Notes of CS 839: Advanced Nonlinear Optimization
Instructor: Jelena Diakonikolas

YI WEI

Sep 2024

Contents

1 Vector Space

[**YW:** TODO: Notes of Sep 4.]

[**Date:** Sep 6, 2024]

Example 1.1. 1. Induced matrix norms $A \in \mathbb{R}^{m\times n}$ Let $\|\cdot\|_a$ be any norm in $\mathbb{R}^n$, $\|\cdot\|_b$ be any norm in R^m, $\|A\|_{a,b} = \max_{x\in\mathbb{R}^n:\|x\|_a\leqslant 1}\|Ax\|_b$

In particular, if $\|\cdot\|_a$ and $\|\cdot\|_b$ are l_p norms:

(a) $a = b = 2 \to$ operator/spectral norm

(b) $a = b = 1$:

$$\|A\|_{1,1} = \max_{x\in\mathbb{R}^n,\|x\|_1\leqslant 1}\|Ax\|_1 \tag{1}$$

$$= \max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|A_{ij}| \tag{2}$$

It's called "max abs column sum"

(c) $a = b = \infty$:

$$\|A\|_{\infty,\infty} = \max_{x\in\mathbb{R}^n,\|x\|_\infty\leqslant 1}\|Ax\|_\infty = \max_{1\leqslant i\leqslant m}\sum_{j=1}^{n}|A_{ij}|$$

It's called "max abs row sum norm".

(d) $a = 1, b = \infty$:

$$\|A\|_{1,\infty} = \max_{x\in\mathbb{R}^n,\|x\|_1\leqslant 1}\|Ax\|_\infty = \max_{1\leqslant i\leqslant m,1\leqslant j\leqslant n}|A_{ij}|$$

where $\|Ax\|_\infty = \begin{bmatrix} A_1x \\ A_2x \\ \vdots \\ A_nx \end{bmatrix}$

1.1 Cartesian Product of Vector Space

Given $m \geqslant 2$ vector spaces $\mathbb{E}_1, \ldots, \mathbb{E}_m$ equipped w/ inner products $\langle\cdot,\cdot\rangle, \ldots, \langle\cdot,\cdot\rangle$, their Cartesian product is the vector space $\mathbb{E} = \mathbb{E}_1 \times \cdots \times \mathbb{E}_n$ containing all m-tuples $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ for which basic operations are defined as:

1. Addition: $(\mathbf{v}_1, \ldots, \mathbf{v}_m) + (\mathbf{w}_1, \ldots, \mathbf{w}_m) =$
2. Scaler multiplication: $\alpha \in \mathbb{R}, \alpha(\mathbf{v}_1, \ldots, \mathbf{v}_m) = (\alpha\mathbf{v}_1, \ldots, \alpha\mathbf{v}_m)$

The inner product on $\mathbb{E}$ is defined by:

$$\langle(\mathbf{v}_1, \ldots, \mathbf{v}_m), (\mathbf{w}_1, \ldots, \mathbf{w}_m)\rangle = \sum_{i=1}^{m}\langle v_i, w_i\rangle_{\mathbb{E}_i}$$

If $\mathbb{E}_i, i \in \{1, \ldots, m\}$ are endowed w/ norms $\|\|_{E_i}$ there a different ways of choosing a norm on $\mathbb{E}$

Example 1.2.

$\|(\mathbf{v}_1, \ldots, \mathbf{v}_m)\| = (\sum_{i=1}^{m} \|v_i\|_{\mathbb{E}_i}^{p})^{\frac{1}{p}}$

$\|(\mathbf{v}_1, \ldots, \mathbf{v}_m)\| = (\sum_{i=1}^{m} w_i\|v_i\|_{\mathbb{E}_i}^{2})$

1.2 Linear Transformation

Definition 1.1. Given two vector spaces $\mathbb{E}$, $\mathbb{V}$, $f : \mathbb{E} \to \mathbb{V}$ is a linear transformation if

$$\forall x, y \in \mathbb{E}, \forall \alpha, \beta \in \mathbb{R} :$$
$$A(\alpha x + \beta y) = \alpha A(x) + \beta A(y)$$

Example 1.3.

1. All linear transformations from $\mathbb{R}^n \to \mathbb{R}^m$ are of the from

$$A(x) = Ax \quad \text{for some matrix } A \in \mathbb{R}^{m\times n}$$

2. All linear transformations from $\mathbb{R}^{n\times n} \to \mathbb{R}^k$ are of the form:

$$A(X) = \begin{bmatrix} \text{trace}(A_1^\top X) \\ \text{trace}(A_2^\top X) \\ \vdots \\ \text{trace}(A_n^\top X) \end{bmatrix} \quad \forall\ X \in \mathbb{R}^{m\times n}$$

some matrices $A_1, \ldots, A_k \in \mathbb{R}^{m\times n}$

3. The identity transformation $\mathcal{I} : \mathbb{E} \to \mathbb{E}$ is defined by $\mathcal{I}(x) = x$

1.3 The Dual Space

Definition 1.2. The dual space of a vector space $\mathbb{E}$ is the space of all linear functionals on $\mathbb{E}$

