Lecture 8: Scalar products, norms and orthogonality

Let U, V and W be three real vector spaces. We say that the transformation[1]

$$G : U \times V \mapsto W \tag{2}$$

is *bilinear* if for all $u_1, u_2 \in U$, all $v_1, v_2 \in V$, and all $c \in \mathbb{R}$

1. $G(u_1 + u_2, v_1) = G(u_1, v_1) + G(u_2, v_1)$,
2. $G(u_1, v_1 + v_2) = G(u_1, v_1) + G(u_1, v_2)$,
3. $G(cu_1, v_1) = G(u_1, cv_1) = cG(u_1, v_1)$.

If the bilinear transformation G is real-valued, i.e.,

$$G : U \times V \mapsto \mathbb{R}, \tag{3}$$

then we say that F is a *bilinear form.*

Example 1: Consider $U = \mathbb{R}^m$, $V = \mathbb{R}^n$ and a matrix $A \in M_{m\times n}(\mathbb{R})$. Then

$$\begin{aligned} G : \mathbb{R}^m \times \mathbb{R}^n &\to \mathbb{R} \\ (u, v) &\to \sum_{i=1}^{m}\sum_{j=1}^{n} u_i A_{ij} v_j \end{aligned}$$

is a bilinear form.

If $U = V$ then we say that $G : V \times V \mapsto \mathbb{R}$ is a bilinear form on V. Moreover, if for all $v_1, v_2 \in V$ we have that

$$G(v_1, v_2) = G(v_2, v_1) \tag{4}$$

then we say that the bilinear form on V is *symmetric.*

Matrix associated to a bilinear form. Similarly to linear transformations, it is possible to define the matrix associated to a bilinear form on V. To this end, suppose that V is n-dimensional and consider the basis $\mathcal{B}_V = \{v_1, \dots, v_n\}$. Let $u, v \in V$ such that

$$u = x_1 v_1 + \cdots + x_n v_n, \qquad v = y_1 v_1 + \cdots + y_n v_n. \tag{5}$$

The coordinates of u and v relative to $\mathcal{B}_V$ are

$$[u]_{B_V} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \qquad [v]_{B_V} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}. \tag{6}$$

[1] The multiplication symbol $\times$ in (2) means "Cartesian product" of two sets. The elements of the set $U \times V$ are of pairs of vectors (u, v) where $u \in V$ and $v \in V$. We have already seen an example of a vector space constructed using multiple Cartesian products, i.e.,

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ times}}. \tag{1}$$

A substitution of u and v into $G(u, v)$ yields

$$G(u, v) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i G(v_i, v_j) y_j. \tag{7}$$

Define the matrix $A_{\mathcal{B}_V}$ associated with $G(u, v)$ relative to B_V as

$$A_{\mathcal{B}_V} = \begin{bmatrix} G(v_1, v_1) & \cdots & G(v_1, v_n) \\ \vdots & \ddots & \vdots \\ G(v_n, v_1) & \cdots & G(v_n, v_n) \end{bmatrix} \tag{8}$$

This allows us to write (7) as

$$G(u, v) = [u]_{B_V}^T A_{\mathcal{B}_V} [v]_{B_V}. \tag{9}$$

If G is symmetric then $A_{\mathcal{B}_V}$ is a symmetric matrix.

Scalar products. A scalar product on a real vector space V is a symmetric bilinear form on V. The scalar product between two vectors in V is a real number[2]. We denote such scalar product as

$$\langle u, v \rangle = G(u, v) \qquad \forall u, v \in V. \tag{10}$$

A scalar product in V is also called "inner product" in V.

Examples of scalar products:

1. $V = \mathbb{R}^n$: $\langle u, v \rangle = \sum_{i=1}^{n} u_i v_i$ (scalar product on $\mathbb{R}^n$). This scalar product is often called "dot-product" and denoted as $u \cdot v$. The matrix associated with the dot product is the identity matrix.

2. $V = \mathbb{R}^n$: $\langle u, v \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} u_i g_{ij} v_j$ where g_{ij} is a $n \times n$ symmetric matrix. In differential geometry and in the theory of general relativity g_{ij} is called *metric tensor* and it represents the metric properties of the space-time (the curvature of the space-time is a nonlinear function of g_{ij}).

3. $V = M_{n\times n}(\mathbb{R})$: $\langle A, B \rangle = \mathrm{Tr}(AB^T) = \sum_{i,j=1}^{n} A_{ij} B_{ij}$ for all $A, B \in M_{n\times n}(\mathbb{R})$ (scalar product between two matrices). Here $\mathrm{Tr}(AB^T)$ denotes the trace of the matrix AB^T, which is clearly a symmetric bilinear form[3].

