Master - June 08

E4.29/ISE 4.55/C1.1

Optimisation - Model answers 2008

(Note to external examiners: all questions involve mostly applications of standard methods and concepts to unseen examples.)

Question 1

a) The stationary points of the function f are computed by solving the equations

$$0 = \nabla f = \begin{bmatrix} x_1^3 - y \\ x_2^3 - y \\ \vdots \\ x_n^3 - y \\ -x_1 - x_2 - \cdots - x_n + ny \end{bmatrix}.$$

The first n equations yield $x_i = y^{1/3}$, hence the last equation becomes

$$0 = -ny^{1/3} + ny = n(y - y^{1/3}).$$

The solutions of this equation are $y = 0$, $y = 1$ and $y = -1$. In summary, the function f has three stationary points

$$P_a = (0, \cdots, 0, 0) \qquad P_b = (1, \cdots, 1, 1) \qquad P_c = (-1, \cdots, -1, -1).$$

b) Note that

$$\nabla^2 f = \begin{bmatrix} 3x_1^2 & 0 & \cdots & 0 & -1 \\ 0 & 3x_2^2 & \cdots & 0 & -1 \\ 0 & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 3x_n^2 & -1 \\ -1 & -1 & \cdots & -1 & n \end{bmatrix}.$$

Hence

$$\nabla^2 f(P_a) = \begin{bmatrix} 0 & 0 & \cdots & 0 & -1 \\ 0 & 0 & \cdots & 0 & -1 \\ 0 & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & -1 \\ -1 & -1 & \cdots & -1 & n \end{bmatrix},$$

which is an indefinite matrix, hence P_a is a saddle point. Finally,

$$\nabla^2 f(P_b) = \nabla^2 f(P_c) = \begin{bmatrix} 3I & -v \\ -v' & n \end{bmatrix},$$

where $v' = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}$. Exploiting the relation

$$\begin{bmatrix} I & 0 \\ v'/3 & 1 \end{bmatrix} \begin{bmatrix} 3I & -v \\ -v' & n \end{bmatrix} \begin{bmatrix} I & v/3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3I & 0 \\ 0 & 2/3n \end{bmatrix},$$

we conclude that P_b and P_c are local minimizers.

c) The function f can be written as

$$f = \frac{1}{4}(x_1^2 - 1)^2 + \cdots + \frac{1}{4}(x_n^2 - 1)^2 + \frac{1}{2}(x_1 - y)^2 + \cdots + \frac{1}{2}(x_n - y)^2 - \frac{n}{4}.$$

Hence $f + n/4$ is a *sum of squares*, and all variables $x_1, x_2, \cdots, x_n, y$ are present in one of the squares. As a result the function is radially unbounded and the local minimum of f is also a global minimum. Note that

$$f(P_b) = f(P_c) = -\frac{n}{4} < 0,$$

hence both P_b and P_c are global minimizers.

d) The direction from P_p to P_m is

$$d = P_m - P_p = -2 \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}.$$

The function f along the direction d at P_p is given by

$$\phi(\alpha) = f(1 - 2\alpha, \cdots, 1 - 2\alpha, 1 - 2\alpha) = \frac{n}{4}(1 - 2\alpha)^4 - \frac{n}{2}(1 - 2\alpha^2) = -\frac{n}{4} + 4n\alpha^2 + \cdots$$

Note that $\phi(0) = -n/4$ and that $\phi(\alpha) > -n/4$ for $\alpha > 0$ and sufficiently small, hence d is an ascent direction for f at P_p.

