Numerical Analysis

Solutions of Nonlinear Equations and Linear Systems

Elmer S. Poliquit

Iterative Methods

Finding one or more roots of an equation

$$f(x) = 0$$

is one of the more commonly occurring problems of applied mathematics. In most cases explicit solutions are not available and we must be satisfied with being able to find a root to any specified degree of accuracy. The numerical methods for finding the roots are called **iterative methods**.
We will discuss the finding of roots in the following conditions:

1. $f(x)$ is any continuously differentiable real valued function of a real variable x.
2. The polynomial equation,

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = 0, \quad \text{where} \quad a_n \neq 0.$$

The Bisection Method

Assume that $f(x)$ is continuous on a given interval $[a, b]$ and that it also satisfies

$$f(a)f(b) < 0$$

Using the intermediate value theorem, the function $f(x)$ must have at least one root in $[a, b]$. Usually $[a, b]$ is chosen to contain only one root α, but the following algorithm for the bisection method will always converge to some root α in [a, b], because of $f(a)f(b) < 0$.

The Bisection Method

Algorithm

Bisect(f, a, b, root, ϵ)

1. Define $c := \frac{a+b}{2}$.
2. If $|b - c| \leq \epsilon$, then accept root $:= c$, and exit.
3. If $f(b) \cdot f(c) \leq 0$, then $a := c$; otherwise, $b := c$.
4. Return to step 1.

The interval $[a, b]$ is halved in size for every pass through the algorithm. Because of step 3, $[a, b]$ will always contain a root of $f(x)$. Since a root α is in $[a, b]$, it must lie within either $[a, c]$ or $[c, b]$; and consequently

$$|c - \alpha| \leq b - c = c - a$$

This is justification for the test in step 2. On completion of the algorithm, c will be an approximation to the root with

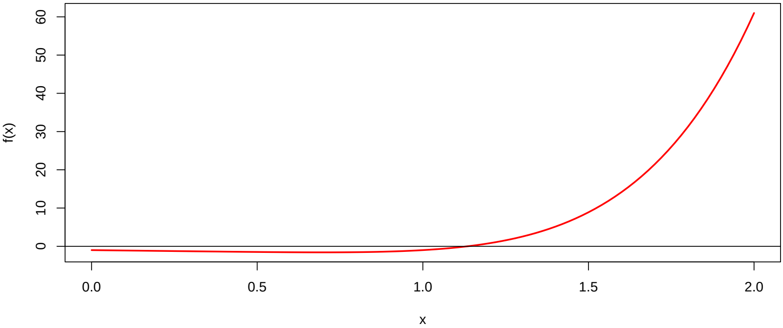
$$|c - \alpha| \leq \epsilon.$$

The Bisection Method

Example

Find the largest real root α of

$$f(x) \equiv x^6 - x - 1 = 0.$$



The Bisection Method

Example

Find the largest real root α of

$$f(x) \equiv x^6 - x - 1 = 0.$$

We can start with $[a, b] = [1, 1.5]$.
Bisect($f(x) \equiv x^6 - x - 1$, 1, 1.5, c_1, $\epsilon = 0.00005$)

1. $c_1 := \frac{1+1.5}{2} = 1.25$.
2. $|b - c_1| \not\leq 0.00005 \to |1.5 - 1.25| = 0.25 \not\leq 0.00005$.
3. $f(1.5) \cdot f(1.25) \not\leq 0 \to b = 1.25$
4. Return to step 1 where $a = 1$ and $b = c_1$

The Bisection Method

Example

Find the largest real root α of

$$f(x) \equiv x^6 - x - 1 = 0.$$

Bisect($f(x) \equiv x^6 - x - 1$, 1, 1.25, $c_2, \epsilon = 0.00005$)

1. $c_2 := \frac{1+1.25}{2} = 1.125$.
2. $|c_1 - c_2| \nleq 0.00005 \rightarrow |1.25 - 1.125| = 0.125 \nleq 0.00005$.
3. $f(1.25) \cdot f(1.125) \leq 0 \rightarrow a = c_2 = 1.125$
4. Return to step 1 where $a = c_2 = 1.125$ and $b = c_1 = 1.25$

To examine the speed of convergence, let c_n denote the nth value of c in the algorithm. Then it is easy to see that $\alpha = \lim_{n\rightarrow\infty} c_n$ and

$$|a - c_n| \leq \left[\frac{1}{2}\right]^n (b - a).$$

where $(b - a)$ denotes the length of the original interval input into Bisect.

Bisection Method

How many iterations do we need?

Since $\epsilon = 0.00005$, so $|a - c_n| = 0.00005$.

$$|a - c_n| \leq \left[\frac{1}{2}\right]^n (b - a) \rightarrow 0.00005 \leq \left[\frac{1}{2}\right]^n (1.5 - 1)$$

$$0.00005 \leq \left[\frac{1}{2}\right]^n (0.5) \rightarrow n \geq 14$$

Bisection Method

n	a	b	c	f(b)	f(c)	f(b)f(c)
1	1.00000	1.50000	1.25000	8.89063	1.56470	13.91113663
2	1.00000	1.25000	1.12500	1.56470	-0.09771	-0.15289200
3	1.12500	1.25000	1.18750	1.56470	0.61665	0.96487530
4	1.12500	1.18750	1.15625	0.61665	0.23327	0.14384599
5	1.12500	1.15625	1.14063	0.23327	0.06158	0.01436419
6	1.12500	1.14063	1.13281	0.06158	-0.01958	-0.00120542
7	1.13281	1.14063	1.13672	0.06158	0.02062	0.00126967
8	1.13281	1.13672	1.13477	0.02062	0.00043	0.00000880
9	1.13281	1.13477	1.13379	0.00043	-0.00960	-0.00000410
10	1.13379	1.13477	1.13428	0.00043	-0.00459	-0.00000196
11	1.13428	1.13477	1.13452	0.00043	-0.00208	-0.00000089
12	1.13452	1.13477	1.13464	0.00043	-0.00083	-0.00000035
13	1.13464	1.13477	1.13470	0.00043	-0.00020	-0.00000009
14	1.13470	1.13477	1.13474	0.00043	0.00011	0.00000005
15	1.13470	1.13474	**1.13472**	0.00011	-0.00004	0.00000000

Bisection Method

```r
f=function(x) x^6-x-1
uniroot(f, c(1,1.5))
```

```
$root
[1] 1.134723

$f.root
[1] -1.061001e-05

$iter
[1] 7

$init.it
[1] NA

$estim.prec
[1] 6.103516e-05
```