For inner product spaces, (Riez Representation) for any linear functional f, $\exists v \in \mathbb{E}$ s.t $f(x) = \langle \mathbf{v}, \mathbf{x} \rangle \quad \forall \mathbf{x} \in \mathbb{E}$.
We write $\mathbf{v} \in \mathbb{E}^*$ (notation).
Elements of $\mathbb{E}^*$ and $\mathbb{E}$ are the same if $\mathbb{E}$ we use a norm $\|\cdot\|$, then in $\mathbb{E}^*$ we use the norm dual to it, defined by (dual norm)

$$\forall \mathbf{y} \in \mathbb{E}^* : \|\mathbf{y}\|_* := \max_{\mathbf{x}\in\mathbb{E}:\|x\|\leqslant 1} \langle \mathbf{y}, \mathbf{x} \rangle$$

Theorem 1.1. Generalized Cauchy-Schwarz:

$$\forall \mathbf{x} \in \mathbb{E}, \forall \mathbf{y} \in \mathbb{E}^* : \|\langle \mathbf{x}, \mathbf{y} \rangle\| \leqslant \|\mathbf{x}\|\|\mathbf{y}\|_*$$

Theorem 1.2. Euclidean norms are self-dual. We say that Euclidean space "self-dual" and write $\mathbb{E} = \mathbb{E}^*$

Example 1.4. 1. In $\mathbb{R}^d$, with $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}$

(a) The norm dual to l_p norm for $p > 1$ is the norm l_p^* where $\frac{1}{p} + \frac{1}{p^*} = 1$. l_1 and l_∞ are dual to each other.

(b) The norm dual to $\| \cdot \|_Q$ for Q symmetric, positive definite is $\| \cdot \|_{Q^{-1}}$

$$\|\mathbf{x}\|_{Q^{-1}} = \left(\mathbf{x}^\top Q^{-1} x\right)^{\frac{1}{2}}$$

If $Q = \mathrm{diag}(w_1, \ldots, w_d)$ for positive $w_1, \ldots, w_d$, then $\|\mathbf{x}\|_{Q^{-1}} = \left(\sum_{i=1}^d \frac{1}{w_i} \mathbf{x}_i^2\right)^{\frac{1}{2}}$

2. $E = E_1 \times \cdots \times E_m$, with $\| \cdot \|_{E_1}, \ldots, \| \cdot \|_{E_m}$

$$\|(\mathbf{v}_1, \ldots, \mathbf{v}_m)\|_{\mathbb{E}} = \left(\sum_{i=1}^m w_i \|\mathbf{v}_i\|_{\mathbb{E}_i}^2\right)^{\frac{1}{2}}$$

$$\|(\mathbf{w}_1, \ldots, \mathbf{w}_m)\|_{\mathbb{E}^*} = \left(\sum_{i=1}^m \frac{1}{w_i} \|\mathbf{u}_i\|_{\mathbb{E}_i^*}^2\right)^{\frac{1}{2}}$$

Theorem 1.3. Bidual space = dual space to $\mathbb{E}^*$.
In finite vector space, $\mathbb{E}^{**} = \mathbb{E}$

Theorem 1.4. $\langle A\mathbf{x}, \mathbf{y} \rangle \leqslant \|A\|_{a,b} \|\mathbf{x}\|_a \|\mathbf{y}\|_b$ if $\| \cdot \|_a$ and $\| \cdot \|_b$ are dual to each other.

1.4 Adjoint Transformation

Definition 1.3. Given vector space $\mathbb{E}$ and $\mathbb{V}$, and a linear transformation $A : \mathbb{E} \to \mathbb{V}$, the adjoint transformation $A^\top : \mathbb{V}^* \to \mathbb{E}^*$ is defined by

$$\langle \mathbf{y}, A(x) \rangle = \langle A^\top(y), \mathbf{x} \rangle$$

Example 1.5. In particular,

1. If $\mathbb{E} = \mathbb{R}^n, \mathbb{V} = \mathbb{R}^m$, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{y}$, then, $A(x) = Ax$ for some $A \in \mathbb{R}^{m \times n}$ and $A^\top(y) = A^\top \mathbf{y}$

2. $\mathbb{E} = \mathbb{R}^{m \times n}, \mathbb{V} = \mathbb{R}^k$

[**Date:** Sep 13, 2024] Given $A : \mathbb{E} \to \mathbb{V}$, $\|\cdot\|_{\mathbb{E}}$, $\|\cdot\|_{\mathbb{E}}$, we define the norm $\|A\| = \sup_{x \in \mathbb{E}, \|x\|_{\mathbb{E}} \leqslant 1} \|A(x)\|_{\mathbb{V}}$

2 Extended Real-Valued Functions

Definition 2.1. functions that map some real vector space $(\mathbb{E}, \langle \cdot, \cdot \rangle), \| \cdot \|$ to the extended real line -either $\mathbb{R} \bigcup \{-\infty, +\infty\} \equiv [-\infty, +\infty]$ or $\mathbb{R} \bigcup \{+\infty\} \equiv (-\infty, +\infty]$

$$\min_{x \in \mathbb{E}} \quad f(x)$$

Consider this problem, why do we even want to include $+\infty$

1. f is not everywhere defined on $\mathbb{E}$, I can assign it to $+\infty$ at points where it's not defined. So when it becomes well-defined on all $\mathbb{E}$.

