Chapter 1

Friday, January 2, 2026 8:02 am

Chapter 1. Linear Equations

§ 1.1 Fields

Story: It is often helpful to "abstractly" describe the key properties of real/complex numbers.

One such useful abstraction is called a "field" which captures "all" the salient features of such "scalars"

Def^n: Let F be a set & suppose two operations "addition" & "multiplication" are given by $+ : F \times F \to F$ & $\cdot : F \times F \to F$.

We say F is a ==field==, if the following conditions hold (together with the operations $+$ & $\cdot$).

1. Addition is commutative,
   $$x + y = y + x \quad \forall\, x, y \in F.$$

2. Addition is associative,
   $$x + (y + z) = (x + z) + z \quad \forall\, x, y, z \in F.$$

3. There is a unique element 0 (zero) in F s.t. $x + 0 = x \quad \forall\, x \in F$

4. For all $x \in F$, $\exists$ a unique element $(-x) \in F$ st $x + (-x) = 0$

5. Multiplication is commutative, i.e. $xy = yx \;\forall\, x, y \in F$

6. Multiplication is associative, i.e.
   $$x(yz) = (xy)z \quad \forall\, x y z \in F.$$

7. $\exists$ a unique non-zero element 1 (one) in F s.t.
   $$x1 = x \quad \forall\, x \in F.$$

8. To each non-zero x in F,
   $\exists$ a corresponding unique element x^{-1} (or $1/x$) in F s.t. $xx^{-1} = 1$.

9. Multiplication distributes over addition
i.e. $x(y+z) = xy + xz \quad \forall\ x, y, z \in F$.

Story: • Intuitively then, a field is a set together with some operations on this set that behave like ordinary addition & multiplication
(subtraction) (division).

• We call these numbers "scalars".

NB. The set of real numbers, with usual addition & multiplication constitute a field.
Similarly for complex numbers.

Story: We now define a "subfield" which as the name suggests, is a "smaller field" contained in a given "field".

Defn: Let $(F, +, \cdot)$ be a field.
Then $(F', +, \cdot)$ is a subfield of $(F, +, \cdot)$
if $F' \subseteq F$ & $(F', +, \cdot)$ is also a field.

NB. The set of real numbers is a subfield of the set of complex numbers.

(Notation: Sometimes we refer to F as the field instead of $(F, +, \cdot)$ which is the formally correct defn.)

Story: The usefulness of a subfield arises from the following:
if one is working with a subfield of say $\mathbb{C}$ (complex numbers)
then performing additions (subtractions), multiplications (additions)
does not take one out of the subfield.

e.g. 1 The set of positive integers:
$1, 2, 3, \ldots$ is not a subfield of $\mathbb{C}$.
(This is for a variety of reasons — 0 not in the set
missing additive &
multiplicative inverses.

e.g. 2 The set of integers: $\ldots, -2, -1, 0, 1, 2, \ldots$
is not a subfield of $\mathbb{C}$
(This is because multiplicative inverses are missing)
(but it satisfies the other requirements)

e.g. 3 The set of rational numbers
(i.e. numbers of the form p/q for p & q integers
w/ $q \neq 0$)
is a subfield of $\mathbb{C}$.

NB: Every subfield of $\mathbb{C}$ must contain the set of rational numbers.

e.g. 4. The set of all complex numbers of the form

$x + y\sqrt{2}$ (where x & y are rational)

is a subfield of $\mathbb{C}$.

Assumption: Henceforth, every field considered is a subfield of the complex numbers.
(unless expressly stated otherwise)

Story: Why?

It so happens that we can add 1 n-many times to obtain a zero (Will see this in Exercise 5, following §1.2) for a finite n.

called Fields of characteristic n.

If this does not happen (as is the case for $\mathbb{C}$ & its subfields)

then F is said to have (somewhat confusingly) characteristic 0.

We won't dwell much on this here, in this course.

§ 1.2 Systems of Linear Equations

Def^n: Suppose F is a field.
A system of m linear equations in n unknowns is any problem of the following form:
find n scalars (elements of F) $x_1 \dots x_n$

s.t.

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \dots + A_{1n}x_n &= y_1 \\ \vdots \qquad\quad \vdots \qquad\qquad\quad & \ \vdots \\ A_{m1}x_1 + A_{m2}x_2 + \dots + A_{mn}x_n &= y_m \end{aligned} \tag{1-1}$$

where $y_1 \dots y_m$ &
$\{A_{ij}\}_{\substack{i \in \{1,\dots,m\} \\ j \in \{1,\dots,n\}}}$ are given elements of F.

Any n-tuple $(x_1 \dots x_n)$ of elements in F satisfying each equation, is called a solution of the system.

If $y_1 = y_2 = \dots = y_m = 0$, the system is called homogeneous.

Story: Can one say, given two systems of equations, that their solutions will be the same?

We begin with an illustration & then try to generalise the idea to answer the question above.

Illustration: Suppose one wants to solve

$$2x_1 - x_2 + x_3 = 0$$
$$x_1 + 3x_2 + 4x_3 = 0$$

• Lets add (-2) times the second eq^n to the first to get
$$-7x_2 - 7x_3 = 0 \iff x_2 = -x_3.$$

• Lets add (3) times the first eq^n to the second to get
$$7x_1 + 7x_3 = 0 \iff x_1 = -x_3.$$

We conclude that any solution x_1, x_2, x_3 of the system must satisfy $x_1 = x_2 = -x_3$. (Conversely) One can verify that for all a, $(a, a, -a)$ is a solution.

