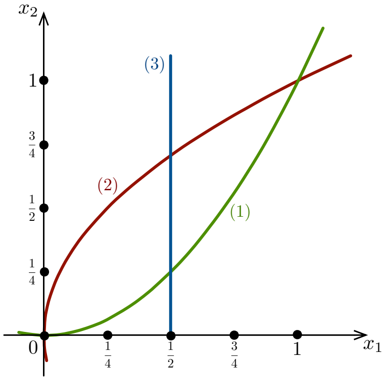
CO 250: Introduction to Optimization

Module 6: Nonlinear Programs (the KKT theorem)

$$\begin{array}{lrcll} \min & -x_1 - x_2 & & & \\ \text{s.t.} & & & & \\ & -x_2 + x_1^2 & \leq & 0 & (1) \\ & -x_1 + x_2^2 & \leq & 0 & (2) \\ & -x_1 + \frac{1}{2} & \leq & 0 & (3) \end{array}$$

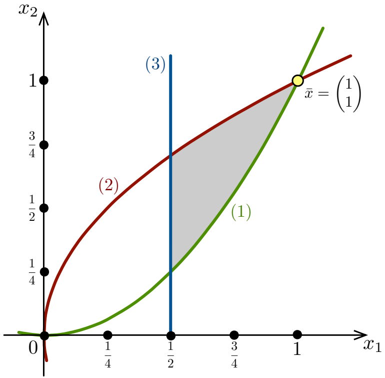


(1) $x_2 \geq x_1^2$;

(2) $x_1 \geq x_2^2$;

(3) $x_1 \geq \frac{1}{2}$.

$$
\begin{array}{llcll}
\min & -x_1 - x_2 & & & \\
\text{s.t.} & & & & \\
& -x_2 + x_1^2 & \leq & 0 & (1) \\
& -x_1 + x_2^2 & \leq & 0 & (2) \\
& -x_1 + \frac{1}{2} & \leq & 0 & (3)
\end{array}
$$



Claim

$\bar{x} = (1, 1)^\top$ is an optimal solution to the NLP.

How do we prove this?

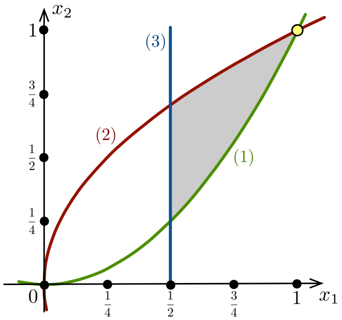
Step 1. Find a relaxation of the NLP.

Step 2. Prove $\bar{x}$ is optimal for the relaxation.

Step 3. Deduce that $\bar{x}$ is optimal for the NLP.

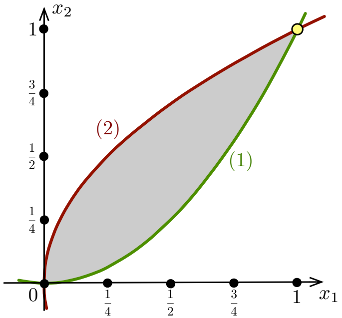
Original NLP

$$\begin{array}{llcll} \min & -x_1 - x_2 & & & \\ \text{s.t.} & & & & \\ & -x_2 + x_1^2 & \leq & 0 & (1) \\ & -x_1 + x_2^2 & \leq & 0 & (2) \\ & -x_1 + \frac{1}{2} & \leq & 0 & (3) \end{array}$$



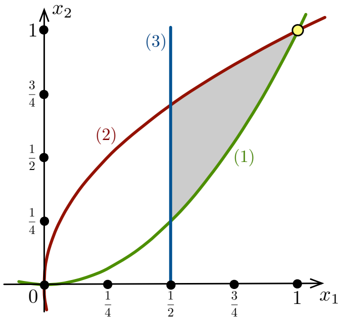
Relaxation

$$\begin{array}{llcll} \min & -x_1 - x_2 & & & \\ \text{s.t.} & & & & \\ & -x_2 + x_1^2 & \leq & 0 & (1) \\ & -x_1 + x_2^2 & \leq & 0 & (2) \end{array}$$



