APPENDIX M

Solution of Sets of Equations

This appendix contains a brief summary of methods of solving linear and nonlinear equations. It is only a summary; for details consult one of the numerous texts on numerical analysis that can be found in any library or one of the tutorials on the Internet.

M.1 Independent Linear Equations

1. What does the term *linear* mean?

The ease and accuracy of solving sets of equations often depends on whether the equations are linear or nonlinear. An equation is *linear* if it has **(a)** the property of proportionality and **(b)** the property of superposition.

Let $\mathcal{L}$ be a linear operator that causes a transformation of a variable x to another variable y. Then, for linearity:

a. Proportionality occurs

$$y = \mathcal{L}(kx) = k\,\mathcal{L}(x)$$

b. Superposition occurs

$$y = \mathcal{L}(x_1 + x_2) = \mathcal{L}(x_1) + \mathcal{L}(x_2)$$

A term in an equation is deemed linear if the operation on the term is linear. Following are some examples of linear and nonlinear operators:

Term	Classification	Example
$3x$	Linear	$3(x_1 + x_2) = 3x_1 + 3x_2$
Σx	Linear	$\Sigma(x_1 + x_2) = x_1 + x_2$
x^2	Nonlinear	$(x_1 + x_2)^2 \neq x_1^2 + x_2^2$
$\ln(x)$	Nonlinear	$\ln(x_1 + x_2) \neq \ln(x_1) + \ln(x_2)$

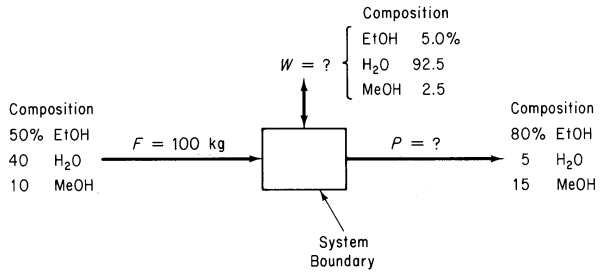


Figure M.1 Data for material balances in a process

2. What does the term *linear equation* mean?

A linear equation is an equation in which the sum of all of the terms equals zero, and each term in the equation is itself linear. For example, Figure M.1 displays a process.

The material balances for the process are

$$\begin{aligned} 0.50(100) &= 0.80(P) + 0.05(W) \\ 0.40(100) &= 0.05(P) + 0.925(W) \\ 0.10(100) &= 0.15(P) + 0.025(W) \end{aligned} \tag{M.1}$$

In general, if you write several *linear* material balances, say, m in number, they will take the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned} \tag{M.2}$$

or in compact matrix notation

$$\mathbf{Ax = b} \tag{M.2a}$$

where $x_1, x_2, \ldots, x_n$ represent the unknown variables, and a_{ii} and b_i represent the constants and known variables.

3. What does the term *set of independent equations* mean?

You know that if you add two equations together to get a third equation, the set of three equations is said to be *not independent;* they are said to be *dependent.* Only two of the equations, any two, are said to be independent because you can formulate the third equation from the other two. Formally, linear

equations are said to be independent if the sum of each equation multiplied by its related coefficient is equal to zero only if all of the coefficients are equal to zero. In symbols this means

$$c_1\,(\text{Eq. } 1) + c_2\,(\text{Eq. } 2) + \cdots + c_n\,(\text{Eq. } N) = 0 \tag{M.3}$$

For example, are the following equations independent?

$$3\,x_1 + 4\,x_2 = 0$$

$$x_2 - 3\,x_2 + 0$$

According to Equation (M.3), you inquire if

$$c_1\,(3\,x_1 + 4\,x_2) + c_2\,(x_1 - 3\,x_2) = 0 \tag{M.4}$$

for any x_i and any values of c_i other than each $c_i = 0$. The answer is no. Multiply the terms out in Equation (M.4) and rearrange to get a better perspective:

$$x_1\,(3\,c_1 + c_2) + x_2\,(4\,c_1 - 3\,c_2) = 0$$

Only $c_i = 0$ makes the equation valid for any x_i.

On the other hand, the following equations

$$y_1 = 2\,x_1 - x_2$$

$$y_2 = x_1 + 2\,x_2$$

$$y_3 = x_1 + 4\,x_2$$

are not linearly independent because

$$c_1\,(2\,x_1 - x_2) + c_2\,(x_1 + 2\,x_2) + c_3\,(x_1 + 4\,x_2) = 0$$

for at least one set of c_i, namely $c_1 = 1$, $c_2 = -9/2$, and $c_3 = 5/2$.

4. How can you tell if a set of linear equations are independent?

The preceding explanation does not provide a convenient way to determine whether a set of linear equations is composed of independent equations. Two classical ways of determining independence are:

- **a.** Determine the rank of the matrix **A** composed of the coefficients of the equations.
- **b.** Determine the order of the largest nonzero determinant that exists for **A** (which is the same as the rank).

The rank is equal to the number of independent equations. We will use the rank in what follows because most software can compute the rank. Consult references on linear algebra for method b. We advise using software to compute the rank of more than three equations to avoid error and save time.

In what follows we will employ the Gauss-Jordan elimination method of solving sets of linear equations to determine the rank of a matrix. If you use a computer program to solve equations like (M.2), you probably will not even be aware of the particular techniques that are used by the program.

The essence of the Gauss-Jordan method is to transform Equation (M.2) into Equation (M.5) by sequential nonunique elementary operations on Equation (M.2):

$$\begin{aligned} x_1 + 0 + \cdots + 0 &= b'_1 \\ 0 + x_2 + \cdots + 0 &= b'_2 \\ &\cdots \\ 0 + 0 + \cdots + x_n &= b'_n \end{aligned} \tag{M.5}$$

Equation (M.5) has a solution for $x_1, \ldots x_n$ that can be obtained by inspection, and the rank of the coefficient matrix is the number of rows that contain an x_i.

