Why do we want vectors to be additive? Because we generalize the properties of the usual 3D vectors we learn in school:

1. Vectors can be added to obtain other vectors, and they can be multiplied by a real number, the multiplication by a number being distributive with respect to the addition.
2. Using this addition, any vector can be decomposed into a sum of those that constitute the reference fame (as in Cartesian coordinates). We need two coordinates to describe the end point of a vector on a plane, and three to describe the end point of a vector in a volume, so we say that planes are two-dimensional and volumes are three-dimensional.
3. If we know that a 3D-vector ends on a certain facet of a polyhedron (or on a plane), we only need two coordinates to uniquely define it, using a reference frame on the facet formed by two adjacent edges (or two non-collinear vectors on the plane). If we know that a 3D-vector ends inside a certain polyhedron, we generally still need three numbers to describe its positions. The numbers can be chosen as coordinates in a reference frame formed by three edges of a trihedral angle.
4. One can measure angles between vectors, and find the projection of one vector onto another. The length of the projecton of v_1 on v_2 is $|v_1|cos(angle(v_1, v_2)) = \frac{(v_1, v_2)}{|v_2|}$ where (v_1, v_2) is the scalar product, so $angle(v_1, v_2) = \arccos(\frac{(v_1, v_2)}{|v_1||v_2|})$

The strict way to formulate these properties of geometrical vectors is the following:

1. Geometrical vectors form a vector space.

2. This vector space contains a set of vectors called **basis**, which we call (in geometry) a reference frame. Each vector of the vector space can be represented in a unique way as a **linear combination** (weighted sum) of elements of the basis. For this to be true, elements of the basis must be **linearly independent**. If a vector space has multiple bases, they all have the same number of elements, which is called the **dimension** of the vector space.

3. Linear combinations of fixed pairs of vectors in 3D Euclidean space form 2D-, 1D- and 0D-subspaces of that space, planes and straight lines that pass through the zero. Such planes and lines are the **spans** of their corresponding pairs of vectors, and we say that these pairs **span** their corresponding spans. Triples of vectors can either span the whole 3D space, a plane, a line or the zero.

4. There is a structure of scalar product (bilinear nondegenerate form) defined on the Euclidean space, with signature $(1, 1, 1)$. That structure gives us handy formulas for defining projections, surface areas and volumes.

Why do we need to generalize these properties on other objects?

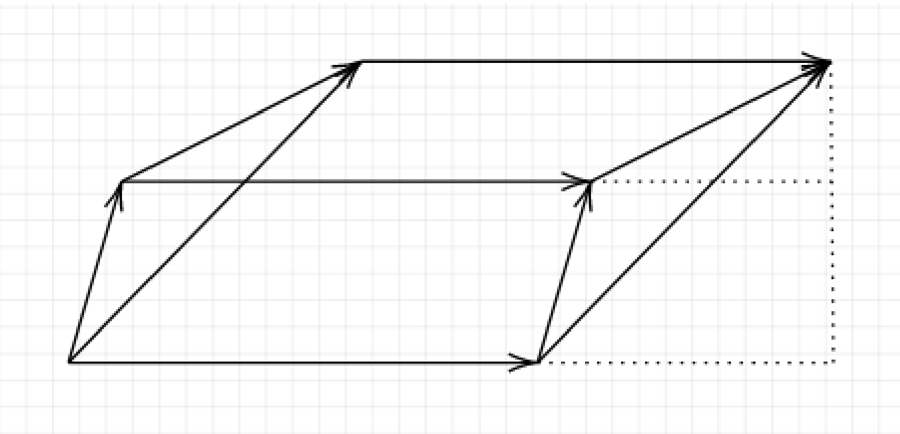
The three reasons why we need linear combination of vectors and bases

Reason 1: Studying linear functions, we only need to understand how they act on each element of the basis

Linearity allows us to work with elements of the basis, rather than with a generic element of the linear space. This is good because one can often choose a convenient basis which is easy to work with.

