

COMP50011 Computational Techniques

Imperial College London

Boxuan Tang

Spring 2023

Contents

| | | |
|----------|---|----------|
| 1 | Part 1 | 2 |
| 1.1 | Vector and Matrix Norms | 2 |
| 1.2 | Complex Matrices | 2 |
| 1.3 | Least Square Method | 2 |
| 1.4 | Spectral Decomposition of Symmetric Matrices | 2 |
| 1.5 | Singular Value Decomposition | 3 |
| 1.6 | Generalised Eigenvectors | 3 |
| 1.7 | Cholesky Decomposition | 3 |
| 1.8 | QR Decomposition | 4 |
| 1.9 | QR Algorithm | 4 |
| 1.10 | Fixed Point and Contracting Mapping Theorem | 4 |
| 2 | Part 2 | 4 |
| 2.1 | Condition Number | 4 |
| 2.2 | Iterative Solutions for Linear Equations | 4 |
| 2.3 | Iterative Techniques for Eigenvalues and Eigenvectors | 4 |
| 2.4 | Functions of Several Variables | 5 |
| 2.5 | Gradient Based Optimisation | 5 |
| 2.6 | Linear Programming | 5 |

1 Part 1

1.1 Vector and Matrix Norms

Definition 1.1 (l_p norm of vector in \mathbb{R}^n) The l_p norm of $x \in \mathbb{R}^n$ is defined as $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$.
 $\|x\|_1 = \sum_{i=1}^n |x_i|$ $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ (Euclidean norm) $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$ with $\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1$

Theorem 1.2 (Equivalence of all vector norms) Any 2 vector norms $\|\cdot\|_a$ and $\|\cdot\|_b$ are equivalent, meaning $\exists r, s > 0, \forall x \in \mathbb{R}^n, r\|x\|_a \leq \|x\|_b \leq s\|x\|_a$

Definition 1.3 (Matrix norm on $\mathbb{R}^{m \times n}$) is a real-valued map $\|\cdot\| : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ satisfying:

- For any non-zero matrix $A \in \mathbb{R}^{m \times n} : \|A\| > 0$
- For any scalar λ and $A \in \mathbb{R}^{m \times n} : \|\lambda A\| = |\lambda| \|A\|$
- For $A, B \in \mathbb{R}^{m \times n} : \|A + B\| \leq \|A\| + \|B\|$ (triangle inequality)
- It is a **sub-multiplicative** norm if $\forall A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}, \|AB\| \leq \|A\| \|B\|$

$\|A\|_1 = \max_j \|a_j\|_1$ (Column sum) $\|A\|_2 = \sigma_1(A)$ (Largest singular value) $\|A\|_\infty = \max_i \|a^i\|_1$ (Row sum)

Definition 1.4 (Consistent and compatible) A matrix norm $\|\cdot\|$ on $\mathbb{R}^{m \times n}$ is consistent with the vector norms $\|\cdot\|_a$ on \mathbb{R}^n and $\|\cdot\|_b$ on \mathbb{R}^m if $\forall A \in \mathbb{R}^{m \times n}, x \in \mathbb{R}^n : \|Ax\|_b \leq \|A\| \|x\|_a$.
 If $a = b$, $\|\cdot\|$ is compatible with $\|\cdot\|_a$.

Definition 1.5 (Subordinate matrix norm) For a vector norm $\|\cdot\|$ on \mathbb{R}^n the subordinate matrix norm $\|\cdot\|$ on $\mathbb{R}^{m \times n}$ is: $\forall A \in \mathbb{R}^{m \times n}, \|A\| = \max\{\|Ax\| : x \in \mathbb{R}^n, \|x\| = 1\}$

Theorem 1.6 A vector norm is compatible with its matrix norm, so $\forall A \in \mathbb{R}^{m \times n}, x \in \mathbb{R}^n \|Ax\| \leq \|A\| \|x\|$

1.2 Complex Matrices

Definition 1.7 (Standard Inner Product & Norm) The inner product $\langle \cdot, \cdot \rangle$ is defined as $\forall u, v \in \mathbb{C}^n, \langle u, v \rangle = \bar{u}^T v = \sum_{i=1}^n \bar{u}_i v_i = \langle v, u \rangle$ (non-symmetric). The **norm** of $u \in \mathbb{C}^n$ is defined as $\|u\| = \sqrt{\langle u, u \rangle}$.

