Maximum Likelihood Data Rectification: Steady-State Systems

Lloyd P. M. Johnston and Mark A. Kramer
Dept. of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

A maximum likelihood rectification (MLR) technique that poses the data-rectification problem in a probabilistic framework and maximizes the probability of the estimated plant states given the measurements is proposed. This approach does not divide the sensors into "normal" and "gross error" classes, but uses all of the data in the rectification, each sensor being appropriately weighted according to the laws of probability. In this manner, the conventional assumption of no sensor bias is avoided, and both random errors (noise) and systematic errors (gross errors) are removed simultaneously. A novel technique is introduced that utilizes historical plant data to determine a prior probability distribution of the plant states. This type of historical plant information, which contains the physical relationships among the variables (mass balances, energy balances, thermodynamic constraints), as well as statistical correlations among the variables, has been ignored in prior data-rectification schemes. This approach can use the historical plant information to solve a new class of data-rectification problems in which there are no known model constraints. The MLR method is demonstrated on data from a simulated flow network and a simulated heat-exchanger network. The MLR technique provides considerably improved performance over existing data-reconciliation schemes in these examples.

Introduction

All chemical processing plants make measurements of a wide variety of process variables, both on- and off-line. These measurements are used to make decisions that affect product quality, plant safety, and plant profitability. Inaccurate process data can lead to poor decisions that can adversely affect many plant functions. Measurement instruments are subject to random errors (noise), and nonrandom errors, such as sensor bias or sensor failure (gross errors). The objective of data rectification is to obtain an estimate of the true state of the plant by removing both the random and nonrandom errors from the data set. The problem of data rectification is posed in this work as a maximum likelihood problem, where the probability of the estimated plant state is maximized given the measurement set.

Although the terms are often used indiscriminately, data rectification and data reconciliation are distinct. Data rectification means literally "to make data right," while data reconciliation refers to adjusting data to conform to prior constraints (Kramer and Mah, 1993). Since making the data conform to constraints will not in itself necessarily correct noisy or biased data, one can view data reconciliation as a special case, or a subset, of the broader problem of data rectification. In this article, we attempt to sharpen the distinction between rectification and reconciliation using the framework of maximum likelihood estimation, and show that, under certain assumptions of the prior knowledge, the rectification problem reduces to the classic reconciliation problem.

The simplest case of data reconciliation is that of weighted least squares applied to steady-state linear systems (Kuehn and Davidson, 1961). The system of interest is modeled by linear steady-state equations (most often mass balances), and the measurements are adjusted to satisfy the model equations. The sum of the squares of the adjustments, each term weighted by the inverse of its measurement noise variance, is minimized. The underlying assumption leading to the weighted least-squares formulation is that the measurement errors are normally distributed (Gaussian), with zero mean. Mathematically the problem is posed as

$$\min_{\hat{x}} (y - \hat{x})^T Q^{-1} (y - \hat{x})$$
$$\text{s.t.} \quad A\hat{x} = 0 \tag{1}$$

where y is the measurement vector, $\hat{x}$ is the rectified plant state vector, Q is the noise covariance matrix, and A is the matrix of linear constraints. This problem has an analytical solution, which makes calculation of the reconciled state quick and suitable for plantwide, on-line application. A projection matrix can be used to recover Eq. 1 when some of the variables in the constraint equations are not measured (Crowe et al., 1983). The solution of Eq. 1 yields unbiased estimates of the plant states when the assumption of normally distributed noise is valid (Mah, 1987).

Sensors can often have errors that do not satisfy the assumptions of zero-mean, Gaussian noise. To treat nonzero mean errors, most past researchers have partitioned measurements into two sets. One set is sensors considered to be operating normally (containing only zero-mean errors with given variance), and the other set contains sensors with "gross errors" (nonzero mean errors). Sensors classified as having gross errors must be excluded from the reconciliation calculation, since including them contradicts the assumptions of the method and leads to inaccurate results (Ripps, 1965).

Great effort has been put into developing auxiliary techniques for identifying gross errors in process measurements. Ripps laid the groundwork for the serial elimination procedure developed by Nogita (1972), which eliminates measurements on a trial basis one at a time, calculating a new test statistic based on the reduced measurement set. The set of measurements that pass the gross error test (comparison of the test statistic with the appropriate standard value, such as a normal variate score) is said not to contain gross errors. The procedure is most effective for a single gross error in the measurement set; as with multiple gross errors the possible combinations of measurements to delete quickly becomes very large. A number of early gross error detection schemes are reviewed by Hlavacek (1977). Other gross error detection tests are reviewed by Mah (1987). The global and nodal tests use statistics based on the constraint residuals $r = Ay$ (Tamhane and Mah, 1985). The measurement test (Mah and Tamhane, 1982; Tamhane and Mah, 1985) is based on classic Neyman–Pearson hypothesis testing, and looks at the size of the adjustments made in the reconciliation $(y - \hat{x})$ to decide if a gross error exists. Iordache et al. (1985) investigated the performance of the measurement test, and found that the performance depended both on the absolute size of the gross error, and on the ratio of the size of the gross error to the size of the noise standard deviation for the measurement. The optimal method of gross error detection is to use combined tests (Rosenberg et al., 1987).

All of the previously mentioned gross error detection schemes, in addition to many other gross error detection methods, are for linear models. This is a restrictive condition for almost any process model other than mass balances. Linearization is the most common approach for dealing with nonlinear constraints. However, by forcing the measurements to conform to the linearized (approximate) model, errors may be introduced. Rollins and Roelfs (1992) approached the nonlinear reconciliation problem by extending the unbiased estimation technique of Rollins and Davis (1992) to bilinear constraints, specifically heat balances.

Work has also been done on solving nonlinear reconciliation problems using nonlinear programming. The formulation of the reconciliation problem as an optimization is similar to the errors-in-variables (EIV) estimation technique. EIV is a method used to regress model parameters, applied when there are measurement errors in the independent variables. Data reconciliation is a special case of EIV estimation where no parameters are to be estimated but the optimization is subject to constraints. Britt and Luecke (1973) applied the EIV method for simultaneous data reconciliation and parameter estimation for nonlinear systems. Subsequently, much attention has been focused on the solution of the EIV optimization problem (Peneloex et al., 1976; Bard, 1974; Tjoa and Biegler, 1992). The performance of the solution techniques has been evaluated by Ricker (1984).

All of the previously discussed data-reconciliation methods for nonlinear systems assume that the measurements only contain zero-centered Gaussian noise. Thus, they must be coupled with gross error identification techniques. Tjoa and Biegler (1991) have attempted to remove this requirement by formulating a consolidated approach to data reconciliation and gross error detection. Instead of minimizing the measurement adjustments in a weighted least-squares approach, a probability distribution of the adjustments is formed, and it is maximized subject to the constraints. The probability distribution of the adjustment contains two parts, a "narrow" Gaussian distribution centered at zero, representing normal measurement noise, and a "broad" Gaussian distribution, also centered at zero, that makes large adjustments somewhat more probable. The two Gaussian distributions are weighted by the probability of not having a gross error in a sensor, and the probability of a gross error in a sensor, respectively. There are no restrictions on the constraints the objective function is optimized subject to, and the method formulation applies to both linear and nonlinear systems. Our approach is somewhat similar to Tjoa and Biegler's; however, we begin from more fundamental considerations (maximum likelihood estimation and probability theory), and the resulting problem formulation includes different prior information and probability distributions. Similar to Tjoa and Biegler, we derive a problem formulation that does not require a separate gross error detection step.

Approach

The proposed approach poses the data-rectification problem in a probabilistic framework and uses the laws of probability and maximum likelihood estimation to determine the most likely rectified states, following the approach originally outlined in Kramer and Mah (1993). The probability distribution of the process states (x), given the process measurements (y) is maximized to find the rectified states. The rectification problem is posed as

$$\max_x P\{x|y\}. \tag{2}$$

By Bayes' theorem, the probability of the states given the measurements can be written in terms of the probability of the measurements given the states, the prior probability density of the states, and the probability density of the measurements:

$$\max_x P\{x|y\} = \max_x P\{y|x\}P\{x\}/P\{y\}. \tag{3}$$

The denominator term $P\{y\}$, independent of x, acts as a normalizing constant (for any given set of measurements) and does not have to be considered further.

