

Day 4: Vectors and Matrices

Ikuma Ogura

Ph.D. student, Department of Government, Georgetown University

August 22, 2019

Today

- Today
 - ▶ Vector & matrix algebra
 - ▶ Matrix calculus
 - ▶ Geometry of matrix algebra

Why Vectors and Matrices?

- Making notations & calculations simple
- Basis of multivariate statistical techniques

Why Vectors and Matrices? (cont.)

- We have so far dealt with algebra and calculus with **scalars**
- Suppose we want to use more than one independent variable (say 10) in regression analysis:

$$y_i = b_0 + b_1x_{1i} + b_2x_{2i} + b_3x_{3i} + b_4x_{4i} + b_5x_{5i} + \cdots + b_{10}x_{10i} + e_i$$

- Let's find $b_0, b_1, b_2, \dots, b_{10}$ which minimize the sum of squared residuals

Why Vectors and Matrices? (cont.)

- As we covered yesterday, we need to solve the system of equations

$$\begin{cases} \frac{\partial}{\partial b_0} \sum_{i=1}^n \{y_i - (b_0 + b_1x_{1i} + b_2x_{2i} + \cdots + b_{10}x_{10i})\}^2 = 0 \\ \frac{\partial}{\partial b_1} \sum_{i=1}^n \{y_i - (b_0 + b_1x_{1i} + b_2x_{2i} + \cdots + b_{10}x_{10i})\}^2 = 0 \\ \vdots \\ \frac{\partial}{\partial b_{10}} \sum_{i=1}^n \{y_i - (b_0 + b_1x_{1i} + b_2x_{2i} + \cdots + b_{10}x_{10i})\}^2 = 0 \end{cases}$$

- Aww...!

Tips

- Keep track of vector/matrix dimensions.
- Become able to connect with scalar algebra/calculus

Vector

- A k -dimensional vector is a list of k numbers.
- Usually numbers are arranged in a column.
- We usually represent a vector using a bold lower case. e.g., \mathbf{a}

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}$$

Vector (cont.)

- To arrange numbers in a row, we **transpose** the vector.
- Transpose of a column vector \mathbf{a} of dimension k , denoted as \mathbf{a}' (also written as \mathbf{a}^T), is a row vector

$$\mathbf{a}' = (a_1 \ a_2 \ \dots \ a_k)$$

- Obviously, the transpose of a row vector is a column vector!

Vector (cont.)

- **Norm** (or length) of vector \mathbf{a} is defined as

$$\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_k^2}$$

- Related types of vectors
 - ▶ Normalized vector: a vector with norm 1
 - ▶ Zero vector ($\mathbf{0}$): a vector whose elements are all 0

Matrix

- A $n \times k$ -dimensional matrix is a rectangle array of numbers

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix}$$

- We usually represent a matrix using a bold upper case. e.g., \mathbf{A}
- By convention, a_{ij} refers to an element in the i th row and the j th column.
- We can think of a k -dimensional column vector as a $k \times 1$ matrix and a k -dimensional row vector as a $1 \times k$ matrix.

Matrix (cont.)

- It is often useful to think of matrices as made up of a collection of column/row vectors.
- For example, we can represent matrix \mathbf{A} as a collection of column vectors.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} = (\mathbf{a}_1 \quad \mathbf{a}_2 \quad \dots \quad \mathbf{a}_k)$$

where

$$\mathbf{a}_i = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{pmatrix}$$

Matrix (cont.)

- Similarly, we can represent matrix \mathbf{A} as a collection of row vectors.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} = \begin{pmatrix} \alpha'_1 \\ \alpha'_2 \\ \vdots \\ \alpha'_n \end{pmatrix}$$

where α_i is a vector where elements of i th row of \mathbf{A} are arranged in a column,

$$\alpha_i = (a_{i1} \quad a_{i2} \quad \dots \quad a_{ik})' = \begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ik} \end{pmatrix}$$

Matrix (cont.)

- Example: Matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 1 & -4 \end{pmatrix}$$

can be represented as a collection of column vectors

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 2 \\ -4 \end{pmatrix}$$

or as a collection of row vectors

$$\boldsymbol{\alpha}_1 = (1 \ 2)' = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \boldsymbol{\alpha}_2 = (1 \ -4)' = \begin{pmatrix} 1 \\ -4 \end{pmatrix}$$

Matrix (cont.)

- A transpose of a matrix can be obtained by flipping its rows and columns.
- Transpose of a $n \times k$ matrix \mathbf{A} is denoted as \mathbf{A}' (also written as \mathbf{A}^T)

$$\mathbf{A}' = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \dots & a_{nk} \end{pmatrix}$$

where the dimension of \mathbf{A}' is $k \times n$.

- Obviously, $(\mathbf{A}')' = \mathbf{A}$

Matrix (cont.)

- Example: The transpose of matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & -4 & -2 \end{pmatrix}$$

is

$$\mathbf{A}' = \begin{pmatrix} 1 & 1 \\ 2 & -4 \\ 3 & -2 \end{pmatrix}$$

Matrix Addition/Subtraction

- Addition/subtraction of vectors of the same dimension is defined as

$$\mathbf{a} \pm \mathbf{b} = \begin{pmatrix} a_1 \pm b_1 \\ a_2 \pm b_2 \\ \vdots \\ a_n \pm b_n \end{pmatrix}$$

- Similarly, if matrices \mathbf{A} and \mathbf{B} are of the same dimensions, we can define addition/subtraction as

$$\mathbf{A} \pm \mathbf{B} = \begin{pmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \dots & a_{1k} \pm b_{1k} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \dots & a_{2k} \pm b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} \pm b_{n1} & a_{n2} \pm b_{n2} & \dots & a_{nk} \pm b_{nk} \end{pmatrix}$$

Matrix Multiplication

- We can multiply a vector \mathbf{a} by a scalar c as

$$c\mathbf{a} = \begin{pmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_k \end{pmatrix}$$

- Similarly, we can define the scalar multiplication of a matrix \mathbf{A} by a scalar c as

$$c\mathbf{A} = \begin{pmatrix} ca_{11} & ca_{12} & \dots & ca_{1k} \\ ca_{21} & ca_{22} & \dots & ca_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \dots & ca_{nk} \end{pmatrix}$$

Matrix Multiplication (cont.)

- **Dot product/inner product** for vectors of the same dimensions, \mathbf{a} and \mathbf{b} , is defined as

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} = a_1b_1 + a_2b_2 + \cdots + a_kb_k = \sum_j^k a_jb_j$$

Therefore, an inner product of two vectors is a scalar.