Definition 1.8 (Common Formulas) $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \times \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

1.3 Least Square Method

Definition 1.9 (Least Square Method) $A^T A x = A^T b$ finds the $x \in \mathbb{R}^n$ where $\|Ax - b\|_2$ is minimised.

Definition 1.10 (Linear Regression) Given $y(t) = \sum_{j=1}^n s_j f_j(t)$. Solve LSM for $As = y$ where

$$A = \begin{bmatrix} f_1(t_1) & f_2(t_1) & \dots & f_n(t_1) \\ \vdots & \vdots & & \vdots \\ f_1(t_m) & f_2(t_m) & \dots & f_n(t_m) \end{bmatrix} \text{ e.g. given } y(t) = h - gt^2/2, \text{ solve } \begin{bmatrix} 1 & -t_1^2/2 \\ \vdots & \vdots \\ 1 & -t_n^2/2 \end{bmatrix} \begin{bmatrix} h \\ g \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

1.4 Spectral Decomposition of Symmetric Matrices

Definition 1.11 (Orthogonal) A matrix $Q \in \mathbb{R}^{n \times n}$ is orthogonal iff it is invertible and $Q^{-1} = Q^T$

- preserve Euclidean **length** and **angle**, so $\|Qu\|_2 = \|u\|_2$ and $\widehat{QuQv} = \widehat{uv}$
- **determinant** of $Q, |Q| = \pm 1$ and all **eigenvalues** $\lambda \in \mathbb{C}$ have $|\lambda| = \pm 1$

Definition 1.12 (Symmetric) A matrix $A \in \mathbb{R}^{n \times n}$ is symmetric iff $A^T = A$. If A is real,

- all **eigenvalues** of A are real
- for all eigenvalues λ_i of A , algebraic multiplicity of λ_i = geometric multiplicity of λ_i
- eigenvectors for distinct eigenvalues are orthogonal

Theorem 1.13 (Spectral Theorem) If $A \in \mathbb{R}^{n \times n}$ is a real symmetric matrix, $A = QDQ^T = QDQ^{-1}$

1. Find the eigenvalues of A from its characteristic polynomial
2. For each eigenvalue, find its eigenspace and an orthonormal basis
3. Combine the bases to form Q . Form the diagonal matrix D using the associated eigenvalues.

1.5 Singular Value Decomposition

Definition 1.14 (Positive (semi-)definite) A **symmetric** real matrix $A \in \mathbb{R}^{n \times n}$ is

- +ve-definite $\Leftrightarrow x \neq 0, x^T A x > 0 \Leftrightarrow$ eigenvalues > 0 (diagonals > 0 and $\max(A_{ii}, A_{jj}) > |A_{ij}|$)
- +ve semi-definite $\Leftrightarrow x^T A x \geq 0 \Leftrightarrow$ eigenvalues ≥ 0 (diagonals ≥ 0 and $\max(A_{ii}, A_{jj}) \geq |A_{ij}|$)

If A is positive (semi-)definite, the sub-matrices in the upper left corner are also positive (semi-)definite. Let $A \in \mathbb{R}^{m \times n}$ then $A^T A, A A^T$ are symmetric and positive semi-definite

Definition 1.15 (Singular Value Decomposition) Let $A \in \mathbb{R}^{m \times n}$. Then $A = U S V^T$ where $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{n \times n}$ are orthogonal $S \in \mathbb{R}^{m \times n} = \text{diag}(\sigma_1, \sigma_2 \dots \sigma_p), p = \min(m, n)$, with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ the **singular values** of A

- If $U = [u_1 \dots u_m]$ and $V = [v_1 \dots v_n]$ then $A = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T$
- positive singular values of A are the positive square roots of the eigenvalues of $A A^T$ or $A^T A$
- number of positive singular value of A equals rank of A and $\|A\|_2 = \sigma_1$, the largest singular value of A
- $\text{span}(u_i)$ where $\sigma_i > 0$ form an orthonormal basis for the $\text{im}(A)$ (U 's left columns reading from $U S V^T$)
- $\text{span}(v_i)$ where $\sigma_i = 0$ form an orthonormal basis for the $\text{ker}(A)$ (V 's bottom rows reading from $U S V^T$)

1. **More rows than columns**, so $A^T A$ is smaller than $A A^T$. Otherwise find SVD for A^T ,
2. Obtain eigenvalues $\sigma_1^2 \geq \dots \sigma_n^2 \geq 0$ and orthonormal eigenvectors $v_1 \dots v_n$ (forming V) of $A^T A$.
3. Let $u_i = \frac{1}{\sigma_i} A v_i$ for $1 \leq i \leq r$. Check $\|u_i\|_2 = 1, u_i^T u_j = 0$ and $A A^T u_i = \sigma_i^2 u_i$ for $1 \leq i, j \leq r, i \neq j$
4. Extend the set $u_1 \dots u_r$ to an orthonormal basis to obtain U

