# Lecture 2: Linear operators and their geometric representation

Francesco Preta

July 2020

## Definition and examples of linear functions

Linear functions (or operators) are one of the fundamental concepts of linear algebra. The definition is the following:

**Definition 1.** Let $V, W$ be vector spaces over $\mathbb{R}$. A function $T : V \to W$ is called *linear* if for every $\mathbf{x}, \mathbf{y} \in V$ and $\alpha, \beta \in \mathbb{R}$, we have

$$T(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha T(\mathbf{x}) + \beta T(\mathbf{y})$$

Examples of linear functions are:

- **Matrix multiplication**: let $A$ be a $n \times m$ matrix. Then for $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^m$ and $\alpha, \beta \in \mathbb{R}$, we have

  $$A(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha A\mathbf{x} + \beta A\mathbf{y}$$

  For instance, take $A = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$, then for every coefficient $\alpha, \beta \in \mathbb{R}$ and vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ we have

  $$A(\alpha\mathbf{x} + \beta\mathbf{y}) = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \end{bmatrix} = \begin{bmatrix} \alpha(x_1 + 2x_2) + \beta(y_1 + 2y_2) \\ \alpha(x_2 - x_1) + \beta(y_2 - y_1) \end{bmatrix}$$

  while

  $$\alpha A\mathbf{x} + \beta A\mathbf{y} = \alpha \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \beta \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \alpha(x_1 + 2x_2) + \beta(y_1 + 2y_2) \\ \alpha(x_2 - x_1) + \beta(y_2 - y_1) \end{bmatrix}$$

- **Inner product with a fixed vector**: let $\mathbf{z} \in \mathbb{R}^n$, then for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, we have

  $$\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle$$

For instance, let $\mathbf{z} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$. Then for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ and $\alpha, \beta \in \mathbb{R}$

$$\langle \alpha \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \beta \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \rangle = \langle \begin{bmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \\ \alpha x_3 + \beta y_3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \rangle = \alpha(x_1 - x_3) + \beta(y_1 - y_3)$$

On the other hand,

$$\alpha \langle \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \rangle + \beta \langle \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \rangle = \alpha(x_1 - x_3) + \beta(y_1 - y_3)$$

- **Integrals on function spaces**: let $f_1, f_2 : [0,1] \to \mathbb{R}$ be Riemann-integrable functions and $\alpha, \beta \in \mathbb{R}$. Then

$$\int_0^1 (\alpha f_1(t) + \beta f_2(t))\, dt = \alpha \int_0^1 f_1(t)\, dt + \beta \int_0^1 f_2(t)\, dt$$

- **Derivatives on function spaces**: let $g_1, g_2 : [0,1] \to \mathbb{R}$ be derivable functions and $\alpha, \beta \in \mathbb{R}$. Then

$$\frac{d}{dt}(\alpha g_1 + \beta g_2)(t) = \alpha \frac{dg_1(t)}{dt} + \beta \frac{dg_2(t)}{dt}$$

Examples of non-linear functions:

- **Euclidean norm of a vector**: for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ in general

$$||\alpha \mathbf{x} + \beta \mathbf{y}|| \neq \alpha ||\mathbf{x}|| + \beta ||\mathbf{y}||.$$

For instance, consider $\mathbf{x} = \begin{bmatrix} 1 & 0 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 0 & 1 \end{bmatrix}$ and $\alpha = \beta = 1$. Then

$$||\alpha \mathbf{x} + \beta \mathbf{y}|| = || \begin{bmatrix} 1 & 1 \end{bmatrix} || = \sqrt{2}$$

but

$$\alpha ||\mathbf{x}|| + \beta ||\mathbf{y}|| = || \begin{bmatrix} 1 & 0 \end{bmatrix} || + || \begin{bmatrix} 0 & 1 \end{bmatrix} || = 2$$

Analogously, if $\alpha = -1, \beta = 0$, we have

$$||\alpha \mathbf{x} + \beta \mathbf{y}|| = || \begin{bmatrix} -1 & 0 \end{bmatrix} || = 1$$

while

$$\alpha ||\mathbf{x}|| + \beta ||\mathbf{y}|| = -|| \begin{bmatrix} 1 & 0 \end{bmatrix} || = -1$$