Bisection Method

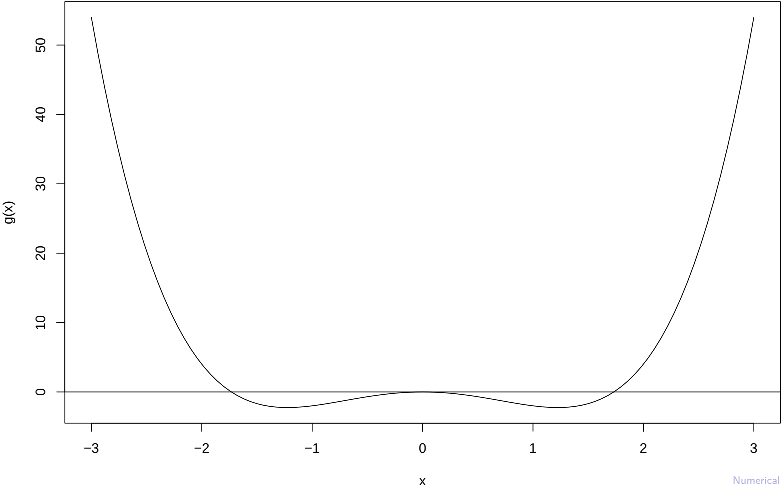
```r
library(cmna)
f=function(x) x^6-x-1
bisection(f, 1, 1.5, tol = 0.00005, m=14)
```

```
[1] 1.13472
```

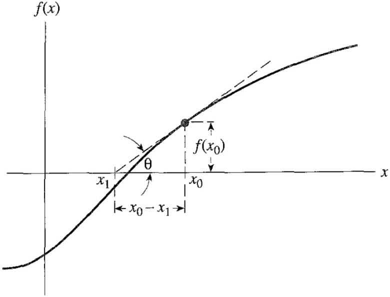
Bisection Method

Try this!

$$x^4 - 3x^2 = 0$$



Newton's Method



The calculation scheme follows immediately from the right triangle shown in the figure above, which has the angle of inclination of the tangent line to the curve at $x = x_0$ as one of its acute angles:

$$\tan\theta = f'(x_0) = \frac{f(x_0)}{x_0 - x_1} \to x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \to x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} \to \cdots$$

Thus, the iteration formula is

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n \geq 0.$$

Newton's Method

Example

Solve for the root of

$$f(x) = x^6 - x - 1 = 0.$$

The $f'(x) = 6x^5 - 1$.

```
6 * x^5 - 1
```

Let the initial estimate be $x_0 = 2$. Then

```
f=function(x){x^6-x-1}
df=function(x){6 * x^5 - 1}
x1=2-(f(2)/df(2))
x1
```

```
[1] 1.680628
```

Newton's Method

```
x2=1.680628-(f(1.680628)/df(1.680628))
x2
```

```
[1] 1.430739
```

```
x3=1.430739-(f(1.430739)/df(1.430739))
x3
```

```
[1] 1.254971
```

```
x4=1.254971-(f(1.254971)/df(1.254971))
x4
```

```
[1] 1.161538
```

```
x5=1.161538-(f(1.161538)/df(1.161538))
x5
```

```
[1] 1.136353
```

Newton's Method

```
x6=1.136353-(f(1.136353)/df(1.136353))
x6
```

```
[1] 1.134731
```

```
x7=1.134731-(f(1.134731)/df(1.134731))
x7 #Same with Bisection Method up to 5 correct decimal dig
```

```
[1] 1.134724
```

```
uniroot(f,c(1,1.5))$root
```

```
[1] 1.134723
```

The convergence is much more rapid than Bisection method.
This formula will be used to show that Newton's method has a quadratic order of convergence, $p = 2$.

Newton's Method

```r
xa=uniroot(f,c(1,1.5))$root
y=c(x1,x2,x3,x4,x5,x6,x7)
data.frame(y,error=round(xa-y,5))
```

```
         y    error
1 1.680628 -0.54591
2 1.430739 -0.29602
3 1.254971 -0.12025
4 1.161538 -0.02682
5 1.136353 -0.00163
6 1.134731 -0.00001
7 1.134724  0.00000
```

Newton's Method

Theorem

Assume $f(x)$, $f'(x)$, and $f''(x)$ are continuous for all x in some neighborhood of α, and assume $f(\alpha) = 0$, $f'(\alpha) \neq 0$. Then if x_0 is chosen sufficiently close to α, the iterates x_n, $n \geq 0$, of the iteration formula will converge to α. Moreover,

$$\lim_{n \to \infty} \frac{\alpha - x_{n+1}}{(\alpha - x_n)^2} = -\frac{f''(\alpha)}{2f'(\alpha)}$$

proving that the iterates have an order of convergence $p = 2$.
Refer the proof to the book.

Newton's Method

Error Estimation

For Newton's method, the standard error estimate is

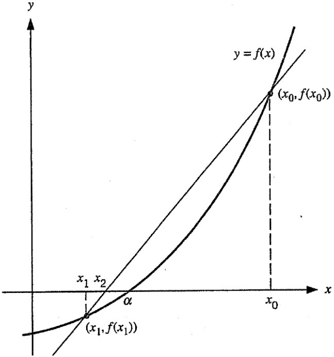
$$\alpha - x_n = x_{n+1} - x_n.$$

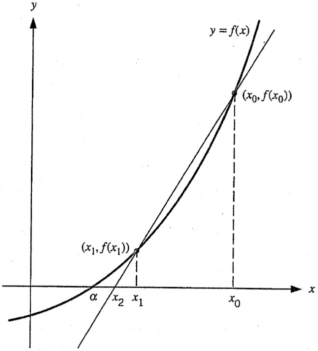
For the relative error, use

$$\frac{\alpha - x_n}{\alpha} = \frac{x_{n+1} - x_n}{x_{n+1}}.$$

Secant Method

Assume that two initial guesses to α are known and denote them by x_0 and x_1. They may occur on opposite sides of α as shown in the left figure, or on the other side of α as shown on the right figure.





Secant Method

The two points $(x_0, f(x_0))$ and $(x_1, f(x_1))$, on the graph of $y = f(x)$, determine a straight line, called a *secant line*. This line is an approximation to the graph of $y = f(x)$ and its root x_2 is an approximation of α.
Using the slope formula with the secant line, we have

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_1) - 0}{x_1 - x_2}$$

Solving for x_2,

$$x_2 = x_1 - f(x_1) \cdot \frac{x_1 - x_0}{f(x_1) - f(x_0)}.$$

Secant Method

Having found x_2, drop x_0 and use x_1, x_2 as a new set of approximate values for α. This leads to an improved value x_3; and this process can be continued indefintely.
Doing so, obtain the general iteration formula

$$x_{n+1} = x_n - f(x_n) \cdot \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1)})}, \quad n \geq 1.$$

This is the **secant method**.