Story (resumed): We followed the process of "eliminating unknowns".

We multiplied & added equations to remove unknowns from our equations.

Let us formalise this process slightly.

Def^n: Recall the previous def^n & (1-1) in particular. Suppose one selects m scalars $c_1 \dots c_m$ multiplies the j^{th} equation by c_j & then adds these to obtain

$$(c_1 A_{11} + \dots + c_m A_{m1}) x_1 + \dots + (c_1 A_{1n} + \dots + c_m A_{mn}) x_n$$
$$= c_1 y_1 + \dots + c_m y_m.$$

Such an equation is called a linear combination of equations in (1-1).

NB: Any solution to (1-1) is also a solution to the equation above.

Story: One can extend this idea to define a notion of equivalence of systems of linear equations.

Def^n: Recall (1-1) & consider another system of equations

$$B_{11}x_1 + \dots + B_{1n}x_n = z_1$$
$$\vdots \qquad\qquad \vdots$$
$$B_{k1}x_1 + \dots + B_{kn}x_n = z_k.$$

These two systems are equivalent, if

each equation in each systems is a linear combination of the equations in the other system.

Story: Clearly, if the two systems are equivalent & $x_1 \dots x_n$ is a solution to the first system it is also a solution to each equation in the second system.

Conversely, if $x_1 \dots x_n$ is a sol^n to the second it is also a sol^n to the first.

We therefore have the following.

> Theorem 1. Equivalent systems of linear equations have exactly the same solutions.

Story: In the next section, we will see how to produce an equivalent system that is easier to solve, starting from an arbitrary initial system.

<u>Rough</u>

$x + y\sqrt{2}$ addition | subtraction | multi^n

Divi ~~suppose~~ Fix x, y, find p, q

$(x + y\sqrt{2})(p + q\sqrt{2}) = 1$

$xp + 2yq + (py + xq)\sqrt{2} = 1$

$xp + 2yq = 1$

$(py + xq) = 0$

$xp + 2yq = 1$

$p = -\frac{xq}{y}$

$-x\left(\frac{xq}{y}\right) + 2yq = 1$

$\left(-\frac{x^2}{y} + 2y\right)q = 1$

$q = \frac{1}{2y - \frac{x^2}{y}}$

$p = \frac{-xq}{y}$

§1.3 Matrices and Elementary Row Operations

Story: We now make our notation more brief.

Notation: We abbreviate the system of equations (1-1) by

$AX = Y$

where $A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}$ $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$

A is called the matrix of coefficients of the system

Remarks:

(1) For now, we take this to simply be a shorthand

We will look at the matrix multiplication notation and definition later.

(2) A is strictly speaking, not a matrix

An $m \times n$ matrix over the field F

is a function A from (i, j) where $1 \leq i \leq m$ and $1 \leq j \leq n$

into the field F

Entries of the matrix A are the scalars $A(i, j) = A_{ij}$

Story:

We now want to consider operations

on the rows of the matrix A

that correspond to forming linear combinations

of the equations in the system

of the equations in the system
$AX = Y$

We first look at three elementary row operation.

Inf Definition. **Elementary row operations** on an $m \times n$ matrix A over the field F
are defined to be the following.
1. Multiplication of one row of A by a non-zero scalar c
2. Replacement of the rth row of A
by row r plus c times row s
where c is any scalar and $r \neq s$
3. Interchange of two rows of A

Story:
An elementary row operation is
thus a special type of function (or rule) e
that associates with each $m \times n$ matrix A
another $m \times n$ matrix $e(A)$.

Using this, one can define these operations more formally.

Definition. Elementary row operations.
Let A be an $m \times n$ matrix as above and
let e be an arbitrary function of the form above.

Then, e must satisfy one of the following
(for some row distinct row indeces r, s and scalar $c \neq 0$):
1. $e(A)_{ij} = A_{ij}$ for $i \in \{1 \ldots m\} \backslash r$ and $e(A)_{rj} = cA_{rj}$
2. $e(A)_{ij} = A_{ij}$ for $i \in \{1 \ldots m\} \backslash r$ and $e(A)_{rj} = A_{rj} + cA_{sj}$
3. $e(A)_{ij} = A_{ij}$ for $i \in \{1 \ldots m\} \backslash \{r, s\}$ and $e(A)_{rj} = A_{sj}$, $e(A)_{sj} = A_{rj}$

These are called types of "e"

Story:
In defining $e(A)$
it is not really important how many columns A has (i.e. n can be arbitrary)
but th enumber of rows of A (i.e. m) is crucial
(in the sense that interchanging rows 5 and 6
of a 5 x 6 matrix makes no sense).

Therefore when we speak of e, we are considering the class of all m-rowed matrices over F.

Why these specific operations?
One reason is that after performing e on a matrix A to obtain $e(A)$,
one can recover A by performing a similar operation on $e(A)$.

Theorem 2.
To each elementary row operation e
there corresponds an elementary row operation e_1
of the same type as e
such that $e_1(e(A)) = e(e_1(A)) = A$ for all A.

