Chapter II: Background Mathematics

Information Retrieval & Data Mining
Universität des Saarlandes, Saarbrücken
Winter Semester 2013/14

Chapter II: Background Mathematics

1. **Linear Algebra**
   Matrices, vectors, and related concepts
2. **Probability Theory and Statistical Inference**
   Events, probabilities, random variables, and limit theorems; likehoods and estimators
3. **Confidence Intervals, Hypothesis Testing and Regression**
   Confidence intervals, statistical tests, linear regression

Chapter II.1: Linear Algebra

Matrices and vectors

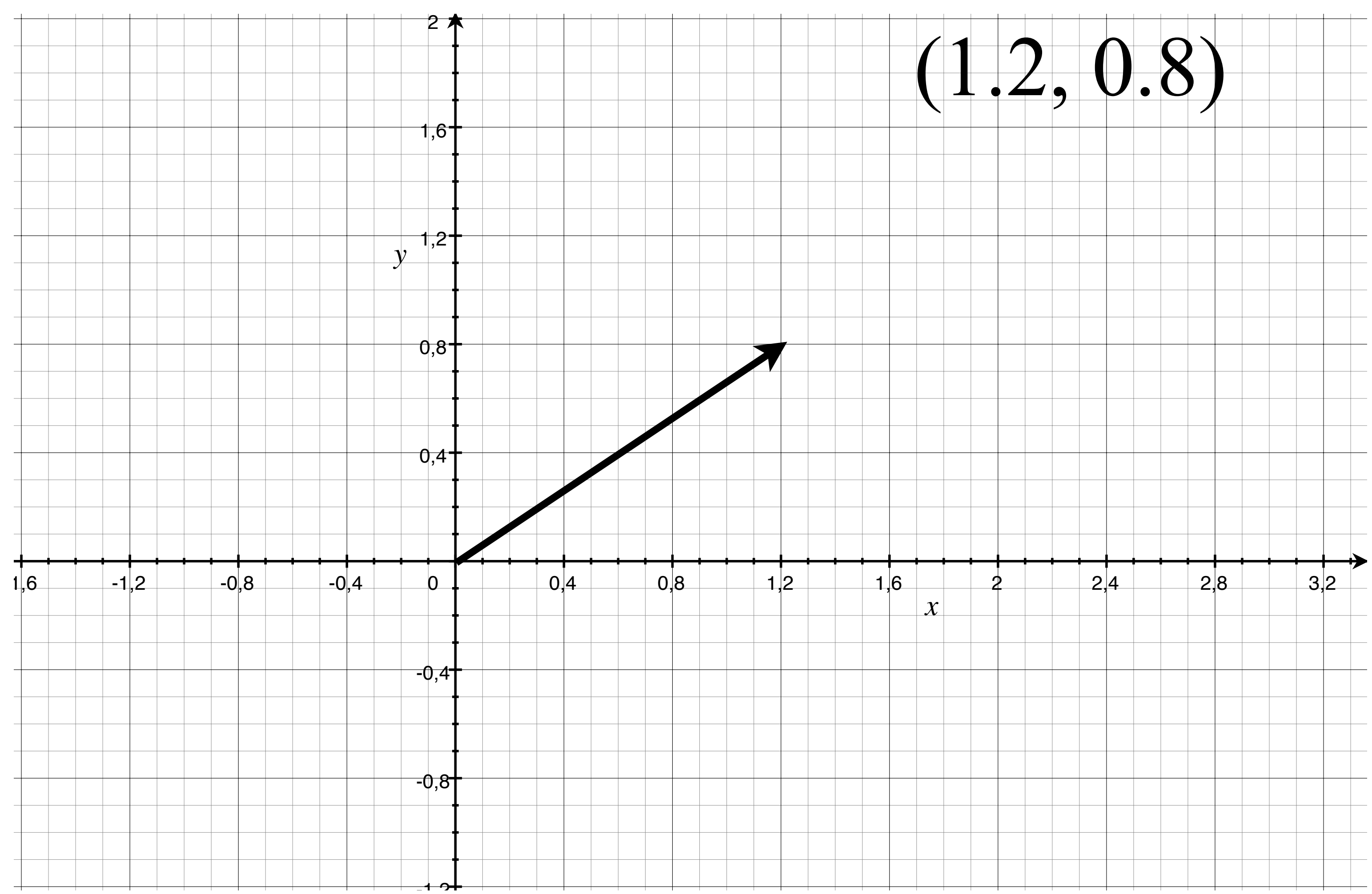
- A **vector** is
  - a 1D array of numbers
  - a geometric entity with magnitude and direction
- The **norm** of the vector defines its magnitude
  - **Euclidean** (L_2) norm:
  - L_p norm ($1 \leq p \leq \infty$)

  $\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$
- The direction is the angle

Matrices and vectors

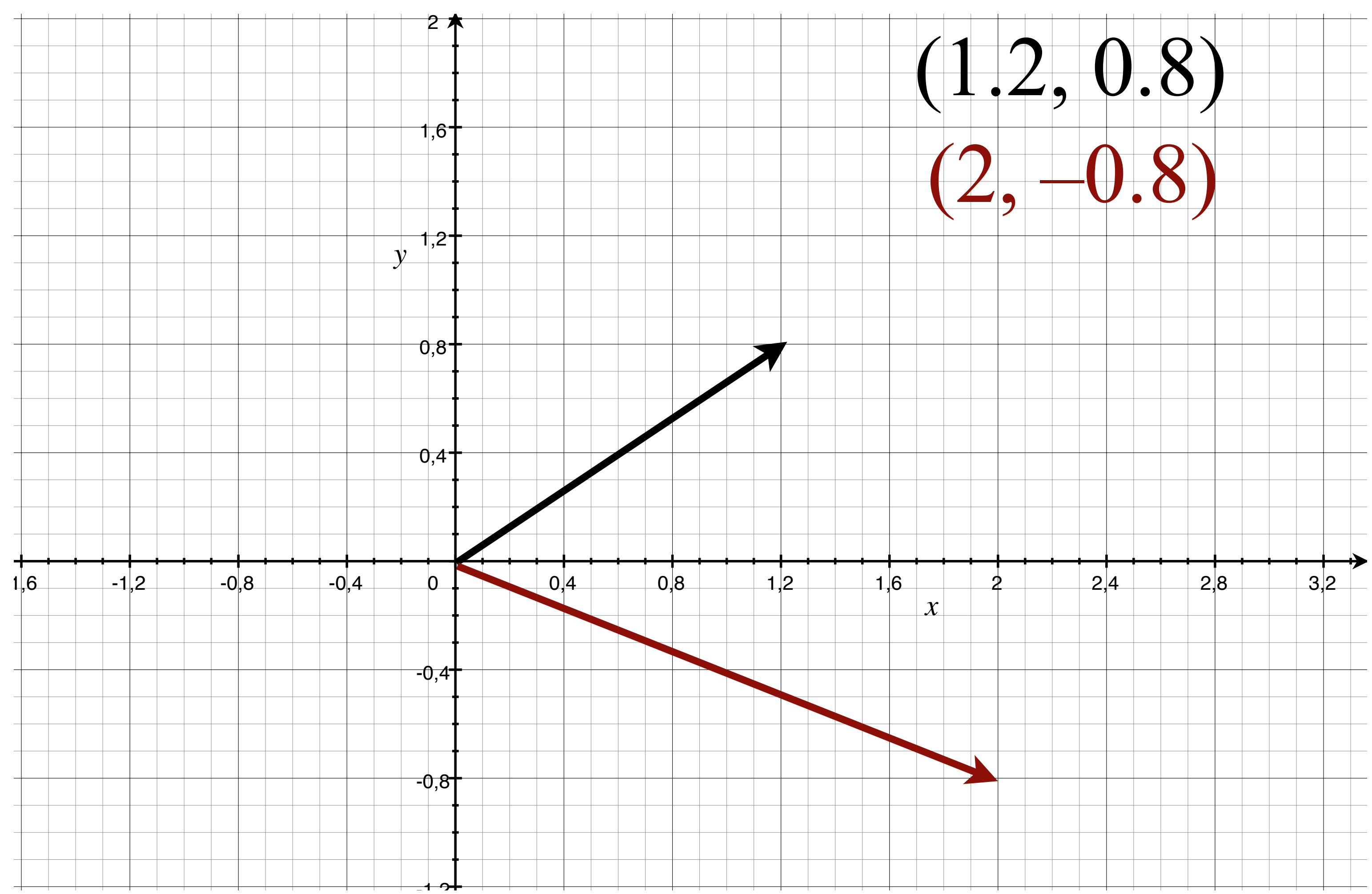
- A **vector** is
  - a 1D array of numbers
  - a geometric entity with magnitude and direction
- The **norm** of the vector defines its magnitude
  - **Euclidean** (L_2) norm:
  - L_p norm ($1 \leq p \leq \infty$)

  $$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$
- The direction is the angle



Matrices and vectors

- A **vector** is
  - a 1D array of numbers
  - a geometric entity with magnitude and direction
- The **norm** of the vector defines its magnitude
  - **Euclidean** (L_2) norm:



  - L_p norm ($1 \leq p \leq \infty$)

