Norm Space

Linear Algebra

Department of Computer Engineering

Sharif University of Technology

Hamid R. Rabiee rabiee@sharif.edu

Maryam Ramezani maryam.ramezani@sharif.edu

P-norm

Vector Norms

- p−norm:

$$\|x\|_p = (|x_1|^p + |x_2|^p + \ \dots + |x_n|^p)^{\frac{1}{p}}$$

$$\textit{subject to } p \geq 1$$

- What is the shape of $\|x\|_p = 1$?
- Properties?

Norm

Definition (Norm)

- A function $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ is a norm if
  1. $f(x) \geq 0$, $f(x) = 0 \Leftrightarrow x = 0$ (positivity)
  2. $f(\alpha x) = |\alpha| f(x)$, $\forall \alpha \in \mathbb{R}$ (homogeneity)
  3. $f(x + y) \leq f(x) + f(y)$ (triangle inequality)

1-norm and 2-norm

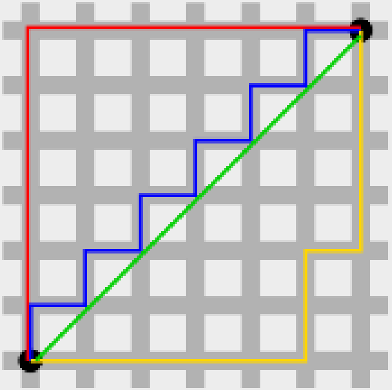
- 1−norm(l_1):

$$\|x\|_1 = (|x_1| + |x_2| + \ ... + |x_n|)$$

- What is the shape of $\|x\|_1 = 1$?
- The distance between two vectors under the l_1 norm is also referred to as the Manhattan Distance.
- Properties?

Example

l_1 distance between $(0, 1)$ and $(1, 0)$?



Norm Derivations

- Square of l_2

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \\ \dots \\ \\ x_n \end{bmatrix}$$

$$\|x\|_2^2 = x_1^2 + x_2^2 + \cdots + x_n^2$$

$$\Rightarrow \begin{cases} \dfrac{d\|x\|_2^2}{dx_1} = 2x_1 \\ \dfrac{d\|x\|_2^2}{dx_2} = 2x_2 \\ \\ \dots \\ \\ \dfrac{d\|x\|_2^2}{dx_n} = 2x_n \end{cases}$$

Norm Derivations

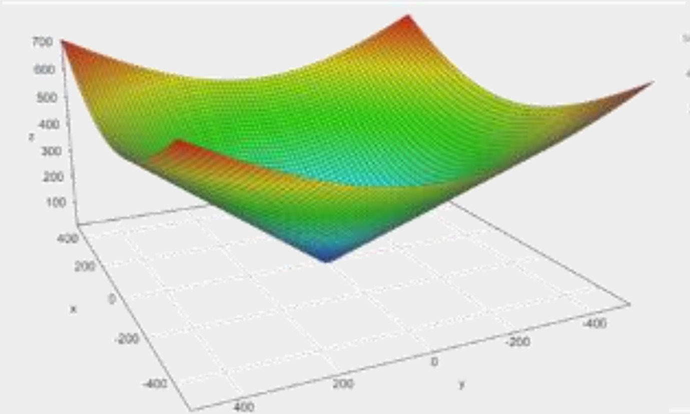
- l_2

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \qquad \|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = (x_1^2 + x_2^2 + \cdots + x_n^2)^{\frac{1}{2}}$$

