Convex Analysis for Optimization

Olga Kuryatnikova (Erasmus University Rotterdam)
kuryatnikova@ese.eur.nl

September-October 2024
Lecture 4

Course plan

- Week 1: Introduction to convexity
- Week 2: More on convex sets
- Week 3: Dual view of convex sets + more on convex functions
- Week 4: Dual view of convex functions
- Week 5: Duality and optimization
- Week 6: Introduction to algorithms, descend methods
- Week 7: Proximal methods, projected gradients
- Weeks 8 - 9: Fix point approach, averaged operators

Dual view on convex functions

- Continuity and closedness
- Differentiability and subgradients
- Conjugate functions
- Prox operators

Types of continuity

Let $S \subseteq \mathbb{R}^n$, consider a function $f{:}S \to \overline{\mathbb{R}}^m$ for some $m \geq 1$.

Def: f is lower semicontinuous in x if $f(x) \leq \liminf_{y \to x} f(y), \forall (y) \subset S$.

Def: f is continuous in $x \in \mathrm{dom}(f)$ if $f(x) = \lim_{y \to x} f(y), \forall (y) \subset \mathrm{dom}(f)$

Def: f is Lipshitz-continuous with constant $L > 0$ if
$\|f(x) - f(y)\|_2 \leq L\|x - y\|_2$ for all $x, y \in \mathrm{dom}(f)$

Semicontinuity and closedness

Def: $f : S \to \overline{\mathbb{R}}$ is closed if its epigraph $\mathrm{epi}(f)$ is a closed set.

Thm: Function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is closed if and only if

$\iff$ f is lower-semicontinuous

$\iff$ level set $V_\gamma =: \{x \in \mathbb{R}^n : \gamma \geq f(x)\}$ is closed for any $\gamma \in \mathbb{R}$

Continuity and convexity

Thm: $f : S \to \overline{\mathbb{R}}$ proper and convex $\Rightarrow$ f continuous over $\mathrm{ri}(\mathrm{dom}(f))$.

Corollary: A convex function $\mathbb{R}^n \to \mathbb{R}$ is continuous.

Lipschitz continuity and fixed points

Def: $f : \mathbb{R}^n \to \mathbb{R}^n$ is a non-expansive mapping if it is Lipschitz continuous with constant $L \leq 1$.

- If also $\|f(x) - f(y)\|_2^2 \leq (f(x) - f(y))^\top (x - y)$ for all $x, y \in \mathrm{dom}(f)$, f is called firmly non-expansive.
- If $L < 1$, f is called a contraction.

Def: x is a fixed point of function $f : \mathbb{R}^n \to \mathbb{R}^n$ if $f(x) = x$.

Banach Fixed Point Thm: Let f be a contraction. Then f admits a unique fixed-point, and an algorithm starting from some x_0 and computing $x_k = f(x_{k-1})$ for $k = 1, \ldots$ converges to that fixed point.

Extension to firmly-non-expansive: $x_k = f(x_{k-1})$ for $k = 1, \ldots$ converges to a fixed point if it exists.

Differentiable functions

Def: $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is differentialble in $\bar{x} \in \mathrm{dom}(f)$ if $\lim\limits_{x \to \bar{x}} \frac{|f(x) - f(\bar{x}) - \nabla f(\bar{x})^\top (x - \bar{x})|}{\|x - \bar{x}\|} = 0$ for all sequences $\{x\}$ converging to $\bar{x}$.

Gradient $\nabla f(\bar{x}) := \left[\frac{\partial f}{\partial x_1}(\bar{x}), \ldots, \frac{\partial f}{\partial x_n}(\bar{x})\right]$ and directional derivative $\nabla_v f(\bar{x}) := \lim\limits_{\alpha \downarrow 0} \frac{f(\bar{x} + \alpha v) - f(\bar{x})}{\alpha} = \nabla f(\bar{x})^\top v$ exist in $\bar{x}$ for all $v \in \mathbb{R}^n$.

