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:

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + b_4 x_{4i} + b_5 x_{5i} + \cdots + b_{10} x_{10i} + e_i$$

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

Why Vectors and Matrices? (cont.)

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

$$\begin{cases} \frac{\partial}{\partial b_0} \sum_{i=1}^{n} \left\{ y_i - (b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_{10} x_{10i}) \right\}^2 = 0 \\ \frac{\partial}{\partial b_1} \sum_{i=1}^{n} \left\{ y_i - (b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_{10} x_{10i}) \right\}^2 = 0 \\ \quad \vdots \\ \frac{\partial}{\partial b_{10}} \sum_{i=1}^{n} \left\{ y_i - (b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_{10} x_{10i}) \right\}^2 = 0 \end{cases}$$

- 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., $\boldsymbol{a}$

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

Vector (cont.)

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

$$\boldsymbol{a}' = \begin{pmatrix} a_1 & a_2 \dots & a_k \end{pmatrix}$$

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

Vector (cont.)

- Norm (or length) of vector $\boldsymbol{a}$ is defined as

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

- 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

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

- We usually represent a matrix using a bold upper case. e.g., $\boldsymbol{A}$

- By convention, a_{ij} refers to an element in the ith row and the jth 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 $\boldsymbol{A}$ as a collection of column vectors.

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a_1} & \boldsymbol{a_2} & \dots & \boldsymbol{a_k} \end{pmatrix}$$

where

$$\boldsymbol{a_i} = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{pmatrix}$$

Matrix (cont.)

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

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha_1'} \\ \boldsymbol{\alpha_2'} \\ \vdots \\ \boldsymbol{\alpha_n'} \end{pmatrix}$$

where $\boldsymbol{\alpha_i}$ is a vector where elements of i th row of $\boldsymbol{A}$ are arranged in a column,

$$\boldsymbol{\alpha_i} = \begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{ik} \end{pmatrix}' = \begin{pmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ik} \end{pmatrix}$$

Matrix (cont.)

- Example: Matrix

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

can be repsented as a collection of column vectors

$$\boldsymbol{a_1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \boldsymbol{a_2} = \begin{pmatrix} 2 \\ -4 \end{pmatrix}$$

or as a collection of row vectors

$$\boldsymbol{\alpha_1} = \begin{pmatrix} 1 & 2 \end{pmatrix}' = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \boldsymbol{\alpha_2} = \begin{pmatrix} 1 & -4 \end{pmatrix}' = \begin{pmatrix} 1 \\ -4 \end{pmatrix}$$

Matrix (cont.)

- A transpose of a matrix can be obtained by flipping its rows and columns.

- Transpose of a $n \times k$ matrix $\boldsymbol{A}$ is denoted as $\boldsymbol{A}'$ (also written as $\boldsymbol{A}^T$)

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

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

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

Matrix (cont.)

- Example: The transpose of matrix

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

is

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

Matrix Addition/Subtraction

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

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

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

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

Matrix Multiplication

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

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

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

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

Matrix Multiplication (cont.)

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

$$\boldsymbol{a} \cdot \boldsymbol{b} = \boldsymbol{b} \cdot \boldsymbol{a} = a_1 b_1 + a_2 b_2 + \cdots + a_k b_k = \sum_j^k a_j b_j$$

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

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

Matrix Multiplication (cont.)

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

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

  - ij element of the resultant matrix $\boldsymbol{C}$ is the innner product of ith row of $\boldsymbol{A}$ and jth column of $\boldsymbol{B}$.
  - Therefore, $\boldsymbol{C}$ is a $n \times m$ matrix.

Matrix Multiplication (cont.)

- Example: Let

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

Then their product $\boldsymbol{AB}$ can be calculated as

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

Matrix Multiplication (cont.)

- Example: Let

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

Then their product $\boldsymbol{AB}$ can be calculated as

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

Matrix Multiplication (cont.)

- Example: Let

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

Then their product $\boldsymbol{AB}$ can be calculated as

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

Matrix Multiplication (cont.)

- Example: Let

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

Then their product $\boldsymbol{AB}$ can be calculated as

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

Matrix Multiplication (cont.)

- Example: Let

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

Then their product $\boldsymbol{AB}$ can be calculated as

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

Matrix Multiplication (cont.)

- $\boldsymbol{AB}$ is generally not equal to $\boldsymbol{BA}$, even if both are defined.

- Common properties of transpose matrices
  1. $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$
  2. $(\boldsymbol{A}'\boldsymbol{A})' = \boldsymbol{A}'(\boldsymbol{A}')' = \boldsymbol{A}'\boldsymbol{A}$

- Thinking of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ as $k \times 1$ matrices, their inner product can also be written as their product

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

Matrix Calculation: Exercises

- Let

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

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

Calculate the following.

