FACULTEIT BIO-INGENIEURSWETENSCHAPPEN

LINEAIRE ALGEBRA

Prof. dr. Willem Waegeman

Bachelor of Science in de biowetenschappen
Academiejaar 2019 – 2020

UNIVERSITEIT GENT

Linear Algebra: Theory and Applications

by
Willem Waegeman, An Schelfaut and Demir Ali Köse
Department of Data Analysis and Mathematical Modelling
Ghent University

Version DRAFT (April 17, 2019)

Edition
Version DRAFT (April 17, 2019).
April 17, 2019.

Publisher
Willem Waegeman, An Schelfaut and Demir Ali Köse
Department of Data Analysis and Mathematical Modelling
Ghent University
Coupure links 653
9000 Ghent
Belgium

Contents

Chapter 1
Systems of Linear Equations

We will motivate our study of linear algebra by studying solutions to systems of linear equations. While the focus of this chapter is on the practical matter of how to find, and describe, these solutions, we will also be setting ourselves up for more theoretical ideas that will appear later.

Section 1.1 What is Linear Algebra?

The subject of linear algebra can be partially explained by the meaning of the two terms comprising the title. “Linear” is a term you will appreciate better at the end of this course, and indeed, attaining this appreciation could be taken as one of the primary goals of this course. However for now, you can understand it to mean anything that is “straight” or “flat.” For example in the xy-plane you might be accustomed to describing straight lines (is there any other kind?) as the set of solutions to an equation of the form $y = mx + b$, where the slope m and the y-intercept b are constants that together describe the line. In multivariate calculus, you may have discussed planes. Living in three dimensions, with coordinates described by triples (x, y, z), they can be described as the set of solutions to equations of the form $ax + by + cz = d$, where a, b, c, d are constants that together determine the plane. While we might describe planes as “flat,” lines in three dimensions might be described as “straight.” From a multivariate calculus course you will recall that lines are sets of points described by equations such as $x = 3t - 4$, $y = -7t + 2$, $z = 9t$, where t is a parameter that can take on any value.

Another view of this notion of “flatness” is to recognize that the sets of points just described are solutions to equations of a relatively simple form. These equations involve addition and multiplication only. We will have a need for subtraction, and occasionally we will divide, but mostly you can describe “linear” equations as involving only addition and multiplication. Here are some examples of typical equations we will see in the next few sections:

$$2x + 3y - 4z = 13 \qquad 4x_1 + 5x_2 - x_3 + x_4 + x_5 = 0 \qquad 9a - 2b + 7c + 2d = -7$$

What we will not see are equations like:

$$xy + 5yz = 13 \qquad x_1 + x_2^3/x_4 - x_3x_4x_5^2 = 0 \qquad \tan(ab) + \log(c - d) = -7$$

The exception will be that we will on occasion need to take a square root.

You have probably heard the word “algebra” frequently in your mathematical preparation for this course. Most likely, you have spent a good ten to fifteen years learning the algebra of the real numbers, along with some introduction to the very similar algebra of complex numbers (see Section 7.1). However, there are many new algebras to learn and use, and likely linear algebra will be your second algebra. Like learning a second language, the necessary adjustments can be challenging at times, but the rewards are many. And it will make learning your third and fourth algebras even easier. In any event, prepare yourself to learn a new algebra and realize that some of the old rules you used for the real numbers may no longer apply to this *new* algebra you will be learning!

The brief discussion above about lines and planes suggests that linear algebra has an inherently geometric nature, and this is true. Examples in two and three dimensions can be used to provide valuable insight into important concepts of this course. However, much of the power of linear algebra will be the ability to work with "flat" or "straight" objects in higher dimensions, without concerning ourselves with visualizing the situation. While much of our intuition will come from examples in two and three dimensions, we will maintain an *algebraic* approach to the subject, with the geometry being secondary. Others may wish to switch this emphasis around, and that can lead to a very fruitful and beneficial course, but here and now we are laying our bias bare.

Section 1.2 Solving Systems of Linear Equations

Definition 1.1 System of Linear Equations

A **system of linear equations** is a collection of m equations in the variable quantities $x_1, x_2, x_3, \ldots, x_n$ of the form,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n &= b_3 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

where the values of a_{ij}, b_i and x_j are from the set of real numbers, $\mathbb{R}$.

Definition 1.2 Solution of a System of Linear Equations

A **solution** of a system of linear equations in n variables, x_1, x_2, x_3, …, x_n (such as the system given in Definition 1.1, is an ordered list of n real numbers, s_1, s_2, s_3, …, s_n such that if we substitute s_1 for x_1, s_2 for x_2, s_3 for x_3, …, s_n for x_n, then for every equation of the system the left side will equal the right side, i.e. each equation is true simultaneously.

More typically, we will write a solution in a form like $x_1 = 12$, $x_2 = -7$, $x_3 = 2$ to mean that $s_1 = 12$, $s_2 = -7$, $s_3 = 2$ in the notation of Definition 1.2. To discuss *all* of the possible solutions to a system of linear equations, we now define the set of all solutions.

Definition 1.3 Solution Set of a System of Linear Equations

The **solution set** of a linear system of equations is the set which contains every solution to the system, and nothing more.

Be aware that a solution set can be infinite, or there can be no solutions, in which case we write the solution set as the empty set, $\varnothing = \{\}$. Here is an example to illustrate using the notation introduced in Definition 1.1 and the notion of a solution (Definition 1.2).

Example 1.1
Given the system of linear equations,

$$\begin{aligned} x_1 + 2x_2 + x_4 &= 7 \\ x_1 + x_2 + x_3 - x_4 &= 3 \\ 3x_1 + x_2 + 5x_3 - 7x_4 &= 1 \end{aligned}$$

we have $n = 4$ variables and $m = 3$ equations. Also,

$$\begin{array}{lllll} a_{11} = 1 & a_{12} = 2 & a_{13} = 0 & a_{14} = 1 & b_1 = 7 \\ a_{21} = 1 & a_{22} = 1 & a_{23} = 1 & a_{24} = -1 & b_2 = 3 \\ a_{31} = 3 & a_{32} = 1 & a_{33} = 5 & a_{34} = -7 & b_3 = 1 \end{array}$$

Additionally, convince yourself that $x_1 = -2$, $x_2 = 4$, $x_3 = 2$, $x_4 = 1$ is one solution (Definition 1.2), but it is not the only one! For example, another solution is $x_1 = -12$, $x_2 = 11$, $x_3 = 1$, $x_4 = -3$, and there are more to be found. So the solution set contains at least two elements. ⊠

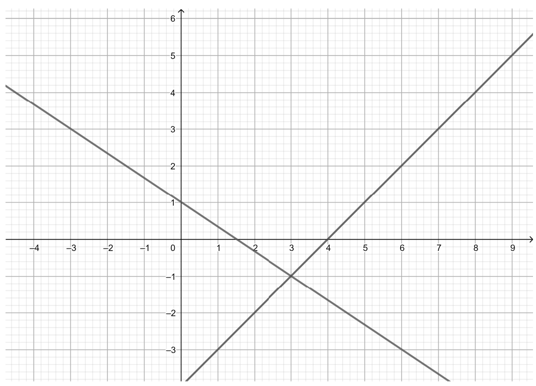
Possibilities for Solution Sets

The next example illustrates the possibilities for the solution set of a system of linear equations. We will not be too formal here, and the necessary theorems to back up our claims will come in subsequent sections. So read for feeling and come back later to revisit this example.

Example 1.2
Consider the system of two equations with two variables,

$$\begin{aligned} 2x_1 + 3x_2 &= 3 \\ x_1 - x_2 &= 4 \end{aligned}$$

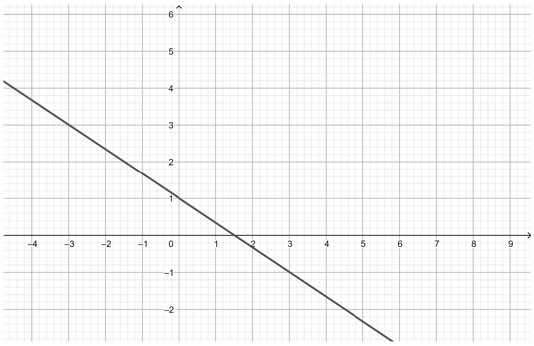
If we plot the solutions to each of these equations separately on the x_1x_2-plane, we get two lines, one with negative slope, the other with positive slope. They have exactly one point in common, $(x_1, x_2) = (3, -1)$, which is the solution $x_1 = 3$, $x_2 = -1$. From the geometry, we believe that this is the only solution to the system of equations, and so we say it is unique.