Convex differentiable functions and optimization

Thm: Let $S \subseteq \mathbb{R}^n$ be convex, f be differentiable over an open set that contains S. Then f is convex over S if and only if

$$f(z) - f(x) \geq \nabla f(x)^\top (z - x) \quad \forall x, z \in S.$$

Corollary: for S and f as above,

- $\nabla f(x^*) = 0 \implies x^*$ minimizes f over $\mathbb{R}^n$;
- x^* minimizes f over S if and only if $\nabla f(x^*)^\top (z - x^*) \geq 0 \quad \forall z \in S$.

Convex twice differentiable functions

Thm: Let $S \subseteq \mathbb{R}^n$ be convex and open, f be twice continuously differentiable over S. Then f is convex over S if and only if

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in S.$$

Subgradient and subdifferential

Def: $g \in \mathbb{R}^n$ is a subgradient of a convex $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ in $\bar{x} \in \operatorname{dom}(f)$ if $f(z) - f(\bar{x}) \geq g^\top (z - \bar{x}) \quad \forall z \in \mathbb{R}^n$.

Def: subdifferential $\partial f(\bar{x})$ is the set of all subgradients of f in $\bar{x}$:

$$\partial f(\bar{x}) := \{g \in \mathbb{R}^n : f(z) - f(\bar{x}) \geq g^\top (z - \bar{x}) \quad \forall z \in \mathbb{R}^n\}.$$

Properties of subdifferential

- $\partial f(\bar{x})$ is closed and convex as an intersection of closed subspaces.
- If f is differentiable in $\bar{x}$, then $\partial f(\bar{x}) = \nabla f(x)$.
- For $S \subseteq \mathbb{R}^n$, $N_S(\bar{x}) = \partial \delta_S(\bar{x})$, where $\delta_S(x) = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{otherwise.} \end{cases}$
- Let f be convex, $A \in \mathbb{R}^{n \times m}$, and $F(x) = f(Ax)$. If f is polyhedral or $\exists \alpha \in \mathbb{R}^m : A\alpha \in \mathrm{ri}(\mathrm{dom}(f))$, then $\partial F(x) = A^\top \partial f(Ax)$.
- Let f, h be convex and $F = f + h$ be proper. If $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(\mathrm{dom}(h)) \neq \emptyset$, then $\partial F(x) = \partial f(x) + \partial h(x)$.
- If $\emptyset \neq S \subseteq \mathrm{dom}(f)$ is compact, f is convex, then $\bigcup_{x \in S} \partial f(x) \neq \emptyset$ and bounded; and f is Lipschitz continuous on S with constant $L = \sup\limits_{g \in \bigcup_{x \in S} \partial f(x)} \|g\|_2$.

Subdifferential in optimization

Let $S \subseteq \mathbb{R}^n, S \neq \emptyset$ be convex and $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper and convex. We know:

- by definition x^* minimizes f on $\mathbb{R}^n$ if and only if $0 \in \partial f(x^*)$;
- $\min_{x \in S} f(x) = \min_{x \in \mathbb{R}^n} F(x), \text{ where} F(x) = f(x) + \delta_S(x)$;
- $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(S) \neq \emptyset \implies \partial F(x) = \partial f(x) + \partial \delta_S(x) = \partial f(x) + N_S(x)$.

Optimality Conditions Thm: Let $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(S) \neq \emptyset$. Then x^* minimizes f over S if and only if $-\partial f(x^*) \cap N_S(x^*) \neq \emptyset$.

Conjugate function

Def: conjugate (aka Fenchel conjugate) function of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is
$f^* : \mathbb{R}^n \to \overline{\mathbb{R}},\ f^*(y) := \sup_{x \in \mathbb{R}^n} (x^\top y - f(x))$

Closed, convex (even if f is not convex), may be not proper.

