Smith, T. L., and R. A. Dickie, in *Block Copolymers*, p. 163, Interscience, New York (1969).
Soen, T., T. Inoue, K. Miyoshi, and H. Kawai, *J. Polymer Sci.*, **10**, Part A2, 1757 (1972).
Soen, T., T. Horino, Y. Ogawa, K. Kyuma, and H. Kawai, *ibid.*, **10**, 1499 (1966).
Soen, T., T. Ono, K. Yamashita, and H. Kawai, *Kolloid Z.u.Z.f. Polymer*, **250**, 459 (1972).
Soen, T., M. Shimomura, T. Uchida, and H. Kawai, *Colloid Polymer Sci.*, **252**, 933 (1974).
Soong, D., and M. Shen, *Macromol.*, **10**, 357 (1977).
Sperling, L. H., ed., *Recent Advances in Polymer Blends, Grafts and Blocks*, Plenum, New York (1973).
Stoelting, J., F. E. Karasz, and W. J. MacKnight, *Polymer Sci. Eng.*, **10**, 133 (1970).
Takayanagi, M., S. Uemura, and S. Minami, *J. Polymer Sci.*, **C5**, 113 (1964).
Tobolsky, A. V., *Properties and Structure of Polymers*, Wiley, New York (1960).
———, and K. Murakami, *J. Polymer Sci.*, **40**, 433 (1959).
Tompa, H., *Trans. Faraday Soc.*, **45**, 1142 (1949).
Toy, L., M. Niinomi, and M. Shen, *J. Macromol. Sci.-Phys.*, **11**, 281 (1975).
Wagner, E. R., *Rubber Chem. Tech.*, **43**, 1129 (1970).
Wang, F. W., *Macromol.*, **8**, 364 (1975).
Wilkes, G. L., and S. L. Samuels, in *Block and Graft Copolymers*, J. J. Burke and V. Weiss, ed., p. 225, Syracuse Univ. Press, N.Y. (1973).
Williams, M. L., R. F. Landel, and J. D. Ferry, *J. Am. Chem. Soc.*, **77**, 3701 (1955).
Williams, M. C., *AIChE J.*, **21**, 1 (1975).
Zimm, B., *J. Chem. Phys.*, **24**, 269 (1956).

THE AUTHORS

Mitchel Shen received his B.S. from Saint Francis College and his M.A. and Ph.D. from Princeton University. Prior to joining the Berkeley faculty in 1969, he was at the Rockwell Science Center in Thousand Oaks, California. His research interests are in the physical properties of polymers, plasma polymerization, and natural macromolecules. In 1975 Mitch was a Visiting Professor of Polymer Chemistry at Kyoto University, where the preparation of this review was initiated.

Hiromichi Kawai received his undergraduate education at the Tokyo Institute of Technology. Since 1946 he has been associated with Kyoto University, where he also received his doctorate degree. At present, Hiro is a Professor in the Department of Polymer Chemistry there. From 1958 to 1960 he worked with R. S. Stein at the University of Massachusetts as a post doctoral fellow and since then has made numerous visits to the United States. His research interests are in the rheooptical properties of semicrystalline and heterophase polymers.

Manuscript received February 17, 1977; revision received July 22, and accepted August 23, 1977.

Evaluation of Parameters for Nonlinear Thermodynamic Models

T. F. ANDERSON
D. S. ABRAMS
and
E. A. GRENS II

Department of Chemical Engineering
University of California
Berkeley, California 94720

A method based on the maximum likelihood principle has been developed for the determination of model parameters from experimental data when all the measured variables are subject to error. In addition to the best estimates of the parameters, this method also yields information useful in selection of appropriate models and evaluation of the accuracy of the data. Application of the method is illustrated in the reduction of binary vapor-liquid equilibrium data.

SCOPE

The estimation of parameters in theoretical and semi-empirical mathematical models from experimental data is an important requirement in many fields of science and engineering. Such models offer useful means of summarizing large amounts of data, allow for interpolation of data and extensions beyond regions in which measurements have been made, and provide insight into fundamental physical or chemical phenomena. Most of the phenomenological models are nonlinear in their adjustable parameters. The best estimates of these parameters can be obtained from a formalized method which correctly treats the statistical behavior of the errors associated with all experimental observations.

It is usually assumed that there are two types of measurement errors: systematic and random. The former are due to an inherent bias in the measurement procedure resulting in a consistent deviation of the observable from its true value. Usually, an experimenter's skill and experience provide the only means of consistently detecting and avoiding systematic errors. Random or statistical errors are assumed to be the result of a large number of small disturbances. Such errors have been found to be distributed ac-

D. S. Abrams is with Buffalo Salt Works and Packing Company, PO Box 3489, Johannesburg, South Africa.

0001-1541/78/9841-0020/$01.25

cording to simple laws that make it possible to treat them with statistical methods. Here, only the effect of these random errors on the parameter estimation process is considered. A major consequence of these random errors is the corresponding presence of errors or uncertainties in the estimated parameters. Because of these errors in the data, and also because of inaccuracies in the model, it is not possible for a model to represent the experimental data exactly. However, a method of parameter estimation which correctly utilizes all the pertinent information available will give a best fit of the model to the data and minimize the parameter uncertainty. This includes taking into account the nature and magnitude of the random errors in all the observables. Such a method will give not only the parameters but also provide estimates of their uncertainties.

Parameter estimation methods are well established for treating data where the model employed is linear in the unknown parameters (Draper and Smith, 1966; Daniel and Wood, 1971). In this case, the least-squares criterion is stated as: the best estimates of the parameters are those which minimize the weighted sum of the squared deviations of the observed values from their calculated values." This same criterion is also utilized, in principle, when the model is nonlinear in its parameters. A discussion of weighted least squares as applied to nonlinear equations is given by Clifford (1973). The least-squares method distinguishes between two types of variables: the independent variables, or factors, and the dependent variables, or responses. A basic assumption of the least-squares criterion is that the values of the independent variables are known without error. This assumption fails to account for the statistical properties of the independent variables, arbitrarily assigning to them a zero standard deviation; therefore, it does not utilize all available information in estimating the parameters.

Demming (1943) formulated the general problem of parameter estimation in models, taking into account the error in all measured variables. An exact solution satisfying his criteria for parameter estimation was nearly impossible at the time; thus he was primarily concerned with making simple adjustments which would give fairly accurate estimates of the desired parameters. Exact solutions satisfying his criteria have since been proposed by a number of authors (York, 1966; Williamson, 1968; Southwell, 1969; O'Neil et al., 1969); however, these methods were concerned only with fitting data by straight lines or polynomials. More recently, Powell and McDonald (1972) and Britt and Luecke (1973) have given algorithms for parameter estimation that take all measurement errors into account. Their methods are completely general in that they do not distinguish between dependent and independent variables; however, in order to converge they do require good initial estimates of the parameters. Fabries and Renon (1975) and Péneloux et al. (1976) have presented related methods. With the exception of these two latter methods, the algorithms are not sufficiently general to treat problems in which the measured variables are related by more than a single equation or constraint. The number of simultaneous constraints that apply is directly related to the number of simultaneous properties measured and the number of properties needed to specify the system. For example, in fitting binary vapor-liquid equilibrium data, four properties are often measured: pressure, temperature, and both phase compositions. Only two of these properties are needed to specify the system, and thus two constraints or relations are required if all experimental data is to be properly utilized. Fabries and Renon discuss the use of two constraints in representing vapor-liquid equilibrium data; however, their maximum likelihood estimate is based on estimated variances of the variables rather than estimated true values for each measurement. Southwell (1976) recently presented a discussion of generalized least-squares parameter estimation methods for nonlinear functions that examines many of the problems involved in these methods.

The present work concerns a parameter estimation procedure based on the maximum likelihood principle; in this respect it is similar to that of Britt and Luecke. However, the procedure is applicable to cases in which there are two constraining equations relating the measured variables (the case of a single constraint equation is then just a special case of the two constraint method). The present method has the additional advantage of being easily fitted with a step-limiting procedure that assures superior convergence behavior.

CONCLUSIONS AND SIGNIFICANCE

The parameter estimation algorithm we have developed, based on the maximum likelihood principle, converges rapidly for almost any initial estimates of the parameters. The rapid convergence is due in part to the similarity to a Gauss-Newton iteration method and in part to the successful application of a step-limiting procedure like that of Law and Bailey (1963).

In addition to the best estimates for the parameters, the procedure yields information useful in analysis of both the model and the data. Part of this additional information is obtained in the form of the variance-covariance matrix of the estimated parameters. This matrix gives an estimate of the uncertainties in the estimated parameters, which may be used to determine the significance of the parameters or to estimate uncertainties in properties calculated using the parameters. Such information is important to the engineer in determining the most appropriate model to be used in design calculations and estimating the uncertainties in these calculations.

Additional information is obtained in the form of estimates of the true values of all measured variables. The deviations of the measured variables from the estimates of these true values are called residuals. These are useful for comparison of different models, detection of obviously bad data points, and discrimination between systematic error and lack of fit of the model employed. Even when data are scarce, examination of residuals obtained in the parameter estimation process, plotted vs. other system variables, can help determine how well the model represents the data. The interpretation and use of this additional information derived from vapor-liquid equilibrium data reduction has been discussed by Fabries and Renon (1975) and Anderson et al. (1976).

