OPTIMIZATION - 1

Prof. Dan A. Simovici

UMB

Definition

A set $C \subseteq \mathbb{R}^n$ is a cone if $\mathbf{0} \in C$ and $c\mathbf{x} \in C$ for every $c \in \mathbb{R}_{\geqslant 0}$ and $\mathbf{x} \in C$.
C is a convex cone if $\mathbf{0} \in C$ and for every $\mathbf{x}, \mathbf{y} \in C$ and any $a, b \in \mathbb{R}_{\geqslant 0}$ we have $a\mathbf{x} + b\mathbf{y} \in C$.

Definition

A conic combination of $\mathbf{x}_1, \ldots, \mathbf{x}_m$ is a vector $\mathbf{x}$ given by

$$\mathbf{x} = a_1\mathbf{x}_1 + \cdots + a_m\mathbf{x}_m,$$

where $a_i \geqslant 0$ for $1 \leqslant i \leqslant m$.
The set $\mathbf{C}(S)$ of conic combinations of all vectors of a set S, $S \subseteq \mathbb{R}^n$ is the conic hull or the conic closure of S.
The ray spanned by $\mathbf{x}$ is the set $\mathbf{C}(\{\mathbf{x}\})$.

Polar Cones

Definition

Let S be a non-empty subset of $\mathbb{R}^n$. The polar cone of S is the set given by

$$S^* = \{\mathbf{p} \in \mathbb{R}^n \mid \mathbf{p}'\mathbf{x} \leqslant 0 \text{ for all } \mathbf{x} \in S\}.$$

Theorem

For every non-empty set S where $S \subseteq \mathbb{R}^n$, S^* is a closed convex cone. Furthermore, $S \subseteq S^{**}$.
If S_1, S_2 are non-empty subsets of $\mathbb{R}^n$, then $S_1 \subseteq S_2$ implies $S_2^* \subseteq S_1^*$.

Proof

If $\mathbf{p} \in S^*$, $\mathbf{p}'\mathbf{x} \leqslant 0$ for all $\mathbf{x} \in S$. Therefore, $(a\mathbf{p})'\mathbf{x} \leqslant 0$ for all $\mathbf{x} \in S$, so $a\mathbf{p} \in S^*$, which implies that S^* is a cone.
If $\mathbf{p}, \mathbf{q} \in S$ and $a, b \geqslant 0$, we have $\mathbf{p}'\mathbf{x} \leqslant 0$ and $\mathbf{q}'\mathbf{x} \leqslant 0$ for all $\mathbf{x} \in S$, so $(a\mathbf{p} + b\mathbf{q})'\mathbf{x} \leqslant 0$ for all $\mathbf{x} \in S$, which implies $a\mathbf{p} + b\mathbf{q} \in S^*$. Thus, S^* is a convex cone.
To prove that S^* is closed we need to show that S^* contains the limit point of every convergent sequence in S^*, which is immediate: if $\lim \mathbf{p}_n = \mathbf{p}$, where $\mathbf{p}_n \in S^*$ for $n \in \mathbb{N}$ we have $\mathbf{p}_n'\mathbf{x} \leqslant 0$, so $\mathbf{p}'\mathbf{x} \leqslant 0$ for $\mathbf{x} \in S^*$.

Proof (cont'd)

Let $\mathbf{u} \in S$. Then, $\mathbf{p}'\mathbf{u} \leqslant 0$ for every $\mathbf{p} \in S^*$. Therefore, $\mathbf{u}'\mathbf{p} \leqslant 0$ for every $\mathbf{p} \in S^*$, which means that $\mathbf{u} \in S^{**}$. Thus, $S \subseteq S^{**}$.
Finally, if $\mathbf{p} \in S_2^*$, $\mathbf{p}'\mathbf{x} \leqslant 0$ for every $\mathbf{x} \in S_2$, so $\mathbf{p}'\mathbf{x} \leqslant 0$ for every $\mathbf{x} \in S_1$, so $\mathbf{p} \in S_1^*$.

Theorem

If C is a non-empty closed convex cone, then $C^{**} = C$.

Proof

We have $C \subseteq C^{**}$. Let $\mathbf{x} \in C^{**}$.
Suppose $\mathbf{x} \notin C$. Then, by the separation theorem applied to the set C and the point $\mathbf{x}$, there exists $\mathbf{p}$ and a such that

$$\mathbf{p}'\mathbf{y} \leqslant a \text{ for } \mathbf{y} \in C \text{ and } \mathbf{p}'\mathbf{x} > a.$$

Since $\mathbf{0} \in C$, $a \geqslant 0$ an $\mathbf{p}'\mathbf{x} > 0$.
We claim that $\mathbf{p} \in C^*$. Suppose that this is not the case. Then, $\mathbf{p}'\overline{\mathbf{y}} > 0$ for some $\overline{\mathbf{y}} \in C$, so $\mathbf{p}'(a\overline{\mathbf{y}})$ can be made arbitrarily large by choosing a large. This contradicts the fact that $\mathbf{p}'\mathbf{y} \leqslant a$ for all $\mathbf{y} \in C$. Thus, $\mathbf{p} \in C^*$.
Since $\mathbf{x} \in C^{**}$, we have $\mathbf{p}'\mathbf{x} \leqslant 0$, which contradicts $\mathbf{p}'\mathbf{x} > 0$. Thus, $\mathbf{x} \in C$.

Example

Let $A \in \mathbb{R}^{m \times n}$. The set

$$C = \{\mathbf{u} \in \mathbb{R}^n \mid \mathbf{u} = A'\mathbf{y} \text{ where } \mathbf{y} \in \mathbb{R}^m, \mathbf{y} \geqslant \mathbf{0}\}$$

is a closed convex cone.
Its polar cone is

$$\begin{aligned} C^* &= \{\mathbf{p} \in \mathbb{R}^n \mid \mathbf{p}'\mathbf{u} \leqslant 0 \text{ for every } \mathbf{u} \in C\} \\ &= \{\mathbf{p} \in \mathbb{R}^n \mid \mathbf{p}'A'\mathbf{y} \leqslant 0 \text{ for every } \mathbf{y} \geqslant \mathbf{0}\} \\ &= \{\mathbf{p} \in \mathbb{R}^n \mid (A\mathbf{p})'\mathbf{y} \leqslant 0 \text{ for every } \mathbf{y} \geqslant \mathbf{0}\} \\ &= \{\mathbf{p} \in \mathbb{R}^n \mid A\mathbf{p} \leqslant 0\}. \end{aligned}$$

Continued example

Example

If $\mathbf{c} \in C^{**}$, $\mathbf{x} \in C^*$ implies $\mathbf{c}'\mathbf{x} \leqslant 0$ or, $A\mathbf{x} \leqslant \mathbf{0}$ implies $\mathbf{c}'\mathbf{x} \leqslant 0$. By the definition of C, $\mathbf{c} \in C$ implies $\mathbf{c} = A'\mathbf{y}$ and $\mathbf{y} \geqslant 0$. Thus, exactly one of the following cases hold:

Case 1: if $\mathbf{c} \notin C^{**} = C$, $A\mathbf{x} \leqslant \mathbf{0}, \mathbf{c}'\mathbf{x} > 0$;

Case 2: if $\mathbf{c} \in C$, $A'\mathbf{y} = \mathbf{c}, \mathbf{y} \geqslant \mathbf{0}$.

