Preliminaries: Topics on Linear Algebra

Pedro Vilanova

January 25, 2024

Contents

Point in $\mathbb{R}^n$

A point in $\mathbb{R}^n$ is a vector $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ where each x_i is a real number, for $i = 1, .., n$.

Note: In Euclidean space, a vector is a geometric object that possesses a magnitude and a direction. A vector can be pictured as an arrow. Its magnitude is its length, and its direction is the direction to which the arrow points.

Vector addition and scalar multiplication are defined for $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$:

$$x + y = (x_1 + y_1, ..., x_n + y_n)^\top,$$
$$\alpha x = (\alpha x_1, ..., \alpha x_n)^\top.$$

Point in $\mathbb{R}^n$

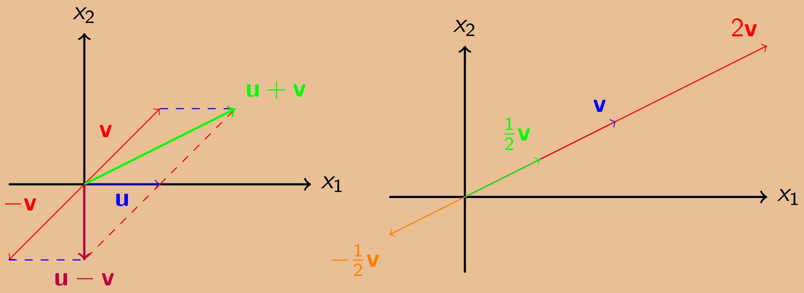


Figure: Vector Addition and Scalar Multiplication in $\mathbb{R}^2$.

Point in $\mathbb{R}^n$

Given any $x, y \in \mathbb{R}^n$ we write

$$\begin{array}{lll} x = y & \text{if} & x_i = y_i \quad \text{for all } i \\ x \geq y & \text{if} & x_i \geq y_i \quad \text{for all } i \\ x > y & \text{if} & x \geq y \text{ and } x \neq y\,. \end{array}$$

Note: x and y may not be comparable under any of the categories above (example: $x = (2, 1)$ and $y = (1, 2)$ are not $=$ nor $x \geq y$ nor $y \geq x$). This is because $\mathbb{R}^n$ for $n > 1$ is not a total order. $\mathbb{R}$ is a total order.

Remark: <u>The space $\mathbb{R}^n$ forms a vector space</u>. In other words, it satisfies commutativity, associativity, and distributive properties, and it has an additive identity, an additive inverse, and a multiplicative identity.

A vector space is usually denoted by V. Functions $f : \mathbb{R}^n \to \mathbb{R}$ also form a vector space.

Subspaces of $\mathbb{R}^n$

Definition (closed under addition)

A set of vectors V is said to be closed under addition if

$$u, v \in V \implies u+v \in V\,, \quad \forall x, v \in V\,.$$

Definition (closed under scalar multiplication)

A set of vectors V is said to be closed under scalar multiplication if

$$u \in V \implies \alpha u \in V\,, \quad \forall x \in V\,,\ \forall \alpha \in \mathbb{R}\,.$$

Proposition

Let $V \subseteq \mathbb{R}^n$, $V \neq \emptyset$, closed under addition and closed under scalar multiplication. Then V is a subspace of $\mathbb{R}^n$.

All possible subspaces in $\mathbb{R}$? A: $\{0\}$ and $\mathbb{R}$. All in $\mathbb{R}^2$? A: $\{0\}$, any line in $\mathbb{R}^2$ through the origin, and $\mathbb{R}^2$. All in $\mathbb{R}^3$?

Note: If V is a subspace of $\mathbb{R}^n$, then the zero vector 0 is in V.

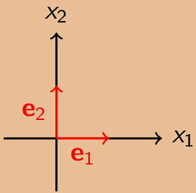
Standard unit vectors

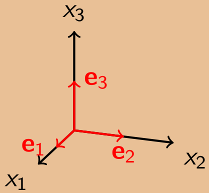
A unit vector is a vector of length 1 (norm 1). A unit vector in the positive direction of a coordinate axis is called a standard unit vector. There are two standard unit vectors in $\mathbb{R}^2$:

- The vector $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is parallel to the x-axis, and
- The vector $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is parallel to the y-axis.

There are three standard unit vectors in $\mathbb{R}^3$:

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





Standard unit vectors

Every vector in $\mathbb{R}^3$ can be written as a sum of scalar multiples of e_1, e_2, and e_3. For example,

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

can be written

$$v = \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$$

$$= 3\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 2\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + (-1)\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 3e_1 + 2e_2 - e_3.$$

The expression $3e_1 + 2e_2 - e_3$ is called a linear combination of e_1, e_2, and e_3.

Question: Is there any way to write any vector in $\mathbb{R}^3$ using less than 3 vectors? Answer is no.

This can be easily generalized to $\mathbb{R}^n$.

Linear combination and Span

Definition (Linear combination)

A linear combination of a set of vectors $\{v_1, \ldots, v_m\}$ *in a subspace of* $\mathbb{R}^n$*, is a vector of the form*

$$\alpha_1 v_1 + \cdots + \alpha_m v_m,$$

where $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$.

Examples: in $\mathbb{R}^3$

- $(17, -4, 2)^\top$ is a linear combination of $(2, 1, -3)^\top, (1, -2, 4)^\top$ because

$$(17, -4, 2)^\top = 6(2, 1, -3)^\top + 5(1, -2, 4)^\top.$$

Linear combination and Span

- $(17, -4, 5)^\top$ is not a linear combination of $(2, 1, -3)^\top, (1, -2, 4)^\top$ because there do not exist scalars α_1, α_2 such that

$$(17, -4, 5)^\top = \alpha_1 (2, 1, -3)^\top + \alpha_2 (1, -2, 4)^\top .$$

In other words, the system of equations

$$\begin{aligned} 17 &= 2\alpha_1 + \alpha_2 \\ -4 &= \alpha_1 - 2\alpha_2 \\ 5 &= -3\alpha_1 + 4\alpha_2 \end{aligned}$$

has no solutions (verify!).

Note: To "quickly" check that a set of n vectors in $\mathbb{R}^n$ are $\ell.\imath.$ compute the determinant of the matrix formed by placing the vectors as columns. If determinant is not zero then vectors are $\ell.\imath.$. For example, check that $(1, 4)^\top$, $(7, 5)^\top$ are $\ell.\imath.$.

Linear combination and Span

Definition (Span)

The set of all linear combinations of a set of vectors $\{v_1, \ldots, v_m\}$*, is called the span and denoted* $\text{Span}\,(v_1, \ldots, v_m)$*. In other words,*

$$\text{span}\,(v_1, \ldots, v_m) = \{a_1 v_1 + \cdots + a_m v_m : \alpha_1, \ldots, \alpha_m \in \mathbb{R}\}\ .$$

The span of the empty set $\{\}$ *(also denoted* $\emptyset$*) is defined to be* $\{0\}$*.*

Example: The previous example shows that

- $(17, -4, 2)^\top \in \text{Span}((2, 1, -3)^\top, (1, -2, 4)^\top)$ since it is a linear combination;
- $(17, -4, 5)^\top \notin \text{Span}((2, 1, -3)^\top, (1, -2, 4)^\top)$.
- Describe $\text{Span}((-3, 1)^\top)$. Only one vector. The span is the set of all vectors of the form $v = a(-3, 1)^\top$, for $a \in \mathbb{R}$.
- Describe $\text{Span}((2, 2)^\top, (-1, 0)^\top)$. Vectors are not collinear, so intuitively it makes sense that the two vectors span $\mathbb{R}^2$.

Linear combination and Span

Formally, consider the vector equation:

$$\begin{pmatrix} 2 & -1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} s \\ t \end{pmatrix}$$

This corresponds to the system:

$$\begin{aligned} 2a_1 - a_2 &= s \\ 2a_1 &= t\,. \end{aligned}$$

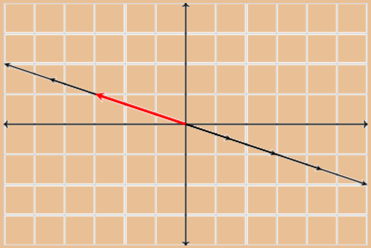
Applying elementary row operations we reach to the equivalent system (reduced row-echelon form):

$$\left(\begin{array}{cc|c} 2 & -1 & s \\ 2 & 0 & t \end{array}\right) \rightsquigarrow_{R_1 \leftrightarrow R_2;\, \frac{1}{2}R_1;\, R_2 - 2R_1;\, -R_2} \rightsquigarrow \left(\begin{array}{cc|c} 1 & 0 & t/2 \\ 0 & 1 & t-s \end{array}\right).$$

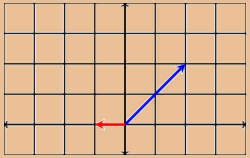
Linear combination and Span

This shows that every vector of $\mathbb{R}^2$ can be written as a linear combination of $(2,2)^\top$ and $(-1,0)^\top$:

$$\frac{t}{2}\begin{pmatrix}2\\2\end{pmatrix} + (t-s)\begin{pmatrix}-1\\0\end{pmatrix} = \begin{pmatrix}s\\t\end{pmatrix}.$$



$\text{Span}((-3,1)^\top)$

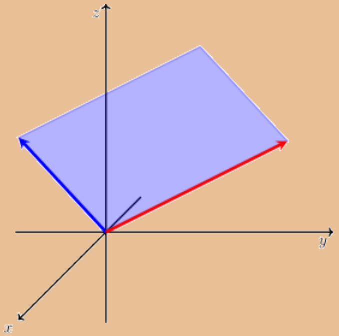


$\text{Span}((2,2)^\top, (-1,0)^\top)$

Note: In $\mathbb{R}^2$, two non-zero vectors can depict a line or the entire plane $\mathbb{R}^2$.

Linear combination and Span

- Describe $\text{Span}((5,0,4)^\top,(0,4,2)^\top)$. Vectors are not collinear and thus define a plane in $\mathbb{R}^3$.



$\text{Span}((5,0,4)^\top,(0,4,2)^\top)$, partially shown. The plane extends infinitely.

Span is a subspace

Proposition

Let S be any set of vectors in $\mathbb{R}^n$. Then $\text{Span}\,(S)$ *is a subspace of* $\mathbb{R}^n$*.*

Since span contains all possible linear combinations it is closed under addition and scalar multiplication.

Now the question is: We know to write any vector in $\mathbb{R}^n$ we need n standard unit vectors, but what about a particular subspace of $\mathbb{R}^n$? The next definition (linear independence) is essential to answer the question.

Linear independence ($\ell.\imath.$)

Given a set of vectors, is there a combination that gives the 0 vector?

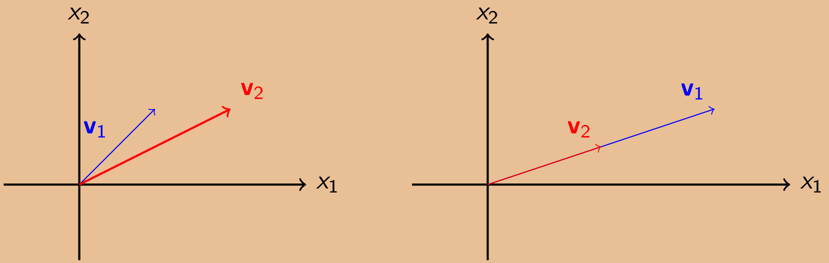
Definition (Linear independence)

A set of vectors $\{v_1, ..., v_k\}$ is said to be linearly independent ($\ell.\imath.$) if the equality

$$\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_k v_k = 0$$

implies that $\alpha_i = 0$ for all i.

A set of the vectors $\{v_1, ..., v_k\}$ is linearly dependent if it is not $\ell.\imath.$.



Left: Linear independent set $\{v_1, v_2\}$. Right: Linear dependent. ($\mathbb{R}^2$)

Linear independence ($\ell.\imath.$)

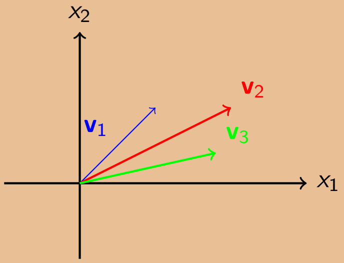
Examples:

- The set composed of the single vector 0 is linearly dependent, for if $\alpha \neq 0$ then $\alpha 0 = 0$. In fact, any set of vectors containing the vector 0 is linearly dependent.
- A set composed of a single nonzero vector $a \neq 0$ is $\ell.\imath.$ since $\alpha a = 0$ implies $\alpha = 0$. Note a singleton set of vectors maybe $\ell.\imath.$ or not.
- Show that $(3, 1)^\top$ and $(1, 2)^\top$ are $\ell.\imath.$. Illustrate.
- The vectors $(1, 0, 0)^\top$, $(0, 1, 0)^\top$, $(0, 0, 1)^\top$ is $\ell.\imath.$ in $\mathbb{R}^3$. This is an important set because forms a basis in $\mathbb{R}^3$ called the canonical basis.

Linear independence ($\ell.\imath.$)

Examples of linearly dependent sets:

- $\{(2,3,1)^\top, (1,-1,2)^\top, (7,3,8)^\top\}$ is linearly dependent in $\mathbb{R}^3$ because $2(2,3,1)^\top + 3(1,-1,2)^\top + (-1)(7,3,8)^\top = (0,0,0)^\top$.
- $\{(2,3,1)^\top, (1,-1,2)^\top, (7,3,c)^\top\}$ is linearly dependent in $\mathbb{R}^3$ iff $c = 8$ (verify!).



