Linear Algebra Summary

Leon Muscat

January 25, 2023

Contents

Chapter 1

Introduction to Vectors and Matrices

1.1 Vectors

A vector has n dimensions based on the number of rows. A 2-dimensional vector looks like this:

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

v_1 is called the first component and v_2 the second component (see Figure 1.1). We differentiate between a column vector and a row vector.

Column vector:

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

Row vector:

$$v = \begin{bmatrix} v_1 & v_2 & v_3 & v_4 \end{bmatrix}$$

1.1.1 Operations

We can operate with vectors.

1. Vector addition

   The sum of two vectors v and w is calculated by adding each component of both vectors. Therefore, both vectors have to of the same dimension:

$$v + w = w + v = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \end{bmatrix}$$

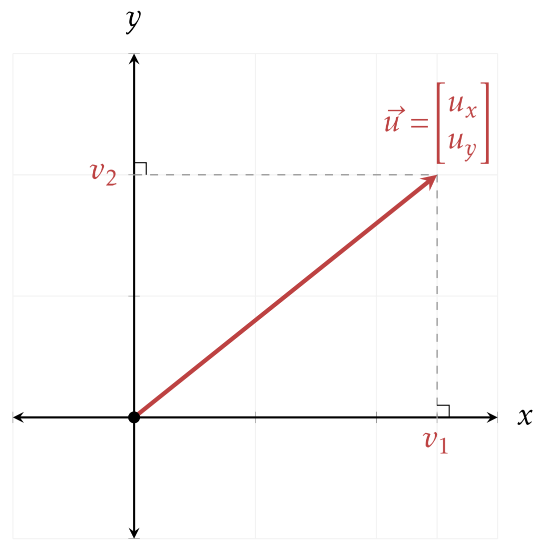


Figure 1.1: Placing a 2-dimensional vector v on the 2-dimensional Cartesian coordinate system, showing its x- and y-components.

2. Scalar multiplication

   Multiplying a vector v by a number c, the scalar, is called scalar multiplication of v by c:

   $$cv = \begin{bmatrix} cv_1 \\ cv_2 \end{bmatrix}$$

3. Linear combination

   Using vector addition and scalar multiplication, we can create linear combinations. The sum of a vector v multiplied by a scalar c and a vector w multiplied by scalar d is a linear combination of v and w: $c * v + d * w$

4. Dot product

   Definition 1.1 Dot Product

   The dot product or inner product of two vectors v and w is calculated as follows: $v \cdot w = \sum_{i=1}^{n} v_i w_i = v_1 w_1 + v_2 w_2 + ... v_n + w_n$.

   π

   Geometrically, the dot product is a product of the length of w and the length of the projection p of v on w. Therefore, if the dot product $\mathbf{x} \cdot \mathbf{y} = 0$, the vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal ($\mathbf{x} \perp \mathbf{y}$).

Note 1.1

The calculation of the dot product given here only applies if the basis of both vectors is orthonormal. If it's not the dot product can be calculated by matrix multiplication.

!

Properties:

i $\mathbf{x} \cdot \mathbf{x} \geq 0$

ii $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ (commutative law)

iii $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$ (distributive law)

iv $(r\mathbf{x}) \cdot \mathbf{y} = r(\mathbf{x} \cdot \mathbf{y})$ (associative law)

1.1.2 Properties

1. Vector length

   A vector with length 0 (all the components are 0) is called a zero vector and a vector with length 1 is called a unit vector.

   The length or norm of a vector v is given by

   $$\|v\| = \sqrt{v \cdot v} = \sqrt{\sum_{i=1}^{n} v_i^2}.$$

   Any vector v can be normalized to a unit vector u by dividing v by its length:

   $$u = \frac{v}{\|v\|}.$$

2. Vector angle

   If the dot product $v \cdot w = 0$, the vectors are orthogonal. This means that v is perpendicular to w and the angle between the two vectors is 90°.

   If $v \cdot w > 0$ then the angle is less than 90°. If $v \cdot w < 0$ then the angle is greater than 90°.

   In general the angle between two vectors v and w can be calculated as follows:

   $$\cos\theta = \frac{v \cdot w}{\|v\| * \|w\|}.$$

   However, there is no normalization needed if v and w are unit vectors. In this case, $\cos\theta = v \cdot w$. How does this work? As explained, the dot product is in part a projection p of v on w. The projection p, the vector

v and the vector u connecting the two form a right triangle. This means we can use trigonometric functions. The definition of cos is

$$\cos = \frac{\text{Adjacent}}{\text{Hypotenuse}}.$$

p is the adjacent and v is the hypotenuse.

3. Schwarz inequality

   According to the Schwarz inequality, it holds

$$|v \cdot w| \leq \|v\| * \|w\|.$$

   Again, let's look at this equation geometrically. The equation says, the product of the length projection of v on w and the length of w is less than the product of the length of v and w. It is obvious that the projection of v on w is shorter than v itself, therefore, the dot product has to be smaller. The only exception is if $v = w$. In this case, the projection of v on w is v, meaning both sides of the inequality are equal.

4. Triangle inequality

   Given two vectors v and w it holds

$$\|v + w\| \leq \|v\| + \|w\|$$

   . If v and w are positive

$$\|v + w\| = \|v\| + \|w\|.$$

1.1.3 Change of coordinates

Within a vector space, every vector $\mathbf{v}$ has a unique representation. For example, lets take a vector $\mathbf{v}$ in the vector space V in $\mathbb{R}^3$ with the standard basis $\mathbf{I}_3$

Note 1.2 Standard basis

The standard basis has all 0 except on the diagonal. That means:

$$\mathbf{I}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

!

