Generalized Least-Squares Parameter Estimation from Multiequation Implicit Models

Simon L. Marshall
Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

Maximum likelihood fit of nonlinear, implicit, multiple-response models to data containing normally distributed random errors can be carried out by a combination of the Gauss-Newton generalized nonlinear least-square algorithm first described by Britt and Luecke in 1973, with a Fletcher-Reeves conjugate gradient search for initial parameter estimates. The convergence of the algorithm is further improved by adding a step-limiting procedure that ensures a reduction in the objective function for each iteration. Multiple-equation regression methods appropriate to the solution of explicit fixed-regressor models are derived from this general treatment as special cases. These include weighted nonlinear least squares (where the covariance matrix of the response is known), and uniformly weighted nonlinear least squares (where the responses are uncorrelated and characterized by a single common variance). Alternative methods for fixed-regressor fits of explicit multiequation models with an unknown covariance matrix of the responses are also considered. The moment-matrix determinant criterion appropriate in such situations is also efficiently minimized by use of the conjugate-gradient algorithm, which is considerably less sensitive to the accuracy of the initial parameter estimate than the more usual Gauss-Newton methods. The performance of the new algorithm for models defined by one, two, and three implicit functional constraints per point is illustrated by random-regressor fits of isothermal $p-X$ and $p-X-Y$ vapor−liquid equilibrium data, and ternary liquid−liquid equilibrium data, respectively.

Introduction

Many data-reduction problems arising in chemical engineering involve fit of an explicit relation of the form $y = f(\boldsymbol{x},z)$ to experimental data $(y_i,\boldsymbol{x}_i)$, $i = 1, \ldots, N$, where y_i are measured values of a single dependent variable, and $\boldsymbol{x}$ and z are vectors of independent (or regressor) variables and model-dependent parameters, respectively. If the $\boldsymbol{x}_i$ are assumed to be error-free and the errors in the y_i are normally distributed with zero mean and standard deviations σ_i, the method of weighted nonlinear least squares (WNLS) can be used to adjust z so that the sum of weighted squared residuals $[y_i - f(\boldsymbol{x}_i,z)]^2/\sigma_i^2$ is minimized. This minimization results in an overdetermined system of nonlinear equations involving the partial derivatives $\partial f/\partial z$ as well as the weights σ_i^{-2}, which requires an iterative numerical solution unless $\partial f/\partial z$ is independent of z (in which case f is said to *linear*). If all the σ_i are equal, the common variance cancels out in the equations for the parameter estimates, and the limiting case of unweighted nonlinear least squares (UNLS)—or, perhaps more correctly, uniformly weighted nonlinear least squares—is obtained. WNLS can itself be regarded as a limiting case of generalized nonlinear least squares (GNLS) or errors-in-variables problems, in which all the data are assumed to be of finite precision, and the distinction between dependent and independent variables disappears. GNLS calculations involve concurrent estimation of the exact values of the data and the best-fit parameter values, and are therefore always nonlinear, even if the postulated functional relationship is linear with respect to its parameters.

Much attention has been devoted to the fit of straight lines to data with errors in both coordinates (Pearson, 1901; Deming, 1943; York, 1966); Macdonald and Thompson (1992) reviewed much of this work, and further discussions and references are to be found in Press et al. (1992, p. 664). Powell and Macdonald (1972) devised the first successful GNLS algorithm for fitting single-response data to explicit functions of general form, and modified Gauss-Newton methods applicable to the fit of either implicit or explicit single-equation models were described by Britt and Luecke (1973), Jefferys (1980), and Lybanon (1984a, 1985). The latter approaches involve minimization of a weighted sum of squares, subject to a vector of functional constraints known as "equations of conditions". Subsequent refinements include application of the Implicit Function Theorem to the calculation of partial derivatives (Sachs, 1976), inclusion of a second vector of functional constraints (Anderson et al., 1978), and application of the Marquardt minimization algorithm (Schwetlick and Tiller, 1985; Valkó and Vajda, 1987) and nonlinear programming techniques (Kim et al., 1990). The literature on chemical engineering applications of multiequation parameter estimation was reviewed by Stewart et al. (1992); the mathematical foundations of nonlinear optimization have been reviewed very thoroughly by Forsgren et al. (2002). Recently, Lisý and Šimon (1998) and Lisý et al. (1999) proposed alternative approaches that do not rely on linearization about a trial solution, and Esposito and Floudas (1998, 2000) have considered parameter estimation in algebraic or algebraic-differential equation models for which multiple-parameter sets are possible.

Single-equation GNLS is the best available method for the correlation of most physical property data. One notable exception is permittivity, where real and imaginary parts must be fitted concurrently to theoretical equations for the frequency response. This can be carried out with dual-response GNLS, but is best achieved with the specialized complex nonlinear least-square techniques developed and extensively applied by Macdonald and collaborators (Macdonald, 1992, 1995; Macdonald and Garber, 1977; Macdonald et al., 1982). The need for multiple-response nonlinear regression in chemical engineering was first recognized by Hunter (1967) with reference to parameter determination in kinetic models. Although most subsequent work has focused on reduction of vapor−liquid equilibrium (VLE) measurements (Anderson et al., 1978; Kemény et al., 1982; Rubio et al., 1983; Skjold-Jørgensen, 1983; Englezos et al., 1991) and on the related problem of assessing the thermodynamic consistency of such data (Dohnal and Fenclová, 1985), these by no means exhaust the range of problems to which such methods can be fruitfully applied. Examples of other applications include analysis of titration curves (Sachs, 1976; Hofman and Krzyżanowska, 1986), and identification of parameters in differential equation reactor models (Kim et al., 1991).

The Gauss-Newton GNLS algorithms described by Britt and Luecke (1973), Anderson et al. (1978), Jefferys (1980), and Lybanon (1984a, 1985) all involve the linearization of the functional constraints about an assumed trial solution followed by solution of the linearized equations for the parameter estimates at each iteration. A consequence of this is that the iteration tends to require rather good initial estimates for the parameters. If the initial parameter estimates are too far from the least-square values for this linearization to be accurate, the parameter increment vector determined by solution of the linearized problem can "overshoot" the least-square point, leading to an increase (rather than a decrease) of the objective function, and divergence of the iteration. This pattern of behavior has given rise to a prevailing view in the literature that Gauss-Newton methods are of limited usefulness, particularly for fit of the highly nonlinear models that are commonly encountered in chemical engineering. On the other hand, as Lybanon (1985) first observed, programs for Gauss-Newton GNLS can be obtained by fairly straightforward adaptation of those designed for conventional nonlinear least-square fitting. This major practical advantage suggests that a reexamination of the applicability of the Gauss-Newton method to the fit of multiequation, implicit models would be worthwhile.

The main purpose of this article is to show that the GNLS algorithm described by Britt and Luecke (1973) for single-equation models can be extended to multiequation models, in the manner suggested by Seber and Wild (1989, p. 502), and that the usefulness of this algorithm can be expanded on the one hand by an improved method for generating initial parameter estimates, and on the other hand by insertion of a step-limiting procedure. The relation of this approach to other multiequation parameter estimation methods is discussed, and the treatment of models defined by one, two, and three equations for each data point is illustrated by reduction of binary VLE and ternary liquid−liquid equilibrium (LLE) data.

Multiresponse Generalized Least Squares

The ML method was first developed by Fisher (1922) and is discussed in many specialist monographs on statistical theory [for example, Cramér (1961)]. The first step in its application to parameter estimation with variables that contain random errors is the assignment of a *likelihood* or joint probability distribution function to all the data. The maximization of this function (or more commonly, its logarithm) is equivalent to the minimization of the weighted sum of squared residuals subject to constraints that express implicit or explicit relationships postulated between the exact values of the variables (Bard, 1974, pp. 67−68). Formulation of the parameter-estimation problem as a constrained optimization is to be contrasted with conventional least-squares approaches, which involve unconstrained minimization of the sum of squared residuals. Since such a treatment was first described in detail by Britt and Luecke (1973) and independently by Jefferys (1980), for the case of a single equation of condition for each datum, it seems appropriate to refer to it as the Britt-Luecke-Jefferys (BLJ) method. Lybanon (1984a) has discussed the relationship of the BLJ treatment to the earlier "effective variance" approach, and has described a realization in the Basic language (Lybanon, 1985). The main purposes of the following derivation are (1) to show how the BLJ approach can be applied to models with multiple equations of condition; (2) to elucidate certain important similarities and differences between GNLS and WNLS; and (3) to define a hierarchy of approximations into which all published nonlinear regression techniques can be placed.

For a data set consisting of N sets of Q variables, the measured values $\boldsymbol{x}_0$, exact values $\boldsymbol{x}$, and residuals $\boldsymbol{v} \equiv \boldsymbol{x} - \boldsymbol{x}_0$ can be stored in NQ-dimensional row vectors. Assuming that the variables follow the multivariate normal distribution, the log-likelihood function is

$$\ln L = -\frac{NQ}{2}\ln(2\pi) - \frac{1}{2}\ln\det\boldsymbol{\sigma} - \frac{1}{2}\boldsymbol{v}\boldsymbol{\sigma}^{-1}\boldsymbol{v}^T, \quad (1)$$

(Anderson, 1958, chap. 2), where $\boldsymbol{\sigma}$ is the covariance matrix of the measurements (assumed to be known) and T denotes the transpose of a matrix. The covariance matrix is always symmetric; the further assumption that the covariance matrix is positive definite covers most cases of physical interest ($\boldsymbol{\sigma}$ is, in fact, usually assumed to be diagonal). Maximization of this expression is equivalent to minimization of the last term on the righthand side, which can be recognized as a weighted sum of squared residuals. Its value provides an overall quantitative measure of the quality of the fit. Imposing an MN-dimensional column vector of constraints $\boldsymbol{G} \equiv \boldsymbol{G}(\boldsymbol{x}_0 + \boldsymbol{v}, \boldsymbol{z}) = \boldsymbol{0}$, where $\boldsymbol{z}$ is a P-dimensional row vector of adjustable parameters, transforms this problem to the unconstrained minimization of the Lagrangian

$$\mathfrak{L} = \frac{1}{2}\mathbf{v}\boldsymbol{\sigma}^{-1}\boldsymbol{v}^T + \boldsymbol{\mu}\mathbf{G} \tag{2}$$

where $\boldsymbol{\mu}$ is a row vector of Lagrange multipliers. This in turn requires that

$$\nabla_x \mathfrak{L} = \boldsymbol{\sigma}^{-1}\boldsymbol{v}^T + \boldsymbol{\mu}\nabla_x\mathbf{G} = \mathbf{0} \qquad \nabla_z \mathfrak{L} = \boldsymbol{\mu}\nabla_z\mathbf{G} = \mathbf{0}$$
$$\nabla_\mu \mathfrak{L} = \mathbf{G} = \mathbf{0}, \tag{3}$$

where the subscripts x, z, and μ are used to identify the gradients taken with respect to the components of the vectors $\boldsymbol{x}$, $\boldsymbol{z}$, and $\boldsymbol{\mu}$, and that the Hessian matrix be positive definite when projected on the constraints. The first step in the solution of this system of equations by the Newton-Raphson method is linearization about a trial solution $\boldsymbol{x} = \boldsymbol{x}_1 + \Delta\boldsymbol{x}$ and $\boldsymbol{z} = \boldsymbol{z}_1 + \Delta\boldsymbol{z}$. Denoting the value of $\boldsymbol{G}$ at $(\boldsymbol{x}_1, \boldsymbol{z}_1)$ by $\boldsymbol{G}_1$, the third of Eq. 3 becomes

$$\boldsymbol{G} = \boldsymbol{G}_1 + \nabla_x\boldsymbol{G}(\Delta\boldsymbol{x})^T + \nabla_z\boldsymbol{G}(\Delta\boldsymbol{z})^T = \boldsymbol{0}, \tag{4}$$

where all the gradients are understood to be evaluated at $(\boldsymbol{x}_1, \boldsymbol{z}_1)$. Substitution of this result into the first of Eqs. 3, together with $\boldsymbol{v} = \boldsymbol{v}_1 + \Delta\boldsymbol{x}$, gives

$$\boldsymbol{\sigma}^{-1}\left(\boldsymbol{v}_1^T + \Delta\boldsymbol{x}^T\right) = -\boldsymbol{\mu}\nabla_x\boldsymbol{G} \tag{5}$$

$$\Rightarrow \Delta\boldsymbol{x}^T = -\boldsymbol{v}_1^T - \boldsymbol{\sigma}\boldsymbol{\mu}\nabla_x\boldsymbol{G} = -\boldsymbol{v}_1^T - \boldsymbol{\sigma}(\nabla_x\boldsymbol{G})^T\boldsymbol{\mu}^T. \tag{6}$$

Combining this with Eq. 4 gives

$$\boldsymbol{G} = \boldsymbol{G}_1 - \nabla_x\boldsymbol{G}\left[\boldsymbol{v}_1^T + \boldsymbol{\sigma}(\nabla_x\boldsymbol{G})^T\boldsymbol{\mu}^T\right] + \nabla_z\boldsymbol{G}(\Delta\boldsymbol{z})^T = \boldsymbol{0}, \tag{7}$$

$$\Rightarrow \boldsymbol{\mu}^T = \boldsymbol{W}\left[\boldsymbol{G}_1 - \nabla_x\boldsymbol{G}\boldsymbol{v}_1^T + \nabla_z\boldsymbol{G}(\Delta\boldsymbol{z})^T\right], \tag{8}$$

where $\boldsymbol{W} \equiv [(\nabla_x\boldsymbol{G})\boldsymbol{\sigma}(\nabla_x\boldsymbol{G})^T]^{-1}$ is known as the effective weighting matrix. The normal equations for $\Delta\boldsymbol{z}$ follow by substituting this expression for $\boldsymbol{\mu}$ into the second of Eqs. 3

$$(\nabla_z\boldsymbol{G})^T\boldsymbol{W}(\nabla_z\boldsymbol{G})(\Delta\boldsymbol{z})^T = -(\nabla_z\boldsymbol{G})^T\boldsymbol{W}\left[\boldsymbol{G}_1 - (\nabla_x\boldsymbol{G})\boldsymbol{v}_1^T\right]. \tag{9}$$

At this point, an important limitation of the classic Gauss-Newton algorithm becomes manifest. For highly nonlinear models, the accuracy of the linearization approximation that forms the basis for the preceding analysis might be insufficient, unless the initial approximation to the parameter vector is very close to the least-square value. One strategy that can be used to overcome the possibility that the updated parameters could result in a deterioration (rather than an improvement) of the residuals is to replace $\boldsymbol{z}_1$ by $\boldsymbol{z}_1 + \alpha\Delta\boldsymbol{z}$, where α is a scalar parameter usually less than 1. The value of α is determined so that the sum of squares corresponding to the corrected parameter values is less than that at the starting point. The Gauss-Newton method described by Anderson et al. (1978) used the step-limiting method devised by Law and Bailey (1963) in connection with the least-squares approximation of system transfer functions in Laplace space. A more general procedure for adjusting α is to define the scalar function $h(\alpha) = |\boldsymbol{G}(\boldsymbol{x}_1, \boldsymbol{z}_1 + \alpha\Delta\boldsymbol{z})|^2$, and to compare the values of $h(0)$ and $h(1)$. If $h(1) \le h(0)$, α is set equal to 1 and the full Gauss-Newton correction is applied. If $h(1) > h(0)$, a one-dimensional search algorithm [as discussed by Brent (1973, chap. 5); see also Forsythe et al. (1977, chap. 8) and Press et al. (1992, pp. 390–398)] is applied to find α corresponding to the minimum value of h. Such *line-search* procedures play a crucial role in "direction set" algorithms for unconstrained function minimization, such as Powell's method (Powell, 1964), as well as variable-metric algorithms for nonlinear programming (nonlinearly constrained optimization) (Han, 1976; Powell, 1978). With α thus determined, the variable increments can now be obtained from

$$(\Delta\boldsymbol{x})^T = \boldsymbol{x}_0^T - \boldsymbol{x}_1^T - \boldsymbol{\sigma}(\nabla_x\boldsymbol{G})^T\boldsymbol{W}$$
$$\times\left[\boldsymbol{G}_1 - (\nabla_x\boldsymbol{G})\boldsymbol{v}_1^T + (\nabla_z\boldsymbol{G})\alpha(\Delta\boldsymbol{z})^T\right], \tag{10}$$

and the entire process is repeated until $\Delta\boldsymbol{z}$, $\Delta\boldsymbol{x}$ satisfy an appropriate convergence criterion.

