KTH Mathematics

Solutions for the exam in Optimization.
wednesday March 14, 2012, time. 14.00–19.00

Instructor: Per Enqvist, tel. 790 62 98

There may be alternative solutions to the problem.

---

1. (a) Introduce the variable x_{ik} which is one if person i is seated at table k and zero otherwise.

The total utility is then $\sum_{i=1}^{18}\sum_{k=1}^{3} x_{ik}u_{ik}$.

Every person should be seated so $\sum_{k=1}^{3} x_{ik} = 1$ for $i = 1, \cdots, 18$.

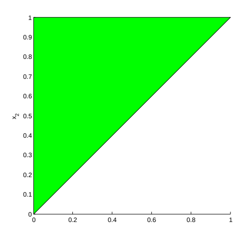
At every table there should be six persons so $\sum_{i=1}^{18} x_{ik} = 6$ for $k = 1, \cdots, 3$.

The binary character of the variables can be relaxed and we consider the following problem with non-negative real variables:

$$(TP)\quad \left[\begin{array}{ll} \min\limits_{x} & \sum_{i=1}^{18}\sum_{k=1}^{3} x_{ik}u_{ik} \\ \text{s.t.} & \\ & \sum_{k=1}^{3} x_{ik} = 1, \quad i = 1, \cdots, 18, \\ & \sum_{i=1}^{18} x_{ik} = 6, \quad k = 1, \cdots, 3. \\ x \geq 0 & \end{array}\right]$$

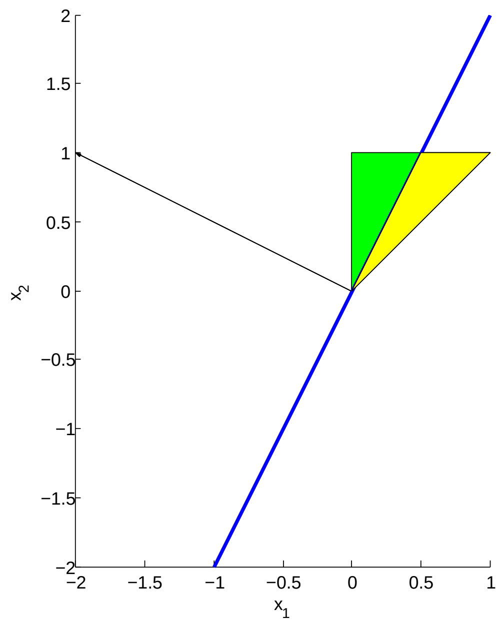
The problem can be seen as a transportation problem and can be solved using the specialized algorithm.

(b) The feasible region of the problem (P) can be described by the figure below.



So the feasible directions are vectors $d = (1, \alpha)$, where $\alpha \in (0, 1)$, or vectors βd, where $\beta > 0$.

The gradient of the objective function is c^T and is described by the arrow in the figure below. Descent directions d are directions such that $f(x+td) < f(x)$ for $t > 0$ small enough. In this linear case d must satisfy $c^T d < 0$, and vectors that satisfy this form an angle larger than 90 degrees to the vector c, i.e. lies to the right under the blue line in the figure below.



The feasible descent directions are those vectors that lie in the yellow cone in the figure above, i.e. vectors $d = (1, \alpha)$, where $\alpha \in (1/2, 1)$, or vectors βd, where $\beta > 0$. The optimum is $\hat{x} = (1, 1)$, and since the problem is convex there exists at every feasible point a feasible descent direction that points to the optimum, in particular $d = (1, 1)$ is a feasible descent direction at x^* pointing to $\hat{x}$.

(c) We should show that $c^T x - b^T y \geq 0$.

Now, using that x is feasible, then $b = Ax$, and

$$c^T x - b^T y = c^T x - (Ax)^T y = x^T c - x^T A^T y = \underbrace{x^T}_{\geq 0} \underbrace{(c - A^T y)}_{\geq 0} \geq 0$$

follows since $x \geq 0$ by primal feasibility and $c - A^T y \geq 0$ due to dual feasibility and the last inequality follows since we multiply two vectors which both have only positive elements.

