Lecture 1 Introduction

Textbook

Students are strongly advised to acquire a copy of the Textbook:


Other editions can be used as well; the book is easily available on Amazon (this includes some very cheap used copies) and in Blackwell’s bookshop on campus.

These lecture notes should be treated only as indication of the content of the course; the textbook contains detailed explanations, worked out examples, etc.

About Homework

The Undergraduate Student Handbook

As a general rule for each hour in class you should spend two hours on independent study. This will include reading lecture notes and textbooks, working on exercises and preparing for coursework and examinations.

In respect of this course, MATH10212 Linear Algebra B, this means that students are expected to spend 8 (eight!) hours a week in private study of Linear Algebra.

Normally, homework assignments will consist of some odd numbered exercises from the sections of the Textbook covered in the lectures up to Wednesday in particular week. The Textbook contains answers to most odd numbered exercises (but the numbering of exercises might change from one edition to another).

Homework N (already posted on the webpage and BB9 space of the course) should be returned to Supervision Classes teachers early in Week N, for marking and discussion at supervision classes on Tuesday and Wednesday.

Communication

The Course Webpage is

http://www.maths.manchester.ac.uk/~avb/math10212-Linear-Algebra-B.html

The Course Webpage page is updated almost daily, sometimes several times a day. Refresh it (and files it is linked to) in your browser, otherwise you may miss the changes.

Email: Feel free to write to me with questions, etc., at the address alexandre.borovik@manchester.ac.uk but only from your university e-mail account.

Emails from Gmail, Hotmail, etc. automatically go to spam.

---

1http://www.maths.manchester.ac.uk/study/undergraduate/information-for-current-students/undergraduatestudenthandbook/teachingandlearning/timetabledclasses/.
What is Linear Algebra?

It is well-known that the total cost of a purchase of amounts $g_1, g_2, g_3$ of some goods at prices $p_1, p_2, p_3$, respectively, is an expression

$$p_1 g_1 + p_2 g_2 + p_3 g_3 = \sum_{i=1}^{3} p_i g_i.$$ 

Expressions of this kind,

$$a_1 x_1 + \cdots + a_n x_n$$

are called **linear forms in variables** $x_1, \ldots, x_n$ **with coefficients** $a_1, \ldots, a_n$.

- Linear Algebra studies the mathematics of linear forms.
- Over the course, we shall develop increasingly compact notation for operations of Linear Algebra. In particular, we shall discover that

$$p_1 g_1 + p_2 g_2 + p_3 g_3$$

can be very conveniently written as

$$\begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}$$

and then abbreviated

$$\begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix}^{T} G,$$

where

$$P = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix},$$

and $T$ (transposition) turns row vectors into column vectors, and vice versa:

$$P^{T} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}^{T} = \begin{bmatrix} p_1 & p_2 & p_3 \end{bmatrix},$$

and

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}^{T} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}.$$

- Physicists use even more short notation and, instead of

$$p_1 g_1 + p_2 g_2 + p_3 g_3 = \sum_{i=1}^{3} p_i g_i$$

write

$$p_1 g_1^1 + p_2 g_2^2 + p_3 g_3^3 = p_i g_i,$$

omitting the summation sign entirely. This particular trick was invented by Albert Einstein, of all people. I do not use “physics” tricks in my lectures, but am prepared to give a few additional lectures to physics students.

- **Warning:** Increasingly compact notation leads to increasingly compact and abstract language.
- Unlike, say, Calculus, Linear Algebra focuses more on the development of a special mathematics language rather than on procedures.

Systems of simultaneous linear equations

A **linear equation** in the variables $x_1, \ldots, x_n$ is an equation that can be written in the form

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b.$$
where \( b \) and the coefficients \( a_1, \ldots, a_n \) are real numbers. The subscript \( n \) can be any natural number.

A system of simultaneous linear equations is a collection of one or more linear equations involving the same variables, say \( x_1, \ldots, x_n \). For example,

\[
\begin{align*}
  x_1 + x_2 &= 3 \\
  x_1 - x_2 &= 1
\end{align*}
\]

We shall abbreviate the words “a system of simultaneous linear equations” just to “a linear system”.

A solution of the system is a list \((s_1, \ldots, s_n)\) of numbers that makes each equation a true identity when the values \( s_1, \ldots, s_n \) are substituted for \( x_1, \ldots, x_n \), respectively. For example, in the system above \((2, 1)\) is a solution.

The set of all possible solutions is called the solution set of the linear system.

Two linear systems are equivalent if they have the same solution set.
Lecture 2 Systems of linear equations: 
Elementary Operations [Lay 1.1]

A solution of the system is a list \((s_1, \ldots, s_n)\) of numbers that makes each equation a true identity when the values \(s_1, \ldots, s_n\) are substituted for \(x_1, \ldots, x_n\), respectively. For example, in the system above \((2, 1)\) is a solution.

The set of all possible solutions is called the solution set of the linear system.

Two linear systems are equivalent if they have the same solution set.

We shall be use the following elementary operations on systems of simultaneous linear equations:

- **Replacement** Replace one equation by the sum of itself and a multiple of another equation.

- **Interchange** Interchange two equations.

- **Scaling** Multiply all terms in a equation by a nonzero constant.

Note: The elementary operations are reversible.

**Theorem:** Elementary operations preserve solutions.

If a system of simultaneous linear equations is obtained from another system by elementary operations, then the two systems have the same solution set.

We shall prove later in the course that a system of linear equations has either

- infinitely many solutions,

under the assumption that the coefficients and solutions of the systems are real or complex numbers.

A system of linear equations is said to be consistent if it has solutions (either one or infinitely many), and a system in inconsistent if it has no solution.

**Solving a linear system**

The basic strategy is to replace one system with an equivalent system (that is, with the same solution set) which is easier to solve.

**Existence and uniqueness questions**

- Is the system consistent?

- If a solution exist, is it unique?

**Equivalence of linear systems**

- When are two linear systems equivalent?

**Checking solutions**

Given a solution of the system of linear equations, how to check that it is correct?
Row reduction and echelon forms [Lay 1.2]

Matrix notation

It is convenient to write coefficients of a linear system in the form of a matrix, a rectangular table. For example, the system

\[
\begin{align*}
x_1 - 2x_2 + 3x_3 &= 1 \\
x_1 + x_2 &= 2 \\
x_2 + x_3 &= 3
\end{align*}
\]

has the matrix of coefficients

\[
\begin{bmatrix}
1 & -2 & 3 \\
1 & 1 & 0 \\
0 & 1 & 1
\end{bmatrix}
\]

and the augmented matrix

\[
\begin{bmatrix}
1 & -2 & 3 & 1 \\
1 & 1 & 0 & 2 \\
0 & 1 & 1 & 3
\end{bmatrix}
\]

notice how the coefficients are aligned in columns, and how missing coefficients are replaced by 0.

The augmented matrix in the example above has 3 rows and 4 columns; we say that it is a \(3 \times 4\) matrix. Generally, a matrix with \(m\) rows and \(n\) columns is called an \(m \times n\) matrix.

Elementary row operations

Replacement Replace one row by the sum of itself and a multiple of another row.

Interchange Interchange two rows.

Scaling Multiply all entries in a row by a nonzero constant.

The two matrices are row equivalent if there is a sequence of elementary row operations that transforms one matrix into the other.

Note:

- The row operations are reversible.
- Row equivalence of matrices is an equivalence relation on the set of matrices.

Theorem: Row Equivalence.

If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

A nonzero row or column of a matrix is a row or column which contains at least one nonzero entry.

We can now formulate a theorem (to be proven later).

Theorem: Equivalence of linear systems.

Two linear systems are equivalent if and only if the augmented matrix of one of them can be obtained from the augmented matrix of another system by row operations and insertion / deletion of zero rows.
A nonzero row or column of a matrix is a row or column which contains at least one nonzero entry. A leading entry of a row is the leftmost nonzero entry (in a non-zero row).

**Definition.** A matrix is in *echelon form* (or *row echelon form*) if it has the following three properties:

1. All nonzero rows are above any row of zeroes.
2. Each leading entry of a row is in column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zeroes.

If, in addition, the following two conditions are satisfied,

4. All leading entries are equal 1.
5. Each leading 1 is the only non-zero entry in its column

then the matrix is in *reduced echelon form*.

An echelon matrix is a matrix in echelon form.

Any non-zero matrix can be row reduced (that, transformed by elementary row operations) into a matrix in echelon form (but the same matrix can give rise to different echelon forms).

**Examples.** The following is a schematic presentation of an echelon matrix:

\[
\begin{bmatrix}
\bullet & \ast & \ast & \ast & \ast \\
0 & \bullet & \ast & \ast & \ast \\
0 & 0 & 0 & \bullet & \ast
\end{bmatrix}
\]

and this is a reduced echelon matrix:

\[
\begin{bmatrix}
1 & 0 & \ast & 0 & \ast \\
0 & 1 & \ast & 0 & \ast \\
0 & 0 & 0 & 1 & \ast
\end{bmatrix}
\]

**Theorem 1.2.1: Uniqueness of the reduced echelon form.**

Each matrix is row equivalent to one and only one reduced echelon form.
Lecture 4 Solution of Linear Systems [Lay 1.2]

Definition. A pivot position in a matrix $A$ is a location in $A$ that corresponds to a leading 1 in the reduced echelon form of $A$. A pivot column is a column of $A$ that contains a pivot position.

Example for solving in the lecture (The Row Reduction Algorithm):

$$
\begin{bmatrix}
0 & 2 & 2 & 2 & 2 \\
1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 3 & 3 \\
1 & 1 & 1 & 2 & 2
\end{bmatrix}
$$

A pivot is a nonzero number in a pivot position which is used to create zeroes in the column below it.

A rule for row reduction:

1. Pick the leftmost non-zero column and in it the topmost non-zero entry; it is a pivot.
2. Using scaling, make the pivot equal 1.
3. Using replacement row operations, kill all non-zero entries in the column below the pivot.
4. Mark the row and column containing the pivot as pivoted.
5. Repeat the same with the matrix made of not pivoted yet rows and columns.
6. When this is over, interchange the rows making sure that the resulting matrix is in echelon form.
7. Using replacement row operations, kill all non-zero entries in the column above the pivot entries.

Solution of Linear Systems

When we converted the augmented matrix of a linear system into its reduced row echelon form, we can write out the entire solution set of the system.

Example. Let

$$
\begin{bmatrix}
1 & 0 & -5 & 1 \\
0 & 1 & 1 & 4 \\
0 & 0 & 0 & 0
\end{bmatrix}
$$

be the augmented matrix of a linear system; then the system is equivalent to

$$
\begin{align*}
x_1 - 5x_3 &= 1 \\
x_2 + x_3 &= 4 \\
0 &= 0
\end{align*}
$$

The variables $x_1$ and $x_2$ correspond to pivot columns in the matrix and are recalled basic variables (also leading or pivot variables). The other variable, $x_3$ is a free variable.

Free variables can be assigned arbitrary values and then leading variables expressed in terms of free variables:

$$
\begin{align*}
x_1 &= 1 + 5x_3 \\
x_2 &= 4 - x_3 \\
x_3 &\text{ is free}
\end{align*}
$$

Theorem 1.2.2: Existence and Uniqueness

A linear system is consistent if and only if the rightmost column of the augmented matrix is not a pivot column—that is, if and only if an echelon form of the augmented matrix has no row of the form

$$
\begin{bmatrix}
0 & \cdots & 0 & b
\end{bmatrix}
$$

with $b$ nonzero

If a linear system is consistent, then the solution set contains either
(i) a unique solution, when there are no free variables, or
(ii) infinitely many solutions, when there is at least one free variable.

Using row reduction to solve a linear system

1. Write the augmented matrix of the system.
2. Use the row reduction algorithm to obtain an equivalent augmented matrix in echelon form. Decide whether the system is consistent.
3. If the system is consistent, get the reduced echelon form.
4. Write the system of equations corresponding to the matrix obtained in Step 3.
5. Express each basic variable in terms of any free variables appearing in the equation.
Lecture 4 Vector equations [Lay 1.3]

A matrix with only one column is called a column vector, or simply a vector. 

\( \mathbb{R}^n \) is the set of all column vectors with \( n \) entries.

A row vector: a matrix with one row.

Two vectors are equal if and only if they have

- the same shape,
- the same number of rows,
- and their corresponding entries are equal.

