Norms

Tuesday, July 7, 2020 16:28

Norm is a way to measure something
↳ could be vectors, matrices, tensors, functions, etc.

ℓ^p vector norm

$$\|v\|_p = \sqrt[p]{|v_1|^p + \ldots + |v_n|^p}$$

This makes good sense for $0 < p < \infty$

For $p = \infty$ → it picks out the biggest element, so:

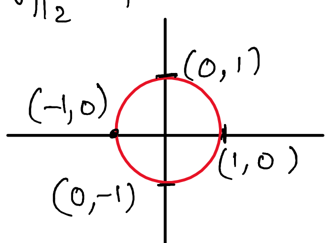
$$\|v\|_\infty = \max_i \{|v_i|\}$$

For $p = 0$ → it gives the number of non-0 elements.
(useful in sparsity questions)
Violates the rule that $\|cv\| = c\|v\|$

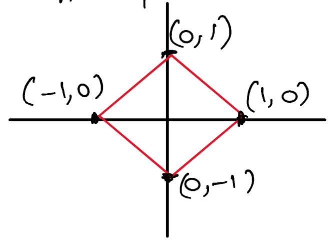
Unit balls in different norms

Consider $\mathbb{R}^2$ — 2D space.

$\|v\|_2 = 1$:

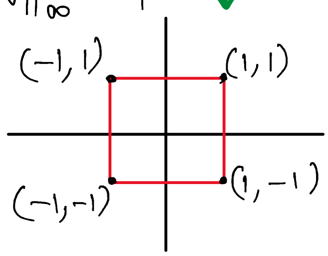


$\|v\|_1 = 1$:



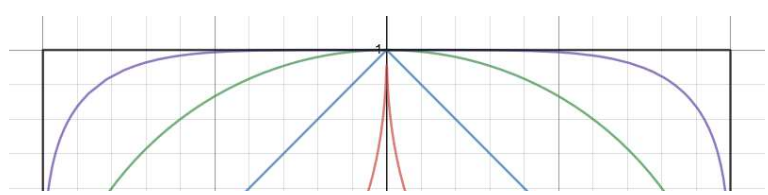
So, the unit ball is a diamond in the ℓ^1 norm.

$\|v\|_\infty = 1$:



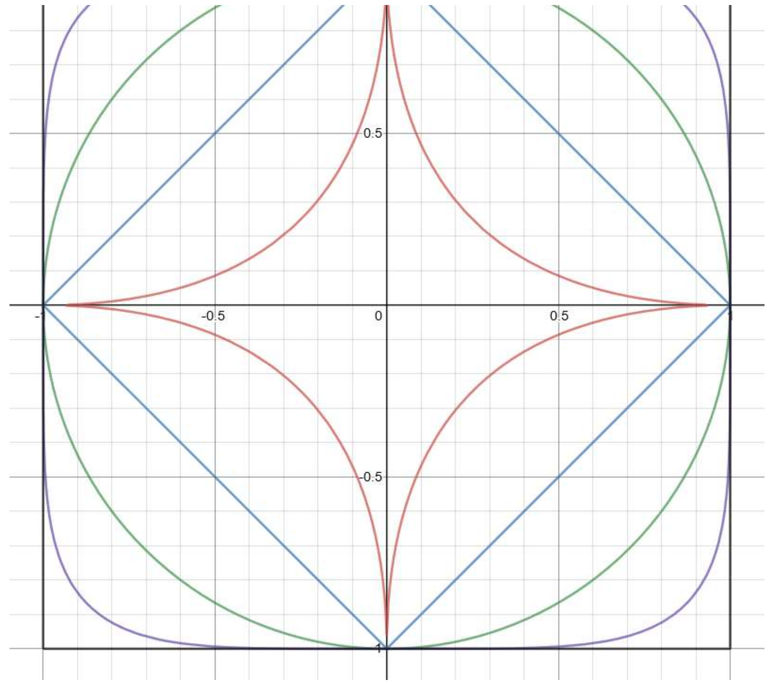
So it kind of swells out from diamond to circle to square as p increases.

p	color
1/2	
1	



1/2	
1	
2	
5	
∞	

Plot of $\|v\|_p = 1$ for $p = \frac{1}{2}, 1, 2, 5, \infty$ in $\mathbb{R}^2$.