To illustrate the elementary operations that are required to execute the Gauss-Jordan method, consider the following set of three equations in three unknowns:

$$4x_1 + 2x_2 + x_3 = 15 \tag{1}$$

$$20x_1 + 5x_2 - 7x_3 = 0 \tag{2}$$

$$8x_1 - 3x_2 + 5x_3 = 24 \tag{3}$$

The augmented matrix includes the right-hand side of the equations as elements of the fourth column.

$$\left[\begin{array}{rrr|r} 4 & 2 & 1 & 15 \\ 20 & 5 & -7 & 0 \\ 8 & -3 & 5 & 24 \end{array}\right]$$

Take the a_{11} element as a pivot. To make it 1 and the other elements in the first column zero, carry out the following elementary operations shown in order for each row:

a. Subtract $\left(\frac{20}{4}\right) \times$ Equation (1) from Equation (2)

b. Subtract $\left(\frac{8}{4}\right) \times$ Equation (1) from Equation (3)

c. Multiply Equation (1) by $\left(\frac{1}{4}\right)$

to get

$$\begin{array}{c} \\ \left[\begin{array}{ccc|c} 1 & \frac{1}{2} & \frac{1}{4} & \frac{15}{4} \\ 0 & -5 & -12 & -75 \\ 0 & -7 & 3 & -6 \end{array}\right] \end{array} \quad \begin{array}{l} \textit{new Equation no.} \\ (1a) \\ (2a) \\ (3a) \end{array}$$

Carry out the following elementary operations to make the pivot element a_{22} equal to 1 and the other elements in the second column equal to zero:

d. Subtract $\left[\left(\frac{1}{2}\right)/-5\right] \times$ Equation (2a) from Equation (1a)
e. Subtract $(-7/-5) \times$ Equation (2a) from Equation (3a)
f. Multiply Equation (2a) by $(1/-5)$

to obtain

$$\begin{array}{c} \\ \left[\begin{array}{ccc|c} 1 & 0 & -\frac{19}{20} & -\frac{15}{4} \\ 0 & 1 & \frac{12}{5} & 15 \\ 0 & 0 & \frac{99}{5} & 99 \end{array}\right] \end{array} \quad \begin{array}{l} \textit{new Equation no.} \\ (1b) \\ (2b) \\ (3b) \end{array}$$

Another series of elementary operations (left for you to propose) leads to a 1 for the element a_{33} and zeros for the other two elements in the third column:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 5 \end{array}\right]$$

The solution to the original set of equations is

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 3 \\ x_3 &= 5 \end{aligned}$$

as can be observed using the augmentation column.

To obtain good accuracy and avoid numerical errors, the choice of the pivot should be made by scanning all of the eligible coefficients and choosing the one with the greatest magnitude for the next pivot. For example, you might choose a_{21} for the first pivot and then find that a_{31} would be the next pivot and finally a_{22} the last pivot to give

$$\left[\begin{array}{ccc|c} 0 & 1 & 0 & 3 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 5 \end{array}\right]$$

By exchanging rows 1 and 2, you get exactly the form of Equation (M.5). The rank of the coefficient matrix is 3. Consequently, the equations are independent.

If the coefficient matrix had been reduced to

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

then the rank of the coefficient matrix would be two, and only two of the equations would be independent.

5. How can you tell if several element balance equations are independent?

Before solving a set of element material balances, you should make sure the equations are independent. The procedure is as follows. Examine the two competing reactions:

$$CO + 2H_2 \rightarrow CH_3OH$$
$$CO + 3H_2 \rightarrow CH_4 + H_2O$$

The number of elements involved is three: H, O, and C. Prepare a species matrix in which the rows (3) are the elements and the columns (5) are the chemical species:

$$\begin{array}{c} \\ \mathrm{H} \\ \mathrm{O} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{ccccc} H_2 & H_2O & CO & CH_4 & CH_3OH \end{array} \\ \left[\begin{array}{ccccc} 2 & 2 & 0 & 4 & 4 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{array}\right] \end{array}$$

Determine the rank of the species matrix. It is 3; thus three element balances will be independent.

The number of independent element balances is not always equal to the number of elements, as shown next. Consider the reaction

$$SO_3 + H_2O \rightarrow H_2SO_4$$

Form the species matrix and determine its rank.

$$\begin{array}{c} \\ \mathrm{H} \\ \mathrm{S} \\ \mathrm{O} \end{array} \begin{array}{c} \begin{array}{ccc} H_2O & SO_3 & H_2SO_4 \end{array} \\ \left[\begin{array}{ccc} 2 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{array}\right] \end{array}$$

Make the transformation to

$$\begin{array}{c} \\ \text{H} \\ \text{S} \\ \text{O} \end{array} \begin{array}{c} \begin{array}{ccc} \text{H}_2\text{O} & \text{SO}_3 & \text{H}_2\text{SO}_4 \end{array} \\ \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

Note that the rank of the matrix is 2, not 3; hence only two independent element balances exist for independent material balances.

By inspection of the species matrix you can see that the elements in column 1 plus those in column 2 equal the elements in column 3, so the column vectors are not independent.

Note that the number of species (N) minus the number of independent element balances (the rank r of the species matrix) equals the number of independent reaction equations R:

$$N - r = R$$

In Example 7.4, $N = 7$ and $r = 4$ so $R = 7 - 4 = 3$ independent reaction equations.