To calculate the coordinates of $\mathbf{v}$, we look for the coefficients a_1 to a_n that satisfy the equation

$$\mathbf{v} = a_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + a_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

These coefficients are the coordinates. Because we used the standard basis, these coordinates are called standard coordinates. It is clear that the coordinates are dependent on the basis we choose. If we choose the basis $2 * \mathbf{I}_3$, the coordinates are twice as high. To transfer those non-standard coordinates back to standard coordinates, we need a transition matrix U:

$$\text{standard coordinates} = U \cdot \text{non-standard coordinates}.$$

We can also take the inverse of U, U^{-1} to transfer standard coordinates to non-standard coordinates.

1.2 Matrices

Example 1.1 Matrix

By combining multiple vectors $v = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, w = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, u = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ we can form a matrix:

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

A matrix is characterized by the number of rows (m) and the number of columns (n). For example, an $m \times n$ matrix a over $\mathbb{R}$ is an $m \times n$ rectangular array of real numbers. An individual entry in row i and column j is denotes as a_{ij}.

The $m \times n$ zero matrix $0_{m,n}$ has all of its entries equal to 0. Adding a zero matrix to an $m \times n$ matrix A is equal to A ($0_{m,n} + A = A$).

A square matrix has a main diagonal, a superdiagonal (above the main diagonal) and a subdiagonal (below the main diagonal).

The $n \times n$ identity matrix $\mathbf{I}_n$ has all of its diagonal entries equal to 1 and all other entries equal to 0. In other words, all $a_{ij} = 1$ when $i = j$ and 0 everywhere else. What makes an identity matrix special? Multiplying an identity matrix with an $n \times n$ matrix A equals A:

$$I_n \cdot A = A \cdot I_n = A.$$

A diagonal matrix is similar to an identity matrix, however, all the diagonal entries can have any value.

1.2.1 Operations

We can also operate with matrices.

1. Matrix addition

   The sum of two matrices A and B is calculated by adding each entry of both matrices. Therefore, both matrices have to be of the same dimension:

   $$A + B = B + A = a_{ij} + b_{ij}$$

2. Scalar multiplication

   Multiplying a matrix A by a number c, the scalar, is called scalar multiplication of A by c:

   $$cA = a_{ij} * c \qquad \text{For all } i, j$$

3. Multiplication of a matrix and a vector

   The multiplication of an $m \times n$ matrix A and a vector x is only defined if x has n dimensions. The resulting vector b has the dimension m. The product is calculated as follows: $b_m = A_{mn} \cdot x_n$ or in full form:

   $$Ax = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

4. Multiplication of matrices

   The multiplication of an $m \times n$ matrix A with a matrix B is only defined if B is $n \times p$. In other words, the number of columns of matrix A has to equal the number of rows of matrix B. The resulting matrix C has the dimensions $m \times p$. The product is calculated as follows:

   $$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} + b_{nj}$$

   or in full form:

   $$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{bmatrix}$$

1.2.2 Laws

Matrices satisfy the following laws provided the sums and products are defined:

i $A + B = B + A$

ii $(A + B)C = AC + BC$

iii $C(A + B) = CA + CB$

iv $(\beta A)B = \beta(AB) = A(\beta B)$

v $(AB)C = A(BC)$

1.2.3 Properties

1. Transpose of a matrix

Definition 1.2 Transpose of a matrix

The transpose of a matrix A, denoted as A^T, is obtained by converting the rows of A into the columns of A^T one at a time. If A has order $m \times n$, then A^T has order $n \times m$.

Example 1.2 Transpose

Lets assume

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

then the transpose of A is

$$A^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

2. Rank of a matrix

Definition 1.3 Rank of a matrix

The rank (or row rank) of a matrix is the number of nonzero rows in the matrix after it has been transformed to row-echelon form.

1.2.4 Determinant of a matrix

Definition 1.4 Determinant

The determinant of a square matrix A is a scalar and denoted as $\det(A)$ or $|A|$.

π

The calculation for the determinant is dependent on the size of the matrix.

1. 1×1 matrices $\det(A) = \begin{vmatrix} a_{11} \end{vmatrix} = a_{11}$
2. 2×2 matrices $\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$
3. 3×3 matrices $\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}$

   Easy way to remember:

$$\begin{pmatrix} {\color{red}*} & * & * \\ * & {\color{red}*} & * \\ * & * & {\color{red}*} \end{pmatrix} + \begin{pmatrix} * & {\color{red}*} & * \\ * & * & {\color{red}*} \\ {\color{red}*} & * & * \end{pmatrix} + \begin{pmatrix} * & * & {\color{red}*} \\ {\color{red}*} & * & * \\ * & {\color{red}*} & * \end{pmatrix} - \begin{pmatrix} * & * & {\color{red}*} \\ * & {\color{red}*} & * \\ {\color{red}*} & * & * \end{pmatrix} - \begin{pmatrix} * & {\color{red}*} & * \\ {\color{red}*} & * & * \\ * & * & {\color{red}*} \end{pmatrix} - \begin{pmatrix} {\color{red}*} & * & * \\ * & * & {\color{red}*} \\ * & {\color{red}*} & * \end{pmatrix}$$

What does the determinant represent in $\mathbb{R}^3$? When by performing a linear transformation with $m \times n$ matrix A, the vector x increased in length. This increase in length is equal for all vectors in $\mathbb{R}^3$. The determinant tells us by what amount the length increases. In this example, the det(A) = 6, which means all vectors are scaled by a factor of 6. However, this not only applies to vectors, but everything in the field $\mathbb{R}^3$, including planes, areas, ...