1. $\boldsymbol{B}' + \boldsymbol{C}$
2. $\boldsymbol{A}\boldsymbol{D}$
3. $\boldsymbol{D}\boldsymbol{A}$
4. $\boldsymbol{B}\boldsymbol{C}$
5. $\boldsymbol{B}'\boldsymbol{C}'$

Matrix Types

- Square Matrix: Number of rows and columns are the same.
- Symmetric Matrix: A square matrix where $\boldsymbol{A} = \boldsymbol{A'}$. 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.

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

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

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

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

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i}^{k} a_{ii}$$

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

Determinant

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

Inverse

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

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

- We call $\boldsymbol{A}^{-1}$ as the inverse of $\boldsymbol{A}$.

- When we cannot define $\boldsymbol{A}^{-1}$, $\boldsymbol{A}$ is called singular.

- Some properties of the inverse
  1. if $\boldsymbol{A}$ is nonsingular, $\boldsymbol{A}^{-1}$ is unique
  2. $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$
  3. $(\boldsymbol{A}^{-1})' = (\boldsymbol{A}')^{-1}$
  4. $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
  5. $\det\boldsymbol{A} \neq 0 \Leftrightarrow \boldsymbol{A}$ 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

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

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

System of Linear Equations (cont.)

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

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

can be represented as

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

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

System of Linear Equations (cont.)

- Example: the system of regression equations

$$\begin{cases} y_1 = b_0 + b_1 x_{11} + \cdots + b_k x_{1k} + e_1 \\ y_2 = b_0 + b_1 x_{21} + \cdots + b_k x_{2k} + e_2 \\ \quad \vdots \\ y_n = b_0 + b_1 x_{n1} + \cdots + b_k x_{nk} + e_n \end{cases}$$

can be written as

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

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

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

System of Linear Equations (cont.)

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

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

  for $\boldsymbol{x}$.

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

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

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 $y = f(x_1, x_2, \cdots, x_n)$ using vector notation:

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

where $\boldsymbol{x}$ is the vector of inputs

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

Matrix Calculus: Preparation (cont.)

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

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = \begin{pmatrix} a_1 & a_2 & \ldots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \boldsymbol{a}'\boldsymbol{x}$$

where $\boldsymbol{a}$ is the vector of coefficients and $\boldsymbol{x}$ 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

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

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

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

Matrix Calculus: Preparation (cont.)

- Quadratic form of two variables $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and coefficient matrix $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ 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 $f(\boldsymbol{x})$ is a column vector of dimention n whose ith element is the partial derivative of $f(\boldsymbol{x})$ with respect to x_i.

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

  - 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 $S(\boldsymbol{b})$ be the function to compute the sum of squared residuals, where $\boldsymbol{b}$ is the vector of coefficients. Then,

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

Hessian

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

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

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

Rules for Matrix Calculus

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

Rules for Matrix Calculus: Exercise

- Let $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. Demonstrate that

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{x}$$

and

$$\frac{\partial^2}{\partial \boldsymbol{x} \partial \boldsymbol{x}'}(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A} + \boldsymbol{A}'$$

Multivariate Optimization

- First order condition

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

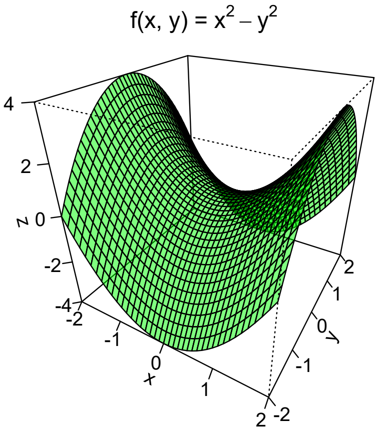
- Second order condition... $\rightarrow$ we use the Hessian matrix to determine local min/max!

Multivariate Optimization (cont.)

- Specifically, we examine the sign of the quadratic form of the Hessian matrix evaluated at $\boldsymbol{x}^\star$ where $\nabla f(\boldsymbol{x}^\star) = \boldsymbol{0}$.
  - $\boldsymbol{x}'\boldsymbol{H}^\star\boldsymbol{x}$ is the function of $\boldsymbol{x}$

- Second order condition:
  - $\boldsymbol{x}'\boldsymbol{H}^\star\boldsymbol{x} > 0$ for any values of $\boldsymbol{x} \to \boldsymbol{x}^\star$ is local min
    - ⋆ In this case, we say $\boldsymbol{H}^\star$ is positive definite
  - $\boldsymbol{x}'\boldsymbol{H}^\star\boldsymbol{x} < 0$ for any values of $\boldsymbol{x} \to \boldsymbol{x}^\star$ is local max
    - ⋆ In this case, we say $\boldsymbol{H}^\star$ is negative definite
  - sign of $\boldsymbol{x}'\boldsymbol{H}^\star\boldsymbol{x}$ depends on the values of $\boldsymbol{x} \to \boldsymbol{x}^\star$ is saddle point
    - ⋆ In this case, we say $\boldsymbol{H}^\star$ is indefinite

Multivariate Opimization (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 Opimization (cont.)

