Optimization 1 — Tutorial 10

December 31, 2020

Consider the problem

$$
\begin{aligned}
(P) \quad \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) \\
\text{s.t.} \quad & g_i(\mathbf{x}) \le 0, \quad i = 1,2,\dots,m, \\
& h_j(\mathbf{x}) \le 0, \quad j = 1,2,\dots,p, \\
& s_k(\mathbf{x}) = 0, \quad k = 1,2,\dots,q,
\end{aligned}
$$

where $f, g_i, h_j, s_k \colon \mathbb{R}^n \to \mathbb{R}$ are functions. We define the Lagrangian $L \colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}$ of problem (P) as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\eta}) = f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}^*) + \sum_{k=1}^{p} \eta_k s_k(\mathbf{x}^*).$$

If the functions are **continuously differentiable**, then

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\eta}) = \nabla f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(\mathbf{x}^*) + \sum_{k=1}^{p} \eta_k \nabla s_k(\mathbf{x}^*).$$

Definitions

Assume that the functions f, g_i, h_j, s_k in problem (P) are **continuously differentiable**.

(i) A feasible point $\mathbf{x}^*$ of (P) is called a **KKT point** if there exist $\boldsymbol{\lambda} \in \mathbb{R}^m_+$, $\boldsymbol{\mu} \in \mathbb{R}^p_+$ and $\boldsymbol{\eta} \in \mathbb{R}^q$ such that

$$\begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\eta}) = \mathbf{0}_n, \\ \lambda_i g_i(\mathbf{x}^*) = 0, \quad \forall i = 1,2,\dots,m, \\ \mu_j h_j(\mathbf{x}^*) = 0, \quad \forall j = 1,2,\dots,p. \end{cases}$$

(ii) A feasible point $\mathbf{x}^*$ of (P) is called a **regular point** if the set

$$\left\{ \nabla g_i(\mathbf{x}^*), \nabla h_j(\mathbf{x}^*), \nabla s_k(\mathbf{x}^*) : i \in I(\mathbf{x}^*), j \in J(\mathbf{x}^*), k \in \{1,2,\dots,q\} \right\},$$

for

$$\begin{aligned} I(\mathbf{x}^*) &= \{ i \in \{1,2,\dots,m\} : g_i(\mathbf{x}^*) = 0 \}, \\ J(\mathbf{x}^*) &= \{ j \in \{1,2,\dots,p\} : h_j(\mathbf{x}^*) = 0 \}, \end{aligned}$$

is linearly independent.

(iii) If the functions are **twice continuously differentiable**, we say that a feasible KKT point $\mathbf{x}^*$ of (P) satisfies the **second-order necessity conditions** if

$$\mathbf{d}^T \nabla^2_{\mathbf{xx}} L(\mathbf{x}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\eta})\, \mathbf{d} \ge 0, \quad \forall \mathbf{d} \in \Lambda(\mathbf{x}^*),$$

where

$$\Lambda(\mathbf{x}^*) = \left\{ \mathbf{d} \in \mathbb{R}^n \colon \begin{array}{c} \nabla g_i(\mathbf{x}^*)^T \mathbf{d} = 0, \quad \forall i \in I(\mathbf{x}^*) \\ \nabla h_j(\mathbf{x}^*)^T \mathbf{d} = 0, \quad \forall j \in J(\mathbf{x}^*) \\ \nabla s_k(\mathbf{x}^*)^T \mathbf{d} = 0, \quad \forall k \in \{1,2,\dots,q\} \end{array} \right\}.$$

Summary of KKT and Second-order Conditions

In the following two cases we assume that the functions f, g_i, h_j, s_k are **continuously differentiable**, and that problem (P) is **feasible**.

