MATRIX ANALYSIS, FALL 2015

VERN I. PAULSEN

Abstract. The lectures follow the text of Horn and Johnson fairly closely. These notes accompany the lectures and show the results that we have covered in class–without proofs.

1. Basics

Assume you know: Fields, vector spaces, bases, linear independence, linear map, matrix of a linear map, determinants, how to compute the inverse of a matrix, from analysis–sequences and series, continuity and derivatives.

The only fields we will consider is either $\mathbb{R}$ or $\mathbb{C}$, when I want to indicate either I'll write $\mathbb{F}$.

$\mathbb{F}^m$ will denote the vector space(over $\mathbb{F}$) of m-tuples. So $v \in \mathbb{F}^m \iff v = (x_1, ..., x_m)$ where $x_i \in \mathbb{F}, \forall 1 \le i \le m$.

The **canonical basis** for $\mathbb{F}^m$ will mean the basis $\{e_1, ..., e_m\}$ where e_i is the vector that is 1 in the i-th entry and 0 in all other entries. So $v = (x_1, ..., x_m) = \sum_{i=1}^m x_i e_i$.

Given vectors $v = (x_1, ..., x_m), w = (y_1, ..., y_m) \in \mathbb{F}^m$ their **dot product** is $v \cdot w = \sum_{i=1}^m x_i y_i$.

We let $M_{m,n}$ denote the vector space over $\mathbb{F}$ consisting of all $m \times n$ matrices, where $m =$ the number or rows, $n =$ number of columns.

We will occasionally write $M_{m,n}(\mathbb{R})$ or $M_{m,n}(\mathbb{C})$ when we want to indicate which underlying field we are allowing the entries of the matrix to belong to. Given $A \in M_{m,n}$ we write

$$A = (a_{i,j}) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}$$

where $a_{i,j} \in \mathbb{F}$ indicates the number in the (i,j)-entry.

There is a canonical basis for the vector space $M_{m,n}$, called the **matrix units**, where $E_{i,j}$ is the $m \times n$ matrix that is 1 in the (i,j)-entry and 0 isn all other entries. Thus, $(a_{i,j}) = \sum_{i=1}^m \sum_{j=1}^n a_{i,j} E_{i,j}$.

Each matrix $A \in M_{m,m}$ defines a linear map $L_A : \mathbb{F}^n \to \mathbb{F}^m$ given by $L_A((x_1, ..., x_m)) = (\sum_{j=1}^n a_{1,j} x_j, \ldots, \sum_{j=1}^n a_{m,j} x_j)$.

Given $A = (a_{i,j}) \in M_{m,n}$ the **conjugate** of A is the matrix denoted $\overline{A} = (b_{i,j}) \in M_{m,n}$ where $b_{i,j} = \overline{a_{i,j}}$. The **transpose** of A is the matrix denoted $A^t = (b_{i,j}) \in M_{n,m}$ where $b_{i,j} = a_{j,i}$. The **adjoint** is the matrix $A^* = (b_{i,j}) \in M_{n,m}$ where $b_{i,j} = \overline{a_{j,i}}$. Thus, $A^* = \overline{A^t} = (\overline{A})^t$.

2. Matrix Product

Given $A = (a_{i,j}) \in M_{m,p}$ and $B = (b_{i,j}) \in M_{p,n}$ their **product** is the matrix $A \cdot B = (c_{i,j}) \in M_{m,n}$ with $c_{i,j} = \sum_{k=1}^{p} a_{i,k}b_{k,j}$. The matrix product has the following properties:

Associative: Given $A \in M_{m,p}, B \in M_{p,n}, C \in M_{n,q}$ we have that $(A \cdot B) \cdot C = A \cdot (B \cdot C)$.

Bilinear: Given $A_1, A_2 \in M_{m,p}, B_1, B_2 \in M_{p,n}$ and $\lambda \in \mathbb{F}$ we have that $(A_1 + A_2) \cdot (B_1 + B_2) = A_1B_1 + A_1B_2 + A_2B_1 + A_2B_2$ and $\lambda(AB) = (\lambda A)B = A(\lambda B)$.

$(AB)^t = B^tA^t$, $\overline{(AB)} = (\overline{A})(\overline{B})$ and $(AB)^* = B^*A^*$.

Note: Most books write a vector in $\mathbb{F}^n$ as a row, but it is better to identify $\mathbb{F}^n \sim M_{n,1}$ and $\mathbb{F}^m = M_{m,1}$ since then $v = (x_1, ..., x_n)^t \in M_{n,1}$ and $L_A : M_{n,1} \to M_{m,1}$ satisfies $L_A(v) = A \cdot v$–the matrix product. For this reason the linear map L_A is really left multiplication by the matrix A, after we make these identifications.

There are several good ways to think of matrix multiplication. First, if we write $A \in M_{m,n}$ in terms of the columns as $A = [C_1 \vdots \dots \vdots C_n]$ where each $C_j \in M_{m,1}$ then $L_A((x_1, ..., x_n)) = x_1C_1 + \cdots x_nC_n$.

This is useful for thinking about **ranges**. Given vector spaces V, W and a linear map, $L : V \to W$ then the **range of L**, is the subspace of W,

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

Thus, we see that $\mathcal{R}(L_A) = span\{C_1, ..., C_n\}$.

Now, if $A \in M_{m,p}$ and $B = [B_1 \vdots \dots \vdots B_n] \in M_{p,n}$ is written in terms of columns, then $A \cdot B = [AB_1 \vdots \dots \vdots AB_n]$. Thus, we see that each column of $A \cdot B$ is a vector in the range of L_A.

Alternatively, if we write $A = \begin{bmatrix} R_1 \\ \dots \\ \vdots \\ \dots \\ R_m \end{bmatrix}$ in terms of its row vectors, and $B = [B_1 \vdots \dots B_m]$ in terms of its columns, then $A \cdot B = (R_i \cdot B_j)$.

Finally, if we write $A = [V_1 \vdots \dots V_p] \in M_{m,p}$ so that $V_k = (a_{1,k}, ..., a_{m,k})^t$ and $B = \begin{bmatrix} W_1 \\ \dots \\ \vdots \\ \dots W_p \end{bmatrix} \in M_{p,n}$ in terms of its rows, so that $W_k = (b_{k,1}, ..., b_{k,n})$ then we have that

$$A \cdot B = (\sum_{k=1}^{p} a_{i,k} b_{k,j}) = \sum_{k=1}^{p} (a_{i,k} b_{k,p}) = \sum_{k=1}^{p} V_k \cdot W_k,$$

since the product of the $m \times 1$ matrix V_k with the $1 \times n$ matrix W_k is the $m \times n$ matrix with (i, j)-entry, $(a_{i,k} b_{k,j})$.

3. Determinants

We assume that you are familiar with determinants and only review a few key facts.

Let $A \in M_n$, then $det(A) \in \mathbb{F}$. There are several formulas for obtaining this number.

3.1. **Laplace Expansion:** Given $A = (a_{i,j}) \in M_n$ we let $A_{i,j} \in M_{n-1}$ be the matrix obtained from A by deleting the i-th row and j-th column. Then, inductively,

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

These two formulas are called the Laplace expansion along the i-row and j-th column, respectively.

3.2. **Permutations:** A map $\sigma : \{1, ..., n\} \to \{1, ..., n\}$ is one-to-one if and only if it is onto. Such a map is called a *permutation*. The set of permutations, with the operation of function composition, forms a group called the *symmetric group on n elements* and denoted S_n. The group S_n has $n!$ elements. A permutation is called a *transposition* if it interchanges two elements and leaves the remaining elements fixed. Every permutation can be written as a composition of transpositions. If σ is written two different ways as a composition of k_1 and k_2 transpositions, then $(-1)^{k_1} = (-1)^{k_2}$ and this number is called the *sign* of the permutation and we set $sgn(\sigma) = (-1)^{k_1}$.

The other formula for the determinant is that

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

The **permanent** of a matrix is the sum

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

It has applications to order statistics and to symmetric products of Hilbert spaces. (Some will understand this.)In particular, if H is a Hilbert space, $H^{\otimes n}$ denotes the tensor product of H with itself n times and P denotes the projection onto the subspace of symmetric tensors, then for any $x_1, ...x_n, y_1, ..., y_n \in H$,

$$\langle P(x_1 \otimes \cdots \otimes x_n), P(y_1 \otimes \cdots \otimes y_n) \rangle = perm((\langle x_i, y_j \rangle)).$$

Some key facts:

(1) $det(A) = det(A^t)$,
(2) $det(AB) = det(A)det(B)$,
(3) A invertible iff $det(A) \neq 0$ and $det(A^{-1}) = det(A)^{-1}$,

3.3. **Multilinear maps and the abstract characterization of determinant:** Given vector spaces $V_1, ..., V_n$ and W a map $L : V_1 \times V_2 \times \cdots V_n \to W$ from the Cartesian product of the V's to W is called **multilinear** provided that:

i) if $1 \leq j \leq n$ and $v_j, \tilde{v}_j \in V_j$ then

$$L(v_1, ..., v_{j-1}, v_j + \tilde{v}_j, v_{j+1}, ..., v_n) = L(v_1, ..., v_{j-1}, v_j, v_{j+1}, ..., v_n) + L(v_1, ..., v_{j-1}\tilde{v}_j, v_{j+1}, ..., v_n),$$

ii) $L(v_1, ..., v_{j-1}, \lambda v_j, v_{j+1}, ..., v_n) = \lambda L(v_1, ..., v_{j-1}, v_j, v_{j+1}, .., v_n)$.

Given $A \in M_n$ write $A = (v_1 \vdots ... \vdots v_n)$ in terms of its columns, so that we may think of $A \in \mathbb{F}^n \times \cdots \times \mathbb{F}^n$,

Theorem 3.1. *$det : M_n \to \mathbb{F}$ is a multilinear map of the columns, it is alternating, i.e., if we transpose two columns then that changes the det by a negative sign, and $det(I_n) = 1$. Moreover, det is the unique multilinear map with these two properties.*

3.4. **Cramer's Rule and the Adjugate:** Let $A \in M_n$ and let $A_{i,j} \in M_{n-1}$ be as before. Set $b_{i,j} = (-1)^{i+j} det(A_{j,i})$. Then the matrix $B = (b_{i,j})$ is called the **adjugate** of A.

Theorem 3.2 (Cramer's Rule)**.** *Let $A \in M_n$ and let B be the adjugate of A. Then $AB = BA = det(A)I_n$. If $det(A) \neq 0$, then $A^{-1} = \frac{1}{det(A)} B$.*

4. Row Reduced Echelon Form(RREF)

Definition 4.1. Let $B \in M_{m,n}$ we say that B is in *RREF* if:

(1) The 1st non-zero of each row of B is a 1. These are called the *leading one's.*

(2) The 1st non-zero entry of the $(i+1)$-st row is strictly to the right of the 1st non-zero entry of the i-th row.
(3) All the other entries in the column of a leading 1 are 0.
(4) All rows with all 0's are at the bottom.

A matrix that only satisfies 1,2,4 is said to be in *row echelon form(REF)*.

4.1. **The Elementary Operations:** There are three types of elementary operations.

Type I: Row interchange, interchange row k with row l.

Type II: Multiply row k by a non-zero constant.

Type III: Add a multiple of row i to row k.

The following theorem can be found in most courses in linear algebra, see for example Hoffman-Kunze.

Theorem 4.2 (RREF)**.** *Given $A \in M_{m,n}$, one can perform a sequence of elementary operations on A to obtain a matrix B that is in RREF. Moreover, if B and B' are both in RREF and obtained from A by performing a sequence of elementary operations, then $B = B'$, i.e., B is unique.*

We sketch the proof of uniqueness. Use the fact that every elementary operation is reversible. So that one can get from B back to A and then to B'. Thus, one could get from B to B' via elementary operations. But since a leading 1 is the only non-zero entry in its column, it can be seen that if the leading 1 in the 1st row for B occurs in the j_1 entry, then the leading 1 for B' must also occur in the j_1 entry. Similarly for the next leading 1, etc. Thus, the leading 1's for B and B' must all occur in the same positions. Next check that if we need to preserve the 1st leading 1 in B then we can not change any of the other entries in that row. So the 1st row of B and B' must be identical.