The maximum likelihood principle provides a powerful and versatile criterion for estimation of model parameters from experimental data. The method is based on known statistical assumptions and allows application of straight-

forward statistical tests in evaluation of the results of the parameter estimation process. The criterion is generally applicable to a wide variety of data reduction tasks. Here, by way of illustration, it has been applied to estimate the van Laar parameters for the acetone/methanol system from published data containing significant random error.

THEORETICAL DEVELOPMENT

The problem considered here is that of estimation of parameters in nonlinear models from experimental data where the number of data points exceeds the number of parameters to be estimated. This is a common situation and arises in almost every field of science and engineering.

Typically, data are obtained for a set of N experiments or observations. For each experiment i, there are M measured variables, denoted by the vector $\mathbf{x}_i^m$. These measured variables are subject to random error. If each experiment were replicated, the average values of the measured variables should approach some true values $\mathbf{x}_i$, and the measured values would be distributed in some manner about these true values. It is assumed that there are two equations (from now on referred to as constraints) which relate the true values corresponding to the measured variables at each observation and which include the parameters being estimated. These constraints can be written as

$$F(\mathbf{x}_i, \boldsymbol{\theta}) = 0 \qquad (1a)$$

$$G(\mathbf{x}_i, \boldsymbol{\theta}) = 0 \qquad (1b)$$

where $\boldsymbol{\theta}$ is a vector of L undetermined parameters. $\mathbf{x}_i$ is also unknown, since it represents the error free values of the observed variables for one experiment. These values for all N experiments can be combined in a single vector $\mathbf{X}$ of length NM. Because the number of constraint equations $2N$ exceeds the number of parameters L, the equations cannot be satisfied in general for the measured variable values $\mathbf{X}^m$. Rather, the true values $\mathbf{X}$ satisfying Equation (1), together with the parameters, must be selected to satisfy some optimality criterion. The criterion used here is based on the maximum likelihood principle.

Maximum Likelihood Principle

The use of the maximum likelihood method as a parameter estimator is not new. Its general form was introduced by Fisher (1922). A more conventional application to regression analysis is discussed in many texts, for example, Kreyszig (1970), and a detailed analysis of its qualities is given by Crámer (1961). Bard and Lapidus (1968) have thoroughly discussed its application to model identification and parameter determination for kinetic data. However, in these applications it has been assumed that only the variables designated as dependent are subject to error.

In effect, the maximum likelihood principle states that the parameters should be so chosen as to make the experimental observations appear to be most likely when taken as a whole. Application of the method requires that a form be specified for the probability density function of the variables. It is known that the distribution of a measured variable about its true values is usually closely approximated by the normal distribution. For a normal distribution the corresponding multivariate joint probability density function (Morrison, 1967) is given by

$$\Phi_i(\mathbf{x}_i^m, \mathbf{x}_i) = \frac{1}{(2\pi)^{1/2M}\,|\boldsymbol{\beta}_i|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{x}_i^m - \mathbf{x}_i)^T \boldsymbol{\beta}_i^{-1} (\mathbf{x}_i^m - \mathbf{x}_i) \right\} \qquad (2)$$

The function Φ_i gives the probability of obtaining the observations $\mathbf{x}_i^m$ given the true values $\mathbf{x}_i$ for experiment i. Here, $\boldsymbol{\beta}_i$ is the $M \times M$ variance-covariance matrix of the measured variables. The off-diagonal elements of $\boldsymbol{\beta}_i$ represent the covariances between the different measured variables. Almost always these can be assumed to be zero. (This will not be true if there is any coupling between the measurement methods, for example, measurements of overlapping peaks on a chromatogram.) The solution technique used here requires that they be zero, in which case $\boldsymbol{\beta}_i$ becomes a diagonal matrix whose elements are variances of the variables. It is necessary that these variances be known or estimated. Ideally, they are obtained from replicated experiments, but they may be estimated from experience or from knowledge of a particular type of experimental apparatus.

If it is assumed that the experiments are independent, the likelihood function can be formulated as the product of the joint probability density functions of the N experiments:

$$\mathcal{L}(\mathbf{X}^m, \mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \Phi_i(\mathbf{x}_i^m, \mathbf{x}_i) \qquad (3)$$

$\mathcal{L}$, the likelihood function, is the probability of observing the whole set of measurements $\mathbf{X}^m$. For a given set of N observations, the value of the likelihood function depends, of course, on the true values of the measured variables $\mathbf{X}$. Since these true values must satisfy the constraints (1), the value is also dependent on the model parameters. The basic assumption of the maximum likelihood principle states that the set of measurements which have been obtained is the most likely set. Therefore, the best estimates of the parameters and the true values of the measured variables are those which maximize the likelihood function subject to the model constraints.

A solution is more conveniently obtained for $\ln \mathcal{L}$ (a maximum of $\ln \mathcal{L}$ corresponds to a maximum of $\mathcal{L}$). Taking the natural logarithm of Equation (3) and substituting into it the density functions (2), one obtains

$$\ln \mathcal{L} = \ln\left[\frac{1}{(2\pi)^{N/2M}} \prod_{i=1}^{N} \left(\frac{1}{|\boldsymbol{\beta}_i|^{1/2}} \right) \right] - \frac{1}{2} \sum_{i=1}^{N} (\mathbf{x}_i^m - \mathbf{x}_i)^T \boldsymbol{\beta}_i^{-1} (\mathbf{x}_i^m - \mathbf{x}_i) \qquad (4)$$

Inspection of Equation (4) shows that maximizing $\ln \mathcal{L}$ will be equivalent to minimizing

$$S = \sum_{i=1}^{N} (\mathbf{x}_i^m - \mathbf{x}_i)^T \boldsymbol{\beta}_i^{-1} (\mathbf{x}_i^m - \mathbf{x}_i) \qquad (5)$$

subject to the constraints given by Equation (1).

Constraint Linearization

The usual method of accomplishing this minimization, where the variables are connected by one or more constraints, is to use a procedure which introduces extraneous parameters, known as Lagrange multipliers. An explana-

tion of this technique is given by Widder (1961). The introduction of the extraneous parameters allows the function to be minimized by equating partial derivatives to zero regardless of which variables are considered independent. However, the relations obtained after taking the partial derivatives still represent a large system of nonlinear equations, which must be solved iteratively. This is the method used by Britt and Luecke. Unfortunately, this method requires good initial estimates of the parameters, is not easily adapted with a step-limiting procedure, and has only been described adequately for the single constraint case.

The procedure used here (similar to that used by Abrams, 1974) linearizes the constraint equations applicable to the system at estimates of the parameters and of the true values for the variables before their substitution into the minimization function. Then the conditions for a minimum are applied, and the equations are solved for new estimates of the variables and parameters. This procedure is repeated until convergence is achieved.

The constraints (1) are first solved, in principle, for two of the system variables in terms of the remaining variables:

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}) \tag{6a}$$

$$\mathbf{Z} = \mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) \tag{6b}$$

$\mathbf{Y}$ and $\mathbf{Z}$ are N vectors that are the true values of two of the measured variables; they can be considered to be dependent variables. $\mathbf{X}$ now becomes an NK vector of the true values of what can be considered to be the independent variables, where K is equal to the number of these independent variables (two less than M). Every set of K elements of $\mathbf{X}$ forms a vector $\mathbf{x}_i$ of these independent variables corresponding to experiment i.

The function to be minimized (5) is now rewritten in terms of the newly defined dependent and independent variables. The assumption that the errors in the measured variables are uncorrelated, that is, that the β_i's are diagonal, allows us to write

$$S = (\mathbf{X} - \mathbf{X}^m)^T \boldsymbol{\lambda} (\mathbf{X} - \mathbf{X}^m) + (\mathbf{Y} - \mathbf{Y}^m)^T \boldsymbol{\gamma} (\mathbf{Y} - \mathbf{Y}^m) + (\mathbf{Z} - \mathbf{Z}^m)^T \boldsymbol{\delta} (\mathbf{Z} - \mathbf{Z}^m) \tag{7}$$

where the superscript m again signifies the experimentally measured value. $\boldsymbol{\lambda}$ is an $NK \times NK$ diagonal matrix whose elements are the reciprocals of the variances σ_X^2 in the variables $\mathbf{X}$. $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are $N \times N$ diagonal matrices whose elements are the reciprocals of the variances σ_Y^2 and σ_Z^2 in the variables $\mathbf{Y}$ and $\mathbf{Z}$, respectively.

