A14538W1

DEGREE OF MASTER OF SCIENCE

Mathematical Modelling and Scientific Computing

Paper B2 Numerical Linear Algebra and Continuous Optimisation

---

TRINITY TERM 2021

Friday, 23 April 2021

Opening Time: 9:30am BST

Mode of Completion: Handwritten

You have 2 hours and 30 minutes of writing time to complete the paper and up to 30 minutes technical time to upload your answer file.

This exam paper contains two sections. You may attempt as many questions as you like but you must answer at least one question in each section. Your best answer in each section will count, along with your next best two answers, making a total of four answers.

Numerical Linear Algebra

1. Let $A \in \mathbb{R}^{n\times n}$ be symmetric positive definite, and let $\|\cdot\|_A$ be the A-norm $\|v\|_A = \sqrt{v^T Av}$.

   (a) [10 marks] Let $AQ_k = Q_kT_k + q_{k+1}[0, \dots, 0, t_{k+1,k}]$ be the Lanczos decomposition after k steps, where $Q_k = [q_1, \dots, q_k] \in \mathbb{R}^{n\times k}, Q_{k+1} = [q_1, \dots, q_{k+1}] \in \mathbb{R}^{n\times(k+1)}$ are orthonormal with $q_1 = b/\|b\|_2 \in \mathbb{R}^n$ and $T_k \in \mathbb{R}^{k\times k}$ is tridiagonal. The CG (conjugate gradient) algorithm finds an approximate solution $\hat{x}$ to the linear system $Ax = b$ in the Krylov subspace $\hat{x} \in \mathcal{K}_k(A, b) := \text{span}([b, Ab, A^2b, \dots, A^{k-1}b])$ by imposing that the residual $r = A\hat{x} - b$ is orthogonal to $\mathcal{K}_k(A, b)$.

   (i) Find an expression of $\hat{x}$ in terms of (a subset of) A, Q_k, Q_{k+1}, T_k and b.

   (ii) Prove that $\hat{x}$ minimises the A-norm of the error in $\mathcal{K}_k(A, b)$, that is, $\|x - \hat{x}\|_A \leqslant \|x - y\|_A$ for any vector $y \in \mathcal{K}_k(A, b)$.

   (b) [5 marks] Show that for any vector $v \in \mathbb{R}^n$, $\sqrt{\lambda_{\min}(A)}\|v\|_2 \leqslant \|v\|_A \leqslant \sqrt{\lambda_{\max}(A)}\|v\|_2$, where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalues of A, respectively.

   (c) [5 marks] Use (b) to show that $\|x - \hat{x}\|_2 \leqslant C\|x - y\|_2$ for any $y \in \mathcal{K}_k(A, b)$, that is, the CG algorithm minimises the the 2-norm of the error up to a constant C (which can depend on A but not on k). Determine the value of C.

   (d) [5 marks] Suppose that A has just five distinct positive eigenvalues. Show that the CG algorithm gives the exact solution of $Ax = b$ in five steps.

2. Let $A \in \mathbb{R}^{n\times n}$, and denote by $A = U\Sigma V^T$ its SVD, where U, V are orthogonal and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ with $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n \geqslant 0$. Let $\|A\|_2 = \sigma_1$ denote the spectral norm, and $\|A\|_F = \sqrt{\sum_{i,j} |A_{ij}|^2}$ the Frobenius norm.

   (a) [6 marks] Let $X, Y \in \mathbb{R}^{n\times k}$ where $k \leqslant n$, and assume that $Y^T X$ is nonsingular. Define $P = X(Y^T X)^{-1} Y^T$.

      (i) Show that P is a projection, i.e., $P^2 = P$.

      (ii) Show that $\|P\|_2 \geqslant 1$.

   (b) [9 marks] Let $Q \in \mathbb{R}^{n\times k}$ be orthonormal $Q^T Q = I_k$, and $[Q\ Q_\perp] \in \mathbb{R}^{n\times n}$ be orthogonal.