Definition 1.16 (Principal Component Analysis) Assume $A \in \mathbb{R}^{m \times n}$ represents m samples of n dimensional data. The principal axes of A are the v_i for $1 \leq i \leq r$. The first principal axis of A , $w_{(1)} = \arg \max_{\|w\|=1} w^T A^T A w = v_1$ the column vector associated with σ_1 . If $\sigma_1 \gg \sigma_2$, then $A \approx \sigma_1 u_1 v_1^T$

1.6 Generalised Eigenvectors

Definition 1.17 (Generalised Eigenvector) Given a square matrix $A \in \mathbb{R}^{n \times n}$, a non-zero vector $v \in \mathbb{C}^n$ is a generalised eigenvector of rank m under eigenvalue $\lambda \in \mathbb{C}$ for A if $(A - \lambda I)^m v = 0$ and $(A - \lambda I)^{m-1} v \neq 0$. For every eigenvalue of algebraic multiplicity k , there are k linearly independent generalised eigenvectors.

Definition 1.18 (Jordan Normal Form) 1. Find the eigenvalues $\lambda_1 \dots \lambda_m$ of A .

2. For each eigenvalue λ_i compute the eigenvectors $v_{i1}^1 \dots v_{ig_i}^1$, noting geometric multiplicity g_i .
3. Find the missing generalised eigenvectors where for each eigenvector $v_{ij}^1 \in \mathbb{C}^n$ associated with λ_i , try to find all $v_{ij}^k \in \mathbb{C}^n$ such that $(A - \lambda_i I) v_{ij}^k = v_{ij}^{k-1}$, solving using Gaussian Elimination if needed.
4. Write $B = [v_{11}^1 \dots v_{1g_1}^1 \dots v_{1g_1}^{k_{1g_1}} \dots v_{m1}^1 \dots v_{m1}^{k_{m1}} \dots v_{mg_m}^1 \dots v_{mg_m}^{k_{mg_m}}]$

$$5. B^{-1} A B = J, \text{ where } J = \begin{bmatrix} J_{k_{11}}(\lambda_1) & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & J_{k_{mg_m}}(\lambda_m) \end{bmatrix} \text{ and } J_{k_{ij}}(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & 0 & 0 \\ 0 & \lambda_i & 1 & 0 \\ 0 & 0 & \lambda_i & 1 \\ 0 & 0 & 0 & \lambda_i \end{bmatrix} \text{ of size } k_{ij}$$

1.7 Cholesky Decomposition

Theorem 1.19 A symmetric positive semi-definite matrix A has decomposition $A = L L^T$.

If A is also positive-definite, L is unique and has positive diagonal elements.

$$L = \begin{bmatrix} l_{11} & 0 & \dots \\ l_{21} & l_{22} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \text{ For } n=3, L L^T = \begin{bmatrix} l_{11}^2 & l_{11} l_{21} & l_{11} l_{31} \\ l_{11} l_{21} & l_{21}^2 + l_{22}^2 & l_{21} l_{31} + l_{22} l_{32} \\ l_{11} l_{31} & l_{21} l_{31} + l_{22} l_{32} & l_{31}^2 + l_{32}^2 + l_{33}^2 \end{bmatrix}$$

Given $Ax = b$, Find $A = L L^T$. Solve $L y = b$ by forward substitution then $L^T x = y$ by backward substitution.

1.8 QR Decomposition

Definition 1.20 (Gram-Schmidt Process) $u_j = v_j - (e_1 \cdot v_j)e_1 - \cdots - (e_{j-1} \cdot v_j)e_{j-1}$ and $e_j = \frac{u_j}{\|u_j\|}$

Definition 1.21 (QR Decomposition using Gram-Schmidt Process)

$$\text{Let } Q = [e_1 \dots e_n] \text{ be an orthonormal basis for } A \in \mathbb{R}^{m \times n}. A = Q \begin{bmatrix} (e_1 \cdot a_1) & \dots & (e_1 \cdot a_n) \\ 0 & \ddots & \vdots \\ 0 & \dots & (e_n \cdot a_n) \end{bmatrix}$$

Theorem 1.22 (Householder Map) $H_u = I - 2uu^T$ reflects a point along a plane through the origin with unit normal u . It is involutory ($H_u = H_u^{-1}$) and orthogonal ($H_u = H_u^T$).

