Linear Algebra

Christos Michalopoulos

NTU, Department of Economics

September 24, 2011

Linear Equations

Definition

A **linear equation** in n-variables $x_1, x_2, ..., x_n$ is one that can be written in the form

$$a_1x_1 + a_2x_2 + ... + a_nx_n = b$$

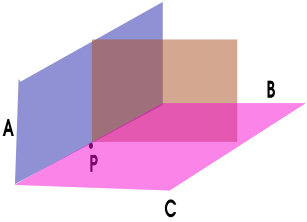
for coefficients a_i, $i = 1, ...n$ and b all constants.

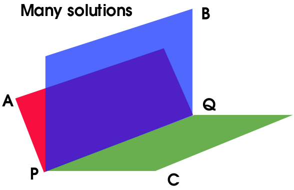
A linear equation in three variables corresponds to a plane in three dimensional space.

Solutions to a system of three equations are points that lie on all three planes

We can have the following type of solutions:

- Unique solution.
- No solution.
- Many solutions





Three planes A, B, and C intersect

Linear Equations: Example

Consider the following system of equations:

$$x_1 + x_2 + x_3 = 2$$

$$2x_1 + 3x_2 + x_3 = 3$$

$$x_1 - x_2 - 2x_3 = -6$$

By substitution we can see that the solution to this system is

$$x := (x_1, x_2, x_3) = (-1, 1, 2).$$

Definition

A **matrix** is a rectangular array of numbers. The numbers in each array are called the elements of the matrix.

In the example above, we get the following *augmented matrix*:

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 2 & 3 & 1 & 3 \\ 1 & -1 & -2 & -6 \end{pmatrix}$$

Elementary Transformations

Elementary transformations can be used to change a system of linear equations into another system of linear equations that has the same solution.

Elementary Transformations

- Interchange two equations.
- Multiply both sides of an equation by a non-zero constant.
- Add a multiple of one equation to another equation.

Systems of equations that are related through elementary transformations are called *equivalent systems*.

Elementary Row Operations

- Interchange two rows of a matrix.
- Multiply the elements of a row by a non-zero constant.
- Add a multiple of the elements of one row to the corresponding elements of another row.

Matrices that are related through elementary row operations are called *row equivalent matrices*.

Equivalence is indicated by the symbol $\approx$.

Definition

Gauss-Jordan Elimination is a method to solve a system of linear equations and is based on the above elementary transformations.

To solve our system of three linear equations above by the Gauss-Jordan method, we follow (in matrix form) the steps below:

- **Step 1**: Create zeros in column 1: Operation needed: R2+(-2)R1 and R3+(-1)R1. We get

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & -2 & -3 & -8 \end{pmatrix}$$

- **Step 2**: Create appropriate zeros in column 2: Operation needed: R1+(-1)R2 and R3+(2)R2. We get

$$\begin{pmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & -5 & 10 \end{pmatrix}$$

- **Step 3**: Make the (3,3)-element 1 (normalizing element): Operation needed: (-1/5)R3. We get

$$\begin{pmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

- **Step 4**: Create zeros in column 3: Operation needed: R1+(-2)R3 and R2+R3. We get

$$\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

- **Step 5**: Get the corresponding solution:

$$x_1 = -1, \;\; x_2 = 1, \;\; x_3 = 2.$$

Gauss-Jordan Elimination

- Write down the augmented matrix for the system of linear equations.
- Derive the reduced echelon form of the augmented matrix using elementary row operations. This is done by creating leading 1s, then zero above and below each each leading 1, column by column, starting with the first column.
- Write down the system of linear equations corresponding to the reduced echelon form. This system gives the solution.

Homogeneous Systems of Linear Equations

Definition

A system of linear equations is called **homogeneous** if all the constant terms are zero, i.e. $b = 0$.

- A homogeneous system of linear equations in n-variables has the trivial solution when $x_1 = x_2 = \cdots = x_n = 0$.
- A homogeneous system of linear equations that has more variables than equations has many solutions. One of these solutions is the trivial one.

- For vectors $\mathbf{u} = (u_1, ..., u_n) \in \mathbb{R}^n$ and $\mathbf{v} = (v_1, ..., v_n) \in \mathbb{R}^n$,we define the operations of addition and scalar multiplication as follows:
  - (i) Addition: $\mathbf{u} + \mathbf{v} = (u_1 + v_1, ..., u_n + v_n)$;
  - (ii) Scalar multiplication: $c\mathbf{u} = (cu_1, ..., cu_n)$.
- Further Properties:
  - (a) Commutativity: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
  - (b) Assosiativity: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.
  - (c) Zero vector: $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$.
  - (d) Negative vector: $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.
  - (e) Distributivity: $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$, $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$ and $c(d\mathbf{u}) = (cd)\mathbf{u}$.
  - (f) Scalar multiplication by 1: $1\mathbf{u} = \mathbf{u}$.

Example

Consider the vectors $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ in $\mathbb{R}^3$. They have the following two properties:

- They **span** $\mathbb{R}^3$: we can write an arbitrary vector (x, y, z) as a linear combination of those three vectors, i.e.

$$(x, y, z) = x(1, 0, 0) + y(0, 1, 0) + z(0, 0, 1).$$

- They are **linearly independent**: the identity

$$p(1, 0, 0) + q(0, 1, 0) + r(0, 0, 1) = 0$$

  for scalars p, q and r, has the unique solution $p = q = r = 0$.

Basis

A set of vectors that satisfies the above two properties is called a **basis**.

- Spanning implies that every vector in $\mathbb{R}^3$ can be expressed in terms of the above three vectors which are called the **standard basis of** $\mathbb{R}^3$.
- Independence means that noone of the three vectors can be written as a linear combination of the other two.

Standard Basis on $\mathbb{R}^n$

The set $\{(1,0,0,...,0),(0,1,0,...,0),...,(0,0,...,1)\}$ of n-vectors is the **standard basis** of $\mathbb{R}^n$ and its dimension is n.

Span, Linear Independence and Basis

Definition

The vectors $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ are said to **span** a space if every vector $\mathbf{v}$ in the space can be expressed as a linear combination of them, i.e.

$$\mathbf{v} = a\mathbf{v}_1 + b\mathbf{v}_2 + c\mathbf{v}_3.$$

Definition

The vectors $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ are said to be **linearly independent** if the identity

$$p\mathbf{v}_1 + q\mathbf{v}_2 + r\mathbf{v}_3 = 0$$

is only true for $p = q = r = 0$. Else they are **linearly dependent**.

Definition

A **basis** for a space is a set that spans the space and is linearly independent. The number of vectors in a basis is called the **dimension** of the space.

Strategy to show linear dependence/independence

For a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n\}$,

1. Write the equation $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$.
2. Express this equation as a system of simultaneous linear equations in the unknowns $c_1, c_2, \ldots, c_n$.
3. Solve these equations.

If the only solution is $c_i = 0$ all i, the set of vectors is *linearly independent*. If in the solution at least one of $c_i, i = 1, \ldots, n$ is not zero, then the set of vectors is *linearly dependent*.

Linear Dependence/Independence: Example

Example: Is the set of vectors $\{(2,0,0),(0,0,1),(-1,2,1)\}$ in $\mathbb{R}^3$ linearly independent?
Follow the strategy of the previous frame. Write,

$$a(2,0,0)+b(0,0,1)+c(-1,2,1)=(0,0,0)$$

which simplifies to $(2a-c,2c,b+c)=(0,0,0)$.
Solve the simultaneous equations,

$$\begin{aligned}2a-c&=0\\2c&=0\\b+c&=0\end{aligned}$$

The solution is $a=b=c=0$, therefore this set of vectors in linearly independent.

Dot Product

Definition

Let $\mathbf{u} = (u_1, \cdots, u_n)$ and $\mathbf{v} = (v_1, \cdots, v_n)$ be two vectors in $\mathbb{R}^n$. The **dot product** of $\mathbf{u}$ and $\mathbf{v}$ is denoted by $\mathbf{u} \cdot \mathbf{v}$ and is defined by

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

The dot product assigns a real number to each pair of vectors.