This is Farkas' Theorem.

Feasible Directions

Definition

Let $S \subseteq \mathbb{R}^n$ and let $\bar{\mathbf{x}} \in \mathbf{K}(S)$. The vector $\mathbf{d} \neq \mathbf{0}$ is a feasible direction in $\bar{\mathbf{x}}$ if there exists $\delta > 0$ such that $\bar{\mathbf{x}} + \lambda \mathbf{d} \in S$ for $0 < \lambda < \delta$.

Theorem

Let $S \subseteq \mathbb{R}^n$, $\overline{\mathbf{x}} \in \mathbf{K}(S)$, $f : S \longrightarrow \mathbb{R}$ *be differentiable at* $\overline{\mathbf{x}}$ *such that* $\overline{\mathbf{x}}$ *is a local optimum. Define*

$$\begin{aligned} D &= \{\mathbf{d} \mid \mathbf{d} \textit{ is a feasible direction in } \overline{\mathbf{x}}\} \\ &\quad \textit{(the cone of feasible directions in } \overline{\mathbf{x}}\textit{)} \\ F_0 &= \{\mathbf{d} \mid (\nabla f)(\overline{\mathbf{x}})'\mathbf{d} < 0\}. \end{aligned}$$

We have $D \cap F_0 = \emptyset$.

Proof

Suppose that there exists $\mathbf{d} \in F_0 \cap D$.
Since f is differentiable at $\overline{\mathbf{x}}$

$$f(\overline{\mathbf{x}} + \lambda\mathbf{d}) = f(\overline{\mathbf{x}}) + \lambda(\nabla f)(\overline{\mathbf{x}})'\mathbf{d} + \lambda^2 \parallel \mathbf{d} \parallel^2 \alpha_{\mathbf{x}}(\lambda\mathbf{d}).$$

If $\mathbf{d} \in F_0$, $(\nabla f)(\overline{\mathbf{x}})'\mathbf{d} < 0$, so $\mathbf{d}$ is an improving direction: a small movement in the direction of $\mathbf{d}$ away from $\overline{\mathbf{x}}$ reduces the value of f. Thus, there exists δ_1 such that $f(\overline{\mathbf{x}} + \lambda\mathbf{d}) < f(\overline{\mathbf{x}})$ for $\lambda \in (0, \delta_1)$
There exists $\delta_2 > 0$ such that $\overline{\mathbf{x}} + \lambda\mathbf{d} \in S$ for $\lambda \in (0, \delta_2)$, which contradicts the optimality of $\overline{\mathbf{x}}$. So, by choosing $\lambda < \min(\delta_1, \delta_2)$ we get $\overline{\mathbf{x}} + \lambda\mathbf{d} \in S$ with $f(\mathbf{x}) < f(\overline{\mathbf{x}})$, which contradicts the optimality of $\overline{\mathbf{x}}$.

Feasible Region

Let $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, where $1 \leqslant i \leqslant m$. The *feasible region* is the set

$$X = \{\mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \leqslant 0 \text{ for } 1 \leqslant i \leqslant m\}.$$

Optimization Problem $\mathcal{O}_{min,\leqslant}(X, f, g_i)$:
minimize $f(\mathbf{x})$ for $\mathbf{x} \in X$;
subject to $g_i(\mathbf{x}) \leqslant 0$ for $1 \leqslant i \leqslant m$ and $\mathbf{x} \in X$.

A Necessary Condition for Optimality

Consider the optimization problem $\mathcal{O}_{min,\leqslant}(X, f, g_i)$, where X is an open set. Define $I_{\overline{\mathbf{x}}} = \{i \mid g_i(\overline{\mathbf{x}}) = 0\}$. The constraints $g_i(\overline{\mathbf{x}}) \leqslant 0$ where $i \in I$ are the binding constraints of the problem.

Theorem

Let $\overline{\mathbf{x}}$ be a feasible point. Suppose that f and g_i are differentiable at the feasible point $\overline{\mathbf{x}}$ for $i \in I_{\overline{\mathbf{x}}}$ and g_i are continuous in $\overline{\mathbf{x}}$ for $i \notin I_{\overline{\mathbf{x}}}$.
If $\overline{\mathbf{x}}$ is a local optimal solution, then $F_0 \cap G_0 = \emptyset$, where

$$\begin{aligned} F_0 &= \{\mathbf{d} \mid (\nabla f)(\overline{\mathbf{x}})'\mathbf{d} < 0\}, \\ G_0 &= \{\mathbf{d} \mid (\nabla g)(\overline{\mathbf{x}})'\mathbf{d} < 0 \text{ for each } i \in I_{\overline{\mathbf{x}}}\}. \end{aligned}$$

Proof

- Let $\mathbf{d} \in G_0$. Since X is open, there exists $\delta_1 > 0$ such that $\overline{\mathbf{x}} + \lambda\mathbf{d} \in X$ for $0 < \lambda < \delta_1$.
- Since $g_i(\overline{\mathbf{x}}) < 0$ and g_i is continuous at $\overline{\mathbf{x}}$ for $i \notin I_{\overline{\mathbf{x}}}$, there exists $\delta_2 > 0$ such that $g_i(\overline{\mathbf{x}} + \lambda\mathbf{d}) < 0$ for $0 < \lambda < \delta_2$ and $i \notin I$.
- Since $\mathbf{d} \in G_0$, $(\nabla g_i)(\overline{\mathbf{x}})'\mathbf{d} < 0$ for $i \in I_{\overline{\mathbf{x}}}$, so there exists $\delta_3 > 0$ such that $g_i(\overline{\mathbf{x}} + \lambda\mathbf{d}) < g_i(\overline{\mathbf{x}}) = 0$ for $0 < \lambda < \delta_3$.

Points of the form $\overline{\mathbf{x}} + \lambda\mathbf{d}$ are feasible for $0 < \lambda < \min\{\delta_1, \delta_2, \delta_3\}$. Thus, $\mathbf{d}$ is a feasible direction and $\mathbf{d} \in G_0$ implies $d \in D$, so $G_0 \subseteq D$.
Since $\overline{\mathbf{x}}$ is a local minimum, $F_0 \cap D = \emptyset$, so $F_0 \cap G_0 = \emptyset$.

Example

minimize $(x_1 - 3)^2 + (x_2 - 2)^2$
subject to

$$\begin{aligned} x_1^2 + x_2^2 &\leqslant 5 \\ x_1 + x_2 &\leqslant 3 \\ x_1 &> 0 \\ x_2 &> 0. \end{aligned}$$

Define

$$\begin{aligned} g_1(\mathbf{x}) &= x_1^2 + x_2^2 - 5 \\ g_2(\mathbf{x}) &= x_1 + x_2 - 3 \\ g_3(\mathbf{x}) &= -x_1 \\ g_4(\mathbf{x}) &= -x_2 \end{aligned}$$

$$
\begin{aligned}
g_1(\mathbf{x}) &= x_1^2 + x_2^2 - 5 & g_2(\mathbf{x}) &= x_1 + x_2 - 3 \\
g_3(\mathbf{x}) &= -x_1 & g_4(\mathbf{x}) &= -x_2
\end{aligned}
$$

For $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ the first two constrains g_1 and g_2 are binding.
The gradients are:

$$
\begin{aligned}
(\nabla f)\begin{pmatrix} 2 \\ 1 \end{pmatrix} &= \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \quad (\nabla g_1)\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}, \\
(\nabla g_2)\begin{pmatrix} 2 \\ 1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix}.
\end{aligned}
$$

Since

$$
\begin{aligned}
F_0 &= \left\{ \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \mid -2d_1 - 2d_2 < 0 \right\} \\
G_0 &= \left\{ \begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \mid 4d_1 + 2d_2 < 0 \text{ and } d_1 + d_2 < 0 \right\}.
\end{aligned}
$$

it is clear that $F_0 \cap G_0 = \emptyset$.