1.9 QR Algorithm

Theorem 1.23 (Algorithm) $A_0 = A$. Apply QR decomposition for $A_k = Q_{k+1}R_{k+1}$. Set $A_{k+1} = R_{k+1}Q_{k+1}$

If A is symmetric, it converges to a diagonal matrix and the eigenvectors are the columns of Q_k

Theorem 1.24 (LU decomposition) A non singular matrix $A \in \mathbb{R}^{n \times n}$ can be factorised $A = LU$ iff A can be reduced into its row echolon form without swapping any two rows.
If the diagonal elements of L are all 1, the decomposition is unique.

1.10 Fixed Point and Contracting Mapping Theorem

Definition 1.25 (Fixed Point) Let $f : S \rightarrow S$ and $p \in S$. If $f(p) = p$ then p is a fixed point.

Definition 1.26 (Contraction) f is a contraction on (S, d) if $\exists 0 \leq \alpha < 1, \forall x, y \in S. d(f(x), f(y)) \leq \alpha d(x, y)$

Theorem 1.27 (Fixed Point Theorem) If f is a contraction on a complete metric space (S, d) , f has a unique fixed point.

2 Part 2

2.1 Condition Number

Definition 2.1 (Condition Number) Let problem P be represented by $Ax = b$. Then $\kappa(P) = \kappa(A) = \|A^\dagger\| \|A\|$ where $A^\dagger = (A^T A)^{-1} A^T$. For square matrices, $\kappa(A) = \|A^{-1}\| \|A\|$.

2.2 Iterative Solutions for Linear Equations

Definition 2.2 (General Method) $A = G + R$. $Ax = b \Leftrightarrow x = (-G^{-1}R)x + (G^{-1}b)$.

- **Jacobi Method** Split $A = D + (L + U)$ with $x = (-D^{-1}(L + U))x + (D^{-1}b)$
- **Gauss-Seidel Method** Split $A = (D + L) + U$ with $x = -(D + L)^{-1}Ux + ((D + L)^{-1}b)$
- **Strictly row diagonally dominant** $\forall i, |a_{ii}| > \sum_{j \neq i} |a_{ij}|$ is sufficient for both convergence (G-S faster)

Theorem 2.3 Let $M \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}^n$. Let $\|\cdot\|$ be a consistent norm on $\mathbb{R}^{n \times n}$. If $\|M\| < 1$, then (x_k) defined by $x_{k+1} = Mx_k + c$ converges for any starting point x_0 .

2.3 Iterative Techniques for Eigenvalues and Eigenvectors

Definition 2.4 (Power Iteration) Let $A \in \mathbb{R}^{n \times n}$ be diagonalisable matrix with eigenvalues of distinct modulus. Set any $x_0 \in \mathbb{R}^n \setminus \{0\}$ and $x_{k+1} = \frac{Ax_k}{\|Ax_k\|}$. As $x \rightarrow \infty, x_k \rightarrow v$ and $\|Ax_k\| \rightarrow |\lambda|$ (dominant eigenvalue).

Definition 2.5 (Inverse Power Iteration) If A is also non-singular, $x_{k+1} = \frac{A^{-1}x_k}{\|A^{-1}x_k\|}$ gives $x_k \rightarrow v$ and $\|A^{-1}x_k\| \rightarrow \left|\frac{1}{\lambda}\right|$ where λ is the eigenvalue of A with the smallest modulus.

Definition 2.6 (Rayleigh Quotient) $R(A, x) = \frac{x^T A x}{x^T x}$ approximates eigenvalue and not its modulus

Definition 2.7 (Deflation) Define H_u with $u = v_1 + \|v_1\|_2 e_1$ where $e_1 = [1, 0, \dots]^T$ (so $Hx_1 = \alpha e_1$).

Then $HAH^{-1} = \begin{bmatrix} \lambda_1 & b^T \\ 0 & B \end{bmatrix}$. Then $x_2 = H^{-1} \begin{bmatrix} \beta \\ z_2 \end{bmatrix}$ where z_2 is dominant eigenvector of B and $\beta = \frac{b^T z_2}{\lambda_2 - \lambda_1}$

2.4 Functions of Several Variables

Theorem 2.8 (Clairaut's Theorem) If f_{xy} and F_{yx} are continuous at (a, b) , $f_{xy}(a, b) = f_{yx}(a, b)$.