- **Maximum function**: Consider $\max : \mathbb{R}^n \to \mathbb{R}$, the function returning the maximum element in a vector. Then for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, in general

$$\max(\alpha \mathbf{x} + \beta \mathbf{y}) \neq \alpha \max(\mathbf{x}) + \beta \max(\mathbf{y})$$

As a counterexample, consider $\mathbf{x} = \begin{bmatrix} 0 & 1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 0 & -1 \end{bmatrix}$, $\alpha = \beta = 1$. Then

$$\max(\alpha \mathbf{x} + \beta \mathbf{y}) = \max \begin{bmatrix} 0 & 0 \end{bmatrix} = 0$$

while

$$\alpha \max(\mathbf{x}) + \beta \max(\mathbf{y}) = \max \begin{bmatrix} 0 & 1 \end{bmatrix} + \max \begin{bmatrix} 0 & -1 \end{bmatrix} = 1 + 0 = 1$$

.

## Matrix representation of linear functions

In Definition 1, $T$ is an operator between two abstract vector spaces. In the specific example of matrix multiplication, a $n \times m$ matrix $A$ defines a function from $\mathbb{R}^m$ to $\mathbb{R}^n$ according to the rule of matrix multiplication. To this extent, inner product can be seen as a particular case of matrix multiplication for $A$ a $1 \times n$ matrix.

A priori, it seems to be the case that matrix multiplication constitutes only one of the possible linear functions between $\mathbb{R}^m$ and $\mathbb{R}^n$. However we will proving that every linear function on $\mathbb{R}^m$ is in some sense a matrix multiplication:

**Theorem 1** (Representation Theorem)**.** *Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator. Then there exists a $n \times m$ matrix $A$ such that $f(\mathbf{x}) = A\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}^m$*

*Proof.* Linearity of $f$ guarantees that the value of $f(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^m$ is uniquely determined by the value of $f$ on a finite set of vectors that span the entire space. In the case of $\mathbb{R}^m$, we introduced in the last lecture the canonical basis $\{\mathbf{e}_i\}_{i=1}^m$, where we recall the definition as

$$(\mathbf{e}_i)_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Then every vector $\mathbf{x} \in \mathbb{R}^m$ is a linear combination of $\mathbf{e}_i$'s, that is $\mathbf{x} = \sum_{i=1}^m x_i \mathbf{e}_i$. Then we have

$$f(\mathbf{x}) = f(\sum_{i=1}^m x_i \mathbf{e}_i) = \sum_{i=1}^m x_i f(\mathbf{e}_i)$$

so that if we know the values of $\{f(\mathbf{e}_i)\}_{i=1}^m$, $f(\mathbf{x})$ can be obtained as a linear combination of those values with $i$-th coefficient $x_i$.

In order to find the $n \times m$ matrix $A$ such that $f(\mathbf{x}) = A\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}^m$, on each $\mathbf{e}_i$ for $i = 1, ..., m$, $A$ would satisfy $A\mathbf{e}_i = f(\mathbf{e}_i)$. But an easy calculation shows that $A\mathbf{e}_i$ is the i-th column of $A$:

$$(A\mathbf{e}_i)_j = \sum_{k=1}^n A_{j,k} (\mathbf{e}_i)_k = A_{j,i}$$

That means that for $A$ given by

$$A = \begin{bmatrix} f(\mathbf{e}_1) & f(\mathbf{e}_2) & ... & f(\mathbf{e}_m) \end{bmatrix}$$

where every $f(\mathbf{e}_i)$ is a column vector of $\mathbb{R}^n$, we have the desired representation matrix. In fact, we have

$$A\mathbf{x} = A(\sum_{i=1}^{m} x_i\mathbf{e}_i) = \sum_{i=1}^{m} x_i A\mathbf{e}_i = \sum_{i=1}^{m} x_i f(\mathbf{e}_i) = f(\sum_{i=1}^{m} x_i\mathbf{e}_i) = f(\mathbf{x})$$

□

An interesting consequence of this construction regards composition of linear functions: let $f_1 : \mathbb{R}^m \to \mathbb{R}^n$ and $f_2 : \mathbb{R}^n \to \mathbb{R}^l$ be two linear functions, then $f_2 \circ f_1 : \mathbb{R}_m \to \mathbb{R}^l$ is also a linear function. This can be proved using the definition since for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ and $\alpha, \beta \in \mathbb{R}$

$$f_2(f_1(\alpha\mathbf{x} + \beta\mathbf{y})) = f_2(\alpha f_1(\mathbf{x}) + \beta f_1(\mathbf{y})) = \alpha f_2(f_1(\mathbf{x})) + \beta f_2(f_1(\mathbf{y}))$$

But then if $A$ is the matrix given by the representation theorem for $f_2 \circ f_1$, $A_1$ is the one for $f_1$ and $A_2$ is the one for $f_2$, $A = A_2A_1$ with the usual matrix product (notice that the dimensions match).

