Algebra of linear maps over vector spaces: Quick reference

vishvAs vAsuki

June 6, 2011

Contents

Part I

Introduction

1 Themes

1.1 References

Based on [4], [5], [1] [2], [3].

1.1.1 Linear algebra over field F

Vector space over field F with all linear transformations.

1.2 Themes

Linear transformations and matrices; their properties. Systems of linear equations. Using decompotions to understand the operation of linear maps better.

1.2.1 Related surveys

For info on vectors, vector spaces, various functionals (including norms), non-linear functions: see survey of vector spaces, their functions and functionals.

For numerical analysis, conditioning, stability, differnce equations, differential equations: see Numerical analysis ref.

1.3 Characterization of research effort

See algorithms ref; Both strategies mentioned there are useful.

Experiment with Matlab.

1.3.1 Matrix algorithm design

The decompoisitional approach to matrix computations is extremely useful. Think of a computation in terms of triangular, diagonal, orthogonal etc.. decompositions. Facilitates error analysis.

1.3.2 Working with algebra

Invlolves much algebra in addition. Reasoning about matrix operations: Understand what is going on by the use of algebra; eg: write scaling rows of A as DA where D is diagonal. Write special matrices algebraically. This highly clarifies things, makes them unambiguous. This is an important skill.

Use induction a lot.

Also see algebra/ mathematical structures ref.

2 Notation

Vectors are small letters. Matrices are capital letters.

2.1 Matrix notation

2.1.1 Dimensions

$A = [a_{ij}]$ is mn matrix: $\in C^{m\times n}$. $q = \min\{m, n\}$. Square A: $m \times m$. $\hat{A}$ is rectangular.

2.1.2 Matrices related to A

$(v_1, v_2 \ldots)$ is a column vector v.

a_i: ith element of vector a or ith column of A. a_i^*: by context: ith row of A or transpose of a_i. $A_{k+1:m,k:f}$: a submatrix; other unambiguous matlab notations $A_{i,:}$ for ith row etc.. $a_{i,j}$ an element of A. $A_{i,j}$: by context: an element of A or a submatrix of A.

Conjugate matrix $\overline{A}$. Adjoint (Hermitian conjugate) of A: $A^* = \overline{A^T}$. $\tilde{A}$: A as stored on computer; or as calculated by alg.

Dilation of matrix A: add rows and cols to A.

2.1.3 Special matrices

Permutation matrix, P. Lower triangular matrix L. Upper triangular matrix U. Diagonal matrix $D = diag(d_i) = diag(d)$. Orthogonal (or Orthonormal) matrix Q, $\hat{Q}$. Identity I.

2.1.4 Special vectors

ith col of I: e_i (Canonical unit vector). e or 1: col vector of 1's. $P_{\perp q}$: projector to space $\perp q$.

2.2 Abused notation

$y = O(\epsilon) \implies \exists c = f(m,n), \forall x \lim_{\epsilon_M \to 0} y \leq c\epsilon_M$. Extra Defn: If $y = \frac{a(x)}{b(x)}$, at $b(x) = 0$, $y = O(\epsilon)$ means $a(x) = O(\epsilon b(x))$.

Part II

Linear operators

3 Linear operators: Matrix representation

3.1 Linear Transformation

3.1.1 Definition and Linearity

3.1.1.1 Linearity

$A(ax + by) = aA(x) + bA(y)$. A called operator, from viewing vector as functions.

3.1.1.2 Mapping between vector spaces

For any field F, can consider linear transformations $A : F^n \to F^m$.

3.1.2 Applications, examples

Vector spaces can model many real world things (see vector spaces ref), even functions. In all of these, linear transformations have deep meanings.

Over function spaces: Differentiation, integration, multiplication by fixed polynomial in P_n.

Geometric operations: Ax.
Rotation, projection, reflection, stretching, shearing.

3.1.3 As a matrix

3.1.3.1 Use action on basis vectors

Take the linear operation A, take standard basis vectors $\{e_i\}$ of dom(A). Take an input x, which, as a combination of basis vectors, is $x = \sum_i x_i e_i$. Now, by linearity, $A(x) = \sum_i x_i A(e_i)$.

So, $A(e_i)$ is the basis of the range space range(A)!

3.1.3.2 Matrix

Now, write a matrix A, a 2-dim array of numbers such that the ith column, $a_i = A_{:,i} = A(e_i)$. This will be a $m \times n$ array. Reason for doing this: Ax now be defined to equal A(x).

3.1.3.3 Matrix * vector multiplication

Define $Ax = \sum_i a_i x_i$. ***Voila - linear operation represented by matrix vector multiplication!***

For other views of Ax, see a later section.

Vector dot product ***This also defines row vector * column vector multiplication!*** This is also the standard inner product.

3.1.3.4 Row view: inner product with rows

Take the standard inner product. Then, Ax is the vector formed by $(\langle A_{i,:}, x\rangle)$.

3.1.4 Changing basis vectors

Representation of the same geometric point can change with the choice of bases.

Take the point $x = \sum_i x_i e_i$, written according to the old standard basis $\{e_i\}$. Express $\{e_i\}$ in terms of the new basis $\{e_i'\}$:, $e_i = \sum_i a_i e_i'$; write it as the vector u_i. Thence, construct the matrix U. Then, $Ux = y$ is the representation of the point in terms of the new basis $\{e_i'\}$.

3.1.5 Representation wrt i/p and o/p bases

Similarly, representation of a linear transformation can also change with choice of bases.

When the right input and output basis is chosen, every matrix is actually diagonal : mere scaling: see SVD.

3.2 Forward operation: Ax

3.2.1 Spaces related to range(A)

3.2.1.1 Range space

As A is linear, $range(A) = \{Ax : \forall x \in dom(A)\}$ is the vector (sub)space $\langle a_i \rangle$ by definition of matrix-vector multiplication. So, aka the column space.

By linearity, A maps $x = x_r + x_n \in C^n, x_n \in N(A), x_r \perp N(A)$ to $Ax_r \in C^m$.

3.2.1.2 Left null space

Also, **orthogonal complement** of $range(A)^{\perp} = N(A^T)$, the Left Null space; both in C^m.

3.2.2 Subspaces of the domain

3.2.2.1 Null space (Kernel)

N(A): $\{x : Ax = 0\}$. By linearity of A, this is a vector space. N(A) dimension = degrees of freedom = free vars' number.

Left null space is $N(A^T)$.

Find null-space Ax=0; Reduce A to U; identify pivot and free vars; find values of pivot vars in terms of f free vars; rewrite as combination of f vectors (basis of null space). If $N(A) = C^0$, every b is unique combo of range(A).

Similarly, Find **left null-space** $N(A^T)$ basis.

3.2.2.2 Row space

This is the space spanned by rows of A, so it is $range(A^*)$. $S \perp N(A)$ wrt standard inner product, $range(A) + N(A) = F^n$.

Every x: $Ax \neq 0$ has a component in the row space.

3.2.2.3 SVD view

See elsewhere.

3.2.3 Rank of A

3.2.3.1 Row and column ranks

The number of linearly independent rows in A is row rank. Similarly column rank is defined. So, column rank = dim(range(A)).

3.2.3.2 Row rank = column rank

Do triangular row elimination to get PA = LU. Then, row rank = number of non-zero rows/ pivots in A.

But, every column a_i of A, corresponding to a 0 pivot, is a linear combination of the non-0 pivot columns: construct the matrix A' with only such columns, and solve $A'x = a_i$ using triangular row elimination. So, column rank = number of non-zero pivots.

3.2.3.3 SVD view

Take SVD: $A = U\Sigma V^*$. Rank r of A corresponds to number of non 0 σ_i; Can take reduced SVD: $A = U_r\Sigma_r V_r^*$. A actually acts between r-dim subspaces.

3.2.3.4 Invertability

Both row space and range(A) must have dimension r. Every x_r in $range(A^T)$ mapped to unique Ax_r. Invertability, I. Every Ax_r in range(A) mapped to unique x_r: Else, if $Ax_{r'} - Ax_r = 0$, $x_{r'} - x_r$ is both in row space and N(A). Also, $A \in C^{m \times n}$ $(m \geq n)$ full column rank iff no x_r mapped to same Ax_r. So, A invertible (both as a function and group theoretically) iff r=m=n. So unique decomposition into basis vectors.

Else, use pseudoinverse to get left inverse if $m > n$, right inverse if $n > m$. To invert Ax=b for $n > m$, identify n-m linearly dependent columns in A, set corresponding $x_i = 0$, drop those columns to get A', solve for x' is $A'x' = b$.

3.2.4 Identity operation

$Ix = x$.

3.3 Matrix multiplication: AB

3.3.1 Composition of transformations

AB is defined to conform to transformation ABx. So, ABCx is associative.

3.3.1.1 As sum of rank 1 matrices

$u \otimes v = uv^T$ makes Rank 1 matrix. $AB = a_1 b_1^* + .. + a_p b_p^*$, where $b_i^* = B_{i,:}$. This ensures that $ABx = \sum_i (B_{i,:}x) a_i$, as intended! ***Remarkable!***

Similarly, for diagonal matrix D: $ADB = \sum_i a_i d_{i,i} b_i^*$. So, for symmetric $S = U\Sigma U^T$, $A^T SB = A^T U \Sigma U^T B = \sum \sigma_i (U^T a_i)^T (U^T b_i) = \sum \sigma_i a_i^T b_i$.

3.3.2 Elementwise description

From sum of rank 1 matrices form: $(AB)_{i,j} = A_{i,:} B_{:,j}$.

3.3.2.1 Columns of the result

A acts on B's rows; B acts on A's columns. Every col of C=AB is a linear combination of A's cols according to some col of B: $c_i = Ab_i$.

3.3.2.2 Form of ABC

Consider expansion of quadratic functional (see vector spaces ref). Similarly, in D = ABC has $D_{i,j} = A_{i,:}(BC_{:,j})$.

So, $D_{i,j}$ is a linear combination of $B_{i,j}$.

3.3.3 Computation

Needs $O(n^3)$ naively. Else needs $O(n^{2.7})$ by Strassen alg. If sparse need $(n^2 \nu)$.

3.3.4 Rank

rank(AB) $\leq$ min (rank(A), rank(B)): Take A=LU factorization, as AB = LUB, rank(AB) $\leq$ rank(B).

