Linear Algebra

Keng-Chi Chang

Department of Political Science
University of California San Diego

2023-09-10

Motivation

- Linear algebra or matrix algebra avoids the mess and lets us solve for things we care about quickly, cleanly and easily

Motivation

- Linear algebra or matrix algebra avoids the mess and lets us solve for things we care about quickly, cleanly and easily
- This is no different than algebra. Consider the difference in the following formulas for the mean:

$$\bar{x} = \frac{x_1 + x_2}{n}$$

$$\bar{x} = \frac{x_1 + x_2 + x_3}{n}$$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Motivation

- Similarly, matrix algebra is a form of notation that cleans up the mess when working with more complex formulas. So suspend disbelief and concern, and treat this as a new language you are learning.
- Think of this as algebra on steroids.

Motivation

Why are we studying matrix algebra?

- Matrices are an intuitive way to think about data.
- We have a set of observations (perhaps individuals) on the row, and observe many different characteristics (such as race, gender, PID, etc.) corresponding to columns.
- We will use matrix algebra to derive the least squares estimator.
- Matrices are useful for solving systems of equations, like multiple regression.
- Notation is much more compact and concise.

Points and Vectors

- A point in $\mathbb{R}$
  - 1
  - π
  - e
- A point in $\mathbb{R}^2$
  - $(1, 2)$
  - (π, e)
- A point $\mathbf{x}$ in $\mathbb{R}^n$
  - $(x_1, x_2, \cdots, x_n)$
- Vector: arrow pointing from the origin to the point
- Draw some examples in 1d, 2d, row and column vectors

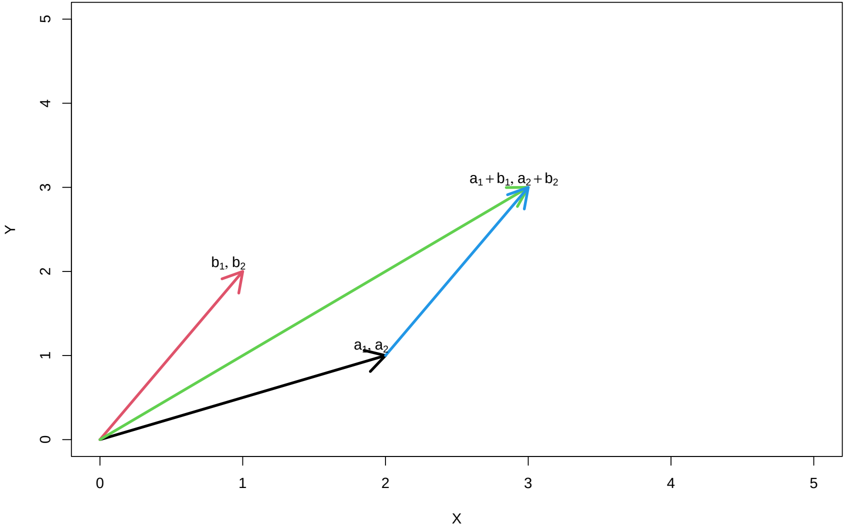
Vector Addition and Subtraction

If two vectors, $\mathbf{u}$ and $\mathbf{v}$, have the same size (i.e. have the same number of elements), they can be added (subtracted) together:

$$\mathbf{u}+\mathbf{v}=\begin{bmatrix} u_1+v_1 \\ u_2+v_2 \\ \vdots \\ u_k+v_n \end{bmatrix} \qquad \mathbf{u}-\mathbf{v}=\begin{bmatrix} u_1-v_1 \\ u_2-v_2 \\ \vdots \\ u_k-v_n \end{bmatrix}$$

- Draw geometrical interpretations of addition and subtraction

Example: Vector Addition



Scalar Multiplication

The product of a scalar c (i.e. a constant) and vector $\mathbf{v}$ is:

$$c\mathbf{v} = \begin{bmatrix} cv_1 \\ cv_2 \\ \dots \\ cv_n \end{bmatrix}$$

- Draw geometrical interpretations scalar multiplication

Exercise

- $\mathbf{a} = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$
- $\mathbf{b} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$
- $\mathbf{a} - \mathbf{b}$=?
- $-3 \cdot (\mathbf{a} - \mathbf{b})$=?

Special Vectors

Zero Vector: A vector of all zeros. Eg. For $\mathbb{R}^3$,

$$\mathbf{0}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Standard Unit Vectors: Vectors whose components are all 0, except one that equals 1.

- For $\mathbb{R}^2$, the standard unit vectors are:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- For $\mathbb{R}^3$, the standard unit vectors are:

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Linear combination

The vector $\mathbf{u}$ is a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$ if

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

Exercise

Represent $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ as linear combination of the unit vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

- Notice that you can always represent one vector into linear combinations of the unit vectors, where the coefficients are the elements in each coordinates.

Matrix

A way to store data. A dataframe is a matrix (with column/row names).

```r
(df = head(mtcars))
```

```
                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```r
dim(df)
```

```
[1]  6 11
```

Matrix

Each row of a matrix is a (row) vector

```
df["Mazda RX4", ] # equivalently mtcars[1, ]
```

```
          mpg cyl disp  hp drat   wt  qsec vs am gear carb
Mazda RX4  21   6  160 110  3.9 2.62 16.46  0  1    4    4
```

Each column of a matrix is also a vector

```
df[ , "mpg"] # equivalently mtcars$mpg or mtcars[ , 1]
```

```
[1] 21.0 21.0 22.8 21.4 18.7 18.1
```

Matrix

A matrix of size $m \times n$ is an array of real numbers arranged in m rows by n columns. The dimensionality of the matrix is defined as the number of rows by the number of columns, $m \times n$.

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

- Number of rows is always the first index
- The element of matrix $\mathbf{A}$ corresponding to *row* i and *column* j is written A_{ij}
- You can think of vectors as special cases of matrices
  - A column vector of length k is a $k \times 1$ matrix
  - A row vector of the same length is a $1 \times k$ matrix.

Exercise

$$\mathbf{W} = \begin{bmatrix} 1 & 3 \\ 2 & -6 \end{bmatrix}$$

- Size?
- We call these kind of matrix “square matrix”

$$\Gamma = \begin{bmatrix} 1 & 4 \\ 1 & 3 \\ 1 & -2 \\ 0 & 3 \end{bmatrix}$$

- Size?

Some Notations

- We can write $\mathbf{A}$ as collection of column vectors:

$$\mathbf{A} = [\mathbf{a}_1 | \mathbf{a}_2 | \dots | \mathbf{a}_n] = \begin{bmatrix} \uparrow & \uparrow & \uparrow & \dots & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \dots & \mathbf{a}_n \\ \downarrow & \downarrow & \downarrow & \dots & \downarrow \end{bmatrix}.$$

- Similarly, we can write $\mathbf{A}$ as a collection of row vectors:

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_m' \end{bmatrix} = \begin{bmatrix} \longleftarrow & \mathbf{a}_1' & \longrightarrow \\ \longleftarrow & \mathbf{a}_2' & \longrightarrow \\ \vdots & \vdots & \vdots \\ \longleftarrow & \mathbf{a}_m' & \longrightarrow \end{bmatrix}.$$

- Sometimes, we will want to refer to both rows and columns in the same context. In these situations, we may use $\mathbf{A}_{i\star}$ to reference the i-th row and $\mathbf{A}_{\star j}$ to reference the j-th column.

