Matrix Methods for Computational Modeling and Data Analytics

Virginia Tech · Spring 2019

Mark Embree
embree@vt.edu

$Ax = b$

version of 21 January 2019

Chapter 2
Linear Systems from Resistor Networks

CIRCUITS provide an elegant source of linear algebraic systems. Here we shall only consider simple systems involving a battery and some resistors, systems that sit at *static equilibrium*. As we proceed, the goal is not so much a model of a given circuit, but a modeling methodology that will apply to a broader set of problems we shall study over the next few weeks.

These notes draw heavily, in spirit, details, and examples, from the texts of Gilbert Strang[1] and Steve Cox[2], and the lab experiments of Cox et al.[3]

[1] Gilbert Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley, MA, 1986

[2] Steven J. Cox. *Matrix Analysis* in Situ. Rice University, 2013

[3] Steven J. Cox, Mark Embree, and Jeffrey M. Hokanson. *Physical Laboratory Manual for CAAM 335*. Rice University, 2013

2.1 Resistor network modeling

We begin with the example shown in Figure 2.1, consisting of six resistors and a constant voltage source. The goal is to determine what the potential (voltage) is at three nodes in the network, x_1, x_2, x_3.

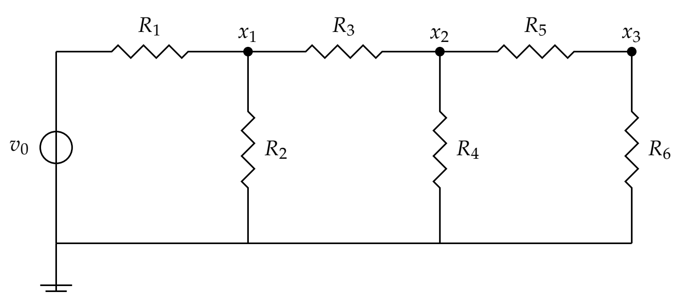

Figure 2.1: A circuit with an input voltage and six resistors. We seek the potential values at the nodes x_1, x_2, x_3. Cox uses this circuit as a primitive model for a neuron, where the horizontal resistors model resistance caused by the intercellular material, and the vertical resistors model leakage through the cell membrane.

While this network is simple, the methodology we shall derive applies to far more complicated circuits. By mastering this systematic approach, you will develop skills of broad applicability – and the linear algebra you need to solve such systems will prove even more useful.

With the constant voltage source v_0, this network sits at equilibrium. The potential at x_1, x_2, and x_3 will depend on v_0 and the relative strength of the resistors. We shall determine the decay of the potential at points farther the voltage source in key modeling four steps.

STEP 1 **Compute voltage drops across resistors.**

Across each of the six resistors, we compute the drop in voltage, denoted $e_1, \ldots, e_6$. As we consider the current flowing forth from the voltage source v_0, we measure the voltage drop across R_1 by the potential before R_1 minus the potential after, i.e.,

$$e_1 = v_0 - x_1.$$

We follow the same approach for the other five resistors. Since the far side of R_2 connects to ground,

$$e_2 = x_1 - 0.$$

Similarly, the drops across $R_3, \ldots, R_6$ are

$$\begin{aligned} e_3 &= x_1 - x_2 \\ e_4 &= x_2 - 0 \\ e_5 &= x_2 - x_3 \\ e_6 &= x_3 - 0. \end{aligned}$$

Even for this simple network, these equations start to get tedious. Just as for the population model in the last lecture, the organization of individual equations into matrix-vector form will illuminate. In this case we seek to relate six potential drops $e_1, \ldots, e_6$ to three potential values x_1, x_2, and x_3, and we must handle the input voltage as well. Collecting like terms in vectors, we have

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix} = \begin{bmatrix} v_0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad (2.1)$$

which we denote as

$$\mathbf{e} = \mathbf{v} - \mathbf{A}\mathbf{x}. \qquad (2.2)$$

STEP 2 **Apply Ohm's Law.**

We use Ohm's Law to relate the voltage drop across each resistor to current. You probably remember "$V = IR$" from physics class.