3.3.5 Inverse and transpose

$(AB)^{-1} = B^{-1}A^{-1}$ (Take ABx to x.). $(AB)^* = B^*A^*$ from defn.

So, as $I = (AA^{-1})^* = (A^{-1})^*A^*$, $(A^{-1})^* = (A^*)^{-1}$ (aka A^{-*}). Also, $(AA^*)^* = AA^*$ and A^*A Hermitian.

3.3.6 Block matrices

Block multiplication works.

3.4 Other matrix products

3.4.1 Entrywise product

Aka Hadamard product, Schur product. $A.A$.

3.4.2 Kronecker product

Aka outer product. A $m \times n$, B $p \times q$; $C =$A$\otimes B$ is $mp \times nq$ block matrix with $C_{i,j} = A_{i,j}B$.

From defn, $\exists A, B : A \otimes B \neq B \otimes$A; $(A \otimes B)^T = B^T \otimes C^T; A \otimes (B + C) =A\otimes B + A \otimes C; aA \otimes bB = abA \otimes B$. $(A \otimes B)(C \otimes D) = AC \otimes BD$ by block multiplicity. So, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$. Also, using $QA = LDU$: $rank(A \otimes B) = rank(A)rank(B)$.

λ vector: $\lambda(A \otimes B) = \lambda(A) \otimes \lambda(B)$: take eigenpairs of A and B: $(\lambda_i, v_i), (\mu_j, u_j)$; $(A \otimes B)(v_i \otimes u_j) = Av_i \otimes Bu_j = \lambda_i \mu_j (v_i \otimes u_j)$.

3.5 Inverse operation: Solve Ax = b for x

3.5.1 Solvability Assumption

$b \in range(A)$.

3.5.2 Left and right inverses

Right inverse: $I = AA^{-1}$. Similarly, left inverse is defined.

3.5.2.1 Existance conditions

The left inverse exists exactly when you can solve Ax = b for all $b \in range(A)$. Right inverse exists when you can solve $x^TA = b^T \forall b \in range(A^T)$.

3.5.2.2 Equivalence if both exist

If both left and right inverses exist, they're the same: $B_l^{-1}BB_r^{-1} = B_l^{-1} = B_r^{-1}$.

3.5.3 Solutions

3.5.3.1 Full column rank

If A has full column rank, you have a unique solution, left inverse exists, $x = A_l^{-1}b$. Proof: Triangularization by row elimination goes through.

3.5.3.2 Rank Defective matrix

If A is column-rank deficient, eg: short and fat, you have an overdetermined set of linear equations: many 'equally good' solutions exist, but you may want one with certain properties (like sparsity). See numerical analysis ref for details.

3.5.3.3 Finding solutions

See numerical analysis reference for various techniques.

3.5.3.4 Finding inverses

Find right inverse: Gauss-Jordan: Make augmented matrix A:I; use row operations to reduce it to $I : A^{-1}$. So, $\exists A^{-1}$ (right) iff r=m. Similar trick for left inverse.

Or make cofactor matrix C, use $A^{-1} = \frac{C^T}{det(A)}$.

3.5.4 Block matrices

3.5.4.1 Block Gaussian elimination

Take $X = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, solve $X \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix}$ for $\begin{bmatrix} x \\ y \end{bmatrix}$, thence derive X^{-1}. Solving for x in the top equation and substituting it in the bottom, you will have reduced the problem to:
$\begin{bmatrix} I & A^{-1}B \\ 0 & S=D-CA^{-1}B \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A^{-1} & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$.

3.5.4.2 Block LU

Thence, get block LU, Aka Leibnitz factoriazation:
$X = \begin{bmatrix} A & 0 \\ C & I \end{bmatrix} * \begin{bmatrix} I & A^{-1}B \\ 0 & S=D-CA^{-1}B \end{bmatrix}$.

3.5.4.3 Schur complement

S is called the Schur complement of A in X.

3.5.4.4 Inverse of X

Do block back substitution to get X^{-1}.

3.5.5 Pseudoinverse for long thin A

Projector + inverse. For rectangular, non column rank deficient matrices: $m \geq n$. Takes $b = A(x_r + x_n)$ to x_r in row space, cannot revert $Ax_n = 0$ ($x_n \in N(A)$).

$A^+ = (A^*A)^{-1}A^* = \hat{R}^{-1}\hat{Q}^*$ (as $A = \hat{R}\hat{Q}$) $= V\hat{\Sigma}^{-1}\hat{U}^*$ (from SVD).

3.6 Restriction to a subspace S

Let Q be the orthogonal matrix formed by an orthonormal basis of S. Then, projection of a_i in S is QQ^Ta_i. So, $Ax = \sum_i x_ia_i$, when restricted to S, becomes $Ax = \sum_i x_iQQ^Ta_i$. So, QQ^TA is the operator A restricted to the subspace S.

3.7 Submatrices

Principle submatrices: $A_{1:k,1:k}\forall k$. For each principle submatrix, schur complement is defined: see determinant section.

4 Approximating complete and incomplete matrices

4.1 Approximating a matrix

4.1.1 The problem

Take set of matrices S. Want $\arg\min_{B\in S} \|A - B\|$ is minimized wrt some $\|\|$ and some set S.

4.1.1.1 The error metric

Also, often $\|.\|$ is an operator norm, as this ensures that $\|(A - B)x\|$ is low wrt the corresponding vector norm. Other times, as in the case of matrix completion problems, it may be desirable for $\|\|$ to be the Frobenius norm.

4.1.1.2 Low rank (k) factorization problem

In many problems, S is the set of rank k matrices, where k is small.

Often, we prefer $A \approx B = XY^T$ rather than computing B, where rank(X), rank(Y) $\leq k$: A may be sparse, but the best B may be dense, so we may run out of memory while storing, and out of time while computing Bx. You can always get $B = XY^T$ from B just by taking SVD of B.

Restriction to subspace view Similarly, may want restriction of A to a low rank subspace: $B = QQ^TA$, Q is low rank, orthogonal; but it should be the best subspace which minimizes the error. A good Q tries to approximate the range space of A.

4.1.1.3 Sparse approximation

Maybe want $\min_B \|B - A\|_2 : \|B\|_0 \leq k$. This can be solved just by setting B $= A$, and then dropping the k smallest $B_{i,j}$ from B.

1 norm regularized form Even the optimization problem $\min_B \|B - A\|_2^2 + l\|B(:)\|_1$ leads to sparse solutions. This form is useful in some sparse matrix completion ideas. Solution: just take B=A, and then drom $B_{i,j} \leq l/2$: ***thresholding!***

Pf: $\min_B \sum_{i,j}(B_{i,j} - A_{i,j})^2 + l\sum_{i,j} sgn(B_{i,j})B_{i,j}$. Let B' be the solution to this. If $A_{i,j} \geq 0$, so is $B_{i,j}$; If $A_{i,j} \leq 0$, so is $B_{i,j}$: if they oppose in sign, setting $B_{i,j} = 0$ definitely lowers the objective. So, get equivalent problem: $\min_B \sum_{i,j}(B_{i,j} - A_{i,j})^2 + l\sum_{i,j} sgn(A_{i,j})B_{i,j}$ subject to $sgn(A) = sgn(B)$. The new objective is differentiable. Its optimality condition: $B_{i,j} = A_{i,j} - l/2$; but the feasible set only includes $sgn(B) = sgn(A)$. So if $A_{i,j} - l/2 \leq 0$, the feasible B closest to the minimum of $\min_B \sum_{i,j}(B_{i,j} - A_{i,j})^2 + l\sum_{i,j} sgn(A_{i,j})B_{i,j}$ has the corresponding $B_{i,j} = 0$.

4.1.2 Best rank t approximation of A from SVD

4.1.2.1 The approximation

Let $A_t = \sum_{j=1}^{t} \Sigma_j u_j v_j^*$ be the approximation. A_t is the best rank t approx to A (wrt $\|.\|_2$, $\|.\|_F$), captures max energy of A possible: $\|A - A_t\| = \inf_B \|A - B\|$.

4.1.2.2 Approximation error

Then, approximation error is $A - A_t = \sum_{i=t+1}^{r} \sigma_i u_i v_i^*$. As $\|X\| = \sigma_1(X)$: $\|A - A_t\|_2 = \sigma_{t+1}$, $\|A - A_t\|_F = \sqrt{\sum_{i=t+1}^{r} \sigma_i^2}$.

4.1.2.3 Geometric interpretation

Approximate hyperellipse by line, 2-dim ellipse etc...

Approximating domain and range spaces Using SVD for example, A can be viewed as a combination of rotation and diagonal matrices. So, getting a low rank approximation of A can be viewed as first getting low rank approximations of the range and domain spaces with orthogonal basis matrices U_t and V_t respectively, and then finding a square $S = U_t^T A V_t^T$ such that $A \approx U_t S V_t$.

4.1.2.4 Proof

Proof by $\Rightarrow\Leftarrow$: dim(N(B))= r-t, Let $\forall w \in N(B), \|Aw\| = \|(A - B)w\| < \Sigma_{t+1}\|w\|$; but $\exists$ t+1 subspace $\{v : \|Av\| \geq \Sigma_{t+1}\}$; $dim(\{w\})+\text{dim}(\{v\}) = r+1$: absurd.

So, $\sigma_k = \min_{S \subset C^n, dim(S)=n-k+1} \max_{x \in S} \|Ax\|_2$
$= \max_{S \subset C^n, dim(S)=k} \min_{x \in S} \|Ax\|_2$.

4.1.2.5 Randomized Approach

Take the $A \approx B = QQ^T A$ view. A good Q must span U_k. Can use something akin to the power method of finding ev. Take random m*k matrix W. $Y = (AA^T)^q AW = U\Sigma^{q+1}VW$, and get $Y = QR$ to get Q. q = 4 or 5 is sufficient to get good approximation of U_k. (Tropp etal) if you aim to get k+p approximation, you get low expected error.

4.1.3 With few observations in A only

Same as the missing value estimation problem.

4.2 Missing value estimation

4.2.1 The problem

May be you have the matrix A, and you have observed only a few entries O. If there were no other conditions, there would be infinite solutions. But, maybe you also know that A has some special structure: low rank, or block structure or smoothness etc..