1. We have
   (a) $\{\text{locally optimal}\} \subseteq \{\text{KKT}\} \cup \{\text{irregular}\}$.
   (b) If all functions are **twice continuously differentiable**, then
   $$\{\text{locally optimal}\} \subseteq \{\text{second order}\} \cup \{\text{irregular}\} \subseteq \{\text{KKT}\} \cup \{\text{irregular}\}\,.$$
2. Assume that g_i are **convex**, h_j, s_k are **affine**.
   (a) If generalized Slater's condition is satisfied, then $\{\text{locally optimal}\} \subseteq \{\text{KKT}\}$.
   (b) If f is convex, then $\{\text{KKT}\} \subseteq \{\text{optimal}\}$.
   (c) If both (a) and (b) hold, then $\{\text{optimal}\} = \{\text{KKT}\}$.

Remark. Notice that

1. When we solve the KKT conditions, we find **all** feasible KKT points (regular and irregular). Therefore, if required, we need to find all other irregular points.
2. The linearly constrained cases are contained in case 2.
3. In case 1 and case 2(a), if $\{\text{local optimal}\} = \emptyset$ (in particular, if there is no optimal solution) – then finding all feasible KKT and irregular points, **does not** guarantee finding a locally optimal or optimal point.
4. In case 2(b), if we find one feasible KKT point – it is an optimal solution.

Problem 1

Consider the optimization problem

$$\begin{aligned}
\min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(x,y,z) = 2xy + \frac{1}{2}z^2 \\
\text{s.t.} \quad & f_1(x,y,z) = 2xz + \frac{1}{2}y^2 \le 0, \\
& f_2(x,y,z) = 2yz + \frac{1}{2}x^2 \le 0.
\end{aligned}$$

(a) Show that the optimal solution is $\mathbf{0}_3$.

(b) Determine whether $\mathbf{0}_3$ satisfies the second-order necessary optimality conditions.

Solution

(a) Notice that $\mathbf{0}_3$ is a feasible point with value 0.

- Assume that there exists a feasible solution (x,y,z) such that $2xy + \frac{1}{2}z^2 < 0$. Then $z^2 < -4xy$. This implies that $x, y \neq 0$ and that $\operatorname{sign}(x) \neq \operatorname{sign}(y)$.
- Moreover, any feasible solution satisfies $y^2 \le -4xz$ and $x^2 \le -4yz$. Therefore, $x,y,z \neq 0$ and they are all with different signs, which is impossible. Therefore $\mathbf{0}_3$ is an optimal solution.

(b) Since $\mathbf{0}_3$ is optimal, it also a locally optimal point. Therefore, if it is a regular point – it must be a KKT point that satisfies the second-order conditions (see case 1(b)).

- So, we check whether $\mathbf{0}_3$ is a regular point: at $\mathbf{0}_3$, both constraints are active and therefore $I(\mathbf{0}_3) = \{1,2\}$. Since the set of the gradients of the active constraints at $\mathbf{0}_3$ contains only the origin, then $\mathbf{0}_3$ is irregular.
- Therefore, we still do not know if $\mathbf{0}_3$ satisfies the second-order conditions. To this end, we first need to check if $\mathbf{0}_3$ is a KKT point.
  - The Lagrangian is

$$L(\mathbf{x},\lambda_1,\lambda_2) = 2xy + \frac{1}{2}z^2 + \lambda_1\left(2xz + \frac{1}{2}y^2\right) + \lambda_2\left(2yz + \frac{1}{2}x^2\right), \quad \lambda_1,\lambda_2 \ge 0.$$