Now adjust the system with a different second equation,

$$\begin{aligned} 2x_1 + 3x_2 &= 3 \\ 4x_1 + 6x_2 &= 6 \end{aligned}$$

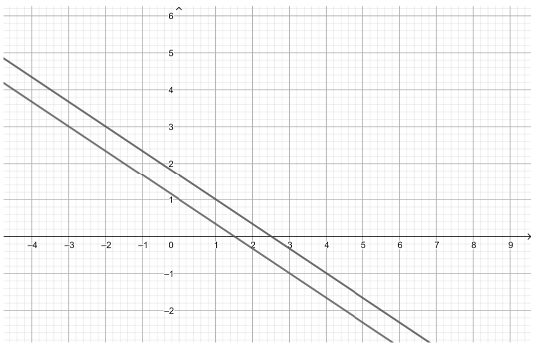
A plot of the solutions to these equations individually results in two lines, one on top of the other! There are infinitely many pairs of points that make both equations true. We will learn shortly how to describe this infinite solution set precisely (see Example 1.13). Notice now how the second equation is just a multiple of the first.



One more minor adjustment provides a third system of linear equations,

$$\begin{aligned} 2x_1 + 3x_2 &= 3 \\ 4x_1 + 6x_2 &= 10 \end{aligned}$$

A plot now reveals two lines with identical slopes, i.e. parallel lines. They have no points in common, and so the system has a solution set that is empty, $S = \varnothing$.



⊠

Equivalent Systems and Equation Operations

Definition 1.4 Equivalent Systems

Two systems of linear equations are **equivalent** if their solution sets are equal.

With this definition, we can begin to describe our strategy for solving linear systems. Given a system of linear equations that looks difficult to solve, we would like to have an *equivalent* system that is easy to solve. Since the systems will have equal solution sets, we can solve the "easy" system and get the solution set to the "difficult" system. Here come the tools for making this strategy viable.

Theorem 1.1 Equation Operations Preserve Solution Sets

If we apply one of the following three equation operations to a system of linear equations, then the original system and the transformed system are equivalent.

1. Swap the locations of two equations in the list of equations.
2. Multiply each term of an equation by a nonzero quantity α.
3. Multiply each term of one equation by some quantity α, and add these terms to a second equation, on both sides of the equality. Leave the first equation the same after this operation, but replace the second equation by the new one.

Proof Proving this theorem in a formal way is a bit technical, but probably the best way to understand the theorem is with an illustrative example. ■

Example 1.3
We solve the following system by a sequence of equation operations.

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ x_1 + 3x_2 + 3x_3 &= 5 \\ 2x_1 + 6x_2 + 5x_3 &= 6 \end{aligned}$$

$\alpha = -1$ times equation 1, add to equation 2:

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ 0x_1 + 1x_2 + 1x_3 &= 1 \\ 2x_1 + 6x_2 + 5x_3 &= 6 \end{aligned}$$

$\alpha = -2$ times equation 1, add to equation 3:

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ 0x_1 + 1x_2 + 1x_3 &= 1 \\ 0x_1 + 2x_2 + 1x_3 &= -2 \end{aligned}$$

$\alpha = -2$ times equation 2, add to equation 3:

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ 0x_1 + 1x_2 + 1x_3 &= 1 \\ 0x_1 + 0x_2 - 1x_3 &= -4 \end{aligned}$$

$\alpha = -1$ times equation 3:

$$x_1 + 2x_2 + 2x_3 = 4$$

$$\begin{aligned} 0x_1 + 1x_2 + 1x_3 &= 1 \\ 0x_1 + 0x_2 + 1x_3 &= 4 \end{aligned}$$

which can be written more clearly as

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ x_2 + x_3 &= 1 \\ x_3 &= 4 \end{aligned}$$

This is now a very easy system of equations to solve. The third equation requires that $x_3 = 4$ to be true. Making this substitution into equation 2 we arrive at $x_2 = -3$, and finally, substituting these values of x_2 and x_3 into the first equation, we find that $x_1 = 2$. Note too that this is the only solution to this final system of equations, since we were forced to choose these values to make the equations true. Since we performed equation operations on each system to obtain the next one in the list, all of the systems listed here are all equivalent to each other by Theorem 1.1. Thus $(x_1,\, x_2,\, x_3) = (2, -3, 4)$ is the unique solution to the *original* system of equations (and all of the other intermediate systems of equations listed as we transformed one into another). We note that $S = \{2, -3, 4\}$. ⊠

Example 1.4
The following system of equations made an appearance earlier in this section (Example 1.1), where we listed *one* of its solutions. Now, we will try to find all of the solutions to this system. Don't concern yourself too much about why we choose this particular sequence of equation operations, just believe that the work we do is all correct.

$$\begin{aligned} x_1 + 2x_2 + 0x_3 + x_4 &= 7 \\ x_1 + x_2 + x_3 - x_4 &= 3 \\ 3x_1 + x_2 + 5x_3 - 7x_4 &= 1 \end{aligned}$$

$\alpha = -1$ times equation 1, add to equation 2:

$$\begin{aligned} x_1 + 2x_2 + 0x_3 + x_4 &= 7 \\ 0x_1 - x_2 + x_3 - 2x_4 &= -4 \\ 3x_1 + x_2 + 5x_3 - 7x_4 &= 1 \end{aligned}$$

$\alpha = -3$ times equation 1, add to equation 3:

$$\begin{aligned} x_1 + 2x_2 + 0x_3 + x_4 &= 7 \\ 0x_1 - x_2 + x_3 - 2x_4 &= -4 \\ 0x_1 - 5x_2 + 5x_3 - 10x_4 &= -20 \end{aligned}$$

$\alpha = -5$ times equation 2, add to equation 3:

$$\begin{aligned} x_1 + 2x_2 + 0x_3 + x_4 &= 7 \\ 0x_1 - x_2 + x_3 - 2x_4 &= -4 \\ 0x_1 + 0x_2 + 0x_3 + 0x_4 &= 0 \end{aligned}$$

$\alpha = -1$ times equation 2:

$$\begin{aligned} x_1 + 2x_2 + 0x_3 + x_4 &= 7 \\ 0x_1 + x_2 - x_3 + 2x_4 &= 4 \\ 0x_1 + 0x_2 + 0x_3 + 0x_4 &= 0 \end{aligned}$$

$\alpha = -2$ times equation 2, add to equation 1:

$$\begin{aligned} x_1 + 0x_2 + 2x_3 - 3x_4 &= -1 \\ 0x_1 + x_2 - x_3 + 2x_4 &= 4 \\ 0x_1 + 0x_2 + 0x_3 + 0x_4 &= 0 \end{aligned}$$

which can be written more clearly as

$$\begin{aligned} x_1 + 2x_3 - 3x_4 &= -1 \\ x_2 - x_3 + 2x_4 &= 4 \\ 0 &= 0 \end{aligned}$$

What does the equation $0 = 0$ mean? We can choose *any* values for x_1, x_2, x_3, x_4 and this equation will be true, so we only need to consider further the first two equations, since the third is true no matter what. We can analyze the second equation without consideration of the variable x_1. It would appear that there is considerable latitude in how we can choose x_2, x_3, x_4 and make this equation true. Let's choose x_3 and x_4 to be *anything* we please, say $x_3 = a$ and $x_4 = b$.

Now we can take these arbitrary values for x_3 and x_4, substitute them in equation 1, to obtain

$$\begin{aligned} x_1 + 2a - 3b &= -1 \\ \Leftrightarrow \quad x_1 &= -1 - 2a + 3b \end{aligned}$$

Similarly, equation 2 becomes

$$\begin{aligned} x_2 - a + 2b &= 4 \\ \Leftrightarrow \quad x_2 &= 4 + a - 2b \end{aligned}$$

So our arbitrary choices of values for x_3 and x_4 (a and b) translate into specific values of x_1 and x_2. The lone solution given in Example 1.1 was obtained by choosing $a = 2$ and $b = 1$. Now we can easily and quickly find many more (infinitely more). Suppose we choose $a = 5$ and $b = -2$, then we compute

$$\begin{aligned} x_1 &= -1 - 2(5) + 3(-2) = -17 \\ x_2 &= 4 + 5 - 2(-2) = 13 \end{aligned}$$

and you can verify that $(x_1, x_2, x_3, x_4) = (-17, 13, 5, -2)$ makes all three equations true. The entire solution set is written as

$$S = \{\,(-1 - 2a + 3b,\ 4 + a - 2b,\ a,\ b) \mid a \in \mathbb{R},\ b \in \mathbb{R}\}$$

It would be instructive to finish off your study of this example by taking the general form of the solutions given in this set and substituting them into each of the three equations and verify that they are true in each case. ⊠

Section 1.3 Reduced Row-Echelon Form

After solving a few systems of equations, you will recognize that it doesn't matter so much *what* we call our variables, as opposed to what numbers act as their coefficients. A system in the variables x_1, x_2, x_3 would behave the same if we changed the names of the variables to a, b, c and kept all the constants the same and in the same places. In this section, we will isolate the key bits of information about a system of equations into something called a matrix, and then use this matrix to systematically solve the equations. Along the way we will obtain one of our most important and useful computational tools.

