Lecture 6: Positive definite matrices

Definition: A *Hermitian* matrix $A \in M_n$ is *positive definite* (pd) if

$$x^* A x > 0 \qquad \forall x \in \mathbf{C}^n,\ x \neq 0$$

A is *positive semidefinite* (psd) if $x^* A x \geq 0$.

Definition: $A \in M_n$ is negative (semi)definite if $-A$ is pd (or psd).

If neither holds: $A \in M_n$ is *indefinite*.

Generating a pd/psd matrix: Choose any $B \in M_n$, then

$$A = B^* B$$

is pd/psd. Possible for all pd/psd matrices!

Positive definite cone

Property: Positive linear combination of pd matrices is pd.

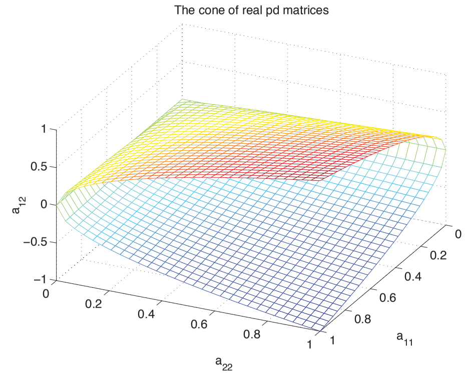
Conclusion: Set of pd matrices is a positive cone in the vector space.

Example:

$A = A^T \in M_2(\mathbf{R})$.

Pd iff $a_{11} > 0, a_{22} > 0$

and $|a_{12}|^2 < a_{11} a_{22}$



Properties of Positive definite matrices

As $x^* A x > 0 \quad \forall x \neq 0$, we have:

- Full rank of pd matrices.
- Any principal sub-matrix of a pd matrix is pd.
- Diagonal elements of a pd matrix are positive.

If $A \in M_n$ is pd and $C \in M_{n,m}$, then

- $C^* A C$ is psd and $\text{rank}(C^* A C) = \text{rank}(C)$
- $C^* A C$ is pd if and only if $\text{rank}(C) = m \leq n$.

Characterizations

How to check if a given matrix is pd / psd?

Based on Eigenvalues:

A Hermitian matrix $A \in M_n$ is pd if and only if $\lambda_i(A) > 0$ for all i.

It is psd iff $\lambda_i(A) \geq 0$.

Based on Determinants:

Let $A_i \in M_i$ denote the leading principal submatrix of a matrix $A \in M_n$. If $A \in M_n$ is Hermitian, then A is pd iff $\det(A_i) > 0$ for all $i = 1, \ldots, n$.

Note: We may permute rows and columns before applying the result.

Matrix Roots

Assume: $A \in M_n$ is psd and k is a positive integer.

Theorem: There exists a unique psd matrix B such that $B^k = A$. It also holds that

- $BA = AB$ and $B = p(A)$ for some polynomial $p(t)$.
- $\text{rank}(B) = \text{rank}(A)$
- B is real if A is real.

Example: If $k = 2$, then B is the unique square root of A.

Cholesky Factorization

Corollary: A matrix $A \in M_n$ is pd iff there exists a lower triangular matrix $L \in M_n$ with positive diagonal elements such that

$$A = LL^*$$

Properties:

- L is called the Cholesky factor
- If A is real then L can be taken to be real.
- Enables solving a linear system of equations by back substitution.

Congruence and Diagonalization

Recall (Similarity): $A, B \in M_n$ are simultaneously diagonalizable if it exists nonsingular $S \in M_n$ such that $S^{-1}AS$ and $S^{-1}BS$ are diagonal. Implication: $AB = BA$.

Can something less strict exist?

Theorem: Suppose $A, B \in M_n$ are Hermitian and there exists a linear combination of A and B which is pd. Then there is a nonsingular $C \in M_n$ such that C^*AC and C^*BC are diagonal.