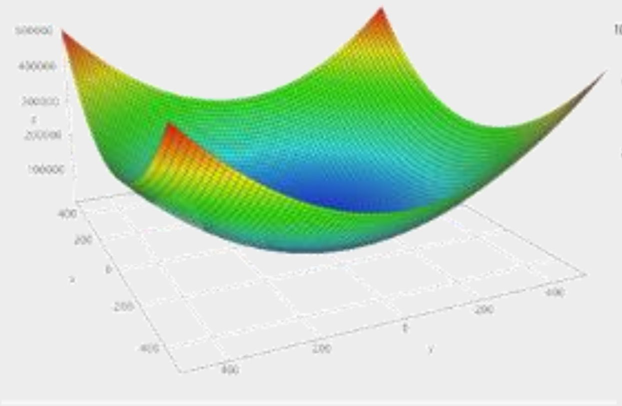
$$\begin{aligned}
\frac{d\|x\|_2}{dx_1} &= \frac{1}{2}(x_1^2 + x_2^2 + \ \dots + x_n^2)^{\frac{1}{2}-1} . \frac{d}{dx_1}(x_1^2 + x_2^2 + \ \dots + x_n^2) \\
&= \frac{1}{2}(x_1^2 + x_2^2 + \ \dots + x_n^2)^{-\frac{1}{2}} . \frac{d}{dx_1}(x_1^2 + x_2^2 + \ \dots + x_n^2) \\
&= \frac{1}{2} . \frac{1}{(x_1^2 + x_2^2 + \ \dots + x_n^2)^{\frac{1}{2}}} . \frac{d}{dx_1}(x_1^2 + x_2^2 + \ \dots + x_n^2) \\
&= \frac{1}{2} . \frac{1}{(x_1^2 + x_2^2 + \ \dots + x_n^2)^{\frac{1}{2}}} . 2 . x_1 \\
&= \frac{x_1}{(x_1^2 + x_2^2 + \ \dots + x_n^2)^{\frac{1}{2}}}
\end{aligned}$$

$$\Rightarrow \begin{cases} \dfrac{d\|x\|_2}{dx_1} = \dfrac{x_1}{(x_1^2 + x_2^2 + \ \dots + x_n^2)^{\frac{1}{2}}} \\ \dfrac{d\|x\|_2}{dx_2} = \dfrac{x_2}{(x_1^2 + x_2^2 + \ \dots + x_n^2)^{\frac{1}{2}}} \\ \qquad \dots \\ \dfrac{d\|x\|_2}{dx_n} = \dfrac{x_n}{(x_1^2 + x_2^2 + \ \dots + x_n^2)^{\frac{1}{2}}} \end{cases}$$

