Chapter 2

Background

§4. Logic

Logical reasoning is as important in linear programming as it is in many other areas of science. It is only in college that problems may be presented to you in the form that you would immediately recognize as a linear program. On the contrary, you will usually be given data, some of which may even be extraneous, and it will be up to you to synthesize and collate the data until you recognize the kind of problem you want to solve. Think back to the examples we presented to you in §2 of Chapter 1. In no case did we start a problem by request that you maximize or minimize a linear form subject to certain linear constraints. Instead, we asked you to analyze a real-life situation, albeit a simple one, in order to come up with a detailed plan for activity. Taking the words and extracting from them the particular linear programming problem demand logical reasoning.

In order to develop your logical reasoning, you must be able to understand the mathematical meaning of certain words that might be different from the common English usage. For example, if we ask you, "Do you want coffee or tea?" we are using the word *or* to indicate that we expect you to choose between two beverages. However, if we told you, "the number -2 or 3 is a solution to the polynomial equation $x^2-x-6=0$," you would understand that both numbers are candidates for the solution. Mathematically speaking, you are not being asked to choose between them. In other words, *or* in mathematics usually means the inclusive disjunction. Notice that it is also the case in certain common everyday situations. If you were asked whether you wanted cream or sugar in your coffee, you would not be expected to choose only one. Some mathematical symbols including *or* are $\geq$, $\leq$, $\pm$.

Another seemingly simple word that has a more precise mathematical definition than it does in English is the word *and*. Think of the various ways in which you use this word. You use it when you are listing a series of objects: "I bought a skirt, a sweater, and a dress." You use this word as a conjunction: "I washed my hair and I brushed my teeth." However, we use the word *and* in the mathematical sense when we want to indicate that two (or more) statements must be satisfied simultaneously. For instance, the use of the word *and* in the statement,

"b is a positive integer
and b is a solution of the equation $x^2 - x - 6 = 0$"

indicates that b must equal 3. Do you see why?

Your understanding of mathematics will be enhanced if you gain facility in using the following words in their mathematical sense and become familiar with their logical symbols:

- and (logical symbol: $\wedge$)
- or (logical symbol: $\vee$)
- implies (logical symbol: $\Rightarrow$)
- follows from (logical symbol: $\Leftarrow$)

Sometimes, the symbols $\Rightarrow$ and $\Leftarrow$ are combined into one symbol $\Leftrightarrow$ which means the same as "if and only if" (iff for short), "that is," "i.e.," "is equivalent," "means," "$\Rightarrow\ \wedge\ \Leftarrow$."

There are other equivalence relations in mathematics and elsewhere besides the logical equivalence. In saying "these theorems are equivalent," we often try to say something different from "either these theorems are both true or both false."

Usually, the comma means "and." For example, $x > 0, y > 0$ means $x > 0$ and $y > 0$. This can be also written as $x, y > 0$. Constraints in linear programs are often connected by commas, which means we want to satisfy all of them. However, in some situations the comma means something else. For example,

$$x = 0, 1, \text{or } 2$$

means that x takes one of these three values. In the example

$$x = 2 + 3, \text{ or } x = 5$$

the comma changes the meaning of *or* to *i.e.*

We often have at our disposal many words that have the same meaning mathematically. For instance, in the following true statement

$$x \geq 0 \text{ if } x = 1 \tag{4.1}$$

we can replace *if* by any of the following words or expressions:

if, when, since, provided that, whenever, is weaker than, follows from, is a consequence of, because, is implied by, $\Leftarrow$.

In particular, we can write (4.1) as follows: $x \geq 0 \Leftarrow x = 1$. We can also rewrite (4.1) as

$$\text{If } x = 1, \text{ then } x \geq 0$$

or, equivalently,

$$x = 1 \text{ implies } x \geq 0. \tag{4.2}$$

We can replace *implies* in (4.2) by any of the following expressions:

implies, is stronger than, only if, so, hence, results in, forces, gives, whence, therefore, thus, consequently, $\Rightarrow$.

Example 4.3. The following seventeen sentences have the same mathematical meaning:

- $x \geq 0$, because $x \geq 2$.
- $x \geq 0$, if $x \geq 2$.
- $x \geq 2$ only if $x \geq 0$.
- If $x \geq 2$, then $x \geq 0$.
- The bound $x \geq 2$ is sharper $x \geq 0$.
- Given $x \geq 2$, we conclude that $x \geq 0$.
- The bound $x \geq 2$ is better than $x \geq 0$.
- The constraint $x \geq 0$ is less tight than $x \geq 2$.
- The bound $x \geq 2$ is more precise than $x \geq 0$.
- In the view of condition $x \geq 2$, we have $x \geq 0$.
- The constraint $x \geq 0$ is less severe than $x \geq 2$.
- The constraint $x \geq 0$ is less strict than $x \geq 2$.
- The condition $x \geq 2$ implies the constraint $x \geq 0$.
- The constraint $x \geq 0$ is less stringent than $x \geq 2$.
- The constraint $x \geq 0$ is less demanding than $x \geq 2$.
- The condition $x \geq 2$ is sufficient to conclude that $x \geq 0$.
- The linear constraint $x \geq 0$ follows from the condition $x \geq 2$.

Remark. In definitions, *if* is used sometimes for "if and only if."

Many statements or conditions in this book are constraints on variables. Every constraint or a system of constraints gives a set, the feasible region. The feasible region corresponding to a stronger system is a part of the feasible region corresponding to a weaker system. Adding new constrains reduces the feasible region. For example, the point $x = 1$ belongs to the ray $x \geq 0$, which means that the condition $x = 1$ implies that $x \geq 0$. Equivalent systems of constraints imply each other, and they have the same feasible region. In the terms of feasible regions, *and* means the intersection, while *or* means the union. For example the condition "$x = 0$ and $y = 0$" gives a point in plane, while the condition "$x = 0$ or $y = 0$" gives the union of two straight lines.

Here is a more sophisticated example. Consider the following statement: $0 = 1$ implies $0 = 0$. This statement is true because any false statement implies everything in general. Also, truth follows from everything in general. In particular, we do not need any conditions to conclude that $0 = 0$. Another way to see that the statement is true is the following argument: given $0 = 1$, we can conclude that $0 = 1 = 0$, hence $0 = 0$. Multiplication of the equation $0 = 1$ by 0 also gives the desired conclusion. Finally, in the terms of the feasible sets our statement is as follows: Nothing is a part of everything.

Now we consider more complicated linear equations. Given two linear equations

$$\begin{cases} x + 3y = 3 \\ x + 2y = 5, \end{cases}$$

we can take their difference and conclude that $y = -2$. Thus, the equation $y = -2$ follows from the system. More generally, given m equations, $f_1 = b_1, \ldots, f_m = b_m$, and any numbers c_i, we can take *linear combination of the given equations with the coefficients* c_i :

$$c_1 f_1 + \cdots + c_m f_m = c_1 b_1 + \cdots + c_m b_m.$$

This equation follows logically from the system, which means that every solution of the system satisfies the equation. Once we know how to solve systems of linear equations, it is easy to show that the converse is true in the case when the system is consistent and all equations are linear equations.

Namely, given a system of linear equations $f_1 = b_1, \ldots, f_m = b_m$, and another linear equation $f_0 = b_0$ in standard form that follows from the system, then either the equation $f_0 = b_0$ or the equation $0 = 1$ is a linear combination of the equations in the system. See §6 of this chapter for the general case and Exercises 38–41 for particular cases. A similar statement is true for linear inequalities if we are careful with the signs of coefficients (see §15 of Chapter 5).

