8. Linear mappings and matrices

A mapping *f* from $\mathbb{R}^n$ to $\mathbb{R}^m$ is called *linear* if it fulfills the following two properties:

(1) $f(\vec{a}+\vec{b}) = f(\vec{a}) + f(\vec{b})$ for all $\vec{a}, \vec{b} \in \mathbb{R}^n$

(2) $f(\lambda \vec{a}) = \lambda f(\vec{a})$ for all $\lambda \in \mathbb{R}$ and all $\vec{a} \in \mathbb{R}^n$

Mappings of this sort appear frequently in the applications. E.g., some important geometrical mappings fall into the class of linear mappings: Rotations around the origin, reflections, projections, scalings, shear mappings...

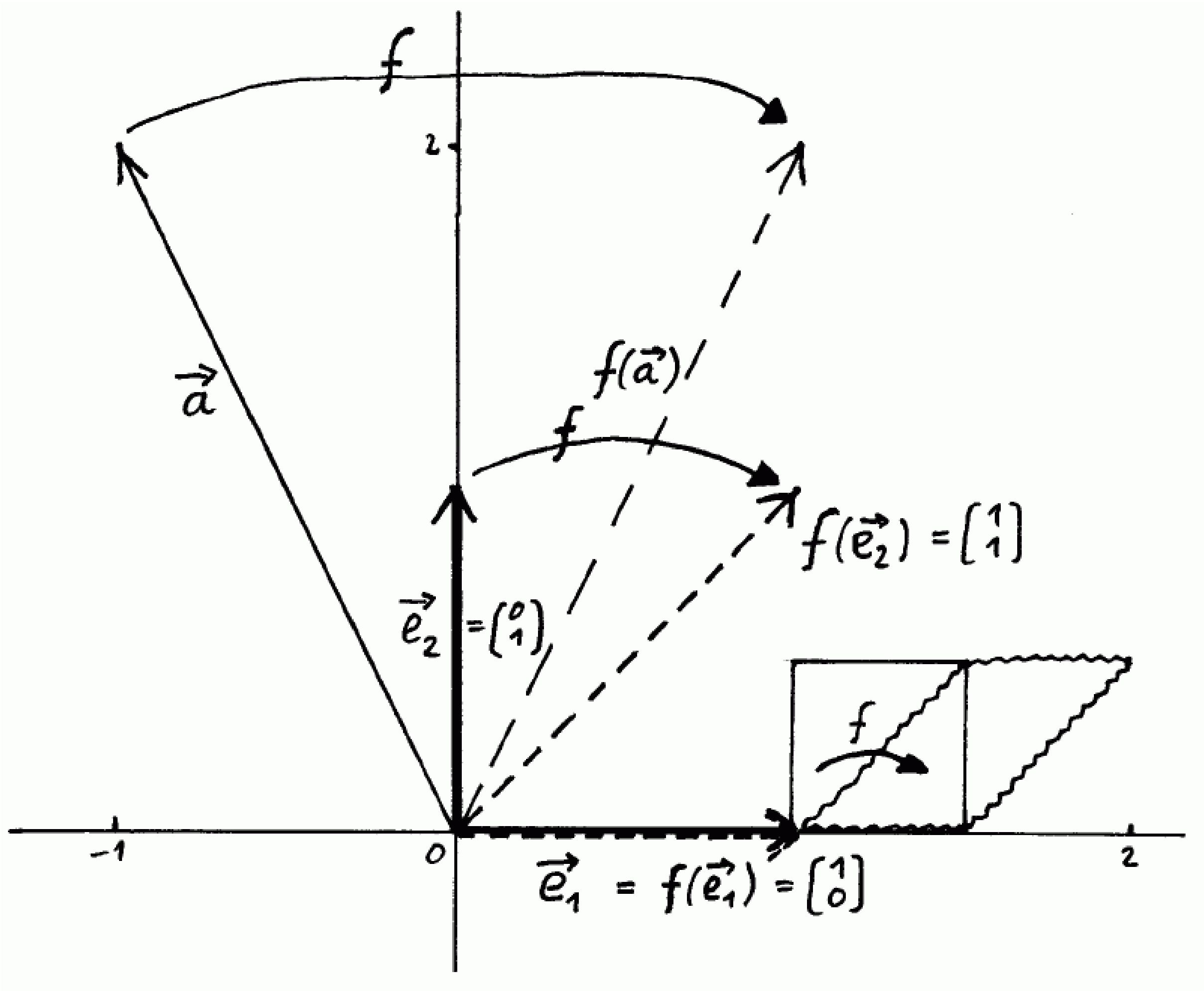
We show at the example of a shear mapping that such a mapping is completely determined (for all input vectors) if its effect on the vectors of the standard basis are known:

Example

Let *f* be the mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$ which performs a *shear* along the *x* axis,

i.e., the image of each point under *f* can be found at the same height as the original point, but shifted along the *x* axis by a length which is proportional (in our example: even equal) to the *y* coordinate.

The figure illustrates the effect of *f* at the examples of the standard basis vectors and an arbitrary vector $\vec{a}$:



We have:

$$f : \mathbb{R}^2 \to \mathbb{R}^2$$
$$f\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}1\\1\end{bmatrix}$$
$$f\begin{bmatrix}0\\2\end{bmatrix} = \begin{bmatrix}2\\2\end{bmatrix} = 2 \cdot f\begin{bmatrix}0\\1\end{bmatrix}$$
$$f\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix}$$

f is indeed a linear mapping, that means:
$f(\vec{a}+\vec{b}) = f(\vec{a}) + f(\vec{b})$ and
$f(c \cdot \vec{a}) = c \cdot f(\vec{a})$ are fulfilled.