Can also view as getting an approximation $B \in S$ for A, where S is the set of matrices having the specified structure.

4.2.2 Smoothness constraint

Sovle min $\sum_{i,j} d(B_{i,j}, B_{i-1,j}) + d(B_{i,j}, B_{i,j-1})$ such that $B_{i,j} = A_{i,j} \forall (i,j) \in O$. If we use l2 or l1 metrics for d(), this can be solved with convex optimization.

4.2.3 Low rank constraint

$rank(B) \leq k$: this is not a convex constraint.

4.2.3.1 Singular value projection (SVP)

(Jain, Meka, Dhillon) Set $B^{(0)}_{i,j} = A_{i,j} \forall (i,j) \in O$, set remaining values in $B^{(0)}$ arbitrarily. Then, in iteration i, do SVD of $B^{(i)}$ to get rank k approximation $B^{(i+1)}$, set $B^{(i+1)}_{i,j} = A_{i,j} \forall (i,j) \in O$.

Projection viewpoint Take $S_1 = \{B : B_{i,j} = A_{i,j} \forall (i,j) \in O\}$,
$S_2 = \{B : rank(B) \leq k\}$. You start with $B_0 \in S_1$, project it to the closest $C \in S_2$, project C to the closest $B_1 \in S_1$ etc..

4.2.4 Applications

The netflix problem.

5 Important decompositions

5.1 Importance of decompositions

Very important in restating and understanding the behavior of a linear operator. Also, important in solving problems: get decomposition, use it repeatedly. For algebraic manipulation: Factor the matrix: QR, LU, Eigenvalue decomposition, SVD.

5.2 EW revealing decompositions

See section on eigenvalues.

5.2.1 Eigenvalue decomposition

Aka Spectral Decomp. Only if A diagonalizable.

Take S: Eigenvectors-as-columns matrix, with independent columns; Λ: Eigenvalue diagonal matrix. Then, AS = SL; So, $S^{-1}AS = L$; $A = SLS^{-1}$: a similarity transformation. Also, If AS=SL, S's columns must be eigenvectors.

A diagonalized into Λ. A and Λ are similar.

5.2.1.1 Non-defectiveness connection

$\exists S^{-1}\Lambda S$ iff A is non defective: If $\exists S^{-1}\Lambda S$: Λ diagonal, non defective, so A non defective; if A non defective: can make non singular S; thence $S^{-1}\Lambda S$.

5.2.1.2 Left ev

$S^{-1}AS = \Lambda$, so $S^{-1}A = \Lambda S^{-1}$. So, the rows of $S^{-1} = L$ are the left ev.

5.2.1.3 Change of basis to ev

$x = SS^{-1}x = SLx = \sum_i \langle x, L_{i,:}\rangle s_i$. Thus, any x can be conveniently rewritten in terms of right ew, with magnitudes of components written in terms of left ew.

5.2.1.4 Unitary (Orth) diagonalizability

For $A = A^*$, $A = -A^*$, Q etc..

A unitarily diagonalizable iff it is normal: $A^*A = AA^*$: From uniqueness of SVD, $US^2U^* = VS^2V^*$; so, $|U| = |V|$; U and V may only differ in sign. So, for some $|\Lambda| = |S|$,A= ULU^*. Aka Spectral theorem.

5.2.2 $A = QUQ^*$ factorization

(Schur). A and upper traingular U similar; all ew on U's diagonal. If $A = A^*$, $U = U^*$, so U is diagonal.

Every A has QUQ^*: by induction: assume true for m; take any λ and corresponding eigenspace $V_\lambda \perp V_\lambda^\perp$; use orth vectors from these spaces as bases; in this basis, operator represented by A has matrix representation $A' = \left[\begin{smallmatrix} \lambda I_\lambda & B \\ O & C \end{smallmatrix}\right] = Q^* A Q$; can then repeat the process for C.

5.3 Singular Value Decomposition (SVD)

5.3.1 Reduced (Thin) SVD

If $m \times n$ A with $m > n$, rank r = n, unitary $n \times n$ $V = [v_i]$, unitary $m \times n : \hat{U} = [u_i], n \times n : \hat{\Sigma} =$ diagonal matrix with $\sigma_i \geq 0$ in descending order, $AV = \hat{U}\hat{\Sigma}$; then $A = \hat{U}\hat{\Sigma}V^*$. If $r < n$; still get reduced SVD by padding V with $\perp$ vectors, $\hat{U}$ with appropriate $\perp$ vectors, Σ with 0 diagonalled columns.

5.3.2 Full SVD

Pad $\hat{U}$ with m-r $\perp$ vectors, $\hat{\Sigma}$ with 0's to make U $m \times m$, invertible; V stays same; $A = U\Sigma V^*$. So, SVD for $m < n$: $A^* = V\Sigma U^*$. So, $A^*U = V\Sigma$. Also, $Av_i = u_i\sigma_i$. So, $range(A) = u_1 \ldots u_r$, $range(A^*) = v_1 \ldots v_r$, $N(A) = v_{r+1} \ldots v_m$ ($Av_{r+1} = 0$), $N(A^*) = u_{r+1} \ldots u_n$.

5.3.3 Geometric view

Take $Ax = U\Sigma V^* x$: V^* rotates the unit ball to unit ball: $v_i \to e_i$, Σ stretches unit ball along axes to make it hyperellipse, U rotates it. Every A with SVD maps unit ball to hyperellipse (Eqn: $\sum \frac{x_i^2}{\sigma_i^2} = 1$): Orthonormal $\{v_i\}$ (left singular vectors) mapped to orthogonal $\{u_i\sigma_i\}$ (Principle semiaxes, orthonormal right singular vectors $\times$ singular values). So $\sigma_1 = \|A\|_2$, $\sigma_n = \left\|A^{-1}\right\|_2$.

From geometric action of $U\Sigma V^* x$, every A with SVD is a diagonal matrix when domain and range bases are altered (See $Ax = b$ as $AVx' = Ub'$, then $\Sigma x' = b'$). 'If ye want to understand A, take its SVD.'

5.3.4 Existance

Every A has a SVD: by induction; prove 1*1 case; assume (m-1)*(n-1) case $B = U_2\Sigma_2V_2^*$; take m*n A; $\sigma_1 = \|A\|_2 = \sup \|Av\|_2$; so $\exists v_1$ in the unit ball with $Av_1 = u_1\sigma_1$; So extend v_1 and u_1 to orthonormal V_1 and U_1, make Σ_1 solving $U_1^* A V_1 = \Sigma_1$; 1st col is σ_1; as $\|A\| = \|\Sigma\| = \sigma_1$, non-diag elements of 1st row gotto be 0; let rest of $\sigma_1 = B$.

5.3.5 Conditional uniqueness up to a sign

SVD is unique if $\{\sigma_i\}$ unique ('up to sign'): write Hyperellipse ellipse semiaxes in descending order; but can reverse sign of u_i, v_i or can multiply them with any suitable pair of complex roots of 1.

5.3.6 Singular value properties

See another section.

5.3.7 Finding SVD using EVD

Use eigenvalue decompositions: $AA^* = U\Sigma^2U^*$, and $A^*A = V\Sigma^2V^*$. Otherwise, find eigenvalue decomposition of $B = \left[\begin{smallmatrix} 0 & A \\ A^T & 0 \end{smallmatrix}\right]$: then ew(A) are composed of zeros and sw(A) repeated with different signs. ev of B is $(\sqrt{2})^{-1}\begin{bmatrix} U_n & V \\ \sqrt{2}U_{m-n} & 0 \end{bmatrix}$.

5.3.8 Polar decomposition

$m \leq n$: take SVD $A = U[\Sigma\ 0][V_1\ V_2]^* = U\Sigma V_1^*$,
$P^2 = AA^* = U\Sigma^2U^*$: +ve semidefinite; take $P = U\Sigma U^*$: Hermitian +ve semidefinite. So, $A = U\Sigma V_1^* = PUV_1^* = PY$, where Y has orthonormal rows.

So, if $m \geq n$: $A = YQ$ for Hermitian +ve semidefinite Q, Y with orth columns: apply thm to A^*.

5.4 PA = LU

Here unit lower triangular L, upper triangular U.

Can also make: PA = LDU: For , unit upper triangular U, diagonal D.

5.4.1 Existance and uniqueness

5.4.1.1 Existance

See triangularization by row elimination algorithm in numerical analysis ref. That this runs to completion proves existance.

5.4.1.2 Uniqueness if P=I

A = LU unique: Else if $LU = L'U'$, $L'^{-1}L = U'U^{-1}$: absurd. So A=LDU unique.

5.4.2 For hermitian positive definite matrices

As $A \succeq 0$, P = I.

As $A = LDU = A^*$, can take $A = RR^*$, where R = $LD^{1/2}$ (Cholesky). It is also unique: $r_{j,j} = \sqrt{d_{j,j}} > 0$ fixed by definition; it inturn fixes rest of R.

5.4.3 Importance

Very useful in solving linear equations invloving the same matrix A: can store L, U for repeated reuse.

5.5 A = QR = LQ'

Express columns of A as linear combinations of orthogonal $\{q_i\}$. For proof of existance, see triangular orthonormalization algorithm in numerical analysis ref.

Taking the QR factorization of A^T, you also get $A = LQ^T$, where L is lower triangular.

5.5.1 Importance

Often, we need to get an orthogonal basis for range(A).

5.5.2 Column Rank deficient A

If A were rank deficient, multiple columns would be linear combinations of same set of q_i's. As Q is square, we would have 0 rows .

5.5.2.1 Rank revealing QR

In such cases, we can always assume that the 0 rows appear at the bottom, revealing the rank.

5.6 Factorization of Hermitian matrices

5.6.1 Unitary diagonalizability, SVD

From Schur factorization: Can write $A = Q\Lambda Q^*$. So, has full set of orthogonal eigenvectors. So, can write: $A = \sum_i \lambda_i q_i q_i^*$.

Also, singular values $s_i = |\lambda_i|$, but can't write $A = U\Sigma V^* = U\Sigma U^*$: there may be sign difference between U and V's columns due to $\lambda_i < 0$.

