Optimization and Computational Linear Algebra for Data Science
Lecture 2: Linear transformations

Léo Miolane · leo.miolane@gmail.com

September 10, 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 Linear transformations

Definition 1.1 (*Linear transformation*)

A function $L : \mathbb{R}^m \to \mathbb{R}^n$ is linear if

(i) for all $v \in \mathbb{R}^m$ and all $\alpha \in \mathbb{R}$ we have $L(\alpha v) = \alpha L(v)$ and

(ii) for all $v, w \in \mathbb{R}^m$ we have $L(v + w) = L(v) + L(w)$.

Notice that $L : \mathbb{R}^m \to \mathbb{R}^n$ is linear if and only if $L(\alpha v + w) = \alpha L(v) + L(w)$ for all $v, w \in \mathbb{R}^m$ and all $\alpha \in \mathbb{R}$.

Proposition 1.1

If $L : \mathbb{R}^m \to \mathbb{R}^n$ and $M : \mathbb{R}^n \to \mathbb{R}^k$ are two linear transformations, then the composite function $M \circ L : \mathbb{R}^m \to \mathbb{R}^k$ is also linear.

Remark 1.1. *Let $L : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. Then*

- $L(0) = L(0.0) = 0.L(0) = 0$.
- $L\big(\sum_{i=1}^k \alpha_i v_i\big) = \sum_{i=1}^k \alpha_i L(v_i)$.

Let $(e_1, \dots, e_m)$ be the canonical basis of $\mathbb{R}^m$. As a consequence of the second point above we have that for all $x = (x_1, \dots, x_m) \in \mathbb{R}^m$:

$$L(x) = L\Big(\sum_{i=1}^m x_i e_i\Big) = \sum_{i=1}^m x_i L(e_i).$$

This means that a linear transformation $L : \mathbb{R}^m \to \mathbb{R}^n$ is uniquely characterized by the vectors $L(e_1), \dots, L(e_m) \in \mathbb{R}^n$.

2 Matrix representation

Definition 2.1

A $n \times m$ matrix is an array (of real numbers) with n rows and m columns. We denote by $\mathbb{R}^{n \times m}$ the set of all $n \times m$ matrices.

We can encode a linear transformation $L : \mathbb{R}^m \to \mathbb{R}^n$ by a $n \times m$ matrix.

Definition 2.2

Let $L : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. The canonical matrix of L is the $n \times m$ matrix (that we will write also L) whose columns are $L(e_1), \dots, L(e_m)$:

$$L = \begin{pmatrix} | & | & & | \\ L(e_1) & L(e_2) & \cdots & L(e_m) \\ | & | & & | \end{pmatrix} = \begin{pmatrix} L_{1,1} & L_{1,2} & \cdots & L_{1,m} \\ L_{2,1} & L_{2,2} & \cdots & L_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ L_{n,1} & L_{n,2} & \cdots & L_{n,m} \end{pmatrix}$$

where we write $L(e_j) = \begin{pmatrix} L_{1,j} \\ L_{2,j} \\ \vdots \\ L_{n,j} \end{pmatrix}$.

Example 2.1 (Homothety). Let $\lambda \in \mathbb{R}$. The mapping (called "homothety of ratio λ")

$$\begin{array}{rccc} L: & \mathbb{R}^n & \to & \mathbb{R}^n \\ & x & \mapsto & \lambda x \end{array}$$

is linear. The canonical matrix of L is

$$\begin{pmatrix} \lambda & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 0 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}.$$

In the case where $\lambda = 1$, L is simply the identity, its matrix is called the identity matrix and denoted by

$$\mathrm{Id}_n \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Definition 2.2 tells us that one can associate a matrix to a linear transformation. The next definition show that one can do the reverse operation: we can obtain a linear transformation from a matrix.

