One of the key driving forces for the development of linear algebra was the profound idea (going back to 17th century and the work of Pierre Fermat and Rene Descartes) that geometry can be studied via algebraic methods by introducing coordinates. Interesting geometric objects and relations between them can be then viewed in terms of equations which can be studied using algebraic methods.

In geometry and in many applications, vectors appear naturally equipped with additional structure: the concept of a length and the concept of an angle between two vectors. To motivate the ideas leading to such concepts let us recall some basic facts from plane geometry.

A starting point is the theorem of cosines, which should be thought of as an extension of Pythagoras theorem, and which (in a slightly different form) has been known at least since the times of Euclid (3rd century B.C.). To recall this theorem, consider a triangle ABC and let α be the angle $\angle BAC$. Then we have

Theorem of cosines. $BC^2 = AB^2 + AC^2 - 2AB \cdot AC \cos\alpha$.

Suppose now that we use cartesian coordinates in which our points are $A = (0, 0)$, $B = (s, t)$ and $C = (p, q)$. Then $AB^2 = s^2 + t^2$, $AC^2 = p^2 + q^2$ and

$$BC^2 = (s - p)^2 + (t - q)^2 = s^2 - 2sp + p^2 + t^2 - 2tq + q^2 = AB^2 + AC^2 - 2(sp + tq).$$

Comparing this to the theorem of cosines, we conclude that

$$AB \cdot AC \cos\alpha = sp + tq.$$

Thinking now in terms of vectors $\overrightarrow{AB} = v = (s, t)$ and $\overrightarrow{AC} = w = (p, q)$, we see that $|v||w| \cos\alpha = v \cdot w$, where on the right we have the dot product of the vectors v and w and on the left $|v|$ and $|w|$ denote the length of v, w and α is the angle between v and w. Note also that $|v| = \sqrt{s^2 + t^2} = \sqrt{v \cdot v}$, and similarly $|w| = \sqrt{w \cdot w}$. Thus, the geometric concepts of length and angle can be expressed in terms of the dot product: the length of a vector v is $|v| = \sqrt{v \cdot v}$ and the angle $\angle(v, w)$ between vectors v and w in the angle $[0, \pi]$ such that

$$\cos\angle(v, w) = \frac{v \cdot w}{|v||w|}.$$

Similar formulas can be easily obtained for vectors in $\mathbb{R}^3$ and we could use dot product in $\mathbb{R}^n$ to define the concept of length and angle using the same formulas. Unfortunately, the dot product makes only sense in $\mathbb{R}^n$, while we would like to extend the ideas of length and angle to all vector spaces. Thus we need to find a good analog of the dot product which works in an arbitrary vector space. A nice example to think about is a subspace V of $\mathbb{R}^n$, for example the plane $x_1 + x_2 + x_3 = 0$ in $\mathbb{R}^3$. Since we can measure length and angle in $\mathbb{R}^3$, we can restrict our attention just to the subspace and study the length and the angles for vectors in the subspace. Suppose now that we forget that our subspace lives inside $\mathbb{R}^3$ (which is equipped with nice system of coordinates) so we need to find formulas for length and angle in the subspace without any reference to the bigger space $\mathbb{R}^3$. For example, we can choose a basis of our subspace, say $(1, 1, -2), (1, -2, 1)$ in our specific example, and then we want to find formulas for length and angle in terms of the coordinates in the chosen basis. In our example, if v has coordinates (a, b) and w has coordinates (e, f) then $v = (a + b, a - 2b, -2a + b)$ and $w = (e + f, e - 2f, -2e + f)$. Thus

$$v \cdot w = (a + b)(e + f) + (a - 2b)(e - 2f) + (-2a + b)(-2e + f) = 6ae - 3af - 3be + 6bf$$

and $|v| = \sqrt{6a^2 - 6ab + 6b^2}$.

After analyzing several similar examples it is not hard to come with the following concept, which will play the role of a dot product in an abstract vector space V.

Definition. An **inner product** on a vector space V is a function $V \times V \longrightarrow \mathbb{R}$ which to any pair of vectors u, w of V assigns a real number, which we will denote by $< u, w >$, such that the following conditions are satisfied:

1. $< u, w >=< w, u >$ for any two vectors $u, w \in V$.
2. $< au_1 + bu_2, w >= a < u_1, w > +b < u_2, w >$ for any vectors $u_1, u_2, w \in V$ and any numbers a, b.
3. $< u, u >$ is positive (i.e. $< u, u >> 0$) for any **non-zero** vector $u \in V$.

