UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

Faculty of Science and Engineering

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327

SEMESTER: Spring 2012

MODULE TITLE: Optimisation

DURATION OF EXAMINATION: 2 1/2 hours

LECTURER: Dr. J. Kinsella

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

EXTERNAL EXAMINER: Prof. T. Myers

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) Prove Zoutendijk's Theorem (Theorem 1 in Appendix D) 15

(b) Explain briefly the significance of the result. 1

(c) Suppose that search directions p_k are generated using a Newton-like method: $p_k = -B_k^{-1} g(x_k)$ where B_k is symmetric and positive definite. For any matrix A let $\|A\|$ be the matrix 2–norm, equal for real symmetric matrices to the absolute value of the largest eigenvalue of A. Show that if $\|B_k\|\|B_k^{-1}\| \leq M$ for all k then $\cos\theta_k \geq 1/M$ where θ_k is defined in Appendix D. 8

(d) Use Zoutendijk's Theorem above to show that in this case $\lim_{k\to\infty} \|g(x_k)\| = 0$. 1

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 C. 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 $0 \leq \theta \leq 2\pi$. 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. 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

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

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

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

Hints: (each step uses some or all of those preceding it)

(i) Use a proof by contradiction — so assume that $\|g_k\|$ is bounded below, say by $\gamma > 0$.

(ii) First use Eq. 2 above to show that

$$\frac{1-2c_2}{1-c_2}\frac{\|g_k\|}{\|p_k\|} \leq \cos\theta_k \leq \frac{1}{1-c_2}\frac{\|g_k\|}{\|p_k\|}.$$

(iii) Then use Zoutendijk's Theorem to show that $\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|p_k\|^4}$ converges.

(iv) Use the strong version of the second Wolfe condition (Eq. 5 in App. A) and the bound on $\cos\theta_k$ above to show that $|g_k^T p_{k-1}| \leq \frac{c_2}{1-c_2}\|g_{k-1}\|^4$.

(v) Use the update rule for p_{k+1} in the FR CGM algorithm (Appendix E) (replacing $k+1$ by k) to show that

$$\|p_k\|^2 \leq c_3\|g_k\|^2 + \beta_k^2\|p_{k-1}\|^2$$

with $c_3 = \frac{1+c_2}{1-c_2}$.

(vi) Finally iterate this inequality to obtain a contradiction with the result of hint (iii) — you can assume that $\|g_k\|$ is bounded above, say by $\bar{\gamma}$.

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

(b) Assume that $m\|z\|^2 \leq z^T\nabla^2 f(x)z \leq M\|z\|^2$ 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 $m_k \geq m$ and $M_k \leq M$ 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 D) 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

(Hint: find an upper bound on $\psi(H_{k+1})$ in terms of $\psi(H_k)$ and iterate it. Show that $\psi(H_{k+1})$ becomes negative for k sufficiently large, contradicting the fact that $\psi(H_{k+1}) > 0$.

5 Consider the general equality-constrained problem (13) as defined in App. G.

(a) Show that the penalty function $F^k(x)$ defined in (12) in App. F has the property that the (unconstrained) minima x_k of F^k converge to a local minimum x^*. 10

(b) Prove the KKT first-order necessary conditions (15a) and (15b) in App. I for an **equality**-constrained problem . 10

(c) Finally; prove that when inequality constraints are introduced, the extra condition $\lambda_i^* \geq 0$ for all $i \in \mathcal{I}$ is necessary for x^* to be an optimal solution of (14) in App. H. Define the functions $c_i^-(x) = \min\{0, c_i(x)\}$, $j \in \mathcal{I}$. Obviously $c_i^-(x) \leq 0$, $j \in \mathcal{I}$.

Adapt (without repeating the details) your argument for part (b) to show that $\lambda_i^* = -\lim_{k\to\infty} kc_i^-(x_k), i \in \mathcal{I}$. 5

6 This question relates to the application of the KKT necessary conditions ((15a)–(15e) in App. I) to the Linear Program (LP) in Standard Form (SF)

$$\min c^T x, \text{ subject to } Ax = b,\ x \geq 0. \tag{3}$$

where c and x are vectors in $\mathbb{R}^n$, b is a vector in $\mathbb{R}^m$, and A is an $m \times n$ matrix.

(a) Show that the KKT conditions for the SF LP (Eq. 3 above) take the form: 5

$$A^T\lambda + s = c \tag{3a}$$
$$Ax = b \tag{3b}$$
$$x \geq 0 \tag{3c}$$
$$s \geq 0 \tag{3d}$$
$$x_i s_i = 0,\ i = 1, \ldots, n. \tag{3e}$$

(b) Show that the KKT conditions for the "dual" LP

$$\max b^T\lambda, \text{ subject to } A^T\lambda \leq c. \tag{4}$$

(when the dual LP is re-fornulated into SF) are the same as those for the primal problem (3). 5

(c) Show that the optimal solutions of the primal and dual problems are equal — $c^T x^* = b^T\lambda^*$ and explain the significance of this result. 2

(d) A vector x is a basic feasible point for the SF LP (3) above if it is feasible and if there exists a subset $\mathcal{B}$ of the index set $\{1, 2, \ldots, n\}$ such that

- $\mathcal{B}$ contains exactly m indices ($m < n$).
- $i \notin \mathcal{B} \Rightarrow x_i = 0$ (i.e. the bound $x_i \geq 0$ is inactive only if $i \in \mathcal{B}$).
- The $m \times m$ matrix B defined by $B = [A_i]_{i\in\mathcal{B}}$ is non-singular (A_i is the i^{th} column of A)

(Part (d) continued overleaf.)