In other words, the inverse operation (function)
of an elementray row operation exists
and it is an elementary row operation

of the same type.

Proof.

We do the proof for each type separately

(1) Suppose e is the operation which
multiplies the rth row of a matrix
by the non-zero scalar c.

Clearly, we can take e_1 to be the operation that multiplies r by c^{-1}.

(2) Suppose e is the operation which replaces row r by row r plus c times row s
($s \neq r$).
Again, clearly, we can take e_1 to be the operation that
replaces row r by row r plus $(-c)$ times row s.

(3) If e interchanges rows r and s, we can take $e_1 = e$.

It is easy to check that we have $e_1\big(e(A)\big) = e\big(e_1(A)\big) = A$ for all A.

∎

Story: We now use these operations to define another notion of equivalence between matrices.

Definition. If A and B are $m \times n$ matrices over the field F
we say that B is row-equivalent to A
if B can be obtained from A
by a finite sequence of elementary row operations.

Story: Here are some simple observations that can be
verified using Theorem 2.

NB.
(1) Each matrix is row-equivalent to itself
(2) if B is row-equivalent to A
then A is row-equivalent to B
(3) if C is row-equivalent to B and B is row-equivalent to A
then C is row-equivalent to A.

In other words, row-equivalence is an equivalence relation.

<---- covered until here on Friday, Jan 2, 2026

Theorem 3. If A and B are row-equivalent $m \times n$ matrices
the homogenous systems of linear equations
$AX = 0$ and $BX = 0$ have exactly the same solutions.

Proof.

Suppose we pass from A to B by a
finite sequence of elementray row operations:

$$A = A_0 \to A_1 \to \cdots \to A_k = B$$

NB. It is enough to prove that the systems $A_iX = 0$ and $A_{i+1}X = 0$

N.B. it is enough to prove that the systems $A_j X = 0$ and $A_{j+1} X = 0$

have the same solutions,

i.e. one elementary row operation leaves the set of solutions unchanged.

This is easy to establish.

Suppose C is obtained from D by a single elementary operation (where C and D are also $m \times n$ matrices)

Now, no matter which of the three types of elementary row operation is used it is clear that each equation in system $CX = 0$ is a linear combination of th eequatinos in $DX = 0$

Furthermore, since the inverse of an elementary row operation is also an elementary row operation each equation in $DX = 0$ will also be a linear combination of the equations in $CX = 0$.

Thus, these two systems are equivalent—and by Theorem 1, have the same solutions.

■

Story:

Now let us look at how one might use these elementary row-operations to potentially solve a homogenous system of linear equations

EXAMPLE 5. Suppose F is the field of rational numbers, and

$$A = \begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix}.$$

We shall perform a finite sequence of elementary row operations on A, indicating by numbers in parentheses the type of operation performed.

$$\begin{bmatrix} 2 & -1 & 3 & 2 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 2 & 6 & -1 & 5 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & -2 & -1 & 7 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 4 & 0 & -1 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & -9 & 3 & 4 \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & \frac{15}{2} & -\frac{55}{2} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(1)}$$

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & -2 & 13 \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & \frac{1}{2} & -\frac{7}{2} \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & -\frac{5}{3} \end{bmatrix}$$

The row-equivalence of A with the final matrix in the above sequence tells us in particular that the solutions of

$$\begin{aligned} 2x_1 - x_2 + 3x_3 + 2x_4 &= 0 \\ x_1 + 4x_2 \qquad\quad - x_4 &= 0 \\ 2x_1 + 6x_2 - x_3 + 5x_4 &= 0 \end{aligned}$$

and

$$\begin{aligned} x_3 - \tfrac{11}{3}x_4 &= 0 \\ x_1 \qquad + \tfrac{17}{3}x_4 &= 0 \end{aligned}$$

$$x_2 \qquad - \tfrac{5}{3}x_4 = 0$$

are exactly the same. In the second system it is apparent that if we assign

any rational value c to x_4 we obtain a solution $(-\frac{17}{3}c, \frac{5}{3}, \frac{11}{3}c, c)$, and also that every solution is of this form.

Here's another example.

Example 6. Suppose F is the field of complex numbers and

$$A = \begin{bmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{bmatrix}.$$

In performing row operations it is often convenient to combine several operations of type (2). With this in mind

$$\begin{bmatrix} -1 & i \\ -i & 3 \\ 1 & 2 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 2+i \\ 0 & 3+2i \\ 1 & 2 \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 0 & 1 \\ 0 & 3+2i \\ 1 & 2 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Thus the system of equations

$$\begin{aligned} -x_1 + ix_2 &= 0 \\ -ix_1 + 3x_2 &= 0 \\ x_1 + 2x_2 &= 0 \end{aligned}$$

has only the trivial solution $x_1 = x_2 = 0$.

Story:

In these examples,

we were trying to simplify the coefficient matrix

in a manner analogous to "eliminating unknowns"

in the system of linear equations.

Let us now make a formal definition of

the type of matrix at which

we were attempting to arrive.

Definition. An $m{\times}n$ matrix R is called **row-reduced** if

(1) the first non-zero entry in

each non-zero row of R is 1

(2) each column of R

which contains the leading non-zero entry of some row

has all its other entries 0.