Matrix Addition and Subtraction

Let $\mathbf{A}$ and $\mathbf{B}$ be two $m \times n$ matrices.

$$\mathbf{A}+\mathbf{B}=\begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}$$

$$\mathbf{A}-\mathbf{B}=\begin{bmatrix} a_{11}-b_{11} & a_{12}-b_{12} & \cdots & a_{1n}-b_{1n} \\ a_{21}-b_{21} & a_{22}-b_{22} & \cdots & a_{2n}-b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}-b_{m1} & a_{m2}-b_{m2} & \cdots & a_{mn}-b_{mn} \end{bmatrix}$$

- Note that matrices $\mathbf{A}$ and $\mathbf{B}$ must have the same size for addition and subtraction to be defined

Scalar Multiplication

Given the scalar s, the scalar multiplication of $s\mathbf{A}$ is

$$s\mathbf{A} = s\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} sa_{11} & sa_{12} & \cdots & sa_{1n} \\ sa_{21} & sa_{22} & \cdots & sa_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ sa_{m1} & sa_{m2} & \cdots & sa_{mn} \end{bmatrix}$$

Exercise

$$\mathbf{A} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 3 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 3 & 9 \\ -1 & 1 \\ 0 & 2 \end{bmatrix}$$

- Find $\mathbf{A} + \mathbf{B}$
- Find $2 \cdot (\mathbf{A} + \mathbf{B})$

Transpose

The transpose of the $m \times n$ matrix A is the $n \times m$ matrix $\mathbf{A}^\top$ (also written $\mathbf{A}'$) obtained by interchanging the rows and columns of A.
In other words, the (i, j)-th element of $\mathbf{A}^\top$ is the (j, i)-th element of $\mathbf{A}$.

$$(\mathbf{A}^\top)_{ij} = \mathbf{A}_{ji}$$

- Example: $\mathbf{A} = \begin{bmatrix} 4 & -2 & 3 \\ 0 & 5 & -1 \end{bmatrix}$, find $\mathbf{A}^\top$
- Example: $\mathbf{B} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$, find $\mathbf{B}^\top$

Properties

1. $(\mathbf{A}+\mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$
2. $(\mathbf{A}^\top)^\top = \mathbf{A}$
3. $(s\mathbf{A})^\top = s\mathbf{A}^\top$
4. $(\mathbf{AB})^\top = \mathbf{B}^\top\mathbf{A}^\top$
   - Also, $(\mathbf{ABC})^\top = \mathbf{C}^\top\mathbf{B}^\top\mathbf{A}^\top$, and so on

Special Matrices

- Square matrix: Matrix with the same number of rows and columns

$$A_{3\times 3} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

- Diagonal matrix: Square matrix with all the elements outside the main diagonal are zero

$$D_{3\times 3} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix}$$

- Identity matrix: Square matrix with ones on the diagonal and zeros elsewhere

$$I_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Special Matrices

- Symmetric matrix: If $\mathbf{A}$ is a symmetric matrix, then

$$\mathbf{A} = \mathbf{A}^\top$$

  - We also know that

$$A_{ij} = A_{ji} \text{ for all } (i, j)$$

- Zero matrix: Matrix of all zeros
  - Ex. The 3×4 zero matrix is given by

$$\mathbf{0}_{3\times4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Vector Inner Product

- We want to summarize the similarity between two vectors
- Draw some examples of vectors pointing in different directions
  - If similar (pointing in similar direction) → positive
  - If pointing in opposite directions → negative
  - If perpendicular → zero

Inner Product: The inner product (dot product) of $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \cdots u_n v_n.$$

- $\mathbf{u}$ and $\mathbf{v}$ must be of the same size
- The inner product is a scalar quantity, $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- If $\mathbf{u} \cdot \mathbf{v} = 0$, the two vectors are orthogonal (or perpendicular).
- Example: $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, find $\mathbf{u} \cdot \mathbf{v}$

Intuition behind Inner Product

Consider the case of two-dimensions. Denote

$$\mathbf{a} = \begin{bmatrix} a_x \\ a_y \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} b_x \\ b_y \end{bmatrix}$$

- Find $\mathbf{a} \cdot \mathbf{b}$
- Draw $\mathbf{a}$ and $\mathbf{b}$ as vector addition on the directions of x and y axes
- Since x and y axis are perpendicular to each other, the inner product between the directions of x and y axes are zero, thus we have

$$\begin{aligned}\mathbf{a} \cdot \mathbf{b} &= a_x b_x \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}^\top \begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{=1\cdot1+0\cdot0=1} + a_x b_y \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}^\top \begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{=1\cdot0+0\cdot1=0} + a_y b_x \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}^\top \begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{=0\cdot1+1\cdot0=0} + a_y b_y \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}^\top \begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{=0\cdot0+1\cdot1=1} \\ &= a_x b_x + a_y b_y\end{aligned}$$

Exercise

Given a data vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- How can you express the sum of the data?
- How can you express the average of the data?
- How can you express the sum of squared of the data?

Vector Norm

The norm of a vector is a measure of its length. There are many different ways to calculate the norm, but the most common is the Euclidean norm (which corresponds to our usual conception of distance):

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{\mathbf{v}^\top \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

This is merely measuring the distance between the point $\mathbf{v}$ and the origin.

To compute the distance between two different points, say $\mathbf{x}$ and $\mathbf{y}$, we'd calculate

$$\begin{aligned}\|\mathbf{x} - \mathbf{y}\| &= \sqrt{(\mathbf{x} - \mathbf{y})^\top (\mathbf{x} - \mathbf{y})} \\ &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}\end{aligned}$$

Matrix Multiplication

- Matrix multiplication is collecting many inner products
- Let $\mathbf{A}$ be a $m \times n$ matrix and $\mathbf{B}$ be a $k \times p$ matrix. The matrix product $\mathbf{AB}$ is possible if and only if $n = k$
- If this condition holds, then the the product, $\mathbf{AB}$, is a $m \times p$ matrix and the (i, j) entry of the product $\mathbf{AB}$ is the inner product of the ith row of $\mathbf{A}$ and the jth column of $\mathbf{B}$:

$$(\mathbf{AB})_{ij} = \mathbf{A}_{i\star} \cdot \mathbf{B}_{\star j}$$

$$\mathbf{AB} = \begin{bmatrix} \longleftarrow & \mathbf{a}_1 & \longrightarrow \\ \longleftarrow & \mathbf{a}_2 & \longrightarrow \\ \vdots & \vdots & \vdots \\ \longleftarrow & \mathbf{a}_m & \longrightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_1 \cdot \mathbf{b}_p \\ \mathbf{a}_2 \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_2 \cdot \mathbf{b}_p \\ \vdots & \ddots & \vdots \\ \mathbf{a}_m \cdot \mathbf{b}_1 & \cdots & \mathbf{a}_m \cdot \mathbf{b}_p \end{bmatrix}$$

where $\mathbf{a}_i$ is the ith row of $\mathbf{A}$ and $\mathbf{b}_j$ is the jth column of $\mathbf{B}$

