Norm Space

CE40282-1: Linear Algebra
Hamid R. Rabiee and Maryam Ramezani
Sharif University of Technology

Vector Norms

- p-norm:

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

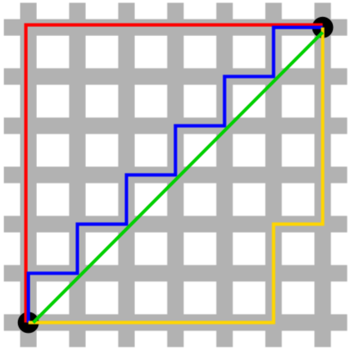
- What is the shape of $||x||_p = 1$?

Vector Norms

- 1-norm: $(\boldsymbol{l_1})$

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

- What is the shape of $||x||_1 = 1$?
- The distance between two vectors under the L1 norm is also referred to as the Manhattan distance
- Example:
  - L1 distance between (0,1) and (1,0)?



Vector Norms

- ∞-norm: ($\boldsymbol{l}_\infty$) (max norm)

$$L_\infty = max(|x_1|, |x_2|, \dots, |x_n|)$$

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

Vector Norms

- $\mathbf{\frac{1}{2}}$-norm: $(\boldsymbol{l}_{\frac{1}{2}})$

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

Vector Norms

- $\boldsymbol{zero}$-norm: $(\boldsymbol{l_0})$

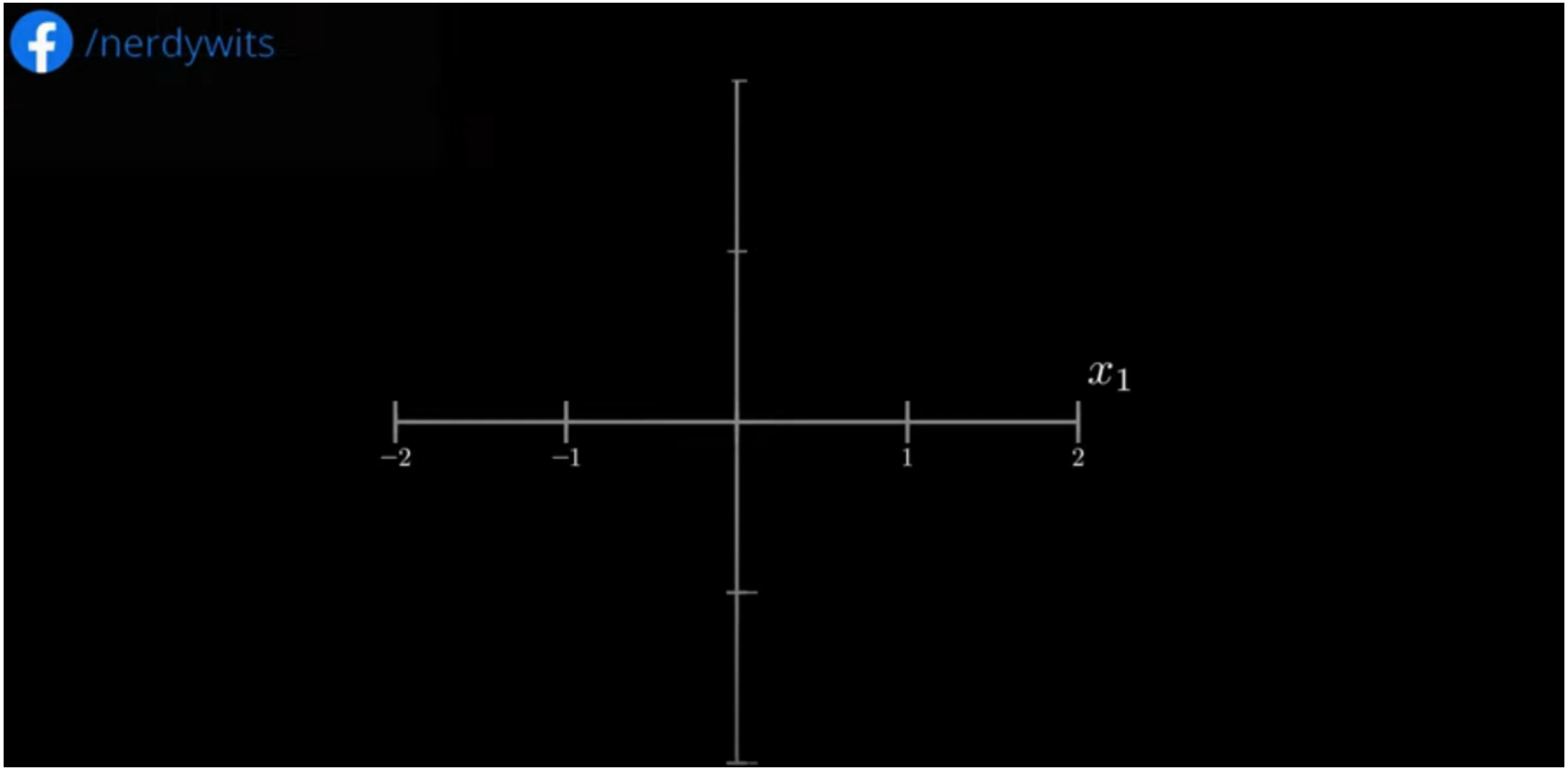
$$\|x\|_0 = \lim_{\alpha \to 0^+} \|x\|_\alpha = \left(\sum_{k=1}^{n} |x|^\alpha \right)^{1/\alpha} = \sum_{k=1}^{n} 1_{(0,\infty)}(|x|)$$

