# Optimization 1 — Homework 10

December 31, 2020

**Problem 1**
Consider the problem

$$\begin{aligned} \min\ & -2x^2 + 2y^2 + 4x \\ \text{s.t. } & x^2 + y^2 - 4 \leq 0, \\ & x^2 + y^2 - 4x + 3 \leq 0. \end{aligned}$$

**(a)** Prove that there exists an optimal solution to the problem.

**(b)** Find all KKT points of the problem.

**(c)** Find the optimal solution of the problem.

**Problem 2**
Consider the optimization problem

$$\begin{aligned} \min\quad & \mathbf{a}^T\mathbf{x} \\ \text{s.t.}\quad & \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c \leq 0. \end{aligned}$$

where $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is PD, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.

**(a)** For which values of $\mathbf{Q}, \mathbf{b}, c$ is the problem feasible?

**(b)** For which values of $\mathbf{Q}, \mathbf{b}, c$ are the KKT conditions necessary?

**(c)** For which values of $\mathbf{Q}, \mathbf{b}, c$ are the KKT conditions sufficient?

**(d)** Under the condition of the third part, find the optimal solution of the problem using the KKT conditions.

**Problem 3**
Consider the problem

$$\begin{aligned} \min\quad & 2x^2 + (y-4)^2 \\ s.t.\quad & -x^2 + 3ky \leq 0, \end{aligned}$$

where $k > 0$.

**(a)** Find all KKT points of the problem.

**(b)** Use necessary second order conditions in order to find the optimal solution of the problem for any $k > 0$.

**Problem 4**

Let $\mathbf{x}^*$ be a local minimum point of the problem

$$\begin{array}{lll} \min & f(\mathbf{x}) & \\ \text{s.t.} & g_i(\mathbf{x}) \le 0, & i = 1,2,\dots,m, \\ & h_j(\mathbf{x}) \le 0, & j = 1,2,\dots,p, \\ & s_k(\mathbf{x}) = 0, & k = 1,2,\dots,q, \end{array}$$

where $f, g_i$ are continuously differentiable functions, $g_i$ are convex, $h_j$ and $s_k$ are affine. Suppose that the generalized Slater's condition is satisfied. Then there exist $\boldsymbol{\lambda} \in \mathbb{R}^m_+$, $\boldsymbol{\eta} \in \mathbb{R}^p_+$ and $\boldsymbol{\mu} \in \mathbb{R}^q$ such that

$$\begin{cases} \nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^p \eta_j \nabla h_j(\mathbf{x}^*) + \sum_{k=1}^q \mu_k \nabla s_k(\mathbf{x}^*) = \mathbf{0}_n, \\ \lambda_i g_i(\mathbf{x}^*) = 0, \quad i = 1,2,\dots,m, \\ \eta_j h_j(\mathbf{x}^*) = 0, \quad j = 1,2,\dots,p. \end{cases}$$

*Hint:* use Motzkin's lemma from HW9.

**Problem 5**

Consider the TRSP

$$\begin{array}{ll} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c \\ \text{s.t.} & \|\mathbf{x}\|^2 \le \alpha^2, \end{array}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$ and $\alpha \in \mathbb{R}_{++}$.
Since $\mathbf{A}$ is symmetric, we can write it using the spectral decomposition as $\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T$, where $\mathbf{Q}$ is orthogonal and $\mathbf{D}$ is a digonal matrix with the eigenavlues $\lambda_1(\mathbf{A}) \le \lambda_2(\mathbf{A}) \le ... \le \lambda_n(\mathbf{A})$ on the diagonal. Then for any $\lambda \ne -\lambda_i(\mathbf{A})$ we can write

$$\mathbf{x}(\lambda) = -\mathbf{Q}(\mathbf{D} + \lambda\mathbf{I}_n)^{-1}\mathbf{Q}^T\mathbf{b} = -\sum_{i=1}^n \frac{\mathbf{Q}_i^T\mathbf{b}}{\lambda_i(\mathbf{A}) + \lambda}\mathbf{Q}_i,$$

where $\mathbf{Q_i}$ is the $i$-th column of $\mathbf{Q}$ (meaning, this is the eigenvector corresponding to $\lambda_i(\mathbf{A})$). Since $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_n$ we have

