MS-E2122 - Nonlinear Optimization
Lecture 7

Fabricio Oliveira

Systems Analysis Laboratory
Department of Mathematics and Systems Analysis

Aalto University
School of Science

November 4, 2021

Outline of this lecture

Optimality for constrained problems

Optimality conditions II

- Fritz-John conditions
- Karush-Kuhn-Tucker (KKT) conditions
- Constraint qualification (CQ)

Outline of this lecture

Optimality for constrained problems

Optimality conditions II

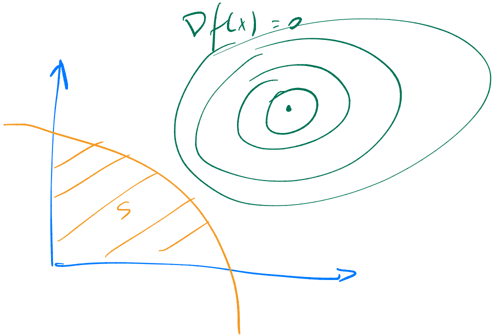
- Fritz-John conditions
- Karush-Kuhn-Tucker (KKT) conditions
- Constraint qualification (CQ)

Optimality for constrained problems

Now we examine how the set S affects the optimality conditions of

$$(P): \ \text{min.} \ \{f(x) : x \in S\}.$$

First, we define two geometric elements:



Optimality for constrained problems

Now we examine how the set S affects the optimality conditions of

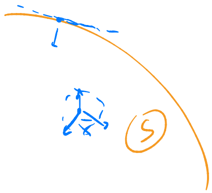
$$(P): \ \text{min. } \{f(x) : x \in S\}.$$

First, we define two geometric elements:

Definition 1 (cone of feasible directions)

Let $S \subseteq \mathbb{R}^n$ be a nonempty set, and let $\overline{x} \in \mathbf{clo}(S)$. The cone of feasible directions D at $\overline{x} \in S$ is given by

$$D = \{d : d \neq 0, \text{ and } \overline{x} + \lambda d \in S \text{ for all } \lambda \in (0, \delta) \text{ for some } \delta > 0\}$$



Optimality for constrained problems

Now we examine how the set S affects the optimality conditions of

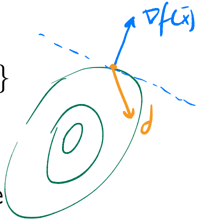
$$(P): \text{ min. } \{f(x) : x \in S\}.$$

First, we define two geometric elements:

Definition 1 (cone of feasible directions)

Let $S \subseteq \mathbb{R}^n$ be a nonempty set, and let $\overline{x} \in \mathbf{clo}(S)$. The cone of feasible directions D at $\overline{x} \in S$ is given by

$$D = \{d : d \neq 0, \text{ and } \overline{x} + \lambda d \in S \text{ for all } \lambda \in (0, \delta) \text{ for some } \delta > 0\}$$



Definition 2 (cone of descent directions)

Let $S \subseteq \mathbb{R}^n$ be a nonempty set, $f : \mathbb{R}^n \to \mathbb{R}$, and $\overline{x} \in \mathbf{clo}(S)$. The cone of improving (i.e., descent) directions F at $\overline{x} \in S$ is

$$F = \{d : f(\overline{x} + \lambda d) < f(\overline{x}) \text{ for all } \lambda \in (0, \delta) \text{ for some } \delta > 0\}.$$

Optimality for constrained sets

Geometrically, the necessary optimality condition is that $D \cap F = \emptyset$.

However, we need an algebraic representation of this condition.

Optimality for constrained sets

Geometrically, the necessary optimality condition is that $D \cap F = \emptyset$. However, we need an algebraic representation of this condition.

For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, Recall that d is a descent direction of f at $\overline{x}$ if $\nabla f(\overline{x})^\top d < 0$. Let us define

$$F_0 = \left\{ d : \nabla f(\overline{x})^\top d < 0 \right\},$$

which is an algebraic representation of F.

$D \cap F_0 = \emptyset$

Optimality for constrained sets

Geometrically, the necessary optimality condition is that $D \cap F = \emptyset$. However, we need an algebraic representation of this condition.

For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, Recall that d is a descent direction of f at $\overline{x}$ if $\nabla f(\overline{x})^\top d < 0$. Let us define

$$F_0 = \left\{ d : \nabla f(\overline{x})^\top d < 0 \right\},$$

which is an algebraic representation of F.

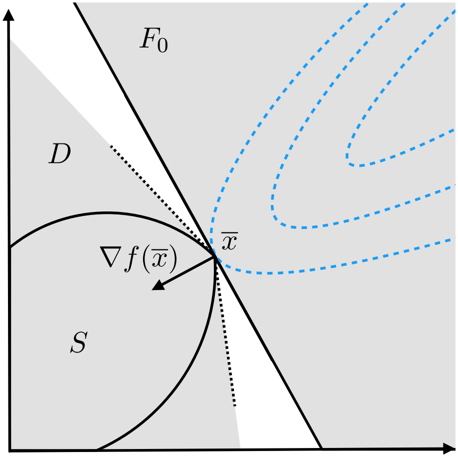
Theorem 3 (geometric necessary condition)