Example: consider $f : \mathbb{R}^3 \to \mathbb{R}^3$ to be the shift-forward operator, that is

$$f(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}) = \begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix}$$

We first need to prove that $f$ is linear and then we need to find the matrix $A \in \mathbb{R}^{3\times3}$ such that $f(\mathbf{x}) = A\mathbf{x}$. In order to prove linearity:

$$f(\alpha\mathbf{x} + \beta\mathbf{y}) = f(\begin{bmatrix} \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \\ \alpha x_3 + \beta y_3 \end{bmatrix}) = \begin{bmatrix} \alpha x_3 + \beta y_3 \\ \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \end{bmatrix}$$

On the other hand

$$\alpha f(\mathbf{x}) + \beta f(\mathbf{y}) = \alpha \begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix} + \beta \begin{bmatrix} y_3 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \alpha x_3 + \beta y_3 \\ \alpha x_1 + \beta y_1 \\ \alpha x_2 + \beta y_2 \end{bmatrix}$$

so that the two quantities coincide for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ and $\alpha, \beta \in \mathbb{R}$. Now, in order to find the operator $A$, consider

$$f(\mathbf{e}_1) = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \mathbf{e}_2$$

$$f(\mathbf{e}_2) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \mathbf{e}_3$$

$$f(\mathbf{e}_3) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{e}_1$$

In this way we have

$$A = \begin{bmatrix} f(\mathbf{e}_1) & f(\mathbf{e}_2) & f(\mathbf{e}_3) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

We can double check that for every $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ we have

$$A\mathbf{x} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_3 \\ x_1 \\ x_2 \end{bmatrix} = f(\mathbf{x})$$

## Injectivity and surjectivity

We recall the following definitions for generic (non-necessarily linear) functions:

**Definition 2.** A function $f : V \to W$ is said to be *injective* if

$$f(\mathbf{x}) = f(\mathbf{y}) \text{ if and only if } \mathbf{x} = \mathbf{y}$$

that is, different points of $V$ have different image in $W$. $f$ is said to be *surjective* if

$$\forall \mathbf{w} \in W, \exists \mathbf{x} \in V \text{ such that } f(\mathbf{x}) = \mathbf{w}$$

that is, all points in $W$ are the image of a point in $V$.

If a function is both injective and surjective it is said to be *bijective.*

In the case of linear functions, the following holds:

**Theorem 2.** *Let $T : V \to W$ be a linear function. Then injectivity of $T$ is equivalent to the condition that $T(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$.*

*Proof.* Consider $\mathbf{x}, \mathbf{y} \in V$ such that $T(\mathbf{x}) = T(\mathbf{y})$. Since $T$ is linear and $T(\mathbf{x}) - T(\mathbf{y}) = \mathbf{0}$, then $T(\mathbf{x} - \mathbf{y}) = 0$. If $T$ is injective, then $\mathbf{x} = \mathbf{y}$, therefore $\mathbf{x} - \mathbf{y} = 0$. On the other hand, if $T(\mathbf{v}) = 0$ if and only if $\mathbf{v} = 0$, then $\mathbf{x} - \mathbf{y} = \mathbf{0}$, so $\mathbf{x} = \mathbf{y}$ and $T$ is injective. □

On the other hand, for a linear operator it is always the case that $T(\mathbf{0}) = \mathbf{0}$. This comes from the definition of linearity, since

$$T(\mathbf{0}) = T(2 \cdot \mathbf{0}) = 2T(\mathbf{0})$$

which is true only if $T(\mathbf{0}) = \mathbf{0}$. Theorem 2 tells us that injectivity of $T$ corresponds to the fact that $\mathbf{0}$ is the only element whose image is $\mathbf{0}$.