5.6.2 Symmetric LDU factorization

(Cholesky). $A = R^*R$. As $A = LDU^* = UDL^*$, $L = U^*$. So, $A = LDL^* = LD^{1/2}D^{1/2}L^* = R^*R$; $d_{j,j} > 0$ as $a_{j,j} > 0$; $r_{j,j} = \sqrt{d_{j,j}} > 0$ chosen.

By SVD, $\|R\|^2 = \|A\|$.

5.6.3 Square root of semidefinite A

$A = (A^{1/2})^* A^{1/2}$. Diagonalize, get $A = QLQ^*$, $A^{1/2} = QL^{1/2}Q^*$: the unique +ve semidefinite solution.

6 Special linear operators

6.1 Orthogonal (Unitary) m*n matrix

Columns orthonormal: $Q^*Q = I$; and $m \leq n$.

6.1.1 Change of basis operation

Qx=b: $x = Q^*b$: so, x has magnitudes of projections of b on q's: Change of basis op.

Alternative view: $Q^*(\sum a_i q_i) = \sum a_i e_i$.

6.1.1.1 Preservation of angles and 2 norms

So, $\langle Qa, Qb \rangle = b^*Q^*Qa = \langle a, b \rangle$. Also, $\|Qx\| = \|x\|$: So, length, angle preserved; analogous to $z \in C$, with $|z| = 1$. If $\|Qx\| = \|x\|$, $Q^*Q = I$.

6.1.2 Square Unitary matrix

6.1.2.1 Orthogonality of rows

If Q square, even rows orthogonal: $Q^*Q = I \implies QQ^* = I$: $q_i^* q_j = \Delta_{i,j}$ (Kronecker $\Delta = 1$ iff i=j, else 0.).

6.1.2.2 Rotation + reflection

Because of its being a change of basis operation, by geometry, Q is rotation or reflection or a combination thereof.

The distinction between orthogonal matrices constructed purely out of rotation matrices (proper rotation), and those involving orthogonal matrices which involve reflections (improper rotation) is important in geometric computations: in applications such as robotics, computational biology etc..

6.1.2.3 Determinant

$det(Q) = \pm 1 : det(Q^*Q) = det(I) = 1$.

Q is a rotation if $|Q| = 1$ or a reflection if $|Q| = -1$: True for m=2; For $m > 2$, see that determinant is multiplicative.

6.1.2.4 Permutation matrix P

A permuted I. Permutes rows (PA) or columns (AP). Partial permutation matrix: every row or column has ≤ 1 (maybe 0) nz value.

6.1.2.5 Rotation matrix

To make a rotation matrix, take new orthogonal basis (u_i): the coordinate system is rotated, $e_i \to u_i$, get matrix U. $(U^*x)_i = u_i^*x$: x rotated. Note: $U^*u_i = e_i$.

6.1.2.6 Reflection matrix

Take reflection accross some basis vector (not any axis). This is just I with -1 instead of some 1.

6.2 Linear Projector P to a subspace S

6.2.1 Using General projection definition

See definition of the generalized projection operation in vector spaces ref. Here, we consider the case where projection happens to be a linear operator: that is, it corresponds to the minimization of a convex quadratic vector functional, where the feasible set is a vector space, the range space of the projector P.

6.2.2 Definition for the linear case

P such that $P^2 = P$: so, a vector already in S is not affected.

(I-P) projects to N(P): If Pv=0, (I-P)v=v; vice versa. Rank r projectors project to r dimension space.

Oblique projectors project along non orthogonal basis.

6.2.3 Orthogonal projector

Here, $(I-P)x \perp Px$: Eg: projectors which arise from solving the least squares problem. Ortho-projectors $\neq$ orthogonal matrices.

If P=P*, P orth projector: If P=P*, $\langle (I-P)x, Px\rangle = 0$. If P orth proj; make orthonormal basis for range(P), N(P); get Q; now $PQ = Q\Sigma$, with σ_i 1 or 0 suitably: SVD! So, if P orth proj, P=P*.

Ergo, (I-P) also orth proj. Also, $P = \hat{Q}\hat{Q}^*$ (As $A = \hat{Q}\hat{R}$): Also from $v = r + \sum(q_i q_i^*)v$. All $\hat{Q}\hat{Q}^*$ orth proj: satisfy props.

$\|P\| = 1$. [**Find proof**]

6.3 Hermitian matrix

Aka Self Adjoint Operator. Symmetric matrix: $A = A^T$. It generalizes to Hermitian matrix $A = A^*$; analogous to $R \subseteq C$. Not all symmetric matrices are Hermitian.

Notation: Symmetric matrices in $R^{n\times n}$: S^n; +ve definite among them: S^n_{++}.

Skew/ anti symmetric matrix: $A = -A^T$, generalizes to skew Hermitian.

$\langle Ax, y\rangle = y^*Ax = \langle x, A^*y\rangle$.

6.3.1 Importance

Appears often in analysis: Any $B = \frac{B+B^*}{2} + \frac{B-B^*}{2}$: Hermitian + Skew Hermitian. Also in projector.

Many applications: Eg: Covariance matrix, adjascency matrix, kernel matrix.

6.3.2 Self adjointness under M

For SPD A, M, $M^{-1}A$ self adjoint under M as $\langle x, M^{-1}Ay\rangle_M = y^*Ax = \langle M^{-1}Ax, y\rangle_M$.

6.3.3 Eigenvalues (ew)

All eigenvalues real:
$\bar{l}x^*x = (lx)^*x = (Ax)^*x = x^*(Ax) = lx^*x$, so $\bar{l} = l$.

ev of Distinct ew are orthogonal: $x_2^*Ax_1 = l_1x_2^*x_1 = l_2x_2^*x_1, \therefore x_2^*x_1(l_1 - l_2) = 0$.

6.3.3.1 sw and norm

Unitary diagonalizability possible for A: see section on factorization of hermitian matrices. Thence, $|\lambda(A)| = \sigma(A)$; so $|\lambda_max(A)| = \|A\|$.

6.3.4 Factorizations

For details, see section on factorization of hermitian matrices.

6.3.5 Skewed inner prod x^*Ax

$x^*Ay = (y^*Ax)^*$. So, $x^*Ax = \overline{x^*Ax}$; so x^*Ax real.

6.4 +ve definiteness

6.4.1 Definition

If $\forall 0 \neq x \in C^n : x^*Ax \in R; x^*Ax \geq 0$, A +ve semi-definite, or non-negative definitene.

If $x^*Ax > 0$, +ve definite: $A \succ 0$.

Similarly, -ve (semi-)definite defined.

6.4.1.1 Importance

Important because Hessians of convex quadratic functionals are +ve semidefinite. Also, it is importance because of its connections with ew(A).

6.4.2 +ve semidefinite cone

The set of +ve semidefinite matrices, is a proper cone. If restricted to symmetric matrices, get S_+^n.

6.4.2.1 Matrix inequalities

Hence, inequalities wrt the cone defined. ***Can write*** $A \succeq 0$ ***to say*** A ***is +ve def.*** This is what is usually assumed by $\succeq$ when dealing with +ve semidefinite matrices - not elementwise inequalities.

Linear matrix inequality (LMI) $A_0 + \sum x_i A_i \preceq 0$. Note that this is equivalent to having $A \preceq 0$ with $A_{i,j} = a_{ij}^T x + b_{ij}$ form. Used in defining semidefinite programming.

6.4.2.2 Analogy with reals

Hermitians analogous to R, +ve semidef like $\{0\} \cup R^+$, +ve def like R^+ in $\mathbb{C}$.

6.4.2.3 Support number of a pencil

$s(A, B) = \arg\min t : tB - \mathrm{A} \succeq 0$.

6.4.3 Non hermitian examples

Need not be hermitian always.
Then, as $x^* Bx = x^* B^* x$, anti symmetric part in $B = \frac{B+B^*}{2} + \frac{B-B^*}{2}$ has no effect.

6.4.4 Diagonal elements, blocks

$e_i^* A e_i = a_{i,i}$. So, $a_{i,i}$ real. $a_{i,i} \geq 0$ if A +ve semidefinite; $a_{i,i} > 0$ if A +ve definite; but converse untrue.

Similarly, for $X \in C^{m \times n}$ invertible: $X^* AX$ has same +ve definiteness as A. Taking X composed of e_i, any principle submatrix $A_{i,i}$ can be writ as $X^* AX$; so $A_{i,i}$ has same positive definiteness as A.

6.4.4.1 Off-diagonal block signs invertible

Off-diagonal block signs are invertible without loosing +ve semidefiniteness. Pf: If $\begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \geq 0 \forall \begin{bmatrix} x \\ y \end{bmatrix}$, then $\begin{bmatrix} x^T & y^T \end{bmatrix} \begin{bmatrix} A & -B \\ -C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \geq 0 \forall \begin{bmatrix} x \\ y \end{bmatrix}$.

6.4.5 Eigenvalues: Real, +ve?

$\forall i : \lambda_i \in R$: take ev x, must be able to compare $x^T Ax = \lambda x^* x$ with 0.

If $A \succeq 0$, ew $\lambda_i \geq 0$: $\lambda_i x^* x = x^* Ax \geq 0$. Also, if $A \succ 0$, $\lambda_i > 0$.

6.4.5.1 Determinant

$det(A) = \prod \lambda_i \geq 0$.

6.4.6 +ve inner products

For +ve definite matrices, get +ve inner products: Take eigenvalue decompositions: $A = \sum_i \lambda_i q_i q_i^T, B = \sum_i l_i p_i p_i^T$.

So, the +ve definite cone is self dual.

6.4.7 Invertibility

If A +ve def., A is invertible: $\forall x \neq 0 : x^*Ax \neq 0$, so $Ax \neq 0$; so A has no nullspace. If A +ve semi-def, can't say this.

6.5 Hermitian +ve definiteness

Also, see properties of not-necessarily symmetric +ve semidefinite matrices.

6.5.1 From general matrices

Any B^*B or BB^* hermitian, +ve semidefinite: $x^*B^*Bx = \|Bx\|$. So, if B invertible, B^*B is +ve definite. So, if B is long and thin, B^*B is +ve definite, but if B is short and fat: so singular, B^*B is +ve semi-definite, also singular.