Let $S \subseteq \mathbb{R}^n$ be a nonempty set, and let $f : S \to \mathbb{R}$ be differentiable at $\overline{x} \in S$. If $\overline{x}$ is a local optimal solution to

$$(P) : \text{ min. } \{f(x) : x \in S\},$$

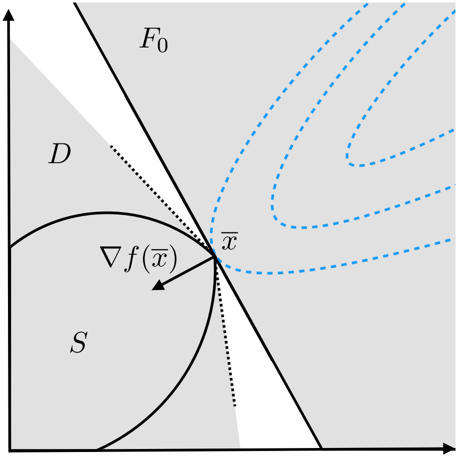
then $F_0 \cap D = \emptyset$, where $F_0 = \left\{ d : \nabla f(\overline{x})^\top d < 0 \right\}$ and D is the cone of feasible directions.

Optimality for constrained sets



Set F_0 and D. Notice that D is an open set.

Optimality for constrained sets

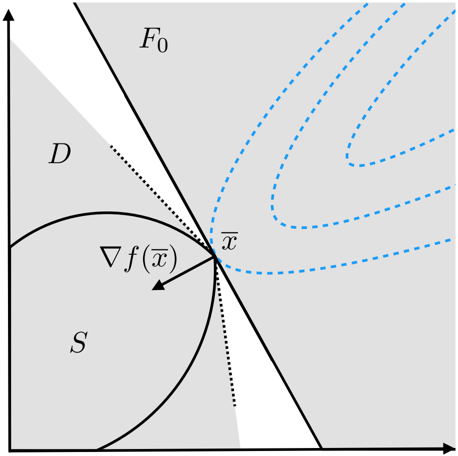


Set F_0 and D. Notice that D is an open set.

Remarks:

1. In presence of convexity, these become sufficient conditions for global optimality;

Optimality for constrained sets

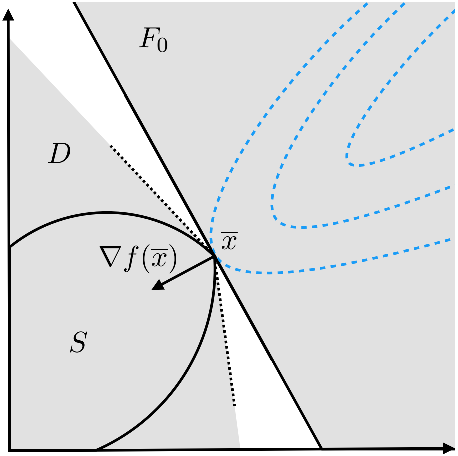


Set F_0 and D. Notice that D is an open set.

Remarks:

1. In presence of convexity, these become sufficient conditions for global optimality;
2. If f is strictly convex, it follows that $F_0 = F$;

Optimality for constrained sets



Set F_0 and D. Notice that D is an open set.

Remarks:

1. In presence of convexity, these become sufficient conditions for global optimality;
2. If f is strictly convex, it follows that $F_0 = F$;
3. If f is linear (i.e., convex and concave), it is worth considering $F_0' = \{d \neq 0 : \nabla f(\overline{x})^\top d \leq 0\}$.

Inequality constrained problems

In mathematical programming applications, S is typically expressed as a set of (in)equalities. That is, problem P is typically defined as

$$
\begin{aligned}
(P) \ : \ \text{min. } & f(x) \\
\text{subject to: } & g_i(x) \leq 0, \ i = 1, \dots, m \\
& x \in X,
\end{aligned}
$$

$g : \mathbb{R}^n \to \mathbb{R}^m$

where $g_i : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function for all $i = 1, \dots, m$ and $X \subset \mathbb{R}^n$ is a nonempty open set.

Inequality constrained problems

In mathematical programming applications, S is typically expressed as a set of (in)equalities. That is, problem P is typically defined as

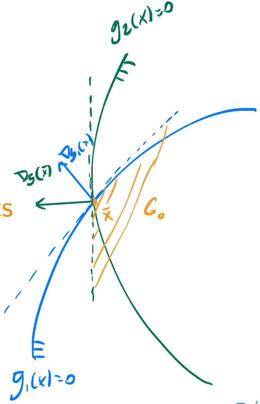
$$
\begin{aligned}
(P) \ : \ \text{min.} \ & f(x) \\
\text{subject to: } & g_i(x) \leq 0, \ i = 1, \dots, m \\
& x \in X,
\end{aligned}
$$

where $g_i : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function for all $i = 1, \dots, m$ and $X \subset \mathbb{R}^n$ is a nonempty open set.

This allows us to define a proxy G_0 for D in terms of the gradients of the binding constraints where $G_0 \subseteq D$ and defined as

active
$g_i(\bar{x}) = 0$

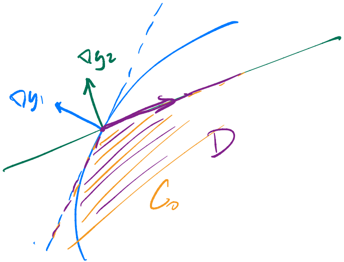
$$
G_0 = \left\{ d : \nabla g_i(\overline{x})^\top d < 0, i \in I \right\}.
$$



Inequality constrained problems

$F \cap D = \emptyset$

Since $F_0 \cap D = \emptyset$ must hold for a local optimal solution $\overline{x} \in S$, it follows that $F_0 \cap G_0 = \emptyset$ must also hold.