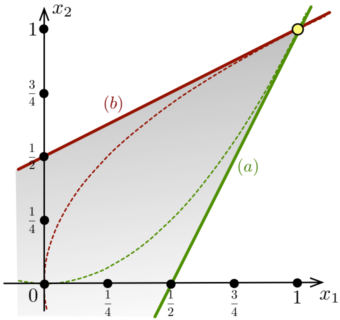
Original NLP

$$\begin{array}{llll} \min & -x_1 - x_2 & & \\ \text{s.t.} & & & \\ & -x_2 + x_1^2 & \leq 0 & (1) \\ & -x_1 + x_2^2 & \leq 0 & (2) \\ & -x_1 + \frac{1}{2} & \leq 0 & (3) \end{array}$$



New relaxation

$$\begin{array}{llll} \min & -x_1 - x_2 & & \\ \text{s.t.} & & & \\ & 2x_1 - x_2 & \leq 1 & (a) \\ & -x_1 + 2x_2 & \leq 1 & (b) \end{array}$$



Claim

$\bar{x} = (1,1)^\top$ is an optimal solution to

$$
\boxed{
\begin{array}{llrcll}
\min & -x_1 - x_2 & & & & \\
\text{s.t.} & & & & & \\
& & 2x_1 - x_2 & \leq & 1 & (a) \\
& & -x_1 + 2x_2 & \leq & 1 & (b)
\end{array}
}
$$

Claim

$\bar{x} = (1,1)^\top$ is an optimal solution to

$$\begin{array}{lrcll} \max & x_1 + x_2 & & & \\ \text{s.t.} & & & & \\ & 2x_1 - x_2 & \leq & 1 & (a) \\ & -x_1 + 2x_2 & \leq & 1 & (b) \end{array}$$

Proof

Tight constraints for $\bar{x}$ are (a) and (b).

Goal: Show that the objective function is in the cone of tight constraints.

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \stackrel{?}{\in} cone \left\{ \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \end{pmatrix} \right\} \quad \Longleftarrow$$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1 \times \begin{pmatrix} 2 \\ -1 \end{pmatrix} + 1 \times \begin{pmatrix} -1 \\ 2 \end{pmatrix} \quad \checkmark$$

Original NLP

$$
\begin{array}{llll}
\min & -x_1 - x_2 & & \\
\text{s.t.} & & & \\
& -x_2 + x_1^2 & \leq 0 & (1) \\
& -x_1 + x_2^2 & \leq 0 & (2) \\
& -x_1 + \frac{1}{2} & \leq 0 & (3)
\end{array}
$$

Relaxation

$$
\begin{array}{llll}
\min & -x_1 - x_2 & & \\
\text{s.t.} & & & \\
& 2x_1 - x_2 & \leq 1 & (a) \\
& -x_1 + 2x_2 & \leq 1 & (b)
\end{array}
$$

$\bar{x} = (1,1)^\top$ is an optimal solution to the relaxation

$\Rightarrow$ $\bar{x}$ is an optimal solution to the *original NLP*

Question

Can we do this in general? YES

The key tool we'll use is subgradients.

Definition

Let $f : \Re^n \to \Re$ be a convex function and $\bar{x} \in \Re^n$.

Then, $s \in \Re^n$ is a subgradient of f at $\bar{x}$ if

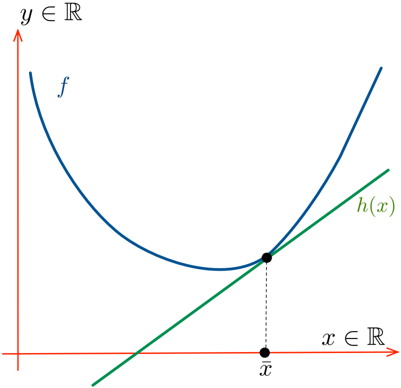
$$h(x) := f(\bar{x}) + s^\top (x - \bar{x}) \leq f(x) \qquad \text{for all } x \in \Re^n.$$

$h(x)$ is affine

$h(\bar{x}) = f(\bar{x})$

h is a lower bound for f

unique subgradient



Definition

Let $f : \Re^n \to \Re$ be a convex function and $\bar{x} \in \Re^n$.

