Pre-sessional Mathematics for Big Data MSc
Class 2: Linear Algebra

Yuri Kalnishkan

September 17, 2017

Linear algebra is vitally important for applied mathematics. We will approach linear algebra from a practical point of view and will be usually sacrificing generality to practicality and simplicity.

Textbooks on linear algebra are plentiful. Browse the library shelf to find what suits you. I will mention some. Textbook [Lip09] is rather elementary; [Poo15] is more advanced but still quite accessible. My personal favourite is [Axl97]: it heavily influenced my perception of the subject and is responsible for such aspects of this tutorial as the focus on linear transformations as opposed to matrices and lack of respect for determinants. The book is not for beginners though.

1 Vectors and Operations on Them

A vector in $\mathbb{R}^n$ is a sequence of n numbers (components, coordinates). In this tutorial we will usually consider *column vectors*

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix};$$

there are also *row vectors*

$$\boldsymbol{v} = \begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix}.$$

Sometimes it does not matter if you are talking about column or row vectors, but as soon as you need to multiply them by matrices, the difference becomes apparent. Check with your course lecturer if the vectors are usually columns or rows.

In a mathematical text it is important to distinguish between vectors and their coordinates. Are v_1, v_2, and v_3 three coordinates of a vector $v = (v_1, v_2, v_3) \in \mathbb{R}^3$ or three vectors in their own right? Some authors use the arrow sign $\vec{v}$ and some use the bar $\bar{v}$ to denote vectors, but those signs are not common outside of school-level textbooks. A more widespread practice is to use bold letters such as $\boldsymbol{v}$ for vectors. This is the convention used in this tutorial.

A vector can be multiplied by a scalar (a scalar is a number as opposed to a vector) and two vectors with the same number of coordinates can be added together. These operations are performed component-wise. If

$$\boldsymbol{v}_1 = \begin{pmatrix} v_1^1 \\ v_2^1 \\ \vdots \\ v_n^1 \end{pmatrix}, \boldsymbol{v}_2 = \begin{pmatrix} v_1^2 \\ v_2^2 \\ \vdots \\ v_n^2 \end{pmatrix} \in \mathbb{R}^n$$

are two vectors and $\alpha_1, \alpha_2 \in \mathbb{R}$ are two scalars then

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 = \begin{pmatrix} \alpha_1 v_1^1 + \alpha_2 v_1^2 \\ \alpha_1 v_2^1 + \alpha_2 v_2^2 \\ \vdots \\ \alpha_1 v_n^1 + \alpha_2 v_n^2 \end{pmatrix} \quad .$$

Note my use of upper indexes to distinguish the coordinates of the first vector from the second one. Notation can be tricky in linear algebra.

Exercise 1. For

$$\boldsymbol{v}_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}$$

and

$$\boldsymbol{v}_2 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

calculate $3\boldsymbol{v}_1$, $-2\boldsymbol{v}_2$, and $\boldsymbol{v}_1 - 4\boldsymbol{v}_2$.

The *scalar product* (also caller *dot product*) of two vectors with the same number of coordinates is calculated as follows:

$$\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = v_1^1 v_1^2 + v_2^1 v_2^2 + \ldots + v_n^1 v_n^2 = \sum_{i=1}^{n} v_i^1 v_i^2 \quad .$$

Notation $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2$ is also used for the scalar product. The sign $\sum$ is used for shortening long sums of similar terms. The summation variable i here is akin to a loop variable in programming and 1 and n are its boundaries.

The *norm* or *length* of a vector is given by

$$\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}\ .$$

The word "length" should be used with care as it is sometimes taken to mean the number of coordinates in a vector.

Exercise 2. Calculate $\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle$, $\|\boldsymbol{v}_1\|$, and $\|\boldsymbol{v}_2\|$ for $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ from Exercise 1.

2 Matrices and Linear Transformation

A *matrix* is a rectangular table of numbers. Here is an $(n \times m)$-matrix A:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix}\ .$$

Notation $a_{i,j}$ is the standard way to refer to the element in row i and column j.

A column vector

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}$$

is an $(m \times 1)$-matrix.