Important: C^*AC or C^*BC need not be the eigenvalue decomposition.

Application: Concavity of $\log\det$

Definition: A function is strictly concave if

$$f(\alpha A + (1-\alpha)B) \geq \alpha f(A) + (1-\alpha) f(B)$$

for $\alpha \in (0,1)$ with equality iff $A = B$.

Theorem: The function $f(A) = \log\det(A)$ is a strictly concave function on the convex set of pd matrices in M_n.

Proof exploits that there exist a nonsingular $C \in M_n$ such that C^*AC and C^*BC are diagonal.

Further Applications of $\log\det$

Theorem: Let $X \in M_n$ be pd. Then
$f(X) = \mathrm{tr}(X) - \log\det(X) \geq n$, with equality iff $X = I$.

Example: “Typical” ML criterion to be minimized:

$$V(\theta) = -\log\det(R^{-1}(\theta)\hat{R}) + \mathrm{tr}(R^{-1}(\theta)\hat{R})$$
$$= -\log\det(\hat{R}^{1/2}R^{-1}(\theta)\hat{R}^{1/2}) + \mathrm{tr}(\hat{R}^{1/2}R^{-1}(\theta)\hat{R}^{1/2})$$

If $R(\theta)$ is any pd matrix, this is a convex problem solved by $R = \hat{R}$ (if $\hat{R} = \hat{R}^{1/2}\hat{R}^{1/2}$ is pd).

Further Applications of $\log\det$ (cont'd)

Theorem: If $A = [a_{ij}] \in M_n$ is pd, then

$$\det(A) \leq \prod_{i=1}^{n} a_{ii}$$

with equality iff A is diagonal.

Example: “Typical” Capacity expression for Multiple Input Multiple Output (MIMO) systems to be maximized:

$$C(H) = \max_{Q:\mathrm{tr}(Q)\leq p} \log\det(I + HQH^*).$$

Convex problem solved by Q that diagonalizes HQH^* (since $\det(I + HQH^*) \leq \prod_{i=1}^{n}(1 + [HQH^*]_{ii})$).

Products

- A, B are pd matrices, then AB is pd if and only if they commute.
- Can we say something more about AB?

Theorem: Let $A \in M_n$ be pd and $B \in M_n$ be Hermitian. Then

1. AB is diagonalizable.
2. AB has the same number of positive, negative and zero eigenvalues as B.

The Schur Product Theorem

Definition: The Schur-Hadamard product of two matrices $A, B \in M_{m,n}$ is

$$A \circ B = [a_{ij}b_{ij}] \in M_{m,n}$$

Also called elementwise multiplication.

Theorem: Let A and B be psd.

- $A \circ B$ is psd.
- If A is pd and all diagonal elements of B are positive,then $A \circ B$ is pd.
- If A and B are pd,then $A \circ B$ is pd.

Positive semidefinite ordering

Observe: Hermitian matrices generalizes real numbers.
Observe: Positive definite matrix generalizes positive real numbers.

How to order the matrices?

Definition: We write $A \geq B$ if $A - B$ is psd, $A > B$ if $A - B$ is pd.

This defines a *partial ordering* of Hermitian matrices.

Theorem: If A, B are pd, then

1. $A \geq B \;\Leftrightarrow\; B^{-1} \geq A^{-1}$
2. If $A \geq B$, then $\det(A) \geq \det(B)$ and $\mathrm{tr}(A) \geq \mathrm{tr}(B)$
3. If $A \geq B$, then $\lambda_k(A) \geq \lambda_k(B)$ for all k (ordered eigenvalues)

Schur complements

Consider a Hermitian matrix partitioned as

$$\begin{bmatrix} A & B \\ B^* & C \end{bmatrix}$$

where A and C are square matrices.

Theorem: This matrix is pd iff $A > 0$ and $C - B^*A^{-1}B > 0$.

Definition: $C - B^*A^{-1}B$ is the Schur complement of A.

