23MAT204

MATHEMATICS FOR INTELLIGENT SYSTEMS - 3

Dr. Sarada Jayan
9448455760
6364333760

Norms of Vectors and Matrices

A way to measure the size of a vector/matrix.

Vector Norms – p norms

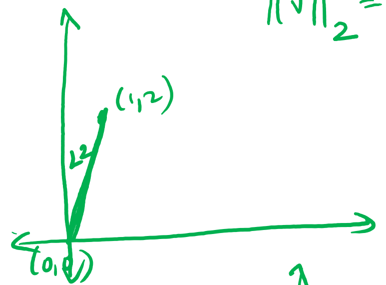
Euclidean Norm $= \|v\|_2 = \sqrt{v_1^2 + v_2^2 + \dots v_n^2}$

if $\bar{v} = (v_1, v_2, \dots v_n)$

Euclidean norm is known as the $L2 / l_2 \mid L_2$ norm.

Find $\|v\|$, $v = [1, 2]$

$\|v\|_2 = \sqrt{1^2 + 2^2} = \sqrt{5}$



Generally, p-norms of a vector.

$$L_p = \|v\|_p = \left[|v_1|^p + |v_2|^p + \dots + |v_n|^p\right]^{1/p}$$

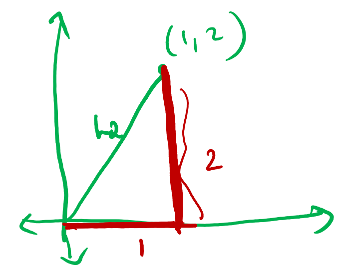
$$L_1 = \|v\|_1 = \left[|v_1| + |v_2| + \dots + |v_n|\right]^{1/1}$$

$$L_\infty = \|v\|_\infty = \lim_{p \to \infty} \left[|v_1|^p + |v_2|^p + \dots + |v_n|^p\right]^{1/p}$$

$$= \max |v_i|$$

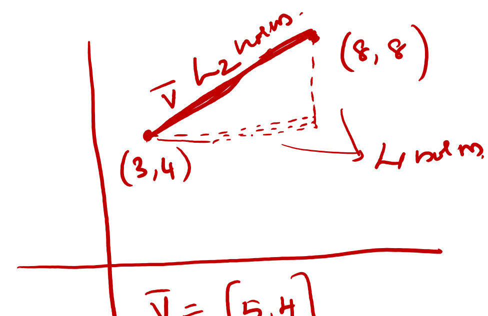
L_{max} norm

L_∞ norm $[1, 2] = 2$



If $\bar{v} = [1, 2]$

$\|v\|_1 =$



$\bar{v} = [5, 4]$

$\|v\|_1 = 5 + 4 = 9$

$\|v\|_2 = \sqrt{41}$

Vector Norms

$\bar{v} = [4, 2, 6, -5, -8]$

$\|v\|_1 = |4| + |2| + |6| + |-5| + |-8| = 25$

$\|v\|_2 = \sqrt{4^2 + 2^2 + 6^2 + (-5)^2 + (-8)^2} = \sqrt{145} = 12.0415$

$\|v\|_3 = \sqrt[3]{4^3 + 2^3 + 6^3 + (-5)^3 + (-8)^3} = (\quad)^{1/3} = 9.7435$

$\|v\|_\infty = \max\{|4|, |2|, |6|, |-5|, |-8|\} = 8$

```
>> v=[4,2,6,-5,-8];
>> nv1=norm(v,1)
nv1 =
   25
>> nv2=norm(v,2)
nv2 =
  12.0416
>> nv3=norm(v,3)
nv3 =
   9.7435
>> nvinf=norm(v,inf)
nvinf =
    8
```

