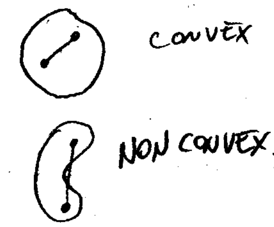
§12 Theory of constrained optimization



Preliminaries:

Def (Convex Set) C is convex iff

$$\forall x, y \in C \;\; \forall \lambda \in [0,1] \;\; \lambda x + (1-\lambda) y \in C$$

Def (Cone) K is a cone iff

$$\forall \lambda \geq 0,\; x \in K \Rightarrow \lambda x \in K$$

Def (Symmetric cone) K is a symm cone iff
K is a cone and $x \in K \Leftrightarrow -x \in K$.

We consider Non-Linear Programming (NLP) of the form:

$$(NLP) \begin{cases} \min & f(x) \\ \text{s.t.} & c_i(x) = 0 \quad i = 1 \ldots m_E \\ & c_i(x) \geq 0 \quad i = m_{E+1} \ldots m \end{cases}$$

$f : \mathbb{R}^n \to \mathbb{R}$
$c_i(x) : \mathbb{R}^n \to \mathbb{R}, \; i = 1 \ldots m$
$m_E = \#$ of equality constraints

$\underline{c} : \mathbb{R}^n \to \mathbb{R}^m$ s.t. $\underline{c}(x) = \begin{pmatrix} c_1(x) \\ c_2(x) \\ \vdots \\ c_m(x) \end{pmatrix}$

Def (Feasible set)

$$\mathcal{F} = \{ x \mid c_i(x) = 0,\; i = 1 \ldots m_E \text{ and } c_i(x) \geq 0,\; i = m_{E+1}, \ldots, m \}$$

Def (local min)
$\bar{x}$ is a local min of (NLP) if $\bar{x} \in \mathcal{F}$ and if there exists $\delta > 0$
s.t. $\forall x \in \mathcal{F},\; \|x - \bar{x}\| < \delta \Rightarrow f(\bar{x}) \leq f(x)$.

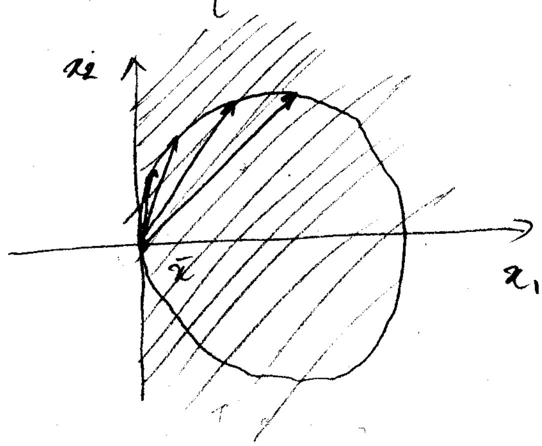
Def (Tangent cone) For $\bar{x} \in \mathcal{F}$, the tangent cone to $\mathcal{F}$ at $\bar{x}$ is:

$$T(\mathcal{F}, \bar{x}) = \left\{ d \mid \exists \{t_k\}_k,\; \exists \{x_k\}_k : t_k \geq 0,\; x_k \in \mathcal{F},\; \lim_{k\to\infty} x_k = \bar{x} \text{ and } \lim_{k \to \infty} t_k (x_k - \bar{x}) = d \right\}$$

Tangent cone = cone of all limiting directions of feasible sequences

Example:

1) $\mathcal{F} = \{x \in \mathbb{R}^2 \mid (x_1-1)^2 + x_2^2 \leq 1\}$



For $\bar{x} = (0,0)$:

