DEGREE OF MASTER OF SCIENCE

MATHEMATICAL MODELLING AND SCIENTIFIC COMPUTING

B2 Further Numerical Linear Algebra and Continuous Optimization

---

HILARY TERM 2017

FRIDAY, 21 April 2017, 9.30am to 11.30am

Candidates should submit answers to a maximum of four questions for credit that include an answer to at least one question in each section.

Please start the answer to each question in a new answer booklet.
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. Suppose that one is given an invertible matrix $A \in \mathbb{R}^{n\times n}$, as well as vectors $r \in \mathbb{R}^n$, $b \in \mathbb{R}^n$, and a starting guess $x_0 \in \mathbb{R}^n$.

   (a) [4 marks] What is the *Krylov subspace* $\mathcal{K}_k(A, r)$? What is a *Krylov subspace method* for the solution of the linear system, $Ax = b$? Show that any such method computes iterates $x_k, k = 1, 2, \ldots$, such that the residuals $r_k = b - Ax_k, k = 0, 1, 2, \ldots$, satisfy

   $$r_k = p(A)r_0, \qquad (1)$$

   where p is a polynomial. Exactly what conditions does p satisfy?

   (b) [5 marks] State Arnoldi's method and say what it achieves.

   (c) [11 marks] If Arnoldi's method is written in the form

   $$AV_k = V_{k+1}\widehat{H}_k,$$

   describe the exact form of the matrices V_k and $\widehat{H}_k$. Hence show how the residuals and iterates of the GMRES method, which minimizes the Euclidean norm of the residual r_k for each k, can be computed via the solution of a linear least squares problem involving $\widehat{H}_k$. If $y \in \mathbb{R}^k$ is the solution of this linear least squares problem, show that

   $$y = V_k^T q(A)r_0$$

   where q is a polynomial that you should express in terms of the polynomial p in (1).

   (d) [5 marks] Calculate the GMRES iterates for the system $Ax = b$ with

   $$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } x_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

2. Let $A \in \mathbb{R}^{n\times n}$ be symmetric and positive definite.

The *Conjugate Gradient method* applied to find the solution, x, of the linear system $Ax = b$ is: choose x_0 , $r_0 = b - Ax_0 = p_0$ and for $k = 0, 1, 2, \ldots$

$$\begin{aligned} \alpha_k &= p_k^T r_k / p_k^T A p_k \\ x_{k+1} &= x_k + \alpha_k p_k \\ r_{k+1} &= b - Ax_{k+1} \qquad (2) \\ \beta_k &= -p_k^T A r_{k+1} / p_k^T A p_k \\ p_{k+1} &= r_{k+1} + \beta_k p_k. \end{aligned}$$

(a) [9 marks] Show that $r_{k+1} = r_k - \alpha_k A p_k$ is an alternative definition of r_{k+1} to (2). What is the advantage of replacing (2) with this formula? Prove that $r_{k+1}^T p_k = 0$ and that $p_{k+1}^T A p_k = 0$. Prove also that $r_{k+1}^T r_k = 0$.

(b) [10 marks] In what sense is the error vector, $x - x_k$, minimized for the conjugate gradient method? Show that

$$\|x - x_k\|_A \leqslant \min_{p\in\Pi_k, p(0)=1} \max_j |p(\lambda_j)| \, \|x - x_0\|_A,$$

where $\lambda_j, j = 1, 2, \ldots, n$, are the eigenvalues of A.

(c) [6 marks] Let $A \in \mathbb{R}^{n\times n}$ be such that all but one of its eigenvalues is contained in $[1, 2]$ and with a single outlying eigenvalue at 10^5. Let $B \in \mathbb{R}^{(n-1)\times(n-1)}$ be another symmetric and positive definite matrix that has all of its eigenvalues contained in $[1, 2]$ being exactly the same as the eigenvalues of A in $[1, 2]$. If the conjugate gradient method with zero starting vector is applied for $Ax = b$ and $By = c$, explain why convergence to some prescribed accuracy can generally only be delayed by at most one iteration for the system involving A compared to the system involving B.

If the outlying eigenvalue were instead at 10^{-5}, would the same statement remain true?

Section B: Continuous Optimization

3. (a) Consider the function

$$p(x) = x_1^n x_2 - 2x_2^2 + x_2^4, \tag{1}$$

where $x = (x_1\ x_2)^T$, n is an integer and $n \geqslant 1$.

(i) [3 marks] Find all stationary points of the function $p(x)$ for $x \in \mathbb{R}^2$.

(ii) [5 marks] Using only second-order optimality conditions, investigate the nature of each of the stationary points, namely, establish, whenever possible, whether they are local minimizers, maximizers or saddle points.

(iii) [5 marks] For $n \geqslant 2$, show that the origin $x^* = (0\ 0)^T$ is a saddle point of $p(x)$.

(b) Consider the constrained optimization problem

$$\min_{x \in \mathbb{R}^3} x_1^2 - x_2^2 - x_3^2 \quad \text{subject to} \quad x_1^4 + x_2^4 + x_3^4 \leqslant 1. \tag{2}$$

(i) [2 marks] Is problem (2) convex? Is any minimizer a KKT point?

(ii) [7 marks] By establishing the nature of the KKT points of problem (2), or otherwise, calculate the global solution(s) of problem (2).

(c) [3 marks] Consider the constrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad c(x) \geqslant 0, \tag{3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^p$ with $c(x) = (c_1(x), c_2(x), \ldots, c_p(x))$ are twice continuously differentiable functions. Assuming a suitable constraint qualification holds (which you do not need to define), briefly state (without proof) the second-order necessary optimality conditions at a local minimizer of (3).