We can multiply an $(n \times m)$-matrix A by an $(m \times 1)$-vector $\boldsymbol{v}$; the result is an $(n \times 1)$-vector

$$A\boldsymbol{v} = \begin{pmatrix} \sum_{j=1}^m a_{1,j} v_j \\ \sum_{j=1}^m a_{2,j} v_j \\ \vdots \\ \sum_{j=1}^m a_{n,j} v_j \end{pmatrix}\ .$$

Note how the middle index m cancels out: $(n \times m) \cdot (m \times 1) = (n \times 1)$. The formula for the multiplication can be memorised as follows: matrices are multiplied *row by column*.

Exercise 3. Let

$$A = \begin{pmatrix} 2 & 1 & 1 \\ -1 & 1 & 0 \end{pmatrix}$$

and $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ be the vectors from Exercise 1. Calculate $A\boldsymbol{v}_1$, $A\boldsymbol{v}_2$, and $A(\boldsymbol{v}_1 - 4\boldsymbol{v}_2)$.

The exercise should convince you that *multiplication is linear*: for any $(n \times m)$-matrix A, $(m \times 1)$-vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, and scalars α_1 and α_2 we have

$$A(\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2) = \alpha_1 A\boldsymbol{v}_1 + \alpha_2 A\boldsymbol{v}_2 \ . \tag{1}$$

The opposite is also true. Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a linear (i.e., satisfying $f(\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2) = \alpha_1 f(\boldsymbol{v}_1) + \alpha_2 f(\boldsymbol{v}_2)$) transformation. Then there is a unique matrix A such that $f(\boldsymbol{v}) = A\boldsymbol{v}$ for all $\boldsymbol{v} \in \mathbb{R}^m$. You will be able to prove this a bit later after we have discussed bases in a linear space. I omit the proof.

3 Composition of Transformations and Matrix Multiplication

Suppose that we have two linear transformations, $f_1 : \mathbb{R}^m \to \mathbb{R}^n$ and $f_2 : \mathbb{R}^n \to \mathbb{R}^k$. We can apply one on top of another, i.e., consider their *composition* $f : \mathbb{R}^m \to \mathbb{R}^k$ given by $f(\boldsymbol{v}) = f_2(f_1(\boldsymbol{v}))$. It is essential that the co-domain of f_1 should match the domain of f_2; otherwise you just will not be able to plug the output of f_1 into f_2!

It is easy to check that if the transformations f_1 and f_2 are linear then the composition f is also linear.

Exercise 4. Check it.

Let A be the matrix corresponding to f_1 and B be the matrix corresponding to f_2. What is the matrix corresponding to the composition then? It is the *matrix product* $C = BA$. It has the size $k \times m$ because it caters for a transformation $\mathbb{R}^m \to \mathbb{R}^k$ (note that we have $(k \times n) \cdot (n \times m) = (k \times m)$) and is given by the following formula. If

$$B = \begin{pmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,n} \\ b_{2,1} & b_{2,2} & \dots & b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k,1} & b_{k,2} & \dots & b_{k,n} \end{pmatrix}$$

and

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix}$$

then

$$C = \begin{pmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ c_{2,1} & c_{2,2} & \dots & c_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{k,1} & c_{k,2} & \dots & c_{k,m} \end{pmatrix} \quad ,$$

where

$$c_{i,j} = \sum_{s=1}^{n} b_{i,s} a_{s,j} \tag{2}$$

(note how the summation index s ranges from 1 to k, which is the common dimension). Again the formula can be memorised as follows: *matrices are multiplied row by column.* To get $c_{i,j}$ you multiply row i of the first matrix by the column j of the second matrix.

Exercise 5. Let A be as in Exercise 3 and $\boldsymbol{v}_1$ as in Exercise 1. Let

$$B = \begin{pmatrix} -1 & 2 \\ 1 & 0 \end{pmatrix} \quad .$$

Calculate $C = BA$. Calculate $B(A\boldsymbol{v}_1)$ and $C\boldsymbol{v}_1$ and check that the results match.

From the discussion above you can see that one can only multiply a $(r \times s)$-matrix B by an $(m \times n)$-matrix A if $s = m$. Otherwise there is no composition of the transformations involved.