Example: volume of a parallelorgram

Volume of a parallelogram is multilinear in the vectors that constitute the parallelogram:



and so should be a linear map \map$Vol$\R$^2 \times$ \R^2\R
The problem: $Vol(-v_1, v_2) = Vol(v_1, v_2)$
So the linearity is broken.
Solution: consider the "oriented volume": $Vol(-v_1, v_2) = -Vol(v_1, v_2)$

Then, due to linearity, we can express $Vol(v_1, v_2)$ as

$$Vol(v_1, v_2) = Vol(v_{11}e_1 + v_{12}e_2, v_{21}e_1 + v_{22}e_2) =$$

$$= v_{11}v_{21}Vol(e_1, e_1) + v_{11}v_{22}Vol(e_1, e_2) + v_{12}v_{21}Vol(e_2, e_1) + v_{12}v_{22}Vol(e_2, e_2)$$

Since $Vol(e_1, e_1) = -Vol(e_1, e_1) = 0 = Vol(e_2, e_2)$, and $Vol(e_2, e_1) = -Vol(e_1, e_2)$,

$$Vol(v_1, v_2) = (v_{11}v_{22} - v_{12}v_{21})Vol(e_1, e_2) = (v_{11}v_{22} - v_{12}v_{21})$$

Introduce the distributive associative operation "outer product":

$$Vol(v_1, v_2) := v_1 \wedge v_2$$

$$e_i \wedge e_j = -e_j \wedge e_i,\ \alpha e_i \wedge e_j = \alpha(e_i \wedge e_j)$$

Example for the example: calculate the oriented volume of $((1,2),(3,4))$ [in orthonormal basis e_1, e_2].

$$(1e_1 + 2e_2) \wedge (3e_1 + 4e_2) = 4e_1 \wedge e_2 + 6e_2 \wedge e_1 = -2e_1 \wedge e_2$$

(note that $e_1 \wedge e_1$ and $e_2 \wedge e_2$ are not present since they equal 0).

What we have, in fact, just did, is introduced a four-dimensional vector space, elements of which are *ordered pairs of vectors in* $\R^3$ *and their linear combinations*, which we then *factorized by the relation* $(v_i, v_j) + (v_j, v_i) = 0$, which made it a one-dimensional vector space with a basis $e_1 \wedge e_2$, and by the choice of a convenient basis in the original $4D-$space we made calculations eazy-peazy lemon squeazy.
Such $4D$ space is, in fact, called the tensor product $\R^2 \otimes \R^2$, and our new $1D-$space is the subpace of this tensor product, called the space of fully anti-symmetric $(0,2)$-tensors (or bivectors), $\R \bigwedge \R$.

We can generalize to higher dimensions, by the same logic.
Then we can calculate volumes of figures in higher dimensions like this:

$$Vol(v_1, v_2, \cdots, v_n) = v_1 \wedge v_2 \wedge \cdots \wedge v_n$$

One can observe that the result is nothing more than the determinant of $(v_1, v_2, \cdots, v_n)$, which is, in fact, the reason why the determinant has such an "arbitrary" definition.

The concept of tensor and outer products plays a foundational role in many fields of mathematics, physics and computer science, including differential geometry, analytical mechanics and dynamical systems, quantum mechanics, thermodynamics and field theory.

