UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

Faculty of Science and Engineering

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327

SEMESTER: Spring 2008

MODULE TITLE: Optimisation

DURATION OF EXAMINATION: 2 1/2 hours

LECTURER: Dr. J. Kinsella

PERCENTAGE OF TOTAL MARKS: 70%

EXTERNAL EXAMINER: Prof. J. King

INSTRUCTIONS TO CANDIDATES: **Answer four questions correctly for full marks; 70%. See the Appendix at the end of the paper for some useful results.**

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

$$\phi(\alpha) = 2 - \alpha + 2\alpha^2 - 1/4\alpha^3;$$

show that the strong Wolfe conditions, (9) in Appendix B, hold in the interval $[0 \cdot 19, 0 \cdot 33]$ (approximately). Take $c_1 = 0.1$ and $c_2 = 0.25$. 13

(b) The function $\phi(\alpha)$ has a local minimum at $\alpha \approx 0{\cdot}26$. Is this consistent with the above result? 1

(c) For any line function ϕ, 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 α, $\Phi(\alpha) \leq 0$ is equivalent to the first Wolfe condition (8a) holding at α. 1

(ii) For any α, $\Phi'(\alpha) = 0$ implies that the second (strong) Wolfe condition (8b) holds at α. 1

(iii) $\Phi(0) = 0$ and $\Phi'(0) < 0$. 1

(d) Use the results from (c) to

(i) sketch the graph of Φ — assume that the first Wolfe condition fails for some $\alpha_0 > 0$, 2

(ii) show that — if the first Wolfe condition fails for some $\alpha_0 > 0$ — the interval $(0, \alpha_0)$ contains values for α that satisfy the strong Wolfe conditions. (Assume that $\phi'(0) < 0$.) Refer to your sketch graph of Φ . 5

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

2 The Two-dimensional subspace Trust Region method is described in Appendix C.

(a) Show that the problem of chosing the optimal value of the parameters α and β can be reduced to that of finding the roots of a quartic (degree 4) polynomial. 24

Hint: write $p = \alpha g + \beta g_1$, where $g_1 = g + \gamma B^{-1} g$ and choose γ so that $g^T g_1 = 0$.

(b) What are the implications of this result for the efficient approximate solution of the sub-problem (10) in App. C? 1

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

(a) Suppose that the algorithm is implemented with a step length α_k that satisfies the strong Wolfe conditions with $0 < c_2 < \frac{1}{2}$ and that the norm of the gradient is bounded above. Assume Zoutendijk's Theorem (Theorem 1 in Appendix D) and the result that

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

and prove that the FR cgm has the global convergence property — i.e. that 23

$$\lim_{k\to\infty} \inf \|g_k\| = 0.$$

(b) Explain the significance of the result. 2

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

$$J_{k+1} = (I - \gamma_k s_k y_k^T) J_k (I - \gamma_k y_k s_k^T) + \gamma_k s_k s_k^T. \quad (1)$$

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

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

$$\min_J \|J - J_k\| \quad (1a)$$

$$\text{subject to} \quad J = J^T, J y_k = s_k. \quad (1b)$$

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 $W s_k = y_k$.

(i) First show that any matrix $J = K^T A K + P$, where $K = I - \gamma_k y_k s_k^T$, $P = \gamma s_k s_k^T$, $\gamma_k = 1/(s_k^T y_k)$ and A is symmetric, satisfies $J = J^T, J y_k = s_k$. 2

(ii) Next show that $\|A\|_W^2 = \text{trace}\, WAWA$ for any symmetric matrix A. 6

(iii) Finally explicitly differentiate $\|J - J_k\|_W^2$ wrt $A_{\alpha\beta}$ — an arbitrary entry of the matrix J — set the result to zero and conclude that $A = J_k$. 10

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

(c) Show that if

- the inverse Hessian approximations are generated using the BFGS algorithm and
- the step lengths α_k satisfy the second Wolfe condition,

then if J_k is positive definite, then the next iterate J_{k+1} is also. 5

(d) Explain the significance of this result. 1

5 Derive the KKT necessary condition — (14a) in App. G — for an equality-constrained optimisation problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad c_i(x) = 0, i \in \mathcal{E} \tag{2}$$

(a) First prove that the penalty function $F_k(x)$ defined in (13) in App. F has the property that the (unconstrained) minima x_k of F_k converge to a local minimum x^*. 13

(b) Now use the result of part (a) to derive the KKT conditions for the equality-constrained problem defined in Eq. 2. 12

