Chalmers/GU
Mathematics

EXAM SOLUTION

TMA947/MMG621
NONLINEAR OPTIMISATION

Date: 16–01–12
Examiner: Michael Patriksson

Question 1

(the simplex method)

(2p) a) We first rewrite the problem in standard form. We introduce slack variables s_1 and s_2. Consider the following linear program:

$$\begin{array}{lrcrcrcrcrcl} \text{minimize} & z = 2x_1 & - & x_2 & + & x_3 & & & & & & \\ \text{subject to} & x_1 & + & 3x_2 & - & x_3 & + & s_1 & & & = & 5, \\ & 2x_1 & - & x_2 & + & 2x_3 & & & - & s_2 & = & 2, \\ & x_1, & & x_2, & & x_3, & & s_1, & & s_2 & \geq & 0. \end{array}$$

Phase I
We introduce an artificial variable a and formulate our Phase I problem.

$$\begin{array}{lrcrcrcrcrcrcl} \text{minimize} & z = & & & & & & & & & & a & & \\ \text{subject to} & x_1 & + & 3x_2 & - & x_3 & + & s_1 & & & & & = & 5, \\ & 2x_1 & - & x_2 & + & 2x_3 & & & - & s_2 & + & a & = & 2, \\ & x_1, & & x_2, & & x_3, & & s_1, & & s_2, & & a & \geq & 0. \end{array}$$

We now have a starting basis (s_1, a). Calculating the reduced costs we obtain $\tilde{\mathbf{c}}_N = (-2, 1, -2, 1)^{\mathrm{T}}$, meaning that x_1 or x_3 should enter the basis. We choose x_3. From the minimum ratio test, we get that a should leave the basis. This concludes Phase I and we now have the basis (s_1, x_3).

Phase II
Calculating the reduced costs, we obtain $\tilde{\mathbf{c}}_N = (1, -\frac{1}{2}, \frac{1}{2})^{\mathrm{T}}$. meaning that x_2 should enter the basis. From the minimum ratio test, we get that the outgoing variable is s_1. Updating the basis we now have (x_2, x_3) in the basis.

Calculating the reduced costs, we obtain $\tilde{\mathbf{c}}_N = (\frac{7}{5}, \frac{1}{5}, \frac{2}{5})^{\mathrm{T}} \geq 0$, meaning that the current basis is optimal. The optimal solution is thus

$$(x_1, x_2, x_3, s_1, s_2)^{\mathrm{T}} = (0, \frac{12}{5}, \frac{11}{5}, 0, 0, 0)^{\mathrm{T}},$$

which in the original variables means $(x_1, x_2, x_3)^{\mathrm{T}} = (0, \frac{12}{5}, \frac{11}{5})^{\mathrm{T}}$ with optimal objective value $f^\star = -\frac{1}{5}$.

(1p) b) Calculating the reduced costs of the modified problem for the optimal basis of the original problem, we obtain $\tilde{\mathbf{c}}_N = (\frac{7}{5}, \frac{1}{5}, \frac{2}{5}, \frac{7}{10})^{\mathrm{T}} \geq 0$ meaning that the the optimal basis from the original problem gives the optimal solution of the modified problem $(x_1, x_2, x_3, x_4)^{\mathrm{T}} = (0, \frac{12}{5}, \frac{11}{5}, 0)^{\mathrm{T}}$ with optimal objective value $f^\star = -\frac{1}{5}$.

Question 2

(Quadratic programming)

Since the objective function is convex (i.e., Hessian matrix $\boldsymbol{A}$ is symmetric positive semidefinite) and the constraints are affine, the KKT conditions are both necessary and sufficient for optimality. Therefore, a point $\boldsymbol{x}$ is a minimum if and only if there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^n$ (Lagrangian multipliers) such that

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} &= \boldsymbol{\mu} \\ \boldsymbol{\mu} &\geq \boldsymbol{0}^n \\ \boldsymbol{x} &\geq \boldsymbol{0}^n \\ \boldsymbol{\mu}_i \boldsymbol{x}_i &= 0, \quad \forall\, i = 1, \dots, n. \end{aligned}$$