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

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

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

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

(*continued from the previous slide*)

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

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

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

Multivariate Opimization: Exercise

- Find $x^\star$ and $y^\star$ in which the gradient of $f(x, y) = x^3 + y^3 - xy$ equals $\mathbf{0}$, and determine whether the point is a local min, local max, or saddle point.

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)

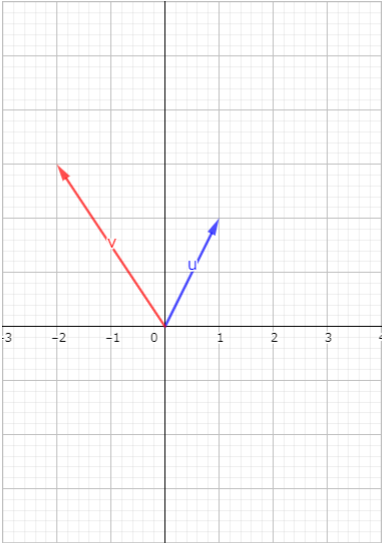
- Global v. local
  - For $(x^\star, y^\star)$ to be the global min/max, the Hessian matrix must be negative/positive definite at points other than $(x^\star, y^\star)$.
  - In such cases, $f(\boldsymbol{x})$ is globally concave/convex.
    - ⋆ e.g., $f(x, y) = -x^2 - y^2$, $f(x, y) = x^4 + 2y^2$

- (*In most (but not all) of the applications you encounter, you don't need to care about the second order condition...*)

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

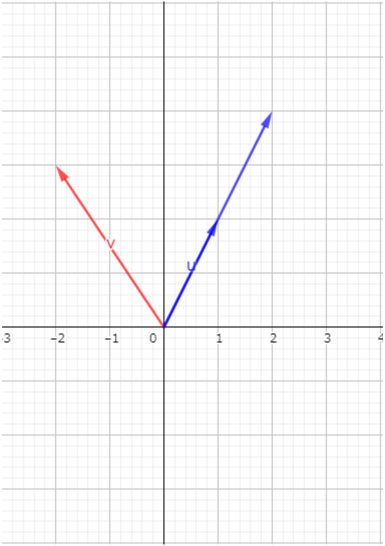


- A d-dimensional vector represents a point (more precisely, an arrow to the point) from the origin on the d-dimensional coordinate (Euclean) space.

- $\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$

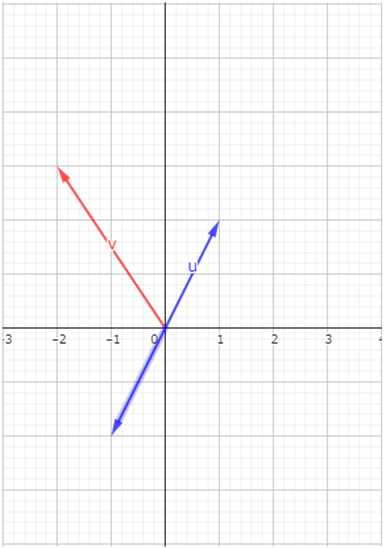
- Vector norm represents the length of a vector
  - We can show this using the Pythagorean theorem

Geometry of Matrix Algebra: Vector (cont.)



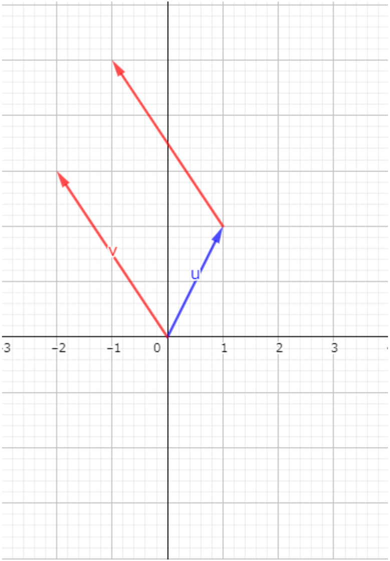
- Scalar product: stretching ($|c| > 1$) or contracting ($|c| < 1$) the original vector based on the size of the scalar.
- When $c < 0$, the original vector is reflected about the origin

Geometry of Matrix Algebra: Vector (cont.)



