DATA8004: Optimization for Statistical Learning

Subject Lecturer: Man-Chung YUE

Lecture 6
Constrained Optimization
KKT conditions

Problem settings

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & f(x) \\ \text{subject to} \quad & h_j(x) = 0, \;\; j = 1, \dots, p, \\ & g_i(x) \leq 0, \;\; i = 1, \dots, m. \end{aligned} \tag{1}$$

Here:

- f, h_j and g_i are all C^1 functions.
- For notational simplicity, we denote

$$J = \{1, \dots, p\}, \quad I = \{1, \dots, m\}.$$

Problem settings

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & f(x) \\ \text{subject to} \quad & h_j(x) = 0, \quad j = 1, \ldots, p, \\ & g_i(x) \leq 0, \quad i = 1, \ldots, m. \end{aligned} \tag{1}$$

Here:

- f, h_j and g_i are all C^1 functions.
- For notational simplicity, we denote

$$J = \{1, \ldots, p\}, \quad I = \{1, \ldots, m\}.$$

- Aim: Find conditions to help characterize local minimizers!

> Definition: We say that x^* is a local minimizer of (1) if x^* is feasible and $\exists \epsilon > 0$ so that $f(x) \geq f(x^*)$ whenever x is feasible and $\|x - x^*\|_2 < \epsilon$.

Problem settings

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{subject to} & h_j(x) = 0, \ \ j = 1, \ldots, p, \\ & g_i(x) \leq 0, \ \ i = 1, \ldots, m. \end{array} \tag{1}$$

Here:

- f, h_j and g_i are all C^1 functions.
- For notational simplicity, we denote
$$J = \{1, \ldots, p\}, \quad I = \{1, \ldots, m\}.$$
- Aim: Find conditions to help characterize local minimizers!

Definition: We say that x^* is a local minimizer of (1) if x^* is feasible and $\exists \epsilon > 0$ so that $f(x) \geq f(x^*)$ whenever x is feasible and $\|x - x^*\|_2 < \epsilon$.

- When f, g_i and h_j are all affine, the above reduces to an LP.

Problem settings

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x) \\ \text{subject to} \quad & h_j(x) = 0, \;\; j = 1, \ldots, p, \\ & g_i(x) \le 0, \;\; i = 1, \ldots, m. \end{aligned} \tag{1}$$

Here:

- f, h_j and g_i are all C^1 functions.
- For notational simplicity, we denote

$$J = \{1, \ldots, p\}, \quad I = \{1, \ldots, m\}.$$

- Aim: Find conditions to help characterize local minimizers!

> Definition: We say that x^* is a local minimizer of (1) if x^* is feasible and $\exists \epsilon > 0$ so that $f(x) \ge f(x^*)$ whenever x is feasible and $\|x - x^*\|_2 < \epsilon$.

- When f, g_i and h_j are all affine, the above reduces to an LP.
- Idea: Derive optimality conditions based on some related LPs?

An LP revisited

Consider the following LP:

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & c^T x \\ \text{subject to} \quad & Bx = d, \\ & Ax \leq b. \end{aligned}$$

where $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{p \times n}$ and $A \in \mathbb{R}^{q \times n}$.

An LP revisited

Consider the following LP:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & c^T x \\ \text{subject to} \quad & Bx = d, \\ & Ax \le b. \end{aligned}$$

where $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{p\times n}$ and $A \in \mathbb{R}^{q\times n}$.

The dual problem is given by

$$\begin{aligned} \underset{\mu\in\mathbb{R}^p,\lambda\in\mathbb{R}^q}{\text{Maximize}} \quad & -d^T\mu - b^T\lambda \\ \text{subject to} \quad & B^T\mu + A^T\lambda + c = 0, \\ & \lambda \ge 0. \end{aligned}$$

Here, we assume familiarity of LP duality theory.

KKT conditions for LP

Theorem 6.1: (Karush-Kuhn-Tucker conditions for the LP)
Consider the linear program

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & c^T x \\ \text{subject to} \quad & Bx = d, \\ & Ax \le b. \end{aligned} \tag{2}$$

where $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{p\times n}$ and $A \in \mathbb{R}^{q\times n}$. Then $x^* \in \mathbb{R}^n$ is an optimal solution of (2) if and only if there exist $\lambda^* \in \mathbb{R}^q$ and $\mu^* \in \mathbb{R}^p$ so that the following conditions hold:

- (Primal feasibility) $Bx^* = d$ and $Ax^* \le b$; and
- (Dual feasibility) $B^T\mu^* + A^T\lambda^* + c = 0$ and $\lambda^* \ge 0$; and
- (Complementary slackness) ${\lambda^*}^T(Ax^* - b) = 0$.

KKT conditions for LP

Theorem 6.1: (Karush-Kuhn-Tucker conditions for the LP)
Consider the linear program

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & c^T x \\ \text{subject to} \quad & Bx = d, \\ & Ax \leq b. \end{aligned} \tag{2}$$

where $c \in \mathbb{R}^n$, $B \in \mathbb{R}^{p \times n}$ and $A \in \mathbb{R}^{q \times n}$. Then $x^* \in \mathbb{R}^n$ is an optimal solution of (2) if and only if there exist $\lambda^* \in \mathbb{R}^q$ and $\mu^* \in \mathbb{R}^p$ so that the following conditions hold:

- (Primal feasibility) $Bx^* = d$ and $Ax^* \leq b$; and
- (Dual feasibility) $B^T \mu^* + A^T \lambda^* + c = 0$ and $\lambda^* \geq 0$; and
- (Complementary slackness) $\lambda^{*T}(Ax^* - b) = 0$.