Remark 4.3. The process of getting from A to a B in RREF is called **Gauss-Jordan elimination**. The process of getting from A to a C in REF is called **Gaussian elimination**.

4.2. **Elementary Matrices:** There are also three types of elementary matrices, corresponding to each elementary operation.

Type I: For $k \neq l$ set $U(k,l) = E_{k,l} + E_{l,k} + \sum_{i:i\neq k, i\neq l} E_{i,i} \in M_m$. Then $U(k,l)^{-1} = U_{k,l}$ and $U(k,l)A$ is the matrix obtained from A by interchanging rows k and l.

Type II: For $\lambda \neq 0$ set $D(k;\lambda) = \lambda E_{k,k} + \sum_{i\neq k} E_{i,i}$. Then $D(k;\lambda)^{-1} = D(k;\lambda^{-1})$ and $D(k;\lambda)A$ is the matrix obtained from A by multiplying the k-th row by λ.

Type III: For $k \neq l$ and $\lambda \in \mathbb{F}$ set $S(k,l;\lambda) = I_m + \lambda E_{l,k}$. Then $S(k,l;\lambda)^{-1} = S(k,l;-\lambda)$ and $S(k,l;\lambda)A$ is the matrix obtained from A by adding λ times row k to row l.

Proposition 4.4. *If B is the RREF of A then $B = WA$ where W is a product of elementary matrices.*

Since each elementary matrix is invertible, this is another way to see that each elementary operation is reversible.

5. Rank

Definition 5.1. Let $A \in M_{m,n}$. Then the **column rank of A** is the dimension of the subspace of $\mathbb{F}^m$ spanned by the columns of A, i.e., the largest number of linearly independent columns. It is denoted $rank_c(A)$.

The **row rank of A** is the dimension of the subspace of $\mathbb{F}^n$ spanned by the rows of A and is denoted $rank_r(A)$.

First we characterize the column rank.

Definition 5.2. Let V, W be vector spaces and let $L : V \to W$ be linear. Then the **range of L** is

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

Note that as sets $\mathcal{R}(L) = L(V)$ and that $\mathcal{R}(L)$ is a vector subspace of W.

Proposition 5.3. *Let $A \in M_{m,n}$ so that $L_A : \mathbb{F}^n \to \mathbb{F}^m$. Then $rank_c(A) = dim(\mathcal{R}(L_A))$.*

Proof. Let $C_1, ..., C_n \in \mathbb{F}^m$ denote the columns of A. Then $L_A((\lambda_1, ..., \lambda_n)) = \sum_{j=1}^n \lambda_j C_j$. Thus, $\mathcal{R}(L_A) = span\{C_1, ..., C_n\}$ and the result follows. □

Theorem 5.4. *Let $A \in M_{m,n}$ and let B be the RREF of A. Then:*

1. $rank_r(A) = rank_r(B)$,
2. $rank_c(A) = rank_c(B)$,
3. $rank_c(B) =$ *the number of rows with leading 1's* $= rank_r(B)$.

Corollary 5.5. *Let $A \in M_{m,n}$. Then $rank_c(A) = rank_r(A)$.*

Definition 5.6. We call this common value, $rank_r(A) = rank_c(A)$ the **rank of A** and denote it by $rank(A)$.

6. Submatrices

Let $A \in M_{m,n}$, let $\alpha = \{\alpha_1, ..., \alpha_k\}$ with $1 \le \alpha_1 < \cdots < \alpha_k \le m$ and let $\beta = \{\beta_1, ..., \beta_l\}$ with $1 \le \beta_1 < \cdots < \beta_l \le n$. We set $A[\alpha, \beta] = (a_{\alpha_i, \beta_j}) \in M_{k.l}$. Such a matrix is called a **submatrix of A**.

A submatrix of the form $A[\alpha, \alpha]$ is called a **principal submatrix of A**.

When $k = l$, so that $A[\alpha, \beta]$ is a square matrix, then we call the number $det(A[\alpha, \beta])$ a **minor of A**. We call $det(A[\alpha, \alpha])$ a **principal minor of A**.

When $\alpha = \{1, ..., k\}$ then we call $A[\alpha, \alpha]$ a **leading principal submatrix of A** and we call $det(A[\alpha, \alpha])$ a **leading principal minor of A**.

When $n = m$ and $\alpha = \{k, k+1, ..., n\}$ then we call $A[\alpha, \alpha]$ a **trailing principal submatrix of A** and we call $det(A[\alpha, \alpha])$ a **trailing principal minor of A**.

A matrix with the property that all of its minors are positive is called **totally positive**.

7. Sums of Spaces

Let V and W be vector spaces over $\mathbb{F}$, on their Cartesian product $V \times W$ define an addition by $(v_1, w_1) = (v_2, w_2) = (v_1 + v_2, w_1 + w_2)$ and scalar multiplication by $\lambda \in \mathbb{F}$ set $\lambda(v, w) = \lambda v, \lambda w)$. Then these operations make $V \times W$ into a vector space called their *direct sum* and denoted $V \oplus W$.

Proposition 7.1. *If $\{v_\alpha : \alpha \in A\}$ is a basis for V and $\{w_\beta : \beta \in B\}$ is a basis for W, then $\{(v_\alpha, 0) : \alpha \in A\} \cup \{(0, w_\beta) : \beta \in B\}$ is a basis for $V \oplus W$. Conseqeuntly, $dim(V \oplus W) = card(A \dot{\cup} B) = card(A) + card(B)$.*

This generalizes to the direct sum of more than two spaces.

When $V = \mathbb{F}^m$ and $W = \mathbb{F}^n$, $v = (x_1, ..., x_m) \in V$ and $w = (y_1, ..., y_n) \in W$ then

$$(v, w) = \big((x_1, ..., x_m), (y_1, ...y_m)\big) \sim (x_1, ..., x_m, y_1, ..., y_n) \in F^{m+n}$$

and we see that $\mathbb{F}^m \oplus \mathbb{F}^n \sim \mathbb{F}^{m+n}$ and the isomorphism is essentially removing extra parentheses. Conversely, if $k = k_1 + k_2$ then we obtain an identification $\mathbb{F}^k \sim \mathbb{F}^{k_1} \oplus \mathbb{F}^{k_2}$ by inserting parentheses.

8. Partitioned Matrices

Suppose that $A = (a_{i,j}) \in M_{m_1+m_2,n_1+n_2}$ so that $L_A : \mathbb{F}^{n_1+n_2} \to \mathbb{F}^{m_1+m_2}$ and as above we may partition these vectors to define identifications with $L_A : \mathbb{F}^{n_1} \oplus \mathbb{F}^{n_2} \to \mathbb{F}^{m_1} \oplus \mathbb{F}^{m_2}$. When we, correspondingly, partition $A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$ with $A_{i,j} \in M_{m_i,n_j}$ then each of these matrices defines a linear map $L_{A_{i,j}} : \mathbb{F}^{n_j} \to \mathbb{F}^{m_i}$. If we write $z = (v, w) \in \mathbb{F}^{n_1} \oplus \mathbb{F}^{n_2}$ as a column vector, then we have that

$$Az = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = \begin{bmatrix} A_{1,1}v + A_{1,2}w \\ A_{2,1}v + A_{2,2}w \end{bmatrix}.$$

When I say that $A \in M_{m_1+\cdots+m_r,n_1+\cdots+n_r}$ is partitioned, I mean that I am writing $A = \big(A_{i,j}\big)$ where $A_{i,j} \in M_{m_i,n_j}$. If $B \in M_{n_1+\cdots+n_r,p_1+\cdots p_r}$ is also partitioned as $B = (B_{i,j}) \in M_{n_i.p_j}$ then it is easy to check that the product $A_{i,k}B_{k,j}$ is defined. Also $AB \in M_{m_1+\cdots+m_r,p_1+\cdots+p_r}$ is the block matrix

$$AB = \big(\sum_{k=1}^{r} A_{i,k}B_{k,j}\big).$$

A partitioned matrix $A = (A_{i,j})$ is **block diagonal** provided that $A_{i,j} = 0 \forall i \neq j$.

Proposition 8.1. *Let $A \in M_{n_1+\cdots+n_k,n_1+\cdots+n_k}$ be block diagonal. Then $det(A) = \prod_{i=1}^{k} det(A_{i,i})$.*

For a matrix of numbers $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ if $a \neq 0$, then $det(A) = a(d - ba^{-1}c)$.

Proposition 8.2. *Let $A \in M_{n_1+n_2,n_1+n_2}$ be partitioned as $A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}$. If $det(A_{1,1}) \neq 0$, then*

$$det(A) = det(A_{1,1})det(A_{2,2} - A_{2,1}A_{1,1}^{-1}A_{1,2}).$$

If $det(A_{2,2}) \neq 0$, then

$$det(A) = det(A_{2,2})det(A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1}).$$

9. Euclidean Norm and Inner product

If $x, y \in \mathbb{F}^n \sim M_{n,1}$ then their **inner product** is

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i\overline{y_i} = y^*x.$$

In the case that $\mathbb{F} = \mathbb{R}$, the inner product is a bilinear map, $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.

In the case that $\mathbb{F} = \mathbb{C}$, the inner product is a **sesquilinear**(which means one and a half linear) since $\langle x, \lambda y \rangle = \overline{\lambda} \langle x, y \rangle$.

The **Euclidean norm of a vector** is $\|x\|_2 = \langle x, x \rangle^{1/2}$. Vectors x, y are called **orthogonal**(written $x \perp y$) provided $\langle x, y \rangle = 0$. Note that every vector is orthogonal to the 0 vector. A set of vectors S is called an **orthogonal set** if $x \perp y$ for every $x, y \in S$ with $x \neq y$. A set of vectors S is called **orthonormal** if it is orthogonal and for every $x \in S$, we have $\|x\|_2 = 1$.

Proposition 9.1 (Cauchy-Schwartz)**.** *For $x, y \in \mathbb{F}^n$, $|\langle x, y \rangle| \leq \|x\|_2 \|y\|_2$.*

Proposition 9.2. *For $x, y \in \mathbb{F}^n$, $\lambda \in \mathbb{F}$ we have that $\|\lambda x\|_2 = |\lambda| \|x\|_2$ and $\|x + y\|_2 \leq \|x\|_2 + \|y\|_2$(**the triangle inequality**).*

9.1. **Gram-Schmidt Orthonormalisation:** Given $\{x_1, ..., x_n\}$ linearly independent, this produces an orthonormal set $\{u_1, ..., u_n\}$ with the property that for every $1 \leq k \leq n$, $span\{u_1, ..., u_k\} = span\{x_1, ..., x_k\}$. The vectors $u_1, ...u_n$ are defined inductively. We described this process in class.

9.2. **Lowden Orthonormalisation:** This is a different way to produce an orthonormal set $\{v_1, ..., v_n\}$ from a linearly independent set $\{x_1, ..., x_n\}$. We will do this later in the course. It produces the unique orthonormal set that minimizes:

$$\sum_{k=1}^{n} \|x_k - v_k\|_2^2.$$

10. Eigenvalues, eigenvectors and spectrum

Proposition 10.1. *Let $S \in M_n$. The following are equivalent.*

1. *$\exists\, T \in M_n$ such that $ST = TS = I_n$,*
2. *$det(S) \neq 0$,*
3. *$L_S : \mathbb{F}^n \to \mathbb{F}^n$ is one-to-one,*
4. *L_S is onto.*

Definition 10.2. We call $S \in M_n$ satisfying any of the above equivalent conditions **non-singular** or **invertible** and we let M_n^{-1} denote the set of invertible matrices in M_n. If S fails to satisfy these conditions, then we call S **singular** or **non-invertible**.

For the rest of this section we assume that $\mathbb{F} = \mathbb{C}$.

