Basics of Linear Algebra

Weikai Chen, 2021/03/11

This is a lecture note for **Marxian Economic Thoery**, a course at Renmin University of China. This note is mainly for senior or graduate students in econ major, so I assume that students have taken a course in linear algebra before.

The purpose of this note is to *review* the basic concepts and methods in linear algebra and prepare the students for the **Perron-Frobenius Theorems** about positive and nonnegative matrices. Nonnegative matrices arise in many areas such as economics, population models, graph theory, Markov chains and so on. The Perron-Frobenius theory is one of the most powerful tools on nonnegative matrices and the workhorse in mathematical Marxian economics. Given its importance and the fact that it is new to most students, I will discuss P-F theorems in a separate **note**.

This Note is written in **Pluto Notebook**, a reactive notebook for **Julia**.

Linear algebra studies the **linear transformation** on **vector spaces**, which can be represented by **matrix**. We will focus on the n-dimensional Euclidean space $\mathbb{R}^n$, though most results discussed in this note can be easily generalized.

Vectors in $\mathbb{R}^n$

A vector in $\mathbb{R}^n$ is a list with n real number. For example, $x = (1, 2, 3)$ is a vector in $\mathbb{R}^3$. Let $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ and $k \in \mathbb{R}$, define **vector addition** and **scalar muliplication** by the following:

$$x + y = (x_1 + y_1, \ldots, x_n + y_n)$$

$$kx = (kx_1, \ldots, kx_n)$$

```
x = ▸Float64[1.5,  2.0]
```

```
x = [1.5, 2.0]
```

```
y = ▸Float64[0.5,  2.0]
```

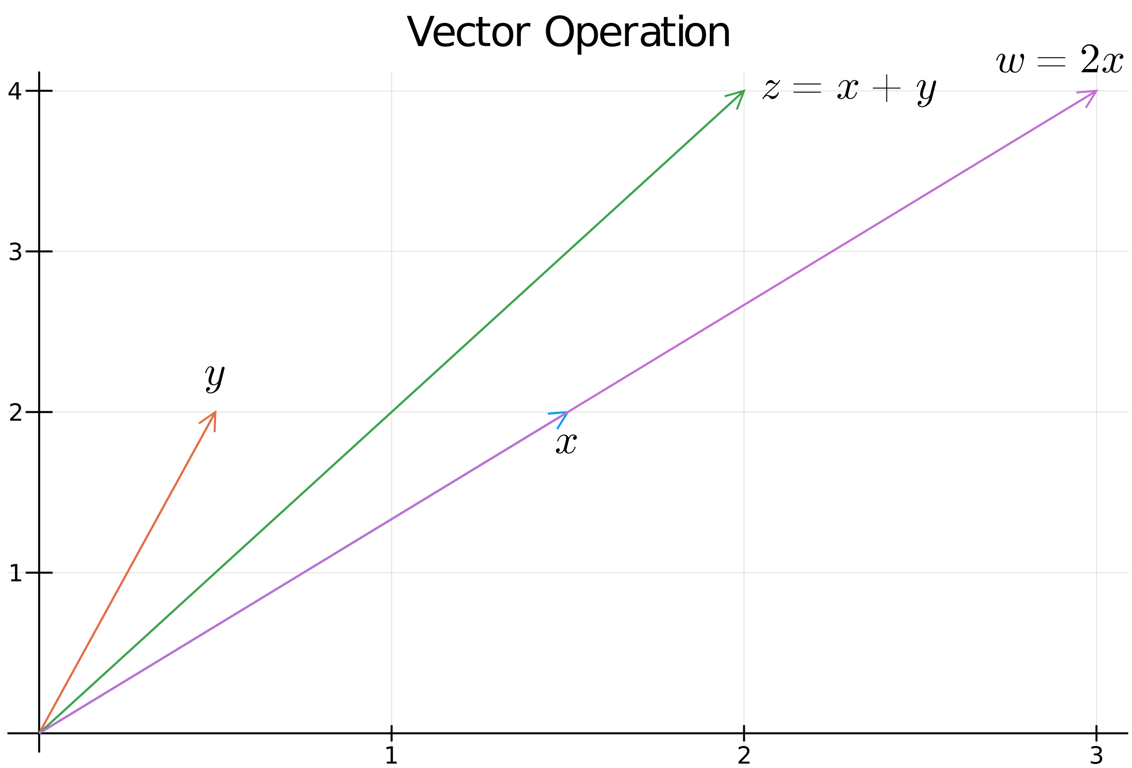
```
z = ▶Float64[2.0, 4.0]
```

```
z = x + y
```

```
w = ▶Float64[3.0, 4.0]
```

```
w = 2*x
```

Now let's plot those vectors.