→ Why is $p < 1$ not giving us a good norm?

Bcoz a "true" norm has a convex unit ball $\|v\| \leqslant 1$

S norm

For a PD symmetric matrix S and a vector v,

$\|v\|_S = \sqrt{v^T S v}$ (related to the energy in vector v)

Shape of $v^T S v = 1$ (similar to the l^p norms above?)

Note : if $S = I$, this is just the l^2 norm!

e.g. $S = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$, $v^T S v = 2v_1^2 + 3v_2^2 = 1$

↳ an ellipse!

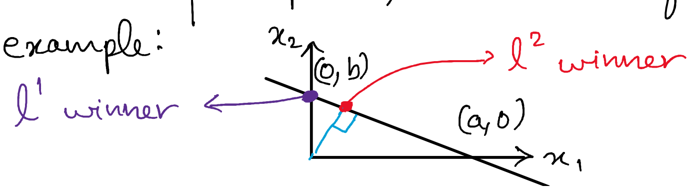
So this norm is kind of a variation of l^2 norm

e.g. problem: minimize $\|x\|_p$ s.t. $Ax = b$

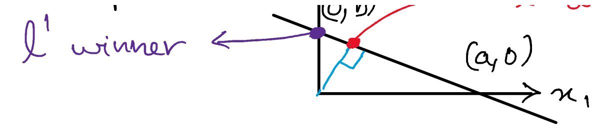
for $p = 1$, it's the Basis Pursuit problem ← as called in optimization

for $p = 2$, it's kind of like the least^2 problem.

2D example:



$\Rightarrow$ l^1 winner is sparse.



$\Rightarrow$ ℓ^1 winner is sparse.

To get the intuition:

→ Imagine a diamond growing out from the center. This diamond will hit $(0,b)$ first among all points on the line.

→ Imagine a circle growing out from the center. This circle will hit the ℓ^2 winner first among all points.

Norms for matrices

1. $$\|A\|_2 = \sigma_1(A) = \frac{\|Ax\|_2}{\|x\|_2}$$ (the blow-up factor)

The max of this ratio over all x gives the $\|A\|_2$ norm of the matrix.

So: $$\|A\|_2 = \max_x \frac{\|Ax\|_2}{\|x\|} = \sigma_1(A)$$

Winner $\Rightarrow$ $x = v_1$ (the first right-singular vector)

Recall that $Av_k = \sigma_k u_k$

↳ Similar to $Av = \lambda v$ eigenvalues, but these are singular values.

$$\Rightarrow \frac{\|Ax\|_2}{\|x\|} = \frac{\|Av_1\|}{\underbrace{\|v_1\|}_{=1}} = \|\sigma_1 u_1\| = \sigma_1$$

2. $$\|A\|_F = \sqrt{\sum a_{ij}^2} \quad \text{(Frobenius norm)}$$
$$= \sqrt{\sum \sigma_i^2}$$

$$A = U\Sigma V^T = \sum \sigma_i u_i v_i^T$$

$$\|A\|_F^2 = \text{Trace}(AA^T) = \text{Trace}(U\Sigma V^T V \Sigma^T U^T) = \text{Trace}(U\Sigma\Sigma^T U^T)$$
$$= \text{Trace}(\Sigma\Sigma^T U U^T) = \text{Trace}(\Sigma\Sigma^T) = \sum \sigma_i^2$$

↳ bcoz $\text{Trace}(AB) = \text{Trace}(BA)$

$= \text{Trace}(\Sigma\Sigma' U U') = \text{Trace}(\Sigma\Sigma') = \sum \sigma_i^2$

↳ bcoz Trace (AB) = Trace (BA)

3. $\|A\|_N$ (Nuclear norm, trace norm, etc)

$= \sum \sigma_i$ → Is being studied in deep-learning gradient descent optimization (look at work by Srebro)