Then, $s \in \Re^n$ is a subgradient of f at $\bar{x}$ if

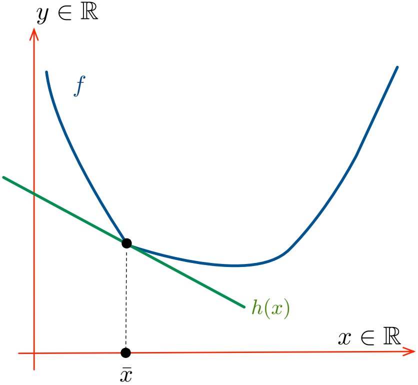
$$h(x) := f(\bar{x}) + s^\top (x - \bar{x}) \leq f(x) \qquad \text{for all } x \in \Re^n.$$

$h(x)$ is affine

$h(\bar{x}) = f(\bar{x})$

h is a lower bound for f

first subgradient



Definition

Let $f : \Re^n \to \Re$ be a convex function and $\bar{x} \in \Re^n$.

Then, $s \in \Re^n$ is a subgradient of f at $\bar{x}$ if

$$h(x) := f(\bar{x}) + s^\top (x - \bar{x}) \leq f(x) \qquad \text{for all } x \in \Re^n.$$

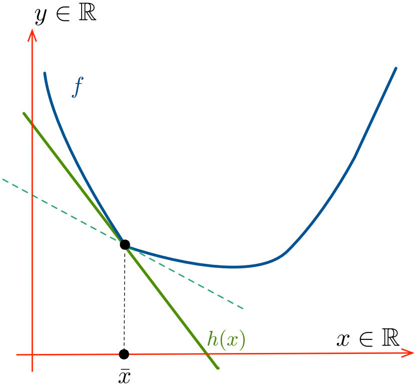
$h(x)$ is affine

$h(\bar{x}) = f(\bar{x})$

h is a lower bound for f

second subgradient

NOT UNIQUE



Definition

$s \in \Re^n$ is a subgradient of f at $\bar{x}$ if

$$h(x) := f(\bar{x}) + s^\top (x - \bar{x}) \leq f(x) \qquad \text{for all } x \in \Re^n.$$

Example

Consider $f : \Re^2 \to \Re$ where $f(x) = -x_1 + x_2^2$ and $\bar{x} = (1, 1)^\top$.

We claim that $(-1, 2)^\top$ is a subgradient of f at $\bar{x}$.

$$\begin{aligned} h(x) &= f(\bar{x}) + s^\top (x - \bar{x}) \\ &= 0 + (-1, 2)(x - (1, 1)^\top) = -x_1 + 2x_2 - 1. \end{aligned}$$

Check: $h(x) \leq f(x)$ for all $x \in \Re^n$.

$$-x_1 + 2x_2 - 1 \overset{?}{\leq} -x_1 + x_2^2$$

or equivalently,

$$x_2^2 - 2x_2 + 1 \overset{?}{\geq} 0,$$

which is the case as $x_2^2 - 2x_2 + 1 = (x_2 - 1)^2 \geq 0$.

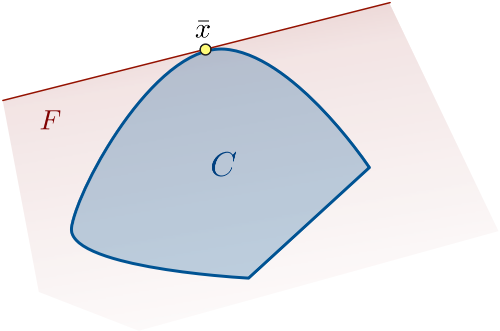
Definition

Let $C \in \Re^n$ be a convex set and let $\bar{x} \in C$.