  - Therefore, the KKT conditions are

$$\begin{cases}
\begin{pmatrix} 2(\lambda_1 z + y) + \lambda_2 x \\ \lambda_1 y + 2\lambda_2 z + 2x \\ 2(\lambda_1 x + \lambda_2 y) + z \end{pmatrix} = \mathbf{0}, & \nabla_{\mathbf{x}} L(\mathbf{x},\boldsymbol{\lambda}) = \mathbf{0}_n \\
\lambda_1\left(2xz + \frac{1}{2}y^2\right) = 0, & \text{complementary slackness} \\
\lambda_2\left(2yz + \frac{1}{2}x^2\right) = 0, & \text{complementary slackness} \\
2xz + \frac{1}{2}y^2 \le 0, & \text{feasibilty} \\
2yz + \frac{1}{2}x^2 \le 0, & \text{feasibilty}
\end{cases}$$

  and we see that $\mathbf{0}_3$ is indeed a feasible KKT point.
- So we are left to check whether $\mathbf{0}_3$ satisfies the second-order conditions.
  - Notice that

$$\nabla^2_{\mathbf{xx}} L(\mathbf{x},\lambda_1,\lambda_2) = \begin{pmatrix} \lambda_2 & 2 & 2\lambda_1 \\ 2 & \lambda_1 & 2\lambda_2 \\ 2\lambda_1 & 2\lambda_2 & 1 \end{pmatrix}.$$

  - We have

$$\begin{aligned}
\Lambda(\mathbf{0}_3) &= \left\{\mathbf{d} \in \mathbb{R}^3 : \nabla f_1(\mathbf{0}_3)^T \mathbf{d} = 0, \nabla f_2(\mathbf{0}_3)^T \mathbf{d} = 0\right\} \\
&= \left\{\mathbf{d} \in \mathbb{R}^3 : \mathbf{0}_3^T \mathbf{d} = 0, \mathbf{0}_3^T \mathbf{d} = 0\right\} = \mathbb{R}^3,
\end{aligned}$$

  which means that the second-order conditions are satisfied if and only if $\nabla^2_{\mathbf{xx}} L \succeq 0$.

- Notice that

$$\begin{cases} \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}^T \nabla^2_{\mathbf{xx}} L \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} = 1 + \lambda_1 - 4\lambda_2, \\ \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}^T \nabla^2_{\mathbf{xx}} L \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 1 - 4\lambda_1 + \lambda_2, \\ \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}^T \nabla^2_{\mathbf{xx}} L \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = -4 + \lambda_1 + \lambda_2. \end{cases}$$

  Summing the above we have $-2(1+\lambda_1+\lambda_2) \geq 0$, which contradicts the fact that $\lambda_1, \lambda_2 \geq 0$.
- Therefore, this is not a PSD matrix and the second-order conditions are not satisfied.

This example demonstrates the fact that not every optimal solution can be attained by the KKT conditions, as some optimal solutions are irregular points.

Conditions for Trust Region Sub-problems

Consider the TRSP

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c \\ \text{s.t.} \quad & \|\mathbf{x}\|^2 \leq \alpha, \end{aligned}$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is symmetric, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$ and $\alpha \in \mathbb{R}_{++}$. Then $\mathbf{x}^*$ is an optimal solution of the problem if and only if there exists $\lambda^* \geq 0$ such that

$$\begin{cases} (\mathbf{A} + \lambda^*\mathbf{I}_n)\mathbf{x}^* = -\mathbf{b}, & (1) \\ \|\mathbf{x}^*\|^2 \leq \alpha, & (2) \\ \lambda^*\left(\|\mathbf{x}^*\|^2 - \alpha\right) = 0, & (3) \\ \mathbf{A} + \lambda^*\mathbf{I}_n \succeq 0. & (4) \end{cases}$$

Problem 2

Devise an algorithm for solving the TRSP, assuming that $-\mathbf{b} \notin \operatorname{Image}(\mathbf{A} - \lambda_{\min}(\mathbf{A})\mathbf{I}_n)$.

Solution

From (1) we obtain that $\lambda^* \neq -\lambda_{\min}(\mathbf{A})$.