Remark: Since $\lambda^* \geq 0$ and $Ax^* \leq b$, we note that $\lambda^{*T}(Ax^* - b) = 0$ means $\lambda_i^*(Ax^* - b)_i = 0$ for each $i = 1, \ldots, q$.

KKT conditions for LP cont.

Proof of Theorem 6.1 sketch: Suppose that x^* solves (2). Then x^* is primal feasible. Moreover, by strong duality, the dual problem

$$\begin{array}{ll} \underset{\mu\in\mathbb{R}^p,\lambda\in\mathbb{R}^q}{\text{Maximize}} & -d^T\mu - b^T\lambda \\ \text{subject to} & B^T\mu + A^T\lambda + c = 0, \\ & \lambda \geq 0, \end{array}$$

has solutions $\mu^* \in \mathbb{R}^p$ and $\lambda^* \in \mathbb{R}^q$ satisfying the dual feasibility condition; furthermore, $c^T x^* = -d^T\mu^* - b^T\lambda^*$. Then

$$\lambda^{*T}(Ax^* - b) = (x^*)^T A^T\lambda^* - b^T\lambda^* = (x^*)^T(-c - B^T\mu^*) - b^T\lambda^*$$
$$= -c^T x^* - (Bx^*)^T\mu^* - b^T\lambda^* = -c^T x^* - d^T\mu^* - b^T\lambda^* = 0.$$

KKT conditions for LP cont.

Proof of Theorem 6.1 sketch: Suppose that x^* solves (2). Then x^* is primal feasible. Moreover, by strong duality, the dual problem

$$\begin{array}{ll} \underset{\mu \in \mathbb{R}^p, \lambda \in \mathbb{R}^q}{\text{Maximize}} & -d^T\mu - b^T\lambda \\ \text{subject to} & B^T\mu + A^T\lambda + c = 0, \\ & \lambda \geq 0, \end{array}$$

has solutions $\mu^* \in \mathbb{R}^p$ and $\lambda^* \in \mathbb{R}^q$ satisfying the dual feasibility condition; furthermore, $c^Tx^* = -d^T\mu^* - b^T\lambda^*$. Then

$$\begin{aligned} {\lambda^*}^T(Ax^* - b) &= (x^*)^T A^T\lambda^* - b^T\lambda^* = (x^*)^T(-c - B^T\mu^*) - b^T\lambda^* \\ &= -c^Tx^* - (Bx^*)^T\mu^* - b^T\lambda^* = -c^Tx^* - d^T\mu^* - b^T\lambda^* = 0. \end{aligned}$$

Conversely, if the three conditions are satisfied, then x^* and (μ^*, λ^*) are feasible for the primal and dual problems respectively. A similar calculation as above shows that $c^Tx^* = -d^T\mu^* - b^T\lambda^*$. Then strong duality shows that x^* and (μ^*, λ^*) are indeed optimal.

Linearizing nonlinear regions?

Back to our problem (1):

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{subject to} & h_j(x) = 0, \quad j \in J, \\ & g_i(x) \leq 0, \quad i \in I. \end{array}$$

Aim: Find conditions to help characterize local minimizers!

Linearizing nonlinear regions?

Back to our problem (1):

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{subject to} & h_j(x) = 0, \quad j \in J, \\ & g_i(x) \leq 0, \quad i \in I. \end{array}$$

Aim: Find conditions to help characterize local minimizers!

Let x^* be a local minimizer of (1).

- Locally approximate f by $f(x^*) + [\nabla f(x^*)]^T(x - x^*)$?

Linearizing nonlinear regions?

Back to our problem (1):

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{subject to} & h_j(x) = 0, \quad j \in J, \\ & g_i(x) \leq 0, \quad i \in I. \end{array}$$

Aim: Find conditions to help characterize local minimizers!

Let x^* be a local minimizer of (1).

- Locally approximate f by $f(x^*) + [\nabla f(x^*)]^T(x - x^*)$?
- Locally approximate $h_j = 0$ by $h_j(x^*) + [\nabla h_j(x^*)]^T(x - x^*) = 0$?

Linearizing nonlinear regions?

Back to our problem (1):

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{subject to} & h_j(x) = 0, \quad j \in J, \\ & g_i(x) \leq 0, \quad i \in I. \end{array}$$

Aim: Find conditions to help characterize local minimizers!

Let x^* be a local minimizer of (1).

- Locally approximate f by $f(x^*) + [\nabla f(x^*)]^T(x - x^*)$?
- Locally approximate $h_j = 0$ by $h_j(x^*) + [\nabla h_j(x^*)]^T(x - x^*) = 0$?
- Define $I(x^*) := \{i \in I : \ g_i(x^*) = 0\}$. (Active index set)

Linearizing nonlinear regions?

Back to our problem (1):

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{subject to} & h_j(x) = 0, \quad j \in J, \\ & g_i(x) \leq 0, \quad i \in I. \end{array}$$

Aim: Find conditions to help characterize local minimizers!

Let x^* be a local minimizer of (1).

- Locally approximate f by $f(x^*) + [\nabla f(x^*)]^T(x - x^*)$?
- Locally approximate $h_j = 0$ by $h_j(x^*) + [\nabla h_j(x^*)]^T(x - x^*) = 0$?
- Define $I(x^*) := \{i \in I : \ g_i(x^*) = 0\}$. (Active index set)

  For each $i \in I(x^*)$,

  locally approximate $g_i \leq 0$ by $g_i(x^*) + [\nabla g_i(x^*)]^T(x - x^*) \leq 0$?

Linearizing nonlinear regions? cont.