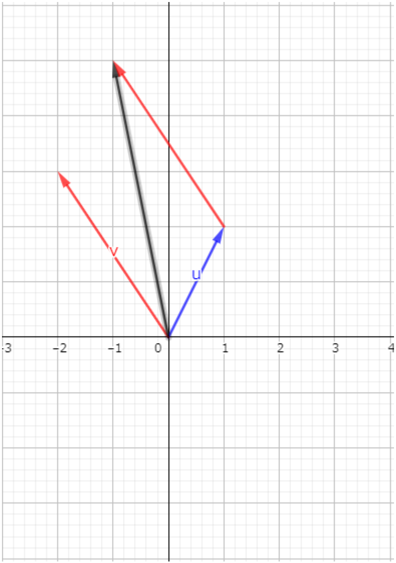
- Scalar product: stretching ($|c| > 1$) or contracting ($|c| < 1$) the original vector based on the size of the scalar.
- When $c < 0$, the original vector is reflected about the origin

Geometry of Matrix Algebra: Vector (cont.)



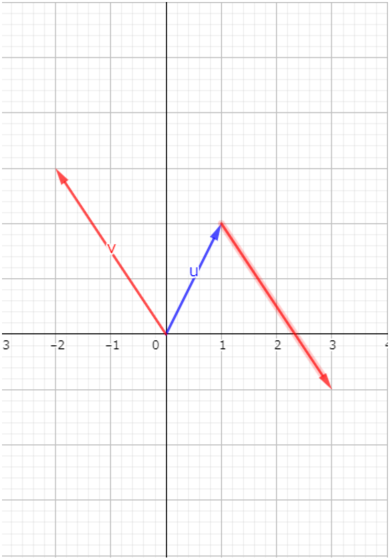
- Vector addition: move the starting point of the second vector to the end of the first vector, and draw a new arrow from the origin to the end of the second
- Vector subtraction: multiply the second vector by -1 and implement the addition

Geometry of Matrix Algebra: Vector (cont.)



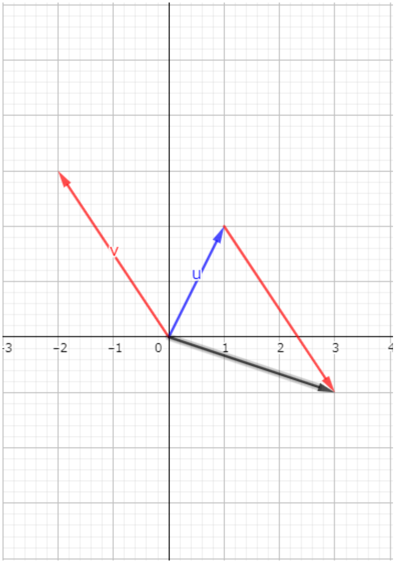
- Vector addition: move the starting point of the second vector to the end of the first vector, and draw a new arrow from the origin to the end of the second
- Vector subtraction: multiply the second vector by -1 and implement the addition

Geometry of Matrix Algebra: Vector (cont.)



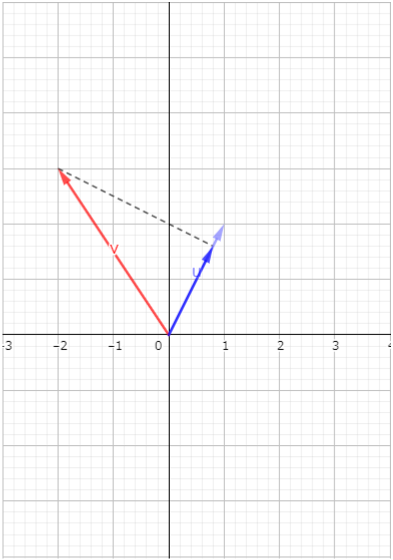
- Vector addition: move the starting point of the second vector to the end of the first vector, and draw a new arrow from the origin to the end of the second
- Vector subtraction: multiply the second vector by -1 and implement the addition

Geometry of Matrix Algebra: Vector (cont.)