Definition 2.9 (Gradient Vector) $\nabla f(x, y) = \langle f_x(x, y), f_y(x, y) \rangle$. For unit u , $f_u(x, y) = \nabla f(x, y) \cdot u$. Directional vector is maximum when it has same direction as ∇f (fastest change of f)

Definition 2.10 (Hessian) $D = \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = f_{xx}f_{yy} - (f_{xy})^2$.

- **Local maximum** $D > 0$ and $f_{xx}(a, b) < 0$ or eigenvalues all negative.
- **Local minimum** $D > 0$ and $f_{xx}(a, b) > 0$ or eigenvalues all positive.
- **Saddle Point** $D < 0$ or eigenvalues both positive and negative

2.5 Gradient Based Optimisation

Definition 2.11 (Quadratic Form) $f(x_1, x_2) = ax_1^2 + bx_1x_2 + cx_2^2 + dx_1 + ex_2$ can be expressed as $\frac{1}{2}x^T Ax - x^T B$ where $A = \begin{bmatrix} 2a & b \\ b & 2c \end{bmatrix}$ and $B = \begin{bmatrix} -d \\ -e \end{bmatrix}$ with $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Then $\nabla f(x_1, x_2) = Ax - B$.

Definition 2.12 (Steepest Descent) $x^{(k+1)} = x^{(k)} - \alpha_k \nabla f(x^{(k)})$.

For quadratic functions $\frac{1}{2}x^T Ax - x^T B$, $\alpha_k = \frac{(g^{(k)})^T (g^{(k)})}{(g^{(k)})^T H g^{(k)}}$ where $g^{(k)} = \nabla f(x)$ and $H = A = \nabla^2 f(x)$

Definition 2.13 (Conjugate Gradient) To solve $Ax = b$:

Set starting point x_0 . Set $r_0 = b - Ax_0$.

For k in 1 to n :

Set $D = \{d_1 \dots d_n\}$, a set of n A -conjugate vectors
($\forall i \neq j, d_i^T A d_j = 0$). Set starting point x_0 .

For k in 1 to n :

1. $\alpha_k = \frac{d_k^T (b - Ax_{k-1})}{d_k^T A d_k}$ and $x_k = x_{k-1} + \alpha_k d_k$

1. if $k = 1$ then $d_k = r_0$ else $d_k = r_{k-1} - \beta_k d_{k-1}$
where $\beta_k = -\frac{r_{k-1}^T r_{k-1}}{r_{k-2}^T r_{k-2}}$

2. $\alpha_k = \frac{r_{k-1}^T r_{k-1}}{d_k^T A d_k}$, $x_k = x_{k-1} + \alpha_k d_k$ and
 $r_k = r_{k-1} - \alpha_k A d_k$

2.6 Linear Programming

Definition 2.14 (Simplex Method) Maximise $z = c_1x_1 + \dots + c_nx_n$ with $a_{11}x_1 + \dots + a_{1n}x_n \leq b_1, \dots$

1. **Identify Column** Choose the most negative entry in the top z row
2. **Identify Row** Evaluate the $\frac{\text{solution column}}{\text{positive entries in that column}}$. Choose the smallest ratio.
3. **Change basis** Change the variable in the left-hand column to the non-basic variable of that column.
4. **Make Pivot = 1** Make the pivot 1 by dividing throughout.
5. **Gaussian Elimination** Clear the rest of elements in that column using GE.

| | x_1 | x_n | x_{n+1} | x_{n+m} | Solution |
|-----------|----------|----------|-----------|-----------|----------|
| z | $-c_1$ | $-c_n$ | 0 | 0 | 0 |
| x_{n+1} | a_{11} | a_{1n} | 1 | 0 | b_1 |
| x_{n+m} | a_{m1} | a_{mn} | 0 | 1 | b_m |

Theorem 2.15 (Duality Principle) Objective function of minimisation problem reaches min iff objective function of its dual reaches max. When they do, they are equal.

| | | | | | | | | | |
|------------------------------|----------|----|----|----|-------------------|-------|----|----|----|
| Minimise $Z = 12x_1 + 16x_2$ | Original | 1 | 2 | 40 | | Dual | 1 | 1 | 12 |
| $x_1 + 2x_2 \geq 40$ | (Min) | 1 | 1 | 30 | \Leftrightarrow | (Max) | 2 | 1 | 16 |
| $x_1 + x_2 \geq 30$ | | 12 | 16 | 0 | | | 40 | 30 | 0 |