- Vector norm can be also expressed as $\|\mathbf{a}\| = \sqrt{\mathbf{a} \cdot \mathbf{a}}$.

Matrix Multiplication (cont.)

- If \mathbf{A} is a $n \times k$ matrix and \mathbf{B} is a $k \times m$ matrix, then we can define their product $\mathbf{C} = \mathbf{AB}$ where

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj}$$

- ▶ ij element of the resultant matrix \mathbf{C} is the inner product of i th row of \mathbf{A} and j th column of \mathbf{B} .
- ▶ Therefore, \mathbf{C} is a $n \times m$ matrix.

Matrix Multiplication (cont.)

- Example: Let

$$\mathbf{A} = \begin{pmatrix} 3 & 8 \\ -5 & 2 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} -4 & 6 \\ 2 & -7 \end{pmatrix}.$$

Then their product \mathbf{AB} can be calculated as

$$\mathbf{AB} = \begin{pmatrix} 3 & 8 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} -4 & 6 \\ 2 & -7 \end{pmatrix} = \begin{pmatrix} 3 \cdot (-4) + 8 \cdot 2 & \\ & \end{pmatrix}$$

Matrix Multiplication (cont.)

- Example: Let

$$\mathbf{A} = \begin{pmatrix} 3 & 8 \\ -5 & 2 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} -4 & 6 \\ 2 & -7 \end{pmatrix}.$$

Then their product \mathbf{AB} can be calculated as

$$\mathbf{AB} = \begin{pmatrix} 3 & 8 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} -4 & 6 \\ 2 & -7 \end{pmatrix} = \begin{pmatrix} 2 & 3 \cdot 6 + 8 \cdot (-7) \end{pmatrix}$$

Matrix Multiplication (cont.)

- Example: Let

$$\mathbf{A} = \begin{pmatrix} 3 & 8 \\ -5 & 2 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} -4 & 6 \\ 2 & -7 \end{pmatrix}.$$

Then their product \mathbf{AB} can be calculated as

$$\mathbf{AB} = \begin{pmatrix} 3 & 8 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} -4 & 6 \\ 2 & -7 \end{pmatrix} = \begin{pmatrix} 2 & -38 \\ (-5) \cdot (-4) + 2 \cdot 2 & \end{pmatrix}$$

Matrix Multiplication (cont.)

- Example: Let

$$\mathbf{A} = \begin{pmatrix} 3 & 8 \\ -5 & 2 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} -4 & 6 \\ 2 & -7 \end{pmatrix}.$$

Then their product \mathbf{AB} can be calculated as

$$\mathbf{AB} = \begin{pmatrix} 3 & 8 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} -4 & 6 \\ 2 & -7 \end{pmatrix} = \begin{pmatrix} 2 & -38 \\ 24 & (-5) \cdot 6 + 2 \cdot (-7) \end{pmatrix}$$

Matrix Multiplication (cont.)

- Example: Let

$$\mathbf{A} = \begin{pmatrix} 3 & 8 \\ -5 & 2 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} -4 & 6 \\ 2 & -7 \end{pmatrix}.$$

Then their product \mathbf{AB} can be calculated as

$$\mathbf{AB} = \begin{pmatrix} 3 & 8 \\ -5 & 2 \end{pmatrix} \begin{pmatrix} -4 & 6 \\ 2 & -7 \end{pmatrix} = \begin{pmatrix} 2 & -38 \\ 24 & -44 \end{pmatrix}$$

Matrix Multiplication (cont.)

- \mathbf{AB} is generally not equal to \mathbf{BA} , even if both are defined.
- Common properties of transpose matrices
 1. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$
 2. $(\mathbf{A}'\mathbf{A})' = \mathbf{A}'(\mathbf{A}')' = \mathbf{A}'\mathbf{A}$
- Thinking of vectors \mathbf{a} and \mathbf{b} as $k \times 1$ matrices, their inner product can also be written as their product

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}'\mathbf{b} = \begin{pmatrix} a_1 & a_2 & \dots & a_k \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix}$$

Matrix Calculation: Exercises

- Let

$$\mathbf{A} = \begin{pmatrix} 2 & 5 \\ -3 & 1 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 4 & -2 & 7 \\ -3 & 1 & 0 \end{pmatrix},$$

$$\mathbf{C} = \begin{pmatrix} 5 & -1 \\ 0 & 4 \\ -2 & 1 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} 2 & 3 \\ -1 & 2 \end{pmatrix}.$$

Calculate the following.

1. $\mathbf{B}' + \mathbf{C}$
2. \mathbf{AD}
3. \mathbf{DA}
4. \mathbf{BC}
5. $\mathbf{B}'\mathbf{C}'$

Matrix Types

- **Square Matrix:** Number of rows and columns are the same.
- **Symmetric Matrix:** A square matrix where $\mathbf{A} = \mathbf{A}'$. Therefore, $a_{ij} = a_{ji}$.
- **Triangular Matrix:** A square matrix in which all the elements above or below the main diagonal are equal to 0.
 - ▶ Main diagonal: a_{ij} s where $i = j$
 - ▶ A square matrix where elements above the main diagonal are 0 are called **lower triangular**, and one where elements below the main diagonal equal 0 are called **upper triangular**.
- **Diagonal Matrix:** A square matrix where off-diagonal elements are all 0.

Matrix Types (cont.)

- **Identity Matrix:** A diagonal matrix where all diagonal elements are 1.

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

- The identity matrix works like scalar 1. That is, for any matrices \mathbf{A} that are conformable with \mathbf{I} ,

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A}$$

Matrix Types: Exercises

- Are the following matrices square, symmetric, triangular, and diagonal?