Gordan's Theorem

Theorem

Let $A \in \mathbb{R}^{m \times n}$*. Exactly one of the following systems has a solution:*

- $A\mathbf{x} < \mathbf{0}$ *for* $\mathbf{x} \in \mathbb{R}^n$*;*
- $A'\mathbf{p} = \mathbf{0}$ *and* $\mathbf{p} \geqslant \mathbf{0}$ *for* $\mathbf{p} \in \mathbb{R}^m - \{\mathbf{0}\}$*.*

Proof

Suppose that $\hat{\mathbf{x}}$ is a solution of the first system and that there exists $\hat{\mathbf{p}} \in \mathbb{R}^m - \{\mathbf{0}\}$ such that $A'\hat{\mathbf{p}} = \mathbf{0}$ and $\hat{\mathbf{p}} \geqslant \mathbf{0}$.
Since $A\hat{\mathbf{x}} < \mathbf{0}$ and $\hat{\mathbf{p}} \geqslant \mathbf{0}$, $\hat{\mathbf{p}} \neq \mathbf{0}$, it follows that $\hat{\mathbf{p}}'A\hat{\mathbf{x}} < 0$, or $\mathbf{x}'A'\hat{\mathbf{p}} < 0$.
This contradicts the fact that $A'\hat{\mathbf{p}} = \mathbf{0}$. Therefore, the second system cannot have a solution.

Proof (cont'd)

Suppose now that the first system has no solution and consider the sets

$$S_1 = \{\mathbf{z} \in \mathbb{R}^m \mid \mathbf{z} = A\mathbf{x}, \mathbf{x} \in \mathbb{R}^n\}, S_2 = \{\mathbf{z} \in \mathbb{R}^m \mid \mathbf{z} < \mathbf{0}\}.$$

S_1, S_2 are convex sets and $S_1 \cap S_2 = \emptyset$.
There exists a hyperplane that separates S_1 from S_2, that is, there exists $\mathbf{p} \neq \mathbf{0}$ such that $\mathbf{p}' A\mathbf{x} \geqslant \mathbf{p}'\mathbf{z}$ for $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{z} \in \mathbf{K}(S_2)$.
Since each component of $\mathbf{z}$ could be an arbitrary small negative number, it follows that $\mathbf{p} \geqslant \mathbf{0}$.
Letting $\mathbf{z} = \mathbf{0}$ implies $\mathbf{p}' A\mathbf{x} \geqslant \mathbf{0}$ for $\mathbf{x} \in \mathbb{R}^n$. By choosing $\mathbf{x} = -A'\mathbf{p}$ it follows that $- \parallel A'\mathbf{p} \parallel^2 \geq 0$, so $A\mathbf{p} = \mathbf{0}$, which proves that the second system has a solution.

The Fritz John Conditions

Theorem

Let $\overline{\mathbf{x}}$ be a feasible solution of $\mathcal{O}_{min,\leqslant}(X, f, g_i)$ and let $I_{\overline{\mathbf{x}}}$ be the set of binding constrains in $\overline{\mathbf{x}}$.
Assume that f and g_i are differentiable at $\overline{\mathbf{x}}$ and g_i are continuous at $\overline{\mathbf{x}}$ for $i \notin I_{\overline{\mathbf{x}}}$.
If $\overline{\mathbf{x}}$ is a local optimum for $\mathcal{O}_{min,\leqslant}(X, f, g_i)$, then there exist the scalars u_0 and u_i for $i \in I_{\overline{\mathbf{x}}}$ such that

$$\begin{aligned} u_0(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} u_i(\nabla g_i)(\overline{\mathbf{x}}) &= \mathbf{0}, \\ u_0, u_i &\geq 0 \text{ for } i \in I_{\overline{\mathbf{x}}} \\ \begin{pmatrix} u_0 \\ \mathbf{u}_{I_{\overline{\mathbf{x}}}} \end{pmatrix} &\neq \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \end{aligned}$$

where $\mathbf{u}_{I_{\overline{\mathbf{x}}}}$ is a vector whose components are u_i for $i \in I_{\overline{\mathbf{x}}}$.

An Equivalent Formulation

u_0 and u_i are called Lagrangean multipliers.
An equivalent form of Fritz John's condition is

$$\begin{aligned} u_0(\nabla f)(\bar{\mathbf{x}}) + \sum_{i \in I_{\bar{\mathbf{x}}}} u_i (\nabla g_i)(\bar{\mathbf{x}}) &= \mathbf{0}, \\ u_i g_i(\bar{\mathbf{x}}) &= 0 \text{ for } 1 \leq i \leq m, \\ u_0, u_i &\geq 0 \text{ for } 1 \leq i \leq m, \\ \begin{pmatrix} u_0 \\ \mathbf{u}_{I_{\bar{\mathbf{x}}}} \end{pmatrix} &\neq \begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}. \end{aligned}$$

The condition

$$u_i g_i(\bar{\mathbf{x}}) = 0 \text{ for } 1 \leq i \leq m,$$

is called the complementary slackness condition. It requires $u_i = 0$ if the corresponding inequality is not binding, that is, $g_i(\bar{\mathbf{x}}) < 0$ and allows $u_i = 0$ only for binding constraints.

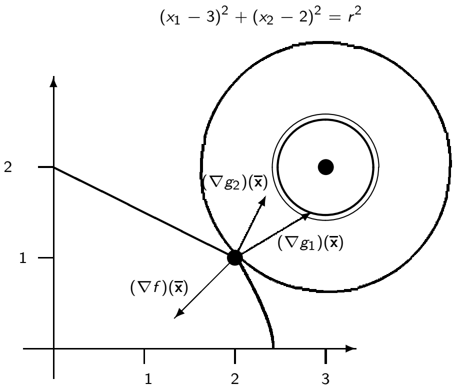
Proof

Since $\overline{\mathbf{x}}$ is a solution for $\mathcal{O}_{min,\leqslant}(X, f, g_i)$, there is no $\mathbf{d}$ such that $(\nabla f)(\overline{\mathbf{x}})'\mathbf{d} < 0$ and $(\nabla g_i)(\overline{\mathbf{x}})'\mathbf{d} < 0$ for $i \in I$. Let

$$A = ((\nabla f)(\overline{\mathbf{x}})\ (\nabla g_1)(\overline{\mathbf{x}})\ \cdots (\nabla g_m)(\overline{\mathbf{x}})).$$

The optimality conditions are equivalent to saying that the system $A'\mathbf{d} < \mathbf{0}$ has no solution in $\mathbf{d}$. Then, by Gordan's Theorem, the system $A\mathbf{p} = 0$ has a solution and we may assume $\mathbf{p} \geq \mathbf{0}$. If u_0, u_i are the components of $\mathbf{p}$, and the result follows.

