UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

Faculty of Science and Engineering

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327

SEMESTER: Spring 2010

MODULE TITLE: Optimisation

DURATION OF EXAMINATION: 2 1/2 hours

LECTURER: Dr. J. Kinsella

GRADING SCHEME: 20% (Project)+80% (Exam)

EXTERNAL EXAMINER: Dr. P. Howell

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) Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is C^1. Let p be a descent direction at x and assume that f is bounded below along the half-line $\{x + \alpha p | \alpha > 0\}$. Show that if $0 < c_1 < c_2 < 1$ there exists at least one interval of search lengths satisfying the Wolfe conditions Eqs. 9a and 9b in Appendix B. 8

(b) A simple Backtracking Line Search algorithm (Algorithm 1) is given in Appendix C. Prove that, provided at least one iteration takes place and for ρ sufficiently close to 1 (i.e. provided the backtracking is sufficiently slow) the algorithm will produce an interval $I = [\alpha_1, \alpha_2]$ such that the Wolfe conditions Eqs. 9a and 9b are satisfied for all $\alpha \in I$ with $c_1 = c$ and for some $c_2 > c_1$. 12

(c) Alg. 1 assumes but does not check that the initial α-value ($\bar{\alpha}$) violates the first Wolfe condition. Write a short piece of Matlab or pseudo-code that computes a suitable value for $\bar{\alpha}$. 5

2 The Trust-Region sub-problem can be stated as:

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

where g is the gradient of the objective function f at the current point and B is either the Hessian of f at the current point or an approximation to it.

(a) The Dogleg Trust Region method is described in Appendix D. Draw a sketch to illustrate the method. 2

(b) If B is positive definite then prove **one of the following**: 10

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

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

(c) What is the significance of these two results? 2

(d) Derive the solution of the Two-Dimensional Subspace Minimisation (TDSM) problem when B is positive definite.

(i) Write p as $p = \alpha g + \beta g_1$ where $g_1 = g + \gamma B^{-1} g$ and γ is chosen so that $g^T g_1 = 0$. 2

(ii) Show that the equation $\|p\|^2 = \Delta^2$ is an ellipse in the α–β plane. 1

(iii) Parameterise α & β appropriately. 1

(iv) Express $m(p)$ as a quadratic in α & β. 1

(v) Finally, use the parameterised form of α and β to express $m(p)$ in terms of a single angle; θ, say, where $0 \le \theta \le 2\pi$. 3

(vi) Show that the equation to be solved can be reduced to a fourth-order polynomial in either $\sin\theta$ or $\cos\theta$. 2

(e) Can you say whether one of the two methods (Dogleg and TDSM) is necessarily better than the other? Why? 1

3 The Fletcher-Reeves (FR) version of the non-linear conjugate gradient algorithm (Alg. 2) is given in Appendix F.

(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 E) 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

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

(b) Explain the significance of the result. 2

4 Consider the **inverse** BFGS formula $H_{k+1} = H_k - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} + \gamma_k y_k y_k^T$. Here $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$, $g_k = \nabla f(x_k)$ and $\gamma_k = \frac{1}{s_k^T y_k}$. You are asked to prove that $\liminf \|g_k\| = 0$.

The proof is broken into steps, only some of which you are asked to check.

(a) Define $\overline{H}_k = \int_0^1 \nabla^2 f(x_k + \tau\alpha_k p_k) d\tau$. Show that $\overline{H}_k s_k = y_k$ (use the Chain Rule). 2

(b) Assume that $m\|z\|^2 \le z^T \nabla^2 f(x) z \le M\|z\|^2$ for all $x, z \in \mathbb{R}^n$ and define $m_k = \frac{y_k^T s_k}{s_k^T s_k}$, $M_k = \frac{y_k^T y_k}{y_k^T s_k}$. **You may assume** that $m_k \ge m$ and $M_k \le M$ for all k using the definitions of s_k and y_k. 0

(c) Show that trace $H_{k+1} =$ trace $H_k - \frac{\|H_k s_k\|^2}{s_k^T H_k s_k} + \frac{\|y_k\|^2}{y_k^T s_k}$ and that det $H_{k+1} =$ det $H_k \left(\frac{y_k^T s_k}{s_k^T H_k s_k}\right)$. (**You may assume** that $\det(I + xy^T + uv^T) = (1 + y^T x)(1 + v^T u) - (x^T v)(y^T u)$ for any vectors x, y, u and v in $\mathbb{R}^n$) 6

(d) **You may assume** that the definition $\cos\theta_k = \frac{s_k^T H_k s_k}{\|s_k\|\|H_k s_k\|}$ is equivalent to the standard definition $\cos\theta_k = -\frac{p_k^T g_k}{\|p_k\|\|g_k\|}$. 0