Matrix multiplication is *associative*: if A, B, and C are such that the products AB and BC exist, then $(AB)C = A(BC)$. This can be checked either from formula (2) or by noting that composing transformations is an associative operation; think of the picture

$$\mathbb{R}^n \xrightarrow{C} \mathbb{R}^m \xrightarrow{B} \mathbb{R}^k \xrightarrow{A} \mathbb{R}^s \quad .$$

Matrix multiplication is not *commutative*. Generally speaking $AB \neq BA$, even if both the products exist (i.e., A and B are square matrices of the same size). Of course $AB = BA$ is possible by chance, but this equality does not have to hold.

4 Linear Independence and Bases

In order to have a meaningful discussion of matrices we need to look into the theory of vector spaces. Some facts in this section will remain unproven; you can find proofs in any good linear algebra textbook.

Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ be vectors with the same number of components. If we take k scalars $\alpha_1, \alpha_2, \ldots, \alpha_k$ (called coefficients in this context), we can construct a *linear combination* $\alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2 + \ldots + \alpha_k\boldsymbol{v}_k$ of the vectors. All linear combinations of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ with all possible coefficients are said to make up the *span* of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$. In other terms, the span of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ is the set

$$\{\alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2 + \ldots + \alpha_k\boldsymbol{v}_k \mid \alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}\} \ .$$

The vector of all zeros

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

must belong to the span. Indeed, it is a linear combination with coefficients $0, 0, \ldots, 0$:

$$\mathbf{0} = 0 \cdot \boldsymbol{v}_1 + 0 \cdot \boldsymbol{v}_2 + \ldots + 0 \cdot \boldsymbol{v}_k \ .$$

Are there any other linear combinations of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ equal to $\mathbf{0}$? This is a very important question and leads to the following important definition.

The vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are *linearly independent* if the equality

$$\alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2 + \ldots + \alpha_k\boldsymbol{v}_k = \mathbf{0} \tag{3}$$

implies that $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$. Another way of putting this is saying that (3) is only possible for $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$.

Exercise 6. Are the vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ from Exercise 1 linearly independent?

Hint. Consider the equality $\alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2 = \mathbf{0}$. It amounts to

$$\alpha_1 \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

or the linear system

$$\begin{cases} 3\alpha_1 + 2\alpha_2 & = 0 \\ -\alpha_1 & = 0 \\ 2\alpha_1 + \alpha_2 & = 0 \end{cases}$$

Solve the system and check if there is a solution apart from $\alpha_1 = \alpha_2 = 0$.

The following property clarifies the meaning of independence. Vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are NOT independent if and only if one of them can be represented as a linear combination of the others.

This is easy to see. Suppose that, say, $\boldsymbol{v}_k$ equals a linear combination of the other vectors:

$$\boldsymbol{v}_k = \alpha_1 \boldsymbol{v}_2 + \ldots + \alpha_{k-1} \boldsymbol{v}_{k-1} \ .$$

Then

$$\alpha_1 \boldsymbol{v}_2 + \ldots + \alpha_{k-1} \boldsymbol{v}_{k-1} - 1 \cdot \boldsymbol{v}_k = \mathbf{0}$$

and the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are not independent.

On the other hand suppose that $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are not linearly independent. Then there are coefficients α_i such that

$$\alpha_1 \boldsymbol{v}_1 + \ldots + \alpha_{k-1} \boldsymbol{v}_{k-1} + \alpha_k \boldsymbol{v}_k = \mathbf{0}$$

and not every coefficient α_i equals 0. There must be a coefficient not equal to 0. Let $\alpha_k \neq 0$. Then we can divide by α_k and write

$$\boldsymbol{v}_k = -\frac{\alpha_1}{\alpha_k} \boldsymbol{v}_1 - \ldots - \frac{\alpha_{k-1}}{\alpha_{k-1}} \boldsymbol{v}_{k-1} \ .$$

The independence thus mean that we cannot omit any of the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ when we build the span. Every vector contributes to the span.

Another important property of independent systems is as follows. By definition vector $\mathbf{0}$ has a unique representation as $\alpha_1 \boldsymbol{v}_1 + \ldots + \alpha_{k-1} \boldsymbol{v}_{k-1} + \alpha_k \boldsymbol{v}_k$. This is also true of any other vector from the span. If $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are linearly independent and

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \ldots + \alpha_k \boldsymbol{v}_k = \beta_1 \boldsymbol{v}_1 + \beta_2 \boldsymbol{v}_2 + \ldots + \beta_k \boldsymbol{v}_k \ ,$$

then $\alpha_1 = \beta_1$, $\alpha_2 = \beta_2$, $\ldots$, $\alpha_k = \beta_k$.