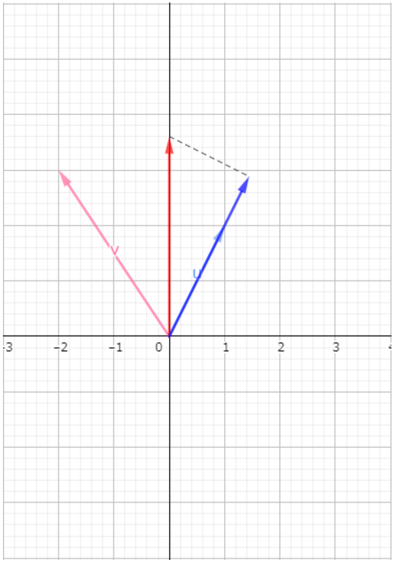
- Vector addition: move the starting point of the second vector to the end of the first vector, and draw a new arrow from the origin to the end of the second
- Vector subtraction: multiply the second vector by -1 and implement the addition

Geometry of Matrix Algebra: Vector (cont.)



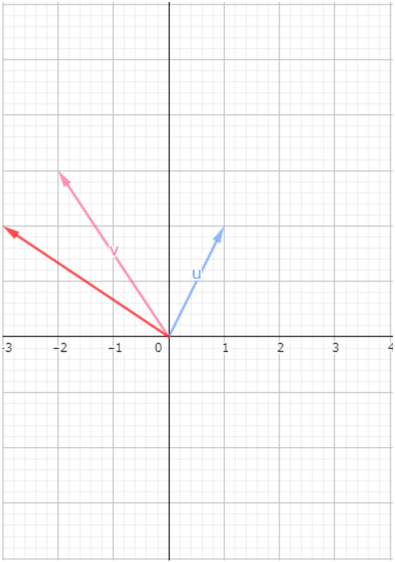
- Dot/inner product: describes the degree to which one vector overlaps another
- Product of the length of the first vector and that of the second one projected onto the first
- Dot product (or length of the projected vector) depends on the angle (θ) between the two
  - $\theta = 90°$: dot product equals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

Vector Space

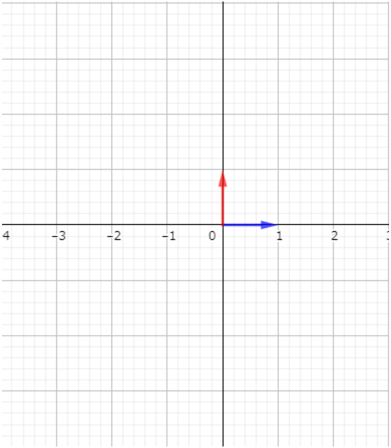
- We denote a set of points (i.e., real-valued vectors) on the d-dimensional coordinate/Euclidean space $\mathbb{R}^d$.
  - $\mathbb{R}$: set of points on a real-number line
  - $\mathbb{R}^2$: set of points on a 2-D plane
  - $\mathbb{R}^3$: set of points in 3-D space
  - ...

- 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: $c_1\boldsymbol{x_1} + c_2\boldsymbol{x_2} + \cdots + c_n\boldsymbol{x_n}$

Vector Space (cont.)

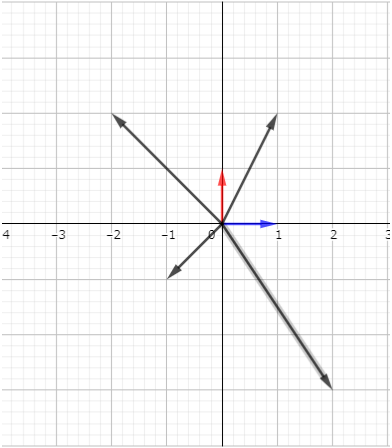
- When we cannot write any vector in a set as a linear combination of the others, we say they set of vectors is linearly independent.
- We call a set of linearly independent vectors which spans a (vector) space a basis.
  - The dimension of a space matches the number of vectors in its basis.

Vector Space (cont.)



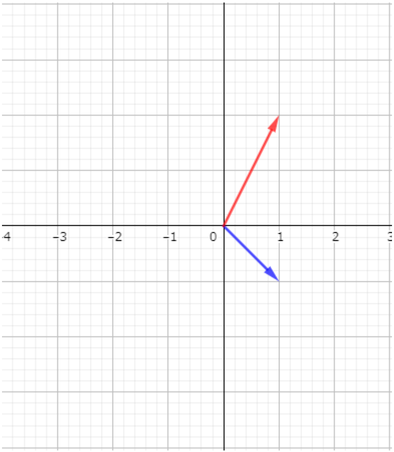
- Example:
$\boldsymbol{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ spans $\mathbb{R}^2$ because all the points in $\mathbb{R}^2$ can be constructed as a linear combination of them.
  - $2\boldsymbol{u} - 3\boldsymbol{v}$
  - $-2\boldsymbol{u} + 2\boldsymbol{v}$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

Vector Space (cont.)



