Pills of Linear Algebra

Davide Galassi

Table of Contents

Vectors

A vector is a geometric entity endowed with **magnitude** and **direction** expressed as a tuple $\langle v_1, \ldots, v_n \rangle$ splitting the entire quantity in its orthogonal axis components.

An *n*-dimensional vector can be written as $\bar{v} = \langle v_1, \ldots, v_n \rangle$, where the numbers v_i are called **elements** of the vector $\bar{v}$.

The components are usually labeled with the same name of the axis to which they corresponds.

A vector $\bar{v}$ can be expressed as a column matrix $\bar{v} = \begin{pmatrix} v_1 \\ \ldots \\ v_n \end{pmatrix}$ or as a row matrix $\bar{v}^T = (v_1 \ldots v_n)$.

Two vectors $\bar{v}$ and $\bar{w}$ are **equal** if and only if the corresponding components are equal.

Euclidean geometric definition

If P and Q are two distinct points in the *xy*-plane there is exactly one **line** passing through P and Q. The points part of the line that joins P to Q form a **line segment** $\overline{PQ}$. If we order the points so that they proceed from P to Q we have a **directed line segment** $\overrightarrow{PQ}$, or a **geometric vector**.

Each vector component is set as the difference between the components of the start and end points:

$$\bar{v} = \overrightarrow{PQ} = \langle Q_1 - P_{1,} \ldots, Q_n - P_n \rangle$$

If $\bar{v}$ is a vector whose initial point is at the origin, then $\bar{v}$ is called a **position vector**. The terminal point of a position vector $\bar{v} = \langle v_1, \ldots, v_n \rangle$ is the point $T = (v_1, \ldots, v_n)$.

If $\bar{v}$ is a vector with initial point $P = (p_1, \ldots, p_n)$, not necessarily the origin, and terminal point $Q = (q_1, \ldots, q_n)$, then $\bar{v} = \overrightarrow{PQ}$ is equal to the position vector $\bar{v} = \langle q_1 - p_1, \ldots, q_n - p_n \rangle$.

By this definition we can replace any geometrically defined vector $\overrightarrow{PQ}$ with a position vector $\bar{v}$.

Arithmetic

Addition

The addition of two vectors $\bar{v}$ and $\bar{w}$ of the same size n is defined as the addition of the respective components

$$\bar{v} + \bar{w} = \langle v_1 + w_1, \ldots, v_n + w_n \rangle$$

Properties

i. *Closure*: $\bar{v}+\bar{w} \in R^n$ (grupoid)

ii. *Associativity*: $\bar{u}+(\bar{v}+\bar{w})=(\bar{u}+\bar{v})+\bar{w}$ (semigroup)

iii. *Identity*: $\exists \bar{x} \in R^n = \bar{0}$: $\bar{v}+\bar{0}=\bar{v}$ (monoid)

iv. *Inverse*: $\forall \bar{x} \in R^n \; \exists \bar{y} \in R^n$: $\bar{x}+\bar{y}=\bar{0}$ (group)

v. *Commutativity*: $\bar{v}+\bar{w}=\bar{w}+\bar{v}$ (abelian group)

With respect to the addition, vector space is an ***Abelian group***

The **difference** between $\bar{v}$ and $\bar{w}$ is equivalent to the addition of $\bar{v}$ with the negative of $\bar{w}$, obtained by inverting the sign of all the components.

$$\bar{v}-\bar{w}=\bar{v}+(-\bar{w})$$

Addition of vector $\bar{v}$ with its opposite vector $-\bar{v}$ yields the identity: $\bar{v}+(-\bar{v})=\bar{0}$

Scalar multiplication

Vectors with a single component are defined as **scalars**.

If α is a scalar and $\bar{v}$ is a vector then $\alpha \cdot \bar{v}$ is a vector where each component of $\bar{v}$ is multiplied by α

$$\alpha \cdot \bar{v}=\langle \alpha \cdot v_1, \ldots, \alpha \cdot v_n \rangle$$

$\bar{0}=0 \cdot \bar{v}$

$-\bar{v}=(-1) \cdot \bar{v}$

$\bar{v}=1 \cdot \bar{v}$

Distributive property

$$(\alpha+\beta) \cdot \bar{v}=\langle (\alpha+\beta) \cdot v_1, \ldots, (\alpha+\beta) \cdot v_n \rangle=\langle \alpha \cdot v_1+\beta \cdot v_1, \ldots, \alpha \cdot v_n+\beta \cdot v_n \rangle=\alpha \bar{v}+\beta \bar{v}$$

$$\alpha \cdot (\bar{v}+\bar{w})=\alpha \cdot \langle v_1+w_1, \ldots, v_n+w_n \rangle=\langle \alpha v_1+\alpha w_1, \ldots, \alpha v_n+\alpha w_n \rangle=\alpha \bar{v}+\alpha \bar{w}$$

Dot product

Given two vectors $\bar{v}$ and $\bar{w}$, the **dot product** is defined as

$$\bar{v} \cdot \bar{w}=v_1 w_1+\ldots+v_n w_n=\sum_{i=1:n} v_i w_i$$

Because the result is a scalar it is also referred as **scalar product**.