In this case we know "V" (the voltage drop) and "R" (the value of the resistor), and seek "I," so we use "$I = V/R$". We shall write the current at the six resistors as $y_1, \ldots, y_6$. Then at each of the resistors Ohm's Law gives

$$y_j = e_j/R_j, \quad j = 1, \ldots, 6.$$

As in Step 1, we want to write this in matrix–vector form, giving

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 1/R_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/R_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/R_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/R_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/R_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/R_6 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \end{bmatrix},$$

which we shall denote as

$$\mathbf{y} = \mathbf{K}\mathbf{e}. \tag{2.3}$$

The direction of these currents is illustrated in Figure 2.2. The diagonal form of the matrix **K** corresponds to the fact that this step of the modeling process does not encode any information about the connectivity of the network: **K** just describes the material properties of individual resistors.

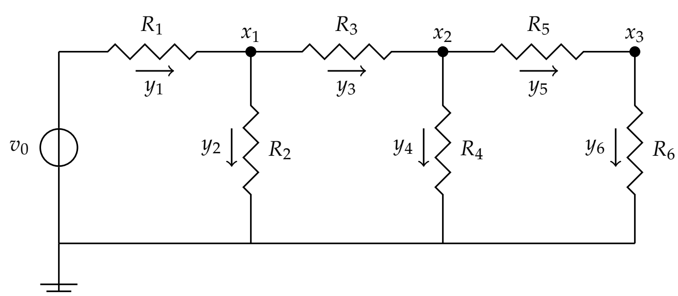

Figure 2.2: The circuit from Figure 2.1, with the current directions noted.

STEP 3 **Apply Kirchhoff's Current Law.**

Having related the potential values to voltage drops, and voltage drops to currents, we can now invoke the equilibrium condition that will allow us to compute the unknown voltages at each node. Kirchhoff's Current Law says that the current entering each node x_1, x_2, and x_3 must sum to zero:

$$\begin{aligned} &\text{at } x_1, \quad y_1 - y_2 - y_3 = 0; \\ &\text{at } x_2, \quad y_3 - y_4 - y_5 = 0; \\ &\text{at } x_3, \quad y_5 - y_6 = 0. \end{aligned}$$

Of course, we write this too as a matrix–vector product:

$$\begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \qquad (2.4)$$

Before giving this matrix a name, please pause to make this key observation: both the voltage drop computation in Step 1 and the current computation in Step 3 are determined by the wiring paths in the circuit – they encode the structure of the circuit. The first case maps the potential values x_1, x_2, and x_3 (via Ohm's law) to the currents $y_1, \ldots, y_6$. The second case does the reverse, in a sense: it imposes a condition on the currents at each of the potentials. Then it is no surprise then that the matrix in (2.4) is precisely the *transpose* of the matrix $\mathbf{A}$ in (2.1). Thus we conserve notation by writing (2.4) as

$$\mathbf{A}^T\mathbf{y} = \mathbf{0}. \qquad (2.5)$$

STEP 4 **Assembly.**

Remember what we are after: given the voltage v_0 (i.e., the vector $\mathbf{v}$ in (2.2), find the potentials x_1, x_2, and x_3. To obtain a clean expression for these potential values, we need to assemble the results of our first three steps.

Insert equation (2.3) for $\mathbf{y}$ into (2.5) to obtain

$$\begin{aligned} \mathbf{0} &= \mathbf{A}^T\mathbf{y} \\ &= \mathbf{A}^T\mathbf{K}\mathbf{e}. \end{aligned}$$

Now insert equation (2.2) for $\mathbf{e}$ into this last result to obtain

$$\begin{aligned} \mathbf{0} &= \mathbf{A}^T\mathbf{K}\mathbf{e} \\ &= \mathbf{A}^T\mathbf{K}(\mathbf{v} - \mathbf{A}\mathbf{x}). \end{aligned}$$

Rewrite this equation, defining $\mathbf{b} := \mathbf{A}^T\mathbf{K}\mathbf{v}$, to get the the fundamental form:

$$\boxed{\mathbf{A}^T\mathbf{K}\mathbf{A}\,\mathbf{x} = \mathbf{b}.} \qquad (2.6)$$

Assuming we know values for the resistances $R_1, \ldots, R_6$, we can assemble the matrix $\mathbf{A}^T\mathbf{K}\mathbf{A}\mathbf{x} = \mathbf{b}$, and arrive at a simple linear system of equations for the unknowns x_1, x_2, and x_3.

