Linear Algebra (not only) for computer scientists

Translation Luděk Kučera

Milan Hladík

25. května 2017

Chapter 8 Scalar product

Vector spaces have been defined in a very general way, so that they cover a large class of objects. On the other hand, if we add other requirements vector spaces have fulfil, it will be possible to derive deeper results. In particular, the scalar product (sometimes called the dot product) makes it possible to define in a natural way the notions of orthogonality, the size and the distance of vectors (and hence limits as well) etc.

8.1. Scalar product and norm

The scalar product (similarly as a group, vector space, etc.,) is defined by a list of properties that it has to fulfill.

Definition 8.1 (Scalar product over $\mathbb{R}$). Let V be a vector space over $\mathbb{R}$. Then a *scalar product* is a mapping $\langle\cdot,\cdot\rangle : V^2 \to \mathbb{R}$ such that
(1) $\langle x,x\rangle \geq 0 \quad \forall x \in V$, and the equality holds only for $x = 0$,
(2) $\langle x+y,z\rangle = \langle x,z\rangle + \langle y.z\rangle \quad \forall x,y,z \in V$,
(3) $\langle \alpha x,y\rangle = \alpha\langle x,z\rangle \quad \forall x,y \in V, \forall \alpha \in \mathbb{R}$,
(4) $\langle x,y\rangle = \langle y,x\rangle \quad \forall x,y \in V$.

A generalization to the field of complex numbers is the following. Let us mention that a *complex conjugate* to $a+bi \in \mathbb{C}$ is defined as $\overline{a+bi} = a - bi$.

Definition 8.2 (Scalar product over $\mathbb{C}$). Let V be a vector space over $\mathbb{C}$. Then a *scalar product* is a mapping $\langle\cdot,\cdot\rangle : V^2 \to \mathbb{C}$ such that
(1) $\langle x,x\rangle \geq 0 \quad \forall x \in V$, and the equality holds only for $x = 0$,
(2) $\langle x+y,z\rangle = \langle x,z\rangle + \langle y.z\rangle \quad \forall x,y,z \in V$,
(3) $\langle \alpha x,y\rangle = \alpha\langle x,z\rangle \quad \forall x,y \in V, \forall \alpha \in \mathbb{C}$,
(4) $\langle x,y\rangle = \overline{\langle y,x\rangle} \quad \forall x,y \in V$.

Because of the property (1), that requires ordering, the scalar product is introduced only for vector spaces over the fields $\mathbb{R}$ and $\mathbb{C}$.

The fourth property of the complex scalar product implies that $\langle x,x\rangle = \overline{\langle x,x\rangle} \in \mathbb{R}$, and hence $\langle x,x\rangle$ is always a real number and therefore we can compare it with zero in the first property.

The properties (2) and (3) say that the scalar product is a linear function of its first coordinate. How it is with the second one?

$$\langle x, y+z\rangle = \overline{y+z, x} = \overline{\langle y, x\rangle} + \overline{\langle z, x\rangle} = \langle x, y\rangle + \langle x, z\rangle,$$

$$\langle x, \alpha y\rangle = \overline{\langle \alpha y, x\rangle} = \overline{\alpha\langle y, x\rangle} = \overline{\alpha}\overline{\langle y, x\rangle} = \overline{\alpha}\langle x, y\rangle.$$

It follows that the complex scalar product is not linear in the second coordinate, while the real scalar product is.

When substituting $\alpha = 0$, we get $\langle o, x\rangle = \langle x, o\rangle = 0$, which means that the scalar product of any vector with the zero one gives zero.

Example 8.3 (Examples of scalar products).

- In the space $\mathbb{R}^n$: the standard scalar product $\langle x, y\rangle = x^T y = \sum_{i=0}^{n} x_i y_i$.
- In the space $\mathbb{C}^n$: the standard scalar product $\langle x, y\rangle = x^T \overline{y} = \sum_{i=0}^{n} x_i \overline{y_i}$.
- In the space $\mathbb{R}^{m\times n}$: the standard scalar product $\langle A, B\rangle = x^T y = \sum_{i=0}^{m}\sum_{j=0}^{n} a_{ij} b_{ij}$.
- In $\mathcal{C}_{[a,b]}$, the space of continuous functions over $[a, b]$: the standard scalar product $\langle f, g\rangle = \int_a^b f(x)g(x)dx$.

The above mentioned scalar products are only examples of possible scalar products on the spaces, there are many other scalar products. Later, in Theorem 11.17, we will describe all scalar products on the space $\mathbb{R}^n$.

Let us consider a vector space V over $\mathbb{R}$ or $\mathbb{C}$. We will first show that a scalar product enables us to introduce the norm, or the length of a vector.

Definition 8.4 (A norm induced by a scalar product). *The norm defined by a scalar product* is defined by $\|x\| := \sqrt{\langle x, x\rangle}$, where $x \in V$.

The norm is well-defined due to the first property from the definition of the scalar product, and it is always non-negative.

When using the standard scalar product in $\mathbb{R}^n$, we get well-known Euclidean norm $\|x\| = \sqrt{x^T x} = \sqrt{\sum_{i=0}^{n} x_i^2}$.

A geometric interpretation of the standard scalar product in $\mathbb{R}^n$ is $\langle x, y\rangle = \|x\| \cdot \|y\| \cos(\varphi)$, where φ is the angle between vectors x and y. In particular, the vectors x and y are orthogonal if and only if $\langle x, y\rangle = 0$. In pther spaces such geometry is missing, and therefore the orthogonality is defined by the equation $\langle x, y\rangle =)$.

Definition 8.5 (Orthogonality). Given vectors x and y are *orthogonal* if $\langle x, y\rangle = 0$. We denote $x \perp y$.

Example 8.6 (Examples of orthogonal vectors for standard scalar products).

- In the space $\mathbb{R}^3$: $(1, 2, 3) \perp (1, 1, -1)$.
- In the space $\mathcal{C}_{[-\pi,\pi]}$: $\sin x \perp \cos x \perp 1$.

Theorem 8.7 (Pythagoras).
If $x, y \in V$ are orthogonal, then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.
Proof.

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle = \langle x,x\rangle + \langle y,y,\rangle = \|x\|^2 + \|y\|^2. \quad ♣$$

Let us mention that in $\mathbb{R}$ the reverse implication is valid as well, but, in general, this is not true in $\mathbb{C}$ (see Problem 8.2).