Properties

i. Commutativity: $\bar{v} \cdot \bar{w} = \bar{w} \cdot \bar{v}$

ii. Distributive: $\bar{u} \cdot (\bar{v} + \bar{w}) = \bar{u} \cdot \bar{v} + \bar{u} \cdot \bar{w}$

iii. $\bar{v} \cdot \bar{v} = \|\bar{v}\|^2$

iv. $\bar{0} \cdot \bar{v} = 0$

Proofs

i. $\bar{v} \cdot \bar{w} = \sum_{i=1:n} v_i w_i = \sum_{i=1:n} w_i v_i = \bar{w} \cdot \bar{v}$

ii. $\bar{u} \cdot (\bar{v} + \bar{w}) = \sum_{i=1:n} u_i (v_i + w_i) = \sum_{i=1:n} u_i v_i + u_i w_i = \sum_{i=1:n} u_i v_i + \sum_{i=1:n} u_i w_i = \bar{u} \cdot \bar{v} + \bar{u} \cdot \bar{w}$

iii. $\bar{v} \cdot \bar{v} = \sum_{i=1:n} v_i v_i = \|\bar{v}\|^2$ (refer to the magnitude paragraph)

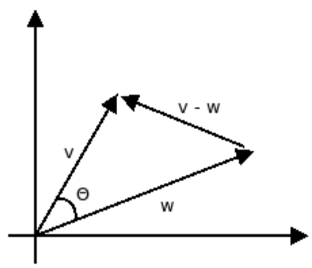
iv. $\bar{0} \cdot \bar{v} = \sum_{i=1:n} 0 v_i = 0$

Geometric interpretation

In Euclidean space, the dot product between $\bar{v}$ and $\bar{w}$ is defined as

$$\bar{v} \cdot \bar{w} = \|\bar{v}\| \|\bar{w}\| \cos \Theta$$

where Θ is the angle between $\bar{v}$ and $\bar{w}$.



The sides of the triangle, formed by the vectors, have lengths $\|\bar{v}\|$, $\|\bar{w}\|$ and $\|\bar{v} - \bar{w}\| = \|\bar{w} - \bar{v}\|$.

Using the law of cosines: $\|\bar{v} - \bar{w}\|^2 = \|\bar{v}\|^2 + \|\bar{w}\|^2 - 2\|\bar{v}\| \|\bar{w}\| \cos \Theta$.

Applying the property iii: $(\bar{v} - \bar{w}) \cdot (\bar{v} - \bar{w}) = \bar{v} \cdot \bar{v} + \bar{w} \cdot \bar{w} - 2\|\bar{v}\| \|\bar{w}\| \cos \Theta$.

Applying the distributive property to the left-hand side:

$$(\bar{v} - \bar{w}) \cdot (\bar{v} - \bar{w}) = (\bar{v} - \bar{w}) \cdot \bar{v} - (\bar{v} - \bar{w}) \cdot \bar{w} = \bar{v} \cdot \bar{v} - 2\bar{v} \cdot \bar{w} + \bar{w} \cdot \bar{w}.$$

Combining the equations we finally get

$$\bar{v} \cdot \bar{v} - 2\bar{v} \cdot \bar{w} + \bar{w} \cdot \bar{w} = \bar{v} \cdot \bar{v} + \bar{w} \cdot \bar{w} - 2\|\bar{v}\| \|\bar{w}\| \cos \Theta \Rightarrow \bar{v} \cdot \bar{w} = \|\bar{v}\| \|\bar{w}\| \cos \Theta.$$

The angle Θ, $0 \le \Theta \le \pi$, between two vectors $\bar{v}$ and $\bar{w}$ is determined by $\Theta = \arccos\left(\frac{\bar{v} \cdot \bar{w}}{\|\bar{v}\| \|\bar{w}\|}\right)$.

Magnitude

The **magnitude** of an n-dimensional vector $\bar{v}$ is denoted as $\|\bar{v}\|$ and is defined as

$$\|\bar{v}\|=\sqrt{v_1^2+\ldots+v_n^2}=\sqrt{\sum_{i=1}^{n} v_i^2}$$

And is the distance from the initial to the terminal vector points.

A vector $\bar{0}$ for which $\|\bar{0}\|=0$ is known as **zero vector**. Note that by the definition all components should be equal to zero. A vector $\bar{u}$ for which $\|\bar{u}\|=1$ is known as **unit vector**.

The **negative** of a vector $\bar{v}$ is $-\bar{v}$ and has the same magnitude of $\bar{v}$ but with the opposite direction. It is obtained by negating each component sign.

Properties

i. $\|\bar{v}\|\geq 0$

ii. $\|\bar{v}\|=0 \Leftrightarrow \bar{v}=\bar{0}$

iii. $\|\alpha\cdot\bar{v}\|=|\alpha|\cdot\|\bar{v}\|$

iv. $\|\bar{v}\|=\|-\bar{v}\|$

Proofs

i. Follows the fact that the magnitude is the square root of sums of squares

ii. If $\|\bar{v}\|=0$ then from the magnitude definition each component has zero value, thus $\bar{v}=\bar{0}$. If $\bar{v}=\bar{0}$ then the the magnitude definition gives a zero value.

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

iv. $\|-\bar{v}\|=\sqrt{\sum_{i=1:n}(-1)^2 v_i^2}=\sqrt{\sum_{i=1:n} v_i^2}=\|\bar{v}\|$

Theorem. For any non-zero vector $\bar{v}$ the vector $\bar{u}=\bar{v}/\|\bar{v}\|$ is a unit vector with same direction of $\bar{v}$.

Proof. Given that $\alpha=\frac{1}{\|\bar{v}\|}>0$ then $\left\|\frac{\bar{v}}{\|\bar{v}\|}\right\|=\|\alpha\cdot\bar{v}\|=|\alpha|\cdot\|\bar{v}\|=\frac{1}{\|\bar{v}\|}\cdot\|\bar{v}\|=1$.

Follows that if $\bar{u}$ is a unit vector with same direction as $\bar{v}$ we can write $\bar{v}=\bar{u}\cdot\|\bar{v}\|$.

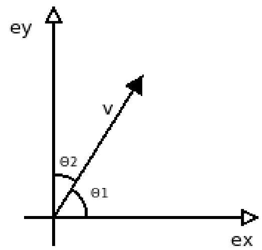
The transformation of a vector into a unit vector with the same direction is called **normalization**.