The set of all vectors with \( n \) entries is denoted \( \mathbb{R}^n \).

The sum \( u + v \) of two vectors \( u \) and \( v \) in \( \mathbb{R}^n \) is obtained by adding corresponding entries in \( u \) and \( v \). For example in \( \mathbb{R}^2 \)

\[
\begin{bmatrix}
1 \\
2
\end{bmatrix}
+ 
\begin{bmatrix}
-1 \\
-1
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
1
\end{bmatrix}
\]

The scalar multiple \( cv \) of a vector \( v \) and a real number ("scalar") \( c \) is the vector obtained by multiplying each entry in \( v \) by \( c \). For example in \( \mathbb{R}^3 \),

\[
1.5 
\begin{bmatrix}
1 \\
0 \\
-3
\end{bmatrix}
= 
\begin{bmatrix}
1.5 \\
0 \\
-2
\end{bmatrix}
\]

The vector whose entries are all zeroes is called the zero vector and denoted \( \mathbf{0} \):

\[
\mathbf{0} = 
\begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

Operations with row vectors are defined in a similar way.

**Algebraic properties of \( \mathbb{R}^n \)**

For all \( u, v, w \in \mathbb{R}^n \) and all scalars \( c \) and \( d \):

1. \( u + v = v + u \)
2. \( (u + v) + w = u + (v + w) \)
3. \( u + 0 = 0 + u = u \)
4. \( u + (-u) = -u + u = 0 \)
5. \( c(u + v) = cu + cv \)
6. \( (c + d)u = cu + du \)
7. \( c(du) = (cd)u \)
8. \( 1u = u \)

(Here \( -u \) denotes \((-1)u\).)

**Linear combinations**

Given vectors \( v_1, v_2, \ldots, v_p \) in \( \mathbb{R}^n \) and scalars \( c_1, c_2, \ldots, c_p \), the vector

\[
y = c_1v_1 + \cdots + c_pv_p
\]

is called a linear combination of \( v_1, v_2, \ldots, v_p \) with weights \( c_1, c_2, \ldots, c_p \).

**Rewriting a linear system as a vector equation**

Consider an example: the linear system

\[
\begin{align*}
x_2 + x_3 &= 2 \\
x_1 + x_2 + x_3 &= 3 \\
x_1 + x_2 - x_3 &= 2
\end{align*}
\]

can be written as equality of two vectors:

\[
\begin{bmatrix}
x_2 + x_3 \\
x_1 + x_2 + x_3 \\
x_1 + x_2 - x_3
\end{bmatrix}
= 
\begin{bmatrix}
2 \\
3 \\
2
\end{bmatrix}
\]

which is the same as

\[
x_1 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}
\]
Let us write the matrix
\[
\begin{bmatrix}
0 & 1 & 1 & 2 \\
1 & 1 & 1 & 3 \\
1 & 1 & -1 & 2
\end{bmatrix}
\]
in a way that calls attention to its columns:
\[
\begin{bmatrix}
a_1 & a_2 & a_3 & b
\end{bmatrix}
\]

Denote
\[
a_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}
\]
and
\[
b = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix},
\]
then the vector equation can be written as
\[
x_1a_1 + x_2a_2 + x_3a_3 = b.
\]
Notice that to solve this equation is the same as

express \( b \) as a linear combination of \( a_1, a_2, a_3 \), and find all such expressions.

Therefore

solving a linear system is the same as finding an expression of the vector of the right part of the system as a linear combination of columns in its matrix of coefficients.

A vector equation
\[
x_1a_1 + x_2a_2 + \cdots + x_na_n = b.
\]
has the same solution set as the linear system whose augmented matrix is
\[
\begin{bmatrix}
a_1 & a_2 & \cdots & a_n & b
\end{bmatrix}
\]

In particular \( b \) can be generated by a linear combination of \( a_1, a_2, \ldots, a_n \) if and only if there is a solution of the corresponding linear system.

**Definition.** If \( v_1, \ldots, v_p \) are in \( \mathbb{R}^n \), then the set of all linear combination of \( v_1, \ldots, v_p \) is denoted by
\[
\text{Span}\{v_1, \ldots, v_p\}
\]
and is called the subset of \( \mathbb{R}^n \) spanned (or generated) by \( v_1, \ldots, v_p \); or the span of vectors \( v_1, \ldots, v_p \).

That is, \( \text{Span}\{v_1, \ldots, v_p\} \) is the collection of all vectors which can be written in the form
\[
c_1v_1 + c_2v_2 + \cdots + c_pv_p
\]
with \( c_1, \ldots, c_p \) scalars.

We say that vectors \( v_1, \ldots, v_p \) span \( \mathbb{R}^n \) if
\[
\text{Span}\{v_1, \ldots, v_p\} = \mathbb{R}^n
\]
We can reformulate the definition of span as follows: if \( a_1, \ldots, a_p \in \mathbb{R}^n \), then
\[
\text{Span}\{a_1, \ldots, a_p\} = \{b \in \mathbb{R}^n \text{ such that } [a_1 \cdots a_p | b] \text{ is the augmented matrix of a consistent system of linear equations}\},
\]
or, in simpler words,
\[
\text{Span}\{a_1, \ldots, a_p\} = \{b \in \mathbb{R}^n \text{ such that the system of equations } x_1a_1 + \cdots + x_pa_p = b \text{ has a solution}\}.
\]
Lecture 5: The matrix equation $Ax = b$ [Lay 1.4]

**Definition.** If $A$ is an $m \times n$ matrix, with columns $a_1, \ldots, a_n$, and if $x$ is in $\mathbb{R}^n$, then the **product of $A$ and $x$**, denoted $Ax$, is the linear combination of the columns of $A$ using the corresponding entries in $x$ as weights:

$$Ax = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

**Example.** The system

$$
\begin{align*}
x_2 + x_3 &= 2 \\
x_1 + x_2 + x_3 &= 3 \\
x_1 + x_2 - x_3 &= 2
\end{align*}
$$

was written as

$$x_1 a_1 + x_2 a_2 + x_3 a_3 = b,$$

where

$$a_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$

and

$$b = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}.$$  

In the matrix product notation it becomes

$$
\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}$$

or

$$Ax = b$$

where

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$  

**Theorem 1.4.3:** If $A$ is an $m \times n$ matrix, with columns $a_1, \ldots, a_n$, and if $x$ is in $\mathbb{R}^n$, the matrix equation

$$Ax = b$$

has the same solution set as the vector equation

$$x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = b$$

which has the same solution set as the system of linear equations whose augmented matrix is

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_n & b \end{bmatrix}.$$  

**Existence of solutions.** The equation $Ax = b$ has a solution if and only if $b$ is a linear combination of columns of $A$.

**Theorem 1.4.4:** Let $A$ be an $m \times n$ matrix. Then the following statements are equivalent.

(a) For each $b \in \mathbb{R}^m$, the equation $Ax = b$ has a solution.

(b) Each $b \in \mathbb{R}^m$ is a linear combination of columns of $A$.

(c) The columns of $A$ span $\mathbb{R}^m$.

(d) $A$ has a pivot position in every row.

**Row-vector rule for computing $Ax$.** If the product $Ax$ is defined then the $i$th entry in $Ax$ is the sum of products of corresponding entries from the row $i$ of $A$ and from the vector $x$.

**Theorem 1.4.5: Properties of the matrix-vector product $Ax$.**

If $A$ is an $m \times n$ matrix, $u, v \in \mathbb{R}^n$, and $c$ is a scalar, then

(a) $A(u + v) = Au + Av$;

(b) $A(cu) = c(Au)$. 
Homogeneous linear systems

A linear system is **homogeneous** if it can be written as

\[ Ax = 0. \]

A homogeneous system always has at least one solution \( x = 0 \) (trivial solution).

Therefore for homogeneous systems an important question is existence of a **nontrivial** solution, that is, a nonzero vector \( x \) which satisfies \( Ax = 0 \):

The homogeneous system \( Ax = b \) has a nontrivial solution if and only if the equation has at least one free variable.

**Example.**

\[
\begin{align*}
x_1 + 2x_2 - x_3 &= 0 \\
x_1 + 3x_3 + x_3 &= 0
\end{align*}
\]
Lecture 6: Solution sets of linear equations [Lay 1.5]

Nonhomogeneous systems
When a nonhomogeneous system has many solutions, the general solution can be written in parametric vector form as one vector plus an arbitrary linear combination of vectors that satisfy the corresponding homogeneous system.

Example.
\[
\begin{align*}
x_1 + 2x_2 - x_3 &= 0 \\
x_1 + 3x_3 + x_3 &= 5
\end{align*}
\]

Theorem 1.5.6: Suppose the equation \( Ax = b \) is consistent for some given \( b \), and \( p \) be a solution. Then the solution set of \( Ax = b \) is the set of all vectors of the form
\[
w = p + v_h,
\]
where \( v_h \) is any solution of the homogeneous equation \( Ax = 0 \).
Lecture 7: Linear independence [Lay 1.7]

**Definition.** An indexed set of vectors
\[ \{ v_1, \ldots, v_p \} \]
in \( \mathbb{R}^n \) is **linearly independent** if the vector equation
\[
x_1 v_1 + \cdots + x_p v_p = 0
\]
has only trivial solution.

The set
\[ \{ v_1, \ldots, v_p \} \]
in \( \mathbb{R}^n \) is **linearly dependent** if there exist weights \( c_1, \ldots, c_p \), **not all zero**, such that
\[
c_1 v_1 + \cdots + c_p v_p = 0
\]

**Linear independence of matrix columns.** The matrix equation
\[
Ax = 0
\]
where \( A \) is made of columns
\[
A = [ a_1 \, \cdots \, a_n ]
\]
can be written as
\[
x_1 a_1 + \cdots + x_n a_n = 0
\]
Therefore the columns of matrix \( A \) are linearly independent if and only if the equation
\[
Ax = 0
\]
has only the trivial solution.

A set of one vectors \( \{ v_1 \} \) is linearly dependent if \( v_1 = 0 \).

A set of two vectors \( \{ v_1, v_2 \} \) is linearly dependent if at least one of the vectors is a multiple of the other.

**Theorem 1.7.7: Characterisation of linearly dependent sets.** An indexed set
\[ S = \{ v_1, \ldots, v_p \} \]
of two or more vectors is linearly dependent if and only if at least one of the vectors in \( S \) is a linear combination of the others.

**Theorem 1.7.8: dependence of “big” sets.** If a set contains more vectors than entries in each vector, then the set is linearly dependent. Thus, any set
\[ \{ v_1, \ldots, v_p \} \]
in \( \mathbb{R}^n \) is linearly dependent if \( p > n \).
Lecture 8: Introduction to linear transformations [1.8]

Transformation. A transformation $T$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x}$ in $\mathbb{R}^n$ a vector $T(\mathbf{x})$ in $\mathbb{R}^m$.

The set $\mathbb{R}^n$ is called the **domain** of $T$, the set $\mathbb{R}^m$ is the **codomain** of $T$.

Matrix transformations. With every $m \times n$ matrix $A$ we can associated the transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$:

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$\mathbf{x} \mapsto Ax.$$  

In short:

$$T(\mathbf{x}) = Ax.$$

The range of a matrix transformation. The **range** of $T$ is the set of all linear combinations of the columns of $A$.

Indeed, this can be immediately seen from the fact that each image $T(\mathbf{x})$ has the form

$$T(\mathbf{x}) = Ax = x_1a_1 + \cdots + x_na_n.$$

**Definition: Linear transformations.** A transformation

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

is **linear** if:

- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$;

- $T(c\mathbf{u}) = cT(\mathbf{u})$ for all vectors $\mathbf{u}$ and all scalars $c$.

Properties of linear transformations. If $T$ is a linear transformation then

$$T(\mathbf{0}) = \mathbf{0}$$

and

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v}).$$

The identity matrix. An $n \times n$ matrix with 1’s on the diagonal and 0’s elsewhere is called the **identity** matrix $I_n$. For example,

$$I_1 = [1], \quad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$  

The columns of the identity matrix $I_n$ will be denoted $e_1, e_2, \ldots, e_n$.