Exercise 7. Why?

Thus every vector $\boldsymbol{v}$ is the span of linearly independent $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ can be uniquely represented as

$$\boldsymbol{v} = \alpha_1 \boldsymbol{v}_1 + \ldots + \alpha_{k-1} \boldsymbol{v}_{k-1} + \alpha_k \boldsymbol{v}_k \ .$$

The coefficients $\alpha_1, \alpha_2, \ldots, \alpha_k$ can be thought of as coordinates of $\boldsymbol{v}$.

Now let us give the following definition. Vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ from $\mathbb{R}^n$ (i.e., with n components) are a *basis* of $\mathbb{R}^n$ if they are linearly independent and their span is the whole of $\mathbb{R}^n$.

It can be shown (I omit the proofs here) that fewer than n vectors can never span the whole of $\mathbb{R}^n$. On the other hand if we take more than n vectors in $\mathbb{R}^n$, they are never linearly independent. Therefore every basis of $\mathbb{R}^n$ must have exactly n vectors. Moreover, for n vectors in $\mathbb{R}^n$ the properties

- to be linearly independent and
- to span the whole of $\mathbb{R}^n$

are equivalent. One holds if and only if the other holds. Thus the definition of the basis can be rewritten as follows. Vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ form a basis in $\mathbb{R}^n$ if they are linearly independent.

There is one important basis in $\mathbb{R}^n$ you already know even if you have not encountered this term.

Consider n vectors in $\mathbb{R}^n$:

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots, \boldsymbol{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

(the vector $\boldsymbol{e}_i$ consists of $n-1$ zeros and 1 in position i). They are linearly independent. Indeed,

$$\alpha_1 \boldsymbol{e}_1 + \alpha_2 \boldsymbol{e}_2 + \ldots + \alpha_n \boldsymbol{e}_n = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$$

and $\alpha_1 \boldsymbol{e}_1 + \alpha_2 \boldsymbol{e}_2 + \ldots + \alpha_n \boldsymbol{e}_n = \mathbf{0}$ immediately implies $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$. Their span coincides with $\mathbb{R}^n$. Indeed, for

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

we have $\boldsymbol{v} = v_1 \boldsymbol{e}_1 + v_2 \boldsymbol{e}_2 + \ldots + \boldsymbol{e}_n$. The numbers $v_1, v_2, \ldots, v_n$ are the coordinates w.r.t. this basis.

5 Inverse Matrices

Consider a square $(n \times n)$-matrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix} \quad .$$

As we discussed previously, it corresponds to a transformation $\mathbb{R}^n \to \mathbb{R}^n$.

Consider the columns of A. There is an important connection between linear independence of the columns and the properties of the transformation. Namely, the columns are linearly independent if and only if the transformation is a bijection.

We can work it out using the facts from the previous section. Let us denote the columns of A by $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$. They are vectors from $\mathbb{R}^n$. The crucial observation is that when we apply A to a vector

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

the result $A\boldsymbol{v}$ is the linear combination $A\boldsymbol{v} = v_1\boldsymbol{a}_1 + v_2\boldsymbol{a}_2 + \dots + v_n\boldsymbol{a}_n$ and thus belongs to the span of $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$.

We know that if $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ are linearly independent, then every vector from the span has a unique representation as a linear combination. Thus if

$$v_1\boldsymbol{a}_1 + v_2\boldsymbol{a}_2 + \dots + v_n\boldsymbol{a}_n = w_1\boldsymbol{a}_1 + w_2\boldsymbol{a}_2 + \dots + w_n\boldsymbol{a}_n \quad ,$$

then $v_1 = w_1$, $v_2 = w_2$, $\dots$, $v_n = w_n$. But this means precisely that A is an injection.

We also noted that n linearly independent vectors span the whole of $\mathbb{R}^n$. This means that A is a surjection. Thus if the vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ are linearly independent, then A is a bijection.