Direction / Magnitude form

In *Euclidean* space a vector $\bar{v}$ can be described in terms of magnitude and direction, rather than in terms of components.

Direction cosines: the cosines of the angles between the vector and the coordinate axis.

For example, in the 2-dimensional *xy-plane*

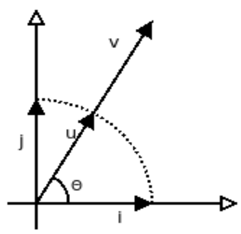


$$\cos\Theta_1 = \frac{\bar{v}}{\|\bar{v}\|}\cdot e_x = \sin\Theta_2$$

$$\cos\Theta_2 = \frac{\bar{v}}{\|\bar{v}\|}\cdot e_y = \sin\Theta_1$$

The direction angles are between *0* and 2π radians.

To express $\bar{v}$ in terms of $\|\bar{v}\|$ and Θ we first need to find the unit vector with the same direction as $\bar{v}$; that is $\bar{u} = \bar{v}/\|\bar{v}\|$.



The coordinates of $\bar{u}$ are thus $\cos\Theta$ and $\sin\Theta$ then follows that $\bar{u} = \cos\Theta\cdot i + \sin\Theta\cdot j$.

Also note that since $\bar{u} = \bar{v}/\|\bar{v}\| = \langle\cos\Theta, \sin\Theta\rangle$ then

$$\bar{u}\cdot\bar{i} = \langle\cos\Theta, \sin\Theta\rangle\cdot\langle 1, 0\rangle = \cos\Theta$$

$$\bar{u}\cdot\bar{j} = \langle\cos\Theta, \sin\Theta\rangle\cdot\langle 0, 1\rangle = \sin\Theta$$

Parallel and orthogonal vectors

Two vectors $\bar{v}$ and $\bar{w}$ are **parallel** if and only if there is a non-zero scalar α such that $\bar{v} = \alpha\bar{w}$. In this case the angle between them is 0 or π .

Two vectors $\bar{v}$ and $\bar{w}$ are **orthogonal** if and only if the angle α between them is $\pi/2$.

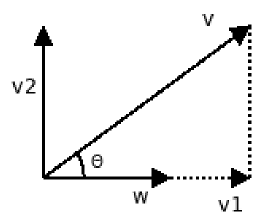
Theorem. The vectors $\bar{v}$ and $\bar{w}$ are orthogonal if and only if $\bar{v}\cdot\bar{w} = 0$

Proof. ($\Rightarrow$) Since $\bar{v}\cdot\bar{w} = \|\bar{v}\|\|\bar{w}\|\cos\Theta$ and $\Theta = \pi/2$ we have that $\bar{v}\cdot\bar{w} = \|\bar{v}\|\|\bar{v}\|\cos\pi/2 = 0$.

($\Leftarrow$) If $\bar{v} \cdot \bar{w} = 0$ then we have $\cos\Theta = 0$ or one of the two vectors is the vector $\bar{0}$. If $\cos\Theta = 0$ then $\Theta = \pi/2$ and the vectors are orthogonal. If one of the vectors is $\bar{0}$ we have that they are orthogonal since, by convention, $\bar{0}$ is assumed orthogonal to every other vector.

Vector projection

The **projection** of a vector $\bar{v}$ into a vector $\bar{w}$, also called the vector component of $\bar{v}$ into the direction of $\bar{w}$) is the **orthogonal projection** of $\bar{v}$ onto a straight line parallel to $\bar{w}$.



$$\|\bar{v}_1\| = \cos\Theta \|\bar{v}\|$$

$$v_1 = \frac{\bar{w}}{\|\bar{w}\|} \cos\Theta \|\bar{v}\|$$

$$\bar{v} \cdot \bar{w} = \|\bar{v}\| \|\bar{w}\| \cos\Theta \;\Rightarrow\; \cos\Theta \|\bar{v}\| = \frac{\bar{v} \cdot \bar{w}}{\|\bar{w}\|}$$

So we finally have: $v_1 = \frac{\bar{w}}{\|\bar{w}\|} \frac{\bar{v} \cdot \bar{w}}{\|\bar{w}\|} = \frac{\bar{v} \cdot \bar{w}}{\|\bar{w}\|^2} \cdot \bar{w} = \alpha \bar{w}$

Note that, because $\alpha = \frac{\bar{v} \cdot \bar{w}}{\|\bar{w}\|^2}$ is a scalar, from the parallel definition $\bar{w}$ and $\bar{v}_1$ are parallel.

The orthogonal component $\bar{v}_2$ of $\bar{v}$ with respect to $\bar{v}_1$ is trivially given by $\bar{v}_2 = \bar{v} - \bar{v}_1$. We've decomposed $\bar{v}$ in two orthogonal vectors.

Basis vectors

Given *n* vectors in $\mathbb{R}^n$ if any other vector in $\mathbb{R}^n$ can be uniquely expressed as a linear combination of them, then they are referred to as a **basis** for the vector space $\mathbb{R}^n$. The basis components of a *n*-dimensional space can be written as $\{\bar{e}_i : 1 \le i \le n\}$.

Every vector in a *n*-dimensional space can be uniquely written as

$$\bar{v} = v_1 \bar{e}_1 + \ldots + v_n \bar{e}_n$$

If the basis vectors are unit vectors then they are called **versors**.

If the versors are mutually **orthogonal** they are referred as a **standard basis**.

Note that the equality $\bar{v} = v_1 \bar{e}_1 + \ldots + v_n \bar{e}_n = \langle v_1, \ldots, v_n \rangle$ holds if and only if each versor $\bar{e}_i$ is part of the a **standard basis**, that is a vector with all components set to 0 except the *i-th* that is set to 1. In our discussion we assume that the versors are part of a standard basis.

