IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2018

MSc and EEE PART IV: MEng and ACGI

Corrected copy

TOPICS IN LARGE DIMENSIONAL DATA PROCESSING

Tuesday, 15 May 10:00 am

Time allowed: 3:00 hours

There are FOUR questions on this paper.

Answer ALL questions.

All questions carry equal marks

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible First Marker(s) : W. Dai
Second Marker(s) : C. Ling

EE4-66 Topics in Large Dimensional Data Processing

Instructions for Candidates

Answer all questions. Each question carries 20 marks.

1. (Linear Algebra)

(a) Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be a matrix. Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ be its singular value decomposition (SVD), where $\boldsymbol{U} = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m]$ and $\boldsymbol{V} = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_m]$ are the left and right singular vector matrices respectively, and $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ is the diagonal matrix of which the diagonal entries are the singular values $\sigma_1 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$.

i State the definition of the eigenvalue decomposition of a square matrix $\boldsymbol{M} \in \mathbb{R}^{m\times m}$. [1]

ii Let $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{A}^T$. State the relationship between the eigenvalue decomposition of the matrix $\boldsymbol{B}$ and the singular value decomposition of the matrix $\boldsymbol{A}$. Justify your answer using the definition of the eigenvalue decomposition. [2]

iii Let $\boldsymbol{A}^\dagger$ be the pseudo-inverse of $\boldsymbol{A}$. Express $\boldsymbol{A}^\dagger$ using the singular value decomposition of $\boldsymbol{A}$. [1]

iv Suppose that $m > n$ and $\text{rank}(\boldsymbol{A}) = n$. For a vector $\boldsymbol{x} \in \mathbb{R}^m$, denote the projection of $\boldsymbol{x}$ on the subspace span $(\boldsymbol{A})$ by $\boldsymbol{x}_p = \text{proj}(\boldsymbol{x}, \boldsymbol{A})$. Write the formula to compute $\boldsymbol{x}_p$ using the matrices $\boldsymbol{A}$ and $\boldsymbol{A}^\dagger$. [1]

v Prove the projection residue vector $\boldsymbol{x}_r = \boldsymbol{x} - \boldsymbol{x}_p$ is orthogonal to $\boldsymbol{A}$. [2]

vi Prove that the shortest distance between $\boldsymbol{x}$ and any vector $\boldsymbol{v} \in \text{span}(\boldsymbol{A})$ is achieved by $\|\boldsymbol{x} - \boldsymbol{x}_p\|_2$, i.e.,

$$\|\boldsymbol{x} - \boldsymbol{x}_p\|_2 = \min\left\{\|\boldsymbol{x} - \boldsymbol{v}\| : \boldsymbol{v} \in \text{span}(\boldsymbol{A})\right\}.$$

[3]

(b)

i For a given square matrix $\boldsymbol{M}$, let $\text{tr}(\boldsymbol{M}) = \sum_i M_{i,i}$ denote its trace function. For any given matrices $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times m}$, prove that $\text{tr}(\boldsymbol{A}\boldsymbol{B}) = \text{tr}(\boldsymbol{B}\boldsymbol{A})$. [2]

ii The Frobenius norm of a matrix $\boldsymbol{A}$ is defined as $\|\boldsymbol{A}\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$. Express the squared Frobenius norm $\|\boldsymbol{A}\|_F^2$ using the singular values of $\boldsymbol{A}$. Justify your answer. [2]

iii The ℓ_2-norm of a matrix $\boldsymbol{A}$ is defined as

$$\|\boldsymbol{A}\|_2 = \sup\left\{\|\boldsymbol{A}\boldsymbol{x}\|_2 : \boldsymbol{x} \in \mathbb{R}^n \text{ and } \|\boldsymbol{x}\|_2 = 1\right\}.$$

Express $\|A\|_2$ using the singular values of A. Justify your answer. [3]

iv The nuclear norm of a matrix A is defined as

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

Let A be symmetric and non-negative definite, i.e., $A = A^T$ and $A \geq 0$. Prove that

$$\|A\|_* = \mathrm{tr}(A)$$

using the singular value decomposition of A. [3]

(Total marks: 20)

2. (Sparse Recovery)

(a) Consider the sparse recovery problem $y = Ax$ where $A \in \mathbb{R}^{m \times n}$ is a flat matrix, and $x \in \mathbb{R}^n$ is the unknown sparse vector.

i The famous Lasso formulation for sparse recovery is given by

$$\min_{x} \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1 , \tag{2.1}$$

where $\lambda > 0$ is a parameter.

A. The soft thresholding function $\eta(\cdot)$ is designed to solve the simplified Lasso problem

$$\min_{x} \frac{1}{2} \|x - z\|_2^2 + \lambda \|x\|_1 . \tag{2.2}$$

State the form of the soft thresholding function $\eta(\cdot)$. (Derivations are not required.) [1]

B. State the Iterative Shrinkage Thresholding (IST) algorithm to solve the Lasso problem (2.1). (Derivations are not required.) [1]

ii Suppose that the sparsity level S of x is given. The Iterative Hard Thresholding (IHT) algorithm is one of the greedy algorithms designed for sparse recovery.