The general formula for this shear mapping is apparently:

$$f\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}x+y\\y\end{bmatrix}$$

To get knowledge about the image $f\begin{bmatrix}x\\y\end{bmatrix}$ of an arbitrary vector $\begin{bmatrix}x\\y\end{bmatrix}\in\mathbb{R}^2$, it is sufficient to know the *images of the vectors of the standard basis*, i.e., $f\begin{bmatrix}1\\0\end{bmatrix}$ and $f\begin{bmatrix}0\\1\end{bmatrix}$:

$$\begin{bmatrix}x\\y\end{bmatrix}=x\cdot\begin{bmatrix}1\\0\end{bmatrix}+y\cdot\begin{bmatrix}0\\1\end{bmatrix}=x\cdot\vec{e}_1+y\cdot\vec{e}_2$$

$$\Rightarrow\quad f\begin{bmatrix}x\\y\end{bmatrix}=f(x\cdot\vec{e}_1+y\cdot\vec{e}_2)\overset{f\text{ is linear}}{=}x\cdot f(\vec{e}_1)+y\cdot f(\vec{e}_2)$$

Here: $f\begin{bmatrix}x\\y\end{bmatrix}=x\cdot\begin{bmatrix}1\\0\end{bmatrix}+y\cdot\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}x+y\\y\end{bmatrix}$,
confirming our formula above.

That means: These images, here $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}1\\1\end{bmatrix}$, describe f completely.
They are put together in a *matrix*:

$$\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}$$ = matrix of f.

In general:
Matrix of a linear mapping $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

has m rows and n columns
$\Rightarrow$ "matrix of type (m; n)"

all entries a_{ij} are real numbers

$$\begin{array}{ccccc} \uparrow & \uparrow & & \uparrow \\ \text{image of} & \text{image of} & & \text{image of} \\ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} & \cdots & \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \end{array}$$

The matrix describes its associated linear mapping completely.

The result of the application of f to a vector $\vec{x} \in \mathbb{R}^n$ can easily be calculated as the *product* of the *matrix* of f *with the vector* $\vec{x}$.

In our example:

$$f\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \cdot x + 1 \cdot y \\ 0 \cdot x + 1 \cdot y \end{bmatrix} = \begin{bmatrix} x + y \\ y \end{bmatrix}$$

In the general case:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11} \cdot x_1 + a_{12} \cdot x_2 + \ldots + a_{1n} \cdot x_n \\ \vdots \\ a_{m1} \cdot x_1 + a_{m2} \cdot x_2 + \ldots + a_{mn} \cdot x_n \end{bmatrix}$$

Example: $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 1\cdot 5 + 2\cdot 6 \\ 3\cdot 5 + 4\cdot 6 \end{bmatrix} = \begin{bmatrix} 17 \\ 39 \end{bmatrix}$

General definition of a matrix:

A *matrix* of type (m;n), also: $m\times n$ matrix ("m cross n"), is a system of $m \cdot n$ numbers a_{ij}, $i = 1, 2, ..., m$ and $j = 1, ..., n$, ordered in m rows and n columns:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

a_{ij} is called the *element* or *entry* of the i-th row and the j-th column. The $m \cdot n$ numbers a_{ij} are the *components* of the matrix.
A matrix of type (m; n) has m rows and n columns.
Each row is an n-dimensional vector (row vector), and each column is an m-dimensional column vector.
The list of elements a_{ii} ($i = 1, 2, ..., r$
with $r = \min(m, n)$) is called the *principal diagonal* of the matrix.

Example:

$$A = \begin{bmatrix} 1 & 4 & -3 & 2 \\ 2 & 3 & 0 & -1 \\ -3 & 4 & 1 & 1 \end{bmatrix}$$

A is of type (3; 4).

A has 3 row vectors:

$\vec{z}_1 = (1, 4, -3, 2)$, $\vec{z}_2 = (2, 3, 0, -1)$, $\vec{z}_3 = (-3, 4, 1, 1)$

and four column vectors:

$$\vec{s}_1 = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} , \quad \vec{s}_2 = \begin{bmatrix} 4 \\ 3 \\ 4 \end{bmatrix} , \quad \vec{s}_3 = \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} , \quad \vec{s}_4 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

Its principal diagonal is 1; 3; 1.

Special forms of matrices:

• square matrix:
If $m = n$, i.e., if the matrix *A* has as many rows as it has columns, *A* is called a *square* matrix.

• $m = 1$: A matrix of type (1; *n*) is a row vector.

• $n = 1$: A matrix of type (*m*; 1) is a column vector.

• $m = n = 1$: A matrix of type (1; 1) can be identified with a single real number (i.e., its single entry).

• diagonal matrix:
If *A* is a square matrix and all elements outside the principal diagonal are 0, *A* is called a *diagonal matrix*.

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}$$

• unit matrix:
The unit matrix E is a diagonal matrix where all elements of the principal diagonal are 1.
It plays an important role: Its associated linear mapping is the *identical mapping* $f(\vec{x}) = \vec{x}$.

$$E = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

• zero matrix:
The matrix where *all* entries are 0 is called the zero matrix.

• triangular matrix:
A matrix where all elements below the principal diagonal are 0 is called an *upper triangular matrix*.

Example:

$$A = \begin{bmatrix} 5 & 2 & -1 & 7 \\ 0 & 3 & 1 & 5 \\ 0 & 0 & -1 & 10 \\ 0 & 0 & 0 & 42 \end{bmatrix}$$

Analogous: A matrix where all elements above the principal diagonal are 0 is called a *lower triangular matrix*.

Addition of matrices and *multiplication* of a matrix *with a scalar*:
These operations are defined in the same way as for vectors, i.e., component-wise.

Example:

$$5\cdot\begin{bmatrix}1 & 3\\ 0 & 2\end{bmatrix}+\begin{bmatrix}-1 & 0\\ 7 & 3\end{bmatrix}=\begin{bmatrix}5\cdot 1-1 & 5\cdot 3+0\\ 5\cdot 0+7 & 5\cdot 2+3\end{bmatrix}=\begin{bmatrix}4 & 15\\ 7 & 13\end{bmatrix}$$

Attention: Only matrices of the same type can be added.

Multiplication of a matrix with a column vector:
Defined as above, i.e.,

$$\begin{bmatrix}a_{11} & \cdots & a_{1n}\\ \vdots & & \vdots\\ a_{m1} & \cdots & a_{mn}\end{bmatrix}\cdot\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix} = \begin{bmatrix}a_{11}\cdot x_1+a_{12}\cdot x_2+\ldots+a_{1n}\cdot x_n\\ \vdots\\ a_{m1}\cdot x_1+a_{m2}\cdot x_2+\ldots+a_{mn}\cdot x_n\end{bmatrix}.$$

The result corresponds to the image of the vector under the corresponding linear mapping.