In 3-dimensional space, the standard basis are defined as

$$\bar{i}=\langle 1,0,0\rangle,\ \bar{j}=\langle 0,1,0\rangle,\ \bar{k}=\langle 0,0,1\rangle$$

Given the above definitions, every vector in 3-dimensional space can be uniquely written as

$$\bar{v}=v_x\bar{i}+v_y\bar{j}+v_z\bar{k}=\langle v_x,0,0\rangle+\langle 0,v_y,0\rangle+\langle 0,0,v_z\rangle=\langle v_x,v_y,v_z\rangle$$

We can easily define the vector arithmetic operations in term of components

$$\bar{v}+\bar{w}=(v_1\bar{e}_1+\ldots+v_n\bar{e}_n)+(w_1\bar{e}_1+\ldots+w_n\bar{e}_n)=(v_1+w_1)\bar{e}_1+\ldots+(v_n+w_n)\bar{e}_n=\langle v_1+w_1,\ldots,v_n+w_n\rangle$$

$$\alpha\bar{v}=\alpha\cdot(v_1\bar{e}_1+\ldots+v_n\bar{e}_n)=\alpha v_1\bar{e}_1+\ldots+\alpha v_n\bar{e}_n=\langle \alpha v_1,\ldots,\alpha v_n\rangle$$

Linear dependence and independence

The vectors in a subset $S=\{\bar{v}_1,\ldots,\bar{v}_n\}$ of a vector space V as $\mathbb{R}^n$ are **linearly dependent** if there exist a set of scalars $\{a_1,\ldots,a_n\}$ not all zeros such that $\bar{0}=a_1\bar{v}_1+\ldots+a_n\bar{v}_n$.

In such a case, at least one element is not zero, say a_1, and the equation can be written as

$$\bar{v}_1=\frac{-a_2}{a_1}\bar{v}_2+\ldots+\frac{-a_n}{a_1}\bar{v}_n$$

Thus one of the vectors can be expressed as a **linear combination** of the others.

The vectors are said to be **linearly independent** if the equation $\bar{0}=a_1\bar{v}_1+\ldots+a_n\bar{v}_n$ is satisfied if and only if $a_i=0$ for $i=1:n$. Thus if they are not linearly dependent.

Geometrically, two vectors $\bar{v}$ and $\bar{w}$ are linearly dependent if one is **parallel** to the other. That is easily seen since in case of linear dependence $\bar{v}=a\bar{w}$ for some non-zero scalar α.

Given three vectors all lying on the same plane, if two of them are not parallel then those two vectors **spans** the entire plane. The other vector is thus a linear combination of them and the vectors are linearly dependent.

If the three vectors don't all lie in the same plane through the origin, none is in the span of the other two, so none is a linear combination of the other two. The three vectors are then linearly independent.

Theorem. If $\{\bar{v}_1,\ldots,\bar{v}_n\}$ are orthogonal and $\bar{v}_i\neq\bar{0},\ \forall i=1:n$ then they are linearly independent.

Proof. The dot product between two vectors is defined as $\bar{v}\cdot\bar{w}=|\bar{v}||\bar{v}|\cos(\alpha)$, with α the angle between them. Thus if they are orthogonal then $\bar{v}\cdot\bar{w}=|\bar{v}||\bar{v}|0=0$, while if $\bar{v}=\bar{w}$ then $\bar{v}\cdot\bar{w}=|\bar{v}||\bar{v}|1=|\bar{v}|^2$. If they are linearly dependent then there are $\{a_1,\ldots,a_n\}$ not all zeros such that $\bar{0}=a_1\bar{v}_1+\ldots+a_n\bar{v}_n$. Multiplying both sides by an arbitrary vector $\bar{v}_i$ with $1\leq i\leq n$ yields $(a_1\bar{v}_1+\ldots+a_n\bar{v}_n)\cdot\bar{v}_i=a_i|\bar{v}_i|^2=0$. Because we've assumed that $\forall i\ \ \bar{v}_i\neq\bar{0}$ then $a_i=0$. The procedure can be repeated for each v_i, resulting that $a_i=0\ \forall i$, that is absurd.

Note that lineal independence doesn't imply orthogonality. To be linear independent is sufficient that no other can be expressed as a linear combination of the others. As an example, in 3-dimensional space the vectors $\bar{i}=\langle 1,0,0\rangle$, $\bar{j}=\langle 0,1,0\rangle$ and $\bar{v}=\langle 2,1,3\rangle$ are not orthogonal but are linearly independent, that is we cannot express $\bar{v}$ as a linear combination of $\bar{i}$ and $\bar{j}$. Note that because of their linear independence the three vectors spans the entire $\mathbb{R}^3$, in other words we can express any arbitrary vector $\bar{x}\in\mathbb{R}^3$ as a linear combination of the three.

Theorem. Every basis for the **vector space** $\mathbb{R}^n$ consists of *n* linearly independent vectors.

Theorem. For any vectors $\bar{v}_1,\ldots,\bar{v}_n$ the following conditions are equivalent

- $\{\bar{v}_1,\ldots,\bar{v}_n\}$ is a **basis** for $\mathbb{R}^n$
- $\{\bar{v}_1,\ldots,\bar{v}_n\}$ is a **spanning set** for $\mathbb{R}^n$
- $\{\bar{v}_1,\ldots,\bar{v}_n\}$ is a linearly independent set

All bases for a vector space *V* has the same cardinality. The **dimension** of a vector space *V*, denoted *dim V*, is the cardinality of its bases. For example, $\mathbb{R}^n$ has cardinality *n*.

