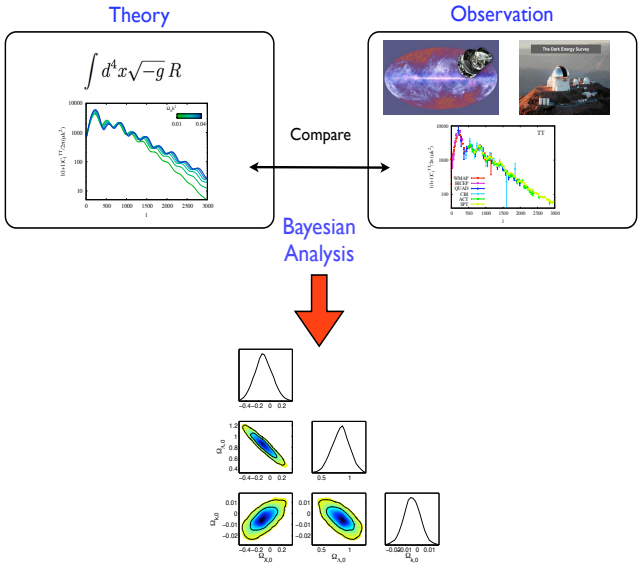
Numerical Methods



José-Alberto Vázquez

ICF-UNAM / Kavli-Cambridge

In progress

August 12, 2021

1

Linear Systems

This chapter deals with simultaneous linear algebraic equations that can be represented generally as

$$\begin{aligned}
E_1: &\quad a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1,\\
E_2: &\quad a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2,\\
&\qquad\qquad\qquad\qquad\qquad\quad \vdots\\
E_n: &\quad a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n,
\end{aligned}$$

1.0.1 The graphical method

A graphical solution is obtainable for two equations by plotting them on Cartesian coordinates with one axis corresponding to x_1 and the other to x_2.

$$a_{11}x_1 + a_{12}x_2 = b_1 \tag{1.1}$$

$$a_{21}x_1 + a_{22}x_2 = b_2 \tag{1.2}$$

Both equations can be solved for x_2:

$$x_2 = -\left(\frac{a_{11}}{a_{12}}\right)x_1 + \frac{b_1}{a_{12}} \tag{1.3}$$

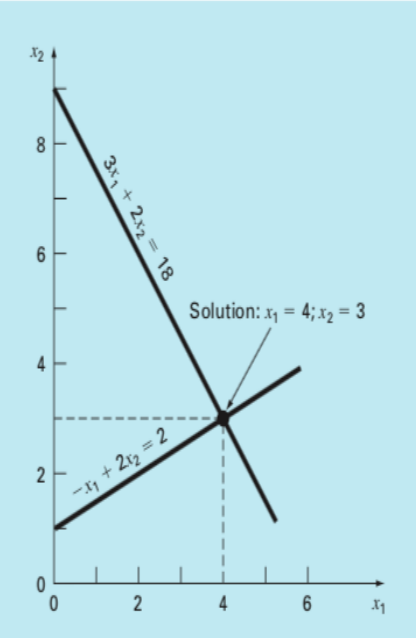
$$x_2 = -\left(\frac{a_{21}}{a_{22}}\right)x_1 + \frac{b_2}{a_{22}} \tag{1.4}$$

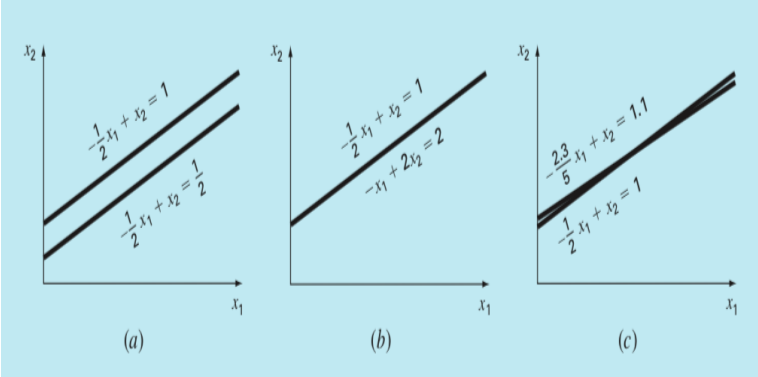
Thus, the equations are now in the form of straight lines; that is, $x_2 = (\text{slope})\, x_1 + \text{intercept}$. These lines can be graphed on Cartesian coordinates with x_2 as the ordinate and x_1 as the abscissa. The values of x_1 and x_2 at the intersection of the lines represent the solution.