Exercise

$$\mathbf{AB} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 & 4 \\ -1 & 2 & 5 \end{bmatrix}$$

Exercise

$$\mathbf{AB} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 & 4 \\ -1 & 2 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} (1 \times 3) + (-2 \times -1) & (1 \times 1) + (-2 \times 2) & (1 \times 4) + (-2 \times 5) \\ (0 \times 3) + (5 \times -1) & (0 \times 1) + (5 \times 2) & (0 \times 4) + (5 \times 5) \\ (4 \times 3) + (3 \times -1) & (4 \times 1) + (3 \times 2) & (4 \times 4) + (3 \times 5) \end{bmatrix}$$

Exercise

$$\mathbf{AB} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 & 4 \\ -1 & 2 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} (1 \times 3) + (-2 \times -1) & (1 \times 1) + (-2 \times 2) & (1 \times 4) + (-2 \times 5) \\ (0 \times 3) + (5 \times -1) & (0 \times 1) + (5 \times 2) & (0 \times 4) + (5 \times 5) \\ (4 \times 3) + (3 \times -1) & (4 \times 1) + (3 \times 2) & (4 \times 4) + (3 \times 5) \end{bmatrix}$$

$$= \begin{bmatrix} 3 + 2 & 1 - 4 & 4 - 10 \\ 0 - 5 & 0 + 10 & 0 + 25 \\ 12 - 3 & 4 + 6 & 16 + 15 \end{bmatrix}$$

Exercise

$$\mathbf{AB} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 & 4 \\ -1 & 2 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} (1 \times 3) + (-2 \times -1) & (1 \times 1) + (-2 \times 2) & (1 \times 4) + (-2 \times 5) \\ (0 \times 3) + (5 \times -1) & (0 \times 1) + (5 \times 2) & (0 \times 4) + (5 \times 5) \\ (4 \times 3) + (3 \times -1) & (4 \times 1) + (3 \times 2) & (4 \times 4) + (3 \times 5) \end{bmatrix}$$

$$= \begin{bmatrix} 3 + 2 & 1 - 4 & 4 - 10 \\ 0 - 5 & 0 + 10 & 0 + 25 \\ 12 - 3 & 4 + 6 & 16 + 15 \end{bmatrix}$$

$$= \begin{bmatrix} 5 & -3 & -6 \\ -5 & 10 & 25 \\ 9 & 10 & 31 \end{bmatrix}$$

Exercise

$$\mathbf{A} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 3 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 3 & 1 & 4 \\ -1 & 2 & 5 \end{bmatrix}$$

- Can we calculate $\mathbf{AB}$?
- What is the size of $\mathbf{AB}$?
- What is $\mathbf{AB}$?
- Can we calculate $\mathbf{BA}$?
- What is the size of $\mathbf{BA}$?
- What is $\mathbf{BA}$?
- Does $\mathbf{AB} = \mathbf{BA}$?

Properties

1. $(A + B) + C = A + (B + C)$
2. $(AB)C = A(BC)$
3. $A + B = B + A$
4. $A(B + C) = AB + AC$
5. $(A + B)C = AC + BC$

Matrix-Vector Product

- What if we want to multiply matrix and vector?
- Matrix-vector product works exactly the same way as matrix multiplication
- For example, if we have an $m \times n$ matrix $\mathbf{A}$, we can
- Multiply by a $1 \times m$ row vector $\mathbf{v}^\top$ on the left

$$\mathbf{v}^\top\mathbf{A} \quad \text{works because} \quad \underset{(1\times m)}{\mathbf{v}^\top}\underset{(m\times n)}{\mathbf{A}} \Rightarrow \text{Getting a } 1 \times n \text{ row vector}$$

- Multiply by an $n \times 1$ column vector $\mathbf{x}$ on the right

$$\mathbf{A}\mathbf{x} \quad \text{works because} \quad \underset{(m\times n)}{\mathbf{A}}\underset{(n\times 1)}{\mathbf{x}} \Rightarrow \text{Getting an } m \times 1 \text{ column vector}$$

Exercise

$$\mathbf{A} = \begin{bmatrix} 2 & 3 \\ -1 & 4 \\ 5 & 1 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad \mathbf{q} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$$

- Find $\mathbf{A}\mathbf{q}$
- Find $\mathbf{A}\mathbf{v}$
- Find $\mathbf{q}^\top\mathbf{A}$

Matrix-Vector Product is Linear Combination

Let $\mathbf{A}$ be an $m \times n$ matrix partitioned into columns,

$$\mathbf{A} = \begin{bmatrix} \uparrow & \uparrow & \dots & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \\ \downarrow & \downarrow & \dots & \downarrow \end{bmatrix}$$

Let $\mathbf{x}$ be a n-dimensional vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Then, $\mathbf{Ax}$ is the linear combination of the columns of $\mathbf{A}$ using coefficients in $\mathbf{x}$:

$$\mathbf{Ax} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

Matrix Multiplication as Linear Combinations

- Matrix multiplication: applying matrix-vector product by columns
- Consider

$$\mathbf{A}_{m\times n}\mathbf{B}_{n\times p} = \mathbf{C}_{m\times p}$$

- The **first column** of C is the linear combination of the columns of A using coefficients in the **first column** of B:

$$\mathbf{A} = \begin{bmatrix} \uparrow & \uparrow & \dots & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \\ \downarrow & \downarrow & \dots & \downarrow \end{bmatrix} \quad \mathbf{b}_1 = \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{bmatrix}$$

$$\mathbf{C}_1 = b_{11}\mathbf{a}_1 + b_{21}\mathbf{a}_2 + \cdots + b_{n1}\mathbf{a}_n = \mathbf{A}\mathbf{b}_1$$

- That is, C_1 is the matrix-vector product of A and $\mathbf{b}_1$
- Doing this for each columns of C, we get

$$\mathbf{C} = \begin{bmatrix} \uparrow & \uparrow & \dots & \uparrow \\ \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_p \\ \downarrow & \downarrow & \dots & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & \dots & \uparrow \\ \mathbf{A}\mathbf{b}_1 & \mathbf{A}\mathbf{b}_2 & \dots & \mathbf{A}\mathbf{b}_p \\ \downarrow & \downarrow & \dots & \downarrow \end{bmatrix}$$

Earlier Example

$$\mathbf{AB} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 & 4 \\ -1 & 2 & 5 \end{bmatrix}$$