The resulting approximating LP:

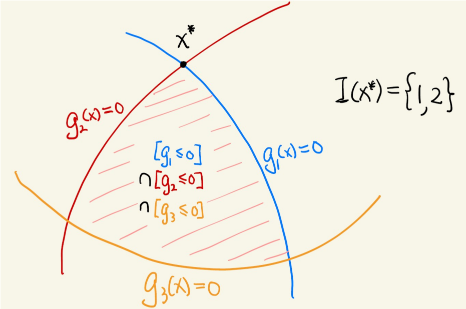
$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & [\nabla f(x^*)]^T (x - x^*) \\ \text{subject to} & h_j(x^*) + [\nabla h_j(x^*)]^T (x - x^*) = 0, \quad j \in J, \\ & g_i(x^*) + [\nabla g_i(x^*)]^T (x - x^*) \leq 0, \quad i \in I(x^*). \end{array}$$

Linearizing nonlinear regions? cont.

The resulting approximating LP:

$$\begin{array}{ll}\underset{x\in\mathbb{R}^n}{\text{Minimize}} & [\nabla f(x^*)]^T(x-x^*)\\ \text{subject to} & h_j(x^*)+[\nabla h_j(x^*)]^T(x-x^*)=0,\quad j\in J,\\ & g_i(x^*)+[\nabla g_i(x^*)]^T(x-x^*)\le 0,\quad i\in I(x^*).\end{array}$$

Why $I(x^*)$?



MFCQ

When is the approximation good?

MFCQ

When is the approximation good?

Definition: (Mangasarian-Fromovitz constraint qualification, MFCQ)
Consider the feasible set of (1) and let x^* be feasible. We say that the Mangasarian-Fromovitz constraint qualification (MFCQ) holds at x^* if the following conditions hold:

- **if**

$$\sum_{i \in J} \mu_j \nabla h_j(x^*) + \sum_{i \in I(x^*)} \lambda_i \nabla g_i(x^*) = 0 \text{ and } \lambda_i \geq 0 \; \forall i \in I(x^*),$$

then $\lambda_i = 0$ for all $i \in I(x^*)$ and $\mu_j = 0$ for all $j \in J$.

Remark:

- If $g_i(x^*) < 0$ for all $i \in I$ so that $I(x^*) = \emptyset$, then MFCQ means $\{\nabla h_j(x^*) : \; j \in J\}$ is linearly independent.
- If $J = \emptyset$, then MFCQ means $\{\nabla g_i(x^*) : \; i \in I(x^*)\}$ is positively independent.

KKT conditions for NLP

Theorem 6.2: (KKT conditions for NLP)
Consider (1) and let x^* be a local minimizer. Suppose that MFCQ holds at x^*. Then there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ so that

- $\nabla f(x^*) + \sum_{i \in I} \lambda_i^* \nabla g_i(x^*) + \sum_{j \in J} \mu_j^* \nabla h_j(x^*) = 0$; and
- $\lambda_i^* \geq 0$ and $\lambda_i^* g_i(x^*) = 0$ for all $i \in I$.

KKT conditions for NLP

Theorem 6.2: (KKT conditions for NLP)
Consider (1) and let x^* be a local minimizer. Suppose that MFCQ holds at x^*. Then there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ so that

- $\nabla f(x^*) + \sum_{i \in I} \lambda_i^* \nabla g_i(x^*) + \sum_{j \in J} \mu_j^* \nabla h_j(x^*) = 0$; and
- $\lambda_i^* \geq 0$ and $\lambda_i^* g_i(x^*) = 0$ for all $i \in I$.

Remarks:

- Under MFCQ, the approximating LP is "good" around x^*. We hence look at the KKT of this LP.
- The first bullet point follows from the dual feasibility condition of the approximating LP, and by defining $\lambda_i^* = 0$ for $i \notin I(x^*)$.
- The second bullet point follows from the definition of $I(x^*)$ and by defining $\lambda_i^* = 0$ for $i \notin I(x^*)$.

KKT conditions for NLP cont.

Remarks cont.:

- The λ^* and μ^* are called Lagrange multipliers at x^*.
- We consider the so-called stationary points:

Definition: (Stationary points)
Consider (1). An $\bar{x}$ is called a stationary point of (1) if it is feasible and there exist $\bar{\lambda} \in \mathbb{R}^m$ and $\bar{\mu} \in \mathbb{R}^p$ so that

- $\nabla f(\bar{x}) + \sum_{i \in I} \bar{\lambda}_i \nabla g_i(\bar{x}) + \sum_{j \in J} \bar{\mu}_j \nabla h_j(\bar{x}) = 0$; and
- $\bar{\lambda} \geq 0$ and $\bar{\lambda}_i g_i(\bar{x}) = 0$ for all $i \in I$.

- According to Theorem 6.2, if x^* is a local minimizer and if the MFCQ holds at x^*, then x^* is a stationary point.

KKT conditions for NLP cont.

Remarks cont.:

- We refer to the following conditions as KKT conditions:

> KKT conditions for (1):
>
> ⋆ $g_i(x) \leq 0$ for all $i \in I$ and $h_j(x) = 0$ for all $j \in J$; and
>
> ⋆ $\nabla f(x) + \sum_{i \in I} \lambda_i \nabla g_i(x) + \sum_{j \in J} \mu_j \nabla h_j(x) = 0$; and
>
> ⋆ $\lambda \geq 0$ and $\lambda_i g_i(x) = 0$ for all $i \in I$.

- According to Theorem 6.2, if x^* is a local minimizer and if the MFCQ holds at x^*, then there exist $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ so that (x^*, λ^*, μ^*) satisfies the KKT conditions.

Example 1

Is MFCQ essential for the existence of Lagrange multipliers?

Example: Consider the following one-dimensional optimization problem:

$$\begin{array}{rl} \underset{x \in \mathbb{R}}{\text{Minimize}} & x \\ \text{subject to} & x^2 = 0. \end{array}$$

Clearly, 0 is the global minimizer. If a Lagrange multiplier exists for $x = 0$, then

$$0 = \left(\frac{d}{dx} x + \mu \frac{d}{dx} x^2 \right) \Big|_{x=0}$$

Example 1

Is MFCQ essential for the existence of Lagrange multipliers?