Inequality constrained problems

Since $F_0 \cap D = \emptyset$ must hold for a local optimal solution $\overline{x} \in S$, it follows that $F_0 \cap G_0 = \emptyset$ must also hold.

Lemma 4

Let $S = \{x \in X : g_i(x) \leq 0$ for all $i = 1, \dots, m\}$, where $X \subset \mathbb{R}^n$ is a nonempty open set and $g_i : \mathbb{R}^n \to \mathbb{R}$ a differentiable function for all $i = 1, \dots, m$. For $\overline{x} \in S$, let $I = \{i : g_i(\overline{x}) = 0\}$ be the index set of the binding (or active) constraints. Let

$$G_0 = \left\{d : \nabla g_i(\overline{x})^\top d < 0, i \in I\right\}$$

Then $G_0 \subseteq D$, where D is the cone of feasible directions.

Inequality constrained problems

Since $F_0 \cap D = \emptyset$ must hold for a local optimal solution $\overline{x} \in S$, it follows that $F_0 \cap G_0 = \emptyset$ must also hold.

Lemma 4

Let $S = \{x \in X : g_i(x) \leq 0$ for all $i = 1, \dots, m\}$, where $X \subset \mathbb{R}^n$ is a nonempty open set and $g_i : \mathbb{R}^n \to \mathbb{R}$ a differentiable function for all $i = 1, \dots, m$. For $\overline{x} \in S$, let $I = \{i : g_i(\overline{x}) = 0\}$ be the index set of the binding (or active) constraints. Let

$$G_0 = \left\{d : \nabla g_i(\overline{x})^\top d < 0, i \in I\right\}$$

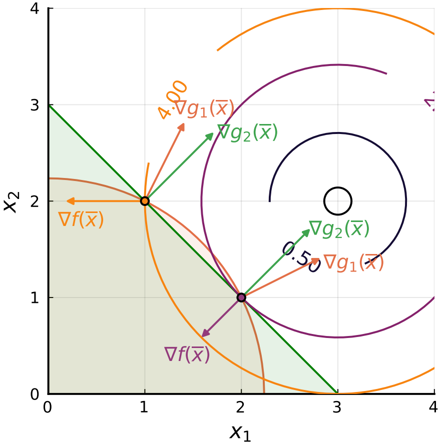
Then $G_0 \subseteq D$, where D is the cone of feasible directions.

Remark: for affine g_i, $D \subseteq G_0' = \left\{d \neq 0 : \nabla g_i(\overline{x})^\top d \leq 0, i \in I\right\}$ might be worth considering.

Inequality constrained problems

Example:

$$
\begin{aligned}
\text{min. } & (x_1-3)^2+(x_2-2)^2 \\
\text{s.t. } & x_1^2+x_2^2 \leq 5 \\
& x_1+x_2 \leq 3 \\
& x_1 \geq 0 \\
& x_2 \geq 0
\end{aligned}
$$



Outline of this lecture

Optimality for constrained problems

Optimality conditions II

- Fritz-John conditions
- Karush-Kuhn-Tucker (KKT) conditions
- Constraint qualification (CQ)

Fritz-John conditions

We will next examine how to represent $F_0 \cap G_0 = \emptyset$ in terms of the gradients of the objective function and binding constraints.

Fritz-John conditions

We will next examine how to represent $F_0 \cap G_0 = \emptyset$ in terms of the gradients of the objective function and binding constraints.

Theorem 5 (Fritz-John necessary conditions)

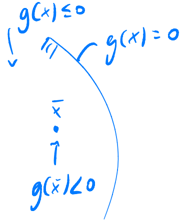
Let $X \subseteq \mathbb{R}^n$ be a nonempty open set, and let $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ be differentiable for all $i = 1, \dots, m$. Additionally, let $\overline{x}$ be feasible and $I = \{i : g_i(\overline{x}) = 0\}$. If $\overline{x}$ solves P locally, there exist scalars u_i, $i \in \{0\} \cup I$, such that

$$u_0 \nabla f(\overline{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{x}) = 0$$

$$u_i g_i(\overline{x}) = 0, \quad i = 1, \dots, m$$

$$u_i \geq 0, \cancel{g_i(\overline{x})}, \quad i = 0, \dots, m$$

$$u = (u_0, \dots, u_m) \neq 0$$



Fritz-John conditions

Proof.

Since $\overline{x}$ solves P locally, Theorem 3 guarantees that there is no d such that $\nabla f(\overline{x})^\top d < 0$ and $\nabla g_i(x)^\top d < 0$ for each $i \in I$. Let A be the matrix whose rows are $\nabla f(\overline{x})^\top$ and $\nabla g_i(\overline{x})^\top$ for $i \in I$.

Using Farkas' theorem, we have that if $Ad < 0$ is inconsistent, then there exists nonzero $p \geq 0$ such that $A^\top p = 0$. Being $I = \{i_1, \dots, i_{|I|}\}$, we let $p = (u_0, u_{i_1}, \dots, u_{i_{|I|}})$ and $u_i = 0$ for $i \notin I$, and the result follows. $\square$

$$\begin{bmatrix} \nabla f(x)^T \\ \nabla g_i(x)^T \\ \vdots \end{bmatrix}^T \begin{bmatrix} d \end{bmatrix} < 0$$

Fritz-John conditions

Proof.