The halfspace $F = \{x : s^\top x \leq \beta\}$ is supporting C at $\bar{x}$ if

(1) $C \subseteq F$ and

(2) $s^\top \bar{x} = \beta$. That is, $\bar{x}$ is on the boundary of F.



UNIQUE SUPPORTING HALFSPACE AT $\bar{x}$.

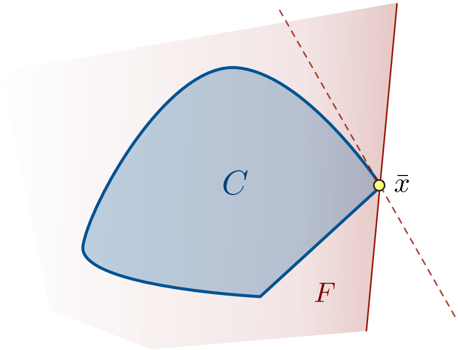
Definition

Let $C \in \Re^n$ be a convex set and let $\bar{x} \in C$.

The halfspace $F = \{x : s^\top x \leq \beta\}$ is supporting C at $\bar{x}$ if

(1) $C \subseteq F$ and

(2) $s^\top \bar{x} = \beta$. That is, $\bar{x}$ is on the boundary of F.



NON-UNIQUE SUPPORTING HALFSPACE AT $\bar{x}$.

Definition

Let $C \in \Re^n$ be a convex set and let $\bar{x} \in C$.

The halfspace $F = \{x : s^\top x \leq \beta\}$ is supporting C at $\bar{x}$ if

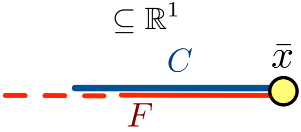
(1) $C \subseteq F$ and

(2) $s^\top \bar{x} = \beta$. That is, $\bar{x}$ is on the boundary of F.

Question

What do we get when $n = 1$?

- C is a segment (or a halfline)
- F is a halfline



Subgradients and Supporting Halfspaces

Proposition

Let $g : \Re^n \to \Re$ be convex and let $\bar{x}$ where $g(\bar{x}) = 0$.

Let s be a subgradient of g at $\bar{x}$.

Let $C = \{x : g(x) \leq 0\}$.

Let $F = \{x : h(x) := g(\bar{x}) + s^\top (x - \bar{x}) \leq 0\}$.

Then, F is a supporting halfspace of C at $\bar{x}$.

Remark

- C is convex, as g is a convex function,
- F is a halfspace, as $h(x)$ is an affine function, and
- $h(\bar{x}) = g(\bar{x}) = 0$ thus, $\bar{x}$ is on the boundary of F.

Proposition

Let $g : \Re^n \to \Re$ be convex and let $\bar{x}$ where $g(\bar{x}) = 0$.

Let s be a subgradient of g at $\bar{x}$.

Let $C = \{x : g(x) \leq 0\}$.

Let $F = \{x : h(x) := g(\bar{x}) + s^\top (x - \bar{x}) \leq 0\}$.

Then, F is a supporting halfspace of C at $\bar{x}$.

Example

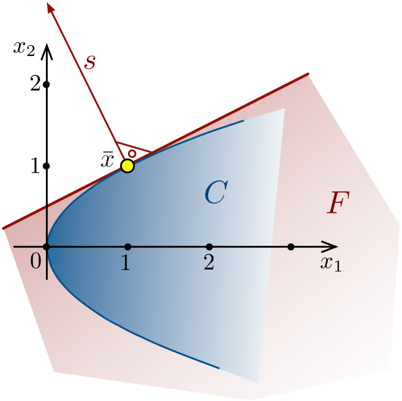
$g(x) = x_2^2 - x_1$

$\bar{x} = (1,1)^\top$

$s = (-1,2)^\top$ subgradient at $\bar{x}$

$$\begin{aligned} h(x) &= 0 + (-1,2)\left[\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] \\ &= -x_1 + 2x_2 - 1 \end{aligned}$$

$F = \{x : -x_1 + 2x_2 \leq 1\}$



Proposition

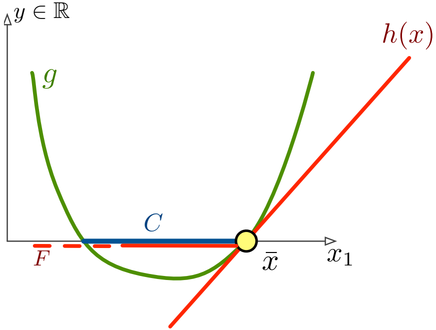
Let $g : \Re^n \to \Re$ be convex and let $\bar{x}$ where $g(\bar{x}) = 0$.