The Augmented Matrix

An $m \times n$ **matrix** is a rectangular layout of numbers from $\mathbb{R}$ having m rows and n columns. We will use upper-case Latin letters from the start of the alphabet (A, B, $C, \dots$) to denote matrices and squared-off brackets to delimit the layout. Many use large parentheses instead of brackets — the distinction is not important. Rows of a matrix will be referenced starting at the top and working down (i.e. row 1 is at the top) and columns will be referenced starting from the left (i.e. column 1 is at the left). For a matrix A, the notations $[A]_{ij}$ and a_{ij} will refer to the real number in row i and column j of A. Note that lower-case symbols are used for the entries in a matrix, while upper-case symbols are used for the matrix itself.

Example 1.5

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

is a matrix with $m = 3$ rows and $n = 4$ columns. We can say that $[B]_{23} = -6$ while $[B]_{34} = -2$. ⊠

A **column vector** of **size** m is an ordered list of m numbers, which is written in order vertically, starting at the top and proceeding to the bottom. At times, we will refer to a column vector as simply a **vector**. Column vectors will be written in bold and with an arrow on top of the symbol, usually with lower case Latin letter from the end of the alphabet such as $\vec{\boldsymbol{u}}$, $\vec{\boldsymbol{v}}$, $\vec{\boldsymbol{w}}$, $\vec{\boldsymbol{x}}$, $\vec{\boldsymbol{y}}$, $\vec{\boldsymbol{z}}$. To refer to the **entry** or **component** that is number i in the list that is the vector $\vec{\boldsymbol{v}}$ we write $[\vec{\boldsymbol{v}}]_i$ or v_i. A vector of size m can thus be written as

$$\vec{\boldsymbol{v}} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_m \end{bmatrix}$$

The **zero vector** of size m is the column vector of size m where each entry is the number zero,

$$\vec{\mathbf{0}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

For a system of linear equations,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \cdots + a_{3n}x_n &= b_3 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

the **coefficient matrix** is the $m \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & & & & \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

and the **vector of constants** is the column vector $\vec{\boldsymbol{b}}$ of size m:

$$\vec{\boldsymbol{b}} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}$$

If A is the coefficient matrix of a system of linear equations and $\vec{\boldsymbol{b}}$ is the vector of constants, then we will write $A\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}$ as a shorthand expression for the system of linear equations, which we will refer to as the **matrix representation** of the linear system. In fact we can see the system as a matrix-vector multiplication $A\vec{\boldsymbol{x}}$ that yields the vector $\vec{\boldsymbol{b}}$. A more formal treatment of matrix-vector multiplication is postponed till Definition 2.5.

Example 1.6
The system of linear equations

$$\begin{aligned} 2x_1 + 4x_2 - 3x_3 + 5x_4 + x_5 &= 9 \\ 3x_1 + x_2 + x_4 - 3x_5 &= 0 \\ -2x_1 + 7x_2 - 5x_3 + 2x_4 + 2x_5 &= -3 \end{aligned}$$

has coefficient matrix

$$A = \begin{bmatrix} 2 & 4 & -3 & 5 & 1 \\ 3 & 1 & 0 & 1 & -3 \\ -2 & 7 & -5 & 2 & 2 \end{bmatrix}$$

and vector of constants

$$\vec{\boldsymbol{b}} = \begin{bmatrix} 9 \\ 0 \\ -3 \end{bmatrix}$$

and so will be referenced as $A\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}$. ⊠

Suppose we have a system of m equations in n variables, with coefficient matrix A and vector of constants $\vec{\boldsymbol{b}}$. Then the **augmented matrix** of the system of equations is the $m \times (n+1)$ matrix whose first n columns are the columns of A and whose last column (number $n + 1$) is the column vector $\vec{\boldsymbol{b}}$. This matrix will be written as $[A \quad \vec{\boldsymbol{b}}]$ or shortly A_b.

Example 1.7
Let us consider the following system of 3 equations in 3 variables.

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 1 \\ 2x_1 + x_2 + x_3 &= 8 \\ x_1 + x_2 &= 5 \end{aligned}$$

Here is its augmented matrix:

$$[A \quad \vec{\boldsymbol{b}}] = \begin{bmatrix} 1 & -1 & 2 & 1 \\ 2 & 1 & 1 & 8 \\ 1 & 1 & 0 & 5 \end{bmatrix}$$

⊠

Row Operations

An augmented matrix for a system of equations will save us the tedium of continually writing down the names of the variables as we solve the system. It will also release us from any dependence on the actual names of the variables. We have seen how certain operations we can perform on equations will preserve their solutions (Theorem 1.1). The next two definitions and the following theorem carry over these ideas to augmented matrices.

Definition 1.5 Row Operations

The following three operations will transform an $m \times n$ matrix into a different matrix of the same size, and each is known as a **row operation**.

1. Swap the locations of two rows.
2. Multiply each entry of a single row by a nonzero quantity.
3. Multiply each entry of one row by some quantity, and add these values to the entries in the same columns of a second row. Leave the first row the same after this operation, but replace the second row by the new values.

We will use a symbolic shorthand to describe these row operations:

1. $R_i \leftrightarrow R_j$: Swap the location of rows i and j.
2. αR_i: Multiply row i by the nonzero scalar α.
3. $R_j + \alpha R_i$: Multiply row i by the scalar α and add to row j.

Definition 1.6 Row-Equivalent Matrices

Two matrices, A and B, are **row-equivalent** if one can be obtained from the other by a sequence of row operations.

Example 1.8
The matrices

$$A = \begin{bmatrix} 2 & -1 & 3 & 4 \\ 5 & 2 & -2 & 3 \\ 1 & 1 & 0 & 6 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 1 & 0 & 6 \\ 3 & 0 & -2 & -9 \\ 2 & -1 & 3 & 4 \end{bmatrix}$$

are row-equivalent as can be seen from

$$\begin{bmatrix} 2 & -1 & 3 & 4 \\ 5 & 2 & -2 & 3 \\ 1 & 1 & 0 & 6 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 1 & 0 & 6 \\ 5 & 2 & -2 & 3 \\ 2 & -1 & 3 & 4 \end{bmatrix} \xrightarrow{R_2 - 2R_1} \begin{bmatrix} 1 & 1 & 0 & 6 \\ 3 & 0 & -2 & -9 \\ 2 & -1 & 3 & 4 \end{bmatrix}$$

We can also say that any pair of these three matrices are row-equivalent. ⊠

Notice that each of the three row operations is reversible, so we do not have to be careful about the distinction between "A is row-equivalent to B" and "B is row-equivalent to A". The preceding definitions are designed to make the following theorem possible. It says that row-equivalent matrices represent systems of linear equations that have identical solution sets.

Theorem 1.2 Row-Equivalent Matrices represent Equivalent Systems

Suppose that A and B are row-equivalent augmented matrices. Then the systems of linear equations that they represent are equivalent systems.

Proof When two rows are swapped, the solution of the system does not change. If one multiplies a constant to a row, then the solution of the system also does not change. The third operation is a bit harder to prove, but examples should make it clear that also in this case the solution of the system does not change. ■

So at this point, our strategy is to begin with a system of equations, represent it by an augmented matrix, perform row operations (which will preserve solutions for the corresponding systems) to get a "simpler" augmented matrix, convert back to a "simpler" system of equations and then solve that system, knowing that its solutions are those of the original system. Here's a rehash of Example 1.3 as an exercise in using our new tools.

Example 1.9
We solve the following system using augmented matrices and row operations. This is the same system of equations solved in Example 1.3 using equation operations.

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ x_1 + 3x_2 + 3x_3 &= 5 \\ 2x_1 + 6x_2 + 5x_3 &= 6 \end{aligned}$$

Form the augmented matrix,

$$[A \quad \vec{\boldsymbol{b}}] = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 1 & 3 & 3 & 5 \\ 2 & 6 & 5 & 6 \end{bmatrix}$$

and apply row operations,

$$\xrightarrow{R_2 - 1R_1} \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 1 & 1 & 1 \\ 2 & 6 & 5 & 6 \end{bmatrix} \xrightarrow{R_3 - 2R_1} \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 2 & 1 & -2 \end{bmatrix}$$
$$\xrightarrow{R_3 - 2R_2} \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -1 & -4 \end{bmatrix} \xrightarrow{-1R_3} \begin{bmatrix} 1 & 2 & 2 & 4 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 4 \end{bmatrix}$$

So the matrix

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

is row-equivalent to A and by Theorem 1.2 the system of equations below has the same solution set as the original system of equations.

$$\begin{aligned} x_1 + 2x_2 + 2x_3 &= 4 \\ x_2 + x_3 &= 1 \\ x_3 &= 4 \end{aligned}$$

Solving this “simpler” system is straightforward and is identical to the process in Example 1.3. ⊠

Reduced Row-Echelon Form

The preceding example amply illustrates the definitions and theorems we have seen so far. But it still leaves two questions unanswered. Exactly what is this “simpler” form for a matrix, and just how do we get it? Here’s the answer to the first question, a definition of reduced row-echelon form.