Matrices

A matrix M is defined as a **bidimensional array** of numbers:

$$M=\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Each matrix **entry** a_{ij} has two indices: a row index (*i*) and column index (*j*).
A matrix with *m* rows and *n* columns is called an $m\times n$ matrix. If $m=n$ the matrix is **square**.

Equality. Two matrices *A* and *B* are equal if *A* and *B* have the same number of rows and columns and each entry a_{ij} in *A* is equal to b_{ij} in *B*.

Arithmetic

Addition

If *A* and *B* have the same dimensions, then the sum *A*+*B* is a matrix *C* with the same dimensions as *A* and *B* where $c_{ij}=a_{ij}+b_{ij}$. The difference is equal to the sum but with the sign inverted in the elements of the second matrix: $A-B=A+(-B)$.

Properties. Given three matrices $A, B, C\in R^{m\times n}$ then the following properties holds

i. *Closure*: $A+B\in R^{m\times n}$
ii. *Associative*: $A+(B+C)=(A+B)+C$
iii. *Neutral element*: there is $Z\in R^{m\times n}$ such that $A+Z=A$
iv. *Inverse element*: for each A there is $N\in R^{m\times n}$ such that $A+N=Z$, with *Z* the neutral element
v. *Commutative*: $A+B=B+A$

With the above properties the matrix set is an **Abelian group**.

Scalar multiplication

If $M \in R^{m \times n}$ and k is a scalar in R, then the matrix $kA \in R^{m \times n}$ is obtained by multiplying each entry of M by k. The matrix kA is a **scalar multiple** of A.

Properties. Given $A \in R^{m \times n}$ and $k, h \in R$

i. $k(h(A)) = (kh)A$
ii. $(k+h)A = kA + hA$
iii. $k(A+B) = kA + kB$

Multiplication

A **row vector** is a $1 \times n$ matrix: $\bar{r} = (r_1 \dots r_n)$

A **column vector** is a $n \times 1$ matrix $\bar{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$

The product between $\bar{r}$ and $\bar{c}$ is equal to a 1×1 matrix, or a scalar, $\bar{r}\,\bar{c} = r_1 c_1 + \dots + r_n c_n$.

In practice, the result is equal to the vectors **dot product**.

Matrix product. Let $A \in R^{m \times r}$ and $B \in R^{r \times n}$. The product AB is a matrix $C \in R^{m \times n}$ whose elements $(AB)_{ij}$ are the dot product of the ***i***-th row of A and the ***j***-th column of B.

If $\bar{A}_i \in R^r$ is the ***i***-th row of A and the $\bar{B}_j \in R^r$ is the ***j***-th column of B, both of length n then

$$(AB)_{ij} = \sum_{k=1}^{r} A_{ik} B_{kj} = \bar{A}_i \cdot \bar{B}_j$$

Properties. Given $A \in R^{m \times r}$, $B \in R^{r \times n}$, $C \in R^{n \times w}$ and a scalar $k \in R$

i. $A(BC) = (AB)C \quad \in R^{m \times w}$
ii. $k(AB) = (kA)B \quad \in R^{m \times n}$

Proof.

i. $$[A(BC)]_{ij} = \sum_{k=1}^{r} A_{ik} (BC)_{kj} = \sum_{k=1}^{r} A_{ik} \left(\sum_{z=1}^{n} B_{kz} C_{zj}\right) = \sum_{z=1}^{n} \sum_{k=1}^{r} (A_{ik} B_{kz}) C_{zj} = \sum_{z=1}^{n} (AB)_{iz} C_{zj} = [(AB)C]_{ij}$$

ii. $$[k(AB)]_{ij} = k(AB)_{ij} = k \sum_{z=1}^{r} A_{iz} B_{zj} = \sum_{z=1}^{r} (kA)_{iz} B_{zj} = [(kA)B]_{ij}$$

The full proof of (i) requires to prove the equality $\sum_{k=1}^{r} A_{ik}(\sum_{z=1}^{n} B_{kz} C_{zj}) = \sum_{z=1}^{n} \sum_{k=1}^{r} (A_{ik} B_{kz}) C_{zj}$, this can be trivially done by expanding the sums on both sides.

Proposition. *The distributive laws hold*

i. $A(B+C) = AB + AC$
ii. $(B+C)A = BC + BA$

Proof. We prove only the first one. The second follows a similar argument.

$$[A(B+C)]_{ij} = \sum_{k=1}^{n} A_{ik}(B+C)_{kj} = \sum_{k=1}^{n} A_{ik}(B_{kj} + C_{kj}) = \sum_{k=1}^{n} A_{ik} B_{kj} + \sum_{k=1}^{n} A_{ik} C_{kj} = AB_{ij} + AC_{ij}$$

The argument is valid because the distributive law holds in *R*.

Transpose. The transpose of an $m \times n$ matrix *M* is an $n \times m$ matrix M^T for which $M_{ij} = M^T_{ji}$.

Theorem. $(AB)^T = B^T A^T$

Proof. $(AB)^T_{ij} = (AB)_{ji} = \sum_{k=1}^{n} A_{jk} B_{ki} = \sum_{k=1}^{n} A^T_{kj} B^T_{ik} = \sum_{k=1}^{n} B^T_{ik} A^T_{kj} = (B^T A^T)_{ij}$

Diagonal matrix. In a square matrix *M* the entries for which *i=j* are called the **main diagonal** entries of *M.* A square matrix whose only non-zero entries appear on the main diagonal is a diagonal matrix.

Identity matrix. A diagonal matrix whose diagonal entries are 1. Often denoted as $I_n \in R^{n \times n}$.

Proposition. If $M \in R^{m \times n}$, then $I_m M = M$ and $M I_n = M$.

