# Optimization and Computational Linear Algebra for Data Science
## Lecture 5: Matrices and orthogonality

Léo MIOLANE · leo.miolane@gmail.com

May 14, 2019

**Warning:** *This material is not meant to be lecture notes. It only gathers the main concepts and results from the lecture, without any additional explanation, motivation, examples, figures...*

# 1 Gram-Schmidt orthogonalisation method

The Gram-Schmidt process takes as input a *linearly independent* family $(x_1, \ldots, x_k)$ of vectors of $\mathbb{R}^n$ and produces as output an *orthonormal basis* $(v_1, \ldots, v_k)$ of $\mathrm{Span}(x_1, \ldots, x_k)$.

In particular in the case $k = n$ where $(x_1, \ldots, x_n)$ is a basis of $\mathbb{R}^n$, it gives us a way to construct an orthonormal basis $(v_1, \ldots, v_n)$ of $\mathbb{R}^n$ from any basis.

The Gram-Schmidt process is iterative and constructs progessively families $(v_1, \ldots, v_i)$ that all verify:

$$\mathcal{H}_i : \ (v_1, \ldots, v_i) \text{ is an orthonormal family and } \mathrm{Span}(v_1, \ldots, v_i) = \mathrm{Span}(x_1, \ldots, x_i).$$

**Construction of $v_1$:** We simply take $v_1 = x_1/\|x_1\|$ and $\mathcal{H}_1$ is obviously verified.

**Construction of $v_{i+1}$ from $(v_1, \ldots, v_i)$:** Suppose that we already constructed $(v_1, \ldots, v_i)$ that verifies $\mathcal{H}_i$. We first substract to $x_{i+1}$ its orthogonal projection on $\mathrm{Span}(v_1, \ldots, v_i)$:

$$\widetilde{v}_{i+1} \overset{\mathrm{def}}{=} x_{i+1} - P_{\mathrm{Span}(v_1, \ldots, v_i)}(x_{i+1}) = x_{i+1} - \sum_{j=1}^{i} \langle v_j, x_{i+1} \rangle v_j. \tag{1}$$

Then we normalize $\widetilde{v}_{i+1}$ to obtain $v_{i+1}$:

$$v_{i+1} = \frac{\widetilde{v}_{i+1}}{\|\widetilde{v}_{i+1}\|}. \tag{2}$$

Let us now verify that $\mathcal{H}_{i+1}$ holds. The fact that $\mathrm{Span}(v_1, \ldots, v_{i+1}) = \mathrm{Span}(x_1, \ldots, x_{i+1})$ can be easily checked using $\mathcal{H}_i$ and equations (1)-(2). It remains thus to prove that $(v_1, \ldots, v_{i+1})$ is orthonormal.

By construction $v_{i+1} = 1$. Then, by Proposition 4.1 from Lecture 4, we have $\widetilde{v}_{i+1} = x_{i+1} - P_{\mathrm{Span}(v_1, \ldots, v_i)}(x_{i+1}) \in \mathrm{Span}(v_1, \ldots, v_i)^{\perp}$. Consequently $v_{i+1}$ is orthogonal to $\mathrm{Span}(v_1, \ldots, v_i)$ and therefore to the vectors $v_1, \ldots, v_i$. This proves that $(v_1, \ldots, v_{i+1})$ is orthonormal.

We conclude:

**Theorem 1.1 (*Gram-Schmidt*)**

*Let $(x_1, \ldots x_k)$ be a linearly independent family of vectors of $\mathbb{R}^n$. The "Gram-Schmidt" procedure described above produces an orthonormal family $(v_1, \ldots, v_k)$ such that for all $i \in \{1, \ldots, k\}$,*
$$\mathrm{Span}(x_1, \ldots, x_i) = \mathrm{Span}(v_1, \ldots, v_i).$$

# 2 Orthogonal matrices

**Definition 2.1 (*Orthogonal matrices*)**

*A matrix $A \in \mathbb{R}^{n \times n}$ is called an* orthogonal matrix *if its columns are an orthonormal family (and therefore a basis of $\mathbb{R}^n$ because it is a linearly independent family of size $n = \dim(\mathbb{R}^n)$).*

**Proposition 2.1**

*Let $A \in \mathbb{R}^{n \times n}$. The following points are equivalent:*

*(i) $A$ is orthogonal.*

*(ii) $A^\mathsf{T} A = \mathrm{Id}_n$.*

*(iii) $A A^\mathsf{T} = \mathrm{Id}_n$*

In particular we get that the inverse of an orthogonal matrix $A$ is $A^\mathsf{T}$. We also get that if $A$ is orthogonal then so is $A^\mathsf{T}$: the lines of an orthogonal matrix are an orthonormal family of vectors.

**Proof.** We first show that (i) $\Leftrightarrow$ (ii). We denote the columns of $A$ by $c_1, \ldots, c_n$. For $i, j \in \{1, \ldots, n\}$ we have

$$(A^\mathsf{T} A)_{i,j} = \langle c_i, c_j \rangle.$$

Consequently, $A^\mathsf{T} A = \mathrm{Id}_n$ if and only if $(c_1, \ldots, c_n)$ is orthonormal. Now, by Proposition 2.4 from Lecture 2 we have

$$A^\mathsf{T} A = \mathrm{Id}_n \;\Leftrightarrow\; A \text{ is invertible with inverse } A^\mathsf{T} \;\Leftrightarrow\; A A^\mathsf{T} = \mathrm{Id}_n,$$

which concludes the proof. $\qquad\square$

**Example 2.1** (Rotation matrices in dimention 2)**.** *For $\theta \in \mathbb{R}$, the matrix*

$$R_\theta \;\overset{\mathrm{def}}{=}\; \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

*is orthogonal. The linear transformation $x \in \mathbb{R}^2 \mapsto R_\theta x$ is the rotation of center $0$ and angle $\theta$.*

**Proposition 2.2 (*Orthogonal matrices preserve the dot product*)**

*Let $A \in \mathbb{R}^{n \times n}$ be an orthogonal matrix. Then $A$ preserves the dot product in the sense that for all $x, y \in \mathbb{R}^n$,*

$$\langle Ax, Ay \rangle = \langle x, y \rangle.$$

*In particular if we take $x = y$ we see that $A$ preserves the Euclidean norm: $\|Ax\| = \|x\|$.*

**Proof.** By Proposition 2.1 $A^\mathsf{T} A = \mathrm{Id}_n$, hence

$$\langle Ax, Ay \rangle = (Ax)^\mathsf{T} Ay = x^\mathsf{T} A^\mathsf{T} Ay = x^\mathsf{T} \mathrm{Id}_n y = x^\mathsf{T} y = \langle x, y \rangle.$$

$\qquad\square$