//TODO: Slides 10-29 on Determinants

1.2.5 Inverse of a matrix

Definition 1.5 Inverse of a matrix

A square matrix A is invertible if there exists a matrix A^{-1} such that $A^{-1}A = I$ and $AA^{-1} = I$. The inverse for a square matrix is unique.
A rectangular matrix can have a right inverse ($AB = I$), which is different from its left inverse ($BA = I$).
A matrix that is invertible is called nonsingular and a non-invertible matrix is called singular.

π

1. Rules for inverse matrices

   i If A is nonsingular, then $(A^{-1})^{-1} = A$

ii If A and B are nonsingular, then so is AB. The inverse of a product AB is $(AB)^{-1} = B^{-1}A^{-1}$ (inverses come in reverse order). This also applies to three or more matrices: $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.

iii If A is nonsingular, then so is A^T. Further, $(A^T)^{-1} = (A^{-1})^T$

2. Properties of Inverses of Triangular Matrices

   i The inverse of an upper/lower triangular matrix with nonzero diagonal elements is itself upper/lower triangular.

   ii If any of diagonal elements of lower or upper triangular matrix is zero, then this matrix is non-invertible (singular).

   iii The inverses of triangular matrices are constructed using row reduction, as seen below.

3. Calculating an inverse of a matrix

   i using row reduction

   We reduce A to the row reduced form. Simultaneously, we apply the same row operations to the identity matrix I_n. These operations transform I_n to A–1. This means an invertible matrix is a product of elementary matrices. If a is not invertible, the rank k of the matrix A is smaller than the number of rows of A.

Example 1.3 Calculating the inverse using row reduction

$$\left[\begin{array}{ccc|ccc} 3 & 2 & 1 & 1 & 0 & 0 \\ 4 & 1 & 3 & 0 & 1 & 0 \\ 2 & 1 & 6 & 0 & 0 & 1 \end{array}\right]$$

$$r_1 \rightarrow \frac{1}{3} r_1$$

$$\downarrow$$

$$\left[\begin{array}{ccc|ccc} 1 & \frac{2}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ 4 & 1 & 3 & 0 & 1 & 0 \\ 2 & 1 & 6 & 0 & 0 & 1 \end{array}\right]$$

$$\vdots$$

$$\downarrow$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{-3}{25} & \frac{11}{25} & \frac{-1}{5} \\ 0 & 1 & 0 & \frac{18}{25} & \frac{-16}{25} & \frac{1}{5} \\ 0 & 0 & 1 & \frac{-2}{25} & \frac{-1}{25} & \frac{1}{5} \end{array}\right]$$

ii using the determinante

1.2.6 LU Decomposition

with LU decomposition, we can rewrite an $n \times n$ matrix A as $A = LU$ (in most cases). L is the lower triangular matrix, meaning all elements above the main diagonal are zero. The upper triangular matrix U has all elements below the main diagonal equal 0.

The factorization is unique, if the elements of the diagonal of U are all equal to 1:

$$U = \begin{bmatrix} 1 & u_{1,2} & u_{1,3} & \dots & u_{1,n} \\ 0 & 1 & u_{2,3} & \dots & u_{2,n} \\ 0 & 0 & 1 & \dots & u_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \quad L = \begin{bmatrix} l_{1,1} & 0 & 0 & \dots & 0 \\ l_{2,1} & l_{2,2} & 0 & \dots & 0 \\ l_{3,1} & l_{3,2} & l_{3,3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n,1} & l_{n,2} & l_{n,3} & \dots & l_{n,n} \end{bmatrix}$$

Calculating the LU decomposition using Crout's reduction algorithm

1.2.7 Eigenvectors and Eigenvalues

When applying a linear transformation with an $n \times n$ matrix A on a vector space, the vectors can change in size and orientation. However, there may be some non zero vectors $\mathbf{v}$ that, after applying the linear transformation, are a scalar multiple of the original vector. Those non zero vectors $\mathbf{v}$ are called eigenvectors or characteristic vectors. The mathematical definition is $A\mathbf{v} = \lambda\mathbf{v}$, where $\lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$. The scalar λ is called the eigenvalue or the characteristic value. There exist at most n eigenvalues.

The vector(s) $\mathbf{v}$ can be calculated using $(A - \lambda\mathbf{I})v = 0$ (Proof: $A\mathbf{v} = \lambda\mathbf{v} \implies A\mathbf{v} = (\lambda\mathbf{I})v \implies (A - \lambda\mathbf{I})v = 0$). Hence $\mathbf{v} \in N(A - \lambda\mathbf{I})$.

If $N(A - \lambda\mathbf{I}) \neq \emptyset$, then the space is called the eigenspace of the matrix A corresponding to the eigenvalue λ.

If $N(A - \lambda\mathbf{I}) = \emptyset$, the matrix $A - \lambda\mathbf{I}$ is singular, therefore, $\det(A - \lambda\mathbf{I}) = 0$ (called the characteristic equation).

1.3 Projections

A projection is a linear transformation that maps a vector onto a subspace. This means that the projection of a vector will be a vector that lies in the subspace. To project a vector $\mathbf{b}$, we multiply it with a projection matrix P to give the projection $\mathbf{p}$.

