DEGREE OF MASTER OF SCIENCE

MATHEMATICAL MODELLING AND SCIENTIFIC COMPUTING

B2 Further Numerical Linear Algebra and Continuous Optimization

---

HILARY TERM 2018

FRIDAY, 20 April 2018, 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. Let Π_m denote the set of real polynomials of degree m or less.

   (a) [6 marks] Given $A \in \mathbb{R}^{n \times n}$ and $r \in \mathbb{R}^n$, what are the *Krylov subspaces* $\mathcal{K}_k(A, r), k = 1, 2, \ldots$? For any non-negative integer ℓ, show that if $s = p(A)r$ with $p \in \Pi_\ell$, then $s \in \mathcal{K}_{\ell+1}(A, r)$. Calculate all of the Krylov subspaces when

   $$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix}, r = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} \qquad (1).$$

   (b) [1 mark] In general, why are Krylov subspaces more convenient for computation when A is a sparse matrix than when A is a dense matrix?

   (c) [9 marks] Without describing the details of the implementation of the method itself, give the criteria which determine the *GMRES* method for the iterative solution of a linear system $Ax = b$, where A is nonsingular. Derive a convergence bound for GMRES which bounds the Euclidean norm of the residual vectors $r_k = b - Ax_k$ for a diagonalisable matrix $A = X\Lambda X^{-1}$, where Λ is the diagonal matrix of eigenvalues of A.

   If it happens that $A = A^T$, what simplifies in the method?

   (d) [9 marks] Now apply GMRES to the specific matrix in (1) together with

   $$b = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \text{ and } x_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

   Show that $x_1 = x_0$, and calculate the GMRES residual vectors r_2 and r_3. In fact the eigenvalues of A are $2, -1, -1$. Reconcile the convergence bound that you have derived with the residuals that you find.

2. The *Conjugate Gradient method*:
choose x_0 , $r_0 = b - Ax_0 = p_0$ and for $k = 0, 1, 2, \ldots$

$$\begin{aligned}
\alpha_k &= p_k^T r_k / p_k^T A p_k \\
x_{k+1} &= x_k + \alpha_k p_k \\
r_{k+1} &= r_k - \alpha_k A p_k \\
&\text{if} \quad r_{k+1} = 0, \text{ stop : } x_{k+1} \text{ is the solution} \\
\beta_k &= -p_k^T A r_{k+1} / p_k^T A p_k \\
p_{k+1} &= r_{k+1} + \beta_k p_k.
\end{aligned}$$

is an iterative method for the solution of linear systems of equations, $Ax = b$.

(a) [3 marks] What properties must the matrix $A \in \mathbb{R}^{n \times n}$ have for the Conjugate Gradient method to be robustly applicable? What quantity is minimised at each iteration of the Conjugate Gradient method?

(b) [16 marks] Given that the residuals, $r_k, k = 0, 1, 2, \ldots$ and search directions, $p_k, k = 0, 1, 2, \ldots$ generated by the Conjugate Gradient method satisfy the orthogonality relations

$$\begin{aligned}
r_k^T p_j &= r_k^T r_j = 0 \ , \ j < k \\
p_k^T A p_j &= 0 \ , \ j < k \ , \ (p_k^T A p_k \neq 0)
\end{aligned}$$

so long as $x_k \neq x$, where x is the solution of $Ax = b$, prove that

$$\begin{aligned}
\text{span}\{r_0, r_1, \ldots, r_{k-1}\} &= \text{span}\{p_0, p_1, \ldots, p_{k-1}\} \\
&= \text{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{k-1} r_0\}.
\end{aligned}$$

Furthermore, prove that $x_k - x_0 \in \text{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{k-1} r_0\}$.

(c) [6 marks] If $P \in \mathbb{R}^{n \times n}$ is a symmetric projection matrix (so that $P = P^T$, $P^2 = P$) and $A = P + 2I$, what is the maximum number of Conjugate Gradient iterations that will be required to solve $Ax = b$ with any starting vector, x_0?

If $Q \in \mathbb{R}^{n \times n}$ is a oblique projection matrix (so that $Q \neq Q^T$, $Q^2 = Q$) and $B = Q + 2I$, what is the maximum number of Conjugate Gradient iterations that will be required to solve $B^T B y = c$ with any starting vector, x_0?

Section B: Continuous Optimization

3. (a) 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 and bounded below. Assume that the gradient ∇f of f is Lipschitz continuous with Lipschitz constant $L > 0$. Assume that the steepest descent method with linesearch is applied to (1) starting from some point $x^0 \in \mathbb{R}^n$, and where on each iteration $k \geqslant 0$, the stepsize $\alpha^k > 0$ is computed so as to satisfy the Armijo condition.

State the Armijo condition. Show that, on each iteration $k \geqslant 0$ for which $\nabla f(x^k) \neq 0$, the Armijo condition is satisfied for all stepsize values $\alpha^k \in (0, \alpha_{\max}]$, where $\alpha_{\max}$ is a constant independent of k but dependent on L. **[Marks 10]**

Assume that $\alpha^k = \alpha_{\max}$ for all $k \geqslant 0$. Show that the resulting steepest descent variant is globally convergent, namely, $\nabla f(x^k) \to 0$ as $k \to \infty$. **[Marks 5]**

(b) Consider the following function of two variables $x = (x_1 \; x_2)^T$,

$$f(x) = \frac{1}{4}x_1^4 + x_1x_2 + \frac{1}{2}(1 + x_2)^2. \tag{2}$$

Show that Newton's method (with or without linesearch) cannot be applied satisfactorily to minimize $f(x)$ if the starting point for Newton's method is $x^0 = (0 \; 0)^T$.