1. $\begin{pmatrix} 1 & 1 \\ 3 & -2 \end{pmatrix}$

2. $\begin{pmatrix} 3 & 6 & 0 \\ 6 & 5 & -7 \\ 0 & -7 & 0 \end{pmatrix}$

3. $\begin{pmatrix} 1 & 4 & 1 \\ 0 & 6 & -5 \\ 0 & 0 & 1 \end{pmatrix}$

Trace

- The trace of a square matrix is the sum of its diagonal elements

$$\text{tr}(\mathbf{A}) = \sum_i^k a_{ii}$$

- Properties of trace
 1. $\text{tr}(c\mathbf{A}) = c\text{tr}(\mathbf{A})$
 2. $\text{tr}(\mathbf{A}') = \text{tr}(\mathbf{A})$
 3. $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$
 4. $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$

Determinant

- The determinant is another widely used operation to transform a square matrix to a scalar.
- The determinant of matrix \mathbf{A} is represented as $\det \mathbf{A}$ or $|\mathbf{A}|$
- Properties of determinant: let \mathbf{A} be a $n \times n$ square matrix,
 1. $\det \mathbf{A} = \det \mathbf{A}'$
 2. if \mathbf{A} is either diagonal or triangular, $\det \mathbf{A} = \prod_{i=1}^n a_{ii}$
 3. $\det(c\mathbf{A}) = c^n \det \mathbf{A}$
 4. $\det \mathbf{AB} = \det \mathbf{BA} = \det \mathbf{A} \det \mathbf{B}$

Inverse

- If a square matrix \mathbf{A} is nonsingular (or invertible), a square matrix \mathbf{A}^{-1} exists which satisfies

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

- We call \mathbf{A}^{-1} as the inverse of \mathbf{A} .
- When we cannot define \mathbf{A}^{-1} , \mathbf{A} is called singular.
- Some properties of the inverse
 1. if \mathbf{A} is nonsingular, \mathbf{A}^{-1} is unique
 2. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
 3. $(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}$
 4. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
 5. $\det \mathbf{A} \neq 0 \Leftrightarrow \mathbf{A}$ is nonsingular

System of Linear Equations

- Matrices help us express and solve system of linear equations.
- For example, the system of linear equations

$$\begin{cases} x - y = 4 \\ 2x + y = 2 \end{cases}$$

can be expressed as

$$\mathbf{Ax} = \mathbf{b}$$

where $\mathbf{A} = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, and $\mathbf{b} = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$.

System of Linear Equations (cont.)

- More generally, a system of n linear equations with n unknowns

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n \end{cases}$$

can be represented as

$$\mathbf{Ax} = \mathbf{b}$$

where $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$

System of Linear Equations (cont.)

- Example: the system of regression equations

$$\begin{cases} y_1 = b_0 + b_1x_{11} + \cdots + b_kx_{1k} + e_1 \\ y_2 = b_0 + b_1x_{21} + \cdots + b_kx_{2k} + e_2 \\ \vdots \\ y_n = b_0 + b_1x_{n1} + \cdots + b_kx_{nk} + e_n \end{cases}$$

can be written as

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

where $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, $\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix}$, and $\mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$.

- ▶ Why is the elements in the first column of \mathbf{X} are all 1?

System of Linear Equations (cont.)

- Let's solve a system of linear equations written as

$$Ax = b$$

for x .

- Assuming that A is invertible, multiply A^{-1} from the left:

$$\begin{aligned} A^{-1}Ax &= A^{-1}b \\ \Rightarrow x &= A^{-1}b \end{aligned}$$

System of Linear Equations (cont.)

- Example: Solving the system of linear equations

$$\begin{cases} x - y = 4 \\ 2x + y = 2 \end{cases}$$

using matrix inversion,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}$$

Matrix Calculus: Preparation

- We can also represent function $y = f(x_1, x_2, \dots, x_n)$ using vector notation:

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

where \mathbf{x} is the vector of inputs

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

Matrix Calculus: Preparation (cont.)

- Example: linear function of x_1, x_2, \dots, x_n can be written using vectors

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = \begin{pmatrix} a_1 & a_2 & \dots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{a}'\mathbf{x}$$

where \mathbf{a} is the vector of coefficients and \mathbf{x} is the vector of variables.

Matrix Calculus: Preparation (cont.)

- Another example: **quadratic form** is a polynomial in which each term is the monomial of degree 2, and can be written using vectors and a matrix

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=j} a_{ii}x_i^2 + \sum_{i \neq j} (a_{ij} + a_{ji})x_i x_j$$

where $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$ is the matrix of coefficients

and $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ is the vector of variables.

Matrix Calculus: Preparation (cont.)

- Quadratic form of two variables $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and coefficient matrix $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is

$$\begin{aligned} & \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} (a_{11}x_1 + a_{12}x_2) & (a_{21}x_1 + a_{22}x_2) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= (a_{11}x_1 + a_{12}x_2)x_1 + (a_{21}x_1 + a_{22}x_2)x_2 \\ &= a_{11}x_1^2 + a_{22}x_2^2 + (a_{12} + a_{21})x_1x_2 \end{aligned}$$

Gradient

- **Gradient** of a function $f(\mathbf{x})$ is a column vector of dimension n whose i th element is the partial derivative of $f(\mathbf{x})$ with respect to x_i .

$$\nabla f(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}} f(\mathbf{x}) = \begin{pmatrix} \frac{\partial}{\partial x_1} f(\mathbf{x}) \\ \frac{\partial}{\partial x_2} f(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} f(\mathbf{x}) \end{pmatrix}$$

- ▶ The operator ∇ is called **nabla**.

Gradient (cont.)

- Example: let $f(x, y) = x^2 - xy + y^2$. Then, its gradient vector is

$$\nabla f(x, y) = \begin{pmatrix} 2x - y \\ -x + 2y \end{pmatrix}$$

Gradient (cont.)

- Another example: we can write the normal equation for deriving the OLS coefficients using gradient. Let $S(\mathbf{b})$ be the function to compute the sum of squared residuals, where \mathbf{b} is the vector of coefficients. Then,

$$\nabla S(\mathbf{b}) = \frac{\partial}{\partial \mathbf{b}} S(\mathbf{b}) = \begin{pmatrix} \frac{\partial}{\partial b_0} S(\mathbf{b}) \\ \frac{\partial}{\partial b_1} S(\mathbf{b}) \\ \vdots \\ \frac{\partial}{\partial b_k} S(\mathbf{b}) \end{pmatrix} = \mathbf{0}$$

Hessian

- **Hessian** of function $f(\mathbf{x})$ is a matrix whose ij entry is the second-order derivative of $f(\mathbf{x})$ with regard to x_i and x_j .

$$\mathbf{H} = \frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}'} f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} f(\mathbf{x}) & \frac{\partial^2}{\partial x_1 \partial x_2} f(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_1 \partial x_k} f(\mathbf{x}) \\ \frac{\partial^2}{\partial x_2 \partial x_1} f(\mathbf{x}) & \frac{\partial^2}{\partial x_2^2} f(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_2 \partial x_k} f(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial x_k \partial x_1} f(\mathbf{x}) & \frac{\partial^2}{\partial x_k \partial x_2} f(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_k^2} f(\mathbf{x}) \end{pmatrix}$$