- $\mathbf{A} \succ 0$:
  - In this case $\mathbf{x}^*$ is an optimal solution of TRSP if and only if there exists $\lambda^* \geq 0$ such that

    $$\begin{cases} (\mathbf{A} + \lambda^*\mathbf{I}_n)\mathbf{x}^* = -\mathbf{b}, & (1) \\ \|\mathbf{x}^*\|^2 \leq \alpha, & (2) \\ \lambda^*\left(\|\mathbf{x}^*\|^2 - \alpha\right) = 0. & (3) \end{cases}$$

  - If $\lambda^* = 0$ then $\mathbf{x}^* = -\mathbf{A}^{-1}\mathbf{b}$ is an optimal solution if and only if $\left\|\mathbf{A}^{-1}\mathbf{b}\right\| \leq \alpha$.
  - If $\lambda^* > 0$ then from (1) we have $\mathbf{x}^* = -(\mathbf{A} + \lambda^*\mathbf{I}_n)^{-1}\mathbf{b}$. Plugging this into (3) we have $\left\|(\mathbf{A} + \lambda^*\mathbf{I}_n)^{-1}\mathbf{b}\right\|^2 = \alpha$.

* $*$ Notice that the function $\phi(\lambda) = \left\|(\mathbf{A}+\lambda\mathbf{I}_n)^{-1}\mathbf{b}\right\|^2 - \alpha$ is strictly decreasing in $(0,\infty)$ (since $\phi(\lambda)$ is a rational function with a power of λ in its denominator), thus has a unique root in $(0,\infty)$ (notice that outside the domain $(0,\infty)$ then $\mathbf{A}+\lambda\mathbf{I}_n$ is not necessarily invertible).
* $*$ Therefore, if $\lambda^* > 0$ then $\mathbf{x}^* = -(\mathbf{A}+\lambda^*\mathbf{I}_n)^{-1}\mathbf{b}$ is an optimal solution, where λ^* is the unique root of $\phi(\lambda)$ in $(0,\infty)$.

- $\mathbf{A} \nsucc 0$:
  - From (4) we have $\lambda^* > -\lambda_{\min}(\mathbf{A}) \geq 0$. Therefore $\mathbf{A}+\lambda^*\mathbf{I}_n \succ 0$.
  - From (3) we have $\|\mathbf{x}^*\|^2 = \alpha$. Now, from (1) we have $\mathbf{x}^* = -(\mathbf{A}+\lambda^*\mathbf{I}_n)^{-1}\mathbf{b}$. Plugging this into $\|\mathbf{x}^*\|^2 = \alpha$ we have $\left\|(\mathbf{A}+\lambda^*\mathbf{I}_n)^{-1}\mathbf{b}\right\|^2 = \alpha$.
  - Notice that the function $\phi(\lambda) = \left\|(\mathbf{A}+\lambda\mathbf{I}_n)^{-1}\mathbf{b}\right\|^2 - \alpha$ is strictly decreasing in $(-\lambda_{\min}(\mathbf{A}),\infty)$, thus has a unique root in $(-\lambda_{\min}(\mathbf{A}),\infty)$ (notice that outside the domain $(-\lambda_{\min}(\mathbf{A}),\infty)$ then $\mathbf{A}+\lambda\mathbf{I}_n$ is not necessarily invertible).

We obtain the following algorithm:

Algorithm 1: TRS

```
Data: A = A^T ∈ R^{n×n}, b ∈ R^m and α ∈ R_{++}
Result: x_trs - a solution to (TRS)
if A ≻ 0 then
    x_naive = −A^{-1} b
    if ||x_naive||^2 ≤ α then
        return x_naive
    else
        find λ > 0 such that φ(λ) = 0
        x = −(A + λI)^{-1} b
    end
else
    find λ > −λ_min(A) such that φ(λ) = 0.
    x = −(A + λI)^{-1} b
end
return x
```

In order to find λ such that $\phi(\lambda) = 0$ we can use the bisection algorithm:

Algorithm 2: Bisection

Input: ε - tolerance parameter. $a < b$ - two numbers satisfying $f(a)f(b) < 0$.
Initialization: take $l_0 = a, u_0 = b$.
General Step: for any $k = 0, 1, 2, \ldots$ execute the following steps:

(a) Take $x_k = \frac{u_k + l_k}{2}$.