What size is the matrix $\mathbf{A}^T\mathbf{K}\mathbf{A}$? We see from the dimensions of the ingredients

$$\Big(3 \times 6\Big)\Big(6 \times 6\Big)\Big(6 \times 3\Big) = \Big(3 \times 3\Big)$$

that $\mathbf{A}^T\mathbf{K}\mathbf{A}$ is a 3×3 matrix. Let us compute it, keeping symbolic values for the resistances. First note that

$$\mathbf{A}^T\mathbf{K} = \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1/R_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/R_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/R_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/R_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/R_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/R_6 \end{bmatrix}$$

$$= \begin{bmatrix} 1/R_1 & -1/R_2 & -1/R_3 & 0 & 0 & 0 \\ 0 & 0 & 1/R_3 & -1/R_4 & -1/R_5 & 0 \\ 0 & 0 & 0 & 0 & 1/R_5 & -1/R_6 \end{bmatrix}.$$

Note: *postmultiplying* $\mathbf{A}^T$ by the diagonal matrix $\mathbf{K}$ *scaled the columns of* $\mathbf{A}^T$ by the diagonal entries. This is a general rule.

Yes, *premultiplying* by a diagonal matrix *scales the rows*. Had we first computed $\mathbf{KA}$, we would have seen this.

Now compute $(\mathbf{A}^T\mathbf{K})\mathbf{A}$:

$$(\mathbf{A}^T\mathbf{K})\mathbf{A} = \begin{bmatrix} 1/R_1 & -1/R_2 & -1/R_3 & 0 & 0 & 0 \\ 0 & 0 & 1/R_3 & -1/R_4 & -1/R_5 & 0 \\ 0 & 0 & 0 & 0 & 1/R_5 & -1/R_6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/R_1 + 1/R_2 + 1/R_3 & -1/R_3 & 0 \\ -1/R_3 & 1/R_3 + 1/R_4 + 1/R_5 & -1/R_5 \\ 0 & -1/R_5 & 1/R_5 + 1/R_6 \end{bmatrix}.$$

It remains to evaluate the right-hand side vector:

$$\mathbf{b} = (\mathbf{A}^T\mathbf{K})\mathbf{v} = \begin{bmatrix} 1/R_1 & -1/R_2 & -1/R_3 & 0 & 0 & 0 \\ 0 & 0 & 1/R_3 & -1/R_4 & -1/R_5 & 0 \\ 0 & 0 & 0 & 0 & 1/R_5 & -1/R_6 \end{bmatrix} \begin{bmatrix} v_0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} v_0/R_1 \\ 0 \\ 0 \end{bmatrix}.$$

We obtain the element-by-element version of the key equation (2.6):

$$\begin{bmatrix} 1/R_1 + 1/R_2 + 1/R_3 & -1/R_3 & 0 \\ -1/R_3 & 1/R_3 + 1/R_4 + 1/R_5 & -1/R_5 \\ 0 & -1/R_5 & 1/R_5 + 1/R_6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} v_0/R_1 \\ 0 \\ 0 \end{bmatrix}.$$

Notice that the matrix in this equation is symmetric.

This must be the case, since

$$(\mathbf{A}^T\mathbf{K}\mathbf{A})^T = \mathbf{A}^T\mathbf{K}^T(\mathbf{A}^T)^T = \mathbf{A}^T\mathbf{K}\mathbf{A}.$$

(We used the fact that $\mathbf{K} = \mathbf{K}^T$ since $\mathbf{K}$ is diagonal, and $(\mathbf{A}^T)^T = \mathbf{A}$ for all $\mathbf{A}$.)

What if all resistors have the same value, R Ohms? Then, clearing $1/R$ terms, our equation becomes

$$\begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} v_0 \\ 0 \\ 0 \end{bmatrix}. \qquad (2.7)$$

Given a value for v_0, you can readily solve for x_1, x_2, and x_3.

student experiments

2.1. Consider the case (2.7) when all resistors have the same values, R. Since this equation does not involve R, there is no way for the resistors to influence the values of x_1, x_2, and x_3. Does this make physical sense?