For example, in $\mathbb{R}^3$

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$  

Observe that if $\mathbf{x}$ is an arbitrary vector in $\mathbb{R}^n$,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix},$$

then

$$\mathbf{x} = x_1e_1 + x_2e_2 + \cdots + x_ne_n.$$  

The identity transformation. It is easy to check that

$$I_n\mathbf{x} = \mathbf{x} \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$
Therefore the linear transformation associated with the identity matrix is the identity transformation of \(\mathbb{R}^n\):

\[
\begin{align*}
\mathbb{R}^n & \rightarrow \mathbb{R}^n \\
x & \mapsto x
\end{align*}
\]

The matrix of a linear transformation [Lay 1.9]

**Theorem 1.9.10: The matrix of a linear transformation.** Let 

\[
T : \mathbb{R}^n \rightarrow \mathbb{R}^m
\]

be a linear transformation. Then there exists a unique matrix \(A\) such that

\[
T(x) = Ax \quad \text{for all } x \in \mathbb{R}^n.
\]

In fact, \(A\) is the \(m \times n\) matrix whose \(j\)th column is the vector \(T(e_j)\) where \(e_j\) is the \(j\)th column of the identity matrix in \(\mathbb{R}^n\):

\[
A = [T(e_1) \quad \cdots \quad T(e_n)]
\]

**Proof.** First we express \(x\) in terms of \(e_1, \ldots, e_n\):

\[
x = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n
\]

and compute, using definition of linear transformation

\[
T(x) = T(x_1 e_1 + x_2 e_2 + \cdots + x_n e_n)
\]

\[
= T(x_1 e_1) + \cdots + T(x_n e_n)
\]

\[
= x_1 T(e_1) + \cdots + x_n T(e_n)
\]

and then switch to matrix notation:

\[
\begin{bmatrix}
x_1 \\
\vdots \\
x_n
\end{bmatrix}
= [T(e_1) \cdots T(e_n)]
\]

\[
= Ax.
\]

The matrix \(A\) is called the **standard matrix** for the linear transformation \(T\).

**Definition.** A transformation 

\[
T : \mathbb{R}^n \rightarrow \mathbb{R}^m
\]

is **onto** \(\mathbb{R}^m\) if each \(b \in \mathbb{R}^m\) is the image of at least one \(x \in \mathbb{R}^n\).

A transformation \(T\) is **one-to-one** if each \(b \in \mathbb{R}^m\) is the image of at most one \(x \in \mathbb{R}^n\).

**Theorem: One-to-one transformations: a criterion.** A linear transformation 

\[
T : \mathbb{R}^n \rightarrow \mathbb{R}^m
\]

is one-to-one if and only if the equation

\[
T(x) = 0
\]

has only the trivial solution.

**Theorem: “One-to-one” and “onto” in terms of matrices.** Let 

\[
T : \mathbb{R}^n \rightarrow \mathbb{R}^m
\]

be linear transformation and let \(A\) be the standard matrix for \(T\). Then:

- \(T\) maps \(\mathbb{R}^n\) onto \(\mathbb{R}^m\) if and only if the columns of \(A\) span \(\mathbb{R}^m\).

- \(T\) is one-to-one if and only if the columns of \(A\) are linearly independent.
Lecture 9: Matrix operations: Addition [Lay 2.1]

Labeling of matrix entries. Let $A$ be an $m \times n$ matrix.

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}$$

Notice that in $a_{ij}$ the first subscript $i$ denotes the row number, the second subscript $j$ the column number of the entry $a_{ij}$. In particular, the column $a_j$ is

$$a_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

Diagonal matrices, zero matrices. The diagonal entries in $A$ are

$$a_{11}, a_{22}, a_{33}, \ldots$$

For example, the diagonal entries of the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

are 1, 5, and 9.

A square matrix is a matrix with equal numbers of rows and columns.

A diagonal matrix is a square matrix whose non-diagonal entries are zeroes.

Matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \begin{bmatrix} \pi & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix}$$

are all diagonal. The identity matrices

$$\begin{bmatrix} 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \ldots$$

are diagonal.

Zero matrix. By definition, 0 is a $m \times n$ matrix whose entries are all zero. For example, matrices

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are zero matrices. Notice that zero square matrices, like

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are diagonal!

Sums. If

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

and

$$B = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}$$

are $m \times n$ matrices then we define the sum $A + B$ as

$$A + B = \begin{bmatrix} a_1 + b_1 & a_2 + b_2 & \cdots & a_n + b_n \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & \cdots & a_{1j} + b_{1j} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} + b_{i1} & \cdots & a_{ij} + b_{ij} & \cdots & a_{in} + b_{in} \\ \vdots & & \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mj} + b_{mj} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$$
Scalar multiple. If \( c \) is a scalar then we define

\[
cA = \begin{bmatrix} ca_1 & ca_2 & \cdots & ca_n \end{bmatrix}
\]

Theorem 2.1.1: Properties of matrix addition. Let \( A, B, \) and \( C \) be matrices of the same size and \( r \) and \( s \) be scalars.

1. \( A + B = B + A \)
2. \( (A + B) + C = A + (B + C) \)
3. \( A + 0 = A \)
4. \( r(A + B) = rA + rB \)
5. \( (r + s)A = rA + sA \)
6. \( r(sA) = (rs)A. \)
Lecture 10: Matrix multiplication [Lay 2.1]

Composition of linear transformations. Let $B$ be an $m \times n$ matrix and $A$ an $p \times m$ matrix. They define linear transformations

$$T : \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad x \mapsto Bx$$

and

$$S : \mathbb{R}^m \rightarrow \mathbb{R}^p, \quad y \mapsto Ay.$$

Their composition

$$(S \circ T)(x) = S(T(x))$$

is a linear transformation

$$S \circ T : \mathbb{R}^n \rightarrow \mathbb{R}^p.$$

From the previous lecture, we know that linear transformations are given by matrices. What is the matrix of $S \circ T$?

Multiplication of matrices. To answer the above question, we need to compute $A(Bx)$ in matrix form.

Write $x$ as

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

and observe

$$Bx = x_1 b_1 + \cdots + x_n b_n.$$ 

Hence

$$A(Bx) = A(x_1 b_1 + \cdots + x_n b_n) = A(x_1 b_1) + \cdots + A(x_n b_n) = x_1 A(b_1) + \cdots + x_n A(b_n) = [A b_1 \ A b_2 \ \cdots \ A b_n] x$$

Therefore multiplication by the matrix

$$C = [A b_1 \ A b_2 \ \cdots \ A b_n]$$

transforms $x$ into $A(Bx)$. Hence $C$ is the matrix of the linear transformation $S \circ T$ and it will be natural to call $C$ the product of $A$ and $B$ and denote

$$C = A \cdot B$$

(but the multiplication symbol “.” is frequently skipped).

Definition: Matrix multiplication. If $A$ is an $p \times m$ matrix and $B$ is an $m \times n$ matrix with columns $b_1, \ldots, b_n$ then the product $AB$ is the $p \times n$ matrix whose columns are

$$Ab_1, \ldots, Ab_n :$$

$$AB = A \begin{bmatrix} b_1 & \cdots & b_n \end{bmatrix} = \begin{bmatrix} Ab_1 & \cdots & Ab_n \end{bmatrix}.$$ 

Columns of $AB$. Each column $Ab_j$ of $AB$ is a linear combination of columns of $A$ with weights taken from the $j$th column of $B$:

$$Ab_j = A \begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix} \begin{bmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{bmatrix} = b_{1j} a_1 + \cdots + b_{mj} a_m$$

Mnemonic rules

$$[m \times n \text{ matrix}] \cdot [n \times p \text{ matrix}] = [m \times p \text{ matrix}]$$

$$\text{column}_j(AB) = A \cdot \text{column}_j(B)$$

$$\text{row}_i(AB) = \text{row}_i(A) \cdot B$$

Theorem 2.1.2: Properties of matrix multiplication. Let $A$ be an $m \times n$ matrix and let $B$ and $C$ be matrices for which indicated sums and products are defined. Then the following identities are true:
1. \( A(BC) = (AB)C \)
2. \( A(B + C) = AB + AC \)
3. \( (B + C)A = BA + CA \)
4. \( r(AB) = (rA)B = A(rB) \) for any scalar \( r \)
5. \( I_mA = A = AI_n \)

**Powers of matrix.** As it is usual in algebra, we define, for a square matrix \( A \),

\[ A^k = A \cdots A \quad (k \text{ times}) \]

If \( A \neq 0 \) then we set

\[ A^0 = I \]

**The transpose of a matrix.** The transpose \( A^T \) of an \( m \times n \) matrix \( A \) is the \( n \times m \) matrix whose rows are formed from corresponding columns of \( A \):

\[
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6
\end{bmatrix}^T = 
\begin{bmatrix}
1 & 4 \\
2 & 5 \\
3 & 6
\end{bmatrix}
\]

**Theorem 2.1.3: Properties of transpose.** Let \( A \) and \( B \) denote matrices whose sizes are appropriate for the following sums and products. Then we have:

1. \((A^T)^T = A\)
2. \((A + B)^T = A^T + B^T\)
3. \((rA)^T = r(A^T)\) for any scalar \( r \)
4. \((AB)^T = B^T A^T\)
Lecture 11: The inverse of a matrix [Lay 2.2]

Invertible matrices

An $n \times n$ matrix $A$ is invertible if there is an $n \times n$ matrix $C$ such that

$$CA = I \text{ and } AC = I$$

$C$ is called the inverse of $A$.

The inverse of $A$, if exists, is unique (!) and is denoted $A^{-1}$:

$$A^{-1}A = I \text{ and } AA^{-1} = I.$$ 

Singular matrices. A non-invertible matrix is called a singular matrix. An invertible matrix is nonsingular.

Theorem 2.2.4: Inverse of a $2 \times 2$ matrix. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

If $ad - bc \neq 0$ then $A$ is invertible and

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

The quantity $ad - bc$ is called the determinant of $A$:

$$\det A = ad - bc.$$ 

Theorem 2.2.5: Solving matrix equations. If $A$ is an invertible $n \times n$ matrix, then for each $b \in \mathbb{R}^n$, the equation $Ax = b$ has the unique solution

$$x = A^{-1}b.$$ 

Theorem 2.2.6: Properties of invertible matrices.

(a) If $A$ is an invertible matrix, then $A^{-1}$ is also invertible and

$$(A^{-1})^{-1} = A.$$ 

(b) If $A$ and $B$ are $n \times n$ invertible matrices, then so is $AB$, and

$$(AB)^{-1} = B^{-1}A^{-1}.$$ 

(c) If $A$ is an invertible matrix, then so is $A^T$, and

$$(A^T)^{-1} = (A^{-1})^T.$$
Lecture 12: Characterizations of invertible matrices.

Definition: Elementary matrices. An elementary matrix is a matrix obtained by performing a single elementary row operation on an identity matrix.

If an elementary row transformation is performed on an \( n \times n \) matrix \( A \), the resulting matrix can be written as \( EA \), where the \( m \times m \) matrix is made by the same row operations on \( I_m \).

Each elementary matrix \( E \) is invertible.

The inverse of \( E \) is the elementary matrix of the same type that transforms \( E \) back into \( I \).

Theorem 2.2.7: Invertible matrices. An \( n \times n \) matrix \( A \) is invertible if and only if \( A \) is row equivalent to \( I_n \), and in this case, any sequence of elementary row operations that reduces \( A \) to \( I_n \) also transforms \( I_n \) into \( A^{-1} \).

Computation of inverses.

- Form the augmented matrix \([A \ I]\) and row reduce it.
- If \( A \) is row equivalent to \( I \), then \([A \ I]\) is row equivalent to \([I \ A^{-1}]\).
- Otherwise \( A \) has no inverse.

The Invertible Matrix Theorem 2.3.8: For an \( n \times n \) matrix \( A \), the following are equivalent:

(a) \( A \) is invertible.
(b) \( A \) is row equivalent to \( I_n \).
(c) \( A \) has \( n \) pivot positions.
(d) \( Ax = 0 \) has only the trivial solution.
(e) The columns of \( A \) are linearly independent.
(f) The linear transformation \( x \mapsto Ax \) is one-to-one.
(g) \( Ax = b \) has at least one solution for each \( b \in \mathbb{R}^n \).
(h) The columns of \( A \) span \( \mathbb{R}^n \).
(i) \( x \mapsto Ax \) maps \( \mathbb{R}^n \) onto \( \mathbb{R}^n \).
(j) There is an \( n \times n \) matrix \( C \) such that \( CA = I \).
(k) There is an \( n \times n \) matrix \( D \) such that \( AD = I \).
(l) \( A^T \) is invertible.