Example: Consider the following one-dimensional optimization problem:

$$\begin{aligned} \underset{x \in \mathbb{R}}{\text{Minimize}} \quad & x \\ \text{subject to} \quad & x^2 = 0. \end{aligned}$$

Clearly, 0 is the global minimizer. If a Lagrange multiplier exists for $x = 0$, then

$$0 = \left(\frac{d}{dx} x + \mu \frac{d}{dx} x^2 \right) \bigg|_{x=0} = 1,$$

leading to a contradiction. Thus, no Lagrange multiplier exists for the global minimizer $x = 0$.

Notice that MFCQ fails: $\frac{d}{dx} x^2 = 2x$, which equals 0 at $x = 0$.

Example 2

Example: Consider the following optimization problem.

$$\begin{array}{rl} \text{Minimize} & x_1 x_2 x_3 \\ \text{subject to} & x_1^2 + x_2^2 + x_3^2 = 1. \end{array}$$

1. Show that the MFCQ holds at every point in the feasible set.
2. Find all stationary points that do not have zero coordinates.

Solution: Let $h(x) = x_1^2 + x_2^2 + x_3^2 - 1$. Then

$$\nabla h(x) = 2x.$$

Hence $\nabla h(x) \neq 0$ at every feasible x because $\|x\|_2 = 1$, which implies $x \neq 0$. Thus, MFCQ holds at every point in the feasible set.

Example 2 cont.

Solution cont.: To find all stationary points, consider the KKT conditions:

$$\begin{cases} x_1^2 + x_2^2 + x_3^2 = 1, \\ x_2 x_3 + 2\mu x_1 = 0, \\ x_1 x_3 + 2\mu x_2 = 0, \\ x_1 x_2 + 2\mu x_3 = 0. \end{cases}$$

Then

$$\begin{cases} x_1^2 + x_2^2 + x_3^2 = 1, \\ x_2 x_3 + 2\mu x_1 = 0, \\ x_1 x_3 + 2\mu x_2 = 0, \\ x_1 x_2 + 2\mu x_3 = 0. \end{cases} \Longrightarrow \begin{cases} x_1^2 + x_2^2 + x_3^2 = 1, \\ x_1 x_2 x_3 + 2\mu x_1^2 = 0, \\ x_1 x_2 x_3 + 2\mu x_2^2 = 0, \\ x_1 x_2 x_3 + 2\mu x_3^2 = 0. \end{cases} \Longrightarrow 3x_1 x_2 x_3 + 2\mu = 0.$$

We only want (x_1, x_2, x_3) to be all nonzero. Hence, $\mu \neq 0$ and $x_1 x_2 x_3 = -\frac{2\mu}{3}$.

Example 2 cont.

Solution cont.:
From $x_2x_3 + 2\mu x_1 = 0$ and $x_1x_2x_3 = -\frac{2\mu}{3}$, we get

$$\frac{x_1x_2x_3}{x_1} + 2\mu x_1 = 0 \;\Rightarrow\; -\frac{2\mu}{3x_1^2} + 2\mu = 0 \;\Rightarrow\; 2\mu\left(1 - \frac{1}{3x_1^2}\right) = 0.$$

Hence, $x_1 = \pm\frac{1}{\sqrt{3}}$. Similarly, $x_2 = \pm\frac{1}{\sqrt{3}}$ and $x_3 = \pm\frac{1}{\sqrt{3}}$. Thus, stationary points with only nonzero entries are

$$\left(\pm\frac{1}{\sqrt{3}}, \pm\frac{1}{\sqrt{3}}, \pm\frac{1}{\sqrt{3}}\right).$$

Example 3

Example: Consider the following optimization problem.

$$\begin{array}{rl} \text{Minimize} & 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \\ \text{subject to} & x_1^2 + x_2^2 - 5 \leq 0, \\ & 3x_1 + x_2 - 6 \leq 0. \end{array}$$

(a) Find all stationary points and the corresponding multipliers.

(b) Verify that the MFCQ holds at each stationary point found in (a).

Solution: The KKT conditions are

$$\begin{cases} 4x_1 + 2x_2 - 10 + 2\lambda_1 x_1 + 3\lambda_2 = 0, \\ 2x_1 + 2x_2 - 10 + 2\lambda_1 x_2 + \lambda_2 = 0, \\ x_1^2 + x_2^2 - 5 \leq 0, \;\; 3x_1 + x_2 - 6 \leq 0, \;\; \lambda_1, \lambda_2 \geq 0, \\ \lambda_1(x_1^2 + x_2^2 - 5) = 0, \;\; \lambda_2(3x_1 + x_2 - 6) = 0. \end{cases}$$

Example 3 cont.

Solution cont.:

Case 1: $\lambda_1 = \lambda_2 = 0$. By solving

$$\begin{cases} 4x_1 + 2x_2 - 10 = 0, \\ 2x_1 + 2x_2 - 10 = 0, \end{cases}$$

we get $x = (0, 5)$, which violates the first constraint. Thus, this case cannot happen.

Example 3 cont.

Solution cont.:

Case 1: $\lambda_1 = \lambda_2 = 0$. By solving

$$\begin{cases} 4x_1 + 2x_2 - 10 = 0, \\ 2x_1 + 2x_2 - 10 = 0, \end{cases}$$

we get $x = (0, 5)$, which violates the first constraint. Thus, this case cannot happen.

