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

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 $Q = \text{diag}(w_1, \ldots, w_d)$ 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 $\|\cdot\|_{E_1}, \ldots, \|\cdot\|_{E_m}$

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

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 $\not\Longleftarrow\!\!\Longrightarrow$ $dom(f)$ 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:

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

2.2 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, $\forall \alpha \in \mathbb{R}$

2.3 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

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

2.4 Closedness vs Continuity

Bottom line: If f has closed domain + continuous over the domain $\Longrightarrow$ closed.
But closed $\not\Longleftrightarrow$ 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. $\Longrightarrow$ the sum of them is closed.