UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

College of Informatics and Electronics

**END OF SEMESTER ASSESSMENT PAPER**

MODULE CODE: MS4327

SEMESTER: Spring 2007

MODULE TITLE: Optimisation

DURATION OF EXAMINATION: 2 1/2 hours

LECTURER: Dr. J. Kinsella

PERCENTAGE OF TOTAL MARKS: 80%

EXTERNAL EXAMINER: Prof. J. King

INSTRUCTIONS TO CANDIDATES: **Answer four questions correctly for full marks; 80%. See the Appendix at the end of the paper for some useful results.**

1 (a) Given the one-dimensional "line" function

$$\phi(\alpha) = 2 - 2\alpha + 2\alpha^2 - 1/2\alpha^3;$$

show that the strong Wolfe conditions (Appendix B) hold in the interval $[0 \cdot 45, 1]$ (approximately). Take $c_1 = 0.1$ and $c_2 = 0.25$. 15

(b) The function $\phi(\alpha)$ has a local minimum at $\alpha = \frac{2}{3}$. Is this consistent with the above result? 1

(c) For any line function $\phi$, define $\Phi(\alpha) = \phi(\alpha) - (\phi(0) + c_1\alpha\phi'(0))$ and show that when $0 < c_1 < c_2 < 1$:

(i) For any $\alpha$, $\Phi(\alpha) \leq 0$ is equivalent to the first Wolfe condition (11a) holding at $\alpha$. 1

(ii) For any $\alpha$, $\Phi'(\alpha) = 0$ implies that the second (weak) Wolfe condition (11b) holds at $\alpha$. 2

(d) Use the results from (c) to show that if the first Wolfe condition fails for some $\alpha_0 > 0$ then the interval $(0, \alpha_0)$ contains values for $\alpha$ that satisfy the strong Wolfe conditions. (Assume that $\phi'(0) < 0$.) Illustrate your proof with a sketch graph of $\Phi$. 5

(e) Explain briefly the significance of this result for step-size selection algorithms. 1

2 The Dogleg Trust Region method is described in Appendix C.

(a) First show that for a positive definite matrix $B$ that the Cauchy-Schwartz inequality

$$(u^T v)^2 \leq (u^T u)^2 (v^T v)^2 \qquad (1)$$

implies that 2

$$(p^T B p)(p^T B^{-1} p) \geq (p^T p)^2 \quad \text{for any vector } p. \qquad (2)$$

(b) Show that if $B$ is positive definite then (you may assume the result from part (a))

(i) $\|\tilde{p}(\tau)\|$ is an increasing function of $\tau$ and 6

(ii) $m(\tilde{p}(\tau))$ is a decreasing function of $\tau$ . 10

(c) Explain why the path $\tilde{p}(\tau)$ intersects the trust region boundary $\|p\| = \Delta$ at exactly one point if $\|\tilde{p}(2)\| \equiv \|p_B\| \geq \Delta$ and nowhere otherwise. 1

(d) What value should be selected for $p$ if $\|p_B\| \leq \Delta$? 1

(e) In the case where the vector $p$ is chosen to be on the boundary, explain carefully how the appropriate value of the parameter $\tau$ is chosen. 2

(f) Finally, if $p_U = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $p_B = \begin{bmatrix} 0.5 \\ 1p.5 \end{bmatrix}$ and $\Delta = 1.5$, find the appropriate value of $\tau$. 3

3 The Fletcher-Reeves (FR) version of the non-linear conjugate gradient algorithm (Alg. 1) is given in Appendix D.

(a) Show that the function $h(t) = \frac{2t-1}{1-t}$ is increasing on $0 \leq t \leq \frac{1}{2}$ and conclude that $-1 \leq h(t) < 0$ on $[0, \frac{1}{2})$. 3

(b) Suppose that the algorithm is implemented with a step length $\alpha_k$ that satisfies the strong Wolfe conditions with $0 < c_2 < \frac{1}{2}$. Use the result of (a) to show that the method generates descent directions $p_k$ that satisfy the following inequalities: 14

$$-\frac{1}{1-c_2} \leq \frac{p_k^T g_k}{\|g_k\|^2} \leq \frac{2c_2 - 1}{1-c_2}, \quad \text{for all } k = 0, 1, \ldots.$$

(c) Explain the significance of the result. 1

(d) Explain carefully why the method "sticks" if a "bad" value of $p_k$ (satisfying $\frac{\|g_k\|}{\|g_k\|} \ll 1$) is generated at iteration $k$. 6

(e) What is the significance of this result for the method? 1

4 The BFGS method updates an estimate $H_k$ of the Inverse Hessian in the form:

$$H_{k+1} = (I - \gamma_k s_k y_k^T) H_k (I - \gamma_k y_k s_k^T) + \gamma_k s_k s_k^T. \tag{3}$$

Here $s_k = x_{k+1} - x_k$, $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$ and $\gamma_k = \frac{1}{s_k^T y_k}$.