2.2. Implement the simple circuit in Figure 2.1 on a breadboard using a 9 V battery (or power supply, etc.) as your voltage source. Use the same type of resistors for $R_1, \ldots, R_6$, say $R = 100\,\Omega$. Measure the potentials x_1, x_2, and x_3 at the nodes. Now solve (2.7) for x_1, x_2, and x_3 (in matlab using the \ command, or by hand using Gaussian elimination, as in the section below). How does your answer compare to the prediction from our model? Now replace the vertical resistors with a different type of resistor, and repeat the experiment.

For example, if $v_0 = 9$, use:

```
S = [3 -1 0; -1 3 -1 ; 0 -1 2];
b = [9; 0; 0];
x = S\b
```

2.3. Imagine the pattern established in the circuit of Figure 2.1 is extended in a regular fashion to have $2N$ resistors (all of the same value, R) with $N-1$ nodes $x_1, \ldots, x_N$. For example, $N = 3$ in Figure 2.1, and Figure 2.3 shows the circuit for $N = 5$. How does the $N = 3$ equation in (2.7) generalize to arbitrary larger values of N? Can you deduce a general form for the matrix? If v_0 and R are held constant, what happens to the value of x_N as $N \to \infty$? Does this agree with your physical intuition?

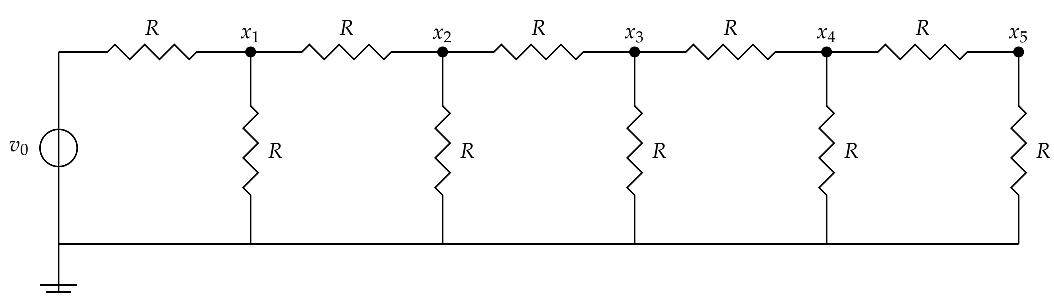

Figure 2.3: Extending the pattern of the $N = 3$ circuit in Figure 2.1 to $N = 5$ nodes and $2N$ resistors.

2.4. Use the four steps developed above to build the linear system $\mathbf{A}^T\mathbf{K}\mathbf{A}\mathbf{x} = \mathbf{b}$ for the branched circuit in Figure 2.4 (which models, in a primitive but suggestive manner, a branched neuron.) Notice how the branch affects the *sparsity* of the resulting matrix $\mathbf{A}^T\mathbf{K}\mathbf{A}$, compared to the unbranched model for large N studied in Experiment 2.3.

The *sparsity* denotes the zero–nonzero structure of the matrix.

