Convex Optimization in Machine Learning and Computational Finance
Lecture 5: Subgradients, Conjugate functions

Dr. Michel Baes, Pr. Patrick Cheridito

RiskLab / ETH Zürich

Quick recall of last week's lecture

- Continuity and differentiability:
  every convex function f is continuous on int dom f
- Necessary/sufficient condition for optimality
  in **unconstrained** optimization.
- Lagrangian duality: strong duality works
  with affine nondecreasing functions
  (whenever inf-sup = sup-inf holds).

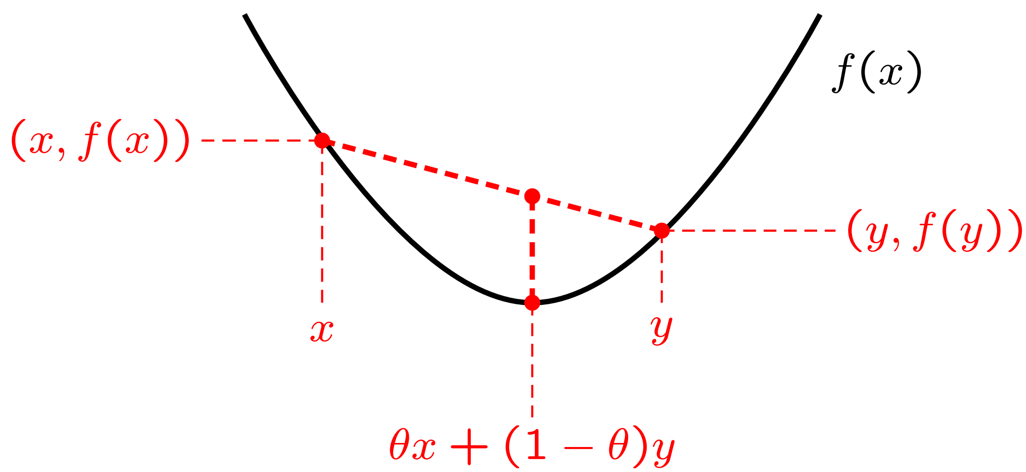
  Examples: convex problems where:

  1. one of the two feasible sets is compact;

  2. or Slater condition holds.
- Conic duality

A summary of everything we know about convex functions 1/2

- A convex function f has a convex epigraph.
- Equivalently, $\forall x, y \in \operatorname{dom} f$, $\lambda \in [0, 1]$:
  $f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y)$.
- Let $(f_\alpha)_{\alpha \in A}$ be convex functions $\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$.
  Then $f = \sup_{\alpha \in A} f_\alpha$ is convex.



Convexity and optimality

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a convex function.
Let $Q \subseteq \mathbb{R}^n$ be a convex set. Let $X^* := \arg\min_{x \in Q} f(x)$.

- X^* is convex: if $x, y \in X^*$, $\lambda \in [0, 1]$, $f^* := \min_{x \in Q} f(x)$, we have $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) = f^*$.
- More generally, *level sets* are convex. If $\mathcal{L}_\alpha(f) := \{x \in Q : f(x) \leq \alpha\}$, $\mathcal{L}_\alpha(f)$ is convex.

Solving a convex optimization problem approximately is finding a point in a convex set – namely $\mathcal{L}_{f^*+\epsilon}(f)$.

A summary of everything we know about convex functions 2/2

- Every convex function f is continuous on $\operatorname{int}\operatorname{dom} f$.
- A closed convex set is the intersection of all the half-spaces containing it.

  For epigraphs of convex LSC functions f, it means:

  $f(x) = \sup\{h(x) : f \geq h, h \text{ affine}\}$.

Note 1: There are non-LSC convex functions:

$$f : \mathbb{R}_+ \to \mathbb{R}, \quad f(x) = 0 \text{ for } x > 0, \quad f(0) = 1.$$

Note 2: There are non-continuous LSC convex functions:

$$f : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}, \quad \begin{aligned} f(x,y) &= y^2/x \text{ for } x \neq 0, \\ f(0,y) &= +\infty \text{ for } y \neq 0, \\ f(0,0) &= 0 \end{aligned}$$

(cfr. Lecture 2: $\operatorname{epi} f \equiv \mathbb{S}^2_+$)

Convex functions are almost differentiable

Theorem 1 *Let* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be convex,* $x \in \text{relint}(\text{dom} f)$*, and* $d \in \text{aff}(\text{dom} f)$*. Then*

$$\nabla f(x)[d] := \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} \text{ exists.}$$

Warning: *$\nabla f(x)[d]$ might not be linear in d.*

e.g. if $f(x) = |x| \in \mathbb{R}$, $\nabla f(0)[d] = |d|$.

