with $x_1 + \cdots + x_m = 1$ such that $Ax = \rho(A)x$. Moreover, $0 \prec x$.*
4. *There is a unique $y \in \mathbb{R}^n$ with $y \cdot x = 1$ and $y^T A = \rho(A) y^T$. Moreover, $0 \prec y$.*

Proof.

1. We know

$$\min_i \sum_{j=1}^n a_{ij} \le \rho(A)$$

If this minimum is 0, then one row is all 0; we permute so that this is the last row. Then

$$P^t A P = \begin{pmatrix} * & * \\ 0 & 0 \end{pmatrix}$$

so A is reducible, a contradiction. So

$$\min_i \sum_{j=1}^n a_{ij} \neq 0$$

So $0 < \rho(A)$.

2. Since $0 \preceq A$, we have that there is $0 \preceq x$ with $x \neq 0$ such that $Ax = \rho(A)x$. So

$$(I+A)^{n-1}x = (1+\rho(A))^{n-1}x$$

But

$$\rho((I+A)^{n-1}) = \sup\{\, |(1+\lambda)^{n-1}| : \lambda \in \sigma(A)\,\} = (1+\rho(A))^{n-1}$$

But A is irreducible; so $0 \prec (I+A)^{n-1}$. So x is the Perron vector and $(1+\rho(A))^{n-1}$ is an eigenvalue of algebraic multiplicity 1 for $(I+A)^{n-1}$.

Suppose for contradiction that $\rho(A)$ has algebraic multiplicity ≥ 2 for A. Then, by Schur, we have

$$U^*AU = \begin{pmatrix} \rho(A) & * & * & * \\ & \rho(A) & * & * \\ & & * & * \\ 0 & & & \ddots \end{pmatrix} = T$$

is upper triangular. But then

$$U^*(I+A)^{n-1}U = (I+T)^{n-1} = \begin{pmatrix} (1+\rho(A))^{n-1} & * & * & * \\ & (1+\rho(A))^{n-1} & * & * \\ & & * & * \\ 0 & & & \ddots \end{pmatrix}$$

So $(1+\rho(A))^{n-1}$ is of algebraic multiplicity at leats 2 in $(I+A)^{n-1}$, contradicting Perron's theorem. So $\rho(A)$ has algebraic multiplicity 1 in A.

3. We know $\rho(A)$ has algebraic multiplicity 1; so $\rho(A)$ is of geometric multplicity 1. So the $\rho(A)$ eigenspace is 1-dimensional. So $Aw = \rho(A)w$ implies $w = \alpha x$, where x is obtained as in (2). So there is a unique multiple to make the coordinates add to 1.

4. Well, $(I+A^T)^{n-1} = ((I+A)^{n-1})^T$. So A^T is irreducible. Applying (2) and (3) to A^T, we get that the dimension of the eigenspace for $\rho(A)$ is also 1 for A^T, and that $A^T\widetilde{y} = \rho(A)\widetilde{y}$ has one solution given by the Perron vector for A^T; so $0 \prec \widetilde{y}$ and any vector w satisfying $A^T w = \rho(A)w$ satisfies $w = \alpha\widetilde{y}$. Pick the unique α such that $(\alpha\widetilde{y}) \cdot x = 1$; let $y = \alpha\widetilde{y}$.

□ Theorem 273

9.2 Stochastic and doubly stochastic matrices

Definition 274. We say $A \in M_n$ is *(row-)stochastic* if $0 \preceq A$ and

$$\sum_{j=1}^{n} a_{ij} = 1$$

for all i. We say A is *column-stochastic* if $0 \preceq A$ and

$$\sum_{i=1}^{n} a_{ij} = 1$$

for all j. We say A is *doubly stochastic* if it is row-stochastic and column-stochastic.

Remark 275. Let $e = (1, \dots, 1)$. Then

1. A is row-stochastic if and only if $0 \preceq A$ and $Ae = e$.

2. A is column-stochastic if and only if $0 \preceq A$ and $e^T A = e^T$.

3. A is doubly stochastic if and only if $0 \preceq A$, $Ae = e$, and $e^T A = e^T$.

4. Suppose A_1, A_2 are stochastic (in one of the three senses). Suppose $t_1 + t_2 = 1$ with $0 \leq t_1$ and $0 \leq t_2$. Then $t_1 A_1 + t_2 A_2$ is stochastic in the same sense.

5. The set of stochastic matrices (in any of the three senses) is convex and compact.

6. Every permutation matrix P_σ is doubly stochastic.

Lemma 276. *Suppose $A = (a_{ij}) \in M_n$ is doubly stochastic with $A \neq I$. Then there is a permutation $\sigma\colon \{1, \dots, n\} \to \{1, \dots, n\}$ such that $a_{1,\sigma(1)} \dots a_{n,\sigma(n)} > 0$.*

Proof. Suppose otherwise. Let $(b_{i,j}) = tI - A$. Then

$$\begin{aligned} p_A(t) &= \det((b_{ij})) \\ &= \sum_\sigma \operatorname{sgn}(\sigma) \prod_{i=1}^n b_{i,\sigma(i)} \\ &= (t - a_{11}) \dots (t - a_{nn}) + \sum_{\sigma \neq \mathrm{id}} \operatorname{sgn}(\sigma) \prod_{i=1}^n b_{i,\sigma(i)} \end{aligned}$$

For fixed σ, we have

$$\prod_{i=1}^n b_{i,\sigma(i)} = \prod_{\sigma(i)=i} (t - a_{ii}) \cdot \prod_{\sigma(i) \neq i}$$

which is 0, since we may preapply a permutation to A so that the diagonal entries of A are non-zero.