Definition 10.3. Given $A \in M_n$, if $x \in \mathbb{C}^n$, $x \neq 0$ and there exists $\lambda \in \mathbb{C}$ such that $Ax = \lambda x$ then we call λ an **eigenvalue of A** and we call X an **eigenvector for A**. The **spectrum of A** is the set denoted $\sigma(A)$ of all eignevalues of A. The **spectral radius of A** is the number $\rho(A) = sup\{|\lambda| : \lambda \in \sigma(A)\}$.

Proposition 10.4. *Let $A \in M_n$ then $\sigma(A) = \{\lambda : \lambda I_n - A$ is non-invertible$\}$.*

Definition 10.5. Let $A \in M_n$ and let $p(t) = a_k t^k + \cdots a_1 t + a_0$ be a polynomial with complex coeffients. Then we set $p(A) = a_k A^k + \cdots a_1 A + a_0 I_n \in M_n$

It is easy to check that if $p(t) = a_k t^k + \cdots + a_0$ and $q(t) = b_m t^m + \cdots + b_0$ are two polynomials and $r(t) = p(t)q(t) = a_k b_m t^{k+m} + \cdots + a_0 b_0$ denotes their product, then $r(A) = p(A)q(A) = q(A)p(A)$ is the product of the matrices $p(A)$ and $q(A)$.

Theorem 10.6 (Spectral Mapping Theorem)**.** *Let $A \in M_n$ and let p be a polynomial. Then $\sigma(p(A)) = \{p(\lambda) : \lambda \in \sigma(A)\}$.*

Definition 10.7. Let $A \in M_n$ then the polynomial $p_A(t) = det(tI_n - A)$ is called the **characteristic polynomial of A**.

Note that $p_A(t)$ is a monic polynomial of degree n.

Proposition 10.8. $\lambda \in \sigma(A) \iff p_A(\lambda) = 0$.

Definition 10.9. Given a linear map $L : V \to W$ the **null space** or **kernel** of L, denoted $\mathcal{N}(L)$ is the subspace $\{x \in V : L(x) = 0\}$. If $B \in M_{m,n}$ then by the null space of B, denoted $\mathcal{N}(B)$ we mean the null space of the map L_B.

Definition 10.10. Let $A \in M_n$ and let $\lambda \in \sigma(A)$. Then the set $\{x \in \mathbb{C}^n : Ax = \lambda x\} = \mathcal{N}(\lambda I_n - A)$ is called the **eigenspace** corresponding to λ. We call the dimension of $\mathcal{N}(\lambda I_n - A)$ the **geometric multiplicity of** λ. We call the number of factors $(t - \lambda)$ that occur in $p_A(t)$ the **algebraic multiplicity of** λ.

Example 10.11. Let $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$ then the geometric multiplicity of λ is 1, and the algebraic multiplicity of λ is 2.

Proposition 10.12. *Let $A \in M_n$ and let $\lambda \in \sigma(A)$. Then the geometric multiplicity of λ is less than or equal to the algebraic multiplicity.*

10.1. **The Elementary Symmetric Functions.**

Definition 10.13. Given an n-tuple of complex numbers, $(\lambda_1, ..., \lambda_n)$, for $1 \leq k \leq n$ we set

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

This is called the **k-th symmetric function**.

In particular,

$$S_1(z_1, ..., z_n) = z_1 + \cdots + z_n,$$

while,

$$\begin{aligned} S_2(z_1, ..., z_n) = z_1 z_2 \quad + \quad & \cdots + z_1 z_n \\ +z_2 z_3 \quad & + \cdots + z_2 z_n \\ & +z_3 z_4 + \cdots + z_3 z_n \\ & \qquad + \cdots \cdots + z_{n-1} z_n \end{aligned}$$

and $S_n(z_1, ..., z_n) = z_1 \cdot z_2 \cdots z_n$.

Note that

$$\begin{aligned} &(t - \lambda_1) \cdots (t_n - \lambda_n) = \\ &\qquad t^n - S_1(\lambda_1, ..., \lambda_n) t^{n-1} + S_2(\lambda_1, ..., \lambda_n) t^{n-2} + \\ &\qquad \cdots + (-1)^k S_k(\lambda_1, ..., \lambda_n) t^{n-k} + \cdots \\ &\qquad + (-1)^{n-1} S_{n-1}(\lambda_1, ..., \lambda_n) t^1 + (-1)^n S_n(\lambda_1, ..., \lambda_n). \end{aligned}$$

Definition 10.14. Let $A \in M_n$ for $1 \leq k \leq n$, we set

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

that is the sum of all $k \times k$ principal minors of A. In particular, we set $tr(A) = \sum_{j=1}^{n} a_{j,j} = E_1(A)$ and call this the **trace** of A.

Given $f : (a, b) \to \mathbb{C}$ we can write $f(t) = g(t) + ih(t)$, where $g, h : (a, b) \to \mathbb{R}$. We say that f is differentiable iff g and h are differentiable and set $f'(t) = g'(t) + ih'(t)$. More generally, if $\vec{f} = (f_1, ..., f_n) : (a, b) \to \mathbb{C}^n$ then we say that $\vec{f}$ is differentiable iff each f_j is differentiable and we set $\vec{f}'(t) = (f_1'(t), ..., f_n'(t))$

Theorem 10.15. *Let $\vec{f_j} = \left(f_{1,j}, ..., f_{n,j}\right)^t : (a, b) \to \mathbb{C}^n$ be differentiable and set $A(t) = [\vec{f_1} \vdots \ldots \vdots \vec{f_n}] : (a, b) \to M_n$. Then $g(t) = det(A(t))$ is* *differentiable and*

$$g'(t) = det\big([\vec{f_1}'\vdots\vec{f_2}\vdots\dots\vdots\vec{f_n}]\big) + det\big([\vec{f_1}\vdots\vec{f_2}'\vdots\dots\vdots\vec{f_n}]\big) + \dots + det\big([\vec{f_1}\vdots\vec{f_2}\vdots\dots\vdots\vec{f_n}']\big)$$

Theorem 10.16. *Let $A \in M_n$ and let $p_A(t) = (t-\lambda_1)\cdots(t-\lambda_n)$. Then for $1 \le k \le n$, we have that*

$$E_k(A) = S_k(\lambda_1, ..., \lambda_n).$$

10.2. **Moments and Newton's Identities.** Given $(\lambda_1, ..., \lambda_n)$ set

$$\mu_k = M_k(\lambda_1, ..., \lambda_n) = \lambda_1^k + \dots + \lambda_n^k.$$

It is easily seen that $\mu_1 = S_1(\lambda_1, ..., \lambda_n)$ and that $\mu_2 = S_1^2 - 2S_2$, $S_2 = 1/2(\mu_1^2 - \mu_2)$

More generally, **Newton's Identities** say that:

$$k(-1)^k S_k + \mu_1(-1)^{k-1}S_{k-1} + \dots + \mu_k = 0, \ \forall 1 \le k \le n.$$

10.3. **Right Multiplication.** Given $A \in M_{m,n}$ we saw that identifying $\mathbb{F}^k \sim M_{k,1}$ then left multiplication by A defines a linear map $L_A : \mathbb{F}^n \to \mathbb{F}^m$.

On the other hand we could identify spaces with rows, i.e., $\mathbb{F}^k \sim M_{1,k}$, in which case we could right multiply row vectors by A to define a linear map, $R_A : \mathbb{F}^m \to \mathbb{F}^n$. Note that for $y \in M_{1,m}$ we have that $(yA)^t = A^t y^t$ where $y^t \in M_{m,1}$. This shows that as linear maps, $R_A = L_{A^t}$.

Thus, if $A \in M_n$ then it really defines two maps, $L_A, R_A = L_{A^t} : \mathbb{F}^n \to \mathbb{F}^n$ and so it is natural to wonder how do eigenvalues, etc. behave for these two maps.

Theorem 10.17. *Let $A \in M_n$. Then*
(1) $p_A(t) = p^{A^t}(t)$,
(2) $\sigma(A) = \sigma(A^t)$,
(3) *for λ in this common spectrum the algebraic multiplicities for A and A^t are the same,*
(4) *for λ in this common spectrum, the geometric multiplicities for A and A^t are the same.*

11. Similarity

11.1. **Change of Basis.** Let $L : V \to V$ be linear, $dim(V) = n$ and let $\mathcal{B} = \{v_1, ..., v_n\}$ be a basis for V. Then we can write $L(v_j) = \sum_{i=1}^n b_{i,j} v_i$

and the matrix $B = (b_{i,j})$ is called the **matrix of L with respect to** $\mathcal{B}$ and it is denoted $B = mat_{\mathcal{B}}(L)$.

If we let $S : \mathbb{F}^n \to V$ be defined by $S(e_j) = v_j$ and let $L_B : \mathbb{F}^n \to \mathbb{F}^n$ then we have that $L = SL_BS^{-1}$.

Now suppose that to begin with $V = \mathbb{F}^n$ so that S is the matrix whose columns are the vectors v_j and we let $L = L_A$, then we have that

$$mat_{\mathcal{B}}(L_A) = B = S^{-1}AS.$$

So we see that when we form $B = S^{-1}AS$ we are just considering the same linear map but represented in a different basis.

Definition 11.1. Let $A, B \in M_n$ the we say that **B is similar to A** and write $B \sim A$ if there exists $S \in M_n^{-1}$ such that $B = S^{-1}AS$.

Proposition 11.2. *Similarity is an equivalence relation, i.e.,* $A \sim A$, $B \sim A \implies A \sim B$ *and* $B \sim A, C \sim B \implies C \sim A$.

Proposition 11.3. *If* $B \sim A$ *then:*

(1) $P_A(t) = p_B(t)$,
(2) $\sigma(A) = \sigma(B)$,
(3) *for each* $\lambda \in \sigma(A) = \sigma(B)$ *its algebraic multiplicity is the same for* A *and* B,
(4) *for each* $\lambda \in \sigma(A) = \sigma(B)$ *its geometric multiplicity is the same for* A *and* B,
(5) $E_k(A) = E_k(B)$ *for all* k.

Example 11.4. These properties do not characterize similarity. Let $N_k = (c_{i,j}) \in M_k$ with $c_{i,i+1} = 1$ and $c_{i,j} = 0, \forall j \neq i+1$, and set $N_1 = 0$. Let $A, B \in M_7$ be defined by $A = N_2 \oplus N_2 \oplus N_3$ and $B = N_1 \oplus N_3 \oplus N_3$. Then $p_A(t) = p_B(t) = t^7$, $\sigma(A) = \sigma(B) = \{0\}$, the algebraic multiplicity of 0 is 7 for both matrices and the geometric multiplicity is 3 for both matrices. Also $E_k(A) = E_k(B) = 0$ for ll k. But $A \nsim B$, since $dim(\mathcal{N}(A^2)) = 6 \neq dim(\mathcal{N}(B^2)) = 5$.

Definition 11.5. $D = (d_{i,j}) \in M_n$ is **diagonal** provided that $d_{i,j} = 0$, $\forall i \neq j$. If $d_{i,i} = \alpha_i$, then we will often write, $D = diag(\alpha_1, ..., \alpha_n)$.

Note that when $D = diag(\lambda_1, ..., \lambda_n)$ then $p_D(t) = (t-\lambda_1)\cdots(t-\lambda_n)$.

Definition 11.6. A matrix $A \in M_n$ is **diagonalizable** provided that there exists $S \in M_n^{-1}$ such that $S^{-1}AS$ is diagonal.

Proposition 11.7. $A \in M_n$ *is diagonalizable* $\iff$ $\mathbb{C}^n$ *has a basis consisting of e-vectors for* A.

Definition 11.8. A set of matrices $\mathcal{F} \subseteq M_n$ is called **simultaneously diagonalizable** if there exists $S \in M_n^{-1}$ such that $S^{-1}AS$ is diagonal $\forall A \in \mathcal{F}$. A set of matrices $\mathcal{F} \subseteq M_n$ is called **commuting** if $\forall A, B \in \mathcal{F},\ AB = BA$.