- Zero-norm, defined as **the number of non-zero elements in a vector**, is an ideal quantity for feature selection. However, minimization of zero-norm is generally regarded as a combinatorically difficult optimization
- $||x||_0 = \sum_{x_i \neq 0} 1$

Vector Norms

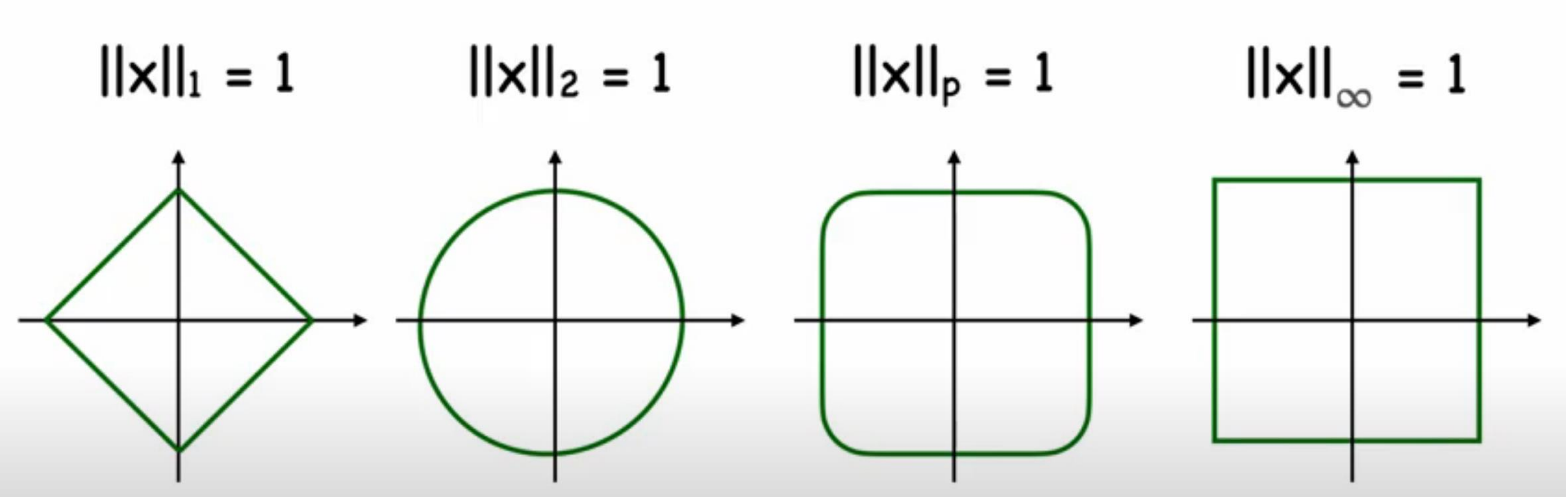
- Is zero-norm a norm??
- What is the shape of $||x||_0 = 1$?
- Examples:
  - L0 distance between (0,0) and (0,5)?
  - L0 distance between (1,1) and (2,2)?
  - (username,password)