We often state property 1 by saying that the inner product is **symmetric**. The symmetry and property 2 easily imply that

$$< w, au_1 + bu_2 >= a < w, u_1 > +b < w, u_2 > \quad \text{for any vectors } u_1, u_2, w \in V \text{ and any numbers } a, b.$$

For that reason we often state property 2 by saying that the inner product is **bilinear**, i.e. when one of the vectors is fixed, the inner product is a linear function of the second vector. Note that property 2 implies that $< 0, w >= 0$ for any vector w. In particular $< 0, 0 >= 0$. Property 3 is often stated by saying that the inner product is **positive definite**. It implies that $< v, v >= 0$ if and only if $v = 0$.

Here are some important examples of inner products.

Example 1. Let $V = \mathbb{R}^n$ and let $< u, w. = u \cdot w$ be the dot product. We leave it as a simple exercise that this is an inner product on $\mathbb{R}^n$.

Example 2. Let V be a vector space with an inner product $< \, , \, >$. If W is a subspace of V then the restriction of the inner product $< \, , \, >$ to the subspace W is an inner product on W. Starting with the dot product on $\mathbb{R}^n$ and restricting it to various subspaces provides a large family of vector spaces with an inner product.

Example 3. Consider the space $\mathbb{P}_n$ of all polynomials of degree at most n. Let $a < b$ we real numbers and define $< f, g >= \int_a^b f(x)g(x)dx$. Again , we leave this as a simple exercise that this defines an inner product on $\mathbb{P}_n$. As a matter of fact, the same formula defines an inner product on the space of all continuous function on an interval containing the numbers a and b.

Example 4. Let $V = \mathbb{R}^2$. Define $< (a, b), (c, d) >= 6ac - 3ad - 3bc + 6bd$. It is not hard to check that this is an inner product on V.

Our next step is to see that the concept of an inner product is indeed a good generalization of the dot product.

Theorem. Let $< \, , \, >$ be an inner product on a vector space V.

1. $2 < u, w >=< u + w, u + w > - < u, u > - < w, w >$ for any vectors $u, w \in V$.
2. If $u, w \in V$ and $< v, u >=< v, w >$ for every $v \in V$ then $u = w$.
3. **Cauchy-Schwartz inequality:** $< u, w >^2 \leq< u, u > \cdot < w, w >$ for any $u, w \in V$.

To see 1. we use the bilinear property of the inner product:

$$< u + w, u + w >=< u, u + w > + < w, u + w >=< u, u > + < u, w > + < w, u > + < w, w >=$$

$$=< u, u > + < w, w > -2 < u, w >$$

which clearly is equivalent to 1.

For 2. note that the assumption means that $< v, u - w >= 0$ for every v. Taking $v = u - w$ we get $< v, v >= 0$ which means that $v = u - w = 0$, i.e. $u = w$ (recall that inner product is positive definite).

The Cauchy-Schwartz inequality is very important. As we will see soon, it is this inequality which will allow us introduce the concepts of a length and an angle and show that they have the desired properties. To justify the Cauchy-Schwartz inequality, consider $u, w \in V$ and note that for every real number t we have $< u + tw, u + tw > \geq 0$. Note that

$$< u + tw, u + tw >=< u, u > +2 < u, w > t+ < w, w > t^2.$$

Now, considered as a function of t, the expression above is a quadratic polynomial, and we know that it is always non-negative. This means that this quadratic polynomial either has no real roots or has exactly one real root. Recall that this happens if and only if the discriminant of the quadratic polynomial is ≤ 0:

$$(2 < u, w >)^2 - 4 < u, u > \cdot < w, w > \leq 0.$$

This is clearly equivalent to the Cauchy-Schwartz inequality.

Exercise. Show that the equality holds in the Cauchy-Schwartz inequality if and only if either $u = 0$ or $w = tu$ for some t (i.e. the vectors u, w are linearly dependent).

We now follow our original discussion an introduce the following definition.

Definition. Let $< \, , \, >$ be an inner product on a vector space V. The **length** $||v||$ of a vector v is defined as $||v|| = \sqrt{< v, v >}$.

The following result should be a convincing evidence that this is the right definition.

Theorem. Let $< \, , \, >$ be an inner product on a vector space V.