Here, the matrix must have as many columns as the vector has components!

Transposition of a matrix:

Let A be a matrix of type (m; n). The matrix A^T of type (n; m), where its k-th row is the k-th column of A (k = 1, ..., m), is called the *transposed matrix* of A. (Transposition = reflection at the principal diagonal.)

$$\begin{bmatrix}a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{m1} & a_{m2} & \ldots & a_{mn}\end{bmatrix}^T = \begin{bmatrix}a_{11} & a_{21} & \ldots & a_{m1}\\ a_{12} & a_{22} & \ldots & a_{m2}\\ \vdots & \vdots & & \vdots\\ a_{1n} & a_{2n} & \ldots & a_{mn}\end{bmatrix}$$

Example:

$$A = \begin{bmatrix} 1 & 4 \\ -2 & 3 \\ 3 & -1 \end{bmatrix}$$ of type (3; 2) $\Rightarrow$

$$A^T = \begin{bmatrix} 1 & -2 & 3 \\ 4 & 3 & -1 \end{bmatrix}$$ of type (2; 3)

Special case: Transposition of a row vector (type (1; m)) gives a column vector (type (m; 1)), and vice versa.

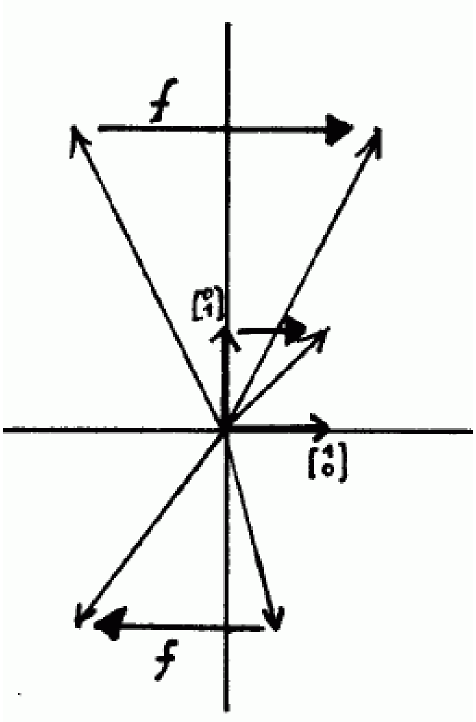
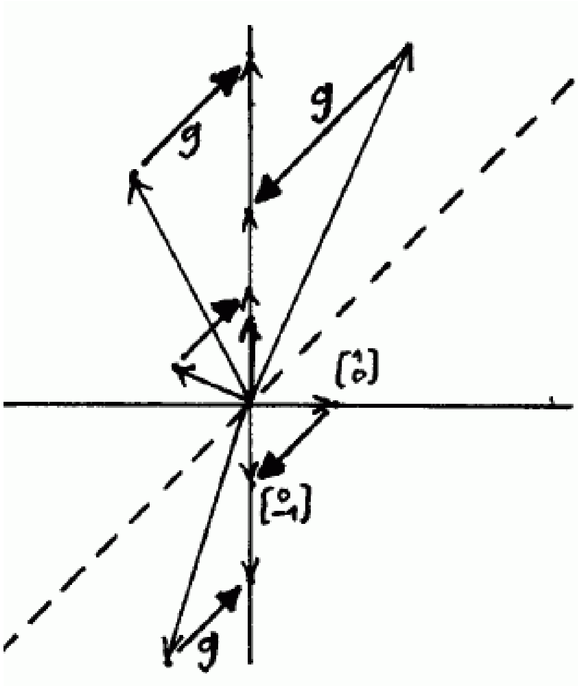
Submatrix:

A *submatrix* of type (m–k; n–p) of a matrix A of type (m; n) is obtained by omitting k rows and p columns from A.
The special submatrix derived from A by omitting the i-th row and the j-th column is sometimes denoted A_{ij} .

We now come back to *linear mappings*, which were our entrance point to motivate the introduction of matrices.
Properties of linear mappings are reflected in numerical attributes of their corresponding matrices.

An important example is the so-called *rank* of a linear mapping.

We demonstrate it at two examples:

$f:\mathrm{IR}^2 \to \mathrm{IR}^2$ shear mapping (= example from above)	$g:\mathrm{IR}^2 \to \mathrm{IR}^2$ projection along the principal diagonal onto the y axis
 $f\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}$ $f\begin{bmatrix}0\\1\end{bmatrix}=\begin{bmatrix}1\\1\end{bmatrix}$ Matrix of f: $\begin{bmatrix}1 & 1\\0 & 1\end{bmatrix}$ The images $f\begin{bmatrix}1\\0\end{bmatrix}, f\begin{bmatrix}0\\1\end{bmatrix}$ (i.e., the column vectors of the matrix of f) are linearly independent, they span the whole plane IR^2	 $g\begin{bmatrix}1\\0\end{bmatrix}=\begin{bmatrix}0\\-1\end{bmatrix}$ $g\begin{bmatrix}0\\1\end{bmatrix}=\begin{bmatrix}0\\1\end{bmatrix}$ Matrix of g: $\begin{bmatrix}0 & 0\\-1 & 1\end{bmatrix}$ The images $g\begin{bmatrix}1\\0\end{bmatrix}, g\begin{bmatrix}0\\1\end{bmatrix}$ (i.e., the column vectors of the matrix of g) are linearly <u>dependent</u>, they are on the same line through 0 (i.e., on the y axis)

⇒ each vector is an image under f (f is surjective) rank f = 2 (= dimension of the plane)	⇒ only the *y* axis is the range of g (g is not surjective) rank g = 1 (= dimension of the line)

Definition:
The rank of a matrix *A* is the maximal number of linearly independent column vectors of *A*.
Notation: *rank* (*A*), r (*A*).

This is consistent with our former definition:
rank (*A*) = rank of the system of column vectors of *A* (as a vector system).

At the same time, it is the dimension of the range of the corresponding linear mapping of *A*.