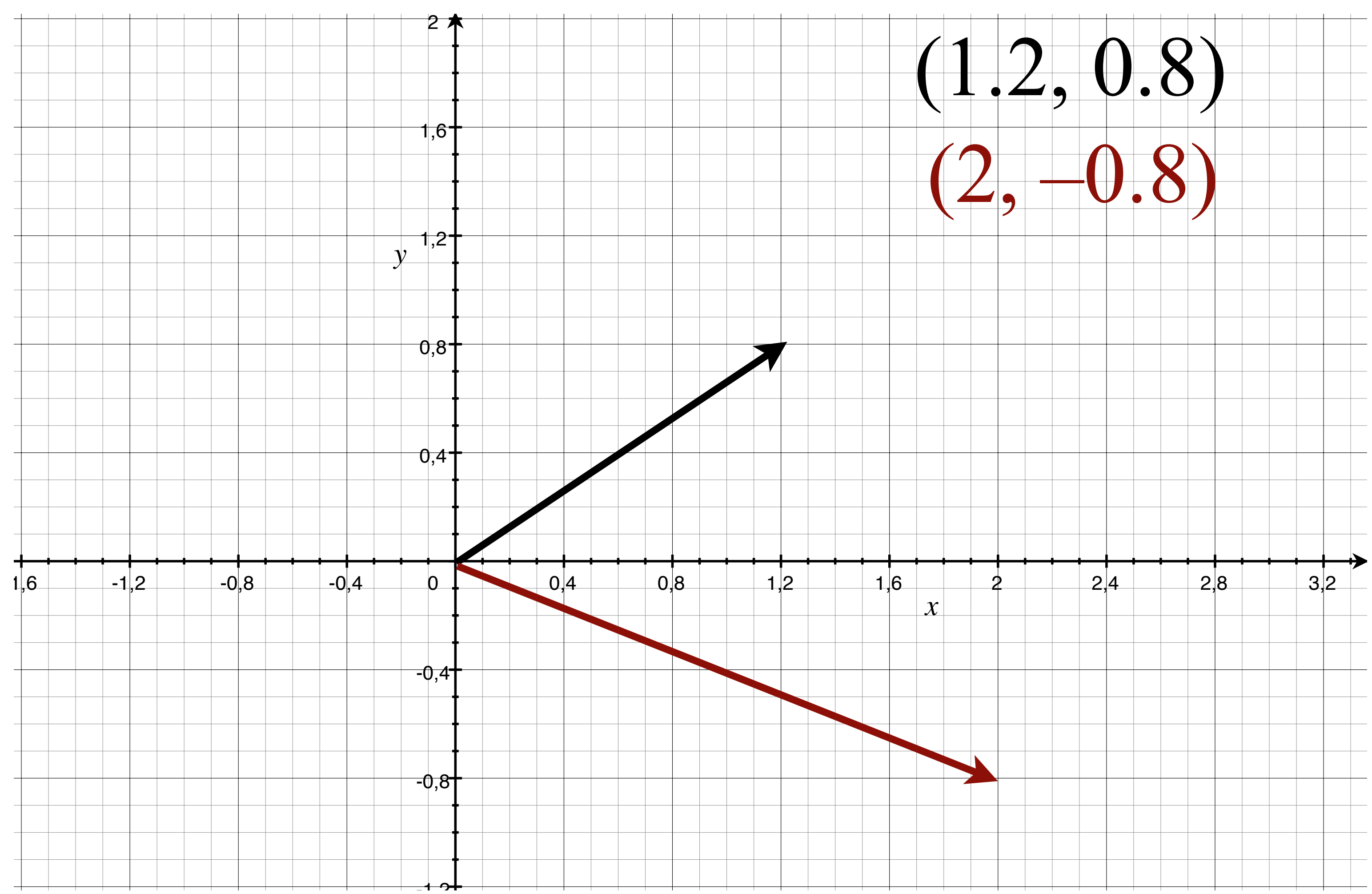
$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

- The direction is the angle

Matrices and vectors

- A **vector** is
  - a 1D array of numbers
  - a geometric entity with magnitude and direction
- The **norm** of the vector defines its magnitude

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

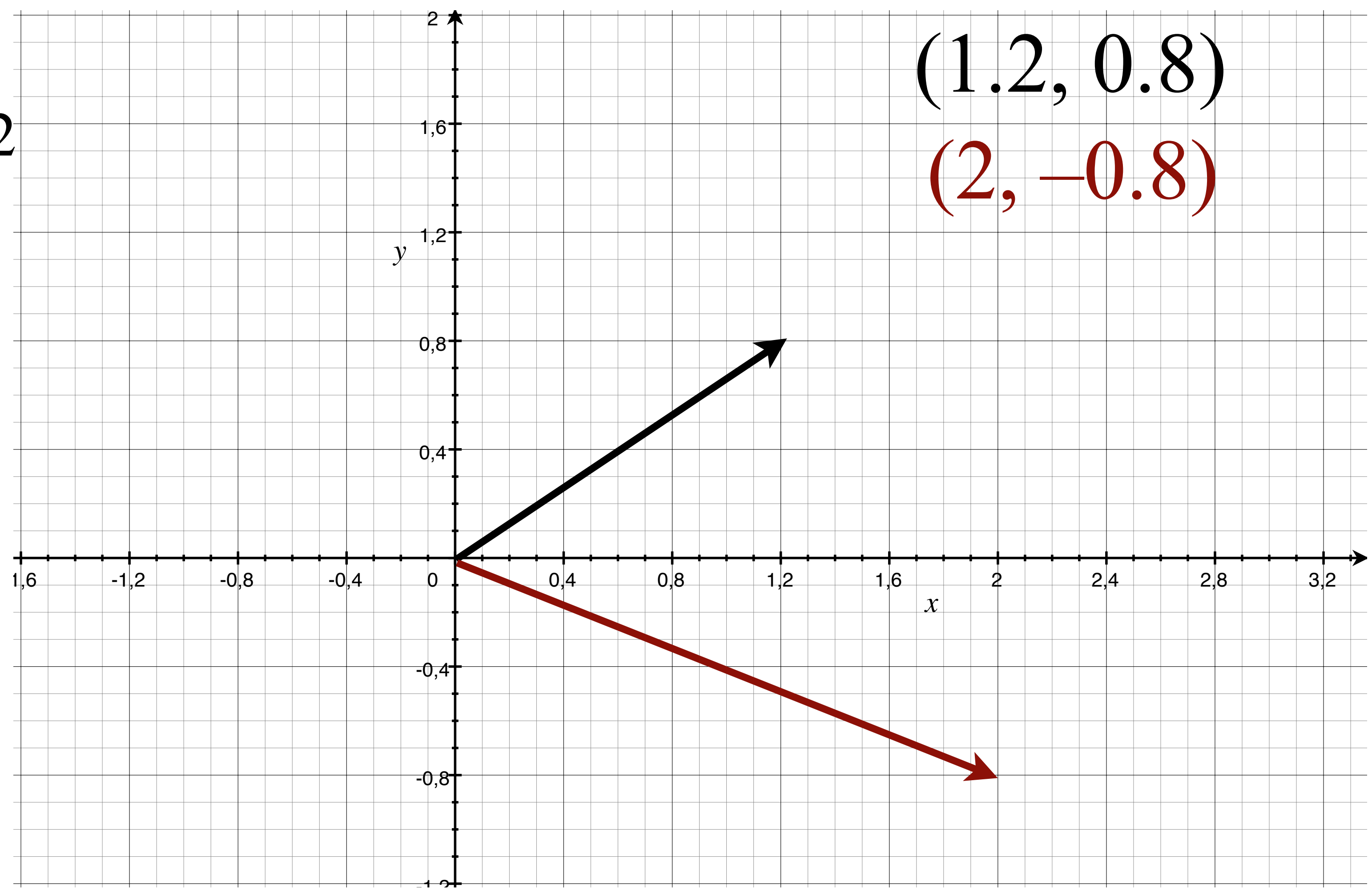


Matrices and vectors

- A **vector** is
  - a 1D array of numbers
  - a geometric entity with magnitude and direction
- The **norm** of the vector defines its magnitude
  - **Euclidean** (L_2) norm:

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

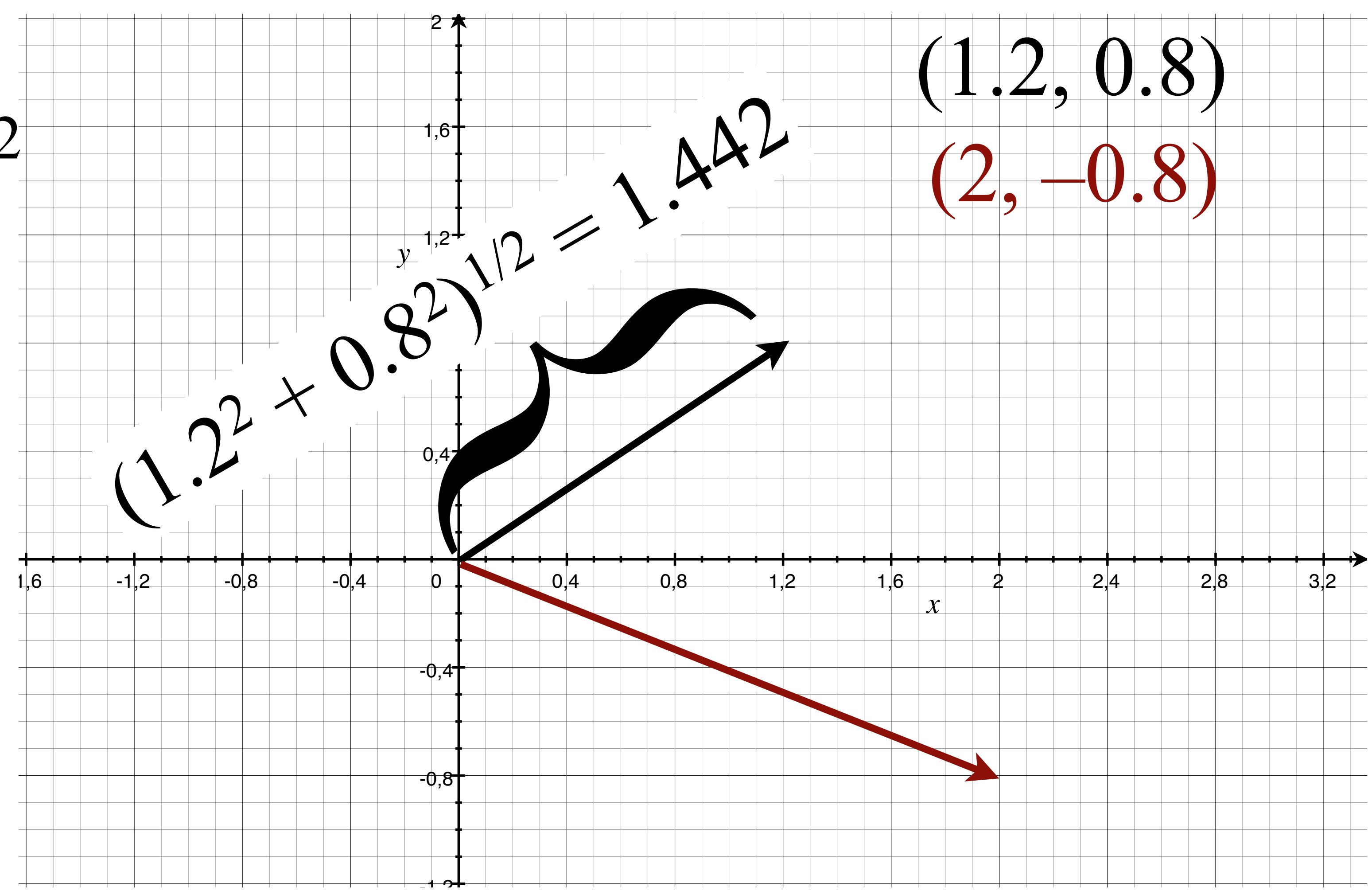
$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$



Matrices and vectors

- A **vector** is
  - a 1D array of numbers
  - a geometric entity with magnitude and direction
- The **norm** of the vector defines its magnitude
  - **Euclidean** (L_2) norm:

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$



$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

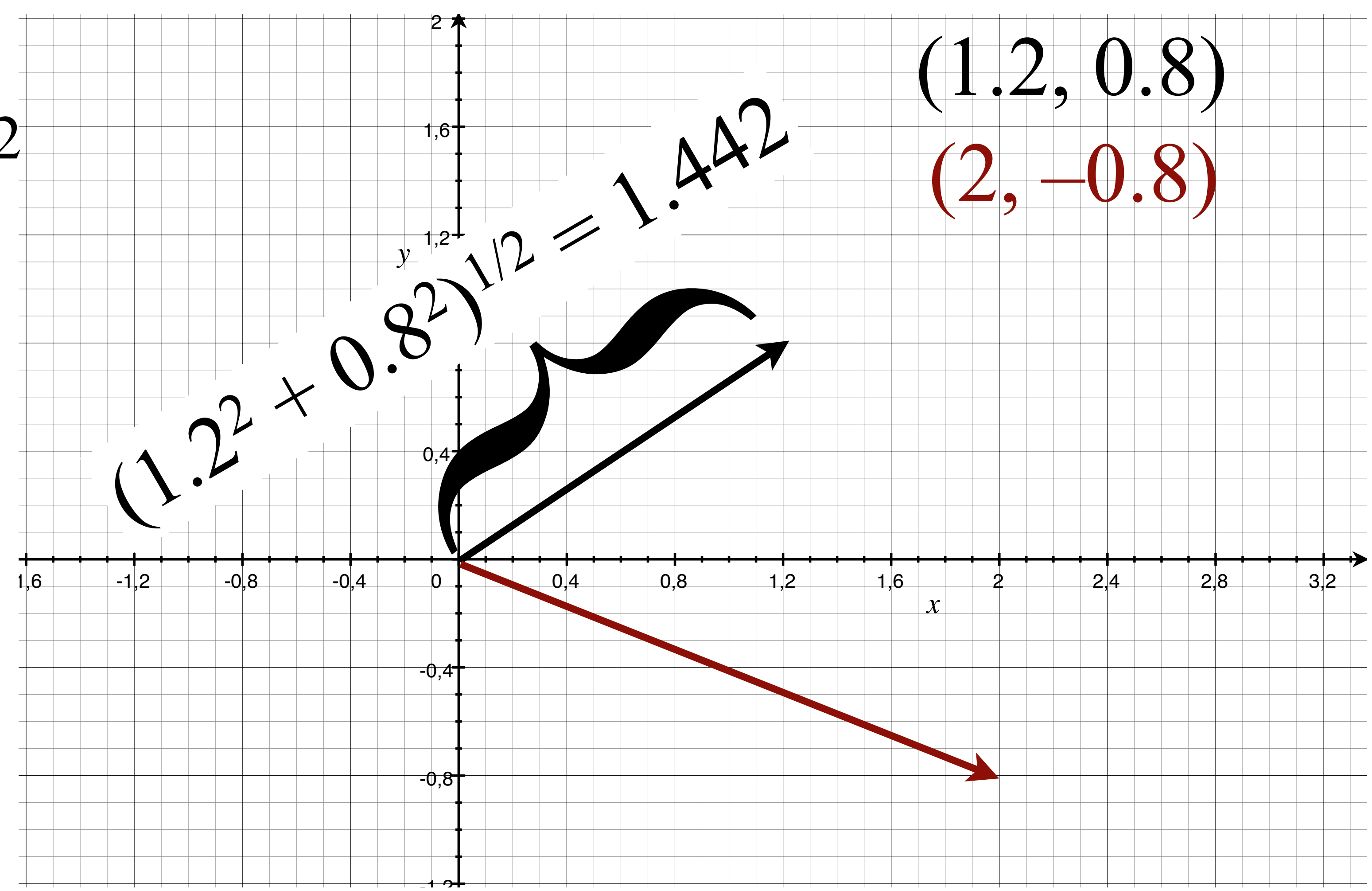
Matrices and vectors

- A **vector** is
  - a 1D array of numbers
  - a geometric entity with magnitude and direction
- The **norm** of the vector defines its magnitude
  - **Euclidean** (L_2) norm:

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

  - L_p norm ($1 \leq p \leq \infty$)

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$



Matrices and vectors

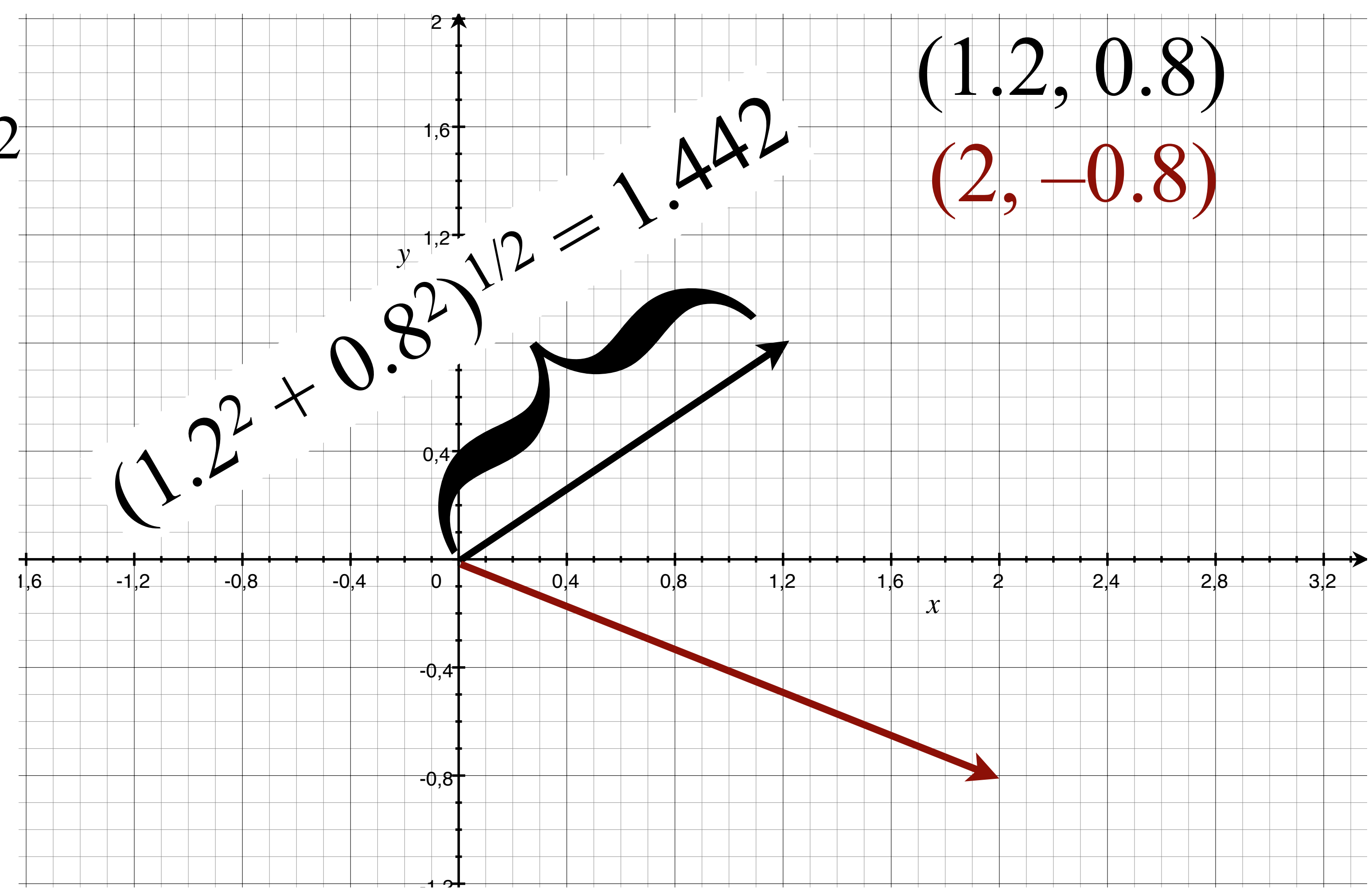
- A **vector** is
  - a 1D array of numbers
  - a geometric entity with magnitude and direction
- The **norm** of the vector defines its magnitude
  - **Euclidean** (L_2) norm:

$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

  - L_p norm ($1 \leq p \leq \infty$)

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

- The direction is the angle



Matrices and vectors

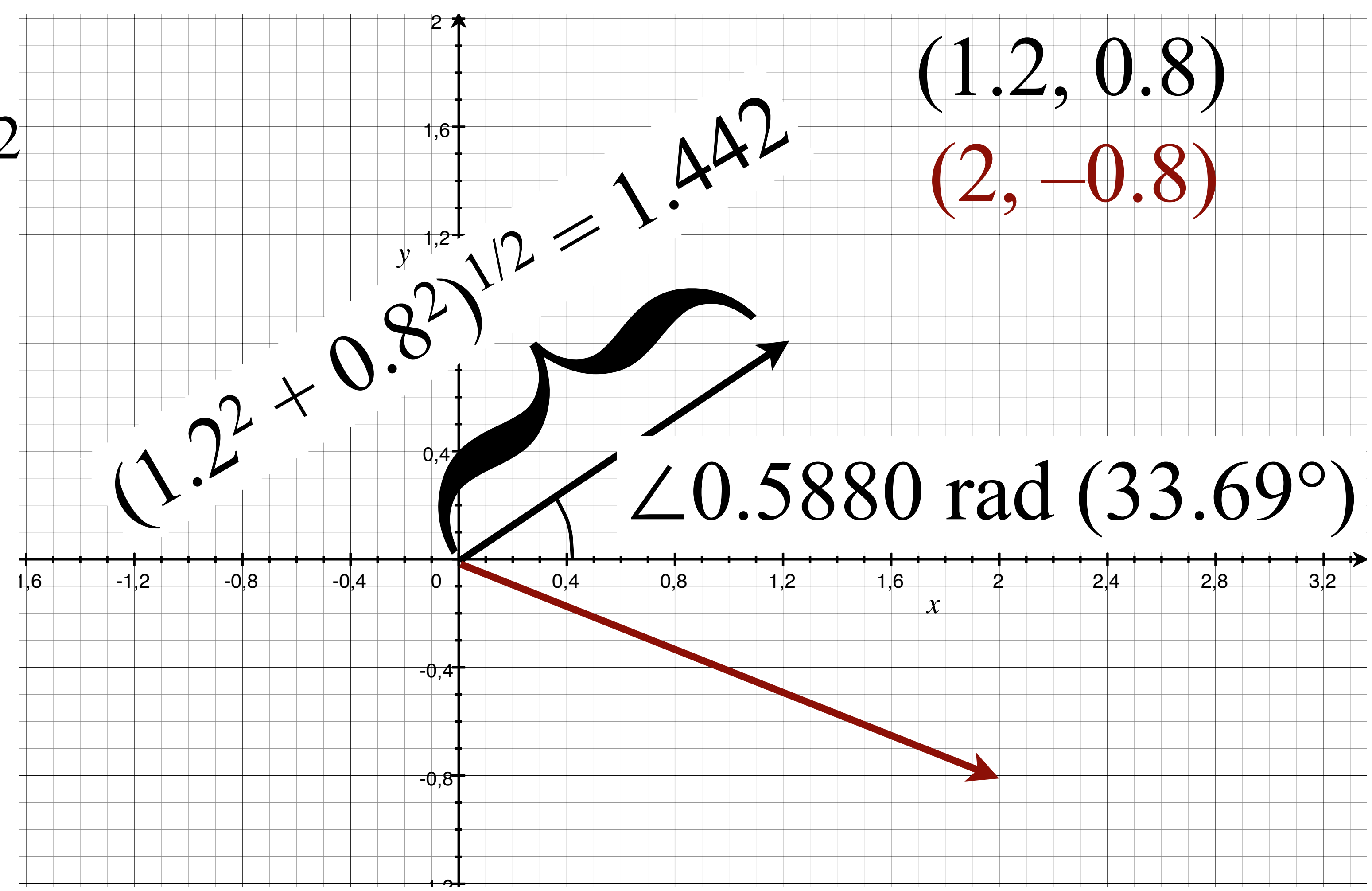
- A **vector** is
  - a 1D array of numbers
  - a geometric entity with magnitude and direction
- The **norm** of the vector defines its magnitude
  - **Euclidean** (L_2) norm:

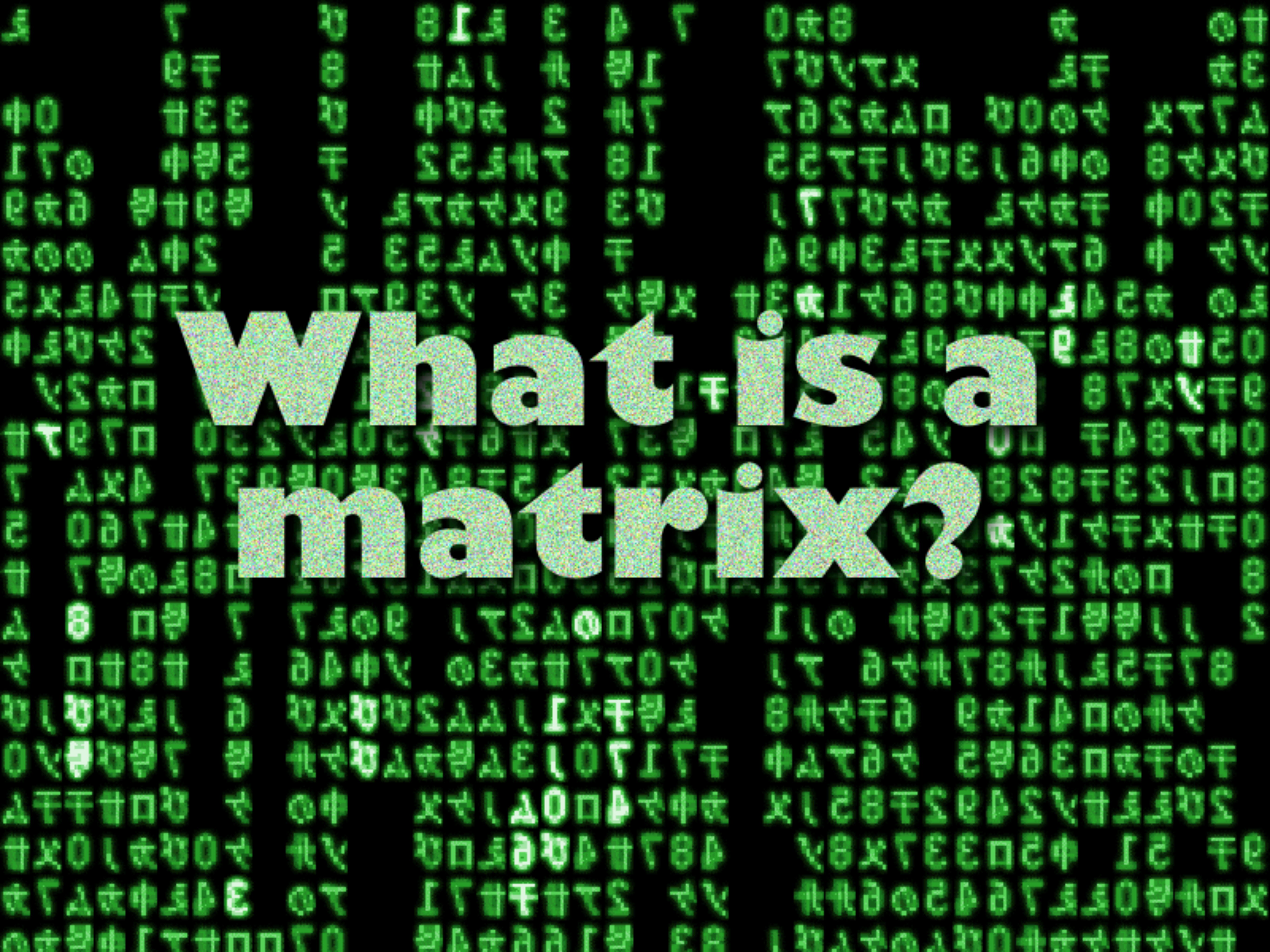
$$\|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$$

  - L_p norm ($1 \leq p \leq \infty$)

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

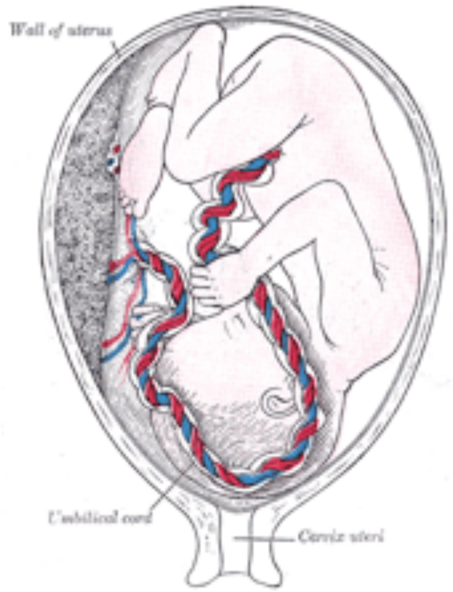
- The direction is the angle




What is a matrix?

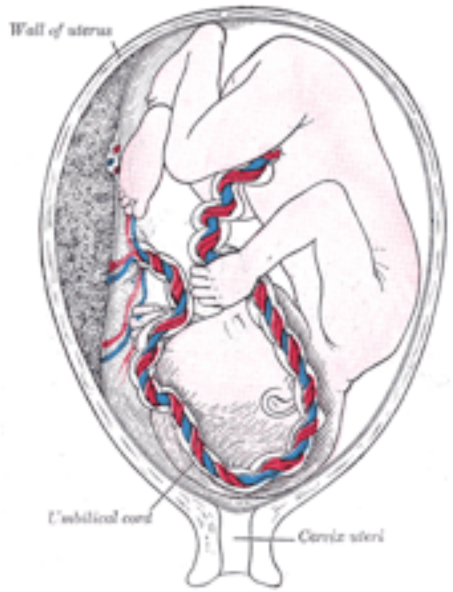
Which of the following are matrices?

Which of the following are matrices?



A womb

Which of the following are matrices?

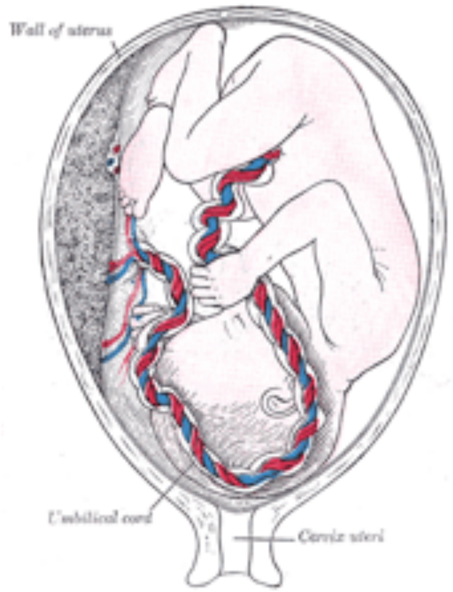


A womb

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

A rectangular array of numbers

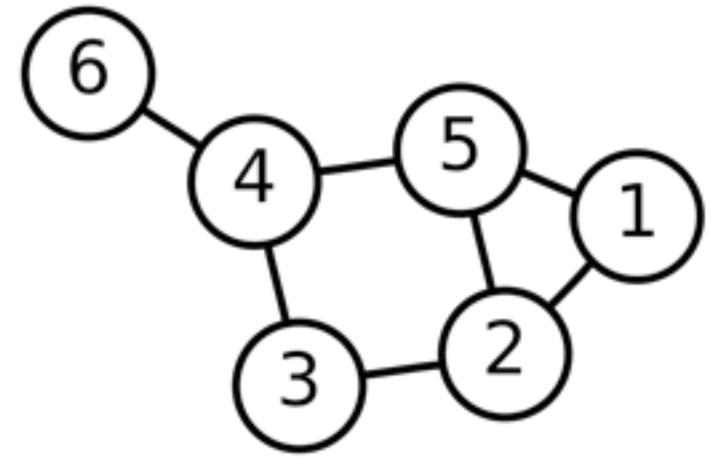
Which of the following are matrices?