Definition 1.7 Reduced Row-Echelon Form

A matrix is in **reduced row-echelon form** if it meets all of the following conditions:

1. If there is a row where every entry is zero, then this row lies below any other row that contains a nonzero entry.
2. The leftmost nonzero entry of a row is equal to 1.
3. The leftmost nonzero entry of a row is the only nonzero entry in its column.
4. Consider any two different leftmost nonzero entries, one located in row i, column j and the other located in row s, column t. If $s > i$, then $t > j$.

A row of only zero entries will be called a **zero row** and the leftmost nonzero entry of a nonzero row will be called a **leading 1**. The number of nonzero rows will be denoted by r.
A column containing a leading 1 will be called a **pivot column**. The set of column indices for all of the pivot columns will be denoted by $D = \{d_1, d_2, d_3, \ldots, d_r\}$ where $d_1 < d_2 < d_3 < \cdots < d_r$, while the columns that are not pivot columns will be denoted as $F = \{f_1, f_2, f_3, \ldots, f_{n-r}\}$ where $f_1 < f_2 < f_3 < \cdots < f_{n-r}$.

The principal feature of reduced row-echelon form is the pattern of leading 1's guaranteed by conditions (2) and (4), reminiscent of a flight of geese, or steps in a staircase, or water cascading down a mountain stream.

There are a number of new terms and notations introduced in this definition, which should make you suspect that this is an important definition. Given all there is to digest here, we will mostly save the use of D and F for the next subsection. However, one important point to make here is that all of these terms and notations apply to a matrix. Sometimes we will employ these terms and sets for an augmented matrix, and other times it might be a coefficient matrix. So always give some thought to exactly which type of matrix you are analyzing.

Example 1.10
The matrix C is in reduced row-echelon form.

$$C = \begin{bmatrix} 1 & -3 & 0 & 6 & 0 & 0 & -5 & 9 \\ 0 & 0 & 0 & 0 & 1 & 0 & 3 & -7 \\ 0 & 0 & 0 & 0 & 0 & 1 & 7 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This matrix has two zero rows and three leading 1's. So $r = 3$. Columns 1, 5, and 6 are pivot columns, so $D = \{1, 5, 6\}$ and then $F = \{2, 3, 4, 7, 8\}$. ⊠

Example 1.11
The matrix E is not in reduced row-echelon form, as it fails each of the four requirements of Definition

1.7 once.

$$E = \begin{bmatrix} 1 & 0 & -3 & 0 & 6 & 0 & 7 & -5 & 9 \\ 0 & 0 & 0 & 5 & 0 & 1 & 0 & 3 & -7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & -4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 7 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

⊠

For some concepts, it will be enough to form only zero entries under a leading 1 (actually it is enough to make the leading element nonzero). In that case, it's even clear which columns are pivot columns. The matrix we find is in **row-echelon form**.

Theorem 1.3 Row-Equivalent Matrix in Echelon Form

Suppose A is a matrix. Then there is a matrix B so that

1. A and B are row-equivalent.
2. B is in reduced row-echelon form.

Proof This theorem can be proven by formally describing the procedure that was outlined in the previous example. A formal description is, however, beyond the scope of this course. The procedure is known as Gauss-Jordan elimination. It will be intensively used in later chapters. ■

So now we can put it all together. Begin with a system of linear equations (Definition 1.1), and represent the system by its augmented matrix. Use row operations (Definition 1.5) to convert this matrix into reduced row-echelon form (Definition 1.7), using the procedure of Gauss-Jordan elimination. Theorem 1.3 also tells us we can always accomplish this, and that the result is row-equivalent (Definition 1.6) to the original augmented matrix. Since the matrix in reduced-row echelon form has the same solution set, we can analyze the row-reduced version instead of the original matrix, viewing it as the augmented matrix of a different system of equations. The beauty of augmented matrices in reduced row-echelon form is that the solution sets to their corresponding systems can be easily determined, as we will see in the next few examples and in the next section.

We will see through the course that almost every interesting property of a matrix can be discerned by looking at a row-equivalent matrix in reduced row-echelon form. For this reason it is important to know that the matrix B guaranteed to exist by Theorem 1.3 is also unique.

Theorem 1.4 Reduced Row-Echelon Form is Unique

Suppose that A is an $m \times n$ matrix and that B and C are $m \times n$ matrices that are row-equivalent to A and in reduced row-echelon form. Then $B = C$.

Proof A formal proof of this theorem is beyond the scope of this course. ■

We will now run through some examples of using these definitions and theorems to solve some systems of equations. From now on, when we have a matrix in reduced row-echelon form, we will mark the leading 1's with a small box. In your work, you can box 'em, circle 'em or write 'em in a different color — just identify 'em somehow. This device will prove very useful later and is a *very good habit* to start developing right now.

Example 1.12
Let's find the solutions to the following system of equations,

$$\begin{aligned} -7x_1 - 6x_2 - 12x_3 &= -33 \\ 5x_1 + 5x_2 + 7x_3 &= 24 \\ x_1 + 4x_3 &= 5 \end{aligned}$$

First, form the augmented matrix,

$$\begin{bmatrix} -7 & -6 & -12 & -33 \\ 5 & 5 & 7 & 24 \\ 1 & 0 & 4 & 5 \end{bmatrix}$$

and work to reduced row-echelon form. Let us start with creating zeros in the first column:

$$\xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 0 & 4 & 5 \\ 5 & 5 & 7 & 24 \\ -7 & -6 & -12 & -33 \end{bmatrix} \xrightarrow{R_2 - 5R_1} \begin{bmatrix} 1 & 0 & 4 & 5 \\ 0 & 5 & -13 & -1 \\ -7 & -6 & -12 & -33 \end{bmatrix}$$
$$\xrightarrow{R_3 + 7R_1} \begin{bmatrix} \boxed{1} & 0 & 4 & 5 \\ 0 & 5 & -13 & -1 \\ 0 & -6 & 16 & 2 \end{bmatrix}$$

Subsequently, we create zeros in the second column:

$$\xrightarrow{\frac{1}{5}R_2} \begin{bmatrix} \boxed{1} & 0 & 4 & 5 \\ 0 & 1 & \frac{-13}{5} & \frac{-1}{5} \\ 0 & -6 & 16 & 2 \end{bmatrix} \xrightarrow{R_3 + 6R_2} \begin{bmatrix} \boxed{1} & 0 & 4 & 5 \\ 0 & \boxed{1} & \frac{-13}{5} & \frac{-1}{5} \\ 0 & 0 & \frac{2}{5} & \frac{4}{5} \end{bmatrix}$$

Finally we create zeros in the third column:

$$\xrightarrow{\frac{5}{2}R_3} \begin{bmatrix} \boxed{1} & 0 & 4 & 5 \\ 0 & \boxed{1} & \frac{-13}{5} & \frac{-1}{5} \\ 0 & 0 & 1 & 2 \end{bmatrix} \xrightarrow{R_2 + \frac{13}{5}R_3} \begin{bmatrix} \boxed{1} & 0 & 4 & 5 \\ 0 & \boxed{1} & 0 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$
$$\xrightarrow{R_1 - 4R_3} \begin{bmatrix} \boxed{1} & 0 & 0 & -3 \\ 0 & \boxed{1} & 0 & 5 \\ 0 & 0 & \boxed{1} & 2 \end{bmatrix}$$

This is now the augmented matrix of a very simple system of equations, namely $x_1 = -3$, $x_2 = 5$, $x_3 = 2$, which has an obvious solution. Furthermore, we can see that this is the *only* solution to this system, so we have determined the entire solution set,

$$S = \left\{ \begin{bmatrix} -3 \\ 5 \\ 2 \end{bmatrix} \right\}$$

You might compare this example with the procedure we used in Example 1.3. ⊠

Example 1.7 and Example 1.12 are meant to contrast each other in many respects. So let's solve Example 1.7 now.

Example 1.13
Let's find the solutions to the following system of equations,

$$\begin{aligned} x_1 - x_2 + 2x_3 &= 1 \\ 2x_1 + x_2 + x_3 &= 8 \\ x_1 + x_2 &= 5 \end{aligned}$$

First, form the augmented matrix,

$$\begin{bmatrix} 1 & -1 & 2 & 1 \\ 2 & 1 & 1 & 8 \\ 1 & 1 & 0 & 5 \end{bmatrix}$$

and work to reduced row-echelon form.

$$\xrightarrow{R_2-2R_1} \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 3 & -3 & 6 \\ 1 & 1 & 0 & 5 \end{bmatrix} \xrightarrow{R_3-1R_1} \begin{bmatrix} \boxed{1} & -1 & 2 & 1 \\ 0 & 3 & -3 & 6 \\ 0 & 2 & -2 & 4 \end{bmatrix}$$

$$\xrightarrow{\frac{1}{3}R_2} \begin{bmatrix} \boxed{1} & -1 & 2 & 1 \\ 0 & 1 & -1 & 2 \\ 0 & 2 & -2 & 4 \end{bmatrix} \xrightarrow{R_1+1R_2} \begin{bmatrix} \boxed{1} & 0 & 1 & 3 \\ 0 & 1 & -1 & 2 \\ 0 & 2 & -2 & 4 \end{bmatrix}$$

$$\xrightarrow{R_3-2R_2} \begin{bmatrix} \boxed{1} & 0 & 1 & 3 \\ 0 & \boxed{1} & -1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The system of equations represented by this augmented matrix needs to be considered a bit differently than that for Example 1.12. First, the last row of the matrix is the equation $0 = 0$, which is *always* true, so it imposes no restrictions on our possible solutions and therefore we can safely ignore it as we analyze the other two equations. These equations are,

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

While this system is fairly easy to solve, it also appears to have a multitude of solutions. For example, choose $x_3 = 1$ and see that then $x_1 = 2$ and $x_2 = 3$ will together form a solution. Or choose $x_3 = 0$, and then discover that $x_1 = 3$ and $x_2 = 2$ lead to a solution. Try it yourself: pick *any* value of x_3 you please, and figure out what x_1 and x_2 should be to make the first and second equations (respectively) true. Because of this behavior, we say that x_3 is a "free" or "independent" variable. But why do we vary x_3 and not some other variable? For now, notice that the third column of the augmented matrix does not have any leading 1's in its column. With this idea, we can rearrange the two equations, solving each for the variable that corresponds to the leading 1 in that row.

