Matrix Algebra

Learner

LaTeX by Xin Wang

Notes

Contents

Chapter 1

Matrices

1.1 Definition of a matrix

Definition 1.1.1: m-by-n matrix

$$\mathbf{A}_{m \times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Here, the matrix element of A in the i-th row and the j-th column is denoted as a_{ij}.

Definition 1.1.2: Column vector

The column vector is n-by-one.

$$\mathbf{q}_{n \times 1} = \begin{pmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{pmatrix}$$

Definition 1.1.3: Row vector

The row vector is one-by-n.

$$\mathbf{p}_{1 \times n} = \begin{pmatrix} p_1 & p_2 & \cdots & p_n \end{pmatrix}$$

1.2 Addition and multiplication of matrices

1.2.1 Addition of matrices

Corollary 1.2.1

$$\mathbf{A}_{m \times n} + \mathbf{B}_{m \times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix} \tag{1.1}$$

$$= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix} \tag{1.2}$$

Corollary 1.2.2

$$k\mathbf{A}_{m \times n} = k \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{1.3}$$

$$= \begin{pmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{pmatrix} \tag{1.4}$$

1.2.2 Multiplication of matrices

Corollary 1.2.3

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

$$c_{ij} = \sum_{i=1}^{n} a_{ik} b_{kj}$$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} a_{1i} b_{i1} & \sum_{i=1}^{n} a_{1i} b_{i2} & \cdots & \sum_{i=1}^{n} a_{1i} b_{ip} \\ \sum_{i=1}^{n} a_{2i} b_{i1} & \sum_{i=1}^{n} a_{2i} b_{i2} & \cdots & \sum_{i=1}^{n} a_{2i} b_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} a_{mi} b_{i1} & \sum_{i=1}^{n} a_{mi} b_{i2} & \cdots & \sum_{i=1}^{n} a_{mi} b_{ip} \end{pmatrix}$$

1.2.3   The associative law for matrix multiplication

Corollary 1.2.4

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

Prove:

$$[\mathbf{A}_{m\times n}(\mathbf{B}_{n\times p}\mathbf{C}_{p\times q}))]_{ij} = \sum_{k=1}^{n} a_{ik}[\mathbf{B}_{n\times p}\mathbf{C}_{p\times q}]_{kj} \tag{1.5}$$

$$= \sum_{k=1}^{n} a_{ik} \sum_{l=1}^{p} b_{kl}c_{lj} \tag{1.6}$$

$$= \sum_{k=1}^{n}\sum_{l=1}^{p} a_{ik}b_{kl}c_{lj} \tag{1.7}$$

$$= \sum_{l=1}^{p}(\sum_{k=1}^{n} a_{ik}b_{kl})c_{lj} \tag{1.8}$$

$$= \sum_{l=1}^{p}[\mathbf{A}_{m\times n}\mathbf{B}_{n\times p}]_{il}c_{lj} \tag{1.9}$$

$$= [(\mathbf{A}_{m\times n}\mathbf{B}_{n\times p})\mathbf{C}_{p\times q}]_{ij} \tag{1.10}$$

1.3   Special matrices

1.3.1   Square matrix

Definition 1.3.1: Square matrix

number of rows equals number of columns.

1.3.2   Zero matrix

Definition 1.3.2: Zero matrix

Zero matrix can be **any size**.

$$\mathbf{0} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

Corollary 1.3.3

$$\mathbf{0}_{p\times m}\mathbf{A}_{m\times n} = \mathbf{0}_{p\times n}$$

$$\mathbf{A}_{m\times n}\mathbf{0}_{n\times p} = \mathbf{0}_{m\times p}$$

1.3.3 Identity matrix

Definition 1.3.4: Identity matrix

Identity matrix is a **square matrix**.

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

Corollary 1.3.5

If **A** and I are the **same sized square matrices**,

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

1.3.4 Diagonal matrix

Definition 1.3.6: Identity matrix

A diagonal matrix has its only nonzero elements on the diagonal.
Usually it is a **square matrix**, but it can also be **rectangular**.

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$$

Corollary 1.3.7

If **A** is a m-by-p diagonal matrix, **B** is a p-by-n diagonal matrix, then

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

is also a m-by-n diagonal matrix.

Proof. The ij element of C is given by

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$$

Since **A** is a diagonal matrix, the only nonzero term in the sum is k = i, and we have $c_{ij} = a_{ii} b_{ij}$. And since **B** is a diagonal matrix, the only nonzero elements of **C** are the diagonal elements $c_{ii} = a_{ii} b_{ii}$. □

1.3.5   Band matrix

Definition 1.3.8: Band matrix

A band (or banded) matrix has nonzero elements only on diagonal bands.

Example.

Tridiagonal matrix, three-by-three band matrix with nonzero diagonals one above and one below a nonzero main diagonal

$$\mathbf{D} = \begin{pmatrix} d_1 & a_1 & \cdots & 0 & 0 \\ b_1 & d_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & d_{n-1} & a_{n-1} \\ 0 & 0 & \cdots & b_{n-1} & d_n \end{pmatrix}$$

1.3.6   Triangular matrix

Definition 1.3.9: Triangular matrix

An upper or lower triangular matrix is a **square matrix** that has zero elements below or above the diagonal.

$$\mathbf{U} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1(n-1)} & a_{1n} \\ 0 & a_{22} & \cdots & a_{2(n-1)} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{(n-1)(n-1)} & a_{(n-1)n} \\ 0 & 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 & 0 \\ a_{21} & a_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{(n-1)1} & a_{(n-1)2} & \cdots & a_{(n-1)(n-1)} & 0 \\ a_{n1} & a_{n2} & \cdots & a_{n(n-1)} & a_{nn} \end{pmatrix}$$

Corollary 1.3.10

If **A** an **B** are both n-by-n upper triangular matrices, then

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

is also a n-by-n upper triangular matrix, with the diagonal elements of the product given by the product of the diagonal elements.

Proof. The ij element of **C** is given by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Since $\mathbf{A}$ and $\mathbf{B}$ are upper triangular, we have $a_{ik} = 0$ when k < i, and $b_{kj} = 0$ when k > j. Excluding the zero terms from the summation, we have

$$c_{ij} = \sum_{k=i}^{j} a_{ik} b_{kj}$$

which is equal to zero when i > j proving that $\mathbf{C}$ is upper triangular. Furthermore, $c_{ii} = a_{ii} b_{ii}$. $\square$

1.4 Transpose matrix

1.4.1 Transpose matrix

Definition 1.4.1: Transpose matrix

The transpose of a matrix $\mathbf{A}$, denoted by $\mathbf{A^T}$ and spoken as $\mathbf{A}$-transpose, switches the rows and columns of $\mathbf{A}$.

$$if\ \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},\ then\ \mathbf{A^T} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}$$

In other words, we write

$$a_{ij}^{\mathbf{T}} = a_{ji}$$

Evidently, if $\mathbf{A}$ is m-by-n then $\mathbf{A^T}$ is n-by-m.

Corollary 1.4.2

$$((\mathbf{A}_{m \times n})^{\mathbf{T}})^{\mathbf{T}} = \mathbf{A}_{m \times n}$$

Proof. Since $a_{ij}^{\mathbf{T}} = a_{ji}$, then $(a_{ij}^{\mathbf{T}})^{\mathbf{T}} = (a_{ji})^{\mathbf{T}} = a_{ij}$ $\square$

Corollary 1.4.3

$$(\mathbf{A}_{m \times n} + \mathbf{B}_{m \times n})^{\mathbf{T}} = (\mathbf{A}_{m \times n})^{\mathbf{T}} + (\mathbf{B}_{m \times n})^{\mathbf{T}}$$

Proof. Since $a_{ij}^{\mathbf{T}} = a_{ji}$ and $b_{ij}^{\mathbf{T}} = b_{ji}$, then $(a_{ij} + b_{ij})^{\mathbf{T}} = a_{ji} + b_{ji} = (a_{ji})^{\mathbf{T}} + b_{ij}^{\mathbf{T}}$ $\square$

Corollary 1.4.4

$$(\mathbf{A}_{m \times n} \mathbf{B}_{m \times n})^{\mathbf{T}} = (\mathbf{B}_{m \times n})^{\mathbf{T}} (\mathbf{A}_{m \times n})^{\mathbf{T}}$$

Proof. Since $a_{ij}^{\mathbf{T}} = a_{ji}$ and $b_{ij}^{\mathbf{T}} = b_{ji}$, then

$$[(\mathbf{A}_{m\times p}\mathbf{B}_{p\times n})^{\mathbf{T}}]_{ij} = [\mathbf{A}_{m\times p}\mathbf{B}_{p\times n}]ji \tag{1.11}$$

$$= \sum_{k=1}^{p} a_{jk}b_{ki} \tag{1.12}$$

$$= \sum_{k=1}^{p} (b_{ik})^{\mathbf{T}}(a_{kj})^{\mathbf{T}} \tag{1.13}$$

$$= [(\mathbf{B}_{p\times n})^{\mathbf{T}}(\mathbf{A}_{m\times p})^{\mathbf{T}}]_{ij} \tag{1.14}$$

$\square$

1.4.2 Symmetric and Skew symmetric

Definition 1.4.5: Symmetric

If $\mathbf{A}$ is a square matrix, and $\mathbf{A}_{n\times n}^{\mathbf{T}} = \mathbf{A}_{n\times n}$, then we say that $\mathbf{A}$ is symmetric.

Definition 1.4.6: Skew symmetric

If $\mathbf{A}$ is a square matrix, and $\mathbf{A}_{n\times n}^{\mathbf{T}} = -\mathbf{A}_{n\times n}$, then we say that $\mathbf{A}$ is skew symmetric.