Dot Product: Properties

Properties of the Dot Product

Let $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$ and c a scalar. Then

(1) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$.

(2) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \ \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$.

(3) $c\mathbf{u} \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot c\mathbf{v}$.

(4) $\mathbf{u} \cdot \mathbf{u} \geq 0$ and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = 0$.

Norm of a vector

Definition

The **norm** of a vector $\mathbf{u} = (u_1, \cdots, u_n)$ in $\mathbb{R}^n$ is denoted by $\|\mathbf{u}\|$ and is defined by

$$\|\mathbf{u}\| = \sqrt{(u_1)^2 + (u_2)^2 + \cdots + (u_n)^2}.$$

Norm and Dot Product

The norm of a vector can also be written in terms of the dot product

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}.$$

Unit vector

A **unit vector** is a vector whose norm is one. If $\mathbf{v}$ is a nonzero vector, then the vector $\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$ is the unit vector in the direction of $\mathbf{v}$.

For the vector $(2, -1, 3)$, its norm is

$$\|(2, -1, 3)\| = \sqrt{2^2 + (-1)^2 + 3^2} = \sqrt{14}.$$

The normalized vector is

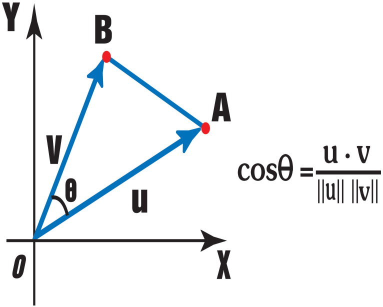
$$\frac{1}{\sqrt{14}}(2, -1, 3) = \left(\frac{2}{\sqrt{14}}, \frac{-1}{\sqrt{14}}, \frac{3}{\sqrt{14}}\right).$$

Angle between vectors

Definition

For two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the cosine of the angle θ between these vectors is

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\;\|\mathbf{v}\|}, \qquad 0 \le \theta \le \pi.$$



Angle between vectors

The **law of cosines** gives $AB^2 = OA^2 + OB^2 - 2(OA)(OB)\cos\theta$. Solving for $\cos\theta$ gives

$$\cos\theta = \frac{OA^2 + OB^2 - AB^2}{2(OA)(OB)}.$$

Then, we have

$$OA^2 + OB^2 - AB^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{v} - \mathbf{u}\|^2 = 2ac + 2db = 2\mathbf{u}\cdot\mathbf{v}.$$

Furthermore $2(OA)(OB) = 2\|\mathbf{u}\|\ \|\mathbf{v}\|$. Hence the angle θ between two vectors in $\mathbb{R}^2$ is given by

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\ \|\mathbf{v}\|}.$$

The Cauchy-Schwartz inequality (later) assures us that

$$\left|\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\ \|\mathbf{v}\|}\right| \leq 1.$$

Orthogonality

Two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

From the above angle formula, this happens when $\theta = \frac{\pi}{2}$.

Orthonormal sets

A set of unit pairwise orthogonal vectors, is called an **orthonornal set**.

For example, the standard basis for $\mathbb{R}^n$,

$$\{(1, 0, \cdots, 0), (0, 1, \cdots, 0), \cdots, (0, 0, \cdots, 1)\}$$

is an orthonormal set.

The Cauchy-Schwartz Inequality

For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, we have

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|.$$

Triangle Inequality

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Pythogonean Theorem

If $\mathbf{u} \cdot \mathbf{v} = 0$, then $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Proof of Cauchy-Schwartz Inequality

Trivial when $\mathbf{u} = 0$ so take $\mathbf{u} \neq 0$. Consider the vector $r\mathbf{u} + \mathbf{v}$, $r \in \mathbb{R}$. By dot product's properties we have

$$(r\mathbf{u}+\mathbf{v})\cdot(r\mathbf{u}+\mathbf{v}) = r^2(\mathbf{u}\cdot\mathbf{u})+2r(\mathbf{u}\cdot\mathbf{v})+\mathbf{v}\cdot\mathbf{v} \text{ and } (r\mathbf{u}+\mathbf{v})\cdot(r\mathbf{u}+\mathbf{v}) \geq 0.$$

Therefore $r^2(\mathbf{u}\cdot\mathbf{u}) + 2r(\mathbf{u}\cdot\mathbf{v}) + \mathbf{v}\cdot\mathbf{v} \geq 0$. Define $a = \mathbf{u}\cdot\mathbf{u}$, $b = 2\mathbf{u}\cdot\mathbf{v}$ and $c = \mathbf{v}\cdot\mathbf{v}$. Then $ar^2 + br + c \geq 0$. This implies the quadratic function $f(r) = ar^2 + br + c$ is never negative, its graph is a parabola and must have either one zero or no zeros. Using the discriminant, we have,

$$\text{no zeros} : b^2 - 4ac < 0, \quad \text{one zero} : b^2 - 4ac = 0.$$

We have, $b^2 - 4ac \leq 0 \Rightarrow (2\mathbf{u}\cdot\mathbf{v})^2 \leq 4(\mathbf{u}\cdot\mathbf{u})(\mathbf{v}\cdot\mathbf{v})$
$\Rightarrow (\mathbf{u}\cdot\mathbf{v})^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2$ and taking the square root on both sides completes the proof.

Proof of Triangle Inequality

By the properties of the norm

$$\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = \mathbf{u}\cdot\mathbf{u} + 2\mathbf{u}\cdot\mathbf{v} + \mathbf{v}\cdot\mathbf{v}$$

$$= \|\mathbf{u}\|^2 + 2\mathbf{u}\cdot\mathbf{v} + \|\mathbf{v}\|^2 \leq \|\mathbf{u}\|^2 + 2|\mathbf{u}\cdot\mathbf{v}| + \|\mathbf{v}\|^2.$$

By the Cauchy-Schwartz inequality, we have

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\ \|\mathbf{v}| + \|\mathbf{v}\|^2 = (\|\mathbf{u}+\mathbf{v}\|)^2.$$

Taking the square root of each side completes the proof.

Proof of the Pythagorean Theorem

By the properties of the dot product and the fact that $\mathbf{u} \cdot \mathbf{v} = 0$, we have

$$\begin{aligned} \|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v}) \cdot (\mathbf{u}+\mathbf{v}) &= \mathbf{u} \cdot \mathbf{u} + 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2. \end{aligned}$$

This completes the proof.

Types of Matrices

Square 3×3 matrix: $\begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix}$.

Diagonal 3×3 matrix: $\begin{pmatrix} \alpha_{11} & 0 & 0 \\ 0 & \alpha_{22} & 0 \\ 0 & 0 & \alpha_{33} \end{pmatrix}$.

This is written as $diag(\alpha_{11}, \alpha_{22}, \alpha_{33})$.

Identity 3×3 matrix: $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Null 3×3 matrix: $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Upper Triangular 3×3 matrix: $\begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ 0 & \alpha_{22} & \alpha_{23} \\ 0 & 0 & \alpha_{33} \end{pmatrix}$.

Lower Triangular 3×3 matrix: $\begin{pmatrix} \alpha_{11} & 0 & 0 \\ \alpha_{21} & \alpha_{22} & 0 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix}$.

Take a complex number $z = a + bi$ where $i = \sqrt{-1}$. Its conjugate is $\overline{z} = a - bi$. A **complex matrix** has complex elements.

Definition

The **conjugate** of a matrix A, denoted by $\overline{A}$ is obtained by taking the conjugate of each element of the matrix. The **conjugate transpose** is defined by $A^* = \overline{A}^{\mathsf{T}}$. A square matrix A is called **Hermitian**, if $A = A^*$.

Example: Matrix $C = \begin{pmatrix} 2 & 3-4i \\ 3+4i & 6 \end{pmatrix}$ is Hermitian. We have,

$$\overline{C} = \begin{pmatrix} 2 & 3+4i \\ 3-4i & 6 \end{pmatrix}, \quad C^* = \overline{C}^{\mathsf{T}} = \begin{pmatrix} 2 & 3-4i \\ 3+4i & 6 \end{pmatrix} = C.$$

Properties of Conjugate Transpose

(a) $(A+B)^* = A^* + B^*$

(b) $(zA)^* = \overline{z}A^*$

(c) $(AB)^* = B^*A^*$

(d) $(A^*)^* = A$

Matrix operations I

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are **equal** if they have the have the same number of rows and columns and if $\alpha_{ij} = \beta_{ij}$ for every i, j.