Case 2: $\lambda_1 > 0, \lambda_2 = 0$. We first consider

$$\begin{cases} 4x_1 + 2x_2 - 10 + 2\lambda_1 x_1 = 0, \\ 2x_1 + 2x_2 - 10 + 2\lambda_1 x_2 = 0, \\ x_1^2 + x_2^2 - 5 = 0, \end{cases}$$

giving $x = (1, 2)$ and $\lambda_1 = 1$. Note that $x = (1, 2)$ is indeed feasible (CHECK!). Hence, $x = (1, 2)$ is stationary, with multipliers $(1, 0)$.

Example 3 cont.

Solution cont.: Case 3: $\lambda_1 = 0, \lambda_2 > 0$. By solving

$$\begin{cases} 4x_1 + 2x_2 - 10 + 3\lambda_2 = 0, \\ 2x_1 + 2x_2 - 10 + \lambda_2 = 0, \\ 3x_1 + x_2 - 6 = 0, \end{cases}$$

we get $x = (0.4, 4.8)$ and $\lambda_2 = -0.4 < 0$. This case cannot happen.

Example 3 cont.

Solution cont.: Case 3: $\lambda_1 = 0, \lambda_2 > 0$. By solving

$$\begin{cases} 4x_1 + 2x_2 - 10 + 3\lambda_2 = 0, \\ 2x_1 + 2x_2 - 10 + \lambda_2 = 0, \\ 3x_1 + x_2 - 6 = 0, \end{cases}$$

we get $x = (0.4, 4.8)$ and $\lambda_2 = -0.4 < 0$. This case cannot happen.

Case 4: $\lambda_1 > 0, \lambda_2 > 0$. We first consider

$$\begin{cases} 4x_1 + 2x_2 - 10 + 2\lambda_1 x_1 + 3\lambda_2 = 0, \\ 2x_1 + 2x_2 - 10 + 2\lambda_1 x_2 + \lambda_2 = 0, \\ x_1^2 + x_2^2 - 5 = 0, \quad 3x_1 + x_2 - 6 = 0. \end{cases}$$

The last 2 equations give $x \approx (2.17, -0.52)$ or $x \approx (1.43, 1.72)$. But

- $x \approx (2.17, -0.52)$ implies $\lambda_1 \approx -2.37 < 0$;
- $x \approx (1.43, 1.72)$ implies $\lambda_2 \approx -1.02 < 0$.

Thus, this case also cannot happen.

Example 3 cont.

Solution cont.: Thus, there is a unique stationary point $x^* = (1, 2)$.
Let $g_1(x) = x_1^2 + x_2^2 - 5$ and $g_2(x) = 3x_1 + x_2 - 6$, then

$$g_1(x^*) = 0 \ \text{ and } \ g_2(x^*) = 3 \cdot 1 + 2 - 6 < 0.$$

Thus, $I(x^*) = \{1\}$. Also, $\nabla g_1(x^*) = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \neq 0$. Thus, $\{\nabla g_1(x^*)\}$ is linearly independent, and hence in particular, positively independent. The MFCQ holds at x^*.

Example 3 cont.

Solution cont.: Thus, there is a unique stationary point $x^* = (1, 2)$. Let $g_1(x) = x_1^2 + x_2^2 - 5$ and $g_2(x) = 3x_1 + x_2 - 6$, then

$$g_1(x^*) = 0 \text{ and } g_2(x^*) = 3 \cdot 1 + 2 - 6 < 0.$$

Thus, $I(x^*) = \{1\}$. Also, $\nabla g_1(x^*) = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \neq 0$. Thus, $\{\nabla g_1(x^*)\}$ is linearly independent, and hence in particular, positively independent. The MFCQ holds at x^*.

Remark: There is a unique stationary point in this example. Since a global minimizer exists (note that the feasible set is compact), can we conclude immediately that $(1, 2)$ is globally optimal?

Example 3 cont.

Solution cont.: Thus, there is a unique stationary point $x^* = (1,2)$. Let $g_1(x) = x_1^2 + x_2^2 - 5$ and $g_2(x) = 3x_1 + x_2 - 6$, then

$$g_1(x^*) = 0 \ \text{ and } \ g_2(x^*) = 3 \cdot 1 + 2 - 6 < 0.$$

Thus, $I(x^*) = \{1\}$. Also, $\nabla g_1(x^*) = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \neq 0$. Thus, $\{\nabla g_1(x^*)\}$ is linearly independent, and hence in particular, positively independent. The MFCQ holds at x^*.

Remark: There is a unique stationary point in this example. Since a global minimizer exists (note that the feasible set is compact), can we conclude immediately that $(1,2)$ is globally optimal?

- GAP: To make such deduction, we need to make sure that global minimizers are stationary points by, for example, showing that the MFCQ holds at the global minimizer(s). — Can we check MFCQ for the whole feasible set easily?

Slater's condition

Theorem 6.3: (MFCQ from Slater)
Consider the set defined by

$$S := \{x \in \mathbb{R}^n : \; g_i(x) \leq 0 \;\; \forall i \in I\},$$

where g_i are convex C^1. Suppose that there exists $\bar{x}$ satisfying

$$g_i(\bar{x}) < 0 \;\; \forall i \in I.$$

Then MFCQ holds at every point in S.

Remark:

- The set S in the above theorem is closed and convex.
- The condition that "there exists $\bar{x}$ satisfying $g_i(\bar{x}) < 0$ for all $i \in I$" is called the Slater's condition. The $\bar{x}$ is called a Slater point.
- One can indeed show that for the above S, the MFCQ holds at every point in S if and only if Slater's condition holds.

Slater's condition cont.

Proof of Theorem 6.3: Let $x \in S$. If $I(x) = \emptyset$, then the MFCQ holds at x. Thus, assume that $I(x) \neq \emptyset$.