Corollary 1.4.7

If $\mathbf{A}$ is a square matrix, then the square matrix $\mathbf{A} + \mathbf{A}^{\mathbf{T}}$ is symmetric, and the square matrix $\mathbf{A} - \mathbf{A}^{\mathbf{T}}$ is skew symmetric.

$$\mathbf{A}_{n\times n} = \frac{1}{2}(\mathbf{A}_{n\times n} + (\mathbf{A}_{n\times n})^{\mathbf{T}}) + \frac{1}{2}(\mathbf{A}_{n\times n} - (\mathbf{A}_{n\times n})^{\mathbf{T}})$$

Corollary 1.4.8

$\mathbf{A}^{\mathbf{T}}\mathbf{A}$ is symmetric, $\mathbf{A}$ can be any size matrix.

$$((\mathbf{A}_{m\times n})^{\mathbf{T}}\mathbf{A}_{m\times n})^{\mathbf{T}} = (\mathbf{A}_{m\times n})^{\mathbf{T}}\mathbf{A}_{m\times n}$$

Proof. Since $(\mathbf{AB})^{\mathbf{T}} = \mathbf{B}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}$ and $(\mathbf{A}^{\mathbf{T}})^{\mathbf{T}} = \mathbf{A}$, then

$$((\mathbf{A}_{m\times n})^{\mathbf{T}}\mathbf{A}_{m\times n})^{\mathbf{T}} = (\mathbf{A}_{m\times n})^{\mathbf{T}}((\mathbf{A}_{m\times n})^{\mathbf{T}})^{\mathbf{T}} = (\mathbf{A}_{m\times n})^{\mathbf{T}}\mathbf{A}_{m\times n}$$

$\square$

1.5   Inner and outer products

u and **v** are both n-by-1 column vectors.

$$\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

Definition 1.5.1: Inner product or dot product or scalar product

The inner product (or dot product or scalar product) between two vectors is obtained from the matrix product of a row vector times a column vector. A row vector can be obtained from a column vector by the transpose operator.

$$\langle \mathbf{uv} \rangle = \mathbf{u} \cdot \mathbf{v} \tag{1.15}$$

$$= \mathbf{u^T v} \tag{1.16}$$

$$= \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \tag{1.17}$$

$$= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \tag{1.18}$$

$$= \sum_{i=1}^{n} u_i v_i \tag{1.19}$$

Corollary 1.5.2

The dot product of two vectors is the product of the magnitude of each vector and the cosine of the angle between them:

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

.

Proof. Place vectors **u** and **v** in standard position and consider the vector $\mathbf{v} - \mathbf{u}$. These three vectors form a triangle with side lengths $\|\mathbf{u}\|$, $\|\mathbf{v}\|$, and $\|\mathbf{v} - \mathbf{u}\|$.
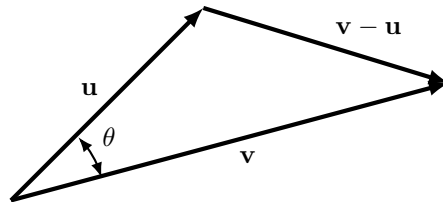

Figure 1.1: The lengths of the sides of the triangle are given by the magnitudes of the vectors that form the triangle.

Recall from trigonometry that the law of cosines describes the relationship among the side lengths of the triangle

and the angle θ. Applying the law of cosines here gives

$$\|\mathbf{v} - \mathbf{u}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{u}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|cos\theta$$

. The dot product provides a way to rewrite the left side of Equation:

$$\|\mathbf{v} - \mathbf{u}\|^2 = (\mathbf{v} - \mathbf{u}) \cdot (\mathbf{v} - \mathbf{u}) \tag{1.20}$$
$$= (\mathbf{v} - \mathbf{u}) \cdot \mathbf{v} - (\mathbf{v} - \mathbf{u}) \cdot \mathbf{u} \tag{1.21}$$
$$= \mathbf{v} \cdot \mathbf{v} - \mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{u} \tag{1.22}$$
$$= \mathbf{v} \cdot \mathbf{v} - \mathbf{u} \cdot \mathbf{v} - \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{u} \tag{1.23}$$
$$= \|\mathbf{v}\|^2 - 2\mathbf{u} \cdot \mathbf{v} + \|\mathbf{u}\|^2 \tag{1.24}$$

Substituting into the law of cosines yields

$$\|\mathbf{v} - \mathbf{u}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta \tag{1.25}$$
$$\|\mathbf{v}\|^2 - 2\mathbf{u} \cdot \mathbf{v} + \|\mathbf{u}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta \tag{1.26}$$
$$-2\mathbf{u} \cdot \mathbf{v} = -2\|\mathbf{u}\|\|\mathbf{v}\|\cos\theta \tag{1.27}$$
$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta \tag{1.28}$$

$$\square$$

Definition 1.5.3: Orthogonal Vector

If the inner product between two nonzero vectors is zero, we say that the vectors are orthogonal.

$$\langle \mathbf{uv} \rangle = \mathbf{u} \cdot \mathbf{v} = \mathbf{u}^{\mathbf{T}}\mathbf{v} = \mathbf{0}$$

Definition 1.5.4: The norm of a vector

$$\|\mathbf{u}\| = (\mathbf{u}^{\mathbf{T}}\mathbf{u})^{1/2} = (u_1^2 + u_2^2 + \cdots + u_n^2)^{1/2}$$

Definition 1.5.5: Normalized Vector

If the norm of a vector is equal to one, we say that the vector is normalized.

$$\|\mathbf{u}\| = 1$$

Definition 1.5.6: Orthonormal Vector

If a set of vectors are mutually orthogonal and normalized, we say that these vectors are orthonormal.

$$\langle \mathbf{u}_i\mathbf{u}_j \rangle = (\mathbf{u}_i)^{\mathbf{T}}\mathbf{u}_j = 0 \ (i \neq j) \ and \ \|\mathbf{u}_n\| = 1$$

Definition 1.5.7: Outer product

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^\mathbf{T} \tag{1.29}$$

$$= \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \tag{1.30}$$

$$= \begin{pmatrix} u_1v_1 & u_1v_2 & \cdots & u_1v_n \\ u_2v_1 & v_2v_2 & \cdots & u_2v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_nv_1 & v_nv_2 & \cdots & u_nv_n \end{pmatrix} \tag{1.31}$$

Notice that every column is a multiple of the single vector $\mathbf{u}$, and every row is a multiple of the single vector $\mathbf{v}^\mathbf{T}$.

Definition 1.5.8: The trace of a square matrix

The trace of a square matrix $\mathbf{A}$, denoted as $\mathrm{Tr}\,\mathbf{A}$, is the sum of the diagonal elements of $\mathbf{A}$.

Corollary 1.5.9

$\mathrm{Tr}(\mathbf{A}^\mathbf{T}\mathbf{A})$ is the sum of the squares of all the elements of $\mathbf{A}$.

Proof. Let $\mathbf{A}$ be an m-by-n matrix. Then

$$\mathrm{Tr}(\mathbf{A}^\mathbf{T}\mathbf{A}) = \sum_{j=1}^{n}(\mathbf{A}^\mathbf{T}\mathbf{A})_{jj} = \sum_{j=1}^{n}\sum_{i=1}^{m}(a_{ji})^\mathbf{T}a_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m}(a_{ij})^2$$

which is the sum of the squares of all the elements of A. $\qquad\square$

1.6 Inverse matrix

Square matrices may have inverses, i.e. some matrices don't have inverses.

Definition 1.6.1: Inverse matrix

When a **square matrix A** has an inverse, we say it is invertible (or nonsingular) and denote its inverse by $\mathbf{A}^{-1}$. The inverse matrix satisfies

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

If a **square matrix A** has no inverse, we say $\mathbf{A}$ is singular or non-invertible.

Corollary 1.6.2

If a **square matrix A** is invertible, then
$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}$$

Corollary 1.6.3

If **A** and **B** are invertible matrices, then

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

Proof. From the definition of an inverse,
$$(\mathbf{AB})^{-1}(\mathbf{AB}) = \mathbf{I}$$
$$(\mathbf{AB})(\mathbf{AB})^{-1} = \mathbf{I}$$

Either multiply on the right by $\mathbf{B}^{-1}$, and then by $\mathbf{A}^{-1}$, or multiply on the left by $\mathbf{A}^{-1}$, and then by $\mathbf{B}^{-1}$, to obtain

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

$\square$

Corollary 1.6.4

if **A** is invertible, then so is $\mathbf{A}^{\mathbf{T}}$, and
$$(\mathbf{A}^{\mathbf{T}})^{-1} = (\mathbf{A}^{-1})^{\mathbf{T}}$$

Proof. We assume that A is invertible so that

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

Taking the transpose of both sides of these two equations, using both $\mathbf{I}^{\mathbf{T}} = \mathbf{I}$ and $(\mathbf{AB})^{\mathbf{T}} = \mathbf{B}^{\mathbf{T}}\mathbf{A}^{\mathbf{T}}$, we obtain

$$(\mathbf{A}\mathbf{A}^{-1})^{\mathbf{T}} = (\mathbf{A}^{-1})^{\mathbf{T}}\mathbf{A}^{\mathbf{T}} = \mathbf{I}^{\mathbf{T}} = \mathbf{I} \ and \ (\mathbf{A}^{-1}\mathbf{A})^{\mathbf{T}} = \mathbf{A}^{\mathbf{T}}(\mathbf{A}^{-1})^{\mathbf{T}} = \mathbf{I}^{\mathbf{T}} = \mathbf{I}$$

. Therefore, since we have found that $(\mathbf{A}^{-1})^{\mathbf{T}}\mathbf{A}^{\mathbf{T}} = \mathbf{I}$ and $\mathbf{A}^{\mathbf{T}}(\mathbf{A}^{-1})^{\mathbf{T}} = \mathbf{I}$, we conclude that $\mathbf{A}^{\mathbf{T}}$ is invertible and that $(\mathbf{A}^{\mathbf{T}})^{-1} = (\mathbf{A}^{-1})^{\mathbf{T}}$. $\square$

Definition 1.6.5: The determinant of the two-by-two matrix

$\mathbf{A}$ is a two-by-two matrix, i.e. $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then the determinant of $\mathbf{A}$ is

$$\det \mathbf{A} = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

Evidently, a two-by-two matrix A is invertible only if $\det \mathbf{A} \neq 0$.