One-sided inverse is the inverse. Let \( A \) and \( B \) be square matrices.

If \( AB = I \) then both \( A \) and \( B \) are invertible and

\[
B = A^{-1} \quad \text{and} \quad A = B^{-1}.
\]

Theorem 2.3.9: Invertible linear transformations. Let

\[
T : \mathbb{R}^n \longrightarrow \mathbb{R}^n
\]

be a linear transformation and \( A \) its standard matrix.

Then \( T \) is invertible if and only if \( A \) is an invertible matrix.

In that case, the linear transformation

\[
S(x) = A^{-1}x
\]

is the only transformation satisfying

\[
S(T(x)) = x \quad \text{for all } x \in \mathbb{R}^n
\]
\[
T(S(x)) = x \quad \text{for all } x \in \mathbb{R}^n
\]
Lecture 13 Partitioned matrices [Lay 2.4]

Example: Partitioned matrix

\[
A = \begin{bmatrix}
1 & 2 & a \\
3 & 4 & b \\
p & q & z
\end{bmatrix} = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix}
\]

\(A\) is a 3 \(\times\) 3 matrix which can be viewed as a 2 \(\times\) 2 partitioned (or block) matrix with blocks

\[
A_{11} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},
A_{12} = \begin{bmatrix} a \\ b \end{bmatrix},
A_{21} = \begin{bmatrix} p & q \end{bmatrix},
A_{22} = \begin{bmatrix} z \end{bmatrix}
\]

Addition of partitioned matrices. If matrices \(A\) and \(B\) are of the same size and partitioned the same way, they can be added block-by-block.

Similarly, partitioned matrices can be multiplied by a scalar blockwise.

Multiplication of partitioned matrices. If the column partition of \(A\) matches the row partition of \(B\) then \(AB\) can be computed by the usual row-column rule, with blocks treated as matrix entries.

\[
A = \begin{bmatrix}
1 & 2 & a \\
3 & 4 & b \\
p & q & z
\end{bmatrix},
B = \begin{bmatrix}
\alpha & \beta \\
\gamma & \delta \\
\Gamma & \Delta
\end{bmatrix}
\]

Example

\[
A = \begin{bmatrix}
1 & 2 & a \\
3 & 4 & b \\
p & q & z
\end{bmatrix},
B = \begin{bmatrix}
\alpha & \beta \\
\gamma & \delta \\
\Gamma & \Delta
\end{bmatrix}
\]

\[
AB = \begin{bmatrix}
1 & 2 & a \\
3 & 4 & b \\
p & q & z
\end{bmatrix} \begin{bmatrix}
\alpha & \beta \\
\gamma & \delta \\
\Gamma & \Delta
\end{bmatrix}
\]

Theorem 2.4.10: Column-Row Expansion of \(AB\). If \(A\) is \(m \times n\) and \(B\) is \(n \times p\) then

\[
AB = \begin{bmatrix}
\text{row}_1(B) \\
\text{row}_2(B) \\
\vdots \\
\text{row}_n(B)
\end{bmatrix}
\]

\[
= \text{col}_1(A)\text{row}_1(B) + \cdots + \text{col}_n(A)\text{row}_n(B)
\]

Example Partitioned matrices naturally arise in analysis of systems of simultaneous linear equations. Consider a system of equations:

\[
\begin{align*}
x_1 + 2x_2 + 3x_3 + 4x_4 &= 1 \\
x_1 + 3x_2 + 3x_3 + 4x_4 &= 2
\end{align*}
\]
In matrix form, it looks as
\[
\begin{bmatrix}
1 & 2 & 3 & 4 \\
1 & 3 & 3 & 4
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix} =
\begin{bmatrix}
1 \\
2
\end{bmatrix}
\]

Denote
\[
A = \begin{bmatrix}
1 & 2 \\
1 & 3
\end{bmatrix}
\]
\[
C = \begin{bmatrix}
3 & 4 \\
3 & 4
\end{bmatrix}
\]
\[
Y_1 = \begin{bmatrix}
x_1 \\
x_2
\end{bmatrix}
\]
\[
Y_2 = \begin{bmatrix}
x_3 \\
x_4
\end{bmatrix}
\]
\[
B = \begin{bmatrix}
1 \\
2
\end{bmatrix}
\]

then we can rewrite the matrices involved in the matrix equation as partitioned matrices:
\[
\begin{bmatrix}
A & C
\end{bmatrix}
\begin{bmatrix}
Y_1 \\
Y_2
\end{bmatrix} = B
\]

and, multiplying matrices as partitioned matrices,
\[
AY_1 + CY_2 = B.
\]

Please notice that matrix \(A\) is invertible – at this stage it should be an easy mental calculation for you. Multiplying the both parts of the last equation by \(A^{-1}\), we get
\[
Y_1 + A^{-1}CY_2 = A^{-1}B
\]

and
\[
Y_1 = A^{-1}B - A^{-1}CY_2
\]

The entries of
\[
Y_2 = \begin{bmatrix}
x_3 \\
x_4
\end{bmatrix}
\]

can be taken for free parameters: \(x_3 = t\), \(x_4 = s\), and \(x_1\) and \(x_2\) expressed as
\[
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = A^{-1}B - A^{-1}C \begin{bmatrix}
t \\
s
\end{bmatrix}
\]

Observe that
\[
A^{-1} = \begin{bmatrix}
3 & -2 \\
-1 & 1
\end{bmatrix}
\]

and
\[
A^{-1}C = \begin{bmatrix}
3 & -2 \\
-1 & 1
\end{bmatrix} \cdot \begin{bmatrix}
3 & 4 \\
3 & 4
\end{bmatrix} = \begin{bmatrix}
3 & 4 \\
0 & 0
\end{bmatrix}
\]

which gives us an answer:
\[
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = \begin{bmatrix}
3 & -2 \\
-1 & 1
\end{bmatrix} \begin{bmatrix}
1 \\
2
\end{bmatrix} - \begin{bmatrix}
3 & 4 \\
3 & 4
\end{bmatrix} \begin{bmatrix}
t \\
s
\end{bmatrix} = \begin{bmatrix}
-1 \\
1
\end{bmatrix} - \begin{bmatrix}
3t + 4s \\
0
\end{bmatrix}
\]
\[
\begin{bmatrix}
x_3 \\
x_4
\end{bmatrix} = \begin{bmatrix}
t \\
s
\end{bmatrix}
\]

We may wish to bring it in a more traditional form:
\[
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix} = \begin{bmatrix}
-1 & 0 & -3t - 4s \\
1 & t & 0 \\
0 & 1 & -4 \\
0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
-1 \\
1 \\
0 \\
0
\end{bmatrix} + \begin{bmatrix}
-3t - 4s \\
t \\
0 \\
0
\end{bmatrix}
\]

Of course, when we solving a singles system of equations, the formulae of the kind
\[
Y_1 = A^{-1}B - A^{-1}CY_2
\]
do not save time and effort, but in many applications you have to solve millions of systems of linear equations with the same matrix of coefficients but different right-hand parts, and in this situation shortcuts based on compound matrices become very useful.
Lecture 14: Subspaces of $\mathbb{R}^n$ [Lay 2.8]

Subspace. A subspace of $\mathbb{R}^n$ is any set $H$ that has the properties

(a) $0 \in H$.

(b) For each vectors $u, v \in H$, the sum $u + v \in H$.

(c) For each vector $u \in H$ and each scalar $c$, $cu \in H$.

$\mathbb{R}^n$ is a subspace of itself.

$\{0\}$ is a subspace, called the zero subspace.

Span. $\text{Span}\{v_1, \ldots, v_p\}$ is the subspace spanned (or generated) by $v_1, \ldots, v_p$.

The column space of a matrix $A$ is the set $\text{Col} A$ of all linear combinations of columns of $A$.

The null space of a matrix $A$ is the set $\text{Nul} A$ of all solutions to the homogeneous system

$$Ax = 0$$

**Theorem 2.8.12:** The null space. The null space of an $m \times n$ matrix $A$ is a subspace of $\mathbb{R}^n$.

Equivalently, the set of all solutions to a system

$$Ax = 0$$

of $m$ homogeneous linear equations in $n$ unknowns is a subspace of $\mathbb{R}^n$.

Basis of a subspace. A basis of a subspace $H$ of $\mathbb{R}^n$ is a linearly independent set in $H$ that spans $H$.

**Theorem 2.8.13:** Pivot columns of a matrix $A$ form a basis of $\text{Col} A$.

Dimension and rank [Lay 2.9]

**Theorem:** Given a basis $b_1, \ldots, b_p$ in a subspace $H$ of $\mathbb{R}^n$, every vector $u \in H$ is uniquely expressed as a linear combination of $b_1, \ldots, b_p$.

Solving homogeneous systems. Finding a parametric solution of a system of homogeneous linear equations

$$Ax = 0$$

means to find a basis of $\text{Nul} A$.

Coordinate system. Let

$$B = \{b_1, \ldots, b_p\}$$

be a basis for a subspace $H$.

For each $x \in H$, the coordinates of $x$ relative to the basis $B$ are the weights $c_1, \ldots, c_p$ such that

$$x = c_1 b_1 + \cdots + c_p b_p.$$ 

The vector

$$[x]_B = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

is the coordinate vector of $x$.

**Dimension.** The dimension of a nonzero subspace $H$, denoted by $\dim H$, is the number of vectors in any basis of $H$.

The dimension of the zero subspace $\{0\}$ is defined to be 0.

**Rank of a matrix.** The rank of a matrix $A$, denoted by $\text{rank} A$, is the dimen-
sion of the column space \( \text{Col } A \) of \( A \):

\[
\text{rank } A = \dim \text{Col } A
\]

**The Rank Theorem 2.9.14:** If a matrix \( A \) has \( n \) columns,

\[
\text{rank } A + \dim \text{Nul } A = n.
\]

**The Basis Theorem 2.9.15:** Let \( H \) be a \( p \)-dimensional subspace of \( \mathbb{R}^n \).

- Any linearly independent set of exactly \( p \) elements in \( H \) is automatically a basis for \( H \).

- Also, any set of \( p \) elements of \( H \) that spans \( H \) is automatically a basis of \( H \).

For example, this means that the vectors

\[
\begin{bmatrix}
1 \\
0 \\
0
\end{bmatrix}, \quad
\begin{bmatrix}
2 \\
3 \\
0
\end{bmatrix}, \quad
\begin{bmatrix}
4 \\
5 \\
6
\end{bmatrix}
\]

form a basis of \( \mathbb{R}^3 \): there are three of them and they are linearly independent (the latter should be obvious to students at this stage).

**The Invertible Matrix Theorem (continued)** Let \( A \) be an \( n \times n \) matrix. Then the following statements are each equivalent to the statement that \( A \) is an invertible matrix.

(m) The columns of \( A \) form a basis of \( \mathbb{R}^n \).

(n) \( \text{Col } A = \mathbb{R}^n \).

(o) \( \dim \text{Col } A = n \).

(p) \( \text{rank } A = n \).

(q) \( \text{Nul } A = \{0\} \).

(r) \( \dim \text{Nul } A = 0 \).
Lecture 15: Introduction to determinants [Lay 3.1]

The determinant det $A$ of a square matrix $A$ is a certain number assigned to the matrix; it is defined recursively, that is, we define first determinants of matrices of sizes $1 \times 1$, $2 \times 2$, and $3 \times 3$, and then supply a formula which expresses determinants of $n \times n$ matrices in terms of determinants of $(n-1) \times (n-1)$ matrices.

The determinant of a $1 \times 1$ matrix

$A = [a_{11}]$ is defined simply as being equal its only entry $a_{11}$:

$$\det [a_{11}] = a_{11}.$$

The determinant of a $2 \times 2$ matrix

$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is defined by the formula

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

The determinant of a $3 \times 3$ matrix

$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$

is defined by the formula

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11} \cdot \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \cdot \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \cdot \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}.$$

Example. The determinant det $A$ of the matrix

$A = \begin{bmatrix} 2 & 0 & 7 \\ 0 & 1 & 0 \\ 1 & 0 & 4 \end{bmatrix}$

equals

$$2 \cdot \det \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} + 7 \cdot \det \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

which further simplifies as

$$2 \cdot 4 + 7 \cdot (-1) = 8 - 7 = 1.$$

Submatrices. By definition, the submatrix $A_{ij}$ is obtained from the matrix $A$ by crossing out row $i$ and column $j$.