   Here we define the domain = effective domain:

$$dom(f) = \{x \in \mathbb{E} : f(x) < +\infty\}$$

2. We can think of all optimization problems whether constrained or unconstrained, as unconstrained optimization problem.

$$\min_{x \in \mathcal{X}} f(x) \iff \min_{x \in \mathbb{E}} f(x) + \delta_{\mathcal{X}}(x)$$
$$\text{where } \delta(x) = \begin{cases} 0, & for x \in \mathcal{X} \\ +\infty, & o.w. \end{cases}$$

"Rules" for dealing with $\pm\infty$ and $a \in \mathbb{R}$:

1. $a + \infty = +\infty + a = +\infty$
2. $a - \infty = -\infty + a = -\infty$
3. $$a \cdot \infty = \begin{cases} \infty, & if\ a > 0 \\ -\infty, & if\ a < 0 \end{cases}$$
4. $0 \cdot \pm\infty = 0$
5. $-\infty < a < \infty \quad \forall a \in \mathbb{R}$

2.1 Closed Functions

Definition 2.2. $epi(f) := \{(x, y) : x \in \mathbb{E}, y \in \mathbb{R}, f(x) \leqslant y\}$

Definition 2.3. A function $f : \mathbb{E} \to [-\infty, \infty]$ is said to be closed if $epi(f)$ is closed.

Proposition 2.1. For $C \subseteq \mathbb{E}$, $\sigma_C(x)$ is closed $\iff$ C is closed.

Proof. $epi(C) = C \times \mathbb{R}_+$ □

Remark. f is closed $\not\Longleftrightarrow$ $dom(f)$ is closed.

Example 2.1.

$$f(x) = \begin{cases} \frac{1}{x}, & x > 0 \\ \infty, & x \leqslant 0 \end{cases}$$

Then $dom(f) = (0, \infty)$ is open. And we see that:

$$epi(f) = \{(x, y) \in \mathbb{R}^2 : x > 0, \frac{1}{x} \leqslant y\}$$

2.1.1 Related Concepts

1. Lower Semicontinuity:

   Definition 2.4. $f : \mathbb{E} \to [-\infty, +\infty]$ is l.s.c. at $x \in \mathbb{E}$ if

   $$f(x) \leqslant \liminf_{n \to \infty} f(x_n)$$

   for any sequence $\{x_n\}_{n \geqslant 1} \in \mathbb{E}$ s.t. $x_n \to x$ as $n \to \infty$.

   f is said to be l.s.c. if it is l.s.c. at all $x \in \mathbb{E}$.

2. Level set: defined for $\alpha \in \mathbb{R}$, $f : \mathbb{E} \to [-\infty, +\infty]$.

   $$Lev(f, \alpha) = \{x \in \mathbb{E} : f(x) \leqslant \alpha\}$$

Theorem 2.2. If $f : \mathbb{E} \to [-\infty, +\infty]$. Then all of the following statements are equivalent:

1. f is l.s.c.
2. f is closed.
3. $Lev(f, \alpha)$ is closed, $\forall \alpha \in \mathbb{R}$

2.1.2 Operations preserving closedness

1. If $f : \mathbb{V} \to [-\infty, +\infty]$ is closed, $A : \mathbb{E} \to \mathbb{V}$ is a linear transformation and $b \in \mathbb{V}$, then

   $$g(x) = f(A(x) + b) \text{ is closed.}$$

2. If $f_1, \ldots, f_m : \mathbb{E} \to (-\infty, +\infty]$ are closed and $\alpha_1, \ldots, \alpha_m \in \mathbb{R}_+$, then

   $$f(x) = \sum_{i=1}^{n} \alpha_i f_i(x) \text{ is closed}$$

3. Given an index set I and functions $f_i : \mathbb{E} \to (-\infty, \infty]$, $i \in I$, that are closed, the function

$$f(x) = \sup_{i \in I} f_i(x) \text{ is closed.}$$

2.1.3 Closedness vs Continuity

Bottom line: If f has closed domain + continuous over the domain $\Longrightarrow$ closed.

But closed $\not\Longleftarrow$ continuous over the domain.

Theorem 2.3. Let $f : \mathbb{E} \to (-\infty, +\infty]$ be continuous over its domain and suppose $dom(f)$ is closed $\Longrightarrow$ f is closed.

Proof. Argue that $epi(f)$ is closed.

Take any sequence $\{(x_n, y_n)\}_{n \geqslant 1} \in epi(f)$ that converges to some (x_*, y_*) as $n \longrightarrow \infty$

To argue: $(x_*, y_*) \in epi(f)$: we know that $x_n \in dom(f)$, $x_n \longrightarrow x_*$, $dom(f)$ is closed $\Longrightarrow x_* \in dom(f)$

By the definition of $epi(f)$:

$$f(x_n) \leqslant y_n$$

Since f is continuous over $dom(f)$ and $\{x_n\}_n, x_* \in dom(f)$ we can take the limit $n \longrightarrow \infty$

$$f(x_*) \leqslant y_*$$
$$\Longrightarrow (x_*, y_*) \in epi(f)$$

$\square$

Example 2.2 (closed $\not\Longrightarrow$ continuous on its domain)**.**

$$f_\alpha(x) = \begin{cases} \alpha, & x = 0 \\ x, & 0 < x \leqslant 1 \\ \infty, & elsewhere \end{cases} \tag{3}$$