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Notice that the inverse of a two-by-two matrix, in words, is found by switching the diagonal elements of the matrix, negating the off-diagonal elements, and dividing by the determinant.

Proof. Let two-by-two matrix $\mathbf{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, and the inverse matrix is $\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}$. Then

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

To solve, we can eliminate y_1 and y_2 using the two homogeneous equations, and find x_1 and x_2 using the two inhomogeneous equations. The solution for the inverse matrix is found to be

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

The term $ad - bc$ is just the definition of the determinant of the two-by-two matrix. $\qquad \square$

Corollary 1.6.6

An n-by-n matrix is invertible if and only if its determinant is nonzero.

$$\det(\mathbf{A}) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(\mathbf{A}_{ij})$$

where $\mathbf{A}_{ij}$ is the (n-1)×(n-1) submatrix obtained by removing the i-th row and j-th column.

Corollary 1.6.7

if a matrix is invertible, then its inverse is unique.

Proof. Let $\mathbf{A}$ be an invertible matrix, and suppose $\mathbf{B}$ and $\mathbf{C}$ are its inverse. To prove that $\mathbf{B} = \mathbf{C}$, we write

$$\mathbf{B} = \mathbf{BI} = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{IC} = \mathbf{C}$$

$\square$

1.7   Orthogonal matrix

Definition 1.7.1: Orthogonal matrix

A square matrix Q with real entries that satisfies

$$\mathbf{Q}^{-1} = \mathbf{Q}^{\mathbf{T}}$$

is called an orthogonal matrix. Another way to write this definition is

$$\mathbf{Q}\mathbf{Q}^{\mathbf{T}} = \mathbf{I} \ and \ \mathbf{Q}^{\mathbf{T}}\mathbf{Q} = \mathbf{I}$$

Therefore, an equivalent definition of an orthogonal matrix is a square matrix with real entries whose columns (and also rows) form a set of orthonormal vectors.

There is a third equivalent definition of an orthogonal matrix. Let $\mathbf{Q}$ be an n-by-n orthogonal matrix, and let $\mathbf{x}$ be an n-by-one column vector. Then the length squared of the vector $\mathbf{Q}\mathbf{x}$ is given by

$$\|\mathbf{Q}\mathbf{x}\|^2 = (\mathbf{Q}\mathbf{x})^{\mathbf{T}}(\mathbf{Q}\mathbf{x}) = \mathbf{x}^{\mathbf{T}}\mathbf{Q}^{\mathbf{T}}\mathbf{Q}\mathbf{x} = \mathbf{x}^{\mathbf{T}}\mathbf{I}\mathbf{x} = \mathbf{x}^{\mathbf{T}}\mathbf{x} = \|\mathbf{x}\|^2$$

The length of $\mathbf{Q}\mathbf{x}$ is therefore equal to the length of $\mathbf{x}$, and we say that an orthogonal matrix is a matrix that preserves lengths.

1.8   Rotation matrices

Corollary 1.8.1

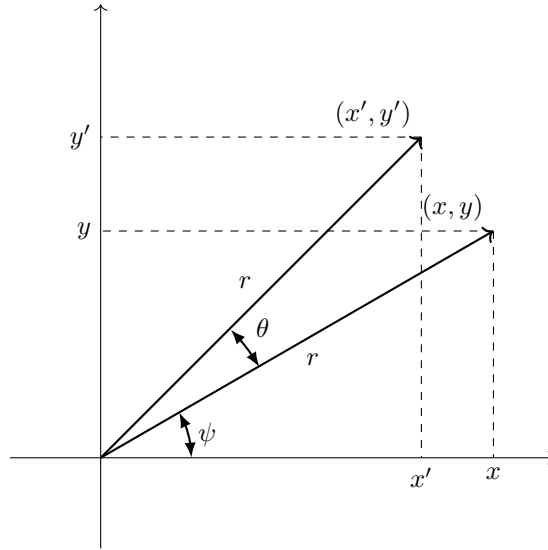A matrix that rotates a vector in space doesn't change the vector's length and so should be an orthogonal matrix.

$$\mathbf{R}_\theta = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}$$

$$(\mathbf{R}_\theta)^{-1} = (\mathbf{R}_\theta)^{\mathbf{T}}$$

$$(\mathbf{R}_\theta)^{\mathbf{T}}\mathbf{R}_\theta = \mathbf{I}$$

$$\mathbf{R}_\theta(\mathbf{R}_\theta)^{\mathbf{T}} = \mathbf{I}$$

$$\|\mathbf{R}_\theta\mathbf{x}\|^2 = (\mathbf{R}_\theta\mathbf{x})^{\mathbf{T}}(\mathbf{R}_\theta\mathbf{x}) = \mathbf{x}^{\mathbf{T}}(\mathbf{R}_\theta)^{\mathbf{T}}\mathbf{R}_\theta\mathbf{x} = \mathbf{x}^{\mathbf{T}}\mathbf{I}\mathbf{x} = \mathbf{x}^{\mathbf{T}}\mathbf{x} = \|\mathbf{x}\|^2$$

Proof.

Figure 1.2: Rotating a vector in the x-y plane.

Consider the two-by-two rotation matrix that rotates a vector counterclockwise through an angle θ in the x-y plane, shown above. Since $\mathbf{x} = \mathbf{r}cos\psi$ and $\mathbf{y} = \mathbf{r}sin\psi$, trigonometry and the addition formula for cosine and sine results in

$$\mathbf{x}^{'} = \mathbf{r}cos(\theta + \psi) \tag{1.32}$$

$$= \mathbf{r}(cos\theta cos\psi - sin\theta sin\psi) \tag{1.33}$$

$$= \mathbf{x}cos\theta - \mathbf{y}sin\theta \tag{1.34}$$

$$\mathbf{y}^{'} = \mathbf{r}sin(\theta + \psi) \tag{1.35}$$

$$= \mathbf{r}(sin\theta cos\psi + cos\theta sin\psi) \tag{1.36}$$

$$= \mathbf{x}sin\theta + \mathbf{y}cos\theta \tag{1.37}$$

Writing the equations for $\mathbf{x}^{'}$ and $\mathbf{y}^{'}$ in matrix form, we have

$$\begin{pmatrix} \mathbf{x}^{'} \\ \mathbf{y}^{'} \end{pmatrix} = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$$

The above two-by-two matrix is a rotation matrix and we will denote it by $\mathbf{R}_\theta$. Observe that the rows and columns of $\mathbf{R}_\theta$ are orthonormal and that the inverse of $\mathbf{R}_\theta$ is just its transpose. The inverse of $\mathbf{R}_\theta$ rotates a vector by $-\theta$.

$\square$

Corollary 1.8.2

$$\mathbf{R}_{(-\theta)} = (\mathbf{R}_\theta)^{-1}$$

Proof. Since $\mathbf{R}_\theta = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix}$ and $(\mathbf{R}_\theta)^{-1} = (\mathbf{R}_\theta)^{\mathbf{T}}$, then

$$\mathbf{R}_{(-\theta)} = \begin{pmatrix} cos(-\theta) & -sin(-\theta) \\ sin(-\theta) & cos(-\theta) \end{pmatrix} \tag{1.38}$$

$$= \begin{pmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{pmatrix} \tag{1.39}$$

$$= (\mathbf{R}_\theta)^{\mathbf{T}} \tag{1.40}$$

$$= (\mathbf{R}_\theta)^{-1} \tag{1.41}$$

$\square$

1.9   Permutation matrices

Definition 1.9.1: Permutation matrix

A permutation matrix is a **square matrix** obtained by permuting the rows or columns of an **identity matrix**.

A permutation matrix is a **square matrix** that has exactly one entry of 1 in each row and each column with all other entries 0.

Corollary 1.9.2

A permutation matrix, when multiplying on the left, permutes the rows of a matrix, and when multiplying on the right, permutes the columns. Clearly, permuting the rows of a column vector will not change its length.

Corollary 1.9.3

The identity matrix is a permutation matrix.

Corollary 1.9.4

A permutation matrix is orthogonal.

Corollary 1.9.5

$$\mathbf{PA} = (\mathbf{PI})\mathbf{A}$$

where P is a permutation matrix and PI is the identity matrix with permuted rows.