2. (a) The standard form is

$$(P_s) \quad \left[\begin{array}{ll} \min\limits_{x} & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array}\right]$$

Define the vector $x = (x_{12}\ x_{13}\ x_{21}\ x_{32})$, then $c = (2\ 1\ -1\ 2)^T$. The Adjacency matrix is

$$\tilde{A} = \begin{bmatrix} 1 & 1 & -1 & 0 \\ -1 & 0 & 1 & -1 \\ 0 & -1 & 0 & 1 \end{bmatrix}, \quad \tilde{b} = \begin{bmatrix} 10 \\ -5 \\ -5 \end{bmatrix}$$

but since there are linearly dependent rows we can eliminate the last one to get

$$A = \begin{bmatrix} 1 & 1 & -1 & 0 \\ -1 & 0 & 1 & -1 \end{bmatrix}, \quad b = \begin{bmatrix} 10 \\ -5 \end{bmatrix}$$

(b) The given flow is $x = (0\ 10\ 0\ 5)$, so x_{13} and x_{32} are the basic variables corresponding to a spanning tree in the graph.

Put the node potential y_3 at node 3 to be 0. Then $y_1 - y_3 = c_{13}$ gives $y_1 = 1$. Then $y_3 - y_2 = c_{32}$ gives $y_2 = -2$.

The reduced costs are now $r_{12} = c_{12} - y_1 + y_3 = -1$ and $r_{21} = c_{21} - y_2 + y_1 = 2$. Since the reduced cost r_{12} is negative the flow in x_{12} should be increased. Increasing the flow in x_{12} to t a cycle in the graph is created and we must compensate to get $x_{32} = 5 - t$ and $x_{13} = 10 - t$. So t can become at most 5 and then x_{32} becomes zero and exits the basis.

In the new flow $x = (5\ 5\ 0\ 0)$, so x_{13} and x_{12} are the basic variables corresponding to a spanning tree in the graph.

Put the node potential y_3 at node 3 to be 0. Then $y_1 - y_2 = c_{12}$ gives $y_1 = 1$. Then $y_1 - y_3 = c_{13}$ gives $y_2 = -1$.

The reduced costs are now $r_{21} = c_{21} - y_2 + y_1 = 1$ and $r_{32} = c_{32} - y_3 + y_2 = 1$. Since the reduced costs are positive the flow is optimal.

(c) The dual problem (D) can be written

$$(D) \quad \left[\begin{array}{ll} \max\limits_{x} & b^T y \\ \text{s.t.} & A^T y \leq c \\ & y \quad \text{free} \end{array}\right]$$

which explicitly amounts to

$$(D) \quad \left[\begin{array}{ll} \max\limits_{x} & 10y_1 - 5y_2 \\ \text{s.t.} & y_1 - y_2 \leq 2 \\ & y_1 \leq 1 \\ & -y_1 + y_2 \leq -1 \\ & y_1, y_2 \quad \text{free} \end{array}\right]$$

The optimal solution is given by the previous part , the node potentials $y_1 = 1$ and $y_2 = -1$.

Complementarity:

Let $s = Ax - b = (0\ 0)^T$ and $r = A^T y - c = (0\ 0\ -1\ -1)$ then $s^T y = 0$ and $r^T x = 0$ should hold. The first holds since $s = 0$ and the second holds since the last two variables in x are zero since they are non-basic variables and the first two elements in r are zero.

3. (a) For f to be convex on the whole $\mathbb{R}^3$ it is necessary that the matrix H is positive semidefinite. Use LDL^T-factorization:

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}.$$

2 Since the last diagonal elements in the D-matrix is negative, the matrix H is not positive semi-definite and the function is not convex on $\mathbb{R}^3$.

(b) Use Gauss-Jordan to determine a basis for the nullspace of A. With the nullspace method a Z-matrix and $\bar{x}$ given by

$$Z = \begin{bmatrix} -1 \\ 1 \\ 3 \end{bmatrix}, \quad \bar{x} = \begin{bmatrix} 0 \\ 0 \\ 1/2 \end{bmatrix},$$

the equation system $(ZHZ^T)v = -Z^T(H\bar{x} + c)$:

$$\begin{bmatrix} 49 \end{bmatrix} v = -\begin{bmatrix} 7 \end{bmatrix},$$

yields $v = [-1/7]$. Therefore, $\hat{x} = \bar{x} + Zv = [1/7\ \ -1/7\ 1/14]^T$ is a global minimum since the problem is convex.