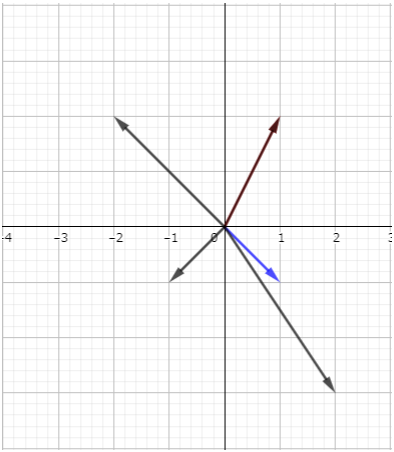
- Example: $\boldsymbol{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \boldsymbol{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ spans $\mathbb{R}^2$ because all the points in $\mathbb{R}^2$ can be constructed as a linear combination of them.
  - $2\boldsymbol{u} - 3\boldsymbol{v}$
  - $-2\boldsymbol{u} + 2\boldsymbol{v}$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

Vector Space (cont.)



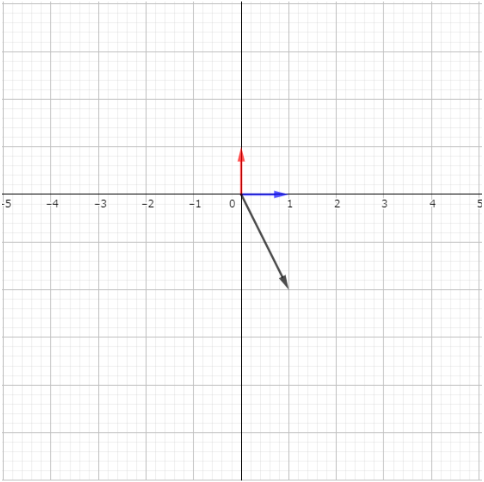
- $\boldsymbol{u}$ and $\boldsymbol{v}$ form a basis for $\mathbb{R}^2$ as they are linearly independent.
- They are not the only basis vectors for $\mathbb{R}^2$.
  - Example:
    $\boldsymbol{a} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$
    spans $\mathbb{R}^2$ and form a basis for $\mathbb{R}^2$
    - ⋆ $\frac{7}{3}\boldsymbol{a} - \frac{1}{3}\boldsymbol{b}$
    - ⋆ $0\boldsymbol{a} + \boldsymbol{b}$
    - ⋆ $-2\boldsymbol{a} + 0\boldsymbol{b}$
    - ⋆ $-\frac{1}{3}\boldsymbol{a} - \frac{2}{3}\boldsymbol{b}$
    - ⋆ ...

Vector Space (cont.)



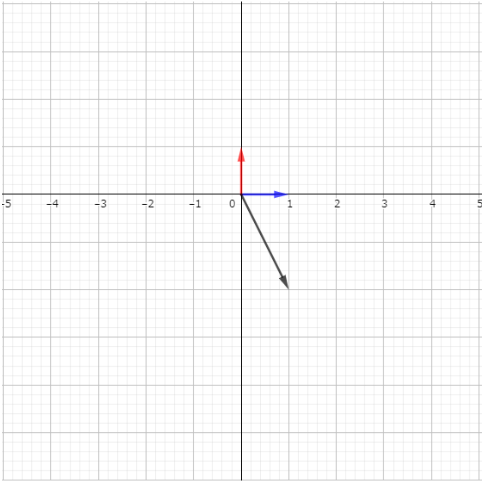
- $\boldsymbol{u}$ and $\boldsymbol{v}$ form a basis for $\mathbb{R}^2$ as they are linearly independent.
- They are not the only basis vectors for $\mathbb{R}^2$.
  - Example:
    $$\boldsymbol{a} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$
    spans $\mathbb{R}^2$ and form a basis for $\mathbb{R}^2$
    - ⋆ $\frac{7}{3}\boldsymbol{a} - \frac{1}{3}\boldsymbol{b}$
    - ⋆ $0\boldsymbol{a} + \boldsymbol{b}$
    - ⋆ $-2\boldsymbol{a} + 0\boldsymbol{b}$
    - ⋆ $-\frac{1}{3}\boldsymbol{a} - \frac{2}{3}\boldsymbol{b}$
    - ⋆ ...

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.)