If the columns of a square matrix A are linearly independent, one says that A is *non-singular*. Otherwise A is *singular*.

There are some equivalent conditions and ways of saying that A is non-singular.

First A is non-singular if and only if its determinant does not equal zero: $\det A \neq 0$. The apparent simplicity of this formula is somewhat misleading.

The determinants are not easy to interpret theoretically and awfully difficult to calculate in practice.

Secondly there is the notion of rank. The rank of A, denoted $\operatorname{rank} A$, is the maximum number of linearly independent columns of A (or linearly independent rows of A: the column rank is equal to the row rank). An (n, n)-matrix A is thus nonsingular if and only if $\operatorname{rank} A = n$.

Thirdly columns of A are linearly independent if and only if rows of A are linearly independent (see the column and row rank above). So singularity can be defined in terms of rows as well.

Finally, A is singular if and only if there is a non-zero vector $\boldsymbol{v}$ such that $A\boldsymbol{v} = \mathbf{0}$.

Exercise 8. Can you prove the last statement?

Bijections have inverse transformations. So if A is nonsingular, it has an inverse transformation. One can show it is also linear. The matrix of the inverse transformation is denoted A^{-1}.

By the definition of an inverse transformation, for every vector $\boldsymbol{v}$ we have $A^{-1}A\boldsymbol{v} = \boldsymbol{v}$. This means that the product $A^{-1}A$ corresponds to the so called *identity transformation*. The identity transformation maps every vector into itself (yes, mathematicians call this a transformation too). The matrix of this transformation is

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} ;$$

it is called the *unit* matrix or *identity* matrix (it has zeros everywhere except on the diagonal going from the upper left to the lower right corner; this diagonal is called the *main diagonal* of a square matrix).

The condition "A^{-1} is the matrix of the inverse transformation" thus can be written as $AA^{-1} = A^{-1}A = I$.

Note 1. Multiplying any matrix by the identity matrix of the corresponding size does not change it: $AI = A$ and $IA = A$. This can be proved directly by multiplying the matrices or by making the following observation. The composition of a transformation with the matching identity transformation is the same transformation. Look at

$$\begin{gathered} \mathbb{R}^n \xrightarrow{A} \mathbb{R}^m \xrightarrow{I} \mathbb{R}^m \\ \mathbb{R}^n \xrightarrow{I} \mathbb{R}^n \xrightarrow{A} \mathbb{R}^m \end{gathered} .$$

Singular matrices have no inverses because the corresponding transformations are not bijections. Non-square matrices have no inverses either because the corresponding transformations are never bijections.

Exercise 9. Explain why a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$ cannot be a bijection when $n \neq m$. Use the facts from the previous section.

Note 2. Calculating the inverse matrix is time-consuming. Creating efficient numerical algorithms for this task is a very important direction of computational mathematics.

The inversion relates to matrix multiplication through the following law. If the matrices A and B are non-singular, then AB is also non-singular and $(AB)^{-1} = B^{-1}A^{-1}$.

This is easy to check. The associativity of the matrix product implies that $(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$. Similarly $(B^{-1}A^{-1})(AB) = I$.

6 Eigenvalues and Eigenvectors

Let A be a square $(n \times n)$-matrix. We say that a number λ is an *eigenvalue* and a vector $\boldsymbol{v} \neq \mathbf{0}$ is an *eigenvector* if $A\boldsymbol{v} = \lambda\boldsymbol{v}$. (We imposed the condition $\boldsymbol{v} \neq \mathbf{0}$ because for the vector of zeros we have $A\mathbf{0} = \mathbf{0} = \lambda\mathbf{0}$ for every λ so zero eigenvectors are not interesting.)

Clearly, if $\boldsymbol{v} \neq \mathbf{0}$ is an eigenvector and $\boldsymbol{w} = \alpha\boldsymbol{v}$, where $\alpha \neq 0$, then $\boldsymbol{w}$ is also an eigenvector: $A\boldsymbol{w} = A\alpha\boldsymbol{v} = \alpha A\boldsymbol{v} = \alpha\lambda\boldsymbol{v} = \lambda\boldsymbol{w}$. Therefore what really matters in an eigenvector is the direction, not the norm.

Eigenvectors are important because they show to us how the linear transformation corresponding to A works. Namely, in the direction of $\boldsymbol{v}$ the transformation stretches the vectors by λ.

