DEGREE OF MASTER OF SCIENCE

MATHEMATICAL MODELLING AND SCIENTIFIC COMPUTING

B2 Further Numerical Linear Algebra and Continuous Optimisation

---

TRINITY TERM 2020

FRIDAY, 15 May 2020
Opening Time: 9.30am (BST)
You have 3 hours 30 minutes to complete the paper and upload your answers.

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.

Please start the answer to each question on a fresh page.
All questions will carry equal marks.

Do not turn this page until you are told that you may do so

Section A: Further Numerical Linear Algebra

1. Let Π_k be the set of real polynomials of degree at most k.

(a) [8 marks] What is the Krylov subspace $\mathcal{K}_k(B, r)$ generated by a matrix $B \in \mathbb{R}^{n\times n}$ and a vector $r \in \mathbb{R}^n$? If B is nonsingular, $r_k = b - Bx_k, k = 0, 1, \ldots$ and $x_k - x_0 \in \mathcal{K}_k(B, r_0)$, show that

$$e_k = p_k(B)e_0$$

where $e_k = B^{-1}b - x_k$ and $p_k \in \Pi_k$ satisfies $p_k(0) = 1$. What is a Krylov subspace method for the solution of the linear system of equations $Bx = b$?

(b) [9 marks] A Krylov subspace method is used which has the defining property that $\|e_k\|^2_{B^TB} := e_k^TB^TBe_k$ is minimal for each k. What is the resulting method called? Prove that, for this method, we must have

$$\|e_k\|_{B^TB} \leqslant \|X\|\|X^{-1}\| \min_{p\in\Pi_k, p(0)=1} \max_j |p(\lambda_j)| \; \|e_0\|_{B^TB}$$

where $B = X\Lambda X^{-1}$ is a diagonalisation of B with Λ being the diagonal matrix of the eigenvalues $\{\lambda_j : j = 1, 2, \ldots, n\}$ of B and $X \in \mathbb{R}^{n\times n}$ is nonsingular. You should prove all results that you need.

(c) [8 marks] If $B = B^T$, prove that

$$\|e_k\|_{B^2} \leqslant \min_{p\in\Pi_k, p(0)=1} \max_j |p(\lambda_j)| \; \|e_0\|_{B^2}.$$

What is the Krylov subspace method called that would be employed in this situation? How is it usually implemented? In the case that $n = 2m$ is even, would you expect faster convergence with this method if the eigenvalues are

$$-m, -m + 1, \ldots, -1, 1, 2, \ldots, m$$

or

$$-1, 1, 2, \ldots, 2m - 1?$$

2. Throughout this question, $A \in \mathbb{R}^{n\times n}$ is a symmetric, positive definite matrix and $b \in \mathbb{R}^n$.

(a) [6 marks] For $x \in \mathbb{R}^n$, define

$$\Phi(x) := \frac{1}{2}x^T Ax - x^T b.$$

Given $x_k, p_k \in \mathbb{R}^n$ with $p_k \neq 0$, define the residual r_k and show that

$$\alpha = \frac{p_k^T r_k}{p_k^T A p_k}$$

minimizes $\Phi(x_k + \alpha p_k)$.

(b) [11 marks] The *conjugate gradient method* is given by
choose x_0 , $r_0 = b - Ax_0 = p_0$ and for $k = 0, 1, 2, \ldots$

$$\begin{aligned}
\alpha_k &= p_k^T r_k / p_k^T A p_k \\
\text{if } \alpha_k &= 0 \text{ stop, else} \\
x_{k+1} &= x_k + \alpha_k p_k \\
r_{k+1} &= b - Ax_{k+1} \\
\beta_k &= -p_k^T A r_{k+1} / p_k^T A p_k \\
p_{k+1} &= r_{k+1} + \beta_k p_k
\end{aligned}$$

It is known that, for each $\ell = 0, 1, \ldots, k$,

$$\text{span}\{p_0, p_1, \ldots, p_\ell\} = \text{span}\{r_0, r_1, \ldots, r_\ell\}.$$

Assuming

$$r_k^T r_j = 0 \quad \text{and} \quad p_k^T A p_j = 0 \quad \text{for all } j < k,$$

show that

$$r_{k+1}^T r_j = 0 \quad \text{and} \quad p_{k+1}^T A p_j = 0 \quad \text{for all } j < k+1.$$

(c) [8 marks] From a starting guess $x_0 = 0$, apply two iterations of the conjugate gradient method to the problem $Ax = b$, where

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

to obtain x_2. State all relevant variables. Calculate the residual for x_2 and comment on your answer.

Section B: Continuous Optimization

3. Consider the unconstrained optimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable with gradient ∇f. We apply a Generic Linesearch Method (GLM) with Armijo linesearch to (1), starting from a point $x^0 \in \mathbb{R}^n$.

(a) [2 marks] State the Armijo linesearch condition that the stepsize α^k is required to satisfy.

(b) [12 marks] For each $k \geqslant 0$, let x^k be the iterate and s^k the search direction generated by the GLM. Assuming that the stepsize α^k satisfies

$$1 \geqslant \alpha^k \geqslant C\frac{-\nabla f(x^k)^T s^k}{\|s^k\|^2},$$

for some constant $C \in (0,1)$ independent of k, show that either there exists an iteration $l \geqslant 0$ such that $\nabla f(x^l) = 0$ or

$$\lim_{k\to\infty} \|\nabla f(x^k)\| \cos\theta^k = 0, \tag{2}$$

where θ^k is the angle between $-\nabla f(x^k)$ and s^k; state all the assumptions on f that are needed.

