Appendix E

E.1 FITTING EQUATIONS TO DATA

Engineers, technologists, and scientists frequently collect paired data in order to understand the behavior of the system. Numerical methods of analysis have been employed to solve a wide range of steady and transient state problems. The fundamentals are essential in the basic operations of curve fitting, approximation, interpolation, numerical solutions of simultaneous linear and non-linear equations, numerical differentiation, and integration. These requirements are greater when new processes are designed.

Various types of software packages are now readily available for scientists and engineers. They must fit a function or functions to measure data that fluctuate, which result from random error of measurement. A Fortran computer program (PROG1) was developed which determines the coefficients that provide the best fit for the following equations:

$$Y = a + bX \tag{E-1}$$

$$Y = a + bX^2 \tag{E-2}$$

$$Y = a + b/X \tag{E-3}$$

$$Y = a + bX^{0.5} \tag{E-4}$$

$$Y = aX^b \tag{E-5}$$

$$Y = ae^{bX} \tag{E-6}$$

$$Y = a + b/X \tag{E-7}$$

$$Y = a + be^X \tag{E-8}$$

The non-linear Eqs. (E-5), (E-6), and (E-8) can be transformed by linearizing as follows:

$$Y = aX^b \quad \ln Y = \ln a + b \ln X \tag{E-9}$$

$$Y = ae^{bX} \quad \ln Y = \ln a + bX \tag{E-10}$$

$$Y = a + be^X \quad \ln Y = \ln a + (\ln b)\ X \tag{E-11}$$

E.2 LINEAR REGRESSION ANALYSIS

Regression analysis uses statistical and mathematical methods to analyze experimental data and to fit mathematical models to these data. The least squares provide the best method for objectively determining the best straight line through a series of points. The method assumes that all deviations from the line are the result of error in the measurement of the dependent variable. The method of least squares yields the parameters which minimize the sum of squares of the residuals (e.g., the deviation of each measurement of the dependent variable from its calculated value). If $\hat{Y}$ is the calculated value and Y is the original value of the dependent variable, then the residual sum of the squares can be minimized by the following expression.

$$\text{SRS} = \sum_{i=1}^{n} r_i^{\,2} = \sum_{i=1}^{n} \varsigma_i^{\,2} = \text{minimum} \tag{E-12}$$

where n = number of observations of X–Y data.

$$\hat{Y}_i = a + bX_i \tag{E-13}$$

$$r_i = a + bX_i - Y_i \tag{E-14}$$

and

$$\text{SRS} = \sum_{i=1}^{n} r_i^{\,2} = \sum_{i=1}^{n} (a + bX_i - Y_i)^2 \tag{E-15}$$

The values of a and b are found by minimizing Eq. (E-15). This involves taking the partial derivative of the equation with respect to each variable a and b, and setting the result to zero.

$$\frac{\partial \sum r_i^{\,2}}{\partial a} = 0 \text{ and } \frac{\partial \sum r_i^{\,2}}{\partial b} = 0 \tag{E-16}$$

Substituting Eq. (E-15) into Eq. (E-16) gives

$$\frac{\partial \sum (a + bX_i - Y_i)^2}{\partial a} = 0 \tag{E-17}$$

and

$$\frac{\partial \sum (a + bX_i - Y_i)^2}{\partial b} = 0 \tag{E-18}$$

This is equivalent to

$$\frac{2 \sum (a + bX_i - Y_i)\ \partial \sum (a + bX_i - Y_i)}{\partial a} = 0 \tag{E-19}$$

and

$$\frac{2X_i \sum (a + bX_i - Y_i)\ \partial \sum (a + bX_i - Y_i)}{\partial b} = 0 \tag{E-20}$$

Since b, X, and Y are not functions of a, and the partial derivative of a with respect to itself is unity, Eq. (E-19) reduces to

$$\sum a + \sum bX_i = \sum Y_i \tag{E-21}$$

Similarly, a, X, and Y are not functions of b. Therefore, Eq. (E-20) becomes

$$\sum aX_i + \sum bX_i^{\,2} = \sum X_i Y_i \tag{E-22}$$

where a and b are constants. Equations (E-17) and (E-18) are expressed as:

$$an + b\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} Y_i \tag{E-23}$$

and

$$a\sum_{i=1}^{n} X_i + b\sum_{i=1}^{n} X_i^{\,2} = \sum_{i=1}^{n} X_i Y_i \tag{E-24}$$

Equations (E-23) and (E-24) apply only when fitting a straight line to a set of X–Y data points. Both equations are linear in X, Y, and n and the unknowns a and b. Using Cramer's rule for the simultaneous equations, we have

$$a = \frac{\begin{vmatrix} \sum Y_i & \sum X_i \\ \sum X_i Y_i & \sum X_i^{\,2} \end{vmatrix}}{\begin{vmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^{\,2} \end{vmatrix}} \tag{E-25}$$

and

$$b = \frac{\begin{vmatrix} n & \sum Y_i \\ \sum X_i & \sum X_i Y_i \end{vmatrix}}{\begin{vmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{vmatrix}} \tag{E-26}$$

Solving Eqs. (E-25) and (E-26) gives

$$a = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - \sum X_i \sum X_i} \tag{E-27}$$

and

$$b = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - \sum X_i \sum X_i} \tag{E-28}$$

respectively, where all the sums are taken over all experimental observations. The developed computer program PROG1 computes the constants a and b for Eqs. (E-1)–(E-8). The program evaluates the correlation coefficient, which gives a measure of the accuracy of fit. Alternatively, a table with column heading as X_i, Y_i, X_i^2, and $X_i Y_i$ can be constructed using the Microsoft Excel spreadsheet; the constants a and b can be evaluated from Eqs. (E-27) and (E-28) respectively. The Excel spreadsheet also employs a *trendline* through the data after being plotted into a worksheet. The following are the steps being employed to fit a straight line to a set of data in the Excel spreadsheet.

1. Open a new worksheet and enter the X-data (the independent variable) in the leftmost column.
2. Enter the Y-data (the dependent variable) in the next column.
3. Plot the data as a line graph (i.e., an X–Y chart) with arithmetic coordinates. *Do not interconnect the individual data points.*
4. Activate the graph for editing. Then click on one of the plotted data points, thus selecting the data set as the active editing object (the data points will appear highlighted when this step is carried out correctly).
5. Choose *Trendline* from the Insert menu. They specify the type of curve, and request any pertinent options when the *Trendline* dialog box appears. These are, displaying the intercept, equation of the curve and its associated correlation coefficient (r^2) squared. Extrapolating the curve fit forward, that is beyond the rightmost data point or backward beyond the leftmost data point, is another option to explore. (Generally the equation of the curve and its associated r^2 value are requested for display on the graph.)
6. Press the OK button. The curve fitting will then be carried out and the results are displayed automatically.

Commercial curve fitting packages are now available in the market and details can be retrieved from the world wide web (www) using the Google search engine by typing *linear regression*. Typical websites are:

- www.ebicom.net/~dhyams
- www.analyse-it.com
- www.wessa.net.

Table E-1 shows the computer results of Example E-1.

Example E-1

The first-order rate constant k for the rotation about the C-N bond in n, n-dimethylnicotinamide measured at different temperatures by nuclear magnetic resonance are

T(°C)	10.0	15.7	21.5	27.5	33.2	38.5	45.7
k(s^{-1})	2.08	4.57	8.24	15.8	28.4	46.1	93.5

(*Source*: B.G. Cox, Modern Liquid Phase Kinetics, Oxford Chemistry Primers, ZENECA, Oxford University Press, 1994.)

Determine the activation energy E and the pre-exponential factor k_o for the rotation.