A womb

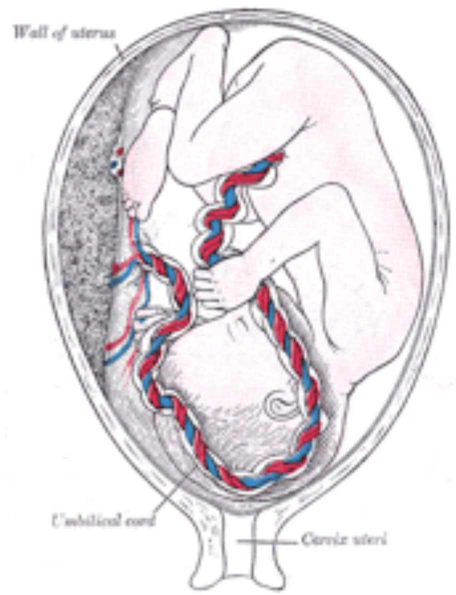
$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

A rectangular array of numbers



A graph

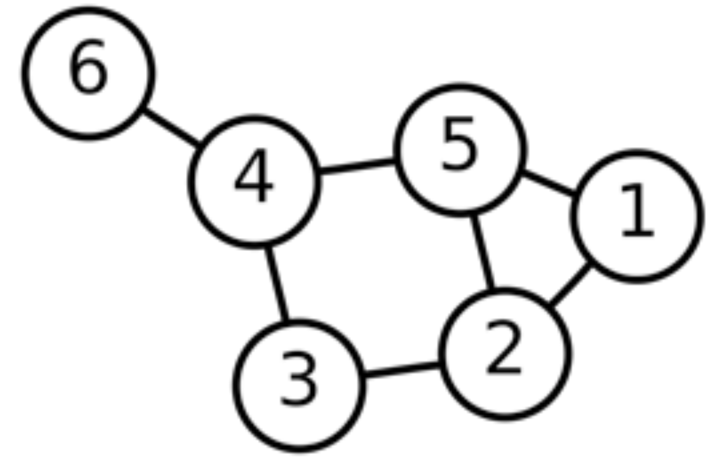
Which of the following are matrices?



A womb

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

A rectangular array of numbers

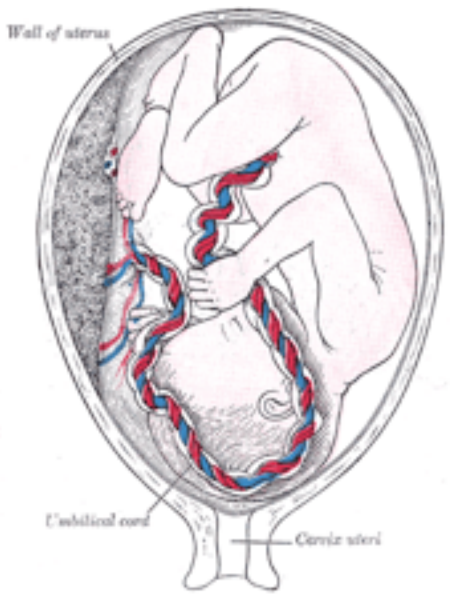


A graph

$$3x + 2y + z = 39$$
$$2x + 3y + z = 34$$
$$x + 2y + 3z = 26$$

A system of linear equations

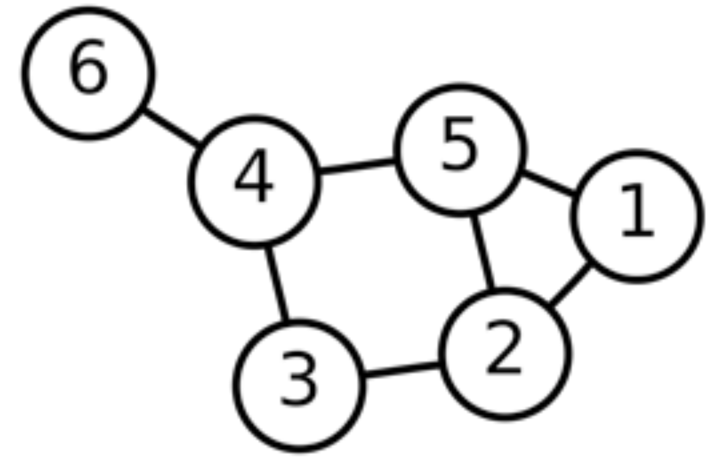
Which of the following are matrices?



A womb

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

A rectangular array of numbers



A graph

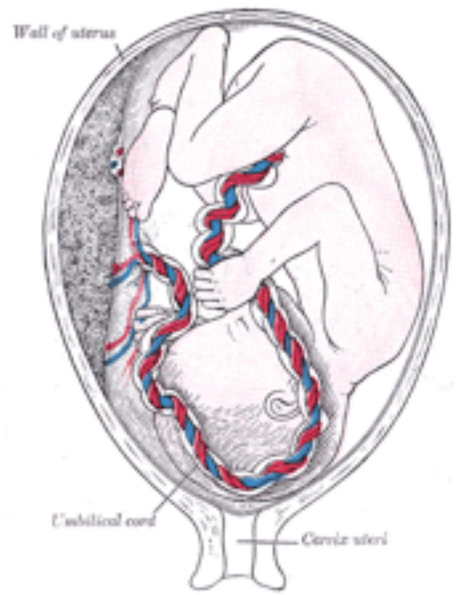
$$\begin{aligned} 3x + 2y + z &= 39 \\ 2x + 3y + z &= 34 \\ x + 2y + 3z &= 26 \end{aligned}$$

A system of linear equations

$$\begin{aligned} f_1(x, y, z) &= 3x + 2y + z \\ f_2(x, y, z) &= 2x + 3y + z \\ f_3(x, y, z) &= x + 2y + 3z \\ f_4(x, y, z) &= x \end{aligned}$$

A linear mapping

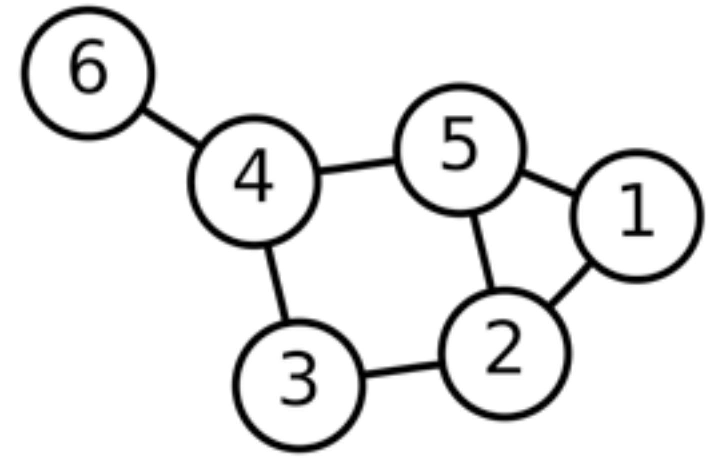
Which of the following are matrices?



A womb

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

A rectangular array of numbers



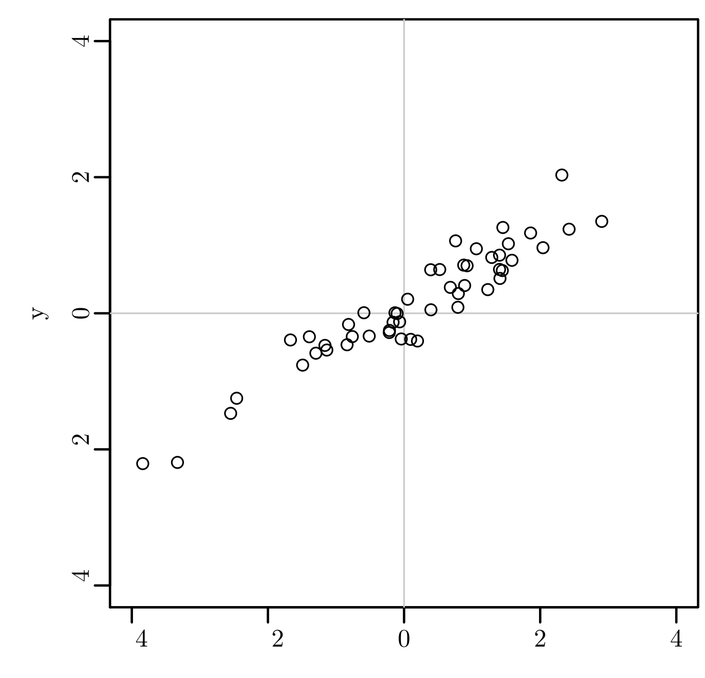
A graph

$$\begin{aligned} 3x + 2y + z &= 39 \\ 2x + 3y + z &= 34 \\ x + 2y + 3z &= 26 \end{aligned}$$

A system of linear equations

$$\begin{aligned} f_1(x, y, z) &= 3x + 2y + z \\ f_2(x, y, z) &= 2x + 3y + z \\ f_3(x, y, z) &= x + 2y + 3z \\ f_4(x, y, z) &= x \end{aligned}$$

A linear mapping



A set of data points

Vectors in IR&DM

- All above meanings of matrices and vectors (and more) are important ways to understand them
  - Different intuitions provide different insights
- In IR&DM, the most important one is the **vector space model**
  - A document in a vocabulary of n terms is represented as an n-dimensional vector
  - A customer transaction in a supermarket selling n items is represented as an n-dimensional vector

Vectors in IR&DM

- All above meanings of matrices and vectors (and more) are important ways to understand them
  - Different intuitions provide different insights
- In IR&DM, the most important one is the **vector space model**
  - A document in a vocabulary of *n* terms is represented as an *n*-dimensional vector
  - A customer transaction in a supermarket selling *n* items is represented as an *n*-dimensional vector

$$\begin{array}{cccccc} \text{Data} & \text{Star} & \text{Trek} & \text{Google} & \text{Information} & \\ (5, & 0, & 0, & 1, & 3, & \ldots) \end{array}$$

Matrices in IR&DM

$$\begin{array}{l} \\ \text{Anna} \\ \text{Bob} \\ \text{Charlie} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Bread} & \text{Butter} & \text{Beer} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \end{array}$$

Customer transactions

$$\begin{array}{l} \\ \text{Book 1} \\ \text{Book 2} \\ \text{Book 3} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Data} & \text{Matrix} & \text{Mining} \end{array} \\ \begin{pmatrix} 5 & 0 & 3 \\ 0 & 0 & 7 \\ 4 & 6 & 5 \end{pmatrix} \end{array}$$

Document-term matrix

$$\begin{array}{l} \\ \text{Alice} \\ \text{Bob} \\ \text{Charlie} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Avatar} & \text{The Matrix} & \text{Up} \end{array} \\ \begin{pmatrix} & 4 & 2 \\ 3 & 2 & \\ 5 & & 3 \end{pmatrix} \end{array}$$

Incomplete rating matrix

$$\begin{array}{l} \\ \text{Saarbrücken} \\ \text{Helsinki} \\ \text{Cape Town} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Jan} & \text{Jun} & \text{Sep} \end{array} \\ \begin{pmatrix} 1 & 11 & 10 \\ 6.5 & 10.9 & 8.7 \\ 15.7 & 7.8 & 8.7 \end{pmatrix} \end{array}$$