Story: Let us look at a very simple example of a row-reduced matrix.

Example 7. One example of a row-reduced matrix is the $n \times n$ (square) **identity matrix** I. This is the $n \times n$ matrix defined by

$$I_{ij} = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases}$$

This is the first of many occasions on which we shall use the **Kronecker delta** (δ).

Story:

Note that in Examples 5 and 6,
the final matrices in the sequence
were row-reduced matrices.

$$\begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & -\frac{5}{3} \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Here are some examples of matrices
that are *not* row-reduced.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix}.$$

The first matrix fails condition (b) in column 3
The second matrix fails condition (a) in row 1

Theorem 4.

Every $m \times n$ matrix over the field F is
row-equivalent to
a row-reduced matrix.

Proof.

Let A be an $m \times n$ matrix over F.

Handling row 1
If every entry in the first row of A is 0
then condition (a) is satisfied
as far as row 1 is concerned.
If row 1 has a non-zero entry
let k be the smallest index for which $A_{1k} \neq 0$

Multiply row 1 by A_{1k}^{-1}
and then condition (a) is satisfied with regard to row 1.

Now, for each $i \geq 2$
add $(-A_{ik})$ times row 1 to row i.

At this point, the leading non-zero entry of row 1
occurs in column k and that entry is 1.
Also, all other entries in column k are 0.

Handling row 2
Consider now the matrix which we obtained above.

If every entry in row 2 is 0, we leave it unchanged.
If some entry in row 2 is not zero
we multiply row 2 by a scalar so that the leading non-zero term
is again 1.

Suppose this leading entry occurs at row k'

It is clear that $k' \neq k$ (from how we handled row 1; recall the definiton of k above)

By adding suitable multiples of row 2 to the various rows
one can ensure that all entries in column k'
are 0 except for row 2 (where it takes value 1).

The crucial observation is that
in carrying out these operations,
the entries of row 1 will remain unchanged
and
the entries of column k remain unchanged.

Of course, if row 1 is identically 0, the operations with row 2
will not affect row 1.

Proceeding like this, one row at a time, it is clear that
in a finite number of steps, one will arrive at a row-reduced matrix.

■

§ 1.4 Row Reduced Echelon Matrices

Story:

- So far,
  we looked at systems of linear equations
  to solve them
  - In §1.3 we established a standard technique
    to obtain such solutions
- We now want to acquire some information
  that is slightly more theoretical
  &
  for this we will help to go beyond row-reduced matrices

Definition.

An $m \times n$ matrix R is called a **row-reduced echelon matrix** if

(a) R is row-reduced
(b) every row of R that has all its entries 0
occurs below every row which
has a non-zero entry
(c) if rows $1, \dots, r$ are the non-zero rows of R
and
if the leading non-zero entry of row i
occurs in column k_i $i \in \{1 \dots r\}$
then $k_1 < \cdots < k_r$

Story:

- We are basically permuting the rows to ensure it has a nice "diagonal-like" form

- Here's another way to define a row reduced echeleon matrix

Alt Definition.

An $m \times n$ matrix R is a **row-reduced echelon matrix**
if the following hold:
Either every entry in R is 0
or
there exists
a positive integer r,
$1 \leq r \leq m$, and
r positive integers $k_1 \ldots k_j$ (with each k_i s.t. $1 \leq k_i \leq n$)
satisfying the following:
(a) $R_{ij} = 0$ for $i > r$ and $R_{ij} = 0$ if $j < k_i$
(b) $R_{ik_i} = \delta_{ij}$ for $1 \leq i \leq r$ and $1 \leq j \leq r$
(c) $k_1 < \cdots < k_r$

$\delta_{ij} = 0 \; if \; i \neq j$
$\delta_{ij} = 1$ if $i = j$

Example 8.

Clearly,
the identity matrix and
the all-zero matrix are row-reduced echelon matrices
For a non-trivial example, consider the following

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

Theorem 5. Every $m \times n$ matrix A is row-equivalent to a row-reduced *echelon* matrix.

Proof.
A is row-equivalent to row-reduced matrix (Thm 4).
A row-reduced matrix is equivalent to a row-reduced echelon matrix
as it only needs its rows to be re-ordered
■

Story:

- In Examples 5 and 6
  we saw how row-reduced matrices
  help in solving homogeneous systems of linear equations

- Let us briefly discuss these when
  the matrices are in row-reduced echelon form,
  i.e.
  $RX = 0$
  where R is in row-reduced echelon

Illustration:

- Notation
  - Let rows $1 \ldots r$ be the non-zero rows of R
    and
    suppose that
    the leading non-zero entry of row i

occurs in column k_i

- Let $u_1 \dots u_{n-r}$ denote the $(n-r)$ unknowns that are different from $x_{k_1}, \dots x_{k_r}$

- The r non-trivial equations in $RX = 0$ are

$$x_{k_1} + \sum_{j=1}^{n-r} C_{1j} u_j = 0$$
$$\vdots \qquad\qquad \vdots \qquad\qquad (1\text{-}3)$$
$$x_{k_r} + \sum_{j=1}^{n-r} C_{rj} u_j$$

$RX = 0$

$$\begin{array}{c} \\ \begin{bmatrix} 0 & 0 & 1 & -\frac{11}{3} \\ 1 & 0 & 0 & \frac{17}{3} \\ 0 & 1 & 0 & -\frac{5}{3} \end{bmatrix} \end{array} \quad \begin{array}{c} x\_1 \\ x\_2 \\ x\_3 \\ u\_1 \end{array}$$

- All solutions to the system of equations $RX=0$ are obtained by assigning arbitrary values to $u_1 \dots u_{n-r}$ and then computing $x_1 \; x_2 \dots x_{k_r}$ using (1-3).