The first term of the numerator of Eq. 3 represents the probability density of the measurements given the states, which is the distribution of the measurement error. Instead of assuming this is given by a normal distribution, we obtain this probability distribution by representing each sensor and its operating modes explicitly. Given an exhaustive set of mutually exclusive modes of sensor operation m_j, a measurement model for sensor i can be expressed by the weighted sum over its modes:

$$P\{y_i|x_i\} = \sum_j P\{y_i|x_i,m_j\}P\{m_j\}. \tag{4}$$

Modes that could be considered include, but are not limited to, normal, biased, and failed. Summing the probability distributions of several modes broadens the basic noise distribution and increases the probability of large errors, relative to the normal mode alone. If sensor errors are independent, the product of this probability over all sensors yields the first term of the rectification objective function:

$$P\{y|x\} = \prod_i P\{y_i|x_i\}. \tag{5}$$

The other term in Eq. 3, $P\{x\}$, is a probability density function (PDF) representing the prior joint distribution of the states. This distribution gives the relative merit of each possible solution to the rectification problem in the absence of measurement data. All previously developed reconciliation schemes carry a critical but implicit assumption about $P\{x\}$, specifically that $P\{x\}$ is a binary variable: that $P\{x\} = 1$ if and only if x satisfies the model equations (equality and inequality constraints), and otherwise $P\{x\} = 0$ (if x fails to satisfy the model constraints). Under this assumption, the $P\{x\}$ term is converted to a set of constraints, and Eq. 3 is converted to the following constrained optimization:

$$\begin{aligned} &\max_x P\{y|x\} \\ \text{s.t.}\quad &h(x) = 0 \\ &g(x) \le 0. \end{aligned} \tag{6}$$

Tjoa and Biegler (1991) modeled $P\{y|x\}$ as a bivariate Gaussian and solved Eq. 6 using sequential quadratic programming. However, all states that satisfy the constraints $g(x) \le 0$, $h(x) = 0$ are not necessarily equally likely, because there may be preferred operating regions within the reduced space described by the constraints. Approximating $P\{x\}$ by a binary variable may introduce unnecessary error into the rectification calculation.

If, in addition to the binary assumption on $P\{x\}$, sensor errors are assumed to follow a normal distribution, taking the logarithm of the objective function in Eq. 6 results in the conventional weighted least-squares formulation of the steady-state rectification problem:

$$\begin{aligned} &\min_x (y-x)^T Q^{-1}(y-x) \\ \text{s.t.}\quad &h(x) = 0 \\ &g(x) \le 0. \end{aligned} \tag{7}$$

For linear equality constraints $Ax = 0$ and in the absence of inequality constraints, this problem has a well-known, closed-form solution (Mah, 1987):

$$\hat{x} = [I - QA^T(AQA^T)^{-1}A]y. \tag{8}$$

Therefore it can be seen that the two key assumptions needed to derive the typical least-squares formulation from the maximum likelihood problem are (1) the distribution of sensor errors are zero-mean and Gaussian; and (2) the prior probability distribution of the states can be reduced to a set of constraints. To overcome limitations of the first assumption, we model the probability distribution of measurement errors by Eq. 4, which considers multiple modes of sensor operation. This results in a model of $P\{y|x\}$ that is typically not normal, even when the random noise contribution is normal. In prior work, the effect of the second assumption has not been clearly explored. To relax the assumption about $P\{x\}$, we model the probability distribution $P\{x\}$ using a combination of constraints and a probability distribution of the states in the subspace of the constraints.

To illustrate the affect of assuming $P\{x\}$ is a binary variable, consider a mixing tee where flow 1 and flow 2 are mixed together, forming flow 3. The mass balance constraint for this problem is $F_1 + F_2 = F_3$. This defines a plane in the measurement space, as shown in Figure 1. The probability of points off the plane is zero. However, not all of the points on the plane are equally likely, and some may be infeasible (negative flows if using positive displacement pumps). If a history of the actual flows were plotted in the measurement space, there would be regions on the plane in which the actual flows would tend to lie, represented by the shaded regions in Figure 1. The separate shaded regions correspond to different modes of plant production. The different likelihood of states in the space of constraints (in this case the plane) is

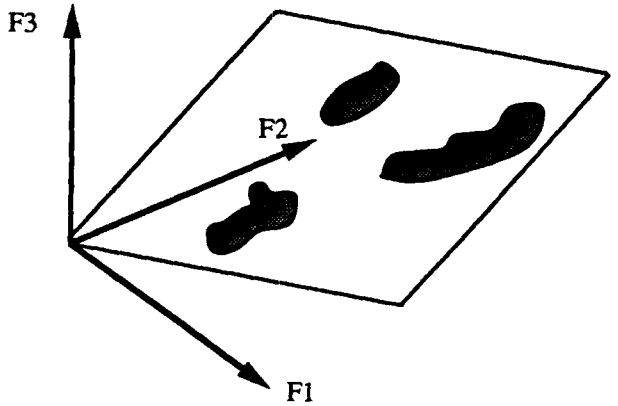

Figure 1. Rectification subspace for a mixing tee where $F1 + F2 = F3$.

Historical data from the process falls in the shaded regions, implying that not all points on the rectification surface are equally likely.

additional information that can be exploited to obtain a better estimate of the true plant state.

Taking into account the probability function of the sensor error modes, given by Eqs. 4 and 5, and the model of $P\{x\}$, the maximum likelihood rectification (MLR) method solves the following problem to obtain the *most likely* rectified states, given the measurements:

$$\max_x P\{x\}P\{y|x\}$$
$$\text{s.t.} \quad h(x) = 0$$
$$g(x) \le 0. \tag{9}$$

In the case where there is no prior knowledge of $P\{x\}$, Eq. 9 reduces to Eq. 6, where only the probability of the adjustments to the measurements $P\{y|x\}$ is maximized subject to the constraints. In this article, we explore the solution of the more general problem in Eq. 9. Solution of Eq. 9 results directly in the plant states, without the requirement for a separate gross error detection step.

One problem immediately encountered in solving Eq. 9 is the modeling of the prior distribution, $P\{x\}$. We assume that this distribution is not known in advance, but must be calibrated from historical data. It cannot be derived directly from measurements because the measurements contain errors, which corrupt the distribution of $P\{x\}$. Our technique uses bootstrapping to find the distribution of $P\{x\}$ indirectly from a calibration measurement set, as described in a subsequent section. Furthermore, the formulation takes into account the statistical tendencies of the process that are represented by the term $P\{x\}$. This information, ignored in other rectification methods, can be exploited to obtain improved estimates of the plant state, as the following simple example shows.

Example 1. Two identical thermocouples measure the temperature of a stream. The sensors normally agree within a degree or two, but the current readings are 20°C and 50°C, respectively. Traditional gross error detection techniques, given the constraint $T_1 = T_2$, can detect that a sensor failure has occurred, but cannot determine which sensor is at fault. Therefore, data cannot be rectified with traditional techniques. However, historical data from the plant has shown that the temperature of the stream follows a Gaussian distribution with a mean of 50°C and a standard deviation of 10°C. Knowing this, an operator would rationally conclude that it is most likely that the 20°C measurement is erroneous, and the actual temperature of the stream is 50°C. Assume that both sensors have Gaussian noise with a standard deviation of 1°C when working normally, and the probability of either sensor failing is 2%, in which case the sensor reports a random value in the range of 0°C to 100°C (modeled as a uniform distribution between 0 and 100). Then the problem can be formally modeled as (with y_i representing the ith temperature measurement and x_i representing the true temperature state):

$$P\{y_i|x_i, \text{normal}_i\} = N(x_i,1) \text{ evaluated at } y_i$$
$$\equiv P_N(y_i - x_i, 1)$$
$$P\{y_i|x_i, \text{failed}_i\} = 0 \text{ if } y_i < 0 \text{ or } y_i > 100$$
$$0.01 \text{ if } 0 \le y_i \le 100$$
$$P\{\text{normal}_i\} = 0.98, P\{\text{failed}_i\} = 0.02.$$