Lemma 11.9. *Let $B = B_1 \oplus B_2 \oplus \cdots \oplus B_k$. Then B is diagonalizable iff each B_i is diagonalizable.*

Theorem 11.10. *Let $\mathcal{F} \subseteq M_n$. Then $\mathcal{F}$ is simultaneously diagonalizable iff $\mathcal{F}$ is commuting and each element of $\mathcal{F}$ is diagonalizable.*

Remark 11.11. Let $A, B \in M_n$. If $A \in M_n^{-1}$ then $A^{-1}(AB)A = BA$, so $AB \sim BA$. In this case we know that the spectral "data" of AB is the same as for BA, i.e., $\sigma(AB) = \sigma(BA)$, $p_{AB}(t) = p_{BA}(t)$ and multiplicities of e-values are the same. So what can be said when neither A nor B is invertible? If we let $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ and let $B = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$, then $AB = 0$, but $BA = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$, so these matrices are not similar. However, $p_{AB}(t) = t^2 = p_{BA}(t)$, so $\sigma(AB) = \sigma(BA) = \{0\}$. But notice that the geometry multiplicity of this e-value is different for the two matrices.

Theorem 11.12. *Let $A \in M_{m,n}$ and let $B \in M_{n,m}$. Then $t^m p_{BA}(t) = t^n p_{AB}(t)$. Hence, $\sigma(AB) \cup \{0\} = \sigma(BA) \cup \{0\}$.*

Corollary 11.13. *Let $A, B \in M_n$. Then $p_{AB}(t) = p_{BA}(t)$ and $\sigma(AB) = \sigma(BA)$.*

11.2. **Persistence of Eigenvalues.**

Lemma 11.14. *Given $\mu_1, ..., \mu_n$, if $S_m(\mu_1, ..., \mu_n) = 0, \forall m > n - k$, then at least k of the μ's are 0.*

Theorem 11.15. *Let $A \in M_n$ and let $\lambda \in \mathbb{C}$. Consider the following statements:*

(1) *λ is an e-value of A of geometric multiplicity greater than or equal to k,*
(2) *for any $m > n - k$ and any $S \subseteq \{1, ..., n\}, |S| = m$, λ is an e-value of $A[S, S]$,*
(3) *λ is an e-value of A of algebraic multiplicity greater than or equal to k.*

Then $(1) \implies (2) \implies (3)$.

Example 11.16. Put example here to show that none of the converses is true.

12. Unitaries and Isometries

Definition 12.1. A linear map $L : \mathbb{F}^n \to \mathbb{F}^m$ is called an **isometry** if $\|Lx\|_2 = \|X\|_2$ for every $x \in \mathbb{F}^n$.

Note: an isometry is 1-1 and so $m \geq n$.

Theorem 12.2. *Let $V \in M_{m,n}$. Then the following are equivalent:*

(1) *L_V is an isometry,*
(2) *the columns of V is an orthonormal set in $\mathbb{F}^m$,*
(3) *$V^*V = I_n$.*

Recall that a set $\{u_1, ..., u_n\}$ is an **orthonormal basis for $\mathbb{F}^n$(o.n.b.)** provided that $u_i \perp u_j$, $\forall i \neq j$ and $\|u_i\| = 1$, $\forall i$. **Parseval's identities** say that if $\{u_1, ..., u_n\}$ is an o.n.b., then for every $x \in \mathbb{F}^n$ we have that

$$\|x\|_2^2 = \sum_{i=1}^{n} |\langle x, u_i\rangle|^2 \text{ and } x = \sum_{i=1}^{n} \langle x, u_i\rangle u_i.$$

Definition 12.3. A set of vectors, $\{r_1, ..., r_m\} \subseteq \mathbb{F}^n$ is called a **Parseval frame** provided that for every $x \in \mathbb{F}^n$ we have $\|x\|_2^2 = \sum_{j=1}^{m} |\langle x, r_j\rangle|^2$.

Proposition 12.4. *Let $\{r_1, ..., r_m\} \subseteq \mathbb{F}^n$. The following are equivalent:*

(1) *$\{r_1, ..., r_m\}$ is a Parseval frame for $\mathbb{F}^n$,*
(2) *the $m \times n$ matrix $[r_1 \vdots \cdots \vdots r_m]^*$ is an isometry,*
(3) *for every $x \in \mathbb{F}^n, x = \sum_{j=1}^{m} \langle x, r_j\rangle r_j$.*

Definition 12.5. A Parseval frame is called **uniform** if $\|r_i\|_2 = \|r_j\|_2$, $\forall i, j$. A Parseval frame is called **equiangular** if it is uniform and for all $i \neq j$, $|\langle r_i, r_j\rangle|$ is a constant independent of i and j.

Remark 12.6. For $\mathbb{R}^n$: if an equiangular Parseval frame of m vectors exists, then $m \leq n(n+1)/2$. Also there are many n for which no equiangular frame for $\mathbb{R}^n$ exists with $m = n(n+1)/2$ vectors. There is a partial classification known of all the pairs (n, m) for which there exists an equiangular Parseval frame for $\mathbb{R}^n$ with m vectors. These are known to exist iff a certain type of completely regular graph on $n-1$ vertices exists.

For $\mathbb{C}^n$: Even less is known about for which pairs (n, m) an equiangular Parseval frame of m vectors exists for $\mathbb{C}^n$. It is known if an equiangular Parseval frame exists then $m \leq n^2$.

Zauner's Conjecture is that for every n, there exists an equiangular Parseval frame of n^2 vectors for $\mathbb{C}^n$. This is only known to be true for a few small values of n.

Remark 12.7. A similar problem is the **mutually unbiased basis problem**. Two o.n.b.'s $\{u_1, ..., u_n\}$ and $\{v_1, ..., v_n\}$ for $\mathbb{C}^n$ are called **mutually unbiased** provided that $|\langle u_i, v_j\rangle|$ is a constant independent of i and j. It is easily shown that necessarily the constant is $1/\sqrt{n}$.

The mutually unbiased basis problem is to find the maximum number $M(n)$ of sets of o.n.b.'s for $\mathbb{C}^n$ that are mutually unbiased. The conjecture is that $M(n) = n+1$. This is known to be true when n is a prime power. It is a pretty bold conjecture because it is still unknown even for $n = 6$. Note that every $n < 6$ is a prime power, so $n = 6$ is the first integer that is not a prime power.

Definition 12.8. A matrix $U \in M_n$ is **unitary** if $U^*U = I_n$. A unitary $U \in M_n(\mathbb{R})$, is also called an **orthogonal matrix**. Since $U^* = U^t$, in this case, we have that $U \in M_n(\mathbb{R})$ is orhtongonal iff $U^tU = I_n$.

Theorem 12.9. *Let $U \in M_n$. The following are equivalent:*

(1) *U is unitary,*
(2) *U is invertible and $U^* = U^{-1}$,*
(3) *$UU^* = I_n$,*
(4) *U^* is unitary,*
(5) *the columns of U are an o.n.b. for $\mathbb{C}^n$,*
(6) *the rows of U are an o.n.b. for $\mathbb{C}^n$,*
(7) *U is an isometry.*

Remark 12.10. The set of invertible $n \times n$ matrices, which I denote M_n^{-1} is a group called the **general linear group** and is generally denoted $GL(n, |bbF)$. The set of unitary matrices is a subgroup of this group, which is generally denoted $\mathcal{U}(n)$ in the complex case and $\mathcal{O}(n)$ in the real case.

If we identify $M_{m,n}$ with $\mathbb{F}^{mn}$ then we can endow matrices with the Euclidean norm, which we write $\|A\|_2 = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{I,j}|^2}$.

In this distance a sequence of matrices $A_k = (a_{i,j}(k))$ converges to a matrix $A = (a_{i,j})$ iff $\lim_k a_{i,j}(k) = a_{i,j}, \forall i, j$. We write $A_k \to A$.

Lemma 12.11. *The set of unitaries, $\mathcal{U}(n)$ is a closed and bounded set and hence compact. In particular, every sequence of unitaries will have a convergent subsequence.*

Lemma 12.12. *If $U \in \mathcal{U}(n)$ then $|det(U)| = 1$. If $U \in \mathcal{O}(n)$ then $det(U) \in \{+1, -1\}$.*

Example 12.13. In $\mathbb{R}^2$ if we rotate counterclockwise through angle θ, then $e_1 \to (cos(\theta), sin(\theta))$ and $e_2 \to (-sin(\theta), cos(\theta))$. Thus, the matrix of this transformation is

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

Note that $R(\theta_1)R(\theta_2) = R(\theta_1 + \theta_2)$ and that $det(R(\theta)) = +1$. In fact, it can be shown that if $U \in \mathcal{O}(2)$ and $det(U) = +1$, then $U = R(\theta)$ for some θ. Also, if $U \in \mathcal{O}(2)$ and $det(U) = -1$, then $U = R(\theta)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ for some θ.

13. Unitary Equivalence

Definition 13.1. $A, B \in M_n$ are **unitarily equivalent** if there exists a unitary U so that $B = U^*AU$. We write $A \sim_{ue} B$.

It is not hard to see that this is another equivalence relation on the matrices and that $A \sim_{ue} B \implies A \sim B$.

Proposition 13.2. *Let $A \sim_{ue} B$. Then*

(1) $\sigma(A) = \sigma(B)$,
(2) $p_A(t) = p_B(t)$,
(3) $\sum_{i,j=1}^n |a_{i,j}|^2 = \sum_{i,j=1}^n |b_{i,j}|^2$.

The easiest way to prove (3) is to use traces. Recall that for $A \in M_n$, $Tr(A) = \sum_{i=1}^n a_{i,i}$.

Proposition 13.3. *Let $A \in M_{m,n}$ and let $B \in M_{n,m}$. Then $Tr(AB) = Tr(BA)$.*

Now to see (3), we write

$$\sum_{i,j=1}^n |a_{i,j}|^2 = Tr(A^*A) = Tr((UBU^*)^*(UBU^*)) = Tr(UB^*U^*UBU^*)$$

$$= Tr(UB^*BU^*) = Tr(B^*BU^*U) = Tr(B^*B) = \sum_{i,j=1}^n |b_{i,j}|^2.$$

Example 13.4. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix}$ and let $B = diag(1, 2, 3)$. Then $A \sim B$, since A is diagonalizable. But they are not unitarily equivalent because they fail (3).

Example 13.5 (Radjavi)**.** Fix distinct numbers $a_1, ..., a_n$. Let $A = (a_{i,j})$ be upper triangular, with $a_{i,i} = a_i$ and $a_{i,i+1} > 0$ for all i. Then two such matrices are unitarily equivalent iff they are equal.

13.1. **The Specht Invariants.** First we need the concept of **free non-commuting variables**. Imagine that we have an alphabet with only two letters, $\{s, t\}$. Then the "word" $sst \neq sts$. This is what we mean by saying that the variables s and t are non-commuting, for if they did commute then $st = ts$ and so $sts = sst$. By "free" we mean that no two words are the same unless they are exactly the same word. Using exponents and introducing s^0 and t^0 to mean that there is no s or t at that spot, then we can write every word as

$$W(s,t) = s^{n_1}t^{m_1} \cdot s^{n_k}t^{m_k},$$

for some k and some choice of $n_1, ..., n_k, m_1, ..., m_k \in \mathbb{N} \cup \{0\} := \mathbb{Z}_+$. The sum $n_1 + m_1 + \cdots + n_k + m_k$ is called the **length of the word W** and is denoted, $|W|$.

The idea of words in two letters is that they can be used to express a general product of matrices. Namely, if $A, B \in M_n$ and $W(s,t)$ is as above, then we set

$$W(A,B) = A^{n_1}B^{m_1} \cdots A^{n_k}B^{m_k}.$$

Thus, for $W(s,t) = s^2t, W(A,B) = AAB = A^2B$, while if $W(s,t) = sts$, then $W(A,B) = ABA$.