Theorem 8.8 (Inequality of Cauchy-Schwartz[1]). For each $x, y \in V$ it is $|\langle x, y\rangle \le \|x\| \cdot \|y\|$.
Proof. (Real version) We will prove the inequaliy over real numbers first, because the proof is quite elegant. For $y = o$ the inequaliy is satisfied trivially, so let us suppose $y \ne o$. Consider a real function $f(t) = \langle x+ty, x+ty\rangle \ge 0$ of the variable t. Then

$$f(t) = \langle x,x\rangle + t\langle x,y\rangle + t\langle y,x\rangle + t^2\langle y,y\rangle = \langle x,x\rangle + 2t\langle x,y\rangle + t^2\langle y,y\rangle.$$

The function is quadratic, and everywhere non-negative, and therefore it does not have two different roots. Therefore the discriminant is non-positive:

$$4\langle x,y\rangle - 4\langle x,x\rangle\langle y,y\rangle \le 0.$$

It follows that $\langle x,y\rangle^2 \le \langle x,x\rangle\langle y,y\rangle$, and taking a square root we get $|\langle x,y\rangle| \le \|x\| \cdot \|y\|$.
Proof. (Complex version) If $\langle x,y\rangle = 0$, the statement is satisfied trivially. Let us assume that $\langle x,y\rangle \ne 0$ and define a vector

$$y = \frac{\langle y,y\rangle}{\langle x,y\rangle} x - y.$$

Then

$$\langle z,y\rangle = \left\langle \frac{\langle y,y\rangle}{\langle x,y\rangle} x - y, y \right\rangle = \frac{\langle y,y\rangle}{\langle x,y\rangle}\langle x,y\rangle - \langle y,y\rangle = 0.$$

It follows that the vectors z and y are orthogonal, and Pythagoras theorem gives

$$\|x+y\|^2 = \|z\|^2 + \|y\|^2,$$

or

$$\frac{\langle y,y\rangle^2}{|\langle x,y\rangle|^2}\|x\|^2 = \|z\|^2 + \|y\|^2 \ge \|y\|^2.$$

By dividing by $\|y\|^2$ and multiplying by $|\langle x,y\rangle|^2$ we get the inequality to be proved $\|x\|^2\|y\|^2 \ge |\langle x,y\rangle^2$ in the second powers. ♣

[1] The inewuality is sometimes called Scwartz, or Cauchy-Bunyakovski, or Cauchy-Schwartz-Bunyakovski. Augustin-louis Cauchy proved it in 1821 for $\mathbb{R}^n$, later it was generalized independently by Hermann Amandus Schwartz (1880) and Viktor Jakovlevich Bunyakovski (1859).

Sometimes the Cauchy-Schwartz inequality is written equivalently as

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle.$$

Cauchy-Schwartz inequality is useful and often used for deriving other general results, and also for particular algebraic expressions. E.g., for the standard scalar product in $\mathbb{R}^n$ we get

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right).$$

Other use, see e.g. Kristl 2008. We will use Cauchy-Schwartz inequality immediately to get the triangle inequality:

Theorem 8.9 (Triangle inequality). For each $x, y \in V$ it is $\|x+y\| \leq \|x\|+\|y\|$.
Proof. Let us first recall that the following holds for any complex number $z = a + bi$¿ $z + \overline{z} = 2a = 2\mathrm{Re}(z)$. Moreover, $a \leq |z|$. Now, we get

$$\begin{aligned} \|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + \langle y, y\rangle + \langle x, y\rangle + \langle y, x\rangle = \\ = \langle x, x\rangle + \langle y, y\rangle + 2\mathrm{Re}(\langle x, y\rangle) \leq \langle x, x\rangle + \langle y, y\rangle + 2|\langle x, y\rangle| \\ \leq \|x\|^2 + \|y\|^2 + 2\|x\| \cdot \|y\| = (\|x\| + \|y\|)^2, \end{aligned}$$

where the last inequality follows from Cauchy-Schwartz inequality. We obtain $\|x+y\|^2 \leq (\|x\|+\|y\|)^2$ and the statement to be proved is obtained by square-rooting.♣

A notm induced by a scalar product is only one of types of a norm, but the notion of a norm is defined in a more general way. However, we will work mostly with a scalar product induced norm, the following definition is a small detour only.

Definition 8.10 (Norm) Let V be a vector space over $\mathbb{R}$ or $\mathbb{C}$. Then a *norm* is a mapping $\|\cdot\| > V \to \mathbb{R}$ such that:
(1) $\|x\| \geq 0$ for all $x \in V$, and the equality holds for $x = o$ only,
(2) $\|\alpha x\| = |\alpha| \cdot \|x\|$ for all $x \in V$ and all $\alpha \in \mathbb{R}$ or $\alpha \in \mathbb{C}$,
(3) $\|x+y\| \leq \|x\| \cdot \|y\|$.

Theorem 8.11 A norm induced by a scalar product is a norm.
Proof. The property (1) is fulfilled in view of the definition of a norm induced by a scalar product. The property (3) follows from 8.9. Now for (2):

$$\|\alpha x\| = \sqrt{\langle \alpha x, \alpha x\rangle} = \sqrt{\alpha\overline{\alpha}\langle x, x\rangle} = \sqrt{\alpha\overline{\alpha}}\sqrt{\langle x, x\rangle} = |\alpha| \cdot \|x\|.$$

Example 8.12 (Examples of norms in $\mathbb{R}^n$). A useful class of norms are so called p-norms. For $1 \leq p$ the p norm of a vector $x \in \mathbb{R}^n$ is defined as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}.$$

Special settings of p give known norms:

- for $p = 2$: the Euclidean norm $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$, which is the norm defined by the standard scalar product,
- for $p = 1$: the sum norm $\|x\|_1 = \sum_{i=1}^n |x_i|$, called also the Manhattan norm, because it represents real distances when walking in a perpendicular network of streets in a town,
- for $p = \infty$ (using limit expression): the maximum (Chebychev) norm $\|x\|_\infty = \max_{i=1,\dots,n} |x_i|$.

Example 8.13 (Examples of norms in $\mathcal{C}_{[a,b]}$). A norm of a continuoius function $f : [a, b] \to \mathbb{R}$ can be defined analogously to the Euclidean space:

- an analogy of the Euclidean norm: $\|f\|_2 = \sqrt{\int_a^b f(x)^2 dx}$,
- an analogy of the sum norm: $\|f\|_1 = \int_a^b |f(x)| dx$,
- an analogy of the maximum norm: $\|f\|_\infty = \max_{x\in[a,b]} |f(x)|$.
- an analogy of the p-norm: $\|f\|_p = \left(\int_a^b |f(x)|^p dx\right)^{\frac{1}{p}}$.

