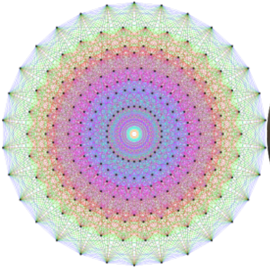
Matrix operations and systems of linear equations

4433LALG3: Linear Algebra

Week 1, Lecture 2, Valente Ramírez

Mathematical & Statistical Methods group — Biometris, Wageningen University & Research



Overview

- Matrix operations
- Systems of linear equations
- Gaussian elimination
- Supplementary material

References:

- Nicholson §2.1, 2.3 & §1.1–1.2
- 3Blue1Brown Ch.4

Section 1

Matrix operations

Matrix operations

Remark

The main use of matrices is to represent linear transformations.

Therefore, all matrix operations are meant to reflect operations relevant to linear transformations.

Addition and scalar multiplication

Consider a linear transformation T and a **scalar** (i.e. a number) k.

- We can create a new linear transformation by re-scaling T:

  $$T_{\text{new}}(\mathbf{x}) = kT(\mathbf{x})$$

Consider two linear transformations T_1 and T_2, both from $\mathbb{R}^n$ to $\mathbb{R}^m$.

- We can create a new linear transformation by adding up the outputs:

  $$T_{\text{new}}(\mathbf{x}) = T_1(\mathbf{x}) + T_2(\mathbf{x})$$

Addition and scalar multiplication

Definition

If $A = [a_{ij}]$ is a matrix and k any number, the **scalar multiple** kA is the matrix obtained from A by multiplying each entry of A by k.

$$kA = [ka_{ij}].$$

Definition

If $A = [a_{ij}]$ and $B = [b_{ij}]$ are matrices of the same size, their **sum** $A + B$ is the matrix formed by adding the corresponding entries.

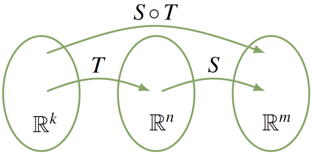
$$A + B = [a_{ij} + b_{ij}].$$

Composition of transformations

Suppose we have two linear transformations:

$$T\colon \mathbb{R}^k \to \mathbb{R}^n \qquad \text{and} \qquad S\colon \mathbb{R}^n \to \mathbb{R}^m.$$

We could **compose** these transformations:



Fact

The composition $S \circ T$ of two linear transformations is itself linear.

Matrix multiplication

Given A, a $m \times n$ matrix, and B, a $n \times k$ matrix, we have

$$T_B \colon \mathbb{R}^k \to \mathbb{R}^n \qquad \text{and} \qquad T_A \colon \mathbb{R}^n \to \mathbb{R}^m.$$

By composition, we obtain a new linear transformation

$$T_A \circ T_B \colon \mathbb{R}^k \to \mathbb{R}^m.$$

This linear transformation is represented by some $m \times k$ matrix.

Definition

If A is an $m \times n$ matrix and B is an $n \times k$ matrix, we define their **product** AB as the $m \times k$ matrix which represents the linear transformation

$$T_{AB} = T_A \circ T_B \colon \mathbb{R}^k \to \mathbb{R}^m.$$

Matrix multiplication

Question

How to express the entries of AB in terms of the entries a_{ij} and b_{ij}?

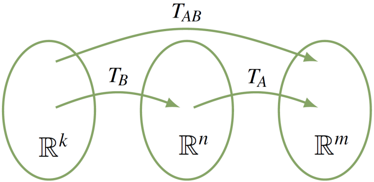
There is a formula for that: The (i,j)–entry of AB equals

$$\sum_{s=1}^{n} a_{is}b_{sj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj}$$

...but that doesn't tell us much.

Matrix multiplication: insight

A better way to think about it, is to keep the composition image in mind:



Start with the basic vector $\mathbf{e}_1 \in \mathbb{R}^k$.

- It is first transformed into $B\mathbf{e}_1 = \mathbf{b}_1$, the first column of B.
- Then it is transformed into $A\mathbf{b}_1$.
- Therefore, $T_{AB}(\mathbf{e}_1) = A\mathbf{b}_1$.

From the last bullet point we conclude that the first column of AB is $A\mathbf{b}_1$.

The same argument shows that the j^{th} column of AB is $A\mathbf{b}_j$.

Matrix multiplication: in practice

Theorem

Consider an $m \times n$ matrix A, and an $n \times k$ matrix B.

Let $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_k$ be the columns of B.

Then AB is the $m \times k$ matrix whose columns are $A\mathbf{b}_1$, $A\mathbf{b}_2$, ..., $A\mathbf{b}_k$.