Example 1



$$(\nabla f)\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}-2\\-2\end{pmatrix}, (\nabla g_1)\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}4\\2\end{pmatrix}, (\nabla g_2)\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}.$$

For $u_0 = 3$, $u_1 = 1$, and $u_2 = 2$ we have $u_0 \begin{pmatrix}-2\\-2\end{pmatrix} + u_1 \begin{pmatrix}4\\2\end{pmatrix} + u_2 \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$.

Fritz John's conditions are satisfied in $\begin{pmatrix}2\\1\end{pmatrix}$.

Example 2

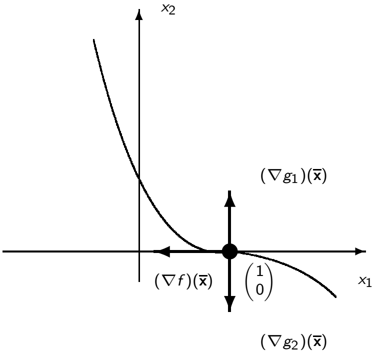
minimize $-x_1$
subject to

$$\begin{aligned} x_2 - (1 - x_1)^3 &\leqslant 0 \\ -x_2 &\leqslant 0 \end{aligned}$$

For $\bar{\mathbf{x}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ we have

$$(\nabla f)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, (\nabla g_1)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, (\nabla g_2)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

as shown in the next slide.



The condition

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

is satisfied only if $u_0 = 0$ and $u_1 = u_2$.

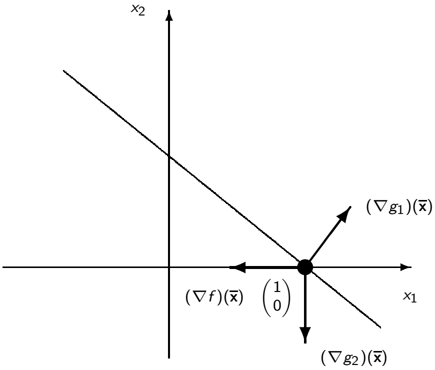
Example 3

minimize $-x_1$
subject to

$$\begin{aligned} x_1 + x_2 - 1 &\leqslant 0 \\ -x_2 &\leqslant 0 \end{aligned}$$

For $\bar{\mathbf{x}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ we have

$$(\nabla f)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, (\nabla g_1)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, (\nabla g_2)\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$



The Fritz John conditions hold for $u_0 = u_1 = u_2 = a$ for any $a > 0$.

Remarks

- The Fritz John's conditions can be extended to equality constraints by replacing $g(\mathbf{x}) = 0$ be the pair $g(\mathbf{x}) \leqslant 0$, $-g(\mathbf{x}) \leq 0$.
- If $\overline{\mathbf{x}}$ is such that $(\nabla f)(\overline{\mathbf{x}}) = \mathbf{0}$ or $(\nabla g_i)(\overline{\mathbf{x}}) = \mathbf{0}$, the corresponding Lagrangian multiplier can be set to any positive number, set all other multipliers to 0 and satisfy Fritz John's conditions.
- In Examples 1 and 3, $u_0 > 0$ at $\overline{\mathbf{x}}$, while in Example 2, $u_0 = 0$. Also, in Example 2, the gradients of the binding constraints are linearly dependent.

An Issue of the Fritz John's Theorem

- If $u_0 = 0$, the FJCs make no use of the gradient of the objective function, and require only the existence of a nonnegative and non-trivial zero linear combination of the gradients of the binding constraints.
- Thus, when $u_0 = 0$, the FJC are of no practical value in seeking an optimal point.

Kuhn-Tucker Necessary Condition

Theorem

Consider $\mathcal{O}_{min,\leqslant}(X, f, g_i)$. Let $\overline{\mathbf{x}}$ be a feasible solution and let $I_{\overline{\mathbf{x}}} = \{i \mid g_i(\overline{\mathbf{x}}) = 0\}$. Suppose that f and g_i are differentiable at $\overline{\mathbf{x}}$ and that g_i are continuous at $\overline{\mathbf{x}}$ for $i \notin I$. Furthermore, suppose that $(\nabla g_i)(\overline{\mathbf{x}})$ for $i \in I_{\overline{\mathbf{x}}}$ are linearly independent.

If $\overline{\mathbf{x}}$ is a local optimum, then there exist scalars u_i for $i \in I_{\overline{\mathbf{x}}}$ such that

$$(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} u_i (\nabla g_i)(\overline{\mathbf{x}}) = \mathbf{0} \text{ and } u_i \geqslant 0 \text{ for } i \in I_{\overline{\mathbf{x}}}.$$

In addition, if g_i are also differentiable at $\overline{\mathbf{x}}$ when $i \notin I_{\overline{\mathbf{x}}}$, then the Kuhn-Tucker can be written as

$$\begin{aligned} (\nabla f)(\overline{\mathbf{x}}) + \sum_{i=1}^{m} (\nabla g_i)(\overline{\mathbf{x}}) &= \mathbf{0} \\ u_i g_i(\overline{\mathbf{x}}) &= 0 \text{ for } 1 \leqslant i \leqslant m \\ u_i &\geqslant 0 \text{ for } 1 \leqslant i \leqslant m. \end{aligned}$$

Proof

By Fritz John's conditions there exist u_0 and $\hat{u}_i$ for $i \in I_{\overline{\mathbf{x}}}$ such that

$$u_0(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} u_i(\nabla g_i)(\overline{\mathbf{x}}) = 0.$$

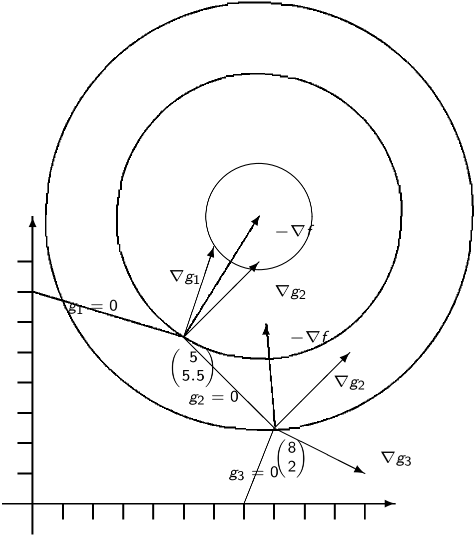
We have $u_0 \neq 0$; otherwise, the linear independence of $\{(\nabla g_i)(\overline{\mathbf{x}}) \mid i \in I_{\mathbf{x}}\}$ would be violated. Thus, the first part follows by defining $u_i = \frac{\hat{u}_i}{u_0}$ for $i \in I_{\overline{\mathbf{x}}}$.

The second part of the theorem follows by letting $u_i = 0$ for $i \notin I_{\overline{\mathbf{x}}}$.

Previous examples and the KT conditions

- In Example 1 at $\bar{\mathbf{x}} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$, the KT conditions are satisfied with $u_1 = 4/3,\ u_2 = 2/3,\ u_3 = u_4 = 0$.
- Example 2 does not satisfy the KT conditions at $\bar{\mathbf{x}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ because $(\nabla g_1)(\bar{\mathbf{x}})$ and $(\nabla g_2)(\bar{\mathbf{x}})$ are linearly dependent.
- Example 3 satisfies KT with $u_1 = u_2 = 1$.