Linear dependent set $\{v_1, v_2, v_3\}$. Recall a set of vectors in $\mathbb{R}^n$ can have at most n independent vectors.

Examples

Proposition

A set of vectors $\{v_1, ..., v_k\}$ is not $\ell.\imath.$ iff one of the vectors is a linear combination of the remaining vectors.

- $\{(2,3,1)^\top, (1,-1,2)^\top, (7,3,8)^\top\}$ is linearly dependent in $\mathbb{R}^3$ because $2(2,3,1)^\top + 3(1,-1,2)^\top = (7,3,8)^\top$.

Proposition

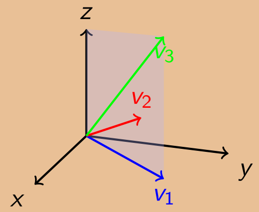
Let $\{u_1, .., u_m\}$ be $\ell.\imath.$ in V. Suppose also that $\{w_1, .., w_n\}$ spans V. Then, we have that $m \leq n$.

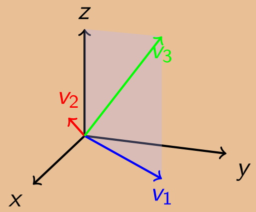
- $\{(1,2,3)^\top, (4,5,8)^\top, (9,6,7)^\top, (-3,2,8)^\top\}$ cannot be $\ell.\imath.$ in $\mathbb{R}^3$ because we know $\{(1,0,0)^\top, (0,1,0)^\top, (0,0,1)^\top\}$ spans $\mathbb{R}^3$.

Illustration: In $\mathbb{R}^3$, let v_1 and v_2 be two non-parallel vectors starting at the origin. If t_1 and t_2 are real numbers, then the vector

$$t_1 v_1 + t_2 v_2$$

is a linear combination of v_1 and v_2. Geometrically, the set of all linear combinations of v_1 and v_2 is called the plane spanned by v_1 and v_2. Any vector in this plane is linearly dependent on v_1 and v_2.
If we take another vector v_3 that is not in this plane then the three vectors are $\ell.\imath.$. In general, three vectors in $\mathbb{R}^3$ are linearly dependent iff they all lie in the same plane.
Three vectors in $\mathbb{R}^3$ are $\ell.\imath.$ iff there is no plane that contains all of them.





We discussed standard unit vectors, $\ell.\imath.$ sets and span. Now we bring these concepts together. The main question is: What is the minimum number of vectors needed to span a vector space?

Definition (Basis)

A basis of a subspace V of $\mathbb{R}^n$ is a $\ell.\imath.$ set of vectors that spans V.

Examples:

- Canonical basis: The set $\{(1,0,\ldots,0)^\top,(0,1,0,\ldots,0)^\top,\ldots,(0,\ldots,0,1)^\top\}$ is the most common used basis of $\mathbb{R}^n$.
- The set $\{(1,2)^\top,(3,5)^\top\}$ is a basis of $\mathbb{R}^2$.
- The set $\{(1,2,-4)^\top,(7,-5,6)^\top\}$ is $\ell.\imath.$ in $\mathbb{R}^3$ but is not a basis of $\mathbb{R}^3$.
- The set $\{(1,2)^\top,(3,5)^\top,(4,13)^\top\}$ spans $\mathbb{R}^2$ but is not a basis of $\mathbb{R}^2$ because it is not $\ell.\imath.$.

- The set $\{(1,-1,0)^\top,(1,0,-1)^\top\}$ is a basis of $W=\left\{(x,y,z)\in\mathbb{R}^3 : x+y+z=0\right\}$.

  1) Show $\ell.\imath.$:

  Two vectors are $\ell.\imath.$ if the only solution to the equation $\alpha_1 v_1+\alpha_2 v_2=0$ is $\alpha_1{=}\alpha_2{=}0$.

  Let $v_1=(1,-1,0)^\top$ and $v_2=(1,0,-1)^\top$ and solve the equation:

  $$\alpha_1(1,-1,0)^\top+\alpha_2(1,0,-1)^\top=(0,0,0)^\top\,.$$

  This gives the system:

  $$\begin{cases}\alpha_1+\alpha_2=0\\ -\alpha_1=0\\ -\alpha_2=0\end{cases}\,.$$

  From the second and third equations, we get $\alpha_1=0$ and $\alpha_2=0$. This shows that the vectors are $\ell.\imath.$.

2) Show $Span(v_1, v_2) = W$:

Take an arbitrary vector $(x, y, z)^\top$ in W. Since $x + y + z = 0$, we can rewrite z as $z = -x - y$.

Now, express $(x, y, z)^\top$ as a linear combination of v_1 and v_2:

$$a(1, -1, 0)^\top + b(1, 0, -1)^\top = (x, y, z)^\top .$$

This gives the system:

$$\begin{cases} a + b = x \\ -a = y \\ -b = z \end{cases} .$$

Substituting $z = -x - y$, we get:

$$\begin{cases} a + b = x \\ -a = y \\ -b = -x - y \end{cases} .$$

Solving this system, we find that $a = -y$ and $b = x + y$.

Since we can express any $(x, y, z)^\top \in W$ as a linear combination of v_1 and v_2 we proved $\mathrm{Span}(v_1, v_2) = W$.

- The set $\{(1,1,0)^\top, (0,0,1)^\top\}$ is a basis of $W = \left\{(x, x, y) \in \mathbb{R}^3 : x, y \in \mathbb{R}\right\}$. To prove the set is $\ell.\imath.$ we argue in the same lines as the previous example. Let $v_1 = (1,1,0)^\top$ $v_2 = (0,0,1)^\top$. To show $\mathrm{Span}(v_1, v_2) = W$. Take an arbitrary vector $(x, x, y)^\top$ in W. Now, express $(x, x, y)^\top$ as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$:

$$a(1,1,0)^\top + b(0,0,1)^\top = (x, x, y)^\top .$$

This gives us the system of equations:

$$\begin{cases} a = x \\ a = x \\ b = y \end{cases}$$

This system is consistent, and we find that $a = x$ and $b = y$.

- The set $\{1, x, \ldots, x^m\}$ is a basis of the space of polynomials of degree m in $\mathbb{R}$.
- A car moves in the directions $(x, y, 0)^\top$, a train moves in the space $(x, 0, 0)^\top$. A helicopter moves in $(x, y, z)^\top$.

A basis allows us to uniquely express every element of V as a linear combination of the elements of the basis:

Proposition (Criterion for basis)

A set of vectors $\{v_1, \ldots, v_n\}$ in V is a basis iff every $v \in V$ can be written <u>uniquely</u> in the form

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n,$$

where $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$.

Definition

Let V be a subspace of $\mathbb{R}^n$. The dimension of V is the number, m, of elements in any basis of V. We write

$$\dim(V) = m\,.$$

Example: We know that vectors $e_1, ...e_n$ form a basis of $\mathbb{R}^n$. Therefore $\dim(\mathbb{R}^n) = n$.

Proposition (Every subspace of $\mathbb{R}^n$ has a basis)

If a linearly independent subset of $\mathbb{R}^n$ contains m vectors, then $m \leq n$.

Later we will see important subspaces of $\mathbb{R}^n$ associated with matrices.

We now study an essential function in linear algebra.

Linear transformations

A function [1] $\mathcal{L} : V \to W$ is called a linear transformation if

1. $\mathcal{L}(\alpha x) = \alpha \mathcal{L}(x)$, for $x \in V$, and $\alpha \in \mathbb{R}$.
2. $\mathcal{L}(x + y) = \mathcal{L}(x) + \mathcal{L}(y)$, for $x, y \in V$.

Here we focus on transformations from $V = \mathbb{R}^n$ to $W = \mathbb{R}^m$, or subspaces of $\mathbb{R}^n$ and $\mathbb{R}^m$.

Examples:

- From $\mathbb{R}^2$ to $\mathbb{R}$: $\mathcal{L}(x_1, x_2) = 2x_1 - 3x_2$
- From $\mathbb{R}^3$ to $\mathbb{R}$: $\mathcal{L}(x_1, x_2, x_3) = x_1 + 2x_2 + 3x_3$
- From $\mathbb{R}$ to $\mathbb{R}^3$: $\mathcal{L}(x) = (x, 2x, 3x)$
- From $\mathbb{R}^2$ to $\mathbb{R}^2$: $\mathcal{L}(x_1, x_2) = (2x_1 + x_2, x_1 + 3x_2)$

Since any linear transformation $\mathcal{L}$ can be represented by a matrix (more on this later) we can use the notation:

$$\mathcal{L}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 \\ x_1 + 3x_2 \end{pmatrix}.$$

[1] check slides `Preliminaries.pdf`

Linear transformations and matrices

Definition (Matrix)

Let m and n denote positive integers. An m-by-n matrix A is a rectangular array of elements in $\mathbb{R}$ with m rows and n columns:

$$A = \begin{pmatrix} a_{1,1} & \dots & a_{n,1} \\ \vdots & & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix},$$

first index refers to the row number and the second index refers to the column number.

Notation: $A \in \mathbb{R}^{m \times n}$.

Same matrix can be denoted as a n-tuple of column vectors in $\mathbb{R}^m$:

$$A = (v_1, \dots, v_n),$$

where $v_i = (a_{1,i}, \dots, a_{m,i})^\top$.

Matrix operations

Marix sum: The sum of two matrices of the same size is the matrix obtained by adding corresponding entries in the matrices.

Matrix scalar product: The product of a scalar and a matrix is the matrix obtained by multiplying each entry in the matrix by the scalar.

The most interesting operation is the **Matrix product:**

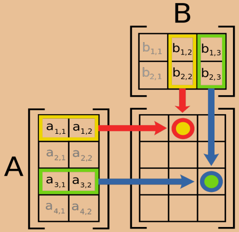
Definition

Let $A \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{n\times p}$. Then $AC \in \mathbb{R}^{m\times p}$ has entry in row j and column k:

$$(AC)_{j,k} = \sum_{r=1}^{n} a_{j,r}c_{r,k}\,.$$

Note: we define the product only when the number of columns of the first matrix equals the number of rows of the second matrix.

Matrix operations



Why the product of matrices is defined in this way?

Consider the linear transformations $\mathcal{L} : U \to V$ and $\mathcal{S} : V \to W$ with associated matrices C and A respectively.

Now observe the composition $\mathcal{S} \circ \mathcal{L}$ is a linear map from U to W. Suppose the associated matrix of this linear transformation is B.

Now the question is

$CA \stackrel{?}{=} B$ answer is yes due to matrix product.

Matrix operations

For $1 \leq k \leq p$, we have [2]

$$
\begin{aligned}
(\mathcal{S}\mathcal{L})u_k &= \mathcal{S}\left(\sum_{r=1}^{n} c_{r,k} v_r\right) \\
&= \sum_{r=1}^{n} c_{r,k} \mathcal{S} v_r \\
&= \sum_{r=1}^{n} c_{r,k} \sum_{j=1}^{m} a_{j,r} w_j \\
&= \sum_{j=1}^{m} \left(\sum_{r=1}^{n} a_{j,r} c_{r,k}\right) w_j \,.
\end{aligned}
$$

Thus

$$B = \sum_{r=1}^{n} a_{j,r} c_{r,k} \,. \qquad \text{(check the derivation)}$$

[2] u_k, v_k, and w_k are the elements of the basis of U, V and W.

Determinant

The determinant of a <u>square</u> matrix $A \in \mathbb{R}^{n\times n}$ is a function that maps A into $\mathbb{R}$.

Notation:

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Properties:

- The determinant of the identity matrix is 1.
- The determinant changes sign when two rows are exchanged.
- The determinant is a linear function of each row separately:

$$\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix},$$

$$\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}.$$

Determinant - Properties

Properties (cont.):

- If two rows of matrix A are equal, then $\det(A) = 0$.
- Subtracting a multiple of one row from another row leaves the determinant of matrix A unchanged.
- A matrix with a row of zeros has $\det(A) = 0$.
- If A is triangular then $\det(A)$ is the product of diagonal entries.
- If A is singular, then $\det(A) = 0$. If A is invertible, then $\det(A) \neq 0$.
- $\det(AB) = \det(A)\det(B)$
- $\det(A^\top) = \det(A)$

Determinant - Particular cases

Mnemonic rules for two particular cases:

The determinant of a 2×2 matrix is

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

and the determinant of a 3×3 matrix is

$$\det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

Determinant - Expansion

Consider a matrix $A \in \mathbb{R}^{n \times n}$. Then for all $j = 1, \ldots, n$:

1. Expansion along column j

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(A_{k,j}).$$

2. Expansion along row j

$$\det(A) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(A_{j,k}).$$

Here $A_{k,j} \in \mathbb{R}^{(n-1) \times (n-1)}$ is the submatrix of A that we can obtain when deleting row k and column j.