Theorem:
rank(*A*) is also the maximal number of linearly independent row vectors of *A*.

"column rank = row rank" !

Special cases:
The rank of the zero matrix is 0 (= smallest possible rank of a matrix).
The rank of *E*, the $n \times n$ unit matrix, is *n* (= largest possible rank of an $n \times n$ matrix).

The rank of an $m \times n$ matrix A can be at most the number of rows and at most the number of columns:
$0 \leq rank(A) \leq \min(m, n)$.

For determining the rank of a matrix, it is useful to know that under certain *elementary operations* the *rank* of a matrix *does not change*:

Elementary row operations

(1) Reordering of rows (particularly, switching of two rows)

(2) multiplication of a complete row by a number $c \neq 0$

(3) addition or omission of a row which is a linear combination of other rows

(4) addition of a linear combination of rows to another row.

Analogous for column operations.

Example:

$$A = \begin{bmatrix} 2 & 6 & -4 \\ 3 & 11 & 1 \\ -4 & -14 & 1 \end{bmatrix}$$

By applying elementary row operations, we transform *A* into an upper triangular matrix (parentheses are omitted for convenience):

$$
\begin{array}{rrrl}
2 & 6 & -4 & :2 \\
3 & 11 & 1 & \\
-4 & -14 & 1 & \\
\hline
1 & 3 & -2 & \cdot 3 \quad \cdot 4 \\
3 & 11 & 1 & - \quad + \\
-4 & -14 & 1 & \\
\hline
1 & 3 & -2 & \\
0 & 2 & 7 & + \\
0 & -2 & -7 & \\
\hline
1 & 3 & -2 & \\
0 & 2 & 7 & \\
0 & 0 & 0 &
\end{array}
$$

The rank of *A* must be the same as the rank of the matrix obtained in the end.

The rank of this triangular matrix can easily seen to be 2 (one zero row; zero rows are always linearly dependent! – The other two rows must be independent because of the first components 1 and 0.)

9. The determinant

The *determinant* is a function (with real numbers as values) which is defined for square matrices.

It allows to make conclusions about the rank and appears in diverse theorems and formulas.

Notation:

$\begin{bmatrix} \cdots \\ \cdots \end{bmatrix}$ matrix, $\begin{vmatrix} \cdots \\ \cdots \end{vmatrix}$ determinant.

Also: *A* matrix, det(*A*) = |*A*| ∈ IR determinant of *A*.

$$det\, A = |A| = \begin{vmatrix} a_{11}a_{12}\cdots a_{1n} \\ \cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots \\ a_{n1}a_{n2}\cdots a_{nn} \end{vmatrix}$$

We call this a determinant of *order* *n*.

Calculation in the special cases $n = 2$ and $n = 3$:

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

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

The formula for the case 3×3 is called "Sarrus' rule".
Other notation for it, using auxiliary columns for better memorizing the products and their signs:

$$|A| = \left|\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array}\right|$$

$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}$$

The calculation formula for the general case requires the notion of a *subdeterminant*:

Let A be an $n \times n$ matrix. Its determinant is $|A|$.
By omitting the ith row and the jth column we obtain a subdeterminant of order $n-1$.
Notation: $|A_{ij}|$.

In the following formulas, the value of this sub-determinant is multiplied by the factor $(-1)^{i+j}$, giving a sign which alternates between rows and columns like in a chessboard:

$$\left|\begin{array}{ccccc} + & - & + & - & \cdots \\ - & + & - & + & \\ + & - & + & - & \\ - & + & - & + & \cdots \\ \vdots & & & \vdots & \end{array}\right|$$

Formulas for calculating determinants of $n \times n$ matrices A of arbitrary size (so-called *development theorems*):

(a) For a fixed jth column:

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|$$

(b) For a fixed ith row:

$$|A| = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{ij}|$$

On the right-hand side we have again determinants, but with smaller size.

We call this "to develop a determinant for a given column (or row)".

Example $n = 3$:

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$= a_{11} \cdot (-1)^{1+1} \cdot \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{12} \cdot (-1)^{1+2} \cdot \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \cdot (-1)^{1+3} \cdot \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

When we have zero entries, it is advantageous to choose the rows or columns with most zeros.

Example for $n = 4$:

$$|A| = \begin{vmatrix} 1 & 2 & 1 & 2 \\ 0 & 1 & 8 & 7 \\ 0 & 3 & 3 & 1 \\ 0 & 1 & 0 & 1 \end{vmatrix} = (-1)^{1+1} \cdot \begin{vmatrix} 1 & 8 & 7 \\ 3 & 3 & 1 \\ 1 & 0 & 1 \end{vmatrix} = 3+8+0-21-0-24 = -34$$

Rules for determinants:

(1) Switching two rows or two columns changes the sign of the determinant.
(2) If a matrix has a zero row (or a zero column), its determinant is 0.
(3) Has a matrix two identical rows (or columns), its determinant is 0.
(4) If a row (column) of a matrix is multiplied by k, the value of its determinant increases also by the factor k.
(5) If some row (column) is a linear combination of the other rows (columns), the determinant is 0.
(6) The determinant does not change its value if some linear combination of the other rows (columns) is added to a row (column).
(7) The determinant of a matrix does not change if the matrix is transposed: $|A| = |A^T|$.
(8) The determinant of a triangular matrix is the product of the elements of the principal diagonal.

Definition "regular" / "singular":
An $n \times n$ matrix A is called *regular* if $rank(A) = n$
(i.e., if it has the maximal possible rank)
– or, expressed in another way: if all its rows (columns) are linearly independent

Otherwise, A is called *singular*.

Theorem: A is regular $\Leftrightarrow$ $|A| \neq 0$.

Corollary: A is singular $\Leftrightarrow$ $|A| = 0$.

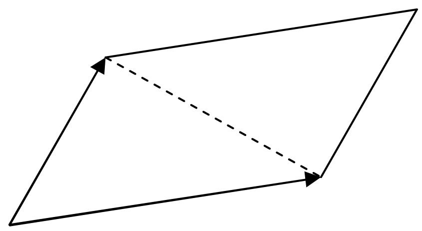
We can thus use the determinant as an indicator of linear independence (of all rows or columns of *A*).