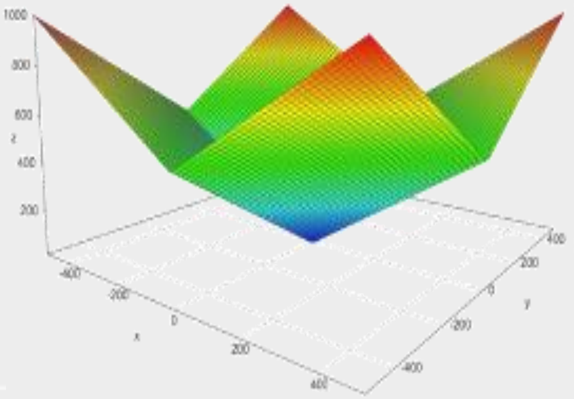
Norm Comparisons



l_2 norm

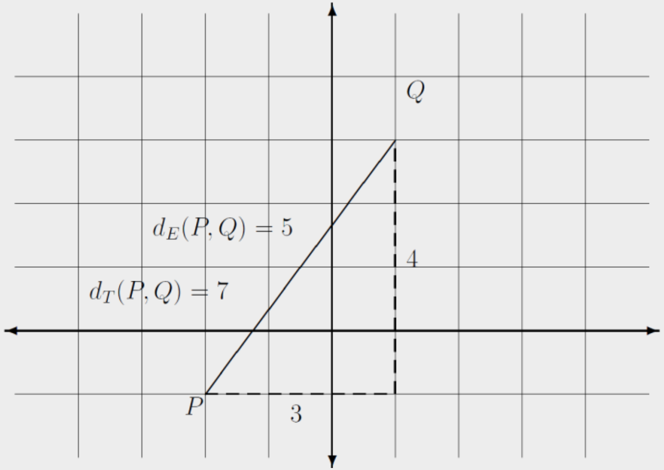


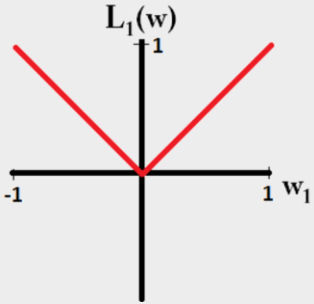
Square l_2 norm

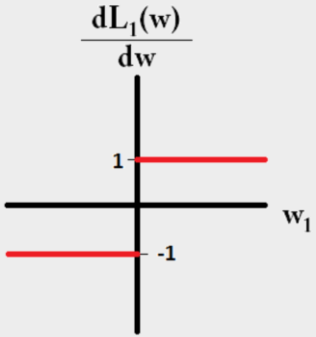


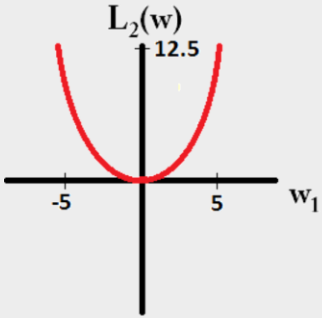
l_1 norm

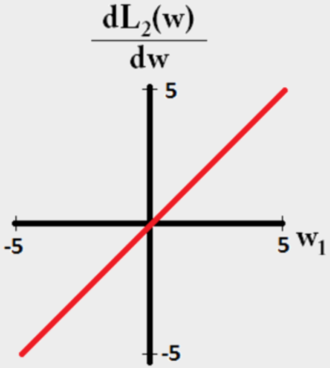
L1 and L2 norm comparisons









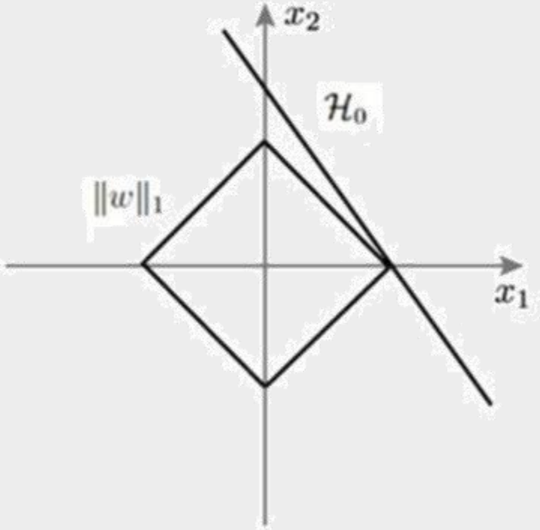


L1 and L2 norm comparisons

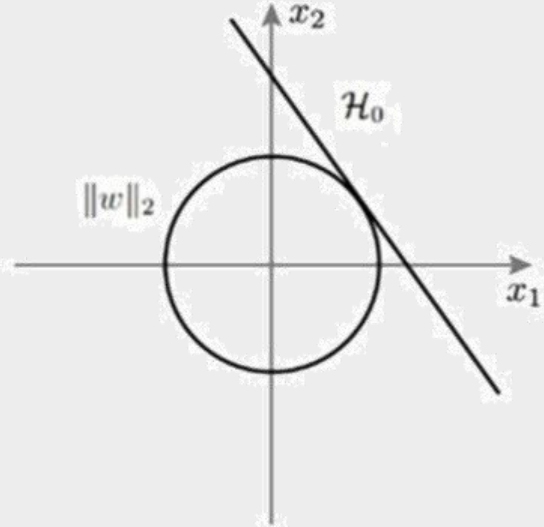
- Robustness is defined as resistance to outliers in a dataset. The more able a model is to ignore extreme values in the data, the more robust it is.
- Stability is defined as resistance to horizontal adjustments. This is the perpendicular opposite of robustness.
- Computational difficulty
- Sparsity

Why is l_1 supposed to lead to sparsity than l_2?

$$\min_x \|x\|_{1\ or\ 2}\,,$$
$$subject\ to\ Ax = b$$



l_1 reqularization



l_2 reqularization

$\infty-$norm

- ∞−norm(l_∞)(max norm):

$$l_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|)$$

- What is the shape of $|x|_\infty = 1$?
- Properties?

$\frac{1}{2}-$norm

Vector Norms

- $\frac{1}{2}$–norm$(l_{\frac{1}{2}})$

- What is the shape of $|x|_{\frac{1}{2}} = 1$?

- Properties?