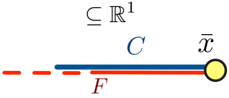
Let s be a subgradient of g at $\bar{x}$.

Let $C = \{x : g(x) \leq 0\}$.

Let $F = \{x : h(x) := g(\bar{x}) + s^\top (x - \bar{x}) \leq 0\}$.

Then, F is a supporting halfspace of C at $\bar{x}$.





Proposition

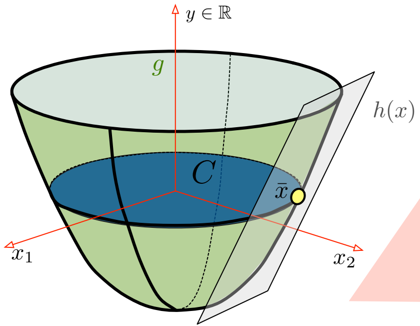
Let $g : \Re^n \to \Re$ be convex and let $\bar{x}$ where $g(\bar{x}) = 0$.

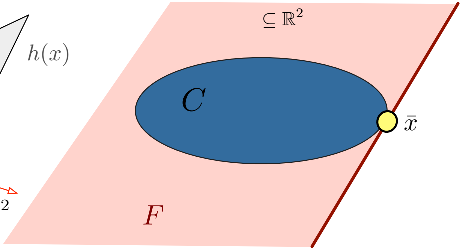
Let s be a subgradient of g at $\bar{x}$.

Let $C = \{x : g(x) \le 0\}$.

Let $F = \{x : h(x) := g(\bar{x}) + s^\top (x - \bar{x}) \le 0\}$.

Then, F is a supporting halfspace of C at $\bar{x}$.





Proposition

Let $g : \Re^n \to \Re$ be convex and let $\bar{x}$ where $g(\bar{x}) = 0$.

Let s be a subgradient of g at $\bar{x}$.

Let $C = \{x : g(x) \leq 0\}$.

Let $F = \{x : h(x) := g(\bar{x}) + s^\top (x - \bar{x}) \leq 0\}$.

Then, F is a supporting halfspace of C at $\bar{x}$.

Proof

Claim: $C \subseteq F$.

Let $x \in C$ and thus, $g(x) \leq 0$

By definition of a subgradient, we know that $h(x) \leq g(x)$.

It follows that $h(x) \leq g(x) \leq 0$.

Hence, $x \in F$.

Claim: $h(\bar{x}) = 0$

$h(\bar{x}) = g(\bar{x}) = 0$.

Proposition

Let $g : \Re^n \to \Re$ be convex and let $\bar{x}$ where $g(\bar{x}) = 0$.

Let s be a subgradient of g at $\bar{x}$.

Let $C = \{x : g(x) \leq 0\}$.

Let $F = \{x : h(x) := g(\bar{x}) + s^\top (x - \bar{x}) \leq 0\}$.

Then, F is a supporting halfspace of C at $\bar{x}$.

Question

Why is this relevant for us?

