Introduction to Nonlinear Optimization

Lecture 01

August 26, 2021

Beck, sections 1.1-1.3

Math hieroglyphs

$\forall$: "for all"

$\exists$: "exists"

$!$: "unique"

$f(x)$

$\forall \varepsilon > 0 \ \exists \delta > 0$ such that

$|x - y| \leq \delta \implies |f(x) - f(y)| \leq \varepsilon$

Number sets

$\mathbb{N} = \{1, 2, 3, \ldots\}$, $\mathbb{N}_0 = W = \{0, 1, 2, \ldots\}$

$\mathbb{Z} = \{\ldots -3, -2, -1, 0, 1, 2, 3, \ldots\}$

$\mathbb{R}$: real number line

Given $n \in \mathbb{N}$, $\mathbb{R}^n$ is n-dim. Euclidean space

$\mathbb{R}^n_+ = \{\underline{x} \in \mathbb{R}^n \mid x_j \geq 0 \quad \forall \ j = 1 .. n\}$,

$\underline{x} = (x_1, x_2 \ldots x_n)$

$$\mathbb{R}^n_{++} = \{ \underline{x} \in \mathbb{R}^n \mid x_j > 0 \quad \forall\, j = 1 \ldots n \}$$

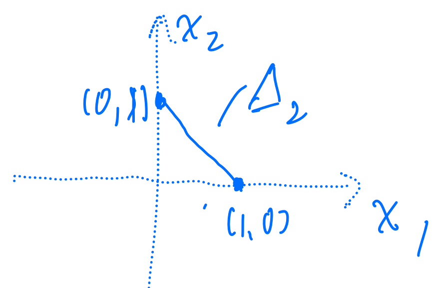
$\mathbb{R}^n_{+}$ and $\mathbb{R}^n_{++}$ are called non-negative and positive orthants, respectively.

Δ_n : unit simplex

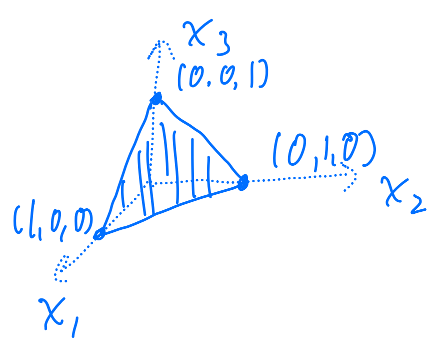
$$\Delta_n = \{ \underline{x} \in \mathbb{R}^n \mid \sum_{i=1}^{n} x_i = 1, \; x_i \geq 0 \;\; \forall\, i = 1, \ldots n \}$$

Ex: $n = 1 \implies \Delta_1 = \{1\}$

$n = 2 \implies \Delta_2$

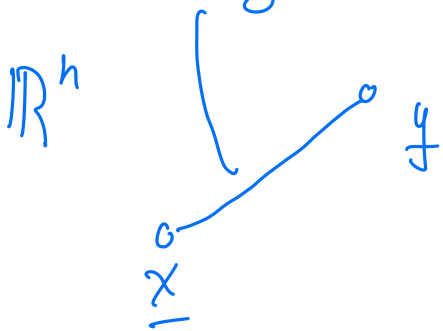


$n = 3 \implies \Delta_3$

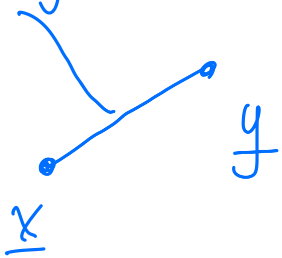


Line segments

$\underline{x}, \underline{y} \in \mathbb{R}^n$: $\quad (\underline{x}, \underline{y}) = \{ t\underline{x} + (1-t)\underline{y} \mid t \in (0,1) \}$



$$[\underline{x}, \underline{y}] = \{ t\underline{x} + (1-t)\underline{y} \mid t \in [0,1] \}$$



$\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$

Matrix: $\underline{\underline{A}} \in \mathbb{R}^{m\times n}$

conceptually, $\mathbb{R}^{m\times n} = \mathbb{R}^{mn}$

On $\mathbb{R}^n$, or $\mathbb{R}^{m\times n}$, addition and scalar multiplication are defined elementwise

$$\lambda \underline{x} = \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}$$