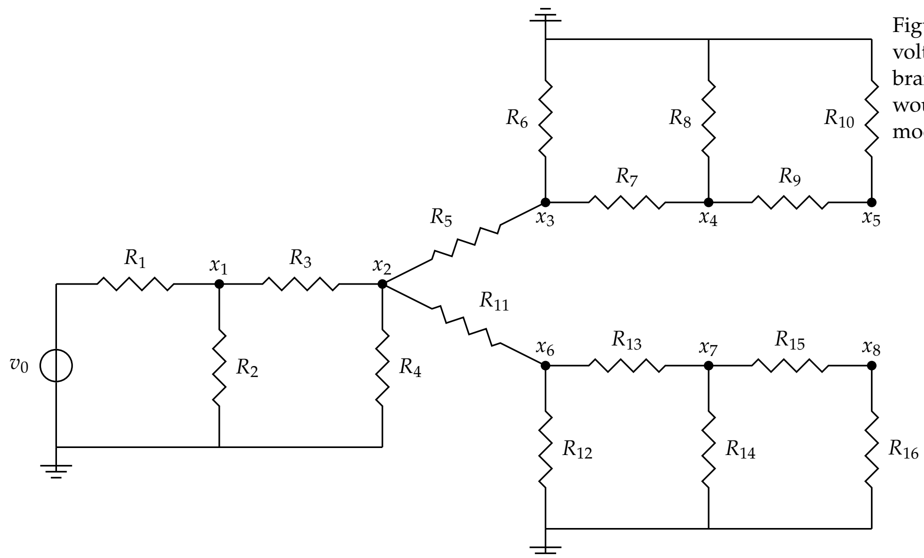

Figure 2.4: A circuit with an input voltage and sixteen resistors in a branched configuration. A neuron would have many such branches modeling its dendrites.

2.2 *Row reduction*

Now it is time to solve our system for x_1, x_2, and x_3. For a concrete example we shall address equation (2.7). The traditional way of solving such a system is *Gaussian elimination*. We presume you are already familiar with this technique, but we will briefly recap it here. Conventionally one writes the matrix and right-hand side together as the *augmented matrix*

"Gaussian" elimination is far more ancient than Gauss; an early user was the 3rd century Chinese mathematician Liu Hui. For more on this history, see:

Joseph Grcar. How ordinary elimination became Gaussian elimination. *Historia Math.*, 38:163–218, 2011

$$\left[\begin{array}{ccc|c} 3 & -1 & 0 & v_0 \\ -1 & 3 & -1 & 0 \\ 0 & -1 & 2 & 0 \end{array}\right]$$

Standard Gaussian elimination proceeds by converting to zero the entries below the diagonal of the matrix in the left of augmented form through use of *elementary row operations*. These operations, which are applied to entire rows of the augmented matrix, consist of three techniques:

1. EXCHANGE TWO ROWS;

2. MULTIPLY A ROW BY A NONZERO SCALAR;

3. ADD ONE ROW TO ANOTHER ROW.

These operations transform the augmented matrix while preserving the solution x_1, x_2, x_3 (provided the operations are applied to both the matrix and the right-hand side in the augmented form). We demonstrate this technique on our 3×3 matrix, and, perhaps unlike your past experience with Gaussian elimination, we shall keep the variable term v_0 in the right hand side of the equation.

Note that the apparently similar operations of exchanging, scaling, and adding *columns* do not preserve x_1, x_2, and x_3. Why?

Multiply row 2 by 3:

$$\left[\begin{array}{ccc|c} 3 & -1 & 0 & v_0 \\ -3 & 9 & -3 & 0 \\ 0 & -1 & 2 & 0 \end{array}\right].$$

Add row 1 to row 2 to zero out the (2,1) entry:

$$\left[\begin{array}{ccc|c} 3 & -1 & 0 & v_0 \\ 0 & 8 & -3 & v_0 \\ 0 & -1 & 2 & 0 \end{array}\right].$$

Multiply row 3 by 8:

$$\left[\begin{array}{ccc|c} 3 & -1 & 0 & v_0 \\ 0 & 8 & -3 & v_0 \\ 0 & -8 & 16 & 0 \end{array}\right].$$

Add row 2 to row 3 to zero out the (3,2) entry:

$$\left[\begin{array}{ccc|c} 3 & -1 & 0 & v_0 \\ 0 & 8 & -3 & v_0 \\ 0 & 0 & 13 & v_0 \end{array}\right]. \tag{2.8}$$