Solution

The Arrhenius equation is expressed as:

$$k = k_o e^{-E/RT}$$

which can be linearized in the form of model Eq. (E-1) as

$$\ln k = \ln k_o - \frac{E}{RT}$$

or used in the form of Eq. (E-6) as:

$$Y = a e^{bX} \tag{E-6}$$

The Excel spreadsheet Prog1a.xls is used to determine the activation energy E and the pre-exponential factor k_o as:

$$E = 78781.8\,\text{J/mol}, \quad k_o = 7.7 \times 10^{14}\,\text{s}^{-1}$$

Figure E-1 shows the snap shot of the spreadsheet.

The Fortran computer program PROG1 also determines the activation energy E and the pre-exponential factor k_o from the linearized data or alternatively from the non-linear equation $Y = ae^{bx}$. The best regression is determined from the various model equations in program PROG1. The regression coefficients from the linearized equation $Y = A + BX$ are as follows.

$$A = 0.3425\text{E}+02$$

$$B = -0.9468\text{E}+04$$

The activation energy E is obtained from the slope of $1/T$ vs $\ln k$

$$\text{Slope: } \frac{-E}{R} = -0.9468 \times 10^4$$

where

$R = 8.314\,\text{J/mol K}$
$E = 78716.9\,\text{J/mol}.$

The pre-exponential factor, $k_o = \exp(A) = \exp(34.25) = 7.49 \times 10^4\,\text{s}^{-1}$.

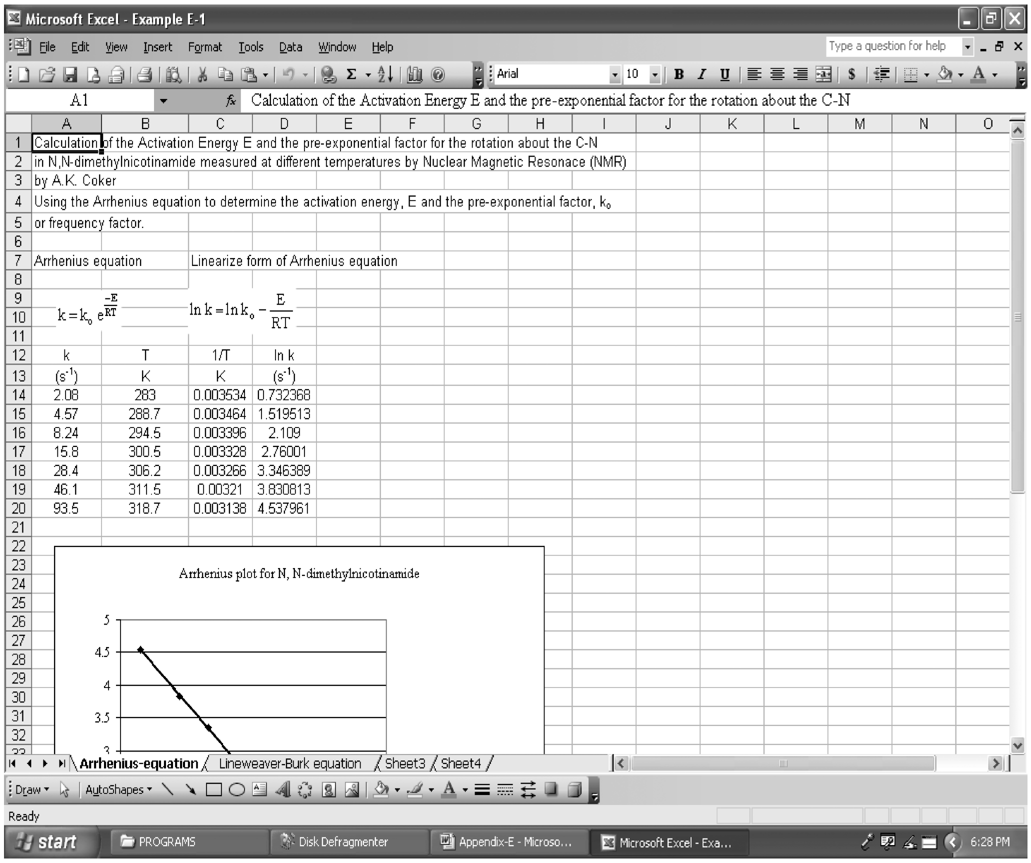


Figure E-1 The Excel spreadsheet snapshot of Example E-1.

Table E-1 Curve Fitting for Two Variables

X	$-1/X$	Y
283.0	−0.00353	2.08
288.7	−0.00346	4.57
294.5	−0.003396	8.24
300.5	−0.003328	15.8
306.2	−0.003266	26.4
311.6	−0.00321	46.1
318.7	−0.003138	93.5

The computer results give the constants A and B from the model Eq. (E-6) as:

$$A = 0.7206\text{E}+15$$

$$B = -0.9457\text{E}+04$$

$$\text{Slope}: \frac{-E}{R} = -0.9457 \times 10^4$$

where

$R = 8.314\,\text{J/mol K}$
$E = 78625\,\text{J/mol}.$

The pre-exponential factor, $k_o = A = 7.206 \times 10^4\,\text{s}^{-1}$.

The Excel spreadsheet Example E-1.xls shows another example using the Lineweaver–Burk equation to determine the constants k_m and V_{max}.

E.3 POLYNOMIAL REGRESSION

Some engineering data are often poorly represented by a linear regression. Some form of non-linear regression can be developed, if the dependence of Y on X is known, although the total convergence of this iterative regression procedure cannot be guaranteed. However, if the form of dependence is unknown, then Y can be treated as a general function of X by trigonometric terms (Fourier analysis) or polynomial function. The least squares procedure can be readily extended to fit the data to an nth-degree polynomial:

$$Y = C_0 + C_1X + C_2X^2 + \cdots + C_nX^n \tag{E-29}$$

where $C_0, C_1, C_2, \ldots, C_n$ are constants.

For this case, the sum of the squares of the residuals is minimized.

$$S = \sum_{j=1}^{N} \left[Y_j - C_0 - C_1X_j - C_2X_j{}^2 - \cdots - C_nX_j{}^n\right]^2 \tag{E-30}$$

At the minimum, all the partial derivatives with respect to the chosen constants are zero, that is

$$\frac{\partial S}{\partial C_0}, \frac{\partial S}{\partial C_1}, \cdots, \frac{\partial S}{\partial C_n} = 0 \tag{E-31}$$

This gives a system of $(n+1)$ linear equations in $(n+1)$ unknowns, $C_0, C_1, C_2, \ldots, C_n$

$$\frac{\partial S}{\partial C_0} = 0 = \sum_{j=1}^{N} 2\left(Y_j - C_0 - C_1 X_j - C_2 X_j^2 - \cdots - C_n X_j^n\right)(-1)$$

$$\frac{\partial S}{\partial C_1} = 0 = \sum_{j=1}^{N} 2\left(Y_j - C_0 - C_1 X_j - C_2 X_j^2 - \cdots - C_n X_j^n\right)\left(-X_j\right)$$

$$\frac{\partial S}{\partial C_2} = 0 = \sum_{j=1}^{N} 2\left(Y_j - C_0 - C_1 X_j - C_2 X_j^2 - \cdots - C_n X_j^n\right)\left(-X_j^2\right)$$

$$\vdots$$

$$\frac{\partial S}{\partial C_n} = 0 = \sum_{j=1}^{N} 2\left(Y_j - C_0 - C_1 X_j - C_2 X_j^2 - \cdots - C_n X_j^n\right)\left(-X_j^n\right) \quad \text{(E-32)}$$

The above equations are set to equal zero, and can be rearranged in the following set of normal equations.