Proof. $(IM)_{ij} = \sum_{k=1}^{m} I_{ik} M_{kj}$. If $i=k$ then $I_{ik} = I_{ii} = 1$ and $I_{ij} M_{kj} = M_{ij}$. If $i \neq k$ then $I_{ik} = 0$ and $I_{ik} M_{kj} = 0$. Thus $(IM)_{ij} = \sum_{k=1}^{n} I_{ik} M_{kj} = M_{ij}$.

The identity matrix is the **neutral element** with respect to the matrix multiplication in $R^{n \times n}$. If $M \in R^{n \times n}$ then $I_n M = M I_n = M$.

Inverse Matrix. Let $M \in R^{n \times n}$. If there exist $M^{-1} \in R^{n \times n}$ such that $M M^{-1} = M^{-1} M = I_n$ then M^{-1} is called the inverse of *M*. If a matrix has no inverse is called **singular**.

Theorem. *A matrix possessing a row or a column consisting of all zeros is singular.*
Proof. If $M \in R^{n\times n}$ has a row $\bar{r}_i$ consisting of all zeros and is not singular, then there exist M^{-1} such that $M M^{-1} = I$. Then $1 = I_{ii} = (M M^{-1})_{ii} = \sum_{k=1}^{n} M_{ik} M_{ki}^{-1} = \sum_{k=1}^{n} 0 M_{ki}^{-1} = 0$, that is impossible so *M* must be singular.

Theorem. *M is invertible if and only if* M^T *is invertible.*
Proof. If *M* is invertible there exist M^{-1} such that $M M^{-1} = M^{-1} M = I$. Because $I = I^T$ we have that $M M^{-1} = I = I^T = (M M^{-1})^T = (M^{-1})^T M^T$ and $M^{-1} M = I = I^T = (M^{-1} M)^T = M^T (M^{-1})^T$.
Since $M^T (M^{-1})^T = (M^{-1})^T M^T = I$, M^T is invertible and $(M^T)^{-1} = (M^{-1})^T$ is its inverse.
Conversely if M^T is invertible then $M^T (M^T)^{-1} = I = I^T = [M^T (M^T)^{-1}]^T = [(M^T)^{-1}]^T [M^T]^T = M^{-1} M$ and, similarly, $(M^T)^{-1} M^T = I = I^T = [(M^T)^{-1} M^T]^T = [M^T]^T [(M^T)^{-1}]^T = M M^{-1}$.
Since $M [(M^T)^{-1}]^T = [(M^T)^{-1}]^T M$, *M* is invertible and $M^{-1} = [(M^T)^{-1}]^T$ is its inverse.

Corollary. $(M^T)^{-1} = (M^{-1})^T$
Proof. Follows from the above theorem proof.

Theorem. *If A and B are invertible matrices then AB is invertible and* $(AB)^{-1} = B^{-1} A^{-1}$.
Proof. $(AB)(B^{-1} A^{-1}) = A (B B^{-1}) A^{-1} = A I A^{-1} = A A^{-1} = I$ and similarly $(B^{-1} A^{-1})(AB) = I$.

Corollary. *The set of the matrix in* $R^{n\times n}$ *is a ring with identity (rarely commutative).*
Proof. All the required properties were already proved.

Elementary operations

There are three kinds of elementary matrix operations. When these operations are performed on rows they are called **elementary row operations**; and then they are performed on columns they are called **elementary column operations**.

Focusing on row operations, we have:
Row switching. A row within the matrix is switched with another row

$$R_i \leftrightarrow R_j$$

Row multiplication. Each element in a row can be multiplied by a non-zero scalar *a*.

$$aR_i \rightarrow R_i, \quad a \neq 0$$

Row addition. A row can be replaced by the sum of that row and a multiple of another row.

$$R_i + aR_j \rightarrow R_i, \quad a \neq 0, i \neq j$$

Elementary row operations are used in Gaussian elimination to reduce a matrix to **row echelon** form, a technique to find the matrix inverse (if exists) and to solve linear equations.

Elementary Matrix

Each type of elementary operation may be performed on a matrix $M \in R^{n \times n}$ by multiplying it by a special matrix $E \in R^{n \times n}$ called an **elementary matrix**.

When *E* is left multiplied it represent an elementary row operation, while when *E* is right multiplied it represent an elementary column operation.

Theorem. *Let H be the* $n \times n$ *matrix resulting from the performance of an elementary row operation on M. Then* $H = EM$ *, where E is the* $n \times n$ *matrix obtained by performing the same row operation on the identity matrix* I_n *.*

Row swap. Elementary matrix after that *I's* row *r* has been swapped with row *s*:

$$E_{ij} = \begin{cases} I_{ij} & , \quad i \neq r, \ i \neq s \\ I_{sj} & , \quad i = r \\ I_{rj} & , \quad i = s \end{cases}$$

$$EM_{ij} = \begin{cases} M_{ij} & , \quad i \neq r, \ i \neq s \\ M_{sj} & , \quad i = r \\ M_{rj} & , \quad i = s \end{cases}$$

If $i = r$, $EM_{ij} = \sum_{k=1}^{n} E_{ik} M_{kj} = M_{sj}$, because $E_{is} = I_{ir}$ is the only non-zero element.

A similar argument holds for $i = s$.

Row mul. Elementary matrix after that *I's* row *r* is multiplied by a scalar *a:*

$$E_{ij} = \begin{cases} I_{ij} & , \quad i \neq r \\ a\, I_{ij} & , \quad i = r \end{cases}$$

$$EM_{ij} = \begin{cases} M_{ij} & , \quad i \neq r \\ aM_{rj} & , \quad i = r \end{cases}$$

Row add. Elementary matrix after that *I's* row *s* has been multiplied by the scalar *a* and then added to *I's* row *r*.

$$E_{ij}=\begin{cases} I_{ij} & , \quad i\neq r \\ I_{ij}+aI_{sj} & , \quad i=r \end{cases}$$

$$EM_{ij}=\begin{cases} M_{ij} & , \quad i\neq r \\ M_{ij}+aM_{sj} & , \quad i=r \end{cases}$$

Determinant

Definition. The determinant is a scalar value derived from the entries of a square matrix. The determinant of a matrix *M* is denoted as $det(A)$ or $|A|$.