Application of line-search minimization procedures to $h = |\boldsymbol{G}|^2$ also proves to be of value in addressing the other widely recognized limitation of Gauss-Newton parameter-estimation algorithms: their reliance on reasonably good initial parameter estimates. For this purpose, the approximation $\boldsymbol{x} \approx \boldsymbol{x}_0$ is made. The simplest choice for the search direction is the gradient $\nabla_z h$, but as is well known, this results in rather poor convergence properties [it can be shown (Wismer and Chattergy, 1978, p. 143) that the rate of convergence becomes infinitely slow as the minimum is approached]. Greatly improved performance can be achieved by the conjugate gradient method (Fletcher and Reeves, 1964), in which a sequence of line searches along gradient vectors are used to set up a mutually conjugate basis of search directions. This consists of the following steps [compare Wismer and Chattergy (1978, pp. 228–229)]:

(1) Let $\boldsymbol{z}^{(0)}$ represent the first approximation. Calculate $\nabla h(\boldsymbol{z}^{(0)})$ and set $\boldsymbol{u}_0 = -\nabla h(\boldsymbol{z}^{(0)})$;

(2) For $i = 1, \ldots, P-1$;

(2.1) Set $\boldsymbol{z}^{(i)} = \boldsymbol{z}^{(i-1)} + \alpha_{i-1}\boldsymbol{u}_{i-1}$, where α_{i-1} is found by minimization of $h(\boldsymbol{z}^{(i-1)} - \alpha\boldsymbol{u}_{i-1})$;

(2.2) Calculate $\nabla h(\boldsymbol{z}^{(i)})$;

(2.3) When $i < n$, define

$$\boldsymbol{u}_i = -\nabla h(\boldsymbol{z}^{(i)}) + \frac{|\nabla h(\boldsymbol{z}^{(i)})|^2}{|\nabla h(\boldsymbol{z}^{(i-1)})|^2}\boldsymbol{u}_{i-1}.$$

(3) Replace $\boldsymbol{z}^{(0)}$ with $\boldsymbol{z}^{(n)}$ and go to step 1 unless a suitable convergence criterion is satisfied.

Further consideration of the theory of this method (or of quadratically convergent minimization algorithms in general) would be out of place here. Details can be found in Wismer and Chattergy (1978, chap. 9) and Press et al. (1992, chap. 10). It produces good initial parameter estimates if all the required partial derivatives are approximated by central difference quotients. The Powell (1964) method (which does not involve numerical differentiation) also works well. But each procedure can be expected to converge, at best, to a *local* minimum; discrimination between multiple parameter sets that are possible in nonlinear models is best achieved on a case-by-case basis, or by application of the global-optimization methods proposed recently by Esposito and Floudas (1998, 2000).

Lybanon (1984a) wrote out explicit equations for the effective weighting matrix and variable increments for the particular case in which the function to be fitted to the data was an explicit functional relation between one independent variable and one dependent variable. The formulas given by Lybanon must be modified if there are several implicit equations of condition for each datum. Thus, if $\boldsymbol{G} \equiv [\boldsymbol{g}^{(1)}\ \boldsymbol{g}^{(2)}\ \cdots\ \boldsymbol{g}^{(N)}]^T$ the Jacobian with respect to the variables $\boldsymbol{x}$ consists of N blocks of dimension $M \times Q$

$$\nabla_x \boldsymbol{G} = \left[\begin{array}{cccccccccc}
\frac{\partial g_1^{(1)}}{\partial x_{11}} & \cdots & \frac{\partial g_1^{(1)}}{\partial x_{1Q}} & 0 & \cdots & 0 & & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\
\frac{\partial g_M^{(1)}}{\partial x_{11}} & \cdots & \frac{\partial g_M^{(1)}}{\partial x_{1Q}} & 0 & \cdots & 0 & & 0 & \cdots & 0 \\
0 & \cdots & 0 & \frac{\partial g_1^{(2)}}{\partial x_{21}} & \cdots & \frac{\partial g_1^{(2)}}{\partial x_{2Q}} & & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \frac{\partial g_M^{(1)}}{\partial x_{21}} & \cdots & \frac{\partial g_M^{(1)}}{\partial x_{2Q}} & & 0 & \cdots & 0 \\
 & \vdots & & & \vdots & & \ddots & & \vdots & \\
0 & \cdots & 0 & 0 & \cdots & 0 & & \frac{\partial g_1^{(N)}}{\partial x_{N1}} & \cdots & \frac{\partial g_1^{(N)}}{\partial x_{NQ}} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 0 & \cdots & 0 & & \frac{\partial g_M^{(N)}}{\partial x_{N1}} & \cdots & \frac{\partial g_M^{(N)}}{\partial x_{NQ}}
\end{array}\right], \tag{11}$$

while that with respect to the parameters $\boldsymbol{z}$ is the $MN \times P$-dimensional matrix

$$\nabla_z \boldsymbol{G} = \begin{bmatrix}
\frac{\partial g_1^{(1)}}{\partial z_1} & \cdots & \frac{\partial g_M^{(N)}}{\partial z_1} \\
\vdots & \ddots & \vdots \\
\frac{\partial g_1^{(1)}}{\partial z_P} & \cdots & \frac{\partial g_M^{(N)}}{\partial z_P}
\end{bmatrix} \tag{12}$$

If the variables of different experiments are uncorrelated, the covariance matrix $\boldsymbol{\sigma}$ consists of N diagonal blocks of dimension $Q \times Q$, that is, $\boldsymbol{\sigma} = \mathrm{diag}_N[\boldsymbol{\sigma}^{(1)}\ \boldsymbol{\sigma}^{(2)}\ \cdots\ \boldsymbol{\sigma}^{(N)}]$, where

$$\boldsymbol{\sigma}^{(i)} \equiv \begin{bmatrix}
\sigma_{11}^{(i)} & \cdots & \sigma_{1Q}^{(i)} \\
\vdots & \ddots & \vdots \\
\sigma_{Q1}^{(i)} & \cdots & \sigma_{QQ}^{(i)}
\end{bmatrix}, \qquad i = 1,\ldots,N$$

With these representations, the weight matrix $\boldsymbol{W}$ also has a block diagonal form, $\boldsymbol{W} = \mathrm{diag}_N[\boldsymbol{w}^{(1)}\ \cdots\ \boldsymbol{w}^{(N)}]$, where each matrix $\boldsymbol{w}^{(i)}$ is itself defined as a matrix triple product

$$\boldsymbol{w}^{(i)} = \left\{ \left(\begin{bmatrix}
\frac{\partial g_1^{(i)}}{\partial x_{11}} & \cdots & \frac{\partial g_1^{(i)}}{\partial x_{1Q}} \\
\vdots & \ddots & \vdots \\
\frac{\partial g_M^{(i)}}{\partial x_{11}} & \cdots & \frac{\partial g_M^{(i)}}{\partial x_{1Q}}
\end{bmatrix}
\begin{bmatrix}
\sigma_{11}^{(i)} & \cdots & \sigma_{1Q}^{(i)} \\
\vdots & \ddots & \vdots \\
\sigma_{Q1}^{(i)} & \cdots & \sigma_{QQ}^{(i)}
\end{bmatrix}
\times \begin{bmatrix}
\frac{\partial g_1^{(i)}}{\partial x_{11}} & \cdots & \frac{\partial g_1^{(i)}}{\partial x_{11}} \\
\vdots & \ddots & \vdots \\
\frac{\partial g_M^{(i)}}{\partial x_{1Q}} & \cdots & \frac{\partial g_M^{(i)}}{\partial x_{1Q}}
\end{bmatrix} \right)^{-1} \right\} \tag{13}$$

Calculation of the variable increments at each iteration can be simplified by introduction of the MN-dimensional vector $\boldsymbol{\phi}' = \boldsymbol{G}_1 - (\nabla_x \boldsymbol{G})\boldsymbol{v}_1^T + (\nabla_z \boldsymbol{G})(\Delta \boldsymbol{z})^T$, in terms of which

$$(\Delta \boldsymbol{x})^T = \boldsymbol{x}_0^T - \boldsymbol{x}_1^T - \boldsymbol{\sigma}(\nabla_x \boldsymbol{G})^T \boldsymbol{W}\boldsymbol{\phi}' \tag{14}$$

The block-diagonal structure of the matrices appearing in this equation again allows the variable increments to be represented as products of smaller matrices. Thus, if the elements of vectors $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$ at point i are represented as $x_{ij}^{(0)}$ and $x_{ij}^{(1)}$, respectively, where $j = 1,\ldots,Q$

$$\begin{bmatrix} \Delta x_{i1} \\ \vdots \\ \Delta x_{iQ} \end{bmatrix} = \begin{bmatrix} x_{i1}^{(0)} - x_{i1}^{(1)} \\ \vdots \\ x_{iQ}^{(0)} - x_{iQ}^{(i)} \end{bmatrix} - \begin{bmatrix} \sigma_{11}^{(i)} & \cdots & \sigma_{1Q}^{(i)} \\ \vdots & \ddots & \vdots \\ \sigma_{Q1}^{(i)} & \cdots & \sigma_{QQ}^{(i)} \end{bmatrix}$$
$$\times \begin{bmatrix} \dfrac{\partial g_1^{(i)}}{\partial x_{i1}} & \cdots & \dfrac{\partial g_M^{(i)}}{\partial x_{i1}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial g_1^{(i)}}{\partial x_{iQ}} & \cdots & \dfrac{\partial g_M^{(i)}}{\partial x_{iQ}} \end{bmatrix} \begin{bmatrix} W_{11}^{(i)} & \cdots & W_{1M}^{(i)} \\ \vdots & \ddots & \vdots \\ W_{M1}^{(i)} & \cdots & W_{MM}^{(i)} \end{bmatrix} \begin{bmatrix} \phi'_{i1} \\ \vdots \\ \phi'_{iM} \end{bmatrix} \tag{15}$$

An alternative expression can be obtained by multiplying out each product, working from right to left. Thus, the variable increment Δx_{im}, $i = 1,\ldots,N$, $m = 1,\ldots,Q$ is given by

$$\Delta x_{im} = x_{im}^{(0)} - x_{im}^{(1)} - \sum_{j=1}^{M} \sum_{k=1}^{M} \sum_{l=1}^{Q} \sigma_{ml}^{(i)} W_{jk}^{(i)} \frac{\partial g_j^{(i)}}{\partial x_{i1}} \phi'_{ik} \tag{16}$$

In some situations it is useful to distinguish between dependent and independent variables. Thus, each group of Q variables in $\boldsymbol{x}$ can be subdivided into M responses y_{ij}, $j = 1,\ldots,M$ and $R \equiv Q - M$ regressors x_{ij}, $j = M+1,\ldots,Q$. The functional constraints that are most often encountered in this case are explicit, that is, $\boldsymbol{G} \equiv \boldsymbol{Y} - \boldsymbol{F}$, where $\boldsymbol{Y} \equiv [\boldsymbol{y}^{(1)}\ \boldsymbol{y}^{(2)} \cdots \boldsymbol{y}^{(N)}]^T$ and $\boldsymbol{F} \equiv [\boldsymbol{f}^{(1)}\ \boldsymbol{f}^{(2)} \cdots \boldsymbol{f}^{(N)}]^T$ are MN-dimensional vectors containing the observed and predicted values of the response variables, respectively. Evidently, $\nabla_z \boldsymbol{G} = -\nabla_z \boldsymbol{F}$. Each block of the weighting matrix assumes one of a variety of simpler forms, depending on the nature of the correlations that exist between the measurements. Thus, if correlations exist between the responses and between the regressors, each block of the covariance matrix consists of two blocks of dimensions $M \times M$ and $R \times R$

$$\boldsymbol{w}^{(i)} = \left\{ \begin{bmatrix} 1 & \cdots & 0 & -\dfrac{\partial f_1^{(i)}}{\partial x_{i,M+1}} & \cdots & -\dfrac{\partial f_1^{(i)}}{\partial x_{iQ}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & -\dfrac{\partial f_M^{(i)}}{\partial x_{i,M+1}} & \cdots & -\dfrac{\partial f_M^{(i)}}{\partial x_{iQ}} \end{bmatrix} \right.$$
$$\times \begin{bmatrix} \sigma_{y11}^{(i)} & \cdots & \sigma_{y1M}^{(i)} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{yM1}^{(i)} & \cdots & \sigma_{yMM}^{(i)} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_{xM+1,1}^{(i)} & \cdots & \sigma_{xM+1,Q}^{(i)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma_{xQ,M+1}^{(i)} & \cdots & \sigma_{xQQ}^{(i)} \end{bmatrix}$$
$$\left. \times \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \\ -\dfrac{\partial f_1^{(i)}}{\partial x_{i,M+1}} & \cdots & -\dfrac{\partial f_M^{(i)}}{\partial x_{i,M+1}} \\ \vdots & \ddots & \vdots \\ -\dfrac{\partial f_1^{(i)}}{\partial x_{iQ}} & \cdots & -\dfrac{\partial f_M^{(i)}}{\partial x_{iQ}} \end{bmatrix} \right\}^{-1}, \tag{17}$$

where $\sigma_{yjk}^{(i)} = \mathrm{cov}(y_{ij}, y_{ik})$, and $\sigma_{xjk}^{(i)} = \mathrm{cov}(x_{ij}, x_{ik})$. After tedious but straightforward algebra, it can be shown that the jkth element of the inverse of this matrix is

$$(\boldsymbol{w}^{(i)-1})_{jk} = \sigma_{yjk}^{(i)} + \sum_{l=M+1}^{Q} \sum_{m=M+1}^{Q} \sigma_{xlm}^{(i)} \frac{\partial f_j^{(i)}}{\partial x_{il}} \frac{\partial f_k^{(i)}}{\partial x_{im}}. \tag{18}$$

