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

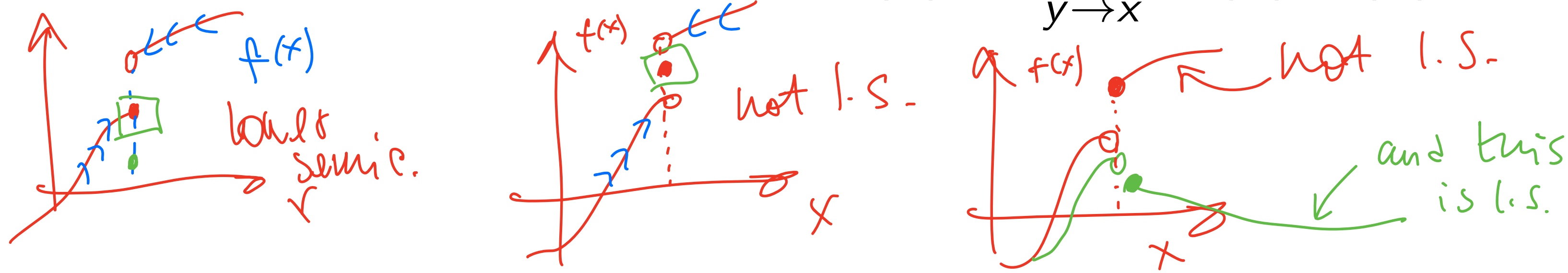
Dual view on convex functions

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

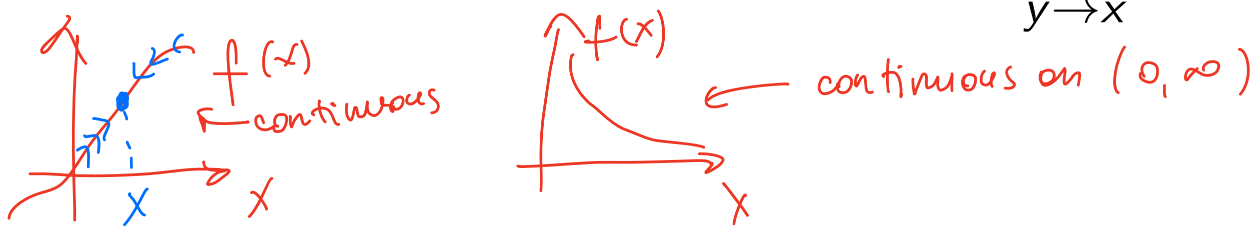
Types of continuity

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 $f(x) \leq \liminf_{y \to x} f(y), \forall (y) \subset S$.



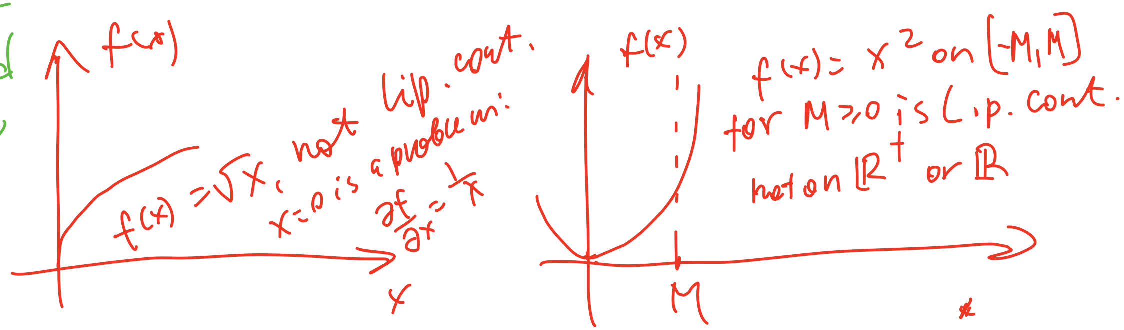
Def: f is continuous in $x \in \mathrm{dom}(f)$ if $f(x) = \lim_{y \to x} f(y), \forall (y) \subset \mathrm{dom}(f)$



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

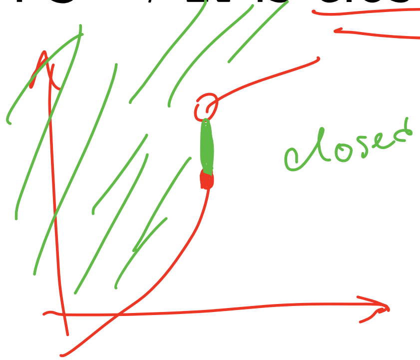
← this implies continuity by ε-δ definition or limits (check).