Example

$$3x_1 + 2x_2 = 18 \tag{1.5}$$

$$-x_1 + 2x_2 = 2 \tag{1.6}$$





For three simultaneous equations, each equation would be represented by a plane in a three-dimensional coordinate system. The point where the three planes intersect would represent the solution.

By a sequence of operation, a linear system can be transformed to a more easily solved linear system that has the same solutions.

$$\begin{array}{lcrcrcrcrcl}
E_1 & : & x_1 & + & x_2 & & & + & 3x_4 & = & 4, \\
E_2 & : & 2x_1 & + & x_2 & - & x_3 & + & x_4 & = & 1, \\
E_3 & : & 3x_1 & - & x_2 & - & x_3 & + & 2x_4 & = & -3, \\
E_4 & : & -x_1 & + & 2x_2 & + & 3x_3 & - & x_4 & = & 4,
\end{array}$$

will be solved for x_1, x_2, x_3 and x_4. We use E_1 to eliminate the unknown x_1 from E_2, E_3 and E_4 by performing $(E_2 - 2E_1) \to (E_2)$, $(E_3 - 3E_1) \to (E_3)$ and $(E_4 + E_1) \to (E_4)$. The resulting system is

$$\begin{array}{lcrcrcrcrcl} E_1 & : & x_1 & + & x_2 & & & + & 3x_4 & = & 4, \\ E_2 & : & & - & x_2 & - & x_3 & - & 5x_4 & = & -7, \\ E_3 & : & & - & 4x_2 & - & x_3 & - & 7x_4 & = & -15, \\ E_4 & : & & + & 3x_2 & + & 3x_3 & + & 2x_4 & = & 8, \end{array}$$

In the new system, E_2 is used to eliminate x_2 from E_3 and E_4 by performing $(E_3 = 4E_2) \to (E_3)$ and $(E_4 + 3E_2) \to (E_4)$

$$\begin{array}{lcrcrcrcrcl} E_1 & : & x_1 & + & x_2 & & & + & 3x_4 & = & 4, \\ E_2 & : & & - & x_2 & - & x_3 & - & 5x_4 & = & -7, \\ E_3 & : & & & & & 3x_3 & + & 13x_4 & = & 13, \\ E_4 & : & & & & & & - & 13x_4 & = & -13, \end{array}$$

The system of equations is now in **triangular** (or **reduced**) **form** and can be solved for the unknowns by a **backward-substitution process**. The solution is therefore $x_4 = -1, x_2 = 2, x_3 = 0$ and $x_4 = 1$.

The only variation from system to system occurred in the coefficients of the unknowns and in the values on the right side. For this reason, a linear system is often replaced by a matrix.

Definition: An $\boldsymbol{n \times m}$ matrix is a rectangular array of elements with $\boldsymbol{n}$ rows and $\boldsymbol{m}$ columns.

The notation for an $n \times m$ matrix will be capital letter such as A for the matrix and lowercase letter with double subscripts, such as a_{ij} to refer to the entry at the intersection of the ith row and jth columns.

$$A = a_{ij} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

The $1 \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{bmatrix}$$

is called an **n-dimensional row vector**, and an $n \times 1$ matrix

$$A = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}$$

is called an **n-dimensional column vector**.

An $n \times (n+1)$ matrix can be used to represent the linear system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, && (1.7) \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, && (1.8) \\ \vdots \quad & \vdots && (1.9) \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n, && (1.10) \end{aligned}$$

by first constructing

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

and then combining these matrices to form the **augmented matrix**

$$[A, b] = \left[\begin{array}{cccc:c} a_{11} & a_{12} & \cdots & a_{1m} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2m} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} & b_n \end{array}\right]$$

So the matrix in the example:

$$\left[\begin{array}{cccc:c} 1 & 1 & 0 & 3 & 4 \\ 2 & 1 & -1 & 1 & 1 \\ 3 & -1 & -1 & 2 & -3 \\ -2 & 2 & 3 & -2 & 4 \end{array}\right] \rightarrow \left[\begin{array}{cccc:c} 1 & 1 & 0 & 3 & 4 \\ 0 & -1 & -1 & -5 & -7 \\ 0 & 0 & 3 & 13 & 13 \\ 0 & 0 & 0 & -13 & -13 \end{array}\right]$$

The procedure involved is called **Gaussian elimination with backward substitution**.