Earlier Example

$$\mathbf{AB} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 & 4 \\ -1 & 2 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} (1 \times 3) + (-2 \times -1) & (1 \times 1) + (-2 \times 2) & (1 \times 4) + (-2 \times 5) \\ (0 \times 3) + (5 \times -1) & (0 \times 1) + (5 \times 2) & (0 \times 4) + (5 \times 5) \\ (4 \times 3) + (3 \times -1) & (4 \times 1) + (3 \times 2) & (4 \times 4) + (3 \times 5) \end{bmatrix}$$

Earlier Example

$$\mathbf{AB} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 & 4 \\ -1 & 2 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} (1 \times 3) + (-2 \times -1) & (1 \times 1) + (-2 \times 2) & (1 \times 4) + (-2 \times 5) \\ (0 \times 3) + (5 \times -1) & (0 \times 1) + (5 \times 2) & (0 \times 4) + (5 \times 5) \\ (4 \times 3) + (3 \times -1) & (4 \times 1) + (3 \times 2) & (4 \times 4) + (3 \times 5) \end{bmatrix}$$

$$= \begin{bmatrix} 3\begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} - 1\begin{pmatrix} -2 \\ 5 \\ 3 \end{pmatrix} & 1\begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} + 2\begin{pmatrix} -2 \\ 5 \\ 3 \end{pmatrix} & 4\begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} + 5\begin{pmatrix} -2 \\ 5 \\ 3 \end{pmatrix} \end{bmatrix}$$

Earlier Example

$$\mathbf{AB} = \begin{bmatrix} 1 & -2 \\ 0 & 5 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 1 & 4 \\ -1 & 2 & 5 \end{bmatrix}$$

$$= \begin{bmatrix} (1\times 3)+(-2\times -1) & (1\times 1)+(-2\times 2) & (1\times 4)+(-2\times 5) \\ (0\times 3)+(5\times -1) & (0\times 1)+(5\times 2) & (0\times 4)+(5\times 5) \\ (4\times 3)+(3\times -1) & (4\times 1)+(3\times 2) & (4\times 4)+(3\times 5) \end{bmatrix}$$

$$= \begin{bmatrix} 3\begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} - 1\begin{pmatrix} -2 \\ 5 \\ 3 \end{pmatrix} & 1\begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} + 2\begin{pmatrix} -2 \\ 5 \\ 3 \end{pmatrix} & 4\begin{pmatrix} 1 \\ 0 \\ 4 \end{pmatrix} + 5\begin{pmatrix} -2 \\ 5 \\ 3 \end{pmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} 5 & -3 & -6 \\ -5 & 10 & 25 \\ 9 & 10 & 31 \end{bmatrix}$$

An Important Motivating Example

- Calculate $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ and $(-1)\begin{bmatrix} 3 \\ 1 \end{bmatrix} + (2)\begin{bmatrix} 1 \\ 2 \end{bmatrix}$
- Are they the same? Why?
- Draw the geometric interpretation of the latter

An Important Motivating Example

- Calculate $$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \quad \text{and} \quad (-1)\begin{bmatrix} 3 \\ 1 \end{bmatrix} + (2)\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$
- Are they the same? Why?
- Draw the geometric interpretation of the latter
- Suppose now
$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix},$$
What is (x, y)?

An Important Motivating Example

- Calculate $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ and $(-1)\begin{bmatrix} 3 \\ 1 \end{bmatrix} + (2)\begin{bmatrix} 1 \\ 2 \end{bmatrix}$
- Are they the same? Why?
- Draw the geometric interpretation of the latter
- Suppose now

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix},$$

  What is (x, y)?
- Interpret the following two expressions geometrically:

$$\begin{bmatrix} 3x + 1y \\ 1x + 2y \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} \quad \text{and} \quad x\begin{bmatrix} 3 \\ 1 \end{bmatrix} + y\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

An Important Motivating Example

- Calculate $\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ and $(-1)\begin{bmatrix} 3 \\ 1 \end{bmatrix} + (2)\begin{bmatrix} 1 \\ 2 \end{bmatrix}$
- Are they the same? Why?
- Draw the geometric interpretation of the latter
- Suppose now

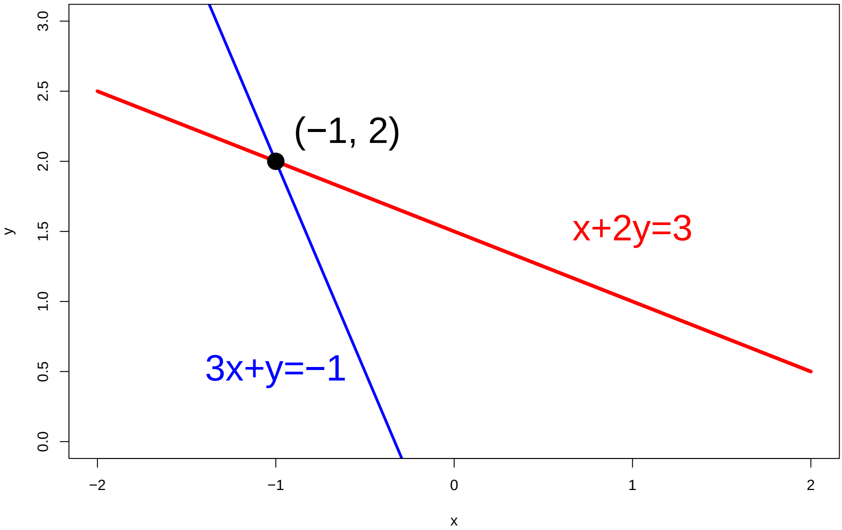
$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix},$$

  What is (x, y)?
- Interpret the following two expressions geometrically:

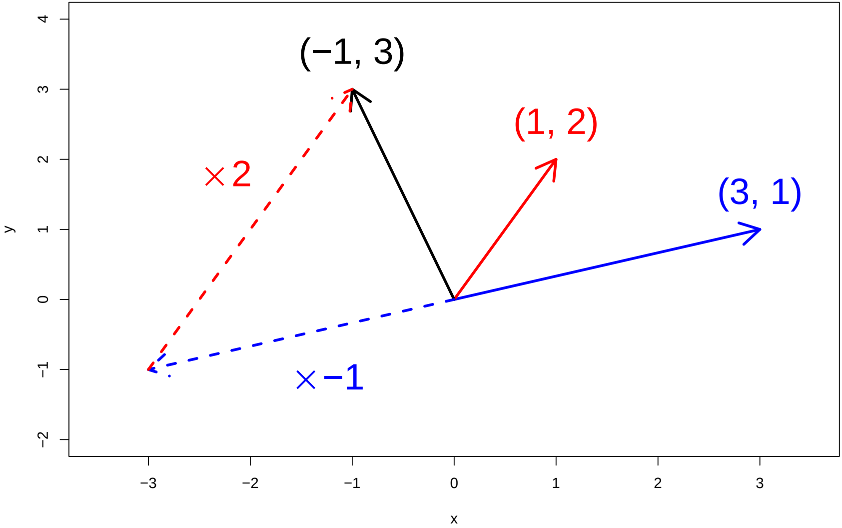
$$\begin{bmatrix} 3x + 1y \\ 1x + 2y \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} \quad \text{and} \quad x\begin{bmatrix} 3 \\ 1 \end{bmatrix} + y\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