Thus,

$$P\{y_i|x_i\} = 0.98P_N(y_i - x_i, 1) + 0.02*0.01$$

and

$$P\{x_i\} = N(50,10) \text{ evaluated at } x_i$$
$$\equiv P_N(50 - x_i, 10).$$

The MLR formulation is:

$$\max_{x1,x2} P_N(50 - x, 10)*[(0.98*P_N(50 - x_1, 1)$$
$$+0.0002)*(0.98*P_N(20 - x_2, 1) + 0.0002)]$$
$$\text{s.t.} \quad x_1 = x_2 \equiv x,$$

same as

$$\max_x 0.9604*P_N(50 - x, 10)*P_N(50 - x, 1)*P_N(20 - x, 1)$$
$$+1.96e-4*P_N(50 - x, 1)*P_N(50 - x, 10)$$
$$+1.96e-4*P_N(20 - x, 1)*P_N(50 - x, 10).$$

For $x = 20$°C, all three terms in the preceding equation are very small, the largest being the third term, which is $4.2e-92$. For $x = 50$°C, the first and last terms are vanishingly small ($< 1e-197$), and the second term is relatively large, $3.1e-6$. Thus the solution to the maximization will be very close to $x_1 = x_2 = x = 50$°C.

From a conceptual point of view, the data rectification in Example 1 can be explained using Occam's razor. Three scenarios are possible: (1) the true temperature is near 50°C, sensor 1 has not failed and sensor 2 has failed with approximately a −30°C bias; (2) the true temperature is near 20°C, sensor 2 has not failed and sensor 1 has failed with approximately a +30°C bias; and (3) the true temperature is neither near 50°C nor 20°C, and both sensors have failed with a large bias (compared to the measurement noise standard deviation of 1°C). Scenario one has two likely events, that the true temperature is 50°C and that sensor 1 has not failed, and one unlikely event, that sensor 2 has failed. Scenario 2 has two unlikely events, that the true temperature is 20°C and that sensor 1 has failed, and one likely event, that sensor 2 has not failed. Scenario three has at least two unlikely events, that sensor 1 and sensor 2 have both failed, and another event that is less likely than the true temperature being 50°C. According to Occam's razor, and to the laws of probability, scenario one is the more likely than scenario two or three because it requires only one unlikely event to explain the observations, whereas the other two scenarios require at least two unlikely events.

This simple example shows how information on the operation of the plant can be exploited to improve the data-rectification process, and in this case solve a problem not solvable by traditional methods.

It is important that the distribution of $P\{x\}$ include a full range of plant behaviors. For plants that operate at a variety of steady states, each of the steady states needs to be represented in the distribution of $P\{x\}$, appropriately weighted by

the prior probability of each plant state. Assuming the plant has a number of mutually exclusive and exhaustive operating modes, the distribution $P\{\boldsymbol{x}\}$ is given by

$$P\{\boldsymbol{x}\} = \sum_j P\{\boldsymbol{x}|o_j\}P\{o_j\}, \tag{10}$$

where the o_j are the different plant operating modes. Such modes could be operation at different rates of production, operation for different product purities, or processing of different feed compositions as in a refinery. If $P\{\boldsymbol{x}\}$ is calculated from historical plant data, then this data set should be a representative set of the plant states, and not just a single operating point. Should a novel operating state not accounted for in the distribution of $P\{\boldsymbol{x}\}$ occur, the MLR technique will make erroneous adjustments to the data because $P\{\boldsymbol{x}\}$ will be underestimated at the measurement point. The resulting rectified state will contain erroneously large adjustments necessary to increase the $P\{\boldsymbol{x}\}$ contribution to the MLR objective function, and thus $P\{\boldsymbol{y}|\boldsymbol{x}\}$ will be small. The result will be an MLR objective function value that is outside the norm, and this small value can be used as a flag to indicate that a novel operating point has been reached and that the rectified values should not be trusted. The threshold value could be easily determined by plotting a histogram of objective function values over a set of calibration cases, and determining the 95% (or similar) point of the distribution.

Sensor Model

The $P\{\boldsymbol{y}|\boldsymbol{x}\}$ term in Eq. 9 is the model of the measurement error. For a particular measurement y_i the measurement can be modeled by

$$y_i = x_i + \delta_i \tag{11}$$

where δ_i is the error. The usual case is that the error in the sensor is an additive random value, but multiplicative errors, where the size of the error is dependent on the size of the signal, are also possible (Liptak and Venczel, 1982). Equation 11 can be used to model multiplicative errors by allowing δ_i to be a function of x_i. For the case where δ_i is independent of x_i, $P\{y_i|x_i\} = P\{\delta_i\}$, which from Eq. 4 is given by

$$P\{\delta_i\} = \sum_j P\{\delta_i|m_j\}P\{m_j\}. \tag{12}$$

When a sensor is in its normal mode of operation the error term δ_i represents the sensor "noise," and is usually modeled as a zero mean Gaussian distribution. The magnitude of the noise term is reported by sensor manufacturers and contains the effects of conformity, hysteresis, dead band, and repeatability errors (Liptak and Venczel, 1982). Because the noise term is a culmination of a number of different errors when the sensor is operating normally, a Gaussian distribution is appropriate for the noise contribution to Eq. 12, as the central limit theorem states that the cumulative error will approach a Gaussian distribution as the number of error sources increase. Additional failed modes can also be characterized, such as failure to a fixed value (modeled as a delta function), and a failure to a random value (modeled as a uniform distribution). However, the exact distributions of other failure modes may be difficult to characterize. If, over a period of time, smaller gross errors are more likely than larger gross errors, the distribution of gross errors might be approximated by a Gaussian distribution whose standard deviation is greater than the standard deviation of the normal noise. Under this assumption, the total distribution of the error term δ_i can be reasonably modeled by a bivariate Gaussian:

$$P\{\delta_i\} = (1-p_i)\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{(x_i-y_i)^2}{2\sigma_i^2}\right) + p_i\frac{1}{\sqrt{2\pi}b_i\sigma_i}\exp\left(-\frac{(x_i-y_i)^2}{2b_i^2\sigma_i^2}\right). \tag{13}$$

In this expression, p_i is the probability of a gross error in sensor i, σ_i is the standard deviation of the normal noise band in sensor i, and b_i is the ratio of the standard deviation of the gross error distribution of sensor i to the standard deviation of the normal noise band for sensor i. Jeffreys (1932) was the first to apply this form of error distribution to regression of experimental data, although applications were limited due to lack of computation power. Fariss and Law (1979) presented a numerical technique for regression using the bivariate Gaussian. Tjoa and Biegler (1991) applied Eq. 13 for data rectification. As mentioned previously, the MLR approach reduces to Tjoa and Biegler's approach if Eq. 13 is used to model the sensor error, and the $P\{\boldsymbol{x}\}$ is modeled simply by constraints (binary assumption).

If one wishes to assume a minimal amount about the distribution of the gross errors, then the second term in Eq. 13 (the broad Gaussian) can be replaced with a uniform distribution. This formulation is analogous to traditional gross error detection methods in that no assumption (other than limits) is made on the distribution of the gross error. In traditional gross error detection schemes (e.g., global, nodal, and measurement test), no information about how gross errors are distributed is required. Modeling the gross error contribution to Eq. 12 using a uniform distribution corresponds to no prior knowledge about the distribution of gross errors, according to the principle of maximum entropy (Levine and Tribus, 1979). The impact of different gross error distributions on the solution to the MLR problem is explored in the "Process Examples" section of the present article.

Regardless of the choice of the individual sensor error distributions, by making the assumption that the sensor biases are independent of each other, the joint distribution of the biases is the product of the individual biases:

$$P\{\boldsymbol{\delta}\} = \prod_i P\{\delta_i\}. \tag{14}$$

This reduces the MLR objective function in Eq. 9 to

$$\begin{aligned} &\max_{\boldsymbol{x}} P\{\boldsymbol{x}\}P\{\boldsymbol{\delta}\} \\ &\text{s.t.} \quad h(\boldsymbol{x}) = 0 \\ &\qquad g(\boldsymbol{x}) \le 0. \end{aligned} \tag{15}$$

The MLR objective function for additive errors has two terms: $P\{x\}$ is the probability of the rectified state; and $P\{\delta\}$ is the probability of the adjustments made to the measurements. Thus there is a tradeoff between likely rectified states and likely adjustments. The optimization seeks a solution that not only ends at a probable rectified state, but also is a probable adjustment to the measurement set.