- Based on the discussion yesterday... \rightarrow Hessian is a symmetric matrix

Rules for Matrix Calculus

- Rules for linear functions and quadratic forms: let \mathbf{a} and \mathbf{A} be vector/matrix of coefficients and \mathbf{x} the vector of variables, then
 - $\frac{\partial}{\partial \mathbf{x}}(\mathbf{a}'\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{a}) = \mathbf{a}$
 - $\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}\mathbf{x}) = (\mathbf{A} + \mathbf{A}')\mathbf{x}$
 - $\frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}'}(\mathbf{x}'\mathbf{A}\mathbf{x}) = \mathbf{A} + \mathbf{A}'$

Multivariate Optimization

- First order condition

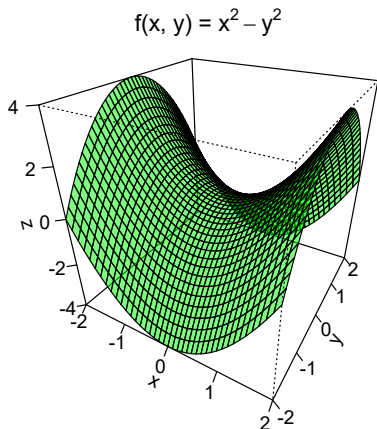
$$\nabla f(\mathbf{x}) = \mathbf{0}$$

- Second order condition... \rightarrow we use the Hessian matrix to determine local min/max!

Multivariate Optimization (cont.)

- Specifically, we examine the sign of the quadratic form of the Hessian matrix evaluated at x^* where $\nabla f(x^*) = 0$.
 - ▶ $x'H^*x$ is the function of x
- Second order condition:
 - ▶ $x'H^*x > 0$ for any values of $x \rightarrow x^*$ is **local min**
 - ★ In this case, we say H^* is **positive definite**
 - ▶ $x'H^*x < 0$ for any values of $x \rightarrow x^*$ is **local max**
 - ★ In this case, we say H^* is **negative definite**
 - ▶ sign of $x'H^*x$ depends on the values of $x \rightarrow x^*$ is **saddle point**
 - ★ In this case, we say H^* is **indefinite**

Multivariate Optimization (cont.)



- Saddle point: a point where satisfies the first order condition but neither local minimum nor local maximum
- Examples
 - ▶ $f(x) = x^3$
 - ▶ $f(x, y) = x^2 - y^2$

Multivariate Optimization (cont.)

- Example: let $f(x, y) = x^2 - y^2$. Then, as

$$\nabla f(x, y) = \begin{pmatrix} 2x \\ -2y \end{pmatrix},$$

the point (x^*, y^*) satisfies the first order condition. Also, the Hessian matrix is

$$\mathbf{H} = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

(continued from the previous slide)

Therefore, the quadratic form of \mathbf{H} at $(x^*, y^*) = (0, 0)$ is

$$\begin{aligned} & (x \ y) \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= (2x \ -2y) \begin{pmatrix} x \\ y \end{pmatrix} \\ &= 2(x^2 - y^2) \end{aligned} \tag{1}$$

Since (1) is positive when $|x| > |y|$ and negative $|x| < |y|$, \mathbf{H} is indefinite, suggesting that $(x^*, y^*) = (0, 0)$ is a saddle point.

Multivariate Optimization: Exercise

- Let $f(x, y) = x^2 + xy + y^2$. Demonstrate that $(x, y) = (0, 0)$ is a local minimum of this function.

Multivariate Optimization (cont.)

- Isn't it too cumbersome?
 - ▶ There's an easier way to determine the definiteness of the Hessian matrix (which uses eigen values)
- Global v. local
 - ▶ For (x^*, y^*) to be the global min/max, the Hessian matrix must be negative/positive definite at points other than (x^*, y^*) .
 - ▶ In such cases, $f(x)$ is globally concave/convex.
- *(In most (but not all) of the applications you encounter, you don't need to care about the second order condition...)*

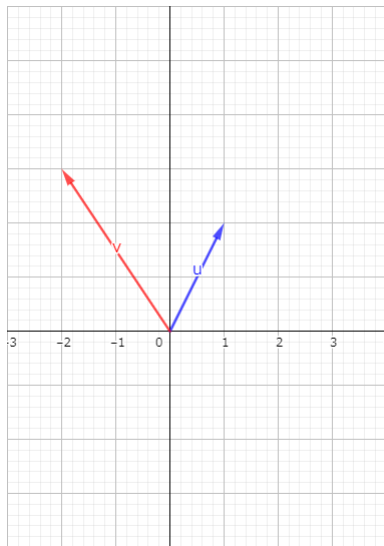
Multivariate Optimization (cont.)

- In the problem set, you're asked to find regression coefficients $\mathbf{b} = (b_0 \ b_1 \ \cdots \ b_k)'$ which best fit the data by minimizing the sum of squared residuals.
 - ▶ First check that the sum of squared residuals can be represented as the dot product of residual vector $\mathbf{e} = (e_1 \ e_2 \ \cdots \ e_n)'$ with itself.
 - ▶ Express sum of squared residuals with \mathbf{y} , \mathbf{X} , and \mathbf{b} .
 - ▶ Differentiate the sum of squared residuals with \mathbf{b} and set the gradient vector to $\mathbf{0}$.
 - ▶ Solve the system of linear equations.

Geometry of Matrix Algebra

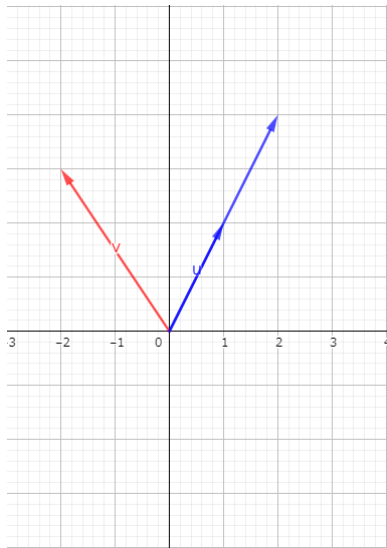
- All vector/matrix operations have geometric meanings.
- Here I use examples in two dimensional space, but the discussion naturally extends to d -dimensional space.

Geometry of Matrix Algebra: Vector



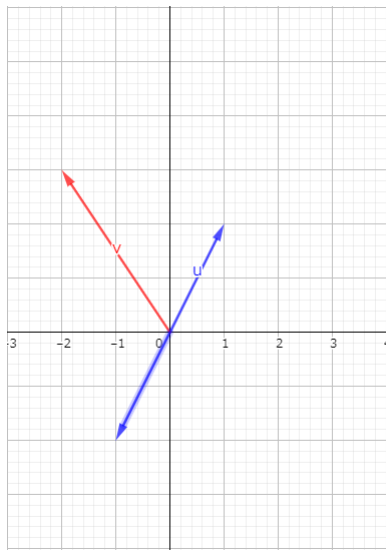
- A d -dimensional vector represents a point (more precisely, an arrow to the point) from the origin on the d -dimensional coordinate (Euclidean) space.
- $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{v} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$
- Vector norm represents the length of a vector
 - ▶ We can show this using the Pythagorean theorem

Geometry of Matrix Algebra: Vector (cont.)