Let us return to our small talk about logic in general. There is no way to list all English equivalents of logical symbols and their uses and abuses. Here is a list of common fallacies:

- $x \geq 0 \Rightarrow x = 0$ or 1 (false dilemma).
- $0 = 0$ so $0 = 1$ (argument from ignorance).
- $x = 1/3$, or $x = 0.33333$ as decimal (slippery slope).
- It is not true that $1 = 1$ or $0 = 1$ (complex conclusion).
- $0 > 1$ because it is what my instructor teaches (appeal to force).
- $0 > 1$ because I spent all night to get this result (appeal to pity).
- The minus in your equation $x^2 = -1$ is a mistake, for otherwise solving it would be waste of time (appeal to consequences).
- Every reasonable person will agree that $1/3 = 0.3$ (prejudicial language).
- Everybody knows that $1/3 = 0.3$ so it is true (appeal to popularity).
- I cannot agree that $0 = 1$ is a linear constraint because you even cannot spell constraint (attacking the person).
- The linear equation $0 = 1$ has a real solution because the *New York Times* wrote so and experts agreed (appeal to authority and anonymous authorities).
- All examples in Chapter 1 have less than 10 constraints so every linear program has less than 10 constraints (hasty generalization).
- I had no difficulties in the first week of classes, so this course is a piece of cake, and I do not need to work hard to pass (unrepresentative sample).
- To solve systems of linear equations, I add and subtract equations, so I will use the same operations to simplify my system of linear constraints (false analogy; while the sum of equations follows from the equations, this is not the case with inequalities of different types).
- Every linear equation for one unknown has a solution (fallacy of exclusion).
- This section follows Chapter 1; therefore, we do not need any logic in Chapter 1 but we need to know what linear programming is to be logical (coincidental correlation).
- Since all constraints in my optimization problem are linear, the objective function should be affine (joint effect; the assumption and the conclusion are both true for linear programs).
- $x > 0$ causes $x > 1$ (wrong direction).
- Complex numbers are not really numbers (equivocation).
- There is no solution for the equation $0 = 1$ (amphiboly, i.e., two different meanings).

• In linear programming all numbers in data and solutions are real, so every linear program has a feasible solution (existential fallacy).

If your head does not spin yet, what do you think about the logic in this statement: "Since I really understood the diet problem with 10 cereals, I am going to eat only cereals from now on." For more examples, see http://www.intrepidsoftware.com/fallacy/toc.php.

Learning logic is similar to learning to walk: It takes patience and practice. We conclude this section with quotations. It is up to the reader to judge whether there is any logic in them.

Jacques Hadamard (1865–1963):
Logic merely sanctions the conquests of the intuition.

Antoine Arnauld (1612–1694):
Common sense is not really so common.

Ludwig Wittgenstein (1889–1951):
There can never be surprises in logic.

Morris Kline (b. 1908):
Logic is the art of going wrong with confidence.

Lord Dunsany (1878–1957):
Logic, like whiskey, loses its beneficial effect when taken in too large quantities.

Oliver Heaviside (1850–1925):
Logic can be patient, for it is eternal.
Why should I refuse a good dinner simply because I don't understand the digestive processes involved?

G. K. Chesterton (1874–1936):
You can only find truth with logic if you have already found truth without it.

Hermann Weyl (1885–1955):
Logic is the hygiene the mathematician practices to keep his ideas healthy and strong.

Richard Feynman (1918–1988):
. . . mathematics is not just another language. . . . it is a language plus logic. Mathematics is a tool for reasoning.

Jeremy Bentham (1748–1832):
O Logic: born gatekeeper to the Temple of Science, victim of capricious destiny: doomed hitherto to be the drudge of pedants: come to the aid of thy master, Legislation.

Jules Henri Poincaré (1854–1912):
It is by logic we prove, it is by intuition that we invent. Thus, be it understood, to demonstrate a theorem, it is neither necessary nor even advantageous to know what it means. The geometer might be replaced by the "logic piano" imagined by Stanley Jevons; or, if you choose, a machine might be imagined where the assumptions were put in at one end, while the theorems came out at the other, like the legendary Chicago machine where the pigs go in alive and come out transformed into hams and sausages. No more than these machines need the mathematician know what he does.

Bertrand Russell (1872–1970):
Ordinary language is totally unsuited for expressing what physics really asserts, since the words of everyday life are not sufficiently abstract. Only mathematics and mathematical logic can say as little as the physicist means to say.

David van Dantzig (1900–1959):
Neither in the subjective nor in the objective world can we find a criterion for the reality of the number concept, because the first contains no such concept, and the second contains nothing that is free from the concept. How then can we arrive at a criterion? Not by evidence, for the dice of evidence are loaded. Not by logic, for logic has no existence independent of mathematics: it is only one phase of this multiplied necessity that we call mathematics. How then shall mathematical concepts be judged? They shall not be judged. Mathematics is the supreme arbiter. From its decisions there is no appeal. We cannot change the rules of the game, we cannot ascertain whether the game is fair. We can only study the player at his game; not, however, with the detached attitude of a bystander, for we are watching our own minds at play.

Exercises

1–31. Now it is your turn! Determine the validity of the following 31 statements. Write down yes or true if you agree with the statement and write no or false otherwise. Explain your reasoning.

1. $|x| = 1$ only if $x \geq 0$.

2. $xy = 0$ only if $y = 0$.

3. $|x| \leq 1$ if $x \leq 1$.

4. If $|x| \leq 1$, then $x \geq -1$.

5. $x \neq 1$ unless $x \geq 0$.

6. $|x| > 1$ hence $x > 0$.

7. $x \geq 0$ provided that $x > 2$.

8. If $0 = 1$, then $2 = 5$.

9. $x \geq 0 \Leftarrow x > 2$.

10. $x^2 = 0 \iff x = 0$.

11. The condition $x = 1$ does not imply the condition $x \geq 1$.

12. The condition $|x| = 1$ is stronger than the condition $x \geq 0$.

13. The condition $x \geq 0$ follows from the condition $x = 5$.

14. $|x| \leq 1$ if and only if $x \leq 1$ *and* $x \geq -1$.

15. $|x| \leq 1$ if and only if $x \leq 1$ *or* $x \geq -1$.

16. The equation $x^2 = 0$ is equivalent to the constraint $x = 0$.

17. The equation $x = y$ is equivalent to the constraint $x - y = 0$.

18. The constraint $|x| \geq 1$ means that either $x \leq -1$ or $x \geq 1$.

19. $|x| \geq 1$ if and only if $x \leq -1$ and $x \geq 1$.

20. If $xy = 0$ then $x = 0$ or $y = 0$.

21. The second condition in the system $x \geq 1, x \geq 0$ is redundant.

22. Given $0 = 1$, we can conclude that $2 = 5$.

23. The condition $x > 10$ is sufficient for the conclusion $x \geq 0$.

24. The constraint $x = 3$ forces $x \geq 0$.

25. The condition $x^2 > 10$ makes x positive.

26. If $|x - 1| > 4$, then either $x > 5$ or $x < -3$.

27. If $x > y$ then $-x < -y$.

28. The condition $x > 0$ is necessary but not sufficient for $x > 2$.

29. The condition $x = 1$ is weaker than the condition $x \geq 1$.

30. The condition $x = 1$ is a consequence of the condition $x \geq 1$.

31. $|x| \leq 1$ means that $x \leq 1$ and $x \geq -1$. ■

32–39. Find all implications between the following four conditions:

32.

(i) $x = 2, y = 3$ (ii) $x \geq 0$

(iii) $y \geq 0$ (iv) $x + y = 5$

33.

(i) $x = 2, y = 3$ or $x = y = 0$ (ii) $x, y \geq 0$;

(iii) x, y are integers (iv) $|x + y| \leq 5$

34.

(i) $0 = 1$ (ii) $0 = 0$

(iii) $1 = 2$ (iv) $x = y$

35.