6. How can you tell if several reaction equations are independent?

You can determine the number of independent reaction equations in a set by treating the equations as algebraic equations. Form the matrix of the coefficients in the equations (the coefficients on the left-hand side of the equation are defined to have negative values), and determine its rank. The rank is the number of independent reaction equations. For example, how many independent reaction equations exist in the following set?

$$\text{CH}_4 + \text{CO}_2 \rightarrow 2\text{CO} + 2\text{H}_2 \quad \text{(a)}$$

$$\text{CO} + \text{H}_2\text{O} \rightarrow \text{CO}_2 + \text{H}_2 \quad \text{(b)}$$

$$\text{CH}_4 + \text{H}_2\text{O} \rightarrow \text{CO} + 3\text{H}_2 \quad \text{(c)}$$

$$\text{CH}_4 + 2\text{H}_2\text{O} \rightarrow \text{CO}_2 + 4\text{H}_2 \quad \text{(d)}$$

The rank of the coefficient matrix

$$\begin{array}{c} \\ \text{Equation (a)} \\ \text{Equation (b)} \\ \text{Equation (c)} \\ \text{Equation (d)} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{CH}_4 & \text{CO}_2 & \text{CO} & \text{H}_2 & \text{H}_2\text{O} \end{array} \\ \begin{bmatrix} -1 & -1 & 2 & 2 & 0 \\ 0 & 1 & -1 & 1 & -1 \\ -1 & 0 & 1 & 3 & -1 \\ -1 & 1 & 0 & 4 & -2 \end{bmatrix} \end{array}$$

is just 2; hence only two of the equations are needed to form an independent set.

7. Solution of a set of linear independent equations

Look at Equation (M.2). With m equations in n unknown variables, three cases can be distinguished:

1. There is no set of xs that satisfies Equation (M.2).
2. There is a unique set of xs that satisfies Equation (M.2).
3. There is an infinite number of sets of xs that satisfy Equation (M.2).

Figure M.2 represents each of the three cases geometrically in two dimensions. Case 1 is usually termed *inconsistent,* whereas Cases 2 and 3 are consistent; but to an engineer who is interested in the solution of practical problems, Case 3 is as unsatisfying as Case 1. Hence Case 2 will be termed *determinate,* and Case 3 will be termed *indeterminate.*

To ensure that a system of equations represented by (M.2) has a unique solution, it is necessary to first show that (M.2) is *consistent,* that is, that the coefficient matrix **A** and the augmented matrix **[A, b]** must have the same rank r. Then, if $n = r$, the system (M.2) is determinate, whereas if $r < n$, as may be the case, then the number $(n - r)$ of variables must be specified in some manner or determined by optimization procedures. If the equations are *independent,* $m = r$. In words, the number of independent equations in a set of equations equals the rank of the coefficient matrix of the equations. We use integers for simplicity as elements of the matrices in the following examples. Noninteger elements may lead to less clear results.

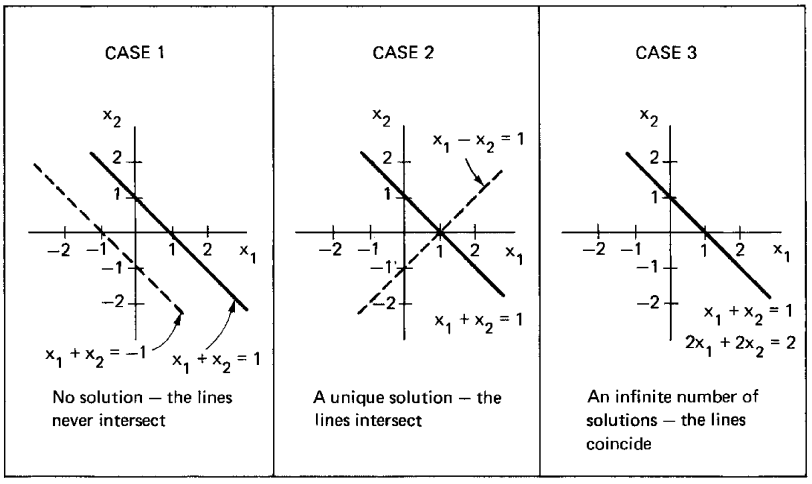


Figure M.2 Types of solutions of linear equations

Let r = number of ***independent*** equations (the rank of the coefficient matrix)
n = number of variables whose values are unknown.

Homogenous Equations (all the right-hand coefficients are zero) Ax = 0

$\boxed{r = n}$ The only solution is $x_1 = x_2 = \ldots = 0$ (a unique solution but a trivial solution).

Example:

$$\left.\begin{aligned} x_1 + x_2 - x_3 &= 0 \\ 2x_1 + 4x_2 - x_3 &= 0 \\ 3x_1 + 2x_2 + 2x_3 &= 0 \end{aligned}\right\} \text{the only solution is } x_1 = x_2 = x_3 = 0$$

Using a computer code, you can find that the rank of the coefficient matrix is 3.

$\boxed{r < n}$ Multiple nonunique solutions exist.

Example 1:

$$\left.\begin{aligned} x_1 + x_2 - x_3 &= 0 \\ x_1 - 5x_2 - 3x_3 &= 0 \end{aligned}\right\} \text{You can combine to get } 3x_2 + 2x_3 = 0, \text{ which has multiple solutions.}$$

The rank of the 2 by 3 coefficient matrix is 2, but $n = 3$; hence multiple solutions exist.

The rank of the augmented matrix is the same as the rank of **A**.

Example 2:

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

The rank of the coefficient matrix is 2, which is the same as the rank of the augmented matrix, but n still equals 3. Thus, multiple solutions exist.

$\boxed{r > n}$ Only the trivial solution exists.

$$\left.\begin{aligned} x + y &= 0 \\ x + 4y &= 0 \\ 3x + 2y &= 0 \end{aligned}\right\} \text{The only solution is the trivial solution } x = y = 0.$$

The rank of $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 4 \\ 3 & 2 \end{bmatrix}$ is 2 and $n = 2$. The rank of the augmented matrix is also 2. A unique solution exists, but it is the trivial solution.