By elimination of $\mathbf{Y}$ and $\mathbf{Z}$ with the constraint Equations (6), S can be written as a function of only $\mathbf{X}$ and $\boldsymbol{\theta}$. Necessary conditions for S to be a minimum require that

$$\frac{\partial S}{\partial \mathbf{X}} = \mathbf{0} \tag{8a}$$

$$\frac{\partial S}{\partial \boldsymbol{\theta}} = \mathbf{0} \tag{8b}$$

Before applying these conditions, we first linearize the constraint Equations (6) by a first-order Taylor's series expansion about the most recent estimates of the parameters and of the true values of the independent variables. Equations (6) become

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}, \boldsymbol{\theta}) \cong \mathbf{f}^r + \mathbf{f}_X(\mathbf{X} - \mathbf{X}^r) + \mathbf{f}_\theta(\boldsymbol{\theta} - \boldsymbol{\theta}^r) \tag{9}$$

and

$$\mathbf{Z} = \mathbf{g}(\mathbf{X}, \boldsymbol{\theta}) \cong \mathbf{g}^r + \mathbf{g}_X(\mathbf{X} - \mathbf{X}^r) + \mathbf{g}_\theta(\boldsymbol{\theta} - \boldsymbol{\theta}^r) \tag{10}$$

The functions $\mathbf{f}^r$ and $\mathbf{g}^r$ are equivalent to $\mathbf{f}(\mathbf{X}^r, \boldsymbol{\theta}^r)$ and $\mathbf{g}(\mathbf{X}^r, \boldsymbol{\theta}^r)$, respectively, where the superscript r refers to the previous iteration number in the procedure (which yielded the values $\mathbf{X}^r$ and $\boldsymbol{\theta}^r$). $\mathbf{f}_X$ and $\mathbf{g}_X$ are sparse, $N \times NK$ Jacobian partial derivative matrices of the two constraint functions taken with respect to the independent variables. $\mathbf{f}_\theta$ and $\mathbf{g}_\theta$ are the Jacobian partial derivative matrices of the two constraint functions taken with respect to the parameters; each is an $N \times L$ matrix. All the Jacobian matrices are evaluated at the point $(\mathbf{X}^r, \boldsymbol{\theta}^r)$ and can be evaluated either analytically or numerically.

The linearized equations for the constraints, (9) and (10), can be substituted into Equation (7); in the resulting expression the only unknowns are the parameters and the true values of the independent variables.

$$\begin{aligned} S = {} & (\mathbf{X} - \mathbf{X}^m)^T \boldsymbol{\lambda} (\mathbf{X} - \mathbf{X}^m) \\ & + [\mathbf{f}^r - \mathbf{Y}^m + \mathbf{f}_X^r(\mathbf{X} - \mathbf{X}^r) + \mathbf{f}_\theta^r(\boldsymbol{\theta} - \boldsymbol{\theta}^r)]^T \\ & \boldsymbol{\gamma}[\mathbf{f}^r - \mathbf{Y}^m + \mathbf{f}_X^r(\mathbf{X} - \mathbf{X}^r) + \mathbf{f}_\theta^r(\boldsymbol{\theta} - \boldsymbol{\theta}^r)] \\ & + [\mathbf{g}^r - \mathbf{Z}^m + \mathbf{g}_X^r(\mathbf{X} - \mathbf{X}^r) + \mathbf{g}_\theta^r(\boldsymbol{\theta} - \boldsymbol{\theta}^r)]^T \\ & \boldsymbol{\delta}[\mathbf{g}^r - \mathbf{Z}^m + \mathbf{g}_X^r(\mathbf{X} - \mathbf{X}^r) + \mathbf{g}_\theta^r(\boldsymbol{\theta} - \boldsymbol{\theta}^r)] \end{aligned} \tag{11}$$

When the conditions (8) for a minimum of S are applied to Equation (11), two matrix equations are obtained:

$$\begin{aligned} & \boldsymbol{\lambda}(\mathbf{X}^{r+1} - \mathbf{X}^m) + \mathbf{f}_X{}^T \boldsymbol{\gamma}[\mathbf{f}^r - \mathbf{Y}^m + \mathbf{f}_X(\mathbf{X}^{r+1} - \mathbf{X}^r) \\ & + \mathbf{f}_\theta(\boldsymbol{\theta}^{r+1} - \boldsymbol{\theta}^r)] + \mathbf{g}_X{}^T \boldsymbol{\delta}[\mathbf{g}^r - \mathbf{Z}^m + \mathbf{g}_X(\mathbf{X}^{r+1} - \mathbf{X}^r) \\ & + \mathbf{g}_\theta(\boldsymbol{\theta}^{r+1} - \boldsymbol{\theta}^r)] = \mathbf{0} \end{aligned} \tag{12}$$

and

$$\begin{aligned} & \mathbf{f}_\theta{}^T \boldsymbol{\gamma}[\mathbf{f}^r - \mathbf{Y}^m + \mathbf{f}_X(\mathbf{X}^{r+1} - \mathbf{X}^r) + \mathbf{f}_\theta(\boldsymbol{\theta}^{r+1} - \boldsymbol{\theta}^r)] \\ & + \mathbf{g}_\theta{}^T \boldsymbol{\delta}[\mathbf{g}^r - \mathbf{Z}^m + \mathbf{g}_X(\mathbf{X}^{r+1} - \mathbf{X}^r) \\ & + \mathbf{g}_\theta(\boldsymbol{\theta}^{r+1} - \boldsymbol{\theta}^r)] = \mathbf{0} \end{aligned} \tag{13}$$

These represent a system of $NK + L$ linear equations. The superscript $r + 1$ is used to designate the new estimates of parameters and of true values of the independent variables to be determined. In application of this procedure, we first solve for $(\mathbf{X}^{r+1} - \mathbf{X}^r)$ in Equation (12). This expression is substituted into Equation (13), which allows a unique solution for $(\boldsymbol{\theta}^{r+1} - \boldsymbol{\theta}^r)$. The necessary intermediate steps are given in Appendix A. The calculations can be summarized as

$$\boldsymbol{\Delta\theta} = \boldsymbol{\theta}^{r+1} - \boldsymbol{\theta}^r = -[\mathbf{T} - \mathbf{R}^T\mathbf{D}^{-1}\mathbf{R}]^{-1}[\mathbf{U} - \mathbf{R}^T\mathbf{D}^{-1}\mathbf{Q}] \tag{14}$$

and

$$\boldsymbol{\Delta X} = \mathbf{X}^{r+1} - \mathbf{X}^r = -\mathbf{D}^{-1}[\mathbf{Q} + \mathbf{R}\boldsymbol{\Delta\theta}] \tag{15}$$

where

$$\mathbf{D} = \boldsymbol{\lambda} + \mathbf{f}_X{}^T \boldsymbol{\gamma} \mathbf{f}_X + \mathbf{g}_X{}^T \boldsymbol{\delta} \mathbf{g}_X \tag{16a}$$

$$\mathbf{R} = \mathbf{f}_X{}^T \boldsymbol{\gamma} \mathbf{f}_\theta + \mathbf{g}_X{}^T \boldsymbol{\delta} \mathbf{g}_\theta \tag{16b}$$

$$\mathbf{T} = \mathbf{f}_\theta{}^T \boldsymbol{\gamma} \mathbf{f}_\theta + \mathbf{g}_\theta{}^T \boldsymbol{\delta} \mathbf{g}_\theta \tag{16c}$$

$$\mathbf{U} = \mathbf{f}_\theta{}^T \boldsymbol{\gamma} \boldsymbol{\Delta} \mathbf{Y}^m + \mathbf{g}_\theta{}^T \boldsymbol{\delta} \boldsymbol{\Delta} \mathbf{Z}^m \tag{16d}$$

$$\mathbf{Q} = \boldsymbol{\lambda} \boldsymbol{\Delta} \mathbf{X}^m + \mathbf{f}_X{}^T \boldsymbol{\gamma} \boldsymbol{\Delta} \mathbf{Y}^m + \mathbf{g}_X{}^T \boldsymbol{\delta} \boldsymbol{\Delta} \mathbf{Z}^m \tag{16e}$$

and

$$\boldsymbol{\Delta} \mathbf{X}^m = \mathbf{X}^r - \mathbf{X}^m \tag{17a}$$

$$\boldsymbol{\Delta} \mathbf{Y}^m = \mathbf{Y}^r - \mathbf{Y}^m \tag{17b}$$

$$\boldsymbol{\Delta} \mathbf{Z}^m = \mathbf{Z}^r - \mathbf{Z}^m \tag{17c}$$

Appendix B gives the derivation of the set of equations similar to Equations (14) and (15) for the one constraint case.

IMPLEMENTATION

Equations (14) and (15) constitute the basis for the algorithm used to determine the maximum likelihood estimates of the parameters. With the estimates $\mathbf{X}^r$ and $\boldsymbol{\theta}^r$, the functions $\mathbf{f}^r$ and $\mathbf{g}^r$ and the Jacobian partial derivative matrices $\mathbf{f}_X$, $\mathbf{g}_X$, $\mathbf{f}_\theta$, and $\mathbf{g}_\theta$ can be evaluated. These quantities are used in Equation (14)to calculate $\boldsymbol{\Delta\theta}$, and then with $\boldsymbol{\Delta\theta}$ in Equation (15) to give $\boldsymbol{\Delta X}$. From these values, new values $\mathbf{X}^{r+1}$ and $\boldsymbol{\theta}^{r+1}$ are directly determined. This procedure is repeated until convergence is achieved.

Computationally, the algorithm poses no serious problems. However, considerable matrix manipulation, including the inversion of the $NK \times NK$ matrix $\mathbf{D}$, is required. This is facilitated by the fact that $\mathbf{D}$ is a symmetric block-diagonal matrix. The individual submatrices have dimensions $K \times K$ and are easily inverted individually by any of a number of efficient symmetric matrix inversion procedures.

The convergence properties of this parameter estimator are similar to those of the standard Newton methods. When good estimates of the parameters and true values of the measured variables are used, convergence is very rapid. However, with poor initial estimates, convergence can be very slow or may not occur. Satisfactory initial estimates for the true values of the measured variables are the corresponding measured values $\mathbf{X}^m$. Good initial estimates for the parameters are more difficult to obtain. In fact, such estimates rely heavily on previous experience, judgment, and chance. A means of insuring convergence for almost any initial estimate of the parameters is required. Quite often, for a Newton iteration method, changes made in the magnitude of a correction (step length) for a given direction of correction (gradient), according to some criterion, will improve convergence. Such a procedure is often called step-limiting.