The general Gaussian procedure: first form the augmented matrix $\tilde{A}$

$$\tilde{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \vdots & a_{1,n+1} \\ a_{21} & a_{22} & \cdots & a_{2n} & \vdots & a_{2,n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \vdots & a_{n,n+1} \end{bmatrix}$$

Provided $a_{11} \neq 0$, the operations corresponding to $(E_j - (a_{j1}/a_{11})E_1) \to (E_j)$ are performed for each $j = 2, 3, \cdots, n$ to eliminate the coefficient of x_1 in each of these rows. With this in mind, we follow the sequential procedure for $i = 2, 3, \cdots, n$ and perform the operation $(E_j - (a_{ji}/a_{ii})E_i) \to (E_j)$ for each $j = i+1, i+2, \cdots, n$.

The resulting matrix has the form

$$\tilde{\tilde{A}} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & \vdots & a_{1,n+1} \\ 0 & a_{22} & \cdots & a_{2n} & \vdots & a_{2,n+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots 0 & a_{nn} & \vdots & a_{n,n+1} \end{bmatrix}$$

where the values of a_{ij} are not expected to agree with those in the original matrix $\tilde{A}$. For these linear system, a backward substitution can be performed. Solving the nth equation for x_n gives

$$x_n = \frac{a_{n,n+1}}{a_{nn}}$$

Solving the $(n-1)$st equation for x_{n-1} and using x_n yields

$$x_{n-1} = \frac{a_{n-1,n+1} - a_{n-1,n}x_n}{a_{n-1,n-1}}$$

Continuing this process, we obtain

$$x_i = \frac{a_{i,n+1} - a_{i,n}x_n - a_{i,n-1}x_{n-1} - \cdots - a_{i,i+1}x_{i+1}}{a_{ii}} = \frac{a_{i,n+1} - \sum_{j=i+1}^{n} a_{ij}x_j}{a_{ii}}$$

for each $i = n-1, n-2, \cdots, 2, 1$.

The Gaussian elimination procedure can be presented more precisely

The procedure will fail if one of the elements $a_{11}^{(1)}, a_{22}^{(2)}, a_{33}^{(3)}, \cdots, a_{n-1,n-1}^{(n-1)}, a_{nn}^{(n)}$ is zero because the step

$$\left(E_i - \frac{a_{i,k}^{(k)}}{a_{kk}^{(k)}} E_k \right) \to E_i$$

cannot be performed.

$$a_{ij}^{(k)} = \begin{cases} a_{ij}^{(k-1)}, & \text{when } i = 1, 2, \ldots, k-1 \text{ and } j = 1, 2, \ldots, n+1. \\ 0, & \text{when } i = k, k+1, \ldots, n \text{ and } j = 1, 2, \ldots, k-1, \\ a_{ij}^{(k-1)} - \dfrac{a_{i,k-1}^{(k-1)}}{a_{k-1,k-1}^{(k-1)}} a_{k-1,j}^{(k-1)}, & \text{when } i = k, k+1, \ldots, n \text{ and } j = k, k+1, \ldots, n+1. \end{cases}$$

$$\tilde{A}^{(k)} = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1,k-1}^{(1)} & a_{1k}^{(1)} & \cdots & a_{1n}^{(1)} & a_{1,n+1}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \cdots & a_{2,k-1}^{(2)} & a_{2k}^{(2)} & \cdots & a_{2n}^{(2)} & a_{2,n+1}^{(2)} \\ \vdots & \ddots & \ddots & & \vdots & \vdots & & \vdots & \vdots \\ & & & & a_{k-1,k-1}^{(k-1)} & a_{k-1,k}^{(k-1)} & \cdots & a_{k-1,n}^{(k-1)} & a_{k-1,n+1}^{(k-1)} \\ & & & & 0 & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} & a_{k,n+1}^{(k)} \\ & & & & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & a_{nk}^{(k)} & \cdots & a_{nn}^{(k)} & a_{n,n+1}^{(k)} \end{bmatrix} \quad (6.6)$$