Since $\overline{x}$ solves P locally, Theorem 3 guarantees that there is no d such that $\nabla f(\overline{x})^\top d < 0$ and $\nabla g_i(x)^\top d < 0$ for each $i \in I$. Let A be the matrix whose rows are $\nabla f(\overline{x})^\top$ and $\nabla g_i(\overline{x})^\top$ for $i \in I$.

Using Farkas' theorem, we have that if $Ad < 0$ is inconsistent, then there exists nonzero $p \geq 0$ such that $A^\top p = 0$. Being $I = \{i_1, \dots, i_{|I|}\}$, we let $p = (u_0, u_{i_1}, \dots, u_{i_{|I|}})$ and $u_i = 0$ for $i \notin I$, and the result follows. □

Remarks:

- The corollary of Farkas' theorem used in the proof is known as *Gordan's theorem*.
- $(u_0, \dots, u_m)$ are called Lagrangian multipliers.
- Notice that, for $i \notin I$, $u_i = 0$.

Fritz-John conditions

The Fritz-John (FJ) conditions:

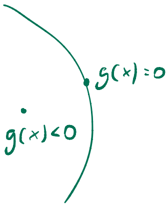
$$\overline{x} \in X, g_i(\overline{x}) \leq 0, \; i = 1, \dots, m \quad \text{(primal feasibility - PF)}$$

$$u_0 \nabla f(\overline{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{x}) = 0 \quad \text{(dual feasibility 1 - DF)}$$

$$u_i g_i(\overline{x}) = 0, \; i = 1, \dots, m \quad \text{(complementary slackness - CS)}$$

$$u_i \geq 0, \cancel{g_i(\overline{x})} \; i = 0, \dots, m \quad \text{(dual feasibility 2)}$$

$$u = (u_0, \dots, u_m) \neq 0 \quad \text{(dual feasibility 3)}$$



Fritz-John conditions

The Fritz-John (FJ) conditions:

$$\overline{x} \in X, g_i(\overline{x}) \leq 0, \ i = 1, \dots, m \quad \text{(primal feasibility - PF)}$$

$$u_0 \nabla f(\overline{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{x}) = 0 \qquad \text{(dual feasibility 1 - DF)}$$

$$u_i g_i(\overline{x}) = 0, \ i = 1, \dots, m \quad \text{(complementary slackness - CS)}$$

$$u_i \geq 0, g_i(\overline{x}) \ i = 0, \dots, m \quad \text{(dual feasibility 2)}$$

$$u = (u_0, \dots, u_m) \neq 0_i \qquad \text{(dual feasibility 3)}$$

Remark: If f if convex and g_i strictly convex for all $i = 1, \dots, m$, the FJ conditions become also sufficient for global optimality.

Fritz-John conditions

Example:

$$
\begin{aligned}
\text{min.} \quad & (x_1 - 3)^2 + (x_2 - 2)^2 \\
\text{subject to:} \quad & x_1^2 + 2x_2^2 \leq 5 \\
& x_1 + 2x_2 \leq 4 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

Fritz-John conditions

Example:

$$
\begin{aligned}
\text{min.} \quad & (x_1 - 3)^2 + (x_2 - 2)^2 \\
\text{subject to:} \quad & x_1^2 + \cancel{2}x_2^2 \leq 5 \\
& x_1 + 2x_2 \leq 4 \\
& x_1, x_2 \geq 0
\end{aligned}
$$

FJ conditions at $\overline{x} = (2, 1)$: $\nabla f(\overline{x}) = (-2, -2)^\top$, $\nabla g_1(\overline{x}) = (4, 2)^\top$, and $\nabla g_2(\overline{x}) = (1, 2)^\top$. We need a nonzero $(u_0, u_1, u_2) \geq 0$ such that

$$
u_0 \begin{pmatrix} -2 \\ -2 \end{pmatrix} + u_1 \begin{pmatrix} 4 \\ 2 \end{pmatrix} + u_2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.
$$

Fritz-John conditions

Example:

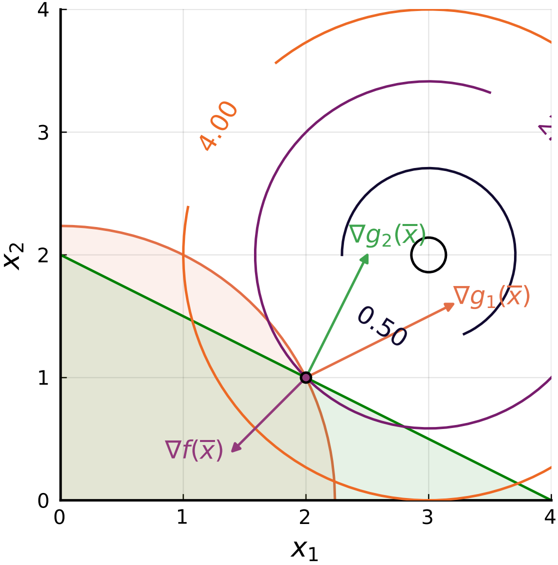
$$\begin{aligned} \text{min.} \quad & (x_1 - 3)^2 + (x_2 - 2)^2 \\ \text{subject to:} \quad & x_1^2 + 2x_2^2 \leq 5 \\ & x_1 + 2x_2 \leq 4 \\ & x_1, x_2 \geq 0 \end{aligned}$$

