Convex Optimization
Lecture 6:
KKT Conditions,
and applications

Dr. Michel Baes,

IFOR / ETH Zürich

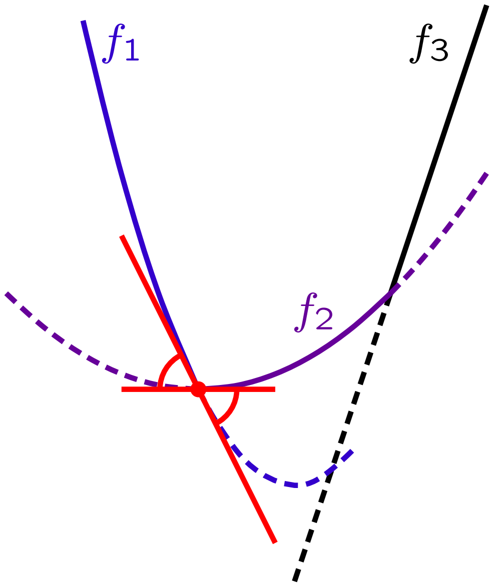
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Quick recall of last week's lecture

- Various aspects of convexity:
  The set of minimizers is convex.
  Convex functions are *line-differentiable*
  (i.e. the limit $\lim_{t \downarrow 0}[f(x+td)-f(x)]/t$ always exists).
  Differentiable convex functions:
  equivalent definitions, easier optimality conditions .
- Subdifferential: a generalization of gradient.
  New optimality conditions.
  Deducing differentiability by looking at $\partial f(x)$.
- Conjugate functions arise naturally from duality.
- $g \in \partial f(x)$ iff $x \in \partial f_*(g)$.
- An easy tool: support functions.
- Support function of subdifferentials.

Combining subdifferentials: Subdifferential of a maximum

Let $f_1, \ldots, f_m : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex,
such that $D := \cap_{i=1}^m \operatorname{int} \operatorname{dom}(f_i) \neq \phi$. Let $f(x) := \max_i f_i(x)$.



Let $I(x) := \{i : f_i(x) = f(x)\}$ for $x \in D$.

$\partial f(x) = C := \operatorname{conv}\{\partial f_i(x) : i \in I(x)\}$.

Proof: (see blackboard). **Key steps**:

- We just need to check $\sigma_C \equiv \sigma_{\partial f(x)}$ as $\partial f(x)$ and C are closed and convex.
- Let $d \in \mathbb{R}^n$. Then $\lim_{t \downarrow 0} I(x + td) \subseteq I(x)$.
- $\sigma_{\partial f(x)}(d) = \nabla f(x)[d] = \max_{i \in I(x)} \nabla f_i(x)[d]$.
- $\nabla f_i(x)[d] = \sigma_{\partial f_i(x)}(d) = \max\{\langle g_i, d \rangle : g_i \in \partial f_i(x)\}$.
- Remember the support function of a k-simplex.

Adapting it slightly, $\sigma_C(d) = \max_{i \in I(x)} \{\langle g_i, d \rangle : g_i \in \partial f_i(x)\}$.

Some examples

- Let $f(t) := |t| = \max\{t, -t\}$.
  Then $\partial f(t) = \text{sign}(t)$ for $t \neq 0$.
  Also, $\partial f(0) = \text{conv}\{-1, 1\} = [-1, 1]$.

- Let $f(x) := \max_{1 \le i \le n} x_i$, and $I(x) := \{i : x_i = f(x)\}$.
  Then $\partial f(x) = \text{conv}\{e_i : i \in I(x)\}$.
  In particular, $\partial f(0) = \Delta_n := \{g \ge 0 : \sum_i g_i = 1\}$.

  Observe that $g \in \partial f(0)$ iff $0 \in \partial f_*(g)$ iff g minimizes f_*.
  Now, f is the support function of Δ_n.
  Thus $f = \chi^*_{\Delta_n}$, and $f^* = \chi^{**}_{\Delta_n} = \chi_{\Delta_n}$,
  which is indeed minimized in Δ_n.

  Generalizable for every support function

Combining subdifferentials:

Subdifferential of a sum

Let $f_1, f_2 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex,
such that $D := \cap_i \text{relint}(\text{dom}(f_i)) \neq \phi$, and $s := f_1 + f_2$.
Then $\partial s(x) = \partial f_1(x) + \partial f_2(x)$ for all $x \in D$.

The proof, due to **Rockafellar**, is far to be trivial.
The direction $\supseteq$ is easy: if $g_i \in \partial f_i(x)$,

$$f_i(y) \geq f_i(x) + \langle g_i, y - x \rangle \quad \forall y, \text{ and } i = 1, 2.$$

Summing up both sides, we get that $g_1 + g_2 \in \partial s(x)$.

