VL13 ↓

5.3 Orthonormal systems and bases

For some applications it it very useful to have a set of vectors $\{\mathbf{u}_1 \dots \mathbf{u}_k\} \subset \mathbb{R}^n$ which are mutually orthogonal:

$$i \neq j \quad \Rightarrow \mathbf{u}_i \perp \mathbf{u}_j \quad \Leftrightarrow \quad \langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$$

and have unit norm:

$$\|\mathbf{u}_i\| = \sqrt{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} = 1.$$

Using the Kronecker symbol:

$$\delta_{ij} = \begin{cases} 1 & : \quad i = j \\ 0 & : \quad i \neq j \end{cases}$$

we may write this in short:

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij}.$$

Definition 5.20. OS, ONS, OB, ONB

Let U be a linear subspace of $\mathbb{R}^n$. A family $\mathcal{F} = (\mathbf{u}_1, \dots, \mathbf{u}_k)$ consisting of vectors from U is called:

- *Orthogonal system (OS) if the vectors in $\mathcal{F}$ are mutually orthogonal: $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ for all $i, j \in \{1, \dots, k\}$ with $i \neq j$;*
- *Orthonormal system (ONS) if $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij}$ for all $i, j \in \{1, \dots, k\}$;*
- *Orthogonal basis (OB) if it is an OS and a basis of U;*
- *Orthonormal basis (ONB) if it is an ONS and a basis of U.*

If $\mathcal{F}$ is an ONB, then the Gram matrix $G(\mathcal{F})$ is the identity matrix and projections are very easily calculable.

ONB: $$G(\mathcal{F}) = \begin{pmatrix} \langle u_1, u_1 \rangle & \cdots & \langle u_k, u_1 \rangle \\ \vdots & & \\ \langle u_1, u_k \rangle & \cdots & \langle u_k, u_k \rangle \end{pmatrix} = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} = \mathbb{1}$$

Example 5.21. Let $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{eukl}$ the standard inner product.

(a) The canonical unit vectors

$$\mathbf{e}_1 = (1, 0, \dots, 0)^T, \quad \mathbf{e}_2 = (0, 1, 0, \dots, 0)^T, \quad \dots, \quad \mathbf{e}_n = (0, \dots, 0, 1)^T$$

$\langle e_i, e_j \rangle = \delta_{ij}$ and $(e_1, \dots, e_n)$ basis of $\mathbb{R}^n$

in $\mathbb{R}^n$ define an ONB for $U = \mathbb{R}^n$.

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \xrightarrow{III - I} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -2 & 0 \end{pmatrix} \xrightarrow{III \leftrightarrow II} \begin{pmatrix} 1 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(b) The family $\mathcal{F} = (\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$ given by

$$\mathbf{u}_1 = (1,0,1)^T, \quad \mathbf{u}_2 = (1,0,-1)^T, \quad \mathbf{u}_3 = (0,1,0)^T$$

defines an OB of $\mathbb{R}^3$. We show this: We immediately have $\langle \mathbf{u}_1, \mathbf{u}_3 \rangle = 0$ and $\langle \mathbf{u}_2, \mathbf{u}_3 \rangle = 0$. Moreover, we find

$$\left\langle \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle \quad \left\langle \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\rangle$$

$$\langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \left\langle \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\rangle = 1 + 0 - 1 = 0.$$

Hence, $\mathcal{F}$ is an OS. It remains to show that $\mathcal{F}$ is also a basis for $\mathbb{R}^3$. Since $\dim(\mathbb{R}^3) = 3$ and $\mathcal{F}$ consists of three linearly independent vectors, we are finished. For showing the linear independence, the next Proposition 5.22 will be always helpful.

(c) Normalising the vectors from (b), we obtain an ONB $(\frac{1}{\sqrt{2}}\mathbf{u}_1, \frac{1}{\sqrt{2}}\mathbf{u}_2, \mathbf{u}_3)$.

$$\Rightarrow \quad v_1 := \frac{u_1}{\|u_1\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad v_3 = u_3$$

Proposition 5.22. An OS is linearly independent.

Let $\mathcal{F} = (\mathbf{u}_1, \ldots, \mathbf{u}_k)$ be an OS in $\mathbb{R}^n$ with $\mathbf{u}_i \neq \mathbf{o}$ for $i = 1, \ldots, k$. Then $\mathcal{F}$ is linearly independent.

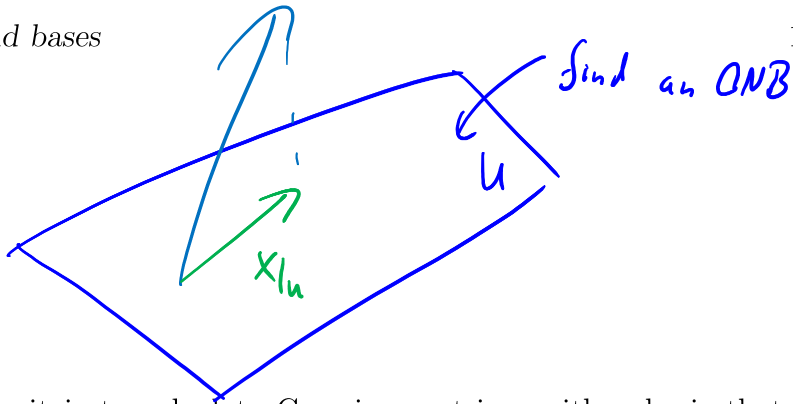
Proof. Let $\mathcal{F}$ be an OS. To show the linear independence of $\mathcal{F}$, we only have to show that $\alpha_1 \mathbf{u}_1 + \ldots + \alpha_k \mathbf{u}_k = \mathbf{o}$ always implies $\alpha_1 = \ldots = \alpha_k = 0$. Using the inner product for $\mathbf{u}_i$ with $i = 1, \ldots, k$, we get:

$$0 = \langle \mathbf{o}, \mathbf{u}_i \rangle = \langle \alpha_1 \mathbf{u}_1 + \ldots + \alpha_k \mathbf{u}_k, \mathbf{u}_i \rangle \overset{(S2),(S3)}{=} \alpha_1 \underbrace{\langle u_1, u_i \rangle}_{=0} + \cdots + \alpha_i \langle u_i, u_i \rangle + \cdots + \alpha_k \underbrace{\langle u_k, u_i \rangle}_{=0}$$

$$= \alpha_i \underbrace{\langle u_i, u_i \rangle}_{\neq 0 \text{ (by (S1))}} \quad \Rightarrow \quad \alpha_i = 0$$

$$\Rightarrow \quad \alpha_1 = \cdots = \alpha_k = 0$$

□



Now we can show, how easy it is to calculate Gramian matrices with a basis that is orthogonal.

Proposition 5.23. Gramian matrix for OB and ONB

The Gramian matrix $G(\mathcal{B})$ for an OB $\mathcal{B} = (\mathbf{u}_1, \dots, \mathbf{u}_k)$ is a diagonal matrix:

$$G(\mathcal{B}) = \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{u}_1 \rangle & \langle \mathbf{u}_2, \mathbf{u}_1 \rangle & \dots & \langle \mathbf{u}_k, \mathbf{u}_1 \rangle \\ \langle \mathbf{u}_1, \mathbf{u}_2 \rangle & \langle \mathbf{u}_2, \mathbf{u}_2 \rangle & \dots & \langle \mathbf{u}_k, \mathbf{u}_2 \rangle \\ \vdots & \vdots & & \vdots \\ \langle \mathbf{u}_1, \mathbf{u}_k \rangle & \langle \mathbf{u}_2, \mathbf{u}_k \rangle & \dots & \langle \mathbf{u}_k, \mathbf{u}_k \rangle \end{pmatrix} = \begin{pmatrix} \|\mathbf{u}_1\|^2 & 0 & & \\ 0 & \|\mathbf{u}_2\|^2 & \ddots & \\ & \ddots & \ddots & 0 \\ & & 0 & \|\mathbf{u}_k\|^2 \end{pmatrix}.$$

If $\mathcal{B}$ actually is an ONB, then we have $G(\mathcal{B}) = \mathbb{1}$.

The orthogonal projection $\mathbf{x}_{|U}$ for a vector $\mathbf{x} \in \mathbb{R}^n$ onto the linear subspace $U =$ Span$(\mathcal{B})$ is then given by the coefficients

$$\begin{pmatrix} \|u_1\|^2 & & \\ & \ddots & \\ & & \|u_k\|^2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} = \begin{pmatrix} \langle x, u_1 \rangle \\ \vdots \\ \langle x, u_k \rangle \end{pmatrix}$$

$$\rightsquigarrow \|u_1\|^2 \alpha_1 = \langle x, u_1 \rangle$$

$$\vdots$$

$$\rightsquigarrow \|u_k\|^2 \alpha_k = \langle x, u_k \rangle$$

$$\alpha_1 = \frac{\langle \mathbf{x}, \mathbf{u}_1 \rangle}{\|\mathbf{u}_1\|^2}, \quad \alpha_2 = \frac{\langle \mathbf{x}, \mathbf{u}_2 \rangle}{\|\mathbf{u}_2\|^2}, \quad \dots, \quad \alpha_k = \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle}{\|\mathbf{u}_k\|^2}$$

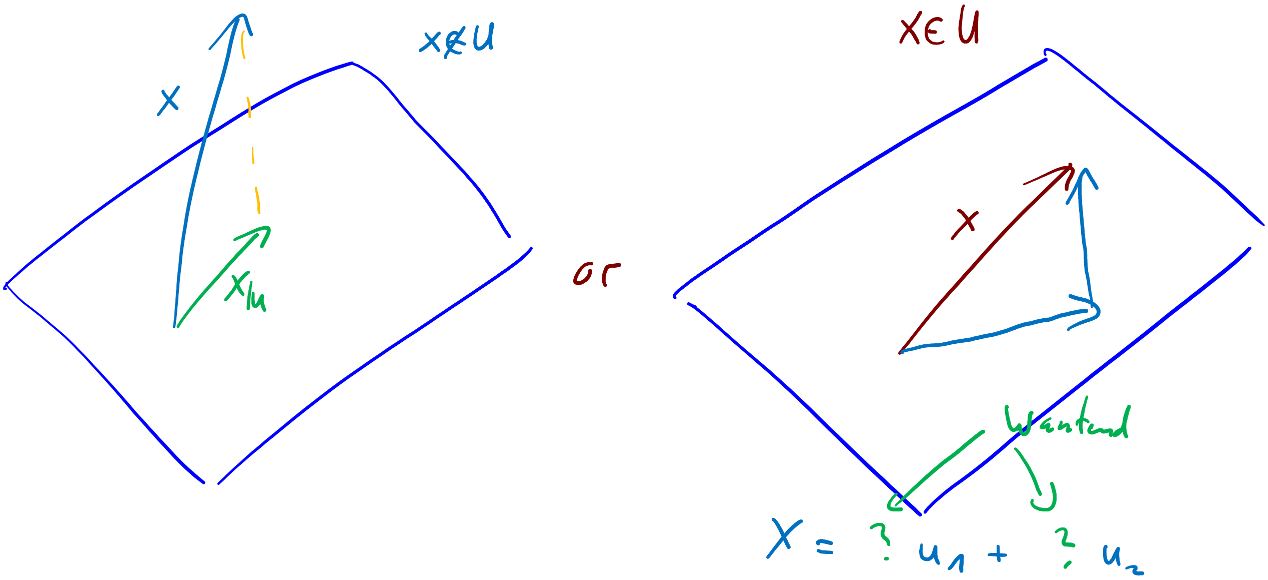
for equation (5.5). We get:

$$\mathbf{x}_{|U} = \frac{\langle \mathbf{x}, \mathbf{u}_1 \rangle}{\|\mathbf{u}_1\|^2} \mathbf{u}_1 + \dots + \frac{\langle \mathbf{x}, \mathbf{u}_k \rangle}{\|\mathbf{u}_k\|^2} \mathbf{u}_k \quad \text{and} \quad \mathbf{x}_{|U^\perp} = \mathbf{x} - \mathbf{x}_{|U}$$

If $\mathcal{B}$ is even an ONB, then all the denominators $\|\mathbf{u}_i\|^2$ are equal to 1.

$$\hookrightarrow x_{|U} = \langle x, u_1 \rangle u_1 + \dots + \langle x, u_k \rangle u_k$$

Even if one is not interested in the projection, this can be helpful for calculating the coefficients for the linear combination.



Corollary 5.24. Fourier expansion w.r.t. an OB or ONB

Let U be a linear subspace of $\mathbb{R}^n$ and $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_k)$ an OB of U. Then the unique linear combination for a vector $\mathbf{x} \in U$ with respect to $\mathcal{B}$ is given by:

$$\mathbf{x} = \alpha_1 \mathbf{u}_1 + \ldots + \alpha_k \mathbf{u}_k \qquad \textit{with} \qquad \alpha_i = \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\|\mathbf{u}_i\|^2} \quad \textit{for all } i \in \{1, \ldots, k\}. \tag{5.6}$$

This formula is called the Fourier expansion of $\mathbf{x}$ with respect to $\mathcal{B}$, and the numbers α_i are called the associated Fourier coefficients. If $\mathcal{B}$ even is an ONB, then

$$\alpha_i = \langle \mathbf{x}, \mathbf{u}_i \rangle \quad \textit{for all} \quad i = 1, \ldots, k.$$

Note that in the case $U = \mathbb{R}^n$, we simply set $k = n$.

Gram-Schmidt

basis of U $\rightsquigarrow$ ONB of U

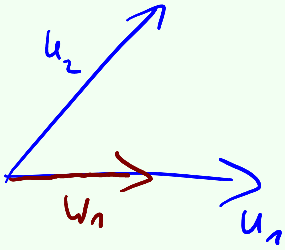
Remark: Gram-Schmidt orthonormalisation

Let U be a linear subspace of $\mathbb{R}^n$ and $(\mathbf{u}_1, \dots, \mathbf{u}_k)$ a basis of U. The following procedure will give us an ONB $(\mathbf{w}_1, \dots, \mathbf{w}_k)$ for U.

(1) Normalise the first vector:

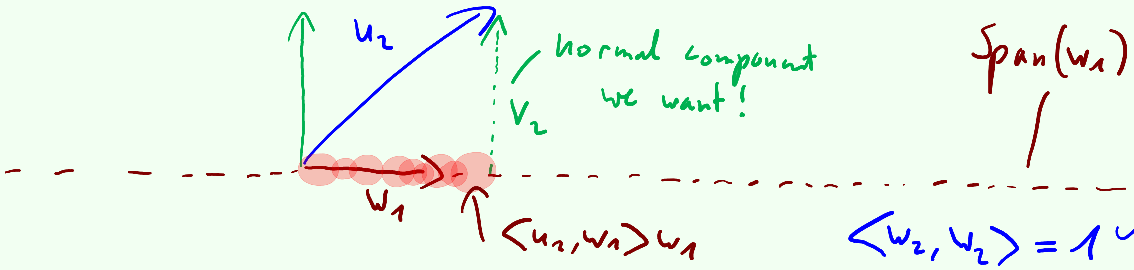
$$\mathbf{w}_1 := \frac{1}{\|\mathbf{u}_1\|}\mathbf{u}_1. \qquad \sqrt{\langle u_1, u_1 \rangle}$$

$\langle w_1, w_1 \rangle = 1$ ✓



(2) Choose the normal component of $\mathbf{u}_2$ with respect to $\mathrm{Span}(\mathbf{w}_1)$

$$\mathbf{v}_2 := \mathbf{u}_2 - \underbrace{\langle \mathbf{u}_2, \mathbf{w}_1 \rangle \mathbf{w}_1}_{\mathbf{u}_2|_{\mathrm{Span}(\mathbf{w}_1)}} \quad \textit{and normalise it:} \quad \mathbf{w}_2 := \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2.$$

