UNIVERSITY *of* LIMERICK
OLLSCOIL LUIMNIGH

Faculty of Science and Engineering
Department of Mathematics and Statistics

END OF SEMESTER ASSESSMENT PAPER

MODULE CODE: MS4327 | SEMESTER: Spring 2016

MODULE TITLE: Optimisation | DURATION OF EXAMINATION: $2\frac{1}{2}$ hours

LECTURER: Dr. J. Kinsella | GRADING SCHEME: 20% +80%

EXTERNAL EXAMINER: Prof. J. King

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

1 (a) Prove Zoutendijk's Theorem (Theorem 1 in Appendix C). 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 C. 8

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

2 (a) Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 and p_k is a descent direction at x_k . Assume that f is bounded below along the ray $\{x_k + \alpha p_k | \alpha > 0\}$. Show that if $0 < c_1 < c_2 < 1$ there exist step lengths α satisfying the Wolfe conditions; Eqs. 3a and 3b in Appendix A. 12

(b) Draw a sketch to illustrate your proof. 4

(c) Consider the Backtracking Line Search algorithm (Algorithm 1 in Appendix D). Prove that provided at least one iteration takes place, then for ρ sufficiently close to 1 (i.e. provided the back-tracking is sufficiently slow), Algorithm 1 will produce an interval $I = [\alpha_1, \alpha_2]$ such that the Wolfe conditions Eqs. 3a and 3b are satisfied for all $\alpha \in I$ with $c_1 = c$ (from Algorithm 1) and for some $c_2 > c_1$. 9

3 Let $f : \mathbb{R}^n \to \mathbb{R}$ be C^3 and let the point x^* satisfy $g(x^*) \equiv \nabla f(x^*) = \mathbf{0}$ and $\nabla^2 f(x^*) \equiv H(x^*)$ be positive definite.

(a) Derive the Newton direction $\mathbf{p}_k = -H^{-1}(\mathbf{x}_k)\mathbf{g}(\mathbf{x}_k)$ by minimising $m_k(\mathbf{p})$, the second-order Taylor series approximation to $f(\mathbf{x_k} + \mathbf{p})$ with respect to $\mathbf{p}$ where $m_k(\mathbf{p}) \equiv f_k + \mathbf{p}^T \mathbf{g_k} + \frac{1}{2}\mathbf{p}^T H(x_k)\mathbf{p}$. 2

(b) Show that if at any point $\mathbf{x}_k$, $H(\mathbf{x}_k)$ is positive definite then the Newton direction $\mathbf{p}_k^N = -H^{-1}(\mathbf{x}_k)\mathbf{g}(\mathbf{x}_k)$ is a descent direction, i.e. that $\mathbf{p}_k^T \mathbf{g}(\mathbf{x}_k) < 0$. 2

(c) Find an explicit negative upper bound for $\mathbf{p}_k^T \mathbf{g}(\mathbf{x}_k)$ in this case. 3

(d) If $H(x_k)$ is not positive definite, describe a simple strategy for obtaining a descent direction by "shifting" the eigenvalues of $H(x_k)$. 2

(e) Let $R = \|x_0 - x^*\|$ be the distance from the start point x_0 to x^*. Assume that R is small enough for the condition $R < \frac{1}{k_1 k_2}$ to hold for some $k_1, k_2 > 0$ defined by the two conditions:

- the inverse Hessian is bounded in norm: $\|H^{-1}(x)\| \leq k_1$
- the remainder term in a second-order Taylor expansion of $g(x^*)$ centred at x is uniformly bounded: $\|g(x^*) - g(x) - H(x)(x^* - x)\| \leq k_2 \|x^* - x\|^2$

for each x inside the ball $B = \{\|x - x^*\| \leq R\}$.

Show that, under these assumptions,

(i) the sequence of vectors $x_{k+1} = x_k + p_k$ with $p_k = -H^{-1}(x_k) g(x_k)$ generated by Newton's method converges to x^*. 12

(ii) the convergence is quadratic. 2

(f) What are the drawbacks of Newton's method compared to other gradient search methods such as Conjugate Gradient and Quasi-Newton methods? 2

4 Derive the DFP Quasi-Newton update formula for $H_k \approx \nabla^2 f(x_k)$

$$\text{DFP}: \quad H_{k+1} = (I - \gamma_k y_k s_k^T) H_k (I - \gamma_k s_k y_k^T) + \gamma_k y_k y_k^T, \qquad (1)$$

where $s_k = \alpha_k p_k$, $y_k = g_{k+1} - g_k$ and $\gamma_k = \frac{1}{y_k^T s_k}$ by showing that H_{k+1} is a solution to the problem $\min_H \|H - H_k\|$ subject to $H = H^T, H s_k = y_k$. (As usual, $g_k \equiv \nabla f(x_k)$.)