The **transpose** of an $m \times n$ matrix $\mathbf{A}$, denoted as A^{T}, is the $n \times m$ matrix whose $(i, j)^{th}$ element is the $(j, i)^{th}$ element of $\mathbf{A}$.

A matrix $\mathbf{A}$ is said to be **symmetric** if $\mathbf{A} = \mathbf{A}^{\mathsf{T}}$. Note that a diagonal matrix is symmetric while a triangular one is not.

Matrix addition works for matrices of the same size and it involves element-by-element addition. Also matrix addition is commutative i.e. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ and associative i.e. $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$. Also $\mathbf{A} + \mathbf{0} = \mathbf{A}$ for the matrix of zeros $\mathbf{0}$.

Furthermore, $c\mathbf{A} = \mathbf{A}c \ \forall c \in \mathbb{R}$, $-\mathbf{A} = -1 \times \mathbf{A}$ and $\mathbf{A} + (-\mathbf{A}) = \mathbf{A} - \mathbf{A} = \mathbf{0}$.

• The matrix multiplication is denoted by $\mathbf{AB}$ and this operation is valid only when the number of columns of $\mathbf{A}$ is the same as the number of rows of $\mathbf{B}$. This is usually expressed by saying that the matrices are *conformable to multiplication*.

• Take $\mathbf{A}_{m\times n}$ and $\mathbf{B}_{n\times p}$. Then $\mathbf{AB}$ works and the resulting matrix has dimension $m \times p$ with elements defined by $\sum_{k=1}^{n} \alpha_{ik}\beta_{kj}$.

• Note that matrix multiplication is NOT commutative, i.e. $\mathbf{AB} \neq \mathbf{BA}$. Actually $\mathbf{BA}$ may not even be defined in some cases.

• Assosiativity of matrix multiplication and distributivity with respect to matrix addition hold, i.e.

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} \text{ and } \mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}.$$

Matrix operations II: Econometrics example

Take the simple two-variable regression model,

$$Y_i = \beta_1 + \beta_2 X_i + u_i.$$

With the help of matrix algebra this can be written as,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$$

if we define, $\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}$, $\mathbf{X} = \begin{pmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ $\mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$.

Here we see the matrix operations of summation and multiplication in action.

Take two d-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$. Their **inner product** is defined as

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\mathsf{T}\mathbf{y} = \sum_{i=1}^{d} x_i y_i = \text{scalar}.$$

For an m-dimensional vector $\mathbf{x}$ and an n-dimensional vector $\mathbf{y}$, we define their **outer product** to be the matrix $\mathbf{x}\mathbf{y}^\mathsf{T}$.

Outer product example:

$$\mathbf{x}\mathbf{y}^\mathsf{T} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & x_1 y_3 \\ x_2 y_1 & x_2 y_2 & x_2 y_3 \\ x_3 y_1 & x_3 y_2 & x_3 y_3 \end{bmatrix}_{3 \times 3}.$$

Note that,

$$\mathbf{x} \cdot \mathbf{x} = \mathbf{x}^\mathsf{T}\mathbf{x} = \sum_{i=1}^{d} x_i^2 \geq 0.$$

The above includes the standard Euclidean norm denoted as

$$\| \mathbf{x} \| = (\mathbf{x}^\mathsf{T}\mathbf{x})^{1/2}.$$

Transformations I

Definition

A transformation (mapping) $T : \mathbb{R}^n \to \mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{u} \in \mathbb{R}^n$ a unique vector $\mathbf{v} \in \mathbb{R}^m$. $\mathbb{R}^n$ is called the domain of T and $\mathbb{R}^m$ is the codomain. We write $T(\mathbf{u}) = \mathbf{v}$ and $\mathbf{v}$ is the image of $\mathbf{u}$ under T.

Dilation and Contraction

Consider the transformation

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = r\begin{bmatrix} x \\ y \end{bmatrix}, \quad r \in \mathbb{R}_+.$$

If $r > 1$, T moves points away from the origin and is called a **dilation** of factor r. If $0 < r < 1$, T moves points closer to the origin and is called a **contraction** of factor r.

Reflection

Consider the transformation

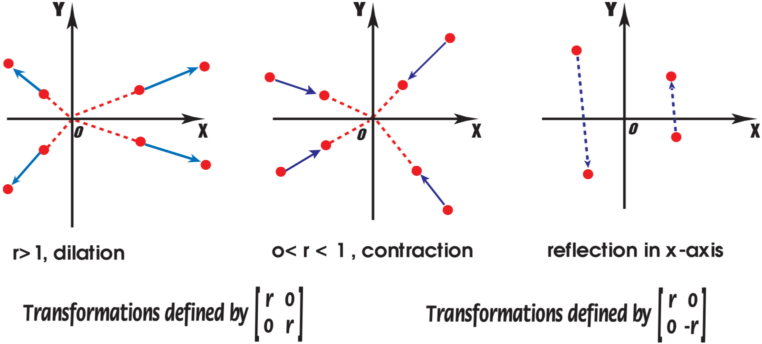
$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ -y \end{bmatrix}.$$

T maps every point in $\mathbb{R}^2$ into its mirror image in the x-axis. T is called a **reflection**.

Rotation about the origin

Consider a rotation T about the origin through an angle θ. T is defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$



Rotation of $\pi/2$ about the origin: Since $\cos(\pi/2) = 0$ and $\sin(\pi/2) = 1$, we have

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Transformations IV

Definition

A **translation** is a transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ defined by

$$T(\mathbf{u}) = \mathbf{u} + \mathbf{v}$$

for a fixed vector $\mathbf{v}$.

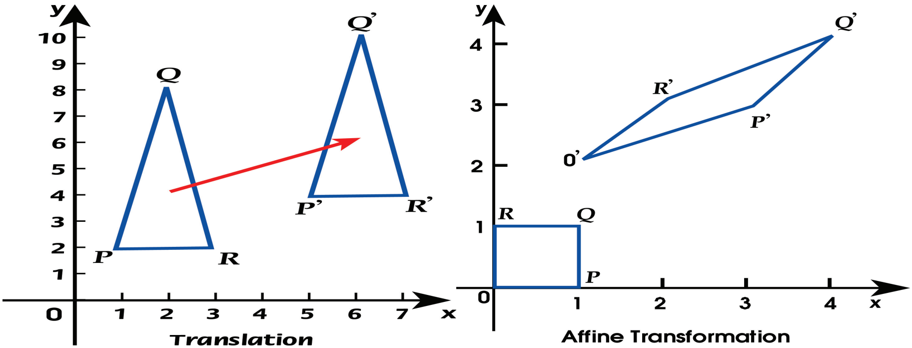
Definition

An **affine transformation** is a transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ defined by

$$T(\mathbf{u}) = A\mathbf{u} + \mathbf{v}$$

for a fixed vector $\mathbf{v}$ and matrix A.

Transformations: Examples



Matrix Transformations

Definition

Let A be an $m \times n$ matrix and $\mathbf{x}$ be an element of $\mathbb{R}^n$ written in column matrix form. A defines a **matrix transformation** $T(\mathbf{x}) = A\mathbf{x}$ of $\mathbb{R}^n$ into $\mathbb{R}^m$ with domain $\mathbb{R}^n$ and codomain $\mathbb{R}^m$. The vector $A\mathbf{x}$ is the image of $\mathbf{x}$.

Matrix transformations map line segments into line segments (or points). If the matrix is invertible the transformation also maps parallel lines into parallel lines.

Definition

A **composite transformation** $T = T_2 \circ T_1$ is given by

$$T(\mathbf{x}) = T_2(T_1(\mathbf{x})) = T_2(A_1\mathbf{x}) = A_2A_1\mathbf{x}$$

for matrix transformations $T_1(\mathbf{x}) = A_1\mathbf{x}$ and $T_2(\mathbf{x}) = A_2\mathbf{x}$.