(e) **You may assume** that setting $q_k = \frac{s_k^T H_k s_k}{\|s_k\|^2}$ allows the results from (c) to be amended to trace $H_{k+1} =$ trace $H_k - \frac{q_k}{\cos^2\theta_k} + M_k$ and det $H_{k+1} =$ det $H_k \left(\frac{m_k}{q_k}\right)$. 0

(f) For any $n \times n$ matrix B, define $\psi(B) = \text{trace}(B) - \ln\det B$ and show that if B is positive definite then $\psi(B) > 0$. 2

(g) Assume Zoutendijk's Theorem (Theorem 1 in Appendix E) and prove that for any starting point x_0, if the objective function f is C^2 and the assumptions and conclusions of the preceding parts of the question hold then the sequence $\{x_k\}$ generated by the **inverse** BFGS formula satisfies $\liminf \|g_k\| = 0$. 15

5 Consider the general equality-constrained problem as defined in App. H.

(a) Show that the penalty function $F^k(x)$ defined in (16) in App. G has the property that the (unconstrained) minima x_k of F^k converge to a local minimum x^*. 13

(b) Prove the KKT first-order necessary conditions (19a) and (19b) in App. J for an **equality**-constrained problem . 12

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

$$\min_x q(x) = \frac{1}{2}x^T Q x + x^T d, \tag{2a}$$
$$\text{subject to} \quad a_i^T x = b_i, \quad i = 1, \dots, k \tag{2b}$$

(a) Let A be the matrix s.t. the vectors $\{a_i\}_{i\in\mathcal{E}}$ are the columns of A^T. Writing the set of k equality constraints (2b) as the matrix equation $Ax - b = 0$, show that if x^* is a local minimum then the KKT conditions (Appendix J) require that there must be a vector λ^* of Lagrange multipliers such that the following system of equations is satisfied: 4

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^* \\ -\lambda^* \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix} \tag{3}$$

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

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -p \\ \lambda^* \end{bmatrix} = \begin{bmatrix} g \\ r \end{bmatrix} \tag{4}$$

where the residual $r = Ax_0 - b$, $g = Qx_0 + d$ (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

$$\left(AQ^{-1}A^T\right)\lambda^* = \left(AQ^{-1}g - r\right). \tag{5}$$

for λ^* and then solving

$$-Qp + A^T\lambda^* = g \tag{6}$$

for p. 3

(**See the following page for part (d) of this question**.)

(d) Solve the **Inequality**-constrained QP:

$$\min_{x \in \mathbb{R}^2} f(x) \text{ subject to } Ax \geq b, \tag{7}$$

with $f(x) = \frac{1}{2}x^T Q x + d^T x$ where $Q = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$ and $d = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ with $A = \begin{bmatrix} 3 & 2 \\ 1 & -3 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ using the following steps.

(i) Find x_u, the **unconstrained** minimum of $f(x)$, ($\nabla f(x_u) = 0$). 3

(ii) Confirm that x_u is infeasible. 2

(iii) Check that $x_0 = \frac{1}{11}\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ is the vertex common to the two constraints — the intersection point of the two lines. (A rough sketch of the two lines is useful in the following.) 1

(iv) Given

$$(AQ^{-1}A^T)^{-1} = \frac{1}{121}\begin{bmatrix} 68 & -6 \\ -6 & 29 \end{bmatrix}$$
$$AQ^{-1} = \frac{1}{16}\begin{bmatrix} 9 & 1 \\ 14 & -18 \end{bmatrix} \quad \text{and}$$
$$g(x_0) = \frac{1}{11}\begin{bmatrix} 29 \\ 36 \end{bmatrix}$$

use the procedure in part (c) of this question to solve the **Equality**-constrained QP

$$\min_{x \in \mathbb{R}^2} f(x) \text{ subject to } Ax = b, \tag{8}$$

as both constraints are binding at x_0. Note that $r = 0$ at x_0. Do not complete the arithmetic necessary to calculate λ^*, just check that its first component is positive and its second is negative. There is no need to calculate p. 4

(v) Discard the constraint corresponding to the negative multiplier (update A and b) and resolve for λ. Confirm that the single Lagrange multiplier λ is positive. You may take 4

$$AQ^{-1}A^T = \frac{29}{16}$$
$$AQ^{-1} = \frac{1}{16}\begin{bmatrix} 9 & 1 \end{bmatrix}$$

(vi) Now solve for p as in Eq. 6 in part (c) of this question. (Just set up the matrix products, do not perform the arithmetic). How would you use the vector p to compute the solution? 2