Linear Combinations

Given a set of vectors $\{a^1, \ldots, a^k\}$ in $\mathbb{R}^n$, the new vectors we can create by performing linear operations are called **linear combinations** of $\{a^1, \ldots, a^k\}$.

That is, $b \in \mathbb{R}^n$ is a linear combination of $\{a^1, \ldots, a^k\}$ if

$$b = x_1 a^1 + \cdots + x_k a^k \text{ for some scalars } x_1, \ldots, x_k$$

In this context, the values $x_1, \ldots, x_k$ are called the *coefficients* of the linear combination.

The set of these linear combinations V is called the **span** of $\{a^1, \ldots, a^k\}$, denoted by $V = \mathrm{span}\{\mathrm{a}^1, \ldots, \mathrm{a}^\mathrm{k}\}$

Note that the cofficients of a linear combinbation $b \in \mathrm{span}\{\mathrm{a}^1, \ldots, \mathrm{a}^\mathrm{k}\}$ may not be unique. If for any $b \in \mathrm{span}\{\mathrm{a}^1, \ldots, \mathrm{a}^\mathrm{k}\}$, the coefficients are unique, then the set of vectors $\{a^1, \ldots, a^k\}$ are said to be **independent**.

A set of vectors $\{a^1, \ldots, a^k\}$ is *dependent* if they are not independent. Then one of them can be expressed as the linear combination of the rest.

A set of vectors $\{a^1, \ldots, a^k\}$ is a **basis** for the span V if it independent.

It can be shown that

1. If $\mathrm{span}\{\mathrm{a}^1, \ldots, \mathrm{a}^\mathrm{k}\} = \mathbb{R}^\mathrm{n}$, then $k \geq n$.
2. If $\{a^1, \ldots, a^k\}$ is independent, then $k \leq n$.

Therefore, $\{a^1, \ldots, a^k\}$ is a basis for $\mathbb{R}^n$ if and only if it is linearly independent and $k = n$.

Below is an example of basis for $\mathbb{R}^3$, called the standard basis:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

For any $x = (x_1, x_2, x_3) \in \mathbb{R}^3$, we can write

$$x = x_1 e_1 + x_2 e_2 + x_3 e_3$$

Inner Product and Norm

The **inner product** of vectors $x, y \in \mathbb{R}^n$ is defined as

$$x \cdot y = \sum_{i=1}^{n} x_i y_i$$

The inner product is also denoted by $x'y$.

Two vectors are called **orthogonal** if their inner product is zero.

The **norm** of a vector x represents its "length" (i.e., its distance from the zero vector) and is defined as

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

The expression $\|x - y\|$ is thought of as the distance between x and y.

```
4.75
dot(x,y) # the inner product of x and y
```

```
4.75
x'*y  # give the same result
```

```
2.5
```

```
norm(x)  # the norm of vector x
```

```
2.5
```

```
sqrt(x'*x) # give the same result
```

```
1.0
```

```
norm(x-y) # the distance between x and y
```

Matrices

A Matrice is a rectangular array of numbers.

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

is called an $m \times n$ matrix. If $m = n$, then A is called *square*.

The matrix formed by replacing a_{ij} by a_{ji} for every i and j is called the *transpose* of A, and denoted A' or $A^{\top}$. If $A = A'$, then A is called *symmetric*.

For a square matrix A, the i elements of the form a_{ii} for $i = 1, \ldots, n$ are called the *principal diagonal*.

A is called *diagonal* if the only nonzero entries are on the principal diagonal, i.e., $a_{ij} = 0$ for all $i \neq j$.

A diagonal matrix A is called the *identity matrix*, and denoted by I if $a_{ii} = 1$ for all i.