Corollary 1.9.6

Six three-by-three permutation matrices:

$$\mathbf{P}_{123} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \ \mathbf{P}_{132} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \ \mathbf{P}_{213} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{P}_{231} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \ \mathbf{P}_{312} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \ \mathbf{P}_{321} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

The inverses of all three-by-three permutation matrices(Since they are all orthogonal matrices, $\mathbf{P}^{-1} = \mathbf{P^T}$):

$$(\mathbf{P}_{123})^{-1} = (\mathbf{P}_{123})^\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{P}_{123}$$

$$(\mathbf{P}_{132})^{-1} = (\mathbf{P}_{132})^\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}^\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \mathbf{P}_{132}$$

$$(\mathbf{P}_{213})^{-1} = (\mathbf{P}_{213})^\mathbf{T} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}^\mathbf{T} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{P}_{213}$$

$$(\mathbf{P}_{231})^{-1} = (\mathbf{P}_{231})^\mathbf{T} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}^\mathbf{T} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = \mathbf{P}_{312}$$

$$(\mathbf{P}_{312})^{-1} = (\mathbf{P}_{312})^\mathbf{T} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}^\mathbf{T} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \mathbf{P}_{231}$$

$$(\mathbf{P}_{321})^{-1} = (\mathbf{P}_{321})^\mathbf{T} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}^\mathbf{T} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \mathbf{P}_{321}$$

Chapter 2

Systems of Linear Equations

2.1 Gaussian elimination

Definition 2.1.1: Gaussian Elimination

Gaussian elimination is a method for solving matrix equations of the form

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

To perform Gaussian elimination starting with the system of equations

$$
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix}
=
\begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_n
\end{bmatrix}
\tag{2.1}
$$

compose the "augmented matrix equation"

$$
\left[
\begin{array}{cccc|c}
a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\
a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn} & b_n
\end{array}
\right]
\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{bmatrix}
\tag{2.2}
$$

Here, the column vector in the variables $\mathbf{x}$ is carried along for labeling the matrix rows. Now, perform elementary row operations to put the augmented matrix into the upper triangular form

$$
\left[
\begin{array}{cccc|c}
a_{11}' & a_{12}' & \cdots & a_{1n}' & b_1' \\
0 & a_{22}' & \cdots & a_{2n}' & b_2' \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & a_{nn}' & b_n'
\end{array}
\right]
\tag{2.3}
$$

Corollary 2.1.2

Solve the equation of the kth row for $\mathbf{x}_k$, then substitute back into the equation of the (k-1)st row to obtain a solution for $\mathbf{x}_{k-1}$, etc., according to the formula

$$x_i = \frac{1}{a'_{ii}} \left(b'_i - \sum_{j=i+1}^{n} a'_{ij} x_j \right)$$

Definition 2.1.3: pivot

When performing Gaussian elimination, the matrix element that is used during the elimination procedure is called the pivot.

Corollary 2.1.4

Gaussian elimination can be used on any rectangular matrix $\mathbf{A_{m \times n}}$.

2.2 Reduced row echelon form

Definition 2.2.1: Reduced row echelon form

A matrix is said to be in reduced row echelon form if the first nonzero entry in every row is a one, all the entries below and above this one are zero, and any zero rows occur at the bottom of the matrix.

Corollary 2.2.2

The row elimination procedure of Gaussian elimination can be continued to bring a matrix to reduced row echelon form. We notate the reduced row echelon form of a matrix $\mathbf{A}$ as rref($\mathbf{A}$).

Corollary 2.2.3

The matrix in reduced row echelon form has the following characteristics:
1. All zero rows are at the bottom of the matrix
2. The leading entry of each nonzero row after the first occurs to the right of the leading entry of the previous row.
3. The leading entry in any nonzero row is 1.
4. All entries in the column above and below a leading 1 are zero.

Definition 2.2.4: Pivot row

Each nonzero row of rref($\mathbf{A}$) is called a pivot row.

Definition 2.2.5: Pivot

The first nonzero in each row of $\mathrm{rref}(\mathbf{A})$ is called a pivot.

Definition 2.2.6: Pivot column

Each column that contains a pivot in $\mathrm{rref}(\mathbf{A})$ is called a pivot column.

Corollary 2.2.7

The reduced row echelon form of a matrix A is unique.

How to prove it?

Corollary 2.2.8

if $\mathbf{A}$ is a square invertible matrix, then $\mathrm{rref}(\mathbf{A})$ is the identity matrix.

How to prove it?

2.3   Computing inverses

Definition 2.3.1: Finding the inverse of a matrix

If $\mathbf{A}$ is an $n \times n$ **square matrix**, then one can use row reduction to compute its inverse matrix, if it exists. First, the $n \times n$ identity matrix is augmented to the right of $\mathbf{A}$, forming an $n \times 2n$ block matrix $[\mathbf{A} \mid \mathbf{I}]$. Now through application of elementary row operations, find the reduced echelon form of this $n \times 2n$ matrix. The matrix $\mathbf{A}$ is invertible if and only if the left block can be reduced to the identity matrix $\mathbf{I}$; in this case the right block of the final matrix is $\mathbf{A}^{-1}$. If the algorithm is unable to reduce the left block to $\mathbf{I}$, then $\mathbf{A}$ is not invertible.

$$[\mathbf{A} \mid \mathbf{I}] = \left[\begin{array}{cccc|cccc} a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 & 0 & \cdots & 1 \end{array}\right] \rightarrow [\mathbf{I} \mid \mathbf{B}] = \left[\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & b_{11} & b_{12} & \cdots & b_{1n} \\ 0 & 1 & \cdots & 0 & b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & b_{n1} & b_{n2} & \cdots & b_{nn} \end{array}\right]$$
$$(2.4)$$

On the left, that $\mathbf{BA} = \mathbf{I}$, therefore $\mathbf{B} = \mathbf{A}^{-1}$. On the right, we kept a record of $\mathbf{BI} = \mathbf{B}$. So

$$\mathbf{A}^{-1} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} \tag{2.5}$$

2.4   Elementary matrices

Definition 2.4.1: Elementary matrix

An elementary matrix $\mathbf{E}$ is a **square matrix** obtained from the application of a single elementary row operation to the **identity matrix**.

Corollary 2.4.2

The row reduction algorithm of Gaussian elimination can be implemented by multiplying elementary matrices.

Example.

$$\mathbf{M_3 M_2 M_1 A = U}$$

where U is an upper triangular matrix.

2.5   LU decomposition

Definition 2.5.1: LU decomposition

A procedure for decomposing an matrix $\mathbf{A}$ into a product of a lower triangular matrix $\mathbf{L}$ and an upper triangular matrix $\mathbf{U}$,

$$\mathbf{LU = A}$$

Example.

If $\mathbf{M_3 M_2 M_1 A = U}$, where U is an upper triangular matrix, then

$$\mathbf{L = M_1^{-1} M_2^{-1} M_3^{-1}}$$

2.6 Solving $(LU)x = b$

> **Definition 2.6.1: Solving $(LU)x = b$**
>
> The **LU** decomposition is useful when one needs to solve
>
> $$\mathbf{Ax} = \mathbf{b}$$
>
> for $\mathbf{x}$ when the matrix $\mathbf{A}$ is fixed and there are many different $\mathbf{b}$'s. First one determines $\mathbf{L}$ and $\mathbf{U}$ using Gaussian elimination. Then one writes
>
> $$(\mathbf{LU})\mathbf{x} = \mathbf{L}(\mathbf{Ux}) = \mathbf{b}$$
>
> We let
>
> $$\mathbf{y} = \mathbf{Ux}$$
>
> and first solve
>
> $$\mathbf{Ly} = \mathbf{b}$$
>
> for $\mathbf{y}$. This can be done by forward substitution
> **Example.**
>
> $$y_1 = \frac{b_1}{l_{11}}$$
>
> $$y_i = \frac{1}{l_{ii}} \left(b_i - \sum_{j=1}^{i-1} l_{ij} y_j \right)$$
>
> for square matrix A with $i = 2, \ldots, n$ and $l_{ii} \neq 0$.
> Then solve
>
> $$\mathbf{Ux} = \mathbf{y}$$
>
> for $\mathbf{x}$. This can be done by back substitution
> **Example.**
>
> $$x_n = \frac{y_n}{u_{nn}}$$
>
> $$x_i = \frac{1}{u_{ii}} \left(y_i - \sum_{j=i+1}^{n} u_{ij} x_j \right)$$
>
> for square matrix A $i = n - 1, \ldots, 1$ and $u_{ii} \neq 0$.
> If we count operations, we can show that solving $(\mathbf{LU})\mathbf{x} = \mathbf{b}$ is a factor of n faster once $\mathbf{L}$ and $\mathbf{U}$ are in hand than solving $\mathbf{Ax} = \mathbf{b}$ directly by Gaussian elimination.

Chapter 3

Vector Spaces

3.1 Vector spaces

<div style="background-color:#d5f5d5; padding:10px;">

Definition 3.1.1: Vector spaces

A vector space consists of a set of vectors and a set of scalars. For the set of vectors and scalars to form a vector space, the set of vectors must be closed under vector addition and scalar multiplication. That is, when you multiply any two vectors in the set by real numbers and add them, the resulting vector must still be in the set.

</div>

Corollary 3.1.2

A vector space V is a set that is **closed under finite vector addition and scalar multiplication**.
For a general vector space, the scalars are members of a field $\mathbf{F}$, in which case $\mathbf{V}$ is called a vector space over $\mathbf{F}$.
In order for $\mathbf{V}$ to be a vector space, the following conditions must hold for all elements $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathbf{V}$ and any scalars $\mathbf{r}, \mathbf{s} \in \mathbf{F}$:

1. **(Additive Closure)** $\mathbf{X} + \mathbf{Y} \in \mathbf{V}$. Adding two vectors gives a vector.

2. **(Additive Commutativity)** $\mathbf{X} + \mathbf{Y} = \mathbf{Y} + \mathbf{X}$. Order of addition doesn't matter.

3. **(Additive Associativity)** $(\mathbf{X} + \mathbf{Y}) + \mathbf{Z} = \mathbf{X} + (\mathbf{Y} + \mathbf{Z})$. Order of adding many vectors doesn't matter.

4. **(Zero)** There is a special vector $\mathbf{0_V} \in \mathbf{V}$ such that $\mathbf{X} + \mathbf{0_V} = \mathbf{0_V} + \mathbf{X} = \mathbf{X}$ for all $\mathbf{X}$ in $\mathbf{V}$.