Determinant - Expansion Example

Consider

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} .$$

The determinant of A is calculated as follows:

$$\begin{aligned} \det(A) &= (-1)^{1+1} \cdot \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+2} \cdot 2 \cdot \begin{vmatrix} 3 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+3} \cdot 3 \cdot \begin{vmatrix} 3 & 1 \\ 0 & 0 \end{vmatrix} \\ &= 1(1 \cdot 1 - 0 \cdot 2) - 2(3 \cdot 1 - 0 \cdot 2) + 3(3 \cdot 0 - 0 \cdot 1) \\ &= 1(1 - 0) - 2(3 - 0) + 3(0 - 0) \\ &= -5 \end{aligned}$$

Linear transformation

Fixing suitable bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, we can write any linear transformation $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ as

$$\mathcal{L}(x) = Ax.$$

Notes:

- The matrix A represents the transformation $\mathcal{L}$ under the particular bases used for $\mathbb{R}^n$ and $\mathbb{R}^m$. If one of the basis is changed, then the matrix that represents the transformation changes.
- Notation: To emphasize that the matrix A is associated with fixed bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ we can write ${}_{\mathcal{W}}(A)_{\mathcal{V}}$ where $\mathcal{V}$ is a basis for $\mathbb{R}^n$ and $\mathcal{W}$ is a basis for $\mathbb{R}^m$.
- The matrix multiplication Ax transforms the vector $x \in \mathbb{R}^n$ into a vector in $\mathbb{R}^m$.
- Note a vector in $\mathbb{R}^n$ is a matrix $n \times 1$.
- Note $Ax + b$ is not a linear transformation, but an *affine* one.

Linear transformation

Example: Given

$$A = \begin{pmatrix} -2 & 1 & 3 \\ 4 & -2 & -6 \end{pmatrix}.$$

(a) Find the domain and the codomain of the transformation $\mathcal{L}$ represented by A.
(b) Find and draw the image of $\mathcal{L}$.

(a) A is a 2×3. So, the domain of $\mathcal{L}$ is $\mathbb{R}^3$ and the codomain is $\mathbb{R}^2$.

(b) Let $v = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ be an arbitrary vector of $\mathbb{R}^3$. The image of v is given by

$$\mathcal{L}(v) = Av = \begin{pmatrix} -2 & 1 & 3 \\ 4 & -2 & -6 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = a \begin{pmatrix} -2 \\ 4 \end{pmatrix} + b \begin{pmatrix} 1 \\ -2 \end{pmatrix} + c \begin{pmatrix} 3 \\ -6 \end{pmatrix}$$

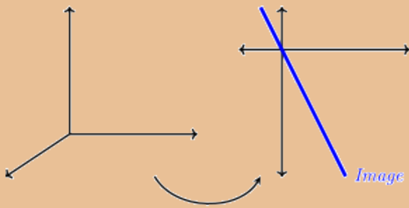
Linear transformation

This shows that the image of every vector in $\mathbb{R}^3$ is a linear combination of the vectors $\begin{pmatrix} -2 \\ 4 \end{pmatrix}$, $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$, and $\begin{pmatrix} 3 \\ -6 \end{pmatrix}$.

To find the image of $\mathcal{L}$, we can observe that the first and third vector are linear combinations of $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

This shows that the image of every vector in $\mathbb{R}^3$ is a multiple of $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$.

This means that the image of $\mathcal{L}$ is a line in $\mathbb{R}^2$.



Definition

The identity transformation on V, denoted by id_V, is a transformation that maps each element of V to itself. In other words,

$$id_V : V \to V$$

is a transformation such that

$$id_V(v) = v\,, \quad \text{for all } v \in V$$

The identity transformation is linear.

Linear transformation

Definition (Linear combination of columns)

Let $A \in \mathbb{R}^{m\times n}$ *and* $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ *is an n-by-1 matrix (or vector). Then*

$$Ax = x_1 a_{\cdot,1} + \cdots + x_n a_{\cdot,n} \,,$$

where $a_{\cdot,k}$ *denotes the k-th column of A.*

In other words, Ax is a linear combination of the columns of A, with the scalars that multiply the columns the components of x.

Review videos:

- Linear combinations and bases: `https://youtu.be/k7RM-ot2NWY`
- Linear transformations and matrices: `https://youtu.be/kYB8IZa5AuE`
- Matrix multiplication and composition: `https://youtu.be/XkY2DOUCWMU`

Geometry of linear transformations

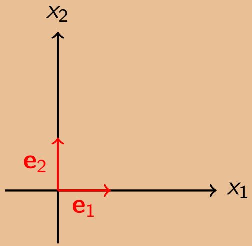
Manipulation of digital images is a good example to visualize linear transformations. Every pixel of the image as a point in $\mathbb{R}^2$. A transformation is applied to each pixel, and the output pixel is colored the same color as the input pixel.

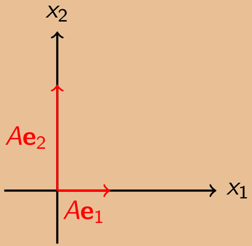
To understand what a linear transformation does, it is enough to understand what it does to basis vectors.

For example, the matrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

does the following transformation to the canonical basis:





This transformation takes point (a, b) to point $(a, 2b)$.

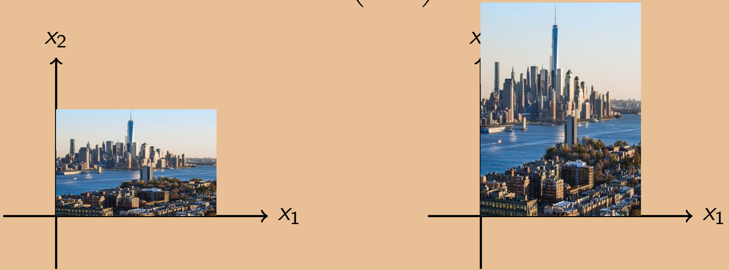
Geometry of linear transformations - scaling

A linear transformation that scales objects in the plane vertically by a factor of $\alpha, \alpha > 0$ is given by

$$A_v = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix}.$$

A linear transformation that scales objects in the plane horizontally by a factor of $\alpha, \alpha > 0$ is given by

$$A_h = \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix}.$$



Vertical scaling of $2\times$.

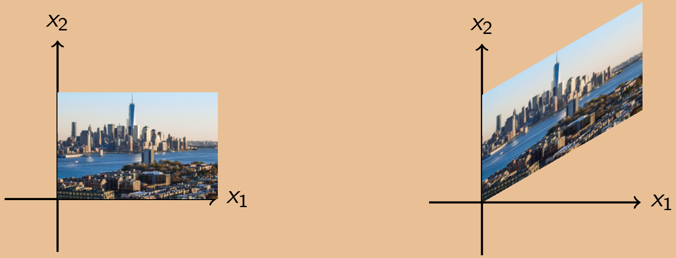
Geometry of linear transformations - shear

A horizontal shear is a transformation that takes an arbitrary point (a, b) and maps it to the point $(a + \alpha b, b)$. This can be accomplished by

$$A_{hs} = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} .$$

A vertical shear is a transformation that takes an arbitrary point (a, b) and maps it to the point $(a, b + \alpha a)$. This can be accomplished by

$$A_{vs} = \begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} .$$

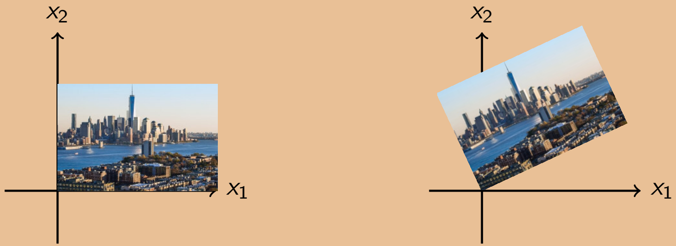


Vertical shear 30 degrees, that is $\alpha = 1/\sqrt{3}$.

Geometry of linear transformations - rotation

A linear transformation that rotates the plane counterclockwise through angle θ about the origin is given by

$$A_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} .$$

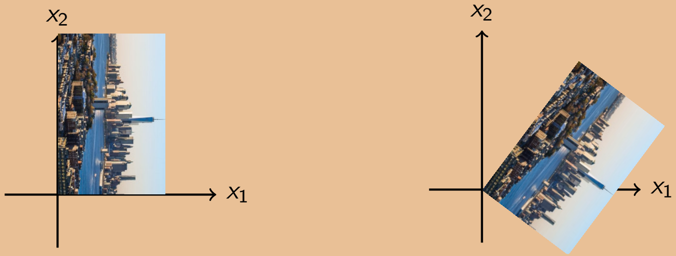


Counterclockwise rotation of 25 degrees about the origin.

Geometry of linear transformations - reflection about line

A linear transformation that reflects the plane about the line $y = mx$ is given by

$$M_l = \frac{1}{1+m^2}\begin{pmatrix} 1-m^2 & 2m \\ 2m & m^2-1 \end{pmatrix}.$$



Left: $m = 1$, right: $m = 1/2$.

Task: An interesting task is to compose many shears, rotations and scalings to obtain certain effects to the images.

Side note: Invertible linear transformation

Definition (Invertible linear transformation)

- *A linear transformation $\mathcal{L} : V \to W$ is called invertible if there exists a linear transformation $\mathcal{S} : W \to V$ such that $\mathcal{SL}$ is the identity transformation on V and $\mathcal{LS}$ equals the identity map on W.*
- *A linear transformation $\mathcal{S} : W \to V$ satisfying $\mathcal{SL} = I$ and $\mathcal{LS} = I$ is called an inverse of $\mathcal{L}$.*

Proposition (Inverse is unique)

An invertible linear transformation has a unique inverse.

Proposition

A linear transformation is invertible iff it is bijective.

Notation: $\mathcal{L}^{-1}$.

Operator and Invertible matrix

Definition (Operator: $\mathbb{R}^n \to \mathbb{R}^n$ linear transformation)

An important particular case of linear transformations are the ones from $\mathbb{R}^n \to \mathbb{R}^n$. Those are called operators and represented by square matrices.

Definition (Invertible matrix)

$A \in \mathbb{R}^{n\times n}$ is called invertible (also nonsingular) if there exists a $B \in \mathbb{R}^{n\times n}$ such that

$$AB = BA = I$$

where $I = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$*, the $n \times n$ identity matrix.*

The matrix B is uniquely determined by A.

Notation: Inverse of A is denoted A^{-1}.

Note An operator $\mathcal{L}$ is invertible iff the matrix A that represents $\mathcal{L}$ is invertible.

Invertible matrix - Useful properties

Let $A \in \mathbb{R}^{n \times n}$. The following statements are equivalent:

- A is invertible, that is there exists a B such that $AB = BA = I$.
- The linear transformation that maps x to Ax is invertible (and bijective).
- A is row-equivalent and column-equivalent to the n-by-n identity matrix I_n.
- A has full rank: $\operatorname{rank}(A) = n$. (more on rank later)
- A has trivial nullspace (or kernel): $\mathcal{N}(A) = \{0\}$. (more on nullspace later)
- The columns of A are $\ell.\imath.$.
- The rows of A are $\ell.\imath.$.
- The determinant of A is nonzero.
- The number 0 is not an eigenvalue of A. (more on this later)
- The transpose $A^\top$ is an invertible matrix.

Inverse of a 2×2 matrix

Let

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

The inverse of A is given by

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} = \frac{1}{\det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} .$$

Inverse of a 3×3 matrix

Given the matrix

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 4 & 4 & 5 \\ 6 & 7 & 7 \end{pmatrix} ,$$

the inverse is given by

$$A^{-1} = \begin{pmatrix} -7 & -7 & 6 \\ 2 & 1 & -1 \\ 4 & 5 & -4 \end{pmatrix} .$$

To compute this inverse we need to calculate:

- The minors matrix: a matrix of determinants.
- The cofactors matrix: the minors matrix element-wise multiplied by a grid of alternating $+1$ and -1.
- The adjugate matrix: the transpose of the cofactors matrix.
- The inverse matrix: the adjugate matrix divided by the determinant.

Inverse of a 3×3 matrix

In the previous example we have:

Minors:

$$\begin{pmatrix} -7 & -2 & +4 \\ +7 & +1 & -5 \\ +6 & +1 & -4 \end{pmatrix},$$

Cofactors:

$$\begin{pmatrix} -7 & +2 & +4 \\ -7 & +1 & +5 \\ +6 & -1 & -4 \end{pmatrix},$$

Adjugate:

$$\begin{pmatrix} -7 & -7 & +6 \\ +2 & +1 & -1 \\ +4 & +5 & -4 \end{pmatrix}.$$

Linear system

Essential feature of matrices: The linear system of equations

$$\begin{pmatrix} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \dots & + & a_{1,n}x_n & = & b_1 \\ \vdots & & \vdots & & \dots & & \vdots & = & \vdots \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \dots & + & a_{m,n}x_n & = & b_m \end{pmatrix}$$

can be compactly written as

$$Ax = b\,,$$

for $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$.

Task: Discuss solving a n-by-n system using the inverse matrix.

Linear system

Definition (Consistent system)

A system of equations is called consistent if there exists at least one solution.

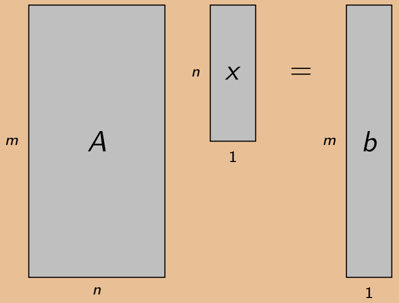
Definition (Inconsistent system)

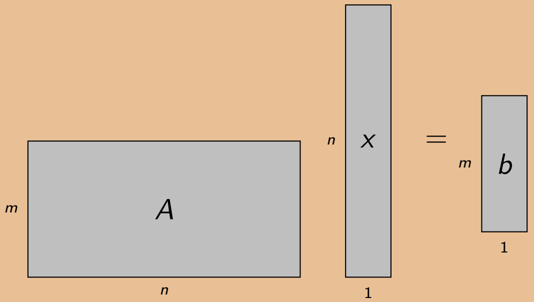
A system of equations is called inconsistent if there is no solution.

