Final.

DEGREE OF MASTER OF SCIENCE

MATHEMATICAL MODELLING AND SCIENTIFIC COMPUTING

B2 Further Numerical Linear Algebra and Continuous Optimization

---

HILARY TERM 2016

FRIDAY, 22 April 2016, 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. (a) [9 marks] State the QR iteration with shifts for computing approximate eigenvalues of a matrix A. Show that the eigenvalues that the QR iteration with shifts converges to are independent of the shifts. Describe how the shift parameter aids in the computation of the approximate eigenvalues.

   (b) [8 marks] Let $P(A) = \sum_{j=0}^{k} \alpha_j A^j$ be a polynomial in the matrix A. Consider a matrix A with all distinct real eigenvalues, one of which is contained in the interval $[c, d]$ and the remaining in $[a, b]$ where $b < c$. Let the polynomial $P(\cdot)$ satisfy the property that when applied to a scalar x, it gives that $1/2 \leqslant P(x) \leqslant 1$ for $x \in [c, d]$ and $|P(x)| \leqslant 1/10$ for $x \in [a, b]$. Describe the convergence properties of the power method for eigenvalues, applied to the matrix $P(A)$. Contrast applying the power method to $P(A)$ as opposed to A directly; specifically comment on convergence rates and what the power method would converge to in the two cases.

   (c) [8 marks] Can a sequence of invertible non-unitary matrices C_j be used to compute approximate eigenvalues of a matrix A by computing $A^{(j)} = C_j A^{(j-1)} C_j^{-1}$ for $j \geqslant 1$ with $A^{(0)} = A$? Contrast such an approach with the Jacobi algorithm for computing approximate eigenvalues of symmetric matrices.

2. Throughout this problem let $A \in \mathbb{R}^{m\times n}$ be an arbitrary nonzero matrix with $m < n$.

   (a) [7 marks] Explain how the power method for computing an eigenvalue can be applied to AA^T in order to compute the largest singular value of A.

   (b) [9 marks] State an algorithm to compute unitary matrices Q and H such that $B = QAH$ is bi-diagonal; that is, $B_{i,j}$ can only be non-zero for indices $i = j$ and $i = j - 1$.

   (c) [9 marks] Let the matrix B computed in part (b) have indices i such that $B_{i,i+1} = 0$. What are the implications of this for computing the singular value decomposition of A both in terms of computational cost and suitability for the computation to be subdivided and implemented across multiple computers simultaneously?

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. We apply a Generic Linesearch Method (GLM) with backtracking-Armijo linesearch to (1), where on each iteration $k \geqslant 0$, the search direction s^k from the current iterate x^k satisfies

$$s^k = -(B^k)^{-1} \nabla f(x^k), \tag{2}$$

where B^k is an $n \times n$ symmetric and positive definite matrix and ∇f is the gradient of f.

(a) [2 marks] Show that s^k defined in (2) is a descent direction provided $\nabla f(x^k) \neq 0$.

(b) [11 marks] State conditions on the matrices B^k, $k \geqslant 0$, and on the objective f, under which the GLM with s^k as in (2) is globally convergent from any $x^0 \in \mathbb{R}^n$, namely, either there exists an iteration $l \geqslant 0$ such that $\nabla f(x^l) = 0$ or $\nabla f(x^k) \longrightarrow 0$ as $k \longrightarrow \infty$. Give a proof of this global convergence result.

Hint: In your proof, you may use the following result. Let $f \in \mathcal{C}^1(\mathbb{R}^n)$ be bounded below on $\mathbb{R}^n$ and ∇f be Lipschitz continuous. Then (any) GLM with backtracking-Armijo linesearch applied to (1) satisfies: either there exists an iteration $l \geqslant 0$ such that $\nabla f(x^l) = 0$ or $\lim_{k \to \infty} \min \left\{ \frac{|\nabla f(x^k)^T s^k|}{\|s^k\|}, |\nabla f(x^k)^T s^k| \right\} = 0$.

(c) [7 marks] Assume that the matrices $(B^k)^{-1}$ in the GLM are updated by a quasi-Newton formula. In particular, on the kth iteration, after calculating x^{k+1}, we set

$$(B^{k+1})^{-1} = (B^k)^{-1} + \frac{\delta^k (\delta^k)^T}{(\gamma^k)^T \delta^k} + \tau v v^T,$$

for some $\tau \in \mathbb{R}$ and $v \in \mathbb{R}^n$, and where $\gamma^k = \nabla f(x^{k+1}) - \nabla f(x^k)$ and $\delta^k = x^{k+1} - x^k$. Find an expression for $(B^{k+1})^{-1}$ such that it satisfies the secant condition $(B^{k+1})^{-1} \gamma^k = \delta^k$, where $\gamma^k = \nabla f(x^{k+1}) - \nabla f(x^k)$ and $\delta^k = x^{k+1} - x^k$. (The expression of $(B^{k+1})^{-1}$ is the so-called DFP formula.)