(i) $x = 1, y = 3$ (ii) $x + y = 3, x - y = 1$

(iii) $x = 1$ or $y = 2$ (iv) $2x + 3y = 8$

36.
(i) $x + y \geq 1, x - y \geq 2$ (ii) $2x \geq 3$
(iii) $2y \geq -1$ (iv) $x, y \geq 0$

37.
(i) $x = 0$ (ii) $x^3 = 0$
(iii) $0 = 0$ (iv) $xy = 0$

38.
(i) $x^2 + y^2 = 1$ (ii) $x^2 + y^2 \leq 1$
(iii) $x \leq 1, y \leq 1$ (iv) $x + y \leq 2$

39.
(i) $|x| + |y| \leq 1$ (ii) $x^2 + y^2 \leq 1$
(iii) $|x| \leq 1, |y| \leq 1$ (iv) $x^4 + y^4 \leq 1$ ■

40. What are other possible replacements for *if* in (4.1)?

41. What are other possible replacements for *implies* in (4.2)?

42–48. Do you agree?
42. This optimization problem is linear, so the objective function must be linear.
43. If every constraint in a linear program is feasible, then the program is feasible.
44. If a linear problem is infeasible, then one of given constraints must be infeasible.
45. If x =3 and $y = -1$, then x is closer to 0 than is y.
46. $x \geq 0$ because $x > y$ and $y > 3$.
47. The constraint $a + 2b + 3c \geq 2$ follows from the system

$$a + b + c \geq 1, b + 2c \geq 1.$$

48. The constraint $a + 2b + 3c \leq 3$ follows from the system

$$a + b + c \leq 1, b + 2c \leq 1. \quad ■$$

49–52. Check whether the third equation follows from the first two equations (i.e., it is redundant in the system). If it is, write it as a linear combination of the first two equations.

49. $\begin{cases} x = 0, \\ 2y = 5, \\ x + y = 2. \end{cases}$ **50.** $\begin{cases} x + y + z = 1, \\ x + 2y + 3z = 3, \\ 2x + 3y + 4z = 4. \end{cases}$

51. $\begin{cases} a - b + c + d = 0, \\ 2a + b - 3c = 1, \\ 3b - 5c - 2d = 1. \end{cases}$ **52.** $\begin{cases} a - b + c + d = 0, \\ 2a + b - 3c = 1, \\ -a - 2b + 4c + d = -1. \end{cases}$

§5. Matrices

Matrices are used often in linear algebra and linear programming. They allow us to write down systems of linear equations and inequalities in an abbreviated form. We will start by defining matrices and then we will see how linear programming problems can be written in a kind of "shorthand" notation by using matrices.

Definition 5.1. A *matrix* is a rectangular array of entries, which can be numbers, variables, polynomials, functions, and so on. ∎

In general, a matrix A can be written in the following form:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

where $[a_{11} \quad a_{12} \quad \ldots \quad a_{1n}]$ is referred to as the first row of the matrix, $[a_{21} \quad a_{22} \quad \ldots \quad a_{2n}]$ is the second row, and so forth. Similarly,

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

is the first column,

$$\begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix}$$

is the second column, and so forth. In this case we say that the matrix A has m rows and n columns. We refer to A as an $m \times n$ matrix. We denote by a_{ij} the entry in the i^{th} row and the j^{th} column.

When $m = 1$ (i.e., our matrix has only one row), we have a row matrix or just a row, also called a row vector or just vector. When $n = 1$, we have a column matrix or a column, also referred to as a column vector or just vector.

Remark. The subscript in a_{ij} means a pair of numbers rather than a product. Use $a_{12,3}$ or $a_{1,23}$ instead of a_{123} for large m, n. We often use commas in row matrices to avoid mix-ups. Compare [1 2 34] and [1, 2, 34]. ∎

Example 5.2.

(*i*) $A = \begin{bmatrix} 1 & -1 & 3 \\ \sqrt{2} & \pi & 6 \\ -4 & 56 & 20 \end{bmatrix}$

is a 3×3 matrix with real entries.

(*ii*) $B = \begin{bmatrix} 1 & x & x^2 \\ 1 & y^2 & y^4 \end{bmatrix}$

is a 2×3 matrix with polynomial entries.

(*iii*) $\begin{bmatrix} x \\ y^2 \end{bmatrix}$ is the second column of B.

(*iv*) $C = \begin{bmatrix} \sin x & e^x & \ln x & x+1 \\ \sqrt{x^2 - \pi^2} & 3/4 & 1/x & 5 \\ e^{7\cos x} & \tan \pi x & 2 & 5 \end{bmatrix}$

is a 3×4 matrix with functional entries.

(*v*) $\begin{bmatrix} \sqrt{x^2 - \pi^2} & \frac{3}{4} & \frac{1}{x} & 5 \end{bmatrix}$ is the second row of C.

(*vi*) If we denote the entries of the matrix C by c_{ij}, where $1 \le i \le 3$ and $1 \le j \le 4$, then, since $\tan \pi x$ is in the third row and the second column, it is the entry c_{32}.

Definition 5.3. Two matrices A and B are *equal* if they have the same size and each entry of A is equal to the corresponding entry of B. ■

In linear programming we also use inequalities between rows (or columns) of the same size. Unless stated otherwise, $A \ge B$ for two vectors of the same size means that every entry of A is greater than or equal to the corresponding entry of B.

Recall your introduction to the set of rational or real numbers. After the set was defined, the operations of addition, subtraction, multiplication, and division and their properties were explained. For instance, both addition and multiplication of numbers are commutative and associative; we can link addition and multiplication via the property known as distributivity of multiplication over addition. Let us do something similar with our newly defined objects called matrices.

Definition 5.4. The operation of *addition* of two $m \times n$ matrices, A and B, is defined component-wise; that is, the entry in the i^{th} row and the j^{th} column of the $m \times n$ matrix $A+B$ is $a_{ij}+b_{ij}$. Subtraction is defined in a similar way. ■

It is easy to verify that matrix addition is commutative and associative.

Definition 5.5. The *product* of matrices, A and B, is defined only when the number of columns of A equals the number of rows of B. If A is an $m \times n$ matrix and B is an $n \times p$ matrix, the product $A \cdot B$ (or just AB) is an $m \times p$ matrix, whose entry in the i^{th} row and the j^{th} column is given by

$$(A \cdot B)_{ij} = \sum_{k=1}^{n} a_{ik}\, b_{kj} \qquad ■$$

Because of the nature of the definition of the product of two matrices, this operation is not commutative. The product in the reverse order, $BA = B \cdot A$, may not be defined, and even if it is, it may differ from $A \cdot B$.

However, matrix multiplication is associative. What do we mean by that? Let A be an $m \times n$ matrix, B an $n \times p$ matrix, and C a $p \times q$ matrix. Then the products $A \cdot B$, $(A \cdot B) \cdot C$, $B \cdot C$, and $A \cdot (B \cdot C)$ are all defined and, moreover,

$$(A \cdot B) \cdot C = A \cdot (B \cdot C).$$

Recall from your knowledge of the properties of real numbers that distributivity of multiplication over addition gives us the equality

$$a \cdot (b + c) = a \cdot b + a \cdot c.$$

Distributivity of matrix multiplication over addition holds as well, provided that the appropriate sums and products of matrices are defined. For example, if A is an $m \times n$ matrix, and B and C are two $n \times p$ matrices, then, $A \cdot (B + C)$ yields the same $m \times p$ matrix as does $A \cdot B + A \cdot C$. Therefore,

$$A \cdot (B + C) = A \cdot B + A \cdot C.$$

Proofs of these claims can be found in any linear algebra textbook, so we will omit them. Here are some more examples.

Example 5.6.
Let A and B be the following matrices:

$$A = \begin{bmatrix} 1 & -1 & 3 \\ 2 & 1/2 & 6 \end{bmatrix}, \qquad B = \begin{bmatrix} 1/2 & -3/4 & 7 \\ 2 & 3 & 8 \end{bmatrix}.$$

(i) Then their sum, $A+B$, and their difference, $A-B$, are given by

$$A + B = \begin{bmatrix} 3/2 & -7/4 & 10 \\ 4 & 7/2 & 14 \end{bmatrix} \text{ and } A - B = \begin{bmatrix} 1/2 & -1/4 & -4 \\ 0 & -5/2 & -2 \end{bmatrix}.$$

Example 5.7.
Let C and D be the following 2×3 and 3×3 matrices:

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \text{ and } D = \begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}.$$

Then the product, $C \cdot D$, is the following 2×3 matrix:

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

$$= \begin{bmatrix} 30 & 24 & 18 \\ 84 & 69 & 54 \end{bmatrix}$$

Notice that the product $D \cdot C$ is not defined. Do you see why?