Denote each column of the matrix A as a vector by a^k for $k = 1, \ldots, n$, then A can be rewritten using these *column vectors* as

$$A = [a^1, \ldots, a^n]$$

Similarly we can write the matrix A using *row vectors*

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

Matrice Operation

Just as was the case for vectors, a number of algebraic operations are defined for matrices.

Scalar multiplication and addition are immediate generalizations of the vector case:

$$\gamma A = \gamma \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nk} \end{bmatrix} = \begin{bmatrix} \gamma a_{11} & \cdots & \gamma a_{1k} \\ \vdots & \vdots & \vdots \\ \gamma a_{n1} & \cdots & \gamma a_{nk} \end{bmatrix}$$

and

$$\begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \vdots & \vdots \\ a_{n1} & \cdots & a_{nk} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ \vdots & \vdots & \vdots \\ b_{n1} & \cdots & b_{nk} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1k} + b_{1k} \\ \vdots & \vdots & \vdots \\ a_{n1} + b_{n1} & \cdots & a_{nk} + b_{nk} \end{bmatrix}$$

In the latter case, the matrices must have the same shape in order for the definition to make sense.

```
A = 2×2 Array{Float64,2}:
     3.0  1.0
     2.0  5.0
```

```
A = [3.0 1
     2 5]
```

```
B = 2×2 Adjoint{Float64,Array{Float64,2}}:
     3.0  2.0
     1.0  5.0
```

```
B = A'
```

```
C = 2×2 Array{Float64,2}:
     6.0   3.0
     3.0  10.0
```

```
C = A + B
```

We also have a convention for *multiplying* matrix with vector.

For a square matrix A and a column vector x, we have

$$Ax = \begin{bmatrix} a_1 \cdot x \\ a_2 \cdot x \\ \vdots \\ a_m \cdot x \end{bmatrix}$$

Another usefull form of Ax is

$$Ax = [a^1, a^2, \ldots, a^n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 a^1 + x_2 a^2 + \cdots + x_n a^n \tag{1}$$

which is a linear combination of the set of column vectors $\{a^1, \ldots, a^n\}$.

```
a1 = ▶Float64[3.0,  2.0]
```

```
a1 = A[:,1] # the first column
```

```
a2 = ▶Float64[1.0,  5.0]
```

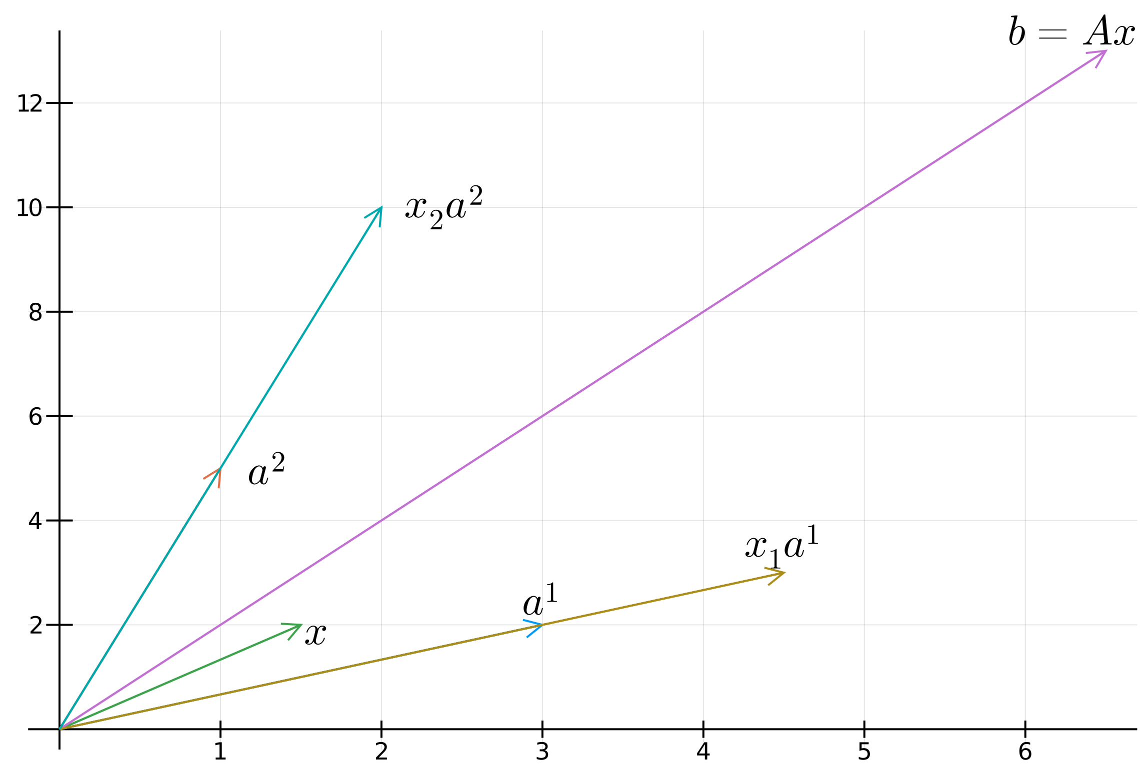
```
a2 = A[:,2] # the second column
```

```
b = ▶Float64[6.5,  13.0]
```

```
b = A*x
```

```
▶Float64[6.5,  13.0]
```

```
a1 * x[1] + a2 * x[2] # the same result
```