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

To write the set of solution vectors in set notation, we have

$$S = \left\{ \begin{bmatrix} 3 - x_3 \\ 2 + x_3 \\ x_3 \end{bmatrix} \middle| \, x_3 \in \mathbb{R} \right\}$$

We'll learn more in the next section about systems with infinitely many solutions and how to express their solution sets. ⊠

Example 1.14
Let's find the solutions to the following system of equations,

$$\begin{aligned} 2x_1 + x_2 + 7x_3 - 7x_4 &= 2 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 &= 3 \\ x_1 + x_2 + 4x_3 - 5x_4 &= 2 \end{aligned}$$

First, form the augmented matrix,

$$\begin{bmatrix} 2 & 1 & 7 & -7 & 2 \\ -3 & 4 & -5 & -6 & 3 \\ 1 & 1 & 4 & -5 & 2 \end{bmatrix}$$

and work to reduced row-echelon form.

$$\xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 1 & 4 & -5 & 2 \\ -3 & 4 & -5 & -6 & 3 \\ 2 & 1 & 7 & -7 & 2 \end{bmatrix} \xrightarrow{R_2 + 3R_1} \begin{bmatrix} 1 & 1 & 4 & -5 & 2 \\ 0 & 7 & 7 & -21 & 9 \\ 2 & 1 & 7 & -7 & 2 \end{bmatrix}$$

$$\xrightarrow{R_3 - 2R_1} \begin{bmatrix} \boxed{1} & 1 & 4 & -5 & 2 \\ 0 & 7 & 7 & -21 & 9 \\ 0 & -1 & -1 & 3 & -2 \end{bmatrix}$$

$$\xrightarrow{R_2 \leftrightarrow R_3} \begin{bmatrix} \boxed{1} & 1 & 4 & -5 & 2 \\ 0 & -1 & -1 & 3 & -2 \\ 0 & 7 & 7 & -21 & 9 \end{bmatrix} \xrightarrow{-1R_2} \begin{bmatrix} \boxed{1} & 1 & 4 & -5 & 2 \\ 0 & 1 & 1 & -3 & 2 \\ 0 & 7 & 7 & -21 & 9 \end{bmatrix}$$

$$\xrightarrow{R_1 - 1R_2} \begin{bmatrix} \boxed{1} & 0 & 3 & -2 & 0 \\ 0 & 1 & 1 & -3 & 2 \\ 0 & 7 & 7 & -21 & 9 \end{bmatrix} \xrightarrow{R_3 - 7R_2} \begin{bmatrix} \boxed{1} & 0 & 3 & -2 & 0 \\ 0 & \boxed{1} & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 & -5 \end{bmatrix}$$

$$\xrightarrow{-\frac{1}{5}R_3} \begin{bmatrix} \boxed{1} & 0 & 3 & -2 & 0 \\ 0 & \boxed{1} & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{R_2 - 2R_3} \begin{bmatrix} \boxed{1} & 0 & 3 & -2 & 0 \\ 0 & \boxed{1} & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} \end{bmatrix}$$

Let's analyze the equations in the system represented by this augmented matrix. The third equation will read $0 = 1$. This is patently false, all the time. No choice of values for our variables will ever make it true. We're done. Since we cannot even make the last equation true, we have no hope of making all of the equations simultaneously true. So this system has no solutions, and its solution set is the empty set, $\varnothing = \{\ \}$.

Notice that we could have reached this conclusion sooner. After performing the row operation $R_3 - 7R_2$, we can see that the third equation reads $0 = -5$, a false statement. Since the system represented by this matrix has no solutions, none of the systems represented has any solutions. However, for this example, we have chosen to bring the matrix fully to reduced row-echelon form for the practice. ⊠

These three examples (Example 1.12, Example 1.13, Example 1.14) illustrate the full range of possibilities for a system of linear equations — no solutions, one solution, or infinitely many solutions. To **row-reduce** the matrix A means to apply row operations to A and arrive at a row-equivalent matrix B in reduced row-echelon form. So the term **row-reduce** is used as a verb. Theorem 1.3 tells us that this process will always be successful and Theorem 1.4 tells us that the result will be unambiguous. Typically, the analysis of A will proceed by analyzing B and applying theorems whose hypotheses include the row-equivalence of A and B.

Consistent Systems

We will now be more careful about analyzing the reduced row-echelon form derived from the augmented matrix of a system of linear equations. In particular, we will see how to systematically handle the situation when we have infinitely many solutions to a system, and we will prove that every system of linear equations has either zero, one or infinitely many solutions. With these tools, we will be able to solve any system by a well-described method.

Definition 1.8 Consistent System

A system of linear equations is **consistent** if it has at least one solution. Otherwise, the system is called **inconsistent**.

We will want to first recognize when a system is inconsistent or consistent, and in the case of consistent systems we will be able to further refine the types of solutions possible. We will do this by analyzing the reduced row-echelon form of a matrix, using the value of r, and the sets of column indices, D and F, first defined back in Definition 1.7. The number r is the single most important piece of information we can get from the reduced row-echelon form of a matrix. It is defined as the number of nonzero rows, but since each nonzero row has a leading 1, it is also the number of leading 1's present. For each leading 1, we have a pivot column, so r is also the number of pivot columns. Repeating ourselves, r is the number of nonzero rows, the number of leading 1's *and* the number of pivot columns. Across different situations, each of these interpretations of the meaning of r will be useful.

Before proving some theorems about the possibilities for solution sets to systems of equations, let's analyze one particular system with an infinite solution set very carefully as an example. We'll use this technique frequently, and shortly we'll refine it slightly.

Example 1.15
We consider a system of $m = 4$ equations in $n = 7$ variables.

$$\begin{aligned} x_1 + 4x_2 - x_4 + 7x_6 - 9x_7 &= 3 \\ 2x_1 + 8x_2 - x_3 + 3x_4 + 9x_5 - 13x_6 + 7x_7 &= 9 \\ 2x_3 - 3x_4 - 4x_5 + 12x_6 - 8x_7 &= 1 \\ -x_1 - 4x_2 + 2x_3 + 4x_4 + 8x_5 - 31x_6 + 37x_7 &= 4 \end{aligned}$$

This system has a 4×8 augmented matrix that is row-equivalent to the following matrix (check this!), and which is in reduced row-echelon form (the existence of this matrix is guaranteed by Theorem 1.3

and its uniqueness is guaranteed by Theorem 1.4),

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

So we find that $r = 3$ and

$$D = \{d_1,\, d_2,\, d_3\} = \{1,\, 3,\, 4\} \qquad\qquad F = \{f_1,\, f_2,\, f_3,\, f_4,\, f_5\} = \{2,\, 5,\, 6,\, 7,\, 8\}$$

Let i denote one of the $r = 3$ non-zero rows, and then we see that we can solve the corresponding equation represented by this row for the variable x_{d_i} and write it as a linear function of the variables $x_{f_1}, x_{f_2}, x_{f_3}, x_{f_4}$. Notice that $f_5 = 8$ does not reference a variable. We'll do this now, but you can already see how the subscripts upon subscripts takes some getting used to.

$$\begin{aligned} x_{d_1} &= x_1 = 4 - 4x_2 - 2x_5 - x_6 + 3x_7 \\ x_{d_2} &= x_3 = 2 - x_5 + 3x_6 - 5x_7 \\ x_{d_3} &= x_4 = 1 - 2x_5 + 6x_6 - 6x_7 \end{aligned}$$

Each element of the set $F = \{f_1,\, f_2,\, f_3,\, f_4,\, f_5\} = \{2,\, 5,\, 6,\, 7,\, 8\}$ is the index of a variable, except for $f_5 = 8$. We refer to $x_{f_1} = x_2$, $x_{f_2} = x_5$, $x_{f_3} = x_6$ and $x_{f_4} = x_7$ as "**free**" (or "**independent**") variables since they are allowed to assume any possible combination of values that we can imagine and we can continue on to build a solution to the system by solving individual equations for the values of the other ("**dependent**") variables.