Proof: Let $\mu(t) := (f(x+td) - f(x))/t$ for all $t \in \mathbb{R}_{++}$.
The function μ is increasing, as it can be easily checked by using convexity of f. Therefore, $\mu(t)$ is bounded from below for $t > 0$.
Since $\mu(t)$ decreases when $t \downarrow 0$, the limit $\lim_{t \downarrow 0} \mu(t)$ exists.

Differentiable convex functions

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a differentiable convex function. For all $x \in \operatorname{relint}(\operatorname{dom} f)$, $y \in \operatorname{dom} f$:

$$\begin{aligned}\langle f'(x), y - x\rangle &= \lim_{t\to 0} \frac{f(x + t(y - x)) - f(x)}{t} \\ &\le \lim_{t\to 0} \frac{(1 - t)f(x) + tf(y) - f(x)}{t} = f(y) - f(x).\end{aligned}$$

- $f(y) \ge f(x) + \langle f'(x), y - x\rangle$.
- **Increasing gradient property**:

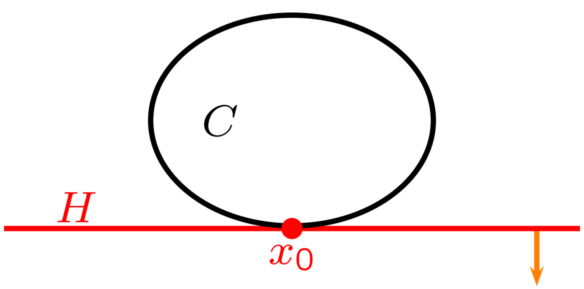
  $\langle f'(y) - f'(x), y - x\rangle \ge 0$.
- If f twice differentiable, $f''(x) \succeq 0 \ \forall x \in \operatorname{int} \operatorname{dom} f$.

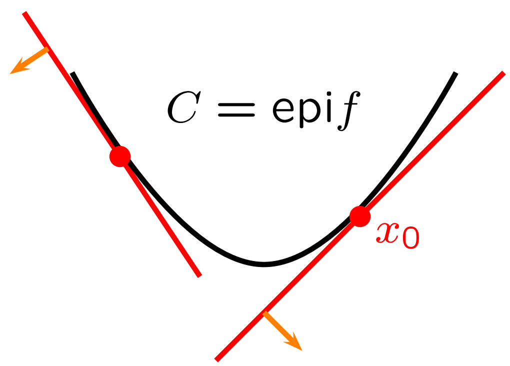
  Indeed, $0 \le \frac{1}{t^2}\langle f'(x + th) - f'(x), th\rangle \to \langle f''(x)h, h\rangle$ as $t \to 0$.

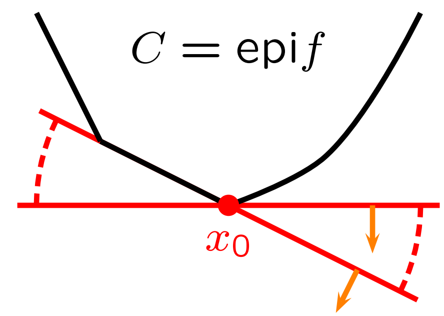
Thus $f'(x^*) = 0$ is sufficient to guarantee optimality!

Supporting hyperplanes

- A *supporting hyperplane* $H = \{x : l(x) = \beta\}$, l linear, of a convex set C at a point $x_0 \in \text{bd}C$, is such that $x_0 \in H$ and $C \subseteq \{x : l(x) \leq \beta\}$.
- We only consider here non-trivial supporting hyperplanes, that is $C \nsubseteq H$.







Supporting hyperplanes of convex epigraphs

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be an LSC convex function.
Let $x_0 \in \text{relint}(\text{dom} f)$. As $(x_0, f(x_0)) \in \text{bd}(\text{epi} f)$,
we can consider all the supporting hyperplanes of epi$f$.
They are not vertical. Let H be one of them:

$$\begin{aligned} H &= \left\{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \langle \begin{pmatrix} \hat{g} \\ -1 \end{pmatrix}, \begin{pmatrix} x \\ t \end{pmatrix} \rangle = \beta \right\} \\ &= \{(x,t) : g(x) - t = \beta\}, \text{ where } g \text{ linear.} \end{aligned}$$