Remark 8.14 (Parallelogram rule). The following statement, so called *parallelogram rule*, is satisfied for a norm induced by a scalar product:

$$\|x - y\|^2 + \|x + y\|^2 = 2\|x\|^2 + \|y\|^2.$$

Proof.

$$\|x-y\|^2+\|x+y\|^2 = \langle x-y, x-y\rangle+\langle x+y, x+y\rangle = 2\langle x, x\rangle+2\langle y, y\rangle = 2\|x\|^2+\|y\|^2. \quad \clubsuit$$

It follows from the theorem that both the sum and the maximum norms are not induced by any scalar product. E.g., for $x = (1, 0)$ and $y = (0, 1)$ the parallelogram rule is not satisfied.

A more general theroem can be proved: if a norm verifies the parallelogram rule, it is induced by some scalar product; see Horn and Johnson (1985).

A notm allows us to define the distance (or metric) of two vectors x and y as a norm $\|x-y\|$/ And when we have distances, we can define limits, etc. I expect that a reader will not be surprised that a metric can be defined axiomatically. Moreover, to define a metric, we do not need a vector space, an arbitrary set is sufficient.

Remark 8.15 (Metric) A metric on a set M is defined as a mapping $d : M^2 \to \mathbb{R}$ verifying:
(1) $d(x, y) \geq 0$ for each $x, y \in M$, and the equality holds only if $x = y$,

(2) $d(x, y) = d(y, x)$ for each $x, y \in M$,
(3) $d(x, z) \leq d(x, y) + d(y, z)$ for each $x, y, z \in M$.

Each norm defines a metric by $d(x, y) = \|x - y\|$, i.e., the distance of x and y is defined as the length of their difference. The reverse statement is not valid in general. There are metric spaces, where the metric is not induces by any norm, e.g., the discrete metric $d(x, y) = \lceil \|x - y\| \rceil$, od the discrete metric defined by $d(x, y) = 1$ for $x \neq y$, and $d(x, y) = 0$ for $x = y$.

Example 8.16 Skipped

8.2. Orthonormal basis, Gram-Schmidt othogonalization

Every vector space has basis. When dealing with a space with scalar product, it is natural to ask whether there is a basis composed of mutually orthogonal vectors. In this section we will show that such a basis exists, and that such a basis has quite a few remarkable properties. We will also derive an algorithm to find such a basis.

Definition 8.17 (An orthogonal and orthonormal system). A system of vectors $z_1, \ldots, z_n$ is *orthogonal*, if $\langle z_i, z_j \rangle = 0$ for all $i \neq j$. A system of vectors $z_1, \ldots, z_n$ is *orthonorma*, if it is orthogonal, and $\|z_i\| = 1$ for all $i = 1, \ldots, n$.

If a system $z_1, \ldots, z_n$ is orthonormal, it is orthogonal. The reversed implication is not valid, but it is not a problem to orthogonalize such a sustem. If $z_1, \ldots, z_n$ are non-zero and orthogonal, then $\frac{1}{\|z_1\|} z_1, \ldots, \frac{1}{\|z_n\|} z_n$ is orthonormal. The proof: $\| \frac{1}{\|z_i\|} z_i \| = \frac{1}{\|z_i\|} \|z_i\| = 1$.

Example 8.18 In the space $\mathbb{R}^n$ with the standard scalar product an example of an orthonormal system is the canonical basis $e_1, \ldots, e_n$. Especially in the plane $\mathbb{R}^2$ an example of an orthonormal system are vectors $(1, 0)^T$ and $(0, 1)^T$. Another example of an orthonormal basis in $\mathbb{R}^2$ is, e.g., $\frac{\sqrt{2}}{2}(1, 1)^T$, $\frac{\sqrt{2}}{2}(1, -1)^T$.

Theorem 8.19 If a system $z_1, \ldots, z_n$ of vectors is orthonormal, it is linearly independent.
Proof. Let us consider a linear combination $\sum_{i-1}^{n} \alpha_i z_i = o$ The for each $k = 1, \ldots, n$ it is:

$$0 = \langle o, z_k \rangle = \left\langle \sum_{i=1}^{n} \alpha_i z_i, z_k \right\rangle = \sum_{i=1}^{n} \alpha_i \langle z_i, z_k \rangle = \alpha_k \langle z_k, z_k \rangle = \alpha_k. \qquad \clubsuit$$

Thus, orthonormality of vectors implies their linear independence plus something else - their orthogonality. And it is this property that allow us to solve certain problems efficiently. E.g., the next theorem tells us that it is easy to compute coordinates of a vector with respect to a basis, which is orthonormal.

Theorem 8.20 (Fourier coefficients). Let $z_1, \ldots, z_n$ be an orthonormal basis of a space V. Then for each $x \in V$ it is $x = \sum_{i=1}^{n} \langle x, z_i \rangle z_i$.
Proof. We know that $x = \sum_{i=1}^{n} \alpha_i z + i$ and the coordinates $\alpha_1, \ldots \alpha_n$ are unique (Theorem 5.28). Now, for each $k = 1, \ldots, n$ it is:

$$\langle x, z_k\rangle = \left\langle \sum_{i=1}^{n} \alpha_i z_i, z_k \right\rangle = \sum_{i=1}^{n} \alpha_i \langle z_i, z_k\rangle = \alpha_k \langle z_k, z_k\rangle = \alpha_k. \quad \clubsuit$$

The expression $x \in V$ it is $x = \sum_{i=1}^{n} \langle x, z_i\rangle z_i$ is called *Fourier series* and the scalars $\langle x, z_i$, $i = 1, \ldots, n$ are called *Fourier coefficients*[2]). Geometric significance of the Fourier coefficient $\langle x, z_i\rangle$ is that the vector $\langle x, z_i\rangle z_i$ is the projection of the vector x to the line $\mathrm{span}\{z_i\}$. In other words, the vector $\langle x, z_i\rangle z_i$ is a vector on the line with the direction z_i, which is the closest to the vector x. Then the vector x can be composed of those partial projections using a simple sum $x = \sum_{i=1}^{n} \langle x, z_i\rangle z_i$ (more about projections in Section 8.3). As illustrated below, if the basis $z_1, \ldots, z_n$ is not orthonormal, the property is not satisfied in general.

The figures at page 94 of the Czech text, z_1 and z_2 orthonormal on the left, but of length 1 and not orthonormal on the right.

