Arrays: computing with many numbers

Some perspective

- We have so far (mostly) looked at what we can do with single numbers (and functions that return single numbers).
- Things can get much more interesting once we allow not just one, but many numbers together.
- It is natural to view an array of numbers as one object with its own rules.
- The simplest such set of rules is that of a **vector.**

Vectors

A vector is an element of a Vector Space

$$\boldsymbol{n}\text{-vector:} \qquad \boldsymbol{x} = \left\{ \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array} \right\} = [x_1 \quad x_2 \cdots x_n]^T$$

Vector space $\mathcal{V}$:

A vector space is a set $\mathcal{V}$ of vectors and a field $\mathcal{F}$ of scalars with two operations:

1) addition: $u + v \in \mathcal{V}$, and $u, v \in \mathcal{V}$

2) multiplication : $\alpha \cdot u \in \mathcal{V}$, and $u \in \mathcal{V}$, $\alpha \in \mathcal{F}$

Vector Space

The addition and multiplication operations must satisfy:

(for $\alpha, \beta \in \mathcal{F}$ and $u, v \in \mathcal{V}$)

Associativity: $u + (v + w) = (u + v) + w$

Commutativity: $u + v = v + u$

Additive identity: $v + 0 = v$

Additive inverse: $v + (-v) = 0$

Associativity wrt scalar multiplication: $\alpha \cdot (\beta \cdot v) = (\alpha \cdot \beta) \cdot v$

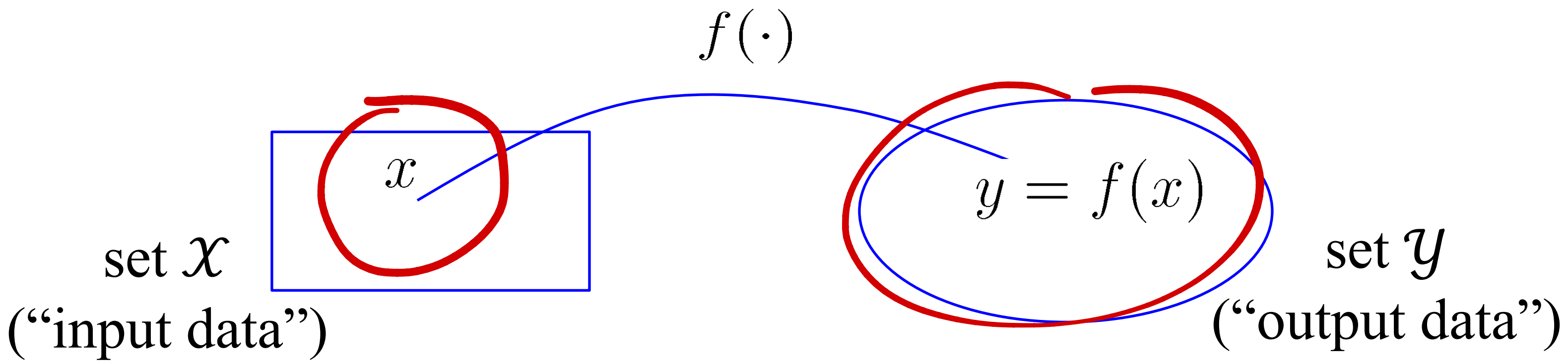
Distributive wrt scalar addition: $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$

Distributive wrt vector addition: $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$

Scalar multiplication identity: $1 \cdot (u) = u$

Linear Functions

Function: $f : \mathcal{X} \to \mathcal{Y}$



The function f takes vectors $\boldsymbol{x} \in \mathcal{X}$ and transforms into vectors $\boldsymbol{y} \in \mathcal{Y}$

A function f is a linear function if

(1) $f(\boldsymbol{u}+\boldsymbol{v}) = f(\boldsymbol{u}) + f(\boldsymbol{v})$

(2) $f(a\boldsymbol{u}) = a\, f(\boldsymbol{u})$ for any scalar a

Linear functions?

$f(x) = \frac{|x|}{x}, \; f: \mathcal{R} \to \mathcal{R}$

$f(u+v) = \frac{|u+v|}{u+v}$

$f(u) = \frac{|u|}{u} \qquad f(v) = \frac{|v|}{v}$

$f(u+v) \neq f(u) + f(v)$

$\boxed{f(x) = \underline{a\,x + b},}\; f: \mathcal{R} \to \mathcal{R}, \;\; a, b \in \mathcal{R}$ and $a, b \neq 0$

$\left.\begin{array}{l} f(u) = au + b \\ f(v) = av + b \end{array}\right\} au + av + 2b$

$= a(u+v) + 2b$

$\underline{\underline{f(u+v)}} = a(u+v) + b$

$\neq$

Matrices

- $m \times n$-matrix

$$\boldsymbol{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix}$$

- Linear functions $f(\boldsymbol{x})$ can be represented by a Matrix-Vector multiplication.

- Think of a matrix $\boldsymbol{A}$ as a linear function that takes vectors $\boldsymbol{x}$ and transforms them into vectors $\boldsymbol{y}$

$$\underline{\boldsymbol{y} = f(\boldsymbol{x})} \rightarrow \underline{\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x}}$$

- Hence we have:

$$\boldsymbol{A}\,(\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{A}\,\boldsymbol{u} + \boldsymbol{A}\,\boldsymbol{v}$$
$$\boldsymbol{A}\,(\alpha\,\boldsymbol{u}) = \alpha\,\boldsymbol{A}\,\boldsymbol{u}$$

A

Matrix-Vector multiplication

$m \times n$

- Recall summation notation for matrix-vector multiplication $\boldsymbol{y} = \boldsymbol{A}\, \boldsymbol{x}$

$m \times 1$ $n \times 1$

$$y_i = \sum_{j=1}^{n} A_{ij} x_j \qquad i = 1,2, \dots, m$$

- You can think about matrix-vector multiplication as:

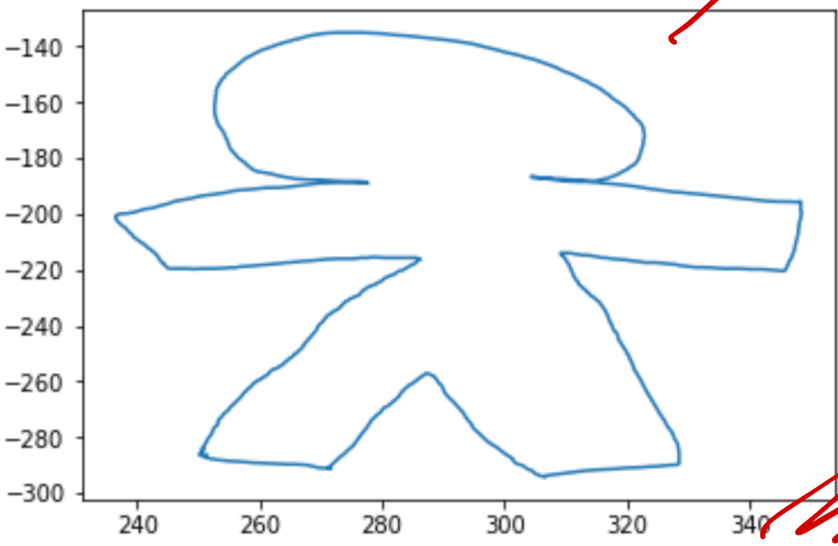
Linear combination of column vectors of **A**

$$\boldsymbol{y} = x_1\, \mathbf{A}[:,1] + x_2\, \mathbf{A}[:,2] + \cdots + x_n\, \mathbf{A}[:,n]$$

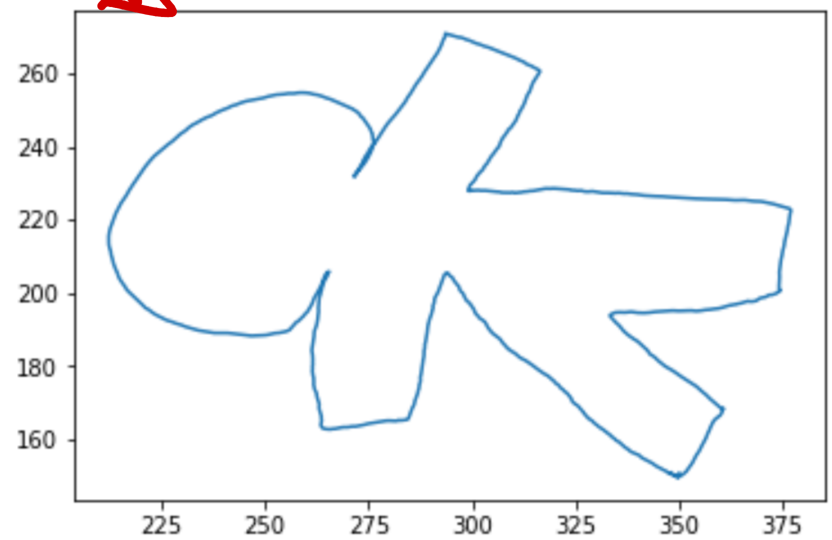
Dot product of $\boldsymbol{x}$ with rows of **A**

$$\boldsymbol{y} = \begin{pmatrix} \mathbf{A}[1,:] \cdot \boldsymbol{x} \\ \vdots \\ \mathbf{A}[m,:] \cdot \boldsymbol{x} \end{pmatrix}$$

Matrices operating on data



Data set: $\boldsymbol{x}$



Data set: $\boldsymbol{y}$

Rotation

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$$

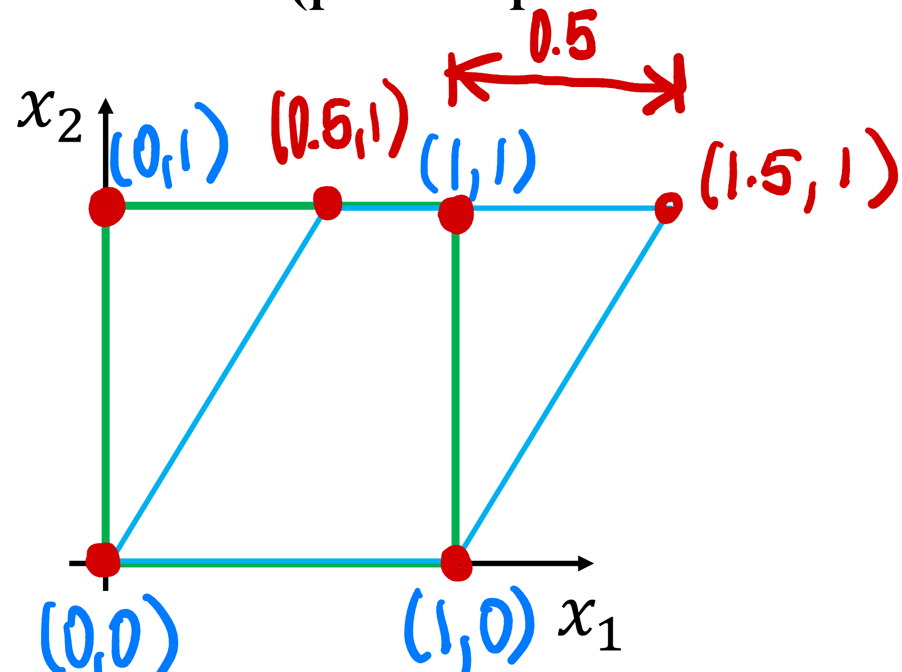
or

$$\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x}$$

Example: Shear operator

Matrix-vector multiplication for each vector (point representation in 2D):

$$\underbrace{\begin{pmatrix}1 & 0.5\\ 0 & 1\end{pmatrix}}_{A}\begin{pmatrix}1\\ 1\end{pmatrix}=\begin{pmatrix}1.5\\ 1\end{pmatrix}$$