Note that is U is a unitary and $B = U^*AU$, then for any word $W(s,t)$ as above,

$$W(B,B^*) = B^{n_1}(B^*)^{m_1} \cdots B^{n_k}(B^*)^{m_k} =$$
$$(U^*AU)^{n_1}(U^*A^*U)^{m_1} \cdots (U^*AU)^{n_k}(U^*A^*U)^{m_k} = U^*W(A,A^*)U.$$

Theorem 13.6 (Specht, 1940)**.** *Let $A, B \in M_n(\mathbb{C})$. Then $A \sim_{ue} B$ iff $Tr(W(A,A^*)) = Tr(W(B,B^*))$ for every word W.*

One implication is easy. If $B = U^*AU$ then $Tr((B,B^*)) = Tr(U^*W(A,A^*)U) = Tr(W(A,A^*)UU^*) = Tr(W(A,A^*))$.

We outline the proof of the converse below.

Lemma 13.7. *Let V, W be vector spaces over the same field, let $\{v_\alpha : \alpha \in A\} \subseteq V$ with span$\{v_\alpha : \alpha \in A\} = V$ and let $\{w_\alpha : \alpha \in A\} \subseteq W$ with span$\{w_\alpha : \alpha \in A\} = W$, be two sets of spanning vectors indexed by the same set. Then there exists a linear map $L : V \to W$ such that $L(v_\alpha) = w_\alpha$ iff whenever $\lambda_i \in \mathbb{F}$ is a finite set of scalars such that $\sum_i \lambda_i v_{\alpha_i} = 0$, it follows that $\sum_i \lambda_i w_{\alpha_i} = 0$.*

If $W_1(s,t) = s^{n_1}t^{m_1} \cdots s^{n_k}t^{m_k}$ nd $W_2(s,t) = s^{p_1}t^{q_1} \cdots s^{p_j}t^{q_j}$ then by their **product** we mean the word obtained by **concatenating** thier letters, i.e.,

$$W_1W_2(s,t) = s^{n_1}t^{m_1} \cdots s^{n_k}t^{m_k}s^{p_1}t^{q_1} \cdots s^{p_j}t^{q_j}.$$

Note that this has the property that if $A, B \in M_n$ then $W_1(A, B) \cdots W_2(A, B) = W_1W_2(A, B)$.

Definition 13.8. Given a matrix $A \in M_n$ then the ***-algebra generated by A** is the linear span of the identity matrix I_n and all the words in A and A^*. We denote this by $C^*(A)$.

Note that since we are in finite dimensions, the vector space $C^*(A)$ is automaticlly closed. Also if $X, Y \in C^*(A)$, then $XY, X^* \in C^*(A)$.

Proposition 13.9. *Let $A, B \in M_n$. If $Tr(W(A, A^*)) = Tr(W(B, B^*))$ for every word W, then there exists a linear map $\pi : C^*(A) \to C^*(B)$ that is one-to-one, onto and satisfies $\pi(W(A, A^*)) = W(B, B^*)$. Moreover, $\pi(XY) = \pi(X)\pi(Y)$, $\pi(X^*) = \pi(X)^*$.*

Proof. First we argue that the linear map exists. We have that $span\{W(A, A^*) : W \text{ is a word }\} = C^*(A)$ and $span\{W(B, B^*) : W \text{ is a word }\} = C^*(B)$. To apply the Lemma we need to prove that $X = \sum_i \lambda_i W_i(A, A^*) = 0$ implies that $Y = \sum_i \lambda_i W_i(B, B^*) = 0$.

But $X = 0$ iff $Tr(X^*X) = 0$ iff $Tr\big(\sum_{i,j} \lambda_i \overline{\lambda_j} W_i(A, A^*)W_j(A, A^*)^*\big) = \sum_{i,j} \lambda_i \overline{\lambda_j} Tr\big(W_i(A, A^*)W_j(A, A^*)\big) = 0$. Note that each product $W_i(A, A^*)W_j(A, A^*)^*$ is just a word in A and A^* and that $W_i(B, B^*)W_j(B, B^*)^*$ is the same word in B and B^*. Thus, the hypothesis guarantees that $Tr(Y^*Y) = 0$ and so there is a well-defined linear map π with $\pi(W(A, A^*)) = W(B, B^*)$.

Moreover, the same argument shows that if $Y = 0$ then $X = 0$ so that π has an inverse and so must be one-to-one and onto.

Finally, since $\pi(W_1(A, A^*)W_2(A, A^*)) = \pi(W_1W_2(A, A^*)) = W_1W_2(B, B^*) = W_1(B, B^*)W_2(B, B^*)$ we have that π behaves multiplicatively on words and from this it follows that it behaves multiplicatively on linear combinations of words. The proof that it preserves adjoints is similar. $\square$

Thus, we see that the hypotheses of Specht's theorem implies the existence of this map $\pi : C^*(A) \to C^*(B)$ and that π will preserve the trace of matrices. To complete his proof Specht used Wedderburn theory to deduce that such a map must be implemented by a unitary conjugation. Those of you familiar with the classification of finite dimensional C*-algebras(which is essentially the Wedderburn theory) should be able to complete this part on your own.

Specht's theorem is lovely, but it requires the checking of the trac of every word, which is not practical. Since the time of his proof, there has been a great deal of effort on finding as few words as possible.

For example, for M_2 it is known that the words s, s^s, st suffice that is:

Proposition 13.10. *Let* $A, B \in M_2$. *Then* $A \sim_{ue} B$ *iff* $Tr(A) = Tr(B), Tr(A^2) = Tr(B^2)$, *and* $Tr(AA^*) = Tr(BB^*)$.

A nice characterisation of unitary equivalence!

The first result to reduce the number of ords from infinitely many to a finite set was:

Theorem 13.11 (Pearcy, 1962)**.** *Let* $A, B \in M_n$. *Then* $A \sim_{ue} B$ *iff* $Tr(W(A, A^*)) = Tr(W(B, B^*)), \forall |W| \leq 2n^2$.

The proof requires two lemas.

Lemma 13.12. *Let* $\mathcal{L}_A(d) = span\{W(A, A^*) : |W| \leq d\}$. *If* $\mathcal{L}_A(d) = \mathcal{L}_A(d+1)$ *then* $\mathcal{L}_A(d) = \mathcal{L}_A(m)$ *for all* $m \geq d$.

Proof. Let $|W| = m = d + 1 + k, k \geq 1$. Then e can write $W = W_1W_2$ where $|W_1| = d{+}1$ and $|W_2| = k$. Then $W(A, A^*) = W_1(A, A^*)W_2(A, A^*)$. But $W_1(A, A^*)$ can be written as a linear combination of words of length at most d. Multiplying this linear combination by $W_2(A, A^*)$ expresses $W(A, A^*)$ as a linear combination of words of length at most $d + k$. Continuing in this manner we reduce the degree needed by one each time. □

Lemma 13.13. *Let* $A \in M_n$, *then* $C^*(A) = \mathcal{L}_A(n^2)$.

Proof. Since $C^*(A) \subseteq M_n$, we have that $dim(C^*(A)) \leq n^2$. Now $1 \leq dim(\mathcal{L}_A(d)) \leq dim(\mathcal{L}_A(d+1)) \leq n^2$ and if they are not equal then the dimension must grow by at least 1. By the pigeonhole principle, for some $d \leq n^2 - 1$, we must have that $dim(\mathcal{L}_A(d)) = dim(\mathcal{L}_A(d{+}1))$. □

To prove Pearcy's theorem we now use that $C^*(A) = \mathcal{L}_A(n^2)$ and $C^*(B) = \mathcal{L}_B(n^2)$. We return to the proof of the existence of π in Specht's theorem. Now we need only argue that if X which is now a sum of words of at most length n^2 is equal to Y which is also a sum of words of length at most n^2. We used that $X = 0$ implies that $Tr(XX^*) = 0$ and XX^* is a sum of words of length at most $2n^2$. So if traces of words of length up to length $2n^2$ agree, then we will get that $X = 0$ implies that $Tr(XX^*) = 0$ implies that $Tr(YY^*) = 0$ which implies that $Y = 0$.

Once one ha the existence of the map π the rest of the proof of Pearcy;s theorem is identical to the proof of the rest of Specht's theorem.

This theorem was the state of the art until recently, when the following was proved:

Theorem 13.14 (Djokovic-Johnson)**.** *For each n there is a set of at most $n\sqrt{\frac{2n^2}{n-1}+\frac{1}{4}}+\frac{n}{2}-2$ words such that for $A, B \in M_n$, we have that $A \sim_{ue} B$ iff $Tr(W(A, A^*)) = Tr(W(B, B^*))$ for all words in this set.*

Here is a set of 7 words that works for M_3: $s, s^2, ts, s^3, s^2t, s^2t^2, s^2t^2st$.

13.2. **Householder Transformations.** Let $w \in \mathbb{F}^n, w \neq 0$. Then the **Householder transformation** is $U_w = I - \frac{2}{\|w\|^2}ww^*$, thus, $U_w x = x - \frac{2}{\|w\|}^2\langle x, w\rangle w$. Note that is $w_1 = rw$ for $r \in \mathbb{F}$, then $U_w = U_{w_1}$. So we will often assume that $\|w\| + 1$.

We now wish to define study this map more closely. Assume that $\|w\| = 1$. We can always decompose $v \in \mathbb{F}^n$ as $v + v_1 + v_2$ with $v_2 \perp w$ and $v_1 = tw$ for some $t \in \mathbb{F}$. In this case we see that

$$U_w(v) = v - 2\langle tw + v_2, w\rangle w = v - 2tw = v_1 + v_2 - 2v_1 = v_2 - v_1.$$

This cn be seen to be the reflection of v about the space perpindicular to w.

Proposition 13.15. *$U_w^* = U_w$ and $U_w^2 = I$. So U_w is a unitary.*

Lemma 13.16. *Let $x, u \in \mathbb{C}^n$ with $\|x\| = \|u\| = 1$, and $\langle x, u\rangle \geq 0$. Set $w = u - x$. Then $U_w(x) = u$ and $U_w(u) = x$.*

Theorem 13.17 (Schur)**.** *Let $A \in M_n$ and let $\lambda_1, ..., \lambda_n$ be the roots of $p_A(t)$. Then there exists a unitray U that is a product of Householder transformations such that $U^*AU = T$ where T is an upper triangular matrix with $t_{i,i} = \lambda_i$, $\forall i$.*

We now look at some consequences of Schur's theorem.

Corollary 13.18. *Let $A \in M_N$ and let $\lambda_1, ..., \lambda_n$ be the roots of $p_A(t)$. Then $Tr(A^k) = \lambda_1^k + \cdots + \lambda_n^k = \mu_k$.*

Remark 13.19. Recall that by Newton's identities,$S_k(\lambda_1, ..., \lambda_n) = E_k(A), 1 \leq k \leq n$ can be expressed as formulas in the $\mu_k, 1 \leq k \leq n$. Thus, the coefficients of $p_A(t)$ can also be expressed as formulas in terms of $Tr(A^k), 1 \leq k \leq n$ and so these numbers determine $p_A(t)$ and consequently, $\lambda_1, ...\lambda_n$.

Theorem 13.20. *Let $A \in M_n$ and let $\epsilon > 0$ be given. Then there exists $B \in M_n$ such that B is invertible and diagonalizble with $\|A - B\|_2 = \sqrt{\sum_{i,j=1}^n |a_{i,j} - b_{i,j}|^2} < \epsilon$.*

Theorem 13.21 (Cayley-Hamilton)**.** *Let $A \in M_n$ then $p_A(A) = 0$.*

Corollary 13.22. *Let $A \in M_n^{-1}$, then $A^{-1} \in span\{I, A, ..., A^{n-1}\}$.*

Given $A_1, ..., A_r \in M_n$ and $B_1, ..., B_r \in M_m$ we can define a linear map $L : M_{n,m} \to M_{n,m}$ by $L(Y) = \sum_{i=1}^r A_i Y B_i$. There is a great deal of research devoted to issues such as determining the norm and the spectrum of such maps, especially the cb-norms of such maps, for those of you who have heard of this concept. Following is one result that e will need.