Matrix and Linear Transformation

Linear transformation and matrix representation

A function $f : \mathbb{R}^n \to \mathbb{R}^n$ is a **linear** if for all $x, y \in \mathbb{R}^n$ and all scalar α and β,

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

For any $n \times n$ matrix A, it is easy to check that the function $f(x) = Ax$ is linear.

In effect, a function f is linear *if and only if* there exists a matrix A such that $f(x) = Ax$ for all x.

Proof

First, let $\alpha = \beta = 0$, we have $f(0) = 0$.

Second, construct a matrix as follows: choose the standard basis $e_1, \ldots, e_n$, let $a^1 = f(e_1), \ldots, a^n = f(e_n)$, and $A = [a^1, \ldots, a^n]$.

Finally, show that $f(x) = Ax$. By $x = \sum_i x_i e_i$ and the linearity of f, we have

$$f(x) = x_1 f(e_1) + \cdots + x_n f(e_n) = x_1 a^1 + \cdots + x_n a^n = Ax$$

Inverse of linear transformation and inverse matrix

What is the range of the function $f(x) = Ax$?

Since $f(x) = Ax = x_1 a^1 + \ldots x_n a^n$, the range is just the span of the columns, i.e.,

$$\mathrm{Range(f) = span\{a^1, \ldots, a^n\}}$$

Moreover, if the columns are linearly independent, then the range is $\mathbb{R}^n$. That is, for any $b \in \mathbb{R}^n$, there exist a unique $x \in \mathbb{R}^n$ such that $f(x) = Ax = b$. Then we say the function f is **invertable**, and denote its inverse function by f^{-1}.

It could be verified that f^{-1} is also linear and thus there exist a matrix A^{-1} such that

$$f^{-1}(y) = A^{-1} y \quad \forall y \in \mathbb{R}^n$$

We called the matrix A^{-1} the **inverse matrix** of A, and by definition we have

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

and then

$$b = Ax \Rightarrow A^{-1} b = A^{-1} A x = I x = x$$

```
2×2 Array{Float64,2}:
  0.384615  -0.0769231
 -0.153846   0.230769
```

```
inv(A) # the inverse of matrix A
```

```
Float64[1.5,  2.0]
```

```
inv(A) * b  # x = inverse(A)*b
```

Composition of linear transformation and matrix multiplication

If $A_{m\times n}$ and $B_{n\times p}$ are two matrices, the linear transformation $f(v) = Av$ mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$, while $g(u) = Bu$ from $\mathbb{R}^p$ to $\mathbb{R}^m$. Then the composition $f \circ g : \mathbb{R}^p \to \mathbb{R}^n$ defined by

$$(f \circ g)(u) = f(g(u)), \quad \forall u \in \mathbb{R}^p$$

is also linear, and thus can be represented by a $m \times p$ matrix D. We say D is the multiplication of A and B, i.e., $D = AB$.

How can we calculate AB? If we write the matrices as

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}, B = [b^1, \ldots, b^p]$$

then their product $D = AB$ is formed by taking as its i, j-th element the inner product of the i-th row of A and the j-th column of B.