$\bar{v} = (v_1, v_2).$

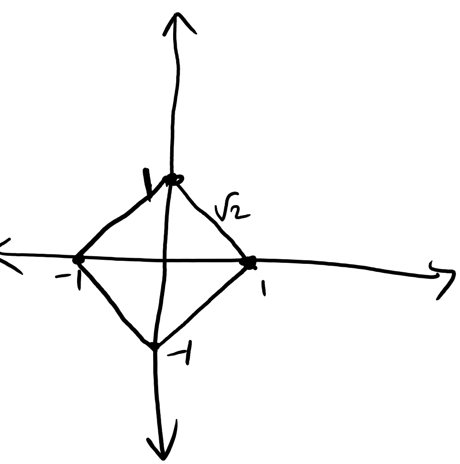
Vector Norms

$\|v\|_1 = 1$

$|v_1| + |v_2| = 1$

$x + y = 1$
$x - y = 1$
$-x + y = 1$
$-x - y = 1$

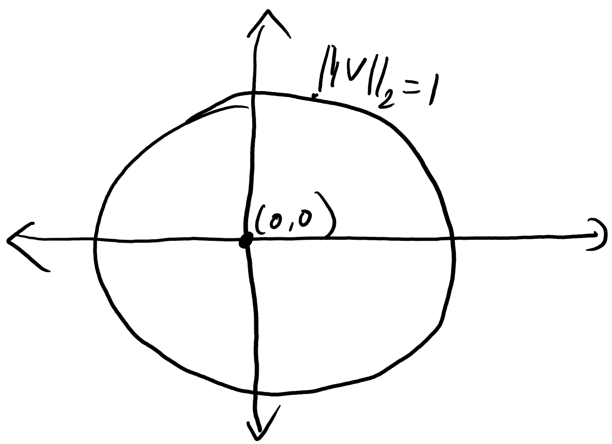
On these lines, $|v_1| + |v_2| = 1$ for $\bar{v} = (v_1, v_2)$



$\|v\|_2 = 1$

$\sqrt{|v_1|^2 + |v_2|^2} = 1$

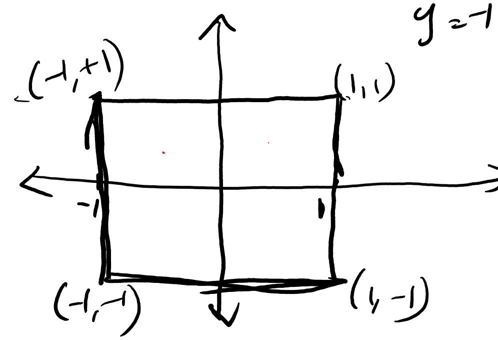
$v_1^2 + v_2^2 = 1$



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

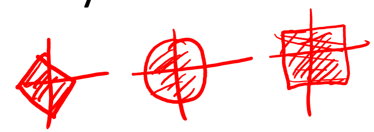
$\max\{|v_1|, |v_2|\} = 1$

$x = 1$
$y = 1$
$x = -1$
$y = -1$



https://www.youtube.com/watch?v=FiSy6zWDfiA

Any true norm $\|v\| \leq 1$ is a convex region.



Vector Norms – S norm

S-Norm:- where S is a symmetric positive matrix.

$$\|v\|_S = v^T S v \longrightarrow \text{Energy of vector } \bar{v}.$$

If $S = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$, $\|v\|_S = [v_1 \ v_2 \ v_3]\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 2v_1^2 + 3v_2^2 + 4v_3^2.$

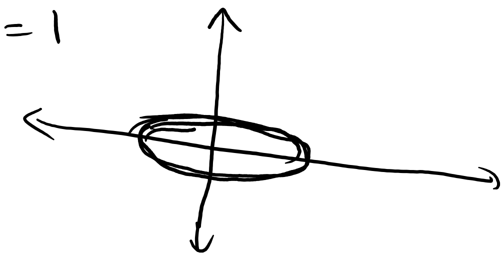
$\bar{v} = [1, 2, 3]$, $\|v\|_S = 2(1)^2 + 3(2)^2 + 4(3)^2 = 50$

$\|v\|_S = 1$, $S = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$

$$\|v\|_S = 2v_1^2 + 3v_2^2.$$

$$\|v\|_S = 1 \longrightarrow 2v_1^2 + 3v_2^2 = 1$$

$$\frac{v_1^2}{1/2} + \frac{v_2^2}{1/3} = 1$$



Vector Norms

- L2 Norm of a vector **v** doesn't change after multiplying a vector by an orthogonal matrix.