[2]It is possible to define scalar products on complex vector spaces. In this setting, the symmetric bilinear form returns a complex number and it is called Hermitian bilinear form.

[3]In fact,

$$\mathrm{Tr}(AB^T) = \mathrm{Tr}(BA^T), \qquad \mathrm{Tr}((A + C)B^T) = \mathrm{Tr}(AB^T) + \mathrm{Tr}(CB^T) = \mathrm{Tr}(B(A + C)^T) \tag{11}$$

4. $V = C^0([0,1])$ (space of continuous functions defined on $[0,1]$). The integral

$$\langle u, v\rangle = \int_0^1 u(x)v(x)dx$$

is a symmetric bilinear form on $C^0([0,1])$ which defines a scalar product. In particular, if u and v are two polynomials of degree at most n defined on $[0,1]$ then $\langle u, v\rangle$ is a scalar product on $\mathbb{P}_n$.

A scalar product on $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ is said to be *non-degenerate* if

$$\langle v, w\rangle = 0 \qquad \text{for all } w \quad \Rightarrow \quad v = 0_V. \tag{12}$$

Theorem 1. Let V be a real vector space of dimension n. A scalar product

$$\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$$

is non-degenerate if and only if the matrix associated with $\langle\cdot,\cdot\rangle$ relative to any basis $\mathcal{B}_V = \{v_1, \dots, v_n\}$ is invertible.

Proof. We know that for all $u, v \in V$

$$\langle u, v\rangle = [u]^T_{\mathcal{B}_V} A_{\mathcal{B}_V} [v]_{\mathcal{B}_V}. \tag{13}$$

If the inner product is non-degenerate then, by definition

$$[u]^T_{\mathcal{B}_V} A_{\mathcal{B}_V} [v]_{\mathcal{B}_V} = 0 \qquad \text{for all} \quad [u]_{\mathcal{B}_V} \quad \Rightarrow \quad [v]_{\mathcal{B}_V} = 0_{\mathbb{R}^n}. \tag{14}$$

This implies that the nullspace of $A_{\mathcal{B}_V}$ reduces to the singleton $\{0_{\mathbb{R}^n}\}$. In fact if there exists another nonzero vector $[w]$ in the nullspace of $A_{\mathcal{B}_V}$ then clearly $A_{\mathcal{B}_V}[w] = 0_{\mathbb{R}^n}$ and the implication in (14) is not true. Conversely, suppose that $A_{\mathcal{B}_V}$ is invertible. Consider the column vector $A_{\mathcal{B}_V}[v]_{\mathcal{B}_V}$ and take all "dot products" with the elements the canonical basis of $\mathbb{R}^n$. This yields the system

$$[e_k]^T A_{\mathcal{B}_V} [v]_{\mathcal{B}_V} = 0 \qquad k = 1, \dots, n, \tag{15}$$

which is a homogeneous linear system of equations in n unknowns $[v]_{\mathcal{B}_V}$ that can be written as

$$A_{\mathcal{B}_V} [v]_{\mathcal{B}_V} = 0_{\mathbb{R}^n}. \tag{16}$$

The solution to this system is clearly $[v]_{\mathcal{B}_V} = 0_{\mathbb{R}^n}$, since $A_{\mathcal{B}_V}$ is invertible by assumption.

□

Examples of non-degenerate scalar products: The scalar products 1., 2. (with g_{ij} invertible) and 3. at page 2, and 4. at page 3 are all non-degenerate scalar products. Let us show that 3. is indeed a non-degenerate scalar product on $M_{n\times n}(\mathbb{R})$. We need to show that

$$\mathrm{Tr}(AB^T) = 0 \quad \text{for all} \quad B \in M_{n\times n}(\mathbb{R}) \quad \text{implies} \quad A = 0_{M_{n\times n}}. \tag{17}$$

The trace of the matrix product can be expressed as

$$\mathrm{Tr}(AB^T) = \sum_{i,j=1}^{n} A_{ij} B_{ij}. \tag{18}$$

Since B is arbitrary it easily follows form $\mathrm{Tr}(AB^T) = 0$ that $A = 0_{M_{n\times n}}$. In fact, evaluate the equation above using B equal to each element of the canonical basis of $M_{n\times n}(\mathbb{R})$. This yields $A_{ij} = 0$.

Examples of degenerate scalar products: Hereafter we provide a few examples of degenerate scalar products.