Geometrical application of the determinant:

When the sign is disregarded, $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ is the area content of the *parallelogram* spanned by the two column vectors $\vec{a} = \begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix}$.

$$\left| det(\vec{a}, \vec{b}) \right| = \left| a_{11} a_{22} - a_{12} a_{21} \right|$$

(Remark: The area of the spanned *triangle* is exactly half of this value!)



Analogously in IR^3 :
Disregarding the sign, $det(\vec{a}, \vec{b}, \vec{c})$ is the volume of the *parallelepiped* spanned by $\vec{a}, \vec{b}, \vec{c}$.

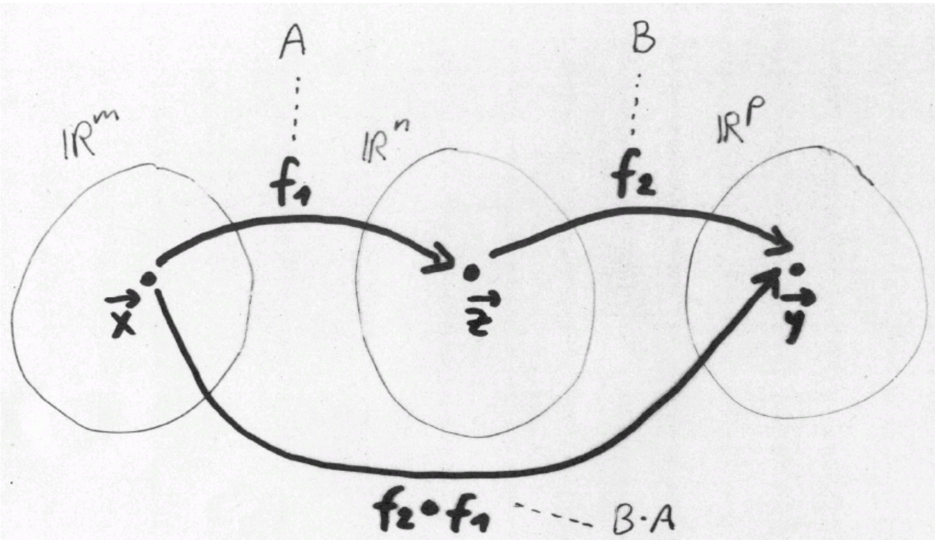
10. More on linear mappings and matrices

Linear mappings can be carried out one after the other (composition of mappings):
Let

$f_1 : \mathrm{I\!R}^m \to \mathrm{I\!R}^n$ be described by the matrix *A*,

$f_2 : \mathrm{I\!R}^n \to \mathrm{I\!R}^p$ be described by the matrix *B*.

By composing both mappings, we obtain a new mapping $f_2 \circ f_1$, the *composition* of f_1 and f_2 (notice the notation from right to left):



The new mapping $f_2 \circ f_1 : \mathrm{I\!R}^m \to \mathrm{I\!R}^p$ is again linear and has also a corresponding matrix (of type (*p*, *m*)).
Its matrix is called the *product* of the matrices *A* and *B*: $B \cdot A$

How is the product of two matrices calculated?

The case of two matrices of type (2, 2):

$$\begin{bmatrix} \boxed{b_{11} \quad b_{12}} \\ b_{21} \quad b_{22} \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \boxed{b_{11} \cdot a_{11} + b_{12} \cdot a_{21}} & b_{11} \cdot a_{12} + b_{12} \cdot a_{22} \\ b_{21} \cdot a_{11} + b_{22} \cdot a_{21} & b_{21} \cdot a_{12} + b_{22} \cdot a_{22} \end{bmatrix}$$

All possible inner products "row of the first matrix" by "column of the second matrix" are calculated and written into the result matrix.

This holds also in the general case:

Definition:
The *product* of two matrices A of type (m, n) and B of type (n, p) is a matrix $C = A \cdot B$ of type (m, p) with the elements

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

The product exists only in the case when the first matrix has as many columns as the second has rows!

Example:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 0 & 2 \\ -1 & -2 & 0 \end{bmatrix} , \quad B = \begin{bmatrix} 2 & -1 \\ 3 & 2 \\ 0 & 1 \end{bmatrix} \quad \Rightarrow$$

$$A \cdot B = \begin{bmatrix} 8 & 6 \\ 6 & -1 \\ -8 & -3 \end{bmatrix}$$

Attention: In the general case, $A \cdot B \neq B \cdot A$.

Example: $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}$, but $\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Transposition of a matrix product:

$$(A \cdot B)^{\mathrm{T}} = B^{\mathrm{T}} \cdot A^{\mathrm{T}}$$

The product of a matrix *A* with a column vector $\vec{v}$ is a special case of the product of two matrices (second factor of type (*n*, 1)).

<u>Application:</u>
Transformation of age-class vectors

We remember:
The age-class structure of a forest at time *t* can be described by an age-class vector.

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

a_1 ← area with trees of age class 1

a_2 ← area with trees of age class 2

a_n ← area with trees of age class *n*

$= \vec{a}_t \in \mathbb{R}^n$ (*n* = number of successive age classes)

The development of the age structure over time can be described by a linear mapping $\mathbb{R}^n \to \mathbb{R}^n$.

Let $p_{j,k}$ be the part of the area of the jth age class which comes into the kth age class.

Example: $p_{3,4} = 0.7$
$p_{3,1} = 0.3$, i.e., 30% of the stand of age class 3 are cut and the free area is immediately reforested with young trees (age class 1)
$p_{3,k} = 0$ for all other k.

Age-class transition matrix:

$$P := \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ \vdots & & & \vdots \\ p_{n,1} & p_{n,2} & \cdots & p_{n,n} \end{bmatrix}$$

In the calculations, more often the transposed matrix P^T is used. In population ecology, it is called the *Leslie matrix*.