When $\alpha < 0$, then it's l.s.c., i.e., closed, but it's not continuous.

l_0 "norm"

$$f(x) = \|\mathbf{x}\|_0 = |\{i : \mathbf{x}_i \neq 0\}|$$

f is not continuous but it's closed.

$$f(x) = \sum_{i=1}^{d} I(\mathbf{x}_i)$$

where

$$I(y) = \begin{cases} 0, & y = 0 \\ 1, & y \neq 0 \end{cases}$$

We know

$$Lev(I, \alpha) = \begin{cases} \varnothing, & \alpha < 0 \\ \{0\}, & 0 \leqslant \alpha < 1 \\ \mathbb{R}, & \alpha \geqslant 1 \end{cases}$$

Then I is closed. $\implies$ the sum of them is closed.

[**Date:** Sep 16, 2024]

Theorem 2.4 (Weierstrass theorem for closed functions)**.** Let $f : \mathbb{E} \to (-\infty, \infty]$ be a $\underbrace{\text{proper}}_{dom(f)\neq\varnothing}$, closed function and let $C \subseteq \mathbb{E}$ be a compact set such that $C \bigcap dom(f) \neq \varnothing$. Then:

1. f is bounded below on C.
2. f attains its minimal value over C.

Proof. 1. Suppose for the purpose of contradiction (FPOC) that f is not bounded below on C. Then $\exists$a sequence $\{x_n\}_{n\geqslant 1}$, $x_n \in C \forall n$, s.t.

$$\lim_{n\to\infty} f(x_n) = -\infty$$

By Bolzano-Weierstrass, since C is compact, there exists a subsequence $\{x_{n_k}\}_{k\geqslant 1}$ that converges to a point $\bar{x} \in C$. Since

$$f \text{ is closed } \iff f \text{ is l.s.c.}$$

We know

$$\begin{aligned} f(\bar{x}) \leqslant \lim_{k\to\infty} f(x_{n_k}) &= -\infty \\ \implies f(\bar{x}) &= -\infty \end{aligned}$$

Contradiction.

2. Let $f_* = \inf_{x \in C} f(x) > -\infty$.

Claim 2.5. $\exists$ a sequence $\{x_n\}_{n \geqslant 1}$ s.t.

$$f(x_n) \to f_* \text{ as } n \to \infty$$

Then $(x_n, f(x_n)) \in epi(f)$. Then take a subsequence $\{x_{n_k}\}_{k \geqslant 1}$ s.t. $x_{n_k} \to \bar{x}$. Then

$$\begin{aligned} f(\bar{x}) &\leqslant \liminf_{k \to \infty} f(n_k) = f_* \\ &\implies \bar{x} \text{ minimizes } f \end{aligned}$$

$\square$

What is we are not optimizing over a compact set.

Definition 2.5. A proper function $f : \mathbb{E} \to (-\infty, \infty]$ is said to be coercive if

$$\lim_{x \in \mathbb{E}: \|x\| \to \infty} f(x) = +\infty$$

Theorem 2.6. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper, closed and coercive function, and let $S \subseteq \mathbb{E}$ be a nonempty closed set that satisfy $S \bigcap dom(f) \neq \varnothing$. Then f attains the minimum over set S.

Proof. Let x_0 be an arbitrary point $\square$

2.2 Convex Function

Definition 2.6 (Equivalent definitions of convexity)**.** f is convex if

1. $epi(f)$ is convex

2. $\forall x, y \in \mathbb{E}, \forall \alpha \in (0, 1)$:

$$f((1-\alpha)x + \alpha y) \leqslant (1-\alpha) f(x) + \alpha f(y)$$

Remark. Notice this induces Jensen's inequality: $\forall x_1, \dots, x_m \in \mathbb{E}, \forall \lambda_1, \dots, \lambda_m \geqslant 0, \sum_{i=1}^n \lambda_i = 1$

$$f(\sum_{i=1}^m \lambda_i x_i) \leqslant \sum_{i=1}^m \lambda_i f(x_i)$$

3. if f is continuously differentiable: $\forall x, y \in \mathbb{E}$

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle$$

4. if $f \in C^2$: $\forall x \in \mathbb{E}$:

$$\nabla^2 f(x) \succcurlyeq 0$$

Theorem 2.7 (Operations preserving convexity). 1. If $A : \mathbb{E} \to \mathbb{V}$ liner transform, $b \in \mathbb{V}$,and $f : \mathbb{V} \to (-\infty, \infty]$ is convex, then $f(A(x) + b)$ is convex.

2. $f_1, \ldots, f_m : \mathbb{E} \to (-\infty, +\infty]$ are convex, $\lambda_1, \ldots, \lambda_n \in \mathbb{R}_+$, then $f(x) = \sum_{i=1}^m \lambda_i f_i(x)$ is convex.

3. I : inded set, $f_i : \mathbb{E} \to (-\infty, \infty]$ convex $\forall i \in I$, then $f(x) = \sup_{i \in I} f_i(x)$ is convex.