$\bar{v} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$

$\|v\|_1 = 7$

$\boxed{\|v\|_2 = 5}$

$\|v\|_\infty = 4$

$\|v\|_S = 116$, if $S = \begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}$

Let $Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$; $\bar{w} = Q\bar{v} = \begin{pmatrix} 7/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$

$\|w\|_1 = 8/\sqrt{2}$.

$\boxed{\|w\|_2 = 5}$

$\|w\|_\infty = 7/\sqrt{2}$.

$\|w\|_S = 4\left(\frac{49}{2}\right) + 5\left(\frac{1}{2}\right) = \frac{204}{2} = 102$

Norms for a matrix

$A = U \Sigma V^T$

$A v_i = \sigma_i u_i$

$\sigma_i = \dfrac{A v_i}{u_i}$

Spectral Norm

$$\|A\|_2 = \max \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1$$

$$= \max \left(\frac{\|Av_1\|}{\|v_1\|}, \frac{\|Av_2\|}{\|v_2\|}, \ldots \right) = \max \{\underline{\sigma_1}, \sigma_2, \ldots \sigma_r\}$$

$(\sigma_1 > \sigma_2 > \ldots > \sigma_r > 0)$

another proof

$$\text{Max } \|A\|_2^2 = \text{Max. } \frac{\|Ax\|_2^2}{\|x\|_2^2} = \max \frac{x^T A^T A x}{x^T x}$$

$\left(\|a\|^2 = a^T a \right)$

$$= \max \frac{x^T S x}{x^T x} = \max \{\lambda_i(S)\} = \lambda_1 = \sigma_1^2$$

$$\text{Max } \|A\|_2^2 = \sigma_1^2$$

$$\Rightarrow \text{Max } \|A\|_2 = \sigma_1$$

Maximize the ratio $\dfrac{\|Ax\|}{\|x\|}$. The maximum is σ_1 at the vector $x = v_1$.

Norms for a matrix

Frobenius Norm

Frobenius norm of a matrix $M = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is $\|M\| = \sqrt{{a_{11}}^2 + {a_{12}}^2 + {a_{21}}^2 + {a_{22}}^2}$

$$\|A\|_F = \sqrt{\sum_i \sum_j |a_{ij}|^2} = a_{11}^2 + a_{12}^2 + \dots + \dots + a_{m1}^2 + a_{m2}^2 + \dots a_{mn}^2$$

$$= \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2}$$

Proof:- $A = U \Sigma V^T$

$$\|A\| = \|U \Sigma V^T\| = \|\Sigma V^T\| = \|\Sigma\|$$

If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $\|A\|_F = \sqrt{1^2 + 2^2 + 3^2 + 4^2} = \sqrt{30}$

Qn.) Find the l_2 norm of the vector with all singular values for $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

$$\sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{2^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2} = \sqrt{9} = 3$$

Frobenius Norm

A =

1	2	8	2	9	4	9
4	7	9	8	3	3	2
9	2	5	2	1	8	7
3	5	1	8	2	5	7
5	6	1	2	6	5	3

```
>> Fnorm=norm(A,'fro')
Fnorm =

   32
>> s=svd(A)
s =
  27.9866
   9.3020
   8.8838
   6.5575
   5.6832
>> norm(s,2)
ans =
  32.0000
```

∴ Frobenius norm of a matrix A is same as the l_2 norm of the vector of singular values of A.

$$\|v\|_F = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots}$$

Norms for a matrix

Nuclear Norm/ Trace norm

$$\|A\|_N = \sigma_1 + \sigma_2 + \cdots + \sigma_r$$

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \|A\|_N = \sigma_1 + \sigma_2 = 3.4392$$

```
>> A=[2,1,1;1,1,1];svd(A)

ans =

   2.9618
   0.4775
>> sum(s)

ans =

   3.4392
```

$A =$

Norms for a matrix

① $A = \begin{bmatrix} 4 & 3 \\ 2 & 7 \\ 1 & 1 \end{bmatrix}$

$\|A\|_2 \rightarrow$ spectral norm $= \sigma_1 = 8.5449$

$\|A\|_F = 8.9443$

$\|A\|_N = \sigma_1 + \sigma_2 = 11.1878$

② $B = \begin{bmatrix} 6 & 7 & 8 & 6 \\ 1 & 2 & 1 & 1 \\ 4 & 3 & 4 & 1 \end{bmatrix}$

spectral $\|B\|_2 = \sigma_1 = 15.1414$

$\|B\|_F = 15.2971$

$\|B\|_N = \sigma_1 + \sigma_2 + \sigma_3 = 17.9883$

③ $C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

spectral $\|C\|_2 = 3.4142 = \sigma_1$

$\|C\|_F = \sqrt{12} = 3.464$

$\|C\|_N = \sigma_1 + \sigma_2 + \underset{0}{\sigma_3} = 4$

Norms for a matrix

Three choices for the matrix norm $||A||$ have special importance and their own names :