Vector Norms

- 0−norm(l_0):

$$\|x\|_0 = \lim_{\alpha \to 0^+} \|x\|_\alpha = \left(\sum_{k=1}^{n} |x|^\alpha \right)^{\frac{1}{\alpha}} = \sum_{k=1}^{n} 1_{(0,\infty)}(|x|)$$

- 0-norm, defined as **the number of non-zero elements in a vector**, is an ideal quantity for feature selection. However, minimization of 0-norm is generally regarded as a combinatorially difficult optimization

- $\|x\|_0 = \sum_{x_i \neq 0} 1$

- Is 0−norm a valid norm?

- What is the shape of $||x||_0 = 1$?

Examples

- l_0 distance between $(0, 0)$ and $(0, 5)$?
- l_0 distance between $(1, 1)$ and $(2, 2)$?
- (username, password)

Class Activity

- l_0 distance between $(0, 0)$ and $(0, 5)$?
- l_0 distance between $(1, 1)$ and $(2, 2)$?
- (username, password)

Or go to the below link
https://forms.gle/xFHSDKJDq1KoL4Kx6

Timer: (2:30 minutes)

Vector Norms

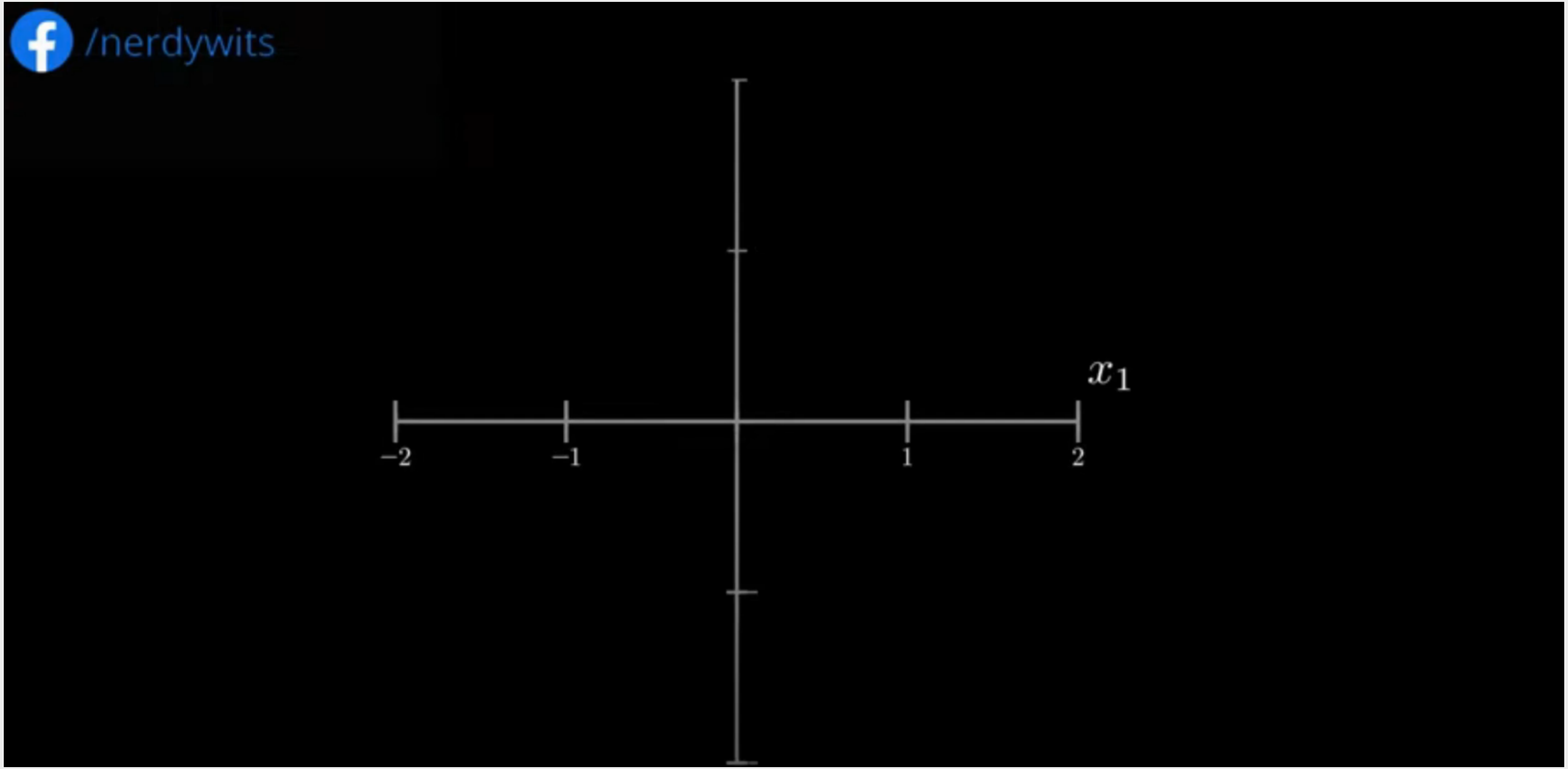
Examples

- l_0 distance between $(0, 0)$ and $(0, 5)$?
- l_0 distance between $(1, 1)$ and $(2, 2)$?
- (username, password)

Solution

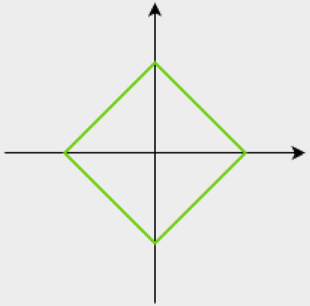
- 1
- 2
- When l_0 is 0, then we can infer that username and password is a match and we can authenticate the user.

Vector Norms Shapes

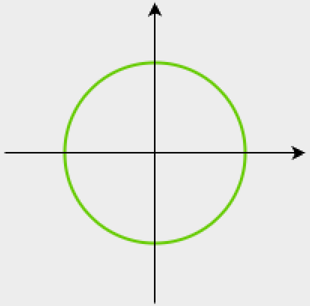


Norms and Convexity

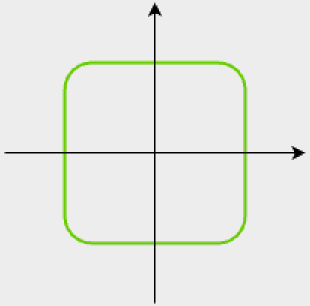
- For $\boldsymbol{p \geq 1}$, $\boldsymbol{l_p}$ norm is convex