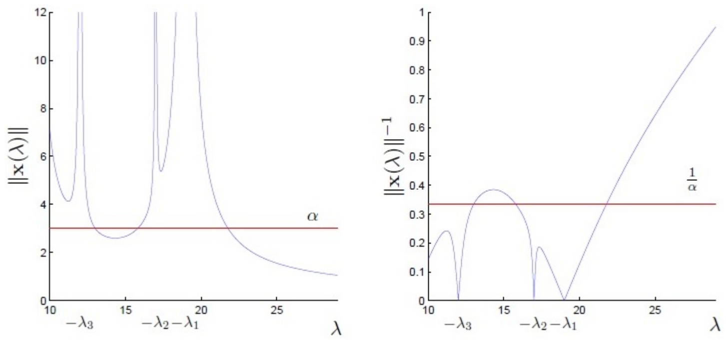
$$\|\mathbf{x}(\lambda)\|^2 = \left\|(\mathbf{A} + \lambda\mathbf{I}_n)^{-1}\mathbf{b}\right\|^2 = \sum_{i=1}^n \frac{\left(\mathbf{Q}_i^T\mathbf{b}\right)^2}{\left(\lambda_i(\mathbf{A}) + \lambda\right)^2}.$$

**(a)** Show that if $\mathbf{Q}_1^T\mathbf{b} \ne 0$ then $-\mathbf{b} \notin \text{Image}(\mathbf{A} - \lambda_{\min}(\mathbf{A})\mathbf{I}_m)$.

Recall that in the solution of the TRSP we find $\lambda \ge 0$ such that $\|\mathbf{x}(\lambda)\|^2 = \alpha^2$, which is the root of the function $\phi_1(\lambda) = \|\mathbf{x}(\lambda)\| - \alpha$. An alternative approach to find such $\lambda \ge 0$ is to use Newton's method. This generates a sequence

$$\lambda^{k+1} = \lambda^k - \frac{\phi_1(\lambda^k)}{\phi_1'(\lambda^k)}.$$

Consider the function $\phi_2(\lambda) = \frac{1}{\|\mathbf{x}(\lambda)\|} - \frac{1}{\alpha}$. In the attached figure the left plot illustrates the value of $\phi_1(\lambda) + \alpha = \|\mathbf{x}(\lambda)\|$, while the red line is $\alpha$. The right plot illustrates the value of $\phi_2(\lambda) + \frac{1}{\alpha} = \frac{1}{\|\mathbf{x}(\lambda)\|}$ while the red line is $\alpha^{-1}$.




**(b)** Based on the figures, explain (no mathematical arguments are required) why applying the Newton's method on $\phi_2$ is preferred to applying the same method on $\phi_1$.

**(c)** Show that the Newton's step applied on the function $\phi_2$ is equivalent to the following algorithm:

- Initialization: choose $\lambda^0 \geq 0$.
- Step:
  - Factor $\mathbf{A} + \lambda^k \mathbf{I}_n = \mathbf{L}^T \mathbf{L}$ (Cholesky factorization).
  - Solve $\mathbf{L}^T \mathbf{L} \mathbf{p}^k = -\mathbf{b}$, $\mathbf{L}^T \mathbf{q}^k = \mathbf{p}^k$.
  - Set

$$\lambda^{k+1} = \lambda^k + \left( \frac{\|\mathbf{p}^k\|}{\|\mathbf{q}^k\|} \right)^2 \left( \frac{\|\mathbf{p}^k\| - \alpha}{\alpha} \right).$$

**(d)** Generate the data $(\mathbf{A}, \mathbf{b})$ according to the following commands

```
randn('seed',317);
rand('seed',317);
n = 10;
Q = orth(randn(n,n));
D = randperm(2*n)';
D = diag(sort(D(1:n).*sign(randn(n,1))));
A = Q*D*Q';
b = 20*rand(n,1)-15;
alpha = 3;
disp(b'*Q(:,1));
```

Implement the algorithm for solving the TRSP for solving this problem. Consider the following two strategies for finding $\lambda$:

1. Bisection.
   *Note:* set the initial lower bound to $l = 10^{-7}$ or $l = -\lambda_{\min}(\mathbf{A}) + 10^{-7}$. For finding the initial upper bound implement
   ```
   u=l+1;
   while phi(u)>0
   u=2*u;
   end
   ```
2. Finding the root of $\phi_2(\lambda)$ using Newton's method.
   *Note:* set $\lambda^0$ as the initial upper bound in the bisection method. To guarantee that $\mathbf{A} + \lambda^k \mathbf{I}_n \succ 0$, update $\lambda^{k+1} = \max\left\{\lambda^{k+1}, 10^{-7}\right\}$ or $\lambda^{k+1} = \max\left\{\lambda^{k+1}, -\lambda_{\min}(\mathbf{A}) + 10^{-7}\right\}$.

Compare the number of iterations required for computing $\lambda$.