In general, we have the following denominations:

**Definition 3.** Let $T : V \to W$ be a linear function. Then the *Kernel* of $T$ is defined as the elements of $V$ whose image is $\mathbf{0}$, that is

$$\mathrm{Ker}(T) = \{\mathbf{x} \in V \mid T(\mathbf{x}) = \mathbf{0}\}$$

and the *image* or *range* of $T$ are the elements of $W$ that are mapped through $T$ by some element in $V$

$$\mathrm{Im}(T) = \{y \in W \,| T(\mathbf{x}) = \mathbf{y} \text{ for some } \mathbf{x} \in V\}$$

Theorem 2 can be restated by saying that $T$ is injective if and only if $\mathrm{Ker}(T) = \{\mathbf{0}\}$.

Linearity can also be used to obtain some result on surjectivity. In fact, notice if by the representation theorem $T(\mathbf{x}) = A\mathbf{x}$ for some matrix $A$ in $\mathbb{R}^{n\times m}$, then every $\mathbf{x}$ can be written as a linear combination of the canonical basis of $\mathbb{R}^m$, and by linearity

$$A\mathbf{x} = A(\sum_{i=1}^{m} x_i\mathbf{e}_i) = \sum_{i=1}^{m} x_i A\mathbf{e}_i$$

However, since $A\mathbf{e}_i$ corresponds to the $i-th$ column of $A$, for every choice of $\mathbf{x}$, $A\mathbf{x}$ will be a linear combination (with coefficients depending on $\mathbf{x}$) of column vectors of $A$. This implies that $\mathrm{Im}(T) = \mathrm{span}(\mathbf{a}_1, ..., \mathbf{a}_m)$, where $\mathbf{a}_i = A\mathbf{e}_i$ is the i-th column vector of $A$. We have proved the following:

**Theorem 3.** *Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear operator and $A$ a $n \times m$ matrix such that $T(\mathbf{x}) = A\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}^m$. Then $T$ is surjective if and only if the columns of $A$ span $\mathbb{R}^n$.*

Example: Consider the following function $f : \mathbb{R}^2 \to \mathbb{R}^3$:

$$f(\mathbf{x}) = (x_1, x_1, x_2)$$

In order to find the corresponding matrix, consider the action on the canonical basis of $\mathbb{R}^2$, that is

$$f(1,0) = (1,1,0), \qquad f(0,1) = (0,0,1)$$

then the matrix is given by

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

In order to prove that $f$ is injective, consider the equation $f(\mathbf{x}) = 0$. Then

$$(x_1, x_1, x_2) = (0,0,0) \implies x_1 = 0, x_2 = 0$$

therefore $f$ is injective. As for surjectivity, consider $\mathrm{span}(A\mathbf{e}_1, A\mathbf{e}_2)$, that is:

$$\mathrm{span}(A\mathbf{e}_1, A\mathbf{e}_2) = \{\alpha \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \mid \alpha, \beta \in \mathbb{R}\} = \{ \begin{bmatrix} \alpha \\ \alpha \\ \beta \end{bmatrix} \mid \alpha, \beta \in \mathbb{R}\}$$

Clearly this is not the whole $\mathbb{R}^3$ because, for instance, it doesn't allow for vectors with first coordinate different from the second coordinate.

In general, to prove injectivity and surjectivity of a linear operator we need a few more notions on linearity and rank of matrices. However, the method to prove injectivity is always the same: prove that the solution to $f(\mathbf{x}) = \mathbf{0}$ is given by the sole vector $\mathbf{0} \in \mathbb{R}^n$, while proving surjectivity involves proving that the span of the column vectors of $A$ is the whole space $\mathbb{R}^n$.

## Linear and affine transformations of $\mathbb{R}^2$

We will now go through some examples of transformations in $\mathbb{R}^2$ and $\mathbb{R}^3$. In these cases, vectors will represent coordinates in the plane and the transformation will bring a point in $\mathbb{R}^2$ (or $\mathbb{R}^3$) to another point in $\mathbb{R}^2$ (or $\mathbb{R}^3$).