Minor. Given an $n\times n$ matrix *M*, if $M^{\{i,j\}}$ denotes the $(n-1)\times(n-1)$ matrix whose entries consists of the original entries of *M* after deleting the *i*-th rown and the *j*-th column, the (*i*,*j*)-minor is the determinant of $M^{\{i,j\}}$.

Cofactor. The (*i*,*j*)-cofactor, denoted as C_{ij} is defined as the (*i*,*j*)-minor multiplied by $(-1)^{i+j}$.

$$C_{ij}(M)=(-1)^{i+j}det(M^{i,j})$$

Laplace Formula

The formula recursively expresses the determinant of a matrix in term of its minors.

$$det(M)=\sum_{i=1}^{n} M_{ik}C_{ik}(M)=\sum_{j=1}^{n} M_{kj}C_{kj}(M)$$

With the determinant of an 1×1 matrix set as the entry of the matrix itself.

Unfortunately, the *Laplace* expansion complexity grows very quickly with the dimension of the matrix. The number of required operations is of the order of *n!*.

Fortunately, determinants are mainly used as theoretical tools and are rarely calculated explicitly in numerical linear algebra.

Example: 2×2 matrix

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}=aC_{11}+bC_{12}=a|d|+b|c|=ad-bc$$

Example: 3×3 matrix

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aC_{11} + bC_{12} + cC_{13} = a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix} = a(ei - fh) - b(di - fg) + c(dh - eg) =$$

$$= aei + bfg + cdh - ceg - afh - bdi$$

The expansion of a 3×3 matrix using the Laplace formula is also known as the **Sarrus rule**.

Properties

Proposition. *The determinant of a triangular or diagonal matrix is the product of the main diagonal.*

Proposition. *The determinant of the identity matrix is 1.*

Proof. Easily follows from the Laplace formula.

Theorem. *If all entries in a row, or a column, are zeros, then the value of determinant is 0.*

Proof. It the *i*-th row is equal to $\bar{0}$. Expand across the zero row (or column) $det(M) = \sum_{j=1}^{n} M_{ij} C_{ij} = 0$.

Theorem 1. *If any two rows, or columns, of a matrix are interchanged, the value of the determinant changes sign.*

Proof. By induction.

Base case: a 2×2 matrix determinant is defined as:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = -(bc - ad) = -\begin{vmatrix} b & a \\ d & c \end{vmatrix} = -\begin{vmatrix} c & d \\ a & b \end{vmatrix}$$

The determinant is equal to the the negative of the determinant of the same matrix with columns and rows swapped, respectively.

Inductive step: Assuming the result is true for all the $(n-1) \times (n-1)$ matrices, let *G* represents the result of exchanging the rows *r* and *s* of an $n \times n$ matrix *F*. Choosing another row *k* such that $k \neq r$ and $k \neq s$, $G_{kj} = F_{kj}$, we have

$$det(G) = \sum_{j=1}^{n} G_{kj} C_{kj}(G) = \sum_{j=1}^{n} (-1)^{k+j} G_{kj} det(G^{\{k,j\}}) = \sum_{j=1}^{n} (-1)^{k+j} F_{kj} det(G^{\{k,j\}})$$

Since $G^{\{k,j\}}$ is an $(n-1) \times (n-1)$ matrix, we have $det(G^{\{k,j\}}) = -det(F^{\{k,j\}})$.

Thus $det(G) = -det(F)$.

Corollary. *The determinant of a matrix with two identical rows, or columns, is 0.*

Proof. If *M* has two identical rows and we exchange these two rows then no exchange has been made to the matrix but, for the above theorem, the sign changes. Should be $det(M) = -det(M) \rightarrow det(M) = 0$.

Theorem 2. *If any row, or column, of a matrix is multiplied by a non-zero number a, the value of the determinant is also changed by a factor of k.*

Proof. Let G represent the result of multiplying the row i of a matrix F by the scalar k.

$$det(G)=\sum_{j=0}^{n} G_{ij} C_{ij}(G)=\sum_{j=1}^{n} k F_{ij} C_{ij}(F)=k\sum_{j=1}^{n} F_{ij} C_{ij}(F)=k\,det(F)$$

Theorem 3. *Adding a multiple of one row to another row has no effect on the determinant.*

Proof. Let G represent the result of adding a scalar k times row r of a matrix F to a row g of F.

$$det(G)=\sum_{j=1}^{n} G_{gj} C_{gj}(G)=\sum_{i=1}^{n} (k F_{rj}+F_{gj}) C_{gj}(F)= k\sum_{i=1}^{n} F_{rj} C_{gj}(F)+\sum_{i=1}^{n} F_{gj} C_{gj}(F)=$$

$$= k\sum_{i=1}^{n} F_{rj} C_{gj}(F)+det(F)$$

The quantity $\sum_{j=1}^{n} F_{rj} C_{gj}(F)$ is equivalent to the determinant of F with the entries in row g replaced by the entries from row r. Such a matrix has two identical rows (g and r) thus its determinant is zero.

Therefore $det(G)=det(F)$.

Theorem. If E is an elementary matrix and M is an arbitrary matrix of the same size then $det(EM)=det(E)det(M)$.

Proof.

If E is obtained from I by swapping two rows, then EM is obtained from M by swapping two rows. For theorem 1 we have that $det(EM)=-det(M)$ and, because $det(E)=-det(I)=-1$, follows that $det(EM)=det(E)det(M)$.