Each element of the set $D = \{d_1,\, d_2,\, d_3\} = \{1,\, 3,\, 4\}$ is the index of a variable. We refer to the variables $x_{d_1} = x_1$, $x_{d_2} = x_3$ and $x_{d_3} = x_4$ as "dependent" variables since they *depend* on the *independent* variables. More precisely, for each possible choice of values for the independent variables we get *exactly one* set of values for the dependent variables that combine to form a solution of the system.

To express the solutions as a set, we write

$$S = \left\{ \left. \begin{bmatrix} 4 - 4x_2 - 2x_5 - x_6 + 3x_7 \\ x_2 \\ 2 - x_5 + 3x_6 - 5x_7 \\ 1 - 2x_5 + 6x_6 - 6x_7 \\ x_5 \\ x_6 \\ x_7 \end{bmatrix} \right|\, x_2,\, x_5,\, x_6,\, x_7 \in \mathbb{R} \right\}$$

The condition that $x_2, x_5, x_6, x_7 \in \mathbb{R}$ is how we specify that the variables x_2, x_5, x_6, x_7 are "free" to assume any possible values.

⊠

Using the reduced row-echelon form of the augmented matrix of a system of equations to determine the nature of the solution set of the system is a very key idea. So let's look at one more example like the last one. But first a definition, and then the example. We mix our metaphors a bit when we call variables free versus dependent.

Definition 1.9 Independent and Dependent Variables

Suppose A_b is the augmented matrix of a consistent system of linear equations and B is a row-equivalent matrix in reduced row-echelon form. Suppose j is the index of a column of B that contains the leading 1 for some row (i.e. column j is a pivot column). Then the variable x_j is **dependent**. A variable that is not dependent is called **independent** or **free**.

If you studied this definition carefully, you might wonder what to do if the system has n variables and column $n+1$ is a pivot column? We will see shortly, by Theorem 1.5, that this never happens for a consistent system.

Example 1.16
Consider the system of five equations in five variables,

$$\begin{aligned} x_1 - x_2 - 2x_3 + x_4 + 11x_5 &= 13 \\ x_1 - x_2 + x_3 + x_4 + 5x_5 &= 16 \\ 2x_1 - 2x_2 + x_4 + 10x_5 &= 21 \\ 2x_1 - 2x_2 - x_3 + 3x_4 + 20x_5 &= 38 \\ 2x_1 - 2x_2 + x_3 + x_4 + 8x_5 &= 22 \end{aligned}$$

whose augmented matrix row-reduces to

$$\begin{bmatrix} \boxed{1} & -1 & 0 & 0 & 3 & 6 \\ 0 & 0 & \boxed{1} & 0 & -2 & 1 \\ 0 & 0 & 0 & \boxed{1} & 4 & 9 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

There are leading 1's in columns 1, 3 and 4, so $D = \{1, 3, 4\}$. From this we know that the variables x_1, x_3 and x_4 will be dependent variables, and each of the $r = 3$ nonzero rows of the row-reduced matrix will yield an expression for one of these three variables. The set F is all the remaining column indices, $F = \{2, 5, 6\}$. That $6 \in F$ refers to the column originating from the vector of constants, but the remaining indices in F will correspond to free variables, so x_2 and x_5 (the remaining variables) are our free variables. The resulting three equations that describe our solution set are then,

$$\begin{aligned} x_1 &= 6 + x_2 - 3x_5 \\ x_3 &= 1 + 2x_5 \\ x_4 &= 9 - 4x_5 \end{aligned}$$

Make sure you understand where these three equations came from, and notice how the location of the leading 1's determined the variables on the left-hand side of each equation. We can compactly describe the solution set as,

$$S = \left\{ \left. \begin{bmatrix} 6 + x_2 - 3x_5 \\ x_2 \\ 1 + 2x_5 \\ 9 - 4x_5 \\ x_5 \end{bmatrix} \right| x_2,\, x_5 \in \mathbb{R} \right\}$$

Notice how we express the freedom for x_2 and x_5: $x_2,\, x_5 \in \mathbb{R}$. ⊠

We can now use the values of m, n, r, and the independent and dependent variables to categorize the solution sets for linear systems through a sequence of theorems. First we have an important theorem that explores the distinction between consistent and inconsistent linear systems.

Theorem 1.5 Recognizing Consistency of a Linear System

Suppose A_b is the augmented matrix of a system of linear equations with n variables. Then the system of equations is inconsistent if and only if the last column of A_b is a pivot column.

Proof A formal proof of this theorem is beyond the scope of this course. ■

The beauty of this theorem being an equivalence is that we can unequivocally test to see if a system is consistent or inconsistent by looking at just a single entry of the reduced row-echelon form matrix. We could program a computer to do it!

Notice that for a consistent system the row-reduced augmented matrix has $n+1 \in F$, so the largest element of F does not refer to a variable. Also, for an inconsistent system, $n+1 \in D$, and it then does not make much sense to discuss whether or not variables are free or dependent since there is no solution. Take a look back at Definition 1.9 and see why we did not need to consider the possibility of referencing x_{n+1} as a dependent variable.

With the characterization of Theorem 1.5, we can explore the relationships between r and n in light of the consistency of a system of equations. First, a situation where we can quickly conclude the inconsistency of a system.

Theorem 1.6 Inconsistent Systems, r and n

Suppose A_b is the augmented matrix of a system of linear equations in n variables. Suppose also that B is a row-equivalent matrix in reduced row-echelon form with r rows that are not completely zeros. If $r = n+1$, then the system of equations is inconsistent.

Proof If $r = n+1$, then $D = \{1, 2, 3, \ldots, n, n+1\}$ and every column of B contains a leading 1 and is a pivot column. In particular, the entry of column $n+1$ for row $r = n+1$ is a leading 1. Theorem 1.5 then says that the system is inconsistent. ■

Next, if a system is consistent, we can distinguish between a unique solution and infinitely many solutions, and furthermore, we recognize that these are the only two possibilities.

Theorem 1.7 Consistent Systems, r and n

Suppose A_b is the augmented matrix of a *consistent* system of linear equations with n variables. Suppose also that B is a row-equivalent matrix in reduced row-echelon form with r rows that are not zero rows. Then $r \leq n$. If $r = n$, then the system has a unique solution, and if $r < n$, then the system has infinitely many solutions.

Proof This theorem contains three implications that we must establish. Notice first that B has $n+1$ columns, so there can be at most $n+1$ pivot columns, i.e. $r \leq n+1$. If $r = n+1$, then Theorem 1.6 tells us that the system is inconsistent, contrary to our hypothesis. We are left with $r \leq n$.

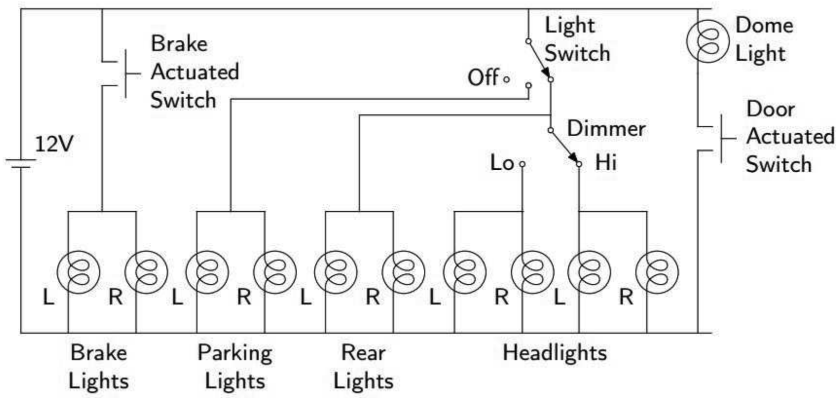
When $r = n$, we find $n - r = 0$ free variables (i.e. $F = \{n+1\}$) and any solution must equal the unique solution given by the first n entries of column $n+1$ of B.

When $r < n$, we have $n - r > 0$ free variables, corresponding to columns of B without a leading 1, excepting the final column, which also does not contain a leading 1 by Theorem 1.5. By varying the values of the free variables suitably, we can demonstrate infinitely many solutions. ■

Section 1.4 Application: Analyzing Networks

Systems of linear equations are used in many areas of science and engineering. Amongst others, they form one of the main tools in the analysis of various types of networks, such as traffic networks, social networks, electrical networks, water networks, ecological networks and metabolic networks. In this section we illustrate the potential of systems of linear equations in the study of electrical networks.

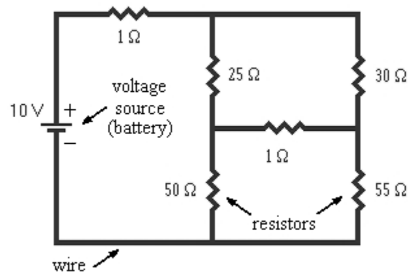
The diagram below shows some of a car's electrical network. The battery is on the left, drawn as stacked line segments. The wires are lines, shown straight and with sharp right angles for neatness. Each light is a circle enclosing a loop.