$\Rightarrow$

WE USE IT TO CONSTRUCT RELAXATIONS OF NLPS

$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & \\ & g_i(x) \leq 0 \qquad (i = 1, \ldots, k) \end{array}$$

$\bar{x}$ is a feasible solution
g_1 is convex
$g_1(\bar{x}) = 0$
s is a subgradient for g_1 at $\bar{x}$

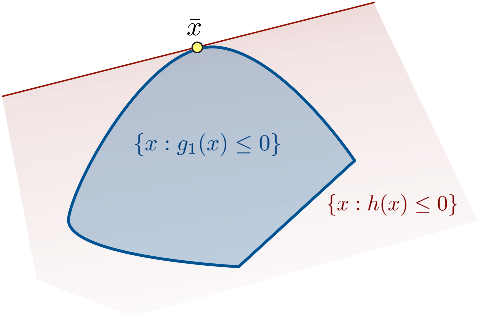
If we replace the nonlinear constraint

$$g_1(x) \leq 0$$

with the linear constraint

$$h(x) = g_1(\bar{x}) + s^\top (x - \bar{x}) \leq 0$$

we get a relaxation.



Proposition

$$\begin{array}{|l|}\hline \min \quad c^\top x \\ \text{s.t.} \\ \quad g_i(x) \le 0 \quad (i = 1, \dots, k) \\ \hline \end{array}$$

$g_1, \dots, g_k$ all convex
$\bar{x}$ is a feasible solution
$\forall i \in I,\ g_i(\bar{x}) = 0$
$\forall i \in I,\ s^{(i)}$ subgradient for g_i at $\bar{x}$

If $-c \in cone \left\{ s^{(i)} : i \in I \right\}$ then $\bar{x}$ is optimal.

Example

$$\begin{array}{|l|}\hline \min \quad -x_1 - x_2 \\ \text{s.t.} \\ \qquad -x_2 + x_1^2 \le 0 \quad (1) \\ \qquad -x_1 + x_2^2 \le 0 \quad (2) \\ \qquad -x_1 + \tfrac{1}{2} \le 0 \quad (3) \\ \hline \end{array}$$

$\bar{x} = (1, 1)^\top$ feasible
$I = \{1, 2\}$
$(2, -1)^\top$ subgradient for g_1 at $\bar{x}$
$(-1, 2)^\top$ subgradient for g_2 at $\bar{x}$

$$-\begin{pmatrix} -1 \\ -1 \end{pmatrix} \in cone \left\{ \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \end{pmatrix} \right\} \quad \Longrightarrow \quad \bar{x} \text{ optimal.}$$

Proposition

$$
\boxed{\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & \\ & g_i(x) \le 0 \quad (i = 1, \dots, k) \end{array}}
$$

$g_1, \dots, g_k$ all convex

$\bar{x}$ is a feasible solution

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

$\forall i \in I,\ s^{(i)}$ subgradient for g_i at $\bar{x}$

If $-c \in \mathit{cone}\left\{s^{(i)} : i \in I\right\}$ then $\bar{x}$ is optimal.

Proof

We have a relaxation

$$
\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & \\ & g_i(x) \le 0 \quad (i \in I) \end{array}
$$

We proved that the set of solutions to $g_i(x) \le 0$

is contained in the set of solutions to $g_i(\bar{x}) + s^{(i)}(x - \bar{x}) \le 0.$

Proposition

$$\boxed{\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & \\ & g_i(x) \le 0 \quad (i = 1, \dots, k) \end{array}}$$

$g_1, \dots, g_k$ all convex

$\bar{x}$ is a feasible solution

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

$\forall i \in I,\ s^{(i)}$ subgradient for g_i at $\bar{x}$

If $-c \in \mathit{cone}\left\{s^{(i)} : i \in I\right\}$ then $\bar{x}$ is optimal.

Proof

We have a relaxation

$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & \\ & g_i(\bar{x}) + s^{(i)}(x - \bar{x}) \le 0 \quad (i \in I) \end{array}$$

$g_i(\bar{x}) + s^{(i)}(x - \bar{x}) \le 0$ can be rewritten as

$s^{(i)}x \le s^{(i)}\bar{x} - g_i(\bar{x})$

Proposition

$$\boxed{\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & \\ & g_i(x) \le 0 \quad (i = 1, \ldots, k) \end{array}}$$

$g_1, \ldots, g_k$ all convex
$\bar{x}$ is a feasible solution
$\forall i \in I$, $g_i(\bar{x}) = 0$
$\forall i \in I$, $s^{(i)}$ subgradient for g_i at $\bar{x}$

If $-c \in cone \left\{ s^{(i)} : i \in I \right\}$ then $\bar{x}$ is optimal.