Definition

An **orthogonal** matrix A is an invertible matrix with the property $A^{-1} = A^{\intercal}$. An **orthogonal transformation** is a transformation $T(\mathbf{u}) = A\mathbf{u}$ where A is an orthogonal matrix.

Theorem

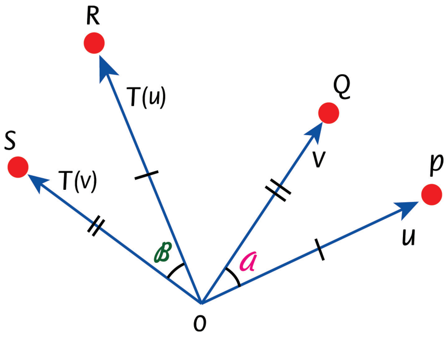
Let T be an orthogonal transformation on $\mathbb{R}^n$ and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Let P, Q be points in $\mathbb{R}^n$ defined by $\mathbf{u}$ and $\mathbf{v}$ and let R, S be their images under T. Then

$$\|\mathbf{u}\| = \|T(\mathbf{u})\|$$

$$\textit{angle between } \mathbf{u} \textit{ and } \mathbf{v} = \textit{angle between } T(\mathbf{u}) \textit{ and } T(\mathbf{v}).$$

Orthogonal transformations preserve norms, angles and distances.

Orthogonal Transformations



T is orthogonal, $\|\mathbf{u}\| = \|T(\mathbf{u})\|$, $\|\mathbf{v}\| = \|T(\mathbf{v})\|$, *angle* $\alpha =$ *angle* β and $d(P, Q) = d(R, S)$ where $d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + \cdots + (u_n - v_n)^2}$.

The Hessian of a matrix I

Take a vector-valued function $\mathbf{f} = (f_1, f_2, \ldots, f_n) : \mathbb{R}^m \to \mathbb{R}^n$. Then, we denote the $m \times n$ matrix of the first-order derivatives of $\mathbf{f}$ with respect to the elements of $\mathbf{x}$ ($\mathbf{x} \in \mathbb{R}^m$) as,

$$\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_n(\mathbf{x})}{\partial x_1} \\ \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_n(\mathbf{x})}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(\mathbf{x})}{\partial x_m} & \frac{\partial f_2(\mathbf{x})}{\partial x_m} & \cdots & \frac{\partial f_n(\mathbf{x})}{\partial x_m} \end{bmatrix}.$$

This becomes a column vector when $n = 1$ and is called the **gradient vector** of $f(\mathbf{x})$.

The $m \times n$ **Hessian matrix** of the second derivatives of the real-valued function $f(\mathbf{x})$ is

$$\nabla^2{}_{\mathbf{x}} f(\mathbf{x}) = \nabla_{\mathbf{x}}(\nabla_{\mathbf{x}} f(\mathbf{x})) \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_m} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_m \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_m \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_m \partial x_m} \end{bmatrix}.$$

Notice that a Hessian matrix is a square matrix.

This is a special type of matrix multiplication without being concerned about matrices' dimension restrictions.
More specifically, define $\mathbf{A} := (\alpha_{ij})$ and $\mathbf{B} := (\beta_{st})$. The Kronecker product transforms these two matrices into a matrix containing all products $\alpha_{ij}\beta_{st}$.

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \alpha_{11}\mathbf{B} & \dots & \alpha_{1n}\mathbf{B} \\ \vdots & & \vdots \\ \alpha_{m1}\mathbf{B} & \dots & \alpha_{mn}\mathbf{B} \end{bmatrix}.$$

The dimension of the above matrix is $mp \times nq$ for $m \times n$ matrix $\mathbf{A}$ and $p \times q$ matrix $\mathbf{B}$.

Let $\mathbf{A} = \begin{bmatrix} 2 & 5 & 2 \\ 0 & 6 & 3 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 2 & 4 & 1 \\ 3 & 5 & 0 \end{bmatrix}$. We have,

$$\mathbf{A}^{\mathsf{T}} \otimes \mathbf{B} = \begin{bmatrix} 2\mathbf{B} & \mathbf{0} \\ 5\mathbf{B} & 6\mathbf{B} \\ 2\mathbf{B} & 3\mathbf{B} \end{bmatrix} = \begin{bmatrix} 4 & 8 & 2 & 0 & 0 & 0 \\ 6 & 10 & 0 & 0 & 0 & 0 \\ 10 & 20 & 5 & 12 & 24 & 6 \\ 15 & 25 & 0 & 18 & 30 & 0 \\ 4 & 8 & 2 & 6 & 12 & 3 \\ 6 & 10 & 0 & 9 & 15 & 0 \end{bmatrix}.$$

The Kronecker product: Properties

The Kronecker product satisfies the following properties:

Kronecker product properties

1. $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}$.
2. $\mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C}$.
3. $a\mathbf{A} \otimes b\mathbf{B} = ab(\mathbf{A} \otimes \mathbf{B})$ for scalars a,b.
4. $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$.
5. $(\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T$.
6. $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$.

The Kronecker product: Properties

The Kronecker product is NOT commutative, i.e.

$$\mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A}.$$

As an example, consider

$$(1 \;\; 0) \otimes (0 \;\; 1) = (0 \;\; 1 \;\; 0 \;\; 0)$$

and

$$(0 \;\; 1) \otimes (1 \;\; 0) = (0 \;\; 0 \;\; 1 \;\; 0).$$

We give here the basic rules one needs to know.
For *d*-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$, we have,

$$\nabla_{\mathbf{y}}(\mathbf{x}^{\mathsf{T}}\mathbf{y}) = \mathbf{x}.$$

For a symmetric matrix $\mathbf{A}$ we have,

$$\nabla_{\mathbf{x}}(\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x} \ (\textit{gradient}), \qquad \nabla^2_{\mathbf{x}}(\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}) = 2\mathbf{A} \ (\textit{Hessian}).$$

Notice that the form $\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x}$ is called **quadratic form**.

Differentiation of Linear Forms

Results

(1) $\frac{\partial \mathbf{c}^\intercal \mathbf{x}}{\partial \mathbf{x}} = \mathbf{c}$ for $\mathbf{c}$ a conformable vector.

(2) $\frac{\partial \mathbf{q}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{q}^\intercal$ for $\mathbf{q}$ a conformable matrix.

(3) $\frac{\partial \mathbf{x}^\intercal \mathbf{q}}{\partial \mathbf{x}} = \mathbf{q}$ for $\mathbf{q}$ a conformable matrix or vector.

(4) $\frac{\partial \mathbf{x}}{\partial \mathbf{x}^\intercal} = \frac{\partial \mathbf{x}^\intercal}{\partial \mathbf{x}} = \mathbf{I}$.

Results

(1) $\frac{\partial \mathbf{x}^\intercal \mathbf{q} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{q}\mathbf{x}$ if $\mathbf{q}$ is a symmetric matrix.

(2) $\frac{\partial \mathbf{x}^\intercal \mathbf{q} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{q} + \mathbf{q}^\intercal)\mathbf{x}$.

(3) $\frac{\partial \mathbf{x}^\intercal \mathbf{q} \mathbf{x}}{\partial \mathbf{x} \partial \mathbf{x}^\intercal} = 2\mathbf{q}$ if $\mathbf{q}$ is a symmetric matrix.

(4) $\frac{\partial \mathbf{x}^\intercal \mathbf{q} \mathbf{x}}{\partial \mathbf{x}} = \mathbf{q} + \mathbf{q}^\intercal$.

Definition

For a square matrix $\mathbf{A}$, let $\mathbf{A}_{ij}$ denote the submatrix obtained from $\mathbf{A}$ by deleting its i^{th} row and j^{th} column. Then, the **determinant** of $\mathbf{A}$ is

$$\det(\mathbf{A}) = \sum_{i=1}^{m} (-1)^{i+j} \alpha_{ij} \det(\mathbf{A}_{ij}),$$

for any $j = 1, \ldots, n$ where $(-1)^{i+j} \det(\mathbf{A}_{ij})$ is called the **cofactor** of α_{ij}.