For example, if

$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

then

$$A_{22} = \begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix}, \quad A_{31} = \begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}$$

Recursive definition of determinant. For $n \geq 3$, the determinant of an $n \times n$ matrix $A$ is defined as the expression

$$\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{1+n}a_{1n} \det A_{1n}$$

$$= \sum_{j=1}^{n} (-1)^{1+j}a_{1j} \det A_{1j}$$
which involves the determinants of smaller \((n - 1) \times (n - 1)\) submatrices \(A_{ij}\), which, in their turn, can be evaluated by a similar formula which reduces the calculations to the \((n - 2) \times (n - 2)\) case, and can be repeated all the way down to determinants of size \(2 \times 2\).

**Cofactors.** The \((i,j)\)-cofactor of \(A\) is

\[
C_{ij} = (-1)^{i+j} \det A_{ij}
\]

Then

\[
\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}
\]

is cofactor expansion across the first row of \(A\).

**Theorem 3.1.1: Cofactor expansion.** For any row \(i\), the result of the cofactor expansion across the row \(i\) is the same:

\[
\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}
\]

The chess board pattern of signs in cofactor expansions. The signs

\[
(-1)^{i+j}
\]

which appear in the formula for cofactors, form the easy-to-recognise and easy-to-remember “chess board” pattern:

\[
\begin{pmatrix}
+ & - & + & \cdots \\
- & + & - & \\
+ & - & + & \\
\vdots & & & \\
\end{pmatrix}
\]

**Theorem 3.1.2: The determinant of a triangular matrix.** If \(A\) is a triangular \(n \times n\) matrix then \(\det A\) is the product of the diagonal entries of \(A\).

**Proof:** This proof contains more details than the one given in the textbook. It is based on induction on \(n\), which, to avoid the use of “general” notation is illustrated by a simple example. Let

\[
A = \begin{bmatrix}
a_{11} & 0 & 0 & 0 \\
a_{21} & a_{22} & 0 & 0 \\
a_{31} & a_{32} & a_{33} & 0 \\
a_{41} & a_{42} & a_{43} & a_{44} \\
a_{51} & a_{52} & a_{53} & a_{54} & a_{55}
\end{bmatrix}
\]

All entries in the first row of \(A\)–with possible exception of \(a_{11}\)–are zeroes,

\[
a_{12} = \cdots = a_{15} = 0,
\]

therefore in the formula

\[
\det A = a_{11} \det A_{11} + \cdots + a_{55} \det A_{55}
\]

all summand with possible exception of the first one,

\[
a_{11} \det A_{11},
\]

are zeroes, and therefore

\[
\det A = a_{11} \det A_{11} = a_{11} \det \begin{bmatrix}
a_{22} & 0 & 0 & 0 \\
a_{32} & a_{33} & 0 & 0 \\
a_{42} & a_{43} & a_{44} & 0 \\
a_{52} & a_{53} & a_{54} & a_{55}
\end{bmatrix}
\]

But the smaller matrix \(A_{11}\) is also lower triangle, and therefore we can conclude by induction that

\[
\det \begin{bmatrix}
a_{22} & 0 & 0 & 0 \\
a_{32} & a_{33} & 0 & 0 \\
a_{42} & a_{43} & a_{44} & 0 \\
a_{52} & a_{53} & a_{54} & a_{55}
\end{bmatrix} = a_{22} \cdots a_{55}
\]

hence

\[
\det A = a_{11} \cdot (a_{22} \cdots a_{55}) = a_{11}a_{22} \cdots a_{55}
\]

is the product of diagonal entries of \(A\).

The basis of induction is the case \(n = 2\); of course, in that case

\[
\det \begin{bmatrix}
a_{11} & 0 \\
a_{21} & a_{22}
\end{bmatrix} = a_{11}a_{22} - 0 \cdot a_{21} = a_{11}a_{22}
\]
is the product of the diagonal entries of $A$.

**Corollary.** The determinant of a diagonal matrix equals the product of its diagonal elements:

$$\begin{vmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & d_n \end{vmatrix} = d_1 d_2 \cdots d_n.$$   

**Corollary.** The determinant of the identity matrix equals 1:

$$\det I_n = 1.$$  

**Corollary.** The determinant of the zero matrix equals 0:

$$\det 0 = 0.$$
Lectures 15–16: Properties of determinants [Lay 3.2]

**Theorem 3.2.3: Row Operations.**
Let $A$ be a square matrix.

(a) If a multiple of one row of $A$ is added to another row to produce a matrix $B$, then
\[
\det B = \det A.
\]

(b) If two rows of $A$ are swapped to produce $B$, then
\[
\det B = -\det A.
\]

(c) If one row of $A$ is multiplied by $k$ to produce $B$, then
\[
\det B = k \det A.
\]

**Theorem 3.2.5: The determinant of the transposed matrix.** If $A$ is an $n \times n$ matrix then
\[
\det A^T = \det A.
\]

**Example.** Let
\[
A = \begin{bmatrix}
1 & 2 & 0 \\
0 & 1 & 0 \\
0 & 3 & 4
\end{bmatrix},
\]
then
\[
A^T = \begin{bmatrix}
1 & 0 & 0 \\
2 & 1 & 3 \\
0 & 0 & 4
\end{bmatrix}
\]
and
\[
\det A = 1 \cdot \det \begin{bmatrix}
1 & 0 \\
3 & 4
\end{bmatrix}
- 2 \cdot \det \begin{bmatrix}
0 & 0 \\
0 & 4
\end{bmatrix}
+ 0 \cdot \det \begin{bmatrix}
0 & 1 \\
0 & 3
\end{bmatrix}
= 1 \cdot 4 - 2 \cdot 0 + 0 \cdot 0
= 4.
\]

Similarly
\[
\det A^T = 1 \cdot \det \begin{bmatrix}
1 & 3 \\
0 & 4
\end{bmatrix} + 0 + 0
= 1 \cdot 4 = 4;
\]
we got the same value.

**Corollary: Cofactor expansion across a column.** For any column $j$,
\[
\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}
\]

**Example.** We have already computed the determinant of the matrix
\[
A = \begin{bmatrix}
2 & 0 & 7 \\
0 & 1 & 0 \\
1 & 0 & 4
\end{bmatrix}
\]
by expansion across the first row. Now we do it by expansion across the third column:
\[
\det A = 7 \cdot (-1)^{1+3} \cdot \det \begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix}
+ 0 \cdot (-1)^{2+3} \cdot \det \begin{bmatrix}
2 & 0 \\
1 & 0
\end{bmatrix}
+ 4 \cdot (-1)^{3+3} \cdot \det \begin{bmatrix}
2 & 0 \\
0 & 1
\end{bmatrix}
= -7 - 0 + 8
= 1.
\]

**Proof:** Columns of $A$ are rows of $A^T$ which has the same determinant as $A$.

**Corollary: Column operations.** Let $A$ be a square matrix.

(a) If a multiple of one column of $A$ is added to another column to produce a matrix $B$, then
\[
\det B = \det A.
\]
(b) If two columns of $A$ are swapped to produce $B$, then
\[ \det B = -\det A. \]

(c) If one column of $A$ is multiplied by $k$ to produce $B$, then
\[ \det B = k \det A. \]

**Proof:** Column operations on $A$ are row operation of $A^T$ which has the same determinant as $A$.

**Computing determinants.** In computations by hand, the quickest method of computing determinants is to work with both columns and rows:

- If a row or a column has a convenient scalar factor, take it out of the determinant.
- If convenient, swap two rows or two columns— but do not forget to change the sign of the determinant.
- Adding to a row (column) scalar multiples of other rows (columns), simplify the matrix to create a row (column) with just one nonzero entry.
- Expand across this row (column).
- Repeat until you get the value of the determinant.

**Example.**

\[
\begin{vmatrix}
1 & 0 & 3 \\
1 & 3 & 1 \\
1 & 1 & -1 \\
\end{vmatrix}
\]

\[ C_3 = 3C_1 \]
\[ \det \begin{bmatrix}
1 & 0 & 0 \\
1 & 3 & -2 \\
1 & 1 & -4 \\
\end{bmatrix} \]
\[ 2 \text{ out of } C_3 \]
\[ \Rightarrow \]
\[ 2 \cdot \det \begin{bmatrix}
1 & 0 & 0 \\
1 & 3 & -1 \\
1 & 1 & -2 \\
\end{bmatrix} \]
\[ -1 \text{ out of } C_3 \]
\[ \Rightarrow \]
\[ -2 \cdot \det \begin{bmatrix}
1 & 0 & 0 \\
1 & 3 & 1 \\
1 & 1 & 2 \\
\end{bmatrix} \]
\[ \text{expand} \]
\[ \equiv \]
\[ -2 \cdot 1 \cdot (-1)^{1+1} \cdot \det \begin{bmatrix}
3 & 1 \\
1 & 2 \\
\end{bmatrix} \]
\[ R_1 \leftarrow 3R_2 \]
\[ \Rightarrow \]
\[ -2 \cdot \det \begin{bmatrix}
0 & -5 \\
1 & 2 \\
\end{bmatrix} \]
\[ R_1 \leftrightarrow R_2 \]
\[ \Rightarrow \]
\[ -2 \cdot (-1) \cdot \det \begin{bmatrix}
1 & 2 \\
0 & -5 \\
\end{bmatrix} \]
\[ = \]
\[ 2 \cdot (-5) \]
\[ = \]
\[ -10. \]

**Example.** Compute the determinant
\[
\Delta = \det \begin{vmatrix}
1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 & 1 \\
1 & 1 & 0 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 \\
\end{vmatrix}
\]

**Solution:** Subtracting Row 1 from Rows 2, 3, and 4, we rearrange the de-
The determinant as
\[
\Delta = \det \begin{bmatrix}
1 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & -1 & 1 \\
0 & 0 & -1 & 0 & 1 \\
0 & -1 & 0 & 0 & 1 \\
0 & 1 & 1 & 1 & 1
\end{bmatrix}
\]

After expanding the determinant across the first column, we get
\[
\Delta = 1 \cdot (-1)^{1+1} \cdot \det \begin{bmatrix}
0 & 0 & -1 & 1 \\
0 & -1 & 0 & 1 \\
-1 & 0 & 0 & 1 \\
1 & 1 & 1 & 1
\end{bmatrix}
\]
\[
= \det \begin{bmatrix}
0 & 0 & -1 & 1 \\
0 & -1 & 0 & 1 \\
-1 & 0 & 0 & 1 \\
1 & 1 & 1 & 1
\end{bmatrix}
\]
\[
R_3 \pm R_2 \quad - \det \begin{bmatrix}
0 & 0 & -1 & 1 \\
0 & -1 & 0 & 1 \\
-1 & 0 & 0 & 1 \\
0 & 1 & 1 & 2
\end{bmatrix}
\]

After expansion across the 1st column, we have
\[
\Delta = -(1) \cdot (-1)^{3+1} \cdot \det \begin{bmatrix}
0 & -1 & 1 \\
-1 & 0 & 1 \\
1 & 1 & 2
\end{bmatrix}
\]
\[
= - \det \begin{bmatrix}
0 & -1 & 1 \\
-1 & 0 & 1 \\
1 & 1 & 2
\end{bmatrix}
\]
\[
R_3 \pm R_2 \quad - \det \begin{bmatrix}
0 & -1 & 1 \\
-1 & 0 & 1 \\
0 & 1 & 3
\end{bmatrix}
\]

As an exercise, I leave you with this question: wouldn’t you agree that the determinant of the similarly constructed matrix of size \(n \times n\) should be \(\pm (n-1)\)? But how can one determine the correct sign?

And one more exercise: you should now be able to instantly compute
\[
\det \begin{bmatrix}
1 & 2 & 3 & 4 & 0 \\
1 & 2 & 3 & 0 & 5 \\
1 & 0 & 3 & 4 & 5 \\
0 & 2 & 3 & 4 & 5
\end{bmatrix}
\]

Theorem 3.2.4: Invertible matrices. A square matrix \(A\) is invertible if and only if
\[
\det A \neq 0.
\]

Recall that this theorem has been already known to us in the case of \(2 \times 2\) matrices \(A\).