$T(\mathcal{F},\bar{x}) = \{d \mid d_1 \geq 0\}$

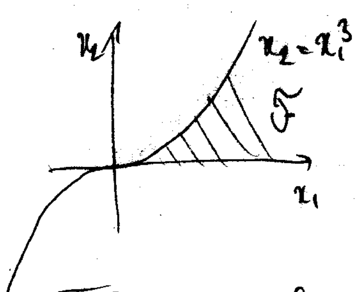
= Right handplane

proof: for $d \in \mathbb{R}^2$ s.t. $d_1 > 0$ we let

$x_k = \bar{x} + \frac{1}{k} d$, $t_k = k$

2) $\mathcal{F} = \{x \in \mathbb{R}^2 \mid x_2 \geq 0,\ x_2 \leq x_1^3\}$

with $\bar{x} = (0,0)$:



$T(\mathcal{F},\bar{x}) = \{d \in \mathbb{R}^2 \mid d_1 \geq 0,\ d_2 = 0\}$

Theorem The set $T(\mathcal{F},\bar{x})$ is a non-empty closed cone.

proof: cone: $d \in T(\mathcal{F},\bar{x}) \Rightarrow \exists\, t_k, x_k$ s.t. $\lim x_k = \bar{x}$, $\lim t_k(x_k - \bar{x}) = d$

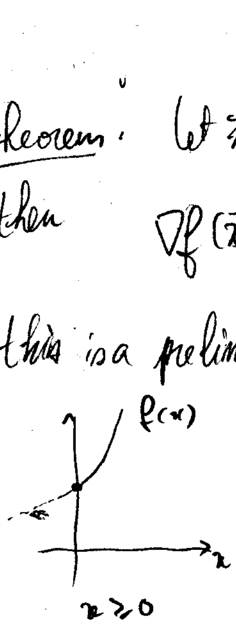
$\Rightarrow \lim \lambda t_k(x_k - \bar{x}) = \lambda d$.

closed: it is possible to show that if $d_k \in T(\mathcal{F},\bar{x})$ then $\lim_{k\to\infty} d_k = d$

Theorem: Let $\bar{x}$ be a local min of (NLP), if f is Frechet diff'ble at $\bar{x}$ then $\nabla f(\bar{x})^T d \geq 0 \quad \forall d \in T(\mathcal{F},\bar{x})$

(this is a preliminary necessary cdt° for $\bar{x}$ to be a minimizer)

e.g.



$\min f(x)$
$x \geq 0$.

have $f'(x) \geq 0$ necessarily.

Proof:

Let $d \in T(\mathcal{F},\bar{x})$, $\{t_k\}_k$, $\{x_k\}_k$ with
$t_k \geq 0$, $x_k \in \mathcal{F}$ s.t. $\lim x_k = \bar{x}$
$\lim t_k (x_k - \bar{x}) = d$

Let δ be s.t. $f(\bar{x}) \leq f(x)$ for all $x \in \mathcal{F}$ s.t. $\|x-\bar{x}\| < \delta$

Since f is Fréchet diff'ble:

$$f(x_k) = f(\bar{x}) + \nabla f(\bar{x})^T (x_k - \bar{x}) + r(x_k)$$

with $\lim\limits_{x_k \to \bar{x}} \dfrac{r(x_k)}{\|x_k - \bar{x}\|} = 0$

$\exists\, k_0 \in \mathbb{N}$ s.t. $\|x_k - \bar{x}\| < \delta$ for all $k \geq k_0$ (since $x_k \to \bar{x}$)

$$\Rightarrow f(\bar{x}) \leq f(x_k) \quad \forall k \geq k_0$$

$$\Rightarrow 0 \leq \nabla f(\bar{x})^T (x_k - \bar{x}) + r(x_k)$$

$$\Rightarrow 0 \leq \underbrace{t_k \nabla f(\bar{x})^T (x_k - \bar{x})}_{\to \nabla f(\bar{x})^T d} + t_k r(x_k)$$

We conclude by noticing

$\lim \dfrac{t_k\, r(x_k)}{t_k \|x_k - \bar{x}\|} = 0$, since $\lim t_k \|x_k - \bar{x}\| = \|d\|$
we must have $\lim t_k r(x_k) = 0$

<u>Def (Linearized cone)</u> Let $\bar{x} \in \mathcal{F}$ and let $\underline{c}$ be Fréchet diff'ble, then the linearized cone is:

$$L(\mathcal{F}, \bar{x}) = \{ d \mid \underline{c}(\bar{x}) + D\underline{c}(\bar{x}) d \in K \}$$

means we look at directions where, according to local linear model at $\bar{x}$, we remain in the feasible set.