As $(x_0, f(x_0)) \in H$, $\beta = g(x_0) - f(x_0)$.
As $(x, f(x)) \in \text{epi} f$ for all $x \in \text{dom} f$, we have:

$$g(x) - f(x) \leq \beta = g(x_0) - f(x_0) \Leftrightarrow f(x_0) + g(x - x_0) \leq f(x).$$

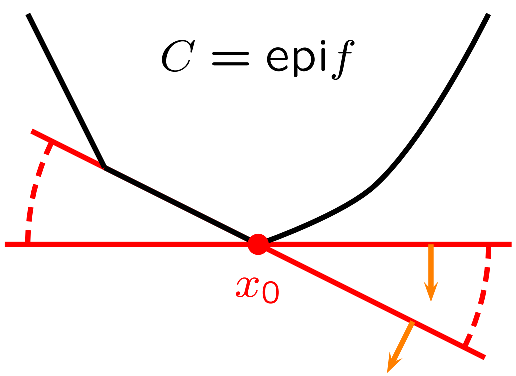
Note: If f is differentiable at x_0, $g := \nabla f(x_0)$ works.

Subgradients of a convex function

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be an LSC convex function.
Let $x_0 \in \text{relint}(\text{dom} f)$.
g completely describes a supp. hyperplane of epi$f$ at x_0.

$$\partial f(x_0) := \{g \text{ linear} : f(x) \geq f(x_0) + g(x - x_0) \forall x \in \mathbb{R}^n\}$$

is the *subdifferential* (= set of *subgradients*) of f at x_0.



Note: We can also see $\partial f(x_0)$ as a set of **points** $\in \mathbb{R}^n$, using a scalar product $\langle \cdot, \cdot \rangle$ of $\mathbb{R}^n$: By Riesz's Theorem, $\exists \hat{g} \in \mathbb{R}^n$: $\langle \hat{g}, x \rangle = g(x)$ for all linear $g : \mathbb{R}^n \to \mathbb{R}$:

$$\partial f(x_0) = \{\hat{g} \in \mathbb{R}^n : f(x) \geq f(x_0) + \langle \hat{g}, x - x_0 \rangle \forall x \in \mathbb{R}^n\}.$$

The subdifferential is rarely empty:

If f is convex, $x \in \text{relint}(\text{dom} f) \Rightarrow \partial f(x) \neq \phi$

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be an LSC convex function,
let $x \in \text{relint}(\text{dom} f)$.
Recall the **affine separation theorem**:

> *Let $C \subseteq \mathbb{R}^N$ convex set with* $\text{relint}(C) = C \neq \phi$
> *and D affine $C \cap D = \phi$. There exists a hyperplane H*
> *of dimension $N - 1$ such that $D \subseteq H$, and $C \subseteq \text{int}\, H^+$.*

Here $C := \text{relint}(\text{epi} f)$, $D = \{(x, f(x))\}$.
Let $H := \{y \in \mathbb{R}^n \times \mathbb{R} : l(y) = c\}$.
As $C \subseteq \{y \in \mathbb{R}^n \times \mathbb{R} : l(y) < c\}$,
epi$f \subseteq \text{cl}(C) \subseteq \{y \in \mathbb{R}^n \times \mathbb{R} : l(y) \leq c\}$.
H is thus a supporting hyperplane of epi$f$ at x.
It is also not vertical, because $x \in \text{relint}(\text{dom} f)$.

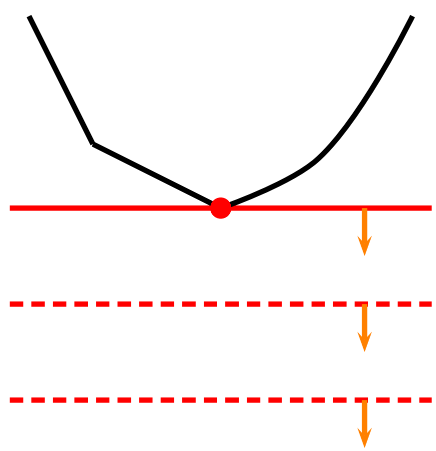
Subdifferential and optimality

Theorem 2 *Let* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be an LSC convex function.* $x^* \in \arg\min_{x \in \mathbb{R}^n} f(x)$ **iff** $0 \in \partial f(x^*)$.

Proof: We have:

$$f(y) \geq f(x^*) + \langle 0, y - x^* \rangle = f(x^*)$$

for every $y \in \mathbb{R}^n$ iff $0 \in \partial f(x^*)$.