Example. The matrix
\[
A = \begin{bmatrix}
2 & 0 & 7 \\
0 & 1 & 0 \\
1 & 0 & 4
\end{bmatrix}
\]
has the determinant
\[
\det A = 2 \cdot \det \begin{bmatrix}
1 & 0 \\
0 & 4
\end{bmatrix} + 7 \cdot \det \begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix}
\]
\[
= 2 \cdot 4 + 7 \cdot (-1)
\]
\[
= 8 - 7 = 1
\]
hence $A$ is invertible. Indeed,

$$A^{-1} = \begin{bmatrix} 4 & 0 & -7 \\ 0 & 1 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

—check!

**Theorem 3.2.6: Multiplicativity Property.** If $A$ and $B$ are $n \times n$ matrices then

$$\det AB = (\det A) \cdot (\det B)$$

**Example.** Take

$$A = \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix},$$

then

$$\det A = 2 \quad \text{and} \quad \det B = 1.$$  

But

$$AB = \begin{bmatrix} 1 & 0 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 5 \\ 3 & 17 \end{bmatrix}$$

**Cramer’s Rule [Lay 3.3]**

Cramer’s Rule is an explicit (closed) formula for solving systems of $n$ linear equations with $n$ unknowns and nonsingular (invertible) matrix of coefficients. It has important theoretical value, but is unsuitable for practical application.

For any $n \times n$ matrix $A$ and any $b \in \mathbb{R}^n$, denote by $A_i(b)$ the matrix obtained from $A$ by replacing column $i$ by the vector $b$:

$$A_i(b) = [a_1 \cdots a_{i-1} \ b \ a_{i+1} \cdots a_n].$$

**Theorem 3.3.7: Cramer’s Rule.** Let $A$ be an invertible $n \times n$ matrix. For any $b \in \mathbb{R}^n$, the unique solution $x$ of the linear system

$$Ax = b$$

is given by

$$x_i = \frac{\det A_i(b)}{\det A}, \quad i = 1, 2, \ldots, n.$$  

For example, for a system

$$x_1 + x_2 = 3$$
$$x_1 - x_2 = 1$$

Cramer’s rule gives the answer

$$x_1 = \frac{\det \begin{bmatrix} 3 & 1 \\ 1 & -1 \end{bmatrix}}{-2} = 2$$
$$x_2 = \frac{\det \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}}{-2} = 1$$
Theorem 3.3.8: An Inverse Formula. Let $A$ be an invertible $n \times n$ matrix. Then

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

where cofactors $C_{ji}$ are given by

$$C_{ji} = (-1)^{j+i} \det A_{ji}$$

and $A_{ji}$ is the matrix obtained from $A$ by deleting row $j$ and column $i$.

Notice that for a $2 \times 2$ matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

the cofactors are

$$C_{11} = (-1)^{1+1} \cdot d = d$$
$$C_{21} = (-1)^{2+1} \cdot b = -b$$
$$C_{12} = (-1)^{1+2} \cdot c = -c$$
$$C_{22} = (-1)^{2+2} \cdot a = a,$$

and we get the familiar formula

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

$$= \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$
Lecture 17: Eigenvalues and eigenvectors

Quiz

What is the value of this determinant?

\[
\Delta = \det \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 3 \\ 1 & 2 & 6 \end{bmatrix}
\]

Quiz

Is \( \Gamma \) positive, negative, or zero?

\[
\Gamma = \det \begin{bmatrix} 0 & 0 & 71 \\ 0 & 93 & 0 \\ 87 & 0 & 0 \end{bmatrix}
\]

Eigenvectors. An eigenvector of an \( n \times n \) matrix \( A \) is a nonzero vector \( x \) such that

\( Ax = \lambda x \)

for some scalar \( \lambda \).

Eigenvalues. A scalar \( \lambda \) is called an eigenvalue of \( A \) if there is a non-trivial solution \( x \) of

\( Ax = \lambda x \);

\( x \) is called an eigenvector corresponding to \( \lambda \).

Eigenvalue. A scalar \( \lambda \) is an eigenvalue of \( A \) iff the equation

\( (A - \lambda I)x = 0 \)

has a non-trivial solution.

Eigenspace. The set of all solutions of

\( (A - \lambda I)x = 0 \)

is a subspace of \( \mathbb{R}^n \) called the eigenspace of \( A \) corresponding to the eigenvalue \( \lambda \).

Theorem: Linear independence of eigenvectors. If \( v_1, \ldots, v_p \) are eigenvectors that correspond to distinct eigenvalues of \( \lambda_1, \ldots, \lambda_p \) of an \( n \times n \) matrix \( A \), then the set

\( \{v_1, \ldots, v_p\} \)

is linearly independent.
Lecture 18: The characteristic equation

Characteristic polynomial

The polynomial
\[ \det(A - \lambda I) \]

in variable \( \lambda \) is characteristic polynomial of \( A \).

For example, in the \( 2 \times 2 \) case
\[
\begin{vmatrix}
a - \lambda & b \\
c & d - \lambda
\end{vmatrix} = (a + d)\lambda + (ad - bc).
\]

\[ \det(A - \lambda I) = 0 \]

is the characteristic equation for \( A \).

Characterisation of eigenvalues

**Theorem.** A scalar \( \lambda \) is an eigenvalue of an \( n \times n \) matrix \( A \) if and only if \( \lambda \) satisfies the characteristic equation
\[ \det(A - \lambda I) = 0 \]

**Zero as an eigenvalue.** \( \lambda = 0 \) is an eigenvalue of \( A \) if and only if \( \det A = 0 \).

The Invertible Matrix Theorem (continued). An \( n \times n \) matrix \( A \) is invertible iff:

- (s) The number 0 is not an eigenvalue of \( A \).
- (t) The determinant of \( A \) is not zero.

**Theorem: Eigenvalues of a triangular matrix.** The eigenvalues of a triangular matrix are the entries on its main diagonal.

**Proof.** This immediately follows from the fact the the determinant of a triangular matrices is the product of its diagonal elements. Therefore, for example, for the \( 3 \times 3 \) triangular matrix
\[
A = \begin{bmatrix}
d_1 & a & b \\
0 & d_2 & c \\
0 & 0 & d_3
\end{bmatrix}
\]

its characteristic polynomial \( \det(A - \lambda I) \) equals
\[
\begin{vmatrix}
d_1 - \lambda & a & b \\
0 & d_2 - \lambda & c \\
0 & 0 & d_3 - \lambda
\end{vmatrix}
\]

and therefore
\[
\det(A - \lambda I) = (d_1 - \lambda)(d_2 - \lambda)(d_3 - \lambda)
\]

has roots (eigenvalues of \( A \))
\[
\lambda = d_1, d_2, d_3.
\]
Lectures 19–22: Diagonalisation • Last change: 20 May 2019

Similar matrices. A is similar to B if there exist an invertible matrix $P$ such that

$$A = PBP^{-1}.$$  

Example:

Occasionally the similarity relation is denoted by symbol $\sim$.

Similarity is an equivalence relation: it is

- reflexive: $A \sim A$
- symmetric: $A \sim B$ implies $B \sim A$
- transitive: $A \sim B$ and $B \sim C$ implies $A \sim C$.

Conjugation. Operation

$$A^P = P^{-1}AP$$

is called conjugation of $A$ by $P$.

Properties of conjugation

- $I^P = I$
- $(AB)^P = A^P B^P$; as a corollary:
- $(A + B)^P = A^P + B^P$
- $(c \cdot A)^P = c \cdot A^P$
- $(A^k)^P = (A^P)^k$
- $(A^{-1})^P = (A^P)^{-1}$
- $(A^P)^Q = A^{PQ}$

Theorem: Similar matrices have the same characteristic polynomial.

If matrices $A$ and $B$ are similar then they have the same characteristic polynomials and eigenvalues.

Diagonalisable matrices. A square matrix $A$ is diagonalisable if it is similar to a diagonal matrix, that is,

$$A = PDP^{-1}$$

for some diagonal matrix $D$ and some invertible matrix $P$.

Of course, the diagonal entries of $D$ are eigenvalues of $A$.

The Diagonalisation Theorem. An $n \times n$ matrix $A$ is diagonalisable iff $A$ has $n$ linearly independent eigenvectors.

In fact,

$$A = PDP^{-1}$$

where $D$ is diagonal iff the columns of $P$ are $n$ linearly independent eigenvectors of $A$.

Theorem: Matrices with all eigenvalues distinct. An $n \times n$ matrix with $n$ distinct eigenvalues is diagonalisable.

Example: Non-diagonalisable matrices

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

has repeated eigenvalues $\lambda_1 = \lambda_2 = 2$.

$A$ is not diagonalisable. Why?
Lectures 23-24: Symmetric matrices and inner product

Symmetric matrices. A matrix $A$ is symmetric if $A^T = A$.

An example:

$$A = \begin{bmatrix}
1 & 2 & 3 \\
2 & 3 & 4 \\
3 & 4 & 5
\end{bmatrix}.$$ 

Notice that symmetric matrices are necessarily square.

Inner (or dot) product. For vectors $u, v \in \mathbb{R}^n$ their inner product (also called scalar product or dot product) $u \cdot v$ is defined as

$$u \cdot v = u^T v = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\
v_2 \\
\vdots \\
v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

Properties of dot product

Theorem 6.1.1. Let $u, v, w \in \mathbb{R}^n$, and $c$ be a scalar. Then

(a) $u \cdot v = v \cdot u$

(b) $(u + v) \cdot w = u \cdot w + v \cdot w$

(c) $(cu) \cdot v = c(u \cdot v)$

(d) $u \cdot u \geq 0$, and $u \cdot u = 0$ if and only if $u = 0$

Properties (b) and (c) can be combined as

$$(c_1u_1 + \cdots + c_pu_p) \cdot w = c_1(u_1 \cdot w) + \cdots + c_p(u_p \cdot w)$$

The length of a vector $v$ is defined as

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

In particular,

$$\|v\|^2 = v \cdot v.$$

We call $v$ a unit vector if $\|v\| = 1$.

Orthogonal vectors. Vectors $u$ and $v$ are orthogonal to each other if

$$u \cdot v = 0$$

Theorem. Vectors $u$ and $v$ are orthogonal if and only if

$$\|u\|^2 + \|v\|^2 = \|u + v\|^2.$$ 

That is, vectors $u$ and $v$ are orthogonal if and only if the Pythagoras Theorem holds for the triangle formed by $u$ and $v$ as two sides (so that its third side is $u + v$).

Proof is a straightforward computation:

$$\|u + v\|^2 = (u + v) \cdot (u + v)$$

$$= u \cdot u + u \cdot v + v \cdot u + v \cdot v$$

$$= u \cdot u + u \cdot v + u \cdot v + v \cdot v$$

$$= u \cdot u + 2u \cdot v + v \cdot v$$

$$= \|u\|^2 + 2u \cdot v + \|v\|^2.$$
Hence
\[ \|u + v\|^2 = \|u\|^2 + \|v\|^2 + 2u \cdot v, \]
and
\[ \|u + v\|^2 = \|u\|^2 + \|v\|^2 \text{ iff } u \cdot v = 0. \]

**Orthogonal sets.** A set of vectors \( \{ u_1, \ldots, u_p \} \) in \( \mathbb{R}^n \) is orthogonal if
\[ u_i \cdot u_j = 0 \text{ whenever } i \neq j. \]

**Theorem 6.2.4.** If
\[ S = \{ u_1, \ldots, u_p \} \]
is an orthogonal set of non-zero vectors in \( \mathbb{R}^n \) then \( S \) is linearly independent.

**Proof.** Assume the contrary that \( S \) is linearly dependent. This means that there exist scalars \( c_1, \ldots, c_p \), not all of them zeroes, such that
\[ c_1 u_1 + \cdots + c_p u_p = 0. \]
Forming the dot product of the left hand side and the right hand side of this equation with \( u_i \), we get
\[ (c_1 u_1 + \cdots + c_p u_p) \cdot u_i = 0 \cdot u_i = 0 \]
After opening the brackets we have
\[ c_1 u_1 \cdot u_i + \cdots + c_{i-1} u_{i-1} + c_i u_i \cdot u_i + c_{i+1} u_{i+1} \cdot u_i + \cdots + c_p u_p \cdot u_i = 0. \]
In view of orthogonality of the set \( S \), all dot products on the left hand side, with the exception of \( u_i \cdot u_i \), equal 0. Hence what remains from the equality is
\[ c_i u_i \cdot u_i = 0 \]
Since \( u_i \) is non-zero,
\[ u_i \cdot u_i \neq 0 \]
and therefore \( c_i = 0 \). This argument works for every index \( i = 1, \ldots, p \). We get a contradiction with our assumption that one of \( c_i \) is non-zero. \( \square \)

An **orthogonal basis** for \( \mathbb{R}^n \) is a basis which is also an orthogonal set.