For each $i \in I(x)$, since $g_i(x) = 0$, for all $t \in [0, 1]$, we have

$$g_i(x + t(\bar{x} - x)) \leq tg_i(\bar{x}) + (1 - t)g_i(x) = tg_i(\bar{x}).$$

Thus,

$$[\nabla g_i(x)]^T(\bar{x} - x) = \lim_{t \downarrow 0} \frac{g_i(x + t(\bar{x} - x)) - g_i(x)}{t} \leq g_i(\bar{x}) < 0.$$

Slater's condition cont.

Proof of Theorem 6.3: Let $x \in S$. If $I(x) = \emptyset$, then the MFCQ holds at x. Thus, assume that $I(x) \neq \emptyset$.

For each $i \in I(x)$, since $g_i(x) = 0$, for all $t \in [0, 1]$, we have

$$g_i(x + t(\bar{x} - x)) \leq tg_i(\bar{x}) + (1 - t)g_i(x) = tg_i(\bar{x}).$$

Thus,

$$[\nabla g_i(x)]^T(\bar{x} - x) = \lim_{t \downarrow 0} \frac{g_i(x + t(\bar{x} - x)) - g_i(x)}{t} \leq g_i(\bar{x}) < 0.$$

Suppose $\lambda_i \geq 0$, $i \in I(x)$ are such that $\sum_{i \in I(x)} \lambda_i \nabla g_i(x) = 0$. Then

$$0 = \sum_{i \in I(x)} \lambda_i [\nabla g_i(x)]^T(\bar{x} - x) \leq \sum_{i \in I(x)} \lambda_i g_i(\bar{x}),$$

forcing $\lambda_i = 0$ for all $i \in I(x)$.

Slater's condition cont.

Proof of Theorem 6.3: Let $x \in S$. If $I(x) = \emptyset$, then the MFCQ holds at x. Thus, assume that $I(x) \neq \emptyset$.

For each $i \in I(x)$, since $g_i(x) = 0$, for all $t \in [0,1]$, we have

$$g_i(x + t(\bar{x} - x)) \leq t g_i(\bar{x}) + (1-t) g_i(x) = t g_i(\bar{x}).$$

Thus,

$$[\nabla g_i(x)]^T(\bar{x} - x) = \lim_{t \downarrow 0} \frac{g_i(x + t(\bar{x} - x)) - g_i(x)}{t} \leq g_i(\bar{x}) < 0.$$

Suppose $\lambda_i \geq 0$, $i \in I(x)$ are such that $\sum_{i \in I(x)} \lambda_i \nabla g_i(x) = 0$. Then

$$0 = \sum_{i \in I(x)} \lambda_i [\nabla g_i(x)]^T(\bar{x} - x) \leq \sum_{i \in I(x)} \lambda_i g_i(\bar{x}),$$

forcing $\lambda_i = 0$ for all $i \in I(x)$. Can you prove the converse as well?

Remark

Back to Example 3 on Slide 15.

- Let $g_1(x) = x_1^2 + x_2^2 - 5$ and $g_2(x) = 3x_1 + x_2 - 6$. Then g_1 is convex because $\nabla^2 g_1(x) = 2I \succ 0$ for all x, and g_2 is convex since it is affine. Moreover,

$$g_1(0) = -5 < 0, \quad g_2(0) = -6 < 0.$$

  Thus, $\bar{x} = (0, 0)$ is a Slater point.
- Using Theorem 6.3, we conclude that MFCQ holds at every point in the feasible set; in particular, at global minimizers.
- Thus, Theorem 6.2 shows that all global minimizers are stationary.
- Since there is only one stationary point in this example, it must be the global minimizer. Recall that a global minimizer must exist because the feasible set is compact.

Generalized Slater's condition

Theorem 6.4: (MFCQ from generalized Slater)
Consider the set defined by

$$S := \{x \in \mathbb{R}^n : \; g_i(x) \leq 0 \;\; \forall i \in I, \; Ax = b\},$$

where g_i are convex C^1 and $A \in \mathbb{R}^{p \times n}$. Suppose that there exists $\bar{x}$ satisfying

$$A\bar{x} = b, \;\; g_i(\bar{x}) < 0 \;\; \forall i \in I,$$

and A has full row rank. Then MFCQ holds at every point in S.

Remark:

- The set S in the above theorem is closed and convex.
- The condition that "there exists $\bar{x}$ satisfying $g_i(\bar{x}) < 0$ for all $i \in I$, $A\bar{x} = b$ and A has full row rank" is called the generalized Slater's condition. The $\bar{x}$ is called a generalized Slater point.
- One can indeed show that for the above S, MFCQ holds at every point in S if and only if generalized Slater's condition holds.

Role of convexity

Consider the following special instance of (1)

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & f(x) \\ \text{subject to} & Ax = b, \\ & g_i(x) \le 0, \quad i \in I, \end{array} \tag{3}$$

here f and g_i are all convex C^1 functions, $A \in \mathbb{R}^{p\times n}$.

Theorem 6.5: (Sufficiency under convexity)
Consider (3). Suppose that there exist $x^* \in \mathbb{R}^n$, $\lambda^* \in \mathbb{R}^m$ and $\mu^* \in \mathbb{R}^p$ so that

- $Ax^* = b$ and $g_i(x^*) \le 0$ for all $i \in I$; and
- $\nabla f(x^*) + \sum_{i\in I} \lambda_i^* \nabla g_i(x^*) + A^T \mu^* = 0$; and
- $\lambda_i^* \ge 0$ and $\lambda_i^* g_i(x^*) = 0$ for all $i \in I$.

Then x^* is a global minimizer of (3).

Role of convexity cont.