Example 5.8.
Let A be the 2×3 matrix

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

x the 3×1 column matrix

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

and b the 2×1 column matrix

$$b = \begin{bmatrix} -5 \\ 6 \end{bmatrix}.$$

Then the product $A \cdot x$ is a well-defined 2×1 column matrix. Therefore, the matrix equality $A \cdot x = b$ makes sense, and, by performing these operations, we obtain the system of linear equations

$$\begin{cases} \quad x_1 - 2x_2 \qquad = -5 \\ -3x_1 + 4x_2 + x_3 = 6. \end{cases}$$

Conversely, the prededing system of equations can be written in the matrix form $A \cdot x = b$. This observation leads the following elementary but important result:

Any system of linear equations can be written in the standard matrix form

$$Ax = b,$$

where x is the column of distinct unknowns (variables), A is a given matrix (the coefficient matrix), and b is a column of given numbers (constant terms).

Note that the number of equations need not be equal to the number of variables. In other words, the coefficient matrix need not be a square matrix.

Definition 5.9. The *transpose* matrix of an $m \times n$ matrix A is the $n \times m$ matrix whose entry in the $i\,j$-position is the entry in the $j\,i$-position of the original matrix A. We denote the transpose of A by A^T. Thus, $(A^T)_{ij} = a_{ji}$. ■

Proposition. Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Then $A \cdot B$ and $B^T \cdot A^T$ are defined and $(A \cdot B)^T = B^T \cdot A^T$.

Sketch of Proof

(i) Compute the product $A \cdot B$.

(ii) Find the transpose of the matrix $A \cdot B$.

(iii) Calculate the product $B^T \cdot A^T$.

(iv) Compare the entries of the matrices $(A \cdot B)^T$ and $B^T \cdot A^T$.

Example 5.10.
The transpose of the 3×4 matrix

$$C = \begin{bmatrix} \sin x & e^x & \ln x & x+1 \\ \sqrt{x^2 - \pi^2} & 3/4 & 1/x & 5 \\ e^{7\cos x} & \tan \pi x & 2 & 5 \end{bmatrix}$$

is the following 4×3 matrix:

$$C^T = \begin{bmatrix} \sin x & \sqrt{x^2 - \pi^2} & e^{7\cos x} \\ e^x & 3/4 & \tan \pi x \\ \ln x & 1/x & 2 \\ x+1 & 5 & 5 \end{bmatrix}. \qquad ■$$

The real numbers 0 and 1 are distinguished elements for addition and multiplication, respectively, because adding 0 to a number does not affect the number and, similarly, multiplication of any number by 1 yields the same number with which we started. We define matrices with similar properties.

Definition 5.11. The zero matrix 0 is the $m \times n$ matrix with entries consisting of zeros only. ■

Remark. The size of a zero matrix is often clear from context and need not to be specified. For example, the first 0 in the inequality $[0, 1, 2, 3] \geq 0$ means a number, while the second zero can be understood as an 1×4 matrix. In the equality

$$[1, -1, 2] \begin{bmatrix} 3 \\ 5 \\ 1 \end{bmatrix} = 0$$

the 0 on the right-hand side means a 1×1 matrix or a number. ■

The *additive inverse* of an $m \times n$ matrix, A, is the $m \times n$ matrix consisting of the additive inverses of the entries of A; we denote the additive inverse of A by $-A$. Thus, $A + (-A) = 0$ for a matrix A of any size. By contrast, not every matrix has a multiplicative inverse. Recall that, even for real numbers, the number 0 is not invertible.

Definition 5.12. For a positive integer n we define the $n \times n$ identity matrix, 1_n, as follows: Its diagonal entries are ones and the other entries are zeros. ■

Recall that the diagonal entries of a matrix $[a_{ij}]$ are a_{ii}. The identity matrix 1_n is a square diagonal matrix. In general, a *diagonal matrix* is defined as a matrix with all nondiagonal entries = 0.

Here is an example of a diagonal 2×3 matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Here are two matrices that are not diagonal (can you see why?):

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 5 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & -1 & \pi & 1/2 \\ 0 & 0 & 1 & 3 & 4 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The 2×2 identity matrix 1_2 is the following matrix:

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Here is the 5×5 identity matrix:

$$I_5 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Definition 5.13. An $n \times n$ matrix, A, is said to be *invertible* if there exists an $n \times n$ matrix B such that $A \cdot B = B \cdot A = I_n$.

Proposition 5.14. Let A be an invertible $n \times n$ matrix. Then the inverse of A is unique.

Proof. Suppose that we have two inverses, B and C, for the invertible matrix A. By the definition, we have $A \cdot B = I_n$ and $C \cdot A = I_n$. Therefore, since, $C = C \cdot I_n$ and $A \cdot B = I_n$, we obtain $C = C \cdot (A \cdot B)$. Since matrix multiplication is associative, the latter product equals $(C \cdot A) \cdot B$ and this equals $I_n \cdot B = B$, because $C \cdot A = I_n$. Combining these equalities, we obtain that $C = B$. ■

From now on, we will use the notation A^{-1} to denote the unique inverse of the $n \times n$ invertible matrix A. Before we continue this presentation of matrices, here are some questions worth thinking about:

(i) How can we calculate the inverse of a matrix?

(ii) Does every nonzero $n \times n$ matrix, A, have an inverse?

(iii) If not, what conditions should the $n \times n$ matrix, A, satisfy in order to have an inverse?

In order to answer these questions, we need to define some operations on matrices that do not have an analog to operations in the set of real numbers.

Definition 5.15. Let A be an $m \times n$ matrix. An *elementary row (column) operation* on A is one of the following procedures, performed on the rows (columns) of A:

(i) Interchange two rows (columns) of A.

(ii) Multiply a row (column) of A by a nonzero number.

(iii) Add a multiple of a row (column) of A by a number to another row (column). ■

Example 5.16. The result of interchanging the first and the third rows of

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

is the matrix

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

Note that $B = PA$ and $A = PB$, where

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

is a permutation matrix. The same permutation operation takes us from B back to A.

Example 5.17. By multiplying the second row of

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

by $\sqrt{2}$, we obtain the matrix

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 4\sqrt{2} & 5\sqrt{2} & 6\sqrt{2} \end{bmatrix}.$$

Note that $B = DA$ and $A = D^{-1}B$, where

$$D = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2} \end{bmatrix}, \quad D^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{2}/2 \end{bmatrix}$$

are diagonal matrices that differ from the identity matrix 1_2 in one diagonal entry. Another row multiplication operation takes us from B back to A.

Example 5.18. The matrix

$$B = \begin{bmatrix} 1 & 2 \\ -12 & -14 \\ 5 & 6 \end{bmatrix}$$

is obtained by replacing the second row of

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

by the sum of its second row and -3 times its third row. A similar row addition operation, with -3 replaced by 3, takes us from B back to A. This row addition operation corresponds to multiplication by an elementary matrix on the left: $B = EA$ and $A = E^{-1}B$, where

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{bmatrix}, \quad E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}.$$

This elementary matrix differs from the identity matrix in one off-diagonal entry. ∎

In general, row (column) elementary operations correspond to multiplication by invertible matrices from the left (right).

Remark 5.19. For an invertible matrix A, the linear system $Ax = b$ has exactly one solution $x = A^{-1}b$. In general, solving the system $Ax = b$ is related with inverting a submatrix A of maximal possible size.

Remark 5.20. Sometimes numbers are called *scalars* to distinguish them from vectors. However, numbers can be also considered as 1×1 matrices; hence they can be considered as both row and column vectors. Addition and multiplication of numbers agrees with matrix addition and multiplication. However, it is tricky to interpret multiplication of a matrix by a scalar in terms of matrix multiplication. An interpretation involves *scalar matrices,* which are square diagonal matrices with the same diagonal entries.

Vectors with n components, written in a row or column, can be thought of as points or arrows (directed line segments) sticking out of the origin in an n-dimensional space. This language came from the cases $n = 1, 2, 3$, where those objects appear in geometry and mechanics. Matrices appear in connection with some geometric transformations. ∎

In conclusion, here is a quotation about matrices by Irving Kaplansky (from *Paul Halmos: Celebrating 50 Years of Mathematics,* New York: Springer-Verlag, 1991):

> We [he and Halmos] share a philosophy about linear algebra: we think basis-free, we write basis-free, but when the chips are down we close the office door and compute with matrices like fury.