Example 1.4 Projection

Say we want to project a vector $\mathbf{b} = \begin{pmatrix} 2 & 3 & 4 \end{pmatrix}$ onto the xy-plane. To do this we multiply it with $P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. This gives us $\mathbf{p} = \begin{pmatrix} 2 & 3 & 0 \end{pmatrix}$.

To calculate the projection matrix, we first need to find the basis vectors S of the subspace V. Then, we compute P using this formula: $P = (SS^T) \cdot (S^T S)^{-1}$.

However, there is a shortcut for projection a vector $\mathbf{v}$ on a vector $\mathbf{w}$: $\text{proj}_{\mathbf{w}} \mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{w}}{|\mathbf{w}|^2} \mathbf{w}$

1.4 Linear Transformation

A matrix is a linear transformation of a given vector space. An easy way to understand this is to look at a matrix as a collection of vectors. Lets say we have a vector $x = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$, a 3×3 matrix A, which has 3 independent vectors $A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and the standard basis for $\mathbb{R}^3$, which is $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. When thinking of a 3 dimensional vector space, like $\mathbb{R}^3$, we always think of the standard basis (I_3). The position of x in the vector space spanned by I_3 can be calculated by taking each entry and multiplying it with the basis. Since we are using I_n as our basis, this equals to x. However, if we apply the linear transformation with A, the standard basis I_3 gets multiplied with A. The product is the new basis for $\mathbb{R}^3$. If we now calculate the position of x, we get $\begin{pmatrix} 3 & 2 & 1 \end{pmatrix}$.

There are different types of linear transformations:

1. A scaling transformation scales the space as shown in the example above. $S_v = \begin{pmatrix} v_x & 0 \\ 0 & v_y \end{pmatrix}$

2. A rotation transformation rotates space. $R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$. θ is the angle by which space is rotated. The rotation works by rotating the basis by θ

   If θ is positive, the rotation is counter-clockwise. Otherwise, if θ is negative, the rotation is clockwise

   The axis of rotation is the eigenvector.

3. A shear transformation only transforms one dimension. $S = \begin{pmatrix} xx & xy \\ yx & yy \end{pmatrix}$.

For example, to transform the y-axis $S = \begin{pmatrix} 1 & 0.4 \\ 0 & 1 \end{pmatrix}$.

4. A symmetric transformation //TODO

1.4.1 Transformation maps

Such a linear transformation can be denoted as a map $T : \mathbb{R}^n \to \mathbb{R}^n$. Or more generally: T : Original Domain $\to$ Target Co-Domain. This means, a linear transformation connects vector spaces, while keeping the structure of the original vector space. Let $T : U \to V$. The vector in the equation $\mathbf{v} = T(\mathbf{u})$ is called image of $\mathbf{u}$ under T and $\mathbf{u}$ is called the pre-image.

The inverse T^{-1} maps V to U: $T^{-1} : V \to U$. If U and V are finite-dimensional vector spaces and a basis is defined for each vector space, then every linear map $T : U \to V$ can be represented by a transformation matrix A. Because A transforms U to V, A is relative to the basis of U.

A transformation map $T : V \to W$ has a kernel $\ker(T)$ and image $\mathrm{im}(T)$.

The $\mathrm{im}(T)$ is defined to be the set $\{T(\mathbf{v}) : \mathbf{v} \in V\}$. The basis for $\mathrm{im}(T)$ is the set $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_s)\}$

The $\ker(T)$ consists of all vectors $\mathbf{v} \in V$ such that $t(\mathbf{v}) = \mathbf{0}$. That is, $\ker(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}$. The basis can be found by row-reduction of the transformation map T.

The $\dim(V) = n = \dim(\mathrm{im}(T)) + \dim(\ker(T))$ (Proof)

1.4.2 Transformation matrix

The transformation matrix A transforms any vector in the original basis to a vector in the new basis. Lets say $T : U \to V$, $U = \mathbb{R}^n$ and $V = \mathbb{R}^m$. T can be represented as a function $T(\mathbf{x}) = A\mathbf{x}$. This transformation matrix A has the dimensions $m \times n$. A is associated with T with respect to both the standard basis of $\mathbb{R}^n$ and $\mathbb{R}^m$. Because A transforms U to V, A is relative to the basis of U.

Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the basis of U. Any vector $\mathbf{u}$ with elements $\mathbf{u}_1, \ldots, \mathbf{u}_n$ can be represented in U as $\mathbf{u} = \alpha_1\mathbf{e}_1 + \cdots + \alpha_n\mathbf{e}_n$. To transform $\mathbf{u}$ to V we put in $\mathbf{u}$ in $T(x)$:

$$T(\mathbf{u}) = T(\alpha_1\mathbf{e}_1 + \cdots + \alpha_n\mathbf{e}_n)$$

Based on rule (i) and (ii) of linear transformations, we can rewrite this:

$$T(\alpha_1\mathbf{e}_1 + \cdots + \alpha_n\mathbf{e}_n) = \alpha_1 T(\mathbf{e}_1) + \cdots + \alpha_n T(\mathbf{e}_n)$$

This means, we are only really interested in the basis, in order to perform a transformation. Lets say the basis for V is $\mathbf{v}_1 + \cdots + \mathbf{v}_m$. Every $T(\mathbf{e}_j)$ can be uniquely represented with $\mathbf{v}_1 + \cdots + \mathbf{v}_m$:

$$T(\mathbf{e}_1) = \beta_{11}\mathbf{v}_1 + \beta_{21}\mathbf{v}_2 + \cdots + \beta_{m1}\mathbf{v}_m$$

$$T(\mathbf{e}_2) = \beta_{12}\mathbf{v}_1 + \beta_{22}\mathbf{v}_2 + \cdots + \beta_{m2}\mathbf{v}_m$$
$$\vdots$$
$$T(\mathbf{e}_n) = \beta_{1n}\mathbf{v}_1 + \beta_{2n}\mathbf{v}_2 + \cdots + \beta_{mn}\mathbf{v}_m$$