Sketch for $\subseteq$: We use $g \in \partial s^*(x)$ iff $s(x) + s^*(g) = \langle g, x \rangle$. It can be proven that $s^*(g) = \inf\{f_1^*(u) + f_2^*(v) : u + v = g\}$ when $D \neq \phi$. Now:

$$g \in \partial s^*(x) \;\Leftrightarrow\; \langle g, x \rangle = f_1(x) + f_2(x) + f_1^*(u^*) + f_2^*(v^*)$$

iff $u^* \in \partial f_1(x)$, $v^* \in \partial f_2(x)$, and $u^* + v^* = g$.

Subdifferential of a sum

The missing part*

The conjugate of a sum [Rockafellar, Th. 16.4]

Let $g_1, g_2 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex.

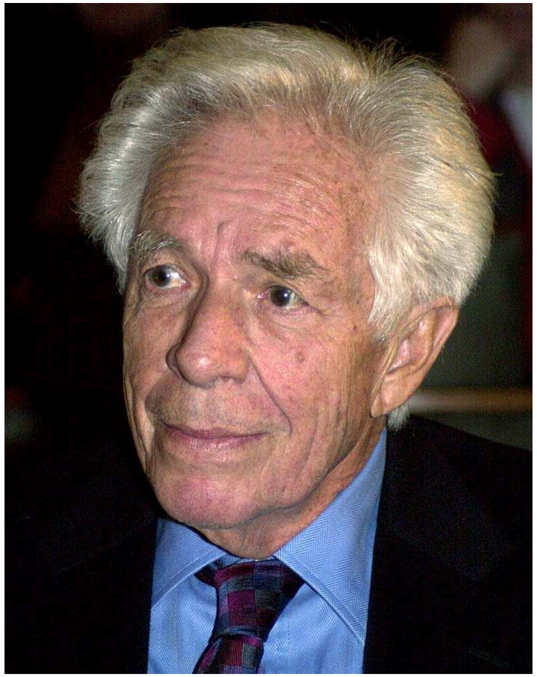
$$
\begin{aligned}
g_1^*(x) + g_2^*(x) &= \sup_{y,z} \langle y+z, x\rangle - g_1(y) - g_2(z) \\
&= \sup_{d} \sup_{y+z=d} \langle y+z, x\rangle - g_1(y) - g_2(z) \\
&= \sup_{d} \langle d, x\rangle - \inf_{y+z=d} g_1(y) + g_2(z) = \phi^*(x),
\end{aligned}
$$

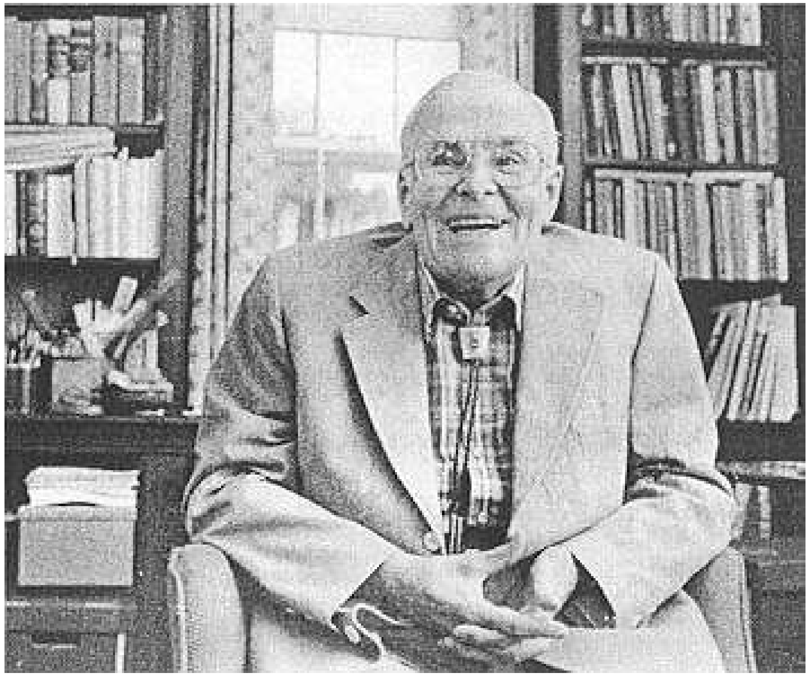
where $\phi(d) := \inf\{g_1(y) + g_2(z) : y + z = d\}$

is the *inf-convolution* of g_1 and g_2.