$$C_0 N + C_1 \sum X_j + C_2 \sum X_j^2 + \cdots + C_n \sum X_j^n = \sum Y_j$$

$$C_0 \sum X_j + C_1 \sum X_j^2 + C_2 \sum X_j^3 + \cdots + C_n \sum X_j^{n+1} = \sum X_j Y_j$$

$$C_0 \sum X_j^2 + C_1 \sum X_j^3 + C_2 \sum X_j^4 + \cdots + C_n \sum X_j^{n+2} = \sum X_j^2 Y_j$$

$$\vdots$$

$$C_0 \sum X_j^n + C_1 \sum X_j^{n+1} + C_2 \sum X_j^{n+2} + \cdots + C_n \sum X_j^{2n} = \sum X_j^n Y_j \quad \text{(E-33)}$$

Equation (E-33) in matrix form becomes

$$UC = V$$

$$U = \begin{bmatrix} N & \sum X_j & \sum X_j^2 & \cdots & \sum X_j^n \\ \sum X_j & \sum X_j^2 & \sum X_j^3 & \cdots & \sum X_j^{n+1} \\ \sum X_j^2 & \sum X_j^3 & \sum X_j^4 & \cdots & \sum X_j^{n+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum X_j^n & \sum X_j^{n+1} & \sum X_j^{n+2} & \cdots & \sum X_j^{2n} \end{bmatrix}$$

$$C = \begin{bmatrix} C_0 \\ C_1 \\ C_2 \\ \vdots \\ C_n \end{bmatrix} \quad V = \begin{bmatrix} \sum Y_j \\ \sum X_j Y_j \\ \sum X_j^2 Y_j \\ \vdots \\ \sum X_j^n Y_j \end{bmatrix}$$

Linear equations generated by polynomial regression can be ill-conditioned when the coefficients have very small and very large numbers. This results in smooth curves that fit poorly. Ill-conditioning often happens if the degree of polynomial is large and if the Y values cover a wide range. The error of the polynomial regression by a standard error of the estimate is shown below:

$$\sigma^2 = \frac{\sum \text{SSE}}{N - n - 1} \quad \text{(E-34)}$$

where

n = the degree of polynomial
N = the number of data points.

The coefficient of determination can be expressed as:

$$r^2 = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\sum_{j=1}^{N}\left(Y_j - \hat{Y}_j\right)^2}{\sum_{j=1}^{N}\left(Y_j - \overline{Y}_j\right)^2} \quad \text{(E-35)}$$

The correlation coefficient is given by

$$r = \left(1 - \frac{\text{SSE}}{\text{SST}}\right)^{0.5} \quad \text{(E-36)}$$

The numerator of Eq. (E-34) should continually decrease as the degree of the polynomial is raised. Alternatively, the denominator of Eq. (E-34) causes σ^2 to increase once there is departure from the optimum degree.

Example E-2

The following data are obtained from $y(x) = x^4 + 3x^3 + 2x^2 + x + 5$. Show that a fourth degree polynomial provides the best least squares approximation to the given data. Determine this polynomial.

Fortran program PROG2 calculates (i) the coefficients for each degree of the polynomial, (ii) the variance, (iii) error sum of squares, (iv) total sum of squares, (v) coefficient of determination, and (vi) the correlation coefficient. Alternatively, the Microsoft Excel spreadsheet can be employed using the *Trendline* from the Insert menu. The program shows that the fourth degree gives the lowest value of variance and therefore shows the best fit. The computer results are shown in Table E-2. The Excel spreadsheet Prog2a.xls was used to determine the polynomial regression of Example E-2. Figure E-2 shows the Excel spreadsheet snap shot of Example E-2.

Table E-2 Polynomial Regression Analysis for an Equation to an Nth Degree

X	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Y	5.123	5.306	5.569	5.938	6.437	7.098	7.949	9.025	10.363

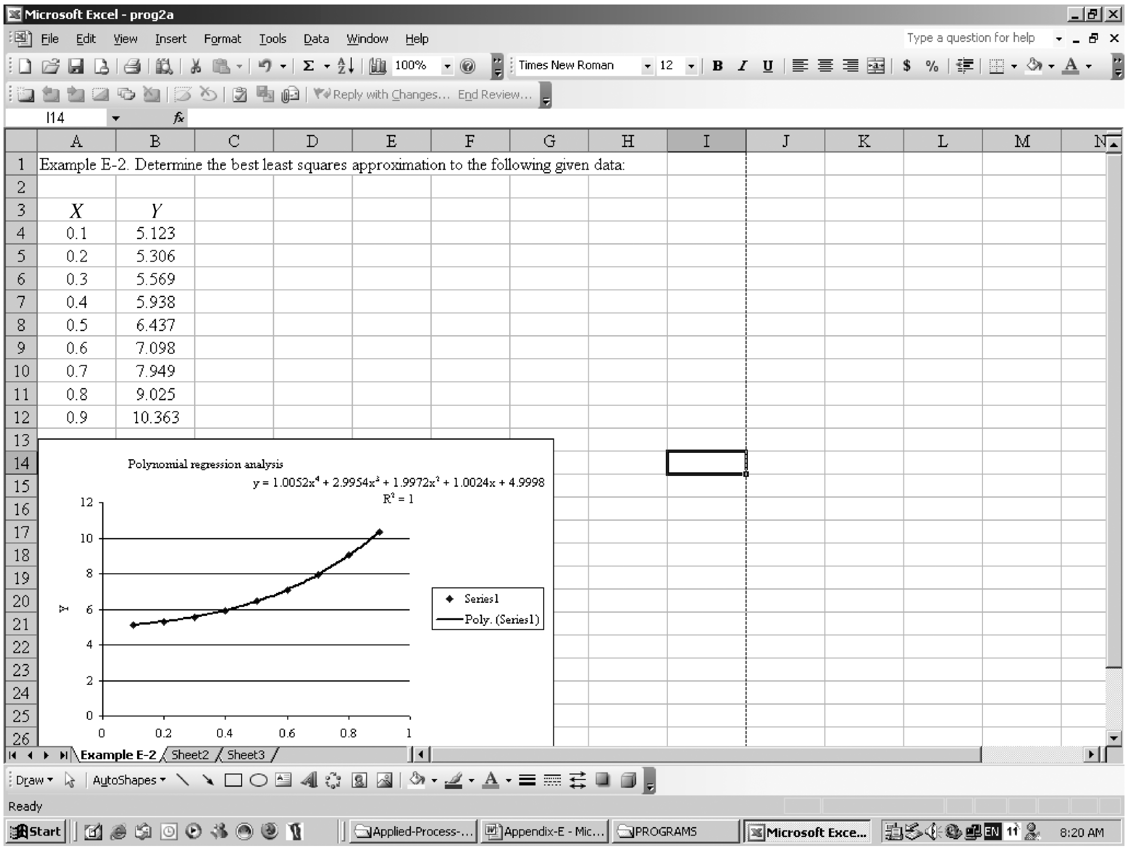


Figure E-2 The Excel spreadsheet snapshot of Example E-2.

E.4 MULTIPLE REGRESSION ANALYSIS

Inadequate results are sometimes obtained with a single independent variable. This shows that one independent variable does not provide enough information to predict the corresponding value of the dependent variable. We can approach this problem, if we use additional independent variables, and develop a multiple regression analysis to achieve a meaningful relationship. Here, we can employ a linear regression model in cases where the dependent variable is affected by two or more controlled variables.

The linear multiple regression equation is expressed as:

$$Y = C_0 + C_1 X_1 + C_2 X_2 + \cdots + C_K X_K \tag{E-37}$$

where

Y = the dependent variable
$X_1, X_2, \cdots, X_K$ = the independent variable
$C_0, C_1, C_2, \cdots, C_K$ = the unknown regression coefficients
K = the number of independent variables.