That is, $D = AB = (d_{ij})_{m\times p}$ where

$$d_{ij} = a_i \cdot b^j = \sum_{k=1}^{n} a_{ik} b_{kj}$$

```
D = 2×2 Array{Float64,2}:
     10.0  11.0
     11.0  29.0
```

```
D = A * B
```

Matrix and System of Linear Equations

Often, the numbers in the matrix represent coefficients in a system of linear equations

$$\begin{aligned} b_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ &\vdots \\ b_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{aligned}$$

The objective here is to solve for the "unknowns" $x_1, \ldots, x_n$ given $a_{11}, \ldots, a_{nn}$ and $b_1, \ldots, b_n$.

This system of equations can be written as

$$Ax = b$$

or

$$x_1a^1 + x_2a^2 + \cdots + x_na^n = b$$

Therefore, to solve $Ax = b$ is to find the coefficients of the linear combination.

Note

(1) If the columns of A are linearly independent, then their span is $\mathbb{R}^n$, so for any $b \in \mathbb{R}^n$ there is a unique solution.

(2) If the columns of A are linearly dependent, then the span is a subset of $\mathbb{R}^n$. If b is in the span, then there are multiple solutions; otherwise, there is no solution.

```
▸Float64[1.5,  2.0]
A\b # solve the system of equations Ax = b
```

Determinant

Given a square matrix A, how can we tell if its columns are linearly independent or not?

There is a function $\det(A)$ or $|A|$ called the **determinant** assigning a real number to any square matrix A, which could help us answer the above question.

In effect, $\det(A) \neq 0$ if and only if the columns of A are linearly indpendent.

When $n = 2$, let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

we have $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

The determinant of a matrix *determines* whether the column vectors are linearly independent or not.

```
determinant = 13.0
determinant = det(A) # the determinant of matrix A
```

```
13.0
A[1,1] * A[2,2] - A[1,2] * A[2,1] # same result
```

I won't dig into details for the calculation of determinants in general. Instead, let's look at its geometric intuition.

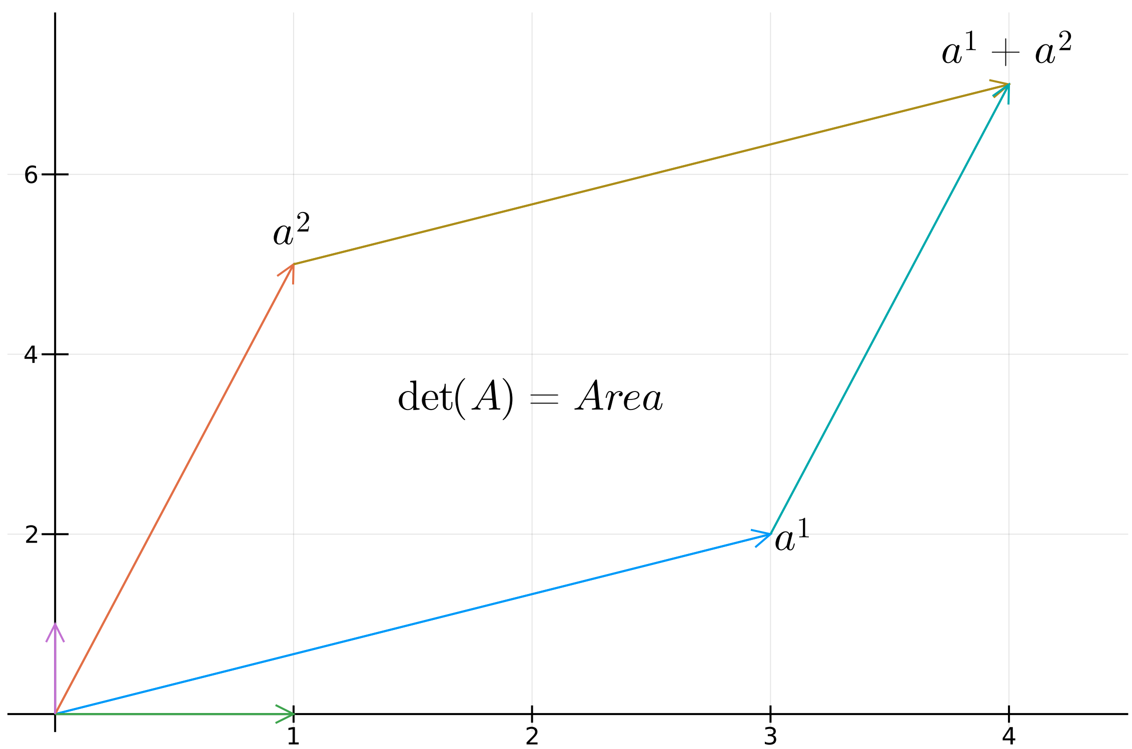
Take the example of $n = 2$. Note that $Ae_1 = a^1$ and $Ae^2 = a^2$. The linear transformation $f(x) = Ax$ transforms the square spanned by e_1 and e_2 into the parallelogram spanned by a^1 and a^2.