We let $g_1 := f_1^*$, $g_2 := f_2^*$. Since $(f_1^{**} + f_2^{**})^* = (f_1 + f_2)^*$ when $\cap_i \text{relint}(\text{dom}(f_i)) \neq \phi$, we get the needed result.

The Karush-Kuhn-Tucker Theorem





- The expression `Kuhn-Tucker` has 185,000 hits on Google.
- Needless to say, it is a cornerstone of Optimization.
- Proved in 1939 in the Master Thesis of Karush, rediscovered in 1951 by Kuhn and Tucker.

The Karush-Kuhn-Tucker Theorem

Theorem 1 (KKT Conditions for Convex Optimization) *Let* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be a convex function,*

$g_1, \ldots, g_m$ *be concave functions,*

$b \in \mathbb{R}^m$ *such that* Slater's condition *holds:*

$$\exists \bar{x} : g_i(\bar{x}) > b_i \text{ for } 1 \leq i \leq m.$$

A point x^* *is a solution to* $f^* = \min\{f(x) : g(x) \geq b\}$

iff $g(x^*) \geq b$, *(Feasibility)*

$\exists h_0 \in \partial f(x^*)$, $h_i \in \partial(-g_i(x^*))$, *("Original"*

$\lambda_i^* \geq 0$ *for* $1 \geq i \geq m$*:* *KKT*

$h_0 + \sum_{i \in I(x^*)} \lambda_i^* h_i = 0$, *Conditions)*

where $I(x^*) := \{i : g_i(x^*) = b_i\}$.

Note: The minus sign ensures that $\partial(-g_i(x^*)) \neq \phi$.

The Karush-Kuhn-Tucker Theorem: the proof is simple with subdifferentials

$$\boxed{f^* = \min\{f(x) : g(x) \geq b\}} \quad (\mathcal{P})$$

- Let $\phi(x) := \max\{f(x)-f^*, b_1-g_1(x), \ldots, b_m-g_m(x)\}$, which is convex.
- x^* is an optimum of $(\mathcal{P})$ iff $x^* \in \arg\min_x \phi(x)$ iff $0 \in \partial\phi(x^*)$
  iff $0 \in \text{conv}\{\partial f(x^*), \partial(-g_i(x^*)) : i \in I(x^*)\}$ (obviously $f(x^*) = f^*$)
  iff $\exists h_0 \in \partial f(x^*), h_i \in \partial(-g_i(x^*))$, $\alpha_i \geq 0$, $\alpha_0 + \sum_{i \in I(x^*)} \alpha_i = 1$
  such that $0 = \alpha_0 h_0 + \sum_{i \in I(x^*)} \alpha_i h_i$.
- $\alpha_0 \neq 0$.
  First, $\langle h_i, y - x^* \rangle \leq g_i(x^*) - g_i(y) = b_i - g_i(y)$ for all y and all $i \in I(x^*)$.
  If $\alpha_0 = 0$, then $0 = \sum_{i \in I(x^*)} \alpha_i \langle h_i, \bar{x} - x^* \rangle \leq \sum_{i \in I(x^*)} \alpha_i (b_i - g_i(\bar{x}))$,
  contradicting Slater's condition, satisfied by $\bar{x}$.
- It remains to let $\lambda_i^* := \alpha_i / \alpha_0$.

This theorem cannot be used!

You need to know $I(x^*)$ in advance!
Easy way out: set $\lambda_i^* := 0$ when $i \notin I(x^*)$.

Theorem 2 (KKT Conditions for Convex Optimization II)
Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be a convex function,*
$g_1, \ldots, g_m$ *be concave functions,*
$b \in \mathbb{R}^m$ *such that* Slater's condition *holds:*

$$\exists \bar{x} : g_i(\bar{x}) > b_i \text{ for } 1 \leq i \leq m.$$

A point x^* *is a solution to* $f^* = \min\{f(x) : g(x) \geq b\}$
iff $g(x^*) \geq b$, *(Feasibility)*
$\exists h_0 \in \partial f(x^*)$, $h_i \in \partial(-g_i(x^*))$, *("Usable"*
$\lambda_i^* \geq 0$ *for* $1 \geq i \geq m$*:* *KKT*
$h_0 + \sum_{i=1}^m \lambda_i^* h_i = 0$, *Conditions)*
and $\lambda_i^*(b_i - g_i(x^*)) = 0$ *for all* i. *(Complementarity)*

When you have a slightly different problem

- **Equality constraints** (necessary **affine** constraints): the same statement holds, but no sign constraint for the corresponding λ_i^*'s, and an extra condition on linear independence of the h_i's.
- A version of the KKT Theorem exists for differentiable non-convex problems. The conditions read the same but **are not sufficient**.
  First find all the *KKT points* (x^*, λ^*),
  then test them all to find the global optimum.
- **Interesting exercise:**
  what happens for general conic inequalities?