Example 2.3. Given $C \subseteq \mathbb{E}$ that is nonempty (but not necessarily convex), let

$$d_C(x) = \inf_{y \in C} \|y - x\|$$

If $\mathbb{E}$ is Euclidean, then $\varphi_C(x) = \frac{1}{2}(\|x\|^2 - d_C^2(x))$ is convex. Notice that

$$\begin{aligned} d_C^2(x) = \inf_{y \in C} \|y - x\|^2 &= \inf_{y \in C} \{\|x\|^2 - 2\langle x, y\rangle + \|y\|^2\} \\ &= \|x\|^2 - \sup_{y \in C} \{2\langle y, x\rangle - \|y\|^2\} \end{aligned}$$

Theorem 2.8 (Convexity under partial minimization). Let $f : \mathbb{E} \times \mathbb{V} \to (-\infty, \infty]$ be a convex function s.t. $\forall x \in \mathbb{E}, \exists y \in \mathbb{V} : \; f(x, y) < \infty$. Let $g : \mathbb{E} \to [-\infty, \infty)$ be defined

$$g(x) := \inf_{y \in \mathbb{V}} f(x, y)$$

Then g is convex.

Proof. To show $\forall x_1, x_2 \in \mathbb{E}, \forall \alpha \in (0, 1)$:

$$g((1 - \alpha)x_1 + \alpha x_2) \leqslant (1 - \alpha)g(x_1) + \alpha g(x_2)$$

Case 1: $g(x_1), g(x_2) > -\infty$. Take any $\epsilon > 0$, then $\exists\, y_1, y_2 \in \mathbb{E}$ s.t.

$$\begin{aligned} f(x_1, y_1) &\leqslant g(x_1) + \epsilon \\ f(x_2, y_2) &\leqslant g(x_2) + \epsilon \end{aligned}$$

f is convex so:

$$\begin{aligned} f((1 - \alpha)x_1 + \alpha x_2, (1 - \alpha)y_1 + \alpha y_2) &\leqslant (1 - \alpha)f(x_1, y_1) + \alpha f(x_2, y_2) \\ &\leqslant (1 - \alpha)g(x_1) + \alpha g(x_2) + \epsilon \end{aligned}$$

Then by the definition of g, we have:

$$g((1-\alpha)x_1 + \alpha x_2) \leqslant (1-\alpha)g(x_1) + \alpha g(x_2) + \epsilon \quad \forall \epsilon > 0$$

Case 2: Assume at least one of $g(x_1), g(x_2)$ is equal $-\infty$. Want to show:

$$g((1-\alpha)x_1 + \alpha x_2) = -\infty$$

Take any $M \in \mathbb{R}$, then $\exists y_1$ s.t. $f(x_1, y_1) \leqslant M$. And $\exists y_2$ s.t. $f(x_2, y_2) < \infty$. Since f is convex

$$\begin{aligned} f((1-\alpha)x_1 + \alpha x_2, (1-\alpha)y_1, \alpha y_2) \\ \leqslant (1-\alpha)f(x_1, y_1) + \alpha f(x_2, y_2) \\ \leqslant (1-\alpha)M + \alpha f(x_2, y_2) \end{aligned}$$

Then by the definition of g, we have

$$g((1-\alpha)x_1 + \alpha x_2) \leqslant (1-\alpha)M + \alpha f(x_2, y_2)$$

M is arbitrary.

$\square$

2.2.1 Infimal Convolution

Definition 2.7. $h_1, h_2 : \mathbb{E} \to (-\infty, \infty]$, both proper

$$h_1 \,\square\, h_2(x) = \inf_{u \in \mathbb{E}} \{h_1(u) + h_2(x-u)\}$$

Remark. It's important for proximal point method. You smoothe functions by infimal convolution with some good functions like quadratic functions.

[**Date:** Sep 20, 2024]

Theorem 2.9. Let $h_1 : \mathbb{E} \to (-\infty, \infty]$ be proper and convex, $h_2 : \mathbb{E} \to \mathbb{R}$ be a real-valued convex function. Then $h_1 \,\square\, h_2$ is convex.

Proof. Define $f(x, y) = h_1(y) + h_2(x-y)$, $g(x) = \inf_{y \in \mathbb{E}} f(x, y) = (h_1 \,\square\, h_2)(x)$

Notice that f is convex since it's the sum of two convex functions, and $h_2 : \mathbb{E} \times \mathbb{E} \to \mathbb{R}$ is convex since $x - y$ is a linear transform of x, y.

Want to show that $\forall \mathbf{x} \in \mathbb{E}$, $\exists \mathbf{y} \in \mathbb{E}$ s.t.

$$h_1(y) + h_2(x - y) < \infty \tag{4}$$

It's obvious since h_1 is proper and h_2 is real-valued.

Then by Theorem 2.8, f is convex. $\square$

Example 2.4.

If $C \subseteq \mathbb{E} \neq \varnothing$ is convex, then

$$d_C(x) = \inf_{y \in C} \|\mathbf{y} - \mathbf{x}\|$$

is convex. (This holds for any norm.)

Remark. $\|\cdot\|$ is convex.

