Revision of norms & orthogonal matrices

For $x \in \mathbb{R}^n$

$\|x\|$ = a measure of the "length of x"

many definitions possible – all satisfying :

(6.1) **Norm axioms** — chosen so $\|x\|$ behaves like length

$\forall\, x, z \in \mathbb{R}^n,\ \forall\, \alpha \in \mathbb{R}$:

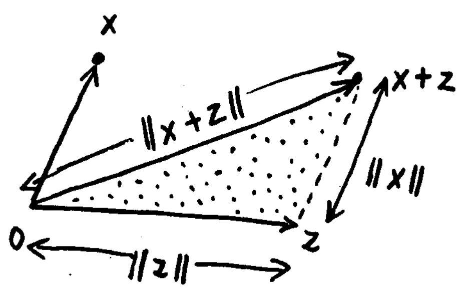
Axiom I : $\|x\| > 0$

Axiom II : $[\,\|x\| = 0\,] \Leftrightarrow [\,x = 0\,]$

Axiom III : $\|\alpha x\| = |\alpha| \cdot \|x\|$

Axiom IV : $\|x+z\| \leq \|x\| + \|z\|$

called triangle inequality -based on



Examples of norms on $\mathbb{R}^n$

1. $\|x\|_1 \triangleq \sum_{i=1}^{n} |x_i|$
2. $\|x\|_2 \triangleq \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$ ← Euclidean norm
3. $\|x\|_\infty \triangleq \max\{|x_1|, |x_2|, \dots, |x_n|\}$

In this lecture always : $\|x\| = \|x\|_2$

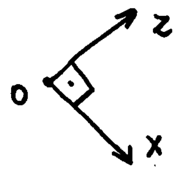
so : $\|x\|^2 = \sum_{i=1}^{n} x_i^2 = x_1^2 + \dots + x_n^2$

$= (x_1, x_2, \dots, x_n)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x^T x$

$\Rightarrow \|x\| = (x^T x)^{1/2}$

Orthogonal vectors :

In $\mathbb{R}^2$:



$x \perp z$ ← orthogonal

In $\mathbb{R}^n$?

(6.2) $[x \perp z] \underset{\text{by def.}}{\Longleftrightarrow} [\,\|x+z\| = \|x-z\|\,]$

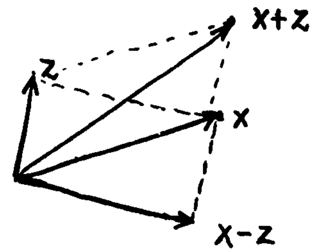
equivalently:

(6.3) $[x \perp z] \underset{\text{by def.}}{\Longleftrightarrow} [\,x^T z = 0\,]$

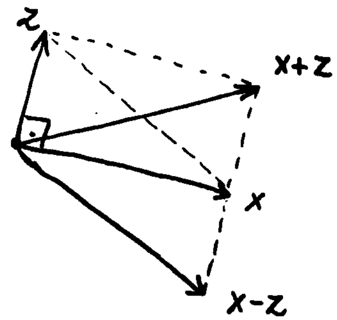
$\boxed{\text{why this definition ?}}$

because it is a key property of orthogonality in $\mathbb{R}^2$

$x \not\perp z$



$\|x+z\| \neq \|x-z\|$



$\|x+z\| = \|x-z\|$

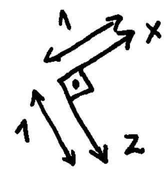
Equivalence of (6.2) & (6.3):

$$\|x+z\| = \|x-z\| \leftarrow (6.2)$$
$$\Updownarrow$$
$$\|x+z\|^2 = \|x-z\|^2$$
$$\Updownarrow$$
$$(x+z)^T(x+z) = (x-z)^T(x-z)$$
$$\| \qquad\qquad \|$$
$$x^Tx + \underbrace{z^Tx}_{x^Tz} + x^Tz + z^Tz = x^Tx - \underbrace{z^Tx}_{x^Tz} - x^Tz + z^Tz$$
$$\Updownarrow$$
$$2x^Tz = -2x^Tz$$
$$\Updownarrow$$
$$x^Tz = 0 \leftarrow (6.3).$$

QED.