Exercises

1–8. Let A equal the 1×4 matrix $[1, 2, 0, -3]$ and B equal the 1×4 matrix $[0, -1, -2, 4]$. Compute

1.	$2A + 3B$	**2.**	AB^T
3.	BA^T	**4.**	$A^T B$
5.	$B^T A$	**6.**	$(A^T B)^2$
7.	$(A^T B)^3$	**8.**	$(A^T B)^{1000}$

9. Could you compute the products AB or BA? Why or why not?

10. Find a square matrix A such that $A \neq 0$ but $A^2 = 0$.

11. Find two matrices, A and B, such that both matrix products AB and BA are defined, $AB = 0$ and $BA \neq 0$.

12–14. Write the system of equations in a matrix form. *Hint* : some of these equations are not linear equations in standard form (see page 4).
12. $5a + 2b + 3c - d = 1, -b + a - 3c = 2, 6c + a - 3c + 1 = a.$
13. $x + 2b + 3c - y = 1, -b + a - 3c = 2y, x + a - 3c + 1 = a.$
14. $x + 2b + 3c - y = 1, -b + a - 3c = 2y + d, x + 6a - y + 1 = a.$

15–17. Solve the systems 12–14.

18–25. Do Exercises 1–8 for matrices

$$A = [0, 1, -1, 0, 0, 2] \text{ and } B = [-2, 1, 0, 0, 3, 2].$$

26–33. Do Exercises 1–8 for matrices

$$A = [1, 1, -1, 0, 0, 2, 0] \text{ and } B = [-1, 1, 0, 3, 3, 2, -1].$$

34. Do Exercises 2–7 for matrices A, B in Example 5.6.

35. For matrix C in Example 5.7, compute E_1C and E_2C, where $E_1 = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$ (a diagonal matrix), $E_2 = \begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix}$ (an elementary matrix). Also compute $(E_1)^2, (E_1)^9, (E_2)^{40}$.
36. For matrices C, D in Example 5.7, compute $CE_1, CE_2, DE_1,$ $DE_2, E_1E_2, E_2E_1, (E_1)^2, (E_1)^9, (E_2)^{40}$, where $E_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}$ (a diagonal matrix), $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$ (an elementary matrix).

The point of Exercises 35–36 was to realize how to multiply by elementary and diagonal matrices (definition will be reminded in the next section). In general, zeros in a matrix help when you multiply by this matrix. ■

37. Let α be an invertible $m \times m$ matrix, β a $m \times n$ matrix, γ a $n \times m$ matrix, and δ a $n \times n$ matrix. Compute

$$\begin{bmatrix} 1_m & 0 \\ -\gamma\alpha^{-1} & 1_n \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} 1_m & -\alpha^{-1}\beta \\ 0 & 1_n \end{bmatrix}.$$

38. A matrix $A = [a_{i,j}]$ is called *diagonal* if $a_{i,j} = 0$ whenever $i \neq j$. Show that the sum of two diagonal matrices of the same size is a diagonal matrix. Show that the product of two diagonal matrices, when defined, is a diagonal matrix. Show that $AB = BA$ for two $n \times n$ diagonal matrices A, B.

39. A matrix $A = [a_{i,j}]$ is called *upper triangular* if $a_{i,j} = 0$ whenever $i > j$. Show that the sum of two upper triangular matrices of the same size is an upper triangular matrix. Show that the product of two upper triangular matrices, when defined, is an upper triangular matrix. Show that $AB \neq BA$ for some $n \times n$ upper triangular matrices A, B.

40. A matrix $A = [a_{i,j}]$ is called *lower triangular* if $a_{i,j} = 0$ whenever $i < j$. Show that the sum of two lower triangular matrices of the same size is a lower triangular matrix. Show that the product of two lower triangular matrices, when defined, is a lower triangular matrix. Show that $AB \neq BA$ for some $n \times n$ lower triangular matrices A, B.

41–44. Obtain a diagonal matrix by row and column elementary operations with the matrix A. *Remark*: The diagonal matrix in the answer is not unique, but the number of zero columns in it is unique. In many textbooks on linear algebra it is proved that this number equals the dimension of the space of solutions of $Ax = 0$.

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

42. $A = \begin{bmatrix} 1 & 0 & -1 \\ -5 & 1 & 3 \\ 2 & 4 & 5 \\ 8 & -2 & -1 \end{bmatrix}$

43. $A = \begin{bmatrix} 1 & -2 & -1 \\ 5 & 1 & 3 \\ 2 & 4 & 5 \\ -1 & -2 & 0 \\ 8 & -2 & -1 \end{bmatrix}$

44. $A = \begin{bmatrix} 1 & -2 & -1 & 0 \\ 5 & 1 & 3 & 1 \\ 5 & 6 & 0 & 7 \end{bmatrix}$

§6. Systems of Linear Equations

The main object of this section is to recall how to solve systems of linear equations. To save paper we use matrices.

One of the main applications of elementary operations is solving systems of linear equations. Every system of linear equations can be written in a matrix form $Ax = b$, where A is a given matrix, x is a column of distinct variables (unknowns), and b is a column of given numbers. The coefficient matrix A and the column b can be put together as the augmented matrix $[A|b]$.

Example 6.1.
The system $\begin{cases} x - 2z = -3 \\ 2x + 5y - 4z = 5 \end{cases}$ can be written as

$$\begin{bmatrix} 1 & 0 & -2 \\ 2 & 5 & -4 \end{bmatrix} \begin{bmatrix} \mathrm{x} \\ y \\ z \end{bmatrix} = \begin{bmatrix} -3 \\ 5 \end{bmatrix}.$$

The augmented matrix is $\begin{bmatrix} 1 & 0 & -2 & -3 \\ 2 & 5 & -4 & 5 \end{bmatrix}$. ■

A row multiplication operation with the augmented matrix corresponds to multiplication of an equation in the system by a nonzero number. A row addition operation corresponds to adding a multiple of an equation to another equation. Interchange of two rows corresponds to interchange of two equations. Interchange of two columns corresponds to interchange of two variables.

These operations with equations take the system to an equivalent system (with the same set of variables and the same solution set). Therefore, elementary row operations, applied to the augmented matrix to bring it to a “simplified” form, can be used to solve an arbitrary system of linear equations. For example, if the coefficient matrix is diagonal (see Exercise 38 of the previous section), then the system splits into independent linear equations in one variable each and hence can be solved easily.

The other column elementary operations correspond to changing of variables. They are not needed to solve systems of linear equations where the coefficients are numbers (rather than functions, differential operators, etc.) but could be useful for this purpose and for other purposes. If they are used for solving a system of linear equation, additional computations should be done to keep track of changes in variables, so in the end we can write the final answer in terms of the original variables. To keep track of those changes, we usually augment the augmented matrix $[A|b]$ by the identity matrix 1_n and start with the matrix

$$\begin{bmatrix} A & b \\ 1_n & \end{bmatrix},$$

where n is the number of variables. By row and column addition operations the coefficient matrix can be taken to diagonal form. Then, using multiplication operations, it is easy to finish solving our system in new variables. Then the matrix in the place of the additional 1_n is used to write a final answer in terms of original variables.

Since changes of variables may result in additional computations, many textbooks avoid them and work only with row operations. By row operations, every coefficient matrix can be brought to the so-called echelon form and then to a reduced echelon form. Then it is not so difficult to finish solving the system. An echelon form is an upper triangular matrix (see Exercise 39 of the previous section) with special properties.

In this section we explain how to use row elementary operations and column interchange operations to make any matrix diagonal. In this way, we keep the same set of variables, and all linear systems on the way from the original one to the final answer are equivalent in the sense that they have the same solution set.