Nonhomogenous Equations Ax = b

$\boxed{r = n}$ A unique solution exists.

Example 1:

$$\left.\begin{aligned} 2x_1 + x_2 - 2x_3 &= 0 \\ 3x_1 + 2x_2 + 2x_3 &= 1 \\ 5x_1 + 4x_2 + 3x_2 &= 4 \end{aligned}\right\}$$ 3 equations, 3 variables, hence a solution exists: $x_1 = 1, x_2 = 3, x_3 = -3$

The rank of the coefficient matrix = 3; the augmented matrix has the same rank.

Example 2:

$$\left.\begin{aligned} x_1 + 2x_2 - 3x_3 &= 4 \\ x_1 + 3x_2 + x_3 &= 11 \\ 2x_1 + 5x_2 - 4x_3 &= 13 \\ 2x_1 + 6x_2 + 2x_2 &= 22 \end{aligned}\right\}$$ 4 equations, 3 variables

The rank of the coefficient matrix is 3 (the maximum would be 3), and the rank of the augmented matrix is 3 (the maximum would be 4), so a unique solution exists ($x_1 = 1, x_2 = 3, x_3 = 1$). Note that only three of the equations are independent.

Example 3:

$$\left.\begin{aligned} x_1 + 2x_2 - 3x_3 &= 1 \\ 3x_1 - x_2 + 2x_3 &= 7 \\ 5x_1 + 3x_2 - 4x_3 &= 2 \end{aligned}\right\}$$ 3 equations, 3 variables

Although superficially it appears that $n = r$, such is not the case because the rank of the coefficient matrix is 2. Is the rank of the augmented matrix

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

equal to 2 or to 3? The rank is 3, and thus there is no unique solution. This is really a case of $r < n$; one equation is not independent.

$\boxed{r < n}$ Multiple solutions exist.

Example 1:

$$x_1 + 2x_2 + x_3 = 3$$
$$2x_1 + x_2 - x_3 = 1$$

The ranks of the coefficient matrix and the augmented matrix are both equal to 2, and $n = 3$. Multiple solutions exist.

Example 2:

$$\left.\begin{aligned} -2x_1 + 5x_2 + 7x_3 &= 6 \\ -x_1 + x_2 + 2x_3 &= 1 \\ x_1 + 2x_2 + x_3 &= 3 \end{aligned}\right\} \text{ 3 equations, 3 variables}$$

The rank of the coefficient matrix is 2.

The rank of the augmented matrix is also 2. Thus, $r = 2$ and $n = 3$. Multiple solutions exist.

$\boxed{r > n}$ No solution exists.

Example 1:

$$\left.\begin{aligned} x_1 + 2x_2 &= 4 \\ x_1 + 3x_2 &= 1 \\ 2x_1 + 5x_2 &= 3 \end{aligned}\right\} \text{ 3 (inconsistent) equations, 2 variables}$$

The rank of the coefficient matrix is 2, the rank of the augmented matrix is 3, and no solution exists.

8. How to determine the number of independent components in the Gibbs phase rule

In Chapter 8, Equation (8.2), the phase rule was stated as

$$\mathcal{F} = 2 - \mathcal{P} + \mathrm{C}$$

where $\mathcal{F}$ = number of degrees of freedom
$\mathcal{P}$ = number of phases that exist at equilibrium
C = number of independent chemical species in the system

Without reaction in the system, the number of species is equal to C. With reaction in the system, the value for the number of species may have to be reduced because C is equal to the rank of the species matrix. For example, consider a system with CO, CO_2, H_2, H_2O, and CH_4 for which the number of components was 5. The species matrix is

	CO_2	CO	H_2	H_2O	H_4
C	1	1	0	0	1
H	0	0	2	2	4
O	1	2	0	1	0

As another example, for ZnO, C, CO, and Zn, the species matrix is 3, hence C = 3.

For Part b of Example 19.2, the number of species was 4. The species matrix is

	ZnO	C	CO	Zn
Zn	1	0	0	1
O	1	0	1	0
C	0	1	1	0

and the rank of the matrix is 3, hence C = 3.

M.2 Nonlinear Independent Equations

The precise criteria used to ascertain if a linear system of equations is determinate cannot be neatly extended to nonlinear systems of equations. Furthermore, the solution of sets of nonlinear equations requires the use of computer codes that may fail to solve your problem for one or more of a variety of reasons, a few of which are mentioned below. The problem to be solved can be written as

$$\left.\begin{aligned} f_1(x_1, \ldots, x_n) &= 0 \\ f_2(x_1, \ldots, x_n) &= 0 \\ &\cdots \\ f_n(x_1, \ldots, x_n) &= 0 \end{aligned}\right\} \mathbf{f}(\mathbf{x}) = \mathbf{0} \tag{M.6}$$

where each function $f_i(x_1, \ldots, x_n)$ corresponds to a nonlinear function containing one or more of the variables whose values are unknown.

Figure M.3 classifies the major general methods of solving systems of nonlinear equations. Within each category and as combinations of categories you can find innumerable variations and submethods in the literature and available as computer codes.

Newton's Method

Refer to Equation (M.6). For a single equation (and variable), $f(x) = 0$, Newton's method uses the expansion of $f(x)$ in a first-order Taylor series about a reference point (a starting guess for the solution) x_0.