Proposition 13.23. *Let $A \in M_n$ and $B \in M_m$ and define $L : M_{n,m} \to M_{n,m}$ by $L(Y) = AY - YB$. If $\sigma(A) \cap \sigma(B)$ is empty, then L is invertible.*

Corollary 13.24. *Let $A \in M_n$, $B \in M_m$, and $X \in M_{n,m}$. If $\sigma(A) \cap \sigma(B)$ is empty, then $\begin{pmatrix} A & X \\ 0 & B \end{pmatrix} \sim \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$.*

Theorem 13.25. *Let $A \in M_n$ and let $p_A(t) = (t - \lambda_1)^{n_1} \cdots (t - \lambda_k)^{n_k}$, then A is similar to a block diagonal matrix with diagonal blocks $T_1, ..., T_k$ of sizes $n_1, ..., n_k$ where each is of the form $T_i = \lambda_i I_{n_i} + N_i$ where each N_i is strictly upper triangular.*

This last reult is quite close to proving the existence of the Jordan canonical form. We let $J_n(\lambda) = \lambda I_n + J_n(0) \in M_n$ where $J_n(0) = \sum_{i=1}^{n-1} E_{i,i+1}$. This matrix is called the **elementary Jordan block of size n and with eigenvalue** λ. The Jordan theorem says that every matrix is similar to a direct sum of such Jordan blocks.

To complete the proof of this theorem from the last result we would only need to show that each T_i is similar to a direct sum of Jordan blocks of the form $J_m(\lambda_i)$ for some m's. To do this it is enough to show that each N_i is similar to a sum of $J_m(0)$'s. This is accomplished by the following theorem, whose proof we omit. Note that because each N_i is strictly upper triangular, $N_i^{n_i} = 0$.

Definition 13.26. A matrix N is called **nilpotent** if there is some K such that $N^K = 0$. The least such K is called the **order of nilpotency** and N is said to be **nilpotent of order K**.

Theorem 13.27. *Let $N \in M_n$ be nilpotent of order K. Then N is similar to a direct sum of Jordan blocks of the form $J_{m_1}(0), ..., J_{m_r}(0)$ with $m_i \le K$ for all i. Moreover, if we let l_i equal the number of Jordan blocks of size $i, 1 \le i \le K$ and let $d_i = dim(\mathcal{N}(N^i)), 1 \le i \le K$, then $d_i = \sum_{j=1}^K min\{i,j\} l_j$. Since the matrix $(min\{i,j\})$ is invertible, this equation determines the numbers l_i uniquely in terms of the numbers d_i.*

13.3. **QR factorization and Gram-Schmidt.** Recall that if $B = [b_1 \vdots \cdots \vdots b_n] \in M_{m,n}$ and $R = (r_{i,j}) \in M_{n,p}$, then $BR = [\sum_i r_{i,1} b_i \vdots \cdots \vdots \sum_j r_{i,p} b_i]$.

Theorem 13.28. *Let* $A \in M_{m,n}$ *then there exists* $U \in \mathcal{U}(m)$ *and an* $R \in M_{m,n}$ *that is upper triangular such that* $A = UR$.

14. Normal and Hermitian Matrices

Proposition 14.1 (Cartesian Decomposition)**.** *Let* $A \in M_n$. *We define*

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

Then $Re(A)^* = Re(A)$, $Im(A)^* = Im(A)$, $A = Re(A) + iIm(A)$ *and if* $H = H^*, K = K^*$ *satisfy* $A = H + iK$, *then* $H = Re(A), K = Im(A)$.

We adopt the notation $[X, Y] = XY - YX$. So that $XY = YX$ iff $[X, Y] = 0$.

Definition 14.2. A matrix A is **normal** if $[A, A^*] = 0$. A matrix B is **unitarily diagonalizable** if there exists a unitary U so that U^*BU is diagonal.

If $A = A^*$ then A is normal.

Theorem 14.3. *Let* $A \in M_n$ *and let* $p_A(t) = (t - \lambda_1) \cdots (t - \lambda_n)$. *The following are equivalent:*

(1) *A is normal,*
(2) $[Re(A), Im(A)] = 0$,
(3) *A is unitarily diagonalizable,*
(4) $\sum_{i,j=1}^{n} |a_{i,j}|^2 = \sum_{j=1}^{n} \lambda_n|^2$,
(5) *there exists an orthonormal basis of for* $\mathbb{C}^n$ *consisting of e-vectors for* A.

Corollary 14.4. *Let* $A \in M_n$. *The following are equivalent:*

(1) $A = A^*$,
(2) *there exists a unitary* U *such that* U^*AU *is a diagonal matrix with real entries,*
(3) *there exists an orthonormal basis for* $\mathbb{C}^n$ *consisting of e-vectors for* A *with real e-values.*

Theorem 14.5. *Let* $A \in M_n$. *Then* $A = A^*$ *iff* $\langle Ax, x \rangle \in \mathbb{R}$ *for every* $x \in \mathbb{C}^n$.

Let $A = A^*$ then we know the roots of $p_A(t)$ are all real. We shall order them so that $\lambda_1 \leq \ldots \leq \lambda_n$.

Theorem 14.6 (Rayleigh-Ritz)**.** *Let* $A = A^* \in M_n$. *Then:*

(1) $\lambda_1 x^*x \leq \langle Ax, x \rangle \leq \lambda_n x^*x$, $\forall x \in \mathbb{C}^n$,

(2) $\lambda_n = \begin{matrix} max \\ x \neq 0 \end{matrix} \frac{\langle Ax,x\rangle}{x^*x} = \begin{matrix} max \\ \|x\| = 1 \end{matrix} \langle Ax, x\rangle,$

(3) $\lambda_1 = \begin{matrix} min \\ x \neq 0 \end{matrix} \frac{\langle Ax,x\rangle}{x^*x} = \begin{matrix} min \\ \|x\| = 1 \end{matrix} \langle Ax, x\rangle.$

Corollary 14.7. *Let $A = A^* \in M_n$, let $x \in \mathbb{C}^n$, $x \neq 0$, and let $\alpha = \frac{\langle Ax,x\rangle}{x^*x}$. Then there is an e-value for A in $(-\infty, \alpha]$ and in $[\alpha, +\infty)$.*

Lemma 14.8 (Subspace Intersection Lemma)**.** *Let V_1, V_2 be finite dimensional subspaces of a vector space W. Then*

$$dim(V_1 + V_2) + dim(V_1 \cap V_2) = dim(V_1) + dim(V_2).$$

Proof. Consider the onto map $L : V_1 \oplus V_2 \to V_1 + V_2$ defined by $L((v_1, v_2)) = v_1 - v_2$ and apply the rank-nullity. □

Corollary 14.9. *If $dim(V_1) + dim(V_2) - dim(V_1 + V_2) \geq 1$, then $V_1 \cap V_2 \neq (0)$.*

Given $A = A^* \in M_n$ if the roots of $p_A(t)$(including multiplicities) are ordered so that $\lambda_1 \leq \ldots \leq \lambda_n$, then we set $\lambda_i(A) = \lambda_i$.

Theorem 14.10 (Courant-Fischer)**.** *Let $A = A^* \in M_n$, let $1 \leq k \leq n$, let S always denote a subspace of $\mathbb{C}^n$ and let $S_1 = \{x \in S : \|x\| = 1\}$ denote the unit sphere. Then*

(1) $\lambda_k(A) = \begin{matrix} min \\ dim(S) = k \end{matrix} \left(\begin{matrix} max \\ x \in S_1 \end{matrix} \langle Ax, x\rangle \right),$

(2) $\lambda_k(A) = \begin{matrix} max \\ dim(S) = n-k+1 \end{matrix} \left(\begin{matrix} min \\ x \in S_1 \end{matrix} \langle Ax, x\rangle \right).$

Theorem 14.11. *Let $A = A^* \in M_n$, let $S \subseteq \mathbb{C}^n$ be a subspace of dimension k, let S_1 denotes its unit sphere, and let $c \in \mathbb{R}$.*

(1) *If $c \leq \langle Ax, x\rangle$, $\forall x \in S_1$, then $c \leq \lambda_{n-k+1}(A)$(if the 1st inequality is strict, then so is the 2nd),*
(2) *If $\langle Ax, x\rangle\rfloor$, $\forall x \in S_1$, then $\lambda_k(A) \leq c$(if the 1st inequality is strict, then so is the 2nd).*

Theorem 14.12 (Weyl)**.** *Let $A = A^*, B = B^*$ both in M_n. Then:*

(1) $\lambda_i(A + B) \leq \lambda_{i+j}(A) + \lambda_{n-j}(B)$, $\forall 0 \leq j \leq n-1, 1 \leq i \leq n$. *Equality holds in this inequality iff $\exists x \neq 0$, such that $Ax = \lambda_{i+j}(A)x, Bx = \lambda_{n-j}(B)x$ and $(A + B)x = \lambda_i(A + B)x$.*
(2) $\lambda_{i-j+1}(A) + \lambda_j(B) \leq \lambda_i(A + B)$, $\forall 1 \leq i \leq n, 1 \leq j \leq i$. *Equality holds in this inequality iff $\exists x \neq 0$ such that $Ax = \lambda_{i-j+1}(A)x, Bx = \lambda_j(B)x$ and $(A + B)x = \lambda_i(A + B)x$.*

To prove we needed:

Lemma 14.13 (2nd Subspace Lemma)**.** *Let $S_1, ..., S_k$ be subspaces of $\mathbb{C}^n$. Let*

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

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

Proof. Consider the linear map $L : S_1 \oplus \cdots S_k \to (S_1 + S_2) \oplus \cdots \oplus (S_{k-1} + S_k)$ defined by $L(v_1, ..., v_k) = (v_1 - v_2, \ldots, v_{k-1} - v_k)$ and apply rank-nullity. □

Theorem 14.14 (Cauchy's Eigenvalue Interlacing Theorem)**.** *Let $A = A^* \in M_n$, let $y \in \mathbb{C}^n$, let $a \in \mathbb{R}$ and define $\hat{A} = \hat{A}^* \in M_{n+1}$ by $\hat{A} = \begin{pmatrix} A & y \\ y^* & a \end{pmatrix}$. Then*

$$\lambda_i(\hat{A}) \leq \lambda_i(A) \leq \lambda_{i+1}(\hat{A}), \forall 1 \leq i \leq n.$$

This yields another persistence theorem.

Corollary 14.15. *Let $A = A^* \in M_n$. Suppose that λ is an eigenvalue of A of geometric multiplicity k. Let $C \in M_{n,k-1}$ let $D \in M_{k-1}$ and let $B = \begin{pmatrix} A & C \\ C^* & D \end{pmatrix}$. Then λ is an eigenvalue of B.*

The following shows that all "interlacings" are attained.

Theorem 14.16. *Let*

$$\mu_1 \leq \lambda_1 \leq \mu_2 \leq \ldots \leq \mu_n \leq \lambda_n \leq \mu_{n+1}.$$

Then there exists $A = A^* \in M_n$, $y_k \geq 0, 1 \leq k \leq n$, $y = (y_1, ..., y_n)^t$ and $a \in \mathbb{R}$ such that for $\hat{A} = \begin{pmatrix} A & y \\ y^* & a \end{pmatrix}$ one has that $\lambda_j(A) = \lambda_j, \forall 1 \leq j \leq n$ and $\lambda_j(\hat{A}) = \mu_j, \forall 1 \leq j \leq n+1$.

Theorem 14.17. *Let $A = A^* = (a_{i,j}) \in M_n$ and let $1 \leq m \leq n$. Then $\sum_{i=1}^m \lambda_i(A) \leq \sum_{i=1}^m a_{i,i}$.*

Proof. Let $A_m = A[\{1, ..., m\}, \{1, ..., m\}]$ be the "upper left" $m \times m$ corner of A. So that A_{m+1} is obtained from A_m by adjoining a row and column. Hence by eigenvalue interlacing, $\lambda_i(A_{m+1}) \leq \lambda_i(A_m)$, for $1 \leq i \leq m$. Inductively, we obtain that for $A = A_n$, $\lambda_i(A) \leq \lambda_i(A_m)$ for $1 \leq i \leq m$. Hence,

$$\sum_{i=1}^m \lambda_i(A) \leq \sum_{i=1}^m \lambda_i(A_m) = Tr(A_m) = \sum_{i=1}^m a_{i,i}.$$

□

Corollary 14.18. *Let $A = A^* \in M_n$ and let $1 \le m \le n$. Then*

$$\sum_{i=1}^{m} \lambda_i(A) = \overset{min}{u_1, ..., u_m} \sum_{i=1}^{m} \langle Au_i, u_i \rangle,$$

where the minimum is taken over all orthonormal sets $\{u_1, ..., u_m\} \in \mathbb{C}^n$.