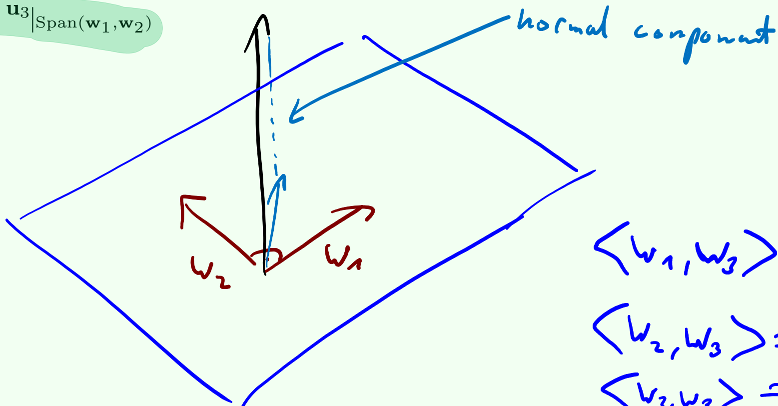


$\langle w_2, w_2 \rangle = 1$ ✓

$\langle w_2, w_1 \rangle = 0$ ✓

(3) Choose the normal component of $\mathbf{u}_3$ with respect to $\mathrm{Span}(\mathbf{w}_1, \mathbf{w}_2)$

$$\mathbf{v}_3 := \mathbf{u}_3 - \underbrace{\Big(\langle \mathbf{u}_3, \mathbf{w}_1 \rangle \mathbf{w}_1 + \langle \mathbf{u}_3, \mathbf{w}_2 \rangle \mathbf{w}_2 \Big)}_{\mathbf{u}_3|_{\mathrm{Span}(\mathbf{w}_1, \mathbf{w}_2)}} \quad \textit{and normalise it:} \quad \mathbf{w}_3 := \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3.$$



$\langle w_1, w_3 \rangle = 0$

$\langle w_2, w_3 \rangle = 0$

$\langle w_3, w_3 \rangle = 1$

$\vdots$

(k) In the last step choose the normal component of $\mathbf{u}_k$ w.r.t. $\mathrm{Span}(\mathbf{w}_1, \dots, \mathbf{w}_{k-1})$

$$\mathbf{v}_k := \mathbf{u}_k - \underbrace{\sum_{i=1}^{k-1} \langle \mathbf{u}_k, \mathbf{w}_i \rangle \mathbf{w}_i}_{\mathbf{u}_k|_{\mathrm{Span}(\mathbf{w}_1, \dots, \mathbf{w}_{k-1})}} \quad \textit{and normalise it:} \quad \mathbf{w}_k := \frac{1}{\|\mathbf{v}_k\|}\mathbf{v}_k.$$

$\rightsquigarrow (w_1, \dots, w_k)$ ONB

Example 5.25. Let $\mathbf{u}_1 = (1,1,0)^T$ and $\mathbf{u}_2 = (2,0,2)^T$ be two vectors in $\mathbb{R}^3$ and $U = \operatorname{Span}(\mathbf{u}_1, \mathbf{u}_2)$ the spanned plane. We calculate an ONB $(\mathbf{w}_1, \mathbf{w}_2)$ for U. The first vector is

$$\mathbf{w}_1 := \frac{1}{\|\mathbf{u}_1\|}\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\end{pmatrix}.$$

For the second vector, we first need to calculate:

$$\mathbf{v}_2 := \mathbf{u}_2 - \langle \mathbf{u}_2, \mathbf{w}_1\rangle \mathbf{w}_1 = \begin{pmatrix}2\\0\\2\end{pmatrix} - \left\langle \begin{pmatrix}2\\0\\2\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\end{pmatrix}\right\rangle \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\\0\end{pmatrix} = \begin{pmatrix}2\\0\\2\end{pmatrix} - \frac{1}{2}2\begin{pmatrix}1\\1\\0\end{pmatrix} = \begin{pmatrix}1\\-1\\2\end{pmatrix}$$

Then $\mathbf{v}_2$ is getting normalised:

$$\mathbf{w}_2 := \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \frac{1}{\sqrt{6}}\begin{pmatrix}1\\-1\\2\end{pmatrix}.$$

Now we have $\|\mathbf{w}_1\| = 1 = \|\mathbf{w}_2\|$ and $\langle \mathbf{w}_1, \mathbf{w}_2\rangle = 0$ and also $\operatorname{Span}(\mathbf{w}_1, \mathbf{w}_2) = U = \operatorname{Span}(\mathbf{u}_1, \mathbf{u}_2)$.

We recall Corollary 5.24: Why are such ONB helpful? Usually, if we want to write a vector $\mathbf{v}$ as a linear combination of basis vectors $\mathcal{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_k)$, we have to solve a linear system:

$$\mathbf{v} = \sum_{i=1}^{k} \lambda_i \mathbf{b}_i\,. \rightsquigarrow \text{Solve LES}$$

If we have an orthonormal basis $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_k)$, then we can dispense with this. We can simply calculate:

$$\langle \mathbf{v}, \mathbf{u}_i\rangle = \left\langle \sum_j \lambda_j \mathbf{u}_j, \mathbf{u}_i \right\rangle = \lambda_i \langle \mathbf{u}_i, \mathbf{u}_i\rangle = \lambda_i.$$

Thus, each coefficient of the linear combination results from a simple inner product.

Remark: Outlook

It is this principle the so called Fourier-Transformation is built on. It decomposes a signal $\mathbf{v}(t)$ into frequencies $\mathbf{u}_i(t) = \sin(\omega_i t)$. This is, however, a problem formulated in a more abstract vector space.

vector space V with $\dim(V) = \infty$

5.4 Orthogonal matrices

Let us now restrict our attention to the standard inner product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle_{euklid} = \mathbf{x}^T \mathbf{y},$$

and write down our results from above in terms of matrices.