Proof of Theorem 6.5 sketch: We make use of the following subdifferential inequality for convex functions:

> Subdifferential inequality: Let h be convex C^1, $x,\ y \in \mathbb{R}^n$. Then
> $$h(y) - h(x) \geq [\nabla h(x)]^T (y - x).$$

Then, for any feasible x, we have

$$\begin{aligned} f(x) - f(x^*) &\geq [\nabla f(x^*)]^T (x - x^*) \\ &= -\left[\sum_{i \in I} \lambda_i^* \nabla g_i(x^*) + A^T \mu^*\right]^T (x - x^*) \\ &= -\left[\sum_{i \in I} \lambda_i^* \nabla g_i(x^*)\right]^T (x - x^*) - \mu^{*T} \underbrace{(Ax - Ax^*)}_{= b - b = 0}. \end{aligned}$$

Role of convexity cont.

Proof of Theorem 6.5 sketch cont.: For each $i \in I$, since $\lambda_i^* \geq 0$, we have in view of the subdifferential inequality that

$$\lambda_i^*(g_i(x) - g_i(x^*)) \geq \lambda_i^*[\nabla g_i(x^*)]^T(x - x^*).$$

Hence,

$$\begin{aligned} f(x) - f(x^*) &\geq -\sum_{i \in I} \lambda_i^*[\nabla g_i(x^*)]^T(x - x^*) \\ &\geq -\sum_{i \in I} \lambda_i^*[g_i(x) - g_i(x^*)] \\ &= -\sum_{i \in I} \lambda_i^* g_i(x) + \sum_{i \in I} \underbrace{\lambda_i^* g_i(x^*)}_{=0} \\ &\geq 0, \end{aligned}$$

since $g_i(x) \leq 0$ for all $i \in I$.

Remark

Back to Example 3 on Slide 15.

- Let $f(x) = 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2$, $g_1(x) = x_1^2 + x_2^2 - 5$ and $g_2(x) = 3x_1 + x_2 - 6$.

  Then g_1 is convex because $\nabla^2 g_1(x) = 2I \succ 0$ for all x; g_2 is convex since it is affine; f is also convex because for all x,

$$\nabla^2 f(x) = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} \succ 0.$$

- Using Theorem 6.5, any stationary point is globally optimal.
- Since $x = (1, 2)$ is a stationary point in this example, it must be a global minimizer.

Example 4

Consider the following optimization problem.

$$\begin{array}{rl} \text{Minimize} & x_1 \\ \text{subject to} & (x_1 - 1)^2 + (x_2 - 1)^2 - 1 \leq 0, \\ & (x_1 - 1)^2 + (x_2 + 1)^2 - 1 \leq 0. \end{array}$$

1. Show that the feasible set is a singleton set.
2. Show that the above problem does not have stationary points.

Example 4

Consider the following optimization problem.

$$\begin{array}{rl} \text{Minimize} & x_1 \\ \text{subject to} & (x_1-1)^2+(x_2-1)^2-1 \leq 0, \\ & (x_1-1)^2+(x_2+1)^2-1 \leq 0. \end{array}$$

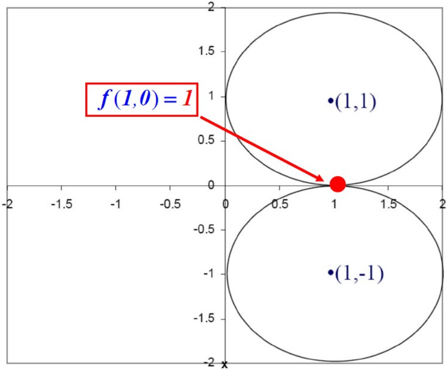
1. Show that the feasible set is a singleton set.
2. Show that the above problem does not have stationary points.

Solution: Let $f(x) = x_1$, $g_1(x) = (x_1-1)^2+(x_2-1)^2-1$ and $g_2(x) = (x_1-1)^2+(x_2+1)^2-1$. Then

$$\nabla f(x) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \ \nabla g_1(x) = \begin{bmatrix} 2(x_1-1) \\ 2(x_2-1) \end{bmatrix}, \ \nabla g_2(x) = \begin{bmatrix} 2(x_1-1) \\ 2(x_2+1) \end{bmatrix}$$

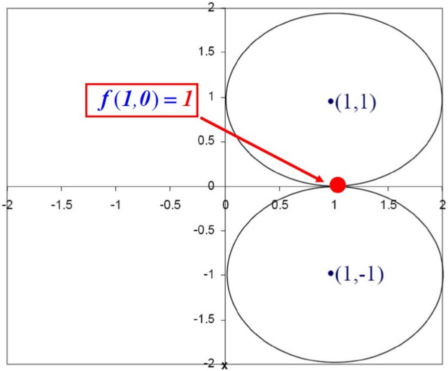
Example 4 cont.

Solution cont.: The feasible set consists only of the point $x^* = (1, 0)$.



Example 4 cont.

Solution cont.: The feasible set consists only of the point $x^* = (1, 0)$.



But $[\nabla g_1(x^*)]_1 = [\nabla g_2(x^*)]_1 = 0$ and $[\nabla f(x^*)]_1 = 1$. Hence, there does not exist $\lambda^* \geq 0$ so that $\nabla f(x^*) + \lambda_1^* \nabla g_1(x^*) + \lambda_2^* \nabla g_2(x^*) = 0$. Thus, x^* is not stationary. There is no stationary point.

Example 5

Example: Consider the following optimization problem.

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{Minimize}} & \frac{1}{2}\|x\|_2^2 \\ \text{subject to} & Ax = b, \end{array}$$

where $A \in \mathbb{R}^{m \times n}$, $m \ll n$, and $b \in \mathbb{R}^m$. Suppose that A has full row rank and the feasible set is nonempty. Write down the KKT conditions and find all stationary points.

Example 5

Example: Consider the following optimization problem.