1. Let $V = \mathbb{R}^2$. Consider two vectors

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{19}$$

the scalar product

$$\langle x, y \rangle = x_1 y_1 \tag{20}$$

is degenerate. In fact, the condition

$$\langle x, y \rangle = 0 \quad \text{for all} \quad y \in \mathbb{R}^2 \quad \text{does not imply} \quad x = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{21}$$

To see this simply consider the vector $x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Alternatively, we observe that the scalar product (20) can be written as

$$\langle x, y \rangle = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}_{A_{\mathcal{B}_V}} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{22}$$

Recalling Theorem 2, we see that the matrix $A_{\mathcal{B}_V}$ associated with the scalar product relative to the canonical basis of $\mathbb{R}^2$ is not invertible and therefore the scalar product is degenerate.

Positive definite scalar products. Let V be a real vector space. A scalar product on V is said to be *positive definite* if

$$\langle v, v \rangle > 0 \quad \text{for all nonzero} \quad v \in V. \tag{23}$$

Clearly, if $v = 0_V$ then $\langle v, v \rangle = 0$.

Examples of positive definite scalar products: The scalar products 1., 3. and 4. defined at page 2 are all non-degenerate and positive definite. The scalar product 2. at page 2 is non-degenerate and positive definite if and only if the matrix g_{ij} is positive definite, i.e., if

$$\sum_{i,j=1}^{n} g_{ij} x_i x_j > 0 \quad \text{for all nozero vectors} \quad x \in \mathbb{R}^n. \tag{24}$$

Positive definite matrices are necessarily invertible. In fact, from condition (24) it follows that for any nonzero vector x, gx is nonzero. Therefore g is full rank ($\Rightarrow$ nullspace reduces to the $0_{\mathbb{R}^n}$) and therefore invertible.

Norms. A norm on a vector space V is is a function $\|\cdot\| : V \to \mathbb{R}$ with the following properties

1. $\|av\| = |a| \, \|v\|$ for all $v \in V$ and for all $a \in \mathbb{R}$ (or $\mathbb{C}$).
2. $\|u + v\| \leq \|u\| + \|v\|$
3. $\|u\| = 0 \quad \Leftrightarrow \quad u = 0_V$
4. $\|u\| > 0 \quad$ for all nonzero $u \in V$.

The norm defines the *length* of vectors in a vector space. We have already seen a norm when we studied complex numbers, i.e., the modulus of a complex number. Let us provide a few examples of norms.

Examples:

- Let $V = \mathbb{R}^n$. For every $v \in \mathbb{R}^n$ we define

$$\|v\|_1 = \sum_{i=1}^{n} |v_i| \qquad \text{(1-norm)} \tag{25}$$

$$\|v\|_2 = \left(\sum_{i=1}^{n} |v_i|^2 \right)^{1/2} \qquad \text{(2-norm)} \tag{26}$$

$$\|v\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p} \qquad (p\text{-norm, } p \geq 1 \text{ real number}) \tag{27}$$

$$\|v\|_\infty = \max_{i=1,\ldots,n} |v_i| \qquad \text{(infinity norm)} \tag{28}$$

  All these norms satisfy properties 1.-4. above. Moreover, it can be shown that

$$\|v\|_\infty = \lim_{p \to \infty} \|v\|_p \,. \tag{29}$$

- Let $V = C^0([0, 1])$ (space of continuous functions in $[0, 1]$). We define

$$\|u\|_\infty = \max_{x \in [0,1]} |u(x)| \qquad \text{(uniform norm)}, \tag{30}$$

$$\|u\|_2 = \int_0^1 |u(x)|^2 \, dx \qquad (L^2\,([0, 1])\text{-norm}). \tag{31}$$

- Let $V = M_{n \times n}(\mathbb{R})$ (space of $n \times n$ matrices with real coefficients). Let us define the following matrix norm

$$\|A\| = \max_{v \neq 0_{\mathbb{R}^n}} \frac{\|Av\|_p}{\|v\|_p} = \max_{\|v\|_p = 1} \|Av\|_p \qquad p \geq 1. \tag{32}$$