Spectral norm $\quad ||A||_2 = \max \dfrac{||Ax||}{||x||} = \boldsymbol{\sigma_1}$ (often called the $\boldsymbol{\ell^2}$ **norm**)

Frobenius norm $\quad ||A||_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$ (12) and (13) also define $||A||_F$

Nuclear norm $\quad ||A||_N = \boldsymbol{\sigma_1} + \boldsymbol{\sigma_2} + \cdots + \boldsymbol{\sigma_r}$ (the trace norm) .

Norms for a matrix

Spectral norm $\quad ||A||_2 = \max \frac{||Ax||}{||x||} = \sigma_1 \quad$ (often called the ℓ^2 **norm**)

Frobenius norm $\quad ||A||_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \quad$ (12) and (13) also define $||A||_F$

Nuclear norm $\quad ||A||_N = \sigma_1 + \sigma_2 + \cdots + \sigma_r \quad$ (the trace norm).

Find the three norms for the identity matrix and any orthogonal matrix

For I_n, $\quad \|I_n\|_2 = 1 \quad ; \quad \|I_n\|_F = \sqrt{n} \quad ; \quad \|I_n\|_N = n$

These norms have different values already for the n by n identity matrix:

$$||I||_2 = 1 \qquad ||I||_F = \sqrt{n} \qquad ||I||_N = n.$$

Replace I by any orthogonal matrix Q and the norms stay the same (because all $\sigma_i = 1$

$$||Q||_2 = 1 \qquad ||Q||_F = \sqrt{n} \qquad ||Q||_N = n.$$

Let $Q = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}, \rightarrow \|Q\|_2 = \sigma_1 = 1 \quad ; \quad \|Q\|_F = \sqrt{2} \quad ; \quad \|Q\|_N = \sigma_1 + \sigma_2 = 2$

Nuclear norm:

>> nucnormA = norm(svd(A),1) %L1 norm of vector of all singular values

```
>> svd(A)
ans =
   27.9526
    9.8184
    6.6884
    1.5859
>> nucnorm = norm(svd(A),1)
nucnorm =
   46.0453
```

Spectral norm:

>> Sv_A=svd(A); specnormA = Sv_A(1) %Linf norm of vector of all singular values (max singular value)