The unknown coefficients are estimated based on n observation for the dependent variable Y, and for each of the independent variables X_i's where $i = 1, 2, 3, \ldots, K$.

These observations are of the form:

$$Y_j = C_0 + C_1 X_{1j} + C_2 X_{2j} + \cdots + C_K C_{Kj} + \xi_j \tag{E-38}$$

For $j = 1, 2, \ldots, N$

where

Y_j = the jth observation of the dependent variable
$X_{1j}, \ldots, X_{Kj}$ = the jth observation of the $X_1, X_2, \ldots, X_K$ independent variables.

We can use a least squares technique to calculate estimates of $\hat{C}_0, \hat{C}_1, \ldots, \hat{C}_K$ by minimizing the following equation:

$$S = \sum_{j=1}^{N} \left[Y_j - \left(\hat{C}_0 + \hat{C}_1 X_{1j} + \cdots + \hat{C}_K X_{Kj} \right) \right]^2 = \sum_{j=1}^{N} \xi_j^{\,2} \tag{E-39}$$

Taking the partial derivatives of S with respect to $\hat{C}_0, \hat{C}_1, \ldots, \hat{C}_K$, that is $\frac{\partial S}{\partial \hat{C}_0}, \frac{\partial S}{\partial \hat{C}_1}, \ldots, \frac{\partial S}{\partial \hat{C}_K}$ and setting them equal to zero, we obtain the following set of equations:

$$\begin{aligned}
&N\hat{C}_0 + \left(\sum X_{1j}\right)\hat{C}_1 + \cdots + \left(\sum X_{Kj}\right)\hat{C}_K = \sum Y_j \\
&\left(\sum X_{1j}\right)\hat{C}_0 + \left(\sum X_{1j}{}^2\right)\hat{C}_1 + \cdots + \left(\sum X_{1j} X_{Kj}\right)\hat{C}_K = \sum X_{1j} Y_j \\
&\left(\sum X_{2j}\right)\hat{C}_0 + \left(\sum X_{1j} X_{2j}\right)\hat{C}_1 + \cdots + \left(\sum X_{2j} X_{Kj}\right)\hat{C}_K = \sum X_{2j} Y_j \\
&\vdots \\
&\left(\sum X_{Kj}\right)\hat{C}_0 + \left(\sum X_{1j} X_{Kj}\right)\hat{C}_1 + \cdots + \left(\sum X_{Kj}{}^2\right)\hat{C}_K = \sum X_{Kj} Y_j
\end{aligned} \tag{E-40}$$

Equation (E-40) can be expressed in matrix form as:

$$U\hat{C} = V$$

where

$$U = \begin{bmatrix} N & \sum X_{1j} & \cdots & \sum X_{Kj} \\ \sum X_{1j} & \sum X_{1j}^2 & \cdots & \sum X_{1j}X_{Kj} \\ \vdots & & & \\ \sum X_{Kj} & \sum X_{1j}X_{Kj} & & \sum X_{Kj}^2 \end{bmatrix} \quad \text{(E-41)}$$

$$\hat{C} = \begin{bmatrix} \hat{C}_0 \\ \hat{C}_1 \\ \vdots \\ \hat{C}_K \end{bmatrix} \quad V = \begin{bmatrix} \sum Y_j \\ \sum X_{1j} Y_j \\ \vdots \\ \sum X_{Kj} Y_j \end{bmatrix} \quad \text{(E-42)}$$

U is a symmetric matrix.

We can obtain estimates for the coefficients $\hat{C}_0,\ \hat{C}_1, \ldots, \hat{C}_K$ by successive elimination or by solving for the inverse of U. That is,

$$\hat{C} = U^{-1} V \quad \text{(E-43)}$$

where

U^{-1} = the inverse of U.

After solving for $\hat{C}_0,\ \hat{C}_1, \ldots, \hat{C}_K$, the estimates of the dependent variable observations $\hat{Y}_j$ can be obtained as follows:

$$\hat{Y}_j = \hat{C}_0 + \hat{C}_1 X_{1j} + \cdots + \hat{C}_K X_{Kj} \quad \text{(E-44)}$$

The power equations have often been derived to calculate the parameters of experimental data. Such an equation can be expressed in the form:

$$Y = C_0 \cdot X_1^{C_1} \cdot X_2^{C_2} \cdot \ldots \cdot X_K^{C_K} \quad \text{(E-45)}$$

We can calculate the coefficients of the independent variables, if Eq. (E-45) is linearized by taking its natural logarithm to give

$$\ln Y = \ln C_0 + C_1 \ln X_1 + C_2 \ln X_2 + \cdots + C_K \ln X_K \quad \text{(E-46)}$$

TABLE E-3 Analysis of Variance Table for Linear Multiple Regression

Source of Variance	Degree of Freedom	Sum of Squares	Mean Squares
Total	$N-1$	$\mathrm{SST} = \sum \left(Y_j - \overline{Y}\right)^2$	$\mathrm{MST} = \mathrm{SST}/(N-1)$
Regression	K	$\mathrm{SSR} = \sum \left(\hat{Y}_j - \overline{Y}\right)^2$	$\mathrm{MSR} = \mathrm{SSR}/K$
Error	$N-K-1$	$\mathrm{SSE} = \sum \left(Y_j - \hat{Y}_j\right)^2$	$\mathrm{MSE} = \mathrm{SSE}/(N-K-1)$

The coefficients $C_0, C_1, C_2, \ldots, C_K$ can then be obtained by Gaussian elimination. Table E-3 shows the variance table for linear multiple regression. The coefficient of determination is

$$r^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \quad \text{(E-47)}$$

and the correlation coefficient is

$$r = \left(1 - \frac{\mathrm{SSE}}{\mathrm{SST}}\right)^{0.5} \quad \text{(E-48)}$$

The test statistic is the F-ratio, which is defined by

$$F = \frac{\mathrm{MSR}}{\mathrm{MSE}} \quad \text{(E-49)}$$

A computer program PROG3 has been developed to determine the coefficients and correlation coefficient of a multiple regression equation.

E.5 SIMULTANEOUS EQUATIONS USING THE MATRIX METHODS

Matrix solution of sets of linear equations are solved by Excel as follows.

First, write a set of simultaneous equations in matrix format

$$Ax = b \quad \text{(E-51)}$$

where A is the coefficient matrix, x is the vector of unknowns, and b is the result vector. To solve this matrix equation, multiply both sides from the left by the inverse of A:

$$A^{-1} A\, x = A^{-1} b \quad \text{(E-52)}$$

Example E-3

In a fluid flow experiment, the volumetric rate of fluid through a pipe is dependent on the pipe diameter and slope by the equation

$$Q = C_0 D^{C_1} S^{C_2} \quad \text{(E-50)}$$

where

Q = flow rate, ft^3/s
D = pipe diameter, ft
S = slope, ft/ft.

Determine the flow rate of fluid for a pipe with a diameter of 3.25 ft and slope of 0.025 ft/ft with the following data:

Table E-4 Linear Multiple Regression Analysis for an Equation

Diameter, ft D	Slope, ft/ft S	Flow rate, ft^3/s Q
1.0	0.001	1.5
2.5	0.005	9.0
3.0	0.010	25.0
4.0	0.010	5.0
1.5	0.050	30.0
3.5	0.050	100.0

Computer program PROG3 is developed to calculate the values of the coefficients C_0, C_1, and C_2 and the correlation coefficient. Table E-4 shows the results of the program.

By definition, a matrix times its inverse gives the identity matrix, and the identity matrix times any vector gives the vector back. The equation above reduces to the following solution:

$$x = A^{-1}\, b \tag{E-53}$$

The following example uses the Fortran program PROG4 and Excel spreadsheet Prog4a.xls.