It is straightforward to show that

$$\|A\|_{\infty} = \max_{i=1,..,n} \left(\sum_{j=1}^{n} |A_{ij}| \right), \tag{33}$$

$$\|A\|_{1} = \max_{j=1,..,n} \left(\sum_{i=1}^{n} |A_{ij}| \right). \tag{34}$$

For example,

$$\|Av\|_{\infty} = \max_{i=1,...,n} \left| \sum_{j=1}^{n} A_{ij} v_j \right| \leq \max_{i=1,...,n} \left(\sum_{j=1}^{n} |A_{ij}| \, |v_j| \right) \leq \|v\|_{\infty} \max_{i=1,...,n} \left(\sum_{j=1}^{n} |A_{ij}| \right) \tag{35}$$

which implies that

$$\frac{\|Av\|_{\infty}}{\|v\|_{\infty}} \leq \max_{i=1,...,n} \left(\sum_{j=1}^{n} |A_{ij}| \right) \qquad \text{for all } v \neq 0_{\mathbb{R}^n}, \tag{36}$$

i.e.,

$$\max_{v \neq 0_{\mathbb{R}^n}} \frac{\|Av\|_{\infty}}{\|v\|_{\infty}} = \max_{i=1,...,n} \left(\sum_{j=1}^{n} |A_{ij}| \right) = \|A\|_{\infty}. \tag{37}$$

The matrix norms (33) and (34) are said to be *compatible* with associated vector norms (or *induced* by the vector norms) since they verify the inequalities

$$\|Av\|_{p} \leq \|A\|_{p} \|v\|_{p} \qquad p = 1, \infty. \tag{38}$$

Norms induced by scalar products. Any non-degenerate positive-definite scalar product on V induces a norm[4]

$$\|v\| = \sqrt{\langle v, v \rangle}. \tag{39}$$

In particular, the standard dot product in $\mathbb{R}^n$

$$\langle v, v \rangle = \sum_{i=1}^{n} v_i^2$$

induces the 2-norm defined in (26). Similarly, the scalar product between two matrices $A, B \in M_{n\times n}(\mathbb{R})$

$$\langle A, B \rangle = \mathrm{Tr}\left(AB^T\right)$$

induces the following norm in the space of matrices $M_{n\times n}(\mathbb{R})$

$$\|A\|_{F} = \sqrt{\sum_{i,j=1}^{n} A_{ij}^2} \qquad \text{(Frobenious norm)}. \tag{40}$$

The Frobenious norm is compatible with the vector norm in the sense that $\|Ax\|_2 \leq \|A\|_F \|x\|_2$.

[4]It is straightforward to show that properties 1.-4. at page 5 are all satisfied by the norm (39).

Theorem 2 (Cauchy-Schwarz inequality)**.** Let V be a real vector space. Then for all $u, v \in V$ we have

$$|\langle u, v\rangle| \leq \|u\| \, \|v\| \,, \tag{41}$$

where $\|u\| = \sqrt{\langle u, u\rangle}$ and $\|v\| = \sqrt{\langle v, v\rangle}$.

Proof. For $v = 0_V$ the inequality reduces to $0 = 0$. Let $u, v \in V$ be nonzero. The vector

$$u - \frac{\langle u, v\rangle}{\langle v, v\rangle} v \tag{42}$$

is orthogonal[5] to v since

$$\left\langle v, u - \frac{\langle u, v\rangle}{\langle v, v\rangle} v \right\rangle = \langle v, u\rangle - \frac{\langle u, v\rangle}{\langle v, v\rangle} \langle v, v\rangle = 0. \tag{43}$$

Next, consider the identity

$$u = u - \frac{\langle u, v\rangle}{\langle v, v\rangle} v + \frac{\langle u, v\rangle}{\langle v, v\rangle} v. \tag{44}$$

We have

$$\begin{aligned} \|u\|^2 &= \langle u, u\rangle \\ &= \left\langle u - \frac{\langle u, v\rangle}{\langle v, v\rangle} v, u - \frac{\langle u, v\rangle}{\langle v, v\rangle} v \right\rangle + \frac{|\langle u, v\rangle|^2}{|\langle v, v\rangle|^2} \langle v, v\rangle \\ &= \left\| u - \frac{\langle u, v\rangle}{\langle v, v\rangle} v \right\|^2 + \frac{|\langle u, v\rangle|^2}{\|v\|^2} \\ &\geq \frac{|\langle u, v\rangle|^2}{\|v\|^2}. \end{aligned} \tag{45}$$

From the last inequality we have $|\langle u, v\rangle|^2 \leq \|u\|^2 \, \|v\|^2$. Taking the square root yields equation (41).