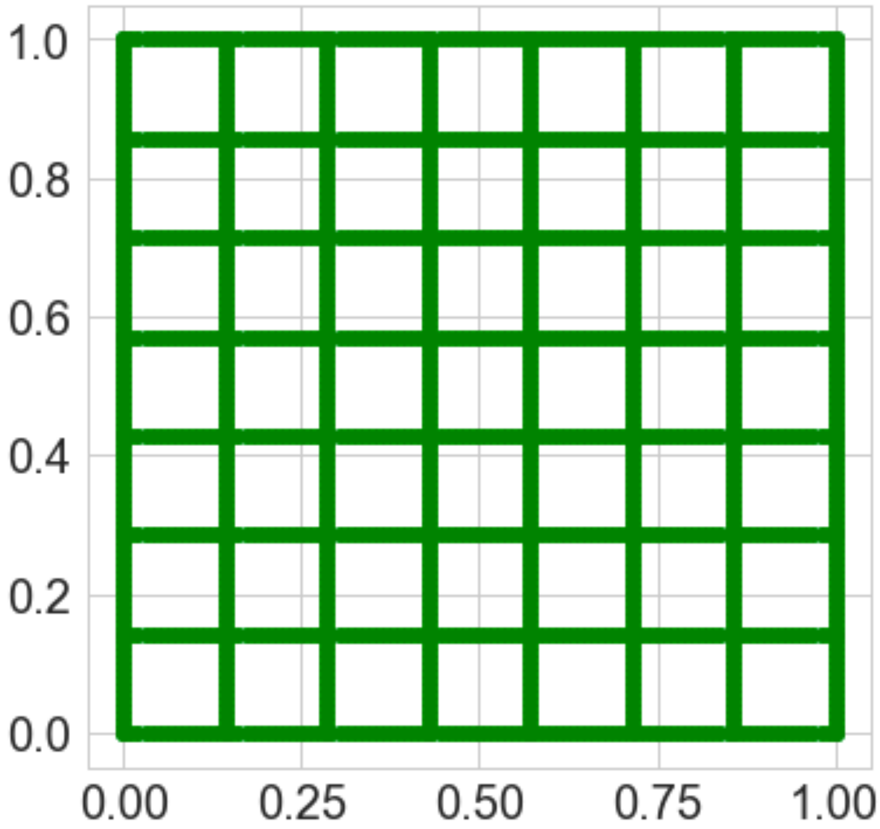


$$\begin{pmatrix}1 & 0.5\\ 0 & 1\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix}=\begin{pmatrix}0.5\\ 1\end{pmatrix}$$

$$A \leftarrow \underbrace{\begin{pmatrix}1 & 0.5\\ 0 & 1\end{pmatrix}}_{2\times 2}\underbrace{\begin{pmatrix}\text{data}\end{pmatrix}}_{2\times 640}=\underbrace{\begin{pmatrix}\text{data final}\end{pmatrix}}_{2\times 640}$$

Matrices as operators

- **Data**: grid of 2D points
- Transform the data using matrix multiply



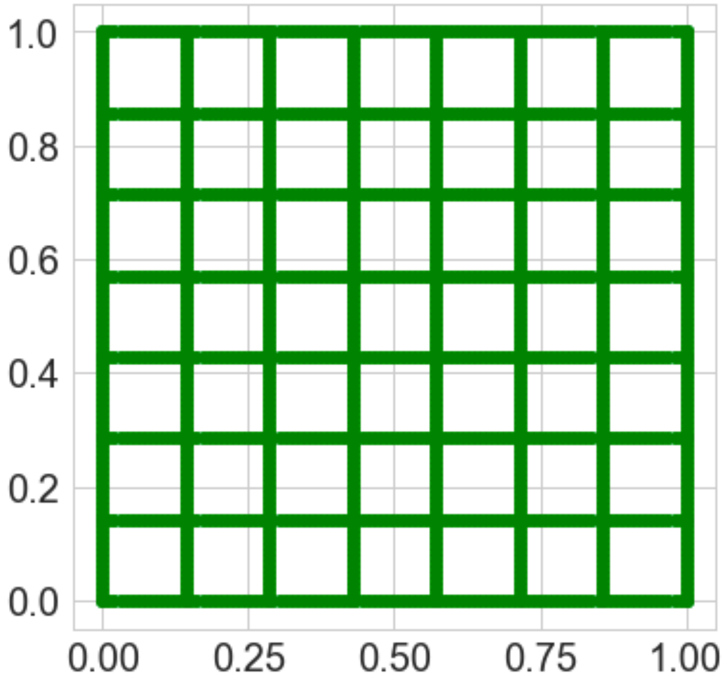
What can matrices do?

1. Shear
2. Rotate
3. Scale
4. Reflect

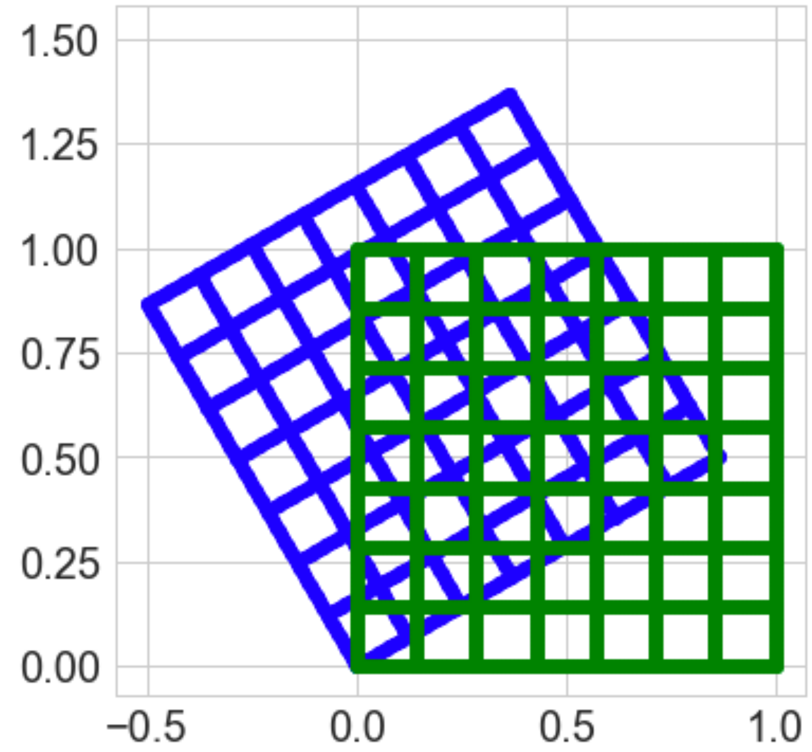
5. Can they translate?