Eliminating $\boldsymbol{\mu}$, the above conditions are equivalent to

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} &\geq \boldsymbol{0}^n \\ \boldsymbol{x} &\geq \boldsymbol{0}^n \\ (\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})_i \boldsymbol{x}_i &= \boldsymbol{0}, \quad \forall\, i = 1, \dots, n. \end{aligned}$$

These are in turn equivalent to

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} &\geq \boldsymbol{0}^n \\ \boldsymbol{x} &\geq \boldsymbol{0}^n \\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} &= \boldsymbol{0}. \end{aligned}$$

Question 3

(characterization of convexity in C^1)

This is Theorem 3.61 (a) in the textbook.

Question 4

(true or false claims in optimization)

(1p) a) The claim is true, as stated by Proposition 9.1 in the textbook.

(1p) b) The claim is false. The point $\bar{\boldsymbol{x}}$ with $\nabla f(\bar{\boldsymbol{x}}) = \mathbf{0}^n$ can also be a local maximum or saddle point.

(1p) c) The claim is false. Consider the problem with one decision variable. $f(x) = \min\{0, -x\}$ and $g(x) = x$. The point $\bar{x} = -1$ is a constrained minimum and $g(\bar{x}) = -1 < 0$. However, removing the constraint $g(x) \leq 0$ will result in a problem whose objective value is unbounded from below.

Question 5

(KKT conditions)

(1p) a) The point $\boldsymbol{x}^* = (1, 1)^{\mathrm{T}}$ is the only feasible point and hence it must be the unique global minimum.

(2p) b) Let $g_1(\boldsymbol{x}) := x_1^2 + x_2^2 - 2$ and $g_2(\boldsymbol{x}) := (x_1 - 2)^2 + (x_2 - 2)^2 - 2$. At $\boldsymbol{x}^* = (1, 1)^{\mathrm{T}}$, both $g_1(\boldsymbol{x}^*) = 0$ and $g_2(\boldsymbol{x}^*) = 0$. That is, both inequality constraints are active. Also, it holds that

$$\nabla f(\boldsymbol{x}^*) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \nabla g_1(\boldsymbol{x}^*) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \quad \nabla g_2(\boldsymbol{x}^*) = -\begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

Therefore, the equality (as part of KKT conditions)

$$-\nabla f(\boldsymbol{x}^*) = \mu_1 \nabla g_1(\boldsymbol{x}^*) + \mu_2 \nabla g_2(\boldsymbol{x}^*), \quad \mu_1 \geq 0, \ \mu_2 \geq 0$$

cannot hold. Hence, the KKT conditions are not satisfied. As a result, the KKT conditions are not necessary for optimality since $\boldsymbol{x}^*$ is a minimum but not a KKT point. This does not contradict any result regarding the necessity of the KKT conditions. For instance, $\nabla g_1(\boldsymbol{x}^*)$ and $\nabla g_2(\boldsymbol{x}^*)$ are not linearly independent, and hence the LICQ constraint qualification does not hold. On the other hand, since the problem is convex, KKT points (if exist) are global optimal solutions.

Question 6

(Frank-Wolfe algorithm)

Figure 1 shows the feasible set of the problem (i.e., the polyhedron with thick black boundary lines) and some contours of the objective function. The optimal solution is denoted by $x^\star$ (i.e., the red dot in the figure).