### Rotations in $\mathbb{R}^2$

A matrix of the form

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{1}$$

represents a rotation around the origin of angle $\theta$. Consider a point $\mathbf{x} = (x_1, x_2)$. Then

$$A\mathbf{x} = (x_1 \cos\theta - x_2 \sin\theta, x_1 \sin\theta + x_2 \cos\theta)$$

First notice that $A\mathbf{x} = 0$ if and only if $\mathbf{x} = (0, 0)$, so that $A$ is injective. An easy calculation shows that $||A\mathbf{x}|| = ||\mathbf{x}||$. That is, $A$ is a norm-preserving matrix (also called *unitary*). In addition, we can calculate the angle between $\mathbf{x}$ and its image as

$$\cos\angle(\mathbf{x}, A\mathbf{x}) = \frac{\langle \mathbf{x}, A\mathbf{x} \rangle}{||\mathbf{x}|| ||A\mathbf{x}||}$$

and we have

$$\langle \mathbf{x}, A\mathbf{x} \rangle = x_1^2 \cos\theta - x_1 x_2 \sin\theta + x_1 x_2 \sin\theta + x_2^2 \cos\theta = ||\mathbf{x}||^2 \cos\theta$$

Therefore $\cos\angle(\mathbf{x}, A\mathbf{x}) = \cos\theta$ which is the algebraic correspondent to the geometric property we wanted.

Alternatively, notice that if $f$ is a linear function on $\mathbb{R}^2$ inducing a counterclockwise rotation around the origin by an angle $\theta$, then

$$f(\mathbf{e}_1) = f(1, 0) = (\cos(\theta), \sin(\theta))$$

and

$$f(\mathbf{e}_2) = f(0, 1) = (\cos(\frac{\pi}{2} + \theta), \sin(\frac{\pi}{2} + \theta)) = (-\sin(\theta), \cos(\theta))$$

so that the representation theorem tells us that $f(\mathbf{x}) = A\mathbf{x}$ for $A$ as in equation 1.

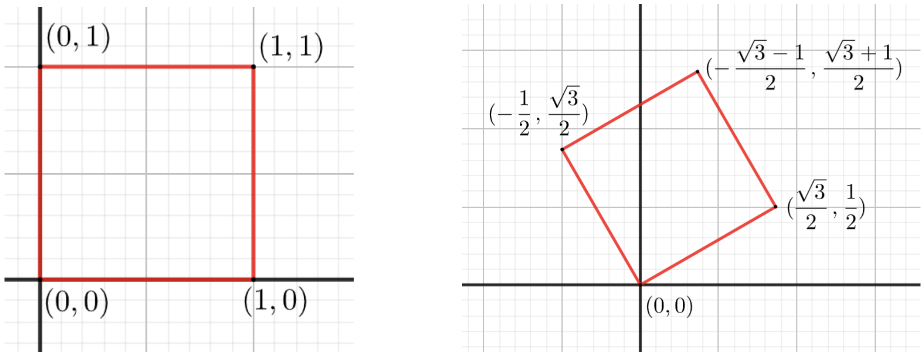



Figure 1: Rotation of $\frac{\pi}{6}$ on the unit square $[0,1] \times [0,1]$.

## Dilatations in $\mathbb{R}^2$

Any diagonal matrix of the form

$$A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$$

for $a, b > 0$ is a dilatation in $\mathbb{R}^2$. $A$ sends a circle of radius one $x^2 + y^2 = 1$ to an ellipsis $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, dilating the horizontal axis by $a$ and the vertical axis by $b$. Each axis is dilated if the corresponding coefficient is greater than 1 and it's contracted if the corresponding coefficient is less than 1.

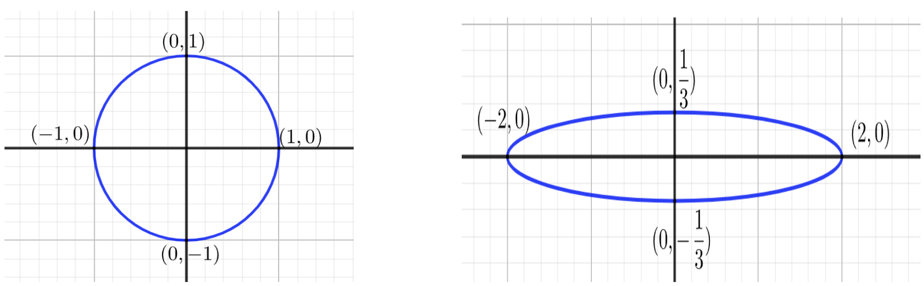



Figure 2: Effect of dilatation by $A = \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$ on the unit circle

## Reflections in $\mathbb{R}^2$

We list here different types of reflections across an axis or a point in $\mathbb{R}^2$. In general, in order to find the matrix form of the reflection, one can either find the image of the canonical basis, or understand where it sends a generic point $(x_1, x_2) \in \mathbb{R}^2$.