FJ conditions at $\overline{x} = (2,1)$: $\nabla f(\overline{x}) = (-2,-2)^\top, \nabla g_1(\overline{x}) = (4,2)^\top$, and $\nabla g_2(\overline{x}) = (1,2)^\top$. We need a nonzero $(u_0, u_1, u_2) \geq 0$ such that

$$u_0 \begin{pmatrix} -2 \\ -2 \end{pmatrix} + u_1 \begin{pmatrix} 4 \\ 2 \end{pmatrix} + u_2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Take $u_1 = u_0/3$ and $u_2 = 2u_0/3$. FJ conditions are then satisfied for any $u_0 > 0$. In fact, $(2,1)$ is the global minimum.

Fritz-John conditions



The issue with Fritz-John conditions

Fritz-John conditions are too weak in general settings; they hold for too many points to be useful.

This arises because $\overline{x}$ is a FJ solution if and only if $F_0 \cap G_0 = \emptyset$, which is trivially satisfied for any feasible $\overline{x}$ at which $G_0 = \emptyset$.

The issue with Fritz-John conditions

Fritz-John conditions are too weak in general settings; they hold for too many points to be useful.

This arises because $\overline{x}$ is a FJ solution if and only if $F_0 \cap G_0 = \emptyset$, which is trivially satisfied for any feasible $\overline{x}$ at which $G_0 = \emptyset$.

Examples where $G_0 = \emptyset$ (making $\overline{x}$ a FJ solution):

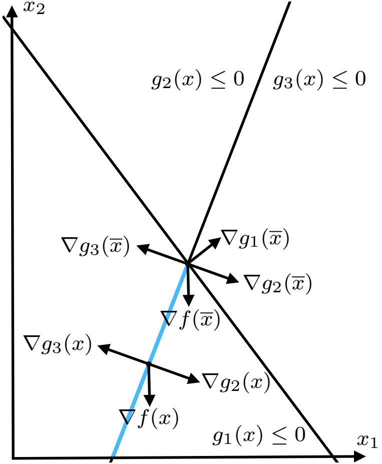
- Gradients that vanish at $\overline{x}$:
  - $\nabla f(\overline{x}) = 0$ or $\nabla g_i(\overline{x}) = 0$ for some $i \in I$;
  - problems with equality constraints: replace $g(x) = 0$ with $g_1(x) \leq 0$ and $-g_2(x) \leq 0$;

    $g_1 \leq 0$

    $g_2 \geq 0$
- Feasible region has no interior in the immediate vicinity of $\overline{x}$.

The issue with Fritz-John conditions

$$
\begin{aligned}
\text{min. } & f(x) = -x_2 \\
& g_1(x) \leq 0 \\
& h(x) = 0 \Rightarrow \\
& \begin{cases} g_2(x) = -h(x) \leq 0 \\ g_3(x) = h(x) \leq 0 \end{cases}
\end{aligned}
$$



All points in the blue segment satisfy FJ conditions, including the minimum $\overline{x}$.

Karush-Kuhn-Tucker conditions

KKT solutions are FJ solutions at which $G_0 \neq \emptyset$. Note that $G_0 \neq \emptyset$ requires $u_0 > 0$ for dual feasibility. This requirement is an example of constraint qualification.

Karush-Kuhn-Tucker conditions

KKT solutions are FJ solutions at which $G_0 \neq \emptyset$. Note that $G_0 \neq \emptyset$ requires $u_0 > 0$ for dual feasibility. This requirement is an example of constraint qualification.

Several conditions imply $G_0 \neq \emptyset$. For example: if $\nabla g_i(\overline{x})$ are linearly independent for all $i \in I$, then $u_0 > 0$ is required and thus implies constraint qualification. This is called the LICQ condition.

Karush-Kuhn-Tucker conditions

KKT solutions are FJ solutions at which $G_0 \neq \emptyset$. Note that $G_0 \neq \emptyset$ requires $u_0 > 0$ for dual feasibility. This requirement is an example of constraint qualification.

Several conditions imply $G_0 \neq \emptyset$. For example: if $\nabla g_i(\overline{x}) = 0$ are linearly independent for all $i \in I$, then $u_0 > 0$ is required and thus implies constraint qualification. This is called the LICQ condition.

We will later see more examples of conditions that imply constraint qualifications. For now, we will use the LICQ condition to express the KKT conditions. Once again, we focus on solving

$$(P) : \ \{\text{min. } f(x) : g_i(x) \leq 0, i = 1, \dots, m, x \in X\}.$$

Karush-Kuhn-Tucker conditions

Theorem 6 (Karush-Kuhn-Tucker necessary conditions)

Let $X \subseteq \mathbb{R}^n$ be a nonempty open set, and let $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ be differentiable for all $i = 1, \dots, m$. Additionally, for a feasible $\overline{x}$, let $I = \{i : g_i(\overline{x}) = 0\}$ and suppose that $\nabla g_i(\overline{x})$ are linearly independent for all $i \in I$. If $\overline{x}$ solves P locally, there exist scalars u_i for $i \in I$ such that

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{x}) = 0$$

$$u_i g_i(\overline{x}) = 0, \ i = 1, \dots, m$$

$$u_i \geq 0, \cancel{g_i(\overline{x})} \ i = 1, \dots, m$$

KKT = FJ + CQ

Karush-Kuhn-Tucker conditions

Proof.