We write

$$d_C(x) = \inf_{\mathbf{y} \in \mathbb{E}} \{\|\mathbf{y} - \mathbf{x}\| + \delta_C(y)\}$$
$$= \underbrace{\delta_C}_{\text{convex \& proper}} \;\square\; \underbrace{\|\cdot\|}_{\text{convex \& real-valued}}$$

Notice that $\delta_C(\cdot)$ is convex when C is a convex set.

2.2.2 Continuity of convex functions

Theorem 2.10. Let $f : \mathbb{E} \to (-\infty, \infty]$ be convex. let $x_0 \in intdom(f)$. Then $\exists \epsilon > 0$ and $L > 0$ s.t. $\underbrace{B[x_0, \epsilon]}_{\text{closed ball, centered at } x_0 \text{ of radius } \epsilon} \subseteq dom(f)$ and

$$\forall x \in B[x_0, \epsilon] : |f(x) - f(x_0)| \leqslant L\|\mathbf{x}_0 - \mathbf{x}_0\|$$

2.3 Support Function

Definition 2.8. Let $C \subseteq \mathbb{E}$ be nonempty. Then the support function of C is defined by

$$\sigma_C : \mathbb{E}^* \to (-\infty, \infty]$$
$$\sigma_C(y) = \sup_{\mathbf{x} \in C} \langle \mathbf{x}, \mathbf{y} \rangle$$
$$\text{Or: } \sigma_C(y) = \sup_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{x}, \mathbf{y} \rangle - \delta_C(x)\}$$

Lemma 2.11. Let $C \subseteq \mathbb{E}$ be a nonempty set. Then σ_C is both closed and convex.

2.3.1 Operations on sets

1. Minkowski sum:

$$A,\, B \subseteq \mathbb{E}, A + B = \{a + b : a \in A, b \in B\}$$

2. for $\alpha \in \mathbb{R},\, A \subseteq \mathbb{E}$:

$$\alpha A = \{\alpha a : a \in A\}$$

Proposition 2.12 (Properties of support functions:). 1. positive homogeneity:

$$\begin{aligned} \forall C \subseteq \mathbb{E}, C \neq \varnothing, \forall \mathbf{y} \in \mathbb{E}^*,\, \forall \alpha \geqslant 0 : \\ \sigma_C(\alpha y) = \alpha \sigma_C(y) \\ \sigma_{\alpha C}(\mathbf{y}) = \alpha \sigma_C(\mathbf{y}) \end{aligned}$$

2. subadditivity: $\forall C \subseteq \mathbb{E},\, C \neq \varnothing$,

$$\forall vy_1,\, vy_2 \in \mathbb{E}^* : \sigma_C(vy_1 + vy_2) \leqslant \sigma_C(\mathbf{y}_1) + \sigma_C(\mathbf{y}_2)$$

3. $\forall A,\, B \subseteq \mathbb{E},\, A \bigcup B \neq \varnothing,\, \forall \mathbf{y} \in \mathbb{E}^* :$

$$\sigma_{A+B}(\mathbf{y}) = \sigma_A(\mathbf{y}) + \sigma_B(\mathbf{y})$$

Example 2.5. 1. $C = conv\{\boldsymbol{b}_1, , \ldots, \boldsymbol{b}_m\},\ \boldsymbol{b}_i \in \mathbb{E} \quad \forall\, i$, then

$$\sigma_C(\mathbf{y}) = \max_{1 \leqslant i \leqslant m} \langle \boldsymbol{b}_i, \mathbf{y} \rangle$$

2. Let $K \subseteq \mathbb{E}$ be a cone set s.t. if $\mathbf{x} \in K$, then $\forall r > 0,\ r\mathbf{x} \in K$.

   The polar cone of K is defined by:

$$K^o := \{\mathbf{y} \in \mathbb{E}^* : \langle \mathbf{y}, \mathbf{x} \rangle \leqslant 0,\ \forall\, \mathbf{x} \in K\}$$

   Then $\sigma_K(\mathbf{y}) = \delta_{K^o}(y)$

3. $\mathbb{E} = \mathbb{R}^d,\ C = \mathbb{R}^d_+$

$$\sigma_{\mathbb{R}^d_+}(\mathbf{y}) = \delta_{\mathbb{R}^d_-}(\mathbf{y})$$

4. $\mathbb{E} = \mathbb{R}^d,\ A \in \mathbb{R}^{n \times d},\ S = \{\mathbf{x} \in \mathbb{R}^d :\ Ax \leqslant 0\}$

$$\sigma_S(\mathbf{y}) = \delta_{S^o}(\mathbf{y})$$

$$\text{where } S^o = \{A^\top \lambda : \lambda \in \mathbb{R}^n_+\}$$

5. $C = B_{\|\cdot\|}[0,1] = \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x}\| \leqslant 1\}$

$$\sigma_C(\mathbf{y}) = \sup_{\mathbf{x}\in\mathbb{E}:\|\mathbf{x}\|\leqslant 1} \langle \mathbf{y}, \mathbf{x}\rangle = \|\mathbf{y}\|_*$$

Proposition 2.13. If $A \subseteq \mathbb{E}$, $A \neq \emptyset$, then

1. $\sigma_A = \sigma_{cl}(A)$ where $cl(A)$ = closure of A
2. $\sigma_A = \sigma_{conv}(A)$ where $conv(A)$ = convex hull of A

If A, $B \subseteq \mathbb{E}$ are closed, convex and nonempty, then

$$A = B \iff \sigma_A = \sigma_B$$

3 Subdifferentiation

Definition 3.1. The directional derivative of a function $f : \mathbb{E} \to [-\infty, \infty]$ at $\bar{\mathbf{x}} \in \mathbb{E}$ in a direction $\boldsymbol{z} \in \mathbb{E}$ is:

$$f'(\bar{\mathbf{x}}; \boldsymbol{z}) = \lim_{\alpha \to 0} \frac{f(\bar{\mathbf{x}} + \alpha \boldsymbol{z}) - f(\bar{\mathbf{x}})}{\alpha}$$

when this limit exists.