Definition

A linear system is called overdetermined if there are more equations than unknowns $m > n$. It is called underdetermined if there are more variables than equations $m < n$. We have a square system when $m = n$.

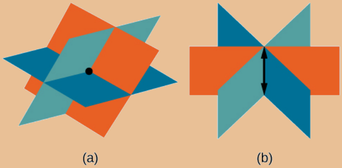
Over/underdetermined system illustration



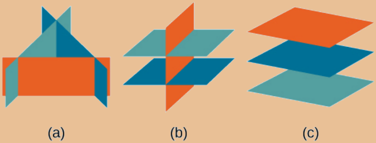


Linear system - Geometric view

The solution, if any, of a linear system represents the points of intersection of the hyperplanes given by the linear equations.



(a) Three planes intersect at a single point, representing a 3×3 system with unique solution. (b) Three planes intersect in a line, representing a 3×3 system with infinite solutions. Same if the three planes coincide.



No solution. (a) The three planes intersect with each other, but not at a common point. (b) Two of the planes are parallel and intersect with the third plane, but not with each other. (c) All three planes are parallel, so there is no point of intersection. (d) Two planes coincide but no intersection with the other.

Linear system - 3×3 example

Example with unique solution:

$$
\begin{aligned}
2x + 3y + z &= 2 && \text{(Magenta)}\\
-3x - 3y + z &= 0 && \text{(Cyan)}\\
-x + 2y + 3z &= -1 && \text{(Lime)}
\end{aligned}
$$

That is

$$Ax = b\,,$$

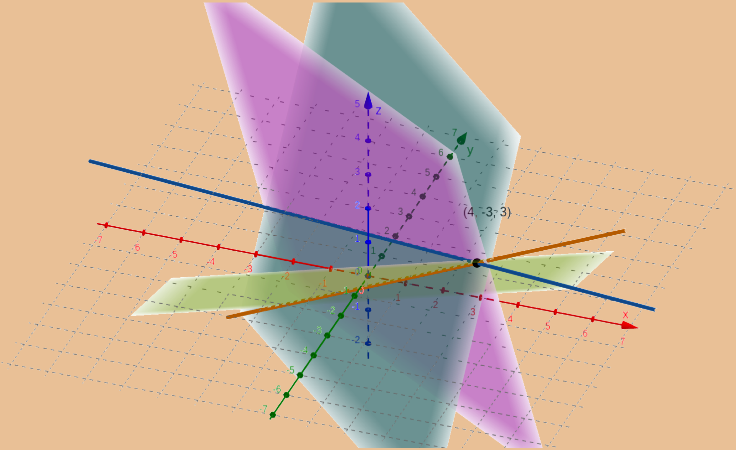
where

$$A = \begin{pmatrix} 2 & 3 & 1 \\ -3 & -3 & 1 \\ -1 & 2 & 3 \end{pmatrix}, \qquad b = \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}.$$

In this case, the inverse of A exists:

$$A^{-1} = \frac{1}{7}\begin{pmatrix} 11 & 7 & -6 \\ -8 & -7 & 5 \\ 9 & 7 & -3 \end{pmatrix}, \quad \text{and} \quad A^{-1}b = \begin{pmatrix} 4 \\ -3 \\ 3 \end{pmatrix}.$$

Linear system - 3×3 example



https://www.geogebra.org/3d/fpwqsuv7

Linear system - 3×3 example

Example with infinite solutions (line):

$$
\begin{aligned}
x + 2y + 3z &= 1 && \text{(Magenta)} \\
2x - y + z &= 3 && \text{(Cyan)} \\
3x + y + 4z &= 4 && \text{(Lime)}
\end{aligned}
$$

The third equation is a linear combination of the first two equations:

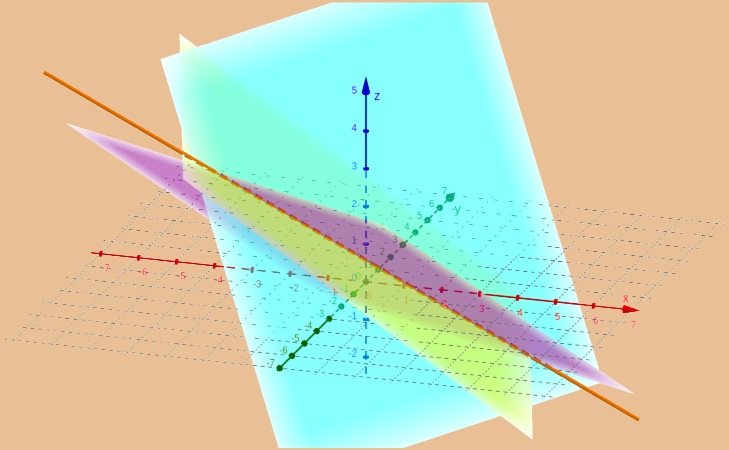
$$3x + y + 4z = (x + 2y + 3z) + (2x - y + z) = 3x + y + 4z$$

When converted to a matrix form, the augmented matrix for this system is:

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 2 & -1 & 1 & 3 \\ 3 & 1 & 4 & 4 \end{array}\right)$$

By performing row reduction, you would be able to find that this system has infinitely many solutions.

Linear system - 3×3 example



https://www.geogebra.org/3d/ftcqynff

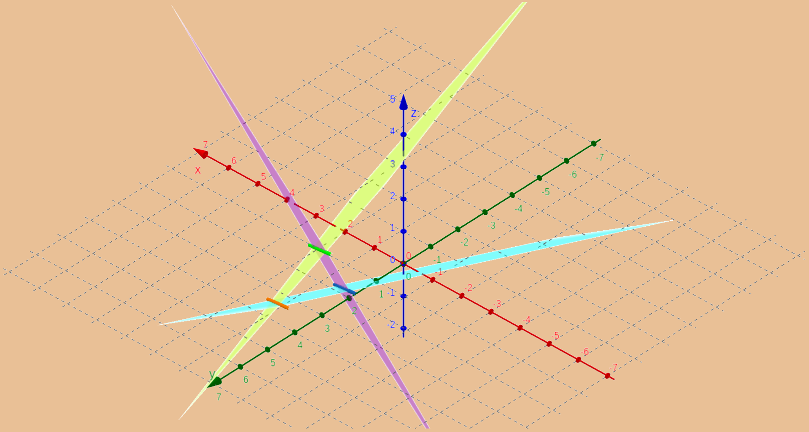
Linear system - 3×3 example

Example with no solutions: the three planes intersect but not in common:

$$x + 2y - z = 4 \qquad \text{(Magenta)}$$

$$2x - y + 3z = -1 \qquad \text{(Cyan)}$$

$$3x + y + 2z = 7 \qquad \text{(Lime)}$$



Linear system - Gaussian elimination

Performs elementary row operations to bring a system of linear equations into reduced row-echelon form. The goal is to write matrix A with the number 1 as the entry down the main diagonal and have all zeros below (row-echelon form).

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \xrightarrow{\text{After Gaussian elimination}} \begin{pmatrix} 1 & b_{12} & b_{13} \\ 0 & 1 & b_{23} \\ 0 & 0 & 1 \end{pmatrix}$$

Elementary row operations to convert A row-echelon form:

- Interchange rows. (Notation: $R_i \leftrightarrow R_j$)
- Multiply a row by a constant. (Notation: cR_i)
- Add the product of a row multiplied by a constant to another row. (Notation: $R_i + cR_j$)

Performing a sequence of elementary row operations on a system of equations produces an equivalent system.

Example: `https://youtu.be/76Y41ncuLeQ`

Gaussian elimination example

$$\begin{pmatrix} 1 & 1 & 3 \\ 1 & 2 & 4 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 15 \\ 21 \\ 13 \end{pmatrix}$$

$$\begin{array}{r} \\ R_2 - R_1 \\ R_3 - R_1 \end{array} \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 15 \\ 6 \\ -2 \end{pmatrix}$$

$$\begin{array}{r} \\ \\ -R_3 \end{array} \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 15 \\ 6 \\ 2 \end{pmatrix}$$

$$\begin{array}{r} R_1 - 3R_3 \\ R_2 - R_3 \\ \\ \end{array} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 9 \\ 4 \\ 2 \end{pmatrix}$$

$$\begin{array}{r} R_1 - R_2 \\ \\ \\ \end{array} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \\ 2 \end{pmatrix} .$$

Linear system - Gaussian elimination

Applications of Gaussian elimination:

- Solving linear systems.
- Finding rank of a matrix.
- Finding basis vectors for a set of given vectors.
- Determining if a symmetric matrix is positive definite.
- Finding the determinant of a matrix.
- Finding LU decomposition of a matrix.

Review the Gaussian elimination method: [3] [4]

[3] https://youtu.be/QVKj3LADCnA

[4] https://math.libretexts.org/@go/page/3101

Subspaces of $\mathbb{R}^n$ associated with matrices

A conclusion of the Gaussian elimination is that there are certain characteristics associated with the rows of a matrix that are not affected by elementary row operations. We now discuss this observation.

Definition (Row space of a matrix)

Let $A \in \mathbb{R}^{m \times n}$. The row space of A, denoted by $\operatorname{row}(A)$, *is the subspace of $\mathbb{R}^n$ spanned by the rows of A.*

Consider

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

Let r_1 and r_2 be the rows of A. Then $\operatorname{row}(A) = \operatorname{Span}(r_1, r_2)$ is a plane through the origin containing r_1 and r_2.

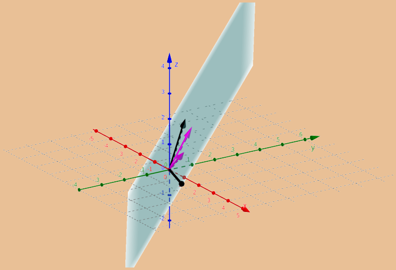
Using elementary row operations we can reduce A to

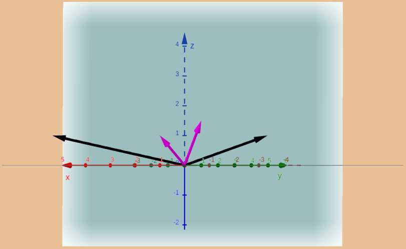
$$\begin{pmatrix} -2 & 2 & 1 \\ 4 & -2 & 1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{3}{2} \end{pmatrix}$$

Let ρ_1 and ρ_2 be the rows of the reduced matrix,

$$\rho_1 = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}, \quad \rho_2 = \begin{pmatrix} 0 & 1 & \frac{3}{2} \end{pmatrix}$$

Vectors ρ_1 and ρ_2 were obtained from r_1 and r_2 by repeated applications of elementary row operations. At every stage of the row reduction process, the rows of the matrix are linear combinations of r_1 and r_2 and thus lie in the span of r_1 and r_2.
In the next image black vectors are r_1 and r_2.





https://www.geogebra.org/3d/txnpgudb

Subspaces of $\mathbb{R}^n$ - $\operatorname{col}(A)$

Definition (Column space of a matrix)

Let $A \in \mathbb{R}^{m \times n}$. The column space of A, denoted by $\operatorname{col}(A)$, is the subspace of $\mathbb{R}^m$ spanned by the columns of A.

Proposition

Let A be a matrix. The dimension of the row space of A, the rank of A, and the dimension of the column space of A are all equal:

$$\dim(\operatorname{row}(A)) = \operatorname{rank}(A) = \dim(\operatorname{col}(A))$$

(we will see later what is $\operatorname{rank}(A)$).

Subspaces of $\mathbb{R}^n$ - Null space

Definition (Null space)

Let $A \in \mathbb{R}^{m \times n}$. The null space of A, denoted by $\mathcal{N}(A)$, is the set of all vectors x in $\mathbb{R}^n$ such that

$$Ax = 0 \, .$$

Notice:

- Vectors in the null space are the ones that get “smashed down” to zero when the transformation A is applied.
- The null space gives us all possible solutions to the homogeneus system $Ax = 0$. This means the null space is never empty, because at least contains the trivial solution $x = 0$.
- If the null space only contains the zero vector, then this means the columns of A are $\ell.\imath.$.
- If a vector $v \neq 0$ is in the null space, the entries of v tell how to create a linear combination of the columns of A that results in the zero vector.

Subspaces of $\mathbb{R}^n$ - Null space

Example: Find the null space of the matrix

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

We present two ways. 1st way) Solve the homogeneous system of linear equations given by $Ax = 0$, that is

$$\begin{aligned} 3x_1 - x_2 &= 0 \\ -6x_1 + 2x_2 &= 0 \, . \end{aligned}$$

Notice that the second equation is just a multiple of the first equation, so we only need to consider one of them. From

$$3x_1 - x_2 = 0 \, .$$

we get $x_2 = 3x_1$.
So the system has infinite solutions since $x_1 \in \mathbb{R}$ is arbitrary.

Subspaces of $\mathbb{R}^n$ - Null space

That is, the solution to the system can be written as:

$$\begin{pmatrix} \alpha \\ 3\alpha \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 3 \end{pmatrix},$$

where $\alpha \in \mathbb{R}$ is a free variable or parameter.

Therefore, the null space of A is the span of the vector $(1, 3)^\top$.

