Lecture 2
Some background on linear algebra

– square matrices, eigenvalues

– general matrices, singular values

– symmetric matrices

– positive semidefinite matrices

– eigenvalues

– Schur complements

– variational characterizations

– singular values

Square matrices, eigenvalues

A square, $n \times n$ matrix A posesses n eigenvalues (in general complex)

If A has real coefficient, the eigenvalues are real, or come in complex-conjuguate pairs

To each eigenvalue λ, we can associate eigenvectors:

$$Ax = \lambda x$$

Interpretation: A behaves like a scalar along direction x

Caracterization: eigenvalues are solutions to *caracteristic equation:*

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

(I denotes identity matrix of size equal to A's)

General matrices, singular values

Every rectangular matrix A cannot be "diagonalized" but:

There exist othorgonal matrices U, V and a diagonal matrix with positive elements such that

$$A = U\Sigma V^T$$

called the *singular value decomposition* (SVD) of A

Interpretation: Σ contains the "gains" (called *singular values*) and U, V are rotations in input and output space induced by A

fact: computing the SVD can be done in $O(n^3)$ flops (can you make this statement precise?)

Rank, condition number

The *rank* of A is the number of non-zero singular values

The condition number of A is

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} \in [1, \ \infty[$$

(equal to ∞ iff A is singular)

Generalized inverses

With SVD, we can define a generalized inverse (denoted $A^\dagger$) of A: if

$$A = U\Sigma V, \text{ with } \Sigma = \mathbf{diag}\,(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$$

and $\sigma_r > 0$ (so that $r = \mathbf{Rank}A$), then let

$$A^\dagger = U\Sigma^\dagger V, \text{ with } \Sigma^\dagger = \mathbf{diag}\,(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0).$$

interpretation: if $A \in \mathbf{R}^{m\times n}$, and $b \in \mathbf{R}^m$, then the problem

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

has more than one solution if A is singular. The minimum-norm solution is

$$x = A^\dagger b$$

(try to prove this)

Symmetric matrices

a square $n \times n$ matrix A is *symmetric* if $A = A^T$
(^{T} denotes transpose)

If A is symmetric,

- every eigenvalue of A is real
- A can be diagonalized in an orthonormal basis:

  $$A = U\Lambda U^T$$

  where Λ is diagonal and $U^TU = UU^T = I$

Notation: if $A = A^T$,

- $\lambda_{\max}(A)$ denotes *largest* eigenvalue of A
- $\lambda_{\min}(A)$ denotes *smallest* eigenvalue of A

(makes sense since eigenvalues are real)

Interpretations

orthonormal basis for $\mathbf{R}^n$

$$U = [u_1 \; u_2 \; \cdots \; u_n] \in \mathbf{R}^{n\times n}$$

– $U^TU = I$, so $U^{-1} = U^T$

– $\{u_1, \ldots, u_n\}$ is an orthonormal basis for $\mathbf{R}^n$

expansion in orthonormal basis: $x = UU^Tx$, *i.e.*,

$$x = \sum_{i=1}^{n} (u_i^T x) u_i$$

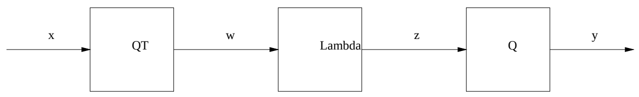
– $u_i^T x$ is the *component* of x in the direction u_i