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

$$\min_x q(x) = \frac{1}{2} x^T G x + x^T d, \tag{3a}$$

$$\text{subject to} \quad a_i^T x = b_i, \quad i = 1, \ldots, k \tag{3b}$$

(a) Let A be the matrix whose columns are the vectrs $\{a_i\}_{i \in \mathcal{E}}$.Writing the set of k equality constraints (3b) as the matrix equation $Ax - b = 0$, show that if x^* is a local minimum then the KKT conditions (Appendix G) require that there must be a vector λ^* of Lagrange multipliers such that the following system of equations is satisfied: 5

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ -\lambda^* \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix} \tag{4}$$

(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 (4) can be re-written as: 2

$$\begin{bmatrix} G & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} g \\ c \end{bmatrix} \tag{5}$$

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

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

$$\left(AG^{-1}A^T\right)\lambda^* = \left(AG^{-1}g - r\right). \tag{6}$$

for λ^* and then solving

$$-Gp + A^T\lambda^* = g \tag{7}$$

for p. 2

(d) Given $q(x) = x^2 - xy + 3y^2 - x + y$ 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.) 15

(e) Comment briefly on the result you found in (d). 1

Appendix of Results

A The Wolfe conditions for the step length α in a line search require that

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha {p_k}^T g(x_k), \quad (7a)$$
$${p_k}^T g(x_k + \alpha p_k) \geq c_2 {p_k}^T g(x_k) \quad (7b)$$

where $g(x) \equiv \nabla f(x)$ and $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (7b) by

$$|{p_k}^T g(x_k + \alpha p_k)| \leq c_2 |{p_k}^T g(x_k)|. \quad (8)$$

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

$$\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0), \quad (8a)$$
$$\phi'(\alpha) \geq c_2 \phi'(0) \quad (8b)$$

where $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (8b) by

$$|\phi'(\alpha)| \leq c_2 |\phi'(0)|. \quad (9)$$

C The 2-dimensional subspace method finds an approximate solution to the problem:

$$\min_{p \in \mathbb{R}^n} m(p) \equiv f_0 + g^T p + \frac{1}{2} p^T B p, \quad \text{such that } \|p\| \leq \Delta, \quad (10)$$

where f_0 is a fixed scalar, g a fixed vector in $\mathbb{R}^n$, B a fixed $n \times n$ positive definite nmatrix and Δ a fixed positive scalar. The method seeks to find the linear combination $p = \alpha g + \beta p_B$ ($p_B \equiv -B^{-1} g$ that minimises $m(p)$ as a function of the real variables α and β.

If $\|p_B\| < \Delta$ the method takes $p = p_B$ so we may re-write the constraint as $\|p\| = \Delta$.

D **Theorem 1 (Zoutendijk)** *Consider any iteration of the form* $x_{k+1} = x_k + \alpha_k p_k$, *where* p_k *is a descent direction and* α_k *satisfies the Wolfe conditions Eqs. 7a and 7b in Appendix A above. Suppose that* f *is bounded below in* $\mathbb{R}^n$ *and that* f *is* C^1 *in an open set* $\mathcal{N}$ *containing the level set* $\mathcal{L} \equiv \{x : f(x) \leq f(x_0)\}$, *where* x_0 *is the starting point. Also assume that* $g(x)$, *the gradient of* f, *is Lipschitz continuous on* $\mathcal{N}$, *i.e. there exists a constant* L *such that*

$$\|g(x) - g(\bar{x})\| \leq L\|x - \bar{x}\|, \quad \textit{for all } x, \bar{x} \in \mathcal{N}. \tag{11}$$

Then

$$\sum_{k \geq 0} \cos^2 \theta_k \, \|g(x_k)\|^2 < \infty, \tag{12}$$

where θ_k *is the angle between* p_k *and the steepest descent direction* $-g(x_k)$.

E **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

F Let x^* be a local minimum of the equality-constrained optimisation defined in Eq. 2 in Q5. For each positive integer k, define a **penalty function**

$$F^k(x) = f(x) + \frac{k}{2}\|c(x)\|^2 + \frac{\alpha}{2}\|x - x^*\|^2, \tag{13}$$

where $\alpha > 0$ is arbitrary.

G The first-order KKT necessary conditions for a point x^* with optimal multipliers λ^* to be a local solution of an inequality-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 λ^*, with components $\lambda_i^*, i \in \mathcal{E} \cup \mathcal{I}$, such that the following conditions are satisfied at (x^*, λ^*)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{14a}$$

$$c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E}, \tag{14b}$$

$$c_i(x^*) \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{14c}$$

$$\lambda_i^* \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{14d}$$

$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E} \cup \mathcal{I} \tag{14e}$$

where $\mathcal{L}(\mathbf{x}, \lambda) \equiv f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$.