Proof

We have a relaxation

$$\begin{array}{ll} \max & -c^\top x \\ \text{s.t.} & \\ & s^{(i)} x \le s^{(i)} \bar{x} - g_i(\bar{x}) \quad (i \in I) \end{array}$$

Then, $\bar{x}$ is optimal for the relaxation if $-c \in cone \left\{ s^{(i)} : i \in I \right\}$.
This means that $\bar{x}$ is also optimal for the NLP.

Question

Is there a converse to this result? YES

Gradients: A Calculus Detour

Proposition

Let $f : \Re^n \to \Re$ be a convex function and let $\bar{x} \in \Re^n$.

If the gradient $\nabla f(\bar{x})$ of f exists at $\bar{x}$, then it is a subgradient.

Proposition

Let $f : \Re^n \to \Re$ be function and let $\bar{x} \in \Re^n$.

If the partial derivative $\frac{\partial f(x)}{\partial x_j}$ exists for f at $\bar{x}$ for all $j = 1, \ldots, n$, then the gradient $\nabla f(\bar{x})$ is obtained by evaluating for $\bar{x}$,

$$\left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n}\right)^\top .$$

Example

Compute the gradient of the convex function

$$f(x) = -x_2 + x_1^2$$

at $\bar{x} = (1,1)^\top$.

We have

$$\left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}\right)^\top = (2x_1, -1)^\top$$

For $\bar{x}$ we get $\nabla f(\bar{x}) = (2,-1)^\top$.

Since $(2,-1)^\top$ is the gradient of f at $\bar{x}$, it is a subgradient as well.

Definition

A feasible solution to $\bar{x}$ is a Slater point of

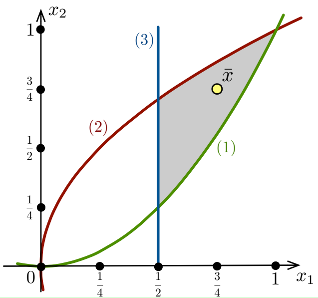
$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & \\ & g_i(x) \le 0 \quad (i = 1, \ldots, k) \end{array}$$

if $g_i(\bar{x}) < 0$ for all $i = 1, \ldots, k$.

Example

$$\begin{array}{lrcll} \min & -x_1 - x_2 & & & \\ \text{s.t.} & & & & \\ & -x_2 + x_1^2 & \le & 0 & (1) \\ & -x_1 + x_2^2 & \le & 0 & (2) \\ & -x_1 + \frac{1}{2} & \le & 0 & (3) \end{array}$$

$\bar{x} = \left(\frac{3}{4}, \frac{3}{4}\right)^\top$ is a Slater point.



The Karush-Kuhn-Tucker (KKT) Theorem

Consider the following NLP:

$$\boxed{\begin{array}{l} \min \quad c^\top x \\ \text{s.t.} \\ \qquad g_i(x) \leq 0 \quad (i = 1, \ldots, k) \end{array}}$$

Suppose that

1. $g_1, \ldots, g_k$ are all convex,
2. there exists a Slater point,
3. $\bar{x}$ is a feasible solution,
4. I is the set of indices i for which $g_i(\bar{x}) = 0$, and
5. for all $i \in I$ there exists a gradient $\nabla g_i(\bar{x})$ of g_i at $\bar{x}$.

Then $\bar{x}$ is optimal $\quad \Longleftrightarrow \quad -c \in cone\{\nabla g_i(\bar{x}) : i \in I\}$.

Remark

We proved the "easy" direction "$\Longleftarrow$".

Recap

- We showed how to prove optimality using relaxations.
- We defined subgradients.
- We defined supporting halfspaces.
- We related subgradients and supporting halfspaces.
- We showed how to relax convex constraints by a linear constraint.
- We gave sufficient conditions for a solution to be optimal.
- We stated the KKT theorem.