One such step-limiting method is described by Law and Bailey (1963). Their method applied here leads to convergence with almost any initial parameter estimates. This procedure always chooses the sign of the corrections in $\boldsymbol{\Delta\theta}$ and $\boldsymbol{\Delta X}$ so that locally the linearized sum of the squared deviations decreases. The magnitudes of the corrections, $\boldsymbol{\Delta\theta}$ and $\boldsymbol{\Delta X}$, are decreased until the new sum of the squared deviations S is less than the previous value. Finally, the magnitude of the corrections are further decreased if it appears that their direction is normal to a valley in the criterion surface.

COMPARISON WITH OTHER METHODS

We mentioned earlier that several authors have previously described other approaches to estimation of parameters in nonlinear models while satisfying Demming's criteria. The algorithm proposed here is similar in many ways to that of Britt and Luecke (1973). Both methods are based on the maximum likelihood principle, and both use a Gauss-Newton type of iteratiion. Britt and Luecke's method could be considered more general in that it can be applied to situations where covariances between the measured variables are nonzero, if they are known. Their algorithm is restrictive in that it is only derived for the case of a single constraint. Britt and Luecke have also pointed out that their algorithm is not easily modified for use with a step limiting procedure. Hence, they require very good initial estimates of the parameters to insure convergence; these estimates are obtained by using Demming's method as a starting point for their algorithm.

The method proposed here can be used when there are two constraints as well as for a single constraint. Although the approach, in its derivation, represented two variables in terms of the remaining variables and parameters, in fact, no such explicit form is required in the implementation. The procedure we have developed cannot treat nonzero covariances in its present implementation; this is not a particularly restrictive limitation because, in the majority of cases, these covariances are either zero or cannot be estimated in any reasonable manner. Finally, the present method is readily modified with common step-limiting procedures, giving it superior convergence properties to that of Britt and Luecke.

The algorithm proposed by Powell and MacDonald is somewhat different in nature. It is of the Newton-Raphson type and hence requires the calculation of the second derivatives of the criterion function. Their method is also only derived for a single constraint, and as with the algorithm of Britt and Luecke, is not easily modified for use with a step-limiting procedure. Thus, their algorithm has convergence problems unless initial parameter estimates are quite good. To overcome this problem, they, too, use Demming's method to obtain starting estimates for their algorithms.

More recently, Fabries and Renon (1975) have developed a method for estimation of parameters in equations representing vapor-liquid equilibrium data, together with an implementing algorithm (Renon, 1977). However, their method is not directly based on the satisfaction of Demming's criteria. As stated by Britt and Luecke (1973) and Southwell (1976), simultaneous convergence on both the parameters and true values of the measured variables are necessary to satisfy these criteria. Instead, Fabries and Renon first determine the optimum parameters and the variances of the measured variables by minimizing a residual function based on the difference between the calculated and measured activity coefficients. Once the parameters and variances are estimated, they then calculate estimates of the true values of the measured variables at each experiment by minimization of residuals for each experiment separately. The method of Péneloux et al. (1976) has a similar basis and a somewhat more general formulation.

ERROR ANALYSIS

One of the important reasons for fitting model equations to experimental data is to obtain a representation that can be used confidently for predictions. The confidence we can place in the predictions depends on the confidence we can place in the data and in the model. Therefore, the method of estimation of parameters should also provide measures of reliability for the predictions. This problem is closely related to that of estimation of uncertainties in the measured variables.

As mentioned earlier, estimates of the variances of each of the observed variables are required. Because of the nature of the maximum likelihood estimation procedure, the absolute values of the variances need not be known; instead, only relative values need be assigned. These are often called weights. The weights corresponding to the variables X, Y, and Z are

$$w_X = \frac{\sigma^2}{\sigma_X^2} \tag{18a}$$

$$w_Y = \frac{\sigma^2}{\sigma_Y^2} \tag{18b}$$

$$w_Z = \frac{\sigma^2}{\sigma_Z^2} \tag{18c}$$

where the common constant σ^2 is called the variance of the observables of unit weight. The estimated parameters are independent of the value of σ^2 and are determined

without its knowledge. However, for an error analysis, it is necessary that σ^2 be evaluated. As discussed by Demming (1943), the sum of the weighted squared deviations S, defined by Equation (7), divided by σ^2 has a χ^2 distribution with $\nu = N - L$ degrees of freedom:

$$\chi^2 = \frac{S}{\sigma^2} \tag{19a}$$

Since the mean, or expected value, of χ^2 is equal to the number of degrees of freedom, the value of σ^2 may be estimated from Equation (19*a*)

$$s^2 = E(\sigma^2) = \frac{S}{N - L} \tag{19b}$$

The quantity s^2 calculated with Equation (19*b*) is useful in itself. For example, if two different models are used with a given set of data, most probably two different values, s_1^2 and s_2^2, will be obtained for this variance. If it is found that s_2^2 is less than s_1^2, then the second model provides a somewhat better representation of the data. Whether the difference in representation is statistically significant must be determined. To test the hypothesis $\sigma_1^2 = \sigma_2^2$ (that the models are equivalent in their representation of the data), we calculate the ratio

$$F = \frac{s_1^2}{s_2^2} \tag{20}$$

which will be distributed as the F distribution (Brownlee, 1965), assuming errors due to lack of fit are normally distributed. At the significance level α, the value can be compared to tabulated $F_{\alpha/2}(\nu, \mu)$ (ν and μ are the degrees of freedom for the two models). If F is greater than $F_{\alpha/2}(\nu, \mu)$, we reject the null hypothesis, concluding that the second model is better than the first. In a similar manner with several sets of data and a single model, we can determine if one set of data is significantly better represented by the model than another. These and other tests are discussed by Crow et al. (1960) and Brownlee (1965).

Parameter Variance-Covariance Matrix

The variance estimated by Equation (19*b*) is further useful in estimation of the errors in the calculated parameters. Britt and Luecke (1973) have shown, for one constraint, how the variance-covariance matrix $\boldsymbol{\Sigma}$ of the parameters can be estimated. In a similar manner, it can be shown that $\boldsymbol{\Sigma}$ for two constraints is given by

$$\boldsymbol{\Sigma} = s^2(\mathbf{T} - \mathbf{R}^T\mathbf{D}^{-1}\mathbf{R})^{-1} \tag{21}$$

The matrix inversion indicated in Equation (21) has already been accomplished in the parameter estimation process [evaluation of $\boldsymbol{\Delta\theta}$ in Equation (14)], and the final inverted matrix is retained for this determination. The sum of the weighted squared deviations S is also calculated at each iteration and is available after convergence is achieved; this is used to calculate s^2 with Equation (19*b*).

Subject to the linear approximations that have been made, the diagonal elements of this matrix represent the variances of the corresponding parameters. The square roots of these variances are estimates of the standard errors in the parameters s_{θ_i} and, in effect, are a measure of the uncertainties in the estimated parameters. They can be used to test the significance of any of the parameters. For example, if s_{θ_i} is the estimated standard error in θ_i, we can test the hypothesis that θ_i is zero. We calculate

$$t_i = \frac{\theta_i}{s_{\theta_i}} \tag{22}$$

This is compared with the value $t_{\alpha/2,\nu}$, determined from a tabulated Student t distribution, for the appropriate number of degrees of freedom and at the desired level of significance. If $|t_i|$ exceeds $t_{\alpha/2,\nu}$, we can reject the hypothesis (concluding that the particular parameter is significantly different from zero).

The off-diagonal elements of the variance-covariance matrix represent the covariances between parameters. These allow us to calculate the corresponding correlation coefficients. When parameters are completely independent, their correlation coefficient is zero. As parameters become more correlated, the correlation coefficient approaches a value of ± 1. With highly correlated parameters, it becomes very difficult to determine them uniquely. In such a case, where the parameters are strongly related, a single parameter, which is some combination of the two, could represent the data just as well.

Finally, the variance-covariance matrix can be used to calculate the uncertainty in any property that is a function of the estimated parameters. If $H(\boldsymbol{\theta})$ is a property we wish to calculate, the variance in this property σ_h^2 is given by

$$\sigma_h^2 = \mathbf{h}_\theta^T \boldsymbol{\Sigma}\, \mathbf{h}_\theta \tag{23}$$

where $\mathbf{h}_\theta$ is an L vector of the partial derivatives of $h(\boldsymbol{\theta})$ with respect to the parameters $\boldsymbol{\theta}$. Such a relation provides a direct means for determination of the expected errors in predictions made with the fitted model.

Residual Analysis

In the procedure developed here, the use of all observed data yields best estimates for the parameters and, at the same time, provides estimates of the true values of all measured variables. These estimated true values can be used to determine the residuals, that is, their differences from the corresponding experimentally measured values. The results of a statistical analysis of these residuals can indicate how well the model represents the data and how precise the data is.

If, however, there are only a few data points, as is often the case with engineering data, then mere examination of the residuals for trends, when plotted vs. other system variables, may provide valuable information. Quite often these plots can indicate at a glance excessive experimental error, systematic error, or lack of fit. Data points which are obviously bad can also be readily detected.