How to construct an otrhonormal basis of some space? The following procedure, Gram-Schmidt orthogonalization, starts with an arbitrary basis, and using subsequent making vectors orthogonal it creates a basis, which is orthonormal. Making vectors orthogonal in the step 2 of the procedure works so that the projection of a vector x_k to the space generated by vectors $x_1, \ldots, x_{k-1}$ is subtracted from the vector x_k, which becomes orthogonal to all preceding vectors. More about projection in Section 8.3.

The figures at page 95 of the Czech text: making the second and the third vector orthogonal to the previous ones.

Theorem 8.21 (Gram-Schmidt orthogonalization). Let $x_1, \ldots, x_n \in V$ be linearly independent.

1: **for** $k := 1$ **to** n **do**
2: $y_k := x_k - \sum_{j=1}^{k-1} \langle x_k, z_j\rangle z_j,$
3: $z_k := \frac{1}{\|y_k\|} y_k,$
4: **end for**

The output: $z_1, \ldots, z_n$ - an orthonormal basis of the space $\mathrm{span}\{x_1, \ldots, x_n\}$.
Proof. (The correctness of Gram-Schmidt orthogonalization.) By mathematical induction on n we will prove that $z_1, \ldots, z_n$ is an orthonormal basis of the space $\mathrm{span}\{x_1, \ldots, x_n\}$. For $n = 1$ it is $y_1 = x_1 \neq o$ and $z_1 := \frac{1}{\|x_k\|} x_1$ is well defined and $\mathrm{span}\{x_1\} = \mathrm{span}\{z_1\}$.

The induction step $n \leftarrow n-1$. Let us assume that $z_1, \ldots, z_{n-1}$ is an orthonormal basis of the space $\mathrm{span}\{x_1, \ldots, x^{n-1}\}$. If $y_n = o$, then $x_n = \sum_{j=1}^{n-1} \langle x_n, z_j\rangle z_j$

[2]Jean Baptiste Joseph Fourier (1768-1830), a French mathematician and physicist. He used the series around the year 1807 to solve the problem of heat conduction in solid bodies.

and hence $x_n \in \text{span}\{z_1, \ldots, z_{n-1}\} = \text{span}\{x_1, \ldots, x_{n-1}\}$, which contradicts linear independence of the vectors $x_1, \ldots, x_n$. Therefore $y_n \neq o$ and $z_n := \frac{1}{\|y_n\|} y_n$ is well defined and its norm is equal to 1.

Now, we will prove that $z_1, \ldots, z_n$ is an orthonormal system. It follows from the induction hypothesis that $z_1, \ldots, z_{n-1} \subseteq \text{span}\{x_1, \ldots, x_n\}$, and therefore $\text{span}\{z_1, \ldots, z_n\} \subseteq \text{span}\{x_1, \ldots, x_n\}$. Since bothh spaces have the same dimension, the equality holds (Theorem 5.42). ♣

The advantage of Gram-Schmidt orthogonalization is that it can be used in any space with scalar product. Especially when using the standard scalar product of $\mathbb{R}^n$, the orthogonalization can be expressed using matrices (see Remark 13.8), but, on the other hand, in this case there are also different methods that have better numerical properties, compare Section 13.3.

Consequence 8.22 (Existence of an orthonormal basis). Every finitely generated space (with scalar product) has orthonormal basis.
Proof. We know (Theorem 5.34) that every finitely generated space has a basis; and the basis can be orthogonalized using Gram-Schmidt method. ♣

Let us note that the statement is not valid for infinitely generated spaces - there are spaces with scalar product that have no orthonormal basis, see Bečvář [2005].

Consequence 8.23 (Extension of an orthonormal system to an orthonormal basis). Every orthonormal system in a finitely generated spacecan be extended to an orthonormal basis.
Proof. We know (Theorem 5.41) that every orthonormal system of vectors $z_1, \ldots, z_m$ can be extended to a basis $z_1, \ldots, z_m, x_{m+1}, \ldots, x_n$, which can be orthogonalize using Gram-Schmidt orthogonalization to $z_1, \ldots, z_m, z_{m+1}, \ldots, z_n$. Note that the orthogonalization doesn't change the first m vectors.

Another useful relation is Bessel inequality and Parseval equality.

Theorem 8.24 Let $z_1, \ldots, z_n$ be an orthonormal system in V, and $x \in V$. Then

$$\text{(1) Bessel inequality:} \quad \|x\|^2 \geq \sum_{j=1}^{n} |\langle x, z_j \rangle|^2,$$

$$\text{(2) Parseval equality:} \quad \|x\|^2 = \sum_{j=1}^{n} |\langle x, z_j \rangle|^2 \quad \text{if and only if } x \in \text{span}\{z_1, \ldots, z_n\}.$$

Proof. (1) follows from

$$0 \leq \left\langle x - \sum_{j-1}^{n} \langle x, z_j \rangle z_j, x - \sum_{j-1}^{n} \langle x, z_j \rangle z_j \right\rangle =$$

$$\langle x, x \rangle - \sum_{j-1}^{n} \overline{\langle x, z_j \rangle} \langle x, z_j \rangle - \sum_{j-1}^{n} \langle x, z_j \rangle \langle z_j, x \rangle + \sum_{j-1}^{n} \langle x, z_j \rangle \overline{\langle x, z_j \rangle} =$$

$$= \|x\|^2 - \sum_{j-1}^{n} |\langle x, z_j\rangle|^2.$$

(2) follows from the preceding computation, because the equality holds if and only if $x = \sum_{j-1}^{n} \langle x, z_j\rangle z_j$. ♣

Parseval equality shows that, in other words, in any finitely generated space V the norm of an arbitrary $x \in V$ can be expressed as the standard Euclidean norm of its vector of coordinates $\|x\| =$
$sqrt[x]_B^T[x]_B$, where B is an orthonormal basis of V. As it will be shown below in 8.25, the property holds analogously for scalar product: $\langle x, y\rangle = [x]_B^T[y]_B$ for a real space, and $\langle x, y\rangle = [x]_B^T\overline{[y]_B}$ for a complex space.

Theorem 8.25 Let $z_1, \ldots, z_n$ be an orthonormal basis of a space V, and $x, y \in V$. Then $\langle x, y\rangle = [x]_B^T\overline{[y]_B}$.
Proof. In view of Theorem 8.20, $[x]_B = (\langle x, z_1\rangle, \ldots, \langle x, z_n\rangle)^T$. Now,

$$\langle x, y\rangle = \left\langle \sum_{j=1}^{n} \langle x, z_j\rangle z_j, y \right\rangle = \sum_{j=1}^{n} \langle x, z_j\rangle\langle z_j, y\rangle =$$

$$= \sum_{j=1}^{n} \langle x, z_j\rangle\overline{\langle y, z_j\rangle} = [x]_B^T\overline{[y]_B}.$$

It is not difficult to see that the theorem holds in the opposite direction as well. Which means that the mapping $\langle\cdot,\cdot\rangle$ is a scalar product on V it and only if it can be expressed as $\langle x, y\rangle = [x]_B^T[y]_B$ for some orthonormal basis B. This implies that each scalar product is the standard scalar product from the point of view of certain orthonormal basis.