Data Rectification and Robust Regression

In this section, we explore the analogy between maximum likelihood rectification and robust regression, a technique used in parameter estimation when the data contain outliers. In steady-state data rectification, the measurement y represent a single point in the measurement space and the goal is to find the point x that lies on a predefined surface (the surface described by the model constraints), that is most likely given y. In regression, there are a large number of points in the measurement space, and the goal is to find a set of parameters $\{\hat{\beta}\}$ that yield the most likely description of the surface (the parameterized model), given the measurements and the model form. Rectification fits a point to a surface and regression fits a surface to a set of points. The maximum likelihood regression problem is posed as

$$\begin{aligned} &\max_{\hat{\beta}} \rho(y, x, \hat{\beta}) \\ \text{s.t.} \quad &y - f(x, \hat{\beta}) = 0 \\ &g(\hat{\beta}) \geq 0, \end{aligned} \tag{16}$$

where ρ is the objective function to be maximized. The data rectification problem (assuming $P\{x\}$ is modeled by a set of constraints) is posed as

$$\begin{aligned} &\max_{\hat{x}} \rho(y, \hat{x}) \\ \text{s.t.} \quad &f(\hat{x}, \beta) = 0 \\ &g(\hat{x}, \beta) \geq 0 \end{aligned} \tag{17}$$

where the model parameters (β) are given.

In ordinary regression or data rectification, weighted least squares is the objective function. For clarity of presentation the regression and rectification problems are posed here assuming that the noise covariance matrices are diagonal. This is not a necessary assumption, but it greatly simplifies the notation. For both problems the weighted least-squares objective function is

$$\rho(z) = \sum_{i=1}^{n} \frac{1}{2} z_i^2 \tag{18}$$

where for regression,

$$z_i = \frac{y_i - \hat{\beta} x_i}{\sigma_i} \tag{19}$$

and for rectification,

$$z_i = \frac{y_i - \hat{x}_i}{\sigma_i}. \tag{20}$$

For the regression case, the z_i are summed over the number of data points, and in the rectification case the z_i are summed over the dimension of the measurement space.

The effect of outliers on the solution of regression problems has been extensively studied in the statistics literature. Hampel (1974) uses the concept of an influence curve to study ordinary and robust regression. The influence of data point i is defined as the derivative of the objective function with respect to the error term z_i:

$$I_i(z_i) = \frac{d\rho}{dz_i}. \tag{21}$$

The influence curve defines the region in which a measurement has an effect on the estimation of the parameters. For weighted least-squares regression, from Eq. 18, the influence curve is $I_i = z_i$. Thus the influence of an error on the estimation of the parameters is unbounded (Cook and Weisberg, 1982). This means that as the error in the ith measurement grows it exerts an increasing influence on the estimated parameters, even though at some level of error it becomes likely that the data point is an outlier. Regressing a data set containing outliers with least squares will yield poorly fit models because the outliers dominate the other data.

The concept of robust regression was developed to negate or limit the effect of outliers on the regressed parameters. Robust regression techniques use probability distribution of errors that have longer tails, thus making large adjustments in some measurements more probable. One such distribution used for robust estimation is the Lorentzian distribution (Huber, 1981)

$$\rho(z) = \sum_{i=1}^{n} \frac{1}{1 + \frac{1}{2} z_i^2}. \tag{22}$$

Figure 2 shows the influence curve for the Lorentzian objective function. The influence of an error in a measurement increases as the error grows from zero, but then decreases and eventually becomes zero once the error becomes large.

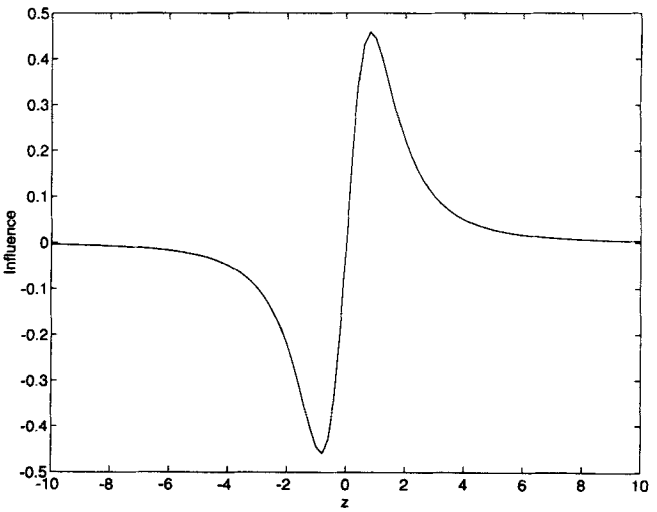

Figure 2. Influence for Lorentzian distribution of errors.

This is a desirable form of the influence curve because severely outlying points are discounted, rather than emphasized as in least squares. The formulation in Eq. 4 using multiple modes of sensor operation achieves a similar broadening of the tails of the distribution while appealing to more fundamental considerations to motivate the form of the distribution (Kramer and Mah, 1993). Thus the distribution of errors given in Eq. 4 could be interpreted as a robust regression probability function, as the following example illustrates.

Example 2. This example is a simple linear regression between one dependent variable (y) and one independent variable (x) with an intercept at the origin,

$$y = \beta x.$$

The data set contains 20 points drawn from a uniform distribution between $x = 0$ and $x = 1$. The y points were generated by adding Gaussian noise

$$y = x + \epsilon, \qquad \epsilon = N(0, 0.1).$$

The $\hat{\beta}$ parameter was regressed using ordinary least-squares regression (Eq. 16) and using the robust bivariate sensor model. For the robust regression, the objective function is

$$\rho(y, x, \hat{\beta}) = \prod_{i=1}^{20} (1-p)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(y_i - \hat{\beta}x_i)^2}{2\sigma^2}\right) + p\frac{1}{\sqrt{2\pi}b\sigma}\exp\left(-\frac{(y-\hat{\beta}x)^2}{2b^2\sigma^2}\right).$$

For this case $\sigma = 0.1$, $p = 0.05$, and $b = 20$.

The estimated $\hat{\beta}$ for both the ordinary and robust regression were calculated as a function of the error in one data point (at $x = 0.341$), by adding errors ranging from -5σ to $+5\sigma$ to the corresponding y value, and maximizing ρ. The results are shown in Figure 3. The compound objective function clearly has the desired effect of discounting outliers, in a manner qualitatively equivalent to the Lorentzian distribution. Errors greater than $\pm 4\sigma$ effectively have no influence on the estimated slope. Errors smaller than $\pm 4\sigma$ are not considered outliers and thus do have an influence on the solution.

For data rectification it is desirable to discount the measurements that contain gross errors, in a manner analogous to outliers in robust regression. The sensor mode model (Eq. 4) gives the same desirable shape of influence curve, as in robust regression, for rectification problems.