The equation $A\boldsymbol{v} = \lambda\boldsymbol{v}$ can be rewritten as $A\boldsymbol{v} = \lambda I\boldsymbol{v}$ and

$$(A - \lambda I)\boldsymbol{v} = \mathbf{0} \ . \tag{4}$$

As discussed in the previous section, the existence of $\boldsymbol{v} \neq \mathbf{0}$ such that (4) holds means that the matrix $(A - \lambda I)$ is singular.

Hence we can give a different definition of an eigenvalue. A number λ is an eigenvalue of a square matrix A if the matrix $A - \lambda I$ is singular.

We mentioned above how singularity can be expressed in terms of determinants. This leads to the popular definition saying that eigenvalues are roots of the equation $\det(A - \lambda I) = 0$. Again, I believe that the apparent simplicity of this statement is misleading.

Exercise 10. Consider a (2×2)-matrix

$$A = \begin{pmatrix} 0 & 2 \\ 2 & 3 \end{pmatrix} .$$

Is -1 an eigenvalue of A? Is 1 an eigenvalue of A?

Note 3. Finding eigenvalues is a difficult (but very important) numerical problem.

In a sense an $(n \times n)$-matrix always has n eigenvalues. But

1. they may be complex numbers and correspond to complex eigenvectors;

2. they should be counted with *multiplicity* (and the count should include generalised eigenvalues).

So it is possible for a matrix to have fewer than n real eigenvalues or no real eigenvalues at all.

Exercise 11. Show that the matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

has no *real* eigenvalues. *Hint:* Try to understand what the transformation corresponding to A does to the plain $\mathbb{R}^2$. Calculate and plot the image of $A\boldsymbol{v}$.

Exercise 12. The vector $\mathbf{0}$ cannot be an eigenvector by definition, but the number $\lambda = 0$ may be an eigenvalue. What does it mean for a matrix A to have an eigenvalue $\lambda = 0$?

7 Transposition

Transposing a matrix is flipping it over the main diagonal. If

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix}$$

then its transpose A' (notation A^T is also used) is the matrix

$$A' = \begin{pmatrix} a_{1,1} & a_{2,1} & \dots & a_{m,1} \\ a_{1,2} & a_{2,2} & \dots & a_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \dots & a_{m,n} \end{pmatrix} .$$

If A was an $(n \times m)$-matrix, then A' is an $(m \times n)$-matrix.

The transpose of a column vector

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

is the row vector

$$\boldsymbol{v}' = \begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix} .$$

This leads to a useful formula for the scalar product

$$\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = \sum_{i=1}^{n} v_i^1 v_i^2 = \boldsymbol{v}_1' \boldsymbol{v}_2 \ ;$$

if you multiply $\boldsymbol{v}_1'$ by $\boldsymbol{v}_2$ as matrices, you get the scalar product.

Exercise 13. Transpose the matrix

$$A = \begin{pmatrix} 2 & 3 \\ -1 & 2 \\ 0 & 1 \end{pmatrix}$$

Transposing a matrix twice gives us the original matrix: $(A')' = A$.

How to transpose a matrix product? We have $(AB)' = B'A'$. This can be checked using the formula for matrix multiplication.

How to transpose the inverse of a square matrix? We have $(A')^{-1} = (A^{-1})'$. Indeed, we have $(A^{-1})'A' = (AA^{-1})' = I' = I$ (the matrix I is *symmetric* and does not change from transposing).

If A is singular, then A' is also singular; if A is non-singular, then A' is non-singular (recall that singularity may be defined either with columns or with rows).

What are the eigenvalues of a transpose? Recall the following definition. If a number λ is an eigenvalue of A then the matrix $A - \lambda I$ is singular.

Then the transpose $(A - \lambda I)'$ is singular too. Let us calculate its transpose: $(A - \lambda I)' = A' - \lambda I' = A' - \lambda I$. But if $A' - \lambda I$ is singular then λ is an eigenvalue of A'. So A and A' have the same eigenvalues.

The eigenvectors may be very different though!

In the literature you may find the term *adjoint*. The linear transformation corresponding to A' is the *adjoint* of A.