Geometric Interpretation of the KT conditions



$-\nabla f$ in the cone of the gradients of constraints at $\begin{pmatrix} 5 \\ 5.5 \end{pmatrix}$ and outside the cone at $\begin{pmatrix} 8 \\ 2 \end{pmatrix}$

KT Sufficient Conditions

Theorem

Consider $\mathcal{O}_{min,\leqslant}(X, f, g_i)$*. Let* $\overline{\mathbf{x}}$ *be a feasible solution and let* $I_{\overline{\mathbf{x}}} = \{i \mid g_i(\overline{\mathbf{x}}) = 0\}$*. Suppose that* f *and* g_i *are differentiable at* $\overline{\mathbf{x}}$*. Furthermore, suppose that* f *and* g_i *are convex and there exist scalars* u_i *for* $i \in I_{\overline{\mathbf{x}}}$ *such that*

$$(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} u_i (\nabla g_i)(\overline{\mathbf{x}}) = \mathbf{0} \text{ and } u_i \geqslant 0 \text{ for } i \in I_{\overline{\mathbf{x}}}.$$

Then $\overline{\mathbf{x}}$ *is a global optimum.*

Proof

Let $\mathbf{x}$ be a feasible solution of $\mathcal{O}_{min,\leqslant}(X, f, g_i)$. For $i \in I_{\overline{\mathbf{x}}}$ we have $g_i(\mathbf{x}) \leqslant g_i(\overline{\mathbf{x}})$ because $g_i(\mathbf{x}) \leqslant 0$ and $g(\mathbf{x}) = 0$. By the convexity of g_i,

$$g_i(\lambda\mathbf{x} + (1-\lambda)\overline{\mathbf{x}}) \leqslant \lambda g_i(\mathbf{x}) + (1-\lambda)g_i(\overline{\mathbf{x}}) \leqslant g_i(\overline{\mathbf{x}}).$$

Thus, g_i does not increase when we move from $\overline{\mathbf{x}}$ along the line $\mathbf{x} - \overline{\mathbf{x}}$. Therefore, $(\nabla g_i)(\overline{\mathbf{x}})'(\mathbf{x} - \overline{\mathbf{x}}) \leqslant 0$. This implies

$$\sum_{i \in I_{\overline{\mathbf{x}}}} u_i (\nabla g_i)(\overline{\mathbf{x}})'(\mathbf{x} - \overline{\mathbf{x}}) \leqslant 0.$$

Since

$$(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} (\nabla g_i)(\overline{\mathbf{x}}) = \mathbf{0} \text{ and } u_i \geqslant 0 \text{ for } i \in I_{\overline{\mathbf{x}}},$$

it follows that $(\nabla f)(\overline{\mathbf{x}})'(\mathbf{x} - \overline{\mathbf{x}}) \geqslant 0$. By the convexity of f, $f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}})$.

Synopsis (Optimization with Inequalities)

Consider $\mathcal{O}_{min,\leqslant}(X, f, g_i)$. Let $\overline{\mathbf{x}}$ be a feasible solution and let $I_{\overline{\mathbf{x}}} = \{i \mid g_i(\overline{\mathbf{x}}) = 0\}$. Suppose that f and g_i are differentiable at $\overline{\mathbf{x}}$.
Necessary Conditions: If $\overline{\mathbf{x}}$ is a local optimum, $(\nabla g_i)(\overline{\mathbf{x}})$ for $i \in I_{\overline{\mathbf{x}}}$ are linearly independent, then there exist scalars u_i for $i \in I_{\overline{\mathbf{x}}}$ such that

$$(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} u_i (\nabla g_i)(\overline{\mathbf{x}}) = \mathbf{0} \text{ and } u_i \geqslant 0 \text{ for } i \in I_{\overline{\mathbf{x}}}.$$

Sufficient Conditions: If f and g_i are convex and there exist scalars u_i for $i \in I_{\overline{\mathbf{x}}}$ such that

$$(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} u_i (\nabla g_i)(\overline{\mathbf{x}}) = \mathbf{0} \text{ and } u_i \geqslant 0 \text{ for } i \in I_{\overline{\mathbf{x}}},$$

then $\overline{\mathbf{x}}$ is a global optimum.

Optimization Problem $\mathcal{O}_{min,\leqslant}(X, f, g_i, h_j)$:

- minimize $f(\mathbf{x})$ for $\mathbf{x} \in X$;
- subject to $g_i(\mathbf{x}) \leqslant 0$ for $1 \leqslant i \leqslant m$, $h_j(\mathbf{x}) = 0$ for $1 \leqslant j \leqslant \ell$, and $\mathbf{x} \in X$.

Let $\overline{\mathbf{x}}$ be a local optimum and $I_{\overline{\mathbf{x}}} = \{i \mid g_i(\overline{\mathbf{x}}) = 0\}$. Define

$$\begin{aligned}
F_0 &= \{\mathbf{d} \mid (\nabla f)(\overline{\mathbf{x}})'\mathbf{d} < 0\}\\
G_0 &= \{\mathbf{d} \mid (\nabla g_i)(\overline{\mathbf{x}})'\mathbf{d} < 0 \text{ for each } i \in I_{\overline{\mathbf{x}}}\}\\
H_0 &= \{\mathbf{d} \mid (\nabla h_j)(\overline{\mathbf{x}})'\mathbf{d} = 0 \text{ for each } 1 \leqslant j \leqslant \ell\}.
\end{aligned}$$

A Precursor to Fritz John's Conditions

Theorem

Let $\overline{\mathbf{x}}$ be a feasible solution of the problem $\mathcal{O}_{min,\leqslant}(X, f, g_i, h_j)$, where X is an open subset of $\mathbb{R}^n$. Let $\overline{\mathbf{x}}$ be a local optimum and $I_{\overline{\mathbf{x}}} = \{i \mid g_i(\overline{\mathbf{x}}) = 0\}$. Suppose that g_i for $i \in I_{\overline{\mathbf{x}}}$ is continuous in $\overline{\mathbf{x}}$, that f and g_i are differentiable at $\overline{\mathbf{x}}$ and that h_j are continuously differentiable at $\overline{\mathbf{x}}$. If the vectors $(\nabla h_j)(\overline{\mathbf{x}})$ for $1 \leqslant j \leqslant \ell$ are linearly independent, then

$$F_0 \cap G_0 \cap H_0 = \emptyset.$$

Fritz John's Conditions

Theorem

Let $\bar{\mathbf{x}}$ *be a feasible solution of the problem* $\mathcal{O}_{min,\leqslant}(X, f, g_i, h_j)$*, where* X *is an open subset of* $\mathbb{R}^n$*.*

Suppose that g_i *for* $i \in I_{\bar{\mathbf{x}}}$ *is continuous in* $\bar{\mathbf{x}}$*, that* f *and* g_i *are differentiable at* $\bar{\mathbf{x}}$ *and that* h_j *are continuously differentiable at* $\bar{\mathbf{x}}$*. If* $\bar{\mathbf{x}}$ *is a local optimum, then there exist* u_0*,* u_i *for* $i \in I_{\bar{\mathbf{x}}}$ *and* v_j *for* $1 \leqslant j \leqslant \ell$ *such that*

$$
\begin{aligned}
u_0(\nabla f)(\bar{\mathbf{x}}) + \sum_{i \in I_{\bar{\mathbf{x}}}} u_i(\nabla g_i)(\bar{\mathbf{x}}) + \sum_{j=1}^{\ell} v_j(\nabla h_j)(\bar{\mathbf{x}}) &= \mathbf{0}, \\
u_0, u_i &\geq 0 \text{ for } i \in I_{\bar{\mathbf{x}}} \\
(u_0, \mathbf{u}, \mathbf{v}) \neq (0, \mathbf{0}, \mathbf{0}),
\end{aligned}
$$

where $\mathbf{u}$ *is a vector whose components are* u_i *for* $i \in I_{\bar{\mathbf{x}}}$ *and* $\mathbf{v} = (v_1, \ldots, v_\ell)$*.*

Note that the signs of multipliers v_j are unrestricted!