Appendix of Results

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

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha {p_k}^T g(x_k), \tag{8a}$$
$${p_k}^T g(x_k + \alpha p_k) \geq c_2 {p_k}^T g(x_k) \tag{8b}$$

where $g(x) \equiv \nabla f(x)$ and $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (8b) by

$$|{p_k}^T g(x_k + \alpha p_k)| \leq c_2 |{p_k}^T g(x_k)|. \tag{9}$$

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

$$\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0), \tag{9a}$$
$$\phi'(\alpha) \geq c_2 \phi'(0) \tag{9b}$$

where $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (9b) by

$$|\phi'(\alpha)| \leq c_2 |\phi'(0)|. \tag{10}$$

C **Algorithm 1 (Backtracking Line Search)**

```
begin
   Choose ᾱ > 0 and ρ, c ∈ (0, 1)
   α := ᾱ
   while φ(α) ≥ φ(0) + cαφ'(0) do
         α := ρα
   end
   α_k := α
end
```

D The Trust Region Method is based on the problem:

$$\min_{p \in \mathbb{R}^n} m(p) \equiv f_0 + g^T p + \frac{1}{2} p^T B p, \quad \text{such that } \|p\| \leq \Delta, \tag{11}$$

where f_0 is a fixed scalar, g a fixed vector in $\mathbb{R}^n$, B a fixed $n \times n$ matrix and Δ a fixed positive scalar. The "dogleg" method finds an approximate solution to (11) 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

$$p^u = -\frac{g^T g}{g^T B g} g \tag{12}$$

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 τ as follows:

$$\tilde{p}(\tau) = \begin{cases} \tau p^U, & 0 \le \tau \le 1, \\ p^U + (\tau - 1)(p^B - p^U), & 1 \le \tau \le 2. \end{cases} \tag{13}$$

E **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. 8a and 8b 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 $\mathcal{L} \equiv \{x : f(x) \le f(x_0)\}$, 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*

$$\|g(x) - g(\bar{x})\| \le L\|x - \bar{x}\|, \quad \textit{for all } x, \bar{x} \in \mathcal{N}. \tag{14}$$

Then

$$\sum_{k \ge 0} \cos^2 \theta_k \, \|g(x_k)\|^2 < \infty, \tag{15}$$

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

F **Algorithm 2 (FR-CGM)**

begin
  Given x_0.
  set $g_0 \leftarrow \nabla f_0, p_0 \leftarrow -g_0, k \leftarrow 0$;
  while $g_k \neq 0$ do
    $\alpha_k \leftarrow$ Result of line search along p_k;
    $x_{k+1} \leftarrow x_k + \alpha_k p_k$;
    $g_{k+1} \leftarrow \nabla f_{k+1}$
    $\beta^{FR}_{k+1} \leftarrow \frac{\|g_{k+1}\|^2}{\|g_k\|^2}$;
    $p_{k+1} \leftarrow -g_{k+1} + \beta^{FR}_{k+1} p_k$;
    $k \leftarrow k + 1$;
  end (while)
end

G Let x^* be a local minimum of the equality-constrained optimisation problem defined in App. H . For each positive integer k, define a **penalty function**

$$F^k(x) = f(x) + \frac{k}{2}\|c(x)\|^2 + \frac{\alpha}{2}\|x - x^*\|^2, \tag{16}$$

where $\alpha > 0$ is arbitrary.

H The general equality constrained optinisation problem is:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \quad c_i(x) = 0, i \in \mathcal{E} \tag{17}$$

I The general inequality constrained optinisation problem is:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \begin{cases} c_i(x) = 0 & i \in \mathcal{E}, \\ c_i(x) \geq 0 & i \in \mathcal{I} \end{cases} \tag{18}$$

J 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 general constrained optimisation (as in App. I) problem and that the LICQ holds at x^* (the active constraint gradients are linearly independant). 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^*)

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{19a}$$

$$c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E}, \tag{19b}$$

$$c_i(x^*) \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{19c}$$

$$\lambda_i^* \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{19d}$$

$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E} \cup \mathcal{I} \tag{19e}$$

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

K The Second-Order necessary conditions for a point x^* with optimal multipliers λ^* to be a local solution of an equality-constrained minimisation problem are as follows: Suppose that x^* is a local solution of an equality-constrained problem as defined in App. H and that the LICQ (see App. J) constraint qualification is satisfied. Let λ^* be a Lagrange multiplier vector such that the first-order (KKT) necessary conditions in App. J are satisfied,

Then $w^T \nabla_{xx} \mathcal{L}(x^*, \lambda^*) w \geq 0$, for all vectors w such that $w \perp \nabla c_i(x^*)$, for each $i \in \mathcal{E}$.