8.3. Orthogonal complement and projection

Orthogonal complement is a useful notion with geometric interpretation. Moreover, an orthogonal projection is a very important tool, its use in many fields overcomes its basic geometric significance.

Definition 8.26 (Orthogonal complement). Let V be a vector space and $M \subseteq V$. Then an *orthogonal complement* of the set M is
$M^\perp := \{x \in V; \langle x, y\rangle = 0 \forall y \in M\}$.

Orthogonal complement $M^\perp$ contains such vectors x, that are orthogonal to all vectors of M (sometimes we say that x is orthogonal to M).

Example 8.27 The orthogonal complement to a vector $(2,5)^T$ is the line $\text{span}\{(5,-2)^T\}$. The orthogonal complement to the whole line $\text{span}\{(2,5)^T\}$is also the line $\text{span}\{(5,-2)^T\}$.

Theorem 8.28 (Properties of orthogonal complement of a set). Let V be a vector space and $M, N \subseteq V$. Then
(1) $M^\perp$ is a subspace of V,

(2) if $M \subseteq N$, then $N^\perp \subseteq M^\perp$,
(3) $M^\perp = \operatorname{span}\{M\}^\perp$.
Proof.
(1) Let us check the properties of a subspace: $o \in M^\perp$ trivially. Now, let $x_1, x_2 \in M^\perp$. Then $\langle x_1, y\rangle = \langle x_2, y\rangle = 0$ for all $y \in M$, so $\langle x_1 + x_2, y\rangle = \langle x_1, y\rangle + \langle x_2, y\rangle = 0$. Finally, let $x \in M^\perp$, which implies $\langle x, y\rangle = 0$ for all $y \in M$. Then for each scalar α it is $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle = 0$.
(2) Let $x \in N^\perp$, which implies $\langle x, y\rangle = 0$ for all $y \in N$. Obviously $\langle x, y\rangle = 0$ for all y in a smaller M, and hence $x \in M^\perp$.
(3) $M \subseteq \operatorname{span}\{M\}$, and (2) implies that $\operatorname{span}\{M\}^\perp \subseteq M^\perp$. The second inclusion follows from the fact that if x is orthogonal to some vectors, it is also orthogonal to their linear combinations, and therefore to their span. Formally¿ let $x \in M^\perp$, i.e., $\langle x, y\rangle = 0$ for all $y \in M$. Especially, $\langle x, y_i\rangle = 0$, where $y_1, \dots, y_n \in M$ is the basis of $\operatorname{span}\{M\}$. Then for an arbitrary $y = \sum_{i=1}^n \in \operatorname{span}\{M\}$ it is

$$\langle x, y\rangle = \left\langle x, \sum_{i=1}^n \alpha_i y_i \right\rangle = \sum_{i=1}^n \overline{\alpha_i}\langle x, y_i\rangle = 0. \quad \clubsuit$$

The property (3) says that the orthogonal complement of a subspace or its basis is the same. This simplifies practival construction of a complement, because it is sufficient to verify orthogonality to vectors of the basis.

While the previous theorem deals with an orthogonal complement of an arbitrary set of vectors, now we will investigate orthogonal complements of subspaces. Note that the proof of the first part is relatively constructive, and gives a method how to compute an orthogonal complement (or its basis).

Theorem 8.29 (Properties of orthogonal complement of a subspace). Let V be a vector space and U its subspace. Then:
(1) If $z_1, \dots, z_m$ is an orthonormal basis of U, and if $z_1, \dots, z_m, z_{m+1}, \dots, z_n$ is its extension to an orthonormal basis of V, then $z_{m+1}, \dots, z_n$ it an orthonormal basis of $U^\perp$.
(2) $\dim V = \dim U + \dim U^\perp$,
(3) $V = U + U^\perp$,
(4) $(U^\perp)^\perp = U$,
(5) $U \cap U^\perp = \{o\}$.
Proof.
(1) It is obvious that $z_{m+1}, \dots, z_n$ is an orthonormal system in V, and therefore it is sufficient to prove that $\operatorname{span}\{z_{m+1}, \dots, z_n\} = U^\perp$.
The inclusion "*supseteq*". Every $x \in V$ has the Fourier series $x = \sum_{i=1}^n \langle x, z_i\rangle z_i$. If $x \in U^\perp$, then $\langle x, z_i\rangle = 0$, $i = 1, \dots, m$, and hence $x = \sum i = m + 1^n \langle x, z_i\rangle z_i \in \operatorname{span}\{z_{m+1}, \dots, z_n\}$.
(2) It follows from (1) that $\dim V = n$, $\dim U = m$, $\dim U^\perp = n - m$.
(3) It follows from (1) that

$$x = \underbrace{\sum_{i=1}^m \langle x, z_i\rangle z_i}_{\in U} + \underbrace{\sum_{i=m+1}^n \langle x, z_i\rangle z_i}_{\in U^\perp} \in U + U^\perp.$$

(4) It follows from (1) that $z_{m+1}, \dots, z_n$ is an orthonormal basis of $U^\perp$, and therefore $z_1, \dots, z_m$ is an orthonormal basis of $(U^\perp)^\perp$.
(5) From the preceding and Theorem 5.48 about the dimension of the union and intersection we get $\dim(U \cap U^\perp) = \dim V - \dim U - \dim U^\perp = 0$. ♣

Another nice property of atrhogonal systems is that they allow to compute in a simple way the projection x_U of a vector x to the subspace U, which is the vector that is closest to x. The next theorem allows us to define a projection as a mapping $V \to U$, defined by $x \mapsto x_U$.

The figure at page 98 of the Czech text.