- Row: Solving system of equations ↔ Col: Finding linear combinations

Row perspective: Solving system of linear equations



Column perspective: Finding linear combinations



Linear Regression

Suppose we have data for n observations. For each observation i, we observe covariates $x_{i1}, x_{i2}, \ldots, x_{ip}$ and dependent variable y_i. Then

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_p x_{1p} \\
y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_p x_{2p} \\
\vdots &\ \vdots\ \vdots \\
y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} \\
\vdots &\ \vdots\ \vdots \\
y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_p x_{np}
\end{aligned}$$

- Example:
  - i: Countries y_i: Democracy x_{i1}: GDP x_{i2}: Gini
- We want to know β_0, β_1, β_2 such that the model fits the data well
  - Solve system of equations (computer to the rescue!)
- We're expressing `Democracy` as a linear combinations of `GDP` and `Gini`

Linear Regression Notations

- Recall that matrix-vector product is linear combination!!
- So we can write the above expression in matrix-vector product

$$\underbrace{\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}}_{\boldsymbol{\beta}}$$

- We can also write the i-th observation in vector inner product

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \\ &= \mathbf{x}_i^\top \boldsymbol{\beta} \end{aligned}$$

where $\mathbf{x}_i = (1, x_{i1}, x_{i2}, \dots, x_{ip})$ is the i-th row of $\mathbf{X}$

System of Linear Equations

- Back to the important example. How would you solve

$$\begin{cases} 3x + y &= -1 \\ x + 2y &= 3 \end{cases}$$

System of Linear Equations

- Back to the important example. How would you solve

$$\begin{cases} 3x + y &= -1 \\ x + 2y &= 3 \end{cases}$$

- You're multiplying each equation by some constants and adding to another
  - This is the so-called Gauss-Jordan elimination
  - You already know how to solve them; complicated ones are solved by computer
- For 2 equations and 2 unknowns
  - Each equation is a line
  - The intersection of the lines is the solution
  - Draw some examples with different number of solutions

System of Linear Equations

- We can write system of linear equations as matrix-vector product

$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix},$$

- We can easily read-off the solution if the matrix is identity matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix},$$

$$\Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

System of Linear Equations

- What about when things are a little more complicated?

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

- We can still write this as matrix-vector product

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

- Our goal is to find some matrix $\mathbf{Q}$ such that

$$\mathbf{Q}\mathbf{A} = \mathbf{I}$$

- Becasue in such case, we calculate $\mathbf{x}$ easily:

$$\mathbf{Q}\mathbf{A}\mathbf{x} = \mathbf{Q}\mathbf{b} \qquad \mathbf{I}\mathbf{x} = \mathbf{Q}\mathbf{b} \qquad \mathbf{x} = \mathbf{Q}\mathbf{b}$$

- When such matrix $\mathbf{Q}$ exists, we call it the **inverse** of $\mathbf{A}$, denoted $\mathbf{A}^{-1}$

Matrix Inverse

An $n \times n$ matrix $\mathbf{A}$ is **invertible** if there exists an $n \times n$ matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$$

- $\mathbf{A}^{-1}$ is called the **inverse** of $\mathbf{A}$
- $\mathbf{A}^{-1}$ is only defined for square matrices
- If there is no such $\mathbf{A}^{-1}$, then $\mathbf{A}$ is called **singular**
- **NOT** all matrices are invertible
  - Idea: For scalar a, $a^{-1}a = 1$ except for $a = 0$
  - For example, $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ is singular and not invertible
  - We will discuss this in more detail later

Properties

- If the inverse exists, it is unique
- $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A}$
- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A}^{\top})^{-1} = (\mathbf{A}^{-1})^{\top}$

Finding Matrix Inverse

```
X = matrix(NA, nrow=3, ncol=3)
X[1,] = c(2,3,4); X[2,] = c(3,1,3); X[3,] = c(2,4,2)
(X.inv = solve(X))
```

```
     [,1] [,2]  [,3]
[1,] -0.5  0.5  0.25
[2,]  0.0 -0.2  0.30
[3,]  0.5 -0.1 -0.35
```

```
round(X.inv%*%X)
```

```
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

Matrix Inverse and Linear Regression

- We can solve the linear regression problem using matrix inverse
- Notice that $\mathbf{X}$ is often not a square matrix, hence not invertible
- However, $\mathbf{X}^\top\mathbf{X}$ is always invertible as long as number of observations (rows) in $\mathbf{X}$ is larger than number of covariates (columns)

$$
\begin{aligned}
\mathbf{X}\boldsymbol{\beta} &= \mathbf{y} \\
\mathbf{X}^\top\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^\top\mathbf{y} \\
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}
\end{aligned}
$$

Example

- Dependent Variable: % Population as gov't employee
- Covariates: Constant (1-vector), Per capita income

$$\%\text{GovEmp} = \beta_0 + \beta_1 \text{PerCapInc}$$

$$\mathbf{y} = \begin{bmatrix} 19.2 \\ 14.5 \\ 16.4 \\ 21.8 \\ 17.3 \\ 18.2 \\ 15.5 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 24,028 \\ 1 & 30,446 \\ 1 & 29,442 \\ 1 & 23,448 \\ 1 & 28,235 \\ 1 & 26,132 \\ 1 & 28,445 \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

- Want to find (β_0, β_1)

```r
x = matrix(c(1, 1, 1, 1, 1, 1, 1,
             24028, 30446, 29442, 23448, 28235, 26132, 28445),
             ncol=2)
y = matrix(c(19.2, 14.5, 16.4, 21.8, 17.3, 18.2, 15.5), ncol=1)
print(x)
```

```
     [,1]  [,2]
[1,]    1 24028
[2,]    1 30446
[3,]    1 29442
[4,]    1 23448
[5,]    1 28235
[6,]    1 26132
[7,]    1 28445
```

```r
print(y)
```

```
     [,1]
[1,] 19.2
[2,] 14.5
[3,] 16.4
[4,] 21.8
[5,] 17.3
[6,] 18.2
[7,] 15.5
```

$$\mathbf{X}^\top\mathbf{X} = \begin{bmatrix} 7 & 190,176 \\ 190,176 & 5,210,158,442 \end{bmatrix}, \quad (\mathbf{X}^\top\mathbf{X})^{-1} = \begin{bmatrix} 17.128 & -0.001 \\ -0.001 & 0.000 \end{bmatrix}$$

$$\mathbf{X}^\top\mathbf{y} = \begin{bmatrix} 1 & 24,028 \\ 1 & 30,446 \\ 1 & 29,442 \\ 1 & 23,448 \\ 1 & 28,235 \\ 1 & 26,132 \\ 1 & 28,445 \end{bmatrix} \begin{bmatrix} 19.2 \\ 14.5 \\ 16.4 \\ 21.8 \\ 17.3 \\ 18.2 \\ 15.5 \end{bmatrix} = \begin{bmatrix} 122.9 \\ 3,301,785.2 \end{bmatrix}$$

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y} \\ &= \begin{bmatrix} 17.603 & -0.001 \\ -0.001 & 0.000 \end{bmatrix} \begin{bmatrix} 122.9 \\ 3,301,785.2 \end{bmatrix} = \begin{bmatrix} 40.7897 \\ -0.0008 \end{bmatrix} \end{aligned}$$