The determinant $\det(A)$ is the area scale factor of the transformation $f(x) = Ax$. That is

$$\det(A) = \frac{\text{Area of the parallelogram}}{\text{Area of the square}}$$

Since the area of the square is 1,

$$\det(A) = \text{Area of the parallelogram}$$



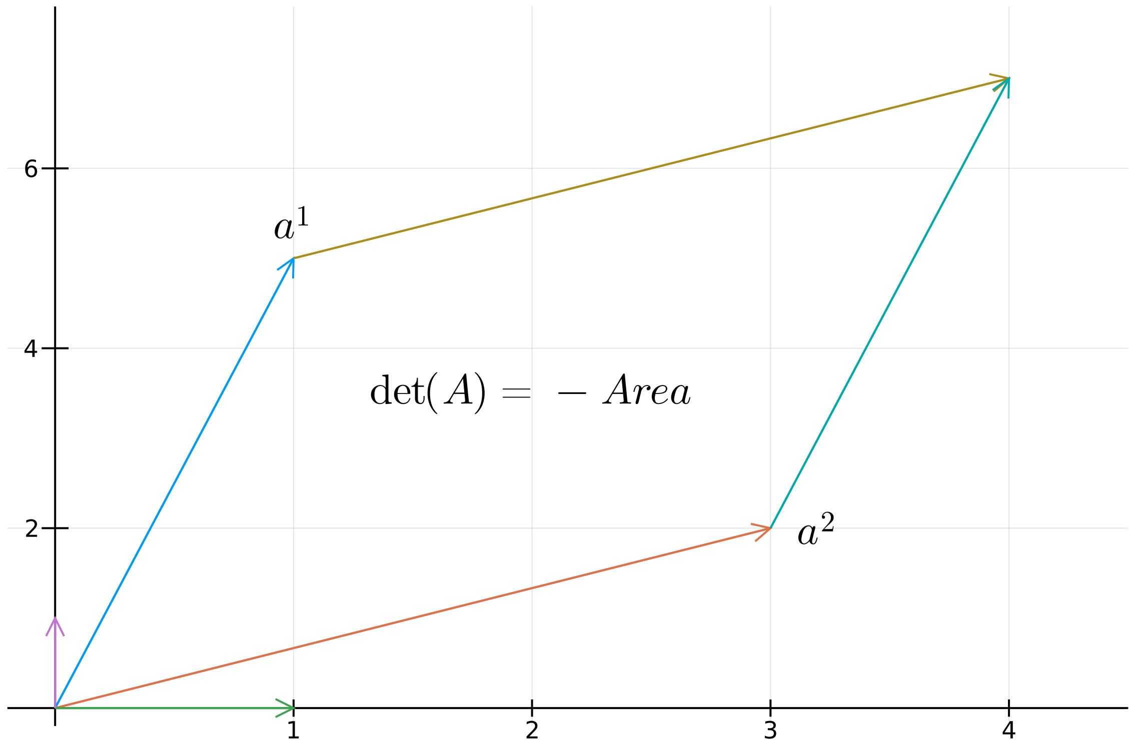
In this case, the area of the parallelogram is $\det(A) =$ 13.0. Therefore, the linear transformation $f(x) = Ax$ *stretch* the space with scale factor 13.0.

If a^1 and a^2 are linearly dependent, the area of the 'parallelogram' becomes zero, i.e., $\det(A) = 0$. If the columns of A are linearly dependent, the linear transformation $f(x) = Ax$ *compresses* the space into a lower-dimensional space, a line in this case.