Matrix multiplication: $\underline{\underline{A}} \in \mathbb{R}^{m \times p}$, $\underline{\underline{B}} \in \mathbb{R}^{p \times n}$

$$\underline{\underline{A}}\,\underline{\underline{B}} \in \mathbb{R}^{m \times n}$$

$$(\underline{\underline{A}}\,\underline{\underline{B}})_{i,j} = \sum_{k=1}^{p} (\underline{\underline{A}})_{i,k} (\underline{\underline{B}})_{k,j}$$

Inner products

Definition: A map $\langle \cdot , \cdot \rangle$ from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ is an inner product if

(i) positive-definite: $\langle \underline{x}, \underline{x} \rangle \geq 0 \;\; \forall \; \underline{x} \in \mathbb{R}^n$ and

$\langle \underline{x}, \underline{x} \rangle = 0$ ~~only if~~ iff $\underline{x} = 0$

(ii) homogeneous: $\forall \lambda \in \mathbb{R}, \; \forall \underline{y}, \underline{x} \in \mathbb{R}^n$, then

$\langle \lambda \underline{x}, \underline{y} \rangle = \lambda \langle \underline{x}, \underline{y} \rangle$

(iii) additive: $\underline{x}, \underline{y}, \underline{z} \in \mathbb{R}^n$, then

$$\langle \underline{x}+\underline{y}, \underline{z} \rangle = \langle \underline{x}, \underline{z} \rangle + \langle \underline{y}, \underline{z} \rangle$$

(iv) symmetry: $\langle \underline{x}, \underline{y} \rangle = \langle \underline{y}, \underline{x} \rangle \quad \forall\, \underline{x}, \underline{y} \in \mathbb{R}^n$.

<u>Ex</u>: $\mathbb{R}^n$,

"dot product": $\langle \underline{x}, \underline{y} \rangle = \sum_{i=1}^{n} x_i y_i$

<u>Ex</u>: $\mathbb{R}^n$

$\underline{w} \in \mathbb{R}^n_{++}$

$$\langle \underline{x}, \underline{y} \rangle_{\underline{w}} = \sum_{i=1}^{n} w_i x_i y_i$$

__Ex__: $\mathbb{R}^2$

$$\langle \underline{x}, \underline{y} \rangle := 2(x_1+x_2)(y_1+y_2) + (x_1-x_2)(y_1-y_2)$$

Notation: dot product $\langle \underline{x}, \underline{y} \rangle = \underline{x}^T \underline{y}$

Inner products yield geometric structure on $\mathbb{R}^n$.

__Definition__: $\underline{x}, \underline{y} \in \mathbb{R}^n$ equipped with a(ny) inner product $\langle \cdot, \cdot \rangle$ are __orthogonal__ if $\langle \underline{x}, \underline{y} \rangle = 0$ (and $\underline{x}, \underline{y} \neq \underline{0}$)

Vector space norms

Defn: A norm $\|\cdot\|$ on $\mathbb{R}^n$ is a function satisfying

(i) positivity: $\|\underline{x}\| \geq 0 \quad \forall\, \underline{x} \in \mathbb{R}^n$ and $\|\underline{x}\| = 0$ iff $\underline{x} = 0$.

(ii) homogeneity: $\|\lambda \underline{x}\| = |\lambda| \cdot \|\underline{x}\| \quad \forall\, \lambda \in \mathbb{R},\ \underline{x} \in \mathbb{R}^n$

(iii) triangle inequality: $\|\underline{x} + \underline{y}\| \leq \|\underline{x}\| + \|\underline{y}\|$
$\forall\, \underline{x}, \underline{y} \in \mathbb{R}^n$

Ex: Norms defined through inner products:

if $\langle \cdot, \cdot \rangle$ is an inner product on $\mathbb{R}^n$, then $\|\underline{x}\| = \sqrt{\langle \underline{x}, \underline{x} \rangle}$ is a norm.

Ex. (ℓ^p norms) $\mathbb{R}^n$

$$\|\underline{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \text{ is a norm \underline{if} } p \geq 1$$

if $p < 1$, this formula is <u>not</u> a norm.