A. State the hard thresholding function $H_S(\cdot)$. Note that soft thresholding function is designed to solve the simplified Lasso problem (2.2). State the optimisation problem that the hard thresholding function is designed to solve. [2]

B. State the IHT algorithm. Note that IST is designed to solve the Lasso problem (2.1). State the optimisation problem that IHT is designed to solve. [2]

(b) Consider the low-rank matrix recovery problem $y = \mathcal{A}(X)$, where $\mathcal{A} : \mathbb{R}^{n_r \times n_c} \to \mathbb{R}^m$ is a linear operator.

i State the counterpart of the IST algorithm designed to solve the low-rank matrix recovery problem. Give the definition of corresponding soft thresholding function used in your algorithm. [2]

ii State the counterpart of the IHT algorithm designed to solve the low-rank matrix recovery problem. Give the definition of corresponding hard thresholding function used in your algorithm. [2]

iii Consider a matrix $\boldsymbol{X}$ with rank 1. The partial observations of $\boldsymbol{X}$ are given as follows:

$$\begin{bmatrix} 1 & ? & ? \\ 1 & -1 & ? \\ ? & 2 & 4 \end{bmatrix}.$$

A. Find the complete $\boldsymbol{X}$ and explain your steps briefly. [3]

B. Let $\boldsymbol{x}_i$ denote the i^{th} column of the matrix $\boldsymbol{X}$. Let $\mathrm{vec}(\boldsymbol{X}) \in \mathbb{R}^9$ be the column vector by stacking the columns of $\boldsymbol{X}$, i.e., $\mathrm{vec}(\boldsymbol{X}) = \left[\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, \boldsymbol{x}_3^T\right]^T$. Find the matrix form $\boldsymbol{A}$ of the linear operator $\mathcal{A}$. Find the observation vector $\boldsymbol{y} = \mathcal{A}(\boldsymbol{X})$. Find the matrix given by $\mathcal{A}^*(\boldsymbol{y})$ where $\mathcal{A}^*$ is the adjoint operator of $\mathcal{A}$ and defined as $\mathrm{vec}(\mathcal{A}^*(\boldsymbol{y})) := \boldsymbol{A}^T\boldsymbol{y}$. [3]

(c) Consider the blind deconvolution problem $y(t) = \sum_\tau h(\tau)\, x(t-\tau)$ where both $h(t)$ and $x(t)$ are unknown time series. For simplicity, assume that $h(t) = 0$ for $t \notin \{1, 2, 3\}$ and write $\boldsymbol{h} = [h(1), h(2), h(3)]^T$. Assume that $x(t) = 0$ for $t \notin \{1, 2, 3\}$ and write $\boldsymbol{x} = [x(1), x(2), x(3)]^T$. As a result, $y(t) = 0$ for $t \notin \{1, \cdots, 5\}$.

Find the convex optimisation formulation for solving the blind deconvolution problem. Find the matrix form $\boldsymbol{A}$ of the linear operator $\mathcal{A}$ involved in your formulation. [4]

(Total marks: 20)

3. (Convex Optimisation)

(a)

i State the definition of a convex set $S \subset \mathbb{R}^n$. [2]

ii State the definition of a convex function $f : \mathbb{R}^n \to \mathbb{R}$. [2]

iii State the standard form of a convex optimisation problem (with equality and inequality constraints). [2]

iv Let $\boldsymbol{Q}$ be positive definite. Solve the following quadratic optimisation problem by minimising its Lagrangian with respect to $\boldsymbol{x}$:

$$\min_{\boldsymbol{x}} \frac{1}{2}\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{subject to } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}.$$

[3]

(b) The α-sublevel set of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$C_\alpha = \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) \le \alpha\}.$$

i Prove that sublevel sets of a convex function f are convex. [2]

ii For a given $y \in \mathbb{R}^+$, define $f(\boldsymbol{x}) = \left(y - \|\boldsymbol{x}\|_2^2\right)^2$ where $\|\boldsymbol{x}\|_2^2 = \sum_i x_i^2$. Show that it is not convex by studying the sublevel set C_0. [2]

(c) As an extension of Problem 3.(b)ii, the function $f(\boldsymbol{x}) = \sum_i \left(y_i - \left(\boldsymbol{a}_i^T \boldsymbol{x}\right)^2\right)^2$ is not convex. The following presents a recently developed technique to optimise it.