Example E-4

The final product from a chemical factory is made by blending four liquids (α, β, γ, δ) together. Each of these liquids contains four components *A*, *B*, *C*, and *D*. The product leaving the factory has to have a closely specified composition. Determine the relative quantities of α, β, γ, and δ required to meet the blend specifications in the following data:

Computer program PROG4 uses the Gaussian elimination method to determine the quantities of each α, β, γ, and δ required to meet the blending specifications. Table E-5 shows the computer results.

TABLE E-5

Component	α	β	γ	δ	composition of specification (W/W)
A	51.30	43.20	56.40	47.40	48.80
B	11.30	11.50	15.50	8.50	11.56
C	29.40	31.50	22.50	30.40	29.43
D	8.00	10.30	5.60	13.70	10.21

(*Source*: B.Sc. Final year 1978, Aston University, Birmingham, UK.)

Answers:

$\alpha = 0.1172$
$\beta = 0.3789$
$\gamma = 0.2117$
$\delta = 0.3054$

Alternatively, the Excel spreadsheet Prog4a.xls is used to solve the above example problem. The procedure is as follows.

1. Start with a new worksheet, and name it Matrix Method.
2. In cell A1, enter solving sets of equations; Matrix inversion.
3. In cell B3, enter $A\mathbf{x} = \mathbf{b}$.
   Enter the coefficient matrix **A** and the results vector **b**.
4. In cell A5, enter input matrix **A**.
5. In cells A6:D9, enter the constants of matrix **A**:

Cell	Constant	Cell	Constant	Cell	Constant	Cell	Constant
A6	51.3	B6	43.2	C6	56.4	D6	47.4
A7	11.3	B7	11.5	C7	15.5	D7	8.5
A8	29.4	B8	31.5	C8	22.5	D8	30.4
A9	8	B9	10.3	C9	5.6	D9	13.7

6. In cell F5, enter result vector **b**
7. In cells F6:F9, enter the constants of the result vector.
8. In cell A12, enter Inverse matrix (1/**A**)
9. Select cells A13:D16, then type

   = MINVERSE (A6:D9)

   and press Ctrl + Shift + Enter to insert the formula into the whole section.
10. In cell F12, enter solution vector $\mathbf{x} = (1/A)\mathbf{b}$.
11. Select cells F13:F16, then type

    + MMULT (A13:D16, F6:F9)

    and press Ctrl + Shift + Enter to insert the formula into the whole selection.
12. Turn off the gridelines, and outline the worksheet cells as shown in Matrix Method.

The worksheet should now look like Matrix Method, with the solution values 0.11722, 0.378939, 0.211747, and 0.305356 for α, β, γ, and δ in cells F13:F16. This is shown in Figure E-3.

E.6 SOLVING SIMULTANEOUS NON-LINEAR EQUATIONS IN EXCEL USING *SOLVER*

Solver can be used to solve a system of simultaneous equations, which can be either linear or non-linear. Suppose the equations are represented as

$$f_1(x_1, x_2, \ldots, x_n) = 0 \tag{E-54}$$

$$f_2(x_1, x_2, \ldots, x_n) = 0 \tag{E-55}$$

$$f_3(x_1, x_2, \ldots, x_n) = 0 \tag{E-56}$$

$$\vdots$$

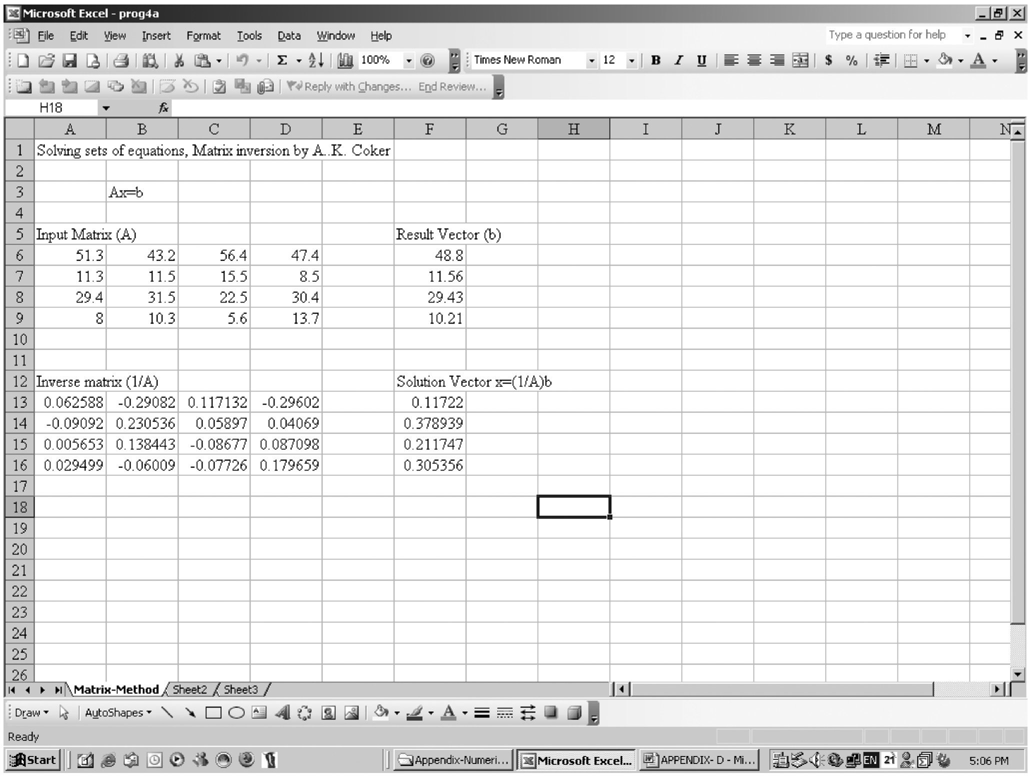
$$f_n(x_1, x_2, \ldots, x_n) = 0 \tag{E-57}$$

Thus, we have a system of *n* equations in *n* unknowns. We can find the values of $x_1, x_2, x_3, \ldots, x_n$ that cause each of the equations to equal zero, if the function

$$y = f_1^{\,2} + f_2^{\,2} + f_3^{\,2} + \ldots + f_n^{\,2} \tag{E-58}$$

is equal to zero; that is, to find the values of $x_1, x_2, x_3, \ldots, x_n$ that cause Eq. (E-58) to equal zero. Since all the terms on the right side of Eq. (E-58) are squares, they will all be greater than or equal to zero. Hence, the only way that *y* can be equal to zero is that each of the individual *f*'s should be equal to zero. Therefore, the values of $x_1, x_2, x_3, \ldots, x_n$ that cause *y* to equal zero will be the solutions to the given system of equations. The general approach used with *Solver* is to define a target function consisting of the squares of the individual equations, as indicated by Eq. (E-58), and to then determine the values of $x_1, x_2, x_3, \ldots, x_n$ that cause the target function to equal zero. To solve a system of simultaneous equations with *Solver*, proceed as follows.

1. Enter an initial guess for each independent variable ($x_1, x_2, x_3, \ldots, x_n$) in a separate cell on the worksheet.
2. Enter the equations for $f_1, f_2, f_3, \ldots, f_n$ and *y* in separate cells, expressed as Excel formulas. Within these formulas, express the unknown quantities $x_1, x_2, x_3, \ldots, x_n$ as the addresses of the cells containing the initial guess.
3. Select *Solver* from the Tools menu.
4. When the *Solver Parameters* dialog box appears, enter the following information.