(6.4) Definitions

1. $x \in \mathbb{R}^n$ is called normalized iff $\|x\| = 1$
2. $x, z \in \mathbb{R}^n$ are called orthonormall iff:
   (a) $x \perp z$
   (b) $\|x\| = \|z\| = 1$



③ $x^1, x^2, \dots, x^k \in \mathbb{R}^n$ are called orthonormal iff:

(a) $x^i \perp x^j, \ \forall\, i,j \ (i \neq j)$

(b) $\|x^i\| = 1, \ \forall\, i$

e.g. the standard basis vectors for $\mathbb{R}^3$:

$$e_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad e_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \quad e_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}$$

<u>Orthogonal matrices</u> (6.5)

very important for numerical algorithms

$P \in \mathbb{R}^{n \times n}$ is called orthogonal iff: its columns are orthonormal.

<u>Some properties of $\perp$ P:</u> (6.6)

① $P^T P = P P^T = I \quad [\Rightarrow P^T = P^{-1}]$

② $\|Px\| = \|x\|$

because $\|Px\|^2 = (Px)^T Px = x^T \underbrace{P^T P}_{= I} x = x^T x = \|x\|^2$

The Cauchy-Schwartz inequality

For $x, y \in \mathbb{R}^n$:

(6.7) $|x^T y| \leq \|x\|_2 \|y\|_2$

equality only when x & y point in the same direction

$\Rightarrow x^T y \leq \|x\|_2 \|y\|_2$

<u>Proof</u>: Obvious if $x = y = 0$, so, say $x \neq y$.

Minimize $\|x + \alpha y\|^2$ with respect to $\alpha \in \mathbb{R}$.

Since $\|x + \alpha y\|^2 = (x + \alpha y)^T (x + \alpha y)$

$= \|x\|^2 + 2\alpha x^T y + \alpha^2 \|y\|^2$

the minimizing α is:

$$\hat{\alpha} = -\frac{x^T y}{\|y\|^2}, \text{ and then}$$

$$0 \leq \|x + \hat{\alpha} y\|^2 = \|x\|^2 - \frac{(x^T y)^2}{\|y\|^2}$$

so it is ≥ 0

$\Rightarrow (x^T y)^2 \leq \|x\|^2 \|y\|^2$

$\Rightarrow |x^T y| \leq \|x\| \|y\|$ QED

Linear least-squared error problems:

(formulation + background leading to Moore-Penrose pseudo-inverse)

Consider

$$\underbrace{y}_{\boxed{\text{output}}} = \underbrace{A}_{\in \mathbb{R}^{n\times n}} \underbrace{x}_{\boxed{\text{input}}}$$

$\boxed{\text{viewed as a transfer function of some non-dynamic system}}$

A control problem:

Given a desired output $y_d \in \mathbb{R}^n$

what makes $Ax \cong y_d$ $\boxed{\text{as near as possible}}$

If $\exists A^{-1} \Rightarrow$ required $x = A^{-1} y_d$

$$\Downarrow$$

$$\{ Ax = y_d \} \quad \boxed{\text{exactly}}$$

But what if $A \neq$ square ? ($\Rightarrow$ when $\nexists A^{-1}$)?

In general:

suitable x can be found by solving:

$$\min_{x \in \mathbb{R}^n} \underbrace{\| y_d - Ax \|}_{\boxed{\text{distance of } Ax \text{ from } y_d}}$$

An obvious approach:

$$\min \| \cdots \| \iff \min \| \cdots \|^2$$

so: take $V(x) = \| y_d - Ax \|^2$

$$= (y_d - Ax)^T (y_d - Ax)$$

$$\Rightarrow \quad V(x) = \underbrace{\| y_d \|^2}_{a} + \underbrace{(-2A^T y_d)^T}_{b} x + \frac{1}{2} x^T \underbrace{[2A^T A]}_{C} x$$

But does $C^T = C$?

$$C^T = (2A^T A)^T \underset{\boxed{(AB)^T = B^T A^T}}{=} 2A^T \underbrace{(A^T)^T}_{A} = C \quad \boxed{\text{YES}}$$

Is $C > 0$?

$$x^T C x = 2\underbrace{x^T A^T}_{(Ax)^T} A x = 2\|Ax\|^2 > 0,\ \forall\, x \neq 0$$

iff: $Ax \neq 0, \forall x \neq 0$

$\Rightarrow$ $C > 0$ iff $Ax \neq 0, \forall x \neq 0$

So, suppose $Ax \neq 0, \forall x \neq 0$

Then $V(x) = \|y_d - Ax\|^2$

$\|$

a case of our standard quadratic

$$\Rightarrow \hat{x} = -C^{-1} b = \underbrace{(A^T A)^{-1} A^T y_d}_{\text{looks nice but ...}}$$

Difficulty 1: not always true that

$Ax \neq 0, \forall x \neq 0$

Difficulty 2: evaluating $A^T A$ very susceptible to errors caused by finite precision computing.