Here $K = \{ x_i = 0,\ i = 1 .. m_E \text{ and } x_i \geq 0,\ i = m_E + 1, \dots, m \}$

Examples

1) $\mathcal{F} = \{x \in \mathbb{R}^2 \mid (x_1-1)^2 + x_2^2 \leq 1\}$

Here $c(x_1, x_2) = 1 - (x_1-1)^2 - x_2^2$

and $K = \{y \in \mathbb{R} \mid y \geq 0\}$

Then for $\bar{x} = \begin{pmatrix}0\\0\end{pmatrix}$,

$$L(\mathcal{F}, \bar{x}) = \{d \mid c(\bar{x}) + \underbrace{Dc(\bar{x})}_{\nabla c(\bar{x})^T} d \geq 0\}$$

$$= \{d \mid 0 + 2d_1 + 0d_2 \geq 0\}$$

$$= \{d \mid d_1 \geq 0\} = T(\mathcal{F}, \bar{x})$$

We do not always have $L(\mathcal{F}, \bar{x}) = T(\mathcal{F}, \bar{x})$,

in general only $L(\mathcal{F}, \bar{x}) \supset T(\mathcal{F}, \bar{x})$

(since the linear model is not perfect, some directions deemed feasible by linear model may not be for the real constraints)

2) $\mathcal{F} = \{x \in \mathbb{R}^2 \mid x_2 \geq 0,\ x_2 \leq x_1^3\}$

$K = \{y \in \mathbb{R}^2 \mid y_1, y_2 \geq 0\}$ (positive orthant)

$$c(x_1, x_2) = \begin{pmatrix} x_2 \\ x_1^3 - x_2 \end{pmatrix}$$

$$Dc(x_1, x_2) = \begin{bmatrix} 0 & 1 \\ 3x_1^2 & -1 \end{bmatrix}$$

For $\bar{x} = (0,0)$: $c(\bar{x}) = (0,0)$

$$c(\bar{x}) + Dc(\bar{x})d = \begin{pmatrix}0\\0\end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} \geq 0$$

$\Rightarrow$

$$L(\mathcal{F}, \bar{x}) = \{d \mid d_1 \in \mathbb{R},\ d_2 = 0\},$$

(compare to $T(\mathcal{F}, \bar{x}) = \{d \mid d_1 \geq 0,\ d_2 = 0\}$)

Since linearized cone is nice to work with we need to impose some <u>constraint qualifications</u> (regularity cdt°) that guarantees $L(\mathcal{F},\bar{x}) = T(\mathcal{F},\bar{x})$.

Recall NLP:

$$\begin{cases} \min f(x) \\ \text{s.t. } c_i(x) = 0, \ i = 1,\dots,m_E \\ \quad\ c_i(x) \geq 0, \ i = m_{E+1},\dots,m \end{cases}$$

$f: \mathbb{R}^n \to \mathbb{R}$

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

$x \to \begin{pmatrix} c_1(x) \\ \vdots \\ c_m(x) \end{pmatrix}$

here $K = \{ y \in \mathbb{R}^m, y_i = 0, i = 1..m_E, y_i \geq 0, i = m_{E+1},\dots,m \}$

The set of <u>active constraints</u>

$$\mathcal{A}(\bar{x}) = \{1,\dots,m_E\} \cup \{ i \in \{m_{E+1}\dots m\} \mid c_i(\bar{x}) = 0 \}$$

= all constraints that are equalities at $\bar{x}$.