Note that the determinant could be negative when the linear transformation *flips* the space. For example,

```
A_flip = 2×2 Array{Int64,2}:
          1  3
          5  2
```

```
A_flip = [1 3; 5 2]        # compare it with A = [3 1; 2 5]
```




```
-13.0
```

```
det(A_flip)   # det(A_flip) = -det(A)
```

```
13.0
```

```
det(A') # det(A') = det(A)
```

Eigenvalue and Eigenvector

Let A be an $n \times n$ square matrix.

If λ is scalar and v is a non-zero vector in $\mathbb{R}^n$ such that

$$Av = \lambda v$$

then we say that λ is an *eigenvalue* of A, and v is an *eigenvector*.

Thus, an eigenvector of A is a vector such that when the map $f(x) = Ax$ is applied, v is merely scaled.

```
Eigen{Float64,Float64,Array{Float64,2},Array{Float64,1}}
values:
2-element Array{Float64,1}:
 2.267949192431123
```

```
 5.732050807568878
vectors:
2×2 Array{Float64,2}:
 -0.806898  -0.343724
  0.59069   -0.939071
```

```julia
evals, evecs = eigen(A)  # find all eigenvalues and corresponding eigenvacters
```

```
v = Float64[-0.806898, 0.59069]
```

```julia
v = evecs[:,1] # one eigenvector
```

```
Float64[-1.83, 1.33966]
```

```julia
A * v  # Av
```

```
Float64[-1.83, 1.33966]
```

```julia
evals[1] * v # lambda_1 * v
```

```
u = Float64[-0.343724, -0.939071]
```

```julia
u =  evecs[:,2] # another eigenvector
```

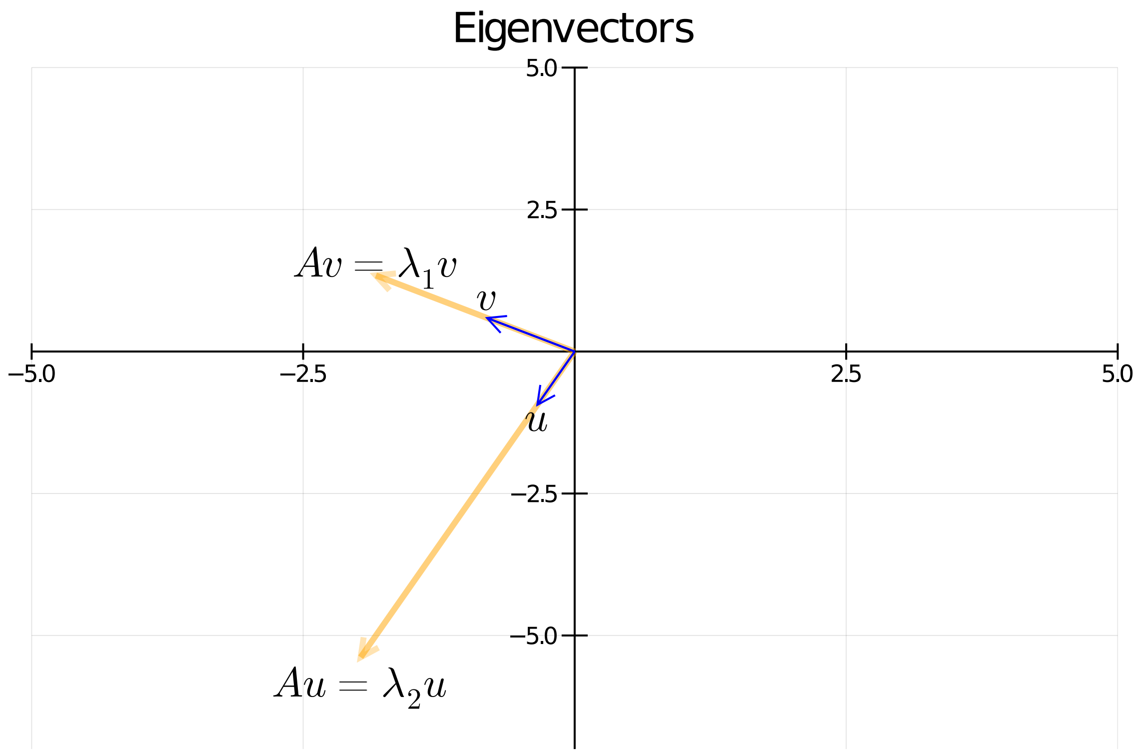
```
Float64[-1.97024, -5.3828]
```

```julia
A * u # Au
```

```
Float64[-1.97024, -5.3828]
```

```julia
evals[2] * evecs[:,2] # lambda_2 u
```

The next figure shows two eigenvectors, v, u, and their images under the linear transformation, Au and Av.