Solving sets of equations, Matrix inversion by A..K. Coker					
	Ax=b				
Input Matrix (A)					Result Vector (b)
51.3	43.2	56.4	47.4		48.8
11.3	11.5	15.5	8.5		11.56
29.4	31.5	22.5	30.4		29.43
8	10.3	5.6	13.7		10.21
Inverse matrix (1/A)					Solution Vector x=(1/A)b
0.062588	-0.29082	0.117132	-0.29602		0.11722
-0.09092	0.230536	0.05897	0.04069		0.378939
0.005653	0.138443	-0.08677	0.087098		0.211747
0.029499	-0.06009	-0.07726	0.179659		0.305356

Figure E-3 The Excel spreadsheet snapshot of Example E-4.

(a) The address of the cell containing the formula for y in the Set Target Cell location.

(b) Select value of y in the Equal to line. Then enter 0 within the associated data area (i.e., determine the values of $x_1, x_2, x_3, \ldots, x_n$ that will drive the target function to zero).

(c) Enter the range of cell addresses containing the initial values of $x_1, x_2, x_3, \ldots, x_n$ in the area labeled *By Changing Cells*.

(d) If you wish to restrict the range of the independent variables, click on the **Add** button under the heading *Subject to the Constraints*. Then provide the following information within the *Add Constraint* dialog box for each of the independent variables.

- **(i)** The cell address containing the initial value of the independent variable in the Cell Reference location.
- **(ii)** The type of constraint (i.e., $\leq$ or $\geq$) from the pull-down menu.
- **(iii)** The limiting value in the constraint data area.
- **(iv)** Select **OK** to return to the *Solver Parameters* dialog box or select **Add** to add another constraint.

Note that you can always change a constraint or delete a constraint after it has been added.

(e) When all the required information has been entered correctly, select **Solve**. This will initiate the actual solution procedure.

A new dialog box labeled *Solver Results* will then appear, telling you whether or not *Solver* has been able to solve the problem. If a solution has been obtained, the desired values of the independent variables will appear in the cells that originally contained the initial values. The cell containing the target function will show a value that is zero or nearly zero.

E.7 GAUSS–SEIDEL ITERATIVE METHOD

Alternatively, Fortran computer program PROG5 employs the Gauss–Seidel method for solving the four material balance equations. The four material balance Eqs. (E-65)–(E-68) are rearranged to solve for the unknown on the diagonal position of each Eq. (E-74).

$$(1000)\,(1) = 1000C_{A1} + (0.1)\,(C_{A1})\,(1000) \tag{E-65}$$

$$1000C_{A1} + 100\,C_{A3} = 1100\,C_{A2} + (0.2)\,(C_{A2})\,(1500) \tag{E-66}$$

$$1100C_{A2} + 100\,C_{A4} = 1200\,C_{A3} + (0.4)\,(C_{A3})\,(100) \tag{E-67}$$

$$1000C_{A3} = 1100\,C_{A4} + (0.3)\,(C_{A4})\,(500) \tag{E-68}$$

$$\begin{array}{lllll}1100C_{A1} & & & & = 1000\\ 1000C_{A1} & -1400C_{A2} & +100C_{A3} & & = 0\\ & 1100C_{A2} & -1240C_{A3} & +100C_{A4} & = 0\\ & & 1100C_{A3} & -1250C_{A4} & = 0\end{array} \tag{E-74}$$

Table E-7 shows the computer results of PROG5.

Example E-5

A chemical reaction takes place in a series of four continuous flow stirred tank reactors arranged as shown. The chemical reaction is a first order irreversible reaction of the type

$$A \xrightarrow{k} B$$

The conditions of temperature in each reactor are such that the values of k_i and V_i are given in Table E-6. Figure E-4 shows four continuous stirred tanks with recycles streams.

Table E-6

Reactor	Volume, $V_i L$	Rate constant, $k_i h^{-1}$
1	1000	0.1
2	1500	0.2
3	100	0.4
4	500	0.3

(*Source*: A. Constantinides, *Applied Numerical Methods With Personal Computers*, McGraw-Hill Book Co., 1987)

The following assumptions are:

1. The system is at steady state.
2. The reactions are in the liquid phase.
3. There is no change in volume or density of the liquid.
4. The rate of disappearance of component A in each reactor is given by

$$(-r_A) = kC_A \tag{E-59}$$

Set up the material balance equations for each of the four reactors, and use the Gauss–Seidel or *Solver* to determine the exit concentration from each reactor.

Solution

Set up the material balance equations for each of the four reactions, and use the Gaus–Seidel method to determine the exit concentrations from each reactor.

The general unsteady state material balance for each reactor is

$$\text{Input by flow} = \text{Output by flow} + \text{disappearance}$$
$$+ \text{accumulation by reaction}$$

Mass balance for reactor 1:

$$u_{A0}\, C_{A0} = u_{A0}\, C_{A1} + (-r_A)\, V_1 + V_1 \frac{dC_{A1}}{dt} \tag{E-60}$$

Because the system is at steady state, the accumulation is zero, the above equation becomes

$$u_{A0}\, C_{A0} = u_{A0}\, C_{A1} + k_1 C_{A1} V_1 \tag{E-61}$$

Mass balance for reactor 2:

$$u_{A0}\, C_{A1} + u_{A3}\, C_{A3} = (u_{A0} + u_{A3})\; C_{A2} + k_2 C_{A2} V_2 \tag{E-62}$$

Mass balance for reactor 3:

$$(u_{A0} + u_{A3})\, C_{A2} + u_{A4}\, C_{A4} = (u_{A0} + u_{A3} + u_{A4})\; C_{A3} + k_3 V_3 C_{A3} \tag{E-63}$$

Mass balance for reactor 4:

$$(u_{A0} + u_{A4})\, C_{A3} = (u_{A0} + u_{A4})\; C_{A4} + k_4 V_4 C_{A4} \tag{E-64}$$

where

$u_{A0} = 1000\,\mathrm{L}/h,\; C_{A0} = 1\,\mathrm{mol/L}$
$u_{A3} = 100\,\mathrm{L/h}$
$u_{A4} = 100\,\mathrm{L/h}.$

$$(1000)\ (1) = 1000C_{A1} + (0.1)\ (C_{A1})\ (1000) \tag{E-65}$$

$$1000C_{A1} + 100\,C_{A3} = 1100\,C_{A2} + (0.2)\ (C_{A2})\,(1500) \tag{E-66}$$

$$1100C_{A2} + 100\,C_{A4} = 1200\,C_{A3} + (0.4)\ (C_{A3})\,(100) \tag{E-67}$$

$$1000C_{A3} = 1100\,C_{A4} + (0.3)\ (C_{A4})\,(500) \tag{E-68}$$

Rearranging the above equations:

$$f\,(C_{A1},\, C_{A2},\, C_{A3},\, C_{A4}) = 1100C_{A1} - 1000 \tag{E-69}$$

$$g\,(C_{A1},\, C_{A2},\, C_{A3},\, C_{A4}) = 1000C_{A1} - 1400C_{A2} + 100C_{A3} \tag{E-70}$$

$$h\,(C_{A1},\, C_{A2},\, C_{A3},\, C_{A4}) = 1100\,C_{A2} - 1240\,C_{A3} + 100\,C_{A4} \tag{E-71}$$

$$j\,(C_{A1},\, C_{A2},\, C_{A3},\, C_{A4}) = 1100C_{A3} - 1250\,C_{A4} \tag{E-72}$$

Thus, we have a system of n equations in four unknowns. We can find the values of C_{A1}, C_{A2}, C_{A3}, C_{A4} that cause each of the equations to equal zero, if the function

$$y = f^2 + g^2 + h^2 + j^2 \tag{E-73}$$

is equal to zero; that is, to find the values of C_{A1}, C_{A2}, C_{A3}, C_{A4} that cause Eq. (E-73) to equal zero. Since all the terms on the right side of Eq.(E-73) are squares, they will all be greater than or equal to zero. Hence, the only way that y can be equal to zero is that each of the individual f, g, h, j should be equal to zero. Therefore, the values of C_{A1}, C_{A2}, C_{A3}, C_{A4} that cause y to equal zero will be the solutions to the given system of equations. The general approach used with *Solver* is to define a target function consisting of the squares of the individual equations, as indicated by Eq. (E-73), and to then determine the values of C_{A1}, C_{A2}, C_{A3}, C_{A4} that cause the target function to equal zero. Figures E-5–E-8 of Excel snapshots show the procedures of determining the final concentrations, after initial guesses.