– $y = U^T x$ *resolves* x into the vector of its u_i components

– $x = Uy$ *reconstitutes* x from its u_i components

– $x = Uy = \sum_i y_i u_i$ is called the $(u_i$-$)$ *expansion* of x

$$A = U\Lambda U^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T$$



linear mapping $y = Ax$ can be decomposed as

– resolve into u_i coordinates

– scale coordinates by λ_i

– reconstitute with basis u_i

or, geometrically,

– rotate by U^T

– diagonal real scale (‘dilation’) by Λ

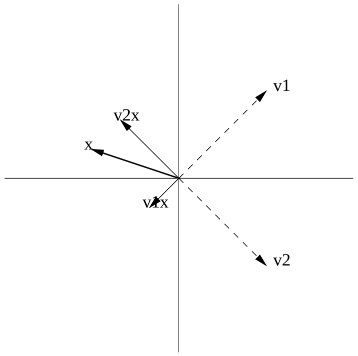
– rotate back by U

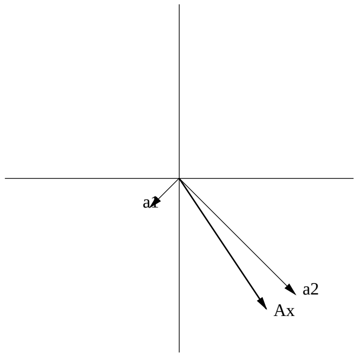
$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^T$$

decomposes A into linear combination of 1-dimensional orthogonal projections

example:

$$\begin{aligned} A &= \begin{bmatrix} -1 & 3 \\ 3 & -1 \end{bmatrix} \\ &= \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right)^T \end{aligned}$$





Positive semidefinite matrices

To each symmetric matrix we can associate a quadratic form

$$f_A(x) := x^T A x$$

(conversely any quadratic form defines a symmetric matrix)

$A = A^T$ is said to be *positive semidefinite* (psd) if associated quadratic form takes only nonnegative values:

$$\forall\ x, \quad x^T A x \geq 0$$

notation: for a given matrix A, $A \succeq 0$ means A is symmetric and psd

(in other words, symmetry of A implicit)

We say A is *positive definite* if strict inequality holds for every $x \neq 0$ (denote $A \succ 0$)

fact: set of psd matrices forms a *cone* (see later)

Trace, inner products

The **trace** of a square matrix is the sum of its diagonal elements

We can define a scalar product in the space of symmetric matrices

$$< A, B > := \mathbf{Tr}\,(AB)$$

lemma: if $A, B \succeq 0$, then

- $< A, B > \geq 0$
  (interpretation: cone of psd matrices forms a 90^o angle at origin)
- $< A, B > = 0$ if and only if $AB = 0$
  (in other words: A, B are "complementary", see later)

Linear maps and their dual

To a linear map Φ from $\mathbf{R}^n$ to $\mathbf{R}^m$,

$$y = \Phi(x) = Ax$$

where A is a $m \times n$ matrix, we can associate a **dual** map:

$$\Phi^* : \mathbf{R}^m \longrightarrow \mathbf{R}^m$$

defined by

$$\text{for every } x, y, \quad y^T \Phi(x) = \Phi^*(y)^T x$$

The dual map is defined by the matrix A^T:

$$\Phi^*(y) = A^T y$$

Dual matrix maps

In this course, we sometimes consider linear maps that take values in the space of symmetric matrices

Consider for example a symmetric $n \times n$ matrix X and a linear function of X that takes values in the space of symmetric, $m \times m$ matrices:

$$\Phi : \ X \longrightarrow Y = \Phi(X)$$

Then the dual map is defined as follows:

$$\text{for every } X, Y, \ \ < Y, \Phi(X) > = < \Phi^*(Y), X >$$

Example: define

$$\begin{array}{rcl} \Phi : \ \mathbf{R}^{n\times n} & \to & \mathbf{R} \\ X & \to & \Phi(X) = \mathbf{Tr}\, X \end{array}$$

then

$$\begin{array}{rcl} \Phi^* : \ \mathbf{R} & \to & \mathbf{R}^{n\times n} \\ s & \to & \Phi^*(s) = s \cdot I_n \end{array}$$