Now the subdiagonal entries, i.e., those in the (2,1), (3,1), and (3,2) positions, have been transformed to zero. The last augmented matrix is equivalent to the linear system

$$\begin{bmatrix} 3 & -1 & 0 \\ 0 & 8 & -3 \\ 0 & 0 & 13 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} v_0 \\ v_0 \\ v_0 \end{bmatrix},$$

and the *upper triangular form* of the matrix means that we can solve the equations from the bottom–up, for the matrix–vector equation is equivalent to the scalar equations

$$3x_1 - x_2 + 0x_3 = v_0 \tag{2.9}$$

$$8x_2 - 3x_3 = v_0 \tag{2.10}$$

$$13x_3 = v_0 \tag{2.11}$$

First solve (2.11) for x_3:

$$x_3 = \frac{v_0}{13}.$$

Substitute this formula for x_3 into (2.10) and solve for x_2:

$$x_2 = \frac{1}{8}\left(1 + \frac{3}{13}\right)v_0 = \frac{2v_0}{13}.$$

Finally, substitute the values of x_1 and x_2 into (2.9) and solve for x_1:

$$x_1 = \frac{1}{3}\left(1 + \frac{2}{13}\right)v_0 = \frac{5v_0}{13}.$$

We summarize the solution:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{v_0}{13}\begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}. \tag{2.12}$$

Now pause for one last essential step: Ask, *Does the answer make sense?* First off, the units are correct: x_1, x_2, and x_3 indeed have the same units as the voltage v_0. Moreover, with $v_0 > 0$ all entries of $\mathbf{x}$ are positive: we do not get negative voltages. And $x_1 > x_2 > x_3$: the potentials decrease with distance from the voltage source. Increasing v_0 uniformly scales the potentials. All this agrees with our physical intuition; the answer seems reasonable.

The application gives this sanity check: made-up problems lack this safeguard.

STUDENT EXPERIMENTS

2.5. Show how each of the three elementary row operations can be encoded in the form of a matrix–matrix product.

(i) Design a matrix $\mathbf{P}_{j,k}$ such that $\mathbf{P}_{j,k}\mathbf{S}$ swaps rows j and k of $\mathbf{S}$. For example, we want

$$\mathbf{P}_{1,2}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} d & e & f \\ a & b & c \\ g & h & i \end{bmatrix}.$$

(ii) Design a matrix $\mathbf{M}_j$ such that $\mathbf{M}_j\mathbf{S}$ multiplies row j of $\mathbf{S}$ by the scalar γ:

$$\mathbf{M}_2\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ \gamma d & \gamma e & \gamma f \\ g & h & i \end{bmatrix}.$$

(iii) Design a matrix $\mathbf{R}_{j,k}$ such that $\mathbf{R}_{j,k}\mathbf{S}$ replaces row j of $\mathbf{S}$ with the sum of rows j and k:

$$\mathbf{R}_{3,1}\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ a+g & b+h & c+i \end{bmatrix}.$$

2.3 *Gaussian elimination is LU factorization*

Now is the time to develop a more mature understanding of Gaussian elimination. This viewpoint was first articulated in the late 1940s, a time when early computer scientists were designing the first codes to solve linear systems on computers.[4] Focus on the matrix

$$\mathbf{S} = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

[4] G. W. Stewart. The decompositional approach to matrix computation. *Computing Sci. Eng.*, 2:50–59, 2000

Since the (3,1) entry is already zero, our first task is to zero out the (2,1) position. Varying slightly from the operations described above, we do so by adding 1/3 times the first row to the second row, consolidating two elementary row operations. Following on from Experiment 2.5, we perform this operation using a matrix-matrix product:

Above we added the first row to three times the second row to preserve integer entries. When designing algorithms for a computer, one prefers regularity to cosmetic beauty. So, we always replace a row by the sum of that row plus a multiple of a different row.

$$\mathbf{L}_{2,1}\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & -1 & 0 \\ 0 & 8/3 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Next, manipulate $\mathbf{L}_{2,1}\mathbf{S}$ to insert a zero in the (3,2) position, adding 3/8 of the new second row to the third row:

$$\mathbf{L}_{3,2}(\mathbf{L}_{2,1}\mathbf{S}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3/8 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 & 0 \\ 0 & 8/3 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & -1 & 0 \\ 0 & 8/3 & -1 \\ 0 & 0 & 13/8 \end{bmatrix}.$$

The resulting matrix is *upper triangular* (zero below the main diagonal), so we call it

$$\mathbf{U} = \begin{bmatrix} 3 & -1 & 0 \\ 0 & 8/3 & -1 \\ 0 & 0 & 13/8 \end{bmatrix}.$$