Example:

$$\begin{array}{lcrcrcrcrcl} E_1 & : & x_1 & - & x_2 & + & 2x_3 & - & x_4 & = & -8, \\ E_2 & : & 2x_1 & - & 2x_2 & - & 3x_3 & - & 3x_4 & = & -20, \\ E_3 & : & x_1 & + & x_2 & + & x_3 & & & = & -2, \\ E_4 & : & x_1 & - & x_2 & + & 4x_3 & + & 3x_4 & = & 4, \end{array}$$

$$\tilde{A} = \tilde{A}^{(1)} = \left[\begin{array}{cccc:c} 1 & -1 & 2 & -1 & -8 \\ 2 & -2 & 3 & -3 & -20 \\ 1 & 1 & 1 & 0 & -2 \\ 1 & -2 & 4 & 3 & 4 \end{array}\right]$$

and performing the operations

$$(E_2 - 2E_1) \to (E_2), (E_3 - E_1) \to (E_3), \quad \text{and} \quad (E_4 - E_1) \to (E_4)$$

gives

$$\tilde{A}^{(2)} = \left[\begin{array}{cccc:c} 1 & -1 & 2 & -1 & -8 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & 2 & 4 & 12 \end{array}\right]$$

Since $a_{22}^{(2)}$, called the **pivot element**, is zero, the procedure cannot continue in its present form. Since $a_{32}^{(2)} \neq 0$, the operation $(E_2) \leftrightarrow (E_3)$ is performed to obtain a new matrix

$$\tilde{A}^{(2)'} = \left[\begin{array}{cccc:c} 1 & -1 & 2 & -1 & -8 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 0 & 2 & 4 & 12 \end{array}\right]$$

$\tilde{A}^{(3)}$ will be $\tilde{A}^{(2)'}$, and the computations continue with the operation $(E_4 + 2E_3) \rightarrow (E_4)$, giving

$$\tilde{A}^{(2)'} = \left[\begin{array}{cccc:c} 1 & -1 & 2 & -1 & -8 \\ 0 & 2 & -1 & 1 & 6 \\ 0 & 0 & -1 & -1 & -4 \\ 0 & 0 & 0 & 2 & 4 \end{array}\right]$$

Finally, the backward substitution is applied

$$x_4 = \frac{4}{2} = 2, \tag{1.11}$$

$$x_3 = \frac{[-4 - (-1)x_4]}{-1} = 2, \tag{1.12}$$

$$x_2 = \frac{[6 - x_4 - (-1)x_3]}{2} = 3, \tag{1.13}$$

$$x_1 = \frac{-8 - (-1)x_4 - 2x_3 - (-1)x_2}{1} = -7, \tag{1.14}$$

If $a_{pk}^{(k)} = 0$ for each p, it can be shown (Thm) that the linear system does not have a unique solution and the procedure stops.

The first system has an infinite number of solutions, and the second leads to a contradiction, hence no solution exists.

https://numpy.org/doc/stable/reference/routines.linalg.html
https://docs.scipy.org/doc/scipy/reference/linalg.html

(hw: from a geometrical standpoint$x_1 + 2x_2 = 3, 2x_1 + 4x_2 = 6$. Using Gaussian elimination with backward sustitution: $4x_1 - x_2 + x_3 = 8, 2x_1 + 5x_2 + 2x_3 = 3, x_1 + 2x_2 + 4x_3 = 11$.
)

$$\begin{aligned} x_1 + x_2 + x_3 &= 4, \\ 2x_1 + 2x_2 + x_3 &= 6, \\ x_1 + x_2 + 2x_3 &= 6, \end{aligned} \quad \text{and} \quad \begin{aligned} x_1 + x_2 + x_3 &= 4, \\ 2x_1 + 2x_2 + x_3 &= 4, \\ x_1 + x_2 + 2x_3 &= 6. \end{aligned}$$

These systems produce matrices

$$\tilde{A} = \left[\begin{array}{ccc:c} 1 & 1 & 1 & 4 \\ 2 & 2 & 1 & 6 \\ 1 & 1 & 2 & 6 \end{array}\right] \quad \text{and} \quad \tilde{A} = \left[\begin{array}{ccc:c} 1 & 1 & 1 & 4 \\ 2 & 2 & 1 & 4 \\ 1 & 1 & 2 & 6 \end{array}\right].$$

Since $a_{11} = 1$, we perform $(E_2 - 2E_1) \rightarrow (E_2)$ and $(E_3 - E_1) \rightarrow (E_3)$ to produce

$$\tilde{A} = \left[\begin{array}{ccc:c} 1 & 1 & 1 & 4 \\ 0 & 0 & -1 & -2 \\ 0 & 0 & 1 & 2 \end{array}\right] \quad \text{and} \quad \tilde{A} = \left[\begin{array}{ccc:c} 1 & 1 & 1 & 4 \\ 0 & 0 & -1 & -4 \\ 0 & 0 & 1 & 2 \end{array}\right].$$