Multiresponse Weighted Least Squares

From Eq. 18 it follows that in fixed-regressor models, for which $\sigma_{xjk}^{(i)} = 0$, the weighting matrix reduces to

$$\boldsymbol{w}^{(i)} = \begin{bmatrix} \sigma_{y11}^{(i)} & \cdots & \sigma_{y1M}^{(i)} \\ \vdots & \ddots & \vdots \\ \sigma_{yM1}^{(i)} & \cdots & \sigma_{yMM}^{(i)} \end{bmatrix}^{-1}. \tag{19}$$

Arranging the errors in the dependent variables (identified by hypothesis with the residuals $\boldsymbol{Y} - \boldsymbol{F}$ of the fit) in a column vector $\boldsymbol{G}$, the log-likelihood function L is (Bard, 1974, pp. 20, 62)

$$\ln L = -\frac{NM}{2}\ln(2\pi) - \frac{1}{2}\ln\det \boldsymbol{V} - \frac{1}{2}\boldsymbol{G}^T \boldsymbol{W} \boldsymbol{G} \tag{20}$$

where $\boldsymbol{W}$ is a block-diagonal matrix with elements defined by Eq. 19, and $\boldsymbol{V} = \boldsymbol{W}^{-1}$. Maximization of L is clearly equivalent to minimization of the quadratic form

$$\boldsymbol{G}^T \boldsymbol{W} \boldsymbol{G} = \sum_{i=1}^{N} \boldsymbol{g}^{(i)T} \boldsymbol{w}^{(i)} \boldsymbol{g}^{(i)} \tag{21}$$

which is the objective function appropriate for multiequation WNLS (Aitken, 1935). Since the regressors are fixed, the calculated values contained in the array $\boldsymbol{F}$ depend only on the parameters z, and no additional constraints on the minimization are required. The application of the Gauss-Newton method to the iterative solution of this problem proceeds almost exactly as before, the only difference being that the vector $\boldsymbol{x}$ does not change. The coefficient matrix of the system of normal equations for the parameter estimates has the jkth element

$$\left((\nabla_z \boldsymbol{F})^T \boldsymbol{W} \nabla_z \boldsymbol{F}\right)_{jk} = \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{f}^{(i)}}{\partial z_j}\right)^T \boldsymbol{w}^{(i)} \frac{\partial \boldsymbol{f}^{(i)}}{\partial z_k} \qquad (j,k=1,\ldots,P) \quad (22)$$

while the righthand side is a column vector with components

$$\left((\nabla_z \boldsymbol{F})^T \boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{F}_1)\right)_k = \sum_{i=1}^{N} \left(\frac{\partial \boldsymbol{f}^{(i)}}{\partial z_k}\right)^T \boldsymbol{w}^{(i)}\left(\boldsymbol{y}^{(i)} - \boldsymbol{f}_1^{(i)}\right) \qquad (k=1,\ldots,P) \quad (23)$$

Several important limiting cases of Eqs. 22 and 23 can be identified. Thus, if all the weighting matrices are equal and proportional to the identity matrix of dimension $M \times M$, the common proportionality constant cancels out and minimization of the quadratic form in Eq. 23 is easily shown to be equivalent to minimizing the sum of squared residuals of all MN responses (Bates and Watts, 1988, chap. 4). This limiting case can be referred to as the diagonal uniform weighting approximation. When $M=1$, the arrays $\boldsymbol{f}_1^{(i)}$, $\boldsymbol{y}^{(i)}$, and $\boldsymbol{w}^{(i)}$ reduce to scalars $f_1^{(i)}$, $y^{(i)}$, and $w^{(i)}$, and the well-known results for single-equation WNLS are recovered [compare Press et al. (1992, pp. 675–683)]. When all the weights are equal, they reduce further to single-equation UNLS [compare Bates and Watts (1988, p. 40)]. The iterative procedures in all these limiting cases are fundamentally similar. It may be noted in passing [see Seber and Wild (1989, pp. 28–30), Gallant (1975)] that if the weighting matrix is positive definite, Cholesky decomposition (Press et al., 1992, pp. 89–91) can be used to transform a WNLS problem to a UNLS fit of linearly transformed response variables.

An important characteristic of the preceding formulation of WNLS can be brought to light by comparison with the unconstrained minimization of the sum of squares $S(z) = (\boldsymbol{Y} - \boldsymbol{F})^T \boldsymbol{W}(\boldsymbol{Y}-\boldsymbol{F})$. The condition for this minimum is that

$$\nabla_z S = -2(\boldsymbol{Y}-\boldsymbol{F})^T \boldsymbol{W} \nabla_z \boldsymbol{F} \quad (24)$$

but the convergence properties of the minimization algorithms are primarily determined by the matrix of second derivatives (or Hessian), which is

$$\nabla_z \nabla_z S = 2\left[(\nabla_z \boldsymbol{F})^T \boldsymbol{W} \nabla_z \boldsymbol{F} - (\boldsymbol{Y}-\boldsymbol{F})^T \boldsymbol{W} \nabla_z \nabla_z \boldsymbol{F}\right] \quad (25)$$

and can be related to the curvature of the response surface. Comparison with Eq. 23 shows that the coefficient matrix of the parameter increments is identical with the first part of the Hessian. The remaining part is seen to involve the residual vector as well the second derivatives of the response function. As discussed by Press et al. (1992, pp. 677–678), the convergence of the Gauss-Newton algorithm deteriorates when this second term in the Hessian is not negligible in comparison with the first (that is, when the coefficient matrix of Eq. 23 is a poor approximation to the Hessian). This can be expected when the residuals are large (or do not approximately cancel out) as a result of some systematic mismatch between the data and the fitting functions, or if the second derivatives are large (as can be expected for highly nonlinear models or poor parameter estimates). Dennis et al. (1981) developed a least-squares algorithm that incorporates an improved estimate of the Hessian of the sum of squared residuals, and uses it to construct a locally quadratic approximation of the objective function in parameter space [as do the quadratically convergent minimization algorithms of Powell (1964) and Fletcher and Reeves (1964)]. Although this possesses intrinsically better convergence properties than the Gauss-Newton algorithm considered in the present article, it is not clear how their approach could be adapted to the fit of random-regressor models.

In the discussion of GNLS, the fitting functions were introduced as functional constraints on the minimization of the weighted sum of squares. For explicit functions, an alternative approach is possible, which involves formulating the log-likelihood function in terms of conditional probabilities. In principle, the parameters in the fitting function are to be estimated conditionally upon the maximum-likelihood values of the regressors. These are unknown, but as suggested by Clutton-Brock (1967), the corresponding value of the fitting function can be approximated by performing a Taylor expansion about the measured values. Consider, for example, the simplest case of a single response to be fitted to a function f of several uncorrelated regressors whose true and measured values at point j are the components of the vectors $\boldsymbol{x}_j$ and $\boldsymbol{x}_{0j}$, respectively. If the corresponding values of the response variable are y_j and y_{0j}, then

$$y_j - f(\boldsymbol{x}_j) = y_j - y_{0j} = f(\boldsymbol{x}_{0j}) - f(\boldsymbol{x}_j) \approx y_j - y_{0j} + \sum_{k=1}^{R} (x_{jk} - x_{0jk}) \left(\frac{\partial f}{\partial x_k}\right)_{\boldsymbol{x}_{0j}} \quad (26)$$

The variance of this random variable is

$$\left\langle \left(y_j - y_{0j} + \sum_{k=1}^{R} (x_{jk} - x_{0jk}) \left(\frac{\partial f}{\partial x_k}\right)_{\boldsymbol{x}_{0j}} \right)^2 \right\rangle = \sigma_{yj}^2 + \sum_{k=1}^{R} \sigma_{xjk}^2 \left(\frac{\partial f}{\partial x_k}\right)_{\boldsymbol{x}_{0j}}^2 \quad (27)$$

and is identical to what could be obtained from Eq. 17; generalization is straightforward but tedious. The "effective variance" method (Box, 1970) is a form of WNLS in which the weighting matrices are defined in this way; in other words, using the weighting matrix appropriate to GNLS, but still treating the regressors as fixed. This approximation is the basis of kinetic model parameterization by Bard and Lapidus (1968), and of the VLE data-reduction method developed by Kemény et al. (1982). The latter authors observed that the complexity of calculations is "considerably reduced" by omitting the determination of the true values of the regressors. As discussed extensively by Lybanon (1984a,b) with particular reference to straight-line fitting, the effective variance method usually results in improvement, but converges to the true least-square solution only in very restricted circumstances (such as when the fitted slope is exactly zero).

Unknown Covariance Matrix

If the response in a fixed-regressor model are characterized by an unknown inverse covariance matrix $\boldsymbol{u}$, which is assumed to be the same for all experiments i, the ML method can be used to provide estimates of these unknown matrix elements in addition to the parameters z. Box and Draper (1965) applied Bayesian arguments to derive the objective function appropriate to this situation. Application of the ML method is described by Bard (1974, pp. 63–66) [see also Stewart et al. (1992), Anderson (1958, p. 347)], whose analysis starts from the observation that the triple-matrix product in Eq. 1 can be written in the alternative form

$$\sum_{i=1}^{N} \boldsymbol{r}^{(i)T}\boldsymbol{w}^{(i)}\boldsymbol{r}^{(i)} = \mathrm{Tr}[\boldsymbol{uM}] \tag{28}$$

where Tr denotes the sum of diagonal matrix elements and $\boldsymbol{M} \equiv \Sigma_{i=1}^{N}\boldsymbol{r}^{(i)}\boldsymbol{r}^{(i)T}$ is the *moment matrix* of the residuals. With these assumptions, the log-likelihood function assumes the form

$$\ln L = -\frac{MN}{2}\ln 2\pi - \frac{N}{2}\ln \det \boldsymbol{u} - \frac{1}{2}\mathrm{Tr}[\boldsymbol{uM}] \tag{29}$$

Maximization of this expression with respect to the matrix elements of $\boldsymbol{u}$ results in

$$\boldsymbol{u} = \frac{1}{N}\boldsymbol{M} \tag{30}$$

which, when used to eliminate the dependence of L on $\boldsymbol{u}$, gives

$$\ln L = \frac{MN}{2}(\ln 2\pi - 1) - \frac{N}{2}\ln \det \boldsymbol{M} \tag{31}$$

Maximization of L therefore requires minimization of the determinant of the moment matrix with respect to the parameters. If the responses are known to be uncorrelated, the moment matrix is diagonal, and the determinant is equal to the product of its diagonal elements. The log-likelihood function in this limit is therefore

$$\ln L = \frac{MN}{2}(\ln 2\pi - 1) - \frac{N}{2}\sum_{j=1}^{M}\ln M_{jj} \tag{32}$$

The use of the determinant criterion as the basis for multiresponse parameter estimation is thoroughly discussed in the monograph by Bates and Watts (1988, chap. 4); for the purposes of the present discussion, it is sufficient to summarize the two most important points of their treatment. The first is that such calculations are inherently nonlinear, even for models that are linear with respect to their parameters. Iterative procedures are always required; Bates and Watts (1987) [compare Bates and Watts (1988, pp. 141–146)] described a generalized Gauss-Newton method for this purpose. Second, the Hessian matrix is not necessarily positive definite. An important situation in which negative eigenvalues of the Hessian are possible is where multiple minima exist in parameter space, as in the differential equation model considered by these authors (Bates and Watts, 1988, pp. 188–194). Convergence difficulties that can arise in such situations can be dealt with by inflating the diagonal of the approximate Hessian, as in the well-known Levenberg (1944) and Marquardt (1963) methods. Most of the applications of the determinant criterion to chemical engineering seem to have involved parameter estimation in kinetic models (see, for example, Hunter (1967)). Schubert (1974) first applied it to VLE data reduction. This latter application is one of the many situations in chemical engineering where heteroscedasticity (unequal variance) is important; applicability of the determinant criterion in such situations is diminished by its underlying assumption that the unknown variance–covariance matrix is the same at each data point.

An alternative approach to the fit of multiresponse models is to estimate the unknown covariance matrix from the moment matrix of residuals. The theory of the estimation of the mean and covariance matrix from multivariate data is outlined by Anderson (1958, chap. 3); application to parameter estimation was first proposed by Zellner (1962) for linear models, and subsequently generalized by Beauchamp and Cornell (1966) to nonlinear models [compare Seber and Wild (1989, pp. 531–536)]. This approach involves fitting each equation separately to the appropriate subset of the data, forming the moment matrix from the residuals, and then carrying out a multiresponse WNLS fit to the entire data set using the inverse of this moment matrix as an estimate of the covariance matrix.

Binary Vapor–Liquid Equilibrium Data Reduction

The correlation of binary VLE measurements provides an interesting test of the WNLS and GNLS approaches to data reduction. Motivation for study of such calculations is provided not only by their considerable practical importance in chemical engineering, but also by the fact that the significance of the various fundamental assumptions just identified can be readily seen. The aim is to characterize the liquid phase by use of empirical or semiempirical representations of excess Gibbs energy as a function of composition. These functions involve adjustable parameters, values of which are to be inferred as reliably as possible from experimental VLE data.