Cities and monthly temperatures

Basic operations on vectors

- A **transpose** $\boldsymbol{v}^T$ transposes a row vector into a column vector and vice versa
- If $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$, $\boldsymbol{v} + \boldsymbol{w}$ is a vector with $(\boldsymbol{v} + \boldsymbol{w})_i = v_i + w_i$
- For vector $\boldsymbol{v}$ and scalar α, $(\alpha\boldsymbol{v})_i = \alpha\boldsymbol{v}_i$
- A **dot product** of two vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ is
$$\boldsymbol{v} \cdot \boldsymbol{w} = \sum_{i=1}^{n} v_i w_i$$
  - A.k.a. **scalar product** or **inner product**
  - Alternative notations: $\langle \boldsymbol{v}, \boldsymbol{w} \rangle$, $\boldsymbol{v}^T\boldsymbol{w}$ (for column vectors), $\boldsymbol{v}\boldsymbol{w}^T$ (for row vectors)
  - In Euclidean space $\boldsymbol{v} \cdot \boldsymbol{w} = \|\boldsymbol{v}\| \, \|\boldsymbol{w}\| \cos\theta$

Basic operations on matrices

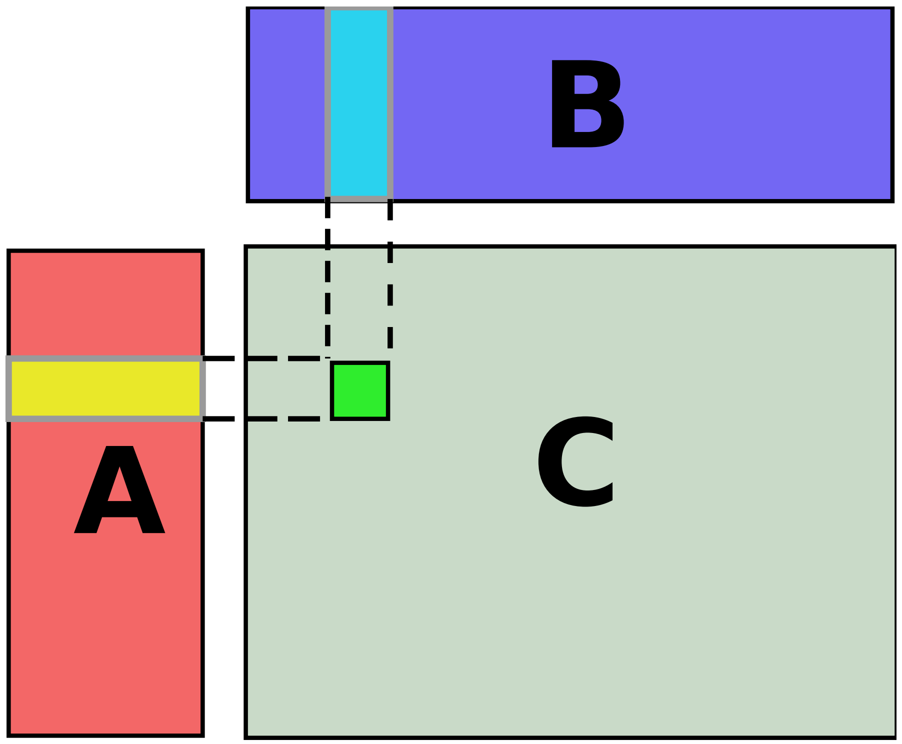
- Matrix transpose $\boldsymbol{A}^T$ has the rows of $\boldsymbol{A}$ as its columns
- If $\boldsymbol{A}$ and $\boldsymbol{B}$ are n-by-m matrices, then $\boldsymbol{A} + \boldsymbol{B}$ is an n-by-m matrix with $(\boldsymbol{A} + \boldsymbol{B})_{ij} = m_{ij} + n_{ij}$
- If $\boldsymbol{A}$ is n-by-k and $\boldsymbol{B}$ is k-by-m, then $\boldsymbol{AB}$ is an n-by-m matrix with

$$(\boldsymbol{AB})_{ij} = \sum_{\ell=1}^{k} a_{i\ell} b_{\ell j}$$

  - The inner dimension (k) must agree
  - **Vector outer product** $\boldsymbol{vw}^T$ (for column vectors) is the matrix product of n-by-1 and 1-by-m matrices

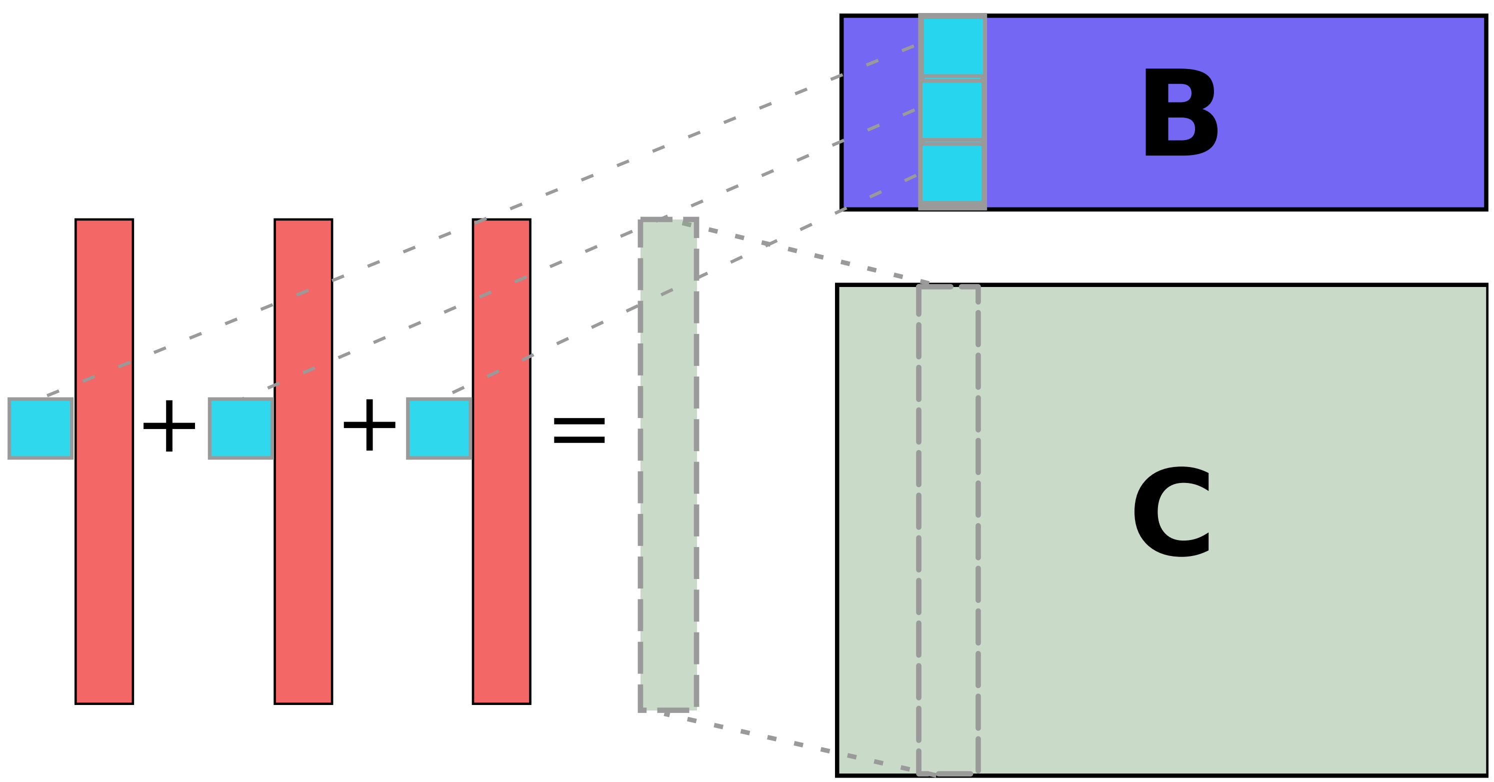
Intuition for matrix multiplication

- Element $(\boldsymbol{AB})_{ij}$ is the inner product of row i of $\boldsymbol{A}$ and column j of $\boldsymbol{B}$



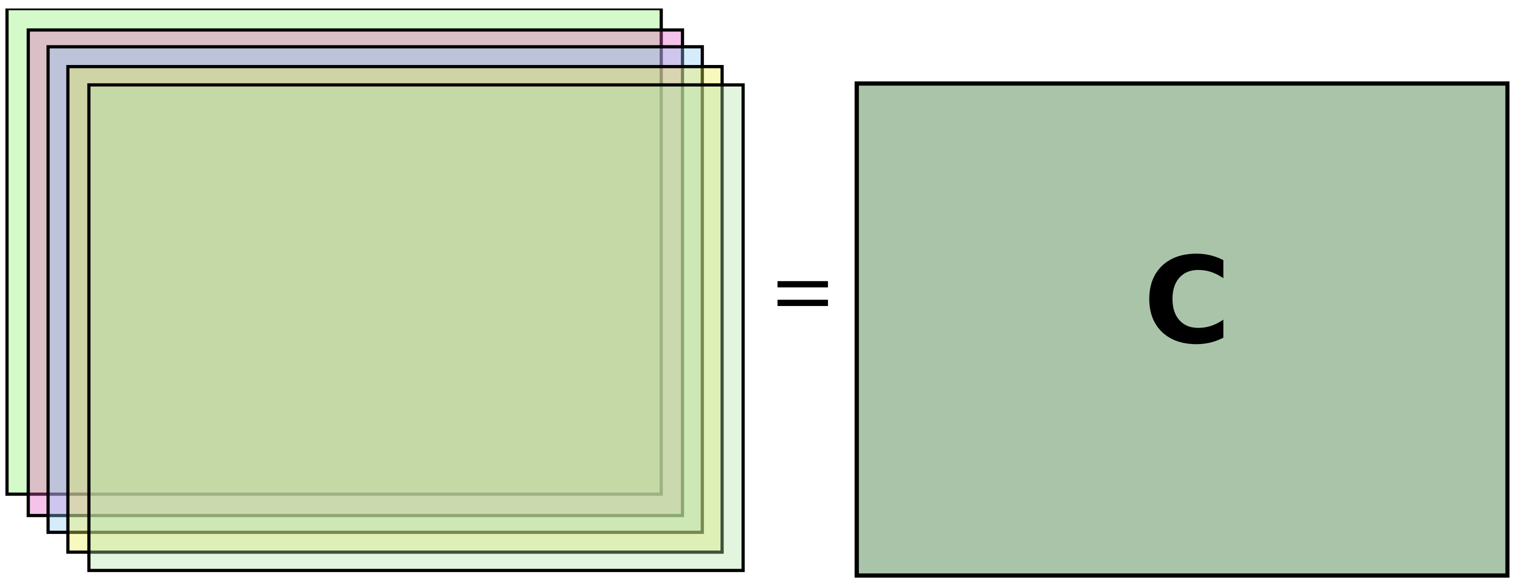
Intuition for matrix multiplication

- Element $(\boldsymbol{AB})_{ij}$ is the inner product of row i of $\boldsymbol{A}$ and column j of $\boldsymbol{B}$
- Column j of $\boldsymbol{AB}$ is the linear combination of columns of $\boldsymbol{A}$ with the coefficients coming from column j of $\boldsymbol{B}$