Secant Method

Example

Solve for the root of

$$f(x) = x^6 - x - 1 = 0.$$

where $x_0 = 1$ and $x_1 = 2$. Then for x_2,

```
f=function(x){x^6-x-1}
x2=2-f(2)*(2-1)/(f(2)-f(1))
x2
```

```
[1] 1.016129
```

```
x3=1.016129-f(1.016129)*(1.016129-2)/(f(1.016129)-f(2))
x3
```

```
[1] 1.030675
```

Secant Method

```
x4=1.030675-f(1.030675)*(1.030675-1.016129)/
  (f(1.030675)-f(1.016129))
x4
```

```
[1] 1.175689
```

```
x5=1.175689-f(1.175689)*(1.175689-1.030675)/
  (f(1.175689)-f(1.030675))
x5
```

```
[1] 1.123679
```

```
x6=1.123679-f(1.123679)*(1.123679-1.175689)/
  (f(1.123679)-f(1.175689))
x6
```

```
[1] 1.133671
```

Secant Method

```r
x7=1.133671-f(1.133671)*(1.133671-1.123679)/
  (f(1.133671)-f(1.123679))
x7
```

```
[1] 1.134753
```

```r
x8=1.134753-f(1.134753)*(1.134753-1.133671)/
  (f(1.134753)-f(1.133671))
x8 #same value with Secant Method at 7th iteration
```

```
[1] 1.134724
```

```r
uniroot(f,c(1,2))$root
```

```
[1] 1.134706
```

Error Analysis

$$x_{n+1} = x_n - f(x_n) \cdot \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1)})}, \quad n \geq 1$$

Multiply both sides of above equation (secant method) by -1 and then add α to both sides, obtaining

$$\alpha - x_{n+1} = -(\alpha - x_{n-1})(\alpha - x_n)\frac{f''(\zeta_n)}{2f'(\xi_n)}$$

with ξ_n between x_{n-1} and x_n, and ζ_n between x_{n-1}, x_n, and α.
Using this error formula, we can examine the convergence of the secant method.
You can check the derivation of the right side of the equation from the book.

Try this in your breakout room!

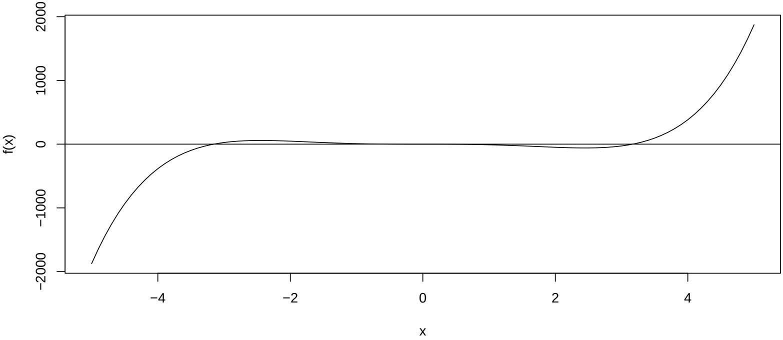
$$x_{n+1} = x_n - f(x_n) \cdot \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1)})}, \quad n \geq 1$$

Example
Solve for the root of

$$f(x) = x^5 - 10x^3 - 1 = 0.$$

where $x_0 = 2$ and $x_1 = 4$. Tolerance $= 0.0001$.

```
[1] 3.16725
```



Solutions of Linear Systems

A **matrix** is a rectangular array of numbers in which not only the value of the number is important but also its position in the array.

- The size of the matrix is described by the number of its rows and columns.
- A matrix of n rows and m columns is said to be $n \times m$.
- The elements of the matrix are generally enclosed in brackets, and double-subscripting is the common way of indexing the elements.
- The first subscript also denotes the row, and the second denotes the column in which the element occurs. Capital letters are used to refer to matrices.

Example

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} = [a_{ij}], i = 1, 2, \cdots, n, j = 1, 2, \cdots, m.$$

Solutions of Linear Systems

Two matrices of the same size may be added or subtracted. The sum of $A = [a_{ij}]$ and $B = [b_{ij}]$ is the matrix whose elements are the sum of the corresponding elements of A and B:

$$C = A + B = [a_{ij}] + [b_{ij}] = [c_{ij}].$$

The difference of $A = [a_{ij}]$ and $B = [b_{ij}]$ is

$$D = A - B = [a_{ij}] - [b_{ij}] = [d_{ij}].$$

Multiplication of two matrices is defined as follows, when A is $n \times m$ and B is $m \times r$:

$$[a_{ij}] \times [b_{ij}] = [c_{ij}].$$

The **transpose** of a matrix $[a_{ij}]$ is $[a_{ij}]^T = [a_{ji}]$.

Solutions of Linear Systems

$A * B$ - number of columns of A = number of rows of B

3×2 by 2×5

$[a_{ij}] \times [b_{ij}] = [c_{ij}] =$

$$\begin{bmatrix} (a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1m}b_{m1}) & \cdots & (a_{11}b_{1r} + \cdots + a_{1m}b_{mr}) \\ (a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2m}b_{m1}) & \cdots & (a_{21}b_{1r} + \cdots + a_{2m}b_{mr}) \\ \vdots & \ddots & \vdots \\ (a_{n1}b_{11} + a_{n2}b_{21} + \cdots + a_{nm}b_{m1}) & \cdots & (a_{n1}b_{1r} + \cdots + a_{nm}b_{mr}) \end{bmatrix}$$

$$[c_{ij}] = \sum_{k=1}^{m} a_{ik}b_{kj}, \quad i = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, r.$$

If $kA = C$, then $c_{ij} = ka_{ij}$.