Question 2

a) Setting $x_{-1} = x_0$ yields

$$\begin{array}{lll} k=0 & \Rightarrow & x_1 = x_0 - \alpha \nabla f(x_0) \\ k=1 & \Rightarrow & x_2 = x_1 - \alpha \nabla f(x_1) + \beta(x_1 - x_0) = x_1 - \alpha(\nabla f(x_1) + \beta \nabla f(x_0)) \\ k=2 & \Rightarrow & x_3 = x_2 - \alpha \nabla f(x_2) + \beta(x_2 - x_1) = x_2 - \alpha(\nabla f(x_2) + \beta \nabla f(x_1) + \beta^2 \nabla f(x_0)) \end{array}$$

from which we deduce the general expression

$$x_{k+1} = x_k - \alpha \left(\nabla f(x_k) + \beta \nabla f(x_{k-1}) + \beta^2 \nabla f(x_{k-2}) + \cdots + \beta^k \nabla f(x_0) \right).$$

b) i) For the considered function the gradient algorithm with constant α is described by the iteration

$$\begin{aligned} x_{1,k+1} &= x_{1,k} - \alpha(4x_{1,k}) &&= (1-4\alpha)x_{1,k}, \\ x_{2,k+1} &= x_{2,k} - \alpha(x_{2,k}) &&= (1-\alpha)x_{2,k}. \end{aligned}$$

The sequences $\{x_{1,k}\}$ and $\{x_{2,k}\}$ converge to 0 if, and only if,

$$-1 < 1 - 4\alpha < 1 \qquad -1 < 1 - \alpha < 1$$

which is equivalent to $\alpha \in (0, 1/2)$.
Setting $\alpha = 1/4$ yields

$$x_{1,k+1} = 0 \qquad x_{2,k+1} = \frac{3}{4} x_{2,k},$$

hence $x_{1,k} = 0$, for all $k \geq 1$.
To determine the speed of convergence note that we can consider only the sequence $\{x_{2,k}\}$, which is such that (recall that the sequence converges to 0)

$$\frac{x_{2,k+1}}{x_{2,k}} = \frac{3}{4},$$

which shows linear speed of convergence.

ii) For the considered function and under the stated conditions the heavy ball algorithm is described by the iteration

$$\begin{aligned} x_{1,k+1} &= x_{1,k} - \alpha(4x_{1,k}) + \beta(x_{1,k} - x_{1,k-1}), \\ x_{2,k+1} &= x_{2,k} - \alpha(x_{2,k}) + \beta(x_{2,k} - x_{2,k-1}). \end{aligned}$$

The first of the equations above, the condition $x_{1,0} = x_{1,-1}$, and $\alpha = 1/4$ imply $x_{1,1} = 0$ and $x_{1,k} = 0$, for all $k \geq 1$.
The second of the equations above, and the results in part a), yield

$$x_{2,k+1} = x_{2,k} - \frac{1}{4} \left(x_{2,k} + \frac{3}{4} x_{2,k-1} + \ldots\ldots \right).$$

Hence

$$\begin{aligned} & x_{2,1} = \frac{3}{4} x_{2,0}, \\ & x_{2,2} = x_{2,1} - \frac{1}{4}(x_{2,1} + 3/4 x_{2,0}) = \frac{1}{2} x_{2,1}, \\ & x_{2,3} = x_{2,2} - \frac{1}{4}(x_{2,2} + \frac{3}{4} x_{2,1} + \frac{9}{16} x_{2,0}) = 0, \\ & x_{2,4} = 0, \end{aligned}$$

which shows that the sequence generated by the heavy ball algorithm converges in finite time.

Question 3

a) The stationary points of the function f are computed solving the equations

$$0 = \nabla f = \begin{bmatrix} 2x_1(2x_1^2 - \delta x_2 + 2x_2^2) \\ 2x_2 - \delta x_1^2 - 3\delta x_2^2 + 4x_2 x_1^2 + 4x_2^3 \end{bmatrix}.$$

From the first equation we have $x_1 = 0$ or $x_1^2 = -x_2^2 + \frac{\delta}{2}x_2$. Replacing $x_1 = 0$ in the second equation yields

$$0 = x_2(2 - 3\delta x_2 + 4x_2^2).$$

Replacing $x_1^2 = -x_2^2 + \frac{\delta}{2}x_2$ in the second equation yields

$$0 = -\frac{1}{2}x_2(\delta - 2)(\delta + 2).$$

In conclusion the function f has the following stationary points.