AppM.qxd 3/12/12 6:54 PM Page 914

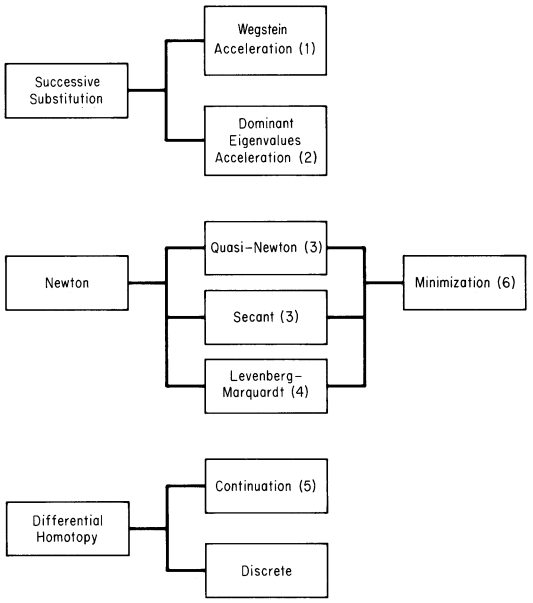


Figure M.3 General categories of techniques to solve nonlinear equations. From 1. J. H. Wegstein, *Commu. ACM.*, **1**, 9 (1958); 2. O. Orbach and C. M. Crowe, *Can. J. Chem. Eng.*, **49**, 509 (1971); 3. J. E. Dennis and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, NJ (1983); 4. D. Marquardt, *SIAM J Appld. Math.*, **11**, 431 (1963); 5. D. Davidenko, *Ukrain Math.*, **5**, 196 (1953); 6. T. F. Edgar, D. M. Himmelblau, and L. Lasdon, *Optimization of Chemical Processes*, 2d ed, McGraw-Hill, New York (2002).

$$f(x) \simeq f(x_0) + \frac{\mathrm{d}f(x_0)}{\mathrm{d}x}(x - x_0) \tag{M.7}$$

Note that Equation (M.7) is a *linear equation* that is tangent to $f(x)$ at x_0. Examine Figure M.4. The right-hand side of Equation (M.7) is equated to zero and the resulting equation solved for $(x - x_0)$.

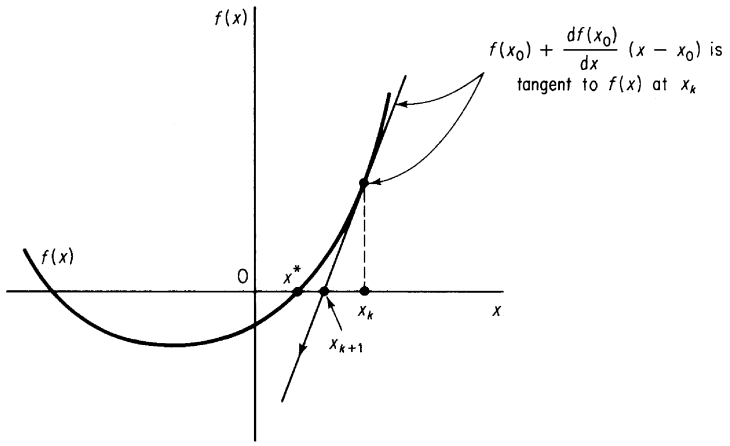


Figure M.4 Newton's method applied to the solution of $f(x) = 0$ starting at x_k; x_{k+1} is the next reference point for linearization. x^* is the solution.

$$x - x_0 = -\frac{f(x_0)}{\mathrm{d}f(x_0)/\mathrm{d}x} \tag{M.8}$$

For example, suppose that $f(x) = 4x^3 - 1 = 0$; hence $\mathrm{d}f(x)/\mathrm{d}x = 12x^2$. The sequence of steps to apply Newton's method using Equation (M.8) starting at $x_0 = 3$ would be

$$\begin{aligned} x_1 &= x_0 - \frac{4x^3 - 1}{12x^2} \\ &= 3 - \frac{107}{108} = 2.009259 \\ x_2 &= 2.00926 - \frac{31.4465}{48.4454} = 1.36015 \end{aligned}$$

Additional iterations yield the following values for x_k:

k	x_k
0	3.00000
1	2.009259
2	1.3601480
3	0.9518103
4	0.7265254
5	0.6422266
6	0.6301933
7	0.6299606
8	0.6299605
9	0.6299605

so that the solution is given with increasing precision as k increases.

Suppose that two independent equations in two variables whose values are to be determined are

$$\begin{aligned} f_1(x_1, x_2) &= 0 \\ f_2(x_1, x_2) &= 0 \end{aligned} \tag{M.9}$$

To apply Newton's method, expand each equation as a first-order Taylor series to get a set of linear equations at the point (x_{10}, x_{20}).

$$\begin{aligned} f_1(x_1, x_2) &\simeq f_1(x_{10}, x_{20}) + \frac{\partial f_1(x_{10}, x_{20})}{\partial x_1}(x_1 - x_{10}) + \frac{\partial f_1(x_{10}, x_{20})}{\partial x_2}(x_2 - x_{20}) \\ f_2(x_1, x_2) &\simeq f_2(x_{10}, x_{20}) + \frac{\partial f_2(x_{10}, x_{20})}{\partial x_1}(x_1 - x_{10}) + \frac{\partial f_2(x_{10}, x_{20})}{\partial x_2}(x_2 - x_{20}) \end{aligned} \tag{M.10}$$

Let the partial derivatives be designated by the constants $a_{ij} = \partial f_i / \partial x_j$ to simplify the notation, and let $(x_i - x_{i0}) = \Delta x_i$. Then, after equating the right-hand side of Equations (M.10) to zero, they become

$$\begin{aligned} a_{11}\,\Delta x_1 + a_{12}\,\Delta x_2 &= -f_1(x_{10}, x_{20}) \\ a_{21}\,\Delta x_1 + a_{22}\,\Delta x_2 &= -f_2(x_{10}, x_{20}) \end{aligned} \tag{M.11}$$

These equations are linear and can be solved by a linear equation solver to get the next reference point (x_{11}, x_{21}). Iteration is continued until a solution of satisfactory precision is reached. Of course, a solution may not be reached, as illustrated in Figure M.5c, or may not be reached because of round-off or truncation errors.