- **Reflection across the $x_1$-axis**. In this case, $(1,0)$ is mapped to itself, while $(0,1)$ is mapped to $(0,-1)$. The matrix $A$ has therefore the form

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

Alternatively, consider the function form $f(x_1, x_2) = (x_1, -x_2)$. Then $A$ is the matrix such that $A\mathbf{x} = f(\mathbf{x})$.

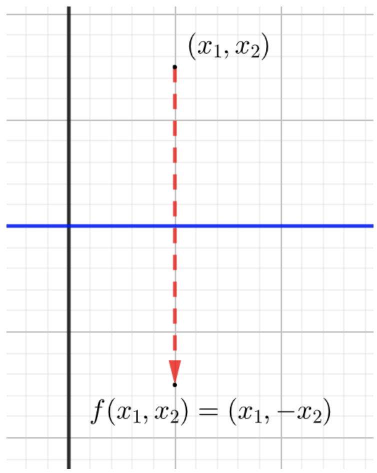



Figure 3: Reflection across the $x_1$ axis.

- **Reflection across the $x_2$-axis**. In this case, $(1,0)$ is mapped to $(-1,0)$, while $(0,1)$ is mapped to itself. The matrix $B$ has therefore the form

$$B = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Alternatively, consider the function form $g(x_1, x_2) = (-x_1, x_2)$. Then $B$ is the matrix such that $B\mathbf{x} = g(\mathbf{x})$.

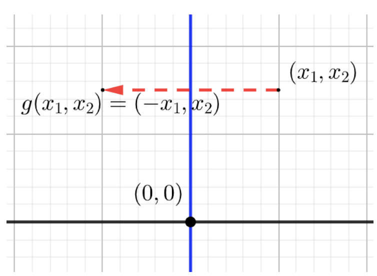



Figure 4: Reflection across the $x_2$ axis.

- **Reflection across the origin**. In this case, $(1, 0)$ is mapped to $(-1, 0)$, while $(0, 1)$ is mapped to $(0, -1)$. The matrix $C$ has therefore the form

$$C = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Alternatively, consider the function form $h(x_1, x_2) = (-x, -y)$. Then $C$ is the matrix such that $C\mathbf{x} = h\mathbf{x}$).

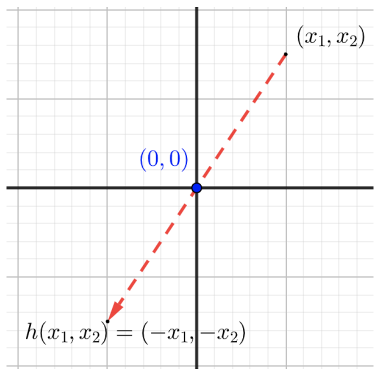



Figure 5: Reflection across the origin.

- **Reflection across the bisector** $x_2 = x_1$. In this case, $(1, 0)$ is mapped to $(0, 1)$, while $(0, 1)$ is mapped to $(1, 0)$. The matrix $D$ has therefore the form

$$D = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Alternatively, consider the function form $i(x_1, x_2) = (x_2, x_1)$. Then $D$ is the matrix such that $D\mathbf{x} = i(\mathbf{x})$.

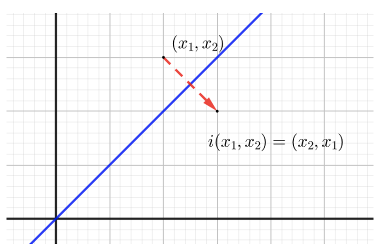



Figure 6: Reflection across the bisector of the first and third quadrant, $x_2 = x_1$.

- **Reflection across the bisector** $y = -x$. In this case, $(1, 0)$ is mapped to $(0, -1)$, while $(0, 1)$ is mapped to $(-1, 0)$. The matrix $E$ has therefore the form

$$E = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

Alternatively, consider the function form $l(x_1, x_2) = (-x_2, -x_1)$. Then $E$ is the matrix such that $E\mathbf{x} = l(\mathbf{x})$.