Duality and differentiability are closely related concepts

Conjugate functions

A very useful related concept: conjugate functions

$$\begin{array}{ll} \inf & f(x) \\ \text{s.t.} & x \succeq_K b \\ & x \in \mathrm{dom}(f) \end{array} \geq \begin{array}{ll} \sup & F(b) \\ \text{s.t.} & F(x) \leq f(x) \ \forall x \in X \\ & F \in \mathcal{F} \end{array}$$

(Here $X = \mathrm{dom}(f)$. And $g \equiv \mathrm{Id}$)

Recall that $\mathcal{F} := \{x \mapsto \langle u, x \rangle - u_0 : u \in K^*\}$
almost ensures strong duality
(check convexity and Slater's conditions).
The dual becomes:

$$\begin{array}{ll} \sup & \langle u, b \rangle - u_0 \\ \text{s.t.} & \langle u, x \rangle - u_0 \leq f(x) \ \forall x \in \mathrm{dom}(f) \\ & u \in K^*, u_0 \text{ free} \end{array}$$

A very useful related concept: conjugate functions

$$\begin{array}{ll} \sup & \langle u, b\rangle - u_0 \\ \text{s.t.} & \langle u, x\rangle - u_0 \le f(x)\ \forall x \in \text{dom}(f) \\ & u \in K^*, u_0 \text{ free} \end{array}$$

Looking at the constraints: $\langle u, x\rangle - u_0 \le f(x)\ \forall x \in \text{dom}(f)$
$\Leftrightarrow\ \langle u, x\rangle - f(x) \le u_0\ \forall x \in \text{dom}(f)$

The function $f_*(u) := \sup_{x\in\mathbb{R}^n} \langle u, x\rangle - f(x)$
is the *conjugate of f*. Note that f_* is LSC and convex.

$$d^* = \sup_{u\in K^*} \quad \langle u, b\rangle - f_*(u) \le \inf_{x \succeq_K b} \quad f(x) \le f(b)$$

Conjugate functions tell you how to package portfolios optimally

We are selling portfolios of d assets, priced $p_1, \ldots, p_d$.
Any random profit/loss goes to the client, but
he must pay a commission for our portfolio management
Let $\xi \in \mathbb{R}^d$ be the portfolio composition
and $f(\xi)$ the commission for managing it.
To minimize our cost:

$$\min_{\xi} p^T \xi - f(\xi) = g(p),$$

where $-g(-p) = (-f)_*(p)$.

How do we determine the optimal ξ^*?

Conjugates of conjugates

By definition, $f_*(u) = \sup_{x\in\mathbb{R}^n} \langle u, x\rangle - f(x)$

Let $f_{**}(b) := \sup_{u\in\mathbb{R}^n} \langle u, b\rangle - f_*(u)$

$$\boxed{\sup_{u\in\mathbb{R}^m} \quad \langle u, b\rangle - f_*(u)} \leq \boxed{\inf_{x\,=\,b} \quad f(x)} = f(b)$$

Theorem 3 *If f is LSC convex, then $f_{**} = f$.*

Proof: When $b \in \operatorname{int}\operatorname{dom}(f)$, Slater's condition holds, yielding strong duality: $f_{**}(b) = f(b)$.
Thus $\operatorname{int}\operatorname{epi}(f_{**}) = \operatorname{int}\operatorname{epi}(f)$. As f_{**} and f are LSC, we get $\operatorname{epi} f_{**} = \operatorname{cl}(\operatorname{int}\operatorname{epi}(f_{**})) = \operatorname{cl}(\operatorname{int}\operatorname{epi}(f)) = \operatorname{epi} f$, and $f_{**} = f$.

Support functions:
the simplest example of conjugate functions

Let $X \subseteq \mathbb{R}^n$ be closed convex. Suppose you have to solve

$$\begin{array}{rl} \text{sup} & \langle u, x \rangle \\ \text{s.t.} & x \in X \end{array}$$

for many u's. The *support function* of X is

$$\sigma_X(u) = \begin{array}{rl} \text{sup} & \langle u, x \rangle \\ \text{s.t.} & x \in X. \end{array}$$

It is the conjugate of the characteristic function of X:

$$\begin{cases} \chi_X(x) = 0, & x \in X \\ \chi_X(x) = +\infty, & x \notin X. \end{cases}$$

Massively used when easily computable.