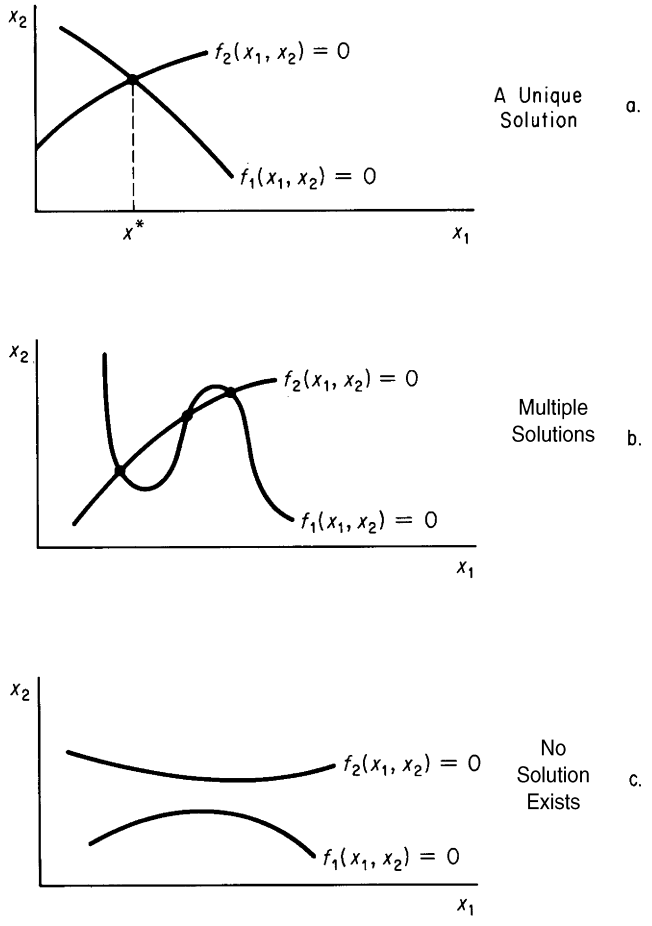


Figure M.5 Possible cases for the solution of two independent nonlinear equations $f_1(x_1, x_2) = 0$ and $f_2(x_1, x_2) = 0$ in two unknown variables

If the Jacobian matrix [see Equation (M.12) below] is singular, the linearized equations may have no solution or a whole family of solutions, and Newton's method probably will fail to obtain a solution. It is quite common for the Jacobian matrix to become ill conditioned because if x_0 is far from the solution or the nonlinear equations are badly scaled, the correct solution will not be obtained.

The analog of (M.11) in matrix notation is

$$\mathbf{J}_k(\mathbf{x} - \mathbf{x}_k) = -\mathbf{f}(\mathbf{x}_k) \tag{M.12}$$

where **J** is the Jacobian matrix (the matrix whose elements are composed of the first partial derivatives of the equations with respect to the variables). For two equations

$$\mathbf{J} = \begin{bmatrix} \partial f_1(\mathbf{x})/\partial x_1 & \partial f_1(\mathbf{x})/\partial x_2 \\ \partial f_2(\mathbf{x})/\partial x_1 & \partial f_2(\mathbf{x})/\partial x_2 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} \qquad \mathbf{f} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

Information concerning the other methods of solving nonlinear equations shown in Figure M.3 can be later in this appendix.

Quasi-Newton Methods

A quasi-Newton method in general is one that imitates Newton's method. If $f(x)$ is not given by a formula, or the formula is so complicated that analytical derivatives cannot be formulated, you can replace df/dx in Eq. (L.7) with a finite difference approximation

$$x_{k+1} = x_k - \frac{f(x)}{[f(x+h) - f(x-h)]/2h} \tag{M.12}$$

A central difference has been used in Eq. (L.12) but forward differences or any other difference scheme would suffice *as long as the step size* h *is selected to match the difference formula and the computer (machine) precision* for the computer on which the calculations are to be executed.

Other than the problem of the selection of the value of h, the only additional disadvantage of a quasi-Newton method is that additional function evaluations are needed on each iteration k. Eq. (L.12) can be applied to sets of equations if the partial derivatives are replaced by finite difference approximations.

Secant Methods

In the secant method the approximate model analogous to the right hand side of Eq. (L.6) (equated to zero) is

$$f(x_k) + m(x - x_k) = 0 \tag{M.13}$$

where m is the slope of the line connecting the point x_k and a second point x_q, given by

$$m = \frac{f(x_q) - f(x_k)}{x_q - x_k}$$

Thus the secant method imitates Newton's method (in this sense the secant method is also a quasi-Newton method (see Figure L.6)

Secant methods start out by using two points x_k and x_q spanning the interval of x, points at which the values of $f(x)$ are of opposite sign. The zero of $f(x)$ is predicted by

$$\tilde{x} = x_q - \frac{f(x_q)}{\dfrac{f(x_q) - f(x_k)}{x_q - x_k}} \tag{M.14}$$

The two points retained for the next step are $\tilde{x}$ and either x_q or x_k, the choice being made so that the pair of values $f(\tilde{x})$, and either $f(x_k)$ or $f(x_q)$, have opposite signs to maintain the bracket on x^*. (This variation is called "regula falsi" or the method of false position.) In Figure M.6, for the $(k + 1)$st stage, $\tilde{x}$ and $x_q$would be selected as the end points of the secant line. Secant methods may seem crude, but they work well in practice. The details of the computational aspects of a sound algorithm to solve multiple equations by the secant method are too lengthy to outline here (particularly the calculation of a new Jacobian matrix from the former one; instead refer to Dennis and Schnabel[1]).