A more compact notation is:

$$T(\mathbf{e}_j) = \sum_{i=1}^{m} \alpha_{ji}\mathbf{v}_i \text{for} 1 \leq j \leq n$$

The coefficients $\alpha_{ij} \in K(1 \leq i \leq m$ and $1 \leq j \leq n)$ can be represented as an $m \times n$ matrix:

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \dots & \alpha_{mn}+ \end{pmatrix}$$

The calculation for $T(\mathbf{e}_j)$ can now be rewritten:

$$T(\mathbf{e}_1) = A\mathbf{e}_1 = \begin{pmatrix} a_{11} \\ a_{12} \\ \dots \\ a_{1n} \end{pmatrix}$$

$$T(\mathbf{e}_2) = A\mathbf{e}_2 = \begin{pmatrix} a_{21} \\ a_{22} \\ \dots \\ a_{2n} \end{pmatrix}$$

$$\vdots$$

$$T(\mathbf{e}_m) = A\mathbf{e}_m = \begin{pmatrix} a_{m1} \\ a_{m2} \\ \dots \\ a_{mn} \end{pmatrix}$$

Example 1.5

1. Example Let the basis vectors of U be $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ be $\mathbf{I}_3$ and the basis of V be $\mathbf{v}_1, \mathbf{v}_2$ be $(\mathbf{I}_2)$. Let $T : U \to V$. To calculate A we write down T:

$$T(\mathbf{u}_1) = \alpha_{11}\mathbf{v}_1 + \alpha_{21}\mathbf{v}_2$$
$$T(\mathbf{u}_2) = \alpha_{12}\mathbf{v}_1 + \alpha_{22}\mathbf{v}_2$$
$$T(\mathbf{u}_3) = \alpha_{13}\mathbf{v}_1 + \alpha_{23}\mathbf{v}_2$$

Then A is:

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

2. Example

However, there is a shorter way to calculate the transformation matrix A. Given the transformation $T : U \to V$, the basis for $U = \mathbf{u}_1, \dots, \mathbf{u}_n$ and the basis of $V = \mathbf{v}_1, \dots \mathbf{v}_n$. The row-reduced form of $\mathbf{v}_1, \dots \mathbf{v}_n | L(\mathbf{u}_1), \dots, L(\mathbf{u}_n)$ is $(\mathbf{I}|A)$.

1.4.3 Rules for linear transformations

The following rules have to be true for a linear transformation:

i $T(\mathbf{u}_1 + \mathbf{u}_2) = T(\mathbf{u}_1) + T(\mathbf{u}_2)$

ii $T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$

These rules can be combined: $T(\alpha\mathbf{u}_1 + \beta\mathbf{u}_2) = \alpha T(\mathbf{u}_1) + \beta T(\mathbf{u}_2)$

What follows from this definition ($T : U \to V$):

i $T(\mathbf{0}_U) = \mathbf{0}_V$

ii $T(-\mathbf{u}) = -T(\mathbf{u})$

1.4.4 Isomorphism and endomorphism

If $U = V$ we call T endomorphism or a linear operator.

If the function $T(\mathbf{x})$ is a bijection, the transformation is called isomorphism (There can be more than one isomorphism for two vector spaces). U and V are then called isomorphic and share many properties.

Example 1.6

Let V be a vector space over K^n with the basis $\mathbf{v}_1, \dots, \mathbf{v}_n$. The linear map $T : K^n \to V$ is then defined as $T(\alpha_1, \dots, \alpha_2) = \alpha_1\mathbf{v}_1 + \dots + \alpha_n\mathbf{v}_n$. This is an isomorphism

Let $T : U \to V$ be an isomorphism and $\mathbf{u}_1, \dots, \mathbf{u}_n$ be vectors in U.

1. The vectors are linearly independent if and only if $T(\mathbf{u}_1), \dots, T(\mathbf{u}_n)$ is linearly independent.
2. The vectors span U if and only if $T(\mathbf{u}_1), \dots, T(\mathbf{u}_n)$ spans V.
3. The vector are a basis of U if and only if $T(\mathbf{u}_1), \dots, T(\mathbf{u}_n)$ is a basis for V

Chapter 2

Vector Spaces and Subspaces

2.1 Fields

A field is a set of numbers K and two maps $K \times K \to K$, which define addition and multiplication within that field. Every field must contain 0 and 1 and satisfy the 8 axioms below.

Example 2.1 Field

For each number p, define $F_p = 0, 1, 2, ..., p-1$ with p elements, where addition and multiplication are carried out modulo p.
The smallest such field is F_2 because it just contains 0 and 1.