KKT and duality

- λ_i^* **is the dual optimum**. Recall:

Theorem 3 (Complementarity conditions) *Suppose that x^* and F^* are feasible for their respective problems, and that $f(x^*) = F^*(b)$. Then*

$$p^* = f(x^*) = F^*(g(x^*)) = F^*(b) = d^*(\mathcal{F}).$$

We take as candidates x^* and $F^*(y) = \langle u, y \rangle + u_0$, with $u := \lambda^*$ and $u_0 := f(x^*) - \langle \lambda^*, b \rangle$.

1. By direct substitution, $F^*(b) = f(x^*)$.

2. F^* is feasible, that is $F^*(g(x)) \le f(x)$ for all x. Fix $x \in \mathbb{R}^n$

First, $f(x^*) \le f(x) - \langle h_0, x - x^* \rangle = f(x) + \sum_{i \in I(x^*)} \lambda_i^* \langle h_i, x - x^* \rangle$

$\le f(x) + \sum_{i \in I(x^*)} \lambda_i^* (g_i(x^*) - g_i(x)) = f(x) + \sum_{i \in I(x^*)} \lambda_i^* (b_i - g_i(x))$,

which is equivalent to $F^*(g(x)) \le f(x)$.

Thus λ^* is the dual optimum,

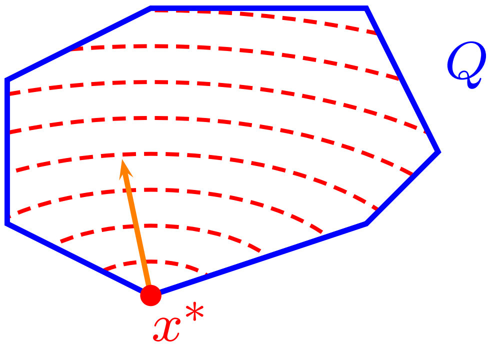
and can be interpreted as the constraints prices.

- The KKT Conditions are nothing but $\partial L(x^*, \lambda^*)/\partial x = 0$

A geometric view of KKT

- For unconstrained problems, we recover the optimality condition $0 \in \partial f(x^*)$.
- When the f is differentiable, and $Q := \{x : g(x) \geq b\}$ has a nonempty interior, we have $x^* \in \arg\min\{f(x) : x \in Q\}$ iff

$$\langle f'(x^*), y - x^* \rangle \geq 0 \quad \forall y \in Q.$$



KKT says $f'(x^*) = -\sum_{i \in I(x^*)} \lambda_i^* h_i$, with $\langle h_i, y - x^* \rangle \leq g_i(x^*) - g_i(y) = b_i - g_i(y)$ and $\lambda_i^* \geq 0$ for $i \in I(x^*)$. Thus:

$$\langle f'(x^*), y - x^* \rangle = -\sum_{i \in I(x^*)} \lambda_i^* \langle h_i, y - x^* \rangle$$
$$\geq -\sum_{i \in I(x^*)} \lambda_i^* (b_i - g_i(y)) \geq 0$$

for all feasible y.

Application

Projecting on a subspace

- One of the **most solved** optimization problems in the world. (Also known as *Least-Squares Problem*)
- Direct applications in meteorology, genomic, statistics, control, signal processing, ...

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, with $n \geq m$.
Find the shortest solution of $Ax = b$:

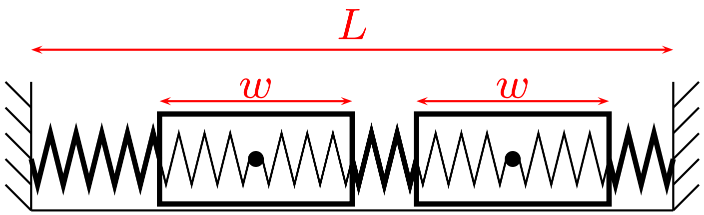
$$\min\{||x||_2^2/2 : Ax = b\}$$

KKT conditions: $Ax^* = b$, $x^* - A^T\lambda^* = 0$
imply $AA^T\lambda^* = b$, and $x^* = A^T(AA^T)^{-1}b$
$A^\dagger := A^T(AA^T)^{-1}$ is the *Moore-Penrose inverse* of A.