Example 3. Consider a process stream whose temperature is measured by three thermocouples. The model equations are $T_1 = T_2 = T_3$. Each measurement of the true temperature T is given by

$$T_i = T + \delta_i,$$

where δ_i is the error in the ith measurement. For simplicity assume the true stream temperature varies uniformly between 50°C and 150°C. Thus $P\{T\}$ is a constant, and the MLR objective function reduces to

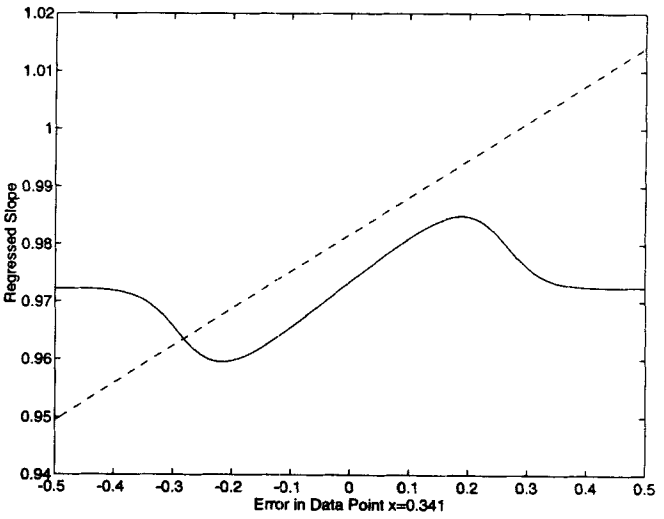

Figure 3. Influence for weighted least squares (---) and robust regression using a bivariate Gaussian error distribution (——).

$$\max_{\hat{T}} \prod_{i=1}^{3} P\{T_i|\hat{T}\}$$

$$\text{s.t.} \qquad 50 \le \hat{T} \le 150.$$

Note that $P\{T_i|\hat{T}\}$ is analogous to $P\{y_i|x_i\}$ where all of the x_i are equal because of the model constraints. The compound objective function in Eq. 13 is used with $p = 0.05$, $b = 20$, and $\sigma = 1$. The nominal measurement point was $T_1 = 100$, $T_2 = 102$, $T_3 = 100$. The error in T_3 was systematically varied from -10°C to $+10$°C and the rectified temperature for each of the error levels is plotted against the error as a solid line in Figure 4. Again the influence curve shows that large errors in T_3 have no effect on the rectified temperature.

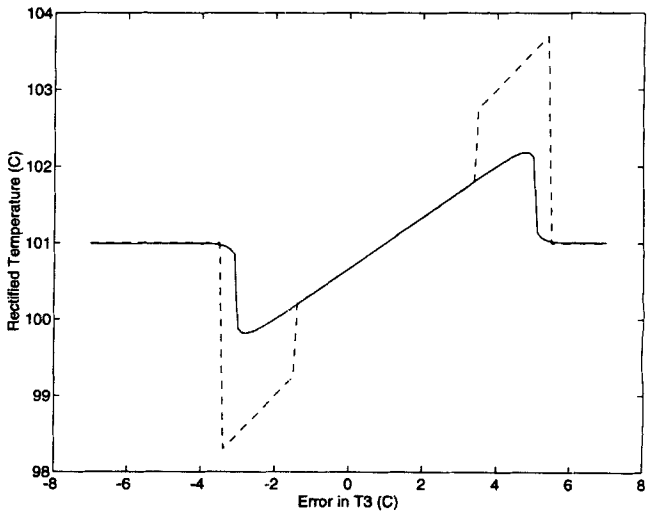

Figure 4. Influence for MLR (——) and traditional linear reconciliation with gross error detection and removal using the global test and serial elimination (---), for Example 3.

We can also compare robust rectification to traditional rectification with auxiliary gross error detection and removal. The dashed line in Figure 4 shows the influence curve for weighted least-squares rectification using the global test at a 0.05 level of significance for gross error detection, and serial elimination for gross error identification. Qualitatively, the effect of gross error detection and removal are the same as robust rectification, for this example. The discontinuities in the shape of the influence curve are due to incorrect identification of a gross error by the serial elimination procedure, for certain levels of error in T_3.

The foregoing example shows that the sensor mode model in Eq. 4 gives the desired influence curve, given a fully specified model and no prior distribution of the underlying process states. We will now show that the MLR objective function gives a similar influence curve when there is a known prior distribution of the states, but the lack of constraints precludes the use of conventional gross error approaches.

Example 4. Consider again the problem in Example 1, where a stream is measured by two thermocouples, and it is known from historical data that the true temperature is distributed normally with a mean of 50°C and a standard deviation $\sigma_T = 10$°C. The MLR objective function for this data-rectification problem is

$$\max_{\hat{T}} \frac{1}{\sigma_T\sqrt{2\pi}} \exp\left(-\frac{(\hat{T}-50)^2}{2\sigma_T^2}\right) \prod_{i=1}^{2}\left[(1-p)\frac{1}{\sigma\sqrt{2\pi}} \times \exp\left(-\frac{(\hat{T}-T_i)^2}{2\sigma^2}\right) + p\frac{1}{b\sigma\sqrt{2\pi}} \exp\left(-\frac{(\hat{T}-T_i)^2}{2b^2\sigma^2}\right)\right]$$

where the measurement errors have been modeled using a bivariate Gaussian (Eq. 13) with $p = 0.05$, $b = 20$, and $\sigma =$ 1°C. With T_1 fixed at 50°C, T_2 was varied from 25°C to 75°C, and the rectified temperature was plotted as a function of the error in T_2 as a solid line in Figure 5. An influence curve that gives little or no weight to measurements with large errors is again observed. The dashed line in Figure 5 is the rectified temperature using weighted least-squares rectification. With only two measurements, gross error tests can be done to signal that a gross error has occurred, but no conventional procedure can identify which sensor has the gross error. Tjoa and Biegler's approach of maximizing the robust objective function (Eq. 13) subject to the $T_1 = T_2$ constraint, without using the distribution of $P\{T\}$ (assuming $P\{T\}$ is uniform and thus equals a constant), also yields the unbounded influence curve shown by the dashed line in Figure 5. The prior distribution of the temperature is required to make correct inferences about which sensor has failed and automatically remove it from the rectification process by giving it zero weight in the solution.

Probability Bootstrapping

The preceding example demonstrates the potential importance of modeling the prior state distribution. $P\{x\}$ captures the physical constraints of the process, such as mass balances, energy balances, thermodynamic relationships, and transport relationships, as well as the statistical tendencies in operation

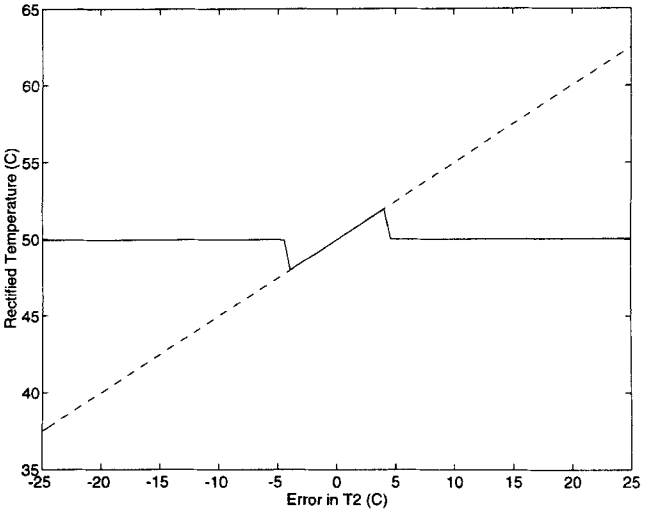

Figure 5. Influence for MLR (——) and traditional linear reconciliation (---), for Example 1.

of the plant (e.g., the distribution over time of setpoints). In this section, we propose a novel technique for estimating $P\{x\}$ from historical plant data, which we call the probability bootstrapping technique.

The historical data that are used to find $P\{x\}$ are termed the calibration set. This set is a matrix of measurements Y, with K rows of N-dimensional measurement vectors y. At the calibration stage, before $P\{x\}$ is known, we are looking for a matrix of the most probable values X given the matrix of measurements Y. Thus we are maximizing $P\{X|Y\} = P\{Y|X\}P\{X\}/P\{Y\}$ as opposed to finding the solutions for an individual x vector from an individual y vector as in Eq. 9. The problem becomes

$$\max_{\hat{x},\beta} P\{\hat{X}\}P\{Y|\hat{X}\}$$
$$\text{s.t.} \quad h(\hat{X}) = 0$$
$$g(\hat{X}) \le 0 \tag{23}$$

where $\hat{X}$ is the estimate of X, and β are the parameters of the $P\{\hat{X}\}$ distribution (means and covariance matrix elements if $P\{\hat{X}\}$ is modeled as being Gaussian). Equation 23 simultaneously estimates the collective probability distribution $P\{\hat{X}\}$ and rectifies the K measurement vectors in Y. (It should be noted that $P\{\hat{X}\}$ is the same as $P\{\hat{x}\}$, but $P\{\hat{X}\}$ is used to clearly state that the density estimate is estimated from the data set contained in the rows of matrix $\hat{X}$.)