For example, the two vectors
\[ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \]
form an orthogonal basis of \( \mathbb{R}^2 \) (check!).

**Theorem: Eigenvectors of symmetric matrices.** Let
\[ v_1, \ldots, v_p \]
be eigenvectors of a symmetric matrix \( A \) for pairwise distinct eigenvalues
\[ \lambda_1, \ldots, \lambda_p. \]
Then
\[ \{ v_1, \ldots, v_p \} \]
is an orthogonal set.

**Proof.** We need to prove the following:

If \( u \) and \( v \) are eigenvectors for \( A \) for eigenvalues \( \lambda \neq \mu \), then
\[ u \cdot v = 0. \]
For a proof, consider the following se-
sequence of equalities:
\[
\begin{align*}
\lambda \mathbf{u} \cdot \mathbf{v} &= (\lambda \mathbf{u}) \cdot \mathbf{v} \\
&= (\mathbf{Au}) \cdot \mathbf{v} \\
&= (\mathbf{Au})^T \mathbf{v} \\
&= (\mathbf{u}^T A^T) \mathbf{v} \\
&= (\mathbf{u}^T A) \mathbf{v} \\
&= \mathbf{u}^T (A \mathbf{v}) \\
&= \mathbf{u}^T (\mu \mathbf{v}) \\
&= \mu \mathbf{u}^T \mathbf{v} \\
&= \mu \mathbf{u} \cdot \mathbf{v}.
\end{align*}
\]

Hence
\[
\lambda \mathbf{u} \cdot \mathbf{v} = \mu \mathbf{u} \cdot \mathbf{v}
\]
and
\[(\lambda - \mu) \mathbf{u} \cdot \mathbf{v} = 0.
\]
Since \(\lambda - \mu \neq 0\), we have \(\mathbf{u} \cdot \mathbf{v} = 0\). \(\Box\)

**Corollary.** If \(A\) is an \(n \times n\) symmetric matrix with \(n\) distinct eigenvalues \(\lambda_1, \ldots, \lambda_n\), then the corresponding eigenvectors \(\mathbf{v}_1, \ldots, \mathbf{v}_n\) form an orthogonal basis of \(\mathbb{R}^n\).

An **orthonormal basis** in \(\mathbb{R}^n\) is an orthogonal basis \(\mathbf{u}_1, \ldots, \mathbf{u}_n\) made of unit vectors,
\[\|\mathbf{u}_i\| = 1 \text{ for all } i = 1, 2, \ldots, n.\]

Notice that this definition can reformulated as
\[
\mathbf{u}_i \cdot \mathbf{u}_j = \begin{cases} 
1, & \text{if } i = j \\
0, & \text{if } i \neq j
\end{cases}.
\]

**Theorem:** Coordinates with respect to an orthonormal basis. If \(\{ \mathbf{u}_1, \ldots, \mathbf{u}_n \}\) is an orthonormal basis and \(\mathbf{v}\) a vector in \(\mathbb{R}^n\), then
\[
\mathbf{v} = x_1 \mathbf{u}_1 + \cdots + x_n \mathbf{u}_n
\]
where
\[x_i = \mathbf{v} \cdot \mathbf{u}_i, \text{ for all } i = 1, 2, \ldots, n.\]

**Proof.** Let
\[
\mathbf{v} = x_1 \mathbf{u}_1 + \cdots + x_n \mathbf{u}_n,
\]
then, for \(i = 1, 2, \ldots, n,\)
\[
\mathbf{v} \cdot \mathbf{u}_i = x_1 \mathbf{u}_1 \cdot \mathbf{u}_i + \cdots + x_n \mathbf{u}_n \cdot \mathbf{u}_i \\
= x_i \mathbf{u}_i \cdot \mathbf{u}_i \\
= x_i.
\]

**Theorem:** Eigenvectors of symmetric matrices. Let \(A\) be a symmetric \(n \times n\) matrix with \(n\) distinct eigenvalues \(\lambda_1, \ldots, \lambda_n\).

Then \(\mathbb{R}^n\) has an orthonormal basis made of eigenvectors of \(A\).
Lecture 25: Orthogonal matrices

A square matrix $A$ is said to be orthogonal if

$$AA^T = I.$$  

Some basic properties of orthogonal matrices:

If $A$ is orthogonal then

- $A^T$ is also orthogonal.
- $A$ is invertible and $A^{-1} = A^T$.
- $\det A = \pm 1$.
- The identity matrix is orthogonal.

An example of an orthogonal matrix:

$$A = \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{bmatrix}$$

**Theorem: Characterisation of Orthogonal Matrices.** For a square $n \times n$ matrix $A$, the following conditions are equivalent:

(a) $A$ is orthogonal.

(b) Columns of $A$ form an orthonormal basis of $\mathbb{R}^n$. 
Lecture 26: Vector spaces

A vector space is a non-empty space of elements (vectors), with operations of addition and multiplication by a scalar.

By definition, the following holds for all \( u, v, w \) in a vector space \( V \) and for all scalars \( c \) and \( d \).

1. The sum of \( u \) and \( v \) is in \( V \).
2. \( u + v = v + u \).
3. \((u + v) + w = u + (v + w)\).
4. There is zero vector \( 0 \) such that \( u + 0 = u \).
5. For each \( u \) there exists \( -u \) such that \( u + (-u) = 0 \).
6. The scalar multiple \( cu \) is in \( V \).
7. \( c(u + v) = cu + cv \).
8. \((c + d)u = cu + du \).
9. \( c(du) = (cd)u \).
10. \( 1u = u \).

Definitions

Most concepts introduced in the previous lectures for the vector spaces \( \mathbb{R}^n \) can be transferred to arbitrary vector spaces. Let \( V \) be a vector space.

Given vectors \( v_1, v_2, \ldots, v_p \) in \( V \) and scalars \( c_1, c_2, \ldots, c_p \), the vector

\[
y = c_1 v_1 + \cdots + c_p v_p
\]

is called a linear combination of \( v_1, v_2, \ldots, v_p \) with weights \( c_1, c_2, \ldots, c_p \).

If \( v_1, \ldots, v_p \) are in \( V \), then the set of all linear combination of \( v_1, \ldots, v_p \) is denoted by \( \text{Span}\{v_1, \ldots, v_p\} \) and is called the subset of \( V \) spanned (or generated) by \( v_1, \ldots, v_p \).

That is, \( \text{Span}\{v_1, \ldots, v_p\} \) is the collection of all vectors which can be written in the form

\[
c_1 v_1 + c_2 v_2 + \cdots + c_p v_p
\]

with \( c_1, \ldots, c_p \) scalars.

\( \text{Span}\{v_1, \ldots, v_p\} \) is a vector subspace of \( V \), that is, it is closed with respect to addition of vectors and multiplying vectors by scalars.

\( V \) is a subspace of itself.

\( \{0\} \) is a subspace, called the zero subspace.

An indexed set of vectors

\[
\{ v_1, \ldots, v_p \}
\]

in \( V \) is linearly independent if the vector equation

\[
x_1 v_1 + \cdots + x_p v_p = 0
\]

has only trivial solution.

Examples. Of course, the most important example of vector spaces is provided by standard spaces \( \mathbb{R}^n \) of column vectors with \( n \) components, with the usual operations of component-wise addition and multiplication by scalar.

Another example is the set \( \mathbb{P}^n \) of polynomials of degree at most \( n \) in variable \( x \),

\[
\mathbb{P}^n = \{ a_0 + a_1 x + \cdots + a_n x^n \}
\]

with the usual operations of addition of polynomials multiplication by constant.

And here is another example: the set \( C[0, 1] \) of real-valued continuous functions on the segment \([0, 1]\), with the usual operations of addition of functions and multiplication by constant.
The set \[ \{ v_1, \ldots, v_p \} \]
in \( V \) is **linearly dependent** if there exist weights \( c_1, \ldots, c_p \), **not all zero**, such that
\[ c_1 v_1 + \cdots + c_p v_p = 0 \]

**Theorem: Characterisation of linearly dependent sets.** An indexed set
\[ S = \{ v_1, \ldots, v_p \} \]
of two or more vectors is linearly dependent if and only if at least one of the vectors in \( S \) is a linear combination of the others.

A **basis** of \( V \) is a linearly independent set which spans \( V \). A vector space \( V \) is called **finite dimensional** if it has a finite basis. Any two bases in a finite dimensional vector space have the same number of vectors; this number is called the **dimension** of \( V \) and is denoted \( \dim V \).

If \( B = (b_1, \ldots, b_n) \) is a basis of \( V \), every vector \( x \in V \) can be written, and in a unique way, as a linear combination
\[ x = x_1 b_1 + \cdots + x_n b_n; \]
the scalars \( x_1, \ldots, x_n \) are called coordinates of \( x \) with respect to the basis \( B \).
Lectures 27: Linear transformations of abstract vector spaces

A transformation $T$ from a vector space $V$ to a vector space $W$ is a rule that assigns to each vector $x$ in $V$ a vector $T(x)$ in $W$.

The set $V$ is called the **domain** of $T$, the set $W$ is the **codomain** of $T$.

A transformation $T : V \rightarrow W$ is **linear** if:

- $T(u + v) = T(u) + T(v)$ for all vectors $u, v \in V$;
- $T(cu) = cT(u)$ for all vectors $u$ and all scalars $c$.

**Properties of linear transformations.** If $T$ is a linear transformation then $T(0) = 0$ and $T(cu + dv) = cT(u) + dT(v)$.

**The identity transformation** of $V$ is

$$V \rightarrow V$$

$$x \mapsto x.$$

**The zero transformation** of $V$ is

$$V \rightarrow V$$

$$x \mapsto 0.$$

If $T : V \rightarrow W$ is a linear transformation, its **kernel**

$$\text{Ker} T = \{x \in V : T(x) = 0\}$$

is a subspace of $V$, while its **image**

$$\text{Im} T = \{y \in W : T(x) = y \text{ for some } x \in V\}$$

is a subspace of $W$. 
Lectures 28 and 29: Inner product spaces

**Inner (or dot) product** is a function which associate, with each pair of vectors \( \mathbf{u} \) and \( \mathbf{v} \) in a vector space \( V \) a real number denoted

\[ \mathbf{u} \cdot \mathbf{v}, \]

subject to the following axioms:

1. \( \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \)
2. \( (\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} \)
3. \( (c\mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) \)
4. \( \mathbf{u} \cdot \mathbf{u} \geq 0 \) and \( \mathbf{u} \cdot \mathbf{u} = 0 \) if and only if \( \mathbf{u} = \mathbf{0} \)

**Inner product space.** A vector space with an inner product is called an **inner product space.**

**Example:** School Geometry. The ordinary Euclidean plane of school geometry is an inner product space:

- Vectors: directed segments starting at the origin \( O \)
- Addition: parallelogram rule
- Product-by-scalar: stretching the vector by factor of \( c \).
- Dot product:
  \[ \mathbf{u} \cdot \mathbf{v} = ||\mathbf{u}|| ||\mathbf{v}|| \cos \theta, \]
  where \( \theta \) is the angle between vectors \( \mathbf{u} \) and \( \mathbf{v} \).

**Example:** \( C[0,1] \). The vector space \( C[0,1] \) of real-valued continuous functions on the segment \([0,1]\) becomes an inner product space if we define inner product by the formula

\[ f \cdot g = \int_0^1 f(x)g(x)dx. \]

In this example, the tricky bit is to show that the dot product on \( C[0,1] \) satisfies the axiom:

\[ \mathbf{u} \cdot \mathbf{u} \geq 0 \] and \( \mathbf{u} \cdot \mathbf{u} = 0 \) if and only if \( \mathbf{u} = \mathbf{0} \),

that is, if \( f \) is a continuous function on \([0,1]\) and

\[ \int_0^1 f(x)^2 dx = 0 \]

then \( f(x) = 0 \) for all \( x \in [0,1] \). This requires the use of properties of continuous functions; since we study linear algebra, not analysis, I am leaving it to the readers as an exercise.