Theorem 8.30 (On orthogonal projection). Let V be a vector space and U be a subspace of V. Then for each $x \in V$ there is the unique x_U such that

$$\|x - x_U\| = \min_{y \in U} \|x - y\|.$$

Moreover, if $z_1, \dots, z_m$ is an orthonormal basis of U, then

$$x_U = \sum_{i=1}^{m} \langle x, z_i \rangle z_i. \qquad 8.1$$

Proof. Let $z_1, \dots, z_m, z_{m+1}, \dots, z_n$ be the extension to an orthonormal basis of V. Let us define $x_U := \sum_{i=1}^m \langle x, z_i \rangle z_i \in U$, and we will show that this is the vector we are looking for. We have $x - x_U = \sum i = 1^n \langle x, z_i \rangle z_i - \sum_{i=1}^m \langle x, z_i \rangle z_i = \sum i = m + 1^n \langle x, z_i \rangle z_i \in U^\perp$. Let $y \in U$ be an arbitrary vector. Since $x_U - y \in U$, we can use Pythagoras theorem that gives

$$\|x - y\|^2 = \|(x - x_U) + (x_U - y)\|^2 = \|x - x_U\|^2 + \|x_U - y\|^2 \geq \|x - x_U\|^2,$$

or, equivalently, $\|x - y\| \geq \|x - x_U\|$, which proves minimality. In order to prove uniqueness, it is sufficient to realize that the equality hold if and only if $\|x_U - y\|^2 = 0$, i.e. if $x_U = y$.♣

If a vector x belongs to a subspace U, its projection is x itself, and the formula 8.1 gives the Fourier series of Theorem 8.20. It is also easy to see that if $x \in U^\perp$, then its projection is o.

Remark 8.31 In view of (3) and (5) of Theorem 8.29 the space V can be expressed as a direct sum of the subspaces U and $U^\perp$ (Remark 5.50). This means, among others, that every vector $v \in V$ has the unique expression as $v = u + u'$, where $u \in U$ and $u' \in U^\perp$. IMoreover, in view of Theorem 8.30, the vector u is the projection of v into U, and the vector u' is the projection of v into $U^\perp$.

We know from the proof 8.30 that $x - x_U \in U^\perp$, but this property is not only a necessary, but also a sufficient condition for x_U being a projection.

Theorem 8.32 In the notation of Theorem 8.30, if some $y \in U$ verifies $x - y \in U^\perp$, then $y = x_U$.

Proof. Since $(x - y) \perp (y - x_U)$, we use Pythagoras theorem that says

$$\|x - x_U\|^2 = \|x - y\|^2 + \|y - x_U\|^@ \geq \|x - y\|^2.$$

We get $\|x - x_U\| \geq \|x - y\|$, and from the properties and uniqueness of the projection we get $y = x_U$. ♣

Paragraphs 8.33, 8.34, 8.35 skipped

8.4. Orthogonal complement and projection in $\mathbb{R}^n$

We know from the previous section how to compute an orthogonal complement and projection for an arbitrtary finitely generated vector space with scalar product, using an orthonormal basis. Now, we will show that in $\mathbb{R}^n$ and for the standard scalar product, the transformations can be formulated explicitly and directly without computing an orthonormal basis.

The following theorem says how to compute the orthogonal complement of any subspace of $\mathbb{R}^n$, if we know its basis or a finite system of generators (they represent rows of the matrix A).

Theorem 8.36 (Orthogonal complement in $\mathbb{R}^n$). Let $A \in \mathbb{R}^{m\times n}$. Then[3] $\mathcal{R}(A) \perp= \mathrm{Ker}(A)$.
Proof. It follows from properties ofan orthogonal complement (Theorem 8.23 (3)) that $\mathcal{R}(A)^\perp = \{A_{1*}, \dots, A_{m*}\}^\perp$. Thus, $x \in \mathcal{R}(A)^\perp$ if and only if x is orthogonal to all rows of the matrix A, i.e., $A_{i*}x = 0$ for all $i = 1, \dots, m$. Equivalently, $Ax = o$, which is $x \in \mathrm{Ker}(A)$. ♣

Example 8.37 Let V be a space generated by vectors $(1, 2, 3)^T$ and $(1, -1, 0)^T$. We want to determine $V^\perp$, and therefore we will use the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & -1 & 0 \end{pmatrix},$$

because $V = \mathcal{R}(A)$. Now, it is sufficient to find the basis of $V^\perp = \mathrm{Ker}(A)$, which is given, e.g., by the vector $(1, 1, -1)^T$. ♣

The characterization of an orthogonal complement has also theoretical consequences, e.g., the relation of the matrix A and the matrix A^TA. Be careful, for column spaces the analogy doesn't hold!

Consequence 8.38 Let $A \in \mathbb{R}^{m\times n}$. Then
(1) $\mathrm{Ker}(A^TA) = \mathrm{Ker}(A)$,
(2) $\mathcal{R}(A^TA) = \mathcal{R}(A)$,
(3) $\mathrm{rank}(A^TA) = \mathrm{rank}(A)$.
Proof.
(1) If $x \in \mathrm{Ker}(A)$, then $Ax = o$, and hence $A^TAx = A^To = o$, and therefore $x \in \mathrm{Ker}(A)$. Conversely, if $x \in \mathrm{Ker}(A^TA)$, then $A^TAx = o$. Multilying by x^T

[3] Let us recall that $\mathcal{R}(A)$ is the space generated by rows of of A, and $\mathcal{S}$ is the space generated by columns of A.

we get $x^T A^T Ax = o$, which is $\|Ax\|^2 = 0$. Taking into account the properties of a norm, we get $Ax = o$, and $x \in \mathrm{Ker}(A)$.
(2) $\mathcal{R}(A^T A) = \mathrm{Ker}(A^T A)^\perp = \mathrm{Ker}(A)^\perp = \mathcal{R}(A)$.
(3) Trivial in view of (2). ♣

Let us now look at the projection of a vector x into the subspace V, for which we will derive an explicit formula. If vectors of a basis of the space V are put into columns of the matrix A, then the projection of x into V can be formulated as a projection of x into $\mathcal{S}(A)$.

medskip**Theorem 8.39** (Orthogonal projection in $\mathbb{R}^m$). Let $A \in \mathbb{R}^{m\times n}$ of the rank n. Then the projection of a vector $x \in \mathbb{R}^m$ into the column vector $\mathcal{S}(A)$ is $x' = A(A^T A)^{-1}A^T x$.
Proof. Let us first note that x' is well defined. The matrix $A^T A$ hes dimension n (Consequence 8.38 (3)), and hence it is regular and has an inversion matrix. According to 8.32 it is sufficient to show that $x' \in \mathcal{S}(A)$ and $x - x' \in \mathcal{S}(A)^\perp$. The first property holds, because $x' = Az$ for $z = (A^T A)^{-1}A^T x$ For the second one it is sufficient to verify that $x - x' \in \mathcal{S}(A)^\perp = \mathcal{R}(A^T)^\perp = \mathrm{Ker}(A^T)$, and this follows from

$$A^T(x-x') = A^T(x-A(A^T A)^{-1}A^T x) = A^T x-A^T A(A^T A)^{-1}A^T x = A^T x-A^T x = o. \quad ♣$$

Let us note that the projection is a linear mapping and in view of the preceding theorem $P := A(A^T A)^{-1}A^T$ is its matrix (with respect to the canonical basis). Moreover, this matrix has a lot of remarkable properties. E.g., it is symmetric, $P^2 = P$, and it is regular only if $m = n$.