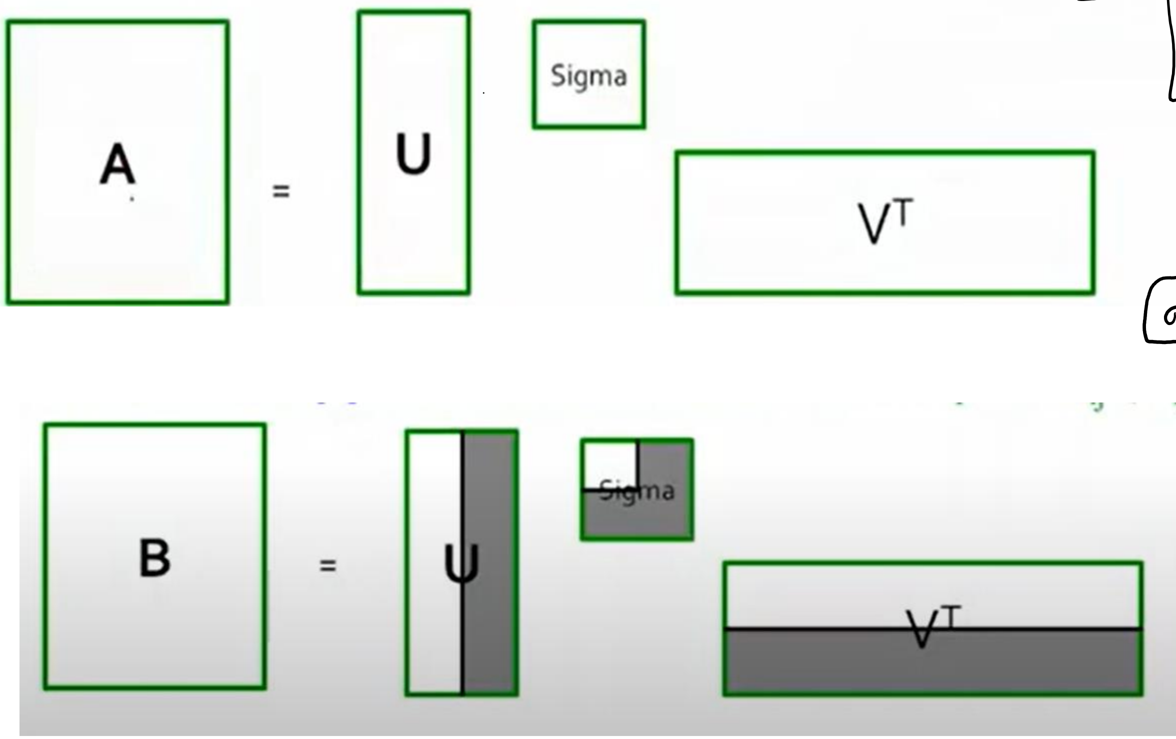
```
>> Sv_A=svd(A), specnormA = Sv_A(1)
Sv_A =
   27.9526
    9.8184
    6.6884
    1.5859
specnormA =
   27.9526
```

Best Low Rank Matrix

SVD gives the best low rank approximation to the matrix

B is best approximation of A: $\|A-B\|_F = \sqrt{\Sigma_{ij} (A_{ij}-B_{ij})^2}$

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ 0 & 0 & \sigma_3 & \cdots \\ & & & 0.001 \\ & & & & 0.0001 \end{bmatrix}$$



$$\boxed{\sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots \sigma_k u_k v_k^T} \cdots + \sigma_n u_n v_n^T$$

SVD gives the best low rank approximation to the matrix

- **Theorem:**
  Let $\mathbf{A} = \mathbf{U}\,\Sigma\,\mathbf{V}^T$ where Σ: $\sigma_1 \geq \sigma_2 \geq \ldots$, and rank($\boldsymbol{A}$)=$\boldsymbol{r}$
  then $\mathbf{B} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^T$ is a **best** rank-$\boldsymbol{k}$ approx. to $\boldsymbol{A}$
  - **Where:**
    S = **diagonal $n \times n$ matrix** where $s_i = \sigma_i$ ($i=1 \ldots k$) else $s_i = 0$

What do we mean by "best":

- B is a solution to $\min_B \|A-B\|_F$ where $\text{rank}(B)=k$

SVD gives the best low rank approximation to the matrix

$$
\begin{array}{c c}
 & \longleftarrow m \longrightarrow \\
n & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}
\end{array}
= \overbrace{\sigma_1 \, \underset{n \times 1}{u_1} \, \underset{1 \times m}{v^T_1} + \sigma_2 \, \underbrace{u_2 \, v^T_2}_{1} + \dots}^{k \text{ terms}}
$$

$+ \sigma_{90} u_{90} v^T_{90} + \sigma_{91} u_{91} v^T_{91}$

Assume: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq 0$

Why is setting small σ_i to 0 the right thing to do?

Vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are unit length, so σ_i scales them.

So, zeroing small σ_i introduces less error.

SVD gives the best low rank approximation to the matrix

Q: How many σs to keep?

A: Rule-of-a thumb:

keep 80-90% of 'energy' $(=\sum \sigma_i^2)$

$$
\begin{array}{cc}
 & \longleftarrow m \longrightarrow \\
n & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}
\end{array}
= \sigma_1 \; u_1 \; v^T_1 + \sigma_2 \; u_2 \; v^T_2 + \ldots
$$

Assume: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots$

$$\frac{\sigma_1^2 + \ldots + \sigma_k^2}{\sigma_1^2 + \ldots + \sigma_n^2} \geq 0.8$$

9

8

5

1

0.5

0.25

0.001