Rotation operator

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

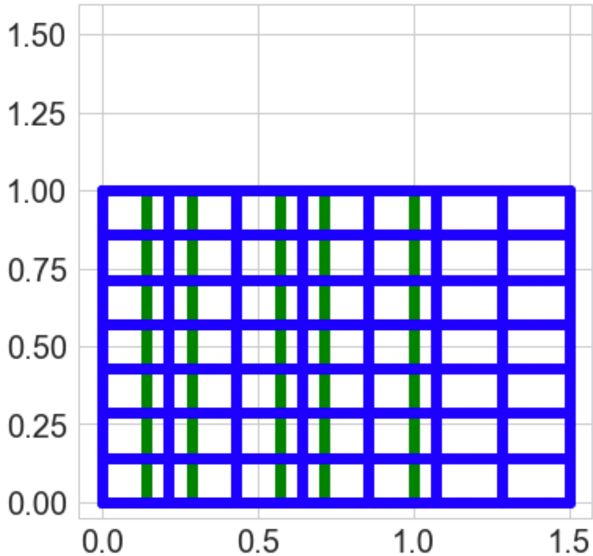


$\theta = \pi/6$



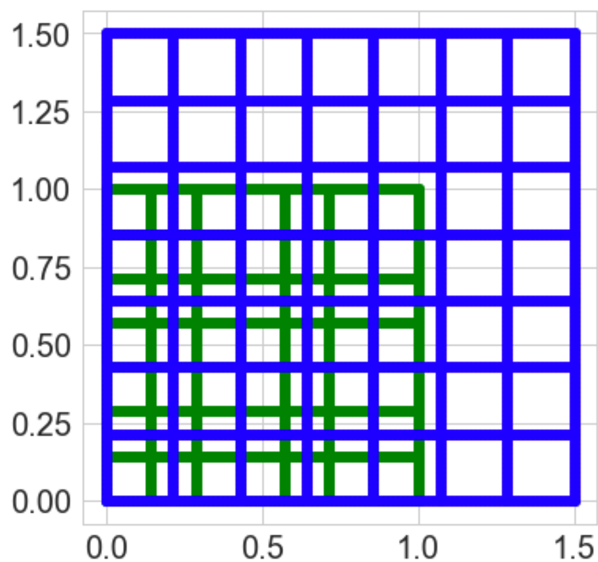
Scale operator

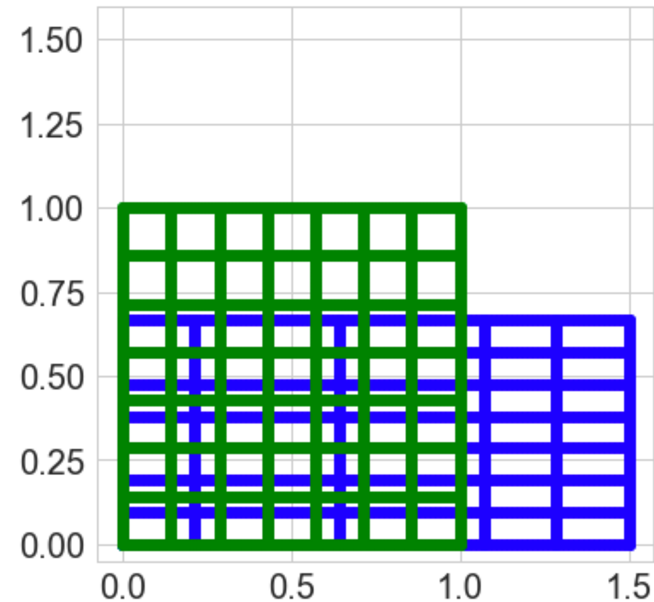
$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$



$$\begin{pmatrix} 3/2 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 3/2 & 0 \\ 0 & 3/2 \end{pmatrix}$$