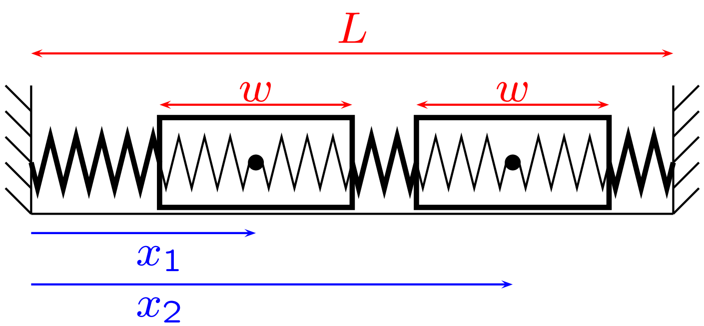
A historical application:
A simple mechanical system

We have on a straight segment between two walls:
two masses each of width w;
three springs of very short length at rest (~ 0)
attached between the walls and the center of the masses,
of rigidity k_1, k_2 k_3 respectively.

What is the equilibrium configuration?
What are the forces on the walls?



A historical application: Modeling as an optimization problem



Potential energy of a spring: rigidity × length^2/2.
$||$**Force**$||$ **exerted by a spring:** rigidity × length.

$$\begin{array}{ll} \min & \frac{1}{2}\left(k_1 x_1^2 + k_2(x_2 - x_1)^2 + k_3(L - x_2)^2\right) \\ \text{s.t.} & x_1 \geq w/2 \\ & x_2 - x_1 \geq w \\ & L - x_2 \geq w/2. \end{array}$$

A historical application: The optimality conditions

$$\begin{array}{ll} \min & \frac{1}{2}\left(k_1 x_1^2 + k_2(x_2 - x_1)^2 + k_3(L - x_2)^2\right) \\ \text{s.t.} & x_1 \geq w/2 \\ & x_2 - x_1 \geq w \\ & L - x_2 \geq w/2. \end{array}$$

Complementarity and KKT Conditions:

$$\lambda_1^*(x_1^* - w/2) = 0, \quad \lambda_2^*(x_2^* - x_1^* - w) = 0, \quad \lambda_3^*(L - x_2^* - w/2) = 0,$$

$$k_1 x_1^* - k_2(x_2^* - x_1^*) - \lambda_1^* + \lambda_2^* = 0,$$

$$k_2(x_2^* - x_1^*) - k_3(L - x_2^*) - \lambda_2^* + \lambda_3^* = 0,$$

$$\lambda_i^* \geq 0, \quad x^* \text{ feasible.}$$

A historical application:
The physical interpretation of dual variables

Complementarity and KKT Conditions:

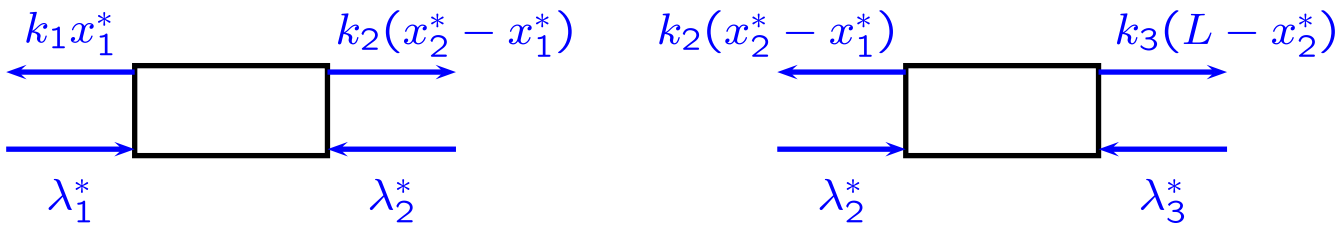
$$\lambda_1^*(x_1^* - w/2) = 0, \quad \lambda_2^*(x_2^* - x_1^* - w) = 0, \quad \lambda_3^*(L - x_2^* - w/2) = 0,$$

$$k_1 x_1^* - k_2(x_2^* - x_1^*) - \lambda_1^* + \lambda_2^* = 0,$$

$$k_2(x_2^* - x_1^*) - k_3(L - x_2^*) - \lambda_2^* + \lambda_3^* = 0,$$

$$\lambda_i^* \geq 0, \quad x^* \text{ feasible.}$$

The KKT Conditions can be interpreted as a force balance equation on both masses.



λ_1^* [λ_3^*] is the force exerted on the left [right] wall
λ_2^* is the force exerted on each block

Applications of KKT's Theorem

are countless

I am sure that each of you

will have to use them some day

(If you stay in engineering)

For next week

Making convex optimization work for you:
Modeling and solving Linear, Second-Order,
and Semidefinite optimization problems.