The set of <u>inactive constraints</u> = $\mathcal{I}(\bar{x}) = \{1..m\} \setminus \mathcal{A}(\bar{x})$

= all constraints that are strict inequalities ($c_i(x) > 0$)

(some room to change these)

<u>Theorem</u> (Mangasarian Fromovitch Constraint Qualification (MFCQ))

Consider NLP, let $\bar{x}$ be a feasible point. If $\exists \bar{w} \in \mathbb{R}^n$ s.t.

$$\nabla c_i(\bar{x})^T \bar{w} > 0 \qquad \forall i \in \{m_{E+1},\dots,m\} \cap \mathcal{A}(\bar{x})$$

$$\nabla c_i(x)^T \bar{w} = 0 \qquad \forall i \in \{1,\dots m_E\}$$

and if $\nabla c_1(\bar{x}), \dots, \nabla c_{m_E}(\bar{x})$ are lin indep

then $L(\mathcal{F},\bar{x}) = T(\mathcal{F},\bar{x})$.

Note: MFCQ implies $m_E < n$ in practice since:

$$\underbrace{\begin{pmatrix} \nabla c_1^T(\bar{x}) \\ \vdots \\ \nabla c_{m_E}^T(\bar{x}) \end{pmatrix}}_{\in \mathbb{R}^{m_E \times n} \text{ with rank} = m_E} \bar{w} = 0$$

the other alternative $m_E \geq n$ is likely to give an empty feasible set.

Note: It is much easier to verify the following (stronger) CQ:

Theorem: Linear Indep. Constr. Qual (LICQ)

Consider (NLP). If $\nabla c_i(\bar{x})$, $i \in \mathcal{A}(\bar{x})$ are linearly indep

$\Rightarrow$ MFCQ is satisfied $\Rightarrow T(\mathcal{F}, \bar{x}) = L(\mathcal{F}, \bar{x})$

proof: If LICQ is satisfied then the lin sys. of eq

$$\begin{cases} \nabla c_i(\bar{x})^T w = 1, & \text{for } i \in \mathcal{A}(\bar{x}) \cap \{m_E+1, \dots, m\} \\ \nabla c_i(\bar{x})^T w = 0, & \text{for } i \in \{1 \dots m_E\} \end{cases}$$

has exactly one solution w $\to$ the w in MFCQ.

§ 12.3 First order optimality conditions

Define the Lagrangian function:

$$L(x, \lambda) = f(x) + \lambda^T \underline{c}(x)$$

Theorem: Assume $\bar{x}$ is a local min of (NLP), with f, c cont diff'ble, and $\bar{x}$ is regular (e.g. LICQ holds at $\bar{x}$). Then

$\exists\ \bar{\lambda} \in \mathbb{R}^m$ s.t. (Lagrange multiplier)

i) $\nabla_x L(\bar{x}, \bar{\lambda}) = \nabla f(\bar{x}) + Dc(\bar{x})^T \bar{\lambda} = 0$

ii) $\bar{\lambda}_i \leq 0$, $i = m_E+1, \dots, m$

iii) $\bar{\lambda}_i c_i(\bar{x}) = 0$, $i = 1 \dots m$ complementarity cdt

Note

- These conditions are known as <u>KKT</u> conditions (as in Karush - Kuhn - Tucker)
- $\bar{\lambda}$ is called "Lagrange multiplier" and there is no real standard for taking it ≤ 0 (as we do here) or ≥ 0 (as in book)
- Condition iii) implies that either $\lambda_i = 0$ or $c_i(\underline{x}) = 0$
- If LICQ holds then $\bar{\lambda}$ is unique.

We will often mention "<u>strict complementarity</u>" which means that for $i = m_E+1, \ldots, m$, exactly one of $\bar{\lambda}_i$ and $c_i(\bar{x})$ is zero ($\Leftrightarrow$ $\bar{\lambda}_i < 0$ for $i \in \{m_E+1, \ldots, m\} \cap \mathcal{A}(\bar{x})$)

- A more general version of the KKT conditions can be formulated for:

$$\begin{aligned} &\min f(x) \\ &c(\underline{x}) \in K \\ &x \in C \end{aligned}$$

, where K = cone $\subset \mathcal{Y}$, C = convex set

is: $\exists\, \bar{\lambda}$ s.t.

