EE4-66
EE9-SO28

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2019

MSc and EEE PART IV: MEng and ACGI

Corrected copy

TOPICS IN LARGE DIMENSIONAL DATA PROCESSING

Friday, 10 May 10:00 am

Time allowed: 3:00 hours

There are THREE 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 three questions. Each question carries 25 marks.

1. (Matrix Analysis)

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ be a matrix.

(a) State the definition of the mutual coherence constance $\mu(\boldsymbol{A})$ of the matrix $\boldsymbol{A}$. [2]

(b)

i State the definitions of Restricted Isometry Property (RIP) and Restricted Isometry Constant (RIC) respectively using squared ℓ_2-norm. [3]

ii State the equivalent definitions of RIP and RIC using the singular values of the relevant matrices. [3]

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

$$|\langle \boldsymbol{A}_{\mathcal{I}}\boldsymbol{a}, \boldsymbol{A}_{\mathcal{J}}\boldsymbol{b}\rangle| \leq c \left\|\boldsymbol{a}\right\|_2 \left\|\boldsymbol{b}\right\|_2, \tag{1.1}$$

for some constant c. Answer the following sub-questions in order to find out the value of the constant c (in terms of RIC).

i Define $\boldsymbol{a}' = \boldsymbol{a}/\left\|\boldsymbol{a}\right\|_2$ and $\boldsymbol{b}' = \boldsymbol{b}/\left\|\boldsymbol{b}\right\|_2$. Compute the squared ℓ_2-norm of the vectors $\begin{bmatrix} \boldsymbol{a}' \\ \boldsymbol{b}' \end{bmatrix}$ and $\begin{bmatrix} \boldsymbol{a}' \\ -\boldsymbol{b}' \end{bmatrix}$. [2]

ii Define $\boldsymbol{x}' = \boldsymbol{A}_{\mathcal{I}}\boldsymbol{a}'$ and $\boldsymbol{y}' = \boldsymbol{A}_{\mathcal{J}}\boldsymbol{b}'$. Find the lower and upper bounds of $\left\|\boldsymbol{x}' + \boldsymbol{y}'\right\|_2^2$ and $\left\|\boldsymbol{x}' - \boldsymbol{y}'\right\|_2^2$ using RIC of the matrix $\boldsymbol{A}$. [3]

iii Using the results of the previous sub-question to derive the constant c in (1.1) in terms of RIC. [3]

iv Based on (1.1), it holds that

$$\left\|\boldsymbol{A}_{\mathcal{I}}^T\boldsymbol{A}_{\mathcal{J}}\boldsymbol{b}\right\|_2 \leq c \left\|\boldsymbol{b}\right\|_2 \tag{1.2}$$

with the same constant c. Prove this result. [2]

(d) Let $\boldsymbol{x} \in \mathbb{R}^n$ be a k-sparse vector with support set $\mathrm{supp}(\boldsymbol{x}) = \mathcal{I}$. Let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.

Remark: For simplicity you can directly use the fact that $\delta_k \leq \delta_{2k} < 1$.

i Establish an upper bound on $\left\|\boldsymbol{A}_{\mathcal{J}}^T\boldsymbol{y}\right\|_2$ using RIC δ_{2k}. [2]

ii Establish a lower bound on $\left\|\boldsymbol{A}_{\mathcal{I}}^T\boldsymbol{y}\right\|_2$ using RIC δ_{2k}. [2]

iii Find the specific range of δ_{2k} so that $\left\|\boldsymbol{A}_{\mathcal{I}}^{T}\boldsymbol{y}\right\|_{2} \geq \left\|\boldsymbol{A}_{\mathcal{J}}^{T}\boldsymbol{y}\right\|_{2}$. [1]

iv Let $\boldsymbol{A}_{\mathcal{J}}^{\dagger}$ and $\boldsymbol{A}_{\mathcal{I}}^{\dagger}$ be the pseudo-inverse of the matrices $\boldsymbol{A}_{\mathcal{J}}$ and $\boldsymbol{A}_{\mathcal{I}}$, respectively. Establish an upper bound on $\left\|\boldsymbol{A}_{\mathcal{J}}^{\dagger}\boldsymbol{y}\right\|_{2}$ using RIC. Find the specific range of RIC δ_{2k} so that $\left\|\boldsymbol{A}_{\mathcal{I}}^{\dagger}\boldsymbol{y}\right\|_{2} \geq \left\|\boldsymbol{A}_{\mathcal{J}}^{\dagger}\boldsymbol{y}\right\|_{2}$. [2]

(Total marks: 25)

2. (Convex Optimisation Basics)

(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]

(b)

i Consider a differentiable function $f:\ \mathbb{R}^n \to \mathbb{R}$. Denote its gradient at $\boldsymbol{x} \in \mathbb{R}^n$ by $\nabla f(\boldsymbol{x})$. For given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, the directional gradient of f along the direction $\boldsymbol{y} - \boldsymbol{x}$ is defined as

$$\nabla_{\boldsymbol{y}-\boldsymbol{x}} f(\boldsymbol{x}) = \lim_{\lambda \to 0} \frac{f(\boldsymbol{x} + \lambda(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x})}{\lambda}.$$