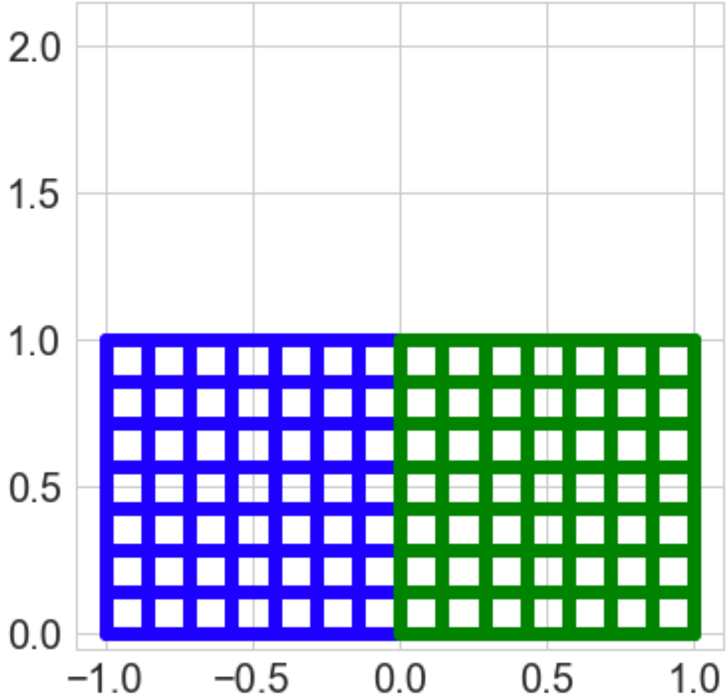


$$\begin{pmatrix} 3/2 & 0 \\ 0 & 2/3 \end{pmatrix}$$

Reflection operator

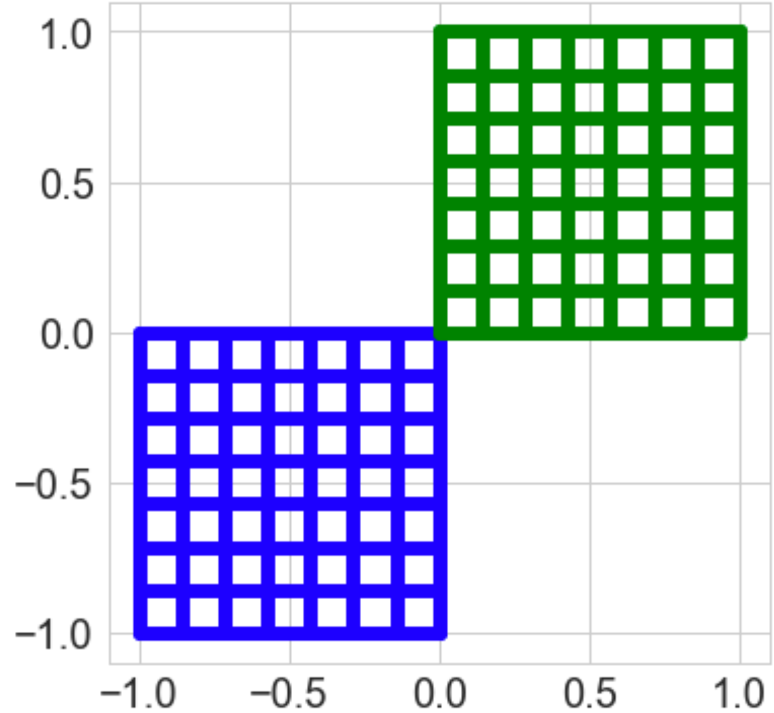
$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} -a & 0 \\ 0 & -b \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$



Reflect about y-axis

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

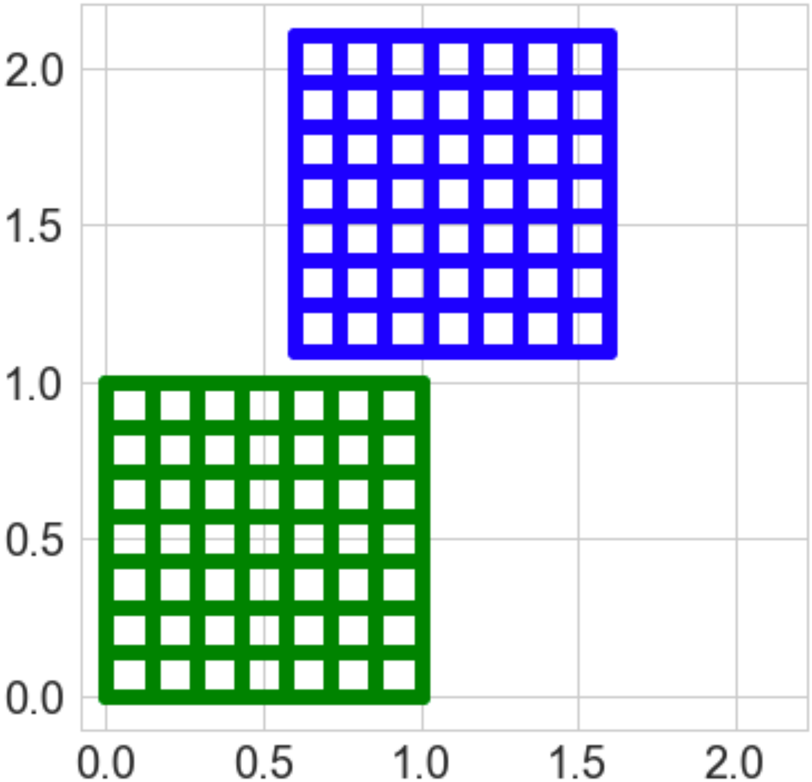


Reflect about x and y-axis

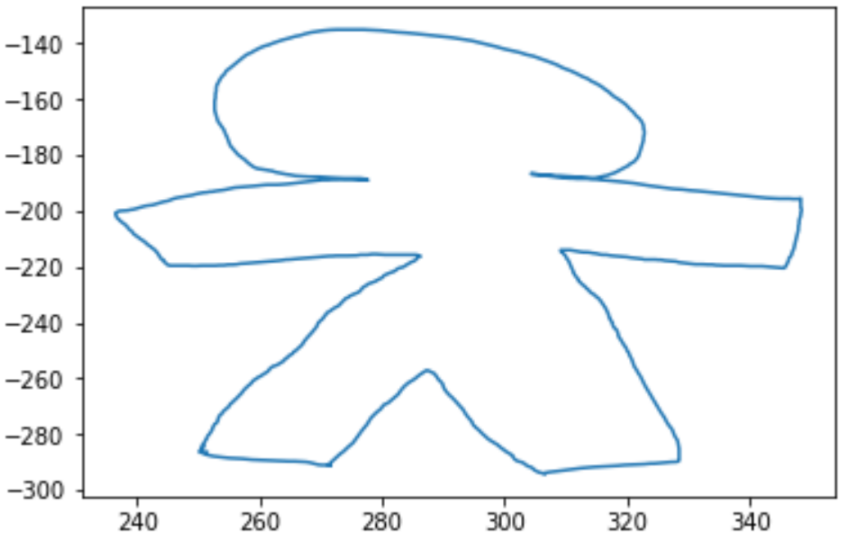
Translation (shift)

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix}$$

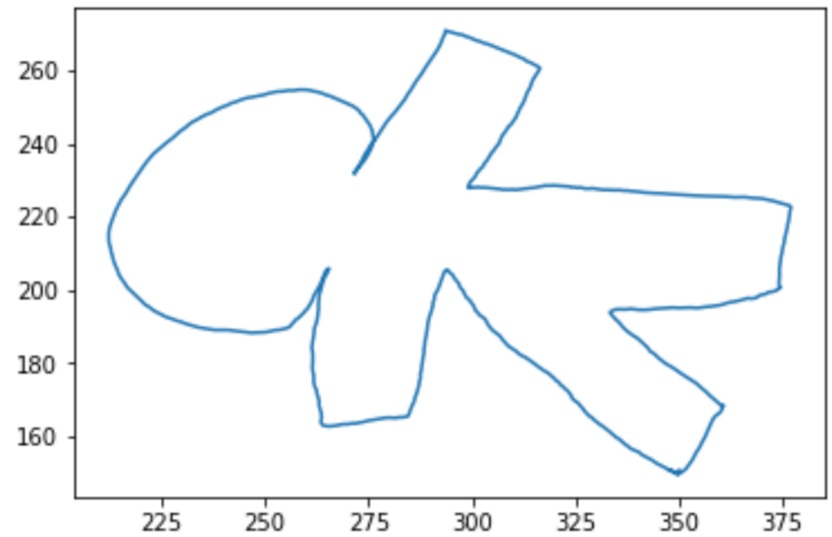
$a = 0.6; b = 1.1$



Matrices operating on data



Data set: *A*



Data set: *B*

Rotation

Norms

What's a norm?