Solutions of Linear Systems

Examples

Suppose

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

```r
A=matrix(c(3,7,1,-2,1,-3), nrow = 2, byrow = T)
B=matrix(c(5,-2,0,3,1,-1), nrow = 3, byrow = T)
A%*%B
```

```
     [,1] [,2]
[1,]   16   14
[2,]  -13   10
```

Solutions of Linear Systems

```r
A=matrix(c(3,7,1,-2,1,-3), nrow = 2, byrow = T)
B=matrix(c(5,-2,0,3,1,-1), nrow = 3, byrow = T)
B%*%A
```

```
     [,1] [,2] [,3]
[1,]   19   33   11
[2,]   -6    3   -9
[3,]    5    6    4
```

```r
k=c(-3,1,4)
A%*%k
```

```
     [,1]
[1,]    2
[2,]   -5
```

Solutions of Linear Systems

```
A=matrix(c(3,7,1,-2,1,-3), nrow = 2, byrow = T)
B=matrix(c(5,-2,0,3,1,-1), nrow = 3, byrow = T)
2*A
```

```
     [,1] [,2] [,3]
[1,]    6   14    2
[2,]   -4    2   -6
```

```
2*A-A
```

```
     [,1] [,2] [,3]
[1,]    3    7    1
[2,]   -2    1   -3
```

Solutions of Linear Systems

The most general form of a linear system is

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$

$$\ldots$$

$$\ldots$$

$$\ldots$$

$$a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n$$

In the matrix notation, we can write this as

$$A\mathbf{x} = \mathbf{b}$$

where A is an $n \times n$ matrix with entries a_{ij} , $\mathbf{b} = (b_1, \cdots, b_n)^T$ and $\mathbf{x} = (x_1, \cdots, x_n)^T$ are n-dimensional vectors.

Solutions of Linear Systems

Set up the linear system for the given.

Lourdes would like to prepare a meal using some combination of three different foods as shown below. She would like the meal to contain 16 grams of protein, 30 grams of carbohydrates, and 3 grams of fat. How many units of each food should she use so that the meal will contain the desired amounts of protein, carbohydrate, and fat?

The table shows the amount in grams of protein, carbohydrate, and fat supplied by one unit (100 grams) of three different foods.

	Food A	Food B	Food C
Protein	15	35	25
Carbohydrate	45	30	50
Fat	6	4	1

Solutions of Linear Systems

Identity Matrix

In $R^{n\times n}$, we define the identity matrix

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix}$$

as the $n \times n$-matrix containing 1 on the *diagonal* and 0 everywhere else.

Solutions of Linear Systems

Determinant

First, the matrix must be square (i.e. have the same number of rows as columns). For a 2×2 matrix (2 rows and 2 columns):

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

The determinant is:

$$det(A) = |A| = ad - bc.$$

The determinant of A equals a times d minus b times c.

Solutions of Linear Systems

Determinant

Let A be a $n \times n$ matrix (with $n \geq 2$). Denote by a_{ij} the entry of A at the intersection of the ith row and jth column. The minor of $a_{ij}(M_{ij})$ is the determinant of the sub-matrix obtained from A by deleting its ith row and its jth column.

Example

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

for $\quad a_{11} = 1$, the minor of

$$a_{11} = \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix} \rightarrow M_{11} = 7.$$

Similarly, for $a_{21} = 4$, the minor of

$$a_{21} = \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix} \rightarrow M_{21} = 3.$$

Solutions of Linear Systems

Determinant

A **cofactor** is a minor whose sign may have been changed depending on the location of the respective matrix entry.

Let A be a $n \times n$ matrix (with $n \geq 2$). Denote M_{ij} by the minor of an entry a_{ij}. The cofactor of a_{ij} is

$$C_{ij} = (-1)^{i+j} M_{ij}.$$

Laplace Expansion

For any row i, the following row expansion holds:

$$det(A) = \sum_{j=1}^{n} a_{ij} C_{ij}.$$

Similarly, for any column j, the following column expansion holds:

$$det(A) = \sum_{j=1}^{n} a_{ij} C_{ij}.$$

Solutions of Linear Systems

Example

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

$det(A) = \sum_{j=1}^{3} a_{1j} C_{1j} = 1C_{11} + 1C_{12} + 1C_{13}$

$$= 1(-1)^{1+1}(7) + 1(-1)^{1+2}(4) + 1(-1)^{1+3}(-8) = -5$$

```
A=matrix(c(1,1,1,4,3,2,0,-2,1), nrow = 3, byrow = T)
det(A)
```

```
[1] -5
```

Solutions of Linear Systems

Theorem

Let n be a positive integer, and let A be given as the general form. Then the following statements are equivalent

I. $det(A) \neq 0$

II. For each right hand side $\mathbf{b}$, the linear system has unique solution $\mathbf{x}$.

III. For $\mathbf{b} = 0$, the only solution for the system is the zero solution.

Solutions of Linear Systems

Gaussian Elimination

Let us introduce the Gaussian Elimination method for $n = 3$. The method for a general $n \times n$ system is similar.

Consider the 3×3 system

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \qquad (E1)$$

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = b_2 \qquad (E2)$$

$$a_{31}x_1 + a_{32}x_2 + a_{33}x_3 = b_3 \qquad (E3)$$

Step 1: Assume that $a_{11} \neq 0$ (otherwise interchange the row for which the coefficient of x_1 is non-zero). Let us eliminate x_1 from $(E2)$ and $(E3)$. For this define

$$m_{21} = \frac{a_{21}}{a_{11}}, m_{31} = \frac{a_{31}}{a_{11}}.$$

Solutions of Linear Systems

Gaussian Elimination

Multiply $(E1)$ with m_{21} and subtract with $(E2)$, and multiply $(E1)$ with m_{31} and subtract with $(E3)$ to give

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \qquad (E1)$$

$$a_{22}^{(2)}x_2 + a_{23}^{(2)}x_3 = b_2^{(2)} \qquad (E2)$$

$$a_{32}^{(2)}x_2 + a_{33}^{(2)}x_3 = b_3^{(2)} \qquad (E3)$$

The coefficients $a_{ij}^{(2)}$ are defined by

$$a_{ij}^{(2)} = a_{ij} - m_{i1}a_{1j}, \qquad i, j = 2, 3$$

$$b_i^{(2)} = b_i - m_{i1}b_1, \qquad i = 2, 3$$

Solutions of Linear Systems

Gaussian Elimination

Step 2: Assume that $a_{22}^{(2)} \neq 0$ and eliminate x_2 from $(E3)$. Define

$$m_{32} = \frac{a_{32}^{(2)}}{a_{22}^{(2)}}.$$

Subtract m_{32} times $(E2)$ from $(E3)$ to get

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \qquad (E1)$$

$$a_{22}^{(2)}x_2 + a_{23}^{(2)}x_3 = b_2^{(2)} \qquad (E2)$$

$$a_{33}^{(3)}x_3 = b_3^{(3)} \qquad (E3)$$

The new coefficients are defined by

$$a_{33}^{(3)} = a_{33}^{(2)} - m_{32}a_{23}^{(2)}, \quad b_3^{(3)} = b_3^{(2)} - m_{32}b_2^{(2)}.$$

Solutions of Linear Systems

Gaussian Elimination

Step 3: Using back substitution to solve successively for x_3, x_2 and x_1, we get

$$x_3 = \frac{b_3^{(3)}}{a_{33}^{(3)}}$$

$$x_2 = \frac{b_2^{(2)} - a_{23}^{(2)} x_3}{a_{22}^{(2)}}$$

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

The algorithm for $n = 3$ is easily extended to a general $n \times n$ non-singular linear system.

