A14538W1

DEGREE OF MASTER OF SCIENCE

Mathematical Modelling and Scientific Computing

Numerical Linear Algebra and Continuous Optimisation

Trinity TERM 2022

Monday, 25 April 2022

Opening Time: 9:30am UK Time

Mode of completion (format in which you will complete this exam): 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.

To upload your work you must submit just one pdf.

Numerical Liner Algebra

1. (a) [12 marks] Let $A \in \mathbb{R}^{m\times n}$ ($m \geqslant n$), and consider the process of computing a bidiagonal matrix $B = UAV^T$ where U, V are orthogonal matrices.
   (i) Explain how to find U, B, and V using Householder reflectors.
   (ii) Establish an upper bound on the operation count of the algorithm in (i) of the form Cm^pn^q, where p, q are (the smallest) integers to be specified, and C is a constant that need not be specified.
   (iii) Show that the singular values of A and B are the same.
   (iv) Assuming $m = n$, show that the eigenvalues of A and B are not necessarily the same.

   (b) [13 marks] Recall that a square matrix P is called a *projector* if $P^2 = P$. Let $M^\dagger$ denote the pseudoinverse of a rectangular matrix M.
   (i) Find $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^\dagger$ and $\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}^\dagger$.
   (ii) Let $P = X(Y^TX)^\dagger Y^T$, where X, Y are $n \times r$, $n > r$. Show that P is a projector.
   (iii) Assuming that Y^TX is nonsingular, show that P is invariant under right-multiplication of X, Y by nonsingular matrices, that is, if $\tilde{X} = XM$ and $\tilde{Y} = YN$ for nonsingular M, N, then $\tilde{X}(\tilde{Y}^T\tilde{X})^\dagger\tilde{Y}^T = P$.
   (iv) Let $M \in \mathbb{R}^{m\times n}$ with $m > n$. Under what condition is $M^\dagger M$ equal to I_n? When is $MM^\dagger$ equal to I_m?

2. Consider the QR algorithm for solving eigenvalue problems $Ax = \lambda x$, where $A \in \mathbb{R}^{n\times n}$. Let A_k denote the iterates generated by the QR algorithm, with $A_0 = A$.

   (a) [5 marks] Briefly describe how the iterates A_k are computed by the QR algorithm without shifts, and with shifts (you need not specify how the shifts are chosen). Prove that in both cases, A_k have the same eigenvalues as A.

   (b) [5 marks] Suppose A is symmetric. Assuming that the iterates A_k in the QR algorithm converge to a diagonal matrix $A_k \to D$ as $k \to \infty$, explain how to find the eigenvalues and eigenvectors of A.

   (c) [5 marks] Let $A \in \mathbb{R}^{n\times n}$ be a matrix of all 1's, $A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \ddots & \cdots & 1 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$. Find the eigenvalues of A by executing one step of the (unshifted) QR algorithm.

   (d) [5 marks] Explain why one step of the QR algorithm provides the eigenvalues of A in (c).

   (e) [5 marks] For a general square matrix $A \in \mathbb{R}^{n\times n}$ and prescribed scalars $s_1, s_2 \in \mathbb{R}$, show how to find a vector parallel to the last (nth) column of $(A-s_1I)^{-1}(A-s_2I)^{-1}$ using two steps of the shifted QR algorithm (and in particular without explicitly inverting a matrix or solving linear systems).

Continuous Optimisation

3. Consider the following unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and bounded below, and its gradient $\nabla f(\cdot)$ is Lipschitz continuous on $\mathbb{R}^n$. Apply the steepest descent method with linesearch to (1).

(a) [13 marks] Assume that the backtracking-Armijo linesearch is used with the steepest descent method.

(i) Show that the stepsize α^k generated by this linesearch is bounded away from zero by a constant that is independent of k, namely, show that there exists a constant $C_1 > 0$ such that

$$\alpha^k \geqslant C_1 \quad \text{for all} \quad k \geqslant 0. \tag{2}$$

[*Hint: Relevant results from the lectures can be used without proof but must be specified and applied carefully.*]

(ii) Using (2), or otherwise, show that there exists a constant $C_2 > 0$ such that

$$f(x^k) - f(x^{k+1}) \geqslant C_2 \|\nabla f(x^k)\|^2 \quad \text{for all} \quad k \geqslant 0, \tag{3}$$

where $\|\cdot\|$ denotes the Euclidean norm.

(iii) Let $\epsilon > 0$. Using (3) or otherwise, show that there exists an iteration $k \geqslant 0$ such that the algorithm terminates finitely, namely, with $\|\nabla f(x^k)\| \leqslant \epsilon$. Provide an upper bound on the number of iterations k that the algorithm takes until termination.

(b) [7 marks] Assume that a constant stepsize is used with the steepest descent method, namely, $\alpha^k = \alpha$ for all $k \geqslant 0$.

(i) Find the largest value of α such that a property of the form (3), for some $C_2 > 0$, holds for this steepest descent variant.

(ii) What conclusions can you draw regarding the finite termination and iteration upper bound for this variant? Briefly justify your answer.