The maximum likelihood method for estimation of parameters can provide a wealth of information useful in assessing the model and the data. This information is available with little or no additional effort once the parameters have been estimated.

VAPOR-LIQUID EQUILIBRIUM DATA REDUCTION

Prediction of vapor-liquid phase equilibrium is an important aspect in the design and successful operation of many separation processes. Such predictions often utilize a semiempirical model with several adjustable parameters to represent liquid phase thermodynamics. Estimates of the model parameters are obtained from binary vapor-liquid equilibrium data.

For binary vapor-liquid equilibrium, the measured properties are pressure P, temperature T, liquid phase mole fraction x, and vapor phase mole fraction y. At equilibrium, the fugacity of each component in the vapor phase is equal to its corresponding fugacity in the liquid phase; this requirement yields the thermodynamic relations (Prausnitz, 1969)

$$y_1\phi_1 P = x_1\gamma_1 {}^0f_1{}^0 \exp\left(\frac{\bar{v}_1 P}{RT}\right) \tag{24a}$$

$$y_2\phi_2 P = x_2\gamma_2{}^0 f_2{}^0 \exp\left(\frac{\bar{v}_2 P}{RT}\right) \qquad (24b)$$

The use of a zero pressure reference state is for convenience and follows the conventions used by Prausnitz et al. (1967). The exponential term corrects the activity coefficient from zero pressure to the system pressure (this correction is often quite small).

The activity coefficients in liquid mixtures can be calculated from a model which represents the excess Gibbs energy

$$RT \ln \gamma_i{}^0 = \left(\frac{\partial n_t g^E}{\partial n_i}\right)_{P,T,n_j} \qquad (25)$$

and

$$\frac{g^E}{RT} = f(x, T, \theta) \qquad (26)$$

Equation (26) indicates that the excess Gibbs energy is a function of composition, temperature, and some unknown parameters. By Equation (27) the activity coefficients will be a function of the same variables, including the unknown parameters. It is these parameters that we want to estimate from binary vapor-liquid equilibrium data.

Constraint Equations

According to the Gibbs phase rule for a two-phase binary system, two variables must be specified to describe the system completely. This implies that two of the measured variables obtained from vapor-liquid equilibrium data can be considered independent; the remaining two variables can then be considered dependent. The desired constraints are just the functional relations between these variables given by Equations (24) through (26). Any two of the measured properties may be chosen as independent variables, but, for convenience, x and T have been chosen here as the independent variables. We let

$$\gamma_i = \gamma_i{}^0 \exp\left(\frac{\bar{v}_i P}{RT}\right) \qquad (27)$$

utilize the relation

$$y_1 + y_2 = 1 \qquad (28)$$

and solve Equations (24a) and (24b) for the total pressure. The desired equation is

$$P = \frac{x_1\gamma_1 f_1{}^0}{\phi_1} + \frac{x_2\gamma_2 f_2{}^0}{\phi_2} \qquad (29)$$

The other constraint is obtained by solving Equations (24a) and (24b) for y_1, again using Equations (27) and (28):

$$y_1 = \frac{1}{1 + \dfrac{x_2\gamma_2 f_2{}^0\phi_1}{x_1\gamma_1 f_1{}^0\phi_2}} \qquad (30)$$

Since ϕ_i is only a weak function of y_1 and P, Equations (29) and (30) represent constraints which allow us to calculate P and y_1 as functions of x and T.

Two cases can be treated. First, when all four variables P, T, x, and y are measured, both constraints are used in the maximum likelihood estimation of the parameters. Second, when only P, T, and x are measured, Equation (29) is used as the only constraint; the maximum likelihood estimates for the parameters are determined by using the simplified equations in Appendix B for the single constraint case. Barker's method of data reduction (Barker, 1953) also utilizes the total pressure Equation (29). However, in this application it is assumed that the measured values of x and T are known without error. This method is often used when vapor compositions have been measured, and yet these measurements do not enter into the parameter estimation process. This is not the case when we use the maximum likelihood method described here. All the pertinent data are used to determine the best estimates of the model parameters.

Application to Parameter Evaluation

Analysis of the isobaric data of Othmer (1928) for the system acetone/methanol illustrates the application of the algorithm for analysis of vapor-liquid equilibrium data. For the illustration, the van Laar equation is used to express the excess Gibbs energy of the liquid phase, although better thermodynamic models are available (Abrams and Prausnitz, 1975). The activity coefficients

Table 1. Vapor-Liquid Equilibrium Data Reduction for Acetone(1)-Methanol(2) System (Othmer, 1928)

Calculated parameters for van Laar equation

$A_{12} = 0.85749$
$A_{21} = 0.68083$

Covariance matrix

$$\begin{bmatrix} 0.00528 & -0.00213 \\ -0.00213 & 0.00187 \end{bmatrix}$$

Correlation coefficient matrix

$$\begin{bmatrix} 1.000 & -0.678 \\ -0.678 & 1.000 \end{bmatrix}$$

Table 2. Measured Variables and Estimates of Their True Values for Acetone(1)/Methanol(2) System (Othmer, 1928)

	Pressure, mm Hg			Temperature, °C			Liquid composition, x_1			Vapor composition, y_1		
	Meas	Calc	Dev	Meas	Calc	Dev	Meas	Calc	Dev	Meas	Calc	Dev
1	755.00	753.85	1.15	61.90	62.22	−0.32	0.05530	0.05996	−0.00466	0.14420	0.15024	−0.00604
2	755.00	754.88	0.12	58.30	58.32	−0.02	0.20460	0.20779	−0.00319	0.37220	0.36056	0.01164
3	755.00	754.99	0.01	56.80	56.79	0.01	0.31270	0.31623	−0.00353	0.47370	0.45428	0.01942
4	755.00	754.73	0.27	55.60	55.67	−0.07	0.44050	0.44131	−0.00081	0.54310	0.53989	0.00321
5	755.00	755.73	−0.73	55.40	55.20	0.20	0.52830	0.52918	−0.00088	0.60480	0.59529	0.00951
6	755.00	755.71	−0.71	54.90	54.70	0.20	0.70710	0.71017	−0.00307	0.73350	0.71615	0.01735
7	755.00	754.64	0.36	54.70	54.79	−0.09	0.83080	0.83040	0.00040	0.81240	0.81341	−0.00101
8	755.00	755.00	−0.00	55.00	55.00	0.00	0.87830	0.87963	−0.00133	0.86540	0.86006	0.00534
9	755.00	754.65	0.35	55.20	55.28	−0.08	0.92640	0.93046	−0.00406	0.92980	0.91399	0.01581
Standard deviations			0.58			0.16			0.00300			0.01240

Dev = Meas − Calc

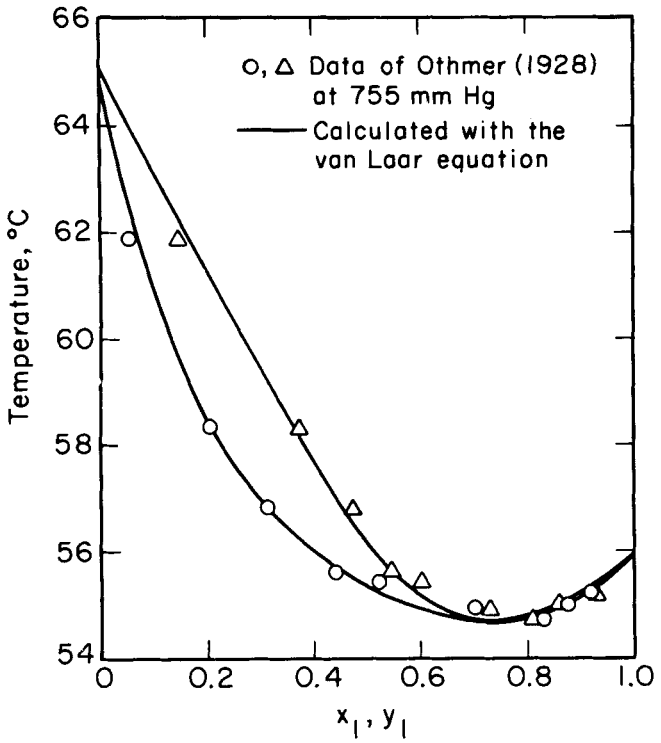

Fig. 1. Calculated phase diagram for acetone(1)/methanol(2) system.