- $P_0 = (0, 0)$, for any value of δ.
- $P_1 = (0, \frac{3\delta + \sqrt{9\delta^2 - 32}}{8})$ and $P_2 = (0, \frac{3\delta - \sqrt{9\delta^2 - 32}}{8})$ if $\delta^2 \geq \frac{32}{9}$. Note that if $\delta = \pm\frac{\sqrt{32}}{3}$ then $P_1 = P_2$.
- If $\delta = \pm 2$ then all points in the set $x_1^2 + x_2^2 - \frac{\delta}{2}x_2 = x_1^2 + x_2^2 \mp x_2 = 0$ are stationary points.

b) If $\delta = \frac{\sqrt{32}}{3}$ then the only stationary points are P_0 and $P_1 = P_2 = (0, \frac{\sqrt{2}}{2})$. From Figure 3.1 we conclude that P_0 is a local minimizer, and $P_1 = P_2$ is a saddle point. (The Hessian matrix is singular at P_0 and P_1, hence it cannot be used to classify these points.)

c) Note that the gradient of f on the x_2-axis is given by

$$\nabla f(0, x_2) = \begin{bmatrix} 0 \\ x_2(2 - \sqrt{32}x_2 + 4x_2^2) \end{bmatrix}.$$

The gradient of f on the x_2-axis is a direction of ascent which is parallel to the x_2-axis. Therefore, the gradient algorithm with exact line search yields the global minimizer in one step for all initial points on the x_2-axis.

d) The set of points such that the gradient algorithm with exact line search yields a sequence which converges to the global minimizer in one step is obtained eliminating α, *i.e.* the line search parameter, from the equation

$$0 = x - \alpha \nabla f(x).$$

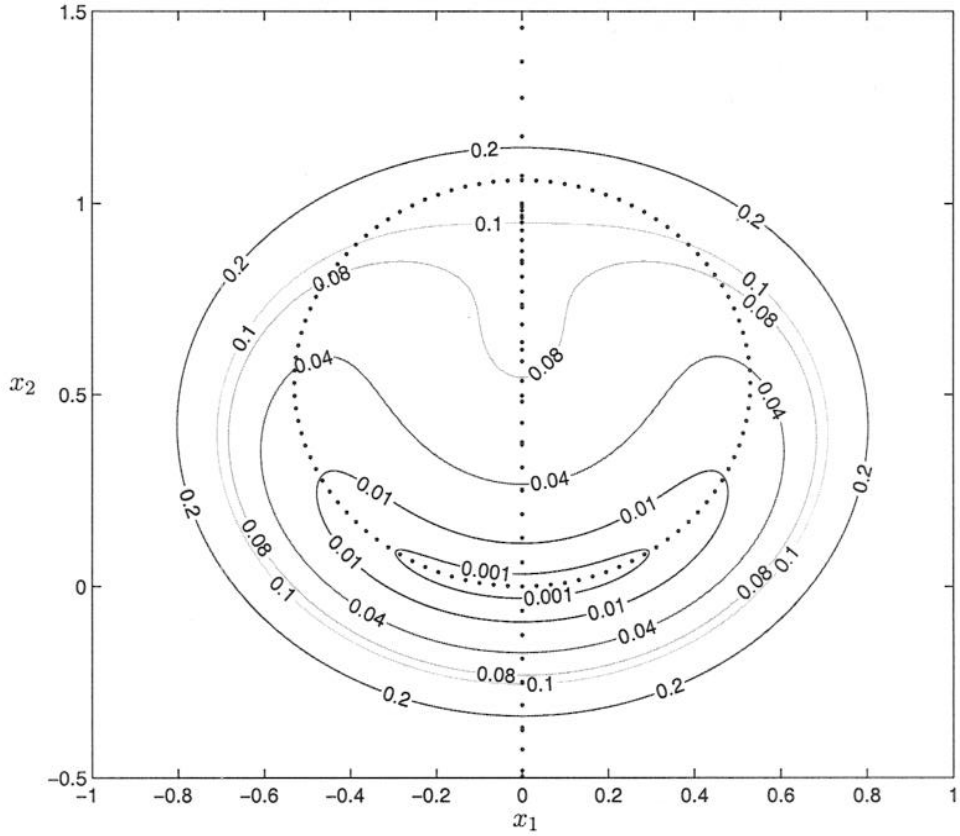
This yields the set of points described by