□

Cosine similarity. The Cauchy-Schwartz inequality (41) implies that

$$-1 \leq \frac{\langle u, v\rangle}{\|u\| \, \|v\|} \leq 1. \tag{46}$$

The quantity

$$\cos(\vartheta) = \frac{\langle u, v\rangle}{\|u\| \, \|v\|} \tag{47}$$

is known as *cosine similarity* between the vectors u and v. In the case where u and v are vectors of $\mathbb{R}^2$, $\mathbb{R}^3$ or $\mathbb{R}^n$, the cosine similarity coincides with cosine of the angle between the two vectors. Such an angle is measures on the two-dimensional plane spanned by the two vectors. From (47) it follows that

$$\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2 \, \|u\| \, \|v\| \cos(\vartheta). \tag{48}$$

which is the well-known law of cosines for triangles. The cosine similarity is practically utilized in many different fields, e.g., natural language processing (similarity between texts) and econometrics (analysis of time series).

[5] We say that to vectors $u, w \in V$ are orthogonal with respect to the scalar product $\langle \cdot, \cdot \rangle$ if $\langle u, w\rangle = 0$.

Orthogonality. Consider a vector space V and a non-degenerate positive definite scalar product $\langle \cdot, \cdot \rangle$ on V. Two vectors $u, v \in V$ are said to be *orthogonal* relative to $\langle \cdot, \cdot \rangle$ if

$$\langle u, v \rangle = 0. \tag{49}$$

Examples:

- $V = \mathbb{R}^2$. The following vectors

$$u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad v = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \tag{50}$$

are orthogonal in $\mathbb{R}^2$ relative to the standard inner product

$$\langle u, v \rangle = \sum_{i=1}^{2} u_i v_i = -1 + 1 = 0. \tag{51}$$

- $V = M_{2\times 2}(\mathbb{R})$. The following matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{52}$$

are orthogonal in $M_{2\times 2}(\mathbb{R})$ relative to the inner product

$$\langle A, B \rangle = \mathrm{Tr}\left(AB^T\right). \tag{53}$$

In fact,

$$\mathrm{Tr}\left(AB^T\right) = \mathrm{Tr}\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right) = \mathrm{Tr}\left(\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right) = 0. \tag{54}$$

- $V = \mathbb{P}_2([-1, 1])$ (vector space of polynomials of degree at most two). The polynomials

$$p_1(x) = x \qquad \text{and} \qquad p_2(x) = \frac{3}{2}x^2 - \frac{1}{2} \tag{55}$$

are orthogonal with respect to the scalar product

$$\langle p_1, p_2 \rangle = \int_{-1}^{1} p_1(x) p_2(x) dx. \tag{56}$$

In fact,

$$\langle p_1, p_2 \rangle = \int_{-1}^{1} p_1(x) p_2(x) dx = \int_{-1}^{1} \left(\frac{3}{2}x^3 - \frac{1}{2}x\right) dx = \left[\frac{3}{8}x^4 - \frac{1}{4}x^2\right]_{-1}^{1} = 0. \tag{57}$$

Orthogonal projections. Consider two vectors u and v in a vector space V. The orthogonal projection of u onto v is defined as

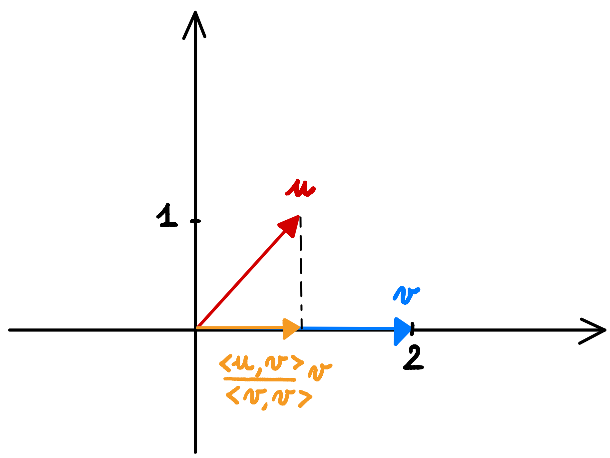
$$P_v u = \frac{\langle u, v \rangle}{\langle v, v \rangle} v = \left\langle u, \frac{v}{\|v\|} \right\rangle \frac{v}{\|v\|}, \tag{58}$$

where $\|\cdot\|$ is the norm induced by the scalar product. Clearly $v/\|v\|$ is a vector with norm equal to one, i.e., a *unit vector*.