Conjugacy Thm: $f^{**}(x) := (f^*)^*(x) \leq f(x),\ \forall x \in \mathbb{R}^n$.
If f is closed, proper, convex, then $f^{**} = f$.

Examples of conjugate functions

- $f(x) = a^\top x + b,\ f^*(y) = \begin{cases} -b & \text{if } y = a \\ \infty & \text{otherwise} \end{cases}$

- $f(x) = \frac{1}{2}x^\top x,\ f^*(y) = \frac{1}{2}y^\top y$

- $f(x) = \|x\|_2,\ f^*(y) = \begin{cases} 0 & \text{if } \|y\|_2 \le 1 \\ \infty & \text{otherwise} \end{cases}$; by Hölder's inequality, if $f(x) = \|x\|_p, p \ge 1$, replace $\|y\|_2$ in f^* by $\|y\|_q$, $\frac{1}{p} + \frac{1}{q} = 1$.

- $f = \delta_S,\ f^*(y) = \sup_{x \in S} x^\top y$; and $f^*(y) = \delta_{S^*}$ if S is a convex cone.

Subgradients and conjugate functions

Conjugate Subgradient Thm: If $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper, convex, then:

$$y \in \partial f(x) \iff x^\top y = f(x) + f^*(y) \text{ for any } x, y \in \mathbb{R}^n$$

$$\iff x \in \partial f^*(y) \text{ if } f \text{ closed.}$$

Corollary: f proper, convex, closed $\implies \arg\min_{x \in \mathbb{R}^n} f(x) = \partial f^*(0)$.

Projection Theorem

Def: projection x^* of vector z on set S: $x^* = \arg\min_{x \in S} \|x - z\|_2$

Thm: projection of any point $z \in \mathbb{R}^n$ on a non-empty closed convex set S is unique and $x^* \in \mathbb{R}^n$ is this projection if and only if

$$(z - x^*)^\top (x - x^*) \leq 0 \;\forall x \in S.$$

Proximal operator

Def: Proximal operator of convex, proper, lower semi-cont. $f:\mathbb{R}^n \to \overline{\mathbb{R}}$ and $\epsilon > 0$: $\mathrm{prox}_{\epsilon,f}:\mathbb{R}^n \to \mathbb{R}^n$, $\mathrm{prox}_{\epsilon,f}(z) := \arg\min_{x\in\mathbb{R}^n} f(x) + \frac{\epsilon}{2}\|x-z\|_2^2$.

Finding prox is unconstrained convex problem, generalized projection:

$\mathrm{prox}_{\delta_S}(z)$ is equal to projection of z on S.

Thm: $\mathrm{prox}_{\epsilon,f}$ exists and is unique for any closed and convex f (extends projection thm).

Examples of prox-operators for $\epsilon = 1$

- $f(x) = 0$: $\text{prox}_f(z) = z$
- $f(x) = \frac{1}{2}x^\top Px + q^\top x + r$: $\text{prox}_f(z) = (I_n + P)^{-1}(v - q)$
- $f(x) = \|x\|_1$: $\text{prox}_f = T_1$, where $T_\epsilon : \mathbb{R}^n \to \mathbb{R}^n$ is a soft-threshold operator $T_\epsilon(z)_i = \begin{cases} x_i - \epsilon & \text{if } x_i \geq \epsilon \\ 0 & \text{if } \epsilon \geq x_i \geq -\epsilon \\ x_i + \epsilon & \text{if } -\epsilon \geq x_i \end{cases}$, $i = 1, \ldots, n$.

Properties of prox-operators

Recall: $\mathrm{prox}_f(z) := \arg\min_{x \in \mathbb{R}^n} f(x) + \frac{1}{2}\|x - z\|_2^2.$

- Fixed points of prox_f are minimizers of f.
- prox_f is firmly non-expansive $\implies$ can iteratively find min of f.
- $y = \mathrm{prox}_f(x) \iff x - y \in \partial f(y).$