$$x_1(2\sqrt{2}(x_1^2 + x_2^2) - 3x_2) = 0,$$

i.e. the x_2-axis and the circle

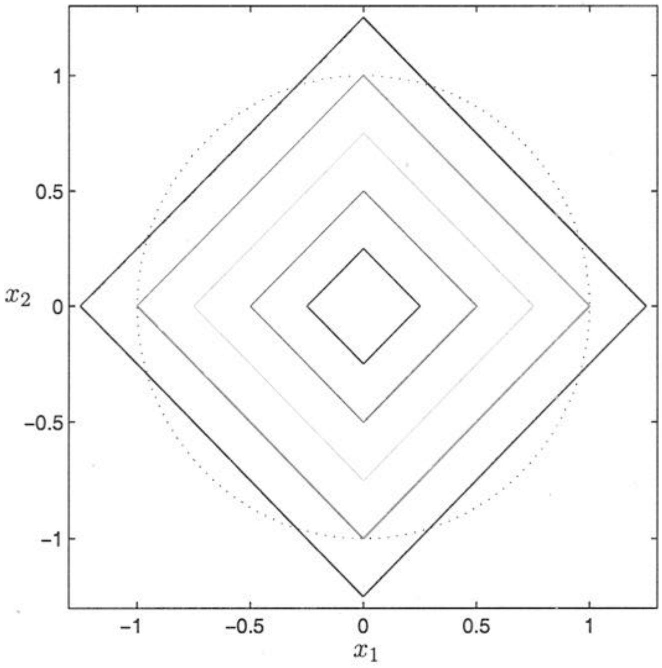
$$x_1^2 + x_2^2 - \frac{3}{4}\sqrt{2}x_2 = 0,$$

which is a circle centered at $P = (0, \frac{3}{8}\sqrt{2})$ and with radius equal to $\frac{3}{8}\sqrt{2})$. The set of all points with the required property is indicated on the figure with "dots".


1.5
1
0.5
0
−0.5
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1
x_1
x_2
0.2
0.1
0.08
0.04
0.01
0.001

Question 4

a) The admissible set is the circle of radius one and with center at $(0,0)$. The level sets of the function $|x_1| + |x_2|$ are squares with their vertices on the x_1- and x_2- axes, as indicated in the figure.



b) The solution to problem P_{min} is obtained considering the smallest square level set intersecting the admissible set. Hence there are four optimal solutions, namely the points $(0, \pm 1)$ and $(\pm 1, 0)$.
The solution to problem P_{max} is obtained considering the largest square level set intersecting the admissible set. Hence there are four optimal solutions, namely the points $(\pm\frac{\sqrt{2}}{2}, \pm\frac{\sqrt{2}}{2})$.

c) Define the Lagrangian

$$L(x_1, x_2, \lambda) = \pm(|x_1| + |x_2|) + \lambda(x_1^2 + x_2^2 - 1),$$

where the $+$ sign has to be used for P_{min} and the $-$ sign has to be used for P_{max}. The first order necessary conditions of optimality are

$$0 = \frac{dL}{dx_1} = \text{sign}(x_1) + 2\lambda x_1 \qquad 0 = \frac{dL}{dx_2} = \text{sign}(x_2) + 2\lambda x_2 \qquad x_1^2 + x_2^2 - 1 = 0$$

and a direct substitution shows that the solutions determined in part b) satisfy the necessary conditions of optimality.