1. $||v|| > 0$ for any non-zero vector $v \in V$ and $||0|| = 0$.
2. If $||av|| = |a|||v||$ for every $v \in V$ every number a.
3. $| < u, w > | \ \leq ||u|| \cdot ||w||$ for any $u, w \in V$.
4. **triangle inequality:** $||u \pm w|| \ \leq ||u|| + ||w||$ for any $u, w \in V$.

Property 1 is just a different way of saying that the inner product is positive definite.

For property 2, note that $< av, av >= a < v, av >= a^2 < v, v >$. Taking square roots yields 2.

Property 3 is just a restatement of the Cauchy-Schwartz inequality (by taking square roots of both sides).

For property 4 note that

$$||u \pm w||^2 =< u \pm w, u \pm w >=< u, u > \pm 2 < u, w > + < w, w >= ||u||^2 \pm 2 < u, w > +||w||^2 \leq$$

$$\leq ||u||^2 \pm 2||u|| \cdot ||w|| + ||w||^2 = (||u|| + ||w||)^2$$

(we used property 3: $\pm < u, w > \leq ||u|| \cdot ||w||$).

Exercise. Show that $||u + w|| = ||u|| + ||w||$ if and only if either $w = 0$ or $u = tw$ for some $t \geq 0$.

We can now define the angle between two vectors.

Definition. Let $< \, , \, >$ be an inner product on a vector space V. The **angle** $\angle(u,w)$ between two vectors $u, w \in V$ is the unique angle in the interval $[0, \pi]$ such that $\cos\angle(u,w) = \frac{< u, w >}{||u|| \cdot ||w||}$.

To see that this definition is meaningful, recall cos defines a bijection between $[0, \pi]$ and $[-1, 1]$. In other words, for any number t in $[-1, 1]$ there is unique $\alpha \in [0, \pi]$ such that $\cos\alpha = t$. Since $\frac{< u, w >}{||u|| \cdot ||w||}$ is in $[-1, 1]$ by property 3. from the last theorem, the angle $\angle(u,w)$ is indeed well defined.

Having defined the angle between vectors we can now speak about perpendicular vectors. Recall that $\cos(\pi/2) = 0$ so $\angle(u,w) = \pi/2$ if and only if $< u, w >= 0$. This leads us to the following definition.

Definition. Let $< \, , \, >$ be an inner product on a vector space V. We say that vectors $u, w \in V$ are **orthogonal** (or **perpendicular**), and write $u \perp w$ if $\angle(u,w) = \pi/2$. Equivalently, $u \perp w$ if and only if $< u, w >= 0$

For example, in $\mathbb{R}^n$ with the dot product as the inner product, any two different vectors in the standard basis are orthogonal. Basis with this property are very convenient when working with inner products. Thus we introduce the following definition.

Definition. Let $< \, , \, >$ be an inner product on a vector space V. A sequence of vectors $v_1, v_2, \ldots, v_k$ is called **orthogonal** if all the vectors are non-zero and they are orthogonal to each other, i.e. $v_i \perp v_j$ for $i \neq j$. We say that a sequence of vectors $v_1, v_2, \ldots, v_k$ is **orthonormal** if it is orthogonal and all the vectors have length 1.

A vector of length 1 is often called a **unit vector**. Note that if v is a non-zero vector then $\frac{1}{||v||}v$ is a unit vector. For simplicity, we write $\frac{v}{||v||}$ for $\frac{1}{||v||}v$. The following observation is straightforward.

If $v_1, v_2, \ldots, v_k$ is orthogonal then $v_1/||v_1||, v_2/||v_2||, \ldots, v_k/||v_k||$ is orthonormal.

The following result will play important role.

Theorem. If $v_1, \ldots, v_k$ is an orthogonal sequence then it is linearly independent.

Indeed, suppose that $a_1v_1 + a_2v_2 + \ldots a_kv_k = 0$. By the bilinear property of the inner product we have

$$0 =< 0, v >=< a_1v_1 + a_2v_2 + \ldots a_kv_k, v >= a_1 < v_1, v > +a_2 < v_2, v > + \ldots + a_k < v_k, v >$$

for any vector v. For $i = 1, \ldots k$, take $v = v_i$ and note that $< v_j, v_i >= 0$ for $j \neq i$. Thus $0 = a_i < v_i, v_i >$. Since $< v_i, v_i >> 0$, we conclude that $a_i = 0$. Thus the only way $a_1v_1 + a_2v_2 + \ldots a_kv_k = 0$ is when $a_1 = \ldots = a_k = 0$. This means that our vectors are linearly independent.