(c) The gradient of f is given by $\nabla f(x^*)^T = Hx^* + c = 22/7(1\ 1\ 0)$ and then applying row operations on

$$\begin{bmatrix} 2 & -4 & 2 \\ 6 & 0 & 2 \\ 22/7 & 22/7 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & -4 & 2 \\ 4 & 4 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

(subtracting the first row from the second and multiplying the last with 7/22) we see that the last two rows are linearly dependent.

It is not a coincidence since $\nabla f(x^*) + \lambda^T \nabla h(x^*)$ should hold at an optimal solution from the Lagrange conditions and $\nabla h(x^*) = A$.

4. (a) For the optimization problem to be convex, it is necessary that the feasible region is convex and that the objective function is convex on the whole feasible region.

The objective function is convex since it is a linear function. The first two constraints are determined by convex functions $g_1(x, y, z) = x^2 + 4y^2 - 1 \leq 0$ and $g_2(x, y, z) = x + z \leq 0$, but the third one $g_3(x, y, z) = xyz \leq 0$ does not determine a convex region. We note that $(x,y,z) = (1/2, 1/10, -1)$ and $(\bar{x},\bar{y},\bar{z}) = (-1/4, -2/10, -1/4)$ are feasible, but

$$\frac{1}{2}(x,y,z)+\frac{1}{2}(\bar{x},\bar{y},\bar{z}) = \frac{1}{2}(1/2,1/10,-1)+\frac{1}{2}(-1/4,-2/10,-1/4) = (1/8,-1/20,-5/8)$$

is not feasible since it does not satisfy the third constraint.

It is not a convex optimization problem, since the constraints does not form a convex feasible set.

(b) The gradients are given by

$$\nabla f(x) = \begin{bmatrix} -2 & 2 & -1 \end{bmatrix}, \quad \nabla g_1(x) = \begin{bmatrix} 2x & 8y & 0 \end{bmatrix}, \quad \nabla g_2(x) = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}.$$

Since constraint 3 is not active, KKT4 tells us that $y_3 = 0$ must hold. Then KKT1 becomes

$$\nabla f(x)+y_1\nabla g_1(x)+y_2\nabla g_2(x) = \begin{bmatrix} -2 & 2 & -1 \end{bmatrix}+y_1\begin{bmatrix} 2x & 8y & 0 \end{bmatrix}+y_2\begin{bmatrix} 1 & 0 & 1 \end{bmatrix} = 0,$$

which tells us that $y_2 = 1$ and $xy_1 = 1/2$ and $yy_1 = -1/4$. Clearly $x \neq 0$ and $y \neq 0$, and then $\hat{x} = \frac{1}{2y_1}$ and $\hat{y} = -\frac{1}{4y_1}$.

We know that $g_1(\hat{x},\hat{y},\hat{z}) = \hat{x}^2 + 4\hat{y}^2 - 1 = \frac{1}{4y_1^2} + 4\frac{1}{4^2y_2^2} - 1 = 0$, which tells us that $y_1 = \pm\frac{1}{\sqrt{2}}$ and from KKT3 we must have $y_1 = \frac{1}{\sqrt{2}}$. Then $\hat{x} = \frac{1}{\sqrt{2}}$ and $\hat{y} = -\frac{1}{2\sqrt{2}}$.

We know that $g_2(\hat{x},\hat{y},\hat{z}) = \hat{x} + \hat{z} = 0$, so $\hat{z} = -\frac{1}{\sqrt{2}}$.

The point $(\hat{x},\hat{y},\hat{z}) = (\frac{1}{\sqrt{2}}, -\frac{1}{2\sqrt{2}}, -\frac{1}{\sqrt{2}})$ does not satsify $g_3(\hat{x},\hat{y},\hat{z}) < 0$, so it does not satisfy all the KKT conditions and could not be a local minimum to the problem since the point is regular - the two active constraints are not linearly dependent.

5. (a) For the optimization problem to be convex, it is necessary that the feasible region is convex and that the objective function is convex on the whole feasible region.