<--- stopped here on Tuesday, Jan 6

- Let us work out Example 8.

$$\text{recall: } \begin{array}{ccccc} & k_1 & & k_2 & \\ & \downarrow & & \downarrow & \end{array}$$
$$\begin{bmatrix} 0 & 1 & -3 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

parameters: $r = 2$ (# non-zero rows)

$k_1 = 2$, $k_2 = 4$ (column of first non-zero entry)

equations: $x_2 - 3\underbrace{x_3}_{u_1} + \frac{1}{2}\underbrace{x_5}_{u_2} = 0 \qquad x_2 = 3u_1 - \frac{1}{2}u_2$

$x_4 + 2\underbrace{x_5}_{u_2} = 0 \qquad x_4 = -2u_2$

$x_1 = u_3$

$\therefore \; (u_3,\; 3u_1 - \frac{1}{2}u_2,\; u_1,\; -2u_2,\; u_2)$

is a sol^n for any u_1, u_2, u_3

Remark

- Observe the following,

for a system of equations $RX = 0$.

- Suppose the number r
  of non-zero rows in R
  is (strictly) less than n.
  Then $RX = 0$
  has a non-trivial solution
  i.e. a solution where $(x_1 \dots x_n)$ is not all zeros.

- Why is that?
  As in the example above
  since $r < n$
  one can choose some x_j
  which is not among the r unknowns
  $x_{k_1} \dots x_{k_r}$ and one can then
  construct a solution as above
  which is $x_j = 1$.

- We thus have the following,
  one of the most fundamental facts about
  systems of homogeneous linear equations.

Theorem 6.
If A is an $m \times n$ matrix and $m < n$
then the homogenous system of linear equations
$AX = 0$ has a non-trivial solution.

Proof.

Let R be a row-reduced echelon matrix which is
row-equivalent to A (existence of R guaranteed by Thm 5)

Then the systems $AX = 0$ and $RX = 0$ have the same solutions
(Thm 3; equivalent systems, same solutions)

If r is the number of rows, then certainly, $r \leq m$.
Since $m < n$, it follows $r < n$.
Using the remark above
it follows that $AX = 0$
has a non-trivial solution.

■

Story:
- One can go further

Theorem 7.
Let A be an $n \times n$ matrix.
Then
(1) A is row-equivalent to the $n \times n$ identity matrix

$\Leftrightarrow$

(2) $AX = 0$ has only the trivial solution

Proof.

(1) ⇒ (2)

If A is row equivalent to I
then $AX = 0$ and $IX = 0$ have the same solution
and
clearly, $IX = 0$ has only the trivial solution.

(2) ⇒ (1)

Suppose $AX = 0$ has only the trivial solution $X = 0$.

Let R be an $n \times n$ row-reduced echelon matrix
which is row equivalent to A
Let r be the number of non-zero rows of R.

Now $RX = 0$ also has no nontrivial solutinos.

Thus, we must have $r \geq n$ (otherwise Thm 6 says
there are non-trivial solutions)

Further, R has n rows, i.e. $r \leq n$, thus we get $r = n$.

Now R has leading non-zero entry of 1 in each of the n rows
and since these 1s occur in a different one of the n columns
it follows that R must be the $n \times n$ identity matrix.

■

Story:

- So far, we have focused on homogeneous equations
- What can we say about
  non-homogeneous equations?

- One immediate difference—
  - homogeneous equations always have the trivial all-zero solution
  - inhomogeneous equations don't always have a solution

Notation.

Consider the system $AX = Y$.
We call A', an $m \times (n+1)$ matrix, the augmented matrix of this system
where
the first n columns of A' are simply those of A
and
the last column of A' is Y.

i.e.

$$A' = \left[\begin{array}{c|c} A & Y \end{array}\right]$$

Story:

Story:

- Suppose we perform a sequence of elementary row operations on A
  arriving at a row-reduced echelon matrix R.

- If this same sequence of operations is applied on A'
  - we will arrive at a matrix R'
    whose first n columns are the columns of R
    and
    whose last column contains certain scalars
    $z_1 \ldots z_m$.
  - Let

$$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_m \end{bmatrix}$$

  - This Z results from applying the same sequence of row operations
    on Y.

NB.

Using the same idea as in the proof of Thm 3,
one can deduce that $AX = Y$ and $RX = Z$ are equivalent
and have the same solutions.

Inf Claim:

It is easy to determine whether $RX = Z$ has any solutions
and
to determine all the solutions, if they exist.

Proof idea:

If R has r non-zero rows
with the leading non-zero entry of row i occuring in column k_i
(here $i \in \{1 \ldots r\}$) then
the first r equations of $RX = Z$ effectively express
$x_{k_1} \ldots x_{k_r}$ in terms of the $(n - r)$ remaining x_js and the scalars $z_1 \ldots z_r$.