Finally, we note that the null space is a one-dimensional subspace of $\mathbb{R}^2$ since is spanned by one vector.

Remark: Observe that the vector $(\frac{1}{3}, 1)^\top$ also satisfies the requirement that the second component is 3 times the first.

2nd way) Use row reduction,

$$\begin{pmatrix} 3 & -1 \\ -6 & 2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -\frac{1}{3} \\ 0 & 0 \end{pmatrix} .$$

From here we obtain again the only equation

$$x_1 - x_2/3 = 0 ,$$

from where we get $x_1 = x_2/3$.

We conclude that $x = x_2 \begin{pmatrix} 1/3 \\ 1 \end{pmatrix}$ for $x_2 \in \mathbb{R}$, which is the same conclusion as before.

Using the following Mathematica commands we can easily compute the row-echelon form:

```
A = {{3, -1}, {-6, 2}};
RowReduce[A]
```

Null space - Example

$$\begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{array}{r} \\ R_2 - 2R_1 \\ R_3 - 3R_1 \end{array} \begin{pmatrix} 1 & 1 & 2 \\ 0 & -1 & -1 \\ 0 & -2 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{array}{r} \\ -R_2 \\ R_3 + 2R_2 \end{array} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{array}{r} R_1 - R_2 \\ \\ \\ \end{array} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

First two columns are pivot and last one free. The null space is spanned by the vectors corresponding to the free ones.

Null space - Example

Here z is a free variable. Express x and y in terms of z. The equations from the reduced form matrix are:

$$\begin{aligned} x + z &= 0 \\ y + z &= 0\,. \end{aligned}$$

To find the null space vector we set $z = t$, where t is a parameter. Then,

$$\begin{aligned} x &= -t \\ y &= -t\,. \end{aligned}$$

So, the null space vector is a function of t, that is $(-t, -t, t)^\top$. If we choose $t = 1$, we get the vector $(-1, -1, 1)^\top$ which is a basis of the null space. This vector represents the direction along which any scalar multiple will still satisfy the equation $Ax = 0$.

So the null space matrix can be written $Z = t\begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$, or $t\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$.

Mathematica:

```
A={{1,1,2},{2,1,3},{3,1,4}}; b={0,0,0}; Solve[A.{x,y,z}==b,{x,y,z}]
```

Rank/nullity theorem

In the example we can observe that the dimension of the null space of a matrix is equal to the number of free variables in the solution vector of the homogeneous system associated with the matrix.

This leads to the following relevant result:

Theorem

Let $A \in \mathbb{R}^{m \times n}$. Then

$$\operatorname{rank}(A) + \dim(\mathcal{N}(A)) = n\,.$$

The four fundamental subspaces

The four fundamental subspaces of $A \in \mathbb{R}^{m \times n}$:

Name	**Notation**	**Note**
Column Space	$\mathrm{col}\,(A) \subseteq \mathbb{R}^m$	All combinations of the columns of A, $\{y : y = Ax\}$.
Null Space	$\mathcal{N}(A) \subseteq \mathbb{R}^n$	All solutions of $Ax = 0$.
Row Space	$\mathrm{col}\,(A^\top) \subseteq \mathbb{R}^n$	All combinations of the rows of A, $\{y : y^\top = x^\top A\}$.
Left Null Space	$\mathcal{N}(A^\top) \subseteq \mathbb{R}^m$	All solutions of $A^\top x = 0$.

Note $\mathrm{row}\,(A) \equiv \mathrm{col}\,(A^\top)$.

Orthogonal Vectors: Two vectors v and w are said to be orthogonal when

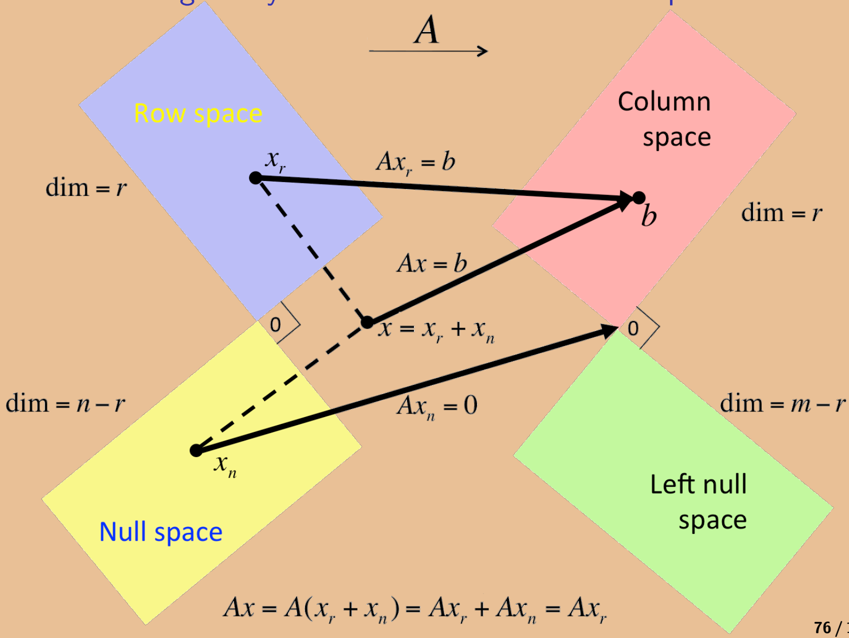
$$v^\top w = 0\,.$$

Definition (Orthogonal Subspaces)

Two subspaces $V, W \subseteq \mathbb{R}^n$ are orthogonal if every vector $v \in V$ is perpendicular to every vector $w \in W$.

$$v^\top w = 0\,, \qquad \textit{for all } v \in V \textit{ and } w \in W\,.$$

Orthogonality of the four fundamental subspaces



$$Ax = A(x_r + x_n) = Ax_r + Ax_n = Ax_r$$

Orthogonality of the four fundamental subspaces - Exercises

Exercise OS1: Let $A \in \mathbb{R}^{m \times n}$. Show that its row space, $\text{row}\,(A)$, and null space, $\mathcal{N}(A)$, are orthogonal.

Solution: Pick arbitrary $x \in \text{col}\,(A)$ and $y \in \mathcal{N}(A)$. We need to show that these two vectors are orthogonal.

Then

$$
\begin{aligned}
& x^\top y \\
& \quad \langle\, x \in \text{row}\,(A) \text{ iff there exists } z \text{ s.t. } x = A^\top z \,\rangle \\
& \quad = (A^\top z)^\top y \\
& \quad = z^\top A y \\
& \quad \langle\, y \in \mathcal{N}(A) \,\rangle \\
& \quad = z^\top 0 \\
& \quad = 0.
\end{aligned}
$$

Orthogonality of the four fundamental subspaces - Exercises

Exercise OS2: Let $A \in \mathbb{R}^{m \times n}$. Show that its column space, $\mathrm{col}\,(A)$, and left null space, $\mathcal{N}(A^\top)$, are orthogonal.

Solution: Pick arbitrary $x \in \mathrm{col}\,(A)$ and $y \in \mathcal{N}(A^\top)$.

Then

$$
\begin{aligned}
& x^\top y \\
& \quad \langle\, x \in \mathrm{col}\,(A) \text{ iff there exists } z \text{ s.t. } x = Az \,\rangle \\
& \quad = (Az)^\top y \\
& \quad = z^\top A^\top y \\
& \quad \langle\, y \in \mathcal{N}(A^\top) \,\rangle \\
& \quad = z^\top 0 \\
& \quad = 0.
\end{aligned}
$$

Orthogonality of the four fundamental subspaces - Exercises

Exercise OS3: Let $\{s_0, \ldots, s_{r-1}\}$ be a basis for subspace $S \subseteq \mathbb{R}^n$ and $\{t_0, \ldots, t_{k-1}\}$ be a basis for subspace $T \subseteq \mathbb{R}^n$. Show that the following are equivalent statements:

1. Subspaces S, T are orthogonal.
2. The vectors in $\{s_0, \ldots, s_{r-1}\}$ are orthogonal to the vectors in $\{t_0, \ldots, t_{k-1}\}$.
3. $s_i^\top t_j = 0$ for all $0 \leq i < r$ and $0 \leq j < k$.
4. $\left(s_0 \mid \ldots \mid s_{r-1}\right)^\top \left(t_0 \mid \ldots \mid t_{k-1}\right) = O$, the zero matrix of appropriate size.

Solution: We are going to prove the equivalence of all the statements by showing that $1 \implies 2$; $2 \implies 3$; $3 \implies 4$; and $4 \implies 1$.

Orthogonality of the four fundamental subspaces - Exercises

Solution:

- $1 \Longrightarrow 2$:
  Subspaces S and T are orthogonal if any vectors $x \in S$ and $y \in T$ are orthogonal. Obviously, this means that s_i is orthogonal to t_j for $0 \leq i < r$ and $0 \leq j < k$.
- $2 \Longrightarrow 3$:
  This is true by definition of what it means for two sets of vectors to be orthogonal.
- $3 \Longrightarrow 4$:

$$\left(s_0 \mid \cdots \mid s_{r-1} \right)^\top \left(t_0 \mid \cdots \mid t_{k-1} \right) = \begin{pmatrix} s_0^\top t_0 & s_0^\top t_1 & \cdots \\ s_1^\top t_0 & s_1^\top t_1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Orthogonality of the four fundamental subspaces - Exercises

- $4 \Longrightarrow 1$:
  We need to show that if $x \in S$ and $y \in T$ then $x^\top y = 0$.
  Notice that
  $$x = \left(s_0 \quad \middle| \quad \cdots \quad \middle| \quad s_{r-1} \right) \begin{pmatrix} \hat{x}_0 \\ \vdots \\ \hat{x}_{r-1} \end{pmatrix}$$
  and
  $$y = \left(t_0 \quad \middle| \quad \cdots \quad \middle| \quad t_{k-1} \right) \begin{pmatrix} \hat{y}_0 \\ \vdots \\ \hat{y}_{k-1} \end{pmatrix}$$
  for appropriate choices of $\hat{x}$ and $\hat{y}$.

Orthogonality of the four fundamental subspaces - Exercises

But then

$$
\begin{aligned}
x^\top y &= \left(\begin{pmatrix} s_0 & | & \cdots & | & s_{r-1} \end{pmatrix} \begin{pmatrix} \hat{x}_0 \\ \vdots \\ \hat{x}_{r-1} \end{pmatrix} \right)^\top \begin{pmatrix} t_0 & | & \cdots & | & t_{k-1} \end{pmatrix} \begin{pmatrix} \hat{y}_0 \\ \vdots \\ \hat{y}_{k-1} \end{pmatrix} \\
&= \begin{pmatrix} \hat{x}_0 \\ \vdots \\ \hat{x}_{r-1} \end{pmatrix}^\top \underbrace{\begin{pmatrix} s_0 & | & \cdots & | & s_{r-1} \end{pmatrix}^\top \begin{pmatrix} t_0 & | & \cdots & | & t_{k-1} \end{pmatrix}}_{0_{r \times k}} \begin{pmatrix} \hat{y}_0 \\ \vdots \\ \hat{y}_{k-1} \end{pmatrix} \\
&= 0_{r \times k} .
\end{aligned}
$$

Orthogonality of the four fundamental subspaces - Exercises

Exercise OS4:
Let $A \in \mathbb{R}^{m \times n}$. Show that any vector $x \in \mathbb{R}^n$ can be written as $x = x_r + x_n$, where $x_r \in \mathcal{R}(A)$ and $x_n \in \mathcal{N}(A)$, and $x_r^\top x_n = 0$.

Solution:
Let r be the rank of matrix A. Then the dimension of $\text{row}(A)$, is r and the dimension of the null space, $\mathcal{N}(A)$, is $n - r$.

Let $\{w_0, \ldots, w_{r-1}\}$ be a basis for $\text{row}(A)$ and $\{w_r, \ldots, w_{n-1}\}$ be a basis for $\mathcal{N}(A)$. Since we know that these two spaces are orthogonal, we know that $\{w_0, \ldots, w_{r-1}\}$ are orthogonal to $\{w_r, \ldots, w_{n-1}\}$. Hence $\{w_0, \ldots, w_{n-1}\}$ are linearly independent and form a basis for $\mathbb{R}^n$. Thus, there exist coefficients $\{a_0, \ldots, a_{n-1}\}$ such that

$$
\begin{aligned}
x &= a_0 w_0 + \ldots + a_{n-1} w_{n-1} \\
&\quad \langle \text{ split the summation } \rangle \\
&= a_0 w_0 + \ldots + a_{r-1} w_{r-1} + a_r w_r + \ldots + a_{n-1} w_{n-1}.
\end{aligned}
$$

Definition (Transformation matrix)

Let $\{b_1, b_2, ..., b_n\}$ *and* $\{\tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_n\}$ *be two bases for* $\mathbb{R}^n$*. Then the matrix*

$$T = (\tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_n)^{-1}(b_1, b_2, ..., b_n)$$

is called a transformation matrix.

Given a vector in $\mathbb{R}^n$, let x be its representation in the basis $\{b_1, b_2, ..., b_n\}$ and $\tilde{x}$ its representation in the basis $\{\tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_n\}$, then

$$\tilde{x} = Tx\,.$$

Note: This result allows a linear transformation to be represented by different matrices. This is crucial in applications since matrix A is usually given by the problem and depending on the applications alternative representations can be found.

