Optimization and Computational Linear Algebra for Data Science
Lecture 3: Rank

Léo Miolane · leo.miolane@gmail.com

September 16, 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 Definition of the rank

Definition 1.1 (*Rank of a family of vectors*)

We define the rank of a family $x_1, \dots, x_k$ of vectors of $\mathbb{R}^n$ as the dimension of its span:

$$\mathrm{rank}(x_1, \dots, x_k) \stackrel{\text{def}}{=} \dim(\mathrm{Span}(x_1, \dots, x_k)).$$

If the vectors $x_1, \dots x_k$ are linearly independent then $\mathrm{rank}(x_1, \dots x_k) = k$. Indeed, in that case $(x_1, \dots, x_k)$ forms a basis of $\mathrm{Span}(x_1, \dots, x_k)$ so $\dim(\mathrm{Span}(x_1, \dots, x_k)) = k$.

Definition 1.2 (*Rank of a matrix*)

Let $M \in \mathbb{R}^{n \times m}$. Let $c_1, \dots, c_m \in \mathbb{R}^n$ be its columns. We define

$$\mathrm{rank}(M) \stackrel{\text{def}}{=} \mathrm{rank}(c_1, \dots, c_m) = \dim(\mathrm{Im}(M)). \tag{1}$$

Proposition 1.1

Let $M \in \mathbb{R}^{n \times m}$. Let $r_1, \dots, r_n \in \mathbb{R}^m$ be the rows of M and $c_1, \dots, c_m \in \mathbb{R}^n$ be its columns. Then we have

$$\mathrm{rank}(r_1, \dots, r_n) = \mathrm{rank}(c_1, \dots, c_m) = \mathrm{rank}(M). \tag{2}$$

Proof. In order to prove (2) it suffices to show (since columns and rows are playing exchangeable roles) that $\mathrm{rank}(\ell_1, \dots, \ell_n) \leq \mathrm{rank}(c_1, \dots, c_m)$. Let $r \stackrel{\text{def}}{=} \mathrm{rank}(\ell_1, \dots, \ell_n)$ and $(x_1, \dots, x_r)$ be a basis of $\mathrm{Span}(\ell_1, \dots, \ell_n)$. We will prove that

$$(Mx_1, \dots, Mx_r) \text{ is linearly independent.} \tag{3}$$

The result follows. Indeed $(Mx_1, \dots, Mx_r)$ is then a linearly independent family of r vectors of $\mathrm{Im}(M) = \mathrm{Span}(c_1, \dots, c_m)$: this implies that $\mathrm{rank}(c_1, \dots, c_m) = \dim(\mathrm{Span}(c_1, \dots, c_m)) \geq r = \mathrm{rank}(\ell_1, \dots, \ell_n)$.

It remains to prove (3). Let $\alpha_1, \dots, \alpha_r \in \mathbb{R}$ such that $\alpha_1 M x_1 + \dots + \alpha_r M x_r = 0$. We will show that in such case the α_i are all zero. Define $v \stackrel{\text{def}}{=} \alpha_1 x_1 + \dots + \alpha_r x_r$. We have by linearity

$$Mv = M(\alpha_1 x_1 + \dots + \alpha_r x_r) = \alpha_1 M x_1 + \dots + \alpha_r M x_r = 0.$$

Since the i^{th} coordinate of Mv is equal to $(Mv)_i = \ell_i \cdot v$, we get that v is orthogonal to all the ℓ_i, and therefore to $\mathrm{Span}(\ell_1, \dots, \ell_n)$. Notice now that $v \in \mathrm{Span}(x_1, \dots, x_r) = \mathrm{Span}(\ell_1, \dots, \ell_n)$ by construction. The vector v is orthogonal to itself, hence $\alpha_1 x_1 + \dots + \alpha_r x_r = v = 0$. Recall that the family $(x_1, \dots, x_r)$ is linearly independent (because it is a basis) so $\alpha_1 = \dots = \alpha_r = 0$. $\square$

Remark 1.1. *For $v_1, \dots, v_k \in \mathbb{R}^n$, and $\alpha \in \mathbb{R} \setminus \{0\}$, $\beta \in \mathbb{R}$ one can easily verify that*

$$\begin{aligned}\operatorname{rank}(v_1, \dots, v_k) &= \operatorname{rank}(v_1, \dots, v_{i-1}, \, \alpha v_i \, , v_{i+1}, \dots, v_k) \\ &= \operatorname{rank}(v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_{j-1}, \, v_j + \beta v_i \, , v_{j+1}, \dots, v_k).\end{aligned}$$

As a consequence, the Gaussian elimination method keeps the rank of a matrix unchanged!