The last theorem implies that an orthogonal sequence can not be longer that the dimension of V. It is natural to ask whether we can always find an orthogonal basis. We will see that the answer is positive, In fact, we will soon learn a very nice procedure which changes and linearly independent sequence of vectors into and orthogonal sequence. The procedure is based on the following observation. Suppose $w_1, \ldots, w_k$ is an orthogonal sequence. Suppose v is a non-zero vector which is not a linear combination of the vectors $w_1, \ldots, w_k$. Can we modify v by some linear combination of $w_1, \ldots, w_k$ to get a vector orthogonal to each of the vectors $w_1, \ldots, w_k$? In other words, we are asking if there are numbers $a_1, \ldots, a_k$ such that $v - (a_1 w_1 + \ldots + a_k w_k)$ is orthogonal to w_i for $i = 1, \ldots, k$. This means that

$$< v - (a_1 w_1 + \ldots + a_k w_k), w_i >=< v, w_i > - a_i < w_i, w_i >= 0 \quad \text{so} \quad a_i = \frac{< v, w_i >}{< w_i, w_i >}.$$

We see that there is a unique choice of the coefficients $a_1, \ldots, a_k$ which satisfies our requirements. This observation leads to the following procedure.

Gramm-Schmidt orthogonalization process. Let $< \, , \, >$ be an inner product on a vector space V and let $v_1, \ldots, v_k$ be linearly independent. Define recursively vectors $w_1, \ldots, w_k$ as follows:

$$w_1 = v_1, \quad w_{j+1} = v_{j+1} - \left(\frac{< v_{j+1}, w_1 >}{< w_1, w_1 >} w_1 + \frac{< v_{j+1}, w_2 >}{< w_2, w_2 >} w_2 + \ldots + \frac{< v_{j+1}, w_j >}{< w_j, w_j >} w_j \right).$$

Then

1. $w_1, w_2, \ldots, w_k$ is orthogonal.
2. $\mathrm{span}\{w_1, \ldots, w_j\} = \mathrm{span}\{v_1, \ldots, v_j\}$ for $j = 1, \ldots, k$.

The justification of 1 and 2 is quite simple. Assuming that $w_1, w_2, \ldots, w_j$ is orthogonal and $\mathrm{span}\{w_1, \ldots, w_j\} = \mathrm{span}\{v_1, \ldots, v_j\}$ we see that $w_{j+1} - v_{j+1} \in \mathrm{span}\{w_1, \ldots, w_j\}$, and therefore

$$\mathrm{span}\{w_1, \ldots, w_j, w_{j+1}\} = \mathrm{span}\{w_1, \ldots, w_j, v_{j+1}\} = \mathrm{span}\{v_1, \ldots, v_j, v_{j+1}\}.$$

Furthermore, w_{j+1} is orthogonal to each of $w_1, \ldots, w_j$ by the computation which led us to the Gramm-Schmidt process.

One could try to perform the Gramm-Schmidt orthogonalization process without knowing up-front that $v_1, \ldots, v_k$ are linearly independent. However, if these vectors are actually linearly dependent then we will get $w_j = 0$ for some j and then the process can not be continued. Thus Gramm-Schmidt orthogonalization process could be used to detect linear dependence.

Example. Consider the basis $v_1 = (1,1,0)$, $v_2 = (1,0,1)$, $v_3 = (1,1,1)$ of $\mathbb{R}^3$ with dot product as the inner product. Let us apply the Gramm-Schmidt orthogonalization process to the vectors v_1, v_2, v_3. We have

<u>step 1:</u>

$$w_1 = v_1 = (1,1,0) \quad \text{and} \quad < w_1, w_1 >= (1,1,0) \cdot (1,1,0) = 2,$$

<u>step 2:</u>

$$< v_2, w_1 >= (1,0,1) \cdot (1,1,0) = 1$$

so

$$w_2 = v_2 - \frac{< v_2, w_1 >}{< w_1, w_1 >} w_1 = (1,0,1) - \frac{1}{2}(1,1,0) = (\frac{1}{2}, \frac{-1}{2}, 1),$$

and

$$< w_2, w_2 >= (\frac{1}{2}, \frac{-1}{2}, 1) \cdot (\frac{1}{2}, \frac{-1}{2}, 1) = \frac{3}{2}.$$