Using the notation $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_k \end{bmatrix}$, we can write:

$$AB = A \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_k \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \ldots & A\mathbf{b}_k \end{bmatrix}.$$

Example

Consider: $A = \begin{bmatrix} 2 & 9 & 0 \\ -4 & 1 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & 5 \\ 0 & -1 \\ 4 & 6 \end{bmatrix}$.

First, compute: $A\mathbf{b}_1 = \begin{bmatrix} 2 & 9 & 0 \\ -4 & 1 & 5 \end{bmatrix} \begin{bmatrix} 7 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 14 \\ -8 \end{bmatrix}$

Then: $A\mathbf{b}_2 = \begin{bmatrix} 2 & 9 & 0 \\ -4 & 1 & 5 \end{bmatrix} \begin{bmatrix} 5 \\ -1 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \end{bmatrix}$

Putting it together: $AB = A\begin{bmatrix} \mathbf{b_1} & \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} 14 & 1 \\ -8 & 9 \end{bmatrix}$.

Compatibility

Warning

The product AB only exists if the matrices are **compatible**:

The number of columns of A equals the number of rows of B.

Notice how the size of AB is determined:

$$\begin{array}{cccc} A & B & = & AB \\ (m \times n) & (n \times k) & & (m \times k) \end{array}$$

Matrix multiplication in R

```
# From the last example:
> A <- matrix(c(2,-4, 9,1, 0,5), 2, 3)
> B <- matrix(c(7,0,4, 5,-1,6), 3, 2)

> A %*% B  # result is a 2x2 matrix
     [,1] [,2]
[1,]   14    1
[2,]   -8    9

> B %*% A  # result is a 3x3 matrix
     [,1] [,2] [,3]
[1,]   -6   68   25
[2,]    4   -1   -5
[3,]  -16   42   30

# The new matrix X is not compatible for the product AX
> X <- matrix(c(1,1,1,1), 2, 2)
> A %*% X
Error in A %*% X : non-conformable arguments
```

Section 2

Linear equations

Linear equations in one variable

A *linear equation in one variable* is an equality of the form

$$ax + c = 0.$$

For example, $2x + 4 = 0$.

Above, x is the *variable* or *unknown*. The variable may take different values, and for each value the equation may or may not be satisfied.

For example,

- If $x = -2$ the equation is **satisfied** because $2 \cdot (-2) + 4 = 0$
- If $x = 1$ the equation is **not satisfied** because $2 \cdot 1 + 4 \neq 0$

A number s is a **solution** whenever the equation is satisfied for $x = s$.

Remark

If we only cared about *one* variable, this slide would be all you need to know...
Luckily, the world is more interesting than that!

Systems of linear equations

A **linear equation** in n variables $x_1, x_2, \ldots, x_n$ is an equality of the form

$$a_1x_1 + a_2x_2 + \ldots a_nx_n = b.$$

A **system** of linear equations is a collection of $m > 1$ linear equations on the same n variables.

Example:

$$\begin{aligned} 3x_1 + 2x_2 - x_3 + x_4 &= -1 \\ 2x_1 - x_3 + 2x_4 &= 0 \\ 3x_1 + x_2 + 2x_3 + 5x_4 &= 2 \end{aligned}$$

Systems of linear equations

Given a system of m linear equations on n unknowns:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots a_{2n}x_n &= b_2, \\ &\dots \\ a_{m1}x_1 + a_{12}x_2 + \dots a_{mn}x_n &= b_m, \end{aligned}$$

there is an underlying linear transformation $T_A \colon \mathbb{R}^n \to \mathbb{R}^m$.

If we set $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$, the system is asking:

For which input $\mathbf{x}$ do we get $\mathbf{b}$ as output under T_A?

This can also be rewritten as:

$$T_A(\mathbf{x}) = \mathbf{b} \quad \text{or} \quad A\mathbf{x} = \mathbf{b}.$$

Application: systems of linear equations

Example

Fertilizers are usually labeled with three numbers indicating the relative content in percentage by weight of the primary macronutrients: nitrogen (N), phosphorus (P), and potassium (K). A fertilizer manufacturer produces three different mixes with different NPK labels, as shown below.

	Mix 1	Mix 2	Mix 3
N	8%	10%	5%
P	4%	4%	5%
K	6%	4%	10%

A farmer growing a polyculture of corn, beans and pumpkin is advised to fertilize with $30\,\mathrm{g}$ nitrogen and $20\,\mathrm{g}$ phosphorus per m^2 (the amount of potassium is not as important).

How can the different mixes be combined to provide suitable fertilization for a $500\,\mathrm{m}^2$ field?