The last $(m - r)$ equations are

$$\begin{aligned} 0 &= z_{r+1} \\ &\vdots \\ 0 &= z_m \end{aligned}$$

Now, the condition for the system to have a solution is
$z_i = 0$ for all $i > r$.

If this condition holds,
one can find all solutions to the system
just as in the homogeneous case
by assigning arbitrary values to the
$(n - r)$ of the x_js and
then computing x_{k_i} from the ith equation.

■

EXAMPLE 9. Let F be the field of rational numbers and

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & 1 & 1 \\ 0 & 5 & -1 \end{bmatrix}$$

and suppose that we wish to solve the system $AX = Y$ for some y_1, y_2, and y_3. Let us perform a sequence of row operations on the augmented matrix A' which row-reduces A:

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 2 & 1 & 1 & y_2 \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 5 & -1 & y_3 \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 5 & -1 & (y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(1)} \begin{bmatrix} 1 & -2 & 1 & y_1 \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix} \xrightarrow{(2)}$$

$$\begin{bmatrix} 1 & 0 & \frac{3}{5} & \frac{1}{5}(y_1 + 2y_2) \\ 0 & 1 & -\frac{1}{5} & \frac{1}{5}(y_2 - 2y_1) \\ 0 & 0 & 0 & (y_3 - y_2 + 2y_1) \end{bmatrix}.$$

The condition that the system $AX = Y$ have a solution is thus

$$2y_1 - y_2 + y_3 = 0$$

and if the given scalars y_i satisfy this condition, all solutions are obtained by assigning a value c to x_3 and then computing

$$\begin{aligned} x_1 &= -\tfrac{3}{5}c + \tfrac{1}{5}(y_1 + 2y_2) \\ x_2 &= \tfrac{1}{5}c + \tfrac{1}{5}(y_2 - 2y_1). \end{aligned}$$

Story:

- We end this subsection
  by making one final observation about the system $AX = Y$

Remark:

- Consider the system $AX = Y$.
  Suppose
  the entries of the matrix A and
  the scalars $y_1 \dots y_m$ happen to lie in a subfield F_1 of F
- If the system of equations has a solution with $x_1 \dots x_m$ in F then
  it has a solution with $x_1 \dots x_n$ in F_1.

Proof idea:

In either field
the condition for the system to have a solution
is that certain relations hold
between $y_1 \dots y_m$ in F_1
(i.e. the relations $z_i = 0$ for all $i > r$ where z_i were described above)

■

E.g. Suppose $AX = Y$ where A and Y are over reals.
If there is a solution in which $x_1 \dots x_n$ are complex
then there is a solution where $x_1 \dots x_n$ are real.

§ 1.5 Matrix Multiplication

Story:

- It would have become evident by now
  that the process of forming linear combinations of
  the rows of a matrix is a fundamental one.

- Therefore, it helps to be more systematic about such operations.

Motivation (matrix products)

- Suppose B is an $n \times p$ matrix
  over a field F with rows $\beta_1 \dots \beta_n$ and
  that from B we construct a matrix C with rows $\gamma_1 \dots \gamma_m$
  by forming certain linear combinations, i.e.
  $\gamma_i = A_{i1}\beta_1 + A_{i2}\beta_2 + \cdots + A_{in}\beta_n$

- The rows of C are determined the mn scalars A_{ij}
  (that we can view as entries of an $(m \times n)$ matrix A).

rows of B

$$C = \begin{matrix} A_{11}\beta_1 + A_{12}\beta_2 + \cdots + A_{1n}\beta_n \\ \vdots \\ A_{i1}\beta_1 + A_{i2}\beta_2 + \cdots + A_{in}\beta_n \\ \vdots \\ A_{m1}\beta_1 + A_{m2}\beta_2 + \cdots + A_{mn}\beta_n \end{matrix}$$

ith row of C

has as many columns as B
(∵ linear combⁿ of rows of B)

$$C_i = (C_{i1}, \dots C_{ip}) = \left(\sum_r A_{ir} B_{r1}, \sum_r A_{ir} B_{r2}, \dots, \sum_r A_{ir} B_{rp}\right)$$

$$= \sum_r (A_{ir} B_{r1}, A_{ir} B_{r2}, \dots, A_{ir} B_{rp})$$

and so, $C_{ij} = \sum_r A_{ir} B_{rj}$

Story:

Motivated by this observation, one can define a product between matrices as follows.

Definition.

Let A be an $m \times n$ matrix over the field F
Let B be an $n \times p$ matrix over F.

The **product** AB is the $m \times p$ matrix C
whose i, j entry is

$$C_{ij} = \sum_{r=1}^{n} A_{ir} B_{rj}.$$

EXAMPLE 10. Here are some products of matrices with rational entries.