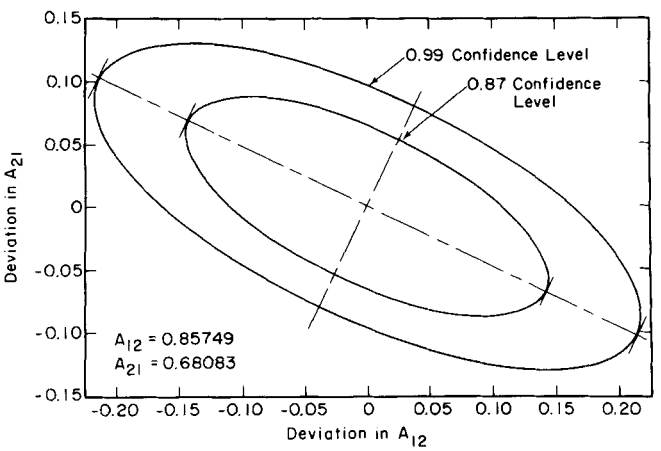

Fig. 2. Confidence ellipses for van Laar parameters. Acetone(1)/methanol(2) system at 755 mm Hg (Othmer, 1928).

are calculated as a function of composition by

$$\ln \gamma_1^0 = \frac{A_{12}}{\left[1 + \frac{A_{12}x_1}{A_{21}x_2}\right]^2} \qquad (31a)$$

$$\ln \gamma_2^0 = \frac{A_{21}}{\left[1 + \frac{A_{21}x_2}{A_{12}x_1}\right]^2} \qquad (31b)$$

Equations (31*a*) and (31*b*) are used with the constraints, Equations (29) and (30), to estimate the two adjustable parameters A_{12} and A_{21}. The assumed standard deviations in the measured variables were: $\sigma_P = 2.0$ mm Hg, $\sigma_T = 0.2°C$, $\sigma_x = 0.005$, and $\sigma_y = 0.01$. Initial parameter estimates were taken as $A_{12} = 1.0$ and $A_{21} = 1.0$. Convergence was achieved in three iterations, giving the results shown in Table 1. Table 2 gives the measured data, the estimates of the true values corresponding to the measurements, deviations of the measured values from model predictions, and estimates of the standard deviations of the measured variables. The ratios of the standard deviations are consistent with those initially assumed, except where pressure measurements are involved (the variance in pressure residuals is appreciably less than the assumed value). If the assumed standard deviations had been based on analysis of replicated experiments, then this inconsistency could indicate either systematic error in the data or lack of fit of the model to the data. In this case, however, they were a priori estimates, and the results of the parameter estimation procedure serve merely to provide better estimates of the standard deviations. The phase diagram corresponding to these parameters, together with the measured data, is shown in Figure 1.

When two parameters are determined, it is useful to plot the confidence ellipses of the estimated parameters. These are obtained from the eigenvalues and eigenvectors of the variance-covariance matrix of the parameters (Bryson and Ho, 1969). For this example, the eigenvalues are 0.00630 and 0.00085, and the major axis of the ellipse is given by the eigenvector $A_{21} = -0.481\ A_{12}$. Confidence ellipses for this example are shown in Figure 2. The regions shown represent areas within which the parameter values can be expected to lie at the confidence level associated with the ellipse (here, 0.87 and 0.99). The fact that large confidence regions are obtained for the parameters is a direct consequence of the random error in the data. For a correct model, the regions become vanishingly small as the random error becomes very small or as the number of experimental measurements becomes very large.

In the presence of significant random error, as in this example, the maximum likelihood approach can lead to considerably better parameter estimates than can less sophisticated methods. When Barker's method was used to estimate the van Laar parameters for the acetone/methanol system from this data, the values obtained were $A_{12} = 0.95991$ and $A_{21} = 0.63334$ (compared with $A_{12} = 0.85749$ and $A_{21} = 0.68083$ using the method proposed here). It should be recalled that Barker's method uses only the *P-T-x* data and assumes the T and x measurements are error free. The relative merit of the parameters obtained by the two methods can be evaluated on the basis of the average absolute difference in the calculated fugacities (over all measurements and both components) between the liquid and vapor phases. Using the parameters determined from Barker's method, this deviation in the fugacities is 12.2 mm Hg; for the parameters obtained with the maximum likelihood method described here, the average deviation is 10.0 mm Hg. This represents a significant improvement in the representation of the data, obtained as the result of use of all measured variables with appropriate consideration of their estimated errors.

ACKNOWLEDGMENT

The authors are grateful to the donors of the Petroleum Research Fund administered by the American Chemical Society for financial support, to the Computer Center, University of California (Berkeley), for the use of its facilities, and to J. M. Prausnitz for his stimulating discussions, which called our attention to the problem considered here.

NOTATION

A_{12}, A_{21} = van Laar parameters
$\mathbf{D} = \boldsymbol{\lambda} + \mathbf{f_X}^T \boldsymbol{\gamma} \mathbf{f_X} + \mathbf{g_X}^T \boldsymbol{\delta} \mathbf{f_X}$
f, g = vectors of constraint equations
$\mathbf{f_X}, \mathbf{f_\theta}, \mathbf{g_X}, \mathbf{g_\theta}$ = Jacobian partial derivative matrices of $\mathbf{f}$ and $\mathbf{g}$ with respect to $\mathbf{X}$ and $\boldsymbol{\theta}$

f_i^0 = reference fugacity for species i
g^E = excess Gibbs energy
$\mathbf{I}$ = identity matrix
K = number of independent variables
L = number of parameters to be estimated
$\mathcal{L}$ = likelihood function
M = total number of variables
N = number of experiments
n_i = number of moles of component i
P = pressure
$\mathbf{Q} = \boldsymbol{\lambda}\mathbf{X}^m + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{Y}^m + \mathbf{g}_X^T\boldsymbol{\delta}\mathbf{Z}^m$
$\mathbf{R} = \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f} + \mathbf{g}_X^T\boldsymbol{\delta}\mathbf{g}$
S = criterion function for minimization
s = estimated standard deviation
T = temperature
$\mathbf{T} = \mathbf{f}_\theta^T\boldsymbol{\gamma}\mathbf{f} + \mathbf{g}_\theta^T\boldsymbol{\delta}\mathbf{f}$
$\mathbf{U} = \mathbf{f}_\theta^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m + \mathbf{g}_\theta^T\boldsymbol{\delta}\boldsymbol{\Delta}\mathbf{Z}^m$
v_i = pure component molar volume of species i
$\bar{v}_i$ = partial molar volume of species i
w_X, w_Y, w_Z = statistical weights proportional to random error in variables X, Y, Z
$\boldsymbol{\Delta}\mathbf{X}$, $\boldsymbol{\Delta}\mathbf{Y}$, $\boldsymbol{\Delta}\mathbf{Z}$, $\boldsymbol{\Delta}\boldsymbol{\theta}$ = changes of deviation in X, Y, Z, $\boldsymbol{\theta}$
x_i = liquid mole fraction of species i
$\mathbf{x}$ = vector of variables for experiment i
$\mathbf{X}$ = vector of all variables or independent variables
y_i = vapor mole fraction of species i
$\mathbf{Y}$, $\mathbf{Z}$ = vectors of dependent variables

Greek Letters

$\boldsymbol{\beta}_i$ = variance-covariance matrix of the measured variables for experiment i
γ_i = activity coefficient of species i corrected to system pressure
γ_i^0 = activity coefficient of species at reference pressure
$\boldsymbol{\theta}$ = vector of model parameters
$\boldsymbol{\lambda}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ = diagonal matrices whose elements are reciprocals of the variances in variables X, Y, and Z, respectively
σ = standard deviation
$\tau = \mathbf{f}_X\boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma}$
$\boldsymbol{\Sigma}$ = variance-covariance matrix of parameters
Φ_i = joint probability density function for experiment i
ϕ_i = fugacity coefficient of species i

Superscripts

m = experimentally measured value
r = iteration number

LITERATURE CITED

Abrams, D. S., and J. M. Prausnitz, "Statistical Thermodynamics of Liquid Mixtures: A New Expression for the Excess Gibbs Energy of Partly or Completely Miscible Systems," *AIChE J.*, **21**, 116 (1975).

Abrams, D. S., "Thermodynamic Properties of Strongly Nonideal Liquid Mixtures," Ph.D. thesis, Univ. Calif., Berkeley (1974).

Anderson, T. F., D. S. Abrams, E. A. Grens, and J. M. Prausnitz, "Maximum Likelihood Estimation of Parameters in Liquid Mixture Models from Binary Vapor-Liquid and Liquid-Liquid Equilibrium Data," paper presented at 69th Annual A.I.Ch.E. Meeting, Chicago, Ill. (1976).

Bard, Y., and L. Lapidus, "Kinetics Analysis by Digital Parameter Estimation," *Catalysis Reviews*, **2**, No. 1, 67 (1968).

Barker, J. A., "Determination of Activity Coefficients from Total Pressure Measurements," *Austral. J. Chem.*, **6**, 207 (1953).

Britt, H. I., and R. H. Luecke, "The Estimation of Parameters in Nonlinear, Implicit Models," *Technometrics*, **5**, No. 2, 233 (1973).

Brownlee, K. A., *Statistical Theory and Methodology in Science and Engineering*, 2 ed., Wiley, New York (1965).

Bryson, A. E., and Y. B. Ho, *Applied Optimal Control; Optimization, Estimation, and Control*, Blaisdell Publishing Co., Waltham, Mass. (1969).

Clifford, A. A., *Multivariate Error Analysis*, Halsted Press, Division of John Wiley & Sons, Inc., New York (1973).

Cramér, H., *Mathematical Methods of Statistics*, 9 printing, Princeton Univ. Press, N. J. (1961).

Crow, E. L., F. A. Davis, and M. W. Maxfield, *Statistics Manual*, Dover Publications, New York (1960).

Daniel, C., and F. S. Wood, *Fitting Equations to Data*, Wiley Interscience, New York (1971).

Demming, W. E., *Statistical Adjustment of Data*, Wiley, New York (1943).

Draper, N. R., and H. Smith, *Applied Regression Analysis*, Wiley, New York (1966).

Fabries, J., and H. Renon, "Method of Evaluation and Reduction of Vapor-Liquid Equilibrium Data of Binary Mixtures," *AIChE J.*, **21**, 735 (1975).

Fisher, R. A., "On the Mathematical Foundations of Theoretical Statistics," *Phil. Trans. Royal Soc. London*, **A222**, 309 (1922).