Application: analysis

Our target is:

- $30\,\mathrm{g/m^2} \times 500\,\mathrm{m^2} = 15\,\mathrm{kg}$ N,
- $20\,\mathrm{g/m^2} \times 500\,\mathrm{m^2} = 10\,\mathrm{kg}$ P.

	Mix 1	Mix 2	Mix 3
N	8%	10%	5%
P	4%	4%	5%

Let x_1, x_2, x_3 denote the amount added (in kg) of each of the mixes.
We want:

$$0.08x_1 + 0.10x_2 + 0.05x_3 = 15$$
$$0.04x_1 + 0.04x_2 + 0.05x_3 = 10$$

The farmer could add $50\,\mathrm{kg}$ of mix 1, $50\,\mathrm{kg}$ of mix 2, and $120\,\mathrm{kg}$ of mix 3:

$$0.08(50) + 0.10(50) + 0.05(120) = 15$$
$$0.04(50) + 0.04(50) + 0.05(120) = 10$$

Therefore, $x_1 = 50, x_2 = 50, x_3 = 120$ is a solution to the system.

Application: some remarks

In the previous slide we found **one** solution: $x_1 = 50, x_2 = 50, x_3 = 120$.

Keep in mind:

- Solutions are not always unique:
  $x_1 = 125, x_2 = 0, x_3 = 100$ is also a solution
- Theoretical solutions are not always useful:
  $x_1 = 500, x_2 = -250, x_3 = 0$ is also a solution
- Solutions don't always exist (depending on the system)

Some definitions

Definition

A system is called **consistent** if it has at least one solution. Otherwise it is **inconsistent**.

Objective

We want to be able to distinguish consistent systems from inconsistent ones, and to systematically find *all* solutions of a consistent system.

Section 3

Gaussian elimination

Matrix notation

A system of equations

$$
\begin{aligned}
0.08x_1 + 0.10x_2 + 0.05x_3 &= 15 \\
0.04x_1 + 0.04x_2 + 0.05x_3 &= 10
\end{aligned}
$$

can be represented in a more compact way using *matrix notation*:

- Coefficient matrix: $\begin{bmatrix} 0.08 & 0.10 & 0.05 \\ 0.04 & 0.04 & 0.05 \end{bmatrix}$
- Constant matrix: $\begin{bmatrix} 15 \\ 10 \end{bmatrix}$
- **Augmented matrix**: $\left[\begin{array}{ccc|c} 0.08 & 0.10 & 0.05 & 15 \\ 0.04 & 0.04 & 0.05 & 10 \end{array}\right]$

Elementary operations

Definition

The following operations, called **elementary operations** can be performed on systems of equations to produce *equivalent systems*.

1. Interchange two equations
2. Multiply one equation by a nonzero number
3. Add a multiple of one equation to a different equation

Definition

The following are called **elementary row operations** on a matrix.

1. Interchange two rows
2. Multiply one row by a nonzero number
3. Add a multiple of one row to a different row

Example: elementary operations

Consider the following system and its augmented matrix.

$$\begin{aligned} x + 2y &= -2 \\ 2x + y &= 7 \end{aligned} \qquad \left[\begin{array}{cc|c} 1 & 2 & -2 \\ 2 & 1 & 7 \end{array}\right]$$

Subtract twice the first equation from the second equation.

$$\begin{aligned} x + 2y &= -2 \\ -3y &= 11 \end{aligned} \qquad \left[\begin{array}{cc|c} 1 & 2 & -2 \\ 0 & -3 & 11 \end{array}\right]$$

Multiplying the second equation by $-\frac{1}{3}$.

$$\begin{aligned} x + 2y &= -2 \\ y &= -\tfrac{11}{3} \end{aligned} \qquad \left[\begin{array}{cc|c} 1 & 2 & -2 \\ 0 & 1 & -\frac{11}{3} \end{array}\right]$$

Subtract twice the second equation from the first equation.

$$\begin{aligned} x &= \tfrac{16}{3} \\ y &= -\tfrac{11}{3} \end{aligned} \qquad \left[\begin{array}{cc|c} 1 & 0 & \frac{16}{3} \\ 0 & 1 & -\frac{11}{3} \end{array}\right]$$

Row-echelon form

Definition

A matrix is said to be in **row-echelon form** if it satisfies:

1. All rows consisting entirely of zeros are at the bottom
2. The first nonzero entry from the left in each nonzero row is a 1, called *the leading 1* for that row
3. Each leading 1 is to the right of all leading 1s in the rows above it

It is moreover in **reduced** row-echelon form if in addition:

4. Each leading 1 is the only nonzero entry in its column

$$\begin{bmatrix} 0 & 1 & * & * & * & * & * \\ 0 & 0 & 0 & 1 & * & * & * \\ 0 & 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & * & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 1 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 1 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Examples: reduced row-echelon form

Example 1: The "perfect" situation

$$\left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & -1 \end{array}\right]$$

Example 2: Row of zeroes

$$\left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right]$$

Example 3: Leading 1 in constant matrix

$$\left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$$

Example 4: Free variables

$$\left[\begin{array}{cccc|c} 1 & 3 & 0 & 1 & 2 \\ 0 & 0 & 1 & 0 & -1 \end{array}\right]$$

Basic variables and free variables

Definition

For a consistent system of linear equations, we distinguish two types of variables:

1. A **basic variable** is a variable that has a leading 1 in its column.
2. A **free variable** is a variable not having a leading 1 in its column.

For example:

$$\begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \left[\begin{array}{cccc|c} 1 & 0 & 5 & 0 & -1 \\ 0 & 1 & 3 & 0 & 3 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right] \end{array}$$

Example: system with free variables

Example

The system below has one free variable: x_3.

$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & & \\ \end{array}$$
$$\left[\begin{array}{cccc|c} 1 & 0 & 5 & 0 & -1 \\ 0 & 1 & 3 & 0 & 3 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right]$$

$$\begin{aligned} x_1 + 5x_3 &= -1 \\ x_2 + 3x_3 &= 3 \\ x_4 &= 2 \end{aligned}$$

We can assign any arbitrary value to the variable x_3, let's say $x_3 = s$, and express each x_i in terms of this value s:

$$\begin{aligned} x_1 &= -1 - 5s \\ x_2 &= 3 - 3s \\ x_3 &= s \\ x_4 &= 2 \end{aligned}$$

Above, the number s is called a **parameter**.

Section 4

Supplementary material

Transpose of a matrix

The **transpose** of a $m \times n$ matrix A is the $n \times m$ matrix $A^\top$ whose rows are formed from the columns of A.

In R, the transpose of a matrix A is computed by `t(A)`.

Example:

If $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$, then $A^\top = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$.

In particular, the transpose of a column vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is

the row vector $\mathbf{x}^\top = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$.

Sometimes, to save space, we write column vectors as: $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^\top$.

Transpose of a matrix

Warning

There are **many** different notations to denote matrix transposition.

For example:

$$\mathbf{x}^\top, \quad \mathbf{x}^T, \quad \mathbf{x}^t, \quad \mathbf{x}'$$

The last one, $\mathbf{x}'$, is very common – it has <u>nothing</u> to do with derivatives!

Matrix transposition seems very simple at a first glance. But it will turn out to be extremely important in subsequent topics.

Some important properties are listed in Nicholson, Theorem 2.1.2 and Theorem 2.3.3 (point 6.).

Partitioned matrices

It is often useful to consider matrices whose entries are themselves matrices (called **blocks**).

We've already seen augmented matrices: $[A \mid \mathbf{b}] = \left[\begin{array}{cc|c} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \end{array}\right]$.

A more elaborate example would be:

$$\begin{bmatrix} I_2 & 0_{23} \\ P & Q \end{bmatrix} = \left[\begin{array}{rr|rrr} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 2 & -1 & 4 & 2 & 1 \\ 3 & 1 & -1 & 7 & 5 \end{array}\right],$$

with

$$P = \begin{bmatrix} 2 & -1 \\ 3 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 4 & 2 & 1 \\ -1 & 7 & 5 \end{bmatrix}.$$

Partitioned matrices

Partitioned matrices are important and appear often in statistics.

We will not discuss much in this course, but you can read about the multiplication of partitioned matrices in Nicholson §2.3, p.73 "Block multiplication".

Gaussian elimination in R

```
# Install and load matlib package
> install.packages("matlib")  # if not already installed
> library(matlib)

# Define coefficient matrix and constant matrix
> A <- matrix(c(1,2,3,4,5,6), ncol=3)
> b <- c(-1,1)

# Recall that by default matrices are filled by column
> A
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6

# Apply Gaussian elimination to obtain the reduced row-echelon form
> gaussianElimination(A, b)
     [,1] [,2] [,3] [,4]
[1,]    1    0   -1  3.5
[2,]    0    1    2 -1.5

# Gaussian elimination is superior to other methods
# (but it requires that you interpret the output correctly)

> solve(A,b)  # can't find solutions
Error in solve.default(A, b) : 'a' (2 x 3) must be square

> qr.solve(A,b)  # returns only one solution
[1]  3.5 -1.5  0.0
```