Similar matrices

Consider the operator $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^n$ and let A be its matrix representation in terms of the basis $\{b_1, b_2, ..., b_n\}$, and B is the matrix representation with respect to $\{\tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_n\}$. Let $y = Ax$ and $\tilde{y} = B\tilde{x}$. We have

$$\tilde{y} = Ty = TAx = B\tilde{x} = BTx$$

and therefore

$$A = T^{-1}BT \, .$$

Then we say A is *similar* to B.

Definition (Similar matrices)

Two square matrices A and B are similar if there exists an invertible matrix T such that $A = T^{-1}BT$.

When <u>A is similar to B</u>, both matrices correspond to the <u>same linear transformation</u> but expressed in different bases.

An important case is when the operator can be represented by a <u>diagonal</u> matrix. This is called the eigenrepresentation.

Eigenvalues (evas) and Eigenvectors (eves)

Let $A \in \mathbb{R}^{n \times n}$ (square). A scalar λ (possibly complex) and a nonzero vector v satisfying the equation

$$Av = \lambda v$$

are said to be, respectively, an eigenvalue and an eigenvector of A.
We have

$$\begin{aligned} Ax - \lambda x &= 0 \\ (A - \lambda I)x &= 0\,. \end{aligned}$$

This shows that any eigenvector x of A is in the null space of $(A - \lambda I)$.

Since eigenvectors are non-zero vectors, this means that A will have eigenvectors iff the null space of $A - \lambda I$ is nontrivial. The only way that $\mathrm{null}(A - \lambda I)$ can be nontrivial is if $(A - \lambda I)$ is singular.

Eigenvalues (evas) and Eigenvectors (eves)

That is, λ is an eigenvalue of A iff

$$\det(A - \lambda I) = 0 \tag{1}$$

This leads to an *n*th-order polynomial equation

$$\det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + ... + a_1\lambda + a_0 = 0\,,$$

where det denotes the determinant of the matrix $\lambda I - A$.

By solving this equation, we obtain the scalars (eigenvalues) $\lambda_1, ..., \lambda_n$.

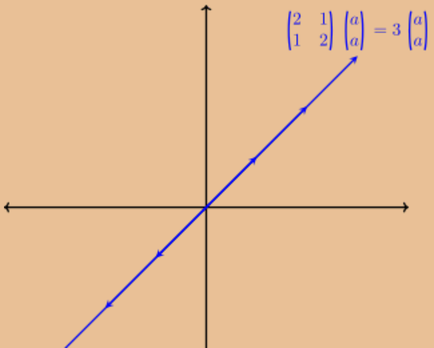
According to the fundamental theorem of algebra, the characteristic equation must have n (possibly nondistinct) roots that are the eigenvalues of A provided that we allow complex numbers. This is why sometimes eigenvalues and their corresponding eigenvectors involve complex numbers. Nevertheless, we will see later why in this course we do not need complex numbers.

evas and eves - Interpretation

Let $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Given a vector x in $\mathbb{R}^2$, the vector Ax is also in $\mathbb{R}^2$.

For many vectors x, Ax will <u>not</u> be pointing in the same direction as x. But for some other vectors, only the magnitudes are changed. See the blue vectors in this geogebra demo: `https://www.geogebra.org/3d/yqpy3jf5`

We observe that vectors parallel to $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ change length but not direction under the linear transformation, that is $Ax = cx$. Those are eigenvectors.



evas and eves - Interpretation

To verify this algebraically, observe that all vectors parallel to $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ can be written in the form $\begin{pmatrix} a \\ a \end{pmatrix}$, for $a \neq 0$. Now compute

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

This shows that any non-zero scalar multiple of $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector of A which has a corresponding eigenvalue of 3.

evas and eves - Computation examples

Example: Consider the matrix

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

To find the eigenvalues, we need to solve the characteristic equation:

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

That is,

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

Computing the determinant we get the polinomial equation:

$$(3-\lambda)(2-\lambda) - (2)(1) = \lambda^2 - 5\lambda + 6 - 2 = \lambda^2 - 5\lambda + 4\,.$$

Factoring:

$$(\lambda - 4)(\lambda - 1) = 0\,.$$

evas and eves - Computation examples

From this, we get the eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 1$.
To find the eigenvector associated with $\lambda_1 = 4$:

$$(A - 4I)v_1 = 0$$

Which gives:

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

From the first equation, we get: $-x + 2y = 0$ or $x = 2y$. Thus, a simple eigenvector for $\lambda_1 = 4$ is $v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

Similarly, for $\lambda_2 = 1$, you can compute and get the eigenvector $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

So, the matrix $A = \begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix}$ has integer eigenvalues $4, 1$ and corresponding integer eigenvectors $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ respectively.

evas and eves - Computation examples

Note: In general, one does not attempt to compute eigenvalues by solving the characteristic equation, as there is no simple way to solve such an equation for $n > 4$. Instead, one can use iterative methods (approximations which are accurate enough).

Review video: `https://youtu.be/PFDu9oVAE-g`

evas and eves - Computation examples

Example 2: Let

$$C = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

So,

$$\det(C - \lambda I) = \det \begin{pmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{pmatrix}$$

Next, we compute the determinant:

$$(2-\lambda)(2-\lambda)(2-\lambda)+1\cdot1\cdot1+1\cdot1\cdot1-1\cdot(2-\lambda)\cdot1-1\cdot1\cdot(2-\lambda)-(2-\lambda)\cdot1\cdot1$$

Expanding and simplifying:

$$= (2-\lambda)^3 + 2 - 3(2-\lambda)$$

$$= -\lambda^3 + 6\lambda^2 - 9\lambda + 4$$

evas and eves - Computation examples

Factoring [5] we get

$$= -(\lambda - 4)(\lambda - 1)^2 .$$

This gives the roots

$$\lambda_1 = 4 \quad \text{(with algebraic multiplicity 1)}$$

$$\lambda_2 = 1 \quad \text{(with algebraic multiplicity 2)}$$

[5] Use Polynomial Division: Divide the cubic polynomial by $(\lambda - 1)$ using polynomial long division or synthetic division. The result is a quadratic polynomial:

$$-\lambda^2 + 5\lambda - 4$$

Proposition

Let T be a triangular matrix. Then the eigenvalues of T are the entries on the main diagonal. In other words, if

$$T = \begin{pmatrix} a & * & * & \cdots & * \\ 0 & b & * & \cdots & * \\ 0 & 0 & c & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & z \end{pmatrix},$$

where the $*$ entries can be any real numbers, then the eigenvalues of T are $a, b, c, \dots, z$. Same applies for the transpose.

Proposition

Let D be a diagonal matrix. Then, the eigenvalues of D are the entries on its main diagonal. In other words, if

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

then the eigenvalues of D are $d_1, d_2, \dots, d_n$.

Task: Choose 2×2 and 3×3 matrices as decribed above, and compute its evas.

Theorem (good news)

All eigenvalues of a real symmetric matrix are real.

Definition (Eigenspace)

The set of all eigenvectors associated with a given eigenvalue of a matrix, together with the zero vector, is known as the eigenspace associated with that eigenvalue.

So, given an eigenvalue λ, there is an associated eigenspace S, and our goal is to find a basis of S. Then, any eigenvector x associated with λ can be expressed as a linear combination of the vectors in that basis. Moreover, we are trying to find a basis for the set of vectors that satisfy

$$(A - \lambda I)x = 0,$$

which means we seek a basis for $\text{null}(A - \lambda I)$. This entails finding the solutions to the homogeneous system of linear equations represented by the above equation, and the non-zero solutions will form the eigenspace associated with λ.

A central goal of linear algebra is to show that a given operator can be represented by a simple matrix, under a certain basis.

Proposition (Diagonalizable / A is similar to a diagonal matrix)

Let $A \in \mathbb{R}^{n\times n}$. Then A is said to be diagonalizable if there exists an invertible matrix P such that

$$P^{-1}AP = D,$$

where D is a diagonal matrix.

Equivalently, this relationship can be rewritten as

$$AP = PD\,. \tag{*}$$

Specifically, the product PD results in multiplying each column of P by the corresponding diagonal entry of D.

Diagonalization

Now,

$$PD = \begin{pmatrix} | & | & & | \\ p_1 & p_2 & \cdots & p_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix} = \begin{pmatrix} | & | & & | \\ d_1 p_1 & d_2 p_2 & \cdots & d_n p_n \\ | & | & & | \end{pmatrix}.$$

On the other hand,

$$AP = A \begin{pmatrix} | & | & & | \\ p_1 & p_2 & \cdots & p_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ Ap_1 & Ap_2 & \cdots & Ap_n \\ | & | & & | \end{pmatrix}.$$

Finally,

$$\begin{pmatrix} | & | & & | \\ Ap_1 & Ap_2 & \cdots & Ap_n \\ | & | & & | \end{pmatrix} = \begin{pmatrix} | & | & & | \\ d_1 p_1 & d_2 p_2 & \cdots & d_n p_n \\ | & | & & | \end{pmatrix}.$$

Diagonalization

For two matrices to be equal, in particular we need each column to be equal. In other words, if we call x_i the i-th column of P and λ_i the i-th diagonal entry of D, then

$$Ax_i = \lambda_i x_i \,.$$

This is the eigenvalue equation.

In summary, associated with every eva λ_k we have the subspace

$$S_k = \mathcal{N}\left(A - \lambda_k I\right).$$

Every non-zero element in S_k is an eve corresponding to λ_k.

Proposition

The matrix A is diagonalizable iff

$$S_1 + S_2 + .. + S_k = \mathbb{R}^n \,.$$

In other words, A is diagonalizable if there are enough $\ell.\imath.$ eigenvectors to form a basis for $\mathbb{R}^n$.

Diagonalization

In effect, if eves satisfy

$$\mathrm{Span}\left(\{v_1, v_2, .., v_n\}\right) = \mathbb{R}^n$$

we have

$$AV = (Av_1 \mid Av_2 \mid \dots \mid Av_n) = (\lambda_1 v_1 \ \dots \ \lambda_n v_n) = V\Lambda\,,$$

where $V = (v_1 \mid .. \mid v_n)$ invertible.

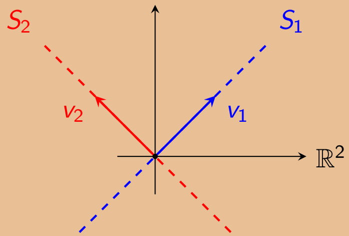


Figure: Matrix $A = \frac{1}{2}\begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$

Diagonalization

Theorem

Suppose the characteristic equation $\det(\lambda I - A) = 0$ *has* n *distinct roots* $\lambda_1, \lambda_2, ..., \lambda_n$. *Then, there exist* n *linearly independent* ($\ell.\imath.$) *vectors* $v_1, v_2, ..., v_n$ *such that*

$$Av_i = \lambda_i v_i, \quad i{=}1, 2, ..., n.$$

In other words, A *is diagonalizable.*

So distinct eigenvalues is a sufficient condition, but not necessary. That is, there are matrices with all eigenvalues equal and still diagonalizable. For example, any diagonal matrix (e.g. the identity matrix) is by definition diagonalizable.

Usage: Consider a basis formed by these $\ell.\imath.$ vectors $v_1, v_2, ..., v_n$. Then, using this basis we can represent A using a diagonal matrix. Let

$$T = (v_1, v_2, ..., v_n).$$

Then

$$\begin{aligned} T^{-1}AT &= T^{-1}A(v_1, v_2, ..., v_n) \\ &= T^{-1}(Av_1, Av_2, ..., Av_n) \\ &= T^{-1}(\lambda_1 v_1, \lambda_2 v_2, ..., \lambda_n v_n) \end{aligned}$$

Finally,

$$T^{-1}AT = T^{-1}T \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Diagonalization - Examples

Example: Let $\mathcal{L} : \mathbb{R}^2 \to \mathbb{R}^2$ by $\mathcal{L}(x, y) = (41x + 7y, -20x + 74y)^\top$.
The associated matrix with respect to the standard basis of $\mathbb{R}^2$ is

$$\begin{pmatrix} 41 & 7 \\ -20 & 74 \end{pmatrix}$$

which is not a diagonal matrix. However, $\mathcal{L}$ is diagonalizable, because the associated matrix with respect to the basis $(1, 4)^\top$, $(7, 5)^\top$ is

$$\begin{pmatrix} 69 & 0 \\ 0 & 46 \end{pmatrix}$$

To verify this, we find the T matrix:

$$T = \begin{pmatrix} 1 & 7 \\ 4 & 5 \end{pmatrix},$$

and

$$T^{-1} = \frac{1}{1 \cdot 5 - 7 \cdot 4} \begin{pmatrix} 5 & -7 \\ -4 & 1 \end{pmatrix} = \frac{1}{-23} \begin{pmatrix} 5 & -7 \\ -4 & 1 \end{pmatrix}.$$

Diagonalization - Examples

Now we write

$$P^{-1}AP = \frac{1}{-23}\begin{pmatrix} 5 & -7 \\ -4 & 1 \end{pmatrix}\begin{pmatrix} 41 & 7 \\ -20 & 74 \end{pmatrix}\begin{pmatrix} 1 & 7 \\ 4 & 5 \end{pmatrix} = \begin{pmatrix} 69 & 0 \\ 0 & 46 \end{pmatrix} = D\,.$$

Review video: `https://youtu.be/13r9QY6cmjc`

Diagonalization - Examples

Example 2: Not every matrix is diagonalizable. Let

$$A = \begin{pmatrix} 1 & 4 \\ 0 & 1 \end{pmatrix}.$$

We see immediately that the eigenvalues of A are $\lambda_1 = 1$ and $\lambda_2 = 1$. The eigenvectors are of the form

$$\begin{pmatrix} t \\ 0 \end{pmatrix} = t \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \text{for } t \in \mathbb{R}.$$

It is easy to see that we cannot form an invertible matrix P, because any two eigenvectors will be of the form $\begin{pmatrix} t \\ 0 \end{pmatrix}$, and so the second row of P would have a row of zeros, and P could not be invertible. Hence A cannot be diagonalized.