(d) [5 marks] Assume we are in the case of part (c), namely, that the matrices $(B^k)^{-1}$ in (2) are updated using the DFP formula. Assume also that B^0 is symmetric and positive definite. Find a necessary condition (involving γ^k and δ^k) for $(B^{k+1})^{-1}$ to be positive definite for all $k \geqslant 0$. State, without proof, whether this condition is also sufficient.

Show that the condition you found holds if on each iteration k, the stepsize $\alpha^k > 0$ is required to satisfy

$$\nabla f(x^k + \alpha^k s^k)^T s^k \geqslant \beta \nabla f(x^k)^T s^k \tag{3}$$

where $0 < \beta < 1$.

4. Consider the trust-region subproblem

$$\min_{s \in \mathbb{R}^n} m(s) = c + s^T g + \frac{1}{2} s^T H s \quad \text{subject to} \quad \|s\| \leqslant \Delta \tag{4}$$

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

(a) [5 marks] State (without proof) the necessary and sufficient conditions that hold at a global minimizer s^* of (4).

(b) [15 marks] In (4), let $n = 3$, $c = 0$, and

$$H = \begin{pmatrix} a & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad g = \begin{pmatrix} b \\ 0 \\ 1 \end{pmatrix}. \tag{5}$$

Using the characterization of global minimizers in part (a) or otherwise, find the global minimizer of (4) in the following cases:

(i) $a \neq 0$ and $b \neq 0$;

(ii) $a < 0$ and $b = 0$.

(c) [5 marks] Consider problem (4) for general $n \geqslant 1$. Prove that if the conditions you stated in (a) hold, then s^* is a global minimizer of the trust-region subproblem (4) (this is the sufficiency part of the conditions you stated in (a)).

Hint: You may use the following fact: for a given (unconstrained) quadratic model $M(s) := c + s^T g + \frac{1}{2} s^T B s$ with B positive semidefinite, every $\overline{s}$ satisfying $B\overline{s} = -g$ is a global minimizer of M.

5. (a) Consider the following constrained optimization problem

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

(i) [3 marks] Find all solutions of problem (6).

(ii) [4 marks] Associate to problem (6), the following penalty function

$$\Phi_\sigma(x) = f(x) + \frac{1}{\sigma}(c(x))^2, \tag{7}$$

where $\sigma > 0$. Find the unconstrained, local or global, minimizers $x(\sigma)$ of $\Phi_\sigma(x)$ for $\sigma > 0$. Show that $x(\sigma)$ converges to a solution of (6) as $\sigma \longrightarrow 0$, and find the rate of this convergence as a function of σ.

(iii) [5 marks] Associate to problem (6), the following barrier function

$$f_\mu(x) = f(x) - \mu \log c(x), \tag{8}$$

where $\mu > 0$. Find the unconstrained, local or global, minimizers $x(\mu)$ of $f_\mu(x)$ for $\mu > 0$. Show that $x(\mu)$ converges to a solution of (6) as $\mu \longrightarrow 0$, and find the rate of this convergence as a function of μ. Briefly compare your findings here with those in (a)(ii).

(b) Consider the inequality-constrained optimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, A is a $p \times n$ matrix and $b \in \mathbb{R}^p$.

(i) [5 marks] Write down the logarithmic barrier function associated with (9) and the conditions under which it is well-defined. Describe its connection to the solution of problem (9), namely, the relation between optimality conditions for the barrier function and those of (9).

(ii) [8 marks] Describe the steps of the basic barrier (also called interior point) algorithm applied to (9). State (without proof) the theorem of global convergence of the basic barrier algorithm. Briefly describe two inefficiencies of the barrier method and one way to overcome them.

6. (a) [6 marks] Consider the quadratic programming problem

$$\min_{s \in \mathbb{R}^n} s^T g + \frac{1}{2} s^T B s \quad \text{subject to} \quad Js = -c, \tag{10}$$

where $g \in \mathbb{R}^n$, $c \in \mathbb{R}^m$, B is an $n \times n$ symmetric matrix and J is an $m \times n$ matrix, with $m \leqslant n$. Write down the KKT conditions for problem (10). Find conditions on the matrix B such that the KKT conditions are sufficient for optimality for problem (10); justify your answer.

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

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

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ and $c : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ are twice continuously differentiable and $m \leqslant n$. Briefly outline the basic steps of the SQP method with linesearch applied to (11), using the quadratic penalty merit function.

Under suitable conditions on the approximation B^k to the Lagrangian's Hessian matrix, show that, while the SQP iterate x^k is not a KKT point, the SQP step s^k is a descent direction for the quadratic penalty function provided the penalty parameter is sufficiently small.

(c) [9 marks] Consider again problem (11). Write down the Newton step s_N^k from some x^k for the quadratic penalty function associated with (11). Compare this Newton step s_N^k to the direction that the SQP method in (b) computes on iteration k, from the same x^k. Assume that the Hessian of the Lagrangian term that occurs in the Hessian of the quadratic penalty is approximated by a matrix B^k. Establish whether the Newton step s_N^k is descent for the quadratic penalty function under the same assumptions on the matrix B^k as those for the SQP step in (b).