Two main approaches to this parameter estimation problem can be identified. The first approach involves use of the measured vapor-phase composition Y and liquid-phase composition X to calculate the liquid-phase activity coefficients from the definition

$$\gamma_i = \frac{Y_i \varphi_i p}{X_i f_i^{0L}}, \tag{33}$$

where $i = 1,2$, p is the total pressure, φ_i is the vapor-phase fugacity coefficient, and f_i^{0L} is the fugacity of pure liquid i. The parameters are then determined either by a single equation fit of the excess molar Gibbs energy $g^E/RT = X_1 \ln \gamma_1 + X_2 \ln \gamma_2$, or by a two-equation fit of the activity coefficients (or some auxiliary function thereof), with covariances estimated by an error-propagation formula based on the Taylor-series approximation used to derive Eq. 27. Thus, if a function f depends on R normally distributed variables x_k with respective means and standard deviations (x_{0k}, σ_{xk}), and $f_0 = f(\boldsymbol{x}_0)$, the variance of f is

$$\begin{aligned} \langle (f - f_0)^2 \rangle &= \left\langle \left(\sum_{k=1}^{R} (x_k - x_{0k}) \left(\frac{\partial f}{\partial x_k} \right)_{x_0} \right)^2 \right\rangle \\ &= \sum_{k=1}^{R} \sigma_{xk}^2 \left(\frac{\partial f}{\partial x_k} \right)_{x_{0j}}^2 . \end{aligned} \tag{34}$$

Ulrichson and Stevenson (1972) used this result in their analysis of the effects of experimental errors on thermodynamic consistency. An example of an auxiliary function f is $\ln(\gamma_1/\gamma_2)$, which possesses the advantage that the dependence on the pressure cancels out (Van Ness et al., 1973; Kemény et al., 1982). Kemény et al. (1982) examined many similar functions; although they did not conclude that any one of them was markedly better than the others, their work showed that inclusion of weights led to consistently superior results. The other approach, advocated by Anderson et al. (1978), involves fit of the data directly by a two-response GNLS procedure with implicit constraints derived from the isofugacity conditions for each component. No consensus appears to have been reached regarding which general approach is better, but Eq. 33 suggests several potential complications associated with the use of activity coefficients. Consider first the uncertainty in γ_1 associated with respective measurement errors δp, δX_1, and δY_1. At constant temperature this can be estimated from

$$\delta\gamma_1 = \gamma_1 \left[\frac{\delta p}{p} + \frac{\delta X_1}{X_1} + \frac{\delta Y_1}{Y_1} \right]; \tag{35}$$

if $X_1, Y_1 \approx 1$, $\delta X_1/X_1$ and $\delta Y_1/Y_1$ are typically between 0.0001 and 0.001, but for mixtures that are very dilute in component 1, they might well be comparable to the value of γ itself. Second, the existence of large differences in the uncertainties of γ_1 and γ_2 at the extremes of the composition range also raises the possibility of correlation between these quantities. Third, use of activity coefficients or auxiliary functions of them as responses amounts to a change in variables from the measured quantities in Eq. 33; the distribution of these transformed variables is related to the distribution of the primary measurements by well-established procedures [see, for example, Anderson (1958, pp. 10–11)]. The question of whether this new distribution is normal (or close to normal) now assumes great importance, considering the fundamental hypothesis of least-squares methods that errors in the responses are normally distributed. Derivation of a good analytical approximation to the joint distribution of γ_1 and γ_2 is feasible, but complicated; the details will be presented in a subsequent article. For present purposes, however, the validity of the error propagation formula and normality of the responses will be assumed.

Table 1. Vapor–Liquid Equilibrium Data in the Chloroform(1)–1,4-Dioxane(2) System at 323 K

p/MPa	X_1	Y_1	$Y_1 p/X_1 p_1^0$	$Y_2 p/X_2 p_2^0$	γ_1*,**,†	γ_2
0.015790	0.0000	0.0000	—	—	1.0000	1.0000
0.017510	0.0932	0.1794	0.4859	1.0035	0.4955	1.0026
0.018150	0.1248	0.2383	0.4997	1.0004	0.5094	0.9991
0.019300	0.1757	0.3302	0.5229	0.9932	0.5329	0.9913
0.019890	0.2000	0.3691	0.5292	0.9934	0.5391	0.9912
0.021370	0.2626	0.4628	0.5430	0.9860	0.5528	0.9830
0.024950	0.3615	0.6184	0.6154	0.9444	0.6256	0.9398
0.029820	0.4750	0.7552	0.6835	0.8806	0.6937	0.8742
0.034800	0.5555	0.8378	0.7567	0.8042	0.7665	0.7963
0.042100	0.6718	0.9137	0.8255	0.7011	0.8339	0.6916
0.060380	0.8780	0.9860	0.9776	0.4388	0.9809	0.4288
0.065390	0.9398	0.9945	0.9976	0.3784	0.9991	0.3687
0.069360	1.0000	1.0000	1.0000	1.0000	—	—

Source: McGlashan and Rastogi (1958).

*Vapor fugacity coefficients calculated from the volume-explicit two-term virial equation.

**Second virial coefficients obtained from the correlation of Pitzer and Curl, Jr. (1957).

†Poynting corrections calculated with liquid molar volumes from the correlation of Yen and Woods (1966).

Example: The Chloroform(1) – Dioxane(2) system

Coexisting vapor- and liquid-phase mole fractions in mixtures of chloroform(1) and 1,4-dioxane(2) at 323 K were measured by McGlashan and Rastogi (1958). These data are reproduced in Table 1, together with activity coefficients calculated according to

$$\gamma_1 = \frac{Y_1 \varphi_1 p}{X_1 f_1^{0L}} \approx \frac{Y_1 p}{X_1 p_1^0} \qquad \gamma_2 = \frac{Y_2 \varphi_2 p}{X_2 f_2^{0L}} \approx \frac{Y_2 p}{X_2 p_2^0}. \tag{36}$$

Assuming the vapors to be described by the volume-explicit virial equation

$$V = \frac{nRT}{p} \left[1 + \frac{\boldsymbol{B} p}{RT} \right], \tag{37}$$

where $\boldsymbol{B} = B_{11} Y_1^2 + 2B_{12} + B_{22} Y_2^2$, the fugacity coefficients in the mixed vapor are

$$\phi_1 = \exp\left[\frac{(B_{11} + \delta_{12} Y_2^2) p}{RT} \right] \qquad \phi_2 = \exp\left[\frac{(B_{22} + \delta_{12} Y_1^2) p}{RT} \right] \tag{38}$$

where $\delta_{12} = 2B_{12} - B_{11} - B_{22}$ is a physical measure of the asymmetry of interactions in the mixture. The corresponding expression for the pure liquid fugacity is

$$f_k^{0L} = \varphi_k^0 p_k^0 \exp\left[\frac{(p - p_k^0)v_k}{RT}\right] \qquad (k = 1,2) \tag{39}$$

where p_k^0 is the vapor pressure and v_k is the molar volume of each saturated pure liquid k (assumed to be incompressible), and the fugacity coefficients of the pure saturated vapors are

$$RT \ln \varphi_k^0 = \int_0^{p_k^0}\left[\frac{V}{n} - \frac{p}{RT}\right]dp = B_{kk}p_k^0 \qquad (k = 1,2) \tag{40}$$

In these calculations, the vapor pressure was calculated from the Antoine equation

$$\ln(p/\text{atm}) = a - \frac{b}{c + (T/K)} \tag{41}$$

in which, following the suggestions of Van Ness et al. (1973), the a parameters were adjusted to the experimentally measured vapor pressures for the pure components. Liquid volumes and virial coefficients were determined as a function of reduced temperature by means of corresponding-states correlations developed by Yen and Woods (1966) and Pitzer and Curl (1957), respectively. Values of the critical constants required for these calculations were obtained from Reid et al. (1977, pp. 629–677), and are given in Table 2.

McGlashan and Rastogi (1958) fitted their data to an associated ideal-solution model; for present purposes it is preferable to assume a two-parameter Redlich-Kister function (Prausnitz, 1969, pp. 205–208), since this representation is far more commonly used in engineering applications. Accordingly, the excess molar Gibbs energy is given by

$$\frac{g^E}{RT} = X_1(1 - X_1)\left[A + B(2X_1 - 1)\right] \tag{42}$$

where A and B are adjustable. If $z \equiv [A,B]$, the activity coefficient for each component (defined with respect to a Raoult Law standard state) can then be obtained from the Gibbs-Duhem equation. With $x_{i1} = X_1^{(i)}$, the responses for a two-equation fixed-regressor fit are

$$y_1^{(i)} \equiv \ln \gamma_1(x_{i1}, z) = (A + 3B)\left(1 - X_1^{(i)}\right)^2 - 4B\left(1 - X_1^{(i)}\right)^3 \tag{43}$$

$$y_2^{(i)} \equiv \ln \gamma_2(x_{i1}, z) = (A - 3B)\left(1 - X_1^{(i)}\right)^2 + 4B\left(1 - X_1^{(i)}\right)^3 \tag{44}$$

Table 2. Relevant Thermophysical Properties of the Components

Property	Chloroform, $CHCl_3$	1,4-Dioxane, $C_4H_8O_2$
Critical temperature, T_c/K	536.40	587.00
Critical pressure, p_c/Pa	$0.54716E+07$	$0.52081E+07$
Critical volume, v_c/m^{3}mol^{-1}	$0.23900E-03$	$0.23800E-03$
Acentric factor, ω	0.2160	0.2880
Antoine a*	15.9732	16.1327
Antoine b	2696.79	2966.88
Antoine c	−46.16	−62.15
T_{max}/K**	370.00	410.00
T_{min}/K**	260.00	275.00
Molar mass/g mol^{-1}	119.377940	88.106320

Source: Reid et al. (1977), p. 629.

*Agreement with the experimental end-member vapor pressures measured by McGlashan and Rastogi (1958) is obtained with $a = 15.9912$ for chloroform and $a = 16.1427$ for dioxane.

**Maximum and minimum temperatures for validity of the Antoine correlation.

Smith and Van Ness (1987, p. 352) fitted these data to the two-suffix Margules equations,

$$\ln \gamma_1 = (1 - X_1)^2\left[A_{12} + 2(A_{21} - A_{12})X_1\right] \tag{45}$$

$$\ln \gamma_2 = X_1^2\left[A_{21} + 2(A_{12} - A_{21})(1 - X_1)\right] \tag{46}$$

which are equivalent to the Redlich-Kister equations with $A = (A_{21} + A_{12})/2$ and $B = (A_{21} - A_{12})/2$. The method they used [compare Van Ness et al. (1973)] involved constructing a plot of $g^E/RTX_1(1 - X_1)$ against X_1 and determining A_{21} and A_{12} by linear regression, which yielded $A_{21} = -1.27$ and $A_{12} = -0.72$, or $A = -0.995$ and $B = -0.275$. These values are reasonable initial estimates of the parameters.

Unknown Covariance Matrix. Equations 43 and 44 can be fitted simultaneously to the data in columns 2, 6, and 7 of Table 1 by conjugate-gradient minimization of the determinant criterion. At $z = (0,0)$, the objective function is 1.2364326 and its gradient is $(-6.8105102, -0.9174459)$. A line-search along this direction by means of the "Golden Section" algorithms described by Press et al. (1992, pp. 387–398) leads to a minimum value of 0.0000147 at $(-0.9699401, -0.1306609)$, and search along the second conjugate gradient direction $(0.0000634, -0.0004707)$ leads to a minimum of 0.0000016 at $(-0.9583910, -0.2164391)$, which satisfies the convergence requirement; the residual standard deviation is $0.1960970E-01$. That the method converges in one cycle of conjugate directions indicates that the objective function can be very closely approximated by a quadratic form [compare Wismer and Chattergy (1978, p. 225)]. Minimization of the objective function corresponding to an assumed diagonal covariance matrix also proceeds very rapidly, giving a minimum of 0.0000043 at $(-0.9703534, -0.2919119)$, with residual standard deviation $0.1583315E-01$. It is interesting to compare these results with those obtained by fitting the approximate activity coefficients in columns 4 and 5 of Table 1 (corresponding to the neglect of Poynting corrections and the assumption of ideal vapor phases). The full determinant criterion has a minimum of 0.0000016 at $(-0.9665991, -0.1771526)$, with residual standard deviation $0.2949802E-01$, while the minimum of the diagonal determinant criterion is 0.0000238 at $(-0.9905318, -0.2958869)$, with residual standard deviation $0.2213936E-01$. Thus, the activity coefficients seem to be better represented by the assumed model than their approximations as $\gamma_1 \approx Y_1 p/X_1 p_1^0$ and $\gamma_2 \approx Y_2 p/X_2 p_2^0$.

The alternative (Beauchamp-Cornell-Zellner) treatment of an unknown covariance matrix leads to a similar value for A, but a markedly different value for B. Single-response UNLS fit of the first response function to the values of γ_1 gives the parameter estimate $z = (-1.0360812 \pm 0.0095324, -0.3056035 \pm 0.0180141)$ with a residual standard deviation of $0.802938E-02$, while similar fit of the second response function to the values of γ_2 gives $z = (-0.9271885 \pm 0.0087831, -0.2927211 \pm 0.0115625)$ with a residual standard deviation of $0.577219E-02$. Convergence in both cases was achieved in three iterations. If the residuals from each fit are contained in the column vectors $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, respectively, the moment matrix is

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{R}_1^T\boldsymbol{R}_1 & \boldsymbol{R}_1^T\boldsymbol{R}_2 \\ \boldsymbol{R}_2^T\boldsymbol{R}_1 & \boldsymbol{R}_2^T\boldsymbol{R}_2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5802382E-03 & -0.2812851E-03 \\ -0.2812851E-03 & 0.2998640E-03 \end{bmatrix}$$

$$= \begin{bmatrix} 0.3160747E+04 & 0.2964914E+04 \\ 0.2964914E+04 & 0.6116060E+04 \end{bmatrix}^{-1} \tag{47}$$

Dual-response GNLS fit of the activity coefficients with each block of the weighting matrix set equal to $\boldsymbol{M}^{-1}$ leads to the least-square estimate $z = (-0.9718955 \pm 0.0002143, -0.2461491 \pm 0.0004068)$. The residual standard deviation is $0.165381E-01$, which is seen to be far worse than obtained for either of the two single-response fits.

