Solutions to the exam Optimization, January 12, 2022

1. (i) Show that the bounded closed interval $[a, b] = \text{conv}(\{a, b\})$.
   (ii) Let X be a nonempty bounded subset of $\mathbb{R}^n$. Define a function $S_X : \mathbb{R}^n \to \mathbb{R}$ by $S_X(x) = \sup\{\langle y, x\rangle : y \in X\}$. Show that $S_X(x)$ is a convex function.
   (iii) Show that $S_X = S_{\text{conv}(X)}$. Determine $S_{[a,b]}$ where $[a, b]$ is a closed interval over the real line.
   (iv) Show that $\langle \xi, d\rangle \leq f'(x; d)$ for all $\xi \in \partial f(x)$ and all $d \in \mathbb{R}^n$.

   Solution. (i) For any $x \in [a, b]$, x can be written as $x = (1 - \lambda)a + \lambda b$ for all $\lambda \in [0, 1]$, i.e. $[a, b] \subseteq \text{conv}(\{a, b\})$. Reverse the argument we get the $\text{conv}(\{a, b\}) \subseteq [a, b]$.

   (ii) The epigraph of S_X is

$$\text{epi}S_X = \{(x, t) : \langle y, x\rangle \leq t, \forall y \in X\} = \bigcap_{y\in X} \{(x, t) : \langle y, x\rangle \leq t\}$$

   is closed and convex since it is the intersection of closed halfspaces in $\mathbb{R}^n \times \mathbb{R}$.

   (iii) Let $Y = \text{conv}X$. Since $X \subseteq Y$, we obviously have $S_X(x) \leq S_Y(x)$. Assume the inequality is strict for some x, i.e. $\langle u, x\rangle < \langle v, x\rangle$ for all $u \in X$ and some $v \in Y$. But v is the convex combination of a set of points $v_i \in X$ by definition, that is, $v = \sum_i \lambda_i v_i$ with $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$. Since $\langle v_i, x\rangle < \langle v, x\rangle$ for alla i this would imply

$$\langle v, x\rangle = \sum_i \lambda_i \langle u_i, x\rangle < \sum_i \lambda_i \langle v, x\rangle = \langle v, x\rangle$$

   a contradiction, proving the equality.

   Clearly $S_{[a,b]}(x) = \{ax, bx\}$.

   (iv) By definition, $\xi \in \partial f(x)$ gives

$$f(y) \geq f(x) + \langle \xi, y - x\rangle,\ \forall x, y$$

   Now, using this inequality for $y = x + \tau d$ for sufficiently small $\tau > 0$, we have

$$\frac{f(x + \tau d) - f(x)}{\tau} \geq \frac{\langle \xi, \tau d\rangle}{\tau} = \langle \xi, d\rangle.$$

   Hence

$$f'(x; d) = \lim_{\tau\searrow 0} \frac{f(x + \tau d) - f(x)}{\tau} \geq \langle \xi, d\rangle.$$

   [Note that the inequality is indeed an equality, whose proof is more involved.]

2. Consider the following problem where $y = (y_1, ..., y_n)^t$, $y_0 = (\frac{1}{n}, ..., \frac{1}{n})^t$, and $e = (1, ..., 1)^t$ belong to $\mathbb{R}^n$: $\min\{y_1 : \|y - y_0\|^2 \leq \frac{1}{n(n-1)}, e^t y = 1\}$. Write the KKT conditions for this problem and verify that $(0, \frac{1}{n-1}, ..., \frac{1}{n-1})^t$ is an optimal solution. Interpret this problem with respect to an inscribed sphere in the simplex defined by $\{y : e^t y = 1, y \geq 0\}$.

   Solution. The KKT conditions are

   - $\|y - y_0\|^2 \leq \frac{1}{n(n-1)}$, $e^t y = 1$
   - $\lambda \geq 0$
   - $\lambda(\|y - y_0\|^2 \frac{1}{n(n-1)}) = 0$
   - $e_1 + 2\lambda(y - \frac{1}{n}e) + \mu e = 0$

   Let $y^* = (0, \frac{1}{n-1}, ..., \frac{1}{n-1})^t$. We see readily

$$e^t y^* = 1 \text{ and } \|y^* - y_0\|^2 = \|(-\frac{1}{n}, \frac{1}{n(n-1)}, \cdots \frac{1}{n(n-1)})\|^2 = \frac{1}{n(n-1)},$$

   meaning y^* is feasible. The last equation in the KKT conditions with $y = y^*$ yields

$$\mu = -\frac{2\lambda}{n(n-1)},\ 1 - \frac{2\lambda}{n} + \mu = 0.$$

Then $\lambda = \frac{n-1}{2} \geq 0$ $\mu = -\frac{1}{n}$. So y^* is a KKT point. Since the objective function is linear, it is convex and the constraints is the intersection of a ball (convex set) and a simplex (convex set) so the feasible set is convex. Thus this optimization problem is a convex program. Therefore the KKT conditions are also sufficient, proving y^* is the optimal solution.