[Marks 6]

Now assume that at $x^0 = (0 \; 0)^T$, a modified Newton direction $\tilde{s}^0$ is computed where in place of $\nabla^2 f(x^0)$, we use the modified Hessian matrix $\nabla^2 f(x^0) + \nu I$ for some scalar ν and where I is the 2×2 identity matrix. Determine the range of ν values that would make this modified Newton direction $\tilde{s}^0$ suitable for minimizing $f(x)$ from x^0 (when a linesearch is allowed along this modified Newton direction). **[Marks 4]**

4. Consider the trust-region subproblem

$$\min_{s\in\mathbb{R}^n} m(s) = c + s^T g + \frac{1}{2} s^T H s \quad \text{subject to} \quad \|s\| \leqslant \Delta \tag{3}$$

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

(a) Let the Cauchy point s_C for (3) be defined as $s_C = -\alpha_C g$ where $\alpha_C = \arg\min_{\alpha>0} m(-\alpha g)$ subject to $\|-\alpha g\| \leqslant \Delta$. Calculate an explicit expression for s_C as a function of g, H and Δ. **[Marks 5]**

(b) State (without proof) the necessary and sufficient optimality conditions that hold at a global minimizer s^* of (3). **[Marks 5]**

(c) In (3), let $n = 3$, $c = 0$, $\Delta = 1$ and

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix} \quad \text{and} \quad g = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}. \tag{4}$$

For these values, calculate

1. the Cauchy point of (3);
2. the global minimizer of (3);
3. the global minimizer of (3) when s is further constrained to belong to the subspace spanned by the vectors g and Hg.

Briefly compare the resulting decreases $m(0) - m(s)$, where s is each of the calculated minimizers in (1), (2) and (3) above.
[Marks 15]

5. (a) A minimization algorithm is applied to

$$\min_{(x_1,x_2,x_3)\in\mathbb{R}^3} x_1^2+x_2^2+x_3^2 \quad \text{subject to } x_1-1 \geqslant 0,\ x_1+4x_2-5 \geqslant 0,\ \text{and } x_1+x_3-2 \geqslant 0. \quad (5)$$

It reaches the point $(\overline{x}_1, \overline{x}_2, \overline{x}_3)^T = (1,1,1)^T$; is this point a (local or global) minimizer of (5)? If not, find a feasible search direction from the point $(\overline{x}_1, \overline{x}_2, \overline{x}_3)^T$ that reduces the objective function. **[Marks 10]**

(b) Consider the equality-constrained optimization problem,

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad c(x) = 0, \qquad (6)$$

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 (6) is a KKT point of (6). **[Marks 15]**

6. (a) Consider the equality-constrained optimization problem,

$$\min_{x\in\mathbb{R}^3} x_1^2 + x_2^2 + x_3^2 \quad \text{subject to} \quad x_1 + 2x_2 + x_3 - 1 = 0, \tag{7}$$

where $x = (x_1\ x_2\ x_3)^T$. Calculate the (unconstrained) global minimizer(s) $x(\sigma)$ of the quadratic penalty function associated to (7), denoted by $\Phi_\sigma(x)$, for any $\sigma > 0$. Show that $x(\sigma)$ converges to the solution x^* of problem (7), as $\sigma \to 0$, and find the rate of this convergence as a function of σ. 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 $\sigma \to 0$. [**Marks 11**] [*Hint for part (a): you may assume (without proof) that the solution of problem* (7) *is* $x^* = \left(\frac{1}{6}\ \frac{1}{3}\ \frac{1}{6}\right)^T$ *with optimal Lagrange multiplier* $y^* = \frac{1}{3}$.]

(b) 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, and $m \leqslant n$. Consider the system

$$F_\sigma(x, y) := \begin{pmatrix} \nabla f(x) - J(x)^T y \\ c(x) + \sigma y \end{pmatrix} = 0, \tag{9}$$

where $(x, y) \in \mathbb{R}^{n+m}$ and $\sigma > 0$, and ∇f and J denote the gradient of f and the Jacobian of the constraints c, respectively.

(i) Establish a connection between solutions of the system (9) and stationary points of the quadratic penalty function $\Phi_\sigma(x)$ associated to (8). [**Marks 4**]

(ii) Consider the following primal-dual quadratic penalty method that starts from some starting point (x^0, y^0) and $\sigma^1 > 0$. On each iteration $k \geqslant 1$, starting from (x^{k-1}, y^{k-1}) and $0 < \sigma^k < \sigma^{k-1}$, it computes an approximate root (x^k, y^k) of $F_{\sigma^k}(x, y) = 0$ such that

$$\|F_{\sigma^k}(x^k, y^k)\| \leqslant \epsilon^k,$$

where $\|\cdot\|$ denotes the Euclidean norm and $\epsilon^k > 0$. By imposing conditions on ϵ^k and σ^k, state a theorem of global convergence for this primal-dual quadratic penalty method. In the conditions of the theorem you state, and assuming that $(x^k, y^k) \to (x^*, y^*)$ as $k \to \infty$, show that x^* is a KKT point of (8) with multiplier y^*. [**Marks 10**]