Known Covariance Matrix. The purpose of this section is to compare the results of dual-response WNLS calculations corresponding to first-principles estimates of the weighting matrices $\boldsymbol{w}^{(i)}$. For a two-equation model, it is necessary to consider defining the weighting matrix as

$$\boldsymbol{w}^{(i)} = \begin{bmatrix} \sigma_{11}^{(i)} & \sigma_{12}^{(i)} \\ \sigma_{21}^{(i)} & \sigma_{22}^{(i)} \end{bmatrix}^{-1} = \frac{1}{\sigma_{11}^{(i)}\sigma_{22}^{(i)} - \sigma_{21}^{(i)}\sigma_{12}^{(i)}} \begin{bmatrix} \sigma_{22}^{(i)} & -\sigma_{12}^{(i)} \\ -\sigma_{21}^{(i)} & \sigma_{11}^{(i)} \end{bmatrix} \tag{48}$$

Specification of the off-diagonal elements of $\boldsymbol{\sigma}$ is equivalent to fixing the value of the correlation coefficient $c^{(i)} = \sigma_{12}^{(i)}/\sqrt{\sigma_{11}^{(i)}\sigma_{22}^{(i)}}$. For example, with $w_{11}^{(i)} = 1/\sigma_{11}^{(i)}$, $w_{22}^{(i)} = 1/\sigma_{22}^{(i)}$, and $w_{12}^{(i)} = w_{21}^{(i)} = \sqrt{w_{11}^{(i)}w_{22}^{(i)}}$, the correlation coefficient $c^{(i)}$ satisfies the equation

$$\frac{1}{\sqrt{\sigma_{11}^{(i)}\sigma_{22}^{(i)}}} = -\frac{c^{(i)}}{(1-c^{(i)2})\sqrt{\sigma_{11}^{(i)}\sigma_{22}^{(i)}}}$$

$$\text{or} \qquad c^{(i)2} - c^{(i)} - 1 = 0 \tag{49}$$

which yields the solution $c^{(i)} = (1-\sqrt{5})/2 \approx -0.618$. In principle, the off-diagonal elements can be estimated by use of a generalization of Eq. 34, but taking into account that the Poynting corrections are only slightly different from 1, and that the vapor-phase fugacity coefficients do not depend strongly on composition, a reasonable approximation is to work out the weights from the approximate formulas $\gamma_1 \approx Y_1 p/X_1 p_1^0$ and $\gamma_2 \approx Y_2 p/X_2 p_2^0$ rather than Eq. 33. The required partial derivatives are thus found to be

$$\frac{\partial\gamma_1}{\partial X_1} = -\frac{\gamma_1}{X_1} \qquad \frac{\partial\gamma_1}{\partial Y_1} = \frac{\gamma_1}{Y_1} \qquad \frac{\partial\gamma_1}{\partial p} = \frac{\gamma_1}{p} \tag{50}$$

$$\frac{\partial\gamma_2}{\partial X_1} = \frac{\gamma_2}{X_2} \qquad \frac{\partial\gamma_2}{\partial Y_1} = -\frac{\gamma_2}{Y_2} \qquad \frac{\partial\gamma_2}{\partial p} = \frac{\gamma_2}{p} \tag{51}$$

so that

$$\sigma_{11} = \gamma_1^2\left[\frac{\sigma_{X_1}}{X_1^2} + \frac{\sigma_{Y_1}}{Y_1^2} + \frac{{\sigma_p}^2}{p}\right] \tag{52}$$

$$\sigma_{22} = \gamma_2^2\left[\frac{\sigma_{X_1}}{X_2^2} + \frac{\sigma_{Y_1}}{Y_2^2} + \frac{{\sigma_p}^2}{p}\right] \tag{53}$$

$$\sigma_{12} = \gamma_1\gamma_2\left[\frac{\sigma_{X_1}}{X_1X_2} + \frac{\sigma_{Y_1}}{Y_1Y_2} + \frac{{\sigma_p}^2}{p}\right]. \tag{54}$$

Setting $\sigma_X = 0$ (for a fixed-regressor model), $\sigma_p = 0.0001p$ and $\sigma_Y = 0.0005$, the weighting matrix elements following from this assumed correlation coefficient can be arranged as rows of the array

$$\boldsymbol{W}^* = \begin{bmatrix} 0.523671E+06 & 0.117840E+07 & 0.117840E+07 & 0.260930E+07 \\ 0.873383E+06 & 0.141631E+07 & 0.141631E+07 & 0.227220E+07 \\ 0.152909E+07 & 0.165999E+07 & 0.165999E+07 & 0.179397E+07 \\ 0.186487E+07 & 0.172706E+07 & 0.172706E+07 & 0.159514E+07 \\ 0.277975E+07 & 0.181205E+07 & 0.181205E+07 & 0.118097E+07 \\ 0.384957E+07 & 0.159047E+07 & 0.159047E+07 & 0.655667E+06 \\ 0.463494E+07 & 0.121046E+07 & 0.121046E+07 & 0.312911E+06 \\ 0.464827E+07 & 0.885744E+06 & 0.885744E+06 & 0.165787E+06 \\ 0.464701E+07 & 0.545177E+06 & 0.545177E+06 & 0.622648E+05 \\ 0.389041E+07 & 0.131204E+06 & 0.131204E+06 & 0.426386E+04 \\ 0.381243E+07 & 0.593814E+05 & 0.593814E+05 & 0.890100E+03 \end{bmatrix} \tag{55}$$

The variability in the diagonal elements (in the first and last columns) lends support to the earlier speculation that heteroscedasticity might be important. Convergence from the initial estimate $z = (-1.0, -0.3)$ requires four iterations, and results in $z = (-1.0190947 \pm 0.0000107,\ -0.3016926 \pm 0.0000215)$, with a residual standard deviation of $0.150619E-01$. On the other hand, assumption of a diagonal weighting model gives $z = (-1.0310451 \pm 0.0000135,\ -0.3192129 \pm 0.0000207)$, with residual standard deviation $0.163342E-01$. In the special case where $\boldsymbol{w}^{(i)} = \sigma^{-2}\boldsymbol{I}_2$, for $i = 1, \ldots, N$, the value of σ has no effect on the parameter estimates, and the objective function reduces to the sum of squared residuals for both responses. The parameter estimates are $z = (-0.9982761 \pm 0.0132031,\ -0.2923256 \pm 0.0214288)$, with residual standard deviation $0.142193E-01$. Repetition of this calculation with columns 4 and 5 used as the responses in place of columns 6 and 7 gives $z = (-1.0100139 \pm 0.0202558,\ -0.3054691 \pm 0.0328380)$, with residual standard deviation $0.216428E-01$. This case is of particular interest, since the objective function corresponds most closely in form to that used in the DECHEMA VLE data collection (Gmehling et al., 1977), where data reduction involved minimization of

$$S = \sum_{i=1}^{N} \left\{ \frac{[\gamma_{1i,\mathrm{exp}} - \gamma_{1i,\mathrm{calc}}]^2}{\gamma_{1i,\mathrm{exp}}^2} + \frac{[\gamma_{2i,\mathrm{exp}} - \gamma_{2i,\mathrm{calc}}]^2}{\gamma_{2i,\mathrm{exp}}^2} \right\} \tag{56}$$

over all data points i. Although this objective function differs from those considered in the sections on multiresponse generalized least squares and multiresponse weighted least squares in its use of relative rather than absolute residuals, its most significant characteristic is the neglect of cross-product terms that arise from correlation of the residuals. Gmehling et al. (1979, p. 441) fitted the same data set to the two-suffix Margules equation and obtained $A_{12} = -0.7761$ and $A_{21} = -1.2291$, which corresponds to the Redlich-Kister parameters $A = -1.0026$ and $B = -0.2265$. The estimate of A is in reasonably good agreement with that obtained here, but that of B is not. These authors used in the "downhill simplex" optimization algorithm of Nelder and Mead (1965) [see also Press et al. (1992, pp. 402–406)] to minimize the objective function. The convergence criterion of that method is based on the difference between the maximum and minimum function values at the vertices of the simplex in parameter space; in this respect it differs fundamentally from the least-square methods used in the present work, which are based on the relative decrease in parameter estimates between iterations. It is conceivable that this could account for the discrepancy in the B values, especially if a relatively liberal convergence criterion were applied to the simplex algorithm. [Gmehling et al. (1977, 1979) did not specify the convergence criterion that they used; the least-square calculations described in the present article required that the maximum relative change in the parameters was 10^{-6}.]

GNLS Fit of Pressure-Composition Data. In the preceding sections, the performance of a variety of methods for simultaneous nonlinear regression of activity coefficient data was compared. The results demonstrate the importance of heteroscedasticity and correlation effects, but the lack of concordance between the estimates produced by different methods suggests that none of the methods by which these effects were taken into account was entirely satisfactory. An apparent advantage of fitting the liquidus and vaporus data directly (without prior calculation of activity coefficients) is that the covariance matrix for the experimental measurements is diagonal under essentially all conditions of interest, with elements that can be unequivocally determined from the known precision of the apparatus. A further advantage of the GNLS approach is the absence of a need for an explicit functional relationship between the variables, which makes it easier to incorporate vapor-phase nonideality and Poynting corrections to the liquid fugacities. This fact gives rise to considerable flexibility in the way the functional constraints can be set up. For example, the equations between the observables can be expressed in the form

$$Y_1 = \frac{f_1^{0L}\gamma_1 X_1/\varphi_1}{f_1^{0L}\gamma_1 X_1/\varphi_1 + f_2^{0L}\gamma_2 X_2/\varphi_2} = \frac{1}{1 + \dfrac{X_2\gamma_2 f_2^{0L}\varphi_1}{X_1\gamma_1 f_1^{0L}\varphi_2}}$$

$$p = \frac{f_1^{0L}\gamma_1 X_1}{\varphi_1} + \frac{f_2^{0L}\gamma_2 X_2}{\varphi_2}, \tag{57}$$

which were used by Anderson et al. (1978, Eqs. 29 and 30). If the vector of regressors at each point i is identified as $(x_{i1}, x_{i2}, x_{i3}, x_{i4}) \equiv (X_1^{(i)}, Y_1^{(i)}, p^{(i)}, T^{(i)})$, the equations of condition to be satisfied are

$$g_1^{(i)} \equiv x_{i2} - \frac{f_1^{0L}\gamma_1 x_{i1}/\varphi_1}{f_1^{0L}\gamma_1 x_{i1}/\varphi_1 + f_2^{0L}\gamma_2(1 - x_{i1})/\varphi_2} = 0$$

$$g_2^{(i)} \equiv x_{i3} - f_1^{0L}\gamma_1 x_{i1}/\varphi_1 + f_2^{0L}\gamma_2(1 - x_{i1})/\varphi_2 = 0, \tag{58}$$

where the additional functional dependences $\gamma_{1,2} \equiv \gamma_{1,2}(\boldsymbol{x}_{i1}, z)$, $f_{1,2}^{0L} \equiv f_{1,2}^{0L}(x_{i3}, x_{i4})$, and $\varphi_{1,2} \equiv \varphi_{1,2}(x_{i2}, x_{i3}, x_{i4})$ have been suppressed for clarity. An alternative formulation is founded on the two isofugacity conditions

$$Y_1\varphi_1 p = f_1^{0L}\gamma_1 X_1 \qquad Y_2\varphi_2 p = f_2^{0L}\gamma_2 X_2 \tag{59}$$

which, with the same identifications of the regressors, give

$$g_1^{(i)} \equiv x_{i2}x_{i3}\varphi_1 - x_{i1}\gamma_1 f_1^{0L} = 0$$

$$g_2^{(i)} \equiv (1 - x_{i2})x_{i3}\varphi_2 - (1 - x_{i1})\gamma_2 f_2^{0L} = 0 \tag{60}$$

The relation between the GNLS approach and the use of activity coefficients can be made clearer by an elementary rearrangement of Eqs. 60 to the form

$$g_1^{(i)} \equiv \frac{x_{i2}x_{i3}\varphi_1}{x_{i1}f_1^{0L}} - \gamma_1 = 0$$

$$g_2^{(i)} \equiv \frac{(1 - x_{i2})x_{i3}\varphi_2}{(1 - x_{i1})f_2^{0L}} - \gamma_2 = 0 \tag{61}$$

Thus, fit of an activity coefficient model can be based either on the *explicit* functional dependence of the activity coefficients on composition, or on an *implicit* relationship between pressure, temperature, and composition. The advantage of the latter is that the precision, and hence the weights, of the primary measurements can be specified more easily than that of the activity coefficients, thereby avoiding the difficulties discussed in the preceding sections.

Consider, for example, the formulation based on Eqs. 58, assuming that the uncertainties of these measurements are $\delta X_1 = 0.0001$, $\delta Y_1 = 0.0005$, $\delta p = 0.00001$ MPa, and $\delta T = 0.01$ K. The diagonal weighting matrix is $\boldsymbol{W}_0 = \mathrm{diag}[\boldsymbol{w}_0^{(1)}\ \boldsymbol{w}_0^{(2)} \cdots \boldsymbol{w}_0^{(13)}]$, where $\boldsymbol{w}_0^{(i)} = \mathrm{diag}_4[0.10000E+09\ \ 0.40000E+07\ \ 0.10000E+11\ \ 0.10000E+05]$. The process is commenced by setting $z = [-0.995,\ -0.275]$, and $\boldsymbol{x}$ equal to the measured values $\boldsymbol{x}_0$. The latter quantities are arranged, as before, in a vector $\boldsymbol{x}_0 = [\boldsymbol{x}_{0,1}\ \boldsymbol{x}_{0,2} \cdots \boldsymbol{x}_{0,13}]$, where, for example, the measurements at the second point are $\boldsymbol{x}_{0,2} = [0.0932\ \ 0.1794\ \ 0.017510\ \ 323.13]$. (The temperature shown here is converted from the 1948 temperature scale to the current standard, ITS-90.) After one iteration, the maximum relative change in the parameters is 0.1174125, so the whole process is repeated with appropriately updated values of the parameters and variables. After the second, third, and fourth iterations, the maximum relative error is 0.0007799, 0.0000095, and 0.0000001, respectively; this can be regarded as satisfactory convergence. The least-square parameter estimates are therefore

$$z \equiv \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} -1.0226337 \\ -0.3075310 \end{bmatrix} \pm \begin{bmatrix} 0.0114681 \\ 0.0209026 \end{bmatrix}$$

and the weighted sum of squared residuals is $0.253479E+04$. These estimates are in close agreement with those resulting from the two-equation WNLS fit of the activity coefficients, based on the nondiagonal weighting matrix obtained from Eqs. 50–55. The variables $\boldsymbol{x}$ are given in Table 3; comparison with the original data in Table 1 shows that the discrepancies between the data and the estimated vector of observables are greatest for the least precise data, as one would expect. Of particular interest is the temperature: although T was held constant in the original experiment, Eq. 11 shows that discrepancies in T can be expected to result from the temperature dependence of the fitting function (as reflected in the Jacobian of $\boldsymbol{G}$ with respect to $\boldsymbol{x}$), as well as the finite precision of this quantity.