You may use the following sequence of steps:

(a) Define $K = I - \gamma_k s_k y_k^T$ and $P = \gamma_k y_k y_k^T$. Show that $K s_k = 0$, $K^T y_k = 0$ and $P s_k = y_k$. 2

(b) Show that any matrix of the form $H = AK + P$ satisfies $H s_k = y_k$ — where A is an arbitrary $n \times n$ matrix. 1

(c) Explain why requiring that H be symmetric tells us that H must take the form $H = K^T A K + P$ — where A is now an arbitrary $n \times n$ symmetric matrix. 1

(d) Show that a positive definite symmetric matrix W may be found s.t. $W y_k = s_k$. 4

(e) Use the weighted Frobenius norm $\|C\|_W \equiv \|W^{\frac{1}{2}} C W^{\frac{1}{2}}\|_F$, where $\|C\|_F^2 \equiv \sum_{i=1}^n \sum_{j=1}^n C_{ij}^2$ for any real square matrix C. Take W as in (d). Show that for any symmetric matrix C, $\|C\|_W^2 = \text{Trace}(WCWC)$. 5

(See part (f) on next page.)

(f) Write $C \equiv H - H_k = K^T A K + P - H_k$ and show that the choice $A = H$ minimises $F(A) \equiv \|H - H_k\|_W^2$. (You can assume that $KWK^T = KW$.) 12

5 Consider the general equality-constrained problem (10) as defined in Appendix G.

(a) Show that the penalty function $F^k(x)$ defined in (9) in Appendix 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 (12a) and (12b) in Appendix 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 (11) in Appendix 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 $\lambda_i^* = -\lim_{k\to\infty} k c_i^-(x_k), i \in \mathcal{I}$. 5

6 This question relates to the application of the KKT necessary conditions ((12a)–(12e) in Appendix I) to the Linear Program (LP) in Standard Form (SF)

$$\min c^T x, \text{ subject to } Ax = b,\ x \geq 0. \tag{2}$$

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. 2 above) take the form: 5

$$A^T \lambda + s = c \tag{2a}$$
$$Ax = b \tag{2b}$$
$$x \geq 0 \tag{2c}$$
$$s \geq 0 \tag{2d}$$
$$x_i s_i = 0,\ i = 1, \dots, n. \tag{2e}$$

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

$$\max b^T \lambda, \text{ subject to } A^T \lambda \leq c. \tag{3}$$

(when the dual LP is re-fornulated into SF) are the same as those for the primal problem (2). 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 (2) above **if it is feasible** and if there exists a subset $\mathcal{B}$ of the index set $\{1, 2, \dots, n\}$ 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)

Show that if we partition the vectors x, s and c according to the index sets $\mathcal{B}$ and $\mathcal{N} = \{1, 2, \dots, n\} \backslash \mathcal{B}$, 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 (Algorithm 2 in Appendix J) may be derived from the KKT conditions (2a)–(2e). 13

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

Appendices of Results

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

$$f(x_k + \alpha p_k) \leq f(x_k) + c_1 \alpha {p_k}^T g(x_k), \tag{3a}$$
$$ {p_k}^T g(x_k + \alpha p_k) \geq c_2 {p_k}^T g(x_k), \tag{3b}$$

where $g(x) \equiv \nabla f(x)$ and $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (3b) by

$$|{p_k}^T g(x_k + \alpha p_k)| \leq c_2 |{p_k}^T g(x_k)|. \tag{4}$$

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

$$\phi(\alpha) \leq \phi(0) + c_1 \alpha \phi'(0), \tag{4a}$$
$$\phi'(\alpha) \geq c_2 \phi'(0), \tag{4b}$$

where $0 < c_1 < c_2 < 1$. The **strong** Wolfe conditions replace (4b) by

$$|\phi'(\alpha)| \leq c_2 |\phi'(0)|. \tag{5}$$

C **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. 3a and 3b 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 sub-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*

$$\|g(x) - g(\bar{x})\| \leq L \|x - \bar{x}\|, \quad \text{for all } x,\ \bar{x} \in \mathcal{N}. \tag{6}$$

Then

$$\sum_{k \geq 0} \cos^2 \theta_k \, \|g(x_k)\|^2 < \infty, \tag{7}$$

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

D Backtracking Line Search Algorithm

Algorithm 1 (Backtracking Line Search)

begin
  Choose $\bar{\alpha} > 0$ and $\rho, c \in (0, 1)$
  $\alpha := \bar{\alpha}$
  while $\phi(\alpha) \geq \phi(0) + c\alpha\phi'(0)$ do
    $\alpha := \rho\alpha$
  end
  $\alpha_k := \alpha$
end

E A sequence $\{x_k\}_{k=1}^{\infty}$ that converges to a point x^* has quadratic convergence if for some positive constant M

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} \leq M, \quad \text{for all } k \text{ sufficiently large.} \tag{8}$$

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

$$F^k(x) = f(x) + \frac{k}{2}\|c(x)\|^2 + \frac{\alpha}{2}\|x - x^*\|^2, \tag{9}$$

where $\alpha > 0$ is arbitrary.

G The general equality constrained optimisation problem is:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad c_i(x) = 0, i \in \mathcal{E}. \tag{10}$$

H The general inequality constrained optimisation problem is:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \begin{cases} c_i(x) = 0 & i \in \mathcal{E}, \\ c_i(x) \geq 0 & i \in \mathcal{I}. \end{cases} \tag{11}$$

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 Appendix 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^*):

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{12a}$$
$$c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E}, \tag{12b}$$
$$c_i(x^*) \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{12c}$$
$$\lambda_i^* \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{12d}$$
$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E} \cup \mathcal{I} \tag{12e}$$

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 $\mathcal{B}$.