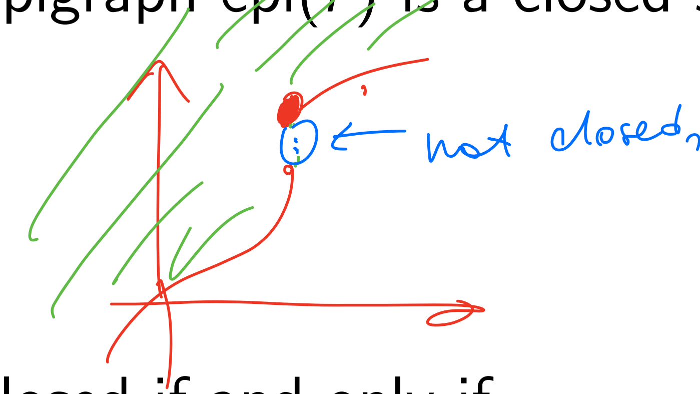
the slope is bounded from above, i.e., never vertical slope



Semicontinuity and closedness

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

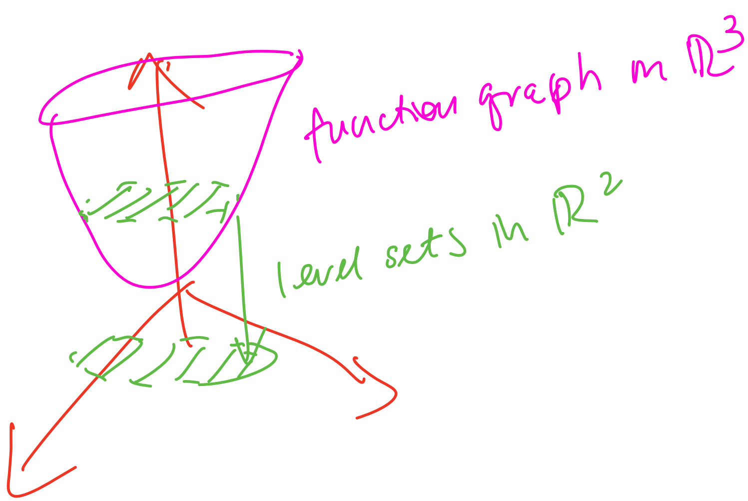




Thm: Function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is closed if and only if

$\iff$ f is lower-semicontinuous

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



Proof: Proposition 1.1.2 in Textbook.

Continuity and convexity

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

Proof: Proposition 1.3.11 in Textbook.

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 $x, y \in \text{dom}(f)$, 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.

explains the convergence of many optimization algorithms, including gradient-based.

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 $\bar{x}$.

$f(x)$

$f(\bar{x}) + \nabla f(\bar{x})^\top (x - \bar{x})$

First-order Taylor approximation in $\bar{x}$ $\iff$ there is a hyperplane that passes through $\bar{x}$ and is tangent to the graph of f in all directions

approaches the function in the limit

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 $v \in \mathbb{R}^n$.

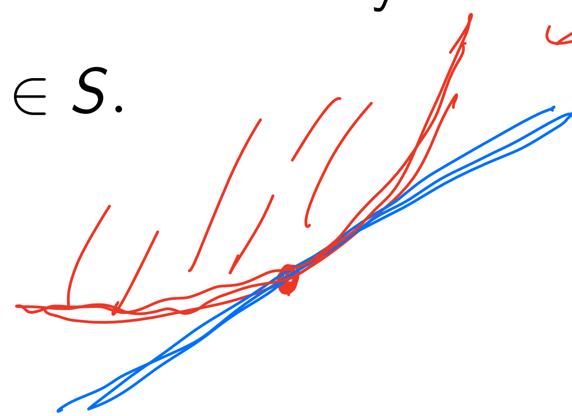
definition of directional derivative

holds for differentiable functions, follows from (*), often used in proofs.

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

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



I.e., a convex function is always above its first-order Taylor approx.

Proof: Proposition 1.1.7 in Textbook.

Idea for „if": take x, y, $p = \alpha x + (1-\alpha) y$; write

$$\left.\begin{aligned} f(x) - f(p) &\geq \nabla f(p)^T (x - p) \quad \Big|\cdot \alpha \\ f(y) - f(p) &\geq \nabla f(p)^T (y - p) \quad \Big|\cdot (1-\alpha) \end{aligned}\right| \text{ add.}$$

Idea for „only if": use directional derivative and its connection to gradient.