6.5.2 Connection to ew

If $A = A^*$, all eigenvalues $l > 0$, then A is +ve definite: $x^*Ax \in R$, $x^*Ax = x^*U\Lambda U^*x = \sum \lambda_i x_i^2$.

Magnitudes of ew same as that of sw: as you can easily derive SVD from eigenvalue decomposition. So, singular value properties carry over.

6.5.3 Connection to the diagonal

6.5.3.1 Diagonal dominance

If $A = A^*$, diagonal dominance and non-negativity of $A_{i,i}$ also holds, then A is +ve semidefinite. See diagonal dominant matrices section for proof.

6.5.3.2 Diagonal heaviness

The biggest element of A is the biggest diagonal element. For some k : $a_{k,k} \geq a_{i,j} \forall i, j$. Pf: Suppose $a_{i,j} > a_{k,k}$; then consider submatrix $B = \left[\begin{smallmatrix} a_{i,i} & a_{i,j} \\ a_{i,j} & a_{j,j} \end{smallmatrix}\right]$; $B \succeq 0$, but due to assumption $|B| \leq 0$, hence $\Rightarrow\Leftarrow$.

6.5.4 Check for +ve (semi) definiteness

Do Gaussian elimination, see if pivots ≥ 0. $x^*Ax = x^*LDL^*x$, if pivots good, can say $= \left\|D^{1/2}L^*x\right\| \geq 0$.

6.5.5 Block matrices: Schur complement connection

Take $X = \left[\begin{smallmatrix} A & B \\ B^T & C \end{smallmatrix}\right]$, $S = C - B^TA^{-1}B$. Then, if $A \succ 0$, $X \succeq 0 \equiv S \succeq 0$. Also, $X \succeq 0 \equiv (A \succ 0 \wedge S \succ 0)$.

6.5.5.1 Proof: rewrite as optimization problem

Take $f(u, v) = u^TAu + 2v^TB^Tu + v^TCv = \left[\begin{smallmatrix} u^T & v^T \end{smallmatrix}\right] X \left[\begin{smallmatrix} u \\ v \end{smallmatrix}\right]$. Solve $\min_u f(u, v)$. By setting $\nabla_u f(u, v) = 0$, get minimizer $u' = -A^{-1}Bv$, $f(u', v) = v^TSv$.

6.5.6 +ve semidefinite cone

Denoted by S^n_{++} and S^n_+.

6.5.6.1 Self duality

If $A, B \succeq 0$, $\langle A, B\rangle = \langle \sum_i \lambda_i q_i q_i^*, \sum_j \lambda_i' q_j q_j^*\rangle \geq 0$. So, dual of S^n_{++} is itself.

When you consider the dual of a semidefinite program, this is important.

6.6 Speciality of the diagonal

6.6.1 Diagonally dominant matrix

$|A_{i,j}| \geq \sum_{j \neq i} A_{i,j}$.

6.6.1.1 Hermitian-ness and +ve semidefiniteness

A hermitian diagonally dominant matrix with non-negative diagonal elements is +ve semi-definite. Pf: Take $x^TAx = \sum_{i,j} A_{i,j}x_ix_j \geq \sum |A_{i,j}|(x_i - x_j)^2$. The decomposition reminds one of properties of the graph laplacian. Alternate pf: take ev u, taking $Au = \lambda u$, show $\lambda \geq 0$.

If symmetry condition is dropped, +ve semidefiniteness need not hold.

6.7 Other Matrices of note

6.7.1 Interesting matrix types

Block matrix; Block tridiagonal matrix.

6.7.2 Triangular matrix

Inverse of L, L' is easy to find: $L'_{i,i} = L^{-1}_{i,i}; L'_{i,j} = -L^{-1}_{i,j}$.

ew are on diagonal.

6.7.3 Polynomial matrix

$P = \sum A(n)x^n$: Also a matrix of polynomials.

6.7.4 Normal matrix

$A^*A = AA^*$. By spectral thm, $A = Q\Lambda Q^*$. Exactly the class of orthogonally diagonalizable matrix.

Let $a = (a_{i,i}), \lambda = (ew_i)$: By direct computation, $a = S\lambda$, where $S = [|q_{ij}|^2] = Q.\bar{Q}$ is stochastic.

6.7.5 Rank 1 perturbation of I

$A = I + uv^*$. Easily invertible: $A^{-1} = I + auv^*$ for some scalar a.

6.7.6 k partial isometry

$A = U\Sigma V^*$ with $\Sigma = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}$.

6.8 Positive matrix A

6.8.1 Get Doubly stochastic matrix

This is important in some applications: like making a composite matrix from the social and affiliation networks.

If $A = A'$, This can be done by first dividing by the largest entry, and then adding appropriate entries to the diagonal.

Can do Sinkhorn balancing: iteratively a] do row normalization, b] column normalization.

6.9 Stochastic matrices

6.9.1 Definition

If A is stochastic, A1 = 1, $A \geq 0$.

6.9.2 Eigenspectrum and norm: Square A

6.9.2.1 1 is an ev

If A is stochastic, A1 = 1. So, 1 is an ew, and 1 is the correspondingn ev.

By Gerschgorin thm, $\lambda(A) \in [-1, 1]$, so, $\lambda_m ax = 1$.

If A is also real and symmetric, can get SVD from eigenvalue decomposition, and $\sigma_m ax = \|A\|_2 = 1$.

6.9.3 Product of stochastic matrices

So, if A, B stochastic, AB1 = 1, $1^*AB = 1^*$: AB also stochastic.

6.9.4 Doubly Stochastic matrix S

S is bistochastic if $S \geq 0, S1 = 1, 1^*S = 1^*$.

(Birkhoff): $\{S\}$ = set of finite convex combos of permutation matrices P_i. Pf of $\to$: If convex combo of $\{P_i\}$, S stochastic. Every P_i is extreme point of $\{S\}$. Every non permutation stochastic matrix A is convex combo of stochastic matrices X = A + aB and Y = A - aB; where B and a are found thus: pick nz a_{ij} in row with ≥ 2 nz entries, then pick nz a_{kj}, then pick nz $a_{k,l}$ etc.. till you hit $a_{i',j'}$ seen before; take this sequence T, set $a = \min T$; make $\pm 1, 0$ matrix B by setting entries corresponding to alternate elements in T 1 or -1. $\{S\}$ is compact and convex set with $\{P_i\}$ as extreme points.

6.9.5 Doubly Substochastic matrix Q

equiv $Q1 \leq 1, 1^*Q \leq 1$.

For permutation matrix P, PQ or QP also substochastic.

Q is dbl substochastic iff B has dbl stochastic dilation S: make deficiency vectors d_r, d_c; get difference matrix $D_r = diag(d_r), D_c = diag(d_c)$; get $S = \begin{bmatrix} Q & D_r \\ D_c^T & Q^T \end{bmatrix}$.

$\{Q\}$ equivalent to set of convex combos of partial permutation matrices: Dilate Q to S, get finite convex combo of P_i, take the convex combo of principle submatrices.

$Q \in C^{nn}$ is dbl substochastic iff $\exists$ dbl stochastic $S \in C^{nn}$ with $A \geq B$: Take any Q, get finite convex combo of partial permutation matrices; alter each to get permutation matrix; their convex combo is S.

6.10 Large matrices

Often an approximation to ∞ size matrices; so have structure or regularity.

6.10.1 Sparsity

Not density. Very few non zero entries per row: ν. Can find Ax in $O(\nu m)$, not $O(m^2)$ flops. Can find AB in $O(mn\nu)$.

6.10.2 Iterative algorithms

See numerical analysis ref.

7 Random matrices

7.1 Applications

7.1.1 Random projections

(Johnson Lindenstrauss) See randomized algorithms survey.

Part III

Matrix functions and functionals

8 Matrix vector spaces and associated functionals

8.1 Vector space of matrices over field F

$M_{m,n}(F), M_{n,n}(F) \equiv M_n(F)$. $M_{m,n}(C) = C^{mn} \equiv C^{m\times n}$.

8.2 Matrix inner products

8.2.1 Trace inner product

$\langle A, B\rangle = tr(B^*A)$: same as taking vectorizing B and A and using vector $\langle ., .\rangle$; also see the elementwise multiplication before addition view. Aka standard inner product.

For symmetric matrices: $\langle A, B\rangle = \sum_i \sum_j X_{ii}Y_{ij}$

8.3 Matrix norms

Obeys all properties of vector norms,
plus sub-multiplicativity: $\|AB\| \leq \|A\|\,\|B\|$. Perhaps $\|A\| = \|A^*\|$ too. Generalized matrix norms need not be submultiplicative.

8.3.1 Unitary invariance

If $\|.\|$ unitary invariant, by SVD, $\|A\| = \|\Sigma\|$.

8.3.1.1 Symmetric gauge fn g

$g : C^q \to R^+$ is a vector norm on C^q which is also an absolute norm, and is permutation invariant: g(Px) = g(x): a fn on a set rather than a seq.

Every unitarily invariant matrix norm $\equiv$ symmetric gauge fn on σ. Pf: Given $\|\|\| : g(x) = \|X\| : X = diag(x)$ is symm gauge: permutation invariance from unitary invariance of $\|\|\|$. $\|X\| = g(\sigma)$ is unitary invariant matrix norm: unitary invariance from invariance of Σ; as g is vector norm, get +ve definiteness, non negativity, homogenousness. $\triangle$ ineq: g is absolute, so monotone; $\sigma(A+B)$ weakly majorized by $\sigma(A)+\sigma(B)$, so $\sigma(A+B) \leq S[\sigma(A)+\sigma(B)]$ for doubly stochastic S; so $g(\sigma(A+B)) \leq g(S(\sigma(A)+\sigma(B))) \leq \sum \alpha_i(g(P_i\sigma(A)) + g(P_i\sigma(B))) \leq g(\sigma(A)) + g(\sigma(B)) = \|A\| + \|B\|$, by Birkhoff.

8.3.2 Max norm

$\max |a_{i,j}|$.

8.3.3 Matrix p norms Induced by vector norms

$\|A\| = \sup_x \frac{\|Ax\|}{\|x\|}$. Obeys triangle ineq! So, get (p, q) norm, p norm.

$\|A\|_p = \|U\Sigma V^*\|_p = \|\Sigma\|_p$.