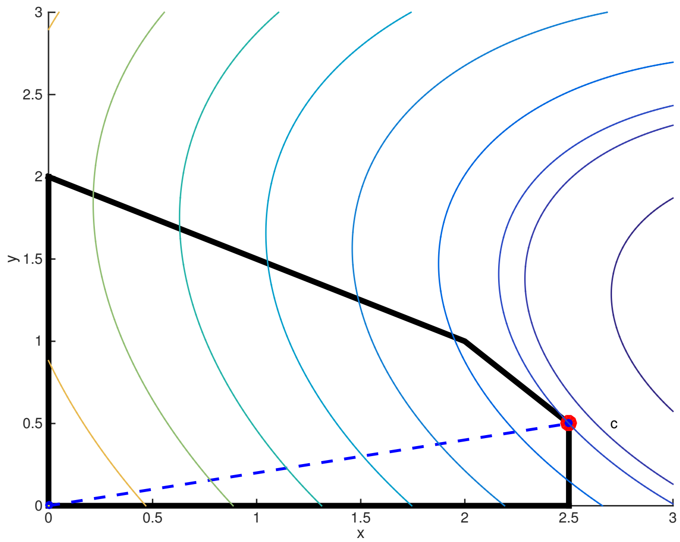

Figure 1: Illustration of the Frank-Wolfe algorithm. The feasible set is a polyhedron with boundary denoted by the thick black lines. Some contours of the objective function are shown. The optimal solution $x^\star = (2.5, 0.5)^{\mathrm{T}}$.

The details of the algorithm steps are as follows. Let X denote the feasible set. Let $f(x_1, x_2)$ denote the objective function. For any given iterate $x^{(k)} = (x_1^{(k)}, x_2^{(k)})^{\mathrm{T}}$. The objective function gradient vector is

$$\nabla f(x_1^{(k)}, x_2^{(k)}) = \begin{bmatrix} 12 & 4 \\ 4 & 18 \end{bmatrix} \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \end{bmatrix} - \begin{bmatrix} 52 \\ 34 \end{bmatrix}.$$

The search direction problem is

$$\operatorname*{minimize}_{x \in X} \quad \nabla f(x_1^{(k)}, x_2^{(k)})^{\mathrm{T}} x. \tag{1}$$

If $\min_{x \in X} \nabla f(x_1^{(k)}, x_2^{(k)})^{\mathrm{T}} x \geq \nabla f(x_1^{(k)}, x_2^{(k)})^{\mathrm{T}} x^{(k)}$, then by optimality conditions (for minimizing a convex function over a convex feasible set) $x^{(k)}$ is optimal. Otherwise, let $y^{(k)}$ denote an optimal solution to the search direction problem. Then the exact minimization line search problem can be expressed into

$$\operatorname*{minimize}_{\alpha \in [0,1]} \quad f(\alpha x^{(k)} + (1-\alpha) y^{(k)}) \iff \operatorname*{minimize}_{\alpha \in [0,1]} \quad g\alpha^2 + h\alpha,$$

where

$$\begin{aligned} g &= \left(x^{(k)} - y^{(k)}\right)^{\mathrm{T}} \begin{bmatrix} 6 & 2 \\ 2 & 9 \end{bmatrix} \left(x^{(k)} - y^{(k)}\right) \\ h &= \left(x^{(k)} - y^{(k)}\right)^{\mathrm{T}} \left(\begin{bmatrix} 12 & 4 \\ 4 & 18 \end{bmatrix} y^{(k)} - \begin{bmatrix} 52 \\ 34 \end{bmatrix} \right). \end{aligned} \tag{2}$$

The minimizing value of α, denoted by $\alpha^{(k)}$, can be found using the optimality condition to be

$$\alpha^{(k)} = \begin{cases} 0 & \text{if} \quad -\frac{h}{2g} < 0 \\ -\frac{h}{2g} & \text{if} \quad 0 \leq -\frac{h}{2g} \leq 1 \\ 1 & \text{if} \quad -\frac{h}{2g} > 1 \end{cases}. \tag{3}$$

The iterate update formula is

$$x^{(k+1)} = \alpha^{(k)} x^{(k)} + (1 - \alpha^{(k)}) y^{(k)}. \tag{4}$$

Now we begin applying the Frank-Wolfe algorithm. At the first iteration with $x^{(0)} = (0, 0)$, the objective function gradient is

$$\nabla f(x_1^{(0)}, x_2^{(0)}) = \begin{bmatrix} 12 & 4 \\ 4 & 18 \end{bmatrix} \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \end{bmatrix} - \begin{bmatrix} 52 \\ 34 \end{bmatrix} = \begin{bmatrix} -52 \\ -34 \end{bmatrix}.$$