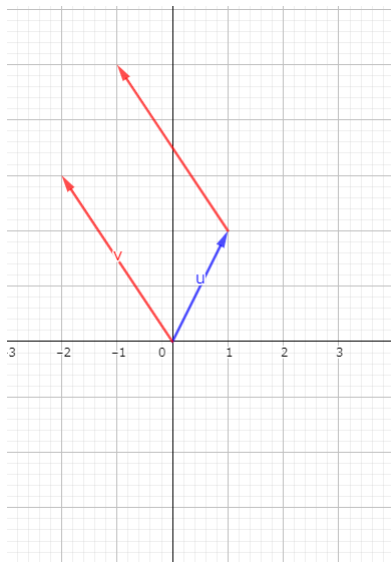
- Scalar product: stretching ($|c| > 1$) or contracting ($|c| < 1$) the original vector based on the size of the scalar.
- When $c < 0$, the original vector is reflected about the origin

Geometry of Matrix Algebra: Vector (cont.)



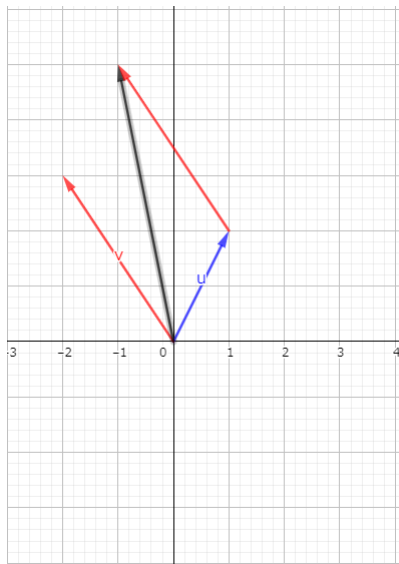
- Scalar product: stretching ($|c| > 1$) or contracting ($|c| < 1$) the original vector based on the size of the scalar.
- When $c < 0$, the original vector is reflected about the origin

Geometry of Matrix Algebra: Vector (cont.)



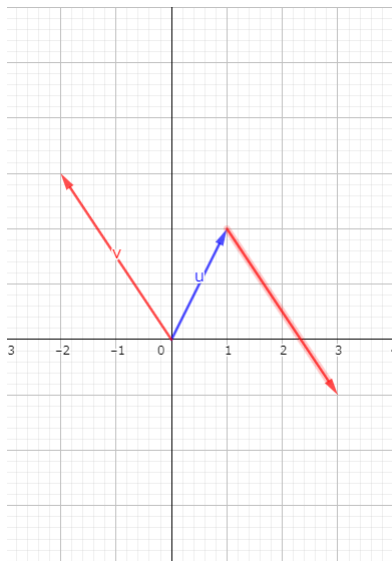
- Vector addition: move the starting point of the second vector to the end of the first vector, and draw a new arrow from the origin to the end of the second
- Vector subtraction: multiply the second vector by -1 and implement the addition

Geometry of Matrix Algebra: Vector (cont.)



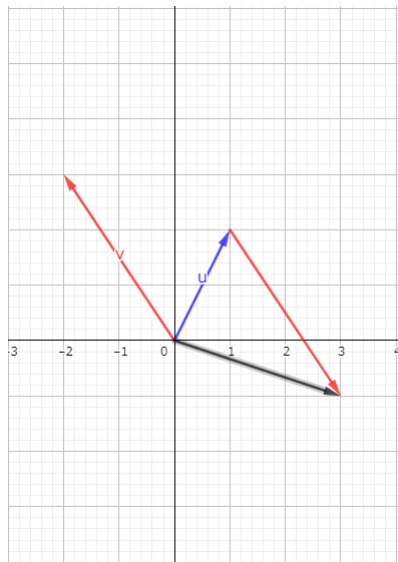
- Vector addition: move the starting point of the second vector to the end of the first vector, and draw a new arrow from the origin to the end of the second
- Vector subtraction: multiply the second vector by -1 and implement the addition

Geometry of Matrix Algebra: Vector (cont.)



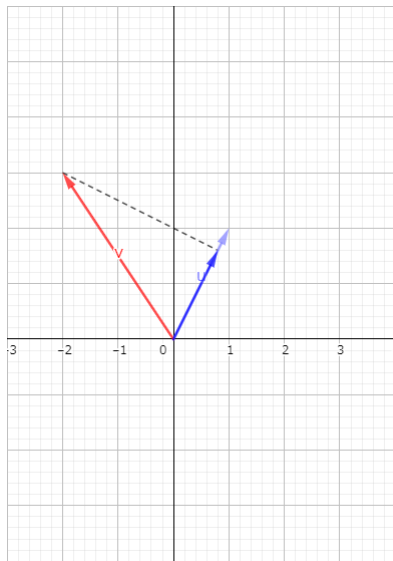
- Vector addition: move the starting point of the second vector to the end of the first vector, and draw a new arrow from the origin to the end of the second
- Vector subtraction: multiply the second vector by -1 and implement the addition

Geometry of Matrix Algebra: Vector (cont.)



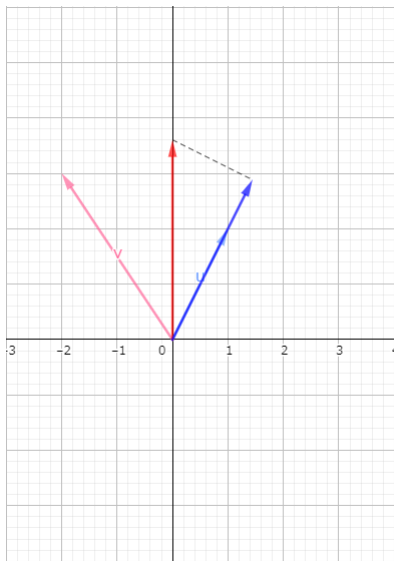
- Vector addition: move the starting point of the second vector to the end of the first vector, and draw a new arrow from the origin to the end of the second
- Vector subtraction: multiply the second vector by -1 and implement the addition

Geometry of Matrix Algebra: Vector (cont.)