Let $\mathcal{B} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ a basis for $\mathbb{R}^n$. Then each $\mathbf{x} \in \mathbb{R}^n$ can be uniquely written as:

$$\mathbf{x} = \alpha_1 \begin{pmatrix} | \\ \mathbf{u}_1 \\ | \end{pmatrix} + \ldots + \alpha_n \begin{pmatrix} | \\ \mathbf{u}_n \\ | \end{pmatrix} = \underbrace{\begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_n \\ | & & | \end{pmatrix}}_{=:A} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

For the so-defined matrix $A = (\mathbf{u}_1 \cdots \mathbf{u}_n)$, we get:

$$\begin{aligned} A^T A &= \begin{pmatrix} — \mathbf{u}_1{}^T — \\ \vdots \\ — \mathbf{u}_n{}^T — \end{pmatrix} \begin{pmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_n \\ | & & | \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1{}^T \mathbf{u}_1 & \cdots & \mathbf{u}_1{}^T \mathbf{u}_n \\ \vdots & & \vdots \\ \mathbf{u}_n{}^T \mathbf{u}_1 & \cdots & \mathbf{u}_n{}^T \mathbf{u}_n \end{pmatrix} \\ &= \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{u}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{u}_1 \rangle \\ \vdots & & \vdots \\ \langle \mathbf{u}_1, \mathbf{u}_n \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{u}_n \rangle \end{pmatrix} = G(\mathcal{B}), \end{aligned} \tag{5.7}$$

$\Rightarrow$ $\mathcal{B}$ is ONB $\Leftrightarrow$ $A^T A = \mathbb{1}$

special name

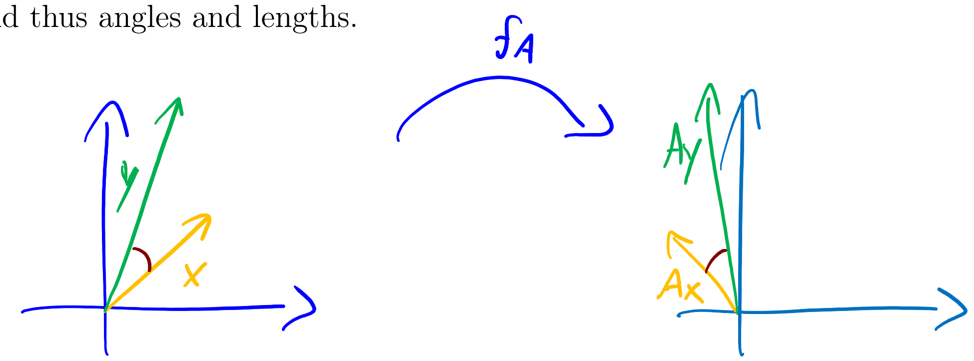
Definition 5.26. Orthogonal matrix

A matrix $A \in \mathbb{R}^{n \times n}$ with the property $A^T A = \mathbb{1}$ is called orthogonal.

We immediately see that an orthogonal matrix A has an ONB as columns and fulfils

$$\langle A\mathbf{x}, A\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle.$$

The last property says that the corresponding linear map f_A preserves the inner product, and thus angles and lengths.



Proposition 5.27. Defining properties of orthogonal matrices

For a matrix $A \in \mathbb{R}^{n \times n}$ the following claims are equivalent:

(a) *A is an orthogonal matrix*

(b) $A^T A = \mathbb{1}$.

(c) $AA^T = \mathbb{1}$. $\longrightarrow$ $(A^T)^T A^T = \mathbb{1}$

(d) $A^{-1} = A^T$.

(e) *A^T is an orthogonal matrix.*

(f) *The columns of A define an ONB of $\mathbb{R}^n$.*

(g) *The rows of A define an ONB of $\mathbb{R}^n$.*

(h) *For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we get $\langle A\mathbf{x}, A\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$.*

(i) *For all $\mathbf{x} \in \mathbb{R}^n$, we get $\|A\mathbf{x}\| = \|\mathbf{x}\|$.*

Polarisation:

$$\langle x, y \rangle = \frac{1}{4}\left(\langle x+y, x+y \rangle - \langle x-y, x-y \rangle\right)$$

Proof. Exercise! □

Such matrices correspond to maps of special geometric interest:

- Rotations
- Reflections
- Special case: permutation matrices

Standard ONB is different order

We also see that solving a LES $A\mathbf{x} = \mathbf{b}$ described by an orthogonal matrix A is easy to solve:

$$\mathbf{x} = A^{-1}\mathbf{b} = A^T\mathbf{b}$$

The inverse is computed now more easily than in the general case.

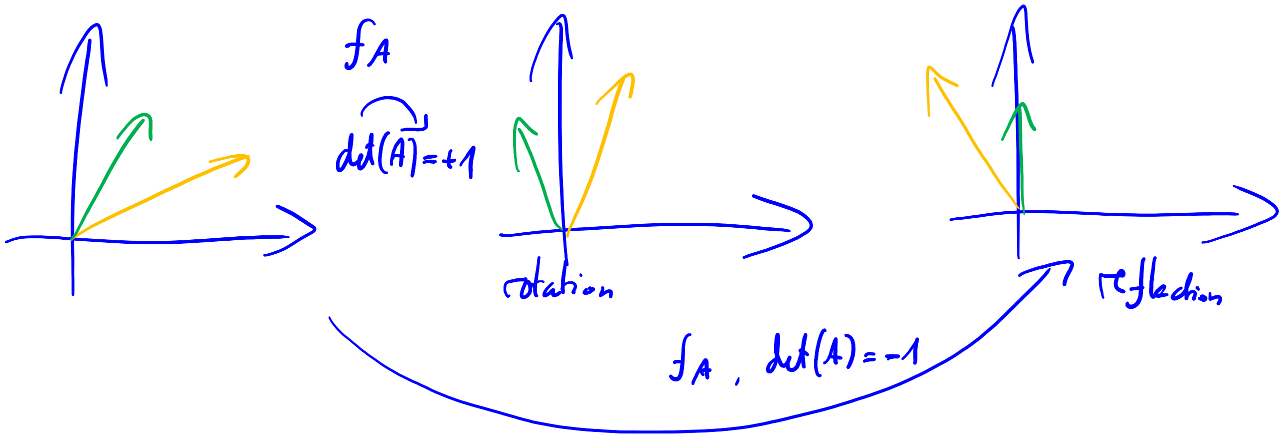
Proposition 5.28. Determinant of orthogonal matrices

For an orthogonal matrix A, we have $\det(A) = \pm 1$.