When the directional derivative $f'(\bar{\mathbf{x}}; \boldsymbol{z})$ is linear in $\boldsymbol{z}$, then we say that f is Gateaux differentiable.

$\mathbb{E} = \mathbb{R}^d$. $\exists g \in \mathbb{E}^*$ s.t. $f'(\mathbf{x}, \boldsymbol{z}) = \langle g, \boldsymbol{z} \rangle$, we say g is the Gateaux derivative.

If f is differentiable on every point of $C \subseteq \mathbb{E}$, we say that f is differentiable on C.

Theorem 3.1. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper convex function and let $\mathbf{x} \in intdom(f)$. Then $\forall \boldsymbol{z} in \mathbb{E}$, the directional derivative $f'(\mathbf{x}; \boldsymbol{z})$ exists.

Exercise 3.1. Show that if f attains $-\infty$, then it would be $-\infty$ anywhere.

Lemma 3.2. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper convex function and let $\mathbf{x} \in intdom(f)$. Then

1. $\boldsymbol{z} \mapsto f'(\mathbf{x}; \boldsymbol{z})$ is convex;

2. $\forall \lambda > 0, \forall \boldsymbol{z} \in \mathbb{E} : f'(\mathbf{x}; \lambda \boldsymbol{z}) = \lambda f'(\mathbf{x}; \boldsymbol{z})$

Proof. 1. Take $\boldsymbol{z}_1; \boldsymbol{z}_2 \in \mathbb{E}$ and $\lambda \in (0, 1)$.

$$\begin{aligned} f'(\mathbf{x}; \lambda \boldsymbol{z} + (1 - \lambda)) &= \lim_{\alpha \to 0} \frac{f(\mathbf{x} + \alpha(\lambda \boldsymbol{z}_1 + (1 - \lambda)\boldsymbol{z}_2) - f(\mathbf{x}))}{\alpha} \\ &= \lim_{\alpha \to 0} \frac{f(\lambda(\mathbf{x} + \alpha \boldsymbol{z}_1) + (1 - \lambda)(\mathbf{x} + \alpha \boldsymbol{z}_2)) - f(\mathbf{x})}{\alpha} \\ \leqslant \lambda \lim_{\alpha \to 0} \frac{f(\mathbf{x} + \alpha \boldsymbol{z}_1) - f(x)}{\alpha} &+ (1 - \lambda) \lim_{\alpha \to 0} \frac{f(\mathbf{x} + \alpha \boldsymbol{z}_2) - f(x)}{\alpha} \\ &= \lambda f'(\mathbf{x}; \boldsymbol{z}_1) + (1 - \lambda) f'(\mathbf{x}; \boldsymbol{z}_2) \end{aligned}$$

2. $\lambda = 0$ Trivial. Assume $\lambda > 0$:

$$f'(\mathbf{x}; \lambda \boldsymbol{z}) = \lambda \lim_{\alpha \to 0} \frac{f(\bar{\mathbf{x}} + \lambda \alpha \boldsymbol{z}) - f(\bar{\mathbf{x}})}{\lambda \alpha}$$

$\square$

Lemma 3.3. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper convex function and let $\mathbf{x} \in intdom(f)$. Then:

$$f(\mathbf{y}) \geqslant f(\mathbf{x}) + f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) \quad \forall \ \mathbf{y} \in dom(f)$$

Proof.

$$
\begin{aligned}
f'(\mathbf{x}; \mathbf{y} - \mathbf{x}) &= \lim_{\alpha \to 0} \frac{f((1-\alpha)\mathbf{x} + \alpha \mathbf{y}) - f(\mathbf{x})}{\alpha} \\
&\leqslant f(\mathbf{y}) - f(\mathbf{x})
\end{aligned}
$$