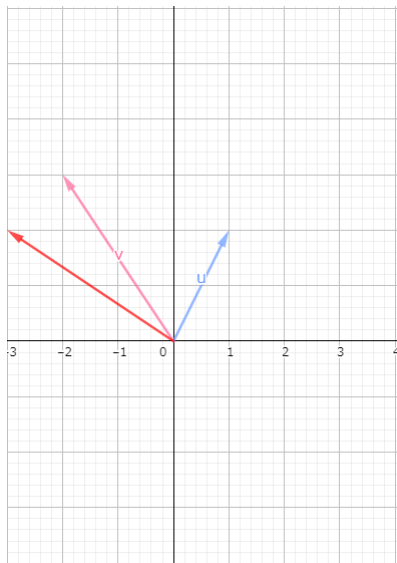
- Dot/inner product: describes the degree to which one vector overlaps another
- Product of the length of the first vector and that of the second one projected onto the first
- Dot product (or length of the projected vector) depends on the angle (θ) between the two
 - ▶ $\theta = 90^\circ$: dot product equals to 0

Geometry of Matrix Algebra: Vector (cont.)



- Dot/inner product: describes the degree to which one vector overlaps another
- Product of the length of the first vector and that of the second one projected onto the first
- Dot product (or length of the projected vector) depends on the angle (θ) between the two
 - ▶ $\theta = 90^\circ$: dot product equals to 0

Geometry of Matrix Algebra: Vector (cont.)



- Dot/inner product: describes the degree to which one vector overlaps another
- Product of the length of the first vector and that of the second one projected onto the first
- Dot product (or length of the projected vector) depends on the angle (θ) between the two
 - ▶ $\theta = 90^\circ$: dot product equals to 0

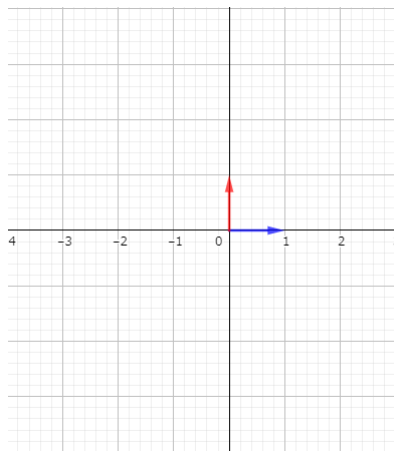
Vector Space

- We denote a set of points (i.e., real-valued vectors) on the d -dimensional coordinate/Euclidean space \mathbb{R}^d .
 - ▶ \mathbb{R} : set of points on a real-number line
 - ▶ \mathbb{R}^2 : set of points on a 2-D plane
 - ▶ \mathbb{R}^3 : set of points in 3-D space
 - ▶ ...
- A set of vectors **spans** a (vector) space if every points/vector in that psace can be written as a linear combination of vectors of that set.
 - ▶ **linear combination**: $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n$

Vector Space (cont.)

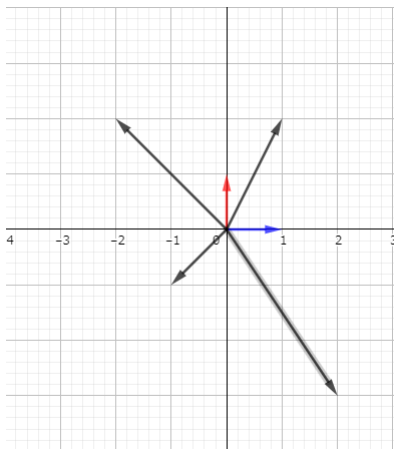
- When we cannot write any vector in a set as a linear combination of the others, we say the set of vectors is **linearly independent**.
- We call a set of linearly independent vectors which spans a (vector) space a **basis**.
 - ▶ The dimension of a space matches the number of vectors in its basis.

Vector Space (cont.)



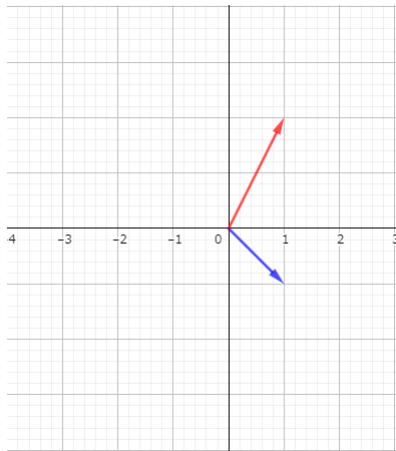
- Example:
 $\mathbf{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ spans \mathbb{R}^2 because all the points in \mathbb{R}^2 can be constructed as a linear combination of them.
 - ▶ $2\mathbf{u} - 3\mathbf{v}$
 - ▶ $-2\mathbf{u} + 2\mathbf{v}$
 - ▶ $\mathbf{u} + 2\mathbf{v}$
 - ▶ $-\mathbf{u} - \mathbf{v}$
 - ▶ ...

Vector Space (cont.)



- Example:
 $\mathbf{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ spans \mathbb{R}^2 because all the points in \mathbb{R}^2 can be constructed as a linear combination of them.
 - ▶ $2\mathbf{u} - 3\mathbf{v}$
 - ▶ $-2\mathbf{u} + 2\mathbf{v}$
 - ▶ $\mathbf{u} + 2\mathbf{v}$
 - ▶ $-\mathbf{u} - \mathbf{v}$
 - ▶ ...

Vector Space (cont.)



- \mathbf{u} and \mathbf{v} form a basis for \mathbb{R}^2 as they are linearly independent.
- They are not the only basis vectors for \mathbb{R}^2 .

► Example:

$$\mathbf{a} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

spans \mathbb{R}^2 and form a basis for \mathbb{R}^2

$$\star \frac{7}{3}\mathbf{a} - \frac{1}{3}\mathbf{b}$$

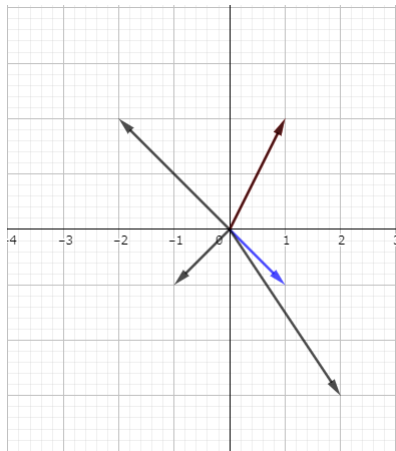
$$\star 0\mathbf{a} + \mathbf{b}$$

$$\star -2\mathbf{a} + 0\mathbf{b}$$

$$\star -\frac{1}{3}\mathbf{a} - \frac{2}{3}\mathbf{b}$$

$$\star \dots$$

Vector Space (cont.)