      (i) Show that $\|QQ^T A - A\|_2 \leqslant \|A\|_2$.

      (ii) State a necessary and sufficient condition (in terms of Q, U, Σ, V) for $\|QQ^T A - A\|_2 = \|A\|_2$ to hold.

      (iii) State a necessary and sufficient condition (in terms of Q, U, Σ, V) for $\|QQ^T A - A\|_F = \|A\|_F$ to hold.

   (c) [10 marks] Let $\mathrm{rank}(A) = r$, and let $A = U_r \Sigma_r V_r^T$ be its reduced SVD, where $U_r, V_r \in \mathbb{R}^{n\times r}$ and Σ is positive definite.

      (i) For an orthonormal $Q \in \mathbb{R}^{n\times k}$, state a necessary and sufficient condition in terms of Q, U_r, Σ_r, V_r for $QQ^T A = A$ to hold, and give a lower bound for k.

      (ii) Let $X \in \mathbb{R}^{n\times r}$ and $AX = QR$ be the thin QR factorisation. State a necessary and sufficient condition in terms of X, U_r, Σ_r, V_r such that $QQ^T A = A$.

      (iii) Describe a (randomised) algorithm that computes orthonormal matrices $Q, \widetilde{Q} \in \mathbb{R}^{n\times r}$ such that $QQ^T A\widetilde{Q}\widetilde{Q}^T = A$ with probability 1, and hence find a reduced SVD of A without using U_r, Σ_r, V_r.
      [*You may use the fact that a square Gaussian matrix G, with i.i.d entries $G_{ij} \sim N(0, 1)$, is nonsingular with probability 1.*]

Continuous Optimisation

3. Consider the following unconstrained problem,

$$\min_{x\in\mathbb{R}^n} f(x) = b^T x + \frac{1}{2} x^T A x, \tag{1}$$

where $b \in \mathbb{R}^n$ and $A = (A_{i,j})$, $i, j \in \{1, \ldots, n\}$ is an $n \times n$ real and diagonal matrix with positive diagonal entries $A_{1,1} = a_1$, $A_{2,2} = a_2$, …, $A_{n,n} = a_n$. Apply a quasi-Newton Generic Linesearch Method (GLM) to (1), starting from a given approximation B^0 to the Hessian A of f, and generating, on each iteration $k \geqslant 0$, a new approximation B^{k+1} to A by updating B^k using x^k, x^{k+1} and the respective gradients of f at these iterates.

(a) [5 marks] At some iteration $k \geqslant 0$, assume that we require B^{k+1} to be a diagonal matrix and that $x_i^{k+1} \neq x_i^k$ for all $i \in \{1, \ldots, n\}$. Is the secant equation sufficient to determine B^{k+1} in this case? Justify your answer.

(b) [7 marks] Without requiring that B^{k+1} be diagonal at some iteration k, we require that $x_i^{k+1} \neq x_i^k$ for all $i \in \{1, \ldots, n\}$, and that $B^k = (B^k_{i,j})$, $i, j \in \{1, \ldots, n\}$, is a diagonal matrix with diagonal entries $B^k_{1,1} = a_1$, $B^k_{2,2} = a_2$, …, $B^k_{n-1,n-1} = a_{n-1}$ and $B_{n,n} = \beta \neq a_n$. Assume also that B^{k+1} is computed from B^k by the Symmetric Rank-One (SR1) formula. Calculate B^{k+1}.

(c) [6 marks] State a theorem of global convergence for the quasi-Newton GLM applied to (1), when backtracking-Armijo linesearch is employed on each iteration, justifying the assumptions that are needed.

(d) [7 marks] In the conditions of the theorem you state in (c), prove that the stepsize α^k generated by the backtracking-Armijo linesearch is bounded away from zero by a constant that is independent of k. *[Hint: Relevant results from the lectures can be used without proof but must be stated and applied carefully.]*

4. Consider the trust-region subproblem

$$\min_{s\in\mathbb{R}^n} m(s) = g^T s + \frac{1}{2} s^T H s \quad \text{subject to} \quad \|s\| \leqslant \Delta \tag{2}$$

where $n \geqslant 1$, $g \in \mathbb{R}^n$ and H is an $n \times n$ real symmetric matrix, where $\|\cdot\|$ denotes the Euclidean vector norm and $\Delta > 0$.