**Length.** The length of the vector \( \mathbf{u} \) is defined as

\[ ||\mathbf{u}|| = \sqrt{\mathbf{u} \cdot \mathbf{u}}. \]

Notice that

\[ ||\mathbf{u}||^2 = \mathbf{u} \cdot \mathbf{u}, \]

and

\[ ||\mathbf{u}|| = 0 \iff \mathbf{u} = \mathbf{0}. \]

**Orthogonality.** We say that vectors \( \mathbf{u} \) and \( \mathbf{v} \) are **orthogonal** if

\[ \mathbf{u} \cdot \mathbf{v} = 0. \]

**The Pythagoras Theorem.** Vectors \( \mathbf{u} \) and \( \mathbf{v} \) are orthogonal if and only if

\[ ||\mathbf{u}||^2 + ||\mathbf{v}||^2 = ||\mathbf{u} + \mathbf{v}||^2. \]

**Proof** is a straightforward computation, exactly the same as in Lectures 25–27:

\[
\begin{align*}
||\mathbf{u} + \mathbf{v}||^2 &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\
&= \mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} \\
&= \mathbf{u} \cdot \mathbf{u} + 2\mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} \\
&= ||\mathbf{u}||^2 + 2\mathbf{u} \cdot \mathbf{v} + ||\mathbf{v}||^2.
\end{align*}
\]
Hence
\[ \|u + v\|^2 = \|u\|^2 + \|v\|^2 + 2u \cdot v, \]
and
\[ \|u + v\|^2 = \|u\|^2 + \|v\|^2 \]
if and only if
\[ u \cdot v = 0. \]

**Theorem:** The Cauchy-Schwarz Inequality. In any inner product space,
\[ |u \cdot v| \leq \|u\|\|v\|. \]

**The Cauchy-Schwarz Inequality: an example.** In the vector space \( \mathbb{R}^n \) with the dot product of
\[ u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \]
defined as
\[ u \cdot v = u^T v = u_1 v_1 + \cdots + u_n v_n, \]
the Cauchy-Schwarz Inequality becomes
\[ |u_1 v_1 + \cdots + u_n v_n| \leq \sqrt{u_1^2 + \cdots + u_n^2} \cdot \sqrt{v_1^2 + \cdots + v_n^2}. \]
or, which is its equivalent form,
\[ (u_1 v_1 + \cdots + u_n v_n)^2 \leq (u_1^2 + \cdots + u_n^2) \cdot (v_1^2 + \cdots + v_n^2). \]

**The Cauchy-Schwarz Inequality: one more example.** In the inner product space \( C[0, 1] \) the Cauchy-Schwarz Inequality takes the form
\[ \left| \int_0^1 f(x)g(x) \, dx \right| \leq \sqrt{\int_0^1 f(x)^2 \, dx} \cdot \sqrt{\int_0^1 g(x)^2 \, dx}, \]
or, equivalently,
\[ \left( \int_0^1 f(x)g(x) \, dx \right)^2 \leq \left( \int_0^1 f(x)^2 \, dx \right) \cdot \left( \int_0^1 g(x)^2 \, dx \right). \]

**Proof of the Cauchy-Schwarz Inequality** given here is different from the one in the textbook by Lay. Our proof is based on the following simple fact from school algebra:

Let
\[ q(t) = at^2 + bt + c \]
be a quadratic function in variable \( t \) with the property
\[ q(t) = 0 \]
for some \( t \). Then
\[ q(t) q(t) = a t^2 + b t + c \cdot a t^2 + b t + c \]
and
\[ q(t) q(t) \leq (a t^2 + b t + c)^2 = a^2 t^4 + 2ab t^3 + (b^2 + 4ac) t^2 + 2bc t + c^2. \]

Since \( f(x) = q(t) \) and \( g(x) = q(t) \), the Cauchy-Schwarz Inequality follows.

\[ \int_0^1 f(x)g(x) \, dx \leq \sqrt{\int_0^1 f(x)^2 \, dx} \cdot \sqrt{\int_0^1 g(x)^2 \, dx}, \]
and
\[ \left( \int_0^1 f(x)g(x) \, dx \right)^2 \leq \left( \int_0^1 f(x)^2 \, dx \right) \cdot \left( \int_0^1 g(x)^2 \, dx \right). \]
that
\[ q(t) \geq 0 \]
for all \( t \). Then
\[ b^2 - 4ac \leq 0. \]

Now consider the function \( q(t) \) in variable \( t \) defined as
\[
q(t) = (u + tv) \cdot (u + tv) \\
= u \cdot u + 2tu \cdot v + t^2v \cdot v \\
= \|u\|^2 + (2u \cdot v)t + (\|v\|^2)t^2.
\]
Notice \( q(t) \) is a quadratic function in \( t \) and
\[ q(t) \geq 0. \]
Hence
\[ (2u \cdot v)^2 - 4\|u\|^2\|v\|^2 \leq 0 \]
which can be simplified as
\[ |u \cdot v| \leq \|u\|\|v\|. \]

\[ \square \]

**Distance.** We define distance between vectors \( u \) and \( v \) as
\[ d(u, v) = \|u - v\|. \]
It satisfies **axioms of metric**:

1. \( d(u, v) = d(v, u) \).
2. \( d(u, v) \geq 0 \).
3. \( d(u, v) = 0 \iff u = v \).
4. The **Triangle Inequality**:
\[ d(u, v) + d(v, w) \geq d(u, w). \]

Axioms 1–3 immediately follow from the axioms for dot product, but the Triangle Inequality requires some attention.

**Proof of the Triangle Inequality.** It suffices to prove
\[ \|u + v\| \leq \|u\| + \|v\|, \]
or
\[ \|u + v\|^2 \leq (\|u\| + \|v\|)^2, \]
or
\[ (u + v) \cdot (u + v) \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2, \]
or
\[ u \cdot u + 2u \cdot v \cdot v \leq u \cdot u + 2\|u\|\|v\| + v \cdot v, \]
or
\[ 2u \cdot v \leq 2\|u\|\|v\| \]
But this is the Cauchy-Schwarz Inequality. Now observe that all rearrangements are reversible, hence the Triangle Inequality holds. \[ \square \]
Lecture 27: Orthogonalisation and the Gram-Schmidt Process

**Orthogonal basis.** A basis $v_1, \ldots, v_n$ in the inner product space $V$ is called **orthogonal** if $v_i \cdot v_j = 0$ for $i \neq j$.

Coordinates in respect to an orthogonal basis:

$$u = c_1 v_1 + \cdots + c_n v_n$$

iff

$$c_i = \frac{u \cdot v_i}{v_i \cdot v_i}$$

**Orthonormal basis.** A basis $v_1, \ldots, v_n$ in the inner product space $V$ is called **orthonormal** if it is orthogonal and $\|v_i\| = 1$ for all $i$.

Equivalently,

$$v_i \cdot v_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Coordinates in respect to an orthonormal basis:

$$u = c_1 v_1 + \cdots + c_n v_n$$

iff

$$c_i = u \cdot v_i$$

The **Gram-Schmidt Orthogonalisation Process** makes an orthogonal basis from a basis $x_1, \ldots, x_p$:

$$v_1 = x_1$$

$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$$

$$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2$$

$$\vdots$$

$$v_p = x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_p \cdot v_2}{v_2 \cdot v_2} v_2 - \cdots - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}$$

**Main Theorem about inner product spaces.** Every finite dimensional inner product space has an orthonormal basis and is isomorphic to one of $\mathbb{R}^n$ with the standard dot product

$$u \cdot v = u^T v = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$
Lecture 30: Revision, Systems of Linear Equations

Systems of linear equations

Here $A$ is an $m \times n$ matrix.

We associate with $A$ systems of simultaneous linear equations

$$Ax = b,$$

where $b \in \mathbb{R}^m$ and

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

is the column vector of unknowns.

Especially important is the homogeneous system of simultaneous linear equations

$$Ax = 0.$$ 

The set

$$\text{null } A = \{ x \in \mathbb{R}^n : Ax = 0 \}$$

is called the null space of $A$; it is a vector subspace of $\mathbb{R}^n$ and coincides with the solution space of the homogeneous system $Ax = 0$.

If $a_1, \ldots, a_n$ are columns of $A$, so that

$$A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix},$$

then the column space of $A$ is defined as

$$\text{Col } A = \{ c_1 a_1 + \cdots + c_n a_n : c_i \in \mathbb{R} \}$$

and equals to the span of the system of vectors $a_1, \ldots, a_n$, that is, the set of all possible linear combinations of vectors $a_1, \ldots, a_n$.

The column space of $A$ has another important characterisation:

$$\text{Col } A = \{ b \in \mathbb{R}^m : Ax = b \text{ has a solution} \}.$$ 

Assume now that $A$ is a square $n \times n$ matrix.

If we associate with $A$ the linear transformation

$$T_A : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

$$v \mapsto Av$$

then the column space of $A$ gets yet another interpretation: it is the image of the linear transformation $T_A$:

$$\text{Col } A = \text{Im } T_A$$

$$= \{ v \in \mathbb{R}^n : \exists v \in \mathbb{R}^n \text{ s.t. } T_A(w) = v \},$$

and the null space of $A$ is the kernel of $T_A$:

$$\text{null } A = \ker T_A$$

$$= \{ v \in \mathbb{R} : T_A(v) = 0 \}.$$
Linear dependence

If \( a_1, \ldots, a_p \) are vectors in a vector space \( V \), an expression
\[
  c_1 a_1 + \cdots + c_p a_p, \quad c_i \in \mathbb{R}
\]
is called a linear combination of the vectors \( a_1, \ldots, a_p \), and coefficients \( c_i \) are called weights in this linear combination. If
\[
  c_1 a_1 + \cdots + c_p a_p = 0
\]
we say that weights \( c_1, \ldots, c_p \) define a linear dependency of vectors \( a_1, \ldots, a_p \). We always have a trivial (or zero) dependency, in which all \( c_i = 0 \). We say that vectors \( a_1, \ldots, a_p \) are linearly dependent if the admit a non-zero (or non-trivial) linear dependency, that is, a dependency in which at least one weight \( c_i \neq 0 \).

Let us arrange the weights \( c_1, \ldots, c_p \) in a column
\[
  c = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}
\]
and form a matrix \( A \) using vectors \( a_1, \ldots, a_p \) as columns, then
\[
  c_1 a_1 + \cdots + c_p a_p = \begin{bmatrix} a_1 & \cdots & a_p \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} = Ac
\]
and linear dependencies between vectors \( a_1, \ldots, a_p \) are nothing else but solutions of the homogeneous system of linear equation
\[
  Ax = 0.
\]

Therefore vectors \( a_1, \ldots, a_p \) are linearly dependent if and only if the system
\[
  Ax = 0
\]
has a non-zero solution.

Elementary row transformation

Elementary row transformations
\[
  R_i \leftrightarrow R_j, \ i \neq j
\]
\[
  R_i \leftarrow R_i + \lambda R_j, \ i \neq j
\]
\[
  R_i \leftarrow \lambda R_i, \ \lambda \neq 0
\]
performed on a matrix \( A \) amount to multiplication of \( A \) on the left by corresponding elementary matrices,
\[
  A \leftarrow EA,
\]
and therefore do not change dependencies between columns:
\[
  Ac = 0 \iff (EA)c = 0.
\]

Therefore if certain columns of \( A \) were linearly dependent (respectively, independent) then the same columns of \( EA \) remain linearly dependent (respectively, independent).

Rank of a matrix

We took it without proof that if \( U \) is a vector subspace of \( \mathbb{R}^n \), then the maximal systems of linearly independent subsets in \( U \) contain the same number of vectors. This number is called the dimension of \( U \) and is denoted \( \text{dim} U \).

A maximal system of linearly independent vectors in \( U \) is called a basis of \( U \).

Each basis of \( U \) contains \( \text{dim} U \) vectors.
An important theorem:
\[ \dim \text{null } A + \dim \text{Col } A = n \]

\[ \dim \text{Col } A \text{ is called the rank of } A \text{ and is denoted } \text{rank } A = \dim \text{Col } A. \]

Therefore
\[ \dim \text{null } A + \text{rank } A = n \]

**Theorem.** If \( \text{rank } A = p \) then among columns of \( A \) there are \( p \) linearly independent columns

**Theorem.** Since elementary row transformations of \( A \) do not change dependency of columns, elementary row operation do not change rank \( A \).