Table 3. Vapor–Liquid Equilibrium in the System Chloroform(1)–1,4-Dioxane(2) at 323 K

X_1	Y_1	p/MPa	T/K
0.0000000	0.0000000	0.0157900	323.1271365
0.0929705	0.1825691	0.0175074	323.1307229
0.1246635	0.2402813	0.0181473	323.1304529
0.1757748	0.3298046	0.0192795	323.1429187
0.1999541	0.3707644	0.0198633	323.1498767
0.2618180	0.4717041	0.0214462	323.0705909
0.3617015	0.6197775	0.0248771	323.1974282
0.4747493	0.7599869	0.0298283	323.1210941
0.5562825	0.8387629	0.0346698	323.2911750
0.6708511	0.9182753	0.0422111	322.9609804
0.8786873	0.9868404	0.0602990	323.2987842
0.9402911	0.9950950	0.0653274	323.2708125
1.0000000	1.0000000	0.0693600	323.1271365

Note: Estimates of measured variables from GNLS fit of data from McGlashan and Rastogi (1958).

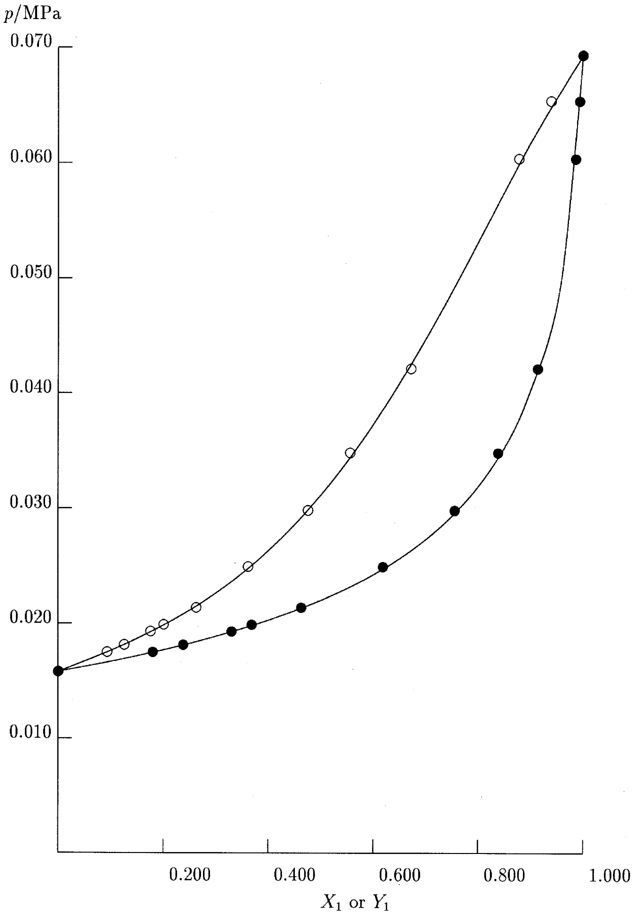

Figure 1. Phase diagram for the system chloroform(1)–1,4-dioxane(2) at 323 K.

Open circles: pressure as a function of liquid composition (McGlashan and Rastogi, 1958); filled circles: pressure as a function of vapor composition (McGlashan and Rastogi, 1958).

Iterative solution of the equations of condition is desirable to verify the success of the fit. "Bubble-pressure" and "dew-pressure" curves calculated as described by Smith and Van Ness (1987, pp. 381–393) are shown with the data in Figure 1, and confirm that the assumed activity coefficient model provides an excellent fit of the data. The quality of the fit and the rapidity of convergence also provide indirect confirmation of the adequacy of the numerical differentiation procedure used to calculate the Jacobians. Experience shows that if the derivatives are not sufficiently accurate, the calculation does not converge but meanders aimlessly through the parameter space, even though the starting estimate was quite good (as is the case in this example). Britt and Luecke (1973) used analytical derivatives, but the very satisfactory results obtained in the present work demonstrate that analytical derivatives are by no means essential for success. The alternative formulation of the functional constraints given in Eqs. 60 leads to the same values of the parameter estimates and standard errors. However, a rigorous comparison with calculations based on Eq. 61 is not possible, because the end-member vapor pressures correspond to compositions at which one of the components of Eq. 61 is undefined. But if the first and last data points in Table 1 are removed and the Antoine constants adjusted to give the correct end-member pressures (Table 2), slightly different results are obtained, viz.

$$z \equiv \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} -1.0227150 \\ -0.3076737 \end{bmatrix} \pm \begin{bmatrix} 0.0126648 \\ 0.0230839 \end{bmatrix}$$

The discrepancy in the standard errors can be attributed largely to the smaller data set.

Example: The Benzene(1)–Cyclohexane(2) system

The BLJ equations also retain the same general form if there is only one equation per point. A worthwhile test of this aspect of the operation of the program is to apply it to the analysis of vapor–liquid equilibrium data consisting of total vapor pressures as a function of liquid composition (with no measurements of vapor composition). Data of this kind are commonly analyzed by use of a procedure developed by Barker (1953), which involves fit of the pressure to the equation

$$p = \frac{f_1^{0L}\gamma_1 X_1}{\varphi_1} + \frac{f_2^{0L}\gamma_2 X_2}{\varphi_2}$$

with vapor-phase compositions and fugacity coefficients obtained iteratively by a bubble-pressure calculation. Prausnitz (1969, pp. 204–206) illustrated Barker's data-reduction scheme by fitting measurements by Hermsen and Prausnitz (1963) of vapor pressures as a function of liquid-phase mole fraction in the benzene(1)–cyclopentane(2) system to the two-parameter Redlich-Kister function, and presented a table of calculated activity coefficients at each experimental composition. An improved version of Barker's method, in which finite precision of all measured quantities is taken into account, can be illustrated by fitting the same data by single-response GNLS; the measurements are listed in Table 4, and the required thermophysical properties of the pure components are listed in Table 5.

These authors estimated the precision of their mole fractions to be ±0.0005, and that of their pressure measurements to be ±0.05 torr = 6.6661 Pa; the relative errors in the data of Table 4 are therefore comparable, which immediately indicates that a GNLS treatment is more appropriate than UNLS. Results are as follows

$$z_1 \equiv A = 0.4568161 \pm 0.0026628$$

$$z_2 \equiv B = -0.0163561 \pm 0.0044087$$

the mole fractions and pressures were assigned weights of 4×10^6 and 2.2504×10^{10}, respectively, and convergence from $A = 0.5$ and $B = -0.1$ required four iterations. The fit is shown in Figure 2.

Table 4. Activity Coefficients and Total Vapor Pressures for Mixtures of Benzene(1)–Cyclopentane(2) at 298 K

x_1	γ_1	γ_2	p/torr
0.1417	1.408	1.010	294.11
0.2945	1.253	1.043	268.36
0.4362	1.151	1.095	243.29
0.5166	1.108	1.135	228.33
0.5625	1.087	1.160	218.98
0.8465	1.010	1.380	148.07

Source: Hermsen and Prausnitz, 1963.

Table 5. Relevant Thermophysical Properties of the Components

Property	Benzene, C_6H_6	Cyclopentane, C_5H_{10}
Critical temperature, T_c/K	562.10	511.60
Critical pressure, p_c/Pa	$0.48940E+07$	$0.45089E+07$
Critical volume, v_c/m^{3}mol^{-1}	$0.25900E-03$	$0.26000E-03$
Acentric factor, ω	0.2120	0.1920
Antoine *a*	15.9008	15.8574
Antoine *b*	2788.51	2588.48
Antoine *c*	−52.36	−41.79
$T_{\max}$/K*	377.00	345.00
$T_{\min}$/K*	280.00	230.00
Molar mass/g mol^{-1}	78.113640	70.134400

Source: Reid et al. (1977), p. 629.
*Maximum and minimum temperatures for validity of the Antoine correlation.

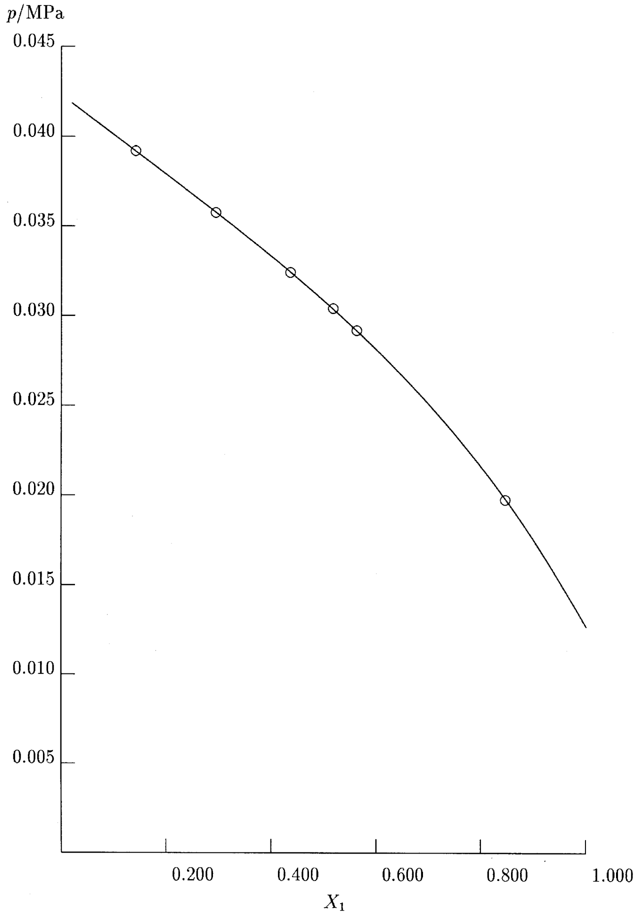

Figure 2. Total pressure in the benzene(1)–cyclopentane(2) system, as measured by Hermsen and Prausnitz (1963) and correlated by a GNLS implementation of Barker's method.

Ternary Liquid–Liquid Equilibrium Data Reduction

In the preceding sections, estimation of parameters in equilibrium models defined by one and two equations was demonstrated. From the phase rule, it follows that three equations are required to describe equilibration of two components between three phases (as in vapor–liquid equilibrium of a binary liquid mixture of limited miscibility), or of three components between two phases (as in the partitioning of a common solute between two liquids). The relevance of the latter situation to the solvent extraction of benzenoid aromatic compounds from gasoline has motivated numerous investigations of ternary LLE, for example, the studies by Lee and Kim (1995, 1998) of the mutual solubilities in the tetrahydrothiophene-1,1-dioxide (sulfolane)(1)–octane(2)–benzene(3) system.

From the data-reduction viewpoint, the need for GNLS is indicated even more strongly in LLE than in VLE, since the limitations of analytical techniques make it inherently impossible to distinguish dependent and independent variables. But GNLS does not appear to have been applied to such problems. For example, Lee and Kim (1995) correlated their ternary LLE measurements with the UNIQUAC and NRTL excess Gibbs energy models, by application of the "direct search" optimization procedure developed by Hooke and Jeeves (1961) to the minimization of the sum of squared differences between the observed and calculated liquid compositions, but no account was taken of the effect of finite precision on the parameter estimates. A further complication that arises is that there are two independent mole fractions in each phase, and only three equilibrium conditions to be satisfied. Use of the same least-squares criterion used by Lee and Kim (1995, 1998) would therefore require simultaneous fit of four equations to the data, in which iterative calculation of the equilibrium composition at each point and each assumed parameter vector would be necessary. A simpler and more efficient alternative would involve adopting the three isofugacity (or isoactivity) conditions as implicit functional constraints to be satisfied by the iterated values of the variables and parameters.

However, a more fundamental problem that must first be addressed is the assignment of appropriate weights to the variables. Lee and Kim (1995) stated that the mass fractions of the coexisting phases were reproducible to about ± 0.003, but presented their results as mole fractions of components 1, 2, and 3, given to four decimal places. To resolve this inconsistency, the mass fractions w_1, w_2, w_3 and mole fractions X_1, X_2, X_3 can be expressed as functions of the actual masses used, m_1, m_2, m_3, according to the equations

$$w_i = \frac{m_i}{m_1 + m_2 + m_3}, \qquad i = 1, 2, 3 \tag{62}$$

$$X_i = \frac{m_i/M_i}{m_1/M_1 + m_2/M_2 + m_3/M_3}, \qquad i = 1, 2, 3 \tag{63}$$

where M_i is the molar mass of component i. Then the partial derivatives of each w_i and each X_i with respect to each m_i can be arranged in the Jacobian matrices

$$\frac{\partial(w_1,w_2,w_3)}{\partial(m_1,m_2,m_3)} = \begin{bmatrix} \dfrac{w_1(1-w_1)}{m_1} & \dfrac{-w_1^2}{m_1} & \dfrac{-w_1^2}{m_1} \\ \dfrac{-w_2^2}{m_2} & \dfrac{w_2(1-w_2)}{m_2} & \dfrac{-w_2^2}{m_2} \\ \dfrac{-w_3^2}{m_3} & \dfrac{-w_3^2}{m_3} & \dfrac{w_3(1-w_3)}{m_3} \end{bmatrix} \tag{64}$$

and

$$\frac{\partial(X_1,X_2,X_3)}{\partial(m_1,m_2,m_3)} = \begin{bmatrix} \dfrac{X_1(1-X_1)}{m_1} & \dfrac{-X_1X_2}{m_2} & \dfrac{-X_1X_3}{m_3} \\ \dfrac{-X_1X_2}{m_1} & \dfrac{X_2(1-X_2)}{m_2} & \dfrac{-X_2X_3}{m_3} \\ \dfrac{-X_1X_3}{m_1} & \dfrac{-X_2X_3}{m_2} & \dfrac{X_3(1-X_3)}{m_3} \end{bmatrix}. \tag{65}$$