The designer of such a network needs to answer questions such as: how much electricity flows when both the hi-beam headlights and the brake lights are on? We will use linear systems to analyze simple electrical networks.

Required Concepts from Physics

Let us start with describing some elementary physical concepts[1]. Here is an electric circuit diagram.

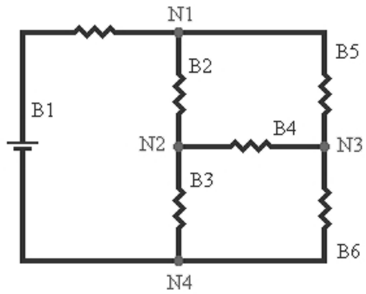


The diagram has a few symbols, such as the symbol for a voltage source (this could be for example a battery). Voltage sources generate a force or "pressure" which causes electric currents to flow in the circuit. The voltage source shown is rated at 10 volts (V). This means that the voltage or "pressure" is 10 volts higher at the $+$ or wide-line side of the battery than at the $-$ or narrow-line side of the battery. As a result, electric charges are given a force in the upward direction (from the - side to the + side).

In the diagram one also finds the symbol for a resistor. Resistors are devices that impede the flow of electric current. The resistor at the lower right has a resistance of 55 ohms (Ω). One also observes

[1]This section is based on http://mathonweb.com/help/backgd2.htm#Kirchoff's%20Voltage%20Law

the symbol for a wire. A wire is assumed to have no resistance. Let's now further annotate the diagram with nodes (a.k.a. vertices) and branches (a.k.a. edges).



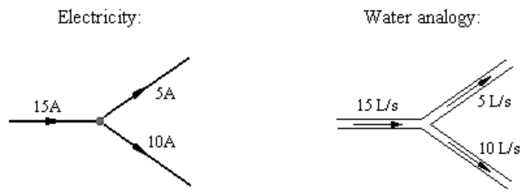
Nodes are points where 3 or more wires meet. This circuit contains 4 of them, denoted N1, ..., N4. A **branch** is any path in the circuit that has a node at each end and contains at least one voltage source or resistor but contains no other nodes. This circuit contains 6 branches, denoted B1, ..., B6. If branch B4 did not contain a resistor then it could be deleted and nodes N2 and N3 could be considered one and the same node.

Electric charge flowing in a branch in a circuit is analogous to water flowing in a pipe. The rate of flow of charge is called the **current**. It is measured in coulombs/second or amperes (A) just as the flow rate of water is measured in litres/second. Water is incompressible, which means that if 1 litre of water enters one end of a length of pipe then 1 litre must exit from the other end. The situation is the same with electric current. If the current is 1A at a certain point in a branch then it is 1A everywhere else in that branch. An immediate consequence of this is **Kirchhoff's Current Law**.

Theorem 1.8 Kirchhoff's Current Law

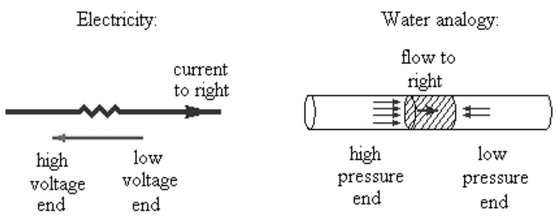
The sum of the currents flowing into a node equals the sum of the currents flowing out of the node.

Here is an example.



This diagram also shows how we draw an arrow on the branch to indicate the current flowing in the branch.

Electric current is the flow of electric charges. Electric **voltage** is the force that causes this flow. Just as a pump pushes a "plug" of water through a pipe by creating a pressure difference between its ends, so a battery pushes charge through a resistor by creating a voltage difference between the two ends of the resistor. The picture shows the analogy.



This diagram also shows how we draw an arrow beside a resistor or any other device to indicate a voltage difference between the two ends of that device. The arrow head is drawn pointing to the higher voltage end.

We have just seen that a voltage difference between the two ends of a resistor causes a current to flow through the resistor. For many substances the voltage and current are proportional. This is expressed in **Ohm's law** and any device that obeys it is called a resistor.

Theorem 1.9 Ohm's Law

Let us consider V as the difference in voltage between the two ends of the resistor (measured in volts), I is the current through the resistor (measured in amperes) and the proportionality constant R is the resistance of the resistor (measured in ohms). Then

$$V = I \times R.$$

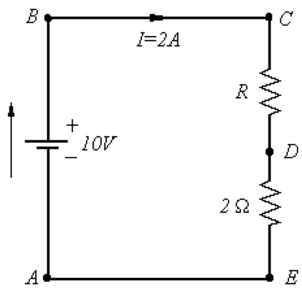
Just as the water pressure drops in a garden hose the farther one moves away from the tap, so the voltage changes as one moves around a circuit away from a voltage source.

Theorem 1.10 Kirchhoff's Voltage Law

Around any closed path in an electric circuit, the sum of the voltage drops through the resistors equals the sum of the voltage rises through the voltage sources.

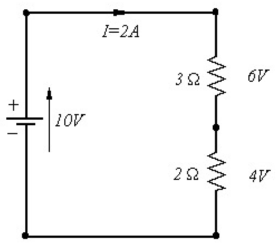
A closed path is a path through a circuit that ends where it starts.

Example 1.17
We'll use Kirchhoff's voltage law and Ohm's law to find the value of the unknown resistor R if it is known that a 2 ampere current flows in the circuit.

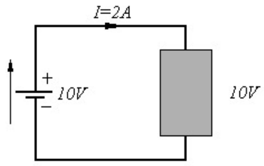


Let's follow the current as it flows clockwise around the circuit. If we start at A and assume the voltage there is 0 then at B the voltage must be 10 volts because the battery behaves like a pump that creates a higher pressure at the + side than the − side. At C the voltage is still 10 volts but it drops going to D through resistor R, and drops again going to E through the 2 ohm resistor. In fact it must return to 0 volts since A and E are at the same voltage (voltage does not change along an ideal wire that has no resistance).

Using Ohm's Law in the form $V = I \times R$, we find that the IR (voltage) drop across the 2 resistor is $(2A) \times (2\Omega) = 4V$. Then by Kirchhoff's Voltage Law the IR drop across the unknown resistor is $10V - 4V = 6V$. Again using $I = 2A$, Ohm's law in the form $R = V/I$ gives $R = 3\Omega$. The results are shown in this picture.



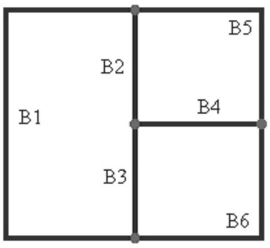
Notice the directions of the voltage arrows across each of the devices. Also notice that the voltage drops across the two resistors are proportional to their resistances. This is called the **Voltage Divider Rule**. This rule is useful in many situations. Suppose that we replaced the above circuit by the one shown here.



Suppose we didn't know what was inside the "black box" but did know that the current flowing into the black box was $2A$ and that the voltage across it was $10V$. Then Ohm's law, $R = V/I$, would tell us that the black box had a resistance of 5Ω. Notice that this is exactly the sum of the two resistances in the original circuit. This is true in general: two resistors R_1 and R_2 in series may be replaced by a single equivalent resistor R_{eq} whose resistance is the sum of the two resistances: $R_{eq} = R_1 + R_2$. ⊠

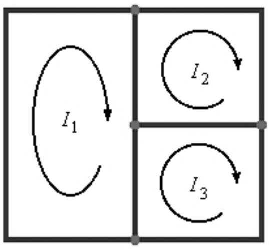
Branch and Loop Currents

In this diagram we have removed all the resistors and voltage sources so that we can focus attention on the topology of the network (i.e. the structure of the circuit) and count its nodes and branches.



"To solve a network" means to find the current flowing in each branch of the network. Since this circuit has 6 branches, this means calculating 6 branch currents.

Loop currents offer a more economical way to describe the current flow in a network. The currents in all 6 branches can be described in terms of just 3 loop currents as shown in the figure below. A loop current is defined as a constant current that flows around a closed path or loop. A closed path is a path through the network that ends where it starts.



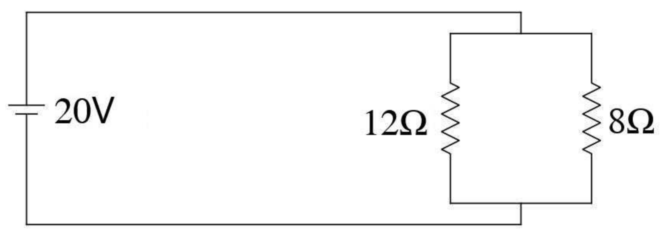
Each branch current is given by the algebraic sum of all the loop currents present in that branch. By algebraic sum we mean that the sign and direction of loop currents must be taken into account in the sum.