Proof. Each such orthonormal set can be extended to a basis $\{u_1, ..., u_n\}$ and the matrix of A with respect to this basis is $((\langle Au + j, u_i \rangle)$. Using the fact that $\lambda_i(A) = \lambda_i(U^*AU)$ for any unitary, and applying the above result, shows that the LHS is less than the RHS. Choosing the onb such that $Au_i = \lambda_i(A)u_i$ shows the equality. □

14.1. **Majorization.** Let $x = (x_1, ..., x_n) \in \mathbb{R}^n$, then we let $x^\downarrow = (x_1^\downarrow, ..., x_n^\downarrow)$ denote the rearrangement of x into decreasing order, i.e., $x_1^\downarrow \ge \cdots \ge x_n^\downarrow$. We also let $x^\uparrow = (x_1^\uparrow, ..., x_n^\uparrow)$ denote the increasing rearrangement of x.

Proposition 14.19. *Let $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathbb{R}^n$ with $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. The following are equivalent:*

(1) $\overset{max}{\{1 \le i_1 < \cdots < i_k \le n\}} \left(\sum_{j=1}^k x_{i_k} \right) \ge$
$\overset{max}{\{1 \le i_1 < \cdots < i_k \le n\}} \left(\sum_{j=1}^k y_{i_j} \right), \forall 1 \le k \le n,$

(2) $\sum_{j=1}^k x_j^\downarrow \ge \sum_{j=1}^k y_j^\downarrow, \forall 1 \le k \le n,$

(3) $\sum_{j=1}^k x_j^\downarrow \le \sum_{j=1}^k y_j^\downarrow, \forall 1 \le k \le n,$

(4) $\overset{min}{\{1 \le i_1 < \cdots < i_k \le n\}} \left(\sum_{j=1}^k x_{i_j} \right) \le$
$\overset{min}{\{1 \le i_1 < \cdots < i_k \le n\}} \left(\sum_{j=1}^k y_{i_j} \right), \forall 1 \le k \le n.$

Definition 14.20. We say that $x = (x_1, ..., x_n)$ **majorizes** $y = (y_1, ..., y_n)$ provided that $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ and any of the four above equivalent conditions holds.

Theorem 14.21 (Schur)**.** *Let $A = A^* = (a_{i,j}) \in M_n$ with $p_A(t) = \prod_{i=1}^n (t - \lambda_i)$. Then $(\lambda_1, ..., \lambda_n)$ majorizes $(a_{1,1}, ..., a_{n,n})$.*

Proof. We can apply a permutation unitary to A so that the diagonal elements of $P^*AP = B = (b_{i,j})$ are arranged in increasing order, i.e., so that $b_{i,i} = a_{i,i}^\uparrow$. By the above theorem,

$$\sum_{i=1}^{m} \lambda_i^\uparrow = \sum_{i=1}^{m} \lambda_i(A) = \sum_{i=1}^{m} \lambda_i(B) \le \sum_{i=1}^{m} b_{i,i} = \sum_{i=1}^{m} a_{i,i}^\uparrow.$$

□

15. Positive Semidefinite Matrices

Definition 15.1. $A \in M_n$ is **positive semidefinite** denoted by $A \geq 0$ or $0 \leq A$, provided that $\forall x \in \mathbb{C}^n$ we have that $\langle Ax, x\rangle \geq 0$. It is called **positive definite** denoted $A > 0$ or $0 < A$ provided that $\langle Ax, x\rangle > 0$ for all $0 \neq x \in \mathbb{C}^n$.

Proposition 15.2. *$A \geq 0$ iff $A = A^*$ and $\lambda_i(A) \geq 0$ for all i.*

Proposition 15.3. *The following are equivalent:*

- $A > 0$,
- $A = A^*$ *and* $\lambda_i(A) > 0$ *for all* i,
- $\exists\, \delta > 0$ *such that* $\langle Ax, x\rangle \geq \delta$ *for all* $x \in \mathbb{C}^n$, $\|x\|_2 = 1$,
- $A \geq 0$ *and* $det(A) > 0$.

Lemma 15.4. *Let $A = A^* \in M_n^{-1}$. Then $A^{-1} = (A^{-1})^*$ and $p_{A^{-1}}(t) = \prod_{i=1}^{n}(t - \lambda_i(A)^{-1})$.*

Proposition 15.5.

- $A \geq 0 \implies \overline{A} = A^t \geq 0$,
- $A > 0 \implies \overline{A} = A^t > 0$ *and* $A^{-1} > 0$,
- $A \geq 0$ *and* $S \subseteq \{1, ..., n\}$ *implies* $A[S, S] \geq 0$,
- $A > 0$ *and* $S \subseteq \{1, ..., n\}$ *implies* $A[S, S] > 0$.

Proposition 15.6. *Let $A \geq 0$(respectively, $A > 0$), then $A^k \geq 0$(resp., $A^k > 0$) for any $k \in \mathbb{N}$.*

Set $A_m = A[\{1, ..., m\}, \{1, ..., m\}]$.

Theorem 15.7. *Let $A = A^* \in M_n$. Then $A > 0$ iff $det(A_m) > 0$ for all $1 \leq m \leq n$.*

Remark 15.8. This result is often used in multivariable calculus. Suppose that $D \subseteq \mathbb{R}^n$ is an open set and that $f : D \to \mathbb{R}$ is in $\mathcal{C}^2(D)$. The **Hessian** of f at a point x is defined by

$$H_f(x) = (\frac{\partial^2 f(x)}{\partial_i \partial_j}).$$

If $f'(x_0) = 0$ and $H_f(x_0) > 0$, then x_0 is a relative minimum. By the above theorem, $H_f(x_0) > 0$ iff each of the subdeterminants is positive. Instead of stating that the matrix must be positive definite, which is a concept not usually introduced in calculus, the positivity of the subdeterminants is often given as the criteria for a relative minimum.

If $f'(x_0) = 0$ and $-H_f(x_0) > 0$, then x_0 is a relative maximum. In terms of subdeterminants, this is equivalent to the signs of the subdeterminants of $H_f(x_0)$ having alternating signs, which again is the criteria given in calculus for a relative maximum.

So the real theorem to remember is the positive definite condition on the Hessian matrix, because this can often be seen more readily then computing so many determinants.

Proposition 15.9. *Let $A \in M_n$, $C \in M_{n,m}$. If $A \geq 0$, then $C^*AC \geq 0$. In particular, $C^*C \geq 0$.*

Proposition 15.10. *$A \in M_n$, $k \in \mathbb{N}$, then there exists $B \in M_n$, $B \geq 0$ with $B^k = A$.*

Proof. Let $U^*AU = D = diag(\lambda_1, ..., \lambda_n)$ and set $B = U^*diag(\lambda_1^{1/k}, ..., \lambda_n^{1/k})U$. □

Definition 15.11. We will denote the B obtained in this was as $A^{1/k}$.

Later we will prove uniqueness of this k-th root.

15.1. **Factorization and Decomposition of Positive semidefinite matrices.**

Proposition 15.12. *Let $A \geq 0$, then there exists C such that $A = C^*C$.*

Proof. Let $C = A^{1/2}$. □

Proposition 15.13. *Let $A \geq 0$ and let $x \in \mathbb{C}^n$. Then $Ax = 0$ iff $\langle Ax, x\rangle = 0$.*

Proposition 15.14. *Let $A \in M_n$. If $A \geq 0$, then there exists an upper triangular matrix R such that $A = R^*R$.*

Proof. Write $A = C^*C$ as above, then QR-factor, $C = QR$ and note that since Q is unitary, $A = R^*R$. □

Proposition 15.15. *Let $P \in M_n, P \geq 0$, then $\mathcal{R}(P) \perp \mathcal{N}(P)$ and $\mathcal{R}(P) + \mathcal{N}(P) = \mathbb{C}^n$.*

Lemma 15.16 (Cholesky's Lemma)**.** *Let $P = P^*$ be written in block form as $P = \begin{pmatrix} A & B \\ B^* & C \end{pmatrix}$ with $A > 0$. Then the following are equivalent:*

1. *$P \geq 0$,*
2. *$P - \begin{bmatrix} A^{1/2} \\ B^*A^{-1/2} \end{bmatrix} \begin{bmatrix} A^{1/2} & A^{-1/2}B \end{bmatrix} \geq 0$,*
3. *$C - B^*A^{-1}B \geq 0$.*

Lemma 15.17. *Let $P = (p_{i,j}) \geq 0$. If there exists i such that $p_{i,i} = 0$, then $p_{i,j} = p_{j,i} = 0$, $\forall j$.*

Remark 15.18 (Cholesky's Algorithm)**.** We described how to apply Cholesky's lemma to obtain a fast algorithm that 1) determines if $P \geq 0$ while at the same time 2) factoring $P = T^*T$ with T upper triangular. This gives a 2nd way to factor, with the first based on QR which in turn is based on the Gram-Schmidt algorithm. In general Cholesky is faster.

15.2. **Decomposition.** Each factorization $P = C^*C$ also yields a way to write P as a sum of rank one positives. To see this recall that when $X = [x_1, ..., x_n]$ in terms of columns and $Y = [r_1, ..., r_n]^t$ is written in terms of rows then $XY = \sum_{i=1}^n x_i y_i^*$ which decomposes the product into a sum of rank ones. Thus, each factorization of $P = XX^* = \sum_{i=1}^n x_i x_i^*$ expresses P as a sum of rank one positives.

There is a third way to factorize/decompose $P \geq 0$ which is called the **spectral factorization/decomposition**. Choose an onb of eigenvectors $\{u_1, ..., u_n\}$ with $Pu_i = \lambda_i u_i$ and note that $P = \sum_{i=1}^n \lambda_i u_i u_i^* = \sum_{i=1}^n (\sqrt{\lambda_i} u_i)(\sqrt{\lambda_i} u_i)^*$.

15.3. **Subspaces and Orthogonal Complements.**

Definition 15.19. Let $V \subseteq \mathbb{C}^n$ be a subspace. Then the **orthogonal complement of** V is the set

$$V^\perp = \{w \in \mathbb{C}^n : w \perp v, \forall v \in V\}.$$

Proposition 15.20. *$V^\perp$ is a subspace and $V \cap V^\perp = (0)$.*

Theorem 15.21. *Let $V \subseteq \mathbb{C}^n$ be a subspace, with dim$(V) = d$, let $\{v_1, ..., v_d\}$ be an onb for V and set $P = \sum_{i=1}^d v_i v_i^*$. Then:*

(1) $P = P^2 = P^*$,
(2) $\mathcal{R}(P) = V$,
(3) $Pv = v, \forall v \in V$,
(4) $P(V^\perp) = (0)$,
(5) $(I - P) = (I - P)^2 = (I - P)^*$,
(6) $\mathcal{R}(I - P) = V^\perp$,
(7) $(I - P)w = w, \forall W \in V^\perp$,
(8) *$V + V^\perp = \mathbb{C}^n$ and for each $x \in \mathbb{C}^n$ there is a unique way to write $x = v + w$ with $v \in V$, and $w \in V^\perp$,*
(9) *if $\{w_1, ..., w_d\}$ is another onb for V, then $P = \sum_{i=1}^d w_i w_i^*$.*

Definition 15.22. We call the P obtained in the above theorem the **orthogonal projection of $\mathbb{C}^n$ onto** V.

Definition 15.23. Let $\{w_1, ..., w_n\} \subseteq \mathbb{C}^k$. The **Gram matrix** or **Grammian** of these vectors is the matrix $G = (g_{i,j}) \in M_n$ with $g_{i,j} = \langle w_j, w_i \rangle$.