Determinant of a 2×2 matrix: $\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}$.

Determinant of a 3×3 matrix:

$\alpha_{11}\alpha_{22}\alpha_{33} - \alpha_{11}\alpha_{23}\alpha_{32} + \alpha_{12}\alpha_{23}\alpha_{31} - \alpha_{12}\alpha_{21}\alpha_{33} + \alpha_{13}\alpha_{21}\alpha_{32} - \alpha_{13}\alpha_{22}\alpha_{31}$.

Matrix Determinant: Example

Consider matrix $\mathbf{A} = \begin{bmatrix} 2 & 3 & 0 & 5 \\ 1 & 4 & 0 & 2 \\ 5 & 4 & 8 & 5 \\ 2 & 1 & 0 & 5 \end{bmatrix}$. It is convenient to use cofactor expansion by column 3 since then,

$$det(\mathbf{A}) = \alpha_{13}C_{13} + \alpha_{23}C_{23} + \alpha_{33}C_{33} + \alpha_{43}C_{43} = 8C_{33}$$

$$= (-1)^{3+3} det \begin{bmatrix} 2 & 3 & 5 \\ 1 & 4 & 2 \\ 2 & 1 & 5 \end{bmatrix} = -16$$

by using the determinant of a 3×3 matrix formula in the previous frame.

Matrix Determinant and Trace: Properties I

It is clear that $\det(\mathbf{A}) = \det(\mathbf{A}^\mathsf{T})$.
Also, $\det(c\mathbf{A}) = c^n \det(\mathbf{A})$ for scalar c and $n \times n$ matrix $\mathbf{A}$.
Some more properties:

1. $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}) = \det(\mathbf{BA})$.
2. $\det(\mathbf{A} \otimes \mathbf{B}) = \det(\mathbf{A})^m \det(\mathbf{B})^p$ for $m \times m$ matrix $\mathbf{A}$ and $p \times p$ matrix $\mathbf{B}$.
3. For an orthogonal matrix $\mathbf{A}$ we have $\det(\mathbf{I}) = \det(\mathbf{A}\mathbf{A}^\mathsf{T}) = [\det(\mathbf{A})]^2$. Such a determinant is either 1 or -1 since the determinant of the identity matrix is always 1.

Definition

The **trace** of a square matrix is the sum of its diagonal elements i.e. $trace(\mathbf{A}) = \sum_i \alpha_{ii}$.

Obviously $trace(\mathbf{I}_n) = n$ and $trace(\mathbf{A}) = trace(\mathbf{A}^T)$.

We list some more properties of the trace of a matrix.

1. $trace(c\mathbf{A} + d\mathbf{B}) = c\ trace(\mathbf{A}) + d\ trace(\mathbf{B})$ for scalars c,d.
2. $trace(\mathbf{AB}) = trace(\mathbf{BA})$.
3. $trace(\mathbf{A} \otimes \mathbf{B}) = trace(\mathbf{A})trace(\mathbf{B})$.

Matrix Inverse

Definition

A matrix is called **nonsingular** if it has a non-sero determinant. A nonsingular matrix $\mathbf{A}$ possesses a unique inverse $\mathbf{A}^{-1}$, calculated as,

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} adj(\mathbf{A}),$$

where the adjoint matrix $adj(\mathbf{A})$ is the transpose of the matrix of cofactors.

For a 2×2 matrix $\mathbf{A}$ we have the simple formula,

$$\mathbf{A}^{-1} = \frac{1}{\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}} \begin{bmatrix} \alpha_{22} & -\alpha_{12} \\ -\alpha_{21} & \alpha_{11} \end{bmatrix}.$$

More general, $adj(\mathbf{A}) = \begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix}$, obtained from matrix $\mathbf{A}$ by replacing each entry by its cofactor and then by transposing it.

Example: Consider $\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 3 \end{bmatrix}$.

Then, the adjoint of A can be written as,

$$adj(A) = \begin{pmatrix} det\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} & -det\begin{pmatrix} -1 & 0 \\ 0 & 3 \end{pmatrix} & det\begin{pmatrix} -1 & 0 \\ 1 & 2 \end{pmatrix} \\ -det\begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix} & det\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} & -det\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \\ det\begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} & -det\begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix} & det\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \end{pmatrix}$$

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

Matrix Inverse: Properties

1. $(A^{-1})^{-1} = A$
2. $(cA^{-1}) = \frac{1}{c}A^{-1}$
3. $(AB)^{-1} = B^{-1}A^{-1}$
4. $(A^n)^{-1} = (A^{-1})^n$
5. $(A^\intercal)^{-1} = (A^{-1})^\intercal$

Matrix Span

Definition

The space spanned by the column vectors of a matrix $\mathbf{A}$ is denoted as *span*$(\mathbf{A})$ and is known as the **column space** of $\mathbf{A}$. The **row space** of $\mathbf{A}$ is defined equivalently as *span*$(\mathbf{A}^\mathsf{T})$.

We denote the span of a set of matrices S by $\langle S \rangle$.

Matrix Span: Example

Suppose our space is that of all 2×3 matrices denoted as $M_{2,3}$. Can we determine the **span** of $\langle S \rangle$ when

$$S = \left\{ \begin{bmatrix} 2 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -2 & 2 \\ 0 & 0 & 0 \end{bmatrix} \right\}?$$

Work as follows:

$$\langle S \rangle = \left\{ a \begin{bmatrix} 2 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + b \begin{bmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & -2 & 2 \\ 0 & 0 & 0 \end{bmatrix} : a, b, c \in \mathbb{R} \right\} =$$

$$\left\{ \begin{bmatrix} 2a + b & -a - 2c & 3b + 2c \\ 0 & 0 & 0 \end{bmatrix} : a, b, c \in \mathbb{R} \right\}.$$

We just considered all linear combinations of the above matrices.

Matrix Span: Example

It is obvious that we have $\langle S \rangle \subseteq \left\{ \begin{bmatrix} \alpha & \beta & \gamma \\ 0 & 0 & 0 \end{bmatrix} : \alpha, \beta, \gamma \in \mathbb{R} \right\}$ and in fact, every 2×3 matrix with zero entries in the second row belongs to $\langle S \rangle$. To show this, we write

$$\begin{bmatrix} \alpha & \beta & \gamma \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2a + b & -a - 2c & 3b + 2c \\ 0 & 0 & 0 \end{bmatrix}$$

and by solving we get $a = \frac{1}{7}(3\alpha - \beta - \gamma)$, $b = \frac{1}{7}(\alpha + 2\beta + 2\gamma)$ and $c = \frac{-1}{14}(3\alpha + 6\beta - \gamma)$.

Therefore $\langle S \rangle = \left\{ \begin{bmatrix} \alpha & \beta & \gamma \\ 0 & 0 & 0 \end{bmatrix} : \alpha, \beta, \gamma \in \mathbb{R} \right\}$.

Matrix Rank

Definition

The column/row **rank** of a matrix $\mathbf{A}$ is the maximum number of linearly independent column/row vectors of $\mathbf{A}$.

Definition

If the column/row rank equals the number of column/row vectors, we say that $\mathbf{A}$ has **full column/row rank**.

It can be shown that the column rank and the row rank of a matrix $\mathbf{A}$ are equal (Lemma 1.3 in the lecture notes).
It follows that the rank of $\mathbf{A}$ is defined as the maximum number of linearly independent column or row vectors of $\mathbf{A}$ and therefore

$$rank(\mathbf{A}) = rank(\mathbf{A}^{\mathsf{T}}).$$

Consider matrix $\mathbf{A} = \begin{bmatrix} 1 & 5 & 6 \\ 2 & 6 & 8 \\ 7 & 1 & 8 \end{bmatrix}$. We want to find its rank. The first two rows of $\mathbf{A}$ are linearly independent hence $rank(\mathbf{A}) \geq 2$. Now, $rank(\mathbf{A}) < 3$ because for vector $\mathbf{u} = (20, -17, 2)^\mathsf{T}$ we have $\mathbf{A}^\mathsf{T}\mathbf{u} = \mathbf{0}$. Also for vector $\mathbf{v} = (1, 1, -1)^\mathsf{T}$ we also have $\mathbf{A}\mathbf{v} = \mathbf{0}$ which implies that if the row rank of $\mathbf{A}$ is 2, the column rank of $\mathbf{A}$ is also 2.