Let $S = \{y : e^t y = 1, y \geq 0\}$. We see that the dimension of Aff(S) is $n-1$. Its center is y_0. Now we want to find the radius r of a sphere centered at y_0 so that the sphere is inscribed with S. So r is the distance from y_0 to the center of a simplex which is of one dimension less that that of S, say, formed in thee $y_2, ..., y_n$-space. So in the whole y space the coordinates are given by $(0, \frac{1}{n-1}, ..., \frac{1}{n-1})$. Hence

$$r^2 = (n-1)\left(\frac{1}{n-1} - \frac{1}{n}\right)^2 + \frac{1}{n^2} = \frac{1}{n(n-1)}.$$

So the given problem examines the $(n-1)$-dimensional sphere formed by the intersection of the sphere given by $\|y - y_0\|^2 \leq r^2$ with the hyperplane $e^t y = 1$, without the nonnegative restrictions $y \geq 0$, and looks for the minimal value of any coordinate in this region. In this problem it is y_1.

3. Consider the binary optimization problem: $\min\{x^T Q x : x_i \in \{-1, 1\}\}$, where $Q \in \mathbb{R}^{n \times n}$ is symmetric positive definite and $x = (x_1, ..., x_n)^t$. Show that its Lagrange dual problem is in the following form: $\max\{\text{tr}\Lambda : Q - \Lambda \geq 0\}$, where Λ is a diagonal matrix.

   Solution. Since $x_i \in \{-1, 1\}$ is equivalent to $x_i^2 - 1 = 0$, the optimization problem at hand can be formulated as $\min\{x^T Q x : x_i^2 = 1\}$. Introduce the Lagrange multiplier vector $\lambda \in \mathbb{R}^n$ and let $\Lambda = \text{diag}(\lambda)$. Now the Lagrange function is

$$L(x, \lambda) = x^t Q x + \sum_{i=1}^n \lambda_i (1 - x_i^2) = x^t Q x - x^t \Lambda x + \text{tr}\Lambda = x^t (Q - \Lambda) x + \text{tr}\Lambda.$$

   Note that $L(x, \lambda) \geq \text{tr}\Lambda$ if $Q - \Lambda \geq 0$. So the dual problem is

$$\max\{\text{tr}\Lambda : Q - \Lambda \geq 0\}.$$

4. Usw Phase I of the simplex method to determine whether the following system of equations has a nonnegative solution.

$$4x_1 + 5x_2 + x_3 + 2x_4 = 0, 3x_1 + 3x_2 + x_3 + x_4 = 1$$

   Find one such solution if it exists

   Solution. Introduce two artificial variables x_5 and x_6 (nonnegative) and minimize $x_5 + x_6$. Let $x = (x_1, ..., x_6)^t$. The Simplex Phase I problem is

$$\min\left\{(0,0,0,0,1,1)x : \begin{pmatrix} 4 & 5 & 1 & 2 & 1 & 0 \\ 3 & 3 & 1 & 1 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}$$

   or

$$\begin{array}{cccccc|c} 4 & 5 & 1 & 2 & 1 & 0 & 0 \\ 3 & 3 & 1 & 1 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{array}$$

   Suitable basic variables are x_5, x_6 and put it in the standard form

$$\begin{array}{cccccc|c} 4 & 5 & 1 & 2 & 1 & 0 & 0 \\ 3 & 3 & 1 & 1 & 0 & 1 & 1 \\ \hline -7 & -8 & -2 & -3 & 0 & 0 & -1 \end{array}$$

   Now choose x_3 as basic variable we get a new tableau

$$\begin{array}{cccccc|c} 4 & 5 & 1 & 2 & 1 & 0 & 0 \\ -1 & -2 & 0 & -1 & -1 & 1 & 1 \\ \hline 1 & 2 & 0 & 1 & 3 & 0 & -1 \end{array}$$

   We see that all reduced costs are ≥ 0, implying that we reach the optimum, which is achieved at $x_6 = 1$ and the other variables are 0. So $\min x_5 + x_6 = 1$, that is the original system does not have a solution.