Proof

We need to examine two cases.

Case I: if $(\nabla h_j)(\bar{\mathbf{x}})$ are linearly dependent, then one can find $v_1, \ldots, v_\ell$ such that

$$\sum_{j=1}^{m} v_j (\nabla h_j)(\bar{\mathbf{x}}) = \mathbf{0}.$$

The conditions of the theorem are trivially satisfied by taking $u_0 = 0$ and $u_i = 0$ for $i \in I_{\bar{\mathbf{x}}}$.

Proof (cont'd)

Case II: if $(\nabla h_j)(\bar{\mathbf{x}})$ are linearly independent, let

$$A_1 = ((\nabla f)(\bar{\mathbf{x}}) \cdots (\nabla g_i)(\bar{\mathbf{x}}) \cdots) \text{ for } i \in I_{\bar{\mathbf{x}}},$$

and

$$A_2 = ((\nabla h_1)(\bar{\mathbf{x}}) \cdots (\nabla h_\ell)(\bar{\mathbf{x}}))\,.$$

By the precursor theorem, the optimality of $\bar{\mathbf{x}}$ implies that the system

$$A_1'\mathbf{d} < \mathbf{0}, A_2'\mathbf{d} = \mathbf{0}.$$

is incompatible, so there is no $\mathbf{d}$ that satisfies the system.

Proof (cont'd)

Consider the sets

$$S_1 = \left\{ \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} \mid \mathbf{z}_1 = A_1'\mathbf{d}, \mathbf{z}_2 = A_2'\mathbf{d} \right\}$$

$$S_2 = \left\{ \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{0} \end{pmatrix} \mid \mathbf{z}_1 < \mathbf{0} \right\}.$$

S_1 and S_2 are convex sets and $S_1 \cap S_2 = \emptyset$, so by the separation theorem, there exists a non-zero vector

$$\mathbf{p} = \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{pmatrix}$$

such that

$$\mathbf{p}_1' A_1' \mathbf{d} + \mathbf{p}_2' A_2' \mathbf{d} \geqslant \mathbf{p}_1' \mathbf{z}_1 + \mathbf{p}_2' \mathbf{z}_2$$

for each $\mathbf{d} \in \mathbb{R}^n$ and $\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} \in \mathbf{K}(S_2)$.

Proof (cont'd)

$$\mathbf{p}_1' A_1' \mathbf{d} + \mathbf{p}_2' A_2' \mathbf{d} \geqslant \mathbf{p}_1' \mathbf{z}_1 + \mathbf{p}_2' \mathbf{z}_2$$

for each $\mathbf{d} \in \mathbb{R}^n$ and $\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} \in \mathbf{K}(S_2)$.

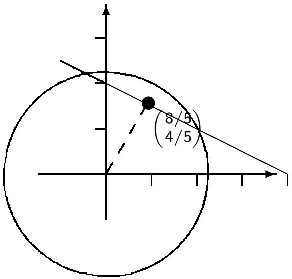
- Let $\mathbf{z}_2 = \mathbf{0}$ and choose $\mathbf{z}_1$ arbitrarily small negative it follows that $\mathbf{p}_1 \geqslant \mathbf{0}$.
- If $\mathbf{z}_1 = \mathbf{0}$ and $\mathbf{z}_2 = \mathbf{0}$ we have

  $$(\mathbf{p}_1' A_1' + \mathbf{p}_2' A_2')\mathbf{d} \geqslant \mathbf{0}$$

  for $\mathbf{d} \in \mathbb{R}^n$. Choosing $\mathbf{d} = -(A_1\mathbf{p}_1 + A_2\mathbf{p}_2)$, we have $- \parallel A_1\mathbf{p}_1 + A_2\mathbf{p}_2 \parallel^2 \geqslant 0$, so $\parallel A_1\mathbf{p}_1 + A_2\mathbf{p}_2 \parallel^2 = 0$, which implies $A_1\mathbf{p}_1 + A_2\mathbf{p}_2 = \mathbf{0}$.
- Denoting the components of $\mathbf{p}_1$ by u_0 and u_i, and letting $\mathbf{v} = \mathbf{p}_2$ the result follows.

Example 4

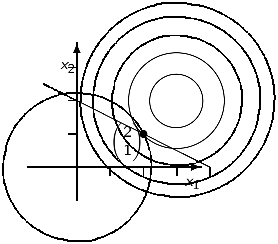
minimize $x_1^2 + x_2^2$
subject to $x_1^2 + x_2^2 \leqslant 5$, $-x_1 \leqslant 0$, $-x_2 \leqslant 0$, $x_1 + 2x_2 = 4$.



Optimal point is $\overline{\mathbf{x}} = \begin{pmatrix} 8/5 \\ 4/5 \end{pmatrix}$ and no binding inequalities exist, so $I_{\overline{\mathbf{x}}} = \emptyset$; multipliers associated to inequality constraints are 0;
$(\nabla f)(\overline{\mathbf{x}}) = \begin{pmatrix} 8/5 \\ 16/5 \end{pmatrix}$ and $(\nabla h_1)(\mathbf{x}) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$; thus, with $u_0 = 5$ and $v_1 = -8$, we have $u_0 \begin{pmatrix} 8/5 \\ 16/5 \end{pmatrix} + v_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \mathbf{0}$.

Example 5

minimize $(x_1 - 3)^2 + (x_2 - 2)^2$
subject to $x_1^2 + x_2^2 \leqslant 5$, $-x_1 \leqslant 0$, $-x_2 \leqslant 0$, $x_1 + 2x_2 = 4$.



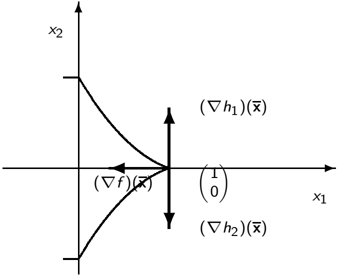
At the optimal point $\bar{\mathbf{x}} = \begin{pmatrix} 2 \end{pmatrix}$ only $x_1^2 + x_2^2 \leqslant 5$ is binding. The Fritz John Condition

$$u_0 \begin{pmatrix} -2 \\ -2 \end{pmatrix} + u_1 \begin{pmatrix} 4 \\ 2 \end{pmatrix} + v_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

can be satisfied, for example, by $u_0 = 3$, $u_1 = 1$, and $v_1 = 2$.