A well-known Gauss-Jordan elimination first uses forward substitutions ("going down" row addition operations) to get an upper diagonal matrix U (sometimes column permutations are needed, and all zeros on the main diagonal of U are in the end), and then it uses backward substitutions ("going up" row addition operations) to get a diagonal matrix (zero rows in the coefficient matrix may give some complications preventing us from obtaining a diagonal matrix). So this method uses column permutations but no column addition or multiplication operations. An interpretation of this method is that we write our coefficient matrix A of size $m \times n$ as $A = LUP$, where L is a lower triangular $m \times m$ matrix with ones along the main diagonal, P an $n \times n$ permutation matrix, and U an upper triangular $m \times n$ matrix.

Actually, use of column operations and the corresponding permutations of variables can be minimized as follows. By row addition operations we can always bring any coefficient matrix of any size $m \times n$ to an upper triangular form (the entries below the main diagonal are zeros). A way to do so is as follows. First we work on the first column and make, if possible, its first entry nonzero. Then we use this entry as a pivot entry to kill (eliminate, i.e., make zero) all other entries. Then we do the same with the submatrix obtained by ignoring the first row, and so on.

Now we consider separately the following four cases:

Case 1: The diagonal entries of this upper triangular matrix U are all nonzero and the matrix is square (that is, the number n of unknowns equals the number of equations m).

Case 2: The diagonal entries of U are all nonzero and $n > m$.

Case 3: The diagonal entries of U are all nonzero and $n < m$.

Case 4: A diagonal entry of U is zero.

In Case 1, by n row multiplication operations we can make all n diagonal entries equal to one. Then by a few [at most $n(n-1)/2$] "going up" row addition operations, we make the coefficient matrix to be 1_n. Thus, the augmented matrix becomes $[1_n|b']$. The system is now solved, and the only solution is $x = b'$. In fact, we just dealt with the case when the coefficient matrix A is invertible, and we described a way to find the answer $x = A^{-1}b = b'$.

In Case 2, we transform as before the submatrix of U consisting of the first m rows and columns to 1_m. The augmented matrix becomes $[1_m, c|b']$ and our final answer is $y = -cz + b'$ with arbitrary z, where y is the first m variables in x and z is the rest of variables.

In Case 3, the last $m-n$ rows of the coefficient matrix are zeros. If the same is true for the last $m-n$ rows of the augmented matrix, these rows say $0 = 0$, and they are redundant. Dropping them, we are reduced to Case 1, so we can find the unique solution by n row multiplication operations and a few row addition operations. On the other hand, if not all last $m-n$ entries in the last column of the augmented matrix are zeros, then we have a linear equation of the type $0 =$ nonzero number; hence our system has no solutions.

Finally, in Case 4, we start to use column permutation operations. To keep track of them, we write variables on top of the coefficient matrix, and when we permute columns we permute the corresponding labels on the top. By row addition and the column permutation operations, we bring the augmented matrix to the form

$$\begin{array}{c} \begin{array}{cc} y^T & z^T \end{array} \\ \begin{bmatrix} U & c & | \; b' \\ 0 & 0 & | \; b'' \end{bmatrix} \end{array},$$

where U is an upper triangular square matrix with nonzero diagonal entries, y is some variables in x labeling U, z is the rest of variables, and [0, 0] stands for a zero row or several zero rows in the new coefficient matrix. If $b'' \neq 0$ (i.e., an entry of b'' is nonzero), then we have an equation of the type $0 =$ nonzero number; hence our system has no solutions. Otherwise, we drop the zero rows of the augmented matrix and we are reduced to Case 2. ∎

Thus, a complete answer to solving a system $Ax = b$ of linear equations has one of the following standard forms:

- $0 = 1$ (i.e., the system has no solutions)
- $x = d$ (the system has exactly one solution)
- $z = Cy + d$ (the system has infinitely many solutions), where the column z consists of some variables in x, the column y consists of the rest of variables.

In the last case C is a constant $k \times l$ matrix, where k is the number of variables in z and l is the number of variables in y, and d is a constant column with k entries.

The number k is known as the rank of A. The number l is called the *dimension of the set of solutions.* The variables in y take arbitrary values, and those values determine the values of variables in z. In the second case, when there is only one solution, we say that the dimension of the solution set is 0.

It is important to understand that in all three cases the equation or system of equations in the answer is equivalent to the original system $Ax = b$ in the sense that these two systems have exactly the same set (or space) of solutions. It may happen that the system $Ax = b$ can be solved for different subsets of variables y, but the numbers k, l are the same for all correct answers.

Remark 6.2. Solving a system of linear equations usually means a complete description of all solutions (or showing that they do not exist). In contrast, solving an optimization problem usually means finding an optimal solution and the optimal value (or showing that they do not exist).

Remark 6.3. In some textbooks, the final answer for solving a system $Ax = b$ is given in the form $x = Ct + d$, where t is a column of l variables (parameters) distinct from all variables in x. In this case, instead of discussing what is the equivalence of two systems with different sets of variables, it is better to talk about a 1-1 correspondence between two solution spaces given in a certain explicit way. Transformations of the form $t \mapsto Ct + d$ are known as *affine transformations.* Thus, the solution space (when nonempty) can be identified with all l-tuples of numbers by an affine transformation. ∎

We described the process of solving systems of linear equations without appealing to the concepts of vector space and dimension.

Humans were solving those systems for at least 2000 years before these concepts and the word *algebra* appeared. By the way, the original meaning of this word was "reduction," and its main subject was solving systems of equations by elimination, reducing the number of variables.

However these concepts do give an additional insight. A linear combination of vectors (say, rows) $v_1, \ldots, v_l$ with coefficients $c_1, \ldots, c_l$ is $c_1 v_1 + \ldots + c_l v_l$. Vectors are called linearly dependent if one of them is a linear combination of others. The rank of a matrix is the maximal number of its linearly independent rows (or columns). A system of $Ax = b$ of linear equations has a solution if and only if the rank of the coefficient matrix A equals the rank of the augmented matrix $[A|b]$. (This is a very easy exercise.) If the ranks are the same, then the dimension of the solution set equals the number of columns in A minus the rank. ■

Now we give a few examples with solutions.

Problem 6.4. Solve the system of linear equations for x, y, z in Example 6.1.

Solution. We start with the augmented matrix

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

Adding the first row to the second row with coefficient -2 (i.e., replacing the second row by the second row minus 2 times the first row), we obtain an upper triangular matrix

$$\begin{bmatrix} 1 & 0 & -2 & -3 \\ 0 & 5 & 0 & 11 \end{bmatrix}$$

with nonzero diagonal entries 1, 5. So we are in Case 2. Multiplying the second row by 1/5, we obtain the final matrix

$$\begin{bmatrix} 1 & 0 & -2 & -3 \\ 0 & 1 & 0 & 11/5 \end{bmatrix}.$$

Now we can write our answer with names of variables : $x = -3 + 2z, y = 11/5 = 2.2, z$ arbitrary. We did not write the names of variables on the top of the coefficient matrix because we did not change them. However, in case we need to be reminded how the augmented matrix is related to the system of equations, we can decorate the matrix with additional information:

$$\begin{array}{cccccc} x & y & z & | & = \\ \end{array}$$
$$\begin{bmatrix} 1 & 0 & -2 & | & -3 \\ 0 & 1 & 0 & | & 11/5 \end{bmatrix}.$$

Problem 6.5. Solve the system $x + 2y = 1, 3x + 6y = 2$.

Solution. Here is the augmented matrix:

$$\begin{bmatrix} 1 & 2 & 1 \\ 3 & 6 & 2 \end{bmatrix}.$$

Adding the first row to the second row with coefficient -3, we obtain an upper triangular matrix

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

Looking at the second row, we conclude the system has no solutions. A shorter way to write down the answer is $0 = 1$.

Problem 6.6. Solve

$$\begin{cases} x + 2y = 1, \\ 3x + 6y = 3. \end{cases}$$

Solution. Doing the same row addition as in Problem 6.5, we obtain a zero row in the augmented matrix. Dropping this row, we obtain the final augmented matrix

$$[1\ 2\ |1].$$

Answer: $x = -2y + 1, y$ arbitrary.

Problem 6.7. Solve $ax = b$ for x, where a, b are given numbers.