Gaussian elimination method is a direct method which solves the linear system exactly. However, sometime, this method fail to give the correct solution as illustrated in the example, pp28-29 of the book.

Solutions of Linear Systems

Gauss-Jordan Method

Solving systems of linear equations $A\mathbf{x} = \mathbf{b}$ will be presented here where

$$A_{nn} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix},$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, and$$

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

Solutions of Linear Systems

Gauss-Jordan Method

This method is another way of solving the system of linear equations. But before we present the Gauss-Jordan method, we begin by knowing the three elementary row operations. The following are

- Exchange any two rows.
- Multiply any row by a nonzero constant.
- Add one row to another row.

Solutions of Linear Systems

Example

Consider the system:

$$\begin{aligned} x + y + z &= 900000 \\ 4x + 3y + 2z &= 2700000 \\ 0x - 2y + z &= 0 \end{aligned}$$

An **augmented matrix** is a matrix formed by combining the columns of two matrices to form a new matrix.
Consider the matrix system below.

$$\begin{pmatrix} 1 & 1 & 1 & | & 900000 \\ 4 & 3 & 2 & | & 2700000 \\ 0 & -2 & 1 & | & 0 \end{pmatrix}$$

Solutions of Linear Systems

Example

The resulting matrix must be in the form of upper triangular matrix.

$$\left(\begin{array}{ccc|c} 1 & * & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 1 & * \end{array}\right).$$

This can be done by doing row operations.

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 900000 \\ \color{blue}{4} & 3 & 2 & 2700000 \\ 0 & -2 & 1 & 0 \end{array}\right)$$

Since row 1 and column 1 entry is 1, so we can now proceed to the next step using the row operation $-4r1 + r2 \to r2$ to reduce row 2 and column 1 to 0.

$$-4[1\ 1\ 1 \mid 900000] + [4\ 3\ 2 \mid 2700000] \to [0\ -1\ -2 \mid -900000]$$

Solutions of Linear Systems

$$\begin{pmatrix} 1 & 1 & 1 & | & 900000 \\ 4 & 3 & 2 & | & 2700000 \\ 0 & -2 & 1 & | & 0 \end{pmatrix}$$

$$-4[1\ 1\ 1 \mid 900000] + [4\ 3\ 2 \mid 2700000] \to [0\ \ -1\ \ -2 \mid\ -900000]$$

The resulting matrix is

$$\begin{pmatrix} 1 & 1 & 1 & | & 900000 \\ 0 & -1 & -2 & | & -900000 \\ 0 & -2 & 1 & | & 0 \end{pmatrix}.$$

Solutions of Linear Systems

Example

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 900000 \\ 0 & -1 & -2 & -900000 \\ 0 & -2 & 1 & 0 \end{array}\right)$$

The first entry of the last row is already 0. Row 2 and column 2 entry must be 1 so we just multiply row 2 by -1. The new matrix is

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 900000 \\ 0 & 1 & 2 & 900000 \\ 0 & -2 & 1 & 0 \end{array}\right).$$

Solutions of Linear Systems

Example

$$\begin{pmatrix} 1 & 1 & 1 & | & 900000 \\ 0 & 1 & 2 & | & 900000 \\ 0 & -2 & 1 & | & 0 \end{pmatrix}$$

We now make the column entry below row 2 and column 2 (row 3 and column 2) to be 0 by applying the row operation 2r2 + r3 $\rightarrow$ r3.

$$2[0\ 1\ 2 \mid 900000] + [0\ \ -2\ 1 \mid 0] \rightarrow [0\ 0\ 5 \mid 1800000]$$

Then we have

$$\begin{pmatrix} 1 & 1 & 1 & | & 900000 \\ 0 & 1 & 2 & | & 900000 \\ 0 & 0 & 5 & | & 1800000 \end{pmatrix}.$$

Solutions of Linear Systems

Example

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 900000 \\ 0 & 1 & 2 & 900000 \\ 0 & 0 & 5 & 1800000 \end{array}\right).$$

Dividing the last row by 5, we have

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 900000 \\ 0 & 1 & 2 & 900000 \\ 0 & 0 & 1 & 360000 \end{array}\right).$$

Solutions of Linear Systems

Example

$$\begin{pmatrix} 1 & 1 & 1 & | & 900000 \\ 0 & 1 & 2 & | & 900000 \\ 0 & 0 & 1 & | & 360000 \end{pmatrix}.$$

The last row gives $x_3 = 360000$. Doing the **back substitution**, row 2 becomes $x_2 + 2(360000) = 900000$ which results to $x_2 = 180000$. Finally for row 1, $x_1 + 180000 + 360000 = 900000$ which results to $x_1 = 360000$. It easy to check that values obtained satisfy the system.
The result in matrix form is

$$\begin{pmatrix} 1 & 0 & 0 & | & 360000 \\ 0 & 1 & 0 & | & 180000 \\ 0 & 0 & 1 & | & 360000 \end{pmatrix}.$$

Solutions of Linear Systems

Example

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 900000 \\ 4 & 3 & 2 & 2700000 \\ 0 & -2 & 1 & 0 \end{array}\right)$$

The solution to the system: $x_1 = 360000$, $x_2 = 180000$, $x_3 = 360000$.

Using R

```
A=matrix(c(1,1,1,4,3,2,0,-2,1),nrow = 3, byrow = T)
b=matrix(c(900000,2700000,0), nrow = 3)
solve(A,b,fractions = T)
```

```
       [,1]
[1,] 360000
[2,] 180000
[3,] 360000
```

Solutions of Linear Systems

LU Factorization

The solution to a system $A\mathbf{x} = \mathbf{b}$ of linear equations can be solved quickly if A can be factored as $A = LU$ where L is the lower triangular and U is the upper triangular. Then the system $A\mathbf{x} = \mathbf{b}$ can be solved in two stages as follows:

1. First solve $Ly = b$ for y by forward substitution.
2. Then solve $Ux = y$ for x by back substitution.

Then x is a solution to $A\mathbf{x} = \mathbf{b}$ because $Ax = LUx = Ly = b$.

Solutions of Linear Systems

LU Factorization

Lemma

Let A and B denote matrices.