2.1.1 Axioms

Axioms for addition:

A1 $\alpha + \beta = \beta + \alpha$ for all $\alpha, \beta \in K$ (commutative law)

A2 $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$ for all $\alpha, \beta, \gamma \in K$ (associative law)

A3 $\alpha + 0 = 0 + \alpha = \alpha$ for all $\alpha \in K$

A4 For each element $\alpha \in K$ there exists an element $-\alpha \in K$ such that $\alpha + (-\alpha) = (-\alpha) + \alpha = 0$

Axioms for multiplication:

M1 $\alpha \cdot \beta = \beta \cdot \alpha$ for all $\alpha, \beta \in K$ (commutative law)

M2 $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$ for all $\alpha, \beta, \gamma \in K$ (associative law)

M3 $\alpha \cdot 1 = 1 \cdot \alpha = \alpha$ for all $\alpha \in K$

M4 For each element $\alpha \in K$ with $a \neq 0$, there exists an element $\alpha^{-1} \in K$ such that $\alpha \cdot \alpha^{-1} = \alpha^{-1} \cdot \alpha = 1$

Axioms for multiplication and addition:

D $(\alpha + \beta) \cdot \gamma = \alpha \cdot \gamma + \beta \cdot \gamma$ for all $\alpha, \beta, \gamma \in K$ (distributive law)

2.1.2 Number systems as fields

Why are the sets $\mathbb{R}$, $\mathbb{Q}$ and $\mathbb{C}$ more interesting for us than $\mathbb{N}$? In $\mathbb{N}$, A1-A3 and M1-M3 hold, but A4 and M4 do not. In $\mathbb{Z}$, A1-A4 and M1-M3 hold, but M4 does not. On the other hand, $\mathbb{R}$, $\mathbb{Q}$ and $\mathbb{C}$ satisfy all axioms and are therefore fields.

2.2 Vector spaces

A vector space over a field K is a set V together with an element $0 \in K$ and two maps $+ : V \times V \to V$ and $\cdot : K \times V \to V$, called addition and scalar multiplication. These maps must satisfy 5 axioms and define $u+v \in V (u, v \in V)$ and $\alpha \cdot v \in V (a \in K)$. If a vector space V is spanned by a finite number of vectors, then it has a basis. A linear combination of all the vectors of a basis, can describe every vector in V. Any n linearly independent vectors in an n-dimensional vector subspace (like $\mathbb{R}^n$) will span that vector space.

//TODO: Extending and reducing linear independence to a spanning set

Elements $v \in V$ of a vector space V are called vectors. Elements of the field K will be called scalars. Both the zero vector and zero scalar are in the V and K.

Matrices are also a vector over a field $K^{(m,n)}$. This can also be denoted as $\mathbb{M}_{m,n}(K)$

Example 2.2 Vector space

Field F: For each prime number p, define $F_p = 0, 1, 2, ..., p-1$ with p elements, where addition and multiplication are carried out modulo p. A vector in F_2^8 is a byte. The 2 means we only have the numbers 0 and 1, while the 8 means the vector space has 8 dimensions.

2.2.1 axioms

For V to be called a vector space, the following axioms must be satisfied for all $\alpha, \beta \in K$ and all $u, v, w \in V$.

i Vector addition satisfies these axioms:

A1 $u + v = v + u$

A2 $(u + v) + w = u + (v + w)$

A3 $v + 0 = 0 + v = v$

A4 There exists $-v \in V$ such that $v + (-v) = (-v) + v = 0$

ii $\alpha u + v = \alpha u + \alpha v$

iii $(\alpha + \beta)v = \alpha u + \alpha v$

iv $(\alpha\beta)v = \alpha(\beta v)$

v $v \cdot 1 = v$

From these axioms we can deduce some useful relations:

i $\alpha \cdot 0 = 0$ for all $\alpha \in K$

ii $0 \cdot v = 0$ for all $v \in V$

iii $-(\alpha v) = (-\alpha)v = \alpha(-v)$ for all $\alpha \in K$ and $v \in V$

iv if $\alpha v = 0$, then $\alpha = 0$ or $v = 0$

2.2.2 Special vector spaces

The vector space V that consists only of a zero vector satisfies all axioms.

The infinite-dimensional space $\mathbb{R}^\infty$) also satisfies all axioms. Its vectors have infinitely many components, as in $u = \begin{pmatrix} 0 & 2 & 1 & 2 & \dots \end{pmatrix}$. The laws for $u+v$ and cu stay unchanged.

2.2.3 Product of two vector spaces

//TODO Product of two vector spaces

2.3 Vector Subspaces

Let V be a vector space over the field K. Certain subsets of V have are closed under addition and scalar multiplication. This means that when adding or taking scalar multiples of vectors in the subset gives which are again in the subset. Such a subset is called a subspace. In other words, if v and w are vectors in a subspace W and c is a scalar in K, then $v + w \in W$ and $cv \in W$. if V is a finite dimensional vector space, then every subspace W of V is also finite dimensional. In addition, the dimension of W is not larger than the dimension of V.

Any vector space V is a subspace of itself. Subspaces other than V are called proper subspaces. The zero subspace is called the trivial subspace.

Example 2.3 Vector subspaces

1. Any line through $\begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$
2. Any plane through $\begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$
3. The whole space of $\mathbb{R}^3$
4. The single vector $\begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$

Inside the vector space $\mathbb{R}^{2,2}$ (or $\mathbb{M}_{2,2}(\mathbb{R})$) there are two subspaces:

1. All upper triangular matrices
2. All diagonal matrices

A vector subspace of a vector space V can be formed by taking the linear combination of the vectors $(v_1, ..., v_n) \in V$. We say the subspace is spanned by the vectors $v_1, ..., v_n$.

Additionally, a vector subspace V with a spanning set S can have an orthogonal complement. The orthogonal complement of a subspace contains every vector in V that is perpendicular to S. This orthogonal subspace is denoted by $S^\perp$ (S perp). Any vector $\mathbf{x}$ that is perpendicular to S is perpendicular to the subspace V ($\mathbf{x} \perp S \implies \mathbf{x} \perp V$). Also, $(S^\perp)^\perp = \text{span}(S)$ and $(V^\perp)^\perp = V$. S and $S^\perp$ can only share the zero vector.

