Diagonalization of Real Symmetric Matrices

A real symmetric matrix $A_{n\times n}$ can be written as:

$$A = \sum_{i=1}^{n} \underbrace{\lambda_i}_{\text{eigen values (real)} \in \mathbb{R}} \underbrace{u_i u_i^T}_{\text{orthonormal e.vectors} \in \mathbb{R}^n}$$

$$= U \Lambda U^T$$

where

$$U = [u_1\ u_2\ \dots\ u_n],\quad \Lambda = \mathrm{diag}(\lambda_1, \dots \lambda_n)$$

Note $\quad U^T U = I \Rightarrow U^T = U^{-1}$ $\quad$ U is an orthogonal matrix

Result $\quad$ For any $x \in \mathbb{R}^n$,

$$\lambda_{min} \|x\|^2 \leq x^T A x \leq \lambda_{max} \|x\|^2$$

Proof $\quad$ Since u_i's are orthonormal,

$$x = \sum_{i=1}^{n} \alpha_i u_i \text{ for some } \alpha_i\text{'s} \in \mathbb{R}.$$

$$\Rightarrow \quad x^T A x = \left(\sum_{i=1}^{n} \alpha_i u_i\right)^T A \left(\sum_{j=1}^{n} \alpha_j u_j\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j u_i^T \underbrace{A u_j}_{= \lambda_j u_j}$$

$$= \sum_{i=1}^{n} \alpha_i^2 \lambda_i$$

$$\Rightarrow \quad \lambda_{min} \underbrace{\sum \alpha_i^2}_{\|x\|^2} \leq x^T A x \leq \lambda_{max} \underbrace{\sum \alpha_i^2}_{\|x\|^2}$$

"=" achieved if x = e. vector for λ_{min}

"=" achieved if x = e. vector for λ_{max}

Trace $A_{n\times n}$, $Tr(A) = \sum_{i=1}^{n} A_{kk}$

Invariance property: $A_{m\times n}$, $B_{n\times k}$, $C_{k\times m}$

$$Tr(ABC) = Tr(CAB) = Tr(BCA)$$

Determinant and Trace in terms of eigenvalues: $A_{n\times n}$

$$\det(A) = \prod_{i=1}^{n} \lambda_i, \quad Tr(A) = \sum_{i=1}^{n} \lambda_i$$

Definitions (Definiteness)

Let $A_{n\times n}$ be a symmetric matrix. Then A is:

- Positive definite (PD): $x^T A x > 0$, $\forall x \in \mathbb{R}^n$, $x \neq 0$
- Positive semi-definite (PSD): $x^T A x \geq 0$, $\forall x \in \mathbb{R}^n$
- Negative definite (ND): $x^T A x < 0$, $\forall x \in \mathbb{R}^n$, $x \neq 0$
- Negative semi-definite (NSD): $x^T A x \leq 0$, $\forall x \in \mathbb{R}^n$
- Indefinite: $x^T A x < 0$ for some x, > 0 for some other x

A is ND (NSD) $\iff$ $-A$ is PD (PSD)

Note: We can extend definitions to non-symmetric $n \times n$ matrices, but

$$x^T A x = x^T A^T x \Rightarrow x^T A x = x^T \underbrace{\left(\frac{A + A^T}{2}\right)}_{\text{symmetric part of } A} x$$

Tests for PD Symmetric matrix A is PD

$\iff$ All eigenvalues of A are >0

$\iff$ All **leading** principal minors of A are >0

Example

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

We saw in lec 1 that $\lambda_1 = 3$, $\lambda_2 = \lambda_3 = 1$

$\Rightarrow$ A is PD

Leading p.m. test:

$$\det([1]) = 1 > 0 \qquad \det\left(\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right) = 2 > 0$$

$$\det(A) = 3 > 0 \qquad \Rightarrow \quad A \text{ is PD}$$

Tests for PSD Symmetric matrix A is PSD

$\iff$ All eigenvalues of A are ≥ 0

$\iff$ All principal minors of A are ≥ 0

(leading not enough)

Examples

$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$
- $\lambda_1 = 1$, $\lambda_2 = 0$ $\Rightarrow$ PSD
- also by p.m. test

$\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$
- $\lambda_1 = 0$, $\lambda_2 = -1$ not PSD
- leading p.m.'s all ≥ 0
- include all p.m.'s:

  $\det([-1]) < 0 \Rightarrow$ not PSD

$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$ is PSD

Principal Minor test for ND, NSD

Test $-A$ for PD, PSD.