i Show that $f(\boldsymbol{x}) = \sum_i \left(y_i - \operatorname{tr}(\boldsymbol{X}\boldsymbol{A}_i)\right)^2$ where $\boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^T$ and $\boldsymbol{A}_i = \boldsymbol{a}_i \boldsymbol{a}_i^T$. You are allowed to use the fact that $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA})$. [2]

ii To solve the original optimisation problem

$$\min_{\boldsymbol{x}} \sum_i \left(y_i - \left(\boldsymbol{a}_i^T \boldsymbol{x}\right)^2\right)^2, \tag{3.3}$$

one can solve the following optimisation problem

$$\min_{\boldsymbol{X}} \sum_i \left(y_i - \operatorname{tr}(\boldsymbol{X}\boldsymbol{A}_i)\right)^2 + \lambda \operatorname{tr}(\boldsymbol{X})$$
$$\text{subject to } \boldsymbol{X} = \boldsymbol{X}^T,\ \boldsymbol{X} \succeq 0, \tag{3.4}$$

where $X \geq 0$ denotes that the matrix X is non-negative definite.

A. Verify the optimisation problem (3.4) is convex. Show your arguments. [3]

B. Explain the motivations of the terms involved in (3.4). [2]

(Total marks: 20)

4. (Performance Analysis)

Consider the problem $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_0$ where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a flat matrix, and $\boldsymbol{x}_0 \in \mathbb{R}^n$ is the unknown S-sparse vector. Assume that the columns of $\boldsymbol{A}$ have been normalised, i.e., $\|\boldsymbol{a}_i\|_2 = 1$, $\forall i$.

(a) State the definition of the mutual coherence constance $\mu(\boldsymbol{A})$ of the matrix $\boldsymbol{A}$. [1]

(b)

i State the definitions of Restricted Isometry Property (RIP) and Restricted Isometry Constant (RIC). [3]

ii RIP implies the near orthogonality of two disjoint submatrices of $\boldsymbol{A}$. Specifically, let $\mathcal{I}, \mathcal{J} \subset \{1, \cdots, n\}$. Assume that $|\mathcal{I}| = k$, $|\mathcal{J}| = \ell$ and $\mathcal{I} \bigcap \mathcal{J} = \phi$. RIP implies that for all $\boldsymbol{a} \in \mathbb{R}^k$ and $\boldsymbol{b} \in \mathbb{R}^\ell$,

$$|\langle \boldsymbol{A}_{\mathcal{I}}\boldsymbol{a}, \boldsymbol{A}_{\mathcal{J}}\boldsymbol{b}\rangle| \leq c \|\boldsymbol{a}\|_2 \|\boldsymbol{b}\|_2, \tag{4.5}$$

for some constant c. Write c in terms of RIC. (No proof is needed.) [1]

iii If $\boldsymbol{A}$ satisfies the RIP with the RIC $\delta_{2S} < 1$, then the solution of the sparse recovery problem is unique, that is, there does not exist another S-sparse vector $\boldsymbol{x}' \neq \boldsymbol{x}_0$ such that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}'$. Prove this claim. [3]

(c) The Gershgorin circle theorem states the following: for any square matrix $\boldsymbol{M}$, every eigenvalue of $\boldsymbol{M}$ satisfies

$$\lambda(\boldsymbol{M}) \in \left[M_{i,i} - \sum_{j \neq i} |M_{i,j}|, M_{i,i} + \sum_{j \neq i} |M_{i,j}| \right].$$

Consider square matrices of the form $\boldsymbol{M} = \boldsymbol{A}_{\mathcal{I}}^T \boldsymbol{A}_{\mathcal{I}}$ where $\boldsymbol{A}_{\mathcal{I}}$ is an S-column submatrix of $\boldsymbol{A}$. Prove that the following relationship between the RIC δ_S and the mutual coherence constant μ.

$$\delta_S \leq S\mu.$$

[3]

(d) Let $\boldsymbol{c} = \boldsymbol{A}^T\boldsymbol{y}$. Define

$$i^* = \arg\max_i |c_i| = \arg\max_i \left|\boldsymbol{a}_i^T \boldsymbol{y}\right|.$$

We seek for an RIP condition to guarantee that $i^* \in \mathcal{T} := \operatorname{supp}(\boldsymbol{x}_0)$ via the following steps.

i For an $i \notin \mathcal{T}$, find an upper bound bound of $\left|\boldsymbol{a}_i^T \boldsymbol{y}\right|$ in terms of the RIP constant δ_{S+1} and $\|\boldsymbol{x}_0\|_2$. [3]

ii Find a lower bound of

$$\max_{i \in \mathcal{T}} \left|\boldsymbol{a}_i^T \boldsymbol{y}\right|$$

in terms of the RIP constant δ_{S+1} and $\|\boldsymbol{x}_0\|_2$. You may need the fact that

$$\max_{i \in \mathcal{T}} |x_{0,i}| \geq \frac{1}{\sqrt{S}} \|\boldsymbol{x}_0\|.$$

[3]

iii Establish an RIP condition to guarantee that $i^* \in \mathcal{T}$. [3]

(Total marks: 20)

Desk 172 - question
on paper