Intuition for matrix multiplication

- Element $(\boldsymbol{AB})_{ij}$ is the inner product of row i of $\boldsymbol{A}$ and column j of $\boldsymbol{B}$
- Column j of $\boldsymbol{AB}$ is the linear combination of columns of $\boldsymbol{A}$ with the coefficients coming from column j of $\boldsymbol{B}$
- Matrix $\boldsymbol{AB}$ is a sum of k matrices $\boldsymbol{a}_l\boldsymbol{b}_l^T$ obtained by multiplying the l-th column of $\boldsymbol{A}$ with the l-th row of $\boldsymbol{B}$



Ring of n-by-n matrices

- Square matrices of same size form a **ring**
  - Operations are addition, subtraction, and multiplication
  - The identity for addition and subtraction (0) is the all-zeros matrix $\boldsymbol{0}$
  - Multiplication doesn't always have an inverse (division)
  - Multiplication isn't commutative ($\boldsymbol{AB} \neq \boldsymbol{BA}$ in general)
  - The identity for multiplication is the **identity matrix** $\boldsymbol{I}$ with 1s on the main diagonal and 0s elsewhere
    - $(\boldsymbol{I})_{ij} = 1$ iff $i = j$; $(\boldsymbol{I})_{ij} = 0$ otherwise
- With these limitations, we can do algebra on matrices as with real numbers

Matrices as linear mappings

- An n-by-m matrix $\boldsymbol{A}$ is a linear mapping from m-dimensional vector space to n-dimensional vector space
  - $A(\boldsymbol{x}) = \boldsymbol{Ax}$, $(\boldsymbol{Ax})_i = \sum_{j=1}^{m} a_{ij} x_j$
  - The transpose $\boldsymbol{A}^T$ is a mapping from n-dimensional to m-dimensional vector space
    - Typically it does **not** hold that if $\boldsymbol{y} = \boldsymbol{Ax}$, then $\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{y}$
- If $\boldsymbol{A}$ is n-by-m and $\boldsymbol{B}$ is m-by-k, then for the concatenated product $(A \circ B)(\boldsymbol{x}) = \boldsymbol{A}(\boldsymbol{Bx}) = (\boldsymbol{AB})\boldsymbol{x}$

Types of matrices

- Diagonal *n*-by-*n* matrix
  - Identity matrix $\boldsymbol{I}_n$ is a diagonal *n*-by-*n* matrix with 1s in diagonal

$$\begin{pmatrix} x_{1,1} & 0 & 0 & & 0 \\ 0 & x_{2,2} & 0 & \cdots & 0 \\ 0 & 0 & x_{3,3} & & 0 \\ & \vdots & & \ddots & \\ 0 & 0 & 0 & & x_{n,n} \end{pmatrix}$$

- Upper triangular matrix
  - Lower triangular is the transpose
  - If diagonal is full of 0s, matrix is *strictly triangular*

$$\begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} & & x_{1,n} \\ 0 & x_{2,2} & x_{2,3} & \cdots & x_{2,n} \\ 0 & 0 & x_{3,3} & & x_{3,n} \\ & \vdots & & \ddots & \\ 0 & 0 & 0 & & x_{n,n} \end{pmatrix}$$

- Permutation matrix
  - Each row and column has exactly one 1, rest are 0
- Symmetric matrix: $\boldsymbol{M} = \boldsymbol{M}^T$

Matrix distances and norms

- Frobenius norm $||\boldsymbol{X}||_F = (\sum_{i,j} x_{ij}^2)^{1/2}$
  - Corresponds to Euclidean norm of vectors
- Sum of absolute values $|\boldsymbol{X}| = \sum_{i,j} |x_{ij}|$
  - Corresponds to L_1-norm of vectors
- The above elementwise norms are sometimes (imprecisely) called L_2 and L_1 norms
  - Matrix L_1 and L_2 norms are something different altogether
- Operator norm $||\boldsymbol{X}||_p = \max_{\boldsymbol{y} \neq 0} ||\boldsymbol{X}\boldsymbol{y}||_p / ||\boldsymbol{y}||_p$
  - Largest norm of an image of a unit norm vector
  - $||\boldsymbol{X}||_2 \leq ||\boldsymbol{X}||_F \leq \sqrt{(\mathrm{rank}(\boldsymbol{X}))}\ ||\boldsymbol{X}||_2$

Basic concepts

- Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthogonal** if their inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is 0
  - Vectors are **orthonormal** if they have unit norm, $\|\boldsymbol{x}\|=\|\boldsymbol{y}\|=1$
  - In Euclidean space, this means that $\|\boldsymbol{x}\|\ \|\boldsymbol{y}\| \cos\theta = 0$ which happens iff $\cos\theta = 0$ which means $\boldsymbol{x}$ and $\boldsymbol{y}$ are perpendicular to each other
- A square matrix $\boldsymbol{X}$ is **orthogonal** if its rows and columns are orthonormal
  - An n-by-m matrix $\boldsymbol{X}$ is row-orthogonal if $n < m$ and its rows are orthogonal and column-orthogonal if $n > m$ and its columns are orthogonal

Linear independency

- Vector $\boldsymbol{v} \in \mathbb{R}^n$ is **linearly dependent** from a set of vectors $W = \{\boldsymbol{w}_i \in \mathbb{R}^n : i = 1, \ldots, m\}$ if there exists a set of coefficients α_i such that $\boldsymbol{v} = \sum_{i=1}^{m} \alpha_i \boldsymbol{w}_i$
  - If $\boldsymbol{v}$ is not linearly dependent, it is **linearly independent**
  - That is, $\boldsymbol{v}$ can't be expressed as a linear combination of the vectors in W
- A set of vectors $V = \{v_i \in \mathbb{R}^n : i = 1, \ldots, m\}$ is **linearly independent** if $\boldsymbol{v}_i$ is linearly independent from $V \setminus \{\boldsymbol{v}_i\}$ for all i

Matrix rank

- The **column rank** of an n-by-m matrix $\boldsymbol{M}$ is the number of linearly independent columns of $\boldsymbol{M}$
- The **row rank** is the number of linearly independent rows of $\boldsymbol{M}$
- The **Schein rank** of $\boldsymbol{M}$ is the least integer k such that $\boldsymbol{M} = \boldsymbol{AB}$ for some n-by-k matrix $\boldsymbol{A}$ and k-by-m matrix $\boldsymbol{B}$
  - Equivalently, the least k such that $\boldsymbol{M}$ is a sum of k vector outer products
- *All these ranks are equivalent*!

The matrix inverse

- The **inverse** of a matrix $\boldsymbol{M}$ is the unique matrix $\boldsymbol{N}$ for which $\boldsymbol{MN} = \boldsymbol{NM} = \boldsymbol{I}$
  - The inverse is denoted by $\boldsymbol{M}^{-1}$
- $\boldsymbol{M}$ has an inverse (is invertible) iff
  - $\boldsymbol{M}$ is square (n-by-n)
  - The rank of $\boldsymbol{M}$ is n (full rank)
- Non-square matrices can have left or right inverses
  - $\boldsymbol{MR} = \boldsymbol{I}$ or $\boldsymbol{LA} = \boldsymbol{I}$
- If $\boldsymbol{M}$ is orthogonal, then (and only then) $\boldsymbol{M}^{-1} = \boldsymbol{M}^T$
  - That is, $\boldsymbol{MM}^T = \boldsymbol{M}^T\boldsymbol{M} = \boldsymbol{I}$

The matrix pseudo-inverse

- The Moore–Penrose **pseudo-inverse** of an n-by-m matrix $\boldsymbol{M}$ is an m-by-n matrix $\boldsymbol{M}^+$ for which
  - $\boldsymbol{MM}^+\boldsymbol{M} = \boldsymbol{M}$ ($\boldsymbol{MM}^+$ doesn't have to be identity)
  - $\boldsymbol{M}^+\boldsymbol{MM}^+ = \boldsymbol{M}^+$ ($\boldsymbol{M}^+\boldsymbol{M}$ doesn't have to be identity)
  - $(\boldsymbol{MM}^+)^T = \boldsymbol{MM}^+$ ($\boldsymbol{MM}^+$ is symmetric)
  - $(\boldsymbol{M}^+\boldsymbol{M})^T = \boldsymbol{M}^+\boldsymbol{M}$ ($\boldsymbol{M}^+\boldsymbol{M}$ is symmetric)
- If the rank of $\boldsymbol{M}$ is m (full column rank, $n \geq m$), then $\boldsymbol{M}^+ = (\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T$
  - If the rank of $\boldsymbol{M}$ is n ($n \leq m$), then $\boldsymbol{M}^+ = \boldsymbol{M}^T(\boldsymbol{MM}^T)^{-1}$

Fundamental decompositions

- A **matrix decomposition** (or **factorization**) presents an n-by-m matrix $\boldsymbol{A}$ as a product of two (or more) **factor matrices**
  - $\boldsymbol{A} = \boldsymbol{BC}$
- For approximate decompositions, $\boldsymbol{A} \approx \boldsymbol{BC}$
- The size of the decomposition is the inner dimension of $\boldsymbol{B}$ and $\boldsymbol{C}$
  - Number of columns in $\boldsymbol{B}$ and number of rows in $\boldsymbol{C}$
  - For exact decompositions, the size is no less than the rank of the matrix

Eigenvalues and eigenvectors

- If $\boldsymbol{X}$ is an n-by-n matrix and $\boldsymbol{v}$ is a vector such that $\boldsymbol{Xv} = \lambda\boldsymbol{v}$ for some scalar λ, then
  - λ is an **eigenvalue** of $\boldsymbol{X}$
  - $\boldsymbol{v}$ is an **eigenvector** of $\boldsymbol{X}$ associated to λ
- That is, eigenvectors are those vectors $\boldsymbol{v}$ for which $\boldsymbol{Xv}$ only changes their magnitude, not direction
  - It is possible to exactly reverse the direction
  - The change in magnitude is the eigenvalue
- If $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{X}$ and α is a scalar, then $\alpha\boldsymbol{v}$ is also an eigenvector with the same eigenvalue

Properties of eigenvalues

- Multiple linearly independent eigenvectors can be associated with the same eigenvalue
  - The **algebraic multiplicity** of the eigenvalue
- Every n-by-n matrix has n eigenvectors and n eigenvalues (counting the multiplicity)
  - But some of these can be complex numbers
- If a matrix is symmetric, then all its eigenvalues are real
- Matrix is invertible iff all its eigenvalues are non-zero