Frequently used

Corollary: for S and f as above,

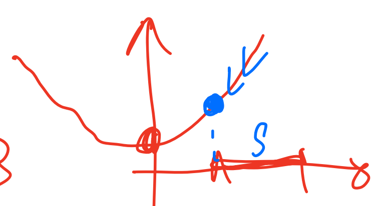
holds too but does not immediately follow from the thm above, follows from

- $\nabla f(x^*) = 0 \Longrightarrow x^*$ minimizes f over $\mathbb{R}^n$;

  this direction follows directly from theorem above, check.

- x^* minimizes f over S if and only if $\nabla f(x^*)^\top (z - x^*) \geq 0 \quad \forall z \in S.$

Proof: Proposition 1.1.8 in Textbook.



$\iff -\nabla f(x^*)^T (z - x^*) \leq 0 \quad \forall z \in S$

$\iff -\nabla f(x^*) \in N_S(x^*)$

„−gradient", aka the direction of steepest descent, points at least 90° outside S $\iff$ all descent directions point outside S.

Convex twice differentiable functions

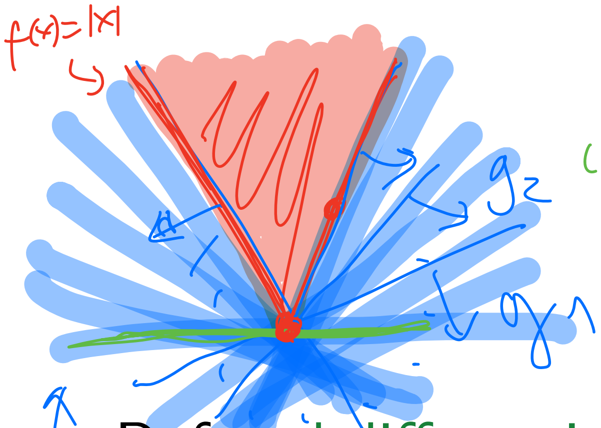
Thm: Let $S \subseteq \mathbb{R}^n$ be convex and open, f be twice continuously differentiable over S. Then f is convex over S if and only if

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

Proof: Proposition 1.1.10 in Textbook

Subgradient and subdifferential

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



normal (i.e. slope) of a supporting hyperplane to epigraph

(i)

$$f(z) - g^T z \geq f(\bar{x}) - g^T \bar{x} \quad \forall z \in \mathbb{R}^n$$

$$\begin{bmatrix} 1 \\ -g \end{bmatrix}^T \begin{bmatrix} f(z) \\ z \end{bmatrix} \geq \begin{bmatrix} 1 \\ -g \end{bmatrix}^T \begin{bmatrix} f(\bar{x}) \\ \bar{x} \end{bmatrix}$$

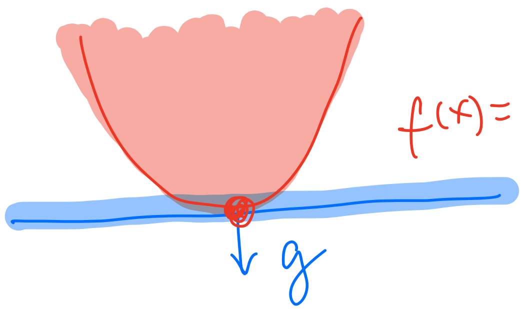
point on the epigraph of f

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

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

subgradient in $\bar{x}$ corresponds to the normal of a supporting hyperplane in $\bar{x}$ to the epigraph of f.

normals of all blue hyperplanes are subgradients.
Subdifferential consists of all these normals.



$f(x) = x^2$ here subdifferential = {gradient}

notice: subdifferential is defined by a set of linear inequalities (infinitely many)

Properties of subdifferential

see the definition of subdif.

- $\partial f(\bar{x})$ is closed and convex as an intersection of closed subspaces.

- If f is differentiable in $\bar{x}$, then $\partial f(\bar{x}) = \nabla f(x)$. Proof: Page 184 of Textbook.

- For $S \subseteq \mathbb{R}^n$, $N_S(\bar{x}) = \partial \delta_S(\bar{x})$, where $\delta_S(x) = \begin{cases} 0 & \text{if } x \in S \\ \infty & \text{otherwise.} \end{cases}$

Proof: follows by definition of subgradient and $\delta_S(x)=0$ for $x \in S$.

- Let f be convex, $A \in \mathbb{R}^{n \times m}$, and $F(x) = f(Ax)$. If f is polyhedral or $\exists \alpha \in \mathbb{R}^m : A\alpha \in \text{ri}(\text{dom}(f))$, then $\partial F(x) = A^\top \partial f(Ax)$.

Proof: Prop. 5.4.5 in Textbook

could be more than 2 functions

- Let f, h be convex and $F = f + h$ be proper. If $\text{ri}(\text{dom}(f)) \cap \text{ri}(\text{dom}(h)) \neq \emptyset$, then $\partial F(x) = \partial f(x) + \partial h(x)$.