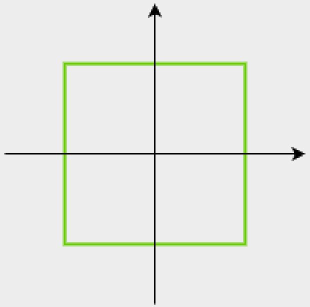
$\|x\|_1 = 1$



$\|x\|_2 = 1$



$\|x\|_p = 1$



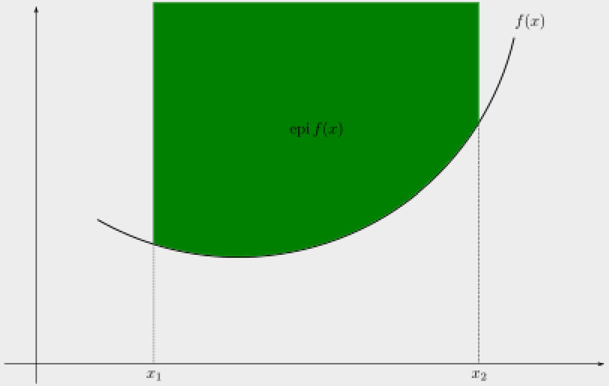
$\|x\|_\infty = 1$

Convex function

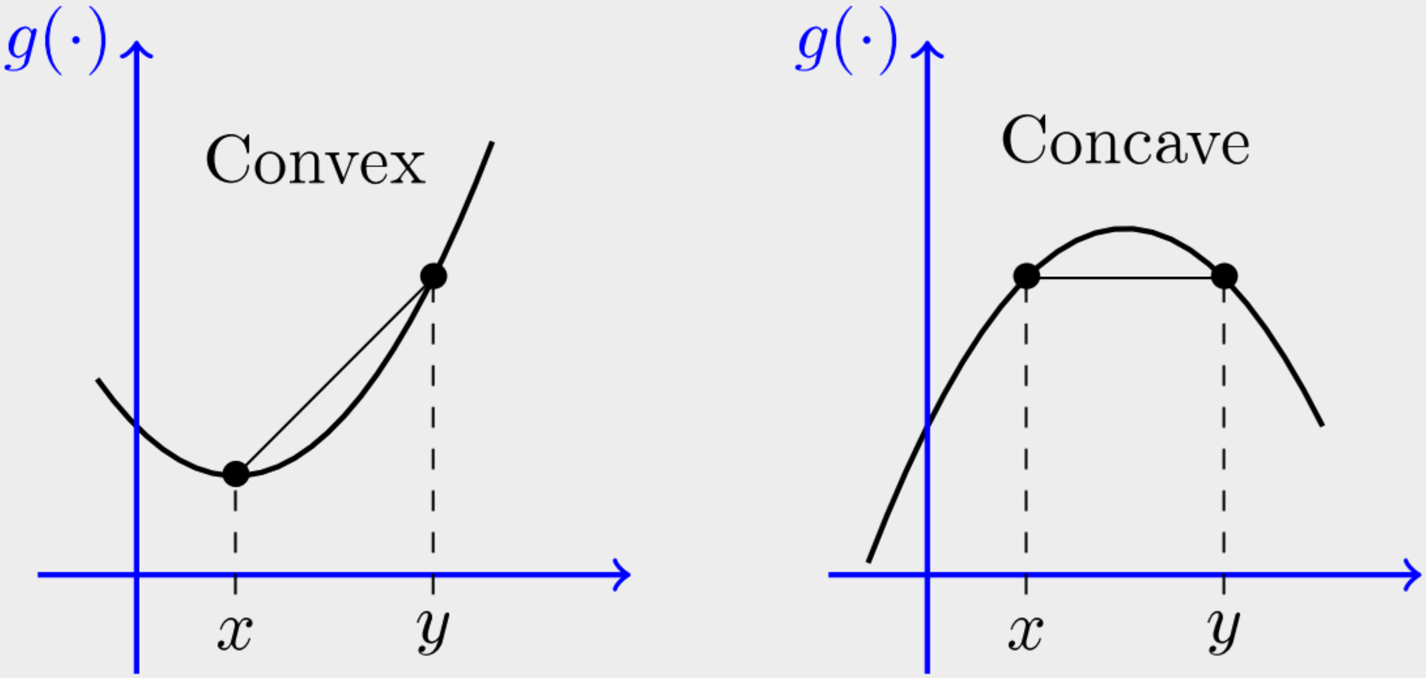
- A function is convex iff its epigraph is a convex set.
- Epigraph or supergraph

$$\mathrm{epi} f = \{(x, \mu) \,:\, x \in \mathbb{R}^n,\, \mu \in \mathbb{R},\, \mu \geq f(x)\} \subseteq \mathbb{R}^{n+1}$$

$$f\left((1-\theta)x^{(0)} + \theta x^{(1)}\right) \leq (1-\theta) f\left(x^{(0)}\right) + \theta f\left(x^{(1)}\right), \quad \forall \theta \in [0, 1]$$