Proof. $1 = \det(\mathbb{1}) = \det(A^T A) = \underbrace{\det(A^T)}_{= \det(A)} \det(A) = \det(A)^2$. □

Definition 5.29. Rotations and reflections

Let $A \in \mathbb{R}^{n\times n}$ be an orthogonal matrix. If $\det(A) = 1$, *we call A a* rotation. *If* $\det(A) = -1$, *we call the matrix a* reflection.

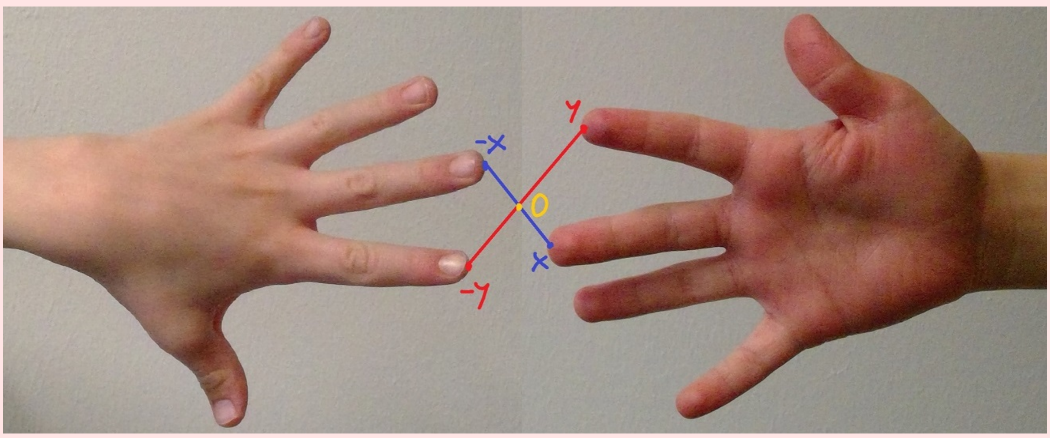


Attention! Notions: Rotation or reflection

(a) *Not every matrix $A \in \mathbb{R}^{n\times n}$ with* $\det(A) = 1$ *(or* $\det(A) = -1$*) is a rotation (or a reflection)!*

$$\det\begin{pmatrix} 1 & 800 \\ 0 & 1 \end{pmatrix} = 1, \quad \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 800 \\ 1 \end{pmatrix} \text{ not orthonormal}$$

(b) *A "reflection" from Definition 5.29 could also be a point reflection in the case* $n \geq 3$.



5.5 Orthogonalisation: the QR-decomposition $\langle \cdot , \cdot \rangle = \langle \cdot , \cdot \rangle_{\text{euklid}}$

$A \in \mathbb{R}^{n \times n} \rightsquigarrow A$ full rank $\iff$ columns form a basis of $\mathbb{R}^n$

basis in $\mathbb{R}^n \xrightarrow{\text{Gram Schmidt}}$ ONB of $\mathbb{R}^n$

QR-decomp. = Gram Schmidt for matrices

$$\underbrace{\begin{pmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{pmatrix}}_{\text{basis}} \rightsquigarrow \underbrace{\begin{pmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{pmatrix}}_{\text{ONB}}, \quad \begin{aligned} a_1 &= \|a_1\| \cdot q_1 \\ a_2 &= ?\, q_1 + ?\, q_2 \\ &\vdots \\ a_n &= \cdots \end{aligned}$$

orthogonal

$$A = Q \cdot R$$

upper triangular

$$\begin{pmatrix} (*) & \cdots \\ 0 & \\ \vdots & \\ 0 & \end{pmatrix}$$

There are at least three alternatives to compute this:

Here ⟸

- "Classical Gram-Schmidt": this is what we learn next (good for pen-and-paper computations), but instable on the computer
- "Modified Gram-Schmidt": equivalent to our Gram-Schmidt, order of loops exchanged, numerically more stable
- "Householder reflections": are cheaper and even more stable. This is the method of choice in numerical computations

If A is square matrix with $\operatorname{rank}(A) = n$, we get

$$A = \begin{pmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{pmatrix}$$

$$= \left(\underbrace{Q \begin{pmatrix} r_{11} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\mathbf{a}_1} \underbrace{Q \begin{pmatrix} r_{12} \\ r_{22} \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\mathbf{a}_2} \cdots \underbrace{Q \begin{pmatrix} r_{1n} \\ r_{2n} \\ r_{3n} \\ \vdots \\ r_{nn} \end{pmatrix}}_{\mathbf{a}_n} \right) = Q \underbrace{\begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ & r_{22} & r_{23} & \cdots & r_{2n} \\ & & r_{33} & \cdots & r_{3n} \\ & & & \ddots & \vdots \\ & & & & r_{nn} \end{pmatrix}}_{=:R} = QR.$$

This defines the so-called *QR-decomposition* of a matrix A.

As a result, we get $(\mathbf{q}_1, \ldots, \mathbf{q}_n)$ as an ONB for the space $\operatorname{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \operatorname{Ran}(A)$. We immediately get $Q^T Q = \mathbb{1}_n$.