Then $p_A(t) = (t - a_{11}) \dots (t - a_{nn})$. But A is stochastic; so 1 is an eigenvalue. So there is i such that $a_{ii} = 1$; say $a_{11} = 1$. Then

$$A = \begin{pmatrix} 1 & 0 \\ 0 & A_1 \end{pmatrix}$$

Then A_1 is doubly stochastic. Similarly, we find i with $2 \leq i \leq n$ such that $a_{ii} = 1$. Iterating, we find $A = I$.

□ Lemma 276

Lemma 277. *Suppose $A = (a_{ij}) \in M_n$ is doubly stochastic. If at most n entries of A are non-zero, then A is a permutation matrix.*

Proof. If $A = I$, then we are done. Suppose then that $A \neq I$. Then there is σ such that $a_{1,\sigma(1)} \dots a_{n,\sigma(n)} > 0$. This is already n non-zero entries. So the only non-zero entries of A are $a_{i,\sigma(i)}$. So A has exactly one non-zero entry in each row and column. So $a_{i,\sigma(i)} = 1$ for all i, and A is a permutation matrix. □ Lemma 277

Theorem 278 (Birkhoff)**.** *Suppose $A \in M_n$. Then A is doubly stochastic if and only if there are permutations $P_1, \dots, P_N$ and $t_i \geq 0$ with $t_1 + \dots + t_N = 1$ such that $A = t_1 P_1 + \dots + t_N P_N$. Moreover, one may take $N \leq n^2 - n + 1$.*

Proof.

($\Longleftarrow$) Clear.

($\Longrightarrow$) If $A = I$, then we are done. Suppose then that $A \neq I$. Pick σ such that $a_{1,\sigma(1)} \dots a_{n,\sigma(n)} \neq 0$. Let $\alpha_1 = \min\{ a_{1,\sigma(1)}, \dots, a_{n,\sigma(n)} \}$. Then

$$A_1 = \frac{1}{1 - \alpha_1} A - \alpha_1 P_{\sigma_1}$$

is doubly stochastic with $A = \alpha_1 P_{\sigma_1} + (1 - \alpha_1) A_1$, and A_1 is 0 in at least 1 entry. If $A_1 = I$, then we are done. Else there is σ_2 such that $a'_{1,\sigma_2(1)} \dots a'_{n,\sigma_2(n)} > 0$. Then $A_1 - \alpha_2 P_{\sigma_2} \succeq 0$. Then

$$A_2 = \frac{1}{1 - \alpha_2} (A_1 - \alpha_2 P_{\sigma_2})$$

is doubly stochastic with $A_1 = \alpha_2 P_{\sigma_2} + (1 - \alpha_2) A_2$; so

$$A = \alpha_1 P_{\sigma_1} + \alpha_2 (1 - \alpha_1) P_{\sigma_2} + (1 - \alpha_2)(1 - \alpha_1) A_2$$

Also A_2 is 0 in at least 2 entries. Continuing N times, we find

$$A = \alpha_1 P_{\sigma_1} + \dots + \alpha_{N-1} P_{\sigma_{N-1}} + \beta A_{N-1}$$

Then A_{N-1} is doubly stochastic and has $N - 1$ entries of 0. If $N - 1 \geq n^2 - n$, then $N \geq n^2 - n + 1$, so A_{N-1} is a permutation matrix, and we're done.

□ Theorem 278

Theorem 279 (Birkhoff)**.** *The premutation matrices are the extreme points of the convex set of doubly stochastic matrices.*

Remark 280. Closed convex sets in finite dimensions are always the convex hull of their extreme points.

Proof of Theorem 279.

($\Longrightarrow$) Suppose P is a permutation matrix; we wish to prove that P is an extreme point. Suppose $P = tA + (1-t)B$ for $0 < t < 1$ with A, B doubly stochastic. Then $p_{ij} = ta_{ij} + (1-t)b_{ij}$ with $0 \preceq A$ and $0 \preceq B$. So if $p_{ij} = 0$, then $a_{ij} = b_{ij} = 0$. So A and B are non-zero for at most 1 entry in each row and column. Hence $A = B = P$.