- A generalization of 'absolute value' to vectors.
- $f(x) : \mathbb{R}^n \to \mathbb{R}_0^+$, returns a 'magnitude' of the input vector
- In symbols: Often written $\|\mathbf{x}\|$.

Define norm.

A function $\|\mathbf{x}\| : \mathbb{R}^n \to \mathbb{R}_0^+$ is called a norm if and only if

1. $\|\mathbf{x}\| > 0 \Leftrightarrow \mathbf{x} \neq \mathbf{0}$.
2. $\|\gamma\mathbf{x}\| = |\gamma| \, \|\mathbf{x}\|$ for all scalars γ.
3. Obeys triangle inequality $\|\mathbf{x} + \mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$

Demo “Vector Norms”

Example of Norms

What are some examples of norms?

The so-called *p-norms*:

$$\left\| \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_p = \sqrt[p]{|x_1|^p + \cdots + |x_n|^p} \quad (p \geqslant 1)$$

$p = 1, 2, \infty$ particularly important

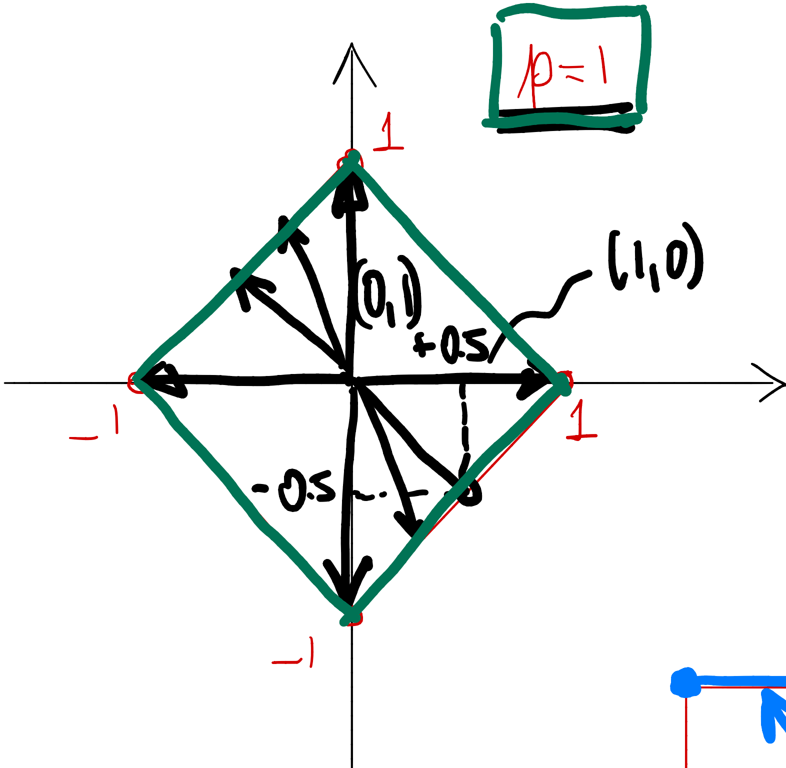
$p=1$: $|x_1| + |x_2| + \cdots + |x_n|$

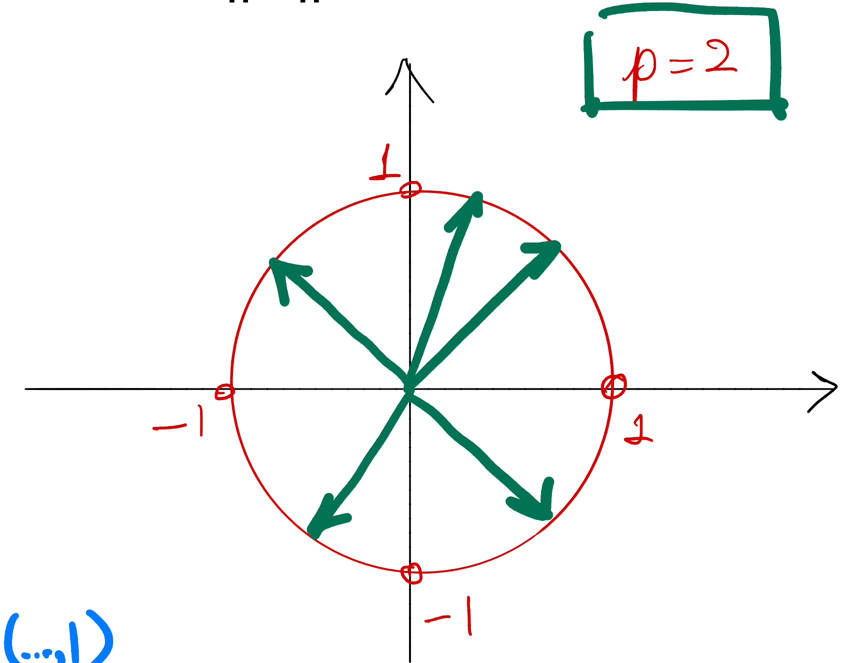
$p=2$: $\sqrt{x_1^2 + x_2^2 + \cdots |x_n|^2}$