Kreyszig, E., *Introductory Mathematical Statistics*, Wiley, New York (1970).

Law, V. J., and R. V. Bailey, "A Method for the Determination of Approximate System Transfer Functions," *Chem. Eng. Sci.*, **18**, 189 (1963).

Morrison, D. F., *Multivariate Statistical Methods*, McGraw-Hill, New York (1967).

O'Neil, M., I. G. Sinclair, and F. J. Smith, "Polynomial Curve Fitting When Abscissas and Ordinates are Both Subject to Error," *Computer J.*, **12**, 52 (1969).

Othmer, D. F., "Compositions of Vapors from Boiling Binary Solutions," *Ind. Eng. Chem.*, **20**, 743 (1928).

Péneloux, A., R. Deyrieux, E. Canals, and E. Neau, "The Maximum Likelihood Test and the Estimation of Experimental Inaccuracies: Application to Data Reduction for Vapor-Liquid Equilibrium," *J. de Chim. Phys.*, **73**, 706 (1976).

Powell, P. R., and J. R. MacDonald, "A Rapidly Convergent Interative Method for the Solution of the Generalized Nonlinear Least Squares Problem," *Computer J.*, **15**, 148 (1972).

Prausnitz, J. M., C. A. Eckert, R. V. Orye, and J. P. O'Connell, *Computer Calculations for Multicomponent Vapor-Liquid Equilibria*, Prentice-Hall, Englewood Cliffs, N.J. (1967).

Prausnitz, J. M., *Molecular Thermodynamics of Fluid-Phase Equilibria*, Prentice-Hall, Englewood Cliffs, N.J. (1969).

Renon, H., Ecole Nationale Supérieure des Mines de Paris, Paris, France, private communication (1977).

Southwell, W. H., "Fitting Experimental Data," *J. Computational Physics*, **4**, 465 (1969).

———, "Fitting Data to Nonlinear Functions with Uncertainties in All Measurement Variables," *Computer J.*, **19**, 69 (1976).

Widder, D. V., *Advanced Calculus*, 2 ed., Prentice-Hall Englewood Cliffs, N.J. (1961).

Williamson, J. H., "Least Squares Fitting of a Straight Line," *Can. J. Physics*, **46**, 1845 (1968).

York, P., "Least Squares Fitting of a Straight Line," *ibid.*, **44**, 1709 (1966).

APPENDIX A. DERIVATION OF EQUATIONS (15) AND (16)

The intermediate steps in the derivation of Equations (14) and (15) are given below. Starting with Equation (12), we add and subtract $\boldsymbol{\gamma}\mathbf{X}^r$, giving

$$\boldsymbol{\lambda}\boldsymbol{\Delta}\mathbf{X} + \boldsymbol{\lambda}\boldsymbol{\Delta}\mathbf{X}^m + \mathbf{f}_X^T\boldsymbol{\gamma}(\mathbf{Y}^m + \mathbf{f}_X\boldsymbol{\Delta}\mathbf{X} + \mathbf{f}_\theta\boldsymbol{\Delta}\boldsymbol{\theta}) + \mathbf{g}_X^T\boldsymbol{\delta}(\boldsymbol{\Delta}\mathbf{Z}^m + \mathbf{g}_X\boldsymbol{\Delta}\mathbf{X} + \mathbf{g}_\theta\boldsymbol{\Delta}\boldsymbol{\theta}) = 0 \quad \text{(A1)}$$

where

$$\boldsymbol{\Delta}\mathbf{X} = \mathbf{X}^{r+1} - \mathbf{X}^r \quad \text{(A2)}$$

$$\boldsymbol{\Delta}\boldsymbol{\theta} = \boldsymbol{\theta}^{r+1} - \boldsymbol{\theta}^r \quad \text{(A3)}$$

$$\boldsymbol{\Delta}\mathbf{X}^m = \mathbf{X}^r - \mathbf{X}^m \quad \text{(A4)}$$

$$\boldsymbol{\Delta}\mathbf{Y}^m = \mathbf{f}^r - \mathbf{Y}^m \tag{A5}$$

$$\boldsymbol{\Delta}\mathbf{Z}^m = \mathbf{g}^r - \mathbf{Z}^m \tag{A6}$$

We then expand and combine terms containing $\boldsymbol{\Delta}\mathbf{X}$, giving

$$(\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X + \mathbf{g}_X^T\boldsymbol{\delta}\mathbf{g}_X)\boldsymbol{\Delta}\mathbf{X} = - [\boldsymbol{\lambda}\boldsymbol{\Delta}\mathbf{X}^m + \mathbf{f}_X^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m + \mathbf{g}_X^T\boldsymbol{\delta}\boldsymbol{\Delta}\mathbf{Z}^m + (\mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_\theta + \mathbf{g}_X^T\boldsymbol{\delta}\mathbf{g}_\theta)\boldsymbol{\Delta}\boldsymbol{\theta}] \tag{A7}$$

When we let

$$\mathbf{D} = \boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X + \mathbf{g}_X^T\boldsymbol{\delta}\mathbf{g}_X \tag{A8}$$

$$\mathbf{Q} = \boldsymbol{\lambda}\boldsymbol{\Delta}\mathbf{X}^m + \mathbf{f}_X^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m + \mathbf{g}_X^T\boldsymbol{\delta}\boldsymbol{\Delta}\mathbf{Z}^m \tag{A9}$$

$$\mathbf{R} = \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_\theta + \mathbf{g}_X^T\boldsymbol{\delta}\mathbf{g}_\theta \tag{A10}$$

Equation (A7) can be solved for $\boldsymbol{\Delta}\mathbf{X} = \mathbf{X}^{r+1} - \mathbf{X}^r$:

$$\boldsymbol{\Delta}\mathbf{X} = - \mathbf{D}^{-1}[\mathbf{Q} + \mathbf{R}\boldsymbol{\Delta}\boldsymbol{\theta}] \tag{A11}$$

This is identical to Equation (15) and gives $\boldsymbol{\Delta}\mathbf{X}$ as a function of $\boldsymbol{\Delta}\boldsymbol{\theta}$ and other known quantities. If we substitute this expression for $\boldsymbol{\Delta}\mathbf{X}$ into Equation (13), we get

$$\mathbf{f}_\theta^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m - \mathbf{f}_\theta^T\boldsymbol{\gamma}\mathbf{f}_X\mathbf{D}^{-1}(\mathbf{Q} + \mathbf{R}\boldsymbol{\Delta}\boldsymbol{\theta}) + \mathbf{f}_\theta^T\boldsymbol{\gamma}\mathbf{f}_\theta\boldsymbol{\Delta}\boldsymbol{\theta} + \mathbf{g}_\theta^T\boldsymbol{\delta}\boldsymbol{\Delta}\mathbf{Z}^m - \mathbf{g}_\theta^T\boldsymbol{\delta}\mathbf{g}_X\mathbf{D}^{-1}(\mathbf{Q} + \mathbf{R}\boldsymbol{\Delta}\boldsymbol{\theta}) + \mathbf{g}_\theta^T\boldsymbol{\gamma}\mathbf{g}_\theta\boldsymbol{\Delta}\boldsymbol{\theta} = 0 \tag{A12}$$

Expanding and combining terms containing $\boldsymbol{\Delta}\boldsymbol{\theta}$, we get

$$[\mathbf{f}_\theta^T\boldsymbol{\gamma}\mathbf{f}_\theta + \mathbf{g}_\theta^T\boldsymbol{\delta}\mathbf{g}_\theta - (\mathbf{f}_\theta^T\boldsymbol{\gamma}\mathbf{f}_X + \mathbf{g}_\theta^T\boldsymbol{\delta}\mathbf{g}_X)\mathbf{D}^{-1}\mathbf{R}]\boldsymbol{\Delta}\boldsymbol{\theta} = - [\mathbf{f}_\theta^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m + \mathbf{g}_\theta^T\boldsymbol{\delta}\boldsymbol{\Delta}\mathbf{Z}^m - (\mathbf{f}_\theta^T\boldsymbol{\gamma}\mathbf{f}_X + \mathbf{g}_\theta^T\boldsymbol{\delta}\mathbf{g}_X)\mathbf{D}^{-1}\mathbf{Q}] \tag{A13}$$

We now define

$$\mathbf{U} = \mathbf{f}_\theta^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m + \mathbf{g}_\theta^T\boldsymbol{\delta}\boldsymbol{\Delta}\mathbf{Z}^m \tag{A14}$$

$$\mathbf{T} = \mathbf{f}_\theta^T\boldsymbol{\gamma}\mathbf{f}_\theta + \mathbf{g}_\theta^T\boldsymbol{\delta}\mathbf{g}_\theta \tag{A15}$$

and note that

$$(\mathbf{f}_\theta^T\boldsymbol{\gamma}\mathbf{f}_X + \mathbf{g}_\theta^T\boldsymbol{\delta}\mathbf{g}_X) \equiv (\mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_\theta + \mathbf{g}_X^T\boldsymbol{\delta}\mathbf{g}_\theta) \equiv \mathbf{R}^T \tag{A16}$$

so that Equation (A13) can be written in the following form:

$$(\mathbf{T} - \mathbf{R}^T\mathbf{D}^{-1}\mathbf{R})\,\boldsymbol{\Delta}\boldsymbol{\theta} = - (\mathbf{U} - \mathbf{R}^T\mathbf{D}^{-1}\mathbf{Q}) \tag{A17}$$

The final form, giving $\boldsymbol{\Delta}\boldsymbol{\theta}$ as a function of known quantities, is

$$\boldsymbol{\Delta}\boldsymbol{\theta} = - (\mathbf{T} - \mathbf{R}^T\mathbf{D}^{-1}\mathbf{R})^{-1}(\mathbf{U} - \mathbf{R}^T\mathbf{D}^{-1}\mathbf{Q}) \tag{A18}$$

This equation is Equation (14).