$$C_{A1} = C_{A2} = C_{A3} = C_{A4} = 0.5$$

TABLE E-7 SOLUTION OF SIMULTANEOUS LINEAR EQUATIONS BY GAUSS-SEIDEL METHOD

```
********************************************************************************************************************

NUMBER OF LINEAR EQUATIONS: 4
MAXIMUM ITERATIONS 15
CONVERGING TOLERANCE: 0.100000E-03

THE COEFFICIENT MATRIX A(1,1)...A(N+1,N+1) IS:
```

1100.0000	0.0000	0.0000	0.0000	1000.0000
1100.0000	−1400.0000	100.0000	0.0000	0.0000
0.0000	1100.0000	−1240.0000	100.0000	0.0000
0.0000	0.0000	1100.0000	−1250.0000	0.0000

```
THE STARTING VECTOR X(1)....X(N) IS:

        5.000000E-01
        5.000000E-01
        5.000000E-01
        5.000000E-01

PROCEDURE CONVERGED AFTER 5 ITERATIONS

SOLUTION VECTOR X(1).....X(N) IS:

        9.090909E-01
        6.968780E-01
        6.654189E-01
        5.855687E-01
```

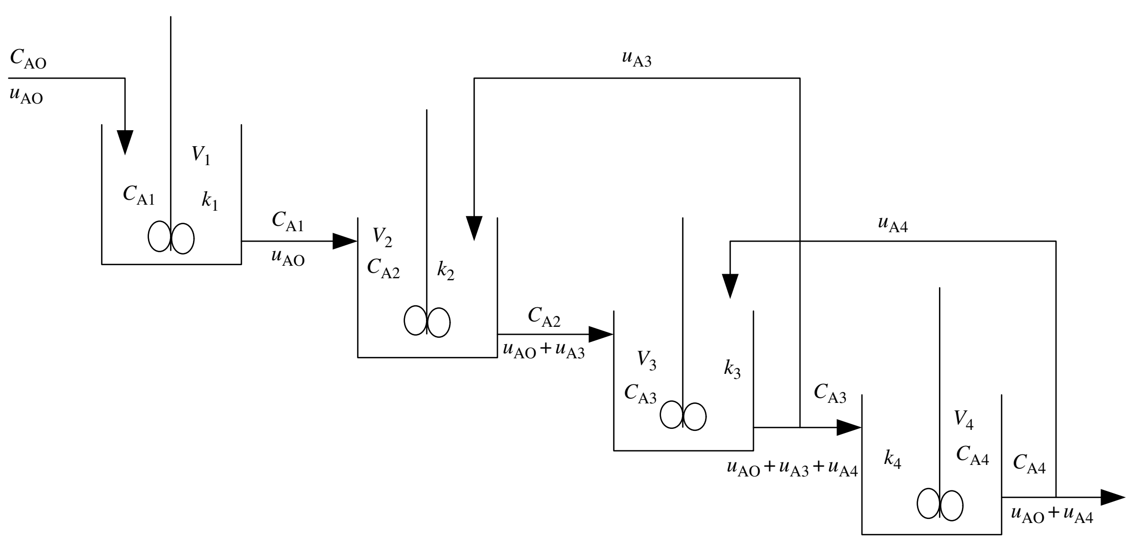


Figure E-4 Chemical reaction with recycles in four continuous stirred tanks.

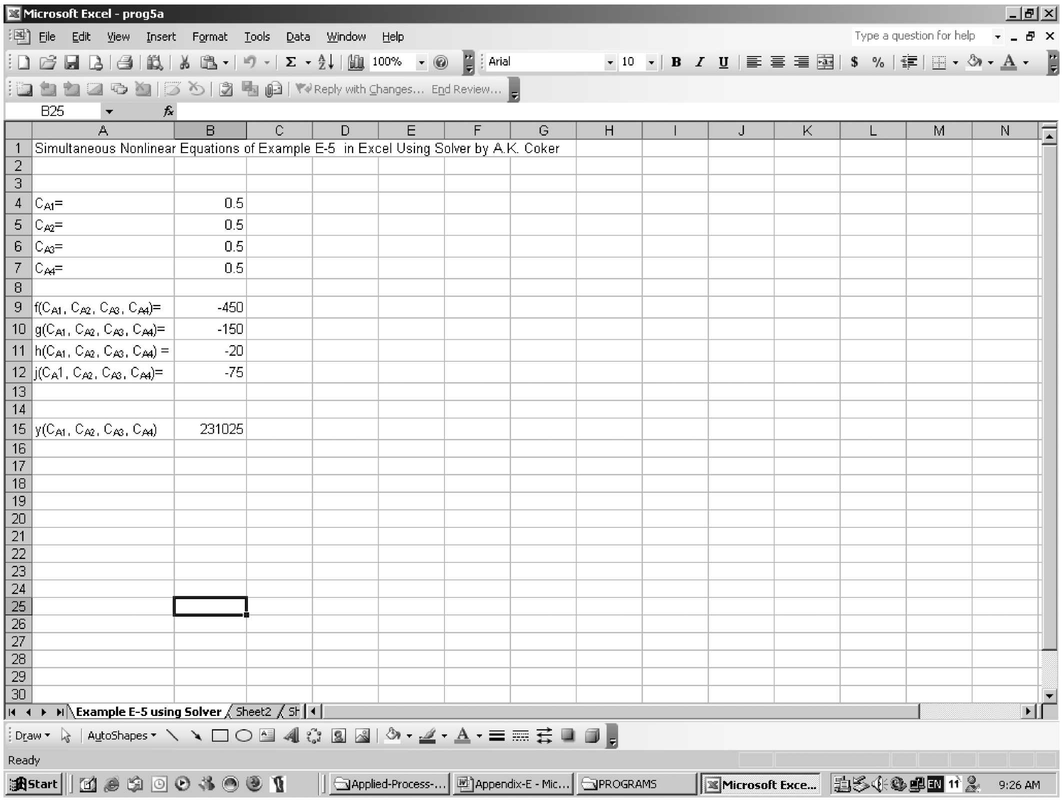


Figure E-5 The Excel spreadsheet snapshot of Example E-5.