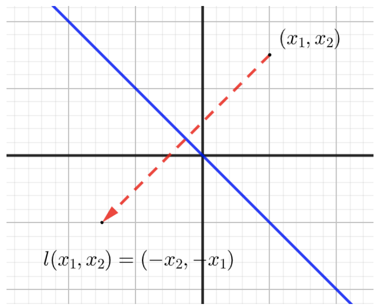



Figure 7: Reflection across the bisector of the second and fourth quadrant, $x_2 = -x_1$.

## Shear transformations

A shear transformation is a transformation of the type

$$S_h = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \qquad \text{(horizontal shear)}$$

or of the type

$$S_v = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix} \qquad \text{(vertical shear)}$$

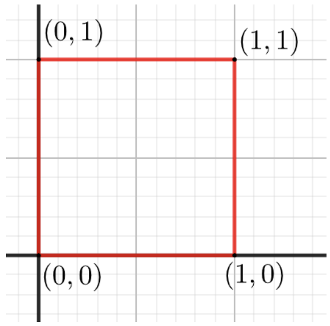
for $k > 0$.

This type of transformation maps the unit square to a parallelograms. In general, a horizontal shear will map horizontal lines to horizontal lines and vertical lines to oblique lines. A vertical shear will do the opposite: horizontal lines will be mapped to oblique lines and vertical lines will be mapped to vertical lines.

Consider the unit square $[0, 1] \times [0, 1]$. In order to understand where such square is mapped, we can consider the image of its vertices. Since both $S_h$ and $S_v$ are linear, $(0, 0)$ is mapped to itself. As for the other vertices, we have

$$\begin{aligned} &S_h(1, 0) = (1, 0), \\ &S_h(0, 1) = (k, 1), \\ &S_h(1, 1) = S_h(0, 1) + S_h(1, 0) = (k + 1, 1). \end{aligned}$$

Thus $S_h$ maps the unit square to a parallelogram of vertices $(0, 0), (1, 0), (k, 1)$ and $(1 + k, 1)$.



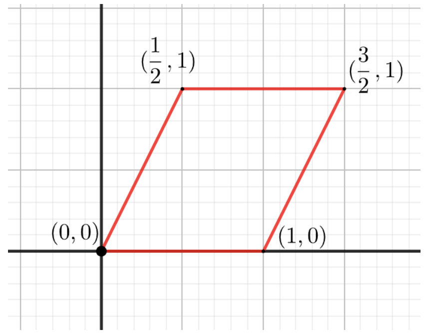




Figure 8: Horizontal shear transformation of the unit square $[0,1] \times [0,1]$ by $S_h = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$.

On the other hand,

$$
\begin{aligned}
&S_v(1,0) = (1,k), \\
&S_v(0,1) = (0,1), \\
&S_v(1,1) = S_v(0,1) + S_v(1,0) = (1, k+1).
\end{aligned}
$$

The unit square is therefore mapped to the parallelogram of vertices $(0,0)$, $(0,1)$, $(1,k)$ and $(1, 1+k)$.

## Affine transformations: translations

An *affine transformation* on $\mathbb{R}^m$ is a function $F : \mathbb{R}^m \to \mathbb{R}^n$ of the form $F(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ for $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. It is therefore a function that differs from a linear function by an additive term $\mathbf{b}$.

In $\mathbb{R}^2$, the additive component $\mathbf{b}$ of the affine transformation can be seen as translation by a vector $\mathbf{b}$. In this sense, if $F$ is an affine function, then for every $\mathbf{x}$, $F(\mathbf{x}) - \mathbf{b}$ is a linear transformation. Given an affine function, it is generally useful to consider its linear component $f(\mathbf{x}) = F(\mathbf{x}) - F(\mathbf{0})$. In fact $F(\mathbf{0}) = A\mathbf{0} + \mathbf{b} = \mathbf{b}$ is nothing else than the translation vector we're considering.