At the single measurement level, we still wish to find the individual rectified values that maximize $P\{x|y\} = P\{y|x\}P\{x\}/P\{y\}$, but the parameters of $P\{x\}$ at this point are unknown, and depend on all K vectors in the calibration data set. In the context of the steady-state plant, the measurement vectors are independent of each other, and thus to form the joint probability of $P\{x_1, x_2, \ldots, x_K|y_1, y_2, \ldots, y_K\} = P\{X|Y\}$, the individual probabilities $P\{y|x\}$ are multiplied together:

$$P\{X|Y\} = \prod_{k=1}^{K} P\{x_k|y_k\}. \tag{24}$$

Using Eq. 24, the calibration data set can be used to solve the nonlinear optimization problem stated in Eq. 23.

Example 5. Assume a Gaussian distribution of actual x values with zero mean and unit standard deviation. The measurements are given by $y = x + \delta$, where δ is zero-centered Gaussian with standard deviation σ_d. We assume the distribution of x to be normal with unknown mean μ, and unknown standard deviation σ to be determined from a calibration data set Y. There are no model constraints imposed on the system. In this case, Eq. 23 becomes

$$\max_{\hat{X},\mu,\sigma} P\{\hat{X}\}P\{Y|\hat{X}\}$$

$$= \max_{\hat{X},\mu,\sigma} \prod_{k=1}^{K} P\{\hat{x}_k;\mu,\sigma\}P\{\delta_k\}$$

$$= \max_{\hat{X},\mu,\sigma} \prod_{k=1}^{K} \left[\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_k-\mu)^2}{2\sigma^2} \right) * \frac{1}{\sqrt{2\pi}\sigma_d} \exp\left(-\frac{(x_k-y_k)^2}{2\sigma_d^2} \right) \right]$$

$$= \max_{\hat{X},\mu,\sigma} \sum_{k=1}^{K} \left[\ln\left(\frac{1}{2\pi\sigma\sigma_d} \right) - \frac{1}{2}\left(\frac{(x_k-\mu)^2}{\sigma^2} + \frac{(x_k-y_k)^2}{\sigma_d^2} \right) \right].$$

The solution of this equation can be obtained by nonlinear optimization with the elements of $\hat{X}$, μ, and σ as the decision variables. Because μ and σ are functions of $\hat{X}$, one can also substitute in the formulas for the mean and standard deviation of $\hat{X}$ and optimize using only $\hat{X}$ as the decision variables. These two techniques converge to the same value of μ and σ as the number of data points increase.

A simulation was performed in which the 750 x values were drawn from a Gaussian distribution with mean 0 and standard deviation 1. The y values were calculated by adding noise from a zero-centered Gaussian distribution with standard deviation σ_d. The preceding equation was used to find σ. Figure 6 shows the average absolute value over 100 independent test cases of the difference between the estimated standard deviation and the actual standard deviation as a function of the noise standard deviation used to corrupt the states. The results show that the technique effectively estimates the standard deviation of the states. Typically one can expect the ratio of noise standard deviation to signal standard deviation in the range of 0.05 to 0.2, and in this range the estimated standard deviation differs from the actual standard deviation by less than 2%. Even when the noise contribution is half the signal contribution ($\sigma_d = 0.5$), the solution yielded errors of only 20%.

Example 5 shows how the simultaneous solution technique for finding the probability distribution of the states can be applied for a one-dimensional distribution. In the general case of multidimensional data with the form of the probability distribution unknown, solving Eq. 23 becomes a very difficult numerical problem. The difficulty arises from having to simultaneously estimate the distribution of $\hat{X}$ using a nonparametric density estimation technique, which depends heavily on the data, and optimizing the objective function in Eq. 23. The following method of probability bootstrapping is suggested as a way to circumvent simultaneous density estimation and optimization.

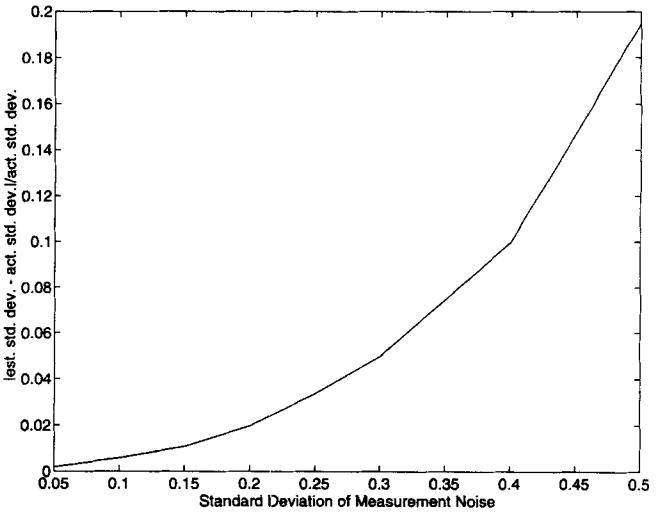

Figure 6. Absolute difference between the standard deviation of the x data and the standard deviation estimated from the y data using the probability bootstrapping technique.

Figure 7 shows the probability bootstrapping technique. The essence of the methodology is that it decouples the density estimation problem from the optimization problem. The initial guess of the rectified states $\hat{X}$ is the measurements in the calibration set Y. The probability distribution of $\hat{X}$ is then estimated using a parametric or nonparametric density estimation technique. Johnston and Kramer (1994) present a nonparametric elliptical basis function density estimation technique, as well as a method based on cross-validation and maximum likelihood for finding the probability density estimator that best fits a particular set of data. This is a suitable technique if the form of the distribution of states is unknown. The resulting probability distribution is then used in the MLR to find a new set of rectified vectors $\hat{X}$, solving for each of the $\hat{x}$ vectors one at a time using Eq. 9. The probability distribution of this set of rectified vectors is then found by repeating the density estimation, and then another set of recti-

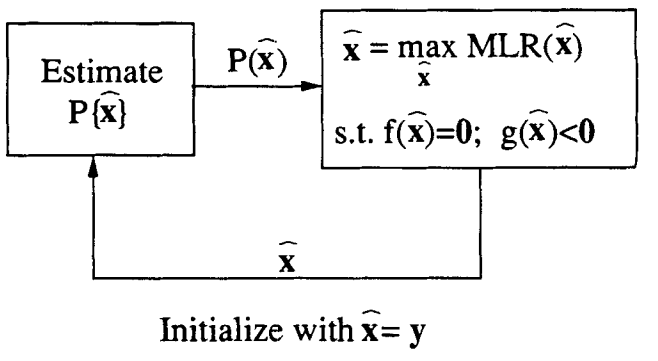

Figure 7. Bootstrapping method to find $P\{x\}$, by making the approximation that $P\{x\} = P\{\hat{x}\}$.

fied vectors is found using MLR. This loop continues until there is no more change in any of the $\hat{x}$ vectors on subsequent iterations. The bootstrapping technique repeatedly inserts the $\hat{X}$ matrix into the objective function given in Eq. 23, and thus the final $\hat{X}$ matrix reached through the bootstrapping should be the same as the $\hat{X}$ matrix reached through the simultaneous approach. By decoupling the density estimation and optimization, the bootstrapping technique allows for more flexible forms of the probability distribution of the states to be used, without the need for solving a very complex nonlinear optimization.

The problem given in Example 5 was used to compare the bootstrapping and simultaneous solution approaches. Because the form of the distribution in the example is given as Gaussian, the simultaneous solution can be achieved by optimizing over $K+2$ variables (the K measurements, the mean μ, and the standard deviation σ). Data sets containing 750 measurement vectors were used to find μ and σ by the simultaneous and bootstrapping techniques. In 100 trials with values of the noise standard deviation (σ_d) ranging from 0.05 to 0.5, the value of σ found by the two techniques differed by an average of 0.5%, and the standard deviation of the difference was 0.5%. The means found by each method were essentially identical. Thus for this one-dimensional example the bootstrapping method arrives at the same distribution as the simultaneous solution method.