- \mathbf{u} and \mathbf{v} form a basis for \mathbb{R}^2 as they are linearly independent.
- They are not the only basis vectors for \mathbb{R}^2 .

► Example:

$$\mathbf{a} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

spans \mathbb{R}^2 and form a basis for \mathbb{R}^2

$$\star \frac{7}{3}\mathbf{a} - \frac{1}{3}\mathbf{b}$$

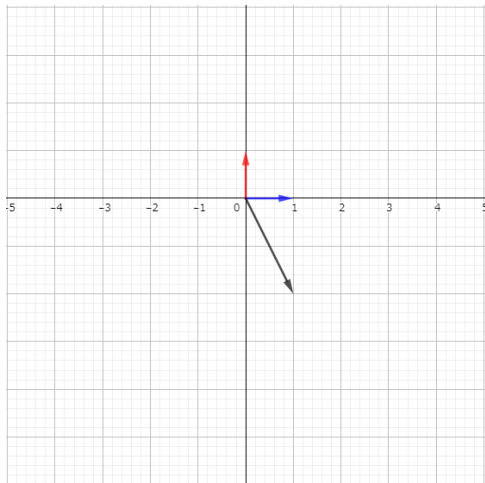
$$\star 0\mathbf{a} + \mathbf{b}$$

$$\star -2\mathbf{a} + 0\mathbf{b}$$

$$\star -\frac{1}{3}\mathbf{a} - \frac{2}{3}\mathbf{b}$$

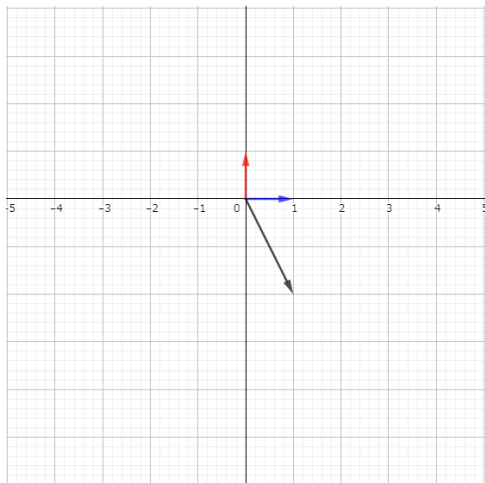
$$\star \dots$$

Matrix



- Geometrically, matrices describe linear transformation of the space/objects on the space
- Linear transformation: transformation of space while holding the origin
 - ▶ rotation
 - ▶ reflection
 - ▶ scaling
 - ▶ squeezing

Matrix (cont.)



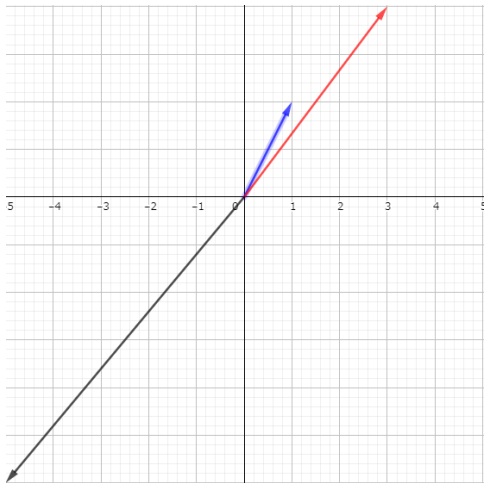
- Example: Matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

reflects and scales the space up.

$$\blacktriangleright \mathbf{A}\mathbf{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{A}\mathbf{v} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Matrix (cont.)



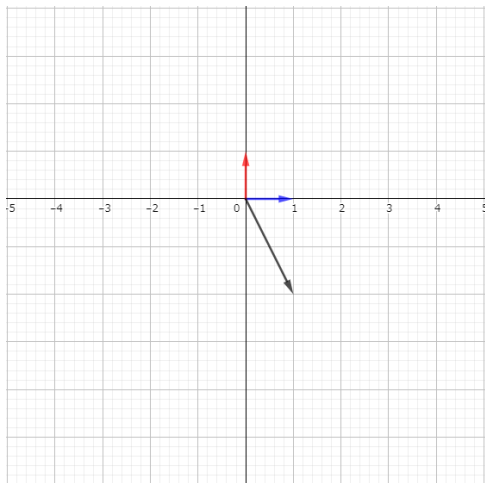
- Example: Matrix

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

reflects and scales the space up.

$$\blacktriangleright Au = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, Av = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Matrix (cont.)



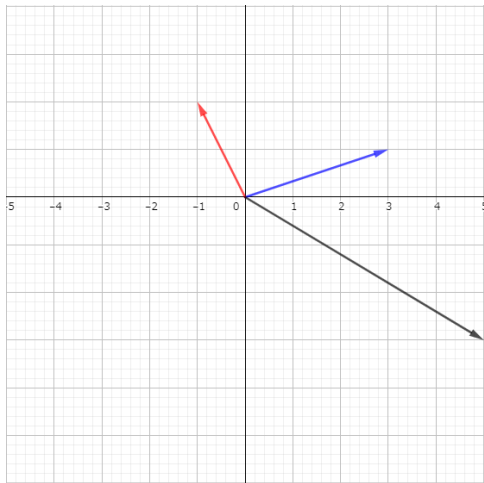
- Example: Matrix

$$B = \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix}$$

rotates and scales the space up.

$$\blacktriangleright Bu = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, Bv = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

Matrix (cont.)



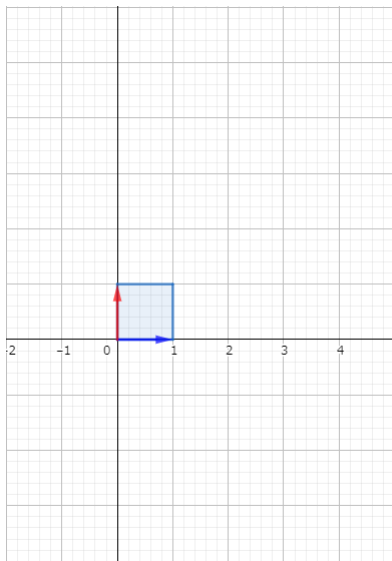
- Example: Matrix

$$B = \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix}$$

rotates and scales the space up.