At this point, $a_{22} = a_{32} = 0$. The algorithm requires that the procedure be halted, and no solution to either system is obtained. Writing the equations for each system gives

$$\begin{aligned} x_1 + x_2 + x_3 &= 4, \\ -x_3 &= -2, \\ x_3 &= 2, \end{aligned} \quad \text{and} \quad \begin{aligned} x_1 + x_2 + x_3 &= 4, \\ -x_3 &= -4, \\ x_3 &= 2. \end{aligned}$$

First we require $(n-i)$ divisions. The replacement of E_j by $(E_j - m_{ij}E_i)$ requires $(n-i)(n-i+1)$ multiplications. Then each term is subtracted, which requires $(n-i)(n-i+1)$ subtractions.

Multiplications/divisions

$$(n-i) + (n-i)(n-i+1) = (n-i)(n-i+2)$$

Additions/subtractions

$$(n-i)(n-i+1)$$

Multiplications/divisions

$$\sum_{i=1}^{n-1}(n-i)(n-i+2) = \frac{2n^3 + 3n^2 - 5n}{6}$$

Additions/subtractions

$$\sum_{i=1}^{n-1}(n-i)(n-i+1) = \frac{n^3 - n}{3}$$

For the backward substitution, requires $(n-i)$ multiplications and $(n-i-1)$ additions for each summation term, and one subtraction and one division.

Multiplications/divisions

$$1+\sum_{i=1}^{n-1}((n-i)+1)=\frac{n^2+n}{2}$$

Additions/subtractions

$$\sum_{i=1}^{n-1}((n-i-1)+1)=\frac{n^2-n}{2}$$

Giving a total number

Multiplications/divisions

$$\frac{2n^3+3n^2-5n}{6}+\frac{n^2+n}{2}=\frac{n^3}{3}+n^2-\frac{n}{3}$$

Additions/subtractions

$$\frac{n^3-n}{3}+\frac{n^2-n}{2}=\frac{n^3}{3}+\frac{n^2}{2}-\frac{5n}{6}$$

n	Multiplications/Divisions	Additions/Subtractions
3	17	11
10	430	375
50	44,150	42,875
100	343,300	338,250

1.0.2 Gauss-Jordan

The GJ method requires: Multiplications/divisions

$$\frac{n^3}{3}+\frac{3n^2}{2}-\frac{5n}{6}$$

Additions/subtractions

$$\frac{n^3}{2}-\frac{n}{2}$$

(hw: make a table comparing the required operation for $n=3,10,50,100$. Which method requires less computation?)

$$\begin{bmatrix} a_{11}^{(1)} & 0 & \cdots & 0 & \vdots & a_{1,n+1}^{(1)} \\ 0 & a_{22}^{(2)} & \ddots & \vdots & \vdots & a_{2,n-1}^{(2)} \\ \vdots & \ddots & \ddots & 0 & \vdots & \vdots \\ 0 & \cdots & 0 & a_{nn}^{(n)} & \vdots & a_{n,n+1}^{(n)} \end{bmatrix}$$

(hw: solve:

$$0.003000x_1 + 59.14x_2 = 59.17 \quad (1.15)$$

$$5.291x_1 - 6.130x_2 = 46.78 \quad (1.16)$$

)

$$\begin{aligned} \pi x_1 - ex_2 + \sqrt{2}x_3 - \sqrt{3}x_4 &= \sqrt{11}, \\ \pi^2 x_1 + ex_2 - e^2 x_3 + \tfrac{3}{7}x_4 &= 0, \\ \sqrt{5}x_1 - \sqrt{6}x_2 + x_3 - \sqrt{2}x_4 &= \pi, \\ \pi^3 x_1 + e^2 x_2 - \sqrt{7}x_3 + \tfrac{1}{9}x_4 &= \sqrt{2}. \end{aligned}$$

Actual solution $(0.78839378, -3.12541367, 0.16759660, 4.55700252)^t$.

1.1 Pivoting Strategies

To reduce roundoff error, it is often necessary to perform row interchanges even when the pivot elements are not zero. If $a_{kk}^{(k)}$ is small in magnitude compared to $a_{jk}^{(k)}$, the magnitude of the multiplier

$$m_{jk} = \frac{a_{jk}^{(k)}}{a_{kk}^{(k)}}$$

will be much larger that 1. Also, when performing the backward substitution.