Theorem:
The age class vector of the stand at time t+1 can be calculated as

$$\vec{a_{t+1}} = P^T \cdot \vec{a}_t$$

In the simple case n=3, this gives

$$\begin{bmatrix} p_{1,1} & p_{2,1} & p_{3,1} \\ p_{1,2} & p_{2,2} & p_{3,2} \\ p_{1,3} & p_{2,3} & p_{3,3} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} p_{1,1} a_1 + p_{2,1} a_2 + p_{3,1} a_3 \\ p_{1,2} a_1 + p_{2,2} a_2 + p_{3,2} a_3 \\ p_{1,3} a_1 + p_{2,3} a_2 + p_{3,3} a_3 \end{bmatrix}$$

Because some of the entries of P^T are necessarily 0 (organisms cannot stop ageing; they cannot overjump some age class or become younger), this can be simplified to:

$$\begin{bmatrix} p_{1,1} & p_{2,1} & p_{3,1} \\ p_{1,2} & 0 & 0 \\ 0 & p_{2,3} & 0 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} p_{1,1}a_1 + p_{2,1}a_2 + p_{3,1}a_3 \\ p_{1,2}a_1 \\ p_{2,3}a_2 \end{bmatrix}$$

We can say that *P* describes a forest management strategy.

Usage of the matrix product in this context:

If between times *t* and *t*+1, strategy *P* is applied, and between times *t*+1 and *t*+2 strategy *Q*, then we have in total:

$$\vec{a_{t+2}} = Q^T \cdot \vec{a_{t+1}} = Q^T \cdot (P^T \cdot \vec{a}_t) = (Q^T \cdot P^T) \cdot \vec{a}_t = (P \cdot Q)^T \cdot \vec{a}_t$$

If the strategy is *the same* in every time step, we have:

$$\begin{aligned} \vec{a}_t &= P^T \cdot \vec{a_{t-1}} \\ &= P^T \cdot P^T \cdot \vec{a_{t-2}} \\ &= P^T \cdot P^T \cdot P^T \cdot \vec{a_{t-3}} \\ &= \ldots \\ &= (P^T)^t \cdot \vec{a}_0 \end{aligned}$$

(Here, $()^T$ means transposition, $()^t$ means the *t*-fold product of a matrix with itself.)

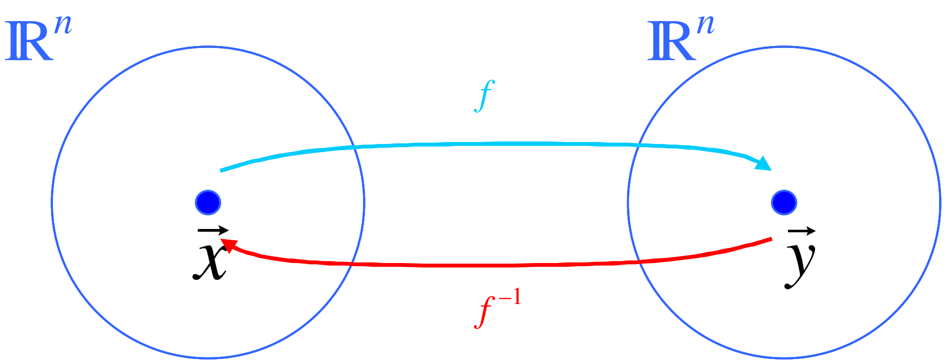
The inverse matrix

Let A be an $n \times n$ matrix.
A^{-1} is called the inverse matrix of A if
$A^{-1} \cdot A = A \cdot A^{-1} = E$ (= the unit matrix).

Not every matrix has an inverse.
If the inverse matrix of A exists, it is unique.
It is always $(A^{-1})^{-1} = A$.

When does A^{-1} exist ?

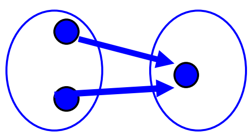
A is a matrix of type (n, n)
corresponding linear mapping $f : \mathbb{R}^n \to \mathbb{R}^n$



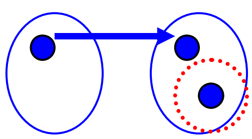
$\vec{y} = f(\vec{x}) = A \cdot \vec{x}$

when does the inverse mapping f^{-1} exist?

If f is bijective, i.e.,
f injective …… not:



and
f surjective ……not:



$$\Leftrightarrow f(\vec{e}_1), \ldots, f(\vec{e}_n) \text{ span } \mathbb{R}^n \text{ completely}$$

$$\vec{y} = m_1 f(\vec{e}_1) + \ldots + m_n f(\vec{e}_n)$$

$$= f(m_1\vec{e}_1 + \ldots + m_n\vec{e}_n) = f\begin{pmatrix} m_1 \\ \vdots \\ m_n \end{pmatrix}$$

$$\Leftrightarrow \; rank(A) = n \;\Leftrightarrow\; \det A \neq 0$$

$$\Leftrightarrow \; \underline{\underline{A \;\; regular}}$$

Exactly the regular $n \times n$ matrices have an inverse matrix.

How to calculate it?
Most efficient way: by elementary row operations.
Concatenate an $n \times n$ unit matrix E to A:

$$M = [\, A \mid E \,] = \left[\begin{array}{cccc|cccc} a_{11} & a_{12} & \ldots & a_{1n} & 1 & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ldots & a_{2n} & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} & 0 & \ldots & \ldots & 1 \end{array}\right]$$

Then transform this larger matrix by elementary row operations in a way that the left part is transformed into the unit matrix. The resulting right part is then A^{-1}: $[\, E \mid A^{-1}]$.

Example:

$$A=\begin{bmatrix}1 & 0 & 0\\ 3 & 1 & 0\\ 0 & 3 & 1\end{bmatrix}$$

The start scheme is

$$\left[\begin{array}{ccc:ccc}1 & 0 & 0 & 1 & 0 & 0\\ 3 & 1 & 0 & 0 & 1 & 0\\ 0 & 3 & 1 & 0 & 0 & 1\end{array}\right]$$

By subtracting the first row 3 times from the second row, and then the second row 3 times from the third, we get

$$\left[\begin{array}{ccc:ccc}1 & 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & -3 & 1 & 0\\ 0 & 0 & 1 & 9 & -3 & 1\end{array}\right] \Rightarrow \quad A^{-1} = \begin{bmatrix}1 & 0 & 0\\ -3 & 1 & 0\\ 9 & -3 & 1\end{bmatrix}$$

It is recommended to check if really $A{\cdot}A^{-1} = E$ (otherwise some error must have occurred).