State the dimension of $\nabla f(\boldsymbol{x})$ and $\nabla_{\boldsymbol{y}-\boldsymbol{x}} f(\boldsymbol{x})$. State how to compute the directional gradient $\nabla_{\boldsymbol{y}-\boldsymbol{x}} f(\boldsymbol{x})$ using the gradient $\nabla f(\boldsymbol{x})$. [2]

ii Assume that a convex function $f:\ \mathbb{R}^n \to \mathbb{R}$ is differentiable. Prove that for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, it holds that

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T (\boldsymbol{y} - \boldsymbol{x}). \tag{2.3}$$

[2]

Hint: Apply the definition of convex function to $f((1-\lambda)\boldsymbol{x} + \lambda \boldsymbol{y})$, $\lambda \in [0,1]$.

iii Assume that a convex function $f:\ \mathbb{R}^n \to \mathbb{R}$ is *not* differentiable at a point $\boldsymbol{x} \in \mathbb{R}^n$. State the definition of the subgradient at the point $\boldsymbol{x}$. Make the dimension of the subgradient explicit in your answer. [2]

iv The set of subgradients at $\boldsymbol{x}$ is called the subdifferential at $\boldsymbol{x}$ and is denoted by $\partial f(\boldsymbol{x})$. Find the subdifferential of $f(x) = |x|$ for all $x \in \mathbb{R}$. [3]

(c)

i Assume that a convex function $f:\ \mathbb{R}^n \to \mathbb{R}$ is differentiable. Prove that $\boldsymbol{x}$ is a global minimiser of f if $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$. [2]
Hint: You are allowed to use the result in (2.3) directly.

ii Assume that a convex function $f:\ \mathbb{R}^n \to \mathbb{R}$ is *not* differentiable at a point $\boldsymbol{x} \in \mathbb{R}^n$. Prove that $\boldsymbol{x}$ is a global minimiser of f if $\boldsymbol{0} \in \partial f(\boldsymbol{x})$. [2]

iii The soft thresholding function $\eta(\cdot)$ is designed to give a global minimiser of the simplified Lasso problem

$$\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{z}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 . \tag{2.4}$$

State the form of the soft thresholding function $\eta(\cdot)$. (Derivations are not required.) [2]

iv The famous Lasso formulation for sparse recovery is given by

$$\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 , \tag{2.5}$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a given matrix and $\lambda > 0$ is a given parameter. State the Iterative Shrinkage Thresholding (IST) algorithm to solve the Lasso problem (2.5). (Derivations are not required.) [3]

v Consider the low-rank matrix recovery problem $\boldsymbol{y} = \mathcal{A}(\boldsymbol{X})$, where $\mathcal{A} : \mathbb{R}^{n_r \times n_c} \to \mathbb{R}^m$ is a linear operator. 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. [3]

(Total marks: 25)

3. (Convex Optimisation)

(a)

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

ii Let u_i be the Lagrange multipliers of the inequality constraints and v_j be the Lagrange multipliers of the equality constraints, respectively. State the corresponding Lagrangian. [2]

iii State the corresponding Lagrange dual function and Lagrange dual problem. [2]

iv State the Karush-Kuhn-Tucker (KKT) conditions for a global optimum. [4]

(b) Alternating direction method of multipliers (ADMM) solves optimisation problems in the form

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{z}} f(\boldsymbol{x}) + g(\boldsymbol{z}) \\ &\text{subject to } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{c}. \end{aligned} \tag{3.6}$$

Consider the equivalent problem

$$\begin{aligned} &\min_{\boldsymbol{x},\boldsymbol{z}} f(\boldsymbol{x}) + g(\boldsymbol{z}) + \frac{\rho}{2}\left\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} - \boldsymbol{c}\right\|_2^2 \\ &\text{subject to } \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{c}, \end{aligned} \tag{3.7}$$

where $\rho > 0$ is a constant.

i State the corresponding Lagrangian of (3.7), denoted by $L_\rho(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{v})$. [1]

ii ADMM is an iterative algorithm with three steps in each iteration:

- Update $\boldsymbol{x}$ to obtain $\boldsymbol{x}^{k+1}$.
- Update $\boldsymbol{z}$ to obtain $\boldsymbol{z}^{k+1}$.
- Update $\boldsymbol{v}$ to obtain $\boldsymbol{v}^{k+1}$ via

$$\boldsymbol{v}^{k+1} = \boldsymbol{v}^k + \rho\left(\boldsymbol{A}\boldsymbol{x}^{k+1} + \boldsymbol{B}\boldsymbol{z}^{k+1} - \boldsymbol{c}\right).$$

State the details of the first two steps. [4]

iii The form of ADMM iterations can be highly simplified by introducing $\boldsymbol{w} = \frac{1}{\rho}\boldsymbol{v}$ (i.e. $\boldsymbol{v} = \rho\boldsymbol{w}$). Rewrite the details of the three steps in each iteration by replacing $\boldsymbol{v}$ with $\boldsymbol{w}$. [2]

In the literature, the simplified form is called the scaled form of ADMM.

(c) In the following, we are applying the scaled form of ADMM to solve two non-smooth convex optimisation problems.

i To solve the Lasso problem

$$\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 ,$$

state the standard ADMM problem formulation in the form of (3.6) and the scaled form of ADMM iterations. [3]

ii To solve the constrained Lasso problem

$$\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1$$
$$\text{subject to } \boldsymbol{B}\boldsymbol{x} \leq \boldsymbol{0},$$

state the standard ADMM problem formulation in the form of (3.6) and the scaled form of ADMM iterations. [5]

(Total marks: 25)