Eigendecomposition

- The (real-valued) **eigendecomposition** of an n-by-n matrix $\boldsymbol{X}$ is $\boldsymbol{X} = \boldsymbol{Q\Lambda Q}^{-1}$
  - $\boldsymbol{\Lambda}$ is a diagonal matrix with eigenvalues in the diagonal
  - Columns of $\boldsymbol{Q}$ are the eigenvectors associated with the eigenvalues in $\boldsymbol{\Lambda}$
- Matrix $\boldsymbol{X}$ has to be diagonalizable
  - $\boldsymbol{PXP}^{-1}$ is a diagonal matrix for some invertible matrix $\boldsymbol{P}$
- Matrix $\boldsymbol{X}$ has to have n real eigenvalues (counting multiplicity)

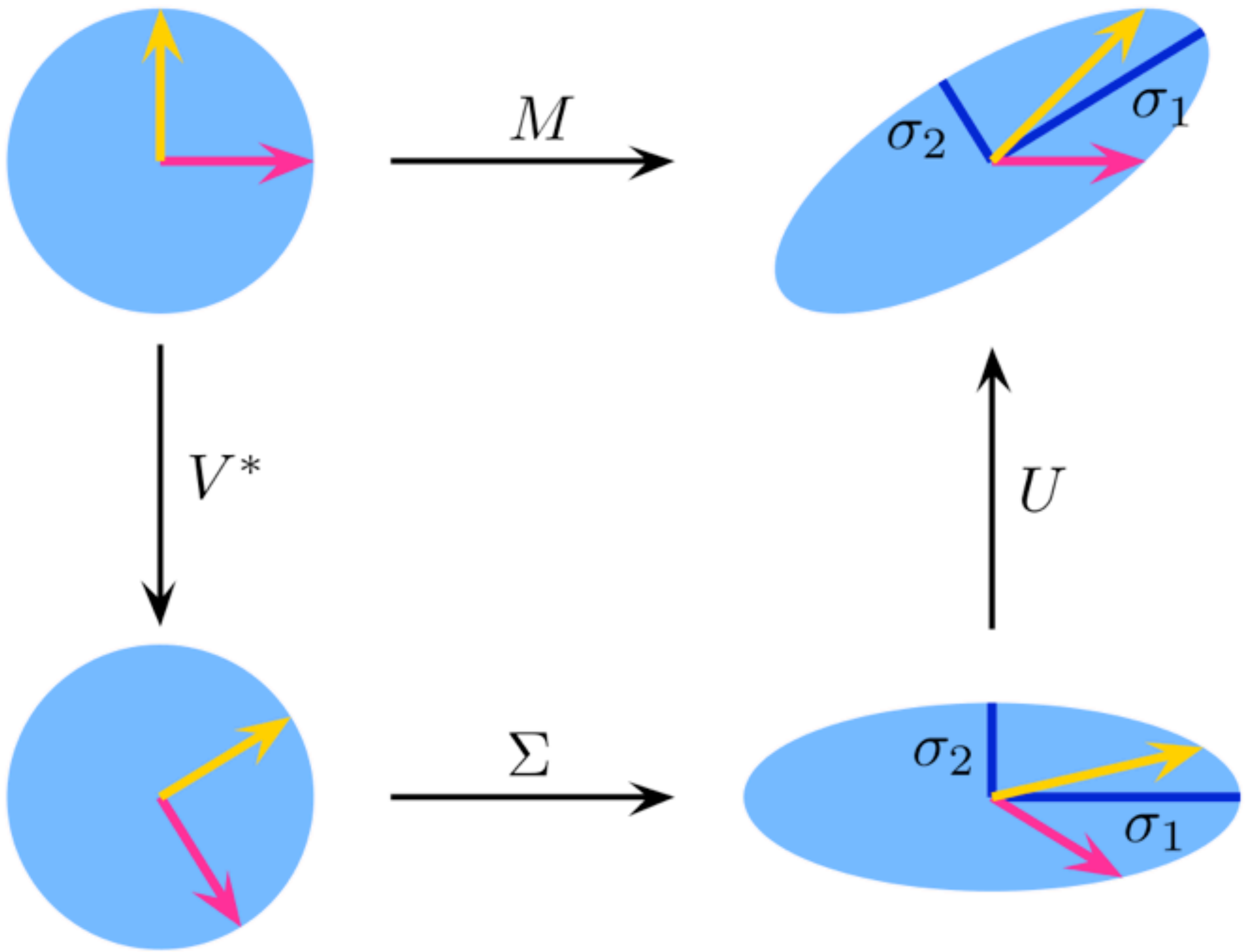
Some useful facts

- Not all matrices have eigendecomposition
  - Not all invertible matrices have eigendecomposition
  - Not all matrices that have eigendecomposition are invertible
  - If $\boldsymbol{X}$ is invertible and has eigendecomposition, then $\boldsymbol{X}^{-1} = \boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^{-1}$
- If $\boldsymbol{X}$ is symmetric and invertible (and real), then $\boldsymbol{X}$ has eigendecomposition $\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^{T}$

Singular value decomposition (SVD)

- Not every matrix has eigendecomposition, but: **Theorem.** If $\boldsymbol{X}$ is n-by-m real matrix, there exists n-by-n orthogonal matrix $\boldsymbol{U}$ and m-by-m orthogonal matrix $\boldsymbol{V}$ such that $\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V}$ is n-by-m matrix $\boldsymbol{\Sigma}$ with values $\sigma_1, \sigma_2, \ldots, \sigma_{\min(n,m)}$, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\min(n,m)} \geq 0$, in its diagonal.
  - In other words, $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$
  - Values σ_i are the **singular values** of $\boldsymbol{X}$
  - Columns of $\boldsymbol{U}$ are the **left singular vectors** and columns of $\boldsymbol{V}$ the **right singular vectors** of $\boldsymbol{X}$

Example



$$M = U \cdot \Sigma \cdot V^*$$

Rank and SVD

- rank($\boldsymbol{X}$) = r iff $\boldsymbol{X}$ has exactly r non-zero singular values
  - $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_{\min(n,m)} = 0$
  - A method to compute the rank of a matrix
- A **truncated SVD** of rank k is obtained by setting all but the first k singular values to 0
  - Typically denoted as $\boldsymbol{U}_k \boldsymbol{\Sigma}_k \boldsymbol{V}_k^T$
  - For the product, we can ignore the columns of U and V corresponding to the zero singular values
    - $\boldsymbol{U}_k$ is n-by-k, $\boldsymbol{V}_k$ is m-by-k, and $\boldsymbol{\Sigma}_k$ is k-by-k

Properties of SVD

- If $\boldsymbol{X}$ is rank-r, then $\mathbf{X} = \sum_{i=1}^{r} \sigma_i u_i v_i^T$
  - $\boldsymbol{X}$ is a sum of r rank-1 matrices scaled with singular values
- $\|\mathbf{X}\|_F^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_{\min(n,m)}^2$
- $\|\mathbf{X}\|_2 = \sigma_1$
- **Eckart–Young theorem.** Let $\boldsymbol{X}$ be of rank-r and let $\boldsymbol{U\Sigma V}^T$ be its SVD. Denote by $\boldsymbol{U}_k$ the first k columns of $\boldsymbol{U}$, by $\boldsymbol{V}_k$ the first k columns of $\boldsymbol{V}$ and by $\boldsymbol{\Sigma}_k$ the upper-left k-by-k corner of $\boldsymbol{\Sigma}$. Then $\boldsymbol{X}_k = \boldsymbol{U}_k\boldsymbol{\Sigma}_k\boldsymbol{V}_k^T$ is the best rank-k approximation of $\boldsymbol{X}$ in the sense that $\|\mathbf{X} - \mathbf{X}_k\|_F \leqslant \|\mathbf{X} - \mathbf{Y}\|_F$ and $\|\mathbf{X} - \mathbf{X}_k\|_2 \leqslant \|\mathbf{X} - \mathbf{Y}\|_2$ for any rank-k matrix $\boldsymbol{Y}$.

SVD and pseudo-inverse

- Recall that if $\boldsymbol{X}$ is n-by-m with $\text{rank}(\boldsymbol{X}) = m \le n$, the *pseudo-inverse* of $\boldsymbol{X}$ is $\boldsymbol{X}^+ = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$
- If $\text{rank}(\boldsymbol{X}) = r$ and $\boldsymbol{X} = \boldsymbol{U\Sigma V}^T$, then we can define $\boldsymbol{X}^+ = \boldsymbol{V\Sigma}^+\boldsymbol{U}^T$
  - $\boldsymbol{\Sigma}^+$ is a diagonal matrix with $1/\sigma_i$ in its ith position (or 0 if $\sigma_i = 0$)
  - More general than the above definition
- This gives the least-squares solution to the following problem: given $\boldsymbol{A}$ and $\boldsymbol{X}$, find $\boldsymbol{Y}$ s.t. $\|\boldsymbol{A} - \boldsymbol{XY}\|_F^2$ is minimized
  - Setting $\boldsymbol{Y} = \boldsymbol{X}^+\boldsymbol{A}$ minimizes the squared Frobenius norm

SVD and eigendecomposition

- Let $\boldsymbol{X}$ be n-by-m and $\boldsymbol{X} = \boldsymbol{U\Sigma V}^T$ its SVD
- The **Gram matrix** of the columns of $\boldsymbol{X}$ is $\boldsymbol{X}^T\boldsymbol{X}$
  - For the rows it is $\boldsymbol{XX}^T$
- Now $\boldsymbol{X}^T\boldsymbol{X} = (\boldsymbol{U\Sigma V}^T)^T(\boldsymbol{U\Sigma V}^T) = \boldsymbol{V\Sigma}^T\boldsymbol{U}^T\boldsymbol{U\Sigma V}^T$
  $= \boldsymbol{V\Sigma}^T\boldsymbol{\Sigma V}^T = \boldsymbol{V\Sigma}_m{}^2\boldsymbol{V}^T$
  - $\Sigma_m{}^2$ is an m-by-m diagonal matrix with $\sigma_i{}^2$ in its ith position
- Similarly $\boldsymbol{XX}^T = \boldsymbol{U\Sigma}_n{}^2\boldsymbol{U}^T$
- Therefore
  - Columns of $\boldsymbol{U}$ are the eigenvectors of $\boldsymbol{XX}^T$
  - Columns of $\boldsymbol{V}$ are the eigenvectors of $\boldsymbol{X}^T\boldsymbol{X}$
  - Singular values are square roots of the associated eigenvalues