By Theorem 5, there exists nonzero $(\hat{u}_i)$ for $i \in \{0\} \cup I$ such that

$$\hat{u}_0 \nabla f(\overline{x}) + \sum_{i=1}^{m} \hat{u}_i \nabla g_i(\overline{x}) = 0$$

$$\hat{u}_i \geq 0, \; i = 0, \ldots, m$$

Note that $\hat{u}_0 > 0$, as the linear independence of $\nabla g_i(\overline{x})$ for all $i \in I$ implies that $\sum_{i=1}^{m} \hat{u}_i \nabla g_i(\overline{x}) \neq 0$. Now, let $u_i = \hat{u}_i / u_0$ for each $i \in I$ and $u_i = 0$ for all $i \notin I$. □

Karush-Kuhn-Tucker conditions

Proof.

By Theorem 5, there exists nonzero $(\hat{u}_i)$ for $i \in \{0\} \cup I$ such that

$$\hat{u}_0 \nabla f(\overline{x}) + \sum_{i=1}^{m} \hat{u}_i \nabla g_i(\overline{x}) = 0$$

$$\hat{u}_i \geq 0, \; i = 0, \dots, m$$

Note that $\hat{u}_0 > 0$, as the linear independence of $\nabla g_i(\overline{x})$ for all $i \in I$ implies that $\sum_{i=1}^{m} \hat{u}_i \nabla g_i(\overline{x}) \neq 0$. Now, let $u_i = \hat{u}_i / u_0$ for each $i \in I$ and $u_i = 0$ for all $i \notin I$. □

Remark: KKT conditions enforce $u_0 > 0$, which can be turned into $u_0 = 1$ with proper scaling. This forces $\nabla f(x)$ to have a role in the optimality conditions.

Karush-Kuhn-Tucker conditions

The Karush-Kuhn-Tucker (KKT) conditions for a general P:

$$(P) \ : \ \{\text{min. } f(x) : g_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, l, x \in X\}$$

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\overline{x}) = 0 \quad \text{(dual feasibility 1)}$$

$\nabla h(x) = 0$

$$u_i g_i(\overline{x}) = 0, \quad i = 1, \dots, m \quad \text{(complementary slackness)}$$

$$\overline{x} \in X, \ g_i(\overline{x}) \leq 0, \ i = 1, \dots, m \quad \text{(primal feasibility)}$$

$$h_i(\overline{x}) = 0, \quad i = 1, \dots, l$$

$$u_i \geq 0, \quad i = 1, \dots, m. \quad \text{(dual feasibility 2)}$$

Karush-Kuhn-Tucker conditions

The Karush-Kuhn-Tucker (KKT) conditions for a general P:

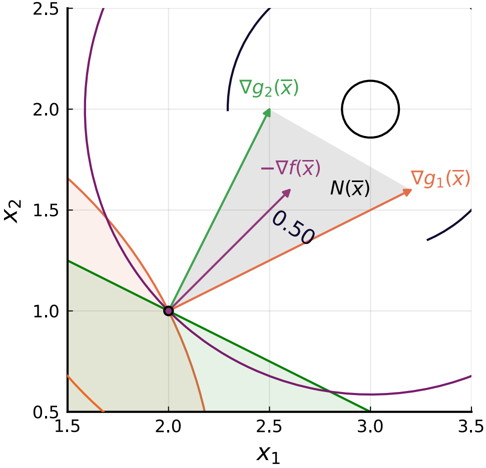
$$(P) \; : \; \{\text{min. } f(x) : g_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, l, x \in X\}$$

$$\begin{array}{lll}
\nabla f(\overline{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\overline{x}) = 0 & & \text{(dual feasibility 1)} \\
u_i g_i(\overline{x}) = 0, & i = 1, \dots, m & \text{(complementary slackness)} \\
\overline{x} \in X, \; g_i(\overline{x}) \leq 0, & i = 1, \dots, m & \text{(primal feasibility)} \\
h_i(x) = 0, & i = 1, \dots, l & \\
u_i \geq 0, & i = 1, \dots, m. & \text{(dual feasibility 2)}
\end{array}$$

Remarks:

1. Multipliers v_i, $i = 1, \dots, l$ are not restricted in sign.
2. For unconstrained problems, KKT conditions are equivalent to the optimality condition $\nabla f(\overline{x}) = 0$.

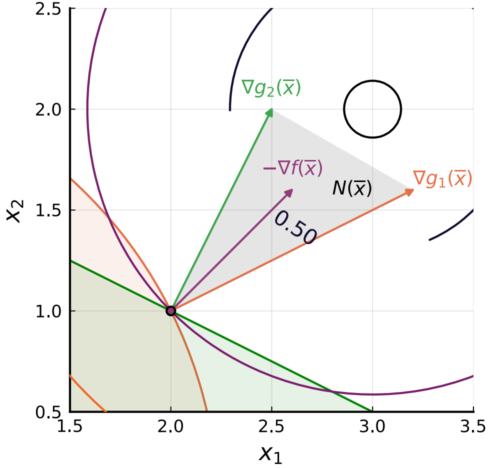
Geometric interpretation of KKT conditions



Graphical illustration of the KKT conditions at the optimal point

$$\begin{cases} u_1 \nabla g_1(x) + u_2 \nabla g_2(x) = -\nabla f(x) \\ u_1, u_2 \geq 0 \end{cases}$$

Geometric interpretation of KKT conditions



Graphical illustration of the KKT conditions at the optimal point

KKT conditions have a geometric interpretation. Let $N(\overline{x}) = \{\sum_{i \in I} u_i \nabla g_i(\overline{x}) : u_i \geq 0\}$ be the cone spanned by the gradient of the active constraints at $\overline{x}$.