Some support functions

- **Point**: $X = \{x\}$. Then $\sigma_X(u) = \langle u, x \rangle \ \forall u \in \mathbb{R}^n$.
- **Unit ball**: $X = \{x : ||x|| \leq 1\}$.
  $\sigma_X(u) = \sup_{||x|| \leq 1} \langle u, x \rangle =: ||u||_* \ \forall u \in \mathbb{R}^n$.
- k-**simplex**: $X = \text{conv}\{v_0, \ldots, v_k\}$.
  $\sigma_X(u) = \sup\{\langle u, \sum_i t_i v_i \rangle : t_i \geq 0, \sum_i t_i = 1\} = \max_i \langle u, v_i \rangle$.
- **Convex cone**:
  $\sigma_K(u)$ is the characteristic function of $-K^*$.
- **Half-space**: $X = \{x : \langle a, x \rangle \leq b\}$.
  $\sigma_X(\gamma a) = \gamma b$ for $\gamma \geq 0$, and $\sigma_X(u) = +\infty$ elsewhere.

Support function of subdifferentials

Theorem 4 *Let* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *be LSC convex. Let* $x \in \text{relint}(\text{dom} f)$*. The support function of* $\partial f(x)$ *is* $h : \text{aff}(\text{dom} f) \to \mathbb{R}$, $d \mapsto \nabla f(x)[d]$.

Proof: We fix $x \in \text{relint}(\text{dom} f)$ and $d \in \text{aff}(\text{dom} f)$. We need to check $h(d) = \sup\{\langle g, d\rangle : g \in \partial f(x)\}$. First, $h(d) \geq \sup\{\langle g, d\rangle : g \in \partial f(x)\}$ as:

$$\forall g \in \partial f(x), \quad h(d) = \lim_{t \downarrow 0} \frac{f(x + td) - f(x)}{t} \geq \lim_{t \downarrow 0} \frac{\langle g, td\rangle}{t} = \langle g, d\rangle.$$

Second, the function h is convex on $\mathbb{R}^n$, by convexity of f.
Thus $\partial h(d) \neq \phi$. Let $g_h \in \partial h(d)$ and $v \in \mathbb{R}^n$. Note that for all $\tau > 0$:

$$\tau h(v) = h(\tau v) \geq h(d) + \langle g_h, \tau v - d\rangle \Rightarrow h(v) \geq \langle g_h, v\rangle = h(0) + \langle g_h, v\rangle,$$

by dividing by τ and letting $\tau \to +\infty$. Hence $g_h \in \partial h(0)$. Also, $\langle g_h, d\rangle \geq h(d)$ by taking $\tau \to 0$. Since $\mu(t) := (f(x + td) - f(x))/t$ is decreasing on $\mathbb{R}_{++}$ (easy to check from the convexity of f) and $g_h \in \partial h(0)$:

$$f(y) - f(x) \geq \lim_{t \downarrow 0} \frac{f(x + t(y - x)) - f(x)}{t} = h(y - x) \geq \langle g_h, y - x\rangle \quad \forall y \in \text{dom} f.$$

Therefore $g_h \in \partial f(x)$ and $\sup\{\langle g, d\rangle : g \in \partial f(x)\} \geq \langle g_h, d\rangle \geq h(d)$. ■

Subdifferential and differentiability

- f is differentiable at x **iff** $\partial f(x) = \{g\}$.
  In that case, $g = \nabla f(x)$.

Proof: Assume that $\partial f(x) = \{g\}$.
The support function of $\partial f(x)$ is thus $\langle g, d\rangle = \nabla f(x)[d]$.
As it is a linear function, f is differentiable at x.
Moreover, $g = f'(x)$.
Conversely, if f is differentiable, $\nabla f(x)[d]$ is linear,
i.e. the support function of a single point.
Thus, $\partial f(x)$ is a singleton. ■

There is a deep link between conjugates and subdifferentials

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be an LSC convex function.
Let $x \in \text{relint}(\text{dom} f)$.

$$\partial f(x) = \{g : f(y) \geq f(x) + \langle g, y - x \rangle \; \forall y \in \mathbb{R}^n\}$$

$$g \in \partial f(x) \;\Leftrightarrow\; \langle g, x \rangle - f(x) \geq \sup_y \langle g, y \rangle - f(y) = f_*(g)$$

But $f_*(g) \geq \langle g, x \rangle - f(x)$; thus

$$g \in \partial f(x) \;\Leftrightarrow\; f_*(g) + f(x) = \langle g, x \rangle.$$

By $f_{**} = f$, you also get:

$$x \in \partial f_*(g) \;\Leftrightarrow\; f_*(g) + f(x) = \langle g, x \rangle.$$

$$\boxed{g \in \partial f(x) \;\Leftrightarrow\; x \in \partial f_*(g)}$$

For next week

- A crucial consequence of the subgradient theory: Optimality conditions for constrained problems (KKT conditions)
- Applications of duality theory and KKT conditions in finance and engineering.