The linear system

$$0.003000x_1 + 59.14x_2 = 59.17 \quad (1.17)$$

$$5.291x_1 - 6.130x_2 = 46.78 \quad (1.18)$$

has exact solution $x_1 = 10.00$ and $x_2 = 1.000$. Performing $(E_2 - m_{21}E_1) \rightarrow (E_2)$ and the appropriate rounding gives

$$0.003000x_1 + 59.14x_2 = 59.17 \tag{1.19}$$

$$-104300x_2 = -104400 \tag{1.20}$$

The disparity in the magnitudes of $m_{21}a_{13}$ and a_{23} has introduced roundoff error, but the roundoff error has not yet been propagated. Backward substitution yields

$$x_2 = 1.001,$$

and

$$x_1 \equiv \frac{59.17 - (59.14)(1.001)}{0.003000} = -10.00,$$

To avoid this problem, pivoting is performed by selecting a larger element $a_{pq}^{(k)}$ for the pivot and interchanging the kth and qth columns, if necessary. We determine the smallest $p \geq k$ such that

$$|a_{pk}^{(k)}|\max_{k\leq i\leq n}|a_{ik}^{k}|,$$

and perform $(E_k) \leftrightarrow (E_p)$.
Consider the same example. The pivoting procedure just described results in first finding

$$\max\left\{|a_{11}^{(1)}|, |a_{21}^{(1)}|\right\} = \max\{|0.003000|.|5.291|\} = |5.291| = |a_{21}^{(1)}|.$$

The operation $(E_2) \leftrightarrow (E_1)$ is then performed to give the system

$$5.291x_1 - 6.130x_2 = 46.78 \tag{1.21}$$

$$0.003000x_1 + 59.14x_2 = 59.17, \tag{1.22}$$

which reduces to

$$5.291x_1 - 6.130x_2 = 46.78 \tag{1.23}$$

$$59.14x_2 = 59.14 \tag{1.24}$$

The correct values $x_1 = 10.00$ and $x_2 = 1.000$. This technique is called **partial pivoting** or *maximal column pivoting.*

1.2 Linear Algebra and Matrix Inversion

Equal matrices, sum, multiplication by scalar, matrix product, transpose,

An **upper-triangular** $n \times n$ matrix $U = (u_{ij})$ has, for each $j = 1, 2, \cdots, n$, the entries

$$u_{ij} = 0, \quad \text{for each } i = j+1, j+2, \cdots, n$$

and a **lower-triangular** matrix $L = (l_{ij})$ has, for each $j = 1, 2, \cdots, n$, the entries

$$l_{ij} = 0, \quad \text{for each } i = j+1, j+2, \cdots, n$$

An $n \times n$ matrix A is said to be **nonsingular** (or invertible) if an $n \times n$ matrix A^{-1} exist with $AA^{-1} = A^{-1}A = I$. The matrix A^{-1} is called the **inverse** of A.

If we have the inverse of A, we can easily solve a linear system of the form $Ax = b$.

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

and then multiply both sides by the inverse, gives the solution $x_1 = 7/9, x_2 = 13/9$ and $x_3 = 5/3$.

Even though it is easier to solve a linear system of the form $Ax = b$ if A^{-1} is known, it is not computationally efficient to determine A^{-1} in order to solve the system.

To determine the inverse of the matrix

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

let us first consider the product AB, where B is an arbitrary 3×3 matrix.

$$AB = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

$$= \begin{bmatrix} b_{11} + 2b_{21} - b_{31} & b_{12} + 2b_{22} - b_{32} & b_{13} + 2b_{23} - b_{33} \\ 2b_{11} + b_{21} & 2b_{12} + b_{22} & 2b_{13} + b_{23} \\ -b_{11} + b_{21} + 2b_{31} & -b_{12} + b_{22} + 2b_{32} & -b_{13} + b_{23} + 2b_{33} \end{bmatrix}.$$

If $B = A^{-1}$, then $AB = I$, so we must have

$$\begin{array}{lll} b_{11} + 2b_{21} - b_{31} = 1, & b_{12} + 2b_{22} - b_{32} = 0, & b_{13} + 2b_{23} - b_{33} = 0, \\ 2b_{11} + b_{21} = 0, & 2b_{12} + b_{22} = 1, & 2b_{13} + b_{23} = 0, \\ -b_{11} + b_{21} + 2b_{31} = 0, & -b_{12} + b_{22} + 2b_{32} = 0, & -b_{13} + b_{23} + 2b_{33} = 1. \end{array}$$