2.3.1 4 fundamental subspaces

The four fundamental subspaces of an $m \times n$ matrix A are the

1. row space $C(A^T)$ with dim r (r refers to the rank of A)

   The row space is spanned by the rows of A. The C in $C(A^T)$ stands for columns. If we take the columns of the transposed matrix, we get the rows. The row space is a subspace of \$^{n} and is also called a coimage. The dimension of the row space is denoted as r or rank(A)

2. column space $C(A)$ with dim r

   The column space is spanned by the columns of A. It is a subspace of $\mathbb{R}^m$

3. null space $N(A)$ with dim $n - r$

   The null space is the set of solutions for the equation $Ax = 0$. The null space meets the row space only at the 0 vector. This represents the trivial solution for matrix A. The null space is a subspace of $\mathbb{R}^n$ and also called a kernel. The dimension of the null space is called nullity.

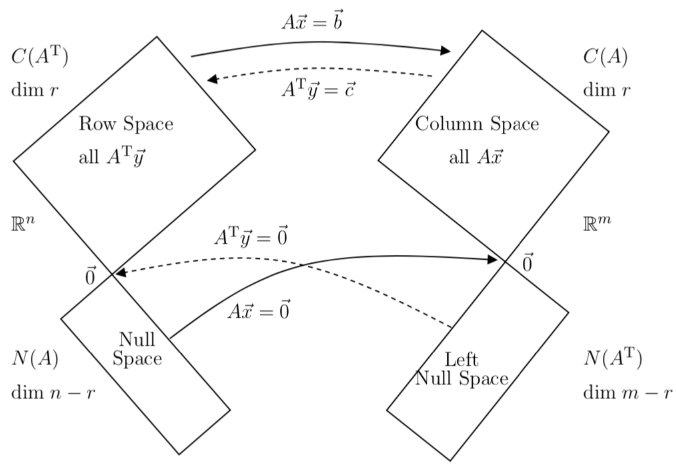
   The null space is orthogonal and linearly independent to the row space. But why? For the equation $Ax = 0$ to be true, the dot product of x and each row in A has to equal 0 (otherwise the solution of $Ax \neq 0$). This means x is orthogonal to each row in A and, therefore, orthogonal to the row space of A. Because of this, the x also has to be linearly independent to the row space of A.

4. left null space $N(A^T)$ with dim m-r

   Similarly to the null space, the left null space is orthogonal to the column space and only meets the columns space at the 0 vector. The left null space is the space spanned by the vectors y that satisfy the equation $A^T y = 0$. The left null space is a subspace of $\mathbb{R}^m$ and also called a cokernel.

The row space and null space a linearly independent and their dimensions add up to R^n. So, together they span $\mathbb{R}^n$. Same with the column space and left null space. They are also linearly independent, but their dimensions add up to R^m. Together they span R^m.

All of this is represented in the following image:



2.3.2 Orthogonal bases and Gram-Schmidt

To understand what an orthogonal basis is, we first need to understand what an orthogonal basis is. A nonempty set $S \in V$ of nonzero vectors is called an orthogonal set if all vectors in S are mutually orthogonal. That is, the dot product of any two vectors $v_k, v_i \in S$ has to be $= 0$. Additionally, all vectors of an orthogonal set are linearly independent.

Furthermore, if every vector $v \in S$ has a length of 1, the set S is called an orthonormal set.

There is an easy way to check if a set S is orthogonal. Create a matrix A with every vector $\mathbf{v} \in S$ and perform a row reduction. If the resulting matrix is an identity matrix, the set S is orthogonal.

To calculate the coordinates of any vector $\mathbf{x} \in V$ with an orthogonal basis, we can do the following calculation:

$$\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{v}_1}{v_1^2} v_1 + \cdots + \frac{\mathbf{x} \cdot \mathbf{v}_n}{v_n^2} v_n$$

If the basis is orthonormal, the normalization is not needed:

$$\mathbf{x} = (\mathbf{x} \cdot \mathbf{v}_1) v_1 + \cdots + (\mathbf{x} \cdot \mathbf{v}_n) v_n$$

To calculate an orthogonal basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ with a non-orthogonal basis $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ for V, we can use the Gram-Schmidt orthogonalization process:

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1^2} \mathbf{v}_1 \\
\mathbf{v}_n &= \mathbf{x}_n - \sum_{i=2}^{n} \frac{\mathbf{x}_i \cdot \mathbf{v}_{i-1}}{\mathbf{v}_{i-1}^2} \mathbf{v}_{i-1}
\end{aligned}$$

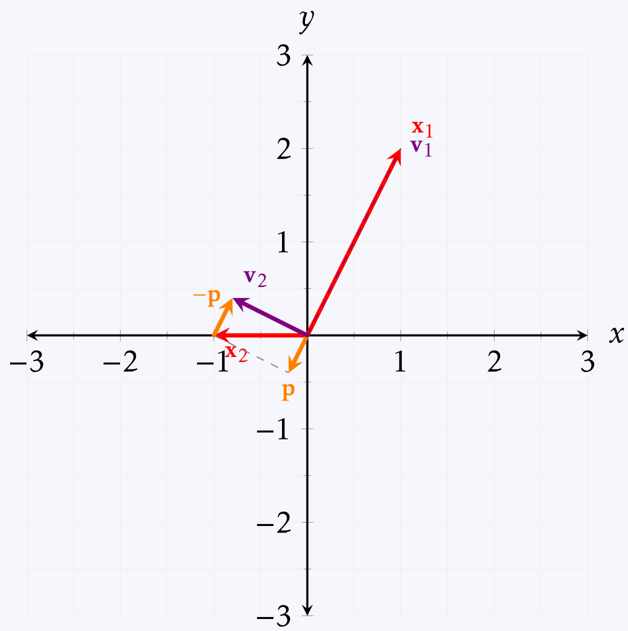
Then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is an orthogonal basis for V.