<u>step 3:</u>

$$< v_3, w_1 >= (1,1,1) \cdot (1,1,0) = 2, \quad < v_3, w_2 >= (1,1,1) \cdot (\frac{1}{2}, \frac{-1}{2}, 1) = 1$$

so

$$w_3 = v_3 - \frac{< v_3, w_1 >}{< w_1, w_1 >} w_1 - \frac{< v_3, w_2 >}{< w_2, w_2 >} w_2 = (1,1,1) - \frac{2}{2}(1,1,0) - \frac{1}{\frac{3}{2}}(\frac{1}{2}, \frac{-1}{2}, 1) = (\frac{-1}{3}, \frac{1}{3}, \frac{1}{3}).$$

We see that w_1, w_2, w_3 is an orthogonal basis of $\mathbb{R}^3$ (verify this!). Thus the vectors $\frac{w_1}{||w_1||} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$, $\frac{w_2}{||w_2||} = (\frac{1}{\sqrt{6}}, \frac{-1}{\sqrt{6}}, \frac{\sqrt{2}}{\sqrt{3}})$, $\frac{w_3}{||w_3||} = (\frac{-1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$ form an orthonormal basis of $\mathbb{R}^3$.

Example. Let $V = \mathbb{P}_3$ be the space of polynomials of degree at most 3 with the inner product $< f, g >= \int_0^1 f(x)g(x)dx$. Consider the basis $v_1 = 1$, $v_2 = x$, $v_3 = x^2$, $v_4 = x^3$ of V. Let us apply the Gramm-Schmidt orthogonalization process to the vectors v_1, v_2, v_3, v_4. We have

step 1:

$$w_1 = v_1 = 1 \quad \text{and} \quad < w_1, w_1 >= \int_0^1 1dx = 1,$$

step 2:

$$< v_2, w_1 >= \int_0^1 xdx = \frac{1}{2}$$

so

$$w_2 = v_2 - \frac{< v_2, w_1 >}{< w_1, w_1 >} w_1 = x - \frac{\frac{1}{2}}{1} \cdot 1 = x - \frac{1}{2},$$

and

$$< w_2, w_2 >= \int_0^1 (x - \frac{1}{2})^2 dx = \frac{1}{12}.$$

step 3:

$$< v_3, w_1 >= \int_0^1 x^2 dx = \frac{1}{3}, \quad < v_3, w_2 >= \int_0^1 x^2(x - \frac{1}{2})dx = \frac{1}{12}$$

so

$$w_3 = v_3 - \frac{< v_3, w_1 >}{< w_1, w_1 >} w_1 - \frac{< v_3, w_2 >}{< w_2, w_2 >} w_2 = x^2 - \frac{\frac{1}{3}}{1} \cdot 1 - \frac{\frac{1}{12}}{\frac{1}{12}}(x - \frac{1}{2}) = x^2 - x + \frac{1}{6}$$

and

$$< w_3, w_3 >= \int_0^1 (x^2 - x + \frac{1}{6})^2 dx = \frac{1}{180}.$$

step 4:

$$< v_4, w_1 >= \int_0^1 x^3 dx = \frac{1}{4}, \ < v_4, w_2 >= \int_0^1 x^3(x - \frac{1}{2})dx = \frac{3}{40}, \ < v_4, w_3 >= \int_0^1 x^3(x^2 - x + \frac{1}{6})dx = \frac{1}{120}$$

so

$$w_4 = v_4 - \frac{< v_4, w_1 >}{< w_1, w_1 >} w_1 - \frac{< v_4, w_2 >}{< w_2, w_2 >} w_2 - \frac{< v_4, w_3 >}{< w_3, w_3 >} w_3 = x^3 - \frac{\frac{1}{4}}{1} \cdot 1 - \frac{\frac{3}{40}}{\frac{1}{12}}(x - \frac{1}{2}) - \frac{\frac{1}{120}}{\frac{1}{180}}(x^2 - x + \frac{1}{6}) =$$

$$= x^3 - \frac{1}{4} - \frac{9}{10}(x - \frac{1}{2}) - \frac{3}{2}(x^2 - x + \frac{1}{6}) = x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20}.$$

Thus $1, x - \frac{1}{2}, x^2 - x + \frac{1}{6}, x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20}$ is an orthogonal basis of $\mathbb{P}_3$.