

Notes of CS 839: Advanced Nonlinear Optimization

Instructor: Jelena Diakonikolas

YI WEI

Sep 2024

Contents

1	Vector Space	1
1.1	Cartesian Product of Vector Space	2
1.2	Linear Transformation	2
1.3	The Dual Space	3
1.4	Adjoint Transformation	4
2	Extended Real-Valued Functions	4
2.1	Closed Functions	5
2.2	Related Concepts	6
2.3	Operations preserving closedness	6
2.4	Closedness vs Continuity	6
2.5	Convex Function	9
2.5.1	Infimal Convolution	11

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 \mathbb{R}^m , $\|A\|_{a,b} = \max_{x \in \mathbb{R}^n: \|x\|_a \leq 1} \|Ax\|_b$
In particular, if $\|\cdot\|_a$ and $\|\cdot\|_b$ are l_p norms:

- (a) $a = b = 2 \rightarrow$ operator/spectral norm
- (b) $a = b = 1$:

$$\|A\|_{1,1} = \max_{x \in \mathbb{R}^n, \|x\|_1 \leq 1} \|Ax\|_1 \quad (1)$$

$$= \max_{1 \leq j \leq n} \sum_{i=1}^n |A_{ij}| \quad (2)$$

It's called "max abs column sum"

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

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

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

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

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

$$\text{where } \|Ax\|_{\infty} = \begin{bmatrix} A_1 x \\ A_2 x \\ \vdots \\ A_n x \end{bmatrix}$$

1.1 Cartesian Product of Vector Space

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

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

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

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

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

Example 1.2.

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

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

1.2 Linear Transformation

Definition 1.1. Given two vector spaces \mathbb{E}, \mathbb{V} , $f : \mathbb{E} \rightarrow \mathbb{V}$ 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 $\mathbb{R}^n \rightarrow \mathbb{R}^m$ are of the form

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

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

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

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

3. The identity transformation $\mathcal{I} : \mathbb{E} \rightarrow \mathbb{E}$ 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)

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

Theorem 1.1. Generalized Cauchy-Schwarz:

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

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 $\|\cdot\|_{Q^{-1}}$

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

If $Q = \text{diag}(w_1, \dots, w_d)$ for positive w_1, \dots, 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. $E = E_1 \times \dots \times E_m$, with $\|\cdot\|_{E_1}, \dots, \|\cdot\|_{E_m}$

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

Theorem 1.3. Bidual space = dual space to \mathbb{E}^* .

In finite vector space, $\mathbb{E}^{**} = \mathbb{E}$

Theorem 1.4. $\langle A\mathbf{x}, \mathbf{y} \rangle \leq \|A\|_{a,b} \|\mathbf{x}\|_a \|\mathbf{y}\|_b$ 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 $A : \mathbb{E} \rightarrow \mathbb{V}$, the adjoint transformation $A^\top : \mathbb{V}^* \rightarrow \mathbb{E}^*$ 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 $A : \mathbb{E} \rightarrow \mathbb{V}$, $\|\cdot\|_{\mathbb{E}}, \|\cdot\|_{\mathbb{V}}$, we define the norm $\|A\| = \sup_{x \in \mathbb{E}, \|x\|_{\mathbb{E}} \leq 1} \|A(x)\|_{\mathbb{V}}$

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 $\mathbb{R} \cup \{-\infty, +\infty\} \equiv [-\infty, +\infty]$ or $\mathbb{R} \cup \{+\infty\} \equiv (-\infty, +\infty]$

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

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:

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

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

$$\min_{x \in \mathcal{X}} f(x) \iff \min_{x \in \mathbb{E}} f(x) + \delta_{\mathcal{X}}(x)$$

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

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

1. $a + \infty = +\infty + a = +\infty$
2. $a - \infty = -\infty + a = -\infty$
- 3.

$$a \cdot \infty = \begin{cases} \infty, & \text{if } a > 0 \\ -\infty, & \text{if } a < 0 \end{cases}$$

4. $0 \cdot \pm\infty = 0$
5. $-\infty < a < \infty \quad \forall a \in \mathbb{R}$

2.1 Closed Functions

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

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

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

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

□

Remark. f is closed $\iff \text{dom}(f)$ is closed.

Example 2.1.

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

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

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

2.2 Related Concepts

1. Lower Semicontinuity:

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

$$f(x) \leq \liminf_{n \rightarrow \infty} f(x_n)$$

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

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 $\alpha \in \mathbb{R}$, $f : \mathbb{E} \rightarrow [-\infty, +\infty]$.

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

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

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

2.3 Operations preserving closedness

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

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

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

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

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

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

2.4 Closedness vs Continuity

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

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

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

Proof. Argue that $\text{epi}(f)$ is closed.

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

To argue: $(x_*, y_*) \in \text{epi}(f)$: we know that $x_n \in \text{dom}(f)$, $x_n \rightarrow x_*$, $\text{dom}(f)$ is closed $\implies x_* \in \text{dom}(f)$

By the definition of $\text{epi}(f)$:

$$f(x_n) \leq y_n$$

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

$$\begin{aligned} f(x_*) &\leq y_* \\ \implies (x_*, y_*) &\in \text{epi}(f) \end{aligned}$$

□

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

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

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

$$Lev(I, \alpha) = \begin{cases} \emptyset, & \alpha < 0 \\ \{0\}, & 0 \leq \alpha < 1 \\ \mathbb{R}, & \alpha \geq 1 \end{cases}$$

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

[Date: Sep 16, 2024]

Theorem 2.4 (Weierstrass theorem for closed functions). Let $f : \mathbb{E} \rightarrow (-\infty, \infty]$ be a $\underbrace{\text{proper}}_{\text{dom}(f) \neq \emptyset}$, closed function and let $C \subseteq \mathbb{E}$ be a compact set such that $C \cap \text{dom}(f) \neq \emptyset$. 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 $\{x_n\}_{n \geq 1}$, $x_n \in C \forall n$, s.t.

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

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

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

We know

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

Contradiction.

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

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

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

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

$$f(\bar{x}) \leq \lim_{k \rightarrow \infty} \inf f(x_{n_k}) = f_*$$

$$\implies \bar{x} \text{ minimizes } f$$

□

What is we are not optimizing over a compact set.

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

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

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

Proof. Let x_0 be an arbitrary point

□

2.5 Convex Function

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

1. $\text{epi}(f)$ is convex
2. $\forall x, y \in \mathbb{E}, \forall \alpha \in (0, 1)$:

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

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

$$f\left(\sum_{i=1}^m \lambda_i x_i\right) \leq \sum_{i=1}^m \lambda_i f(x_i)$$

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

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

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

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

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

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

3. I : inded set, $f_i : \mathbb{E} \rightarrow (-\infty, \infty]$ 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 $f : \mathbb{E} \times \mathbb{V} \rightarrow (-\infty, \infty]$ be a convex function s.t. $\forall x \in \mathbb{E}, \exists y \in \mathbb{V} : f(x, y) < \infty$. Let $g : \mathbb{E} \rightarrow [-\infty, \infty)$ be defined

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

Then g is convex.

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

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

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

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

f is convex so:

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

Then by the definition of g , we have:

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

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. $f(x_1, y_1) \leq M$. And $\exists y_2$ s.t. $f(x_2, y_2) < \infty$. Since f is convex

$$f((1 - \alpha)x_1 + \alpha x_2, (1 - \alpha)y_1 + \alpha y_2)$$

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

Then by the definition of g , we have

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

M is arbitrary.

□

2.5.1 Infimal Convolution

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

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