Examples:

- Let $V = \mathbb{R}^2$ and consider the following vectors

$$u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad v = \begin{bmatrix} 2 \\ 0 \end{bmatrix}. \tag{59}$$



The projection of u onto v is

$$P_v u = \frac{\langle u, v\rangle}{\langle v, v\rangle} v = \frac{2}{4}\begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{60}$$

Note that if we subtract $\dfrac{\langle u, v\rangle}{\langle v, v\rangle} v$ from u we obtain a vector that is orthogonal to v.

$$u - \frac{\langle u, v\rangle}{\langle v, v\rangle} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{61}$$

- $V = \mathbb{R}^3$. Given three vectors $v_1, v_2, v_3 \in \mathbb{R}^3$ we can compute the orthogonal projection of any vector onto any other vector, e.g., the orthogonal projection of v_2 onto v_1

$$P_{v_1} v_2 = \frac{\langle v_2, v_1\rangle}{\langle v_1, v_1\rangle} v_1. \tag{62}$$

We can also construct an orthogonal set of vector by transforming the given set of linearly independent vectors $\{v_1, v_2, v_3\}$ as follows

$$\begin{aligned} u_1 &= v_1, \\ u_2 &= v_2 - \frac{\langle v_2, u_1\rangle}{\langle u_1, u_1\rangle} u_1, \\ u_3 &= v_3 - \frac{\langle v_3, u_1\rangle}{\langle u_1, u_1\rangle} u_1 - \frac{\langle v_3, u_2\rangle}{\langle u_2, u_2\rangle} u_2. \end{aligned}$$

This procedure is known as *Gram-Schmidt orthogonalization*, and it allows us to transform any set of linearly independent vectors into an orthogonal one. Such set of ortogonal vectors can be then normalized.

Gram-Schmidt orthogonalization. The previous example suggests that we can transform any basis $\{v_1, \ldots, v_n\}$ of a n-dimensional vector space V into an orthonormal basis[6] by using the Gram-Schmidt procedure. In fact, we can first compute

$$
\begin{aligned}
u_1 =& v_1, \\
u_2 =& v_2 - \frac{\langle v_2, u_1\rangle}{\langle u_1, u_1\rangle} u_1, \\
u_3 =& v_3 - \frac{\langle v_3, u_1\rangle}{\langle u_1, u_1\rangle} u_1 - \frac{\langle v_3, u_2\rangle}{\langle u_2, u_2\rangle} u_2, \\
&\vdots \\
u_n =& v_n - \frac{\langle v_n, u_1\rangle}{\langle u_1, u_1\rangle} u_1 - \cdots \frac{\langle v_n, u_{n-1}\rangle}{\langle u_{n-1}, u_{n-1}\rangle} u_{n-1}.
\end{aligned}
$$

and then normalize the vectors $\{u_1, \ldots, u_n\}$ to obtain the orthonormal basis

$$
\left\{ \frac{u_1}{\|u_1\|}, \ldots, \frac{u_n}{\|u_n\|} \right\}. \tag{63}
$$

Alternatively, we can normalize each vector u_i right after we compute it. This reduces the number of calculations in the Gram-Schmidt procedure as we can write

$$
\begin{aligned}
u_1 &= v_1, & \widehat{u}_1 &= u_1 / \|u_1\|, \\
u_2 &= v_2 - \langle v_2, \widehat{u}_1\rangle \widehat{u}_1 & \widehat{u}_2 &= u_2 / \|u_2\|, \\
u_3 &= v_3 - \langle v_3, \widehat{u}_1\rangle \widehat{u}_1 - \langle v_3, \widehat{u}_2\rangle \widehat{u}_2 & \widehat{u}_3 &= u_3 / \|u_3\|, \\
&\cdots
\end{aligned}
$$

It is straightforward to show that

$$
\langle u_i, u_j\rangle = \delta_{ij} \|u_j\|^2, \tag{64}
$$

where δ_{ij} is the Kronecker delta function[7]. For example,

$$
\langle u_1, u_2\rangle = \left\langle v_1, v_2 - \frac{\langle v_2, v_1\rangle}{\langle v_1, v_1\rangle} v_1 \right\rangle = \langle v_1, v_2\rangle - \frac{\langle v_2, v_1\rangle}{\langle v_1, v_1\rangle} \langle v_1, v_1\rangle = 0. \tag{66}
$$

Example: Let us use the Gram-Schmidt procedure to orthogonalize the following vectors in $\mathbb{R}^2$

$$
v_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \qquad v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.
$$

[6]Note that the orthogonal basis we obtain from the Gram-Schmidt procedure is not unique. In fact a reordering of the vectors $\{v_1, \ldots, v_n\}$ yields a different orthogonal basis at the end of the procedure.