The required derivatives of X_1, X_2, X_3 with respect to w_1, w_2, w_3 follow from the chain rule for Jacobians

$$\frac{\partial(X_1, X_2, X_3)}{\partial(w_1, w_2, w_3)} = \frac{\partial(X_1, X_2, X_3)}{\partial(m_1,m_2,m_3)} \left[\frac{\partial(w_1, w_2, w_3)}{\partial(m_1, m_2, m_3)} \right]^{-1}. \tag{66}$$

The actual values of (m_1, m_2, m_3) appearing in these equations are arbitrary, and without loss of generality can be set equal to the mass percentages; the molar masses are $C_4H_8SO_2 = 120.166320\ \mathrm{g \cdot mol^{-1}}$, $C_8H_{18} = 114.230920\ \mathrm{g \cdot mol^{-1}}$, and $C_6H_6 = 78.113640\ \mathrm{g \cdot mol^{-1}}$. For example, at the point $(X_1, X_2, X_3) = (0.9237, 0.0064, 0.0699)$, the preceding formulas give

$$\frac{\partial(w_1,w_2,w_3)}{\partial(m_1, m_2, m_3)} = \begin{bmatrix} 0.0006 & -0.0077 & -0.0077 \\ -0.0077 & 0.0083 & -0.0077 \\ -0.0004 & -0.0004 & 0.0079 \end{bmatrix}$$

$$= \begin{bmatrix} -126.0409 & -128.3924 & -246.2073 \\ -128.3924 & -4.2897 & -128.3924 \\ -12.1113 & -6.3158 & 108.0550 \end{bmatrix}^{-1}$$

and

$$\frac{\partial(X_1,X_2,X_3)}{\partial(m_1,m_2,m_3)} = \begin{bmatrix} 0.0006 & -0.0081 & -0.0118 \\ -0.0001 & 0.0087 & -0.0001 \\ -0.0006 & -0.0006 & 0.0119 \end{bmatrix},$$

so that

$$\frac{\partial(X_1,X_2,X_3)}{\partial(w_1,w_2,w_3)}=\begin{bmatrix} 1.1014 & 0.0278 & -0.3959 \\ -1.1091 & -0.0300 & -1.1125 \\ 0.0077 & 0.0021 & 1.5084 \end{bmatrix}$$

The increment in (X_1, X_2, X_3) corresponding to $(\delta w_1, \delta w_2, \delta w_3) = (0.003, 0.003, 0.003)$ is therefore

$$\begin{bmatrix} \delta X_1 \\ \delta X_2 \\ \delta X_3 \end{bmatrix}=\begin{bmatrix} 1.1014 & 0.0278 & -0.3959 \\ -1.1091 & -0.0300 & -1.1125 \\ 0.0077 & 0.0021 & 1.5084 \end{bmatrix}\begin{bmatrix} 0.003 \\ 0.003 \\ 0.003 \end{bmatrix}$$

$$=\begin{bmatrix} 0.0022001 \\ 0.0067506 \\ 0.0045545 \end{bmatrix},$$

which shows that the results given by Lee and Kim (1995) do not justify being quoted to four decimal places. For the purposes of least-squares analysis, the weight applied to each component of (X_1, X_2, X_3) can be set equal to the reciprocal square of the corresponding component of the vector $(\delta X_1, \delta X_2, \delta X_3)$, obtained as described earlier at each data point.

According to Novák et al. (1987, p. 291), the generalized Redlich-Kister formula appropriate for the molar excess Gibbs energy in a three-component system is

$$\begin{aligned}\frac{G^E}{(n_1+n_2+n_3)RT} &= X_1X_2X_3[C+C_1X_1+C_2X_2] \\ &\quad + X_1X_2[b_{12}+c_{12}(X_1-X_2)] \\ &\quad + X_1X_3[b_{13}+c_{13}(X_1-X_3)] \\ &\quad + X_2X_3[b_{23}+c_{23}(X_2-X_2)] \qquad (67)\end{aligned}$$

where n_1, n_2, and n_3 are mole numbers and the last term represents the contribution of ternary interactions. The corresponding expressions for the activity coefficients are as follows:

$$\begin{aligned}\ln\gamma_1 &= b_{12}X_2(1-X_1)+b_{13}X_3(1-X_1)-b_{23}X_2X_3 \\ &\quad + c_{12}X_2[X_1+(X_1-X_2)(1-2X_1)] \\ &\quad + c_{13}X_3[X_1+(X_1-X_3)(1-2X_1)] \\ &\quad - 2c_{23}X_2X_3(X_2-X_3)+X_2X_3[C(1-2X_1) \\ &\quad + C_1X_1(2-3X_1)+C_2X_2(1-3X_1)] \qquad (68)\end{aligned}$$

$$\begin{aligned}\ln\gamma_2 &= b_{12}X_1(1-X_2)-b_{13}X_1X_3+b_{23}X_3(1-X_2) \\ &\quad + c_{12}X_1[-X_2+(X_1-X_2)(1-2X_2)] \\ &\quad - 2c_{13}X_1X_3(X_1-X_3) \\ &\quad + c_{23}X_3[X_2+(X_2-X_3)(1-2X_2)] \\ &\quad + X_1X_3[C(1-2X_2)+C_1X_1(1-3X_1) \\ &\quad + C_2X_2(2-3X_2)] \qquad (69)\end{aligned}$$

Table 6. Ternary LLE for the System Sulfolane(1)–Octane(2)–Benzene(3) at 298.15 K (Lee and Kim, 1995): Least-Squares Estimates of Parameters for the Ternary Redlich-Kister Equation

Parameter	Estimate	Standard Error
$z_1 = b_{12}$	5.093131	0.191669
$z_2 = b_{13}$	1.609184	0.215888
$z_3 = b_{23}$	1.455389	0.230983
$z_4 = c_{12}$	−0.152620	0.135304
$z_5 = c_{13}$	−1.345499	0.238131
$z_6 = c_{23}$	−1.557121	0.261735

$$\begin{aligned}\ln\gamma_3 &= -b_{12}X_1X_2+b_{13}X_1(1-X_3)+b_{23}X_2(1-X_3) \\ &\quad - 2c_{12}X_1X_2(X_1-X_2)+c_{13}X_1[-X_3+(X_1-X_3) \\ &\quad \times(1-2X_3)]+c_{23}X_2[-X_3+(X_2-X_3)(1-2X_3)] \\ &\quad + X_1X_2[C(1-2X_3)+C_1X_1(1-3X_3) \\ &\quad + C_2X_2(2-3X_3)]. \qquad (70)\end{aligned}$$

This particular representation of the excess molar Gibbs energy can be fitted to ternary LLE data by three-equation GNLS, in which the ith subvector of observables is $\boldsymbol{x}_i = (X_{1a}^{(i)}, X_{2a}^{(i)}, X_{1b}^{(i)}, X_{2b}^{(i)}, T^{(i)})$ (the subscripts a and b are used to distinguish between the two liquid phases), and the corresponding vector of functional constraints to be satisfied by the components of $\boldsymbol{x}$ is

$$\boldsymbol{g}^{(i)} = \left(\frac{\gamma_{1a}X_{1a}}{\gamma_{1b}X_{1b}}-1,\quad \frac{\gamma_{2a}X_{2a}}{\gamma_{2b}X_{2b}}-1,\quad \frac{\gamma_{3a}X_{3a}}{\gamma_{3b}X_{3b}}-1\right)^T. \qquad (71)$$

Neglecting the effects of ternary interactions in the two phases (that is, setting $C = C_1 = C_2 = 0$) leaves six adjustable parameters for each isotherm. Least-squares estimates of these parameters derived from the 298.15 K measurements by Lee and Kim (1995), with mole fractions weighted as described earlier, are given in Table 6, and the least-squares estimates of the mole fractions in each liquid phase are given in Tables 7 and 8. (In contrast to Table 3, the temperature is constant, since the partial derivatives of the assumed model equations with respect to this variable are all zero.) The standard deviations of unweighted residuals for X_1 and X_2 are

Table 7. Ternary LLE for the System Sulfolane(1)–Octane(2)–Benzene(3) at 298.15 K (Lee and Kim, 1995): Calculated and Experimental Compositions of Phase *a*

X_1 (calc.)	X_1 (expt.)	Residual	X_2 (calc.)	X_2 (expt.)	Residual	T/K
0.9218	0.9237	−0.0019	0.0083	0.0064	0.0019	298.15
0.8113	0.8093	0.0020	0.0099	0.0095	0.0004	298.15
0.7550	0.7541	0.0009	0.0113	0.0114	−0.0001	298.15
0.7274	0.7267	0.0007	0.0117	0.0113	0.0004	298.15
0.6177	0.6167	0.0010	0.0161	0.0174	−0.0013	298.15
0.5671	0.5689	−0.0018	0.0189	0.0201	−0.0012	298.15
0.4873	0.4869	0.0004	0.0250	0.0337	−0.0087	298.15

Table 8. Ternary LLE for the System Sulfolane(1)–Octane(2)–Benzene(3) at 298.15 K (Lee and Kim, 1995): Calculated and Experimental Compositions of Phase *b*

X_1 (calc.)	X_1 (expt.)	Residual	X_2 (calc.)	X_2 (expt.)	Residual	T/K
0.0064	0.0045	0.0019	0.9100	0.9119	−0.0019	298.15
0.0084	0.0104	−0.0020	0.7937	0.7941	−0.0004	298.15
0.0095	0.0104	−0.0009	0.7436	0.7435	0.0001	298.15
0.0110	0.0117	−0.0007	0.7037	0.7041	−0.0004	298.15
0.0145	0.0155	−0.0010	0.6106	0.6093	0.0013	298.15
0.0165	0.0147	0.0018	0.5714	0.5702	0.0012	298.15
0.0199	0.0203	−0.0004	0.5207	0.5120	0.0087	298.15

0.0011 and 0.0028, respectively, which are comparable to the estimated uncertainties in the mole fractions.

Discussion

The calculations presented in the preceding sections demonstrate the successful extension of the BLJ approach to GNLS parameter estimation in multiequation implicit models. The applications to the reduction of binary VLE data also suggest that the long-recognized problem of weighting the activity coefficients in the fit of excess Gibbs energy models can be resolved by use of an implicit formulation of the equations of condition. Moreover, the change of variables that is implied in the use of the activity coefficients themselves as responses is thereby avoided, as is the need for replicate measurements to estimate the elements of the variance–covariance matrix. Although in the particular example considered, the GNLS results agreed well with the WNLS fit of activity coefficients incorporating a covariance matrix estimated from the Gaussian error-propagation formula, it is by no means obvious that this will always be true.

The approach described in the present article invites comparison with the dual-response GNLS method described by Anderson et al. (1978) (AAG) and applied by them to VLE data reduction; some discussion of the relation between the two treatments is appropriate. Both are based on the assumption that the regressor variables can be described by a multivariate normal distribution, and both formulate the parameter estimation problem as the constrained maximization of the corresponding log-likelihood function in the form of Eq. 1. The derivations diverge in the formulation of the functional constraints: as a single vector, or as two separate vectors (each with its own set of Lagrange multipliers). Both involve the same linearizations about a trial solution, but the solution of the linearized equations results in far more complex vector algebra when the constraint vectors are considered separately, as in AAG. A further, perhaps less significant point is that the solution method proposed by these authors requires the additional assumption of a diagonal weighting matrix—a condition that is nearly always satisfied in practice—while the generalized BLJ equations remain valid even if the covariance matrices are not diagonal. But the most significant respect in which the BLJ and AAG methods differ is that, with appropriate redefinition of the various arrays, the former is applicable to problems involving more than two constraints at each point, while the approach used by AAG would become unmanageably complex if a third vector of functional constraints were to be incorporated.

It also seems appropriate to discuss briefly the approximation of partial derivatives with respect to parameters and variables, especially considering that this point was not discussed by AAG. A detailed analytical investigation of the effects of numerical differentiation errors on least-squares calculations would be beyond the scope of this article, but experience suggests that the central difference approximation, viz.

$$\left(\frac{\partial f}{\partial x}\right)_{x=x_0} \approx \frac{f(x_0+h)-f(x_0-h)}{2h} \qquad (72)$$

gives sufficient accuracy. Following the suggestions of Press et al. (1992, pp. 182–184) and Bates and Watts (1988, p. 82), the quantity h is in turn obtained from x_0 by

$$h = \begin{cases} x_0\sqrt[3]{\epsilon}: & x_0 \neq 0 \\ \sqrt[3]{\epsilon}: & x_0 = 0, \end{cases} \qquad (73)$$

where ϵ is the relative machine error (the smallest number for which $1+\epsilon$ is detectably different from 1). In Fortran, the scale factor can be determined by a single call to a subroutine of the sort described by Press et al. (1992, pp. 881–886) and Forsythe et al. (1977, pp. 13–14), and supplied to the numerical differentiation subroutines through a COMMON block. For the Macintosh iBook SE on which the program was developed, the cube root of ϵ is found to be $4.806217383937358D-06$ in double precision (use of which is to be strongly recommended in any calculation involving numerical differentiation). The accuracy of the difference approximation can be seriously vitiated if ϵ is not consistent with the word length of the processor; it is a good idea to determine it every time the program is run, especially if it is desired to port the code to different machines.

In each of the models considered in the present work, the required number of equations of condition was the same throughout the entire data set. While this is true for many models encountered in chemical engineering practice, it is sometimes of interest to estimate parameters by simultaneously fitting multiple properties described by different numbers of equations. This situation frequently arises in the multiproperty fit of equations of state to some assumed representation of the Helmholtz energy. For example, data sets consisting of different combinations of vapor pressures, saturated liquid densities, and saturated vapor densities could be associated with up to three equations of condition per point, while the pressure-density isotherms and speed-of-sound measurements would be associated with only one. The different precision achievable in each of these types of measurements clearly indicates the need for GNLS. But the diagonal blocks of the effective weighting matrix for the entire data set would be of different sizes. The matrix manipulations involved in solving the normal equations for the parameter increments would therefore require some modifications, although the basic iterative procedure would remain the same. The application of GNLS to the multiproperty fit of equations of state will be treated in a future article.

Conclusions

In this article, an extension of the generalized Gauss-Newton method first described by Britt and Luecke (1973) and subsequently by Jefferys (1980) and Lybanon (1984a, 1985) has been described. This allows multiresponse data to be simultaneously fitted to models containing several equations of condition per point, and is considerably simpler than the dual-response GNLS algorithm developed by Anderson et al. (1978). The block-diagonal weighting matrix arising in the solution of the normal equations for the parameter increments is formed in exactly the same way as in the earlier work of Kemény et al. (1982) based on the effective-variance approximation, but the algorithm differs from that work in involving the simultaneous estimate of the regressors as well as the parameters. The algorithm can be applied without modification to simpler problems, such as those with fixed regressors or involving only one equation of condition per point. Central-difference numerical approximations to the required partial derivatives, with respect to both the parameters and the variables, were found to be sufficiently accurate, at least for the simple analytical functions considered in the calculations.