```
print(t(x)%*%x)
```

```
       [,1]       [,2]
[1,]      7     190176
[2,] 190176 5210158442
```

```
solve(t(x)%*%x)
```

```
              [,1]          [,2]
[1,] 17.1275170151 -6.251715e-04
[2,] -0.0006251715  2.301132e-08
```

```r
print(t(x)%*%y)
```

```
          [,1]
[1,]     122.9
[2,] 3301785.2
```

```r
solve(t(x)%*%x)%*%t(x)%*%y
```

```
              [,1]
[1,] 40.7897669064
[2,] -0.0008551466
```

```r
summary(lm(y ~ x - 1))
```

```
Call:
lm(formula = y ~ x - 1)

Residuals:
      1       2       3       4       5       6       7 
-1.0423 -0.2540  0.7875  1.0617  0.6553 -0.2431 -0.9651 

Coefficients:
     Estimate Std. Error t value Pr(>|t|)    
x1 40.7897669  3.8460306  10.606 0.000129 ***
x2 -0.0008551  0.0001410  -6.066 0.001758 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When is a Matrix Invertible?

The following statements for an $n \times n$ square matrix A are **equivalent**:

- A is **invertible**:

$$\mathrm{A}^{-1} \text{ exists and } \mathrm{A}^{-1}\mathrm{A} = \mathbf{I}_n$$

- The system $\mathrm{A}\mathbf{x} = \mathbf{b}$ has a **unique solution** for all $\mathbf{b} \neq \mathbf{0}$ (zero vector)
- If $\mathrm{A}\mathbf{x} = \mathbf{0}$ (zero vector), it implies that $\mathbf{x} = \mathbf{0}$ (zero vector)
- The column vectors in A are **linearly independent** and **spans** $\mathbb{R}^n$
- The **rank** of A is n:

$$\mathrm{rank}(\mathrm{A}) = n$$

- The **determinant** of A is not zero:

$$\det(\mathrm{A}) \neq 0$$

Some Motivating Facts

- Identity Matrices are always invertible

$$\mathbf{I}_n^{-1} = \mathbf{I}_n \qquad \mathbf{I}_n^{-1}\mathbf{I}_n = \mathbf{I}_n$$

- Consider the 4-dimensional vectors, i.e, $\mathbf{x} = (x_1, x_2, x_3, x_4)$ in $\mathbb{R}^4$
  - We can always express $\mathbf{x}$ as a linear combination of columns in $\mathbf{I}_4$

$$\mathbf{x} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

  - Can you express any $\mathbf{x}$ using just 3 of the 4 elementary vectors?
    - NO! Not enough vectors to reach all points (**span**) in $\mathbb{R}^4$
  - To reach all points in $\mathbb{R}^4$, we need at least 4 vectors
  - To reach all points in $\mathbb{R}^n$, we need at least n vectors

Another Motivating Fact

- Can we always reach all points in $\mathbb{R}^n$ using at least n vectors?
- Consider the case in $\mathbb{R}^2$, can you reach all points using these vectors

$$\left\{\begin{bmatrix}2\\1\end{bmatrix}, \quad \begin{bmatrix}2\\1\end{bmatrix}\right\}?$$

Another Motivating Fact

- Can we always reach all points in $\mathbb{R}^n$ using at least n vectors?
- Consider the case in $\mathbb{R}^2$, can you reach all points using these vectors

$$\left\{\begin{bmatrix}2\\1\end{bmatrix}, \quad \begin{bmatrix}2\\1\end{bmatrix}\right\}? \qquad \left\{\begin{bmatrix}2\\1\end{bmatrix}, \quad \begin{bmatrix}4\\2\end{bmatrix}, \quad \begin{bmatrix}-6\\-3\end{bmatrix}\right\}?$$

Another Motivating Fact

- Can we always reach all points in $\mathbb{R}^n$ using at least n vectors?
- Consider the case in $\mathbb{R}^2$, can you reach all points using these vectors

$$\left\{\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right\}? \qquad \left\{\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}4\\2\end{bmatrix}, \begin{bmatrix}-6\\-3\end{bmatrix}\right\}?$$

  - NO as well. Sometimes the vectors are *repeating the information*
- How about these vectors in $\mathbb{R}^3$? Can you reach all points in $\mathbb{R}^3$ using these vectors?

$$\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\2\end{bmatrix}\right\}?$$

Another Motivating Fact

- Can we always reach all points in $\mathbb{R}^n$ using at least n vectors?
- Consider the case in $\mathbb{R}^2$, can you reach all points using these vectors

$$\left\{\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix}\right\}? \qquad \left\{\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}4\\2\end{bmatrix}, \begin{bmatrix}-6\\-3\end{bmatrix}\right\}?$$

  - NO as well. Sometimes the vectors are *repeating the information*
- How about these vectors in $\mathbb{R}^3$? Can you reach all points in $\mathbb{R}^3$ using these vectors?

$$\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\2\end{bmatrix}\right\}?$$

  - Still NO. The summation of the first two vectors is the third vector. The third vector is redundant and not adding new information. We can already point to the first two directions

Linear Independence

- Let's write down the formal definition of "no redundant vectors"
- A set of vectors

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

  is **linearly independent** if and only if
  - $\mathbf{v}_i$ is **not** a *linear combination* of the *other vectors* $\mathbf{v}_j$, i.e.,

$$\mathbf{v}_i \neq \sum_{j \neq i} c_j \mathbf{v}_j = \underbrace{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n}_{\text{linear combination excluding } \mathbf{v}_i} \qquad \text{for all } i$$

Linear Independence

- Let's write down the formal definition of "no redundant vectors"
- A set of vectors

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$$

  is **linearly independent** if and only if

  - $\mathbf{v}_i$ is **not** a *linear combination* of the *other vectors* $\mathbf{v}_j$, i.e.,

$$\mathbf{v}_i \neq \sum_{j \neq i} c_j \mathbf{v}_j = \underbrace{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n}_{\text{linear combination excluding } \mathbf{v}_i} \qquad \text{for all } i$$

  - Subtracting $\mathbf{v}_i$, this is equivalent to

$$\underbrace{\mathbf{0}}_{\text{zero vector!}} \neq -\mathbf{v}_i + \sum_{j \neq i} c_j \mathbf{v}_j = \underbrace{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + (-1)\mathbf{v}_i + \cdots + c_n\mathbf{v}_n}_{\text{linear combination of } \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n} \quad \text{for all } i$$