Examples

1. $A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$ $\quad -A = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ is PSD

$\Rightarrow$ A is NSD

2. $B = \begin{bmatrix} -3 & -3 & 0 \\ -3 & -10 & -1 \\ 0 & -1 & -8 \end{bmatrix}$ $\quad \det([-3]) < 0$

$\Rightarrow$ B is not PSD

$-B = \begin{bmatrix} 3 & 3 & 0 \\ 3 & 10 & 1 \\ 0 & 1 & 8 \end{bmatrix}$ $\quad \det([3]) > 0$

$\det\left(\begin{bmatrix} 3 & 3 \\ 3 & 10 \end{bmatrix}\right) > 0$

$\det(-B) = 3(79) - 3(24) > 0$

$\Rightarrow$ $-B$ is PD $\Rightarrow$ B is ND.

3. $C = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$ leading p.m.'s: $\det([1]) > 0$

$\det(C) = -8 < 0$

$\Rightarrow$ C is not PSD

$-C = \begin{bmatrix} -1 & -3 \\ -3 & -1 \end{bmatrix}$ $\quad \det([-1]) < 0$

$\Rightarrow$ $-C$ is not PSD

$\Rightarrow$ C is not NSD

C is neither PSD nor NSD

C is indefinite

Matrix Norm (Induced Norm) Given any vector norm in $\mathbb{R}^n$ we have the following induced norm on $A_{n\times n}$:

$$\|A\| = \max_{\{x\in\mathbb{R}^n : \|x\|=1\}} \|Ax\| = \max_{x\neq 0} \frac{\|Ax\|}{\|x\|}$$

$\uparrow$ set is compact $\Rightarrow$ max attained
$\|Ax\|$ continuous

Spectral Radius For $n\times n$ matrix A,

$$\rho(A) = \max_{i=1,\dots n} |\lambda_i|$$

Result $\rho(A) \leq \|A\|$

Proof Let λ be an eigenvalue of A and u corresponding unit eigenvector

Then $\|Au\| = |\lambda|\,\|u\| = |\lambda|$

$$\max_{\{x : \|x\|=1\}} \|Ax\| \geq \|Au\| = |\lambda| \quad \forall \text{ eigenvalues } \lambda$$

$$\Rightarrow \quad \|A\| \geq \max_i |\lambda_i| = \rho(A)$$

Result For symmetric $A_{n\times n}$, $\|A\| = \rho(A)$

Proof Since $\rho(A) \leq \|A\|$ for all $A_{n\times n}$, we need to show $\rho(A) \geq \|A\|$ for symmetric A

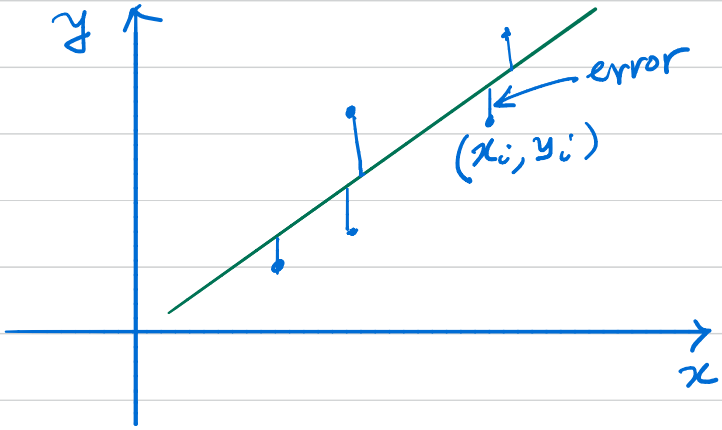
For $x \in \mathbb{R}^n$, $x = \sum_{i=1}^n \alpha_i u_i \leftarrow$ orthonormal e. vectors

$$\|Ax\|^2 = \Big\|\sum_{i=1}^n \alpha_i A u_i\Big\|^2 = \Big\|\sum_{i=1}^n \alpha_i \lambda_i u_i\Big\|^2$$

$$= \sum_{i=1}^n |\alpha_i|^2 |\lambda_i|^2 \leq \underbrace{\Big(\sum_{i=1}^n |\alpha_i|^2\Big)}_{\|x\|^2} \rho(A)^2$$

Examples of Optimization Problems

1. Least Squares (Linear Regression):



Need to find "best" linear fit for data of form:

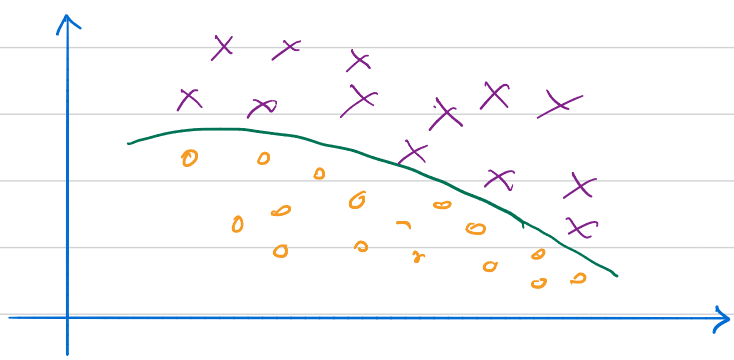
$$y_i = a^T x_i + b$$

scalar (y_i), vector (x_i)

Residual Sum of Squares (RSS): $\sum_{i=1}^{N} \left(y_i - (a^T x_i + b)\right)^2$

Optimization Problem: $\min_{\underline{a},\, b} \; RSS(\underline{a}, b)$

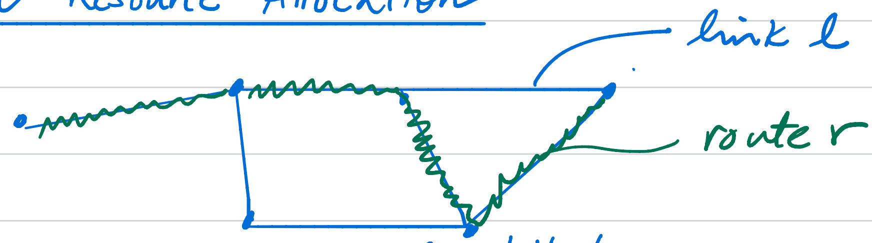
2. Supervised Learning – Binary Classification



Goal: Learn "best" boundary separating classes

Examples: Support Vector Machine, Neural Network posed as optimization problems

3. Internet Resource Allocation



C_l : Capacity of link l bits/s

$l \in r$ if link l belongs to route r

x_r : data rate (bits/s) for files on route r

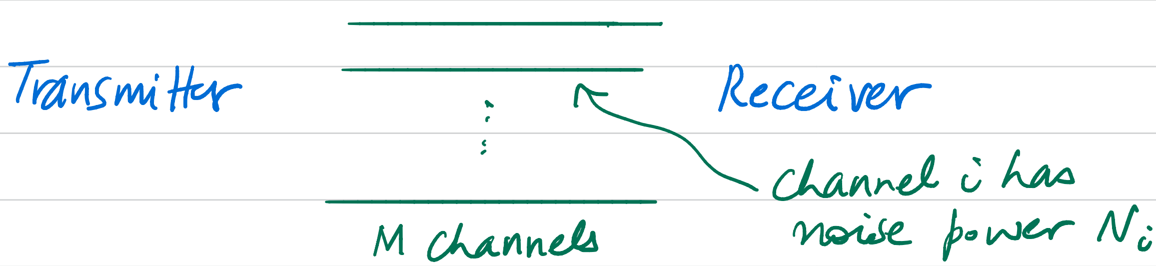
$U_r(x_r)$: Utility to user on route r due to transmission at rate x_r

Optimization Problem : $\max\limits_{\{x_r\}} \sum\limits_r U_r(x_r)$

$$\text{s.t.} \quad \sum_{r: l \in r} x_r \leq C_l \quad \forall l$$

$$x_r \geq 0 \quad \forall r$$

4. Power Allocation in wireless Networks



Total power P needs to be allocated on M channels

P_i = Power on channel i $\Rightarrow$ Rate on i = $\log\left(1+\frac{P_i}{N_i}\right)$

Optimization Problem : $\max \sum\limits_{i=1}^{M} \log\left(1+\frac{P_i}{N_i}\right)$

$$\text{s.t.} \quad \sum P_i \leq P, \quad P_i \geq 0 \quad \forall i$$

Conditions for Optimality

Consider $\min\limits_{x \in S}$ or $\max\limits_{x \in S} f(x)$ $\quad x \in \mathbb{R}^n$, $S \subseteq \mathbb{R}^n$

f : real-valued objective function

S : Constraint set

$x^* \in S$ is called a global minimum if

$$f(x^*) \leq f(x) \quad \forall x \in S$$

$x^* \in S$ is a unique global min. if

$$f(x^*) < f(x) \quad \forall x \neq x^*$$

$x^* \in S$ is a local min. if $\exists \varepsilon > 0$ s.t.

$$f(x^*) \leq f(x) \quad \forall x \in S : \|x - x^*\| < \varepsilon$$

(Inequalities reversed for max)

All global min. are local min. but not all local min are global min.