Espacially, the matrix of projection to a one-dimensional space (a line) has form $P = a(a^T a) - 1a^T$, where $a \in \mathbb{R}^n$ is the direction of the line. If, moreover, the vector in the direction of the line is normed so that $\|a\|_2 = 1$, then $a^T a = 1$, and the projection gets a simple form $P = aa^T$.

Theorem 8.40 (Orthogonal projection into the complement). Let $P \in \mathbb{R}n \times n$ be the matrix of the projection onto a space V, which is a subspace of $\mathbb{R}^n$. Then $I - P$ is the matrix of the projection onto $V^\perp$.
Proof. According to 8.29, every vector $x \in \mathbb{R}^n$ can be uniquely decomposed to $x = y + z$, where $y \in V$ and $z \in V^\perp$. In view of 8.30, y is the projection of x into V, and z is the projection into $V^\perp$. It follows that $z = x-y = x-Px = (I-P)x$. ♣

Example 8.41 (The matrrix of the projection into $\mathrm{Ker}(A)$). Theorem 8.40 allows us to formulate in an elegant way the projection into the kernel of a matrix $A \in \mathbb{R}^{m\times n}$. Let us suppose that $\mathrm{rank}(A) = m$. Since $\mathrm{Ker}(A)^\perp = \mathcal{R}(A) = \mathcal{S}(A^T)$, the matrix of projection into $\mathrm{Ker}(A)$ is given by $I-A^T(AA^T)^{-1}A$, where $A^T(AA^T)^{-1}A$ is the matrix of the projection into $\mathcal{S}(A^T)$. ♣

Remark 8.42 Skipped

8.5. Least squares method

The section is skipped

8.4. Orthogonal matrices

Let us consider a linear mapping in the space $\mathbb{R}^n$. How this mapping (or its matrix) should look like not to deform geometrical objects? A rotation or a flip horizontally or vertically or along another axis are examples of such mappings, but we would like to analyze them in general. We will show that the property is related to so called orthogonal matrices. But such matrices are far more important. Since they have good numerical properties (see section 1.2 and 3.5.1), we encounter them ofthe in different numerical algorithms.

Also in this section we consider the standard scalar product in $\mathbb{R}^n$ and Euclidean norm.

Definition 8.46 (Orthogonal and unitary matrices). A matrix $Q \in \mathbb{R}^{n\times n}$ is *orthogonal*, if $Q^TQ = I_n$. A matrix $Q \in \mathbb{C}^{n\times n}$ is unitary, if $\overline{Q}^TQ = I_n$.

The notion of unitary matrix is a generalization of orthogonal matrices for complex numbers. However, in the sequel, we will work with orthogonal matrices only.

Theorem 8.47 (Characterization of orthogonal matrices). Let $Q \in \mathbb{R}^{n\times n}$. Then the following statements are equivalent:
(1) Q is orthogonal,
(2) Q is regular and $Q^{-1} = Q^T$,
(3) $QQ^T = I_n$,
(4) Q^T is orthogonal,
(5) Q^{-1} exists and is orthogonal,
(6) columns of Q represent an orthogonal basis of $\mathbb{R}^n$,
(7) rows of Q represent an orthogonal basis of $\mathbb{R}^n$.
Proof. Briefly (1)-(5). If Q is orthogonal, then $Q^TQ = I$ and therefore $Q^{-1} = Q^T$, and conversely. Using properties of the inversion, we get $QQ^T + I$, or $(Q^T)^TQ^T = I$ and therefore Q^T is orthogonal.

(6) It follows from $Q^TQ = I$ (by comparing elements at the position i, j, that $\langle Q_{*i}, Q_{*j}\rangle = 1$ if $i = j$, and $\langle Q_{*i}, Q_{*j}\rangle = 0$ if $i \neq j$. This means that the colums form an orthonormal system. The converse is analogous. ♣

In view of (6) it sould be more appropriate to say “orthonormal matrix”, but the term orthogonal matrix is commonly used.

Theorem 8.48 (A product of orthogonal matrices). If $Q_1, Q_2 \in \mathbb{R}^{n\times n}$ are orthogonal, then Q_1Q_2 is orthogonal as well.
Proof. $(Q_1Q_2)^TQ_1Q_2 = Q_2^TQ_1^TQ_1Q_2 = Q_2^TQ_2 = I_n$.

Example 8.49 (Examples of orthogonal matrices).

- Identity matrix I_n or the matrix $-I_n$.
- *Householder matrix*: $H(a) := I_n - \frac{2}{a^Ta}aa^T$, where $o \neq a \in \mathbb{R}^n$. Its geometric meaning is the following: Let x' be the projection of the point x to tohe line span$\{a\}$, and consider linear mapping of rotating the point x along the axis span$\{a\}$ by 180^o. Using Theorem 8.39 on projection, we get

that the point x is mapped to the vector

$$x + 2(x' - x) = 2x' - x = 2a(a^T a)^{-1} a^T x = \left(2\frac{aa^T}{a^T a} - I\right) x.$$

Thus, the matrix of the rotation is $\frac{2}{a^T a} aa^T - I$. Let us now the mirror image according to a hyperplane with the normale a. This can be represented as the rotation by 180^o along a and then flipping by the origin. This means that the matrix of this mapping is $I - \frac{2}{a^T a} aa^T = H(a)$.

Figures at page 104:
Left: Rotation along the line a by 180^o,
Right: Mirror by the hyperplane with the normale a

It can be shown that any orthogonal matrix of the rank n can be written as the product of at most n appropriate Householder matrices.

- *Givens matrix*[4]: For $n = 2$ it is the matrix of rotation by the angle α counterclockwise

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}.$$

It is a matrix of the form $\begin{pmatrix} c & -s \\ s & c \end{pmatrix}$, where $c^2 + s^2 = 1$ and any such matrix represents rotation. More generally, in the dimansion n, it is the matrix representing rotation by α in the plane given by axis' x_i, x_j, i.e.,

$$G_{i,j}(c, s) = \begin{pmatrix} I & & & & \\ & c & & -s & \\ & & I & & \\ & s & & c & \\ & & & & I \end{pmatrix}.$$

Also Givens matrices have the property that any orthogonal matrix is a product of Givens matrices, but in general we need up to $\binom{n}{2}$ factors and possibly one diagonal matrix with $+1$ and -1 on the diagonal.