Now since $\mathbf{L}_{3,2}\mathbf{L}_{2,1}\mathbf{S} = \mathbf{U}$, we can write

$$\mathbf{S} = \mathbf{L}_{2,1}^{-1}\mathbf{L}_{3,2}^{-1}\mathbf{U}. \tag{2.13}$$

The inverses of the *lower triangular* matrices $\mathbf{L}_{3,2}$ and $\mathbf{L}_{2,1}$ are incredibly simple:

$$\mathbf{L}_{2,1}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{L}_{3,2}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3/8 & 1 \end{bmatrix}.$$

Check that the proposed inverses give $\mathbf{L}_{2,1}\mathbf{L}_{2,1}^{-1} = \mathbf{I}$ and $\mathbf{L}_{3,2}\mathbf{L}_{3,2}^{-1} = \mathbf{I}$.

Now compute the product $\mathbf{L}_{2,1}^{-1}\mathbf{L}_{3,2}^{-1}$ in (2.13):

$$\mathbf{L}_{2,1}^{-1}\mathbf{L}_{3,2}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3/8 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ 0 & -3/8 & 1 \end{bmatrix}.$$

This product of lower triangular matrices is also lower triangular, so we call it

$$\mathbf{L} := \mathbf{L}_{2,1}^{-1}\mathbf{L}_{3,2}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ 0 & -3/8 & 1 \end{bmatrix}. \tag{2.14}$$

We arrive at $\mathbf{S} = \mathbf{L}_{2,1}^{-1}\mathbf{L}_{3,2}^{-1}\mathbf{U} = \mathbf{L}\mathbf{U}$, so

$$\mathbf{S} = \mathbf{L}\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 \\ -1/3 & 1 & 0 \\ 0 & -3/8 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 & 0 \\ 0 & 8/3 & -1 \\ 0 & 0 & 13/8 \end{bmatrix} = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 3 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

This formula $\mathbf{S} = \mathbf{L}\mathbf{U}$ opens innumerable doors. For example, using the properties of inverses described in Lecture 1,

$$\mathbf{S}^{-1} = (\mathbf{L}\mathbf{U})^{-1} = \mathbf{U}^{-1}\mathbf{L}^{-1}.$$

Maybe you are dubious that we have helped the situation by transforming the problem of inverting one matrix $\mathbf{S}$ into the problem of inverting two matrices $\mathbf{U}$ and $\mathbf{L}$, then multiplying the results. But since $\mathbf{U}$ and $\mathbf{L}$ are upper and lower triangular matrices, they are easy to invert. For $\mathbf{U}$ we find

$$\mathbf{U}^{-1} = \begin{bmatrix} 1/3 & 1/8 & 1/13 \\ 0 & 3/8 & 3/13 \\ 0 & 0 & 8/13 \end{bmatrix}$$

while for $\mathbf{L}$, use equation (2.14) to see

$$\mathbf{L}^{-1} = (\mathbf{L}_{2,1}^{-1}\mathbf{L}_{3,2}^{-1})^{-1} = \mathbf{L}_{3,2}\mathbf{L}_{2,1} = \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 1/8 & 3/8 & 1 \end{bmatrix},$$

Indeed, the inverses of upper/lower triangular matrices are also upper/lower triangular.

so we arrive at

$$\mathbf{S}^{-1} = \begin{bmatrix} 1/3 & 1/8 & 1/13 \\ 0 & 3/8 & 3/13 \\ 0 & 0 & 8/13 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 1/8 & 3/8 & 1 \end{bmatrix} = \frac{1}{13}\begin{bmatrix} 5 & 2 & 1 \\ 2 & 6 & 3 \\ 1 & 3 & 8 \end{bmatrix}.$$