We also know that $rank(\mathbf{A}) = rank(\mathbf{A}^\mathsf{T})$ and we can verify that we also have

$$rank(\mathbf{A}\mathbf{A}^\mathsf{T}) = rank(\mathbf{A}^\mathsf{T}\mathbf{A}) = rank(\mathbf{A}) = rank(\mathbf{A}^\mathsf{T}).$$

Finally, we cannot construct a 3×4 matrix of rank 4 because there are only three rows and these cannot generate a four-dimensional space.

Matrix Rank: Properties

For two $n \times k$ matrices $\mathbf{A}$ and $\mathbf{B}$, the following relations hold:

1. $rank(\mathbf{A} + \mathbf{B}) \leq rank(\mathbf{A}) + rank(\mathbf{B})$.
2. $rank(\mathbf{A} \otimes \mathbf{B}) = rank(\mathbf{A}) rank(\mathbf{B})$.

For $n \times k$ and $k \times m$ matrices $\mathbf{A}$ and $\mathbf{B}$ respectively we have,

$$rank(\mathbf{A}) + rank(\mathbf{B}) - k \leq rank(\mathbf{AB}) \leq \min[rank(\mathbf{A}), rank(\mathbf{B})].$$

From this it follows that,

$$rank(\mathbf{AB}) \leq rank(\mathbf{B}) = rank(\mathbf{A}^{-1}\mathbf{AB}) \leq rank(\mathbf{AB}),$$

$$rank(\mathbf{BC}) = rank(\mathbf{C}^{\mathsf{T}}\mathbf{B}^{\mathsf{T}}) = rank(\mathbf{B}^{\mathsf{T}}) = rank(\mathbf{B}),$$

i.e. the rank of a matrix is preserved under nonsingular transformations (note that $\mathbf{A}$ and $\mathbf{C}$ are nonsingular).

Definition

Let $\mathbf{A}$ be any square matrix. A non-zero vector $\mathbf{v}$ is an **eigenvector** of $\mathbf{A}$ if $\mathbf{Av} = \lambda\mathbf{v}$ for some number λ. The number λ is the corresponding **eigenvalue**.

The system of equations $\mathbf{Av} = \lambda\mathbf{v}$ has a non-trivial solution if and only if

$$det(\mathbf{A} - \lambda\mathbf{I}) = 0.$$

This is the **characteristic equation** of $\mathbf{A}$.
Intuitively, we can consider the matrix $\mathbf{A}$ as a transformation of a vector $\mathbf{v}$ i.e we have the transformation $\mathbf{Av}$. Now, non-zero vectors that are transformed to scalar multiples of themselves i.e. $\mathbf{Av} = \lambda\mathbf{v}$ are called eigenvectors and the corresponding scalar multipliers, λ in our case, are called eigenvalues.

Characteristic equation: $p_{\mathbf{A}}(\lambda) = det(\mathbf{A} - \lambda\mathbf{I}) = 0$.
The function $p_{\mathbf{A}}(\lambda)$ is called the **characteristic polynomial** of $\mathbf{A}$.
From the *Fundamental Theorem of Algebra*, we know that the n^{th}-degree equation $p_{\mathbf{A}}(\lambda) = 0$ has n roots which may be real or complex, multiple or unique. We can write the factorization,

$$p_{\mathbf{A}}(\lambda) = p_0 + p_1\lambda + \cdots + p_{n-1}\lambda^{n-1} + p_n\lambda^n = (\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n).$$

The numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ denote the eigenvalues of $\mathbf{A}$.

Eigenvalues and Eigenvectors: Examples I

• Example 1: Take matrix $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ -1 & 0 \end{bmatrix}$. We need to compute the determinant of $\begin{bmatrix} \lambda - 2 & 1 \\ 1 & \lambda \end{bmatrix}$ which gives the characteristic polynomial $\lambda^2 - 2\lambda + 1$. Solving this gives us the eigenvalues of $\mathbf{A}$ and having those, we can compute the eigenvectors of $\mathbf{A}$.

• Example 2: Suppose two matrices $\mathbf{A}$ and $\mathbf{B}$ have the same characteristic polynomial. Is it necessarily true that $\mathbf{A} = \mathbf{B}$? The answer is NO. Take $\mathbf{A} = \begin{bmatrix} 1 & \alpha \\ 0 & 2 \end{bmatrix}$. Then $p_{\mathbf{A}}(\lambda) = (\lambda - 1)(\lambda - 2)$ for every α. Therefore, although the eigenvalues is a good way of characterizing a matrix, they do NOT characterize a matrix completely.

• Example 3: Let's find the eigenvectors of $\mathbf{A} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$. The characteristic equation is $(a - \lambda)^2 = 0$ for which $\lambda = a$ is a repeated root. Therefore every vector $\mathbf{x} \neq \mathbf{0}$ satisfies

$$\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} ax_1 \\ ax_2 \end{bmatrix},$$

and we can write the complete set of eigenvectors as,

$$\mathbf{x} := k \begin{bmatrix} 1 \\ 0 \end{bmatrix} + l \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

with $(k, l) \neq (0, 0)$.
We have two eigenvectors associated with the multiple eigenvalue.

Eigenvalues and Eigenvectors: Examples II

Why we chose the above eigenvectors? Recall that the eigenvector equations are

$$(a-\lambda)x_1 = 0 \ \text{ and } \ (a-\lambda)x_2 = 0.$$

The eigenvector equations become $0 = 0$ and $0 = 0$ which are satisfied for all values of x_1 and x_2. So the eigenvectors are all non-zero vectors of the form $(k, l)^\mathsf{T}$.
BUT since any non-zero vector is an eigenvector, it is possible to choose two eigenvectors that are linearly independent, for example $(1, 0)^\mathsf{T}$ and $(0, 1)^\mathsf{T}$ as above.

What the last example is telling us is that eigenvectors are NOT unique.
We can have two distinct vectors associated with the same eigenvalue.
As another example, take the identity matrix $\mathbf{I}_n$. This matrix has 1 as its only eigenvalue hence every $n \times 1$ vector $\mathbf{x} \neq \mathbf{0}$ is an eigenvector.
BUT an eigenvector cannot be associated with two distinct eigenvalues.
For suppose $\lambda_1 \neq \lambda_2$ are eigenvalues of a matrix $\mathbf{A}$ and that $\mathbf{x} \neq \mathbf{0}$ satisfies $\mathbf{Ax} = \lambda_1 \mathbf{x}$ and $\mathbf{Ax} = \lambda_2 \mathbf{x}$. Then $\lambda_1 \mathbf{x} = \lambda_2 \mathbf{x}$ and therefore $\mathbf{x} = \mathbf{0}$. But $\mathbf{x} = \mathbf{0}$ is not permitted as an eigenvector, so we have a contradiction.

If a matrix $\mathbf{A}$ has n distinct eigenvalues, then it also has n distinct eigenvectors unique up to scalar multiplication. Such eigenvectors also have the property that they are linearly independent. This is stated as a theorem.

Theorem

If $\mathbf{p_1}, \ldots, \mathbf{p_n}$ are eigenvectors of $\mathbf{A}$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then $\{\mathbf{p_1}, \ldots, \mathbf{p_n}\}$ is a linearly independent set.

Let $\mathbf{V}$ denote the matrix of all the eigenvectors and Λ the diagonal matrix with the eigenvalues as their diagonal elements. We can write, $\mathbf{AV} = \mathbf{V}\Lambda$. Since $\mathbf{V}$ is nonsingular, we can write,

$$\mathbf{V}^{-1}\mathbf{AV} = \Lambda \;\; \textit{or} \;\; \mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}.$$

We say that $\mathbf{A}$ is **similar** to Λ.

The above can be written as a theorem.