In several computational examples, the probability bootstrapping method has been found to offer a practical way to estimate $P\{X\}$ from a data set corrupted with sensor errors, including gross errors. Once $P\{X\}$ has been found, it can be used to rectify future measurements by using the MLR technique, unless there is a change in the operation of the plant that was not included in the calibration data set. If this occurs, then $P\{X\}$ can be reevaluated with the new data, and then used for rectification of future measurements using MLR.

Process Examples

In this section, two chemical engineering examples are presented to demonstrate data rectification by the MLR technique. The first is a flow system from Mah (1987), and the second example is a heat exchanger network that was taken, with some modification, from Tjoa and Biegler (1991). The performance criterion used to gauge the efficacy of the rectification is the mean-squared error (MSE)

$$\text{MSE} = \frac{1}{N^* K} \sum_{k=1}^{K} \sum_{n=1}^{N} \frac{(\hat{x}_{n,k} - x_{n,k})^2}{\sigma_n^2}, \qquad (25)$$

where N the dimension of the measurement vector, K is the number of measurement vectors used to calculate the MSE, $\hat{x}$ is the rectified state, x is the true plant state, and σ_n is the standard deviation of the Gaussian noise on sensor n. The MSE is structured such that if there are no sensor failures (the measurements only contain random Gaussian noise with standard deviations σ_n), then the MSE = 1.0 (for $\hat{x} = y$). If there are gross errors, the MSE can be greater than 1.0. If the rectification is perfect (the true state of the plant was found for every measurement vector) then the MSE = 0.0. Thus the goal is to drive the MSE toward zero.

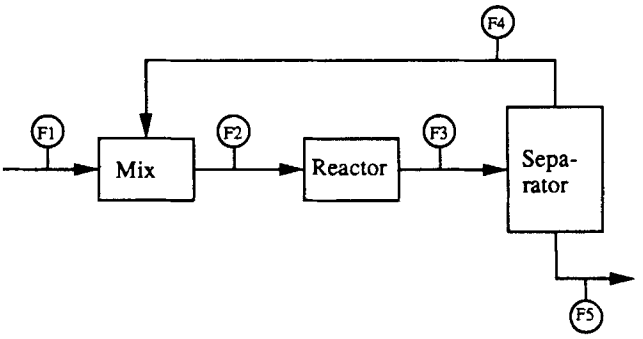

Figure 8. Flow network used in data rectification.

All of the sensors in both examples are assumed to have a bivariate Gaussian distribution (Eq. 13). For all sensors $p_i = 0.05$, and $b_i = 20$.

Figure 8 shows the flow network from Mah (1987). The total flow of each stream is measured. Data were generated by sampling F_1 from $U(15,40)$, calculating the remaining flows, and generating the y values by adding random, zero-centered, Gaussian noise with standard deviations 1.0, 4.0, 4.0, 3.0, and 1.0, respectively, for the five sensors. A training data set of 100 points was generated to estimate $P\{x\}$ using the bootstrapping procedure described previously, and a separate 1,000-point test set was also generated for the performance studies. A multivariate Gaussian distribution was used for $P\{x\}$, which, in spite of the mismatch from the actual distribution, was found to be the best estimator for this data set using cross-validation (see Johnston and Kramer, 1994). Another 1,000-point data set was also created, and in this set a failure of the F_3 sensor was simulated by adding a bias sampled from $N(0,20)$.

Although this example involves simple linear balance equations, the example is used to demonstrate how the number of known constraints affects the data-rectification process. There are three independent mass balances that can be written for this system. By assuming that only a subset of these balances is known, we can demonstrate how different degrees of redundancy affect the rectification. In general, the more relationships that are known, the better the data rectification will be. Table 1 shows the MSE results for the MLR method as well as the traditional linear weighted least-squares method as a function of the number of constraints used. In the case of simulated failures in F_3, the global test and serial elimination were used to detect and remove gross errors before reconciling the data. As the number of constraints increases, the MSE is driven lower for both the linear reconciliation and

Table 1. MSE Results for Rectification of Flow Example Data, $P\{x\}$ Calibrated from Normal Data Only

	No Failures (mse = 1.02)		Failure of F3 (mse = 19.93)	
Number of Constraints	Linear	MLR	Linear*	MLR
0	—	0.431	—	0.485
1	0.793	0.392	—	0.426
2	0.597	0.337	3.51	0.362
3	0.408	0.263	1.56	0.276

*With serial elimination

the MLR method. In all cases the MLR method outperforms the linear method. In the case of the failure of F_3 the linear reconciliation technique, including serial elimination, is unable to satisfactorily remove the errors from the measurements (MSE > 1.0). However, the results show that even in the complete absence of known process constraints, the data are rectified reasonably well by the MLR method. The MSE is reduced from 1.02 to 0.431 in the case of no gross error, and from 19.93 to 0.485 for simulated failure in F_3. The capture of the physical constraints between the process variables in $P\{x\}$ allows the MLR method to rectify the data in the absence of analytical constraints, and the added information about the probability distribution of the rectified states causes the MLR method to outperform the linear reconciliation even when constraints are used.

Figure 9 shows a heat-exchanger network that was modified from Tjoa and Biegler (1991). Process stream A is heated by process stream B in two exchangers, and with utility stream C in one exchanger. The flow of stream D is adjusted so that the outlet temperature of stream A, TA8, will be 615°C. The plant operates at low-, medium-, and high-capacity levels where process streams A and B are at low, medium, and high flow rates. The temperatures of all streams are at constant values, with some process noise. The operating conditions are summarized in Figure 9. Using these operating conditions, a training set of 120 points was generated, from which $P\{x\}$ was estimated using the bootstrapping technique. The noise added to the training and test sets had a standard deviation of 0.75 for the temperature measurements, and standard deviations of 4.0, 2.0, 3.0, and 3.0 for flows FA1, FB1, FC1, and FD1, respectively. In this case the density estimator that best fit the data was an elliptical basis function estimator with five units and overlap of 1 (see Johnston and Kramer, 1994). Three separate test sets were created for performance

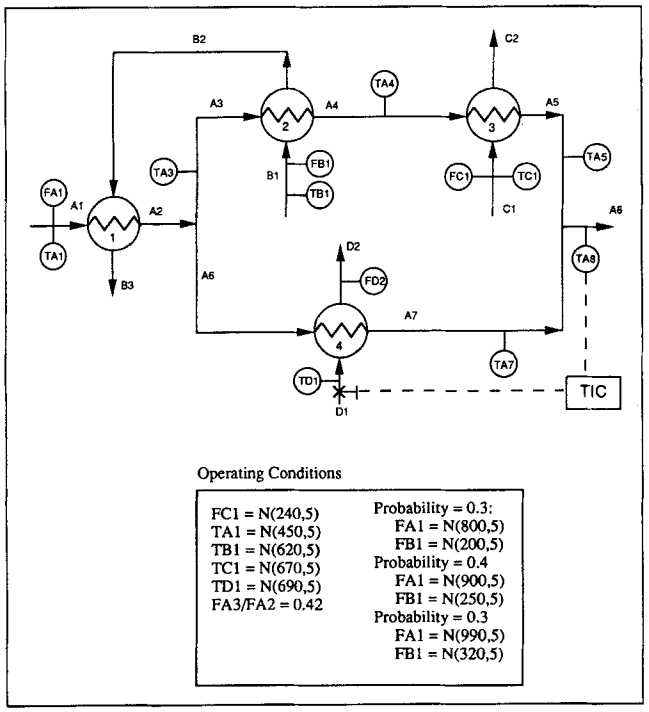

Figure 9. Heat-exchanger network example.

Table 2. MSE Results for Rectification of Heat-Exchanger Network Data by MLR Technique

	No Failures	Failure of TA4	Failure of TA4 and TB1
Unrectified	1.016	10.31	20.62
Rectified	0.527	0.724	0.864

testing. One set had only Gaussian noise in the sensors. Another set simulated a failure in sensor TA4, by adding a bias from a $N(0,10)$ distribution, and the final test set simulated a failure in both TA4 and TB1 by adding biases drawn from a $N(0,10)$ distribution.