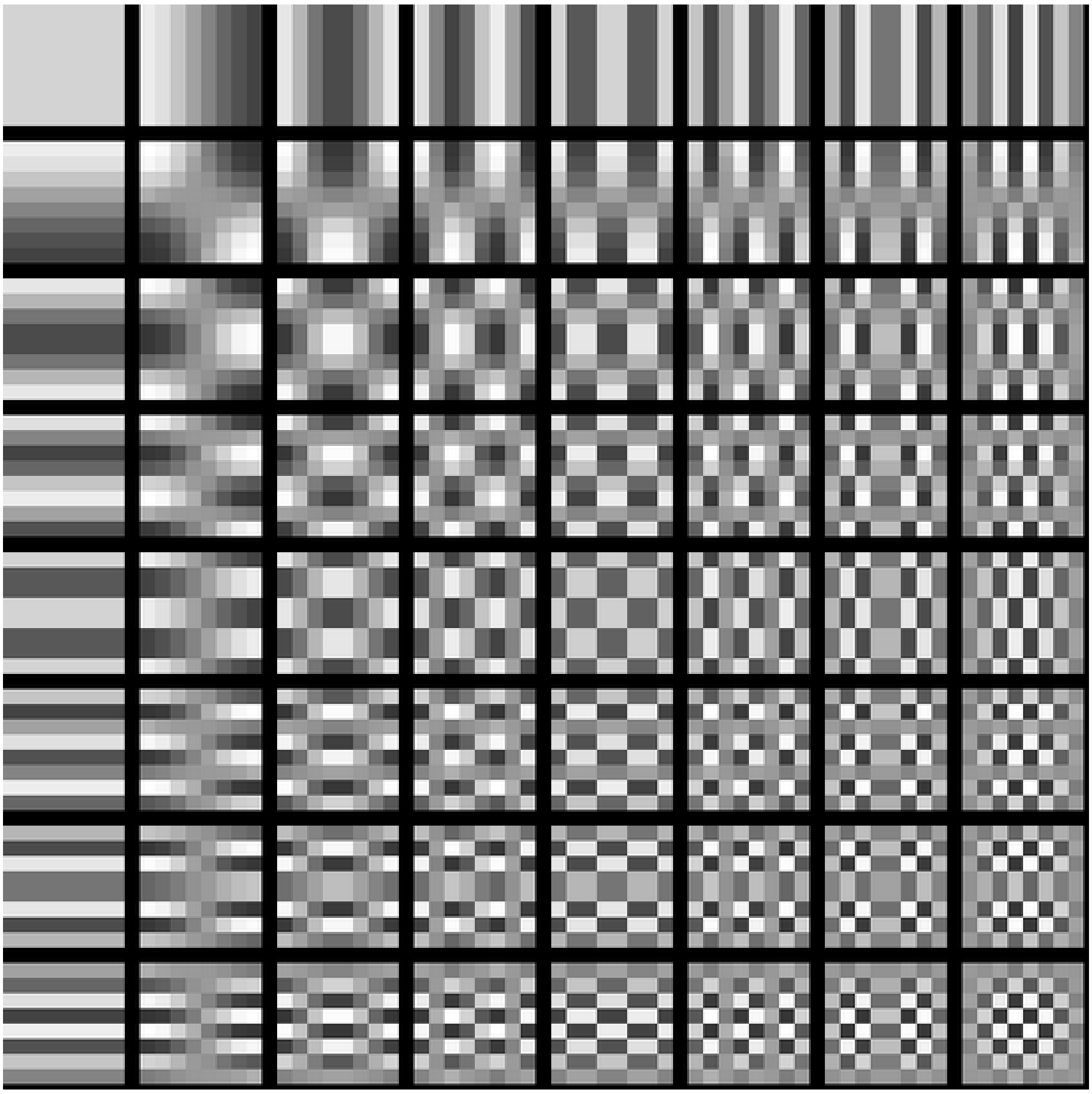
Reason 2: We can study the structure of objects, represented by vectors, by studying their decompositions in convenient bases

You can decompose objects from linear space into their "substituent parts", giving you some insight about the structure of those objects, and allowing you to manipulate them in a convenient way.

Example: DFT and image processing.

One can (obviously) represent an arbitrary $n \times n$ black-and-white image as an element of an n^2-dimensional "vector space"[1] over the "field"[2] $\mathbb{F}_{256}$ into a sum of n^2 vertical and horizontal "harmonics" of different frequencies. This can be used, for example, for

high-frequency noize filtering or effective data compression.

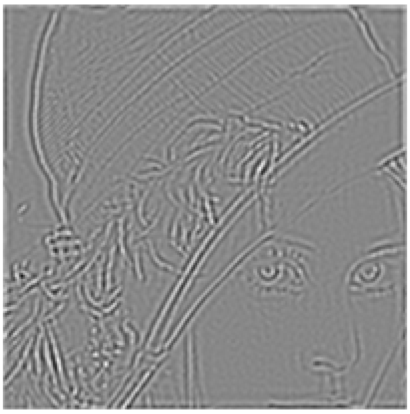


On the top: the 64 harmonics of an 8×8 image. Credit

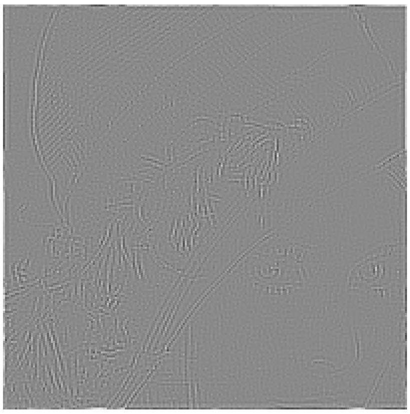
On the bottom: an original an its Fourier-blurred version. Credit

(a) Original image

(b) Low frequency component



(c) Middle frequency component



(d) High frequency component

How exactly one can perform such a decomposition is, again, defined by a certain structure we impose on $\mathbb{F}_{256}$, namely, again a certain "inner product", which is a topic for another time.