i) $Df(\bar{x})(x-\bar{x}) + \langle \bar{\lambda}, Dc(\bar{x})(x-\bar{x}) \rangle \geq 0, \forall x \in C$

ii) $\bar{\lambda} \in K^D = \{ y \in \mathcal{Y} \mid \langle y, x \rangle \leq 0 \ \forall x \in K \}$ = dual cone

iii) $\langle \bar{\lambda}, \underline{c}(\bar{x}) \rangle = 0$ (complementarity condition)

Second order optimality conditions :

$x \in \mathbb{R}^n$

$$\text{(NLP)} \quad \begin{aligned} &\min f(x) \\ &\text{s.t.} \quad c_i(x) = 0, \; i = 1 \dots m_E \\ &\qquad\; c_i(x) \geq 0, \; i = m_E+1, \dots, m \end{aligned}$$

Thm (Second order necessary optimality conditions)

Let $\bar{x} \in \mathcal{F}$ be a local min of (NLP). If $\exists \bar{\lambda}$

i) $\bar{\lambda} \leq 0$, $i = m_E+1, \dots, m$

ii) $\bar{\lambda}_i \, c_i(\bar{x}) = 0 \quad i = 1 \dots m$

iii) $\nabla_x L(\bar{x}, \bar{\lambda})$

(guaranteed if $\bar{x}$ is a regular point, i.e. CQ hold at $\bar{x}$).

then $\quad d^T \nabla_x^2 L(\bar{x}, \bar{\lambda}) \, d \geq 0 \quad \forall d \in T(\mathcal{F}_1, \bar{x})$

where $\mathcal{F}_1 = \{ x \in \mathcal{F} \mid c_i(x) = 0, \; i \in \{m_{E+1}, \dots, m\} \text{ s.t. } \bar{\lambda}_i < 0 \}$.

We have $\mathcal{F}_1 \subset \mathcal{F}$ and assuming regularity of $\bar{x}$:

$$T(\mathcal{F}_1, \bar{x}) = L(\mathcal{F}_1, \bar{x}) = \left\{ d \;\middle|\; \begin{aligned} &\nabla c_i(\bar{x})^T d = 0, \; i = 1 \dots m_E, \\ &\nabla c_i(\bar{x})^T d = 0, \; i \in \{m_{E+1}, \dots, m\} \cap \mathcal{A}(\bar{x}) \text{ with } \bar{\lambda}_i < 0 \\ &\nabla c_i(\bar{x})^T d \geq 0, \; i \in \{m_{E+1} \dots m\} \cap \mathcal{A}(\bar{x}) \text{ with } \bar{\lambda}_i = 0. \end{aligned} \right\}$$

The cone $T(\mathcal{F}_1, \bar{x})$ contains directions for which it is not clear to first order whether the function will increase or decrease.

if $d \in T(\mathcal{F}_1, \bar{x})$ then:

$$d^T \nabla f(\bar{x}) = -\sum_{i=1}^{m} \bar{\lambda}_i \, d^T \nabla c_i(\bar{x}) = 0$$

$f(x+d) \simeq f(\bar{x}) + \underbrace{d^T \nabla f(\bar{x})}_{=0}$ so linear model is insufficient for these directions

Second order sufficient optimality conditions

Let $\bar{x} \in \mathcal{F}$. Let f be twice cont. diff'ble in a nbd of $\bar{x}$.

If there is $\bar{\lambda} \in \mathbb{R}^m$ s.t.

i) $\bar{\lambda}_i \leq 0$, $i = m_E, \dots m$

ii) $\bar{\lambda}_i c_i(\bar{x}) = 0$, $i = 1 .. m$

iii) $\nabla_x L(\bar{x}, \bar{\lambda}) = 0$

and if for all $d \in L(\mathcal{F}_1, \bar{x})$, $d \neq 0$:

$$d^T \nabla^2 L(\bar{x}, \bar{\lambda})\, d > 0$$

Then there is $\delta > 0$, $\varepsilon > 0$ s.t.

$$f(x) \geq f(\bar{x}) + \varepsilon \|x - \bar{x}\|^2 \quad \forall x \in B_\delta(\bar{x}) \cap \mathcal{F}$$