The application of Eq. (M.14) yields the following results for $f(x) = 4x^3 - 1 = 0$ starting at $x_k = -3$ and $x_q = 3$. Some of the values of $f(x)$ and x during the search are shown below; note that x_q remains unchanged in order to maintain the bracket with $f(x) > 0$.

k	x_q	x_k	$f(x_k)$
0	3	−3	−109.0000
1	3	0.0277778	− 0.9991
2	3	0.055296	− 0.9992
3	3	0.0825434	− 0.9977
4	3	0.1094966	− 0.9899
5	3	0.1361213	− 0.9899
20	3	0.4593212	− 0.6124
50	3	0.6223007	− 0.0360
100	3	0.6299311	$-\ 1.399 \times 10^{-4}$
132	3	0.6299597	$-\ 3.952 \times 10^{-6}$

Brent's and Brown's Methods

Brent's[2] and Brown's[3] methods are variations of Newton's method that improve convergence. The calculation of the elements in $\mathbf{J}_k$ in Eq. (M.11) and the solving of the linear equations are intermingled. Each row of $\mathbf{J}_k$ is obtained as needed

[1]Dennis, J. E. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Appendix A) Englewood Cliffs, NJ: Prentice-Hall, 1983.

[2]Brent, R. P. *SIAM J. Num. Anal.*, *10* (1973): 327.

[3]Brown, K. M., PhD Dissertation, Purdue University, 1966.

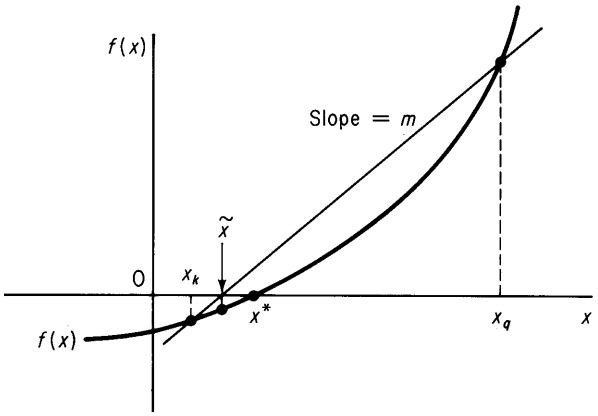


Figure M.6 Secant Method for the solution of $f(x) = 0$. x^* is the solution, $\tilde{x}$ the approximate to x^*, and x_q and x_k the starting points for iteration k of the secant method.

using the latest information available. Then one more step in the solution of the linear equations is executed. Brown's method is an extension of Gaussian elimination; Brent's method is an extension of QR factorization. Computer codes are generally implemented by using numerical approximations for the partial derivatives in $\mathbf{J}_k$.

Powell Hybrid and Levenberg-Marquardt Methods

Powell[4] and Levenberg[5]-Marquardt[6] calculated a new point $\mathbf{x}^{(k+1)}$ from the old one $\mathbf{x}^{(k)}$ by (note in the next two equations the superscript (k) is used instead of a subscript k to denote the stage of iteration so that the notation is less confusing)

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta x^{(k)} \tag{M.15}$$

where $\Delta x^{(k)}$ was obtained by solving the set of linear equations

$$\sum_{j=1}^{n} [\mu^{(k)} I_{ij} + \sum_{t=1}^{n} J_{ti}^{(k)} J_{tj}^{(k)}] \Delta x_j^{(k)} = -\sum_{t=1}^{n} J_{ti}^{(k)} f_t(x^{(k)}) \qquad i = 1, \ldots, n \tag{M.16}$$

where I_{ij} is an element of the unit matrix $\mathbf{I}$ and $\mu^{(k)}$ is a non-negative parameter whose value is chosen to reduce the sum of squares of the deviations (f_i-0) on each stage of the calculations. Powell used numerical approximations for the elements of $\mathbf{J}$. In matrix notation Eq. (L.16) can be derived by premultiplying Eq. (M.11) by $\mathbf{J}_k^T$.

[4]Powell, M.J.D. In *Numerical Methods for Nonlinear Algebraic Equations*, ed. P. Rabinowitz, Chapt. 6. New York: Gordon Breach, 1970.

[5]Levenberg, K. *Quart. Appld. Math.*, *2* (1944): 164.

[6]Marquardt, D. W. *J. SIAM*, *11* (1963): 431.

$$\mathbf{J}_k^T \mathbf{J}_k \Delta \mathbf{x}_k = -\mathbf{J}_k^T f(\mathbf{x}_k)$$

and adding a weighting factor $\mu_k \mathbf{I}$ to the left hand side to insure $\mathbf{J}_k^T \mathbf{J}$ is positive definite.

Minimization Methods

The solution of a set of nonlinear equations can be accomplished by minimizing the sum of squares of the deviations between the function values and zero.

$$\text{Minimize } F = \sum_{i=1}^{n} f_i^2(\mathbf{x}) \qquad i = 1, \ldots, m$$

Edgar and Himmelblau[7] list a number of codes to minimize F, including codes that enable you to place constraints on the variables.