Useful to rewrite optimization constraints.

SVD: Singular Value Decomposition

Theorem: Any $A \in M_{m,n}$ can be decomposed as

$$A = V\Sigma W^*$$

- $V \in M_m$: Unitary with columns being eigenvectors of AA^*.
- $W \in M_n$: Unitary with columns being eigenvectors of A^*A.
- $\Sigma = [\sigma_{ij}] \in M_{m,n}$ has $\sigma_{ij} = 0, \; \forall\, i \neq j$

Suppose $\mathrm{rank}(A) = k$ and $q = \min\{m, n\}$, then

- $\sigma_{11} \geq \cdots \geq \sigma_{kk} > \sigma_{k+1,k+1} = \cdots = \sigma_{qq} = 0$
- $\sigma_{ii} \equiv \sigma_i$ square roots of non-zero eigenvalues of AA^* (or A^*A)
- Unique : σ_i, Non-unique : V, W

If A is real then V and W can be taken to be real.

SVD cont'd

Observation: SVD relies on eigendecompositions of AA^* and A^*A

Consequence: Many results for eigenvalues of Hermitian matrices can be converted to results for the singular values.

Examples:

1. Perturbations: Small error in A results in small error in singular values $\Rightarrow$ Well conditioned for computation.
2. Interlacing: $A \in M_{m,n}$ is given and $\hat{A}$ is obtained by deleting any one column of A. Denote the singular values of A by σ_i, the singular values of $\hat{A}$ by $\hat{\sigma}_i$, and set $q = \min\{m, n\}$:
   $\sigma_1 \geq \hat{\sigma}_1 \geq \sigma_2 \geq \hat{\sigma}_2 \geq \cdots \geq \hat{\sigma}_{q-1} \geq \sigma_q \geq 0$

SVD cont'd

Observation 2: Singular values from eigendecomposition
Let $A \in M_{m,n}$ have singular values $\sigma_1 \geq \ldots \geq \sigma_q$.

Define the Hermitian matrix

$$B = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$$

Ordered eigenvalues of B are

$$-\sigma_1 \leq \ldots \leq -\sigma_q \leq 0 = \ldots = 0 \leq \sigma_q \leq \ldots \leq \sigma_1$$

Useful connection for analysis or algorithm development.

Inequalities

Example: If $A, B \in M_n$ and $\sigma_i()$ are the singular values, then

$$\mathrm{Re}\Big(\mathrm{tr}(AB^*)\Big) \leq \sum_{i=1}^{n} \sigma_i(A)\sigma_i(B).$$

with equality iff SVDs are $A = V\Sigma_A W^*$ and $B = V\Sigma_B W^*$.

Many more examples: “Inequalities” by Marshall, Olkin, and Arnold.

SVD Applications: Perturbation

Problem: Find smallest (in e.g. Frobenius norm) perturbation E to the nonsingular matrix $A \in M_n$ such that $A + E$ is singular.

Answer: Let $A = V\Sigma W^*$ be the SVD of A. Choose $E = -v_n \sigma_n w_n^*$ where σ_n is the smallest singular value of A and v_n, w_n are the corresponding left and right singular vectors, respectively.

SVD Applications: Minimal difference

Problem: Let $A, B \in M_{m,n}$, calculate

$$\min_{\mathrm{rank}(B)=k} ||A - B||_F$$

Answer: Let $A = V\Sigma W^*$ be the SVD of A. Choose $B = V_k \Sigma_k W_k^*$ where V_k is the matrix with the k left singular vectors corresponding to the k largest singular values etc..

If σ_l are the singular values of A, then,

$$\min_{\mathrm{rank}(B)=k} ||A - B||_F^2 = \sum_{l=k+1}^{n} \sigma_l^2$$

These results can be generalized to all unitarily invariant norms.