Properties of the Gram-Schmidt process:

1. The span of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ and $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is the same.
2. $\mathbf{v}_k$ is orthogonal to $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}$
3. $\mathbf{v}_k = \mathbf{x}_k - \mathbf{p}_k$, where $\mathbf{p}_k$ is the orthogonal projection of the vector $\mathbf{x}_k$ on the subspace spanned by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}$
4. $\|\mathbf{v}_k\|$ is the distance from $\mathbf{x}_k$ to the subspace spanned by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}$

Example 2.4

Let $x_1 = \begin{bmatrix} 1 & 2 \end{bmatrix}, x_2 = \begin{bmatrix} -1 & 0 \end{bmatrix}$ span $\mathbb{R}^2$



The calculation can be simplified:

$$\mathbf{w}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$$

$$\mathbf{v}_2 = \mathbf{x}_2 - (\mathbf{x}_2 \cdot \mathbf{w}_1)\mathbf{w}_1$$

$$\mathbf{v}_n = \mathbf{x}_n - (\mathbf{x}_n \cdot \mathbf{w}_1)\mathbf{w}_1$$

Then $\mathbf{w}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis for V, $\|\mathbf{w}_1\| = 1$, and $\mathbf{w}_1$ is orthogonal to $\mathbf{v}_2, \ldots, \mathbf{v}_n$.

Gram-Schmidt can be used to check linear independence:

The vectors $x_1, \ldots, x_n$ stay the same, if they are already linearly independent

//TODO

Chapter 3

Systems of linear equations

A system of linear equations $A\mathbf{x} = \mathbf{b}$ has m equations and n unknowns:

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m$$

If $\mathbf{b} = \mathbf{0}$, the system is homogeneous. Otherwise, the system is nonhomogeneous.

A system can have 1. no solution, 2. exactly one solution, 3. more than one solution. If a system has at least one solution, it is consistent. Otherwise, it is inconsistent. If a solution is $\mathbf{x} = \mathbf{0}$, it is called a trivial solution. Otherwise, it is a nontrivial solution. A system with no solutions or infinitely many solutions is singular.

A system $A\mathbf{x} = \mathbf{b}$ can also be written as $(A|\mathbf{b})$:

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right]$$

3.1 Elementary operations

All elementary row and column operations are the same. All operations can be achieved by left multiplying a corresponding elementary matrix (linear transformation).

R1 Add a multiple of $\mathbf{r}_j$ to $\mathbf{r}_i$ ($i \neq j$). The elementary matrix is $E(n)^1_{\lambda,i,j}$. It is an identity matrix, which has a non-zero entry λ at the position (i,j).

R2 Interchange two rows. The elementary matrix is $E(n)^2_{i,j}$. It is an identity matrix with rows i and j interchanged.

R3 Multiplying a row by a non-zero scalar. The elementary matrix is $E(n)^3_{\lambda,i}$. It is an identity matrix with the position (i,i) replaced by λ.

3.2 Solving linear equations

3.2.1 row reduced form / upper echelon form

In general, to solve a system of linear equations using row reduction, we can use the following steps:

1. Write the system of equations in augmented matrix form $(A|b)$.
2. Put the matrix in upper-echelon form by performing row operations on the matrix.
3. Put the matrix in row-reduced form by performing further row operations on the matrix.
4. Read off the solutions from the row reduced form of the matrix.

Example 3.1

Given the following equation:

$$\begin{aligned} x + y &= 2 \\ 3x - y &= 5 \end{aligned}$$

We can solve this equation using row-reduction like this:

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 3 & -1 & 5 \end{array}\right]$$
$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & -4 & -1 \end{array}\right]$$
$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & 1/4 \end{array}\right] = \text{Upper-echelon form}$$
$$\left[\begin{array}{cc|c} 1 & 0 & 7/4 \\ 0 & 1 & 1/4 \end{array}\right] = \text{Row-reduced form}$$
$$\Rightarrow (x,y) = (\frac{7}{4}, \frac{1}{4})$$

3.2.2 Solving using the inverse

The equation system $A\mathbf{x} = \mathbf{b}$ has a unique solution if and only if A is nonsingular (=invertible).

The homogeneous system of equations $A\mathbf{x} = 0$ has a non-zero solution if and only if A is singular. Otherwise, it has only a trivial solution.

To solve an equation system using A^{-1}, we can use the following equation:

$$\mathbf{x} = \mathbf{b}A^{-1} = A^{-1}\mathbf{b}$$

3.2.3 Solving using LU Decomposition

To solve the system of equations using LU decomposition, we can first rearrange the original equation as $LUx = b$, where x is the vector of variables $x_1, x_2, \ldots, x_n$. Since L is a lower triangular matrix, we can solve for the vector y in the equation $Ly = b$ by performing row-reduction. Once we have solved for y, we can then solve for the vector x in the equation $Ux = y$ by again performing row-reduction.