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 2015

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 Appendix at the end of the paper for some useful results.**

1 (a) Suppose that a function $f : \mathbb{R}^n \to \mathbb{R}$ is C^1 on $\mathbb{R}^n$ and has gradient $\nabla f(x) \equiv g(x)$. If a vector $\mathbf{p}$ satisfies the condition $\mathbf{p}^\mathbf{T}\mathbf{g}(\mathbf{x}) < \mathbf{0}$, $\mathbf{p}$ is called a "descent direction" for f at x. Explain carefully the importance of this definition for optimisation algorithms, justifying any statement that you make. 4

(b) 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. 10a and 10b in Appendix A. 12

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

(d) Show that if $c_1 > \dfrac{1}{2}$, a line search satisfying the Wolfe conditions would exclude the minimiser of a strictly convex quadratic objective function so the restriction $c_1 \le \dfrac{1}{2}$ is reasonable. 5

Hint: write $\phi(\alpha) = \phi_0 + \alpha\phi_0' + \frac{1}{2}\alpha^2\phi_0''$, find the α–range determined by the first Wolfe condition and compare with α^*, the α–value that minimises $\phi(\alpha)$. Take $\phi_0'' > 0$ as given f strictly convex.

2 (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 $\|B_k\|\|B_k^{-1}\| \le M$ for all k then $\cos\theta_k \ge 1/M$ 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

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 C) and the result that

$$-\frac{1}{1-c_2} \le \frac{p_k^T g_k}{\|g_k\|^2} \le \frac{2c_2-1}{1-c_2}, \text{for all } k = 0, 1, \dots \qquad (1)$$

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

(i) First use Eq. 1 above to show that

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

3

(ii) Then use Zoutendijk's Theorem to show that $\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|p_k\|^4}$ converges. 3

(iii) Use the strong version of the second Wolfe condition (Eq. 11 in App. A) and the bound on $\cos\theta_k$ above to show that $|g_k^T p_{k-1}| \le \frac{c_2}{1-c_2}\|g_{k-1}\|^2$. 4

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

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

with $c_3 = \frac{1+c_2}{1-c_2}$. 4

(v) 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 $\bar{\gamma}$. 9

(b) Explain the significance of the result. 2

4 (a) Prove the Sherman-Morrison-Woodbury formula for the inverse of a matrix $A + \Delta A$ — stated in Appendix F. 6

(b) Given the DFP update formula:

$$\textbf{DFP}: \quad H_{k+1} = (I - \gamma_k y_k s_k^T) H_k (I - \gamma_k s_k y_k^T) + \gamma_k y_k y_k^T, \qquad (2)$$

where

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

show that the DFP-update formula may be written as $H_{k+1} = H_k + \Delta H_k$, with 6

$$\Delta H_k = RSR^T,$$

where

$$R = \begin{bmatrix} y_k & Hs_k \end{bmatrix}, \quad S = \gamma_k \begin{bmatrix} 1 + \gamma_k s_k^T H_k s_k & -1 \\ -1 & 0 \end{bmatrix}.$$

(c) Apply the SMW formula from part (a) to the formula $H_{k+1} = H_k + \Delta H_k$ from part (b) to derive the following equation for the update of the inverse Hessian approximation, J_k , that corresponds to the DFP update of H_k in Eq. 2; 13

$$\textbf{DFP} - \textbf{Inverse} \quad J_{k+1} = J_k - \frac{J_k y_k y_k^T J_k}{y_k^T J_k y_k} + \frac{s_k s_k^T}{y_k^T s_k}. \qquad (3)$$

Remember that for any 2×2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{(ad - bc)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

5 Consider the general equality-constrained problem (18) as defined in Appendix H.

(a) Show that the penalty function $F^k(x)$ defined in (17) in Appendix G 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 (20a) and (20b) in Appendix J 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 (19) in Appendix I. 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 (a) Consider the equality-constrained Quadratic Program (QP):

$$\min_x q(x) = \frac{1}{2} x^T Q x + x^T \mathbf{d}, \tag{4a}$$
$$\text{subject to} \quad a_i^T x = b_i, \quad i = 1, \ldots, k \tag{4b}$$

Let A be the matrix s.t. the vectors $\{a_i\}_{i \in \mathcal{E}}$ are the columns of A^T.

(i) Writing the set of k equality constraints (4b) 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 2

$$Q\mathbf{x} + \mathbf{d} - \sum_{i \in \mathcal{E}} \lambda_i^* \mathbf{a}_i = \mathbf{0}. \tag{5}$$

(ii) Writing $x^* = x_0 + p$, where x_0 is any estimate of the solution and p the required step to the solution, show that (6) and the constraints $Ax - b = 0$ can be re-written as: 3

$$AQ^{-1}A^T\lambda = AQ^{-1}\mathbf{g} - \mathbf{r}, \tag{6}$$
$$Q\mathbf{p} = A^T\lambda - \mathbf{g}, \tag{7}$$

where the residual $r = Ax_0 - b$ and $g = Qx_0 + d$ (the gradient of $q(x_0)$). Explain how the two equations can be used to solve for λ and $\mathbf{p}$.

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

(b) Consider the **inequality**-constrained QP:

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

with $q(x) = \frac{1}{2}x^T Q x + d^T x$ where $Q = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ and $\mathbf{d} = \begin{bmatrix} -2 \\ -5 \end{bmatrix}$ with $A = \begin{bmatrix} 1 & -2 \\ -1 & -2 \\ -1 & 2 \\ 1 & 2 \end{bmatrix}$ and $b = \begin{bmatrix} -2 \\ -6 \\ -2 \\ 2 \end{bmatrix}$.