Vector Norms Shapes



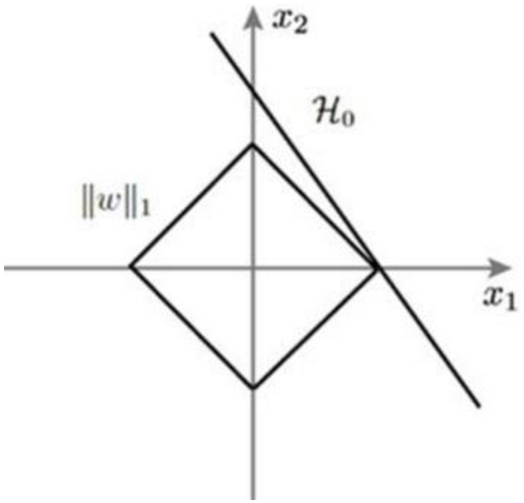
Norms and Convexity

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



Norms and Convexity

***Theorem:* If A is a convex subset of a normed linear space B whose norm is strictly convex, then, for every f in B, there exists a unique best approximation a* in A to f**



Norm Derivations

- Square of l_2

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{bmatrix} \quad ||u||_2 = u_1^2 + u_2^2 + \cdots + u_n^2$$

$$\begin{cases} \dfrac{d||u||_2}{du_1} = 2u_1 \\ \dfrac{d||u||_2}{du_2} = 2u_2 \\ \cdots \\ \dfrac{d||u||_2}{du_n} = 2u_n \end{cases}$$

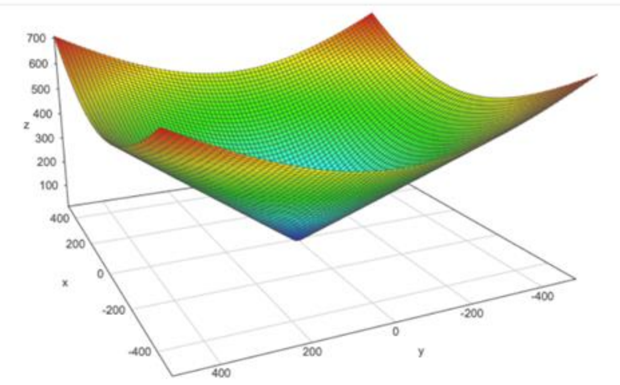
- l_2

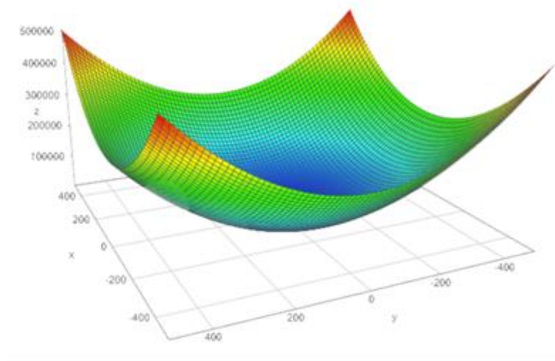
$$||u||_2 = \sqrt{(u_1^2 + u_2^2 + \cdots + u_n^2)} = (u_1^2 + u_2^2 + \cdots + u_n^2)^{\frac{1}{2}}$$

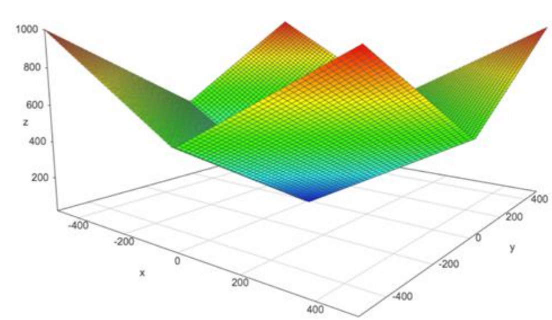
$$\begin{aligned} \frac{d||u||_2}{du_1} &= \frac{1}{2}(u_1^2 + u_2^2 + \cdots + u_n^2)^{\frac{1}{2}-1} \cdot \frac{d}{du_1}(u_1^2 + u_2^2 + \cdots + u_n^2) \\ &= \frac{1}{2}(u_1^2 + u_2^2 + \cdots + u_n^2)^{-\frac{1}{2}} \cdot \frac{d}{du_1}(u_1^2 + u_2^2 + \cdots + u_n^2) \\ &= \frac{1}{2} \cdot \frac{1}{(u_1^2 + u_2^2 + \cdots + u_n^2)^{\frac{1}{2}}} \cdot \frac{d}{du_1}(u_1^2 + u_2^2 + \cdots + u_n^2) \\ &= \frac{1}{2} \cdot \frac{1}{(u_1^2 + u_2^2 + \cdots + u_n^2)^{\frac{1}{2}}} \cdot 2 \cdot u_1 \\ &= \frac{u_1}{\sqrt{(u_1^2 + u_2^2 + \cdots + u_n^2)}} \end{aligned}$$

