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

Léo Miolane · leo.miolane@gmail.com

September 29, 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 orthogonalization method

The Gram-Schmidt process takes as input a *linearly independent* family $(x_1, \dots, x_k)$ of vectors of $\mathbb{R}^n$ and produces as output an *orthonormal basis* $(v_1, \dots, v_k)$ of $\mathrm{Span}(x_1, \dots, x_k)$. In particular in the case where $(x_1, \dots, x_k)$ is a basis of a subspace S, it gives us a way to construct an orthonormal basis $(v_1, \dots, v_k)$ of S.

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

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

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 $(v_1, \dots, v_i)$: Suppose that we already constructed $(v_1, \dots, v_i)$ that verifies $\mathcal{H}_i$. We first subtract to x_{i+1} its orthogonal projection on $\mathrm{Span}(v_1, \dots, v_i)$:

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

By Corollary 4.1 from Lecture 4, we have $\widetilde{v}_{i+1} = x_{i+1} - P_{\mathrm{Span}(v_1,\dots,v_i)}(x_{i+1}) \perp \mathrm{Span}(v_1, \dots, v_i)$. Consequently $\widetilde{v}_{i+1}$ is orthogonal to the vectors $v_1, \dots, v_i$. It remains to normalize $\widetilde{v}_{i+1}$ to obtain an orthonormal family:

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

Notice that $\widetilde{v}_{i+1} \neq 0$ because otherwise we would have that $x_{i+1} = P_{\mathrm{Span}(v_1,\dots,v_i)}(x_{i+1}) \in \mathrm{Span}(v_1, \dots, v_i) = \mathrm{Span}(x_1, \dots, x_i)$, which contradicts the linear independence of $(x_1, \dots, x_{i+1})$.

We have seen that $(v_1, \dots, v_{i+1})$ is orthonormal. The fact that $\mathrm{Span}(v_1, \dots, v_{i+1}) = \mathrm{Span}(x_1, \dots, x_{i+1})$ can be easily checked using $\mathcal{H}_i$ and equations (1)-(2). We conclude that $(v_1, \dots, v_{i+1})$ verifies $\mathcal{H}_{i+1}$.

Theorem 1.1 (*Gram-Schmidt*)

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

$$\mathrm{Span}(x_1, \dots, x_i) = \mathrm{Span}(v_1, \dots, v_i).$$

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) $A^\top A = \mathrm{Id}_n$.

(iii) $AA^\top = \mathrm{Id}_n$

In particular we get that the inverse of an orthogonal matrix A is $A^\top$. We also get that if A is orthogonal then so is $A^\top$: 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 $c_1, \dots, c_n$. For $i, j \in \{1, \dots, n\}$ we have

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

Consequently, $A^\top A = \mathrm{Id}_n$ if and only if $(c_1, \dots, c_n)$ is orthonormal. Now, by Proposition 2.2 from Lecture 2 we have

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

which concludes the proof. □

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

$$R_\theta \stackrel{\text{def}}{=} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

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

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 $A^\top A = \mathrm{Id}_n$, hence

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

□