5. **(Additive Inverse)** For every $\mathbf{X} \in \mathbf{V}$ there exists $-\mathbf{X} \in \mathbf{V}$ such that $\mathbf{X} + (-\mathbf{X}) = \mathbf{0_V}$.

6. **(Multiplicative Closure)** $\mathbf{rX} \in \mathbf{V}$. Scalar times a vector is a vector.

7. **(Distributivity)** $(\mathbf{r} + \mathbf{s})\mathbf{X} = \mathbf{rX} + \mathbf{sX}$. Scalar multiplication distributes over addition of scalars.

8. **(Distributivity)** $\mathbf{r}(\mathbf{X} + \mathbf{Y}) = \mathbf{rX} + \mathbf{rY}$. Scalar multiplication distributes over addition of vectors.

9. **(Associativity)** $(\mathbf{rs})\mathbf{X} = \mathbf{r}(\mathbf{sX})$.

10. **(Unity)** $\mathbf{1X} = \mathbf{X}$ for all $\mathbf{X} \in \mathbf{V}$.

3.2 Linear independence

Definition 3.2.1: Linearly independent

The vectors $\{\mathbf{u_1}, \mathbf{u_2}, \dots, \mathbf{u_n}\}$ are linearly independent if for any scalars $c_1, c_2, \dots, c_n$, the equation $c_1\mathbf{u_1} + c_2\mathbf{u_2} + \dots + c_n\mathbf{u_n} = 0$ has only the solution $c_1 = c_2 = \dots = c_n = 0$. What this means is that one is unable to write any of the vectors $\mathbf{u_1}, \mathbf{u_2}, \dots, \mathbf{u_n}$ as a linear combination of any of the other vectors.

Corollary 3.2.2

place the vectors as the rows of a matrix and compute the reduced row echelon form. If the last row becomes all zeros, then the vectors are linearly dependent, and if not all zeros, then they are linearly independent

Definition 3.2.3: Linearly independent Vectors

A list of vectors $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ is called linearly independent if the only solution for $a_1, \ldots, a_m \in \mathbf{F}$ to the equation

$$a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m = 0$$

is $a_1 = \ldots = a_m = 0$. In other words, the zero vector can only trivially be written as a linear combination of $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$.

Definition 3.2.4: Linearly dependent Vectors

A list of vectors $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ is called linearly dependent if it is not linearly independent. That is, $(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ is linear dependent if there exist $a_1, \ldots, a_m \in \mathbf{F}$, not all zero, such that

$$a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m = 0$$

3.3   Span, basis and dimension

Definition 3.3.1: Span of vectors

The span of the set of vectors $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is the vector space consisting of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. We say that a set of vectors spans a vector space.
Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subseteq V$, then

$$\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} = \left\{ \sum_{i=1}^{n} c_i\mathbf{v}_i, c_i \in \mathbf{F} \right\}$$

Definition 3.3.2: Basis and Dimension

The smallest set of vectors needed to span a vector space forms a basis for that vector space.
The number of vectors in a basis gives the dimension of the vector space.
Let $\mathbf{V}$ be a vector space.

- Then a set $\mathbf{S}$ is a basis for $\mathbf{V}$ if $\mathbf{S}$ is linearly independent and $\mathbf{V} = \text{span}\{S\}$.

- The number of vectors in $\mathbf{S}$ is the dimension of $\mathbf{V}$.

Corollary 3.3.3

Let $\mathbf{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for a vector space V. Then every vector $\mathbf{w} \in \mathbf{V}$ can be written uniquely as a linear combination of vectors in the basis $\mathbf{S}$:

$$\mathbf{w} = c_1\mathbf{v}_1 + \ldots + c_n\mathbf{v}_n$$

3.4 Gram-Schmidt process

Definition 3.4.1: Gram-Schmidt process

Given any basis for a vector space, we can use an algorithm called the Gram-Schmidt process to construct an orthonormal basis for that space. Let the vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be a basis for some n-dimensional vector space. We will assume here that these vectors are column matrices, but this process also applies more generally. We will construct an orthogonal basis $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$, and then normalize each vector to obtain an orthonormal basis. First, define $\mathbf{u}_1 = \mathbf{v}_1$. To find the next orthogonal basis vector, define

$$\mathbf{u}_2 = \mathbf{v}_2 - \frac{(\mathbf{u}_1^{\mathbf{T}} \mathbf{v}_2)\mathbf{u}_1}{\mathbf{u}_1^{\mathbf{T}} \mathbf{u}_1}$$

. Observe that $\mathbf{u}_2$ is equal to $\mathbf{v}_2$ minus the component of $\mathbf{v}_2$ that is parallel to $\mathbf{u}_1$. By multiplying both sides of this equation with $\mathbf{u}_1^{\mathbf{T}}$, it is easy to see that $\mathbf{u}_1^{\mathbf{T}} \mathbf{u}_2 = 0$ so that these two vectors are orthogonal. The next orthogonal vector in the new basis can be found from

$$\mathbf{u}_3 = \mathbf{v}_3 - \frac{(\mathbf{u}_1^{\mathbf{T}} \mathbf{v}_3)\mathbf{u}_1}{\mathbf{u}_1^{\mathbf{T}} \mathbf{u}_1} - \frac{(\mathbf{u}_2^{\mathbf{T}} \mathbf{v}_3)\mathbf{u}_2}{\mathbf{u}_2^{\mathbf{T}} \mathbf{u}_2}$$

Here, $\mathbf{u}_3$ is equal to $\mathbf{v}_3$ minus the components of $\mathbf{v}_3$ that are parallel to $\mathbf{u}_1$ and $\mathbf{u}_2$. We can continue in this fashion to construct n orthogonal basis vectors.

$$\mathbf{u}_k = \mathbf{v}_k - \sum_{i=1}^{k-1} \frac{(\mathbf{u}_i^{\mathbf{T}} \mathbf{v}_n)\mathbf{u}_i}{\mathbf{u}_i^{\mathbf{T}} \mathbf{u}_i}$$

These vectors can then be normalized via

$$\mathbf{u}_k = \frac{\mathbf{u}_k}{(\mathbf{u}_k^{\mathbf{T}} \mathbf{u}_k)^{1/2}}, \text{ etc.}$$

Since $\mathbf{u}_k$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k$, the vector subspace spanned by the first k basis vectors of the original vector space is the same as the subspace spanned by the first k orthonormal vectors generated through the Gram-Schmidt process. We can write this result as

$$\text{span}\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$$

3.5 Null space

Definition 3.5.1: Null space

The null space of a matrix $\mathbf{A}$, which we denote as $\text{Null}(\mathbf{A})$, is the vector space spanned by all column vectors $\mathbf{x}$ that satisfy the matrix equation

$$\mathbf{Ax} = \mathbf{0}$$

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, the null space of A is:

$$\text{Null}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\}$$

Corollary 3.5.2

Clearly, if $\mathbf{x}$ and $\mathbf{y}$ are in the null space of $\mathbf{A}$, then so is $\mathbf{ax} + \mathbf{by}$ so that the null space is closed under vector addition and scalar multiplication.

Corollary 3.5.3

If the matrix $\mathbf{A}$ is m-by-n, then Null$(\mathbf{A})$ is a vector subspace of all n-by-one column matrices.

Corollary 3.5.4

If $\mathbf{A}$ is a square invertible matrix, then Null$(\mathbf{A})$ consists of just the zero vector.

Definition 3.5.5: Basic variables

The variables associated with the pivot columns are called basic variables.

Definition 3.5.6: Free variables

The variables associated with the non-pivot columns are called free variables.

Corollary 3.5.7

In general, the dimension of Null$(\mathbf{A})$ is equal to the number of non-pivot columns of rref$(\mathbf{A})$.

$$\text{Null}(\mathbf{A}) = \text{Null}(\text{rref}(\mathbf{A}))$$

$$\dim(\text{Null}(\mathbf{A})) = \dim(\text{Null}(\text{rref}(\mathbf{A})))$$

Definition 3.5.8: Basis for null space

To find the null space of a matrix A, follow these steps:

1. Write the system of linear equations $\mathbf{Ax} = \mathbf{0}$.
   Set up the matrix equation $\mathbf{Ax} = \mathbf{0}$, where $\mathbf{A}$ is the given matrix and $\mathbf{x} = [x_1, x_2, \cdots, x_n]^{\mathbf{T}}$ is the vector of variables.

2. Convert the augmented matrix $[\mathbf{A} \mid \mathbf{0}]$ to reduced row echelon form (RREF) rref($[\mathbf{A} \mid \mathbf{0}]$).
   Apply row operations to transform the matrix $\mathbf{A}$ into its RREF form. The augmented matrix $[\mathbf{A} \mid \mathbf{0}]$ will become $[\mathbf{B} \mid \mathbf{0}]$, where $\mathbf{B}$ is the RREF of $\mathbf{A}$.

3. Identify the free variables in the RREF system.
   The free variables are the variables that are not associated with the pivot columns in the RREF.

4. Express the basic variables in terms of the free variables.
   Write the solution to the system of linear equations $\mathbf{Ax} = \mathbf{0}$ by expressing the basic variables (the variables associated with the pivot columns) in terms of the free variables.
   The general solution will have the form $\mathbf{x} = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$, where $v_1, v_2, \cdots, v_k$ are the vectors corresponding to the free variables, and $c_1, c_2, \cdots, c_k$ are the free variable coefficients.