The performance of this method in the direct reduction of VLE data was compared with that of various multiresponse nonlinear regression methods in the correlation of activity coefficients derived therefrom. These included minimization of the weighted sum of squares (Aitken, 1935) with weights estimated from the error propagation formula, the diagonal uniform weighting model, and the determinant criterion appropriate to problems where the covariances between the responses are unknown. This last objective function has been shown to be very efficiently minimized by application of the conjugate gradient algorithm. On the one hand, the discrepancies between the parameter estimates yielded by these calculations is a clear illustration of the importance of weighting in such fits. On the other hand, the lack of agreement with the GNLS calculation suggests a fundamental inequivalence between the ML formulations underlying each. One way in which this discrepancy can be removed is to replace the explicit dependence of the activity coefficients on composition with implicit functional constraints involving the composition, total pressure, and temperature at each data point. Another way is to reformulate the ML theory in terms of the exact joint distribution functions for the mole-fraction ratios. A new analytical theory that achieves this will be described in a subsequent article.

Acknowledgments

It is a pleasure to express my sincere thanks to Dr. J. G. Blencoe, without whose support and continuing interest this work could not have been performed. I am also grateful to Dr. M. R. Ally and an anonymous referee for helpful comments and suggestions. Funding for the research was provided by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Services, U.S. Department of Energy under contract DE-AC05-00OR22725. Oak Ridge National Laboratory is managed and operated by UT-Battelle, LLC.

Literature Cited

Aitken, A. C., "On Least Squares and Combinations of Observations," *Proc. Roy. Soc. Edinburgh*, **55**, 42 (1935).

Anderson, T. W., *An Introduction to Multivariate Statistical Analysis*, Wiley, New York (1958).

Anderson, T. F., D. S. Abrams, and E. A. Grens II, "Evaluation of Parameters for Nonlinear Thermodynamic Models," *AIChE J.*, **24**, 20 (1978).

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

Bard, Y., *Nonlinear Parameter Estimation*, Academic Press, New York (1974).

Bard, Y., and L. Lapidus, "Kinetic Analysis by Digital Parameter Estimation," *Catal. Rev.*, **2**, 67 (1968).

Bates, D. M., and D. G. Watts, "A Generalized Gauss-Newton Procedure for Multi-Response Parameter Estimation," *SIAM J. Sci. Stat. Comput.*, **8**, 49 (1987).

Bates, D. M., and D. G. Watts, *Nonlinear Regression Analysis and Its Applications*, Wiley, New York (1988).

Beauchamp, J. J., and R. G. Cornell, "Simultaneous Nonlinear Estimation," *Technometrics*, **8**, 319 (1966).

Box, M. J., "Improved Parameter Estimation," *Technometrics*, **12**, 219 (1970).

Box, G. E. P., and N. R. Draper, "The Bayesian Estimation of Common Parameters from Several Responses," *Biometrika*, **52**, 355 (1965).

Brent, R. P., *Algorithms for Minimization Without Derivatives*, Prentice Hall, Englewood Cliffs, NJ (1973).

Britt, H. I., and R. H. Luecke, "The Estimation of Parameters in Nonlinear, Implicit Models," *Technometrics*, **15**, 233 (1973).

Clutton-Brock, M., "Likelihood Distributions for Estimating Functions when both Variables Are Subject to Error," *Technometrics*, **9**, 9261, 261 (1967).

Cramér, H., *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton, NJ (1961).

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

Dennis, J. E., Jr., D. M. Gay, and R. E. Welsch, "An Adaptive Nonlinear Least-Squares Algorithm," *ACM Trans. Math. Softw.*, **7**(3), 348 (1981).

Dohnal, V., and D. Fenclová, "A New Procedure for Consistency Testing of Binary Vapor-Liquid Equilibrium Data," *Fluid Phase Equilib.*, **21**, 211 (1985).

Englezos, P., N. Kalogerakis, and P. R. Bishnoi, "Simultaneous Regression of Binary VLE and VLLE Data," *Fluid Phase Equilib.*, **61** 1 (1991).

Esposito, W. R., and C. A. Floudas, "Global Optimization in Parameter Estimation of Nonlinear Algebraic Models via the Error-in-Variables Approach," *Ind. Eng. Chem. Res.*, **37**, 1841 (1998).

Esposito, W. R., and C. A. Floudas, "Global Optimization for the Parameter Estimation of Differential-Algebraic Systems," *Ind. Eng. Chem. Res.*, **39**, 1291 (2000).

Fisher, R. A., "On the Mathematical Foundations of Theoretical Statistics," *Philos. Trans. R. Soc. Lond. A*, **222**, 309 (1922).

Fletcher, R., and C. M. Reeves, "Function Minimization by Conjugate Gradients," *Comput. J.*, **7**, 149 (1964).

Forsgren, A., P. E. Gill, and M. H. Wright, "Interior Methods for Nonlinear Optimization," *SIAM Rev.*, **44**, 525 (2002).

Forsythe, G. E., M. A. Malcolm, and C. B. Moler, *Computer Methods for Mathematical Calculations*, Prentice Hall, Englewood Cliffs, NJ (1977).

Gallant, A. R., "Seemingly Unrelated Nonlinear Regressions," *J. Econometrics*, **3**, 35 (1975).

Gmehling, J., U. Onken, W. Arlt, and U. Weidlich, *Vapor-Liquid Equilibrium Data Collection*, Vol. 1, Pt. 1, DECHEMA, Frankfurt am Main, Germany (1977).

Gmehling, J., U. Onken, W. Arlt, and U. Weidlich, *Vapor-Liquid Equilibrium Data Collection*, Vol. 1, Pt. 4, DECHEMA, Frankfurt am Main, Germany (1979).

Han, S.-P., "Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems," *Math. Program.*, **11**, 263 (1976).

Hermsen, R. W., and J. M. Prausnitz, "Thermodynamic Properties of the Benzene and Cyclopentane System," *Chem. Eng. Sci.*, **18**, 485 (1963).

Hofman, T., and M. Krzyżanowska, "Determination of Stability Constants of Complexes from the Titration Curve by the Maximum-Likelihood Principle," *Talanta*, **33**, 851 (1986).

Hooke, R., and T. A. Jeeves, "'Direct Search' Solution of Numerical and Statistical Problems," *J. Assoc. Comput. Mach.*, **8**, 212 (1961).

Hunter, W. G., "Estimation of Unknown Constants from Multiresponse Data," *Ind. Eng. Chem. Fundam.*, **6**, 461 (1967).

Jefferys, W. H., "On the Method of Least Squares," *Astrophys. J.*, **85**, 177 (1980).

Kemény, S., J. Manczinger, S. Skjold-Jørgensen, and M. Tóth, "Reduction of Thermodynamic Data by Means of the Multiresponse Maximum Likelihood Principle," *AIChE J.*, **28**, 20 (1982).

Kim, I.-W., M. J. Liebman, and T. F. Edgar, "Robust Error-in-Variables Estimation Using Nonlinear Programming Techniques," *AIChE J.*, **36**(7), 985 (1990).

Kim, I.-W., T. F. Edgar, and N. H. Bell, "Parameter Estimation for a Laboratory Water-Gas-Shift Reactor Using a Nonlinear Error-in-Variables Method," *Comput. Chem. Eng.*, **15**, 361 (1991).

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

Lee, S., and H. Kim, "Liquid-Liquid Equilibria for the Ternary Systems Sulfolane + Octane + Benzene, Sulfolane + Octane + Toluene and Sulfolane + Octane + *p*-Xylene," *J. Chem. Eng. Data*, **40**, 499 (1995).

Lee, S., and H. Kim, "Liquid-Liquid Equilibria for the Ternary Systems Sulfolane + Octane + Benzene, Sulfolane + Octane + Toluene, and Sulfolane + Octane + *p*-Xylene at Elevated Temperatures," *J. Chem. Eng. Data*, **43**, 358 (1998).

Levenberg, K., "A Method for the Solution of Certain Nonlinear Problems in Least Squares," *Q. Appl. Math.*, **2**, 164 (1944).

Lisý, J. M., and P. Šimon, "Evaluation of Parameters in Nonlinear Models by the Least Squares Method," *Comput. Chem.*, **22**, 509 (1998).

Lisý, J. M., P. Šimon, S. Bafrncová, and E. Graczová, "Least Squares Estimation of Parameters in Implicit Models," *Comput. Chem.*, **23**, 587 (1999).

Lybanon, M., "A Better Least-Squares Method When Both Variables have Uncertainties," *Amer. J. Phys.*, **52**, 22 (1984a).

Lybanon, M., "Comment on 'Least Squares when Both Variables have Uncertainties'," *Amer. J. Phys.*, **52**, 276 (1984b).

Lybanon, M., "A Simple Generalized Least-Squares Algorithm," *Comput. Geosci.*, **11**, 501 (1985).

Macdonald, J. R., "Comparison and Application of Two Methods for the Least Squares Analysis of Immittance Data," *Solid State Ionics*, **58**, 97 (1992).

Macdonald, J. R., "Exact and Approximate Nonlinear Least-Squares Inversion of Dielectric Relaxation Spectra," *J. Chem. Phys.*, **102**, 6241 (1995).

Macdonald, J. R., and J. A. Garber, "Analysis of Impedance and Admittance Data for Solids and Liquids," *J. Electrochem. Soc.*, **124**, 1022 (1977).

Macdonald, J. R., J. Schoonman, and A. P. Lehnen, "The Applicability and Power of Complex Nonlinear Least Squares for the Analysis of Impedance and Admittance Data," *J. Electroanal. Chem. Interfacial Electrochem.*, **131**, 77 (1982).

Macdonald, J. R., and W. J. Thompson, "Least-Squares Fitting When Both Variables Contain Errors: Pitfalls and Possibilities," *Amer. J. Phys.*, **60**, 66 (1992).

Marquardt, D. W., "An Algorithm for the Estimation of Nonlinear Parameters," *J. SIAM*, **11**, 431 (1963).

McGlashan, M. L., and R. P. Rastogi, "The Thermodynamics of Associated Mixtures. Part I—Dioxan + Chloroform," *Trans. Faraday Soc.*, **54**, 496 (1958).

Nelder, J. A., and R. Mead, "A Simplex Method for Function Minimization," *Comput. J.*, **7**, 308 (1965).

Novák, J. P., J. Matouš, and J. Pick, *Liquid-Liquid Equilibria*, Elsevier, Amsterdam, The Netherlands (1987).

Pearson, K., "On Lines and Planes of Closest Fit to Points in Space," *Philos. Mag.*, **2**, 559 (1901).

Pitzer, K. S., and R. F. Curl, Jr., "The Volumetric and Thermodynamic Properties of Fluids. III. Empirical Equation for the Second Virial Coefficient," *J. Amer. Chem. Soc.*, **79**, 2369 (1957).

Powell, M. J. D., "An Efficient Method for Finding the Minimum of a Function Without Calculating Derivatives," *Comput. J.*, **7**, 155 (1964).

Powell, M. J. D., "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," *Numerical Analysis*, Vol. 630, *Lecture Notes in Mathematics*, G. A. Watson, ed., Springer-Verlag, Berlin, p. 144 (1978).

Powell, D. R., and J. R. Macdonald, "A Rapidly Convergent Iterative Method for the Solution of the Generalized Nonlinear Least Squares Problem," *Comput. J.*, **15**, 148 (1972).

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

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, 2nd ed., Cambridge Univ. Press, Cambridge (1992).

Reid, R. C., J. M. Prausnitz, and T. K. Sherwood, *The Properties of Gases and Liquids*, McGraw-Hill, New York (1977).

Rubio, R. G., J. A. R. Renuncio, and M. Diaz-Peña, "Regression of Vapor-Liquid Equilibrium Data Based on Application of the Maximum-Likelihood Principle," *Fluid Phase Equilib.*, **12**, 217 (1983).

Sachs, W. H., "Implicit Multifunctional Nonlinear Regression Analysis," *Technometrics*, **18**, 161 (1976).

Schubert, E., "Vergleich verschiedener Methoden zur Ermittlung der Parameter in der Wilson- und NRTL-Gleichung," *Chem.-Ing. Tech.*, **46**, 73 (1974).

Schwetlick, H., and V. Tiller, "Numerical Methods for Estimating Parameters in Nonlinear Models with Errors in the Variables," *Technometric*, **27**, 17 (1985).

Seber, G. A. F., and C. J. Wild, *Nonlinear Regression*, Wiley, New York (1989).

Skjold-Jørgensen, S., "On Statistical Principles in Reduction of Thermodynamic Data," *Fluid Phase Equilib.*, **14**, 273 (1983).

Smith, J. M., and H. C. Van Ness, *Introduction to Chemical Engineering Thermodynamics*, 4th ed., McGraw-Hill, New York (1987).

Stewart, W. E., M. Caracotsios, and J. P. Sørensen, "Parameter Estimation from Multiresponse Data," *AIChE J.*, **38**(5), 641 (1992).

Ulrichson, D. L., and F. D. Stevenson, "Effects of Experimental Errors on Thermodynamic Consistency and on Representation of Vapor-Liquid Equilibrium Data," *Ind. Eng. Chem. Fundam.*, **11**, 287 (1972).

Valkó, P., and S. Vajda, "An Extended Marquardt-Type Procedure for Fitting Error-in-Variables Models," *Comput. Chem. Eng.*, **11**(1), 37 (1987).

Van Ness, H. C., S. M. Byers, and R. E. Gibbs, "Vapor-Liquid Equilibrium: Part I. An Appraisal of Data Reduction Methods," *AIChE J.*, **19**, 238 (1973).

Wismer, D. A., and R. Chattergy, *Introduction to Nonlinear Optimization: A Problem Solving Approach*, North-Holland, New York (1978).

Yen, L. C., and S. S. Woods, "A Generalized Equation for Computer Calculation of Liquid Densities," *AIChE J.*, **12**, 95 (1966).

York, D., "Least-Squares Fitting of a Straight Line," *Can. J. Phys.*, **44**, 1079 (1966).

Zellner, A., "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *J. Amer. Stat. Assoc.*, **57**, 348 (1962).

Manuscript received Sept. 13, 2002, and revision received Mar. 10, 2003.