Answer. If $a \neq 0$, then $x = b/a$. If $a = b = 0$, then x is arbitrary. If $a = 0 \neq b$, there are no solutions.

Problem 6.8. Solve the system

$$\begin{cases} x + 2y + 3z = 4, \\ z = 2. \end{cases}$$

Solution. If we write the augmented matrix for x, y, z, namely,

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 \end{bmatrix},$$

we find that we are in Case 4. Permuting y and z, we obtain the matrix

$$\begin{array}{ccc} x & z & y \end{array}$$
$$\left[\begin{array}{ccc|c} 1 & 3 & 2 & 4 \\ 0 & 1 & 0 & 2 \end{array}\right].$$

Adding the second row to the first one with coefficient -3, we obtain our final matrix:

$$\begin{array}{ccc} x & z & y \end{array}$$
$$\left[\begin{array}{ccc|c} 1 & 0 & 2 & -2 \\ 0 & 1 & 0 & 2 \end{array}\right].$$

Now we can write our final answer in the standard form

$$\begin{cases} x = -2y - 2, \\ z = 2 \end{cases}$$

where y is arbitrary. ■

Note that in this case passing to matrices did not result in saving time and room. The addition operation we used is the same as substitution of the second equation into the first one. However, for large systems with many nonzero coefficients, we save time and room when we use matrices to display information, by avoiding writing names of variables unnecessarily.

Problem 6.9. Solve for x_1, x_3, x_5:

$$\begin{cases} 2x_1 + 3x_2 + x_5 + 5 - 2x_2 = x_1 + x_4 + 1, \\ 2x_3 + x_5 + 2 = -3x_1 - 5x_2 + 3, \\ 3x_3 + 3x_2 - 1 = -x_5 - 6x_4 - 1. \end{cases}$$

Solution. The linear equations are not in standard form. We write them in standard form, and here is the augmented matrix:

$$\left[\begin{array}{ccccc|c} x_1 & x_3 & x_5 & x_2 & x_4 & = \\ 1 & 0 & 1 & 1 & -1 & -4 \\ 3 & 2 & 1 & 5 & 0 & 1 \\ 0 & 3 & 1 & 3 & 6 & 0 \end{array}\right].$$

By two downward addition operations, we make the matrix upper triangular:

$$\begin{array}{r} -3\downarrow \begin{array}{c}\leftarrow \\ \rightarrow\end{array} \\ \\ \end{array} \left[\begin{array}{ccccc|c} 1 & 0 & 1 & 1 & -1 & -4 \\ 3 & 2 & 1 & 5 & 0 & 1 \\ 0 & 3 & 1 & 3 & 6 & 0 \end{array}\right]$$

$$\Downarrow$$

$$\begin{array}{r} \\ -3/2\downarrow\leftarrow \\ \rightarrow \end{array}\begin{bmatrix} 1 & 0 & 1 & 1 & -1 & | & -4 \\ 0 & 2 & -2 & 2 & 3 & | & 13 \\ 0 & 3 & 1 & 3 & 6 & | & 0 \end{bmatrix}$$

$$\Downarrow$$

$$\begin{bmatrix} 1 & 0 & 1 & 1 & -1 & | & -4 \\ 0 & 2 & -2 & 2 & 3 & | & 13 \\ 0 & 0 & 4 & 0 & 1.5 & | & -39/2 \end{bmatrix}.$$

Now we do two row multiplication operations to make the diagonal entries = 1:

$$\begin{array}{r} \\ 1/2\cdot \\ 1/4\cdot \end{array}\begin{bmatrix} 1 & 0 & 1 & 1 & -1 & | & -4 \\ 0 & 2 & -2 & 2 & 3 & | & 13 \\ 0 & 0 & 4 & 0 & 1.5 & | & -39/2 \end{bmatrix}$$

$$\Downarrow$$

$$\begin{bmatrix} 1 & 0 & 1 & 1 & -1 & | & -4 \\ 0 & 1 & -1 & 1 & 3/2 & | & 13/2 \\ 0 & 0 & 1 & 0 & 3/8 & | & -39/8 \end{bmatrix}.$$

Now we do two upward row addition operations (back substitution):

$$\begin{array}{r} \longrightarrow\quad\longrightarrow \\ -1\uparrow\qquad\rightarrow \\ \longleftarrow\ 1\uparrow\leftarrow \end{array}\begin{bmatrix} 1 & 0 & 1 & 1 & -1 & | & -4 \\ 0 & 1 & -1 & 1 & 3/2 & | & 13/2 \\ 0 & 0 & 1 & 0 & 3/8 & | & -39/8 \end{bmatrix}$$

$$\Downarrow$$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & -11/8 & | & 7/8 \\ 0 & 1 & 0 & 1 & 15/8 & | & 13/8 \\ 0 & 0 & 1 & 0 & 3/8 & | & -39/8 \end{bmatrix}.$$

Now we can write the answer in standard form:

$$\begin{cases} x_1 = -x_2 + 11x_4/8 + 7/8 \\ x_3 = -x_2 - 15x_4/8 + 13/8 \\ x_5 = -3x_4/8 - 39/8. \end{cases}$$

Problem 6.10. Solve for x_1, x_2, x_3 the same system of linear equations (Problem 6.9).

Solution. We consider the last augmented matrix with columns permuted:

$$\begin{array}{ccccccc} x_1 & x_2 & x_3 & x_5 & x_4 & | & = \end{array}$$
$$\left[\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & -11/8 & 7/8 \\ 0 & 1 & 1 & 0 & 15/8 & 13/8 \\ 0 & 0 & 0 & 1 & 3/8 & -39/8 \end{array}\right].$$

We cannot write the answer in the form

$$\begin{cases} x_1 = & \text{an affine function of } x_4, x_5 \\ x_2 = & \text{an affine function of } x_4, x_5 \\ x_3 = & \text{an affine function of } x_4, x_5. \end{cases}$$

So it is not clear what the problem exactly means. However, if it means that we have to solve the system for unknown x_1, x_2, x_3 with given numbers x_4, x_5, then here is a way to solve the problem.

If $x_5 + 3x_4/8 \neq -39/8$, then there are no solutions. Otherwise, we drop the last row in the matrix, switch the x_2-column with the x_3-column, and get the final matrix

$$\begin{array}{ccccccc} x_1 & x_3 & x_2 & x_5 & x_4 & | & = \end{array}$$
$$\left[\begin{array}{ccccc|c} 1 & 0 & 1 & 0 & -11/8 & 7/8 \\ 0 & 1 & 1 & 0 & 15/8 & 13/8 \end{array}\right].$$

Therefore, in the case $x_5 + 3x_4/8 = -39/8$, our answer in standard form is

$$\begin{cases} x_1 = -x_2 + 11x_4/8 + 7/8 \\ x_3 = -x_2 - 15x_4/8 + 13/8 \end{cases}$$

with an arbitrary x_2. ■

We call two systems of equations equivalent if they have the same solutions. When we perform row elementary operations with an augmented matrix, we change the system of linear equations to an equivalent system. Is the converse true? See Exercises 40 and 41 at the end of this section. How about equivalence of systems involving different sets of variables? We leave this tricky question to the reader as a brain teaser. We do not need to address this issue if we do not change variables and do not consider elimination of variables as passing to an independent system with a smaller number of variables.

In §4 we stated the following fact:

Theorem 6.11. Given a system of linear equations $Ax = b$ and another linear equation $f_0 = b_0$ in standard form that follows from the system, then either the equation $f_0 = b_0$ or the equation $0 = 1$ is a linear combination of the equations in the system.

Proof. Note that if we do an elementary row operation, then all equations in the new system are linear combinations of old ones. In fact, only one of them is new, and it is a linear combination of an old equation in the case of multiplicative operation, and a linear combination of two old equations in the case of addition operation. Since linear combinations of linear combinations are linear combinations, after any number of elementary row operations, every new equation is a linear combination of the original equations.

If the system has no solutions, then we saw that we obtain the equation of the form $0 = c$ with $c \neq 0$, and a multiplication operation makes it $0 = 1$.