The constraints are linear inequality constraints in p_i and therefore they form a convex feasible set.

The objective function is a sum of functions of the form $f_i(p_i)$ and the functions f_i are two times differentiable for positive values of p_i. Since, $f_i'(p_i) = 1 + \log p_i$ and $f_i''(p_i) = \frac{1}{p_i} > 0$ for positive p_i the functions f_i are convex and there sum is also convex.

So (P) is a convex optimization problem.

(b) Introduce the Lagrange function

$$L(p,y) = \sum_{i=1}^{n} p_i \log p_i + y_1\left(\sum_{i=1}^{n} p_i - 1\right) + y_2\left(\sum_{i=1}^{n} ip_i - \mu\right) + y_3\left(\sum_{i=1}^{n}(i-\mu)^2 p_i - \sigma^2\right)$$

i.e.

$$L(p,y) = \sum_{i=1}^{n}\left(p_i \log p_i + y_1 p_i + y_2 i p_i + y_3(i-\mu)^2 p_i\right) - y_1 - \mu y_2 - \sigma^2 y_3,$$

where we assume that the variables y_1, y_2, y_3 are non-negative.

We note that minimizing $L(p, y)$ over p_i for fixed y we can minimize the sum term-by-term, i.e. for each p_i individually.

The function $l_i(p_i) = p_i \log p_i + y_1 p_i + y_2 i p_i + y_3(i-\mu)^2 p_i$ is convex for positive p_i, and the derivative $l_i'(p_i) = 1 + \log p_i + y_1 + y_2 i + y_3(i-\mu)^2 = 0$ for $\hat{p}_i(y) = \exp -1 - y_1 + y_2 i + y_3(i-\mu)^2$ which will be the minimum of $L(p, y)$. Note that the "probabilities" p_i are always non-negative for any choice of y, and the optimal y is determined by solving the dual optimization problem.

Let $\varphi(y) := L(\hat{p}(y), y)$, which will be

$$L(\hat{p}(y), y) = \sum_{i=1}^{n} \big(\exp\{-1 - (y_1 + y_2 i + y_3(i-\mu)^2)\} \left[-1 - (y_1 + y_2 i + y_3(i-\mu)^2)\right] + \dots$$

$$+ (y_1 + y_2 i + y_3(i-\mu)^2) \exp\{-1 - (y_1 + y_2 i + y_3(i-\mu)^2)\}\big) - y_1 - \mu y_2 - \sigma^2 y_3,$$

i.e.

$$\varphi(y) = L(\hat{p}(y), y) = -\sum_{i=1}^{n} \exp\{-1 - (y_1 + y_2 i + y_3(i-\mu)^2)\} - y_1 - \mu y_2 - \sigma^2 y_3,$$

The dual optimization problem is

$$(D) \quad \begin{bmatrix} \max\limits_{y} & \varphi(y) \\ \text{s.t.} & y \geq 0. \end{bmatrix}$$

(c) If there is a point $\hat{y}$ such that the gradient of the dual is zero then

1. the derivative w.r.t. y_1 equal to zero say that

$$\sum_{i=1}^{n} \exp\{-1 - (\hat{y}_1 + \hat{y}_2 i + \hat{y}_3(i-\mu)^2)\} - 1 = 0$$

i.e. constraint one in the primal (P) with $p_i(\hat{y})$ is satisfied with equality.

2. the derivative w.r.t. y_2 equal to zero say that

$$\sum_{i=1}^{n} i \exp\{-1 - (\hat{y}_1 + \hat{y}_2 i + \hat{y}_3(i-\mu)^2)\} - \mu = 0$$

i.e. constraint two in the primal (P) with $p_i(\hat{y})$ is satisfied with equality.

3. the derivative w.r.t. y_3 equal to zero say that

$$\sum_{i=1}^{n} (i-\mu)^2 \exp\{-1 - (\hat{y}_1 + \hat{y}_2 i + \hat{y}_3(i-\mu)^2)\} - \sigma^2 = 0$$

i.e. constraint three in the primal (P) with $p_i(\hat{y})$ is satisfied with equality.

For continuous stochastic variables, where sums are replaced by integrals, this is used to show that the Normal distribution maximizes the entropy.