(i) Sketch the diamond-shaped feasible region, labelling the constraints 1–4 in order of their appearance in A and b, e.g. constraint #1 is the region **below** the line $x - 2y = -2$ and constraint #3 is the region **above** the line $-x + 2y = -2$. 2

(ii) Starting at the vertex $x_0 = (2, 0)$ where constraints #3 and #4 are binding/active, solve the problem by solving a succession of equality-constrained QPs

$$\min_{x \in \mathbb{R}^2} q(x) \text{ subject to } \hat{A}x = \hat{b}, \tag{9}$$

where $\hat{A}$ consists of the rows of A corresponding to the indices in the working set $\mathcal{W}$ and similarly for $\hat{b}$. Note that **r** is the residual $\hat{A}\mathbf{x} - \hat{\mathbf{b}}$.

Use the following steps:

A. Explain why **r** is the zero vector at the starting point $\mathbf{x_0}$ and calculate $g(x_0)$. (You should find that $\mathbf{g} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$.)
Take the working set $\mathcal{W} = \{3, 4\}$, use Eq. 6 in Part (a) of this question to show $\lambda = \begin{bmatrix} -9/4 \\ -1/4 \end{bmatrix}$ (take $\left(\hat{A}Q^{-1}\hat{A}^T\right)^{-1} = \frac{1}{8}\begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}$).
Give a simple argument as to why the step **p** **must** be the zero vector. 4

B. Remove the index corresponding to the most negative multiplier from $\mathcal{W}$, update $\hat{A}$ and $\hat{b}$ and resolve for λ and **p** using Eqs. 6 and 7 from Part (a) of this question. (You should find $\lambda = -8/5$ and $\mathbf{p} = \begin{bmatrix} -9/5 \\ 9/10 \end{bmatrix}$.) 3

C. Determine the steplength α using the formula

$$\alpha \equiv \min\left(1, \min_{i \in \mathcal{W}^c, a_i^T p < 0} \frac{b_i - a_i^T x}{a_i^T p}\right). \tag{10}$$

You should find that $\alpha = \min\{1, 10/9\} = 1$ so no constraint is added to $\mathcal{W}$. 3

D. Update x to $x + \alpha p$ and recalculate $\mathbf{r}$ and $\mathbf{g}$ at the new point $\mathbf{x} = \begin{bmatrix} 1/5 \\ 9/10 \end{bmatrix}$. (You should find that r is still zero and that $\mathbf{g} = \begin{bmatrix} -8/5 \\ -16/5 \end{bmatrix}$.) 3

E. At the next step (**do not perform the calculation**) λ and $\mathbf{p}$ are found to be $\lambda = -8/5$ and $\mathbf{p} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Explain why you should drop the (only) constraint from the working set. Solve the **unconstrained** problem $\min \frac{1}{2} p^T Q p + \mathbf{g}^\mathbf{T}\mathbf{p}$. You should find $\mathbf{p} = \begin{bmatrix} 4/5 \\ 8/5 \end{bmatrix}$. 2

F. At a later iteration, $\mathcal{W} = \{1\}$ and $\mathbf{x} = \begin{bmatrix} 7/5 \\ 17/10 \end{bmatrix}$, $\lambda = 4/5$ and $\mathbf{p} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Explain carefully why the first-order KKT conditions for the inequality-constrained problem Eq. 8 are now satisfied. 2

G. Why is it obvious that the second-order conditions are also satisfied? 1

Appendix of Results

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

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

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 α 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 **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. 10a and 10b 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*

$$\|g(x) - g(\bar{x})\| \leq L\|x - \bar{x}\|, \quad \text{for all } x,\ \bar{x} \in \mathcal{N}. \tag{13}$$

Then

$$\sum_{k \geq 0} \cos^2 \theta_k \, \|g(x_k)\|^2 < \infty, \tag{14}$$

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

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

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} \leq M, \quad \text{for all } k \text{ sufficiently large.} \tag{15}$$

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 The Sherman-Morrison-Woodbury (SMW) formula states that if a square non-singular matrix A is updated by

$$A' = A + \Delta A, \quad \text{where } \Delta A = RST^T$$

where R, T are $n \times p$ matrices for $1 < p < n$ and S is $p \times p$ then

$$A'^{-1} \equiv (A + \Delta A)^{-1} = A^{-1} - A^{-1} R U^{-1} T^T A^{-1}, \tag{16}$$

where $U = S^{-1} + T^T A^{-1} R$.

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**

$$F^k(x) = f(x) + \frac{k}{2}\|c(x)\|^2 + \frac{\alpha}{2}\|x - x^*\|^2, \tag{17}$$

where $\alpha > 0$ is arbitrary.

H The general equality constrained optimisation problem is:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad c_i(x) = 0, i \in \mathcal{E}. \tag{18}$$

I The general inequality constrained optimisation problem is:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \begin{cases} c_i(x) = 0 & i \in \mathcal{E}, \\ c_i(x) \geq 0 & i \in \mathcal{I}. \end{cases} \tag{19}$$

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^* (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^*):

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = 0, \tag{20a}$$
$$c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E}, \tag{20b}$$
$$c_i(x^*) \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{20c}$$
$$\lambda_i^* \geq 0, \quad \text{for all} \quad i \in \mathcal{I}, \tag{20d}$$
$$\lambda_i^* c_i(x^*) = 0, \quad \text{for all} \quad i \in \mathcal{E} \cup \mathcal{I} \tag{20e}$$

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

K 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 $\mathcal{B}$.