Systems of linear equations

A system of *m* linear equations with *n* unknowns can always be ordered and rewritten in the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \vdots \qquad\qquad\qquad & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

If we collect the unknowns x_k in a column vector

$\vec{x} \in \mathbb{R}^n$

and the numbers b_i on the right-hand side (called absolute terms) also in a column vector

$\vec{b} \in \mathbb{R}^m$

and the coefficients a_{ik} in a matrix A of type (m, n), we can write the whole system *as a single equation*:

$$A \cdot \vec{x} = \vec{b}$$

$A = \begin{bmatrix} a_{11} & \ldots & a_{1n} \\ a_{21} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{bmatrix}$ is called the *coefficient matrix*

of the system,

A_{ext} = $\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right]$ the *extended matrix* of the system.

Notice:
$A \cdot \vec{x} = \vec{b}$ can also be interpreted as $f(\vec{x}) = \vec{b}$, with f the linear mapping described by the matrix *A*.

Finding a solution of the linear system means thus to find a vector which is mapped to $\vec{b}$.

For each system of linear equations, there are three possibilities:
(1) The system has exactly one solution $\vec{x}$,
(2) the system has infinitely many solutions,
(3) the system has no solutions at all (it is then called *inconsistent*).

Examples:
(1) The system $\left\{\begin{matrix} x_1 + x_2 = 5 \\ 2x_1 + x_2 = 7 \end{matrix}\right\}$ has exactly one solution: $\vec{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, that means, $x_1 = 2$ and $x_2 = 3$.

Indeed, $2 + 3 = 5$ and $2 \cdot 2 + 3 = 7$, and there are no other combinations of numbers which fulfill both equations simultaneously.

(2) The system $\begin{Bmatrix} x_1 + x_2 = 5 \\ 2x_1 + 2x_2 = 10 \end{Bmatrix}$ has infinitely many solutions, which all have the form $\vec{x} = \begin{bmatrix} 5-a \\ a \end{bmatrix}$ ($a \in$ IR), that means, $x_1 = 5 - a$ and $x_2 = a$.

(3) The system $\begin{Bmatrix} x_1 + x_2 = 5 \\ 2x_1 + 2x_2 = 7 \end{Bmatrix}$ has no solution.
Both equations contradict each other.

Frobenius' Theorem:

The system of m linear equations with n unknowns which is described by the vector equation $A \cdot \vec{x} = \vec{b}$ has solutions if and only if $rank(A) = rank(A_{ext})$.
More precisely:

(1) If $rank(A) = rank(A_{ext}) = n$, the system has exactly one solution.

(2) If $rank(A) = rank(A_{ext}) < n$, the system has infinitely many solutions. In this case, the values of $n - rank(A)$ of the unknowns can be chosen arbitrarily.

(3) If $rank(A) \neq rank(A_{ext})$, the system has no solutions at all.

We can check the theorem at the examples from above:

(1) $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}$, $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\vec{b} = \begin{bmatrix} 5 \\ 7 \end{bmatrix}$

so, $A_{ext} = \begin{bmatrix} 1 & 1 & 5 \\ 2 & 1 & 7 \end{bmatrix}$, and we have
rank(*A*) = *rank*(A_{ext}) = 2 = *n*, there is exactly one solution.

(2) $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$, $A_{ext} = \begin{bmatrix} 1 & 1 & 5 \\ 2 & 2 & 10 \end{bmatrix}$
Here, *rank*(*A*) = *rank*(A_{ext}) = 1 < 2 = *n* (the second row is a multiple of the first one), and we have infinitely many solutions (2–1 = 1 unknown can be put to an arbitrary value).

(3) $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$, $A_{ext} = \begin{bmatrix} 1 & 1 & 5 \\ 2 & 2 & 7 \end{bmatrix}$
Here, *rank*(*A*) = 1 < *rank*(A_{ext}) = 2. There is no solution.

How to solve systems of linear equations?

"*Gaussian method of elimination*":
most effective method in the general case.

By *elementary row operations*, the extended matrix of the system is transformed into an upper triangular matrix. The solutions of the corresponding system of equations remain the same!
The system corresponding to the upper triangular matrix can easily be solved "bottom-up" by successive insertion and elimination of unknowns.

Example: Solve the system

$$
\begin{aligned}
x_1-2x_2+x_3 &= 0\\
3x_1-5x_2-2x_3 &= -3\\
7x_1-3x_2+x_3 &= 16
\end{aligned}
$$

Its extended matrix is $A_{ext} = \left[\begin{array}{ccc|c} 1 & -2 & 1 & 0\\ 3 & -5 & -2 & -3\\ 7 & -3 & 1 & 16 \end{array}\right]$

By applying elementary row operations, one gets

$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 0\\ 0 & 1 & -5 & -3\\ 0 & 0 & 49 & 49 \end{array}\right]$ (upper triangular matrix).

From this, we can immediately see that
$rank(A) = rank(A_{ext}) = 3 = n$,
and following Frobenius we can conclude that the system has exactly one solution.

The system of equations corresponding to the transformed matrix is

$$
\begin{aligned}
x_1-2x_2+x_3 &= 0\\
x_2-5x_3 &= -3\\
49x_3 &= 49
\end{aligned}
$$

and this can be solved easily from the third row up to the second and first row by elimination of variables. We obtain:

$x_3=1,\quad x_2=2,\quad x_1=3$

and thus the unique vector solution $\vec{x}=\begin{bmatrix}3\\2\\1\end{bmatrix}$.