$r_{11} = \|a_1\|$, $q_1 = \dfrac{a_1}{\|a_1\|}$, $a_1 = \overbrace{\|a_1\|}^{r_{11}} \cdot q_1$

$$\frac{a_2 - \langle a_2, q_1 \rangle q_1}{\|a_2 - \langle a_2, q_1 \rangle q_1\|} = q_2$$

$$a_2 = \overbrace{\|(*)\|}^{r_{22}} \cdot q_2 + \overbrace{\langle a_2, q_1 \rangle}^{r_{12}} q_1$$

Example 5.30. Consider

$$A = \begin{pmatrix} 2 & -1 & 8 \\ 1 & 1 & 1 \\ -2 & 4 & 4 \end{pmatrix}$$

$$\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} = \frac{\mathbf{a}_1}{3} = \begin{pmatrix} 2/3 \\ 1/3 \\ -2/3 \end{pmatrix}, \qquad r_{11} = 3$$

$$r_{12} = \langle \mathbf{a}_2, \mathbf{q}_1 \rangle = -3, \qquad \mathbf{q}_2 = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}, \qquad r_{22} = 3$$

$$r_{13} = \langle \mathbf{a}_3, \mathbf{q}_1 \rangle = 3, \quad r_{13} = \langle \mathbf{a}_3, \mathbf{q}_1 \rangle = 6, \quad \mathbf{q}_3 = \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix}, \quad r_{33} = 6;$$

$$Q = \begin{pmatrix} 2/3 & 1/3 & 2/3 \\ 1/3 & 2/3 & -2/3 \\ -2/3 & 2/3 & 1/3 \end{pmatrix} \qquad R = \begin{pmatrix} 3 & -3 & 3 \\ & 3 & 6 \\ & & 6 \end{pmatrix}$$

$A = Q \cdot R$

Example 5.31.

$$\text{For} \quad A = \begin{pmatrix} 2 & -1 & 8 \\ 1 & 1 & 1 \\ -2 & 4 & 4 \end{pmatrix} \quad \text{Gram-Schmidt gives us} \quad \mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} = \begin{pmatrix} 2/3 \\ 1/3 \\ -2/3 \end{pmatrix},$$

$$\mathbf{q}_2 = \frac{\mathbf{a}_2 - (\mathbf{a}_2)_{|\mathrm{Span}(\mathbf{q}_1)}}{\|\dots\|} = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}, \quad \mathbf{q}_3 = \frac{\mathbf{a}_3 - (\mathbf{a}_3)_{|\mathrm{Span}(\mathbf{q}_1,\mathbf{q}_2)}}{\|\dots\|} = \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix}$$

$$\text{Hence:} \quad Q = \frac{1}{3}\begin{pmatrix} 2 & 1 & 2 \\ 1 & 2 & -2 \\ -2 & 2 & 1 \end{pmatrix} \quad \text{and} \quad R = Q^T A = \begin{pmatrix} 3 & -3 & 3 \\ & 3 & 6 \\ & & 6 \end{pmatrix}.$$

As we have seen in the LR-decomposition, we can also use the QR-decomposition for solving an LES $A\mathbf{x} = \mathbf{b}$. If A is a square matrix ($m = n$), we know:

$$A\mathbf{x} = \mathbf{b} \quad \Longleftrightarrow \quad QR\mathbf{x} = \mathbf{b} \quad \overset{Q^{-1}=Q^T}{\Longleftrightarrow} \quad R\mathbf{x} = Q^T\mathbf{b} \qquad (5.8)$$

The last system has a triangle form and is solved by backwards substitution. A QR-decomposition is also possible in the non-square case as we will see later in detail.

5.6 Distances: points, lines and planes (for reading at home)

Recall that we call an affine subspace H in $\mathbb{R}^n$ with dimension $n-1$ a hyperplane. This is, for example, a line in $\mathbb{R}^2$ or a plane in $\mathbb{R}^3$.

Definition 5.32. Hesse normal form (HNF), distance $\mathrm{dist}(\cdot,\cdot)$

For each hyperplane in $\mathbb{R}^n$, there exists a normal form

$$\{\mathbf{v} \in \mathbb{R}^n : \langle \mathbf{n}, \mathbf{v} - \mathbf{p}\rangle = 0\}$$

where $\mathbf{p} \in \mathbb{R}^n$ is one chosen point and $\mathbf{n} \in \mathbb{R}^n$ a normal vector. We call it Hesse normal form (HNF) if $\|\mathbf{n}\| = 1$ holds.

For a given point $\mathbf{q} \in \mathbb{R}^n$ and affine subspaces S, T in $\mathbb{R}^n$, we write:

$$\mathrm{dist}(\mathbf{q}, T) := \min_{\mathbf{t}\in T} \|\mathbf{q} - \mathbf{t}\| \quad \textit{and} \quad \mathrm{dist}(S, T) := \min_{\mathbf{s}\in S} \mathrm{dist}(\mathbf{s}, T) = \min_{\mathbf{s}\in S}\min_{\mathbf{t}\in T} \|\mathbf{s} - \mathbf{t}\|$$

for the shortest distance between $\mathbf{v}$ and T and the shortest distance between S and T, respectively.

If we are using the HNF for a hyperplane, then the expression $\langle \mathbf{n}, \mathbf{v} - \mathbf{p} \rangle$ can indeed measure the distances:

Proposition 5.33.