The results are shown in Table 2. The MLR was done without any constraints because the instrumentation on the network does not allow for any mass or energy balances to be written. Again the MLR shows its power in estimating the state of the plant, without any prior constraints. The MSE was reduced from 1.0 to 0.53, from 10.31 to 0.724, and from 20.62 to 0.864, for the three data sets. Even when two sensors had failed, the MLR was able to find the rectified values and lower the MSE below the random noise level of 1.0. The results show that even in a 13-dimensional problem the MLR is capable of finding the rectified values without any prior process model. The addition of constraints would improve the performance, as shown in the previous example.

In the foregoing examples, the $P\{x\}$ was generated from data that contained no sensor failures. In general there could be some sensor failures in the data set used to calibrate $P\{x\}$. The probability bootstrapping technique is capable of estimating $P\{x\}$ accurately, even when some sensor failures are present in the calibration data. The data rectification for the flow network shown in Figure 8 was also performed when 10% of the 100-point training set contained F_3 measurements from a failed sensor. The results for the MLR rectification are shown in Table 3, and compared to the case where there were no sensor failures in the calibration data used to find $P\{x\}$. The bootstrapping methodology removed the sensor failures in the calibration data, and the $P\{x\}$ that was finally arrived at was essentially the same as the $P\{x\}$ found from the data set without the sensor failures. The performance of the MLR was essentially unchanged.

The effect of different gross error distributions on the MLR solution was also investigated. The MLR of data from the flow network were calculated for different values of b and p, the parameters of the bivariate Gaussian sensor model (Eq. 13). Separate 1,000-point test sets, simulated the same way as the data used for the results in Table 1, were used and the

Table 3. MSE Results for Rectification of Flow Example Data, $P\{x\}$ Calibrated from Data Corrupted with 10% Gross Errors

	No Failures (mse = 1.02)	Failure of $F3$ (mse = 19.93)
Number of Constraints	MLR	MLR
0	0.433	0.490
1	0.395	0.428
2	0.338	0.365
3	0.263	0.277

Table 4. MSE Values for Data Rectified Using the MLR Method with No Constraints, for Various Combinations of b and p Parameters, for 1,000-Point Data Set Containing No Gross Errors

	b				
p	5	10	15	20	25
0.02	0.401	0.402	0.402	0.403	0.403
0.05	0.400	0.401	0.402	0.402	0.402
0.10	0.397	0.400	0.401	0.401	0.402
0.15	0.392	0.398	0.400	0.400	0.401
0.20	0.388	0.395	0.397	0.399	0.400

Table 5. MSE Values for Data Rectified Using the MLR Method with No Constraints, for Various Combinations of b and p Parameters, for 1,000-Point Data Set Containing Gross Errors in Sensor F3

	b				
p	5	10	15	20	25
0.02	0.583	0.578	0.569	0.564	0.571
0.05	0.529	0.531	0.524	0.515	0.523
0.10	0.511	0.501	0.502	0.495	0.503
0.15	0.482	0.476	0.465	0.468	0.479
0.20	0.436	0.441	0.444	0.452	0.463

mean-square error for the different combinations of b and p are shown in Tables 4 and 5. The MLR was done without analytic constraints. The solution of the MLR is not sensitive to the choice of the probability of a gross error (p) or to the width of the gross error distribution (b) for the ranges investigated.

Conclusions

A new method for data rectification for steady-state systems, based on maximum likelihood estimation, has been introduced and demonstrated. The maximum likelihood rectification (MLR) method is a generalized framework that allows for various sensor behaviors and any number of process model constraints, including no constraints. The method uses historical plant data to find the prior probability distribution of the true plant states, using a probability bootstrapping methodology. This methodology is robust enough to find the distribution of the underlying states from historical data, even when there are measurement biases (gross errors) in the historical data. The MLR technique was demonstrated on two examples, and was able to obtain good estimates of the true plant state, even when no process model was used in the rectification. This constitutes a new class of problem that can now be solved.

Literature Cited

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

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

Cook, R. D., and S. Weisberg, *Residuals and Influence in Regression*, Chapman & Hall, New York (1982).

Crowe, C. M., Y. A. Garcia-Campos, and A. Hrymak, "Reconciliation of Process Flow Rates by Matrix Projection: I. Linear Case," *AIChE J.*, **29**, 881 (1983).

Fariss, R. H., and V. H. Law, "An Efficient Computational Technique for Generalized Application of Maximum Likelihood to Improve Correlation of Experimental Data," *Comput. Chem. Eng.*, **3**, 95 (1979).

Hampel, F. R., "The Influence Curve and Its Role in Robust Estimation," *J. Amer. Stat. Assoc.*, **69**, 383 (1974).

Hlavacek, V., "Analysis of Complex Plant—Steady-State and Transient Behavior," *Comput. Chem. Eng.*, **1**, 75 (1977).

Huber, P. J., *Robust Statistics*, Wiley, New York (1981).

Iordache, C., R. S. H. Mah, and A. C. Tamhane, "Performance Studies of the Measurement Test for Detection of Gross Errors in Process Data," *AIChE J.*, **31**, 1187 (1985).

Jeffreys, H., "An Alternative to the Rejection of Observations," *Proc. R. Soc. (London)*, **A137**, 78 (1932).

Johnston, L. P. M., and M. A. Kramer, "Probability Density Estimation Using Elliptical Basis Functions," *AIChE J.*, **40**, 1639 (1994).

Kramer, M. A., and R. S. H. Mah, "Model-Based Monitoring," Foundations of Computer Aided Process Operations, Crested Butte, CO (1993).

Kuehn, D. R., and H. Davidson, "Computer Control II: Mathematics of Control," *Chem. Eng. Prog.*, **57**, 44 (1961).

Levine, R. D., and M. Tribus, eds., *The Maximum Entropy Formalism*, MIT Press, Cambridge, MA (1979).

Liptak, B. G., and K. Venczel, *Instrument Engineers Handbook*, Chilton, Randor, PA (1982).

Mah, R. S. H., "Data Screening," *Foundations of Computer Aided Process Operations*, G. V. Reklaitus and H. D. Spriggs, eds., Elsevier, New York (1987).

Mah, R. S. H., and A. C. Tamhane, "Detection of Gross Errors in Process Data," *AIChE J.*, **28**, 828 (1982).

Nogita, S., "Statistical Test and Adjustment of Process Data," *Ind. Eng. Chem. Proc. Des. Dev.*, **11**, 197 (1972).

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

Ricker, N. L., "Comparison of Methods for Nonlinear Parameter Estimation," *Ind. Eng. Chem. Proc. Des. Dev.*, **23**, 283 (1984).

Ripps, D. L., "Adjustment of Experimental Data," *Chem. Eng. Prog. Symp. Ser.*, **61**, 8 (1965).

Rollins, D. K., and J. F. Davis, "Unbiased Estimation of Gross Errors in Process Measurements," *AIChE J.*, **38**, 563 (1992).

Rollins, D. K., and S. D. Roelfs, "Gross Error Detection When Constraints are Bilinear," *AIChE J.*, **38**, 1295 (1992).

Rosenberg, J., R. S. H. Mah, and C. Iordache, "Evaluation of Schemes for Detecting and Identifying Gross Errors in Process Data," *Ind. Eng. Chem. Res.*, **26**, 555 (1987).

Tamhane, A. C., and R. S. H. Mah, "Data Reconciliation and Gross Error Detection in Chemical Process Networks," *Technometrics*, **27**, 409 (1985).

Tjoa, I. B., and L. T. Biegler, "Simultaneous Strategies for Data Reconciliation and Gross Error Detection of Nonlinear Systems," *Comput. Chem. Eng.*, **15**, 679 (1991).

Tjoa, I. B., and L. T. Biegler, "Reduced Successive Quadratic Programming Strategy for Errors in Variables Estimation," *Comput. Chem. Eng.*, **16**, 525 (1992).

Manuscript received July 29, 1994, and revision received Dec. 2, 1994.