Theorem 8.50 (Properties of orthogonal matrices). Let $Q \in \mathbb{R}^{n\times n}$ be an orthogonal matrix. Then
(1) $\langle Qx, Qy\rangle = \langle x, y\rangle$ for each $x, y \in \mathbb{R}^n$,
(2) $\|Qx\| = \|x\|$ for each $x \in \mathbb{R}^n$,
(3) $|Q_{ij}| \le 1$ and $|Q^{-1}_{ij}| \le 1$ for each $i, j = 1, \ldots, n$,
(4) $\begin{pmatrix} 1 & o^T \\ o & Q \end{pmatrix}$ is an orthogonal matrix.

[4] James Wallace Givens, Jr., (1910-1993), American mathematician

Proof.
(1) $\langle Qx, Qy\rangle = (Qx)^T Qy = x^T Q^T Qy = x^T Iy = x^T y = \langle x, y\rangle$.
(2) $\|Qx\| = \sqrt{\langle Qx, Qx\rangle} = \sqrt{\langle x, x\rangle} = \|x\|$.
(3) In view of (6) of Theorem 8.47, it is $\|Q_{*j}\|1$ for each $j = 1, \ldots, n$. Therefore $1 = \|Q_{*j}\|^2 = \sum_{i=1}^n q_{ij}^2$ qhich means that $q_{ij}^2 \leq 1$, and hence $|q_{ij}| \leq 1$. The matrix Q^{-1} is orthogonal, and the statement holds for it as well.
(4) From the definition

$$\begin{pmatrix} 1 & o^T \\ o & Q \end{pmatrix}^T \begin{pmatrix} 1 & o^T \\ o & Q \end{pmatrix} = \begin{pmatrix} 1 & o^T \\ o & Q^T Q \end{pmatrix} = I_{n+1}. \quad ♣$$

If we see Q as the matrix of the correesponding linear mapping $x \mapsto Qx$, then the property (1) of Theorem 8.50 says that the mapping preserves angles, and (2) says that the lengths are preserved. The statement holds also in the opposite direction: matrices preserving scalar product must be orthogonal (see Theorem 8.51) and even matrices preserving Euclidean norm must be orthogonal (Horn and Johnson, 1985). The property (3) is appreciated in numerical mathematics, because Q and Q^{-1} have bounded elements. The most important for numerical computing is (2), because multiplying by an orthogonal matrix (and also rounding errors) have a tendency not to grow.

Finally let us show some generalization of the aove properties to an arbitrary scalar product and a general linear mapping.

medskip**Theorem 8.51** (Orthogonal matrices and linear mappings) Let U, V be spaces over $\mathbb{R}$ with an arbitrary scalar product, and $f : U \to V$ be a linear mapping. Let B_1 is an orthonormal basis of U, and B_2 is an orthonormal basis of V. Then the matrix ${}_{B_2}[f]_{B_1}$ is orthogonal if and only if $\langle f(x), f(y)\rangle = \langle x, y\rangle$ for each $x, y \in U$.
Proof. According to 8.25 and the properties of the matrix of a linear mapping it is

$$\langle x, y\rangle = [x]_{b_1}^T [y]_{B_1},$$

$$\langle f(x), f(y)\rangle = [f(x)]_{B_2}^T [f(y)]_{B_2} = ({}_{B_2}[f]_{B_1} \cdot [x]_{B_1})_{B_2}^T [f]_{B_1} \cdot [y]_{B_1} =$$

$$= [x]_{B_1}^T {}_{B_2}[f]_{B_1} [y]_{B_1}.$$

Therefore, if ${}_{B_2}[f]_{B_1}$ is orthogonal, then $\langle f(x), f(y)\rangle = \langle x, y\rangle$. Conversely, if $\langle f(x), f(y)\rangle = \langle x, y\rangle$ holds for each $x, y \in U$, the equality holds especially for vectors with coordinates representing by unit vectors. If we substitute the i-th vector of the basis B_1 for x and the j-th vector of the basis B_2 for y, we get $[x]_{b_1} = e_i$, $[y]_{B_2} = e_j$ and therefore

$$(I_n)_{ij} = e_i^T e_j = [x]_{B_1}^T [y]_{B_1} = \langle x, y\rangle = \langle f(x), f(y)\rangle = [x]_{B_1}^T \cdot {}_{B_2}[f]_{B_1}^T \cdot {}_{B_2}[f]_{B_1} \cdot [y]_{B_1} =$$

$$= e_i^T \cdot_{B_2} [f]_{B_1}^T \cdot_{B_2} [f]_{B_1} \cdot e_j = \left({}_{B_2}[f]_{B_1}^T \cdot {}_{B_2}[f]_{B_1}\right)_{ij}.$$

Considering particular elements of the matrices, we get $I_n =_{B_2} [f]_{B_1}^T {}_{B_2}[f]_{B_1}$.

medskip**Theorem 8.52** (Orthogonal matrices and matrices of transfer) Let V be a space ove $\mathbb{R}$ with an arbitrary scalar product, and b_1, b_2 be two of its

basis. The arbitrary two of the following properties imply the third one:
(1) B_1 is an orthonormal basis,
(2) B_2 is an orthonormal basis,
(3) ${}_{B_2}[id]_{B_1}$ is an orthonormal basis.
Proof.
The implication "(1),(2) $\Rightarrow$ (3)". It follows from Theorem 8.51, because the identity preserves the scalar product.
The implication "(2),(3) $\Rightarrow$ (1)". Let $B_1 = \{x_1, \ldots, x_n\}$. From the definition the columns of ${}_{B_2}[id]_{B_1}$ form vectors $[x_i]_{B_2}$ that are (due to the orthogonality of the transfer matrix) orthonormal under the standard scalar product in $\mathbb{R}^n$. In view of Theorem 8.25, $\langle x_i, x_j \rangle = [x_i]_{B_2}^T [x_j]_{B_2}$, which is 0 for $i = j$ and 1 otherwise.
The implication "(3),(1) $\Rightarrow$ (2)". It follows from the above using symmery, because ${}_{B_1}[id]_{B_2} ={}_{B_2} [id]_{B_1}^{-1}$. ♣