Linear Independence

- Let's write down the formal definition of "no redundant vectors"
- A set of vectors

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$$

is **linearly independent** if and only if

- $\mathbf{v}_i$ is **not** a *linear combination* of the *other vectors* $\mathbf{v}_j$, i.e.,

$$\mathbf{v}_i \neq \sum_{j \neq i} c_j \mathbf{v}_j = \underbrace{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n}_{\text{linear combination excluding } \mathbf{v}_i} \qquad \text{for all } i$$

- Subtracting $\mathbf{v}_i$, this is equivalent to

$$\underbrace{\mathbf{0}}_{\text{zero vector!}} \neq -\mathbf{v}_i + \sum_{j \neq i} c_j \mathbf{v}_j = \underbrace{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + (-1)\mathbf{v}_i + \cdots + c_n\mathbf{v}_n}_{\text{linear combination of } \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n} \quad \text{for all } i$$

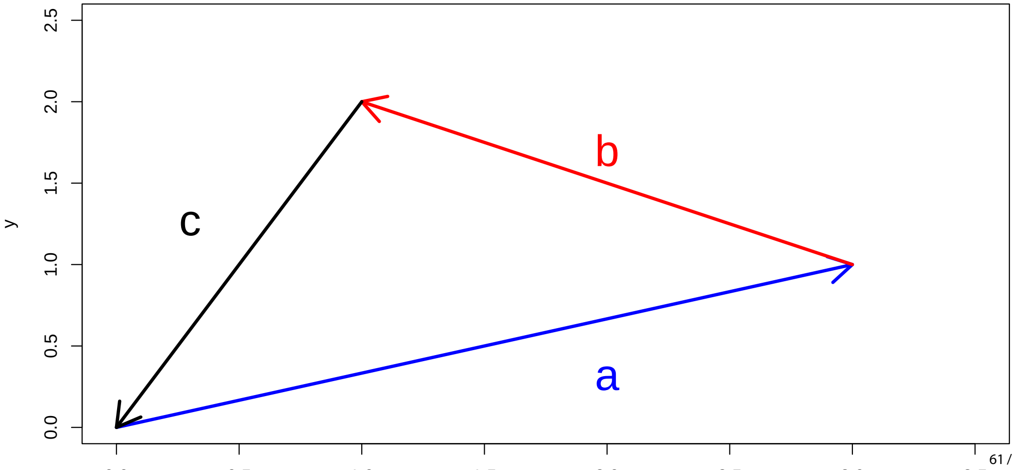
- This is equivalent to saying: The only solution to

$$\underbrace{\mathbf{0}}_{\text{zero vector!}} = \underbrace{c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + \mathbf{v}_i + \cdots + c_n\mathbf{v}_n}_{\text{linear combination of } \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_i, \dots, \mathbf{v}_n}$$

is $c_1 = c_2 = \cdots = c_n = 0$, i.e., $\mathbf{A}\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$

Another way of looking at the last definition

$$\mathbf{a} + \mathbf{b} = -\mathbf{c} \qquad 1 \cdot \mathbf{a} + 1 \cdot \mathbf{b} + 1 \cdot \mathbf{c} = \mathbf{0}$$



Linear Independence and Rank

- We say a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is **linearly dependent** if some vector $\mathbf{v}_i$ is a *linear combination* of the *other vectors* $\mathbf{v}_j$
- If all vectors **cannot** be written as linear combinations of the other vectors, then the set of vectors are **linearly independent**
- Are these vectors linearly independent?

$$\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right)?$$

Linear Independence and Rank

- We say a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is **linearly dependent** if some vector $\mathbf{v}_i$ is a *linear combination* of the *other vectors* $\mathbf{v}_j$
- If all vectors **cannot** be written as linear combinations of the other vectors, then the set of vectors are **linearly independent**
- Are these vectors linearly independent?

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right)? \qquad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\right)?$$

Linear Independence and Rank

- We say a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is **linearly dependent** if some vector $\mathbf{v}_i$ is a *linear combination* of the *other vectors* $\mathbf{v}_j$
- If all vectors **cannot** be written as linear combinations of the other vectors, then the set of vectors are **linearly independent**
- Are these vectors linearly independent?

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right)? \quad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\right)? \quad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}\right)?$$

Linear Independence and Rank

- We say a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is **linearly dependent** if some vector $\mathbf{v}_i$ is a *linear combination* of the *other vectors* $\mathbf{v}_j$
- If all vectors **cannot** be written as linear combinations of the other vectors, then the set of vectors are **linearly independent**
- Are these vectors linearly independent?

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right)? \quad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\right)? \quad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}\right)?$$

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}\right)?$$

Linear Independence and Rank

- We say a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is **linearly dependent** if some vector $\mathbf{v}_i$ is a *linear combination* of the *other vectors* $\mathbf{v}_j$
- If all vectors **cannot** be written as linear combinations of the other vectors, then the set of vectors are **linearly independent**
- Are these vectors linearly independent?

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right)? \quad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\right)? \quad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}\right)?$$

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}\right)? \quad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}\begin{bmatrix}4\\5\\6\end{bmatrix}\right)?$$

Linear Independence and Rank

- We say a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is **linearly dependent** if some vector $\mathbf{v}_i$ is a *linear combination* of the *other vectors* $\mathbf{v}_j$
- If all vectors **cannot** be written as linear combinations of the other vectors, then the set of vectors are **linearly independent**
- Are these vectors linearly independent?

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right)? \quad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\right)? \quad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}\right)?$$

$$\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}\right)? \quad \left(\begin{bmatrix}1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}\begin{bmatrix}4\\5\\6\end{bmatrix}\right)?$$

- The *more* vectors added, the *easier they become redundant*!

Linear Independence, Rank, and Span

- **Rank** of a Matrix: *Number* of linearly independent columns in a matrix
  - Rank for the examples above: 1, 2, 3, 3, 3

Linear Independence, Rank, and Span

- **Rank** of a Matrix: *Number* of linearly independent columns in a matrix
  - Rank for the examples above: 1, 2, 3, 3, 3
- We want to reach all n coordinates using columns in $\mathrm{A}_{n \times n}$

Linear Independence, Rank, and Span

- **Rank** of a Matrix: *Number* of linearly independent columns in a matrix
  - Rank for the examples above: 1, 2, 3, 3, 3
- We want to reach all n coordinates using columns in $\mathrm{A}_{n \times n}$
  - The more linearly independent vectors, the more coordinates we can reach
    - We want to **span** as much coordinate as possible

Linear Independence, Rank, and Span

- **Rank** of a Matrix: *Number* of linearly independent columns in a matrix
  - Rank for the examples above: 1, 2, 3, 3, 3
- We want to reach all n coordinates using columns in $\mathrm{A}_{n \times n}$
  - The more linearly independent vectors, the more coordinates we can reach
    - We want to **span** as much coordinate as possible
    - We want $\mathrm{rank}(\mathrm{A}) \geq n$