SVD Applications: Least Squares/Curve Fitting

Problem: Let $A \in M_{m,n}$, $b \in \mathbf{C}^m$, and $x \in \mathbf{C}^n$. Solve

$$\min_x ||Ax - b||_2$$

Answer: Let $A = V\Sigma W^*$ be the SVD of A and define

$\Sigma^\dagger$ = transpose of Σ in which $\sigma_l > 0$ is replaced by $1/\sigma_l$

$A^\dagger = W\Sigma^\dagger V^*$

($A^\dagger$ is called the Moore-Penrose pseudo inverse of A)

One solution is $x = A^\dagger b$. It is the unique solution if $\mathrm{rank}(A) = n$.

If $\mathrm{rank}(A) < n$, it is the solution with minimum (Euclidean) norm.

SVD Applications: Procrustes Problem

Problem: Let $A, B \in M_{m,n}$ be given. Find a unitary matrix $U \in M_m$ such that

$$||A - UB||_F$$

is minimized.

Answer: Let $AB^* = V\Sigma W^*$ be the SVD of AB^*. Then the minimum is obtained by letting $U = VW^*$.

SVD Applications: Total least squares (TLS)

Let $A, E \in M_{m,n}$ and $B, R \in M_{m,k}$.

Find X that solves the linear system of equations

$$(A + E)X = B + R$$

when E and R are as “small” as possible. More precisely, solve

$$\min_{E,R} ||[E, R]||_F$$

subject to $\mathrm{range}(B + R) \subseteq \mathrm{range}(A + E)$. If $[E_0, R_0]$ is a solution, then X is a TLS solution if it solves

$$(A + E_0)X = B + R_0$$

Solved using SVD; see “Matrix Computations” by Golub & Van Loan.

SVD Applications: Nuclear norm

Let $Y \in M_{m,n}$ have a singular value decomposition $Y = U\Sigma V^*$ where $\Sigma = \mathrm{diag}(\{\sigma_i\}_{i=1}^{\min(m,n)})$.

The nuclear norm of Y is

$$\|Y\|_* = \sum_{i=1}^{\min(m,n)} \sigma_i$$

(also called trace norm, Ky-Fan norm).

SVD Applications: Shrinkage

Consider the problem

$$Z = \arg\min_X \frac{1}{2}\|X - Y\|_F^2 + \|X\|_*$$

Solution

$$Z = \text{shrink}(Y, \tau) = U S_\tau(\Sigma) V^*$$

where

$$S_\tau(\Sigma) = \text{diag}(\{(\sigma_i - \tau)_+\})$$

and $(t)_+ = \max(0, t)$.

SVD Applications: Matrix completion

Matrix completion problem:

$$\min_X \ \text{rank}\, X$$

subject to $X_{ij} = M_{ij},\ (i,j) \in \Omega$

Convex relaxation:

$$\min_X \|X\|_*$$

subject to $X_{ij} = M_{ij},\ (i,j) \in \Omega$

SVD Applications: Singular value thresholding

Modified problem:

$$\min_X \ \tau\|X\|_* + \frac{1}{2}\|X\|_F^2$$

subject to $X_{ij} = M_{ij},\ (i,j) \in \Omega$

Algorithm ($Y_0 = 0$):

$$X_k = \text{shrink}(Y_{k-1}, \tau)$$
$$Y_k = Y_{k-1} + \delta_k P_\Omega(M - X_k)$$

where $P_\Omega(\cdot)$ “projects” on Ω, and δ_k is a step-size parameter.

The Polar Decomposition

Theorem: Let $A \in M_{m,n}$ with $m \leq n$. Then A may be factored as

$$A = PU$$

where

- $P \in M_m$ is psd (and hence Hermitian),
- $\text{rank}(P) = \text{rank}(A)$
- U has orthonormal rows ($UU^* = I$)

Observation: Always unique $P = (AA^*)^{1/2}$.
If A has full rank, then $U = P^{-1}A$ is unique.