5. Formulate the minimization problem Minimize $\|Ax - b\|_\infty$ (ℓ_∞-norm approximation) as an LP problem. Explain in detail the relation between the optimal solution and the solution of its equivalent LP.

Solution. It is equivalent to the LP

$$\min\{y : Ax - b \le ye, Ax - b \ge -ye\}$$

in the variables x, y, where e is an all 1 vector. Now we show the equivalence. Assume x is fixed in this problem, and we optmized only over y. The constraints say that $-y \le a_k^t x - b_k \le y$, for each k, i.e., $y \ge |a_k^t x - b_k|$. Then $y \ge \max_k |a^t x - b| = \|Ax - b\|_\infty$. It says the optimal value of the LP is $\|Ax - b\|_\infty$ if x is fixed. Hence optimizing over x and y at the same time is equivalent to the original problem.

6. Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at x and let the vectors $d_1, ..., d_n$ in $\mathbb{R}^n$ be linearly independent. Assume that the minimum of $f(x + \lambda d_j)$ over $\lambda \in \mathbb{R}$ occurs at $\lambda = 0$ for $j = 1, ..., n$. Show that $\nabla f(x) = 0$. Does this imply that f has a local minimum at x?

   Solution. Since the minimum of $f(x + \lambda d_j)$ over $\lambda \in \mathbb{R}$ attained at $\lambda = 0$ for $j = 1, ..., n$, we have that
   $$\frac{d}{d\lambda} f(x + \lambda d_j)\big|_{\lambda=0} = \langle \nabla f(x), d_j \rangle, \ \forall j = 1, ..., n,$$
   i.e.this homogeneous system of linear equations (for $\nabla f(x)$) with the coefficient matrix A with d_j^t as rows. Now $d_1, ..., d_n$ are linearly independent, i.e. it has a unique solution which is trivial. Hence $\nabla f(x) = 0$. However this does not imply that f has a local minimum at x. For example Figure 4.1 (BSS, p. 172) with $x = (0, 0)$, $d_1 = (1, 0)$, and $d_2 = (0, 1)$.

7. Assume $x_i > 0$, $i = 1, ..., n$. Let $A = \frac{x_1 + \cdots + x_n}{n}$, and $G = (x_1 \cdot ... \cdot x_n)^{\frac{1}{n}}$.

   (i) Show, using the theory developed in this course, $G(x) \le A(x)$

   (ii) Justify if the set $\{x \in \mathbb{R}^n_{++} : G(x) \ge A(x)\}$ is convex or not. Is this set a cone if if we define $0^{\frac{1}{n}} = 0$?

   Solution. We show that $\frac{1}{n}(\log x_1 + \cdots \log x_n) \le \log \frac{x_1 + \cdots + x_n}{n}$. Now log cot is a concave function so the inequality follows from Jensen's inequality with $\lambda_1 = ... = \lambda_n = \frac{1}{n}$.

   Since $A(x)$ is convex and $-G(x)$ (See lecture notes) is convex, the function $A(x) - G(x)$ is a convex function. Now the set is a level set of a convex function, so it is convex. Indeed this is a convex cone because for anly $\lambda \ge 0$ $A(\lambda x) - G(\lambda x) = A(\lambda x) - G(\lambda x) = \lambda(A(x) - G(x)) \le 0$.

8. In line search to find optimal solution to nonlinear optimization problems we often need to solve the the following optimization problem
   $$\min\{\| -\nabla f(x) - d\|^2 : A_1 d = 0\},$$
   where A_1 is a $\nu \times n$ matrix with rank ν and x is fixed.

   (i) Find the optimal solution $\bar{d}$ is an optimal solution without using the KKT conditions or Lagrange relaxation.

   (ii) Solve $\bar{d}$ from the KKT system.

   (iii) Find $\bar{d}$ in case $\nabla f(x) = (2, -3, 3)^t$ and $A_1 = \begin{pmatrix} 2 & 2 & -3 \\ 2 & 1 & 2 \end{pmatrix}$.

   Solution. By the projection theorem we know that $\bar{d}$ is optimal if and only if $\bar{d}$ is a projection of $-\nabla f(x)$ onto the nullspace of A_1. So $\bar{d} = -(I - A_1^t (A_1 A_1^t)^{-1} A_1) \nabla f(x)$. This can be obtained by the KTT system: $-\nabla f(x) = \bar{d} - A_1^t u, A_1 \bar{d} = 0$ by multiplying the first equation by A_1 together with the second equation and A_1 has full row rank. A straightforward computation gives $d = (-266, 380, 76)^t / 153$.