Theorem 15.24. *Let* $\{w_1, ..., w_n\} \subseteq \mathbb{C}^k$ *and let* $W = [w_1 \vdots \dots \vdots w_n] \in M_{k,n}$. *Then:*

(1) $G = W^*W$ *and hence* $G \geq 0$,
(2) $G > 0$ *iff* $\{w_1, ..., w_n\}$ *is a linearly independent set,*
(3) $rank(G) = rank(W) = dim(span\{w_1, ..., w_n\})$.

15.4. **The Polar Form and the Singular Value Decomposition.**

Definition 15.25. Given $A \in M_{m,n}$ we set $|A| = (A^*A)^{1/2} \in M_n$ which we call the **absolute value of** A. The numbers, $s_i(A) := \lambda_i^{\downarrow}(|A|)$, $1 \leq i \leq n$ are called the **singular values of** A.

Lemma 15.26. *Let* $A \in M_{m,n}$ *and let* $x \in \mathbb{C}^n$. *Then* $\|Ax\|_2 = \||A|x\|_2$ *and so* $Ax = 0$ *iff* $|A|x = 0$.

Theorem 15.27 (Polar Decomposition Theorem I)**.** *Let* $A \in M_{m,n}$. *Then there exists a unique isometry* $W : \mathcal{R}(|A|) \to \mathcal{R}(A)$ *such that* $A = W|A|$.

Theorem 15.28 (Polar Decomposition Theorem II)**.** *Let* $A \in M_n$ *then there exists a unitary* U *such that* $A = U|A|$.

Corollary 15.29. *Let* $A \in M_n$ *then there exists a unitary* V *such that* $A = |A^*|V$.

Corollary 15.30 (Singular Value Decomposition I)**.** *Let* $A \in M_n$ *let* $S = diag(s_1(A), ..., s_n(A))$. *Then there exists unitaries* U, V *such that* $A = USV$.

Corollary 15.31 (Singular Value Decomposition II)**.** *Let* $A \in M_n$ *then there exists onb's* $\{u_1, ..., u_n\}$ *and* $\{v_1, ..., v_n\}$ *for* $\mathbb{C}^n$ *such that*

$$A = \sum_{i=1}^{n} s_i(A) u_i v_i^*.$$

15.5. **Schur Products.**

Definition 15.32. Let $A = (a_{i,j}), B = (b_{i,j}) \in M_{m,n}$. Then their **Schur**(or **Hadamard** or **freshman**) **product** is the matrix

$$A \circ B = (a_{i,j} b_{i,j}) \in M_{m,n}.$$

Properties: $A \circ B = B \circ A$, $(A \circ B) \circ C = A \circ (B \circ C)$, $A \circ (B + C) = A \circ B + A \circ C$, $\lambda(A \circ B) = (\lambda A) \circ B = A \circ (\lambda B)$.

These properties imply that the map $\circ : M_{m,n} \times M_{m,n} \to M_{m,n}$ is bilinear. Finally, note that if J is the matrix of all 1's and the appropriate size, then $J \circ A = A \circ J = A$.

Theorem 15.33. *Let* $A, B \in M_n$. *If* $A \geq 0$ *and* $B \geq 0$, *then* $A \circ B \geq 0$.

15.6. **The Positive Semidefinite Ordering.** Given $A = A^*, B = B^* \in M_n$ we write $A \geq B$ (or $B \leq A$) iff $A - B$ is positive semidefinite. We write $A > B$(or $B < A$) iff $A - B$ is positive definite.

Proposition 15.34. *Let $A = A^*, B = B^*, C = C^* \in M_n$. Then*

(1) $A \leq B$ *and* $B \leq A$ *implies* $A = B$
(2) $A \leq B, B \leq C$ *implies* $A \leq C$,
(3) $A \leq B$ *implies* $A + C \leq B + C$,
(4) $A \leq B$ *and* $X \in M_{n,m}$ *implies* $X^*AX \leq X^*BX$.

Lemma 15.35. $A \geq I$ *implies* $A^{-1} \leq I$.

Theorem 15.36. *Let $A = A^*, B = B^* \in M_n$. Then*

(1) $A \geq B > 0$ *implies* $A^{-1} \leq B^{-1}$,
(2) $A \geq B \geq 0$ *implies* $det(A) \geq det(B)$ *and* $Tr(A) \geq Tr(B)$
(3) $A \geq B$ *implies* $\lambda_k(A) \geq \lambda_k(B)$ *for* $1 \leq k \leq n$.

Theorem 15.37. *Let $P \in M_n$ and let $S \subseteq \{1, ..., n\}$. If $P > 0$ then $P^{-1}[S, S] \geq P[S, S]^{-1}$.*

16. Matrix Norms

Definition 16.1. V a vector space, then a map $\|\cdot\| : V \to \mathbb{R}$ is called a norm provided:

(1) $\|v\| \geq 0 \ \forall v$,
(2) $\|v\| = 0$ iff $v = 0$,
(3) $\|\lambda v\| = |\lambda|\|v\|$,
(4) $\|v + w\| \leq \|v\| + \|w\|$.

Example 16.2. $\|v\|_2 = \left(\sum_{i=1}^n |\lambda_i|^2\right)^{1/2}$, $\|v\|_1 = \sum_{i=1}^n |\lambda_i|$ and $\|v\|_\infty = max\{|\lambda_1|, ..., |\lambda_n|\}$ are three norms on $\mathbb{F}^n$ that we will use.

Remark 16.3. One fact that we will use without proof is that given any pair of norms on $\mathbb{F}^n$ then there are constants $C_1, C_2 > 0$ such that $\|v\|_1 \leq C_2\|v\|_2$ and $\|v\|_2 \leq C_1\|v\|_1$ for all $v \in \mathbb{F}^n$.

Definition 16.4. A norm $\|\cdot\|$ on M_n is called a **matrix norm** provided that $\|AB\| \leq \|A\|\|B\|$.

Careful: In functional analysis courses they also require that $\|I_n\| = 1$, which we do not require.

Example 16.5. For $A \in M_n$ if we set $\|A\|_2 = \left(\sum_{i,j+1}^n |a_{i,j}|^2\right)^{1/2}$, then this defines a matrix norm with $\|I_n\| = \sqrt{n}$. So is $\|A\|_1 = \sum_{i,j=1}^m |a_{i,j}|$ and $\|I_n\|_1 = n$.

Example 16.6. If we set $\|A\|_\infty = max\{|a_{i,j}| : 1 \le i, j \le n\}$ then this is not a matrix norm. An easy way to see is to let J_n denote the matrix of all 1's. Then $J_n^2 = nJ_n$. Hence, $n = \|J_n^2\|_\infty > \|J_n\|_\infty \|J_n\|_\infty$.

Example 16.7. Given any norm $\|\cdot\|$ on $\mathbb{F}^n$ if we set $\|A\| = sup\{\|Ax\| : \|x\| = 1\}$ then this defines a matrix norm on M_n that we call the **induced operator norm**. For such norms we always have that $\|I_n\| = 1$.

Proposition 16.8. *Let $\|\cdot\|$ be any matrix norm on M_n. Then:*

(1) $\|I_n\| \ge 1$,
(2) *if* $A \in M_n^{-1}$ *then* $\|A^{-1}\| \ge \frac{\|I\|}{\|A\|}$,
(3) $\forall \lambda \in \sigma(A)$, *we have that* $|\lambda| \le \rho(A) \le \|A\|$.

Theorem 16.9. *Let $A \in M_n$. Then*

$$\rho(A) = inf\{\|A\| : \|\cdot\| \text{ is a matrix norm on } M_n\}.$$

Since all norms on M_n are equivalent, a sequence of matrices converges to 0 in a particular norm iff their entries converge to 0, i.e., convergence to 0 has the same meaning independent of the particular norm.

Corollary 16.10. *Let $A \in M_n$. If $\rho(A) < 1$ then $A^k \to 0$.*

Theorem 16.11 (Gelfand)**.** *Let $\|\cdot\|$ be any matrix norm on M_n and let $A \in M_n$. Then $\rho(A) = \lim_n \|A^n\|^{1/n}$.*

Given a power series $p(z) = \sum_{k=0}^{+\infty} p_k z^k$ and a matrix $A \in M_n$ we let $p_N(z) = \sum_{k=0}^{N} p_k z^k$. We say that $p(A)$ converges if there exists a matrix B such that $B - p_N(A) \to 0$ as $N \to +\infty$.

Theorem 16.12. *Let $p(z) = \sum_{k=0}^{+\infty} p_k z^k$ with radius of convergence R and let $A \in M_n$. If $\rho(A) < R$ then the power series $p(A) = \sum_{k=0}^{+\infty} p_k A^k$ converges.*

Corollary 16.13. *Let $A \in M_n$. If there exists a matrix norm such that $\|I_n - A\| < 1$, then $A \in M_n^{-1}$ and the series $\sum_{k=0}^{+\infty}(I - A)^k$ converges to A^{-1}.*

17. Gersgorin Type Theorems

Given $A \in M_n$ we write $A = D + B$ where $D = diag(a_{1,1}, ..., a_{n,n})$ so that B is a matrix with 0 diagonal. Set $R_i^{|prime} = \sum_{j \ne i} |a_{i,j}|$ and let $\Omega_i = \{z \in \mathbb{C} : |z - a_{i,i}| \le R_i^{|prime}\}$.

Theorem 17.1 (Gersgorin)**.** *Let $A = (a_{i,j}) \in M_n$ and let $G(A) = \cup_{i=1}^n \Omega_i$. Then $\sigma(A) \subseteq G(A)$.*

Corollary 17.2. *Let* $C'_j = \sum_{i \neq j} |a_{i,j}|$ *and let* $\Omega' = \{z \in \mathbb{C} : |z - a_{j,j}| \leq C'_j\}$. *Then* $\sigma(A) \subseteq \big(\cup_{i=1}^n \Omega_i\big) \cap \big(\cup_{j=1}^n \Omega'_j\big)$.

The next result needed a little background in contour integration first.

Theorem 17.3. *Let* $A \in M_n$ *and assume that there is a set of* $\{i_1, ..., i_k\} \subseteq \{1, ..., n\}$ *has the property that* $\big(\cup_{l=1}^k \Omega_{i_l}\big) \cap \big(\cup_{j \neq i_l, \forall l} \Omega_j\big)$ *is empty. Then* $p_A(t)$ *has exactly* k *roots in the set* $\cup_{l=1}^k \Omega_{i_l}$.

Note that if we let $S = diag(p_1, ..., p_n)$ with $p_i > 0$, $\forall i$ then $S^{-1}AS = (p_i^{-1} a_{i,j} p_j)$. Hence,

$$\sigma(A) = \sigma(S^{-1}AS) \subseteq G(S^{-1}AS) = \cup_{i=1}^n \{z : |z - a_{i,i}| \leq p_i^{-1} \sum_{j \neq i} p_j |a_{i,j}|\}.$$

In particular, $\sigma(A)$ must be int the intersection of all such sets as we vary $p_1, ..., p_n$.

Corollary 17.4. *Let* $A \in M_n$ *then*

$$\rho(A) \leq \overset{min}{\{p_1, ..., p_n > 0\}} \; \overset{max}{\{1 \leq i \leq n\}} \;\; p_i^{-1} \sum_{j \neq i} p_j |a_{i,j}|.$$

Definition 17.5. A matrix $A = (a_{i,j})$ is **diagonally dominant(d.d)** provided that $|a_{i,i}| \geq \sum_{j \neq i} |a_{i,j}|$, $\forall i$ and **strictly diagonally dominant(s.d.d)** provided that $|a_{i,i}| > \sum_{j \neq i} |a_{i,j}|, \forall i$.

Let $RHP = \{z \in \mathbb{C} : Re(Z) > 0\}$ which we call the **right half plane**.

Theorem 17.6. *Let* $A \in M_n$ *be s.d.d. then:*

(1) $A \in M_n^{-1}$,
(2) *if* $a_{i,i} > 0, \forall i$ *then* $\sigma(A) \subseteq RHP$,
(3) *if* $A = A^*$ *and* $a_{i,i} > 0, \forall i$, *then* A *is positive definite.*

Department of Pure Mathematics, University of Waterloo
E-mail address: vpaulsen@uwaterloo.ca