$$\begin{cases} \dfrac{d||u||_2}{du_1} = \dfrac{u_1}{\sqrt{(u_1^2 + u_2^2 + \cdots + u_n^2)}} \\ \dfrac{d||u||_2}{du_2} = \dfrac{u_2}{\sqrt{(u_1^2 + u_2^2 + \cdots + u_n^2)}} \\ \cdots \\ \dfrac{d||u||_2}{du_n} = \dfrac{u_n}{\sqrt{(u_1^2 + u_2^2 + \cdots + u_n^2)}} \end{cases}$$

Norm Comparisons





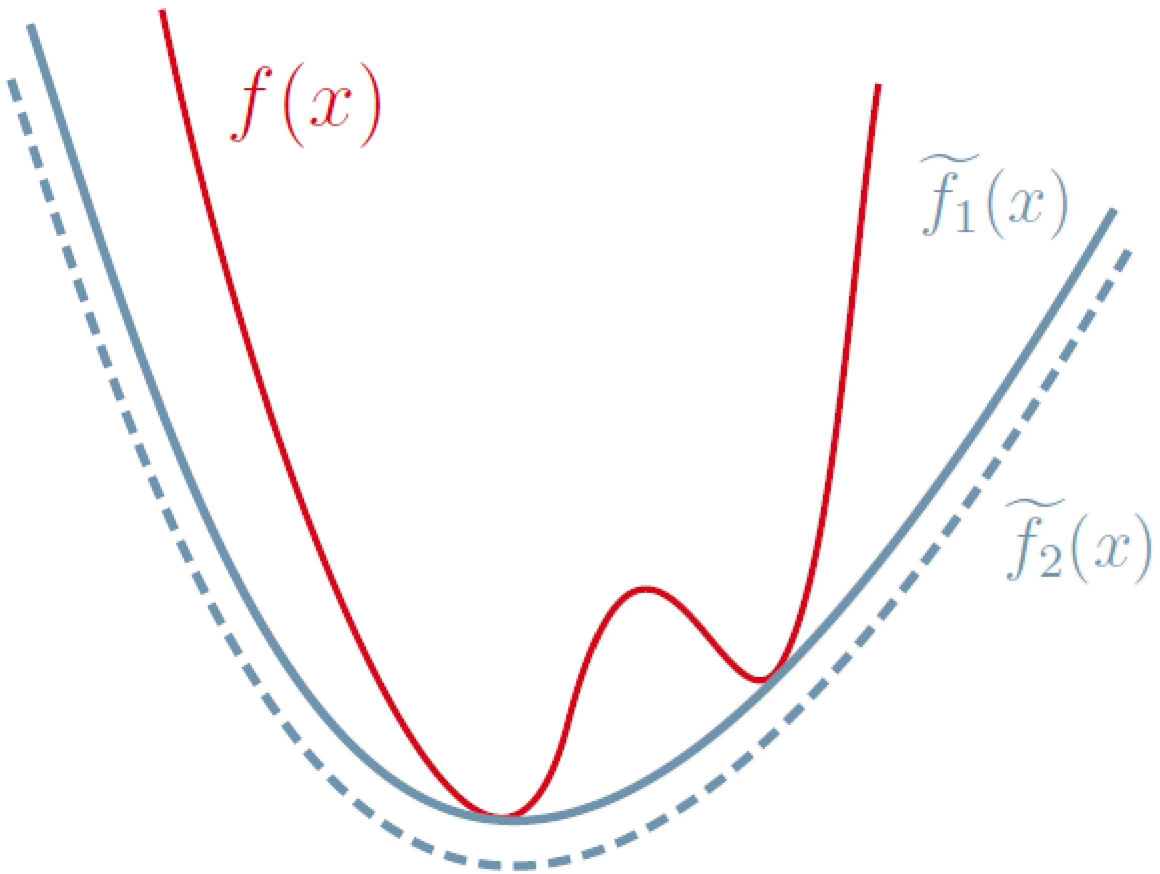


l_2 norm

Square l_2 norm

l_1 norm

Convex Relaxation

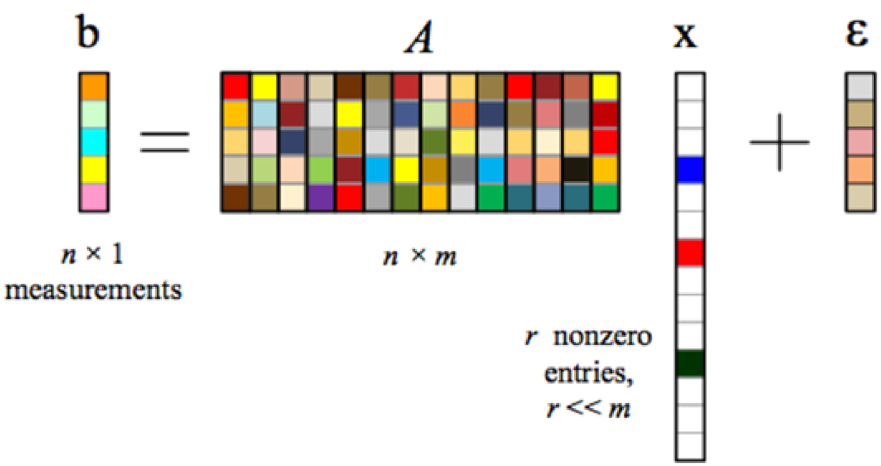


Sparse applications

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

$$\min_{x} \|\mathbf{x}\|_0 \qquad \text{subject to} \;\; \|\mathbf{Ax}-\mathbf{b}\|_2 < \varepsilon$$

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



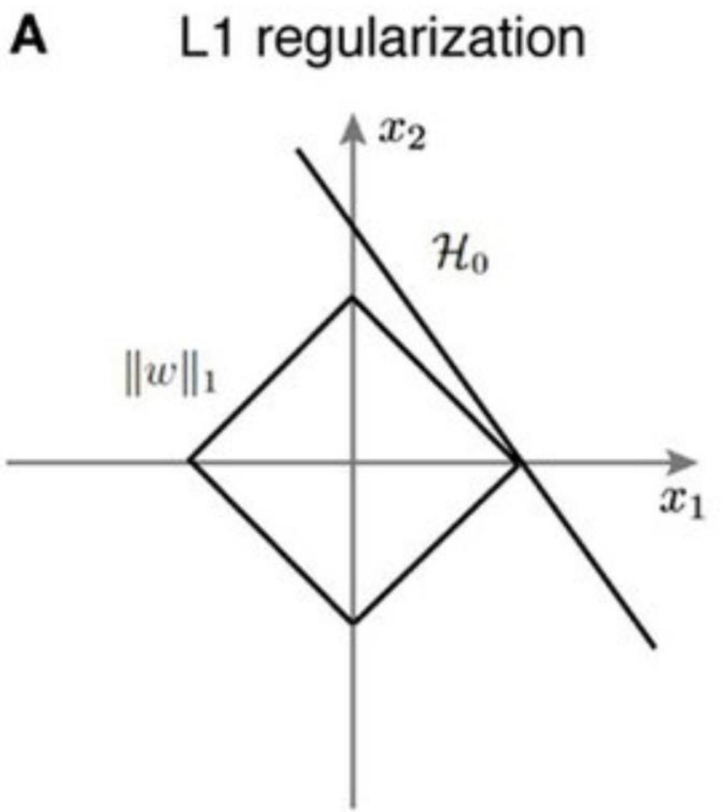
Sparse Solution

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

$$\min_x \| x \|_1 \quad \text{subject to } \| Ax - b \|_2 < \varepsilon$$

$$\min_x \| x \|_0 \quad \text{subject to } \| Ax - b \|_2 < \varepsilon$$

Why is L1 supposed to lead to sparsity than L2?



Vector Norms

- Which norm is the convex hull of the intersection between the L0 norm ball and L2 norm ball?

- Any valid norm ||.|| is a convex function.
  - Proof?

- The L0 norm is not convex.
  - Proof?

Complexity

- Norm: 2n flops. O(n)
- RMS: 2n flops. O(n)
- Distance: 3n flops. O(n)
- Angle: 6n flops. O(n)
- Standard deviation: 4n flops. O(n) can reduce to 3n flops $\mathbf{std}(x)^2 = \mathbf{rms}(x)^2 - \mathbf{avg}(x)^2,$
- Standardizing: 5n flops. O(n)
- Correlation coefficient: 10n flops. O(n)

Reference

- 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