Example 6

minimize $f(\mathbf{x}) = -x_1$
subject to $h_1(\mathbf{x}) = x_2 - (1 - x_1)^3 = 0$, $h_2(\mathbf{x}) = -x_2 - (1 - x_1)^3 = 0$,



The problem has only one feasible point $\bar{\mathbf{x}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; at this point we have

$$(\nabla f)(\bar{\mathbf{x}}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, (\nabla h_1)(\bar{\mathbf{x}}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, (\nabla h_2)(\bar{\mathbf{x}}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

The condition

$$u_0 \begin{pmatrix} -1 \\ 0 \end{pmatrix} + v_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \mathbf{0}$$

implies $u = 0$, $v_1 = v_2 = a$ for any real a. The Fritz John conditions are met at $\bar{\mathbf{x}}$.

Kuhn-Tucker Necessary Conditions

u_0 is not necessarily positive in the FJ conditions. This is remedied by adding additional qualifications on restrictions.

Theorem

Let $\overline{\mathbf{x}}$ be a feasible solution of the problem $\mathcal{O}_{min,\leqslant}(X, f, g_i, h_j)$, where X is an open subset of $\mathbb{R}^n$.
Suppose that f and g_i are differentiable at $\overline{\mathbf{x}}$ and that h_j are continuously differentiable at $\overline{\mathbf{x}}$, and, for $i \not\in I_{\overline{\mathbf{x}}}$, g_i is continuous at $\overline{\mathbf{x}}$. Further, suppose that $h_j(\mathbf{x})$ is continuously differentiable at $\overline{\mathbf{x}}$.
If $(\nabla g_i)(\overline{\mathbf{x}})$ and $(\nabla h_j)(\overline{\mathbf{x}})$ are linearly independent, then there exist u_i for $i \in I_{\overline{\mathbf{x}}}$ and v_j for $1 \leqslant j \leqslant \ell$ such that

$$(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} u_i (\nabla g_i)(\overline{\mathbf{x}}) + \sum_{j=1}^{\ell} v_j (\nabla h_j)(\overline{\mathbf{x}}) = \mathbf{0},$$

and $u_i \geqslant 0$ for $i \in I_{\overline{\mathbf{x}}}$.

If g_i is differentiable at $\mathbf{\overline{x}}$ for $i \notin I_{\mathbf{\overline{x}}}$, then KT conditions can be written as

$$\begin{aligned}
(\nabla f)(\mathbf{\overline{x}}) + \sum_{i \in I_{\mathbf{\overline{x}}}} u_i (\nabla g_i)(\mathbf{\overline{x}}) + \sum_{j=1}^{\ell} v_j (\nabla h_j)(\mathbf{\overline{x}}) &= \mathbf{0}, \\
u_i g_i(\mathbf{\overline{x}}) &= 0 \text{ for } 1 \leqslant i \leqslant m, \\
u_i &\geqslant 0 \text{ for } 1 \leqslant i \leqslant m.
\end{aligned}$$

Proof

By Fritz John's Theorem, there exist u_0, $\hat{u}_i$ for $i \in I_{\overline{\mathbf{x}}}$, and $\hat{v}_j$ for $1 \leqslant j \leqslant \ell$, *not all zero* such that

$$u_0(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} \hat{u}_i (\nabla g_i)(\overline{\mathbf{x}}) + \sum_{j=1}^{\ell} \hat{v}_j (\nabla h_j)(\overline{\mathbf{x}}) = \mathbf{0},$$

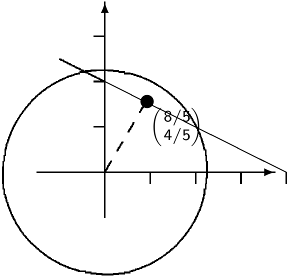
$u_0, \hat{u}_i \geqslant 0$ for $i \in I_{\overline{\mathbf{x}}}$.
Note that $u_0 > 0$ because $u_0 = 0$ would contradict the linear independence of $(\nabla g_i)(\overline{\mathbf{x}})$ for $i \in I_{\overline{\mathbf{x}}}$ and $(\nabla h_j)(\overline{\mathbf{x}})$ for $1 \leqslant j \leqslant \ell$. By defining

$$u_i = \frac{\hat{u}_i}{u_0} \text{ and } v_j = \frac{\hat{v}_i}{u_0},$$

we obtain the desired equality.

Example 4 Revisited

minimize $x_1^2 + x_2^2$
subject to $x_1^2 + x_2^2 \leqslant 5$, $-x_1 \leqslant 0$, $-x_2 \leqslant 0$, $x_1 + 2x_2 = 4$.



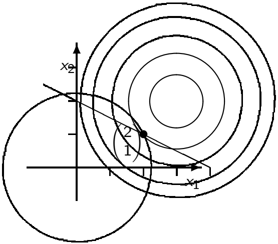
Optimal point is $\bar{\mathbf{x}} = \begin{pmatrix} 8/5 \\ 4/5 \end{pmatrix}$ and $I_{\bar{\mathbf{x}}} = \emptyset$.

$(\nabla f)(\bar{\mathbf{x}}) = \begin{pmatrix} 8/5 \\ 16/5 \end{pmatrix}$, $(\nabla h_1)(\mathbf{x}) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. KT conditions are satisfied at $\bar{\mathbf{x}}$ with

$$u_1 = u_2 = u_3 = 0 \text{ and } v_1 = -\frac{8}{5}.$$

Example 5 Revisited

minimize $(x_1 - 3)^2 + (x_2 - 2)^2$
subject to $x_1^2 + x_2^2 \leqslant 5$, $-x_1 \leqslant 0$, $-x_2 \leqslant 0$, $x_1 + 2x_2 = 4$.

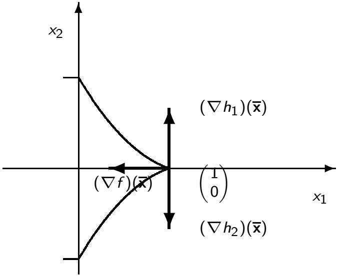


Optimal point is $\overline{\mathbf{x}} = \left(2\right)$. only $x_1^2 + x_2^2 \leqslant 5$ is binding. The KT conditions are satisfied by

$$u_1 = \frac{1}{3}, u_2 = u_3 = 0, v_1 = \frac{2}{3}$$

Example 6 (revisited)

minimize $f(\mathbf{x}) = -x_1$
subject to $h_1(\mathbf{x}) = x_2 - (1 - x_1)^3 = 0,\ h_2(\mathbf{x}) = -x_2 - (1 - x_1)^3 = 0,$



The problem has only one feasible point $\bar{\mathbf{x}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$; at this point we have

$$(\nabla f)(\bar{\mathbf{x}}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, (\nabla h_1)(\bar{\mathbf{x}}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, (\nabla h_2)(\bar{\mathbf{x}}) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

The KT conditions are not met because $(\nabla h_1)(\mathbf{x})$ and $(\nabla h_2)(\mathbf{x})$ are not linearly independent.