$\Rightarrow$ a better minimization method is desirable

– based on orthogonal matrices.

$\downarrow$

A geometrical interpretation

Set of all possible outputs $= \{Ax : x \in \mathbb{R}^n\}$

$\triangleq \mathcal{R}[A] \subset \mathbb{R}^m$

general case
A not square

the range of A

Computation of Ax

we know: $(Ax)_i = \sum_{k=1}^{n} a_{ik} x_k$

Also:

column formula

$$Ax = \sum_{k=1}^{n} a_{*k} x_k$$

$\|$

$[a_{*1}, \dots, a_{*n}]$

column 1 of A

row formula

$$Ax = \begin{bmatrix} a_{1*} x \\ \vdots \\ a_{m*} x \end{bmatrix}$$

$\|$

$\begin{bmatrix} a_{1*} \\ \vdots \\ a_{m*} \end{bmatrix}$ $\leftarrow$ row 1 of A

So:

$$\mathcal{R}[A] = \{Ax : x \in \mathbb{R}^n\}$$

$$= \left\{ \sum_{k=1}^{n} a_{*k} x_k : x_k \in \mathbb{R}, \forall k \right\}$$

$\Rightarrow \mathcal{R}[A] =$ set of all linear combinations of $a_{*1}, \dots, a_{*n}$

$= \mathcal{L}[a_{*1}, \dots, a_{*n}]$ ← the linear space spanned by the columns of A

Ideally would like an $x \in \mathbb{R}^n$ to exist such that $y_d = Ax$ which happens iff

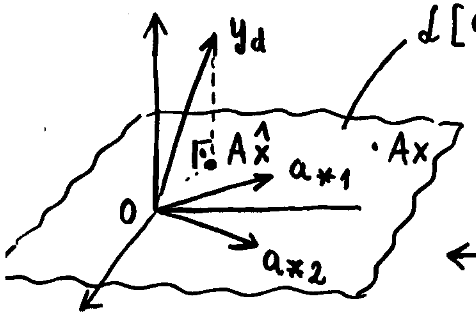
$y_d =$ a possible system output, i.e. iff

$y_d \in \mathcal{R}[A]$ – set of all possible outputs

$\Downarrow$ then

$$\min_{x \in \mathbb{R}^n} \| y_d - Ax \|^2 = 0$$

<u>Example:</u> $(m = 3, n = 2)$



obvious geometrically that $y_d - A\hat{x}$ is $\perp$ to $\mathcal{R}[A]$, i.e is $\perp$ to any vector in $\mathcal{R}[A]$.

Actually, for a linear subspace like the range of A, the set of all vectors which are orthogonal to every vector in $\mathcal{R}[A]$ is called the <u>orthogonal complement</u> of the range of A, denoted $^{\perp}\mathcal{R}[A]$

Symbolically:

$$^{\perp}\mathcal{R}[A] = \{ z \in \mathbb{R}^n : z^T y = 0, \forall y \in \mathcal{R}[A] \}$$

Hence the fact that

$$(y_d - A\hat{x}) \perp \mathcal{R}[A]$$

can be expressed more neatly by:

$$(y_d - A\hat{x}) \in {}^{\perp}\mathcal{R}[A]$$

Algebraic proof that :

$\hat{x}$ minimizes $\|y_d - Ax\|^2$ on $\mathbb{R}^n$

if $(y_d - A\hat{x}) \perp \mathcal{R}[A]$ (even if $y_d \in \mathcal{R}[A]$)

$$\|y_d - Ax\|^2 = \|y_d - A(\hat{x} + x - \hat{x})\|^2$$

$$= \|(y_d - A\hat{x}) - A(x - \hat{x})\|^2$$

$$= \|y_d - A\hat{x}\|^2 - \underbrace{2(y_d - A\hat{x})^T \underbrace{A\underbrace{(x - \hat{x})}_{\in \mathbb{R}^n}}_{\in \mathcal{R}[A]}}_{=0} + \underbrace{\|A(x - \hat{x})\|^2}_{\geqslant 0}$$

$$\geqslant \|y_d - A\hat{x}\|^2 , \ \forall x \in \mathbb{R}^n$$

$\Rightarrow$ $\hat{x}$ is a global minimizer if $(y_d - A\hat{x}) \perp \mathcal{R}[A]$.

QED