$\|A\|_\infty$ is max row sum: use suitable $x = |1|^n$; thence get $\|Ax\|_\infty$.

$\|A\|_1$ is max col sum.

8.3.3.1 Unitary invariance: 2 norm only

Change of orth basis. $\|QA\|_2 = \|A\|_2$ as $\|Qx\|_2 = \|x\|_2$.

But, $\|QA\|_p \neq \|A\|_p$ as $\|Qx\|_p \neq \|x\|_p$. By SVD, $\|A\|_2 = \|A^*\|_2$.

8.3.3.2 Comaprison of norms

$\|A\|_\infty \leq \sqrt{n}\,\|A\|_2$: Take x with $\|x\|_2 = 1$, for which $\|Ax\|_2 = \|A\|_2$; then $n\,\|Ax\|_2^2 = n\,\|A\|_2^2 = \sum_j(\sum_i nx_iA_{j,i})^2$; $nx_i^2 \geq 1$, so this exceeds every row sum.

Similarly, $\frac{\|A\|_F}{\sqrt{n}} \leq \|A\|_2$.

$\|A\|_2 \leq \sqrt{m}\,\|A\|_\infty$: For $\|x\|_2 = 1, Ax_i \leq$ max row sum of A.

Indicate matrix **energy**, consider sphere mapped to ellipse.

8.3.3.3 Connection with spectral radius

$\|A\| \geq |\lambda_{max}(A)|$ as $\sup_x \frac{\|Ax\|}{\|x\|} \geq |\lambda_{max}(A)|$. ***Wonderful!***

8.3.3.4 Find p norm of A

For $\|A\|_2$ use SVD; aka spectral norm if A square.

Take x with $\|x\|_p = 1$, maximize $\|Ax\|_p$. Use Triangle inequality: $\|Ax\|_1 = \|\sum x_ia_i\|$
$\leq \sum \|x_ia_i\|$, so $\|Ax\|_1 = max\,\|x_i\|$.

Similarly use Cauchy Schwartz ineq. By $\|A\| \geq \frac{\|Ax\|}{\|x\|}$,
$\|ABx\| \leq \|A\|\,\|Bx\| \leq \|A\|\,\|B\|\,\|x\|$; so $\|AB\| \leq \|A\|\,\|B\|$ (in general a loose bound).

8.3.4 Matrix (p, q) induced norm

Aka operator norm. $\max_{\|q\|=1}\|Ax\|_p$. Check $\triangle$ ineq.

8.3.5 Ky Fan (p,k) norms

Take σ_i in descending order. $\|A\|_{p,k} = (\sum_{i=1}^k \sigma_i^p)^{1/p}$ for $p \geq 1$: p norm to top k σ.

$\triangle$ ineq for (1, k) norm from Σ inequalities. Vector normness for $\|A\|_{p,k}$: $\|x\|_{p,k}$ a symmetric gauge fn: $\triangle$ ineq: take A, b in descending order to get $a' = (a_{[i]}), b' = (b_{[i]})$; $\sum_i^k (a_{[i]} + b_{[i]}) \geq \sum_i^k (a+b)_{[i]}$; so by weak majorization lore, for $p \geq 1$: $\sum_i^k (a_{[i]} + b_{[i]})^p \geq \sum_i^k (a+b)_{[i]}^p$; thence see: $\|a' + b'\|_{p,k} \leq \|a'\|_{p,k} + \|b'\|_{p,k}$ from p-norm properties.

Matrix normness:
$\sum_{i=1}^k \sigma_i(AB)^p \leq \sum_{i=1}^k \sigma_i(A)^p \sigma_i(B)^p \leq \sum_{i=1}^k \sigma_i(A)^p \sum_{i=1}^k \sigma_i(B)^p$.

$\|A\|_{1,1} = \|A\|_2$.

8.3.6 Schatten p norms

Apply p norm to singular values. Special case of Ky Fan norm: $\|A\|_{p,q} = \|A\|_{Sp} = (\sum \sigma_i^p)^{1/p}$. Vector normness from seeing that this is a symmetric gauge fn.

8.3.6.1 Frobenius (Hilbert-Schmidt, Euclidian) norm

$\|A\|_{S2} = \|A\|_F$.

$(\sum a_{i,j}^2)^{\frac{1}{2}} = (\sum \|a_j\|^2)^{\frac{1}{2}} = (trA^*A)^{\frac{1}{2}} = (trAA^*)^{\frac{1}{2}} = (tr\Sigma^*\Sigma)^{1/2} = A_F$. So, based on matrix inner product: $\langle A, B\rangle = tr(B^*A)$.

So, $\|QA\|_F = \|A\|_F$. By Cauchy Schwartz, $\|C\|_F^2 = \|AB\|_F^2 = \sum_i \sum_j (a_i^* b_j)^2 \leq \sum_i \sum_j \|a_i\|_2^2 \|b_j\|_2^2 = \|A\|_F \|B\|_F$.

8.3.6.2 Trace (Nuclear) norm

$\|A\|_{S1} = \|A\|_{tr} = \sum \sigma_i = tr((A^*A)^{1/2})$. Corresponds to the trace inner product.

In finding $C, D : \min \|A - CD\|_{tr}$, using trace norm often yields low rank solutions. [**Check**]

9 Other functionals

9.1 Functionals over square matrices

Also see functionals over +ve definite A.

9.1.1 Determinant of square A

9.1.1.1 Definitions

Det(A) or $|A|$: The recursive defn. **Cofactor** of $A_{1,1}$: $C_{1,1}$ = det of submatrix (minor) of A. Defn by properties: 1: det I = 1. 2: If 2 rows are equal, det(A) = 0. 3: det A depends linearly on first row.

9.1.1.2 Properties

So, there is sign change in Det(A) with row exchanges. Also, Det(A) is unchanged with $row_1 += krow_2$. So, 0 row means det $A = 0$. Also, for L or U, just multiply diagonal. Also, det $A = 0$ iff A singular.

Multiplicativity and consequences $|A||B| = |AB|$: See that $\frac{|AB|}{|B|}$ has the 3 properties $|A|$ should have.

So, $|A^{-1}| = \frac{1}{|A|}$. So, $|Q| = \pm 1$. So, considering $|PA| = |LU|$ and $|A^T P^T| = |U^T L^T|$, $Det(A^T) = Det(A)$. So, column operations don't alter det A, there is sign change with col exchanges, write det A with column cofactors etc..

9.1.1.3 Connection with ew, sw

Take characteristic polynomial p_A. $det(A) = p_A(0) = \prod_i (l_i - l)]_{l=0}$: also by considering $A = QTQ^*$. Also, $|det(A)| = \prod \sigma_i(A)$.

Connection with rank and semidefiniteness Thence, det(A) $= 0 \equiv A$ is rank deficient. $det(A) < 0 \implies$ A$\not\succeq 0$: implication is one directional.

9.1.1.4 Find Det(A)

Reduce A to U, find det U. For each of the n! diagonals, multiply elements; add after accounting for permutation sign: From linearity property, get row exchanged diagonal matrices, consider relation between determinant sign and row exchanges; Hence note equivalence of definitions. Use eigenvalue decomposition: $|A| = |S\Lambda S^{-1}| = |S||\Lambda||S|^{-1} = |\Lambda|$.

Det(A) = volume spanned by row (or col) vectors: orthogonalize (parallelopiped changed to cube of equal volume): $|A|$ unchanged; now AA^T is diagonal matrix; $|AA^T| = |A|^2 = vol^2$. **Perm(A)**: Same as Det(A), ignore sign.

Block matrices Leibnitz:
$X = \left[\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right] = \left[\begin{smallmatrix} A & 0 \\ C & I \end{smallmatrix}\right] * \left[\begin{smallmatrix} I & A^{-1}B \\ 0 & D-CA^{-1}B \end{smallmatrix}\right]$. Use this if A square, invertible, thence find determinant.

If X symmetric, ie if $B = C^T$: $D - CA^{-1}B$ is the Schur complement of A.

9.1.2 Trace of A

$tr(A) = \sum_i a_{i,i}$.

A linear map: tr(kA + lB) = k tr(A) + l tr(B): so convex.

9.1.2.1 Trace of AB

Trace is an inner product on matrices. It is same as vectorizing A and B and applying the standard inner product.

$tr(AB) = \sum_i \sigma_i \sigma_{\pi(i)}$ for some permutation π: from $|A||B| = |AB|$.
(Von Neumann): $tr(AB) \leq \prod_i \sigma_i(A)\sigma_i(B)$ [**Find proof**].

9.1.2.2 Cyclicity and similarity invariance

$tr(AB) = tr(BA) = \sum_i \sum_j a_{i,j} b_{j,i}$; but $tr(ABC) = tr(BCA) \neq tr(ACB)$.
Similarity invariant: $tr(P^{-1}AP) = tr(APP^{-1}) = tr(A)$.

9.1.2.3 Trace of outer products

$tr(ab^T) = b^T a$ by algebra.

9.1.2.4 Connection with ew

Take characteristic polynomial p_A. tr(A) = coefficient of l^{m-1} in $p_A = \sum_i l_i$.

9.1.2.5 Gradient to tr

By considering $tr(f(X + \Delta X)) - tr(f(X))$, get: $\nabla_X tr(X) = I, \nabla_X tr(XA) = \nabla_X tr(AX) = A^T, \nabla_X tr(BXA) = \nabla_X tr(ABX) = B^T A^T$.

By considering $tr(f(X + \Delta X)g(X + \Delta X)) - tr(f(X)g(X))$, $\nabla_X tr(AXBX) = B^T X^T A^T + A^T X^T B^T$.

9.2 Functionals over +ve definite matrices

9.2.1 Log det divergence

$f(A) = \log det(A) = \sum_i \log \lambda_i(A) = tr(\log(A))$;
often used because it is convex. $\nabla \log det(A) = A^{-1}$ [**Find proof**].

9.3 Singular values (sw)

See SVD section.

9.3.1 Unitary invariance

$\Sigma \in R^{mn}$ always, so $\Sigma = \Sigma^*$. $\Sigma = U^* AV$: so, Σ is unitary invariant: $\sigma_i(A) = \sigma_i(Q_1 A Q_2)$.