4. Consider

$$\min_{x \in \mathbb{R}^n} f(x), \tag{4}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, and let $\nabla f(\cdot)$ denote the gradient of f. Apply a generic trust-region method to (4), where at the kth iterate x^k, the step s^k is calculated by solving (exactly) the following trust-region subproblem

$$\min_{s \in \mathbb{R}^n} m_k(s) := f(x^k) + s^T \nabla f(x^k) \quad \text{subject to} \quad \|s\| \leqslant \Delta_k, \tag{5}$$

where $\Delta_k > 0$ is the trust-region radius.

(a) [8 marks] Assuming that $\nabla f(x^k) \neq 0$, calculate an expression for the step s_k as defined by (5), depending on $\nabla f(x^k)$ and Δ_k. Using this expression for s_k, or otherwise, calculate a lower bound on the model decrease $f(x^k) - m_k(s^k)$.

(b) [3 marks] Define the Cauchy point s_C^k for (5). Relate s_C^k to the step s^k calculated in part (a).

(c) [10 marks] State a global convergence theorem for the generic trust-region method with subproblem (5).

Prove the theorem you state; you may assume that $\nabla f(x^k) \neq 0$ for all k, that there are infinitely many successful iterations and that (under the assumptions you have stated in your theorem) there exists a constant $\kappa_d > 0$ independent of k such that

$$\Delta_k \geqslant \kappa_d \inf_{k \geqslant 0} \|\nabla f(x^k)\| \quad \text{for all} \quad k \geqslant 0.$$

(d) [4 marks] Briefly describe two benefits and two disadvantages of using the step s^k as defined by (5) in each iteration of a generic trust-region method applied to (4).

5. (a) Consider the nonlinear least-squares problem

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}\|r(x)\|^2, \tag{6}$$

where $r : \mathbb{R}^n \to \mathbb{R}^m$ is twice-continuously differentiable with $m \geqslant n$ and $\|\cdot\|$ denotes the Euclidean norm.

(i) [7 marks] Write down an expression for the gradient $\nabla f(x)$ and the Hessian matrix $\nabla^2 f(x)$ of f in (6) as a function of the residual $r(x)$ and its derivatives. Derive the expression of the Gauss-Newton iteration (without linesearch) for (6). Show that the Gauss-Newton search direction is a descent direction from some x with $\nabla f(x) \neq 0$, provided the Jacobian of r at x has full column rank.

(ii) [9 marks] In (6), let

$$r(x) = \begin{pmatrix} x^2 - 1 \\ x^2 - 2x + \lambda \end{pmatrix}, \tag{7}$$

where $x \in \mathbb{R}$ and $\lambda \in \mathbb{R}$. Find all x^* such that $r(x^*) = 0$. Show that the local rate of convergence of the Gauss-Newton method (without linesearch) applied to (6) with r defined as in (7) is quadratic, when the starting point is close to each of the zero-residual solutions x^* that you found. Briefly justify this rate by comparison to the local rate of convergence of Newton's method.

(b) [9 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 twice continuously differentiable functions.

Write down the quadratic penalty function associated with (8). State (without proof) the global convergence theorem for the quadratic penalty method applied to (8).

Assume that on each major iteration of the quadratic penalty method, Newton's method with backtracking-Armijo linesearch is employed to minimize the corresponding quadratic penalty function. State (without proof) conditions under which this (inner) minimization can be terminated successfully, irrespective of the starting point.

6. Consider the inequality-constrained optimization problem,

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad c(x) \geqslant 0, \tag{9}$$

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

(a) (i) [2 marks] Write down the logarithmic barrier function associated with (9) and the conditions under which it is well-defined.

(ii) [5 marks] State the global convergence theorem of the basic barrier algorithm applied to (9).

(iii) [8 marks] In the conditions of the theorem you state in part (a)(ii), prove that the multiplier estimates at the kth iteration, $\lambda_i^k = \frac{\mu^k}{c_i(x^k)}$ converge to the optimal multipliers λ_i^* as $k \to \infty$, for $i \in \{1, \ldots, p\}$, where x^k are the (major) iterates of the basic barrier method that are assumed to converge to some point x^* and μ^k is the barrier parameter. *Hint*: Recall that $\lambda^* = (\lambda_{\mathcal{A}}^* \ \lambda_{\mathcal{I}}^*)$ where the partition is according to active and inactive constraints at the KKT point x^*, respectively. Also, under suitable assumptions, $\lambda_{\mathcal{I}}^* = 0$ and $\lambda_{\mathcal{A}}^* = J_{\mathcal{A}}(x^*)^+ \nabla f(x^*)$ where $J_{\mathcal{A}}(x^*)^+$ is the pseudo-inverse of $J_{\mathcal{A}}(x^*)^T$.

(b) [10 marks] In (9), let

$$f(x) = x_1^2 + (x_2 - 1)^2 \quad \text{and} \quad c(x) = x_1 + x_2 - 2, \tag{10}$$

where $x = (x_1 \ x_2)^T$ and $p = 1$. Find the minimizer(s) $x(\mu)$ of the logarithmic barrier function f_μ associated with (9) and (10), for any value $\mu > 0$ of the barrier parameter.

Let $\nabla_{xx}^2 f_\mu(x(\mu))$ be the Hessian matrix of the logarithmic barrier function f_μ evaluated at $x(\mu)$. Show that the condition number of $\nabla_{xx}^2 f_\mu(x(\mu))$ grows unboundedly as $\mu \to 0$. Briefly describe why this ill-conditioning causes difficulties in interior point methods, as well as a way to overcome it.

Hint: you may assume (without proof) that the solution of problem (9) with data (10) is $x^* = \left(\frac{1}{2} \ \frac{3}{2}\right)^T$.

 End of Last Page