If E is obtained from I by multiplying a row by a scalar k, then EM is obtained from M by multiplying a row by a scalar k. For theorem 2 we have that $det(EM)=k\,det(M)$ and, since $det(E)=k\,det(I)=k$, follows that $det(EM)=det(E)det(M)$.

If E is obtained from I by adding a multiple of one row to another row then the matrix EM is obtained from M by adding a multiple of one row to another row (of M). For theorem 3 we have that $det(EM)=det(M)$ and, because $det(E)=det(I)=1$, follows that $det(EM)=det(E)det(M)$.

Theorem 1. *A square matrix M is invertible if and only if* $det(M)\neq 0$.

Proof. (relying on the Gauss-Jordan elimination algorithm)

If M is invertible then it can be written as the product of elementary matrices each having a non-zero determinant. Since the determinant of the product of elementary matrices is equal to the product of the determinant of the matrices, then the determinant of M cannot be zero.

Conversely, if the matrix is not invertible then it can be written as the product of elementary matrices and a matrix having a row of zeros (because the rows are linearly dependent). Since the determinant of a matrix possessing a row of zeros is zero, the determinant of the product is zero.

Proposition. *If the square matrix M is singular then for any matrix A with the same dimension of M, AM is singular.*

Proof. If *M* is singular, it can be written as the product of elementary matrices and a matrix *S* having a row of zeros. If *A* is not singular then it can be written as the product of elementary matrices. Thus $AM=E_1\dots E_k S$, and $det(AM)=0$. If A is singular a similar argument shows that $det(AM)=0$.

Theorem. *For two square matrices F and G,* $det(FG)=det(F)det(G)$

Proof. If either *F* or *G* is singular, then *FG* is singular and the equation holds since both sides are zero. Otherwise both *F* and *G* can be factored into elementary matrices and the determinant of *FG* is the product of the determinant of the elementary matrices.

Another proof for Theorem 1. If *M* is invertible then there exist M^{-1} such that $MM^{-1}=I$. Given that $1=det(I)=det(MM^{-1})=det(M)det(M^{-1})$ then both $det(M)$ and $det(M^{-1})$ should be non zero.

Theorem. *If F is an* $n\times n$ *square matrix and* $C_{ji}(F)=C_{ij}(F^T)$ *is the (i-j)-cofactor of* F^T.
Setting $G_{ij}=C_{ji}(F)/det(F)$ *we have that* $G=F^{-1}$.

Proof. $(FG)_{ij}=\sum_{k=1}^{n} F_{ik}G_{kj}=\sum_{k=1}^{n} F_{ik}\frac{C_{jk}(F)}{det(F)}=\frac{1}{det(F)}\sum_{k=1}^{n} F_{ik}C_{jk}(F)$

If $i=j$ then the sum gives the $det(F)$, so $(FG)_{ij}=1$.

If $i\neq j$ the sum gives the determinant of a matrix equal to *F* but with row *j* equal to the entries of row *i*. And we know that the determinant of a matrix with two equal rows is 0, so $(FG)_{ij}=0$.

Follows that $FG=I$.

Using the theorem above we can derive the explicit formulas for the inverse matrices.

Example. 2×2 matrix inverse

$$A=\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad A^T=\begin{pmatrix} a & c \\ b & d \end{pmatrix} \quad A^{-1}=\frac{1}{det(A)}\begin{pmatrix} C_{11}(A^T) & C_{12}(A^T) \\ C_{21}(A^T) & C_{22}(A^T) \end{pmatrix}=\frac{1}{det(A)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Example. 3 times 3 matrix inverse

$$A=\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \quad A^T=\begin{pmatrix} a & d & g \\ b & e & h \\ c & f & i \end{pmatrix}$$

$$A^{-1}=\frac{1}{det(A)}\begin{pmatrix} C_{11}(A^T) & C_{12}(A^T) & C_{13}(A^T) \\ C_{21}(A^T) & C_{22}(A^T) & C_{23}(A^T) \\ C_{31}(A^T) & C_{32}(A^T) & C_{33}(A^T) \end{pmatrix}=\frac{1}{det(A)}\begin{pmatrix} ei-hf & bi-hc & ei-hf \\ di-gf & ai-gc & ai-gc \\ dh-\mathrm{ge} & ah-gb & ae-db \end{pmatrix}$$

Given a matrix M, the matrix of the cofactors of the transpose matrix is called **adjugate** matrix and is written as $adj(M)$.

$$adj(A)=A^{-1}det(A)$$

Theorem. *Given a matrix* $A\in\mathbb{R}^{mxn}$ *then* $rank(A)\leq min(m,n)$.

Proof. (Informal) If $m\leq n$ then the result is pretty obvious and $rank(A)\leq m$. Let's focus on the case where $m<n$. Elementary row operations don't change the rows vector space, that is, the set of linear combinations of the rows. The number of leading ones in the elimination is at most equal to the number of columns, because they fall in distinct columns. So the number of nonzero rows in the row echelon form of *A* (which is the row rank) is at most equal to the number of columns *n*.

Any vector $\bar{v}\in R^n$ when multiplied by a matrix $A\in\mathbb{R}^{mxn}$ with $m<n$, is mapped to R^m vector space, thus loosing information about one or more dimensions.

Decomposition methods

Given a matrix *M*, some methods compute its determinant by writing *M* as a product whose determinants can be more easily computed.

The ***LU* decomposition** expresses *M* in terms of a lower **triangular** matrix *L* and an upper triangular matrix *U*: $M=LU$

Determinant of *M* is thus $det(M)=det(LU)=det(L)\cdot det(U)$ and this can be easily computed given that the determinant of the triangular matrix is the product the respective diagonal entries.