Theorem

Let $\mathbf{A}$ be an $n \times n$ matrix with distinct eigenvalues. Then there exists a nonsingular $n \times n$ matrix $\mathbf{V}$ and a diagonal $n \times n$ matrix Λ whose diagonal elements are the eigenvalues of $\mathbf{A}$ such that $\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \Lambda$.

Proof: For distinct eigevalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$, let $\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$, $i = 1, \ldots, n$. Let $\mathbf{V} := \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$. Then

$$\mathbf{A}\mathbf{V} = (\mathbf{A}\mathbf{x}_1, \ldots, \mathbf{A}\mathbf{x}_n) = (\lambda_1 \mathbf{x}_1, \ldots, \lambda_n \mathbf{x}_n) = \mathbf{V}\Lambda,$$

where $\mathbf{V} := diag(\lambda_1, \ldots, \lambda_n)$. The eigenvectors $\mathbf{x}_i$ are linearly independent because the eigenvalues are distinct by the previous theorem. Therefore $\mathbf{V}$ is nonsingular and we get $\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \Lambda$. This concludes the proof.

More properties:

When $\mathbf{A}$ has n distinct eigenvalues, we have,

$$det(\mathbf{A}) = det(\mathbf{V}\Lambda\mathbf{V}^{-1}) = det(\Lambda)det(\mathbf{V})det(\mathbf{V}^{-\mathbf{1}}) = det(\Lambda),$$

$$trace(\mathbf{A}) = trace(\mathbf{V}\Lambda\mathbf{V}^{-1}) = trace(\mathbf{V}^{-1}\mathbf{V}\Lambda) = trace(\Lambda).$$

Therefore we get the following important result:

Lemma

When $\mathbf{A}$ has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, we have

$$det(\mathbf{A}) = det(\Lambda) = \prod_{i=1}^{n} \lambda_i \;\; and \;\; trace(\mathbf{A}) = trace(\Lambda) = \sum_{i=1}^{n} \lambda_i.$$

• Note:
$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1} \Rightarrow \mathbf{A}^{-1} = \mathbf{V}\Lambda^{-1}\mathbf{V}^{-1}$: The eigenvectors of $\mathbf{A}^{-1}$ are the same as those of $\mathbf{A}$ and the corresponding eigenvalues are the reciprocals of the eigenvalues of $\mathbf{A}$.

• Note:
$\mathbf{A}^2 = (\mathbf{V}\Lambda\mathbf{V}^{-1})(\mathbf{V}\Lambda\mathbf{V}^{-1}) = \mathbf{V}\Lambda^2\mathbf{V}^{-1}$: The eigenvectors of $\mathbf{A}^2$ are the same as those of $\mathbf{A}$ and the corresponding eigenvalues are the squares of the eigenvalues of $\mathbf{A}$.

Orthogonal eigenvectors with distinct eigenvalues

When our matrix $\mathbf{A}$ is symmetric, we can show that the eigenvectors associated with distinct eigenvalues are orthogonal to each other (not only linearly independent as we have seen before).
Take distinct eigenvalues λ_i, λ_j and corresponding eigenvectors $\mathbf{v_i}$ and $\mathbf{v_j}$ of matrix $\mathbf{A}$. We have,

$$\mathbf{A}\mathbf{v_i} = \lambda_i \mathbf{v_i} \;\; \textit{and} \;\; \mathbf{A}\mathbf{v_j} = \lambda_j \mathbf{v_j}.$$

Because $\mathbf{A}$ is symmetric, i.e. $\mathbf{A} = \mathbf{A}^\mathsf{T}$ we have,

$$\lambda_i \mathbf{v_j}^\mathsf{T} \mathbf{v_i} = \mathbf{v_j}^\mathsf{T} \mathbf{A} \mathbf{v_i} = \mathbf{v_i}^\mathsf{T} \mathbf{A}^\mathsf{T} \mathbf{v_j} = \mathbf{v_i}^\mathsf{T} \mathbf{A} \mathbf{v_j} = \lambda_j \mathbf{v_i}^\mathsf{T} \mathbf{v_j}.$$

But this says that $\mathbf{v_i}^\mathsf{T} \mathbf{v_j} = 0$ since $\lambda_i \neq \lambda_j$ (*orthogonal eigenvectors*).

From the above we conclude that a symmetric matrix $\mathbf{A}$ is *orthogonally diagonalizable* since,

$$\mathbf{V}^\intercal \mathbf{A} \mathbf{V} = \Lambda \quad \textit{or} \quad \mathbf{A} = \mathbf{V} \Lambda \mathbf{V}^\intercal.$$

Here $\mathbf{V}$ is the orthogonal matrix of associated eigenvectors.

Theorem

Let A be a square matrix. A is orthogonally diagonalizable if and only if it is a symmetric matrix.

Take $\mathbf{A} = \mathbf{V} \Lambda \mathbf{V}^\intercal$ with diagonal matrix Λ. Then

$$\mathbf{A}^\intercal = (\mathbf{V} \Lambda \mathbf{V}^\intercal)^\intercal = \mathbf{V} \Lambda \mathbf{V}^\intercal = \mathbf{A}.$$

This completes the proof.

Some more results for a symmetric matrix $\mathbf{A}$:

1. $rank(\mathbf{A}) = rank(\Lambda)$ which equals the number of non-zero eigenvalues of $\mathbf{A}$.
2. $det(\mathbf{A}) = det(\Lambda) = \prod_{i=1}^{n} \lambda_i$.
3. $trace(\mathbf{A}) = trace(\Lambda) = \sum_{i=1}^{n} \lambda_i$.

Therefore a symmetric matrix is non-singular if its eigenvalues are all non-zero.

Also it can be proved that the eigenvalues of a symmetric matrix are real (for proof notice that the discriminant of the characteristic equation cannot be negative).

Definition

A symmetric matrix $\mathbf{A}$ is said to be **positive definite** if $\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} > 0$ for all vectors $\mathbf{x} \neq 0$.

If we also have equality in the above condition, we talk about $\mathbf{A}$ being a **positive semi-definite** matrix.
Reversing the above inequalities, we get **negative definite** and **negative semi-definite** matrix $\mathbf{A}$.

Symmetric Matrices V

Since we have seen that we can diagonalize a matrix $\mathbf{A}$ as $\mathbf{V}^{\mathsf{T}}\mathbf{A}\mathbf{V} = \Lambda$, we have that for any $\mathbf{x} \neq 0$

$$\mathbf{x}^{\mathsf{T}}\Lambda\mathbf{x} = \mathbf{x}^{\mathsf{T}}(\mathbf{V}^{\mathsf{T}}\mathbf{A}\mathbf{V})\mathbf{x} = \tilde{\mathbf{x}}^{\mathsf{T}}\mathbf{A}\tilde{\mathbf{x}} \geq 0,$$

where $\tilde{\mathbf{x}} = \mathbf{V}\mathbf{x}$. Therefore, Λ is also positive semi-definite and its diagonal elements are non-negative. This is stated as a lemma.

Lemma

A symmetric matrix is positive definite (positive semi-definite) if and only if its eigenvalues are all positive (non-negative).

Idempotent matrices: $\mathbf{A}^2 = \mathbf{A}$.
Take a vector $\mathbf{x}$ in the Euclidean space $\mathbf{X}$. The projection of $\mathbf{x}$ **onto** (i.e. range = codomain, $f(\mathbf{X}) = \mathbf{S}$) a subspace $\mathbf{S}$ of $\mathbf{X}$, is a linear transformation of $\mathbf{x}$ to $\mathbf{S}$. The resulting projected vector is written as $\mathbf{Px}$ where $\mathbf{P}$ is the projection/transformation matrix.
Since a further projection of $\mathbf{x}$ onto $\mathbf{S}$ should have no effect on $\mathbf{Px}$, we deduce that the projection/transformation matrix $\mathbf{P}$ must be idempotent.

$$\mathbf{P}(\mathbf{Px}) = \mathbf{P}^2\mathbf{x} = \mathbf{Px}.$$

Orthogonal Projection

A projection of $\mathbf{x}$ onto $\mathbf{S}$ is orthogonal if the projection $\mathbf{Px}$ is orthogonal to the difference between $\mathbf{x}$ and $\mathbf{Px}$, i.e.

$$(\mathbf{x} - \mathbf{Px})^{\mathsf{T}}\mathbf{Px} = \mathbf{x}^{\mathsf{T}}(\mathbf{I} - \mathbf{P})^{\mathsf{T}}\mathbf{Px} = \mathbf{0}.$$

Since we have equality iff $(\mathbf{I} - \mathbf{P})^{\mathsf{T}}\mathbf{P} = 0$, we deduce that $\mathbf{P}$ must be symmetric i.e. $\mathbf{P} = \mathbf{P}^{\mathsf{T}}\mathbf{P}$ and $\mathbf{P}^{\mathsf{T}} = \mathbf{P}^{\mathsf{T}}\mathbf{P}$.
Note that the orthogonal projection must be unique.