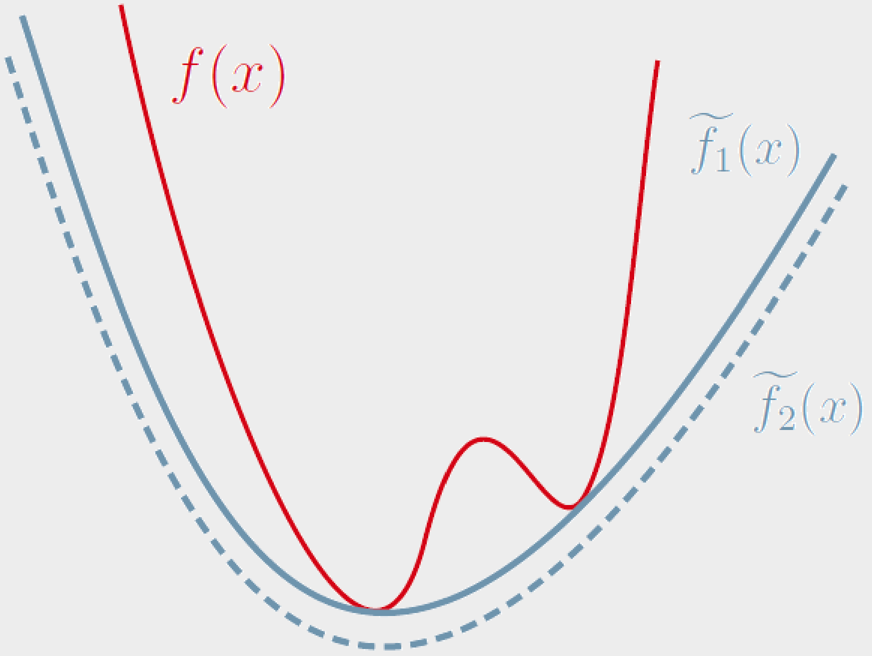


Convex and Concave Function



second derivative is nonnegative on its entire domain

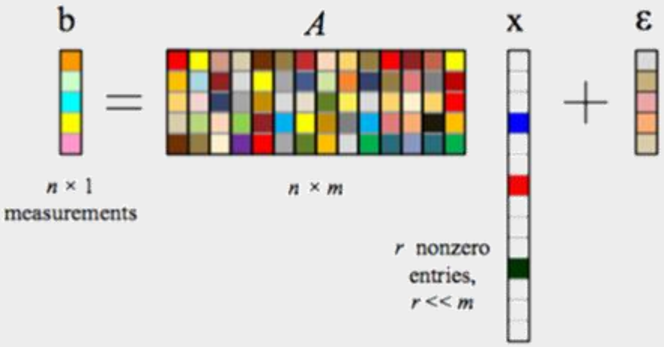
Convex Relaxation



- **Alternative viewpoint:** We try to find the sparsest solution which explains our noisy measurements

$$\min_{x} \|x\|_0 , \qquad subject\ to\ \|Ax - b\|_2 < \epsilon$$

- Here, the l_0-norm is a shorthand notation for counting the number of non-zero elements in x.



Sparse Solution

- l_0 optimization is np−hard.
- Convex relaxation for solving the problem.

$$\min_{1} \|x\|_1$$

$$subject\ to\ \|Ax\ - b\|_2 < \epsilon$$

$$\min_{1} \|x\|_0$$

$$subject\ to\ \|Ax\ - b\|_2 < \epsilon$$

Theorem

For all $x \in \mathbb{R}^d$:

$$||x||_2 \leq ||x||_1 \leq \sqrt{d}||x||_2$$

Proof

$$\sum_i |x_i| \sum_i |x_i| = \sum_i x_i^2 + \sum_{i \neq j} |x_i|\,|x_j|$$

Max norm inequality

Theorem

For all $x \in \mathbb{R}^d$:

$$||x||_\infty \leq ||x||_1 \leq d||x||_\infty$$

$$||x||_\infty \leq ||x||_2 \leq \sqrt{d}||x||_\infty$$

Proof

Conclusion

- By a normed linear space (briefly normed space) is meant a real or complex vector space E in which every vector x is associated with a real number $|x|$, called its absolute value or norm, in such a manner **that the properties** $(a') - (c')$ holds. That is, for any vectors $x, y \subset E$ and scalar α we have:

i. $|x| \geq 0$

ii. $|x| = 0 \; iif \; x = \vec{0}$

iii. $|\alpha x| = |\alpha||x|$

iv. $|x + y| \leq |x| + |y|$

Inner product and norm

Theorem

Take any inner product $\langle \cdot, \cdot \rangle$ and define $f(x) = \sqrt{\langle x, x \rangle}$. Then f is a norm.

Proof

Note

Every inner product gives rise to a norm, but not every norm comes from an inner product. (Think about norm 2 and norm max)

Entry-wise matrix norms

Definition

$$\|A\|_{p,p} = \|vec(A)\|_p = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^p\right)^{\frac{1}{p}}$$

Special Cases

- Frobenius (Euclidian, Hilbert Schmidt) norm:(p = 2)

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}} = \sqrt{trace(A^*A)}$$

- Max norm ($p = \infty$)

$$\|A\|_{max} = \max_{ij}|a_{ij}|$$

- Sum-absolute-value norm

$$\|A\|_{sav} = \sum_{i,j}|A_{i,j}|$$

Theorem

- Invariant under rotations (**unitary operations = orthogonal matrices**)

$$\|A\|_F = \|AU\|_F = \|UA\|_F$$
$$\|A+B\|_F^2 = \|A\|_F^2 + \|B\|_F^2 + 2\langle A, B\rangle$$
$$\|A^*A\|_F = \|AA^*\|_F \le \|A\|_F^2$$