Kuhn-Tucker Sufficient Conditions

Consider the problem $\mathcal{O}_{min,\leqslant}(X, f, g_i, h_j)$, where X is an open subset of $\mathbb{R}^n$, let $\overline{\mathbf{x}}$ be a feasible solution and let $I_{\overline{\mathbf{x}}} = \{i \mid g_i(\overline{\mathbf{x}}) = 0\}$.
Suppose that there exist $\overline{u_i} \geqslant 0$ for $i \in I_{\overline{\mathbf{x}}}$ and $\overline{v_j}$ for $1 \leqslant j \leqslant \ell$ such that

$$(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} \overline{u_i}(\nabla g_i)(\overline{\mathbf{x}}) + \sum_{j=1}^{\ell} \overline{v_j}(\nabla h_j)(\overline{\mathbf{x}}) = \mathbf{0}.$$

Let

$$J = \{j \mid \overline{v_j} > 0\} \text{ and } K = \{j \mid \overline{v_j} < 0\}.$$

If f, g_i are convex, h_j is convex for $j \in J$, and concave for $j \in K$, then $\overline{\mathbf{x}}$ is a global optimum for f.

Proof

Let $\mathbf{x}$ be a feasible solution. We have $g_i(\mathbf{x}) \leqslant g_i(\overline{\mathbf{x}})$ because $g_i(\mathbf{x}) \leqslant 0$ and $g_i(\overline{\mathbf{x}}) = 0$ for $i \in I_{\overline{\mathbf{x}}}$.
By the convexity of g_i, we have

$$g_i(\overline{\mathbf{x}} + \lambda(\mathbf{x} - \overline{\mathbf{x}})) = g_i((1-\lambda)\overline{\mathbf{x}} + \lambda\mathbf{x}) \leqslant (1-\lambda)g_i(\overline{\mathbf{x}}) + \lambda g_i(\mathbf{x}) \leqslant g_i(\overline{\mathbf{x}}),$$

for $\lambda \in (0,1)$. So, g_i does not increase when we move from $\overline{\mathbf{x}}$ in the direction of $\mathbf{x} - \overline{\mathbf{x}}$. Thus,

$$(\nabla g_i)(\overline{\mathbf{x}})'(\mathbf{x} - \overline{\mathbf{x}}) \leqslant 0 \text{ for } i \in I_{\overline{\mathbf{x}}}.$$

Proof (cont'd)

Since h_j is convex for $j \in J$ and concave for $j \in K$,

$$\begin{aligned} (\nabla h_j)(\bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}}) &\leqslant 0 \quad \text{for } j \in J, \\ (\nabla h_j)(\bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}}) &\geqslant 0 \quad \text{for } j \in K. \end{aligned}$$

By applying the multiplications

$$\begin{array}{llll} (\nabla g_i)(\bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}}) & \leqslant & 0 \text{ for } i \in I_{\bar{\mathbf{x}}} & (\text{by } \overline{u_i} \geq 0) \\ (\nabla h_j)(\bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}}) & \leqslant 0 & \text{for } j \in J, & (\text{by } \overline{v_j} > 0) \\ (\nabla h_j)(\bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}}) & \geqslant 0 & \text{for } j \in K. & (\text{by } \overline{v_j} < 0) \end{array}$$

and adding, we get

$$\left[\sum_{i \in I_{\bar{\mathbf{x}}}} \overline{u_i} (\nabla g_i)(\mathbf{x}) + \sum_{j \in J \cup K} \overline{v_j} (\nabla h_j)(\bar{\mathbf{x}}) \right]' (\mathbf{x} - \bar{\mathbf{x}}) \leqslant 0,$$

Proof (cont'd)

By multiplying the equality of the hypothesis

$$(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} \overline{u_i}(\nabla g_i)(\overline{\mathbf{x}}) + \sum_{j=1}^{\ell} \overline{v_j}(\nabla h_j)(\overline{\mathbf{x}}) = \mathbf{0}.$$

by $(\mathbf{x} - \overline{\mathbf{x}})$ we have

$$\left[(\nabla f)(\overline{\mathbf{x}}) + \sum_{i \in I_{\overline{\mathbf{x}}}} \overline{u_i}(\nabla g_i)(\overline{\mathbf{x}}) + \sum_{j=1}^{\ell} \overline{v_j}(\nabla h_j)(\overline{\mathbf{x}})\right]' (\mathbf{x} - \overline{\mathbf{x}}) = \mathbf{0}.$$

By the inequality from the previous slide and the fact that $\overline{v_j} = 0$ if $j \notin J \cup K$,

$$\left[\sum_{i \in I_{\overline{\mathbf{x}}}} \overline{u_i}(\nabla g_i)(\mathbf{x}) + \sum_{j \in J \cup K} \overline{v_j}(\nabla h_j)(\overline{\mathbf{x}})\right]' (\mathbf{x} - \overline{\mathbf{x}}) \leqslant 0,$$

we obtain $(\nabla f)(\overline{\mathbf{x}})'(\mathbf{x} - \overline{\mathbf{x}}) \geqslant 0$, which by the convexity of f implies $f(\mathbf{x}) \geqslant f(\overline{\mathbf{x}})$, so $\overline{\mathbf{x}}$ is optimal.

Synopsis (Optimization with inequalities and equalities)

Let $\bar{\mathbf{x}}$ be a feasible solution of the problem $\mathcal{O}_{min,\leqslant}(X, f, g_i, h_j)$, where X is an open subset of $\mathbb{R}^n$.

Suppose that f and g_i are differentiable at $\bar{\mathbf{x}}$ and that h_j are continuously differentiable at $\bar{\mathbf{x}}$, and, for $i \notin I_{\bar{\mathbf{x}}}$, g_i is continuous at $\bar{\mathbf{x}}$. Further, suppose that $h_j(\mathbf{x})$ is continuously differentiable at $\bar{\mathbf{x}}$.

Necessary Conditions: If $(\nabla g_i)(\bar{\mathbf{x}})$ and $(\nabla h_j)(\bar{\mathbf{x}})$ are linearly independent, then there exist u_i for $i \in I_{\bar{\mathbf{x}}}$ and v_j for $1 \leqslant j \leqslant \ell$ such that

$$(\nabla f)(\bar{\mathbf{x}}) + \sum_{i \in I_{\bar{\mathbf{x}}}} u_i (\nabla g_i)(\bar{\mathbf{x}}) + \sum_{j=1}^{\ell} v_j (\nabla h_j)(\bar{\mathbf{x}}) = \mathbf{0},$$

and $u_i \geqslant 0$ for $i \in I_{\bar{\mathbf{x}}}$.

Sufficient Conditions: If $\overline{u_i} \geqslant 0$ for $i \in I_{\bar{\mathbf{x}}}$ and $\overline{v_j}$ for $1 \leqslant j \leqslant \ell$ such that

$$\begin{aligned} &(\nabla f)(\bar{\mathbf{x}}) + \textstyle\sum_{i \in I_{\bar{\mathbf{x}}}} \overline{u_i} (\nabla g_i)(\bar{\mathbf{x}}) + \sum_{j=1}^{\ell} \overline{v_j} (\nabla h_j)(\bar{\mathbf{x}}) = \mathbf{0}, \\ &J = \{j \mid \overline{v_j} > 0\} \text{ and } K = \{j \mid \overline{v_j} < 0\}, \end{aligned}$$

f, g_i are convex, h_j is convex for $j \in J$, and concave for $j \in K$, then $\bar{\mathbf{x}}$ is a global optimum for f.