1. If A and B are both lower (upper) triangular, the same is true of AB.
2. If A is $n \times n$ and lower (upper) triangular, then A is invertible if and only if every main diagonal entry is nonzero. In this case A^{-1} is also lower (upper) triangular.

Solutions of Linear Systems

The process of getting an **inverse** of matrix A using elementary row operations is given by
$[A|I] \to$ doing some row operation $\to [I|A^{-1}]$.

Example

$$\begin{pmatrix} 6 & 3 & | & 1 & 0 \\ 4 & 5 & | & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & | & 5/18 & -1/6 \\ 0 & 1 & | & -2/9 & 1/3 \end{pmatrix}$$

Thus,

$$A^{-1} = \begin{pmatrix} 5/18 & -1/6 \\ -2/9 & 1/3 \end{pmatrix}.$$

```r
library(matlib);library(MASS) #inv for fractions
```

```
Warning: package 'MASS' was built under R version 4.2.3
```

```r
A=matrix(c(6,3,4,5), nrow = 2, byrow = T)
fractions(inv(A))
```

```
     [,1]  [,2]
```

Solutions of Linear Systems

Example

If A and B are both lower (upper) triangular, the same is true of AB.

```r
A=matrix(c(1,1,1,0,1,2,0,0,1),nrow = 3, byrow = T)
A
```

```
     [,1] [,2] [,3]
[1,]    1    1    1
[2,]    0    1    2
[3,]    0    0    1
```

```r
A%*%A
```

```
     [,1] [,2] [,3]
[1,]    1    2    4
[2,]    0    1    4
[3,]    0    0    1
```

Solutions of Linear Systems

Example

If A is $n \times n$ and lower (upper) triangular, then A is invertible if and only if every main diagonal entry is nonzero. In this case A^{-1} is also lower (upper) triangular.

```r
library(matlib)
A=matrix(c(1,1,1,0,1,2,0,0,1),nrow = 3, byrow = T)
inv(A)
```

```
     [,1] [,2] [,3]
[1,]    1   -1    1
[2,]    0    1   -2
[3,]    0    0    1
```

Solutions of Linear Systems

LU Factorization

If A can be lower reduced to a row-echelon matrix U, then $A = LU$ where L is lower triangular and invertible and U is upper triangular and row-echelon. A factorization $A = LU$ as in theorem is called an LU-factorization of A.

LU Factorization

From the previous row reduced,

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 4 & 3 & 2 \\ 0 & -2 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & -2 & 1 \end{pmatrix} \to U_A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 5 \end{pmatrix}$$

The lower triangular matrix is

$$L_A = \begin{pmatrix} 1 & 0 & 0 \\ 4 & -1 & 0 \\ 0 & -2 & 5 \end{pmatrix}.$$

To check if the product of two matrices is equal to matrix A, we can see it below.

Solutions of Linear Systems

LU Factorization

```
library(matlib)
A=matrix(c(1,1,1,4,3,2,0,-2,1), nrow = 3, byrow = T)
L=matrix(c(1,0,0,4,-1,0,0,-2,5),nrow = 3, byrow = T)
U=matrix(c(1,1,1,0,1,2,0,0,1), nrow = 3, byrow = T)
A==(L%*%U)
```

```
     [,1] [,2] [,3]
[1,] TRUE TRUE TRUE
[2,] TRUE TRUE TRUE
[3,] TRUE TRUE TRUE
```

Solutions of Linear Systems

Using LU:

First solve $Ly = b$ for y by forward substitution.

```r
library(matlib)
A=matrix(c(1,1,1,4,3,2,0,-2,1), nrow = 3, byrow = T)
L=matrix(c(1,0,0,4,-1,0,0,-2,5),nrow = 3, byrow = T)
U=matrix(c(1,1,1,0,1,2,0,0,1), nrow = 3, byrow = T)
b=matrix(c(900000,2700000,0))
y=solve(L,b)
y
```

```
       [,1]
[1,] 900000
[2,] 900000
[3,] 360000
```

Solutions of Linear Systems

Using LU:

Then solve $Ux = y$ for x by back substitution.

```
x=solve(U,y)
x
```

```
       [,1]
[1,] 360000
[2,] 180000
[3,] 360000
```

Solutions of Linear Systems

Jacobi Method

Consider to solve an $n \times n$ size system of linear equations $A\mathbf{x} = \mathbf{b}$ with

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

and

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

for

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Solutions of Linear Systems

We split A into

$$A = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix} + \begin{pmatrix} 0 & \cdots & 0 & 0 \\ a_{21} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{n(n-1)} & 0 \end{pmatrix} +$$

$$\begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & a_{(n-1)n} \\ 0 & 0 & \cdots & 0 \end{pmatrix} = D + L + U.$$

$A\mathbf{x} = \mathbf{b}$ is transformed into $(D + L + U)\mathbf{x} = \mathbf{b}$.
$D\mathbf{x} = -(L + U)\mathbf{x} + \mathbf{b}$.

Solutions of Linear Systems

Jacobi Method

$$D\mathbf{x} = -(L+U)\mathbf{x} + \mathbf{b}$$

Assume D^{-1} exists and

$$D^{-1} = \begin{pmatrix} \frac{1}{a_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{a_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{a_{nn}} \end{pmatrix}.$$

Then

$$\mathbf{x} = -D^{-1}(L+U)\mathbf{x} + D^{-1}\mathbf{b}.$$

The matrix form of Jacobi iterative method is

$$\mathbf{x}^k = -D^{-1}(L+U)\mathbf{x}^{k-1} + D^{-1}\mathbf{b}, \quad k = 1, 2, 3, \cdots$$

Solutions of Linear Systems

Example

$$A\mathbf{x} = \mathbf{b}, \begin{pmatrix} 6 & -2 & 1 \\ -2 & 7 & 2 \\ 1 & 2 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \\ -1 \end{pmatrix}.$$

The solution to this system is:

```r
A=matrix(c(6,-2,1,-2,7,2,1,2,-5), nrow = 3, byrow = T)
b=matrix(c(11,5,-1), nrow = 3, byrow = T)
solve(A,b)
```

```
     [,1]
[1,]    2
[2,]    1
[3,]    1
```

Solutions of Linear Systems

Example

$$A\mathbf{x} = \mathbf{b}, \begin{pmatrix} 6 & -2 & 1 \\ -2 & 7 & 2 \\ 1 & 2 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \\ -1 \end{pmatrix}.$$

Let $A = L + D + U$, where

$$L = \begin{pmatrix} 0 & 0 & 0 \\ -2 & 0 & 0 \\ 1 & 2 & 0 \end{pmatrix}, D = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -5 \end{pmatrix}, U = \begin{pmatrix} 0 & -2 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Let $x^0 = (0, 0, 0)$.