Linear Independence, Rank, and Span

- **Rank** of a Matrix: *Number* of linearly independent columns in a matrix
  - Rank for the examples above: 1, 2, 3, 3, 3
- We want to reach all n coordinates using columns in $A_{n \times n}$
  - The more linearly independent vectors, the more coordinates we can reach
    - We want to **span** as much coordinate as possible
    - We want $\text{rank}(A) \geq n$
  - But once we reached all n coordinates, newly added vectors are redundant
    - We also want to make sure all vectors are **linearly independent**

Linear Independence, Rank, and Span

- **Rank** of a Matrix: *Number* of linearly independent columns in a matrix
  - Rank for the examples above: 1, 2, 3, 3, 3
- We want to reach all n coordinates using columns in $A_{n \times n}$
  - The more linearly independent vectors, the more coordinates we can reach
    - We want to **span** as much coordinate as possible
    - We want $\mathrm{rank}(A) \geq n$
  - But once we reached all n coordinates, newly added vectors are redundant
    - We also want to make sure all vectors are **linearly independent**
    - We can only have $\mathrm{rank}(A) \leq n$

Linear Independence, Rank, and Span

- **Rank** of a Matrix: *Number* of linearly independent columns in a matrix
  - Rank for the examples above: 1, 2, 3, 3, 3
- We want to reach all n coordinates using columns in $A_{n \times n}$
  - The more linearly independent vectors, the more coordinates we can reach
    - We want to **span** as much coordinate as possible
    - We want $\text{rank}(A) \geq n$
  - But once we reached all n coordinates, newly added vectors are redundant
    - We also want to make sure all vectors are **linearly independent**
    - We can only have $\text{rank}(A) \leq n$
  - The only choice we left is to have $\text{rank}(A) = n$

Reiterate the equivalent conditions

The following statements for an $n \times n$ square matrix A are **equivalent**:

- A is **invertible**:

$$\mathrm{A}^{-1} \text{ exists and } \mathrm{A}^{-1}\mathrm{A} = \mathbf{I}_n$$

- The system $\mathrm{A}\mathbf{x} = \mathbf{b}$ has a **unique solution** for all $\mathbf{b} \neq \mathbf{0}$ (zero vector)
- If $\mathrm{A}\mathbf{x} = \mathbf{0}$ (zero vector), it implies that $\mathbf{x} = \mathbf{0}$ (zero vector)
- The column vectors in A are **linearly independent** and **spans** $\mathbb{R}^n$
- The **rank** of A is n:

$$\mathrm{rank}(\mathrm{A}) = n$$

- The **determinant** of A is not zero (more on this)

$$\det(\mathrm{A}) \neq 0$$

Determinant

- Usually not easy to calculate except for matrices of size 2×2

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\det(\mathbf{A}) = ad - bc$$

- In 2×2 can think about it as the **area** of the parallelogram spanned by the column vectors
  - If the area is zero, then the parallelogram is a line
  - Then we cannot reach all points and is thus not invertible
  - If the area is not zero, then we can reach all points and is thus invertible

Determinant

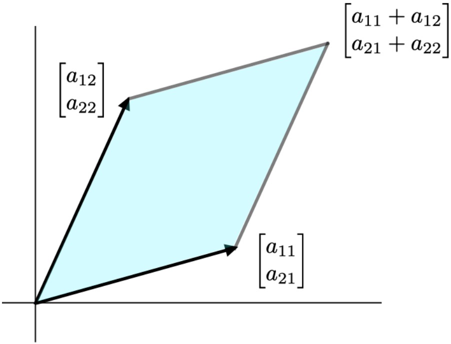
- Usually not easy to calculate except for matrices of size 2×2

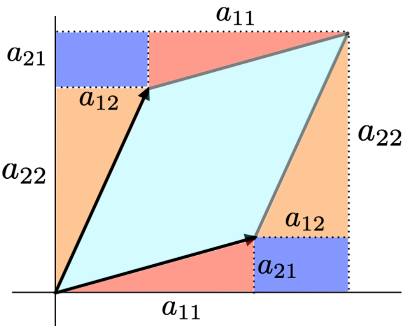
$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\det(\mathbf{A}) = ad - bc$$

- In 2×2 can think about it as the **area** of the parallelogram spanned by the column vectors
  - If the area is zero, then the parallelogram is a line
  - Then we cannot reach all points and is thus not invertible
  - If the area is not zero, then we can reach all points and is thus invertible

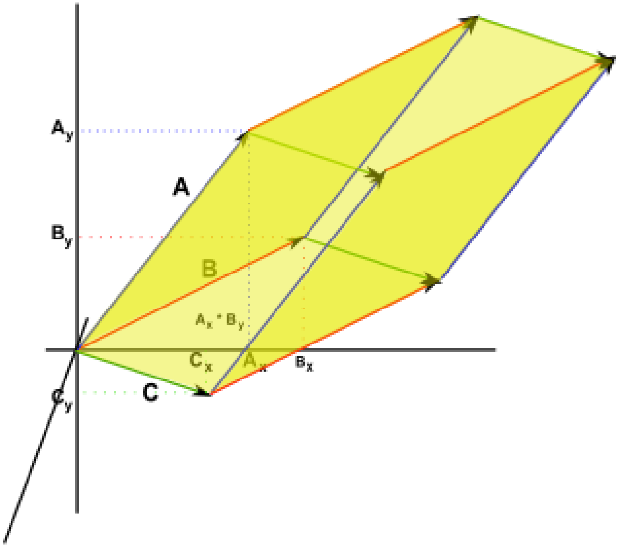
Determinant for 2×2 Matrix





$$
\begin{aligned}
\text{area of parallelogram} &= (a_{11} + a_{12})(a_{21} + a_{22}) - a_{12}a_{22} - a_{11}a_{21} - 2a_{21}a_{12} \\
&= \cancel{a_{11}a_{21}} + a_{11}a_{22} + a_{12}a_{21} + \cancel{a_{12}a_{22}} - \cancel{a_{12}a_{22}} - \cancel{a_{11}a_{21}} - 2a_{21}a_{12} \\
&= a_{11}a_{22} - a_{21}a_{12}
\end{aligned}
$$

Determinant for 3 × 3 Matrix



Properties

- $\det(\mathbf{I}) = 1$
- $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
- $\det(\mathbf{A}) = \det(\mathbf{A}^\top)$
- $\det(\mathbf{A}^{-1}) = \frac{1}{\det(\mathbf{A})}$

Inverse for 2×2 Matrix

- Usually not easy to calculate inverse except for matrices of size 2×2

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\det(\mathbf{A}) = ad - bc$$

- If $\det(\mathbf{A}) \neq 0$, then $\mathbf{A}$ is invertible, and

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Inverse for 2×2 Matrix

- Usually not easy to calculate inverse except for matrices of size 2×2

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$\det(\mathbf{A}) = ad - bc$$

- If $\det(\mathbf{A}) \neq 0$, then $\mathbf{A}$ is invertible, and

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$