More importantly, we can now use the $\mathbf{LU}$ factorization of $\mathbf{S}$ to solve the equation $\mathbf{S}\mathbf{x} = \mathbf{b}$, i.e. (2.6): for one thing, we could write

$$\mathbf{x} = \mathbf{S}^{-1}\mathbf{b} = \frac{1}{13}\begin{bmatrix} 5 & 2 & 1 \\ 2 & 6 & 3 \\ 1 & 3 & 8 \end{bmatrix} \begin{bmatrix} v_0 \\ 0 \\ 0 \end{bmatrix} = \frac{v_0}{13}\begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}$$

in agreement with (2.12). However, we can arrive at $\mathbf{x}$ without explicitly computing $\mathbf{S}^{-1}$. Use $\mathbf{S} = \mathbf{L}\mathbf{U}$ to arrive at

$$\mathbf{L}\mathbf{U}\mathbf{x} = \mathbf{b},$$

or

$$\begin{aligned}\mathbf{U}\mathbf{x} &= \mathbf{L}^{-1}\mathbf{b} \\ &= \mathbf{L}_{3,2}\mathbf{L}_{2,1}\mathbf{b}.\end{aligned}$$

Take a moment to savor this equation. Why? Because (aside from the different way we chose to scale the rows when we added them), this last equation is equivalent to the "augmented matrix" result in (2.8)! When you applied elementary row operations to the augmented matrix (starting with $[\mathbf{S}|\mathbf{b}]$), you reduced $\mathbf{S}$ to the upper triangular form $\mathbf{U}$, and applied those same reducing transformations to $\mathbf{b}$, giving $\mathbf{L}_{3,2}\mathbf{L}_{2,1}\mathbf{b}$. When you then found $\mathbf{x}$ by solving equations (2.9)–(2.9), you were inverting the triangular matrix $\mathbf{U}$, i.e.,

$$\begin{aligned}\mathbf{x} &= \mathbf{U}^{-1}(\mathbf{L}_{3,2}\mathbf{L}_{2,1})\mathbf{b} \\ &= \mathbf{U}^{-1}\mathbf{L}^{-1}\mathbf{b} \\ &= \mathbf{S}^{-1}\mathbf{b}.\end{aligned}$$

STUDENT EXPERIMENTS

2.6. We seek a general formula for the inverse of a matrix that has simple structure like $\mathbf{L}_{2,1}$ and $\mathbf{L}_{3,2}$, i.e., an identity matrix with a single off-diagonal entry set to α (the others being zero). We can write such a matrix as

$$\mathbf{I} + \alpha\mathbf{e}_j\mathbf{e}_k^T,$$

where α is the entry that goes in the (j,k) position. (Recall from Lecture 1 that $\mathbf{e}_\ell$ denotes the ℓth column of the identity matrix, so $\mathbf{e}_j\mathbf{e}_k^T$ is the matrix that is zero in all entries, save for a 1 in the (j,k) position.) Inspired by the form of $\mathbf{L}_{2,1}^{-1}$ and $\mathbf{L}_{3,2}^{-1}$, guess a formula for $(\mathbf{I} + \alpha\mathbf{e}_j\mathbf{e}_k)^{-1}$ (for an arbitrary dimension) and show that it works.

2.4 *Epilogue: pivoted LU factorizations*

Thinking of Gaussian elimination as the matrix factorization $\mathbf{S} = \mathbf{L}\mathbf{U}$ is a higher form of thinking that has broad consequences for both theory and numerical algorithms. However, not every matrix has a decomposition of this form. For example, if

$$\mathbf{S} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$

then $\mathbf{S}$ is an invertible matrix, but there is no way to write $\mathbf{S} = \mathbf{L}\mathbf{U}$ with $\mathbf{L}$ lower triangular and $\mathbf{U}$ upper triangular. If you tried to apply the style of Gaussian elimination you learned in high school to this equation, you would start by swapping the rows. That is the key problem: row swaps destroy the lower-triangular structure in $\mathbf{L}$. In a course in numerical analysis, you will learn that you can encode the row swaps by premultiplying $\mathbf{S}$ by a matrix $\mathbf{P}$ whose columns are the same as the identity matrix, but arranged in a different order. For any invertible matrix $\mathbf{S}$ we can always factor

$$\mathbf{PS} = \mathbf{LU}.$$

There is so much more to say about this factorization, but in this course we must move on now to other topics. . . .

For more, see, e.g., the textbook:

Lloyd N. Trefethen and David Bau, III. *Numerical Linear Algebra*. SIAM, Philadelphia, 1997