For a hyperplane $T = \{\mathbf{v} \in \mathbb{R}^n : \langle \mathbf{n}, \mathbf{v} - \mathbf{p} \rangle = 0\}$ *with* $\|\mathbf{n}\| = 1$ *(this is the HNF), we have*

$$\langle \mathbf{n}, \mathbf{q} - \mathbf{p} \rangle \quad = \quad \pm \mathrm{dist}(\mathbf{q}, T) \tag{5.9}$$

where the sign "+" holds if $\mathbf{q}$ *lies on the same side of* T *as the normal vector* $\mathbf{n}$*, and "−" holds if* $\mathbf{q}$ *lies on the other side of* T*.*

Proof. This is an exercise where you should use

$$\langle \mathbf{n}, \mathbf{v} - \mathbf{p} \rangle = \frac{\langle \mathbf{n}, \mathbf{v} - \mathbf{p} \rangle}{1} = \frac{\langle \mathbf{v} - \mathbf{p}, \mathbf{n} \rangle}{\langle \mathbf{n}, \mathbf{n} \rangle}$$

and use projections. □

Distances in $\mathbb{R}^3$

- *Point/Point:* $\operatorname{dist}(\mathbf{p},\mathbf{q}) = \|\mathbf{p}-\mathbf{q}\|$, *(for completeness's sake),*
- *Point/Plane:* $\operatorname{dist}(\mathbf{q}, \underbrace{\mathbf{p}+\operatorname{Span}(\mathbf{a},\mathbf{b})}_{E}) = |\langle \mathbf{n}, \mathbf{q}-\mathbf{p}\rangle|$, *cf.* (5.9).
- *Line/Plane:*

$$\operatorname{dist}(\underbrace{\mathbf{p}+\operatorname{Span}(\mathbf{a})}_{g}, \underbrace{\mathbf{q}+\operatorname{Span}(\mathbf{b},\mathbf{c})}_{E}) = \operatorname{dist}(\mathbf{p}, E),$$

  if g is parallel with respect to E. In the other case, g and E intersect, and, hence, $\operatorname{dist}(g,E) = 0$. *If g is parallel with respect to E, one has* $\mathbf{a}$ *in* $\operatorname{Span}(\mathbf{b},\mathbf{c})$*: the family* $(\mathbf{a},\mathbf{b},\mathbf{c})$ *is linearly dependent, i.e.* $\det(\mathbf{a}\ \mathbf{b}\ \mathbf{c}) = 0$.
- *Plane/Plane:*

$$\operatorname{dist}(\underbrace{\mathbf{p}+\operatorname{Span}(\mathbf{a},\mathbf{b})}_{E_1}, \underbrace{\mathbf{q}+\operatorname{Span}(\mathbf{c},\mathbf{d})}_{E_2}) = \operatorname{dist}(\mathbf{p}, E_2),$$

  if E_1 is parallel to E_2. Otherwise, $\operatorname{dist}(E_1,E_2) = 0$, *which means that E_1 and E_2 have an intersection. If E_1 and E_2 are parallel, then the normal vectors* $\mathbf{n}_1 := \mathbf{a}\times\mathbf{b}$ *and* $\mathbf{n}_2 := \mathbf{c}\times\mathbf{d}$ *of E_1 and E_2, respectively, are linearly dependent.*

- *Line/Line: Let* $g_1 = \mathbf{p}+\operatorname{Span}(\mathbf{a})$ *and* $g_2 = \mathbf{q}+\operatorname{Span}(\mathbf{b})$ *be lines in* $\mathbb{R}^3$.
  - *1st case: If* $\mathbf{a}$ *and* $\mathbf{b}$ *are parallel, then g_1 and g_2 are parallel.*
  - *2nd case: If the vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *are not parallel, then*

$$\operatorname{dist}(g_1,g_2) = \operatorname{dist}(\mathbf{p}, \underbrace{\mathbf{q}+\operatorname{Span}(\mathbf{a},\mathbf{b})}_{=:E}).$$

- *Point/Line: Let* $\mathbf{p}$ *be a point in* $\mathbb{R}^3$ *and* $g = \mathbf{q}+\operatorname{Span}(\mathbf{a})$ *a line in* $\mathbb{R}^3$. *Define* $\mathbf{b} := (\mathbf{p}-\mathbf{q})\times\mathbf{a}$. *If* $\mathbf{b} = \mathbf{o}$, *then* $\mathbf{p}$ *lies in g, and, hence,* $\operatorname{dist}(\mathbf{p},g) = 0$. *On the other hand,* $\mathbf{b}$ *can be perpendicular to the plane, defined by* $\mathbf{p}$ *and g. In this case:*

$$\operatorname{dist}(\mathbf{p},g) = \operatorname{dist}(\mathbf{p}, \underbrace{\mathbf{q}+\operatorname{Span}(\mathbf{a},\mathbf{b})}_{=:E}).$$

  Alternatively: Norm of the normal component of $\mathbf{p}-\mathbf{q}$ *with respect to* $\operatorname{Span}(\mathbf{a})$ *(Proposition 5.7).*

Summary

- Each vector $\mathbf{x} \in \mathbb{R}^n$ can be uniquely decomposed into
  - a vector $\mathbf{x}_{|U}$ in a given subspace U and
  - a vector $\mathbf{n}$ that is orthogonal to U.

  The vector $\mathbf{x}_{|U}$ is called the orthogonal projection of $\mathbf{x}$ onto U. $\mathbf{n}$ is equal to $\mathbf{x}_{|U^\perp}$.
- If $\dim(U) = 1$, the calculation of $\mathbf{x}_{|U}$ is very easy, while one can use Proposition 5.7; If $\dim(U) \geq 2$, the one has to choose a basis $\mathcal{B}$ for U and either
  - solve an LES with the help of the Gramian matrix $G(\mathcal{B})$ (Proposition 5.16) or
  - build an ONS or ONB with the help of the Gram-Schmidt procedure and use Proposition 5.23.
- A matrix $A \in \mathbb{R}^{n \times n}$ with $A^{-1} = A^T$ is called orthogonal. The determinant is ± 1. Depending on the sign of $\det(A)$, the matrix A describes a reflection or a rotation.