In the case when the system has exactly one solution $x = d$, the Gauss-Jordan method gives every equation $x_i = d_i$ as a linear combination of the original equations. Since $x = d$ satisfies the linear equation $f_0 = cx = b_0$, we have $cd = b_0$. Now we combine the equations $x = d$ with the coefficients c_i and obtain $cx = cd = b_0$.

Finally, assume that the system has infinitely many solutions. Then the Gauss-Jordan method gives the system of the form $z - Cy = d$ (see the standard answer on page 57). Substitution of this into $f_0 = cx = b_0$ gives a relation among C, d, c, b. Namely, writing $f_0 = cx = c'y + c''z$, the relation is $c'(Cz + d) + c''z = 0$ for all z; hence $c'C + c'' = 0$ and $c'd = 0$. Now taking the linear combination $c''(z - Cy) = cd$ of the equations $z - Cy = d$, we obtain the equation $f_0 = b_0$.

Remark 6.12. Can we solve systems of linear inequalities eliminating one variable after another like we did for systems of linear equations? Fourier asked this question and answered positively about 200 years ago. See Section A10 in the Appendix for the Fourier-Motzkin method. However, no one was able to make the method practical enough to compete with other methods. The problem with the method is that the number of constraints may grow very fast. Here is an example of when the elimination method works well.

Problem 6.13. Find a feasible solution for the system

$$\begin{cases} x_1 + x_2 + x_3 + x_4 \le 3, \\ x_1 + 2x_2 + 3x_3 \ge 1, \\ \text{all } x_i \ge 0. \end{cases} \tag{6.14}$$

Solution. We write down all two constraints involving x_4 in the form

$$0 \le x_4 \le 3 - x_1 - x_2 - x_3 \tag{i}$$

and eliminate x_4 from the system. So we obtain the following system

$$\begin{cases} 0 \le 3 - x_1 - x_2 - x_3 \\ x_1 + 2x_2 + 3x_3 \ge 1, \\ \text{all } x_i \ge 0 \end{cases}$$

for three variables. Now we write down all three constraints involving x_3 in the form

$$0, (1 - x_1 - 2x_2)/3 \le x_3 \le 3 - x_1 - x_2 \tag{ii}$$

and eliminate x_3 from the system. So we obtain the system

$$\begin{cases} 0, (1 - x_1 - 2x_2)/3 \le 3 - x_1 - x_2 \\ x_1, x_2 \ge 0 \end{cases}$$

for two variables. Finally, to finish our forward substitution, we write down all three constraints involving x_2 in the form

$$0 \le x_2 \le 3 - x_1, 8 - 2x_1 \tag{iii}$$

and eliminate x_2 from the system. So we obtain the system

$$0 \le 3 - x_1, 8 - 2x_1;\ \ x_1 \ge 0$$

for one variable. The last system is equivalent to $0 \le x_1 \le 3$. Taking any feasible value for x_1, we find one after another feasible values for x_2, x_3, x_4. By this back substitution we can find all feasible solutions to the original system (6.14).

For example, we start with $x_1 = 2$. The constraints (iii) for x_2 become $0 \le x_2 \le 1$. Let us pick $x_2 = 0$. The constraints (ii) are now $0 \le x_3 \le 1$. We pick $x_3 = 0$. The constraints (i) are now $0 \le x_4 \le 1$.

Let us pick $x_4 = 1$, and we are done. A feasible solution for (6.14) is found.

Remark 6.15. Most linear systems in real life have rational data. Therefore the answer involves only rational numbers. However, the rational numbers in the answer could have very large numerators and denominators, so their computation may take too much time. Usually systems are solved on a computer with limited precision. Approximate solutions are usually sufficient for practical applications. Uncertainty in data gives a good excuse for not finding exact solutions.

Exercises

1–8. Find whether the matrix A is invertible, reducing it to a diagonal matrix by row and column addition operations. *Remark*: The diagonal matrix in the answer is not unique, but the product of the diagonal entries in it is unique. In many textbooks on linear algebra it is proved that this product equals the *determinant* of A.

1. $A = \begin{bmatrix} 1 & 2 \\ 6 & 8 \end{bmatrix}$

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

3. $A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 1 \\ 0 & 0 & c \end{bmatrix}$

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

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

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

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

8. $A = \begin{bmatrix} 1 & -2 & -1 & 0 \\ 5 & 1 & 3 & 1 \\ 0 & 1 & 0 & 1 \\ 5 & 6 & 0 & 7 \end{bmatrix}$

9–16. Solve for x, y. *Hints*: Equations could be given not in the standard form for linear equations. Treat b, t, u, z as given numbers.

9. $\begin{cases} x + 2y = 1 \\ 2x + 4y = 3 \end{cases}$

10. $\begin{cases} x + 2y = 3 \\ 2x + 4y = 6 \end{cases}$

11. $\begin{cases} 2x + 3y + 5z = 2, \\ 3x + 5y + 8z = b \end{cases}$

12. $\begin{cases} 2x + 3y + 1 = 2 + x - y, \\ 3x + 5y + 8 = 2x \end{cases}$

13. $\begin{cases} x + 2y = 3 + u \\ 2x + 4y = t \end{cases}$

14. $\begin{cases} x + ty = 3 \\ 2x + 4y = 1 \end{cases}$

15. $\begin{cases} x + t^2 y = 1 \\ x + y = t \end{cases}$ **16.** $\begin{cases} x + ty = t^2 \\ 2x + uy = 1 \end{cases}$

17. Solve the system in Exercise 11 for y and z.

18. Solve the system in Exercise 11 for x and z.

19. Is there a system of linear equations with exactly three solutions?

20–23. Find A^{-1} (if it exists) for the matrix A in Exercises 5–8.

24–27. Write the matrix A in Exercises 1–4 as $A = LU$ with an upper triangular matrix U and a lower triangular matrix L, if possible.

28–31. Write the matrix A in Exercises 5–8 as $A = LU$ with an upper triangular matrix U and a lower triangular matrix L, if possible. Then compute the matrix UL.

32–35. Solve each system for x, y, z:

32. $\begin{cases} x + y + z = a + b^2, \\ x + 2y + 3z = c^3, \\ x + 3y + 4z = d. \end{cases}$ **33.** $\begin{cases} x + 5y + z = y, \\ x + 2y + 3 = z, \\ x + 3y - 4z = d. \end{cases}$

34. $\begin{cases} x + y + z = u_1, \\ x + 2y + 3z = u_2, \\ x + 3y + 4z = u_3. \end{cases}$ **35.** $\begin{cases} u + y + z = x, \\ x + 2u + 3z = y, \\ x + 3y + 4z = v. \end{cases}$

36–39. Find a feasible solution for the system (or prove that the system is infeasible).

36. $x,\ y \geq 0,\ x + y \leq 2,\ z = 3x - 4y + 5.$

37. $x,\ y,\ z \geq 0,\ x + y + z \leq 2,\ x + 2y + 3z \geq 3.$

38. $x,\ y,\ z \geq 0,\ x + y - z \leq 2,\ x + 2y + 3z \geq 3.$

39. $x - 2y + 3z \geq 5,\ 3x + 4y - z \leq 8,\ x - 2y \geq -3.$

40. Given two systems of linear equations $Ax = b$ and $A'x = b'$ with the same column of n distinct variables x, the same nonempty solution set, and the same number m of equations, show that we can obtain the augmented matrix $[A'|b']$ from $[A|b]$ by row addition operations and a row multiplication operation. In particular, there is an invertible $m \times m$ matrix C such that $CA = A', Cb = b'$.

41. Given two systems of linear equations $Ax = b$ and $A'x = b'$ with the same column of n distinct variables x, the same nonempty solution set, and different numbers $m > m'$ of equations, show that we can obtain the matrix $\begin{bmatrix} A' & b' \\ 0 & 0 \end{bmatrix}$ of size $m \times n$, with last $m - m'$ rows being zero rows, from the matrix $[A|b]$ by row addition operations. In particular, there is an $m \times m'$ matrix C such that $CA = A', Cb = b'$. Show also that there is an $m' \times m$ matrix C' such that $C'A' = A, C'b' = b$.