($\Longleftarrow$) Suppose A is doubly stochastic and not a permutation matrix. We wish to show that A is not an extreme point; i.e. there are A_+ and A_- such that $A = \frac{1}{2}A_+ + \frac{1}{2}A_-$ with $A_+ \neq A$ and $A_- \neq A$. Since A is not a permutation matrix, there is some row i_1 such that A is non-zero in 2 entries of the i_1 row; say $a_{i_1,j_1} \neq 0$ and $a_{i_1,j_2} \neq 0$; then $0 < a_{i_1,j_1} < 1$ and $0 < a_{i_1,j_2} < 1$. Now, in the j_2 column there must be another non-zero entry; say $0 < a_{i_2,j_2} < 1$. Continue until we return to an entry in column j_1; we get $(i_1, j_1), (i_1, j_2), (i_2, j_2), \ldots, (i_k, j_1)$, with all of these entries in $(0, 1)$. Let $B = E_{(i_1,j_1)} - E_{(i_1,j_2)} + E_{(i_2,j_2)} - \ldots$.

Claim 281. *Each row and column of B sums to* 0*.*

Proof. Picture. □ Claim 281

Let $\alpha = \min\{\, a_{i_1,j_1}, \ldots, a_{i_k,j_1} \,\} > 0$. Let

$$A_1 = A - \alpha B$$
$$A_2 = A + \alpha B$$

Then each row and column of A_1, A_2 still sums to 1. By our choice of α we have $0 \preceq A$ and $0 \preceq A_2$. So A_1 and A_2 are doubly stochastic. But $A = \frac{1}{2}A_1 + \frac{1}{2}A_2$. So A is not extreme.

□ Theorem 279

Corollary 282. *Suppose $A \neq I$ is doubly stochastic. Then there is a permutation $\sigma \neq \mathrm{id}$ such that* $a_{1,\sigma(1)} \ldots a_{n,\sigma(n)} > 0$.

Proof. By the theorem and the remark, we have

$$A = t_1 P_{\sigma_1} + \cdots + t_m P_{\sigma_m}$$

where each $t_i > 0$ and $t_1 + \cdots + t_m = 1$. Since $A \neq I$ there is at least one $\sigma_{i_0} \neq \mathrm{id}$. So $a_{i,\sigma_{i_0}(i)} \geq t_{i_0} \cdot 1 > 0$. So $a_{1,\sigma_{i_0}(1)} \ldots a_{n,\sigma_{i_0}(n)} > 0$. □ Corollary 282

10 Matroids

Definition 283 (H. Whitney (1935))**.** Suppose X is a (finite) set. A *matroid* on X is a $\mathcal{I} \subseteq \mathcal{P}(X)$ satisfying

1. $\emptyset \in \mathcal{I}$.
2. If $B \in \mathcal{I}$ and $A \subseteq B$ then $A \in \mathcal{I}$.
3. If $A, B \in \mathcal{I}$ and $|A| < |B|$ then there is $b \in B \setminus A$ such that $A \cup \{\, b \,\} \in \mathcal{I}$.

Example 284. Suppose V is a vector space with $X \subseteq V$. Given $A \subseteq X$, we define $A \in \mathcal{I}$ if A is linearly independent. Then $\mathcal{I}$ is a matroid on X. (1) and (2) are clear; to see (3), note that

$$\dim(\mathrm{span}(A)) = |A| < |B| = \dim(\mathrm{span}(B))$$

So there is $b \in B$ such that $b \notin \mathrm{span}(A)$, in which case $A \cup \{ b \}$ is linearly independent .

Definition 285. Suppose $\mathcal{I}$ is a matroid on X; suppose $E \subseteq X$. We define $\mathrm{rank}(E) = \max\{ |A| : A \subseteq E, A \in \mathcal{I} \}$.

Example 286. Continuing the previous example, we have $\dim(\mathrm{span}(E))$ is the maximum cardinality of a linearly independent subset of E, which is just $\mathrm{rank}(E)$.

Theorem 287 (Horn 1955, Rado 1962)**.** *Suppose V is a vector space with $X \subseteq V$. Then X can be partitioned into k linearly indepdendent subsets if and only if for all finite $E \subseteq X$ we have*

$$\frac{|E|}{\dim(\mathrm{span}(E))} \leq k$$

Much later, the same theorem was proven by Rado and collaborators for arbitrary matroids, where $\dim(\mathrm{span}(E))$ is replaced by $\mathrm{rank}(E)$.

Proof.

($\implies$) Suppose $X = A_1 \cup \cdots \cup A_k$ where the A_i are pairwise disjoint and linearly independent. Suppose $E \subseteq X$. Then

$$E = (E \cap A_1) \cup \cdots \cup (E \cap A_k)$$

where the $E \cap A_i$ are linearly independent. So $\dim(\mathrm{span}(E \cap A_i)) = |E \cap A_i|$. So

$$\begin{aligned} |E| &= \sum_{i=1}^{k} |E \cap A_i| \\ &= \sum_{i=1}^{k} \dim(\mathrm{span}(E \cap A_i)) \\ &\leq \sum_{i=1}^{k} \dim(\mathrm{span}(E)) \\ &= k \dim(\mathrm{span}(E)) \end{aligned}$$

So

$$\frac{|E|}{\dim(\mathrm{span}(E))} \leq k$$

□ Theorem 287

TODO 1. *Last lecture.*