(a) $$\begin{bmatrix} 5 & -1 & 2 \\ 0 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 5 & -1 & 2 \\ 15 & 4 & 8 \end{bmatrix}$$

Here

$$\begin{aligned} \gamma_1 &= (5 \quad -1 \quad 2) = 1 \cdot (5 \quad -1 \quad 2) + 0 \cdot (15 \quad 4 \quad 8) \\ \gamma_2 &= (0 \quad 7 \quad 2) = -3(5 \quad -1 \quad 2) + 1 \cdot (15 \quad 4 \quad 8) \end{aligned}$$

(b) $$\begin{bmatrix} 0 & 6 & 1 \\ 9 & 12 & -8 \\ 12 & 62 & -3 \\ 3 & 8 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 3 \\ 5 & 4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 6 & 1 \\ 3 & 8 & -2 \end{bmatrix}$$

Here

$$\begin{aligned} \gamma_2 &= (9 \quad 12 \quad -8) = -2(0 \quad 6 \quad 1) + 3(3 \quad 8 \quad -2) \\ \gamma_3 &= (12 \quad 62 \quad -3) = 5(0 \quad 6 \quad 1) + 4(3 \quad 8 \quad -2) \end{aligned}$$

(c) $$\begin{bmatrix} 8 \\ 29 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 5 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix}$$

(d) $$\begin{bmatrix} -2 & -4 \\ 6 & 12 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} \begin{bmatrix} 2 & 4 \end{bmatrix}$$

Here

$$\gamma_2 = (6 \quad 12) = 3(2 \quad 4)$$

(e) $$\begin{bmatrix} 2 & 4 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \end{bmatrix}$$

(f) $$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -5 & 2 \\ 2 & 3 & 4 \\ 9 & -1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(g) $$\begin{bmatrix} 1 & -5 & 2 \\ 2 & 3 & 4 \\ 9 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 9 & 0 \end{bmatrix}$$

Remarks.

1. Note that the product of two matrices need not be defined
   the product is defined $\Leftrightarrow$ the number of columns in the first matrix
   coincides with the number of rows in the second

2. When clear from the context,
   we write the products as AB
   without explicitly stating their sizes

3. Even when AB and BA are well defined
   it could be that $AB \neq BA$ (i.e. matrix multiplication is *not commutative*).

   See
   (d) and (e)
   (f) and (g)
   in the example above.

EXAMPLE 11.

(a) If I is the $m \times m$ identity matrix and A is an $m \times n$ matrix

(a) If I is the $m \times m$ identity matrix and A is an $m \times n$ matrix, $IA = A$.

(b) If I is the $n \times n$ identity matrix and A is an $m \times n$ matrix, $AI = A$.

(c) If $0^{k,m}$ is the $k \times m$ zero matrix, $0^{k,n} = 0^{k,m}A$. Similarly, $A0^{n,p} = 0^{m,p}$.

EXAMPLE 12. Let A be an $m \times n$ matrix over F. Our earlier shorthand notation, $AX = Y$, for systems of linear equations is consistent with our definition of matrix products. For if

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

with x_i in F, then AX is the $m \times 1$ matrix

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

such that $y_i = A_{i1}x_1 + A_{i2}x_2 + \cdots + A_{in}x_n$.

Notation:

One frequently uses the following notation.

Suppose B is an $n{\times}p$ matrix where the columns $B_1 \dots B_p$ of B, are $1{\times}n$ matrices given by

$$B_j = \begin{bmatrix} B_{1j} \\ \vdots \\ B_{nj} \end{bmatrix}, \qquad 1 \leq j \leq p.$$

The matrix B is the succession of these columns:

$B = [B_1 \dots B_p]$.

One can check that $AB = [AB_1, \dots, AB_p]$.

Story:

- Even though matrix multiplication turns out to not be commutative
  it is associative.

Theorem 8.

Let A,B,C be matrices (over the field F)
such that BC, and $A(BC)$ are well-defined.

Then,
the products AB and $(AB)C$ are also well defined
and
$A(BC) = (AB)C$.

Proof.

Suppose B is an $n \times p$ matrix.
Since BC is defined,
C is a matrix with p rows and
BC has n rows

Further, since $A(BC)$ is defined
we can assume A is an $m \times n$ matrix.

In summary,
A is an $m \times n$ matrix
B is an $n \times p$ matrix
C is a $p \times \ell$ matrix

Clearly then,
AB is well defined and is an $m \times p$ matrix
And $(AB)C$ is also well defined and is an $m \times \ell$ matrix.

It remains to show that $A(BC) = (AB)C$.

We have

$$
\begin{aligned}
[A(BC)]_{ij} &= \sum_r A_{ir} (BC)_{rj} && \text{by def}^n \\
&= \sum_r A_{ir} \sum_s B_{rs} C_{sj} && \text{by def}^n \\
&= \sum_r \sum_s A_{ir} B_{rs} C_{sj} \\
&= \sum_s \sum_r A_{ir} B_{rs} C_{sj} \\
&= \sum_s (AB)_{is} C_{sj} \\
&= [(AB)C]_{ij} .
\end{aligned}
$$

■

Remark/Notation:

- When A is an $n \times n$ matrix
  the product AA is well defined
  and
  we denote it by A^2.
- In fact, the general product $AA \dots A$ (k times)
  is also well defined
  and
  we denote it by A^k.

Story

- Fix some matrix B.

Now if C is obtained from B using elementary row operations
then clearly
each row of C is a linear combination of rows of B.

- Thus, there is a matrix A such that $AB = C$.