$$\mathbf{x}^1 = -D^{-1}(L + U)(0, 0, 0)^T + D^{-1}\mathbf{b}$$

Solutions of Linear Systems

Example

Let $x^0 = (0, 0, 0)$.

$$\mathbf{x}^1 = -D^{-1}(L+U)(0,0,0)^T + D^{-1}\mathbf{b}$$

```r
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,6,-1),nrow = 3, byrow = T)
A==L+D+U
```

```
     [,1] [,2] [,3]
[1,] TRUE TRUE TRUE
[2,] TRUE TRUE TRUE
[3,] TRUE TRUE TRUE
```

Solutions of Linear Systems

Example

```
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,5,-1),nrow = 3, byrow = T)
x=matrix(c(2,1,1), nrow = 3, byrow = T)
x0=matrix(c(0,0,0), nrow = 3, byrow = T)
x1=(-inv(D))%*%(L+U)%*%x0+inv(D)%*%b
x1
```

```
          [,1]
[1,] 1.8333334
[2,] 0.7142857
[3,] 0.2000000
```

Solutions of Linear Systems

Example

```
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,5,-1),nrow = 3, byrow = T)
x2=(-1*inv(D))%*%(L+U)%*%x1+inv(D)%*%b
x2
```

```
         [,1]
[1,] 2.038095
[2,] 1.180952
[3,] 0.852381
```

Solutions of Linear Systems

Example

```
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,5,-1),nrow = 3, byrow = T)
x3=(-1*inv(D))%*%(L+U)%*%x2+inv(D)%*%b
x3
```

```
         [,1]
[1,] 2.084921
[2,] 1.053061
[3,] 1.080000
```

Solutions of Linear Systems

Example

```
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,5,-1),nrow = 3, byrow = T)
x4=(-1*inv(D))%*%(L+U)%*%x3+inv(D)%*%b
x4
```

```
         [,1]
[1,] 2.004354
[2,] 1.001406
[3,] 1.038209
```

Solutions of Linear Systems

Example

```
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,5,-1),nrow = 3, byrow = T)
x5=(-1*inv(D))%*%(L+U)%*%x4+inv(D)%*%b
x5
```

```
          [,1]
[1,] 1.9941006
[2,] 0.9903272
[3,] 1.0014331
```

Solutions of Linear Systems

Example

```
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,5,-1),nrow = 3, byrow = T)
x6=(-1*inv(D))%*%(L+U)%*%x5+inv(D)%*%b
x6
```

```
         [,1]
[1,] 1.996537
[2,] 0.997905
[3,] 0.994951
```

Solutions of Linear Systems

Example

```
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,5,-1),nrow = 3, byrow = T)
x7=(-1*inv(D))%*%(L+U)%*%x6+inv(D)%*%b
x8=(-1*inv(D))%*%(L+U)%*%x7+inv(D)%*%b
x8
```

```
         [,1]
[1,] 2.000406
[2,] 1.000478
[3,] 1.000210
```

Solutions of Linear Systems

Gauss-Seidel Method

For Gauss-Seidel, $A\mathbf{x} = \mathbf{b}$ can be rewritten as

$$(L + D)\mathbf{x} = -U\mathbf{x} + \mathbf{b},$$

and from this we get

$$x^{n+1} = -(L + D)^{-1}U\mathbf{x}^n + (L + D)^{-1}\mathbf{b}.$$

We will use the same example and see if it works faster.

$$A\mathbf{x} = \mathbf{b}, \begin{pmatrix} 6 & -2 & 1 \\ -2 & 7 & 2 \\ 1 & 2 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \\ -1 \end{pmatrix}.$$

Solutions of Linear Systems

Example

$$x^{n+1} = -(L+D)^{-1}U\mathbf{x}^n + (L+D)^{-1}\mathbf{b}.$$

```
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,5,-1),nrow = 3, byrow = T)
x0=matrix(c(0,0,0), nrow = 3, byrow = T)
x1=(-1*inv(L+D))%*%U%*%x0+inv(L+D)%*%b
x1
```

```
         [,1]
[1,] 1.833333
[2,] 1.238095
[3,] 1.061905
```

Solutions of Linear Systems

Example

$$x^{n+1} = -(L + D)^{-1} U\mathbf{x}^n + (L + D)^{-1}\mathbf{b}.$$

```
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,5,-1),nrow = 3, byrow = T)
x2=-1*inv(L+D)%*%U%*%x1+inv(L+D)%*%b
x2
```

```
         [,1]
[1,] 2.069048
[2,] 1.002041
[3,] 1.014626
```

Solutions of Linear Systems

Example

$$x^{n+1} = -(L+D)^{-1}U\mathbf{x}^n + (L+D)^{-1}\mathbf{b}.$$

```
L=matrix(c(0,0,0,-2,0,0,1,2,0), nrow = 3, byrow = T)
D=matrix(c(6,0,0,0,7,0,0,0,-5), nrow = 3, byrow = T)
U=matrix(c(0,-2,1,0,0,2,0,0,0), nrow = 3, byrow = T)
b=matrix(c(11,5,-1),nrow = 3, byrow = T)
x3=-inv(L+D)%*%U%*%x2+inv(L+D)%*%b
x4=-inv(L+D)%*%U%*%x3+inv(L+D)%*%b
x5=-inv(L+D)%*%U%*%x4+inv(L+D)%*%b
x5
```

```
         [,1]
[1,] 2.000119
[2,] 1.000068
[3,] 1.000051
```

Matrix Norm

A **vector norm** on R^n is a function from R^n to $[0, \infty)$ denoted by $||\cdot||$ that satisfies the following properties:
For any $\mathbf{x}$, $\mathbf{y} \in R^n$, $\alpha \in R$,

I. $||\mathbf{x}|| \geq 0$
II. $||\mathbf{x}|| = 0$ if and only if $\mathbf{x} = 0$
III. $||\alpha\mathbf{x}|| = |\alpha|||\mathbf{x}||$
IV. $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$

Solutions of Linear Systems

Some examples of vector norm are given here.

I. The **sum of magnitudes** is defined as

$$||x||_1 = \sum_{i=1}^{n} |x_i|.$$

II. The **Euclidean norm** is defined as

$$||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

III. The **maximum norm** is defined as

$$||\mathbf{x}||_\infty := max_{1\leq i\leq n}|x_i|, \mathbf{x} = (x_1, x_2, \cdots, x_n).$$

Solutions of Linear Systems

Example

Compute the 1-, 2-, and ∞-norms of the vector x, if $x = (1.25, 0.02, -5.15, 0)$.