**Example**: Consider an affine function mapping the points $(0,0)$, $(1,0)$ and $(0,1)$ respectively to $(3,2)$, $(3,3)$ and $(7,3)$. Since $f(0,0) = (3,2) \neq (0,0)$, this function is not linear, but rather it has the form $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ for $\mathbf{b} = f(0,0) = (3,2)$. Then it makes sense to consider the linear component $f_l(\mathbf{x}) = f(\mathbf{x}) - f(0,0)$. Such component maps $(0,0)$ to itself, $(1,0)$ to $(0,1)$ and $(0,1)$ to $(4,1)$. The matrix form of $f$ is therefore given by

$$
f(x_1, x_2) = \begin{bmatrix} 0 & 4 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 3 \\ 2 \end{bmatrix}
$$

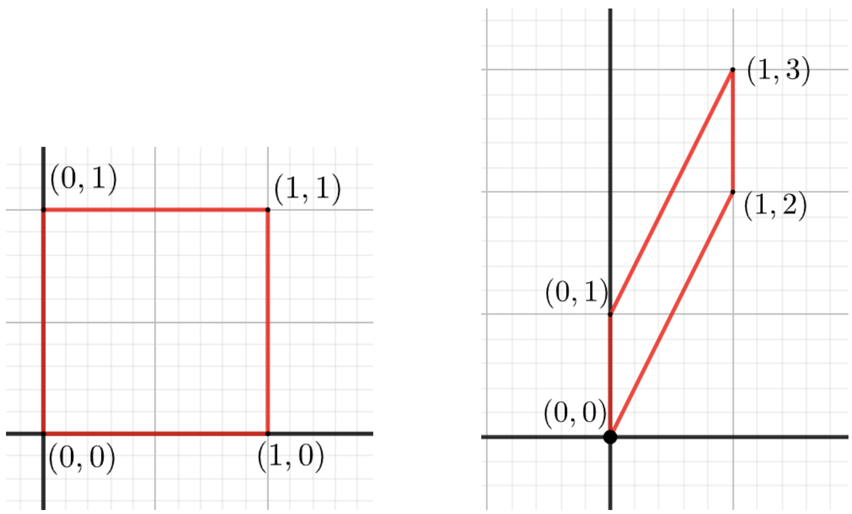



Figure 9: Vertical shear transformation of the unit square $[0, 1] \times [0, 1]$ by $S_h = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$.

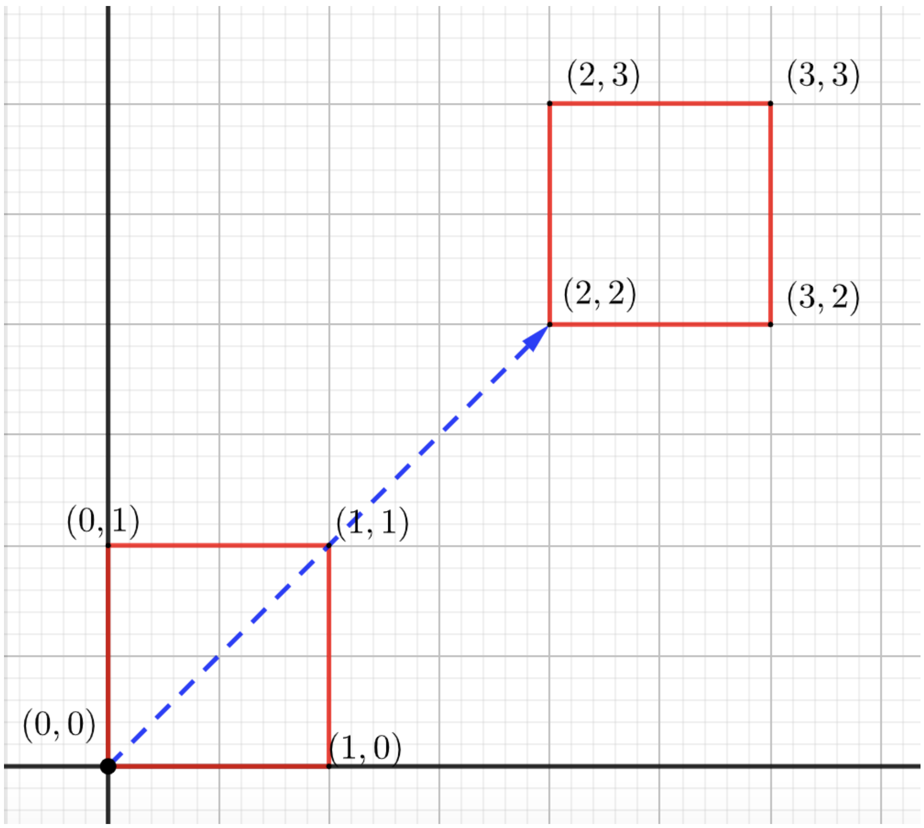



Figure 10: Translation of the unit square $[0, 1] \times [0, 1]$ by the vector $b = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$