Computation of Branch Currents in Electrical Networks

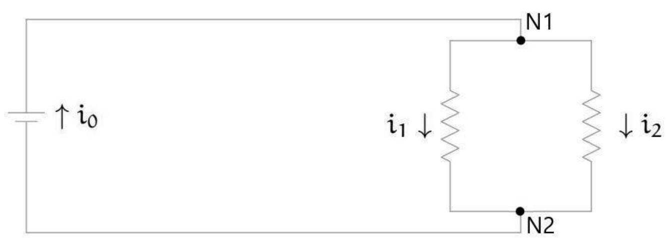
Systems of linear equations are used for computing the branch and loop current of electrical networks. We start with the computation of the branch current, which is the most easy task of the two. In this method, we set up and solve a system of equations in which the unknowns are branch currents. The steps in the branch current method are:

1. Count the number of branch currents required. Call this number n.
2. Call the n branch currents i_1, i_2,..., i_n and draw them on the circuit diagram.
3. Write down Kirchhoff's Current Law for each node and Kirchhoff's Voltage Law for each closed path. The result, after simplification, is a system of linear equations.
4. Solve the system of linear equations with one of the methods that we discuss in this course.

Example 1.18
We start with the analysis of a network that has two resistors in parallel.



As we see on the figure, there are 3 branches. We begin by labeling the branches as below. Let the current through the left branch of the parallel portion be i_1 and that through the right branch be i_2, and also let the current through the battery be i_0. Note that we don't need to know the actual direction of flow if current flows in the direction opposite to our arrow then we will get a negative number in the solution.



First, we apply Kirchoff's Current Law in each node. The split point in the upper right, N1, gives that $i_0 = i_1 + i_2$. Applied to the split point in the lower right, N2, it gives $i_1 + i_2 = i_0$.

Second, we apply Kirchoff's Voltage Law in each closed path. In the circuit that loops out of the top of the battery, down the left branch of the parallel portion, and back into the bottom of the battery, the voltage rise is 20 while the voltage drop is $i_1 \cdot 12$, so the Voltage Law gives that $12i_1 = 20$.

Similarly, the circuit from the battery to the right branch and back to the battery gives that $8i_2 = 20$. And, in the circuit that simply loops around in the left and right branches of the parallel portion (we arbitrarily take the direction of clockwise), there is a voltage rise of 0 and a voltage drop of $8i_2 - 12i_1$ so $8i_2 - 12i_1 = 0$.

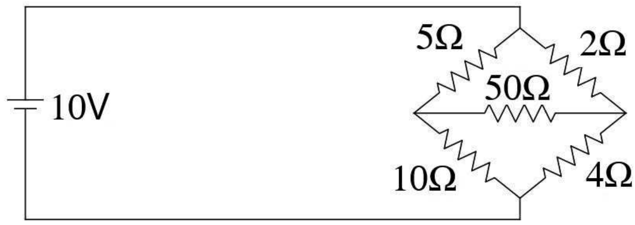
At the end we get:

$$\begin{aligned} i_0 - i_1 - i_2 &= 0 \\ -i_0 + i_1 + i_2 &= 0 \\ 12i_1 &= 20 \\ 8i_2 &= 20 \\ -12i_1 + 8i_2 &= 0 \end{aligned}$$

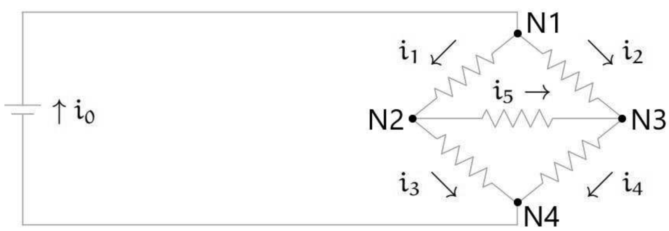
The solution is $i_0 = 25/6$, $i_1 = 5/3$, and $i_2 = 5/2$, all in amperes. (Incidentally, this illustrates that redundant equations can arise in practice.) ⊠

Example 1.19

Kirchhoff's laws can establish the electrical properties of very complex networks. The next diagram shows five resistors, whose values are in ohms, wired in series-parallel.



This is a Wheatstone bridge. To analyze it, we can place the arrows in this way.



Kirchhoff's Current Law, applied to the top node N1, the left node N2, the right node N3, and the bottom node N4 gives these.

$$\begin{aligned} i_0 &= i_1 + i_2 \\ i_1 &= i_3 + i_5 \\ i_2 + i_5 &= i_4 \\ i_3 + i_4 &= i_0 \end{aligned}$$

Kirchhoff's Voltage Law, applied to the inside loop (the i_0 to i_1 to i_3 to i_0 loop), the outside loop (the i_0 to i_2 to i_4 to i_0 loop), and the upper and lower loop not involving the battery, gives these.

$$5i_1 + 10i_3 = 10$$

$$
\begin{aligned}
2i_2 + 4i_4 &= 10 \\
5i_1 + 50i_5 - 2i_2 &= 0 \\
50i_5 + 4i_4 - 10i_3 &= 0
\end{aligned}
$$

Those suffice to determine the solution $i_0 = 7/3$, $i_1 = 2/3$, $i_2 = 5/3$, $i_3 = 2/3$, $i_4 = 5/3$, and $i_5 = 0$. ⊠

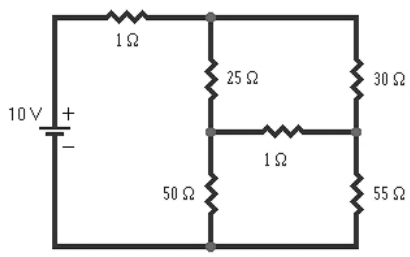
We can understand many kinds of networks in this way. For instance, we can analyze some networks of streets[2].

Computation of Loop Currents in Electrical Networks

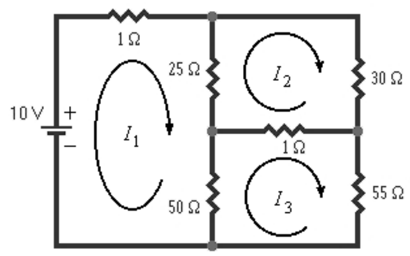
In this method, we set up and solve a system of equations in which the unknowns are loop currents. The currents in the various branches of the circuit are then easily determined from the loop currents. The steps in the loop current method are:

1. Count the number of loop currents required. Call this number m.
2. Choose m independent loop currents, call them I_1,I_2,..., I_m and draw them on the circuit diagram.
3. Write down Kirchhoff's Voltage Law for each loop. The result, after simplification, is a system of linear equations.
4. Solve the system of linear equations with one of the methods that we discuss in this course.
5. Reconstruct the branch currents from the loop currents.

Example 1.20
We search for the current flowing in each branch of this circuit.



The number of loop currents required is 3. We will choose the loop currents shown here.



[2]Watch for example https://www.youtube.com/watch?v=8Kg21jBCm-k for a nice introductory example.

Write down Kirchoff's Voltage Law for each loop. In particular, we see in the left loop that the voltage of the battery is $10V$. So, this is the voltage through the resistors of the left loop in total. There are 3 resistors in this left loop. The resistor of 1Ω is only being influenced by I_1. The other 2 resistors are being influenced by 2 current loops, in opposite direction through the resistor. So we have to take the difference between the loop currents to take into account the directions. At the end, for the left loop we find that $1I_1 + 25(I_1 - I_2) + 50(I_1 - I_3) = 10$. In total, we get the following system of equations.

$$\begin{aligned} 1I_1 + 25(I_1 - I_2) + 50(I_1 - I_3) &= 10 \\ 25(I_2 - I_1) + 30I_2 + 1(I_2 - I_3) &= 0 \\ 50(I_3 - I_1) + 1(I_3 - I_2) + 55I_3 &= 0 \end{aligned}$$

Collecting terms this becomes:

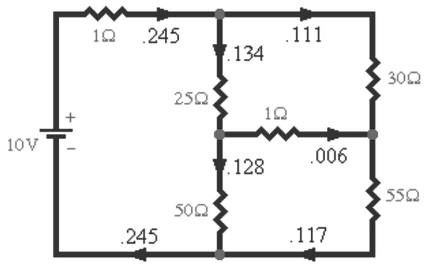
$$\begin{aligned} 76I_1 - 25I_2 - 50I_3 &= 10 \\ -25I_1 + 56I_2 - 1I_3 &= 0 \\ -50I_1 - 1I_2 + 106I_3 &= 0 \end{aligned}$$

Solving the system of equations using computer software gives the following loop currents (measured in amperes):

$$I_1 = 0.245\,, \qquad I_2 = 0.111\,, \qquad I_3 = 0.117\,.$$

Reconstructing the branch currents from the loop currents gives the results shown in the figure below. To explain the branch currents, we set

$$\begin{aligned} i_1 &= I_1 - I_2 = 0.134 \\ i_2 &= I_1 - I_3 = 0.128 \\ i_3 &= I_3 - I_2 = 0.006. \end{aligned}$$



⊠