EGT2
ENGINEERING TRIPOS PART IIA

Monday 1 May 2023 14.00 to 15.40

Module 3M1

MATHEMATICAL METHODS

Answer all the questions.

All questions carry the same number of marks.

*The **approximate** percentage of marks allocated to each part of a question is indicated in the right margin.*

*Write your candidate number **not** your name on the cover sheet.*

STATIONERY REQUIREMENTS
Write on single-sided paper.

SPECIAL REQUIREMENTS TO BE SUPPLIED FOR THIS EXAM
CUED approved calculator allowed.
Engineering Data Books.

10 minutes reading time is allowed for this paper at the start of the exam.

You may not start to read the questions printed on the subsequent pages of this question paper until instructed to do so.

You may not remove any stationery from the Examination Room.

1 (a) For a function $f : \mathbb{R}^n \to \mathbb{R}$, consider its Taylor series expansion about $\boldsymbol{x} \in \mathbb{R}^n$,

$$f(\boldsymbol{x} + \boldsymbol{v}) = f(\boldsymbol{x}) + (\nabla f)^T \boldsymbol{v} + \frac{1}{2}\boldsymbol{v}^T \boldsymbol{H} \boldsymbol{v} + O\left(\|\boldsymbol{v}\|^3\right),$$

where the matrix $\boldsymbol{H} := \partial^2 f / \partial x_i \partial x_j$ is known as the *Hessian* of f.

(i) Show that the Hessian is symmetric. [5%]

(ii) At a stationary point of f we have $(\nabla f)^T \boldsymbol{v} = 0$ for all $\boldsymbol{v}$. Explain how the nature of a stationary point can be characterised in terms of properties of $\boldsymbol{H}$. [15%]

(iii) For a constant Hessian, consider $\|\boldsymbol{x}\|_{\boldsymbol{H}}^2 := \boldsymbol{x}^T \boldsymbol{H} \boldsymbol{x}$. Under what conditions is $\|\boldsymbol{x}\|_{\boldsymbol{H}}$ a norm? Explain why. [10%]

(b) A particular iterative scheme for finding approximate solutions to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ uses the update:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + w(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_k)$$

where $w > 0$.

(i) Find an expression for the residual $\boldsymbol{r}_k := \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_k$ in terms of the initial residual $\boldsymbol{r}_0$. [20%]

(ii) Find an expression for the error $\boldsymbol{e}_k := \boldsymbol{x}_k - \boldsymbol{x}$ in terms of $\boldsymbol{e}_0$. [20%]

(iii) For symmetric positive-definite $\boldsymbol{A}$, what range of values for w guarantees convergence, and what is the optimal value of w for convergence in the 2-norm? [30%]

2 For D-dimensional column vector $\mathbf{x} = [x_1, x_2, \cdots, x_D]^\mathsf{T} \in \mathbb{R}^D$ consider the product of logistic functions

$$s(\mathbf{x}) = \prod_{d=1}^{D} \frac{1}{1 + \exp(-x_d)}$$

In addition consider the Gaussian distribution over $\mathbf{x} \in \mathbb{R}^D$

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^D |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\mathsf{T} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

with mean vector $\boldsymbol{\mu} \in \mathbb{R}^D$ and $D \times D$ covariance matrix $\boldsymbol{\Sigma}$. Consider the probability

$$P(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\mathcal{Z}(\boldsymbol{\mu}, \boldsymbol{\Sigma})} s(\mathbf{x}) \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

where $\mathcal{Z}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int_{\mathbb{R}^D} s(\mathbf{x}) \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) d\mathbf{x}$

(a) Show that the point $\mathbf{x}^* \in \mathbb{R}^D$ yielding the maximum value of $P(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be obtained by minimising the function

$$f(\mathbf{x}) = \mathbf{1}^\mathsf{T} \mathbf{g}(\mathbf{x}) + \frac{1}{2} \mathbf{x}^\mathsf{T} \boldsymbol{\Sigma}^{-1} \mathbf{x} - \mathbf{x}^\mathsf{T} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$$

where the D-dimensional vector of ones is denoted as $\mathbf{1} = [1, 1, \cdots, 1]^\mathsf{T}$ and $\mathbf{g}(\mathbf{x})$ is an element-wise application of the function $g(\cdot)$ which acts on each component of the vector $\mathbf{x}$ such that $\mathbf{g}(\mathbf{x}) = [g(x_1), g(x_2), \cdots, g(x_D)]^\mathsf{T}$ with each $g(\cdot)$ defined appropriately. [20%]

(b) Derive and justify the necessary condition for a point $\mathbf{x}^* \in \mathbb{R}^D$ to be a strong local minimum of the function $f(\mathbf{x})$. [15%]

(c) Derive and justify the sufficient condition for a point $\mathbf{x}^* \in \mathbb{R}^D$ to be a strong local minimum of the function $f(\mathbf{x})$. [15%]

(d) Derive a Newton optimisation method for the function $f(\mathbf{x})$ acting on $\mathbf{x} \in \mathbb{R}^D$. [20%]

(e) For the specific case where $D = 2$, and $\boldsymbol{\Sigma}$ is a diagonal covariance matrix, assess whether $f(\mathbf{x})$ is convex in $\mathbb{R}^2$ and state the implication on the nature of the point $\mathbf{x}^*$. [30%]

3 The transitions in a finite state space Markov chain are governed by an $N \times N$ matrix, $\mathbf{P}$, with the following form

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{0} & \mathbf{P}_{22} \end{bmatrix}$$

where element i, j of the matrix $\mathbf{P}$ is the probability of transitioning to state j if the system is in state i. $\mathbf{P}_{11}$ is a square $m \times m$ matrix. The elements of $\mathbf{P}_{11}$, $\mathbf{P}_{12}$ and $\mathbf{P}_{22}$ are non-zero.

(a) What constraints must be satisfied by $\mathbf{P}$ for it to be a valid transition matrix for a finite state-space Markov chain? [10%]

(b) What equation must the *stationary distribution*, $\boldsymbol{\pi}$, satisfy? $\boldsymbol{\pi}$ is a row vector where

$$\boldsymbol{\pi} = \begin{bmatrix} \boldsymbol{\pi}_1 & \boldsymbol{\pi}_2 \end{bmatrix}$$

and $\boldsymbol{\pi}_1$ is an m-dimensional row vector. The simplest form of the expression should be given and clearly motivated. Does the stationary distribution depend on the intial state? [20%]

(c) If all elements of $\mathbf{P}_{12}$ are zero, how does this change the answer to Part (b)? [10%]

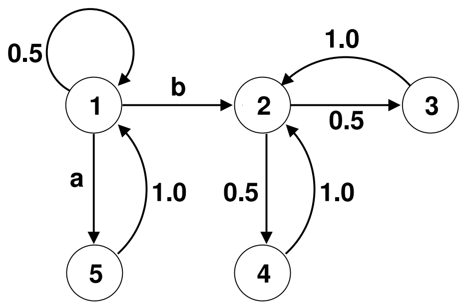

Fig. 1

(d) Figure 1 shows the state-diagram for a particular process.

(i) Write down the transition matrix for this process. [15%]

(ii) Find an expression for all stationary distributions for this process when $a = 0.5$. Does the stationary distribution depend on the initial state in the context of your answers to Parts (b) and (c)? [25%]

(iii) The network is now modified so that $b = 0.1$. Discuss the periodicity of this network. Give two eigenvalues of the transition matrix, justifying your answer. [20%]

END OF PAPER