5. The vectors $v_1, v_2, \cdots, v_k$ form a basis for the null space of A.
   The null space of $\mathbf{A}$ is the set of all vectors $\mathbf{x}$ that satisfy $\mathbf{Ax} = \mathbf{0}$.
   The vectors $v_1, v_2, \cdots, v_k$ span the null space and represent a basis for the null space.

$$\text{Null}(\mathbf{A}) = \text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$$

3.6 Application of the null space

Corollary 3.6.1

An under-determined system of linear equations $\mathbf{Ax} = \mathbf{b}$ with more unknowns than equations may not have a unique solution. If $\mathbf{u}$ is the general form of a vector in the null space of $\mathbf{A}$, and $\mathbf{v}$ is any vector that satisfies $\mathbf{Av} = \mathbf{b}$, then $\mathbf{x} = \mathbf{u} + \mathbf{v}$ satisfies $\mathbf{Ax} = \mathbf{A}(\mathbf{u} + \mathbf{v}) = \mathbf{Au} + \mathbf{Av} = \mathbf{0} + \mathbf{b} = \mathbf{b}$. The general solution of $\mathbf{Ax} = \mathbf{b}$ can therefore be written as the sum of a general vector in as Null($\mathbf{A}$) and a particular vector that satisfies the under-determined system.

3.7   Column space

Definition 3.7.1: Column space

The column space of a matrix is the vector space spanned by the columns of the matrix.

For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, the column space of A is:

$$\text{Col}(\mathbf{A}) = \{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n\} \tag{3.1}$$

$$= \{c_1 x_1 + \cdots + c_n x_n \mid x_1, \cdots, x_n \in \mathbb{R}\} \tag{3.2}$$

where $c_1, \cdots, c_n$ are the columns of $\mathbf{A}$.

Since $\mathbf{A}$ has m rows, the column space of $\mathbf{A}$ is a subspace of all m-by-one column matrices.

Corollary 3.7.2

When a matrix is multiplied by a column vector, the resulting vector is in the column space of the matrix.

Corollary 3.7.3

It should be self-evident that only the pivot columns of $\text{rref}(\mathbf{A})$ are linearly independent, and the dimension of the column space of $\mathbf{A}$ $\text{Col}(\mathbf{A})$ is therefore equal to its number of pivot columns.

$$\text{Col}(\mathbf{A}) = \text{Col}(\text{rref}(\mathbf{A})), \text{ if } \mathbf{A} \text{ full rank.}$$

$$\dim(\text{Col}(\mathbf{A})) = \dim(\text{Col}(\text{rref}(\mathbf{A})))$$

Definition 3.7.4: Basis for column space

Suppose $\mathbf{A}$ is m-by-n. If columns $\{c_j \mid j = 1, \cdots, r\}$ are the pivot columns of $\text{rref}(\mathbf{A})$ then columns $\{c_j \mid j = 1, \cdots, r\}$ of A constitute a basis for $\text{Col}(\mathbf{A})$

Corollary 3.7.5

Recall that the dimension of the null space is the number of non-pivot columns—equal to the number of free variables—so that the sum of the dimensions of the null space and the column space is equal to the total number of columns. A statement of this theorem is as follows. Let A be an m-by-n matrix. Then

$$\dim(\text{Col}(\mathbf{A})) + \dim(\text{Null}(\mathbf{A})) = \mathbf{n}$$

3.8 Row space, left null space and rank

Definition 3.8.1: Left null space

The left null space of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the null space of its transpose, i.e.,

$$\text{LeftNull}(\mathbf{A}) = \text{Null}(\mathbf{A^T}) \tag{3.3}$$

$$= \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{A^T y = 0}\} \tag{3.4}$$

The word "left" in this context stems from the fact that $\mathbf{A^T y = 0}$ is equivalent to $\mathbf{y^T A = 0}$ where $\mathbf{y}$ "acts" on $\mathbf{A}$ from the left.

Corollary 3.8.2

$$\text{LeftNull}(\mathbf{A}) = \text{Null}(\mathbf{A^T})$$

$$\text{LeftNull}(\mathbf{A^T}) = \text{Null}((\mathbf{A^T})^T) = \text{Null}(\mathbf{A})$$

Definition 3.8.3: Row space

The column space of $\mathbf{A^T}$, which are called the row space.
For a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, the row space of $\mathbf{A}$ is:

$$\text{Row}(\mathbf{A}) = \text{Col}(\mathbf{A^T}) \tag{3.5}$$

$$= \{\mathbf{A^T y} \mid \mathbf{y} \in \mathbb{R}^m\} \tag{3.6}$$

$$= \{c_1 y_1 + \cdots + c_n y_n \mid y_1, \cdots, y_n \in \mathbb{R}\} \tag{3.7}$$

where $c_1, \cdots, c_n$ are the columns of $\mathbf{A^T}$, i.e., the rows of $\mathbf{A}$.
The row space of the m-by-n matrix A is simply the span of its rows, i.e., Since $\mathbf{A}$ has n columns, the row space of $\mathbf{A}$ is a subspace of all one-by-n row matrices.

Corollary 3.8.4

$$\text{Row}(\mathbf{A}) = \text{Col}(\mathbf{A^T})$$

$$\text{Row}(\mathbf{A^T}) = \text{Col}((\mathbf{A^T})^T) = \text{Col}(\mathbf{A})$$

Definition 3.8.5: Basis for row space

Suppose $\mathbf{A}$ is m-by-n. The pivot rows of rref$(\mathbf{A})$ constitute a basis for Col$(\mathbf{A^T})$, that is Row$(\mathbf{A})$.

Corollary 3.8.6

If **A** is an m-by-n matrix, then the row space and the null space are subspaces of all n-by-one column matrices, and the column space and the left null space are subspaces of all m-by-one column matrices.

Corollary 3.8.7

The null space consists of all vectors $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{0}$, that is, the null space is the set of all vectors that are orthogonal to the row space of **A**. We say that these two vector spaces are orthogonal.

$$\mathrm{Null}(\mathbf{A}) \perp \mathrm{Row}(\mathbf{A})$$

Corollary 3.8.8

A basis for the row space of a matrix can be found from computing rref(A), and is found to be rows of rref(A) (written as column vectors) with pivot columns. The dimension of the row space of A is therefore equal to the number of pivot columns, while the dimension of the null space of A is equal to the number of nonpivot columns.

$$\mathrm{Null}(\mathbf{A}) = \mathrm{Null}(\mathrm{rref}(\mathbf{A}))$$

$$\mathrm{Row}(\mathbf{A}) = \mathrm{Row}(\mathrm{rref}(\mathbf{A}))$$

The union of these two subspaces make up the vector space of all n-by-one matrices and we say that these subspaces are orthogonal complements of each other.

$$\mathrm{Null}(\mathbf{A}) = \mathrm{Row}(\mathbf{A})^{\perp}$$

$$\mathrm{Row}(\mathbf{A}) = \mathrm{Null}(\mathbf{A})^{\perp}$$

Definition 3.8.9

The nullity of a matrix $\mathbf{A}_{m \times n}$, written nullity($\mathbf{A}$), is the dimension of the null space Null($\mathbf{A}$).

Definition 3.8.10: Rank

The rank of a matrix $\mathbf{A}_{m \times n}$, written rank($\mathbf{A}$), is the dimension of the column space Col($\mathbf{A}$), which is also equal to the number of pivot columns, so that the dimensions of the column space and the row space of a matrix are equal.

$$\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{A^T}) = \dim(\mathrm{Col}(\mathbf{A})) = \dim(\mathrm{Row}(\mathbf{A})) \leq \min\{m, n\}$$

We call this dimension the rank of the matrix A.

Definition 3.8.11: Full rank

When the equality rank($\mathbf{A}$) = $\min\{m, n\}$ holds, we say that the matrix $\mathbf{A}_{m \times n}$ is of full rank.

Corollary 3.8.12

$$\mathbf{A}_{m \times n}(m \leq n) \text{ is full row rank} \iff \text{All rows of matrix } \mathbf{A}_{m \times n} \text{ are linearly independent} \tag{3.8}$$

$$\iff \text{rank}(\mathbf{A}_{m \times n}) = m \tag{3.9}$$

$$\iff \text{left nullity}(\mathbf{A}_{m \times n}) = 0 \tag{3.10}$$

$$\iff \mathbf{A}_{m \times n}\mathbf{A}_{m \times n}^{-1} = \mathbf{I} \tag{3.11}$$

Corollary 3.8.13

$$\mathbf{A}_{m \times n}(m \geq n) \text{ is full column rank} \iff \text{All columns of matrix } \mathbf{A}_{m \times n} \text{ are linearly independent} \tag{3.12}$$

$$\iff \text{rank}(\mathbf{A}_{m \times n}) = n \tag{3.13}$$

$$\iff \text{nullity}(\mathbf{A}_{m \times n}) = 0 \tag{3.14}$$

$$\iff \mathbf{A}_{m \times n}^{-1}\mathbf{A}_{m \times n} = \mathbf{I} \tag{3.15}$$

Corollary 3.8.14

$$\mathbf{A}_{n \times n} \text{ is square full rank} \iff \text{All rows of matrix } \mathbf{A}_{n \times n} \text{ are linearly independent} \tag{3.16}$$

$$\iff \text{All columns of matrix } \mathbf{A}_{n \times n} \text{ are linearly independent} \tag{3.17}$$

$$\iff \text{rank}(\mathbf{A}_{n \times n}) = n \tag{3.18}$$

$$\iff \text{nullity}(\mathbf{A}_{n \times n}) = \text{left nullity}(\mathbf{A}_{n \times n}) = 0 \tag{3.19}$$

$$\iff \det(\mathbf{A}_{n \times n}) \neq n \tag{3.20}$$

$$\iff \mathbf{A}_{n \times n} \text{ is invertible: } \mathbf{A}_{n \times n}\mathbf{A}_{n \times n}^{-1} = \mathbf{A}_{n \times n}^{-1}\mathbf{A}_{n \times n} = \mathbf{I} \tag{3.21}$$

Definition 3.8.15: Non-zero Determinant

Corollary 3.8.16

when A is a square matrix and of full rank, then the dimension of the null space is zero and A is invertible.