2 Properties of the rank

Theorem 2.1 (*Rank-nullity theorem*)

Let $L : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Then

$$\operatorname{rank}(L) + \dim(\operatorname{Ker}(L)) = m.$$

Theorem 2.1 is proved at the end of these notes.

Proposition 2.1

Let $A \in \mathbb{R}^{n \times m}$ and $B \in \mathbb{R}^{m \times k}$. Then the following holds

(i) $\operatorname{rank}(A) \leq \min(n, m)$.

(ii) $\operatorname{rank}(AB) \leq \min(\operatorname{rank}(A), \operatorname{rank}(B))$.

Exercise 2.1 (Important)**.** *Let $M \in \mathbb{R}^{n \times m}$ and $r = \operatorname{rank}(M)$. Show that there exist $A \in \mathbb{R}^{n \times r}$ and $B \in \mathbb{R}^{r \times m}$ such that $M = AB$.*

3 Invertible matrices

Definition 3.1 (*Matrix inverse*)

A matrix $M \in \mathbb{R}^{n \times n}$ is called invertible *if there exists a matrix $M^{-1} \in \mathbb{R}^{n \times n}$ such that*

$$MM^{-1} = M^{-1}M = \mathrm{Id}_n.$$

Such matrix M^{-1} is unique and is called the inverse *of M.*

Remark 3.1. *$M \in \mathbb{R}^{n \times n}$ is invertible if and only if the linear transformation associated to M is a bijection. In that case, M^{-1} is the matrix associated to the inverse transformation.*

Theorem 3.1

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

(i) *M is invertible.*

(ii) $\operatorname{rank}(M) = n$.

(iii) $\operatorname{Ker}(M) = \{0\}$.

Proof. Points (ii) and (iii) are equivalent by Theorem 2.1. It remains to prove that (i) $\Leftrightarrow$ [(ii)-(iii)].

We start by proving that (i) implies (iii). Assume that M is invertible and consider $x \in \mathrm{Ker}(M)$. Since $M^{-1}M = \mathrm{Id}_n$, we have $M^{-1}Mx = x$, which leads to $0 = x$ because $Mx = 0$. Hence $\mathrm{Ker}(M) = \{0\}$.

Conversely assume that [(ii)-(iii)] hold. Then, as seen at the end of Lecture 2, for all $y \in \mathbb{R}^n$ there exists a unique $x^{(y)} \in \mathbb{R}^n$ such that $Mx^{(y)} = y$. One can verify easily that the map $L : y \mapsto x^{(y)}$ is linear and verifies (by construction) $L \circ M = M \circ L = \mathrm{Id}$. Consequently, the corresponding matrices verify: $LM = ML = \mathrm{Id}_n$. $\square$

Exercise 3.1. *Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times m}$. Show that if B is invertible then* $\mathrm{rank}(AB) = \mathrm{rank}(A)$. *Similarly for $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$, show that if A is invertible then* $\mathrm{rank}(AB) = \mathrm{rank}(B)$.

4 Transpose of a matrix, symmetric matrices

Definition 4.1 (*Transpose*)

Let $M \in \mathbb{R}^{n\times m}$. We define its transpose $M^\top \in \mathbb{R}^{m\times n}$ *by*

$$(M^\top)_{i,j} = M_{i,j}$$

for all $i \in \{1,\dots,m\}$ and $j \in \{1,\dots,n\}$.

Remark 4.1.

- *We have $(M^\top)^\top = M$.*
- *The mapping $M \mapsto M^\top$ is linear.*

We remark also that the rows of M become the columns of $M^\top$ and that the columns of M become the rows of $M^\top$. By Definition 1.2, this gives:

Proposition 4.1

$\mathrm{rank}(M) = \mathrm{rank}(M^\top)$.

Proposition 4.2

Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{m\times k}$. Then

$$(AB)^\top = B^\top A^\top.$$

Corollary 4.1

If $M \in \mathbb{R}^{n\times n}$ is invertible, then so is $M^\top$ and

$$(M^\top)^{-1} = (M^{-1})^\top.$$

Proof. We compute, using Proposition 4.2:

$$M^\top (M^{-1})^\top = (M^{-1}M)^\top = \mathrm{Id}_n^\top = \mathrm{Id}_n.$$

This proves that $M^\top$ is invertible with inverse $(M^{-1})^\top$. $\square$

Definition 4.2 (***Symmetric matrix***)

A square matrix $A \in \mathbb{R}^{n\times n}$ is said to be symmetric *if*

$$\forall i,j \in \{1,\dots,n\},\ A_{i,j} = A_{j,i}$$

or, equivalently if $A = A^\top$.