(a) [3 marks] State (without proof) the necessary and sufficient optimality conditions that hold at a global minimizer s^* of (2).

(b) [5 marks] Calculate the first-order (namely, KKT) and the second-order necessary optimality conditions that hold at a local minimizer of problem (2).

(c) [5 marks] Compare the local optimality conditions in (b) with the characterization of the global minimizer in (a). Are there KKT points that are not global minimizers of problem (2)?

(d) [12 marks] In (2), suppose H is the diagonal matrix

$$H = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 2 & 0 & \dots & 0 & 0 \\ \dots & & & & & \\ 0 & 0 & 0 & \dots & n-1 & 0 \\ 0 & 0 & 0 & \dots & 0 & \pm n \end{pmatrix} \quad \text{and} \quad g = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \alpha \end{pmatrix}, \tag{3}$$

where $\pm$ denotes the sign of the two cases that need to be addressed. Using the characterization of global minimizers in part (a) or otherwise, find the global minimizer of (2) when $\alpha \neq 0$.

Briefly describe when and why the case $\alpha = 0$ may be difficult.

5. (a) [13 marks] Consider the optimization problem below that has only one equality constraint,

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad c(x) = 0, \tag{4}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}$ are continuously differentiable functions.

(i) Write down the quadratic penalty function $\Phi_\sigma(x)$ associated with (4).

(ii) State the theorem of global convergence for the quadratic penalty method in the special case of problem (4).

(iii) Assuming that the stated conditions hold, prove that the Lagrange multiplier estimate $y^k \in \mathbb{R}$ generated at iteration k of the penalty method when applied to (4) converges to the optimal Lagrange multiplier y^* of the constraint as $k \to \infty$.

(b) [12 marks] Consider the constrained optimization problem,

$$\min_{x\in\mathbb{R}^n} x_1^2 + x_2^2 + \ldots + x_n^2 \quad \text{subject to} \quad a_1x_1 + a_2x_2 + \ldots + a_nx_n = 1, \tag{5}$$

where for all $i \in \{1, \ldots, n\}$, $a_i \neq 0$ are given constants.

(i) Calculate the (unconstrained) global minimizer(s) $x(\sigma)$ of the quadratic penalty function associated to (5), denoted by $\Phi_\sigma(x)$, for any $\sigma > 0$.

(ii) Show that $x(\sigma)$ converges to the solution x^* of problem (5), as $\sigma \to 0$.

(iii) Let $\nabla^2_{xx}\Phi_\sigma(x(\sigma))$ be the Hessian matrix of Φ_σ evaluated at $x(\sigma)$. Show that the condition number of $\nabla^2_{xx}\Phi_\sigma(x(\sigma))$ grows unboundedly as $\sigma \to 0$.

6. (a) [15 marks] Consider the inequality-constrained optimization problem

$$\min_{x\in\mathbb{R}^2} -x_1 - 2x_2 \quad \text{subject to} \quad x_1 + x_2 \leqslant 1, \quad x_1 \geqslant 0, \quad x_2 \geqslant 0. \tag{6}$$

(i) Write down the logarithmic barrier function $f_\mu(x)$ associated with (6), where $\mu > 0$ and x is any strictly feasible point of (6). Show that f_μ is a convex function in the domain where it is well-defined.

(ii) Does the central path of global minimizers $x(\mu)$ of f_μ exist for all values of μ? Briefly relate your findings to the theorem of local existence of central path in the lectures.

(b) [10 marks] Consider the inequality-constrained optimization problem,

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad c(x) \geqslant 0, \tag{7}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable functions.

Describe the steps of the basic barrier (also called interior point) algorithm applied to (7). Briefly describe two difficulties that the barrier method encounters and a way to overcome them.