- In general, there are many such matrices A
  and among these, it helps to choose one with
  a few special properties.

- We quickly introduce a class of matrices and then
  resume the discussion above.

Definition.

An $m \times n$ matrix is said to be an **elementary matrix** if
it can be obtained from the $m \times m$ identity matrix
by means of a *single* elementary row operation.

EXAMPLE 13. A 2×2 elementary matrix is necessarily one of the following:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$$

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}, \quad c \neq 0, \quad \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}, \quad c \neq 0.$$

Theorem 9.

Let e be an elementary row operation and
let E be the $m \times m$ elementary matrix $E = e(I)$.

Then, for every $m \times n$ matrix A,
$e(A) = EA$.

Proof (partial).

The idea is that the entry in the ith row and jth column of
the product matrix EA is
obtained from the ith row of E and jth column of A.

The three types of elementary row operations
can be considered separately.

We do the proof for an operation of type (ii)
(the rest are easier and left as an exercise, by the textbook)

Recall, a type (ii) elementary row operation is
replace the rth row
with
row r + c times row s
(for $s \neq r$)

Suppose e is indeed such an operation

Suppose e is indeed such an operation.

Then,

$$E_{ik} = \begin{cases} \delta_{ik}, & i \neq r \\ \delta_{rk} + c\,\delta_{sk} & i = r \end{cases}$$

(recall $E_{ik} = e(\mathbb{1})_{ik}$

Therefore,

$$(EA)_{ij} = \sum_{k=1}^{m} E_{ik} A_{kj} = \begin{cases} A_{ik}, & i \neq r \\ A_{rj} + c\,A_{sj} & i = r \end{cases}$$

In other words, $EA = e(A)$.

■

Corollary.

Let A and B be $m \times n$ matrices over the field F.

Then

B is row-equivalent to A

$\Leftrightarrow$

$B = PA$ where P is a product of $m \times n$ elementary matrices

Proof.

The "⇒" direction (i.e. $\mathrm{B} = \mathrm{PA} \Rightarrow B$ is row equivalent to A)

Suppose $B = PA$ where

$P = E_s \dots E_2 E_1$

and

E_i are $m \times m$ elementary matrices.

Then, $E_1 A$ is row-equivalent to A

$E_2(E_1 A)$ is row-equivalent to $E_1 A$ and so on

Thus, $(E_s \dots E_1)A$ is row-equivalent to A.

The "⇐" direction (i.e. B is row equivalent to A implies $B = PA$)

Suppose B is row-equivalent to A.

Let $E_1 \dots E_s$ be the elementary matrices

corresponding to some sequence of elementary row operations

that take A to B.

Then, $B = (E_s \dots E_1)A$

(using Theorem 9).

■

§ 1.6 Invertible Matrices

Story.

- Since we have multiplicative inverses for Fields,
  one may wonder if what a reasonable notion of inverses
  should be for matrices.

- Indeed, such notion can be defined but one has to be a bit careful
  since matrix multiplication is not commutative.

Definition.

Let A be an $n \times n$ (square) matrix over the field F.
An $n \times n$ matrix B such that $BA = I$
is called a **left inverse** of A;

An $n \times n$ matrix B such that $AB = I$
is called a **right inverse** of A.

If $AB = BA = I$
then B is called a **two-sided inverse** of A
and
A is called **invertible**.

Story.

- It appears that we implicitly assumed that
  the left and right inverses must be the same.
- Turns out, that if both left and right inverses exist
  they must, in fact, be the same.

Lemma.

If A has a left inverse B and a right inverse C,
then $B = C$.

Proof.

Suppose $BA = I$ and $AC = I$.
Then

$$B = BI = B(AC) = (BA)C = IC = C$$

■

Remark.

If A has a left and a right inverse, A is invertible
and has a unique two-sided inverse.

Notation.

This is denoted by A^{-1} and simply called
the inverse of A.

Story:

- The following theorem states some properties

- The following theorem states some properties
  that we expect "inverses" to satisfy

Theorem 10.
Let A and B be $n\times n$ matrices over F.

(i) If A is invertible, so is A^{-1} and $(A^{-1})^{-1} = A$
(ii) If both A and B are invertible, so is AB
and $(AB)^{-1} = B^{-1}A^{-1}$.

Proof. Let E be an elementary matrix corresponding to the elementary row operation e. If e_1 is the inverse operation of e (Theorem 2) and $E_1 = e_1(I)$, then

$$EE_1 = e(E_1) = e(e_1(I)) = I$$

and

$$E_1E = e_1(E) = e_1(e(I)) = I$$

so that E is invertible and $E_1 = E^{-1}$. ∎

[the following story segment
seems confusing; skipped]

[to skip] Story:

- Suppose P is an $m\times m$ matrix
  which is a product of elementary matrices

- For each $m\times n$ matrix A
  the matrix $B = PA$ is row-equivalent to A

  Hence, A is row-equivalent to B
  and there is a product Q of elementary matrices
  such that
  $A = QB$.

- In particular,
  this is true when A is the $m\times m$ identity matrix.

  i.e. there is an $m\times m$ matrix Q (which itself is a product of elementary matrices)
  such that $QP = I$.

- As we shall see soon
  $QP = I$ is equivalent to the fact that
  P is a product of elementary matrices.