Proof: Prop. 5.4.6. in Textbook

f must be convex

- If $\emptyset \neq S \subseteq \text{dom}(f)$ is compact, then $\bigcup_{x \in S} \partial f(x) \neq \emptyset$ and bounded; and f is Lipschitz continuous on S with constant $L = \sup_{g \in \bigcup_{x \in S} \partial f(x)} \|g\|_2$.

Proof: Proposition 5.4.2 in Textbook

Subdifferential in optimization

S must be convex.

Let $S \subseteq \mathbb{R}^n, S \neq \emptyset$ 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 $0 \in \partial f(x^*)$;
- $\min_{x \in S} f(x) = \min_{x \in \mathbb{R}^n} F(x)$, where $F(x) = f(x) + \delta_S(x)$;

  Frequently used result

- $\text{ri}(\text{dom}(f)) \cap \text{ri}(S) \neq \emptyset \implies \partial F(x) = \partial f(x) + \partial \delta_S(x) = \partial f(x) + N_S(x)$.

follows from the above 3 points

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

This is almost the same condition as for differentiable functions on slide 9, but the set of g to choose from could contain many vectors.

$\exists g: -g \in \partial f(x^*)$ and $g^T(z - x^*) \leq 0$ for all $z \in S$

$\Leftrightarrow \exists g \in \partial f(x^*)$ with $g^T(z - x^*) \geq 0$ for all $z \in S$

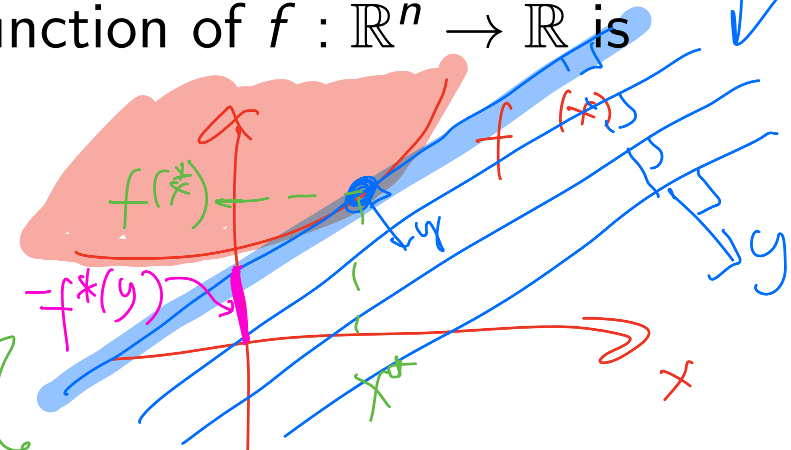
Same intuition: need to leave S if we want to decrease f since all descent directions point outside S.

Conjugate function

Among all hyperplanes normal to y, we find one supporting function f, this gives us x^* in which the hyperplane is supporting, so $y \in \partial f(x^*)$, and $-f^*(y)$ is the intercept, so the hyperplane is defined as $\begin{bmatrix}-1\\ y\end{bmatrix}^T \begin{bmatrix}t(x)\\ x\end{bmatrix} - f^*(y) = 0 \iff f^*(y) = y^T x^* - f(x^*)$.

Def: conjugate (aka Fenchel conjugate) function of $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is

$$f^* : \mathbb{R}^n \to \overline{\mathbb{R}}, \quad f^*(y) := \sup_{x \in \mathbb{R}^n} \left(x^\top y - f(x)\right)$$



if $f^*(y) \neq \infty \iff \exists x^* : f^*(y) = x^{*T} y - f(x^*)$

and $x^{*T} y - f(x^*) \geq z^T y - f(z) \quad \forall z \in \mathbb{R}^n$ $\Big\}\ y \in \partial f(x^*)$

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

supremum over affine functions

let $f(x) = \begin{cases} \infty, & x > 0 \\ -\infty, & x \leq 0 \end{cases}$, then $f^*(y) = \infty$ for all y.

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

this part follows by definition:

$$f^{**}(x) = \sup_y \left(x^T y - f^*(y)\right) \leq \sup_y \big(x^T y \underbrace{- x^T y + f(x)}_{\geq\, -f^*(y) \text{ by defn.}}\big) = f(x)$$

Proof of this part: Proposition 1.6.1 in Textbook.

f^{**} is the best convex underestimator for any function f.