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{Minimize}} & \frac{1}{2}\|x\|_2^2 \\ \text{subject to} & Ax = b, \end{array}$$

where $A \in \mathbb{R}^{m\times n}$, $m \ll n$, and $b \in \mathbb{R}^m$. Suppose that A has full row rank and the feasible set is nonempty. Write down the KKT conditions and find all stationary points.

Solution: The KKT conditions are

$$x + A^T\mu = 0 \text{ and } Ax = b.$$

Multiplying both sides of the first equality from the left by A, we get $b = -AA^T\mu$. Since A has full row rank, AA^T is invertible. Thus, $\mu = -(AA^T)^{-1}b$ and the unique stationary point is given by $A^T(AA^T)^{-1}b$.

Set representation affects CQ I

Example: Consider $C := \{x \in \mathbb{R}^2 : \; x_1 \leq 0, x_2 \leq 0\}$. Notice that

$$C = \{x \in \mathbb{R}^2 : \; x_1 \leq 0, x_2 \leq 0\} = \{x \in \mathbb{R}^2 : \; (x_1)_+^2 + (x_2)_+^2 \leq 0\}.$$

$g_1(x) = x_1,\; g_2(x) = x_2$	$g_1(x) = (x_1)_+^2 + (x_2)_+^2$
g_1, g_2 are convex	g_1 is convex
$(-1,-1)$ is a Slater point	$\nabla g_1(x) \equiv 0$ in the feasible set
MFCQ holds everywhere in C	MFCQ fails everywhere in C

Set representation affects CQ II

Example: Consider the following set

$$\widetilde{S} = \{x \in \mathbb{R} : \; x^2 - x \le 0, x \ge 0\} = [0, 1].$$

Let $g_1(x) = x^2 - x$ and $g_2(x) = -x$. Then

- $g_1(0) = g_2(0) = 0$; and
- $g_1'(0) = g_2'(0) = -1$.

Hence, $I(0) = \{1, 2\}$. Since

$$\lambda_1 g_1'(0) + \lambda_2 g_2'(0) = 0 \text{ and } \lambda_i \ge 0, i = 1, 2, \implies \lambda_i = 0, i = 1, 2,$$

the MFCQ holds at 0.

Set representation affects CQ II

Example: Consider the following set

$$\widetilde{S} = \{x \in \mathbb{R} : \ x^2 - x \leq 0, x \geq 0\} = [0, 1].$$

Let $g_1(x) = x^2 - x$ and $g_2(x) = -x$. Then

- $g_1(0) = g_2(0) = 0$; and
- $g_1'(0) = g_2'(0) = -1$.

Hence, $I(0) = \{1, 2\}$. Since

$$\lambda_1 g_1'(0) + \lambda_2 g_2'(0) = 0 \text{ and } \lambda_i \geq 0, i = 1, 2, \implies \lambda_i = 0, i = 1, 2,$$

the MFCQ holds at 0.

At x satisfying $x^2 - x < 0$ and $x > 0$, $I(x) = \emptyset$ and MFCQ holds at x.

Set representation affects CQ II

Example: Consider the following set

$$\widetilde{S} = \{x \in \mathbb{R} : \; x^2 - x \leq 0, x \geq 0\} = [0, 1].$$

Let $g_1(x) = x^2 - x$ and $g_2(x) = -x$. Then

- $g_1(0) = g_2(0) = 0$; and
- $g_1'(0) = g_2'(0) = -1$.

Hence, $I(0) = \{1, 2\}$. Since

$$\lambda_1 g_1'(0) + \lambda_2 g_2'(0) = 0 \text{ and } \lambda_i \geq 0, i = 1, 2, \implies \lambda_i = 0, i = 1, 2,$$

the MFCQ holds at 0.

At x satisfying $x^2 - x < 0$ and $x > 0$, $I(x) = \emptyset$ and MFCQ holds at x.

At $x = 1$, $I(1) = \{1\}$ and $g_1'(1) = 1 \neq 0$. Hence, MFCQ holds at 1.

Set representation affects CQ II

Example: Consider the following set

$$\widetilde{S} = \{x \in \mathbb{R} : \; x^2 - x \leq 0, x \geq 0\} = [0, 1].$$

Let $g_1(x) = x^2 - x$ and $g_2(x) = -x$. Then

- $g_1(0) = g_2(0) = 0$; and
- $g_1'(0) = g_2'(0) = -1$.

Hence, $I(0) = \{1, 2\}$. Since

$$\lambda_1 g_1'(0) + \lambda_2 g_2'(0) = 0 \text{ and } \lambda_i \geq 0, i = 1, 2, \implies \lambda_i = 0, i = 1, 2,$$

the MFCQ holds at 0.

At x satisfying $x^2 - x < 0$ and $x > 0$, $I(x) = \emptyset$ and MFCQ holds at x.

At $x = 1$, $I(1) = \{1\}$ and $g_1'(1) = 1 \neq 0$. Hence, MFCQ holds at 1.

Thus, MFCQ holds at every point in $\widetilde{S}$.

Set representation affects CQ II cont.

Example cont.: Consider

$$\begin{array}{ll} \underset{x \in \mathbb{R}}{\text{Minimize}} & f(x) \\ \text{subject to} & x^2 - x \leq 0, x \geq 0. \end{array}$$

We know that MFCQ holds at every point in the feasible set.

Note that the above problem can be rewritten equivalently as

$$\begin{array}{ll} \underset{u \in \mathbb{R}}{\text{Minimize}} & f(u^2) \\ \text{subject to} & u^4 - u^2 \leq 0. \end{array}$$

However, with $g_1(u) = u^4 - u^2$, $g_1(0) = g_1'(0) = 0$ and the MFCQ fails at 0! Thus, this latter reformulation may not be desirable.