Derive a suitable assumption on the GLM such that it is a globally convergent method, namely, $\lim_{k\to\infty} \|\nabla f(x^k)\| = 0$.

(c) [11 marks] In (1), let

$$f(x) = \frac{1}{2}\left(ax_1^2 + x_2^2 + x_3^2\right), \tag{3}$$

where a is a positive constant.

Assume that the starting point x^0 in GLM applied to (3) has $x_i^0 \neq 0$ for all $i \in \{1,2,3\}$. On each iteration $k \geqslant 0$ of the GLM, let s^k be chosen as

$$s^k = -\begin{pmatrix} ax_1^k & x_2^k & x_3^k \end{pmatrix}^T. \tag{4}$$

Show that s^k in (4) is a suitable choice of direction in the GLM, and that it yields a globally convergent method.

Find another set of suitable directions $\{s^k\}$, $k \geqslant 0$, that make the GLM globally convergent.

4. Consider the trust-region subproblem

$$\min_{s \in \mathbb{R}^n} m(s) = f + s^T g + \frac{1}{2} s^T H s \quad \text{subject to} \quad \|s\| \leqslant \Delta \tag{5}$$

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

(a) [9 marks] State (without proof) the necessary and sufficient optimality conditions that hold at a global minimizer s^* of (5).

Describe how to use these conditions to calculate s^*, identifying the two cases that may occur.

(b) [10 marks] In (5), let $n = 3$, $f = 0$, and

$$H = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \quad \text{and} \quad g = \begin{pmatrix} a \\ 0 \\ d \end{pmatrix}. \tag{6}$$

Using the characterization of global minimizers in part (a) or otherwise, find the global minimizer of (5) when

$$a > 0, \; b > 0, \; c \neq 0 \quad \text{and} \quad d \neq 0.$$

Briefly describe why the case $c < 0$ and $d = 0$ may be difficult.

(c) [6 marks] Consider problem (5) for general $n \geqslant 1$, and assume that the optimality conditions you stated in (a) hold at a point s^*. Prove that s^* is an unconstrained global minimizer of a quadratic model

$$M(s) = f + s^T g + \frac{1}{2} s^T B s,$$

for a certain positive semidefinite matrix B that you must identify.

Using the above, or otherwise, prove that s^* is a global minimizer of the trust-region subproblem (5) (this is the sufficiency part of the conditions you stated in (a)).

5. (a) [9 marks] Consider the following function

$$f(x) = ax_1^6 + bx_2^6 + cx_3^6, \tag{7}$$

where $x = (x_1, x_2, x_3)^T \in \mathbb{R}^3$ and a, b and c are positive constants. Apply Newton's method (without linesearch or trust-region) to minimizing f, starting at $x^0 \in \mathbb{R}^3$, where $x_i^0 \neq 0$ for all $i \in \{1, 2, 3\}$.
Calculate the Newton iterates, and establish their convergence and rate of convergence. Explain why this convergence rate is not quadratic.

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

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad c(x) = 0, \tag{8}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ with $c(x) = (c_1(x), \ldots, c_m(x))^T$ are continuously differentiable functions, and $m \leqslant n$.
Assuming a suitable constraint qualification holds (that you do not need to define), show that any local minimizer of (8) is a KKT point of (8).

(c) [5 marks] Let A be a $p \times n$ real matrix with rows a_i^T, $i \in \{1, \ldots, p\}$, and $b = (b_1, \ldots, b_p)^T \in \mathbb{R}^p$, where $p \leqslant n$. For the following problem

$$\max_{x \in \mathbb{R}^n} \prod_{i=1}^{p} (a_i^T x - b_i) \quad \text{subject to} \quad Ax - b > 0, \tag{9}$$

show that any local maximizer is also a global maximizer.

6. (a) [10 marks] Consider the minimization problem

$$\min_{x \in \mathbb{R}^2} x_1^2 + x_2^2 - 4x_1 - 6x_2 \quad \text{subject to} \quad x_2 - x_1^2 \geqslant 0, \tag{10}$$

where $x = (x_1\ x_2)^T$. Using direct calculations or otherwise, show that (10) has a unique global minimizer.

(b) Consider the equality-constrained optimization problem,

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad c(x) = 0, \tag{11}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ with $c(x) = (c_1(x), \ldots, c_m(x))^T$ are continuously differentiable functions, and $m \leqslant n$.

(i) [3 marks] Write down the quadratic penalty function $\Phi_\sigma(x)$ associated with (11) and describe the connection between the stationary points of $\Phi_\sigma(x)$ and a KKT solution of problem (11).

(ii) [5 marks] State the theorem of global convergence for the quadratic penalty method. Prove that the approximate minimizer x^k of the quadratic penalty function $\Phi_{\sigma^k}(x)$ that is calculated on each iteration k of the method is asymptotically feasible for problem (11) as $k \to \infty$.
Hint: in your proof, you may assume that the Lagrange multiplier estimate y^k generated by the penalty method converges to the optimal Lagrange multiplier of the constraints as $k \to \infty$.

(iii) [5 marks] Give conditions on (11) such that the Lagrange multiplier estimate y^k grows unboundedly with k. Assuming the iterates x^k of the penalty method converge to a point x^*, establish the nature of x^* in this case.

(iv) [2 marks] Briefly describe a difficulty that quadratic penalty method may encounter and a way to overcome it.