[7]The Kronecher delta is defined as:

$$
\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{65}
$$

We have

$$u_1 = v_1, \qquad u_2 = v_2 - \frac{\langle v_2, u_1 \rangle}{||u_1||^2} u_1.$$

The norm of $u_1 = v_1$ is

$$\|u_1\|^2 = \|v_1\|^2 = \langle v_1, v_1 \rangle = 2^2 + 1^2 = 5. \tag{67}$$

This implies that

$$u_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \frac{4}{5} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 - 8/5 \\ 2 - 4/5 \end{bmatrix} = \begin{bmatrix} -3/5 \\ 6/5 \end{bmatrix}$$

Note that u_1 and u_2 are orthogonal. In fact,

$$\langle u_1, u_2 \rangle = 2 \times \left(-\frac{3}{5} \right) + \frac{6}{5} = 0. \tag{68}$$

The norm of u_2 is

$$\|u_2\| = \sqrt{\langle u_2, u_2 \rangle} = \sqrt{\frac{9}{25} + \frac{36}{25}} = \frac{3\sqrt{5}}{5}. \tag{69}$$

This means that

$$\left\{ \frac{u_1}{\|u_1\|}, \frac{u_2}{\|u_2\|} \right\} = \left\{ \frac{\sqrt{5}}{5} \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \frac{\sqrt{5}}{5} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\} \tag{70}$$

is an orthonormal basis of $\mathbb{R}^2$. Note that $u_2/\|u_2\|$ can be obtained by rotating $u_1/\|u_1\|$ by 90 degrees counterclockwise.

Representation of vectors relative to orthonormal bases. Let $\mathcal{B}_V = \{\widehat{u}_1, \ldots, \widehat{u}_n\}$ be an orthonormal basis of a n-dimensional vector space V. Any vector $v \in V$ can be represented relative to the basis $\mathcal{B}_V$ as

$$v = x_1 \widehat{u}_1 + \cdots + x_n \widehat{u}_n. \tag{71}$$

by projecting the vector v onto $\widehat{u}_i$ and taking into account the orthonormality conditions $\langle u_i, u_j \rangle = \delta_{ij}$ yields

$$\begin{aligned} \langle v, \widehat{u}_j \rangle &= \langle x_1 \widehat{u}_1 + \cdots + x_n \widehat{u}_n, \widehat{u}_j \rangle \\ &= x_1 \langle \widehat{u}_1, \widehat{u}_j \rangle + \cdots + x_n \langle \widehat{u}_n, \widehat{u}_j \rangle \\ &= x_j \langle \widehat{u}_j, \widehat{u}_j \rangle \\ &= x_j \end{aligned} \tag{72}$$

i.e., the j-the coordinate of v relative to $\mathcal{B}_V$ coincides with the projection of v onto $\widehat{u}_j$. On the other hand, if we consider an orthogonal basis $\{u_1, \ldots, u_n\}$ we obtain

$$v = y_1 u_1 + \cdots + y_n u_n \quad \Rightarrow \quad x_j = \frac{\langle v, u_j \rangle}{\langle u_j, u_j \rangle}. \tag{73}$$

Theorem 3. Let $\mathcal{B}_V = \{\widehat{u}_1, \ldots, \widehat{u}_n\}$ be an orthonormal basis of a n-dimensional vector space V. Then for any vector $v \in V$ we have

$$v = x_1 \widehat{u}_1 + \cdots + x_n \widehat{u}_n \quad \text{and} \quad \|v\|^2 = \sum_{k=1}^{n} x_k^2. \tag{74}$$

Orthogonal complement. Let S be a subspace of V. The orthogonal complement of S in V is defined as

$$S^{\perp} = \{v \in V : \langle v, w\rangle = 0 \quad \text{for all} \quad w \in S\}$$

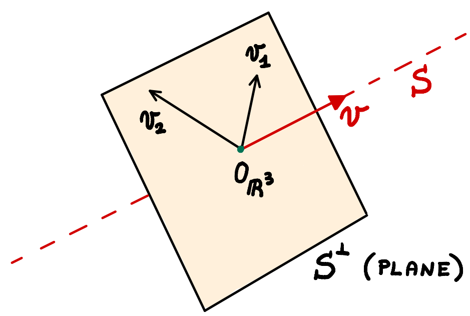
It can be shown that $S^{\perp}$ is a vector subspace of V. Any vector $v \in V$ can be expressed as a sum of two vectors $w_1 \in S$ and $w_2 \in S^{\perp}$, i.e.,

$$v = w_1 + w_2 \tag{75}$$

Equivalently, we say that V is the direct sum of S and $S^{\perp}$, and write

$$V = S \oplus S^{\perp}. \tag{76}$$

For example, any vector $v \in V = \mathbb{R}^3$ defines a one-dimensional vector subspace S. The orthogonal complement of S in $\mathbb{R}^3$ is a plane orthogonal S. Such plane is denoted by $S^{\perp}$