Examples of conjugate functions

These conditions describe the subdifferential: if $f^*(y)$ is finite, then $y \in \partial f(x^*)$ for some x^*.
In this particular case: subdifferential = {gradient} = normal of the original line. (i.e. its slope)

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

Proof:
$f^*(y) = \sup_x x^T y - a^T x - b = \sup_x x^T(y-a) - b$

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

$\nabla f(x) = 0$

Proof:
$f^*(y) = \sup_x x^T y - \frac{1}{2}x^T x \to y - x = 0 \Rightarrow y = x \Rightarrow y^T y - \frac{1}{2}y^T y = \frac{1}{2}y^T y$

no special conditions here: every y belongs to $\partial f(x)$ for some x.

- $f(x) = \|x\|_2, \;\; f^*(y) = \begin{cases} 0 & \text{if } \|y\|_2 \le 1 \\ \infty & \text{otherwise} \end{cases}$; by Hölder's inequality,

$|x^T y| \le \|x\|_p \cdot \|y\|_q$

if $f(x) = \|x\|_p, p \ge 1$, replace $\|y\|_2$ in f^* by $\|y\|_q$, $\frac{1}{p} + \frac{1}{q} = 1$.

- $f = \delta_S, \;\; f^*(y) = \underbrace{\sup_{x \in S} x^\top y}^{\text{as finite } f(x)=0}$; and $f^*(y) = \delta_{S^\circ}$ if S is a convex cone.

polar

Proof: $\sup_{x\in S} x^T y = \begin{cases} 0 & \text{if } x^T y \le 0 \text{ for all } x \in S \\ \infty & \text{otherwise} \end{cases} = \begin{cases} 0 & \text{if } y \in S^\circ \\ \infty & \text{otherwise} \end{cases} = \delta_{S^\circ}(y)$

Subgradients and conjugate functions

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

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

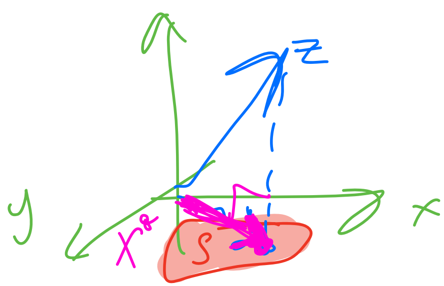
Proof: $x = \arg\sup_{x \in \mathbb{R}^n} \{ x^T y - f(x) \} \iff x^T y - f(x) \geq z^T y - f(z) \quad \forall z \in \mathbb{R}^n$

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

here $y = 0$, so $x^T y$ above disappears

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

$$\underbrace{(z - x^*)}_{\in N_S(x^*)}^\top (x - x^*) \leq 0 \; \forall x \in S.$$

Proof: Proposition 1.1.9

Proximal operator

often set to 1
if ε is omitted,
it is $= 1$

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

Finding prox is unconstrained convex problem, generalized projection:

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

Proof: Indeed, $f(x) = 0$ if it is finite, and it is finite if $x \in S$, so we have

$\text{prox}_{\varepsilon,\delta_S}(z) = \arg\min_{x \in S} \frac{\varepsilon}{2}\|x - z\|_2^2$, which is the projection.

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

Proof: exercise.

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

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

Proof: exercise. Idea: solve the optimization problem, similar to how we found conjugate functions.

Properties of prox-operators (we assume f is proper, convex)

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

- Fixed points of prox_f are minimizers of f.

Proof: 1) x^* fixed point $\Rightarrow \text{prox}_f(z) = z \Rightarrow \min_{x\in\mathbb{R}^n} f(x) + \frac{1}{2}\|x-z\| = \min_{x\in\mathbb{R}^n} f(x)$

2) x^* minimizer $\overset{\text{By def of subgradient}}{\Rightarrow} 0 \in \partial f(x^*) \Rightarrow 0 \in \partial f(x^*) + x^* - x^* \overset{\text{By (*) below and the rule of subgr. sum}}{\Rightarrow} 0 \in \partial g_{x^*}(x^*) \Rightarrow x^*$ is fixed point of prox_f.

- prox_f is firmly non-expansive $\implies$ can iteratively find min of f.

(Extension of Banach FPT)

Proof of firm non-expansiveness: exercise.

- $y = \text{prox}_f(x) \iff x - y \in \partial f(y)$.

so, ri(dom(f)) ≠ ∅ (*)

Proof: f is proper, convex $\Rightarrow f(x) + \frac{1}{2}\|x-z\|_2^2$ is so too, and $\text{dom}(f) \neq \emptyset$.

Hence, $y = \text{prox}_f(x)$ iff $0 \in \partial g_x(y) \overset{\text{By (*) and the rule of subgradient sum}}{\iff} 0 \in \partial f(y) + y - x \iff x - y \in \partial f(y)$.