Show that if we partition the vectors x, s and c according to the index sets $\mathcal{B}$ and $\mathcal{N} = \{1, 2, \ldots, n\}\backslash\mathcal{B}$, using the notation

$$\begin{aligned} x_B &= [x_i]_{i\in\mathcal{B}}, & x_N &= [x_i]_{i\in\mathcal{N}} \\ s_B &= [s_i]_{i\in\mathcal{B}}, & s_N &= [s_i]_{i\in\mathcal{N}} \\ c_B &= [c_i]_{i\in\mathcal{B}}, & c_N &= [c_i]_{i\in\mathcal{N}} \end{aligned}$$

the Simplex algorithm (Alg. 2 in App. J) may be derived from the KKT conditions (3a)– (3e). 13

(Take $\mathcal{B} = \{1, 2, \ldots, m\}$, $\mathcal{N} = \{m + 1, \ldots, n\}$ and $N = [A_i]_{i\in\mathcal{N}}$.)

Appendix of Results

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

$$\begin{aligned} f(x_k + \alpha p_k) &\leq f(x_k) + c_1 \alpha {p_k}^T g(x_k), && (4a) \\ {p_k}^T g(x_k + \alpha p_k) &\geq c_2 {p_k}^T g(x_k) && (4b) \end{aligned}$$

where $g(x) \equiv \nabla f(x)$ and $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (4b) by

$$|{p_k}^T g(x_k + \alpha p_k)| \leq c_2 |{p_k}^T g(x_k)|. \tag{5}$$

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

$$\begin{aligned} \phi(\alpha) &\leq \phi(0) + c_1 \alpha \phi'(0), && (5a) \\ \phi'(\alpha) &\geq c_2 \phi'(0) && (5b) \end{aligned}$$

where $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (5b) by

$$|\phi'(\alpha)| \leq c_2 |\phi'(0)|. \tag{6}$$

C The Trust Region Method is based on the problem:

$$\min_{p \in \mathbb{R}^n} m(p) \equiv f_0 + g^T p + \frac{1}{2} p^T B p, \quad \text{such that } \|p\| \leq \Delta, \tag{7}$$

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 (7) 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

$$p^U = -\frac{g^T g}{g^T B g} g \tag{8}$$

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:

$$\tilde{p}(\tau) = \begin{cases} \tau p^U, & 0 \leq \tau \leq 1, \\ p^U + (\tau - 1)(p^B - p^U), & 1 \leq \tau \leq 2. \end{cases} \tag{9}$$

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. 4a and 4b 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*

$$\|g(x) - g(\bar{x})\| \le L\|x - \bar{x}\|, \quad \textit{for all } x, \bar{x} \in \mathcal{N}. \tag{10}$$

Then

$$\sum_{k \ge 0} \cos^2\theta_k \, \|g(x_k)\|^2 < \infty, \tag{11}$$

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 $g_0 \leftarrow \nabla f_0, p_0 \leftarrow -g_0, k \leftarrow 0$;
  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

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

$$F^k(x) = f(x) + \frac{k}{2}\|c(x)\|^2 + \frac{\alpha}{2}\|x - x^*\|^2, \tag{12}$$

where $\alpha > 0$ is arbitrary.

G The general equality constrained optinisation problem is:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad c_i(x) = 0, i \in \mathcal{E} \tag{13}$$

H The general inequality constrained optinisation problem is:

$$\min_{x\in\mathbb{R}^n} f(x) \quad \text{subject to} \begin{cases} c_i(x) = 0 & i \in \mathcal{E}, \\ c_i(x) \geq 0 & i \in \mathcal{I} \end{cases} \tag{14}$$

I 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. H) problem and that the LICQ holds at x^* (i.e. the active constraint gradients are linearly independent). 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^*)

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, && && \text{(15a)} \\ c_i(x^*) &= 0, && \text{for all} \quad i \in \mathcal{E}, && \text{(15b)} \\ c_i(x^*) &\geq 0, && \text{for all} \quad i \in \mathcal{I}, && \text{(15c)} \\ \lambda_i^* &\geq 0, && \text{for all} \quad i \in \mathcal{I}, && \text{(15d)} \\ \lambda_i^* c_i(x^*) &= 0, && \text{for all} \quad i \in \mathcal{E} \cup \mathcal{I} && \text{(15e)} \end{aligned}$$

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

J The (revised) Simplex algorithm may be stated as:

Algorithm 2

Given $\mathcal{B}, \mathcal{N}, x_B = B^{-1}b \geq 0, x_N = 0$;
Solve $B^T\lambda = c_B$ for λ
$s_N = c_N - N^T\lambda$; (pricing)
if $s_N \geq 0$
  then STOP; (optimal point found)
fi
Select $q \in \mathcal{N}$ with $s_q < 0$ as the entering index
Solve $Bd = A_q$ for d;
if $d \leq 0$
  then STOP; (problem is unbounded)
fi
Calculate $x_q^+ = \min_{i|d_i>0}(x_B)_i/d_i$ and use p to denote the minimizing i;
Update $x_B^+ = x_B - dx_q^+, x_N^+ = (0, \ldots, 0, x_q^+, 0, \ldots, 0)^T$;
Change $\mathcal{B}$ by adding q and removing the basic variable
corresponding to column p of $\mathcal{B}$.