$-\nabla f(\overline{x}) = \sum_{i=1}^{m} u_i \nabla g_i(\overline{x})$ is the same as requiring that $-\nabla f(\overline{x}) \in N(\overline{x})$.

Constraint qualification

We will next examine cases where constraint qualification is guaranteed to hold.

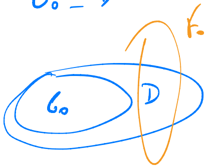
- Can be seen as a certification that the geometry of the feasible space and gradient information from the binding constraints are related at an optimal solution.

$$F \cap D = \emptyset$$

$$\parallel \qquad \parallel$$

$$F_0 \cap G_0 = \emptyset$$

$G_0 \subseteq D$



$G_0 = D$

Constraint qualification

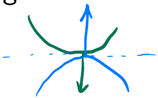
We will next examine cases where constraint qualification is guaranteed to hold.

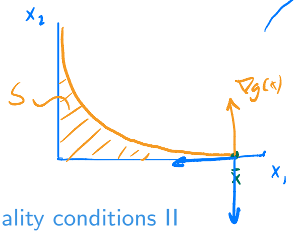
- Can be seen as a certification that the geometry of the feasible space and gradient information from the binding constraints are related at an optimal solution.
- Gradients provide a first-order approximation of feasible region.

Constraint qualification

We will next examine cases where constraint qualification is guaranteed to hold.

- Can be seen as a certification that the geometry of the feasible space and gradient information from the binding constraints are related at an optimal solution.
- Gradients provide a first-order approximation of feasible region.
- Often an issue for regions with cusps or those consisting of single points, for example.





Constraint qualification

We will next examine cases where constraint qualification is guaranteed to hold.

- Can be seen as a certification that the geometry of the feasible space and gradient information from the binding constraints are related at an optimal solution.
- Gradients provide a first-order approximation of feasible region.
- Often an issue for regions with cusps or those consisting of single points, for example.

There are several conditions that imply constraint qualification. We will focus on those most often used in practice.

Constraint qualification (CQ)

Constraint qualifications can be seen as certificates for proper relationships between the set of feasible directions

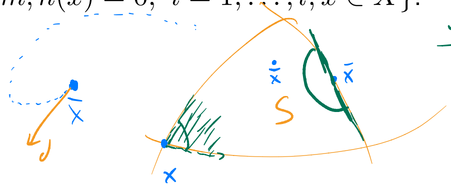
$$G_0' = \left\{ d \neq 0 : \nabla g_i(\overline{x})^\top d \leq 0, i \in I \right\}$$

and the cone of tangents (or tangent cone)

$$T = \{d : d = \lim_{k \to \infty} \lambda_k (x_k - \overline{x}), \lim_{k \to \infty} x_k = \overline{x}, x_k \in S, \lambda_k > 0, \forall k\},$$

with $S = \{g_i(x) \leq 0,\ i = 1, \dots, m; h(x) = 0,\ i = 1, \dots, l; x \in X\}$.

$\underline{D} = T$



$D = G_0$

$\shortparallel$

$\boxed{T = G_0}$

Constraint qualification (CQ)

Constraint qualifications can be seen as certificates for proper relationships between the set of feasible directions

$$G'_0 = \left\{d \neq 0 : \nabla g_i(\overline{x})^\top d \leq 0, i \in I\right\}$$

and the cone of tangents (or tangent cone)

$$T = \{d : d = \lim_{k\to\infty} \lambda_k(x_k - \overline{x}), \lim_{k\to\infty} x_k = \overline{x}, x_k \in S, \lambda_k > 0, \forall k\},$$

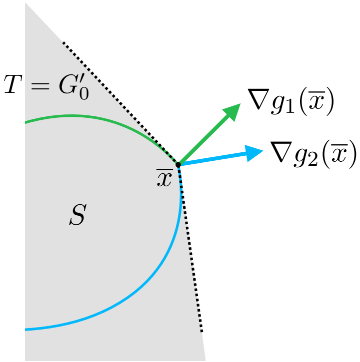
with $S = \{g_i(x) \leq 0,\ i = 1, \dots, m; h(x) = 0,\ i = 1, \dots, l; x \in X\}$.

Definition 7 (Abadie constraint qualification)

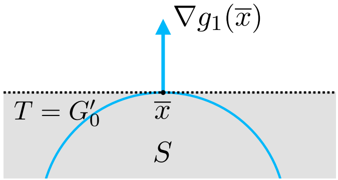
Abadie constraint qualification holds at $\overline{x}$ if $T = G'_0$.

Remark: with equality constraints, Abadie CQ may be rewritten as $T = G'_0 \cap H_0$, with $H_0 = \left\{d : \nabla h_i(\overline{x})^\top d = 0, i = 1, \dots, l\right\}$.