□

3.1 Directional derivative of a max-type function

Theorem 3.4. Suppose that $f(\mathbf{x}) = \max_{1 \leqslant i \leqslant m} f_i(\mathbf{x})$ where $f_1, \dots, f_m : (-\infty, \infty]$ are proper. Let $\mathbf{x} \in \bigcap_{i=1}^{m} intdom(f_i)$ and let $\boldsymbol{z} \in \mathbb{E}$. Assume that $f_i'(\mathbf{x}; \boldsymbol{z})$ exists, $\forall\, i \in \{1, \dots, m\}$. Then:

$$
f'(\mathbf{x}; \boldsymbol{z}) = \max_{i \in I(\mathbf{x})} f_i'(\mathbf{x}; \boldsymbol{z}),
$$

where $I(\mathbf{x}) = \{i : s.t\ f_i(\mathbf{x}) = f(\mathbf{x})\}$

Proof. For any $i \in \{1, \dots, m\}$:

$$
\begin{aligned}
\lim_{\alpha \to 0^+} f_i(\mathbf{x} + \alpha \boldsymbol{z}) = \lim_{\alpha \to 0} \left\{ \alpha \frac{f_i(\mathbf{x} + \alpha \boldsymbol{z}) - f_i(\mathbf{x})}{\alpha} + f_i(\mathbf{x}) \right\} \\
= 0 \cdot f_i'(\mathbf{x}; \boldsymbol{z}) + f_i(\mathbf{x}) \\
= f_i(\mathbf{x})
\end{aligned}
$$

By the definition of $I(\mathbf{x})$, $f_i(\mathbf{x}) > f_j(\mathbf{x})$, $\forall i \in I(\mathbf{x})$, $j \neq I(\mathbf{x})$,
$\implies \exists \epsilon > 0,\ \forall\, \alpha \in (0, \epsilon]$ s.t.

$$
\begin{aligned}
f_i(\mathbf{x} + \alpha \boldsymbol{z}) > f_j(\mathbf{x} + \alpha \boldsymbol{z}) \quad \forall i \in I(\mathbf{x}), j \notin I(\mathbf{x}) \\
\implies \forall\, \alpha \in (0, \epsilon] :\ f(\mathbf{x} + \alpha \boldsymbol{z}) = \max_{i \leqslant i \leqslant m} f_i(\mathbf{x} + \alpha \boldsymbol{z}) \\
= \max_{i \in I(\mathbf{x})} f_i(\mathbf{x} + \alpha \boldsymbol{z}) \\
\implies \forall \alpha \in (0, \epsilon] : \\
\frac{f(\mathbf{x} + \alpha \boldsymbol{z}) - f(\mathbf{x})}{\alpha} = \frac{\max_{i \in I(\mathbf{x})} (f_i(\mathbf{x} + \alpha \boldsymbol{z}) - f_i(\mathbf{x}))}{\alpha}
\end{aligned}
$$

We obtain:

$$
\begin{aligned}
\lim_{\alpha \to 0} \frac{f(\mathbf{x} + \alpha \boldsymbol{z}) - f(\mathbf{x})}{\alpha} &= \lim_{\alpha \to 0} \max_{i \in I(\mathbf{x})} \frac{f_i(\mathbf{x} + \alpha \boldsymbol{z}) - f_i(\mathbf{x})}{\alpha} \\
&= \max_{i \in I(\mathbf{x})} \lim_{\alpha \to 0} \frac{f_i(\mathbf{x} + \alpha \boldsymbol{z}) - f_i(\mathbf{x})}{\alpha} \\
&= \max_{i \in I(\mathbf{x})} f_i'(\mathbf{x}; \boldsymbol{z})
\end{aligned}
$$

□

3.2 Subgradient

Definition 3.2. Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper function and let $\mathbf{x} \in dom(f)$. A vector $g \in \mathbb{E}^*$ is a subgradient of f at x if

$$\forall\, \mathbf{y} \in \mathbb{E} : \underbrace{f(\mathbf{y}) \geqslant f(\mathbf{x}) + \langle g, \mathbf{y} - \mathbf{x} \rangle}_{\text{subgradient inequality}}$$

The set of all subgradient of f at $\mathbf{x}$ is called the subdifferential of f at $\mathbf{x}$ and denoted by $\partial f(\mathbf{x})$. If $\partial f(\mathbf{x}) \neq \varnothing$, we say that f is subdifferentiable at $\mathbf{x}$.

$$\partial f(\mathbf{x}) = \{g \in \mathbb{E}^* : f(\mathbf{y}) \geqslant f(\mathbf{x}) + \langle g, \mathbf{y} - \mathbf{x} \rangle, \quad \forall\, \mathbf{y} \in \mathbb{E}\}$$

Example 3.1. 1. Let $f : \mathbb{E} \to \mathbb{R}$ be defined by

$$f(\mathbf{x}) = ||\mathbf{x}||, \qquad \text{where } ||\cdot|| \text{ is the norm at } \mathbb{E}$$

Then:

$$\partial f(\vec{0}) = B_{\|\cdot\|_*}[0, 1] = \{g \in \mathbb{E}^* : \|g\|_* \leqslant 1\}$$

Proof. By the def of a subgradient and subdifferential, $g \in \partial f(\vec{0})$ if and only if

$$\begin{aligned} \forall \mathbf{y} \in \mathbb{E} :\ & f(\mathbf{y}) \geqslant f(\vec{0}) + \langle g, \mathbf{y} \rangle \\ \iff & \langle g, \mathbf{y} \rangle \leqslant \|\mathbf{y}\| \end{aligned}$$

($\Longrightarrow$) want to show: $\|g\|_* \leqslant 1 \Longrightarrow \langle g, \mathbf{y} \rangle \leqslant \|\mathbf{y}\|$

($\Longleftarrow$) want to show: $\|g\|_* \leqslant 1 \Longleftarrow \langle g, \mathbf{y} \rangle \leqslant \|\mathbf{y}\|$ $\square$