d) A penalty function for problem P_{max} is

$$F_\epsilon(x_1, x_2) = -(|x_1| + |x_2|) + \frac{1}{\epsilon}(x_1^2 + x_2^2 - 1)^2.$$

The stationary points of F_ϵ are the solutions of the equations

$$0 = -\text{sign}(x_1) + \frac{4}{\epsilon}x_1(x_1^2 + x_2^2 - 1) \qquad 0 = -\text{sign}(x_2) + \frac{4}{\epsilon}x_2(x_1^2 + x_2^2 - 1).$$

If we assume that the stationary points of F_ϵ, for ϵ sufficiently small, are away from $x_1 = 0$ and from $x_2 = 0$, then the stationary points are such that

$$\frac{\text{sign}(x_1)}{x_1} = \frac{\text{sign}(x_2)}{x_2},$$

which implies $x_2 = \pm x_1$. Replacing this in the first of the equations above yields

$$0 = -\text{sign}(x_1) + \frac{4}{\epsilon}x_1(2x_1^2 - 1),$$

or equivalently

$$\frac{\epsilon}{4}\text{sign}(x_1) = x_1(2x_1^2 - 1).$$

For ϵ sufficiently small the solutions of this equations are of the form

$$x_1 = \pm\frac{\sqrt{2}}{2} + o(\epsilon).$$

As a result, the stationary points of F_ϵ are of the form

$$\left(\pm\left(\frac{\sqrt{2}}{2} + o(\epsilon)\right), \pm\left(\frac{\sqrt{2}}{2} + o(\epsilon)\right)\right),$$

i.e. they are close to the optimal solutions of the problem P_{max} for ϵ sufficiently small.

Question 5

a) Define the Lagrangian

$$L(x_1, x_2, \rho_1, \rho_2) = x_1^3 - x_1^2 x_2 + 2x_2^2 + \rho_1(-x_1) + \rho_2(-x_2).$$

The first order necessary conditions of optimality are

$$0 = \frac{dL}{dx_1} = 3x_1^2 - 2x_1 x_2 - \rho_1 \qquad 0 = \frac{dL}{dx_2} = -x_1^2 + 4x_2 - \rho_2$$

$$-x_1 \le 0 \qquad -x_2 \le 0 \qquad \rho_1 > 0 \qquad \rho_2 > 0$$

$$-x_1\rho_1 = 0 \qquad -x_2\rho_2 = 0.$$

b) Using the complementarity conditions, *i.e.* the last two conditions, we have four possibilities.

- $\rho_1 = 0$ and $\rho_2 = 0$. This yields the candidate optimal solutions $(x_1, x_2) = (0, 0)$ and $(x_1, x_2) = (6, 9)$.
- $\rho_1 = 0$ and $x_2 = 0$. This yields the candidate optimal solution $(x_1, x_2) = (0, 0)$.
- $x_1 = 0$ and $\rho_2 = 0$. This yields the candidate optimal solution $(x_1, x_2) = (0, 0)$.
- $x_1 = 0$ and $x_2 = 0$.

In summary there are two candidate optimal solutions: the point $(0, 0)$, on the boundary of the admissible set, and the point $(3, 9/2)$ in the interior of the admissible set.

c) The second order sufficient condition of optimality for the candidate point in the interior of the admissible set is

$$\nabla^2 L(3, 9/2) > 0.$$

Note that

$$\nabla^2 L(3, 9/2) = \begin{bmatrix} 9 & -6 \\ -6 & 2 \end{bmatrix},$$

and that $\det\nabla^2 L(3, 9/2) < 0$, which implies that $\nabla^2 L(3, 9/2)$ is not positive definite. Hence the candidate optimal point in the interior of the admissible set is not a local minimizer. (It is a saddle point.).