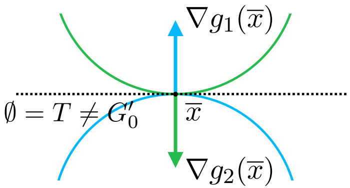
Constraint qualification (CQ)



(a)



(b)



(c)

CQ holds for 1a and 1b, but not for 1c.

Constraint qualification (CQ)

KKT conditions can be expressed more generally, assuming that Abadie CQ holds.

Constraint qualification (CQ)

KKT conditions can be expressed more generally, assuming that Abadie CQ holds.

Theorem 8 (Karush-Kuhn-Tucker necessary conditions II)

Consider the problem

$$(P): \ \{\text{min. } f(x) : g_i(x) \leq 0, i = 1, \dots, m, x \in X\}.$$

Let $X \subseteq \mathbb{R}^n$ be a nonempty open set, and let $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ be differentiable for all $i = 1, \dots, m$. Additionally, for a feasible $\overline{x}$, let $I = \{i : g_i(\overline{x}) = 0\}$ and suppose that Abadie CQ holds at $\overline{x}$. If $\overline{x}$ solves P locally, there exist scalars u_i for $i \in I$ such that

$T = G_0'$

D

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{x}) = 0$$

$$u_i g_i(\overline{x}) = 0, \ i = 1, \dots, m$$

$$u_i \geq 0, \cancel{g_i(\overline{x})} \ i = 1, \dots, m.$$

Constraint qualification

Verifying if Abadie CQ holds is not practical. Typically, we look for other conditions that imply Abadie CQ. Most useful are:

1. **Linear independence (LI)CQ:** holds at $\overline{x}$ if $\nabla g_i(\overline{x})$, for $i \in I$, as well as $\nabla h_i(\overline{x})$, $i = 1, \dots, l$ are linearly independent.

Constraint qualification

Verifying if Abadie CQ holds is not practical. Typically, we look for other conditions that imply Abadie CQ. Most useful are:

1. **Linear independence (LI)CQ:** holds at $\overline{x}$ if $\nabla g_i(\overline{x})$, for $i \in I$, as well as $\nabla h_i(\overline{x})$, $i = 1, \dots, l$ are linearly independent.
2. **Affine CQ:** holds for all $x \in S$ if g_i, for all $i = 1, \dots, m$, and h_i, for all $i = 1, \dots, l$, are affine.

Constraint qualification

Verifying if Abadie CQ holds is not practical. Typically, we look for other conditions that imply Abadie CQ. Most useful are:

1. **Linear independence (LI)CQ:** holds at $\overline{x}$ if $\nabla g_i(\overline{x})$, for $i \in I$, as well as $\nabla h_i(\overline{x})$, $i = 1, \dots, l$ are linearly independent.
2. **Affine CQ:** holds for all $x \in S$ if g_i, for all $i = 1, \dots, m$, and h_i, for all $i = 1, \dots, l$, are affine.
3. **Slater's CQ:** holds for all $x \in S$ if g_i is a convex function for all $i = 1, \dots, m$, h_i is an affine function for all $i = 1, \dots, l$, and there exists $x \in S$ such that $g_i(x) < 0$ for all $i = 1, \dots, m$.

Constraint qualification

Verifying if Abadie CQ holds is not practical. Typically, we look for other conditions that imply Abadie CQ. Most useful are:

1. **Linear independence (LI)CQ:** holds at $\overline{x}$ if $\nabla g_i(\overline{x})$, for $i \in I$, as well as $\nabla h_i(\overline{x})$, $i = 1, \dots, l$ are linearly independent.
2. **Affine CQ:** holds for all $x \in S$ if g_i, for all $i = 1, \dots, m$, and h_i, for all $i = 1, \dots, l$, are affine.
3. **Slater's CQ:** holds for all $x \in S$ if g_i is a convex function for all $i = 1, \dots, m$, h_i is an affine function for all $i = 1, \dots, l$, and there exists $x \in S$ such that $g_i(x) < 0$ for all $i = 1, \dots, m$.

Remark: Slater's CQ is by far the most frequently used.

KKT as necessary and sufficient conditions

Under convexity, KKT conditions are only sufficient for (global) optimality, which highlights the importance of Slater's CQ.

Consider, for example: $P = \left\{\text{min. } x_1 : x_1^2 + x_2 \leq 0, x_2 \geq 0\right\}$. The KKT system for P is

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 0; u_1, u_2 \geq 0,$$

which has no solution. Thus, KKT are not necessary for the global optimal $(0,0)$. This is due to the lack of CQ.

KKT as necessary and sufficient conditions

Under convexity, KKT conditions are only sufficient for (global) optimality, which highlights the importance of Slater's CQ.

Consider, for example: $P = \left\{\text{min. } x_1 : x_1^2 + x_2 \leq 0, x_2 \geq 0\right\}$. The KKT system for P is

$$\begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0 & 0\\1 & -1\end{pmatrix}\begin{pmatrix}u_1\\u_2\end{pmatrix} = 0; u_1, u_2 \geq 0,$$

which has no solution. Thus, KKT are not necessary for the global optimal $(0,0)$. This is due to the lack of CQ.

$F \cap D = \emptyset$

Corollary 9 (Necessary and sufficient KKT conditions)

Suppose that Slater's CQ holds. Then, if f is convex, the conditions of Theorem 8 are necessary and sufficient for $\overline{x}$ to be a global optimal solution.