Notice that the coefficients in each of the systems of equations are the same, the only change in the systems occurs on the right side of the equations, therefore

$$\left[\begin{array}{ccc:ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ -2 & 1 & 2 & 0 & 0 & 1 \end{array}\right]$$

First performing $(E_2 - 2E_1) \to (E_2)$ and $(E_3 + E_1) \to (E_3)$, followed by $(E_3 + E_2) \to (E_3)$ produces

$$\left[\begin{array}{ccc:ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -3 & 2 & -2 & 1 & 0 \\ 0 & 3 & 1 & 1 & 0 & 1 \end{array}\right] \quad \text{and} \quad \left[\begin{array}{ccc:ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -3 & 2 & -2 & 1 & 0 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array}\right]$$

Backward substitution gives (three systems of eqns.)

$$A^{-1} = \begin{bmatrix} -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

As we saw in that example, it is convenient to set up a larger augmented matrix, $[A\vdots I]$. Upon performing the elimination in accordance, we obtain an augmented matrix of the form $[U\vdots L]$.

1.3 Matrix Factorization

The steps used to solve a system of the form $Ax = b$ can be used to factor a matrix. The factorization is particularly useful when it has the form $A = LU$.

If A has been factored into the triangular form $A = LU$, then we can solve for x more easily by using a two-step process. First we let $y = Ux$ and solve the system $Ly = b$ for y. Since L is triangular, determining y from this equation requires only $\mathcal{O}(n^2)$ operations. Once y is known, the upper triangular system $Ux = y$ requires only an additional $\mathcal{O}(n^2)$ to determine the solution x. Then, the number of operations is reduced from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$.

If Gaussian elimination can be performed on the linear system $A\mathbf{x} = \mathbf{b}$ without row interchanges, then the matrix A can be factored into the product of a lower-triangular matrix L and an upper-triangular matrix U,

$$A = LU,$$

where $m_{ji} = a_{ji}^{(i)}/a_{ii}^{(i)}$,

$$U = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1,n}^{(n-1)} \\ 0 & \cdots & 0 & a_{nn}^{(n)} \end{bmatrix}, \quad \text{and} \quad L = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ m_{21} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ m_{n1} & \cdots & m_{n,n-1} & 1 \end{bmatrix}. \quad \blacksquare$$

The linear system

$$\begin{array}{rcrcrcrcr} x_1 & + & x_2 & & & + & 3x_4 & = & 4, \\ 2x_1 & + & x_2 & - & x_3 & + & x_4 & = & 1, \\ 3x_1 & - & x_2 & - & x_3 & + & 2x_4 & = & -3, \\ -x_1 & + & 2x_2 & + & 3x_3 & - & x_4 & = & 4, \end{array}$$

The system can be converted to the triangular system

$$\begin{array}{rcrcrcrcr} x_1 & + & x_2 & & & + & 3x_4 & = & 4, \\ & - & x_2 & - & x_3 & - & 5x_4 & = & -7, \\ & & & & 3x_3 & + & 13x_4 & = & 13, \\ & & & & & - & 13x_4 & = & -13, \end{array}$$

The multipliers m_{ij} and the upper triangular matrix produce the factorization

$$A = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -2 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix} = LU$$

A matrix A is **positive definite** if its symmetric and if $x^t Ax > 0$ for every n-dimensional vector $x \neq 0$.

The matrix A is **positive definite** if and only if A can be factored in the form LDL^t, where L is lower triangular with 1's on its diagonal and D is a diagonal matrix with positive diagonal entries.

The matrix A is **positive definite** if and only if A can be factored in the form LL^t, where L is lower triangular with nonzero diagonal entries.
They offer computational advantages because only half the storage is needed and, in most cases, only half of the computation time is required for their solution.

The matrix

$$A = \begin{bmatrix} 4 & -1 & 1 \\ -1 & 4.25 & 2.75 \\ 1 & 2.75 & 3.5 \end{bmatrix}$$

is positive definite. The factorization LDL^t of A given in Algorithm 6.5 is

$$A = LDL^t = \begin{bmatrix} 1 & 0 & 0 \\ -0.25 & 1 & 0 \\ 0.25 & 0.75 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -0.25 & 0.25 \\ 0 & 1 & 0.75 \\ 0 & 0 & 1 \end{bmatrix}.$$

and Choleski's Algorithm 6.6 produces the factorization

$$A = LL^t = \begin{bmatrix} 2 & 0 & 0 \\ -0.5 & 2 & 0 \\ 0.5 & 1.5 & 1 \end{bmatrix} \begin{bmatrix} 2 & -0.5 & 0.5 \\ 0 & 2 & 1.5 \\ 0 & 0 & 1 \end{bmatrix}.$$