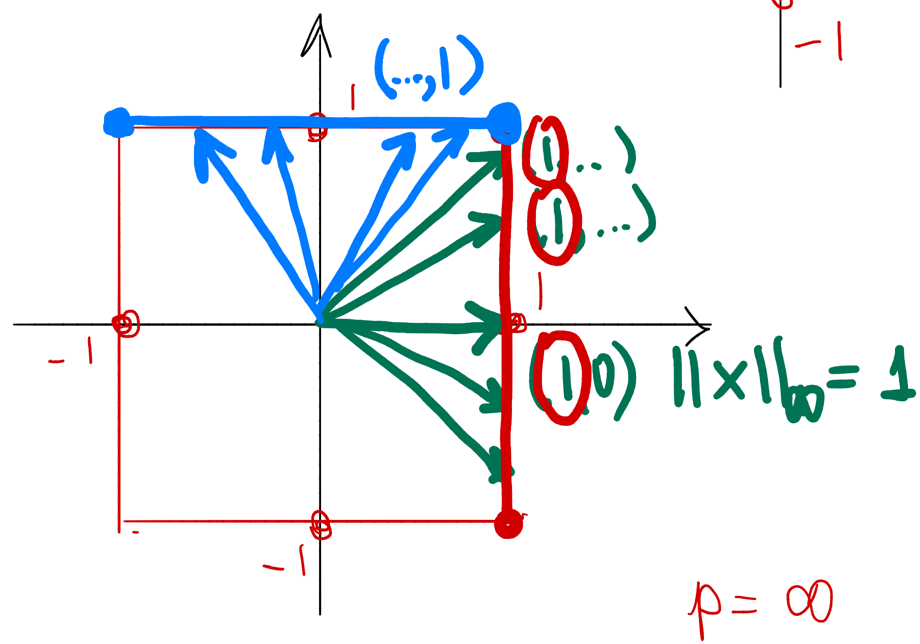
$p=\infty$: $\|x\|_\infty = \max_i |x_i|$

Unit Ball: Set of vectors $\boldsymbol{x}$ with norm $\|\boldsymbol{x}\| = 1$

in 2D







Norms and Errors

If we're computing a vector result, the error is a vector.
That's not a very useful answer to 'how big is the error'.
What can we do?

Apply a norm!

How? Attempt 1:

$$\text{Magnitude of error} \neq \cancel{\|\text{true value}\| - \|\text{approximate value}\|}$$ **WRONG!**

Attempt 2:

$$\text{Magnitude of error} = \|\underbrace{\text{true value} - \text{approximate value}}\|$$

Absolute and Relative Errors

What are the absolute and relative errors in approximating the location of Siebel center (40.114, −88.224) as (40, −88) using the 2-norm?

$x_{true} = (40.114, -88.224)$

$x_{mea} = (40, -88)$

$\underset{\sim}{e_a} = (0.114, -0.224)$

$\|e_a\|_{p=2} = \sqrt{0.114^2 + 0.224^2} = 0.2513$

$\|e_r\|_{p=2} = \dfrac{\|e_a\|_{p=2}}{\|x_{true}\|_{p=2}}$

$= \dfrac{0.2513}{\sqrt{40.114^2 + 88.224^2}}$

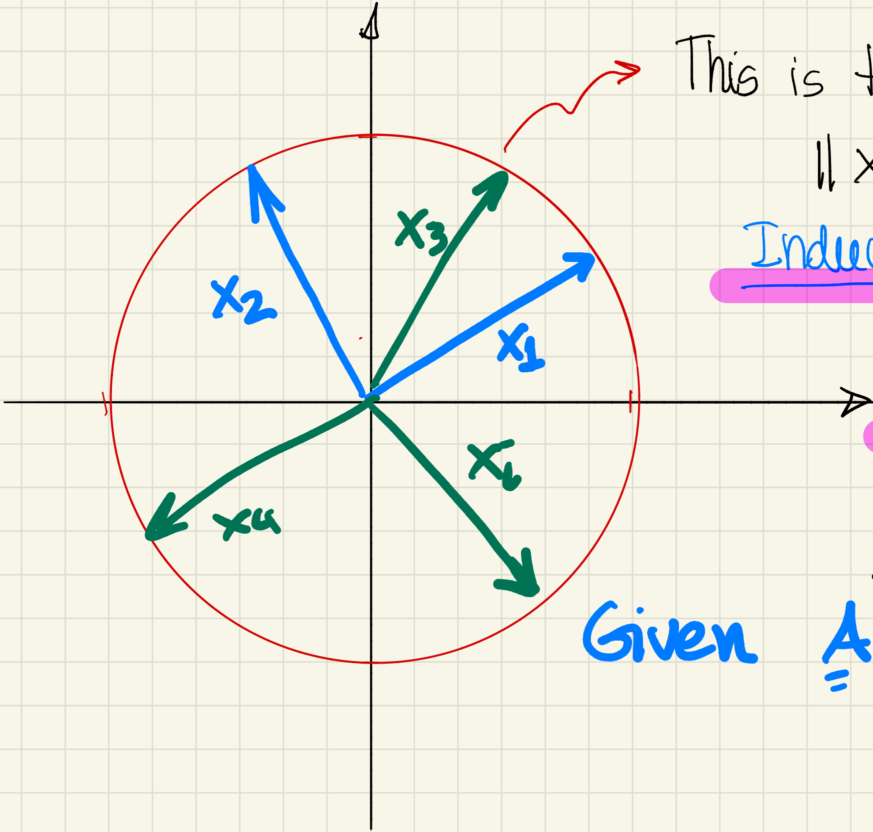
$\|e_r\|_{p=2} = 2.593 \times 10^{-3}$

Matrix Norms

What norms would we apply to matrices?

- Easy answer: '*Flatten*' matrix as vector, use vector norm. This corresponds to an entrywise matrix norm called the Frobenius norm,

$$\|A\|_F := \sqrt{\sum_{i,j} a_{ij}^2}.$$



This is the collection of all vectors $\underset{\sim}{x}$ such that

$$\|x\|_2 = 1$$

Induced matrix norm

$$\|A\|_p = \max_{\|x\|_p = 1} \|Ax\|_p \quad //$$

max

Given $\underset{=}{A}$:

$y_1 = Ax_1 \to \|y_1\|_p = \_\_\_\_$

$y_2 = Ax_2 \to \|y_2\|_p = \_\_\_\_$

$\vdots$

$y_i = Ax_i \to \|y_i\|_p = \_\_\_\_$

Matrix Norms

However, interpreting matrices as linear functions, what we are really interested in is the maximum amplification of the norm of any vector multiplied by the matrix,

$$\|A\| := \max_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| .$$

These are called induced matrix norms, as each is associated with a specific vector norm $\|\cdot\|$.