The following example is fundamental:

Example 4.1 (Gram matrices). Let $M \in \mathbb{R}^{k\times n}$. Then the $n \times n$ "Gram matrix" $A \stackrel{\text{def}}{=} M^\top M$ is symmetric.

Proof of Theorem 2.1

We will need the following result.

Proposition 4.3

Let V be a vector space of dimension n. Let $x_1,\dots,x_k \in V$. If $x_1,\dots,x_k$ are linearly independent then one can find vectors $x_{k+1},\dots,x_n \in V$ such that $(x_1,\dots,x_n)$ forms a basis of V.

Let us write $k = \dim(\mathrm{Ker}(L))$ and let us fix a basis $(x_1,\dots,x_k)$ of $\mathrm{Ker}(L)$. By Proposition 4.3 one can complete this family into a basis $(x_1,\dots,x_k,x_{k+1},\dots x_m)$ of $\mathbb{R}^m$. We will show that

(i) $\mathrm{Span}(L(x_{k+1}),\dots,L(x_m)) = \mathrm{Im}(L)$.

(ii) the family $(L(x_{k+1}),\dots,L(x_m))$ is linearly independent.

By proving (i) and (ii) we will get that $(L(x_{k+1}),\dots,L(x_m))$ is a basis of $\mathrm{Im}(L)$ which implies that

$$\mathrm{rank}(L) = \dim(\mathrm{Im}(L)) = m - k = m - \dim(\mathrm{Ker}(L)),$$

hence the result.

We start by proving (i). Since $L(x_{k+1}),\dots,L(x_m)$ are all in $\mathrm{Im}(L)$ (which is a linear subspace) any linear combination of these vectors belongs to $\mathrm{Im}(L)$, hence $\mathrm{Span}(L(x_{k+1}),\dots,L(x_m)) \subset \mathrm{Im}(L)$.

Let us prove the converse inclusion. Let $y \in \mathrm{Im}(L)$, which means that we can find $z \in \mathbb{R}^m$ such that $y = L(z)$. Let $(\alpha_1,\dots,\alpha_m) \in \mathbb{R}^m$ be the coordinates of z in the basis $(x_1,\dots,x_m)$: $z = \alpha_1 x_1 + \cdots + \alpha_m x_m$. We have then by linearity of L

$$y = L(z) = L(\alpha_1 x_1 + \cdots + \alpha_m x_m) = \alpha_1 L(x_1) + \cdots + \alpha_m L(x_m).$$

Recall now that $x_1,\dots,x_k$ belong to $\mathrm{Ker}(L)$. Therefore $L(x_1) = \cdots = L(x_k) = 0$. We get

$$y = \alpha_{k+1} L(x_{k+1}) + \cdots + \alpha_m L(x_m),$$

hence $y \in \mathrm{Span}(L(x_{k+1}),\dots,L(x_m))$: $\mathrm{Im}(L) \subset \mathrm{Span}(L(x_{k+1}),\dots,L(x_m))$. We conclude that $\mathrm{Im}(L) = \mathrm{Span}(L(x_{k+1}),\dots,L(x_m))$.

Let us now prove (ii). To prove that $(L(x_{k+1}),\dots,L(x_m))$ are linearly independent, we consider scalars $\alpha_{k+1},\dots,\alpha_m \in \mathbb{R}$ such that $\alpha_{k+1}L(x_{k+1}) + \cdots + \alpha_m L(x_m) = 0$. Our goal is to show that $\alpha_{k+1} = \cdots = \alpha_m = 0$. We have by linearity of L:

$$0 = \alpha_{k+1}L(x_{k+1}) + \cdots + \alpha_m L(x_m) = L(\alpha_{k+1}x_{k+1} + \cdots + \alpha_m x_m)$$

which gives that $\alpha_{k+1}x_{k+1} + \cdots + \alpha_m x_m \in \mathrm{Ker}(L)$. Recall that $(x_1, \ldots, x_k)$ is a basis of $\mathrm{Ker}(L)$, so there exists scalars $\alpha_1, \ldots, \alpha_k$ such that $\alpha_1 x_1 + \cdots + \alpha_k x_k = \alpha_{k+1}x_{k+1} + \cdots + \alpha_m x_m$. We obtain

$$\alpha_1 x_1 + \cdots + \alpha_k x_k - \alpha_{k+1}x_{k+1} - \cdots - \alpha_m x_m = 0$$

which implies that $\alpha_1 = \cdots = \alpha_m = 0$ because $(x_1, \ldots, x_m)$ is a basis of $\mathbb{R}^m$. This proves (ii). $\square$