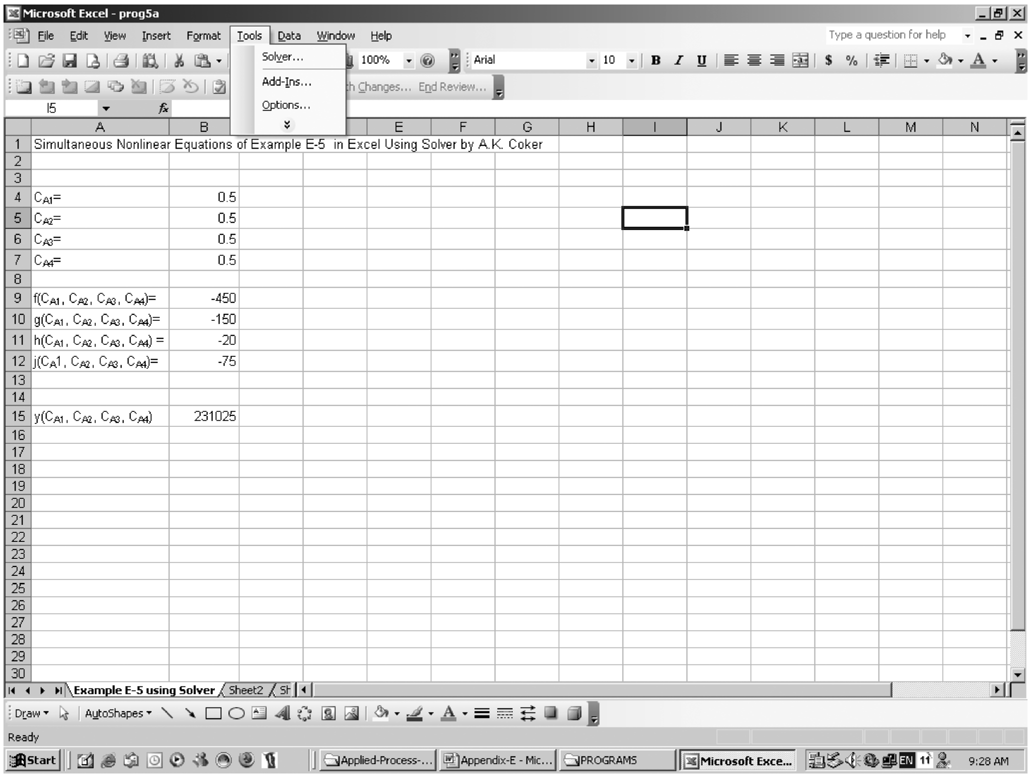


Figure E-6 The Excel spreadsheet snapshot of Example E-5 (continued).

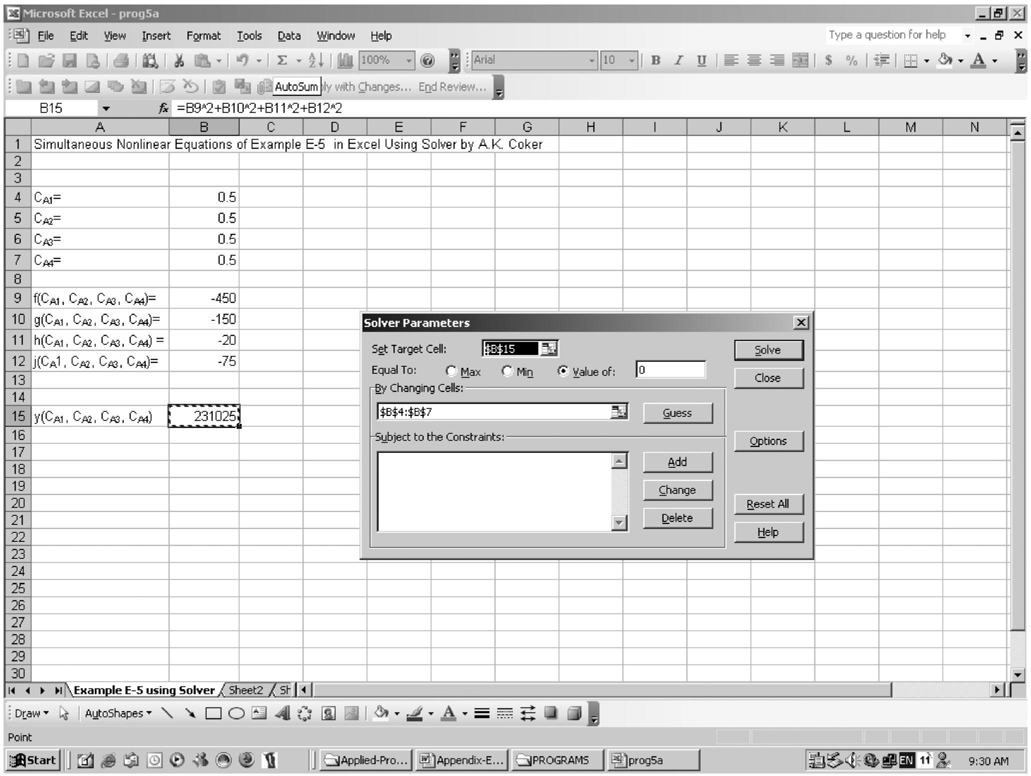


Figure E-7 The Excel spreadsheet snapshot of Example E-5 (continued).

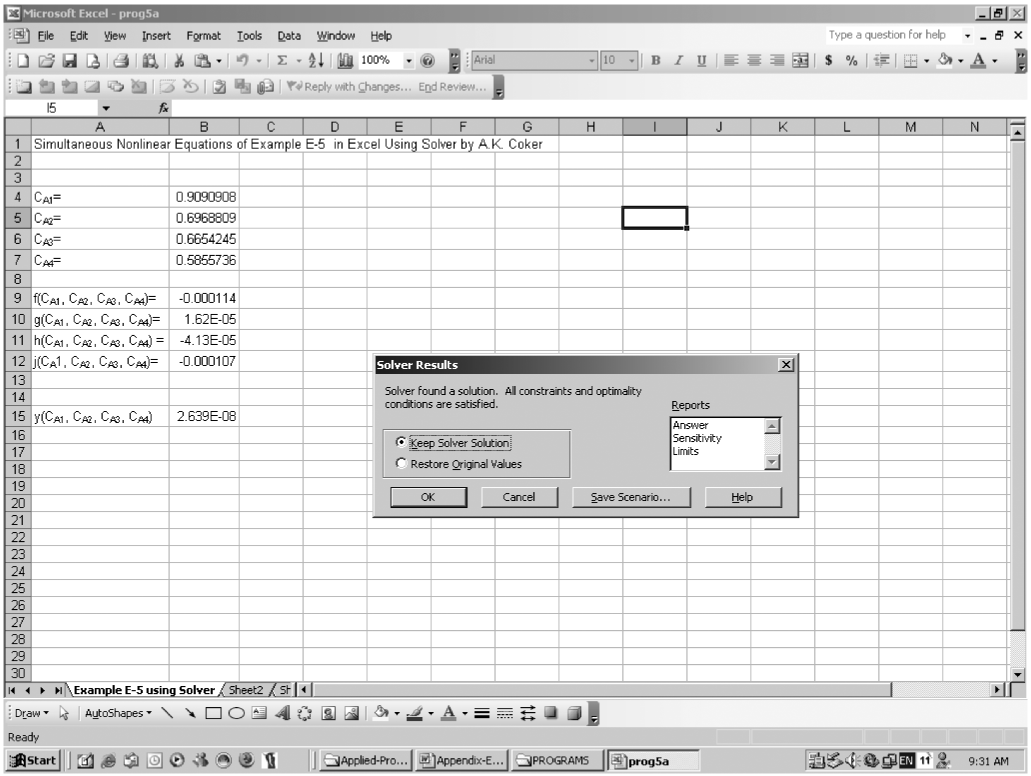


Figure E-8 The Excel spreadsheet snapshot of Example E-5 (continued).

NOMENCLATURE

F = F-ratio
MSE = Error mean squares
MSR = Regression mean squares
MST = Total mean squares
r = Coefficient of determination
SSE = Error sum of squares
SSR = Regression sum of squares
SST = Total sum of squares

FURTHER READING

1. Coker, A.K., *Fortran Programs For Chemical Process Design, Analysis and Simulation*, Gulf Publishing Co., Houston, Tx, 1995.
2. Gottfield, B.S., *Spreadsheet Tools for Engineers: Excel 5.0 Version*, The McGraw-Hill Co. Inc., 1996.
3. Billo, E.J., *Excel for Chemists – A Comprehensive Guide*, Wiley-VCH, Inc., 1997.
4. Gottfried, B.S., *Spreadsheet Tools for Engineers: Excel 2000 Version*, McGraw-Hill's Best – Basic Engineering Series and Tools, 2000.