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}} = \sqrt{trace(A^*A)}$$

Frobenius (Euclidean norm)

Theorem

Let $b_1, b_2, \cdots, b_n$ denote the columns of B. Then

$$\|AB\|_{HS}^2 = \sum_{i=1}^{n} \|Ab_i\|^2 \leq \sum_{i=1}^{n} \|A\|^2 \|b_i\|^2 = \|A\|^2 \|B\|_{HS}^2$$

Using Cauchy-Schawrtz Inequality

Definition

$$\|A\|_p = \max_{\vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|_p}{\|\vec{x}\|_p} = \max_{\|\vec{x}\|_p=1} \|A\vec{x}\|_p$$

Theorem

1. $\|Ax\| \leq \|A\|\|x\|$ for all vectors $\|x\|$
2. For all matrices A, B: $\|AB\| \leq \|A\|\|B\|$

Matrix norms induced by vector norms

Definition

- The norm of a matrix is a real number which is a measure of the magnitude of the matrix.
- Norm 1:

$$\|A\|_1 = \max_{1\leq j\leq n}\left(\sum_{i=1}^{n}|a_{ij}|\right)$$

- Norm max:

$$\|A\|_\infty = \max_{1\leq i\leq n}\left(\sum_{j=1}^{n}|a_{ij}|\right)$$

Example

$$B = \begin{bmatrix} 5 & -4 & 2 \\ -1 & 2 & 3 \\ -2 & 1 & 0 \end{bmatrix}$$

SVD and Norm

- One common definition for the norm of a matrix is the Frobenius norm:

$$||A||_F^2 = \sum_{i=1:m} \sum_{j=1:n} a_{ij}^2$$

- Frobenius norm can be computed from SVD

$$||A||_{\mathrm{F}}^2 = \sum_{i=1:p} \Sigma_i{}^2 \quad where\, p = \min(n, m)$$

- So changes to a matrix can be evaluated by looking at changes to singular values

Theorem

1) Orthogonal matrices, they preserve the Euclidean norm

2) $$\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1$$

Norms Compare

The 2-norm (spectral norm) of a matrix is the greatest distortion of the unit circle/sphere/hyper-sphere. It corresponds to the largest singular value (or |eigenvalue| if the matrix is symmetric/hermitian).

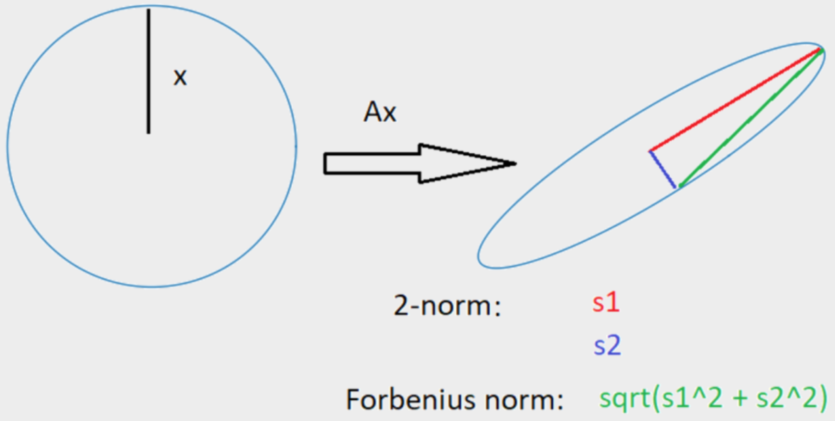
The Forbenius norm is the "diagonal" between all the singular values.

i.e.

$$||A||_2 = s_1 \quad , \quad ||A||_F = \sqrt{s_1^2 + s_2^2 + \ldots + s_r^2}$$

(r being the rank of A).

Here's a 2D version of it: x is any vector on the unit circle. Ax is the deformation of all those vectors. The length of the red line is the 2-norm (biggest singular value). And the length of the green line is the Forbenius norm (diagonal).



References

- Linear Algebra and Its Applications, David C. Lay
- Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares
- https://www.youtube.com/watch?v=76B5cMEZA4Y