Corollary 3.8.17

Concept	Notation	Mathematical Definition	Subspace of
Null Space	$\text{Null}(\mathbf{A}) = \text{LeftNull}(\mathbf{A^T})$	All vectors x that $text satisfy \mathbf{Ax} = \mathbf{0}$ $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\}$	$\mathbb{R}^n$ (Domain of $\mathbf{A}$)
Left Null Space	$\text{LeftNull}(\mathbf{A}) = \text{Null}(\mathbf{A^T})$	All vectors y that satisfy $\mathbf{A^T y} = \mathbf{0}$ $\{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{A^T y} = \mathbf{0}\}$	$\mathbb{R}^m$ (Codomain of $\mathbf{A^T}$)
Column Space	$\text{Col}(\mathbf{A}) = \text{Row}(\mathbf{A^T})$	All linear combinations of the columns of $\mathbf{A}$ $\{\mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n\}$	$\mathbb{R}^m$ (Codomain of $\mathbf{A}$)
Row Space	$\text{Row}(\mathbf{A}) = \text{Col}(\mathbf{A^T})$	All linear combinations of the rows of $\mathbf{A}$ $\{\mathbf{A^T y} \mid \mathbf{y} \in \mathbb{R}^m\}$	$\mathbb{R}^n$ (Domain of $\mathbf{A^T}$)

Table 3.1: Comparison of Null Space, Left Null Space, Column Space, and Row Space for an $m \times n$ matrix $\mathbf{A}$

Concept	Dimension	Relationships	Relationships2
Null Space	$\dim(\text{Null}(\mathbf{A})) = \dim(\text{Null}(\text{rref}(\mathbf{A})))$ $= \dim(\text{LeftNull}(\mathbf{A^T}))$ $= \text{nullity}(\mathbf{A})$ $= \text{nullity}(\text{rref}(\mathbf{A}))$ $\text{nullity}(\mathbf{A}) + \text{rank}(\mathbf{A}) = n$	$\text{Null}(\mathbf{A}) \perp \text{Row}(\mathbf{A})$ $\text{Null}(\mathbf{A}) = \text{Row}(\mathbf{A})^\perp$	$\text{Null}(\mathbf{A}) = \text{Null}(\text{rref}(\mathbf{A}))$
Left Null Space	$\dim(\text{LeftNull}(\mathbf{A})) = \dim(\text{LeftNull}(\text{rref}(\mathbf{A})))$ $= \dim(\text{Null}(\mathbf{A^T}))$ $= \text{left nullity}(\mathbf{A})$ $= \text{left nullity}(\text{rref}(\mathbf{A}))$ $\text{left nullity}(\mathbf{A}) + \text{rank}(\mathbf{A}) = m$	$\text{LeftNull}(\mathbf{A}) \perp \text{Col}(\mathbf{A})$ $\text{LeftNull}(\mathbf{A}) = \text{Col}(\mathbf{A})^\perp$	$\text{LeftNull}(\mathbf{A}) = \text{LeftNull}(\text{rref}(\mathbf{A}))$
Column Space	$\dim(\text{Col}(\mathbf{A})) = \dim(\text{Col}(\text{rref}(\mathbf{A})))$ $= \text{rank}(\mathbf{A})$ $= \text{rank}(\text{rref}(\mathbf{A}))$ $= \dim(\text{Row}(\mathbf{A}))$ $\text{left nullity}(\mathbf{A}) + \text{rank}(\mathbf{A}) = m$ $\text{nullity}(\mathbf{A}) + \text{rank}(\mathbf{A}) = n$	$\text{Col}(\mathbf{A}) \perp \text{LeftNull}(\mathbf{A})$ $\text{Col}(\mathbf{A}) = \text{LeftNull}(\mathbf{A})^\perp$	$\text{Col}(\mathbf{A}) = \text{Col}(\text{rref}(\mathbf{A}))$, if $\mathbf{A}$ is full rank.
Row Space	$\dim(\text{Row}(\mathbf{A})) = \dim(\text{Row}(\text{rref}(\mathbf{A})))$ $= \text{rank}(\mathbf{A})$ $= \text{rank}(\text{rref}(\mathbf{A}))$ $= \dim(\text{Col}(\mathbf{A}))$ $\text{left nullity}(\mathbf{A}) + \text{rank}(\mathbf{A}) = m$ $\text{nullity}(\mathbf{A}) + \text{rank}(\mathbf{A}) = n$	$\text{Row}(\mathbf{A}) \perp \text{Null}(\mathbf{A})$ $\text{Row}(\mathbf{A}) = \text{Null}(\mathbf{A})^\perp$	$\text{Row}(\mathbf{A}) = \text{Row}(\text{rref}(\mathbf{A}))$

Table 3.2: Comparison of Null Space, Left Null Space, Column Space, and Row Space for an $m \times n$ matrix $\mathbf{A}$

3.9 Orthogonal projections

Definition 3.9.1: Orthogonal projections

Suppose that $\mathbf{V}$ is an n-dimensional vector space and W is a p-dimensional subspace of $\mathbf{V}$. Let $\{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_p\}$ be an orthonormal basis for $\mathbf{W}$. Extending the basis for $\mathbf{W}$, let $\{\mathbf{s}_1, s2_2, \cdots, \mathbf{s}_p, \mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_{n-p}\}$ be an orthonormal basis for $\mathbf{V}$.

Any vector $\mathbf{v}$ in V can be written in terms of the basis for $\mathbf{V}$ as

$$\mathbf{v} = a_1\mathbf{s}_1 + a_2\mathbf{s}_2 + \cdots + a_p\mathbf{s}_p + b_1\mathbf{t}_1 + b_2\mathbf{t}_2 + b_{n-p}\mathbf{t}_{n-p}$$

The orthogonal projection of $\mathbf{v}$ onto $\mathbf{W}$ is then defined as

$$\mathbf{v}_{proj_W} = a_1\mathbf{s}_1 + a_2\mathbf{s}_2 + \cdots + a_p\mathbf{s}_p$$

, that is, the part of $\mathbf{v}$ that lies in $\mathbf{W}$. If you only know the vector $\mathbf{v}$ and the orthonormal basis for $\mathbf{W}$, then the orthogonal projection of $\mathbf{v}$ onto $\mathbf{W}$ can be computed from

$$\mathbf{v}_{proj_W} = (\mathbf{v}^{\mathbf{T}}\mathbf{s}_1)\mathbf{s}_1 + (\mathbf{v}^{\mathbf{T}}\mathbf{s}_2)\mathbf{s}_2 + \cdots + (\mathbf{v}^{\mathbf{T}}\mathbf{s}_p)\mathbf{s}_p$$

, that is, $a_1 = \mathbf{v}^{\mathbf{T}}\mathbf{s}_1, a2 = \mathbf{v}^{\mathbf{T}}\mathbf{s}_2$, etc.

Corollary 3.9.2

The vector $\mathbf{v}_{proj_W}$ is the vector in $\mathbf{W}$ that is closest to $\mathbf{v}$.

Proof. Let $\mathbf{w}$ be any vector in W different than $\mathbf{v}_{proj_W}$, and expand $\mathbf{w}$ in terms of the basis vectors for $\mathbf{W}$:

$$\mathbf{w} = c_1\mathbf{s}_1 + c_2\mathbf{s}_2 + \cdots + c_p\mathbf{s}_p$$

The distance between $\mathbf{v}$ and $\mathbf{w}$ is given by the norm $\|\mathbf{v} - \mathbf{w}\|$, and we have

$$\|\mathbf{v} - \mathbf{w}\|^2 = (a_1 - c_1)^2 + (a_2 - c_2)^2 + \cdots + (a_p - c_p)^2 + b_1^2 + b_2^2 + \cdots + b_{n-p}^2 \geq b_1^2 + b_2^2 + \cdots + b_{n-p}^2 = \|\mathbf{v} - \mathbf{v}_{proj_W}\|^2$$

, or

$$\|\mathbf{v} - \mathbf{v}_{proj_W}\|^2 \leq \|\mathbf{v} - \mathbf{w}\|^2$$

, a result that will be used later in the problem of least squares. □

3.10   The least-squares problem

Claim: The least-squares problem

In general, let the data consist of a set of n points given by $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$. Here, the x values are exact, and the y values are noisy. We assume that a line of the form

$$y = \beta_0 + \beta_1 x$$

is the best fit to the data. Although we know that the line will not go through all of the data points, we can still write down the resulting equations. We have

$$y_1 = \beta_0 + \beta_1 x_1$$

$$y_2 = \beta_0 + \beta_1 x_2$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_n$$

These equations are a system of n equations in the two unknowns β_0 and β_1. The corresponding matrix equation is given by

$$\begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \tag{3.22}$$

This is an overdetermined system of equations that obviously has no solution. The problem of least-squares is to find the best solution of these equations for β_0 and β_1.

We can generalize this problem as follows. Suppose we are given the matrix equation

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

that has no solution because $\mathbf{b}$ is not in the column space of $\mathbf{A}$. Instead of exactly solving this matrix equation, we want to solve another approximate equation that minimizes the error between $\mathbf{Ax}$ and $\mathbf{b}$. The error can be defined as the norm of $\mathbf{Ax} - \mathbf{b}$, and the square of the error is just the sum of the squares of the components.