- $||x||_1 = |1.25| + |0.02| + |-5.15| + |0| = 6.42$
- $||x||_2 = \sqrt{(1.25)^2 + (0.02)^2 + (-5.15)^2 + (0)^2} = 5.299566$
- $||x||_\infty = |-5.15| = 5.15$

Matrix Norms

$||A||_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}| =$ maximum column sum
$||A||_\infty = \max_{1 \leq j \leq n} \sum_{j=1}^{n} |a_{ij}| =$ maximum row sum
For an $m \times n$, the *Frobenius* norm is defined as

$$||A||_f = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}.$$

Solutions of Linear Systems

Example

Compute the Frobenius norms of A and B, and the ∞-norms, given that

$$A = \begin{pmatrix} 5 & 9 \\ -2 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0.1 & 0 \\ 0.2 & 0.1 \end{pmatrix}.$$

$||A||_f = \sqrt{25 + 81 + 4 + 1} = 10.54$; $||A||_\infty = 14.0$;
$||B||_f = \sqrt{0.01 + 0 + 0.04 + 0.01} = 0.2449$; $||B||_\infty = 0.3$.

Solutions of Linear Systems

Error in Solving Linear Systems

Let $\bar{x}$ be the computed solution, an approximation to the true solution. Define the *residual*, r, as $r = b - A\bar{x}$, the difference between the b-vector and what we get when the approximate $\bar{x}$ is substituted into the equations. Let e be the error in $\bar{x}$ and x be the true solution to the system (which we don't know), $e = x - \bar{x}$. Because $Ax = b$, we have

$$r = b - A\bar{x} = Ax - A\bar{x} = A(x - \bar{x}) = Ae.$$

Hence,

$$e = A^{-1}r.$$

Taking norms for a product, we write

$$||e|| = ||A^{-1}|| \cdot ||r||.$$

Solutions of Linear Systems

Example

We are given from Gauss-Seidel Method:

$$A\mathbf{x} = \mathbf{b}, \begin{pmatrix} 6 & -2 & 1 \\ -2 & 7 & 2 \\ 1 & 2 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \\ -1 \end{pmatrix},$$

whose true solution is

$$\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}.$$

The estimated value is $\bar{x} = (2.000119, 1.000068, 1.000051)^T$.

Solutions of Linear Systems

Then $r = b - A\bar{x}$ is

```r
A=matrix(c(6,-2,1,-2,7,2,1,2,-5), nrow = 3, byrow = T)
b=matrix(c(11,5,-1), nrow = 3, byrow = T)
x_bar=matrix(c(2.000119, 1.000068, 1.000051))
r=b-A%*%x_bar
r
```

```
          [,1]
[1,] -0.000629
[2,] -0.000340
[3,]  0.000000
```

Solutions of Linear Systems

We now solve $A\bar{e} = r$, again using three-digit precision, and get

```
library(matlib)
A=matrix(c(6,-2,1,-2,7,2,1,2,-5), nrow = 3, byrow = T)
b=matrix(c(11,5,-1), nrow = 3, byrow = T)
x_bar=matrix(c(2.000119, 1.000068, 1.000051))
r=b-A%*%x_bar
e_bar=inv(A)%*%r
abs(e_bar)
```

```
         [,1]
[1,] 0.000119
[2,] 0.000068
[3,] 0.000051
```

Solutions of Linear Systems

Finally, correcting $\bar{x}$ with $\bar{x} + \bar{e}$ gives almost exactly the correct solution:

```
library(matlib)
A=matrix(c(6,-2,1,-2,7,2,1,2,-5), nrow = 3, byrow = T)
b=matrix(c(11,5,-1), nrow = 3, byrow = T)
x_bar=matrix(c(2.000119, 1.000068, 1.000051))
r=b-A%*%x_bar
e_bar=inv(A)%*%r
x_barcorrected=x_bar+e_bar
x_barcorrected
```

```
     [,1]
[1,]    2
[2,]    1
[3,]    1
```

Problem Set 2

1. Use the bisection method with a hand calculator or computer to find the real root of $x^3 - x - 3 = 0$. Use an error tolerance of $\epsilon = 0.0001$. Graph the function $f(x) = x^3 - x - 3$ and label the root.
2. The function $f(x) = -3x^3 + 2e^{x^2/2} - 1$ has values of zero near $x = -0.5$ and $x = 0.5$.
   a. What is the derivative of f?
   b. If you begin Newton's method at $x = 0$, which root is reached? How many iterations to achieve an error less than 10^{-5}?
   c. Begin Newton's method at another starting point to get the other zero.
3. Use the function from no. 2 and find the root using the secant method where $x_0 = 0$ and $x_1 = 1$. Use an error tolerance of $\epsilon = 0.001$.

Problem Set 2

4. Consider the system

$$10.2x + 2.4y - 4.5z = 14.067,$$
$$-2.3x - 7.7y + 11.1z = -0.996,$$
$$-5.5x - 3.2x + 0.9z = -12.645.$$

   a. Present the augmented matrix of the system.
   b. Solve the system using $Ax = LUx = Ly = b$ and round the final answer to 4 decimal digits.
   c. Find the residual vector if the correct solution is $x = 1.4531001$, $y = -1.5891949$, $z = -0.2748947$.

5. Compute the Frobenius norm, maximum column sum, and maximum row sum of the matrix:

$$\begin{pmatrix} 10.2 & 2.4 & 4.5 \\ -2.3 & 7.7 & 11.1 \\ -5.5 & -3.2 & 0.9 \end{pmatrix}.$$

Problem Set 2

6. Solve the system of equations given in no. 4, starting with the initial vector of $[0, 0, 0]$:
   a. Solve using the Jacobi method with 2-digit precision.
   b. Solve using Gauss-Seidel method with 2-digit precision.
   c. Solve for $\bar{e}$ if the true solution is $x = (1.5, 0.33, 0.45)^T$.

References

- Atkinson, K.E. (1989). An Introduction to Numerical Analysis. John Wiley and Sons, New York.
- Gerald, C.F. and Wheatly, P.O. (2004). Applied Numerical Analysis. Pearson Education, Inc.
- Kreyszig, H. (2011). Advanced Engineering Mathematics. John Wiley & Sons, Inc.
- Sastry, S.S. (2012). Introductory Methods of Numerical Analysis. Rajkamal Electric Press
- Bloomfield, V. A (2014). Using R for Numerical Analysis in Science and Engineering. Taylor & Francis Group, LLC

Thank You!