Reason 3: Vectors are geometrical objects that don't depend on the choice of coordinates

A vector is an abstract object; here vectors are represented by lists of numbers, but in reality lists of numbers are only representations which depend on the basis, the vector itself has no intrinsic list of coordinates. Different bases represent "different ways to look at the system".

Exmaple: 3D rotations, Lorenz boosts

If we choose a reference frame in our familiar $3D$ Euclidean space which is aligned with our orientation in space, we can describe the position of a certain object as a list of numbers. However, if we turn, the position of the object remains the same, but the list of numbers we use to describe it changes. That is an example of the same geometrical object (vector, describing the position of the object) having two different representations in two different bases, which corresponds to the different ways in

which the object can be observed. It turns out that the relativistic effects of the Special Relativity have the same nature, but the rotations are of a different kind, they are the so-called "Lorenz boosts", which correspond to moving in a certain direction with a nonzero speed, hence the name, and the "rotation" is in fact in the plane which includes the time axis. The difference between the two kinds of rotations is again defined by the structure of the "inner product" introduced in the linear space in question ($\R^4$). Moreover, the local effects of the General Relativity are also defined by the structure of the inner product, which changes between different points of spacetime.

[1] In fact, module, but in this context the difference does not matter.
[2] Well, in fact, ring, which is why "the vector space" is not a real vector space, but rather a module, an analogous to vector spaces structure over rings.

Strict definitions and important theorems

Suppose V is a vector space over a field F. Suppose $v_1, v_2, \cdots v_i \in V$, $a_1, a_2, \cdots, a_i \in F$, $i \in \mathbb{N}$.

A *linear combination* of the list of vectors $v_1, v_2, \cdots, v_i$ with coefficients $a_1, a_2, \cdots, a_i$ is defined as

$$a_1 v_1 + a_2 v_2 + \cdots v_i a_i$$

The span \bra (v_1, \cdots, v_i) \ket is the set of all linear combinations of $\{v_k\}_{k=1}^{i}$.

It is trivial to check that the span of a list is a vector space, which is a subspace of V. We say that a list spans a subspace $W \subset V$ if its span contains W.

A spanning list of V is a list that spans V. It is trivial to check that if $W_1, W_2 \subset V$ are subspaces of V, then $W_1 \cap W_2$ (and, in fact, of any number of $W's$) is likewise a subspace of V. Take the intersection of all subspaces of V that contain a list $\{v_k\}$. It is also a subspace, and so it must contain the span of $\{v_k\}$. But it is the smallest such subspace that contains $\{v_k\}$, so it is the span of $\{v_k\}$ itself (since, if the span were smaller, it wouldn't contain some of the vectors in $\{v_k\}$), and vice versa, the span of $\{v_k\}$ is the smallest subspace that contains all of $\{v_k\}$.

If a list $\{v_k\}$ spans V, any vector of V can be represented as a linear combination of vectors in the list. If this representation is unique, we say that $\{v_k\}$ forms a basis in V. This is analogous to all vectors of $\{v_k\}$ being linearly independent, which means no vector on the list can be expressed as a linear combination of all the others:

$$\sum_{k=1} a_k v_k = 0 \iff \forall k : a_k = 0$$

Indeed, if it was possible to represent one vector of the list as a linear combination of the others, we could just toss if off, and repeat this action until all vectors left are

linearly independent without losing any vectors of the span. If a vector space has a finite spanning list, we say it is finite-dimensional. By the procedure above any finite-dimensional vector space has a basis, since the number of vectors in the list can't become negative. Any two bases have the same number of elements: indeed, if we had two bases of different dimensions, we colud express each vector of the bigger in terms of the smaller, and then

$$1^i b_i = k_i^j a_j = k'^i a_i$$