To solve the search direction problem in (1), it is sufficient to restrict the feasible set to the set of all extreme points. That is,

$$\underset{x \in V}{\text{minimize}} \quad \nabla f(x_1^{(0)}, x_2^{(0)})^{\mathrm{T}} x, \tag{5}$$

where V is the set of all extreme points defined as

$$V = \left\{ (0,0)^{\mathrm{T}}, (0,2)^{\mathrm{T}}, (2,1)^{\mathrm{T}}, (2.5, 0.5)^{\mathrm{T}}, (2.5, 0)^{\mathrm{T}} \right\}.$$

This amounts to finding the minimum among five numbers: 0, -68, -138, -147, -130. The result is that $y^{(0)} = (2.5, 0.5)^{\mathrm{T}}$. Applying the formula in (2) yields

$$\begin{aligned} g &= \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.5 \\ 0.5 \end{bmatrix} \right)^{\mathrm{T}} \begin{bmatrix} 6 & 2 \\ 2 & 9 \end{bmatrix} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.5 \\ 0.5 \end{bmatrix} \right) &= 44.75 \\ h &= \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 2.5 \\ 0.5 \end{bmatrix} \right)^{\mathrm{T}} \left(\begin{bmatrix} 12 & 4 \\ 4 & 18 \end{bmatrix} \begin{bmatrix} 2.5 \\ 0.5 \end{bmatrix} - \begin{bmatrix} 52 \\ 34 \end{bmatrix} \right) &= 57.5 \end{aligned}$$

According to (3), $\alpha^{(0)} = 0$. Hence, by (4)

$$x^{(1)} = y^{(0)} = (2.5, 0.5)^{\mathrm{T}}.$$

This is shown in Figure 1.

At the next iteration with $x^{(1)} = (2.5, 0.5)^{\mathrm{T}}$, we have

$$\nabla f(x_1^{(1)}, x_2^{(1)}) = \begin{bmatrix} -20 \\ -15 \end{bmatrix}.$$

Solving (5) leads to $y^{(1)} = x^{(1)} = (2.5, 0.5)^{\mathrm{T}}$. Thus, it holds that

$$\min_{x \in X} \nabla f(x_1^{(1)}, x_2^{(1)})^{\mathrm{T}} x \geq \nabla f(x_1^{(1)}, x_2^{(1)})^{\mathrm{T}} x^{(1)}.$$

By optimality conditions, $x^{(1)} = (2.5, 0.5)^{\mathrm{T}}$ is the optimal solution to our problem.

Question 7

(LP duality)

Since $P = \{\boldsymbol{y} \mid \boldsymbol{A}\boldsymbol{y} \geq \boldsymbol{b},\ \boldsymbol{y} \geq \boldsymbol{0}^n\}$ is assumed to be nonempty and bounded, strong duality implies that, for any fixed $\boldsymbol{x}$, the minimum objective value of

$$\begin{aligned} \inf_{\boldsymbol{y}} \quad & \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{y} \geq \boldsymbol{b} \\ & \boldsymbol{y} \geq \boldsymbol{0}^n \end{aligned}$$

is the same as the maximum objective value of

$$\begin{aligned} \sup_{\boldsymbol{z}} \quad & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{z} \\ \text{subject to} \quad & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z} \leq \boldsymbol{x} \\ & \boldsymbol{z} \geq \boldsymbol{0}^m. \end{aligned} \tag{1}$$

Substituting (1) into the original problem in the statement of Problem 7 results in

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{maximize}} \quad & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \\ \text{subject to} \quad & \sup_{\boldsymbol{z}}\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{z} \geq d \\ & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{z} \leq \boldsymbol{x} \\ & \boldsymbol{x} \geq \boldsymbol{0}^n, \boldsymbol{z} \geq \boldsymbol{0}^m. \end{aligned}$$

This problem is equivalent to the second problem in the statement of Problem 7.