Notice that $\mathbf{I} - \mathbf{P}$ is also idempotent and symmetric therefore it is an orthogonal projection. By symmetry we know that $(\mathbf{I} - \mathbf{P})\mathbf{P} = \mathbf{0}$ and therefore the projections $\mathbf{P}\mathbf{x}$ and $(\mathbf{I} - \mathbf{P})\mathbf{x}$ must be orthogonal.
We deduce that any vector $\mathbf{x}$ can be uniquely decomposed into two orthogonal components:

$$\mathbf{x} = \mathbf{P}\mathbf{x} + (\mathbf{I} - \mathbf{P})\mathbf{x}.$$

Notice that $(\mathbf{I} - \mathbf{P})\mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto $\mathbf{S}^{\perp}$, where $\mathbf{S}^{\perp}$ is the orthogonal complement of a subspace $\mathbf{S}$ i.e.

$$\mathbf{S}^{\perp} = \{\mathbf{x} \in \mathbf{X} : \mathbf{x}^{\mathsf{T}}\mathbf{s} = \mathbf{0},\ \textit{all}\ \mathbf{s} \in \mathbf{S}\}$$

Intuitively, the orthogonal projection $\mathbf{Px}$ can be interpreted as the "best approximation" of $\mathbf{x}$ in $\mathbf{S}$ in the sense that $\mathbf{Px}$ is the closest to $\mathbf{x}$ in terms of the Euclidean norm. We have for any $\mathbf{s} \in \mathbf{S}$,

$$\| \mathbf{x} - \mathbf{s} \|^2 = \| \mathbf{x} - \mathbf{Px} + \mathbf{Px} - \mathbf{s} \|^2$$

$$= \| \mathbf{x} - \mathbf{Px} \|^2 + \| \mathbf{Px} - \mathbf{s} \|^2 + 2(\mathbf{x} - \mathbf{Px})^\mathsf{T} (\mathbf{Px} - \mathbf{s})$$

$$= \| \mathbf{x} - \mathbf{Px} \|^2 + \| \mathbf{Px} - \mathbf{s} \|^2 .$$

We therefore get the following lemma.

Lemma

Let $\mathbf{x}$ *be a vector in* $\mathbf{X}$ *and* $\mathbf{Px}$ *its orthogonal projection onto* $\mathbf{S} \subseteq \mathbf{X}$*. Then, for any* $\mathbf{s} \in \mathbf{S}$*,*

$$\| \mathbf{x} - \mathbf{Px} \| \leq \| \mathbf{x} - \mathbf{s} \| .$$

Orthogonal Projection: More results

As we have seen before, there exists an orthogonal matrix $\mathbf{V}$ that diagonalizes a symmetric and idempotent matrix $\mathbf{A}$ to Λ. Now we can get,

$$\Lambda = \mathbf{V}^\mathsf{T}\mathbf{A}\mathbf{V} = \mathbf{V}^\mathsf{T}\mathbf{A}(\mathbf{V}\mathbf{V}^\mathsf{T})\mathbf{A}\mathbf{V} = \Lambda^2,$$

which is possible iff the eigenvalues of $\mathbf{A}$ are zero and one. This gives the following lemma:

Lemma

A symmetric and idempotent matrix is positive semi-definite with eigenvalues 0 and 1.

Orthogonal Projection: More results

Since the trace of Λ is the number of non-zero eigenvalues we have that $trace(\Lambda) = rank(\Lambda)$.
We have also seen before that $rank(\mathbf{A}) = rank(\Lambda)$ and $trace(\mathbf{A}) = trace(\Lambda)$. This gives us the following lemma:

Lemma

For a symmetric and idempotent matrix $\mathbf{A}$, $rank(\mathbf{A}) = trace(\mathbf{A})$, the number of non-zero eigenvalues of $\mathbf{A}$.

It can be shown (see lecture notes) that the orthogonal complement of the row space of $\mathbf{A}$ is the same as the orthogonal complement of the row space of $\mathbf{A}^\mathsf{T}\mathbf{A}$ and that the column space of $\mathbf{A}$ is the same as the column space of $\mathbf{A}\mathbf{A}^\mathsf{T}$. This gives the following property:

$$rank(\mathbf{A}) = rank(\mathbf{A}^\mathsf{T}\mathbf{A}) = rank(\mathbf{A}\mathbf{A}^\mathsf{T}).$$

Also if $\mathbf{A}$ of dimension $n \times k$ is of full column rank $k < n$, then $\mathbf{A}^\mathsf{T}\mathbf{A}$ is $k \times k$ and hence of full rank while $\mathbf{A}\mathbf{A}^\mathsf{T}$ is $n \times n$ and singular, therefore,

Lemma

If $\mathbf{A}$ of dimension $n \times k$ is a matrix of full column rank $k < n$, then $\mathbf{A}^\mathsf{T}\mathbf{A}$ is symmetric and positive definite.

Orthogonal Projection: More results

For an $n \times k$ matrix $\mathbf{A}$ with full column rank $k < n$, the matrix $\mathbf{P} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$ is symmetric and idempotent therefore is an orthogonal projection matrix.
Also,

$$trace(\mathbf{P}) = trace(\mathbf{A}^\top\mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}) = trace(\mathbf{I_k}) = k,$$

and we deduce from the previous lemmas (1.11 and 1.12 in lecture notes) that also $rank(\mathbf{P}) = k$, the k eigenvalues that equal one. Similarly, $rank(\mathbf{I} - \mathbf{P}) = n - k$.

Orthogonal Projection: More results

For any vector $\mathbf{x} \in \mathit{span}(\mathbf{A})$ we can write $\mathbf{x}$ as $\mathbf{Ab}$ for any non-zero vector $\mathbf{b}$ such that,

$$\mathbf{Px} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top(\mathbf{Ab}) = \mathbf{Ab} = \mathbf{x},$$

which implies that $\mathbf{P}$ projects vectors onto $\mathit{span}(\mathbf{A})$.
Equivalently for $\mathbf{x} \in \mathit{span}(\mathbf{A})^\perp$, we can see that $\mathbf{I} - \mathbf{P}$ must project vectors onto $\mathit{span}(\mathbf{A})^\perp$. Therefore we have the following lemma.

Lemma

Let $\mathbf{A}$ be an $n \times k$ matrix with full column rank k. Then, $\mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$ orthogonally projects vectors onto $\mathit{span}(\mathbf{A})$ and has rank k.
On the other hand, $I_n - \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$ orthogonally projects vectors onto $\mathit{span}(\mathbf{A})^\perp$ and has rank $n - k$.

Further References

1. Abadir, K. M. and J. R. Magnus, “Matrix Algebra”, 2005, Econometric Exercises I, Cambridge University Press.

2. Kuan, C-M., “Elements of Matrix Algebra”, 2009, Lecture Notes, NTU, Dept of Finance.

3. Lang, S., “Linear Algebra”, 2004, 2^{nd} edition, Undergraduate Texts in Mathematics, Springer.

4. Magnus, J. R. and H. Neudecker, “Matrix Differential Calculus with Applications in Statistics and Econometrics”, 1999, 2^{nd} Edition, Wiley series in probability and statistics.

5. Strang, G., “Linear Algebra and Its Applications”, 1988, 3^{rd} Edition, Brooks Cole.