Cholesky decomposition, square root

if $A = A^T \succeq 0$ there exist a (non unique) matrix R such that

$$A = R^T R$$

R is called a **Cholesky factor** of A

there exist a **unique** Cholesky factor with the property that $R = R^T \succeq 0$: the **square root** of A, denoted $A^{1/2}$

if A admits the eigenvalue decomposition

$$A = U \Lambda U^T,$$

where

$$\Lambda = \mathbf{diag}\,(\lambda_1, \ldots, \lambda_n),$$

with $\lambda_i \geq 0$, then the square root is

$$A^{1/2} = U^T \Lambda^{1/2} U,$$

where

$$\Lambda^{1/2} = \mathbf{diag}\,(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$$

Generalized eigenvalues

if $A = A^T$, $B = B^T \succ 0$, the *generalized eigenvalues* of (A, B) are the eigenvalues of symmetric matrix

$$M = B^{-1/2} A B^{-1/2}$$

notation:

- $\lambda_{\max}(A, B)$ denotes **largest** eigenvalue of M
- $\lambda_{\min}(A, B)$ denotes **smallest** eigenvalue of M
- $A \succ B$ (resp. $A \succeq B$) means $\lambda_{\min}(A, B) > 0$ (resp. ≥ 0)

  (note above only holds when $B \succ 0$)

Variational caracterization of eigenvalues

if $A = A^T$,

$$\begin{aligned} \lambda_{\max}(A) &= \max_{x \neq 0} \frac{x^T A x}{x^T x} \\ &= \max_{\|x\|_2 = 1} x^T A x \\ &= \min\{\lambda \ : \ \lambda I - A \succeq 0\} \end{aligned}$$

hence: $\lambda > \lambda_{\max}(A)$ iff $\lambda I \succ A$

similar characterizations for minimum eigenvalue, maximum generalized eigenvalue, and other (eg, eigenvalue sums)

hence: if $B \succ 0$ $\lambda > \lambda_{\max}(A, B)$ iff $\lambda B \succ A$

Matrix norms

For general, rectangular matrices, induced norm:

$$\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

When both vector norm in above ratio are taken to be Euclidean-norm:

$$\|A\| = \sigma_{\max}(A)$$

interpretation: $\sigma_{\max}(A)$ is maximum "gain"

Other common norms include

$$\|A\|_F = \sqrt{\textbf{Tr}\, A^T A}$$

interpretation: can you find it? (think about expected values)

Schur complements

consider partioned symmetric matrix

$$X = X^T = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

$S = C - B^T A^{-1} B$ is the *Schur complement* of A in X (provided $\det A \neq 0$)

Schur complement lemma:

- $X \succ 0$ if and only if $A \succ 0$ and $S \succ 0$
- if $A \succ 0$, then $X \succeq 0$ if and only if $S \succeq 0$

Example

the (convex) quadratic inequality on vector x

$$(Ax+b)^T(Ax+b) - c^Tx - d \leq 0$$

is equivalent to the matrix inequality

$$\begin{bmatrix} I & Ax+b \\ (Ax+b)^T & c^Tx+d \end{bmatrix} \succeq 0$$

note: x enters

- nonlinearly in scalar appearing in first constraint
- linearly in matrix appearing in second constraint

Variational characterizations of singular values

for a general (rectangular) matrix M:

$$\begin{aligned} \sigma_{\max}(M) &= \max_{x \neq 0} \frac{\|Mx\|}{\|x\|} \\ &= \max_{\|x\|_2=1} \|Mx\| \\ &= \min\{\lambda \ : \ \lambda^2 I - M^T M \succeq 0\} \\ &= \min\{\lambda \ : \ \begin{bmatrix} \lambda I & M^T \\ M & \lambda I \end{bmatrix} \succeq 0\} \end{aligned}$$

(note use of Schur complements here)

similar characterizations for minimum singular value

Notations

In this course:

- For a **vector** x, the notation

$$x \succeq 0$$

  means x is positive componentwise, *i.e.*:

$$x_i \geq 0 \text{ for every } i$$

- For a **matrix** X, the notation

$$X \succeq 0$$

  means X is positive semidefinite ($u^T X u \geq 0$ for every vector u)

Same distinction with strict inequalities