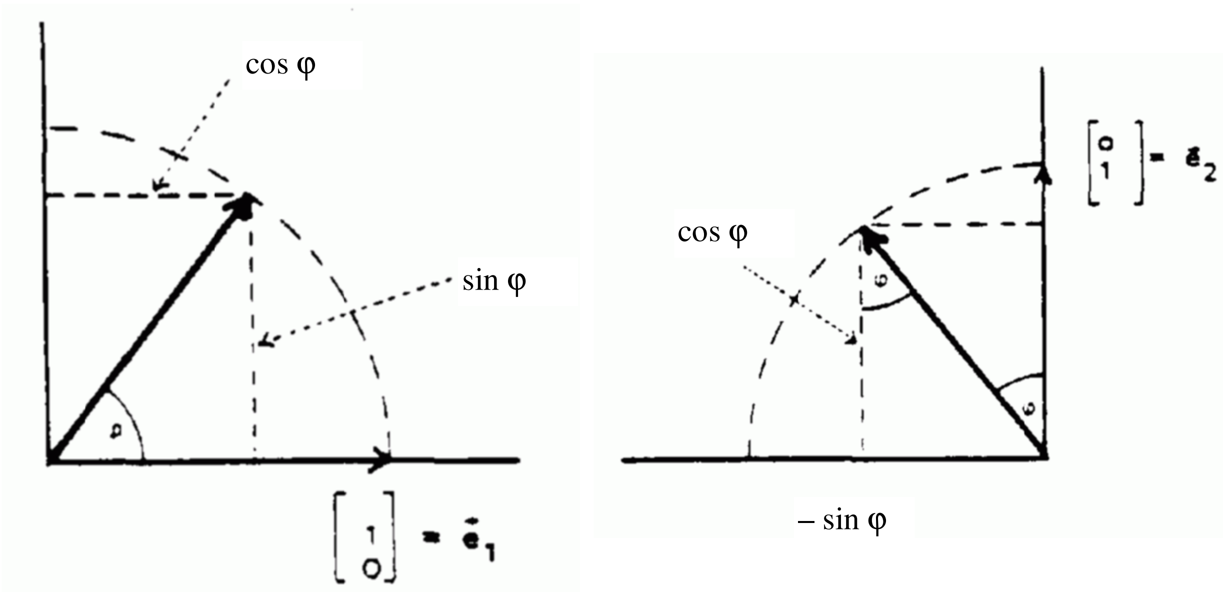
Special cases of linear mappings

(a) Rotations around the origin

Let f be the counterclockwise rotation by the angle φ around the zero point (0; 0) (origin of the cartesian coordinate system).

Each vector is rotated by φ, its image has the same length as before.

Image vectors of the standard basis vectors = ?



We obtain as image vectors: $\begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix}$ and $\begin{pmatrix} -\sin\varphi \\ \cos\varphi \end{pmatrix}$.

The matrix of f is thus:

$$A = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

We call this a *rotation matrix*.

(b) Scaling

Let f be the mapping which enlarges (or shrinks) every vector by a certain, fixed factor $\lambda \neq 0$.
Its corresponding matrix is

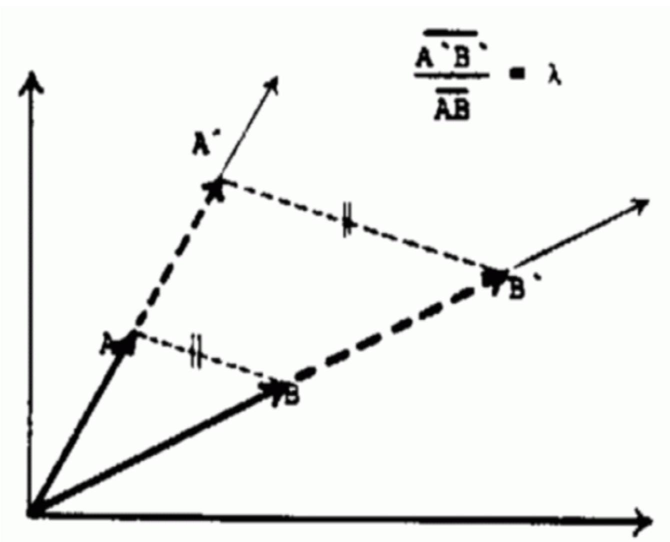
$$A = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix},$$

called a *scaling matrix* with factor λ.

Indeed, we have

$$A\vec{x} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \end{bmatrix}$$

The image of each vector has the same direction as before (or the opposite direction, if λ is negative), but a length which is modified by the factor $|\lambda|$.
Parallelism and all angles remain unchanged under this mapping.



(c) Centroaffine mapping

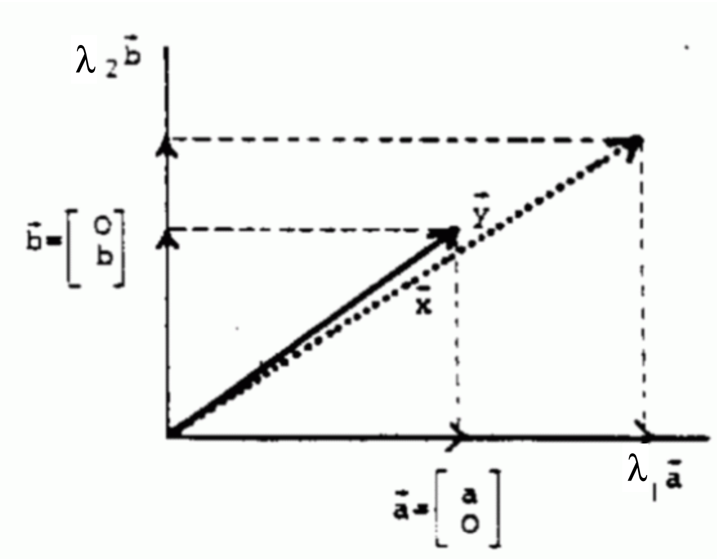
Let f act as a scaling by λ_1 on the *x* axis and as a scaling by λ_2 on the *y* axis, with two different numbers $\lambda_1 \neq \lambda_2$.
The corresponding matrix is

$$A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

This mapping is called *centroaffine*.

Its effect: $\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \lambda_1 a \\ \lambda_2 b \end{bmatrix}$

It works as a pure scaling on the *x* and *y* axis, but not for vectors which are outside these coordinate axes (they are also rotated a bit):



The centroaffine mapping is thus not a scaling for all vectors.
Certain vectors play a special role for this mapping, namely, those on the coordinate axes: They are only scaled, the others are also rotated.