Method of Successive Substitutions

Successive substitution (or resubstitution) starts by solving each equation $f_i(x)$ for a single (different) output variable. For example, for three equations, you solve for an output variable (f_2 for x_1, f_3 for x_3, and f_1 for x_2)

$$f_1(x) = 3x_1 + x_2 + 2x_3 - 3 = 0$$

$$f_2(x) = -3x_1 + 5x_2^2 + 2x_1x_3 - 1 = 0$$

$$f_3(x) = 25x_1x_2 + 20x_3 + 12 = 0$$

and rearrange them as follows

$$\left.\begin{aligned} x_1 &= F_1(x) = \frac{5x_2^2}{3} + \frac{2x_3x_1}{3} - \tfrac{1}{3} \\ x_2 &= F_2(x) = -3x_1 - 2x_3 + 3 \\ x_3 &= F_3(x) = -\frac{25x_1x_2}{20} - \frac{12}{20} \end{aligned}\right\} \mathbf{x} = \mathbf{F(x)} \qquad \text{(M.17)}$$

An initial vector (x_{10}, x_{20}, x_{30}) is guessed and then introduced into the right-hand side of (L.17) to get the next vector (x_{11}, x_{21}, x_{31}), which is in turn introduced into the right-hand side and so on. In matrix notation the iteration from k to k+1 is

$$\mathbf{x}_{k+1} = \mathbf{F}(\mathbf{x}_k) \qquad \text{(M.18)}$$

[7]Edgar, T. F., and D. M. Himmelblau. *Optimization of Chemical Processes*. (Chapters 6 and 8). New York: McGraw-Hill, 1988.

For the procedure of successive substitution to be guaranteed to converge, the value of the largest absolute eigenvalue of the Jacobian matrix of **F(x)** evaluated at each iteration point must be less than (or equal to) one. If more than one solution exists for Eqs. (M.17), the starting vector and the selection of the variable to solve for in an equation controls the solution located. Also, different arrangements of the equations and different selection of the variable to solve for may yield different convergence results.

The Wegstein and Dominant Eigenvalue methods listed in Figure M.3 are useful techniques to speed up convergence (or avoid nonconvergence) of the method of successive substitutions. Consult the references cited in Figure M.3 for the specific details.

Wegstein's method, which is used in many flowsheeting codes, accelerates the convergence of the method of successive substitutions on each iteration. In the secant method, the approximate slope is

$$m = \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}},$$

where x_k is the value of x on the kth (current) iteration and x_{k-1} is the value of x on the $(k - 1)$st (previous) iteration. The equation of a line through x_k and $f(x_k)$ with slope m is

$$f(x) - f(x_k) = m(x - x_k).$$

In successive substitutions we solve $x = f(x)$. Introduce $x = f(x)$ into the equation for the line

$$x - f(x_k) = m(x - x_k)$$

and solve for x:

$$x = \frac{1}{1-m} f(x_k) - \frac{m}{1-m} x_k.$$

Let $t = \dfrac{1}{1-m}$. Then for the $(k+1)st$ (next) iteration,

$$x_{k+1} = (1 - t)x_k + tf(x_k).$$

For the solution of several equations simultaneously, each x is treated independently, a procedure that may possibly cause some instability if the x's interact. In such cases, an upper limit should be placed on t, say $0 \le t \le 1$. The Wegstein algorithm is for stage k.

1. Calculate x_k from the previous stage.
2. Evaluate $f(x_k)$.

3. Calculate m and t.
4. Calculate x_{k+1}.
5. Set $x_k \rightarrow x_{k+1}$ and repeat the above, starting with 2.
6. Terminate when $x_{k+1} - x_k <$ tolerance assigned.

Homotopy (Continuation) Methods

Homotopy methods[8] can be viewed as methods that widen the domain of convergence of any particular method of solving nonlinear equations, or, alternatively, as a method of obtaining starting guesses satisfactorily close to the desired solution. A set of functions $\mathbf{F(x)}$ is modified as follows to be a linear combination of a parameter t:

$$\mathbf{F}(\mathbf{x},t) = \mathbf{F(x)} + (1-t)\mathbf{F}(\mathbf{x}_0) = \mathbf{0} \qquad 0 \le t \le 1 \tag{M.19}$$

where t is a scalar parameter such that when t is a fixed number the trajectory $\mathbf{F}(\mathbf{x},t) = 0$ occurs [a mapping of $\mathbf{F(x)}$] and when $t = 1$ the trajectory of the set of equations reaches the desired solution $\mathbf{F}(\mathbf{x}^*) = 0$. With this definition $\mathbf{x}$ describes a curve in space (a continuous mapping called a homotopy) with one end point at a known value (the starting guess) for $\mathbf{x}$, namely $\mathbf{x}_0$, and the other end point at a solution of $\mathbf{F(x) = 0}$, $\mathbf{x}^*$.

Lin et al. outline the procedure, which is first to determine $\mathbf{x}$ and t as functions of the arc length of the homotopy trajectory. Then Eq. (M.19) is differentiated with respect to the arc length to yield an initial value problem in ordinary differential equations. Starting at $\mathbf{x}_0$ and t_0, the initial value problem is transformed by using Euler's method to a set of linear algebraic equations that yield the next step in the trajectory. The trajectory may reach some or all of the solutions of $\mathbf{F(x) = 0}$; hence several starting points may have to be selected to create paths to all the solutions, and many undesired solutions (from a physical viewpoint) will be obtained. A number of practical matters to make the technique work can be found in the review by Seydel and Hlavacek.[9]

SUPPLEMENTARY REFERENCE

KELLEY, C.T. *Iterative Methods for Linear and Nonlinear Equations*. SIAM, Philadelphia, 1995.

RHEINBOLT, W. C. *Numerical Analysis of Parameterized Nonlinear Equations*. New York, Wiley-Interscience, 1986.

SIKORSKI, K.A. *Optimal Solution of Nonlinear Equations*. Oxford Univ. Press, Oxford, U.K., 2000.

[8]M. Kubicek. "Algorithm 502." *ACM Trans. Math. Software*, *2* (1976):98; W. J. Lin, J. D. Seader, and T. L. Wayburn. *AIChE J.*, *33* (1987):33.

[9]R. Seydel, and V. Hlavacek. *Chem. Eng. Sci.*, *42* (1987): 1281.