Let $a = (a_{i,i}), \sigma = (\sigma_i)$: By direct computation using $U\Sigma V^*$, $a = Z\sigma$, where $Z = [u_{ij} v_{ij}] = (U.\bar{V})_{1:q,1:q}$; take $Q = |Z|$; get $|a| \leq Q\sigma$. $\|q_i^*\|_1^2 = (\sum_j |u_{ij} v_{ij}|)^2 \leq \sum_j |u_{ij}|^2 \sum_j |v_{ij}|^2 \leq 1$, by induction; also $\|q_i\|_1 \leq 1$; so Q is substochastic.

9.3.2 Effect of row or column deletion

A_r: A with r rows or cols deleted; $\sigma_k(A) \geq \sigma_k(A_r) \geq \sigma_{k+r}(A)$. Prove for r=1, get general case thence. Pf where sth col is deleted: For upper bound, use

$\sigma_k(A_1) = \max_{S\subset C^n, dim(S)=k} \min_{x\in S} \|Ax\|_2$ with extra cond: $x \perp e_s$; for lower bound use
$\sigma_k(A_1) = \min_{S\subset C^n, dim(S)=n-k} \max_{x\in S} \|Ax\|_2$ with extra cond. For row deletion, consider A^*.

9.3.2.1 Sum, product of sw of square A

ew sw sum comparison By block multiplicity, for any arbit sq orth x, Y: $S_k = X_k^* A Y_k^*$: upper left submatrix of $S = X^*AY$; So, $\sigma_i(X_k^* A Y_k^*) \leq \sigma_i(X^*AY) = \sigma_i(A)$. So, $|det(X_k^* A Y_k)| = \prod_{i=1}^k \sigma_i(X_k^* A Y_k) \leq \prod_{i=1}^k \sigma_i(A)$.

For square A, take $A = QTQ^*$, $T = Q^*AQ$, $(\lambda_i)\downarrow$; so take k-principle submatrix, use block multiplicity to get: $T_k = Q_k^* A Q_k$
So, $|det(T_k)| = \prod_{i=1}^k |\lambda_i(A)| = |det(Q_k^* A Q_k)| \leq \prod_{i=1}^k \sigma_i(A)$: = for k=m.

By majorization lore, $|tr(A)| \leq \sum_{i=1}^q |\lambda_i(A)| \leq \sum_{i=1}^q \sigma_i(A)$. Also, for any $p > 0, \sum_{i=1}^q |\lambda_i(A)|^p \leq \sum_{i=1}^q \sigma_i(A)^p$.

Sum of sw of matrix products $A \in C^{mp}, B \in C^{pn}$.

Take $AB = U\Sigma V^*$; $U_k^* ABV_k = \Sigma_k$; polar decomposition of $BV_k = X_k Q$; $Q^2 = V_k^* B^* B V_k$; $det(Q^2) \leq \prod_{i=1}^k \sigma_i(B^*B) = \prod_{i=1}^k \sigma_i(B)^2$.

So, $\prod_{i=1}^k \sigma_i(AB) = |det(U_k^* ABV_k)| = |det(U_k^* AX_k) det(Q)|$
$\leq \prod_{i=1}^k \sigma_i(A)\sigma_i(B)$.

By majorization lore, for $p > 0, \sum_{i=1}^q \sigma_i(AB)^p \leq \sum_{i=1}^q \sigma_i(A)^p \sigma_i(B)^p$.

Sum of sw as trace maximizer of SVD-like decompositions .
$\sum_{i=1}^k \sigma_i(A) = \max \left\{ |tr(X^*AY)| : X^*X = Y^*Y = I, X \in C^{mk}, Y \in C^{nk} \right\}$
$= \max \{ |tr(AC)| C \in C^{nm} \texttt{ is rank k partial isometry} \}$.

Pf: Can get C from Y, X: $C = YX^*, \sigma_i(C^*C) = \sigma_i(XX^*) = \sigma_i(X^*X) = 1$; Can get Y, X from C: use SVD: $C = U\Sigma V^* = U_k V_k^*$. $|tr(AC)| = |\sum \lambda_i(AC)| \leq \sum_{i=1}^q \sigma_i(AC) \leq \sum_{i=1}^q \sigma_i(A)\sigma_i(C) = \sum_{i=1}^k \sigma_i(A)$. Take $A = U\Sigma V^*$, then for rank k isometry $C = V\hat{I}_k U^*$, $tr(AC) = tr(U\Sigma \hat{I}_k U^*) = \sum_{i=1}^k \sigma_i(A)$.

Triangle inequality All σ_i in descending order; then:
$\sum_{i=1}^k \sigma_i(A+B) \leq \sum_{i=1}^k \sigma_i(A) + \sum_{i=1}^k \sigma_i(B)$.

Pf: Let C be some rank k partial isometry:
$\sum_{i=1}^k \sigma_i(A+B) = \max \{|tr((A+B)C)|\} \leq \max \{|tr(AC)| + |tr(BC)|\}$
$\leq \max \{|tr(AC)|\} + \max \{|tr(BC)|\} = RHS$.

9.3.3 Convexity, concavity

$\sigma_1(X)$ is convex, but $\sigma_n(X)$ is concave: $\sigma_1(tA + (1-t)B) = \sup_{\|x\|=1} tAx + (1-t)Bx \leq \sigma_1(A)x + \sigma_1(B)x$.

9.3.4 Properties of A discerned from SVD

Rank(A) = num +ve singular values: By geometric action, or by algebra U, V unitary, Σ diagonal: $|A| = |U\Sigma V^*| = |U||\Sigma||V^*| = |\Sigma|$. Similarly, numerical Rank of A = num +ve not-close-to-0 σ. Good way to find rank.

SVD is most accurate method for finding orthonormal basis for N(A) and range(A).

$A = \sum \sigma_i u_i v_i^*$: $\sum$ rank 1 matrices.

10 Eigenvalues and relatives

10.1 Eigenvalue (ew)

10.1.1 ew problem

Aka: ew or eigenwert. For square A only. These are solutions to the eigenvalue problem: $Ax = \lambda x$ (Eigen = distinctive). This defines the Eigen pair: (λ, x), where λ is ew, x is a (right) ev (eigen vector).

10.1.1.1 Left and right eigenpairs

Also, can define left ev and ew by the relation $xA = \lambda x$. ew of A and left ew of A^T are same.

ew of A and ew of A^T are the same if A is real: By Schur, $A = QUQ^*$, $A^* = QU^*Q^*$. As QU^*Q^* is a similarity transformation, and U^* is triangular, the ew of A are still $diag(U)$.

In the case of diagonalizable A, ew decomposition $AS = S\Lambda$ reveals that rows of S^{-1} are the left ev: For details see ew decomposition section.

10.1.2 Characteristic polynomial

As $(A - \lambda I)x = 0$, $det(A - \lambda I) = 0$. 0 is always an ev.

Other things apart, this implies that ew of a triangular matrix are on its diagonal.

10.1.2.1 Mapping polynomials to matrices

Every polynomial $p(\lambda) = \lambda^m + a_{m-1}\lambda^{m-1} .. + a_0$ determinant of some matrix; Eg: $\begin{bmatrix} \lambda & a_0 \\ -1 & \lambda+a_1 \end{bmatrix}$ for m=2; which is $A - \lambda I$ for some companion matrix A.

10.1.3 Applications

Domain and range of A are the same space. Useful where iterative calculations: A^k or e^{tA} occur.

Physics: evolving systems generated by linear equations; Eg: resonance, stability.

Simpler algorithms: Reduces coupled system into a collection of scalar systems. [**Find proof**]

10.2 ew and ev properties

10.2.1 ew properties

For connection with matrices, distribution etc.. see later section.

10.2.1.1 Number of ew

There are n ew in C. Algebraic multiplicity of ew: number of times an ew appears as root of $p(\lambda)$.

10.2.1.2 0 as an ew

If A full rank, 0 is not an ew: $Ax = \lambda x$ else A's columns would be dependent. If 0 is an ew, $E_0 = N(A)$.

10.2.2 Eigenspace of an ew

If x is an ev corresponding to λ, so is -x and kx for any scalar k. For every λ, ev span the null space of $A - \lambda I$.

An invariant subspace: $AE_\lambda \subseteq E_\lambda$. $E_{\lambda_1} = N(A - \lambda_1 I)$.

10.2.2.1 Independence of ev

ev x_1, x_2 (non-0) for distinct ew λ_1, λ_2 are linearly independent: otherwise action of A on collinear x_1, x_2 would involve the same scaling.

10.2.2.2 Defective matrices

Geometric multiplicity of $\lambda \leq$ algebraic multiplicity: Let n be geometric multiplicity of λ in A; Select orth vectors in E_λ to get $\hat{V}$; extend it to square V; take $B = V^*AV = \left[\begin{smallmatrix} \lambda I & C \\ 0 & D \end{smallmatrix}\right]$; but $det(B - zI) = det(\lambda I - zI)det(D - zI) = (\lambda - z)^n det(D - zI)$: so algebraic multiplicity of λ in B $\geq n$, same in A.

If algebraic multiplicity of ew exceeds geometric multiplicity, ew is defective. If A has defective ew, it is defective. Eg: Diagonal matrix never defective.

10.2.3 Rayleigh quotient of x

$r(x) = \frac{x^T Ax}{x^T x}$. Like solving the least squares problem: $xa \approx Ax$ for a; thereby approx eigenvalue. Range(r(x)): Field of values of numerical range of A: W(A). Includes Convex hull of $\Lambda(A)$.

10.2.3.1 EV as stationary points

$\frac{\partial r(x)}{\partial x_j} = \frac{2(Ax-r(x)x)_j}{\|x\|^2}$; their vector, the gradient $\nabla r(x) = \frac{2(Ax-r(x)x)}{\|x\|^2}$. ev are the stationary points: $\nabla r(x) = 0$ iff x is ev and r(x) is ew.

10.2.3.2 Geometry

$r(x) : S^{n-1} \to R$. Normalized ev of A are stationary points of r(x) for $x \in S^{n-1}$.

10.2.3.3 maxmin thm for symmetric A

[Courant Fishcer]. λ_i descending. Then $\lambda_i = \max_{S:dim(S)=i} \min_{y\neq 0 \in S} R_A(y)$.