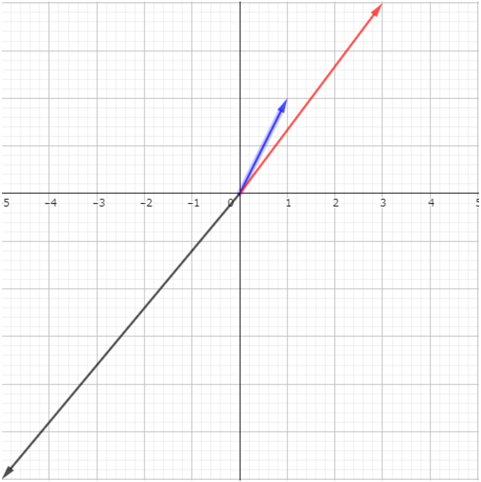
- Example: Matrix

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

reflects and scales the space up.

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

Matrix (cont.)

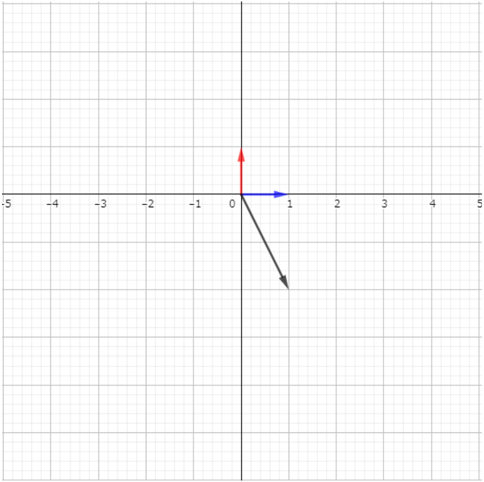


- Example: Matrix

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

  reflects and scales the space up.
  - $\boldsymbol{A}\boldsymbol{u} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \boldsymbol{A}\boldsymbol{v} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$

Matrix (cont.)



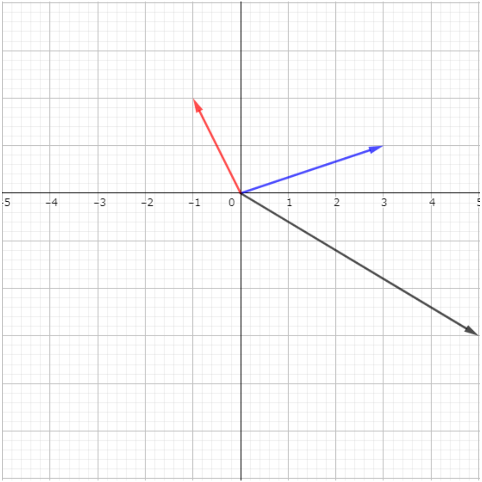
- Example: Matrix

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

  rotates and scales the space up.

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

Matrix (cont.)

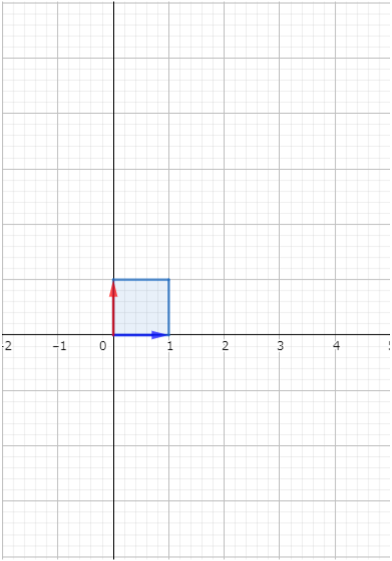


- Example: Matrix

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

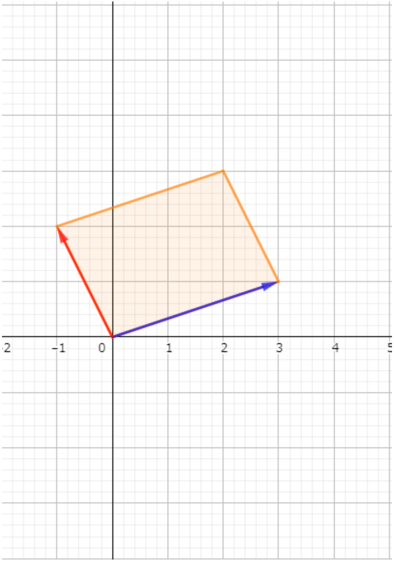
  rotates and scales the space up.
  - $Bu = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, Bv = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$

Determinant



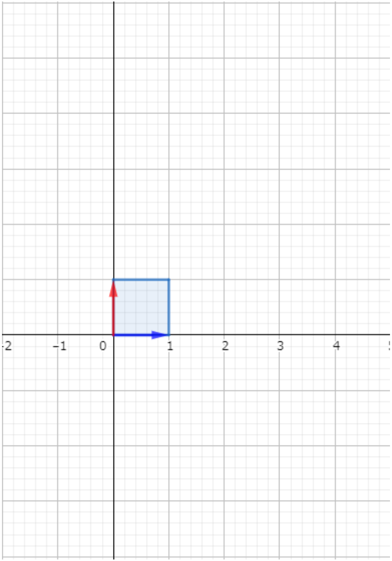
- Determinant of a square matrix represents the scale factor and the reflection of the linear transformation defined by the matrix.
- Example: $|\boldsymbol{B}| = 7$
  - compare the area of rectangles defined by the basis vactors and one of the parallelogram by the transformed vectors