APPENDIX B. CASE OF ONE CONSTRAINT

The one constraint case is an important simplification of the algorithm discussed in the text. It is, as might be expected, computationally faster than the two constraint case. The savings in time is derived from not having to invert a $NK \times NK$ matrix.

If we have only one constraint $\mathbf{Y} = f(\mathbf{X}, \boldsymbol{\theta})$, then the terms associated with the constraint $\mathbf{Z} = g(\mathbf{X}, \boldsymbol{\theta})$ are zero. Equation (14) can then be written as

$$\boldsymbol{\Delta}\boldsymbol{\theta} = [\mathbf{f}_\theta^T\boldsymbol{\gamma}\mathbf{f}_\theta - \mathbf{f}_\theta^T\boldsymbol{\gamma}\underline{\mathbf{f}_X\,(\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}}\,\mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_\theta]^{-1} [\mathbf{f}_\theta^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m - \mathbf{f}_\theta^T\boldsymbol{\gamma}\underline{\mathbf{f}_X\,(\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}} (\boldsymbol{\lambda}\boldsymbol{\Delta}\mathbf{X}^m + \mathbf{f}_X^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m)] \tag{B1}$$

The first step in simplifying this equation is to note that the underlined quantities can be simplified in the following steps:

$$\begin{aligned}
&\mathbf{f}_X(\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1} \\
&\quad \equiv \mathbf{f}_X[\boldsymbol{\lambda}(\mathbf{I} + \boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)]^{-1} \\
&\quad \equiv \mathbf{f}_X(\mathbf{I} + \boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}\boldsymbol{\lambda}^{-1} \\
&\quad \equiv (\mathbf{I} + \mathbf{f}_X\boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma})^{-1}(\mathbf{I} + \mathbf{f}_X\boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma})\,\mathbf{f}_X(\mathbf{I} + \boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}\boldsymbol{\lambda}^{-1} \\
&\quad \equiv (\mathbf{I} + \mathbf{f}_X\boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma})^{-1}\mathbf{f}_X\boldsymbol{\lambda}^{-1}
\end{aligned} \tag{B2}$$

The reason for making this transformation is that $(\mathbf{I} + \mathbf{f}_X\boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma})$ is an $N \times N$ diagonal matrix whose inverse is simply the reciprocal of the diagonal elements. $(\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)$, however, is an $NK \times NK$ block-diagonal matrix which is more difficult to invert.

We let

$$\boldsymbol{\tau} = \mathbf{f}_X\boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma} \tag{B3}$$

make the desired substitution of expression (B2) into (B1), and get

$$\boldsymbol{\Delta}\boldsymbol{\theta} = [\mathbf{f}_\theta^T\boldsymbol{\gamma}\mathbf{f}_\theta - \mathbf{f}_\theta^T\boldsymbol{\gamma}(\mathbf{I} + \boldsymbol{\tau})^{-1}\boldsymbol{\tau}\mathbf{f}_\theta]^{-1} [\mathbf{f}_\theta^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m - \mathbf{f}_\theta^T\boldsymbol{\gamma}(\mathbf{I} + \boldsymbol{\tau})^{-1}\mathbf{f}_X\boldsymbol{\lambda}^{-1}(\boldsymbol{\lambda}\boldsymbol{\Delta}\mathbf{X}^m + \mathbf{f}_X^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m)] \tag{B4}$$

This further simplifies to

$$\boldsymbol{\Delta}\boldsymbol{\theta} = [\mathbf{f}_\theta^T\boldsymbol{\gamma}\{\mathbf{I} - (\mathbf{I} + \boldsymbol{\tau})^{-1}\boldsymbol{\tau}\}\mathbf{f}_\theta]^{-1} [\mathbf{f}_\theta^T\boldsymbol{\gamma}\{\boldsymbol{\Delta}\mathbf{Y}^m - (\mathbf{I} + \boldsymbol{\tau})^{-1}\mathbf{f}_X\boldsymbol{\Delta}\mathbf{X}^m - (\mathbf{I} + \boldsymbol{\tau})^{-1}\boldsymbol{\tau}\boldsymbol{\Delta}\mathbf{Y}^m\}] \tag{B5}$$

Finally, noting that $(\mathbf{I} - (\mathbf{I} + \boldsymbol{\tau})^{-1}\boldsymbol{\tau}) = (\mathbf{I} + \boldsymbol{\tau})^{-1}$, we get

$$\boldsymbol{\Delta}\boldsymbol{\theta} = [\mathbf{f}_\theta^T\boldsymbol{\gamma}(\mathbf{I} + \boldsymbol{\tau})^{-1}\mathbf{f}_\theta]^{-1}\,\mathbf{f}_\theta^T\boldsymbol{\gamma}(\mathbf{I} + \boldsymbol{\tau})^{-1}[\boldsymbol{\Delta}\mathbf{Y}^m - \mathbf{f}_X\boldsymbol{\Delta}\mathbf{X}^m] \tag{B6}$$

Similar simplifications can be made in solving for the new true values of the variables $\mathbf{X}^{r+1}$. For one constraint, Equation (15) becomes

$$\mathbf{X}^{r+1} - \mathbf{X}^r = - (\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}(\boldsymbol{\lambda}\boldsymbol{\Delta}\mathbf{X}^m + \mathbf{f}_X^T\boldsymbol{\gamma}\boldsymbol{\Delta}\mathbf{Y}^m + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_\theta\boldsymbol{\Delta}\boldsymbol{\theta}) \tag{B7}$$

Here, it is more convenient to solve for $\mathbf{X}^{r+1} - \mathbf{X}^m$. We add the quantity $\boldsymbol{\Delta}\mathbf{X}^m = \mathbf{X}^r - \mathbf{X}^m$ to both sides of Equation (B7) giving

$$\mathbf{X}^{r+1} - \mathbf{X}^m = - (\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}[\boldsymbol{\lambda}\boldsymbol{\Delta}\mathbf{X}^m + \mathbf{f}_X^T\boldsymbol{\gamma}(\boldsymbol{\Delta}\mathbf{Y}^m + \mathbf{f}_\theta\boldsymbol{\Delta}\boldsymbol{\theta}) + \boldsymbol{\Delta}\mathbf{X}^m] \tag{B8}$$

When $\boldsymbol{\Delta}\mathbf{X}^m$ is brought inside the parentheses, we get

$$\mathbf{X}^{r+1} - \mathbf{X}^m = - (\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}[\boldsymbol{\lambda}\boldsymbol{\Delta}\mathbf{X}^m - (\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)\boldsymbol{\Delta}\mathbf{X}^m + \mathbf{f}_X^T\boldsymbol{\gamma}(\boldsymbol{\Delta}\mathbf{Y}^m + \mathbf{f}_\theta\boldsymbol{\Delta}\boldsymbol{\theta})] \tag{B9}$$

which simplifies to

$$\mathbf{X}^{r+1} - \mathbf{X}^m = - \underline{(\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}\,\mathbf{f}_X^T\boldsymbol{\gamma}}(\boldsymbol{\Delta}\mathbf{Y}^m - \mathbf{f}_X\boldsymbol{\Delta}\mathbf{X} + \mathbf{f}_\theta\boldsymbol{\Delta}\boldsymbol{\theta}) \tag{B10}$$

We further note that the underlined quantity can be transformed as follows:

$$\begin{aligned}
&(\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}\mathbf{f}_X^T \\
&\quad = (\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}\mathbf{f}_X^T\boldsymbol{\gamma}(\mathbf{I} + \mathbf{f}_X\boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma})(\mathbf{I} + \mathbf{f}_X\boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma})^{-1} \\
&\quad = (\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)^{-1}(\boldsymbol{\lambda} + \mathbf{f}_X^T\boldsymbol{\gamma}\mathbf{f}_X)\boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma}(\mathbf{I} + \mathbf{f}_X\boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma})^{-1} \\
&\quad = \boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma}(\mathbf{I} + \boldsymbol{\tau})^{-1}
\end{aligned} \tag{B11}$$

This transformation allows Equation (B10) in the final form:

$$\mathbf{X}^{r+1} - \mathbf{X}^m = - \boldsymbol{\lambda}^{-1}\mathbf{f}_X^T\boldsymbol{\gamma}(\mathbf{I} + \boldsymbol{\tau})^{-1}(\boldsymbol{\Delta}\mathbf{Y}^m - \mathbf{f}_X\boldsymbol{\Delta}\mathbf{X} + \mathbf{f}_\theta\boldsymbol{\Delta}\boldsymbol{\theta}) \tag{B12}$$

Equations (B6) and (B12) can be used to determine the maximum likelihood estimates of the parameters for the one constraint case in the same way that Equations (14) and (15) are used for the two constraint case.

Manuscript received February 15, 1977; revision received August 24, and accepted September 14, 1977.