1.4 Iterative techniques in Matrix Algebra

1.4.1 Gauss-Seidel

Suppose that for conciseness we limit ourselves to a 3×3 set of equations. The equations to solve yield ($Ax = b$):

$$x_1 = \frac{b_1 - a_{12}x_2 - a_{13}x_3}{a_{11}} \tag{1.25}$$

$$x_2 = \frac{b_2 - a_{21}x_1 - a_{23}x_3}{a_{22}} \tag{1.26}$$

$$x_3 = \frac{b_3 - a_{31}x_1 - a_{32}x_2}{a_{33}} \tag{1.27}$$

Now, we can start the solution process by choosing guesses for the x's. A simple way to obtain initial guesses is to assume that they are all zero. These zeros can be substituted into Eqs. which can be used to calculate a new value for $x_1 = b_1/a_{11}$. Example:

$$\begin{array}{rcrcrcr} 3x_1 & - & 0.1x_2 & - & 0.2x_3 & = & 7.85 \\ 0.1x_1 & + & 7x_2 & - & 0.3x_3 & = & -19.3, \\ 0.3x_1 & - & 0.2x_2 & + & 10x_3 & = & 71.4, \end{array}$$

Solve each of the equations for its unknown on the diagonal

$$x_1 = \frac{7.85 + 0.1x_2 + 0.2x_3}{3} \tag{1.28}$$

$$x_2 = \frac{-19.3 - 0.1x_1 + 0.3x_3}{7} \tag{1.29}$$

$$x_3 = \frac{71.4 - 0.3x_1 + 0.2x_2}{10} \tag{1.30}$$

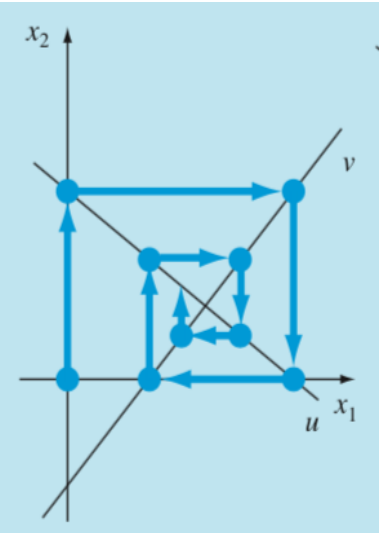
By assuming that x_2 and x_3 are zero, the first Eq. can be used to compute

$$x_1 = \frac{7.85 + 0 + 0}{3} = 2.616667$$

This value, along with the assumed value of $x_3 = 0$, can be substituted into the second to calculate

$$x_2 = \frac{-19.3 - 0.1(2.616667) + 0}{7} = -2.794524$$

and then use both of them to compute x_3 and so on. The true solution is $x_1 = 3, x_2 = -2.5, x_3 = 7$.



The diagonal coefficient in each of the equations must be larger that the sum of the absolute values of the other coefficients in the equation.

(hw: Tarea)

**Find α's such that has no solution, infinite number of solutions, the solution.

$$x_1 - x_2 + \alpha x_3 = -2 \tag{1.31}$$

$$-x_1 + 2x_2 - \alpha x_3 = 3 \tag{1.32}$$

$$\alpha x_1 + x_2 + x_3 = 2 \tag{1.33}$$

Prove:

- The product of two symmetric matrices is symmetric
- The product of two $n \times n$ lower (upper) triangular matrix is lower (upper) triangular.

***Find all values of α that make the matrix singular

$$A = \begin{bmatrix} 1 & -1 & \alpha \\ 2 & 2 & 1 \\ 0 & \alpha & -3/2 \end{bmatrix}$$

**Find α so that is positive definite

$$A = \begin{bmatrix} 2 & \alpha & -1 \\ \alpha & 2 & 1 \\ -1 & 1 & 4 \end{bmatrix}$$

**Find all values of α and β s.t. is singular, strictly diagonally dominant, symmetric, positive definite.

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

a) Factor the following matrix into LU decomposition

$$\begin{bmatrix} 2.1756 & 4.0231 & -2.1732 & 5.1967 \\ -4.0231 & 6.0000 & 0 & 1.1973 \\ -1.0000 & -5.2107 & 1.1111 & 0 \\ 6.0235 & 7.0000 & 0 & -4.1561 \end{bmatrix}$$