Eigenvectors symmetric matrices

Real symmetric matrices, which play a special role in this course, are diagonalizable. But not only that: the eves are mutually orthogonal.

Theorem

Any real symmetric $n \times n$ matrix has a set of n eigenvectors that are mutually orthogonal.

These eigenvectors form an orthogonal basis for $\mathbb{R}^n$. If the basis $\{v_1, v_2, ..., v_n\}$ is normalized so that each element has norm 1, then for

$$T = [v_1, v_2, ..., v_n],$$

we have

$$T^\top T = I,$$

and hence

$$T^\top = T^{-1}.$$

A matrix whose transpose is equal to its inverse is said to be an orthogonal matrix (extremely convenient: compare the complexity of inverting a matrix versus transpose).

evas and eves - Experiment

MATLAB/Octave experiment.

```
n = 8;
Q = randn(n);
% make Q symmetric
Q = (Q+Q')/2;
% since Q is randomly generated, all columns are l.i.
rank(Q)

[X, D] = eig(Q);
% check eigendecomposition (all entries are numerical zeros)
Q - X*D*inv(X);
% since Q is real symmetric, evectors form an orthogonal matrix
id = X'*X;

% clean up numerical zeros
id(abs(id)<1E-14)=0
```

Rank of a matrix

Rank of a matrix provides crucial information for solving a system of equations. The rank is a measure of "non-degeneracy".

In some applications, especially in machine learning, the rank can be interpreted as the amount of "information" or "features" it contains.

Definition

Column-rank of matrix A is the maximal number of $\ell.\imath.$ columns of A.

Definition

Row-rank of matrix A is the maximal number of $\ell.\imath.$ rows of A.

Proposition (Column-rank and row-rank are equal)

The column-rank and the row-rank are equal.

Definition (Rank of $A \in \mathbb{R}^{m\times n}$)

The rank of $A \in \mathbb{R}^{m\times n}$ is the maximal number of $\ell.\imath.$ columns of A.

Rank of a matrix

Proposition

$$\operatorname{rank}(A) \leq \min(m, n)\,.$$

Definition (Full rank matrix)

A matrix A such that $\operatorname{rank}(A) = \min(m, n)$ *is said to have full rank; otherwise, the matrix is rank deficient.*

A has full row-rank (column-rank), if $\operatorname{rank}(A) = m$ ($\operatorname{rank}(A) = n$).

Proposition

Only the zero matrix has rank zero.

Computing rank using Gaussian elimination

Let

$$A = \begin{pmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{pmatrix} \qquad \begin{matrix} (R1) \\ (R2) \\ (R3) \end{matrix}$$

can be put in reduced row-echelon form by using the following elementary row operations:

$$\begin{pmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{pmatrix} \xrightarrow{2R_1+R_2\to R_2} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 3 & 5 & 0 \end{pmatrix} \xrightarrow{-3R_1+R_3\to R_3} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & -1 & -3 \end{pmatrix}$$

$$\xrightarrow{R_2+R_3\to R_3} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{-2R_2+R_1\to R_1} \begin{pmatrix} 1 & 0 & -5 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix} .$$

The final matrix (in row echelon form) has two non-zero rows and thus $\operatorname{rank}(A) = 2$.

Linear transformation and rank of a matrix

Let $\mathcal{L}$ be the linear transformation $\mathcal{L}(x) = Ax$.

Proposition (Column rank - dimension of column space)

The rank of A is the dimension of the image of $\mathcal{L}$, which is spanned by the columns of A.

The image of $\mathcal{L}$ is the set of all possible vectors you can get in $\mathbb{R}^m$ after applying $\mathcal{L}$ to every vector in $\mathbb{R}^n$: $\text{Image}(\mathcal{L}) = \{\mathcal{L}(x) : x \in \mathbb{R}^n\}$

Proposition (Injective $\mathcal{L}$ and rank)

$\mathcal{L}$ is injective iff A has rank n (in this case, we say that A has full column rank).

Proposition (Surjective $\mathcal{L}$ and rank)

$\mathcal{L}$ is surjective iff A has rank m (in this case, we say that A has full row rank).

Proposition (Rank and invertibility)

If $A \in \mathbb{R}^{n\times n}$, then A is invertible iff A has rank n (that is, A has full rank).

Rank and nullspace

Recall the nullspace of A is the set of vectors

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}.$$

Proposition (Injective $\mathcal{L}$ and nullspace)

$\mathcal{L}$ is injective iff $\mathcal{N}(A)$ contains only the 0 vector.

In other words, $\mathcal{L}$ is injective iff ($Ax = 0$ implies $x = 0$).

Rank can be easily computed using MATLAB with the `rank` command. Then we can determine if the linear transformation represented by A is injective or surjective. Also we can determine the dimension of null, and the dimension of the image of the transformation.

Rank and linear systems

Another useful use of `rank` is the following:

Proposition

A necessary and sufficient condition for a linear system of equations to be consistent is that the rank of the coefficient matrix A is equal to the rank of the augmented matrix. That is

$$Ax = b \quad \textit{iff} \quad \operatorname{rank}(A) = \operatorname{rank}\left(\left(A, b\right)\right).$$

The matrix $\left(A, b\right)$ is called the augmented matrix and denoted A_b.

Rank and linear systems

Proposition

Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Suppose the system $Ax = b$ is compatible and that $\operatorname{rank}(A) = \operatorname{rank}(A_b) = k$

(a) *If $k < m$ (rank is less than the number of equations m), then $m - k$ equations are superfluous in the sense that if we choose any subsystem of equations corresponding to k $\ell.\imath.$ rows, then any solution of these k equations also satisfies the remaining $m - k$ equations.*

(b) *If $k < n$ (rank is less than the number of unknowns n), then there exist $n - k$ variables that can be chosen freely, whereas the remaining k variables are uniquely determined by the choice of these $n - k$ free variables. The system then has $n - k$ degrees of freedom.*

Example 1: (Redundant equaiton) Consider

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 3 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix},$$

or

$$\begin{cases} x_1 + 2x_2 = 1 \\ 2x_1 + x_2 = 2 \\ 3x_1 + 3x_2 = 2 \end{cases} .$$

Observe last row in the augmented matrix is the sum of the other two, which are $\ell.\imath.$. Using the first two rows we find $x_1 = x_2 = 1/3$, which also satisfies the last equation.

Also observe the inverse is

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

Rank and linear systems

Example 2: Consider

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

The augmented matrix is

$$A_b = \begin{pmatrix} 1 & 2 & 3 & | & 1 \\ 2 & 4 & 6 & | & 2 \end{pmatrix}.$$

The second row is a multiple of the first row, so they are not $\ell.\imath.$. Thus, $\text{rank}\,(A) = \text{rank}\,(A_b) = 1$.
Since $k = 1$, $n = 3$, and $k < n$, we have $n - k = 3 - 1 = 2$ variables that can be chosen freely.
Let's choose the variables x_2 and x_3 freely, and express x_1 in terms of x_2 and x_3.
From the first equation: $x_1 + 2x_2 + 3x_3 = 1$, we get $x_1 = 1 - 2x_2 - 3x_3$.
Here, $x_2, x_3 \in \mathbb{R}$, and they will determine the value of x_1.

Rank and linear systems

Example 3: Let

$$\begin{aligned} x_1 + x_2 - 2x_3 + x_4 + 3x_5 &= 1 \\ 2x_1 - x_2 + 2x_3 + 2x_4 + 6x_5 &= 2 \\ 3x_1 + 5x_2 - 10x_3 - 3x_4 - 9x_5 &= 3 \\ 3x_1 + 2x_2 - 4x_3 - 3x_4 - 9x_5 &= 3 \end{aligned}$$

We have the reduced form

$$A = \begin{pmatrix} 1 & 1 & -2 & 1 & 3 & | & 1 \\ 2 & -1 & 2 & 2 & 6 & | & 2 \\ 3 & 5 & -10 & -3 & -9 & | & 3 \\ 3 & 2 & -4 & -3 & -9 & | & 3 \end{pmatrix} \quad \to \quad \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & | & 1 \\ 0 & 1 & -2 & 0 & 0 & | & 0 \\ 0 & 0 & 0 & 1 & 3 & | & 0 \\ 0 & 0 & 0 & 0 & 0 & | & 0 \end{pmatrix},$$

so rank is 3. If we remove the third row, all remaining are $\ell.\imath.$ (check it). We can also remove 4th row to get 3 $\ell.\imath.$.

Rank and linear systems

We re-write the system by exchanging the order of the sum:

$$\begin{cases} x_1 - 2x_3 + x_4 + x_2 + 3x_5 = 1 \\ 2x_1 + 2x_3 + 2x_4 - x_2 + 6x_5 = 2 \\ 3x_1 - 4x_3 - 3x_4 + 2x_2 - 9x_5 = 3 \end{cases},$$

or in matrix form

$$\begin{pmatrix} 1 & -2 & 1 \\ 2 & 2 & 2 \\ 3 & -4 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} 1 & 3 \\ -1 & 6 \\ 2 & -9 \end{pmatrix} \begin{pmatrix} x_2 \\ x_5 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Rank and linear systems

The 3×3 coefficient matrix corresponding to x_1, x_3, and x_4 has a determinant different from 0, so it has an inverse.

Now we solve for x_1, x_3, x_4 as a function of the free variables:

$$\begin{pmatrix} x_1 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 & -2 & 1 \\ 2 & 2 & 2 \\ 3 & -4 & -3 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 1 & -2 & 1 \\ 2 & 2 & 2 \\ 3 & -4 & -3 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 3 \\ -1 & 6 \\ 2 & -9 \end{pmatrix} \begin{pmatrix} x_2 \\ x_5 \end{pmatrix}.$$

It is easy to verify that

$$\begin{pmatrix} 1 & -2 & 1 \\ 2 & 2 & 2 \\ 3 & -4 & -3 \end{pmatrix}^{-1} = \frac{1}{18} \begin{pmatrix} -1 & 5 & 3 \\ -6 & 3 & 0 \\ 7 & 1 & -3 \end{pmatrix}.$$

Rank and linear systems

Finally,

$$\begin{pmatrix} x_1 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ -\frac{1}{2}x_2 \\ 3x_5 \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1}{2}x_2 \\ -3x_5 \end{pmatrix} .$$

So if $x_2 = a$ and $x_5 = b$ are arbitrary real numbers, then there is a solution $x_1 = 1$, $x_2 = a$, $x_3 = \frac{1}{2}a$, $x_4 = -3b$, and $x_5 = b$. This confirms that there are two degrees of freedom. (You should verify that the values found for x_1, x_2, x_3, x_4, and x_5 do satisfy the original system of equations for all values of a and b.)

Rank facts

When we multiply matrices, the rank cannot increase.

Here are five key facts for inequalities and equalities for the rank:

1. $\text{rank}(AB) \leq \text{rank}(A)$
2. $\text{rank}(AB) \leq \text{rank}(B)$
3. $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$
4. $\text{rank}(A^\top A) = \text{rank}(AA^\top) = \text{rank}(A) = \text{rank}(A^\top)$
5. If $A \in \mathbb{R}^{m \times r}$ and $B \in \mathbb{R}^{r \times n}$ (both with rank r) then $\text{rank}(AB) = r$

Quadratic Forms

We now study the second most important function in this course.
A quadratic form $f : \mathbb{R}^n \to \mathbb{R}$ is a function

$$f(x) = x^\top Q x,$$

where Q is an $n \times n$ real matrix. There is no loss of generality in assuming Q to be symmetric, that is, $Q = Q^\top$. For if the matrix Q is not symmetric, we can always replace it with the symmetric matrix

$$Q_0 = Q_0^\top = \frac{1}{2}(Q + Q^\top).$$

Note that

$$x^\top Q x = x^\top Q_0 x = x^\top \left(\frac{1}{2} Q + \frac{1}{2} Q^\top \right) x.$$

Note: Since we are considering real symmetric matrices, we know the eigenvalues are real (recall "good news" theorem).

Quadratic Forms - Example

Consider a general quadratic function in $\mathbb{R}^2$:

$$f(x, y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} .$$

This expands to:

$$f(x, y) = ax^2 + (b + c)xy + dy^2 .$$

Let

$$a = 1, b = 1, c = 2, \text{ and } d = 4 ,$$

so Q is not symmetric.
Our quadratic function is:

$$f(x, y) = x^2 + 3xy + 4y^2$$

The non-symmetric matrix Q corresponding to this quadratic is:

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

Can we represent the same f using a symmetric matrix? Yes. Choose

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

Quadratic Forms

Definition (Positive (semi)definite)

A quadratic form $x^\top Qx$ *for* Q *symmetric is said to be positive definite if* $x^\top Qx > 0$ *for all nonzero vectors* x*. It is positive semidefinite if* $x^\top Qx \geq 0$ *for all* x*.*

Note: Every pd form is also psd. The converse is not true.