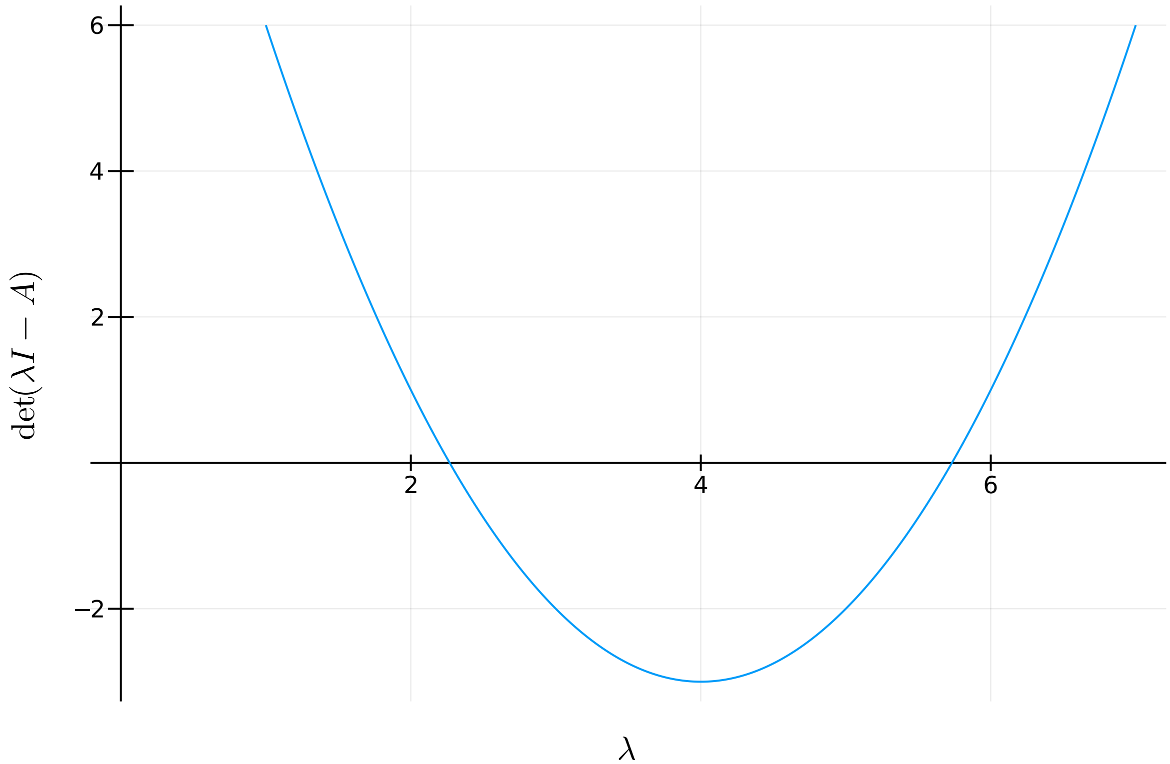
Suppose that $Av = \lambda v$, then the system of equatins

$$(\lambda I - A)v = 0$$

has a non-zero solution. Or the columns of the matrix $(\lambda I - A)$ is linearly dependent. Therefore,

$$\det(\lambda I - A) = 0$$

The next figure shows the plot of the *characteristic polynominal* $\det(\lambda I - A)$ as a function of λ.




```
begin
    Determinant(λ;matrix=A) = det(λ * Matrix(I,2,2) - matrix)
    λ = 1:0.01:7;
    determinant_λ = [Determinant(i) for i in λ];
    plot(λ, determinant_λ,legend = false, framestyle = :origin)
    plot!(xlab = L"\lambda", ylab = L"\det(\lambda I- A)")
end
```