10.2.3.4 Quadratic Accuracy of ew wrt ev error

For $A = A^T$, let q_i be ev. Then, any $x = \sum_i a_i q_i$, $r(x) = \frac{\sum a_i^2 \lambda_i}{\sum a_i^2}$. So, W(A) = Convex hull of $\Lambda(A)$. So, $max_x r(x) = \lambda_{max}$. Also, $r(x) - r(q_i) = O(\|x - q_i\|^2)$ as $x \to q_i$ or $\forall j : |a_j/a_i| \leq \epsilon$.

10.2.3.5 Rayleigh quotient of M: Generalization

If $A =$ m*m, M is m*n: thin, tall, required to be full rank.
$r(M) = tr((M^T M)^{-1}(M^T AM))$.

Take svd: $M = U\Sigma V^*$. Then
$\max r(M) = \max_U |tr(U^T AU)| = \max_U \prod |R(u_i)|$. This happens when U is top k unique ev(A).

10.3 ew and matrices: other connections

10.3.1 ew distribution

Set of ew: Spectrum of A: $\Lambda(A)$. Spectral radius: $\rho(A) = |\lambda_{max}|$. ew follows wigner's semicircle distribution for random matrices. [**Find proof**]

10.3.2 ew and the diagonal

10.3.2.1 In Disks around the diagonal

(Gerschgorin) In complex plane, take disks with center $A_{i,i}$, radii $\sum_{j\neq i} A_{i,j}$; each ew lies in atleast one disk: Take ev v and ew l; take $i = argmax_i\ v_i$; then $|\lambda - a_{i,i}| \leq |\sum_{j\neq i} a_{i,j} \frac{v_j}{v_i}| \leq |\sum_{j\neq i} a_{i,j}|$. Good estimate of ew!

10.3.2.2 Monotonic dependence on diagonal

Take $A = QTQ^*$, take any +ve diagonal D; get $A + D = Q(T + D)Q^*$. Used to take $A \notin S_+ to B \in S_+$.

10.3.3 Effect of transformations

10.3.3.1 Similarity transformation

X nonsingular, square; $A \to B = X^{-1}AX$. A, B similar if $\exists X : B = X^{-1}AX$.

Change of basis op: See Ax=b as $AXx' = Xb'$ when $Xx' = x$, $Xb' = b$, so $Bx' = b'$.

$p_{X^{-1}AX} = det(\lambda I - X^{-1}AX) = det(X^{-1}(\lambda I - A)X) = det(\lambda I - A) = p_A$. So, ew's, their agebraic multiplicities same. Geometric multiplicities same: $X^{-1}E_\lambda$ is eigenspace of $X^{-1}AX$: if $Ax = \lambda x$, $BX^{-1}x = \lambda X^{-1}x$.

10.3.3.2 Eigenpairs of A^k

$A^k = (QUQ^*)^k = QU^kQ^*$. So, ew are λ_i^k. So, ew of A^{-1} are $\frac{1}{\lambda_i}$: $L^{-1} = (S^{-1}AS)^{-1} = S^{-1}A^{-1}S$.

10.3.4 Matrix construction

10.3.4.1 Matrix with certian ew and diagonal entries

If real λ majorizes A, $\exists$ real symmetric A with $a_{i,i} = a_{(i)}$, such that $\lambda_{(i)}$ are ew of A. Pf: By induction; assume true for n-1; Take $g \in R^{n-1}$ interleaved amongst λ which majorizes a; by ind hyp, take real symmetric $B = QGQ^*$ with diagonal a and ew g; can make $A' = \begin{bmatrix} G & y \\ y^T & \alpha \end{bmatrix}$ with ew λ; take $A = \begin{bmatrix} Q & 0 \\ 0 & 1 \end{bmatrix} A' \begin{bmatrix} Q^T & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} B & Qy \\ y^TQ^T & \alpha \end{bmatrix}$ with ew λ and diag a.

10.3.4.2 Extending Λ to A with certain interleaving $\lambda(A)$

Take $\lambda \in R^n$ interleaved between $l \in R^{n+1}$. Can make real $A = \begin{bmatrix} \Lambda & y \\ y^T & a \end{bmatrix}$ with $\lambda(A) = l$.

Pf where λ_i differ: $a = tr(A) - tr(\Lambda) = \sum l_i - \sum \lambda_i$. Take $det(tI - A) = det[\begin{bmatrix} I & (tI-\Lambda)^{-1}y \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} tI-\Lambda & -y \\ -y^T & t-a \end{bmatrix} \begin{bmatrix} \Lambda & y \\ y^T & a \end{bmatrix}]$ (multiplicity two matrices with det = 1) $= [(t-a) - y^T(tI-\Lambda)^{-1}y]det(tI-\Lambda) = (t-a) - \sum_i y_i^2(t-\lambda_i)^{-1}\prod_i(t-\lambda_i) = p_A(t)$.

Find y to make $p_A(l_i) = 0$. Characterize y, show it exists: Take $f(t) = \prod(t-l_i), g(t) = \prod(t-\lambda_i)$, so $f(t) = g(t)(t-c) + r(t)$, where r(t) has degree $\leq n-1$. $c = \sum_i l_i - \sum_i \lambda_i = a$; $f(\lambda_i) = r(\lambda_i)$, so get Lagrange interpolation form of $r(t) = \sum_i f(\lambda_i)\frac{g(t)}{g'(\lambda_i)(t-\lambda_i)}$. So, $\frac{f(t)}{g(t)} = (t-a) - \sum_i \frac{-f(\lambda_i)}{g'(\lambda_i)(t-\lambda_i)}$; this is $0 \forall t = l_i$ and $y_i^2 = \frac{-f(\lambda_i)}{g'(\lambda_i)}$. But, by interlacing, $f(\lambda_i) = (-1)^{n-i+1}\prod_j |\lambda_i - l_j|$ and $g'(\lambda_i) = (-1)^{n-i}\prod_{j\neq i}|\lambda_i - \lambda_j|$ oppose in sign, so $\exists y$.

Pf where ew_i recurs: Divide out $(t-\lambda_i)^k$ from all, proceed as usual.

10.3.5 EW of special matrices

10.3.5.1 Real A: complex ew in pairs

If A is real, p_A has real coefficients.

So, if l is ew of A, so is $\bar{\lambda}$: complex roots of P appear in pairs. ew simple if its algebraic multiplicity 1.

10.3.5.2 Triangular and diagonal A

For triangular A, from $det(A - \lambda I) = 0$, ew are on diagonal. So, $tr(A) = \sum A_{i,i} = \sum \lambda_i$. Also, $\prod \lambda_i = |A|$. For diagonal A, eigenpairs are diagonal element $(A_{i,i}, ke_i)$.

10.3.5.3 Nilpotent matrix A

A is a nilpotent op (see algebra ref). So, all ew are 0; $|A|$, tr(A) are 0. Any singular $A =$ product of nilpotent matrices.

10.3.5.4 Singular A

0 is an ew.

10.3.5.5 Semidefiniteness and hermitianness

See other sections.

10.4 Generalized eigenvalue problem

$Az = \lambda Bz$. $det(A - \lambda B) = 0$. $\{A - \lambda B\}$ is called pencil.

10.4.1 General Rayleigh quotient of x

$R(x) = \frac{x^*Ax}{x^*Bx}$. $\nabla R(x) = 2(Ax - R(x)Bx)$ is 0 exactly where R(x) is λ, x the ev.

10.4.1.1 Of M

M is tall, thin, with independent columns.
$R(M) = tr((M^*BM)^{-1}(M^*AM))$. Using svd $M = U\Sigma V^*$:
get $R(M) = tr_U(U^*B^{-1}AU)$: same as common rayleigh quotient of $B^{-1}A$: see ev section. Maximized when U is the top ev's of $B^{-1}A$.

10.4.2 Reductions to common ew problem

If B is invertible: $B^{-1}Az = \lambda z$: so now an λ problem. Don't want to invert: so solve some other way.

So, if B is invertible, symmetric, +ve definite: $R(x) = \frac{x^*Ax}{x^*B^{1/2}B^{1/2}x}$; taking $z = B^{1/2}x$, get $R(z) = \frac{z^*B^{-1/2}AB^{-1/2}z}{z^*z}$. Max of R(x) is achieved somewhere in R(z) as $z \to x$ is a 1 to 1 map. ew of $B^{-1/2}AB^{-1/2}$ are easier to find as it is symmetric.

11 Matrix to matrix functions

11.1 $(I - A)^{-1}$ series for square A

Aka Neumann series. $(I - A)^{-1} = \sum_{k=0}^{\infty} A^k$, if it converges. It converges when A has norm < 1.

11.2 Matrix exponentiation for square A

$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$: always converges. Also defines $\log A$. If A nilpotent, series is finite.

Using expansion, aggregating suitably: $e^{aX}e^{bX} = e^{(a+b)X}$; If $XY = YX$: $e^X e^Y = e^{X+Y}$.

$e^X e^{-X} = I$: So exponential of X is always invertible.

$e^{YXY^{-1}} = Ye^X Y^{-1}$.

$e^{X^*} = (e^X)^*$.

If D diagonal, easy to get e^D. Thence easily get e^A if $A = XDX^{-1}$ (A diagonalizable).

11.2.1 Relationship among ew

As $e^{\lambda(A)} = \lambda(e^A), det(e^A) = e^{\sum \lambda_i(A)} = e^{tr(A)}$.

Part IV

Generalizations

12 Tensors

Independent of chosen frame of reference. An array of numbers is a mere representation (components of the tensor) in a certain basis, not the tensor itself.

12.1 Order

Number of indices for each entry. Observe the difference from dimensions!
Scalar: order 0 tensor. Vector: order 1. Matrices: order 2.

12.2 Rank

[**Incomplete**]

Bibliography

[1] Horn and Johnson. *Matrix Analysis.* CUP, 1985.

[2] Horn and Johnson. *Topics in Matrix Analysis.* CUP, 1991.

[3] K. B. Petersen and M. S. Pedersen. The matrix cookbook, oct 2008. Version 20081110.

[4] Gilbert Strang. *Linear Algebra and its applications, 4e.* Thomson Brooks/Cole, 2006.

[5] Lloyd N. Trefethen and David Bau III. *Numerical Linear Algebra.* Siam, 1997.