(a) Derive this update rule by solving the problem: 18

$$\min_H \|H - H_k\| \tag{3a}$$

$$\text{subject to} \quad H = H^T, H y_k = s_k. \tag{3b}$$

Use the "weighted Frobenius norm": $\|A\|_W \equiv \|W^{\frac{1}{2}} A W^{\frac{1}{2}}\|_F$, where $\|C\|_F^2 \equiv \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2$. The weight $W$ is an unspecified positive definite symmetric matrix satisfying $W s_k = y_k$.

(b) Explain briefly how you could use the BFGS method to generate search directions. 1

(c) Show that (when the inverse Hessian approximations are generated using the BFGS algorithm) if $H_k$ is positive definite, then the next iterate $H_{k+1}$ is also ). 5

(d) Explain the significance of this result. 1

5 (a) Consider the constrained minimisation problem:

$$\min_{x \in \mathbb{R}^2} f(x) = (x_1 - a)^2 + (x_2 - b)^2 + x_1 x_2$$
$$\text{subject to} \quad 0 \le x_i \le 1, \text{for } i = 1, 2.$$

(i) Show that the unconstrained minimum of $f$ is at 3

$$x = \frac{2}{3}(2a - b, 2b - a).$$

(ii) Write the first-order necessary (KKT) conditions for this problem. (See Appendix E.) 4

(iii) When $a = \frac{7}{4}$ and $b = \frac{5}{4}$, show that the unconstrained minimum of $f$ is at $x = \left(\frac{3}{2}, \frac{1}{2}\right)$ — an infeasible point. 2

(iv) Show that because of the location of the unconstrained minimum either one or two of the constraints $x_1 \le 1$, $x_2 \ge 0$ and $x_2 \le 1$ must be active at the optimal point. Explain why the solution must either be at (A) the vertex $(1, 1)$, (B) on the edge $x_1 = 1$ or (C) at the vertex $(1, 0)$ — use a sketch to illustrate your argument. 4

(v) Solve the KKT equations to find the solution to the constrained problem for these values of $a$ and $b$ — check the three cases A, B and C separately. 8

(b) Consider the problem

$$\min_{x \in \mathbb{R}^2} f(x) = -2x_1 + x_2 \quad \text{subject to} \begin{cases} (1 - x_1)^3 - x_2 & \ge 0 \\ x_2 + 0.25x_1^2 - 1 & \ge 0. \end{cases} \quad (4)$$

(i) Show that the LICQ applies at $x^* = (0, 1)$ (active constraints are linearly independant). 2

(ii) Show that the KKT conditions are satisfied at $x^*$. 2

6 Consider the equality-constrained Quadratic Program (Q.P.):

$$\min_x q(x) = \frac{1}{2}x^T G x + x^T d, \tag{5a}$$

$$\text{subject to} \quad a_i^T x = b_u i, \quad i = 1, \dots, k \tag{5b}$$

(a) Writing the set of $k$ equality constraints (5b) as the matrix equation $Ax - b = 0$, show that if $x^*$ is a local minimum then the KKT conditions (Appendix E) require that there must be a vector $\lambda^*$ of Lagrange multipliers such that the following system of equations is satisfied: 3

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix} \tag{6}$$

(b) If we write $x^* = x_0 + p$, where $x_0$ is any estimate of the solution and $p$ the required step to the solution, show that (6) can be re-written as: 1

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} g \\ c \end{bmatrix} \tag{7}$$

where $c = Ax_0 - b$, $g = Gx_0 + d$ (the gradient of $q(x_0)$) and $p = x^* - x_0$.