Matrix Norms

The following are equivalent:

$$\max_{\|x\|\neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = \max_{\|x\|\neq 0} \left\| A \underbrace{\frac{\mathbf{x}}{\|\mathbf{x}\|}}_{y} \right\| \overset{\|\mathbf{y}\|=1}{=} \max_{\|\mathbf{y}\|=1} \|A\mathbf{y}\| = \|A\| \, .$$

Logically, for each vector norm, we get a different matrix norm, so that, e.g. for the vector 2-norm $\|\mathbf{x}\|_2$ we get a matrix 2-norm $\|A\|_2$, and for the vector ∞-norm $\|\mathbf{x}\|_\infty$ we get a matrix ∞-norm $\|A\|_\infty$.

Induced Matrix Norms

$$\underset{m\times n}{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ +A_{21} & A_{22} & & \\ +A_{31} & & \ddots & \\ \vdots & & & \\ +A_{m1} & & & A_{mn} \end{bmatrix}$$

$$\|\boldsymbol{A}\|_1 = \max_j \sum_{i=1}^{n} |A_{ij}|$$

Maximum absolute column sum of the matrix $\boldsymbol{A}$

$$\max\{\ \boxed{}\ \boxed{}\ \cdots\ \boxed{}\ \}$$

$$\|\boldsymbol{A}\|_\infty = \max_i \sum_{j=1}^{n} |A_{ij}|$$

Maximum absolute row sum of the matrix $\boldsymbol{A}$

$$\underset{m\times n}{A} = \begin{bmatrix} A_{11} + A_{12} + \cdots + A_{1n} \\ A_{21} \quad A_{22} \\ A_{31} \quad \ddots \\ \vdots \\ A_{m1} \qquad\qquad A_{mn} \end{bmatrix} \begin{matrix} = \boxed{} \\ = \boxed{} \\ = \vdots \\ = \boxed{} \end{matrix}$$

max

$$\|\boldsymbol{A}\|_2 = \max_k \sigma_k$$

σ_k are the singular value of the matrix $\boldsymbol{A}$

Properties of Matrix Norms

Matrix norms inherit the vector norm properties:

1. $||A|| > 0 \Leftrightarrow A \neq \mathbf{0}$.
2. $||\gamma A|| = |\gamma| \, ||A||$ for all scalars γ.
3. Obeys triangle inequality $||A + B|| \leqslant ||A|| + ||B||$

$\boxed{\text{But also some more properties that stem from our definition:}}$

1. $||A\mathbf{x}|| \leqslant ||A|| \, ||\mathbf{x}||$
2. $||AB|| \leqslant ||A|| \, ||B||$ (easy consequence)

Both of these are called submultiplicativity of the matrix norm.

Examples

Determine the norm of the following matrices:

1) $\left\| \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \right\|_\infty \rightarrow \boxed{\begin{matrix} 3 \\ 7 \end{matrix}} \Rightarrow \| \quad \|_\infty = 7$

2) $\left\| \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \right\|_1$

$\boxed{4 \quad 6} \Rightarrow \| \quad \|_{p=1} = 6$

Matrix Norm Approximation

Suppose you know that for a given matrix A three vectors $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ for the vector norm $||\cdot||$,

$||\mathbf{x}|| = 2$, $||\mathbf{y}|| = 1$, $||\mathbf{z}|| = 3$,

and for corresponding induced matrix norm,

$||A\mathbf{x}|| = 20$, $||A\mathbf{y}|| = 5$, $||A\mathbf{z}|| = 90$.

What is the largest lower bound for $||A||$ that you can derive from these values?

$$\left\{ \frac{||Ax||}{||x||}, \frac{||Ay||}{||y||}, \frac{||Az||}{||z||} \right\}$$

$$= \{10, 5, 30\} \qquad ||A|| \rightarrow 30$$

Induced Matrix Norm of a Diagonal Matrix

What is the 2-norm-based matrix norm of the diagonal matrix

$$A = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 13 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}?$$

$$\sigma = [100, 13, 0.5]$$

$$\|A\|_{p=2} = \max_i \sigma_i = 100$$

Induced Matrix Norm of an Inverted Diagonal Matrix

What is the 2-norm-based matrix norm of the **inverse** of the diagonal matrix

$$A = \begin{bmatrix} 100 & 0 & 0 \\ 0 & 13 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}?$$

$\|A^{-1}\|_{p=2}$

sing values A^{-1} ?

$$A^{-1} = \begin{bmatrix} 1/100 & & \\ & 1/13 & \\ & & 1/0.5 \end{bmatrix}$$

$$\sigma = \left[\frac{1}{100}, \frac{1}{13}, \boxed{\frac{1}{0.5}}\right] \to 2$$

Notation and special matrices

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- Square matrix: $m = n$
- Zero matrix: $A_{ij} = 0$
- Identity matrix $[\boldsymbol{I}] = [\delta_{ij}]$
- Symmetric matrix: $A_{ij} = A_{ji}$ $[\boldsymbol{A}] = [\boldsymbol{A}]^T$
- Permutation matrix:
  - Permutation of the identity matrix
  - Permutes (swaps) rows

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} c \\ a \\ b \end{pmatrix}$$

- Diagonal matrix: $A_{ij} = 0, \ \forall i, j \mid i \neq j$
- Triangular matrix:

Lower triangular: $L_{ij} = \begin{cases} L_{ij}, i \geq j \\ 0, \quad i < j \end{cases}$ Upper triangular: $U_{ij} = \begin{cases} U_{ij}, i \leq j \\ 0, \quad i > j \end{cases}$

More about matrices

- Rank: the rank of a matrix $\boldsymbol{A}$ is the dimension of the vector space generated by its columns, which is equivalent to the number of linearly independent columns of the matrix.
- Suppose $\boldsymbol{A}$ has shape $m \times n$:
  - $rank(\boldsymbol{A}) \leq \min(m, n)$
  - Matrix $\boldsymbol{A}$ is **full rank**: $rank(\boldsymbol{A}) = \min(m, n)$. Otherwise, matrix $\boldsymbol{A}$ is **rank deficient**.
- Singular matrix: a square matrix $\boldsymbol{A}$ is invertible if there exists a square matrix **B** such that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$. If the matrix is not invertible, it is called singular.