(b) If $f(l_k) \cdot f(x_k) > 0$, define $l_{k+1} = x_k, u_{k+1} = u_k$. Otherwise, define $l_{k+1} = l_k, u_{k+1} = x_k$.

(c) if $u_{k+1} - l_{k+1} \leq \varepsilon$, then STOP and x_k is the output.

A MATLAB implementation of the bisection algorithm is given in moodle.

Problem 3

Consider the optimization problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n} \quad & f(x,y,z) = 2x + 3y - z \\ \text{s.t.} \quad & f_1(x,y,z) = x^2 + y^2 + z^2 = 1, \\ & f_2(x,y,z) = x^2 + 2y^2 + 2z^2 = 2. \end{aligned}$$

Find its optimal solution.

Solution

- Since the problem satisfies case 1(a) (and 1(b)) we have {locally optimal} $\subseteq$ {KKT} $\cup$ {irregular}.
- The problem indeed has an optimal solution since f is continuous over a non-empty and compact set (for example, $(0, 1, 0)$ is in the set).
- Since {locally optimal} $\neq \emptyset$, we need to find all KKT and irregular points.
- Writing the KKT conditions we have

$$\begin{cases} 2\mu_1 x + 2\mu_2 x + 2 = 0 & (i) \\ 2\mu_1 y + 4\mu_2 y + 3 = 0, & (ii) \\ 2\mu_1 z + 2\mu_2 z - 1 = 0 & (iii) \\ x^2 + y^2 + z^2 = 1, & (iv) \\ x^2 + 2y^2 + 2z^2 = 2, & (v) \end{cases}$$

  - From $(i), (i), (iii)$ we have $x, y, z \neq 0$. Then $\mu_1 + \mu_2 = -\frac{1}{x}$ and $\mu_1 + \mu_2 = \frac{1}{2z}$. Therefore $x = -2z$. Plugging into $(iv), (v)$ we have

$$\begin{cases} y^2 + 5z^2 = 1 \\ 2y^2 + 6z^2 = 2 \end{cases} \Rightarrow z^2 = 0$$

  - This contradicts $z \neq 0$ and there are no feasible KKT points.
- Therefore, the optimal solution is a feasible irregular point.
- The gradients of the (active) constraints are

$$\begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}, \begin{pmatrix} 2x \\ 4y \\ 4z \end{pmatrix}.$$

  These two vectors are linearly dependent if and only if $x = 0$, $y, z \in \mathbb{R}$.
- We need to find the feasible irregular points. In this case the problem becomes

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} \quad & 3y - z \\ \text{s.t.} \quad & y^2 + z^2 = 1. \end{aligned}$$

- We saw in class that

$$3y - z = \begin{pmatrix} y \\ z \end{pmatrix}^T \begin{pmatrix} 3 \\ -1 \end{pmatrix} \geq - \left\| \begin{pmatrix} y \\ z \end{pmatrix} \right\| \left\| \begin{pmatrix} 3 \\ -1 \end{pmatrix} \right\| = -\sqrt{10}\sqrt{y^2 + z^2} = -\sqrt{10}.$$

- This inequality is satisfied with equality if and only if there exists $\alpha \leq 0$ such that $\left(\begin{smallmatrix} y \\ z \end{smallmatrix}\right) = \alpha \left(\begin{smallmatrix} 3 \\ -1 \end{smallmatrix}\right)$.
- Solving $9\alpha^2 + \alpha^2 = 1$ for $\alpha \leq 0$ we have $\alpha = -\frac{1}{\sqrt{10}}$. So $\left(0, -\frac{3}{\sqrt{10}}, \frac{1}{\sqrt{10}}\right)$ is the only feasible irregular point, and thus a unique optimal solution.