(c) [5 marks] Let $f(x) = \frac{1}{2}(a_1 x_1^2 + a_2 x_2^2 + a_3 x_3^2)$, where $a_1 > a_2 > a_3 > 0$ and $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$. Calculate the constant stepsize α in (b)(i) for this objective function f.
Briefly comment on the difficulty of calculating this stepsize in the case of a general convex quadratic objective.
[*Hint: the required stepsize α depends on the reciprocal of the Lipschitz constant of the gradient.*]

4. (a) [13 marks] Consider the least-squares problem

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2}\|r(x)\|^2, \tag{1}$$

where $r : \mathbb{R}^n \to \mathbb{R}^m$ with $m \geqslant n$ and $\|\cdot\|$ denotes the Euclidean norm. Consider minimizing the following local quadratic approximation of $f(x+s)$ in (1) for some $x \in \mathbb{R}^n$, within a trust region of radius $\Delta > 0$, namely,

$$\min_{s \in \mathbb{R}^n} q(s) = \frac{1}{2}\|J(x)s + r(x)\|^2 \quad \text{subject to} \quad \|s\| \leqslant \Delta, \tag{2}$$

where $J(x)$ denotes the $m \times n$ Jacobian matrix of $r(x)$.

(i) Write down (necessary and sufficient) optimality conditions for a global minimizer s^* of (2). Are there Karush-Kuhn-Tucker (KKT) points that are not global minimizers of problem (2)?

(ii) Find an expression for s^* in (a)(i) in the case when the trust-region constraint is inactive at s^*.

(iii) Find sufficient conditions on problems (1) and (2) such that s^* is a descent direction for f from x.

(b) [12 marks] Consider the following function in one variable $x \in \mathbb{R}$,

$$f(x) = -x^6 + 48x^2. \tag{3}$$

(i) Calculate the stationary points of this problem and establish whether they are (local) minimizers or maximizers. Estimate the local rate(s) of convergence of Newton's method for optimization (without linesearch) applied to f, when the starting point is close to each of the stationary points of f that you found.

(ii) Are there starting points x^0 for Newton's method for optimization (without linesearch) applied to f in (3) such that the ensuing iterates fail to converge to a stationary point of f? Justify your answer.

5. (a) [13 marks] Consider the constrained optimization problem

$$\min_{x\in\mathbb{R}^n} x_1+x_2+\cdots+x_n \text{ subject to } x_1\cdot x_2\cdot\ldots\cdot x_n = 1 \text{ and } x_i \geqslant \alpha_i,\ i\in\{1,\ldots,n\}, \quad (1)$$

where $x = (x_1, x_2, \ldots, x_n)^T$, and the given constants satisfy $\alpha_i > 0$, $i \in \{1, \ldots, n\}$ and $\alpha_1 \cdot \alpha_2 \cdot \ldots \cdot \alpha_n < 1$.

(i) Show that problem (1) has a unique Karush-Kuhn-Tucker (KKT) point.

(ii) Write down the second-order optimality conditions at this KKT point. Is this point a local minimizer?

(b) [12 marks] Consider the unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \quad (2)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

(i) Assume that there exists a stationary point $\overline{x} \in \mathbb{R}^n$ of f such that the Hessian matrix $\nabla^2 f(\overline{x})$ is indefinite (that is, it has both positive and negative eigenvalues). Find a direction $s \in \mathbb{R}^n$, $s \neq 0$, such that f decreases along s from its value $f(\overline{x})$, namely, $f(\overline{x} + \alpha s) < f(\overline{x})$ for all $\alpha > 0$ sufficiently small.

(ii) Assume that there exists a stationary point $\hat{x} \in \mathbb{R}^n$ of f such that the Hessian matrix $\nabla^2 f(\hat{x})$ is positive semidefinite. Discuss whether and when $\hat{x}$ is a local minimizer, maximizer or saddle point, justifying your answer carefully.

6. (a) [14 marks] Consider the optimization problem below that has only one equality constraint,

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad c(x) = 0, \tag{1}$$

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

(i) State the theorem of global convergence for the augmented Lagrangian method in the special case of problem (1).

(ii) Assuming that the conditions of the theorem in (a)(i) hold, prove that the Lagrange multiplier estimate $y^k \in \mathbb{R}$ generated at iteration k of the augmented Lagrangian method when applied to (1) converges to the optimal Lagrange multiplier y^* of the constraint, as $k \to \infty$.

(iii) Assume now that the quadratic penalty term in the augmented Lagrangian function $\Phi(x, u, \sigma)$ is replaced by a penalty term of the form $|c(x)|^{2p}/(2p\sigma)$ for integer $p > 1$. Does this generate a well-defined penalty function and does the convergence theorem in (a)(i) holds in this case? Justify your answer, outlining potential similarities and differences.

(b) [11 marks] Consider the constrained optimization problem,

$$\min_{x\in\mathbb{R}^n} x_1^2 + x_2^2 + \ldots + x_n^2 \quad \text{subject to} \quad c(x) := x_1 + 2x_2 + \ldots + nx_n - 1 = 0, \tag{2}$$

where $n \geqslant 2$.

(i) Write down the augmented Lagrangian function $\Phi(x, u, \sigma)$ associated with problem (2) and calculate its (unconstrained) global minimizer(s) $x(u, \sigma)$, for any $u \in \mathbb{R}$ and $\sigma > 0$.

(ii) Let σ be fixed. Let u be updated by the formula

$$u^{k+1} = u^k - \frac{c(x(u^k, \sigma))}{\sigma}, \quad k \geqslant 0,$$

starting from some u^0, and where $x(u^k, \sigma)$ is, like above, the minimizer of $\Phi(x, u^k, \sigma)$. Show that, as $k \to \infty$, $\{u^k\}$ converges to the Lagrange multiplier y^* of the constraint in (2) and $\{x(u^k, \sigma)\}$ to the global minimizer x^* of (2).

(iii) Let $u \in \mathbb{R}$ be fixed. Briefly describe a difficulty that the augmented Lagrangian method encounters when $\sigma \to 0$.