3.11 Solution of the least-squares problem

Claim: Solution of the least-squares problem

We want to find the least-squares solution to an overdetermined matrix equation $\mathbf{Ax} = \mathbf{b}$. We write $\mathbf{b} = \mathbf{b}_{proj_{\mathrm{Col}(\mathbf{A})}} + (\mathbf{b} - \mathbf{b}_{proj_{\mathrm{Col}(\mathbf{A})}})$, where $\mathbf{b}_{proj_{\mathrm{Col}(\mathbf{A})}}$ is the projection of b onto the column space of $\mathbf{A}$.

$$\mathbf{b}_{proj_{\mathrm{Col}(\mathbf{A})}} = \mathbf{Ax}$$

Since $(\mathbf{b} - \mathbf{b}_{proj_{\mathrm{Col}(\mathbf{A})}})$ is orthogonal to the column space of $\mathbf{A}$,

$$(\mathbf{b} - \mathbf{b}_{proj_{\mathrm{Col}(\mathbf{A})}}) \perp \mathrm{Col}(\mathbf{A}) \iff (\mathbf{b} - \mathbf{b}_{proj_{\mathrm{Col}(\mathbf{A})}}) \perp \mathrm{Row}(\mathbf{A^T}) \tag{3.23}$$

$$\iff (\mathbf{b} - \mathbf{b}_{proj_{\mathrm{Col}(\mathbf{A})}}) \in \mathrm{Null}(\mathbf{A^T}) \tag{3.24}$$

it is in the nullspace of $\mathbf{A^T}$, that is,

$$\mathbf{A^T}(\mathbf{b} - \mathbf{b}_{proj_{\mathrm{Col}(\mathbf{A})}}) = \mathbf{0}$$

$$\mathbf{A^T}(\mathbf{b} - \mathbf{Ax}) = \mathbf{0}$$

then

$$\mathbf{A^T Ax} = \mathbf{A^T b}$$

It results in a solvable set of equations, called the **normal equations** for Ax = b.

A unique solution to this matrix equation exists when the columns of $\mathbf{A}$ are linearly independent. An interesting formula exists for the matrix which projects $\mathbf{b}$ onto the column space of $\mathbf{A}$. Multiplying the normal equations on the left by $\mathbf{A(A^T A)^{-1}}$, we obtain

$$\mathbf{A(A^T A)^{-1} A^T Ax} = \mathbf{A(A^T A)^{-1} A^T b} \tag{3.25}$$

$$\mathbf{Ax} = \mathbf{A(A^T A)^{-1} A^T b} \tag{3.26}$$

$$\mathbf{b}_{proj_{\mathrm{Col}(\mathbf{A})}} = \mathbf{Ax} \tag{3.27}$$

$$= \mathbf{A(A^T A)^{-1} A^T b} \tag{3.28}$$

Notice that the **projection matrix** $\mathbf{P} = \mathbf{A(A^T A)^{-1} A^T}$ satisfies $\mathbf{P}^2 = \mathbf{P}$, that is, two projections is the same as one.

$$\mathbf{P}^2 = (\mathbf{A(A^T A)^{-1} A^T})(\mathbf{A(A^T A)^{-1} A^T}) \tag{3.29}$$

$$= \mathbf{A} \left[(\mathbf{A^T A})^{-1} (\mathbf{A^T A}) \right] (\mathbf{A^T A})^{-1} \mathbf{A^T} \tag{3.30}$$

$$= \mathbf{A(A^T A)^{-1} A^T} \tag{3.31}$$

$$= \mathbf{P} \tag{3.32}$$

$$\mathbf{P^T} = (\mathbf{A(A^T A)^{-1} A^T})^{\mathbf{T}} \tag{3.33}$$

$$= (\mathbf{A^T})^{\mathbf{T}} \left[(\mathbf{A^T A})^{-1} \right]^{\mathbf{T}} \mathbf{A^T} \tag{3.34}$$

$$= \mathbf{A} \left[(\mathbf{A^T A})^{\mathbf{T}} \right]^{-1} \mathbf{A^T} \tag{3.35}$$

$$= \mathbf{A} \left[\mathbf{A^T (A^T)^T} \right]^{-1} \mathbf{A^T} \tag{3.36}$$

$$= \mathbf{A(A^T A)^{-1} A^T} \tag{3.37}$$

$$= \mathbf{P} \tag{3.38}$$

If $\mathbf{A}$ itself is a square invertible matrix, then $\mathbf{P}^2 = \mathbf{I}$ and $\mathbf{b}$ is already in the column space of $\mathbf{A}$.

Chapter 4

Eigenvalues and Eigenvectors

4.1 Two-by-two and three-by-three determinants

> **Definition 4.1.1: Two-by-two matrix determinant**
>
> If the two-by-two matrix $\mathbf{A}$ is $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then its determinant is
>
> $$\det \mathbf{A} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

> **Definition 4.1.2: Three-by-three matrix determinant**
>
> If the three-by-three matrix $\mathbf{A}$ is $\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$, then its determinant is
>
> $$\det \mathbf{A} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh$$

Corollary 4.1.3

The square matrix $\mathbf{A}_{n \times n}$ is invertible (or nonsingular)

$\Longleftrightarrow \mathbf{A}_{n \times n} \mathbf{A}_{n \times n}^{-1} = \mathbf{A}_{n \times n}^{-1} \mathbf{A}_{n \times n} = \mathbf{I}$

$\Longleftrightarrow \det(\mathbf{A}_{n \times n}) \neq 0$

$\Longleftrightarrow \mathbf{A}_{n \times n}$ is full rank, that is $\operatorname{rank}(\mathbf{A}_{n \times n}) = \dim(\operatorname{Col}(\mathbf{A}_{n \times n})) = \dim(\operatorname{Row}(\mathbf{A}_{n \times n})) = n$

$\Longleftrightarrow \dim(\operatorname{Null}(\mathbf{A}_{n \times n})) = 0$

$\Longleftrightarrow \dim(\operatorname{LeftNull}(\mathbf{A}_{n \times n})) = 0$

$\Longleftrightarrow$ The columns of $\mathbf{A}_{n \times n}$ are linearly independent

$\Longleftrightarrow$ The rows of $\mathbf{A}_{n \times n}$ are linearly independent

$\Longleftrightarrow$ The homogeneous equation $\mathbf{A}_{n \times n}\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{A}_{n \times n}^{-1}(\mathbf{A}_{n \times n}\mathbf{x}) = \mathbf{A}_{n \times n}^{-1}\mathbf{0} = \mathbf{0}$

$\Longleftrightarrow$ The equation $\mathbf{A}_{n \times n}\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = \mathbf{A}_{n \times n}^{-1}(\mathbf{A}_{n \times n}\mathbf{x}) = \mathbf{A}_{n \times n}^{-1}\mathbf{b}$

$\Longleftrightarrow$ The reduced row echelon form of $\mathbf{A}_{n \times n}$ is the identity matrix: $\operatorname{rref}(\mathbf{A}_{n \times n}) = \mathbf{I}_{n \times n}$

$\Longleftrightarrow 0$ is not eigenvalue of $\mathbf{A}_{n \times n}$

Corollary 4.1.4

The square matrix $\mathbf{A}_{n \times n}$ is not invertible (or singular)

$\Longleftrightarrow \det(\mathbf{A}_{n \times n}) = 0$

$\Longleftrightarrow \mathbf{A}_{n \times n}$ is not full rank, that is $\operatorname{rank}(\mathbf{A}_{n \times n}) = \dim(\operatorname{Col}(\mathbf{A}_{n \times n})) = \dim(\operatorname{Row}(\mathbf{A}_{n \times n})) < n$

$\Longleftrightarrow \dim(\operatorname{Null}(\mathbf{A}_{n \times n})) > 0$

$\Longleftrightarrow \dim(\operatorname{LeftNull}(\mathbf{A}_{n \times n})) > 0$

$\Longleftrightarrow$ The columns of $\mathbf{A}_{n \times n}$ are linearly dependent

$\Longleftrightarrow$ The rows of $\mathbf{A}_{n \times n}$ are linearly dependent

$\Longleftrightarrow$ The homogeneous equation $\mathbf{A}_{n \times n}\mathbf{x} = \mathbf{0}$ has non-trivial solution

- i.e., there exists a non-zero vector $\mathbf{v}$ such that $\mathbf{A}_{n \times n}\mathbf{v} = \mathbf{0}$

$\Longleftrightarrow$ The equation $\mathbf{A}_{n \times n}\mathbf{x} = \mathbf{b}$ may have no solution or an infinite number of solutions for some $\mathbf{b}$.

$\Longleftrightarrow$ The reduced row echelon form of $\mathbf{A}_{n \times n}$ has at least one row of zeros

$\Longleftrightarrow \mathbf{A}_{n \times n}$ cannot be reduced to the identity matrix through elementary row operations

$\Longleftrightarrow 0$ is an eigenvalue of $\mathbf{A}_{n \times n}$

- equivalently, the characteristic equation $\det(\mathbf{A}_{n \times n} - \lambda \mathbf{I}) = 0$ has $\lambda = 0$ as a root

Corollary 4.1.5

If we have square matrices $\mathbf{A}_{n \times n}$ and $\mathbf{B}_{n \times n}$, then

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$$

4.2   Laplace expansion

4.3   Leibniz formula

4.4   Properties of a determinant

Claim: Properties of a determinant

The determinant is a function that maps a **square matrix** to a scalar. It is uniquely defined by the following three properties:

Property 1: The determinant of the identity matrix is one;

$$\det \mathbf{I_{n \times n}} = = 1$$

Property 2: The determinant changes sign under row interchange;

Property 3: The determinant is a linear function of the first row, holding all other rows fixed.