d) To show that the point $(0, 0)$ is a local minimizer note that the function f to be minimized is such that $f(0, 0) = 0$, $f(x_1, 0) > 0$ for $x_1 > 0$, and $f(0, x_2) > 0$ for $x_2 > 0$. Consider now straight lines described by $x_2 = \alpha x_1$, with $\alpha > 0$. Then

$$f(x_1, \alpha x_1) = \alpha^2 \left(\frac{1-\alpha}{\alpha^2} x_1^3 + 2x_1^2 \right),$$

which is positive for all $\alpha > 0$ and all $x_1 > 0$ and sufficiently small. Since the function f is zero at the candidate optimal solution $(0, 0)$ and strictly positive in all admissible point in a neighborhood of this point, then the point is a local minimizer.

e) The function f along the line $x_2 = 2x_1$ is given by

$$f(x_1, 2x_1) = -x_1^3 + 4x_1^2,$$

and this function is not bounded from below, *i.e.* $\lim_{x_1 \to \infty} f(x_1, 2x_1) = -\infty$. This implies that the considered optimization problem does not have a global solution.

Question 6

a) Define the Lagrangian (note the $-$ sign due to the transformation of the maximization problem into a minimization problem)

$$L(x_1, x_2, x_3, \lambda) = -(x_1x_2 + x_2x_3 + x_1x_3) + \lambda(x_1 + x_2 + x_3 - 3).$$

The first order necessary conditions of optimality are

$$0 = \frac{dL}{dx_1} = -x_2 - x_3 + \lambda \qquad 0 = \frac{dL}{dx_2} = -x_1 - x_3 + \lambda$$

$$0 = \frac{dL}{dx_3} = -x_2 - x_1 + \lambda \qquad 0 = x_1 + x_2 + x_3 - 3.$$

This is system a linear equations with the unique solution $(x_1, x_2, x_3, \lambda) = (1, 1, 1, 2)$. Hence the problem has only one candidate optimal solution.

b) Note that

$$\nabla^2 L = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}$$

and

$$\frac{\partial g}{\partial x} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}.$$

The candidate optimal solution is a minimizer if $s'\nabla^2 Ls > 0$ for all $s \neq 0$ such that $s'\frac{\partial g}{\partial x} = 0$. The set of such s's can be described by linear combinations of the vectors

$$s_1' = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \qquad s_2' = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}.$$

Note that

$$[s_1, s_2]'\nabla^2 L[s_1, s_2] = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} > 0,$$

hence the candidate optimal solution is a local minimizer.

c) An exact penalty function for a constraint optimization problem with equality constraints is

$$G(x) = f(x) - g'(x)(\frac{\partial g}{\partial x}\frac{\partial g'}{\partial x})^{-1}\frac{\partial g}{\partial x}\nabla f + \frac{1}{\epsilon}\|g(x)\|^2,$$

with $\epsilon > 0$.

i) For the considered problem we have

$$G(x_1, x_2, x_3) = -(x_1x_2 + x_2x_3 + x_1x_3) + \frac{2}{3}(x_1 + x_2 + x_3 - 3)(x_1 + x_2 + x3) + \frac{1}{\epsilon}(x_1 + x_2 + x_3 - 3)^2.$$

ii) The function is well-defined for all (x_1, x_2, x_3) since $\frac{\partial g}{\partial x}\frac{\partial g'}{\partial x}$ is a full rank matrix (it is a nonzero constant).

iii) The stationary points of the function $G(x_1, x_2, x_3)$ are the solutions of the equations

$$0 = \nabla G = \begin{bmatrix} \frac{1}{3}(4x_1 + x_2 + x_3) - 2 + \frac{2}{\epsilon}(x_1 + x_2 + x_3 - 3) \\ \frac{1}{3}(x_1 + 4x_2 + x_3) - 2 + \frac{2}{\epsilon}(x_1 + x_2 + x_3 - 3) \\ \frac{1}{3}(x_1 + x_2 + 4x_3) - 2 + \frac{2}{\epsilon}(x_1 + x_2 + x_3 - 3) \end{bmatrix}.$$

These equations have a unique solution $(x_1, x_2, x_3) = (1, 1, 1)$ which does not depend upon ϵ and coincides with the optimal solution determined in part b).