(c) Let $A$ have full row rank and let $Z$ be the $n \times (n - m)$ matrix whose columns form a basis for the null space of $A$. $(AZ = 0.)$ Show that, if the "reduced-Hessian" matrix $Z^T G Z$ is positive definite, the KKT matrix

$$K = \begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \tag{8}$$

is non-singular. 8

(d) Briefly explain the significance of this result. 1

(e) Show that this block matrix equation can be reduced to the problem of solving

$$\left(AG^{-1}A^T\right)\lambda^* = \left(AG^{-1}g - c\right). \tag{9}$$

for $\lambda^*$ and then solving

$$-Gp + A^T\lambda^* = g \tag{10}$$

for $p$. 2

(f) Given $q(x) = 2x^2 + 2xy + 2y^2 - x + y$ and the **inequality** constraint $2x + 3y \geq 1$, solve the QP using the method you derived in (e). (First check whether the unconstrained solution is feasible — if it is, you are finished; if not the inequality constraint must be active at the optimal point.) 10

**Appendix of Results**

A The Wolfe conditions for the step length $\alpha$ in a line search require that

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha {p_k}^T g(x_k), \tag{10a}$$
$${p_k}^T g(x_k + \alpha p_k) \geq c_2 {p_k}^T g(x_k) \tag{10b}$$

where $g(x) \equiv \nabla f(x)$ and $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (10b) by

$$|{p_k}^T g(x_k + \alpha p_k)| \leq c_2 |{p_k}^T g(x_k)|. \tag{11}$$

B In terms of a "line" function $\phi(\alpha) \equiv f(x + \alpha p)$; the Wolfe conditions for the step length $\alpha$ in a line search require that

$$\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0), \tag{11a}$$
$$\phi'(\alpha) \geq c_2 \phi'(0) \tag{11b}$$

where $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (11b) by

$$|\phi'(\alpha)| \leq c_2 |\phi'(0)|. \tag{12}$$

C The Trust Region Method is based on the problem:

$$\min_{p \in \mathbb{R}^n} m(p) \equiv f_0 + g^T p + \frac{1}{2} p^T B p, \quad \text{such that } \|p\| \leq \Delta, \tag{13}$$

where $f_0$ is a fixed scalar, $g$ a fixed vector in $\mathbb{R}^n$, $B$ a fixed $n \times n$ matrix and $\Delta$ a fixed positive scalar. The "dogleg" method finds an approximate solution to (13) by replacing the (unknown) curved trajectory for $p^*(\Delta)$ with a path consisting of two line segments. The first line segment runs from the starting point to the unconstrained minimiser along the steepest descent direction defined by

$$p^U = -\frac{g^T g}{g^T B g} g \tag{14}$$

while the second line segment runs from $p^U$ to $p^B \equiv -B^{-1} g$. We can define the trajectory as a path $\tilde{p}(\tau)$ parameterised by $\tau$ as follows:

$$\tilde{p}(\tau) = \begin{cases} \tau p^U, & 0 \leq \tau \leq 1, \\ p^U + (\tau - 1)(p^B - p^U), & 1 \leq \tau \leq 2. \end{cases} \tag{15}$$

D **Algorithm 1 (FR-CGM)**

**begin**
Given $x_0$.
**set** $r_0 \leftarrow \nabla f_0, p_0 \leftarrow -r_0, k \leftarrow 0$;
**while** $r_k \neq 0$ **do**
$\alpha_k \leftarrow$ Result of line search along $p_k$;
$x_{k+1} \leftarrow x_k + \alpha_k p_k$;
$r_{k+1} \leftarrow \nabla f_{k+1}$
$\beta^{FR}_{k+1} \leftarrow \frac{r^T_{k+1} r_{k+1}}{r_k{}^T r_k} \equiv \frac{\|r_{k+1}\|^2}{\|r_k\|^2}$;
$p_{k+1} \leftarrow -r_{k+1} + \beta^{FR}_{k+1} p_k$;
$k \leftarrow k+1$;
**end** (while)
**end**

E The first-order KKT necessary conditions for a point $x^*$ with optimal multipliers $\lambda^*$ to be a local solution of an equality-constrained minimisation problem are as follows: Suppose that $x^*$ is a local solution of a constrained minimisation problem and that the LICQ holds at $x^*$. Then there is a Lagrange multiplier vector $\lambda^*$, with components $\lambda^*_i, i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at $(x^*, \lambda^*)$

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{16a}$$
$$c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E}, \tag{16b}$$
$$c_i(x^*) \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{16c}$$
$$\lambda^*_i \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{16d}$$
$$\lambda^*_i c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E} \cup \mathcal{I} \tag{16e}$$