Determinant



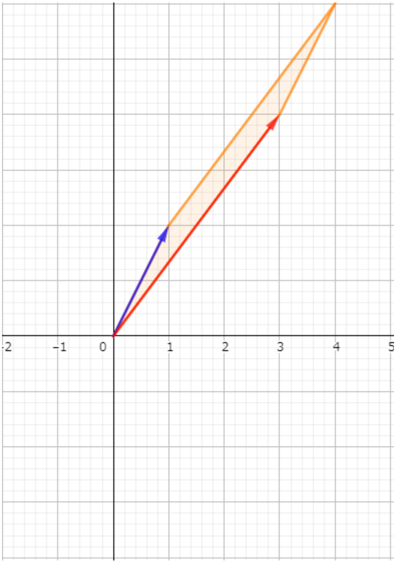
- Determinant of a square matrix represents the scale factor and the reflection of the linear transformation defined by the matrix.
- Example: $|\boldsymbol{B}| = 7$
  - compare the area of rectangles defined by the basis vactors and one of the parallelogram by the transformed vectors

Determinant (cont.)



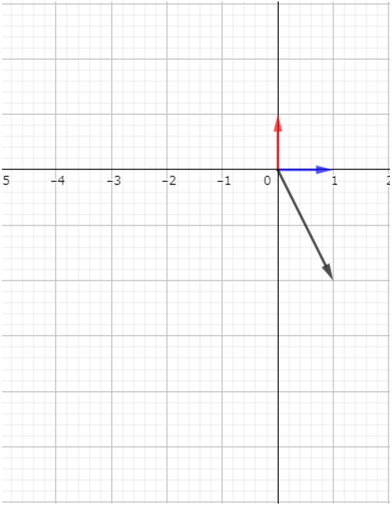
- Example: $|\boldsymbol{A}| = -2$
  - area of orange parallelogram is 2
  - negative sign means that the linear transformation defined by $\boldsymbol{A}$ reflects the space

Determinant (cont.)



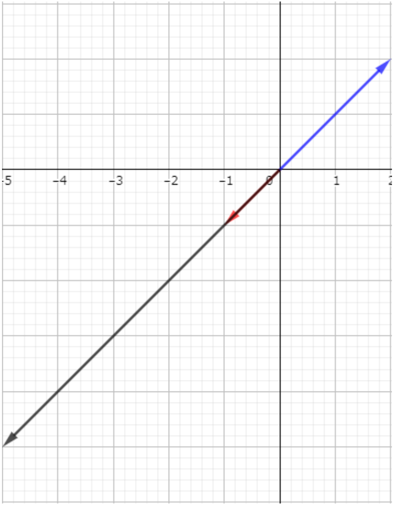
- Example: $|\boldsymbol{A}| = -2$
  - area of orange parallelogram is 2
  - negative sign means that the linear transformation defined by $\boldsymbol{A}$ reflects the space

Determinant (cont.)



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

Determinant (cont.)



- What happens when the determinant equals to 0?
- Example: The determinant of $\boldsymbol{C} = \begin{pmatrix} 2 & -1 \\ 2 & -1 \end{pmatrix}$ 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 $\mathrm{rank}(\boldsymbol{A}) = \min(n, k)$, we say $\boldsymbol{A}$ is full rank.
- Rank, determinant, invertibility
  - when a square matrix $\boldsymbol{A}$ is full rank ($=$ all the row/column vectors are linearly independent), $|\boldsymbol{A}| \neq 0$, so we can invert the matrix.
  - when $\boldsymbol{A}$ is not full rank, $|\boldsymbol{A}| = 0$ and $\boldsymbol{A}$ is singular.

Matrix Rank (cont.)

- Properties of matrix rank
  1. $\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{A'})$
  2. $\text{rank}(\boldsymbol{A}\boldsymbol{A}') = \text{rank}(\boldsymbol{A}'\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$

- Practical implications
  - if the coefficient matrix $\boldsymbol{A}$ 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
  - $\boldsymbol{X}'\boldsymbol{X}$ is not full rank either $\rightarrow$ $\boldsymbol{X}'\boldsymbol{X}$ is singular

Tomorrow

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