Definition (Negative (semi)definite)

A quadratic form $x^\top Qx$ *for* Q *symmetric is said to be negative definite if* $x^\top Qx < 0$ *for all nonzero vectors* x*. It is negative semidefinite if* $x^\top Qx \leq 0$ *for all* x*.*

Note: Every nd is also a nsd. The converse is not true.

Example: The matrix in the previous example is pd.

We now present a criterion to classify a matrix as pd. We first recall what are the leading principal minors: For $Q = \begin{pmatrix} q_{11} & q_{12} & \dots & q_{1n} \\ q_{21} & q_{22} & \dots & q_{2n} \\ \vdots & \vdots & & \vdots \\ q_{n1} & q_{n2} & \dots & q_{nn} \end{pmatrix}$

the leading principal minors are

$$\Delta_1 = q_{11}\,, \Delta_2 = \det \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}, \Delta_3 = \det \begin{pmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{pmatrix}, .., \Delta_n = \det Q$$

Theorem (Sylvester's Criterion)

A quadratic form $x^\top Qx$ for Q is pd iff the leading principal minors of Q are positive.

Note: If Q is not symmetric, Sylvester's criterion cannot be used to check positive definiteness of the quadratic form $x^\top Qx$. However, considering $Q_0 = \frac{1}{2}(Q + Q^\top)$ the criterion can be used for any quadratic form.

Example: Consider $Q = \begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix}$. Minors are positive, however if $x = (1,1)^\top$, $x^\top Qx = -2$. However, consider $Q_0 = \frac{1}{2}(Q + Q^\top) = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ -4 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & -4 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}$ and apply the criterion.

Note: A necessary condition for a quadratic form to be psd is that the leading principal minors be nonnegative. However, this is not a sufficient condition. Example: The indefinite matrix $\begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ has $\Delta_k \geq 0$.

Positive definite matrix

Definition (Positive definite matrix Q)

A real symmetric matrix Q is said to be pd if the quadratic form $x^\top Qx$ is pd.

Similarly, we define a symmetric matrix Q to be psd.

Definition

A real symmetric matrix Q is indefinite if it is neither psd nor nsd.

Proposition

A real symmetric matrix Q is pd (or psd) iff the matrix $-Q$ is nd (or nsd).

Positive definite matrix and diagonalization

Through diagonalization, we can show that a symmetric psd matrix Q has a psd (symmetric) square root $Q^{1/2}$ satisfying $Q^{1/2}Q^{1/2} = Q$. For this, we use T as above and define

$$Q^{1/2} = T \begin{pmatrix} \lambda_1^{1/2} & & & 0 \\ & \lambda_2^{1/2} & & \\ & & \ddots & \\ 0 & & & \lambda_n^{1/2} \end{pmatrix} T^\top .$$

Note that a quadratic form $x^\top Qx$ can be expressed as $\|Q^{1/2}x\|^2$. (Later we will define norms).

Some properties

- Let Q be pd. Then the diagonal elements are positive.
- Let Q be psd. Then the diagonal elements are nonnegative.
- Let Q be a symmetric matrix. If there exist positive and negative elements in the diagonal, then Q is indefinite.
- Let Q be pd. Then Q has inverse and is pd.
- Let $A \in \mathbb{R}^{m \times n}$, $m > n$. Then $A^\top A$ is pd iff $\operatorname{rank}(A) = n$. If $\operatorname{rank}(A) < n$ then $A^\top A$ is psd. This expression will be important later, when solving least-squared problems.
- The $n \times n$ zero matrix is psd and nsd.

Examples

$$Q_1 = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

The diagonal elements are all positive. Conclusion?

$$Q_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

All diagonal elements are non-negative. Conclusion?

$$Q_3 = \begin{pmatrix} 2 & 1 & 0 \\ 1 & -3 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

It's a symmetric matrix with both positive and negative elements on the diagonal, so it's indefinite.

Examples

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{pmatrix}$$

$A_1^T A_1$ is 2-by-2 symmetric pd because A_1 has full column rank (i.e., rank 2). The matrix is invertibe $A_1^T A_1$.

$$A_2 = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix}$$

$A_2^T A_2$ is a 3-by-3 matrix psd because A_2 does not have full column rank (i.e., rank less than 3). In this case 0 is an eigenvalue and the matrix is not invertible.

Examples

Another pd matrix:

$$Q_4 = \begin{pmatrix} 5 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 3 \end{pmatrix}$$

All diagonal elements are positive.
Another indefinite matrix with eigenvalue 0:

$$Q_5 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Another Indefinite matrix:

$$Q_6 = \begin{pmatrix} -1 & 2 & 0 \\ 2 & 1 & 2 \\ 0 & 2 & -1 \end{pmatrix}$$

Note: The definiteness of these matrices can be determined by computing eigenvalues (see next slide).

2×2 matrix definiteness

Consider the important case

$$Q = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

and let

$$q(x, y) = \begin{pmatrix} x & y \end{pmatrix} Q \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + 2bxy + cy^2.$$

If $a \neq 0$, then we may complete the square to obtain

$$\begin{aligned} q(x, y) &= a\left(x^2 + \frac{2b}{a}xy\right) + cy^2 = a\left(\left(x + \frac{b}{a}y\right)^2 - \frac{b^2}{a^2}y^2\right) + cy^2 \\ &= a\left(x + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2 = a\left(x + \frac{b}{a}y\right)^2 + \frac{ac - b^2}{a}y^2 \\ &= a\left(x + \frac{b}{a}y\right)^2 + \frac{\det(Q)}{a}y^2. \end{aligned}$$

2×2 matrix definiteness

Suppose $\det(Q) > 0$: Then we have that $q(x, y) > 0$ for all $(x, y) \neq (0, 0)$ if $a > 0$ and $q(x, y) < 0$ for all $(x, y) \neq (0, 0)$ if $a < 0$. That is, Q is pd if $a > 0$ and nd if $a < 0$.

Suppose $\det(Q) < 0$, then $q(1, 0)$ and $q\left(-\frac{b}{a}, 1\right)$ will have opposite signs, and so Q is indefinite.

Suppose $\det(Q) = 0$: Then

$$q(x, y) = a\left(x + \frac{b}{a}y\right)^2,$$

so $q(x, y) = 0$ when $x = -\frac{b}{a}y$. Moreover, $q(x, y)$ has the same sign as a for all other values of (x, y). Hence, in this case Q is psd if $a > 0$ and nsd if $a < 0$.

Finally, if $\det(Q) = 0$ and $a = 0$ we have 3 cases:

- If $c > 0$ then Q is psd.
- If $c < 0$ then Q is nsd.
- If $c = 0$ then Q is the zero matrix and thus psd and nsd.

$2{\times}2$ matrix definiteness

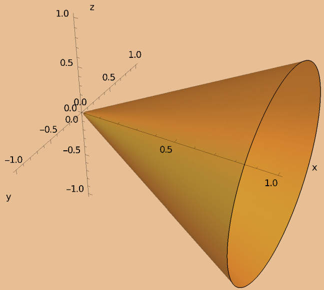
An alternative parameterization is

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} = \begin{pmatrix} x+y & z \\ z & x-y \end{pmatrix}.$$

(take $z = b$, $x = \frac{a+c}{2}$, and $y = \frac{a-c}{2}$).

In this parameterization the condition for the matrix to be pd is $x > \sqrt{y^2+z^2}$. Task: derive this condition from $\det(Q) > 0$ and $a > 0$.

This describes a convex cone in the 3-d space x, y, and z:



Matrix definiteness and eigenvalues

Let $x^\top Qx$ be a quadratic form, and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the (real) eigenvalues of Q. Then:

(a) Q is positive definite $\Leftrightarrow \lambda_1 > 0,\ \lambda_2 > 0, \ldots, \lambda_n > 0$

(b) Q is positive semidefinite $\Leftrightarrow \lambda_1 \geq 0,\ \lambda_2 \geq 0, \ldots, \lambda_n \geq 0$

(c) Q is negative definite $\Leftrightarrow \lambda_1 < 0,\ \lambda_2 < 0, \ldots, \lambda_n < 0$

(d) Q is negative semidefinite $\Leftrightarrow \lambda_1 \leq 0,\ \lambda_2 \leq 0, \ldots, \lambda_n \leq 0$

(e) Q is indefinite $\Leftrightarrow A$ has both positive and negative eigenvalues

Matrix definiteness and eigenvalues

$$Q1 = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \text{ evas: } 3,\ 3,\ -1.$$

$$Q2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \text{ evas: } 0,\ 2,\ 3.$$

$$Q3 = \begin{pmatrix} 2 & 1 & 0 \\ 1 & -3 & 1 \\ 0 & 1 & 2 \end{pmatrix}, \text{ evas: } \tfrac{-1-\sqrt{33}}{2},\ \tfrac{-1+\sqrt{33}}{2},\ 2.$$

$$Q4 = \begin{pmatrix} 5 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 3 \end{pmatrix}, \text{ evas: } 4+\sqrt{3},\ 4,\ 4-\sqrt{3}.$$

$$Q5 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \text{ evas: } -\sqrt{2},\ \sqrt{2},\ 0.$$

$$Q6 = \begin{pmatrix} -1 & 2 & 0 \\ 2 & 1 & 2 \\ 0 & 2 & -1 \end{pmatrix}, \text{ evas: } -3,\ 3,\ -1.$$

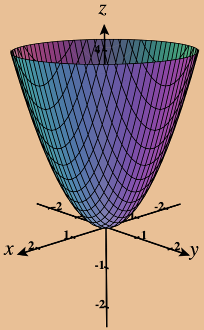
Quadratic form graphs

In the following slides we show the shape of "canonical" $\mathbb{R}^2$ quadratic form graphs. The graphs of other quadratic forms look similar, though they may be stretched in various directions. Later we will study quadratic functions.

Task: Visualize the following quadratic forms in your favourite plotting software. Also try examples where Q is non-diagonal.

Quadratic form graphs - pd - "basin" or "bowl"

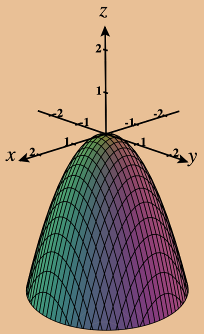
The quadratic form $z \equiv f(x, y) = x^2 + y^2$ is pd. Since $Q = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, its eigenvalue is 1 (duplicated). Recall the e-vas of a diagonal matrix is the diagonal.



Note here is a strict (unique) minimum at the origin.

Quadratic form graphs - nd

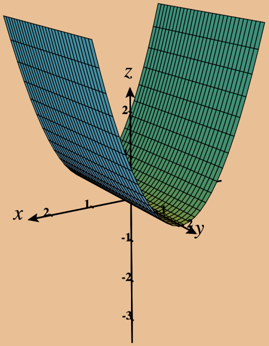
The quadratic form $f(x, y) = -x^2 - y^2$ is nd. Since $Q = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$, its eigenvalue is -1 (duplicated).



Note there is a strict (unique) maximum at the origin.

Quadratic form graphs - psd - "creek"

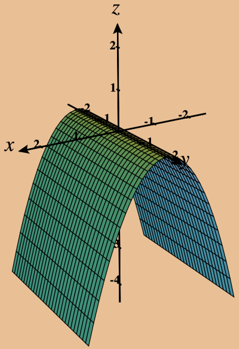
The quadratic form $f(x, y) = x^2$ is psd. Since $Q = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, its eigenvalues are 0 and 1.



Note there is a non-strict (non-unique) minimum at the origin.

Quadratic form graphs - nsd

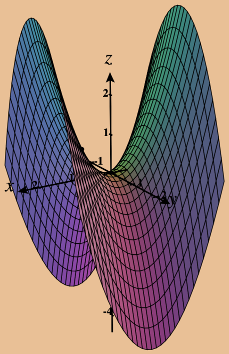
The quadratic form $f(x,y) = -x^2$ is nsd. Since $Q = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$, its eigenvalues are 0 and -1.



Note there is a non-strict (non-unique) maximum at the origin.

Quadratic form graphs - indefinite - "pringle chip"

The quadratic form $p(x, y) = x^2 - y^2$ is indefinite. Since $Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, its eigenvalues are 1 and -1.



Note there is no maximum nor minimum at the origin. This type of point is called saddle.

References

Online: Introduction to Linear Algebra, Vladimir Dobrushkin.
`https://www.cfm.brown.edu/people/dobrush/cs52/index.html`

Videos: Linear Algebra Done Right, Sheldon Axler.
`https://www.youtube.com/@sheldonaxler5197/videos`

- Axler, S. (2014).
  Linear Algebra Done Right.
  Springer.
- Bazaraa, M. S. and Shetty, C. M. (1979).
  Nonlinear Programming: Theory and Algorithms.
  Wiley, 1 edition.
- Beck, A. (2014).
  Introduction to Nonlinear Optimization.
  MOS-SIAM Series on Optimization. SIAM.
- Bertsekas, D. P. (1999).
  Nonlinear Programming.
  Athena Scientific, 2 edition.
- Boyd, S. and Vandenberghe, L. (2004).
  Convex Optimization.
  Cambridge University Press.
- Chong, E. K. P. and Żak, S. H. (2013).
  An Introduction to Optimization.
  Wiley, 4 edition.
- Fletcher, R. (1987).
  Practical Methods of Optimization.
  John Wiley & Sons, New York, NY, USA, second edition.
- Nocedal, J. and Wright, S. (2006).
  Numerical Optimization.