Definition 2.3

The linear transformation associated to a matrix $L \in \mathbb{R}^{n \times m}$ is the map

$$\begin{array}{rccc} L: & \mathbb{R}^m & \to & \mathbb{R}^n \\ & x & \mapsto & Lx \end{array}$$

where the "matrix-vector" product $Lx \in \mathbb{R}^n$ is defined by

$$(Lx)_i = \sum_{j=1}^{m} L_{i,j} x_j \qquad \text{for all } i \in \{1, \dots, n\}.$$

Definition 2.4 (*Matrix product*)

Let $L \in \mathbb{R}^{n\times m}$ and $M \in \mathbb{R}^{m\times k}$. We define the matrix product LM as the $n \times k$ matrix of the linear transformation $L \circ M$. His coefficients are given by the formula:

$$(LM)_{i,j} = \sum_{\ell=1}^{m} L_{i,\ell} M_{\ell,j} \quad \text{for all} \quad 1 \le i \le n, \quad 1 \le j \le k.$$

Proposition 2.1

Let $A \in \mathbb{R}^{p\times q}$, $B \in \mathbb{R}^{q\times r}$ and $C \in \mathbb{R}^{r\times s}$. Then

$$(AB)C = A(BC).$$

Proposition 2.2 (*Matrix inverse*)

Let $M \in \mathbb{R}^{n\times n}$. Assume that there exists a matrix $M^{-1} \in \mathbb{R}^{n\times n}$ such that

$$MM^{-1} = \mathrm{Id}_n \quad \text{or, such that} \quad M^{-1}M = \mathrm{Id}_n.$$

Then $MM^{-1} = M^{-1}M = \mathrm{Id}_n$ and M^{-1} is the unique matrix that verifies this property. We say that M is invertible *and the matrix M^{-1} is called the* inverse *of M.*

Remark 2.1. *$M \in \mathbb{R}^{n\times n}$ 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.*

3 Kernel and image

Definition 3.1 (*Kernel*)

The kernel $\mathrm{Ker}(L)$ *(or nullspace) of a linear transformation $L : \mathbb{R}^m \to \mathbb{R}^n$ is defined as the set of all vectors $v \in \mathbb{R}^m$ such that $L(v) = 0$, i.e.*

$$\mathrm{Ker}(L) \stackrel{\text{def}}{=} \{v \in \mathbb{R}^m \,|\, L(v) = 0\}.$$

Definition 3.2 (*Image*)

The image $\mathrm{Im}(L)$ *(or column space) of a linear transformation $L : \mathbb{R}^m \to \mathbb{R}^n$ is defined as the set of all vectors $u \in \mathbb{R}^n$ such that there exists $v \in \mathbb{R}^m$ such that $L(v) = u$.* $\mathrm{Im}(L)$ *is also the Span of the columns of the matrix representation of L.*

Proposition 3.1

$\mathrm{Ker}(L)$ *and* $\mathrm{Im}(L)$ *are subspaces of respectively $\mathbb{R}^m$ and $\mathbb{R}^n$. We have*

$$L \text{ injective} \iff \mathrm{Ker}(L) = \{0\}$$

and

$$L \text{ surjective} \iff \mathrm{Im}(L) = \mathbb{R}^n.$$

Application: Solutions of a linear system. We are interested into solving the system of equations in $x = (x_1, \dots, x_m) \in \mathbb{R}^m$

$$\left\{ \begin{array}{ccccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \dots & + & a_{1,m}x_m & = & y_1 \\ \vdots & & & & & & & & \\ a_{n,1}x_1 & + & a_{n,2}x_2 & + & \dots & + & a_{n,m}x_m & = & y_n \end{array} \right. \tag{1}$$

where $a_{i,j} \in \mathbb{R}$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$. If we define the matrix $A \in \mathbb{R}^{n \times m}$ by $A_{i,j} = a_{i,j}$ the system (1) can be rewritten as

$$Ax = y.$$

Solving (1) precisely mean « finding the inverse image of y by A ». From the definition of $\mathrm{Im}(A)$ we get that **the equation $Ax = y$ admits (at least) a solution x_0 if and only if** $y \in \mathrm{Im}(A)$.

We suppose now to be in that case. We would now like to know if there are other solutions. Let x be another solution to $Ax = y$. By subtraction we get

$$A(x - x_0) = y - y = 0.$$

This means that $(x - x_0) \in \mathrm{Ker}(A)$: any solution of $Ax = y$ can therefore be written as $x = x_0 + v$ with $v \in \mathrm{Ker}(A)$. Conversely, one can verify easily that any vector of this form is a solution. We conclude that if the equation $Ax = y$ admits a solution x_0, then the set of **all** solutions is

$$x_0 + \mathrm{Ker}(A) \overset{\text{def}}{=} \{x_0 + v \,|\, v \in \mathrm{Ker}(A)\}.$$

In particular, x_0 **is the unique solution if and only if** $\mathrm{Ker}(A) = \{0\}$.