Ex. If $\langle \cdot, \cdot \rangle$ is the dot product, then the norm $\|\cdot\|$ induced by this inner product is $\|\cdot\|_2$.

Ex If $p = \infty$, $\|\underline{x}\|_\infty = \max_{i=1..n} |x_i|$

Matrix norms – induced and entrywise

$\underline{\underline{A}} \in \mathbb{R}^{m\times n}$, the definition of a norm $\|\underline{\underline{A}}\|$ is the same as for vectors (positivity, homogeneity, triangle inequality).

There are <u>many</u> kinds of matrix norms

<u>Ex</u>. "Entrywise norms"

"Frobenius" norm: $\|\underline{\underline{A}}\|_F^2 = \sum_{i,j} |(\underline{\underline{A}})_{ij}|^2$

<u>Ex</u> "Induced" norms

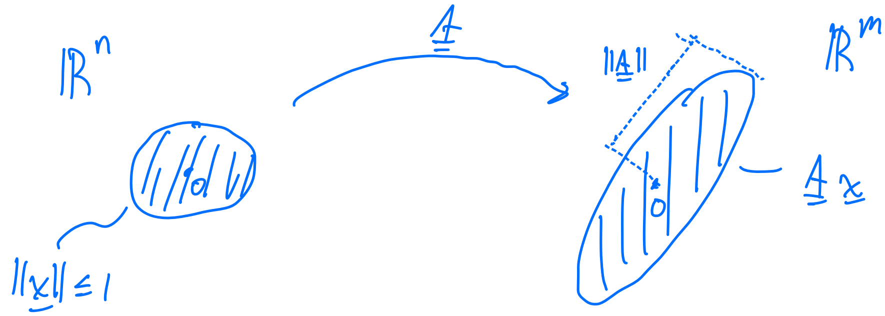
On $\mathbb{R}^{m\times n}$, we will derive a matrix norm from a norm on $\mathbb{R}^m$ and a norm on $\mathbb{R}^n$.

Given $p, q \geq 1$

$$\|\underline{\underline{A}}\|_{p,q} = \max_{\underline{x} : \|\underline{x}\|_p \leq 1} \|\underline{\underline{A}}\,\underline{x}\|_q$$

This norm measure how much $\underline{\underline{A}}$ "inflates" vectors.

If $p = q = 2$: $\|\underline{\underline{A}}\|_{2,2} = \|\underline{\underline{A}}\|_2 = \max_{\underline{x} : \|\underline{x}\|_2 \leq 1} \|\underline{\underline{A}}\,\underline{x}\|_2$



$\|\underline{\underline{A}}\|_2$ is difficult to compute.

Turns out: $\|\underline{\underline{A}}\|_2 = \sqrt{\lambda_{max}(\underline{\underline{A}}^T \underline{\underline{A}})}$

max. eigenvalue.

Cauchy-Schwarz inequality: for any inner product and its derived norm, then $|\langle \underline{x}, \underline{y} \rangle| \leq \|\underline{x}\| \cdot \|\underline{y}\|$

Ex: dot product

$$\underline{x}^T \underline{y} = \|\underline{x}\| \cdot \|\underline{y}\| \cos\theta \qquad \theta: \text{angle between them.}$$

$$\frac{|\underline{x}^T \underline{y}|}{\|\underline{x}\| \cdot \|\underline{y}\|} = |\cos\theta| \leq 1$$

Cauchy-Schwarz is "transparent".

Orthogonal matrices / "Unitary" matrices

Definition: $\underline{U} \in \mathbb{R}^{n \times n}$ is orthogonal if the columns of $\underline{\underline{U}}$ are orthonormal (with respect to the dot product)

Ex. $\underline{\underline{U}} = \underline{\underline{I}} = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & \ddots & \\ & & & & 1 \end{pmatrix}$

Implications of $\underline{\underline{U}}$ being orthogonal

- $\underline{\underline{U}}^T \underline{\underline{U}} = \underline{\underline{I}}$
- $\underline{\underline{U}}^{-1} = \underline{\underline{U}}^T$
- rows of $\underline{\underline{U}}$ are (also) orthonormal

- columns of $\underline{U}$ are an orthonormal basis for $\mathbb{R}^n$.