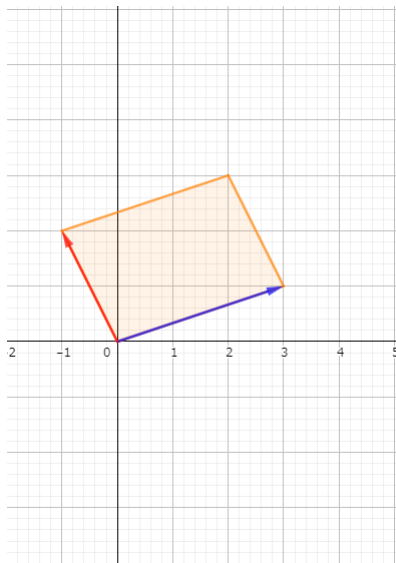
$$\blacktriangleright Bu = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, Bv = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$$

Determinant



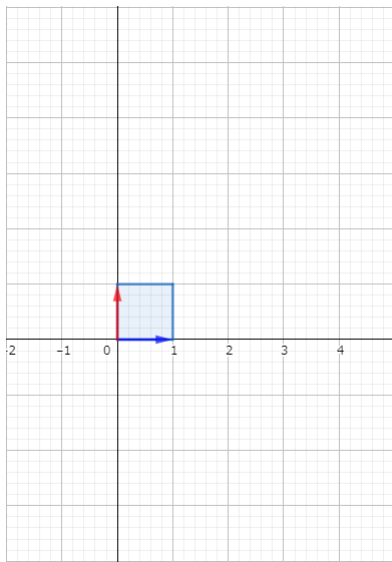
- Determinant of a square matrix represents the scale factor and the reflection of the linear transformation defined by the matrix.
- Example: $|B| = 7$
 - ▶ compare the area of rectangles defined by the basis vectors and one of the parallelogram by the transformed vectors

Determinant



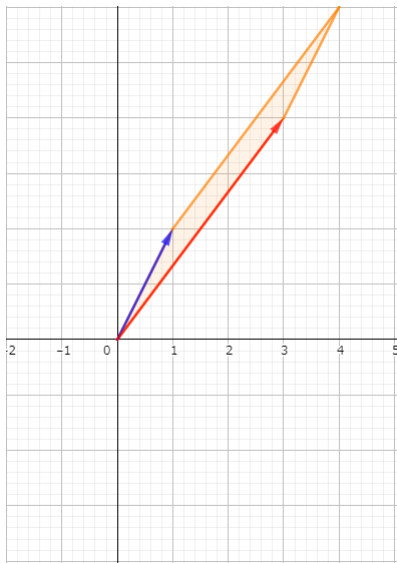
- Determinant of a square matrix represents the scale factor and the reflection of the linear transformation defined by the matrix.
- Example: $|B| = 7$
 - ▶ compare the area of rectangles defined by the basis vectors and one of the parallelogram by the transformed vectors

Determinant (cont.)



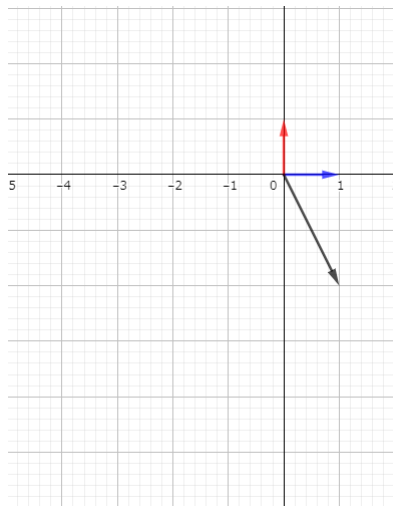
- Example: $|A| = -2$
 - ▶ area of orange parallelogram is 2
 - ▶ negative sign means that the linear transformation defined by A reflects the space

Determinant (cont.)



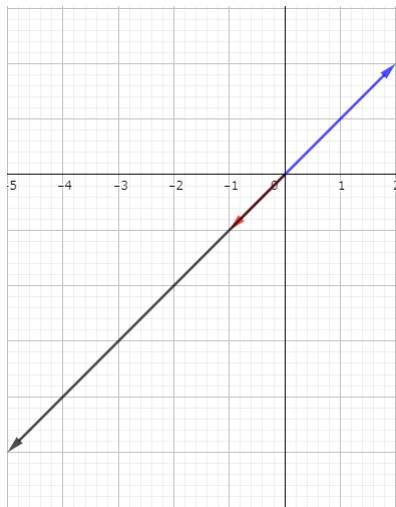
- Example: $|\mathbf{A}| = -2$
 - ▶ area of orange parallelogram is 2
 - ▶ negative sign means that the linear transformation defined by \mathbf{A} reflects the space

Determinant (cont.)



- What happens when the determinant equals to 0?
- Example: The determinant of $C = \begin{pmatrix} 2 & -1 \\ 2 & -1 \end{pmatrix}$ is 0.
- Transformed space degenerates to a lower dimensional space!

Determinant (cont.)



- What happens when the determinant equals to 0?
- Example: The determinant of $C = \begin{pmatrix} 2 & -1 \\ 2 & -1 \end{pmatrix}$ is 0.
- Transformed space degenerates to a lower dimensional space!

Matrix Rank

- Rank of a matrix is the number of linearly independent rows/columns.
- When $\text{rank}(\mathbf{A}) = \min(n, k)$, we say \mathbf{A} is **full rank**.
- Rank, determinant, invertibility
 - ▶ when a square matrix \mathbf{A} is full rank (= all the row/column vectors are linearly independent), $|\mathbf{A}| \neq 0$, so we can invert the matrix.
 - ▶ when \mathbf{A} is not full rank, $|\mathbf{A}| = 0$ and \mathbf{A} is singular.

Matrix Rank (cont.)

- Properties of matrix rank
 1. $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}')$
 2. $\text{rank}(\mathbf{A}\mathbf{A}') = \text{rank}(\mathbf{A}'\mathbf{A}) = \text{rank}(\mathbf{A})$
- Practical implications
 - ▶ if the coefficient matrix \mathbf{A} of a system of linear equations is singular...
 - ▶ some of its row vectors can be written as a linear combination of others
 - ▶ number of equations is smaller than the number of unknowns!

Relationship with Statistical Analysis

- Dot product as a measure of similarity
 - ▶ between variables (e.g., correlation coefficient)
 - ▶ between observations (e.g., cosine similarity)
- Multicollinearity: data matrix is not full rank
 - ▶ some column (= variable) can be written as a linear combination of others
 - ▶ $\mathbf{X}'\mathbf{X}$ is not full rank either $\rightarrow \mathbf{X}'\mathbf{X}$ is singular

Tomorrow

- Tomorrow
 - ▶ Probability
 - ▶ Random variable
 - ▶ Probability distribution
 - ▶ Moore and Siegel, Chapters 9-11.