The plane, i.e., the vector space $S^{\perp}$, is identified mathematically by the condition

$$\langle x, v\rangle = 0 \qquad (v \text{ is given, } x \in \mathbb{R}^3 \text{ is arbitrary}) \tag{77}$$

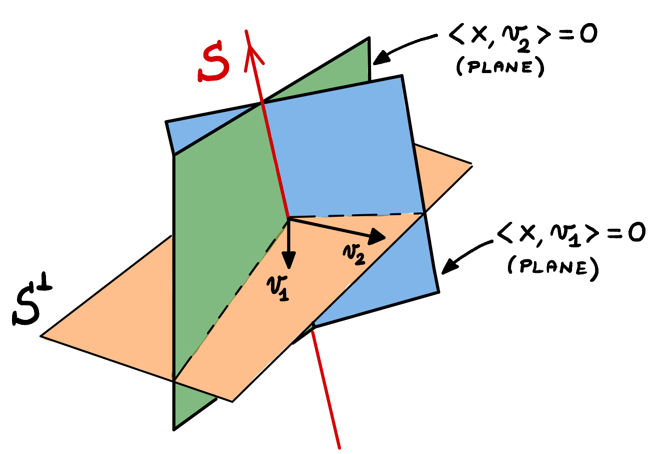
i.e.,

$$v_1x_1 + v_2x_2 + v_3x_3 = 0 \tag{78}$$

We know this expression very well, but now we learned something new, i.e., that the coefficients of v_1, v_2 and v_3 are the components of a vector that is orthogonal to the plane. Similarly, given two linearly independent vectors vectors v_1 and v_2 in $\mathbb{R}^3$, it is possible to determine the space that is orthogonal to the span of v_1 and v_2 by solving the system of equations

$$\langle x, v_1\rangle = 0, \qquad \langle x, v_2\rangle = 0 \qquad x \in \mathbb{R}^3. \tag{79}$$

This system represents the intersection of two planes orthogonal to v_1 and v_2.



Orthogonal complement of the range and the nullspace of a matrix. Next consider a $m \times n$ matrix A and let $\{v_1, \ldots, v_n\}$ be the columns of A. Denote by

$$R(A) = \text{span}\{v_1, \ldots, v_n\} \tag{80}$$

the column space of A, i.e., the range of the matrix A. We known that such space is a vector subspace of $\mathbb{R}^m$. The orthogonal complement of $R(A)$ is

$$[R(A)]^\perp = \{v \in \mathbb{R}^m : \langle v, w \rangle = 0 \quad \text{for all} \quad w \in R(A)\}. \tag{81}$$

Let us write the condition $\langle v, w \rangle = 0$ a more explicitly.

To this end, we notice that $R(A)$ can be characterized as the set of vectors $w \in \mathbb{R}^m$ such that such that $w = Ax$. Hence,

$$\begin{aligned} v \in [R(A)]^\perp \quad &\Leftrightarrow \quad \langle v, Ax \rangle = 0 \quad \text{for all } x \in \mathbb{R}^n \\ &\Leftrightarrow \quad \left\langle A^T v, x \right\rangle = 0 \quad \text{for all } x \in \mathbb{R}^n \\ &\Leftrightarrow \quad A^T v = 0_{\mathbb{R}^n} \\ &\Leftrightarrow \quad v \in N(A^T). \end{aligned}$$

This means that

$$[R(A)]^\perp = N(A^T). \tag{82}$$

In other words, the orthogonal complement of the column space of a matrix coincides with the nullspace of the matrix transpose. We can also prove the equality the other way around, i.e.,

$$\begin{aligned} v \in N(A^T) \quad &\Leftrightarrow \quad A^T v = 0 \qquad v \in \mathbb{R}^m \\ &\Leftrightarrow \quad \left\langle w, A^T v \right\rangle = 0 \quad \text{for all } w \in \mathbb{R}^n \\ &\Leftrightarrow \quad \langle Aw, v \rangle = 0 \quad \text{for all } w \in \mathbb{R}^n \\ &\Leftrightarrow \quad v \in [R(A)]^\perp. \end{aligned}$$

Repeating this simple proof for $N(A)$ yields

$$\left[R(A^T)\right]^\perp = N(A). \tag{83}$$

i.e., the orthogonal complement of the *row space* of A (i.e., the column space of A^T) coincides with the nullspace of A. Similarly, it can be shown that

$$[N(A)]^\perp = R(A^T) \qquad \text{and} \qquad \left[N(A^T)\right]^\perp = R(A). \tag{84}$$