8 Symmetric Matrices

A particularly important class of matrices is symmetric matrices. A square matrix A is *symmetric* if $A' = A$. In other terms it does not change if you flip it around the main diagonal.

This property has important implications for eigenvalues. Symmetric matrices have full sets of real eigenvalues (well, counting multiplicities). An eigenvalue λ of a symmetric matrix is always real and eigenvectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ corresponding to different eigenvalues λ_1 and λ_2 must be orthogonal: $\boldsymbol{v}_1'\boldsymbol{v}_2 = 0$. Hence for a symmetric $(n \times n)$-matrix A there is a basis consisting of orthogonal vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ such that in the direction of $\boldsymbol{v}_i$ the transformation A is a stretch plus, perhaps, a mirroring. This fact is called the *spectral theorem for self-adjoint operators*.

Some sub-classes of symmetric matrices are even more special. A symmetric $(n \times n)$-matrix M is called symmetric positive semi-definite, if for any vector $\boldsymbol{v} \in \mathbb{R}^n$ we get $\boldsymbol{v}'M\boldsymbol{v} \geq 0$.

Can the expression $\boldsymbol{v}'M\boldsymbol{v}$ equal zero? Of course it can: $\mathbf{0}'M\mathbf{0} = 0$. But are there other such vectors apart from $\mathbf{0}$? If for every $\boldsymbol{v} \neq \mathbf{0}$ we have $\boldsymbol{v}'M\boldsymbol{v} > 0$, then the symmetric matrix M is called *symmetric positive definite*.

All eigenvalues of a symmetric positive semi-definite matrix M are non-negative and all eigenvalues of a symmetric positive-definite M are strictly positive.

Exercise 14. Why?

A symmetric positive definite matrix is always non-singular. If it is only known that the matrix is symmetric positive semi-definite, it can be singular.

Exercise 15. Why?

The importance of symmetric positive semi-definite and positive definite matrices for statistics and data analysis is due to the fact that covariance matrices are always symmetric positive semi-definite.

Let us prove this (covariances will only be discussed in the next class; feel free to skip this now and get back here later).

Consider n random variables $\xi_1, \xi_2, \ldots, \xi_n$. Their covariance matrix M has in row i and column j the number

$$\mathrm{cov}(\xi_i, \xi_j) = \mathrm{E}[(\xi_i - \mathrm{E}\,\xi_i)(\xi_j - \mathrm{E}\,\xi_j)] \enspace .$$

Covariance is symmetric and $\mathrm{cov}(\xi_i, \xi_j) = \mathrm{cov}(\xi_j, \xi_i)$. Thus M is symmetric.

Take

$$\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

and consider the random variable $\xi = v_1\xi_1 + v_2\xi_2 + \ldots + v_n\xi_n$. Let us calculate the variance of ξ by opening up the brackets:

$$\begin{aligned}
\mathrm{var}\,\xi &= \mathrm{E}(\xi - \mathrm{E}\,\xi)^2 \\
&= \mathrm{E}\left[(v_1\xi_1 - \mathrm{E}\,v_1\xi_1) + \ldots + (v_n\xi_n - \mathrm{E}\,v_n\xi_n)\right]^2 \\
&= \mathrm{E}\left[\sum_{i,j=1}^{n} (v_i\xi_i - \mathrm{E}\,v_i\xi_i)(v_j\xi_j - \mathrm{E}\,v_j\xi_j)\right] \\
&= \sum_{i,j=1}^{n} v_i v_j\,\mathrm{E}\left[(\xi_i - \mathrm{E}\,\xi_i)(\xi_j - \mathrm{E}\,\xi_j)\right] \\
&= \boldsymbol{v}' M \boldsymbol{v} \enspace .
\end{aligned}$$

Since the variance is always non-negative, $\mathrm{var}\,\xi = \boldsymbol{v}'M\boldsymbol{v} \geq 0$ and M is symmetric positive semi-definite.

References

[Axl97] S. Axler. *Linear algebra done right.* Springer, 2nd edition, 1997.

[Lip09] S. Lipschutz. *Schaum's outline of theory and problems of linear algebra.* McGraw-Hill, 2009.

[Poo15] D. Poole. *Linear algebra: a modern introduction.* Cengage Learning, 4th edition, 2015.